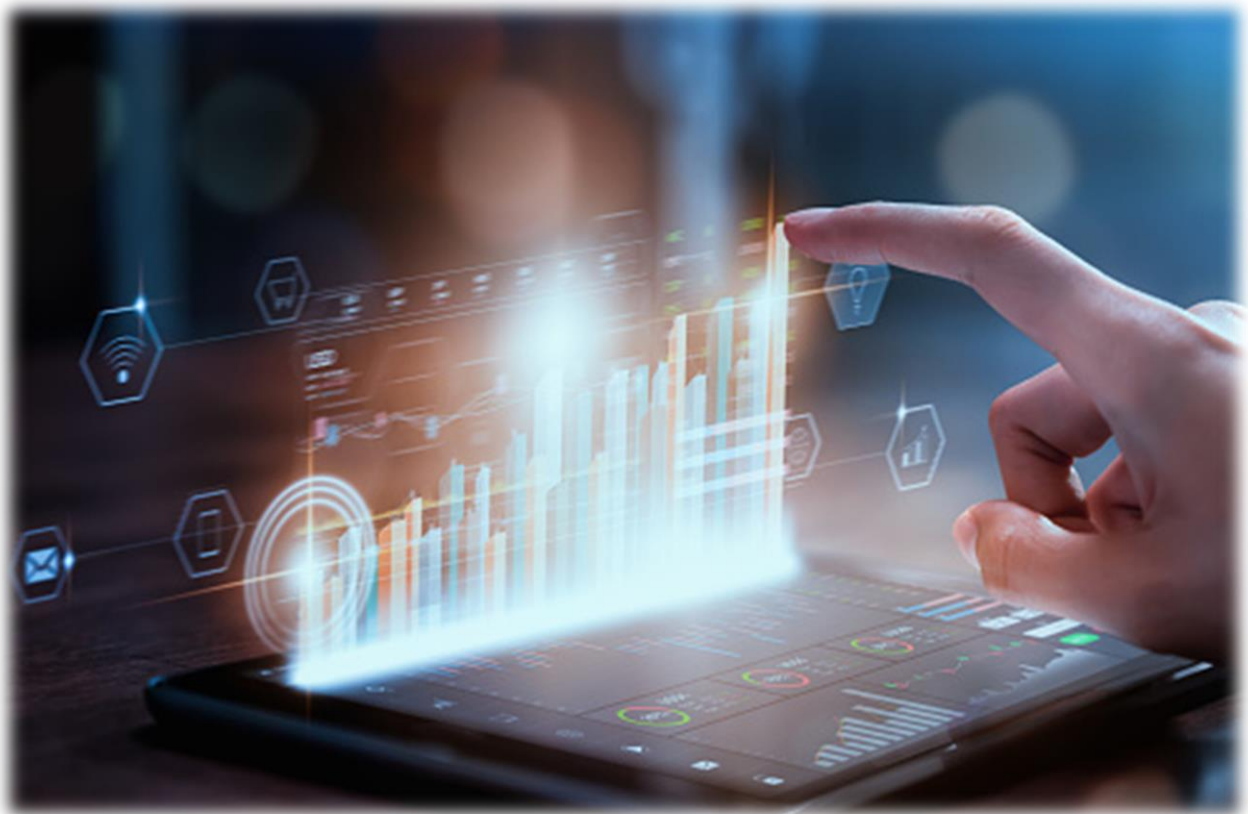




ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΚΑΙ ΥΔΡΟΛΟΓΙΚΗ ΕΡΜΗΝΕΙΑ ΤΗΣ ΔΙΑΙΤΑΣ ΤΩΝ ΗΜΕΡΗΣΙΩΝ ΠΑΡΟΧΩΝ ΣΕ ΜΕΓΑΛΟ ΔΕΙΓΜΑ ΠΟΤΑΜΩΝ ΤΗΣ ΕΥΡΩΠΗΣ



ΠΡΟΜΠΟΝΑΣ ΕΜΜΑΝΟΥΗΛ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Επιβλέπων: Ανδρέας Ευστρατιάδης, Επίκουρος Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2021



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

ΤΟΜΕΑΣ ΥΔΑΤΙΚΩΝ ΠΟΡΩΝ ΚΑΙ ΠΕΡΙΒΑΛΛΟΝΤΟΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΚΑΙ ΥΔΡΟΛΟΓΙΚΗ ΕΡΜΗΝΕΙΑ
ΤΗΣ ΔΙΑΙΤΑΣ ΤΩΝ ΗΜΕΡΗΣΙΩΝ ΠΑΡΟΧΩΝ ΣΕ
ΜΕΓΑΛΟ ΔΕΙΓΜΑ ΠΟΤΑΜΩΝ ΤΗΣ ΕΥΡΩΠΗΣ

DIPLOMA THESIS

STATISTICAL ANALYSIS AND HYDROLOGICAL
INTERPRETATION OF THE DAILY FLOW RATE DIET IN
A LARGE SAMPLE OF EUROPE RIVERS

ΠΡΟΜΠΟΝΑΣ ΕΜΜΑΝΟΥΗΛ

Επιβλέπων: Ανδρέας Ευστρατιάδης, Επίκουρος Καθηγητής Ε.Μ.Π.

Copyright © Μανόλης Προμπονάς, 2021

Με επιφύλαξη παντός δικαιώματος

Απαγορεύεται η αντιγραφή, αποθήκευση σε αρχείο πληροφοριών, διανομή, αναπαραγωγή, μετάφραση ή μετάδοση της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό, υπό οποιαδήποτε μορφή και με οποιοδήποτε μέσο επικοινωνίας, ηλεκτρονικό ή μηχανικό, χωρίς την προηγούμενη έγγραφη άδεια του Μανόλη Προμπονά. Επιτρέπεται η αναπαραγωγή, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν στη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Η έγκριση της διπλωματικής εργασίας από τη Σχολή Πολιτικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου δεν υποδηλώνει αποδοχή των απόψεων του (Ν. 5343/1932, Άρθρο 202).

Copyright © Manolis Promponas, 2021

All Rights Reserved

Neither the whole nor any part of this diploma thesis may be copied, stored in a retrieval system, distributed, reproduced, translated, or transmitted for commercial purposes, in any form or by any means now or hereafter known, electronic or mechanical, without the written permission from the author. Reproducing, storing and distributing this thesis for non-profitable, educational or research purposes is allowed, without prejudice to reference to its source and to inclusion of the present text. Any queries in relation to the use of the present thesis for commercial purposes must be addressed to its author.

Approval of this diploma thesis by the School of Civil Engineering of the National Technical University of Athens (NTUA) does not constitute in any way an acceptance of the views of the author contained herein by the said academic organization (L. 5343/1932, art. 202)

Φωτογραφία εξωφύλλου: Στατιστική Ανάλυση Δεδομένων (Πηγή:
<https://www.istockphoto.com>)

Δεν ελπίζω τίποτα, δεν φοβάμαι τίποτα

είμαι λεύτερος

N. Καζαντζάκης (1883 – 1957)

*Δεν θέλω τα μεγάλα πράγματα απ' τη ζωή
αλλά εκείνα τα μικρά που κάνουν τη ζωή μεγάλη*

Ernesto Che Guevara (1928 – 1967)

ΕΥΧΑΡΙΣΤΙΕΣ

Η ολοκλήρωση της συγγραφής της παρούσας διπλωματικής εργασίας σηματοδοτεί το τέλος της πενταετούς μου φοίτησης στη σχολή Πολιτικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου. Ένα ταξίδι που μου προσέφερε πολλές χαρές και γνώσεις γύρω από το αντικείμενο του πολιτικού μηχανικού, το οποίο γνώρισα και αγάπησα βαθιά. Το συγκεκριμένο κεφάλαιο της ζωής μου δεν θα μπορούσε να κλείσει χωρίς να ευχαριστήσω κάποιους από τους ανθρώπους που έκαναν αυτήν την εμπειρία τόσο σημαντική.

Αρχικά οφείλω ένα μεγάλο ευχαριστώ στον επιβλέποντα της εργασίας κ. Ανδρέα Ευστρατιάδη, Επικ. Καθηγητή Ε.Μ.Π. για την ιδέα και ανάθεση αυτή της διπλωματικής. Τον ευχαριστώ από καρδιάς που με εμπιστεύθηκε και με τις γνώσεις του, τις αρχές του και την ευφυΐα του, μου έδινε έμπνευση να βάζω τα δυνατά μου και να εξελίσσομαι. Ένας καθηγητής που συγκεντρώνει τις αρετές τόσο του διακεκριμένου επιστήμονα - ερευνητή, που με ζήλο πρωτοπορεί στην έρευνα, όσο και του δασκάλου που επιθυμεί να είναι πάντα δίπλα στο φοιτητή, να του μεταδώσει γνώση, αξίες αλλά και την επιθυμία για περαιτέρω έρευνα και εμβάθυνση στο εκάστοτε αντικείμενο. Το ακαδημαϊκό και προσωπικό του ήθος, η στάση του προς τη ζωή και η αγάπη του για τους φοιτητές τον καθιστούν πρότυπο καθηγητή και πηγή έμπνευσης.

Στα πλαίσια της εργασίας σημαντικό ρόλο, είχαν ο Γιάννης Τσουκαλάς (Πολιτικός Μηχανικός ΕΜΠ, MSc, Δρ. Μηχανικός), ο Παναγιώτης Κοσσιέρης (Πολιτικός Μηχανικός ΕΜΠ, MSc, Δρ. Μηχανικός), η Άνυ Ηλιοπούλου (Πολιτικός Μηχανικός ΕΜΠ, MSc, Δρ. Μηχανικός) και η Τζωρτζίνα Σακκή (Πολιτικός Μηχανικός ΕΜΠ, Υποψ. Δρ. Μηχανικός) που ήταν πάντα εκεί για τις απορίες μου και πάντα πρόθυμοι να με βοηθήσουν.

Επίσης, θα ήθελα να ευχαριστήσω όλα τα μέλη της ΙΤΙΑΣ, προερχόντων των κ. Κουτσογιάννη, κ. Μαμάση, κ. Παπανικολάου για την προσφορά τους στην ακαδημαϊκή μου πορεία μέσω των προπτυχιακών τους μαθημάτων και για τις καθημερινές τηλεδιασκέψεις κατά την διάρκεια του κορονοϊού στο 8^ο και 9^ο εξάμηνο των σπουδών μου, μέσω των οποίων μου μετέδωσαν γνώσεις και συνέβαλαν ουσιαστικά αλλά και ηθικά στην εκπόνηση της παρούσας διπλωματικής εργασίας.

Σε προσωπικό επίπεδο, θα ήθελα να ευχαριστήσω τους φίλους μου που ήταν δίπλα μου σε όλα τα χρόνια της φοιτητικής μου πορείας, στα δύσκολα και εύκολα και που χωρίς εκείνους όλο αυτό το ταξίδι θα ήταν φτωχότερο, Κωνσταντίνο Κούντρα, Χρήστο Μιχαλόπουλο, Αναστάση Μαυρογιάννη και Ιάσωνα Νίκα Νασσιούλη. Ακόμα θα ήθελα

να ευχαριστήσω, τον καλό μου φίλο Δημήτρη Μουρλά, για τη πολύτιμη βοήθεια που μου παρείχε σε όλα τα έτη των σπουδών με τις γνώσεις του και για την αμέριστη συμπαράσταση και ανοχή που επέδειξε στις ιδιοτροπίες μου. Επιπλέον θα ήταν παράλειψη να μην ευχαριστήσω τον φίλο μου Βαλάντη Ξουρή που ήταν πάντα διαθέσιμος και πρόθυμος να επιλύσει οποιαδήποτε απορία μου σε πρακτικά ζητήματα στα 2 τελευταία έτη των σπουδών συμβάλλοντας έτσι καθοριστικά στην ολοκλήρωση της προπτυχιακής μου φοίτησης.

Τέλος, το μεγαλύτερο ευχαριστώ ανήκει στην οικογένεια μου, στον πατέρα μου Βασίλη, την μητέρα μου Μαριέττα και την αδερφή μου Σοφία για την αγάπη, την αδιάκοπη ηθική στήριξη και την πίστη στις δυνατότητες μου. Η στάση ζωής και φιλοσοφίας τους αποτέλεσε κίνητρο και πρότυπο για εμένα, που μου έδωσε τη απαραίτητη δύναμη που χρειαζόμουν για να συνεχίσω. Τους ευχαριστώ για όλα.

Μανόλης Προμπονάς
Αθήνα, Νοέμβριος 2021

ΠΕΡΙΛΗΨΗ

Ο κλάδος της υδρολογίας στηρίζεται στην μελέτη-παρατήρηση των φυσικών-υδρολογικών διεργασιών μέσω της επεξεργασίας υπάρχουσών μετρήσεων, ώστε να περιγράψει ικανοποιητικά το χαοτικό χαρακτήρα των υδρολογικών συστημάτων. Οι μετρήσεις λοιπόν έχουν μια άμεση και προφανή χρησιμότητα αφού η μελέτη των φυσικών διεργασιών-φαινομένων, η κατανόηση τους και η γνώση της χρονικής τους εξέλιξης βασίζεται στη δημιουργία χρονοσειρών με μετρήσεις και στην επεξεργασία αυτών των χρονοσειρών. Τα τελευταία χρόνια ένα από τα συνηθέστερα προβλήματα το οποίο καλείται να αντιμετωπίσει όποιος ασχολείται με την επιστήμη της υδρολογίας και συγκεκριμένα με τις ημερήσιες παροχές ποταμών είναι η έλλειψη (συστηματική ή μη) χρονοσειρών μετρήσεων. Κρίνεται λοιπόν απαραίτητη η στατιστική ανάλυση των υπάρχοντων δεδομένων ώστε να αποκτήσουμε μια γενική εικόνα για την στατιστική συμπεριφορά των ημερήσιων παροχών και ενδεχομένως να ανιχνεύσουμε κάποια ομαδοποίηση συμπεριφορών ανάλογα με τα κλιματικά χαρακτηριστικά κάθε περιοχής.

Για τη συγκεκριμένη στατιστική ανάλυση χρησιμοποιήθηκαν δεδομένα ημερήσιων παροχών από το Παγκόσμιο Ιστορικό Κλιματολογικό Δίκτυο (Global Historical Climatology Network, GHCN-Daily της N.O.A.A.). Επιλέχθηκαν 209 σταθμοί στην εδαφική επικράτεια της Ευρώπης, βάσει κριτηρίων ποιότητας και πληρότητας των καταγραφών τους. Η ανάλυση των δεδομένων πραγματοποιήθηκε και ελέγχθηκε στο μεγαλύτερο μέρος της, μέσω του πηγαίου κώδικα που εφαρμόστηκε στη γλώσσα προγραμματισμού R και συγκεκριμένα στο ελεύθερο προγραμματιστικό περιβάλλον της R-Studio. Υπολογίστηκαν τα στατιστικά χαρακτηριστικά των χρονοσειρών και έγινε η εκτίμηση των παραμέτρων των κατανομών μέσω της μεθόδου των L-ροπών. Στα δεδομένα προσαρμόστηκαν πέντε συναρτήσεις κατανομής πιθανότητας (PBF, Generalized Gamma, Lognormal, Weibull, Gamma), με κριτήριο την καταλληλότητα τους βάσει της διαθέσιμης βιβλιογραφίας και από την άλλη την δυνατότητα σύγκρισης δι-παραμετρικών και τρι-παραμετρικών κατανομών. Η καταλληλότητα της προσαρμογής τους ελέγχθηκε με κριτήριο το Μέσο Τετραγωνικό Σφάλμα (MSE) αλλά και από τα διαγράμματα λόγου ροπής στην κλίμακα L. Τέλος πραγματοποιήθηκε σύνθεση εποπτικών χαρτών χωρικής κατανομής στο ελεύθερο λογισμικό QGIS, που συνέβαλαν στην εξαγωγή συμπερασμάτων και ερμηνείας των αποτελεσμάτων που προέκυψαν από την ανάλυση των δεδομένων.

ABSTRACT

The field of hydrology relies on the study-observation of natural-hydrological processes through the elaboration of existing measurements, in order to satisfactorily describe the chaotic nature of hydrological systems. Measurements therefore have an immediate and obvious utility since the study of natural processes-phenomena, their understanding and knowledge of their temporal evolution is based on the creation of time series with measurements and the processing of these time series. In recent years, one of the most common problems that anyone involved in the science of hydrology and in particular with the daily flow of rivers has to face is the lack (systematic or not) of time series measurements. It is therefore necessary to statistically analyze the existing data in order to obtain an overview of the statistical behavior of daily benefits and possibly to detect some grouping of behaviors according to the climatic characteristics of each region.

For this statistical analysis, were used daily river flow data from the Global Historical Climatology Network (GHCN-Daily of N.O.A.A.). 209 stations were selected in the territory of Europe, based on quality criteria and completeness of their recordings. The data analysis was performed and controlled for the most part, through the source code applied in the R programming language and specifically in the free programming environment of R-Studio. The statistical characteristics of the time series were calculated and the distribution parameters were estimated using the L-moments method. Five probability distribution functions (PBF, Generalized Gamma, Lognormal, Weibull, Gamma) were adapted to the data, based on their suitability based on the available literature and on the other hand the ability to compare bi-parametric and tri-parametric distributions. The appropriateness of their adjustment was checked by the criterion of Mean Square Error (MSE) but also by the ratio diagrams on the scale L. Finally, supervisory spatial distribution maps were synthesized in the free QGIS software, which helped to draw conclusions and interpret the results data analysis.

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΥΧΑΡΙΣΤΙΕΣ	ix
ΠΕΡΙΛΗΨΗ	xi
ABSTRACT	xii
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ	xvi
ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ	xvii
ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ	xx
1 ΕΙΣΑΓΩΓΗ	1
1.1 Γενική τοποθέτηση.....	1
1.2 Αντικείμενο της εργασίας.....	2
1.3 Διάρθρωση της εργασίας	3
2 ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ	5
2.1 Βασικά στοιχεία στατιστικής ανάλυσης	5
2.1.1 Αναμενόμενες τιμές και παράμετροι κατανομών	5
2.1.2 Από κοινού και περιθώριες συναρτήσεις κατανομής.....	7
2.2 Πιθανοτικές κατανομές για την περιγραφή ημερήσιων παροχών	9
2.2.1 Η κατανομή Pareto-Burr-Feller (PBF)	9
2.2.2 Η κατανομή Generalized Gamma.....	12
2.2.3 Η κατανομή Weibull	14
2.2.4 Η κατανομή Lognormal.....	15
2.2.5 Η κατανομή Gamma	16
2.3 Μέθοδος των L Ροπών	18
2.4 Κριτήρια Επιλογής Εκτιμητών.....	22
2.4.1 Καταλληλότητα στατιστικής συνάρτησης	22
2.4.2 Μέσο τετραγωνικό Σφάλμα (MSE).....	23
3 ΥΠΟΛΟΓΙΣΤΙΚΑ ΕΡΓΑΛΕΙΑ	25
3.1 Εργαλείο Ανάλυσης Δεδομένων – Στατιστικό Πακέτο R	25
3.1.1 Τι είναι η R.....	25
3.1.2 Πλεονεκτήματα της γλώσσας R	26
3.1.3 R-Studio.....	27
3.2 Το ελεύθερο λογισμικό QGIS.....	29

3.2.1	Τι είναι το QGIS	29
3.2.2	Εγκατάσταση και εκκίνηση του προγράμματος	31
3.2.3	Το γραφικό περιβάλλον διεπαφής του QGIS.....	31
3.2.4	Σημαντικά χαρακτηριστικά γνωρίσματα	32
3.2.5	Πλεονεκτήματα του QGIS	35
4	ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ	37
4.1	Συλλογή Δεδομένων	37
4.2	Πηγή Δεδομένων GHCN-Daily	38
4.2.1	Γενικές Πληροφορίες.....	38
4.2.2	Δεδομένα	40
5	ΑΝΑΠΤΥΞΗ ΜΕΘΟΔΟΛΟΓΙΑΣ ΣΤΗΝ R	43
5.1	Εισαγωγή δεδομένων στην R	43
5.2	Υπολογισμός στατιστικών χαρακτηριστικών χρονοσειρών	48
5.3	Προσαρμογή κατανομών lognormal, gamma, weibull	58
5.4	Προσαρμογή κατανομών PBF, Generalized Gamma	66
5.5	Εκτίμηση παραμέτρων κατανομών PBF και Generalized Gamma	71
6	ΔΙΑΓΡΑΜΜΑΤΑ ΛΟΓΟΥ ΡΟΠΗΣ.....	79
6.1	Γενικά Στοιχεία.....	79
6.2	Διαγράμματα L-skewness LCV.....	80
7	ΧΩΡΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΣΤΑΤΙΣΤΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ	86
7.1	Διαγράμματα – Χάρτες χωρικής κατανομής	86
8	ΣΥΜΠΕΡΑΣΜΑΤΑ	91
8.1	Σύνοψη μελέτης και συμπεράσματα.....	91
8.2	Προτάσεις για μελλοντική έρευνα.....	93
9	ΒΙΒΛΙΟΓΡΑΦΙΑ.....	94
9.1	Ξενόγλωσση Βιβλιογραφία	94
9.2	Ελληνόγλωσση Βιβλιογραφία	95
9.3	Ιστοσελίδες.....	96
	ΠΑΡΑΡΤΗΜΑ I: Οι σταθμοί που χρησιμοποιήθηκαν.....	98
	ΠΑΡΑΡΤΗΜΑ II: Ο κώδικας στην γλώσσα R.....	103
	ΠΑΡΑΡΤΗΜΑ III.....	120

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 3-1: Χαρακτηριστικά γνωρίσματα και λειτουργίες του QGIS	32
Πίνακας 5-1: Δεδομένα της βάσης GHCN-Daily για τον ποταμό Po (Ιταλία)	43
Πίνακας 5-2: Αποτελέσματα Μέσης τιμής απορροής (mm) για τις πρώτες 20 χρονοσειρές ανά μήνα και ετήσια	55
Πίνακας 5-3: Αποτελέσματα συντελεστή μεταβλητότητας απορροής για τις πρώτες 20 χρονοσειρές ανά μήνα και ετήσια	55
Πίνακας 5-4: Αποτελέσματα Ασυμμετρίας απορροής (mm) για τις πρώτες 20 χρονοσειρές ανά μήνα και ετήσια	56
Πίνακας 5-5: Αποτελέσματα quantile 50%/Μέσης Τιμής απορροής για τις πρώτες 20 χρονοσειρές ανά μήνα και ετήσια	56
Πίνακας 5-6: Αποτελέσματα quantile 90%/Μέσης Τιμής απορροής για τις πρώτες 20 χρονοσειρές ανά μήνα και ετήσια	57
Πίνακας 5-7: Αποτελέσματα υστέρησης 1 απορροής (mm) για τις πρώτες 20 χρονοσειρές ανά μήνα και ετήσια	57
Πίνακας 5-8: Ποσοστά επιτυχίας κατανομών lognormal, gamma, weibull ανά χρονοσειρά	63
Πίνακας 5-9: Ποσοστά επιτυχίας κατανομών PBF, Generalized Gamma ανά χρονοσειρά	68
Πίνακας 6-1: Ποσοστά επιτυχίας κατανομών PBF, Generalized Gamma ανά μήνα για όλες τις χρονοσειρές	84

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Σχήμα 2.1: Μορφές της Pdf της κατανομής PBF για διαφορετικές τιμές των παραμέτρων a, c, k . (Πηγή: Wikipedia).....	10
Σχήμα 2.2: Μορφές της Cdf (αριστερά) και Pdf (δεξιά) της κατανομής PBF για διαφορετικές τιμές της παραμέτρου a και για σταθερές τιμές των άλλων 2 ($c=5$ $k=2$). (Πηγή: Wikipedia).....	11
Σχήμα 2.3: Συναρτήσεις που προκύπτουν με κατάλληλη ρύθμιση των παραμέτρων της Generalized Gamma (όπου $\lambda = 1/a$) (Πηγή: Wikipedia).....	12
Σχήμα 2.4: Μορφή της Pdf της κατανομής Generalized Gamma για διαφορετικές τιμές των παραμέτρων a, b, k (Στα σχήματα όπου $d=b$ και όπου $p=k$). (Πηγή:Wikipedia)	13
Σχήμα 2.5: Μορφή της Cdf της κατανομής Generalized Gamma για διαφορετικές τιμές των παραμέτρων a, b, k (Στα σχήματα όπου $d=b$ και όπου $p=k$). (Πηγή:Wikipedia)	13
Σχήμα 2.6: Μορφές της Cdf (αριστερά) και Pdf (δεξιά) της κατανομής Weibull για διαφορετικές τιμές των παραμέτρων a, b . (Στα σχήματα όπου $\lambda=a$ και όπου $k=b$). (Πηγή:Wikipedia)	15
Σχήμα 2.7: Μορφές της Cdf (αριστερά) και Pdf (δεξιά) της κατανομής Lognormal για διαφορετικές τιμές των παραμέτρων μ, σ . (Πηγή:Wikipedia)	16
Σχήμα 2.8: Μορφές της Cdf (αριστερά) και Pdf (δεξιά) της κατανομής Gamma για διαφορετικές τιμές των παραμέτρων k, θ . (Πηγή:Wikipedia)	17
Σχήμα 2.9: Παράδειγμα προσαρμογής κατανομής για διάφορες τιμές του στατιστικού μέτρου της ασυμμετρίας με (Γ) αρνητική ασυμμετρία, (A), (B) μηδενική ασυμμετρία, (Δ) θετική ασυμμετρία (Πηγή:Wikipedia).....	20
Σχήμα 2.10: Παράδειγμα προσαρμογής κατανομής για διάφορες τιμές του στατιστικού μέτρου της κυρτότητας (Πηγή:Wikipedia).....	21
Σχήμα 3.1: Προγραμματιστικό περιβάλλον R-Studio	28
Σχήμα 3.2: Το γραφικό περιβάλλον διεπαφής (graphical user interface) του QGIS (Πηγή: http://www.qgis.org)	31
Σχήμα 3.3: Screenshots από το μενού του QGIS plugin QuickMapServices (QMS), για τα χαρτογραφικά δεδομένα των πηγών: OSM-Open Street Map (αριστερά) και MapSurfer (Δεξιά)	35
Σχήμα 4.1: Ενημερωμένος κλιματικός χάρτης Köppen – Geiger για την περίοδο 1951–2000 (Kottek et al., 2006) που δείχνει τη θέση των 209 σταθμών μέτρησης ποταμών (Πηγή: Θ. Ηλιοπούλου, Διερεύνηση της ύπαρξης μακροπρόθεσμης εμμονής στην ετήσια βροχόπτωση από μία παγκόσμια βάση δεδομένων,	

Διπλωματική εργασία, 146 pages, Τομέας Υδατικών Πόρων και Περιβάλλοντος – Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, Μάρτιος 2013.).....	41
Σχήμα 4.2: Γεωγραφική κατανομή σταθμών (Google Earth).....	42
Σχήμα 4.3: Γεωγραφική κατανομή σταθμών (QGIS)	42
Σχήμα 5.1: Προσαρμογή κατανομής gamma για τη χρονοσειρά $i=1$ και το μήνα $j=159$	
Σχήμα 5.2: Προσαρμογή κατανομής Weibull για τη χρονοσειρά $i=1$ και το μήνα $j=159$	
Σχήμα 5.3: Προσαρμογή κατανομής Lognormal για τη χρονοσειρά $i=1$ και το μήνα $j=1$	60
Σχήμα 5.4: Αριθμός σταθμών που επικρατούν οι κατανομές lognormal, gamma, weibull	64
Σχήμα 5.5: Επικρατούσες κατανομές στους σταθμούς (QGIS) - MSE	65
Σχήμα 5.6: Προσαρμογή κατανομής PBF – Generalized Gamma για τη χρονοσειρά $i=1$ και το μήνα $j=1$	66
Σχήμα 5.7: Προσαρμογή κατανομής PBF – Generalized Gamma για τη χρονοσειρά $i=1$ και το μήνα $j=2$	67
Σχήμα 5.8: Προσαρμογή κατανομής PBF – Generalized Gamma για τη χρονοσειρά $i=1$ και το μήνα $j=3$	67
Σχήμα 5.9: Προσαρμογή κατανομής PBF – Generalized Gamma για τη χρονοσειρά $i=1$ ετήσια	68
Σχήμα 5.10: Αριθμός σταθμών που κυριαρχούν οι κατανομές PBF, Generalized Gamma	69
Σχήμα 5.11: Επικρατούσες κατανομές στους σταθμούς (QGIS) - MSE	70
Σχήμα 5.12: Ιστόγραμμα παραμέτρου shape1 κατανομής Generalized Gamma.....	71
Σχήμα 5.13: Γεωγραφική κατανομή τιμών παραμέτρου shape1 κατανομής Generalized Gamma	72
Σχήμα 5.14: Ιστόγραμμα παραμέτρου shape2 κατανομής Generalized Gamma.....	73
Σχήμα 5.15: Γεωγραφική κατανομή τιμών παραμέτρου shape2 κατανομής Generalized Gamma	73
Σχήμα 5.16: Ιστόγραμμα παραμέτρου shape1 κατανομής PBF.....	74
Σχήμα 5.17: Γεωγραφική κατανομή τιμών παραμέτρου shape1 κατανομής PBF.....	75
Σχήμα 5.18: Ιστόγραμμα παραμέτρου shape2 κατανομής PBF.....	76
Σχήμα 5.19: Γεωγραφική κατανομή τιμών παραμέτρου shape2 κατανομής PBF.....	77
Σχήμα 5.20: Τιμές της παραμέτρου shape2 της κατανομής PBF που είναι μεγαλύτερες ή ίσες με 0.5.....	78

Σχήμα 6.1: Διάγραμμα L-skewness, L-CV για το μήνα Ιανουάριο.....	81
Σχήμα 6.2: Διάγραμμα L-skewness, L-CV για το μήνα Φεβρουάριο.....	82
Σχήμα 6.3: Διάγραμμα L-skewness, L-CV για το μήνα Μάρτιο.....	83
Σχήμα 6.4: Διάγραμμα L-skewness, L-CV για όλους τους μήνες	83
Σχήμα 6.5: Ποσοστά Επιτυχίας Κατανομών PBF, GG	84
Σχήμα 7.1: Ιστόγραμμα Μέσης τιμής ημερήσιας Απορροής όλων των σταθμών για τις χρονοσειρές σε ετήσια κλίμακα	87
Σχήμα 7.2: Γεωγραφική κατανομή τιμών μέσης τιμής απορροής.....	87
Σχήμα 7.3: Ιστόγραμμα Συντελεστή Μεταβλητότητας όλων των σταθμών για τις χρονοσειρές σε ετήσια κλίμακα	88
Σχήμα 7.4: Γεωγραφική κατανομή τιμών συντελεστή μεταβλητότητας.....	89
Σχήμα 7.5: Ιστόγραμμα Ασυμμετρίας όλων των σταθμών για τις χρονοσειρές σε ετήσια κλίμακα	90
Σχήμα 7.6: Γεωγραφική κατανομή τιμών ασυμμετρίας.....	90

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 5.1: Η συνάρτηση <code>setwd</code> στην R.....	44
Εικόνα 5.2: Φόρτωση πακέτων στην R	45
Εικόνα 5.3: Δημιουργία λίστας, διανυσμάτων, πίνακα, σειράς για την εισαγωγή των δεδομένων από το <code>for loop</code>	45
Εικόνα 5.4: <code>For loop</code> για την εισαγωγή χρονοσειρών απορροής, έκτασης τους και ονομασίας σταθμών.....	46
Εικόνα 5.5: Ημερήσιες τιμές παροχής σε μονάδες απορροής (mm) για τη χρονοσειρά $i=1$	47
Εικόνα 5.6: Έκταση λεκάνης απορροής και ονομασία σταθμού χρονοσειράς για $i=1$	47
Εικόνα 5.7: Συνάρτηση διαχωρισμού των ημερήσιων δεδομένων ανά μήνα	48
Εικόνα 5.8: <code>For loop</code> για τον υπολογισμό των στατιστικών χαρακτηριστικών ανά μήνα.....	49
Εικόνα 5.9: Αποτελέσματα στατιστικών χαρακτηριστικών για τους 3 πρώτους μήνες της χρονοσειράς $i=1$	50
Εικόνα 5.10: Αποτελέσματα δεδομένων παροχής για τον 1 ^ο μήνα της χρονοσειράς $i=1$	50
Εικόνα 5.11: Χρονοσειρές και μήνες με πιθανότητα ξηρασίας μη μηδενική	51
Εικόνα 5.12: Δημιουργία <code>data frame</code> με τα στατιστικά μεγέθη ανά μήνα για κάθε χρονοσειρά	51
Εικόνα 5.13: Φόρτωση αποτελεσμάτων του κώδικα που ήδη έχει «τρέξει»	52
Εικόνα 5.14: <code>For loop</code> για τον υπολογισμό των στατιστικών χαρακτηριστικών ετήσια	52
Εικόνα 5.15: Δημιουργία ενός μεγάλου <code>data.frame</code> με όλα τα αποτελέσματα στατιστικών μεγεθών	54
Εικόνα 5.16: <code>For loop</code> για τον υπολογισμό του μέσου τετραγωνικού σφάλματος κάθε κατανομής ανά μήνα χρονοσειράς.....	58
Εικόνα 5.17: Αποτελέσματα μέσου τετραγωνικού σφάλματος για χρονοσειρά $i=1$ και μήνα $j=1$ για τις κατανομές <code>gamma</code> , <code>weibull</code> και <code>lognormal</code>	58
Εικόνα 5.18: <code>For loop</code> για την ένωση των αποτελεσμάτων των 12 μηνών κάθε χρονοσειράς.....	60
Εικόνα 5.19: Αποτελέσματα μέσου τετραγωνικού σφάλματος για την χρονοσειρά $i=1$	61

Εικόνα 5.20: Υπολογισμός καταλληλότερης κατανομής και ποσοστά επιτυχίας κάθε κατανομής..... 62

Εικόνα 5.21: Εξαγωγή των ποσοστών επιτυχίας κάθε κατανομής σε .txt αρχεία..... 62

1 ΕΙΣΑΓΩΓΗ

1.1 Γενική τοποθέτηση

Οι θετικές επιστήμες έχοντας ως κύριο αντικείμενο την μελέτη φυσικών διεργασιών, βασίζονται στην παρατήρηση των φυσικών φαινομένων. Ο κλάδος της υδρολογίας επίσης, στηρίζεται στην παρατήρηση των φυσικών-υδρολογικών διεργασιών (βροχόπτωση, εξάτμιση, διαπνοή κτλ.). Ο λόγος για τον οποίο η υδρολογία και γενικότερα οι γεωφυσικές επιστήμες στηρίχτηκαν σε μεγάλο βαθμό στην παρατήρηση των φυσικών φαινομένων και στη μελέτη και επεξεργασία των υπαρχουσών μετρήσεων είναι το γεγονός ότι οι απλοί νόμοι της κλασικής μηχανικής (π.χ. διατήρηση μάζας, ορμής και ενέργειας) αδυνατούν να περιγράψουν ικανοποιητικά το χαοτικό χαρακτήρα των υδρολογικών συστημάτων (υδρολογικές λεκάνες, ποταμοί, λίμνες, ατμόσφαιρα). Συνεπώς, η μελέτη της συμπεριφοράς ενός υδρολογικού συστήματος στο παρελθόν αποτελεί μονόδρομο για την κατανόηση της σημερινής και της μελλοντικής του συμπεριφοράς.

Ειδικότερα, όταν μελετάται ένα έργο που στοχεύει στην ανάπτυξη και αξιοποίηση των υδατικών πόρων το βασικότερο στοιχείο που καθορίζει το σχεδιασμό του έργου είναι η ποσότητα του νερού που μπορεί να αξιοποιηθεί. Επίσης, στην περίπτωση μελέτης ενός συστήματος έργων, υδρευτικών, αρδευτικών, ή υδροηλεκτρικών, πάλι θα πρέπει να γνωρίζουμε τη χρονική διακύμανση των ποσοτήτων νερού. Ο σωστός λοιπόν σχεδιασμός ενός υδραυλικού έργου που θα οδηγήσει στη βέλτιστη λύση, προϋποθέτει την άριστη γνώση των υδρολογικών φαινομένων που επικρατούν στη συγκεκριμένη περιοχή ώστε να είναι δυνατή η ακριβέστερη πρόβλεψη της εξέλιξης των φυσικών διεργασιών. Επομένως οι μετρήσεις έχουν μια άμεση και προφανή χρησιμότητα στο προγραμματισμό, το σχεδιασμό και στη λειτουργία των έργων.

Η μελέτη λοιπόν των φυσικών διεργασιών και η κατανόηση τους βασίζεται σε μετρήσεις που γίνονται, σε διάφορες χρονικές κλίμακες, σε ειδικούς σταθμούς, με τις απαραίτητες υποδομές, εξοπλισμένους με κατάλληλα όργανα για τη μέτρηση της βροχόπτωσης και άλλων μετεωρολογικών συνθηκών (μετεωρολογικοί και βροχομετρικοί σταθμοί). Για τη μελέτη και την ορθή καταγραφή της κάθε φυσικής διεργασίας χρησιμοποιούνται διάφορα όργανα. Οι νέες τεχνολογίες μετρήσεων που έχουν αναπτυχθεί (αυτόματοι ψηφιακοί αισθητήρες μετρήσεων, τηλεμετρία, μετεωρολογικά ραντάρ κ.ά.) καθώς επίσης και το έμπυχο δυναμικό των αντίστοιχων σταθμών (καταρτισμένοι παρατηρητές, εκπαιδευμένα συνεργεία) βοηθούν ακόμη περισσότερο στην ακριβέστερη παρατήρηση των φυσικών διεργασιών. Η παρατήρηση αυτή των φυσικών φαινομένων και η γνώση της χρονικής τους εξέλιξης έχει ως αποτέλεσμα τη δημιουργία χρονοσειρών με μετρήσεις. Με τον όρο

χρονοσειρά, εννοούμε ένα σύνολο από παρατηρήσεις οι οποίες αναφέρονται σε κάποιο συγκεκριμένο φυσικό μέγεθος και λαμβάνονται σε ορισμένες χρονικές στιγμές ή περιόδους οι οποίες ισαπέχουν μεταξύ τους.

Η περιγραφή των υδρολογικών διεργασιών, όπως ήδη αναφέραμε, βασίζεται στη μελέτη και επεξεργασία των υπαρχουσών χρονοσειρών. Η επεξεργασία αυτών των χρονοσειρών οδηγεί στη δημιουργία μοντέλων που αναπαράγουν τα στατιστικά χαρακτηριστικά των παρατηρήσεων με αποτέλεσμα την περιγραφή της εκάστοτε φυσικής διεργασίας. Συμπεραίνουμε πως το πρόβλημα της περιγραφής μιας φυσικής διεργασίας σχετίζεται άμεσα με την επεξεργασία και τη μελέτη των χρονοσειρών. Είναι λοιπόν προφανές πως όσο πιο αξιόπιστες είναι οι τιμές της χρονοσειράς που εξετάζουμε και όσο πιο μεγάλο είναι το μέγεθος του υδρολογικού δείγματος τόσο πιο ολοκληρωμένη εικόνα θα αποκτήσουμε για την αντίστοιχη φυσική διεργασία και τόσο πιο αξιόπιστες θα είναι και οι εκτιμήσεις μας.

Τα τελευταία χρόνια ένα από τα συνηθέστερα προβλήματα το οποίο καλείται να αντιμετωπίσει όποιος ασχολείται με την επιστήμη της υδρολογίας και συγκεκριμένα με τις ημερήσιες παροχές ποταμών είναι η έλλειψη (συστηματική ή μη) χρονοσειρών μετρήσεων. Κρίνεται λοιπόν απαραίτητη η στατιστική ανάλυση των υπάρχοντων δεδομένων ώστε να αποκτήσουμε μια γενική εικόνα για την στατιστική συμπεριφορά των ημερήσιων παροχών, την υδρολογική ερμηνεία που προκύπτει από την εξαγωγή χαρακτηριστικών όπως η μέση τιμή, η τυπική απόκλιση, η ασυμμετρία, η αυτοσυσχέτιση και ενδεχομένως να ανιχνεύσουμε κάποια ομαδοποίηση συμπεριφορών ανάλογα με τα κλιματικά χαρακτηριστικά κάθε περιοχής.

1.2 Αντικείμενο της εργασίας

Στόχος της παρούσας διπλωματικής εργασίας είναι η στατιστική ανάλυση και υδρολογική ερμηνεία της δίκαιτας των ημερήσιων παροχών σε μεγάλο δείγμα σταθμών. Οι χρονοσειρές που επεξεργάζονται και αναλύονται για την επίτευξη του παραπάνω στόχου αντιστοιχούν σε δεδομένα παρατηρήσεων ημερήσιων παροχών σε μεγάλο δείγμα ποταμών που ανήκουν στην εδαφική επικράτεια της Ευρώπης.

Η μεθοδολογία της στατιστικής ανάλυσης εν συντομία μπορεί να διαιρεθεί σε τρεις επιμέρους ενότητες. Η πρώτη αφορά την συλλογή, επεξεργασία και διόρθωση των χρονοσειρών ημερήσιων παροχών από την πηγή δεδομένων Global Historical Climatology Network (GHCN-Daily) και η εκλογή του επιθυμητού μήκους χρονοσειράς για την εξαγωγή αξιόπιστων αποτελεσμάτων.

Η δεύτερη ενότητα αφορά στην ανάλυση των στατιστικών χαρακτηριστικών, την προσαρμογή των κατανομών PBF, Generalized Gamma, lognormal, weibull, gamma

στα δεδομένα των ημερήσιων παροχών με κριτήριο το μέσο τετραγωνικό σφάλμα (MSE) αλλά και τα διαγράμματα λόγου ροπής στην κλίμακα L καθώς και την εκτίμηση των παραμέτρων των κατανομών μέσω της μεθόδου των L-ροπών.

Η τελευταία ενότητα αφορά στην καθολική εφαρμογή της διαδικασίας για το σύνολο των σταθμών της βάσης δεδομένων Global Historical Climatology Network (GHCN-Daily) και την δημιουργία εποπτικών χαρτών χωρικής κατανομής για τα διάφορα στατιστικά μεγέθη που υπολογίσθηκαν από όπου προέκυψαν τα συμπεράσματα και η υδρολογική ερμηνεία των στατιστικών μεγεθών.

1.3 Διάρθρωση της εργασίας

Όπως αναφέρθηκε παραπάνω, κεντρικός στόχος της παρούσας έρευνας είναι η λεπτομερής διερεύνηση της στατιστικής συμπεριφοράς και η υδρολογική ερμηνεία των ημερήσιων παροχών σε διάφορες περιοχές της Ευρώπης. Η στατιστική αυτή διερεύνηση διαμορφώνεται και αναλύεται στην παρούσα διπλωματική εργασία στο διάστημα των 9 κεφαλαίων και τριών Παραρτημάτων των οποίων δίνεται μια συνοπτική αιτιολογική οριοθέτηση ως εξής:

Με τον παρόν εισαγωγικό κεφάλαιο (**Κεφάλαιο 1**) επιδιώκεται μια σύντομη απεικόνιση της έλλειψης των δεδομένων των ημερήσιων παροχών γενικότερα, πράγμα που καθιστά την στατιστική ανάλυση των υπάρχοντων χροοσειρών αναγκαία και που οδηγεί στο τελικό κίνητρο αλλά και αρχική παρότρυνση σύλληψης της παρούσας διπλωματικής θέσης.

Στο **Κεφάλαιο 2**, ανατέμνεται το βασικό θεωρητικό υπόβαθρο της στατιστικής ανάλυσης που εφαρμόστηκε. Παραδίδονται παραδείγματα και ορισμοί με σκοπό την εναργή κατανόηση και ερμηνεία της μεθοδολογίας και των συμπερασμάτων που θα παρατεθούν στη συνέχεια.

Στο **Κεφάλαιο 3**, αναπτύσσονται τα υπολογιστικά εργαλεία που χρησιμοποιήθηκαν. Η ανάγκη εφαρμογής της προαναφερθείσας στατιστικής ανάλυσης σε μια πληθώρα δειγμάτων κατέστησε θεμελιώδη την ανάγκη εξοικείωσης και χρήσης ενός προγραμματιστικού περιβάλλοντος. Για την παρούσα εργασία επιλέχθηκε η επιλογή του προγραμματιστικού περιβάλλοντος της R Software, κυρίως λόγω της ευκολίας εξάσκησης της και των υπεράφθονων δυνατοτήτων της σε επίπεδο στατιστικής ανάλυσης. Ακόμη αναλύεται διεξοδικά το σύγχρονο εργαλείο ανοιχτού κώδικα QGIS μέσω του οποίου έγινε η παραγωγή εποπτικών χαρτών που συνέβαλαν στην εξαγωγή συμπερασμάτων και ερμηνείας των αποτελεσμάτων.

Στο **Κεφάλαιο 4**, γίνεται συνοπτική περιγραφή της βάσης δεδομένων που χρησιμοποιήθηκε στην εργασία, καθώς και οι κατάλληλοι έλεγχοι που πραγματοποιήθηκαν για την τελική επιλογή σταθμών, για περαιτέρω μελέτη.

Στο **Κεφάλαιο 5**, γίνεται λόγος για τα βήματα που ακολουθήθηκαν στη γλώσσα προγραμματισμού R για την στατιστική ανάλυση των δεδομένων. Συγκεκριμένα παρατίθενται μέσα από εικόνες τα σημαντικότερα στοιχεία του κώδικα καθώς και τα αποτελέσματα αυτών σε πίνακες. Κατά σειρά θα παρουσιαστούν η εισαγωγή των δεδομένων στο περιβάλλον της R, ο υπολογισμός των στατιστικών χαρακτηριστικών όλων των χρονοσειρών, η εξαγωγή των αποτελεσμάτων σε μορφή αρχείου .txt και η προσαρμογή των κατανομών lognormal, gamma, weibull, PBF και Generalized Gamma στις χρονοσειρές της βάσης δεδομένων όπως και η εκτίμηση των παραμέτρων τους από όπου προκύπτουν κάποια αρχικά συμπεράσματα για το σύνολο των δειγμάτων.

Στο **Κεφάλαιο 6**, παρουσιάζονται τα διαγράμματα λόγου ροπής, μια ευρέως χρησιμοποιούμενη προσέγγιση για την προκαταρκτική διάκριση των πιο συνεπών κατανομών δηλαδή για να περιγράψει την καταλληλότητα των κατανομών μέσω μιας γραφικής σύγκρισης μεταξύ των στατιστικών δειγμάτων, που υπολογίζονται από τα δεδομένα της βάσης και των θεωρητικών, όπως δίδονται από τους τύπους των υποψήφιων παραμετρικών κατανομών.

Στο **Κεφάλαιο 7**, παρουσιάζονται ορισμένα διαγράμματα των στατιστικών χαρακτηριστικών που προέκυψαν από την ανάλυση καθώς και οι εποπτικοί χάρτες που παρήχθησαν στο ελεύθερο λογισμικό QGIS και που συνέβαλαν στην εξαγωγή συμπερασμάτων και ερμηνείας των αποτελεσμάτων.

Στο **Κεφάλαιο 8**, παρουσιάζονται τα τελικά συμπεράσματα της στατιστικής ανάλυσης και της υδρολογικής ερμηνείας του δείγματος των σταθμών της Ευρώπης και διατυπώνονται προτάσεις για τη μετεξέλιξη της έρευνας.

Στο **Κεφάλαιο 9**, γίνεται λόγος για τις ελληνόγλωσσες και ξενόγλωσσες βιβλιογραφικές αναφορές καθώς και τις ιστοσελίδες που χρησιμοποιήθηκαν για άντληση στοιχείων.

Τέλος, η διπλωματική εργασία συνοδεύεται επίσης από 3 Παραρτήματα. Στο **Παράρτημα I** παρατίθενται όλοι οι σταθμοί που χρησιμοποιήθηκαν στην μελέτη με την ονομασία τους, την γεωγραφική τους θέση και την έκτασή τους, στο **Παράρτημα II** παρουσιάζεται ο πηγαίος κώδικας που εφαρμόστηκε γραμμένος στη γλώσσα προγραμματισμού R και στο **Παράρτημα III** τα ετήσια αποτελέσματα της προσαρμογής των κατανομών Burr και Generalized Gamma στα δεδομένα των χρονοσειρών ημερήσιων παροχών σε μορφή διαγραμμάτων για όλες τις χρονοσειρές.

2 ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

2.1 Βασικά στοιχεία στατιστικής ανάλυσης

Οι διάφορες υδρολογικές μεταβλητές παρουσιάζονται και αντιμετωπίζονται ως τυχαίες μεταβλητές. Για την ανάλυση των υδρολογικών μεταβλητών, η επιστήμη της υδρολογίας κάνει χρήση διαφόρων εργαλείων της στατιστικής και της θεωρίας πιθανοτήτων. Σκοπός μας εδώ δεν είναι να αναλύσουμε πλήρως τη θεωρία της στατιστικής, αλλά μόνο να περιγράψουμε τα εργαλεία που χρησιμοποιούνται στα επόμενα κεφάλαια και βοηθούν στην κατανόηση των διαφόρων υδρολογικών διεργασιών.

2.1.1 Αναμενόμενες τιμές και παράμετροι κατανομών

Η κατανομή μιας τυχαίας μεταβλητής X περιγράφεται πλήρως από τις συναρτήσεις πυκνότητας πιθανότητας και κατανομής. Συχνά όμως, ενδιαφερόμαστε για ορισμένες σταθερές οι οποίες περιγράφουν με απλό τρόπο τα βασικά χαρακτηριστικά κάποιας κατανομής και μας επιτρέπουν να έχουμε μια γρήγορη και συνοπτική εικόνα της πιθανοτικής συμπεριφοράς της τυχαίας μεταβλητής X . Οι σταθερές αυτές καλούνται παράμετροι της κατανομής και οι πιο σημαντικές είναι η μέση τιμή, η διασπορά, η ασυμμετρία και η κύρτωση.

Ορίζεται ως αναμενόμενη τιμή ή προσδοκία της $g(X)$ το μέγεθος:

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)f_X(x)dx \quad (2.1)$$

όπου X είναι μια συνεχής τυχαία μεταβλητή και $g(X)$ μια πραγματική συνάρτηση της X . Η παραπάνω σχέση είναι γνωστή ως νόμος του αφηρημένου στατιστικού και πρακτικά σημαίνει ότι για να βρούμε τη μέση τιμή της τυχαίας μεταβλητής $Y = g(X)$, δεν είναι ανάγκη να βρούμε τη σ.π.π. της τ.μ. Y , αρκεί να γνωρίζουμε τη σ.π.π. της τ.μ. X και τον τύπο της συνάρτησης $g(X)$.

Για $g(X) = X^r$, όπου $r = 0, 1, 2, \dots$, το μέγεθος:

$$m_X^{(r)} = E[X^r] \quad (2.2)$$

ονομάζεται ροπή περί την αρχή (ή απλώς ροπή) τάξης r της τ.μ. X .

Ειδικότερα, για $r = 1$ που σημαίνει $g(X) = X$, το μέγεθος:

$$m_X = E[X] \quad (2.3)$$

ονομάζεται αναμενόμενη τιμή ή μέση τιμή της τ.μ. X . Η μέση τιμή αποτελεί παράμετρο θέσης της τυχαίας μεταβλητής X και περιγράφει τη θέση του κέντρου βάρους του σχήματος που ορίζει η συνάρτηση πυκνότητας πιθανότητας με τον οριζόντιο άξονα. Άλλοι παράμετροι θέσης είναι η πιθανότερη τιμή ή κορυφή, συμβολικά x_p , που ορίζεται ως η τιμή της μεταβλητής x για την οποία η σ.π.π. $f_x(x)$ γίνεται μέγιστη, και η διάμεσος, $x_{0.5}$, που ορίζεται ως εκείνη η τιμή της μεταβλητής για την οποία ισχύει $P(X \leq x_{0.5}) = P(X \geq x_{0.5}) = 0.5$. Στην περίπτωση όπου η κατανομή είναι συμμετρική και μονοκόρυφη τότε η μέση τιμή, η διάμεσος και η πιθανότερη τιμή ταυτίζονται.

