ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΚΑΙ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Πιθανοτικές προβλέψεις σε εφαρμογές παραγωγής και κατανάλωσης ενέργειας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Φώτιος Καραμπλιάς

Επιβλέπων: Βασίλειος Ασημακόπουλος

Καθηγητής Ε.Μ.Π.

Υπεύθυνος: Ευάγγελος Σπηλιώτης

Διδάκτωρ Ε.Μ.Π

Αθήνα, Σεπτέμβριος 2019

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΚΑΙ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Πιθανοτικές προβλέψεις σε εφαρμογές παραγωγής και κατανάλωσης ενέργειας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Φώτιος Καραμπλιάς

Επιβλέπων: Βασίλειος Ασημακόπουλος

Καθηγητής Ε.Μ.Π.

Υπεύθυνος: Ευάγγελος Σπηλιώτης

Διδάκτωρ Ε.Μ.Π

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 3η Οκτωβρίου 2019.

_____	_____	_____
Βασίλειος Ασημακόπουλος	Ιωάννης Ψαρράς	Δημήτριος Ασκούνης
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

Αθήνα, Σεπτέμβριος 2019

Φώτιος Ε. Καραμπλιάς

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Περίληψη

Στόχος της παρούσας εργασίας είναι η ανάπτυξη μιας μεθοδολογίας η οποία θα προβλέπει με ακρίβεια μέσω πιθανοτικών μοντέλων την ηλεκτρική ενέργεια που παράγεται από εγκαταστάσεις που αξιοποιούν Ανανεώσιμες Πηγές Ενέργειας (ΑΠΕ) και πιο συγκεκριμένα την ηλιακή. Το πρόβλημα της πρόβλεψης ενέργειας που προέρχεται από ΑΠΕ αντιμετωπίζεται σήμερα σε ικανοποιητικό βαθμό από αρκετά μοντέλα, τα οποία ωστόσο περιορίζονται συνήθως στην εξαγωγή σημειακών προβλέψεων. Ο εν λόγω περιορισμός αποτελεί σημαντικό πρόβλημα για αρκετές εφαρμογές της αγοράς ενέργειας καθώς δεν επιτρέπει στον αποφασίζοντα να ενημερώνεται σχετικά με το αναμενόμενο εύρος της παραγωγής, πράγμα απαραίτητο σε περιπτώσεις όπου απαιτείται να εκτιμηθούν σχετικά ρίσκα και να μελετηθούν οι αντίστοιχες επιπτώσεις τους.

Πιο συγκεκριμένα, η παρούσα διπλωματική καταπιάνεται με την εκτίμηση της εμπειρικής συνάρτησης πυκνότητας πιθανότητας της παραγόμενης ηλεκτρικής ενέργειας ενός ελληνικού φωτοβολταϊκών πάρκου χρησιμοποιώντας ένα σύνολο από διαθέσιμα παρελθοντικά δεδομένα. Η εν λόγω συνάρτηση αξιοποιείται για την ανάπτυξη μίας στοχαστικής μεθόδου πρόβλεψης η οποία μας δίνει τη δυνατότητα να εκτιμούμε τον κίνδυνο η εγκατάσταση να παρουσιάσει χαμηλά επίπεδα παραγωγής ή να ξεπεράσει κάποια άλλα, την πιθανότητα δηλαδή να εμφανιστεί ρίσκο.

Αρχικά εισάγεται η έννοια της εκτιμήτριας συναρτήσεων πυκνότητας πιθανότητας και η σημασία των στοχαστικών προβλέψεων στην παραγωγή ενέργειας. Ακολουθεί μία σύντομη βιβλιογραφική επισκόπηση σχετικά με τον τρόπο που παράγονται οι συγκεκριμένες προβλέψεις, καθώς και για το ποιοι δείκτες σφάλματος αξιοποιούνται συνήθως για την αξιολόγηση της ακρίβειάς τους. Στη συνέχεια παρουσιάζεται η μεθοδολογία που αναπτύχθηκε για την επεξεργασία των δεδομένων και την παραγωγή προβλέψεων, ενώ γίνεται και αναφορά στη γλώσσα προγραμματισμού R που χρησιμοποιήθηκε για την εξαγωγή των αποτελεσμάτων. Στο τελευταίο κομμάτι της εργασίας εξετάζεται η ακρίβεια της προτεινόμενης μεθοδολογίας και συγκρίνεται με αυτήν άλλων γνωστών εναλλακτικών, παραθέτοντας σχετικά συμπεράσματα και μελλοντικές προεκτάσεις.

Λέξεις κλειδιά: Τεχνικές Προβλέψεων, Πιθανότητες, Ανανεώσιμες Πηγές Ενέργειας, Ηλεκτρική Ενέργεια, Φωτοβολταϊκά Συστήματα

Abstract

The aim of the present study is to develop a methodology that accurately predicts through the use of probabilistic forecast methods the electric energy produced by installations using Renewable Energy Sources (RES) and more specifically solar energy. The problem of forecasting energy produced from RES is currently being well addressed by several forecasting methods, which however are usually limited to point forecasts. This limitation is a major problem for several applications of the energy market as it does not allow decision makers to be informed about the expected range of production, which is necessary in cases where it is necessary to assess relative risks and to study their impacts.

More specifically, this thesis deals with the estimation of the empirical probability density function of the generated electric power of a Greek photovoltaic park using a set of available past data. This function is used to develop a predictive method that allows us to evaluate the risk that the installation will exhibit low levels of production or overcome some others, in other words the likelihood of risk occurring.

Initially, we introduced the concept of probability density estimator and the importance of stochastic predictions in energy production. Then we made a brief bibliographic overview of how these predictions are produced, and which error indicators are commonly used to evaluate their accuracy. Next, the methodology developed for data processing and forecasting is presented, with a reference to the programming language R which is used to extract the results. The last part of the thesis examines the accuracy of the proposed methodology and compares it with other known alternatives, listing relevant conclusions and future extensions.

Keywords: Forecasting Techniques, Probabilities, Renewable Energy Sources, Electricity, Photovoltaic Systems

Ευχαριστίες

Η διπλωματική αυτή εργασία εκπονήθηκε στα πλαίσια των ερευνητικών δραστηριοτήτων της Μονάδας Προβλέψεων και Στρατηγικής κατά το ακαδημαϊκό έτος 2017-2018. Η μονάδα υπάγεται στον Τομέα Βιομηχανικών Διατάξεων και Συστημάτων Αποφάσεων της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου.

Αρχικά, θα ήθελα να ευχαριστήσω τον Καθηγητή κ. Βασίλειο Ασημακόπουλο για την ευκαιρία που μου έδωσε να ασχοληθώ με το αντικείμενο των προβλέψεων και τη συγκεκριμένη εργασία, καθώς και τον Καθηγητή κ. Ιωάννη Ψαρρά και τον Αν. Καθηγητή κ. Δημήτριο Ασκούνη για την συμμετοχή τους στην τριμελή εξεταστική επιτροπή.

Επίσης, θα ήθελα να ευχαριστήσω τον διδάκτορα της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου Ευάγγελο Σπηλιώτη για την παρακολούθηση και τις χρήσιμες συμβουλές που πρόσφερε καθ' όλη τη διάρκεια εκπόνησης της εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους μου, που είναι κοντά μου και με στηρίζουν όλα αυτά τα χρόνια.

Καραμπλιάς Φώτιος, Σεπτέμβριος 2019

Πίνακας Περιεχομένων

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κεφάλαιο 1: Εισαγωγή

1.1 Αντικείμενο της εργασίας

Στις μέρες μας, οι ανανεώσιμες πηγές ενέργειας καθώς και οι σταθμοί αξιοποίησης τους, διαδραματίζουν σημαντικό ρόλο στην πορεία της παγκόσμιας οικονομίας. Κύρια και απόλυτη ανανεώσιμη πηγή ενέργειας αποτελεί ο ήλιος με χωρητικότητα ενέργειας μεγαλύτερης κατά 160 φορές από την αποθηκευμένη ενέργεια στη γη. Συνεπώς οι εγκαταστάσεις που αξιοποιούν την ηλιακή ακτινοβολία και παράγουν ηλεκτρική ενέργεια αποτελούν ακρογωνιαίο λίθο της σύγχρονης βιομηχανικής εποχής.

Η πρόβλεψη της παραγωγής ηλεκτρικής ενέργειας είναι ιδιαίτερα σημαντική, καθώς συμβάλλει στον ορθό προγραμματισμό και στην κατάλληλη λήψη αποφάσεων, ειδικά αν αναλογιστούμε τις οικονομικές και περιβαλλοντικές προεκτάσεις του συγκεκριμένου τομέα. Ο ηλεκτρισμός όπως γνωρίζουμε δεν αποθηκεύεται. Η περιορισμένη χωρητικότητα των μονάδων παραγωγής και ο κίνδυνος υποεκτίμησης ή υπερεκτίμησης της ποσότητας ενέργειας που θα καλύψει τις ανάγκες του πελάτη όπως συμβαίνει για παράδειγμα σε μία σύμβαση υποχρέωσης πλήρους εξυπηρέτησης συνιστούν τον κίνδυνο ποσοτήτων και πρέπει να ληφθεί υπ’ όψιν όταν η αγορά προσπαθεί να εξισορροπήσει την προσφορά με τη ζήτηση.

Μέχρι σήμερα υπάρχουν πολλά μοντέλα πρόβλεψης για την παραγόμενη ηλεκτρική ενέργεια. Ωστόσο, τα εν λόγω μοντέλα περιορίζονται συχνά στην παραγωγή σημειακών προβλέψεων που σε αρκετές εφαρμογές είναι ανεπαρκείς καθώς δεν μας ενημερώνουν σχετικά με το αναμενόμενο εύρος της παραγωγής και της κατανάλωσης αντίστοιχα. Αυτό είναι σημαντικό ειδικά σε περιπτώσεις όπου επιθυμούμε να υπολογίσουμε ενδεχόμενα ρίσκα και τις επιπτώσεις τους. Επίσης αν αναλογιστούμε την τυχαιότητα στις προβλέψεις για την αναμενόμενη ηλιακή ακτινοβολία και τη μη-γραμμική σχέση ηλιακής ακτινοβολίας και αντίστοιχης παραγόμενης ενέργειας, αντιλαμβανόμαστε τη δυσκολία του συγκεκριμένου προβλήματος, και την ανάγκη επίλυσης του με στοχαστικά μοντέλα πρόβλεψης.

Αντικείμενο, λοιπόν, της παρούσας εργασίας είναι η πρόβλεψη της συνάρτησης πυκνότητας πιθανότητας της παραγόμενης ηλεκτρικής ενέργειας αξιοποιώντας δεδομένα από σταθμούς αξιοποίησης της ηλιακής ακτινοβολίας με την βοήθεια των εκτιμητριών με πυρήνα(Kernel). Κατ’ επέκταση ελέγχουμε την ακρίβεια και την αξιοπιστία των προβλέψεων με βάση συγκεκριμένους στατιστικούς δείκτες, συγκρίνουμε τα αποτελέσματά μας με άλλες γνωστές μεθόδους κι εξάγουμε χρήσιμα συμπεράσματα.

Η προσέγγιση αυτή έχει το πλεονέκτημα ότι βοηθάει στον καλύτερο προγραμματισμό και διαχείριση από το σταθμό παραγωγής καθώς οι κατάλληλες προβλέψεις συμβάλλουν στην ορθή διαχείριση του αποθέματος, στη μειωμένη εξάρτηση από ορυκτά καύσιμα, και στην αποφυγή των οικονομικών κυρώσεων που επιφέρει το έλλειμα ενέργειας.

1.2 Δομή της εργασίας

Στο δεύτερο κεφάλαιο της παρούσας εργασίας, πραγματοποιείται μια επισκόπηση των Τεχνικών Προβλέψεων. Πιο συγκεκριμένα γίνεται μια σύντομη περιγραφή της ανάλυσης χρονοσειρών καθώς και μια αναφορά σε κάποιες βασικές στατιστικές μεθόδους πρόβλεψης. Στη συνέχεια δίνεται έμφαση στη Γραμμική Παλινδρόμηση μιας και αυτή μας απασχολεί στην παρούσα διπλωματική, αφού εφαρμόζεται ως τεχνική πρόβλεψης για την παραγόμενη ηλεκτρική ενέργεια από μία ΑΠΕ, δηλαδή τον ήλιο.

Το τρίτο κεφάλαιο αναφέρεται στα πιθανοτικά μοντέλα πρόβλεψης. Πιο συγκεκριμένα ξεκινά με μια εισαγωγή στη θεωρία πιθανοτήτων. Στη συνέχεια παρουσιάζονται τα κυριότερα πιθανοτικά μοντέλα πρόβλεψης και τέλος αναλύεται εκτενέστερα η χρησιμοποίηση των εκτιμητριών συναρτήσεων στην παραγωγή προβλέψεων. Η συγκεκριμένη θεωρία εφαρμόζεται στη μελέτη μας.

Το τέταρτο κεφάλαιο ασχολείται αποκλειστικά με σημεία αναφοράς και σφάλματα. Πιο συγκεκριμένα γίνεται διαχωρισμός των σφαλμάτων μεταξύ σημειακών προβλέψεων που προκύπτουν από στατιστικές μεθόδους και συναρτήσεων πυκνότητας πιθανότητας που προκύπτουν από πιθανοτικές μεθόδους. Στο κεφάλαιο αυτό, περιέχονται όλα τα σφάλματα και οι στατιστικοί δείκτες που θα χρησιμοποιηθούν στη συνέχεια για την εκτίμηση και την αξιολόγηση της ακρίβειας των προβλέψεων.

Το πέμπτο και έκτο κεφάλαιο παρουσιάζουν τις μεθοδολογίες που εφαρμόσθηκαν πάνω σε αληθινά δεδομένα ενεργειακής παραγωγής από έναν ελληνικό σταθμό με φωτοβολταϊκά. Στόχος των 2 κεφαλαίων είναι η ανάδειξη της αποτελεσματικότητας της πιθανοτικής μεθόδου πρόβλεψης με βάση την τεχνική που αναφέρθηκε και η αξιολόγησή αυτών υπό ρεαλιστικές συνθήκες, παρατηρώντας, δηλαδή, τις αποκλίσεις των προβλέψεων από τα πραγματικά δεδομένα.

Στο έβδομο και τελευταίο κεφάλαιο της εργασίας, εξάγονται τα κύρια συμπεράσματα με βάση τα αποτελέσματα. Τέλος, παρατίθενται προτάσεις που χρίζουν περαιτέρω μελέτης και έρευνας.

Κεφάλαιο 2: Τεχνικές Προβλέψεων

2.1 Γενικά για τις προβλέψεις

Η προσπάθεια για παραγωγή προβλέψεων αποτελούσε ανέκαθεν ένα αναπόσπαστο κομμάτι της ανθρώπινης φύσης , τόσο σε καθημερινό και ατομικό επίπεδο για αποφάσεις οι οποίες επηρεάζονται από γεγονότα για τα οποία έχουμε κάποια εκτίμηση και όχι βεβαιότητα , όπως την εξέλιξη των καιρικών φαινομένων , όσο και σε μακροπρόθεσμο και συλλογικό επίπεδο , στον τομέα των επενδύσεων ή και την χάραξη πολιτικής μιας ολόκληρης εταιρείας ή οργανισμού. Έτσι, από το 1980 και μετά, ο τομέας των προβλέψεων έχει αναπτυχθεί σημαντικά βρίσκοντας εφαρμογή τόσο σε ακαδημαϊκό επίπεδο, όσο και στο επίπεδο των επιχειρήσεων. Σε ακαδημαϊκό επίπεδο αναπτύσσονται και βελτιώνονται μοντέλα και μέθοδοι προβλέψεων, ενώ σε επίπεδο επιχειρήσεων εφαρμόζονται οι μέθοδοι προβλέψεων και φανερώνονται στην πράξη τα πραγματικά τους αποτελέσματα.

Χρησιμοποιώντας διάφορες τεχνικές προβλέψεων , ο κυριότερος σκοπός μας είναι η μείωση του σφάλματος, της απόκλισης δηλαδή της προβλεπόμενης από την πραγματική τιμή ενός μεγέθους. Στην προσπάθεια αυτή συμβάλλουν και ο ακαδημαϊκός τομέας, με ανάπτυξη και βελτίωση διαφόρων τεχνικών πρόβλεψης, αλλά και ο τομέας των επιχειρήσεων μέσω της πρακτικής εφαρμογής των τεχνικών αυτών σε πραγματικά δεδομένα και εξαγωγής αποτελεσμάτων για την ακρίβειά τους.

Το ενδιαφέρον για την παραγωγή προβλέψεων, πηγάζει κυρίως από την ανασφάλεια η οποία δημιουργείται εξαιτίας της αβεβαιότητας που υπάρχει για μελλοντικές καταστάσεις. Από την καθημερινή ζωή του καθενός, στις πολιτικές αποφάσεις που καλούνται να λάβουν εκλεγμένα σύνολα ανθρώπων ή αποφάσεις που αφορούν χάραξη πολιτικής επενδυτικού χαρακτήρα σε ιδιωτικούς ή κρατικούς τομείς, η αβεβαιότητα αποτελεί το μεγαλύτερο μειονέκτημα.

Ειδικότερα στον τομέα της οικονομίας, πέραν των ενδεχομένως λάθος αποφάσεων που θα μπορούσαν να ληφθούν έχοντας ως αποτέλεσμα μείωση κερδών ή ζημίες, η οικονομική ανασφάλεια από μόνη της θα μπορούσε να συμβάλλει σε αποσταθεροποίηση διαφόρων αγορών, οδηγώντας τελικά σε ζημιογόνες καταστάσεις και επιπτώσεις μεγάλης κλίμακας. Όλα τα παραπάνω καταδεικνύουν την ανάγκη για μεθόδους και τεχνικές προβλέψεων οι οποίες θα μπορούν να είναι αποδοτικές, χρήσιμες και ακριβείς.

Η έννοια της αβεβαιότητας , έχει κατηγοριοποιηθεί από τον Σ. Μακρυδάκη και τους συνεργάτες του σε δύο είδη, την «αβεβαιότητα του μετρό» και την «αβεβαιότητα της καρύδας». Το πρώτο είδος αναφέρεται σε μικρές αλλά συνεχείς διακυμάνσεις οι οποίες παρατηρούνται στην ιδιωτική αλλά και την επιχειρηματική καθημερινότητα. Η ονομασία πηγάζει από τη διακύμανση στο χρόνο που μπορεί να έχει ένας συρμός κατά τη μετάβασή του από έναν σταθμό σε έναν άλλον, η οποία μπορεί να οφείλεται σε πολύ μεγάλο πλήθος επιβατών, μειωμένο προσωπικό ή σε κάποιο τεχνικό πρόβλημα. Η «αβεβαιότητα της καρύδας» αναφέρεται σε εντελώς απρόσμενα γεγονότα τα οποία συμβαίνουν σπάνια, αλλά μπορούν να

έχουν εξαιρετικά σημαντική επίπτωση σε μελλοντικές τιμές ενός μεγέθους και αντιπαρατίθενται με μεγάλες οικονομικές και φυσικές καταστροφές. Η ονομασία προέρχεται από το μη προβλέψιμο και με μικρή πιθανότητα σενάριο του να πέσει μία καρύδα στο κεφάλι κάποιου.

Οι επιστημονικές μέθοδοι πρόβλεψης μπορούν να βελτιώσουν προβλήματα αστοχίας μόνο σαν αυτά του μετρό, και αυτό όμως είναι από μόνο του πολύ σημαντικό. Αβεβαιότητες, λοιπόν, σαν αυτές του μετρό και της καρύδας, που κάνουν τα μοντέλα προβλέψεων συχνά να αποκλίνουν σημαντικά ή και να αστοχούν πλήρως, έχουν κατά καιρούς κάνει τους ανθρώπους να αντιμετωπίζουν τον τομέα των προβλέψεων με καχυποψία. Απ' την άλλη όμως, όσο περισσότερο αλλάζει απρόβλεπτα και πολύπλοκα το περιβάλλον, τόσο πιο αδύναμος είναι ο καθένας να προβλέψει απλοϊκά μόνος του και τόσο πιο αναγκαία φαντάζει η ανάγκη παραγωγής προβλέψεων μέσω συστηματικών μεθόδων

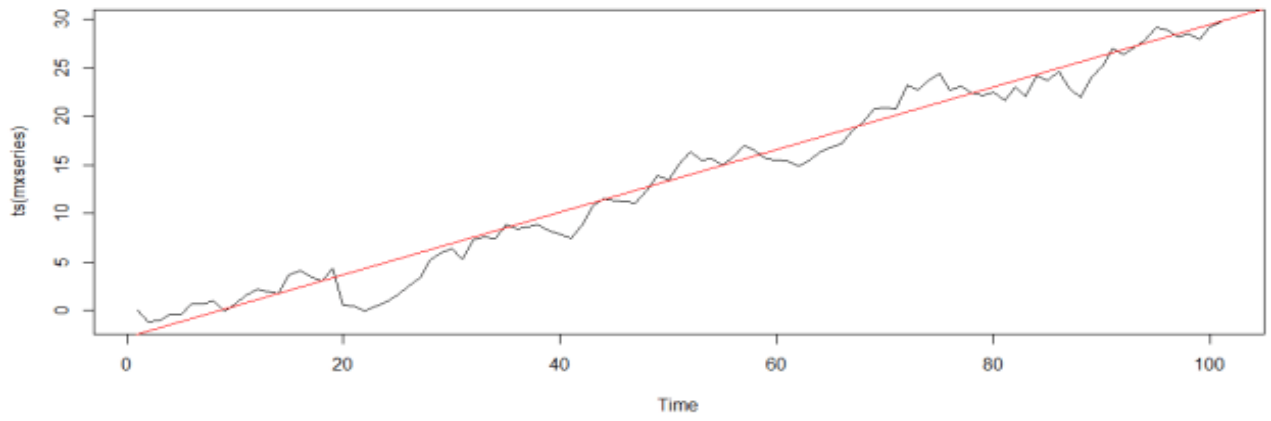
Ο συνεχής κύκλος μεταξύ ακαδημαϊκής μελέτης αλλά και βελτίωσης των διάφορων τεχνικών προβλέψεων και της πρακτικής εφαρμογής τους από ιδιωτικούς και κρατικούς φορείς οδηγεί σε συνεχή εξέλιξη της επιστήμης. Η εξέλιξη βασίζεται μεταξύ άλλων και στην εισαγωγή νέων μεθόδων για την βελτίωση των εκάστοτε τεχνικών προβλέψεων αλλά και στην αξιολόγηση των διαδικασιών με τις οποίες γίνεται η κατάταξη των μεθόδων.

2.2 Ποιοτικά χαρακτηριστικά χρονοσειρών

Οι χρονοσειρές αποτελούνται από ένα σύνολο παρατηρήσεων ενός συγκεκριμένου μεγέθους συναρτήσει του χρόνου. Μπορεί να αναφέρονται σε οποιονδήποτε μετρήσιμο τομέα και να έχουν συχνότητα από υποδιαιρέσεις του δευτερολέπτου έως και πολλαπλάσια ετών. Ανάλογα με τη συσχέτιση των διαδοχικών τιμών μίας χρονοσειράς μπορούν να διαχωριστούν σε ντετερμινιστικές, όπου οι διαδοχικές παρατηρήσεις είναι συσχετισμένες με αποτέλεσμα οι μελλοντικές τιμές να μπορούν να υπολογιστούν από τις προηγούμενες και στοχαστικές, όπου οι μελλοντικές τιμές προκύπτουν από μία στοχαστική διαδικασία με αποτέλεσμα το πλήθος των προηγούμενων τιμών να μη μπορούν να δώσουν σαφές αποτέλεσμα για τις μελλοντικές.

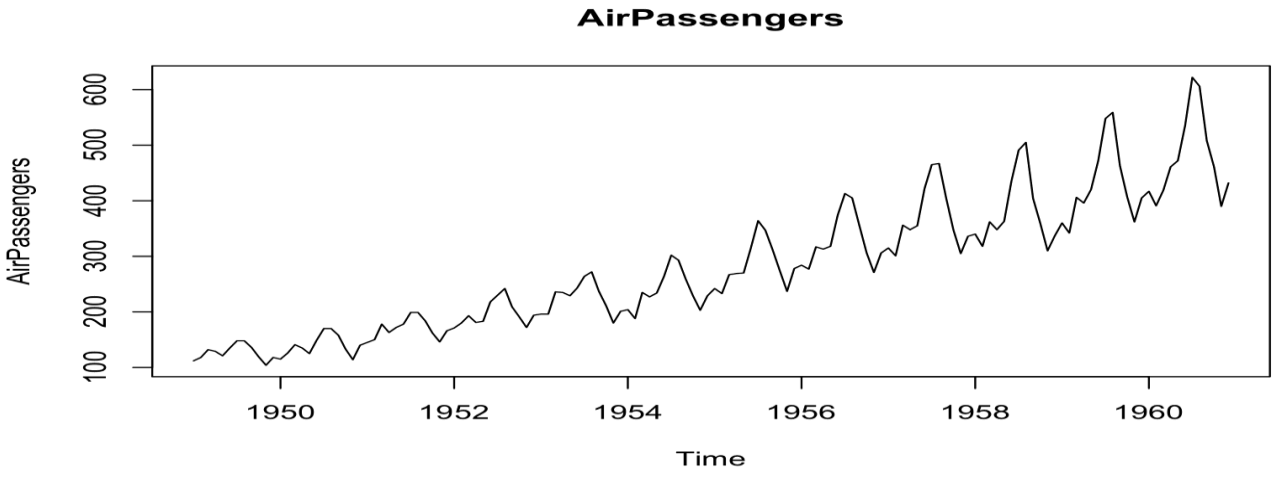
Στα πραγματικά δεδομένα, μπορούμε να καθορίσουμε μόνο εν μέρει τις μελλοντικές τιμές μίας χρονοσειράς βάσει των προηγούμενων δεδομένων καθώς η συντριπτική πλειοψηφία τους επηρεάζεται και από ένα τυχαίο παράγοντα. Σε μία προσπάθεια για αποσύνθεση των χρονοσειρών ώστε να μπορέσουμε να έχουμε μία καλύτερη εικόνα για τη μελλοντική συμπεριφορά τους, έχουν εισαχθεί οι έννοιες ορισμένων ποιοτικών χαρακτηριστικών οι οποίες βοηθούν στην επιλογή της κατάλληλης μεθόδου αλλά και τον καθορισμό των παραμέτρων της ώστε να έχουμε το καλύτερο δυνατό αποτέλεσμα. Τα βασικότερα από αυτά είναι τα εξής :

Τάση: Ορίζεται ως μία μακροπρόθεσμη μεταβολή του μέσου επιπέδου των τιμών της χρονοσειράς. Βέβαια, το πρόβλημα εδώ είναι το πώς κρίνεται κάτι ως μακροπρόθεσμο. Η απάντηση έχει να κάνει ξεκάθαρα με τη φύση των εκάστοτε δεδομένων και για αυτό το σκοπό θα πρέπει κανείς να διαθέτει στα χέρια του έναν ικανοποιητικό αριθμό δεδομένων για να μπορεί με ασφάλεια να αποφανθεί για την τάση και να μην υπάρξει παρερμηνεία των στοιχείων. Μία τάση μπορεί να είναι ανοδική, σταθερή ή πτωτική και να εκτιμηθεί ανάλογα τη μορφή της από μία ευθεία ή εκθετική καμπύλη.



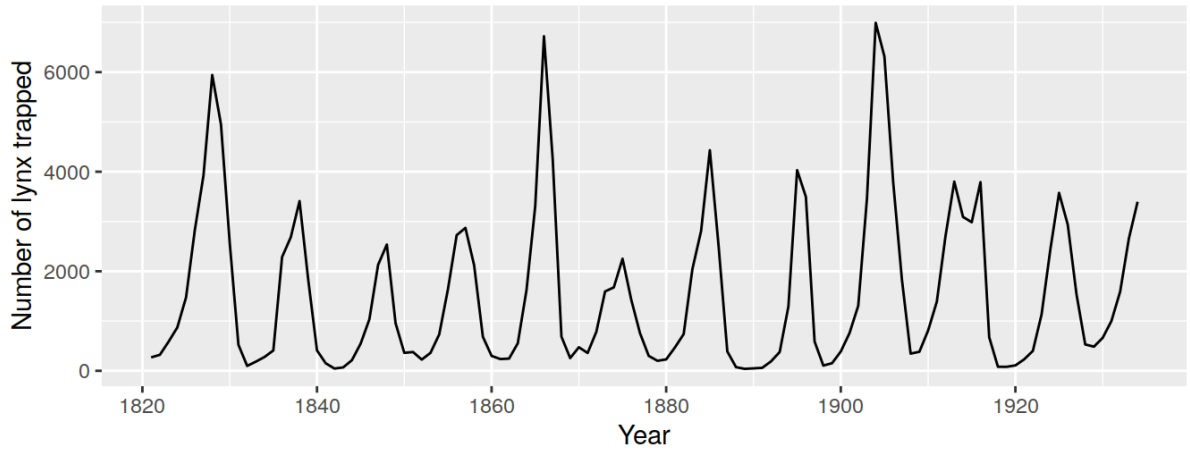
Σχήμα 2.1: Παράδειγμα Χρονοσειράς με γραμμική αύξουσα τάση

Εποχικότητα: Ορίζεται ως μια περιοδική διακύμανση η οποία έχει σταθερό και μικρότερο ή ίσο μήκος από ένα έτος. Είναι, μαζί με την τάση, το πιο εύκολα οπτικά αναγνωρίσιμο χαρακτηριστικό μιας χρονοσειράς λόγω του επαναληπτικού μοτίβου που παρουσιάζει, ενώ εύκολα μπορεί κάποιος να αντιμετωπίσει την επίδρασή της δεδομένου ότι γνωρίζει πότε και σε τι βαθμό αυτή επηρεάζει τα δεδομένα. Συγκεκριμένα, η εποχικότητα αντιμετωπίζεται με την εύρεση των δεικτών εποχικότητας για τα αντίστοιχα χρονικά διαστήματα και τη διαίρεση αυτών με τα πραγματικά δεδομένα. Η νέα χρονοσειρά που προκύπτει ονομάζεται αποεποχικοποιημένη χρονοσειρά.



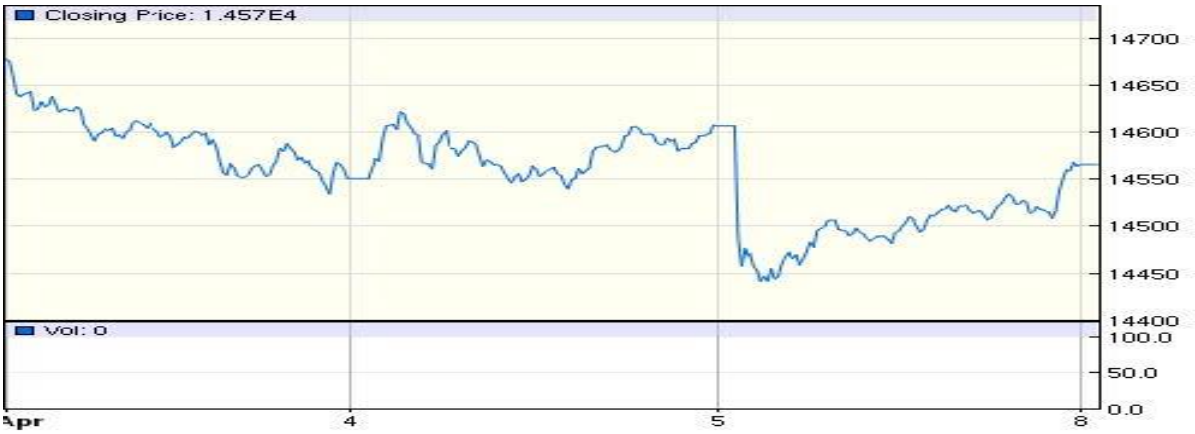
Σχήμα 2.2: Παράδειγμα Χρονοσειράς με σταθερή εποχικότητα

Κυκλικότητα: Αναφέρεται σε μία κυματοειδή μεταβολή η οποία δεν εμφανίζεται σε σταθερές χρονικές περιόδους που κατά κανόνα διαρκούν περισσότερο από ένα έτος. Οφείλεται κυρίως σε εξωγενείς παράγοντες οι οποίοι επηρεάζουν τη χρονοσειρά και εμφανίζεται συνήθως σε οικονομικά μεγέθη καθώς και σε χρονοσειρές οι οποίες επηρεάζονται άμεσα από αυτά. Κάποια από αυτά τα μεγέθη είναι το Ακαθάριστο Εθνικό Προϊόν , οι δείκτες βιομηχανικής παραγωγής, οι τιμές των μετοχών καθώς και οι τιμές του πετρελαίου και του χρυσού. Οι διακυμάνσεις αυτές, απορρέουν από διαδοχικές περιόδους ανόδου και ύφεσης των παγκόσμιων και εγχώριων οικονομιών αλλά και σχέσεων μεταξύ οικονομιών διάφορων χωρών και είναι γνωστές με τον όρο επιχειρηματικοί κύκλοι.



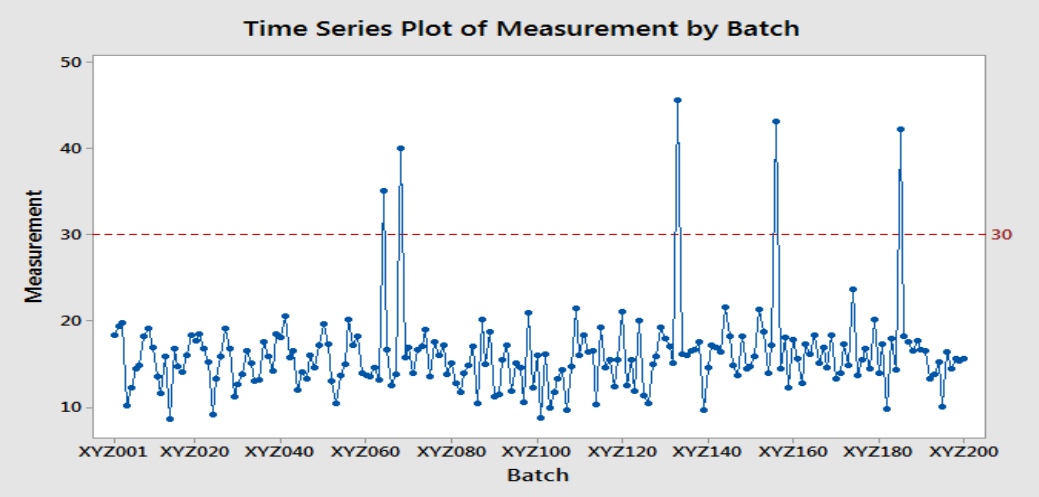
Σχήμα 2.3: Παράδειγμα Χρονοσειράς με κυκλικότητα

Τυχαιότητα: αποτελεί τη διαφορά ανάμεσα στην συνδυασμένη επίδραση των τριών πρώτων συνιστωσών των χρονοσειρών (τάση, κυκλικότητα και εποχικότητα) και των πραγματικών δεδομένων. Μπορεί να χαρακτηριστεί λοιπόν ως κάτι το στοχαστικό και να αντιμετωπιστεί ανάλογα.



Σχήμα 2.4: Παράδειγμα Χρονοσειράς με τυχαιότητα

Ασυνέχειες : Είναι μεμονωμένες παρατηρήσεις οι οποίες εμφανίζονται στα γραφήματα των χρονοσειρών ως απότομες αλλαγές στο πρότυπο της συμπεριφοράς των δεδομένων. Κάποια πολύ απότομα αυξημένη ή μειωμένη τιμή η οποία μπορεί να οφείλεται σε ένα μεμονωμένο απρόβλεπτο γεγονός. Οι απότομες αυτές αλλαγές μπορεί να έχουν παροδική διάρκεια (outliers ή special events) όπως για παράδειγμα μία έντονη μείωση στη παραγωγή εξαιτίας μίας απεργίας ή μόνιμο χαρακτήρα (level shifts) όπου επιδρούν για μεγαλύτερο διάστημα ή και αλλάζουν εντελώς το μέσο όρο των τιμών μίας χρονοσειράς, όπως για παράδειγμα το αποτέλεσμα στο πλήθος των πελατών μιας μεταφορικής εταιρείας μετά από ένα ατύχημα κατά τη μεταφορά σημαντικού φορτίου ή την εισαγωγή σε ένα ανταγωνιστικό χώρο μίας νέας εταιρείας η οποία μπορεί να προκαλέσει πτώση των πωλήσεων των ήδη υπαρχόντων εταιρειών και σταθεροποίηση τους σε νέο χαμηλότερο μέσο επίπεδο. Η ερμηνεία τέτοιων απότομων μεταβολών στην περίπτωση των special events απαιτεί σημαντική θεωρητική και κριτική ικανότητα και δε μπορεί να προβλεφθεί με μελέτη ιστορικών δεδομένων.



