

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ



ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΚΑΙ ΠΡΟΒΛΕΨΕΙΣ ΣΕ
ΠΕΡΙΒΑΛΛΟΝΤΟΛΟΓΙΚΑ ΜΟΝΤΕΛΑ ΜΕ ΤΗ
ΧΡΗΣΗ ΤΗΣ ΘΕΩΡΙΑΣ ΑΚΡΑΙΩΝ ΤΙΜΩΝ

ΒΑΣΙΛΕΙΟΣ Ε. ΠΑΡΤΣΑΚΟΥΛΑΚΗΣ

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

ΠΕΙΡΑΙΑΣ
ΜΑΡΤΙΟΣ 2012

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ



ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΚΑΙ ΠΡΟΒΛΕΨΕΙΣ ΣΕ
ΠΕΡΙΒΑΛΛΟΝΤΟΛΟΓΙΚΑ ΜΟΝΤΕΛΑ ΜΕ ΤΗ
ΧΡΗΣΗ ΤΗΣ ΘΕΩΡΙΑΣ ΑΚΡΑΙΩΝ ΤΙΜΩΝ

ΒΑΣΙΛΕΙΟΣ Ε. ΠΑΡΤΣΑΚΟΥΛΑΚΗΣ

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

ΠΕΙΡΑΙΑΣ

ΜΑΡΤΙΟΣ 2012

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίστηκε από τη Γ.Σ.Ε.Σ. του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμό συνεδρίαση του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική.

Τα μέλη της Επιτροπής ήταν:

- (Επιβλέπων)
-
-

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

**UNIVERSITY OF PIRAEUS
DEPARTMENT OF STATISTICS AND INSURANCE SCIENCE**

**POSTGRADUATE PROGRAM IN
APPLIED STATISTICS**



**STATISTICAL ANALYSIS AND FORECASTS IN
ENVIRONMENTAL MODELS USING EXTREME
VALUE THEORY**

By

Vasileios E. Partsakoulakis

MSc Dissertation

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

PIRAEUS, GREECE

MARCH 2012

This thesis was approved unanimously by the three-member committee appointed by the Department of Statistics and Insurance Science, University of Piraeus, in accordance with the rules of the MSc program in Applied Statistics.

Committee members were:

- (Supervisor)

-

-

Approval of this Thesis from the Department of Statistics and Insurance Science, University of Piraeus, does not imply any endorsement of the opinions of the author.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΔΑΙΑ

Στην οικογένεια μου

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά όσους συνετέλεσαν στην ολοκλήρωση της παρούσης διπλωματικής εργασίας. Αρχικά θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή κ. Μιχάλη Μπούτσικα, για τη γνώση και τις συμβουλές που μου προσέφερε καθ'όλη την διάρκεια των μεταπτυχιακών σπουδών μου. Η συμβολή του στην κατανόηση του υπό διαπραγμάτευση αντικειμένου κρίνεται καθοριστική και ιδιαίτερα σημαντική. Επίσης θα ήθελα να ευχαριστήσω τα μέλη της τριμελούς επιτροπής Αναπληρωτή Καθηγητή κ. Κωνσταντίνο Πολίτη και Επίκουρο Καθηγητή κ. Γεώργιο Τζαβελά για τη συμμετοχή τους στην εξεταστική επιτροπή.

Τέλος θα ήθελα να ευχαριστήσω θερμά την οικογένεια μου για την ηθική και υλική συμπαράσταση καθ'όλη τη διάρκεια των σπουδών μου, καθώς και τους φίλους και τους συμφοιτητές μου για την ηθική συμπαράσταση τους.

Περίληψη

Τα τελευταία χρόνια, η συμβολή της θεωρίας ακραίων τιμών στην πρόβλεψη και πρόληψη ακραίων περιβαλλοντολογικών φαινομένων είναι καίρια, με ότι αυτό συνεπάγεται στην πρόληψη τεράστιων φυσικών και οικονομικών καταστροφών. Κύριοι στόχοι της παρούσας διπλωματικής εργασίας είναι (α) η παρουσίαση στατιστικών μεθόδων που βασίζονται στη θεωρία ακραίων τιμών, οι οποίες είναι κατάλληλες για την στατιστική ανάλυση ακραίων περιβαλλοντολογικών δεδομένων, και (β) η εφαρμογή τους χρησιμοποιώντας πραγματικά δεδομένα από τον Ελλαδικό χώρο

Αρχικά, παρατίθενται τα βασικά πιθανοθεωρητικά αποτελέσματα της θεωρίας ακραίων τιμών καθώς και η μέθοδος στατιστικής ανάλυσης ακραίων παρατηρήσεων που βασίζεται στα μέγιστα κλάσεων των δεδομένων (Block Maxima) για στάσιμες παρατηρήσεις (Κεφάλαια 1, 2).

Στη συνέχεια αναλύονται οι βασικές ιδιαιτερότητες που παρουσιάζουν τα περιβαλλοντολογικά ακραία δεδομένα, όπως η τοπική εξάρτηση των παρατηρήσεων, η περιοδικότητα και η τάση, και στη συνέχεια προτείνονται τρόποι μοντελοποίησης τους στα πλαίσια της θεωρίας ακραίων τιμών, μέσω της μεθόδου Block Maxima (Κεφάλαιο 3).

Ακολούθως, γίνεται εμπειρική αξιολόγηση της μεθόδου Block Maxima που αναπτύχθηκε για τη μοντελοποίηση των περιβαλλοντολογικών δεδομένων, η οποία λαμβάνει υπόψη τυχόν περιοδικότητα και τάση στις παρατηρήσεις (Κεφάλαιο 4).

Στο Κεφάλαιο 5, πραγματοποιείται στατιστική ανάλυση περιβαλλοντολογικών δεδομένων με βάση τη θεωρία και τις μεθόδους που αναπτύχθηκαν στα προηγούμενα κεφάλαια. Συγκεκριμένα, αναλύονται οι μέγιστες βροχοπτώσεις σε 21 μετεωρολογικούς σταθμούς του Ελλαδικού χώρου για μια χρονική περίοδο 50 ετών καθώς και οι μέγιστες θερμοκρασίες σε 18 μετεωρολογικούς σταθμούς για την ίδια χρονική περίοδο. Οι εκτιμήσεις των παραμέτρων των μοντέλων όπως και η εκτίμηση της μηνιαίας στάθμης απόδοσης για 10 και 50 έτη επιτυγχάνεται με τη βοήθεια του στατιστικού πακέτου R

Τέλος παρουσιάζονται ποιοτικά συμπεράσματα για την μέθοδο που χρησιμοποιήθηκε και ποσοτικά συμπεράσματα για τις μέγιστες βροχοπτώσεις και θερμοκρασίες.

Abstract

In the last years, the contribution of extreme value theory in prediction and prevention of extreme environmental phenomena is very crucial, resulting to the prevention of huge natural and economic disasters. The main target of this dissertation is to present statistical methods from extreme value theory, appropriate for the statistical analysis of extreme environmental data.

Initially, we present the basic probabilistic results from extreme value theory and the main statistical method of analyzing extremes (Block Maxima) for stationary observations.

Next, we analyze the special features and properties of environmental extreme data, such as time dependence, seasonality and trend, and show how these characteristics can be modeled in the context of extreme value theory.

Subsequently, we empirically study the performance of MLE's of Block Maxima method (incorporating seasonality and trend) that was developed for the study of environmental extremes.

Two data sets are being used for the application of the statistical models presented. The first set contains extreme precipitations of 21 meteorological stations over Greece for a time period of 50 years, and the second set contains extreme high temperatures of 18 meteorological stations over Greece for the same time period. For all necessary computations and numerical estimations we use the software package R (www.r-project.org/). More specifically we estimate the parameters of the models which were developed in the previous sessions, as well as the monthly return level for 10 and 50 years (and also compute the respective variance of estimates).

Finally, we discuss qualitative conclusions about the method we used, and some general conclusions about extreme precipitation and temperature in Greece.

Περιεχόμενα

Κατάλογος Πινάκων	xvi
Κατάλογος Σχημάτων	xviii
Κατάλογος Συντομογραφιών	xx
1. Εισαγωγή.....	1
1.1 Ιστορική αναδρομή	1
1.2 Η σημασία της θεωρίας ακραίων τιμών στις περιβαλλοντολογικές επιστήμες	3
2.Θεωρία ακραίων τιμών –Στατιστική ανάλυση ακραίων παρατηρήσεων	4
2.1 Εισαγωγή στη θεωρία ακραίων τιμών	4
2.1.1 Το γενικό πρόβλημα.....	4
2.1.2 Πιθανοθεωρητική θεμελίωση της θεωρίας ακραίων τιμών.....	5
2.1.3 Περιοχές έλξης κατανομής ακροτάτων	10
2.2 Στατιστική ανάλυση ακραίων παρατηρήσεων	14
2.2.1 Η μέθοδος Block Maxima	14
2.2.2 Η στάθμη απόδοσης	16
2.2.3 Εκτιμήσεις των παραμέτρων της GEV	18
2.2.4 Έλεγχος καλής προσαρμογής.....	21
3. Θεωρία ακραίων τιμών σε περιβαλλοντολογικά μοντέλα	23
3.1 Εισαγωγή.....	23
3.2.Θεωρία ακραίων τιμών για ακολουθίες εξαρτημένων τ.μ.....	24
3.2.1 Μοντελοποίηση στάσιμων χρονολογικών σειρών.....	29
3.3 Θεωρία ακραίων τιμών σε μη-στάσιμες ακολουθίες τ.μ	30
3.3.1 Εισαγωγή.....	30
3.3.2 Μέθοδοι μοντελοποίησης της περιοδικότητας στις ακραίες παρατηρήσεις.....	31
3.3.3 Μέθοδοι μοντελοποίησης της τάσης στις ακραίες παρατηρήσεις.....	36
3.3.4 Μέθοδοι εκτίμησης παραμέτρων σε μη-στάσιμες ακολουθίες τ.μ	39
3.3.5 Στάθμη απόδοσης σε μη-στάσιμες ακολουθίες τ.μ	40
3.3.6 Επιλογή καταλληλότερου μοντέλου.....	42
3.3.7 Έλεγχος καλής προσαρμογής για μη-στάσιμες ακολουθίες τ.μ	44

4. Εμπειρική αξιολόγηση της μεθόδου Block Maxima σε περιβαλλοντολογικά δεδομένα	46
4.1 Εμπειρική αξιολόγηση της μεθόδου Block Maxima σε περιβαλλοντολογικά δεδομένα.....	46
5. Εφαρμογή της Θεωρίας ακραίων τιμών σε περιβαλλοντολογικά δεδομένα	56
5.1 Περιβαλλοντολογικά δεδομένα και μέθοδος ανάλυσης	56
5.1.1 Ανάλυση μέγιστων βροχοπτώσεων.....	59
5.1.2 Ανάλυση μέγιστων θερμοκρασιών.....	96
6. Συμπεράσματα	124
6.1 Ποσοτικά συμπεράσματα	124
6.2 Ποσοτικά συμπεράσματα	125
6.2.1 Μέγιστες Βροχοπτώσεις.....	125
6.2.2 Μέγιστες θερμοκρασίες.....	127
Παράρτημα	130
Βιβλιογραφία	135

РАНЕЕЗНАМО ТЕРРА

Κατάλογος Πινάκων

- 5.1** Οι 21 μετεωρολογικοί σταθμοί με τις γεωγραφικές συντεταγμένες τους καθώς και το χρονικό εύρος των παρατηρήσεων που εξετάζουμε.
- 5.2** Οι 21 μετεωρολογικοί σταθμοί και το μέγεθος κάθε κλάσης από όπου προέρχονται οι μέγιστες βροχοπτώσεις.
- 5.3** Εκτιμήσεις παραμέτρων $\mu_0, \mu_1, \mu_2, \mu_3$ κατανομής *GEV* για τις μέγιστες βροχοπτώσεις και τυπικά σφάλματα εκτιμητών.
- 5.4** Εκτιμήσεις παραμέτρων $\sigma_0, \sigma_1, \sigma_2, \xi_0$ κατανομής *GEV* για τις μέγιστες βροχοπτώσεις και τυπικά σφάλματα εκτιμητών.
- 5.5** Πίνακας με τις εκτιμήσεις των αναμενόμενων μέγιστων βροχοπτώσεων για κάθε μήνα του έτους και για κάθε μετεωρολογικό σταθμό.
- 5.6** Οι 18 μετεωρολογικοί σταθμοί και το μέγεθος κάθε κλάσης από όπου προέρχονται οι μέγιστες θερμοκρασίες.
- 5.7** Εκτιμήσεις παραμέτρων $\mu_0, \mu_1, \mu_2, \mu_3$ κατανομής *GEV* για τις μέγιστες θερμοκρασίες και τυπικά σφάλματα εκτιμητών.
- 5.8** Εκτιμήσεις παραμέτρων $\sigma_0, \sigma_1, \sigma_2, \xi_0$ κατανομής *GEV* για τις μέγιστες θερμοκρασίες και τυπικά σφάλματα εκτιμητών.
- 5.9** Πίνακας με τις εκτιμήσεις των αναμενόμενων μέγιστων θερμοκρασιών για κάθε μήνα του έτους και για κάθε μετεωρολογικό σταθμό.

РАНЕЕЗНАМО ПЕРПАА

Κατάλογος Σχημάτων

- 4.1 Εκτιμήσεις των παραμέτρων της *GEV* κατανομής για μέγεθος δείγματος $n = 100, 200, 500, 1000$.
- 5.1 Χρονολογική σειρά μηνιαίων μέγιστων βροχοπτώσεων για τους 21 μετεωρολογικούς σταθμούς.
- 5.2 Θηκογράμματα μηνιαίων μέγιστων βροχοπτώσεων για τους 21 μετεωρολογικούς σταθμούς.
- 5.3 Γράφημα της εκτιμώμενης παραμέτρου $\mu(i)$ για τις κατανομές των μέγιστων βροχοπτώσεων, και 95% δ.ε. για $i = 1, 2, \dots, 12$ για καθένα από τους 21 μετεωρολογικούς σταθμούς.
- 5.4 Γράφημα της εκτιμώμενης παραμέτρου $\sigma(i)$ για τις κατανομές των μέγιστων βροχοπτώσεων, και 95% δ.ε. για $i = 1, 2, \dots, 12$ για τους 2 μετεωρολογικούς σταθμούς όπου η περιοδικότητα στην παράμετρο κλίμακος είναι στατιστικά σημαντική.
- 5.5 Γράφημα της εκτιμώμενης μέσης μηνιαίας μέγιστης βροχόπτωσης για $i = 1, 2, \dots, 12$ για καθένα από τους 21 μετεωρολογικούς σταθμούς.
- 5.6 Γράφημα μηνιαίας στάθμης απόδοσης για 10 έτη και 95 % δ.ε. για τις μέγιστες βροχοπτώσεις για τους 21 μετεωρολογικούς σταθμούς.
- 5.7 Γράφημα μηνιαίας στάθμης απόδοσης 50 ετών και 95 % δ.ε. για τις μέγιστες βροχοπτώσεις για τους 21 μετεωρολογικούς σταθμούς.
- 5.8 Q-Q Plots μέγιστων βροχοπτώσεων για τους 21 μετεωρολογικούς σταθμούς.
- 5.9 Χρονολογική σειρά μηνιαίων μέγιστων θερμοκρασιών για τους 18 μετεωρολογικούς σταθμούς.
- 5.10 Θηκογράμματα μηνιαίων μέγιστων θερμοκρασιών για τους 18 μετεωρολογικούς σταθμούς.
- 5.11 Γράφημα της εκτιμώμενης παραμέτρου $\mu(i)$ για τις κατανομές των μέγιστων θερμοκρασιών και 95% δ.ε. για $i = 1, 2, \dots, 12$ για καθένα από τους 18 μετεωρολογικούς σταθμούς.
- 5.12 Γράφημα της εκτιμώμενης παραμέτρου $\sigma(i)$ για $i = 1, 2, \dots, 12$ για τις κατανομές των μέγιστων θερμοκρασιών και 95% δ.ε. για τους 4 μετεωρολογικούς σταθμούς όπου η περιοδικότητα στην παράμετρο κλίμακος είναι στατιστικά σημαντική.
- 5.13 Γράφημα της εκτιμώμενης μέσης μηνιαίας μέγιστης θερμοκρασίας για $i = 1, 2, \dots, 12$ για καθένα από τους 18 μετεωρολογικούς σταθμούς.
- 5.14 Γράφημα μηνιαίας στάθμης απόδοσης για 10 έτη και 95 % δ.ε. για τις μέγιστες θερμοκρασίες για τους 18 μετεωρολογικούς σταθμούς.
- 5.15 Γράφημα μηνιαίας στάθμης απόδοσης 50 ετών και 95 % δ.ε. για τις μέγιστες θερμοκρασίες για τους 18 μετεωρολογικούς σταθμούς.
- 5.16 Q-Q Plots μέγιστων θερμοκρασιών για τους 18 μετεωρολογικούς σταθμούς.
- 6.1 Εκτιμήσεις στάθμης απόδοσης 50 ετών για τις μέγιστες βροχοπτώσεις, για το μήνα Ιανουάριο.
- 6.2 Εκτιμήσεις στάθμης απόδοσης 50 ετών για τις μέγιστες βροχοπτώσεις, για το μήνα Ιούλιο.
- 6.3 Εκτιμήσεις στάθμης απόδοσης 50 ετών για τις μέγιστες θερμοκρασίες, για το μήνα Ιανουάριο.
- 6.4 Εκτιμήσεις στάθμης απόδοσης 50 ετών για τις μέγιστες θερμοκρασίες, για το μήνα Ιούλιο.

РАНЕЕЗНАМО ПЕРПАА

Κατάλογος Συντομογραφιών

- τ.μ.** τυχαία μεταβλητή
- σ.κ.** συνάρτηση κατανομής
- σ.π.π.** συνάρτηση πυκνότητας πιθανότητας
- iid** independent identically distributed
(ανεξάρτητες και ισόνομες τυχαίες μεταβλητές)
- GEV** Generalized Extreme Value Distribution
(Γενικευμένη κατανομή ακροτάτων)
- ε.μ.π.** εκτιμήτρια μέγιστης πιθανοφάνειας
- δ.ε.** διάστημα εμπιστοσύνης

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

1.1 Ιστορική αναδρομή

Η θεωρία ακραίων τιμών τα τελευταία 50 χρόνια, έχει αποτελέσει ένα πολύ βασικό εργαλείο που χρησιμοποιείται σε αρκετούς τομείς των εφαρμοσμένων επιστήμων. Αντικείμενο της θεωρίας ακραίων τιμών αποτελεί η ανάπτυξη και μελέτη στοχαστικών μοντέλων με σκοπό την επίλυση προβλημάτων που σχετίζονται με την εμφάνιση πολύ μεγάλων ή πολύ μικρών τιμών σε τυχαία πειράματα. Το παραπάνω μπορεί να διατυπωθεί πιο φORMALISΤΙΚά ως εξής: έστω X_1, X_2, \dots, X_n ανεξάρτητες και ισόνομες τυχαίες μεταβλητές από κάποια κατανομή με συνάρτηση κατανομής F και το ενδιαφέρον έγκειται στην στοχαστική συμπεριφορά της τυχαίας μεταβλητής $M_n = \max\{X_1, X_2, \dots, X_n\}$ ή της τυχαίας μεταβλητής $M_n' = \min\{X_1, X_2, \dots, X_n\}$.

Ιστορικά, η πρώτη εργασία πάνω σε προβλήματα ακραίων τιμών χρονολογείται γύρω στα 1700, και συγκεκριμένα στο 1709, όπου ο Nicola Bernoulli έθεσε το πρόβλημα της εύρεσης της μέσης μέγιστης απόστασης μεταξύ της αρχής και n σημείων τυχαία διασκορπισμένων πάνω σε ένα ευθύγραμμο τμήμα μήκους t . Μετά το 19^ο αιώνα και την ευρεία χρήση της κανονικής κατανομής το ενδιαφέρον στράφηκε στις ακραίες παρατηρήσεις που προέρχονται από την κανονική κατανομή. Η ανάγκη για μελέτη προβλημάτων ακραίων τιμών άρχισε από αστρονόμους, οι οποίοι ασχολούνταν με την συλλογή επαναλαμβανόμενων παρατηρήσεων του ίδιου αντικείμενου, π.χ. δεδομένα για τη διάμετρο ενός αστεριού. Με βάση τις μετρήσεις αυτές ήθελαν να κατασκευάσουν ένα κριτήριο για την αποδοχή ή την απόρριψη μιας ακραίας παρατήρησης. Ένα τέτοιο κριτήριο έδωσε ο Pierce το 1852. Οι προσπάθειες συνεχιστήκαν για την κατασκευή κριτηρίων και τα επόμενα χρόνια με τις εργασίες των Fuller(1914) και Griffith(1920) χρησιμοποιώντας όμως μεθόδους μαθηματικής ανάλυσης. Η πρώτη εργασία που ασχολείται με την πιθανοθεωρητική μεριά των ακραίων τιμών είναι του L.von Bortkiewitz(1922) ο οποίος μελετά την κατανομή του εύρους τυχαίων μεταβλητών που προέρχονται από κανονική κατανομή. Η σημαντικότητα της εργασίας αυτής έγκειται στο γεγονός ότι για πρώτη φορά στην βιβλιογραφία εμφανίζεται η έννοια της κατανομής της μέγιστης παρατήρησης (largest value's distribution). Στη συνέχεια ο M.

Frechet(1927) πρώτος ερεύνησε την οριακή κατανομή από ακραίες παρατηρήσεις που δεν προέρχονται από την κανονική κατανομή και αναγνώρισε μία πιθανή κατανομή για την μέγιστη παρατήρηση. Την επομένη χρονιά οι R.A. Fischer και L.H.C. Tippett(1928) απέδειξαν το γνωστό θεώρημα που φέρει το όνομα τους, και αναφέρει ότι η οριακή κατανομή της μέγιστης παρατήρησης ενός συνόλου από ανεξάρτητες και ισόνομες τ.μ. (από οποιαδήποτε κατανομή και αν προέρχονται οι παρατηρήσεις), μπορεί να είναι μόνο μία από τρεις συγκεκριμένες κατανομές ακροτάτων. Έπειτα ο R.Von Mises έδωσε κάποιες συνθήκες για την ασθενή σύγκλιση διατεταγμένων τυχαίων μεταβλητών, και το 1943 ο B.Gnedenko θεμελίωσε την θεωρία ακραίων τιμών δίνοντας ικανές και αναγκαίες συνθήκες για την σύγκλιση των ακραίων διατεταγμένων παρατηρήσεων. Τα χρόνια που ακολούθησαν οι εφαρμογές της θεωρίας ακραίων τιμών και η στατιστική ανάλυση ακραίων παρατηρήσεων παρουσίασαν αλματώδη άνοδο. Ενδεικτικά, μερικοί κλάδοι όπου χρησιμοποιούνται τα αποτελέσματα της, ευρέως είναι οι ακόλουθοι:

- στις περιβαλλοντολογικές επιστήμες, όπου μελετάται η συχνότητα εμφάνισης και η ένταση ακραίων φαινομένων, όπως βροχοπτώσεων, θερμοκρασιών, συγκέντρωσης όζοντος, τυφώνων, σεισμικών δονήσεων κ.α..
- στην Αστρονομία, π.χ. για την πρόβλεψη μεγάλων ηλιακών καταιγίδων που επηρεάζουν διάφορες επιστημονικές και επιχειρηματικές δραστηριότητες.
- σε χρηματοοικονομικά και ασφαλιστικά προβλήματα, όπου η θεωρία ακραίων τιμών είναι από τα πιο βασικά εργαλεία για την μέτρηση των χρηματοοικονομικών, και κεφαλαιακών κινδύνων, καθώς και των κινδύνων μεγάλων αποζημιώσεων, που διατρέχουν τα τραπεζικά ιδρύματα και οι ασφαλιστικοί οργανισμοί.
- στην αντοχή υλικών όπου μελετάται η αντοχή υλικών σε έντονες καταπονήσεις και η χρονική περίοδος αντοχής τους στις καταπονήσεις αυτές.
- στις Τηλεπικοινωνίες όπου το ενδιαφέρον βρίσκεται στην συχνότητα και την ένταση υπερφόρτωσης των τηλεπικοινωνιακών δικτύων, και αποτελεί χρήσιμο εργαλείο για τον καλύτερο σχεδιασμό των δικτύων.

Και σε πολλούς άλλους τομείς των επιστήμων όπως, στην θεωρία αξιοπιστίας, στην ανάλυση επιβίωσης, στην βιοτεχνολογία κ.α. κάτι το οποίο δείχνει την σημαντικότητα του συγκεκριμένου κλάδου της στατιστικής.

Για μια πιο ενδελεχή παρουσίαση ιστορικών στοιχείων που αφορούν την εξέλιξη της θεωρίας ακραίων τιμών παραπέμπουμε στους Kotz et al.(2000) και Gumbel (1958).

1.2 Η σημασία της θεωρίας ακραίων τιμών στις περιβαλλοντολογικές επιστήμες

Στους περισσότερους τομείς των περιβαλλοντολογικών επιστήμων, ειδικότερα στην πρόληψη ακραίων περιβαλλοντικών φαινομένων, η θεωρία ακραίων τιμών διαδραματίζει σημαντικό ρόλο. Φαινόμενα όπως σεισμοί, τυφώνες, πλημμύρες, καύσωνες, τσουνάμι, μπορούν να έχουν δυσβάστακτα αποτελέσματα στις περιοχές που πλήττονται με συνέπεια την καταστροφή οικοσυστημάτων, μεγάλες οικονομικές απώλειες, ακόμη και απώλεια ανθρώπινων ζώων. Όλοι οι κίνδυνοι που αναφέρονται παραπάνω μπορεί να προληφθούν ή να μετριαστούν με τη χρήση των αποτελεσμάτων της θεωρίας ακραίων τιμών. Πως όμως μπορεί να χρησιμοποιηθεί η θεωρία ακραίων τιμών για την αποτροπή όλων αυτών των περιβαλλοντικών καταστροφών; Για την χάραξη περιβαλλοντικής πολιτικής αλλά και πολιτικών προστασίας από ακραία φαινόμενα, απαιτείται η ανάπτυξη κανόνων και ο καθορισμός ανώτερων και κατώτερων επιπέδων για μεταβλητές που σχετίζονται με το περιβάλλον, όπως η στάθμη υδάτων, η ταχύτητα του ανέμου κ.α.. Πιο συγκεκριμένα ο καθορισμός ανώτερων και κατώτερων επιπέδων εμπλέκει την επιλογή υψηλών ή χαμηλών ορίων (σταθμών) έτσι ώστε να ορίζεται αν ένα φυσικό γεγονός που ξεπερνά αυτά τα όρια είναι ακραίο ή όχι. Πάνω σε αυτό τον τομέα η θεωρία ακραίων τιμών διαδραματίζει σημαντικότατο ρόλο με την ανάπτυξη στατιστικών τεχνικών για την επιλογή των παραπάνω ορίων, και την συχνότητα εμφάνισης των ακραίων φυσικών φαινομένων. Επομένως τα αποτελέσματα από τη χρήση της θεωρίας ακραίων τιμών χρησιμοποιούνται ως πυλώνες για την κατασκευή αντιπλημμυρικών έργων, ζωνών πυροπροστασίας, ανθεκτικών κτηρίων στους σεισμούς, θαλάσσιων φραγμάτων κ.α. για την αποτροπή των φυσικών καταστροφών.

Επίσης η θεωρία ακραίων τιμών και οι τεχνικές που έχουν αναπτυχθεί χρησιμοποιούνται ευρέως για την ανίχνευση αλλαγών στο παγκόσμιο κλίμα, αλλά και την επίδραση του φαινομένου του θερμοκηπίου στην ατμόσφαιρα της γης, μέσω των αλλαγών που παρατηρούνται στην συχνότητα και την ένταση των ακραίων καιρικών και κλιματικών φαινομένων. Για μια εκτενέστερη παρουσίαση των εφαρμογών της θεωρίας ακραίων τιμών παραπέμπουμε στο βιβλίο των Reiss et al.(2007) καθώς και στις εργασίες των Buishand(1989) και Katz et al.(2002).

ΚΕΦΑΛΑΙΟ 2

Θεωρία ακραίων τιμών

Στατιστική ανάλυση ακραίων παρατηρήσεων

2.1 Εισαγωγή στην θεωρία ακραίων τιμών

2.1.1 Το γενικό πρόβλημα

Στη θεωρία ακραίων τιμών αρχικά επικεντρωνόμαστε στη στατιστική συμπεριφορά της μέγιστης τυχαίας μεταβλητής από μία ακολουθία ανεξάρτητων τ.μ.. Το πρόβλημα που τίθεται είναι το εξής : Έστω X_1, X_2, \dots, X_n ανεξάρτητες και ισόνομες τ.μ. που ακολουθούν κάποια κατανομή με σ.κ. F_θ . Πολλές φορές ζητείται να υπολογιστεί μία τιμή -ένα κατώφλι- c_p το οποίο δεν θα ξεπερνά η μέγιστη τ.μ. με πιθανότητα $1 - p$ ή ισοδύναμα θα το ξεπερνά με πιθανότητα p . Το πρόβλημα μπορεί να τεθεί και ως εξής: (π.χ. βλ. Μπούτσικας (2008)).

Έστω $M_n = \max\{X_1, X_2, \dots, X_n\}$, τότε για την κατανομή της μέγιστης παρατήρησης βασιζόμενοι στην κατανομή F_θ θα έχουμε τα εξής \square

$$\begin{aligned} P(M_n < c_p) = 1 - p &\Leftrightarrow P(\max\{X_1, X_2, \dots, X_n\} < c_p) = 1 - p \\ &\Leftrightarrow P(X_1 < c_p, X_2 < c_p, \dots, X_n < c_p) = 1 - p \\ &\Leftrightarrow P(X_1 < c_p)P(X_2 < c_p) \dots P(X_n < c_p) = 1 - p \\ &\Leftrightarrow F_\theta(c_p)^n = 1 - p \\ &\Leftrightarrow c_p = F_\theta^{-1}((1 - p)^{1/n}) \end{aligned}$$

Επομένως θα πρέπει να εκτιμηθούν με κάποιο τρόπο οι παράμετροι θ της κατανομής F_θ για να μπορεί στη συνέχεια να εκτιμηθεί το κατώφλι c_p που δεν θα υπερβεί η μέγιστη παρατήρηση με πιθανότητα $1 - p$. Η τιμή όμως $(1 - p)^{1/n}$ θα συγκλίνει στο 1 και επομένως η εκτίμηση του c_p θα βασίζεται στην οριακή συμπεριφορά της F_θ , δηλαδή στην μορφή της δεξιάς ουράς της πραγματικής κατανομής των δεδομένων.

Παρατηρείται όμως ότι:

(1) Η επιλογή της F_θ είναι συνήθως αυθαίρετη ή προσεγγιστική και τίποτα δεν εγγυάται ότι η δεξιά ουρά της F_θ παρουσιάζει την ίδια συμπεριφορά με την πραγματική ουρά των δεδομένων.

(2) Οι εκτιμήτριες των παραμέτρων θ πιθανότατα έχουν γίνει με βάση το σύνολο των παρατηρήσεων, επομένως ελάχιστα θα βασίζονται στην συμπεριφορά των ακραίων παρατηρήσεων που υποδηλώνουν τη μορφή της δεξιάς ουράς της πραγματικής κατανομής των δεδομένων.

(3) Ακόμα και να έχει γίνει σωστή επιλογή της F_θ , ελάχιστες αποκλίσεις των εκτιμητών $\hat{\theta}$ από τις πραγματικές τιμές των παραμέτρων θ , μπορούν να οδηγήσουν σε πολύ μεγάλες αποκλίσεις που γίνονται «κοντά» στο δεξί άκρο της F_θ .

(4) Κατά τη μελέτη και πρόβλεψη ακραίων τιμών μίας μεταβλητής, συνήθως έχουμε στη διάθεση μας δεδομένα που αφορούν μόνο μεγάλες ή ακραίες τιμές της μεταβλητής αυτής. Άρα είτε έχουμε μέγιστες τιμές ανά χρονικές υποπεριόδους, είτε περικομμένα δεδομένα με αποτέλεσμα να μην μπορούμε να εκτιμήσουμε την σ.κ. F_θ .

Από τα παραπάνω γίνεται αντιληπτό ότι για να απαντηθούν ερωτήσεις που σχετίζονται με ακραίες παρατηρήσεις θα πρέπει η μελέτη να επικεντρωθεί στη συμπεριφορά της δεξιάς ουράς της κατανομής των παρατηρήσεων. Ειδικότερα, θα πρέπει να μελετηθεί η κατανομή των μέγιστων παρατηρήσεων ή η κατανομή των παρατηρήσεων πάνω από ένα υψηλό κατώφλι. Οι κατανομές αυτές προφανώς εξαρτώνται από την κατανομή F_θ των παρατηρήσεων X_1, X_2, \dots, X_n . Από τη στιγμή που αυτή δεν είναι γνωστή ή δεν μπορεί εύκολα να εκτιμηθεί θα πρέπει η μελέτη να βασιστεί στην οριακή κατανομή των μέγιστων παρατηρήσεων, και την οριακή κατανομή των παρατηρήσεων πάνω από ένα κατώφλι.

Επομένως το ερώτημα που τίθεται είναι, αν υπάρχουν οι οριακές κατανομές των μέγιστων παρατηρήσεων, και αν αυτές εξαρτώνται από την αρχική κατανομή F_θ των παρατηρήσεων. Στην επόμενη ενότητα θα προσπαθήσουμε να απαντήσουμε σε μερικά από αυτά τα ερωτήματα.

2.1.2 Πιθανοθεωρητική θεμελίωση της θεωρίας ακραίων τιμών

Αρχικά θα πρέπει να προσδιοριστεί η μορφή της κατανομής της μέγιστης παρατήρησης. Έστω X_1, X_2, \dots, X_n ανεξάρτητες και ισόνομες τ.μ. που ακολουθούν κάποια κατανομή με σ.κ. F_θ . Συμβολίζουμε με $M_n = \max\{X_1, X_2, \dots, X_n\}$ την τ.μ. που δηλώνει την μέγιστη

παρατήρηση. Στη συνέχεια η προσοχή θα επικεντρωθεί στην μελέτη της μέγιστης παρατήρησης M_n , αφού η μελέτη της ελάχιστης παρατήρησης $M_n' = \min\{X_1, X_2, \dots, X_n\}$ γίνεται όμοια χρησιμοποιώντας τη σχέση \square

$$\min\{X_1, X_2, \dots, X_n\} = -\max\{-X_1, -X_2, \dots, -X_n\}.$$

Η ακριβής σ.κ. της τ.μ. M_n θα είναι \square

$$\begin{aligned} P(M_n \leq x) &= P(\max\{X_1, X_2, \dots, X_n\} \leq x) \\ &= P(X_1 \leq x)P(X_2 \leq x) \dots P(X_n \leq x) \\ &= F(x)F(x) \dots F(x) = (F(x))^n \end{aligned}$$

Όμως η σ.κ. F είναι άγνωστη (δεν έχουμε υποθέσει κάποια κατανομή F για τις αρχικές παρατηρήσεις) και επομένως η κατανομή της τ.μ. M_n , θα γίνει προσπάθεια να προσεγγιστεί για μεγάλο μέγεθος δείγματος n (ασυμπτωτικά) χωρίς να βασιστούμε στην κατανομή των X_i . Στη συνέχεια θα εισάγουμε μία έννοια που θα φανεί αρκετά χρήσιμη για τη θεμελίωση της θεωρίας ακραίων τιμών.

Συμβολίζουμε με x_F το **δεξιό άκρο του στηρίγματος** της κατανομής με σ.κ. F να είναι το supremum του παρακάτω συνόλου \square

$$x_F = \sup\{x \in \mathbb{R}, F(x) < 1\}$$

το οποίο είναι το δεξί άκρο των τιμών που μπορούν να πάρουν τα X_i με θετική πιθανότητα. Το x_F μπορεί να είναι πεπερασμένο ή άπειρο.

Αν $x_F = \infty$ τότε, για κάθε πραγματικό x θα ισχύει ότι \square

$$P(M_n \leq x) \rightarrow 0 \text{ καθώς } n \rightarrow \infty,$$

ενώ αν $x_F < \infty$ τότε έχουμε ότι \square

$$P(M_n \leq x) = F(x)^n \xrightarrow{n \rightarrow \infty} \begin{cases} 0, & x \leq x_F \\ 1, & x > x_F \end{cases}$$

Από τα παραπάνω συνεπάγεται ότι η ακολουθία των τ.μ. M_n συγκλίνει στο δεξί άκρο του στηρίγματος x_F με πιθανότητα 1, ή με άλλα λόγια το δειγματικό μέγιστο M_n συγκλίνει στο πληθυσμιακό μέγιστο x_F . Μία όμως προσεκτικότερη ματιά στην παραπάνω πρόταση ίσως θυμίσει τον νομό των μεγάλων αριθμών (NMA) όπου ο δειγματικός μέσος \bar{X} οποιαδήποτε κατανομής συγκλίνει στον πληθυσμιακό μέσο μ . Η ταχύτητα σύγκλισης, και η οριακή κατανομή που συγκλίνει περιγράφεται από το κεντρικό οριακό θεώρημα (ΚΟΘ). Επομένως θα μπορούσαμε να προσπαθήσουμε να κανονικοποιήσουμε το M_n για να πάρουμε ανάλογα αποτελέσματα, δηλαδή να προσπαθήσουμε να πάρουμε αποτελέσματα της μορφής

$$\frac{M_n - d_n}{c_n} \rightarrow G$$

για κάποιες ακολουθίες πραγματικών αριθμών $d_n, c_n, n = 1, 2, \dots$ και κάποια μη - εκφυλισμένη κατανομή G (κατανομή όπου δεν συγκεντρώνεται όλη η μάζα πιθανότητας σε ένα σημείο). Η παραπάνω παρατήρηση θα αποτελέσει τον ακρογωνιαίο λίθο για το κτίσιμο του πιθανοθεωρητικού μέρους της θεωρίας ακραίων τιμών. Για μία πιο ενδελεχή παρουσίαση των πιθανοθεωρητικών αποτελεσμάτων που ακολουθούν, παραπέμπουμε στα βιβλία των Beirlant et al.(2004), Embrechts et al.(1997), και Finkenstädt et al.(2003) από όπου έχουν παρθεί και τα θεωρήματα-προτάσεις που αναφέρονται παρακάτω.

Συνεχίζοντας τον παραπάνω συλλογισμό για την εύρεση οριακών κατανομών για την μέγιστη παρατήρηση έχουμε τα εξής \square

Για κάποιες ακολουθίες $d_n, c_n, n = 1, 2, \dots$ και $x \in \mathbb{R}$ θα ισχύει ότι \square

$$\begin{aligned} P\left(\frac{M_n - d_n}{c_n} \leq x\right) &= P(M_n \leq c_n x + d_n) = P(M_n \leq u_n(x)) \\ &= P(X_1 \leq u_n(x), X_2 \leq u_n(x), \dots, X_n \leq u_n(x)) \\ &= F(u_n(x))^n = (1 - (1 - F(u_n(x))))^n = (1 - \bar{F}(u_n(x)))^n \\ &= \left(1 - \frac{n\bar{F}(u_n(x))}{n}\right)^n \end{aligned}$$

όπου θέτουμε $u_n(x) = c_n x + d_n$ και συμβολίζουμε την δεξιά ουρά της κατανομής με $\bar{F}(x) = 1 - F(x)$. Χρησιμοποιώντας το αποτέλεσμα ότι \square

$$\left(1 - \frac{a_n}{n}\right)^n \xrightarrow{n \rightarrow \infty} e^{-\lim a_n}$$

καταλήγουμε στην ακόλουθη πρόταση (βλ. π.χ. Μπούτσικας (2008)).

Πρόταση 2.1 \square Για $n \rightarrow \infty$ και $x \in \mathbb{R}$, θα ισχύει η ισοδυναμία \square

$$P\left(\frac{M_n - d_n}{c_n} \leq x\right) \xrightarrow{n \rightarrow \infty} e^{-\lambda(x)} \Leftrightarrow n\bar{F}(u_n(x)) \xrightarrow{n \rightarrow \infty} \lambda(x), \quad \lambda(x) \in [0, \infty]$$

Από την παραπάνω πρόταση παίρνουμε μία πρώτη εικόνα για την μορφή της οριακής κατανομής της κανονικοποιημένης μέγιστης παρατήρησης, ότι θα έχει την μορφή μίας εκθετικής συνάρτησης, κάτι το οποίο είναι δικαιολογημένο λόγω των ιδιοτήτων της εκθετικής συνάρτησης στο άπειρο.

Τίθεται τώρα το ερώτημα, αν η μέγιστη παρατήρηση συγκλίνει πάντοτε σε κάποια κατανομή G , ανεξαρτήτως της κατανομής από την οποία προέρχονται οι αρχικές παρατηρήσεις. Μία απάντηση δίδει το ακόλουθο πόρισμα.

Πόρισμα 2.2 \square Αν $x_F < \infty$ και $P(X = x_F) = F(x_F) - F(x_{F^-}) > 0$ τότε για κάθε ακολουθία $u_n(x) \in R$ για την οποία ισχύει ότι $P(M_n \leq u_n(x)) \xrightarrow{n \rightarrow \infty} p(x)$ θα είναι είτε $p(x) = 0$ είτε $p(x) = 1$.

Με άλλα λόγια αν τα X_i έχουν μία κατανομή η οποία λαμβάνει μία μέγιστη τιμή με θετική πιθανότητα τότε δεν υπάρχει μη-εκφυλισμένη κατανομή G στην οποία συγκλίνει η M_n , οποιαδήποτε κανονικοποίηση και να γίνει.

Αποδεικνύεται ότι για να υπάρχει μία μη-εκφυλισμένη κατανομή G ως όριο της M_n , θα πρέπει τα άλματα της F πλησιάζοντας στο δεξί άκρο του στηρίγματος x_F να φθίνουν με κάποια ταχύτητα. Αυτό περιγράφεται και στο παρακάτω θεώρημα.

Θεώρημα 2.3 \square Έστω F σ.κ. με $x_F \leq \infty$. Θα υπάρξει μία ακολουθία $u_n(x)$ $n = 1, 2, \dots$ τέτοια ώστε $n\bar{F}(u_n(x)) \xrightarrow{n \rightarrow \infty} \lambda(x) \in (0, \infty)$ αν και μόνο αν $\frac{\bar{F}(x)}{\bar{F}(x^-)} \xrightarrow{x \rightarrow x_F} 1$.

Έπειτα από τις συνθήκες που παρατέθηκαν για την ύπαρξη ή μη, οριακής κατανομής ακροτάτων παρουσιάζονται κάποιες έννοιες που θα φανούν αρκετά χρήσιμες στην συνέχεια για την διατύπωση του θεωρήματος των Fischer –Tippet.

Ορισμός 2.4 \square Συμβολίζουμε με MD το σύνολο των μη-εκφυλισμένων κατανομών που μπορούν να προκύψουν ως όριο κάποιου κανονικοποιημένου μεγίστου. Δηλαδή μία κατανομή H ανήκει στην κλάση MD αν και μόνο αν ισχύει ότι \square

$$H \in MD \Leftrightarrow P\left(\frac{M_n - d_n}{c_n} \leq x\right) \xrightarrow{n \rightarrow \infty} H(x), x \in R.$$

για κάποιες ακολουθίες d_n, c_n .

Επίσης ορίζουμε και μία ακόμη κλάση κατανομών με την εξής ιδιότητα: αν $X_i \sim F$ τότε και το κανονικοποιημένο μέγιστο των X_i θα ακολουθεί την ίδια κατανομή F .

Ορισμός 2.5 \square *Max-stable (μέγιστο-ευσταθής) κατανομή καλείται η κατανομή μίας τ.μ. X με την παρακάτω ιδιότητα*

$$M_n \stackrel{=}_d c_n X + d_n \quad n = 1, 2, \dots$$

για κάποιες ακολουθίες d_n, c_n , όπου M_n είναι το μέγιστο από ανεξάρτητα αντίγραφα της X (με $\stackrel{=}_d$ συμβολίζεται η ισότητα κατά κατανομή).

Το επόμενο θεώρημα ταυτίζει τις δύο παραπάνω κλάσεις κατανομών.

Θεώρημα 2.6 \square *Η κλάση των max-stable κατανομών συμπίπτει με την κλάση των μη-εκφυλισμένων κατανομών που μπορούν να προκύψουν ως όριο κανονικοποιημένου μεγίστου, δηλαδή $\text{max-stable} \equiv \text{MD}$.*

Ορισμός 2.7 \square *Θα λέμε ότι δύο τ.μ. X, Y (ή αντίστοιχα οι κατανομές τους) είναι του ίδιου τύπου αν ισχύει ότι $Y \stackrel{=}_d cX + d$, για κάποια $c, d \in \mathbb{R}$.*

Αποδεικνύεται ότι αν ένα κανονικοποιημένο μέγιστο συγκλίνει σε μια κατανομή G και με διαφορετική κανονικοποίηση συγκλίνει σε μία άλλη κατανομή G' , αυτή θα είναι του ίδιου τύπου με την G . Επομένως οι ακολουθίες c_n, d_n δεν είναι μοναδικές, αλλά καταλήγουμε πάντα σε οριακές κατανομές του ίδιου τύπου. Από το προηγούμενο θεώρημα βρήκαμε ένα χαρακτηριστικό των κατανομών που μπορούν να προκύψουν ως όριο ενός κανονικοποιημένου μεγίστου. Είναι οι max-stable κατανομές. Ποιες κατανομές όμως είναι max-stable; Την απάντηση σε αυτό το ερώτημα δίδει το επόμενο θεώρημα των Fisher-Tippett που είναι το βασικότερο θεώρημα της θεωρίας ακραίων τιμών.

Θεώρημα 2.8 (Fisher-Tippett) □ Έστω X_1, X_2, \dots, X_n μία ακολουθία από ανεξάρτητες και ισόνομες (iid) τ.μ.. Αν υπάρχουν ακολουθίες $c_n > 0, d_n \in \mathbb{R}$ και μία μη-εκφυλισμένη κατανομή G ώστε

$$\frac{M_n - d_n}{c_n} = \frac{\max\{X_1, X_2, \dots, X_n\} - d_n}{c_n} \xrightarrow{d} G$$

τότε η G θα είναι του ίδιου τύπου με μία από τις τρεις ακόλουθες κατανομές □

$$\text{Frechet} \quad \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ e^{-x^{-\alpha}}, & x > 0 \end{cases} \quad \alpha > 0$$

$$\text{Weibull} \quad \Psi_\alpha(x) = \begin{cases} e^{-(-x)^\alpha} & x \leq 0 \\ 1 & x > 0 \end{cases} \quad \alpha > 0$$

$$\text{Gumbel} \quad \Lambda(x) = e^{-e^{-x}} \quad x \in \mathbb{R}$$

Το παραπάνω θεώρημα αποδεικνύεται, δείχνοντας ότι οι μόνες μη-εκφυλισμένες λύσεις της συναρτησιακής εξίσωσης που προκύπτει από τον ορισμό των max-stable κατανομών είναι οι τρεις παραπάνω. Οι κατανομές $\Phi_\alpha, \Psi_\alpha, \Lambda$, καλούνται τυπικές κατανομές ακροτάτων, ενώ κατανομές που είναι του ίδιου τύπου με αυτές καλούνται κατανομές ακροτάτων. Οι κατανομές Φ_α, Λ έχουν $x_F = \infty$, ενώ η κατανομή Ψ_α έχει $x_F = 0$.

Επίσης είναι ενδιαφέρον το γεγονός ότι οι τρεις οριακές κατανομές που προβλέπονται από το θεώρημα Fischer-Tippett συνδέονται ως εξής □

$$X \sim \Lambda \Leftrightarrow e^{X/\alpha} \sim \Phi_\alpha \Leftrightarrow -e^{-X/\alpha} \sim \Psi_\alpha.$$

2.1.3 Περιοχές έλξης μιας κατανομής ακροτάτων

Σε αυτή την ενότητα θα προσπαθήσουμε να κάνουμε μία αντιστοίχιση μεταξύ των οριακών κατανομών των κανονικοποιημένων μέγιστων παρατηρήσεων και των αρχικών κατανομών που προέρχονται τα δεδομένα. Επομένως αν το κανονικοποιημένο μέγιστο, των αρχικών παρατηρήσεων X_i , συγκλίνει σε μία οριακή κατανομή αυτή θα είναι μία από τις $\Lambda, \Psi_\alpha, \Phi_\alpha$ ή κατανομή του ίδιου τύπου με αυτές. Τίθεται λοιπόν το ερώτημα σε ποιες αρχικές κατανομές αντιστοιχεί η Λ , σε ποιες η Ψ_α , και σε ποιες η Φ_α ; Για μία αναλυτικότερη πιθανοθεωρητική παρουσίαση της έννοιας της περιοχής έλξης της κατανομής ακροτάτων

παραπέμπουμε στο βιβλίο του De Haan et al.(2006) από όπου έχουν παρθεί και τα θεωρήματα που ακολουθούν.

Αρχικά δίνεται ο ορισμός της περιοχής έλξης μίας κατανομής ακροτάτων, της βασικότερης έννοιας της ενότητας αυτής.

Ορισμός 2.9 \square Μία τ.μ X (ή αντίστοιχα η κατανομή της F) ανήκει στην περιοχή έλξης μίας κατανομής ακροτάτων G αν υπάρχουν ακολουθίες $c_n > 0, d_n \in R$ έτσι ώστε η

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} G.$$

(Συμβολικά X ή $F \in MDA(G)$), όπου ως συνήθως, M_n είναι το μέγιστο από ανεξάρτητα αντίγραφα της X)

Το χαρακτηριστικό μίας κατανομής F που καθορίζει σε ποια περιοχή έλξης ανήκει είναι προφανώς η μορφή της δεξιάς ουράς της, δηλαδή πόσο γρήγορα συγκλίνει στο 0 η $\bar{F}(x)$ όταν $x \rightarrow x_F$. Αυτή η παρατήρηση θα αποτελέσει τη βάση για την κατάταξη των κατανομών στις περιοχές έλξης των οριακών κατανομών. Επομένως, θα πρέπει να εξεταστεί η συμπεριφορά των συναρτήσεων κατανομής στο ∞ , και με την ευκαιρία αυτή θα δοθούν κάποιοι χρήσιμοι ορισμοί.

Ορισμός 2.10 \square Μία θετική (Lebesgue - μετρήσιμη) συνάρτηση h στο $(0, \infty)$ καλείται ομαλής κύμανσης (regularly varying) στο ∞ με δείκτη a (συμβ $h \in \mathfrak{R}_a$) αν

$$\lim_{x \rightarrow \infty} \frac{h(tx)}{h(x)} = t^a, \quad t \geq 0$$

Αν $a = 0$ τότε καλείται βραδείας κύμανσης (slowly varying). Επίσης αν $a = -\infty$ τότε καλείται ταχείας κύμανσης (rapidly varying).

Αρκετά χρήσιμη είναι η ακόλουθη πρόταση, η οποία αναφέρεται στην αναπαράσταση μίας συνάρτησης ομαλής, βραδείας ή ταχείας κύμανσης.

Θεώρημα 2.11 \square Αν $h \in \mathfrak{R}_a, a \in R \cup \{-\infty\}$ τότε η h μπορεί να αναπαρασταθεί ως εξής

$$h(x) = c(x)e^{\int_z^x \frac{\delta(u)}{u} du}, \quad x \geq z$$

για κάποιο $z \geq 0$, και για κάποιες μετρήσιμες συναρτήσεις c, δ για τις οποίες ισχύει ότι $c(x) \rightarrow c_0 \in (0, \infty), \delta(x) \rightarrow a$, όταν $x \rightarrow \infty$.

Είναι φανερό ότι αν $\bar{F} \in \mathfrak{R}_\alpha$ με $\alpha > 0$ τότε $h(x) \rightarrow \infty$, ενώ αν $\bar{F} \in \mathfrak{R}_\alpha$ με $\alpha < 0$ τότε $h(x) \rightarrow 0$. Εδώ το ενδιαφέρον έγκειται στην περίπτωση που $\bar{F} \in \mathfrak{R}_\alpha$ με $\alpha < 0$. Ένα κριτήριο για να δείξουμε αν η δεξιά ουρά μίας κατανομής \bar{F} είναι ομαλής κύμανσης δίνεται στην επόμενη πρόταση.

Πρόταση 2.12 \square Έστω F μία απόλυτα συνεχής σ.κ. με $F(x) < 1$ για $x \geq 0$, και f η αντίστοιχη σ.π.π.. Αν $\lim_{x \rightarrow \infty} \frac{xf(x)}{\bar{F}(x)} = a$ για κάποιο $a > 0$ τότε $\bar{F} \in \mathfrak{R}_{-a}$.

Δύο σ.κ. F, G μπορεί να έχουν διαφορετική μορφή στο R αλλά η δεξιά ουρά τους να έχει την ίδια συμπεριφορά. Θα καλούμε δύο τέτοιες κατανομές ισοδύναμης ουράς (tail equivalent). Αναλυτικότερα θα έχουμε την παρακάτω πρόταση.

Πρόταση 2.13 \square Δύο σ.κ. F, G έχουν ισοδύναμη ουρά αν $x_F = x_G$ και για κάποιο $c > 0$ ισχύει

$$\lim_{x \rightarrow x_F} \frac{\bar{F}(x)}{\bar{G}(x)} = c.$$

Είναι εύκολο να δειχθεί ότι αν δύο σ.κ. έχουν ισοδύναμη ουρά τότε ανήκουν στην ίδια περιοχή έλξης μίας κατανομής ακροτάτων.

Μία έννοια που θα χρησιμοποιήσουμε στην συνέχεια είναι της γενικευμένης αντιστροφής της σ.κ. F . Ειδικότερα έχουμε τον ακόλουθο ορισμό.

Ορισμός 2.14 \square Ως γενικευμένη αντίστροφη μίας σ.κ. F ορίζεται η

$$F^{\leftarrow}(t) = \inf F^{-1}([t,1]) = \inf\{x \in R : F(x) \in [t,1]\}, \quad 0 < t < 1.$$

Ο παραπάνω ορισμός καλύπτει και την περίπτωση που η F δεν είναι αντιστρέψιμη. Στην περίπτωση που η F είναι αντιστρέψιμη τότε η γενικευμένη αντίστροφη συμπίπτει με την κλασική αντίστροφη της σ.κ. F .

Τέλος παραθέτουμε συνοπτικά τις περιοχές έλξης των τριών οριακών κατανομών ακροτάτων και τις απαραίτητες συνθήκες που πρέπει να ικανοποιούν οι σ.κ. F για να ανήκουν στην περιοχή έλξης τους.

Η περιοχή έλξης της Frechet Φ_a

Θεώρημα 2.15 \square Μία σ.κ. $F \in MDA(\Phi_a)$ αν και μόνο αν $\bar{F} \in \mathfrak{R}_{-\alpha}$, $\alpha > 0$. Επίσης αν $F \in MDA(\Phi_a)$ τότε

$$\frac{M_n}{c_n} \xrightarrow{d} \Phi_a$$

όπου $c_n = F^{\leftarrow}(1-n^{-1})$, $d_n = 0$.

Ένα απλούστερο κριτήριο για να ανήκει μία κατανομή στην περιοχή έλξης της Frechet δίνεται στην επομένη πρόταση.

Πόρισμα 2.16 \square Αν F είναι μία απόλυτα συνεχής σ.κ. όπου f η αντίστοιχη σ.π.π. και $\lim_{x \rightarrow \infty} \frac{xf(x)}{\bar{F}(x)} = a$, για κάποιο $a > 0$ τότε $F \in MDA(\Phi_a)$.

Στην $MDA(\Phi_a)$ ανήκουν οι κατανομές *Pareto*, *Cauchy*, *Log-gamma* κ.α..

Η περιοχή έλξης της Weibull Ψ_a

Θεώρημα 2.17 \square Μία σ.κ. $F \in MDA(\Psi_a)$ αν και μόνο αν ισχύει ότι $x_F < \infty$ και $\bar{F}(x_F - x^{-1}) \in \mathfrak{R}_{-\alpha}$, $\alpha > 0$. Επίσης αν $F \in MDA(\Psi_a)$ τότε

$$\frac{M_n - x_F}{c_n} \xrightarrow{d} \Psi_a$$

όπου $c_n = x_F - F^{\leftarrow}(1-n^{-1})$, $d_n = x_F$.

Και ένα απλούστερο κριτήριο για να ανήκει μία κατανομή στην περιοχή έλξης της Weibull είναι το ακόλουθο.

Πόρισμα 2.18 □ Αν F μία απόλυτα συνεχής σ.κ. με σ.π.π. f και

$$\lim_{x \rightarrow x_F} \frac{(x_F - x)f(x)}{\bar{F}(x)} = \alpha > 0$$

τότε η $F \in MDA(\Psi_\alpha)$.

Στην $MDA(\Psi_\alpha)$ ανήκουν η *Ομοιόμορφη* κατανομή, η *Βήτα* κατανομή κ.α..

Η περιοχή έλξης της Gumbel Λ

Θεώρημα 2.19 □ Η σ.κ. F με $x_F \leq \infty$ ανήκει στην $MDA(\Lambda)$ αν και μόνο αν μπορεί να αναπαρασταθεί ως εξής

$$\bar{F}(x) = c(x) e^{\int_z^x \frac{g(t)}{\alpha(t)} dt} \quad z < x < x_F$$

για κάποιο $z < x$ όπου c, g είναι μετρήσιμες συναρτήσεις με $c(x) \rightarrow c_0 > 0$, $g(x) \rightarrow 1$ όταν $x \rightarrow x_F$ και α είναι μία θετική απόλυτα συνεχής συνάρτηση με πυκνότητα α' και $\lim_{x \rightarrow x_F} \alpha'(x) = 0$.

Επίσης αν η $F \in MDA(\Lambda)$ έχοντας την παραπάνω αναπαράσταση τότε

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} \Lambda$$

όπου $d_n = F^{\leftarrow}(1 - n^{-1})$, $c_n = \alpha(d_n)$.

Στην $MDA(\Lambda)$ ανήκουν οι κατανομές *Εκθετική*, *Κανονική*, *Λογαριθμοκανονική*, *Weibull* κ.α.

2.2 Στατιστική ανάλυση ακραίων παρατηρήσεων

2.2.1 Η μέθοδος block-maxima

Στην ενότητα αυτή θα ασχοληθούμε με το στατιστικό μέρος της θεωρίας ακραίων τιμών, που αποτελείται από μοντέλα ερμηνείας, εκτιμήσεων και προβλέψεων ακραίων συμβάντων από πραγματικά δεδομένα. Στις αρχικές εφαρμογές της θεωρίας ακραίων τιμών οι ερευνητές αρχικά επέλεξαν μία από τις τρεις οριακές κατανομές ακροτάτων και την προσαρμόζαν στα δεδομένα. Αυτή η μέθοδος παρουσιάζει μειονεκτήματα διότι δεν υπάρχει κάποια ασφαλής τεχνική επιλογής της οριακής κατανομής ακροτάτων που είναι κατάλληλη για τα δεδομένα $(\Lambda, \Phi_\alpha$ ή $\Psi_\alpha)$, και επομένως η συμπερασματολογία που βασίζεται σε αυτή την επιλογή μπορεί να είναι εντελώς λανθασμένη λόγω της αρχικής επιλογής της οριακής κατανομής.

Λύση σε αυτό το πρόβλημα δίνεται με την χρησιμοποίηση της γενικευμένης κατανομής ακροτάτων (Generalized Extreme Value distribution *GEV*) όπου οι τρεις αρχικές οριακές κατανομές ακροτάτων ενοποιούνται σε μία. Η κατανομή *GEV* περιέχει τρεις παραμέτρους, την παράμετρο θέσης μ , την παράμετρο κλίμακος σ , και την παράμετρο σχήματος ξ . Η αβεβαιότητα που ενδεχομένως να υπάρχει για τον τύπο της οριακής κατανομής ακροτάτων θα εκφράζεται φυσιολογικά μέσα από την αβεβαιότητα εκτίμησης της παραμέτρου σχήματος ξ . Επαναδιατυπώνουμε το θεώρημα Fischer – Tippet μέσω της κατανομής *GEV*.

Θεώρημα 2.20 \square Έστω X_1, X_2, \dots, X_n μία ακολουθία από ανεξάρτητες και ισόνομες (*iid*) τ.μ.. Αν υπάρχουν ακολουθίες $c_n > 0$, $d_n \in \mathbb{R}$ και μία μη-εκφυλισμένη κατανομή G ώστε

$$\frac{M_n - d_n}{c_n} = \frac{\max\{X_1, X_2, \dots, X_n\} - d_n}{c_n} \xrightarrow{d} G$$

τότε η G θα είναι του ίδιου τύπου με την

$$G(z, \mu, \sigma, \xi) = \exp\left[-\left(1 + \xi \left(\frac{z - \mu}{\sigma}\right)\right)^{-1/\xi}\right]$$

και ορίζεται στο σύνολο $\{z : 1 + \xi \frac{z - \mu}{\sigma} > 0\}$.

Για $\xi = 0$ θεωρούμε ως G το όριο της για $\xi \rightarrow 0$ και έχουμε την κατανομή

$$G(z, \mu, \sigma) = \exp\left(-\exp\left(-\frac{z - \mu}{\sigma}\right)\right)$$

Για τις παραμέτρους ισχύει ότι $\mu, \xi \in \mathbb{R}$ και $\sigma > 0$.

Η κατανομή *GEV* για $\xi > 0$ αντιστοιχεί στην κατανομή *Frechet*, για $\xi < 0$ αντιστοιχεί στην κατανομή *Weibull*, ενώ για $\xi = 0$ αντιστοιχεί στην κατανομή *Gumbel*. Φυσικό επακόλουθο είναι ότι και σε αυτή τη περίπτωση δεν χρειάζεται να γνωρίζουμε την αρχική κατανομή από όπου προέρχονται τα δεδομένα.

Ένα παροδικό πρόβλημα που εμφανίζεται κατά την διατύπωση και εφαρμογή του θεωρήματος είναι ότι δεν είναι γνώστες οι ακολουθίες οι οποίες κανονικοποιούν το μέγιστο. Δηλαδή δεν γνωρίζουμε τις ακολουθίες d_n, c_n τέτοιες ώστε να ισχύει ότι

$$P\left(\frac{M_n - d_n}{c_n} \leq z\right) \approx G(z).$$

Από το παραπάνω προκύπτει ότι

$$P(M_n \leq z) \approx G\left(\frac{z-d_n}{c_n}\right) \approx G'(z)$$

και για συγκεκριμένο μέγεθος δείγματος n , η G' είναι και αυτή μέλος της οικογενείας GEV και αφού οι παράμετροι της G' θα εκτιμηθούν, θα εκτιμηθούν μαζί τους και οι τιμές d_n, c_n . Επομένως, αρκεί να θεωρήσουμε ότι η κατανομή της M_n είναι προσεγγιστικά η GEV με κάποιες άγνωστες παραμέτρους μ, σ, ξ που απομένει να εκτιμηθούν.