Για $g(X) = (X - m_X)^r$, όπου $r=0,1,2,\dots$, το μέγεθος:

$$\mu_X^{(r)} = E[(X - m_X)^r] \quad (2.4)$$

Για $r=2$, και $g(X) = (X - m_X)^2$, το μέγεθος:

$$\sigma_X^2 = \mu_X^{(2)} = E[(X - m_X)^2] \quad (2.5)$$

ονομάζεται διασπορά της τυχαίας μεταβλητής X και συμβολίζεται με $\text{Var}[X]$. Η διασπορά μιας τυχαίας μεταβλητής αποτελεί μέτρο μεταβλητότητας της κατανομής της και δείχνει το μέγεθος της συγκέντρωσης της πυκνότητας πιθανότητας γύρω από τη μέση τιμή. Έτσι μικρή τιμή διασποράς δείχνει συγκεντρωμένη κατανομή γύρω από τη μέση τιμή ενώ μεγάλη τιμή διασποράς υποδηλώνει μια πιο «απλωμένη» κατανομή.

Μέτρο μεταβλητότητας της κατανομής είναι επίσης, η τετραγωνική ρίζα της διασποράς, που ονομάζεται τυπική απόκλιση, συμβολικά σ_X . Η τυπική απόκλιση έχει ίδιες διαστάσεις με τη τυχαία μεταβλητή και δίνεται από τον τύπο:

$$\sigma_X = \sqrt{\text{Var}[X]} \quad (2.6)$$

Η αδιάστατη παράμετρος:

$$C_{vX} = \frac{\sigma_X}{m_X} \quad (2.7)$$

ονομάζεται συντελεστής μεταβλητότητας και αποτελεί μέτρο αβεβαιότητας ως προς την εκτίμηση της τυχαίας μεταβλητής X . Μεγάλη τιμή του συντελεστή μεταβλητότητας, σημαίνει μεγάλη διασπορά σε σχέση με τη μέση τιμή και συνεπάγεται μεγάλη αβεβαιότητα ως προς την εκτίμηση της m_X .

Η κεντρική ροπή τρίτης τάξης είναι το μέγεθος:

$$\mu_X^{(3)} = E[(X - m_X)^3] \quad (2.8)$$

και σχετίζεται με την ασυμμετρία της τυχαίας μεταβλητής X . Μηδενική τιμή της τρίτης κεντρικής ροπής δείχνει συμμετρική κατανομή. Αν η τρίτη κεντρική ροπή είναι μεγαλύτερη ή μικρότερη του μηδενός σημαίνει ότι η κατανομή είναι θετικά ή αρνητικά ασύμμετρα, αντίστοιχα. Οι υδρολογικές μεταβλητές, στην πλειονότητά τους, είναι θετικά ασύμμετρα, καθώς οι εξαιρετικά υψηλές τιμές εμφανίζονται με χαμηλή συχνότητα. Στην περίπτωση θετικά ασύμμετρης κατανομής, η μέση τιμή της μεταβλητής είναι μεγαλύτερη από τη διάμεσο.

Αποτελεσματικότερο μέτρο της ασυμμετρίας της κατανομής είναι ο αδιάστατος συντελεστής ασυμμετρίας που ορίζεται από τη σχέση:

$$C_{S_X} = \frac{\mu_X^{(3)}}{\sigma_X^{(3)}} \quad (2.9)$$

2.1.2 Από κοινού και περιθώριες συναρτήσεις κατανομής

Στην υδρολογία, πολύ συχνά, μας ενδιαφέρει η ταυτόχρονη μελέτη δυο τυχαίων μεταβλητών. Έστω ζεύγος τυχαίων μεταβλητών (X, Y) που, σύμφωνα με τα παραπάνω, είναι συναρτήσεις των δειγματικών χώρων (Ω_X, Ω_Y) , αντίστοιχα. Η από κοινού συνάρτηση κατανομής ζεύγους μεταβλητών (X, Y) ορίζεται ως:

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) \quad (2.10)$$

Σημειώνεται ότι οι συναρτήσεις $F_X(x) = P(X \leq x)$ και $F_Y(y) = P(Y \leq y)$ ονομάζονται περιθώριες συναρτήσεις κατανομής των X και Y , αντίστοιχα.

Υπό την προϋπόθεση ότι η F_{XY} είναι παραγωγίσιμη, η από κοινού συνάρτηση πυκνότητας πιθανότητας των μεταβλητών (X, Y) ορίζεται ως:

$$f_{XY}(x, y) = \frac{d^2 F_{XY}(x, y)}{dx dy} \quad (2.11)$$

Προφανώς ισχύει:

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^y \int_{-\infty}^x f_{XY}(u, v) du dv \quad (2.12)$$

Κατά παρόμοιο τρόπο, όπως προηγουμένως, μπορούμε να ορίσουμε τις από κοινού κεντρικές ροπές των μεταβλητών.

Η αναμενόμενη τιμή ή προσδοκία της συνάρτησης $g(X, Y)$ ορίζεται ως:

$$E[g(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) dy dx \quad (2.13)$$

Το μέγεθος $E[X^p Y^q]$ ονομάζεται από κοινού ροπή τάξης $p+q$ των μεταβλητών X και Y . Αντίστοιχα, το μέγεθος $E[(X - m_X)^p (Y - m_Y)^q]$ ονομάζεται από κοινού κεντρική ροπή τάξης $p+q$ των μεταβλητών X και Y . Από τις από κοινού κεντρικές ροπές συχνότερα χρησιμοποιείται η:

$$\rho_{XY} = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]\text{Var}[Y]}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad (2.14)$$

Ο συντελεστής συσχέτισης λαμβάνει τιμές στο διάστημα $[-1, 1]$ και αποτελεί σημαντική παράμετρο για τη μελέτη της γραμμικής συσχέτισης δύο μεταβλητών. Αν η τιμή του συντελεστή ρ_{XY} είναι κοντά στο 1, τότε οι μεταβλητές παρουσιάζουν ισχυρή θετική συσχέτιση. Αν η τιμή του συντελεστή ρ_{XY} είναι κοντά στο -1, τότε οι μεταβλητές παρουσιάζουν ισχυρή αρνητική συσχέτιση. Και τέλος, αν η τιμή του συντελεστή ρ_{XY} είναι κοντά στο μηδέν, τότε οι μεταβλητές ονομάζονται ασυσχέτιστες και δεν παρουσιάζουν καμιά γραμμική συσχέτιση.

Θα πρέπει να τονίσουμε πως ασυσχέτιστες μεταβλητές δεν σημαίνει ανεξάρτητες μεταβλητές. Μηδενική τιμή συντελεστή αυτοσυσχέτισης σημαίνει πως ο βαθμός της γραμμικής σχέσης μεταξύ των μεταβλητών X και Y είναι μηδέν (μπορεί να υπάρχει κάποια άλλη μη γραμμική σχέση που να συνδέει τις μεταβλητές). Οι μεταβλητές X και Y λέγονται ανεξάρτητες, αν για κάθε ζεύγος τιμών (x, y) ισχύει:

$$F_{XY}(x, y) = F_X(x)F_Y(y) \quad (2.15)$$

Σε αυτή την περίπτωση ισχύει επίσης η σχέση:

$$f_{XY}(x, y) = f_X(x)f_Y(y) \quad (2.16)$$

2.2 Πιθανοτικές κατανομές για την περιγραφή ημερήσιων παροχών

Στην συνέχεια σκοπός είναι η εύρεση και προσαρμογή, μίας κατάλληλης περιθώριας συνάρτησης κατανομής (marginal distribution function) ικανής να περιγράψει όλο το φάσμα των ημερήσιων παροχών των χρονοσειρών με μηνιαίο βήμα. Στην βιβλιογραφία, η πιο ευρέως χρησιμοποιούμενη κατανομή για να περιγράψει τις ημερήσιες παροχές είναι η lognormal (Λογαριθμική κατανομή). Σε διάφορες αναφορές υπήρχε προσαρμογή των ημερήσιων παροχών στις: PBF, Weibull, Gamma και Generalized Gamma. Επομένως, επειδή είναι μείζονος σημασίας η προσαρμογή στην σωστή κατανομή και η εξαγωγή των κατάλληλων παραμέτρων της, έτσι ώστε να προσομοιάζει όσο το δυνατόν καλύτερα την εμπειρική κατανομή, έγινε διερεύνηση και των πέντε προαναφερόμενων κατανομών, με μοναδικό σκοπό την προσαρμογή χρονοσειρών όσο το δυνατόν πιο κοντά στην πραγματικότητα.

2.2.1 Η κατανομή Pareto-Burr-Feller (PBF)

Η κατανομή Burr εισήχθη αρχικά από τον Burr (1942) έχοντας δύο παραμέτρους. Μια πρόσθετη παράμετρος κλίμακας εισήχθη από τον Tadikamalla (1980). Περιλαμβάνει, επικαλύπτει, ή έχει ως οριακή περίπτωση, πολλές κοινώς χρησιμοποιούμενες κατανομές όπως η Gamma, η Lognormal και η Loglogistic. Η κατανομή Burr έχει δύο ασυμπτωτικές οριακές περιπτώσεις: την Weibull και την Pareto Type I. Η ονομασία για την συγκεκριμένη κατανομή ποικίλλει ανάλογα με τον συγγραφέα. Χρησιμοποιούνται ονομασίες όπως Pareto type IV, Burr type VII ή και Burr type XII. Για την αποφυγή τέτοιων συγχύσεων, οι Koutsoyiannis et al. (2016) πρότειναν την ονομασία Pareto-Burr-Feller (PBF), προς τιμήν του Ιταλού πολιτικού μηχανικού, οικονομολόγου και κοινωνιολόγου V. Pareto, ο οποίος ανακάλυψε μία οικογένεια κατανομών δύναμης για τη διερεύνηση της κατανομής του μεγέθους των εισοδημάτων σε μια κοινωνία, του Burr ο οποίος αναγνώρισε και ανέλυσε (αλλά χωρίς να δώσει αιτιολόγηση) μια συνάρτηση που προτάθηκε αρχικά σε αλγεβρική μορφή από τον Bierens de Haan, και στον Feller ο οποίος τη συνέδεσε με τη συνάρτηση και κατανομή Beta.

Η κατανομή PBF είναι μια κατανομή τριών παραμέτρων και είναι ορισμένη στο $(0, +\infty)$. Παρατίθενται οι κλασσικοί αναλυτικοί τύποι της συνάρτησης πυκνότητας πιθανότητας (Pdf) και της συνάρτησης αθροιστικής πιθανότητας (Cdf).

Η Cumulative Distribution Function (Cdf) είναι η :

$$F(x|a, c, k) = 1 - \frac{1}{\left(1 + \left(\frac{x}{a}\right)^c\right)^k}, x > 0, a > 0, c > 0, k > 0 \quad (2.17)$$

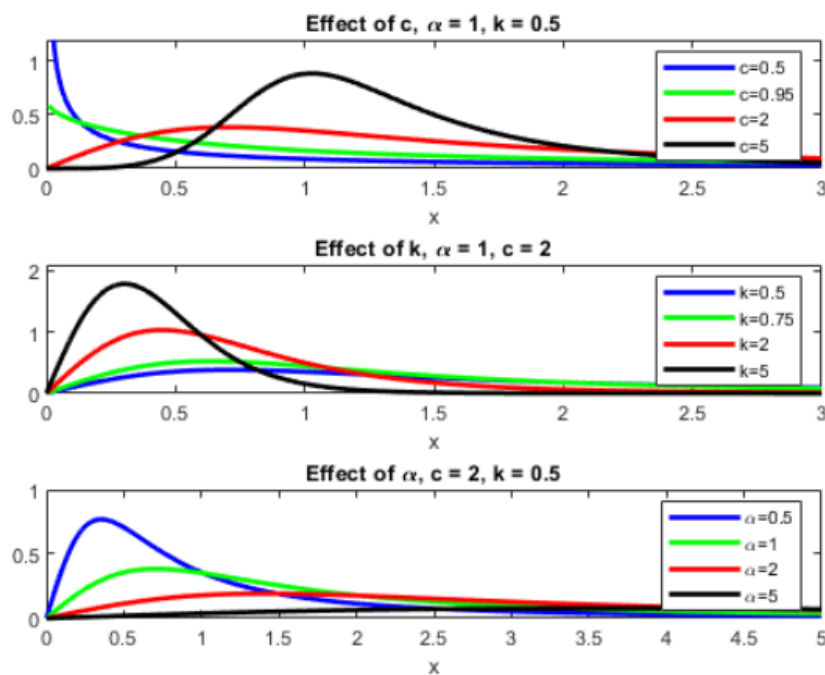
Η Probability Density Function (Pdf) είναι η :

$$f(x|a, c, k) = \frac{kc \left(\frac{x}{a}\right)^{c-1}}{\left(1 + \left(\frac{x}{a}\right)^c\right)^{k+1}}, x > 0, a > 0, c > 0, k > 0 \quad (2.18)$$

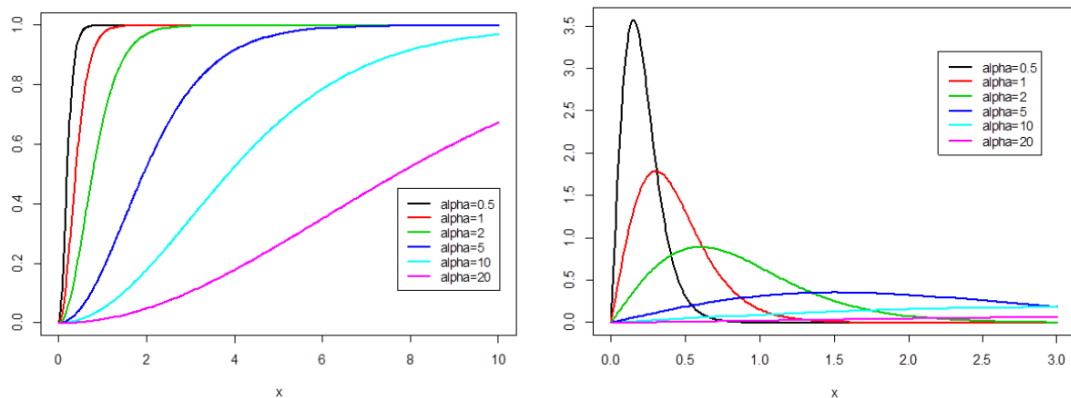
όπου:

a : είναι η παράμετρος κλίμακας

c, k : είναι οι παράμετροι σχήματος



Σχήμα 2.1: Μορφές της Pdf της κατανομής PBF για διαφορετικές τιμές των παραμέτρων a, c, k . (Πηγή: Wikipedia)



Σχήμα 2.2: Μορφές της Cdf (αριστερά) και Pdf (δεξιά) της κατανομής PBF για διαφορετικές τιμές της παραμέτρου α και για σταθερές τιμές των άλλων 2 ($c=5$ $k=2$).
(Πηγή: Wikipedia)

Αν στην κατανομή PBF $k \rightarrow \infty$, $a \rightarrow \infty$, $\frac{a}{k^{1/c}} = \theta$, τότε η τελευταία μετατρέπεται σε μία κατανομή 2 παραμέτρων με Cdf :

$$F(x|\theta, c) = 1 - e^{-(x/\theta)^c}, x > 0, \theta > 0, c > 0 \quad (2.19)$$

μετατρέπεται δηλαδή στην κατανομή Weibull.

Ωστόσο στην παρούσα εργασία χρησιμοποιήθηκαν οι παρακάτω τύποι της συνάρτησης πυκνότητας πιθανότητας (Pdf) και της συνάρτησης αθροιστικής πιθανότητας (Cdf) της κατανομής PBF.

Η Cumulative Distribution Function (Cdf) είναι η :

$$F(x|\lambda, \zeta, \xi) = 1 - \left(1 + \xi \zeta \left(\frac{x}{\lambda}\right)^\zeta\right)^{-\frac{1}{\xi}}, \quad x \geq 0 \quad (2.20)$$

Η Probability Density Function (Pdf) είναι η :

$$f(x|\lambda, \zeta, \xi) = \frac{1}{\lambda \xi} \left(\frac{x}{\lambda}\right)^{\zeta-1} \left(1 + \left(\frac{x}{\lambda}\right)^\zeta\right)^{-\frac{1}{\xi}-1}, \quad x \geq 0 \quad (2.21)$$

όπου:

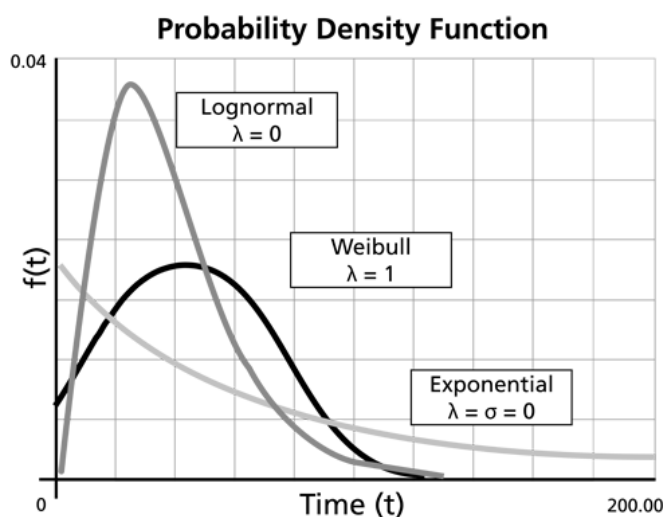
λ : η παράμετρος κλίμακας, ενώ

ζ (*shape1*) and ξ (*shape2*): είναι οι παράμετροι σχήματος

Η παράμετρος ξ είναι γνωστή και ως δείκτης ουράς(*tail index*)

2.2.2 Η κατανομή Generalized Gamma

Μια από τις «ελαφριάς ουράς» κατανομές είναι η Generalized Gamma. Είναι μια κατανομή τριών παραμέτρων. Ένα πλήθος από γνωστές συναρτήσεις προκύπτουν από αυτή με κατάλληλη ρύθμιση των παραμέτρων της (κατανομή Gamma, Weibull, Nakgami, half-normal και η εκθετική κατανομή). Συγκεκριμένα στο παρακάτω σχήμα απεικονίζονται οι συναρτήσεις που προκύπτουν από την Generalized Gamma:



Σχήμα 2.3: Συναρτήσεις που προκύπτουν με κατάλληλη ρύθμιση των παραμέτρων της Generalized Gamma (όπου $\lambda = 1/\sqrt{a}$) (Πηγή: Wikipedia)

Η κατανομή Generalized Gamma είναι μια κατανομή ορισμένη στο $(0, +\infty)$. Στη συνέχεια παρατίθενται οι αναλυτικοί τύποι της συνάρτησης αθροιστικής πιθανότητας (Cumulative Distribution Function CDF) και της συνάρτησης πυκνότητας πιθανότητας (Probability Density Function PDF), (Mert and Karakus, 2014).

Η Cumulative Distribution Function (Cdf) είναι η :

$$F(x|a, b, k) = \frac{\gamma\left(\left(\frac{x}{b}\right)^k\right)}{\Gamma(\alpha)}, x > 0, a > 0, b > 0, k > 0 \quad (2.22)$$

Η Probability Density Function (Pdf) είναι η :

$$f(x|a, b, k) = kx^{a-1} \frac{\exp\left(-\left(\frac{x}{b}\right)^k\right)}{b^ka\Gamma(\alpha)}, x > 0, \alpha > 0, b > 0, k > 0 \quad (2.23)$$

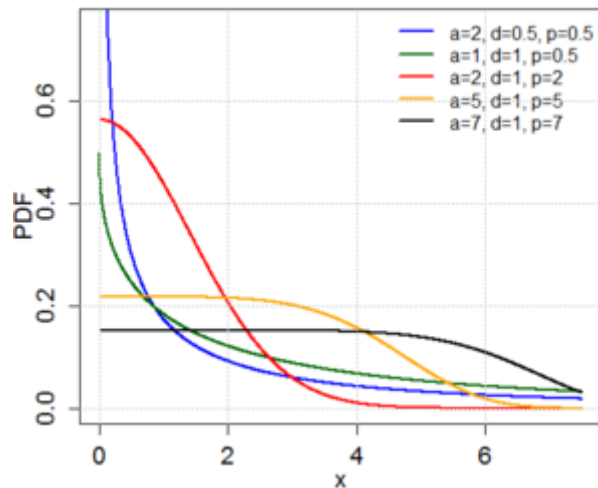
με

- $\Gamma(a)$: η κατανομή Γάμμα (Gamma) στο σημείο a
- $\gamma\left(\left(\frac{x}{b}\right)^k\right)$: η ατελής κατανομή Γάμμα (Gamma)

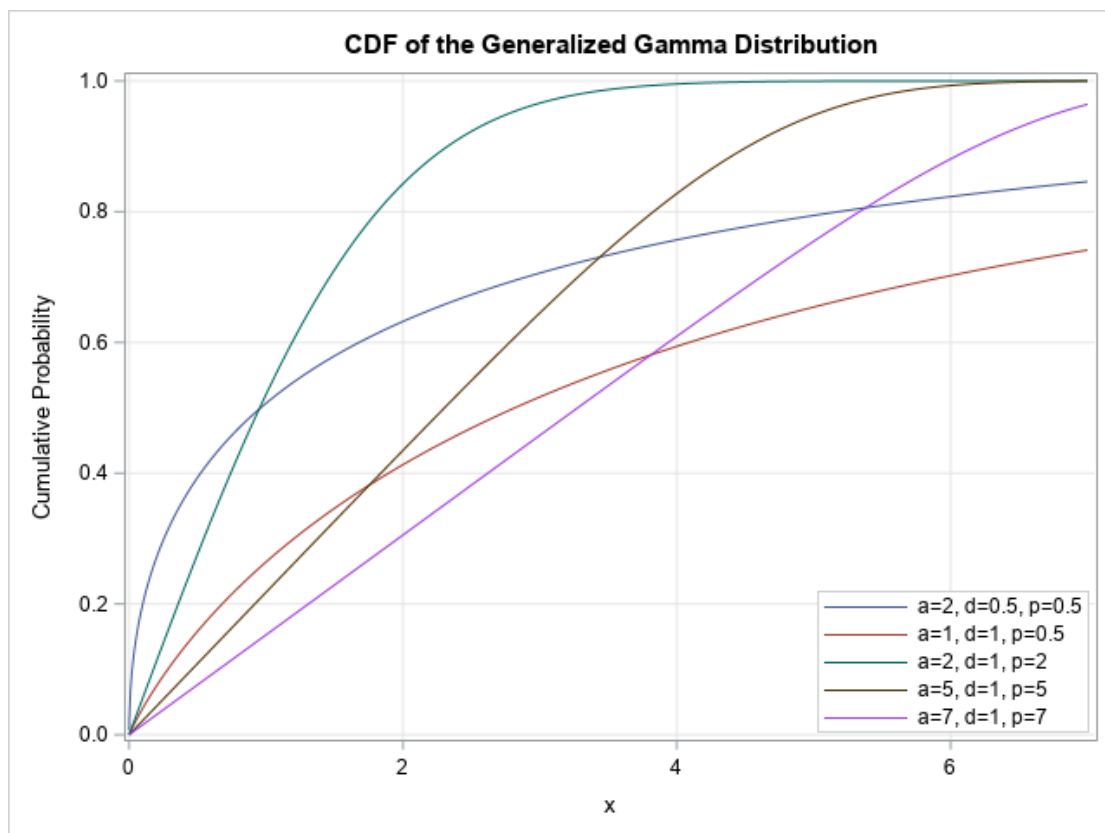
όπου:

b : είναι η παράμετρος κλίμακας

a, k : είναι οι παράμετροι σχήματος



Σχήμα 2.4: Μορφή της Pdf της κατανομής Generalized Gamma για διαφορετικές τιμές των παραμέτρων a, b, k (Στα σχήματα όπου $d=b$ και όπου $p=k$).
(Πηγή:Wikipedia)



Σχήμα 2.5: Μορφή της Cdf της κατανομής Generalized Gamma για διαφορετικές τιμές των παραμέτρων a, b, k (Στα σχήματα όπου $d=b$ και όπου $p=k$).
(Πηγή:Wikipedia)

Ωστόσο στην παρούσα εργασία χρησιμοποιήθηκαν οι παρακάτω τύποι της συνάρτησης πυκνότητας πιθανότητας (Pdf) και της συνάρτησης αθροιστικής πιθανότητας (Cdf) της κατανομής PBF.

Η Cumulative Distribution Function (Cdf) είναι η :

$$F(x|b, a_1, a_2) = 1 - \frac{\Gamma\left[\frac{a_1}{a_2}, \left(\frac{x}{b}\right)^{a_2}\right]}{\Gamma\left(\frac{a_1}{a_2}\right)}, \quad x \geq 0 \quad (2.24)$$

Η Probability Density Function (Pdf) είναι η :

$$f(x|b, a_1, a_2) = \frac{a_2}{b\Gamma(a_1/a_2)} \left(\frac{x}{b}\right)^{a_1-1} \exp\left[-\left(\frac{x}{b}\right)^{a_2}\right], \quad x \geq 0 \quad (2.25)$$

όπου:

b: η παράμετρος κλίμακας, ενώ

α_1 (shape1) and α_2 (shape2): είναι οι παράμετροι σχήματος

2.2.3 Η κατανομή Weibull

Η κατανομή Weibull είναι μία συνεχής κατανομή πιθανότητας. Πήρε το όνομά της από τον Σουηδό μαθηματικό Waloddi Weibull, ο οποίος την περιέγραψε λεπτομερώς το 1951, παρόλο που για πρώτη φορά εμφανίστηκε από τον Fréchet (1927) και εφαρμόστηκε τελικά από τους Rosin & Rammler (1933) για την περιγραφή της κατανομής μεγέθους σωματιδίων.

Η κατανομή Weibull είναι μια κατανομή δύο παραμέτρων και είναι ορισμένη στο $(0, +\infty)$. Παρατίθενται οι αναλυτικοί τύποι της συνάρτησης πυκνότητας πιθανότητας (Pdf) και της συνάρτησης αθροιστικής πιθανότητας (Cdf).

Η Cumulative Distribution Function (Cdf) είναι η :

$$F(x|a, b) = 1 - e^{-\left(\frac{x}{a}\right)^b}, \quad x > 0, a > 0, b > 0 \quad (2.26)$$

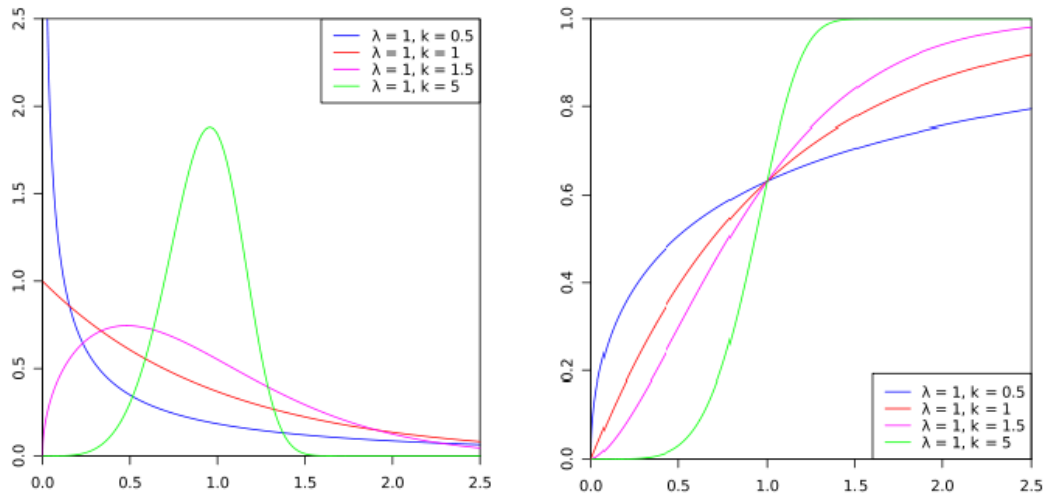
Η Probability Density Function (Pdf) είναι η :

$$f(x|a, b) = \frac{b}{a} \left(\frac{x}{a}\right)^{b-1} e^{-\left(\frac{x}{a}\right)^b}, \quad x > 0, a > 0, b > 0 \quad (2.27)$$

όπου:

a : είναι η παράμετρος κλίμακας

b : είναι η παράμετρος σχήματος



Σχήμα 2.6: Μορφές της Cdf (αριστερά) και Pdf (δεξιά) της κατανομής Weibull για διαφορετικές τιμές των παραμέτρων a, b . (Στα σχήματα όπου $\lambda=a$ και όπου $k=b$).
(Πηγή:Wikipedia)

2.2.4 Η κατανομή Lognormal

Η κατανομή Lognormal είναι μια συνεχής κατανομή πιθανότητας μιας τυχαίας μεταβλητής, της οποίας ο λογάριθμος είναι κανονικά κατανεμημένος. Μια τυχαία μεταβλητή, η οποία ακολουθεί την Lognormal κατανομή, λαμβάνει μόνο θετικές πραγματικές τιμές, αφού είναι ορισμένη στο $(0, +\infty)$. Παρατίθενται οι αναλυτικοί τύποι της συνάρτησης πυκνότητας πιθανότητας (Pdf) και της συνάρτησης αθροιστικής πιθανότητας (Cdf).

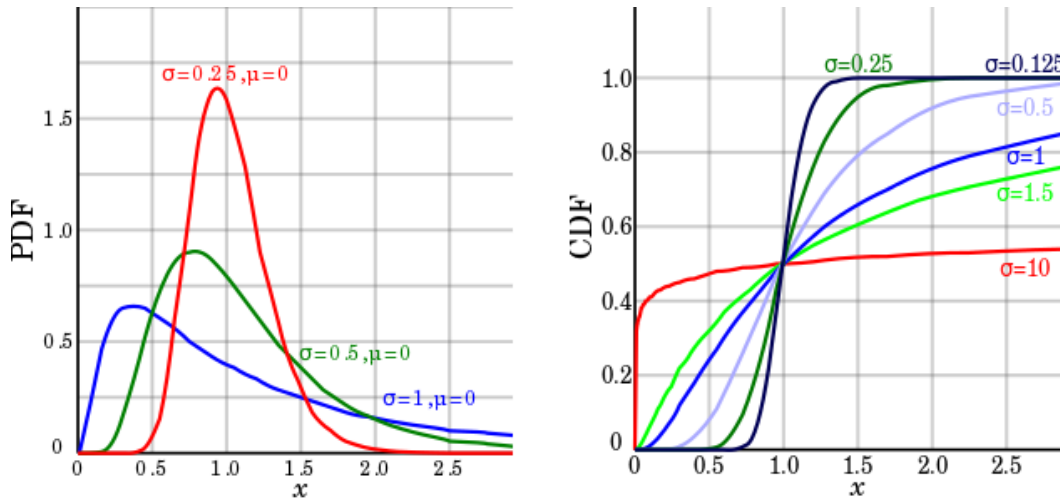
Η Cumulative Distribution Function (Cdf) είναι η :

$$F(x|\mu, \sigma) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left[\frac{\ln(x) - \mu}{\sqrt{2}\sigma} \right], x > 0, -\infty < \mu < +\infty, \sigma > 0 \quad (2.28)$$

Η Probability Density Function (Pdf) είναι η :

$$f(x|\mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}, x > 0, -\infty < \mu < +\infty, \sigma > 0 \quad (2.29)$$

Με δεδομένη μια τυχαία μεταβλητή X που ακολουθεί την κατανομή Lognormal και δύο παραμέτρους μ και σ που είναι αντίστοιχα η μέση τιμή και η τυπική απόκλιση του φυσικού λογαρίθμου της μεταβλητής, τότε ο λογάριθμος του X κατανέμεται κανονικά και μπορούμε να γράψουμε το X ως : $X = e^{\mu+\sigma Z}$, με το Z μία μεταβλητή που ακολουθεί κανονική κατανομή με μέση τιμή 0 και τυπική απόκλιση 1.



Σχήμα 2.7: Μορφές της Cdf (αριστερά) και Pdf (δεξιά) της κατανομής Lognormal για διαφορετικές τιμές των παραμέτρων μ, σ . (Πηγή: Wikipedia)

2.2.5 Η κατανομή Gamma

Η κατανομή Gamma βασίζεται σε μία οικογένεια κατανομών με δύο παραμέτρους. Οι chi-square1 και οι εκθετικές κατανομές, οι οποίες είναι «παιδιά» της κατανομής Gamma, είναι κατανομές μίας παραμέτρου, σταθεροποιώντας μία από τις δύο παραμέτρους Gamma. Συγκεκριμένα η κατανομή chi-square είναι μια ειδική περίπτωση της κατανομής Gamma και είναι μία από τις πιο διαδεδομένες κατανομές πιθανοτήτων στην επαγωγική στατιστική, ιδίως σε δοκιμές υποθέσεων ή στην κατασκευή διαστημάτων εμπιστοσύνης.

Η κατανομή Gamma είναι μια κατανομή δύο παραμέτρων και είναι ορισμένη στο $(0, +\infty)$. Παρατίθενται οι αναλυτικοί τύποι της συνάρτησης πυκνότητας πιθανότητας (Pdf) και της συνάρτησης αθροιστικής πιθανότητας (Cdf).

Η Cumulative Distribution Function (Cdf) είναι η :

$$F(x|k, \theta) = \frac{1}{\Gamma(k)} \gamma\left(k, \frac{x}{\theta}\right), x > 0, k > 0, \theta > 0 \quad (2.30)$$

Η Probability Density Function (Pdf) είναι η :

$$f(x|k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}, x > 0, k > 0, \theta > 0 \quad (2.31)$$

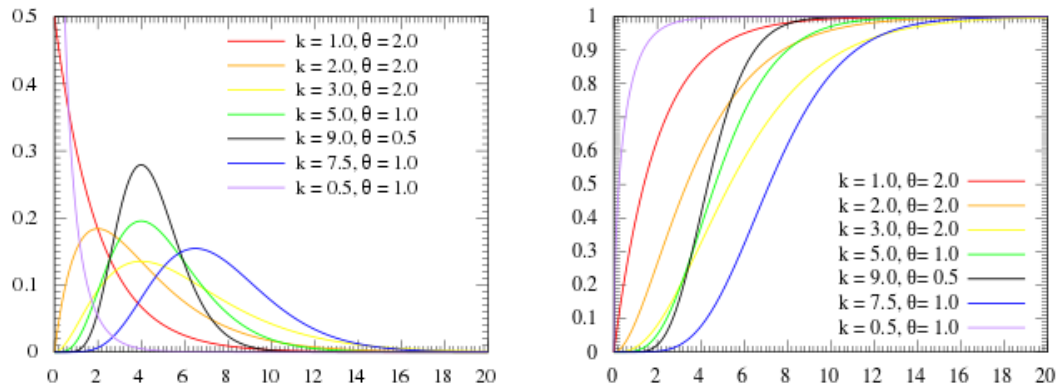
όπου :

$\Gamma(k)$: η συνάρτηση Γάμμα υπολογισμένη στο σημείο k ,

$\gamma\left(k, \frac{x}{\theta}\right)$: η ατελής συνάρτηση Γάμμα.

θ : είναι η παράμετρος κλίμακας

k : είναι η παράμετρος σχήματος



Σχήμα 2.8: Μορφές της Cdf (αριστερά) και Pdf (δεξιά) της κατανομής Gamma για διαφορετικές τιμές των παραμέτρων k, θ . (Πηγή: Wikipedia)

2.3 Μέθοδος των L Ροπών

Η μέθοδος των L-ροπών είναι μια στατιστική μεθοδολογία για την περιγραφή του σχήματος και των χαρακτηριστικών μίας κατανομής πιθανότητας.

Για μια τυχαία μεταβλητή X , η $r^{\text{η}}$ L-ροπή είναι:

$$\lambda_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} EX_{r-k:r} \quad (2.32)$$

όπου:

- $X_{k:n}$ είναι η $k^{\text{η}}$ στατιστική τάξη από ένα ανεξάρτητο δείγμα μεγέθους n από μία κατανομή X
- E είναι η αναμενόμενη τιμή

Συγκεκριμένα, οι πρώτες τέσσερις ροπές δίνονται από τις σχέσεις:

$$\lambda_1 = EX \quad (2.33)$$

$$\lambda_2 = (EX_{2:2} - XE_{1:2})/2 \quad (2.34)$$

$$\lambda_3 = (EX_{3:3} - 2XE_{2:3} + EX_{1:3})/3 \quad (2.35)$$

$$\lambda_4 = (EX_{4:4} - 3XE_{3:4} + 3EX_{2:4} - EX_{1:4})/4 \quad (2.36)$$

Υπάρχουν δύο συνηθισμένοι τρόποι με τους οποίους χρησιμοποιούνται οι L-ροπές, και στις δύο περιπτώσεις ανάλογα με τις κλασσικές ροπές:

- 1) Ως συνοπτικά στατιστικά στοιχεία για τα δεδομένα.
- 2) Να εξαχθούν εκτιμήσεις για τις παραμέτρους κατανομής πιθανοτήτων, εφαρμόζοντας τη μέθοδο των L ροπών και όχι των κλασσικών ροπών.

Εκτός από τις κλασσικές ροπές, η 2^η περίπτωση (εκτίμηση) γίνεται συχνότερα χρησιμοποιώντας μεθόδους μέγιστης πιθανοφάνειας. Ωστόσο, η χρήση των L-ροπών παρέχει πολλά πλεονεκτήματα. Συγκεκριμένα, οι L-ροπές είναι πιο ισχυρές από τις κλασσικές ροπές και η ύπαρξη υψηλότερων L-ροπών απαιτεί μόνο η τυχαία μεταβλητή να έχει πεπερασμένο μέσο όρο. Ένα μειονέκτημα των L-ροπών για εκτίμηση είναι η τυπικά μικρότερη ευαισθησία τους. Για παράδειγμα, η κατανομή Laplace έχει κύρτωση 6 και αδύναμες εκθετικές ουρές, αλλά μεγαλύτερη αναλογία 4ης L-ροπής από π.χ. η κατανομή Student-t με βαθμούς ελευθερίας 3, η οποία έχει άπειρη κύρτωση και πολύ πιο βαριές ουρές.

Ως παράδειγμα, λαμβάνουμε υπόψη ένα σύνολο δεδομένων με λίγα σημεία δεδομένων και μία απομακρυσμένη τιμή δεδομένων. Εάν ληφθεί η συνηθισμένη τυπική απόκλιση αυτού του συνόλου δεδομένων, θα επηρεαστεί πολύ από αυτό το ένα σημείο: ωστόσο, εάν ληφθεί η κλίμακα L, θα είναι πολύ λιγότερο ευαίσθητη σε αυτήν την τιμή δεδομένων. Κατά συνέπεια, οι L-ροπές είναι πολύ πιο σημαντικές

όταν αντιμετωπίζουμε δεδομένα ακραίων τιμών σε σχέση με τις κλασσικές ροπές. Ωστόσο, υπάρχουν και άλλες πιο κατάλληλες μέθοδοι για την επίτευξη μιας ακόμη υψηλότερης ευρωστίας από την απλή αντικατάσταση των ροπών με τις L-ροπές. Ένα παράδειγμα αυτού είναι η χρήση των L-ροπών ως συνοπτικών στατιστικών στη θεωρία ακραίων τιμών (Extreme Value Theory). Αυτή η εφαρμογή δείχνει την περιορισμένη ευρωστία των L-ροπών, δηλαδή τα στατιστικά L δεν είναι ανθεκτικά στατιστικά, καθώς μια ακραία τιμή μπορεί να τα αποβάλει, αλλά επειδή είναι μόνο γραμμικά (όχι στατιστικά υψηλότερης τάξης), επηρεάζονται λιγότερο από ακραίες τιμές από ότι οι κλασσικές ροπές.

Ένα άλλο πλεονέκτημα που έχουν οι L-ροπές έναντι των κλασσικών ροπών είναι ότι η ύπαρξή τους απαιτεί μόνο η τυχαία μεταβλητή να έχει πεπερασμένο μέσο όρο, οπότε οι L-ροπές υπάρχουν ακόμη και αν δεν υπάρχουν οι υψηλότερες κλασσικές ροπές (για παράδειγμα, για την κατανομή t του Student με χαμηλούς βαθμούς ελευθερίας). Απαιτείται επιπλέον μια πεπερασμένη διακύμανση προκειμένου τα τυπικά σφάλματα των εκτιμήσεων των L-ροπών να είναι πεπερασμένα.

Συνοψίζοντας ένα από τα βασικά πλεονεκτήματα της μεθοδολογίας των L-ροπών είναι πως δεν επηρεάζεται σημαντικά από τις ακραίες τιμές και το μήκος του δείγματος. Αυτό είναι ιδιαίτερα σημαντικό στις περιπτώσεις που διαθέτουμε χρονοσειρές με μικρό πλήθος παρατηρήσεων. Για αυτούς τους λόγους, η χρήση τους τείνει να υπερισχύει την χρήση άλλων μεθόδων, όπως τον κλασσικών ροπών (Perez et al., 2003).

Μέσω της διαδικασίας L-ροπών υπολογίζονται οι στατιστικές εκτιμήτριες $\lambda_1, \lambda_2, \lambda_3$ και λ_4 , που με την χρήση των παρακάτω εξισώσεων αντιστοιχούν στις στατιστικές παραμέτρους της μέσης τιμής, του συντελεστής μεταβλητότητας, της παραμέτρου της ασυμμετρίας και της κύρτωσης για το δείγμα που μελετάμε.

Για την μέση τιμή (L-moment mean) :

$$\tau_1 = \lambda_1 \quad (2.37)$$

Η μέση αριθμητική τιμή ή μέσος όρος προκύπτει από το ηλίκο διαίρεσης του αθροίσματος των τιμών μιας μεταβλητής δια του συνολικού πλήθους τους, δηλαδή του συνόλου των συχνοτήτων τους. Αποτελεί το σπουδαιότερο και χρησιμότερο μέτρο της στατιστικής, και αποτελεί μέτρο θέσης, δηλαδή δείχνει σχετικά τις θέσεις των αριθμών στους οποίους αναφέρεται.

Για τον συντελεστή μεταβλητότητας (L-moment coefficient of variation):

$$\tau_2 = \frac{\lambda_2}{\lambda_1} \quad (2.38)$$

Ο συντελεστής μεταβλητότητας είναι ένα στατιστικό μέτρο που είναι χρήσιμο στις περιπτώσεις στην σύγκριση ομάδων τιμών, που είτε εκφράζονται σε διαφορετικές μονάδες μέτρησης είτε εκφράζονται στην ίδια μονάδα μέτρησης αλλά έχουν σημαντικά διαφορετικές μέσες τιμές.

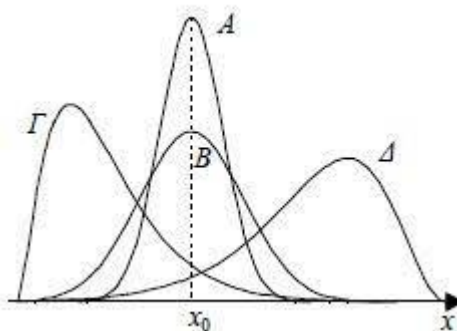
Ο συντελεστής μεταβλητότητας εκφράζεται επί τοις εκατό, είναι συνεπώς ανεξάρτητος από τις μονάδες μέτρησης και παριστάνει ένα μέτρο σχετικής διασποράς των τιμών και όχι της απόλυτης διασποράς. Εκφράζει, δηλαδή, τη μεταβλητότητα των δεδομένων απαλλαγμένη από την επίδραση της μέσης τιμής.

Για το μέτρο ασυμμετρίας (L-moment Coefficient of skewness):

$$\tau_3 = \frac{\lambda_3}{\lambda_2} \quad (2.39)$$

Η κατανομή του πληθυσμού που μελετάται μπορεί να είναι είτε συμμετρική είτε μη συμμετρική. Στην περίπτωση συμμετρικής κατανομής η κορυφή, διάμεσος και μέση τιμή συμπίπτουν, ενώ στην περίπτωση της μη συμμετρικής κατανομής μια από τις τρεις παραμέτρους η κορυφή της κατανομής την διασπά σε ασύμμετρα, ως προς το πλήθος των παρατηρήσεων, τμήματα.

Το είδος της ασυμμετρίας χωρίζεται σε δύο κατηγορίες, την θετική και την αρνητική, οι οποίες ορίζουν μια κατανομή στην οποία οι περισσότερες παρατηρήσεις τοποθετούνται στη δεξιά και αριστερή, αντίστοιχα, πλευρά της κορυφής της. (Κοκολάκης, 2009).



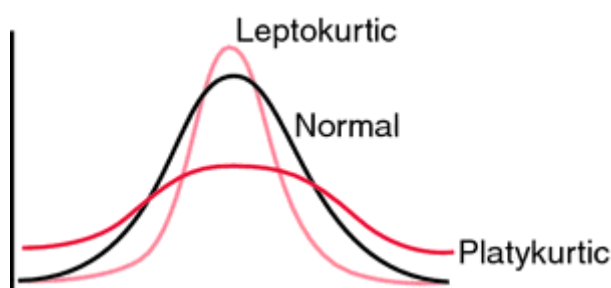
Σχήμα 2.9: Παράδειγμα προσαρμογής κατανομής για διάφορες τιμές του στατιστικού μέτρου της ασυμμετρίας με (Γ) αρνητική ασυμμετρία, (Α), (Β) μηδενική ασυμμετρία, (Δ) θετική ασυμμετρία (Πηγή:Wikipedia)

Για το μέτρο κυρτότητας (L-moment of Kurtosis):

$$\tau_4 = \frac{\lambda_4}{\lambda_2^2} \quad (2.40)$$

Μια κατανομή η οποία έχει σχετικά μεγάλη μέγιστη συχνότητα (κορυφή) και επομένως, υπάρχει μια μεγάλη συγκέντρωση τιμών γύρω από το μέσο λέγεται λεπτόκυρτη. Αντιθέτως στην περίπτωση που η μέγιστη συχνότητα της είναι σχετικά μικρή λέγεται πλατύκυρτη, ενώ την περίπτωση που η κατανομή προσεγγίζει την κανονική ονομάζεται μεσόκυρτη.

Έχοντας υπόψη πως οι κανονικές κατανομές έχουν τιμή κυρτότητας ίση με 3, είναι σύνηθες να ορίζεται η κυρτότητα μιας κατανομής με βάση την απόκλιση της από την παραπάνω σταθερά. Δηλαδή, κατανομές με κυρτότητα μεγαλύτερη της τιμής 3 ονομάζονται λεπτόκυρτες, ενώ με αρνητική διαφορά πλατύκυρτες.



Σχήμα 2.10: Παράδειγμα προσαρμογής κατανομής για διάφορες τιμές του στατιστικού μέτρου της κυρτότητας (Πηγή:Wikipedia)

2.4 Κριτήρια Επιλογής Εκτιμητών

2.4.1 Καταλληλότητα στατιστικής συνάρτησης

Πώς θα μπορούσαμε να αποφανθούμε για την καταλληλότητα μιας στατιστικής συνάρτησης $T(\underline{X})$ ως εκτιμητή του $g(\theta)$; Φαινομενικά, το απόλυτο σφάλμα εκτίμησης $|T(\underline{X}) - g(\theta)|$, δίνει απάντηση σε αυτό το ερώτημα: 'Όσο πιο «μικρό», τόσο πιο «κοντά» είναι ο $T(\underline{X})$ στο $g(\theta)$. Όμως, οφείλουμε να αναγνωρίσουμε ότι η παράσταση, $|T(\underline{X}) - g(\theta)|$ έχει άγνωστη και τυχαία τιμή, επειδή περιέχει την άγνωστη τιμή $g(\theta)$, και τυχαία επειδή, είναι τυχαία μεταβλητή ως συνάρτηση του \underline{X} . Αν λοιπόν X είναι το σύνολο τιμών του \underline{X} , οποιαδήποτε από τις τιμές $|T(\underline{Y}) - g(\theta)|$, $\underline{y} \in X$, $\theta \in \Theta$ θα μπορούσε να είναι η τιμή του απολύτου σφάλματος. Δεν γνωρίζουμε όμως ποια είναι. Επομένως, το απόλυτο σφάλμα ως μη «μετρήσιμο», δεν εξυπηρετεί ως κριτήριο επιλογής ή σύγκρισης εκτιμητών. Προφανώς, στο ίδιο συμπέρασμα θα καταλήξουμε αν αποπειραθούμε να αξιολογήσουμε τον εκτιμητή $T(\underline{X})$ χρησιμοποιώντας παρόμοιες ποσότητες, όπως το τετραγωνικό σφάλμα εκτίμησης $(T(\underline{X}) - g(\theta))^2$ ή ακόμη μια δύναμη του απολύτου σφάλματος $(T(\underline{X}) - g(\theta))^k$, $k > 0$.

Από την παραπάνω ανάλυση γίνεται αντιληπτό ότι, για να καταλήξουμε σε ένα χρήσιμο κριτήριο θα πρέπει κατά κάποιο τρόπο να διαχειριστούμε την τυχαιότητα του σφάλματος εκτίμησης, αλλά και το άγνωστο (του) θ . Προς αυτήν την κατεύθυνση μερικά σημαντικά κριτήρια είναι τα εξής.