Σχήμα 2.5: Παράδειγμα Χρονοσειράς με ασυνέχειες

2.3 Κατηγορίες μεθόδων πρόβλεψης

Οι μέθοδοι πρόβλεψης, ανάλογα τη σκοπιά που εξετάζουν τα δεδομένα και τον τρόπο που τα επεξεργάζονται, χωρίζονται σε τρεις μεγάλες κατηγορίες: τις ποσοτικές (quantitative), τις κριτικές (judgmental) και τις τεχνολογικές (technological). Στη παρούσα διπλωματική θα χρησιμοποιήσουμε μόνο ποσοτικές μεθόδους πρόβλεψης, οι οποίες και θα αναλυθούν εκτενέστερα στη συνέχεια. Αξίζει ωστόσο και μία σύντομη βιβλιογραφική αναφορά και στις άλλες δύο κατηγορίες μεθόδων.

Στις ποσοτικές μεθόδους, βασικά εργαλεία για ένα ερευνητή είναι η στατιστική και τα μαθηματικά μοντέλα. Με βάση αυτά επεξεργάζεται τις χρονοσειρές και εξάγει προβλέψεις. Όταν όμως υπάρξουν κάποιες αλλαγές στο μοτίβο που ακολουθεί η χρονοσειρά, όπως special events, τότε οι ποσοτικές μέθοδοι αδυνατούν να τις κατανοήσουν και ακόμα περισσότερο να τις προβλέψουν. Σε αυτό το σημείο πρέπει να επεμβαίνει η ανθρώπινη κρίση και να αποφασίζει το αν και πόσο θα επηρεάσει η κάθε αλλαγή τις προβλέψεις. Βέβαια, κάτι τέτοιο απαιτεί καλές γνώσεις, μεγάλη εμπειρία και κριτική ικανότητα απ' το άτομο που προβλέπει και για αυτό το σκοπό το ρόλο αυτό αναλαμβάνουν συνήθως επιτροπές ή άτομα που συμβουλεύονται παράλληλα από άλλους ειδικούς και managers. Αυτές είναι οι κριτικές προβλέψεις.

Οι τεχνολογικές προβλέψεις, με τη σειρά τους, χρησιμοποιούνται για μακροπρόθεσμες προβλέψεις σχετικά με τεχνολογικά, οικονομικά, κοινωνικά και πολιτικά θέματα. Χωρίζονται σε διερευνητικές (exploratory) και κανονιστικές (normative) μεθόδους. Οι διερευνητικές μέθοδοι ξεκινούν από το παρελθόν ή το παρόν και εξετάζοντας όλες τις πιθανές περιπτώσεις οδηγούνται στο μέλλον, ενώ οι κανονιστικές μέθοδοι πρώτα καθορίζουν όλους τους μελλοντικούς στόχους και έπειτα εξετάζουν τη δυνατότητα επίτευξης τους λαμβάνοντας υπόψη τα δεδομένα.

2.4 Βασικές στατιστικές μέθοδοι πρόβλεψης

2.4.1 Απλοϊκή Μέθοδος (Naive)

Είναι η πιο απλή στατιστική μέθοδος πρόβλεψης. Ως πρόβλεψη θεωρείται η τελευταία διαθέσιμη παρατήρηση. Δηλαδή:

$$F_t = Y_{t-1}$$

Η τεχνική αυτή ενδείκνυται για περιπτώσεις που τα δεδομένα δεν παρουσιάζουν τάση και για μικρούς ορίζοντες πρόβλεψης. Χρησιμοποιείται κυρίως ως μέτρο σύγκρισης για την ακρίβεια άλλων μεθόδων (benchmark) και για την τεχνική back-casting. Η τεχνική back-casting χρησιμοποιείται σε περιπτώσεις που εμφανίζονται κενές τιμές σε μια χρονοσειρά και για την συμπλήρωση των κενών αυτών.

2.4.2 Μέθοδοι εκθετικής εξομάλυνσης

Η εκθετική εξομάλυνση είναι μια μέθοδος πρόβλεψης η οποία προεκτείνει στοιχεία του προτύπου των ιστορικών δεδομένων, όπως τάσεις και εποχιακούς κύκλους, στο μέλλον. Οι προβλέψεις υπολογίζονται μετά από εξομάλυνση των δεδομένων, προκειμένου να

απομονωθούν τα πραγματικά πρότυπα από τις τυχαίες διακυμάνσεις. Η δημοτικότητα των μεθόδων αυτών οφείλεται στην απλότητα των μοντέλων που υιοθετούν, τις περιορισμένες απαιτήσεις τους σε αποθήκευση δεδομένων και τον μειωμένο υπολογιστικό φόρτο. Εμπειρικές μελέτες αποδεικνύουν ότι οι μέθοδοι εκθετικής εξομάλυνσης παρουσιάζουν ικανοποιητικά ποσοστά ακρίβειας σε σχέση με πιο πολύπλοκες μεθόδους πρόβλεψης. Το γεγονός αυτό οφείλεται στο ότι οι μέθοδοι εκθετικής εξομάλυνσης δεν επηρεάζονται από τις ιδιομορφίες των προτύπων των δεδομένων ή από περιστασιακά εμφανιζόμενες ακραίες τιμές, οι οποίες παρατηρούνται σε επιχειρησιακά δεδομένα.

2.4.2.1 Απλή Εκθετική Εξομάλυνση (Simple Exponential Smoothing)

Το μοντέλο αυτό ονομάζεται και μοντέλο απλής εκθετικής εξομάλυνσης ή SES. Οι ακόλουθες εξισώσεις περιγράφουν το μοντέλο σταθερού επιπέδου.

$$e_t = Y_t - F_t$$

$$S_t = S_{t-1} - a * e_t$$

$$F_{t+1} = S_t$$

Στις παραπάνω εξισώσεις, το e δηλώνει το σφάλμα πρόβλεψης, το S το επίπεδο, F την πρόβλεψη και α μια σταθερά εξομάλυνσης που λαμβάνει οποιαδήποτε τιμή στο διάστημα [0,1]. Το μοντέλο αυτό είναι κατάλληλο για δεδομένα που δεν παρουσιάζουν έντονο το στοιχείο της τάσης.

Το αρχικό επίπεδο ορίζεται με διάφορούς τρόπους και συνήθως χρησιμοποιείται ο μέσος όρος όλων ή κάποιων αρχικών παρατηρήσεων, η πρώτη παρατήρηση ή το σταθερό επίπεδο από το μοντέλο της απλής γραμμικής παλινδρόμησης. Ο σημαντικότερος όμως παράγοντας είναι ο καθορισμός του συντελεστή εξομάλυνσης α. Ο καθορισμός του α εξαρτάται από τον θόρυβο που έχουν τα δεδομένα και από (όσο περισσότερος τόσο μικρότερη τιμή παίρνει ο α) και από την σταθερότητα του μέσου όρου της χρονοσειράς (μεγάλες μεταβολές αντιμετωπίζονται με μεγαλύτερο α). Υπάρχουν διάφοροι αλγόριθμοι εύρεσης του κατάλληλου α, συνήθως με την εύρεση εκείνου που ελαχιστοποιεί κάποιο δείκτη σφάλματος. Για τις ακραίες τιμές α=1, η πρόβλεψη γίνεται ίδια με την naive ενώ για α=0, η πρόβλεψη παραμένει ίδια και ίση με το αρχικό επίπεδο. Είναι εμφανές πως σε περιπτώσεις που απαιτούνται προβλέψεις ορίζοντα μεγαλύτερου από μια χρονική περίοδο, όλες οι προβλέψεις είναι ίδιες με την τελευταία.

2.4.2.2 Μοντέλο Γραμμικής Τάσης (Holt Exponential Smoothing)

Το μοντέλο γραμμικής τάσης αποτελεί μια εξέλιξη του μοντέλου σταθερού επιπέδου και δίνει τη δυνατότητα διαχείρισης δεδομένων που παρουσιάζουν το στοιχείο της τάσης. Το μοντέλο περιγράφεται από τις παρακάτω εξισώσεις:

$$e_t = Y_t - F_t$$

$$S_t = S_{t-1} + T_{t-1} + \alpha \cdot e_t$$

$$T_t = T_{t-1} + \beta \cdot e_t$$

$$F_{t+m} = S_t - m \cdot T_t$$

Η νέα παράμετρος β που εμπεριέχεται στις εξισώσεις ονομάζεται συντελεστής εξομάλυνσης τάσης και λαμβάνει τιμές στο διάστημα [0,1]. Σε αυτό το μοντέλο χρειάζεται αρχικοποίηση τόσο του επιπέδου όσο και της τάσης. Το αρχικό επίπεδο ορίζεται όπως στην απλή εκθετική εξομάλυνση, ενώ η αρχική τάση ως η διαφορά της ν-στής και της πρώτης παρατήρησης διαιρεμένης με ν-1, ή ως η σταθερά κλίσης από το μοντέλο της απλής γραμμικής παλινδρόμησης. Το αρχικό επίπεδο και η αρχική τάση πρέπει να καθορίζονται με προσοχή καθώς επηρεάζουν αρκετά την τελική πρόβλεψη. Η μεγάλη διαφορά του μοντέλου αυτού από τη μέθοδο SES είναι η παραγωγή προβλέψεων με χρονικό ορίζοντα μεγαλύτερο της μονάδας. Λόγω της θεώρησης πως τα δεδομένα έχουν μια σταθερά ανοδική τάση, οι προβλέψεις για ορίζοντα μεγαλύτερο της μονάδας προκύπτουν με τη χρήση των τελευταίων διαθέσιμων τιμών για το επίπεδο και την τάση και αύξηση του δείκτη m.

2.4.2.3 Μοντέλο Μη Γραμμικής Τάσης (Damped Exponential Smoothing)

Το μοντέλο φθίνουσας γραμμικής τάσης είναι μία υποπερίπτωση του μοντέλου μη γραμμικής τάσης. Το μοντέλο μη γραμμικής τάσης έχει τη δυνατότητα μεταβολής της μορφής της χρονοσειράς και της προσαρμογής της σε μη γραμμικές τάσεις. Η προσαρμογή αυτή γίνεται μέσω μιας μεταβλητής που ονομάζεται παράμετρος διόρθωσης της τάσης φ. Το μοντέλο μη γραμμικής τάσης περιγράφεται μαθηματικά από τις παρακάτω εξισώσεις:

$$e_t = Y_t - F_t$$

$$S_t = S_{t-1} + T_{t-1} + \alpha \cdot e_t$$

$$T_t = T_{t-1} + \beta \cdot e_t$$

$$F_{t+1} = S_t + \sum_{i=1}^{m} \varphi^i \cdot T_t$$

Όπου, t η χρονική περίοδος, Y_t η πραγματική τιμή των δεδομένων, F_t η πρόβλεψη τη χρονική στιγμή t, e_t το σφάλμα (απόκλιση πραγματικής τιμής από πρόβλεψη), S_t το επίπεδο της χρονοσειράς, T_t η τάση της χρονοσειράς, α ο συντελεστής εξομάλυνσης επιπέδου, λαμβάνει τιμές στο διάστημα [0,1], β ο συντελεστής εξομάλυνσης της τάσης, λαμβάνει τιμές στο διάστημα [0,1], φ ο συντελεστής διόρθωσης της τάσης, λαμβάνει τιμές στο διάστημα (0,1) και m χρονικός ορίζοντας της πρόβλεψης.

Εύκολα γίνεται αντιληπτό, ότι οι εξισώσεις είναι πανομοιότυπες με αυτές του γραμμικού μοντέλου πλην της τελευταίας, όπου αντί να υπολογίζεται μια γραμμική αύξηση μέσω του συντελεστή m, πραγματοποιείται ένας μη γραμμικός υπολογισμός αυτής, γεγονός που οφείλεται στην παράμετρο εξομάλυνσης φ. Η παράμετρος φ, σε αντίθεση με τις παραμέτρους α και β, δύναται να λάβει τιμές μεγαλύτερες του μηδενός, χωρίς κάποιο άνω όριο αλλά είναι πολύ σημαντική η επιβολή άνω και κάτω ορίων ανάλογα με την εκάστοτε περίπτωση.

Όπως αναφέρεται και παραπάνω για 0<φ<1 προκύπτει το μοντέλο της φθίνουσας τάσης (Damped Exponential Smoothing). Ανάλογα την τιμή που παίρνει η παράμετρος φ, το μοντέλο της μη γραμμικής τάσης μπορεί να πάρει περεταίρω τις μορφές:

1. Για φ=0 προκύπτει το μοντέλο της απλής εκθετικής εξομάλυνσης (Simple Exponential Smoothing), αφού η τάση δεν συμμετέχει στην παραγωγή προβλέψεων.
2. Για φ=1 προκύπτει το μοντέλο της γραμμικής τάσης (Holt Exponential Smoothing), καθώς στην εξίσωση υπολογισμού της πρόβλεψης, τη θέση του αθροίσματος παίρνει το γινόμενο της μεταβλητής χρονικού ορίζοντα m κα της προηγούμενης τάσης Tt.
3. Για φ>1 προκύπτει το μοντέλο της εκθετικής τάσης, το οποίο χαρακτηρίζεται από μεγάλη προκατάληψη.

Σχετικά με την επιλογή του αρχικού επιπέδου (S_0), της αρχικής τάσης (T_0) και την βελτιστοποίηση των παραμέτρων εξομάλυνσης, ισχύουν τα ίδια που αναφέρθηκαν παραπάνω για την περίπτωση του μοντέλου γραμμικής τάσης. Συγκεκριμένα για την μη γραμμική τάση προτείνεται ωστόσο η εφαρμογή της γραμμικής παλινδρόμησης με ανεξάρτητη μεταβλητή το χρόνο t για τον προσδιορισμό των S_0 και T_0. Για την εύρεση των βέλτιστων συνδυασμών των παραμέτρων α, β, φ εφαρμόζεται και πάλι η διαδικασία της γραμμικής αναζήτησης, ελαχιστοποιώντας το μέσο τετραγωνικό σφάλμα (MSE).

Λόγω της θετικής προκατάληψης που περιέχει το μοντέλο εκθετικής τάσης χρησιμοποιείται σε ορισμένες μόνο ειδικές περιπτώσεις, όπως η εισαγωγή ενός προϊόντος στην αγορά. Θετική προκατάληψη εντοπίζεται και στα μοντέλα γραμμικής τάσης. Γι' αυτό το λόγο τα μοντέλα φθίνουσας τάσης τυγχάνουν μεγάλης αποδοχής ιδιαίτερα για προβλέψεις μεγάλου χρονικού ορίζοντα. Εμπειρικά αποτελέσματα φαίνεται να δικαιολογούν την επιλογή αυτή.

2.4.3 Αυτοπαλινδρομικά μοντέλα κινητού μέσου όρου (μέθοδος ARIMA)

Τα αυτοπαλινδρομικά μοντέλα κινητού μέσου όρου, ανήκουν στα στοχαστικά μαθηματικά μοντέλα και με την βοήθειά τους μπορούμε να περιγράψουμε την διαχρονική εξέλιξη φυσικών μεγεθών, που εξαρτώνται από μη ντετερμινιστικούς παράγοντες. Είναι αρκετά διαδεδομένα, και ειδικά σε περιπτώσεις που εμπεριέχονται φυσικά μεγέθη, τα οποία δεν τα γνωρίζουμε απόλυτα, και επιπλέον όταν δεν γνωρίζουμε τους παράγοντες οι οποίοι τα επηρεάζουν. Τα στοχαστικά αυτά μοντέλα περιέχουν τον τυχαίο παράγοντα, τις τιμές του μεγέθους οι οποίες εμφανίστηκαν σε παρελθοντικές χρονικές στιγμές και μπορεί και κάποιους επιπλέον στοχαστικούς παράγοντες. Πιο συγκεκριμένα, με την χρήση αυτών των μοντέλων, δυνάμεθα να υπολογίσουμε την πιθανότητα ή την τιμή του μεγέθους που εξετάζουμε να βρίσκεται σε ένα συγκεκριμένο διάστημα. Ως απόρροια όλων των παραπάνω, μπορούμε να αντιληφθούμε ότι τα Αυτοπαλινδρομικά μοντέλα κινητού μέσου όρου είναι αρκετά αποτελεσματικά κυρίως σε βραχυπρόθεσμες προβλέψεις, εφόσον δίνουν μεγαλύτερη έμφαση στις πιο πρόσφατες παρελθοντικές παρατηρήσεις. Βασική προϋπόθεση για την καλύτερα αποτελέσματα στα εξής μοντέλων είναι να εφαρμόζονται σε χρονοσειρές οι οποίες είναι στάσιμες και διακριτές. Διακριτές είναι οι χρονοσειρές που όλες οι παρατηρήσεις τους έχουν ληφθεί σε χρονικές στιγμές που ισαπέχουν μεταξύ τους, ενώ στάσιμες θεωρούνται αυτές που η μέση τιμή, η διακύμανσή τους και η συνάρτηση αυτοσυσχέτισής τους είναι σταθερές σε όλη την διάρκεια του χρόνου.

2.5 Γραμμική Παλινδρόμηση

Η γραμμική παλινδρόμηση αποτελεί μια ποσοτική μέθοδο πρόβλεψης, η οποία ανήκει στην κατηγορία των αιτιοκρατικών μοντέλων. Θα επικεντρωθούμε σε αυτήν μιας και θα αποτελέσει το μέτρο σύγκρισης στην παρούσα διπλωματική σε σχέση με τα πιθανοτικά μοντέλα πρόβλεψης που θα αναλυθούν στο επόμενο κεφάλαιο.

Στη στατιστική, η γραμμική παλινδρόμηση είναι μια προσέγγιση για τη μοντελοποίηση της σχέσης μεταξύ μιας βαθμωτής εξαρτημένης μεταβλητής Y και μία ή περισσότερες επεξηγηματικές μεταβλητές (ή ανεξάρτητη μεταβλητή) συμβολίζεται X. Περίπτωση μιας επεξηγηματικής μεταβλητής ονομάζεται απλή γραμμική παλινδρόμηση. Για περισσότερες από μία επεξηγηματικές μεταβλητές, η διαδικασία ονομάζεται πολλαπλή γραμμική παλινδρόμηση.

Στην γραμμική παλινδρόμηση, οι σχέσεις μοντελοποιούνται με την χρήση γραμμικών συναρτήσεων (μοντέλων) πρόβλεψης, των οποίων οι άγνωστες παράμετροι μοντελοποίησης υπολογίζονται από τα δεδομένα. Αυτά τα μοντέλα καλούνται γραμμικά μοντέλα. Συνήθως, η μέση τιμή της εξαρτημένης μεταβλητής Y, δεδομένης της X (ή των X_i), υποτίθεται να είναι μια συνάρτηση συσχετισμού της X, ενώ πιο σπάνια, η διάμεση τιμή ή κάποια άλλη ποσόστωση της κατανομής της Y υποτίθεται να είναι μια γραμμική συνάρτηση της X.

Όπως συμβαίνει σε όλους τους τύπους παλινδρόμησης, η γραμμική παλινδρόμηση εστιάζει στην κατανομή πιθανότητας y, δεδομένης της X, και όχι στην απο κοινού κατανομή πιθανότητας των y και X, αντικείμενο της ανάλυσης πολλών μεταβλητών. Η γραμμική παλινδρόμηση ήταν ο πρώτος τύπος της ανάλυσης παλινδρόμησης που μελετήθηκε διεξοδικά και εφαρμόστηκε εκτενώς σε πρακτικές εφαρμογές. Αυτό οφείλεται στο γεγονός ότι τα μοντέλα που εμφανίζουν γραμμική εξάρτηση από τις άγνωστες παραμέτρους τους παρουσιάζουν καλύτερη εφαρμογή από μοντέλα τα οποία εμφανίζουν μη γραμμική εξάρτηση από τις παραμέτρους τους, καθώς επίσης και στο γεγονός ότι είναι ευκολότερος ο προσδιορισμός των στατιστικών ιδιοτήτων των προκυπτουσών εκτιμητριών.

2.5.1 Απλή Γραμμική Παλινδρόμηση

Η απλή γραμμική παλινδρόμηση είναι μία ευθεία προσέγγιση που προβλέπει μία ποσοτική απόκριση Y (εξαρτημένη μεταβλητή) με βάση μία μόνο ανεξάρτητη μεταβλητή X. Υποθέτει δηλαδή ότι υπάρχει γραμμική σχέση ανάμεσα σε X και Y και κάνει προβλέψεις που βασίζονται σε αυτήν τη σχέση. Μαθηματικά, η γραμμική αυτή σχέση εκφράζεται ως εξής::

$$Y = a + \beta \cdot X$$

Με a συμβολίζεται η τεταγμένη του σημείου τομής της ευθείας με τον άξονα των εξαρτημένων μεταβλητών, ενώ με β συμβολίζεται η κλίση της ευθείας. Αν θεωρήσουμε ότι κάθε εκτιμώμενη τιμή $\hat{Y}_i$ έχει ένα σφάλμα e_i, που ορίζεται ως η απόσταση της ευθείας μας από την πραγματική τιμή, τότε ισχύει ότι:

$$e_i = Y_i - \alpha - \beta \cdot X_i$$

Ορίζουμε το άθροισμα των τετραγώνων των σφαλμάτων e_i ως RSS(Residual Sum of Squares).

$$RSS = \sum_{i=1}^{n} (Y_i - \alpha - \beta \cdot X_i)^2$$

Οι τιμές των συντελεστών a και β υπολογίζονται με βάση την αρχή των ελαχίστων τετραγώνων, επιλέγονται δηλαδή οι συντελεστές που ελαχιστοποιούν το άθροισμα των τετραγώνων των διαφορών των πραγματικών τιμών Y_i από τις προβλεπόμενες σε κάθε χρονική περίοδο $\hat{Y}_i$,δηλαδή το RSS, όπως φαίνεται και στην επόμενη σχέση:

$$(\alpha, \beta) | \min \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

Με βάση τη λογική της ελαχιστοποίησης της απόστασης των πραγματικών παρατηρήσεων Yi από τη βέλτιστη γραμμή παλινδρόμησης προκύπτουν οι εξισώσεις υπολογισμού των συντελεστών a και β, όπως παρουσιάζονται στη συνέχεια :

$$\beta = \frac{\frac{\sum_{i=1}^{n} Y_i X_i}{n} - \bar{X}\bar{Y}}{\frac{\sum_{i=1}^{n} (X_i)^2}{n} - \bar{X}^2} = \frac{\sum_{i=1}^{n} [(X_i - \bar{X}) \cdot (Y_i - \bar{Y})]}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

$$a = \bar{Y} - \beta \cdot \bar{X}$$

Με $\bar{X}$ και $\bar{Y}$ συμβολίζονται οι μέσες τιμές των μεταβλητών X και Y αντίστοιχα. Ο αριθμός των παρατηρήσεων, βάσει των οποίων υπολογίζεται η ευθεία παλινδρόμησης, συμβολίζεται με n.

Η κλίση της ευθείας παλινδρόμησης ισούται με το γινόμενο της συσχέτισης των μεταβλητών X και Y και του λόγου των τυπικών αποκλίσεων σ_Y και σ_X, σύμφωνα με την παρακάτω σχέση:

$$\beta = r_{xy} \cdot \frac{\sigma_Y}{\sigma_X}$$

Η πρόβλεψη με χρήση της μεθόδου της απλής γραμμικής παλινδρόμησης δίνει μια καλή εικόνα της μέσης και της μακροπρόθεσμης συμπεριφοράς του υπό μελέτη μεγέθους.

Σε περίπτωση που η σχέση ανάμεσα σε δύο μεταβλητές, την εξαρτημένη και την ανεξάρτητη, δεν είναι γραμμική, μπορεί και πάλι να εφαρμοστεί η μέθοδος της απλής γραμμικής παλινδρόμησης, αφού πρώτα γίνει μετασχηματισμός της σχέσης των δύο μεταβλητών σε γραμμική.

2.5.2 Αξιολόγηση παραμέτρων μοντέλου παλινδρόμησης

Έχουμε υπολογίσει τις τιμές των συντελεστών α και β, ώστε η τιμή του RSS να ελαχιστοποιείται. Επομένως με βάση την ανάλυση της προηγούμενης ενότητας έχουμε υποθέσει ότι υπάρχει μία ευθεία που ορίζεται από το α και το β που υπολογίσαμε, η οποία διέρχεται όσο το δυνατόν πλησιέστερα από όλα τα ζεύγη πραγματικών τιμών (X_i, Y_i) που έχουμε στη διάθεση μας. Έχουμε δηλαδή υποθέσει εξαρχής ότι οι πραγματικές τιμές του Y που διαθέτουμε, σχετίζονται γραμμικά με τις αντίστοιχες τιμές του x. Ωστόσο, καθώς η ευθεία ποτέ δεν μπορεί να διέρχεται ακριβώς από όλα τα σημεία των πραγματικών ζευγών (X_i, Y_i), θεωρούμε πως στο γραμμικό μοντέλο που εξαγάγαμε υπάρχει πάντα ένα σφάλμα. Το σφάλμα αυτό θα το εισάγουμε στη γραμμική μας σχέση με τον όρο ε, δηλαδή έχουμε:

$$Y = a + \beta \cdot X + \varepsilon$$

Καθώς το γραμμικό μοντέλο είναι στην ουσία πολύ απλουστευμένο, στην πράξη οι τιμές των X και Y πιθανότατα να μη συνδέονται εντελώς γραμμικά. Μπορεί δηλαδή αύξηση ή μείωση του X να συνεπάγεται αντίστοιχα αύξηση ή μείωση του Y, χωρίς όμως οι μεταβολές να γίνονται με γραμμικό τρόπο. Επομένως είναι βέβαιο ότι θα υπάρχει διαφορά ανάμεσα στην ευθεία που εξαγάγαμε και στις πραγματικές τιμές των ζευγών $(X_i\ , Y_i)$, γι αυτό και εισάγουμε στη γραμμική σχέση και τον όρο ε, ο οποίος θεωρούμε ότι είναι ανεξάρτητος του X.

Στο σημείο αυτό κρίνεται σκόπιμο να ορίσουμε κάποια υπολογιστικά σφάλματα για την τιμή του α και του β. Για κάθε μία από αυτές τις 2 τιμές που εκτιμήσαμε, με τη βοήθεια του γραμμικού μοντέλου το οποίο εξήχθη από τα δεδομένα μας, ορίζεται ένα σφάλμα Standard Error (SE) με βάση τους παρακάτω τύπους:

$$SE(a)^2 = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]$$

$$SE(\beta)^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

η μοναδική τιμή που δε γνωρίζουμε είναι το σ^2, το οποίο το ορίζουμε ως Residual Standard Error(RSE) και μπορούμε να το υπολογίσουμε από τον τύπο:

$$RSE = \sqrt{\frac{1}{n-2}RSS} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$$

,όπου το RSS έχει υπολογιστεί παραπάνω. Με τη βοήθεια των Standard Errors μπορούμε να ορίσουμε τα αντίστοιχα διαστήματα εμπιστοσύνης για τις τιμές των α και β και συγκεκριμένα τα 95% διαστήματα εμπιστοσύνης για την κάθε σταθερά. Ένα διάστημα εμπιστοσύνης 95% ορίζεται ως ένα διάστημα τιμών με άνω και κάτω όριο, μέσα στο οποίο υπάρχει 95% πιθανότητα να βρίσκεται πραγματική τιμή της σταθεράς του γραμμικού μας μοντέλου.

Το 95% διάστημα εμπιστοσύνης δίνεται από τον τύπο $\alpha' \ + \ 2 \cdot SE(\alpha')$, δηλαδή ορίζεται ως:

$$[\ \alpha' \ - \ 2 \cdot SE(\alpha')\ , \alpha' \ + \ 2 \cdot SE(\alpha')\ ]$$

Κατ' αντιστοιχία, το 95% διάστημα εμπιστοσύνης για τη σταθερά β ορίζεται ως:

$$[\ \beta' \ - \ 2 \cdot SE(\beta'), \beta' \ + \ 2 \cdot SE(\beta')]$$

Έτσι είμαστε βέβαιοι κατά 95% ότι οι πραγματικές τιμές των α και β, θα βρίσκονται στα παραπάνω διαστήματα τα οποία ορίζονται από τις εκτιμώμενες τιμές α' και β' αντίστοιχα.

Ο υπολογισμός των Standard Errors μας επιτρέπει εκτός από την εκτίμηση των 95% διαστημάτων εμπιστοσύνης για τις πραγματικές τιμές των α και β, να εξετάσουμε και το κατά πόσο οι τιμές του x σχετίζονται γραμμικά με τις τιμές του Y μέσω του ελέγχου μηδενικής υπόθεσης. Ο έλεγχος μηδενικής υπόθεσης είναι μία συνήθης μέθοδος που χρησιμοποιείται για έλεγχο γραμμικές σχέσης βασίζεται σε απλές υποθέσεις:

H_0: Δεν υπάρχει καμία σχέση μεταξύ x και Y , ή εναλλακτικά

H_α: Υπάρχει κάποια σχέση μεταξύ x και Y

Εάν υποτεθεί ότι ισχύει η συνθήκη H_0:, τότε θεωρούμε ότι $\beta = 0$ και επομένως ισχύει $Y = \alpha + \varepsilon$ και κατά συνέπεια η τιμή του Y είναι εντελώς ανεξάρτητη από την τιμή του x.

Εάν από την άλλη υποτεθεί ότι ισχύει η συνθήκη H_α, ότι δηλαδή υπάρχει κάποια γραμμική σχέση μεταξύ x και Υ, τότε η τιμή του β θα είναι διάφορη του μηδενός, επομένως $Y = a + \beta \cdot X + \varepsilon$

Για να μπορέσουμε να απαντήσουμε εάν αυτή η υπόθεση είναι σωστή, θα πρέπει να δούμε εάν η τιμή του β', δηλαδή η εκτιμώμενη βέλτιστη τιμή για το β που υπολογίσαμε, είναι αρκετά μακριά από το 0, ώστε να είμαστε σίγουροι ότι το β δεν έχει μηδενική τιμή. Το πόσο μακριά πρέπει να βρίσκεται η τιμή του β' από το μηδέν είναι κάτι σχετικό και εξαρτάται από την τιμή του SE(β'). Εάν η τιμή του SE(β') είναι μικρή, τότε ακόμα και μία σχετικά μικρή τιμή για το β' θα αρκούσε στο να φανεί ότι β ≠0 και επομένως υπάρχει κάποια γραμμική σχέση μεταξύ x και Υ. Εάν η τιμή του SE(β') είναι μεγάλη, αυτό σημαίνει πως χρειαζόμαστε μεγάλη τιμή για το β', ώστε να φαίνεται ότι αυτό απέχει πολύ από το μηδέν και επομένως είναι διάφορο του μηδενός και έτσι θα ισχύει η υπόθεση ότι υπάρχει γραμμικής σχέση μεταξύ x και Υ. Σε κάθε περίπτωση μία μεγάλη τιμή του για το β' είναι επιθυμητή προκειμένου να αποδειχθεί η ύπαρξη γραμμικότητας.

Για την ευκολότερη απόδειξη των παραπάνω, χρησιμοποιούμε τον όρο t-statistic ο οποίος εκφράζεται με τον τύπο:

$$t = \frac{\beta' - 0}{SE(\beta')}$$

Ο όρος αυτός εκφράζει το κατά πόσο η τιμή β' απέχει από το 0. Ξεκινώντας με τον έλεγχο μηδενικής υπόθεσης και θεωρώντας ότι β' = 0, είναι απλό να υπολογίσουμε την πιθανότητα του να εντοπίσουμε κάποια τιμή ίση ή μεγαλύτερη του |t|. Η πιθανότητα αυτή ορίζεται ως p-value και ερμηνεύεται ως εξής: μία μικρή τιμή της p-value υποδεικνύει πως είναι απίθανο να παρατηρήσουμε κάποια ουσιώδη σχέση μεταξύ κάποιων μεμονωμένων ανεξάρτητων (x) και εξαρτημένων μεταβλητών (Υ) λόγω τύχης και όχι στο σύνολό τους. Είναι δηλαδή απίθανο τα x και Υ να μη σχετίζονται γραμμικά μεταξύ τους στο σύνολό τους, αλλά να βρούμε κάποια ελάχιστα ζεύγη που σχετίζονται κατά τύχη. Κατά συνέπεια, μία μικρή τιμή της πιθανότητας p-value μας βοηθάει να συμπεράνουμε πως υπάρχει γραμμική σχέση μεταξύ των x και των αντίστοιχων Υ.

Επομένως σε περίπτωση που έχουμε μικρή τιμή για την p-value (κοντά στο 0), μπορούμε να απορρίψουμε τη μηδενική υπόθεση H_0, η οποία ξεκινάει με την παραδοχή ότι δεν υπάρχει καμία σχέση ανάμεσα στην ανεξάρτητες και τις εξαρτημένες τιμές. Ένα τυπικό άνω όριο για την τιμή της p-value ώστε να απορρίψουμε με ασφάλεια τη μηδενική υπόθεση, είναι το 1%, δηλαδή εάν η τιμή της p-value είναι μικρότερη από 0,001, τότε μπορούμε να οδηγηθούμε στο συμπέρασμα πώς δεν ισχύει η μηδενική υπόθεση.

2.5.3 Αξιολόγηση μοντέλου παλινδρόμησης βάσει προσαρμογής

Σε μία οποιαδήποτε μελέτη ελέγχου γραμμικότητας δεν αρκεί μόνο να αποδειχθεί ότι απορρίπτεται η μηδενική υπόθεση. Βρισκόμαστε στο σημείο όπου έχουμε καταλήξει στο συμπέρασμα πως τα 2 παραπάνω μεγέθη x-Y σχετίζονται γραμμικά μεταξύ τους και μάλιστα έχουμε κατασκευάσει ένα θεωρητικό γραμμικό μοντέλο (δηλαδή τις τιμές α' και β') που προσπαθεί να εκφράζει όσο καλύτερα μπορεί τη γραμμική αυτή σχέση. Το επόμενο στάδιο της ανάλυσης μας είναι να αξιολογήσουμε το μοντέλο μας, δηλαδή να εξετάσουμε το κατά πόσο τα σημεία της θεωρητικής αυτής ευθείας προσεγγίζουν τις αντίστοιχες πραγματικές τιμές των δεδομένων μας.

Η ακρίβεια ενός γραμμικού μοντέλου που έχει προκύψει με τη μέθοδο της γραμμικής παλινδρόμησης, αξιολογείται πρακτικά με τη βοήθεια 2 όρων οι οποίοι σχετίζονται μεταξύ τους:

- Residual Standard Error (RSE)

Παρουσιάστηκε προηγουμένως και υπολογίζεται από τα σφάλματα ανάμεσα στις πραγματικές τιμές των δεδομένων μας και τις αντίστοιχες τιμές του γραμμικού μοντέλου και εκφράζεται μαθηματικά από τη σχέση:

$$RSE = \sqrt{\frac{1}{n-2}RSS} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}.$$

Ο όρος αυτός υποδηλώνει πως παρόλο που έχουμε υπολογίσει μία βέλτιστη τιμή για το α και μία αντίστοιχα για το β, πρακτικά οι τιμές της εξαρτημένης μεταβλητής του γραμμικού μοντέλου δε συμπίπτουν με τις αντίστοιχες πραγματικές τιμές, για δεδομένες τιμές της ανεξάρτητης μεταβλητής. Αυτό κατ' επέκταση σημαίνει πως με τη βοήθεια του γραμμικού μοντέλου μπορούμε να προβλέπουμε την τιμή του Y για μια δεδομένη τιμή του x, χωρίς ωστόσο η πρόβλεψη μας να είναι απόλυτα ακριβής. Ο όρος RSE στην ουσία υπολογίζει τον μέσο όρο της απόκλισης των τιμών που βρίσκονται πάνω στην ευθεία του μοντέλου μας, από τις αντίστοιχες πραγματικές τιμές του Y.

Με άλλα λόγια το RSE υφίσταται λόγω του ότι στην πράξη, όσο ακριβές και να είναι το γραμμικό μοντέλο, οι εκτιμώμενες τιμές της εξαρτημένης μεταβλητής ποτέ δεν συμπίπτουν ακριβώς με τις αντίστοιχες πραγματικές τιμές και κατά συνέπεια η πρόβλεψη για μελλοντικές τιμές της εξαρτημένης μεταβλητής όσο καλή κι αν είναι, στην πράξη δε συμπίπτει ακριβώς με την πραγματική τιμή που θα έρθει στο μέλλον. Προφανώς, ελλείψει μίας τέλειας πρόβλεψης, σκοπός μας είναι το γραμμικό μοντέλο να έχει RSE όσο το δυνατόν πλησιέστερα στο 0, δηλαδή οι τιμές που προβλέπει το μοντέλο να είναι όσο το δυνατόν πιο κοντά με τις αντίστοιχες πραγματικές τιμές.

