



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΔΙΕΠΙΣΤΗΜΟΝΙΚΟ - ΔΙΑΤΜΗΜΑΤΙΚΟ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΕΠΙΣΤΗΜΗ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ
ΥΔΑΤΙΚΩΝ ΠΟΡΩΝ»

**ΔΙΕΡΕΥΝΗΣΗ ΜΟΝΤΕΛΩΝ ΠΟΛΛΑΠΛΗΣ
ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ ΓΙΑ ΤΗ
ΧΩΡΙΚΗ ΕΚΤΙΜΗΣΗ ΤΗΣ ΒΡΟΧΟΠΤΩΣΗΣ ΣΤΗΝ
ΑΤΤΙΚΗ**

Αθηνά Παππά

Επιβλέπων: Ευάγγελος Μπαλτάς, Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2022



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΔΙΕΠΙΣΤΗΜΟΝΙΚΟ - ΔΙΑΤΜΗΜΑΤΙΚΟ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

«ΕΠΙΣΤΗΜΗ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ ΥΔΑΤΙΚΩΝ ΠΟΡΩΝ»

ΔΙΕΡΕΥΝΗΣΗ ΜΟΝΤΕΛΩΝ ΠΟΛΛΑΠΛΗΣ ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ ΓΙΑ ΤΗ ΧΩΡΙΚΗ ΕΚΤΙΜΗΣΗ ΤΗΣ ΒΡΟΧΟΠΤΩΣΗΣ ΣΤΗΝ ΑΤΤΙΚΗ

Αθηνά Παππά

Επιβλέπων: Ευάγγελος Μπαλτάς, Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβρης 2022

ΕΥΧΑΡΙΣΤΙΕΣ

Ολοκληρώνοντας τον κύκλο των μεταπτυχιακών μου σπουδών, μέσω της παρούσας εργασίας, οφείλω και θέλω να ευχαριστήσω αυτούς που με στήριξαν.

Πρωτίστως, θα ήθελα να ευχαριστήσω τον επί σειρά ετών καθηγητή μου και επιβλέποντα της παρούσας και της προηγούμενης διπλωματικής μου εργασίας κ. Ευάγγελο Μπαλτά, για την ανάθεση του ερευνητικού θέματος με το οποίο ασχολήθηκα. Αλλά κυρίως τον ευχαριστώ θερμά, για την άψογη συνεργασία μας, με την ευγένεια και την αμεσότητα που σαν άνθρωπο τον διακρίνουν και για την διαρκή επιστημονική καθοδήγηση που μου παρείχε, καθ' όλη τη διάρκεια φοίτησης μου πίσω από τα έδρανα του Πολυτεχνείου.

Ιδιαίτερες ευχαριστίες οφείλω επίσης να απευθύνω στην Διδάκτορα Ελισσάβητ Φελώνη, για την συνεχή ενασχόληση της και την ειλικρινώς άμεση ανταπόκριση της, σε οποιοδήποτε ζήτημα ανέκυπτε, καθώς και για τις πολύτιμες συμβουλές και διορθώσεις της, για την αποπεράτωση του παρόντος τεύχους.

Ευχαριστώ επίσης, το τρίτο μέλος της εξεταστικής επιτροπής και Αναπληρωτή Καθηγητή, κ. Μαλαμή, διδάσκοντα του Δ.Π.Μ.Σ, το διδακτικό έργο του οποίου εκτιμώ ειλικρινά.

Κλείνοντας, θέλω να ευχαριστήσω και να αφιερώσω τη συγκεκριμένη δουλειά σε όλους τους κοντινούς μου ανθρώπους και φίλους, που μου στάθηκαν ανιδιοτελώς και με βοήθησαν με κάθε μέσο, σε αυτή τη δύσκολη για μένα περίοδο.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ	v
EXTENDED SUMMARY	vi
Introduction.....	vi
Study Area and Data used	vi
Study Area: Athens basin, Attica region	vi
Raw Data: Precipitation Time Series and Spatial Data	viii
Methods	ix
Results	xiv
Conclusions and Recommendations for Future Research.....	xxvi
Selected References	xxvii
1 ΕΙΣΑΓΩΓΗ.....	1
1.1 Αντικείμενο Εργασίας και Εισαγωγικά Στοιχεία	1
1.2 Πρωτότυπα σημεία	2
1.3 Διάρθρωση εργασίας	3
2 ΠΕΡΙΟΧΗ ΜΕΛΕΤΗΣ ΚΑΙ ΠΡΩΤΟΓΕΝΗ ΔΕΔΟΜΕΝΑ.....	4
2.1 Εισαγωγή	4
2.2 Χαρακτηριστικά Περιοχής Μελέτης.....	4
2.2.1 Γεωγραφικά και Γεωμορφολογικά Χαρακτηριστικά της Περιοχής Μελέτης.....	4
2.2.2 Κλίμα.....	6
2.3 Διαθέσιμα Πρωτογενή Δεδομένα	10
2.3.1 Βροχομετρικά Δεδομένα	10
2.3.2 Γεωχωρικά Δεδομένα	12
3 ΜΕΘΟΔΟΛΟΓΙΑ.....	13
3.1 Γενική Επισκόπηση	13
3.2 Αναλυτική Περιγραφή	14
3.2.1 Επιλογή βροχομετρικών σταθμών	14
3.2.2 Επεξεργασία πρωτογενών χρονοσειρών.....	15
3.2.3 Επιλογή ανεξάρτητων μεταβλητών και επεξεργασία γεωχωρικών δεδομένων.....	20
3.2.4 Συμπλήρωση χρονοσειρών με την μέθοδο Kriging.....	21

3.2.5	Ανάπτυξη μοντέλων παλινδρόμησης.....	25
3.2.6	Καθορισμός κριτηρίων αξιολόγησης των μοντέλων.....	28
4	ΑΠΟΤΕΛΕΣΜΑΤΑ.....	38
4.1	Συμπλήρωση και ομαδοποίηση βροχομετρικών δεδομένων.....	38
4.2	Τιμές Ανεξάρτητων Μεταβλητών.....	43
4.3	Μοντέλα Παλινδρόμησης.....	47
4.4	Στατιστικοί δείκτες και γραφήματα στατιστικά σημαντικών μοντέλων.....	50
4.4.1	Μοντέλο Παλινδρόμησης Οκτωβρίου.....	51
4.4.2	Μοντέλο Παλινδρόμησης Δεκεμβρίου.....	54
4.4.3	Μοντέλο Παλινδρόμησης Έτους.....	56
4.4.4	Μοντέλο Παλινδρόμησης Ξηρής Περιόδου.....	58
4.4.5	Μοντέλο Παλινδρόμησης Υγρής Περιόδου.....	60
4.4.6	Συγκριτικά Αποτελέσματα.....	62
5	ΣΥΜΠΕΡΑΣΜΑΤΑ – ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ.....	70
5.1	Σύνοψη.....	70
5.2	Συμπεράσματα.....	70
5.3	Μελλοντική Έρευνα.....	72
6	ΒΙΒΛΙΟΓΡΑΦΙΑ.....	74

ΠΕΡΙΛΗΨΗ

Η παρούσα μεταπτυχιακή εργασία αποτελεί μία πρώτη προσπάθεια διαμόρφωσης και εφαρμογής ενός μεθοδολογικού πλαισίου για την διερεύνηση και ανάπτυξη μοντέλων παλινδρόμησης με στόχο την χωρική εκτίμηση βροχόπτωσης, στην περιοχή μελέτης του λεκανοπεδίου της Αττικής.

Η επιλογή των βροχομετρικών σταθμών προς εξέταση έγινε με κριτήριο την διαθεσιμότητα, την ποιότητα των δεδομένων και την ομοιόμορφη χωρική κατανομή εντός και πλησίον της περιοχής μελέτης του λεκανοπεδίου. Για τον σκοπό αυτό αξιοποιήθηκαν τα βροχομετρικά δεδομένα από ένα σύνολο 21 σταθμών, αντιπροσωπευτικά κατανεμημένων στην περιοχή έρευνας, μέρος των οποίων ανήκουν στο δίκτυο σταθμών του Υδρολογικού Παρατηρητήριου Αθήνας (δίκτυο ΗΟΑ) και οι λοιποί στο δίκτυο σταθμών του Εθνικού Αστεροσκοπείου Αθηνών (δίκτυο ΝΟΑ).