Αφού εισάγαμε την κατανομή GEV μπορούμε τώρα να αναλύσουμε τη μέθοδο block maxima. Έστω ότι έχουμε X_1, X_2, \dots, X_m ανεξάρτητες παρατηρήσεις. Η μέθοδος λειτουργεί ως εξής : χωρίζουμε τα δεδομένα σε k blocks (κλάσεις) μεγέθους n το καθένα έτσι ώστε $m = nk$. Από κάθε block παίρνουμε την μέγιστη παρατήρηση. Με αυτόν τον τρόπο δημιουργείται η ακολουθία Z_1, Z_2, \dots, Z_k από τα block-maxima (ή μέγιστα κλάσεων) και σε αυτήν την ακολουθία προσαρμόζεται η κατανομή GEV (δηλαδή εκτιμώνται τα μ, σ, ξ). Συνήθως τα blocks επιλέγονται έτσι ώστε να αντιστοιχούν σε παρατηρήσεις μίας χρονικής περιόδου (π.χ. ένα έτος, ένα μήνα ή οποιαδήποτε άλλη χρονική περίοδο). Αφού τα X_1, X_2, \dots, X_m είναι ανεξάρτητα μεταξύ τους και τα Z_1, Z_2, \dots, Z_k θα είναι ανεξάρτητα μεταξύ τους. Η επιλογή του μεγέθους του block είναι μία επιλογή που αμφιταλαντεύεται μεταξύ σφάλματος (bias) και διακύμανσης (variance) των εκτιμητών. Διότι μικρά μεγέθη block οδηγούν σε όχι τόσο καλή προσέγγιση της οριακής κατανομής GEV κάτι που θα οδηγήσει σε σφάλματα στην εκτίμηση και στην στατιστική συμπερασματολογία. Αντίθετα μεγάλα μεγέθη block οδηγούν σε λίγα block-maxima που συνεπάγεται μικρό μέγεθος δείγματος από την GEV κάτι που έχει ως αποτέλεσμα μεγάλη διακύμανση των εκτιμήσεων. Για περισσότερες λεπτομέρειες παραπέμπουμε στο βιβλίο του Coles S. (2001) καθώς και στις σημειώσεις του Μπούτσικα Μ.(2008).

2.2.2 Στάθμη απόδοσης

Μία ποσότητα που είναι πολύ σημαντική κατά την μελέτη ακραίων παρατηρήσεων είναι η στάθμη z_p , την οποία δεν θα υπερβούν τα Z_i (μέγιστα κλάσεων), και κατά συνέπεια τα X_i , με πιθανότητα $1 - p$. Αν συμβολίσουμε με T την τ.μ. που μετρά το πλήθος των χρονικών περιόδων (block) που θα περάσουν έως ότου να υπάρξει παρατήρηση Z_i πάνω από την z_p τότε η τ.μ. $T \sim \text{Γεωμετρική}(p)$. Επομένως $E(T) = 1/p$ και άρα θα εμφανίζονται μέγιστα κλάσεων (ή ισοδύναμα παρατηρήσεις) που θα ξεπερνούν την στάθμη z_p κατά μέσο όρο κάθε $1/p$

χρονικές περιόδους ή ισοδύναμα κάθε μέγιστο κλάσης θα ξεπερνά την στάθμη z_p με πιθανότητα p . Αναλυτικότερα έχουμε τον παρακάτω ορισμό.

Ορισμός 2.21 □ Η ποσότητα z_p την οποία υπερβαίνουν τα μέγιστα κλάσεων κατά μέσο όρο κάθε $1/p$ χρονικές περιόδους καλείται άνω όριο απόδοσης ή στάθμη απόδοσης (return level) για $1/p$ χρονικές περιόδους απόδοσης (return period).

Έχουμε ότι $G(z_p) = P(M_n \leq z_p) = 1 - p$ όπου G η σ.κ. της GEV και για να προσδιορίσουμε τη στάθμη απόδοσης z_p αντιστρέφουμε την σ.κ. G :

$$G(z_p) = 1 - p \Rightarrow G^{-1}(1 - p) = z_p$$

επομένως αντιστρέφοντας παίρνουμε τον ακόλουθο τύπο για την στάθμη απόδοσης:

$$z_p = \mu - \frac{\sigma}{\xi} [1 - (-\ln(1 - p))^{-\xi}] \text{ για } \xi \neq 0$$

$$z_p = \mu - \sigma \ln[-\ln(1 - p)] \text{ για } \xi = 0.$$

Θέτουμε $y_p = -\ln(1 - p)$ και οι παραπάνω εξισώσεις μετατρέπονται στις ακόλουθες □

$$z_p = \mu - \frac{\sigma}{\xi} (1 - y_p^{-\xi}) \text{ για } \xi \neq 0$$

$$z_p = \mu - \sigma \ln y_p \text{ για } \xi = 0$$

Οι παράμετροι μ , σ , ξ είναι οι παράμετροι της κατανομής GEV που ακολουθούν τα μέγιστα των κλάσεων. Μπορούμε να κατασκευάσουμε το γράφημα της στάθμης απόδοσης z_p (return level plot) σε λογαριθμική κλίμακα, δηλαδή να απεικονίζονται τα σημεία $(-\ln y_p, z_p)$, από όπου μπορούμε να αποκομίσουμε αρκετές χρήσιμες πληροφορίες. Έτσι όταν $\xi > 0$ η συνάρτηση z_p , και επομένως και το γράφημα της, είναι μία αύξουσα κυρτή συνάρτηση του $-\ln y_p$ όταν $\xi < 0$ η z_p είναι κοίλη συνάρτηση του $-\ln y_p$ ενώ όταν $\xi = 0$ τότε η z_p είναι γραμμικά αύξουσα συνάρτηση του $-\ln y_p$. Μπορούμε να κατασκευάσουμε το εμπειρικό γράφημα της στάθμης απόδοσης από τις παρατηρήσεις που έχουμε και να πάρουμε μία γραφική εκτίμηση για την παράμετρο σχήματος ξ , και για την καλή προσαρμογή των δεδομένων στην GEV .

2.2.3 Εκτιμήσεις των παραμέτρων της GEV

Πολλές μέθοδοι έχουν προταθεί για την εκτίμηση των παραμέτρων της κατανομής *GEV*.

Κάποιες από αυτές είναι \square

- μέσω γραφημάτων (χρησιμοποιούνται κατάλληλα Probability-Plots)
- μέσω της μεθόδου των ροπών (method of moments)
- μέσω της μεθόδου των σταθμισμένων ροπών (method of probability weighted moments)
- μέσω μεθόδων που βασίζονται σε κατάλληλες συναρτήσεις από διατεταγμένες παρατηρήσεις (method of order statistics)
- μέσω μεθόδων μέγιστης πιθανοφάνειας (MLE)

Κάθε μέθοδος έχει τα πλεονεκτήματα και τα μειονεκτήματα της, όμως η προσαρμογή των εκτιμητών μέγιστης πιθανοφάνειας πάνω σε συνθετότερα μοντέλα της θεωρίας ακραίων τιμών κάνει την μέθοδο αυτή ελκυστικότερη από τις άλλες. Επομένως, στην συνέχεια θα βασιστούμε στην μέθοδο μέγιστης πιθανοφάνειας για την εκτίμηση των παραμέτρων της *GEV*.

Ένα πρόβλημα που ενδέχεται να προκύψει με τη χρήση των ε.μ.π. είναι κατά πόσο η κατανομή *GEV* ικανοποιεί τις συνθήκες ομαλότητας έτσι ώστε να ισχύουν οι ασυμπτωτικές ιδιότητες των εκτιμητών. Πιο συγκεκριμένα οι συνθήκες ομαλότητας δεν ικανοποιούνται σε όλο τον παραμετρικό χώρο διότι το στήριγμα της *GEV* εξαρτάται από τις παραμέτρους της κατανομής. Επομένως δεν μπορούμε αυτόματα να βασιστούμε στις ασυμπτωτικές ιδιότητες των ε.μ.π.. Γενικότερα όμως έχει αποδειχθεί ότι \square

- όταν $\xi > -0.5$ οι ε.μ.π. ικανοποιούν τις συνθήκες ομαλότητας και επομένως έχουν τις γνώστες ασυμπτωτικές ιδιότητες τους.
- όταν $-1 < \xi < -0.5$ οι ε.μ.π. μπορούν να υπολογιστούν άλλα δεν έχουν τις ασυμπτωτικές τους ιδιότητες τους.
- όταν $\xi < -1$ οι ε.μ.π. δεν μπορούν να υπολογιστούν καν.

Η περίπτωση όμως $\xi < -0.5$ αντιστοιχεί σε κατανομές με $x_F < \infty$, δηλαδή με λεπτή δεξιά ουρά και σπάνια εμφανίζονται στις εφαρμογές.

Ας εξετάσουμε τώρα συγκεκριμένα τις ε.μ.π. των μ , σ , ξ της κατανομής *GEV*.

Έστω Z_1, Z_2, \dots, Z_k τα μέγιστα των κλάσεων τα οποία είναι ανεξάρτητα μεταξύ τους και ακολουθούν την κατανομή *GEV*.

Η σ.π.π. της *GEV* (για $\xi \neq 0$) είναι \square

$$g(z, \mu, \sigma, \xi) = \frac{1}{\sigma} \left(1 + \xi \frac{z - \mu}{\sigma}\right)^{-\frac{1}{\xi}} e^{-\left(1 + \xi \frac{z - \mu}{\sigma}\right)^{-\frac{1}{\xi}}} I(z, \mu, \sigma, \xi), \quad z \in \mathbb{R}$$

όπου για την δείκτρια συνάρτηση έχουμε ότι $I(z, \mu, \sigma, \xi) = 1$ ή 0 ανάλογα με το αν η συνθήκη $1 + \xi \frac{z - \mu}{\sigma} > 0$ ισχύει ή όχι.

Ο λογάριθμος της πιθανοφάνειας των Z_1, Z_2, \dots, Z_k θα είναι (για $\xi \neq 0$) \square

$$\begin{aligned} l(\mu, \sigma, \xi) &= \ln \prod_{i=1}^k g(Z_i : \mu, \sigma, \xi) = \sum_{i=1}^k \ln g(Z_i : \mu, \sigma, \xi) \\ &= \sum_{i=1}^k \ln I(Z_i : \mu, \sigma, \xi) - k \ln \sigma - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^k \ln \left(1 + \xi \frac{Z_i - \mu}{\sigma}\right) - \sum_{i=1}^k \left(1 + \xi \frac{Z_i - \mu}{\sigma}\right)^{-\frac{1}{\xi}} \end{aligned}$$

Το άθροισμα των $\ln I(Z_i, \mu, \sigma, \xi)$ θα είναι μηδέν αν όλα τα Z_i ικανοποιούν την συνθήκη $1 + \xi \frac{Z_i - \mu}{\sigma} > 0$ και $-\infty$ διαφορετικά.

Αντίστοιχα η σ.π.π. της *GEV* (για $\xi = 0$) είναι η εξής \square

$$g(z, \mu, \sigma, 0) = \frac{1}{\sigma} e^{-\frac{z - \mu}{\sigma}} e^{-e^{-\frac{z - \mu}{\sigma}}}, \quad z \in \mathbb{R}$$

και ο λογάριθμος της πιθανοφάνειας για $\xi = 0$ είναι \square

$$l(\mu, \sigma, 0) = \ln \prod_{i=1}^k g(Z_i : \mu, \sigma, 0) = \sum_{i=1}^k \ln g(Z_i : \mu, \sigma, 0) = -k \ln \sigma - \sum_{i=1}^k \frac{Z_i - \mu}{\sigma} - \sum_{i=1}^k e^{-\frac{Z_i - \mu}{\sigma}}$$

Οι ε.μ.π. $\hat{\mu}, \hat{\sigma}, \hat{\xi}$, είναι οι τιμές των μ, σ, ξ που μεγιστοποιούν την $l(\mu, \sigma, \xi)$ για $\mu, \xi \in \mathbb{R}$ και $\sigma > 0$. Για την εύρεση των ε.μ.π. από τις παραπάνω εξισώσεις δεν υπάρχουν αναλυτικές εκφράσεις αλλά χρησιμοποιούνται επαναληπτικοί αλγόριθμοι βελτιστοποίησης. Ιδιαίτερη προσοχή απαιτείται κατά την εφαρμογή των αλγορίθμων για την εύρεση των ε.μ.π. έτσι ώστε να μην εισέρχονται οι αλγόριθμοι σε ατέρμονες επαναλήψεις. Αυτό μπορεί να συμβεί α) όταν παραβιάζεται η υπόθεση ότι $1 + \xi(Z_i - \mu)/\sigma > 0, i = 1, 2, \dots, n$ και β) όταν εξετάζονται τιμές του ξ κοντά στο μηδέν διότι ο τύπος της $l(\mu, \sigma, \xi)$ μπορεί να δώσει μεγάλα σφάλματα στρογγύλευσης.

Έχοντας κατά νου τους περιορισμούς στις ασυμπτωτικές ιδιότητες των ε.μ.π. που προκύπτουν από τις τιμές του ξ , οι ε.μ.π. $\hat{\mu}, \hat{\sigma}, \hat{\xi}$, θα ακολουθούν μία πολυδιάστατη κανονική κατανομή με μέση τιμή $(\mu, \sigma, \xi)^T$ και πίνακα διασποράς τον αντίστροφο του πίνακα πληροφορίας (αρκεί να ισχύει ότι $\xi > -0.5$) ο οποίος εκτιμάται από τον \square

$$\hat{I} = - \begin{bmatrix} \frac{\partial^2 l(\hat{\mu}, \hat{\sigma}, \hat{\xi})}{\partial \mu^2} & \frac{\partial^2 l(\hat{\mu}, \hat{\sigma}, \hat{\xi})}{\partial \mu \partial \sigma} & \frac{\partial^2 l(\hat{\mu}, \hat{\sigma}, \hat{\xi})}{\partial \mu \partial \xi} \\ \frac{\partial^2 l(\hat{\mu}, \hat{\sigma}, \hat{\xi})}{\partial \sigma \partial \mu} & \frac{\partial^2 l(\hat{\mu}, \hat{\sigma}, \hat{\xi})}{\partial \sigma^2} & \frac{\partial^2 l(\hat{\mu}, \hat{\sigma}, \hat{\xi})}{\partial \sigma \partial \xi} \\ \frac{\partial^2 l(\hat{\mu}, \hat{\sigma}, \hat{\xi})}{\partial \xi \partial \mu} & \frac{\partial^2 l(\hat{\mu}, \hat{\sigma}, \hat{\xi})}{\partial \xi \partial \sigma} & \frac{\partial^2 l(\hat{\mu}, \hat{\sigma}, \hat{\xi})}{\partial \xi^2} \end{bmatrix},$$

Αν $s_{i,i}$ είναι το i -διαγώνιο στοιχείο του \hat{I}^{-1} , τότε προσεγγιστικά ισχύει ότι \square

$$\hat{\mu} \sim N(\mu, s_{1,1}), \hat{\sigma} \sim N(\sigma, s_{2,2}), \hat{\xi} \sim N(\xi, s_{3,3})$$

Και αρά τα $(\hat{\mu} \pm z_{\alpha/2} \sqrt{s_{1,1}})$, $(\hat{\sigma} \pm z_{\alpha/2} \sqrt{s_{2,2}})$, $(\hat{\xi} \pm z_{\alpha/2} \sqrt{s_{3,3}})$ είναι προσεγγιστικά δ.ε. συντελεστού $1-\alpha$ για τις παραμέτρους μ , σ , ξ . Επίσης για την κατασκευή δ.ε. μπορούμε να χρησιμοποιήσουμε τις profile Deviance Functions $D_p(\mu)$, $D_p(\sigma)$, $D_p(\xi)$. Για παράδειγμα αν θέλουμε ένα δ.ε. για το μ θα ισχύει ότι \square

$$D_p(\mu) = 2(l(\hat{\mu}, \hat{\sigma}, \hat{\xi}) - \max_{\sigma, \xi} l(\mu, \sigma, \xi)) \sim \chi_1^2$$

Η τιμή της $D_p(\mu)$ μπορεί να βρεθεί αριθμητικά για ένα σύνολο δεδομένων τιμών M της μ . Το σύνολο $\{\mu : D_p(\mu) \leq \chi_1^2(\alpha)\}$ είναι ένα δ.ε. συντελεστού $1-\alpha$ για το μ . Για να βρούμε το σύνολο αυτό κατασκευάζουμε το γράφημα της συνάρτησης $D_p(\mu)$ ως προς τις τιμές $\mu \in M$ και προσθέτουμε την οριζόντια γραμμή με τεταγμένη $\chi_1^2(\alpha)$. Οι τετμημένες των σημείων τομής της ευθείας και της καμπύλης προσδιορίζουν τα άκρα του δ.ε. για το μ . Αντίστοιχα ισχύουν και για την κατασκευή δ.ε. για τις άλλες παραμέτρους της κατανομής GEV . Η παραπάνω μέθοδος κατασκευής δ.ε. είναι συνήθως ακριβέστερη της μεθόδου που βασίζεται στην ασυμπτωτική κανονικότητα των εκτιμητών.

Τέλος μπορούμε να εκτιμήσουμε τη στάθμη απόδοσης z_p για $1/p$ χρονικές περιόδους αφού έχουν εκτιμηθεί οι ε.μ.π. των παραμέτρων μ , σ , ξ . Η ποσότητα αυτή είναι μία συγκεκριμένη συνάρτηση των παραμέτρων της GEV κατανομής. Επομένως, αρκεί να αντικαταστήσουμε τις μ , σ , ξ με τις ε.μ.π. τους και έτσι θα έχουμε \square

$$\hat{z}_p(\mu) = h(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} (1 - y_p^{-\hat{\xi}}) & \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma} \ln y_p & \hat{\xi} = 0 \end{cases}$$

όπου $y_p = -\ln(1-p)$.

Και σε αυτή την περίπτωση χρησιμοποιώντας την μέθοδο Δέλτα η ε.μ.π. z_p ακολουθεί προσεγγιστικά κανονική κατανομή με μέση τιμή $h(\mu, \sigma, \xi)$ και διακύμανση \square

$$\text{Var}(\hat{z}_p) = \left(\frac{\partial h}{\partial \mu}, \frac{\partial h}{\partial \sigma}, \frac{\partial h}{\partial \xi} \right) \cdot \hat{I}^{-1} \cdot \left(\frac{\partial h}{\partial \mu}, \frac{\partial h}{\partial \sigma}, \frac{\partial h}{\partial \xi} \right)^T$$

όπου $\left(\frac{\partial h}{\partial \mu}, \frac{\partial h}{\partial \sigma}, \frac{\partial h}{\partial \xi} \right) = (1, -\xi^{-1}(1-y_p^{-\xi}), \sigma\xi^{-2}(1-y_p^{-\xi}) - \sigma\xi^{-1}y_p^{-\xi} \ln y_p)$ επομένως η κατανομή

του ε.μ.π. της στάθμης απόδοσης είναι η $\hat{z}_p \sim N(h(\mu, \sigma, \xi), \text{Var}(\hat{z}_p))$.

Εναλλακτικά μπορούμε να χρησιμοποιήσουμε την profile Deviance Function $D_p(z_p)$ για την κατασκευή δ.ε. για το z_p . Για να γίνει αυτό θα πρέπει να πραγματοποιήσουμε μία νέα παραμετροποίηση έτσι ώστε το z_p να είναι μία από τις παραμέτρους της κατανομής *GEV* (η παραμετροποίηση π.χ. μπορεί να είναι $\mu = z_p + \frac{\sigma}{\xi}(1 - (-\ln(1-p))^{-\xi})$).

Αν $l'(z_p, \sigma, \xi)$ η πιθανοφάνεια των μέγιστων των κλάσεων ως προς την νέα παραμετροποίηση τότε θα ισχύει ότι \square

$$D_p(z_p) = 2(l'(\hat{z}_p, \hat{\mu}, \hat{\xi}) - \max_{\sigma, \xi} l'(z_p, \hat{\mu}, \hat{\xi})) \sim \chi_1^2$$

Συνεπώς το σύνολο $\{z_p : D_p(z_p) \leq \chi_1^2(\alpha)\}$ είναι ένα δ.ε. συντελεστού $1-\alpha$ για την στάθμη απόδοσης. Για περισσότερες λεπτομέρειες για την μέθοδο εκτίμησης των παραμέτρων των κατανομών ακροτάτων παραπέμπουμε στο βιβλίο του Coles(2001).

2.2.4 Έλεγχοι καλής προσαρμογής

Στις προηγούμενες ενότητες δικαιολογήθηκε ότι για $n \rightarrow \infty$ τα μέγιστα των κλάσεων προέρχονται από την κατανομή *GEV*. Στις εφαρμογές όμως το n είναι πεπερασμένο και επομένως θα πρέπει να βεβαιωθούμε ότι τα μέγιστα των κλάσεων προσαρμόζονται ικανοποιητικά σε κάποια κατανομή *GEV* ώστε να έχει νόημα στη συνέχεια η εκτίμηση των παραμέτρων μ, σ, ξ . Αυτό μπορεί να γίνει εύκολα χρησιμοποιώντας P-P Plots ή Q-Q Plots. Μάλιστα εδώ προτιμούμε τα Q-Q Plots έναντι των P-P Plots διότι στα P-P Plots οι ακραίες παρατηρήσεις εμφανίζονται όλες κοντά στο (1,1) και επομένως τυχόν αποκλίνουσες παρατηρήσεις στο δεξιό άκρο της κατανομής που είναι μείζονος σημασίας στην ανάλυση δεν

είναι ευδιάκριτες. Επομένως το κύριο βάρος για την εξαγωγή συμπερασμάτων κατά την ανάλυση ακραίων παρατηρήσεων θα δίνεται στα Q-Q Plots ενώ τα P-P Plots θα αναφέρονται απλώς.

Έστω Z_1, Z_2, \dots, Z_k τα μέγιστα των κλάσεων και $Z_{(1)} \leq Z_{(2)} \leq \dots \leq Z_{(k)}$ τα διατεταγμένα μέγιστα των κλάσεων. Το Q-Q Plot ως προς την $GEV(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ θα αποτελείται από τα σημεία

$$\{Z_{(i)}, \hat{G}^{-1}\left(\frac{i}{k+1}\right) : i=1, 2, \dots, k\}$$

όπου $\hat{G}^{-1}\left(\frac{i}{k+1}\right) = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left(1 - \left(-\ln \frac{i}{k+1}\right)^{-\hat{\xi}}\right)$ είναι η αντίστροφη σ.κ. της $GEV(\hat{\mu}, \hat{\sigma}, \hat{\xi})$, και

το P-P Plot από τα σημεία \square

$$\left\{\frac{i}{k+1}, \hat{G}(Z_{(i)}) : i=1, 2, \dots, k\right\}$$

όπου αναφερόμαστε στην σ.κ. της $GEV(\hat{\mu}, \hat{\sigma}, \hat{\xi})$.

Εάν τα μέγιστα των κλάσεων προέρχονται από μία κατανομή GEV θα πρέπει τα k σημεία του Q-Q Plot να βρίσκονται «κοντά» στην κύρια διαγώνιο του γραφήματος, και πιο συγκεκριμένα «κοντά» στην ευθεία $y = \frac{x - \mu}{\sigma}$. Δηλαδή οι παράμετροι μ, σ μπορούν να εκτιμηθούν πρόχειρα μέσω του γραφήματος, το σ από την κλίση της ευθείας και το μ από το σημείο τομής της ευθείας με τον κάθετο άξονα. Επίσης το γράφημα μπορεί να χρησιμοποιηθεί εμμέσως και για μία πρόχειρη εκτίμηση του ξ . Μπορούμε να δοκιμάσουμε διαφορές τιμές του ξ και να δούμε για ποια τιμή του ξ τα k σημεία ευθυγραμμίζονται. Αυτή είναι προσεγγιστικά η πραγματική τιμή του ξ .

ΚΕΦΑΛΑΙΟ 3

Θεωρία ακραίων τιμών σε περιβαλλοντολογικά μοντέλα

3.1 Εισαγωγή

Στο προηγούμενο κεφάλαιο παρουσιάστηκαν τα βασικά πιθανοθεωρητικά αποτελέσματα της θεωρίας ακραίων τιμών καθώς και η μέθοδος ανάλυσης ακραίων παρατηρήσεων block maxima. Οι βασικές υποθέσεις που γίνονται για την εξαγωγή των θεμελιωδών ασυμπτωτικών αποτελεσμάτων που σχετίζονται με τις κατανομές ακροτάτων είναι ότι οι τ.μ. X_1, X_2, \dots, X_n

□

- α) είναι ανεξάρτητες μεταξύ τους,
- β) προέρχονται από την ίδια κατανομή,
- γ) η κατανομή δεν αλλάζει με τον χρόνο (στάσιμη).

Φυσική συνέπεια αυτών είναι ότι και για τα μέγιστα των κλάσεων, τα οποία για την μέθοδο block maxima υποθέτουμε ότι προέρχονται από την γενικευμένη κατανομή ακροτάτων, θα πρέπει να ισχύουν οι τρεις παραπάνω συνθήκες. Όμως στις εφαρμογές της θεωρίας ακραίων τιμών, οι ακραίες παρατηρήσεις και πιο ειδικά τα ακραία δεδομένα που σχετίζονται με περιβαλλοντολογικά φαινόμενα δεν ικανοποιούν αυτές τις υποθέσεις. Πιο συγκεκριμένα οι ακραίες παρατηρήσεις από περιβαλλοντολογικά φαινόμενα παρουσιάζουν τις παρακάτω ιδιαιτερότητες □

- βραχυχρόνια εξάρτηση (short-term dependence), δηλαδή διαδοχικές τιμές της χρονολογικής σειράς των ακραίων παρατηρήσεων είναι εξαρτημένες μεταξύ τους, ενώ παρατηρήσεις που ο χρονικός ορίζοντας μεταξύ τους είναι μεγάλος είναι ανεξάρτητες.
- μακροχρόνια τάση (long-term trend), αύξουσα ή φθίνουσα, όπου η κατανομή των ακραίων παρατηρήσεων αλλάζει σταδιακά με την πάροδο του χρόνου.
- περιοδικές διακυμάνσεις (seasonal variation), όπου η κατανομή των ακραίων παρατηρήσεων αλλάζει περιοδικά ανάλογα με τον χρονική περίοδο (εποχή, μήνα, έτος).

Όλες αυτές οι ιδιαιτερότητες που παρουσιάζονται στα ακραία περιβαλλοντολογικά φαινόμενα μπορούν να αντιμετωπιστούν με την ταυτόχρονη επέκταση της ασυμπτωτικής θεωρίας και των τεχνικών μοντελοποίησης των ακραίων παρατηρήσεων. Η αντιμετώπιση των ιδιαιτεροτήτων αυτών σε περιβαλλοντολογικά φαινόμενα θα αποτελέσει το περιεχόμενο του κεφαλαίου αυτού. Για αναλυτικότερη παρουσίαση των ιδιαιτεροτήτων που παρουσιάζουν οι

ακραίες παρατηρήσεις από περιβαλλοντολογικά φαινόμενα παραπέμπουμε στο βιβλίο των Reiss et al.(2007).

3.2 Θεωρία ακραίων τιμών για ακολουθίες εξαρτημένων τυχαίων μεταβλητών

3.2.1 Θεωρία ακραίων τιμών για ακολουθίες εξαρτημένων τυχαίων μεταβλητών

Τα στατιστικά μοντέλα της θεωρίας ακραίων τιμών υποθέτουν ότι οι τ.μ. που εμπλέκονται σε αυτά είναι χρονικά ανεξάρτητες. Αυτή η υπόθεση στις περιβαλλοντολογικές εφαρμογές δεν ισχύει. Πιο συγκεκριμένα τα ακραία φαινόμενα τείνουν να εμφανίζονται σε διαδοχικές χρονικές στιγμές, π.χ. μία ημέρα με πολύ υψηλή θερμοκρασία τείνει να ακολουθείται από μία ή περισσότερες ημέρες με παρόμοιες υψηλές θερμοκρασίες. Αντίστοιχες συνθήκες ισχύουν για όλα τα περιβαλλοντολογικά – κλιματολογικά φαινόμενα. Επομένως η έννοια της ακολουθίας των ανεξάρτητων τ.μ. θα πρέπει να γενικευτεί με κάποιο τρόπο έτσι ώστε να περιλαμβάνει και τις παραπάνω ιδιαιτερότητες. Η πιο φυσική γενίκευση μίας ακολουθίας από ανεξάρτητες τ.μ. είναι μία στάσιμη ακολουθία από τ.μ.. Η στασιμότητα η οποία είναι μία πιο ρεαλιστική υπόθεση για πολλές φυσικές διαδικασίες, αντιστοιχεί σε μία ακολουθία όπου οι τ.μ. είναι μεταξύ τους εξαρτημένες αλλά οι στοχαστικές τους ιδιότητες είναι ομοιογενείς κατά τη διάρκεια του χρόνου. Για παράδειγμα αν τα X_1, X_2, \dots, X_n είναι μία στάσιμη ακολουθία από τ.μ. τότε η τ.μ. X_1 θα έχει την ίδια κατανομή με την τ.μ. X_{101} , η κοινή κατανομή των (X_1, X_2) θα είναι ίδια με των (X_{101}, X_{102}) χωρίς όμως να απαιτείται η τ.μ. X_1 να είναι ανεξάρτητη με την X_2 ή την X_{102} .

Η έννοια της εξάρτησης στις στάσιμες ακολουθίες τ.μ. μπορεί να έχει πολλές διαφορετικές μορφές, κάνοντας έτσι αδύνατη την ανάπτυξη μίας γενικής θεωρίας για στάσιμες εξαρτημένες τ.μ., εκτός και αν τεθούν κάποιοι περιορισμοί. Επομένως θα πρέπει να ορίσουμε κάποιες συνθήκες ώστε να γίνεται εμφανής η διαφορά ανάμεσα στις έννοιες της μακροχρόνιας και της βραχυχρόνιας εξάρτησης. Για παράδειγμα αν γνωρίζουμε ότι σήμερα υπήρξε έντονη βροχόπτωση μπορεί να επηρεάσει την πιθανότητα για έντονη βροχόπτωση σε

μία ή δύο ημέρες άλλα όχι την πιθανότητα για έντονες βροχοπτώσεις π.χ. σε τρεις μήνες. Δηλαδή τα ακραία γεγονότα είναι κοντά στο να είναι χρονικά ανεξάρτητα όταν ο χρονικός ορίζοντας πραγματοποίησης τους είναι μεγάλος. Επομένως η θεωρία που θα παρουσιαστεί στην συνέχεια θα βασίζεται στις στάσιμες ακολουθίες τ.μ. και στις έννοιες της βραχυχρόνιας και μακροχρόνιας εξάρτησης. Για μία πιο ενδελεχή πιθανοθεωρητική παρουσίαση των παρακάτω αποτελεσμάτων παραπέμπουμε στην εργασία του Leadbetter(1983).

Αρχικά θα ορίσουμε μία συνθήκη που θα επεξηγεί τις έννοιες της βραχυχρόνιας και μακροχρόνιας εξάρτησης.

Ορισμός 3.1 □ Μία στάσιμη ακολουθία τ.μ. X_1, X_2, \dots, X_n λέμε ότι ικανοποιεί την συνθήκη $D(u_n)$ αν για κάθε $i_1 < \dots < i_p < j_1 < \dots < j_q$ με $j_1 - i_p > l$

$$|P(X_{i_1} \leq u_n, \dots, X_{i_p} \leq u_n, X_{j_1} \leq u_n, \dots, X_{j_q} \leq u_n) - P(X_{i_1} \leq u_n, \dots, X_{i_p} \leq u_n)P(X_{j_1} \leq u_n, \dots, X_{j_q} \leq u_n)| \leq a(n, l)$$

όπου $a(n, l_n) \rightarrow 0$ για κάποια ακολουθία l_n τ.ω. $l_n/n \rightarrow 0$ καθώς το $n \rightarrow \infty$.

Για ακολουθίες ανεξάρτητων τ.μ. η παραπάνω διαφορά είναι ακριβώς μηδέν για κάθε ακολουθία u_n . Πιο γενικά θα απαιτούμε η συνθήκη $D(u_n)$ να ισχύει για μία συγκεκριμένη ακολουθία u_n η οποία είναι αύξουσα καθώς το n μεγαλώνει. Για μία τέτοια ακολουθία η συνθήκη $D(u_n)$ αναφέρει, ότι η παραπάνω διαφορά δεν θα είναι ακριβώς μηδέν, άλλα πολύ κοντά στο μηδέν έτσι ώστε να μην υπάρχει καμία επίδραση στα οριακά θεωρήματα και στα ασυμπτωτικά αποτελέσματα που αφορούν τις ακραίες τ.μ.. Αυτή η ιδιότητα της συνθήκης $D(u_n)$ αναφέρεται και στο επόμενο θεώρημα (Coles(2001)).

Θεώρημα 3.2 □ Έστω X_1, X_2, \dots, X_n μία στάσιμη ακολουθία τ.μ. και $M_n = \max\{X_1, X_2, \dots, X_n\}$. Αν υπάρχουν ακολουθίες $a_n > 0$ και $b_n \in \mathbb{R}$ τ.ω. να ισχύει

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) \xrightarrow{n \rightarrow \infty} G(z)$$

όπου G είναι μία μη-εκφυλισμένη σ.κ., και επιπλέον η συνθήκη $D(u_n)$ ικανοποιείται με $u_n = a_n z + b_n$ για κάθε πραγματικό αριθμό z , τότε η G είναι μέλος της οικογένειας των GEV κατανομών.

Η σημαντικότητα του παραπάνω θεωρήματος έγκειται στο γεγονός ότι για μία ακολουθία τ.μ. με περιορισμένη μακροχρόνια εξάρτηση, στις μέγιστες παρατηρήσεις ισχύουν τα ίδια οριακά θεωρήματα όπως και για ακολουθίες ανεξάρτητων τ.μ.. Οι παράμετροι όμως των οριακών κατανομών επηρεάζονται από την εξάρτηση που περιέχεται στην ακολουθία. Αυτό μπορούμε να το δούμε εξετάζοντας την κατανομή του μέγιστου μίας στάσιμης ακολουθίας τ.μ. και του μέγιστου μιας ακολουθίας από ανεξάρτητες τ.μ. που έχουν την ίδια αρχική κατανομή F . Έστω X_1, X_2, \dots, X_n μία στάσιμη ακολουθία τ.μ. με $X_i \sim F$ $i = 1, 2, \dots, n$ και $X_1^*, X_2^*, \dots, X_n^*$ μία ακολουθία από ανεξάρτητες τ.μ. με $X_i^* \sim F$ $i = 1, 2, \dots, n$. Θα συγκρίνουμε την οριακή κατανομή του $M_n = \max\{X_1, X_2, \dots, X_n\}$ και του $M_n^* = \max\{X_1^*, X_2^*, \dots, X_n^*\}$ καθώς το $n \rightarrow \infty$. Αφού τα X_i, X_i^* προέρχονται από την ίδια κατανομή F , η διαφορά στις οριακές κατανομές των μέγιστων θα οφείλεται στην εξάρτηση που περιέχουν οι τ.μ. X_i . Ο παραπάνω συλλογισμός γίνεται πιο εύκολα αντιληπτός με την παράθεση του παρακάτω παραδείγματος.

Έστω ότι Y_1, Y_2, \dots, Y_n είναι μία ακολουθία από ανεξάρτητες τ.μ. με σ.κ.

$$F_Y(y) = \exp\left(-\frac{1}{(a+1)y}\right), \quad y > 0$$

και για την παράμετρο a της κατανομής θα ισχύει ότι $0 \leq a \leq 1$. Ορίζουμε τη στοχαστική διαδικασία $X_0 = Y_0, X_i = \max\{aY_{i-1}, Y_i\}$ $i = 1, 2, \dots, n$ και έτσι για κάθε $i = 1, 2, \dots, n$ ισχύει ότι

$$P(X_i \leq x) = P(aY_{i-1} \leq x, Y_i \leq x) = P(Y_{i-1} \leq \frac{x}{a})P(Y_i \leq x) = \exp(-1/x)$$

για κάθε $x > 0$.

Επομένως η οριακή κατανομή των τ.μ. X_i είναι *Frechet* και η ακολουθία των τ.μ. είναι στάσιμη διότι οι τ.μ. X_i έχουν την ίδια κατανομή αλλά δεν είναι ανεξάρτητες μεταξύ τους. Ορίζουμε επίσης τις τ.μ. $X_1^*, X_2^*, \dots, X_n^*$ να είναι μία ακολουθία από ανεξάρτητες τ.μ. που

ακολουθούν την κατανομή *Frechet*. Η υπόθεση αυτή γίνεται για να έχουν οι τ.μ. την ίδια αρχική κατανομή. Στη συνέχεια θα μελετήσουμε την οριακή κατανομή των μέγιστων από τις δύο ακολουθίες τ.μ.. Αρχικά για την τ.μ. $M_n^* = \max\{X_1^*, X_2^*, \dots, X_n^*\}$ που αναφέρεται στην ακολουθία από ανεξάρτητες τ.μ. θα ισχύει ότι \square

$$P(M_n^* \leq nz) = [F^*(nz)]^n = [\exp(-1/nz)]^n = \exp(-1/z)$$

όπου F^* η σ.κ. της *Frechet*. Έπειτα, για την ακολουθία των εξαρτημένων τ.μ. έχουμε ότι για την μέγιστη παρατήρηση $M_n = \max\{X_1, X_2, \dots, X_n\}$ ισχύει ότι \square

$$\begin{aligned} P(M_n \leq nz) &= P(X_1 \leq nz, \dots, X_n \leq nz) = P(Y_1 \leq nz, aY_1 \leq nz, \dots, Y_n \leq nz, aY_n \leq nz) \\ &= P(Y_1 \leq nz, \dots, Y_n \leq nz) = [F_Y(nz)]^n = \left[\exp\left(-\frac{1}{(a+1)nz}\right)\right]^n = [\exp(-1/z)]^{\frac{1}{a+1}} \end{aligned}$$

όπου F_Y η σ.κ. των τ.μ. Y_i . Η παραπάνω σχέση ισχύει διότι $a \leq 1$. Από τις παραπάνω σχέσεις καταλήγουμε στην παρακάτω εξίσωση

$$P(M_n^* \leq nz) = (P(M_n \leq nz))^{\frac{1}{a+1}}$$

όπου είναι εμφανής η διαφορά στην οριακή κατανομή των μέγιστων που προέρχονται από ακολουθίες ανεξάρτητων τ.μ. η μία, και από ακολουθίες εξαρτημένων τ.μ. η άλλη.

Το παραπάνω αποτέλεσμα μπορεί να γενικευτεί για μια πιο ευρεία κλάση στάσιμων ακολουθιών τ.μ.. Αυτό αναφέρεται στο επόμενο θεώρημα(Coles(2001)).

Θεώρημα 3.3 \square Έστω X_1, X_2, \dots, X_n μία στάσιμη ακολουθία τ.μ. που προέρχονται από μία κατανομή F και $X_1^*, X_2^*, \dots, X_n^*$ μία ακολουθία από ανεξάρτητες τ.μ. με την ίδια κατανομή F . Ορίζουμε $M_n = \max\{X_1, X_2, \dots, X_n\}$ και $M_n^* = \max\{X_1^*, X_2^*, \dots, X_n^*\}$. Κάτω από κατάλληλες συνθήκες ομαλότητας θα ισχύει ότι \square

$$P\left(\frac{M_n^* - b_n}{a_n} \leq z\right) \xrightarrow{n \rightarrow \infty} G_1(z)$$

(όπου για τις ακολουθίες a_n, b_n έχουμε ότι $a_n > 0$ και $b_n \in \mathbb{R}$ και G_1 είναι μία μη-εκφυλισμένη σ.κ.) αν και μόνο αν ισχύει ότι \square

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) \xrightarrow{n \rightarrow \infty} G_2(z)$$

όπου για τις οριακές κατανομές των μέγιστων ισχύει ότι $G_2(z) = G_1^\theta(z)$ και θ είναι μία σταθερά τ.ω. $0 < \theta \leq 1$.

Από το παραπάνω θεώρημα, η επίδραση της εξάρτησης σε μία στάσιμη ακολουθία τ.μ. οδηγεί στην αντικατάσταση της G_1 με G_1^θ στον τύπο της οριακής κατανομής της μέγιστης τ.μ.. Έτσι αν η G_1 αντιστοιχεί π.χ. στην κατανομή GEV με παραμέτρους (μ, σ, ξ) και $\xi \neq 0$ τότε για τις παραμέτρους της G_1^θ θα ισχύει ότι \square

$$G_1^\theta(z) = \exp\left(-\left(1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right)^{\frac{1}{\xi}}\right)^\theta = \exp\left(-\theta\left(1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right)^{\frac{1}{\xi}}\right) = \exp\left(-\left(1 + \xi\left(\frac{z-\mu^*}{\sigma^*}\right)\right)^{\frac{1}{\xi}}\right)$$

όπου για τις τυποποιημένες παραμέτρους της κατανομής ισχύει ότι \square

$$\mu^* = \mu - \frac{\sigma}{\xi}(1 - \theta^{-\xi}) \quad \text{και} \quad \sigma^* = \sigma\theta^\xi$$

Αν η ασυμπτωτική κατανομή της τ.μ. M_n^* είναι μία $GEV(\mu, \sigma, \xi)$ τότε η ασυμπτωτική κατανομή της τ.μ. M_n είναι $GEV(\mu^*, \sigma^*, \xi)$. Παρατηρούμε ότι ενώ οι παράμετροι θέσης και κλίμακος είναι διαφορετικοί, η παράμετρος σχήματος είναι ίδια και στις δύο κατανομές. Για την απόδειξη του Θεωρήματος 3.3 καθώς και για τις συνθήκες ομαλότητας που πρέπει να ικανοποιούνται παραπέμπουμε στην εργασία του Leadbetter(1983). Η σταθερά θ καλείται **δείκτης ακρότατης εξάρτησης** (extremal index) και παίρνει τιμές μεταξύ μηδέν και ένα. Αποδεικνύεται ότι ο δείκτης ακρότατης εξάρτησης έχει σχέση με το πλήθος των υπερβάσεων πάνω από ένα κατώφλι διαδοχικών τιμών της εξαρτημένης ακολουθίας (συστάδα υπερβάσεων). Στο παράδειγμα που αναφέραμε στην αρχή της ενότητας ισχύει για το δείκτη ακρότατης εξάρτησης ότι $\theta = (\alpha+1)^{-1}$. Πιο συγκεκριμένα, εναλλακτικά ο ορισμός του δείκτη ακρότατης εξάρτησης δίνεται και από την παρακάτω σχέση \square

$$\theta = (\text{οριακό μέσο μέγεθος συστάδας υπερβάσεων})^{-1}$$

Αν π.χ. $\alpha = 1$ τότε $\theta = 0.5$ και το μέσο μέγεθος συστάδας είναι 2 ενώ αν $\alpha = 5$ τότε $\theta = 0.166$ και το μέσο μέγεθος συστάδας είναι 6. Επομένως όσο μειώνεται η τιμή του δείκτη ακρότατης εξάρτησης αυξάνεται το μέγεθος της συστάδας των υπερβάσεων που είναι εξαρτημένες μεταξύ τους. Για ανεξάρτητες ακολουθίες τ.μ. συνεπάγεται ότι $\theta = 1$, το αντίστροφο όμως δεν ισχύει. Για μια πιο λεπτομερή περιγραφή του δείκτη ακρότατης εξάρτησης και τρόπους

εκτίμησης του παραπέμπουμε στην Διπλωματική Εργασία του Λ. Κατσάπα (2010) και στις εκεί παραπομπές

3.2.2 Μοντελοποίηση στάσιμων χρονολογικών σειρών

Γνωρίζοντας ότι η μακροχρόνια εξάρτηση στα ακραία επίπεδα είναι ασθενής, και επομένως οι παρατηρήσεις ικανοποιούν μία κατάλληλη συνθήκη $D(u_n)$, η κατανομή των μέγιστων των κλάσεων από στάσιμες ακολουθίες τ.μ. ανήκει στην ίδια οικογένεια κατανομών, όπως και η κατανομή των μέγιστων των κλάσεων από ακολουθίες ανεξάρτητων τ.μ.. Επομένως μπορούμε να μοντελοποιήσουμε τα μέγιστα των κλάσεων, από περιβαλλοντολογικά φαινόμενα, των στάσιμων χρονολογικών σειρών χρησιμοποιώντας τις μεθόδους που αναφέρθηκαν στο Κεφάλαιο 2. Οι παράμετροι της κατανομής, από όπου προέρχονται τα μέγιστα των κλάσεων από μία στάσιμη ακολουθία τ.μ., είναι διαφορετικοί από ότι αν προέρχονταν από μία ακολουθία ανεξάρτητων τ.μ. άλλα αφού θα εκτιμηθούν αυτό δεν παίζει κάποιο σημαντικό ρόλο. Επομένως αν ο χρονικός ορίζοντας μεταξύ των κλάσεων είναι μεγάλος, (π.χ. οι κλάσεις να είναι τριμηνιαίες, εξαμηνιαίες, ετήσιες και όχι ημερήσιες οι εβδομαδιαίες) έτσι ώστε η εξάρτηση μεταξύ των μέγιστων να είναι ασθενής, η μέθοδος block maxima αποτελεί ακριβέστερη μέθοδο ως προς την στατιστική συμπερασματολογία σε σχέση με άλλες μεθόδους ανάλυσης ακραίων παρατηρήσεων. Διότι οι αντίστοιχες μέθοδοι ανάλυσης ακραίων παρατηρήσεων, όπως η μέθοδος των υπερβάσεων, ή η μέθοδος σημειακών διαδικασιών αντιμετωπίζουν το εξής πρόβλημα \square έχοντας ορίσει ένα κατώφλι, οι υπερβάσεις που ξεπερνούν αυτό το κατώφλι, θα εμφανίζονται, λόγω της ιδιαιτερότητας των περιβαλλοντολογικών φαινομένων, σε διαδοχικές χρονικές στιγμές, σχηματίζοντας συστάδες εξαρτημένων υπερβάσεων κάνοντας με αυτόν τον τρόπο αρκετά δύσκολη την εύρεση και την μοντελοποίηση των συστάδων των υπερβάσεων. Έχουν προταθεί αρκετές μέθοδοι για την εύρεση και την μοντελοποίηση των συστάδων των υπερβάσεων οι περισσότερες όμως με αρκετά μειονεκτήματα. Εν κατακλείδι, η μέθοδος block maxima αποτελεί την ακριβέστερη, από στατιστικής πλευράς μέθοδο εκτίμησης των ακραίων περιβαλλοντολογικών συμβάντων από όσες έχουν προταθεί μέχρι σήμερα.

3.3 Θεωρία ακραίων τιμών σε μη-στάσιμες ακολουθίες τ.μ.

3.3.1 Εισαγωγή

Οι μη-στάσιμες διαδικασίες έχουν χαρακτηριστικά που μεταβάλλονται συστηματικά με την πάροδο του χρόνου. Στα περιβαλλοντολογικά φαινόμενα, η μη-στασιμότητα εμφανίζεται και στην χρονολογική σειρά των αρχικών παρατηρήσεων αλλά και στην χρονολογική σειρά των ακραίων παρατηρήσεων, η οποία είναι υποσύνολο της αρχικής χρονολογικής σειράς. Το φαινόμενο της μη-στασιμότητας, στα περιβαλλοντολογικά φαινόμενα, οφείλεται κυρίως σε δύο παράγοντες □

α) στην περιοδικότητα, δηλαδή στην ύπαρξη εποχικών ή ετήσιων κύκλων, λόγω των διαφορετικών κλιματικών μοτίβων που εμφανίζονται κατά τη διάρκεια του έτους και τα οποία επαναλαμβάνονται περιοδικά,

β) στην παρουσία μακροχρόνιων τάσεων λόγω των πολυσυνθέτων αλλαγών που παρατηρούνται στο κλίμα.

Οι ίδιοι παράγοντες παίζουν καταλυτικό ρόλο και στην μη-στασιμότητα των ακραίων συμβάντων. Όλα αυτά συνεπάγονται την παραβίαση της υπόθεσης της ισονομίας των τ.μ., κάνοντας έτσι αδύνατη την απευθείας χρήση των στατιστικών μοντέλων που αναπτύχθηκαν στο Κεφάλαιο 2. Η πολυπλοκότητα των μη-στάσιμων διαδικασιών καθιστά επίσης ανέφικτη την κατασκευή μίας γενικής θεωρίας για μη-στάσιμες ακραίες τ.μ. σε αντίθεση με ότι συνέβη στην προηγούμενη ενότητα με την έννοια της χρονικής εξάρτησης των ακραίων τ.μ.. Επομένως ο κύριος στόχος της ενότητας αυτής είναι η παράθεση και η ανάλυση μη-τετριμμένων στατιστικών τεχνικών για την μοντελοποίηση των παραγόντων που οδηγούν στην μη-στασιμότητα τα περιβαλλοντολογικά ακραία φαινόμενα, δηλαδή στην μοντελοποίηση της περιοδικότητας και των τάσεων των ακραίων παρατηρήσεων.

3.3.2 Μέθοδοι μοντελοποίησης της περιοδικότητας στις ακραίες παρατηρήσεις

Η περιοδικότητα με την οποία εμφανίζονται τα περιβαλλοντολογικά φαινόμενα αποτελεί βασικό χαρακτηριστικό των διαδικασιών που παράγουν τα φαινόμενα αυτά. Έτσι οι ετήσιοι κύκλοι και οι εποχικές διακυμάνσεις που παρατηρούνται σε πολλά φυσικά φαινόμενα όπως, στις βροχοπτώσεις, στην στάθμη των υδάτων των ποταμών και των λιμνών, στο ύψος των κυμάτων των θαλασσών, στην ένταση των ανέμων, και στις θερμοκρασίες κάνουν επιτακτική την ανάγκη μοντελοποίησης του φαινομένου της περιοδικότητας στις κλασσικές περιβαλλοντολογικές χρονολογικές σειρές. Σε αντιστοιχία με τα παραπάνω, περιοδικότητα τείνουν να έχουν και τα ακραία περιβαλλοντολογικά φαινόμενα χωρίς όμως να υπακούν στους ίδιους νομούς, για την περιοδικότητα, που υπακούει και η αρχική χρονολογική σειρά. Πιο συγκεκριμένα η περιοδικότητα που εμφανίζουν π.χ. οι βροχοπτώσεις κατά τη περίοδο της Άνοιξης κάθε έτους δεν συμπίπτει με την περιοδικότητα που εμφανίζουν οι ακραίες βροχοπτώσεις την ίδια περίοδο.

Η κλασσική μέθοδος ανάλυσης της περιοδικότητας των περιβαλλοντολογικών ακραίων παρατηρήσεων θα περιελάμβανε την προσαρμογή μίας περιοδικής συνάρτησης, π.χ. μίας συνάρτησης ημιτόνου ή συνημιτόνου, στο μέσο της αρχικής χρονολογικής σειράς με τη μέθοδο των ελαχίστων τετραγώνων. Έπειτα θα εφαρμόζονταν οι μέθοδοι μοντελοποίησης της θεωρίας ακραίων τιμών στα κατάλοιπα που θα προέκυπταν από την παλινδρόμηση. Η μέθοδος αυτή όμως έχει το μειονέκτημα ότι υποθέτει ότι η περιοδικότητα που εμφανίζεται στην αρχική χρονολογική σειρά θα εμφανίζεται με την ίδια μορφή και στην χρονολογική σειρά των ακραίων παρατηρήσεων, κάτι που δεν ισχύει απαραίτητα. Επομένως θα πρέπει με κάποιο τρόπο να επεκτείνουμε τις στατιστικές τεχνικές που αναπτύχθηκαν στο Κεφάλαιο 2 έτσι ώστε να συμπεριλαμβάνεται η έννοια της περιοδικότητας. Στην παρούσα Παράγραφο θα αναπτύξουμε τις τεχνικές αυτές για την μέθοδο μοντελοποίησης ακραίων παρατηρήσεων

block maxima. Αντίστοιχες τεχνικές μπορούν να αναπτυχθούν και για τις άλλες μεθόδους στατιστικής ανάλυσης των ακραίων παρατηρήσεων. Δύο είναι οι τεχνικές που χρησιμοποιούνται ευρέως για την μοντελοποίηση της περιοδικότητας. Και στις δύο περιπτώσεις οι παράμετροι της κατανομής *GEV*, από όπου υποθέτουμε ότι προέρχονται τα μέγιστα των κλάσεων, μεταβάλλονται ανάλογα με την χρονική περίοδο που εξετάζουμε την περιοδικότητα π.χ. ανά εποχή, ανά μήνα, ή όποια άλλη χρονική περίοδο θέλουμε να εξετάσουμε τον τρόπο με τον οποίο επιδρά στα ακραία φαινόμενα.

Η πρώτη τεχνική, η οποία παρουσιάζεται στον Coles(2001), αναλύεται ως εξής □

Έστω Z_1, Z_2, \dots, Z_k τα μέγιστα των κλάσεων τα οποία υποθέτουμε ότι προέρχονται από μία *GEV* κατανομή της οποίας οι παράμετροι μεταβάλλονται ανάλογα με την χρονική περίοδο. Επίσης έστω m οι χρονικές περίοδοι των οποίων θέλουμε να μελετήσουμε την επίδραση της περιοδικότητας στην εμφάνιση ακραίων συμβάντων. Σε κάθε μία από τις m χρονικές περιόδους αντιστοιχούν r μέγιστα κλάσεων τ.ω. $m \cdot r = k$. Επομένως σε κάθε μία από τις m χρονικές περιόδους προσαρμόζουμε μία κατανομή *GEV*, και εκτιμούμε τις παραμέτρους του κάθε μοντέλου ξεχωριστά. Έτσι σε κάθε μία χρονική περίοδο τα μέγιστα των κλάσεων τα οποία ανήκουν σε αυτήν είναι ισόνομα. Για παράδειγμα έστω ότι θέλουμε να μελετήσουμε την μηνιαία περιοδικότητα που εμφανίζουν οι ακραίες βροχοπτώσεις σε μία περιοχή για μία περίοδο πενήντα ετών. Τα μέγιστα των κλάσεων θα παρθούν να είναι μηνιαία επομένως θα έχουμε $m = 12$ χρονικές περιόδους (μήνες) όπου σε κάθε χρονική περίοδο θα αντιστοιχούν $r = 50$ μέγιστα κλάσεων, με $12 \cdot 50 = 600$ όπου 600 είναι ο συνολικός αριθμός των μέγιστων για τα πενήντα έτη. Τέλος σε κάθε μία από τις δώδεκα χρονικές περιόδους θα προσαρμόσουμε μία κατανομή *GEV*. Με αυτόν τον τρόπο οι διαφορές που εμφανίζονται σε κάθε χρονική περίοδο θα είναι ορατές μέσω των διαφορών που θα παρουσιάζονται στις εκτιμήσεις των παραμέτρων του μοντέλου και στις εκτιμήσεις των σταθμών απόδοσης για κάθε χρονική περίοδο. Έτσι αν θέλουμε να δώσουμε μία πιο φορμαλιστική μορφή σε αυτό το μοντέλο έχουμε την εξής διατύπωση □

Έστω Z_1, Z_2, \dots, Z_k έτσι ώστε $Z_i \sim GEV(\mu(i), \sigma(i), \xi(i))$ όπου για τις παραμέτρους του μοντέλου ισχύει ότι \square

$$\mu(i) = \mu_{((i-1) \bmod m)+1}, \sigma(i) = \sigma_{((i-1) \bmod m)+1}, \xi(i) = \xi_{((i-1) \bmod m)+1}, i = 1, 2, \dots, k$$

Έτσι π.χ. $\mu(1) = \mu(m+1) = \dots = \mu(rm+1) = \mu_1, \mu(2) = \mu(m+2) = \dots = \mu(rm+2) = \mu_2$ κ.ο.κ.

Το μειονέκτημα της μεθόδου αυτής είναι ότι εισέρχονται πολλές παράμετροι στο μοντέλο $(\mu_1, \dots, \mu_m, \sigma_1, \dots, \sigma_m, \xi_1, \dots, \xi_m)$ κάτι το οποίο δυσκολεύει την ερμηνεία των αποτελεσμάτων που παίρνουμε για την περιοδικότητα.

Η δεύτερη τεχνική μοντελοποίησης της περιοδικότητας, η οποία παρουσιάζεται σε πολλές εργασίες π.χ. βλ. Rust et al.(2009), είναι να μοντελοποιήσουμε τις παραμέτρους της κατανομής ακροτάτων, στην περίπτωση μας της κατανομής GEV , ως περιοδικές συναρτήσεις του χρόνου. Συνεπώς με αυτόν τον τρόπο κάθε ακραία παρατήρηση θα ακόλουθει μία διαφορετική κατανομή ακροτάτων. Δηλαδή στην περίπτωση που αναφερόμαστε στην κατανομή GEV , η παράμετρος θέσης μ και ο λογάριθμος της παραμέτρου κλίμακος σ (έτσι ώστε να διασφαλίζουμε ότι $\sigma > 0$) θα μοντελοποιηθούν ως γραμμικές συναρτήσεις του ημιτόνου και του συνημιτόνου ενώ η παράμετρος σχήματος ξ θα θεωρηθεί σταθερή και όχι χρονικά μεταβαλλόμενη. Ο λόγος που δεν μοντελοποιείται η παράμετρος σχήματος ως περιοδική συνάρτηση του χρόνου είναι ότι υπάρχει δυσκολία στην ακριβή εκτίμηση της, και επομένως η μοντελοποίηση της ως συνάρτησης του χρόνου θα περιέπλεκε ακόμα περισσότερο την διαδικασία εκτίμησης της. Εξάλλου, η παράμετρος ξ εξαρτάται από την οριακή συμπεριφορά της ουράς της κατανομής των αρχικών παρατηρήσεων, η οποία στις περισσότερες εφαρμογές μπορεί να θεωρηθεί ότι δεν μεταβάλλεται, παρότι οι αρχικές παρατηρήσεις μπορεί να παρουσιάζουν μία περιοδικότητα (ως προς την θέση και την κλίμακα).

Έστω Z_1, Z_2, \dots, Z_k τα μέγιστα των κλάσεων τα οποία προέρχονται από την GEV έτσι ώστε $Z_i \sim GEV(\mu(i), \sigma(i), \xi(i))$ και έστω m οι χρονικές περίοδοι που θέλουμε να μελετήσουμε το φαινόμενο της περιοδικότητας και την επίδραση που παρουσιάζει στα ακραία γεγονότα. Οι

χρονικές περίοδοι μπορούν να αντιστοιχούν σε μήνες π.χ. για την μελέτη της περιοδικότητας σε διαφορετικούς μήνες της ίδιας εποχής, σε έτη π.χ. για τον ενδεκαετή κύκλο του Ήλιου, σε εποχές ή σε όποια άλλη χρονική περίοδο θέλουμε να μελετήσουμε την περιοδικότητα της. Επομένως έχουμε ότι τα $Z_i \sim GEV(\mu(i), \sigma(i), \zeta(i))$ όπου οι παράμετροι θα έχουν την έξης μοντελοποίηση \square

$$\mu(i) = \mu_0 + \mu_1 \sin\left(\frac{2\pi i}{m}\right) + \mu_2 \cos\left(\frac{2\pi i}{m}\right)$$

$$\ln \sigma(i) = \sigma_0 + \sigma_1 \sin\left(\frac{2\pi i}{m}\right) + \sigma_2 \cos\left(\frac{2\pi i}{m}\right)$$

$$\zeta(i) = \zeta$$

όπου το i παίρνει διαδοχικά τις τιμές από 1 έως k και m η περίοδος που φαινόμενου που μελετάμε.

Η παραπάνω μοντελοποίηση επιλέχθηκε διότι η συνάρτηση του ημιτόνου είναι περιοδική συνάρτηση και επομένως μέσω αυτής προσπαθούμε να μοντελοποιήσουμε την επαναληπτικότητα. Αν χρησιμοποιούσαμε μόνο την συνάρτηση του ημιτόνου θα έπρεπε να είναι της μορφής $\sin(2\pi i/m + b)$ όπου b είναι το σημείο από όπου ξεκίνα η ημιτονοειδής καμπύλη διότι δεν είναι απαραίτητο ότι ξεκίνα από το μηδέν. Επομένως το b θα έπρεπε να εκτιμηθεί με κάποιο τρόπο κάτι που δεν είναι εύκολο να πραγματοποιηθεί διότι εμφανίζεται μέσα σε ένα ημίτονο. Λύση σε αυτό το πρόβλημα δίνει η εισαγωγή στην εξίσωση του συνημιτόνου όπου υπολογίζει το σημείο από όπου ξεκίνα η καμπύλη χωρίς να δημιουργείται ιδιαίτερο πρόβλημα στην εκτίμηση των παραμέτρων (δεν υπάρχουν παράμετροι προς εκτίμηση μέσα σε τριγωνομετρικές συναρτήσεις).

Για να κάνουμε την παραπάνω μοντελοποίηση πιο κατανοητή θα παραθέσουμε ένα παράδειγμα. Έστω ότι θέλουμε να μελετήσουμε την μηνιαία περιοδικότητα των ακραίων βροχοπτώσεων σε μία περιοχή για μία χρονική περίοδο πενήντα ετών. Για την ανάλυση θα εργαστούμε με μηνιαίες κλάσεις όπου θα πάρουμε το μέγιστο τους. Επομένως θα έχουμε $m = 12$ μέγιστα κλάσεων για κάθε έτος και συνολικά $50 \cdot 12 = 600$ μέγιστα κλάσεων, $k = 600$, κατά την ανάλυση μας. Η περίοδος θα ισούται με $m = 12$ και το i θα παίρνει διαδοχικά τις τιμές

από το 1 έως το 600. Έτσι με αυτόν τον τρόπο περιμένουμε ότι οι ακραίες βροχοπτώσεις θα εμφανίζουν μία μορφή επαναληπτικότητας κάθε 12 χρονικές περιόδους δηλαδή στην περίπτωση μας κάθε 12 μήνες. Επίσης βασικό ρόλο στην μορφή της περιοδικότητας που θα εμφανίζουν οι ακραίες παρατηρήσεις παίζει και η επιλογή των χρονικών περιόδων που παίρνονται τα μέγιστα των κλάσεων στην μέθοδο block maxima.

Από την παραπάνω μοντελοποίηση με τις παραμέτρους της κατανομής *GEV* να είναι περιοδικές συναρτήσεις του χρόνου μπορούν να προκύψουν πολλά εμφωλευμένα μοντέλα, δηλαδή μοντέλα που περιέχονται, στο αρχικό μοντέλο που δημιουργήσαμε. Έχοντας λάβει υπόψη μας και την φύση των δεδομένων μπορούμε να διενεργήσουμε αρκετούς στατιστικούς ελέγχους έτσι ώστε να επιλέξουμε το καταλληλότερο μοντέλο. Μερικοί από τους ελέγχους που μπορούμε να πραγματοποιήσουμε παρατίθενται παρακάτω \square

- i) Ως μηδενική υπόθεση εξετάζεται η υπόθεση ότι $\mu_1 = \mu_2 = \sigma_1 = \sigma_2 = 0$ δηλαδή δεν υπάρχει ένδειξη περιοδικότητας ούτε στην παράμετρο θέσης, ούτε στην παράμετρο κλίμακος, επομένως αν δεχθούμε την μηδενική υπόθεση σημαίνει ότι τα δεδομένα που εξετάζονται δεν παρουσιάζουν κάποιο είδος επαναληπτικότητας.
- ii) Ως μηδενική υπόθεση εξετάζεται η υπόθεση ότι $\mu_1 \neq 0, \mu_2 \neq 0, \sigma_1 = \sigma_2 = 0$, δηλαδή δεν υπάρχει ένδειξη περιοδικότητας στην παράμετρο κλίμακος αλλά υπάρχει στην παράμετρο θέσης. Αυτό σημαίνει ότι αν δεχθούμε την μηδενική υπόθεση τα μέγιστα των κλάσεων τα οποία δεν ακολουθούν την ίδια κατανομή, παρουσιάζουν μία μορφή επαναληπτικότητας στην παράμετρο θέσης, αλλά δεν παρουσιάζουν κάποια μορφή επαναληπτικότητας στην παράμετρο κλίμακος.
- iii) Όπου ως μηδενική υπόθεση εξετάζεται η υπόθεση ότι $\mu_1 \neq 0, \mu_2 \neq 0, \sigma_1 \neq 0, \sigma_2 \neq 0$, δηλαδή υπάρχει ένδειξη περιοδικότητας και στις δύο παραμέτρους της κατανομής κάτι που σημαίνει ότι τα μέγιστα των κλάσεων έχουν μία μορφή επαναληπτικότητας.