- R^2 statistic

Ο όρος RSE όπως εξηγήθηκε, είναι μια απόλυτη τιμή που ουσιαστικά εκφράζει την απουσία μιας τέλειας προσαρμογής των τιμών που προβλέπονται από το γραμμικό μοντέλο με τις αντίστοιχες πραγματικές τιμές. Ωστόσο, καθώς είναι όρος που εκφράζεται στη μονάδα μέτρησης του Y, δεν είναι ξεκάθαρο ότι πάντα η τιμή του θα μας βοηθά να κρίνουμε αν το μοντέλο κάνει καλές προβλέψεις ή όχι. Είπαμε ότι σε γενικές γραμμές θέλουμε να έχει όσο το δυνατόν μικρότερη τιμή, όμως αυτό είναι κάποιες φορές σχετικό. Εξαρτάται καθαρά από τη φύση του προβλήματος για το αν μία μεγάλη τιμή του RSE μπορεί να είναι αποδεκτή ή αντίστροφα μία σχετικά μικρή τιμή του να βγαίνει εκτός αποδεκτών ορίων. Για το λόγο είναι προτιμότερο να μελετάμε τον όρο R^2 ο οποίος εκφράζεται σε ποσοστό, δηλαδή λαμβάνει τιμές από 0 έως 1 και επομένως μας βοηθά να αξιολογήσουμε πιο σωστά το γραμμικό μας μοντέλο, εφόσον δεν εξαρτάται από τη μονάδα μέτρησης της μεταβλητής Y.

Η τιμή του R^2 υπολογίζεται από τον τύπο:

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

όπου $TSS = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$.

Ο όρος R^2 θα μπορούσαμε να πούμε ότι εκφράζει την ποιότητα της διακύμανσης των εκτιμώμενων τιμών του Y, βάσει των τιμών της ανεξάρτητης μεταβλητής x. Μία τιμή του R^2 κοντά στο 1, υποδηλώνει πως ένα μεγάλο ποσοστό της συνολικής διακύμανσης στην εξαρτημένη μεταβλητή, δικαιολογείται από τις δεδομένες τιμές του x. Αντιθέτως, μία τιμή του R^2 κοντά στο 0 ότι η παλινδρόμηση δεν μπορεί να δικαιολογήσει τη διακύμανση ανάμεσα στις διάφορες τιμές του Y, βάσει των τιμών του x, είτε διότι το γραμμικό μοντέλο που έχουμε εξάγει δεν είναι σωστό είτε γιατί το RSE και κατ' επέκταση το RSS έχουν μεγάλη τιμή.

Σημείωση: Γενικά, όταν θέλουμε να διερευνήσουμε την ύπαρξη γραμμικής σχέσης ανάμεσα σε μία εξαρτημένη μεταβλητή Y και μία ανεξάρτητη μεταβλητή X, χρησιμοποιούμε την έννοια της συσχέτισης (correlation) η οποία δίνεται από τον παρακάτω τύπο:

$$Cor(X,Y) = \frac{\sum_{i=1}^{n}[(X_i - \bar{X}) \cdot (Y_i - \bar{Y})]}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Όμως αποδεικνύεται ότι στην περίπτωση της γραμμικής παλινδρόμησης και μόνο, ο όρος $Cor^2(X,Y)$ ισούται με τον όρο R^2. Επομένως όποιον από τους δύο όρους και να υπολογίσουμε για την αξιολόγηση του μοντέλου που προέκυψε από απλή παλινδρόμηση, έχουμε το ίδιο αποτέλεσμα. Ωστόσο στην περίπτωση της πολλαπλής παλινδρόμησης που θα αναλύσουμε στην επόμενη ενότητα, όπου εξετάζεται η σχέση που συνδέει μία εξαρτημένη μεταβλητή με περισσότερες από μία ανεξάρτητες μεταβλητές, προφανώς δεν μπορεί να γίνει χρήση της παραπάνω σχέσης η οποία περιλαμβάνει μόνο μία ανεξάρτητη μεταβλητή. Για το λόγο αυτό θα πρέπει να υπολογίσουμε αναγκαστικά μόνο την τιμή του R^2

2.5.4 Πολλαπλή Γραμμική Παλινδρόμηση

Όπως αναφέρθηκε προηγουμένως, σε περιπτώσεις που απαιτούνται περισσότερες από μια ανεξάρτητες μεταβλητές, το μοντέλο γραμμικής παλινδρόμησης μπορεί να γενικευθεί μέσω της τεχνικής της πολλαπλής παλινδρόμησης προκειμένου να συμπεριληφθούν σε αυτό όλες οι μεταβλητές που επηρεάζουν την τιμή της μεταβλητής πρόβλεψης.

Στην πολλαπλή παλινδρόμηση υπάρχει εξαρτημένη μεταβλητή, της οποίας η τιμή πρέπει να προβλεφθεί βάσει των τιμών δύο ή περισσοτέρων ανεξάρτητων μεταβλητών. Έτσι, η γενική μορφή της πολλαπλής παλινδρόμησης είναι:

$$Y = b_0 + b_1X_1 + b_2X_2 + \cdots + b_kX_k + e$$

Στην παραπάνω εξίσωση, οι μεταβλητές $X_1, X_2, \ldots, X_k$ εκφράζουν τις ανεξάρτητες μεταβλητές ενώ η μεταβλητή Υ εκφράζει την εξαρτημένη μεταβλητή. Οι συντελεστές b_0 , b_1 , b_2 , . . . , b_k είναι σταθερές παράμετροι ενώ το e δηλώνει τον τυχαίο παράγοντα, ο οποίος θεωρείται κανονικά κατανεμημένος γύρω από το μηδέν.

Η εξίσωση της πολλαπλής γραμμικής παλινδρόμησης είναι γραμμική ως προς τους συντελεστές. Κάθε συντελεστής b_i έχει εκθέτη ίσο με τη μονάδα, γεγονός που εξασφαλίζει την γραμμικότητα. Οι τιμές των συντελεστών αυτών μπορούν να προκύψουν με εφαρμογή της μεθόδου ελαχίστων τετραγώνων. Το σχήμα της συνάρτησης που συνδέει τις ανεξάρτητες με τις εξαρτημένες μεταβλητές δεν είναι εύκολο να περιγραφεί. Στην περίπτωση μιας ανεξάρτητης μεταβλητής, το σχήμα της συνάρτησης είναι μια ευθεία γραμμή. Στην περίπτωση δύο ανεξάρτητων μεταβλητών, η Y παριστάνεται στο επίπεδο που σχηματίζουν οι δύο ανεξάρτητες μεταβλητές. Στην περίπτωση περισσοτέρων από δύο μεταβλητές, η Y παριστάνεται σε υπερεπίπεδο (επιφάνεια με περισσότερες από δύο διαστάσεις).

Στην πράξη, η διαδικασία της πολλαπλής παλινδρόμησης αποσκοπεί στον προσδιορισμό των αγνώστων παραμέτρων συντελεστές b_0 , b_1 , b_2 , . . . , b_k του μοντέλου και της διακύμανσης του τυχαίου παράγοντα, δεδομένου ενός συγκεκριμένου συνόλου δεδομένων, στο οποίο μπορεί να εφαρμοστεί η μέθοδος ελαχίστων τετραγώνων. Συνεπώς η μορφή του στατιστικού μοντέλου παλινδρόμησης είναι :

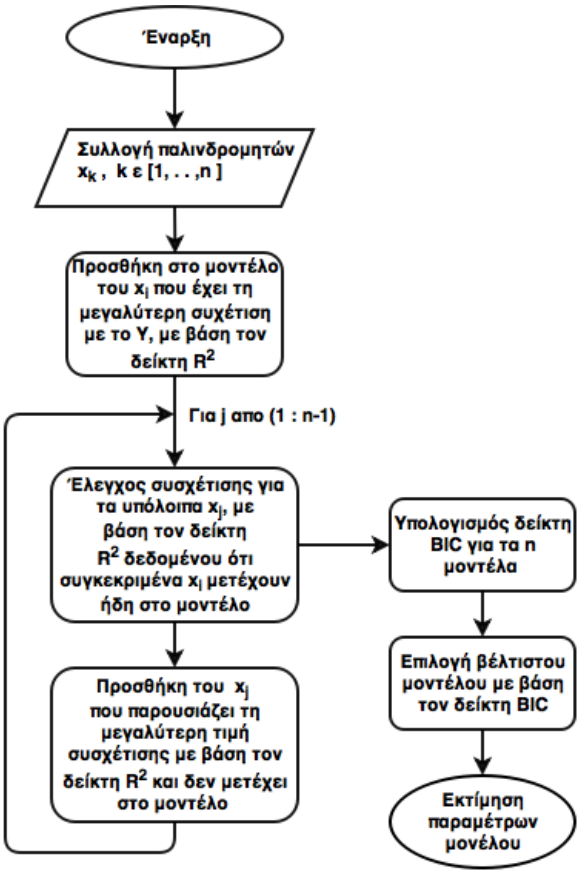
$$Y_i = b_0 + b_1X_{1,i} + b_2X_{2,i} + \cdots + b_kX_{k,i} + e_i$$

2.5.5 Διαδικασία επιλογής ανεξάρτητων μεταβλητών

Υπάρχουν αρκετές μεθοδολογίες, οι οποίες προσεγγίζουν με βηματικό τρόπο την ανάπτυξη ενός μοντέλου παλινδρόμησης, έτσι ώστε να μην είναι απαραίτητος ο εξαντλητικός έλεγχος όλων των διαθέσιμων δεδομένων. Τέτοιου είδους μεθοδολογίες είναι και αυτές που ανήκουν στην οικογένεια της κλασσικής βηματικής ανάλυσης παλινδρόμησης (stepwise regression analysis). Πιο αναλυτικά, υπάρχουν και χρησιμοποιούνται σχετικά συχνά σε διάφορες εφαρμογές, η προς τα εμπρός επιλογή (forward selection), η προς τα πίσω απαλοιφή (backward elimination) και η βηματική παλινδρόμηση (stepwise regression).

2.5.5.1 Forward selection

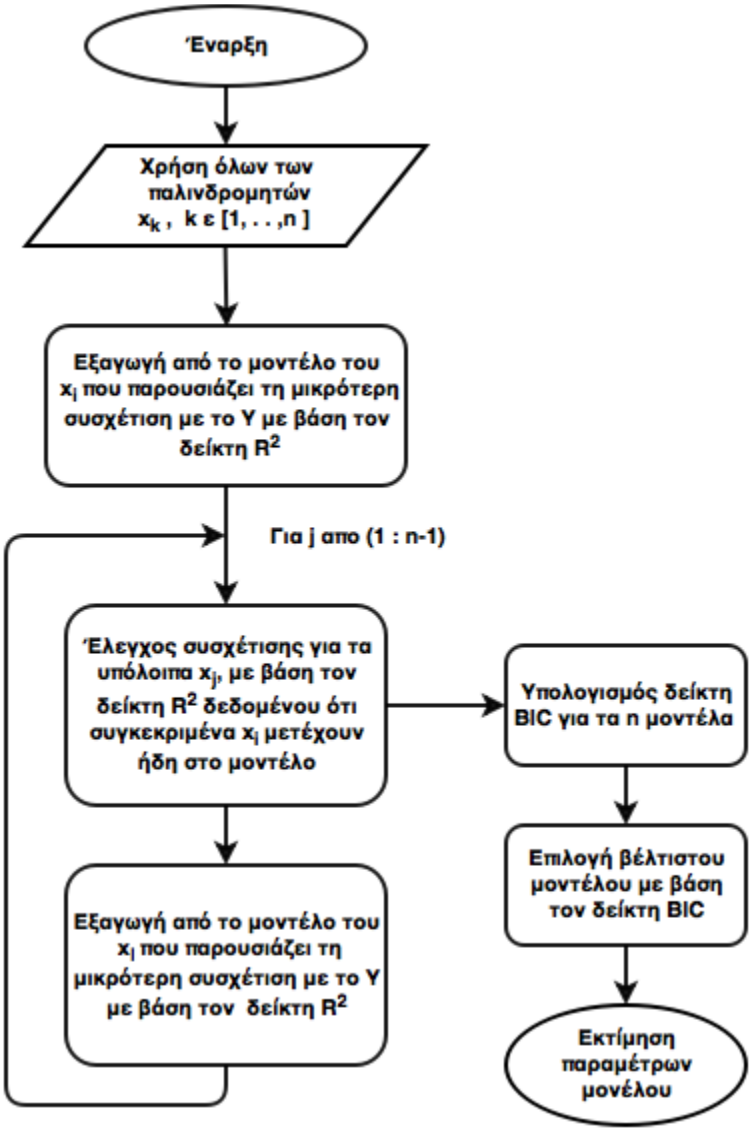
Η προς τα εμπρός επιλογή ξεκινά από τη μη συμμετοχή καμιάς ανεξάρτητης μεταβλητής στο μοντέλο παλινδρόμησης και βήμα- βήμα εξετάζεται η συμμετοχή ή όχι κάποιας από τις διαθέσιμες μεταβλητές σε αυτό. Η εισαγωγή αυτή στηρίζεται, σε κάθε βήμα (εκτός του αρχικού), στο γεγονός ότι είναι γνωστό ότι ήδη κάποιες συγκεκριμένες μεταβλητές συμμετέχουν σε αυτό και η συμμετοχή μιας άλλης πραγματοποιείται μέσω στατιστικών υποθέσεων, οι οποίες επιβεβαιώνονται ή όχι με βάση ένα προκαθορισμένο κατώφλι αποδοχής (π.χ. χρησιμοποίηση του δείκτη R^2 ή της στατιστικής F).



Σχήμα 2.6: Αλγόριθμος Forward Selection

2.5.5.2 Backward selection

Στην περίπτωση της απαλοιφής προς τα πίσω, λειτουργούμε με την αντίθετη λογική της προς τα εμπρός επιλογής. Η διαδικασία ξεκινά με την υπόθεση ότι όλες οι διαθέσιμες μεταβλητές συμμετέχουν στο μοντέλο παλινδρόμησης, οπότε σε κάθε επόμενο βήμα, εξετάζεται, πάλι με στατιστικές υποθέσεις και ένα προκαθορισμένο κατώφλι απόρριψης, η εξαγωγή μιας μεταβλητής από το μοντέλο.



Σχήμα 2.7: Αλγόριθμος Backward Selection

2.5.6 Επιλογή βέλτιστου μοντέλου

Έχοντας σχηματίσει n μοντέλα παλινδρόμησης με διαφορετικό πλήθος παλινδρομητών (που επιλέγονται σύμφωνα με κάποια από τις διαδικασίες επιλογής παλινδρομητών που περιγράφηκαν στην προηγούμενη ενότητα), χρειαζόμαστε κάποιον δείκτη σχετικής ποιότητας των n στατιστικών μοντέλων για ένα δοσμένο σύνολο δεδομένων, ο οποίος θα μας οδηγήσει στην επιλογή του βέλτιστου μοντέλου παλινδρόμησης. Η σχετική ποιότητα των στατιστικών μοντέλων δίνεται με χρήση κάποιου κριτήριου πληροφορίας. Ακολουθεί η αναλυτική περιγραφή δύο τέτοιων κριτηρίων: του κριτηρίου πληροφορίας Akaike (AIC) και του κριτήριου πληροφορίας Bayes (BIC). Προτιμούμε τα συγκεκριμένα αντί του δείκτη R^2 καθώς μία τέτοια προσέγγιση απαιτεί μεγάλο πλήθος υπολογισμών.

2.5.6.1 Κριτήριο Πληροφορίας AIC (Akaike Information Criterion)

Ο δείκτης AIC αποτελεί ένα δείκτη σχετικής ποιότητας των στατιστικών μοντέλων για ένα δοσμένο σύνολο πληροφοριών. Για δεδομένη λίστα μοντέλων για τα δεδομένα, ο δείκτης αυτός εκτιμά την απόδοση κάθε μοντέλου σχετικά με κάθε άλλο μοντέλο.

Βασίζεται στη θεωρία δεδομένων και προσφέρει μια σχετική εκτίμηση των χαμένων πληροφοριών όταν χρησιμοποιείται ένα συγκεκριμένο δοσμένο μοντέλο για την αναπαράσταση της διαδικασίας παραγωγής δεδομένων. Συμβιβάζει την καλή δυνατότητα εφαρμογής ενός μοντέλου και την πολυπλοκότητα του. Ως σχετικός δείκτης εξετάζει τη σχετική και όχι την απόλυτη απόδοση των μοντέλων. Δίνεται από την ακόλουθη σχέση :

$$AIC = 2k - 2ln(L)$$

,όπου:

L : μέγιστη τιμή συνάρτησης πιθανότητας και

k : αριθμός εκτιμώμενων παραμέτρων του μοντέλου

Όσο μεγαλύτερη είναι η τιμή του δείκτη AIC, τόσο καλύτερη σχετική απόδοση παρουσιάζει το μοντέλο. Συνεπώς, ο AIC :

- «επιβραβεύει» την καλή εφαρμογή του μοντέλου (σύμφωνα με τη συνάρτηση πιθανότητας)
- περιλαμβάνει «ποινή», η οποία είναι αύξουσα συνάρτηση του αριθμού των εκτιμώμενων παραμέτρων

Η ποινή αυτή είναι ανασταλτικός παράγοντας για το overfitting. Σημειώνεται ότι, όσο αυξάνεται ο αριθμός των παραμέτρων στο μοντέλο, σχεδόν πάντα βελτιώνεται η δυνατότητα εφαρμογής του μοντέλου.

Έστω ότι μια άγνωστη διαδικασία f παράγει δεδομένα κι έστω ότι έχουμε στη διάθεση μας δύο υποψήφια μοντέλα g_1 και g_2 .Εάν γνωρίζαμε την f, θα μπορούσαμε να διαλέξουμε το μοντέλο που θα εμφάνιζε τις ελάχιστες χαμένες πληροφορίες (ελάχιστη απόκλιση), έχοντας πρώτα υπολογίσει τις αποκλίσεις Kullback-Leibel μέσω του τύπου :

$$D_{KL_i} = \int_{-\infty}^{\infty} log \frac{f(x)}{g_i(x)} dx$$

Μη γνωρίζοντας την f, δεν είναι δυνατόν να στηριχτεί με βεβαιότητα ποιο μοντέλο είναι καταλληλότερο. Με βάση τον δείκτη AIC, ωστόσο, πόσο περισσότερη ή λιγότερη πληροφορία χάνεται μεταξύ των g_1 και g_2.

Εφαρμογή AIC

Έστω R υποψήφια μοντέλα, τα οποία ελέγχονται πάνω σε συγκεκριμένα δεδομένα ως προς την ελάχιστη απώλεια πληροφοριών. Έστω $AIC_1, AIC_2, \ldots, AIC_R$ οι τιμές του δείκτη για τα μοντέλα και AIC_{min} η ελάχιστη τιμή του δείκτη που εμφανίζεται. Τότε, ο όρος $e^{\frac{AIC_{min}-AIC_i}{2}}$ αποτυπώνει τη σχετική πιθανότητα το i-μοντέλο να ελαχιστοποιεί την απώλεια πληροφοριών.

Αν το μοντέλο είναι γραμμικό και τα σφάλματα ακολουθούν κανονική κατανομή, ισχύει :

$$AIC_L = AIC + \frac{2k(k+1)}{n-k+1}$$

,όπου:

L : μέγιστη τιμή συνάρτησης πιθανότητας και

k : αριθμός εκτιμώμενων παραμέτρων του μοντέλου

Με βάση τον παραπάνω τύπο, γίνεται αντιληπτό, ότι όταν το n δεν είναι πολλές φορές μεγαλύτερο του k^2 και για κάθε επιπλέον παράμετρο, αυξάνεται η πιθανότητα να επιλεχθεί μοντέλο που έχει πάρα πολλές παραμέτρους (overfitting). Υποθέτοντας ότι τα σφάλματα ακολουθούν κανονική κατανομή, προκύπτει :

$$AIC = n \cdot ln(MSE) + 2 \cdot k$$

2.5.6.2 Κριτήριο Πληροφορίας BIC (Bayesian Information Criterion)

Ο δείκτης BIC αποτελεί ένα δείκτη επιλογής μοντέλου για ένα δοσμένο σύνολο πληροφοριών. Βασίζεται στη θεωρία της συνάρτησης πιθανότητας. Όσο μικρότερος υπολογιστεί ο δείκτης για ένα μοντέλο, τόσο καταλληλότερο θεωρείται το μοντέλο. Η βασική διαφορά του συγκεκριμένου δείκτη με τον δείκτη AIC είναι ότι εισάγει μεγαλύτερη ποινή για κάθε επιπλέον

παράμετρο που λαμβάνεται υπόψιν στο μοντέλο πρόβλεψης. Υπολογίζεται με βάση τον ακόλουθο τύπο:

$$BIC = -2 \cdot ln(\hat{L}) + k \cdot ln(n)$$

,όπου:

L : μέγιστη τιμή συνάρτησης πιθανότητας και

k : αριθμός εκτιμώμενων παραμέτρων του μοντέλου

Εφαρμογή BIC

Το ολοκλήρωμα της συνάρτησης πιθανότητας $P(x|\theta, m)$ μετρά την εκ των προτέρων κατανομή πιθανότητας $P(\theta|m)$ επί των παραμέτρων θ του μοντέλου m για δεδομένο x και υπολογίζεται ως εξής :

$$BIC = -2\ln\big(P(x|m)\big) \approx 2\ln\big(\hat{L}\big) + k \cdot (\ln(n) - 2\ln(2\pi))$$

,όπου:

x : παρατηρήσεις

θ : παράμετροι μοντέλου

n : μέγεθος δείγματος

k : αριθμός εκτιμώμενων παραμέτρων του μοντέλου

L : μέγιστη τιμή συνάρτησης πιθανοφάνειας μοντέλου $\mathcal{L}$ = P (x|θ,m) και

θ : τιμές παραμέτρων που μεγιστοποιούν τη συνάρτηση πιθανότητας

Περιορισμοί :

- $n >> k$
- Ο δείκτης BIC δεν μπορεί να διαχειριστεί σύνθετες και μεγάλες συλλογές μοντέλων

Υποθέτοντας ότι τα σφάλματα ακολουθούν κανονική κατανομή, προκύπτει :

$$BIC = n \cdot ln(MSE) + k \cdot ln(n)$$

2.5.7 Υπολογισμός συντελεστών παλινδρόμησης

Έστω ότι το μοντέλο παλινδρόμησης περιέχει k ανεξάρτητες μεταβλητές. Τότε:

$$Y = b_0 + b_1X_1 + b_2X_2 + \cdots + b_kX_k + e$$

Για κάθε διαφορετικό διάνυσμα παρατηρήσεων χρησιμοποιείται μια ξεχωριστή τιμή του δείκτη i, όπως παρατηρείται στην παρακάτω εξίσωση :

$$Y_i = b_0 + b_1X_{1,i} + b_2X_{2,i} + \cdots + b_kX_{k,i} + e_i = \hat{Y}_i + e_i$$

,όπου $\hat{Y}_i$ είναι μια εκτίμηση της τιμής της μεταβλητής Y, η οποία βασίζεται στις τιμές των X_1 , $X_2, \ldots, X_k$. Επομένως, το σφάλμα e_i ισούται με : $e_i = Y_i - \hat{Y}_i$.

Εφαρμόζοντας τη μέθοδο ελαχίστων τετραγώνων, υπολογίζουμε το ελάχιστο άθροισμα των τετραγώνων των σφαλμάτων e_i ,δηλαδή :

$$b_0, b_1, \ldots, b_k \mid \min \left[\sum_{i=1}^{n} {e_i}^2\right]$$

Όμως,

$\sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n}(Y_i - b_0 - b_1X_{1,i} - b_2X_{2,i} - \cdots - b_kX_{k,i})^2$

Προκειμένου να προσδιορίσουμε τους άγνωστους συντελεστές b_0 , b_1 , . . , b_k, οι οποίοι ελαχιστοποιούν την παραπάνω ποσότητα, υπολογίζουμε τις μερικές παραγώγους αυτής για κάθε έναν από τους συντελεστές και θέτοντας αυτές ίσες με το μηδέν, λύνουμε ένα γραμμικό σύστημα k εξισώσεων με k αγνώστους.

2.5.8 Πολλαπλή συσχέτιση και συντελεστής R^2

Η συσχέτιση ανάμεσα στην πραγματική τιμή της μεταβλητής Y και στην υπολογισμένη τιμή Ŷ με βάση την εξίσωση παλινδρόμησης δίνεται από την εξίσωση :

$$R_{Y\hat{Y}} = \frac{n\cdot\sum_{i=1}^{n}(Yi\cdot\hat{Y}i) - (\sum_{i=1}^{n}Yi)\cdot(\sum_{i=1}^{n}\hat{Y}i)}{\sqrt{n\cdot\sum_{i=1}^{n}Yi - (\sum_{i=1}^{n}Yi)^2}\cdot\sqrt{n\cdot\sum_{i=1}^{n}\hat{Y}i - (\sum_{i=1}^{n}\hat{Y}i)^2}}$$

Το τετράγωνο του $R_{Y\hat{Y}}$ καλείται coefficient of determination. Το $R_{Y\hat{Y}}$ είναι γνωστό ως συντελεστής πολλαπλής συσχέτισης και εκφράζει τη συσχέτιση ανάμεσα στην εξαρτημένη μεταβλητή Υ και την εκτίμηση της $\hat{Y}$ με βάση τις ανεξάρτητες μεταβλητές. Για τον υπολογισμό του R^2 χρησιμοπιείται η ίδια εξίσωση που χρησιμοποιείται και στην περίπτωση της απλής γραμμικής παλινδρόμησης :

$$R^2=\frac{\text{ερμηνευθείσα διακύμανση των τιμών Y}}{\text{συνολική διακύμανση των τιμών Y}}=\frac{\sum_{i=1}^{n}(\hat{Y}i-\bar{Y})^2}{\sum_{i=1}^{n}(Yi-\bar{Y})^2}$$

Ωστόσο, στην προηγούμενη εξίσωση δε λαμβάνεται υπ' όψιν ο αριθμός των ανεξάρτητων μεταβλητών και ο αριθμός του συνόλου των παρατηρήσεων. Προκειμένου να ξεπεραστεί αυτό το πρόβλημα, υπολογίζεται ένας «διορθωμένος» συντελεστής R^2 από την εξίσωση:

$$\hat{R}^2 = 1-(1-R^2)\cdot\frac{n-1}{n-k-1}$$

Ο νέος διορθωμένος συντελεστής $\hat{R}^2$ εκφράζει το ποσοστό της διασποράς της μεταβλητής Υ που ερμηνεύεται από τις ανεξάρτητες μεταβλητές X_1 , X_2, . . . , X_k . Η διαφορά (n-1) εκφράζει τους συνολικούς βαθμούς ελευθερίας της συνολικής διακύμανσης του μοντέλου ενώ ο όρος (n-k-1) εκφράζει τους βαθμούς ελευθερίας της ερμηνευθείσας διακύμανσης.

2.5.9 Ο στατιστικός δείκτης F (F-test)

Ο στατιστικός δείκτης F αποτελεί ένα μέτρο της σημαντικότητας του μοντέλου παλινδρόμησης και υπολογίζεται από αντίστοιχες, όπως στην απλή παλινδρόμηση, εξισώσεις :

$$F=\frac{\frac{\sum_{i=1}^{n}(\hat{Y}i-\bar{Y})^2}{k}}{\frac{\sum_{i=1}^{n}(Yi-\hat{Y}i)^2}{n-k-1}}=\frac{\frac{R^2}{k}}{\frac{1-R^2}{n-k-1}}$$

Αν η μη ερμηνευθείσα διακύμανση (διακύμανση σφαλμάτων) είναι μεγάλη, τότε αντίστοιχα και ο παρονομαστής της παραπάνω εξίσωσης είναι μεγάλος και ο δείκτης F γίνεται μικρότερος, γεγονός που σημαίνει ότι το μοντέλο παλινδρόμησης δεν είναι επιτυχημένο. Αντίθετα, αν η ερμηνευθείσα διακύμανση, που εμφανίζεται στον αριθμητή, είναι σχετικά μεγαλύτερη, τότε και ο δείκτης F είναι μεγαλύτερος.

Όπως φαίνεται και στο δεύτερο σκέλος της παραπάνω εξίσωσης υπάρχει στενή σχέση ανάμεσα στο συντελεστή R^2 και στο στατιστικό δείκτη F.

2.5.10 Ο στατιστικός δείκτης t (t-test)

Αφού πρώτα ελεγχθεί η συνολική σημαντικότητα του μοντέλου παλινδρόμησης, είναι συχνά χρήσιμο να εξεταστεί η σημαντικότητα καθενός από τους συντελεστές παλινδρόμησης. Στην περίπτωση της πολλαπλής παλινδρόμησης, ο στατιστικός δείκτης t για κάθε συντελεστή αποτελεί εκτίμηση της σημαντικότητας του συντελεστή αυτού με την παρουσία όλων των άλλων ανεξάρτητων μεταβλητών. Για κάθε συντελεστή παλινδρόμησης b_j, μπορεί να οριστεί ένα τυπικό σφάλμα ως μέτρο της σταθερότητας του συντελεστή και με βάση την υπόθεση της κανονικότητας του μοντέλου παλινδρόμησης, ο δείκτης t ακολουθεί την t-κατανομή με (n-k-1) βαθμούς ελευθερίας και δίνεται από τον παρακάτω τύπο:

$$tb_j = \frac{b_j}{SE_{b_j}}$$

Υπολογίζοντας τον δείκτη t για κάθε συντελεστή του μοντέλου παλινδρόμησης, υπολογίζεται η σημαντικότητα του, μέσω της σύγκρισης της τιμής του συντελεστή αυτού με το μηδέν, τιμή για την οποία η αντίστοιχη ανεξάρτητη μεταβλητή δε συνεισφέρει στην πρόβλεψη της εξαρτημένης μεταβλητής Y, με δεδομένη την παρουσία των άλλων ανεξάρτητων μεταβλητών.

Στο σημείο αυτό αξίζει να σημειωθούν δύο βασικά θέματα σχετικά με τους στατιστικούς δείκτες των συντελεστών παλινδρόμησης. Πρώτον, η σταθερότητα των συντελεστών παλινδρόμησης εξαρτάται από τη συσχέτιση των ανεξάρτητων μεταβλητών. Για δύο ανεξάρτητες μεταβλητές X_1 και X_2, όσο μεγαλύτερη είναι η μεταξύ τους συσχέτιση τόσο πιο ασταθείς θα είναι οι δυο συντελεστές b_1 και b_2 που θα υπολγιστούν για τις μεταβλητές αυτές. Δεύτερον, στην πρακτική μορφή του μοντέλου παλινδρόμησης οι συντελεστές b_0 έως b_k είναι όλοι τυχαίες μεταβλητές, δηλαδή οι τιμές τους κυμαίνονται από δείγμα σε δείγμα, ενώ ακολουθούν μια κατανομή πιθανότητας. Συνεπώς, είναι δυνατόν να υπολογισθούν οι συσχετίσεις ανάμεσα στους συντελεστές.

2.5.11 Έλεγχος υπολοιπόμενων σφαλμάτων (Residual Errors)

Η μελέτη των υπολοιπόμενων σφαλμάτων αναφέρεται στη μελέτη των σφαλμάτων προσαρμογής του μοντέλου στα πραγματικά δεδομένα και είναι ιδιαίτερα σημαντική για να αποφασιστεί η καταλληλότητα ενός μοντέλου πρόβλεψης. Αν τα σφάλματα είναι επαρκώς τυχαία, τότε το μοντέλο μπορεί να θεωρηθεί ικανοποιητικό. Αν τα σφάλματα ακολουθούν οποιοδήποτε πρότυπο, τότε το μοντέλο αδυνατεί να εκμεταλλευτεί όλη τη συστηματική πληροφορία που εμπεριέχεται στα δεδομένα. Μερικές από τις πιο πιθανές αναλύσεις των σφαλμάτων είναι οι ακόλουθες:

(α) διαγραμματική αναπαράσταση των σφαλμάτων για οπτική επισκόπηση και εύρεση της κατανομής που ακολουθούν

(β) μελέτη της αυτοσυσχέτισης των υπολοιπόμενων σφαλμάτων

(γ) υπολογισμός του στατιστικού δείκτη Durbin-Watson

Ο στατιστικός δείκτης Durbin-Watson δίνεται από την παρακάτω εξίσωση :

$$DW = \frac{\sum_{t=2}^{N}(e_t - e_{t-1})^2}{\sum_{t=1}^{N} e_t^2}$$

Στον αριθμητή εμφανίζονται οι διαφορές ανάμεσα σε διαδοχικά σφάλματα, ενώ ο παρονομαστής ισούται με το άθροισμα των τετραγωνικών σφαλμάτων. Σε κάθε συνδυασμό αριθμού παρατηρήσεων, αριθμού συντελεστών παλινδρόμησης και επιπέδου εμπιστοσύνης, αντιστοιχεί ένα ζευγάρι αριθμητικών τιμών DWL και DWU. Ανάλογα με την υπολογισμένη τιμή του στατιστικού δείκτη, τα σφάλματα του εκάστοτε μοντέλου παλινδρόμησης χαρακτηρίζονται ως:

- Σημαντικά θετικά συσχετισμένα, αν DW≤DW_L
- Ασυσχέτιστα, αν DW_U≤DW≤4-DW_L
- Σημαντικά αρνητικά συσχετισμένα, αν DW≥4-DW_L

Αν DW_L≤DW≤DW_U ή DW_U≤DW4-DW_L τότε δεν μπορεί να εξαχθεί ασφαλές συμπέρασμα από το στατιστικό δείκτη Durbin-Watson σχετικά με την τυχαιότητα των σφαλμάτων.

Κεφάλαιο 3: Πιθανοτικές Προβλέψεις

3.1 Εισαγωγή

Βασικό μειονέκτημα των σημειακών προβλέψεων είναι ότι δεν παρέχουν πληροφορία για τη διασπορά των παρατηρήσεων γύρω από τη μέση τιμή. Για παράδειγμα, θα ήταν χρήσιμη η εύρεση ενός επιπέδου ηλεκτρικής ισχύος από ηλιακή ενέργεια το οποίο με μεγάλη πιθανότητα δεν υπερβαίνεται. Για το σκοπό αυτό όμως, απαιτείται η επιπλέον πληροφορία της αβεβαιότητας που σχετίζεται με τις προβλέψεις της μελλοντικής παραγωγής ηλιακής ενέργειας. Πρόσφατες έρευνες έχουν εστιάσει στο συσχετισμό εκτιμήσεων αβεβαιότητας με σημειακές προβλέψεις, λαμβάνοντας υπόψη τη μορφή πιθανοτικών προβλέψεων, δείκτες ρίσκου ή σενάρια βραχυπρόθεσμης παραγωγής της ηλεκτρικής ισχύος.

Πολλές μελέτες έχουν αναδείξει τα πλεονεκτήματα που προκύπτουν από την πληροφορία της αβεβαιότητας των προβλέψεων. Στην [29], το βέλτιστο επίπεδο εφεδρικής παραγωγής υπολογίζεται με χρήση της αβεβαιότητας των αιολικών προβλέψεων. Στις [27,41] ερευνάται ο βέλτιστος συντονισμός υδροπαραγωγής και αιολικής ενέργειας χρησιμοποιώντας τις προβλέψεις ενός μοντέλου πιθανοτικής πρόβλεψης. Στη [12] αναδεικνύονται τα οικονομικά οφέλη που προκύπτουν όταν οι στρατηγικές προσφοράς βασίζονται σε προβλεπόμενες συναρτήσεις πυκνότητας πιθανότητας της αιολικής ισχύος σε αγορές ηλεκτρικής ενέργειας με ορίζοντα ημέρας (day-ahead electricity markets).