1. Μέσο απόλυτο σφάλμα – μέσο τετραγωνικό σφάλμα – Μέση ζημία

(Διαχείριση της τυχαιότητας του σφάλματος)

Αντί της τυχαίας τιμής (τυχαίας μεταβλητής) $|T(\underline{X}) - g(\theta)|$, θεωρούμε την κατά μέσο όρο τιμή της ως προς την κατανομή του \underline{X} , δηλαδή λαμβάνουμε ως κριτήριο το μέσο απόλυτο σφάλμα $E_{\theta}|T(\underline{X}) - g(\theta)|$, το οποίο εξαρτάται μόνον από το θ . Τώρα, ναι μεν το θ έχει μία και μοναδική τιμή στο Θ , όμως αυτή μπορεί να είναι οποιοδήποτε στοιχείο του Θ . Επομένως, για να μελετήσουμε το μέσο απόλυτο σφάλμα, επιβάλλεται να το θεωρήσουμε ως συνάρτηση του θ , αφήνοντας το θ να διατρέχει το Θ (δηλαδή το πεδίο ορισμού της συνάρτησης είναι το Θ). Μάλιστα, ο δείκτης στην παραπάνω μέση τιμή τονίζει την εξάρτησή της από το θ . Στη συνέχεια, λοιπόν, χωρίς να υπάρχει κίνδυνος παρερμηνείας, θα συμβολίζουμε με θ , όχι μόνον την άγνωστη παράμετρο (που έχει μία και μοναδική τιμή), αλλά και οποιοδήποτε στοιχείο του Θ . Με το ίδιο σκεπτικό, ως κριτήριο μπορεί να ληφθεί το μέσο τετραγωνικό σφάλμα $E_{\theta}(T(\underline{X}) - g(\theta))^2$ ή ακόμη $E_{\theta}(T(\underline{X}) - g(\theta))^k$, $k > 0$, καθένα μάλιστα θεωρούμενο ως συνάρτηση του $\theta \in \Theta$. Ακόμη γενικότερα μπορούμε να θεωρήσουμε μία

συνάρτηση $L(t, \theta)$, που την ονομάζουμε συνάρτηση ζημίας και η οποία παριστάνει, ποσοτικά, τη «ζημία μας» (ή την ποινή ή το σφάλμα), εάν εκτιμήσουμε το $g(\theta)$ με την τιμή t . Συνήθως η συνάρτηση $L(t, \theta)$ λαμβάνεται τέτοια ώστε

$$L(t, \theta) \geq 0, \quad \forall \theta, t \quad (2.41)$$

$$L(g(\theta), \theta) = 0, \quad \forall \theta \in \Theta. \quad (2.42)$$

Οι ποιοτικές αυτές ιδιότητες δηλώνουν ότι η «ζημία μας» είναι μηδέν ή θετική και ειδικά, στην περίπτωση που η εκτίμηση t συμπίπτει με την τιμή $g(\theta)$, μηδενίζεται. Εάν $T(\underline{X})$ είναι εκτιμητής του $g(\theta)$, η συνάρτηση (του θ):

$$R(T, \theta) = E_{\theta} \left(L \left(T(\underline{X}), \theta \right) \right) \quad (2.43)$$

λέγεται συνάρτηση κινδύνου του εκτιμητή $T(\underline{X})$ και παριστάνει τη «μέση ζημία μας» από τη χρησιμοποίηση του $T(\underline{X})$ ως εκτιμητή του $g(\theta)$. Προφανώς, «μικρές» τιμές $R(T, \theta)$, $\theta \in \Theta$ συνηγορούν υπέρ της καταλληλότητας του εκτιμητή T . Σημειώνουμε ότι για $L(t, \theta) = |t - g(\theta)|^r$ με $r = 1, 2, \dots$ προκύπτουν τα παραπάνω αναφερθέντα κριτήρια επιλογής. Ιδανικά, θα θέλαμε να βρούμε έναν εκτιμητή $T(\underline{X})$ που ελαχιστοποιεί το $R(T, \theta)$ οποιαδήποτε και εάν είναι η τιμή του θ , δηλαδή για κάθε $\theta \in \Theta$. Για λόγους απλότητας και ευκολίας στον υπολογισμό του, επιλέχθηκε όπως θα αναλυθεί στην επόμενη υποενότητα, το μέσο τετραγωνικό σφάλμα $E_{\theta} \left(T(\underline{X}) - g(\theta) \right)^2$, το οποίο είναι ίσως το πιο ευρέως χρησιμοποιούμενο στην πράξη κριτήριο σύγκρισης εκτιμητών.

2.4.2 Μέσο τετραγωνικό Σφάλμα (MSE)

Στην υποενότητα αυτή μελετάμε το μέσο τετραγωνικό σφάλμα, το πιο γνωστό και συνάμα ευρέως χρησιμοποιούμενο, στη θεωρία και στις εφαρμογές της Στατιστικής, κριτήριο σύγκρισης και αξιολόγησης εκτιμητών.

Το μέσο τετραγωνικό σφάλμα (ΜΤΣ) εκτιμητή $T(\underline{X})$ της τιμής $g(\theta)$ ορίστηκε στην υποενότητα 2.5.1 και δίνεται από τη σχέση:

$$ΜΤΣ(T(\underline{X}), \theta) = E_{\theta} \left(T(\underline{X}) - g(\theta) \right)^2 \quad (2.44)$$

Επίσης, το τυπικό σφάλμα (ΤΣ) του $T(\underline{X})$ ορίζεται από τη σχέση

$$ΤΣ(T(\underline{X}), \theta) = \sqrt{ΜΤΣ(T(\underline{X}), \theta)} \quad (2.45)$$

Σε αριθμητικούς υπολογισμούς, θεωρείται προτιμότερο να ανακοινώνεται η τιμή του τυπικού σφάλματος αντί της τιμής του μέσου τετραγωνικού σφάλματος. Αυτό δικαιολογείται επειδή το ΜΤΣ αξιολογεί τις τετραγωνικές αποκλίσεις των τιμών του $T(\underline{X})$ από το $g(\theta)$, ενώ υπολογίζοντας την τετραγωνική ρίζα του ΜΤΣ, το ΤΣ είναι άμεσα συγκρίσιμο με την εκτίμηση του $g(\theta)$ και επιπλέον έχει τις ίδιες μονάδες μέτρησης με το $g(\theta)$. Σημειώνουμε επίσης ότι το ΤΣ ή ένας εκτιμητής του ΤΣ καθορίζει σε ορισμένες περιπτώσεις το εύρος του διαστήματος εμπιστοσύνης για το $g(\theta)$.

Ο συμβολισμός $MTS(T(\underline{X}), \theta)$ ή πιο απλά $MTS(T, \theta)$ δηλώνει ότι το ΜΤΣ εξαρτάται, εκτός φυσικά από τον συγκεκριμένο εκτιμητή $T(\underline{X})$, εν γένει, και από την άγνωστη παράμετρο $\theta \in \Theta$, είναι δηλαδή συνάρτηση του $\theta \in \Theta$. Η φυσική του ερμηνεία είναι έκδηλη στον ορισμό του : παριστάνει την «κατά μέσο όρο» τετραγωνική απόκλιση του εκτιμητή από την τιμή $g(\theta)$, που έχει κληθεί να εκτιμήσει. Η καθιέρωση του ΜΤΣ ως κριτηρίου οφείλεται σε ένα μεγάλο βαθμό στο γεγονός ότι μπορεί να υπολογιστεί σχετικά εύκολα μέσω ενός απλού και χρήσιμου τύπου ο οποίος ισχύει για οποιονδήποτε εκτιμητή. Πιο συγκεκριμένα, στην επόμενη πρόταση παρατηρούμε ότι ο υπολογισμός του ΜΤΣ (ως συνάρτησης του θ) απαιτεί μόνον την εύρεση δύο βασικών παραμέτρων, της μέσης τιμής και της διασποράς του εκτιμητή και όχι κατ' ανάγκη τη γνώση της κατανομής του. Ας προσθέσουμε ακόμη ότι η άμεση σύνδεση του ΜΤΣ με τις δύο αυτές παραμέτρους, γνωστές σε κάθε χρήστη Στατιστικής, συνεισφέρει από μόνη της στη δημοφιλία και στην κατανόηση του ΜΤΣ. Μάλιστα, για εκτιμητές που έχουν την ιδιότητα της αμεροληψίας, το ΜΤΣ συμπίπτει με τη διασπορά τους. Τονίζουμε εδώ ότι το πλεονέκτημα του γενικού απλού τύπου υπολογισμού δεν το έχουν άλλα επίσης ορθολογικά κριτήρια, όπως, π.χ. το μέσο απόλυτο σφάλμα $E_{\theta}|T(\underline{X}) - g(\theta)|$. Τέλος, όσον αφορά τον συμβολισμό, ας μην ξεχνάμε ότι $T(\underline{X})$ ή T παριστάνουν τον ίδιο εκτιμητή.

3 ΥΠΟΛΟΓΙΣΤΙΚΑ ΕΡΓΑΛΕΙΑ

3.1 Εργαλείο Ανάλυσης Δεδομένων – Στατιστικό Πακέτο R

Στην παρούσα διπλωματική εργασία η ανάλυση των δεδομένων πραγματοποιήθηκε και ελέγχθηκε σε μεγάλο μέρος της με τη γλώσσα προγραμματισμού R.

3.1.1 Τι είναι η R

Η γλώσσα R είναι μια γλώσσα προγραμματισμού συνοδευόμενη από ένα περιβάλλον για επεξεργασία δεδομένων, υπολογισμών και γραφημάτων. Αν και χρησιμοποιείται κυρίως στην στατιστική οι δημιουργοί του προτιμούν να το αποκαλούν εργαλείο για ανάλυση δεδομένων τονίζοντας ότι περιλαμβάνει και μοντέρνες και παλιές στατιστικές μεθοδολογίες. Είναι μια γλώσσα ελεύθερου λογισμικού (GPL) με πολλά πακέτα διαθέσιμα στους χρήστες για την εκτέλεση πολλών στατιστικών (στατιστικά τεστ, γραμμική και μη γραμμική ανάλυση, κλασική στατιστική ανάλυση, ανάλυση χρονοσειρών, ταξινόμηση και ομαδοποίηση) και γραφικών τεχνικών.

Είναι παρόμοια με τη γλώσσα και το περιβάλλον S που αναπτύχθηκε στα εργαστήρια της Bell (στο παρελθόν AT&T, τώρα Lucent Technologies) από τον John Chambers και τους συνεργάτες του. Η γλώσσα R μπορεί να θεωρηθεί ως μια διαφορετική υλοποίηση του S. Υπάρχουν μερικές σημαντικές διαφορές, αλλά ο περισσότερος κώδικας που γράφεται για το S τρέχει αμετάβλητος και στην R (όμως το S και η R δεν είναι 100% συμβατά). Η R συγκρινόμενη με το εμπορικό πακέτο S-plus, που είναι βασισμένο και αυτό στην γλώσσα προγραμματισμού S, παρουσιάζει τη διαφορά ότι δεν έχει γραφικό περιβάλλον εργασίας.

Η R γλώσσα προγραμματισμού ξεκίνησε από τους Robert Gentleman και Ross Ihaka του “Statistics Department of the University of Auckland” το 1995. Πλέον, συντηρείται από το R-core-development team που αποτελεί ένα διεθνές κύκλωμα εθελοντών για την ανάπτυξη και συντήρησή του. Στηρίζεται δε στην ανάπτυξη προγραμμάτων, μέσω πακέτων (packages), τα οποία διατίθενται πάλι ελεύθερα από χρήστες ανά τον κόσμο. Αξιοσημείωτο είναι ότι δεν υπάρχει εμπορική υποστήριξη οπότε αρωγός σε όλα είναι η διεθνής επιστημονική κοινότητα. Την R μπορεί κάποιος να την αποκτήσει στην ιστοσελίδα του: <http://www.r-project.org/> ή από ένα από τα πολλά πρότυπα (mirrors) του Cran (Comprehensive R Archive) <http://www.cran.r-project.org/> το οποίο είναι ένα δίκτυο διανομής του R σε πολλά μέρη του κόσμου μέσω διαδικτύου. Χρησιμοποιείται ευρέως ως εκπαιδευτική γλώσσα και ως ερευνητικό εργαλείο και

μπορεί να μεταγλωττιστεί και να εκτελεστεί σε μεγάλη γκάμα λειτουργικών συστημάτων, όπως τα Windows της Microsoft, το MacOS της Apple και τα Ubuntu του Linux.

3.1.2 Πλεονεκτήματα της γλώσσας R

Η γλώσσα προγραμματισμού R παρέχει ένα πλήρες σύστημα βοήθειας ενσωματωμένο στο λογισμικό και πλήθος εντολών και στατιστικών συναρτήσεων. Ωστόσο, η δυνατότητα που παρέχει για τη συγγραφή νέων συναρτήσεων, την κάνει ένα πολύ ισχυρό εργαλείο προγραμματισμού. Τα τελευταία χρόνια η R έχει αναπτυχθεί ραγδαία σε όλες τις επιστήμες που ασχολούνται με την στατιστική έρευνα, συμπεριλαμβανομένης και της υδρολογίας. Το γεγονός αυτό δικαιολογείται από την ικανότητα που έχει, αφενός να διαχειρίζεται μεγάλο όγκο δεδομένων και αφετέρου να αναλύει και να επεξεργάζεται δεδομένα με μεγάλη ταχύτητα. Μπορεί να χρησιμοποιηθεί είτε με απευθείας εντολές που υπάρχουν είτε με προγράμματα τα οποία ο χρήστης μπορεί να αναπτύξει για επίλυση πιο πολύπλοκων στατιστικών προβλημάτων. Επίσης, ο χρήστης μπορεί να δανειστεί και έτοιμα προγράμματα τα οποία είναι ενσωματωμένα μέσα σε πακέτα που διατίθενται ελεύθερα. Η ποικιλία τέτοιων προγραμμάτων είναι τεράστια.

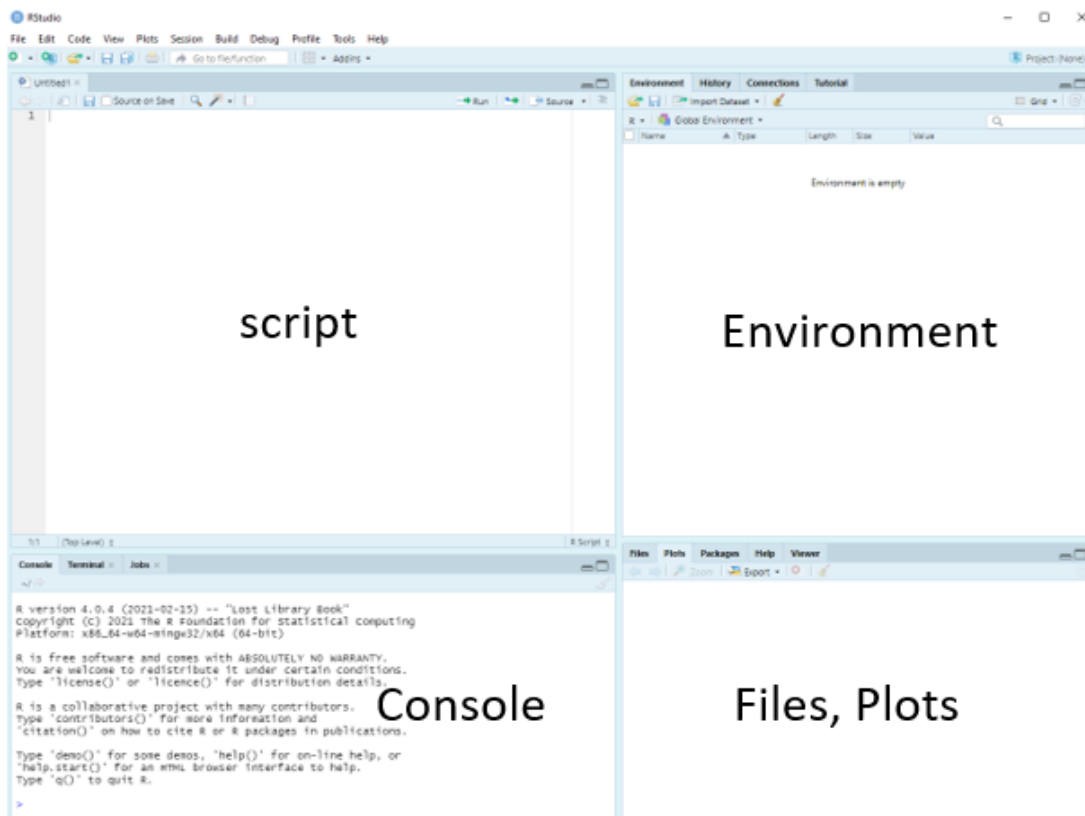
Αναλυτικότερα, είναι μια γλώσσα προγραμματισμού που χρησιμεύει κατεξοχήν στην επεξηγηματική ανάλυση δεδομένων τα οποία εισάγονται με τη μορφή διανύσματος ή πίνακα, καθώς και στην εφαρμογή διαφόρων στατιστικών μοντέλων, λόγοι που έπαιξαν καθοριστικό ρόλο στην επιλογή του εν λόγω εργαλείου. Μπορεί να χρησιμοποιηθεί είτε με κατευθείαν εντολές, είτε με προγράμματα τα οποία μπορούν να αναπτυχθούν και να δοθούν για εκτέλεση. Δίνει την εξαιρετική δυνατότητα στο χρήστη να προγραμματίζει βάση των ερευνητικών του αναγκών και να κατασκευάζει ειδικές συναρτήσεις (functions), οι οποίες χρησιμεύουν για ανάπτυξη ιδίων προγραμμάτων. Η φιλοσοφία του προγραμματισμού στην R βασίζεται στη δημιουργία καινούριων συναρτήσεων, οι οποίες θα χρησιμοποιηθούν για περαιτέρω ανάπτυξη της γλώσσας. Τα κύρια δομικά στοιχεία της είναι οι υπάρχουσες συναρτήσεις (functions), που εκτελούνται μέσω λογικών τελεστών και τελεστών σύγκρισης, και τα πακέτα (packages) εκτέλεσης προσχεδιασμένων σειρών συναρτήσεων. Η R μπορεί να επεκταθεί (εύκολα) μέσω των πακέτων. Υπάρχουν περίπου οκτώ πακέτα που παρέχονται με την ελεύθερη εγκατάσταση της R. Πολλά ακόμη, όμως, είναι διαθέσιμα μέσα από την οικογένεια CRAN των δικτυακών τόπων που καλύπτουν ένα ευρύτατο φάσμα σύγχρονων στατιστικών και εντοπίζονται στην επίσημη ιστοσελίδα της R.

Επιπροσθέτως, η R αποτελεί μία γλώσσα προγραμματισμού - διερμηνέα. Αυτό σημαίνει ότι οι εντολές διαβάζονται και μετά εκτελούνται αμέσως. Αντίθετα, η C και η Fortran είναι γλώσσες προγραμματισμού - μεταγλωττίστριες στις οποίες ολοκληρωμένα προγράμματα μεταφράζονται με τη βοήθεια ενός μεταγλωττιστή στην κατάλληλη γλώσσα μηχανής. Το μεγάλο πλεονέκτημα των διερμηνέων γλωσσών προγραμματισμού είναι ότι επιτρέπουν τη σταδιακή ανάπτυξη. Με άλλα λόγια, μια συνάρτηση μπορεί να δημιουργηθεί, να εκτελεσθεί και μετά να δημιουργηθεί μια καινούργια συνάρτηση η οποία να καλεί την προηγούμενη κ.ο.κ. (Φωκιανός Κ., Χαραλάμπους Χ., 2010). Αυτό δίνει την επιπλέον δυνατότητα στο χρήστη της R να πληκτρολογεί εκφράσεις οι οποίες άμεσα εκτελούνται μία προς μία.

Ένα επιπλέον δυνατό σημείο της R είναι η ευκολία με την οποία παράγει καλά σχεδιασμένα διαγράμματα οπτικής απεικόνισης των Στατιστικών και Μαθηματικών Διαδικασιών που εκτελούνται, σε άριστη ποιότητα ηλεκτρονικής ή έντυπης εκτύπωσης, τα οποία μπορούν εκτός των άλλων να περιλαμβάνουν μαθηματικά σύμβολα και τύπους, όπου χρειάζεται. Δεδομένου ότι τα γραφήματα είναι πολύ σημαντικά για την οπτική αναπαράσταση των δεδομένων και καθοδηγούν τον στατιστικό στη διαδικασία της μοντελοποίησης και αξιολόγησης της ανάλυσης, η ευκολία της δημιουργίας των γραφημάτων αυτών και η ποικιλία των εναλλακτικών επιλογών που παρέχονται από το πρόγραμμα ήταν ανέκαθεν βασικό ζητούμενο της διεκπεραίωσης της έρευνάς μας. Ο χρήστης διατηρεί τον πλήρη έλεγχο της δημιουργίας των οπτικών απεικονίσεων καθ' όλη τη διάρκεια της ανάλυσης των δεδομένων.

3.1.3 R-Studio

Στη συγκεκριμένη διπλωματική εργασία χρησιμοποιήθηκε το πρόγραμμα RStudio, ένα ελεύθερο προγραμματιστικό περιβάλλον για την γλώσσα R, το οποίο παρέχει μια οργανωμένη διάταξη και διάφορες πρόσθετες επιλογές. Το πρόγραμμα αυτό παρέχει μια ευρεία γκάμα δυνατοτήτων για την δημιουργία, ανάλυση, πρόβλεψη και απεικόνιση χρονοσειρών. Το RStudio είναι ένα εξαιρετικά χρήσιμο εργαλείο, καθώς δίνει τη δυνατότητα γραφής και εκτέλεσης του κώδικα, καθώς και εξαγωγής αποτελεσμάτων και γραφημάτων. Στο παρακάτω σχήμα φαίνεται το προγραμματιστικό περιβάλλον RStudio, το οποίο είναι χωρισμένο σε τέσσερα τμήματα:



Σχήμα 3.1: Προγραμματιστικό περιβάλλον R-Studio

Στο πάνω αριστερά παράθυρο βρίσκονται τα αρχεία κειμένου (script) που περιέχουν τις εντολές σε γλώσσα R και στο κάτω αριστερά φαίνεται η κονσόλα της R δηλαδή το παράθυρο των αποτελεσμάτων των γραπτών εντολών (commands) που δίνονται είτε απευθείας γραμμή προς γραμμή στην κονσόλα είτε μέσω ενός αρχείου text (script) από το πάνω αριστερά παράθυρο. Στο πάνω δεξιά είναι η καρτέλα Environment όπου φαίνονται όλες οι αποθηκευμένες μεταβλητές και συναρτήσεις, καθώς και το ιστορικό των πρόσφατων εντολών που έχουν εκτελεστεί και στο κάτω δεξιά φαίνεται ένα σύνολο καρτελών, όπως η καρτέλα αρχεία (Files) που δείχνει κάθε φάκελο και αρχείο στο καθορισμένο χώρο εργασίας του προγράμματος (workspace), η καρτέλα γραφήματα (Plots) που δείχνει όλα τα γραφήματα, η καρτέλα πακέτα (Packages) που περιέχει τη λίστα των βιβλιοθηκών που είναι εγκατεστημένες αλλά όχι ενεργές κατ' ανάγκη και η καρτέλα βοήθεια (Help) που δείχνει τη βοήθεια για την R (διάφορες θεωρητικές πληροφορίες για τα πακέτα, τις συναρτήσεις και γενικά για το RStudio).

3.2 Το ελεύθερο λογισμικό QGIS

Στην παρούσα διπλωματική εργασία η παραγωγή εποπτικών χαρτών, που συνέβαλαν στην εξαγωγή συμπερασμάτων και ερμηνείας των αποτελεσμάτων που προέκυψαν από την ανάλυση των δεδομένων, πραγματοποιήθηκε στο ελεύθερο λογισμικό QGIS.

3.2.1 Τι είναι το QGIS

Τα Γεωγραφικά Συστήματα Πληροφοριών (Γ.Σ.Π.) – Geographical Information Systems (G.I.S.) είναι μια σύγχρονη τεχνολογία, που αναπτύσσεται ραγδαία και εμφανίζει τη δυσκολία προσέγγισης ενός και μόνο αυστηρού και ολοκληρωμένου ορισμού της, που να περιέχει όλα τα στοιχεία που τη προσδιορίζουν. Κατά καιρούς έχουν δοθεί αρκετοί και διάφοροι ορισμοί για τα ΓΣΠ, ορισμοί που διαμορφώθηκαν ανάλογα με τις εμπειρίες τις εφαρμογές και τις ανάγκες αυτών που χρησιμοποιούσαν τα συστήματα αυτά (Maquire, 1991). Έτσι λοιπόν μέχρι σήμερα έχουν καταγραφεί οι κάτωθι ορισμοί:

- Μορφή Συστήματος Διαχείρισης Πληροφοριών (MIS), που επιτρέπει τη δημιουργία – παρουσίαση χαρτών από γενικές πληροφορίες (Devine & Field, 1986).
- Δυναμικό σύνολο εργαλείων για συλλογή, αποθήκευση, πρόσβαση, μετασχηματισμό και παρουσίαση χωρικών δεδομένων (Burrough, 1986) .
- Τεχνολογία πληροφοριών η οποία αποθηκεύει, αναλύει και παρουσιάζει ταυτόχρονα χωρικές και μη χωρικές πληροφορίες (Parker, 1988).
- Σύστημα που υποβοηθά τη λήψη αποφάσεων, σχετικά με ένα πρόβλημα που έχει τεθεί, βάσει της διαχείρισης χωρικών δεδομένων (Cowen, 1988).
- Οποιοδήποτε, καθοδηγούμενο από τον άνθρωπο ή βασισμένο σε υπολογιστή, σύνολο διαδικασιών που χρησιμοποιείται για την αποθήκευση και διαχείριση γεωγραφικών δεδομένων (Aronoff, 1989).
- Σύστημα με ανεπτυγμένες δυνατότητες γεω-μοντελοποίησης (Koshkariion, et al., 1989).
- Μια οργανωμένη συλλογή υπολογιστικού εξοπλισμού, λογισμικού, γεωγραφικών δεδομένων και προσωπικού, σχεδιασμένο να συλλαμβάνει, να αποθηκεύει, να ενημερώνει, να χειρίζεται, να αναλύει και να εμφανίζει όλες τις μορφές των γεωγραφικά αναφερόμενων πληροφοριών (ESRI, 1990).

Από αυτούς τους ορισμούς διαφαίνεται η μετάβαση από την θεώρηση του GIS ως λογιστικού συστήματος για μια συγκεκριμένη εφαρμογή, σε ένα γενικότερο σύνολο εργαλείων υλικού και λογισμικού που χρησιμοποιούνται για τη διευκόλυνση της

χρήσης γεωγραφικών πληροφοριών για ανάλυση και μοντέλα δεδομένων και για την επίλυση προβλημάτων (Wieczorek & Delmerico, 2009).

Ένας πλήρης - σύγχρονος, ορισμός που περιλαμβάνει όλες τις πτυχές ενός Σύστηματος Γεωγραφικών Πληροφοριών θα μπορούσε να είναι: Το Σύστημα Γεωγραφικών Πληροφοριών (ΣΓΠ), γνωστό ευρέως και ως G.I.S. (Geographic Information Systems), είναι σύστημα διαχείρισης χωρικών δεδομένων (spatial data) και συσχετισμένων ιδιοτήτων. Στην πιο αυστηρή μορφή του είναι ένα ψηφιακό σύστημα, ικανό να ενσωματώσει, αποθηκεύσει, προσαρμόσει, αναλύσει και παρουσιάσει γεωγραφικά συσχετισμένες (geographically-referenced) πληροφορίες. Σε πιο γενική μορφή, ένα ΣΓΠ είναι ένα εργαλείο "έξυπνου χάρτη", το οποίο επιτρέπει στους χρήστες του να αποτυπώσουν μια περίληψη του πραγματικού κόσμου, να δημιουργήσουν διαδραστικά ερωτήσεις χωρικού ή περιγραφικού χαρακτήρα (αναζητήσεις δημιουργούμενες από τον χρήστη), να αναλύσουν τα χωρικά δεδομένα (spatial data), να τα προσαρμόσουν και να τα αποδώσουν σε αναλογικά μέσα (εκτυπώσεις χαρτών και διαγράμμάτων) ή σε ψηφιακά μέσα (αρχεία χωρικών δεδομένων, διαδραστικοί χάρτες στο Διαδίκτυο) (WIKIPEDIA, 2021).

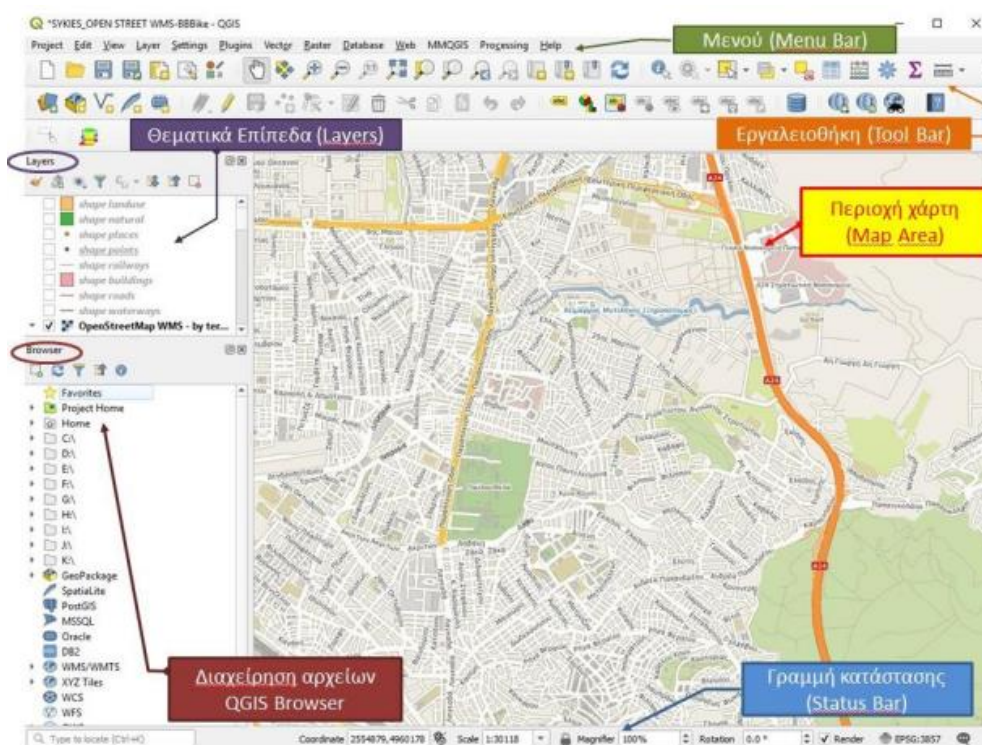
Το Quantum GIS ή QGIS (<https://www.qgis.org/en/site/>) είναι ένα φιλικό ανοιχτού κώδικα (open source) Σύστημα Γεωγραφικών Πληροφοριών (GIS) αναπτυγμένο κάτω από την άδεια GNU (<https://www.gnu.org/licenses/licenses.html>) General Public License (GPL). Η συγκεκριμένη άδεια παρέχει τη δυνατότητα επιθεώρησης και τροποποίησης του πηγαίου κώδικα, δίνοντας την εγγύηση στον χρήστη πως θα έχει πάντα πρόσβαση σε ένα δωρεάν λογισμικό GIS, που μπορεί να τροποποιήσει ελεύθερα. Το QGIS project ξεκίνησε τον Μάιο του 2002 με σκοπό την ανάπτυξη ενός ενημερωμένου και δωρεάν λογισμικού GIS για οποιονδήποτε έχει πρόσβαση σε έναν Προσωπικό Υπολογιστή (Personal Computer - PC). Εδραιώθηκε επίσημα τον Ιούνιο του ίδιου έτους μέσω της διαδικτυακής υπηρεσίας SourceForge (<https://sourceforge.net/>), που προσφέρει στους προγραμματιστές λογισμικού μια κεντρική online θέση για τον έλεγχο και τη διαχείριση έργων λογισμικού ελεύθερης και ανοικτής πηγής. Επίσης, υποστηρίζεται από το Open Source Geospatial Foundation (OSGeo) (<https://www.osgeo.org/>). Το QGIS εκδίδεται για πολλά λειτουργικά συστήματα (Linux, Unix, MacOSX, Windows, Android), υποστηρίζοντας μια μεγάλη σειρά μορφότυπων (format) και λειτουργιών (functionalities) για τη διαχείριση, ενημέρωση και επεξεργασία χωρικών δεδομένων. Εκτός από τις θεμελιώδεις λειτουργίες του QGIS, υπάρχει ένα αποθετήριο προσθέτων (plugin repository) ανεπτυγμένο από την κοινότητα των χρηστών του με στόχο την συνεχή επέκταση των δυνατοτήτων του για όλους (<https://plugins.qgis.org/plugins/>).

3.2.2 Εγκατάσταση και εκκίνηση του προγράμματος

Η εγκατάσταση του QGIS είναι πολύ απλή και υλοποιείται με το πρόγραμμα εγκατάστασης που διατίθεται, για κάθε λειτουργικό σύστημα, από την ιστοσελίδα <https://www.qgis.org/en/site/forusers/download.html>. Η πιο πρόσφατη έκδοση του προγράμματος είναι QGIS 3.20.3 released on 18.05.2020. Αφού ολοκληρωθεί η εγκατάσταση, η εκκίνηση του προγράμματος QGIS πραγματοποιείται από το εικονίδιο.



3.2.3 Το γραφικό περιβάλλον διεπαφής του QGIS



Σχήμα 3.2: Το γραφικό περιβάλλον διεπαφής (graphical user interface) του QGIS (Πηγή: <http://www.qgis.org>)

Η αρχική εμφάνιση του γραφικού περιβάλλοντος ενδέχεται να είναι διαφορετική για κάθε χρήστη ανάλογα με τη γλώσσα και τις εργαλειοθήκες που εμφανίζονται. Γενικά, στο γραφικό περιβάλλον διακρίνονται οι ακόλουθες πέντε περιοχές: Μενού (Menu Bar), Εργαλειοθήκη (Tool Bar), Θεματικά Επίπεδα (Layers), Περιοχή χάρτη (Map Area) και Γραμμή Κατάστασης (Status Bar) (Σχήμα 3.2).

Μενού (Menu Bar): Το μενού αυτό δίνει πρόσβαση σε διαφορετικά εργαλεία του QGIS ακολουθώντας μια ιεραρχική δομή (Σχήμα 3.2). Πλήρης αναφορά των επιλογών

του μενού βρίσκεται στην ιστοσελίδα του QGIS (http://docs.qgis.org/2.6/en/docs/user_manual/introduction/qgis_gui.html).

Συνοπτικά τα εργαλεία του Menu Bar είναι:

- *Έργο (Project)* - για το άνοιγμα, την αποθήκευση, τον καθορισμό των ιδιοτήτων και την εκτύπωση ενός έργου.
- *Επεξεργασία (Edit)* - για την επεξεργασία των δεδομένων ενός θεματικού επιπέδου π.χ. προσθήκη, διόρθωση, διαγραφή κ.ά.
- *Εμφάνιση (View)* - για την πλοήγηση εντός του χάρτη, την αλλαγή του βαθμού μεγέθυνσης, του ορισμού του εύρους της περιοχής σχεδίασης κ.ά.
- *Επίπεδο (Layer)* - για την εισαγωγή και διαχείριση των θεματικών επιπέδων.
- *Ρυθμίσεις (Settings)* - για τη διαχείριση των ρυθμίσεων του περιβάλλοντος.
- *Πρόσθετα (Plugins)* - για τη διαχείριση των υπαρχόντων πρόσθετων εργαλείων και την προσθήκη νέων.
- *Διανύσματα (Vector), Ψηφιδωτό (Raster), Βάση δεδομένων (Database), και Processing* - για την διαχείριση των δεδομένων.
- *Διαδίκτυο (Web)* για την εισαγωγή διαδικτυακών δεδομένων με plugins για διασύνδεση με metadata, open layers, quick map services.
- *MMQGIS* για την διαχείριση των διανυσματικών επιπέδων μέσω διαφόρων Python plugins: CSV input/output/join, geocoding, geometry conversion.

3.2.4 Σημαντικά χαρακτηριστικά γνωρίσματα

Το QGIS έχει πολλά κοινά χαρακτηριστικά γνωρίσματα και λειτουργίες με τα υπόλοιπα GIS, τα σημαντικότερα αναφέρονται παρακάτω στον Πίνακα 4-1:

Πίνακας 3-1: Χαρακτηριστικά γνωρίσματα και λειτουργίες του QGIS

Κοινά χαρακτηριστικά γνωρίσματα και λειτουργίες του QGIS με τα υπόλοιπα GIS	
Ψηφιοποίηση Δεδομένων	Επιλογή χαρακτηριστικών γνωρισμάτων Τοποθέτηση ετικέτας στα γνωρίσματα
Επισκόπηση και Σύνθεση χαρτών	Διαρκείς επιλογές (persistent selections)
Πρόσθετα - Plugins	Αποθήκευση και αποκατάσταση των εργασιών (save and restore)
Υποστήριξη ενσωμάτωσης του GRASS	Αλλαγή raster συμβόλων (grayscale, pseudocolor, multiband RGB)
Αναγνώριση γνωρισμάτων (identify features)	On the fly προβολή των vector layers
Προβολή πίνακα χαρακτηριστικών (attribute table)	Εξαγωγή σε μορφή αρχείου Map server

Υποστήριξη των shapfiles της ESRI και άλλων vector μορφότυπων από τη βιβλιοθήκη OGR, συμπεριλαμβανομένου και του MapInfo. SVG σύμβολα(single, graduated, unique value)	Αλλαγή συμβολισμού των vector γεωμετρικών συμβόλων (single, graduated, unique value, continuous) ανάλογα με την Θεματική τους πληροφορία
Υποστήριξη των πινάκων της χωρικά επεκτάσιμης βάσης PostgreSQL με τη χρήση PostGIS	Υποστήριξη μορφότυπων raster που υποστηρίζει η βιβλιοθήκη GDAL. Απεικόνιση raster δεδομένων, όπως αεροφωτογραφίες

Οι νέες εκδόσεις (μετέπειτα της 1.8) του QGIS περιλαμβάνουν επιπλέον χαρακτηριστικά γνωρίσματα όπως: Βελτιωμένη επεξεργασία διανυσματικών αρχείων καθώς και των ιδιοτήτων τους, Βελτιωμένα εργαλεία μέτρησης, Βελτιωμένο εργαλείο εξαγωγής MapServer, Νέα δομή για την εισαγωγή και την επεξεργασία λεζάντας, Υποστήριξη του GRASS σε όλες τις πλατφόρμες, Ενισχυμένες εντολές υποστήριξης των εργαλειοθηκών του GRASS, Επεξεργασία Shapefile/OGRIaye, και Υποστήριξη WMS, WFS.

Ωστόσο, ένα από τα πλεονεκτήματα του QGIS είναι τα ένθετα (plugin) που προσφέρει στους χρήστες του, τα οποία και σχεδόν καθημερινά αυξάνονται. Με τον όρο ένθετο στους ηλεκτρονικούς υπολογιστές, ονομάζεται ένα στοιχείο λογισμικού που προσθέτει ένα ιδιαίτερο χαρακτηριστικό σε ένα υπάρχον πρόγραμμα υπολογιστή.

OpenLayers Plugin : επιτρέπει στον χρήστη να προσθέτει μια σειρά από υπηρεσίες βασικού χάρτη σε ψηφιακούς χάρτες, όπως: Google, OpenStreetMap, Bing, MapQuest

QGIS Cloud Plugin : είναι μια προσωπική υποδομή χωρικών δεδομένων (SDI) στο διαδίκτυο όταν θέλει ο χρήστης να δημοσιεύσει τα έργα / προϊόντα του QGIS ή να μοιραστεί τα χωρικά δεδομένα με άλλους.

To QGIS Cloud : Ο χρήστης μπορεί να δημοσιεύει όλους τους χάρτες και άλλες πληροφορίες, μπορεί να χρησιμοποιεί το OGC, με το Web GIS και μια βάση δεδομένων PostGIS2.

Qgis2web : μπορεί να δημιουργήσει ένα χάρτη διαδικτύου (web) από την τρέχουσα εφαρμογή του χρήστη, είτε ως OpenLayers3 είτε ως Leaflet.

Qgis2threejs : ο χρήστης μπορεί να στείλει πληροφορίες 3D σε οποιοδήποτε web browser που υποστηρίζει GL Web. Είναι ένα χρήσιμο εργαλείο για τους κατασκευαστές ιστοσελίδων που τώρα μπορούν να έχουν συνδέσεις σε πραγματικό χρόνο για πληροφορίες ΓΣΠ.

GEarthView : επιτρέπει την σύνδεση του QGIS με το Google Earth.

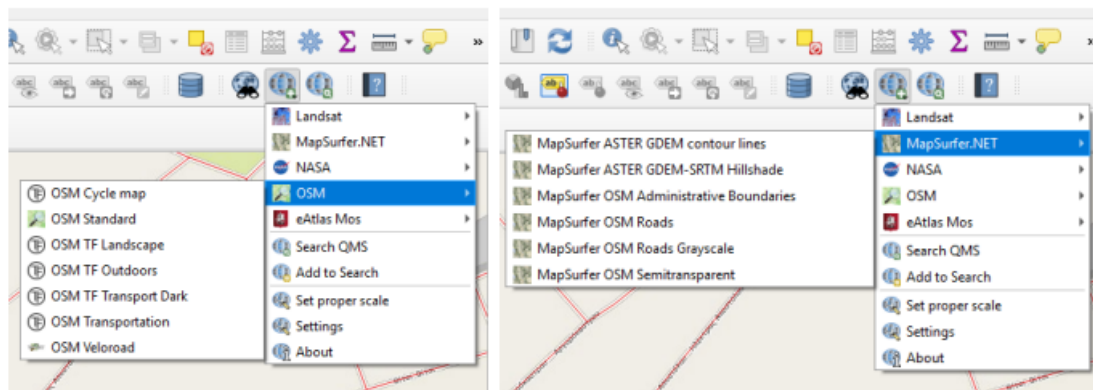
Online Routing Mapper : Μία από τις από πλέον χρησιμοποιούμενες εφαρμογές χαρτών δρομολόγησης που χρησιμοποιεί μια online υπηρεσία από επιλεγμένες του προμηθευτή και καθορίζεται από τον χρήστη το σημείο εκκίνησης και του τέλους.

MetaSearch Catalogue Client : επιτρέπει την αλληλεπίδραση με τις υπηρεσίες καταλόγου μεταδεδομένων, υποστηρίζοντας την Υπηρεσία Καταλόγου OGC για το πρότυπο Web (CSW).

Το χαρτογραφικό υπόβαθρο είναι συχνά το πρώτο επίπεδο που προστίθεται στην αρχή ενός έργου GIS. Οι βάσεις χαρτών μπορούν να αποκτηθούν και να εισαχθούν στο σύστημα μέσω διαφόρων υπηρεσιών ιστού: TMS, WMS, WMTS, ESRI ArcGIS Services ή απλώς ως tiles XYZ.

Με τις τόσες πολλές υπηρεσίες geodata που υπάρχουν σήμερα στο διαδίκτυο ήταν αναμενόμενο να αναπτυχθούν από τις εταιρείες GIS εργαλεία γρήγορης αναζήτησης, χωρίς να χρειάζεται να απομνημονεύουμε διευθύνσεις URL, παραμέτρους κ.λπ. Το κάθε GIS Desktop έχει διαφορετικά εργαλεία για την προσθήκη βασικών χαρτών. Το QGIS, διαθέτει ειδικά μενού και 3rd-party plugins για πρόσβαση σε αυτές τις υπηρεσίες. Τα πρωτόκολλα WMS και WMTS είναι τα βασικά αλλά είναι δυνατή και η πρόσβαση των tiles XYZ μέσω του προγράμματος οδήγησης WMS, γράφοντας ένα ειδικό αρχείο XML που είναι κατανοητό από το GDAL. Αυτό βέβαια δεν είναι καθόλου εύχρηστο, καθώς απαιτεί αφενός γνώση όλων αυτών των διευθύνσεων και αφετέρου χρόνο για να δημιουργηθούν αυτά τα αρχεία.

Τα υπάρχοντα πρόσθετα προσφέρουν κάποια από τις απαραίτητες λειτουργίες, αλλά πρόσφατα αναπτύχθηκε από το QGIS, ένα νέο εύχρηστο plugin το QuickMapServices (QMS), που είναι ένα ειδικό plugin που διευκολύνει την εργασία με webbased basemaps εξασφαλίζοντας την πρόσβαση σε διαθέσιμα, προκαθορισμένα χαρτογραφικά δεδομένα (Σχήμα 3.3).



Σχήμα 3.3: Screenshots από το μενού του QGIS plugin QuickMapServices (QMS), για τα χαρτογραφικά δεδομένα των πηγών: OSM-Open Street Map (αριστερά) και MapSurfer (Δεξιά)

3.2.5 Πλεονεκτήματα του QGIS

Το Quantum GIS ή QGIS (<http://www.qgis.org>) είναι ένα λογισμικό ανοιχτού κώδικα ΓΣΠ που λειτουργεί σε όλα τα λειτουργικά συστήματα. Διατίθεται με την άδεια GNU GeneralPublic License (GPL). Το QGIS διαθέτει φιλικό γραφικό περιβάλλον επικοινωνίας, επιτρέποντας στους χρήστες να αναλύουν και να επεξεργάζονται χωρικές πληροφορίες, εκτός από τη σύνθεση και την εξαγωγή γραφικών χαρτών. Προσφέρει εργαλεία γεωεπεξεργασίας, στατιστικής και χωρικής ανάλυσης, τοπογραφικών εφαρμογών, εργαλεία χαρτοσύνθεσης και δημιουργίας βάσεων δεδομένων.

Το γεγονός πως το QGIS είναι ελεύθερο λογισμικό και παράλληλα απολύτως νόμιμο είναι ένας σημαντικός λόγος για την επιλογή του ως Γεωγραφικό Σύστημα Πληροφοριών. Παρότι δεν χρειάζεται η αγορά κάποιας άδειας για τη χρήση του δεν υστερεί ιδιαίτερα με άλλα εμπορικά μη ελεύθερα λογισμικά, καθώς έχει μεγάλη ταχύτητα και απόδοση προσφέροντας πληθώρα δυνατοτήτων ανάλυσης και επεξεργασίας, μέσω συνεργασιών με το GRASS και άλλα εργαλεία. Το QGIS υποστηρίζει πολλές απαραίτητες μορφές αρχείων, όπως shapfiles, dxf κ.α.. Υποστηρίζονται επίσης υπηρεσίες Web, για να επιτρέπεται η εισαγωγή δεδομένων από εξωτερικές πηγές.

Στο QGIS ενσωματώνονται και προσθήκες (plugins) ανοιχτού κώδικα. Αυτές οι προσθήκες γραμμένες σε Python ή C ++ επεκτείνουν τις δυνατότητες του QGIS. Οι προσθήκες μπορούν να πραγματοποιήσουν γεωκωδικοποίηση χρησιμοποιώντας το API Google Geocoding, να εκτελούν λειτουργίες παρόμοιες με εκείνες των τυποποιημένων εργαλείων που υπάρχουν σε εμπορικά λογισμικά, όπως το ArcGIS, και να διασυνδέονται με βάσεις δεδομένων στο διαδίκτυο. Έτσι, ένας χρήστης με

γνώσεις προγραμματισμού έχει τη δυνατότητα να γράψει και να ενσωματώσει το δικό του script για κάποια ειδική λειτουργία ή αυτοματοποίηση λειτουργιών σώζοντας χρόνο. Με αυτόν τον τρόπο, το QGIS γίνεται ένα «προσωπικό» ΣΓΠ που ανταπεξέρχεται στις ανάγκες του χρήστη, γεγονός το οποίο δεν γίνεται σε εμπορικά λογισμικά. Σε σχέση με άλλα ΓΣΠ απαιτεί λιγότερη μνήμη RAM και επεξεργαστική ισχύ. Ως εκ τούτου, μπορεί να χρησιμοποιηθεί σε παλαιότερα υπολογιστικά συστήματα (hardware) ή να λειτουργεί ταυτόχρονα με άλλες εφαρμογές όπου η ισχύς της CPU ενδέχεται να είναι περιορισμένη.

Ακόμα το QGIS παρουσιάζει πολύ καλή υποστήριξη: Οι οδηγίες χρήσης, τα εγχειρίδια και γενικότερα οι εμπειρίες των χρηστών από την εφαρμογή είναι πάρα πολλές και δίνουν σημαντική βοήθεια στους νέους και όχι μόνο χρήστες. Τέλος η εξέλιξη του QGIS είναι πολύ γρήγορη. Κάθε φορά που υπάρχει μια νέα έκδοση του λογισμικού, υποδεικνύεται με ένα αναδυόμενο παράθυρο που δίνει τη δυνατότητα λήψης, παρουσιάζοντας όλα τα νέα χαρακτηριστικά μαζί με οδηγίες χρήσης.

4 ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ

4.1 Συλλογή Δεδομένων

Πρέπει να γίνει σαφές πως καθοριστικό ρόλο, πέρα από τα εργαλεία που θα χρησιμοποιηθούν κατά την ανάλυση, παίζουν αδιαμφισβήτητα η επιλογή των κατάλληλων δεδομένων από τον μελετητή, ο τρόπος συλλογής τους αλλά και η ποιότητα στις μεθόδους καταγραφής τους, προκειμένου να υπάρξουν αξιόπιστα αποτελέσματα της στατιστικής ανάλυσης των υδρολογικών δεδομένων.

Η προκαταρκτική ανάλυση δεδομένων αποτελεί μία προ-επεξεργασία της βάσης δεδομένων που διατίθεται. Αποτελεί βασικό στοιχείο σε οποιαδήποτε ανάλυση δεδομένων και περιλαμβάνει την εξαγωγή γραφημάτων για την καλύτερη διερεύνηση, κατανόηση και παρουσίαση των δεδομένων. Μία μελέτη με επίκεντρο την στατιστική ανάλυση και υδρολογική ερμηνεία, που δεν περιλαμβάνει μία τέτοια ενδελεχή διερεύνηση των δεδομένων δεν μπορεί να θεωρηθεί πλήρης. Το πρώτο βήμα είναι η εξέταση των ακατέργαστων δεδομένων με σκοπό την αναγνώριση ιδιαίτερων χαρακτηριστικών, όπως προβλήματα των μετρήσεων (αδικαιολόγητα υψηλές τιμές, μεγάλα κενά στις μετρήσεις, κ.α.), εμφανή χρονικά μοτίβα (π.χ. εποχικότητα) και χωρικά μοτίβα. Μία τέτοια πρωταρχική ανάλυση παίζει επίσης σημαντικό ρόλο στην τεκμηρίωση ή όχι βασικών υποθέσεων της κλασικής στατιστικής, όπως η υπόθεση της ανεξαρτησίας των δεδομένων ή η αναγνώριση στατιστικών κατανομών που μπορούν να αναπαραστήσουν τα δεδομένα. Όπως προαναφέρθηκε μία τέτοια προεπεξεργασία μπορεί να βοηθήσει τον εκάστοτε μελετητή να εντοπίσει προβληματικές τιμές του δείγματος, όμως σε καμία περίπτωση δεν υπονοείται ότι μπορεί να αντικαταστήσει τους ελέγχους ποιότητας των δεδομένων. Η πλέον κατάλληλη και αξιόπιστη διερεύνηση των δεδομένων περιλαμβάνει τη δημιουργία και τη μελέτη διαγραμμάτων με σκοπό να αναδείξουν τα βασικά χαρακτηριστικά των δεδομένων, αλλά και να υποδείξουν περαιτέρω διαγράμματα που ενδεχομένως χρειάζονται για την κυρίως ανάλυση. Συνηθισμένοι τύποι διαγραμμάτων που είναι χρήσιμα κυρίως στην στατιστική υδρολογία είναι τα ιστογράμματα, τα διαγράμματα πιθανότητας (P-P plots/Q-Q plots), τα διαγράμματα χρονοσειρών, τα αυτοσυσχετογράμματα κ.α. Περισσότερες λεπτομέρειες και εφαρμογές για την προ-επεξεργασία υδρολογικών δεδομένων δίνονται από τους Grubb and Robson(2000) και Tufte(1983).

Το προκαταρκτικό στάδιο της παρούσας διπλωματικής εργασίας εστίασε στην εύρεση χρονοσειρών παροχών ποταμών σε κατάλληλη χρονική κλίμακα (ημερήσια), από ένα μεγάλο δείγμα σταθμών.

Συγκεκριμένα, αναζητήθηκαν πηγές που διαθέτουν αξιόπιστα δεδομένα μεγάλου μήκους (κατ' ελάχιστο 50 έτη), καθώς και όσο τον δυνατόν περισσότερες πληροφορίες για τις ιδιαιτερότητες των αντίστοιχων σταθμών.

Ύστερα από εκτενή έρευνα, αποφασίστηκε η επιλογή της πηγής δεδομένων Global Historical Climatology Network (GHCN-Daily), η οποία περιέχει καθημερινά δεδομένα για περισσότερους από 50.000 χειρσαίους σταθμούς σε όλο τον κόσμο. Οι σταθμοί αυτοί ανήκουν στην εδαφική επικράτεια των Ηνωμένων Πολιτειών της Αμερικής και της Ευρώπης κυρίως και πληρούν τα ανωτέρω κριτήρια χρονικής διάρκειας και αξιοπιστίας.