Η μορφή των ελέγχων για την επιλογή του καταλληλότερου μοντέλου θα παρουσιαστούν σε επομένη ενότητα.

3.3.3 Μέθοδοι μοντελοποίησης της τάσης στις ακραίες παρατηρήσεις

Η ανίχνευση και η παρακολούθηση τάσεων που εμφανίζονται στα περιβαλλοντολογικά ακραία φαινόμενα, αποτελεί μία από τις κύριες ερευνητικές δραστηριότητες των επιστημόνων που ασχολούνται με το περιβάλλον και το κλίμα, λόγω του ότι μπορεί να αποτελέσει μία πρόμη ένδειξη για μία ευρύτερη κλιματική αλλαγή. Έτσι η έγκυρη ανίχνευση τάσεων θα αποτελέσει σημαντικό παράγοντα πρόληψης από τα ακραία περιβαλλοντολογικά φαινόμενα. Πιο συγκεκριμένα θα μπορούσαμε να αναφέρουμε ότι τα οικοσυστήματα σε μία περίοδο μετασχηματισμού της λειτουργίας τους και των ιδιοτήτων τους, αρχικά αντιδρούν με την εμφάνιση έντονων ακραίων γεγονότων από ότι με φυσιολογικά γεγονότα, κάνοντας έτσι πιθανότερη την ανίχνευση τάσεων, δηλαδή σημαδιών αλλαγής, στα ακραία γεγονότα από ότι στις φυσιολογικές συνθήκες. Μία κλασική τεχνική ανάλυσης της τάσης στα περιβαλλοντολογικά φαινόμενα θα περιελάμβανε την προσαρμογή μίας συνάρτησης, γραμμικής ή άλλης μορφής, τάσης στο μέσο της αρχικής χρονολογικής σειράς. Έπειτα στα κατάλοιπα της παλινδρόμησης θα εφαρμόζονταν οι τεχνικές της θεωρίας ακραίων τιμών για την ανίχνευση τάσης. Η τεχνική αυτή, όπως και στην περίπτωση της περιοδικότητας, έχει το μειονέκτημα ότι υποθέτει ότι η μορφή της τάσης που υπάρχει στην αρχική χρονολογική σειρά θα έχει την ίδια μορφή και στις ακραίες παρατηρήσεις κάτι που δεν ισχύει πάντοτε. Επομένως για την μοντελοποίηση της τάσης στις ακραίες παρατηρήσεις θα ακολουθήσουμε τις τεχνικές που αναπτύχθηκαν στην προηγούμενη ενότητα. Δηλαδή οι παράμετροι των κατανομών ακρότατων, εδώ η κατανομή *GEV*, θα μεταβάλλονται με το χρόνο έτσι ώστε να μοντελοποιηθεί η μη-στασιμότητα της χρονολογικής σειράς των ακραίων

παρατηρήσεων. Όπως και στην περίπτωση της περιοδικότητας οι παράμετροι θέσης μ και κλίμακες σ της κατανομής GEV θα μοντελοποιηθούν ως συναρτήσεις του χρόνου ενώ η παράμετρος σχήματος ξ θα θεωρηθεί σταθερή. Στη συνέχεια θα παρουσιαστούν τα βασικότερα μοντέλα ανάλυσης της τάσης στις ακραίες παρατηρήσεις. Για εκτενέστερη παρουσίαση των τεχνικών ανάλυσης της τάσης στα ακραία περιβαλλοντολογικά φαινόμενα παραπέμπουμε στα βιβλία των Coles(2001), Reiss et al.(2007) καθώς και στην εργασία του Smith(1989).

Έστω Z_1, Z_2, \dots, Z_k τα μέγιστα των κλάσεων και $Z_i \sim GEV(\mu(i), \sigma(i), \xi(i))$.

Τότε οι παράμετροι της κατανομής GEV μπορεί να μοντελοποιηθούν ως γραμμικές συναρτήσεις της τάσης, δηλαδή να έχουν την μορφή

$$\xi(i) = \xi$$

$$\mu(i) = \mu_0 + \mu_1 i$$

$$\ln \sigma(i) = \sigma_0 + \sigma_1 i$$

όπου $i = 1, 2, \dots, k$ είναι ο χρονικός ορίζοντας που έχουν ληφθεί τα μέγιστα των κλάσεων. Έτσι αν έχουμε να μελετήσουμε ως προς την τάση μέγιστες παρατηρήσεις 20 ετών, το i θα παίρνει τιμές από το 1 έως το 20. Εδώ ο συντελεστής μ_1 αναφέρεται στην μεταβολή που εμφανίζεται στα μέγιστα των κλάσεων από την μία χρονική περίοδο έως την άλλη, ως συνέπεια της τάσης που υπάρχει στα ακραία φαινόμενα. Π.χ. αν μελετάμε ένα πρόβλημα που σχετίζεται με ακραίες βροχοπτώσεις, μία τιμή του συντελεστή μ_1 ίση με 0,2 mm ερμηνεύεται ως αύξησης στις ακραίες βροχοπτώσεις κατά 0,2 mm ανά χρονική περίοδο όπου έχουν ληφθεί τα μέγιστα των κλάσεων. Ο συντελεστής μ_1 μπορεί να είναι είτε θετικός είτε αρνητικός.

Επίσης μία διαφορετική μοντελοποίηση των παραμέτρων της κατανομής GEV μπορεί να είναι ως τετραγωνικές (quadratic) συναρτήσεις της τάσης, δηλαδή να είναι της μορφής \square

$$\xi(i) = \xi$$

$$\mu(i) = \mu_0 + \mu_1 i + \mu_2 i^2$$

$$\ln \sigma(i) = \sigma_0 + \sigma_1 i + \sigma_2 i^2$$

όπου $i = 1, 2, \dots, k$ ο χρονικός ορίζοντας που λήφθησαν τα μέγιστα των κλάσεων. Με αυτό το μοντέλο η τάση που έχουν οι ακραίες παρατηρήσεις παρουσιάζει μία πιο γενική μορφή καθώς σε αυτό το είδος της μοντελοποίησης μπορεί να συνυπάρχουν και αυξητικές και φθίνουσες τάσεις.

Επίσης τα ακραία συμβάντα μπορεί να μην έχουν την μορφή μίας αύξουσας ή φθίνουσας τάσης, αλλά να χωρίζονται σε υποπεριόδους, όπου σε κάθε υποπερίοδο οι ακραίες παρατηρήσεις είναι ομοιογενείς μεταξύ τους, αλλά κάθε υποπερίοδος να έχει διαφορετικά χαρακτηριστικά κάνοντας έτσι το συνολικό μοντέλο που προκύπτει να είναι μη-ομοιογενές. Αυτή η συμπεριφορά των ακραίων παρατηρήσεων μπορεί να ερμηνευθεί με ένα μοντέλο αλλαγής καταστάσεων όπου, π.χ. στην περίπτωση που έχουμε δύο υποπεριόδους και το χρονικό σημείο που αλλάζει κατάσταση το μοντέλο είναι το i_0 , τότε ένα μοντέλο που θα περιγράφει την διαδικασία των ακραίων παρατηρήσεων θα είναι το παρακάτω \square

$$\mu(t) = \begin{cases} \mu_1, & i \geq i_0 \\ \mu_0, & i < i_0 \end{cases}$$

$$\sigma(t) = \begin{cases} \sigma_1, & i \geq i_0 \\ \sigma_0, & i < i_0 \end{cases}$$

$$\xi(t) = \begin{cases} \xi_1, & i \geq i_0 \\ \xi_0, & i < i_0 \end{cases}$$

όπου θα πρέπει να ισχύει ότι $\sigma_1 > 0$ και $\sigma_0 > 0$.

Τέλος μπορεί να αναπτυχθούν και συνθετότερα μοντέλα για την μοντελοποίηση της τάσης των ακραίων συμβάντων με μη-παραμετρικές ή ημι-παραμετρικές τεχνικές κάτι που ξεφεύγει από τους σκοπούς της διπλωματικής αυτής εργασίας.

Για τα μοντέλα που περιέχουν την γραμμική τάση και την τετραγωνική τάση στις παραμέτρους της κατανομής GEV μπορούμε να πραγματοποιήσουμε στατιστικούς ελέγχους για την επιλογή του καταλληλότερου μοντέλου από όλα τα εμφωλευμένα (nested) μοντέλα που περιέχονται στις παραπάνω εξισώσεις. Μερικοί από τους ελέγχους που μπορούμε να πραγματοποιήσουμε είναι οι παρακάτω \square

i) Ως μηδενική υπόθεση εξετάζεται η υπόθεση ότι $\mu_1 = \mu_2 = \sigma_1 = \sigma_2 = 0$ δηλαδή δεν υπάρχει ένδειξη τάσης σε καμιά από τις δύο παραμέτρους του μοντέλου επομένως αποδοχή της μηδενικής υπόθεσης σημαίνει ότι δεν υπάρχει ένδειξη μη-στασιμότητας των ακραίων παρατηρήσεων.

ii) Ως μηδενική υπόθεση εξετάζεται η υπόθεση ότι $\mu_2 = \sigma_2 = 0, \mu_1 \neq 0, \sigma_1 \neq 0$ δηλαδή ότι δεν υπάρχει ένδειξη για μοντέλο τετραγωνικής τάσης στα δεδομένα, ενώ υπάρχει ένδειξη μιας μορφής γραμμικής τάσης στα δεδομένα. Επομένως αποδοχή της μηδενικής υπόθεσης σημαίνει ότι υπάρχει ένδειξη για μία μορφή γραμμικής τάσης στις ακραίες παρατηρήσεις.

iii) Ως μηδενική υπόθεση εξετάζεται η υπόθεση ότι $\mu_1 \neq 0, \mu_2 \neq 0, \sigma_1 = \sigma_2 = 0$ δηλαδή δεν υπάρχει ένδειξη τάσης στην παράμετρο κλίμακος ενώ υπάρχει ένδειξη τάσης στην παράμετρο θέσης.

Η μορφή της στατιστικής συνάρτησης για την διεξαγωγή των παραπάνω ελέγχων θα παρουσιαστεί στην ενότητα 3.3.6.

3.3.4 Μέθοδοι εκτίμησης παραμέτρων σε μη-στάσιμες ακολουθίες ακραίων τ.μ.

Το πλεονέκτημα της μεθόδου μέγιστης πιθανοφάνειας έναντι των άλλων μεθόδων εκτίμησης, των παραμέτρων των κατανομών ακροτάτων, είναι η προσαρμοστικότητα σε συνθετότερα μοντέλα από ότι τα μοντέλα που συζητήθηκαν στο κεφάλαιο 2. Επομένως και για τις μη-στάσιμες διαδικασίες οι ε.μ.π. είναι οι μοναδικές που προσαρμόζονται στις ιδιαιτερότητες των μοντέλων αυτών. Την μέθοδο αυτή λοιπόν θα υιοθετήσουμε και θα γίνει μία συνοπτική περιγραφή της παρακάτω.

Έστω Z_1, Z_2, \dots, Z_k τα μέγιστα των κλάσεων για τα οποία ισχύει ότι $Z_i \sim GEV(\mu(i), \sigma(i), \zeta(i))$ όπου οι παράμετροι $\mu(i), \sigma(i), \zeta(i)$ είναι συναρτήσεις του χρόνου ή περιοδικά μεταβαλλόμενες συναρτήσεις. Συμβολίζουμε με β το διάνυσμα που περιέχει όλες τις παραμέτρους του μοντέλου, και η πιθανοφάνεια δίνεται από τον παρακάτω τύπο \square

$$L(\boldsymbol{\beta}) = \prod_{i=1}^k g(z_i, \mu(i), \sigma(i), \xi(i))$$

όπου $g(z_i, \mu(i), \sigma(i), \xi(i))$ είναι η σ.π.π. της κατανομής *GEV* υπολογισμένης στο σημείο z_i . Ανάλογα με το Κεφάλαιο 2, ο λογάριθμος της πιθανοφάνειας, όπου κάθε μία από τις συνιστώσες του διανύσματος της παραμέτρου $\xi(i)$ θα πρέπει να είναι διάφορη του μηδενός, είναι \square

$$l(\mu(i), \sigma(i), \xi(i)) = - \sum_{i=1}^k \left[\ln \sigma(i) + \left(1 + \frac{1}{\xi(i)}\right) \ln \left(1 + \xi(i) \left(\frac{z_i - \mu(i)}{\sigma(i)}\right)\right) + \left(1 + \xi(i) \left(\frac{z_i - \mu(i)}{\sigma(i)}\right)\right)^{\frac{1}{\xi(i)}} \right]$$

και για κάθε ένα από τα σημεία z_i θα πρέπει να ισχύει ότι \square

$$1 + \xi(i) \left(\frac{z_i - \mu(i)}{\sigma(i)}\right) > 0$$

για κάθε $i = 1, 2, \dots, k$. Σε όλες τις εξισώσεις οι παράμετροι $\mu(i)$, $\sigma(i)$, $\xi(i)$ αντικαθιστώνται από τις αναλυτικές τους εκφράσεις. Αντίστοιχα αν κάποιος από τους όρους του διανύσματος της παραμέτρου $\xi(i)$ ισούται με μηδέν αντικαθιστούμε τον τύπο της κατανομής *GEV* με τον οριακό τύπο για $\xi(i) \rightarrow 0$ τον τύπο δηλαδή της κατανομής *Gumbel* και υπολογίζουμε την αντίστοιχη συνάρτηση πιθανοφάνειας. Ούτε και σε αυτή την περίπτωση, όπως και στο Κεφάλαιο 2, όπου οι παράμετροι είναι χρονικά μεταβαλλόμενες, δεν υπάρχουν αναλυτικές εκφράσεις για τους ε.μ.π. και επομένως χρησιμοποιούνται επαναληπτικοί αλγόριθμοι βελτιστοποίησης για την εύρεση τους. Η διασπορά των εκτιμητών προκύπτει από τον αντίστροφο του εκτιμημένου πίνακα πληροφορίας ο οποίος και αυτός υπολογίζεται αριθμητικά. Τα αντίστοιχα προσεγγιστικά δ.ε προκύπτουν με τον ίδιο τρόπο όπως και στο Κεφάλαιο 2.

3.3.5 Στάθμη απόδοσης για μη-στάσιμες ακολουθίες ακραίων τ.μ.

Στην περίπτωση όπου έχουμε μη-στάσιμες ακραίες παρατηρήσεις η στάθμη απόδοσης δεν μπορεί να οριστεί με τον τρόπο που ορίζεται στην στάσιμη περίπτωση. Αυτό συμβαίνει διότι οι ακραίες παρατηρήσεις Z_1, Z_2, \dots, Z_k δεν προέρχονται από την ίδια κατανομή ακροτάτων, αλλά από διαφορετικές κατανομές ακροτάτων που μεταβάλλονται με το χρόνο. Πιο

συγκεκριμένα σε ότι αφορά τα περιβαλλοντολογικά φαινόμενα όπου η μη-στασιμότητα εμφανίζεται με τη μορφή περιοδικότητας ή τάσης, μπορούμε να παράξουμε τροποποιημένες στάθμες απόδοσης σε μη-στάσιμες παρατηρήσεις που εμφανίζουν περιοδικότητα, δεν έχει όμως νόημα να παράξουμε στάθμη απόδοσης σε ακραίες παρατηρήσεις που εμφανίζουν τάση. Στην περίπτωση της περιοδικότητας η μη-στασιμότητα των ακραίων παρατηρήσεων μπορεί να «παρακαμφθεί» παράγοντας στάθμες απόδοσης για κάθε μία από τις m χρονικές περιόδους που μελετάμε την περιοδικότητα αφού οι παρατηρήσεις που ανήκουν σε κάθε μία από τις m χρονικές περιόδους θεωρούνται στάσιμες (π.χ. $m = 12$ μήνες). Αντίθετα στην περίπτωση ύπαρξης τάσης δεν μπορεί να «παρακαμφθεί» με κάποιο τρόπο η μη-στασιμότητα επομένως δεν έχει κάποιο νόημα η ύπαρξη στάθμης απόδοσης ως το κατώφλι που υπερβαίνουν τα μέγιστα κατά μέσο όρο κάθε $1/p$ περιόδους σε μη-στάσιμες περιπτώσεις. Έστω $Z_i \sim GEV(\mu(i), \sigma(i), \xi(i))$ $i = 1, \dots, k$ όπου οι παράμετροι $\mu(i), \sigma(i), \xi(i)$ είναι περιοδικά μεταβαλλόμενες συναρτήσεις του χρόνου. Αφού έχουν υπολογιστεί οι ε.μ.π. των παραμέτρων $\mu(i), \sigma(i), \xi(i)$, σύμφωνα με την ενότητα 3.3.4, μπορούμε να εκτιμήσουμε την στάθμη απόδοσης για κάθε μία από τις m χρονικές περιόδους. Η ποσότητα αυτή είναι μια συνάρτηση των παραμέτρων της $GEV(\mu(i), \sigma(i), \xi(i))$ επομένως αρκεί να αντικαταστήσουμε τις παραμέτρους με τις ε.μ.π. τους και έτσι θα έχουμε \square

$$\hat{z}_p(i) = h(\hat{\mu}(i), \hat{\sigma}(i), \hat{\xi}(i)) = \begin{cases} \hat{\mu}(i) - \frac{\hat{\sigma}(i)}{\hat{\xi}(i)} (1 - y_p^{-\hat{\xi}(i)}), & \hat{\xi}(i) \neq 0 \\ \hat{\mu}(i) - \hat{\sigma}(i) \ln y_p, & \hat{\xi}(i) = 0 \end{cases}$$

όπου $y_p = -\ln(1-p)$. Σημειώνεται ότι εδώ η στάθμη απόδοσης εξαρτάται και από το $i \in \{1, 2, \dots, m\}$. Αν π.χ. η περίοδος m αποτελείται από 12 μηνιαίες μέγιστες παρατηρήσεις τότε η στάθμη απόδοσης εξαρτάται από τον μήνα $i = 1, 2, \dots, 12$ και φυσικά το p (πιθανότητα υπέρβασης). Έτσι το $z_p(i)$ είναι το κατώφλι που ξεπερνούν οι παρατηρήσεις του μήνα i με πιθανότητα p .

Και εδώ, όπως και στην στάσιμη περίπτωση χρησιμοποιώντας την μέθοδο Δέλτα η ε.μ.π. $z_p(i)$ ακολουθεί προσεγγιστικά κανονική κατανομή με μέση τιμή $h(\mu(i), \sigma(i), \xi(i))$ και διακύμανση \square

$$\text{Var}(\hat{z}_p(i)) = \left(\frac{\partial h}{\partial \mu(i)}, \frac{\partial h}{\partial \sigma(i)}, \frac{\partial h}{\partial \xi(i)} \right)^T \cdot \hat{I}^{-1} \cdot \left(\frac{\partial h}{\partial \mu(i)}, \frac{\partial h}{\partial \sigma(i)}, \frac{\partial h}{\partial \xi(i)} \right)$$

όπου \hat{I} ο εκτιμημένος πίνακας πληροφορίας που ορίστηκε στην ενότητα 2.2.3. Επομένως μέσω της μεθόδου Δέλτα μπορούμε να κατασκευάσουμε δ.ε., χρησιμοποιώντας την ασυμπτωτική κανονικότητα των εκτιμητών για τη στάθμη απόδοσης όπως και στην ενότητα 2.2.3.

Για να γίνουμε πιο σαφείς ως προς την εκτίμηση των m διαφορετικών σταθμών απόδοσης στην περίπτωση της περιοδικότητας θα εξετάσουμε το ακόλουθο παράδειγμα \square Έστω ότι έχουμε ακραίες παρατηρήσεις $Z_i \sim GEV(\mu(i), \sigma(i), \xi(i))$ $i = 1, 2, \dots, k$ και έχει επιλεγεί η εξής μοντελοποίηση για τις παραμέτρους της κατανομής \square

$$\mu(i) = \mu_0 + \mu_1 \sin\left(\frac{2\pi i}{m}\right) + \mu_2 \cos\left(\frac{2\pi i}{m}\right)$$

$$\ln \sigma(i) = \sigma_0 + \sigma_1 \sin\left(\frac{2\pi i}{m}\right) + \sigma_2 \cos\left(\frac{2\pi i}{m}\right)$$

$$\xi(i) = \xi_0$$

όπου m είναι η περίοδος του φαινομένου που μελετάμε. Σε αυτή τη περίπτωση μπορούμε να εκτιμήσουμε m διαφορετικές στάθμες απόδοσης, μία για κάθε χρονική περίοδο που μελετάμε την περιοδικότητα αφού για κάθε χρονική περίοδο θα αντιστοιχούν διαφορετικές παράμετροι $\mu(i)$, $\sigma(i)$ και έτσι αντικαθιστώντας στην παραπάνω εξίσωση θα έχουμε την σημειακή εκτίμηση για τη στάθμη απόδοσης. Έτσι στην περίπτωση που μελετάμε την ετήσια εποχικότητα κάποιου φαινομένου μπορούμε να έχουμε 12 διαφορετικές στάθμες απόδοσης όπου κάθε μία θα αντιστοιχεί σε ένα μήνα του έτους. Με τη μέθοδο Δέλτα μπορούμε να κατασκευάσουμε δ.ε. χρησιμοποιώντας την ασυμπτωτική κανονικότητα των εκτιμητών. Για εκτενέστερη παρουσίαση παραπέμπουμε στις εργασίες των Carter et al.(1981) και Rust et al.(2009).

3.3.6 Επιλογή καταλληλότερου μοντέλου

Για την μοντελοποίηση των μη-στάσιμων ακραίων παρατηρήσεων που προέρχονται από περιβαλλοντολογικά φαινόμενα έχει υιοθετηθεί, λόγω ευκολίας στην χρήση της, η μέθοδος

όπου οι παράμετροι των κατανομών ακροτάτων είναι είτε συναρτήσεις του χρόνου είτε περιοδικές συναρτήσεις. Η μοντελοποίηση αυτή παράγει πολλά διαφορετικά μοντέλα τα περισσότερα από τα οποία είναι εμφωλευμένα (nested) μεταξύ τους κάνοντας έτσι την επιλογή του κατάλληλου μοντέλου ένα σημαντικό θέμα για την στατιστική ανάλυση των ακραίων παρατηρήσεων και την εξαγωγή των ζητούμενων αποτελεσμάτων. Βασική υπόθεση για την κατασκευή στατιστικών ελέγχων για την επιλογή του καταλληλότερου μοντέλου είναι τα μοντέλα να είναι εμφωλευμένα μεταξύ τους. Ένα παράδειγμα για την επεξήγηση της έννοιας του εμφωλευμένου μοντέλου είναι το εξής \square αν οι παράμετροι της κατανομής *GEV* έχουν μοντελοποιηθεί ως συναρτήσεις του χρόνου, τότε το μοντέλο όπου οι παράμετροι είναι της μορφής $\mu(i) = \mu_0 + \mu_1 i$, $\ln \sigma(i) = \sigma_0 + \sigma_1 i$, είναι υποσύνολο του μοντέλου, όπου οι παράμετροι είναι συναρτήσεις της μορφής $\mu(i) = \mu_0 + \mu_1 i + \mu_2 i^2$, $\ln \sigma(i) = \sigma_0 + \sigma_1 i + \sigma_2 i^2$.

Η στατιστική συνάρτηση (deviance statistic) του έλεγχου για δύο μοντέλα τέτοια ώστε $M_0 \subset M_1$ δηλαδή το ένα να είναι υποσύνολο του άλλου θα είναι της μορφής \square

$$D = 2 \ln \frac{L(M_1)}{L(M_0)}$$

οπού $L(M_1)$, $L(M_0)$ είναι οι μεγιστοποιημένες πιθανοφάνειες των δύο μοντέλων, δηλαδή έχοντας αντικαταστήσει τις εκτιμήτριες μέγιστης πιθανοφάνειας στην θέση των παραμέτρων. Και η στατιστική συνάρτηση καταλήγει στην απλούστερη μορφή

$$D = 2(l_1(M_1) - l_0(M_0))$$

όπου $l_1(M_1)$, $l_0(M_0)$ είναι οι μεγιστοποιημένοι λογάριθμοι των πιθανοφανειών των δύο μοντέλων. Μεγάλες τιμές της στατιστικής συνάρτησης D δείχνουν ότι το μοντέλο M_1 επεξηγεί περισσότερη από την μεταβλητότητα των δεδομένων από ότι το μοντέλο M_0 ενώ αντίθετα μικρές τιμές της D δείχνουν ότι η αύξηση στο μέγεθος του μοντέλου M_1 δεν έφερε σημαντικές βελτιώσεις στην ικανότητά του να επεξηγεί περισσότερη μεταβλητότητα από ότι το M_0 . Πως όμως μπορούμε να ορίσουμε τις μικρές ή τις μεγάλες τιμές της στατιστικής συνάρτησης D ; Λύση σε αυτό το πρόβλημα δίνει η ασυμπτωτική κατανομή της στατιστικής συνάρτησης D η οποία αποδεικνύεται ότι είναι μία χ^2 κατανομή με κ βαθμούς ελευθερίας,

οπού κ είναι η διάφορα στις διαστάσεις του μοντέλου M_1 από το μοντέλο M_0 . Έτσι, στο παραπάνω παράδειγμα, όπου οι παράμετροι είναι συναρτήσεις του χρόνου, στο M_0 θα είναι

$$\mu(i) = \mu_0 + \mu_1 i \quad \text{και} \quad \ln \sigma(i) = \sigma_0 + \sigma_1 i,$$

στο μοντέλο M_1 θα είναι

$$\mu(i) = \mu_0 + \mu_1 i + \mu_2 i^2 \quad \text{και} \quad \ln \sigma(i) = \sigma_0 + \sigma_1 i + \sigma_2 i^2$$

και θέλουμε να ελέγξουμε ποιο από τα δύο μοντέλα είναι καταλληλότερο, λαμβάνοντας υπόψη ότι η στατιστική συνάρτηση D σε αυτή την περίπτωση ακολουθεί μία χ^2 κατανομή με 2 β.ε. Επομένως το κριτήριο για την απόρριψη του μοντέλου M_0 σε επίπεδο σημαντικότητας α είναι

$$D > \chi_{\kappa}^2(1 - \alpha) : \text{άνω } \alpha \text{ ποσοστημόριο της } \chi^2 \text{ κατανομής με } \kappa \text{ β.ε.},$$

Τέλος, ένας έλεγχος που ξεφεύγει από τις έννοιες της ενότητας αυτής αλλά είναι πολύ σημαντικός για την στατιστική ανάλυση των ακραίων παρατηρήσεων είναι ο έλεγχος για το αν η παράμετρος σχήματος ξ είναι μηδέν ή όχι. Αν η παράμετρος σχήματος ξ είναι μηδέν αναφερόμαστε σε μία κατανομή *Gumbel* ενώ αν είναι διάφορη του μηδενός αναφερόμαστε σε μία κατανομή *GEV*. Με την ορολογία της ενότητας αυτής το M_1 είναι το μοντέλο οπού έχουμε υποθέσει ότι οι ακραίες παρατηρήσεις προέρχονται από μία κατανομή *GEV* και M_0 το μοντέλο οπού έχουμε υποθέσει ότι οι ακραίες παρατηρήσεις προέρχονται από μία κατανομή *Gumbel* τότε σε αντιστοιχία με τα προηγούμενα η στατιστική συνάρτηση του έλεγχου είναι

$$D = 2(l_1(M_1) - l_0(M_0))$$

όπου $l_1(M_1)$, $l_0(M_0)$ είναι οι μεγιστοποιημένοι λογάριθμοι των πιθανοφανειών των δυο μοντέλων.

3.3.7 Έλεγχοι καλής προσαρμογής σε μη-στάσιμες παρατηρήσεις

Έχοντας δοκιμάσει διάφορα μοντέλα και έχοντας επιλέξει μεταξύ αυτών το καταλληλότερο, θα πρέπει με κάποιο τρόπο να διασφαλίσουμε ότι το τελικό μοντέλο που υιοθετήσαμε προσαρμόζεται ικανοποιητικά στα δεδομένα. Στο προηγούμενο κεφάλαιο

αναφερθήκαμε σε ελέγχους καλής προσαρμογής χρησιμοποιώντας όμως την υπόθεση ότι οι παρατηρήσεις προέρχονται από την ίδια κατανομή, είναι δηλαδή ισόνομες. Όμως τώρα οι ακραίες παρατηρήσεις είναι μη-στάσιμες, επομένως δεν προέρχονται από την ίδια κατανομή, και άρα οι μέθοδοι αυτοί δεν μπορούν να χρησιμοποιηθούν. Λύση σε αυτό το πρόβλημα δίνει η τυποποίηση των αρχικών μη-στάσιμων παρατηρήσεων και ο μετασχηματισμός τους σε παρατηρήσεις που κατανέμονται ισόνομα (βλ. π.χ. Coles(2001)).

Έτσι $Z_i \sim GEV(\mu(i), \sigma(i), \xi(i))$ όπου οι παράμετροι $\mu(i), \sigma(i), \xi(i)$ είναι συναρτήσεις του χρόνου ή άλλων μεταβλητών. Τότε η τυποποιημένη τ.μ. που ορίζεται ως \square

$$Z'_i = \frac{1}{\xi(i)} \ln\left(1 + \xi(i) \left(\frac{Z_i - \mu(i)}{\sigma(i)}\right)\right)$$

ακολουθεί την κατανομή Gumbel που έχει σ.κ.

$$F(x) = \exp(-e^{-x}), x \in \mathbf{R}$$

Από αυτή την τυποποίηση μπορούμε να κατασκευάσουμε κατάλληλα P-P Plots και Q-Q Plots για τις τυποποιημένες τ.μ. Z'_i που θα ακολουθούν μία κατανομή *Gumbel*. Συμβολίζουμε με $Z'_{(1)}, Z'_{(2)}, \dots, Z'_{(k)}$ τις διατεταγμένες παρατηρήσεις της τυποποιημένης τ.μ. τότε το P-P Plot θα αποτελείται από τα σημεία

$$\left\{ \frac{i}{k+1}, \exp(-\exp(-Z'_{(i)})) : i = 1, 2, \dots, k \right\}$$

και το Q-Q Plot από τα σημεία

$$\left\{ Z'_{(i)}, -\ln(-\ln(\frac{i}{k+1})) : i = 1, 2, \dots, k \right\}$$

Επομένως τα τυποποιημένα δεδομένα θα προσαρμόζονται ικανοποιητικά στην κατανομή *Gumbel* και κατά συνέπεια σε μία μη-στάσιμη κατανομή *GEV* αν τα k σημεία του γραφήματος βρίσκονται «κοντά» στην κύρια διαγώνιο.

ΚΕΦΑΛΑΙΟ 4

Εμπειρική αξιολόγηση της μεθόδου Block-Maxima σε περιβαλλοντολογικά δεδομένα

4.1 Εμπειρική αξιολόγηση της μεθόδου Block-Maxima σε περιβαλλοντολογικά δεδομένα

Πριν προχωρήσουμε στην στατιστική ανάλυση και συμπερασματολογία των περιβαλλοντολογικών δεδομένων που έχουμε στην διάθεση μας, θα ήταν αρκετά χρήσιμο να αξιολογηθούν εμπειρικά οι εκτιμήσεις των παραμέτρων της κατανομής GEV , που προκύπτει ως φυσική συνέπεια της χρησιμοποίησης της μεθόδου Block-Maxima κατά την ανάλυση περιβαλλοντολογικών δεδομένων. Το βασικό ερώτημα που θέλουμε να εξετάσουμε είναι αν η επαναληπτική μέθοδος που θα χρησιμοποιήσουμε για τον υπολογισμό των εκτιμήσεων των παραμέτρων (μέσω του πακέτου R) είναι αξιόπιστη, δηλαδή εκτιμά αποτελεσματικά την τάση και την εποχικότητα τα δύο βασικά χαρακτηριστικά που εμφανίζονται στα περιβαλλοντολογικά δεδομένα, και επομένως τα συμπεράσματα που θα εξαχθούν μέσω αυτής της μεθόδου είναι ασφαλή. Επίσης μας ενδιαφέρει να εξετάσουμε, αν και κατά πόσο βελτιώνονται ή χειροτερεύουν οι εκτιμήσεις των παραμέτρων καθώς μεταβάλλεται το

μέγεθος του δείγματος. Με βάση όλα τα παραπάνω το ενδιαφέρον έγκειται στην εμπειρική αξιολόγηση των εκτιμήσεων των παραμέτρων από μη-στάσιμες κατανομές μεγίστων, μιας και αυτές παρουσιάζουν τα χαρακτηριστικά που θέλουμε να μελετήσουμε.

Η γενική διαδικασία που θα ακολουθήσουμε για να απαντήσουμε στα δύο παραπάνω ερωτήματα θα είναι η εξής □

- 1) Αρχικά θα παράγουμε n τυχαίους αριθμούς από την μη-στάσιμη κατανομή μεγίστων GEV μέσω προσομοίωσης, χρησιμοποιώντας την μέθοδο της αντιστροφής (για περισσότερες λεπτομέρειες βλ. Μπούτσικας(2004)) δίνοντας συγκεκριμένες τιμές στις παραμέτρους της κατανομής.
- 2) Στη συνέχεια με βάση την εξίσωση πιθανοφάνειας για μη-στάσιμες κατανομές μεγίστων, που αναφέρθηκε στην Ενότητα 3.3.4., εκτιμούμε αριθμητικά μέσω του πακέτου R τις παραμέτρους της κατανομής από όπου προέρχονται τα δεδομένα.
- 3) Επαναλαμβάνουμε τα βήματα (1) και (2) 1000 φορές και αποθηκεύουμε τις εκτιμήσεις των παραμέτρων που παίρνουμε για κάθε n τυχαίους αριθμούς (διαφορετικούς κάθε φορά) από την συγκεκριμένη κατανομή.

Την παραπάνω μέθοδο θα την εφαρμόσουμε για την κατανομή GEV όπου για την GEV η παράμετρος θέσης θα μοντελοποιείται ως γραμμική συνάρτηση της τάσης και περιοδική συνάρτηση του χρόνου,

Το μέγεθος δείγματος, που εδώ ταυτίζεται με το πόσους τυχαίους αριθμούς θα παράξουμε στο βήμα (1) της επαναληπτικής μεθόδου, θα παίρνει τέσσερις διαφορετικές τιμές 100, 200, 500, 1000 έτσι ώστε να εξετάζουμε την μεταβολή των εκτιμήσεων καθώς μεταβάλλεται το μέγεθος του δείγματος. Τέλος η ακρίβεια των εκτιμητών θα εξετάζεται μέσω της κατασκευής ιστογράμματος όπου θα απεικονίζονται οι 1000 εκτιμήσεις των παραμέτρων των κατανομών.

Ειδικότερα, θα παράξουμε n τυχαίους αριθμούς από την $GEV(\mu(i), \sigma(i), \zeta(i))$ με σ.κ.

$$G(z_i, \mu(i), \sigma(i), \xi(i)) = \exp[-(1 + \xi(i) (\frac{z_i - \mu(i)}{\sigma(i)}))^{-1/\xi(i)}]$$

όπου οι παράμετροι της κατανομής θα έχουν την εξής μοντελοποίηση \square

$$\mu(i) = \mu_0 + \mu_1 \sin(\frac{2\pi i}{m}) + \mu_2 \cos(\frac{2\pi i}{m}) + \mu_3 i$$

$$\sigma(i) = \sigma_0$$

$$\xi(i) = \xi_0$$

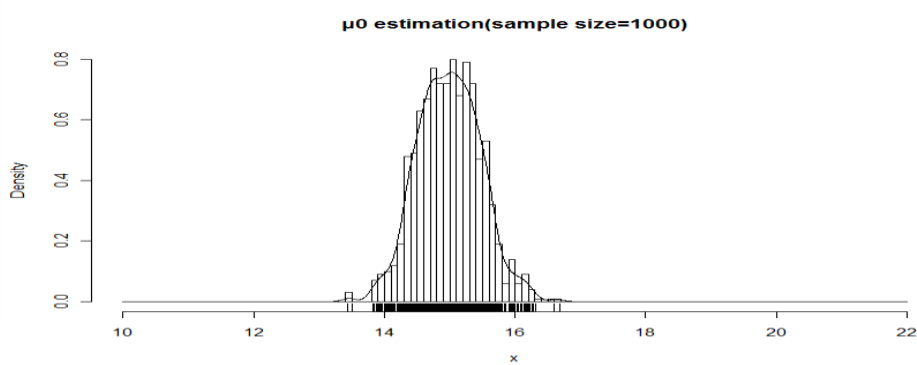
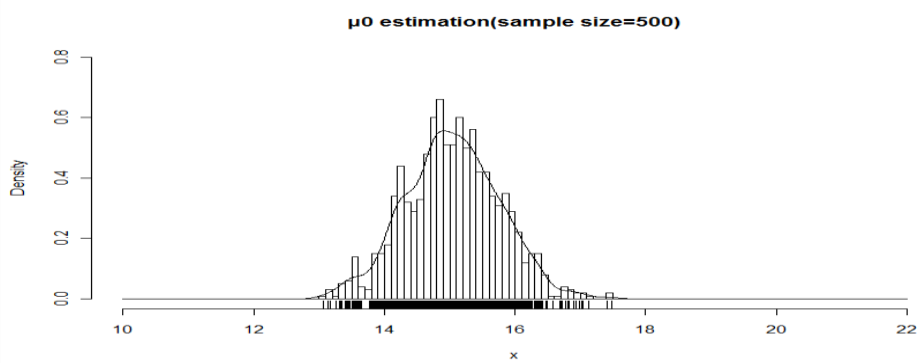
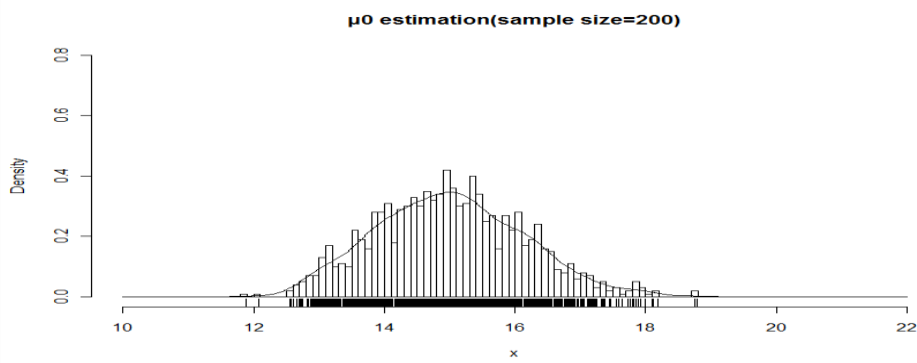
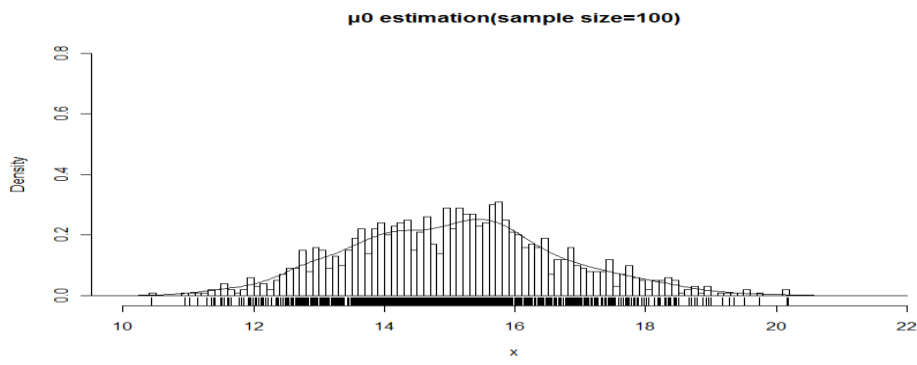
όπου για το i θα ισχύει ότι $i = 1, 2, \dots, n$ και οι τιμές που θα έχουν οι παράμετροι θα είναι $\mu_0 = 15, \mu_1 = 2, \mu_2 = 3, \mu_3 = -0.004, \sigma_0 = 8, \xi_0 = 0.2, m = 12$. Οι τιμές επιλέχτηκαν έτσι ώστε να αντικατοπτρίζουν πραγματικά περιβαλλοντολογικά φαινόμενα.

Στη συνέχεια ακολουθούν τα ιστογράμματα των εκτιμήσεων των παραμέτρων της $GEV(\mu(i), \sigma(i), \xi(i))$ για 4 διαφορετικά μεγέθη δείγματος ($n = 100, 200, 500, 1000$) (Παράρτημα Κώδικας 1).

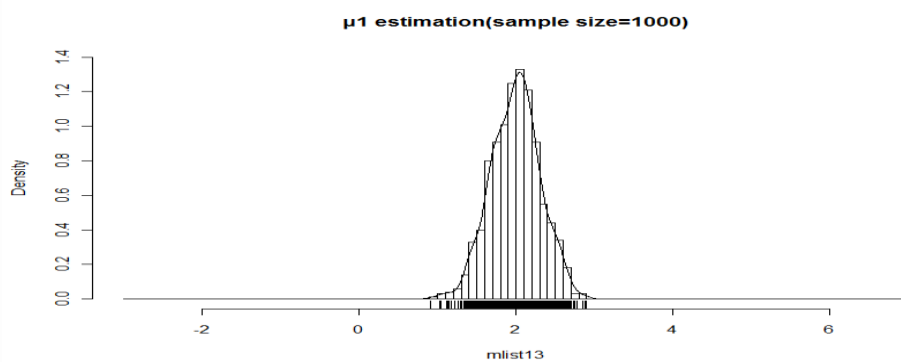
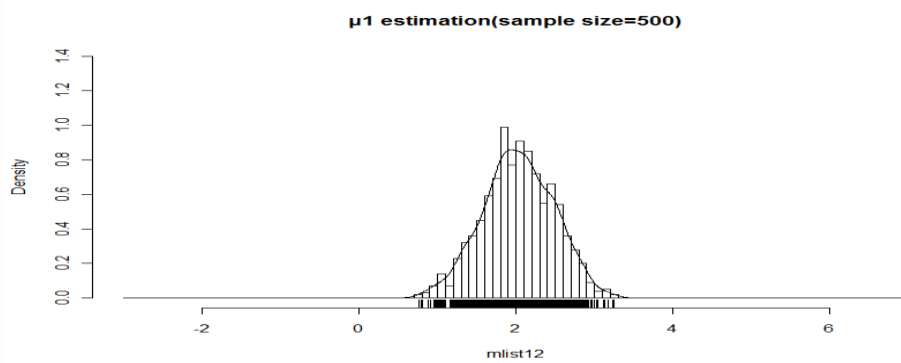
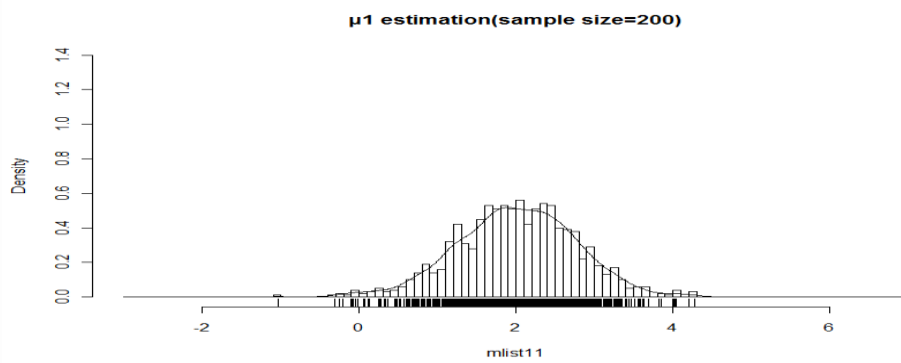
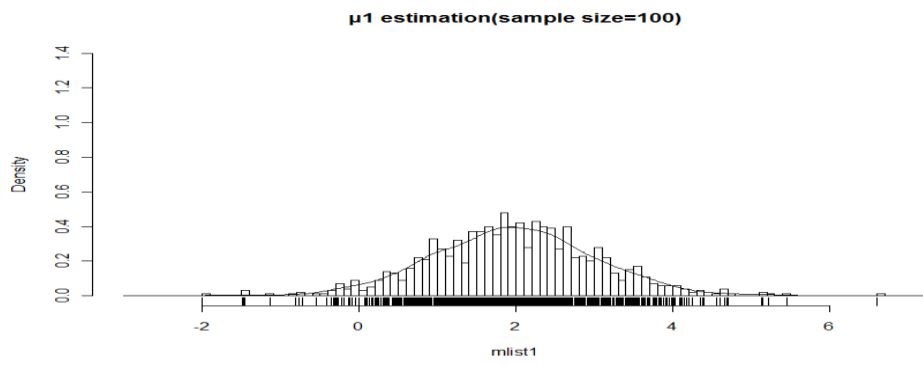
Σχήμα 4.1

Εκτιμήσεις των παραμέτρων της GEV κατανομής για μέγεθος δείγματος $n = 100, 200, 500, 1000$.

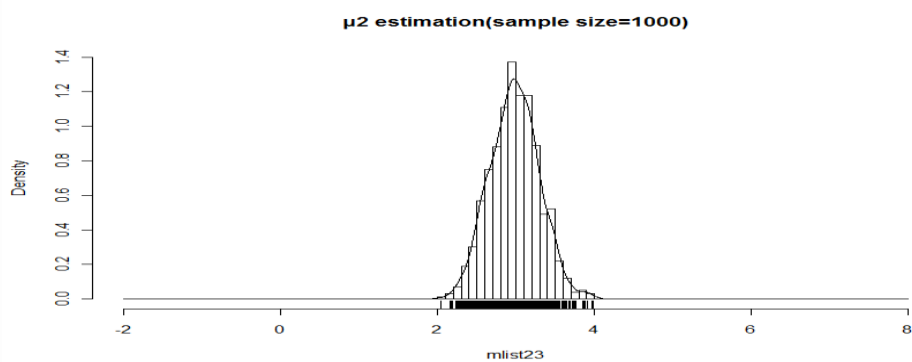
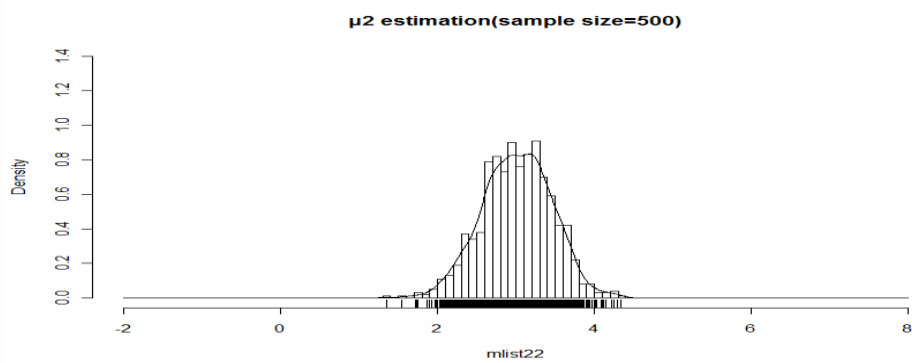
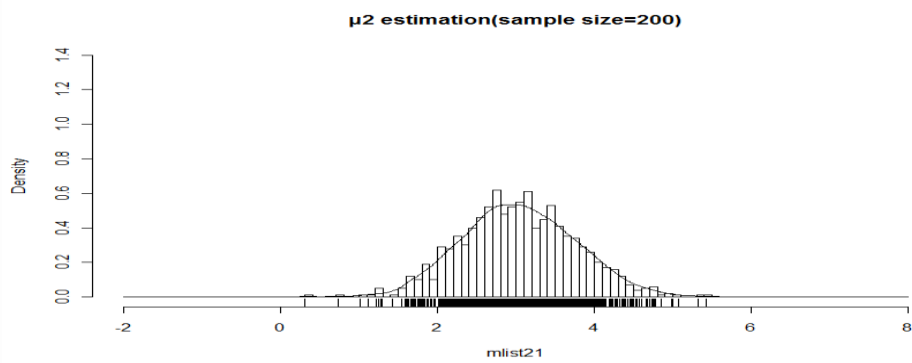
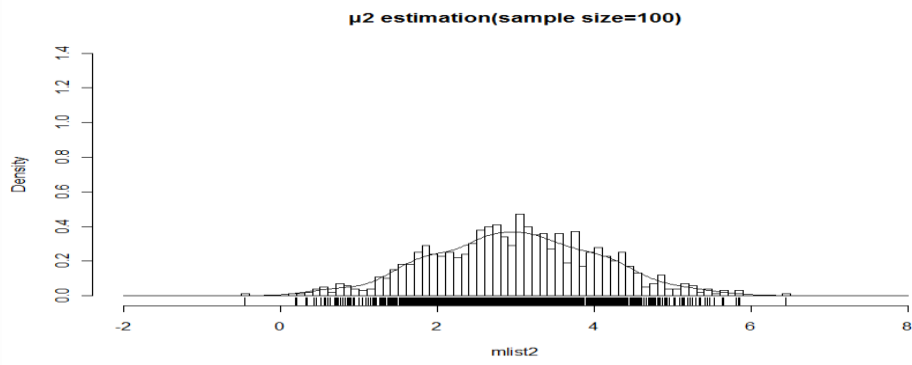
- 1) Εκτιμήσεις της παραμέτρου μ_0 για μέγεθος δείγματος $n = 100, 200, 500, 1000$.



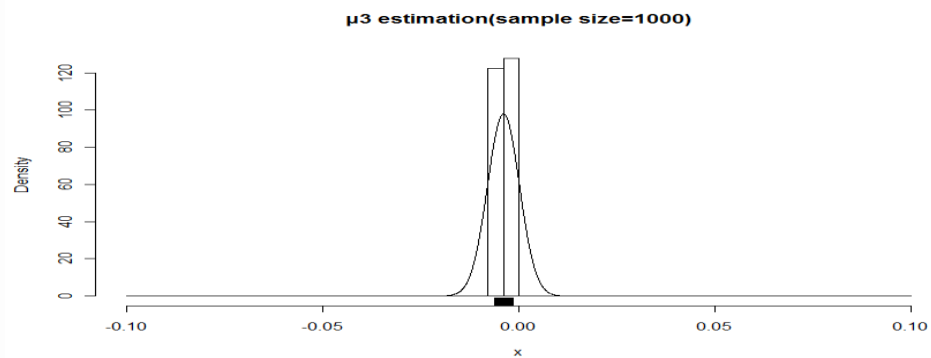
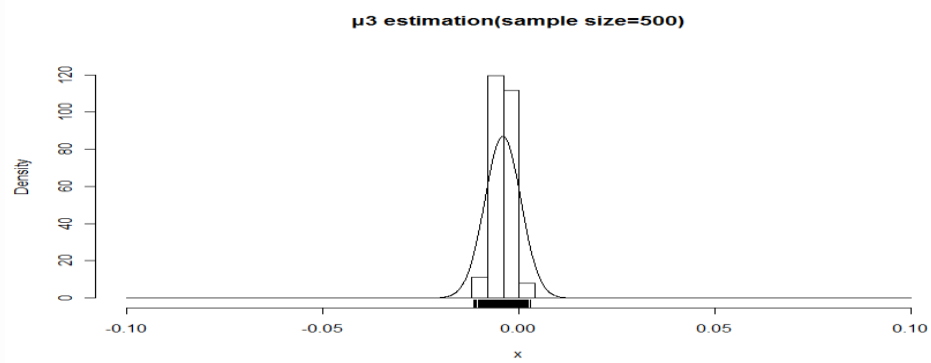
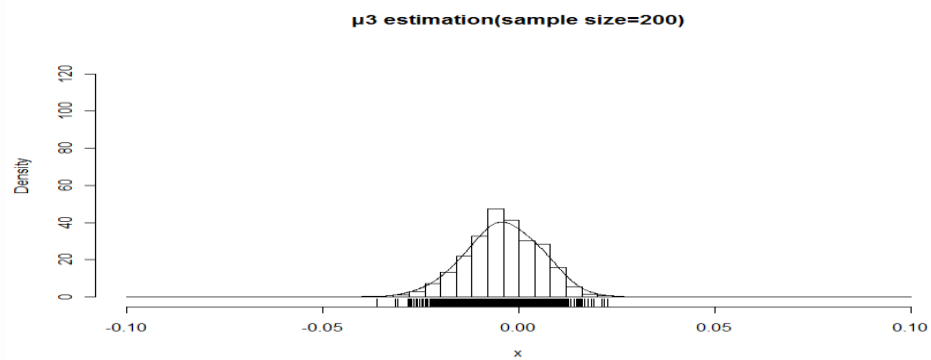
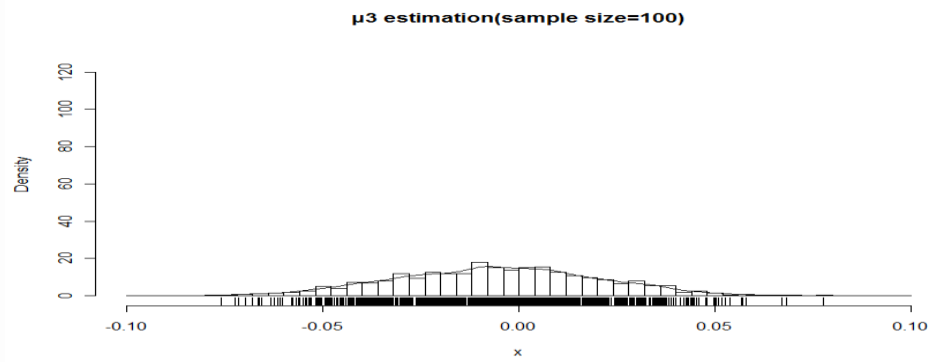
2) Εκτιμήσεις της παραμέτρου μ_1 για μέγεθος δείγματος $n = 100, 200, 500, 1000$.



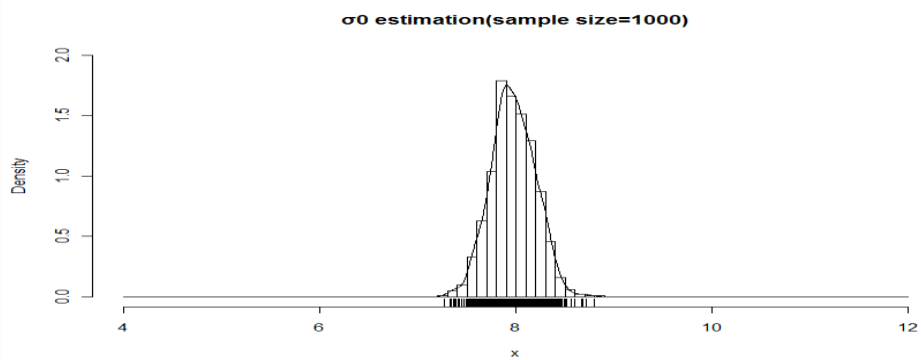
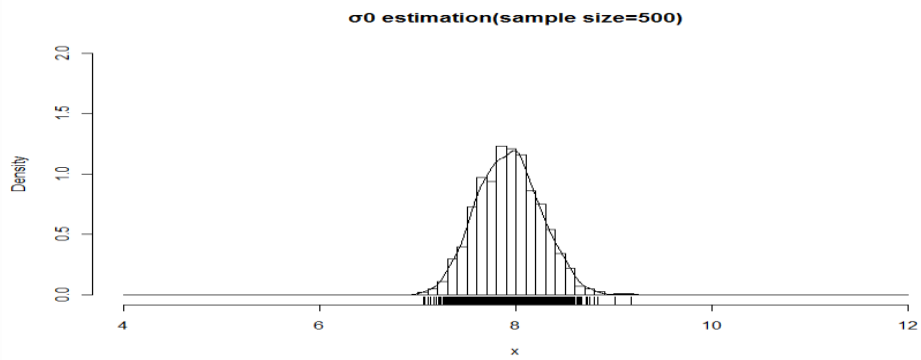
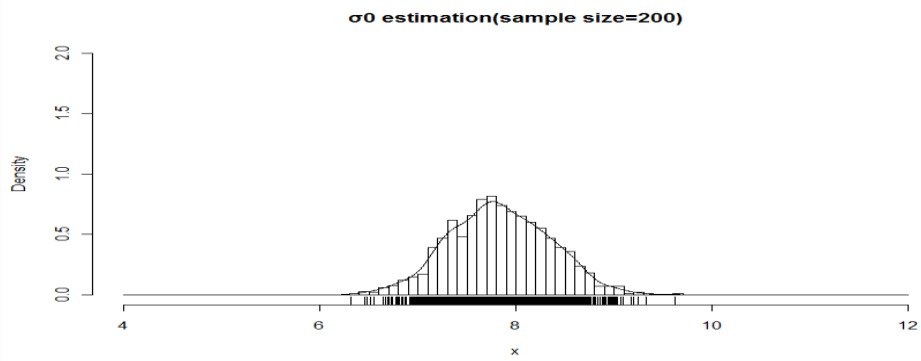
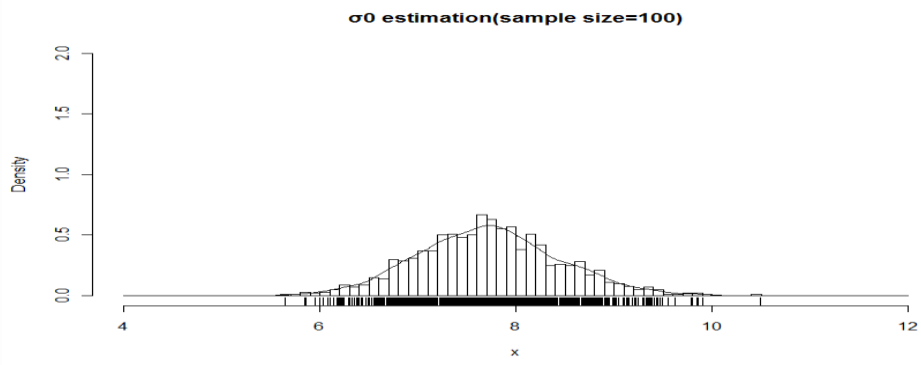
3) Εκτιμήσεις της παραμέτρου μ_2 για μέγεθος δείγματος $n = 100, 200, 500, 1000$.



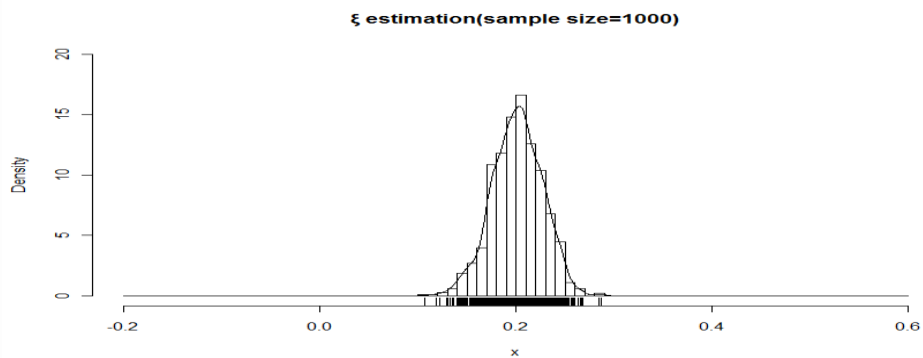
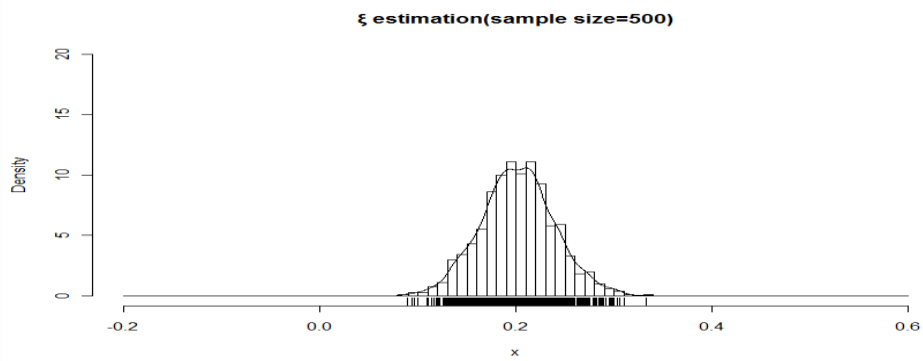
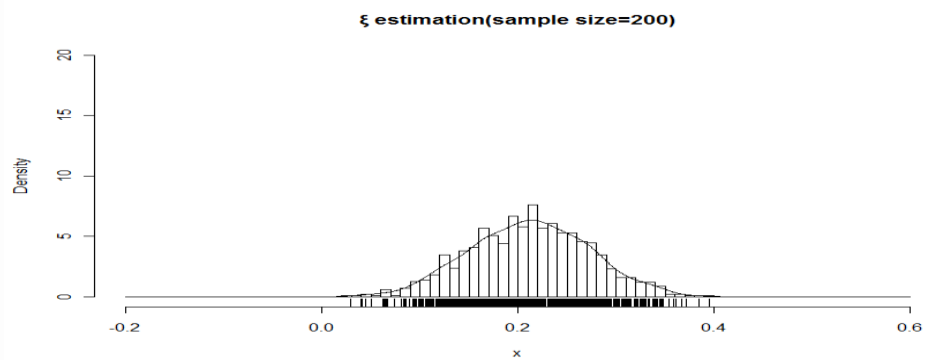
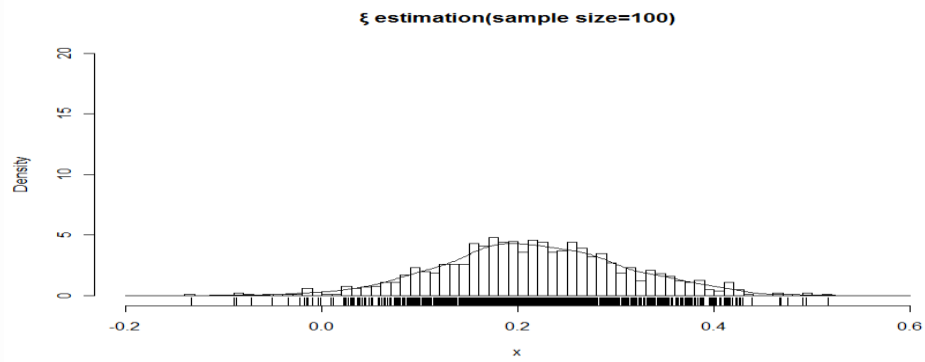
4) Εκτιμήσεις της παραμέτρου μ_3 για μέγεθος δείγματος $n = 100, 200, 500, 1000$.



5) Εκτιμήσεις της παραμέτρου σ_0 για μέγεθος δείγματος $n=100, 200, 500, 1000$.



6) Εκτιμήσεις της παραμέτρου ξ για μέγεθος δείγματος $n=100, 200, 500, 1000$.