Τα μοντέλα πιθανοτικής πρόβλεψης αιολικής ισχύος χρησιμοποιούν μετεωρολογικά σύνολα (meteorological ensembles) που αποκτώνται από ένα μετεωρολογικό μοντέλο υψηλής ανάλυσης [31],[32] ή από παραδοσιακές χρονοσειρές αιολικής ισχύος και αριθμητικών προβλέψεων καιρού (NWP). Στη δεύτερη περίπτωση εφαρμόζεται μια στατιστική μέθοδος για να εκτιμηθούν οι κατανομές πρόβλεψης στη μορφή εκατοστημορίων ή διαστημάτων. Έτσι, στην [33] εφαρμόζεται μια τοπική γραμμική παλινδρόμηση εκατοστημορίων για τον υπολογισμό δέκα διαφορετικών εκατοστημορίων. Παρόμοια μέθοδος χρησιμοποιείται στη μελέτη [34] όπου γίνεται συνδυασμός με ομαλές πολυωνυμικές συναρτήσεις για την εκτίμηση του σφάλματος πρόβλεψης του μοντέλου WWPT [35]. Στην εργασία [36] μια μέθοδος που παρέχει τη συνεχή συνάρτηση πυκνότητας πιθανότητας της αιολικής ισχύος προτείνεται βάσει των εκτιμητριών πυκνότητας πυρήνα (kernel density estimators). Παρόμοια μεθοδολογία ακολουθείται και στην [37], όπου χρησιμοποιούνται χρονικά προσαρμοζόμενοι πυρήνες. Στην [38] τα διαστήματα πρόβλεψης εκτιμώνται με προσαρμόσιμη αναδειγματοληψία, όπου γίνεται χρήση της πληροφορίας αβεβαιότητας ενός κατάλληλου δείκτη ρίσκου που προκύπτει από συνεχόμενες προβλέψεις καιρού. Ορισμένα μοντέλα χρησιμοποιούν τις σημειακές προβλέψεις που προκύπτουν από κάποιο μοντέλο σημειακής πρόβλεψης της αιολικής παραγωγής και αντλούν την πληροφορία της αβεβαιότητας από τις χρονοσειρές NWP [34,40].

Οι Pinson και Kariniotakis (2004) ορίζουν ένα δείκτη μετεωρολογικού ρίσκου (meteo-risk index-MRI) για τη μέτρηση της διασποράς των προβλέψεων καιρού σε δεδομένο χρόνο. Αυτό επιτυγχάνεται μετρώντας τη διακύμανση των προβλέψεων προηγούμενων αναβαθμίσεων του παρόχου. Μια σχετική προσέγγιση είναι η συσχέτιση της διάδοσης των συνόλων (ensembles) αιολικής ισχύος με το σφάλμα πρόβλεψης ελέγχου της αιολικής ισχύος. Από έρευνες σχετικά με τον τρόπο που μπορούν οι επαγγελματίες να χρησιμοποιούν πιθανοτικές προβλέψεις για τη λήψη αποφάσεων, φαίνεται ότι τέτοιοι δείκτες ρίσκου μπορεί να είναι επωφελείς ώστε να εκφραστεί το επίπεδο της αβεβαιότητας στην πρόβλεψη. Η προσέγγιση αυτή αναπτύχθηκε περαιτέρω στη μελέτη των Pinson et al. (2009a), όπου ως είσοδοι θεωρούνται διαφορετικοί τύποι συνόλου προβλέψεων μετεωρολογικών μεταβλητών. Αυτές περιλαμβάνουν τις ECMWF προβλέψεις συνόλου, καθώς επίσης και μια εναλλακτική μέσου όρου χρονικής υστέρησης (lagged-average) που αποτελείται από τις χρονικά καθυστερημένες ECMWF προβλέψεις ελέγχου (5 μέλη). Προτού υπολογιστούν οι δείκτες ρίσκου, μετατρέπονται όλες σε προβλέψεις συνόλου αιολικής ισχύος. Σε αυτή τη μελέτη, διαπιστώθηκε ότι οι προβλέψεις μέσου όρου χρονικής υστέρησης επιτρέπουν την επίλυση καταστάσεων με ποικίλα επίπεδα αβεβαιότητας πρόγνωσης.

Στην εργασία [26] ένα νέο regime switching μοντέλο βασιζόμενο στην τεχνητή νοημοσύνη προτείνεται για την παροχή προβλέψεων με χρονικό ορίζοντα μεγαλύτερο της μίας ημέρας με ιδιαίτερη μέριμνα για την πρόβλεψη ακραίων γεγονότων. Το μοντέλο που προτάθηκε κατάφερε να βελτιώσει την προβλεψιμότητα της παραγωγής αιολικής ενέργειας θεωρώντας τα ακραία γεγονότα ως ξεχωριστό regime σχετιζόμενο με την αβεβαιότητα των NWP. Για τον υπολογισμό του regime εφαρμόζεται ένα νευρωνικού δίκτυου που στηρίζεται στη θεωρία προσαρμόσιμου συντονισμού (adaptive reasonance theory-ART). Το επονομαζόμενο RBF-pARTMAP εκτιμά την πιθανότητα εμφάνισης των regime. Οι προβλέψεις αιολικής ισχύος παράγονται από νευρωνικά δίκτυο ακτινωτής βάσης RBFNN (Radial Basis Function Neural Netwroks), όπου το καθένα εκπαιδεύεται με δεδομένα που αντιστοιχούν σε διαφορετικά regime. Η τελική έξοδος του προτεινόμενου μοντέλου αποκτάται συνδυάζοντας τις προβλέψεις των RBFNN με τις πιθανότητες των regime όπως αυτές εκτιμώνται από το δίκτυο RBF-pARTMAP. Για τον εμπλουτισμό των RBFNN με νέα πληροφορία εφαρμόζεται ένα υβριδικό μοντέλο που βασίζεται στη συνδυασμένη χρήση αλγορίθμου MRAN (Minimal Resource Allocation Network)[24] και γενετικών αλγορίθμων GMRAN [25].

3.2 Βασικές Έννοιες Πιθανοτήτων

Μπορούμε να καταλάβουμε την έννοια της πιθανότητας ως τη σχετική συχνότητα εμφάνισης n_i κάποιας τιμής x_i μιας διακριτής τυχαίας μεταβλητής(τ.μ.) X . Αν είχαμε τη δυνατότητα να συλλέξουμε αυθαίρετα πολλές n παρατηρήσεις (n→∞), τότε το όριο της σχετικής συχνότητας είναι η πιθανότητα η τ.μ. X να πάρει την τιμή x_i

$$P(x_i) = P(X = x_i) = \lim_{n\to\infty} \frac{n_i}{n}$$

Για συνεχή τ.μ. X δεν έχει νόημα να μιλάμε για την πιθανότητα η X να πάρει μία συγκεκριμένη τιμή αλλά για την πιθανότητα η X να ανήκει σε ένα διάστημα τιμών dx, δηλαδή

$$P(a < X < b) = \int_a^b f(x)dx\,, \qquad \forall a < b$$

Η πιθανότητα η τ.μ. X να πάρει κάποια τιμή x_i , αν είναι διακριτή, ή να βρίσκεται σε ένα διάστημα τιμών dx, αν είναι συνεχής, μπορεί να μμεταβάλλεται στο σύνολο των διακεκριμένων τιμών ή σε διαφορετικά διαστήματα και δίνεται ως συνάρτηση της τ.μ. X . Για διακριτή τ.μ. X που παίρνει τις τιμές $x_1, x_2, \ldots, x_m$, η συνάρτηση αυτή λέγεται συνάρτηση μάζας πιθανότητας, ορίζεται ως

$$f_X(x_i) = P(X = x_i)$$

και ικανοποιεί τις συνθήκες:

$$f_X(x_i) \geq 0 \text{ και } \sum_{i=1}^{m} f_X(x_i) = 1.$$

Αντίστοιχα, για συνεχή τ.μ. X ($X \in R$) ορίζεται η συνάρτηση πυκνότητας πιθανότητας $f_X(x)$ που ικανοποιεί τις συνθήκες:

$$f_X(x) \geq 0 \text{ και } \int_{-\infty}^{+\infty} f_X(x)dx = 1$$

Η κατανομή πιθανότητας της τ.μ. X ορίζεται επίσης από την αθροιστική συνάρτηση κατανομής $F_X(x)$ και δηλώνει την πιθανότητα η τ.μ. X να πάρει τιμές μικρότερες ή ίσες από κάποια τιμή x.

Για διακριτή τ.μ. X είναι

$$F_X(x_i) = P(X \leq x_i) = \sum_{x\leq xi} f_X(x)$$

Και για συνεχή τ.μ. Χ είναι

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(u)du$$

Σημειώνεται ότι μία συνεχής μεταβλητή μπορεί να μετατραπεί σε διακριτή με κατάλληλη διαμέριση του πεδίου τιμών της. Αν η συνεχής τ.μ. Χ ορίζεται στο διάστημα [a, b] μια διαμέριση Σ σε m κελιά δίνεται ως

$\Sigma = \{a = r_0, r_1, \ldots, r_{m-1}, r_m = b\}$, όπου $r_0 < r_1 < \ldots < r_m$.

Αντιστοιχίζοντας διακεκριμένες τιμές x_i , i = 1, . . . , m, σε κάθε κελί (διάστημα) $[r_i-1, r_i)$, η πιθανότητα εμφάνισης μιας τιμής x_i της διακριτικοποιημένης τ.μ. Χ ′, $f_{X'}(x_i) = P(X' = x_i)$, δίνεται από την πιθανότητα η συνεχής τ.μ. Χ να παίρνει τιμές στο διάστημα $[r_i-1, r_i)$, $P(r_i-1 \leq X \leq r_i) = F_X(r_i) - F_X(r_{i-1})$.

3.3 Δεσμευμένη Πιθανότητα

Ο ορισμός της δεσμευμένης πιθανότητας θα μας χρειαστεί στη συνέχεια της παρούσας διπλωματικής καθώς θα αναζητούμε την πιθανότητα να έχουμε μία τιμή παραγόμενης ηλεκτρικής ισχύος, δεδομένης μιας συγκεκριμένης τιμής ηλιακής ακτινοβολίας.

Δειγματικός Χώρος (ή δειγματοχώρος) Ω (≠ ∅) ονομάζεται το σύνολο των δυνατών αποτελεσμάτων ενός πειράματος τύχης

π.χ. Ω = {Κ,Γ}, {1,2,3,4,5,6}, [0,1]}.

Ενδεχόμενα του Ω ονομάζονται μια «συλλογή» F από υποσύνολα του Ω που έχουν τις ακόλουθες ιδιότητες (σ-άλγεβρα):

(i) Ω ∈ F.

(ii) Αν A ∈ F τότε και A^c ∈ F.

(iii) Αν $A_1, A_2, \ldots$ ∈ F τότε και $U_{i=1}^{\infty} A_i$ ∈ F

Έστω δύο ενδεχόμενα Α και Β ενός δειγματικού χώρου Ω και P(B) > 0. Η πιθανότητα να πραγματοποιηθεί το ενδεχόμενο Α δεδομένου ότι έχει (ή ότι θα) πραγματοποιηθεί το ενδεχόμενο Β ορίζεται:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Η P(A|B) καλείται και δεσμευμένη πιθανότητα του Α δοθέντος του Β.

Παράδειγμα. Μία οικογένεια έχει δύο παιδιά. Ποια είναι η πιθανότητα να είναι και τα δύο αγόρια δεδομένου ότι τουλάχιστον ένα από αυτά είναι αγόρι;

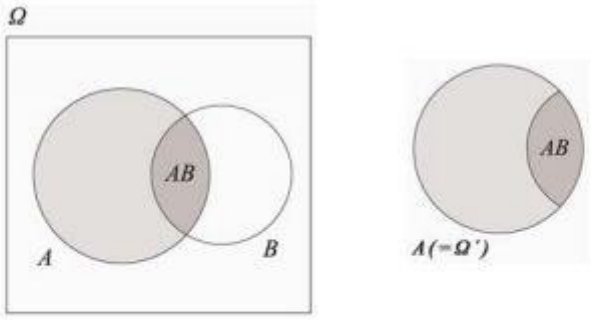
Εδώ Ω={(α,α),(α,κ),(κ,α),(κ,κ)} και θεωρούμε τα ενδεχόμενα

Α={και τα δύο παιδιά είναι αγόρια}={(α,α)}

Β={τουλάχιστον ένα από τα παιδιά είναι αγόρι}={(α,α),(α,κ),{κ,α)}

Ζητείται η:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{(a,a)\})}{P(\{(a,a),(a,k),(k,a)\})} = \frac{1/4}{3/4} = \frac{1}{3}$$



Σχήμα
Για τη δεσμευμένη πιθανότητα $P(B \mid A)$ ο δειγματικός χώρος Ω περιορίζεται στο ενδεχόμενο Α και το ενδεχόμενο Β των ευνοϊκών αποτελεσμάτων, περιορίζεται στο ενδεχόμενο ΑΒ

Σχήμα 3.1 Δειγματικός χώρος δεσμευμένης πιθανότητας

3.4 Παράμετροι

Η κατανομή πιθανότητας περιγράφει πλήρως τη συμπεριφορά της τ.μ., αλλά συνήθως στην πράξη δεν είναι γνωστή ή απαραίτητη. Όταν μελετάμε μια τ.μ. μας ενδιαφέρει κυρίως να προσδιορίζουμε κάποια βασικά χαρακτηριστικά της κατανομής της, όπως η κεντρική τάση και η μεταβλητότητα της τ.μ.. Αυτά τα χαρακτηριστικά είναι οι παράμετροι της κατανομής της τ.μ.

3.4.1 Μέση Τιμή

Τα μέτρα θέσης-κεντρικής τάσης μας δίνουν πληροφορίες για τη θέση της κατανομής των παρατηρήσεων. Τα πλέον χρησιμοποιούμενα είναι η μέση τιμή, η διάμεσος, η κορυφή και τα ποσοστημόρια.

Αν X είναι μια διακριτή τ.μ. που παίρνει m διακριτές τιμές x_1 , x_2 , … , x_m , με σμπ $f_X(x)$, η μέση τιμή της που συμβολίζεται $\mu_x \equiv E[X]$ ή απλά μ, δίνεται ως

$$\mu \equiv E[X] = \sum_{i=1}^{m} x_i f_X(x_i).$$

Αν η X είναι συνεχής τ.μ. με σππ $f_X(x)$, η μέση τιμή της δίνεται ως

$\mu \equiv E[X] = \int_{-\infty}^{\infty} x f_X(x) dx.$

Κάποιες βασικές ιδιότητες της μέσης τιμής είναι :

- Αν η τ.μ. X παίρνει μόνο μια σταθερή τιμή c είναι $E[X] = c$.
- Αν X είναι μια τ.μ. και c είναι μια σταθερά : $E[cX] = cE[X]$
- Αν X και Y είναι δύο τ.μ.: $E[X + Y] = E[X] + E[Y]$.
- Αν X και Y είναι δύο ανεξάρτητες τ.μ.: $E[XY] = E[X]\,E[Y]$.

Οι ιδιότητες (2) και (3) δηλώνουν πως η μέση τιμή έχει τη γραμμική ιδιότητα, δηλαδή ισχύει

$$E[aX + bY] = aE[X] + bE[Y]$$

3.4.2 Κορυφή ή Επικρατούσα τιμή

Η κορυφή του δείγματος συμβολίζεται με M_0. Είναι η τιμή που εμφανίζεται στο δείγμα με την μεγαλύτερη συχνότητα.

3.4.3 Διάμεσος

Η διάμεσος του δείγματος συμβολίζεται με δ . Είναι η τιμή x, για την οποία ισχύει ότι: το 50% των παρατηρήσεων είναι μικρότερες από αυτή και το υπόλοιπο 50% των παρατηρήσεων είναι μεγαλύτερες από αυτή. Εκφράζει την κεντρική θέση της κατανομής των παρατηρήσεων και γι΄ αυτό στη βιβλιογραφία συναντάται και ως μέσος θέσης (position average).

Αν το πλήθος n των παρατηρήσεων είναι αριθμός περιττός τότε $\delta = x_{\frac{n+1}{2}}$, ενώ αν είναι άρτιος τότε $\delta = \frac{x_{\frac{n}{2}} + x_{(\frac{n}{2}+1)}}{2}$ (με x_n συμβολίζουμε τη n-οστή παρατήρηση, σε αύξουσα διάταξη παρατηρήσεων).

3.4.4 Ποσοστημόρια

Το ποσοστημόριο p_a είναι η τιμή x, για την οποία ισχύει ότι: το α% των παρατηρήσεων είναι μικρότερες από αυτή και το υπόλοιπο (1-α)% των παρατηρήσεων είναι μεγαλύτερες από αυτή. Τα ποσοστημόρια διακρίνονται σε:

Εκατοστημόρια (percentiles) $p_1, p_2, \dots, p_{99}$

Δεκατημόρια (deciles) αν $p_{10}, p_{20}, \dots, p_{90}$

Τεταρτημόρια (quartiles) $p_{25} = Q_1$, $p_{50} = Q_2$, $p_{75} = Q_3$

3.4.5 Εύρος

Ορίζεται ως η διαφορά της μικρότερης από τη μεγαλύτερη παρατήρηση του δείγματος.

$$R = x_{max} - x_{min}$$

3.4.6 Τυπική απόκλιση

Η τυπική απόκλιση του πληθυσμού συμβολίζεται με σ και του δείγματος με s .

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n \cdot \bar{x}^2\right)} \quad \text{ή}$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{k}(y_i - \bar{x})^2 \cdot \nu_i} = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{k} y_i^2 \cdot \nu_i - n \cdot \bar{x}^2\right)}$$

3.4.7 Διασπορά

Το τετράγωνο της τυπικής απόκλισης των παρατηρήσεων ονομάζεται διασπορά και συμβολίζεται με σ^2 για τον πληθυσμό και με s^2 για το δείγμα. Δηλαδή η διασπορά δίνεται από τον τύπο:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \nu \cdot \bar{x}^2\right) \quad \text{ή}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k}(y_i - \bar{x})^2 \cdot \nu_i = \frac{1}{n-1}\left(\sum_{i=1}^{k} y_i^2 \cdot \nu_i - n \cdot \bar{x}^2\right)$$

3.5 Είδη πιθανοτικών προβλέψεων

Η πιθανοτική πρόβλεψη (probabilistic forecasting) παρέχει πληροφορίες για τη μελλοντική πιθανότητα ενός ή περισσοτέρων γεγονότων και έρχεται σε αντίθεση με τη ντετερμινιστική πρόβλεψη, όπου παρέχεται μία προβλεπόμενη τιμή για τον υπό μελέτη χρονικό ορίζοντα της πρόβλεψης. Οι πιθανοτικές προβλέψεις μπορούν να έχουν διαφορετικές «μορφές» (forms), ανάλογα με τη φύση των μεταβλητών για τις οποίες γίνεται η πρόβλεψη. Για τις διακριτές (discrete) μεταβλητές, π.χ. για ένα συγκεκριμένο αριθμό πιθανών γεγονότων, οι πιθανοτικές προβλέψεις ονομάζονται προβλέψεις πιθανοτήτων (probability forecasts). Στην περίπτωση της πρόβλεψης συνεχών μεταβλητών, υπάρχουν διάφορα είδη προβλέψεων. Η πρόβλεψη εκατοστημορίου (quantile forecast) είναι η τιμή που η παρατήρηση έχει προκαθορισμένη πιθανότητα να είναι μικρότερη ή ίση από αυτή. Τα διαστήματα πρόβλεψης (predicti intervals) παρέχουν το κατώτερο και το ανώτερο όριο ενός διαστήματος, στο οποίο αναμένεται να ανήκει η παρατήρηση με κάποια προκαθορισμένη πιθανότητα. Συνεπώς, οι προβλέψεις εκατοστημορίων προβλέψεις μπορούν να θεωρηθούν σαν ανοιχτά προγνωστικά διαστήματα.

3.5.1 Προβλέψεις Συνόλου (Ensemble Forecasts)

Αρκετοί πάροχοι μετεωρολογικών προβλέψεων παράγουν πλέον πολλαπλές προσομοιώσεις του μοντέλου τους σχετικά με τις αριθμητικές προβλέψεις καιρού (NWP), καταλήγοντας σε μία πρόβλεψη συνόλου. Το σύνολο παρέχει διαφορετικά, εξίσου πιθανά, σενάρια δεδομένων των αρχικών συνθηκών και του δυναμικού μοντέλου της ατμόσφαιρας. Διαφορετικά μέρη του συνόλου επιχειρούν να εκφράσουν την εξάπλωση της αβεβαιότητας στη διαδικασία μοντελοποίησης. Αυτή η αβεβαιότητα προκύπτει από εκλιπούσες ή λανθασμένες

παρατηρήσεις, παραμετρική αβεβαιότητα και σφάλμα προτύπου (model error). Υπάρχουν ποικίλες διαφορετικές διαθέσιμες τεχνικές για την κατασκευή προβλέψεων συνόλου, όπως τα «μοναδιαία διανύσματα» και τα «αναπαραχθέντα διανύσματα», που παρουσιάζονται από τους Leutbecher and Palmer (2008). Συγκεκριμένα το ECMWF παρέχει μια πρόβλεψη συνόλου από 50 μέλη, μια πρόβλεψη ελέγχου και μια ντετερμινιστική πρόβλεψη υψηλής ανάλυσης. Οι προβλέψεις συνόλου χρησιμοποιούνται ως ακατέργαστη είσοδος από το μοντέλο NWP και μπορεί να απαιτούν βαθμονόμηση για βελτίωση των στατιστικών τους ιδιοτήτων, όπως συμβαίνει στην περίπτωση της αιολικής παραγωγής.

3.5.2 Προβλέψεις εκατοστημορίων (Quantile Forecasts)

Έστω, ότι έχουμε το πρόβλημα του υπολογισμού της τιμής της παραγόμενης ηλεκτρικής ισχύος από ένα σταθμό που αξιοποιεί την ηλιακή ακτινοβολία. Με τη μέθοδο παλινδρόμησης εκατοστημορίων (Quantile Regression) υπολογίζεται ένας πεπερασμένος αριθμός από εκατοστημόρια (quantiles) της πιθανολογικής κατανομής της παραγόμενης ενέργειας, χρησιμοποιώντας ιστορικά δεδομένα. Το εκατοστημόριο θ ορίζεται ως η τιμή όπου η πιθανότητα για παραγωγή ισχύος μικρότερη από αυτή, ισούται με θ. Δεδομένου ότι η F_t είναι γνησίως αύξουσα συνάρτηση, το εκατοστημόριο (quantile) ${q_t}^{(a)}$ της τυχαίας μεταβλητής P_t (με $a \in [0,1]$) ορίζεται μονοσήμαντα ως η τιμή του *x* τέτοια ώστε:

$$P(P_t < x) = a$$

Ή ισοδύναμα,

$${q_t}^{(a)} = {F_t}^{-1}(a)$$

Με άλλα λόγια στο πρόβλημα της πρόβλεψης της ηλιακής παραγωγής, το εκατοστημόριο ${q_t}^{(a)}$ ορίζεται ως η τιμή της ηλεκτρικής ισχύος ,από την οποία η παραγωγή ισχύος είναι μικρότερη, με καθορισμένη πιθανότητα η οποία ισούται με a.

3.5.3 Διαστήματα Πρόβλεψης (Prediction Intervals)

Τα διαστήματα πρόβλεψης χρησιμοποιούνται για να εκφράσουν ένα εύρος πιθανών τιμών μέσα στο οποίο το πραγματικό γεγονός P_t , αναμένεται να βρίσκεται με μια συγκεκριμένη πιθανότητα. Αυτή η πιθανότητα μπορεί να καθοριστεί ως ένας δείκτης κάλυψης (coverage rate) 1-β τέτοιος ώστε $\beta \in [0,1]$. Ένα διάστημα πρόβλεψης , που παρέχεται τη χρονική στιγμή t και με χρόνο οδήγησης t+k, καθορίζεται από τη διαφορά του άνω από του κάτω ορίου,

τα οποία όρια στην πραγματικότητα είναι εκατοστημόρια :

$$\hat{I}_{t+k|t} = [\hat{q}^{(a_l)}{}_{t+k|t}, \hat{q}^{(a_u)}{}_{t+k|t}]$$

, όπου για τα a_l και a_u ισχύει η σχέση: $a_u - a_l = 1 - \beta$. Έτσι από τον παραπάνω ορισμό προκύπτει ότι ένα διάστημα πρόβλεψης δεν ορίζεται μονοσήμαντα από το ονοματικό ποσοστό κάλυψής του. Για τις περισσότερες εφαρμογές πρόβλεψης, ένα πολύ σημαντικό ζήτημα που προκύπτει για τα διαστήματα πρόβλεψης είναι η επιλογή του βέλτιστου ποσοστού κάλυψης. Για να παρέχεται ένας μονοσήμαντος ορισμός πρέπει να αποφασιστεί με ποιο τρόπο θα κεντραριστεί το διάστημα στην πυκνότητα πρόβλεψης. Μια τυπική προσέγγιση είναι ο ορισμός κεντρικών διαστημάτων πρόβλεψης κεντράροντας τα διαστήματα στη διάμεσο, γεγονός που εξασφαλίζει την ύπαρξη ίσης πιθανότητας ότι η επαλήθευση θα επεκτείνεται πάνω ή κάτω του διαστήματος πρόβλεψης. Αυτό θέτει έναν επιπλέον περιορισμό στις παραμέτρους της μορφής:

$$a_l = 1 - a_u = \frac{1 - \beta}{2}$$

3.5.4 Προβλεπόμενες κατανομές (Density Forecasts)

Η προβλεπόμενη κατανομή αναφέρεται σε μια συνάρτηση συνεχούς πυκνότητας πιθανότητας που αφορά την τυχαία μεταβλητή και παρέχει μια εκτενή περιγραφή του μέλλοντος για δεδομένο χρονικό διάστημα. Η διακύμανση της προβλεπόμενης κατανομής μπορεί να χρησιμοποιηθεί για την έκφραση της αντίστοιχης με την πρόβλεψη αβεβαιότητας. Με το συμβολισμό $\hat{f}_{t+k|t}$ αναπαρίσταται η προβλεπόμενη κατανομή για την τυχαία μεταβλητή σε χρόνο t και χρόνο οδήγησης t+k. Όμοια, με $\hat{F}_{t+k|t}$ δηλώνεται η αντίστοιχη συνάρτηση αθροιστικής κατανομής (Cumulative Density Function-CDF) μιας πιθανοτικής πρόβλεψης.

3.6 Εκτιμήτριες Συναρτήσεις

Η συνάρτηση πυκνότητας πιθανότητας είναι μία θεμελιώδης έννοια στην στατιστική. Θεωρούμε μία τυχαία μεταβλητή X η οποία έχει συνάρτηση πυκνότητας πιθανότητας(σ.π.π) f. Η σ.π.π f μας δίνει την πλήρη περιγραφή της κατανομής της τ.μ. X και βοηθάει στην εύρεση των πιθανοτήτων με την χρήση της σχέσης:

$$P(a < X < b) = \int_a^b f(x)dx\,, \qquad \forall a < b$$

Υποθέτουμε ότι έχουμε ένα σύνολο παρατηρήσεων που προέρχονται από μια άγνωστη σ.π.π. f. Με τον όρο εκτιμήτρια $\hat{f}$ της σ.π.π. f εννοούμε την κατασκευή μίας εκτίμησης της άγνωστης σ.π.π. από τις δοθέντες παρατηρήσεις, ή με άλλα λόγια μια συνάρτηση του τυχαίου δείγματος που χρησιμοποιείται για την εκτίμηση μιας άγνωστης παραμέτρου μιας συνάρτησης κατανομής.

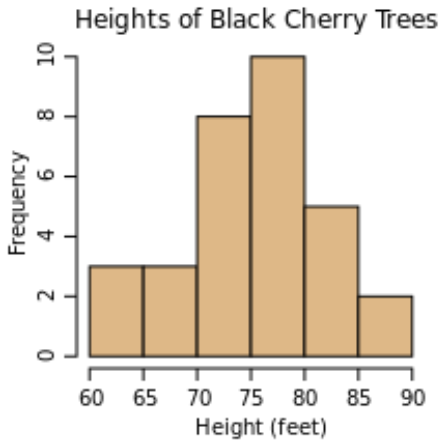
Μια από τις προσεγγίσεις της εκτίμησης της σ.π.π είναι η παραμετρική. Με την υπόθεση ότι οι παρατηρήσεις μας ανήκουν σε μια γνωστή οικογένεια κατανομών, για παράδειγμα των κανονικών με άγνωστη μέση τιμή μ και συνδιακυμανση σ^2, η σ.π.π. f θα μπορούσε να εκτιμηθεί βρίσκοντας εκτιμήσεις για τις παραμέτρους μ και σ^2 και αντικαθιστώντας αυτές στην σχέση της εκτίμησης για τις κανονικές κατανομές.

Μια ακόμα προσέγγιση, με την οποία θα ασχοληθούμε , είναι της μη-παραμετρικής στατιστικής, όπου δεν κάνουμε υποθέσεις για την κατανομή των παρατηρήσεων, αλλά τα στοιχεία από μόνα τους αποφασίζουν ποια κατανομή τους ταιριάζει καλύτερα. Δύο άλλες προσεγγίσεις, που δεν θα μας απασχολήσουν είναι της ευσταθούς και της ημιπαραμετρικής στατιστικής.

Οι βασικοί λόγοι της χρησιμοποίησης μη-παραμετρικών μεθόδων είναι ότι αξιοποιούνται για την εξερεύνηση είτε για παρουσίαση των δεδομένων. Μας δείχνουν διάφορα χαρακτηριστικά των δεδομένων, όπως λοξότητα και πολυπλοκότητα, που βοηθάν στην μετέπειτα επιλογή ενός κατάλληλου παραμετρικού μοντέλου. Επιπλέον, βοηθάν στον έλεγχο και στην εξαγωγή στατιστικών συμπερασμάτων κάτω από ελάχιστες συνθήκες.

3.6.1 Το ιστόγραμμα

Το ιστόγραμμα είναι γραφική απεικόνιση στατιστικών συχνοτήτων περιοχών τιμών ενός μεγέθους. Σχηματίζεται από παρακείμενα ορθογώνια. Η επιφάνεια κάθε ορθογωνίου είναι μέτρο της συχνότητας εμφάνισης της συγκεκριμένης περιοχής τιμών ενώ το ύψος του ισούται με το λόγο της συχνότητας προς το εύρος των τιμών που αντιπροσωπεύει το ορθογώνιο. Πρόκειται για τη συνηθέστερη επιλογή γραφικής παράστασης συνεχών μεταβλητών. Στα συνεχή δεδομένα, οι τιμές της μεταβλητής ομαδοποιούνται και οι ομάδες διατάσσονται στον οριζόντιο άξονα κατ' αύξουσα σειρά. Στη συνέχεια από κάθε ομάδα υψώνουμε ορθογώνια ,το ύψος των οποίων αντιστοιχεί στη συχνότητα κάθε ομάδας.



Σχήμα 3.2 Παράδειγμα ιστογράμματος

Πολλοί θα αναρωτιούνται γιατί να μην μπορούμε να χρησιμοποιούμε το ιστόγραμμα σε όλες τις στατιστικές διαδικασίες και πολλές φορές ψάχνουμε μεθόδους πιο προχωρημένες. Αυτό οφείλεται στο γεγονός ότι το ιστόγραμμα έχει ένα σημαντικό μειονέκτημα, το οποίο μεταφράζεται ως αναποτελεσματική χρήση των δεδομένων, κυρίως όταν χρησιμοποιείται ως εκτιμήτρια σε διαδικασίες όπως η αθροιστική ανάλυση ή η μη-παραμετρική διακριτή ανάλυση. Το μειονέκτημα του είναι η ασυνέχεια που παρουσιάζει, η οποία προκαλεί δυσκολίες όταν απαιτούνται παράγωγα της εκτίμησης, και για αυτό το λόγο εάν θέλουμε η εκτιμήτρια να χρησιμοποιηθεί ως ενδιάμεση διαδικασία σε άλλες μεθόδους, προκύπτει η ανάγκη να χρησιμοποιήσουμε μια άλλη εκτιμήτρια αντί του ιστογράμματος. Γενικά, το ιστόγραμμα είναι ικανοποιητικό μόνο για εξερεύνηση ή παρουσίαση των δεδομένων και κυρίως για μονομεταβλητές περιπτώσεις, γιατί σε διμεταβλητά ή τριμεταβλητά δεδομένα, είναι δύσκολο να το σχεδιάσουμε τρισδιάστατα αφού η εκτιμήτρια δεν εξαρτάται μόνο από την αρχική τιμή x_0 αλλά και από τις διευθύνσεις των κελιών.

3.6.2 Ο απλοϊκός εκτιμητής

Γνωρίζουμε ότι εάν έχουμε μια τυχαία μεταβλητή X, η σ.π.π. ορίζεται από την σχέση:

$$f(x) = lim_{h\to 0}\frac{1}{2h}P(x-h<X<x+h).$$

Για δοσμένο h, μπορούμε να εκτιμήσουμε την πιθανότητα P(x − h < X < x + h) από την αναλογία του δείγματος που ανήκει στο διάστημα (x− h, x + h). Έτσι προκύπτει και ο ορισμός του απλοϊκού εκτιμητή:

$$\hat{f}(x) = \frac{1}{2nh}[o\ αριθ.\ των\ παρατ.\ X_i\ που\ ανήκουν\ στο\ διάστ\ (x-h, x+h)].$$

Για να εκφράσουμε καλύτερα τον απλοϊκό εκτιμητή πρέπει να ορίσουμε μια συνάρτηση

βάρους w ως εξής:

$$w(x) = \begin{cases} \frac{1}{2}, & \varepsilon\alpha\nu\ |x| < 1 \\ 0, & \delta\iota\alpha\varphi o\rho\varepsilon\tau\iota\kappa\alpha \end{cases},$$

όπου w(x) είναι η σ.π.π. μιας συνεχούς ομοιόμορφης κατανομής, οπότε, ο απλοϊκός εκτιμητής θα πάρει την μορφή :

$$\hat{f}(x) = \frac{1}{n}\sum_1^n \frac{1}{h} w(\frac{x-X_i}{h}).$$

Από τις παραπάνω σχέσεις φαίνεται ότι αντικαθιστώντας κάθε παρατήρηση με ένα «κουτί» πλάτους $2h$ και ύψους $(2nh)^{-1}$ και αθροίζοντας αυτά τα κουτιά προκύπτει η εκτίμηση μας.

3.6.3 Η εκτιμήτρια με την μέθοδο του πυρήνα(Kernel density estimation)

Η εκτιμήτρια με την μέθοδο του πυρήνα είναι μια γενίκευση του απλοϊκού εκτιμητή που σκοπό έχει να προσεγγίσει καλύτερα το πρόβλημα. Έτσι αντικαθιστώντας την συνάρτηση βάρους w(x) με μια συνάρτηση πυρήνα K, που θα ικανοποιεί την συνθήκη $\int_{-\infty}^{\infty} K(x)dx = 1$,η εκτιμήτρια με την μέθοδο του πυρήνα έχει την μορφή:

$$\hat{f}(x) = \frac{1}{nh}\sum_1^n K(\frac{x-X_i}{h}).$$

Στην περίπτωση που έχουμε δεσμευμένη πιθανότητα $f(y|x)$ τότε η εκτιμήτρια $\hat{f}(y|x)$ παίρνει τη μορφή:

$$\tilde{f}(y \mid x) = \frac{\tilde{f}(y,x)}{\tilde{f}(x)} = \frac{\sum_{i=1}^{n} K_{h_2}(x - X_i)\, K_{h_1}(y - Y_i)}{\sum_{i=1}^{n} K_{h_2}(x - X_i)}.$$

$$\tilde{f}(y,x) = \frac{1}{n}\sum_{i=1}^{n} K_{h_2}(x - X_i)\, K_{h_1}(y - Y_i)$$

$$\tilde{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{h_2}(x - X_i)$$

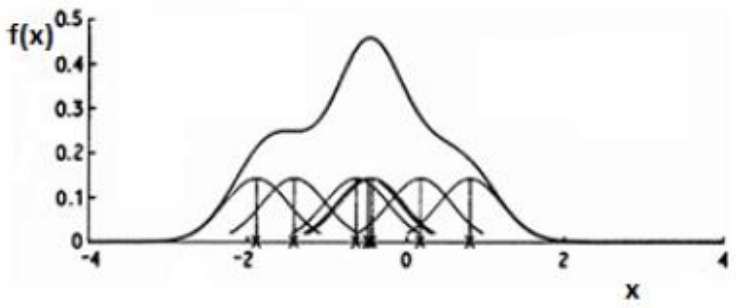
Όπου $K_h(.) = K(\frac{\cdot}{h})/h$ είναι η συνάρτηση πυρήνα με πλάτος h.