Η μεθοδολογία που ακολουθήθηκε διακρίνεται σε τρία βασικά σκέλη: α) τον έλεγχο, την ομαδοποίηση και την συμπλήρωση των χρονοσειρών βροχόπτωσης, β) τον καθορισμό των ανεξάρτητων μεταβλητών προς εξέταση και τον προσδιορισμό των τιμών τους μέσω επεξεργασίας γεωχωρικής πληροφορίας και γ) την ανάπτυξη και διερεύνηση μοντέλων πολλαπλής γραμμικής παλινδρόμησης και την αξιολόγηση της μέσω κατάλληλων στατιστικών δεικτών, βαθμωτών μέτρων ακριβείας και διαγραμμάτων.

Αρχικά, τα διαθέσιμα πρωτογενή δεδομένα χρονοσειρών βροχόπτωσης χρονικού βήματος δεκαλέπτου, εισάχθηκαν, ελέγχθηκαν ως προς την ύπαρξη κενών καταγραφών, επεξεργάστηκαν και ομαδοποιήθηκαν σε μηνιαίο χρονικό βήμα καταλλήλως, μέσω της σύνταξης και εκτέλεσης κώδικα στο προγραμματιστικό περιβάλλον MATLAB. Παράλληλα η συμπλήρωση των χρονοσειρών, όπου αυτό κρίθηκε απαραίτητο, έγινε μέσω της μεθόδου γεωστατιστικής τεχνικής παρεμβολής – Kriging, εντός συστήματος γεωγραφικών πληροφοριών (GIS) και πιο συγκεκριμένα μέσω του λογισμικού ArcGIS.

Μέσω της ολοκληρωμένης βιβλιογραφικής επισκόπησης που επιχειρήθηκε, καθορίστηκαν οι προς εξέταση ανεξάρτητες μεταβλητές που ενδέχεται να συσχετίζονται με το μέγεθος της βροχόπτωσης και στη συνέχεια αξιοποιήθηκε ο αλγόριθμος Stepwise Procedure – Backward Elimination για την εύρεση του βέλτιστου μοντέλου πολλαπλής γραμμικής παλινδρόμησης.

Από τα εξεταζόμενα μοντέλα, επιλέγονται τα στατιστικώς σημαντικά με την βοήθεια του παρατηρούμενου επιπέδου σημαντικότητας (p -value). Τα ίδια στη συνέχεια αξιολογούνται, βάσει στατιστικών δεικτών και διαγραμμάτων που σύμφωνα με τη βιβλιογραφία χρησιμοποιούνται σε ανάλογες εφαρμογές. Τέλος, παράγονται τα αντίστοιχα γεωχωρικά επίπεδα εκτίμησης βροχόπτωσης για την περιοχή μελέτης στο περιβάλλον του λογισμικού ArcGIS.

Λέξεις κλειδιά: πολλαπλή γραμμική παλινδρόμηση, χωρική παλινδρόμηση, βροχόπτωση, μηνιαία βροχόπτωση, Αττική

EXTENDED SUMMARY

Introduction

Knowledge and understanding of temporal and spatial patterns of precipitation, on a climate scale, is essential for forecasting and protection against extreme flood events, as well as for drought monitoring, general atmosphere circulation and the proper management of watersheds. In recent past, meteorological data were obtained by rain gauges. Rain gauges are capable of providing high accuracy point precipitation data, however these data refer to the narrow area of each rain gauge and, therefore, they are spatially limited. Thus estimation and modeling rainfall via suitable methods and corresponding spatial variability are a very challenging scientific field.

This study aims to develop and implement an integrated methodological framework for the spatial estimation rainfall in different periods (monthly, seasonal and annual), which is also useful for downscaling, through regression models application. This approach emphasizes the evaluation process of the individual statistical models. This research is conducted in a sub-area of Attica region, named Athens basin, an area characterised by intense urbanisation and variability in terms of its geomorphology and meteorological/climatological characteristics.

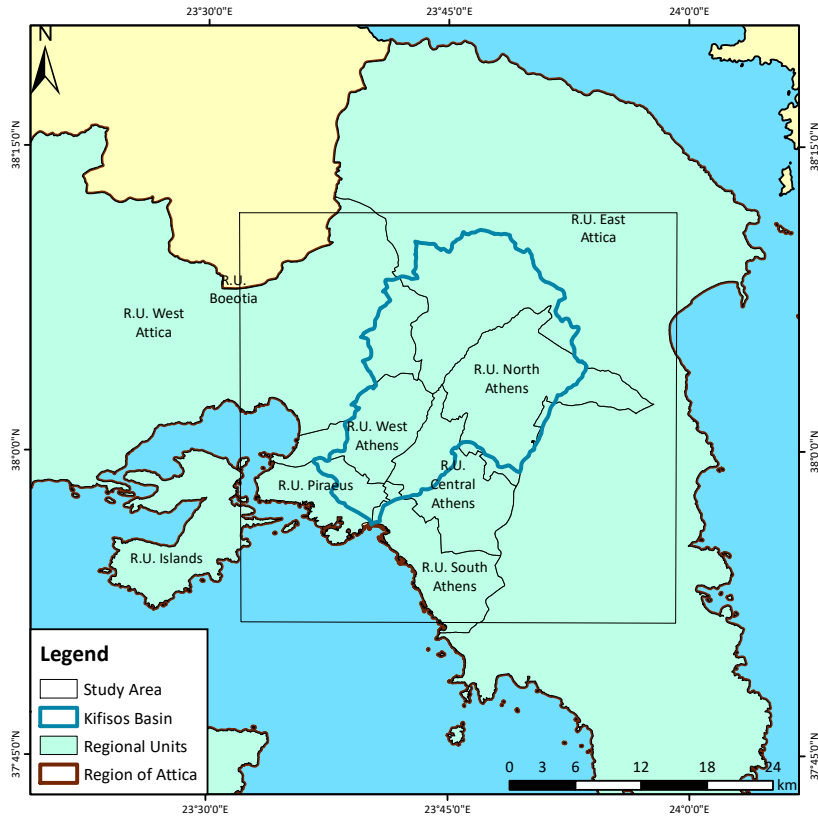
Keywords: multiple linear regression, spatial regression, precipitation, monthly rainfall, Attica

Study Area and Data used

Study Area: Athens basin, Attica region

Attica region is an administrative unit of Greece that encompasses the entire metropolitan area of Athens, which is the largest city and the country's capital. The study area selected includes a central continental part of Attica Region, enclosed by the frame, as shown in the following map (1Figure 1). The same area contains sub-sections of all the individual Regional Units of Attica, as well as the entire watershed of Kifisos river.

Relief of Attica region varies from mainly plain to semi-mountainous lands. The main mountains and hills located in the area are shown in map of Figure 2. Kifisos basin is surrounded by the mountain ranges of Parnitha to the North, Pentelis mountain (East), Aegaleo (West) and Hymettus in the south-eastern part. Parnitha, with a maximum altitude of 1413 m, is the highest of the mountains of Attica Region. Coastline in southern and western parts of the region are surrounded by Saronic Gulf, whilst at the eastern part, by the Southern Euboean Gulf.