Από τα παραπάνω ιστογράμματα για την εμπειρική αξιολόγηση της μη-στάσιμης κατανομής *GEV* μπορούμε να πούμε ότι η γενική εικόνα για την επαναληπτική μέθοδο

εκτίμησης των παραμέτρων είναι αρκετά καλή. Για μέγεθος δείγματος $n = 100$ και 200 οι εκτιμήσεις για όλες τις παραμέτρους και κυρίως οι εκτιμήσεις για τις παραμέτρους που σχετίζονται άμεσα με την εποχικότητα έχουν μεγαλύτερη διασπορά από ότι στις περιπτώσεις με μεγαλύτερο n . Αυτό οφείλεται στο ότι τα δεδομένα πάνω στα οποία προσαρμόζεται η επαναληπτική μέθοδος έχουν ως κύριο χαρακτηριστικό την περιοδικότητα (η περίοδος των δεδομένων είναι 12 παρατηρήσεις) και αφού το μέγεθος του δείγματος είναι μικρό η περιοδικότητα είναι δύσκολο να εκτιμηθεί ακριβώς, εφόσον το φαινόμενο επαναλαμβάνεται λίγες φορές. Όμως στην περίπτωση που το μέγεθος δείγματος είναι μεγάλο δηλαδή $n = 500$ ή 1000 η διασπορά των εκτιμητών είναι ελάχιστη επομένως έχουμε μεγάλη ακρίβεια της μεθόδου στην εκτίμηση των παραμέτρων, κάτι το οποίο σημαίνει ότι ο επαναληπτικός αλγόριθμος είναι αρκετά αποτελεσματικός. Επίσης πρέπει να σημειωθεί ότι ο αλγόριθμος φαίνεται να δίνει αμερόληπτες εκτιμήσεις (δεν παρατηρείται σημαντική μεροληψία) και δεν οδηγεί σε ακραίες εκτιμήσεις που απέχουν αρκετά από τις πραγματικές τιμές των παραμέτρων, κάτι που θα οδηγούσε σε εντελώς εσφαλμένη εκτίμηση και των αποτελεσμάτων και φυσικά σε λανθασμένη συμπερασματολογία. Τέλος από τα ιστογράμματα μπορούμε να παρατηρήσουμε ότι στις περισσότερες των περιπτώσεων, η κατανομή των εκτιμητών φαίνεται να είναι κανονική, επομένως ίσως να μπορούσαμε να ισχυριστούμε ότι οι εκτιμητές είναι ασυμπτωτικά κανονικοί ιδιότητα η οποία είναι επιθυμητή στα πλαίσια της αξιολόγησης των εκτιμητών.

ΚΕΦΑΛΑΙΟ 5

Εφαρμογή της θεωρίας ακραίων τιμών σε περιβαλλοντολογικά δεδομένα της Ελλάδας

5.1 Περιβαλλοντολογικά δεδομένα και μέθοδος ανάλυσης

Στο παρόν κεφάλαιο, θα αναλυθούν περιβαλλοντολογικά δεδομένα και πιο συγκεκριμένα μετεωρολογικά-κλιματολογικά δεδομένα από τον Ελλαδικό χώρο με τις μεθόδους που αναπτύχθηκαν στα Κεφάλαια 2 και 3. Το σύνολο των δεδομένων που θα αναλυθούν, χορηγήθηκε από την Εθνική Μετεωρολογική Υπηρεσία (Ε.Μ.Υ) για τους σκοπούς της συγκεκριμένης Διπλωματικής Εργασίας. Τα δεδομένα προέρχονται από 21 μετεωρολογικούς σταθμούς που καλύπτουν όλο τον Ελλαδικό χώρο. Η επιλογή των μετεωρολογικών σταθμών έγινε με τρία κριτήρια □

α) Να καλύπτουν όσο το δυνατόν μεγαλύτερη γεωγραφική επιφάνεια της Ελλάδος έτσι ώστε να είναι διακριτές οι κλιματολογικές διαφορές που υπάρχουν στην Ελλάδα, κατά την εκτίμηση των παραμέτρων του μοντέλου,

β) Να περιέχουν όσο το δυνατόν μεγαλύτερο χρονικό εύρος παρατηρήσεων ώστε τα αποτελέσματα που θα εξαχθούν μέσω της ανάλυσης να είναι ασφαλή και η συμπερασματολογία που θα στηριχτεί πάνω σε αυτά να είναι στέρεη,

γ) Να έχουν όσο το δυνατόν λιγότερες ελλειψείς τιμές, έτσι ώστε τα αποτελέσματα που θα εξαχθούν να μην επηρεάζονται από τις ελλειψείς τιμές του κάθε μετεωρολογικού σταθμού.

Στο τέλος αυτής της ενότητας παρατίθεται ο πίνακας με τους 21 μετεωρολογικούς σταθμούς, με τις γεωγραφικές συντεταγμένες τους καθώς και το χρονικό εύρος των παρατηρήσεων του κάθε σταθμού.

Τα περιβαλλοντολογικά – μετεωρολογικά δεδομένα που θα αναλυθούν σε αυτό το κεφάλαιο αφορούν μέγιστες βροχοπτώσεις και μέγιστες θερμοκρασίες στους 21 σταθμούς που έχουν επιλεγεί για τις βροχοπτώσεις, και στους 18 σταθμούς για τις θερμοκρασίες. Η επιλογή αυτού του είδους των δεδομένων έγινε λόγω των ιδιαίτερων χαρακτηριστικών που εμφανίζουν τα δύο αυτά φυσικά φαινόμενα στον Ελλαδικό χώρο, άποψη η οποία ενισχύεται και από διαφορές εργασίες που έχουν γίνει για την ανάλυση αυτών των φαινομένων, είτε σε μεμονωμένες περιοχές της Ελλάδος (Koutsouyiannis et al.(2000)) είτε σε όλο τον Ελλαδικό χώρο (Kioutsoukakis et al.(2009)).

Το χρονικό εύρος των μεγίστων των κλάσεων πάρθηκε να είναι μηνιαίο και όχι ετήσιο, έτσι ώστε να μοντελοποιείται κατάλληλα ο ετήσιος κλιματικός κύκλος που εμφανίζουν τα φαινόμενα αυτά, παράγοντας ο οποίος είναι πολύ σημαντικός στην ανάλυση ακραίων περιβαλλοντολογικών φαινομένων. Επίσης μέσω αυτής της επιλογής του χρονικού εύρους μπορούμε να αντλήσουμε περισσότερη πληροφορία από το σύνολο των δεδομένων, αφού σε ετήσια βάση αντιστοιχούν δώδεκα ακραίες παρατηρήσεις, αντί μίας παρατήρησης που θα αντιστοιχούσε αν το χρονικό εύρος των μεγίστων ήταν ετήσιο.

Τέλος, αποτέλεσμα αυτής της μοντελοποίησης είναι να μπορούμε να παράγουμε στάθμες απόδοσης για κάθε μήνα του έτους, κάτι το οποίο είναι πιο χρήσιμο σε διάφορους τομείς της οικονομικής και κοινωνικής δραστηριότητας σε αντίθεση με την παράγωγη ετήσιας στάθμης απόδοσης.

Κλείνοντας, τα μηνιαία μέγιστα κλάσεων και για τις δύο κατηγορίες δεδομένων θα μοντελοποιηθούν με τις αντίστοιχες μη-στάσιμες κατανομές μέγιστων, *GEV* και *Gumbel*, όπου η παράμετρος θέσης θα είναι περιοδική συνάρτηση του χρόνου και της τάσης, και η παράμετρος κλίμακος θα είναι περιοδική συνάρτηση του χρόνου, ενώ η παράμετρος σχήματος θα θεωρηθεί σταθερή σε σχέση με το χρόνο. Η μοντελοποίηση αυτή επιλέχθηκε, από όλες τις μεθόδους που αναφέρθηκαν στο Κεφάλαιο 3 για τη μοντελοποίηση της περιοδικότητας και της τάσης, λόγω του ότι με αυτό το τρόπο δεν εισέρχεται μεγάλος αριθμός παραμέτρων (κάτι που βοηθά στην ευκολότερη εύρεση των ε.μ.π.), και οδηγεί σε ευκολότερη ερμηνεία των αποτελεσμάτων. Επομένως θεωρούμε ότι οι μηνιαίες μέγιστες παρατηρήσεις $Z_i \sim GEV(\mu(i), \sigma(i), \zeta(i))$ όπου οι παράμετροι της κατανομής *GEV* θα έχουν την εξής μοντελοποίηση \square

$$\begin{aligned}\mu(i) &= \mu_0 + \mu_1 \sin\left(\frac{2\pi i}{12}\right) + \mu_2 \cos\left(\frac{2\pi i}{12}\right) + \mu_3 i \\ \sigma(i) &= \exp\left(\sigma_0 + \sigma_1 \sin\left(\frac{2\pi i}{12}\right) + \sigma_2 \cos\left(\frac{2\pi i}{12}\right)\right) \\ \zeta(i) &= \zeta_0\end{aligned}$$

Η επιλογή του καταλληλότερου μοντέλου, θα γίνεται για κάθε σταθμό και για κάθε σύνολο δεδομένων μέσω των ελέγχων του λόγου πιθανοφανειών που ορίστηκε στην Ενότητα 3.3.6.

Πίνακας 5.1

Οι 21 μετεωρολογικοί σταθμοί με τις γεωγραφικές συντεταγμένες τους καθώς και το χρονικό εύρος των παρατηρήσεων που εξετάζουμε.

Μετεωρολογικός Σταθμός	Γεωγραφικό Πλάτος	Γεωγραφικό Μήκος	Χρονική περίοδος
1) Αλεξανδρούπολη	40.51 ⁰	25.53 ⁰	1951-2003
2) Αεροδρόμιο Μακεδονία	40.31 ⁰	22.58 ⁰	1959-2004
3) Φλώρινα	40.48 ⁰	21.25 ⁰	1961-2004

4) Κοζάνη	40.18 ⁰	21.47 ⁰	1955-2004
5) Ιωάννινα	39.41 ⁰	20.49 ⁰	1957-2004
6) Κέρκυρα	39.37 ⁰	19.55 ⁰	1955-2004
7) Λάρισα	39.38 ⁰	22.15 ⁰	1955-2004
8) Αγχιάλος	39.13 ⁰	22.48 ⁰	1956-2004
9) Σκύρος	38.58 ⁰	24.29 ⁰	1955-2004
10) Ελληνικό	37.44 ⁰	23.41 ⁰	1956-2004
11) Ν. Φιλαδέλφεια	38.03 ⁰	23.73 ⁰	1955-2004
12) Πειραιάς	37.59 ⁰	23.34 ⁰	1956-2004
13) Αγρίνιο	38.36 ⁰	21.21 ⁰	1956-2004
14) Αίγιο	38.14 ⁰	22.44 ⁰	1974-2003
15) Αργοστόλι	38.11 ⁰	20.29 ⁰	1970-2003
16) Τρίπολη	37.32 ⁰	22.24 ⁰	1957-2004
17) Καλαμάτα	37.40 ⁰	22.10 ⁰	1956-2004
18) Νάξος	37.40 ⁰	25.22 ⁰	1955-2004
19) Μυτιλήνη	39.34 ⁰	26.35 ⁰	1955-2004
20) Ρόδος	36.24 ⁰	28.5 ⁰	1955-2004
21) Ηράκλειο	35.19 ⁰	25.10 ⁰	1955-2004

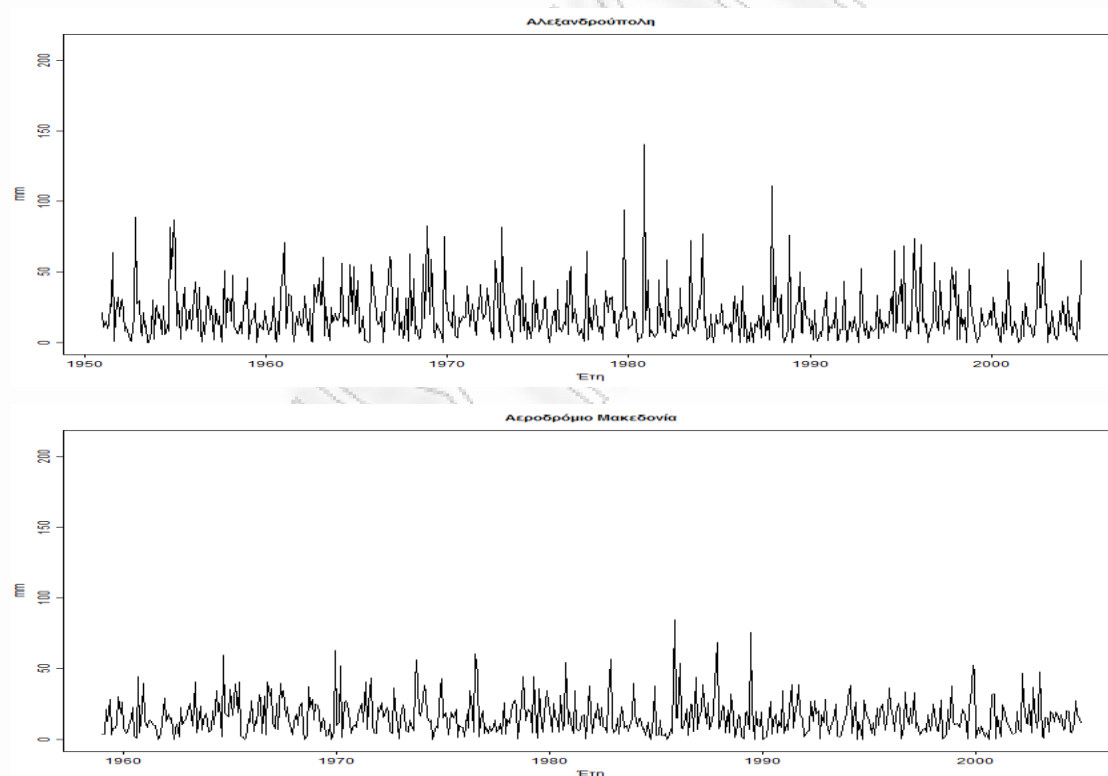
5.1.1 Ανάλυση μέγιστων βροχοπτώσεων

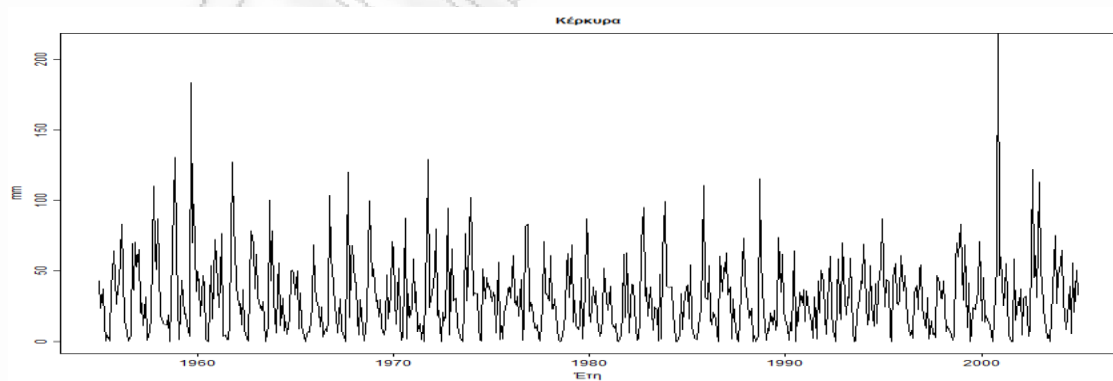
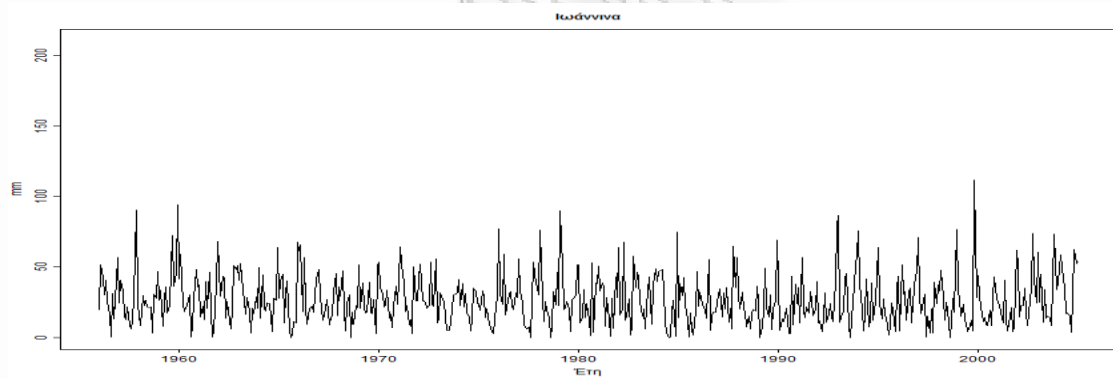
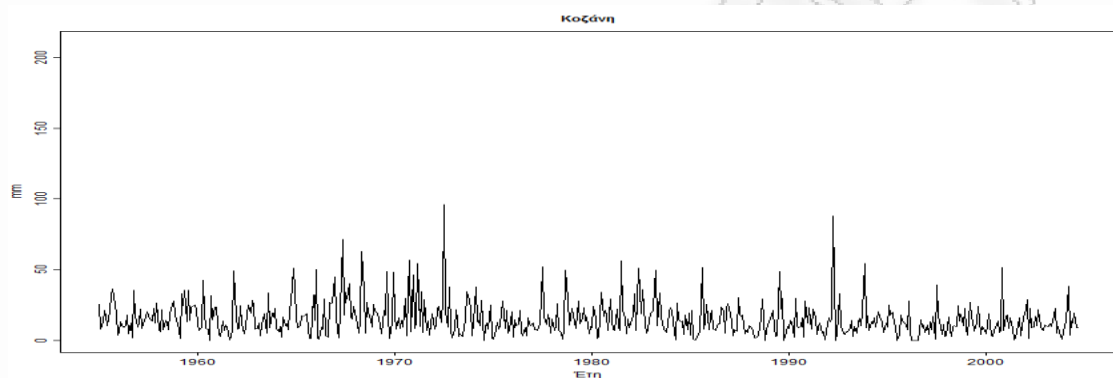
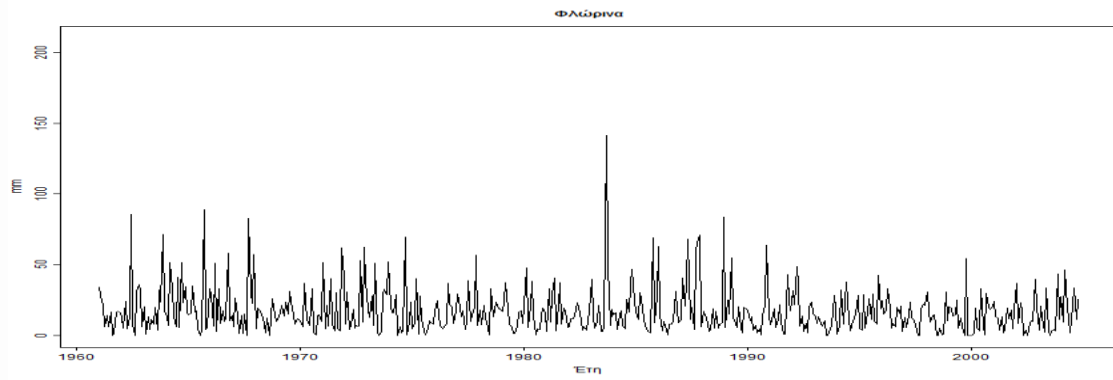
Από τον πίνακα 5.1 μπορούμε να παρατηρήσουμε ότι η πλειονότητα των μετεωρολογικών σταθμών έχει εύρος παρατηρήσεων που ξεπερνά τα 45 χρόνια, επομένως με την επιλογή των μέγιστων των κλάσεων να είναι μηνιαία, έχουμε σε κάθε σύνολο δεδομένων αριθμό παρατηρήσεων μεγαλύτερο από 500, κάτι το οποίο θα μας βοηθήσει να οδηγηθούμε σε ασφαλή συμπεράσματα κατά την ανάλυση των ακραίων βροχοπτώσεων. Στο σχήμα 5.2 παρατίθεται το γράφημα της χρονολογικής σειράς των μηνιαίων μέγιστων βροχοπτώσεων για κάθε έναν από τους 21 μετεωρολογικούς σταθμούς. Από τα γραφήματα αυτά μπορούμε να παρατηρήσουμε την έντονη περιοδικότητα που εμφανίζουν οι μηνιαίες μέγιστες βροχοπτώσεις. Επίσης στο σχήμα 5.3

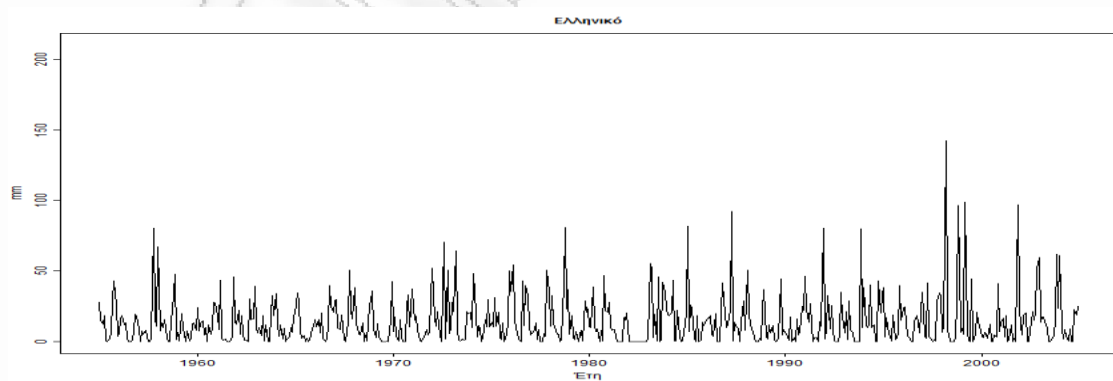
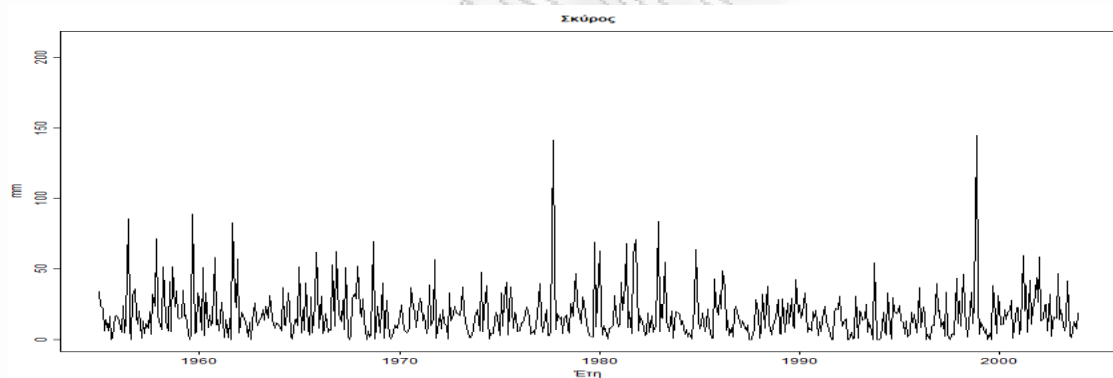
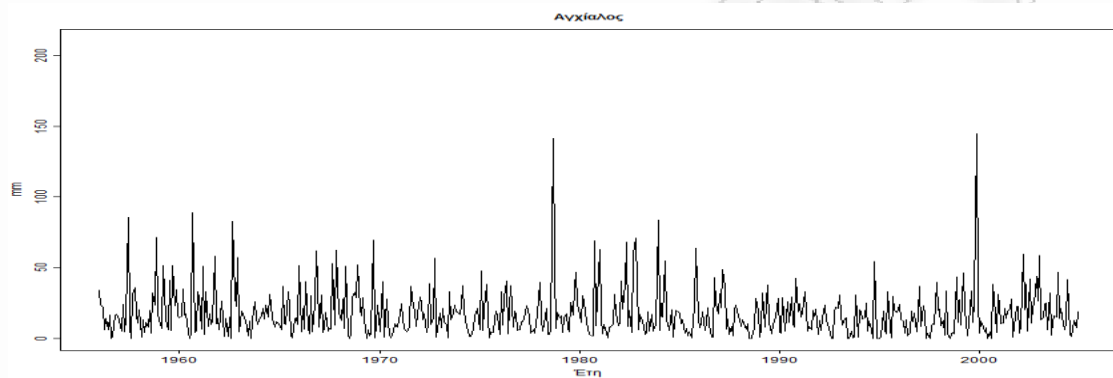
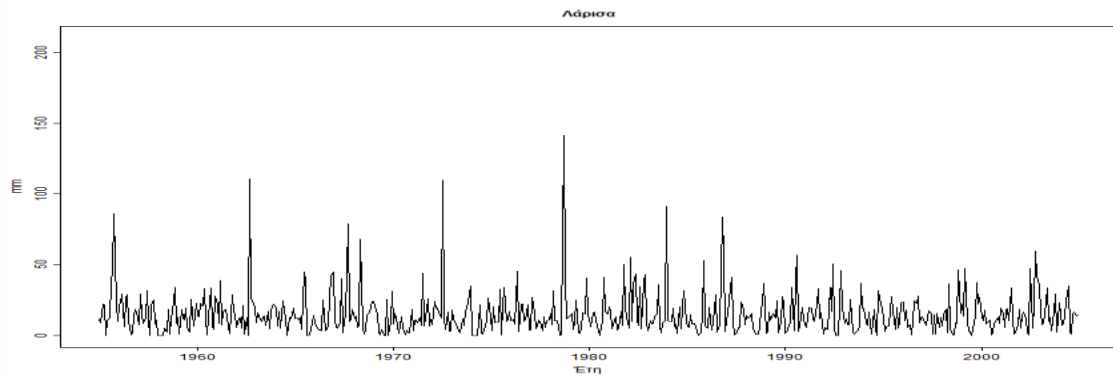
παρατίθενται τα θηκογράμματα των μηνιαίων μέγιστων βροχοπτώσεων για κάθε μετεωρολογικό σταθμό όπου είναι εμφανείς οι κλιματολογικές διαφορές ανάμεσα στους σταθμούς. Αξίζει να αναφέρουμε ότι η βροχόπτωση έχει ως μονάδα μέτρησης το χιλιοστό ύψους βροχής (mm). Στην πράξη βροχόπτωση ύψους 1 χιλιοστού ισοδυναμεί με 1 λίτρο νερού σε 1 τετραγωνικό μέτρο (m^2) ή με 1 κυβικό μέτρο(m^3) νερού σε 1 στρέμμα (1 στρέμμα= $1000 m^2$).

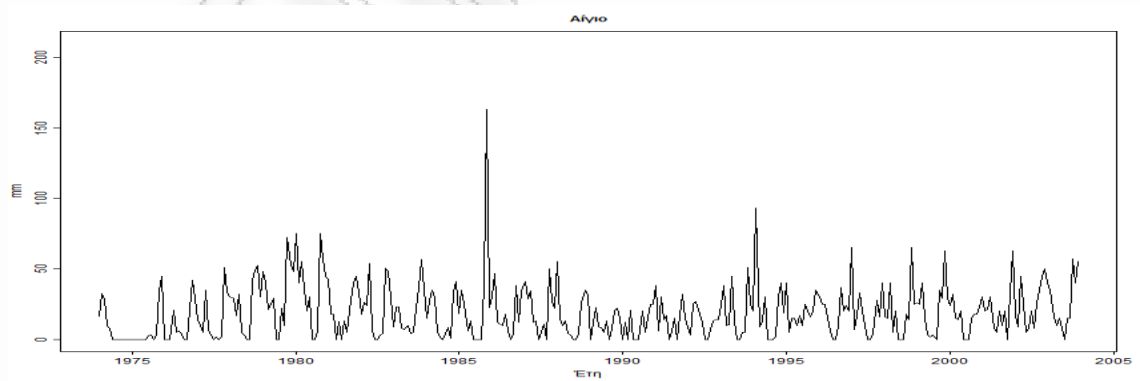
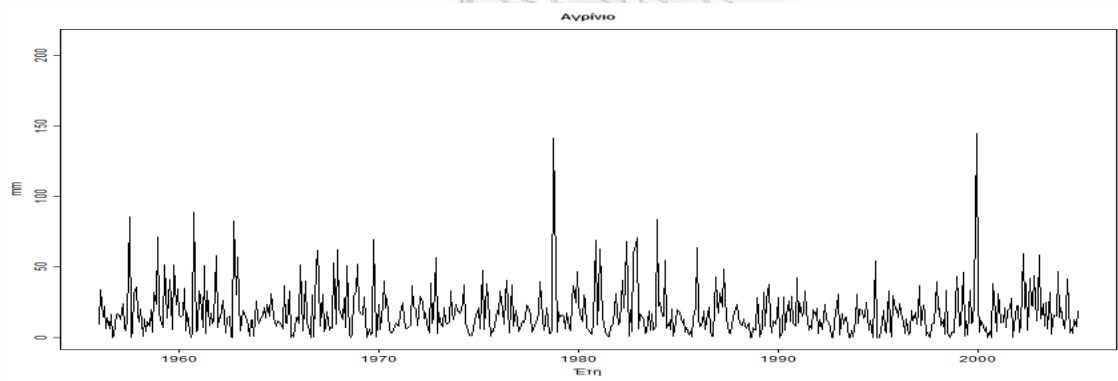
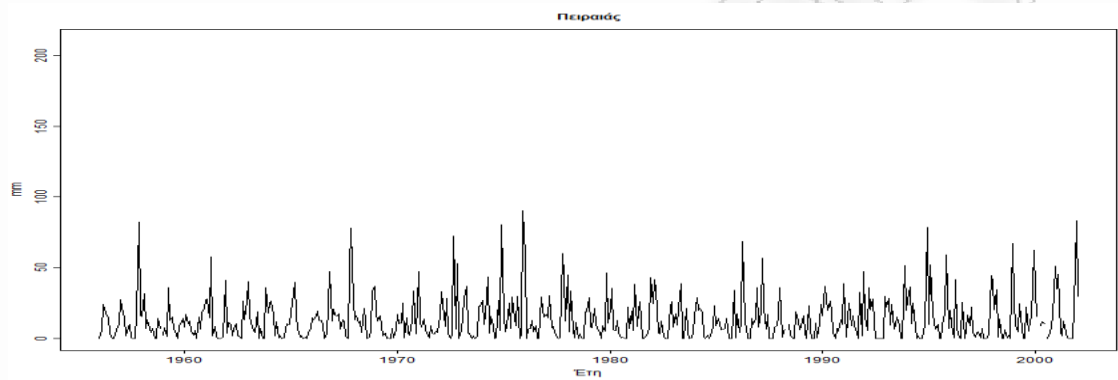
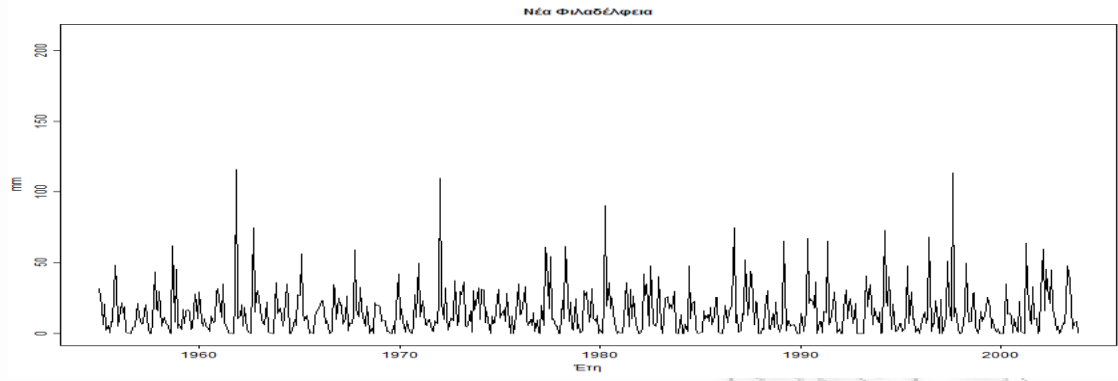
Σχήμα 5.1

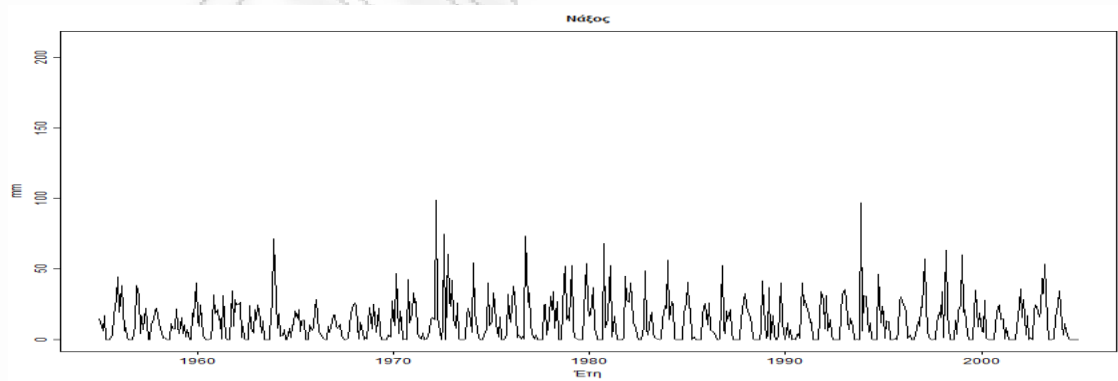
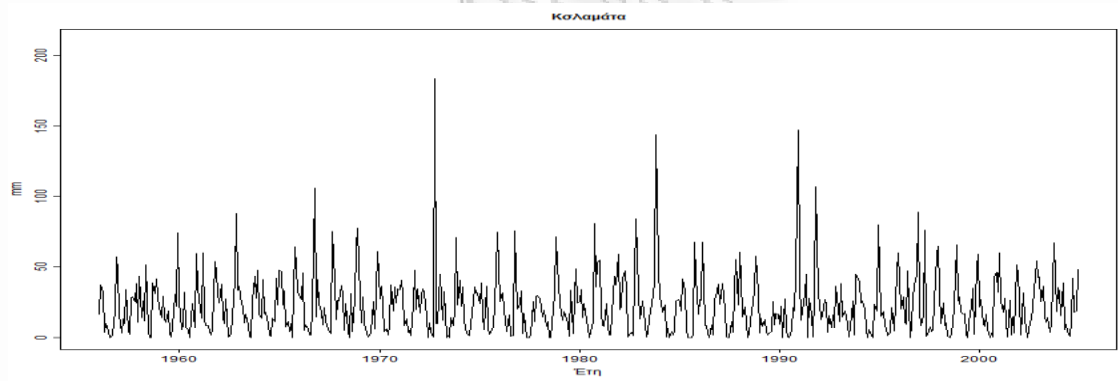
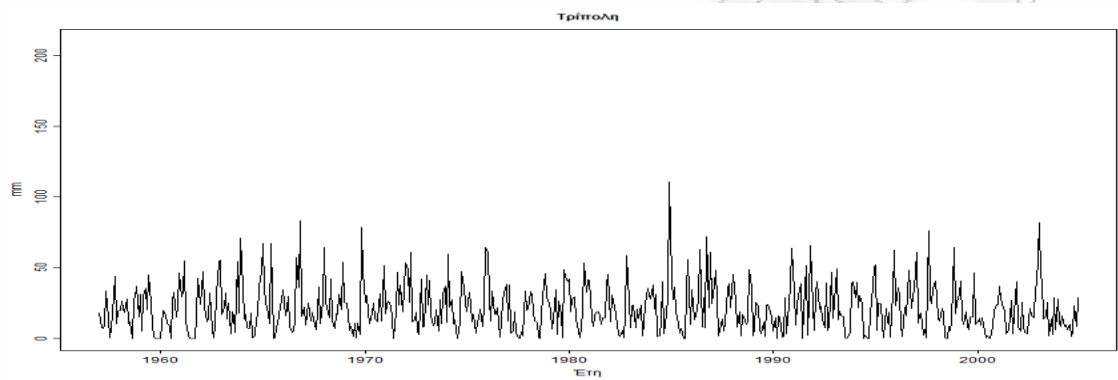
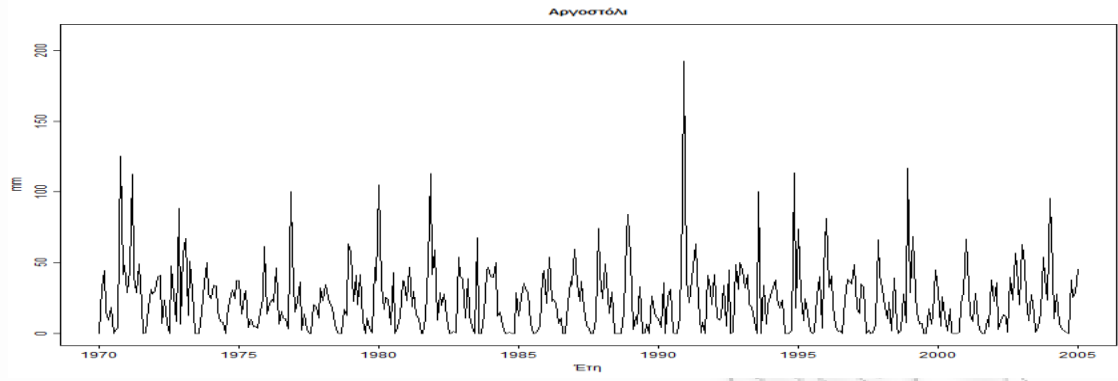
Χρονολογική σειρά μηνιαίων μέγιστων βροχοπτώσεων για τους 21 μετεωρολογικούς σταθμούς.

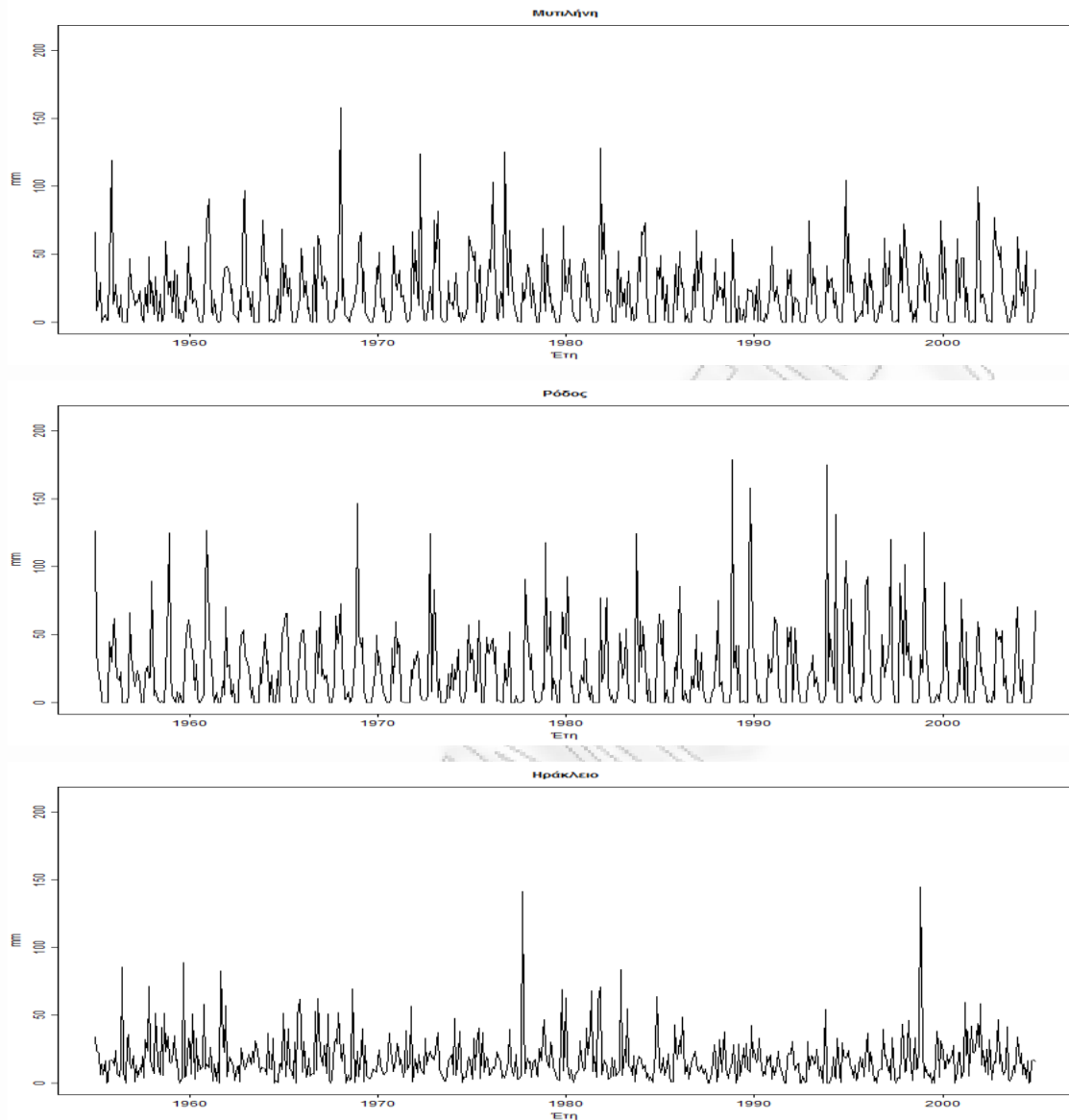






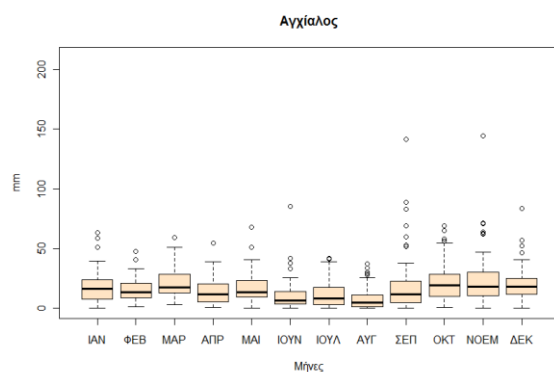
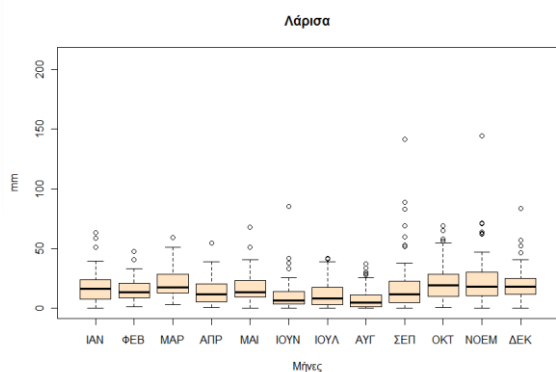
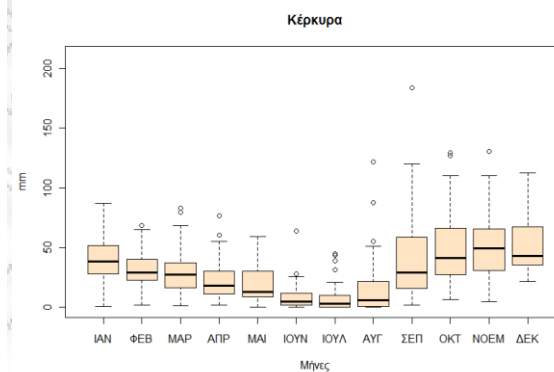
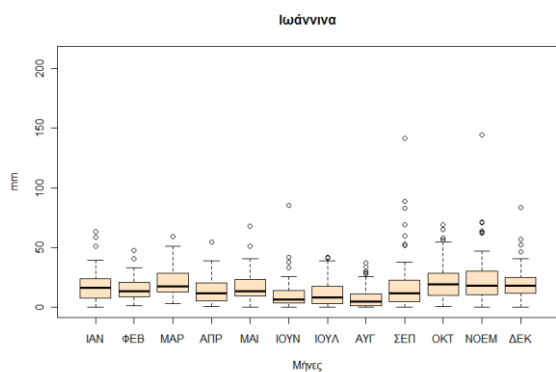
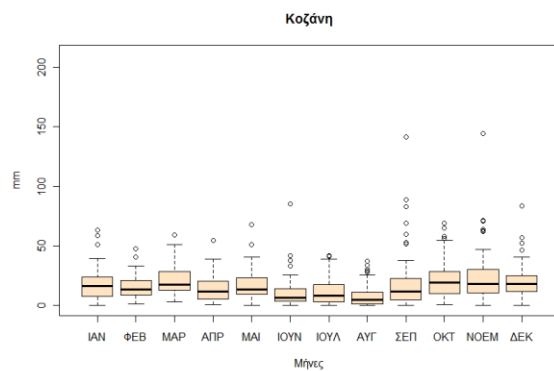
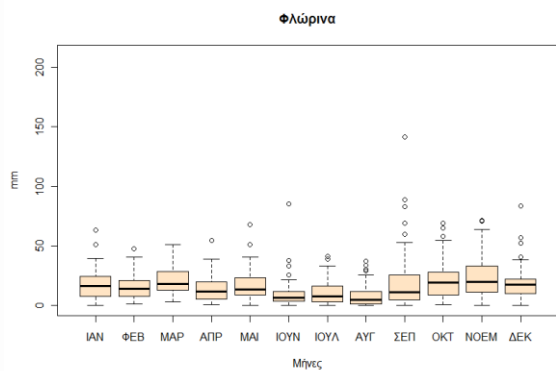
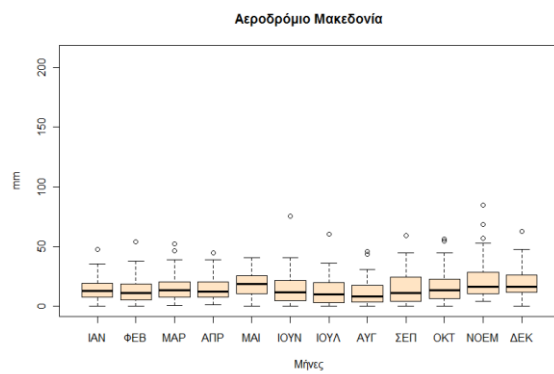
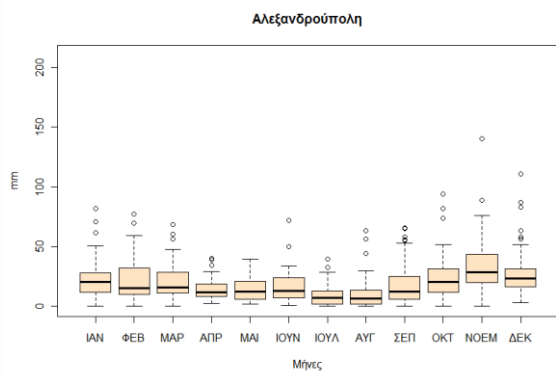


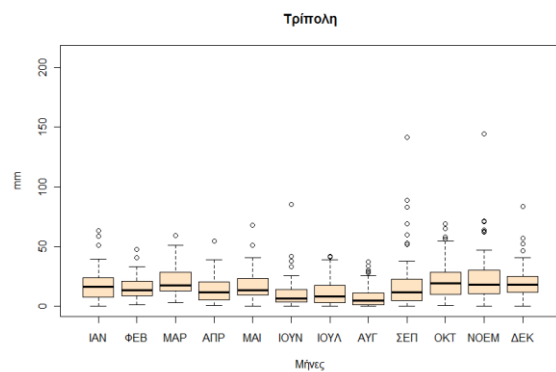
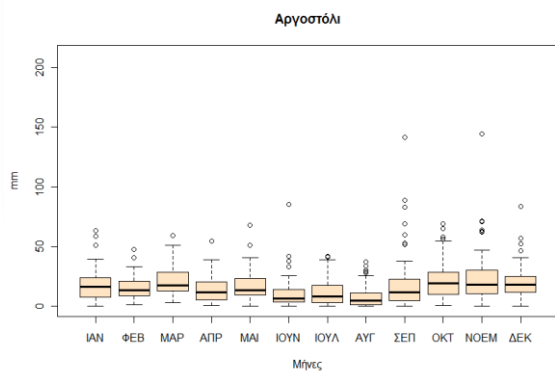
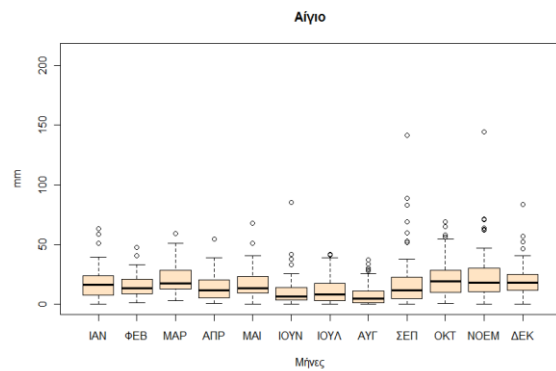
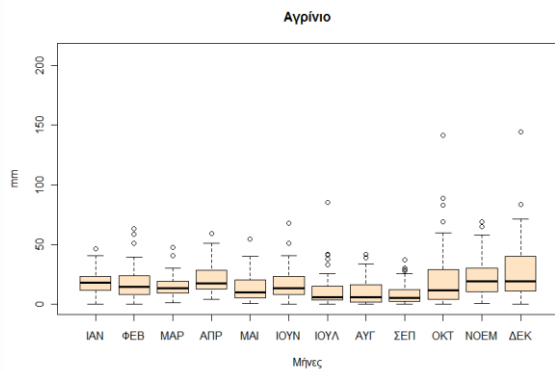
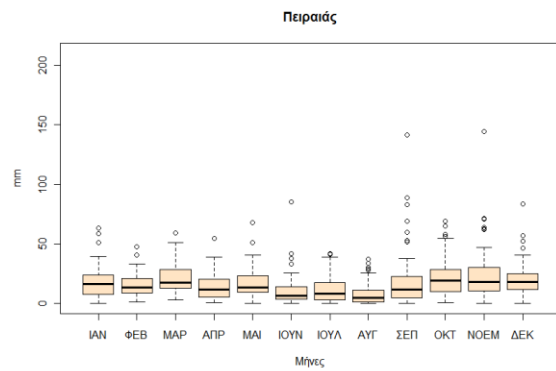
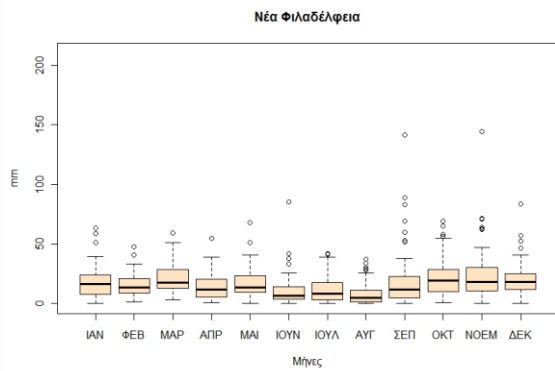
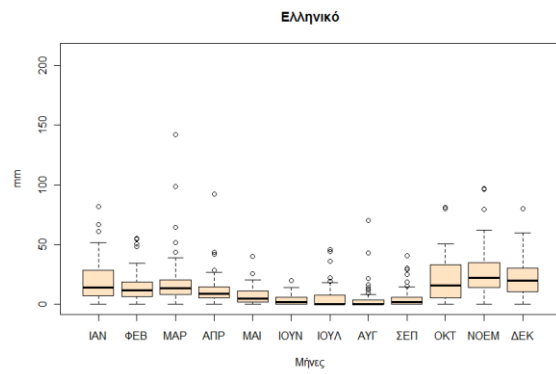
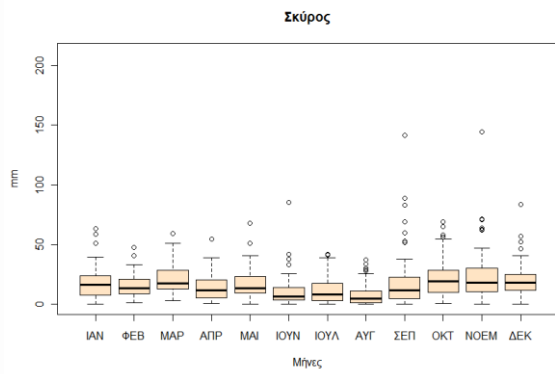


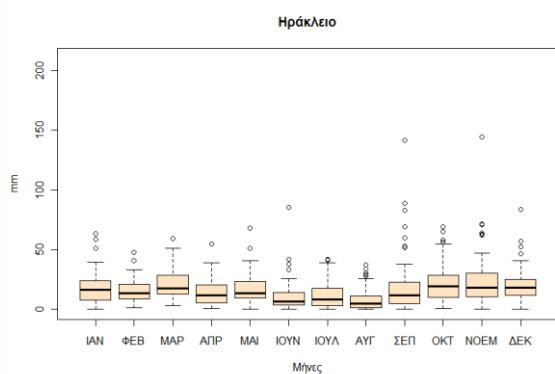
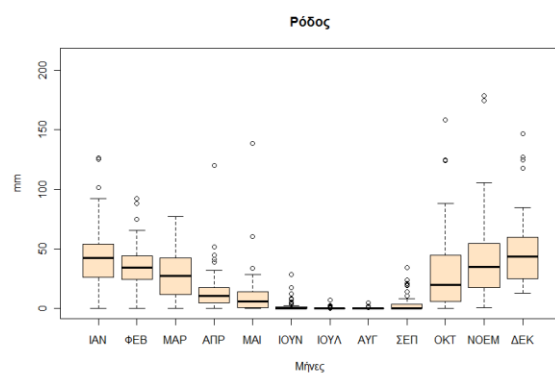
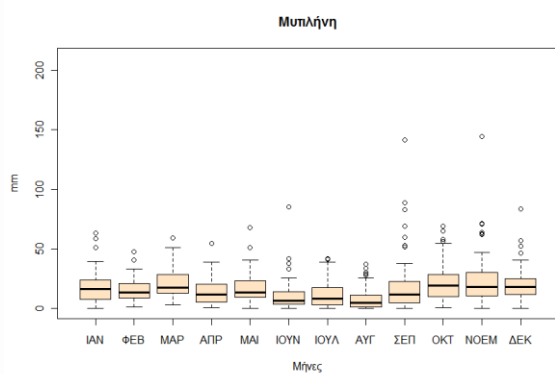
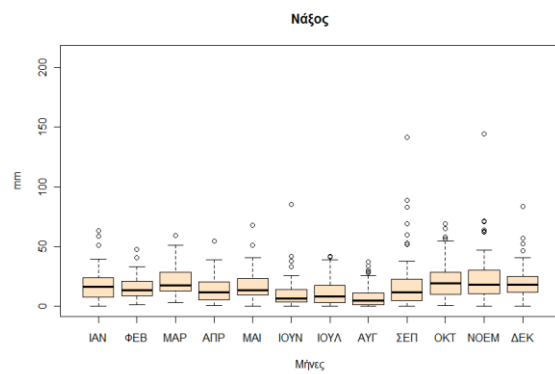
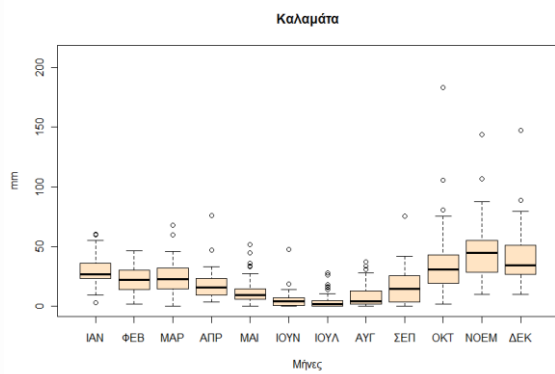


Σχήμα 5.2

Θηκογράμματα μηνιαίων μέγιστων βροχοπτώσεων για τους 21 μετεωρολογικούς σταθμούς.







Οι ιδιαίτερες κλιματολογικές συνθήκες που επικρατούν σε πολλές περιοχές της Ελλάδας κατά τους καλοκαιρινούς μήνες, όπου οι βροχοπτώσεις είναι ελάχιστες έως μηδαμινές, έχουν ως αποτέλεσμα σε αρκετούς μετεωρολογικούς σταθμούς τα μηνιαία μέγιστα να μην έχουν καθόλου καλή προσαρμογή στις κατανομές ακροτάτων τις οποίες χρησιμοποιούμε στην μέθοδο block maxima, με τελική συνέπεια τα αποτελέσματα τα οποία θα εξάγαμε από την μέθοδο να μην είναι καθόλου αξιόπιστα. Για να αντιμετωπίσουμε αυτό το πρόβλημα εφαρμόσαμε την ακόλουθη τεχνική, χωρίς να επηρεαστεί καθόλου η περιοδικότητα των δεδομένων. Συμπτύξαμε ίδιους μήνες δύο, τριών ή τεσσάρων διαδοχικών ετών σε μία κλάση (block) αυξάνοντας με αυτό το τρόπο το μέγεθος της κλάσης από 30 παρατηρήσεις, όπου

ήταν η μηνιαία κλάση, σε 60, 90 ή και 120 παρατηρήσεις. Έτσι π.χ. στην περίπτωση όπου αναφερόμαστε σε μετεωρολογικό σταθμό όπου το μέγεθος της κλάσης είναι ίσο με 60, και ο χρονικός ορίζοντας των παρατηρήσεων ξεκινά από το 1955 ακολουθήσαμε την έξη διαδικασία για να πάρουμε τα μέγιστα των κλάσεων □ συμπτύξαμε τον Ιανουάριο του 1955 και τον Ιανουάριο του 1956 σε μία κλάση και πήραμε την μέγιστη παρατήρηση, τον Φεβρουάριο του 1955 και του 1956 και κάναμε την ίδια διαδικασία και παρόμοια δουλέψαμε για όλους τους μήνες του έτους και όλα τα έτη των παρατηρήσεων. Αντίστοιχα εργαζόμαστε και στην περίπτωση όπου το μέγεθος της κλάσης είναι ίσο με 90 ή 120 παρατηρήσεις. Με αυτή τη διαδικασία η περιοδικότητα δεν επηρεάζεται καθόλου από την αλλαγή στο μέγεθος των κλάσεων, η προσαρμογή των μέγιστων στις κατανομές ακρότατων βελτιώνεται αισθητά, μειώνεται όμως και ο αριθμός των παρατηρήσεων, δηλαδή το μέγεθος του δείγματος κατά το $\frac{1}{2}$ το $\frac{1}{3}$ ή το $\frac{1}{4}$. Επίσης η ερμηνεία των διάφορων αποτελεσμάτων δεν είναι ίδια για μετεωρολογικούς σταθμούς με διαφορετικό μέγεθος κλάσης, κάτι το οποίο θα φανεί και στην συνέχεια. Στη συνέχεια ακολουθεί ο πίνακας 5.2 με τους μετεωρολογικούς σταθμούς και τον αριθμό των παρατηρήσεων όπου περιέχονται σε κάθε κλάση (block) έτσι ώστε να έχουμε ικανοποιητική προσαρμογή των δεδομένων. Μία αναλυτικότερη παρουσίαση της παραπάνω μεθόδου υπάρχει στις εργασίες των Engeland et al.(2004) και Rust et al.(2009).

Πίνακας 5.2

Οι 21 μετεωρολογικοί σταθμοί και το μέγεθος κάθε κλάσης από όπου προέρχονται οι μέγιστες βροχοπτώσεις.

Μετεωρολογικός Σταθμός	Μέγεθος κλάσης
1) Αλεξανδρούπολη	30
2) Αεροδρ. Μακεδονία	30
3) Φλώρινα	30
4) Κοζάνη	30
5) Ιωάννινα	30
6) Κέρκυρα	30
7) Λάρισα	30
8) Αγχιάλος	30
9) Σκύρος	120
10) Ελληνικό	120
11) Ν. Φιλαδέλφεια	60
12) Πειραιάς	60
13) Αγρίνιο	90
14) Αίγιο	30
15) Αργοστόλι	90
16) Τρίπολη	60
17) Καλαμάτα	60
18) Νάξος	120
19) Μυτιλήνη	90
20) Ρόδος	90
21) Ηράκλειο	120

Στους πίνακες 5.3 και 5.4 παρουσιάζονται οι εκτιμήσεις των παραμέτρων μέγιστης πιθανοφάνειας της κατανομής *GEV* για τις μέγιστες βροχοπτώσεις με την μοντελοποίηση που εισήχθη στην Ενότητα 5.1, δηλαδή οι παράμετροι της κατανομής θα έχουν την παρακάτω μορφή □

$$\mu(i) = \mu_0 + \mu_1 \sin\left(\frac{2\pi i}{12}\right) + \mu_2 \cos\left(\frac{2\pi i}{12}\right) + \mu_3 i$$

$$\sigma(i) = \exp\left(\sigma_0 + \sigma_1 \sin\left(\frac{2\pi i}{12}\right) + \sigma_2 \cos\left(\frac{2\pi i}{12}\right)\right)$$

$$\xi(i) = \xi_0$$

και τα τυπικά σφάλματα αυτών για τους 21 μετεωρολογικούς σταθμούς που παρουσιάστηκαν παραπάνω. Η σειρά με την οποία εμφανίζονται οι σταθμοί στους παρακάτω πίνακες, είναι η ίδια με την οποία εμφανίζονται στον πίνακα 5.1. Στους σταθμούς όπου υπάρχει η τιμή μηδέν στην παράμετρο σχήματος ξ_0 σημαίνει ότι τα δεδομένα σε αυτό τον σταθμό προέρχονται από κατανομή *Gumbel*, ενώ αντίστοιχα στους σταθμούς όπου η παράμετρος μ_3 είναι μηδέν η ύπαρξη τάσης, δεν ήταν στατιστικά σημαντική σε επίπεδο σημαντικότητας 5%. Επίσης στους σταθμούς όπου οι εκτιμημένες τιμές των παραμέτρων περιέχουν την ένδειξη (*), σημαίνει ότι οι παράμετροι αυτοί δεν ήταν στατιστικά σημαντικοί για να εισέλθουν στο μοντέλο σε επίπεδο σημαντικότητας 5%.

Πίνακας 5.3

Εκτιμήσεις παραμέτρων $\mu_0, \mu_1, \mu_2, \mu_3$ κατανομής *GEV* για τις μέγιστες βροχοπτώσεις και τυπικά σφάλματα εκτιμητών.

	$\hat{\mu}_0$	<i>s.e.</i> ($\hat{\mu}_0$)	$\hat{\mu}_1$	<i>s.e.</i> ($\hat{\mu}_1$)	$\hat{\mu}_2$	<i>s.e.</i> ($\hat{\mu}_2$)	$\hat{\mu}_3$	<i>s.e.</i> ($\hat{\mu}_3$)
1)	12.812	0.731	1.536	0.447	3.316	0.557	-0.0042	0.0017
2)	9.406	0.401	0.958	0.415	0.859*	0.452	0	-
3)	12.105	0.437	1.057*	0.504	2.117	0.515	0	-
4)	9.555	0.347	-0.706*	0.367	0.424*	0.425	0	-
5)	19.456	0.542	5.937	0.717	7.137	0.767	0	-
6)	19.880	0.680	0.692*	0.799	12.856	0.953	0	-
7)	8.019	0.336	0.856	0.341	1.759	0.375	0	-
8)	9.509	0.404	1.615	0.408	2.031	0.431	0	-
9)	27.172	2.337	2.517*	1.583	10.567	1.506	-0.082	0.028
10)	19.294	1.151	4.450	1.141	10.071	1.422	0	-
11)	13.389	0.666	2.397	0.713	6.843	0.960	0	-
12)	12.240	0.626	2.082	0.685	7.075	0.833	0	-
13)	31.309	1.016	-0.255*	1.274	15.113	1.271	0	-
14)	7.201	1.122	2.354	0.742	6.878	0.999	0.023	0.004
15)	27.168	1.418	10.713	1.376	14.335	1.950	0	-
16)	22.343	0.702	4.167	0.914	10.124	0.971	0	-
17)	14.13	0.500	2.397	0.537	9.583	0.735	0	-
18)	15.769	0.727	6.702	0.670	13.133	0.879	0	-
19)	20.579	0.971	1.423*	1.079	6.176	1.134	0	-
20)	23.750	1.144	9.660	1.253	15.805	1.814	0	-
21)	19.204	1.037	7.073	0.941	12.568	1.162	0	-

Πίνακας 5.4

Εκτιμήσεις παραμέτρων σ_0 , σ_1 , σ_2 , ξ_0 κατανομής *GEV* για τις μέγιστες βροχοπτώσεις και τυπικά σφάλματα εκτιμητών.