Συνηθώς ως συνάρτηση πυρήνα Κ χρησιμοποιείται η κανονική κατανομή Gauss:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$$

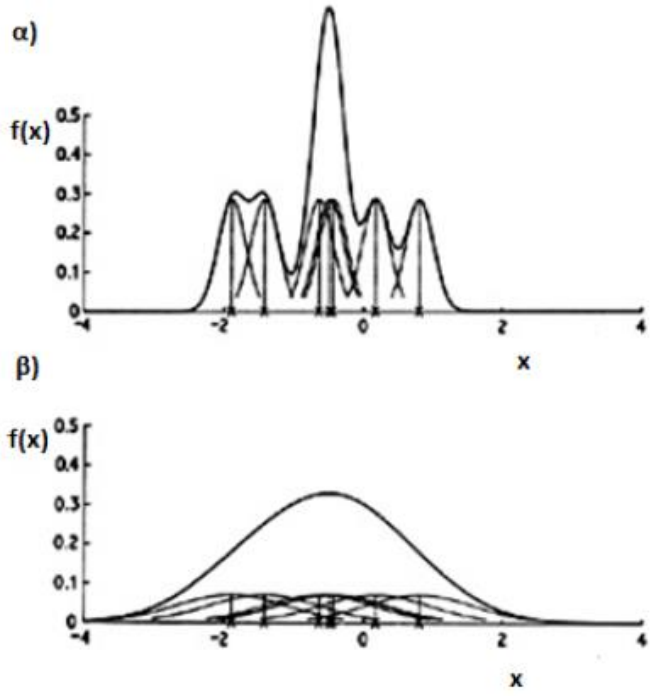
Όπως ο απλοϊκός εκτιμητής μπορεί να θεωρηθεί ως πρόσθεση των ΄΄κουτιών΄΄ έτσι και η εκτιμήτρια με την μέθοδο του πυρήνα μπορεί να θεωρηθεί ως πρόσθεση των καμπύλων (bumps). Η συνάρτηση πυρήνα Κ καθορίζει το σχήμα της καμπύλης ενώ το h(bandwidth) καθορίζει το πλάτος της.

Στην συνέχεια θα παραθέσουμε κάποια γραφήματα από τα οποία θα δούμε πως η εκτιμήτρια με την μέθοδο του πυρήνα κατασκευάζεται και πως επηρεάζεται από την αλλαγή του h.



Εκτιμήτρια πυρήνα που προκύπτει από τις μεμονωμένες καμπύλες. Πλάτος κελιού $h = 0.4$.

Σχήμα 3.3 Εκτιμήτρια πυρήνα με h=0.4



Εκτιμήτριες πυρήνα που προκύπτουν από τις μεμονωμένες καμπύλες. Πλάτος κελιού α) $h = 0.2$ β) $h = 0.8$.

Σχήμα 3.4 Εκτιμήτριες πυρήνα με διαφορετικές τιμές h

Στο πρώτο γράφημα φαίνεται ότι η εκτιμήτρια πυρήνα κατασκευάζεται από το άθροισμα των μεμονωμένων καμπύλων. Στο δεύτερο γράφημα, που ακολουθεί, φαίνεται η επίδραση της αλλαγής του h. Το όριο καθώς το h τείνει στο 0 είναι το άθροισμα των αιχμών της δέλτα Dirac συνάρτησης (Σχήμα 3.4.α), ενώ καθώς το h μεγαλώνει πολλές λεπτομέρειες της κατανομής δεν γίνονται φανερές (Σχήμα 3.4β). Γι' αυτό και είναι βαρύνουσας σημασίας η εύρεση της βέλτιστης τιμής για το πλάτος κελιού h, ώστε να επιτύχουμε το βέλτιστο αποτέλεσμα.

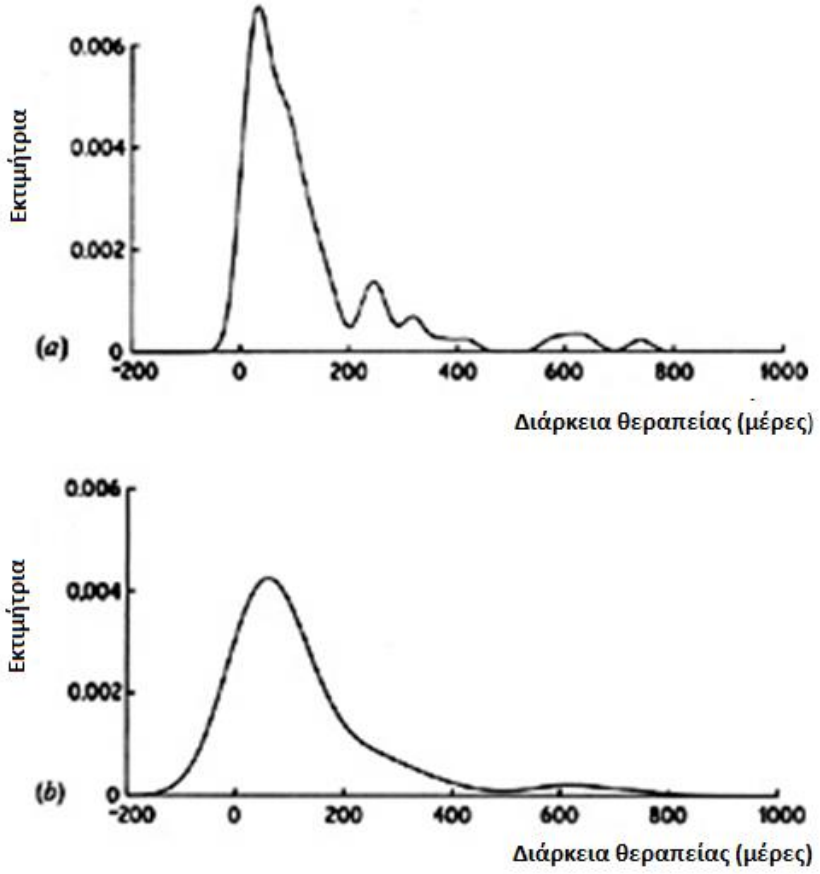
<u>Βασικές ιδιότητες της εκτιμήτριας με την μέθοδο του πυρήνα</u>

Δεδομένου ότι η συνάρτηση πυρήνα K είναι παντού μη-αρνητική και ικανοποιεί τη συνθήκη $\int_{-\infty}^{\infty} K(x)dx = 1$,δηλαδή είναι μια σ.π.π., συνεπάγεται ότι και η εκτιμήτρια $\hat{f}$ είναι και αυτή μια σ.π.π.. Επιπλέον η $\hat{f}$ κληρονομεί όλες τις ιδιότητες της συνάρτησης πυρήνα K όπως π.χ. την συνέχεια και την διαφορισιμότητα. Έτσι εάν η K είναι μια κανονική σ.π.π. τότε συνεπάγεται ότι και η $\hat{f}$ θα είναι μια ομαλή καμπύλη.

Τέλος, η εκτιμήτρια με την μέθοδο του πυρήνα είναι η εκτιμήτρια που χρησιμοποιείται πιο πολύ από όλες και για αυτό το λόγο είναι και η πιο μελετημένη μαθηματικώς. Ωστόσο, αυτή η

μέθοδος πάσχει από ένα σοβαρό μειονέκτημα όταν εφαρμόζεται σε δεδομένα που ανήκουν σε κατανομές με μακριές ουρές. Επειδή το h είναι σταθερό σε όλο το μήκος των παρατηρήσεων, υπάρχει μια τάση ανωμαλίας στην ουρά της εκτίμησης. Ωστόσο, εάν η εκτίμηση είναι εξομαλυμένη ώστε να μπορεί να αντιμετωπίσει αυτό το πρόβλημα, τότε σημαντικές ιδιότητες από το κεντρικό μέρος της κατανομής δεν γίνονται φανερές

Η εκτίμηση που φαίνεται στο Γράφημα (α) με $h = 20$ παρουσιάζει θόρυβο στην δεξιά ουρά ενώ η εκτίμηση στο Γράφημα (b) με $h = 60$ παρουσιάζει μια πιο ομαλή καμπύλη στην ουρά. Ωστόσο, στην δεύτερη περίπτωση το μήκος της καμπύλης στο κεντρικό μέρος της κατανομής μεγαλώνει. Γι' αυτό και στη προσέγγιση μας στην παρούσα διπλωματική θα δώσουμε ιδιαίτερη έμφαση στην εύρεση των βέλτιστων τιμών για το συντελεστή h. Θα ψάξουμε δηλαδή αυτόν που μας παρέχει την καλύτερη δυνατή καμπύλη, ή με άλλα λόγια αυτόν που έχει το μικρότερο σφάλμα.



Σχήμα 3.5 Διαφόρα καμπύλων στην ουρά για διαφορετικές τιμές του h.

Κεφάλαιο 4: Ακρίβεια Πρόβλεψης

4.1 Εισαγωγή

Η σπουδαιότητα της επιλογής του κατάλληλου μοντέλου για την ελαχιστοποίηση του σφάλματος έγκειται στο γεγονός ότι κατά την πρόβλεψη της ηλεκτρικής ισχύος, η απόκλιση από τις πραγματικές τιμές συνεπάγεται οικονομικές απώλειες για του τελικούς χρήστες των προβλέψεων. Συνεπώς, η αποτίμηση των προβλέψεων είναι ύψιστης σημασίας μέρος της διαδικασίας της πρόβλεψης, όχι μόνο για τη διαμόρφωση πλήρους άποψης σχετικά με τη λειτουργία της επιλεγμένης προσέγγισης, αλλά επίσης και για την απόκτηση βαθύτερης διορατικότητας σε ό,τι αφορά την αβεβαιότητα της πρόβλεψης.

Στο κεφάλαιο αυτό παρατίθεται ο ορισμός του σφάλματος της πρόβλεψης και στη συνέχεια παρουσιάζεται η μέθοδος αξιολόγησης των σημειακών και πιθανοτικών προβλέψεων. Κατά την παράθεση αυτή γίνεται αναφορά σε ένα σύνολο εκ των σημαντικότερων δεικτών σφάλματος που απασχολούν τη βιβλιογραφία των προβλέψεων και διακρίνοντάς τους σε συγκεκριμένες κατηγορίες, γίνεται μια προσπάθεια εντοπισμού των βασικών προτερημάτων και αδυναμιών του καθενός.

4.2 Ορισμός σφάλματος πρόβλεψης

Ένας αξιοσημείωτος αριθμός από διαφορετικούς στατιστικούς δείκτες μέτρησης του σφάλματος έχουν κατά καιρούς προταθεί. Το απλό σφάλμα της πρόβλεψης ενός μεγέθους ορίζεται ως η διαφορά της προβλεπόμενης τιμής που προέρχεται από το μοντέλο πρόβλεψης και της παρατηρούμενης τιμής, όπως αυτή προκύπτει από τα όργανα μέτρησης. Έτσι, το σφάλμα της πρόβλεψης της ηλεκτρικής ισχύος ενός μοντέλου που αντιστοιχεί στις επόμενες k ώρες μπορεί να γραφεί:

$$\varepsilon_{t+k|t} = \hat{y}_{t+k|t} - y_{t+k|t}$$

,όπου $\hat{y}_{t+k|t}$ είναι η πρόβλεψη της ηλεκτρικής ενέργειας μετά από k ώρες και $y_{t+k|t}$ είναι η πραγματική παραγωγή ισχύος k ώρες μετά. Στη συνέχεια αναλύουμε τα σημεία αναφοράς για τις προβλέψεις μας και κάποιους βασικούς στατιστικούς δείκτες σφάλματος.

4.3 Σημεία αναφοράς(benchmarks) σημειακών προβλέψεων

Η πιο απλή και συνηθισμένη μορφή πρόβλεψης είναι η παροχή της καλύτερης εικασίας για τη μελλοντική τιμή της πρόβλεψης μιας τυχαίας μεταβλητής. Έστω ότιη τυχαία μεταβλητή είναι η ηλιακή ενέργεια. Έστω ότι οι χρονοσειρές που σχηματίζονται από την πρόβλεψη της ηλιακής παραγωγής σε διακριτές χρονικές στιγμές t εκφράζονται με τη μεταβλητή y_t . Τότε κάθε σημειακή πρόβλεψη με ορίζοντα k βημάτων στο μέλλον μπορεί να γραφεί ως

$$\hat{y}_{t+k|t} = f(\Omega_t, k),$$

όπου Ω_t είναι το σύνολο πληροφορίας τη χρονική στιγμή t και αποτελείται από όλες τις παρατηρήσεις που είναι διαθέσιμες μέχρι εκείνη τη χρονική στιγμή. Η ιδέα αυτή της παροχής της καλύτερης εικασίας υπονοεί την ύπαρξη μιας συνάρτησης ωφελείας (utility function) του επαγγελματία που θα δράσει στη μεταφερόμενη με την πρόβλεψη πληροφορία.

Για την ανάδειξη του καλύτερου μοντέλου πρόβλεψης για κάποια συγκεκριμένη χρονοσειρά παραγωγής ηλιακής ενέργειας, πραγματοποιείται σύγκριση μεταξύ τους, εξετάζοντας κάποιο συγκεκριμένο μέτρο ακρίβειας της πρόβλεψης. Για τη διευκόλυνση της διαδικασίας σύγκρισης των μοντέλων χρησιμοποιούνται ορισμένα απλά σημεία αναφοράς (benchmarks) που στόχο έχουν να προσδιορισθεί το επίπεδο ακρίβειας που θα πρέπει να αναμένεται. Για το λόγο αυτό χρησιμοποιούνται τα παρακάτω σημεία αναφοράς:

- **Παραμένουσα τιμή (persistence):** Αποτελεί την πιο απλή στατιστική μέθοδο. Η πρόβλεψη που προκύπτει από τη μέθοδο αυτή (στη βιβλιογραφία αναφέρεται και ως Naive) για μια χρονική στιγμή $t + k$ είναι ίση με την πραγματική παρατήρηση της τελευταίας διαθέσιμης χρονικής περιόδου t :

  $$\hat{y}^{per}_{t+k|t} = y_t$$

  Για προβλέψεις μικρού χρονικού ορίζοντα η παραμένουσα τιμή παρέχει ένα καλό σημείο αναφοράς.

- **Απλός κινητός μέσος όρος (simple moving average):** Χρησιμοποιώντας τον απλό κινητό μέσο όρο των παρατηρήσεων που έγιναν κατά τα τελευταία m χρονικά βήματα επιτυγχάνεται συνήθως βελτίωση του σημείου αναφοράς persistence:

  $$\hat{y}^{sma}_{t+k|t} = \frac{1}{m}\sum_{i=1}^{m} y_{t-i+1}$$

- **Απόλυτος μέσος όρος (unconditional mean):** Αποτελεί ειδική περίπτωση του απλού κινητού μέσου όρου και προκύπτει όταν το m ισούται με το σύνολο των διαθέσιμων παρατηρήσεων, δηλαδή το μήκος της χρονοσειράς. Αυτό το σημείο αναφοράς αντιστοιχεί στο μακροπρόθεσμο μέσο όρο και συμβολίζεται ως $\bar{y}$. Για μεγάλους ορίζοντες πρόβλεψης παρέχει ένα καλό σημείο αναφοράς.

- **Σταθμισμένο σημείο αναφοράς (weighted benchmark):** Το σημείο αυτό αναφοράς δίνει καλά αποτελέσματα για ενδιάμεσους ορίζοντες πρόβλεψης. Μπορεί να κατασκευαστεί θεωρώντας έναν σταθμισμένο μέσο όρο όπου τα βάρη είναι συνάρτηση του ορίζοντα πρόβλεψης:

$$\hat{y}_{t+k|t}^{wb} = a_k y_t + (1 - a_k)\bar{y}$$

, όπου οι παράμετροι a_k πρέπει να εκτιμώνται χρησιμοποιώντας το διαθέσιμο σύνολο εκπαίδευσης δεδομένων και με βάση το χρονικό ορίζοντα της πρόβλεψης. Παρατηρείται άλλωστε ότι για $a_k = 1$ έχουμε τη μέθοδο της παραμένουσας τιμής, ενώ για $a_k = 0$ προκύπτει το σημείο αναφοράς του απόλυτου μέσου όρου.

4.4 Δείκτες σφάλματος σημειακών προβλέψεων

Στη συνέχεια παρατίθενται ορισμένοι στατιστικοί δείκτες σφάλματος που χρησιμοποιούνται σε μεγάλο βαθμό για τη σύγκριση μεταξύ των σημειακών μοντέλων πρόβλεψης:

- **Μέσο σφάλμα (Mean Error - Bias):** Αναφέρεται στο συστηματικό λάθος που παρατηρείται κατά την πρόβλεψη. Η ποσότητα αυτή εκτιμάται ως το μέσο λάθος κατά την περίοδο αποτίμησης και υπολογίζεται ξεχωριστά για κάθε ορίζοντα πρόβλεψης. Ο τύπος του συγκεκριμένου δείκτη είναι:

$$Bias(k) = \bar{\varepsilon}_\kappa = \frac{1}{N}\sum_{t=1}^{N} \varepsilon_{t+k|t}$$

- **Μέσο τετραγωνικό σφάλμα (root mean square error):** Αυτό το μέτρο ακρίβειας της πρόβλεψης δίνει πολύ μεγάλο βάρος στα μεγάλα σφάλματα και μικρότερο βάρος στα μικρά, δεδομένου ότι το σφάλμα τετραγωνίζεται. Υπολογίζεται από τον τύπο:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N} \varepsilon^2{}_{t+k|t}}$$

Ακόμη, το RMSE σχετίζεται με την υπόθεση των ανεξάρτητων και κατανεμημένων λαθών πρόβλεψης. Αν το μοντέλο παράγει κανονικά κατανεμημένα λάθη πρόβλεψης τότε η εκτιμήτρια μέγιστης πιθανοφάνειας των παραμέτρων συμπίπτει με τις τιμές των παραμέτρων που προέκυψαν που προέκυψαν από την ελαχιστοποίηση του RMSE χρησιμοποιώντας το σύνολο εκπαίδευσης δεδομένων (training data set), όπως θα εξηγηθεί και στη συνέχεια.

- **Μέσο απόλυτο σφάλμα (Mean Absolute Error) :** Εκφράζει ένα μέτρο της ακρίβειας της πρόβλεψης έναντι των πραγματικών τιμών διατηρώντας τις μονάδες μέτρησης της αρχικής χρονοσειράς. Δηλώνει ένα μέτρο της αστοχίας της πρόβλεψης, χωρίς να δίνεται έμφαση στην κατεύθυνση της πρόβλεψης. Όσο μεγαλύτερη είναι η τιμή του δείκτη, τόσο μικρότερη προκύπτει η ακρίβεια της μεθόδου που εφαρμόσθηκε. Υπολογίζεται από τον τύπο:

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|\varepsilon_{t+k|t}|$$

Εφόσον είναι επιθυμητό να ελαχιστοποιηθεί το MAE, η σημειακή πρόβλεψη θα πρέπει να αντιστοιχεί στη διάμεσο της προβλεπόμενης κατανομής.

4.5 Δείκτες σφάλματος πιθανοτικών προβλέψεων

Με την αύξηση των πιθανοτικών μοντέλων πρόβλεψης της καθίσταται απαραίτητη η ανάγκη κατάταξής τους για την επιλογή του μοντέλου εκείνου που επιφέρει τα καλύτερα αποτελέσματα. Ωστόσο, ενώ η επιλογή μεταξύ σημειακών μοντέλων πρόβλεψης είναι μια σχετικά ξεκάθαρη διαδικασία, τα πράγματα περιπλέκονται στην περίπτωση των πιθανοτικών προβλέψεων. Στη συνέχεια περιγράφονται τα σημαντικά διαγνωστικά εργαλεία για την αποτίμηση των χαρακτηριστικών ενός συστήματος πιθανοτικής πρόβλεψης.

Διαγράμματα αξιοπιστίας (reliability diagram)

Το διάγραμμα αξιοπιστίας παρέχει ένα μέσο οπτικοποίησής του (πιθανοτικού) bias του συστήματος πιθανοτικής πρόβλεψης. Μπορεί να κατασκευάζεται με διαφορετικούς τρόπους ανάλογα με το αν θεωρηθούν γεγονότα πολλών κατηγοριών (multi-categorical events) ή συνεχείς μεταβλητές. Στην πρώτη περίπτωση, το διάγραμμα κατασκευάζεται σχεδιάζοντας την παρατηρηθείσα συχνότητα του γεγονότος σε συνάρτηση με την προβλεπόμενη πιθανότητα, όπου το εύρος των προβλεπόμενων πιθανοτήτων διαιρείται σε διαστήματα (π.χ. 0-5%, 5-10% κτλ.). Η διαγώνια γραμμή δείχνει την τέλεια αξιοπιστία (η μέση παρατηρηθείσα συχνότητα ισούται με την προβλεφθείσα πιθανότητα για κάθε κατηγορία) και η οριζόντια γραμμή αναπαριστά την κλιματολογική συχνότητα. Στη δεύτερη περίπτωση, όταν δηλαδή πρόκειται για συνεχείς μεταβλητές, τα διαγράμματα αξιοπιστίας είναι όμοια με τις γραφικές εκατοστημορίων-εκατοστημορίων στο ότι δίνουν το παρατηρηθέν τμήμα των διαφόρων εκατοστημορίων που αποτελούν τις προβλεπόμενες πυκνότητες σε συνάρτηση με τις ονομαστικές.

Ιστόγραμμα κατάταξης (Rank Histogram)

Μια διαφορετική προσέγγιση στην ανάλυση της βαθμονόμησης μιας πιθανοτικής πρόβλεψης ενός συστήματος συνόλου (ensemble system) είναι η κατασκευή ενός ιστογράμματος κατάταξης. Τα ιστογράμματα κατάταξης συνήθως παράγονται για συστήματα συνόλου με περιορισμένο αριθμό μελών. Αν η πιθανοτική πρόβλεψη ενός τέτοιου συνόλου είναι καλά βαθμονομημένη, η παρατήρηση είναι εξίσου πιθανό να βρίσκεται μεταξύ δύο οποιονδήποτε διατεταγμένων διπλανών μελών, συμπεριλαμβανομένων των περιπτώσεων όπου η παρατήρηση θα βρίσκεται έξω από το εύρος του συνόλου σε οποιαδήποτε πλευρά της κατανομής. Τότε το ιστόγραμμα κατάταξης θα πρέπει να είναι επίπεδο με τον ίδιο αριθμό επιβεβαιώσεων σε κάθε διάστημα. Λόγω του περιορισμένου μεγέθους του συνόλου, η παρατήρηση μπορεί να βρίσκεται εκτός του εύρους του συνόλου. Για παράδειγμα, στο ECMWF που είναι ένα σύνολο με 51 μέλη, αυτό θα συμβεί για 2/51 ή περίπου 4% του χρόνου.

Για συνεχείς προβλεπόμενες κατανομές, απαιτείται διαφορετική μέθοδος για την αξιολόγηση της βαθμονόμησης. Αυτό επιτυγχάνεται με το μετασχηματισμό του πιθανοτικού ολοκληρώματος. Ο μετασχηματισμός που αντιστοιχεί στην CDF πρόβλεψη, $\hat{F}_{t+k|t}$ και στην πραγματική ηλεκτρική παραγωγή $y_{t+k|t}$ δίνεται από τον τύπο $z_{t,k} = \hat{F}_{t+k|t}$.

Λογαριθμικό αποτέλεσμα (Logarithmic score)

Η λογαριθμική πιθανότητα για μια πιθανοτική πρόβλεψη που παράγεται τη χρονική στιγμή t και με χρονικό ορίζοντα t+k δίνεται από τη σχέση:

$$L_{t+k|t} = \ln\left(\hat{f}_{t+k|t}\left(y_{t+k|t}\right)\right)$$

Όπου, $\hat{f}_{t+k|t}\left(y_{t+k|t}\right)$ είναι η εκτίμηση της πιθανότητας την οποία παρέχει η προβλεπόμενη κατανομή $\hat{f}_{t+k|t}(y)$, υπολογισμένη στη συγκεκριμένη τιμή της παρατήρησης y_{t+k}. Αυτή μπορεί να υπολογιστεί εμπειρικά εκτιμώντας την παράγωγο της CDF πρόβλεψης $\hat{F}_{t+k|t}(y)$ της ηλεκτρικής ισχύος. Ο μέσος όρος των λογαριθμικών πιθανοτήτων κάθε ζευγαριού πρόβλεψης/επιβεβαίωσης παρέχει ένα αποτέλεσμα για κάθε ορίζοντα πρόβλεψης:

$$LS(k) = \frac{1}{N}\sum_{t=1}^{N} \hat{L}_{t+k|t}$$

Continuous ranked probability score (CRPS)

Το CRPS χρησιμοποιείται ευρύτατα ως μέσο αποτίμησης πιθανοτικών προβλέψεων. Το CRPS για μια CDF πρόβλεψη $\hat{F}_{t+k|t}(y)$ και η αντίστοιχη επιβεβαίωση y_{t+k} ορίζονται ως εξής:

$$crps(\hat{F}_{t+k|t}(y), y_{t+k}) = \int_{-\infty}^{+\infty} (F_{t+k|t}(y) - I(y \geq y_{t+k}))^2 dy$$

Όπου $I(.)$ είναι μια ενδεικτική συνάρτηση, που ισούται με 1 αν το γεγονός εντός της παρενθέσεως είναι αληθές και με 0 αλλιώς. Ο μέσος όρος αυτών των CRPS τιμών πάνω σε κάθε ζευγάρι πρόβλεψης/επιβεβαίωσης παρέχει ένα αποτέλεσμα για κάθε ορίζοντα πρόβλεψης:

$$CRPS(k) = \frac{1}{N}\sum_{t=1}^{N} crps(\hat{F}_{t+k|t}(y), y_{t+k})$$

Το CRPS είναι το βασικότερο μέτρο σύγκρισης για τις πιθανοτικές προβλέψεις, με την έννοια ότι μεταξύ ανταγωνιστικών μεθόδων πρόβλεψης επιλέγεται αυτή που ελαχιστοποιεί το συγκεκριμένο δείκτη. Αξίζει επίσης να σημειωθεί ότι για σημειακές προβλέψεις , το CRPS εκφυλίζεται στο μέσο απόλυτο σφάλμα (MAE).

Ranked Probability Score (RPS)

Είναι ο δείκτης σφάλματος που χρησιμοποιήσαμε. Το RPS (Epstein, 1969) χρησιμοποιείται ευρύτατα ως μέσο αποτίμησης πιθανοτικών προβλέψεων, και πιο συγκεκριμένα αποτελεί εξειδίκευση του CPRS όταν έχουμε διακριτές τιμές. Ουσιαστικά υπολογίζει τη διαφορά των συναρτήσεων κατανομών της προβλεπόμενης τυχαίας μεταβλητής και των παρατηρήσεων αντίστοιχα. Ο συγκεκριμένος δείκτης δίνει βάρος στην σχετική απόσταση μεταξύ της προβλεπόμενης τιμής και του πραγματικού αποτελέσματος. Η σχέση που υπολογίζεται είναι η εξής:

$$RPS = \frac{1}{r-1}\sum_{i=1}^{r}\left(\sum_{j=1}^{i} p_j - \sum_{j=1}^{i} e_j\right)^2$$

Όπου r: το πλήθος των πιθανών αποτελεσμάτων

p_j: η πιθανότητα να έρθει αποτέλεσμα j

e_j: 0 ή 1, αναλόγως αν ήρθε η τιμή j στα πραγματικά δεδομένα

Συνάρτηση απώλειας εκατοστημορίων (quantile loss function)

Η συνάρτηση απώλειας εκατοστημορίων γνωστή και ως «συνάρτηση ελέγχου», χρησιμοποιείται τυπικά για τον ορισμό ενός συγκεκριμένου εκατοστημορίου μιας κατανομής. Για ένα συγκεκριμένο τμήμα $\alpha \in [0,1]$, η συνάρτηση απώλειας εκατοστημορίων είναι μια τμηματικά γραμμική συνάρτηση που δίνεται από τη σχέση:

$$\rho_\alpha(u) = u\big(a - I(u < 0)\big)$$

, όπου u είναι η διαφορά ανάμεσα στην παρατηρηθείσα και την υπολογισμένη τιμή. Το πρόβλημα της εκτίμησης του εκατοστημορίου με τμήμα α μπορεί να γραφεί ως:

$$\hat{\mathrm{q}}^{(\alpha)} = \min_q \sum_{t=1}^{N} \rho_\alpha (y_t - q)$$

Εκτός της χρησιμοποίησης της συνάρτησης απώλειας εκατοστημορίων για εκτίμηση του εκατοστημορίου με αυτόν τον τρόπο, μπορεί επιπλέον να χρησιμοποιηθεί για την αποτίμηση των προβλέψεων εκατοστημορίων. Μια σειρά προβλέψεων εκατοστημορίων $\hat{q}_{t+k|t}^{(\alpha)}(y)$, εκδοθείσες σε χρόνους t, με ορίζοντα k και τμήμα α, μπορεί να αποτιμηθεί χρησιμοποιώντας την:

$$QL(k,\alpha) = \sum_{t=1}^{N} \rho_\alpha (y_{t+k} - \hat{q}_{\mathbf{t+k|t}}^{(\alpha)}(y))$$

Στην απλή περίπτωση όπου α=0,5 , το αποτέλεσμα αυτό εκφυλίζεται στο μισό του MAE.

Αποτίμηση πρόβλεψης εκατοστημορίων

Για μια συγκεκριμένη πρόβλεψη εκατοστημορίων $\hat{q}_{t+k|t}^{(\alpha)}$, εκδοθείσα σε χρόνο t, με χρονικό ορίζοντα t+k και επαλήθευση y_{t+k} , ορίζουμε τη μεταβλητή ένδειξης:

$$\xi_{\mathbf{t,k}}^{(\alpha)} = I(y_{t+k} < \hat{q}_{\mathbf{t+k|t}}^{(\alpha)})$$

Η χρονοσειρά που αποτελείται από τα $\xi_{t,k}^{(\alpha)}$, αποτελεί μια δυαδική ακολουθία που αντιστοιχεί στα "hits" αν η επαλήθευση είναι κάτω από την πρόβλεψη εκατοστημορίων, αλλιώς καταγράφεται ως "miss". Για κάθε ορίζοντα k, μπορούμε να υπολογίσουμε την πραγματική κάλυψη της πρόβλεψης εκατοστημορίων θεωρώντας ένα μέσο όρο του συνόλου αποτίμησης:

$$\hat{a}_{\mathbf{k}}^{(\alpha)} = \frac{1}{N} \sum_{t=1}^{N} \xi_{\mathbf{t,k}}^{(\alpha)}$$

Για την ποσοτικοποίηση της *αξιοπιστίας* μπορούμε να μετρήσουμε την πόλωση (bias) του συστήματος πρόβλεψης:

$$b_{\mathbf{k}}^{(\alpha)} = \alpha - \hat{a}_{\mathbf{k}}^{(\alpha)}$$

Για την αξιολόγηση της απόδοσης της πρόβλεψης μπορεί να είναι χρήσιμη η παροχή τιμών bias για κάθε ονομαστικό τμήμα εκατοστημορίου, ως ένας μέσος όρος πάνω από ολόκληρο το μήκος των αντίστοιχων οριζόντων πρόβλεψης:

$$\bar{b}^{(a)} = \frac{1}{k_{max}} \sum_{k=1}^{k_{max}} b_{\mathrm{k}}^{(\alpha)}$$

Αποτίμηση διαστημάτων πρόβλεψης

Εφόσον ένα διάστημα πρόβλεψης περιλαμβάνει 2 εκατοστημόρια, ό,τι αφορά ένα μόνο quantile σχετίζεται με την αξιολόγηση της απόδοσης ενός διαστήματος πρόβλεψης. Πράγματι, δεν αρκεί απλώς να ελεγχθεί η κάλυψη που παρέχεται από το διάστημα πρόβλεψης, αλλά είναι σημαντικό να αξιολογηθεί αν και τα δύο quantiles που απαιτούνται για τον ορισμό της πρόβλεψης διαστημάτων είναι αμερόληπτα (unbiased). Μια προσέγγιση για τον έλεγχο της *αιχμηρότητας* της πρόβλεψης διαστημάτων είναι η εστίαση στο πλάτος τους. Για διαστήματα πρόβλεψης κεντραρισμένα στη διάμεσο με τιμή κάλυψης (1-β), το πλάτος δίνεται από τη σχέση:

$$\delta_{t,\mathrm{k}}^{(\beta)} = \hat{q}_{\mathrm{t+k|t}}^{(1-\frac{\beta}{2})} - \hat{q}_{\mathrm{t+k|t}}^{(\frac{\beta}{2})}$$

και το μέτρο της αιχμηρότητας των διαστημάτων αυτών δίνεται από τη σχέση:

$$\bar{\delta}_{\mathrm{k}}^{(\beta)} = \frac{1}{N} \sum_{t=1}^{N} \delta_{t,\mathrm{k}}^{(\beta)}$$

Μπορεί ακόμη να είναι χρήσιμη η πληροφορία σχετικά με την αιχμηρότητα των διαστημάτων πρόβλεψης σε ένα εύρος οριζόντων θεωρώντας το μέσο όρο:

$$\bar{\delta} = \frac{1}{k_{max}} \sum_{k=1}^{k_{max}} \bar{\delta}_{\mathrm{k}}^{(\alpha)}$$

Κεφάλαιο 5: Προτεινόμενη μεθοδολογία και δεδομένα εξεταζόμενου προβλήματος

5.1 Εξετάζομενο πρόβλημα και μέθοδοι πρόβλεψης

Στόχος της παρούσας διπλωματικής είναι η παραγωγή προβλέψεων για την παραγόμενη ηλεκτρική ενέργεια ενός σταθμού παραγωγής που αξιοποιεί τις Α.Π.Ε. (Ανανεώσιμες Πηγές Ενέργειας). Πιο συγκεκριμένα, προσεγγίζουμε το πρόβλημα πιθανοτικά, δηλαδή εξάγουμε τις συναρτήσεις πυκνότητας πιθανότητας για την αναμενόμενη παραγωγή ενέργειας συναρτήσει τις ηλιακής ακτινοβολίας, με την τεχνική των εκτιμητριών συναρτήσεων με πυρήνα (Kernel Density Estimation), που παρουσιάστηκε στο 2ο κεφάλαιο. Η συγκεκριμένη τεχνική προσφέρει πληθώρα πλεονεκτημάτων έναντι της κλασσικής σημειακής πρόβλεψης, τα όποια αναλύθηκαν παραπάνω. Αφού λοιπόν εξάγουμε τις πιθανοτικές μας προβλέψεις, μελετάμε την ακρίβεια των προβλέψεων μας με συγκεκριμένους στατιστικούς δείκτες και τη συγκρίνουμε με άλλες ήδη γνωστές μεθόδους. Τέλος, εξάγουμε κάποια βασικά συμπεράσματα καθώς και προτείνουμε ορισμένες ιδέες για το μέλλον.

5.2 Δομή του πειράματος

5.2.1 Χρονικό Διάστημα Συλλογής Πληροφοριών

Το χρονικό διάστημα συλλογής των πληροφοριών επιλέχθηκε να είναι από 20 Απριλίου 2016 και ώρα 0:00 έως 13 Δεκεμβρίου 2016 και ώρα 23:00. Η επιλογή αυτή στηρίχθηκε στο γεγονός ότι το διάστημα οφείλει να είναι αρκετό για να υπάρχουν συνολικά 4968 τιμές, άρα οι αριθμητικές μέθοδοι να έχουν την αξιοπιστία που απαιτείται. Πιο συγκεκριμένα, τα δεδομένα που αξιοποιούμε ωριαία είναι η τιμή της ηλιακής ακτινοβολίας(Irradiation) και της παραγωγής ηλεκτρικής ενέργειας(PV Production).