1Figure 1 Administrative boundaries of regional units of Athens, Kifisos basin and study area

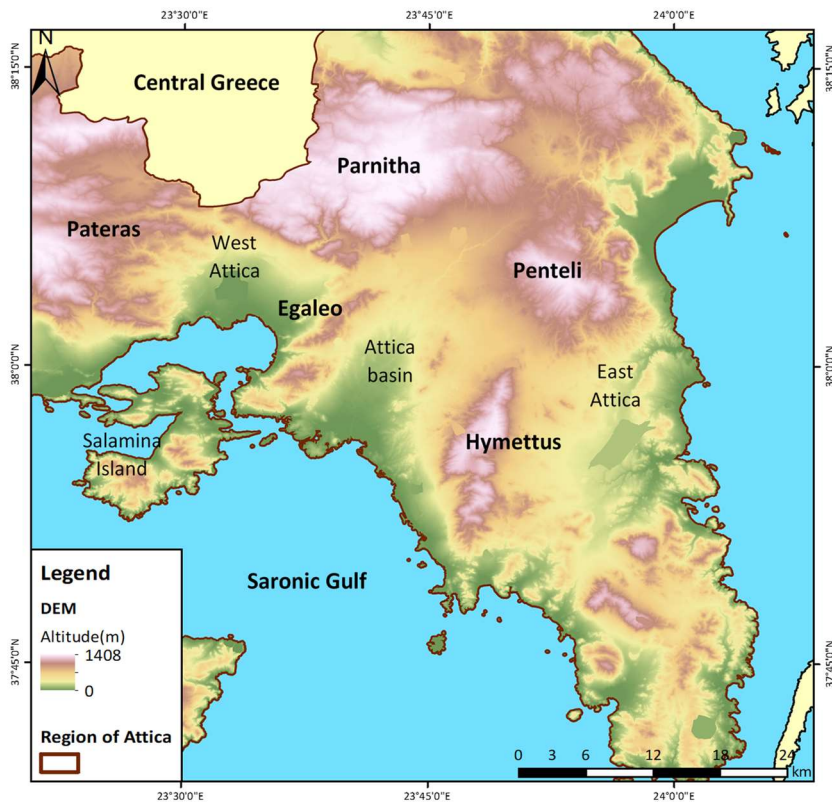


Figure 2 Terrain relief in the Region of Attica

Attica has a Mediterranean climate according to the Köppen Classification, with long periods of intermittent rainfall and temperature fluctuations during the cold period of the year, while in the warm period a steadier weather pattern prevails (Maheras & Anagnostopoulou, 2003), with intense convective precipitation episodes (Apostolopoulos., 2014).

This climatic pattern, combined with the aforementioned geomorphology, creates a set of river water bodies with torrential characteristics, which, due to intense urbanization, often lead to an increased probability of flooding events (Feloni et al., 2020). Particularly for the Kifissos river basin, it is estimated that the built-up area exceeds 70%, which is constantly increasing, while the mountainous and semi-mountainous areas of the basin in Parnitha and Penteli have occasionally been affected by fires, with the most destructive one in the summer of 2007 (Koutsogiannis et al., 2010).

Raw Data: Precipitation Time Series and Spatial Data

In order to conduct spatial rainfall estimation models for the study area, primary precipitation data of fine temporal scale are collected, using stations from a dense and spatially uniform network of rain gauges that operate in the same area. Initially, raw rainfall data were obtained for 30 telemetric hydrometeorological stations. Each one of these stations belongs either to the Hydrological Observatory of Athens network (HOA) (Lagouvardos et al., 2017) or to the National Observatory of Athens network (NOA) (Grammatikogiannis et al., 2005). In the Table 1, the selected stations, their origin and their coordinates in the Hellenic Geodetic Reference System 1987 (Greek Grid/ EGSA'87) are listed. The primary rainfall time-series of each station, regardless its network, have a ten-minute temporal resolution, while the acquisition period differs.

For the purpose of this study, a set of geospatial data were also collected and used. The main spatial dataset used is the layer of Digital Elevation Model (DEM), provided by the National Agency for Cadastre & Mapping S.A. (Copyright © 2012), with a spatial accuracy of a 5m x 5m cell size.

In addition to the digital terrain model, the geospatial polygon encompassing the Kifissos River Basin was used (Baltas et al. 2015), as well as, the polygon layers of the coastline and the administrative boundaries of the Attica Region, as provided by the National Open Geospatial Data Catalogue (geodata.gov.gr)

Table 1 Location of available rain gauges, latitude, longitude and their network origin

Location	X	Y	Network
Ampelokipoi	478631	4203530	NOA
Anavyssos	491500	4175991	NOA
Athens	474861	4203170	NOA
Dionysos	490196	4217492	NOA
Ekali	484311	4215491	NOA
Faliro	472875	4197735	NOA
Ippokrateios	481516	4227284	NOA
Kantza	488048	4203239	NOA
Lavrio	504963	4174086	NOA
Markopoulo	494311	4191882	NOA
Marousi	483450	4211276	NOA
Nea Makri	497811	4212060	NOA
Nea Smyrni	474805	4200123	NOA
Parnitha	475189	4223648	NOA
Patisia	476134	4207998	NOA
Peristeri	473845	4205768	NOA
Porto Rafti	501228	4192916	NOA
Psychico	480593	4207503	NOA
Spata	493449	4203828	NOA
Vari	482516	4187017	NOA
Ag. Kosmas	475060	4193690	HOA
Ano Liosia	470968	4213918	HOA
Galatsi	478564	4208803	HOA
Ilioupoli	478838	4196512	HOA
Mandra	461615	4219262	HOA
Menidi	476080	4217305	HOA
Penteli	487813	4215150	HOA
Pikermi	493583	4205666	HOA
Psittalia	464231	4199302	HOA
Zografou	480493	4203306	HOA

Methods

After the initial data collection, as described above, the final selection of the stations to be studied follows, depending on their spatial location and data quality. The available time series are clustered to convert them into monthly time series, while any missing values are detected. The missing values were then filled in, using the Kriging geostatistical interpolation technique, applied in the ArcGIS software. The same software is used to spatially express the values of the independent variables corresponding to the location of each station, since the set of independent variables to be studied has already been determined after relevant literature review. Finally, with the aggregated data and through regression

analysis, the rainfall estimation models are derived, which are evaluated through statistical coefficients and charts. The general methodology framework for the purposes of this thesis is described in the following flowchart (Figure 3).

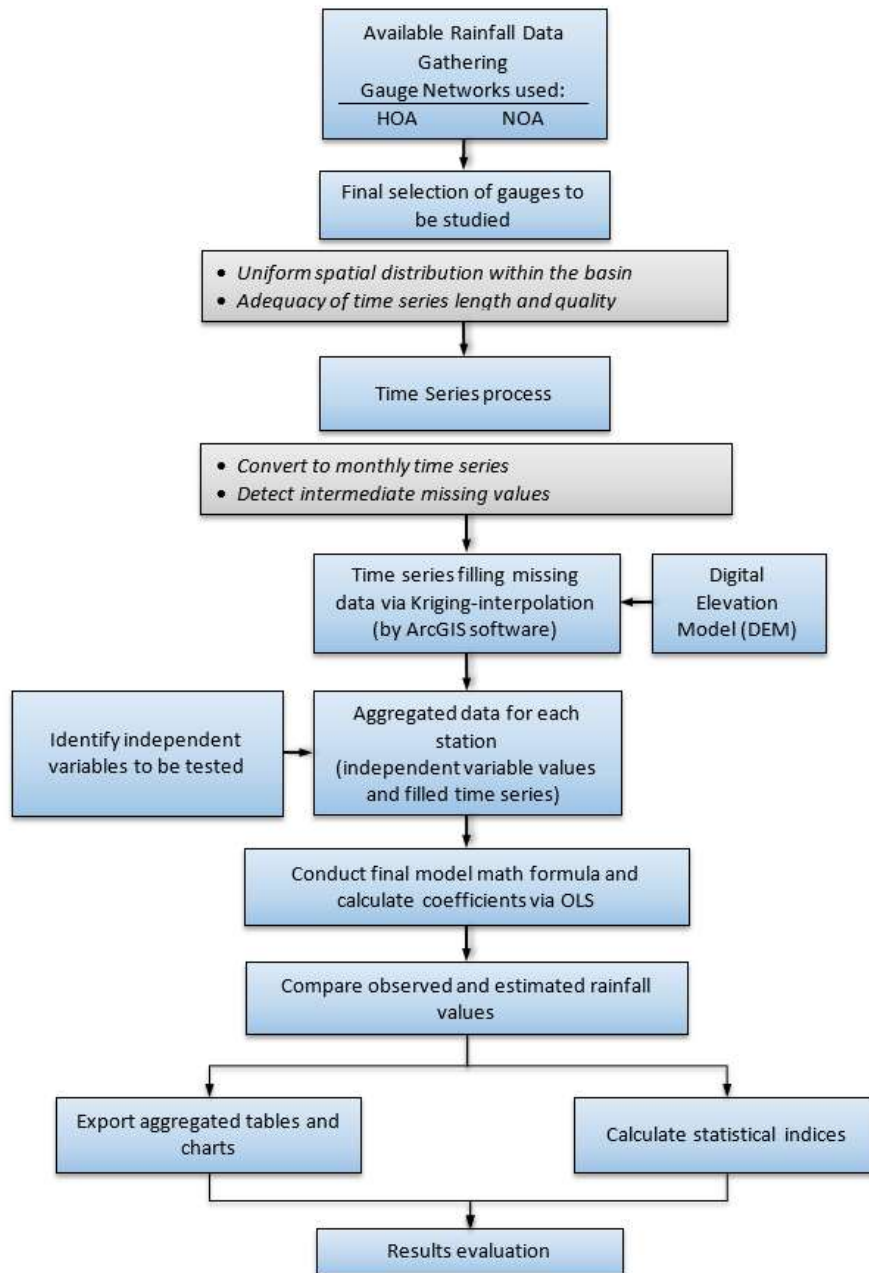


Figure 3 Flowchart of general methodology followed

More specifically, the selection of the rain gauges and their time series to be processed was based on the time range of the available recordings of each station, as well as their continuity. Stations with partial recordings and many missing values in the time series were considered unsuitable, since even their subsequent completion would lead to data with strong uncertainty, which would be inappropriate