4.2 Πηγή Δεδομένων GHCN-Daily

4.2.1 Γενικές Πληροφορίες

Τα δεδομένα που μελετήθηκαν προήλθαν από την επεξεργασία ημερήσιων τιμών παροχής από το Παγκόσμιο Ιστορικό Δίκτυο Κλιματολογίας – Ημερήσια Βάση Δεδομένων (Global Historical Climatology Network, GHCN-Daily, <https://www.ncdc.noaa.gov/ghcnd-data-access>) το οποίο περιέχει δεδομένα από σταθμούς σε όλο τον κόσμο. Συγκεκριμένα, το GHCN-Daily αποτελεί μια ενσωματωμένη βάση δεδομένων που περιέχει ημερήσια κλιματικά αρχεία ελεγμένης ποιότητας. Περιέχει δεδομένα από περίπου 75.000 σταθμούς από 180 χώρες που περιλαμβάνουν γεωφυσικές μετρήσεις όπως μέγιστη και ελάχιστη θερμοκρασία, ολική ημερήσια βροχόπτωση, ημερήσια παροχή, χιονόπτωση και ύψος χιονιού. Περίπου όμως το 2/3 των σταθμών καταγράφει μόνο τη βροχόπτωση. Το μήκος των αρχείων καθώς και η χρονική περίοδος των καταγραφών ποικίλουν από λιγότερο από ένα χρόνο μέχρι περισσότερο από 175 χρόνια.

Το δείγμα δεδομένων υπόκειται σε ελέγχους ποιότητας που λαμβάνουν χώρα περίπου κάθε σαββατοκύριακο κατά τη διάρκεια των οποίων τα δεδομένα ανακατασκευάζονται από τις πηγές δεδομένων που τα συνθέτουν. Η βάση ενημερώνεται έτσι όταν είναι δυνατό από μια ποικιλία νέων πληροφοριών όταν οι τελευταίες είναι διαθέσιμες και έχουν επίσης υποστεί ελέγχους ποιότητας. Κάθε δείγμα συνοδεύεται από ένα αρχείο στο οποίο ορίζεται η ποιότητα των δεδομένων και αξιολογείται η πηγή προέλευσης με χρήση «σημαιών ποιότητας». Αυτή η βάση δεδομένων είναι επίσης η επίσημη βάση δεδομένων που χρησιμοποιεί το U.S. Cooperative Observer data.

Τα δεδομένα έχουν συλλεχθεί μέσω επαφών με αντιπροσώπους των Εθνικών Μετεωρολογικών και Υδρολογικών Κέντρων (National Meteorological and

Hydrological Centers) παγκοσμίως και προέρχονται από τέσσερις βασικές πηγές: (α) την βάση δεδομένων των Η.Π.Α. (U.S. Collection), (β) τη διεθνή βάση δεδομένων εκτός Η.Π.Α. (International Collection), (γ) κρατικά δεδομένα (Government Exchange Data) και (δ) ημερήσια δεδομένα (Global Summary of the Day) που μεταφέρονται μέσω του δικτύου GTS (Global Telecommunications System).

Η πυκνότητα του δικτύου σταθμών είναι γενικά μεγαλύτερη στις Η.Π.Α., στον Καναδά, στην Αυστραλία και στην Ευρώπη και μικρότερη σε περιοχές όπως η Αφρική, η Νότια Αμερική ή η Ανταρκτική. Κάποιοι σταθμοί παρόλα αυτά προέρχονται από τη Βραζιλία, την Ινδία και τη Νότια Αφρική. Η χρονική εξέλιξη του δικτύου των σταθμών είναι ιδιαίτερα έντονη. Χαρακτηριστικό είναι ότι πριν το 1890 ο αριθμός των σταθμών που κατέγραφαν την ημερήσια παροχή αποτελούσε μόνο το 8,9% του μέγιστου αριθμού σταθμών του επόμενου αιώνα.

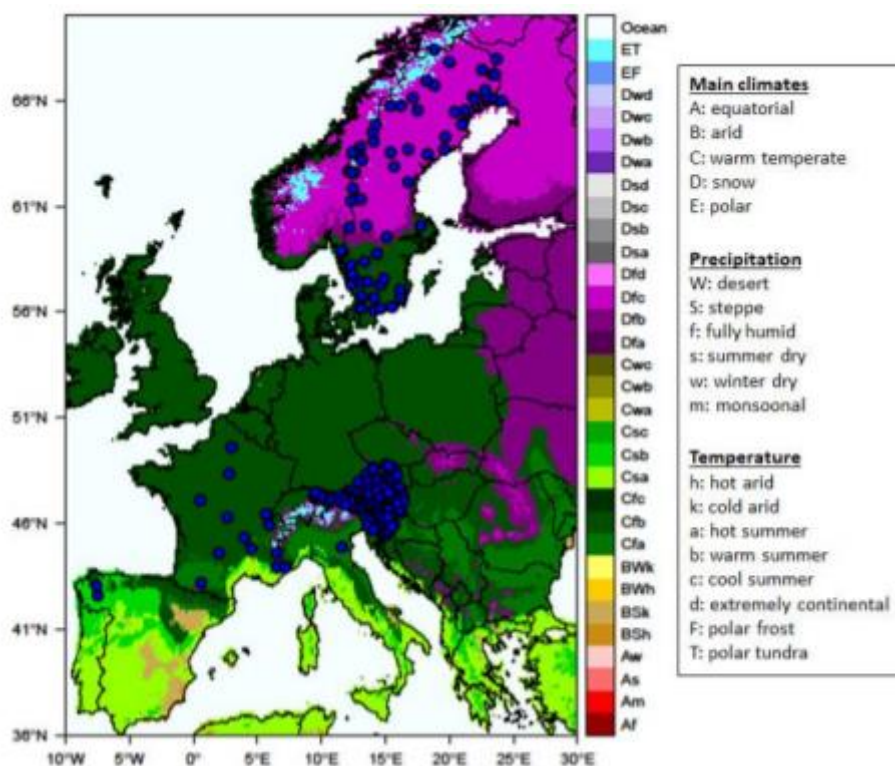
Η διαδικασία της ενσωμάτωσης νέων δεδομένων στο δίκτυο περιλαμβάνει γενικά 3 στάδια: α) το στάδιο της αρχικής διαλογής δεδομένων βάσει προϋποθέσεων (π.χ. ύπαρξη επαρκών γεωγραφικών και ιστορικών πληροφοριών για το σταθμό) β) το στάδιο της ταξινόμησης σταθμού είτε ως τοποθεσία για την οποία ήδη υπάρχουν δεδομένα είτε ως νέα τοποθεσία και γ) το στάδιο του συγκερασμού των σταθμών με χρήση ιεραρχικής ταξινόμησης βάσει κριτηρίων ποιότητας των πηγών προέλευσης των δεδομένων. Σε περιπτώσεις αλληλοεπικάλυψης των δεδομένων χρησιμοποιούνται δεδομένα από την πιο αξιόπιστη πηγή. Επίσης, δεδομένα από μια νέα πηγή για την ίδια τοποθεσία μπορεί να συγκριθούν με τα υπάρχοντα δεδομένα και ανάλογα με το βαθμό ταύτισης και την απόσταση μεταξύ των δύο πηγών, να δημιουργηθεί ή όχι νέα βάση δεδομένων για αυτή την πηγή.

Αξίζει να σημειωθεί ότι το GHCN-Daily δεν περιέχει διορθώσεις για τη μεροληψία που προκύπτει από ιστορικές αλλαγές στα όργανα και τις πρακτικές μέτρησης. Επίσης, οι σταθμοί δεν είναι δυνατόν να ικανοποιήσουν όλα τα απαιτούμενα κριτήρια για την παρατήρηση του κλίματος καθώς οι περισσότεροι έχουν ως στόχο την ικανοποίηση μετρητικών αναγκών για την γεωργία, την υδρολογία, την πρόγνωση καιρού, την αεροπορία κλπ. Περισσότερες πληροφορίες για τη μεθοδολογία της βάσης δεδομένων μπορεί κανείς να βρει στην ιστοσελίδα <https://www.ncdc.noaa.gov/ghcnd-data-access> καθώς και στη βιβλιογραφία (Durre, Menne, and Vose 2008; Durre et al. 2010; Menne et al. 2012).

4.2.2 Δεδομένα

Για τη μελέτη των δεδομένων ημερήσιων παροχών χρησιμοποιήθηκαν χρονοσειρές μετρήσεων από την πηγή δεδομένων Global Historical Climatology Network, GHCN-Daily. Ενα σημαντικό ποσοστό αυτών των δεδομένων παρουσιάζουν τα τυπικά ζητήματα των περισσότερων διαθέσιμων συνόλων δεδομένων, δηλαδή τιμές που λείπουν, μικρό μήκος μήκους καταγραφής και τιμές παροχής αμφιβόλου ποιότητας, όπως πολύ υψηλές τιμές. Για να περιορίσουμε την ποιότητα των δεδομένων σε σημαντικά υψηλό επίπεδο, πρέπει να φιλτράρουμε το σύνολο δεδομένων χρησιμοποιώντας ορισμένα κριτήρια. Ειδικότερα, μελετάμε μόνο τους σταθμούς που πληρούν τις ακόλουθες προϋποθέσεις: α) μήκος εγγραφής πάνω από 50 χρόνια, β) οι τιμές που λείπουν να είναι λιγότερες από 20% και, γ) ύποπτες τιμές με σημαία ποιότητας μικρότερες από 0,1%. Αρχικά, για να κατασκευάσουμε την ετήσια σειρά διαγράφουμε όλες τις καθημερινές τιμές που έχουν εκχωρηθεί σημαίες ποιότητας, υποδεικνύοντας μη ρεαλιστικά μεγάλες τιμές και, στη συνέχεια, εκτιμούμε τη μέση ημερήσια αξία ανά έτος. Συγκεκριμένα, λόγω της ύπαρξης τιμών που λείπουν στα περισσότερα αρχεία, η σύνοψη όλων των ημερήσιων τιμών ενός έτους θα είχε μικρότερες τιμές ετήσιες εκτιμήσεις από τις πραγματικές σε βαθμό που εξαρτάται από τον αριθμό των αγνοουμένων αξίες. Θα ήταν σαφώς πιο ισχυρό να εκτιμηθούν οι ημερήσιες μέσες τιμές ανά έτος. Αυτό εκτελείται μόνο για τα έτη που έχουν λιγότερες από 20 ημερήσιες τιμές που λείπουν, ενώ το τα υπόλοιπα θεωρούνται αγνοούμενα. Στη συνέχεια, στους σταθμούς που οι τιμές που λείπουν ετησίως είναι πάνω από 20% οι τιμές αφαιρούνται. Προφανώς, αυτή η επιλογή διασφαλίζει ένα σύνολο δεδομένων υψηλότερης ποιότητας για την ανάλυσή μας. Το σύνολο δεδομένων λοιπόν περιλαμβάνει 209 εγγραφές που καλύπτουν περισσότερα από 50 χρόνια καθημερινών παρατηρήσεων ροής ποταμών από σταθμούς μέτρησης, κυρίως από μη ρυθμιζόμενα ρεύματα. Μεταξύ των 209 ποταμών, 102 βρίσκονται στην Αυστρία, 69 στη Σουηδία, 31 στη Σλοβενία, 4 στη Γαλλία, δύο στην Ισπανία και ένα στην Ιταλία. Οι περιοχές απορροής ποικίλλουν σημαντικά, η μεγαλύτερη είναι η λεκάνη του ποταμού Ρο στην Ιταλία (70.091 km²) και η μικρότερη είναι η λεκάνη του ποταμού Halabäcken στη Σουηδία (4,7 km²). Η γεωγραφική θέση των σταθμών μέτρησης του ποταμού καθώς και η κλιματολογική ταξινόμησή τους φαίνονται στο Σχήμα 4.1. Τα περισσότερα από τα ποτάμια που εξετάστηκαν ανήκουν είτε σε θερμούς εύκρατους (Γ) είτε σε υδάτινο καθεστώς λόγω της τοπογραφικής τους μεταβλητότητας. Στα χαμηλότερα υψόμετρα, το καθεστώς που κυριαρχεί είναι το Cfb, αλλά καθώς αυξάνεται η εγγύτητα προς τις Άλπεις, κυριαρχεί ένα καθεστώς Dfc και προοδευτικά, στις λεκάνες του υψηλότερου υψομέτρου, το κλίμα γίνεται ένας τύπος πολικής τούνδρας (Et), που χαρακτηρίζεται κυρίως από τις πολύ χαμηλές

θερμοκρασίες. Τα χαρακτηριστικά όλων των κλιματικών καθεστώτων των ποταμών που μελετήθηκαν δίνονται στο υπόμνημα του Σχήματος 4.1.

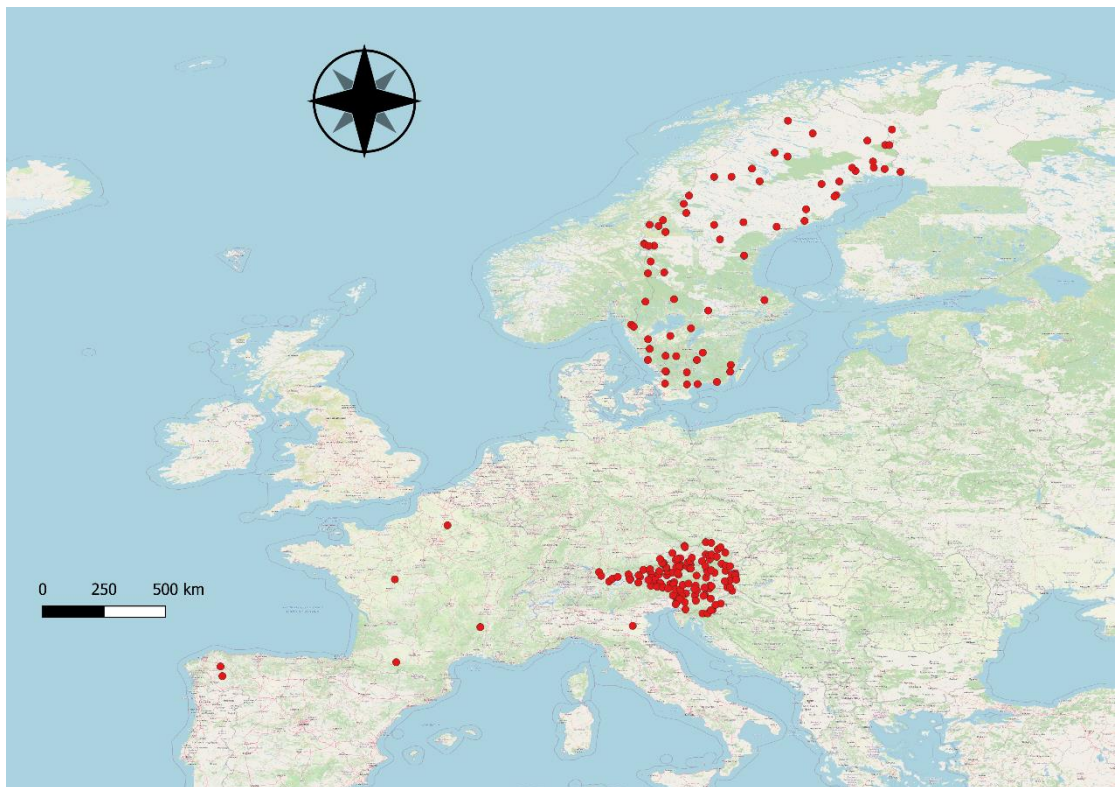


Σχήμα 4.1: Ενημερωμένος κλιματικός χάρτης Köppen – Geiger για την περίοδο 1951–2000 (Kottek et al., 2006) που δείχνει τη θέση των 209 σταθμών μέτρησης ποταμών (Πηγή: Θ. Ηλιοπούλου, Διερεύνηση της ύπαρξης μακροπρόθεσμης εμμονής στην ετήσια βροχόπτωση από μία παγκόσμια βάση δεδομένων, Διπλωματική εργασία, 146 pages, Τομέας Υδατικών Πόρων και Περιβάλλοντος – Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, Μάρτιος 2013.)

Στο Σχήμα 4.2 της επόμενης σελίδας φαίνονται οι απεικονίσεις της γεωγραφικής κατανομής των σταθμών πάνω στη γη με χρήση του προγράμματος Google Earth ενώ στο Σχήμα 4.3 φαίνεται πάλι η γεωγραφική κατανομή των σταθμών σε χάρτη Google Map με φόντο την Ευρώπη με χρήση του προγράμματος QGIS. Η πυκνότητα του δικτύου των σταθμών είναι μεγαλύτερη στη Κεντρική και Βόρεια Ευρώπη και μικρότερη σε περιοχές όπως η Ισπανία, η Ιταλία και η Γαλλία.



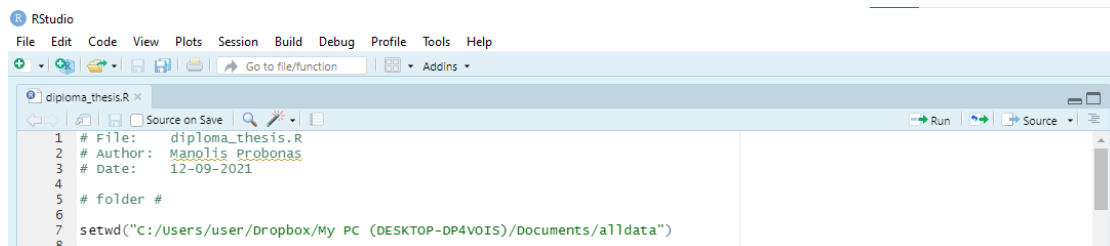
Σχήμα 4.2: Γεωγραφική κατανομή σταθμών (Google Earth)



Σχήμα 4.3: Γεωγραφική κατανομή σταθμών (QGIS)

Πρόκειται για 209 δεδομένα (μετρήσεις) χρονοσειρών που προέκυψαν από επεξεργασία της βάσης GHCN-Daily. Στο περιβάλλον της R πρωταρχικό μέλημα είναι ο καθορισμός του καταλόγου εργασίας (working directory) και η αποστολή (install) και η φόρτωση (load) των πακέτων (packages) μέσω των οποίων καλούμε τις συναρτήσεις (functions) που είναι απαραίτητες για τον υπολογισμό των διάφορων στατιστικών χαρακτηριστικών των χρονοσειρών.

Ο καθορισμός του καταλόγου εργασίας γίνεται με την εντολή `setwd(" ")` όπου ανάμεσα στα αυτάκια βάζουμε το path δηλαδή τη διαδρομή που θα ακολουθήσει για να φτάσει στον κατάλογο εργασίας όπου βρίσκονται τα δεδομένα μας. Χαρακτηριστικά η εντολή φαίνεται στην Εικόνα 5.1:



```
1 # File: diploma_thesis.R
2 # Author: Manolis Probonas
3 # Date: 12-09-2021
4
5 # folder #
6
7 setwd("C:/Users/user/Dropbox/My PC (DESKTOP-DP4V0IS)/Documents/alldata")
8
```

Εικόνα 5.1: Η συνάρτηση `setwd` στην R

Όσο αναφορά τα πακέτα που φορτώσαμε για την ανάλυση τα κυριότερα παρουσιάζονται παρακάτω και είναι: `datasets`, `data.table`, `dplyr`, `FAdist`, `fitdistrplus`, `ggplot2`, `gtools`, `hydroTSM`, `moments`, `lubridate`, `PerformanceAnalytics`, `readr`, `tidyverse`, `tseries`, `writexl`, `xts`. Η συνάρτηση που χρησιμοποιήθηκε για τη φόρτωση όλων των πακέτων είναι η `library()` όπου στο εσωτερικό της παρένθεσης τοποθετείται το όνομα του πακέτου. Στην Εικόνα 5.2 απεικονίζονται τα σημαντικότερα από τα πακέτα που φορτώσαμε στην R.

```

9 # LOAD PACKAGES #
10
11 library(Benchmarking)
12 library(datasets)
13 library(data.table)
14 library(dplyr)
15 library(FAdist)
16 library(fitdistrplus)
17 library(flexsurv)
18 library(forecast)
19 library(ggplot2)
20 library(ggrepel)
21 library(grid)
22 library(gridExtra)
23 library(gtools)
24 library(hydroTSM)
25 library(logspline)
26 library(matrixStats)
27 library(moments)
28 library(lmomco)
29 library(lmom)
30 library(lmomRFA)
31 library(lubridate)
32 library(openxlsx)
33 library(pacman)
34 library(PerformanceAnalytics)
35 library(psych)
36 library(quantmod)
37 library(RColorBrewer)
38 library(readr)

```

Εικόνα 5.2: Φόρτωση πακέτων στην R

Στη συνέχεια φτιάχνουμε μια κενή λίστα (list) με τα αρχεία .txt της βάσης δεδομένων, ένα διάνυσμα (vector) με την ονομασία των σταθμών, ένα διάνυσμα με τις ημερήσιες παροχές, ένα πίνακα (matrix) με τις συντεταγμένες των σταθμών και μια σειρά με την έκταση των λεκανών απορροής που θα μας χρησιμεύσουν στην εισαγωγή όλων των δεδομένων μέσα από ένα μεγάλο for loop. Οι συναρτήσεις δημιουργίας των διανυσμάτων, της λίστας, του πίνακα φαίνονται στην Εικόνα 5.3.

```

55 names_files<-list.files(path="c:/Users/user/Dropbox/My PC (DESKTOP-DP4V0IS)/Documents/alldata",pattern=".txt")
56 names_files=gtools::mixedsort(names_files)
57 xtslist<-vector("list", length=length(names_files))
58 data_length=length(names_files)
59 name1list<-c()
60 dataXY=matrix(NA, nrow = data_length, ncol = 2)
61 areaSize=rep(NA, data_length)
62 colnames(dataXY)=c('X', 'Y')

```

Εικόνα 5.3: Δημιουργία λίστας, διανυσμάτων, πίνακα, σειράς για την εισαγωγή των δεδομένων από το for loop

Τέλος μέσα από μια μεγάλη επανάληψη γίνεται η ανάγνωση όλων των προαναφερθέντων δεδομένων έτσι ώστε μετά το τρέξιμο της επανάληψης να είμαστε σε θέση να καλέσουμε οποιαδήποτε χρονοσειρά με την έκταση της, τις ημερήσιες παροχές κάθε χρονοσειράς όπως και το όνομα κάθε χρονοσειράς. Πρέπει να τονιστεί ότι μέσα στην επανάληψη έγινε μετατροπή των ημερήσιων παροχών (m^3/d) σε απορροή (mm) ώστε τα αποτελέσματα να είναι πιο εποπτικά στη συνέχεια. Στην Εικόνα 5.4 απεικονίζεται σημαντικό μέρος του κώδικα της επανάληψης:

```

64 ~ for(i in 1:data_length){
65
66     x1 <- read_csv(names_files[i], col_names = FALSE)
67
68     # read the coordinates
69     coordTemp=as.character(x1[2,])
70     coordTemp= stringr::str_split(coordTemp, ':')[[1]][2]
71     coordTemp= stringr::str_split(coordTemp, ',')[[1]]
72
73 ~ if (length(coordTemp)==1) {
74     coordTemp= stringr::str_split(coordTemp, ' ')[[1]]
75     coordTemp=na.omit(as.numeric(coordTemp))
76 ^ }
77
78     dataXY[i,]= as.numeric(coordTemp)
79
80     # read the catchment size - [km2]
81     areaTemp=as.character(x1[3,])
82     areaTemp= stringr::str_split(areaTemp, ':')[[1]][2]
83     areaTemp= stringr::str_split(areaTemp, 'km2')[[1]][1]
84     areaSize[i]= as.numeric(areaTemp)
85
86     startdate<-as.character(x1[5,])
87     enddate<-as.character(x1[7,])
88     leapy<-as.character(x1[9,])
89     title<-as.character(x1[1,])
90     namelist<-c(namelist,title)
91     type<-as.character(x1[10,])
92
93     ts<-as.vector(x1[11:nrow(x1),])
94     ts<-as.numeric(ts$x1)
95     ts<-ts*86.4/areaSize[i]
96
97
98     # convert negative values to NAs
99     ts=ifelse(ts<0, NA, ts)
100
101
102     # create empty xts full series to put variables
103     starttime=as.POSIXct(startdate,format="%Y/%m/%d",tz = 'UTC')
104     endtime=as.POSIXct(enddate,format="%Y/%m/%d", tz = 'UTC')
105
106     datesNA=seq(from= starttime, to=endtime, by='1 day')
107
108     xtsNA=xts(x = matrix(NA,ncol=1,nrow=length(datesNA)), order.by = datesNA)
109
110 ~ if(leapy=="T"){
111
112     myDates <- !(format(datesNA,"%m") == "02" & format(datesNA, "%d") == "29")
113     leapdayspos<-which(myDates==FALSE)
114     xtsNA<-xtsNA[-leapdayspos]
115
116     xtsNA[,1]<-ts
117
118 ~ } else {
119

```

Εικόνα 5.4: For loop για την εισαγωγή χρονοσειρών απορροής, έκτασης τους και ονομασίας σταθμών

Το αποτέλεσμα είναι με τις εντολές `xtslist[[i]]`, `areaSize[[i]]` και `namelist[[i]]` όπου `i` είναι η κάθε χρονοσειρά, να παίρνουμε μια λίστα με τις ημερήσιες παροχές σε αποροή, την έκταση της λεκάνης της χρονοσειράς καθώς και το όνομα του σταθμού της χρονοσειράς. Τα αποτελέσματα των εντολών όπως τυπώνονται στην κονσόλα της R εμφανίζονται παρακάτω για `i=1`:

```
Console Terminal x Jobs x
R 4.1.1 · C:/Users/user/Dropbox/My PC (DESKTOP-DP4VOIS)/Documents/alldata/ ↗
1920-01-04 1.6863107
1920-01-05 1.9648971
1920-01-06 2.3531923
1920-01-07 3.2382588
1920-01-08 4.0419683
1920-01-09 4.7026865
1920-01-10 4.8444451
1920-01-11 4.2774108
1920-01-12 3.6462770
1920-01-13 3.0730793
1920-01-14 2.6662938
1920-01-15 2.4283860
1920-01-16 2.3470289
1920-01-17 2.3605884
1920-01-18 2.2940235
1920-01-19 2.1350073
1920-01-20 1.9846200
1920-01-21 1.8749112
1920-01-22 1.8182078
1920-01-23 1.7491775
1920-01-24 1.6924741
1920-01-25 1.6172804
1920-01-26 1.5618096
1920-01-27 1.4952448
1920-01-28 1.4644277
1920-01-29 1.4225164
1920-01-30 1.3978628
1920-01-31 1.4348433
```

Εικόνα 5.5: Ημερήσιες τιμές παροχής σε μονάδες απορροής (mm) για τη χρονοσειρά $i=1$

```
Console Terminal x Jobs x
R 4.1.1 · C:/Users/user/Dropbox/My PC (DESKTOP-DP4VOIS)/Documents/alldata/ ↗
> i=1
> areaSize[[i]]
[1] 70091
> namelist[[i]]
[1] "Po River at Pontelagoscuro (Italy)"
> |
```

Εικόνα 5.6: Έκταση λεκάνης απορροής και ονομασία σταθμού χρονοσειράς για $i=1$

5.2 Υπολογισμός στατιστικών χαρακτηριστικών χρονοσειρών

Αρχικά επειδή θέλουμε υπολογισμό στατιστικών χαρακτηριστικών ανά μήνα και τα δεδομένα μας είναι ημερήσια πρέπει με μια συνάρτηση να μπορούμε να καλέσουμε σε κάθε χρονοσειρά οποιοδήποτε μήνα. Έτσι δημιουργήσαμε τη συνάρτηση `sub_month(x, Month)` στην οποία δίνουμε την χρονοσειρά `x` και το μήνα `Month` και επιστρέφει για παράδειγμα τον 1^ο μήνα της 1^{ης} χρονοσειράς. Η συνάρτηση φαίνεται στην Εικόνα 5.7:

```
138 # Function to select a subset of daily data - subsetting by month
139
140 sub_month=function(x,Month){
141   Dates=index(x)
142   # D=format(D,"%m")=="Month" # Month is a variable specified by the user.
143   Dates=format(Dates,"%m")
144   Dates=as.numeric(Dates)
145   # select_month=which(D==TRUE)
146   select_month=which(Dates==Month)
147   tsd=x[
148     select_month]
149   return(tsd)
150 }
```

Εικόνα 5.7: Συνάρτηση διαχωρισμού των ημερήσιων δεδομένων ανά μήνα

Έπειτα μπορούμε να προχωρήσουμε στον υπολογισμό των στατιστικών χαρακτηριστικών των χρονοσειρών ανά μήνα. Τα στατιστικά μεγέθη που υπολογίστηκαν είναι η μέση τιμή, η τυπική απόκλιση, η διασπορά, ο συντελεστής μεταβλητότητας, η Τρίτη κεντρική ροπή, η ασυμμετρία, η κύρτωση καθώς και τα αντίστοιχα μεγέθη στην κλίμακα L (L ροπές) δηλαδή οι L1, L2, L3, L4, LCV, TAU3 και TAU4. Ακόμη υπολογίστηκαν τα quantiles, τα οποία είναι σημεία κοπής που διαιρούν το εύρος μιας κατανομής πιθανότητας σε συνεχή διαστήματα με ίσες πιθανότητες, και συγκεκριμένα για πιθανότητα 5%, 20%, 50%, 80%, 90%, όπως και η αυτοσυσχέτιση για υστέρηση (lag) ίση με 1, 2 και 3. Ομοίως με παραπάνω στόχος αποτελεί με μια εντολή να λαμβάνουμε τα στατιστικά χαρακτηριστικά μιας χρονοσειράς συγκεκριμένου μήνα. Για αυτό το λόγο δημιουργήσαμε πάλι πριν το for loop κενές λίστες τόσο των δεδομένων ανά μήνα όσο και των στατιστικών χαρακτηριστικών. Τέλος σαν μέγιστη υστέρηση ορίστηκε ο αριθμός 3. Σημαντικό μέρος του κώδικα απεικονίζεται στην Εικόνα 5.8:

```

153 # Calculating statistical characteristics by month of all values
154 statslist=vector("list",data_length)
155 statslistNonZero=vector("list",data_length)
156 flow_data_per_month_list=vector("list",data_length)
157 maxlag=3
158 for(i in 1:data_length){
159   flow_data=statslist[[i]]
160   currmat=c()
161   currmatNonZero=c()
162   for(j in 1:12){
163     flow_data_per_month=sub_month(x=flow_data, Month=j)
164     df=as.data.frame(flow_data_per_month)
165     df=na.omit(df)
166     df_flow_data=df$v1
167     Pdr<-round(mean(df_flow_data<=0.01,na.rm=T),5)
168     lmom=lmom.ub(df_flow_data)
169     LMean=lmom$L1
170     LScale=lmom$L2
171     L3=lmom$L3
172     L4=lmom$L4
173     LVariation=lmom$LCV
174     LSkewness=lmom$STAU3
175     Lkurtosis=lmom$STAU4
176     Min=round(min(df_flow_data,na.rm=TRUE),4)
177     Max=round(max(df_flow_data,na.rm=TRUE),4)
178     Mean=round(mean(df_flow_data,na.rm=TRUE),4)
179     Var=round(var(df_flow_data,na.rm=TRUE),4)
180     StDev=round(sd(df_flow_data,na.rm=TRUE),4)
181     Variation=round(StDev/Mean,4)
182     Mom3=round(sum((df_flow_data-Mean)^3,na.rm=T)*(1/(sum(!is.na(df_flow_data))-1)),5)
183     Skewness= round(Mom3/StDev^3,4)
184     Kurtosis=round(moments::kurtosis(df_flow_data,na.rm=TRUE))
185     Month=j
186     Q5<-round(quantile(df_flow_data,probs=c(0.05),na.rm=TRUE,names=FALSE),5)
187     Q20<-round(quantile(df_flow_data,probs=c(0.20),na.rm=TRUE,names=FALSE),5)
188     Q50<-round(quantile(df_flow_data,probs=c(0.50),na.rm=TRUE,names=FALSE),5)
189     Q80<-round(quantile(df_flow_data,probs=c(0.80),na.rm=TRUE,names=FALSE),5)
190     Q90<-round(quantile(df_flow_data,probs=c(0.90),na.rm=TRUE,names=FALSE),5)
191     Autocorrelation<-acf(df_flow_data,lag.max=maxlag,type="correlation",plot=FALSE,na.action = na.pass)
192     Autocorrelation<-as.vector(Autocorrelation[[1]])[2:(maxlag+1)]
193     Autocovariance<-acf(df_flow_data,lag.max=maxlag,type="covariance",plot=FALSE,na.action = na.pass)
194     Autocovariance<-as.vector(Autocovariance[[1]])[2:(maxlag+1)]
195     Description_Analysis=c(LMean,LScale,L3,L4,LVariation,LSkewness,
196                           Lkurtosis,Min,Max,Mean,Var,StDev,variation,Mom3,
197                           skewness,kurtosis,Q5,Q20,Q50,Q80,Q90,Pdr,Autocorrelation,
198                           Autocovariance)
199     currmat=cbind(currmat,Description_Analysis)
200     rownames(currmat)=c("LMean","LScale","L3","L4","LVariation","LSkewness",
201                       "Lkurtosis","Min","Max","Mean","Var","StDev","Variation","Mom3",
202                       "skewness","kurtosis","Q5","Q20","Q50","Q80","Q90","Pdr","Autocorrelation1",
203                       "Autocorrelation2","Autocorrelation3","Autocovariance1","Autocovariance2",
204                       "Autocovariance3")
205     flow_data_per_month_list[[i]][[j]]=flow_data_per_month

```

Εικόνα 5.8: For loop για τον υπολογισμό των στατιστικών χαρακτηριστικών ανά μήνα

Το αποτέλεσμα είναι με τις εντολές `statslist[[i]]`, `flow_data_per_month_list[[i]][[j]]` όπου *i* είναι η κάθε χρονοσειρά και *j* ο κάθε μήνας, να παίρνουμε μια λίστα με τα στατιστικά μεγέθη κάθε χρονοσειράς για τους 12 μήνες και τα δεδομένα παροχής ανά μήνα κάθε χρονοσειράς σε μονάδες απορροής (mm) αντίστοιχα. Τα αποτελέσματα των εντολών όπως τυπώνονται στην κονσόλα της R εμφανίζονται παρακάτω για *i*=1 και *j*=1:

```

Console Terminal x Jobs x
R 4.1.1 · C:/Users/user/Dropbox/My PC (DESKTOP-DP4VOIS)/Documents/alldata/ ↗
> i=1
> statslist[[i]]
Description_Analysis Description_Analysis Description_Analysis
LMean 1.53113130 1.58468864 1.87034805
LScale 0.38273073 0.39775010 0.50218069
L3 0.13739548 0.13538534 0.16329449
L4 0.08773241 0.09281083 0.08271987
LVariation 0.24996598 0.25099574 0.26849585
LSkewness 0.35898732 0.34037789 0.32517079
Lkurtosis 0.22922751 0.23333954 0.16472133
Min 0.48570000 0.51400000 0.66440000
Max 6.66020000 7.16190000 7.32210000
Mean 1.53110000 1.58470000 1.87030000
Var 0.62490000 0.68730000 0.97990000
StDev 0.79050000 0.82900000 0.98990000
Variation 0.51630000 0.52310000 0.52930000
Mom3 1.07018000 1.35479000 1.70928000
Skewness -2.16650000 -2.37800000 -1.76210000
Kurtosis 9.00000000 11.00000000 7.00000000
Q5 0.80494000 0.79138000 0.89302000
Q20 0.97382000 1.01179000 1.12914000
Q50 1.29802000 1.34362000 1.51620000
Q80 1.93038000 1.97229000 2.51467000
Q90 2.54216000 2.52700000 3.16800000
Pdr 0.00000000 0.00000000 0.00000000
Autocorrelation1 0.92821395 0.92783883 0.93292433
Autocorrelation2 0.82400059 0.81046856 0.83350517
Autocorrelation3 0.72530028 0.69199110 0.74406384
Autocovariance1 0.57983798 0.63744473 0.91388864
Autocovariance2 0.51473783 0.55680889 0.81649807
Autocovariance3 0.45308159 0.47541239 0.72888173

```

Εικόνα 5.9: Αποτελέσματα στατιστικών χαρακτηριστικών για τους 3 πρώτους μήνες της χρονοσειράς $i=1$

```

Console Terminal x Jobs x
R 4.1.1 · C:/Users/user/Dropbox/My PC (DESKTOP-DP4VOIS)/Documents/alldata/ ↗
1920-01-04 1.6863107
1920-01-05 1.9648971
1920-01-06 2.3531923
1920-01-07 3.2382588
1920-01-08 4.0419683
1920-01-09 4.7026865
1920-01-10 4.8444451
1920-01-11 4.2774108
1920-01-12 3.6462770
1920-01-13 3.0730793
1920-01-14 2.6662938
1920-01-15 2.4283860
1920-01-16 2.3470289
1920-01-17 2.3605884
1920-01-18 2.2940235
1920-01-19 2.1350073
1920-01-20 1.9846200
1920-01-21 1.8749112
1920-01-22 1.8182078
1920-01-23 1.7491775
1920-01-24 1.6924741
1920-01-25 1.6172804

```

Εικόνα 5.10: Αποτελέσματα δεδομένων παροχής για τον 1^ο μήνα της χρονοσειράς $i=1$

Επιπλέον δεν πρέπει να παραληφθεί ο υπολογισμός της πιθανότητας κάποια έτη χρονοσειρών να είναι ξηρά δηλαδή να έχουν μηδενικές τιμές κάποιοι μήνες (probability dry). Σε περίπτωση που σε κάποιες χρονοσειρές η πιθανότητα ξηρασίας είναι μη μηδενική θα πρέπει να υπολογιστούν τα στατιστικά χαρακτηριστικά ανά μήνα και για τις μη μηδενικές τιμές. Η επανάληψη μέσω της οποίας υπολογίζεται σε ποια χρονοσειρά και σε ποιους μήνες εμφανίζονται ξηρά έτη φαίνεται στην Εικόνα 5.11:

```

291 #in which stations and months appears Pdr
292 df_Pdr_0=vector("list",data_length)
293 for(i in 1:data_length){
294     df_Pdr_0[[i]]=which(df_statslist[[i]]["Pdr",]!=0)
295
296 }
297 i=187
298 df_Pdr_0[[i]]
299 df_statslist[[i]]["Pdr",]
300

```

Εικόνα 5.11: Χρονοσειρές και μήνες με πιθανότητα ξηρασίας μη μηδενική

Οι χρονοσειρές στις οποίες η πιθανότητα ξηρασίας είναι μη μηδενική είναι οι ακόλουθες: 6, 47, 61,89, 149, 150, 153, 170, 187, 188, 189, 195, 200, 205, 210. Επομένως είναι απαραίτητος ο υπολογισμός των στατιστικών χαρακτηριστικών και γι' αυτές τις χρονοσειρές με τις μη μηδενικές τιμές πιθανότητας ξηρασίας. Ο κώδικας στην ουσία είναι όμοιος με τον παραπάνω απλά από τα δεδομένα των τιμών ημερήσιων παροχών θέλουμε μόνο τις θετικές τιμές δηλαδή `df_flow_data[df_flow_data>0]`.

Τέλος κρίνεται αναγκαία η δημιουργία data.frame με τα στατιστικά χαρακτηριστικά όλων των μηνών κάθε χρονοσειράς ώστε να μπορεί να γίνει εξαγωγή στη συνέχεια σε .txt αρχείο ή σε αρχείο .xlsx. Η δημιουργία data.frame γίνεται μέσα από την παρακάτω επανάληψη όπου σε κάθε στήλη είναι τα στατιστικά χαρακτηριστικά κάθε μήνα για τη συγκεκριμένη χρονοσειρά:

```

267 # Make data frame with statistical characteristics by month of all values
268 df_statslist=vector("list",data_length)
269 for(i in 1:data_length){
270     df_statslist[[i]]=data.frame(statslist[[i]])
271     colnames(df_statslist[[i]])=c("Month1","Month2","Month3","Month4","Month5","Month6",
272                                 "Month7","Month8","Month9","Month10",
273                                 "Month11","Month12")
274 }
275 i=1 #time series data
276 df_statslist[[i]]["LMean","Month1"]
277

```

Εικόνα 5.12: Δημιουργία data frame με τα στατιστικά μεγέθη ανά μήνα για κάθε χρονοσειρά

Ακόμη στο προγραμματιστικό περιβάλλον της R δίνεται η δυνατότητα σε περίπτωση μεγάλου κώδικα να μην «τρέχουμε» τον κώδικα από την αρχή κάθε φορά αλλά να αποθηκεύουμε στο περιβάλλον εργασίας τα αποτελέσματα και να τα φορτώνουμε αυτόματα κάθε φορά. Αυτό εκτελείται με τη συνάρτηση `load(" ")` όπου ανάμεσα στα αυτάκια βάζουμε το `path` δηλαδή τη διαδρομή που θα ακολουθήσει για να φτάσει στο φάκελο που έχουμε αποθηκεύσει τα αποτελέσματα του κώδικα. Η συνάρτηση απεικονίζεται στην Εικόνα 5.13:

```
54 # LOAD RESULTS CODE #
55 load("C:/Users/user/Dropbox/My PC (DESKTOP-DP4VOIS)/Documents/alldata/teliko.RData")
56
```

Εικόνα 5.13: Φόρτωση αποτελεσμάτων του κώδικα που ήδη έχει «τρέξει»

Ωστόσο μας ενδιαφέρει ο υπολογισμός των στατιστικών χαρακτηριστικών και για ολόκληρη τη χρονοσειρά και όχι μόνο ανά μήνα, δηλαδή ετήσια. Παρομοίως με τον υπολογισμό των στατιστικών χαρακτηριστικών των χρονοσειρών ανά μήνα παραπάνω, δημιουργήσαμε έναν κώδικα για στατιστική ανάλυση σε ετήσια βάση δηλαδή για ολόκληρη τη χρονοσειρά, τον οποίο βλέπουμε παρακάτω:

```
3 # Calculating statistical characteristics of timeseries #
4 statslist_ts=vector("list",data_length)
5 flow_data_per_ts_list=vector("list",data_length)
6 maxlag=3
7 for(i in 1:data_length){
8   flow_data_ts=xtslist[[i]]
9   currmat_ts=c()
10  df_ts=as.data.frame(flow_data_ts)
11  df_ts=na.omit(df_ts)
12  df_flow_data_ts=df_ts$V1
13  Pdr_ts<-round(mean(df_flow_data_ts<=0.01,na.rm=T),5)
14  lmom_ts=lmom.ub(df_flow_data_ts)
15  lMean_ts=lmom_ts$L1
16  lScale_ts=lmom_ts$L2
17  l3_ts=lmom_ts$L3
18  l4_ts=lmom_ts$L4
19  lVariation_ts=lmom_ts$LCV
20  lSkewness_ts=lmom_ts$TAU3
21  lKurtosis_ts=lmom_ts$TAU4
22  Min_ts=round(min(df_flow_data_ts,na.rm=TRUE),4)
23  Max_ts=round(max(df_flow_data_ts,na.rm=TRUE),4)
24  Mean_ts=round(mean(df_flow_data_ts,na.rm=TRUE),4)
25  var_ts=round(var(df_flow_data_ts,na.rm=TRUE),4)
26  StDev_ts=round(sd(df_flow_data_ts,na.rm=TRUE),4)
27  Variation_ts=round(StDev_ts/Mean_ts,4)
28  Mom3_ts=round(sum((df_flow_data_ts-Mean_ts)^3,na.rm=T)*(1/(sum(!is.na(df_flow_data_ts))-1)),5)
29  skewness_ts= round(Mom3_ts/StDev_ts^3,4)
30  Kurtosis_ts=round(moments:kurtosis(df_flow_data_ts,na.rm=TRUE))
31  Q5_ts<-round(quantile(df_flow_data_ts,probs=c(0.05),na.rm=TRUE,names=FALSE),5)
32  Q20_ts<-round(quantile(df_flow_data_ts,probs=c(0.20),na.rm=TRUE,names=FALSE),5)
33  Q50_ts<-round(quantile(df_flow_data_ts,probs=c(0.50),na.rm=TRUE,names=FALSE),5)
34  Q80_ts<-round(quantile(df_flow_data_ts,probs=c(0.80),na.rm=TRUE,names=FALSE),5)
35  Q90_ts<-round(quantile(df_flow_data_ts,probs=c(0.90),na.rm=TRUE,names=FALSE),5)
36  Autocorrelation_ts<-acf(df_flow_data_ts,lag.max=maxlag,type="correlation",plot=FALSE,na.action = na.pass)
37  Autocorrelation_ts<-as.vector(Autocorrelation_ts[[1]])[2:(maxlag+1)]
38  Autocovariance_ts<-acf(df_flow_data_ts,lag.max=maxlag,type="covariance",plot=FALSE,na.action = na.pass)
39  Autocovariance_ts<-as.vector(Autocovariance_ts[[1]])[2:(maxlag+1)]
40  Description_Analysis_ts=c(lMean_ts,lScale_ts,l3_ts,l4_ts,lVariation_ts,lSkewness_ts,
41                            lKurtosis_ts,Min_ts,Max_ts,Mean_ts,Var_ts,StDev_ts,Variation_ts,Mom3_ts,
42                            Skewness_ts,Kurtosis_ts,Q5_ts,Q20_ts,Q50_ts,Q80_ts,Q90_ts,Pdr_ts,Autocorrelation_ts,
43                            Autocovariance_ts)
44  currmat_ts=cbind(currmat_ts,Description_Analysis_ts)
45  rownames(currmat_ts)=c("lMean","lScale","l3","l4","lVariation","lSkewness",
46                        "lKurtosis","Min","Max","Mean","Var","StDev","Variation","Mom3",
47                        "Skewness","Kurtosis","Q5","Q20","Q50","Q80","Q90","Pdr","Autocorrelation1",
48                        "Autocorrelation2","Autocorrelation3","Autocovariance1","Autocovariance2",
49                        "Autocovariance3")
50  flow_data_per_ts_list[[i]]=flow_data_ts
51  statslist_ts[[i]]=currmat_ts
52 }
53
```

Εικόνα 5.14: For loop για τον υπολογισμό των στατιστικών χαρακτηριστικών ετήσια

Προκειμένου να μπορέσουμε να εξάγουμε τα αποτελέσματα σε μορφή αρχείου .txt ή σε αρχείο .xlsx είναι απαραίτητο τα αποτελέσματα των στατιστικών μεγεθών να τα κάνουμε data frame. Στόχος λοιπόν είναι να δημιουργήσουμε 10 στήλες όσα και τα στατιστικά μεγέθη που υπολογίζουμε και σε κάθε στήλη να έχουμε τα αποτελέσματα του αντίστοιχου στατιστικού μεγέθους για όλες τις χρονοσειρές. Την ίδια διαδικασία εφαρμόζουμε και για τα στατιστικά αποτελέσματα των χρονοσειρών ανα μήνα όπου ως στήλες έχουμε τα αποτελέσματα του αντίστοιχου στατιστικού μεγέθους για όλους τους μήνες και για όλες τις χρονοσειρές. Αυτό το πετυχαίνουμε με τον παρακάτω κώδικα:

```
58 #Mean for all time series data
59 df_Mean_ts=c()
60 for(i in 1:data_length){
61   df_Mean_ts=rbind(df_Mean_ts,statslist_ts[[i]]["Mean",])
62 }
63 df_Mean_ts=as.data.frame(df_Mean_ts)
64
65
66 #Mean for all time series data
67 df_StDev_ts=c()
68 for(i in 1:data_length){
69   df_StDev_ts=rbind(df_StDev_ts,statslist_ts[[i]]["StDev",])
70 }
71 df_StDev_ts=as.data.frame(df_StDev_ts)
72
73
74 #skewness for all time series data
75 df_skewness_ts=c()
76 for(i in 1:data_length){
77   df_skewness_ts=rbind(df_skewness_ts,statslist_ts[[i]]["skewness",])
78 }
79 df_skewness_ts=as.data.frame(df_skewness_ts)
80
81
82 #Q5 for all time series data
83 df_Q5_ts=c()
84 for(i in 1:data_length){
85   df_Q5_ts=rbind(df_Q5_ts,statslist_ts[[i]]["Q5",])
86 }
87 df_Q5_ts=as.data.frame(df_Q5_ts)
88
89
90 #Q20 for all time series data
91 df_Q20_ts=c()
92 for(i in 1:data_length){
93   df_Q20_ts=rbind(df_Q20_ts,statslist_ts[[i]]["Q20",])
94 }
95 df_Q20_ts=as.data.frame(df_Q20_ts)
96
97
98 #Q50 for all time series data
99 df_Q50_ts=c()
100 for(i in 1:data_length){
101   df_Q50_ts=rbind(df_Q50_ts,statslist_ts[[i]]["Q50",])
102 }
103 df_Q50_ts=as.data.frame(df_Q50_ts)
104
105
106 #Q80 for all time series data
107 df_Q80_ts=c()
108 for(i in 1:data_length){
109   df_Q80_ts=rbind(df_Q80_ts,statslist_ts[[i]]["Q80",])
110 }
111 df_Q80_ts=as.data.frame(df_Q80_ts)
112
```



```

114 #Q90 for all time series data
115 df_Q90_ts=c()
116 for(i in 1:data_length){
117   df_Q90_ts=rbind(df_Q90_ts,statslist_ts[[i]]["Q90",])
118 }
119 df_Q90_ts=as.data.frame(df_Q90_ts)
120
121
122 #Autocorrelation1 for all time series data
123 df_lag1_ts=c()
124 for(i in 1:data_length){
125   df_lag1_ts=rbind(df_lag1_ts,statslist_ts[[i]]["Autocorrelation1",])
126 }
127 df_lag1_ts=as.data.frame(df_lag1_ts)
128
129
130 #Autocorrelation2 for all time series data
131 df_lag2_ts=c()
132 for(i in 1:data_length){
133   df_lag2_ts=rbind(df_lag2_ts,statslist_ts[[i]]["Autocorrelation2",])
134 }
135 df_lag2_ts=as.data.frame(df_lag2_ts)
136
137
138 #Autocorrelation3 for all time series data
139 df_lag3_ts=c()
140 for(i in 1:data_length){
141   df_lag3_ts=rbind(df_lag3_ts,statslist_ts[[i]]["Autocorrelation3",])
142 }
143 df_lag3_ts=as.data.frame(df_lag3_ts)
144
145
146 df_teliko_ts=cbind(df_Mean_ts,df_StDev_ts,df_skewness_ts,df_Q5_ts,df_Q20_ts,
147                  df_Q50_ts,df_Q80_ts,df_Q90_ts,df_lag1_ts,df_lag2_ts,
148                  df_lag3_ts)
149

```

Εικόνα 5.15: Δημιουργία ενός μεγάλου data.frame με όλα τα αποτελέσματα στατιστικών μεγεθών

Καταλήγοντας η εξαγωγή (export) του data frame σε .txt αρχείο γίνεται μέσω της συνάρτησης write.table () όπου εντός των παρενθέσεων τοποθετούμε το data.frame και τη διαδρομή του αρχείου μέχρι το φάκελο αποθήκευσης όπως και το όνομα του .txt αρχείου. Επειδή όμως θέλουμε τα αποτελέσματα σε xls. αρχείο μεταφέρουμε το .txt αρχείο στο excel και το αποθηκεύουμε. Στο excel επίσης διαιρέσαμε την τυπική απόκλιση και τα quantiles με τη μέση τιμή ώστε να έχουμε αδιάστατα μεγέθη. Τα τελικά αποτελέσματα ενδεικτικά για τα στατιστικά μεγέθη της μέσης τιμής, του συντελεστή μεταβλητότητας, της ασυμετρίας, του λόγου Q50/Mean, του λόγου Q90/Mean και της Αυτοσυσχέτισης για υστέρηση ίση με 1 στο excel για τις πρώτες 20 χρονοσειρές είναι τα εξής:

Πίνακας 5-2: Αποτελέσματα Μέσης τιμής απορροής (mm) για τις πρώτες 20 χρονοσειρές ανά μήνα και ετήσια