	$\hat{\sigma}_0$	<i>s.e.</i> ($\hat{\sigma}_0$)	$\hat{\sigma}_1$	<i>s.e.</i> ($\hat{\sigma}_1$)	$\hat{\sigma}_2$	<i>s.e.</i> ($\hat{\sigma}_2$)	$\hat{\xi}_0$	<i>s.e.</i> ($\hat{\xi}_0$)
1)	2.255	0.036	0.012*	0.050	0.396*	0.049	0.22	0.035
2)	2.072	0.040	-0.013*	0.056	-0.011*	0.052	-0.20	0.043
3)	2.181	0.038	-0.202*	0.051	-0.034*	0.051	0.143	0.034
4)	2.00	0.027	-0.256*	0.053	-0.110*	0.051	0.158	0.034
5)	2.464	0.032	0.078*	0.047	0.343	0.046	0	-
6)	2.684	0.035	-0.298*	0.052	0.118*	0.361	0.130	0.030
7)	2.110	0.039	-0.147*	0.052	0.147*	0.052	0.281	0.037
8)	2.119	0.033	-0.064*	0.052	0.215*	0.053	0.294	0.040
9)	2.556	0.066	-0.123*	0.095	0.247*	0.104	0	-
10)	2.456	0.080	-0.026*	0.046	0.411*	0.053	0.254	0.080
11)	2.270	0.051	-0.075*	0.079	0.471*	0.089	0.188	0.049
12)	2.195	0.054	0.107*	0.057	0.662*	0.062	0.182	0.050
13)	2.517	0.063	-0.366*	0.087	0.209*	0.094	0.172	0.055
14)	2.357	0.030	0.173*	0.061	0.751*	0.068	0.116	0.047
15)	2.652	0.079	-0.508*	0.119	-0.025*	0.398	0.219	0.072
16)	2.364	0.045	-0.148	0.069	0.346	0.063	0	-
17)	2.347	0.036	-0.093*	0.05	0.629*	0.058	0.149	0.032
18)	2.044	0.061	0.024*	0.046	0.426*	0.058	0.298	0.069
19)	2.597	0.031	-0.235*	0.071	0.004*	0.086	0.145	0.058
20)	2.452	0.067	0.278*	0.097	0.783*	0.105	0.32	0.064
21)	2.255	0.036	-0.083*	0.050	0.367*	0.058	0.220	0.035

Στους 3 μετεωρολογικούς σταθμούς (Αλεξανδρούπολη, Σκύρος, Αίγιο) όπου η ύπαρξη τάσης είναι στατιστικά σημαντική σε επίπεδο σημαντικότητας 5%, η παράμετρος της τάσης δεν λήφθηκε υπόψη κατά τους υπολογισμούς για την κατασκευή των γραφημάτων για την παράμετρο θέσης, της μέσης μέγιστης μηνιαίας βροχόπτωσης, και των σταθμών απόδοσης λόγω της πολύ μικρής τιμής της. Θεωρήθηκε αμελητέα για να μπορούν να παραχθούν τα γραφήματα και οι στάθμες απόδοσης.

Στο σχήμα 5.3 απεικονίζονται οι σημειακές εκτιμήσεις της παραμέτρου θέσεως $\mu(i)$ για κάθε μήνα του έτους μαζί με τα προσεγγιστικά 95% διαστήματα εμπιστοσύνης που προκύπτουν μέσω της ασυμπτωτικής κανονικότητας που υποθέτουμε για τις παραμέτρους του μοντέλου. Για την ερμηνεία του γραφήματος θα πρέπει να θυμηθούμε ότι τα μέγιστα των κλάσεων παρουσιάζουν περιοδικότητα, επομένως δεν ακολουθούν όλα τα μέγιστα την ίδια κατανομή ακροτάτων. Έτσι για παράδειγμα, οι μέγιστες βροχοπτώσεις τον Ιανουάριο έχουν παράμετρο θέσης ίση με

$$\hat{\mu}(1) = \hat{\mu}_0 + \hat{\mu}_1 \sin\left(\frac{2\pi \cdot 1}{12}\right) + \hat{\mu}_2 \cos\left(\frac{2\pi \cdot 1}{12}\right)$$

η οποία είναι μία σταθερή τιμή για όλες τις παρατηρήσεις που ανήκουν στον μήνα Ιανουάριο, και αντίστοιχα ισχύουν και για τους άλλους μήνες του έτους. Για να γίνουμε πιο κατανοητοί, έχουμε π.χ. για το μετεωρολογικό σταθμό των Ιωαννίνων ότι η σημειακή εκτίμηση της παραμέτρου θέσεως για τον μήνα Ιανουάριο είναι ίση με 28.6 και το 95% δ.ε. ίσο με [26.411 , 30.792], για τον μήνα Φεβρουάριο η σημειακή εκτίμηση είναι ίση με 28.1 και το 95% δ.ε. ίσο με [26.102 , 30.231] κτλ. Αντίστοιχη ερμηνεία ισχύει και για τους άλλους μήνες του έτους και για τους υπόλοιπους μετεωρολογικούς σταθμούς. Στο γράφημα η συνεχής γραμμή απεικονίζει την μηνιαία σημειακή εκτίμηση του $\mu(i)$ που δίνεται από την συνάρτηση \square

$$\hat{\mu}(i) = \hat{\mu}_0 + \hat{\mu}_1 \sin\left(\frac{2\pi i}{12}\right) + \hat{\mu}_2 \cos\left(\frac{2\pi i}{12}\right), \quad i = 1, 2, \dots, 12.$$

ενώ οι ασυνεχείς γραμμές αναφέρονται στο άνω και κάτω όριο του δ.ε. μέσω της κανονικής προσέγγισης που δίνεται από τον τύπο $\square \hat{\mu}(i) \pm z_{\alpha/2} \cdot \sqrt{\text{Var}(\hat{\mu}(i))}$ για $i = 1, 2, \dots$,

12 όπου α το άνω ποσοστιαίο σημείο της κανονικής κατανομής $N(0,1)$. Η διακύμανση του $\hat{\mu}(i)$ δίνεται από τον ακόλουθο τύπο \square

$$\hat{V}ar(\hat{\mu}(i)) = \hat{V}ar(\hat{\mu}_0 + \hat{\mu}_1 \sin(\frac{2\pi i}{12}) + \hat{\mu}_2 \cos(\frac{2\pi i}{12})), \quad i = 1, 2, \dots, 12$$

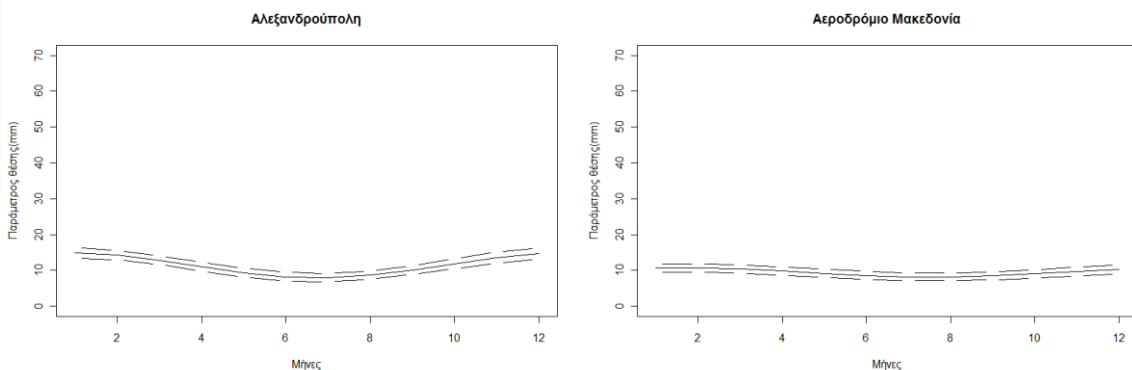
όπου ο παραπάνω τύπος αναλύεται ως εξής (το i παίρνει διαδοχικά τις τιμές 1, 2, ..., 12) \square

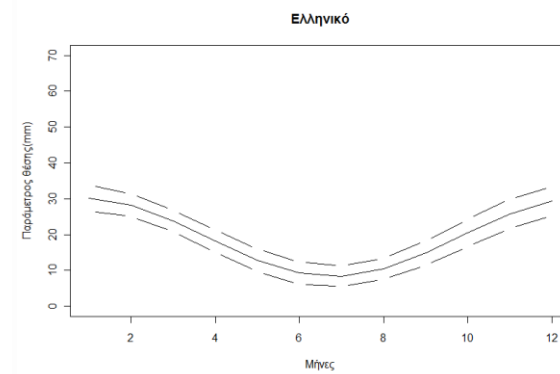
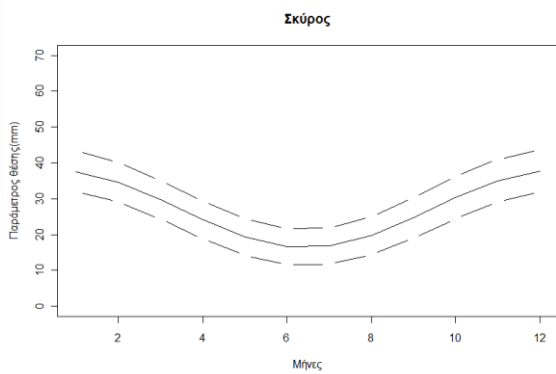
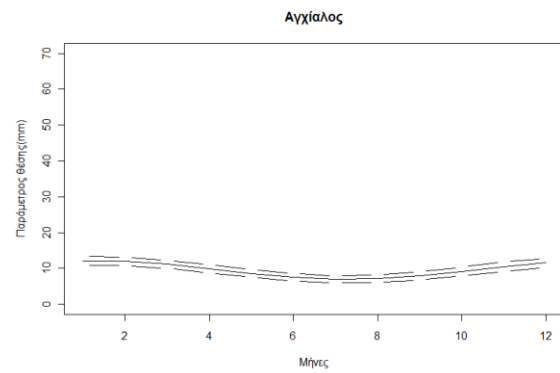
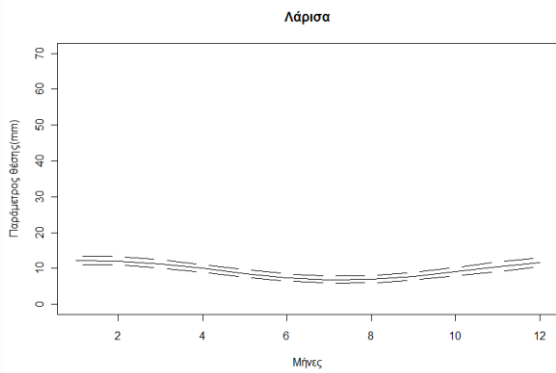
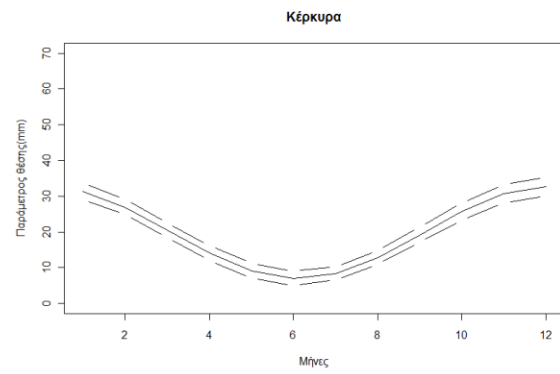
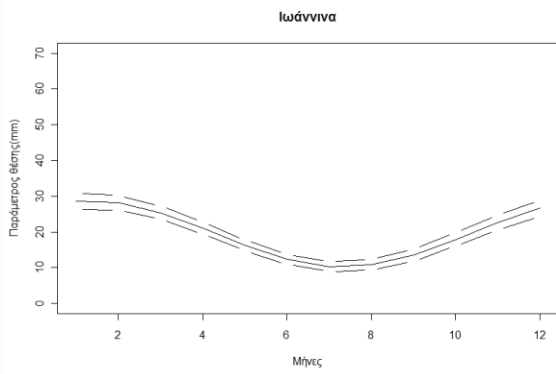
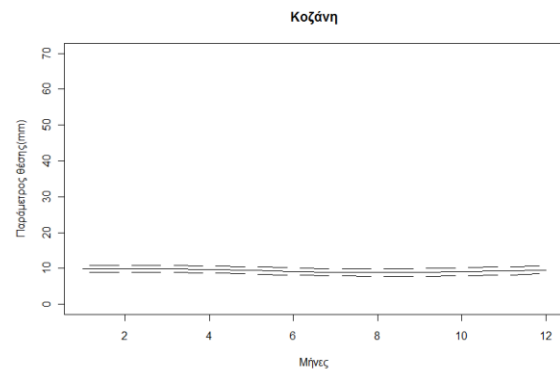
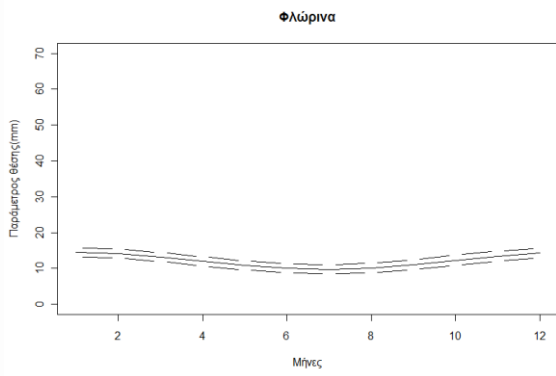
$$\begin{aligned} Var(\hat{\mu}(i)) = & Var(\hat{\mu}_0) + Var(\hat{\mu}_1) \cdot \sin^2(\frac{2\pi i}{12}) + Var(\hat{\mu}_2) \cdot \cos^2(\frac{2\pi i}{12}) + \\ & + 2 \cdot \sin(\frac{2\pi i}{12}) \cdot Cov(\hat{\mu}_0, \hat{\mu}_1) + 2 \cdot \cos(\frac{2\pi i}{12}) \cdot Cov(\hat{\mu}_0, \hat{\mu}_2) + 2 \cdot \sin(\frac{2\pi i}{12}) \cdot \cos(\frac{2\pi i}{12}) \cdot Cov(\hat{\mu}_1, \hat{\mu}_2) \end{aligned}$$

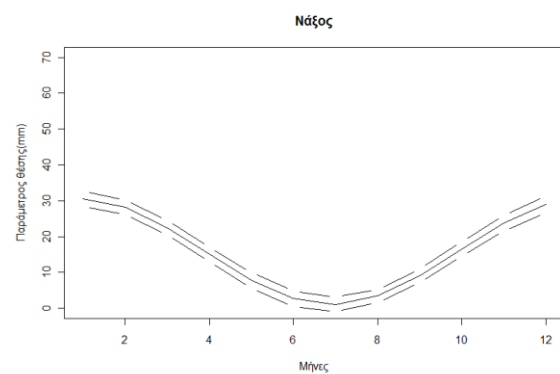
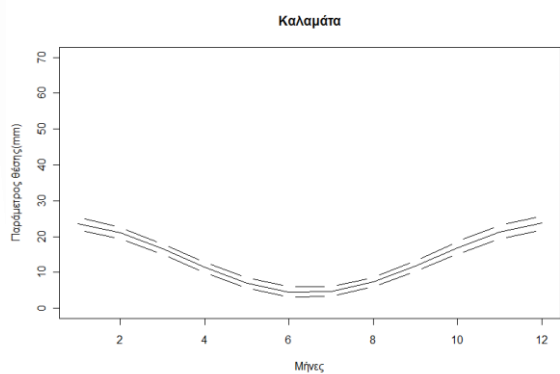
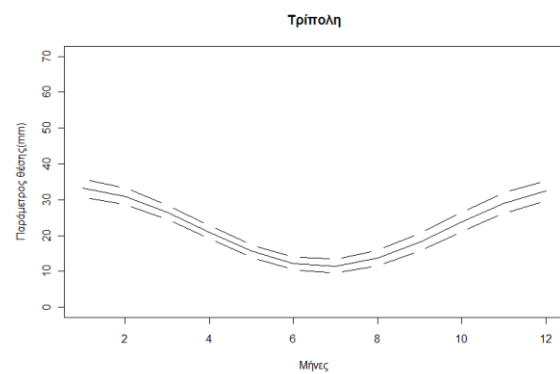
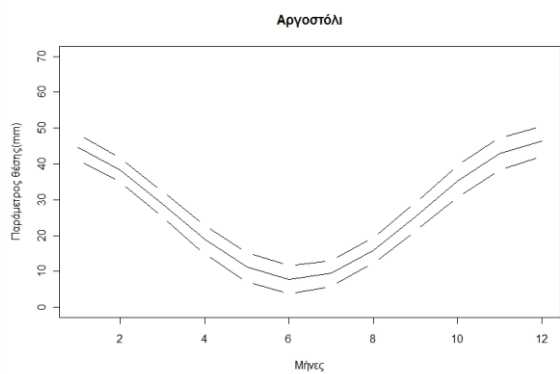
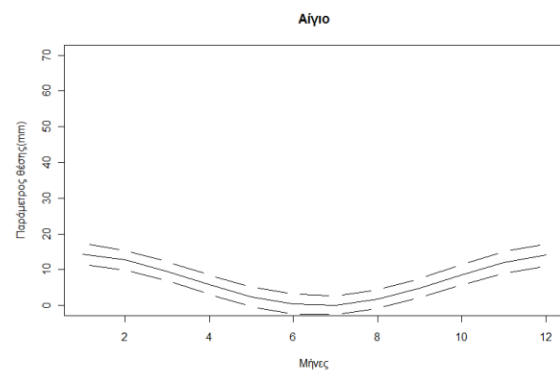
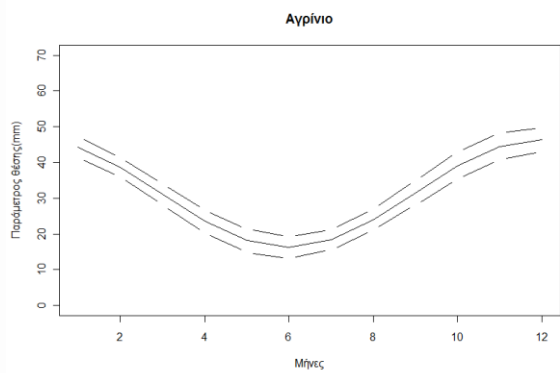
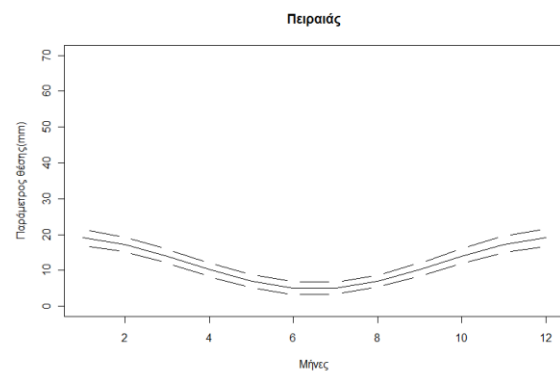
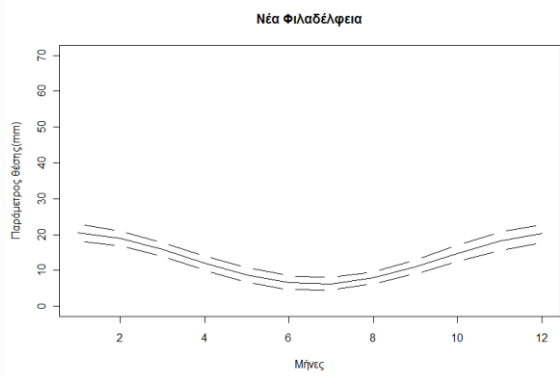
όπου η εκτίμηση για την διακύμανση του $\hat{\mu}_0$ αντιστοιχεί στο στοιχείο που βρίσκεται στην θέση (1,1) του εκτιμημένου πίνακα διασποράς \hat{I}^{-1} (όπου \hat{I} ο εκτιμημένος πίνακας πληροφορίας που έχει οριστεί στην ενότητα 2.2.3), για το $\hat{\mu}_1$ η εκτίμηση της διακύμανσης αντιστοιχεί στο στοιχείο (2,2) του \hat{I}^{-1} , και για το $\hat{\mu}_2$ αντιστοιχεί στο στοιχείο (3,3), και αντίστοιχα η εκτίμηση της συνδιακύμανσης των $\hat{\mu}_0, \hat{\mu}_1$ αντιστοιχεί στο στοιχείο (1,2) (ή αντίστοιχα στο στοιχείο (2,1)) του πίνακα, η εκτίμηση της συνδιακύμανσης των $\hat{\mu}_0, \hat{\mu}_2$ στο στοιχείο (1,3) και τέλος η εκτίμηση της συνδιακύμανσης των $\hat{\mu}_1, \hat{\mu}_2$ αντιστοιχεί στο στοιχείο (2,3) (Παράρτημα Κώδικας 2).

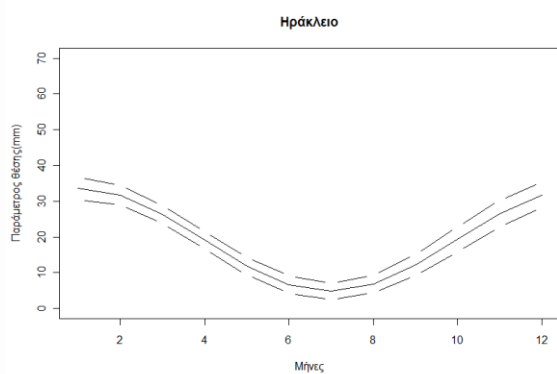
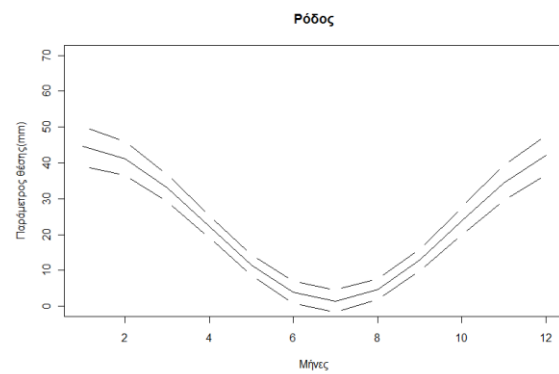
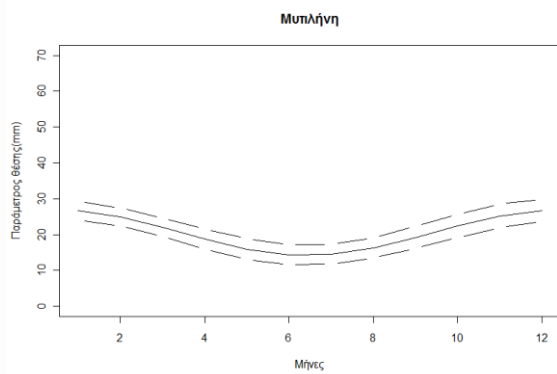
Σχήμα 5.3

Γράφημα της εκτιμώμενης παραμέτρου $\mu(i)$ για τις κατανομές των μέγιστων βροχοπτώσεων, και 95% δ.ε. για $i = 1, 2, \dots, 12$ για καθένα από τους 21 μετεωρολογικούς σταθμούς.









Στο σχήμα 5.4 απεικονίζονται οι σημειακές εκτιμήσεις της παραμέτρου κλίμακος για κάθε μήνα του έτους, καθώς και τα προσεγγιστικά δ.ε. που παρήχθησαν μέσω της ασυμπτωτικής κανονικότητας που υποθέτουμε για τους εκτιμητές, στους σταθμούς όπου, βάσει του δείγματος που είχαμε, η περιοδικότητα στην παράμετρο κλίμακος ήταν στατιστικά σημαντική σε επίπεδο σημαντικότητας 5%. Η ερμηνεία που δίνεται στα παρακάτω γραφήματα είναι ανάλογη της ερμηνείας που δώσαμε για τα γραφήματα της παραμέτρου θέσεως, δηλαδή οι μέγιστες βροχοπτώσεις που ανήκουν σε διαφορετικούς μήνες του έτους έχουν διαφορετική παράμετρο κλίμακος, και όλες οι παρατηρήσεις που ανήκουν στον συγκεκριμένο μήνα έχουν την ίδια παράμετρο κλίμακος. Έτσι π.χ. για τον μετεωρολογικό σταθμό των Ιωαννίνων, έχουμε ότι η σημειακή εκτίμηση για τον μήνα Ιανουάριο για την παράμετρο κλίμακος είναι ίση με 16.461 και το 95% δ.ε. ίσο με [14.692 , 18.234] για το μήνα Φεβρουάριο η σημειακή εκτίμηση είναι ίση με 14.941 και το 95% δ.ε. ίσο με [13.321, 16.561]. Αντίστοιχη ερμηνεία δίνεται και για τους άλλους μήνες του έτους και για τους άλλους μετεωρολογικούς σταθμούς. Στο γράφημα που ακολουθεί η συνεχής γραμμή απεικονίζει την μηνιαία σημειακή εκτίμηση που δίνεται από την συνάρτηση \square

$$\hat{\sigma}(i) = \exp\left(\hat{\sigma}_0 + \hat{\sigma}_1 \sin\left(\frac{2\pi i}{12}\right) + \hat{\sigma}_2 \cos\left(\frac{2\pi i}{12}\right)\right) \quad i = 1, 2, \dots, 12.$$

και η ασυνεχής γραμμή το άνω και κάτω όριο του δ.ε. που δίνεται μέσω της κανονικής προσέγγισης από τον τύπο $\hat{\sigma}(i) \pm z_{\alpha/2} \cdot \sqrt{\hat{Var}(\hat{\sigma}(i))}$ για $i = 1, 2, \dots, 12$. Για να υπολογίσουμε την διασπορά του εκτιμητή $\hat{\sigma}(i)$ θα χρησιμοποιήσουμε τη μέθοδο Δέλτα. Έτσι η διασπορά θα δίνεται από τον ακόλουθο τύπο \square

$$\hat{Var}(\hat{\sigma}(i)) = \left(\frac{\partial \hat{\sigma}(i)}{\partial \sigma_0}, \frac{\partial \hat{\sigma}(i)}{\partial \sigma_1}, \frac{\partial \hat{\sigma}(i)}{\partial \sigma_2} \right) \cdot \hat{I}'' \cdot \left(\frac{\partial \hat{\sigma}(i)}{\partial \sigma_0}, \frac{\partial \hat{\sigma}(i)}{\partial \sigma_1}, \frac{\partial \hat{\sigma}(i)}{\partial \sigma_2} \right)^T \quad i = 1, 2, \dots, 12.$$

όπου οι μερικές παράγωγοι της συνάρτησης $\hat{\sigma}(i)$ ως προς τις παραμέτρους $\hat{\sigma}_0, \hat{\sigma}_1, \hat{\sigma}_2$, δίνονται από τους παρακάτω τύπους \square

$$\frac{\partial \hat{\sigma}(i)}{\partial \sigma_0} = \exp(\hat{\sigma}_0 + \hat{\sigma}_1 \cdot \sin(\frac{2\pi i}{12}) + \hat{\sigma}_2 \cdot \cos(\frac{2\pi i}{12})),$$

$$\frac{\partial \hat{\sigma}(i)}{\partial \sigma_1} = \exp(\hat{\sigma}_0 + \hat{\sigma}_1 \cdot \sin(\frac{2\pi i}{12}) + \hat{\sigma}_2 \cdot \cos(\frac{2\pi i}{12})) \cdot \sin(\frac{2\pi i}{12}),$$

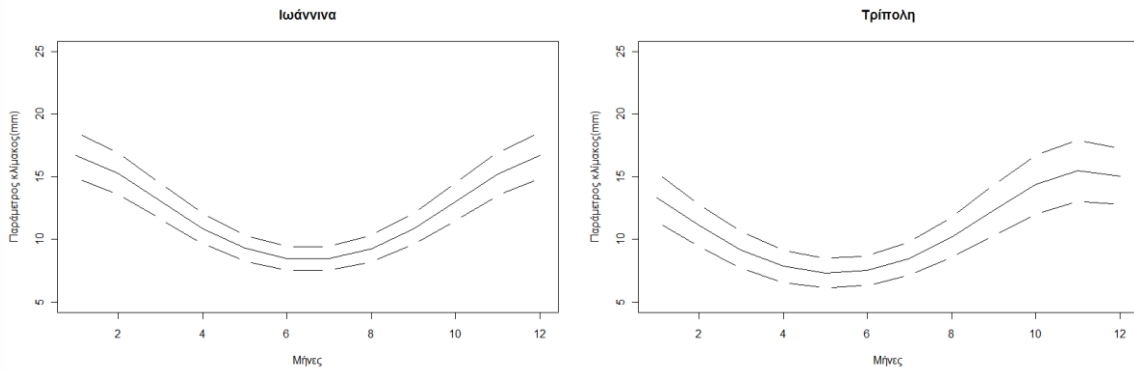
$$\frac{\partial \hat{\sigma}(i)}{\partial \sigma_2} = \exp(\hat{\sigma}_0 + \hat{\sigma}_1 \cdot \sin(\frac{2\pi i}{12}) + \hat{\sigma}_2 \cdot \cos(\frac{2\pi i}{12})) \cdot \cos(\frac{2\pi i}{12}).$$

Ο πίνακας \hat{I}'' , είναι υποπίνακας του εκτιμημένου πίνακα διασποράς \hat{I}^{-1} του αρχικού μοντέλου και περιέχει ως στοιχεία του τις (εκτιμημένες) διακυμάνσεις και τις συνδιακυμάνσεις των $\hat{\sigma}_0, \hat{\sigma}_1, \hat{\sigma}_2$ παρουσιάζεται παρακάτω (Παράρτημα Κώδικας 3).

$$\hat{I}'' = \begin{vmatrix} Var(\hat{\sigma}_0) & Cov(\hat{\sigma}_0, \hat{\sigma}_1) & Cov(\hat{\sigma}_0, \hat{\sigma}_2) \\ Cov(\hat{\sigma}_1, \hat{\sigma}_0) & Var(\hat{\sigma}_1) & Cov(\hat{\sigma}_1, \hat{\sigma}_2) \\ Cov(\hat{\sigma}_2, \hat{\sigma}_0) & Cov(\hat{\sigma}_2, \hat{\sigma}_1) & Var(\hat{\sigma}_2) \end{vmatrix}$$

Σχήμα 5.4

Γράφημα της εκτιμώμενης παραμέτρου $\sigma(i)$ για τις κατανομές των μέγιστων βροχοπτώσεων, και 95% δ.ε. για $i = 1, 2, \dots, 12$ για τους 2 μετεωρολογικούς σταθμούς όπου η περιοδικότητα στην παράμετρο κλίμακος είναι στατιστικά σημαντική.



Ιδιαίτερο ενδιαφέρον παρουσιάζει και η εκτιμώμενη μέση μηνιαία μέγιστη βροχόπτωση, η οποία αντιστοιχεί στην εκτιμημένη μέση τιμή της κατανομής *GEV* ή *Gumbel*. Έστω Z_1, Z_2, \dots, Z_k τα μέγιστα των κλάσεων τα οποία προέρχονται από την *GEV* ή την *Gumbel* έτσι ώστε $Z_i \sim GEV(\hat{\mu}(i), \hat{\sigma}(i), \hat{\xi}_i)$ τότε η εκτιμημένη μέση τιμή της *GEV* θα δίνεται από την ακόλουθη έκφραση \square

$$E(Z_i) = \hat{\mu}(i) + \hat{\sigma}(i) \left(\frac{\Gamma(1 - \frac{\hat{\xi}_i}{\hat{\sigma}(i)}) - 1}{\hat{\xi}_i} \right) \quad i = 1, 2, \dots, 12$$

όπου με Γ συμβολίζουμε την συνάρτηση Γάμμα που ορίζεται από τον τύπο \square

$$\Gamma(a) = \int_0^{+\infty} e^{-x} x^{a-1} dx$$

για $a \in \mathbb{R}_+$. Αντίστοιχα για την *Gumbel* κατανομή έχουμε ότι η εκτιμημένη μέση της τιμή δίνεται από τον τύπο \square

$$E(Z_i) = \hat{\mu}(i) + \gamma \hat{\sigma}(i) \quad i = 1, 2, \dots, 12$$

όπου γ είναι η σταθερά του Euler και ισούται περίπου με 0.5722. Για την ερμηνεία των γραφημάτων που παραθέτουμε στη συνέχεια καθώς και του πίνακα με τις εκτιμώμενες μέσες τιμές των κατανομών για κάθε μήνα του έτους και για κάθε μετεωρολογικό σταθμό θα πρέπει

να θυμηθούμε ότι το μέγεθος των κλάσεων των μέγιστων διαφέρει από σταθμό σε σταθμό επομένως θα διαφέρει και η ερμηνεία. Έτσι για τους σταθμούς όπου το μέγεθος των κλάσεων είναι ίσο με 30 παρατηρήσεις, η εκτιμώμενη μέση τιμή της κατανομής μας δίνει την αναμενόμενη μηνιαία μέγιστη ημερήσια βροχόπτωση για το μήνα αυτό. Αντίθετα για τους σταθμούς όπου το μέγεθος των κλάσεων είναι μεγαλύτερο από 30 παρατηρήσεις, δηλαδή είναι 60 90 ή 120 παρατηρήσεις, η ερμηνεία της εκτιμώμενης μέσης τιμής δεν είναι ίδια. Η ερμηνεία θα γίνει ευκολότερα κατανοητή με την παράθεση παραδειγμάτων από σταθμούς με διαφορετικό μέγεθος κλάσης. Έτσι για το μετεωρολογικό σταθμό των Ιωαννίνων (που αντιστοιχεί στην 6^η γραμμή του πίνακα 5.5) όπου το μέγεθος των κλάσεων είναι ίσο με 30 παρατηρήσεις, η αναμενόμενη μέγιστη ημερήσια βροχόπτωση για τον μήνα Ιανουάριο είναι ίση με 38.027 χιλιοστά, για τον μήνα Φεβρουάριο είναι ίση με 36.717 κοκ. Από την άλλη μεριά για το μετεωρολογικό σταθμό της Σκύρου (που αντιστοιχεί στην 9^η γραμμή του πίνακα 5.5) όπου το μέγεθος των κλάσεων είναι ίσο με 120 παρατηρήσεις, η εκτιμώμενη μέση τιμή που αντιστοιχεί στον Ιανουάριο και είναι ίση με 44.961 χιλιοστά, είναι η αναμενόμενη μέγιστη ημερήσια βροχόπτωση που αναμένεται να συμβαίνει ανά 4 Ιανουαρίους, για το μήνα Φεβρουάριο είναι ίση με 42.015 χιλιοστά και αναμένεται να συμβαίνει ανά 4 Φεβρουάριους κοκ. Ακολουθεί ο πίνακας 5.5 με τις εκτιμήσεις των αναμενόμενων μέγιστων βροχοπτώσεων, καθώς και τα αντίστοιχα γραφήματα με τις σημειακές εκτιμήσεις.

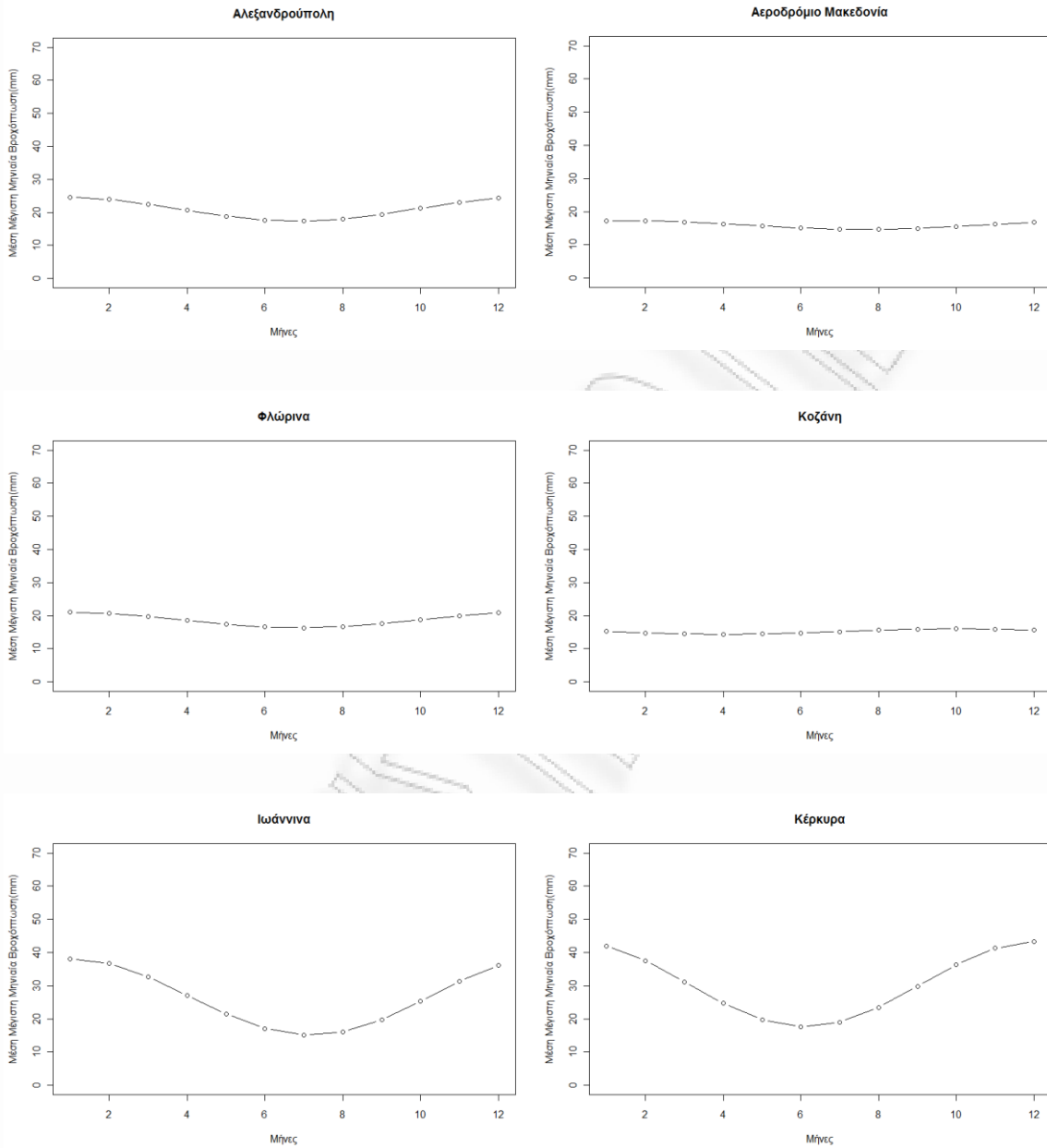
Πίνακας 5.5

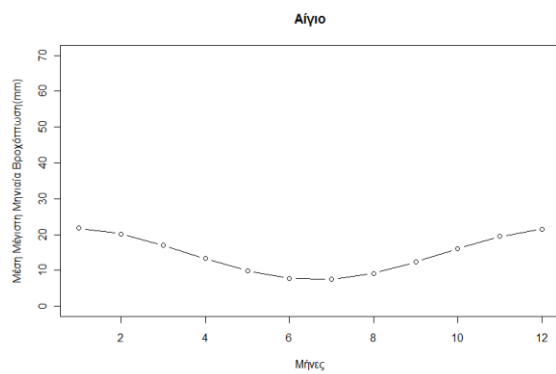
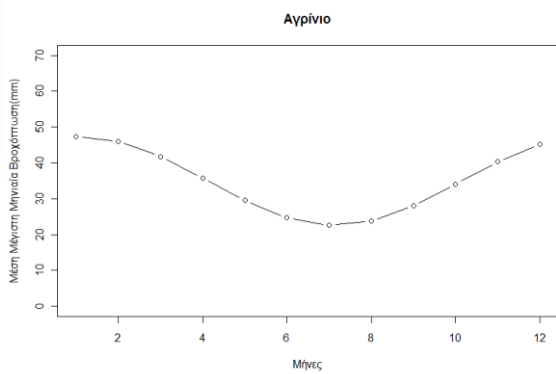
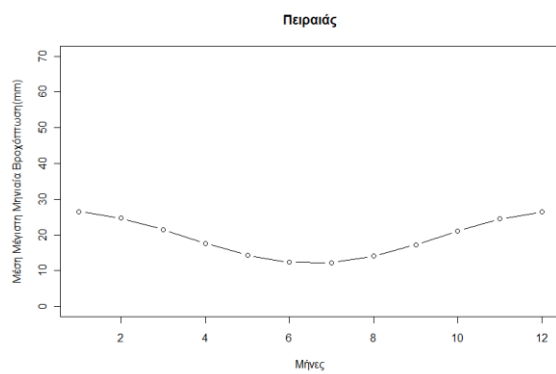
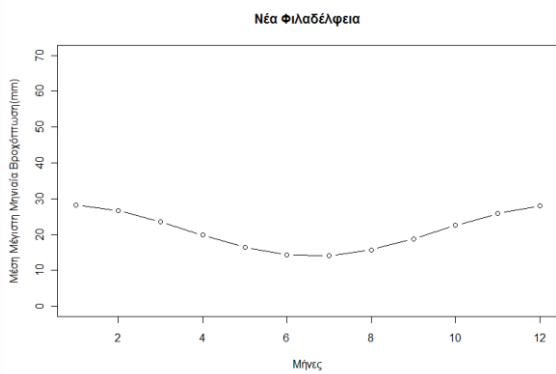
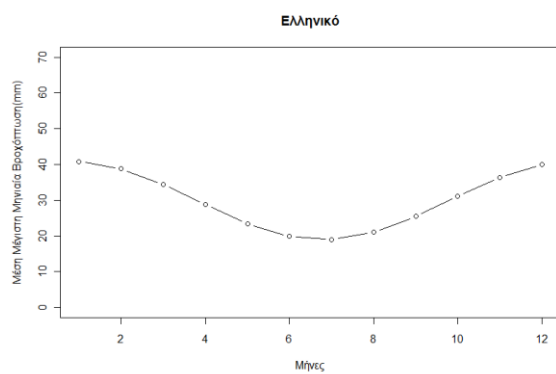
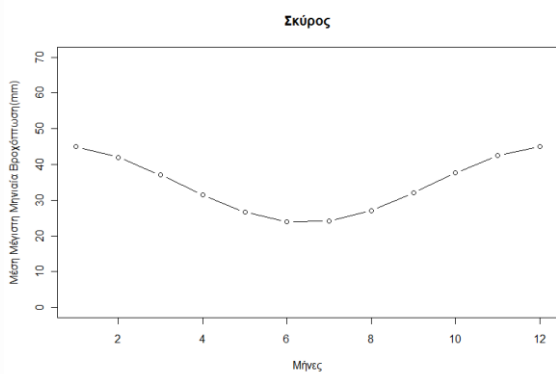
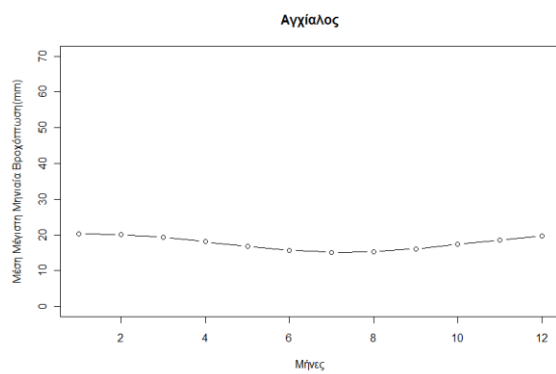
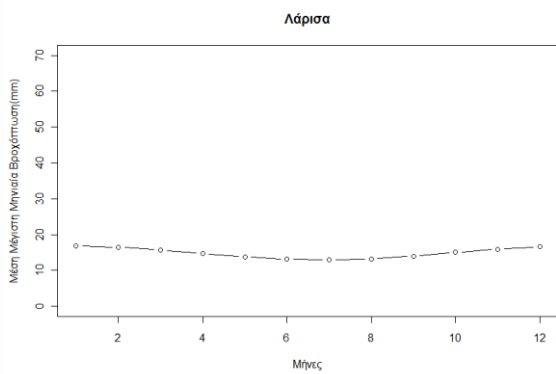
Πίνακας με τις εκτιμήσεις των αναμενόμενων μέγιστων βροχοπτώσεων για κάθε μήνα του έτους και για κάθε μετεωρολογικό σταθμό.

	Ιαν	Φεβ	Μαρ	Απρ	Μαι	Ιουν	Ιουλ	Αυγ	Σεπ	Οκτ	Νοεμ	Δεκ
1)	24.588	23.937	22.484	20.620	18.843	17.630	17.306	17.958	19.410	21.274	23.051	24.264
2)	17.160	17.196	16.895	16.337	15.671	15.077	14.713	14.676	14.978	15.536	16.202	16.796
3)	21.029	20.641	19.724	18.524	17.362	16.549	16.305	16.693	17.610	18.810	19.972	20.784
4)	15.55	15.648	15.603	15.426	15.165	14.890	14.674	14.576	14.621	14.798	15.059	15.334
5)	38.027	36.717	32.671	27.094	21.440	17.091	15.111	16.039	19.739	25.346	31.378	36.078
6)	41.982	37.530	31.195	24.673	19.713	17.642	19.017	23.468	29.804	36.325	41.286	43.356
7)	16.805	16.475	15.711	14.716	13.758	13.094	12.901	13.230	13.995	14.990	15.947	16.612
8)	20.267	20.115	19.316	18.084	16.749	15.669	15.134	15.286	16.085	17.317	18.652	19.732
9)	44.961	42.015	37.069	31.448	26.659	23.984	24.142	27.088	32.034	37.655	42.444	45.119
10)	40.854	38.797	34.358	28.726	23.410	19.836	18.960	21.017	25.457	31.089	36.404	39.978
11)	28.320	26.693	23.590	19.843	16.456	14.336	14.053	15.680	18.783	22.530	25.917	28.037
12)	26.555	24.728	21.469	17.653	14.301	12.311	12.218	14.045	17.304	21.120	24.472	26.461
13)	53.959	48.334	40.743	33.220	27.781	25.884	28.036	33.662	41.253	48.776	54.214	56.112
14)	21.801	20.145	17.021	13.267	9.888	7.789	7.533	9.189	12.312	16.067	19.446	21.545
15)	57.202	50.776	41.352	31.456	23.738	20.268	21.975	28.401	37.825	47.721	55.438	58.908
16)	40.823	37.377	31.755	25.388	19.843	16.521	16.346	19.492	25.237	32.028	37.878	41.075
17)	42.595	38.456	31.990	24.930	19.167	16.247	16.950	21.089	27.555	34.615	40.378	43.299
18)	64.017	61.663	55.994	48.529	41.269	36.158	34.566	36.920	42.589	50.054	57.314	62.425
19)	52.255	50.006	43.397	34.200	24.879	17.932	15.219	17.468	24.076	33.273	42.594	49.542
20)	58.995	55.696	47.505	36.614	25.943	18.352	15.873	19.171	27.363	38.253	48.924	56.516
21)	44.992	42.487	37.151	29.918	22.729	17.508	15.655	17.667	23.004	30.236	37.425	42.646

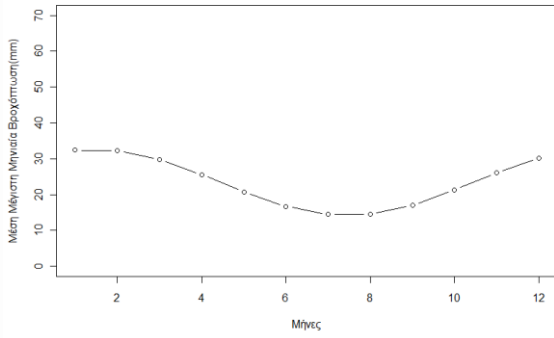
Σχήμα 5.5

Γράφημα της εκτιμώμενης μέσης μηνιαίας μέγιστης βροχόπτωσης για $i = 1, 2, \dots, 12$ για καθένα από τους 21 μετεωρολογικούς σταθμούς.

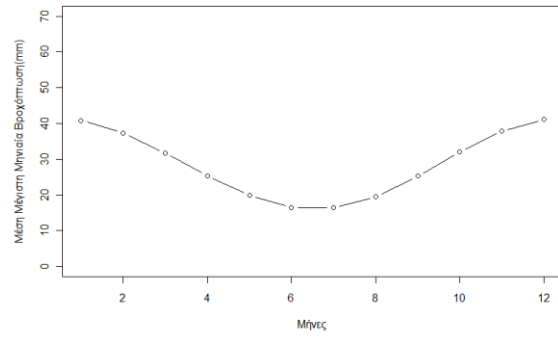




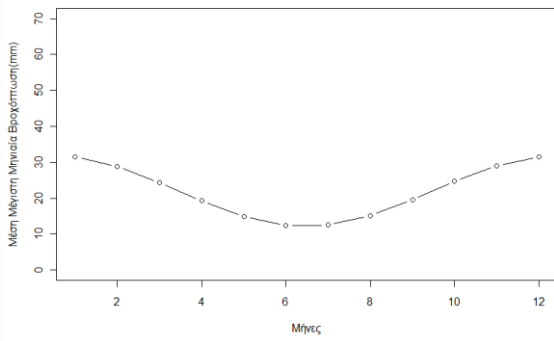
Αργιστόλι



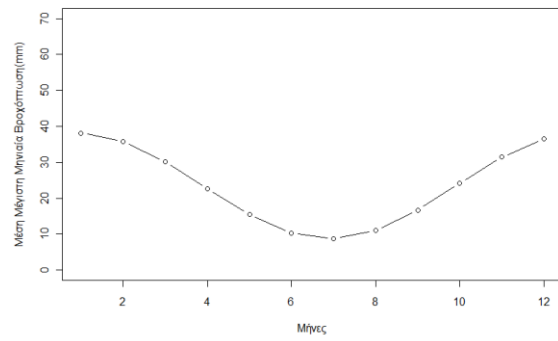
Τρίπολη



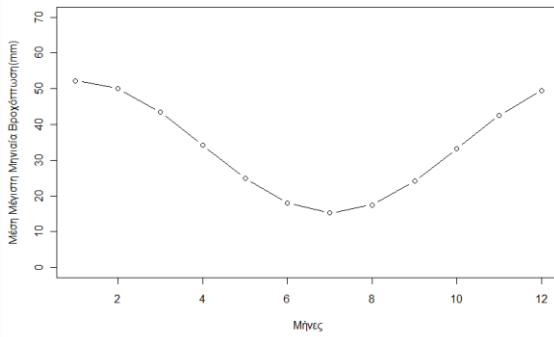
Καλαμάτα



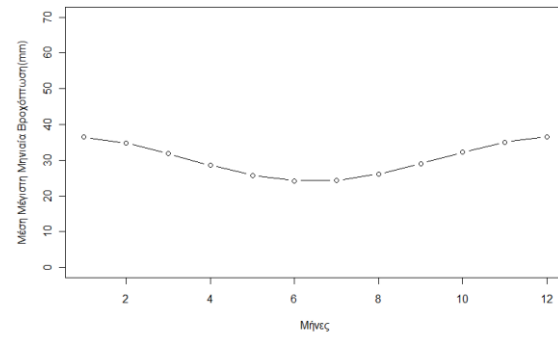
Νάξος



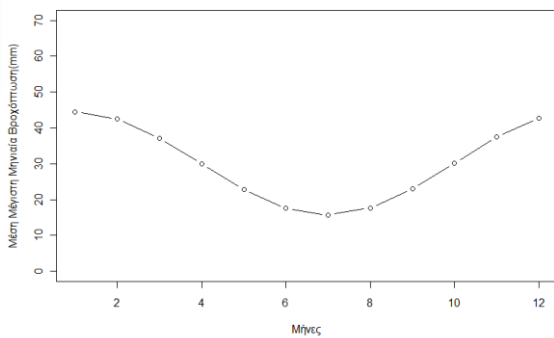
Μυτιλήνη



Ρόδος



Ηράκλειο



Χρησιμοποιώντας τις εκτιμήσεις των παραμέτρων από τους πίνακες 5.3 και 5.4 και τα όσα αναπτύχθηκαν στις παραγράφους 2.2.3 και 3.3.5, όσον αφορά την εκτίμηση της στάθμης απόδοσης καθώς και την κατασκευή προσεγγιστικών δ.ε. μέσω της ασυμπτωτικής κανονικότητας, υπολογίσαμε την σημειακή εκτίμηση της στάθμης απόδοσης και τα δ.ε. και κατασκευάσαμε τα γραφήματα που παρουσιάζονται στα σχήματα 5.6 και 5.7 και αφορούν την μηνιαία στάθμη απόδοσης για 10 και 50 χρόνια. Εδώ θα πρέπει να αναφέρουμε ότι για να αντιστοιχούν όλες οι στάθμες απόδοσης, από όλους τους μετεωρολογικούς σταθμούς, στην ίδια χρονική περίοδο, δηλαδή σε 10 και 50 έτη, πήραμε διαφορετική πιθανότητα υπέρβασης p για μετεωρολογικούς σταθμούς με διαφορετικό μέγεθος κλάσης, αφού εξ ορισμού το μέγεθος της κλάσης είναι αλληλένδετο με τον χρονικό ορίζοντα της στάθμης απόδοσης. Έτσι π.χ. για να έχουμε στάθμη απόδοσης 10 έτων σε όλους τους μετεωρολογικούς σταθμούς, πήραμε πιθανότητα υπέρβασης $p = 0.1$ για μετεωρολογικούς σταθμούς με μέγεθος κλάσης ίσο με 30 παρατηρήσεις, $p = 0.05$ για σταθμούς με μέγεθος κλάσης ίσο με 60 παρατηρήσεις, $p = 0.0333$ για σταθμούς με μέγεθος κλάσης ίσο με 90 παρατηρήσεις και $p = 0.025$ για σταθμούς με μέγεθος κλάσης ίσο με 120 παρατηρήσεις. Σε κάθε μήνα του έτους αντιστοιχίζουμε μια στάθμη απόδοσης z_p , ένα κατάφλι το οποίο οι μέγιστες ημερήσιες βροχοπτώσεις του συγκεκριμένου μήνα θα ξεπερνούν κατά μέσο όρο μία φορά κάθε 10 ή 50 χρόνια. Επομένως η συνεχής γραμμή στα παρακάτω γραφήματα απεικονίζει την σημειακή εκτίμηση που δίνεται από τον τύπο (για $i = 1, 2, \dots, 12$) :

$$\hat{z}_p(i) = h(\hat{\mu}(i), \hat{\sigma}(i), \hat{\xi}(i)) = \begin{cases} \hat{\mu}(i) - \frac{\hat{\sigma}(i)}{\hat{\xi}(i)}(1 - y_p^{-\hat{\xi}(i)}), & \hat{\xi}(i) \neq 0 \\ \hat{\mu}(i) - \hat{\sigma}(i) \ln y_p, & \hat{\xi}(i) = 0 \end{cases}$$

$$\mu(i) = \mu_0 + \mu_1 \sin\left(\frac{2\pi i}{12}\right) + \mu_2 \cos\left(\frac{2\pi i}{12}\right)$$

$$\sigma(i) = \exp\left(\sigma_0 + \sigma_1 \sin\left(\frac{2\pi i}{12}\right) + \sigma_2 \cos\left(\frac{2\pi i}{12}\right)\right)$$

$$\xi(i) = \xi_0$$

όπου $y_p = -\ln(1-p)$, και η ασυνεχής γραμμή το άνω και κάτω όριο του δ.ε. που δίνεται μέσω της κανονικής προσέγγισης από τον τύπο : $\hat{z}_p(i) \pm z_{\alpha/2} \cdot \sqrt{\text{Var}(\hat{z}_p(i))}$ για $i = 1, 2, \dots, 12$,

όπου a το άνω ποσοστιαίο σημείο της κανονικής κατανομής $N(0,1)$. Η διακύμανση του $\hat{z}_p(i)$ δίνεται από τον ακόλουθο τύπο \square

$$\text{Var}(\hat{z}_p(i)) = \left(\frac{\partial h}{\partial \mu(i)}, \frac{\partial h}{\partial \sigma(i)}, \frac{\partial h}{\partial \xi(i)} \right) \cdot \hat{I}^{-1} \cdot \left(\frac{\partial h}{\partial \mu(i)}, \frac{\partial h}{\partial \sigma(i)}, \frac{\partial h}{\partial \xi(i)} \right)^T, \quad i = 1, 2, \dots, 12$$

όπου \hat{I}^{-1} ο εκτιμημένος πίνακας διασποράς. Οι μερικές παράγωγοι δίνονται από τους παρακάτω τύπους (το i παίρνει διαδοχικά τις τιμές 1, 2, ..., 12). \square

$$\frac{\partial h}{\partial \mu_0} = 1$$

$$\frac{\partial h}{\partial \mu_1} = \sin\left(\frac{2\pi i}{12}\right)$$

$$\frac{\partial h}{\partial \mu_2} = \cos\left(\frac{2\pi i}{12}\right)$$

$$\frac{\partial h}{\partial \sigma_0} = -\sigma(i) \cdot \xi^{-1} \cdot (1 - y_p^{-\xi})$$

$$\frac{\partial h}{\partial \sigma_1} = -\sigma(i) \cdot \xi^{-1} \cdot (1 - y_p^{-\xi}) \cdot \sin\left(\frac{2\pi i}{12}\right)$$

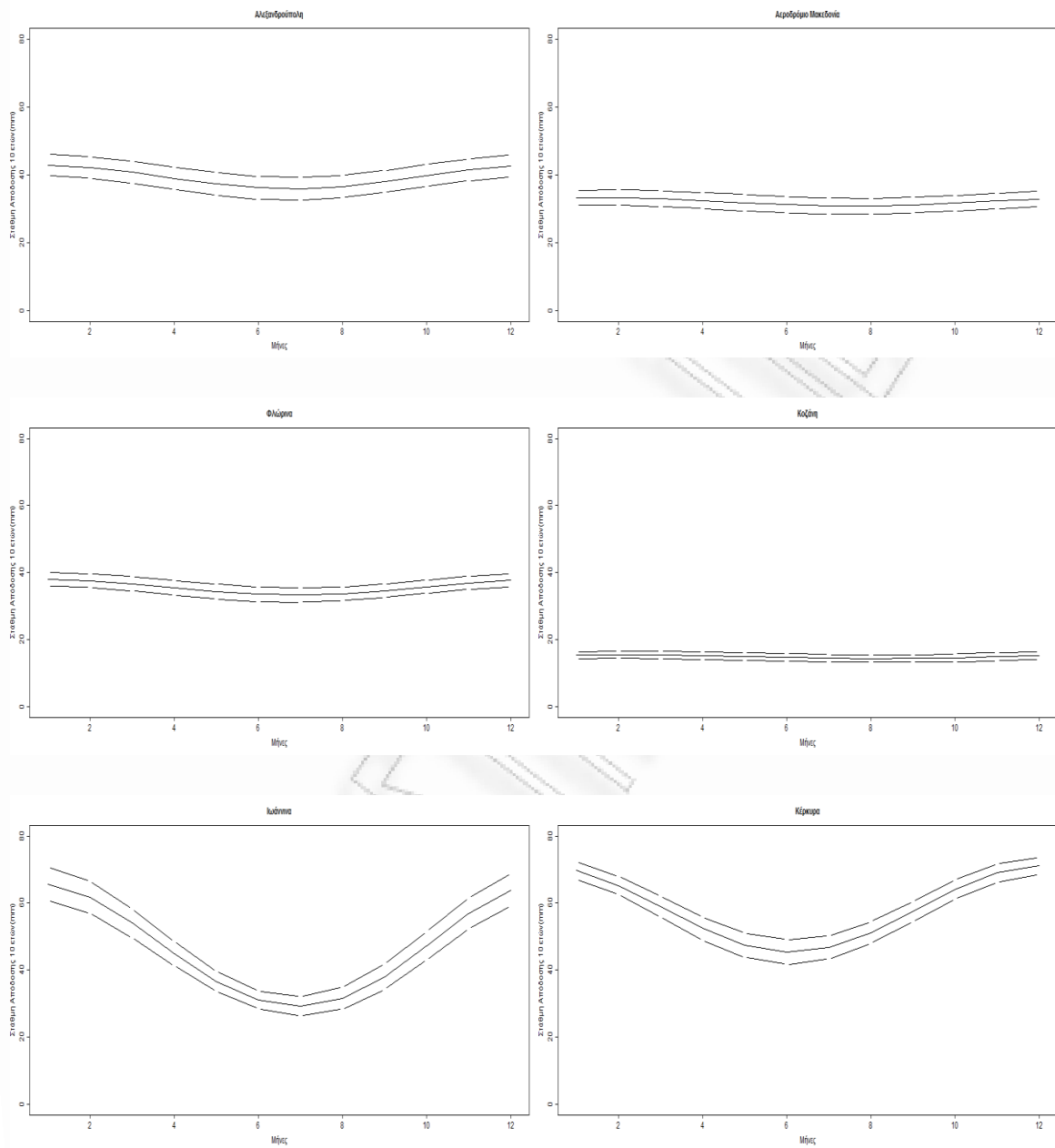
$$\frac{\partial h}{\partial \sigma_2} = -\sigma(i) \cdot \xi^{-1} \cdot (1 - y_p^{-\xi}) \cdot \cos\left(\frac{2\pi i}{12}\right)$$

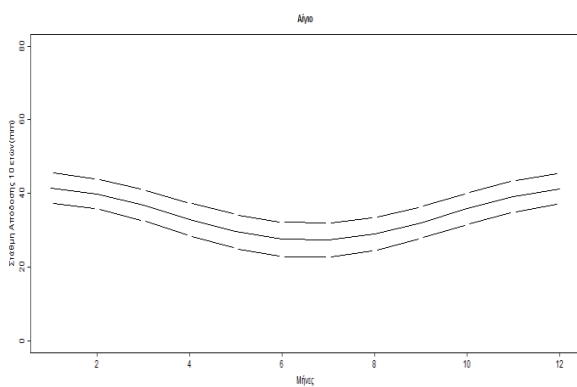
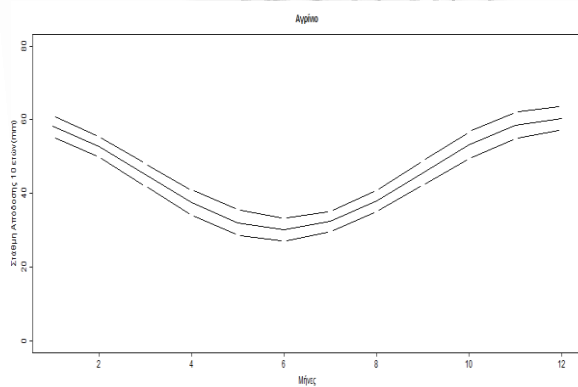
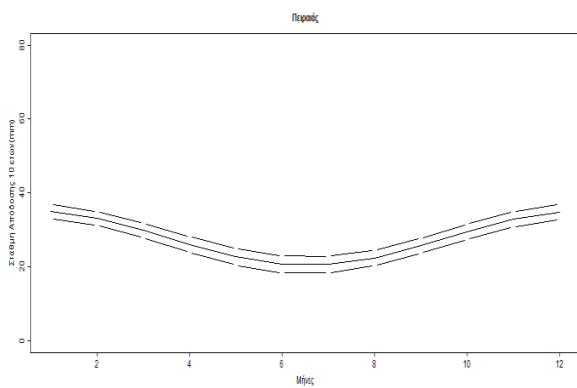
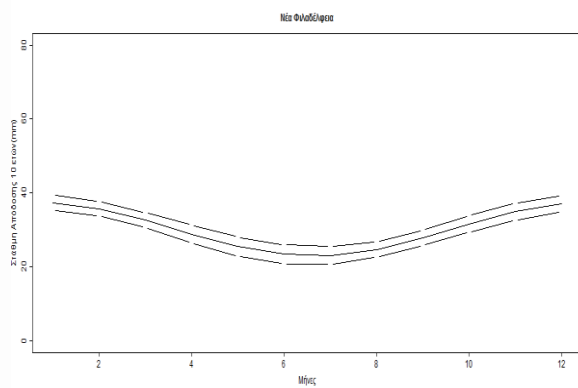
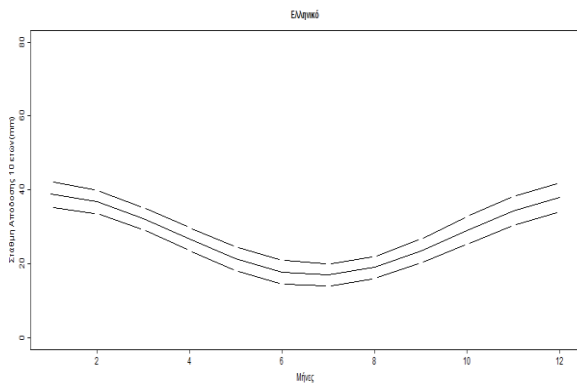
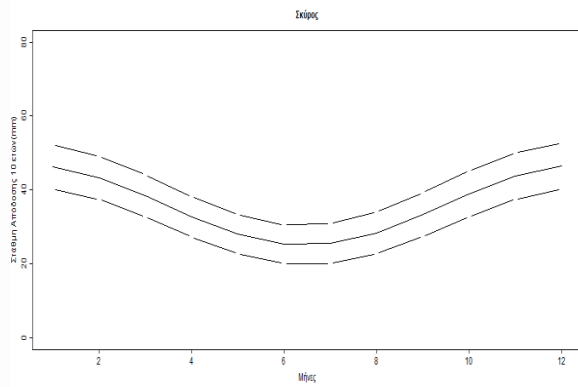
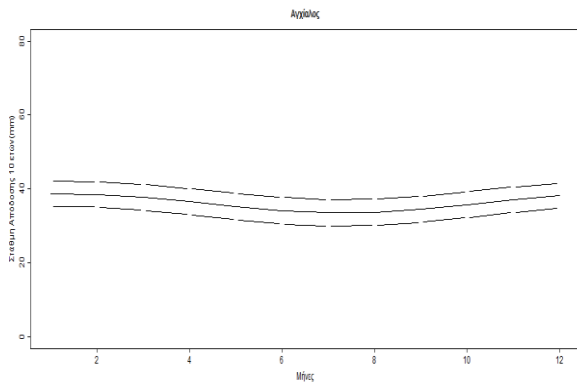
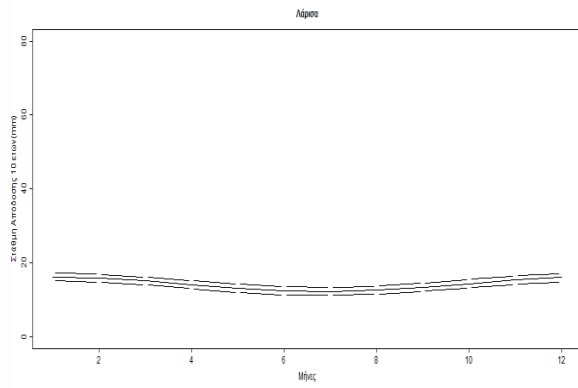
$$\frac{\partial h}{\partial \xi} = \sigma(i) \cdot \xi^{-2} (1 - y_p^{-\xi}) - \sigma(i) \cdot \xi^{-1} \cdot y_p^{-\xi} \cdot \ln y_p$$

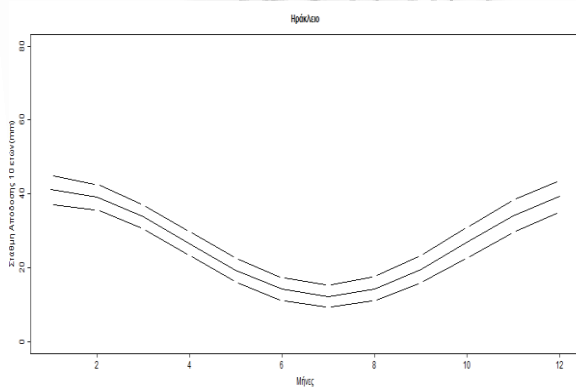
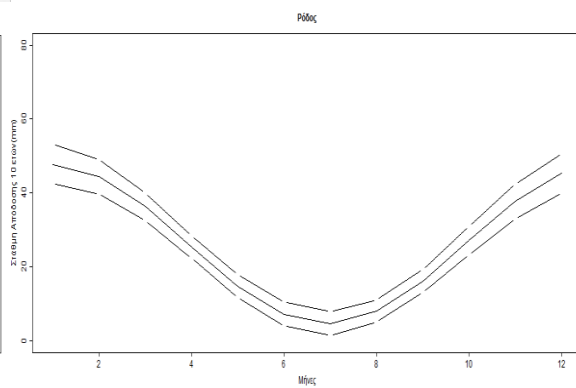
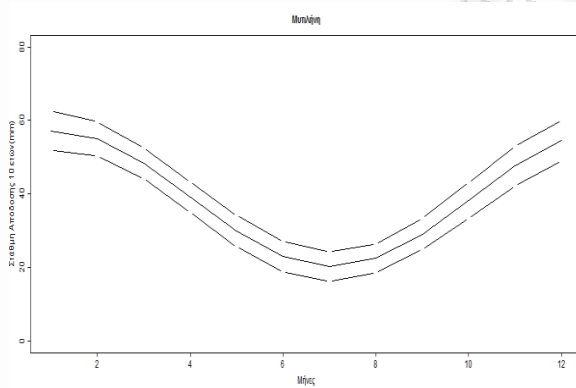
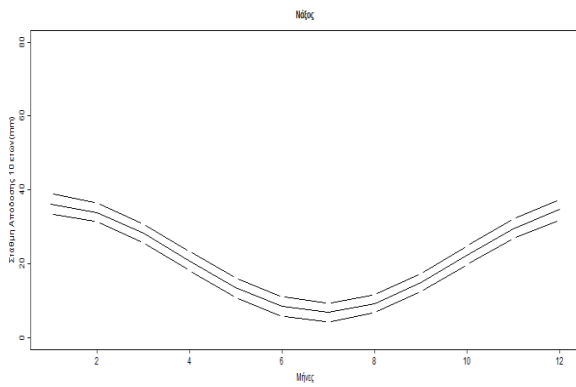
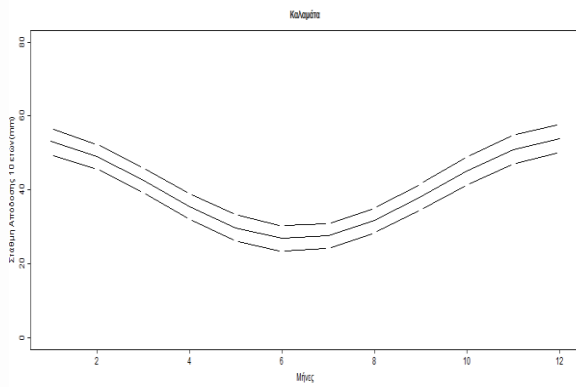
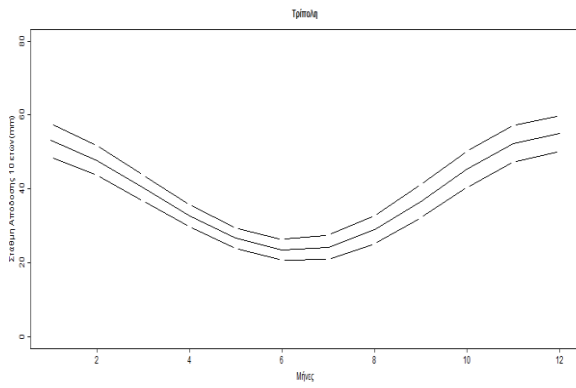
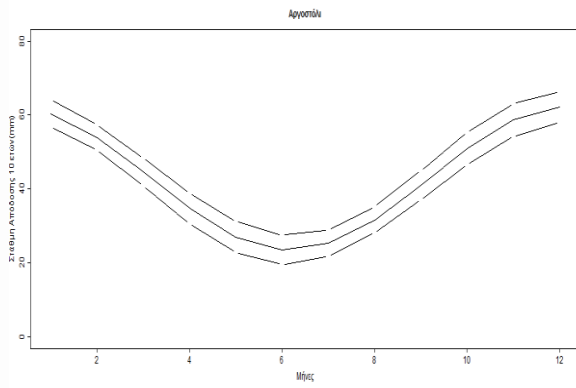
Έτσι πιο συγκεκριμένα για το μετεωρολογικό σταθμό των Ιωαννίνων, η σημειακή εκτίμηση της στάθμης απόδοσης για 10 χρόνια για τον μήνα Ιανουάριο είναι ίση με 65.656 χιλιοστά βροχόπτωσης, που σημαίνει ότι κατά μέσο όρο, η ημερήσια βροχόπτωση τον μήνα Ιανουάριο στα Ιωάννινα υπερβαίνει τα 65.656 χιλιοστά μία φορά κάθε 10 Ιανουαριους, για τον μήνα Φεβρουάριο έχουμε σημειακή εκτίμηση ίση με 61.792 χιλιοστά που σημαίνει ότι στα Ιωάννινα η ημερήσια βροχόπτωση υπερβαίνεται 61.792 χιλιοστά κατά μέσο όρο μία φορά κάθε δέκα Φεβρουάριους. Με αυτό τον τρόπο έχουμε τοπική εκτίμηση της στάθμης απόδοσης για κάθε μήνα του έτους. Αντίστοιχη ερμηνεία ισχύει και για την στάθμη απόδοσης για τους άλλους μήνες του έτους καθώς και για την στάθμη απόδοσης για 50 χρόνια σε όλους τους μετεωρολογικούς σταθμούς (Παράρτημα Κώδικας 4).