for the generation of subsequent rainfall estimation models. At the same time, another significant criterion was to achieve a spatially well-defined network of stations with as uniform spatial distribution as possible, in and around the Kifisos river basin. Selected and rejected rain gauges and their network origin are illustrated in the following map (Figure 4).

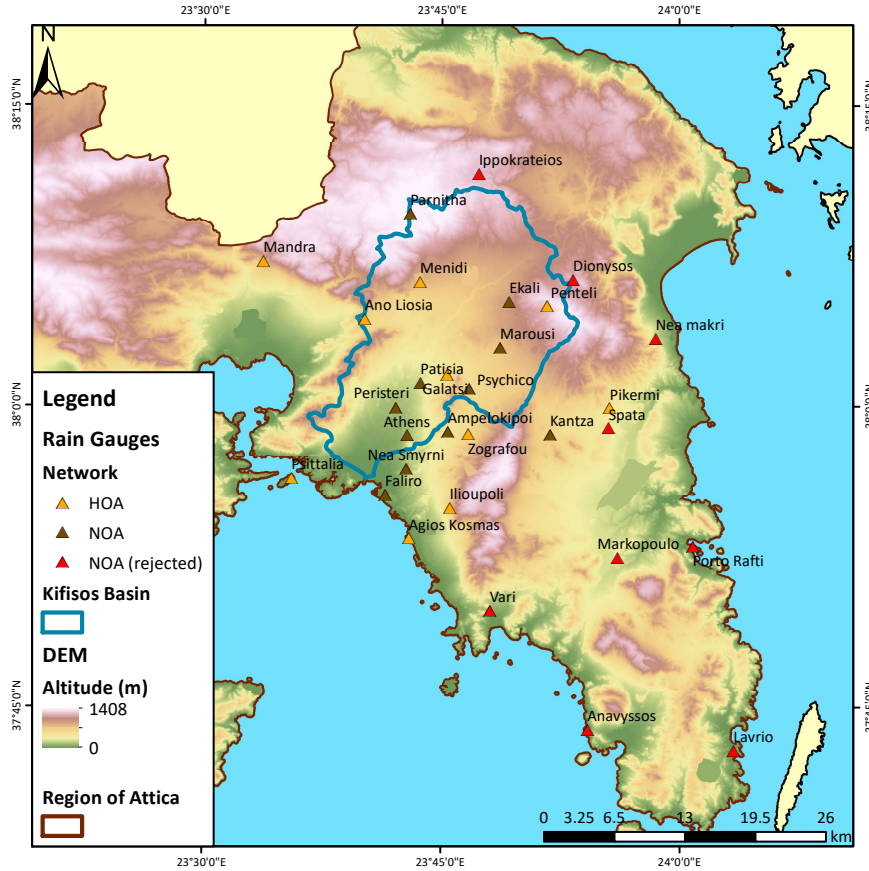


Figure 4 HOA & NOA rain gauge networks, rejected stations and boundaries of Kifisos basin.

After the final selection of stations, their time series are processed. The processing of the primary rainfall data, in the form of time series, was done entirely through writing and executing code using MATLAB programming environment. The available time series have a time step of 10 minutes, which is a high accuracy data scale. However, due to the large number of intermediate gaps identified in time series of stations derived from NOA network, as well as, due to the uncertainty in modeling the spatial distribution of fine temporal scale rainfall, it was considered more appropriate to form models with a monthly (and larger) time scale response, in order to draw safer conclusions. Given the data availability for each station, the range period to be studied is limited from October 2006 to December 2016, in order to at least use a decade-long data set.

In order to create models for monthly rainfall, the available primary data are first aggregated per day and then aggregated to the monthly values. However, this process cannot be done serially and seamlessly due to missing values. Thus, the positions of missing recording are detected and filled with those of nearest stations or with the average of the start and end values from the beginning to the end

of the unrecorded interval, if possible. Otherwise, the whole day shall be considered as non-recorded. Depending on the number of non-recorded days, the month will be considered as out of data or receive the value for the sum of the rainfall for the days corresponding to this month. The algorithm followed for this process is shown in detail, in the following flowchart (Figure 5).

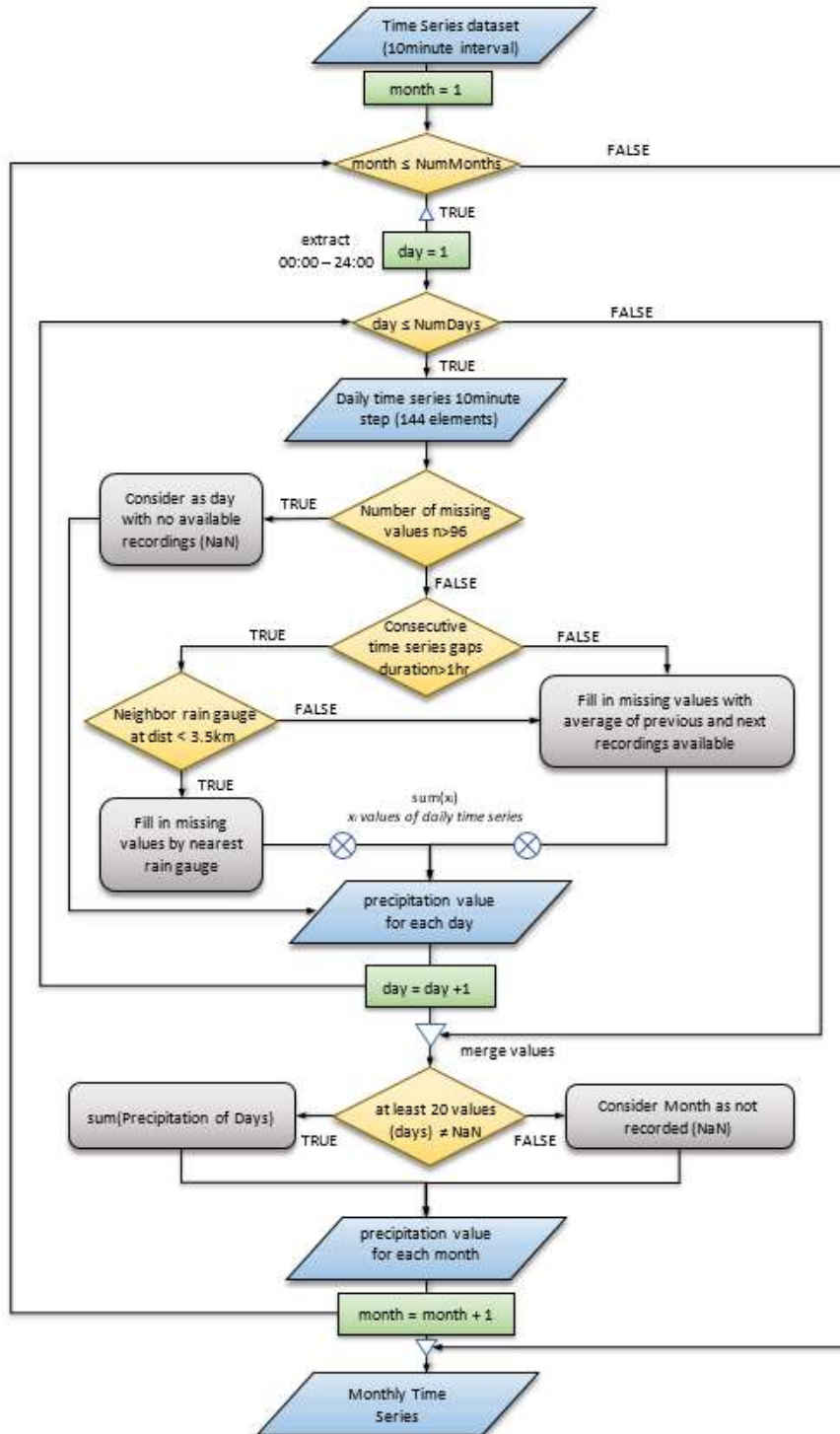


Figure 5 Flowchart for the process of raw data, in order to form monthly time series