Στη συνέχεια παρουσιάζονται τα δεδομένα μας, όπως αύτα εμφανίζονται στη διεπαφή του περιβάλλοντος RStudio. Ενδεικτικά παρουσιάζουμε τις πρώτες 23 ληφθείσες μετρήσεις για λόγους κατανόησης της οργάνωσης των δεδομένων πριν την επεξεργασία τους.

	datetime	Irradiation	Temperature	PV production	Storage	Grid	CHP1	CHP2	Electricity demand
1	20-04-16 0:00	0.00	13.40	0.00	-1.63	88.64	0.09	0.15	90.04
2	20-04-16 1:00	0.00	12.30	0.00	-1.62	83.73	0.09	0.15	85.11
3	20-04-16 2:00	0.00	11.55	0.00	-1.61	79.74	0.08	0.15	81.11
4	20-04-16 3:00	0.00	11.00	0.00	-1.62	81.25	0.09	0.15	82.63
5	20-04-16 4:00	0.00	10.70	0.00	-1.61	81.75	0.08	0.15	83.13
6	20-04-16 5:00	0.00	10.40	0.00	-1.62	77.20	0.08	0.15	78.58
7	20-04-16 6:00	0.00	10.15	0.00	-1.63	80.68	0.08	0.15	82.07
8	20-04-16 7:00	0.00	9.95	7.09	-1.62	104.33	0.08	0.15	112.80
9	20-04-16 8:00	0.00	9.75	24.45	-1.63	151.23	0.08	0.15	177.07
10	20-04-16 9:00	13.30	9.75	39.26	-1.63	209.38	0.08	0.15	250.04
11	20-04-16 10:00	108.90	10.70	49.44	-1.63	202.15	0.08	0.15	253.00
12	20-04-16 11:00	294.50	14.67	50.16	-1.62	223.93	0.08	0.14	275.48
13	20-04-16 12:00	459.08	17.70	51.54	-1.63	183.31	0.09	0.15	236.25
14	20-04-16 13:00	601.65	19.95	46.85	-1.61	202.66	0.08	0.14	250.88
15	20-04-16 14:00	717.28	21.70	46.25	-1.61	193.11	0.08	0.14	240.73
16	20-04-16 15:00	788.33	22.95	45.19	-1.62	193.18	0.08	0.14	239.76
17	20-04-16 16:00	798.45	23.65	37.03	-1.61	181.82	0.08	0.14	220.23
18	20-04-16 17:00	747.13	23.90	19.31	-1.63	118.29	0.08	0.14	139.00
19	20-04-16 18:00	643.35	23.88	10.94	-1.63	95.28	0.08	0.14	107.62
20	20-04-16 19:00	499.40	23.73	2.77	-1.62	101.30	0.08	0.14	105.46
21	20-04-16 20:00	336.80	23.13	0.00	-1.60	98.80	0.08	0.14	100.17
22	20-04-16 21:00	172.78	21.80	0.00	-1.63	100.08	0.08	0.14	101.49
23	20-04-16 22:00	37.90	19.68	0.00	-1.65	90.89	0.08	0.14	92.31

Πίνακας 5.1 Δεδομένα επεξεργασίας

5.2.2 Ανάλυση δεδομένων

Τα βασικά μεγέθη που μας απασχολούν από τα δεδομένα μας όπως είπαμε είναι η ηλιακή ακτινοβολία(Irradiation) και η παραγωγή ηλεκτρικής ενέργειας(PV Production). Η μέση και η μέγιστη τιμή των παραπάνω μεγεθών είναι:

- ✓ Μέση τιμή(PV Production) = 13.46264 W, Μέγιστη τιμή(PV Production) = 72.22 W
- ✓ Μέση τιμή(Irradiation) = 217.4458 W/m^2 , Μέγιστη τιμή(Irradiation) = 933.63 W/m^2

Αναλυτικότερα θα διαχωρίσουμε τα δεδομένα ανά ώρα και θα δούμε τις αντίστοιχες μέσες τιμές στον παρακάτω πίνακα:

Time	Mean (PV Production)
0	0.023
1	0.023
2	0.023
3	0.023
4	0.023
5	0.023
6	0.549
7	4.876
8	13.901
9	24.908
10	35.418
11	41.652
12	44.169
13	43.067
14	38.134
15	31.049
16	22.952
17	14.114
18	6.225
19	1.680
20	0.191
21	0.023
22	0.023
23	0.023

Πίνακας 5.2 Ωριαίες μέσες τιμές Παραγόμενης Ηλεκτρικής Ενέργειας

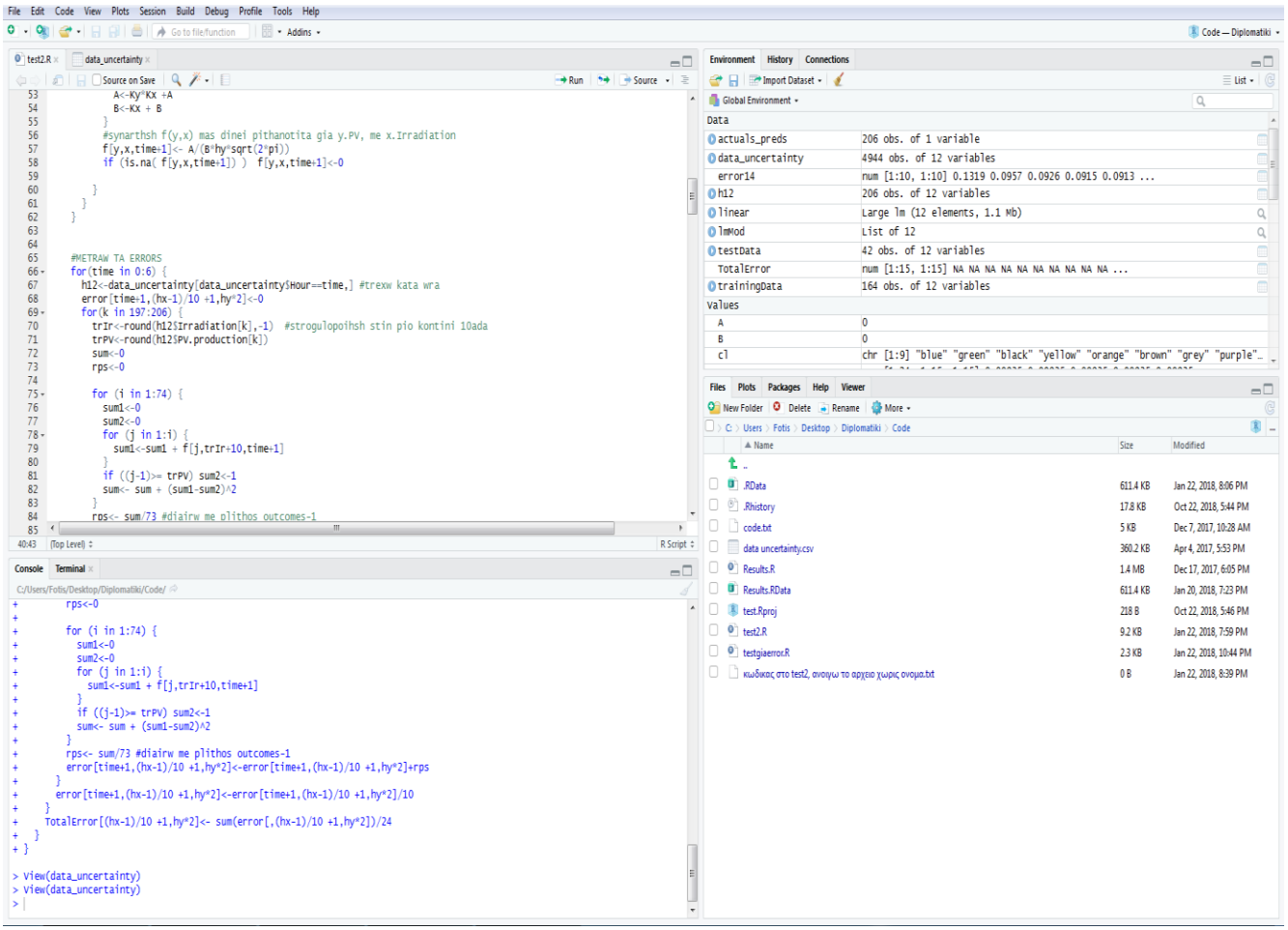
Time	Mean(Irradiation)
0	0
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	11.42
9	65.71
10	182.28
11	320.36
12	449.61
13	561.59
14	624.35
15	644.79
16	636.12
17	563.71
18	457.04
19	344.69
20	217.49
21	105.98
22	32.13
23	1.36

Πίνακας 5.3 Ωριαίες μέσες τιμές Παραγόμενης Ηλεκτρικής Ενέργειας

Από την παραπάνω διερεύνηση των δεδομένων μας προκύπτει ότι για τις ώρες: 0:00-7:00 και 18:00-23:00 παρατηρούνται ιδιαιτέρως μικρές τιμές παραγωγής ενέργειας αλλά και ακτινοβολίας, πράγμα λογικό μιας και τις συγκεκριμένες ώρες δεν έχουμε ηλιοφάνεια. Η ανάλυση αυτή θα διαδραματίσει καθοριστικό ρόλο στην περαιτέρω ανάπτυξη της μεθοδολογίας μας.

5.2.2 Λίγα λόγια για το RStudio

Η R είναι μια ισχυρή γλώσσα και ένα ισχυρό περιβάλλον ανάπτυξης για στατιστικούς υπολογισμούς και γραφικά. Τα κύρια πλεονεκτήματα της R είναι το γεγονός ότι η R είναι ελεύθερο λογισμικό και ότι υπάρχει πολύ βοήθεια διαθέσιμη στο διαδίκτυο. Είναι αρκετά παρόμοια με άλλα προγραμματιστικά πακέτα όπως η MATLAB (που δεν είναι ελεύθερο λογισμικό), αλλά πιο φιλική προς τον χρήστη από γλώσσες προγραμματισμού όπως η C++ και η Fortran. Ο όγκος των δεδομένων μας καθώς και η πολυπλοκότητα των υπολογιστικών πράξεων και των δοκιμών που έγιναν στην παρούσα διπλωματική, καθιστούν αναγκαία την αξιοποίηση ενός εργαλείου όπως το RStudio. Το RStudio είναι γραμμένο σε R και αποτελεί ένα ολοκληρωμένα γραφικό περιβάλλον στατιστικής επεξεργασίας. Η έκδοση που χρησιμοποιήσαμε στην παρούσα διπλωματική είναι η 1.2.1355.



Σχήμα 5.1 Το περιβάλλον του RStudio.

5.2.3 Προτεινόμενη μεθοδολογία

Αρχικά, τα δεδομένα (σε μορφή .csv) εισήχθησαν στο περιβάλλον της R, με αρχική μορφή όπως αυτή που παρουσιάστηκε παραπάνω. Οι μεταβλητές ενδιαφέροντος στα πλαίσια της εν λόγω διπλωματικής εργασίας είναι, πρώτον, η παραγόμενη ηλεκτρική ενέργεια και, δεύτερον, η αντίστοιχη τιμή της ηλιακής ακτινοβολίας για κάθε χρονική στιγμή. Και για τις δύο μεταβλητές, η σχετική πληροφορία παρέχεται στα αρχεία csv, την οποία και θα αξιοποιήσουμε για περαιτέρω ανάλυση.

Η ανάλυση αυτή αφορά την υλοποίηση μιας εκτιμήτριας πυρήνα (Conditional Kernel Density Estimation) με σκοπό τη δημιουργία συναρτήσεων πυκνότητας πιθανότητας $\hat{f}\,(y|x)$ της Παραγόμενης Ηλεκτρικής Ενέργειας (ΠΗΕ) , έστω y, συναρτήσει της Ηλιακής Ακτινοβολίας (ΗΑ), έστω x. Αναλυτικότερα, αναζητούμε τη συχνότητα εμφάνισης μιας δεδομένης τιμής της ΠΗΕ, για μια δεδομένη τιμή της ΗΑ. Ο υπολογισμός της ανωτέρω συνάρτησης πυκνότητας πιθανότητας έγινε με χρήση του μαθηματικού τύπου που παρουσιάστηκε στο τρίτο κεφάλαιο της παρούσας διπλωματικής. Οι υπολογισμοί που έλαβαν μέρος στη διαδικασία εύρεσης της πυκνότητας έγιναν με έναν αλγόριθμο που υλοποιήθηκε στη γλώσσα R. Υπενθυμίζουμε τον τύπο της εκτιμήτριας συνάρτησης πυκνότητας πιθανότητας για δύο τυχαίες μεταβλητές x, y:

$$\hat{f}\,(y|x) = \frac{\sum_{i=1}^{n} K_{h_x}(X_t - x) K_{h_y}(Y_t - y)}{\sum_{i=1}^{n} K_{h_x}(X_t - x)}$$

Ερμηνεύοντας με λόγια της προηγούμενη εξίσωση, βλέπουμε ότι για τις διάφορες (γνωστές) τιμές της x, υπολογίζεται κάθε φορά η πιθανότητα εμφάνισης της y. Τελικά, δημιουργείται ένα δυσδιάστατος πίνακας που περιλαμβάνει τις πιθανότητες της y δεδομένης της x. Αξίζει να σημειωθεί ότι οι στήλες του πίνακα αθροίζουν (η καθεμιά) στη μονάδα.

Επιπλέον, από τον παραπάνω μαθηματικό τύπο είναι προφανές ότι είναι γνώστες όλες οι μεταβλητές, εκτός των παραμέτρων h_x, h_y (bandwidth). Η συγκεκριμένη τιμή καθορίζει το πλάτος του κελιού κατά τον υπολογισμό της εκτιμήτριας συνάρτησης και κατ' επέκταση το πλάτος της τελικής καμπύλης. Όσο μικρότερη η τιμή του, τόσο πιο αιχμηρό προκύπτει το γράφημα, ενώ για μεγάλες τιμές του προκύπτει υπερβολικά εξομαλυμένη καμπύλη. Ο τελικός στόχος μας ήταν η βελτιστοποίηση των παραμέτρων h_x, h_y, ώστε να έχουμε την καλύτερη δυνατή προσέγγιση.

Αναλυτικότερα η μεθοδολογία που επιλέχθηκε είναι η εξής: Αρχικά απαιτείται η ομαδοποίηση των δεδομένων κατά ώρα. Αυτή η ανάγκη προέκυψε καθώς διαφορετικές ώρες της ημέρας έχουμε διαφορετικές τιμές ηλιακής ακτινοβολίας άρα και ηλεκτρικής παραγωγής. Αυτό είναι προφανές αφού τις βραδινές ώρες για παράδειγμα δεν έχουμε ηλιοφάνεια. Όταν χρησιμοποιηθούν όλα τα δεδομένα χωρίς ομαδοποίηση τα αποτελέσματα δεν έχουν καμία σημαντικότητα καθώς παρουσιάζουν μεγάλα σφάλματα, γι' αυτό και προέκυψε η ανάγκη της ωριαίας ομαδοποίησης. Συνεπώς για κάθε ώρα(0- 23) έχουμε και την αντίστοιχη εκτιμήτρια συνάρτηση $\hat{f}\,(y|x)$.

Εν συνεχεία, εξάγουμε από τα δεδομένα μόνο τις τιμές εκείνες που αφορούν το πρόβλημα, δηλαδή η παραγόμενη ηλεκτρική ενέργεια και η αντίστοιχη τιμή της ηλιακής ακτινοβολίας. Το επόμενο στάδιο είναι η βελτιστοποίηση των ευρών ζώνης. Εφόσον δε γνωρίζουμε εκ των προτέρων τις τιμές για τα εύρη ζώνης(bandwidth) h_x και h_y, παράγουμε 225(15X15) διαφορετικούς συνδυασμούς Kernel, για την κάθε ώρα, και αναζητούμε το τοπικό ελάχιστο σφάλματος, στο πλέγμα αυτό τιμών.

- ✓ Το h_x πήρε τιμές από 1 έως 141 με βήμα 10.
- ✓ Το h_y πήρε τιμές από 0.5 έως 7.5 με βήμα 0.5.

Σε περιπτώσεις όπου δεν βρέθηκε το τοπικό ακρότατο μέσα σε αυτό το εύρος τιμών έγινε επιπλέον διερεύνηση στις τιμές που κρίθηκε αναγκαίο ώστε να εντοπιστεί το τοπικό ελάχιστο. Ο δείκτης σφάλματος που χρησιμοποιήσαμε για να συγκρίνουμε τα αποτελέσματα και να βρούμε τον βέλτιστο συνδυασμό για κάθε ώρα ήταν ο RPS που αναλύθηκε στο τέταρτο κεφάλαιο.

$$RPS = \frac{1}{r-1}\sum_{i=1}^{r}\left(\sum_{j=1}^{i} p_j - \sum_{j=1}^{i} e_j\right)^2$$

Εδώ αξίζει να τονίσουμε ότι εφόσον οι τιμές της Παραγόμενης Ενέργειας και της Ακτινοβολίας είναι συνεχείς, για να επιτευχθεί η προσομοίωση του προβλήματος αλγοριθμικά απαιτείται μια διακριτοποίηση των παραπάνω τιμών. Συνεπώς αφού βρεθούν οι μέγιστες τιμές του PV(72.22) και του Irradiation(933.63), τίθεται το εξής βήμα για την κάθε μια.

- ✓ Το PV πήρε τιμές από 0 έως 73 με βήμα 1.
- ✓ Το Irradiation πήρε τιμές από 0 έως 940 με βήμα 10.

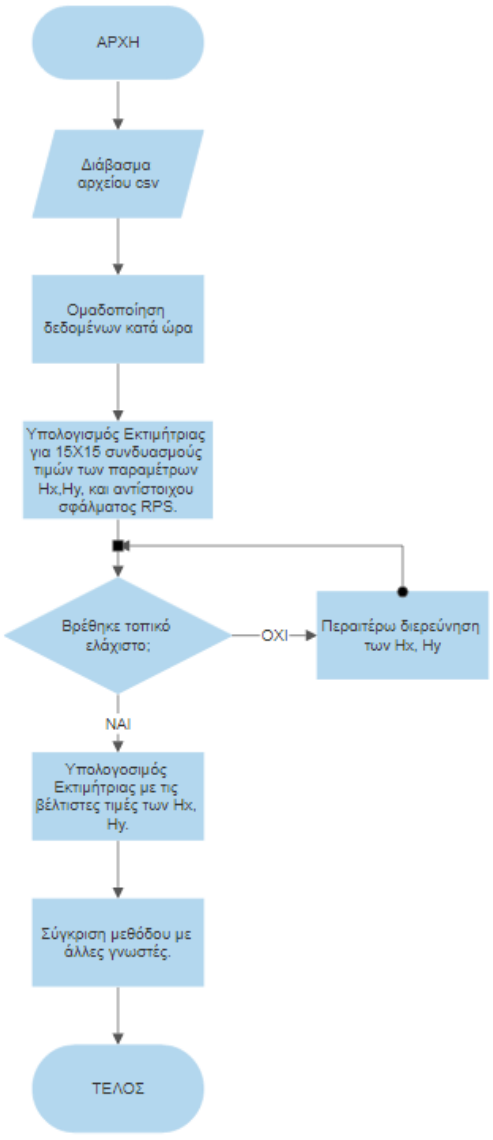
Τέλος, ο έλεγχος των σφαλμάτων έγινε με cross-validation. Δηλαδή για κάθε ώρα έχουμε 296 τιμές και κρατάμε τα πρώτα 196 για τον υπολογισμό της εκτιμήτριας και στα επόμενα 10 ελέγχουμε τα σφάλματα. Αφού βρούμε τις βέλτιστες τιμές των bandwidth για την κάθε ώρα ξανατρέχουμε τον αλγόριθμο της εκτιμήτριας σε όλα τα δεδομένα και έτσι προκύπτουν οι τελικές εκτιμήτριες.

Στη συνέχεια της παρούσας διπλωματικής εφαρμόσαμε κι άλλες ήδη γνωστές τεχνικές προβλέψεων στα δεδομένα μας, ώστε να συγκρίνουμε την ακρίβεια της προτεινόμενης μεθόδου με αυτές. Οι τεχνικές που αναλύθηκαν στα προηγούμενα κεφάλαια και εφαρμόσαμε είναι οι εξής:

- Εκτιμήτρια με πυρήνα (Kernel Density Estimation)- με διαχωρισμό ανά ώρα
- Εκτιμήτρια με πυρήνα (Kernel Density Estimation)- χωρίς διαχωρισμό ανά ώρα
- Απλή γραμμική παλινδρόμηση- χωρίς διαχωρισμό ανά ώρα
- Απλή γραμμική παλινδρόμηση- με διαχωρισμό ανά ώρα

Αξίζει σε αυτό το σημείο να τονίσουμε ότι μπορούμε να χωρίσουμε τις ώρες που μελετάμε σε δύο κατηγορίες αναλόγως με το πόσο σημαντικές ή μη είναι. Η πρώτη κατηγορία είναι οι ώρες που επικρατεί ελάχιστη ή και καθόλου ηλιοφάνεια και κατά συνέπεια η ηλιακή ακτινοβολία έχει κυρίως μηδενικές τιμές. Αυτές οι ώρες είναι από τις 18:00 το απόγευμα έως τις 7:00 το επόμενο πρωί για την κάθε ημέρα. Αντίστοιχα από τις 8:00 το πρωί μέχρι 17:00 το απόγευμα είναι οι πλέον σημαντικές ώρες.

Για να γίνει πιο κατανοητή η μεθοδολογία που ακολουθήσαμε παρουσιάζουμε σε αυτό το σημείο και το διάγραμμα ροής της αλγοριθμικής ιδέας:



Σχήμα 5.2 Διάγραμμα ροής αλγορίθμου μεθοδολογίας

Κεφάλαιο 6: Παρουσίαση αποτελεσμάτων και σύγκριση μεθόδων

6.1 Υπολογισμός εκτιμήτριας συνάρτησης

Αρχικά όπως τονίστηκε στο προηγούμενο κεφάλαιο η πρώτη ενέργεια είναι η ομαδοποίηση των δεδομένων ανά ώρα του εικοσιτετράωρου. Ο κώδικας στην R που υλοποιεί το συγκεκριμένο σκοπό είναι ο εξής:

```
for (i in 1:nrow(data_uncertainty)){
if (nchar(data_uncertainty$datetime[i])==13){
  data_uncertainty$Hour[i]<-as.numeric(substr(data_uncertainty$datetime[i],10,10))
}
  else{
  data_uncertainty$Hour[i]<-as.numeric(substr(data_uncertainty$datetime[i],10,11))
  }
}
```

Σχήμα 6.1 Κώδικας ομαδοποίησης δεδομένων κατά ώρα στην R.

Στη συνέχεια, ακολουθεί η εύρεση των βέλτιστων τιμών για τα h_x και h_y. Η επιλογή τους θα γίνει με βάση το δείκτη ακρίβειας RPS. Πιο συγκεκριμένα, δοκιμάζονται διάφοροι συνδυασμοί των τιμών h_x και h_y. Κρατούνται N-h ιστορικές παρατηρήσεις και παράγονται προβλέψεις για h τιμές μπροστά, όπου h : ορίζοντας πρόβλεψης. Ειδικότερα, για κάθε ώρα από τα δεδομένα υπάρχουν ως N=206 παρατηρήσεις και h=10 τιμές που θα γίνει η πρόβλεψη. Συνεπώς, παράγοντας 10 προβλέψεις και υπολογίζοντας το σφάλμα με τις τελευταίες 10 παρατηρήσεις που θεωρήθηκαν κρυφές, προκύπτουν τα RPS σφάλματα.

```
for (hx in seq(1, 141, by = 10)){
  for (hy in seq(0.5, 7.5 , by = 0.5)){

    A<-0
    B<-0
    f<-array(NA, dim=c(74,950,24))

    for(time in 0:23) {
      h12<-data_uncertainty[data_uncertainty$Hour==time,]
      for(y in 1:74) {
        for(x in seq(10, 950, by = 10))
          A<-0
          B<-0
          for(i in 1:(nrow(h12)-10)) {
            Xt<- h12$Irradiation[i]
            Yt<- h12$PV.production[i]
            Ky<-exp(-0.5* ((Yt-(y-1))/hy)^2)
            Kx<-exp(-0.5* ((Xt-(x-10))/hx)^2)
            A<-Ky*Kx +A
            B<-Kx + B
          }
          f[y,x,time+1]<- A/(B*hy*sqrt(2*pi))
          if (is.na( f[y,x,time+1]) )  f[y,x,time+1]<-0

        }
      }
    }
```

Σχήμα 6.2 Κώδικας υπολογισμού εκτιμήτριας συνάρτησης

```
    for(time in 0:23) {
      h12<-data_uncertainty[data_uncertainty$Hour==time,]
      error[time+1,(hx-1)/10 +1,hy*2]<-0
      for(k in 197:206) {
        trIr<-round(h12$Irradiation[k],-1)
        trPV<-round(h12$PV.production[k])
        sum<-0
        rps<-0

        for (i in 1:74) {
          sum1<-0
          sum2<-0
          for (j in 1:i) {
            sum1<-sum1 + f[j,trIr+10,time+1]
          }
          if ((j-1)>= trPV) sum2<-1
          sum<- sum + (sum1-sum2)^2
        }
        rps<- sum/73
        error[time+1,(hx-1)/10 +1,hy*2]<-error[time+1,(hx-1)/10 +1,hy*2]+rps
      }
      error[time+1,(hx-1)/10 +1,hy*2]<-error[time+1,(hx-1)/10 +1,hy*2]/10
    }
    TotalError[(hx-1)/10 +1,hy*2]<- sum(error[,(hx-1)/10 +1,hy*2])/24
  }
```

Σχήμα 6.3 Κώδικας υπολογισμού δείκτη σφάλματος RPS

6.2 Εύρεση βέλτιστων τιμών εύρους ζώνης

Παρακάτω παρουσιάζουμε αναλυτικά τα σφάλματα που προέκυψαν για διάφορους συνδυασμούς τιμών h_x και h_y. Οι πίνακες έχουν τη μορφή heatmap, δηλαδή οι χαμηλότερες τιμές σφαλμάτων εμφανίζονται με πράσινο χρώμα ενώ όσο μεγαλύτερες τείνουν προς το κόκκινο, για ευκολότερη ανάγνωση. Στον οριζόντιο άξονα μεταβάλλεται η τιμή του h_y από 0.5 αριστερά μέχρι 7.5 προς τα δεξιά με βήμα 0.5, ενώ στον κατακόρυφο η τιμή του h_x από 1 πάνω έως 141 προς τα κάτω με βήμα 10. Αξίζει να σημειωθεί ότι σε περιπτώσεις ωρών που δε βρέθηκε τοπικό ελάχιστο στον πίνακα αναζήτησης, πραγματοποιήθηκε περαιτέρω διερεύνηση και τα αποτελέσματα αυτά παρουσιάζονται σε παρακάτω πίνακα. Τέλος, όπως έχει ήδη τονιστεί ορισμένες ώρες ομαδοποιήθηκαν αφού είχαν ακριβώς ίδιες τιμές σφαλμάτων, λόγω μηδενικής ηλιοφάνειας.

0:00-5:00

H_x \ H_y	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7,5
1	0,0094	0,0951	0,1434	0,1726	0,1927	0,2078	0,2198	0,2299	0,2386	0,2463	0,2533	0,2598	0,2658	0,2715	0,2769
11	0,0094	0,0951	0,1434	0,1726	0,1927	0,2078	0,2198	0,2299	0,2386	0,2463	0,2533	0,2598	0,2658	0,2715	0,2769
21	0,0094	0,0951	0,1434	0,1726	0,1927	0,2078	0,2198	0,2299	0,2386	0,2463	0,2533	0,2598	0,2658	0,2715	0,2769
31	0,0094	0,0951	0,1434	0,1726	0,1927	0,2078	0,2198	0,2299	0,2386	0,2463	0,2533	0,2598	0,2658	0,2715	0,2769
41	0,0094	0,0951	0,1434	0,1726	0,1927	0,2078	0,2198	0,2299	0,2386	0,2463	0,2533	0,2598	0,2658	0,2715	0,2769
51	0,0094	0,0951	0,1434	0,1726	0,1927	0,2078	0,2198	0,2299	0,2386	0,2463	0,2533	0,2598	0,2658	0,2715	0,2769
61	0,0094	0,0951	0,1434	0,1726	0,1927	0,2078	0,2198	0,2299	0,2386	0,2463	0,2533	0,2598	0,2658	0,2715	0,2769
71	0,0094	0,0951	0,1434	0,1726	0,1927	0,2078	0,2198	0,2299	0,2386	0,2463	0,2533	0,2598	0,2658	0,2715	0,2769
81	0,0094	0,0951	0,1434	0,1726	0,1927	0,2078	0,2198	0,2299	0,2386	0,2463	0,2533	0,2598	0,2658	0,2715	0,2769
91	0,0094	0,0951	0,1434	0,1726	0,1927	0,2078	0,2198	0,2299	0,2386	0,2463	0,2533	0,2598	0,2658	0,2715	0,2769
101	0,0094	0,0951	0,1434	0,1726	0,1927	0,2078	0,2198	0,2299	0,2386	0,2463	0,2533	0,2598	0,2658	0,2715	0,2769
111	0,0094	0,0951	0,1434	0,1726	0,1927	0,2078	0,2198	0,2299	0,2386	0,2463	0,2533	0,2598	0,2658	0,2715	0,2769
121	0,0094	0,0951	0,1434	0,1726	0,1927	0,2078	0,2198	0,2299	0,2386	0,2463	0,2533	0,2598	0,2658	0,2715	0,2769
131	0,0094	0,0951	0,1434	0,1726	0,1927	0,2078	0,2198	0,2299	0,2386	0,2463	0,2533	0,2598	0,2658	0,2715	0,2769
141	0,0094	0,0951	0,1434	0,1726	0,1927	0,2078	0,2198	0,2299	0,2386	0,2463	0,2533	0,2598	0,2658	0,2715	0,2769

Πίνακας 6.1 Σφάλματα ωρών 00:00-05:00

Παρατηρούμε ότι το σφάλμα είναι ανεξάρτητο του hx, και εξαρτάται μόνο από το hy. Επιλέγουμε ως hx την τιμή 31 αφού δίνει τα καλύτερα γραφικά αποτελέσματα.

Απ' τις δοκιμές παρατηρείται ελαχιστοποίηση του σφάλματος για hy=0.5, που ήταν το κατώτερο άκρο των δοκιμών μας. Άρα θα αναζητήσουμε και για μικρότερα hy με hx=31.

6:00

H_x \ H_y	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7,5
1	0,007	0,0511	0,0867	0,1157	0,1394	0,159	0,1754	0,1895	0,2017	0,2125	0,2221	0,2309	0,239	0,2466	0,2536
11	0,007	0,0511	0,0867	0,1157	0,1394	0,159	0,1754	0,1895	0,2017	0,2125	0,2221	0,2309	0,239	0,2466	0,2536
21	0,007	0,0511	0,0867	0,1157	0,1394	0,159	0,1754	0,1895	0,2017	0,2125	0,2221	0,2309	0,239	0,2466	0,2536
31	0,007	0,0511	0,0867	0,1157	0,1394	0,159	0,1754	0,1895	0,2017	0,2125	0,2221	0,2309	0,239	0,2466	0,2536
41	0,007	0,0511	0,0867	0,1157	0,1394	0,159	0,1754	0,1895	0,2017	0,2125	0,2221	0,2309	0,239	0,2466	0,2536
51	0,007	0,0511	0,0867	0,1157	0,1394	0,159	0,1754	0,1895	0,2017	0,2125	0,2221	0,2309	0,239	0,2466	0,2536
61	0,007	0,0511	0,0867	0,1157	0,1394	0,159	0,1754	0,1895	0,2017	0,2125	0,2221	0,2309	0,239	0,2466	0,2536
71	0,007	0,0511	0,0867	0,1157	0,1394	0,159	0,1754	0,1895	0,2017	0,2125	0,2221	0,2309	0,239	0,2466	0,2536
81	0,007	0,0511	0,0867	0,1157	0,1394	0,159	0,1754	0,1895	0,2017	0,2125	0,2221	0,2309	0,239	0,2466	0,2536
91	0,007	0,0511	0,0867	0,1157	0,1394	0,159	0,1754	0,1895	0,2017	0,2125	0,2221	0,2309	0,239	0,2466	0,2536
101	0,007	0,0511	0,0867	0,1157	0,1394	0,159	0,1754	0,1895	0,2017	0,2125	0,2221	0,2309	0,239	0,2466	0,2536
111	0,007	0,0511	0,0867	0,1157	0,1394	0,159	0,1754	0,1895	0,2017	0,2125	0,2221	0,2309	0,239	0,2466	0,2536
121	0,007	0,0511	0,0867	0,1157	0,1394	0,159	0,1754	0,1895	0,2017	0,2125	0,2221	0,2309	0,239	0,2466	0,2536
131	0,007	0,0511	0,0867	0,1157	0,1394	0,159	0,1754	0,1895	0,2017	0,2125	0,2221	0,2309	0,239	0,2466	0,2536
141	0,007	0,0511	0,0867	0,1157	0,1394	0,159	0,1754	0,1895	0,2017	0,2125	0,2221	0,2309	0,239	0,2466	0,2536

Πίνακας 6.2 Σφάλματα ώρας 06:00

7:00

H_x \ H_y	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7,5
1	0.0341	0.0564	0.0728	0.0854	0.0956	0.1044	0.1124	0.12	0.1273	0.1346	0.1418	0.1491	0.1564	0.1638	0.1711
11	0.0341	0.0564	0.0728	0.0854	0.0956	0.1044	0.1124	0.12	0.1273	0.1346	0.1418	0.1491	0.1564	0.1638	0.1711
21	0.0341	0.0564	0.0728	0.0854	0.0956	0.1044	0.1124	0.12	0.1273	0.1346	0.1418	0.1491	0.1564	0.1638	0.1711
31	0.0341	0.0564	0.0728	0.0854	0.0956	0.1044	0.1124	0.12	0.1273	0.1346	0.1418	0.1491	0.1564	0.1638	0.1711
41	0.0341	0.0564	0.0728	0.0854	0.0956	0.1044	0.1124	0.12	0.1273	0.1346	0.1418	0.1491	0.1564	0.1638	0.1711
51	0.0341	0.0564	0.0728	0.0854	0.0956	0.1044	0.1124	0.12	0.1273	0.1346	0.1418	0.1491	0.1564	0.1638	0.1711
61	0.0341	0.0564	0.0728	0.0854	0.0956	0.1044	0.1124	0.12	0.1273	0.1346	0.1418	0.1491	0.1564	0.1638	0.1711
71	0.0341	0.0564	0.0728	0.0854	0.0956	0.1044	0.1124	0.12	0.1273	0.1346	0.1418	0.1491	0.1564	0.1638	0.1711
81	0.0341	0.0564	0.0728	0.0854	0.0956	0.1044	0.1124	0.12	0.1273	0.1346	0.1418	0.1491	0.1564	0.1638	0.1711
91	0.0341	0.0564	0.0728	0.0854	0.0956	0.1044	0.1124	0.12	0.1273	0.1346	0.1418	0.1491	0.1564	0.1638	0.1711
101	0.0341	0.0564	0.0728	0.0854	0.0956	0.1044	0.1124	0.12	0.1273	0.1346	0.1418	0.1491	0.1564	0.1638	0.1711
111	0.0341	0.0564	0.0728	0.0854	0.0956	0.1044	0.1124	0.12	0.1273	0.1346	0.1418	0.1491	0.1564	0.1638	0.1711
121	0.0341	0.0564	0.0728	0.0854	0.0956	0.1044	0.1124	0.12	0.1273	0.1346	0.1418	0.1491	0.1564	0.1638	0.1711
131	0.0341	0.0564	0.0728	0.0854	0.0956	0.1044	0.1124	0.12	0.1273	0.1346	0.1418	0.1491	0.1564	0.1638	0.1711
141	0.0341	0.0564	0.0728	0.0854	0.0956	0.1044	0.1124	0.12	0.1273	0.1346	0.1418	0.1491	0.1564	0.1638	0.1711

Πίνακας 6.3 Σφάλματα ώρας 07:00

8:00

Hx \ Hy	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7,5
1	0,0599	0,0677	0,0742	0,0801	0,0856	0,0909	0,0963	0,1018	0,1076	0,1134	0,1193	0,1252	0,1311	0,1371	0,143
11	0,0712	0,0787	0,0849	0,0905	0,0957	0,1007	0,1058	0,111	0,1163	0,1216	0,127	0,1323	0,1377	0,143	0,1483
21	0,0764	0,0836	0,0897	0,0951	0,1001	0,1049	0,1098	0,1148	0,1198	0,1249	0,13	0,1351	0,1402	0,1453	0,1504
31	0,0825	0,0896	0,0956	0,1008	0,1057	0,1104	0,115	0,1197	0,1245	0,1293	0,1341	0,1389	0,1438	0,1485	0,1533
41	0,0868	0,0939	0,0998	0,105	0,1097	0,1143	0,1188	0,1233	0,1279	0,1325	0,1372	0,1418	0,1464	0,151	0,1555
51	0,0896	0,0966	0,1024	0,1075	0,1122	0,1167	0,1211	0,1256	0,1301	0,1346	0,1391	0,1435	0,148	0,1525	0,1569
61	0,0913	0,0983	0,1041	0,1092	0,1138	0,1182	0,1226	0,127	0,1314	0,1358	0,1403	0,1447	0,1491	0,1535	0,1578
71	0,0924	0,0994	0,1052	0,1102	0,1149	0,1193	0,1236	0,1279	0,1323	0,1367	0,1411	0,1454	0,1498	0,1541	0,1584
81	0,0932	0,1002	0,1059	0,111	0,1156	0,12	0,1243	0,1286	0,1329	0,1373	0,1416	0,1459	0,1502	0,1545	0,1588
91	0,0938	0,1007	0,1065	0,1115	0,1161	0,1204	0,1247	0,129	0,1334	0,1377	0,142	0,1463	0,1506	0,1549	0,1591
101	0,0942	0,1011	0,1069	0,1119	0,1165	0,1208	0,1251	0,1294	0,1337	0,138	0,1423	0,1466	0,1508	0,1551	0,1593
111	0,0945	0,1014	0,1072	0,1122	0,1167	0,1211	0,1254	0,1296	0,1339	0,1382	0,1425	0,1468	0,151	0,1553	0,1595
121	0,0947	0,1016	0,1074	0,1124	0,117	0,1213	0,1256	0,1298	0,1341	0,1384	0,1427	0,1469	0,1512	0,1554	0,1596
131	0,0949	0,1018	0,1076	0,1126	0,1171	0,1215	0,1257	0,13	0,1342	0,1385	0,1428	0,147	0,1513	0,1555	0,1597
141	0,0951	0,102	0,1077	0,1127	0,1173	0,1216	0,1259	0,1301	0,1344	0,1386	0,1429	0,1471	0,1514	0,1556	0,1598