Σχήμα 5.6

Γράφημα μηνιαίας στάθμης απόδοσης και 95 % δ.ε. για τις μέγιστες βροχοπτώσεις 10 ετών για τους 21 μετεωρολογικούς σταθμούς.

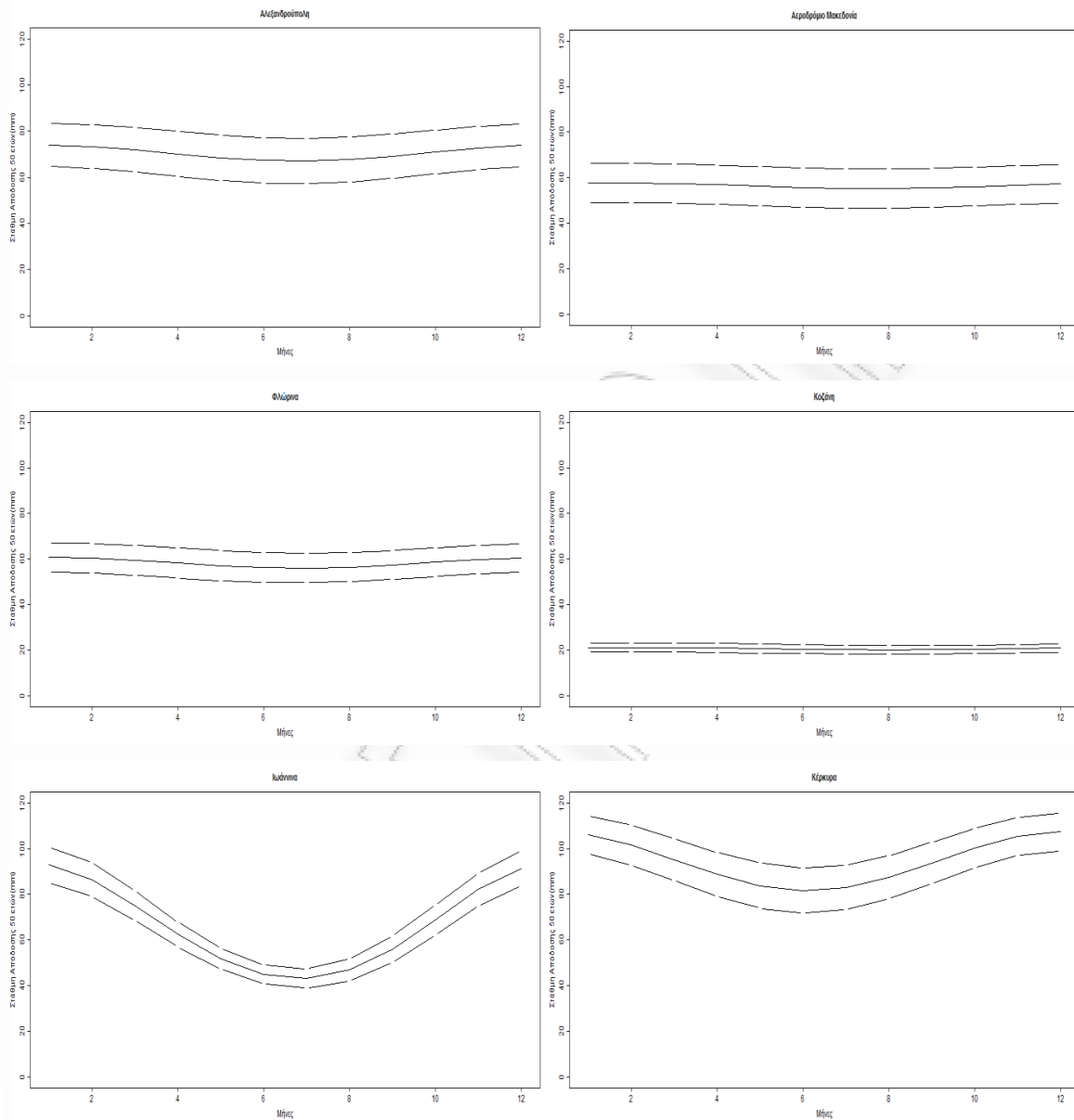


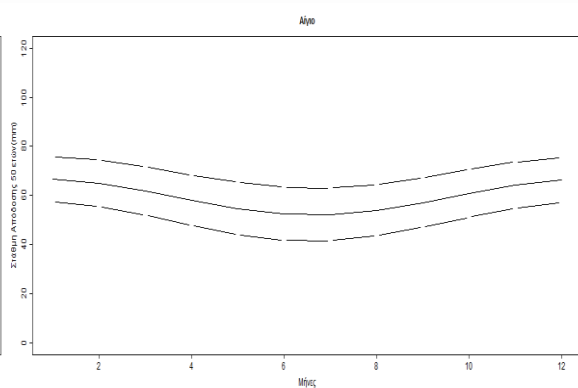
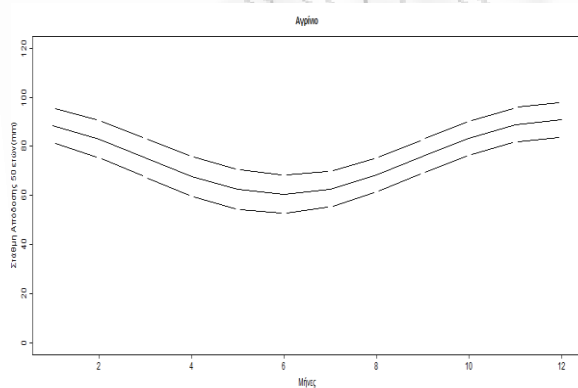
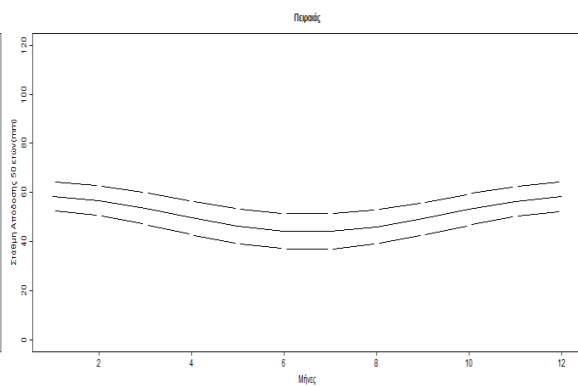
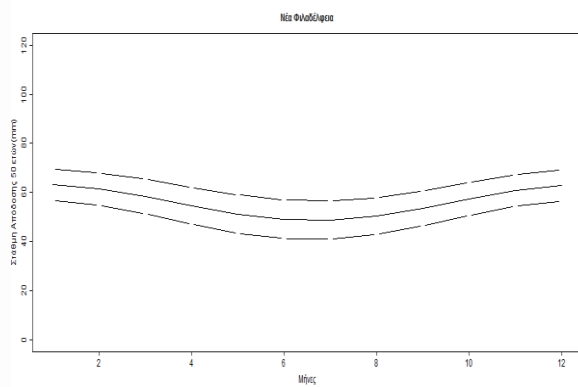
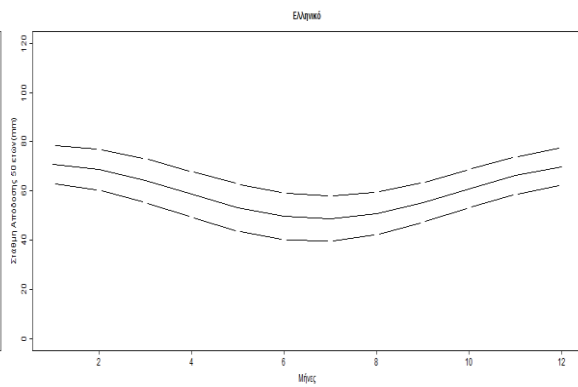
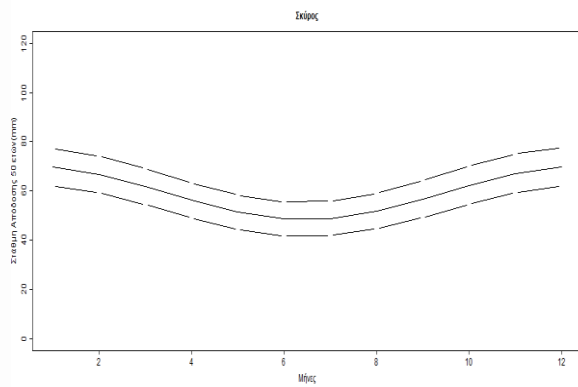
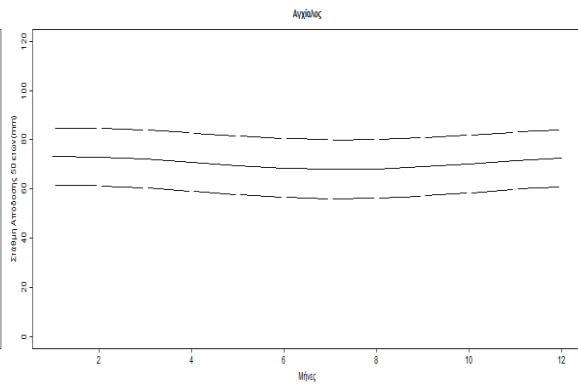
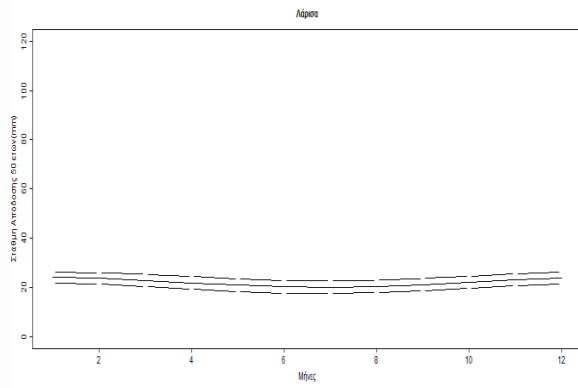


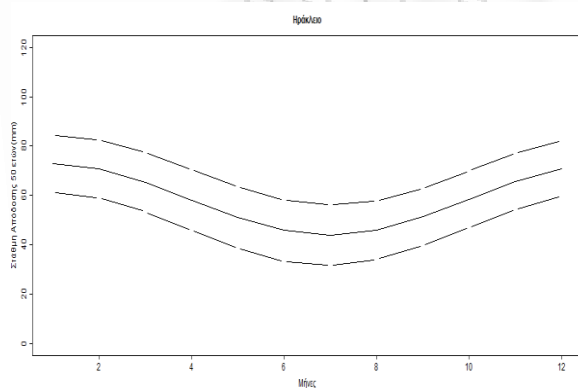
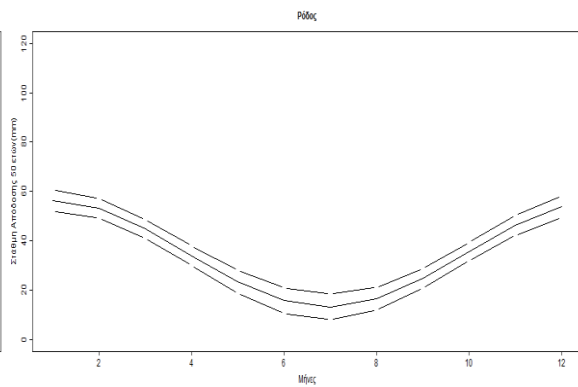
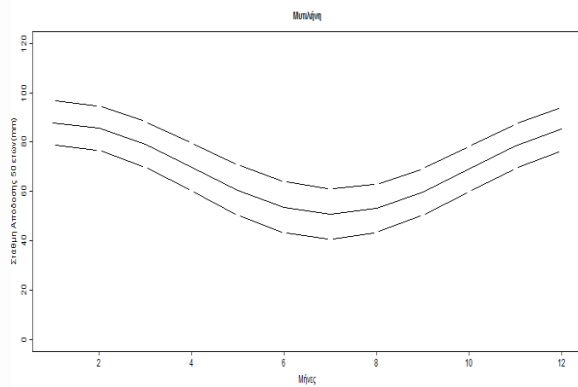
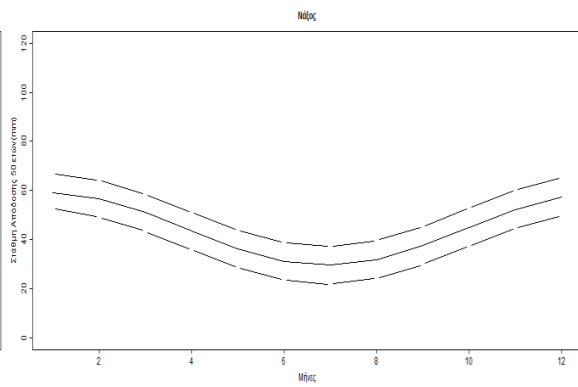
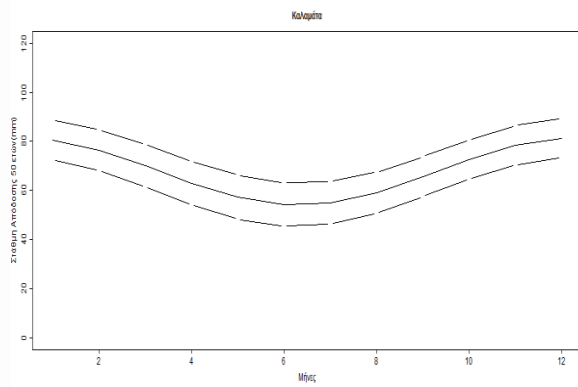
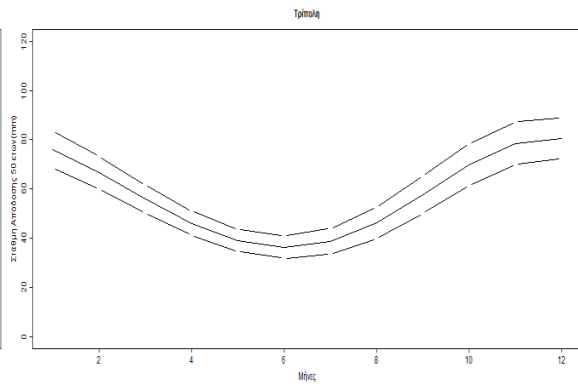
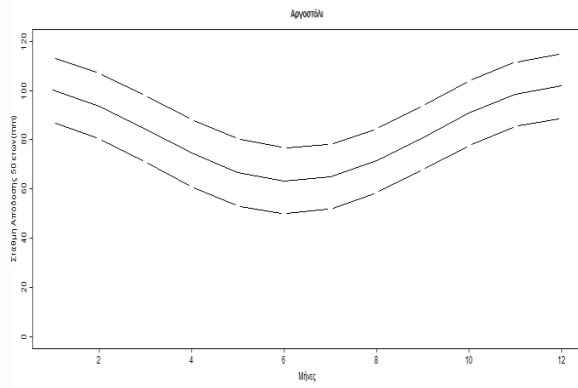


Σχημα 5.7

Γράφημα μηνιαίας στάθμης απόδοσης 50 ετών και 95 % δ.ε. για τις μέγιστες βροχοπτώσεις, για τους 21 μετεωρολογικούς σταθμούς.



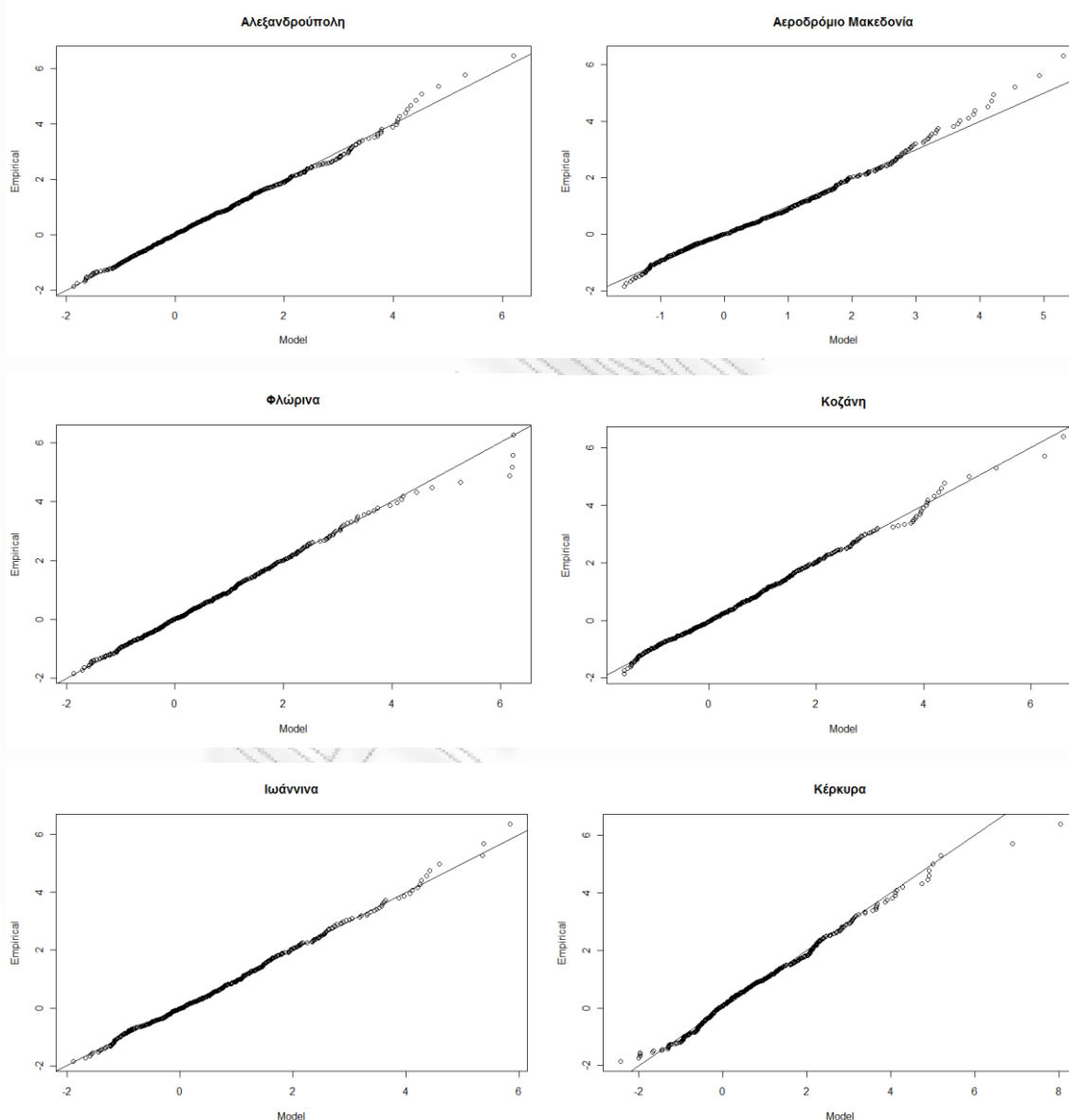


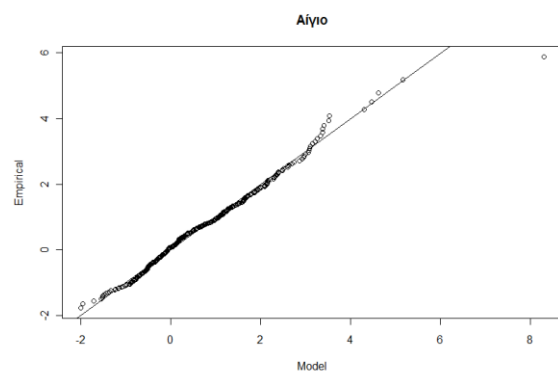
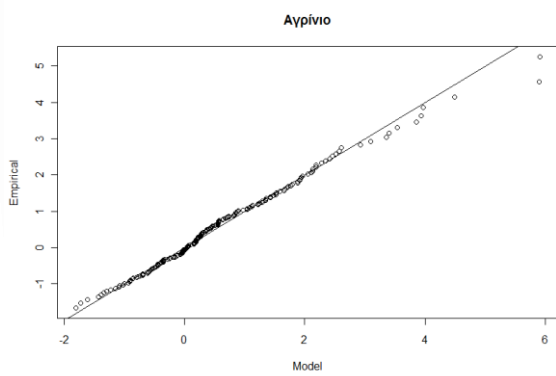
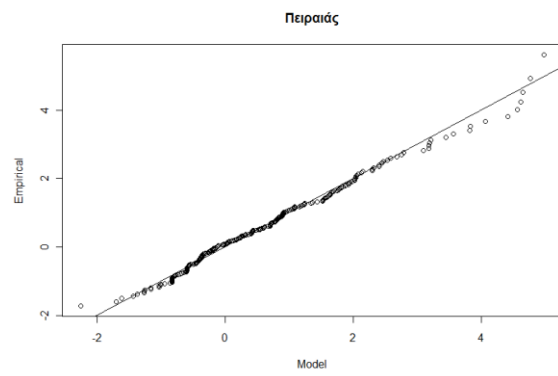
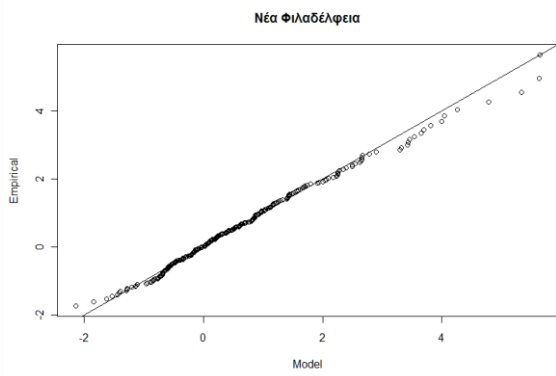
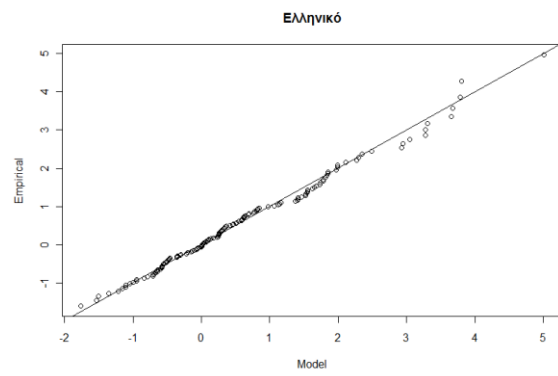
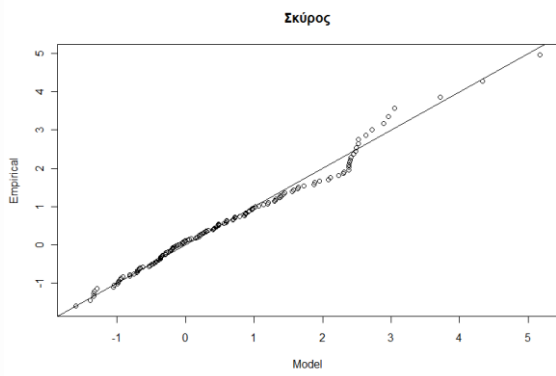
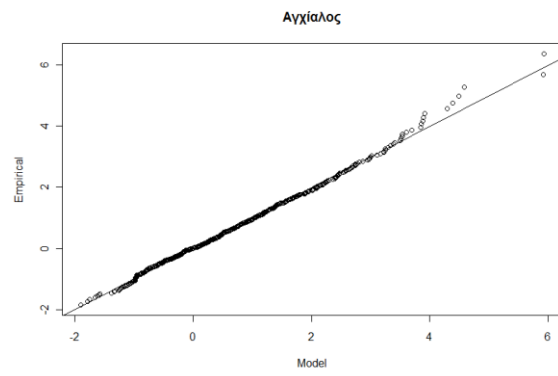
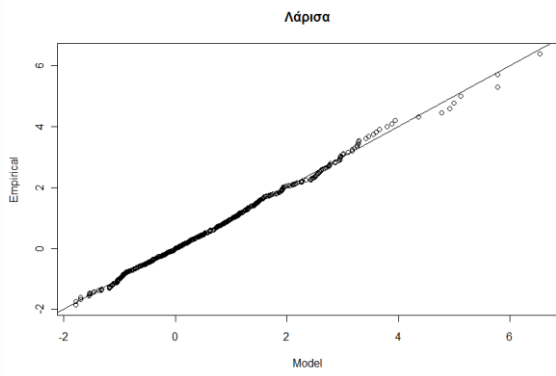


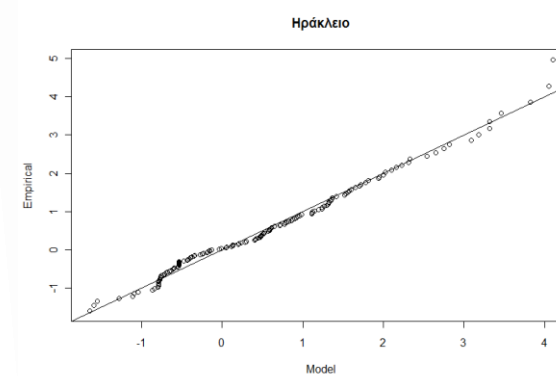
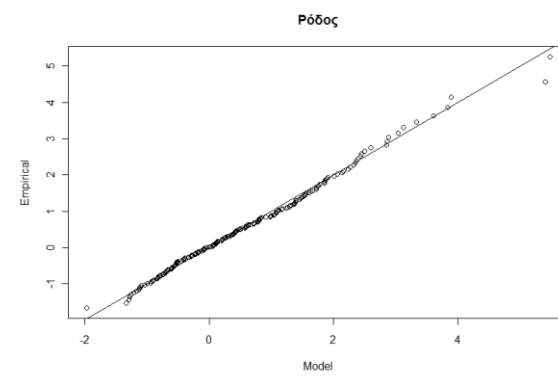
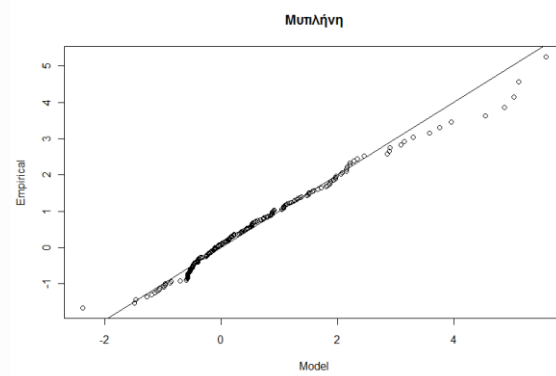
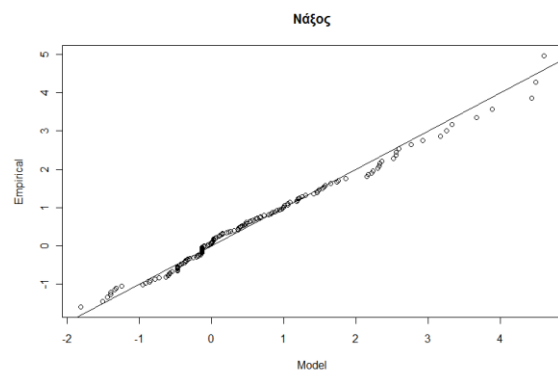
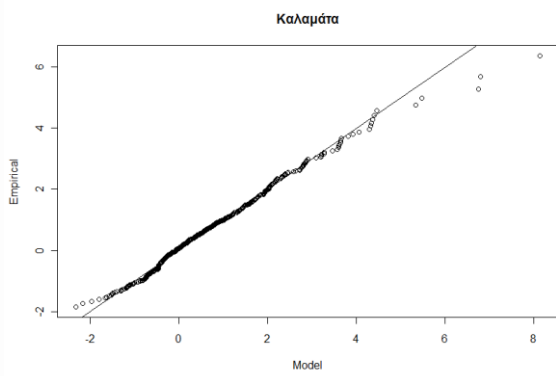
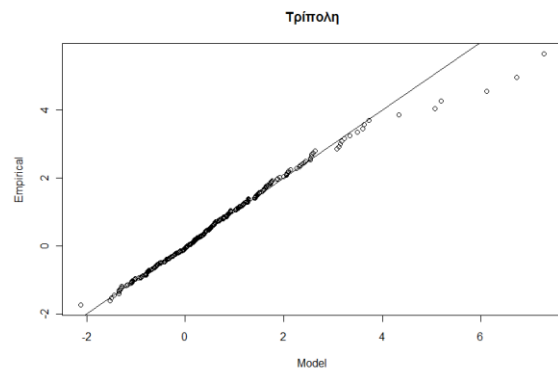
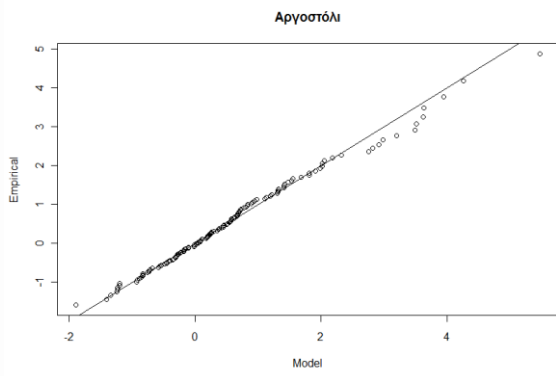
Ο έλεγχος για το αν τα δεδομένα προσαρμόζονται καλά στα μοντέλα που έχουμε εκτιμήσει παραπάνω για τις μέγιστες βροχοπτώσεις θα γίνει μέσω κατάλληλων Q-Q Plots που έχουν περιγραφεί στην Ενότητα 3.3.7 όπου οι αρχικές μη-στάσιμες παρατηρήσεις τυποποιούνται, έτσι ώστε να προέρχονται από μία στάσιμη τυπική κατανομή *Gumbel*. Κατασκευάζουμε τα επόμενα γραφήματα και παρατηρούμε ότι τα δεδομένα των μέγιστων βροχοπτώσεων προσαρμόζονται ικανοποιητικά στους σταθμούς που έχουν επιλεγεί (Παράρτημα Κώδικας 5).

Σχήμα 5.8

Q-Q Plots μέγιστων βροχοπτώσεων για τους 21 μετεωρολογικούς σταθμούς.







5.1.2 Αναλυση μέγιστων θερμοκρασιών

Από τους 21 μετεωρολογικούς σταθμούς που αναλύσαμε στην ενότητα με τις μέγιστες βροχοπτώσεις, τρεις, τα Ιωάννινα το Αίγιο και το Αργοστόλι δεν θα αναλυθούν σε αυτή την ενότητα λόγω πολλών ελλειπών τιμών που υπήρχαν σε αυτούς τους 3 σταθμούς στην μεταβλητή που αφορούσε τις μέγιστες θερμοκρασίες. Αξίζει να αναφέρουμε εδώ ότι αντίστοιχο ενδιαφέρον παρουσιάζει και η μοντελοποίηση των μηνιαίων ελάχιστων θερμοκρασιών. Σε αυτή την περίπτωση αντί για την μοντελοποίηση της μέγιστης παρατήρησης θα μοντελοποιούνται η ελάχιστη παρατήρηση. Την ίδια μεθοδολογία και συμπερασματολογία θα ακολουθούσαμε και σε αυτή την περίπτωση.

Ακολουθεί ο πίνακας με τους μετεωρολογικούς σταθμούς που θα αναλυθούν σε αυτή την ενότητα και το μέγεθος της κλάσης (του block) για κάθε μετεωρολογικό σταθμό.

Πίνακας 5.6

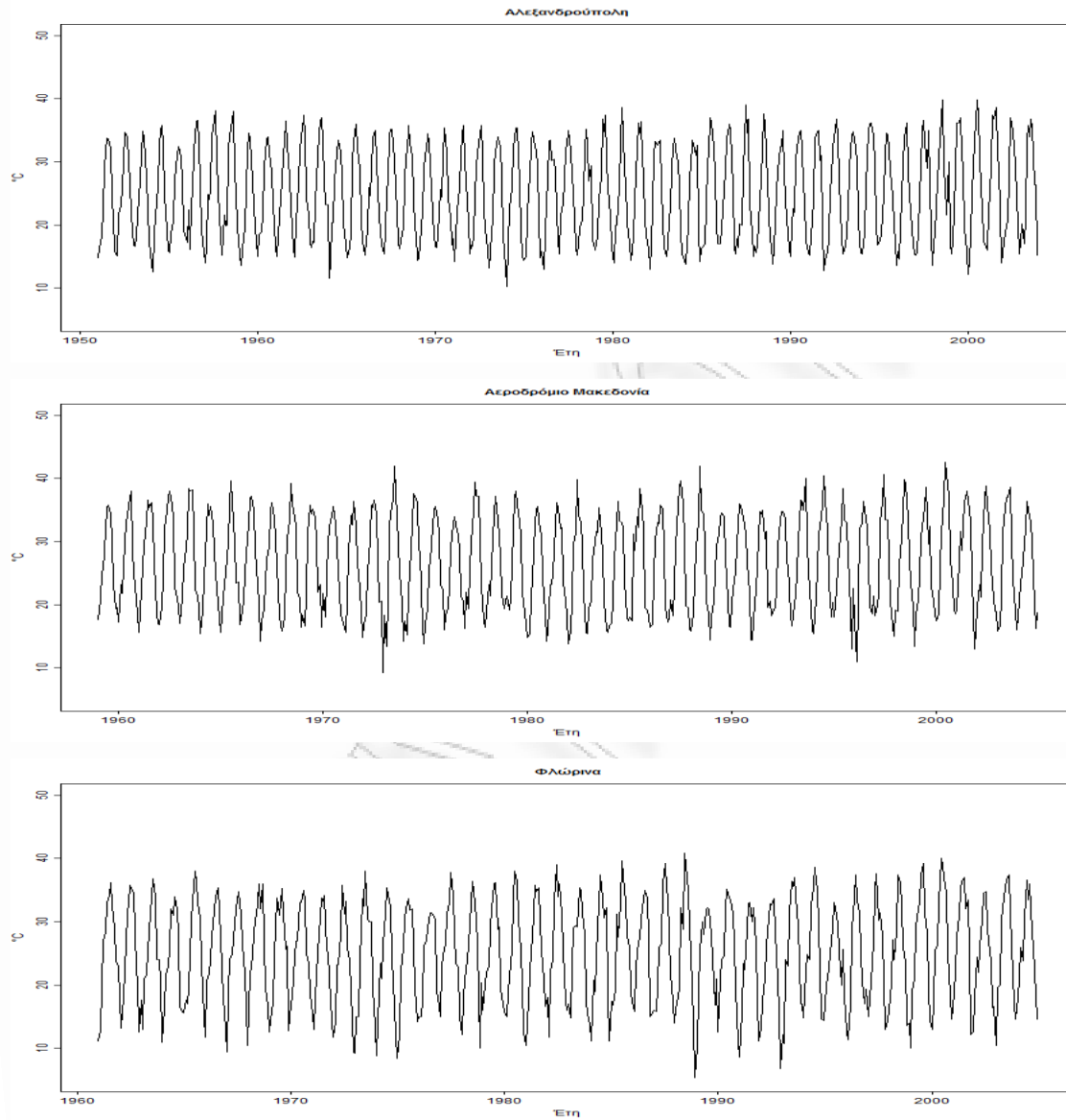
Οι 18 μετεωρολογικοί σταθμοί και το μέγεθος κάθε κλάσης από όπου προέρχονται οι μέγιστες θερμοκρασίες.

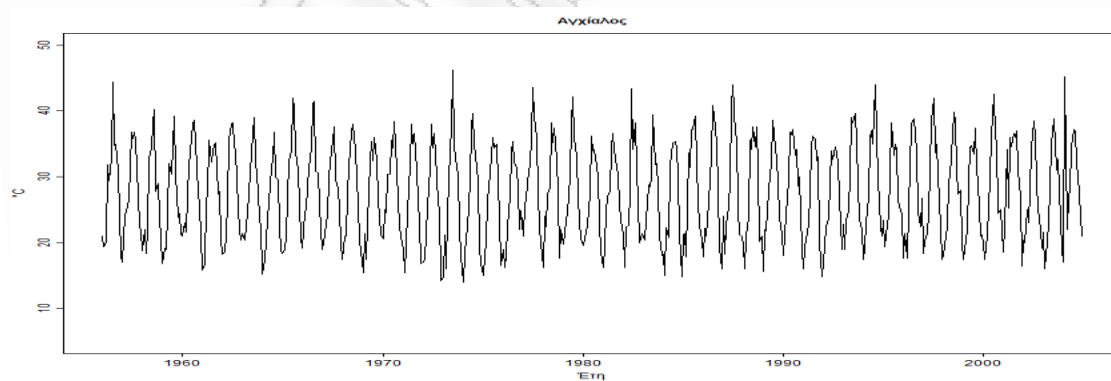
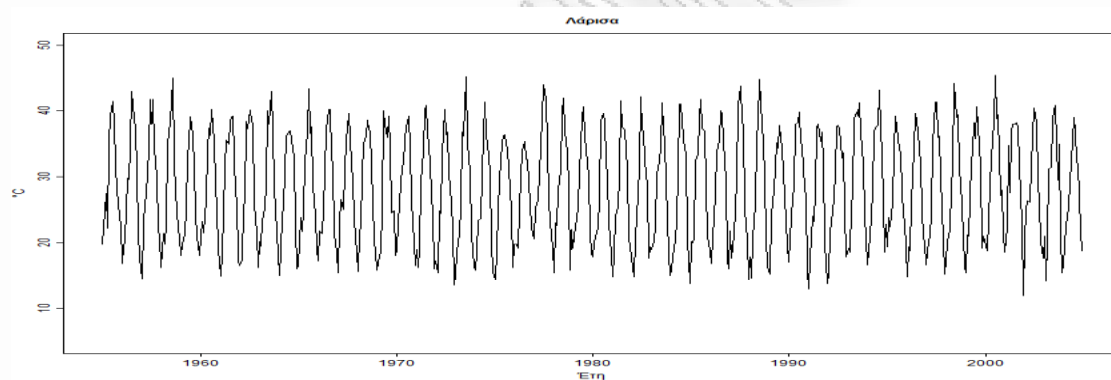
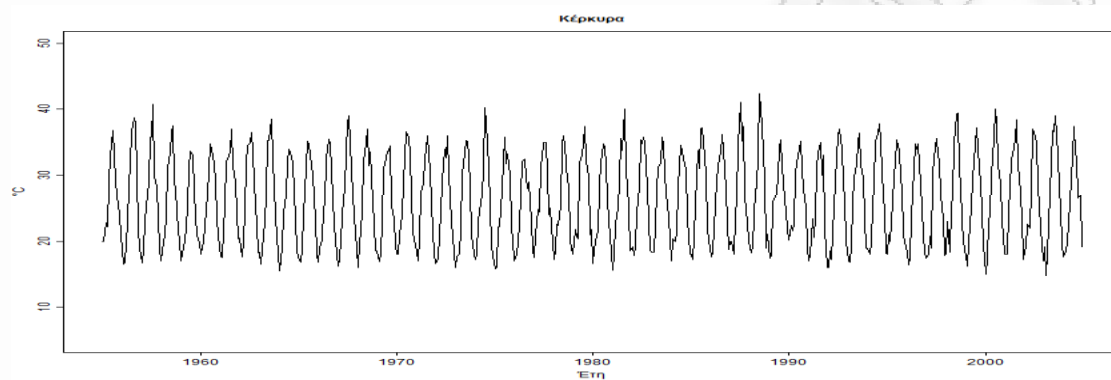
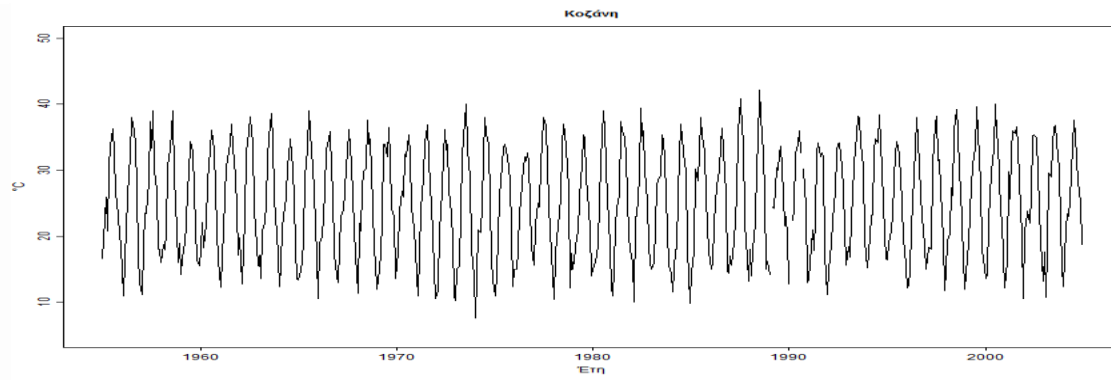
Μετεωρολογικός Σταθμός	Μέγεθος κλάσης
1) Αλεξανδρούπολη	120
2) Αεροδρ. Μακεδονία	120
3) Φλώρινα	60
4) Κοζάνη	120
5) Κέρκυρα	60
6) Λάρισα	30
7) Αγχιάλος	60
8) Σκύρος	60
9) Ελληνικό	30
10) Ν. Φιλαδέλφεια	30
11) Πειραιάς	30
12) Αργίριο	30
13) Τρίπολη	30
14) Καλαμάτα	30
15) Νάξος	30
16) Μυτιλήνη	60
17) Ρόδος	30
18) Ηράκλειο	30

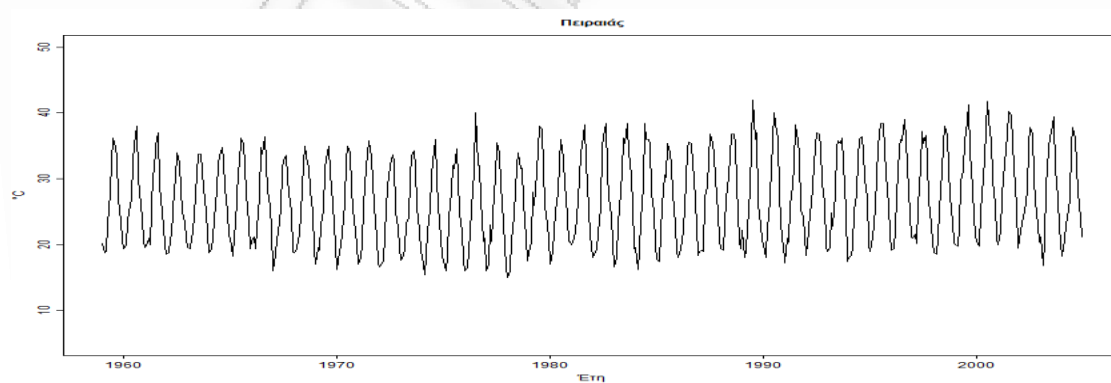
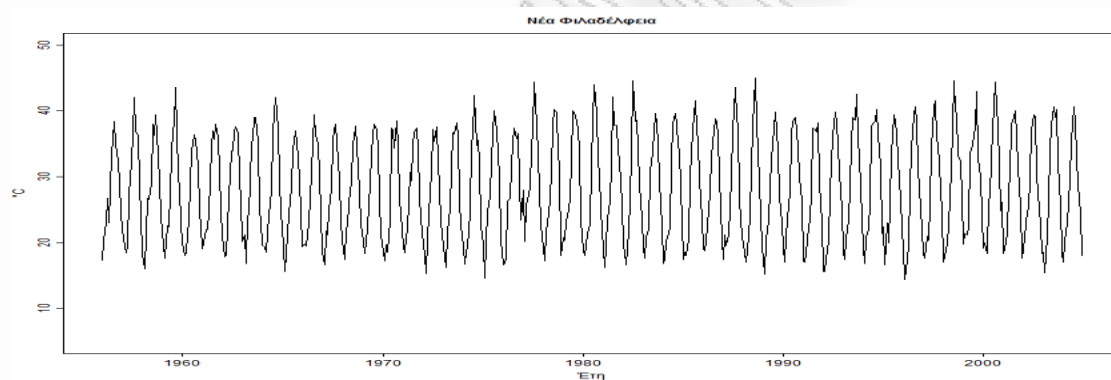
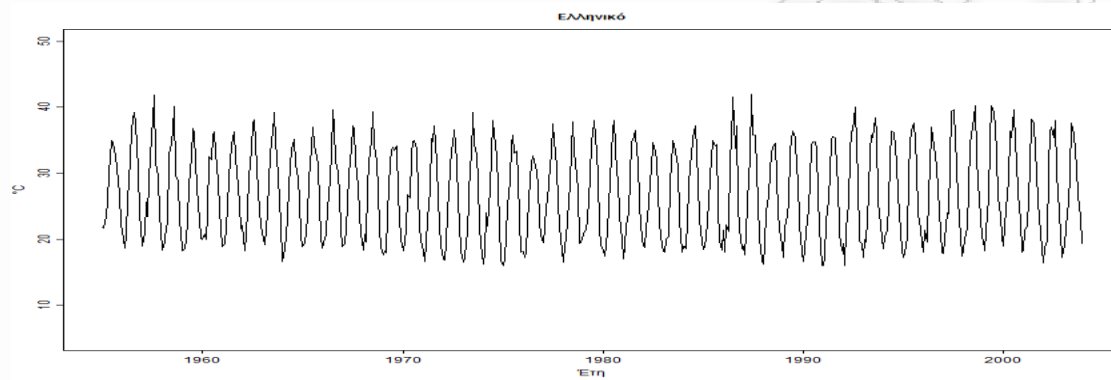
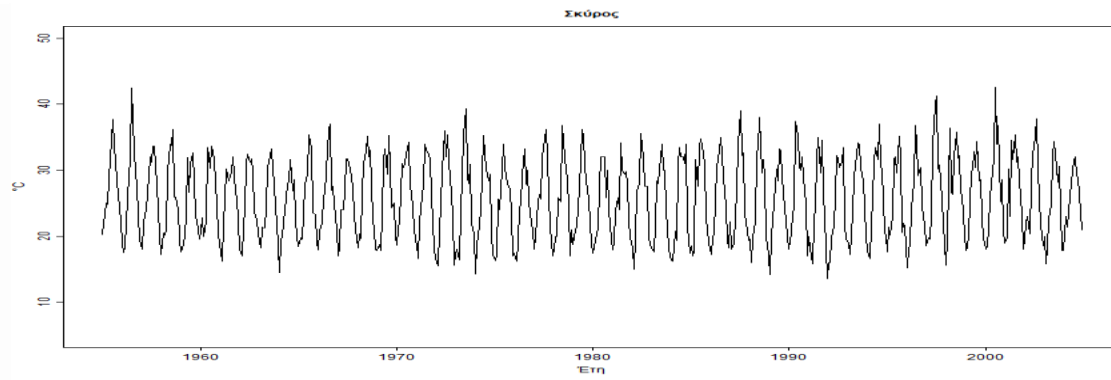
Στο σχήμα 5.9 απεικονίζεται η χρονολογική σειρά των μηνιαίων μέγιστων θερμοκρασιών για τους 18 μετεωρολογικούς σταθμούς για την χρονική περίοδο που υπάρχουν παρατηρήσεις. Αντίστοιχα, στο σχήμα 5.10 παρουσιάζονται τα θερμογράμματα των μηνιαίων μέγιστων θερμοκρασιών για κάθε σταθμό. Σε αυτό το σχήμα είναι εμφανής η περιοδικότητα που υπάρχει ανάμεσα στα μέγιστα των μηνιαίων θερμοκρασιών.

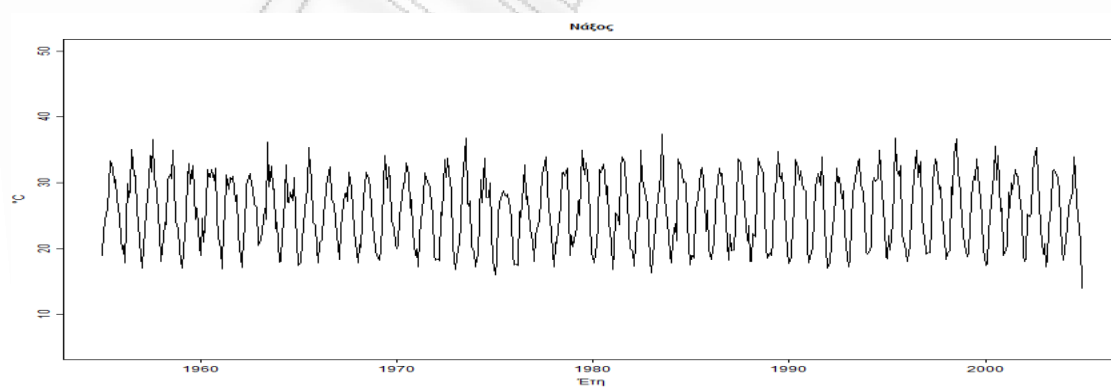
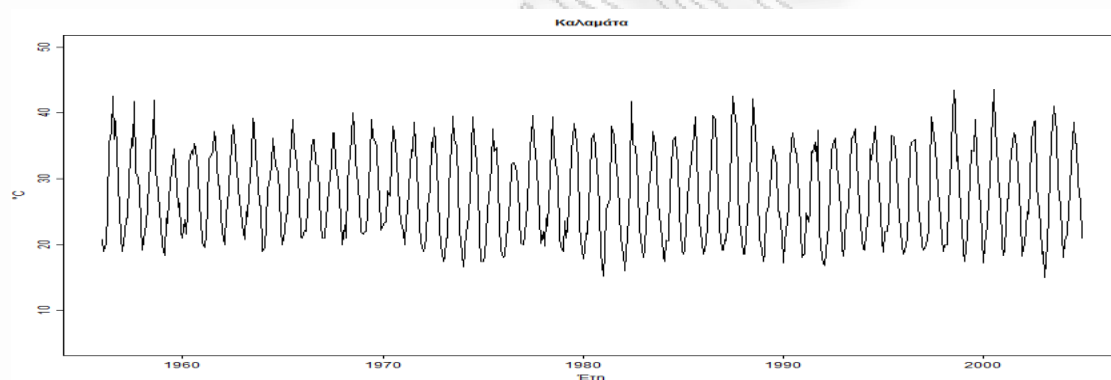
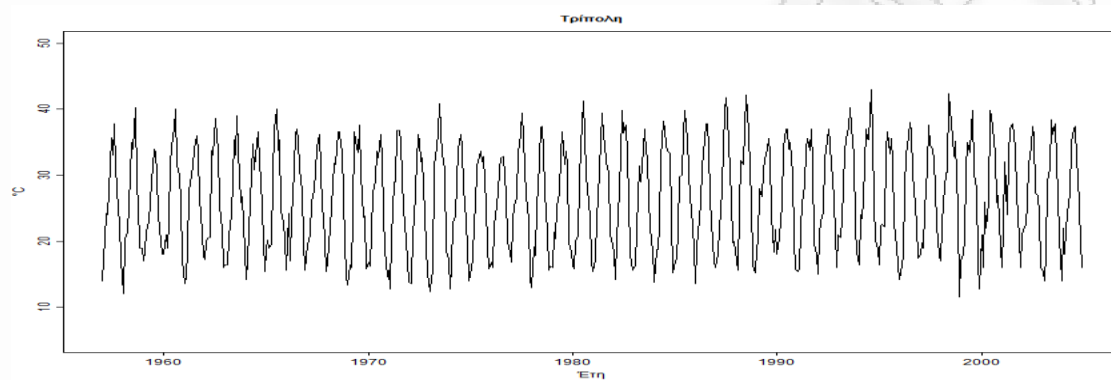
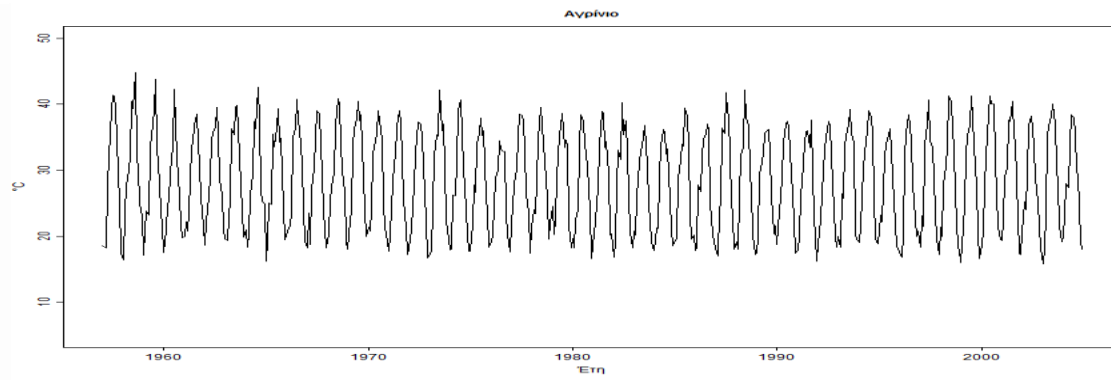
Σχημα 5.9

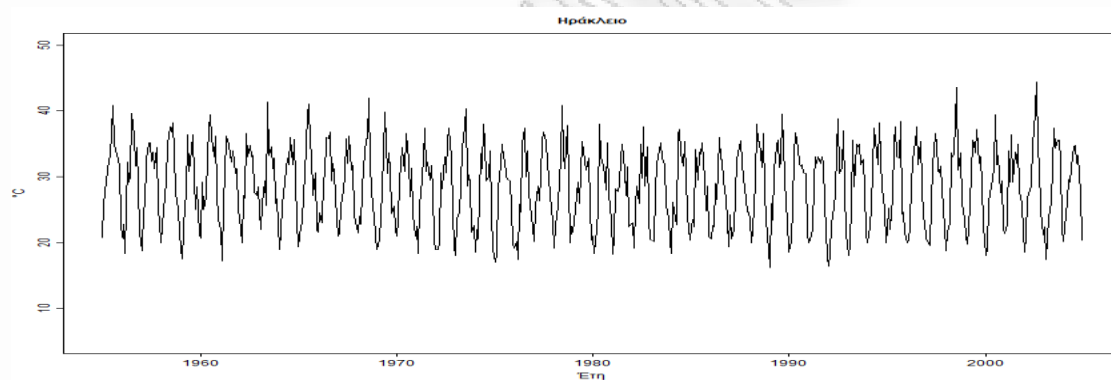
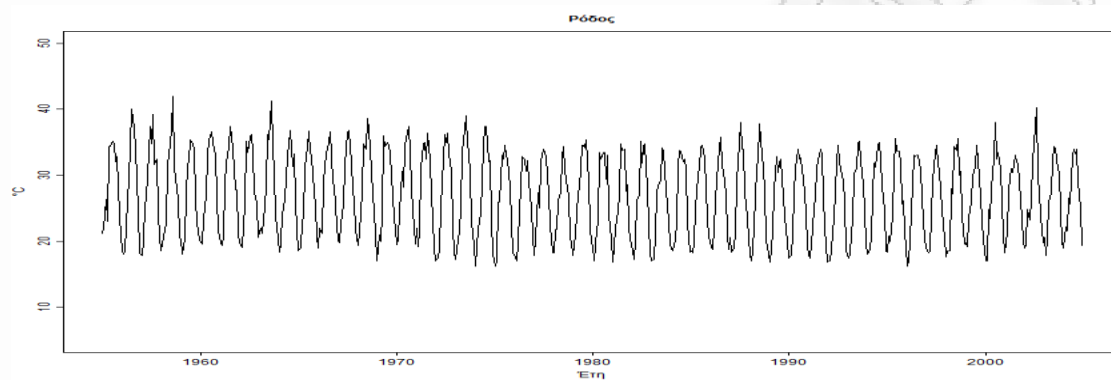
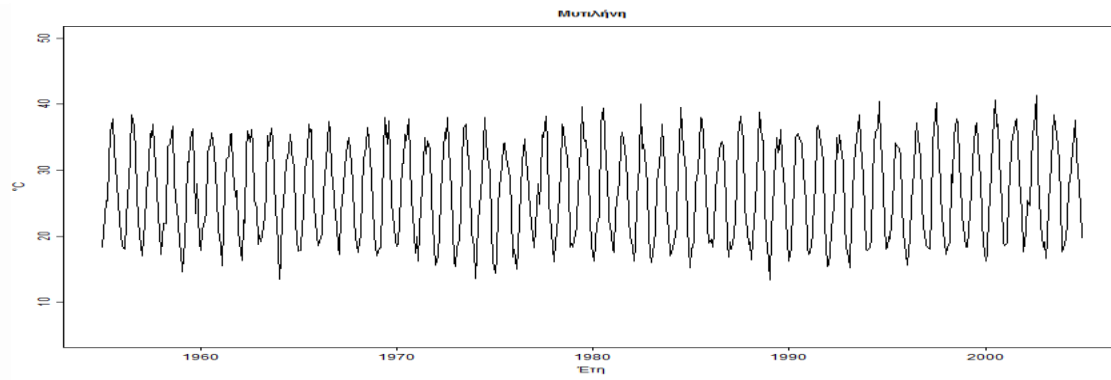
Χρονολογική σειρά μηνιαίων μέγιστων θερμοκρασιών για τους 18 μετεωρολογικούς σταθμούς.