Then the missing monthly values (either intermediate or values that were not initially available due to the station’s operation period) were completed for the ten-year period). These values are estimated via Kriging method, using the ArcGIS software. Specifically, and based on the conceptual nature of the variables considered in this thesis, as well as, the scientific publication by Earl and Dixon (2007), the Ordinary Kriging variant was used and the exponential semi-barogram model was selected for this purpose.

A set of geomorphological factors that are potentially related to the amount of precipitation are identified through the review on the relevant scientific literature. The final independent variables chosen to be considered are: elevation, X and Y coordinates according to the Greek Grid '87 projection system, terrain slope, aspect, closest distance to a river, as well as closest distance to the coastline.

Once the values of independent and dependent variables have been collected, the aim is to develop 12 different models for each month of the year, as well as one model corresponding the annual precipitation and two more referring to the wet and dry periods of the year, respectively, using a total of 15 models. The next flowchart (Figure 6) shows the process followed to form the final regression model. This methodology is based on the algorithm of Stepwise Procedure with backward elimination and is often applied to similar problems (Hession, et al., 2011; Marquinez et al., 2003).

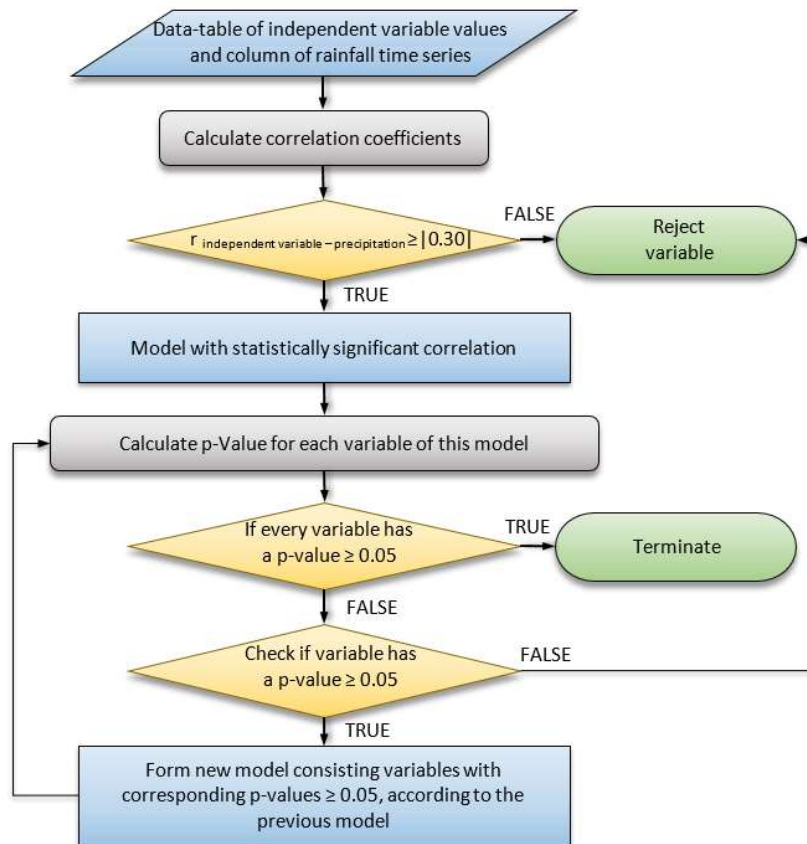


Figure 6 Flowchart of stepwise procedure (backwards elimination)

After finalizing the 15 regression models, the geospatial layers of predicted precipitation were created for each model case using the Raster Calculator toolkit in ArcGIS. Through these layers the set of the estimated ('forecasted') rainfall values corresponding to each station were extracted, by detecting the cells, in which the requested stations are located within. These values are compared to the actual – observed rainfall, in order to evaluate the accuracy of each model, through statistical indicators such as; R^2 ordinary and adjusted, observed significance p-value, MAE, MSE, RMSE, ME (bias) as well as a set of verification measures derived through the formation of a contingency table¹ (Probability of Detection, False Alarm Ratio and Critical Success Index). Verification of continuous variable prediction models, as in this case (amount of precipitation), often involves some exploratory charts, as suggested (Collaboration for Australian Weather and Climate Research - Australian Government, Bureau of Meteorology). Thus scatterplots, boxplots and bar charts are also formed, comparing the pairs of estimated and observed values for each model.

Results

After processing and filling the time series via Kriging method, through the clustering of the time series rainfall data, the average annual 10-year rainfall for each of the 21 stations is extracted. Furthermore, the average monthly rainfall for the decade 2007-2016 was calculated, considering all stations. These results are shown respectively in the following charts.