Στατιστικά Χαρακτηριστικά ανά σταθμό και ανά μήνα												
Mean												
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	All
1.5311	1.5847	1.8703	1.9339	2.4484	2.2154	1.4077	1.1475	1.6398	2.1762	2.3610	1.8377	1.8462
2.7816	2.8788	2.1192	1.6494	1.1979	0.7733	0.4317	0.3360	0.3474	0.6525	1.2753	2.2184	1.3804
3.2720	3.4623	2.4305	1.8300	1.1763	0.7270	0.3571	0.2463	0.3183	0.7044	1.4266	2.7849	1.551
0.7357	0.7660	1.1021	1.5839	2.0924	1.9474	1.7097	1.4445	1.2404	1.1391	1.0919	0.9317	1.3183
0.7665	0.9737	1.1586	0.8636	0.5999	0.5421	0.4665	0.4030	0.4295	0.5414	0.8804	0.9541	0.7132
0.4905	0.6634	0.7893	0.5674	0.4060	0.4070	0.4058	0.3638	0.2976	0.3589	0.5959	0.6209	0.4962
1.3713	1.0194	1.1687	1.8073	2.0194	1.9243	1.9564	1.6743	1.6341	1.5439	1.4768	1.4243	1.5883
2.1667	1.7313	2.8442	5.5684	6.9519	5.6760	4.2847	3.3192	4.1897	5.2605	5.8232	3.5985	4.2942
2.8725	2.1477	3.3496	7.9958	15.8104	12.9075	7.7444	5.5167	8.1001	9.4086	9.6800	5.0801	7.5723
2.5913	2.2174	3.0610	4.2118	4.2331	3.8662	3.1718	2.4913	2.9400	4.0532	5.1865	3.9364	3.5011
1.3601	1.5921	1.9196	1.5309	1.1172	1.0531	0.8427	0.5810	0.8342	1.2349	1.7102	1.8251	1.2979
5.2701	5.0863	5.6562	6.9221	4.3591	2.9070	1.8335	1.7064	3.8588	6.0098	7.5287	7.4991	4.8802
4.1749	4.1767	4.6843	5.3202	3.1967	2.2226	1.4481	1.2969	2.6756	4.1944	5.7354	5.6259	3.7228
3.4564	3.5307	4.3396	4.3439	3.0179	2.0728	1.3797	1.1879	2.2080	3.6753	5.0015	4.6236	3.2322
2.4925	2.6030	3.4756	3.1944	2.0274	1.3863	0.8834	0.7901	1.4503	2.3018	3.1218	3.4167	2.2588
2.9382	2.7706	3.3627	3.4443	2.5474	2.1282	1.5119	1.2433	1.8798	2.9307	4.0776	3.7580	2.7138
2.4398	2.2662	3.4062	4.4597	4.3328	3.4219	2.5921	2.1844	2.8100	3.8214	4.7974	3.4237	3.3324
1.7317	1.7549	2.2856	2.5128	2.0700	1.9417	1.6666	1.2955	1.6309	2.1666	2.7892	2.3717	2.0181
3.0469	2.7339	3.7923	4.9265	3.5977	3.0476	2.5966	2.1915	3.0258	4.0003	5.0487	4.4684	3.5411
2.0014	2.1374	2.7583	2.7819	1.8578	1.6861	1.3470	1.0197	1.4790	2.1537	2.9184	2.7811	2.0748

Πίνακας 5-3: Αποτελέσματα συντελεστή μεταβλητότητας απορροής για τις πρώτες 20 χρονοσειρές ανά μήνα και ετήσια

Stdev/Mean												
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	All
0.5163	0.5231	0.5293	0.5671	0.6077	0.5338	0.5739	0.5770	0.6391	0.7558	0.7834	0.6130	0.6724
1.2082	1.3315	1.2226	0.9300	0.8842	0.8575	0.7399	0.6497	0.9162	1.9140	1.3690	1.2616	1.5930
0.9760	1.0021	0.9173	0.9028	0.8434	1.0056	1.0227	0.9972	1.4169	1.8827	1.2527	1.1850	1.4681
0.3211	0.4099	0.4909	0.4115	0.3771	0.4716	0.5746	0.6502	0.5385	0.6081	0.5136	0.4665	0.6116
1.3320	1.3445	1.2382	1.3304	1.8300	1.7375	2.2279	2.2419	2.2210	1.7484	1.6288	1.3250	1.6303
1.4198	1.6357	1.5338	1.8467	2.2133	2.6887	3.2698	3.2202	2.7228	2.6088	2.0881	1.5900	2.1493
0.7254	0.4867	0.5253	0.5905	0.6196	0.7078	0.7848	0.6215	0.6094	0.6000	0.5431	0.5222	0.6747
1.2305	1.0291	1.0999	0.6853	0.6891	0.6648	0.6849	0.8695	1.1663	1.1535	1.1809	1.1463	1.0466
2.1002	1.3843	1.1317	0.7814	0.5765	0.7149	0.8457	1.2015	1.4827	1.4277	1.5894	1.7022	1.3067
1.4374	1.1531	1.3695	0.8421	0.6778	0.8287	0.7810	0.8646	1.3762	1.4094	1.2593	1.2411	1.1783
1.3195	1.2277	1.1224	1.2392	1.6906	1.5942	2.1633	1.8084	2.5367	2.2182	1.6062	1.3575	1.6309
1.3533	1.3794	1.2172	0.7482	1.0037	1.1096	1.2782	1.7955	1.7946	1.4667	1.2261	1.3813	1.4264
1.2505	1.2439	1.1067	0.7906	1.1137	1.2173	1.4536	1.8100	1.7311	1.4567	1.1378	1.2474	1.3556
1.1085	1.0839	0.9792	0.8116	1.0989	1.2344	1.3705	1.8692	1.6750	1.3782	1.0421	1.1325	1.2560
1.1293	1.0608	1.1388	1.0196	1.5641	1.5499	1.8437	2.4642	2.1978	1.7371	1.2578	1.1723	1.4623
0.7977	0.8469	0.7491	0.6351	0.7094	0.8537	0.9243	1.1379	1.2200	0.9551	0.7129	0.7061	0.8891
1.2322	1.0784	1.0165	0.6554	0.7333	0.9093	1.0925	1.3754	1.5176	1.3059	1.1043	1.1216	1.1202
1.2377	1.0976	1.0037	0.8086	0.9104	1.1886	1.3348	1.4096	1.6684	1.5783	1.2625	1.1283	1.2435
1.6442	1.2252	1.1323	0.7522	1.0294	1.1522	1.2103	1.5762	1.8822	1.7498	1.4678	1.2958	1.4028
0.9611	0.9350	0.8787	0.7808	0.9774	1.0147	1.1697	1.1320	1.3190	1.1666	0.9029	0.9138	1.0433

Πίνακας 5-4: Αποτελέσματα Ασυμμετρίας απορροής (mm) για τις πρώτες 20 χρονοσειρές ανά μήνα και ετήσια

												2
Skewness												
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	All
2.1665	2.3780	1.7621	1.5046	1.3645	1.0588	1.2096	2.3104	2.7266	2.1035	2.0028	2.4686	2.2943
3.8336	4.6632	5.0104	2.8168	3.0542	3.6849	2.3677	1.4621	11.8232	14.0028	4.6374	4.0475	6.1071
2.3098	2.3472	2.9517	3.2701	2.1938	5.0995	4.2544	3.1085	5.9723	6.3192	3.1098	2.8706	3.7308
1.5769	2.6774	2.0698	2.7731	1.5093	2.4410	2.7237	3.6927	2.6086	3.3390	2.7624	2.8932	2.5689
3.8012	3.4226	2.6903	4.1395	6.3265	4.4944	6.2897	6.8423	6.7700	4.8922	4.0390	3.6740	4.3841
4.9937	4.9632	4.1689	7.7342	10.1722	10.1416	11.8779	9.2870	8.0379	10.6663	5.7587	5.7401	8.1226
3.9220	1.6857	1.6085	2.3080	1.1325	4.1856	3.5127	2.4184	2.0741	2.1624	2.0588	1.8422	3.1671
7.6533	6.5213	6.1179	2.8903	3.2854	2.6952	3.7129	6.3593	4.0430	3.5985	4.1315	4.9002	4.3069
11.3103	7.1455	4.8137	3.6210	1.6569	1.9998	3.0798	5.8198	4.4635	3.3014	4.1883	6.3354	4.2301
12.0182	7.9274	10.1241	7.6395	3.8990	11.4232	4.5235	6.6036	9.4441	7.9185	5.0429	6.6486	8.2767
4.4309	3.6152	2.5178	4.1837	6.4999	4.9258	8.0563	6.5854	8.3415	6.2175	4.8398	3.5317	5.3324
3.6480	3.6860	3.5890	2.1967	3.5375	3.4368	5.4131	6.7222	4.6752	3.5889	2.9296	3.3970	4.1022
3.0559	2.9484	2.7866	1.9732	3.7508	3.4913	5.5127	5.9022	4.0136	3.1894	2.5756	2.8577	3.3754
2.9388	2.7384	2.6849	2.5783	3.9032	4.6641	5.5394	8.4624	4.2968	2.9829	2.2929	2.7235	3.2406
3.1665	2.4806	3.7297	2.3684	5.6100	4.6298	6.7981	9.2242	6.3504	3.3618	2.6440	2.5795	3.7842
1.5061	1.4822	1.2001	1.0198	1.7257	2.0410	2.8138	3.3014	2.4929	1.3617	0.8079	1.0933	1.5368
5.2944	4.1695	4.0800	2.5041	2.8543	3.2859	4.5409	12.4490	5.9879	3.5749	3.2866	4.0532	4.5047
5.1561	4.5208	3.5176	3.1586	4.1581	5.3630	5.5370	6.7102	7.5147	5.3914	4.7630	3.7673	5.3704
8.8688	4.4655	4.9380	3.4538	5.2611	4.8128	4.3820	8.5507	8.5023	6.0521	5.3718	4.3048	6.6497
2.4729	1.9222	1.7358	1.5072	2.8303	2.7993	3.9664	4.3430	3.3201	2.1320	1.5469	1.6865	2.2465

Πίνακας 5-5: Αποτελέσματα quantile 50%/Μέσης Τιμής απορροής για τις πρώτες 20 χρονοσειρές ανά μήνα και ετήσια

												5
Q50/Mean												
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	All
0.8478	0.8479	0.8107	0.8404	0.8458	0.9070	0.8538	0.8395	0.8119	0.7307	0.6996	0.8251	0.7932
0.5983	0.6003	0.6447	0.7012	0.6988	0.8134	0.9196	0.9470	0.8455	0.6752	0.5374	0.5798	0.5154
0.6650	0.6285	0.7795	0.7583	0.6857	0.6709	0.7669	0.7692	0.6365	0.4607	0.5049	0.6230	0.4789
0.9340	0.8904	0.8545	0.9094	0.9152	0.8884	0.8475	0.8427	0.8402	0.8194	0.8536	0.8776	0.8291
0.5997	0.5323	0.5336	0.5873	0.5369	0.4579	0.4254	0.3776	0.4429	0.4901	0.4357	0.5762	0.4796
0.5702	0.5067	0.4942	0.5924	0.5213	0.4384	0.3375	0.2954	0.3486	0.3584	0.3983	0.5520	0.4600
0.7653	0.8617	0.8424	0.9297	0.9080	0.8588	0.8076	0.8662	0.8587	0.8974	0.8823	0.8610	0.8352
0.7005	0.7046	0.6885	0.8326	0.8569	0.8154	0.8034	0.7850	0.6182	0.5820	0.6120	0.6694	0.6755
0.5904	0.6354	0.6573	0.7818	0.8779	0.7549	0.7531	0.6490	0.5130	0.4578	0.4482	0.5139	0.5526
0.6826	0.7041	0.6583	0.8491	0.8321	0.8156	0.7953	0.7965	0.6459	0.6110	0.6495	0.6760	0.6898
0.5707	0.5582	0.5626	0.6209	0.5153	0.4463	0.4559	0.4913	0.3903	0.3261	0.4416	0.5157	0.4648
0.5132	0.4690	0.5778	0.8032	0.6894	0.6254	0.6300	0.5221	0.3592	0.4231	0.5538	0.4677	0.4965
0.5157	0.4985	0.5854	0.7772	0.6462	0.5712	0.5861	0.5079	0.3606	0.4018	0.5612	0.4894	0.4911
0.6006	0.5873	0.6772	0.7657	0.6435	0.6205	0.6068	0.5276	0.3929	0.4439	0.6295	0.5563	0.5572
0.5889	0.6223	0.6204	0.6080	0.5006	0.4928	0.5228	0.4150	0.3229	0.3121	0.5065	0.5770	0.4666
0.7125	0.7025	0.7538	0.8427	0.7871	0.7274	0.6972	0.6349	0.5242	0.6156	0.8549	0.7893	0.6793
0.6443	0.6553	0.6934	0.8558	0.8189	0.7400	0.6757	0.6765	0.5379	0.5389	0.6475	0.6526	0.6577
0.6372	0.6511	0.6953	0.8040	0.7475	0.6740	0.6419	0.6414	0.5191	0.4796	0.6033	0.6175	0.6316
0.5849	0.6017	0.6575	0.8339	0.7121	0.6520	0.6467	0.6295	0.4870	0.4587	0.5600	0.5657	0.5976
0.6416	0.6376	0.6839	0.7382	0.6673	0.6342	0.6447	0.7000	0.5456	0.4858	0.6632	0.6135	0.5917

Πίνακας 5-6: Αποτελέσματα quantile 90%/Μέσης Τιμής απορροής για τις πρώτες 20 χρονοσειρές ανά μήνα και ετήσια

													7
Q90/Mean													
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	All	
1.6603	1.5946	1.6938	1.7408	1.7925	1.6839	1.8039	1.6973	1.7064	2.0109	2.1200	1.7172	1.8161	
2.2981	2.0688	1.9402	1.9588	2.0024	1.7668	1.8143	1.7483	1.6909	1.7724	2.4520	2.3390	2.2712	
2.2817	2.3486	2.0607	1.9271	1.9455	2.2448	1.7545	1.5841	1.9212	2.0505	2.3642	2.3586	2.2836	
1.3910	1.4601	1.5990	1.5086	1.5064	1.5424	1.6533	1.5838	1.6119	1.6530	1.6297	1.5369	1.7161	
1.9847	2.3312	2.5255	2.1706	1.7862	2.1489	1.8699	2.0837	1.9164	2.2627	2.4306	2.1532	2.2670	
1.9213	2.1199	2.3048	1.7040	1.6652	1.7343	1.7302	1.6425	1.5945	2.0051	1.9589	2.1280	1.9402	
1.7197	1.6454	1.7152	1.6735	1.8247	1.8018	1.8804	1.7409	1.7173	1.6917	1.6448	1.6806	1.7817	
1.5696	1.6250	1.9639	1.7861	1.6989	1.7159	1.6599	1.5598	2.0413	2.1620	2.1181	1.7718	2.0330	
1.5972	1.7126	1.9935	1.8854	1.7154	1.9514	1.8857	1.9029	2.2357	2.6086	2.4715	1.8819	2.3429	
1.5835	1.7318	1.9048	1.7021	1.7260	1.6911	1.6731	1.5794	1.7789	1.9326	1.8981	1.7776	1.7781	
2.1848	2.3549	2.5414	2.2485	2.0439	2.3017	1.9371	2.0565	1.7757	2.3423	2.4539	2.3131	2.3719	
2.3414	2.4039	2.1920	1.8809	2.0251	2.1522	1.7210	1.7413	2.6094	2.5453	2.4331	2.3794	2.3404	
2.4743	2.4837	2.3210	2.0044	2.0923	2.2136	1.8435	1.8194	2.6708	2.7375	2.3779	2.5145	2.4943	
2.2475	2.3615	2.1656	1.9473	2.0020	2.0304	1.7917	1.9175	2.4628	2.5598	2.2872	2.3951	2.3500	
2.2729	2.2796	2.2841	2.3464	2.2227	2.1337	1.7683	1.8499	2.3400	2.7136	2.5330	2.3742	2.5060	
2.1734	2.2650	2.1490	1.8931	1.9631	2.1171	1.9618	2.0174	2.6341	2.4758	2.0319	2.0482	2.2943	
1.8353	1.9759	2.0825	1.7539	1.7617	1.9129	1.9667	1.8596	2.1455	2.2984	2.1345	1.9789	2.0581	
1.8847	1.9966	2.0678	1.8790	1.8371	1.8568	1.8841	1.8052	2.0069	2.1238	2.1227	2.1503	2.0003	
1.9070	2.0841	2.0131	1.7519	1.9124	1.9747	1.8735	1.7398	1.9864	2.0517	2.2232	2.1265	2.0138	
2.2567	2.3841	2.2934	2.1167	2.0813	2.1460	1.9035	1.8145	2.2400	2.6387	2.3281	2.3459	2.4002	

Πίνακας 5-7: Αποτελέσματα υστέρησης 1 απορροής (mm) για τις πρώτες 20 χρονοσειρές ανά μήνα και ετήσια

													8
lag1													
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	All	
0.9282	0.9278	0.9329	0.9145	0.9358	0.9322	0.9330	0.9271	0.9185	0.9446	0.9462	0.9379	0.971092	
0.7461	0.7971	0.7816	0.8305	0.8440	0.8259	0.9131	0.9400	0.6354	0.6904	0.7068	0.7338	0.831273	
0.8370	0.8516	0.8045	0.8026	0.8586	0.8031	0.8836	0.9259	0.8792	0.8397	0.7949	0.8210	0.894748	
0.8618	0.8630	0.8733	0.8590	0.8741	0.8305	0.8677	0.8690	0.8118	0.8562	0.8135	0.8524	0.914775	
0.8688	0.8374	0.8667	0.8292	0.6973	0.7527	0.7950	0.7691	0.7215	0.7739	0.8193	0.7916	0.83386	
0.7480	0.7248	0.8168	0.6660	0.5065	0.4552	0.7081	0.6897	0.6300	0.5474	0.6449	0.6346	0.678032	
0.8726	0.8991	0.9006	0.8589	0.9060	0.7215	0.8317	0.7872	0.7977	0.8722	0.8485	0.8674	0.867111	
0.7221	0.7889	0.7906	0.8043	0.7956	0.7593	0.7134	0.6651	0.7407	0.7191	0.7042	0.7706	0.787888	
0.7882	0.8280	0.7862	0.8084	0.8165	0.8069	0.7542	0.7706	0.7669	0.7400	0.7421	0.7815	0.81942	
0.6083	0.6790	0.6732	0.6991	0.7137	0.5212	0.6628	0.5687	0.5668	0.5347	0.6002	0.6725	0.648753	
0.7784	0.7742	0.8352	0.7793	0.7377	0.7392	0.6874	0.6484	0.7084	0.6904	0.7423	0.7574	0.775084	
0.6531	0.6673	0.6830	0.6839	0.6891	0.6535	0.6164	0.6603	0.5462	0.6163	0.6260	0.6509	0.689149	
0.7036	0.7132	0.7512	0.7291	0.6988	0.6409	0.6153	0.6555	0.6433	0.6770	0.6635	0.7035	0.741767	
0.7454	0.7388	0.7815	0.7620	0.7059	0.6560	0.7052	0.6299	0.6933	0.6972	0.7006	0.7283	0.770705	
0.7737	0.8067	0.8402	0.7851	0.7493	0.7316	0.7169	0.7719	0.6985	0.7334	0.7378	0.7829	0.809698	
0.9131	0.9101	0.9026	0.8939	0.8708	0.8317	0.8321	0.8193	0.8650	0.8968	0.9007	0.9044	0.929211	
0.7373	0.7282	0.7322	0.7667	0.7741	0.6438	0.7229	0.4880	0.6803	0.6959	0.6712	0.7346	0.729832	
0.7236	0.7228	0.7609	0.7421	0.6924	0.5491	0.6267	0.5942	0.6252	0.6198	0.6109	0.6723	0.682347	
0.5745	0.7053	0.6364	0.6690	0.6399	0.5662	0.6198	0.5052	0.5906	0.5861	0.5279	0.6037	0.624714	
0.8916	0.8944	0.9031	0.8903	0.8642	0.8600	0.8771	0.8332	0.8725	0.8858	0.8833	0.9090	0.924254	

5.3 Προσαρμογή κατανομών lognormal, gamma, weibull

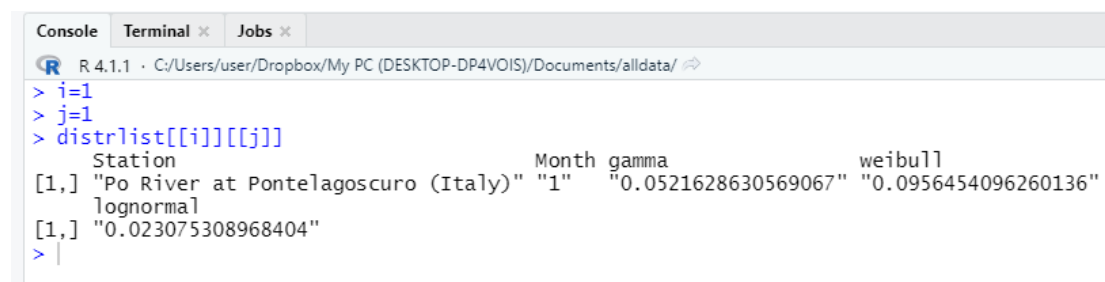
Στην ενότητα αυτή παρουσιάζονται τα βασικότερα στοιχεία του κώδικα στη γλώσσα προγραμματισμού R για την προσαρμογή των κατανομών lognormal, gamma, weibull στα δεδομένα ημερήσιων παροχών ανά μήνα. Το κριτήριο σύγκρισης με βάση το οποίο συμπεραίνουμε την καταλληλότητα της κάθε κατανομής είναι το μέσο τετραγωνικό σφάλμα (MSE).

Αρχικά δημιουργήθηκε μια κενή λίστα όπου στη συνέχεια μέσα από το for loop θα υπολογιστεί για κάθε μια από τις εξεταζόμενες κατανομές το μέσο τετραγωνικό σφάλμα με το οποίο έγινε η σύγκριση. Η διαδικασία αυτή γίνεται για όλες τις χρονοσειρές ανά μήνα. Επίσης μέσα από την επανάληψη δίνεται η δυνατότητα εκτύπωσης διαγραμμάτων ανά μήνα και χρονοσειρά όπου φαίνεται η προσαρμογή της κάθε κατανομής στα δεδομένα της βάσης. Παρακάτω φαίνεται ο κώδικας επανάληψης από όπου εξάγεται το μέσο τετραγωνικό σφάλμα για κάθε κατανομή:

```
259 #Fitting gamma, weibull, lognormal distributions
260 distrlist=vector("list",data_length)
261 for(i in 1:data_length){
262   for(j in 1:12){
263     flow_data_per_monthz=flow_data_per_month_list[[i]][[j]]
264     flow_data_per_monthz=cordata(flow_data_per_monthz)
265     flow_data_per_monthz=na.omit(flow_data_per_monthz)
266     flow_data_per_monthz=flow_data_per_monthz[flow_data_per_monthz>0]
267     fit_gam=elmgamma(x = as.vector(flow_data_per_monthz), plot = 0) # Fit the Gamma distribution
268     fit_weib=elmweibull(x = as.vector(flow_data_per_monthz), plot = 0) # Fit the weibull distribution
269     fit_lnorm=elmlnorm(x = as.vector(flow_data_per_monthz), plot = 0) # Fit the Lognormal distribution
270     gam=fit_gam$Eval$MSEquant
271     weib=fit_weib$Eval$MSEquant
272     lnorm=fit_lnorm$Eval$MSEquant
273     results=cbind(namelist[i],j,gam,weib,lnorm)
274     colnames(results)=c("Station","Month","gamma","weibull","lognormal")
275     distrlist[[i]][[j]]=results
276   }
277 }
```

Εικόνα 5.16: For loop για τον υπολογισμό του μέσου τετραγωνικού σφάλματος κάθε κατανομής ανά μήνα χρονοσειράς

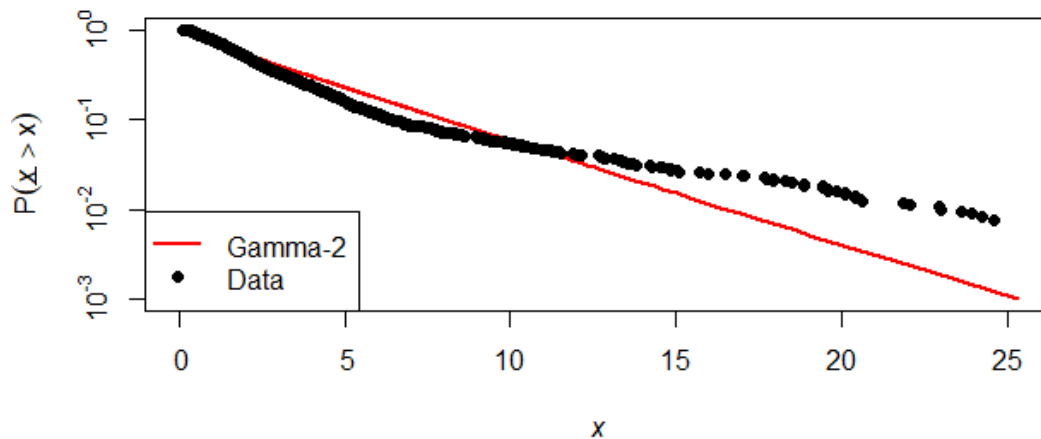
Έτσι λοιπόν για $i=1$ και $j=1$ τα αποτελέσματα της εντολής `distrlist[[i]][[j]]` είναι τα εξής:



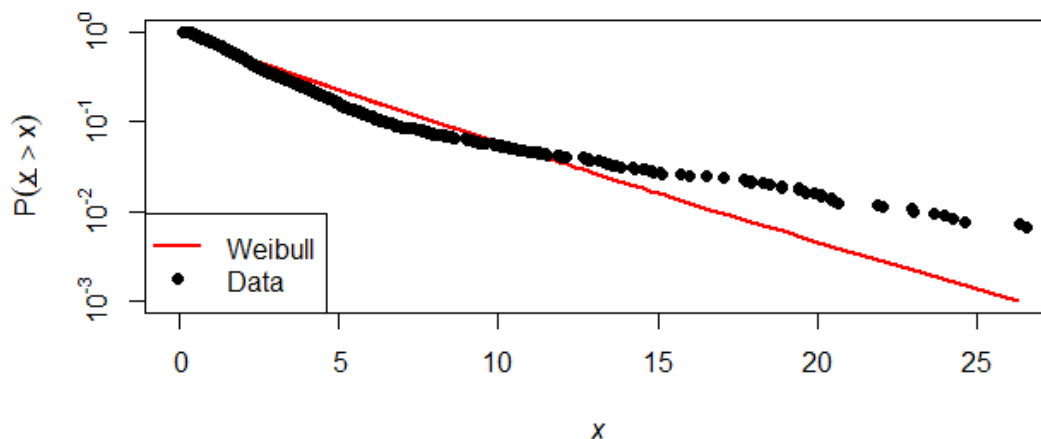
```
R 4.1.1 · C:/Users/user/Dropbox/My PC (DESKTOP-DP4VOIS)/Documents/alldata/ ↗
> i=1
> j=1
> distrlist[[i]][[j]]
      Station      Month gamma      weibull
[1,] "Po River at Pontelagoscuro (Italy)" "1"  "0.0521628630569067" "0.0956454096260136"
      lognormal
[1,] "0.023075308968404"
> |
```

Εικόνα 5.17: Αποτελέσματα μέσου τετραγωνικού σφάλματος για χρονοσειρά $i=1$ και μήνα $j=1$ για τις κατανομές gamma, weibull και lognormal

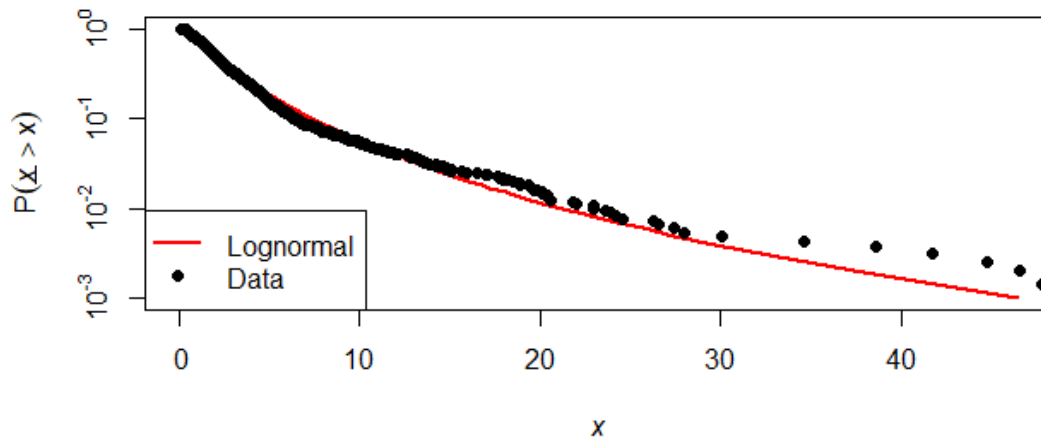
Τα διαγράμματα προσαρμογής των κατανομών gamma, weibull, lognormal που προκύπτουν από τις εντολές `fit_gam=elmgamma(x = as.vector(flow_data_per_monthz), plot = 1)`, `fit_wei=elmweibull(x = as.vector(flow_data_per_monthz), plot = 1)`, `fit_lnorm=elmlnorm(x = as.vector(flow_data_per_monthz), plot = 1)` απεικονίζονται στα παρακάτω Σχήματα:



Σχήμα 5.1: Προσαρμογή κατανομής gamma για τη χρονοσειρά $i=1$ και το μήνα $j=1$



Σχήμα 5.2: Προσαρμογή κατανομής Weibull για τη χρονοσειρά $i=1$ και το μήνα $j=1$



Σχήμα 5.3: Προσαρμογή κατανομής Lognormal για τη χρονοσειρά $i=1$ και το μήνα $j=1$

Το επόμενο βήμα είναι η δημιουργία μιας λίστας η οποία θα παίρνει σαν όρισμα μόνο την χρονοσειρά και θα δίνει σαν αποτέλεσμα για κάθε χρονοσειρά τα μέσα τετραγωνικά σφάλματα των 12 μηνών. Στις παρακάτω εικόνες φαίνονται τόσο το for loop με τις συναρτήσεις όσο και τα αποτελέσματα για τη χρονοσειρά $i=1$.

```

283 #data frame with fitting results per time series data
284 df_distrlist=vector("list",data_length)
285 for(i in 1:data_length){
286   for(j in 1:12){
287     df_distrlist[[i]]=rbind(df_distrlist[[i]],distrlist[[i]][[j]])
288   }
289 }

```

Εικόνα 5.18: For loop για την ένωση των αποτελεσμάτων των 12 μηνών κάθε χρονοσειράς

```

> df_distrlist[[i]]
      Station Month      gamma      weibull
1 Po River at Pontelagoscuro (Italy) 1 0.0521628630569067 0.0956454096260136
2 Po River at Pontelagoscuro (Italy) 2 0.0621202791479821 0.110433502910887
3 Po River at Pontelagoscuro (Italy) 3 0.0421331372389967 0.0919240220215967
4 Po River at Pontelagoscuro (Italy) 4 0.0174703534941668 0.0534453165637611
5 Po River at Pontelagoscuro (Italy) 5 0.0096148945054954 0.0449212147972697
6 Po River at Pontelagoscuro (Italy) 6 0.00496330941855724 0.0181278717890217
7 Po River at Pontelagoscuro (Italy) 7 0.00264101837831382 0.0120930447446881
8 Po River at Pontelagoscuro (Italy) 8 0.0267505650089666 0.0505230434479548
9 Po River at Pontelagoscuro (Italy) 9 0.113755222376507 0.180226692762832
10 Po River at Pontelagoscuro (Italy) 10 0.15872025340314 0.240519764455276
11 Po River at Pontelagoscuro (Italy) 11 0.21438053616443 0.298736648617872
12 Po River at Pontelagoscuro (Italy) 12 0.101564908791454 0.174303154749495
      Lognormal
1 0.023075308968404
2 0.0275623691311204
3 0.0135569083823206
4 0.00891319217624236
5 0.0407013132376175
6 0.0336748494891833
7 0.0147365987811806
8 0.00659674000437858
9 0.0440232955805192
10 0.0809173567787635
11 0.15997539413948
12 0.0322701652547494
> |

```

Εικόνα 5.19: Αποτελέσματα μέσου τετραγωνικού σφάλματος για την χρονοσειρά $i=1$

Συνεχίζοντας στόχος είναι να γνωρίζουμε σε κάθε χρονοσειρά ποια κατανομή επικρατεί και ποιο είναι το ποσοστό επιτυχίας της κάθε κατανομής. Γι' αυτό το λόγο δημιουργήσαμε έναν κώδικα στον οποίο με βάση ποιο είναι το μικρότερο μέσο τετραγωνικό σφάλμα βρίσκει την καταλληλότερη κατανομή ανά μήνα κάθε χρονοσειράς και στη συνέχεια το ποσοστό επιτυχίας. Ακολούθως ενώσαμε τα ποσοστά επιτυχίας κάθε κατανομής σε ένα μεγάλο data frame. Έτσι προέκυψαν τα data frames `df_dist_3l`, `df_dist_3g`, `df_dist_3w`. Ο κώδικας φαίνεται στην Εικόνα 5.20:

```

295 #make data frame with posability success every distribution
296 df_dist_wins=vector("list",data_length)
297 df_dist_gamma=vector("list",data_length)
298 df_dist_weibull=vector("list",data_length)
299 for(i in 1:data_length){
300   for(j in 1:12){
301     df_distrlist[[i]]=as.data.frame(df_distrlist[[i]])
302     which_month=which(df_distrlist[[i]][,"Month"]==j)
303     df_distrlist_month=df_distrlist[[i]][which_month,]
304     df_distrlist_month$Min=apply(df_distrlist_month[,c("gamma","weibull","lognormal")],1,min)
305     count_gamma=length(which(df_distrlist_month$Min==df_distrlist_month$gamma))
306     count_wei=length(which(df_distrlist_month$Min==df_distrlist_month$weibull))
307     count_lnorm=length(which(df_distrlist_month$Min==df_distrlist_month$lognormal))
308     if(count_gamma==max(count_gamma,count_wei,count_lnorm)){
309       print(paste0("gamma"))
310     } else if(count_wei==max(count_gamma,count_wei,count_lnorm)){
311       print(paste0("wei"))
312     } else {
313       print(paste0("lognormal"))
314     }
315     df_dist_wins[[i]][[j]]=count_lnorm/sum(count_gamma+count_wei+count_lnorm)*100
316     df_dist_gamma[[i]][[j]]=count_gamma/sum(count_gamma+count_wei+count_lnorm)*100
317     df_dist_weibull[[i]][[j]]=count_wei/sum(count_gamma+count_wei+count_lnorm)*100
318   }
319 }
320
321 i=87#Timeseriesdata
322 j=12#Month
323 df_dist_wins[[i]]
324 df_dist_gamma[[i]][[j]]
325 df_dist_weibull[[i]][[j]]
326
327
328 df_dist_3l=c()
329 df_dist_3g=c()
330 df_dist_3w=c()
331 for(i in 1:data_length){
332   df_dist_wins[[i]]=as.data.frame(df_dist_wins[[i]])
333   colnames(df_dist_wins[[i]])=c("Month1","Month2","Month3","Month4","Month5","Month6",
334     "Month7","Month8","Month9","Month10","Month11","Month12")
335   df_dist_3l=rbind(df_dist_3l,df_dist_wins[[i]])
336
337   df_dist_gamma[[i]]=as.data.frame(df_dist_gamma[[i]])
338   colnames(df_dist_gamma[[i]])=c("Month1","Month2","Month3","Month4","Month5","Month6",
339     "Month7","Month8","Month9","Month10","Month11","Month12")
340   df_dist_3g=rbind(df_dist_3g,df_dist_gamma[[i]])
341
342   df_dist_weibull[[i]]=as.data.frame(df_dist_weibull[[i]])
343   colnames(df_dist_weibull[[i]])=c("Month1","Month2","Month3","Month4","Month5","Month6",
344     "Month7","Month8","Month9","Month10","Month11","Month12")
345   df_dist_3w=rbind(df_dist_3w,df_dist_weibull[[i]])
346 }

```

Εικόνα 5.20: Υπολογισμός καταλληλότερης κατανομής και ποσοστά επιτυχίας κάθε κατανομής

Καταλήγοντας η εξαγωγή (export) των data frame σε .txt αρχεία γίνεται μέσω της συνάρτησης write.table () όπου εντός των παρενθέσεων τοποθετούμε ξεχωριστά το κάθε data.frame και τη διαδρομή του αρχείου μέχρι το φάκελο αποθήκευσης όπως και το όνομα του .txt αρχείου. Επειδή όμως θέλουμε τα αποτελέσματα σε xlsx. αρχείο μεταφέρουμε το .txt αρχείο στο excel και το αποθηκεύουμε. Η εντολή εμφανίζεται στην Εικόνα 5.21:

```

351 write.table(df_dist_3w, file = "C:/Users/user/Desktop/weibull.txt", append = FALSE,
352             quote = TRUE, sep = " ", eol = "\n", na = "NA", dec = ",", row.names = TRUE,
353             col.names = TRUE)
354

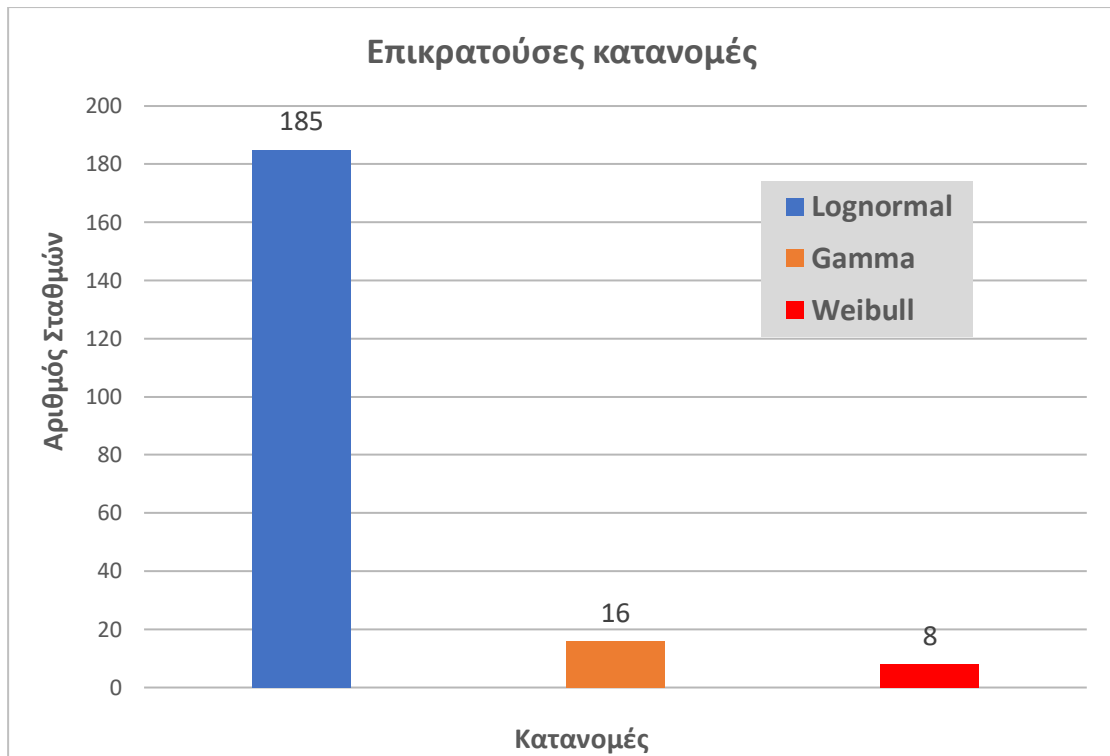
```

Εικόνα 5.21: Εξαγωγή των ποσοστών επιτυχίας κάθε κατανομής σε .txt αρχεία

Αφού αποθηκεύτηκαν τα αποτελέσματα σε xlsx. αρχείο έγινε επεξεργασία και τα τελικά αποτελέσματα για όλες τις χρονοσειρές είναι τα εξής: σε σύνολο 209 χρονοσειρών στις 185 επικρατεί η κατανομή Lognormal, στις 16 επικρατεί η κατανομή Gamma και στις 8 επικρατεί η κατανομή weibull. Συγκεκριμένα ο πίνακας με τα ποσοστά επιτυχίας κάθε κατανομής για τις πρώτες 30 χρονοσειρές απεικονίζεται παρακάτω:

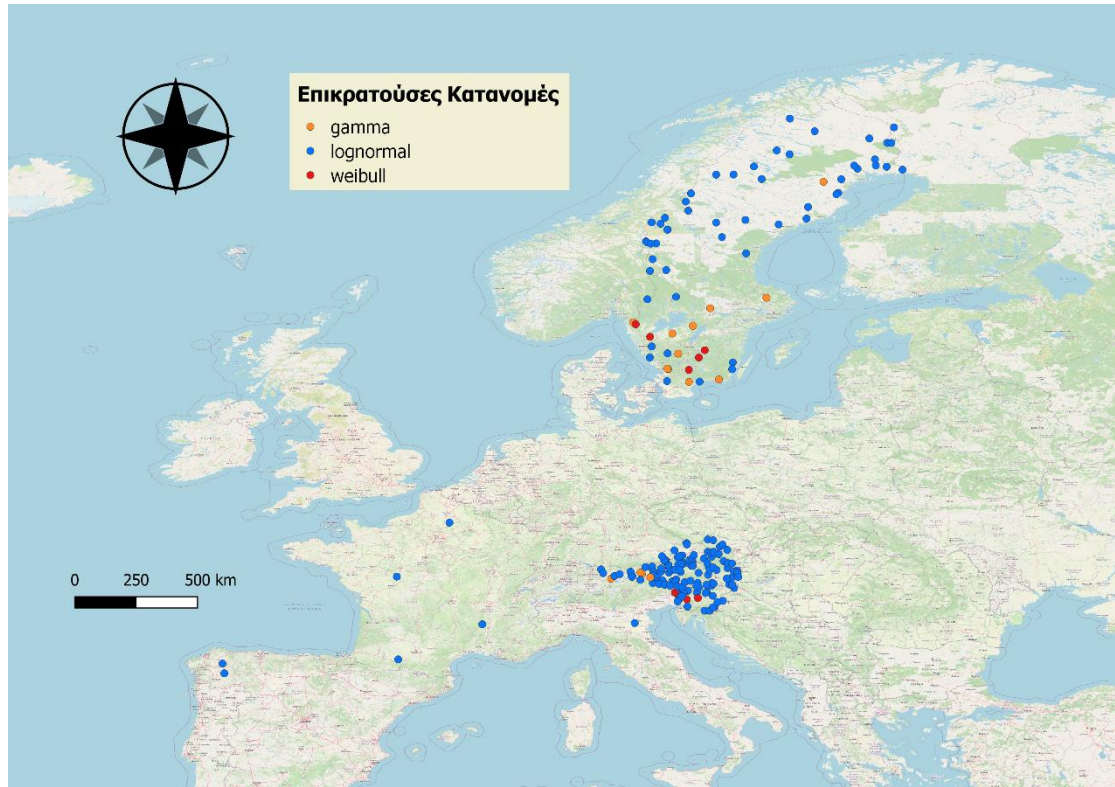
Πίνακας 5-8: Ποσοστά επιτυχίας κατανομών lognormal, gamma, weibull ανά χρονοσειρά

Ποσοστά Επιτυχίας κάθε κατανομής ανά χρονοσειρά				
Time Series	lnorm	gamma	weibull	winner
1	75.00	25.00	0.00	lognormal
2	91.67	8.33	0.00	lognormal
3	66.67	16.67	16.67	lognormal
4	100.00	0.00	0.00	lognormal
5	91.67	0.00	8.33	lognormal
6	100.00	0.00	0.00	lognormal
7	91.67	8.33	0.00	lognormal
8	100.00	0.00	0.00	lognormal
9	66.67	8.33	25.00	lognormal
10	91.67	0.00	8.33	lognormal
11	91.67	0.00	8.33	lognormal
12	83.33	0.00	16.67	lognormal
13	50.00	8.33	41.67	lognormal
14	75.00	8.33	16.67	lognormal
15	58.33	16.67	25.00	lognormal
16	16.67	25.00	58.33	weibull
17	100.00	0.00	0.00	lognormal
18	100.00	0.00	0.00	lognormal
19	91.67	0.00	8.33	lognormal
20	50.00	33.33	16.67	lognormal
21	83.33	0.00	16.67	lognormal
22	75.00	8.33	16.67	lognormal
23	66.67	16.67	16.67	lognormal
24	100.00	0.00	0.00	lognormal
25	33.33	25.00	41.67	weibull
26	25.00	16.67	58.33	weibull
27	75.00	8.33	16.67	lognormal
28	75.00	0.00	25.00	lognormal
29	58.33	0.00	41.67	lognormal
30	100.00	0.00	0.00	lognormal



Σχήμα 5.4: Αριθμός σταθμών που επικρατούν οι κατανομές lognormal, gamma, weibull

Ακόμη μέσω του ελεύθερου λογισμικού QGIS δημιουργήσαμε έναν χάρτη όπου φαίνεται και η γεωγραφική κατανομή των σταθμών και ποια κατανομή επικρατεί στις περισσότερες χρονοσειρές ώστε να έχουμε και ένα εποπτικό αποτέλεσμα. Παρατηρούμε την κυριαρχία της log normal κατανομής στους περισσότερους σταθμούς στην Κεντρική, Δυτική και Βόρεια Ευρώπη, η κατανομή gamma φαίνεται να κυριαρχεί στη νότια Σουηδία ενώ η κατανομή Weibull έχει πολύ χαμηλά ποσοστά επικράτησης στην Κεντρική και βόρεια Ευρώπη.

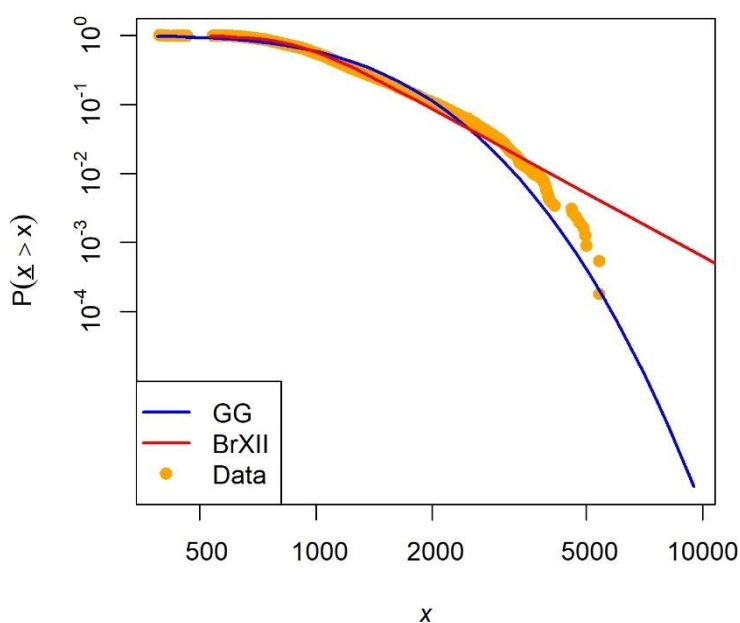


Σχήμα 5.5: Επικρατούσες κατανομές στους σταθμούς (QGIS) - MSE

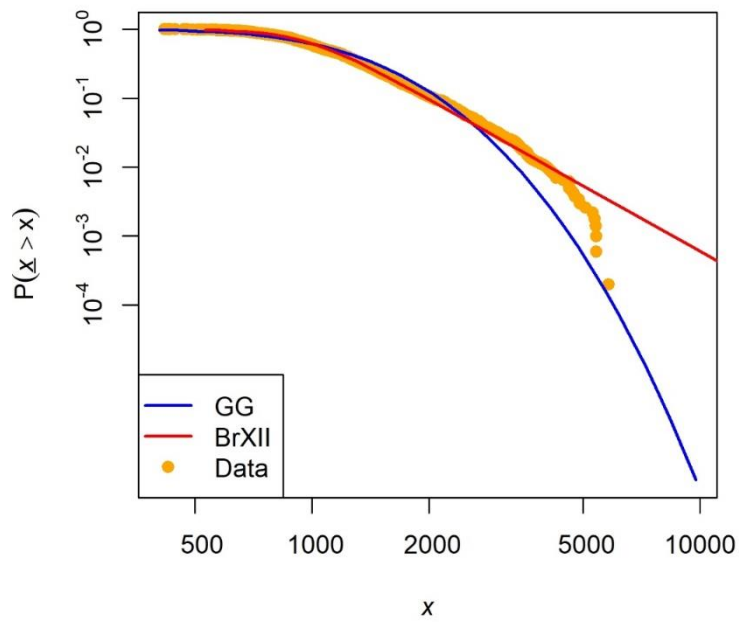
5.4 Προσαρμογή κατανομών PBF, Generalized Gamma

Η ίδια ακριβώς μεθοδολογία προσαρμογής ακολουθήθηκε στη γλώσσα προγραμματισμού R για τις κατανομές Pareto – Burr – Feller και Generalized Gamma οι οποίες εξετάστηκαν ξεχωριστά από τις προηγούμενες, ενώ η σύγκριση όλων μαζί έγινε με τα λεγόμενα «διαγράμματα λόγου ροπής» (moment ratio diagrams) τα οποία μέσω μιας γραφικής σύγκρισης των στατιστικών των δειγμάτων, παρέχουν μια ένδειξη καταλληλότητας των κατανομών για να περιγράψουν τις χρονοσειρές της βάσης δεδομένων. Αυτά ωστόσο θα αναλυθούν στο επόμενο κεφάλαιο. Κριτήριο σύγκρισης καταλληλότητας των κατανομών αποτέλεσε το Μέσο Τετραγωνικό Σφάλμα.