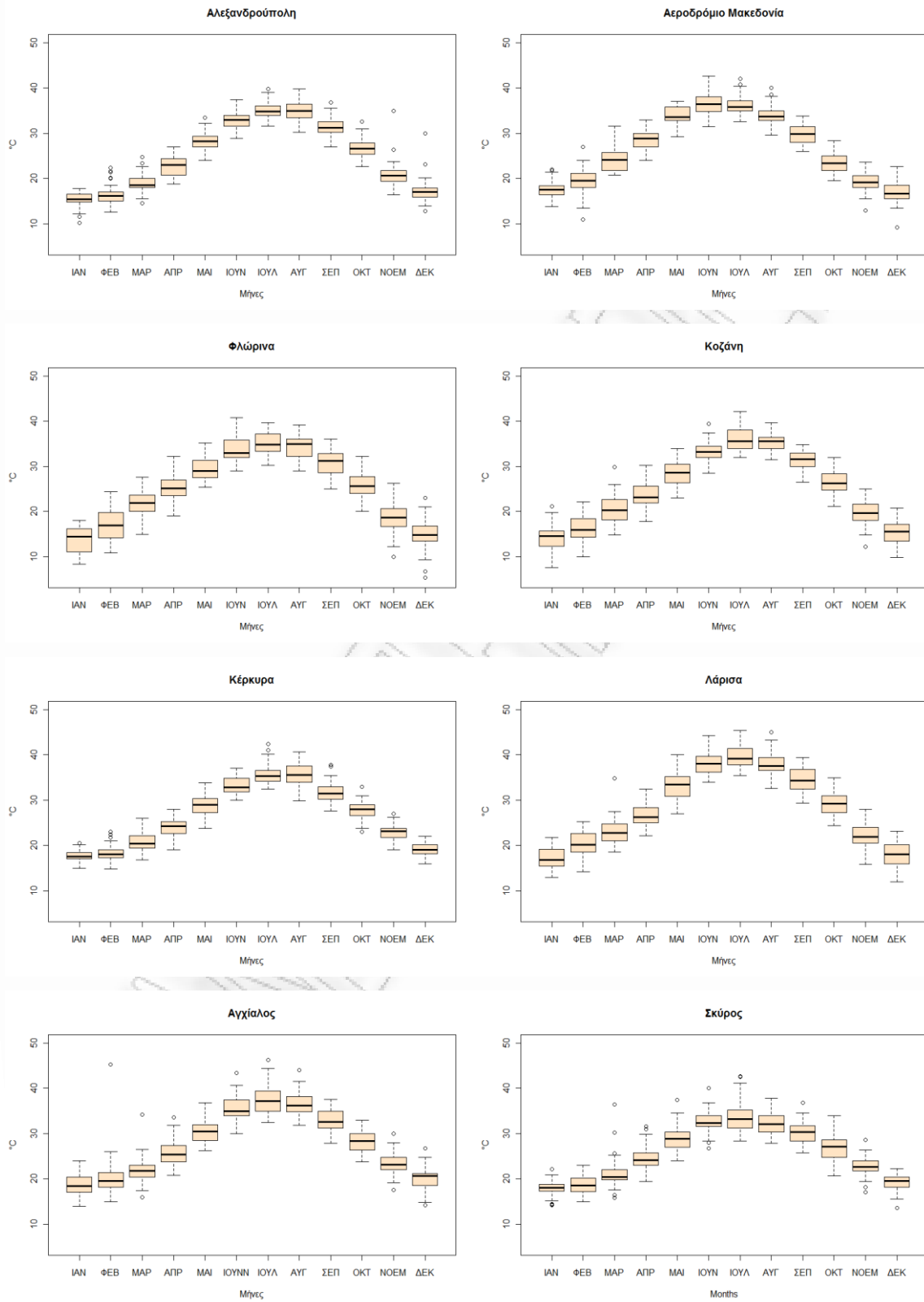


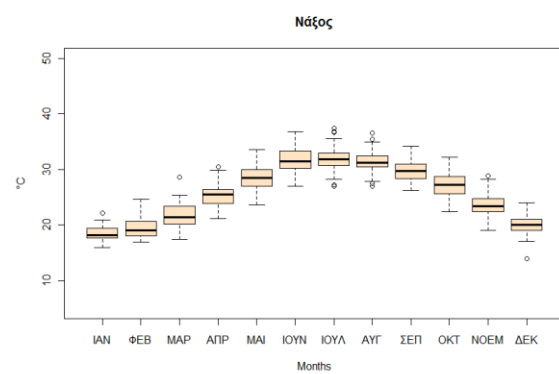
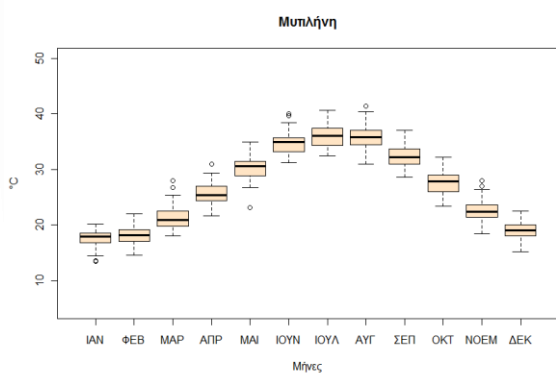
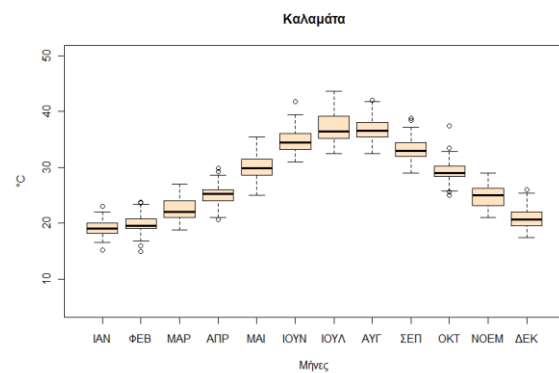
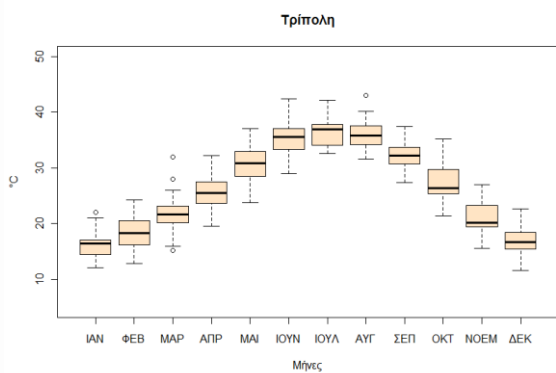
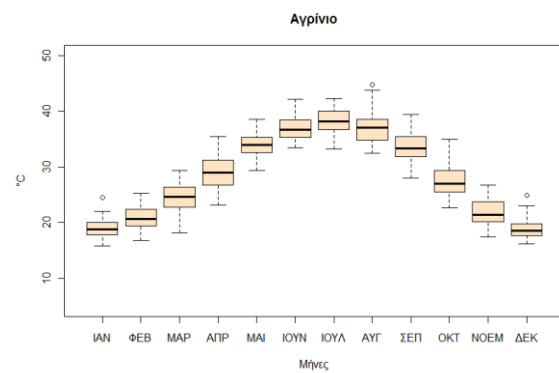
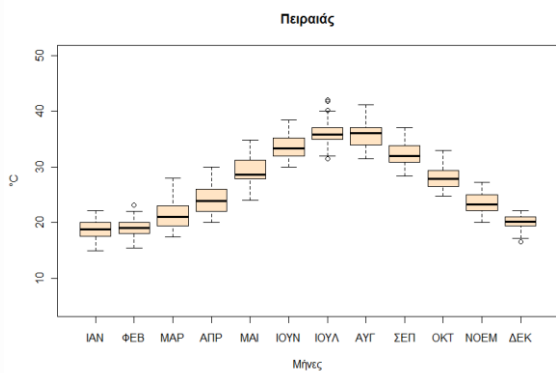
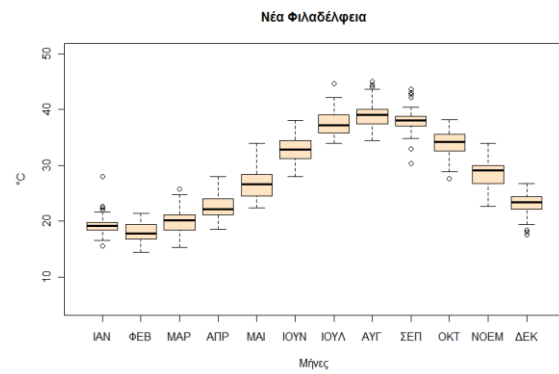
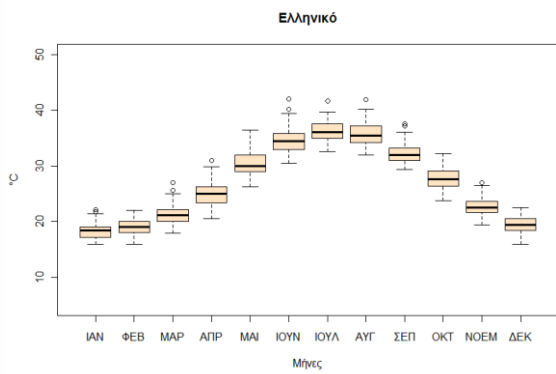


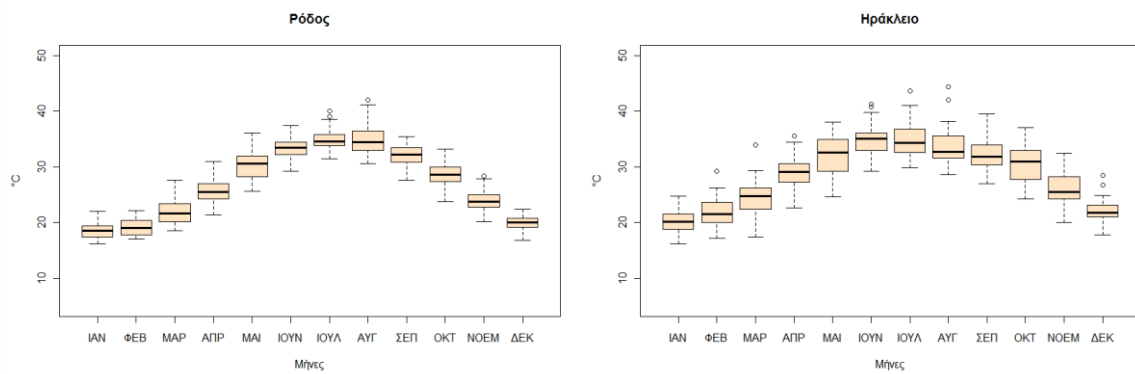


Σχήμα 5.10

Θηκογράμματα μηνιαίων μέγιστων θερμοκρασιών για τους 18 μετεωρολογικούς σταθμούς.







Στους πίνακες 5.6 και 5.7 παρουσιάζονται οι εκτιμήσεις των παραμέτρων μέγιστης πιθανοφάνειας της κατανομής *GEV* για τις μέγιστες θερμοκρασίες με την μοντελοποίηση που εισήχθη στην Ενότητα 5.1 και τα τυπικά σφάλματα αυτών για τους 18 σταθμούς που παρουσιάστηκαν παραπάνω. Η σειρά με την οποία εμφανίζονται οι σταθμοί είναι η ίδια με την οποία εμφανίζονται στον πίνακα 5.6. Στους σταθμούς όπου υπάρχει η τιμή μηδέν στην παράμετρο σχήματος ξ_0 σημαίνει ότι τα δεδομένα προέρχονται από κατανομή *Gumbel*, ενώ αντίστοιχα στους σταθμούς όπου η παράμετρος μ_3 είναι μηδέν η ύπαρξη τάσης δεν ήταν στατιστικά σημαντική σε επίπεδο σημαντικότητας 5%. Επίσης στους σταθμούς όπου στην εκτίμηση των παραμέτρων υπάρχει η ένδειξη(*) σημαίνει ότι η παρουσία των παραμέτρων στο μοντέλο δεν ήταν στατιστικά σημαντική σε επίπεδο σημαντικότητας 5%.

Πίνακας 5.7

Εκτιμήσεις παραμέτρων $\mu_0, \mu_1, \mu_2, \mu_3$ κατανομής *GEV* για τις μέγιστες θερμοκρασίες και τυπικά σφάλματα εκτιμητών.

	$\hat{\mu}_0$	<i>s.e.</i> ($\hat{\mu}_0$)	$\hat{\mu}_1$	<i>s.e.</i> ($\hat{\mu}_1$)	$\hat{\mu}_2$	<i>s.e.</i> ($\hat{\mu}_2$)	$\hat{\mu}_3$	<i>s.e.</i> ($\hat{\mu}_3$)
1)	26.234	0.111	-6.057	0.160	-8.045	0.142	0	-
2)	28.148	0.165	-2.442	0.219	-9.858	0.221	0	-
3)	25.701	0.186	-4.724	0.231	-8.788	0.238	0	-
4)	26.213	0.188	-5.961	0.262	-8.778	0.244	0	-
5)	26.643	0.102	-5.866	0.138	-6.982	0.140	0	-
6)	27.253	0.109	-5.807	0.154	-9.472	0.147	0	-
7)	28.082	0.1387	-5.435	0.184	-7.599	0.189	0	-
8)	25.953	0.114	-4.701	0.153	-6.142	0.157	0	-
9)	26.054	0.083	-5.388	0.114	-7.373	0.117	0	-
10)	27.409	0.103	-9.325	0.141	-4.763	0.135	0	-
11)	25.964	0.095	-8.280	0.130	-2.800	0.128	0	-
12)	27.480	0.105	-4.255	0.140	-8.928	0.139	0	-
13)	25.385	0.114	-5.517	0.158	-8.608	0.151	0	-
14)	26.767	0.082	-5.884	0.119	-6.464	0.118	0	-
15)	24.903	0.085	-3.960	0.112	-5.175	0.112	0	-
16)	27.052	0.100	-5.382	0.134	-7.748	0.134	0	-
17)	27.205	0.155	-5.345	0.102	-6.488	0.103	-0.003	0.0004
18)	27.315	0.123	-3.953	0.161	-5.783	0.165	0	-

Πίνακας 5.8

Εκτιμήσεις παραμέτρων σ_0 , σ_1 , σ_2 , ξ_0 κατανομής *GEV* για τις μέγιστες θερμοκρασίες και τυπικά σφάλματα εκτιμητών.

	$\hat{\sigma}_0$	<i>s.e.</i> ($\hat{\sigma}_0$)	$\hat{\sigma}_1$	<i>s.e.</i> ($\hat{\sigma}_1$)	$\hat{\sigma}_2$	<i>s.e.</i> ($\hat{\sigma}_2$)	$\hat{\xi}_0$	<i>s.e.</i> ($\hat{\xi}_0$)
1)	0.283	0.062	-0.046*	0.088	0.136*	0.094	0	-
2)	0.548	0.066	0.038*	0.058	-0.039*	0.068	-0.182	0.052
3)	0.984	0.049	0.079*	0.065	0.174*	0.067	-0.311	0.042
4)	0.714	0.064	0.4*	0.086	0.058*	0.091	-0.252	0.056
5)	0.478	0.043	-0.039*	0.054	-0.228*	0.061	-0.170	0.032
6)	0.916	0.030	0.042*	0.043	0.026*	0.043	-0.158	0.015
7)	0.767	0.044	0.072*	0.062	-0.119*	0.061	-0.116	0.031
8)	0.594	0.044	-0.027	0.061	-0.298	0.063	0	-
9)	0.592	0.031	0.043	0.044	-0.108	0.046	-0.114	0.030
10)	0.843	0.030	-0.052*	0.045	-0.025*	0.042	-0.208	0.018
11)	0.708	0.033	0.022*	0.043	-0.201*	0.045	-0.184	0.027
12)	0.833	0.031	0.051*	0.040	-0.078*	0.047	-0.194	0.023
13)	0.926	0.031	0.056*	0.042	-0.048*	0.043	-0.164	0.021
14)	0.632	0.029	0.006*	0.041	-0.078*	0.041	-0.113	0.001
15)	0.611	0.030	-0.061	0.041	-0.129	0.042	-0.122	0.029
16)	0.455	0.044	-0.076*	0.055	-0.147*	0.060	-0.168	0.034
17)	0.531	0.031	-0.064	0.042	-0.145	0.046	-0.096	0.030
18)	0.994	0.032	0.013*	0.043	-0.173*	0.047	-0.159	0.028

Η ύπαρξη τάσης στον μετεωρολογικό σταθμό της Ρόδου αν και στατιστικά σημαντική σε επίπεδο σημαντικότητας 5% είναι πολύ μικρή και δεν την λαμβάνουμε υπόψη μας κατά τον υπολογισμό της παραμέτρου θέσεως, της μέσης μηνιαίας μέγιστης θερμοκρασίας και των σταθμών απόδοσης.

Στο σχήμα 5.11 απεικονίζονται οι σημειακές εκτιμήσεις της παραμέτρου θέσεως $\mu(i)$ για κάθε μήνα του έτους μαζί με τα προσεγγιστικά 95% διαστήματα εμπιστοσύνης που προκύπτουν μέσω της ασυμπτωτικής κανονικότητας των εκτιμητών. Η ερμηνεία των γραφημάτων είναι ταυτόσημη με τα αντίστοιχα γραφήματα των μέγιστων βροχοπτώσεων.

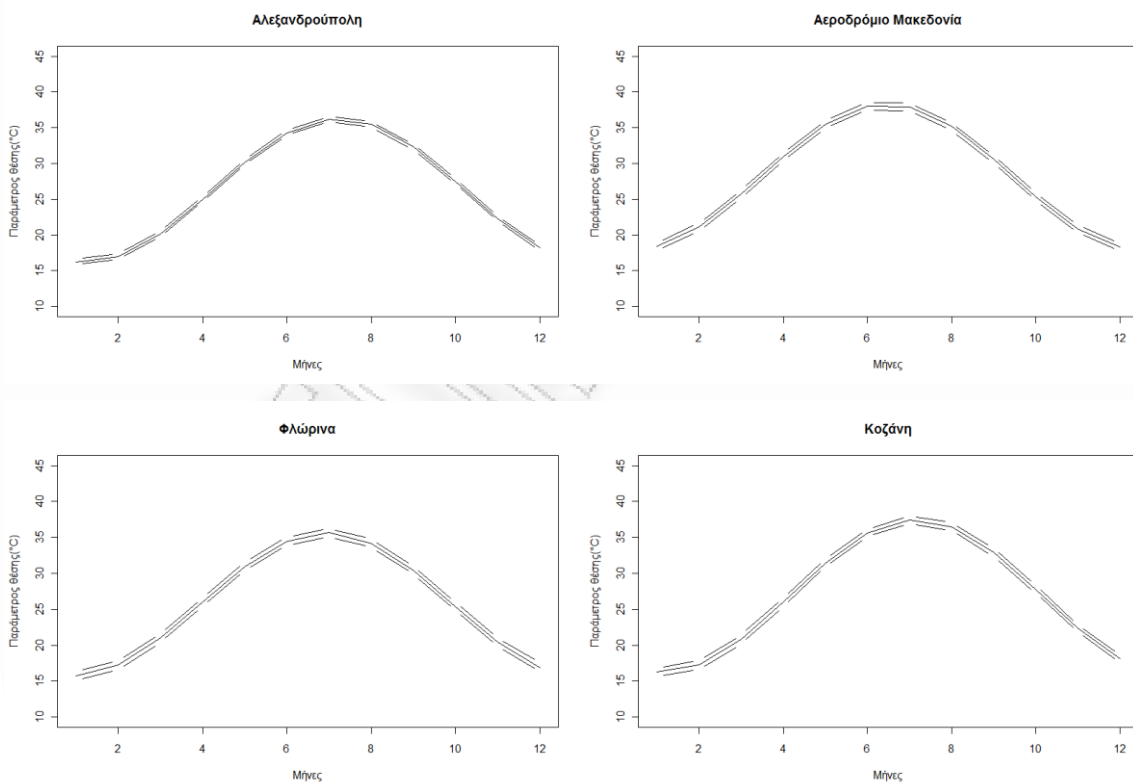
Στο γράφημα η συνεχής γραμμή απεικονίζει την μηνιαία σημειακή εκτίμηση του $\mu(i)$ που δίνεται από την συνάρτηση \square

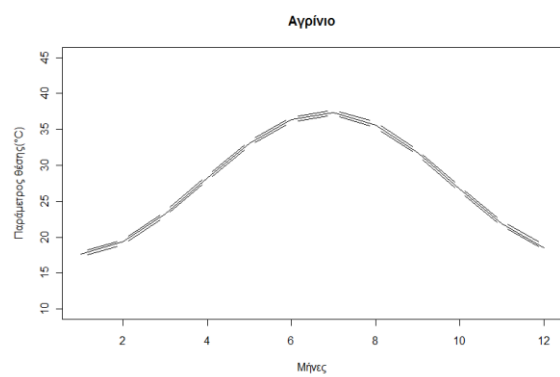
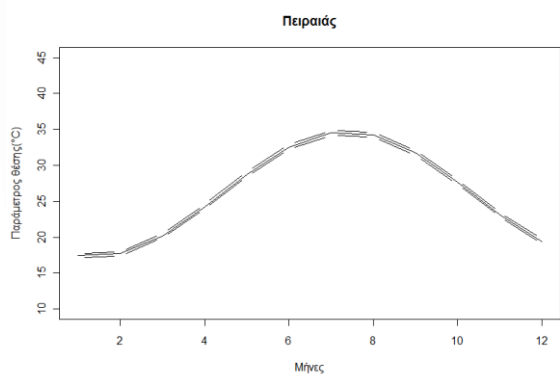
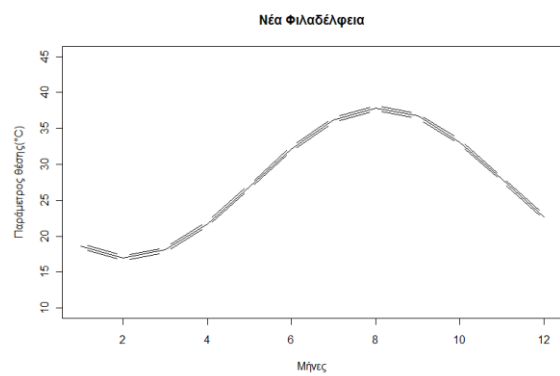
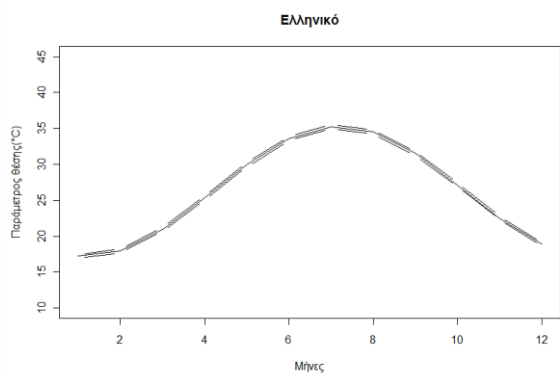
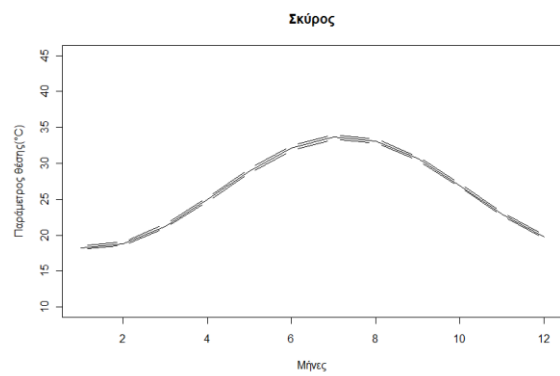
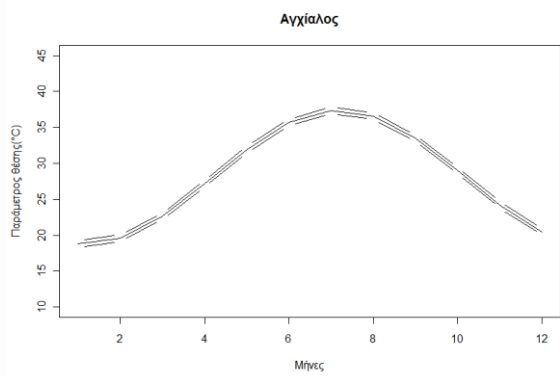
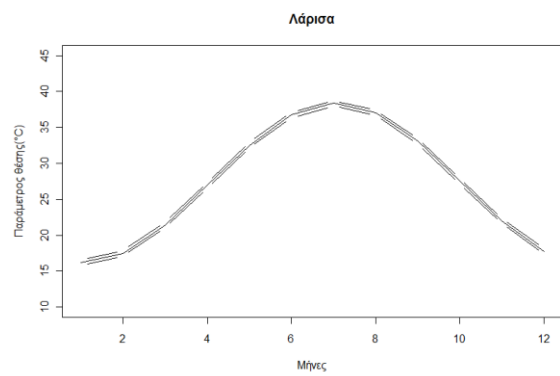
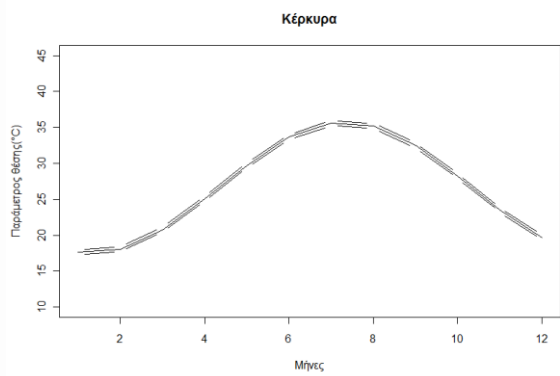
$$\hat{\mu}(i) = \hat{\mu}_0 + \hat{\mu}_1 \sin\left(\frac{2\pi i}{12}\right) + \hat{\mu}_2 \cos\left(\frac{2\pi i}{12}\right), \quad i = 1, 2, \dots, 12.$$

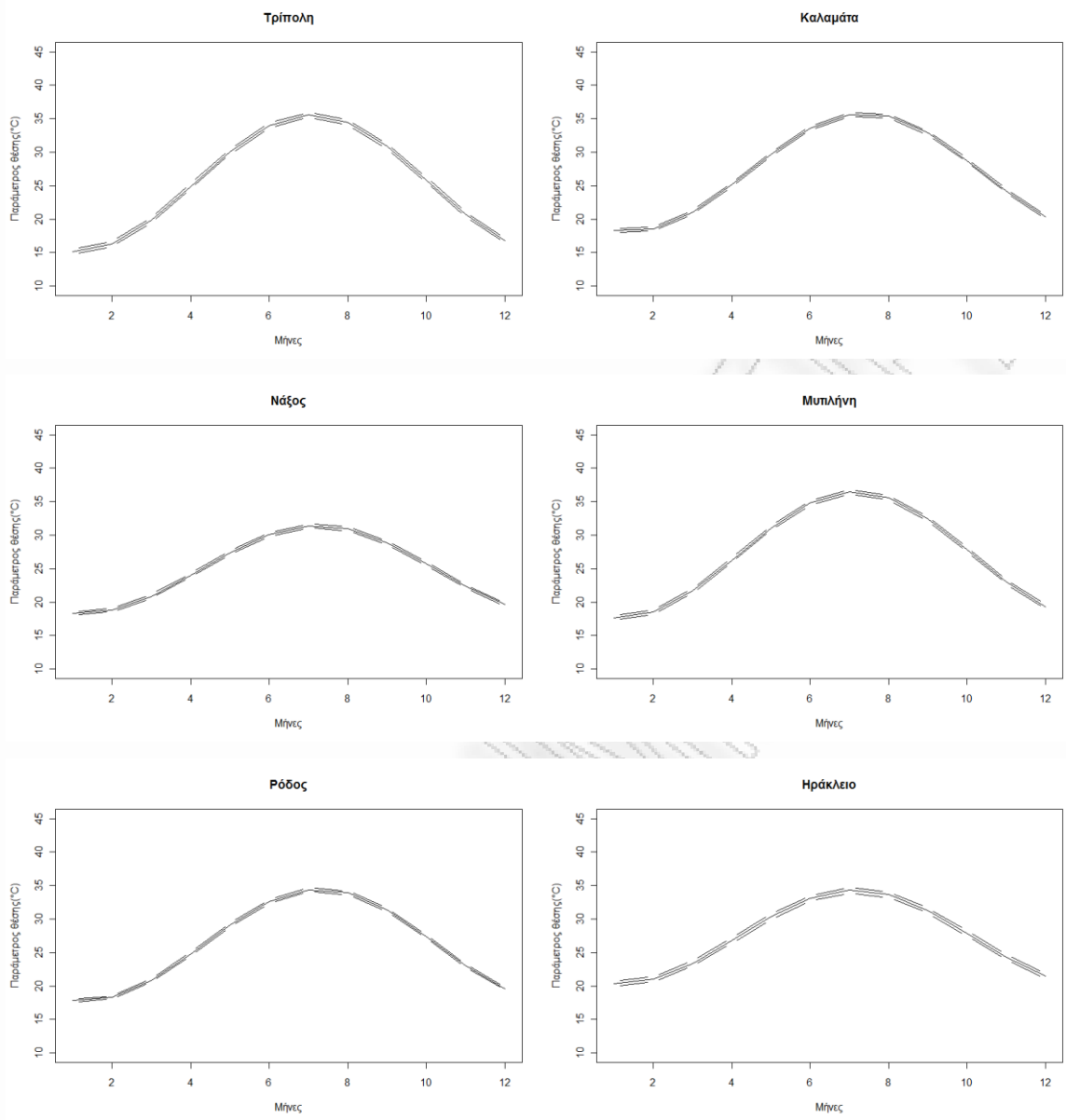
ενώ οι ασυνεχείς γραμμές αναφέρονται στο άνω και κάτω όριο του δ.ε. μέσω της κανονικής προσέγγισης που πήραμε από τον τύπο $\square \hat{\mu}(i) \pm z_{\alpha/2} \cdot \sqrt{\text{Var}(\hat{\mu}(i))}$ για $i = 1, 2, \dots, 12$ όπου α το άνω ποσοστιαίο σημείο της κανονικής κατανομής $N(0,1)$. Οι τύποι για την διασπορά του εκτιμητή και η μεθοδολογία υπολογισμού του είναι ταυτόσημοι με την περίπτωση των μέγιστων βροχοπτώσεων.

Σχήμα 5.11

Γράφημα της εκτιμώμενης παραμέτρου $\mu(i)$ και 95% δ.ε. για τις μέγιστες θερμοκρασίες, για $i = 1, 2, \dots, 12$ για καθένα από τους 18 μετεωρολογικούς σταθμούς.







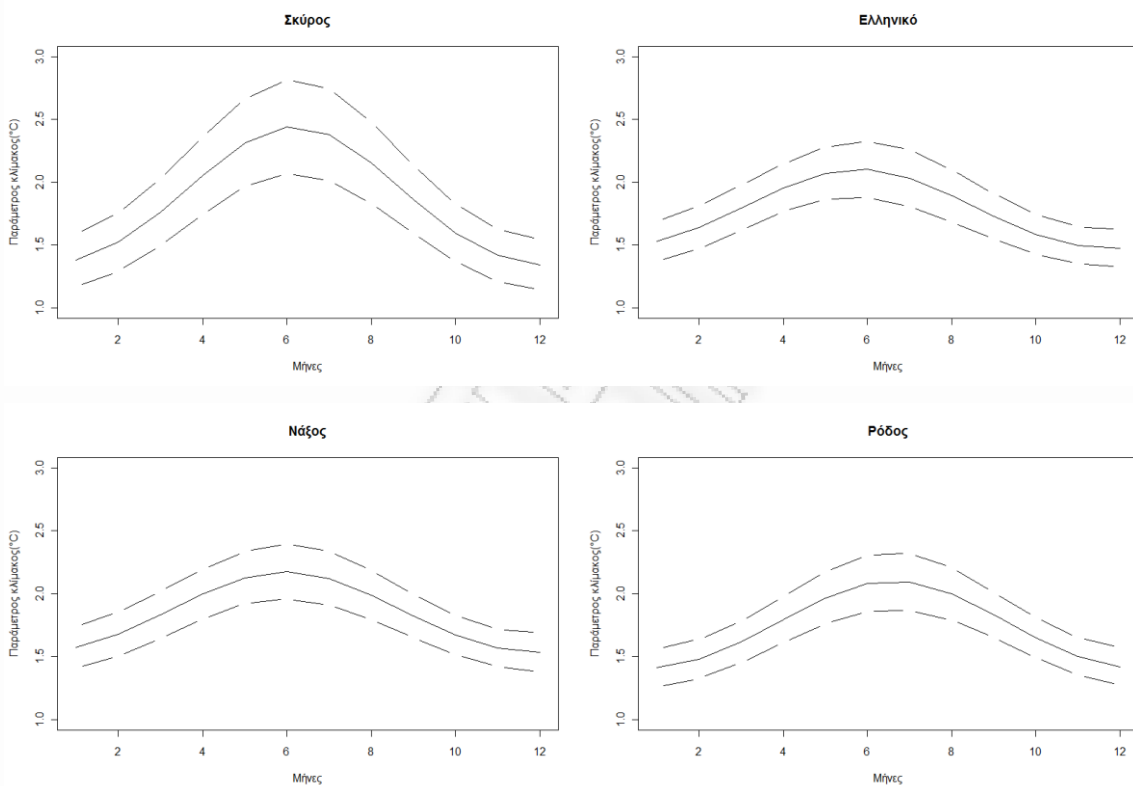
Στο σχήμα 5.12 απεικονίζονται οι εκτιμημένες τιμές της παραμέτρου κλίμακος για κάθε μήνα του έτους καθώς και τα προσεγγιστικά 95% δ.ε. που παρήχθησαν μέσω της ασυμπτωτικής κανονικότητας, στους σταθμούς όπου βάσει του δείγματος που είχαμε, η περιοδικότητα στην παράμετρο κλίμακος ήταν στατιστικά σημαντική σε επίπεδο σημαντικότητας 5%. Η ερμηνεία και σε αυτή την περίπτωση ταυτίζεται με την ερμηνεία που δώσαμε στις μέγιστες βροχοπτώσεις. Στο γράφημα που ακολουθεί η συνεχής γραμμή απεικονίζει την μηνιαία σημειακή εκτίμηση που δίνεται από την συνάρτηση \square

$$\hat{\sigma}(i) = \exp(\hat{\sigma}_0 + \hat{\sigma}_1 \sin(\frac{2\pi i}{12}) + \hat{\sigma}_2 \cos(\frac{2\pi i}{12})) \quad i = 1, 2, \dots, 12.$$

και η ασυνεχής γραμμή το άνω και κάτω όριο του δ.ε. που δίνεται μέσω της κανονικής προσέγγισης από τον τύπο $\hat{\sigma}(i) \pm z_{\alpha/2} \cdot \sqrt{\text{Var}(\hat{\sigma}(i))}$ για $i = 1, 2, \dots, 12$. Η διασπορά του εκτιμητή υπολογίζεται όπως και στην περίπτωση των βροχοπτώσεων.

Σχήμα 5.12

Γράφημα της εκτιμώμενης παραμέτρου $\sigma(i)$ και 95% δ.ε. για τις μέγιστες θερμοκρασίες, για $i = 1, 2, \dots, 12$ για τους 4 μετεωρολογικούς σταθμούς όπου η περιοδικότητα στην παράμετρο κλίμακος είναι στατιστικά σημαντική.



Στη συνέχεια ακολουθεί ο πίνακας με τις εκτιμώμενες μέσες μέγιστες θερμοκρασίες καθώς και το γράφημα της μέσης μηνιαίας μέγιστης θερμοκρασίας για τους 18 μετεωρολογικούς σταθμούς που αντιστοιχεί στην εκτιμημένη μέση τιμή της κατανομής *GEV* ή *Gumbel* ανάλογα σε ποια κατανομή ανήκουν τα δεδομένα. Οι τύποι για την μέση τιμή των κατανομών δόθηκαν στην προηγούμενη ενότητα. Η ερμηνεία και σε αυτή την περίπτωση είναι διαφορετική όσο αφορά σταθμούς όπου το μέγεθος της κλάσης είναι ίσο με 30

παρατηρήσεις και σταθμούς με μεγαλύτερο μέγεθος κλάσης από 30. Αντίστοιχες ερμηνείες όπως και στην περίπτωση των μέγιστων βροχοπτώσεων ισχύουν και εδώ.

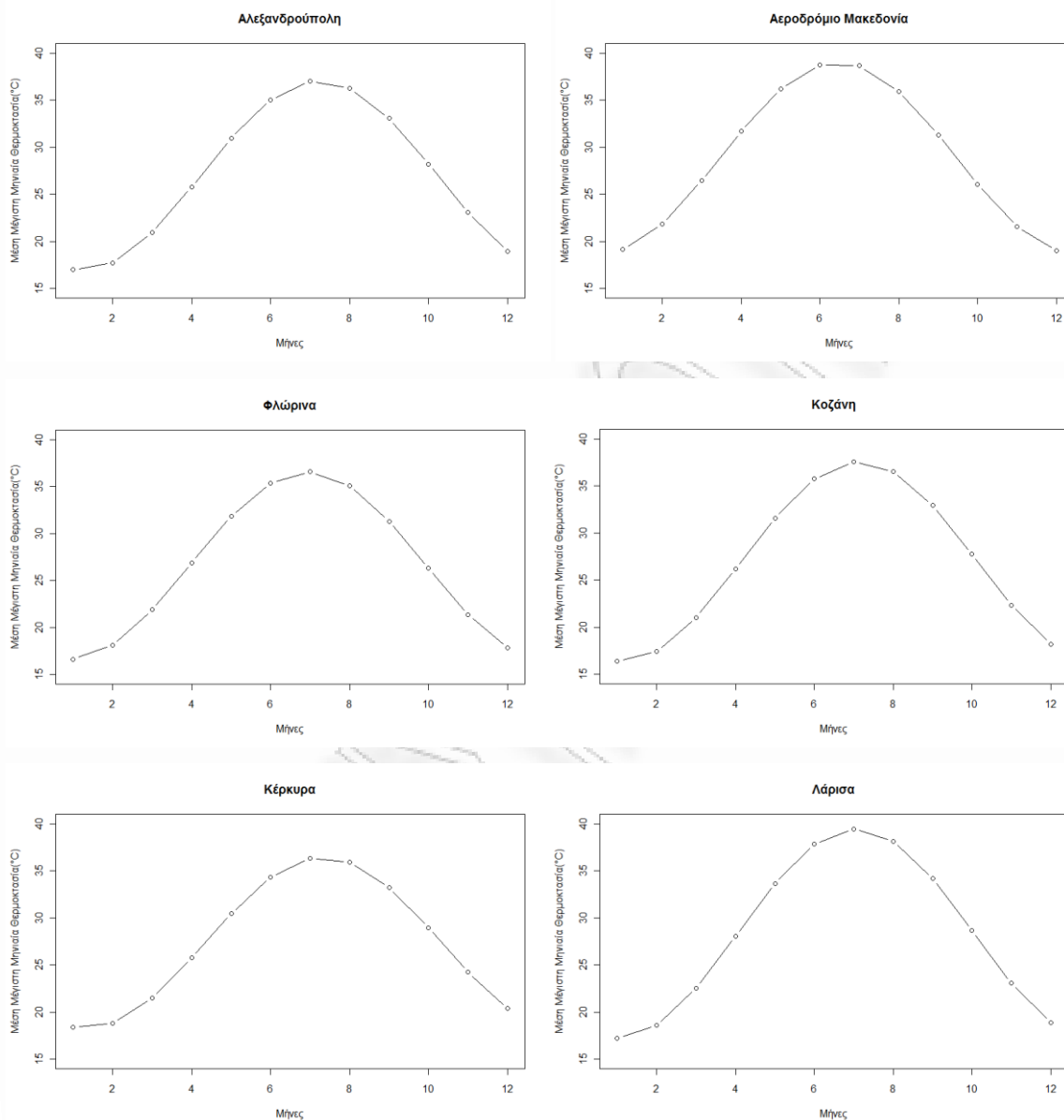
Πίνακας 5.9

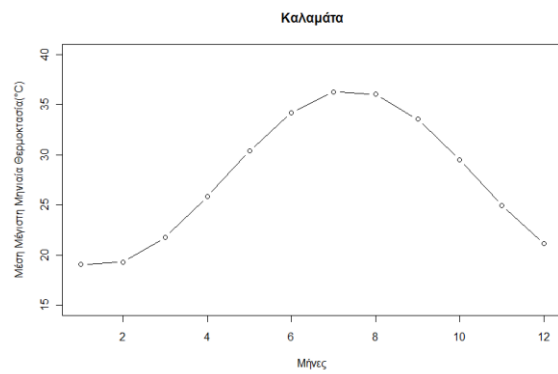
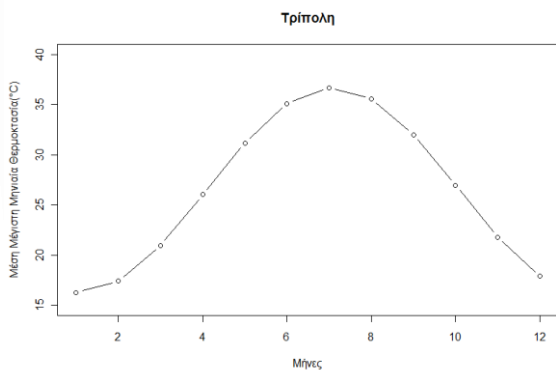
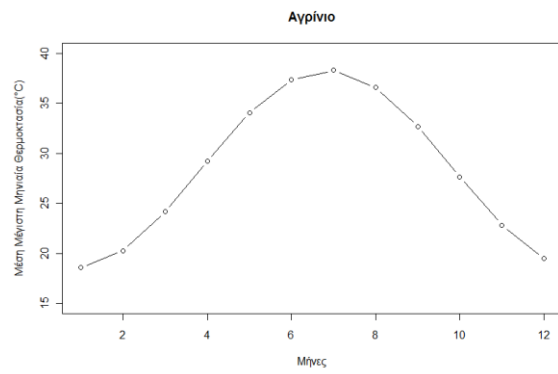
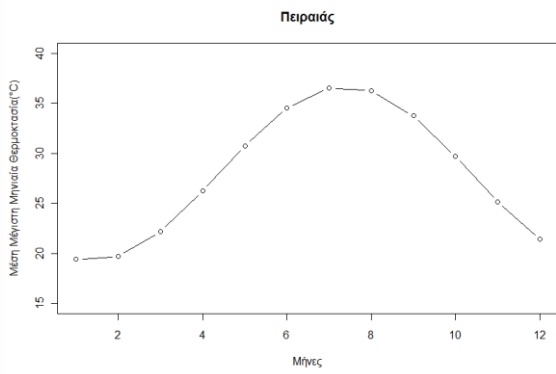
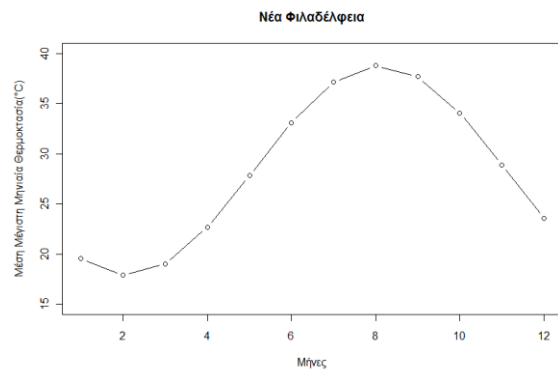
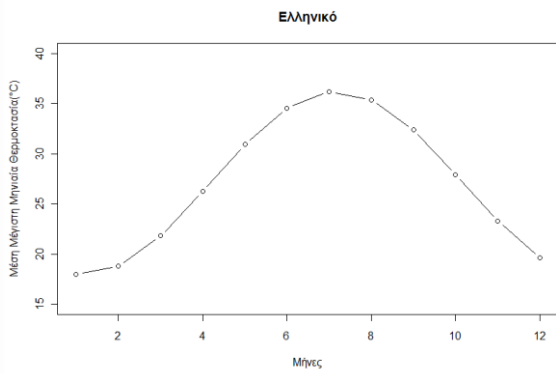
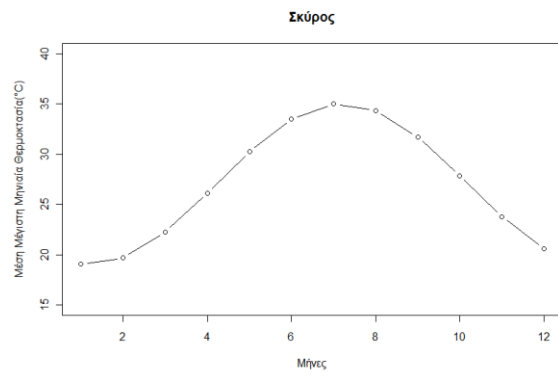
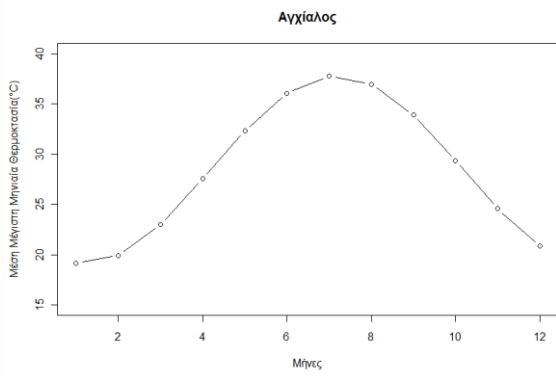
Πίνακας με τις εκτιμήσεις των αναμενόμενων μέγιστων θερμοκρασιών για κάθε μήνα του έτους και για κάθε μετεωρολογικό σταθμό.

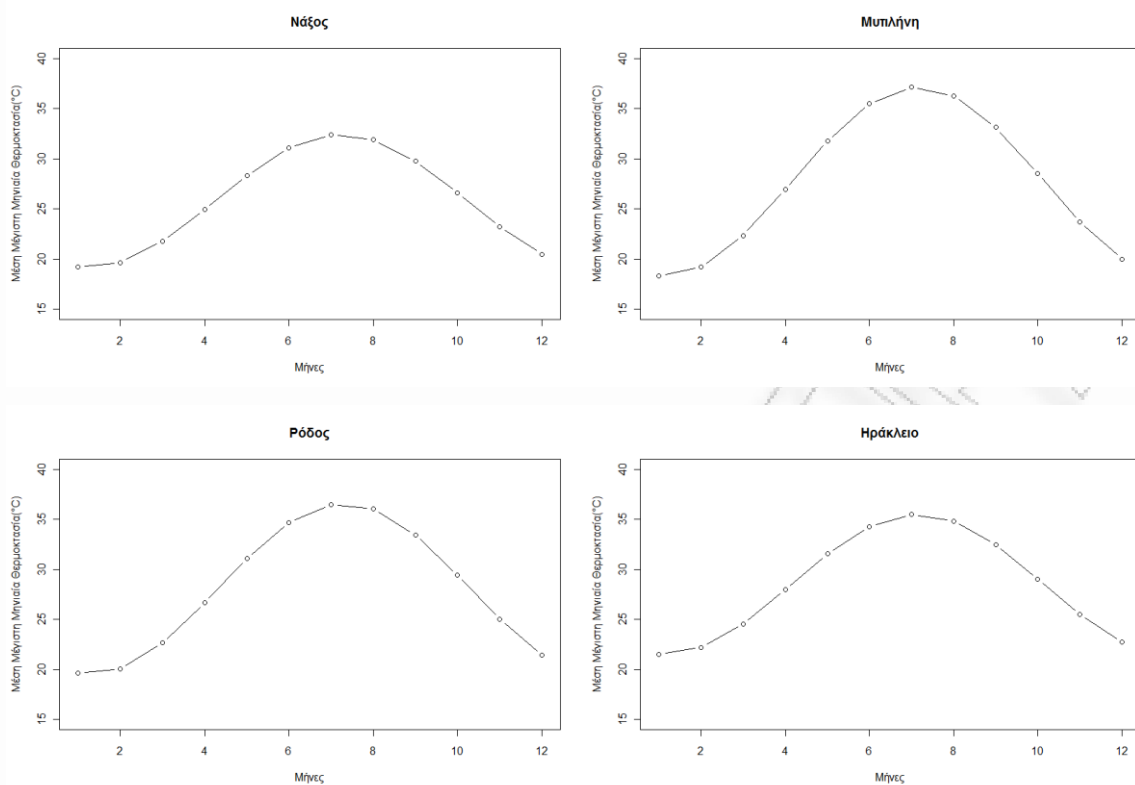
	Ιαν	Φεβ	Μαρ	Απρ	Μαι	Ιουν	Ιουλ	Αυγ	Σεπ	Οκτ	Νοεμ	Δεκ
1)	16.997	17.725	20.936	25.771	30.933	35.039	36.990	36.262	33.051	28.216	23.054	18.948
2)	19.119	21.834	26.436	31.693	36.195	38.737	38.637	35.923	31.321	26.064	21.562	19.020
3)	16.623	18.110	21.872	26.899	31.845	35.385	36.570	35.082	31.321	26.293	21.347	17.807
4)	17.122	18.104	21.664	26.848	32.267	36.469	38.328	37.346	33.786	28.602	23.183	18.981
5)	18.358	18.766	21.471	25.749	30.452	34.321	36.319	35.910	33.205	28.928	24.225	20.356
6)	17.246	18.588	22.546	28.060	33.653	37.826	39.461	38.120	34.161	28.647	23.054	18.881
7)	19.146	19.938	23.010	27.537	32.309	36.044	37.744	36.952	33.881	29.353	24.582	20.846
8)	19.073	19.682	22.260	26.128	30.246	33.493	34.985	34.329	31.720	27.868	23.796	20.580
9)	17.928	18.764	21.794	26.206	30.813	34.378	35.946	35.102	32.074	27.672	23.073	19.506
10)	19.559	17.889	19.021	22.653	27.809	33.110	37.135	38.804	37.672	34.041	28.884	23.583
11)	18.255	18.544	21.051	25.104	29.617	33.381	35.387	35.098	32.591	28.538	24.025	20.261
12)	18.571	20.281	24.175	29.210	34.035	37.359	38.291	36.581	32.687	27.653	22.827	19.503
13)	15.575	16.706	20.271	25.314	30.485	34.397	36.002	34.871	31.306	26.262	21.092	17.180
14)	19.146	19.397	21.905	25.999	30.582	34.425	36.499	36.248	33.740	29.646	25.063	21.220
15)	19.187	19.651	21.753	24.933	28.340	31.060	32.360	31.891	29.781	26.598	23.196	20.485
16)	17.847	18.713	21.866	26.462	31.268	34.998	36.651	35.785	32.632	28.036	23.230	19.500
17)	19.490	19.948	22.491	26.439	30.735	34.224	35.969	35.504	32.955	29.008	24.720	21.237
18)	20.768	21.437	23.799	27.220	30.783	33.535	34.737	34.067	31.706	28.285	24.721	21.970

Σχήμα 5.13

Γράφημα της εκτιμώμενης μέσης μηνιαίας μέγιστης θερμοκρασίας για $i = 1, 2, \dots, 12$ για καθένα από τους 18 μετεωρολογικούς σταθμούς.



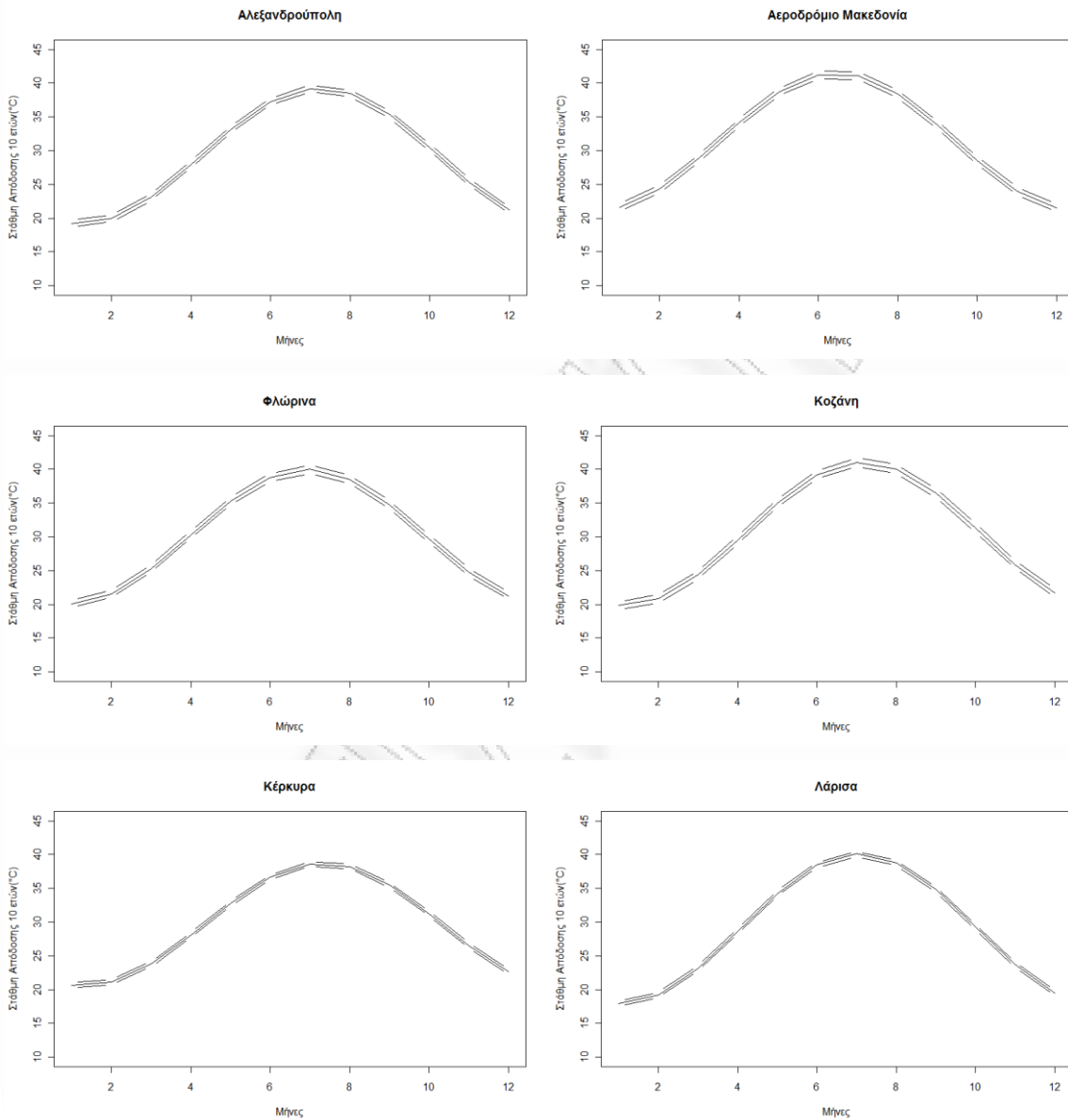


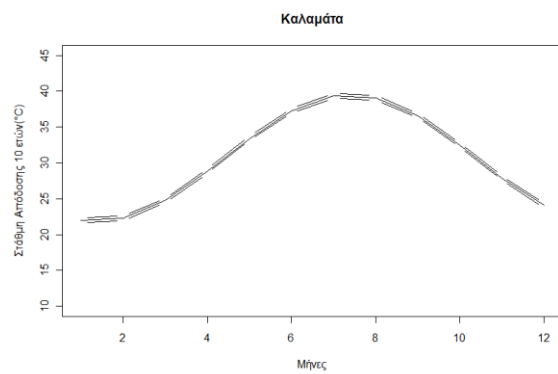
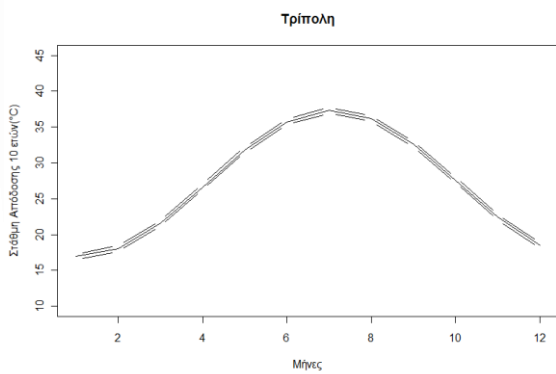
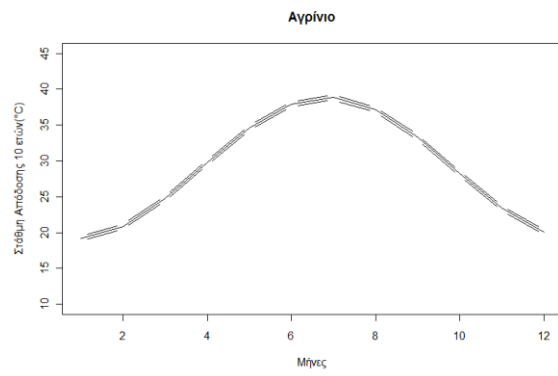
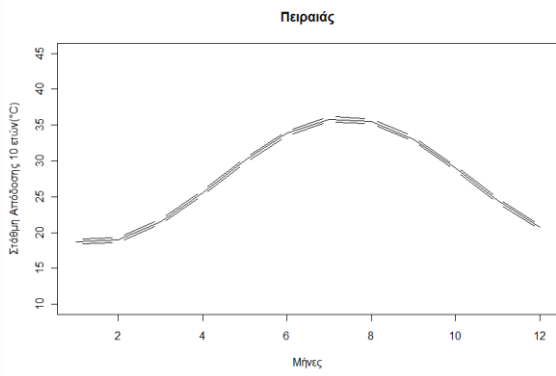
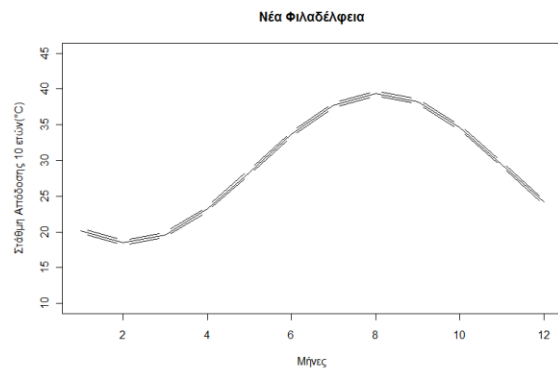
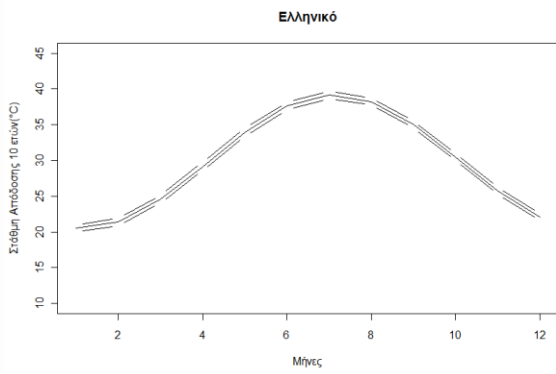
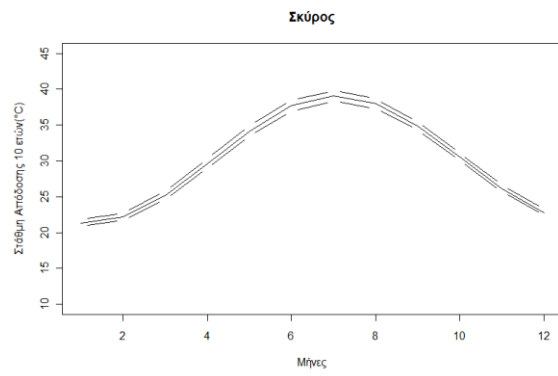
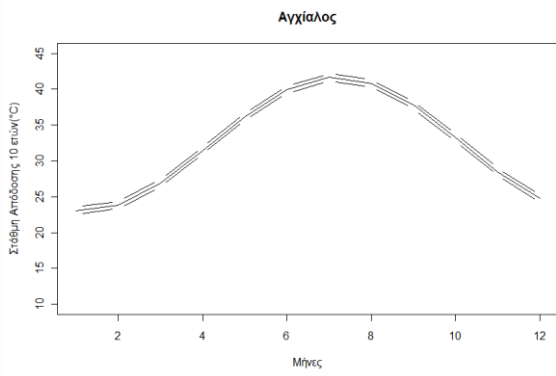


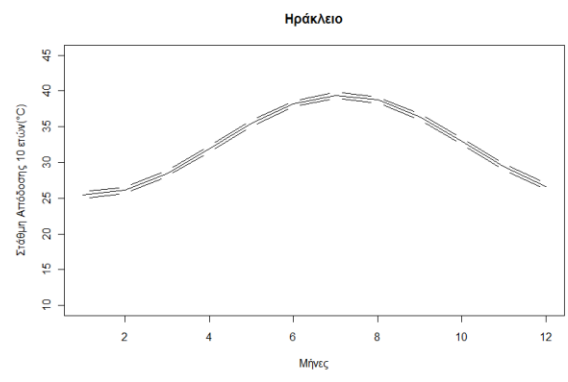
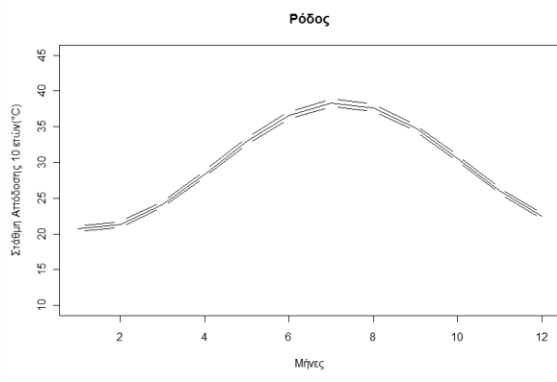
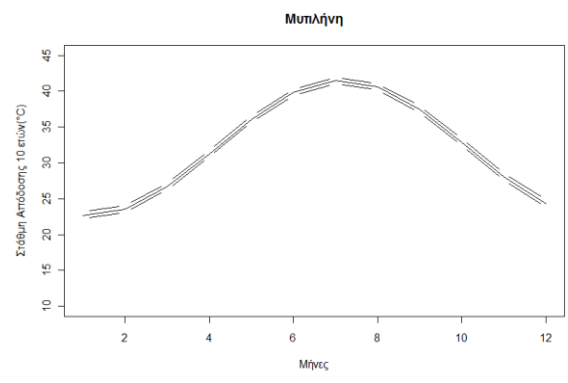
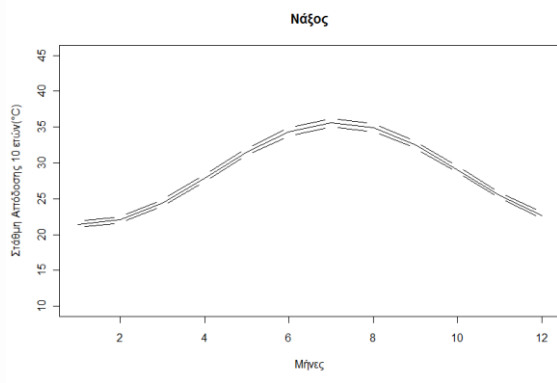
Στο σχήμα 5.14 απεικονίζονται τα γραφήματα των μηνιαίων σταθμών απόδοσης καθώς και τα προσεγγιστικά διαστήματα εμπιστοσύνης τάξεως 95%, για 10 και 50 έτη για τους 18 μετεωρολογικούς σταθμούς. Και σε αυτά τα γραφήματα όπως και στα γραφήματα που απεικονίζουν την στάθμη απόδοσης των μέγιστων βροχοπτώσεων έχουμε παράγει στάθμες απόδοσης για κάθε μήνα του έτους. Και εδώ όπως και στην περίπτωση των μέγιστων βροχοπτώσεων η πιθανότητα υπέρβασης διαφέρει ανάλογα το μέγεθος της κλάσης του κάθε σταθμού. Έτσι για παράδειγμα για τον μετεωρολογικό σταθμό του Ελληνικού η σημειακή εκτίμηση της στάθμης απόδοσης 50 ετών για τον μήνα Ιανουάριο είναι ίση με 22.39°C και το αντίστοιχο 95% δ.ε. ίσο με $[21.53, 23.25]$ που σημαίνει ότι κατά μέσο όρο η μέγιστη ημερήσια θερμοκρασία τον μήνα Ιανουάριο υπερβαίνει τους 22.39 βαθμούς Κελσίου μια φορά κάθε 50 Ιανουαρίους, και αντίστοιχα για το καλοκαίρι θα έχουμε ότι τον Ιούνιο η σημειακή εκτίμηση είναι ίση με 40.068°C που σημαίνει ότι η ημερήσια θερμοκρασία τον Ιούνιο υπερβαίνει τους 40 βαθμούς Κελσίου κατά μέσο όρο μία φορά κάθε 50 Ιουνίους, και το 95% δ.ε. ίσο με $[39.141, 40.99]$. Αντίστοιχη ερμηνεία δίνεται και για τους άλλους μήνες του έτους και για τους άλλους μετεωρολογικούς σταθμούς και για την στάθμη απόδοσης 10 ετών.