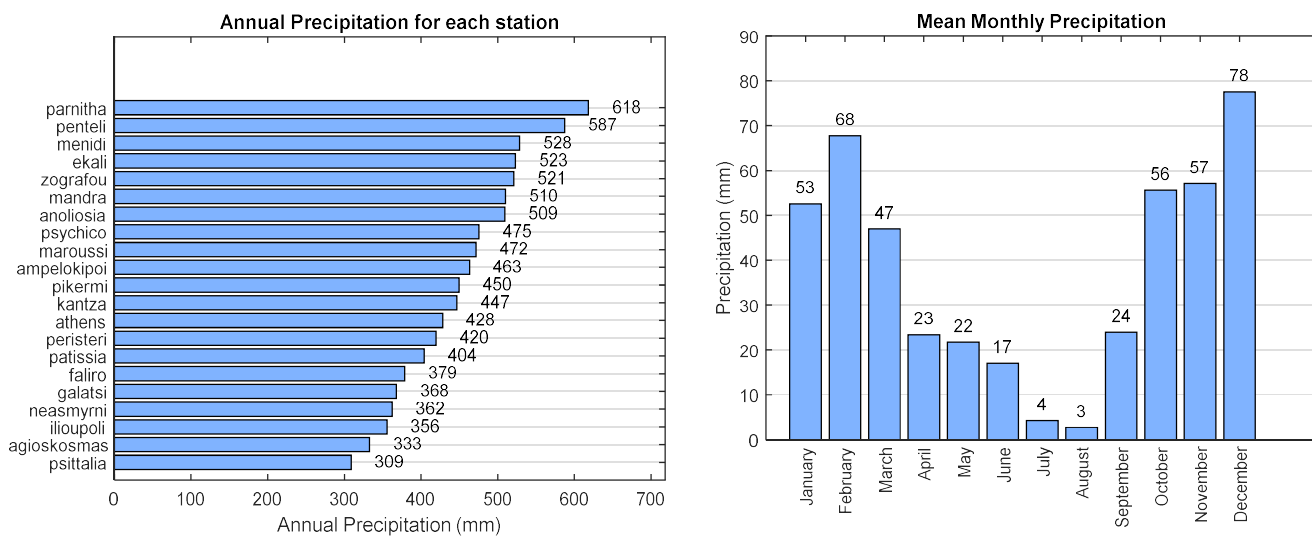


Figure 7 Values of dependent variables per station

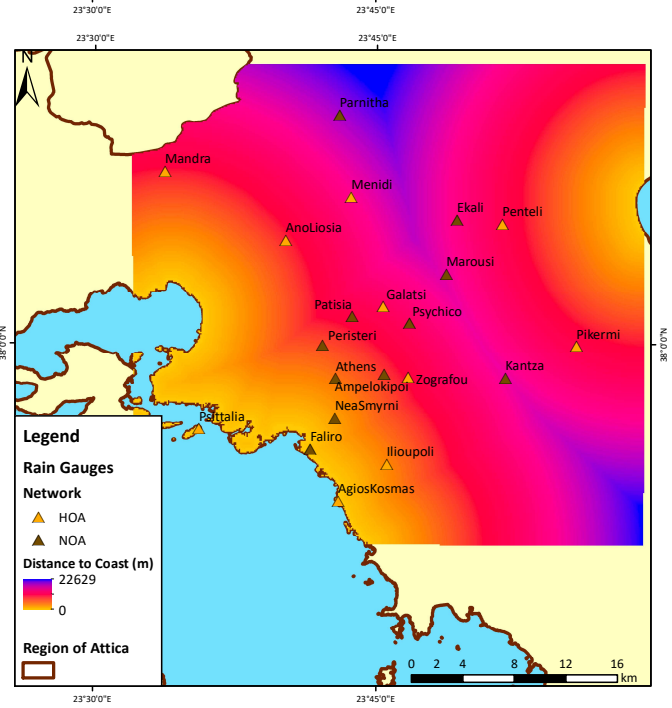
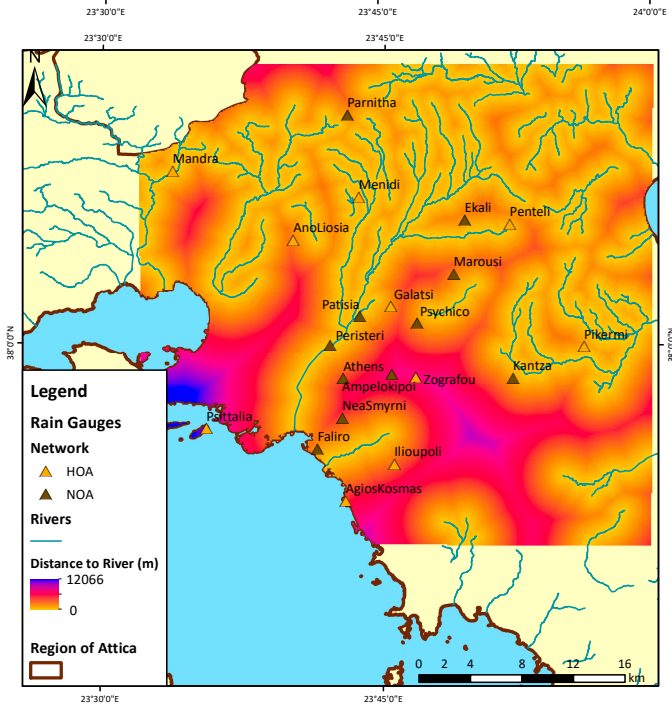
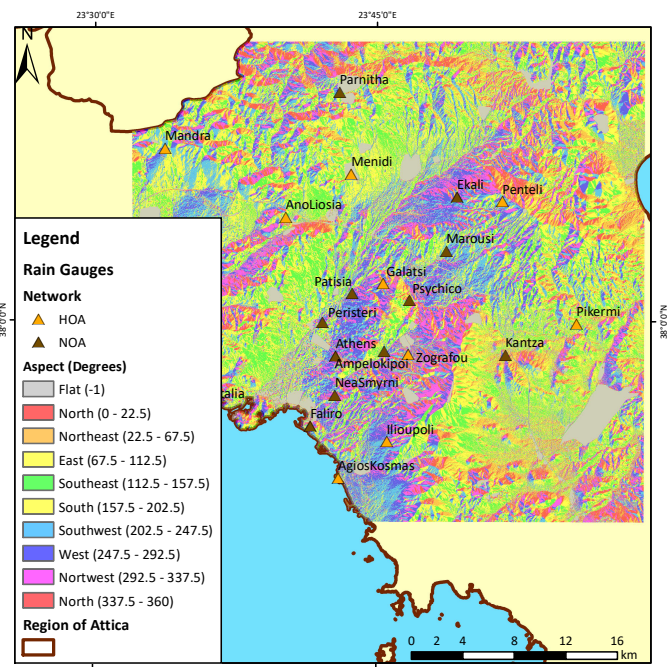
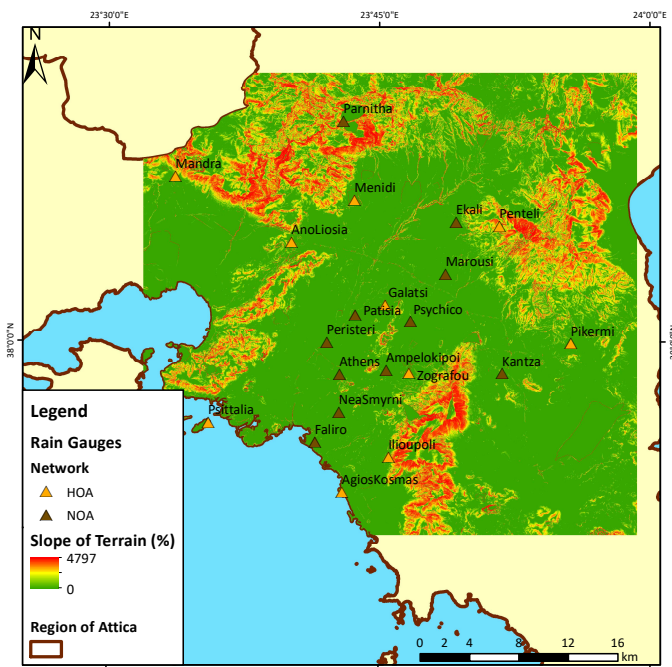
The geospatial layers for each of the selected independent variables are formed (in spatial accuracy of a 100m x 100m cell size) and the values corresponding to each station are extracted. The corresponding layers were created using the ArcGIS software are shown below.

¹ Jolliffe, I.T., and D.B. Stephenson, 2003

Wilks, D.S., 2011

Rossa, A., Nurmi, P., & Ebert, E., 2008

Nastos, P. T., Kapsomenakis, J., & Philandras, K. M., 2016



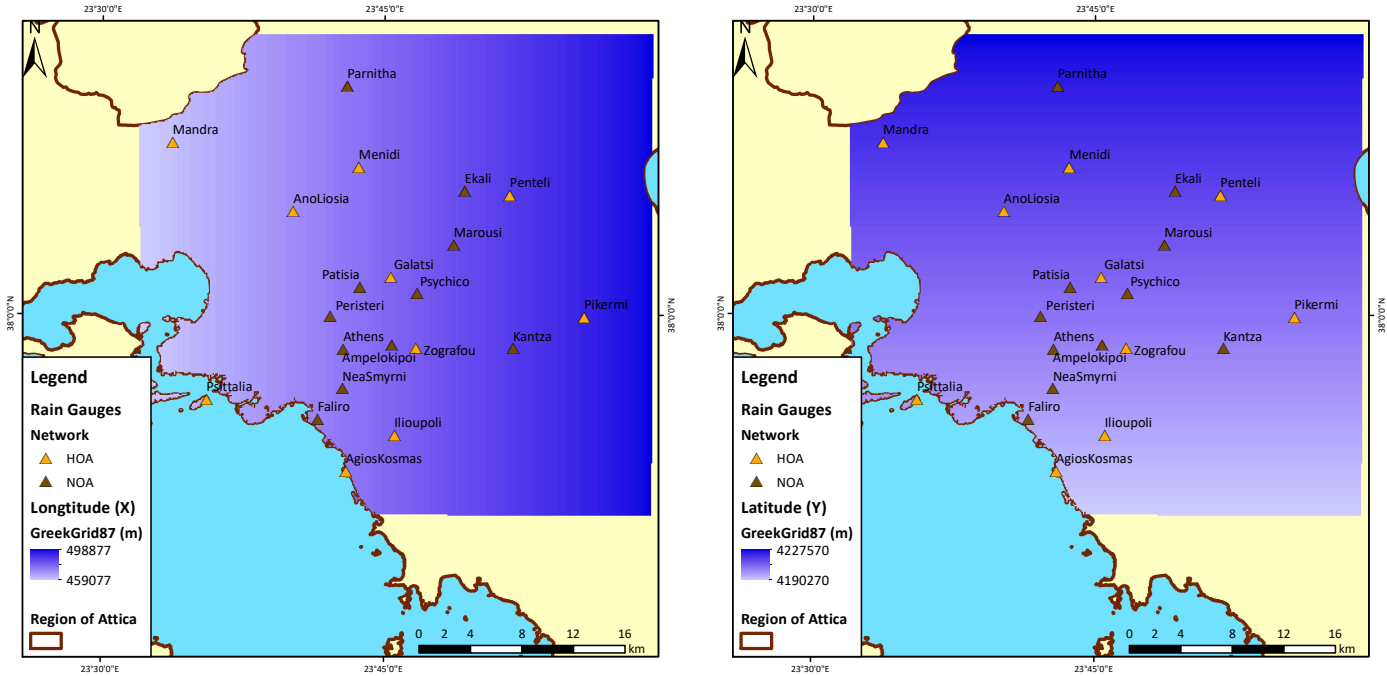


Figure 8 Maps for the studied independent variables; terrain slope (%), terrain aspect (degrees °), closest distance to a river, closest distance to the coastline, longitude and latitude (from up to down and left to right)