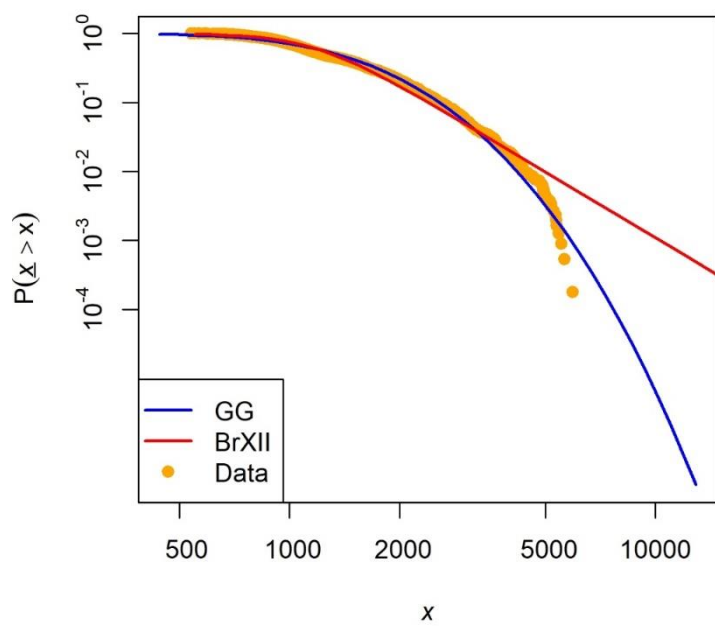
Τα διαγράμματα προσαρμογής των κατανομών PBF και Generalized Gamma απεικονίζονται για την χρονοσειρά $i=1$ και για τους 3 πρώτους μήνες δηλαδή για $j=1, 2, 3$ καθώς και ετήσια στα παρακάτω Σχήματα:



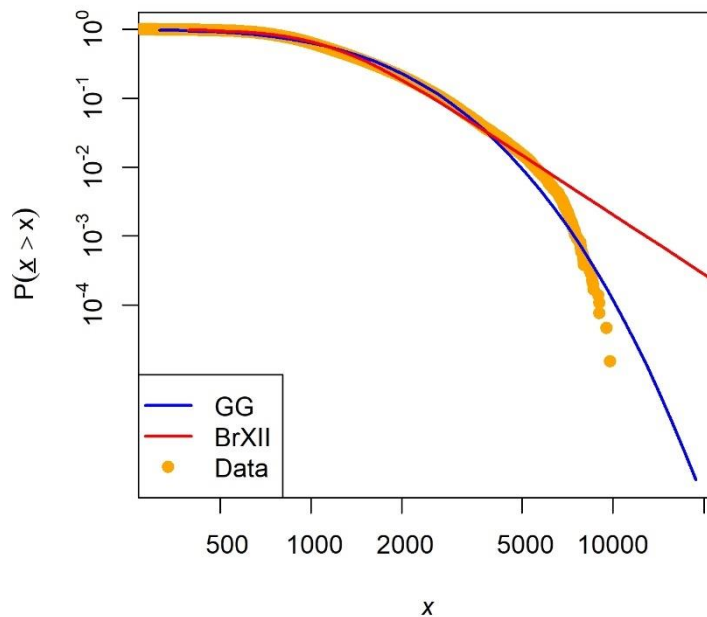
Σχήμα 5.6: Προσαρμογή κατανομής PBF – Generalized Gamma για τη χρονοσειρά $i=1$ και το μήνα $j=1$



Σχήμα 5.7: Προσαρμογή κατανομής PBF – Generalized Gamma για τη χρονοσειρά $i=1$ και το μήνα $j=2$



Σχήμα 5.8: Προσαρμογή κατανομής PBF – Generalized Gamma για τη χρονοσειρά $i=1$ και το μήνα $j=3$



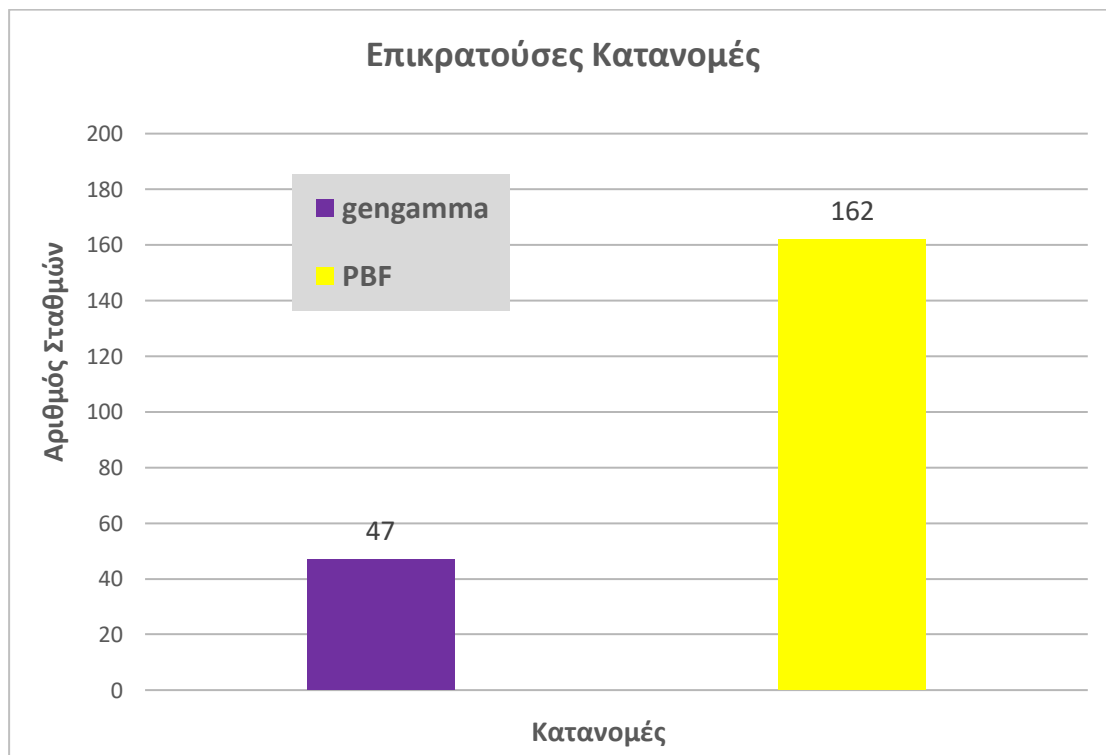
Σχήμα 5.9: Προσαρμογή κατανομής PBF – Generalized Gamma για τη χρονοσειρά $i=1$ ετήσια

Παρομοίως με τις προηγούμενες κατανομές αφού αποθηκεύτηκαν τα παραγόμενα αποτελέσματα σε xlsx. αρχείο έγινε επεξεργασία και τα τελικά αποτελέσματα για όλες τις χρονοσειρές είναι τα εξής: σε σύνολο 209 χρονοσειρών στις 47 επικρατεί η κατανομή Generalized Gamma και στις 162 επικρατεί η κατανομή PBF. Συγκεκριμένα ο πίνακας με τα ποσοστά επιτυχίας κάθε κατανομής για τις πρώτες 30 χρονοσειρές απεικονίζεται παρακάτω:

Πίνακας 5-9: Ποσοστά επιτυχίας κατανομών PBF, Generalized Gamma ανά χρονοσειρά

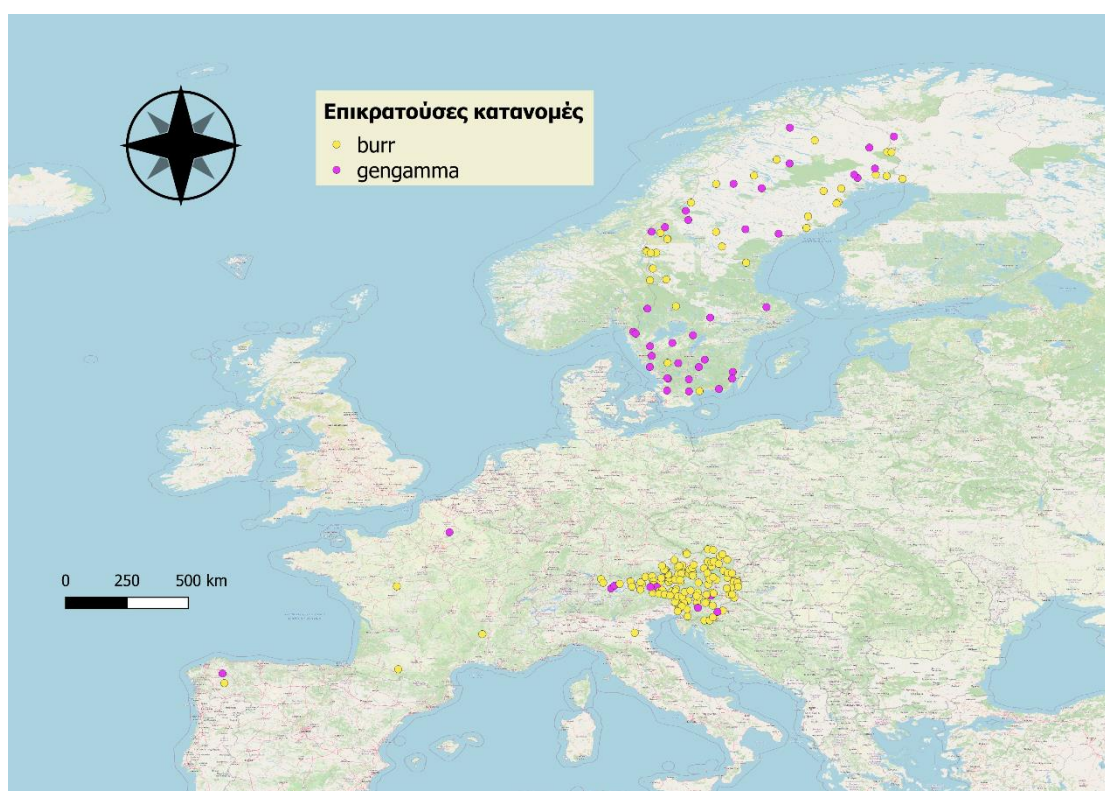
Ποσοστά Επιτυχίας κάθε κατανομής ανά χρονοσειρά			
Time Series	gengamma	burr	winner
1	83.33	16.67	gengamma
2	50.00	50.00	gengamma
3	75.00	25.00	gengamma
4	33.33	66.67	burr
5	100.00	0.00	gengamma
6	58.33	41.67	gengamma
7	33.33	66.67	burr
8	41.67	58.33	burr
9	58.33	41.67	gengamma

10	8.33	91.67	burr
11	100.00	0.00	gengamma
12	91.67	8.33	gengamma
13	100.00	0.00	gengamma
14	83.33	16.67	gengamma
15	91.67	8.33	gengamma
16	100.00	0.00	gengamma
17	75.00	25.00	gengamma
18	50.00	50.00	gengamma
19	41.67	58.33	burr
20	100.00	0.00	gengamma
21	50.00	50.00	gengamma
22	33.33	66.67	burr
23	41.67	58.33	burr
24	41.67	58.33	burr
25	50.00	50.00	gengamma
26	83.33	16.67	gengamma
27	75.00	25.00	gengamma
28	83.33	16.67	gengamma
29	66.67	33.33	gengamma
30	33.33	66.67	burr



Σχήμα 5.10: Αριθμός σταθμών που κυριαρχούν οι κατανομές PBF, Generalized Gamma

Ακόμη μέσω του ελεύθερου λογισμικού QGIS δημιουργήσαμε έναν χάρτη όπου φαίνεται και η γεωγραφική κατανομή των σταθμών και ποια κατανομή επικρατεί στις περισσότερες χρονοσειρές ώστε να έχουμε και ένα εποπτικό αποτέλεσμα. Παρατηρούμε ότι όπως στο Σχήμα 5.4 όπου ήταν ολοφάνερη η κυριαρχία της κατανομής lognormal σε σύγκριση με τις κατανομές gamma και weibull, έτσι κι εδώ είναι ξεκάθαρη η επικράτηση της κατανομής PBF έναντι της κατανομής Generalized Gamma. Συγκεκριμένα υπερτερεί στην Κεντρική Ευρώπη, στους λίγους σταθμούς της Ιταλίας και Γαλλίας ενώ στην Βόρεια Ευρώπη υπερτερεί στην εδαφική επικράτεια της Βόρειας Σουηδίας. Αντίθετα η κατανομή Generalized Gamma έχει μεγάλα ποσοστά επιτυχίας στην Βόρεια Ευρώπη και σχεδόν μηδαμινά στην Κεντρική και Δυτική.

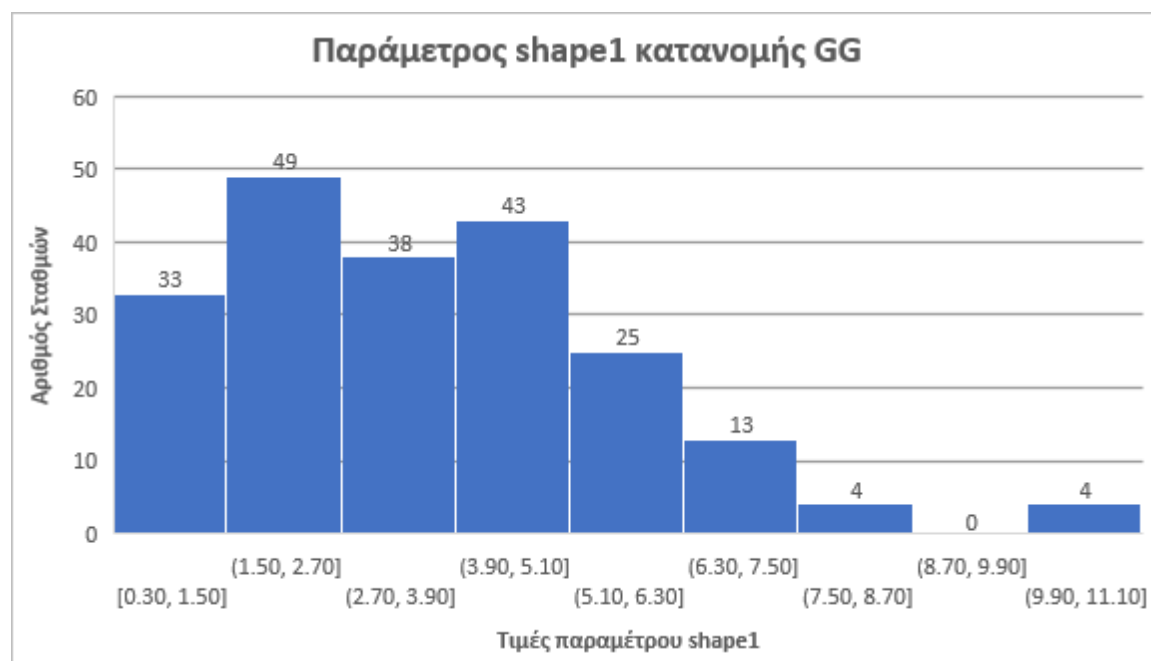


Σχήμα 5.11: Επικρατούσες κατανομές στους σταθμούς (QGIS) - MSE

5.5 Εκτίμηση παραμέτρων κατανομών PBF και Generalized Gamma

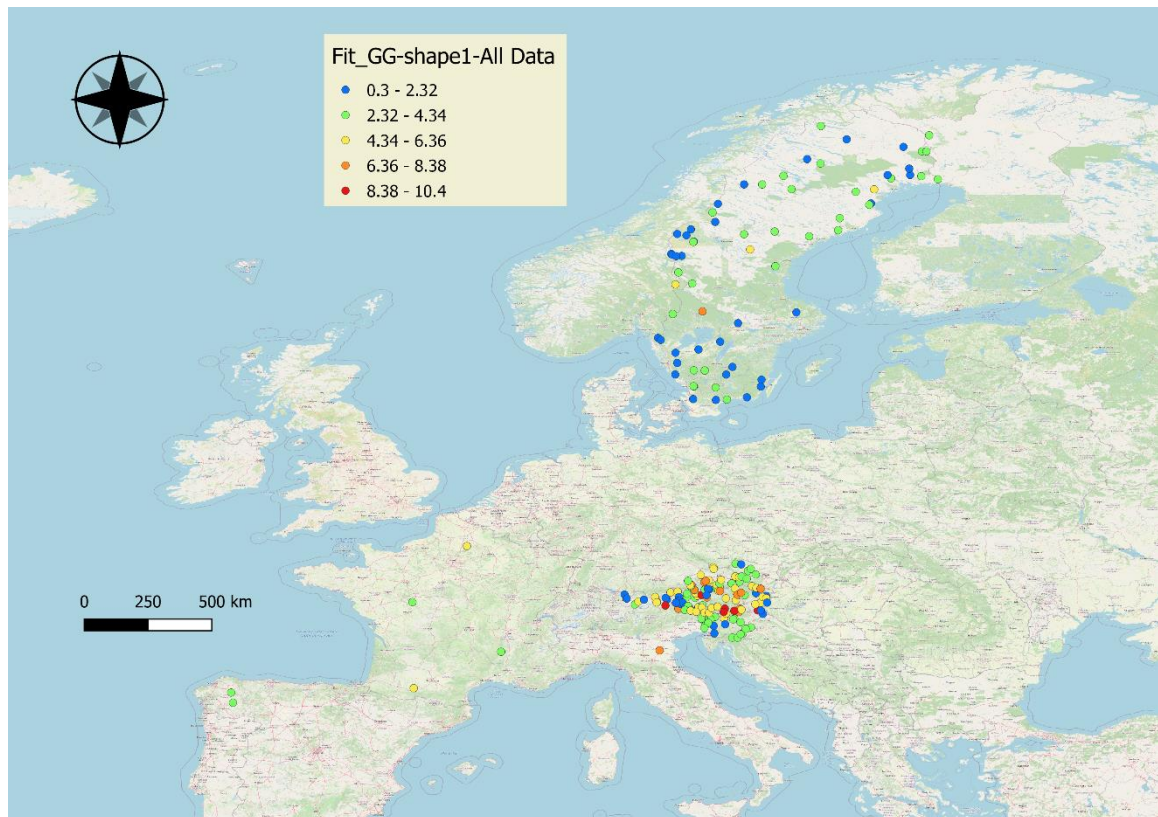
Από την ανάλυση στη γλώσσα προγραμματισμού R με τη χρήση της μεθόδου των L ροπών εκτιμήθηκαν οι παράμετροι όλων των παραπάνω κατανομών. Τα αποτελέσματα των παραμέτρων σχήματος για τις 2 κατανομές που προσαρμόζονται καλύτερα στα δεδομένα των ημερήσιων παροχών δηλαδή για την κατανομή PBF και την κατανομή Generalized Gamma, τόσο σε μορφή ιστογράμματος όσο και σε χάρτη χωρικής κατανομής όπου φαίνεται που βρίσκονται οι υψηλότερες και που οι χαμηλότερες τιμές των παραμέτρων, παρατίθενται παρακάτω.

Αρχικά για την κατανομή Generalized Gamma, που είναι μια κατανομή τριών παραμέτρων με 2 παραμέτρους σχήματος οι οποίες μας ενδιαφέρουν περισσότερο για την προσαρμογή στα δεδομένα, τα αποτελέσματα για τις παραμέτρους shape1 και shape2 ολόκληρης της χρονοσειράς είναι τα εξής:



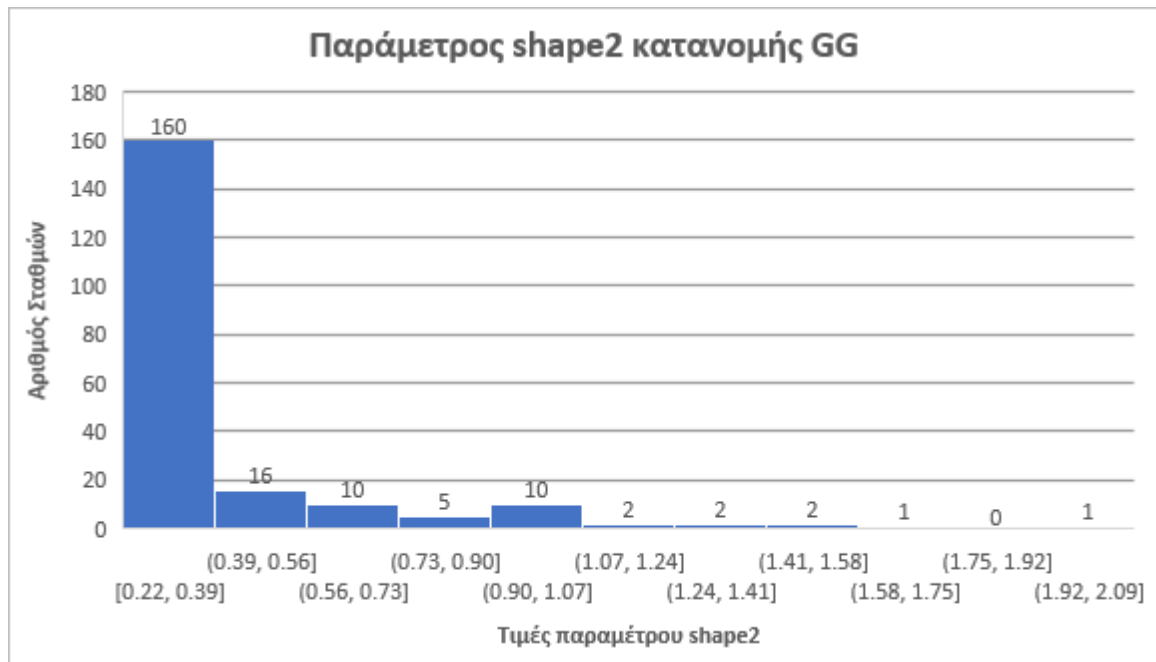
Σχήμα 5.12: Ιστόγραμμα παραμέτρου shape1 κατανομής Generalized Gamma

Από το ιστόγραμμα παρατηρείται ότι σε 82 σταθμούς οι τιμές της παραμέτρου shape1 είναι χαμηλές, σε 38 σταθμούς οι τιμές βρίσκονται σε μια μέση κατάσταση και σε 89 σταθμούς οι τιμές της παραμέτρου είναι υψηλές.



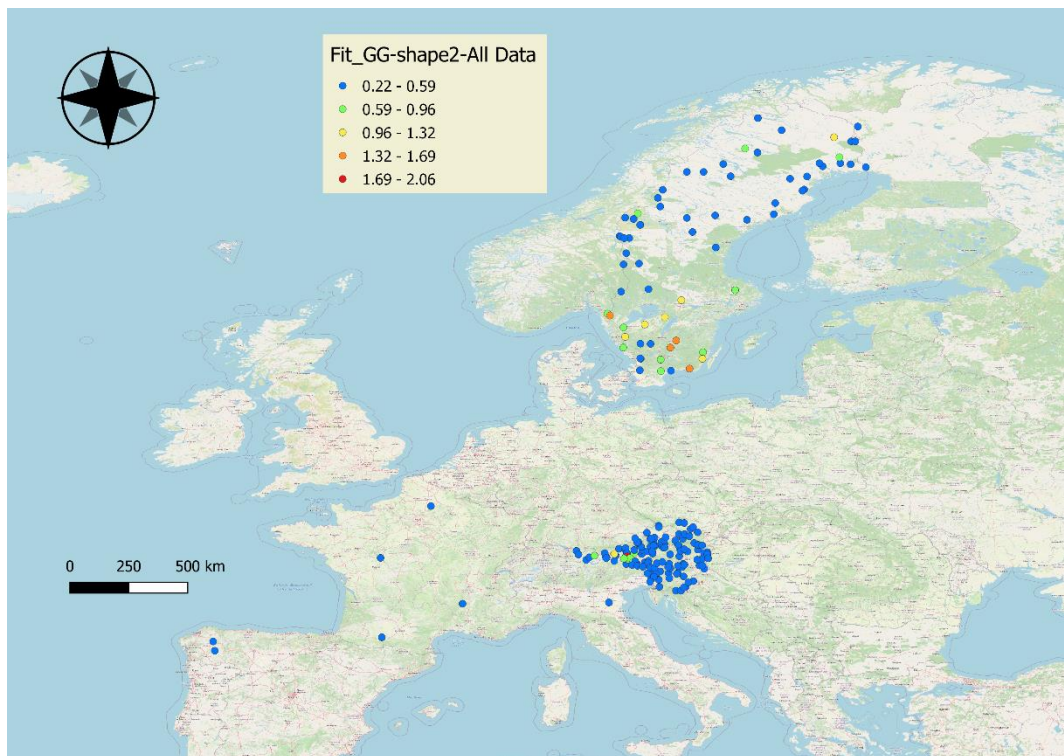
Σχήμα 5.13: Γεωγραφική κατανομή τιμών παραμέτρου shape1 κατανομής Generalized Gamma

Από το χάρτη χωρικής κατανομής παρατηρείται ότι οι υψηλότερες τιμές της παραμέτρου shape1 εμφανίζονται σε περιοχές στην Κεντρική Ευρώπη και στην Ιταλία, στην Ισπανία έχει χαμηλές τιμές η παράμετρος shape1 ενώ στην Βόρεια Ευρώπη στις περισσότερες περιοχές έχει χαμηλές τιμές η παράμετρος shape1 και λίγες περιοχές όπου οι τιμές θεωρούνται μέσης κατάστασης.



Σχήμα 5.14: Ιστόγραμμα παραμέτρου shape2 κατανομής Generalized Gamma

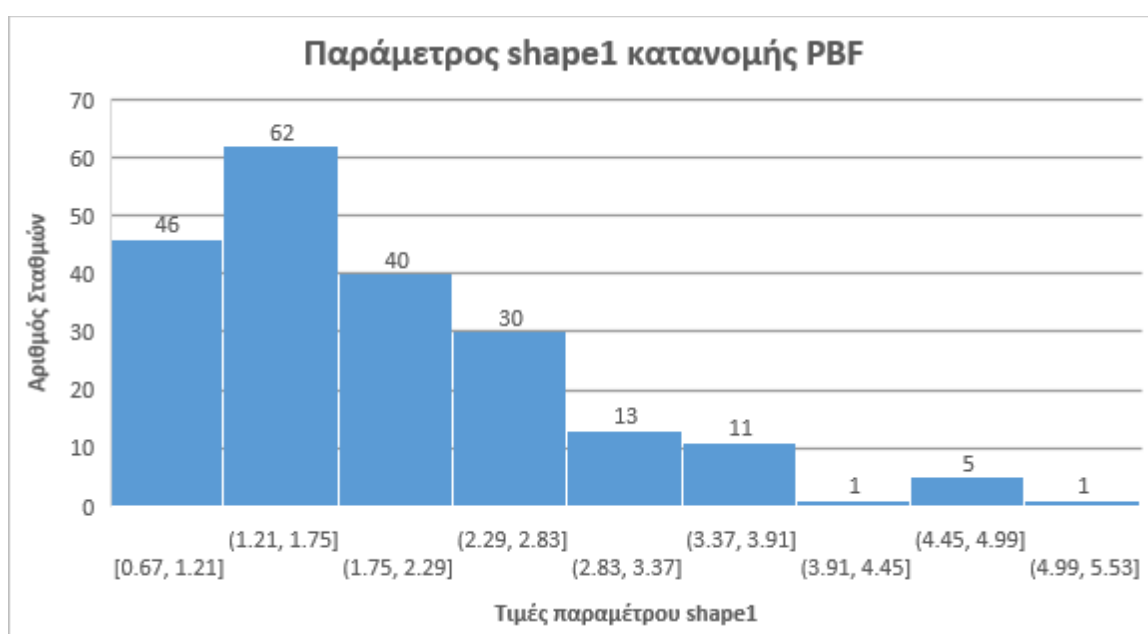
Από το ιστόγραμμα παρατηρείται ότι σε 160 σταθμούς οι τιμές της παραμέτρου shape1 είναι χαμηλές και μέσης κατάστασης ενώ σε 49 σταθμούς οι τιμές της παραμέτρου είναι υψηλές.



Σχήμα 5.15: Γεωγραφική κατανομή τιμών παραμέτρου shape2 κατανομής Generalized Gamma

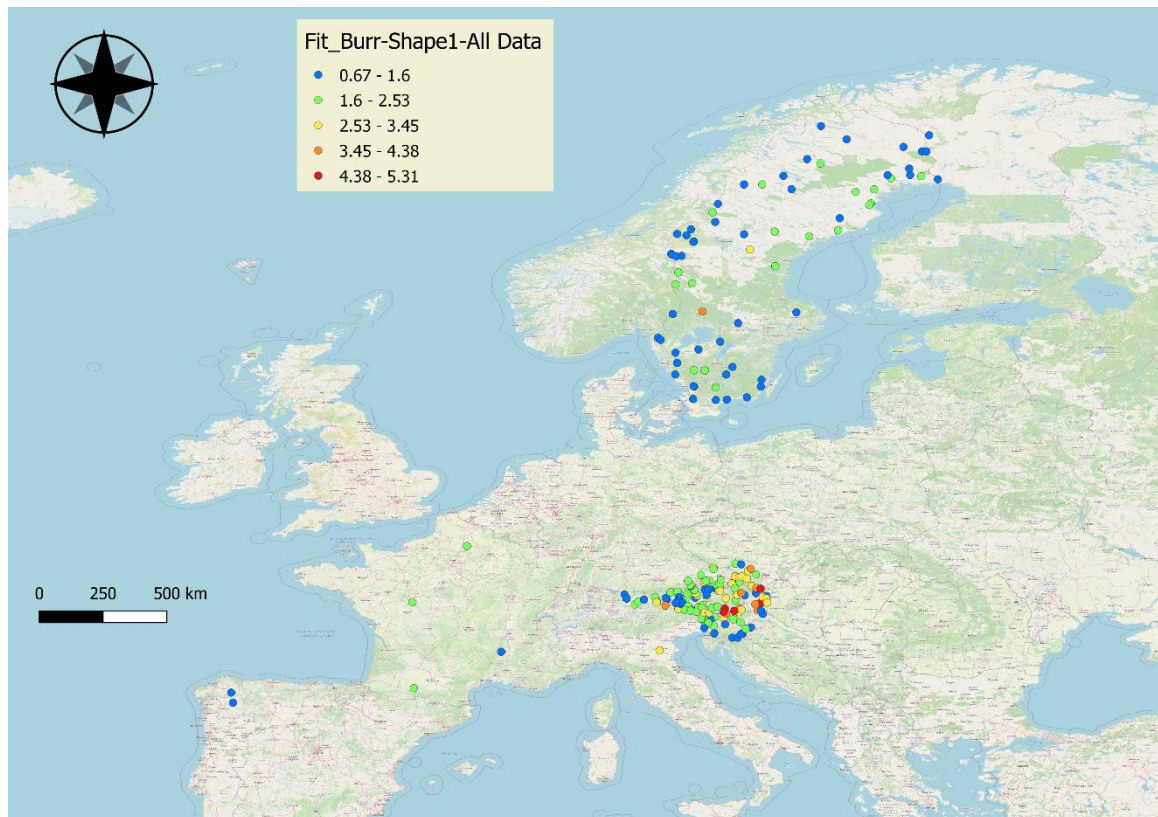
Από το χάρτη χωρικής κατανομής παρατηρείται ότι οι υψηλότερες τιμές της παραμέτρου *shape2* εμφανίζονται σε περιοχές στην Βόρεια Ευρώπη, στην Ισπανία, στην Ιταλία και στη Γαλλία έχει χαμηλές τιμές η παράμετρος *shape2* ενώ στην Κεντρική Ευρώπη έχει αρκετές περιοχές με χαμηλές τιμές της παραμέτρου *shape2*, κάποιες όπου οι τιμές θεωρούνται μέσης κατάστασης και λιγότερες με υψηλές τιμές.

Όσο αναφορά την κατανομή PBF που είναι μια κατανομή τριών παραμέτρων με 2 παραμέτρους σχήματος *shape1* και *shape2*, παρουσιάζει ιδιαίτερο ενδιαφέρον καθώς η 2^η παράμετρος σχήματος *shape2* που ονομάζεται και δείκτης ουράς όταν είναι μεγαλύτερη ή ίση με 0.5 τότε η 2^η ροπή είναι άπειρη. Γι' αυτό το λόγο όσες τιμές η 2^η παράμετρος *shape2* είναι μεγαλύτερη ή ίση με 0.5 τις αφαιρούμε.



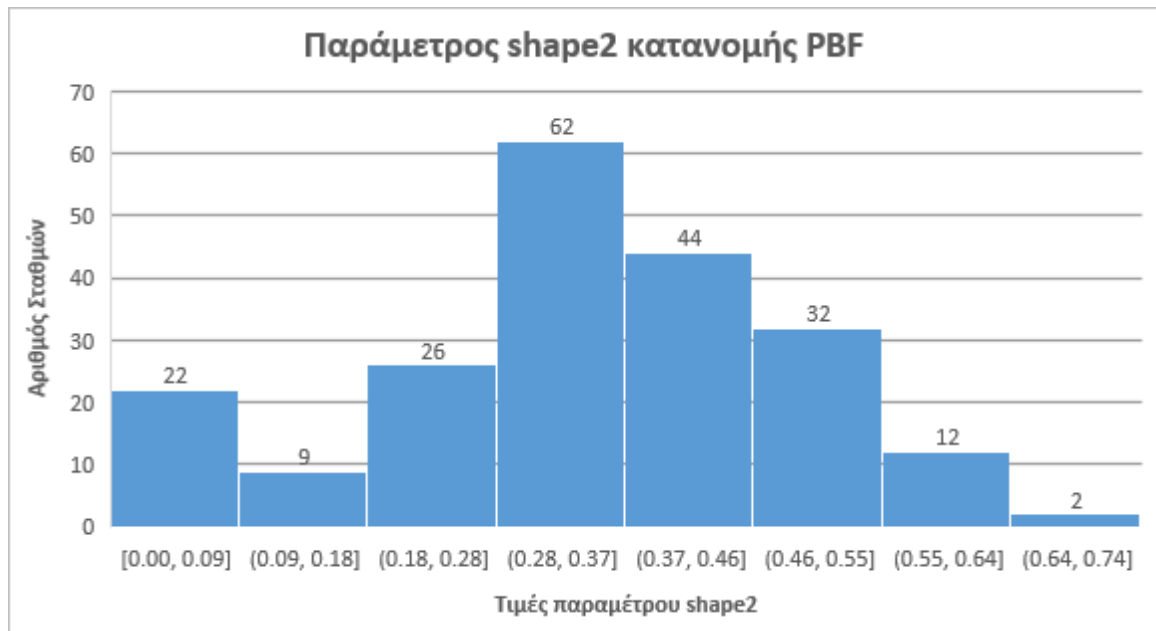
Σχήμα 5.16: Ιστόγραμμα παραμέτρου *shape1* κατανομής PBF

Από το ιστόγραμμα παρατηρείται ότι σε 102 σταθμούς οι τιμές της παραμέτρου *shape1* είναι χαμηλές και σε 101 σταθμούς οι περισσότερες τιμές είναι υψηλές ενώ λίγες τιμές βρίσκονται σε μέση κατάσταση.



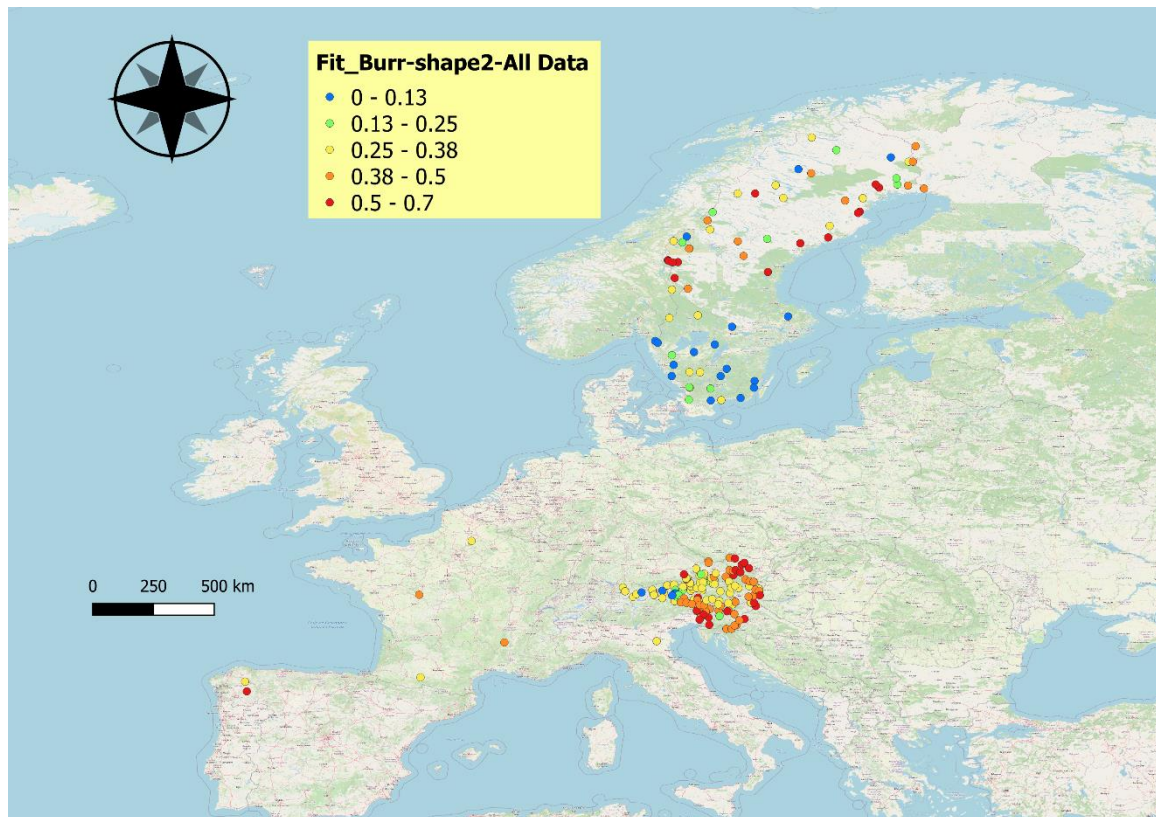
Σχήμα 5.17: Γεωγραφική κατανομή τιμών παραμέτρου shape1 κατανομής PBF

Από το χάρτη χωρικής κατανομής παρατηρείται ότι οι υψηλότερες τιμές της παραμέτρου shape1 εμφανίζονται σε περιοχές στην Κεντρική Ευρώπη και στην Ιταλία, στην Ισπανία έχει χαμηλές τιμές η παράμετρος shape1 ενώ στην Βόρεια Ευρώπη έχει αρκετές περιοχές με χαμηλές τιμές της παραμέτρου shape1.



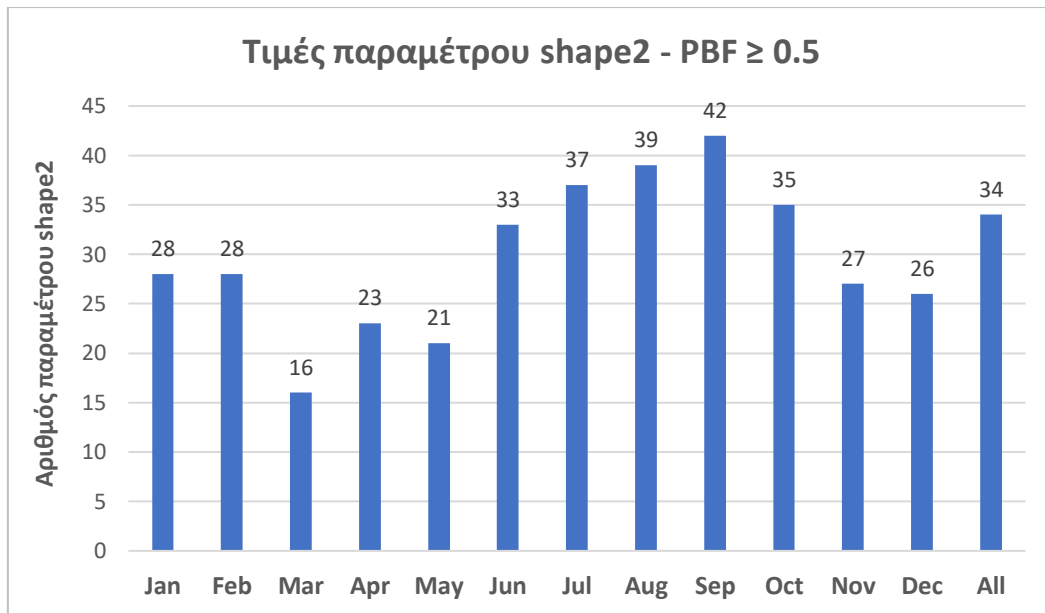
Σχήμα 5.18: Ιστόγραμμα παραμέτρου shape2 κατανομής PBF

Από το ιστόγραμμα παρατηρείται ότι σε 102 σταθμούς οι τιμές της παραμέτρου shape1 είναι χαμηλές και σε 101 σταθμούς οι περισσότερες τιμές είναι υψηλές ενώ λίγες τιμές βρίσκονται σε μέση κατάσταση.



Σχήμα 5.19: Γεωγραφική κατανομή τιμών παραμέτρου shape2 κατανομής PBF

Από το χάρτη χωρικής κατανομής παρατηρείται ότι οι υψηλότερες τιμές της παραμέτρου shape2 εμφανίζονται σε περιοχές στην Κεντρική Ευρώπη και στην Ισπανία, στην Ιταλία έχει χαμηλές τιμές η παράμετρος shape2 ενώ στην Βόρεια Ευρώπη έχει αρκετές περιοχές με χαμηλές τιμές, αλλά και κάποιες υψηλές τιμές της παραμέτρου shape1. Ωστόσο όπως αναφέραμε παραπάνω οι τιμές της shape2 από 0.5 και πάνω δημιουργεί ροπή άπειρη οπότε τις απορρίπτουμε και δεν τις λαμβάνουμε υπ' όψιν. Οι τιμές που είναι μεγαλύτερες ή ίσες με 0.5 είναι 34 για ολόκληρη τη χρονοσειρά και βρίσκονται στην Βόρεια και Κεντρική Ευρώπη. Παρακάτω απεικονίζεται το διάγραμμα με τις τιμές της παραμέτρου shape2 που είναι μεγαλύτερες ή ίσες με 0.5 κάθε μήνα για όλες τις χρονοσειρές αλλά και για ολόκληρη την χρονοσειρά:



Σχήμα 5.20: Τιμές της παραμέτρου shape2 της κατανομής PBF που είναι μεγαλύτερες ή ίσες με 0.5

6 ΔΙΑΓΡΑΜΜΑΤΑ ΛΟΓΟΥ ΡΟΠΗΣ

6.1 Γενικά Στοιχεία

Μια ευρέως χρησιμοποιούμενη προσέγγιση για την προκαταρκτική διάκριση των πιο συνεπών κατανομών δηλαδή για να περιγράψει την καταλληλότητα των κατανομών είναι τα λεγόμενα «διαγράμματα λόγου ροπής» (“moment-ratio diagrams”). Αυτός ο τύπος ανάλυσης έχει εφαρμοστεί ευρέως στην πιθανολογική μελέτη υδρολογικών μεταβλητών, ειδικά σε περιπτώσεις όπου εξετάζεται μεγάλος αριθμός αρχείων, καθώς παρέχει μια γρήγορη και άμεση αξιολόγηση της καταλληλότητας πολλών κατανομών από στατιστική άποψη. Τα διαγράμματα λόγου ροπής παρέχουν μια γραφική σύγκριση μεταξύ των στατιστικών δειγμάτων, που υπολογίζονται από τα δεδομένα της βάσης και των θεωρητικών, όπως δίδονται από τους τύπους των υποψήφιων παραμετρικών κατανομών. Η εγγύτητα των δειγμάτων στατιστικών στον θεωρητικό τόπο (δηλ. Σημείο, καμπύλη ή περιοχή) μιας συγκεκριμένης κατανομής παρέχει μια ένδειξη της καταλληλότητας αυτής της κατανομής για να περιγράψει καλύτερα τις καταγραφές δεδομένων, η οποία υπονοεί εάν υπάρχει ένα συγκεκριμένο σύνολο παραμέτρου κατανομής έτσι ώστε να αναπαράγει αυτές τις καταγραφές.

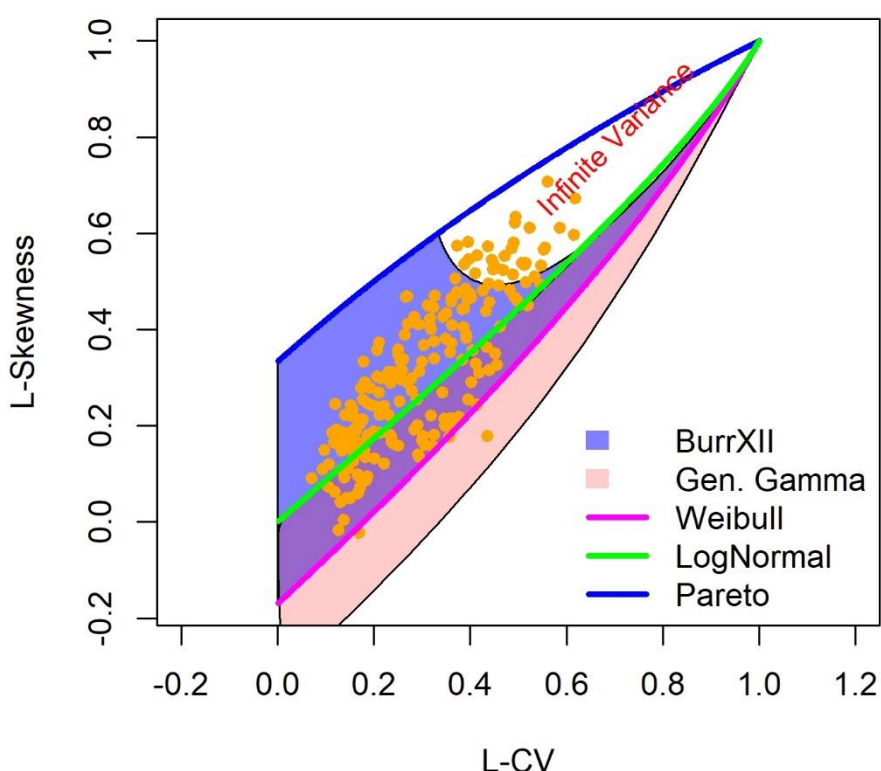
Σε τέτοια διαγράμματα, τυπικά απεικονίζεται ένας συντελεστής μεταβλητότητας έναντι ενός συντελεστή ασυμμετρίας ή ένας συντελεστής ασυμμετρίας έναντι ενός συντελεστή κυρτώσεως ενώ θα μπορούσε επίσης να χρησιμοποιηθεί οποιοδήποτε ζευγάρι τυποποιημένων ροπών. Μια ολοκληρωμένη ανασκόπηση των συμβατικών διαγραμμάτων αναλογίας ροπής προϊόντος που συνοψίζουν 37 θεωρητικά πιθανολογικά μοντέλα μπορεί να βρεθεί στο Vargo και Leemis. Τις τελευταίες δεκαετίες, τα διαγράμματα αναλογίας L-ροπών, τα οποία εισήχθησαν για πρώτη φορά από τον Hosking, έχουν αποκτήσει δημοτικότητα σε σχέση με τα συμβατικά και έχουν χρησιμοποιηθεί ευρέως σε πολλές υδρολογικές εφαρμογές λόγω των καλύτερων στατιστικών ιδιοτήτων των L ροπών. Τα ζεύγη L-μεταβλητότητας έναντι L-ασυμμετρίας και L-ασυμμετρίας έναντι L-κύρτωση χρησιμοποιούνται κυρίως, ενώ το τελευταίο είναι πιο δημοφιλές αφού η διακύμανση L δεν είναι καλά καθορισμένη για κατανομές με μέση τιμή ίση με το μηδέν ή είναι αρνητική. Ο θεωρητικός τόπος κατανομής σε διαγράμματα λόγου L καθώς και σε συμβατικά διαγράμματα ροπών μπορεί να είναι ένα σημείο, μια καμπύλη ή μια περιοχή ανάλογα με τον αριθμό των παραμέτρων της κατανομής. Στο διάγραμμα L-skewness έναντι L-kurtosis, ο τύπος του τόπου εξαρτάται από τον αριθμό των παραμέτρων σχήματος, δηλαδή οι κατανομές χωρίς παράμετρο σχήματος σχηματίζουν ένα σημείο (π.χ. Κανονική κατανομή) με μία παράμετρο σχήματος να σχηματίζει μια καμπύλη (π.χ. Κατανομή

γάμμα) και με δύο παραμέτρους σχήματος που σχηματίζουν μια περιοχή (π.χ. Burr Type XII). Αντίθετα, ο θεωρητικός τύπος κατανομών σε ένα διάγραμμα L-μεταβλητότητας έναντι L-ασυμμετρίας εξαρτάται επίσης από τις παραμέτρους τοποθεσίας και κλίμακας (εάν είναι προσαρμοσμένες ώστε να λαμβάνουν οποιαδήποτε τιμή) και στη συνέχεια τον αριθμό των παραμέτρων σχήματος.

6.2 Διαγράμματα L-skewness LCV

Η παρούσα ανάλυση βασίστηκε στο διάγραμμα του λόγου L-ασυμμετρία έναντι του συντελεστή μεταβλητότητας L για θετικές τυχαίες μεταβλητές. Πρώτον, οι μελετημένες κατανομές πρέπει να ορίζονται στον πραγματικό θετικό άξονα. Επιπλέον, οι υψηλοί συντελεστές L ασυμμετρίας και L μεταβλητότητας μαζί με τη μεγάλη μεταβλητότητα αυτών των χαρακτηριστικών σχήματος από ποταμό σε ποταμό και από ημέρα σε ημέρα συνεπάγονται τη χρήση κατανομών των οποίων οι παράμετροι σχήματος υποστηρίζουν ένα εύρος τιμών για αυτές τις στατιστικές. Η ανάλυση πραγματοποιήθηκε για ξεχωριστά για κάθε μήνα των χρονοσειρών όπως και για όλες τις χρονοσειρές. Χαρακτηριστικό παράδειγμα όπου φαίνονται κατά πόσο το ζεύγος L-skewness, L-CV των δεδομένων «πέφτει» μέσα στον αντίστοιχο L-skewness, L-CV χώρο της κάθε κατανομής φαίνεται στο παρακάτω σχήμα που αφορά τον Ιανουάριο όλων των χρονοσειρών.

Data: Jan



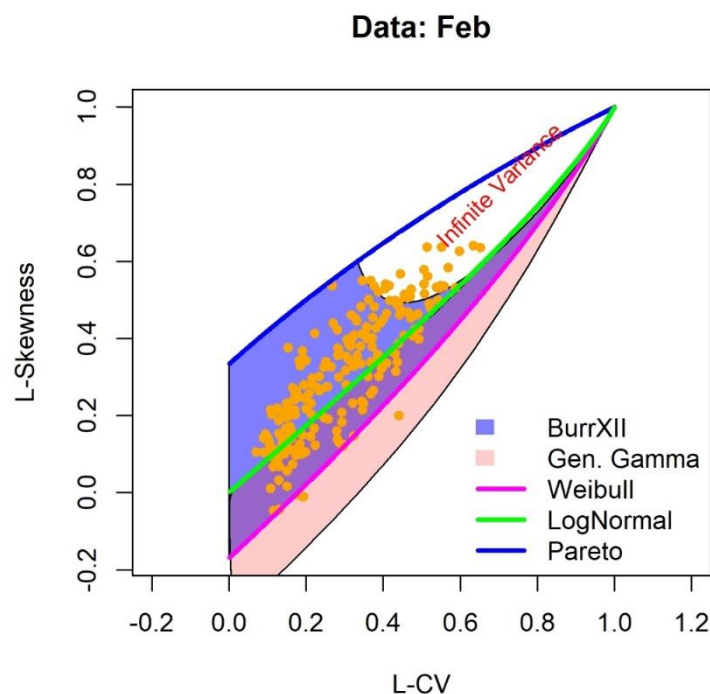
Σχήμα 6.1: Διάγραμμα L-skewness, L-CV για το μήνα Ιανουάριο

Όπως βλέπουμε στο διάγραμμα L-ασυμετρίας έναντι L-μεταβλητότητας του Σχήματος 6.1, ο τόπος όλων των κατανομών ελέγχου είναι μια καμπύλη εκτός από αυτόν της PBF κατανομής και της κατανομής Generalized Gamma που είναι μια περιοχή και όχι σημείο ή καμπύλη. Η κατανομή gamma δεν φαίνεται αλλά εμπεριέχεται στην Generalized Gamma ενώ η περιοχή της Generalized Gamma είναι από την καμπύλη της lognormal (GG για $\lambda=0$) και κάτω. Ακόμη παρατηρούμε ότι στο χώρο της κατανομής PBF υπάρχει ένα άσπρο κομμάτι που έχει άπειρη διασπορά. Αυτό οφείλεται σε μια ιδιομορφία της κατανομής PBF η οποία έχει μια παράμετρο shape 2 καλούμενη και ως δείκτης ουράς (tail index) που όταν είναι μεγαλύτερη ή ίση με 0.5 δεν υπάρχει η 2^η ροπή, είναι άπειρη. Όταν σε μια τυχαία μεταβλητή η κατανομή που της ορίζω έχει άπειρη διασπορά αυτό σημαίνει αυτόματα ότι δεν μπορείς να ορίσεις και άλλα μεγέθη όπως η αυτοσυσχέτιση και επομένως δημιουργεί πολλά προβλήματα. Επιπλέον σύμφωνα με το paper "Entropy Production in Stochastics", Demetris Koutsoyiannis 2017 σε περίπτωση που σε μια στοχαστική ανέλιξη προκύψει άπειρη διασπορά, αυτό συνεπάγεται άπειρη ενέργεια πράγμα που

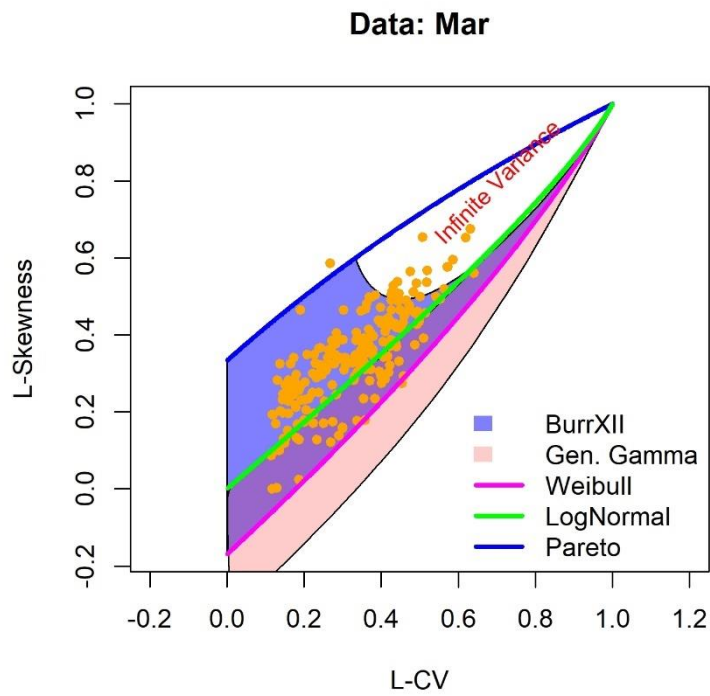
είναι ανέφικτο στη φύση. Συμπερασματικά λοιπόν δεν υπάρχει περίπτωση η διασπορά (variance) να είναι άπειρη.