Πίνακας 6.4 Σφάλματα ώρας 08:00

9:00

Hx \ Hy	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	
1	0,0416	0,0418	0,0435	0,0457	0,048	0,0504	0,0529	0,0556	0,0587	0,062	0,0656	0,0694	0,0735	0,0778	0,0823
11	0,0547	0,058	0,0614	0,0644	0,0671	0,0697	0,0722	0,0748	0,0775	0,0803	0,0833	0,0863	0,0894	0,0927	0,0961
21	0,0678	0,0714	0,0749	0,078	0,0807	0,0832	0,0857	0,0882	0,0908	0,0935	0,0963	0,0992	0,1021	0,1052	0,1083
31	0,0786	0,0823	0,0858	0,0888	0,0914	0,0939	0,0963	0,0988	0,1013	0,104	0,1066	0,1094	0,1122	0,115	0,118
41	0,0877	0,0913	0,0947	0,0976	0,1002	0,1026	0,105	0,1074	0,1098	0,1123	0,1148	0,1174	0,1201	0,1228	0,1256
51	0,0952	0,0988	0,1021	0,1049	0,1074	0,1097	0,112	0,1143	0,1167	0,119	0,1215	0,124	0,1265	0,1291	0,1317
61	0,1015	0,105	0,1082	0,111	0,1134	0,1156	0,1178	0,1201	0,1223	0,1247	0,127	0,1294	0,1318	0,1343	0,1368
71	0,1069	0,1103	0,1134	0,1161	0,1185	0,1207	0,1228	0,125	0,1272	0,1295	0,1317	0,134	0,1364	0,1388	0,1412
81	0,1116	0,1149	0,118	0,1206	0,1229	0,1251	0,1272	0,1293	0,1315	0,1337	0,1359	0,1381	0,1404	0,1427	0,1451
91	0,1156	0,1189	0,1219	0,1245	0,1268	0,1289	0,131	0,1331	0,1352	0,1373	0,1395	0,1417	0,1439	0,1462	0,1485
101	0,1191	0,1224	0,1253	0,1278	0,1301	0,1322	0,1342	0,1363	0,1384	0,1405	0,1426	0,1448	0,1469	0,1491	0,1514
111	0,1221	0,1253	0,1282	0,1307	0,1329	0,135	0,137	0,1391	0,1411	0,1432	0,1453	0,1474	0,1495	0,1517	0,1539
121	0,1246	0,1278	0,1307	0,1331	0,1353	0,1374	0,1394	0,1414	0,1435	0,1455	0,1476	0,1496	0,1517	0,1538	0,156
131	0,1268	0,1299	0,1327	0,1352	0,1374	0,1394	0,1414	0,1434	0,1455	0,1475	0,1495	0,1515	0,1536	0,1557	0,1578
141	0,1286	0,1317	0,1345	0,137	0,1391	0,1411	0,1431	0,1451	0,1471	0,1491	0,1511	0,1532	0,1552	0,1572	0,1593

Πίνακας 6.5 Σφάλματα ώρας 09:00

10:00

Hx \ Hy	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7.5
1	0,0985	0,0972	0,0961	0,0955	0,0954	0,0956	0,0961	0,0968	0,0975	0,0981	0,0988	0,0995	0,1003	0,101	0,1019
11	0,0789	0,0791	0,0793	0,0798	0,0807	0,0819	0,0833	0,0849	0,0866	0,0882	0,0899	0,0916	0,0933	0,0951	0,0968
21	0,0764	0,0767	0,0771	0,0779	0,0791	0,0807	0,0826	0,0845	0,0865	0,0884	0,0904	0,0924	0,0943	0,0962	0,0982
31	0,0809	0,0817	0,0826	0,0839	0,0856	0,0875	0,0896	0,0917	0,0937	0,0957	0,0977	0,0996	0,1015	0,1034	0,1053
41	0,0881	0,0894	0,0907	0,0923	0,0943	0,0964	0,0986	0,1007	0,1027	0,1047	0,1065	0,1083	0,1101	0,1119	0,1136
51	0,0962	0,0977	0,0991	0,1009	0,103	0,1052	0,1073	0,1094	0,1113	0,1132	0,1149	0,1166	0,1182	0,1198	0,1214
61	0,1043	0,1059	0,1074	0,1093	0,1114	0,1135	0,1156	0,1176	0,1194	0,1212	0,1228	0,1243	0,1258	0,1272	0,1287
71	0,1122	0,1139	0,1154	0,1173	0,1194	0,1215	0,1235	0,1254	0,1271	0,1287	0,1302	0,1317	0,133	0,1343	0,1356
81	0,1198	0,1215	0,123	0,1249	0,1269	0,1289	0,1309	0,1327	0,1344	0,1359	0,1373	0,1386	0,1399	0,1411	0,1423
91	0,1269	0,1286	0,1301	0,1319	0,1339	0,1359	0,1378	0,1395	0,1411	0,1425	0,1438	0,1451	0,1463	0,1474	0,1485
101	0,1334	0,1351	0,1366	0,1384	0,1403	0,1422	0,144	0,1457	0,1472	0,1485	0,1498	0,151	0,1521	0,1531	0,1542
111	0,1393	0,1409	0,1424	0,1441	0,146	0,1479	0,1496	0,1512	0,1526	0,1539	0,1551	0,1563	0,1573	0,1583	0,1592
121	0,1445	0,1461	0,1476	0,1493	0,1511	0,1529	0,1546	0,1561	0,1575	0,1587	0,1599	0,1609	0,1619	0,1629	0,1638
131	0,1491	0,1507	0,1522	0,1538	0,1556	0,1573	0,1589	0,1604	0,1618	0,163	0,1641	0,1651	0,166	0,1669	0,1678
141	0,1532	0,1548	0,1562	0,1578	0,1595	0,1612	0,1628	0,1642	0,1655	0,1667	0,1677	0,1687	0,1696	0,1705	0,1713

Πίνακας 6.6 Σφάλματα ώρας 10:00

11:00

Hx \ Hy	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7.5
1	0,154	0,1379	0,1326	0,1279	0,1223	0,1166	0,1114	0,1071	0,1037	0,1012	0,0995	0,0983	0,0976	0,0972	0,0972
11	0,1128	0,1096	0,1081	0,1066	0,1047	0,1029	0,1012	0,0998	0,0987	0,0979	0,0974	0,0971	0,097	0,0971	0,0973
21	0,1046	0,1026	0,1013	0,1	0,0988	0,0976	0,0966	0,0958	0,0953	0,0949	0,0948	0,0948	0,095	0,0954	0,0959
31	0,1004	0,0989	0,0979	0,0968	0,0959	0,095	0,0943	0,0939	0,0936	0,0936	0,0937	0,094	0,0944	0,0949	0,0956
41	0,0966	0,0955	0,0947	0,094	0,0934	0,0929	0,0926	0,0925	0,0925	0,0927	0,0931	0,0936	0,0942	0,095	0,0958
51	0,0941	0,0934	0,0929	0,0925	0,0922	0,0921	0,0921	0,0922	0,0925	0,093	0,0936	0,0943	0,095	0,0959	0,0969
61	0,0931	0,0928	0,0925	0,0924	0,0924	0,0925	0,0928	0,0931	0,0937	0,0943	0,095	0,0958	0,0967	0,0977	0,0987
71	0,0935	0,0934	0,0934	0,0935	0,0937	0,094	0,0945	0,095	0,0956	0,0964	0,0972	0,0981	0,099	0,1	0,1011
81	0,095	0,0952	0,0953	0,0956	0,0959	0,0964	0,097	0,0976	0,0983	0,0991	0,1	0,1009	0,1019	0,1029	0,104
91	0,0974	0,0977	0,098	0,0984	0,0988	0,0994	0,1	0,1008	0,1015	0,1024	0,1033	0,1042	0,1052	0,1062	0,1072
101	0,1003	0,1009	0,1012	0,1017	0,1022	0,1028	0,1035	0,1043	0,1051	0,1059	0,1068	0,1077	0,1087	0,1097	0,1107
111	0,1037	0,1043	0,1048	0,1053	0,1058	0,1065	0,1072	0,108	0,1088	0,1096	0,1105	0,1114	0,1123	0,1133	0,1142
121	0,1072	0,1079	0,1084	0,109	0,1096	0,1102	0,111	0,1118	0,1126	0,1134	0,1142	0,1151	0,116	0,1169	0,1178
131	0,1108	0,1116	0,1121	0,1126	0,1133	0,1139	0,1147	0,1154	0,1162	0,117	0,1178	0,1186	0,1195	0,1203	0,1212
141	0,1143	0,1151	0,1156	0,1162	0,1168	0,1175	0,1182	0,1189	0,1197	0,1205	0,1212	0,122	0,1228	0,1236	0,1245

Πίνακας 6.7 Σφάλματα ώρας 11:00

12:00

Hx \ Hy	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7.5
1	0,2013	0,2011	0,1981	0,1951	0,1916	0,1875	0,1829	0,1779	0,1729	0,168	0,1634	0,1591	0,1552	0,1516	0,1483
11	0,1381	0,1375	0,1361	0,1347	0,133	0,131	0,1287	0,1265	0,1243	0,1224	0,1207	0,1192	0,1178	0,1167	0,1156
21	0,1267	0,1249	0,1237	0,1227	0,1214	0,1198	0,1181	0,1164	0,1148	0,1134	0,1122	0,1111	0,1103	0,1095	0,1089
31	0,1251	0,1225	0,1212	0,1201	0,1187	0,1172	0,1156	0,1141	0,1127	0,1115	0,1105	0,1097	0,109	0,1085	0,1081
41	0,1213	0,1185	0,1172	0,116	0,1147	0,1133	0,1119	0,1106	0,1095	0,1085	0,1078	0,1072	0,1067	0,1064	0,1062
51	0,1166	0,1139	0,1127	0,1116	0,1104	0,1092	0,108	0,107	0,1061	0,1054	0,1049	0,1045	0,1043	0,1041	0,1041
61	0,1119	0,1095	0,1084	0,1075	0,1064	0,1054	0,1045	0,1037	0,1031	0,1026	0,1023	0,1021	0,102	0,1021	0,1022
71	0,1076	0,1055	0,1046	0,1038	0,103	0,1022	0,1015	0,101	0,1006	0,1003	0,1002	0,1002	0,1002	0,1004	0,1006
81	0,1039	0,1021	0,1014	0,1008	0,1002	0,0996	0,0992	0,0989	0,0987	0,0986	0,0986	0,0987	0,0989	0,0992	0,0995
91	0,101	0,0995	0,099	0,0986	0,0982	0,0978	0,0976	0,0974	0,0974	0,0975	0,0976	0,0979	0,0982	0,0986	0,099
101	0,0989	0,0978	0,0974	0,0971	0,0969	0,0967	0,0967	0,0967	0,0968	0,097	0,0973	0,0976	0,098	0,0985	0,099
111	0,0976	0,0968	0,0966	0,0965	0,0964	0,0964	0,0965	0,0966	0,0969	0,0972	0,0976	0,098	0,0985	0,099	0,0995
121	0,0971	0,0966	0,0965	0,0965	0,0965	0,0967	0,0969	0,0972	0,0975	0,0979	0,0983	0,0988	0,0993	0,0999	0,1005
131	0,0972	0,097	0,097	0,0971	0,0973	0,0975	0,0978	0,0982	0,0986	0,0991	0,0996	0,1001	0,1006	0,1012	0,1018
141	0,098	0,098	0,0981	0,0983	0,0985	0,0988	0,0992	0,0996	0,1001	0,1006	0,1011	0,1017	0,1023	0,1028	0,1035

Πίνακας 6.8 Σφάλματα ώρας 12:00

13:00

Hx \ Hy	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7.5
1	0,1911	0,1865	0,1835	0,1801	0,1746	0,1674	0,1596	0,1522	0,1457	0,1402	0,1357	0,132	0,1289	0,1265	0,1245
11	0,1386	0,1355	0,1334	0,1307	0,1274	0,1238	0,1202	0,117	0,1143	0,112	0,1101	0,1087	0,1075	0,1067	0,1061
21	0,1026	0,1009	0,0998	0,0986	0,0972	0,0958	0,0944	0,0932	0,0921	0,0913	0,0908	0,0904	0,0903	0,0904	0,0907
31	0,0959	0,0946	0,0938	0,0928	0,0919	0,0908	0,0899	0,089	0,0883	0,0878	0,0875	0,0875	0,0876	0,0878	0,0883
41	0,0933	0,0923	0,0915	0,0907	0,0898	0,089	0,0881	0,0874	0,0869	0,0866	0,0864	0,0864	0,0866	0,087	0,0875
51	0,0919	0,0909	0,0902	0,0895	0,0888	0,088	0,0873	0,0867	0,0863	0,0861	0,086	0,0861	0,0864	0,0868	0,0874
61	0,0906	0,0898	0,0892	0,0885	0,0879	0,0873	0,0867	0,0863	0,086	0,0859	0,0859	0,0861	0,0865	0,087	0,0876
71	0,0895	0,0888	0,0882	0,0877	0,0872	0,0867	0,0863	0,086	0,0859	0,0859	0,0861	0,0864	0,0868	0,0874	0,0881
81	0,0886	0,088	0,0876	0,0872	0,0868	0,0865	0,0862	0,0861	0,0861	0,0863	0,0865	0,0869	0,0874	0,0881	0,0888
91	0,088	0,0876	0,0873	0,0871	0,0868	0,0866	0,0865	0,0866	0,0867	0,087	0,0874	0,0878	0,0884	0,0891	0,0899
101	0,088	0,0878	0,0876	0,0874	0,0873	0,0873	0,0873	0,0875	0,0878	0,0882	0,0886	0,0892	0,0899	0,0906	0,0915
111	0,0885	0,0885	0,0884	0,0884	0,0884	0,0885	0,0887	0,089	0,0894	0,0898	0,0904	0,091	0,0918	0,0926	0,0934
121	0,0896	0,0897	0,0898	0,0898	0,09	0,0902	0,0905	0,0909	0,0914	0,092	0,0926	0,0933	0,0941	0,095	0,0959
131	0,0912	0,0915	0,0917	0,0919	0,0921	0,0924	0,0929	0,0934	0,094	0,0946	0,0953	0,0961	0,0969	0,0978	0,0988
141	0,0934	0,0939	0,0942	0,0944	0,0947	0,0952	0,0957	0,0963	0,097	0,0977	0,0985	0,0993	0,1002	0,1011	0,1021

Πίνακας 6.9 Σφάλματα ώρας 13:00

14:00

Hx \ Hy	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7.5
1	0,1889	0,1871	0,1838	0,18	0,1751	0,1696	0,1641	0,159	0,1544	0,1504	0,147	0,1443	0,1421	0,1404	0,1392
11	0,0761	0,0753	0,0743	0,0735	0,0725	0,0715	0,0706	0,07	0,0697	0,0699	0,0704	0,0713	0,0725	0,074	0,0758
21	0,0573	0,0574	0,0578	0,0583	0,0586	0,0589	0,0593	0,0598	0,0607	0,0618	0,0632	0,0648	0,0666	0,0686	0,0708
31	0,0545	0,0545	0,0547	0,055	0,0552	0,0555	0,056	0,0567	0,0576	0,0588	0,0602	0,0618	0,0637	0,0657	0,0679
41	0,0536	0,0534	0,0535	0,0536	0,0538	0,0541	0,0545	0,0552	0,0562	0,0574	0,0588	0,0604	0,0623	0,0643	0,0665
51	0,0533	0,0532	0,0532	0,0533	0,0534	0,0537	0,0542	0,0549	0,0559	0,0571	0,0585	0,0602	0,062	0,064	0,0662
61	0,0534	0,0533	0,0533	0,0534	0,0536	0,054	0,0545	0,0552	0,0562	0,0575	0,0589	0,0606	0,0624	0,0644	0,0665
71	0,0537	0,0536	0,0537	0,0539	0,0541	0,0545	0,0551	0,0559	0,057	0,0582	0,0597	0,0614	0,0632	0,0652	0,0673
81	0,0542	0,0541	0,0543	0,0545	0,0549	0,0554	0,056	0,0569	0,058	0,0593	0,0609	0,0625	0,0644	0,0664	0,0685
91	0,0549	0,0549	0,0552	0,0555	0,0559	0,0565	0,0572	0,0582	0,0594	0,0607	0,0623	0,064	0,0659	0,0679	0,07
101	0,0559	0,056	0,0563	0,0567	0,0572	0,0578	0,0587	0,0597	0,061	0,0624	0,064	0,0658	0,0677	0,0697	0,0718
111	0,0571	0,0573	0,0578	0,0583	0,0588	0,0596	0,0605	0,0616	0,0629	0,0644	0,0661	0,0679	0,0698	0,0718	0,0739
121	0,0588	0,0591	0,0596	0,0602	0,0608	0,0617	0,0627	0,0639	0,0652	0,0668	0,0685	0,0703	0,0722	0,0743	0,0764
131	0,0608	0,0612	0,0618	0,0625	0,0632	0,0641	0,0652	0,0665	0,068	0,0696	0,0713	0,0732	0,0751	0,0771	0,0793
141	0,0632	0,0637	0,0645	0,0653	0,0661	0,0671	0,0682	0,0696	0,0711	0,0727	0,0745	0,0764	0,0784	0,0804	0,0826

Πίνακας 6.10 Σφάλματα ώρας 14:00

15:00

Hx \ Hy	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7.5
1	0,0538	0,0488	0,0469	0,0467	0,0473	0,0484	0,0501	0,0527	0,0562	0,0604	0,0651	0,0704	0,0759	0,0817	0,0877
11	0,0405	0,0429	0,0453	0,0475	0,0501	0,0534	0,0572	0,0613	0,0658	0,0704	0,0752	0,0801	0,0851	0,0902	0,0954
21	0,0379	0,0411	0,0436	0,046	0,0489	0,0524	0,0564	0,0608	0,0654	0,0701	0,075	0,0799	0,085	0,0902	0,0955
31	0,0365	0,0394	0,0417	0,0441	0,047	0,0505	0,0545	0,0588	0,0633	0,0679	0,0727	0,0776	0,0827	0,0879	0,0931
41	0,0357	0,0384	0,0405	0,0428	0,0456	0,049	0,0528	0,0569	0,0612	0,0658	0,0705	0,0753	0,0803	0,0855	0,0908
51	0,0354	0,0377	0,0397	0,0419	0,0446	0,0478	0,0514	0,0554	0,0596	0,064	0,0686	0,0734	0,0784	0,0835	0,0887
61	0,0352	0,0375	0,0393	0,0414	0,044	0,0471	0,0506	0,0545	0,0586	0,0629	0,0674	0,0722	0,0771	0,0822	0,0874
71	0,0354	0,0375	0,0393	0,0414	0,0439	0,047	0,0504	0,0542	0,0582	0,0625	0,0669	0,0716	0,0765	0,0815	0,0868
81	0,0358	0,0379	0,0397	0,0418	0,0443	0,0473	0,0507	0,0545	0,0584	0,0626	0,0671	0,0717	0,0766	0,0816	0,0868
91	0,0364	0,0387	0,0405	0,0426	0,0451	0,0481	0,0515	0,0552	0,0592	0,0633	0,0677	0,0723	0,0771	0,0821	0,0873
101	0,0374	0,0397	0,0416	0,0437	0,0463	0,0494	0,0527	0,0564	0,0603	0,0645	0,0689	0,0734	0,0782	0,0831	0,0882
111	0,0386	0,041	0,043	0,0452	0,0479	0,0509	0,0543	0,058	0,0619	0,066	0,0704	0,0749	0,0796	0,0845	0,0895
121	0,0401	0,0427	0,0448	0,0471	0,0498	0,0529	0,0563	0,0599	0,0638	0,0679	0,0722	0,0767	0,0814	0,0862	0,0912
131	0,042	0,0448	0,047	0,0493	0,0521	0,0552	0,0586	0,0623	0,0662	0,0702	0,0745	0,0789	0,0835	0,0883	0,0932
141	0,0443	0,0473	0,0495	0,052	0,0548	0,0579	0,0614	0,065	0,0689	0,0729	0,0771	0,0815	0,086	0,0907	0,0955

Πίνακας 6.11 Σφάλματα ώρας 15:00

16:00

Hx \ Hy	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7.5
1	0,0168	0,0252	0,0368	0,0514	0,0681	0,0854	0,1023	0,1182	0,133	0,1467	0,1593	0,171	0,1819	0,192	0,2015
11	0,0137	0,0239	0,0413	0,0617	0,0823	0,1015	0,119	0,1347	0,1487	0,1614	0,173	0,1835	0,1931	0,2021	0,2105
21	0,0119	0,0213	0,0379	0,0582	0,079	0,0985	0,1163	0,1322	0,1465	0,1594	0,1711	0,1817	0,1915	0,2005	0,2089
31	0,0111	0,0202	0,0367	0,0569	0,0777	0,0972	0,115	0,1309	0,1452	0,158	0,1697	0,1803	0,1901	0,1991	0,2076
41	0,0109	0,0199	0,0365	0,0567	0,0774	0,0969	0,1145	0,1304	0,1446	0,1574	0,169	0,1796	0,1893	0,1983	0,2068
51	0,0107	0,0199	0,0367	0,0569	0,0775	0,0968	0,1143	0,1301	0,1442	0,1569	0,1684	0,179	0,1887	0,1977	0,2061
61	0,0108	0,0201	0,037	0,0572	0,0776	0,0968	0,1141	0,1297	0,1438	0,1564	0,1679	0,1784	0,1881	0,1971	0,2055
71	0,0109	0,0205	0,0375	0,0576	0,0778	0,0967	0,1139	0,1294	0,1433	0,1558	0,1672	0,1777	0,1874	0,1963	0,2048
81	0,0112	0,0211	0,0382	0,0581	0,0781	0,0967	0,1137	0,1289	0,1427	0,1551	0,1665	0,1769	0,1865	0,1955	0,2039
91	0,0117	0,0219	0,0389	0,0587	0,0783	0,0967	0,1134	0,1285	0,142	0,1544	0,1656	0,1759	0,1855	0,1944	0,2028
101	0,0123	0,0227	0,0398	0,0593	0,0786	0,0967	0,1131	0,1279	0,1413	0,1534	0,1645	0,1747	0,1842	0,1931	0,2014
111	0,0132	0,0238	0,0407	0,06	0,079	0,0967	0,1128	0,1273	0,1405	0,1524	0,1634	0,1734	0,1828	0,1916	0,1998
121	0,0142	0,025	0,0419	0,0608	0,0794	0,0967	0,1125	0,1267	0,1396	0,1513	0,1621	0,172	0,1813	0,1899	0,1981
131	0,0154	0,0265	0,0432	0,0618	0,08	0,0969	0,1122	0,1262	0,1388	0,1503	0,1608	0,1706	0,1797	0,1882	0,1963
141	0,017	0,0282	0,0447	0,0629	0,0807	0,0972	0,1122	0,1257	0,138	0,1493	0,1596	0,1692	0,1781	0,1865	0,1945

Πίνακας 6.12 Σφάλματα ώρας 16:00

17:00

Hx \ Hy	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7.5
1	0,0094	0,0958	0,1445	0,174	0,1942	0,2093	0,2213	0,2313	0,2399	0,2476	0,2545	0,2609	0,2669	0,2725	0,2779
11	0,0088	0,0934	0,1424	0,1722	0,1927	0,208	0,2202	0,2303	0,239	0,2468	0,2538	0,2603	0,2663	0,272	0,2774
21	0,0087	0,093	0,1421	0,1719	0,1924	0,2078	0,22	0,2301	0,2389	0,2466	0,2537	0,2601	0,2662	0,2719	0,2773
31	0,0087	0,0927	0,1416	0,1715	0,1919	0,2072	0,2194	0,2296	0,2383	0,2461	0,2531	0,2596	0,2657	0,2714	0,2768
41	0,0087	0,0921	0,1406	0,1701	0,1904	0,2056	0,2177	0,2279	0,2367	0,2445	0,2516	0,2581	0,2642	0,27	0,2754
51	0,0087	0,091	0,1388	0,168	0,1881	0,2031	0,2152	0,2254	0,2342	0,242	0,2492	0,2558	0,262	0,2679	0,2735
61	0,0087	0,0894	0,1365	0,1653	0,1852	0,2002	0,2123	0,2225	0,2314	0,2393	0,2466	0,2533	0,2597	0,2656	0,2713
71	0,0087	0,0877	0,1338	0,1623	0,182	0,197	0,2091	0,2194	0,2284	0,2365	0,2439	0,2507	0,2572	0,2633	0,2691
81	0,0087	0,0857	0,1308	0,1588	0,1784	0,1934	0,2055	0,2159	0,2251	0,2333	0,2409	0,2479	0,2545	0,2607	0,2667
91	0,0087	0,0834	0,1273	0,1549	0,1743	0,1893	0,2016	0,2121	0,2215	0,2299	0,2376	0,2448	0,2516	0,2579	0,264
101	0,0088	0,081	0,1236	0,1505	0,1697	0,1848	0,1972	0,2079	0,2174	0,2261	0,234	0,2414	0,2483	0,2548	0,2611
111	0,0089	0,0785	0,1196	0,1458	0,1648	0,1799	0,1924	0,2033	0,2131	0,2219	0,23	0,2376	0,2447	0,2514	0,2578
121	0,0091	0,0759	0,1154	0,1409	0,1597	0,1747	0,1873	0,1984	0,2083	0,2174	0,2257	0,2335	0,2408	0,2477	0,2542
131	0,0094	0,0734	0,1112	0,1359	0,1543	0,1693	0,182	0,1932	0,2033	0,2126	0,2211	0,2291	0,2366	0,2436	0,2503
141	0,0098	0,071	0,1071	0,131	0,149	0,1638	0,1765	0,1879	0,1981	0,2075	0,2162	0,2244	0,232	0,2392	0,2461

Πίνακας 6.13 Σφάλματα ώρας 17:00

18:00-23:00

Hx \ Hy	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7.5
1	0,0094	0,0958	0,1445	0,174	0,1942	0,2093	0,2213	0,2313	0,2399	0,2476	0,2545	0,2609	0,2669	0,2725	0,2779
11	0,0094	0,0958	0,1445	0,174	0,1942	0,2093	0,2213	0,2313	0,2399	0,2476	0,2545	0,2609	0,2669	0,2725	0,2779
21	0,0094	0,0958	0,1445	0,174	0,1942	0,2093	0,2213	0,2313	0,2399	0,2476	0,2545	0,2609	0,2669	0,2725	0,2779
31	0,0094	0,0958	0,1445	0,174	0,1942	0,2093	0,2213	0,2313	0,2399	0,2476	0,2545	0,2609	0,2669	0,2725	0,2779
41	0,0094	0,0958	0,1445	0,174	0,1942	0,2093	0,2213	0,2313	0,2399	0,2476	0,2545	0,2609	0,2669	0,2725	0,2779
51	0,0094	0,0958	0,1445	0,174	0,1942	0,2093	0,2213	0,2313	0,2399	0,2476	0,2545	0,2609	0,2669	0,2725	0,2779
61	0,0094	0,0958	0,1445	0,174	0,1942	0,2093	0,2213	0,2313	0,2399	0,2476	0,2545	0,2609	0,2669	0,2725	0,2779
71	0,0094	0,0958	0,1445	0,174	0,1942	0,2093	0,2213	0,2313	0,2399	0,2476	0,2545	0,2609	0,2669	0,2725	0,2779
81	0,0094	0,0958	0,1445	0,174	0,1942	0,2093	0,2213	0,2313	0,2399	0,2476	0,2545	0,2609	0,2669	0,2725	0,2779
91	0,0094	0,0958	0,1445	0,174	0,1942	0,2093	0,2213	0,2313	0,2399	0,2476	0,2545	0,2609	0,2669	0,2725	0,2779
101	0,0094	0,0958	0,1445	0,174	0,1941	0,2092	0,2212	0,2312	0,2399	0,2475	0,2545	0,2609	0,2669	0,2725	0,2779
111	0,0094	0,0958	0,1444	0,1739	0,1941	0,2092	0,2212	0,2312	0,2398	0,2475	0,2544	0,2608	0,2668	0,2725	0,2778
121	0,0094	0,0957	0,1444	0,1738	0,194	0,2091	0,2211	0,2311	0,2398	0,2474	0,2544	0,2608	0,2668	0,2724	0,2778
131	0,0094	0,0957	0,1443	0,1738	0,1939	0,209	0,221	0,231	0,2397	0,2474	0,2543	0,2607	0,2667	0,2724	0,2777
141	0,0094	0,0957	0,1443	0,1737	0,1939	0,2089	0,2209	0,231	0,2396	0,2473	0,2543	0,2607	0,2667	0,2723	0,2777

Πίνακας 6.14 Σφάλματα ωρών 18:00-23:00

Στη συνέχεια για τις ώρες στις οποίες δεν προέκυψε τοπικό ελάχιστο κάνουμε μία περαιτέρω διερεύνηση. Όπως είναι φανερό από τους παραπάνω πίνακες έχουν βρεθεί οι βέλτιστες τιμές των H_x για όλες τις ώρες, άρα αναζητούμε τα βέλτιστα H_y για ορισμένες εξ' αυτών. Οι ώρες που χρειάζονται περαιτέρω διερεύνηση είναι όλες εκτός από 11:00-14:00. Συνεπώς υπολογίζουμε τα σφάλματα για διάφορα H_y, με δεδομένο το βέλτιστο H_x που αντιστοιχεί στην κάθε ώρα

Time \ Hy	0.05	0.1	0.2	0.3	0.4	0.5
0-5	-	-	0.9947	0.1127	**0.0017**	0.0094
6	-	-	0.3321	0.0384	**0.0024**	0.007
7	-	-	0.0650	**0.0256**	0.0282	0.0341
8	0.0744	**0.0261**	0.0378	0.0506	0.0569	0.0599
9	-	-	0.0436	0.0421	0.0418	**0.0416**
10	-	-	0.0851	0.0761	**0.0759**	0.0764
15	-	-	0.0544	**0.0334**	0.0338	0.0352
16	-	-	0.0544	0.0107	**0.0101**	0.0107
17	-	-	0.9053	0.1037	**0.0016**	0.0087
18-23	-	-	0.9998	0.1137	**0.0017**	0.0094

Πίνακας 6.15 Διερεύνηση για βέλτιστα Hy

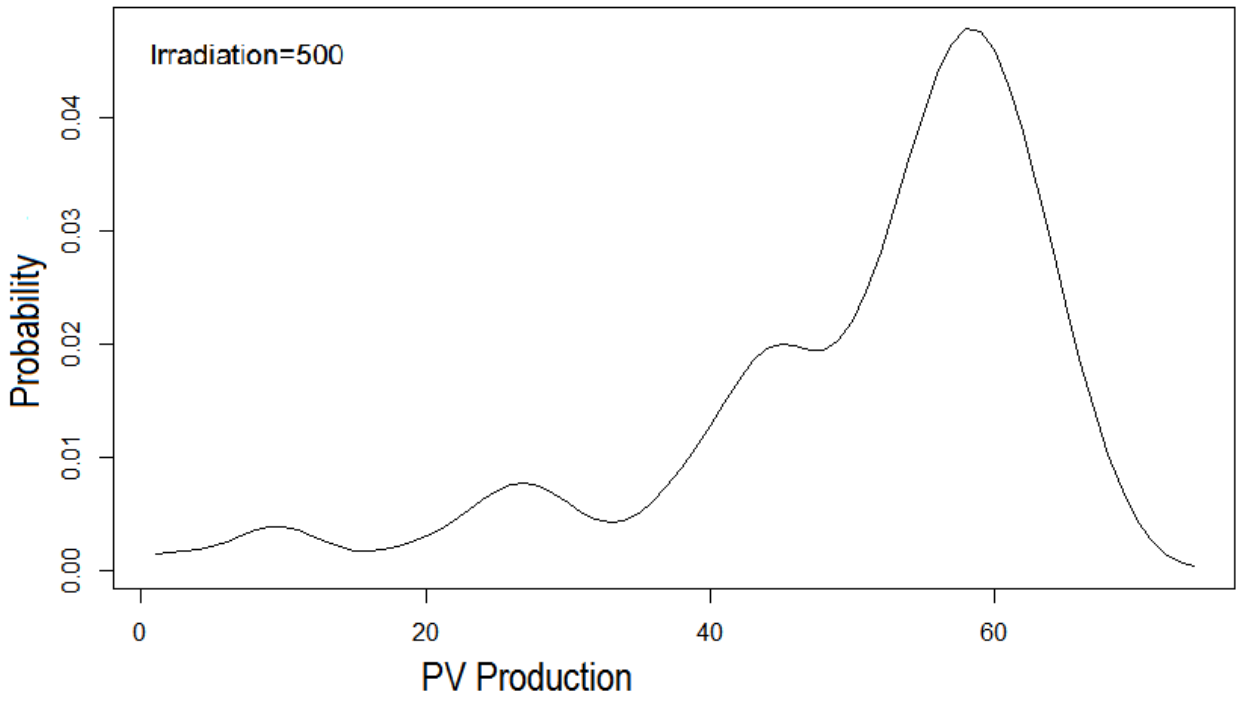
Συνοπτικά έχουμε τον εξής πίνακα τιμών

Time	Hx	Hy	Error
0	31	0.4	0.0017
1	31	0.4	0.0017
2	31	0.4	0.0017
3	31	0.4	0.0017
4	31	0.4	0.0017
5	31	0.4	0.0017
6	31	0.4	0.0024
7	31	0.3	0.0256
8	1	0.1	0.0261
9	1	0.5	0.0416
10	21	0.4	0.0759
11	51	3	0.0921
12	111	3	0.0964
13	61	5	0.0859
14	51	1	0.0532
15	61	0.3	0.0334
16	51	0.4	0.0101
17	31	0.4	0.0016
18	31	0.4	0.0017
19	31	0.4	0.0017
20	31	0.4	0.0017
21	31	0.4	0.0017
22	31	0.4	0.0017
23	31	0.4	0.0017

Πίνακας 6.16 Αναλυτικά σφάλματα ωρών

Το συνολικό σφάλμα της μεθόδου είναι ο μέσος όρος των παραπάνω Error για τις βέλτιστες τιμές παραμέτρων, δηλαδή RPS=0.0244.

Τέλος, παρουσιάζουμε ενδεικτικά τη μορφή μιας καμπύλης εκτιμήτριας συνάρτησης, με βάση τους βέλτιστους συντελεστές που επιλέχθηκαν παραπάνω. Επιλέγοντας τυχαία ως ώρα 12:00 και με δεδομένη ακτινοβολία τιμής $500 W/m^2$



Σχήμα 6.4 Παράδειγμα γραφικής παράστασης εκτιμήτριας συνάρτησης.

6.3 Σύγκριση με άλλες μεθόδους

Αφού βρήκαμε τους κατάλληλους συντελεστές μπορούμε να υπολογίσουμε το συνολικό σφάλμα της μεθόδου μας ως το μέσο όρο όλων των σφαλμάτων για την κάθε ώρα.

Στη συνέχεια θα συγκρίνουμε αυτό το σφάλμα με κάποιες μεθόδους που θα χρησιμοποιηθούν ως σημεία αναφοράς όπως αναλύθηκε στα παραπάνω κεφάλαια. Αυτές οι μέθοδοι, οι οποίες αναλύθηκαν στα προηγούμενα κεφάλαια της θεωρίας είναι οι εξής:

- Εκτιμήτρια με πυρήνα (Kernel Density Estimation)- χωρίς διαχωρισμό ανά ώρα
- Απλή γραμμική παλινδρόμηση- χωρίς διαχωρισμό ανά ώρα
- Απλή γραμμική παλινδρόμηση- με διαχωρισμό ανά ώρα

6.3.1 Εκτιμήτρια με πυρήνα (Kernel Density Estimation)- χωρίς διαχωρισμό ανά ώρα

Η συγκεκριμένη τεχνική δεν κατηγοριοποιεί τα δεδομένα μας ανά ώρα, άλλα τα επεξεργάζεται στο σύνολό τους. Είναι η συνήθης τεχνική που εφαρμόζεται στην περίπτωση των εκτιμητριών. Ωστόσο, όπως διαπιστώσαμε στη συνέχεια τα αποτελέσματα που παρουσίαζε ήταν στατιστικά ασήμαντα και το σφάλμα της μεγάλο, αφού αναλόγως την ώρα αλλάζει αισθητά και η διακύμανση των βασικών μας μεταβλητών. Τρέχουμε λοιπόν ακριβώς τον ίδιο αλγόριθμο που αναλύθηκε παραπάνω παραλείποντας το πρώτο βήμα που διαχωρίζει τα δεδομένα ανά ώρα.