The extracted values of these independent variables as well as the dependent precipitation values are entered in order to create the models. The regression models are created in the MATLAB programming environment, following the methodology developed in detail in the previous chapter. In addition to the multiple linear regression models, as derived from the relevant scientific literature (Ranhao et al., 2008), polynomial and logarithmic regression models were also considered. These models did not satisfactorily fit the observed values compared to the multiple linear regression models, as indicated by the statistical indicators of correlation coefficient, coefficients of ordinary and adjusted R^2 and observed significance p-value, therefore were finally rejected. The best fitted multiple linear regression relationship is the following.

$$Precipitation = b_0 + b_1 \cdot Altitude + b_2 \cdot Y_{coordinate}$$

This relationship is applied to all 15 different models, for each of which the input values of the dependent variable (rainfall) are varied, thus extracting the values of the coefficients b_0, b_1, b_2 corresponding to each model, as well as a set of statistical indicators on the basis of which the models are evaluated.

The table below lists the values of the coefficients of each model, as well as the statistical indicators R_{adj}^2 and p-value, which are the main criteria for evaluating the models.

Table 2 Coefficients , $R^2_{adjusted}$, p-values, for each regression model

id	Model	Coefficients			$R^2_{adjusted}$	p-values		
		Intercept	Altitude	Coord. Y		Intercept	Altitude	Coord. Y
1	January	-2249	1.29E-02	5.46E-04	0.52	0.048	0.153	0.044
2	February	-3378	1.25E-02	8.18E-04	0.57	0.014	0.232	0.013
3	March	-1408	3.01E-02	3.44E-04	0.55	0.310	0.014	0.297
4	April	-1051	3.15E-03	2.55E-04	0.48	0.026	0.382	0.023
5	May	-846	1.54E-02	2.06E-04	0.78	0.068	0.000	0.063
6	June	-2656	6.75E-03	6.35E-04	0.79	0.000	0.139	0.000
7	July	-83	3.74E-03	2.05E-05	0.08	0.833	0.251	0.826
8	August	-794	-1.32E-03	1.89E-04	0.32	0.015	0.587	0.014
9	September	-1785	-2.70E-03	4.30E-04	0.31	0.017	0.631	0.016
10	October	-3555	1.35E-02	8.58E-04	0.80	0.000	0.045	0.000
11	November	-2760	1.17E-02	6.69E-04	0.39	0.074	0.337	0.070
12	December	-3971	2.76E-02	9.61E-04	0.77	0.004	0.011	0.003
13	Annual	-24536	1.33E-01	5.93E-03	0.75	0.002	0.031	0.002
14	Dry Season	-7214	2.50E-02	1.74E-03	0.79	0.000	0.046	0.000
15	Wet Season	-18241	1.19E-01	4.42E-03	0.77	0.003	0.013	0.003

In the previous table, R^2_{adj} values below the 0.70 threshold are highlighted in color, to highlight models that do not demonstrate a satisfactory fit to the observed values. At the same time, color-coding is also applied to the observed significance values corresponding to the b_i coefficients of independent variables, which are not statistically significant, i.e. for p-value values greater than the threshold of 0.05^2 . The latter criterion is crucial for separating the final statistically significant models that are further evaluated. More specifically, although some models (e.g. those corresponding to the mean monthly precipitation of May and June) show a high R^2_{adj} coefficient (0.78 and 0.79, respectively), they are rejected since they are not statistically significant, since the p-values do not meet the abovementioned threshold. Therefore, in the following paragraph, results only for the models of: October, December, annual, dry and wet season will be presented.

² for significance level, $\alpha=5\%$ (95% confidence interval)

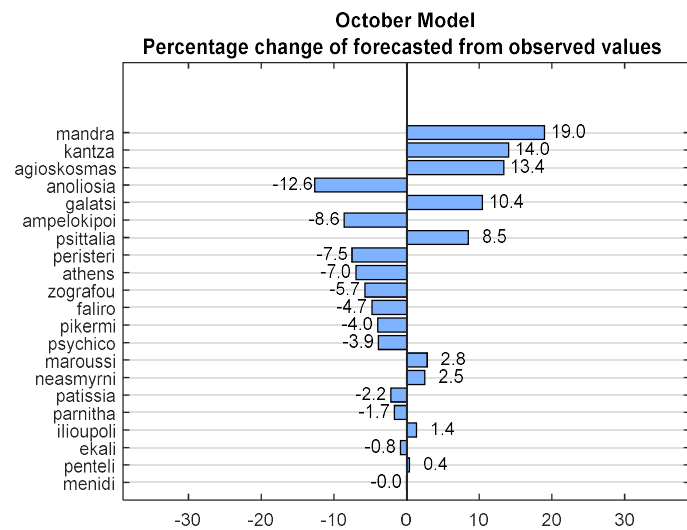
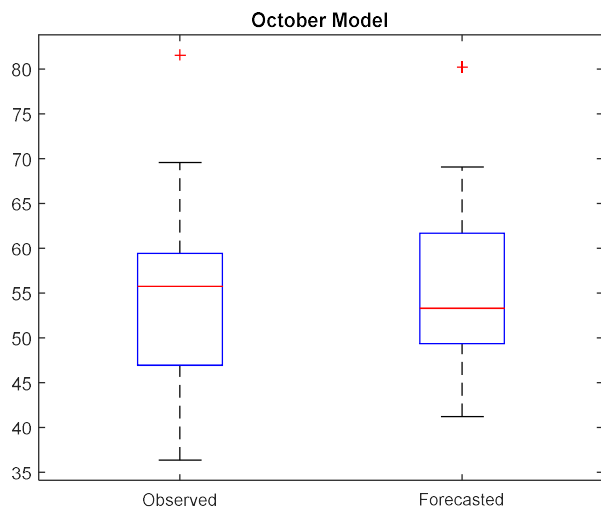
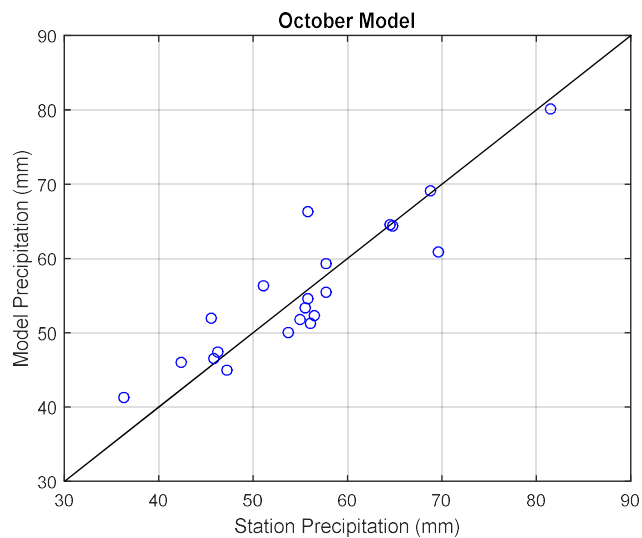
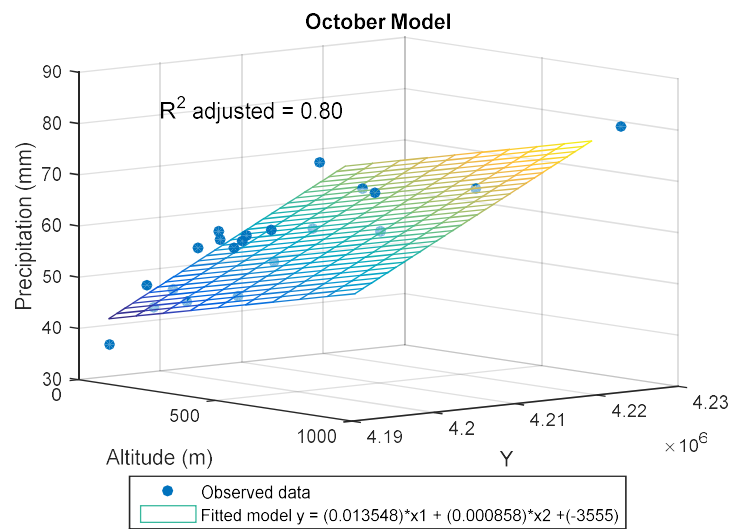