Το άσπρο κομμάτι θα μπορούσε να μοντελοποιηθεί από την κατανομή PBF με μια παράμετρο που θα έδινε άπειρη διασπορά ωστόσο αυτό δεν έγινε. Επομένως όσα σημεία πέφτουν στο χώρο της άπειρης διασποράς δεν μετριοούνται σαν ποσοστό επιτυχίας της κατανομής PBF αλλά αφαιρούνται. Πρέπει επίσης να τονισθεί ότι η PBF είναι η μοναδική από τις εξεταζόμενες κατανομές που μπορεί να δώσει άπειρες ροπές, όλες οι υπόλοιπες δίνουν πεπερασμένες ροπές.

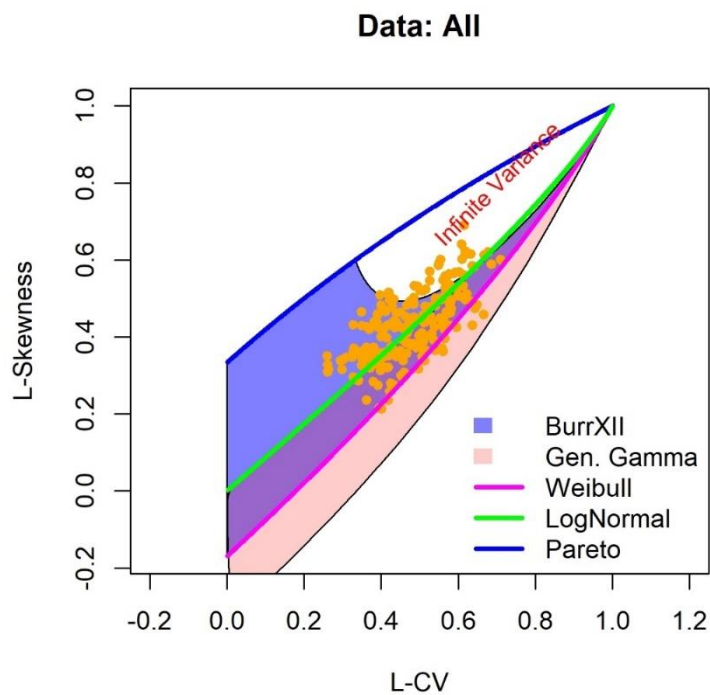
Παρακάτω απεικονίζονται τα διαγράμματα λόγου ροπής και για τις 5 κατανομές για τον μήνα Φεβρουάριο, τον μήνα Μάρτιο και για όλους τους μήνες όλων των χρονοσειρών :



Σχήμα 6.2: Διάγραμμα L-skewness, L-CV για το μήνα Φεβρουάριο



Σχήμα 6.3: Διάγραμμα L-skewness, L-CV για το μήνα Μάρτιο

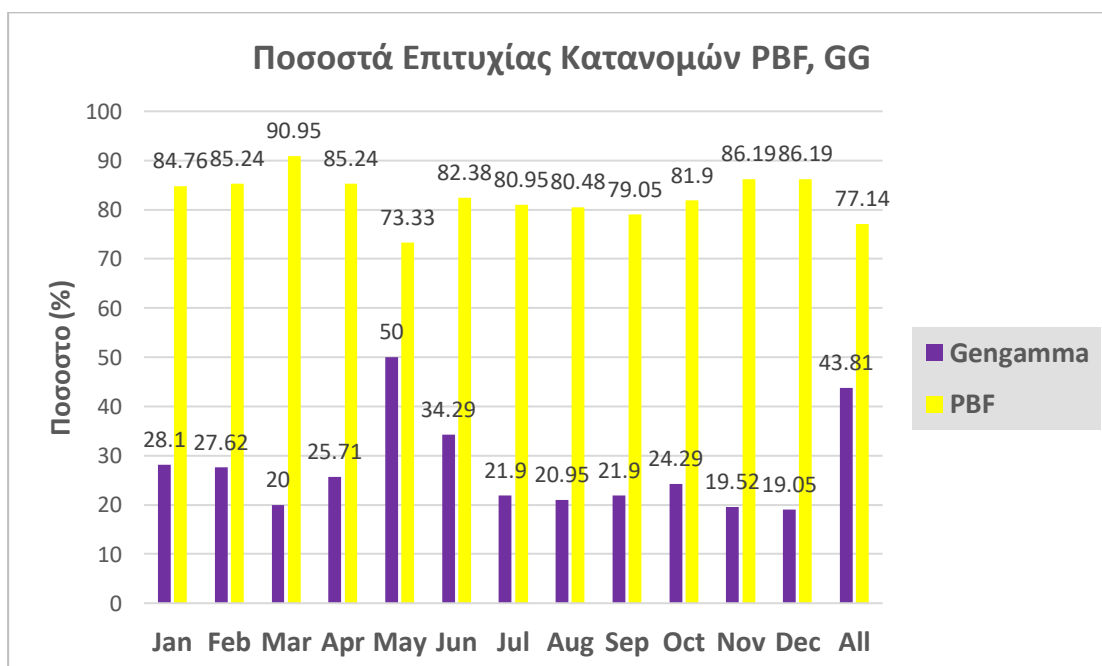


Σχήμα 6.4: Διάγραμμα L-skewness, L-CV για όλους τους μήνες

Παρατηρείτε ότι σε όλα τα διαγράμματα τα περισσότερα σημεία των ζευγών L-skewness, L-CV των δεδομένων πέφτουν είτε στην περιοχή της κατανομής PBF είτε στην περιοχή της κατανομής Generalized Gamma. Συγκεκριμένα τα ποσοστά επιτυχίας της κάθε κατανομής PBF και Generalized Gamma φαίνονται στον παρακάτω Πίνακα:

Πίνακας 6-1: Ποσοστά επιτυχίας κατανομών PBF, Generalized Gamma ανά μήνα για όλες τις χρονοσειρές

GG_in	Br_in	type
28.10	84.76	Jan
27.62	85.24	Feb
20.00	90.95	Mar
25.71	85.24	Apr
50.00	73.33	May
34.29	82.38	Jun
21.90	80.95	Jul
20.95	80.48	Aug
21.90	79.05	Sep
24.29	81.90	Oct
19.52	86.19	Nov
19.05	86.19	Dec
43.81	77.14	all



Σχήμα 6.5: Ποσοστά Επιτυχίας Κατανομών PBF, GG

Όπως παρατηρούμε από το Σχήμα 6.5 τα ποσοστά των 2 κατανομών ανά μήνα αλλά και για ολόκληρη τη χρονοσειρά δεν αθροίζουν το 100% και αυτό συμβαίνει γιατί υπάρχει μια κοινή περιοχή για τις 2 κατανομές και συγκεκριμένα μεταξύ της πράσινης καμπύλης (lognormal) και της ροζ καμπύλης (weibull).

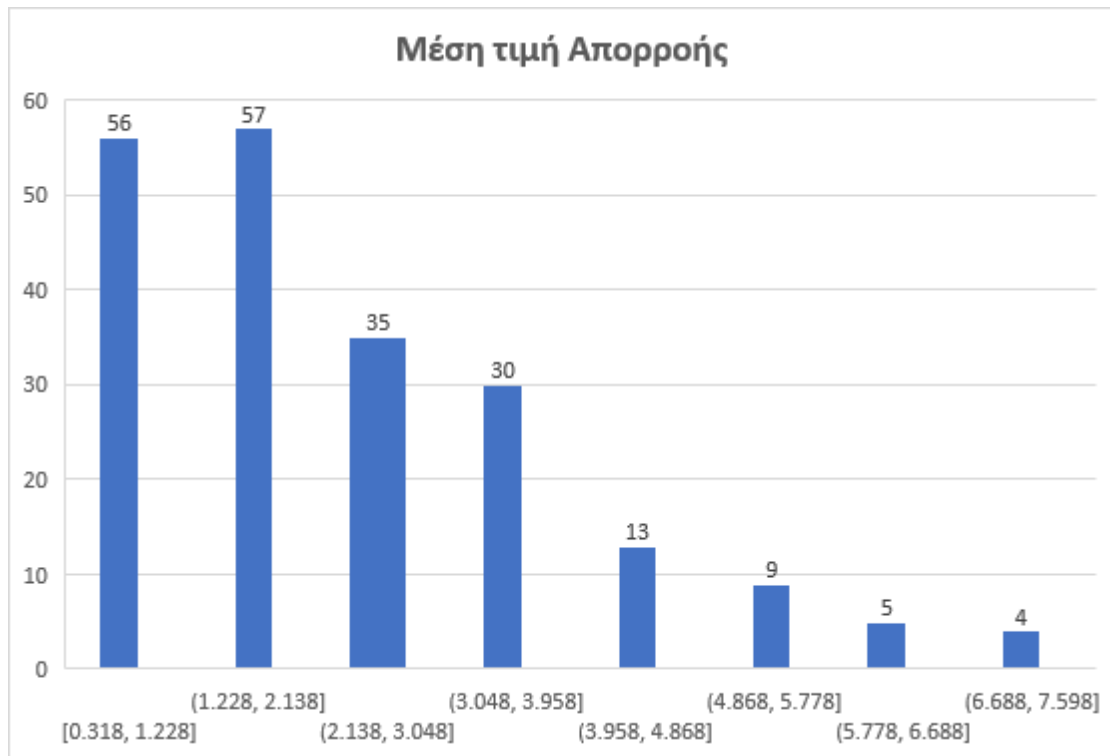
Σύμφωνα με τα διαγράμματα λόγου ροπής στην κλίμακα L (Lratio diagrams) η καταλληλότερη κατανομή για τα δεδομένα μας είναι η κατανομή PBF (Pareto-Burr-Feller) σε κάθε μήνα ξεχωριστά για όλες τις χρονοσειρές αλλά και για όλους τους μήνες μαζί όπως φαίνεται και από τα ποσοστά στο Σχήμα 6.5 σε σύγκριση με την κατανομή Generalized Gamma. Αυτό μάλιστα συμφωνεί και με τα αποτελέσματα με τη χρήση του Μέσου τετραγωνικού Σφάλματος όπου καταλληλότερη κρίθηκε πάλι η κατανομή PBF σε σύγκριση με την κατανομή Generalized Gamma.

7 ΧΩΡΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΣΤΑΤΙΣΤΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

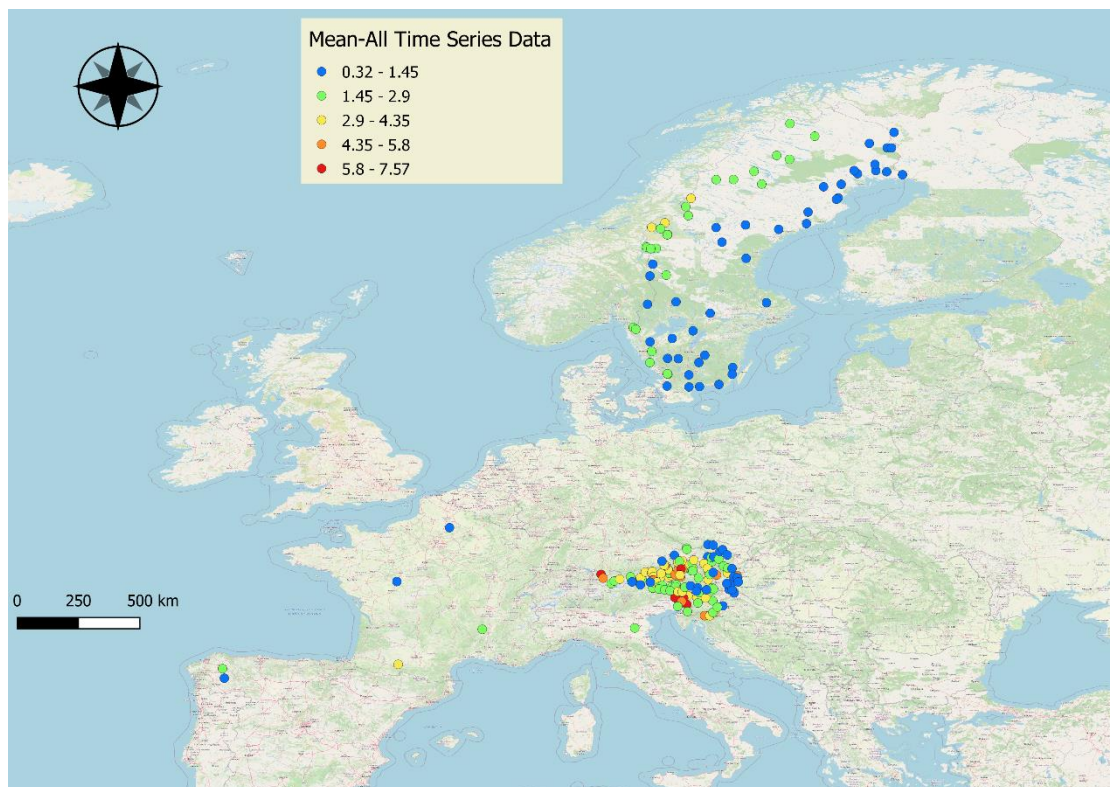
7.1 Διαγράμματα – Χάρτες χωρικής κατανομής

Στην παρούσα διπλωματική εργασία παρήχθησαν μέσω του ελεύθερου λογισμικού QGIS χάρτες χωρικής απεικόνισης των στατιστικών χαρακτηριστικών που υπολογίστηκαν από την ανάλυση των δεδομένων. Παρακάτω παρουσιάζονται διαγράμματα των στατιστικών μεγεθών που προέκυψαν από την ανάλυση με τους αντίστοιχους χάρτες χωρικής κατανομής μέσω των οποίων προέκυψαν συμπεράσματα για την ομαδοποίηση στατιστικών συμπεριφορών ανάλογα με τα κλιματικά χαρακτηριστικά κάθε περιοχής.

Τα αποτελέσματα της ανάλυσης είναι σε μορφή απορροής (mm). Όσο αναφορά την μέση τιμή ημερήσιας απορροής παρατηρείται ότι η μέγιστη τιμή είναι 7.57mm και η ελάχιστη 0.32mm. Ακόμα από το ισόγραμμα φαίνεται η έντονη διακύμανση της μέσης τιμής απορροής ανάμεσα στους σταθμούς πράγμα που οφείλεται στα ιδιαίτερα κλιματικά χαρακτηριστικά της περιοχής που βρίσκεται ο κάθε σταθμός. Έτσι όπως φαίνεται από τη γεωγραφική κατανομή των τιμών μέσης απορροής στο Σχήμα 7.2 υψηλές τιμές άνω των 5.8mm παρατηρούνται στην Κεντρική Ευρώπη και συγκεκριμένα σε 6 σταθμούς στη Σλοβενία και σε 3 στην Αυστρία ενώ στην Βόρεια Ευρώπη, στην Ιταλία, στην Γαλλία και στην Ισπανία παρατηρούνται χαμηλές τιμές απορροής κάτω των 3mm. Αυτό οφείλεται στο ότι στην Κεντρική Ευρώπη σύμφωνα με τον ενημερωμένο κλιματικό χάρτη Köppen – Geiger για την περίοδο 1951–2000 (Kottek et al., 2006) κυριαρχεί το καθεστώς Cfb, το οποίο αντιπροσωπεύει το θαλάσσιο κλίμα που χαρακτηρίζεται από χαμηλά σύννεφα, ομίχλη και ψιλή βροχή όλο το χρόνο που συνδέεται με τις υγρές θαλάσσιες αέριες μάζες. Δηλαδή υγρό κλίμα όλο το χρόνο με ήπιους χειμώνες και μεγάλα και δροσερά καλοκαίρια. Αντίθετα στην Βόρεια Ευρώπη επικρατεί το πολικό κλίμα και ειδικότερα η Τούνδρα (Et). Ο τύπος Et χαρακτηρίζεται από έδαφος που είναι μόνιμα παγωμένο, ενώ το καλοκαίρι όπου ο θερμός καιρός λειώνει τον πάγο που υπάρχει στο ανώτερο εδαφικό στρώμα, η τούνδρα γίνεται βαλτώδης και λασπώδης. Σε αυτό το κλίμα κυριαρχεί η πολύ μικρή εξάτμιση με αποτέλεσμα να επικρατεί μεγάλη υγρασία στην ατμόσφαιρα.

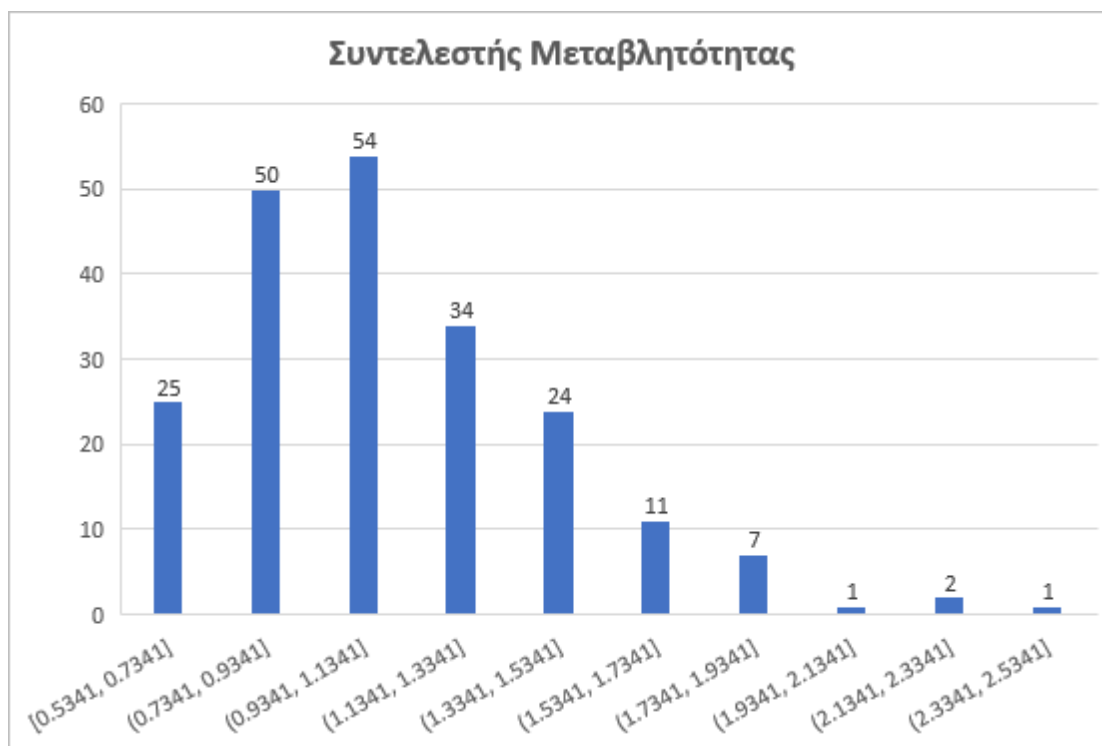


Σχήμα 7.1: Ιστόγραμμα Μέσης τιμής ημερήσιας Απορροής όλων των σταθμών για τις χρονοσειρές σε ετήσια κλίμακα

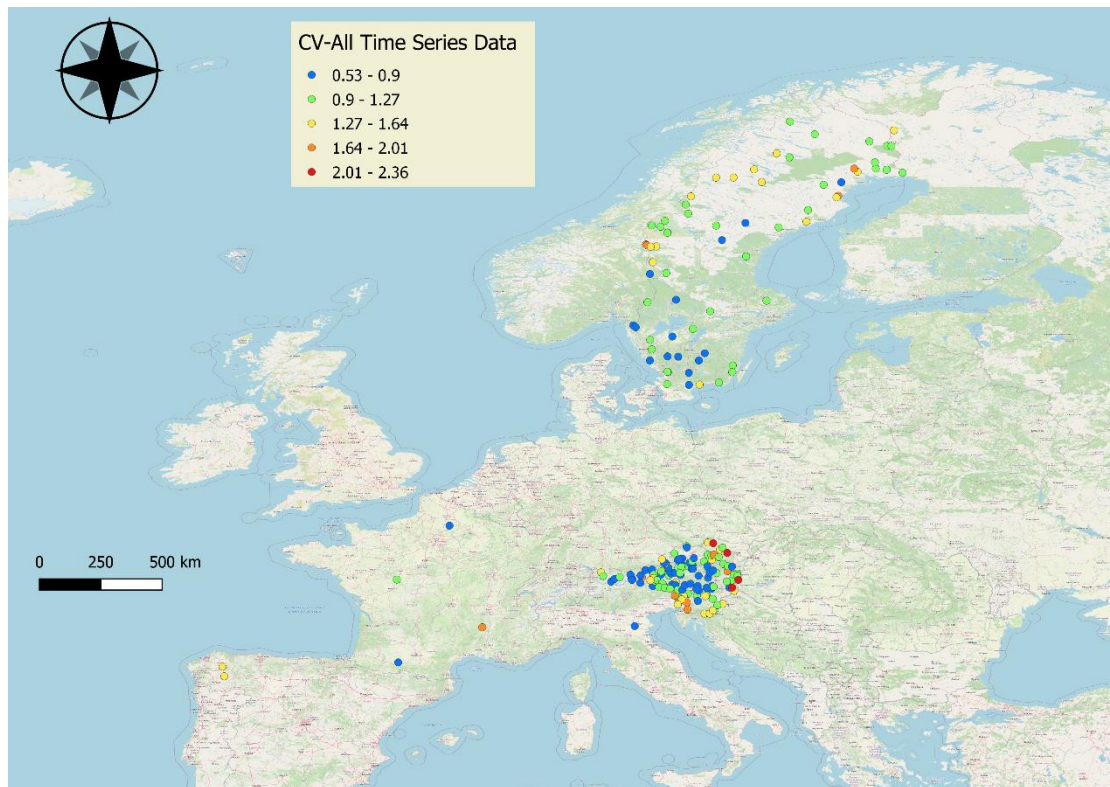


Σχήμα 7.2: Γεωγραφική κατανομή τιμών μέσης τιμής απορροής

Στη συνέχεια για τον Συντελεστή μεταβλητότητας παρατηρείται από το ιστόγραμμα ότι η μέγιστη τιμή του συντελεστή μεταβλητότητας είναι 2.36 και η ελάχιστη τιμή 0.53. Ο συντελεστής μεταβλητότητας που ισούται με την τυπική απόκλιση δια τη μέση τιμή, αποτελεί μέτρο αβεβαιότητας ως προς την εκτίμηση της ημερήσιας παροχής. Μεγάλη τιμή του συντελεστή μεταβλητότητας, σημαίνει μεγάλη διασπορά σε σχέση με τη μέση τιμή και συνεπάγεται μεγάλη αβεβαιότητα ως προς την εκτίμηση της μ_x . Από το ιστόγραμμα φαίνεται ότι αρκετοί σταθμοί έχουν συντελεστή μεταβλητότητας πάνω από το 1 και άρα οι τιμές της απορροής δεν είναι κοντά στον μέση τιμή που αναφέρθηκε παραπάνω. Συγκεκριμένα σε 119 σταθμούς ο συντελεστής μεταβλητότητας είναι μεγαλύτερος από τη μονάδα και άρα υπάρχει μεγάλη διασπορά σε σχέση με τη μέση τιμή και άρα μεγαλύτερη αβεβαιότητα ως προς την αξιοπιστία των αποτελεσμάτων. Όσο αφορά τη χωρική κατανομή του συντελεστή μεταβλητότητας παρατηρούνται μεγαλύτερες τιμές στην εδαφική επικράτεια της Αυστρίας και συγκεκριμένα κοντά στην Βιέννη αλλά και σε κάποιες περιοχές της Σλοβενίας, ενώ στην υπόλοιπη Κεντρική Ευρώπη αλλά και στη Βόρεια Ευρώπη παρατηρούνται χαμηλότερες τιμές.

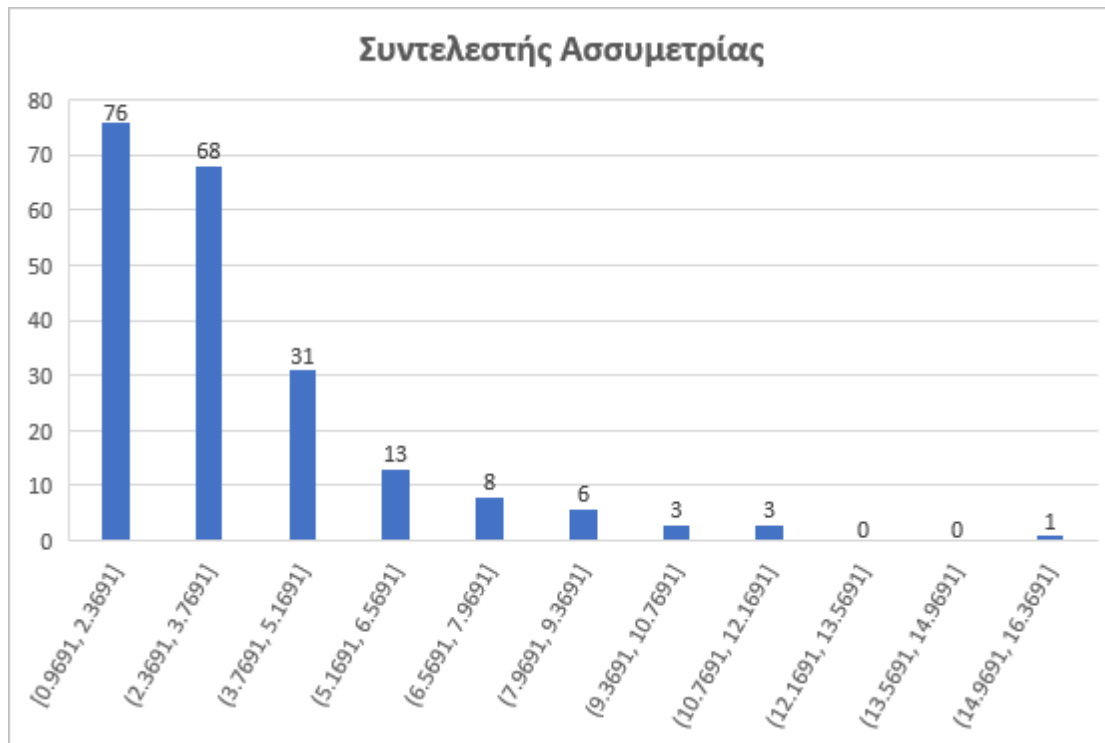


Σχήμα 7.3: Ιστόγραμμα Συντελεστή Μεταβλητότητας όλων των σταθμών για τις χρονοσειρές σε ετήσια κλίμακα

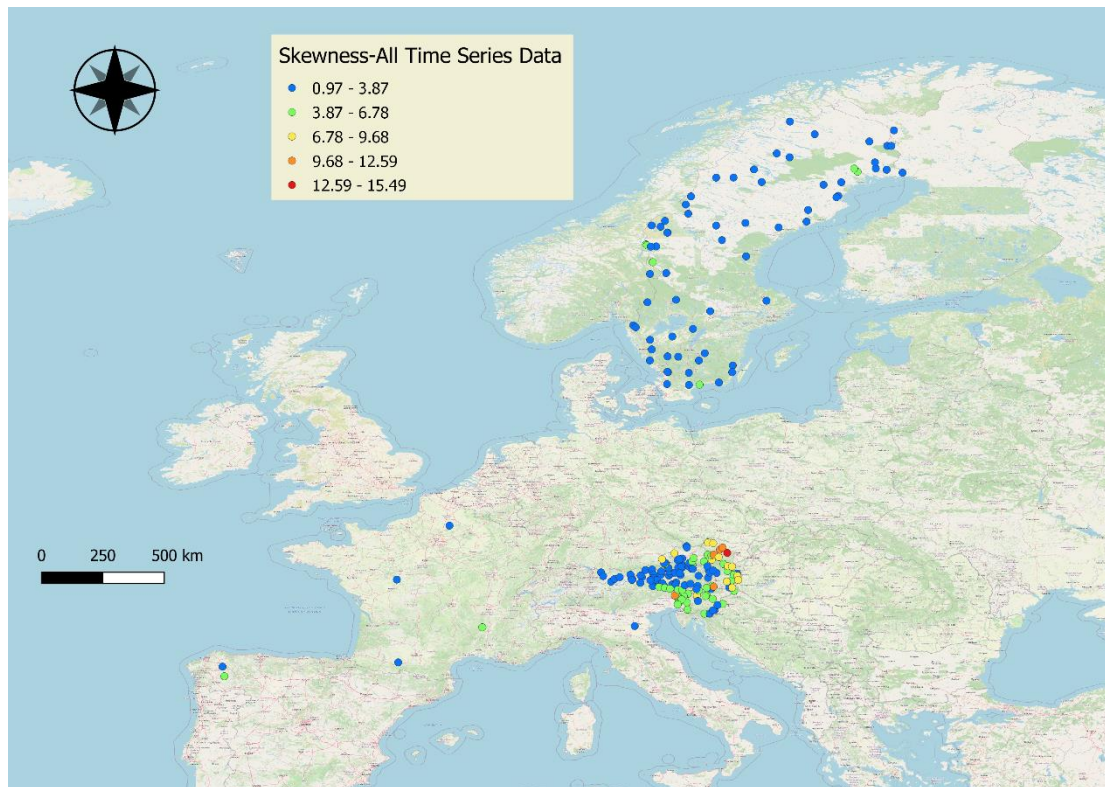


Σχήμα 7.4: Γεωγραφική κατανομή τιμών συντελεστή μεταβλητότητας

Για το στατιστικό μέγεθος της ασυμμετρίας παρατηρούμε ότι όλες οι τιμές είναι θετικές, πράγμα που συμβαίνει στις υδρολογικές μεταβλητές στην πλειονότητά τους, να είναι δηλαδή θετικά ασύμμετρες. Ακόμη οι εξαιρετικά υψηλές τιμές εμφανίζονται με χαμηλή συχνότητα. Στην περίπτωση θετικά ασύμμετρης κατανομής, η μέση τιμή της μεταβλητής είναι μεγαλύτερη από τη διάμεσο. Στους περισσότερους σταθμούς οι τιμές της ασυμμετρίας είναι χαμηλές δηλαδή κάτω από 3.9mm ενώ εξαιρετικά υψηλές τιμές παρατηρούνται στην Κεντρική Ευρώπη πάλι κοντά στην περιοχή της Βιέννης.



Σχήμα 7.5: Ιστόγραμμα Ασσυμετρίας όλων των σταθμών για τις χρονοσειρές σε ετήσια κλίμακα



Σχήμα 7.6: Γεωγραφική κατανομή τιμών ασυμετρίας

8 ΣΥΜΠΕΡΑΣΜΑΤΑ

8.1 Σύνοψη μελέτης και συμπεράσματα

Σκοπός της παρούσας εργασίας ήταν η στατιστική ανάλυση και υδρολογική ερμηνεία μιας βάσης δεδομένων ημερήσιων παροχών σε μεγάλο δείγμα σταθμών της Ευρώπης. Προσαρμόστηκαν πέντε διαφορετικές κατανομές, ενώ επίσης ελέγχθηκε και με διάφορους τρόπους η καταλληλότητα της προσαρμογής των κατανομών αυτών. Χρησιμοποιήθηκαν δεδομένα από το Παγκόσμιο Ιστορικό Κλιματολογικό Δίκτυο (Global Historical Climatology Network, GHCN-Daily της N.O.A.A.) και συγκεκριμένα 209 σταθμοί που πληρούσαν ορισμένα κριτήρια ποιότητας και πληρότητας καταγραφών. Το δείγμα των σταθμών αναλύθηκε αρχικά ως προς τα στατιστικά χαρακτηριστικά της ημερήσιας παροχής (μέση τιμή, διασπορά, τυπική απόκλιση, ασυμμετρία, κύρτωση) και παρατηρήθηκε η μεγάλη διασπορά σε σχέση με τη μέση τιμή απορροής των ημερήσιων παροχών σε σταθμούς κοντά στη Βιέννη και σε κάποιους στη Σλοβενία αλλά και η θετική ασυμμετρία μεγαλύτερη του 0.5 όλων των διαθέσιμων χρονοσειρών. Τα αποτελέσματα της ανάλυσης αυτής παρουσιάζονται σε διαγράμματα και πίνακες στο κεφάλαιο 5.

Συγκεκριμένα σε σχέση με τη στατιστική ανάλυση των δεδομένων παρατηρήθηκε η έντονη διακύμανση της μέσης τιμής απορροής ανάμεσα στους σταθμούς πράγμα που οφείλεται στα ιδιαίτερα κλιματικά χαρακτηριστικά της περιοχής που βρίσκεται ο κάθε σταθμός. Υψηλές τιμές μέσης απορροής άνω των 5.8mm παρατηρούνται στην Κεντρική Ευρώπη και συγκεκριμένα σε 6 σταθμούς στη Σλοβενία και σε 3 στην Αυστρία ενώ στην Βόρεια Ευρώπη, στην Ιταλία, στην Γαλλία και στην Ισπανία παρατηρούνται χαμηλές τιμές απορροής κάτω των 3mm. Αυτό οφείλεται στο ότι στην Κεντρική Ευρώπη σύμφωνα με τον ενημερωμένο κλιματικό χάρτη Köppen – Geiger για την περίοδο 1951–2000 (Kottek et al., 2006) κυριαρχεί το καθεστώς Cfb που συνδέεται με υγρό κλίμα όλο το χρόνο, ενώ στην Βόρεια Ευρώπη επικρατεί το πολικό κλίμα και ειδικότερα η Τούνδρα (E_T). Έτσι η Σκανδιναβία έχει χαμηλή απορροή ενώ και στην Κεντρική Ευρώπη υπάρχει μια ζώνη όπου παρατηρείται χαμηλή απορροή. Όσο αναφορά το συντελεστή μεταβλητότητας παρατηρούνται μεγαλύτερες τιμές στην εδαφική επικράτεια της Αυστρίας και συγκεκριμένα κοντά στην Βιέννη αλλά και σε κάποιες περιοχές της Σλοβενίας, ενώ στην υπόλοιπη Κεντρική Ευρώπη αλλά και στη Βόρεια Ευρώπη παρατηρούνται χαμηλότερες τιμές. Αυτό σημαίνει ότι για τις υψηλές τιμές υπάρχει μεγάλη διασπορά σε σχέση με τη μέση τιμή και άρα μεγαλύτερη αβεβαιότητα ως προς την αξιοπιστία των αποτελεσμάτων. Επιπλέον υπάρχουν περιοχές που έχουν υψηλή απορροή και χαμηλή μεταβλητότητα πράγμα που είναι δείκτης ξηρού κλίματος και άρα το ύψος απορροής δεν δείχνει μια πλήρη

εικόνα του κλιματικού καθεστώτος. Σχετικά με το στατιστικό μέγεθος της ασυμμετρίας παρατηρείται ότι όλες οι τιμές είναι θετικές ενώ οι εξαιρετικά υψηλές τιμές εμφανίζονται με χαμηλή συχνότητα. Στους περισσότερους σταθμούς οι τιμές της ασυμμετρίας είναι χαμηλές δηλαδή κάτω από 3.9mm ενώ εξαιρετικά υψηλές τιμές παρατηρούνται στην Κεντρική Ευρώπη πάλι κοντά στην περιοχή της Βιέννης. Ακόμη όλες οι τιμές του συντελεστή ασυμμετρίας είναι μεγαλύτερες από 0.5 και άρα αποκλείεται να προσαρμοστεί στα δεδομένα η κανονική κατανομή αν θεωρήσουμε ότι από 0-0.5 τείνουν τα δεδομένα προς την κανονική κατανομή.

Σχετικά με την προσαρμογή κατανομών στα δεδομένα επιλέχθηκαν πέντε διαφορετικές κατανομές, η λογαριθμοκανονική κατανομή (lognormal), η κατανομή γάμμα (Gamma), η κατανομή Weibull, η κατανομή PBF (Pareto-Burr-Feller) και η Γενικευμένη κατανομή γάμμα (Generalized Gamma). Βασικά κριτήρια για την επιλογή προς προσαρμογή αυτών των κατανομών ήταν η καταλληλότητα τους βάσει της διαθέσιμης βιβλιογραφίας και από την άλλη η δυνατότητα σύγκρισης δι-παραμετρικών και τρι-παραμετρικών κατανομών. Η μέθοδος που υιοθετήθηκε για τον υπολογισμό των παραμέτρων των κατανομών ήταν η Μέθοδος των L ροπών, η οποία αναλύεται στο Κεφάλαιο 2. Τέλος, ελέγχθηκε η καταλληλότητα της προσαρμογής αυτών των κατανομών με κριτήριο το Μέσο τετραγωνικό Σφάλμα. Η διαφορά αυτού του κριτηρίου σε σχέση με τις κλασικές στατιστικές δοκιμές καταλληλότητας προσαρμογής μιας κατανομής είναι κυρίαρχα η δυνατότητα που δίνει στον μελετητή για την διάκριση ως καταλληλότερης μίας και μόνο κατανομής έναντι υπολοίπων, γεγονός που αποτέλεσε και το βασικό κριτήριο για την επιλογή του στην παρούσα εργασία.

Από τη διαδικασία αυτή προέκυψε ότι οι τρεις από τις πέντε επιλεγμένες κατανομές προσαρμόζονται σε αρκετούς σταθμούς με σχετικά ικανοποιητικά ποσοστά. Συγκεκριμένα από τη σύγκριση των δι-παραμετρικών κατανομών με κριτήριο το μέσο Τετραγωνικό Σφάλμα επικράτησε η κατανομή lognormal στους 185 σταθμούς που αντιστοιχεί σε ποσοστό 88.51%. Από τη σύγκριση των τρι-παραμετρικών κατανομών με κριτήριο το μέσο Τετραγωνικό Σφάλμα επικράτησε η κατανομή PBF στους 162 σταθμούς που αντιστοιχεί σε ποσοστό 77.51%. Τέλος από τα διαγράμματα λόγου ροπής στην κλίμακα L προέκυψε ότι η κατανομή PBF προσαρμόζεται καλύτερα στα δεδομένα με ποσοστό 77.14% ενώ δεύτερη έρχεται η Γενικευμένη Κατανομή Γάμμα αφού όντας κατανομή τριών παραμέτρων (1 επιπλέον παράμετρο σχήματος) απεικονίζεται το ζεύγος L-skewness L-CV της κατανομής σαν μια περιοχή και όχι σαν καμπύλη όπως της δι-παραμετρικής κατανομής lognormal.

8.2 Προτάσεις για μελλοντική έρευνα

Για την επέκταση της μελέτης και τη γενίκευση των συμπερασμάτων της παρούσας διπλωματικής εργασίας προτείνονται κάποιες περαιτέρω διερευνήσεις.

Σε πρώτο στάδιο, θα ενίσχυε την ποιότητα και αξιοπιστία των συμπερασμάτων η εύρεση χρονοσειρών με μήκος παρατηρήσεων μεγαλύτερο των 60 ετών. Αυτό προβλέπει την επέκταση της βάσης δεδομένων GCHN-Daily με χρονοσειρές που θα καλύπτουν μεγαλύτερο χρονικό εύρος παρατηρήσεων ή ακόμα της προμήθειας δεδομένων από διαφορετικές πηγές. Ακόμα, επιβάλλεται ο περαιτέρω ποιοτικός έλεγχος των δεδομένων της βάσης και η διόρθωση τυχών σφαλμάτων στις μετρήσεις των καταγραφών.

Ακόμα καλύτερα, κρίνεται απαραίτητη η επέκταση των δειγμάτων μας σε περισσότερες περιοχές της Ευρώπης με άλλα κλιματικά χαρακτηριστικά, καθώς οι αναλύσεις αφορούν αποκλειστικά σε παρατηρήσεις κυρίως στην Κεντρική και Βόρεια Ευρώπη. Έτσι, θα ήταν δυνατή η διεξοδικότερη διερεύνηση της στατιστικής δομής των υδρολογικών διεργασιών, σε πολλαπλές κλίμακες, και η υδρολογική ανάλυση θα είχε ευρύτερα αποτελέσματα. Εκτός αυτού όμως, εξέχουσας σημασίας θα ήταν η εφαρμογή των παραπάνω αναλύσεων σε βάσεις δεδομένων που έχουν επεξεργαστεί άλλες μελέτες, με την προϋπόθεση ότι έχει ληφθεί υπόψη η όποια δομή συσχέτισης των δεδομένων που ενδεχομένως παραβιάζει τις υποθέσεις της κλασικής στατιστικής. Διαδικασία που αναμφισβήτητα είναι χρονοβόρα και επίπονη αλλά με πολύ πιο αξιόπιστα αποτελέσματα.

Επιπλέον, ζητούμενο είναι η αξιοποίηση των αποτελεσμάτων τέτοιων διευρυμένων στατιστικών αναλύσεων σε έργα υδρολογικού σχεδιασμού, όπως τα Μικρά Υδροηλεκτρικά Έργα (Μ.Υ.Η.Ε.) με στόχο τον σωστό και ακριβή σχεδιασμό υδραυλικών έργων καθώς και σε Μελέτες Περιβαλλοντικών Επιπτώσεων (ΜΠΕ). Εξέχουσας σημασίας θα ήταν και η εξαγωγή κάποιων περιοχικών σχέσεων ανάλογα με τα φυσιογραφικά χαρακτηριστικά κάθε περιοχής.

Τέλος θα είχε ειδική σημασία η επανάληψη της στατιστικής ανάλυσης και των υπολογισμών σε βάθος χρόνου, όταν θα είναι διαθέσιμα πολύ μεγαλύτερα μήκη καταγραφών του δείγματος και η σύγκριση των εξαχθέντων αποτελεσμάτων.

9 ΒΙΒΛΙΟΓΡΑΦΙΑ

9.1 Ξενόγλωσση Βιβλιογραφία

- ❖ Aronoff, S., 1989. Geographic Information Systems: A management perspective. Ottawa, Canada: WDL Publications
- ❖ Beirlant, J., Matthys, G., & Dierckx, G., 2001. Heavy-Tailed Distributions and Rating. *ASTIN Bulletin*, 31(1), 37-58, doi:10.2143/AST.31.1.993
- ❖ Brouers, F., 2015. The Burr XII distribution family and the maximum entropy principle: power-law phenomena are not necessarily nonextensive. *Open J Stat*, Volume 5, pp. 730- 741
- ❖ Burr, I., 1942. Cumulative Frequency Functions. *Annals of Mathematical Statistics*, Volume 13, pp. 215-235
- ❖ Burrough, A. P., 1986. Principles of Geographic Information Systems for Land Resource Assessment. Monographs on Soil and Resources Survey No 12. New York: Oxford Science Publications
- ❖ Cowen, D. J., 1988. GIS versus CAD versus DBMS: what are the differences ?. *Photogrammetric Engineering and Remote Sensing*, Volume 54, pp. 1551-4
- ❖ Devine, H. A. & Field, R. C., 1986. The gist of GIS. *Journal of Forestry*, August, pp. 17-22
- ❖ ESRI, 1990. Understanding GIS - The Arc/Info Method. Redlands, California: Environmental Systems Research Institute Inc.
- ❖ Ioannis Tsoukalas, Christos Markopoulos, and Demetris Koutsoyiannis, 2018, Simulation of Stochastic Processes Exhibiting Any-Range Dependence and Arbitrary Marginal Distributions, Department of Water Resources, School of Civil Engineering, National Technical University of Athens, Zographou, Greece
- ❖ Koutsoyiannis Dimitris, *Statistical Hydrology*, Edition 4, 312 pages, National Technical University of Athens, Athens, 1997
- ❖ Koutsoyiannis Demetris, 2017, Article: Entropy Production in Stochastics, Department of Water Resources, School of Civil Engineering, National Technical University of Athens, 15780 Athina, Greece
- ❖ Koshkariov, V. A., Tikunov, S. V. & Trofimov, M. A., 1989. The current state and the main trends in the development of geographical information systems in the USSR. *International journal of geographical information systems*, 3(3), pp. 257-272
- ❖ Maquire, D. J., 1991. An Overview and Definition of GIS. In: Maguire, D.J., Goodchild, M.F. and Rhind, D.W., Eds., *Geographical Information Systems: Principles and Applications*. Hoboken: Wiley

- ❖ Mert, I. and Karakus, C., 2015. A statistical analysis of wind speed data using Burr, generalized gamma, and Weibull distributions in Antakya, Turkey, Tubitak, doi:10.3906/elk-1402-66
- ❖ Panagiotis Kossieris, and Christos Markopoulos, 2017, Article: Exploring the Statistical and Distributional Properties of Residential Water Demand at Fine Time Scales, Department of Water Resources, School of Civil Engineering, National Technical University of Athens, Heroon Polytechniou 5, GR-157 80 Zographou, Greece
- ❖ Parker, D. H., 1988. The unique qualities of a geographic information system: a commentary. *Photogrammetric Engineering and Remote Sensing*, 54(11), pp. 1547-1549
- ❖ T. Iliopoulou, October 2020, Stochastic investigation of hydrological extremes: influence of temporal variability and dependence, PhD thesis, 237 pages, Department of Water Resources and Environmental Engineering – National Technical University of Athens
- ❖ W. N. Venables, D. M. Smith and the R Development Core Team, 2005, *An Introduction to R – Notes on R: A Programming Environment for Data Analysis and Graphics Version 2.2.0* (91pages), England
- ❖ Wieczorek, W. F. & Delmerico, A. M., 2009. *Geographic Information Systems. Comput Stat*, 1(2), pp. 167-186
- ❖ WIKIPEDIA, 2021. QGIS. [Online] Available at: <https://en.wikipedia.org/wiki/QGIS>

9.2 Ελληνόγλωσση Βιβλιογραφία

- ❖ Θ. Ηλιοπούλου, Διερεύνηση της ύπαρξης μακροπρόθεσμης εμμονής στην ετήσια βροχόπτωση από μία παγκόσμια βάση δεδομένων, Διπλωματική εργασία, 146 pages, Τομέας Υδατικών Πόρων και Περιβάλλοντος – Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, Μάρτιος 2013.
- ❖ Κάβουρας, Μ., Δάρρα, Α., Κονταξάκη, Σ., Τομαή, Ε., 2016. *Επιστήμη Γεωγραφικής Πληροφορίας - Αρχές και Τεχνολογίες*. [ηλεκτρ. βιβλ.] Αθήνα: Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Διαθέσιμο στο: <http://hdl.handle.net/11419/6392>
- ❖ Καλογήρου Σ, Χαρτογράφηση με Ανοιχτό Λογισμικό GIS, http://gisc.gr/sac/docs/SKalogirou_QGIS.pdf, 2013
- ❖ Κοκολάκης Γ. - Φουσκάκης Δ., 2009, *Στατιστική Θεωρία & Εφαρμογές* (σελ. 370), Εκδόσεις Συμεών, Αθήνα, ΕΜΠ

- ❖ Κουτσογιάννης, Δ., και Θ. Ξανθόπουλος, Τεχνική Υδρολογία, Έκδοση 3, (418 σελ.), Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, 1999
- ❖ Κουτσογιάννης, Δ., 1996. Στατιστική Υδρολογία, Έκδοση 3, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, σελ. 76.
- ❖ Κουτσογιάννης Δ., 2013. «Σημειώσεις Στοχαστικών Μεθόδων στους Υδατικούς Πόρους», Επανεκδοση, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα.
- ❖ Μαρία Νέζη, Πολυμεταβλητή στατιστική ανάλυση ακραίων βροχοπτώσεων και παροχών σε δείγμα 400 λεκανών απορροής των ΗΠΑ από τη βάση δεδομένων ΜΟΡΕΧ, Διπλωματική Εργασία, 97 σελίδες, Εθνικό Μετσόβιο Πολυτεχνείο, Σχολή Πολιτικών Μηχανικών, Τομέας Υδατικών Πόρων και Περιβάλλοντος
- ❖ Μιμίκου Μ.Α., Ε.Α. Μπαλάς, Τεχνική Υδρολογία, 5η Έκδοση, (416 σελ.), Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, 2012
- ❖ Νάκος Β., Φιλιππακοπούλου, Θεματική χαρτογραφία, ΕΜΠ, Αθήνα 1992
- ❖ Τσιρτσής Γ. – Κολοβογιάννης Β., 2015, Προγραμματισμός στη Γλώσσα R – Εισαγωγικές Σημειώσεις (Έκδοση 2.0) (σελ. 62), Μυτιλήνη, Πανεπιστήμιο Αιγαίου – Σχολή Περιβάλλοντος Τμήμα Επιστημών της Θάλασσας
- ❖ Τσούλος, Λ., Σκοπελίτη, Α., Στάμου, Λ. 2015. Χαρτογραφική σύνθεση και απόδοση σε ψηφιακό περιβάλλον. [ηλεκτρ. βιβλ.] Αθήνα:Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών. Διαθέσιμο στο: <http://hdl.handle.net/11419/2506>
- ❖ Φουσκάκης Δ., 2013, Ανάλυση Δεδομένων με Χρήση της R (σελ. 504), Εκδόσεις Τσότρας Αθήνα, ΕΜΠ
- ❖ Φωκιανός Κ., Χαραλάμπους Χ., 2010, Εισαγωγή στην R - Πρόχειρες Σημειώσεις (σελ.282), Λευκωσία, Τμήμα Μαθηματικών & Στατιστικής Πανεπιστημίου Κύπρου
- ❖ Χαλκιάς, 2019, Κάνοντας γεωαναφορά και ψηφιοποίηση με το ανοιχτού κώδικα λογισμικό GIS QGIS, Σημειώσεις Μαθήματος Συστήματα Γεωγραφικών Πληροφοριών, Χαροκόπειο Πανεπιστήμιο