Ο κώδικας υπολογισμού είναι ο εξής:

```
error<-array(NA, dim=c(15,15))
f<-array(NA, dim=c(74,950))
data_uncertainty<-read.csv("C:/Users/Fotis/Desktop/Diplomatiki/Code/data uncertainty.csv",stringsAsFactors = F)
data_uncertainty<-na.omit(data_uncertainty)
for (hx in seq(0.4, 0.9, by = 0.1)){    #10X10epanalipseis
  for (hy in seq(0.1, 0.4 , by = 0.1)){
  h12<-data_uncertainty
  for(y in 1:74) {   #74 epanalipseis mexri to maxPV(72.2) apo 0 mexri 73
    for(x in seq(10, 950, by = 10)) { #935 epanalipseis mexti to maxIrradiation
      A<-0
      B<-0
      for(i in 1:(nrow(h12)-10)) {
        Xt<- h12$Irradiation[i]
        Yt<- h12$PV.production[i]
        Ky<-exp(-0.5* ((Yt-(y-1))/hy)^2)
        Kx<-exp(-0.5* ((Xt-(x-10))/hx)^2)
        A<-Ky*Kx +A
        B<-Kx + B
      }
      f[y,x]<- A/(B*hy*sqrt(2*pi)) #synarthsh f(y,x) mas dinei pithanotita gia y.PV, me x.Irradiation
      if (is.na( f[y,x]) )  f[y,x]<-0
     }
  }
#METRAW TA ERRORS
  error[hx*10,hy*10]<-0
  for(k in 197:206) {
    trIr<-round(h12$Irradiation[k],-1)  #strogulopoihsh stin pio kontini 10ada
    trPV<-round(h12$PV.production[k])
    sum<-0
    rps<-0
    for (i in 1:74) {
      sum1<-0
      sum2<-0
      for (j in 1:i) {
        sum1<-sum1 + f[j,trIr+10]
      }
      if ((j-1)>= trPV) sum2<-1
      sum<- sum + (sum1-sum2)^2
    }
    rps<- sum/73 #diairw me plithos outcomes-1
    error[hx*10,hy*10]<-error[hx*10,hy*10]+rps
  }
  error[hx*10,hy*10]<-error[hx*10,hy*10]/10
}
}
```

Σχήμα 6.5 Κώδικας Kernel χωρίς ομαδοποίηση δεδομένων κατά ώρα στην R.

Και στη συγκεκριμένη περίπτωση αναζητούμε τα βέλτιστα εύρη ζώνης. Στον παρακάτω πίνακα υπολογίσαμε τα σφάλματα και αναζητούμε το ελάχιστο σφάλμα, το οποίο μας παρέχει και το βέλτιστο συνδυασμό. Και αυτή τη φορά έχουμε τα εξής όρια:

- ✓ Το h_x πήρε τιμές από 1 έως 141 με βήμα 10.
- ✓ Το h_y πήρε τιμές από 0.5 έως 7.5 με βήμα 0.5.

Σε περίπτωση που δε βρέθηκε μέσα στο πλέγμα το τοπικό ελάχιστο, τότε εφαρμόζουμε περαιτέρω διερεύνηση στις τιμές που παρουσιάζεται το μεγαλύτερο ενδιαφέρον.

Όπως φαίνεται και στους παρακάτω πίνακες οι βέλτιστες τιμές παραμέτρων σε αυτή τη μέθοδο είναι h_x=1 και h_y=0.5. Το σφάλμα σε αυτήν την περίπτωση είναι RPS=0.1152.

Σε σχέση με το 0.0244 της προτεινόμενης μεθόδου είναι 5 φορές μεγαλύτερο, άρα η μέθοδος του διαχωρισμού ανά ώρα είναι πολύ πιο αποδοτική και ακριβής, όπως εξάλλου ήταν αναμενόμενο.

Hy Hx	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**	**4**	**4.5**	**5**
1	***0.1152***	0.1333	0.1470	0.1556	0.1615	0.1659	0.1694	0.1723	0.1749	0.1773
11	0.1189	0.1354	0.1483	0.1564	0.1620	0.1662	0.1696	0.1724	0.1749	0.1772
21	0.1246	0.1409	0.1540	0.1622	0.1678	0.1719	0.1750	0.1777	0.1800	0.1822
31	0.1263	0.1431	0.1561	0.1643	0.1698	0.1738	0.1769	0.1794	0.1817	0.1838
41	0.1273	0.1442	0.1573	0.1654	0.1708	0.1747	0.1778	0.1803	0.1826	0.1846
51	0.1284	0.1452	0.1581	0.1662	0.1716	0.1755	0.1785	0.1811	0.1833	0.1853
61	0.1302	0.1465	0.1592	0.1672	0.1725	0.1764	0.1795	0.1820	0.1842	0.1862
71	0.1320	0.1477	0.1602	0.1681	0.1734	0.1773	0.1804	0.1829	0.1851	0.1872
81	0.1336	0.1488	0.1611	0.1689	0.1742	0.1781	0.1811	0.1836	0.1859	0.1879
91	0.1349	0.1496	0.1618	0.1695	0.1748	0.1787	0.1817	0.1843	0.1865	0.1885

Πίνακας 6.17 Σφάλματα Kernel χωρίς ομαδοποίηση δεδομένων κατά ώρα(i)

Διερευνούμε και για μικρότερες τιμές.

Hy Hx	0.1	0.2	0.3	0.4
0.4	4.3600	0.6970	0.2275	0.1438
0.5	4.2212	0.6721	0.2186	0.1380
0.6	4.1030	0.6507	0.2114	0.1337
0.7	4.0045	0.6318	0.2051	0.1303
0.8	3.9252	0.6158	0.1996	0.1274
0.9	3.8628	0.6027	0.1950	0.1249

Πίνακας 6.18 Σφάλματα Kernel χωρίς ομαδοποίηση δεδομένων κατά ώρα(ii)

6.3.2 Απλή γραμμική παλινδρόμηση(χωρίς διαχωρισμό ανά ώρα)

Στη συγκεκριμένη τεχνική όπως αναφέρθηκε και στο 2ο κεφάλαιο της παρούσας εργασίας, παράγουμε προβλέψεις αξιοποιώντας μία ευθεία προσέγγιση που προβλέπει μία ποσοτική απόκριση Y(Παραγόμενη Ενέργεια) με βάση μία μόνο ανεξάρτητη μεταβλητή X(Ακτινοβολία).

Αρχικά εφαρμόσαμε την απλή γραμμική παλινδρόμηση πάνω σε όλα τα δεδομένα μας χωρίς να κάνουμε διαχωρισμό ανά ώρα. Ο κώδικας που τρέξαμε ήταν ο εξής:

```
1 data_uncertainty<-read.csv("C:/Users/Fotis/Desktop/Diplomatiki/Code/data uncertainty.csv",stringsAsFactors = F)
2 data_uncertainty<-na.omit(data_uncertainty)
3 h12<-data_uncertainty
4 simple.fit = lm(PV.production~Irradiation, data=h12)
5 summary(simple.fit)
```

Σχήμα 6.6 Κώδικας απλής γραμμικής παλινδρόμησης στην R.

Λάβαμε τις εξής τιμές:

```
Residual standard error: 14.89 on 4942 degrees of freedom
Multiple R-squared:  0.4142,    Adjusted R-squared:  0.4141
F-statistic:  3494 on 1 and 4942 DF,  p-value: < 2.2e-16
```

Δηλαδή παρατηρούμε ότι επιβεβαιώνεται ο έλεγχος μηδενικής υπόθεσης αφού το p-value λαμβάνει πολύ μικρή τιμή. Συνεπώς συμπεραίνουμε ότι υπάρχει γραμμική συσχέτιση των μεταβλητών μας. Επίσης το F-statistic>>1 άρα κι αυτό επαληθεύει ότι το μοντέλο μας είναι αποδοτικό. Ωστόσο οι τιμές του R^2=0.4142 δείχνουν κάποια απόκλιση από την επιθυμητή τιμή 1. Συνεπώς θα δούμε το σφάλμα που μας δίνει η συγκεκριμένη τεχνική και στη συνέχεια επιλέγουμε το διαχωρισμό ανά ώρα και τον υπολογισμό μιας ξεχωριστής ευθείας για την κάθε ώρα της μέρας, ώστε να ελαχιστοποιήσουμε τα σφάλματα.

Για να συγκρίνουμε αυτή τη μέθοδο που παράγει σημειακές προβλέψεις με την προηγούμενη μέθοδο που είναι πιθανοτική, βρίσκουμε το διάστημα εμπιστοσύνης του 95% για την προβλεπόμενη τιμή και στη συνέχεια μοιράζουμε το συγκεκριμένο διάστημα σε πιθανότητες. Θεωρούμε ότι στο συγκεκριμένο διάστημα, ακολουθώντας το νόμο των μεγάλων αριθμών, η τυχαία μεταβλητή μας(προβλεπόμενη τιμή) ακολουθεί την κανονική κατανομή. Στη συνέχεια, διακριτοποιούμε το συγκεκριμένο διάστημα, για να έχει την ίδια μορφή πίνακα με αυτήν της εκτιμήτριας συνάρτησης. Με αυτόν τον τρόπο συγκρίνουμε εν τέλει το σφάλμα μεταξύ δύο συναρτήσεων πυκνότητας πιθανότητας, που μας παρέχουν πληροφορία για το επίπεδο της παραγόμενης ηλεκτρικής ενέργειας και την αντίστοιχη πιθανότητα του, με δεδομένο το επίπεδο της ακτινοβολίας.

Οφείλουμε να τονίσουμε σε αυτό το σημείο ότι για ομοιομορφία στη σύγκριση των αποτελεσμάτων, κρατάμε και πάλι κρυφές τις τελευταίες 10 παρατηρήσεις των δεδομένων μας, πάνω στις οποίες θα γίνει ο υπολογισμός του σφάλματος.

Έτσι ο κώδικας που τρέξαμε τελικά είναι ο εξής:

```
data_uncertainty<-read.csv("C:/Users/user/Desktop/Diplomatiki/Code/data uncertainty.csv",stringsAsFactors = F)
data_uncertainty<-na.omit(data_uncertainty)
h12<-data_uncertainty
error<-array(NA, dim=c(10))

trainset<-data_uncertainty
#model <- lm(PV.production~Irradiation, data=trainset)

data12=trainset[1:(nrow(trainset)-10),]
test12=trainset[(nrow(trainset)-9):nrow(trainset),]
model2=lm(PV.production~Irradiation, data=data12)
summary(model2)
frc <- predict(model2,interval="predict", newdata = test12)
mean_frc <- frc[,1]
low_frc <- frc[,2]
up_frc <- frc[,3]
sd=(up_frc-low_frc)/4
for(ia in 1:10) {

  dist <- rnorm(1000, mean_frc[ia], sd[ia])
  model <- density(dist)
  production <- model$x
  probability <- model$y/sum(model$y)
  solution <- data.frame(production, probability)
  if (nrow(solution[solution$production<0,])>0){
    solution[solution$production<0,]$production <- 0
  }

  k<-array(NA, dim=c(74,950))

  for(y in 1:74) {   #74 epanalipseis mexri to maxPV(72.2) apo 0 mexri 73
    for(x in seq(10, 950, by = 10)) { #935 epanalipseis mexti to maxIrradiation

      k[y,x]<-0
    }
  }
```

```
  TrIr<-round(h12$Irradiation[4934+ia],-1)
  for(y in 1:74) {
    for( i in 1:512)
      if(solution$production[i]<=y && solution$production[i]>=y-1 )
        k[y,TrIr+10]<-k[y,TrIr+10]+solution$probability[i]
  }


  error[ia]<-0
  kl=4934+ia
    TrIr<-round(h12$Irradiation[kl],-1)  #strogulopoihsh stin pio kontini 10ada
    TrPV<-round(h12$PV.production[kl])
    sum<-0
    rps<-0

    for (cnt in 1:74) {
      sum1<-0
      sum2<-0
      for (j in 1:cnt) {
        sum1<-sum1 + k[j,TrIr+10]
      }
      if ((j-1)>= TrPV) sum2<-1
      sum<- sum + (sum1-sum2)^2
    }
    rps<- sum/73 #diairw me plithos outcomes-1
    error[ia]<-error[ia]+rps

}
error1<-0
for(ioi in 1:10) {
  error1<-error1+error[ioi]
}
error1<-error1/10
```

Σχήμα 6.7 Κώδικας απλής γραμμικής παλινδρόμησης και υπολογισμού RPS στην R.

Το συνολικό σφάλμα της συγκριμένης μεθόδου είναι RPS=0.051. Είναι μεγαλύτερο της εκτιμήτριας συνάρτησης που διαχωρίζει ανά ώρα τα δεδομένα, αλλά μικρότερο αυτής που τα υπολογίζει συνολικά.

6.3.3 Απλή γραμμική παλινδρόμηση(με διαχωρισμό ανά ώρα)

Ακολουθούμε ακριβώς την ίδια μέθοδο με το προηγούμενο κεφάλαιο, με μόνη διαφορά ότι τώρα κατηγοριοποιούμε τα δεδομένα ανά ώρα ώστε να επιτύχουμε καλύτερη προσέγγιση. Ο κώδικας που τρέξαμε για να διαχωρίσουμε ανά ώρα είναι ο εξής:

```
data_uncertainty<-read.csv("C:/Users/user/Desktop/Diplomatiki/Code/data uncertainty.csv",stringsAsFactors = F)
data_uncertainty<-na.omit(data_uncertainty)
h12<-data_uncertainty
error<-array(NA, dim=c(24, 10))
for (i in 1:nrow(data_uncertainty)){
  if (nchar(data_uncertainty$datetime[i])==13){
    data_uncertainty$Hour[i]<-as.numeric(substr(data_uncertainty$datetime[i],10,10))
  }
  else{
    data_uncertainty$Hour[i]<-as.numeric(substr(data_uncertainty$datetime[i],10,11))
  }
}
  f<-array(NA, dim=c(74,950,24))

  for(time in 0:23) {
    #if (time>=18 | time<=7) hy<-0.39687
    h12<-data_uncertainty[data_uncertainty$Hour==time,] #trexw kata wra
    trainset<-h12

    #predictor PV -dependent IR
    #simple.fit = lm(PV.production~Irradiation, data=h12)
    #summary(simple.fit)

    data12=h12[1:(nrow(h12)-10),]
    test12=h12[(nrow(h12)-9):nrow(h12),]
    model2=lm(PV.production~Irradiation, data=data12)
    summary(model2)
    #pred12 <- predict(model2, newdata = test12)
    frc <- predict(model2,interval="predict", newdata = test12)
    mean_frc <- frc[,1]
    low_frc <- frc[,2]
    up_frc <- frc[,3]
    sd=(up_frc-low_frc)/4


        for(ia in 1:10) {

          dist <- rnorm(1000, mean_frc[ia], sd[ia])
          model <- density(dist)
          production <- model$x
          probability <- model$y/sum(model$y)
          solution <- data.frame(production, probability)
          if (nrow(solution[solution$production<0,])>0){
            solution[solution$production<0,]$production <- 0
          }

          k<-array(NA, dim=c(74,950))

          for(y in 1:74) {   #74 epanalipseis mexri to maxPV(72.2) apo 0 mexri 73
            for(x in seq(10, 950, by = 10)) { #935 epanalipseis mexti to maxIrradiation

              k[y,x]<-0
            }
          }
          kapa<-nrow(h12)-9
          TrIr<-round(h12$Irradiation[kapa],-1)
          for(y in 1:74) {
            for( i in 1:512)
              if(solution$production[i]<=y && solution$production[i]>=y-1 )
                k[y,TrIr+10]<-k[y,TrIr+10]+solution$probability[i]
          }


          error[time+1, ia]<-0

          TrIr<-round(h12$Irradiation[kapa],-1)  #strogulopoihsh stin pio kontini 10ada
          TrPV<-round(h12$PV.production[kapa])
          sum<-0
          rps<-0
```

```
    for (cnt in 1:74) {
      sum1<-0
      sum2<-0
      for (j in 1:cnt) {
        sum1<-sum1 + k[j,TrIr+10]
      }
      if ((j-1)>= TrPV) sum2<-1
      sum<- sum + (sum1-sum2)^2
    }
    rps<- sum/73 #diairw me plithos outcomes-1
    error[time+1, ia]<-error[time+1,ia]+rps

  }
  error1[time+1]<-0
  for(ioi in 1:10) {
    error1[time+1]<-error1[time+1]+error[time+1,ioi]
  }
  error1[time+1]<-error1[time+1]/10
  kapa<-kapa+1

}
```

Σχήμα 6.8 Κώδικας απλής γραμμικής παλινδρόμησης στην R με διαχωρισμό ανά ώρα και υπολογισμού σφάλματος.

Τα αναλυτικά ωριαία σφάλματα της συγκεκριμένης μεθόδου είναι τα εξής:

Time	Error (LRL ανά ώρα)	Error (Kernel ανά ώρα)
0	0.0001	0.0017
1	0.0001	0.0017
2	0.0001	0.0017
3	0.0001	0.0017
4	0.0001	0.0017
5	0.0001	0.0017
6	0.0015	0.0024
7	0.0363	0.0256
8	0.0786	0.0261
9	0.0929	0.0416
10	0.0538	0.0759
11	0.0718	0.0921
12	0.1032	0.0964
13	0.0491	0.0859
14	0.0552	0.0532
15	0.0569	0.0334
16	0.0200	0.0101
17	0.0046	0.0016
18	0.0009	0.0017
19	0.0003	0.0017
20	0.0001	0.0017
21	0.0001	0.0017
22	0.0001	0.0017
23	0.0001	0.0017

Πίνακας 6.19 Σφάλματα γραμμικής παλινδρόμησης και εκτιμήτριας συνάρτησης με ομαδοποίηση δεδομένων κατά ώρα

Το συνολικό σφάλμα της μεθόδου είναι RPS=0.0261. Παρατηρούμε ότι είναι μεγαλύτερο από την εκτιμήτρια που διαχωρίζει ανά ώρα τα δεδομένα μας αλλά καλύτερο από την αντίστοιχη εκτιμήτρια που δε διαχωρίζει ανά ώρα καθώς και από την αντίστοιχη γραμμική παλινδρόμηση που δε διαχωρίζει ανά ώρα, όπως και ήταν αναμενόμενο.

Ειδικότερα, αν συγκρίνουμε την ακρίβεια των δύο μεθόδων που διαχωρίζουν ανά συγκεκριμένες ώρες, παρατηρούμε ότι η προτεινόμενη μέθοδος της εκτιμήτριας συνάρτησης παρουσιάζει καλύτερα αποτελέσματα, τις ώρες αιχμής που μας ενδιαφέρουν, δηλαδή τις ώρες που παρατηρείται ηλιοφάνεια. Από τις 7:00 έως τις 17:00 παρατηρούμε μεγαλύτερη ακρίβεια με την εκτιμήτρια συνάρτηση με εξαίρεση τις ώρες 10:00, 11:00 και 13:00, όπου οριακά η γραμμική παλινδρόμηση παράγει καλύτερα αποτελέσματα. Το γεγονός αυτό δικαιολογείται από την τυχαιότητα των δεδομένων αλλά και της μεγάλης διακύμανσης τους που οφείλεται στο μεγάλο χρονικό διάστημα της συλλογής τους, που περιλαμβάνει εαρινούς και θερινούς μήνες, δηλαδή εποχές με μεγάλη διαφορά στην ηλιοφάνεια. Τις ώρες που δεν παρατηρείται ηλιοφάνεια, βλέπουμε ότι η γραμμική παλινδρόμηση είναι καλύτερη, μιας και η μέθοδος μας είναι πιο ευαίσθητη σε τιμές που παρατηρείται ηλιακή ακτινοβολία ενώ το επίπεδο παραγωγής είναι μηδενικό. Ωστόσο οι συγκεκριμένες ώρες δεν παρουσιάζουν πρακτικό ενδιαφέρον στη μελέτη μας. Σε μια πιο εξειδικευμένη μελέτη με ετήσια δεδομένα, η εφαρμογή της μεθόδου της εκτιμήτριας θα μπορούσε να παράγει ακόμα καλύτερα αποτελέσματα, με μεγαλύτερη ακρίβεια. Συνολικά έχουμε τα εξής:

Method	RPS Error
Kernel(Ανά ώρα)	0.0240
Kernel(καθολικό)	0.1152
Linear regression(καθολικό)	0.0521
Linear regression(ανά ώρα)	0.0261

Πίνακας 6.20 Συγκεντρωτικός πίνακας σφαλμάτων

Συγκεντρωτικά φαίνεται ξεκάθαρα ότι η μέθοδός μας δίνει τα καλύτερα αποτελέσματα, έχοντας το μικρότερο σφάλμα, και όπως αναλύθηκε προηγουμένως λειτουργεί καλύτερα από όλες στις ώρες αιχμής, που μας ενδιαφέρουν πραγματικά.

Κεφάλαιο 7: Συμπεράσματα και Προεκτάσεις

7.1 Σύνοψη αποτελεσμάτων

Η παρούσα εργασία αποτελείται από δύο βασικά μέρη. Στο πρώτο μέρος γίνεται μία βιβλιογραφική επισκόπηση στατιστικών τεχνικών που χρησιμοποιούνται ευρέως από οργανισμούς, εταιρείες και ερευνητές για την παραγωγή προβλέψεων. Πιο συγκεκριμένα βλέπουμε πως αυτές οι προβλέψεις αξιοποιούνται στον ενεργειακό τομέα και αναδεικνύουμε τα πλεονεκτήματα που η καθεμία εξ αυτών προσφέρει. Στο δεύτερο μέρος αναπτύσσεται και εφαρμόζεται μία μεθοδολογία παραγωγής προβλέψεων ενεργειακής παραγωγής με σκοπό την βέλτιστη οργάνωση ενός σταθμού που αξιοποιεί ΑΠΕ και πιο συγκεκριμένα την ηλιακή ακτινοβολία.

Βασικό σημείο της μελέτης είναι η επιλογή μιας τεχνικής πρόβλεψης που περιγράφει καλύτερα τα επίπεδα παραγωγής, λαμβάνοντας υπ' όψιν τη στοχαστική φύση των μετεωρολογικών προβλέψεων. Αυτό γίνεται όπως είδαμε με τη χρήση στατιστικών μεθόδων πρόβλεψης, επιλογή που από τα αποτελέσματα της εργασίας φαίνεται να μας δικαιώνει. Έχοντας ως κριτήριο το δείκτη σφάλματος RPS, η τεχνική που εφαρμόσαμε παρουσίασε μικρά σφάλματα και σε σύγκριση με διάφορες άλλες γνωστές μεθόδους ήταν πιο αποδοτική, ειδικά στις ώρες μεγάλου ενδιαφέροντος.

Πιο συγκεκριμένα, η μέθοδος της εκτιμήτριας συνάρτησης που διαχωρίζει ανά ώρα τα δεδομένα είχε μικρότερο σφάλμα τόσο από την εκτιμήτρια που αναλύει τα δεδομένα ολιστικά, χωρίς δηλαδή κάποια κατηγοριοποίηση, όσο και από την απλή γραμμική παλινδρόμηση. Σε ότι αφορά τη σύγκριση της προτεινόμενης μεθόδου με την απλή γραμμική παλινδρόμηση που διαχωρίζει τα δεδομένα ανά ώρα, παρατηρήθηκε ότι στις περισσότερες ώρες αιχμής η μέθοδος μας είχε μεγαλύτερη ακρίβεια, με εξαίρεση τις νυχτερινές ώρες όπου η ευθεία της γραμμικής παλινδρόμησης προσαρμόζεται καλύτερα. Επίσης, ήταν οριακά πιο ακριβής και τις ώρες 10:00, 11:00 και 13:00, γεγονός που οφείλεται στη μεγάλη διακύμανση των δεδομένων αλλά και του χρονικού διαστήματος συλλογής τους που εκτείνεται σε 4 μήνες με αρκετά διαφορετικές καιρικές συνθήκες.

Θέλοντας κάποιος να κάνει μία γενική αποτίμηση της μεθόδου, θα μπορούσε να πει ότι δεδομένης των διαθέσιμων στοιχείων και του βάθους μελέτης που μία διπλωματική εργασία μπορεί να έχει, η μέθοδος βρέθηκε να έχει εξαιρετική απόδοση εφόσον γνωρίζαμε την επικείμενη ακτινοβολία και ανοίγει ένα πιο ενδιαφέρον πεδίο στον τομέα των Τεχνικών Προβλέψεων αφού ξεφεύγει απ' τη λογική των σημειακών προβλέψεων και εισέρχεται στη σύγχρονη εποχή των πιθανοτήτων, αντιμετωπίζοντας το πρόβλημα στοχαστικά με συναρτήσεις κατανομής. Η ακρίβεια μας ήταν αρκετά ικανοποιητική και γίνεται φανερό πως αν υπήρχαν περισσότερα στοιχεία για την παραγωγή πιο αντιπροσωπευτικών προβλέψεων, τα αποτελέσματα της μεθόδου θα ήταν σίγουρα ακόμα καλύτερα.

Σημαντικό βήμα λοιπόν στην εφαρμογή μιας τέτοιας μεθόδου για τη παραγωγή ορθών προβλέψεων, είναι όχι μόνο η σωστή εφαρμογή της, αλλά και η ύπαρξη πληθώρας στοιχείων στα χέρια του μελετητή για την εύρεση των πιο αντιπροσωπευτικών συναρτήσεων κατανομής και την αξιοποίησή τους. Σύμμαχός του σε αυτή την έρευνα είναι πάντα οι διαθέσιμοι σε βιβλιογραφία δείκτες αν και όπως φάνηκε και παραπάνω ο πειραματισμός και η εμπειρία του είναι εκείνα που θα του υποδείξουν το βέλτιστο δρόμο στην επιλογή των καταλληλότερων δεικτών.

7.2 Μελλοντικές προεκτάσεις

Όπως αναφέρθηκε στην εισαγωγή, η χρήση ΑΠΕ στην ενεργειακή παραγωγή είναι μία ταχέως εξελισσόμενη τεχνική και το μέλλον της είναι βαρύνουσας σημασίας για την παγκόσμια οικονομία. Η αβεβαιότητα που εμπεριέχουν τα μετεωρολογικά μοντέλα καθιστούν απαραίτητη την όσο το δυνατόν καλύτερη ακρίβεια προβλέψεων ενεργειακής παραγωγής με στόχο την καλύτερη διαχείριση και αξιοποίηση των υπαρχουσών εγκαταστάσεων και τεχνολογιών.

Από οικονομικής απόψεως, η πρόβλεψη ενεργειακών παραγωγών μπορεί να αποτελέσει και ένα βασικό εργαλείο στα χέρια του manager της κάθε επιχείρησης, διευκολύνοντας τον να προβλέπει τις οικονομικές υποχρεώσεις του ως προς τις εταιρείες παροχής ενέργειας και να σχεδιάζει αποτελεσματικότερα το οικονομικό και ενεργειακό πλάνο της επιχείρησής του.

Ως προς τις μελλοντικές επεκτάσεις της μεθόδου που παρουσιάστηκε στην παρούσα διπλωματική εργασία, μπορούμε να αναφέρουμε αρκετά πράγματα. Αρχικά, η μέθοδος μπορεί να τροποποιηθεί ώστε να εφαρμοστεί και σε άλλες εγκαταστάσεις που αξιοποιούν ΑΠΕ (πχ. ανεμογεννήτριες, γεωθερμικές πηγές, υδροηλεκτρικές εγκαταστάσεις).

Αρκετό ενδιαφέρον θα μπορούσε να έχει και ο έλεγχος της αποδοτικότητας του μοντέλου σε περίπτωση που οι προβλέψεις των παραγωγών γινόντουσαν συνδυάζοντας περισσότερες εξαρτημένες μεταβλητές, πέρα της ηλιακής ακτινοβολίας, όπως ο άνεμος, η θερμοκρασία, η πίεση του αέρα κ.α.

Επιπλέον, η επεξεργασία δεδομένων μεγαλύτερου όγκου, που θα περιλάμβαναν μετρήσεις στη διάρκεια ενός ολόκληρου ημερολογιακού έτους, θα μπορούσε να συμβάλλει στην ακόμα πιο εξειδικευμένη εφαρμογή της μεθόδου ώστε να διαχωρίζονται τα δεδομένα μας, όχι μόνο ωριαία αλλά και μηνιαία με στόχο την απεμπλοκή του μοντέλου μας από περιττά σφάλματα που προκύπτουν από τις διαφορετικές καιρικές συνθήκες που επικρατούν τις ίδιες ώρες της μέρας αλλά σε διαφορετικούς μήνες.

Τέλος, βασική επέκταση στο να φανεί η συγκεκριμένη μεθοδολογία πρακτικά χρήσιμη θα ήταν η δημιουργία μία ηλεκτρονικής εφαρμογής η οποία θα λάμβανε πραγματικά στοιχεία από το σύστημα monitoring της εγκατάστασης και που θα εκτελούσε τη μέθοδο για τη παραγωγή προβλέψεων, ώστε να επιτυγχάνει την κατάλληλη διαχείριση των επιπέδων παραγωγής.

Βιβλιογραφία

[1] Raid B. I. Salha «Εκτίμηση πυρήνων των δεσμευμένων ποσοστημορίων και επικρατούσας τιμής χρονικών σειρών», Διδακτορική διατριβή, Θεσσαλονική 2006

[2] B.W. Silverman «Denisty Estimation for statistics and data analysis» London: Chapman and Hall, 1986.

[3] Φώτιος Χ. Καράμερος. *«Πιθανοτική πρόβλεψη Ηλιακής παραγωγής με χρήση ARTMAP»*, Διπλωματική εργασία, Αθήνα 2013

[4] Μιχ. Π. Παπαδόπουλος. *«Παραγωγή Ηλεκτρικής Ενέργειας από Ανανεώσιμες Πηγές».* ΕΜΠ. Αθήνα 1997.

[5] Pierre Pinson. *"Estimation of the uncertainty in wind power forecasting".* 23/3/2006. Paris.

[6] Γκίκας Αντώνιος, *«Ατμοσφαιρικά μοντέλα πρόγνωσης καιρού»,* πτυχιακή εργασία, Χανιά 2004

[7] Σταύρος Παπαθανασίου. *«Εκτίμηση της ενεργειακής απόδοσης αιολικών πάρκων»*. ΕΜΠ. Νοέμβριος 2005

[8] Δημήτριος Ν. Τόλης. *«Πιθανοτική πρόβλεψη της αιολικής ισχύος κοντά στο όριο αποκοπής της ταχύτητας του ανέμου»*, Διπλωματική εργασία, Αθήνα 2011

[9] Bruce E. Hansen "Nonparametric Conditional Density Estimation", University of Wisconsin, November 2004

[10] James W. Taylor , Jooyoung Jeon "Using Conditional Kernel Density Estimation for Wind Power Density Forecasting", Journal of the American Statistical Association, 2012, Vol. 107, pp. 66-79

[11] James W. Taylor , Jooyoung Jeon "Forecasting Wind Power Quantiles Using Conditional Kernel Estimation", Renewable Energy, 2015, Vol. 80, pp. 370-379

[12] P. Pinson, C. Chevallier, G. Kariniotakis, '*Trading wind generation with short-term probabilistic forecasts of wind power*', IEEE Trans. on Power Systems 22 (3), pp. 1148-1156, 2007

[13] Hyndman et al. 1996 –" Estimating and visualizing conditional densities"

[14] Fan et al. 1996- "Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems"

[15] Edward S. Epstein, "Stochastic dynamic prediction", University of Michigan 1969

[16] David M. Bashtannyk, Rob J. Hyndman, "Bandwidth selection for kernel conditional density estimation", Department of Econometrics and Business Statistics, Monash University, Clayton, Melbourne, Vic. 3800, Australia 2000

[17] Μπουτσικούδη Χ. Σοφία, «Πιθανοτική πρόβλεψη Αιολικής ισχύος», Διπλωματική εργασία, Αθήνα 2009

[18] P. Pinson, G. Kariniotakis, H. Madsen, H.Aa. Nielsen Jan K. Moller "*Non-Parametric Probabilistic Forecasts of Wind Power: Required Properties and Evaluation*" Research Article 18 April 2007.

[19] G. Kariniotakis et al. "Evaluation of Advanced Wind Power Forecasting Models –Results of the Anemos Project", EWEC, Athens 2006 128

[20] V. Guénard, G. Kariniotakis, I. Martí, , "ANEMOS Advanced Wind Power Forecasting. Operational Challenges and On-line Performance". Proc. of EWEC'07, Milan, Italy, 7-10 May 2007

[21] Β. Ασημακόπουλος, Φ. Πετρόπουλος, «Επιχειρησιακές προβλέψεις», Αθήνα 2011

[22] G. Kariniotakis, P. McSharry, P. Pinson, R. Girard, "*Methodology for the evaluation of probabilistic Forecasts*", SafeWind deliverable, October 2009

[23] Σπηλιώτης, Γ. Κοκκολάκης «*Εισαγωγή στη θεωρία των πιθανοτήτων και Στατιστική, Έκδοση 3η* » Εκδόσεις Συμεών ,Αθήνα 1999.

[24] McCulloch, W.S. & W. Pitts,1943. "*A logical calculus of the ideas immanent in nervous activity*", Bulletin of Mathematical Biophysics, issue 5, pp. 115-133

[25] Churchland, P.S. & T.J. Sejnowski, 1992. "*The computational Brain*", Cambridge, MA: MIT Press

[26] G. Siderados and N.D. Hatziargyriou, "Wind power forecasting focused on extreme power system events".

[27] E.D. Castronuovo, J.A. Pecas Lopes, "On the optimization of the daily operation of a wind-hydro power plant", IEEE Trans. on Power Systems, vol. 19, pp. 1599–1606, 2004.

[28] G. N. Bathurst, J.Weatherhill, and G. Strbac, "Trading wind generation in short-term energy markets," IEEE Trans. Power System, vol. 17, no. 3, pp. 782–789, Aug. 2002. 130

[29] R. Doherty, M. O'Malley, "A new approach to quantify reserve demand in systems with significant installed wind capacity", IEEE Trans. On Power Systems, vol. 20, pp. 587-5952005

[30] T.S.Karakatsanis,N.D.Hatziargyriou, "Probabilistic Constrained Load Flow based on Sensitivity Analysis" ,*IEEE Trans on Power Systems*, Vol 9,No 4, November 1994,pp 1853-1860

[31] G. Giebel, L. Landberg, J. Badger, K. Sattler, H. Feddersen, T.S. Nielsen, H.Aa. Nielsen, H. Madsen, "Using Ensemble Forecasting for Wind Power", EWEC'03, Madrid, 16-20 June, 2003

[32] J.W. Taylor, P.E. McSharry, and R. Buizza, "Wind Power Density Forecasting Using Ensemble Predictions and Time Series Models", IEEE Trans. on Energy Conversion, vol. 24, pp. 775-782, 2009

[33] J. B. Bremnes, "Probabilistic wind power forecasts using local quantile regression", Wind Energy, vol. 7, no 1, pp. 47–54, 2004.

[34] H. A. Nielsen, H. Madsen, and T. S. Nielsen, "Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts", Wind Energy, vol 9, no 1-2, pp. 95–108, 2006.

[35] T. S.Nielsen and H.Madsen, 'Statistical methods for predicting wind power', in Proceedings of the 1997 EuropeanWind Energy Conference, EWEC'97, Dublin, Ireland, pp. 755–758, 1997

[36] J. Juban, L. Fugon, and G. N. Kariniotakis, "Probabilistic short-term wind power forecasting based on kernel density estimators," Proc. of the EWEC'07, Milan, Italy, May 2007.

[37] R.J. Bessa, V. Miranda, A. Botterud, Z. Zhou, J. Wang, "Time-Adaptive Quantile-Copula for Wind Power Probabilistic Forecasting," Renewable Energy, Vol. 40, No. 1, pp. 29-39, 2012.

[38] P. Pinson, G. Kariniotakis, "On-line assessment of prediction risk for wind power production forecasts" Wind Energy, vol 7, pp. 119–132, 2004.

[39] M. Lange, "Analysis of the Uncertainty of Wind Power Predictions" PhD dissertation, Carl von Ossietzky Oldenburg University, 2003.

[40] Giebel G., "The State of the Art in Short-Term Prediction of Wind Power - A Literature Overview, 2nd Edition" Deliverable 1.2b of the ANEMOS.plus project. Available: Anemos-plus.eu.

[41] J. Usaola, O. Ravelo, G. Gonzalez, F. Soto, C. Davila, B.Diaz-Guerra, "Benefits for wind energy in electricity markets from using short term wind power prediction tools: a simulation study", Wind Engineering, vol. 28, no.1, pp. 119-128, 2004