Σχήμα 5.14

Γράφημα μηνιαίας στάθμης απόδοσης για 10 έτη και 95% δ.ε. για τις μέγιστες θερμοκρασίες για τους 18 μετεωρολογικούς σταθμούς.

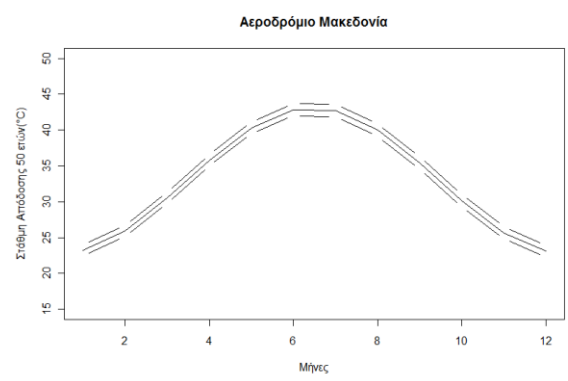
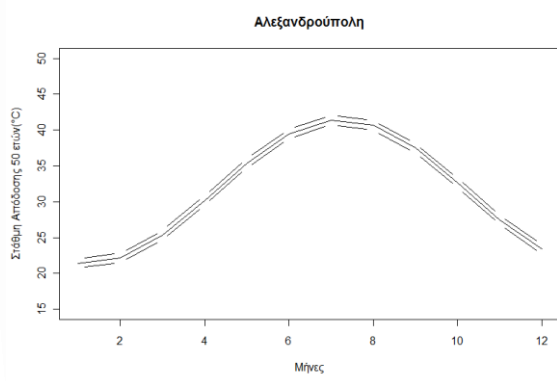


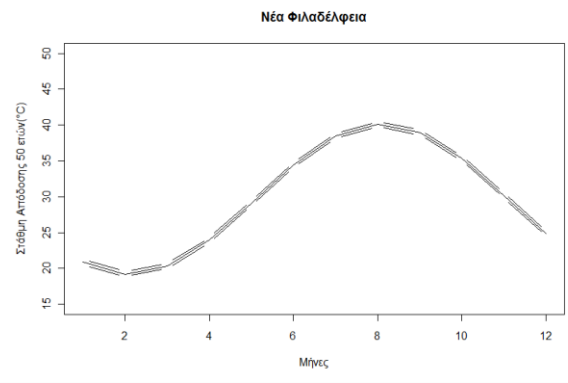
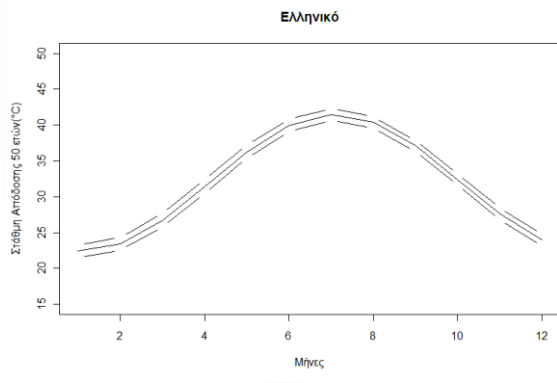
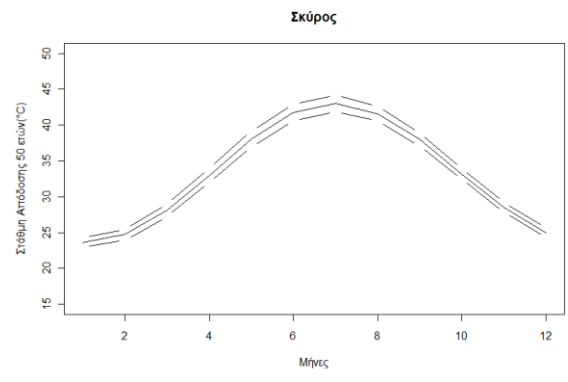
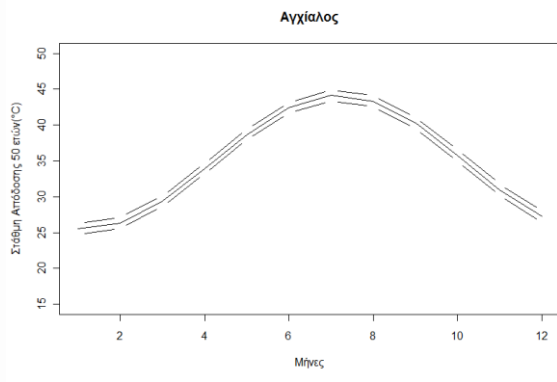
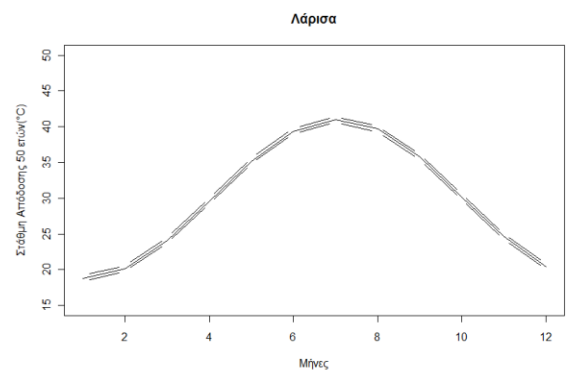
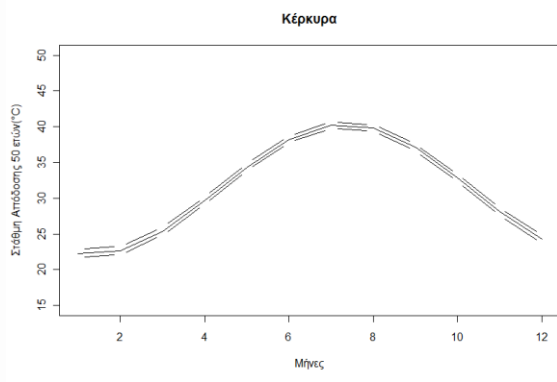
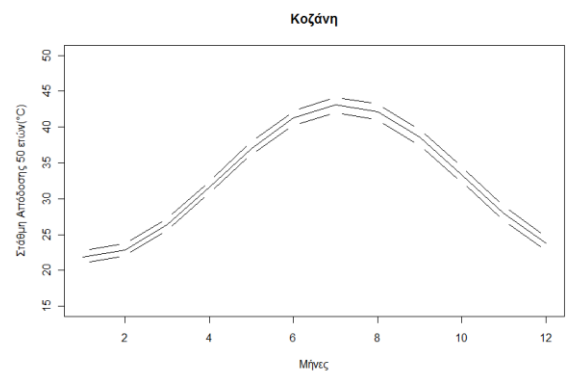
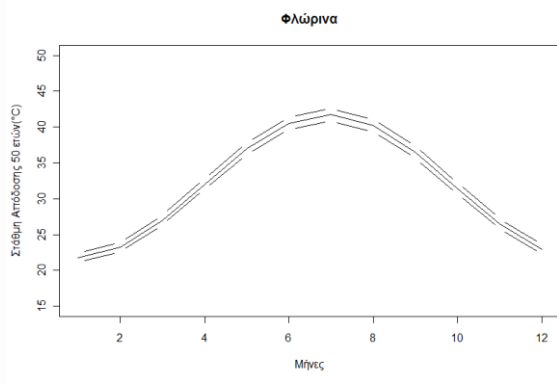


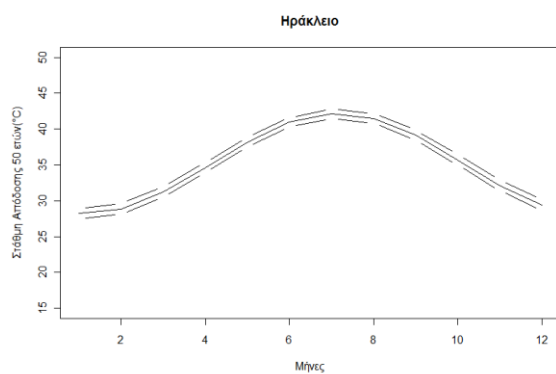
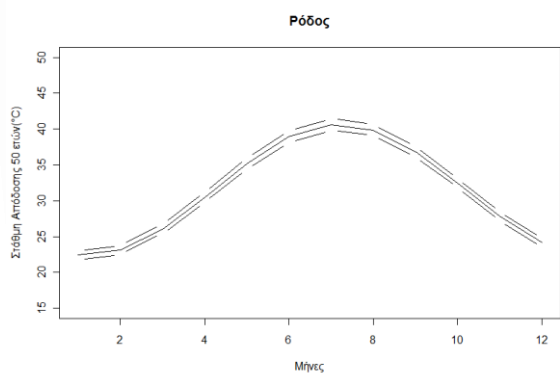
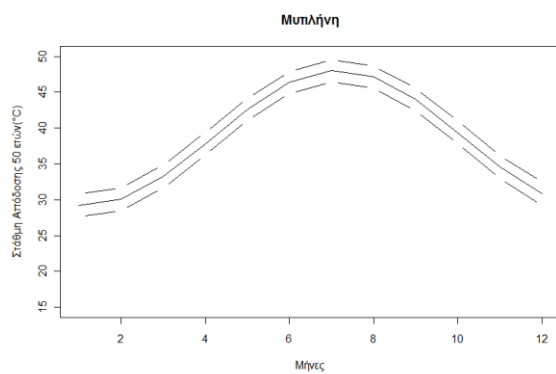
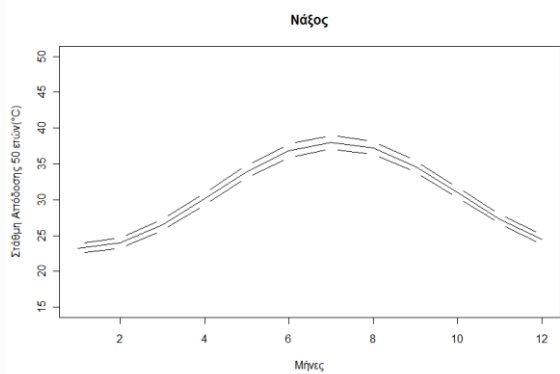
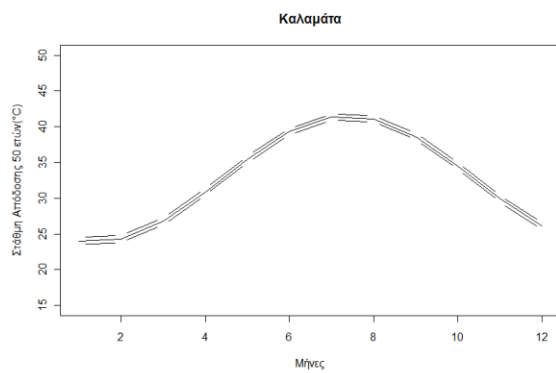
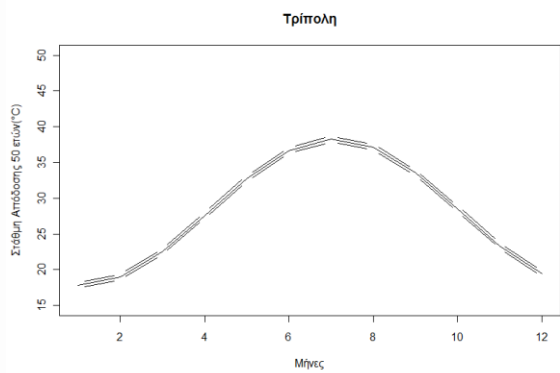
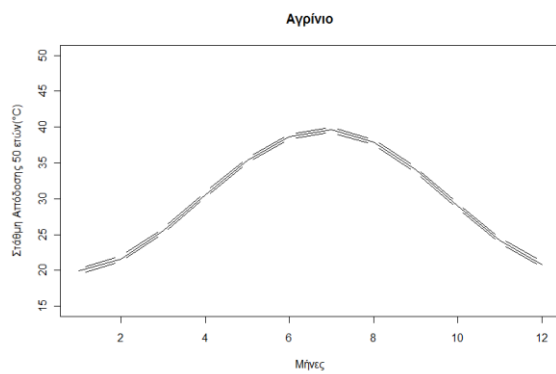
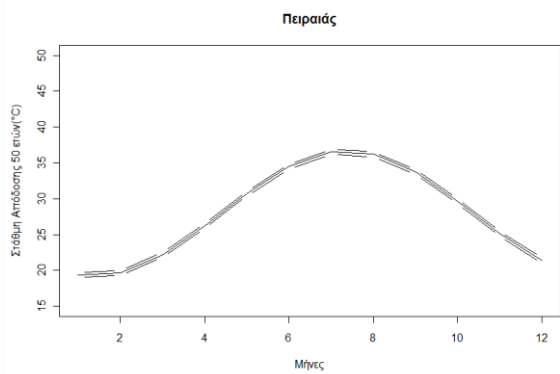


Σχήμα 5.15

Γράφημα μηνιαίας στάθμης απόδοσης 50 ετών και 95% δ.ε. για τις μέγιστες θερμοκρασίες για τους 18 μετεωρολογικούς σταθμούς.



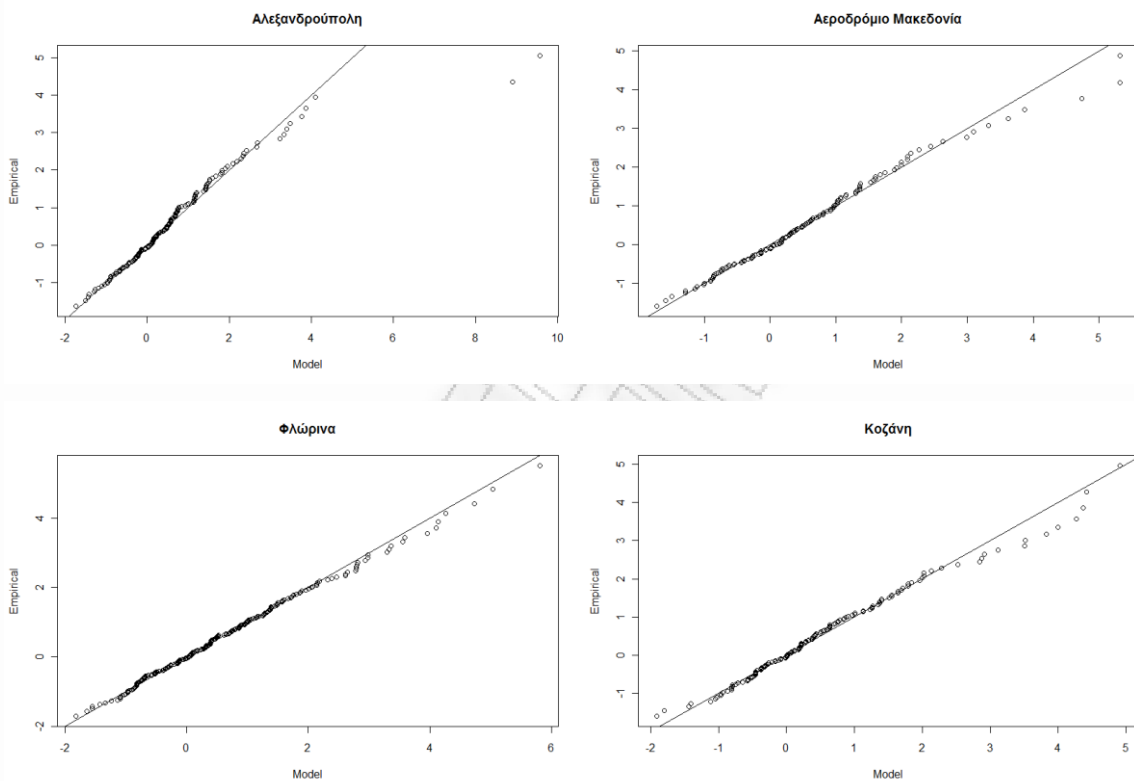


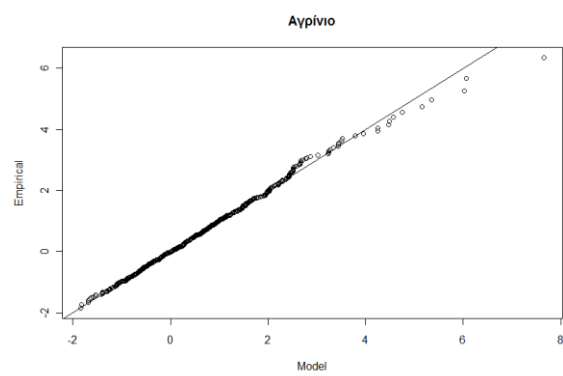
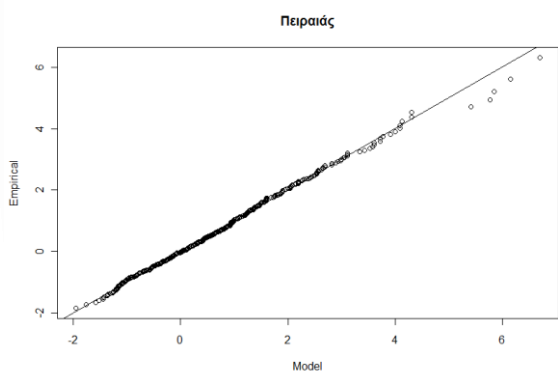
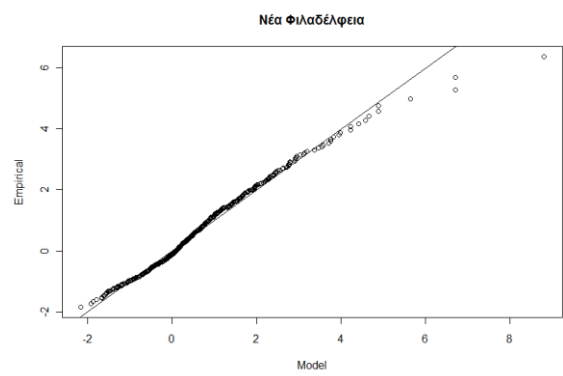
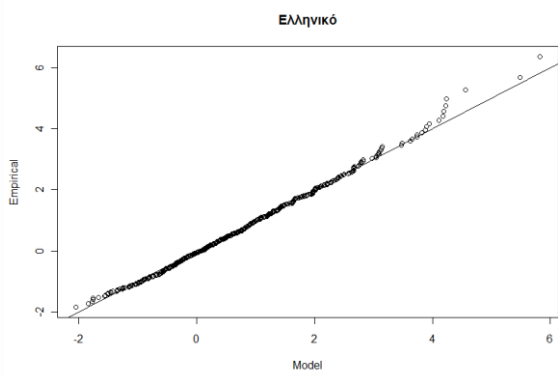
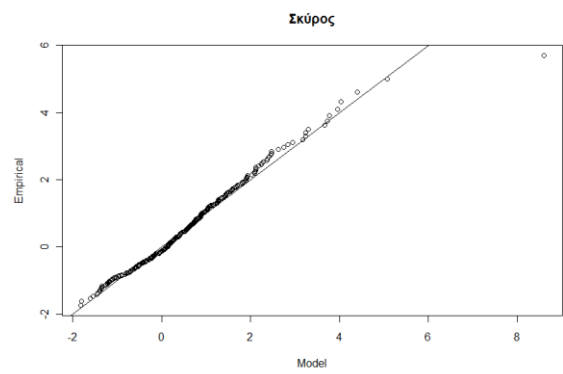
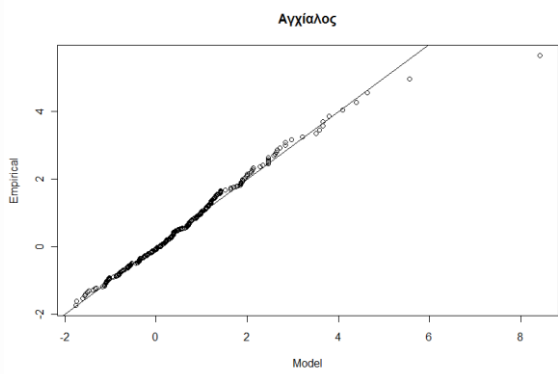
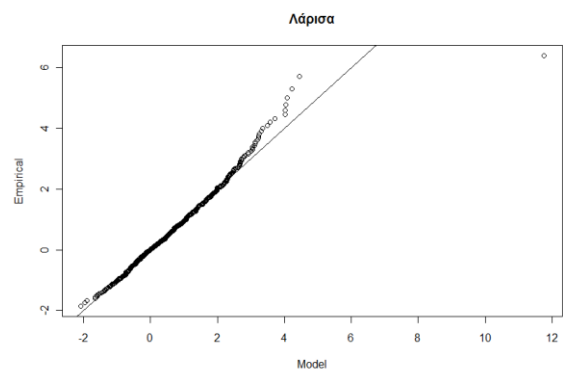
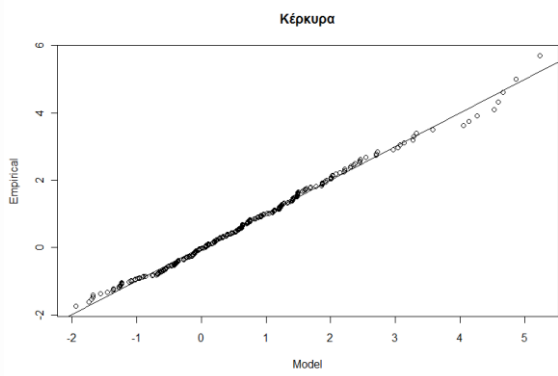


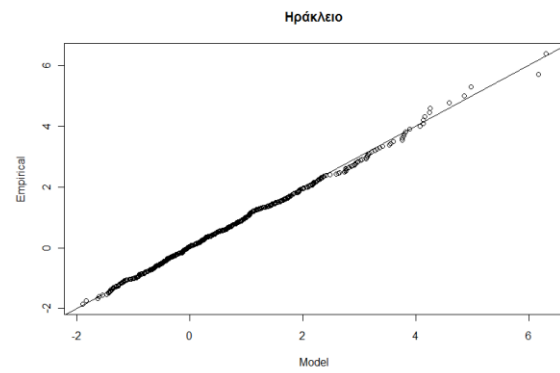
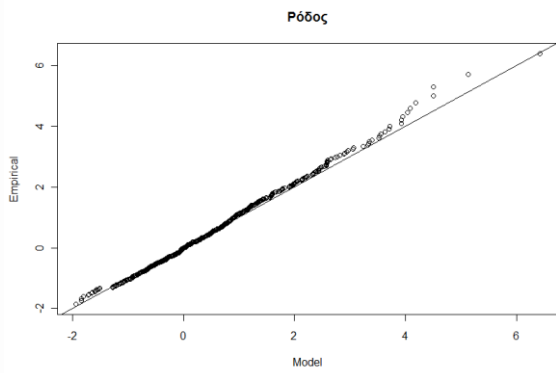
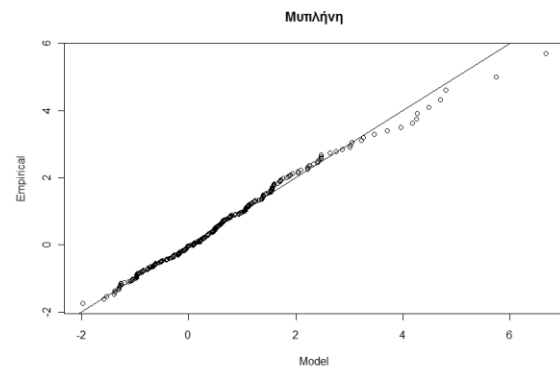
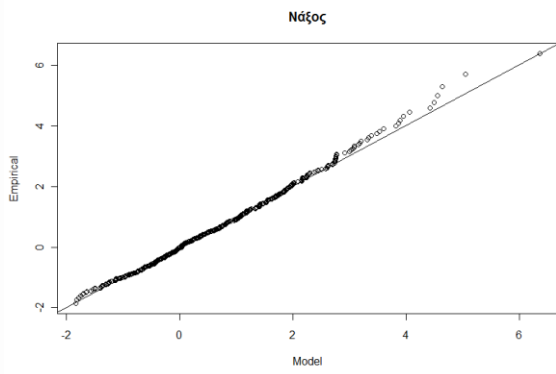
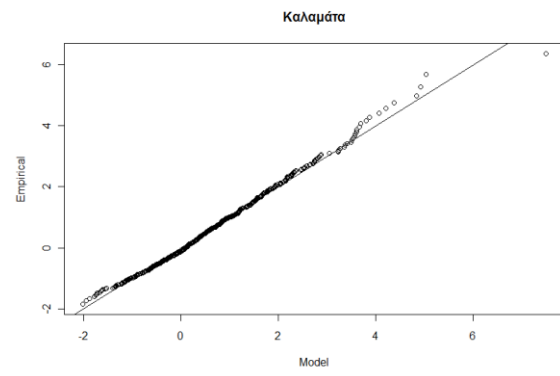
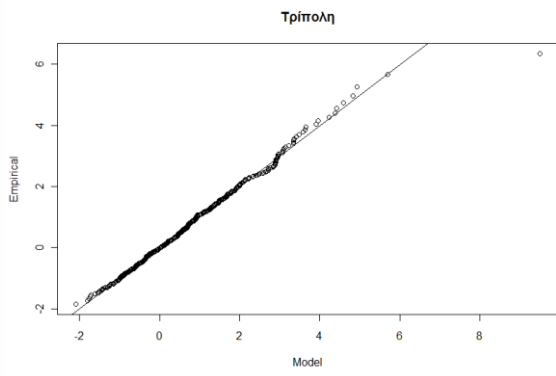
Ο έλεγχος για το αν τα δεδομένα των μέγιστων θερμοκρασιών προσαρμόζονται καλά στο μοντέλο θα γίνει μέσω κατάλληλων Q-Q Plots που έχουν περιγραφεί στην Ενότητα 3.3.7 όπου οι αρχικές μη-στάσιμες παρατηρήσεις τυποποιούνται έτσι ώστε να προέρχονται από μια τυπική κατανομή *Gumbel* ακριβώς όπως και στην περίπτωση των μέγιστων βροχοπτώσεων.. Κατασκευάζουμε τα επόμενα γραφήματα και παρατηρούμε ότι τα δεδομένα προσαρμόζονται ικανοποιητικά στους σταθμούς που έχουν επιλεγεί.

Σχήμα 5.16

Q-Q Plots μέγιστων θερμοκρασιών για τους 18 μετεωρολογικούς σταθμούς.







ΠΑΝΕΠ

ΚΕΦΑΛΑΙΟ 6

Συμπεράσματα

6.1 Ποιοτικά συμπεράσματα

Η στατιστική μέθοδος που επιλέξαμε για να μοντελοποιήσουμε τα περιβαλλοντολογικά ακραία φαινόμενα μέσω της επιλογής μηνιαίων κλάσεων μεγίστων και όχι ετήσιων κλάσεων, παρουσιάζει αρκετά πλεονεκτήματα. Αρχικά μέσω της εφαρμογής της παραπάνω μεθόδου καταφέραμε να μοντελοποιήσουμε τον ετήσιο κλιματικό κύκλο των μεγίστων βροχοπτώσεων και μεγίστων θερμοκρασιών και να γίνουν ορατές οι ιδιαιτερότητες και οι διαφορές που εμφανίζουν τα δύο αυτά φυσικά φαινόμενα στον Ελλαδικό χώρο. Επίσης μέσω της χρησιμοποίησης της παραπάνω μεθόδου εκμεταλλευτήκαμε καλύτερα το σύνολο των δεδομένων που είχαμε στην διάθεση μας, αντλώντας περισσότερη πληροφορία από τα δεδομένα, και καταλήξαμε σε ασφαλέστερες προβλέψεις για μακροπρόθεσμα ακραία μετεωρολογικά φαινόμενα. Κλείνοντας με τα πλεονεκτήματα της μεθόδου αξίζει να αναφέρουμε ότι καταφέραμε να παράγουμε στάθμες απόδοσης για κάθε μήνα του έτους, κάνοντας έτσι τις προβλέψεις συγκεκριμένες για κάθε μήνα, λύνοντας με αυτό τον τρόπο το πρόβλημα που εμφανίζονταν με τις ετήσιες στάθμες απόδοσης όπου το κατώφλι που εκτιμούσαμε δεν γνωρίζαμε ποια χρονική περίοδο του έτους ήταν πιθανόν να εμφανιστεί. Όσο αφορά τις ιδιαιτερότητες της μεθόδου οι σημαντικότερες είναι οι εξής □ με την παραπάνω μεθοδολογία πήραμε μικρότερες παρατηρήσεις στο δείγμα μας, λόγω του ότι πήραμε μέγιστες παρατηρήσεις σε περιόδους όπου το φαινόμενα της μέγιστης βροχόπτωσης και της μέγιστης θερμοκρασίας δεν ήταν σε έξαρση και έτσι πήραμε μικρότερες εκτιμήσεις για τις παραμέτρους. Επίσης και οι τάσεις που εμφανίστηκαν σε ορισμένους μετεωρολογικούς σταθμούς ήταν πολύ μικρές λόγω και της φύσεως των δεδομένων αφού η περιοδικότητα των δεδομένων υπερίσχυε των τάσεων. Τέλος δεν χρησιμοποιήσαμε την μέθοδο ανάλυσης ακραίων παρατηρήσεων POT με την παραπάνω μοντελοποίηση λόγω των προβλημάτων που θα εμφάνιζε η μέθοδος εξαιτίας της εξάρτησης που θα παρουσίαζαν οι υπερβάσεις (παρατηρήσεις) που θα ξεπερνούσαν πάνω από το κατώφλι που θα είχαμε επιλέξει.

6.2 Ποσοτικά συμπεράσματα

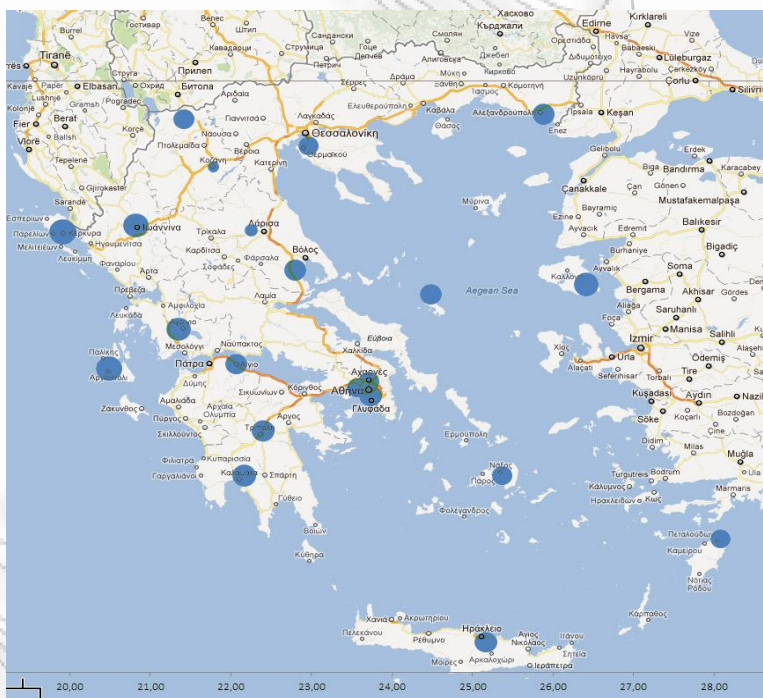
6.2.1 Μέγιστες βροχοπτώσεις

Αρχικά το πρώτο συμπέρασμα για τις μέγιστες βροχοπτώσεις παρατηρώντας την ανάλυση που πραγματοποιήθηκε στο Κεφάλαιο 5 αφορά τις διαφορές στο μέγεθος των κλάσεων (blocks) στους μετεωρολογικούς σταθμούς ανά τον Ελλαδικό χώρο, κάτι που θα παίζει σημαντικό ρόλο στη συνέχεια της ανάλυσης. Πιο συγκεκριμένα οι μετεωρολογικοί σταθμοί που βρίσκονται στην Βόρεια Ελλάδα και στην Θεσσαλία προσαρμόζονται αρκετά ικανοποιητικά στις οριακές κατανομές ακροτάτων για μέγεθος κλάσης ίσο με 30 παρατηρήσεις, ενώ οι μετεωρολογικοί σταθμοί που βρίσκονται στην Κεντρική Ελλάδα στην Πελοπόννησο και στα νησιά του Αιγαίου προσαρμόζονται ικανοποιητικά στις οριακές κατανομές ακροτάτων για μεγαλύτερο μέγεθος κλάσης της τάξης των 60-120 παρατηρήσεων. Ένας από τους λόγους που υπάρχει αυτή η διαφορά στο μέγεθος των κλάσεων στους μετεωρολογικούς σταθμούς είναι το ότι στις περιοχές της Βόρειας Ελλάδας και της Θεσσαλίας υπάρχουν βροχοπτώσεις και κατά την διάρκεια των καλοκαιρινών μηνών κάτι που δεν ισχύει για τις υπόλοιπες περιοχές όπου οι βροχοπτώσεις εκείνο το διάστημα είναι ελάχιστες ως μηδαμινές, παράγοντας έτσι αρκετές παρατηρήσεις ίσες με το μηδέν κάτι που δυσκολεύει την προσαρμογή των κατανομών ακροτάτων. Με την προσαρμογή της πιθανότητας υπέρβασης στο μέγεθος της κλάσης το κάθε σταθμού, πήραμε στάθμες απόδοσης οι οποίες είναι απολύτως συγκρίσιμες μεταξύ τους, παράγοντας με αυτό τον τρόπο αρκετά χρήσιμα συμπεράσματα. Επίσης κατά την εκτίμηση των παραμέτρων των κατανομών των μέγιστων βροχοπτώσεων εμφανίζονται μεγάλα τυπικά σφάλματα των εκτιμητών, κάτι που επιδρά στην συνέχεια και στην κατασκευή των δ.ε. για την παράμετρο θέσης και την παράμετρο κλίμακος καθώς και για την στάθμη απόδοσης. Παρατηρούμε από τους πίνακες 5.3 και 5.4 μια πολύ μικρή τάση μείωσης των μέγιστων βροχοπτώσεων, κατά την χρονική περίοδο που μελετάμε, στους σταθμούς της Αλεξανδρούπολης και της Σκύρου, ενώ αντίθετα πολύ μικρή τάση αύξησης των μέγιστων βροχοπτώσεων παρουσιάζει ο σταθμός του Αιγίου. Έστω και μικρές οι τάσεις αυτές φανερώνουν αλλαγή των κλιματολογικών συνθηκών όσο αφορά τις συγκεκριμένες περιοχές, η ερμηνεία όμως των αλλαγών αυτών ξεφεύγει από τους σκοπούς της διπλωματικής αυτής εργασίας καθώς απαιτούνται εξειδικευμένες κλιματολογικές γνώσεις. Η παράμετρος σχήματος ξ που δείχνει την συμπεριφορά της δεξιάς ουράς της κατανομής μεγίστων, είναι θετική στην πλειονότητα των μετεωρολογικών

σταθμών, κάτι που δείχνει ότι οι κατανομές ακροτάτων των μέγιστων βροχοπτώσεων έχουν βαριές ουρές. Επίσης όσο αφορά τις προβλέψεις για τη στάθμη απόδοσης, ακραίες βροχοπτώσεις εκτιμάται ότι θα έχουν κυρίως οι σταθμοί στα Ιωάννινα, στην Κέρκυρα, στο Αγρίνιο, στο Αργοστόλι και στην Μυτιλήνη. Εν κατακλείδι η ανάλυση που παρουσιάστηκε στο Κεφάλαιο 5 για τις μέγιστες βροχοπτώσεις μας έδωσε μία σαφή εικόνα του ετήσιου κλιματικού κύκλου των μέγιστων βροχοπτώσεων και μέσω των κατανομών ακροτάτων που προσαρμόσαμε σε κάθε σταθμό πήραμε εκτιμήσεις για την μηνιαία στάθμη απόδοσης για 10 και 50 έτη. Τέλος παρουσιάζουμε δύο χάρτες του Ελλαδικού χώρου όπου πάνω έχουν απεικονιστεί οι 21 μετεωρολογικοί σταθμοί καθώς και η εκτίμηση της στάθμης απόδοσης 50 ετών για τις μέγιστες βροχοπτώσεις για τους μήνες Ιανουάριο και Ιούλιο.

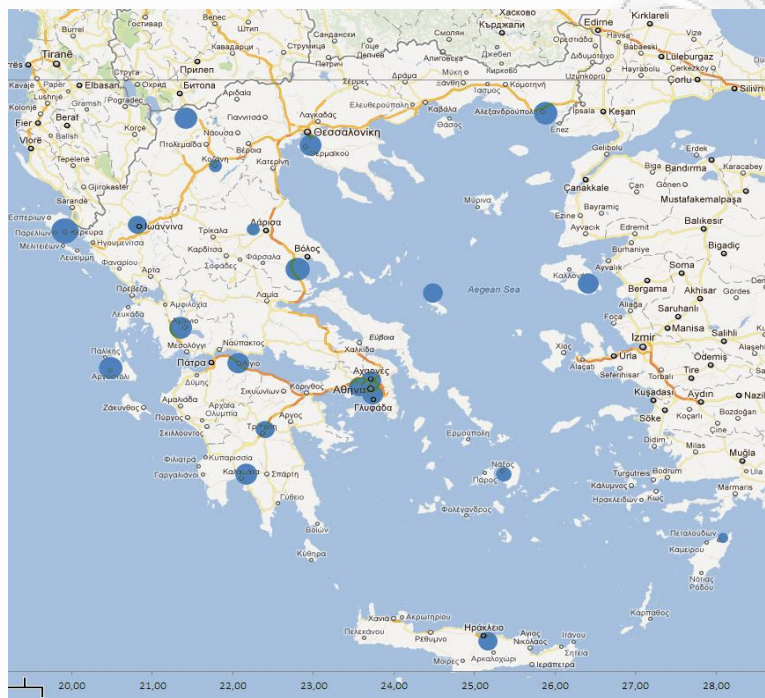
Σχήμα 6.1

Εκτιμήσεις στάθμης απόδοσης 50 ετών για τις μέγιστες βροχοπτώσεις, για το μήνα Ιανουάριο.



Σημα 6.2

Εκτιμήσεις στάθμης απόδοσης 50 ετών για τις μέγιστες βροχοπτώσεις, για το μήνα Ιούλιο.



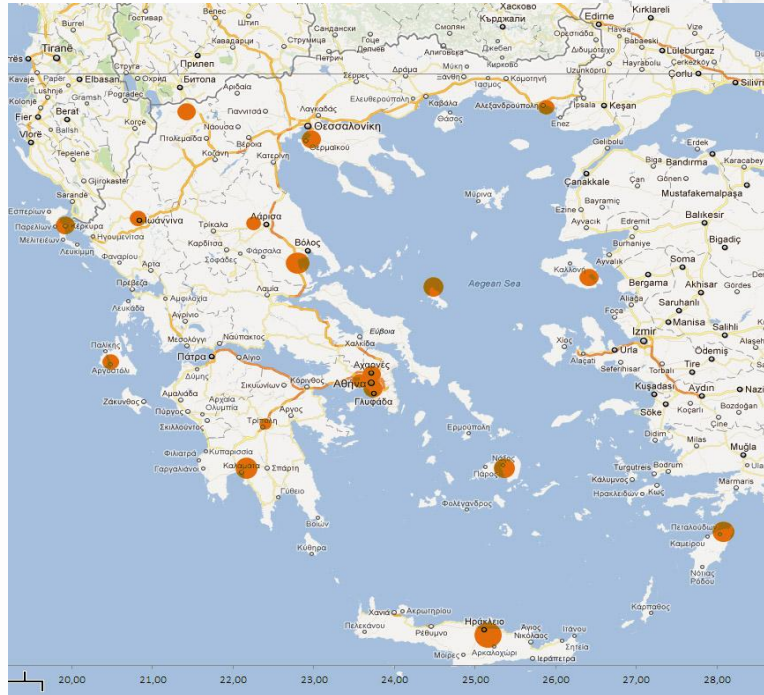
6.2.2 Μέγιστες θερμοκρασίες

Αντίστοιχο φαινόμενο, όπως και στις μέγιστες βροχοπτώσεις, όσο αφορά τις διαφορές στο μέγεθος των κλάσεων των μετεωρολογικών σταθμών (έτσι ώστε να έχουμε ικανοποιητική προσαρμογή στις κατανομές ακροτάτων στους μετεωρολογικούς σταθμούς) συναντάμε και στην περίπτωση των μέγιστων θερμοκρασιών. Πιο συγκεκριμένα οι μετεωρολογικοί σταθμοί που βρίσκονται στην Κεντρική Ελλάδα, την Πελοπόννησο, και στα νησιά του Κεντρικού Αιγαίου και της Κρήτης προσαρμόζονται αρκετά ικανοποιητικά στις οριακές κατανομές ακροτάτων για μέγεθος κλάσης ίσο με 30 παρατηρήσεις, ενώ οι μετεωρολογικοί σταθμοί που βρίσκονται στην Βόρεια Ελλάδα και το Βόρειο Αιγαίο προσαρμόζονται ικανοποιητικά στις οριακές κατανομές ακροτάτων για μεγαλύτερο μέγεθος κλάσης της τάξης των 60-120 παρατηρήσεων. Ένας από τους λόγους που υπάρχει αυτή η διαφορά στο μέγεθος των κλάσεων στους μετεωρολογικούς σταθμούς είναι ότι κατά την διάρκεια των χειμώνων στα

νησιά του Βορείου Αιγαίου και στις περιοχές της Βόρειας Ελλάδος έχουμε πολύ χαμηλές θερμοκρασίες με συνέπεια στο δείγμα αυτών των μετεωρολογικών σταθμών να υπάρχουν πολύ μικρές παρατηρήσεις κάτι που δυσκολεύει την προσαρμογή των κατανομών ακροτάτων. Από τους πίνακες των εκτιμήσεων των παραμέτρων για τις μέγιστες θερμοκρασίες 5.6 και 5.7, παρατηρούμε ότι η εποχικότητα, που φαίνεται μέσα από τις εκτιμήσεις των παραμέτρων μ_1 και μ_2 είναι πιο έντονη στους μετεωρολογικούς σταθμούς της Βόρειας Ελλάδος λόγω των διαφορών που υπάρχουν στις μέγιστες θερμοκρασίες από το χειμώνα στο καλοκαίρι ενώ αντίθετα στις νησιωτικές περιοχές η περιοδικότητα δεν είναι τόσο έντονη. Τάση μείωσης των μηνιαίων μέγιστων θερμοκρασιών παρουσιάζει μόνο ο σταθμός της Ρόδου, όμως αν και η τάση είναι στατιστικά σημαντική, σε επίπεδο σημαντικότητας 5%, είναι πολύ μικρή, δεν παύει όμως να αποτελεί ένδειξη αλλαγής του κλίματος στην περιοχή αυτή. Οι προβλέψεις που αφορούν τις μηνιαίες στάθμες απόδοσης, που μέσω της τροποποίησης των πιθανοτήτων υπέρβασης ανάλογα το μέγεθος της κλάσης του κάθε σταθμού είναι απόλυτα συγκρίσιμες μεταξύ τους, για 10 έως 50 χρόνια είναι ομοιογενείς σε όλους τους μετεωρολογικούς σταθμούς χωρίς να υπάρχει καμιά ιδιαίτερη πρόβλεψη για κάποια πολύ μεγάλη θερμοκρασία σε κάποιο από τους σταθμούς. Κλείνοντας όπως και στην περίπτωση των μέγιστων βροχοπτώσεων μέσω της ανάλυσης των μέγιστων θερμοκρασιών στο Κεφάλαιο 5 καταφέραμε να μοντελοποιήσουμε τον ετήσιο κλιματικό κύκλο των μέγιστων θερμοκρασιών και να παράγουμε μηνιαίες στάθμες απόδοσης για 10 και 50 έτη. Τέλος παρουσιάζουμε δύο χάρτες του Ελλαδικού χώρου όπου πάνω έχουν απεικονιστεί οι 18 μετεωρολογικοί σταθμοί καθώς και η εκτίμηση της στάθμης απόδοσης 50 ετών για τις μέγιστες θερμοκρασίες για τους μήνες Ιανουάριο και Ιούλιο.

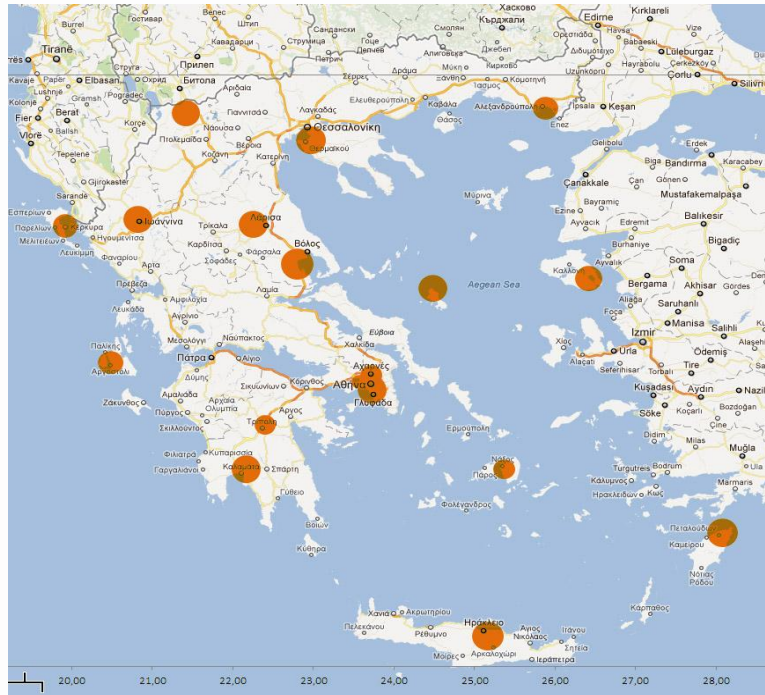
Σχήμα 6.3

Εκτιμήσεις στάθμης απόδοσης 50 ετών για τις μέγιστες θερμοκρασίες, για το μήνα Ιανουάριο.



Σχήμα 6.4

Εκτιμήσεις στάθμης απόδοσης 50 ετών για τις μέγιστες θερμοκρασίες, για το μήνα Ιούλιο.



Παράρτημα

Κώδικας 1 □ Με το ακόλουθο script έγινε η εμπειρική αξιολόγηση της μεθόδου Block Maxima στο κεφάλαιο 4.

Για τον υπολογισμό των εκτιμήσεων χρησιμοποιήσαμε την συνάρτηση mle του πακέτου stats4 η οποία ελαχιστοποιεί την $-\log L$ ως προς τις παραμέτρους της κατανομής.

```
library(stats4)
m1ist0=rep(1:1000)
m1ist1=rep(1:1000)
m1ist2=rep(1:1000)
m1ist3=rep(1:1000)
s1ist=rep(1:1000)
x1ist=rep(1:1000)
for(n in 1:4)
{
for(j in 1:1000)
{
m0=15;m1=2;m2=3;m3= -4;str=8;xtr=0.2;m=12;
k=1000/n;
hktr=rep(1:k);simbm=rep(1:k);t=1:k;
for(i in 1:k)
{hktr[i]=runif(1,0,1)
simbm[i]=(m0+m1*sin(2*pi*t[i]/m)+m2*cos(2*pi*t[i]/m)+m3*i/1000)-((str/xtr)*(1-(-log(hktr[i]))^(-xtr)))
}
n2=length(simbm);t=1:n2
mle12=function(mu0,mu1,mu2,mu3,lsu,xu)
```

```

(if(prod(1+(xu*(simbm-
(mu0+mu1*sin(2*pi*t/m)+mu2*cos(2*pi*t/m)+mu3*t/1000))/exp(lsu)))>0)==1)n2*lsu else
100000)+(if(prod(1+(xu*(simbm-
(mu0+mu1*sin(2*pi*t/m)+mu2*cos(2*pi*t/m)+mu3*t/1000))/exp(lsu)))>0)==1)sum(((1+(1/xu))*log(1+(
xu*(simbm-(mu0+mu1*sin(2*pi*t/m)+mu2*cos(2*pi*t/m)+mu3*t/1000))/exp(lsu)))))+
sum((1+(xu*(simbm-(mu0+mu1*sin(2*pi*t/m)+mu2*cos(2*pi*t/m)+mu3*t/1000))/exp(lsu))))^(-
(1/xu))) else 1000000)
a=mean(simbm)-0.57722*(sqrt((6*var(simbm)))/pi)
b=sqrt((6*var(simbm)))/pi
c=0.1
modell=mle(mle12, start=list(mu0=a, mu1=0, mu2=, mu3=1, lsu=log(b), xu=c))
m1list0[j]=coef(modell)[1]
m1list1[j]=coef(modell)[2]
m1list2[j]=coef(modell)[3]
m1list3[j]=coef(modell)[4]
s1list[j]=coef(modell)[5]
x1list[j]=coef(modell)[6]
}
}
hist(m1list0, seq(10, 22, 0.1), prob=TRUE, ylim=c(0, 0.8), xlab="x", main="μ0 estimation (sample
size=1000/n)")
lines(density(m1list3))
rug(m1list3)
hist(m1list1, seq(-3, 7, 0.1), prob=TRUE, ylim=c(0, 1.4), main="μ1 estimation (sample size=1000/n)")
lines(density(m1list3))
rug(m1list3)
hist(m1list2, seq(-2, 8, 0.1), prob=TRUE, ylim=c(0, 1.4), main="μ2 estimation (sample size=1000/n)")
lines(density(m1list2))
rug(m1list2)
m6=m1list3/1000
hist(m6, seq(-0.1, 0.1, 0.004), prob=TRUE, ylim=c(0, 130), xlab="x", main="μ3 estimation (sample
size=1000/m)")
lines(density(m6, bw=0.004))
rug(m6)
s4=exp(s1list)
hist(s4, seq(4, 12, 0.1), prob=TRUE, ylim=c(0, 2), xlab="x", main="σ0 estimation (sample size=1000/n)")
lines(density(s4))
rug(s4)
hist(x1list, seq(-0.2, 0.6, 0.01), prob=TRUE, ylim=c(0, 20), xlab="x", main="ξ estimation (sample
size=1000/n)")
lines(density(x1list))
rug(x1list)

```

Σημείωση □ Για τον υπολογισμό των εκτιμήσεων των παραμέτρων των κατανομών ακροτάτων για τις μέγιστες βροχοπτώσεις και θερμοκρασίες χρησιμοποιήσαμε την συναρτήρηση mle του πακέτου stats4.

Κώδικας 2 □ Το επόμενο script υπολογίζει την παράμετρο θέσεως καθώς και το 95% δ.ε. της για $i = 1, 2, \dots, 12$ και κατασκευάζει και τα γραφήματά τους.

```

a1=vcov(model1)
mestpoint=rep(1:12)
varr=rep(1:12)
lowerb=rep(1:12)
upperb=rep(1:12)
for(i in 1:12)
{
mestpoint[i]=coef(model1)[1]+coef(model1)[2]*sin(2*pi*i/12)+coef(model1)[3]
*cos(2*pi*i/12)
var[i]=a1[1,1]+((sin(2*pi*i/12))^2)*a1[2,2]+((cos(2*pi*i/12))^2)*a1[3,3]+2*
sin(2*pi*i/12)*a1[2,1]+2*cos(2*pi*i/12)*a1[1,3]+2*sin(2*pi*i/12)*cos(2*pi*i
/12)*a1[2,3]
lowerb[i]=mestpoint[i]-1.96*sqrt(varr[i])
upperb[i]=mestpoint[i]+1.96*sqrt(var[i])
}
plot(mestpoint,type="l",ylim=c(0,70),xlab="Μήνες",ylab="Παράμετρος
θέσης(mm)")
lines(lowerb,type="c")
lines(upperb,type="c")

```

Κώδικας 3 □ Το επόμενο script υπολογίζει την παράμετρο κλίμακος καθώς και το 95% δ.ε. της για $i = 1, 2, \dots, 12$ και κατασκευάζει και τα γραφήματά τους.

```

a1=vcov(model2)
c1=c(a1[4,4],a1[4,5],a1[4,6])
c2=c(a1[5,4],a1[5,5],a1[5,6])
c3=c(a1[6,4],a1[6,5],a1[6,6])
q=c(c1,c2,c3)
d=matrix(q,nrow=3,byrow=T)
grs0=rep(1:12)
grs1=rep(1:12)
grs2=rep(1:12)
sestpoint=rep(1:12)
lowerb=rep(1:12)
upperb=rep(1:12)
var=rep(1:12)

```



```

for(i in 1:12)
{
sestpoint[i]=exp(coef(model2)[4]+coef(model2)[5]*sin(2*pi*i/12)+coef(model2)
)[6]*cos(2*pi*i/12))
grs0[i]=exp(coef(modelt2p)[4]+coef(modelt2p)[5]*sin(2*pi*i/12)+coef(modelt2
p)[6]*cos(2*pi*i/12))
grs1[i]=exp(coef(modelt2p)[4]+coef(modelt2p)[5]*sin(2*pi*i/12)+coef(modelt2
p)[6]*cos(2*pi*i/12))*sin(2*pi*i/12)
grs2[i]=exp(coef(modelt2p)[4]+coef(modelt2p)[5]*sin(2*pi*i/12)+coef(modelt2
p)[6]*cos(2*pi*i/12))*cos(2*pi*i/12)
grd=rep(1:3)
grd=c(grs0[i],grs1[i],grs2[i])
var[i]=t(grd)%*%d%*%grd
lowerb[i]=sestpoint[i]-1.96*sqrt(var[i])
upperb[i]=sestpointt[i]+1.96*sqrt(var[i])
}
plot(sestpoint,type="l",ylim=c(5,25),ylab="Παράμετρος
κλίμακος (mm)",xlab="Μήνες")
lines(lowerb,type="c")
lines(upperb,type="c")

```

Κώδικας 4 □ Εκτίμηση της μηνιαίας στάθμης απόδοσης για 10 χρόνια, και υπολογισμός 95% δ.ε. και κατασκή γραφημάτων.

```

a1=vcov(model1)
m=rep(1:12)
s=rep(1:12)
x=rep(1:12)
zp=rep(1:12)
dm0=rep(1:12)
dm1=rep(1:12)
dm2=rep(1:12)
ds=rep(1:12)
dx=rep(1:12)
var=rep(1:12)
upperb=rep(1:12)
lowerb=rep(1:12)
for(i in 1:12)
{

```

```

m[i]=coef(model1)[1]+coef(model1)[2]*sin(2*pi*i/12)+coef(model2)[3]*cos(2*pi*i/12)
s[i]=coef(model1)[4]
x[i]=coef(model1)[5]
p=1/10;
w=-log(-log(1-p));
zp[i]=m[i]-s[i]/x[i]*(1-exp(x[i]*w))
dm0[i]=1
dm1[i]=sin(2*pi*i/12)
dm2[i]=cos(2*pi*i/12)
ds[i]=-(1/x[i])*(1-exp(x[i]*w))
dx[i]=s[i]/(x[i]^2)*(1-exp(x[i]*w))+(s[i]/x[i])*w*exp(x[i]*w)
grandz1=rep(1:5)
grandz1=c(dm0[i],dm1[i],dm2[i],ds[i],dx[i])

var[i]=t(grandz1)*%*%a1*%*%grandz1
lowerbb[i]=zp[i]-1.96*sqrt(var[i])
upperb[i]=zp[i]+1.96*sqrt(var[i])
}
plot(zp,type="l",ylim=c(0,160),xlab="Μήνες",ylab="Στάθμη Απόδοσης 10
ετών (mm)")
lines(lowerb,type="c")
lines(upperb,type="c")

```

Κώδικας 5 □ Με το ακόλουθο script κατασκευάζουμε το Q-Q Plot και μετατρέπουμε τις μη-στασιμες παρατηρήσεις σε στάσιμες παρατηρήσεις που ακολουθούν μια τυποποιημένη κατανομή *Gumbel*.

```

m1=coef(model1)[1]
m2=coef(model1)[2]
m3=coef(model1)[3]
s=coef(model1)[4]
x=coef(model1)[5]
zt=rep(1:length(blockmonthly))
for(i in 1:length(blockmonthly))
{
zt[i]=(1/x1)*log(1+x1*((blockmonthly[i]-m1-m2*sin(2*pi*i/12)-
m3*cos(2*pi*i/12))/s1))
}

```

```
k=length(zt)
plot(sort(zstand1),-log(-log(1:k/(k+1))),xlab="Model",ylab="Empirical")
abline(0,1)
```

Βιβλιογραφία

Ελληνόγλωσση

Κατσάπας Λ. (2010), *Θεωρία ακραίων τιμών σε μοντέλα εξαρτημένων τυχαίων μεταβλητών*, Διπλωματική εργασία για το ΠΜΣ «Αναλογιστική Επιστήμη και Διοικητική Κινδύνου» Πανεπιστήμιο Πειραιώς.

Μπούτσικας Μ. (2004), *Μέθοδοι προσομοίωσης και υπολογιστικές στατιστικές τεχνικές*, Πανεπιστημικές σημειώσεις ΠΜΣ «Εφαρμοσμένη Στατιστική» Πανεπιστήμιο Πειραιώς.

Μπούτσικας Μ. (2008), *Θεωρία ακραίων τιμών*, Πανεπιστημικές σημειώσεις ΠΜΣ «Αναλογιστική Επιστήμη και Διοικητική Κινδύνου» Πανεπιστήμιο Πειραιώς.

Ξενόγλωσση

Beirlant J., Goegebeur Y., Teugels J., Segers J., De Waal D., Ferro C. (2004), *Statistics of Extremes* □ *Theory and Applications*, John Wiley & Sons, Ltd.

Buishand T.A. (1989), Statistics of extremes in climatology, *Statistica Neerlandica*, 43.

Carter D.J.T. and Challenger P. G. (1981), Estimating return values of environmental parameters, *Quartely Journal of Royal Meteorological Society*, 107, 259-266.

Coles S. (2001), *An introduction to statistical modelling of extreme values*, Springer Verlag.

- De Haan L., Ferreira, A. (2006), *Extreme Value Theory: An Introduction*, Springer, Boston.
- Embrechts P., Kluppelberg C., Mikosch T. (1997), *Modelling Extremal Events for Insurance and Finance*, Springer Verlag.
- Engeland K, Hisdal H., Firgessi A. (2004), Practical Extreme Value Modelling of Hydrological Floods and Droughts: A Case Study, *Extremes*, 7, 5–30.
- Fisher R.A. and Tippett LHC (1928), Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Mathematical Proceedings of the Cambridge Philosophical Society* 24, issue 02, 180-190.
- Finkenstädt B. and Rootzén H. (2003), *Extreme Value in Finance Telecommunications, and the Environment*, Chapman and Hall, CRC.
- Gumbel E.J. (1958), *Statistics of extremes*, Columbia University Press, New York.
- Katz R.W., Parlange M.B., Naveau P. (2002), Statistics of extremes in hydrology, *Advances in Water Resources*, 25, 1287-1304.
- Kioutsioukis I., Melas D. and Zerefos C. (2010), Statistical assessment of changes in climate extremes over Greece(1955-2004), *International Journal of Climatology*, 30, 1723-1737.
- Kotz S. and Nadarajah S. (2001), *Extreme Value Distributions: Theory and applications*, Imperial College Press London.
- Koutsouyiannis D. Baloutsos G. (2000), Analysis of a Long Record of Annual Maximum Rainfall in Athens, Greece, and Design Rainfall Inferences, *Natural Hazards*, 29, 29–48.
- Leadbetter M.R. (1983), Extremes and local dependence in stationary sequences, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 65, 291-306.
- Reiss R.D. and Thomas M, (2007), *Statistical Analysis of Extreme Values(third edition)*, Birkhäuser Verlag AG.
- Rust H.W., Maraun D. and Osborn T.J. (2009), Modeling seasonality in extreme precipitation: A UK case study, *The European Physical Journal special topics*, 174, 99-111.
- Smith R.L. (1989), Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-Level Ozone, *Statistical Science*, 4, 367-377.

РАНЕЕЗНАМО ТЕРРА