9.3 Ιστοσελίδες

- ❖ <http://cran.r-project.org/>
- ❖ <http://manuals.bioinformatics.ucr.edu/home/programming-in-r>
- ❖ <https://www.qgis.org/en/site/forusers/download.html>
- ❖ <http://www.qgis.org>
- ❖ <https://www.gnu.org/licenses/licenses.html>
- ❖ <https://www.ncdc.noaa.gov/ghcnd-data-access>
- ❖ <https://www.osgeo.org/>

- ❖ <https://ourworldindata.org/>
- ❖ <https://plugins.qgis.org/plugins/>
- ❖ <https://www.researchgate.net/>
- ❖ <https://www.wikipedia.org/>
- ❖ <https://sourceforge.net/>

ΠΑΡΑΡΤΗΜΑ Ι: Οι σταθμοί που χρησιμοποιήθηκαν

A/A	Stations	Area (km ²)	Coordinates	
			X	Y
1	Po River at Pontelagoscuro (Italy)	70091.00	44.89	11.61
2	Cabe River at Rivas Altas	353.00	42.55	-7.48
3	Minho river at Lugo (Spain)	2303.00	43.00	-7.56
4	Mura River at Gornja Radgona	10197.00	46.68	16.00
5	Scavnica River at Pristava (Slovenia)	272.54	46.52	16.23
6	Ledava River at Polana (Slovenia)	208.21	46.68	16.14
7	Bistrica River at Muta (Slovenia)	146.55	46.61	15.16
8	Radovna River at Podhom (Slovenia)	166.79	46.39	14.09
9	Sava Bohinjka River at Sveti	93.99	46.28	13.89
10	Kokra River at Kokra (Slovenia)	112.34	46.31	14.50
11	Sotla River at Rakovec (Slovenia)	559.99	45.92	15.70
12	Kolpa River at Petrina (Slovenia)	460.00	45.46	14.86
13	Kolpa River at Radenci (Slovenia)	1191.00	45.46	15.10
14	Kolpa River at Metlika (Slovenia)	2002.00	45.64	15.33
15	Lahinja River at Gradac (Slovenia)	221.32	45.62	15.24
16	Ljubljanica River at Moste (Slovenia)	1762.32	46.06	14.54
17	Savinja River at Nazarje (Slovenia)	457.30	46.32	14.95
18	Savinja River at Lasko (Slovenia)	1663.60	46.15	15.23
19	Dreta River at Krase (Slovenia)	100.84	46.28	14.90
20	Krka River at Podbocje (Slovenia)	2238.12	45.86	15.46
21	Soca River at Krsovec (Slovenia)	158.06	46.34	13.60
22	Soca River at Log Cezsoksi	324.74	46.31	13.49
23	Soca River at Kobarid (Slovenia)	437.02	46.25	13.59
24	Soca River at Solkan (Slovenia)	1572.80	45.98	13.65
25	Ucja River at Zaga (Slovenia)	50.21	46.31	13.48
26	Idrijca River at Podroteja (Slovenia)	112.84	45.99	14.03
27	Idrijca River at Hotesk (Slovenia)	442.83	46.13	13.79
28	Vipava River at Miren (Slovenia)	589.97	45.89	13.61
29	Reka River at Cerkevnikov Mlin	332.37	45.65	14.06
30	Alfenz river at Lorfc>ns (Austria)	172.60	47.13	15.02
31	Ill river at Gisingen (Austria)	1281.00	47.26	13.73
32	Bregenzera river at Au (Austria)	149.30	47.33	16.37
33	Bregenzera river at Mellau (Austria)	228.60	47.36	15.42
34	Subersach river at Schf6>nenbac (Austria)	31.10	47.39	10.04
35	Subersach river at Lingenau (Austria)	111.60	47.45	15.92
36	Bregenzera river at Kennelbach (Austria)	826.30	47.48	15.05
37	Lech river at Lech (Tann	84.30	47.21	10.14
38	Inn river at Kajetansbr (Austria)	2162.00	46.95	10.51
39	Inn river at Prutz (Austria)	2461.50	47.08	10.66

40	Otztaler A river at Tumpen	785.50	47.16	10.91
41	Obernberge river at Gries am	58.30	47.04	11.48
42	Schmirnbac river at St. Jodok	108.80	47.06	11.50
43	Gschnitzba river at Steinach a	111.40	47.09	11.47
44	Sill river at Puig (Austria)	341.80	47.11	11.45
45	Sill river at Innsbruck- (Austria)	854.40	47.28	11.41
46	Obernberge river at Gries am	200.50	47.04	11.48
47	Ziller river at Klaushof (135.30	47.16	11.94
48	Ziller river at Zell am	696.30	47.24	11.88
49	Brixentale river at Bruckh<e4>usl (Austria)	322.30	47.49	12.10
50	Kitzb<fc>hele river at Kitzb<fc>hel (Austria)	153.00	47.46	12.39
51	Kitzb<fc>hele river at St. Johann	332.40	47.52	12.42
52	Salzach river at Mittersill (Austria)	582.60	47.28	12.48
53	Salzach river at Bruck (Sal	1168.70	47.29	12.83
54	Fuscher Ac river at Bruck	161.00	47.27	12.83
55	Gro<df>arler river at Gro<df>arl (Austria)	144.60	47.24	13.20
56	Kleinarler river at Wagrain (Austria)	91.40	47.33	13.30
57	Fritzbach river at Kreuzbergm (Austria)	154.80	47.44	13.22
58	Lammer river at Oberg<e4>u (Austria)	394.50	47.58	13.18
59	Salzach river at Golling (Austria)	3555.70	47.60	13.16
60	Salzach river at Salzburg (Austria)	4425.70	47.82	13.04
61	Glanbach river at Moos (Austria)	26.50	47.77	13.00
62	Urslaubach river at Saalfelden (Austria)	119.50	47.43	12.84
63	Saalach mi river at Siezenheim	1139.10	47.81	12.98
64	Moosache river at Au-St.Geor (Austria)	88.10	48.00	12.88
65	Taurach river at L<f6>benau (Austria)	86.50	47.37	13.48
66	Antiesen river at Haging (Austria)	164.90	48.27	13.45
67	Gro<df>e M<fc>hl river at Furtm<fc>hle	256.40	48.61	14.02
68	Gro<df>e M<fc>hl river at Teufelm<fc>hl	453.50	48.54	14.03
69	Traun river at Obertraun (Austria)	317.40	47.56	13.72
70	Traun river at Steeg (See	615.20	47.61	13.63
71	Traun river at Ebensee (Austria)	1223.40	47.81	13.77
72	Traun river at Roitham (Austria)	1458.40	48.03	13.82
73	Ager river at Raudaschls (Austria)	462.90	47.95	13.61
74	Ager river at Schalchham (Austria)	949.90	48.00	13.68
75	Steyr river at Dietlgut (Austria)	25.20	47.67	14.11
76	Teichl river at Teichlbr<fc>c (Austria)	148.60	47.73	14.30
77	Steyr river at Klaus an	542.40	47.83	14.16
78	Krumme Ste river at Molln	129.40	47.89	14.28
79	Steyr river at Pergern (Austria)	898.10	48.05	14.36
80	Isper river at Isperdorf (Austria)	164.70	48.20	15.00
81	Ybbs river at Opponitz (506.90	47.89	14.83
82	Je<df>nitz river at Neubruck (Austria)	32.20	47.97	15.19
83	Erlauf river at Niederndor (Austria)	595.30	48.17	15.17
84	Melk river at Matzleinsd (Austria)	284.60	48.20	15.28

85	Pielach river at Hofstetten (Austria)	289.50	48.10	15.51
86	Krems river at Imbach (Austria)	305.90	48.45	15.57
87	Kamp river at Stiefern (Austria)	1493.30	48.54	15.69
88	Perschling river at Atzenbrugg (Austria)	268.30	48.29	15.91
89	Braunabac river at Hoheneich (Austria)	291.50	48.77	15.01
90	Deutsche T river at Schwarzena	175.50	48.74	15.26
91	Schwarza river at Singerin (252.30	47.77	15.74
92	Schwarza river at Gloggnitz (Austria)	472.20	47.68	15.94
93	Pitten river at Warth (Austria)	277.00	47.65	16.13
94	Lafnitz river at Dobersdorf (Austria)	925.10	47.02	16.14
95	Pinka river at Oberwart (171.70	47.28	16.22
96	Pinka river at Woppendorf (Austria)	416.90	47.20	16.39
97	Tauchenbac river at Hannersdor (Austria)	175.40	47.23	16.38
98	Strem river at Heiligenbr (Austria)	400.40	47.03	16.42
99	Traun river at Archkogel (Austria)	125.10	47.62	13.83
100	Altaussee river at Altaussee (Austria)	54.50	47.64	13.77
101	Unterthale river at Tetter (Austria)	65.20	47.36	13.71
102	Enns river at Liezen (R<f6>	2116.20	47.55	14.25
103	Paltenbach river at Selzthal (Austria)	368.70	47.55	14.31
104	Salza river at Wildalpen (Austria)	592.30	47.67	14.98
105	Raab river at Feldbach (Austria)	689.40	46.96	15.89
106	P<f6>ls river at P<f6>ls (Austria)	422.00	47.22	14.58
107	Mur river at Leoben (Austria)	4391.90	47.38	15.09
108	Th<f6>rlbach river at Hansenh<fc>tt (Austria)	324.20	47.47	15.26
109	Schwarze S river at Schwanberg	75.20	46.76	15.21
110	Wei<df>e Sulm river at Wies	69.70	46.72	15.27
111	Drau river at Rabland (Austria)	374.00	46.75	12.45
112	Isel river at Hinterbich (Austria)	107.00	47.02	12.34
113	Islitz river at Hinterbich (Austria)	35.00	47.02	12.34
114	Gschl<f6><df>bac river at Innergschl (Austria)	39.30	47.12	12.45
115	Tauernbach river at Matreier T	59.90	47.12	12.50
116	Isel river at Br<fc>hl (Austria)	518.40	46.97	12.55
117	Dorferbach river at Sp<f6>ttling (Austria)	47.00	47.03	12.63
118	Teischnitz river at Sp<f6>ttling- (Austria)	14.20	47.03	12.63
119	Isel river at Lienz (Austria)	1198.70	46.83	12.77
120	Islitz river at Hinterbich (Austria)	500.20	47.02	12.34
121	Drau river at Oberdraubu (Austria)	2112.00	46.75	12.97
122	Drau river at Sachsenbur (Austria)	2561.40	46.83	13.35
123	M<f6>ll river at Winklern (Austria)	412.10	46.88	12.89
124	Malta river at Sandriesen (Austria)	266.00	46.91	13.53
125	Lieser river at Spittal (F	1035.50	46.81	13.50
126	Gail river at Maria Lugg	146.10	46.70	12.74
127	Gail river at Mauthen (Austria)	348.60	46.67	13.00
128	Gail river at Rattendorf (Austria)	594.90	46.62	13.25
129	Gail river at N<f6>tsch (Austria)	908.50	46.58	13.62

130	Gail river at Federaun (Austria)	1304.90	46.57	13.82
131	Vellach river at Miklauzhof (Austria)	194.30	46.54	14.59
132	Gurk river at Maitratten (Austria)	201.40	46.79	13.93
133	Gurk river at Weitensfel (Austria)	431.80	46.85	14.20
134	Görschitz river at Brückl (Austria)	315.60	46.76	14.53
135	Gurk river at Gumisch (Austria)	2555.40	46.61	14.48
136	Lavant river at St.Gertrau (Austria)	380.20	46.90	11.87
137	Lavant river at Krottendor (Austria)	954.50	46.66	14.95
138	Oise River at Sempigny (France)	4320.00	49.56	2.99
139	Vienne River at Nouatre (France)	19895.00	47.05	0.55
140	Gave d'Oloron River at Oloron-Sainte-Marie	1101.00	43.20	0.61
141	Rötkfors River at X (Sweden)	23102.90	66.17	22.82
142	Niemisel River at X (Sweden)	3780.80	66.02	21.97
143	Gimdalsby River at X (Sweden)	2163.60	62.84	15.66
144	Rörvik River at X (Sweden)	158.70	57.24	14.59
145	Solberg River at X (Sweden)	1080.90	65.75	15.39
146	Niavve River at X (Sweden)	1717.50	66.88	18.21
147	Fuluns River at X (Sweden)	883.10	61.31	13.07
148	Hälabäck River at X (Sweden)	4.70	56.12	14.62
149	Vassbotten River at X (Sweden)	624.10	58.88	11.54
150	Getebro River at X (Sweden)	1332.70	57.01	16.17
151	Möckeln River at X (Sweden)	1026.00	56.66	14.12
152	Lillglån River at X (Sweden)	64.60	62.64	12.13
153	Ytterholmen River at X (Sweden)	1012.00	66.17	21.81
154	Kukkasjövi River at X (Sweden)	494.30	66.12	23.32
155	Assmebro River at X (Sweden)	652.80	57.43	13.14
156	Ljusnedal-vre River at X (Sweden)	340.30	62.55	12.60
157	Grötsjö River at X (Sweden)	565.00	61.81	12.45
158	Fjällnäs River at X (Sweden)	108.40	62.60	12.18
159	Görsilt River at X (Sweden)	55.10	56.70	13.15
160	Töndalen River at X (Sweden)	226.60	62.54	12.35
161	Brücklingen River at X (Sweden)	155.80	63.74	13.01
162	Rengen River at X (Sweden)	1110.10	64.07	14.10
163	Flötemarken River at X (Sweden)	229.70	58.79	11.66
164	Rön River at X (Sweden)	583.80	63.64	16.76
165	Karats River at X (Sweden)	1173.90	66.69	18.81
166	Kaalasjövi River at X (Sweden)	1472.50	67.77	19.98
167	Ankarvattnet River at X (Sweden)	427.80	64.88	14.22
168	Sikfors River at X (Sweden)	10816.10	65.53	21.21
169	Torsebro River at X (Sweden)	3664.50	56.10	14.13
170	Muonio River at X (Sweden)	9595.04	67.95	23.67
171	Murusjö River at X (Sweden)	345.60	64.49	13.98
172	Nybergssund River at X (Sweden)	4428.00	61.27	12.32
173	Magnor River at X (Sweden)	360.00	59.95	12.20
174	Kukkolankoski-vre River at X (Sweden)	33930.00	65.98	24.06

175	Junosuando-Tornegrenen River at X (Sweden)	4351.00	67.43	22.52
176	Pajala River at X (Sweden)	11038.10	67.23	23.35
177	Rdupp River at X (Sweden)	15733.00	66.46	22.78
178	Bjrkkliden River at X (Sweden)	621.80	65.42	20.39
179	Vindeln River at X (Sweden)	11850.50	64.24	19.67
180	Anundsjön River at X (Sweden)	1465.20	63.43	18.30
181	Åster-Noren River at X (Sweden)	2384.40	63.47	12.81
182	Medstugusjön River at X (Sweden)	224.10	63.53	12.40
183	Ottsjön River at X (Sweden)	730.50	63.19	13.14
184	Fångeån River at X (Sweden)	163.60	63.18	13.12
185	Solbergsvattnet River at X (Sweden)	2428.40	63.52	15.39
186	Velen River at X (Sweden)	45.00	58.71	14.32
187	Nåmmen River at X (Sweden)	157.20	57.58	14.87
188	Källstorp River at X (Sweden)	342.40	56.70	16.14
189	Klippan River at X (Sweden)	241.30	56.14	13.11
190	Simlingen River at X (Sweden)	259.70	56.72	13.13
191	Vikaresjön River at X (Sweden)	825.80	57.42	13.63
192	Ås River at X (Sweden)	2160.20	57.24	12.31
193	Edebäck River at X (Sweden)	8594.00	60.07	13.53
194	Velanda River at X (Sweden)	277.80	58.21	12.32
195	Torrbjelle River at X (Sweden)	2859.00	63.70	19.60
196	Hasselasjön River at X (Sweden)	651.20	62.09	16.79
197	Vattholma River at X (Sweden)	293.80	60.02	17.73
198	Hammarby River at X (Sweden)	891.40	59.54	15.12
199	Nåttaby River at X (Sweden)	437.70	56.22	15.52
200	Osttråsket River at X (Sweden)	150.50	64.90	21.07
201	Kånge River at X (Sweden)	905.50	64.84	20.99
202	Åverstjuktan River at X (Sweden)	418.40	65.75	16.20
203	Sorsele River at X (Sweden)	6056.30	65.54	17.51
204	Tvårsjön River at X (Sweden)	5.20	57.76	12.40
205	Tornetråsk River at X (Sweden)	3345.50	68.36	18.83
206	Kallio River at X (Sweden)	14477.10	67.23	23.55
207	Laisan River at X (Sweden)	1773.60	66.13	17.15
208	Marum River at X (Sweden)	662.20	58.36	13.35
209	Glueyre River at Gluiras (France)	72.00	44.83	4.52

ΠΑΡΑΡΤΗΜΑ ΙΙ: Ο κώδικας στην γλώσσα R

Ο κώδικας που χρησιμοποιήθηκε για την εισαγωγή των δεδομένων και τον υπολογισμό των στατιστικών χαρακτηριστικών των χρονοσειρών ανά μήνα είναι:

```
# required packages
library(Benchmarking)
library(datasets)
library(data.table)
library(dplyr)
library(FAdist)
library(fitdistrplus)
library(flexsurv)
library(forecast)
library(ggplot2)
library(ggrepel)
library(grid)
library(gridExtra)
library(gtools)
library(hydroTSM)
library(logspline)
library(matrixStats)
library(moments)
library(lmomco)
library(lmom)
library(lmomRFA)
library(lubridate)
library(openxlsx)
library(pacman)
library(PerformanceAnalytics)
library(psych)
library(quantmod)
library(RColorBrewer)
library(readr)
library(readxl)
library(reshape2)
library(rio)
library(runner)
library(scales)
library(stringi)
library(tidyverse)
library(tseries)
library(writexl)
library(xts)
```


#Read the data

```
names_files<-list.files(path="C:/Users/user/Dropbox/My PC (DESKTOP-
DP4VOIS)/Documents/allldata",pattern=".txt")
names_files=gtools::mixedsort(names_files)
xtslist<-vector("list", length=length(names_files))
data_length=length(names_files)
namelist<-c()
dataXY=matrix(NA, nrow = data_length, ncol = 2)
areaSize=rep(NA, data_length)
colnames(dataXY)=c('X', 'Y')
for(i in 1:data_length){
X1 <- read_csv(names_files[i], col_names = FALSE)
```

read the coordinates

```
coordTemp=as.character(X1[2,])
coordTemp= stringr::str_split(coordTemp, ':')[[1]][2]
coordTemp= stringr::str_split(coordTemp, ',')[[1]]
```

```
if (length(coordTemp)==1) {
  coordTemp= stringr::str_split(coordTemp, ' ')[[1]]
  coordTemp=na.omit(as.numeric(coordTemp))
}
```

```
dataXY[i,]= as.numeric(coordTemp)
```

read the catchment size - [km2], startdate, enddate, leapy, title

```
areaTemp=as.character(X1[3,])
areaTemp= stringr::str_split(areaTemp, ':')[[1]][2]
areaTemp= stringr::str_split(areaTemp, 'km2')[[1]][1]
areaSize[i]= as.numeric(areaTemp)
```

```
startdate<-as.character(X1[5,])
enddate<-as.character(X1[7,])
leapy<-as.character(X1[9,])
title<-as.character(X1[1,])
namelist<-c(namelist,title)
type<-as.character(X1[10,])
```

```
ts<-as.vector(X1[11:nrow(X1),])
ts<-as.numeric(ts$X1)
ts<-ts*86.4/areaSize[i]
```

convert negative values to NAs

```
ts=ifelse(ts<0, NA, ts)
```

```

# create empty xts full series to put variables
starttime=as.POSIXct(startdate,format="%Y/%m/%d",tz = 'UTC')
endtime=as.POSIXct(enddate,format="%Y/%m/%d", tz = 'UTC')
datesNA=seq(from= starttime, to=endtime, by='1 day')
xtsNA=xts(x = matrix(NA,ncol=1,nrow=length(datesNA)), order.by = datesNA)
if(leapy=="T"){
  myDates <- !(format(datesNA,"%m") == "02" & format(datesNA, "%d") == "29")
  leapdayspos<-which(myDates==FALSE)
  xtsNA<-xtsNA[-leapdayspos]
  xtsNA[,1]<-ts
} else {

  xtsNA[,1]<-ts
  myDates <- !(format(datesNA,"%m") == "02" & format(datesNA, "%d") == "29")
  leapdayspos<-which(myDates==FALSE)
  xtsNA<-xtsNA[-leapdayspos]
  }

xtslist[[i]]<-xtsNA

print( paste0("Importing file: ", i, ' from ', data_length) )
}

names(xtslist)<-namelist

```

Function to select a subset of daily data - Subsetting by month

```

sub_month=function(x,Month){
  Dates=index(x)
  # D=format(D,"%m")==="Month" # Month is a variable specified by the user.
  Dates=format(Dates,"%m")
  Dates=as.numeric(Dates)
  # Select_month=which(D==TRUE)
  Select_month=which(Dates==Month)
  tsd=x
  tsd=tsd[Select_month]
  return(tsd)
}

```

Calculating statistical characteristics by month of all values

```

statslist=vector("list",data_length)
statslistNonZero=vector("list",data_length)
flow_data_per_month_list=vector("list",data_length)
maxlag=3
for(i in 1:data_length){
  flow_data=xtslist[[i]]

```

```

currmat=c()
currmatNonZero=c()
for(j in 1:12){
flow_data_per_month=sub_month(x=flow_data, Month=j)
df=as.data.frame(flow_data_per_month)
df=na.omit(df)
df_flow_data=df$V1
Pdr<-round(mean(df_flow_data<=0.01,na.rm=T),5)
lmom=lmom.ub(df_flow_data)
LMean=lmom$L1
LScale=lmom$L2
L3=lmom$L3
L4=lmom$L4
LVariation=lmom$LCV
LSkewness=lmom$TAU3
Lkurtosis=lmom$TAU4
Min=round(min(df_flow_data,na.rm=TRUE),4)
Max=round(max(df_flow_data,na.rm=TRUE),4)
Mean=round(mean(df_flow_data,na.rm=TRUE),4)
Var=round(var(df_flow_data,na.rm=TRUE),4)
StDev=round(sd(df_flow_data,na.rm=TRUE),4)
Variation=round(StDev/Mean,4)
Mom3=round(sum((df_flow_data-
Mean)^3,na.rm=T)*(1/(sum(!is.na(df_flow_data))-1)),5)
Skewness=round(Mom3/StDev^3,4)
Kurtosis=round(moments::kurtosis(df_flow_data,na.rm=TRUE))
Month=j
Q5<-round(quantile(df_flow_data,probs=c(0.05),na.rm=TRUE,names=FALSE),5)
Q20<-round(quantile(df_flow_data,probs=c(0.20),na.rm=TRUE,names=FALSE),5)
Q50<-round(quantile(df_flow_data,probs=c(0.50),na.rm=TRUE,names=FALSE),5)
Q80<-round(quantile(df_flow_data,probs=c(0.80),na.rm=TRUE,names=FALSE),5)
Q90<-round(quantile(df_flow_data,probs=c(0.90),na.rm=TRUE,names=FALSE),5)
Autocorrelation<-
acf(df_flow_data,lag.max=maxlag,type="correlation",plot=FALSE,na.action=
na.pass)
Autocorrelation<-as.vector(Autocorrelation[[1]][2:(maxlag+1)])
Autocovariance<-
acf(df_flow_data,lag.max=maxlag,type="covariance",plot=FALSE,na.action=
na.pass)
Autocovariance<-as.vector(Autocovariance[[1]][2:(maxlag+1)])
Description_Analysis=c(LMean,LScale,L3,L4,LVariation,LSkewness,
Lkurtosis,Min,Max,Mean,Var,StDev,Variation,Mom3,
Skewness,Kurtosis,Q5,Q20,Q50,Q80,Q90,Pdr,Autocorrelation,
Autocovariance)
currmat=cbind(currmat,Description_Analysis)
rownames(currmat)=c("LMean","LScale","L3","L4","LVariation","LSkewness",
"Lkurtosis","Min","Max","Mean","Var","StDev","Variation","Mom3",

```

```

"Skewness","Kurtosis","Q5","Q20","Q50","Q80","Q90","Pdr","Autocorrelation1",
"Autocorrelation2","Autocorrelation3","Autocovariance1","Autocovariance2"
"Autocovariance3")
flow_data_per_month_list[[i]][j]=flow_data_per_month

# Calculating statistical characteristics by month of non zero values
MinNonZero<-round(min(df_flow_data[df_flow_data>0],na.rm=TRUE),5)
MaxNonZero<-round(max(df_flow_data[df_flow_data>0],na.rm=TRUE),5)
MeanNonZero<-round(mean(df_flow_data[df_flow_data>0],na.rm=T),5)
VarNonZero<-round(var(df_flow_data[df_flow_data>0],na.rm=T),5)
StDevNonZero<-round(sd(df_flow_data[df_flow_data>0],na.rm=T),5)
VariationNonZero=round(StDevNonZero/MeanNonZero,5)
Mom3NonZero<-round(sum((df_flow_data[df_flow_data>0]-
  MeanNonZero)^3,na.rm=T)*(1/(sum(!is.na(df_flow_data[df_flow_data>0]))-1)),6)
SkewnessNonZero<-round(Mom3NonZero/VarNonZero^1.5,5)
KurtosisNonZero=round(moments::kurtosis(df_flow_data[df_flow_data>0],na.rm=T
RUE))
Q5NonZero<-round(quantile(df_flow_data[df_flow_data>0],probs=c(0.05),
na.rm=TRUE,names=FALSE),5)
Q20NonZero<-round(quantile(df_flow_data[df_flow_data>0],probs=c(0.20),
na.rm=TRUE,names=FALSE),5)
Q50NonZero<-round(quantile(df_flow_data[df_flow_data>0],probs=c(0.50),
na.rm=TRUE,names=FALSE),5)
Q80NonZero<-round(quantile(df_flow_data[df_flow_data>0],probs=c(0.80),
na.rm=TRUE,names=FALSE),5)
Q90NonZero<-round(quantile(df_flow_data[df_flow_data>0],probs=c(0.90),
na.rm=TRUE,names=FALSE),5)
AutocorrelationNonZero<-
acf(df_flow_data[df_flow_data>0],lag.max=maxlag,type="correlation",
plot=FALSE,na.action = na.pass)
AutocorrelationNonZero<-as.vector(Autocorrelation[[1]][2:(maxlag+1)])
AutocovarianceNonZero<-
acf(df_flow_data[df_flow_data>0],lag.max=maxlag,type="covariance",
plot=FALSE,na.action = na.pass)
AutocovarianceNonZero<-as.vector(Autocovariance[[1]][2:(maxlag+1)])
tsss<-df_flow_data[!is.na(df_flow_data) & df_flow_data>0]
if(length(tsss)>4 && length(unique(tsss))>1){
  lmom<-lmom.ub(tsss)
  LMeanNoNZero<-lmom$L1
  LScaleNoNZero<-lmom$L2
  L3NoNZero<-lmom$L3
  L4NoNZero<-lmom$L4
  LVariationNoNZero<-lmom$LCV
  LSkewnessNoNZero<-lmom$TAU3
  LkurtosisNoNZero<-lmom$TAU4
}

```

```

Description_AnalysisNonZero=c(LMeanNoZero,LScaleNoZero,L3NoZero,L4NoZero,
LVariationNoZero,LSkewnessNoZero,LkurtosisNoZero,
MinNonZero,MaxNonZero,MeanNonZero,VarNonZero,StDevNonZero,
Mom3NonZero,SkewnessNonZero,KurtosisNonZero,Q5NonZero,Q20NonZero,
Q50NonZero,Q80NonZero,Q90NonZero,AutocorrelationNonZero,
AutocovarianceNonZero)
  currmatNonZero=cbind(currmatNonZero,Description_AnalysisNonZero)
rownames(currmatNonZero)=c("LMeanNonZero","LScaleNonZero","L3NonZero","L4
NonZero","LVariationNonZero","LSkewnessNonZero","LkurtosisNonZero",
"MinNonZero","MaxNonZero","MeanNonZero","VarNonZero","StDevNonZero",
"Mom3NonZero","SkewnessNonZero","KurtosisNonZero","Q5NonZero","Q20NonZero",
"Q50NonZero","Q80NonZero","Q90NonZero","AutocorrelationNonZero1",
"AutocorrelationNonZero2","AutocorrelationNonZero3","AutocovarianceNonZero1",
"AutocovarianceNonZero2","AutocovarianceNonZero3")
}
statslist[[i]]=currmat
statslistNonZero[[i]]=currmatNonZero
}

```

Make data frame with statistical characteristics by month of all values

```

df_statslist=vector("list",data_length)
for(i in 1:data_length){
  df_statslist[[i]]=data.frame(statslist[[i]])
colnames(df_statslist[[i]])=c("Month1","Month2","Month3","Month4","Month5","M
onth6","Month7","Month8","Month9","Month10","Month11","Month12")
}

```

Make data frame with statistical characteristics by month of non zero values

```

df_statslistNonZero=vector("list",data_length)
for(i in 1:data_length){
  df_statslistNonZero[[i]]=data.frame(statslistNonZero[[i]])
colnames(df_statslistNonZero[[i]])=c("Month1","Month2","Month3","Month4","Mo
nth5","Month6","Month7","Month8","Month9","Month10","Month11","Month12")
}

```

#in which stations and months appears Pdr (propability dry)

```

df_Pdr_0=vector("list",data_length)
for(i in 1:data_length){
  df_Pdr_0[[i]]=which(df_statslist[[i]]["Pdr",]!=0)
}

```

Ο κώδικας που χρησιμοποιήθηκε για την προσαρμογή των κατανομών lognormal, gamma και Weibull σε κάθε χρονοσειρά ανά μήνα και για την εξαγωγή των αποτελεσμάτων σε data frame και στη συνέχεια σε .txt file είναι:

#Fit Distributions

Gamma

```
dgamma=function(x,scale,shape) {
  fx=stats::dgamma(x=x,shape=shape,scale=scale)
  return(fx)
}
pgamma=function(q,scale,shape) {
  FX=stats::pgamma(q=q,shape=shape,scale=scale)
  return(FX)
}
qgamma=function(p,scale,shape) {
  x=stats::qgamma(p=p,shape=shape,scale=scale)
  return(x)
}
rgamma=function(n,scale,shape) {
  x=stats::rgamma(n=n,shape=shape,scale=scale)
  return(x)
}
```

Parameter estimation for the Gamma distribution (using L-moments)

```
elmgamma=function(x,plot=0) {
  sam=lmom::samlm(x)
  par=lmom::pelgam(sam)
  names(par)=NULL
  fit=list()
  fit$scale = par[2]
  fit$shape = par[1]

  if (plot==1) {
    u=seq(0,0.999,by = 0.001)
    Td1=qgamma(p=u,scale=fit$scale,shape=fit$shape)
    plot(Td1,1-u,type='l',col='red',log='y',lwd=2,
         yaxt="n",
         xaxt="n",
         xlab=bquote(italic(x)),
         ylab=bquote(P(italic(underline(x)>plain(x))))))
    ticks <- c(0,-1,-2,-3,-4)
    labels <- sapply(ticks, function(i) as.expression(bquote(10^ .(i))))
    axis(2, at=c(1,0.1,0.01,0.001,0.0001), labels=labels)
    labelx=axTicks(1)
    axis(1,at=labelx,labels=format(labelx,scientific=FALSE),1)
  }
}
```

```

points(sort(x),1-ppoints(x),pch=19)
legend('bottomleft',c('Gamma-2','Data'),col=c("red",'black'),
      lwd=c(2,NA), lty=c(1,NA),pch=c(NA,19),
      box.lty=1,
      cex=1)
}

MLE<-sum(log(dgamma(x,scale=fit$scale,shape=fit$shape)))
plotpos<-pp(x=x,a=0,sort=FALSE)
theorquantiles<-qgamma(p=plotpos,scale=fit$scale,shape=fit$shape)
theorquantilessort<-sort(theorquantiles,decreasing=TRUE)[1:10]
samplesort<-sort(x,decreasing=TRUE)[1:10]
MSEquant<-sum((theorquantiles-sort(x))^2)/length(x)
DiffOfMax<-((max(theorquantiles)-max(x))/max(x))*100
MeanDiffOf10Max<-sum(abs(theorquantilessort-samplesort))/10
Eval<-list(MLE=MLE,MSEquant=MSEquant,DiffOfMax=DiffOfMax,
          MeanDiffOf10Max=MeanDiffOf10Max)
Res<-list()
Res$Distribution<-list(FXs="qgamma")
Res$Param<-fit
Res$TheorLMom<-lmrgam(par,nmom=4)
Res$DataLMom<-sam
Res$Eval<-Eval
return(Res)
}

```

Weibull, 3-parameter

identical runs

```

dweibull=function(x,location=0,scale,shape) {
  fx=FAdist::dweibull3(x=x,shape=shape,scale=scale,thres=location)
  return(fx)
}
pweibull=function(q,location=0,scale,shape) {
  FX=FAdist::pweibull3(q=q,shape=shape,scale=scale,thres=location)
  return(FX)
}
qweibull=function(p,location=0,scale,shape) {
  x=FAdist::qweibull3(p=p,shape=shape,scale=scale,thres=location)
  return(x)
}
rweibull=function(n,location=0,scale,shape) {
  x=FAdist::rweibull3(n=n,shape=shape,scale=scale,thres=location)
  return(x)
}

```

Parameter estimation for the Weibull distribution (using L-moments)

```
elmweibull=function(x,bound=0,plot=0) {
  sam=lmom::samlm(x)
  par=lmom::pelwei(sam, bound=bound)
  names(par) = NULL
  fit=list()
  fit$location=par[1]
  fit$scale = par[2]
  fit$shape = par[3]

  if (plot==1) {
    u=seq(0,0.999,by = 0.001)
    Td1=lmom::quawei(f = u, para=unlist(fit))
    plot(Td1,1-u,type='l',col='red',log='y',lwd=2,
        yaxt="n",
        xaxt="n",
        xlab=bquote(italic(x)),
        ylab=bquote(P(italic(underline(x)>plain(x))))))
    ticks <- c(0,-1,-2,-3,-4)
    labels <- sapply(ticks, function(i) as.expression(bquote(10^ .(i))))
    axis(2, at=c(1,0.1,0.01,0.001,0.0001), labels=labels)
    labelx=axTicks(1)
    axis(1,at=labelx,labels=format(labelx,scientific=FALSE),1)
    points(sort(x),1-ppoints(x),pch=19)
    legend('bottomleft',c('Weibull','Data'),col=c("red","black"),
        lwd=c(2,NA), lty=c(1,NA),pch=c(NA,19),
        box.lty=1,
        cex=1)
  }

  MLE<-sum(log(dweibull(x,location=fit$location,scale=fit$scale,shape=fit$shape)))
  plotpos<-pp(x=x,a=0,sort=FALSE)
  theorquantiles<-qweibull(p=plotpos,location=fit$location,
  scale=fit$scale,shape=fit$shape)
  theorquantilessort<-sort(theorquantiles,decreasing=TRUE)[1:10]
  samplesort<-sort(x,decreasing=TRUE)[1:10]
  MSEquant<-sum((theorquantiles-sort(x))^2)/length(x)
  DiffOfMax<-((max(theorquantiles)-max(x))/max(x))*100
  MeanDiffOf10Max<-sum(abs(theorquantilessort-samplesort))/10
  Eval<-list(MLE=MLE,MSEquant=MSEquant,DiffOfMax=DiffOfMax,
  MeanDiffOf10Max=MeanDiffOf10Max)
  Res<-list()
  Res$Distribution<-list(FXs="qweibull",bound=bound)
  Res$Param<-fit
  Res$TheorLMom<-lmrwei(par,nmom=4)
  Res$DataLMom<-sam
  Res$Eval<-Eval
```



```

return(Res)
}

```

LogNormal, 3-parameter

```

dlnorm=function(x,location=0,scale,shape) {
  fx=FAdist::dlnorm3(x=x,shape=shape,scale=scale,thres=location)
  return(fx)
}
plnorm=function(q,location=0,scale,shape) {
  FX=FAdist::plnorm3(q=q,shape=shape,scale=scale,thres=location)
  return(FX)
}
qlnorm=function(p,location=0,scale,shape) {
  x=FAdist::qlnorm3(p=p,shape=shape,scale=scale,thres=location)
  return(x)
}
rlnorm=function(n,location=0,scale,shape) {
  x=FAdist::rlnorm3(n=n,shape=shape,scale=scale,thres=location)
  return(x)
}

```

Parameter estimation for the log-normal distribution (using L-moments)

```

elmInorm=function(x,bound=0,plot=0) {
  sam=lmom::samlnu(x)
  par=lmom::pelln3(sam,bound=bound)
  names(par) = NULL
  fit = list()
  fit$location=par[1]
  fit$scale = par[2]
  fit$shape = par[3]

  if (plot==1) {
    u=seq(0,0.999,by = 0.001)
    Td1=lmom::qualn3(f = u, para=unlist(fit))
    plot(Td1,1-u,type='l',col='red',log='y',lwd=2,
         yaxt="n", xaxt="n", xlab=bquote(italic(x)),
         ylab=bquote(P(italic(underline(x)>plain(x))))))
    ticks <- c(0,-1,-2,-3,-4)
    labels <- sapply(ticks, function(i) as.expression(bquote(10^ .(i))))
    axis(2, at=c(1,0.1,0.01,0.001,0.0001), labels=labels)
    labelx=axTicks(1)
    axis(1,at=labelx,labels=format(labelx,scientific=FALSE),1)
    points(sort(x),1-ppoints(x),pch=19)
    legend('bottomleft',c('Lognormal','Data'),col=c("red","black"),
          lwd=c(2,NA), lty=c(1,NA),pch=c(NA,19),

```

```

    box.lty=1,
    cex=1)
}

MLE<-sum(log(dlnorm(x,location=fit$location,scale=fit$scale,shape=fit$shape)))
plotpos<-pp(x=x,a=0,sort=FALSE)
theorquantiles<-qlnorm(p=plotpos,location=fit$location,
scale=fit$scale,shape=fit$shape)
theorquantilessort<-sort(theorquantiles,decreasing=TRUE)[1:10]
samplesort<-sort(x,decreasing=TRUE)[1:10]
MSEquant<-sum((theorquantiles-sort(x))^2)/length(x)
DiffOfMax<-((max(theorquantiles)-max(x))/max(x))*100
MeanDiffOf10Max<-sum(abs(theorquantilessort-samplesort))/10
Eval<-list(MLE=MLE,MSEquant=MSEquant,DiffOfMax=DiffOfMax,
MeanDiffOf10Max=MeanDiffOf10Max)
Res<-list()
Res$Distribution<-list(FXs="qlnorm",bound=bound)
Res$Param<-fit
Res$TheorLMom<-lmrln3(par,nmom=4)
Res$DataLMom<-sam
Res$Eval<-Eval
return(Res)
}

```

Demonstration of function

Now we fit three distributions, namely: Gamma, Lognormal and Weibull

If you don't want the plot, set: plot = 0

#fit_gam=elmgamma(x = x, plot = 1) # Fit the Gamma distribution

#fit_Inrom=elmlnorm(x = x, plot = 1) # Fit the Lognormal distribution

#fit_wei=elmweibull(x = x, plot = 1) # Fit the Weibull distribution

Among the three "fits", select the distribution with lowest quantiles Mean Squared Error (MSE)

#fit_gam\$Eval\$MSEquant

#fit_wei\$Eval\$MSEquant

#fit_Inrom\$Eval\$MSEquant

#Fitting gamma, weibull, lognormal distributions

distrlist=vector("list",data_length)

for(i in 1:data_length){

 for(j in 1:12){

 flow_data_per_monthz=flow_data_per_month_list[[i]][[j]]

 flow_data_per_monthz=coredata(flow_data_per_monthz)

 flow_data_per_monthz=na.omit(flow_data_per_monthz)

 flow_data_per_monthz=flow_data_per_monthz[flow_data_per_monthz>0]

```

fit_gam=elmgamma(x = as.vector(flow_data_per_monthz), plot = 0) # Fit the Gamma
distribution
fit_wei=elmweibull(x = as.vector(flow_data_per_monthz), plot = 0) # Fit the Weibull
distribution
fit_Inorm=elmlnorm(x = as.vector(flow_data_per_monthz), plot = 0) # Fit the
Lognormal distribution
gam=fit_gam$Eval$MSEquant
wei=fit_wei$Eval$MSEquant
lnorm=fit_Inorm$Eval$MSEquant
results=cbind(namelist[i],j,gam,wei,lnorm)
colnames(results)=c("Station","Month","gamma","weibull","lognormal")
distrlist[[i]][[j]]=results
}
}

```

#data frame with fitting results per time series data

```

df_distrlist=vector("list",data_length)
for(i in 1:data_length){
  for(j in 1:12){
    df_distrlist[[i]]=rbind(df_distrlist[[i]],distrlist[[i]][[j]])
  }
}
df_distrlist=as.data.frame(df_distrlist)

```

#make data frame with possibility success every distribution

```

df_dist_wins=vector("list",data_length)
df_dist_gamma=vector("list",data_length)
df_dist_weibull=vector("list",data_length)
for(i in 1:data_length){
  for(j in 1:12){
    df_distrlist[[i]]=as.data.frame(df_distrlist[[i]])
    which_month=which(df_distrlist[[i]]["Month"]==j)
    df_distrlist_month=df_distrlist[[i]][which_month,]
    df_distrlist_month$Min=apply(df_distrlist_month[,c("gamma","weibull","lognormal"
)],1,min)
    count_gamma=length(which(df_distrlist_month$Min==df_distrlist_month$gamma))
    count_wei=length(which(df_distrlist_month$Min==df_distrlist_month$weibull))
    count_Inorm=length(which(df_distrlist_month$Min==df_distrlist_month$lognormal)
)
    if(count_gamma==max(count_gamma,count_wei,count_Inorm)){
      print(paste0("gamma"))
    } else if(count_wei==max(count_gamma,count_wei,count_Inorm)){
      print(paste0("wei"))
    } else {
      print(paste0("lognormal"))
    }
  }
}

```

```

df_dist_wins[[i]][j]=count_Inorm/sum(count_gamma+count_wei+count_Inorm)*100
df_dist_gamma[[i]][j]=count_gamma/sum(count_gamma+count_wei+count_Inorm)*100
df_dist_weibull[[i]][j]=count_wei/sum(count_gamma+count_wei+count_Inorm)*100
}
}

```

```

df_dist_3l=c()
df_dist_3g=c()
df_dist_3w=c()
for(i in 1:data_length){
  df_dist_wins[[i]]=as.data.frame(df_dist_wins[[i]])

  colnames(df_dist_wins[[i]])=c("Month1","Month2","Month3","Month4","Month5","Month6","Month7","Month8","Month9","Month10","Month11","Month12")
  df_dist_3l=rbind(df_dist_3l,df_dist_wins[[i]])
  df_dist_gamma[[i]]=as.data.frame(df_dist_gamma[[i]])
  colnames(df_dist_gamma[[i]])=c("Month1","Month2","Month3","Month4","Month5","Month6","Month7","Month8","Month9","Month10","Month11","Month12")
  df_dist_3g=rbind(df_dist_3g,df_dist_gamma[[i]])
  df_dist_weibull[[i]]=as.data.frame(df_dist_weibull[[i]])
  colnames(df_dist_weibull[[i]])=c("Month1","Month2","Month3","Month4","Month5","Month6","Month7","Month8","Month9","Month10","Month11","Month12")
  df_dist_3w=rbind(df_dist_3w,df_dist_weibull[[i]])
}

```

#Export data.frame to .txt file

```

write.table(df_dist_3w, file = "C:/Users/user/Desktop/weibull.txt", append = FALSE,
  quote = TRUE, sep = " ", eol = "\n", na = "NA", dec = ".", row.names = TRUE,
  col.names = TRUE)

```

Ο κώδικας που χρησιμοποιήθηκε για τον υπολογισμό των στατιστικών χαρακτηριστικών κάθε χρονοσειράς ετήσια και για την εξαγωγή των αποτελεσμάτων σε data frame και στη συνέχεια σε .txt file είναι:

```
# Calculating statistical characteristics of timeseries yearly
statslist_ts=vector("list",data_length)
flow_data_per_ts_list=vector("list",data_length)
maxlag=3
for(i in 1:data_length){
  flow_data_ts=xtslist[[i]]
  currmat_ts=c()
  df_ts=as.data.frame(flow_data_ts)
  df_ts=na.omit(df_ts)
  df_flow_data_ts=df_ts$V1
  Pdr_ts<-round(mean(df_flow_data_ts<=0.01,na.rm=T),5)
  lmom_ts=lmom.ub(df_flow_data_ts)
  LMean_ts=lmom_ts$L1
  LScale_ts=lmom_ts$L2
  L3_ts=lmom_ts$L3
  L4_ts=lmom_ts$L4
  LVariation_ts=lmom_ts$LCV
  LSkewness_ts=lmom_ts$TAU3
  Lkurtosis_ts=lmom_ts$TAU4
  Min_ts=round(min(df_flow_data_ts,na.rm=TRUE),4)
  Max_ts=round(max(df_flow_data_ts,na.rm=TRUE),4)
  Mean_ts=round(mean(df_flow_data_ts,na.rm=TRUE),4)
  Var_ts=round(var(df_flow_data_ts,na.rm=TRUE),4)
  StDev_ts=round(sd(df_flow_data_ts,na.rm=TRUE),4)
  Variation_ts=round(StDev_ts/Mean_ts,4)
  Mom3_ts=round(sum((df_flow_data_ts-Mean_ts)^3,
  na.rm=T)*(1/(sum(!is.na(df_flow_data_ts))-1)),5)
  Skewness_ts=-round(Mom3_ts/StDev_ts^3,4)
  Kurtosis_ts=round(moments::kurtosis(df_flow_data_ts,na.rm=TRUE))
  Q5_ts<-round(quantile(df_flow_data_ts,probs=c(0.05),
  na.rm=TRUE,names=FALSE),5)
  Q20_ts<-round(quantile(df_flow_data_ts,probs=c(0.20),
  na.rm=TRUE,names=FALSE),5)
  Q50_ts<-round(quantile(df_flow_data_ts,probs=c(0.50),
  na.rm=TRUE,names=FALSE),5)
  Q80_ts<-round(quantile(df_flow_data_ts,probs=c(0.80),
  na.rm=TRUE,names=FALSE),5)
  Q90_ts<-round(quantile(df_flow_data_ts,probs=c(0.90),
  na.rm=TRUE,names=FALSE),5)
  Autocorrelation_ts<-acf(df_flow_data_ts,lag.max=maxlag,type="correlation",
  plot=FALSE,na.action = na.pass)
  Autocorrelation_ts<-as.vector(Autocorrelation_ts[[1]])[2:(maxlag+1)]
```

```

Autocovariance_ts<-acf(df_flow_data_ts,lag.max=maxlag,type="covariance",
plot=FALSE,na.action = na.pass)
Autocovariance_ts<-as.vector(Autocovariance_ts[[1]])[2:(maxlag+1)]
Description_Analysis_ts=c(LMean_ts,LScale_ts,L3_ts,L4_ts,LVariation_ts,LSkewness_
ts,Lkurtosis_ts,Min_ts,Max_ts,Mean_ts,Var_ts,StDev_ts,Variation_ts,Mom3_ts,
Skewness_ts,Kurtosis_ts,Q5_ts,Q20_ts,Q50_ts,Q80_ts,Q90_ts,Pdr_ts,Autocorrelatio
n_ts,Autocovariance_ts)
currmat_ts=cbind(currmat_ts,Description_Analysis_ts)
rownames(currmat_ts)=c("LMean","LScale","L3","L4","LVariation","LSkewness",
"Lkurtosis","Min","Max","Mean","Var","StDev","Variation","Mom3",
"Skewness","Kurtosis","Q5","Q20","Q50","Q80","Q90","Pdr","Autocorrelation1",
"Autocorrelation2","Autocorrelation3","Autocovariance1","Autocovariance2",
"Autocovariance3")
flow_data_per_ts_list[[i]]=flow_data_ts
statslist_ts[[i]]=currmat_ts
}

```

#Mean for all time series data

```

df_Mean_ts=c()
for(i in 1:data_length){
df_Mean_ts=rbind(df_Mean_ts,statslist_ts[[i]]["Mean",])
}
df_Mean_ts=as.data.frame(df_Mean_ts)

```

#Stdev for all time series data

```

df_StDev_ts=c()
for(i in 1:data_length){
df_StDev_ts=rbind(df_StDev_ts,statslist_ts[[i]]["StDev",])
}
df_StDev_ts=as.data.frame(df_StDev_ts)

```

#Skewness for all time series data

```

df_skewness_ts=c()
for(i in 1:data_length){
df_skewness_ts=rbind(df_skewness_ts,statslist_ts[[i]]["Skewness",])
}
df_skewness_ts=as.data.frame(df_skewness_ts)

```

#Q5 for all time series data

```

df_Q5_ts=c()
for(i in 1:data_length){
df_Q5_ts=rbind(df_Q5_ts,statslist_ts[[i]]["Q5",])
}
df_Q5_ts=as.data.frame(df_Q5_ts)

```

#Q20 for all time series data

```
df_Q20_ts=c()
for(i in 1:data_length){
  df_Q20_ts=rbind(df_Q20_ts,statslist_ts[[i]]["Q20",])
}
df_Q20_ts=as.data.frame(df_Q20_ts)
```

#Q50 for all time series data

```
df_Q50_ts=c()
for(i in 1:data_length){
  df_Q50_ts=rbind(df_Q50_ts,statslist_ts[[i]]["Q50",])
}
df_Q50_ts=as.data.frame(df_Q50_ts)
```

#Q80 for all time series data

```
df_Q80_ts=c()
for(i in 1:data_length){
  df_Q80_ts=rbind(df_Q80_ts,statslist_ts[[i]]["Q80",])
}
df_Q80_ts=as.data.frame(df_Q80_ts)
```

#Q90 for all time series data

```
df_Q90_ts=c()
for(i in 1:data_length){
  df_Q90_ts=rbind(df_Q90_ts,statslist_ts[[i]]["Q90",])
}
df_Q90_ts=as.data.frame(df_Q90_ts)
```

#Autocorrelation1 for all time series data

```
df_lag1_ts=c()
for(i in 1:data_length){
  df_lag1_ts=rbind(df_lag1_ts,statslist_ts[[i]]["Autocorrelation1",])
}
df_lag1_ts=as.data.frame(df_lag1_ts)
```

#Autocorrelation2 for all time series data

```
df_lag2_ts=c()
for(i in 1:data_length){
  df_lag2_ts=rbind(df_lag2_ts,statslist_ts[[i]]["Autocorrelation2",])
}
df_lag2_ts=as.data.frame(df_lag2_ts)
```

#Autocorrelation3 for all time series data

```
df_lag3_ts=c()
for(i in 1:data_length){
  df_lag3_ts=rbind(df_lag3_ts,statslist_ts[[i]]["Autocorrelation3",])
}
```

```
df_lag3_ts=as.data.frame(df_lag3_ts)

df_teliko_ts=cbind(df_Mean_ts,df_StDev_ts,df_skewness_ts,df_Q5_ts,df_Q20_ts,
  df_Q50_ts,df_Q80_ts,df_Q90_ts,df_lag1_ts,df_lag2_ts,
  df_lag3_ts)

#Export data.frame to .txt file
write.table(df_teliko_ts, file = "C:/Users/user/Desktop/Results_ts.txt", append =
FALSE,quote = TRUE,sep = " ",eol = "\n", na = "NA", dec = ",", row.names = TRUE,
col.names = TRUE)

# CLEAN UP

# Clear environment
rm(list = ls())

# Clear plots
dev.off()

# Clear console
cat("\014")

# Clear mind :)
```


ΠΑΡΑΡΤΗΜΑ ΙΙΙ