Figure 9 3-D regression model, scatter plot, boxplot of observed and estimated values, bar plot of percentage change between forecasted and observed values for October model

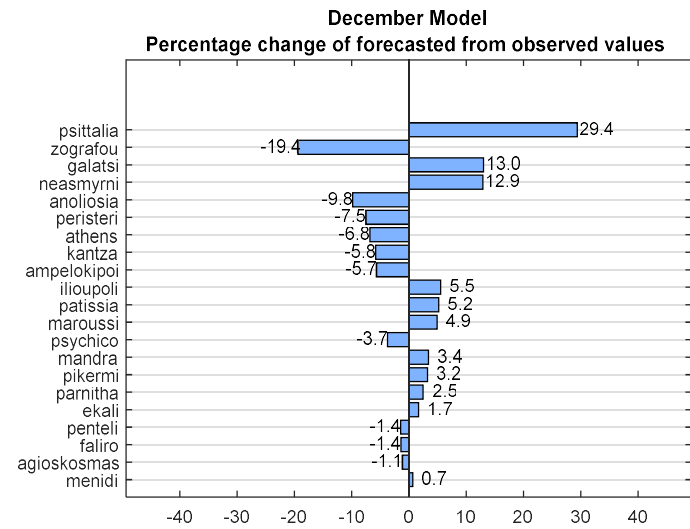
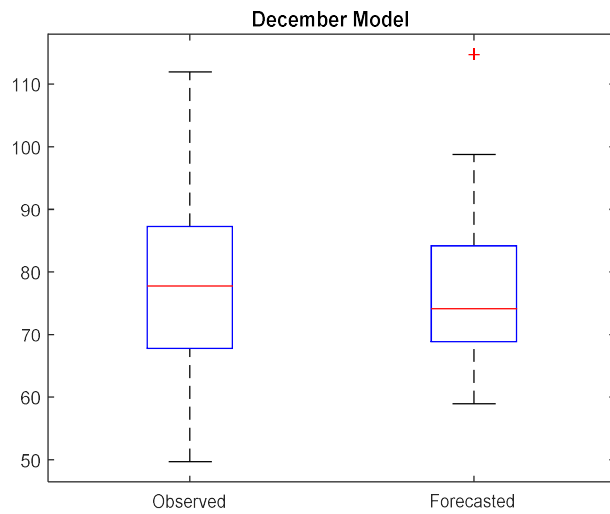
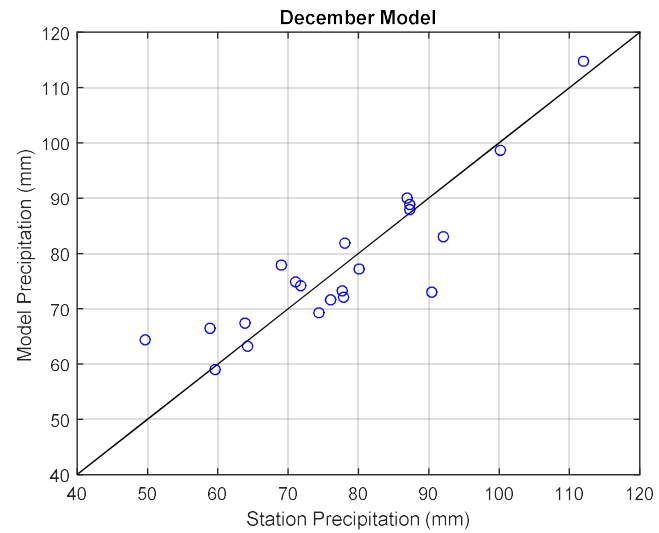
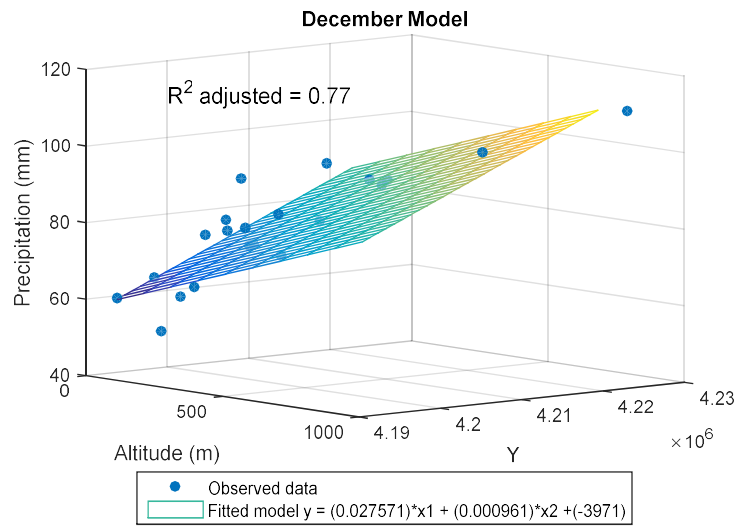


Figure 10 3-D regression model, scatter plot, boxplot of observed and estimated values, bar plot of percentage change between forecasted and observed values for December model

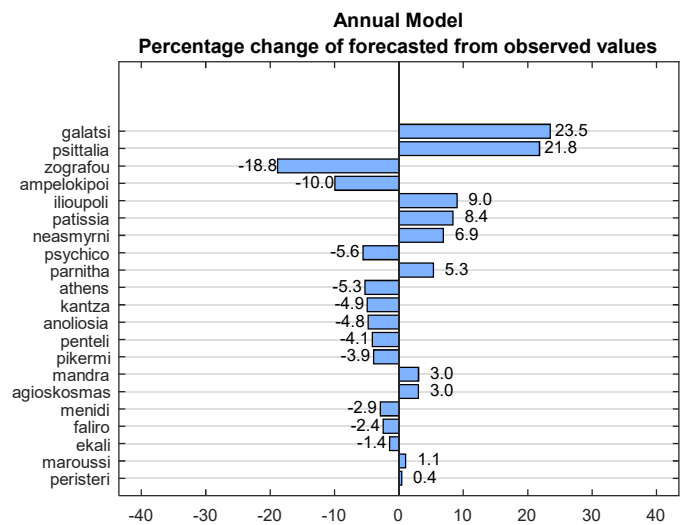
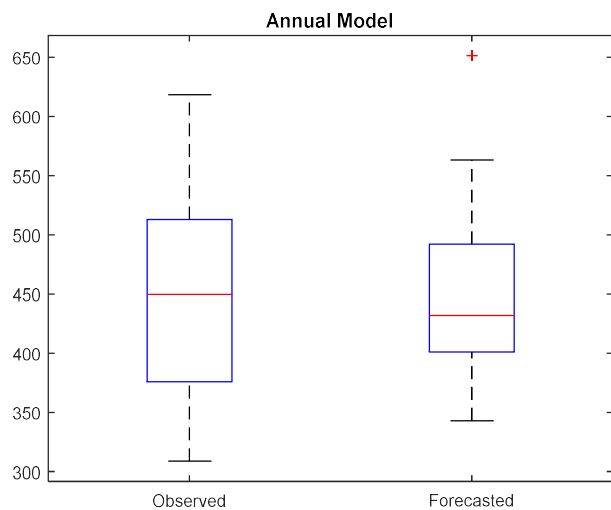
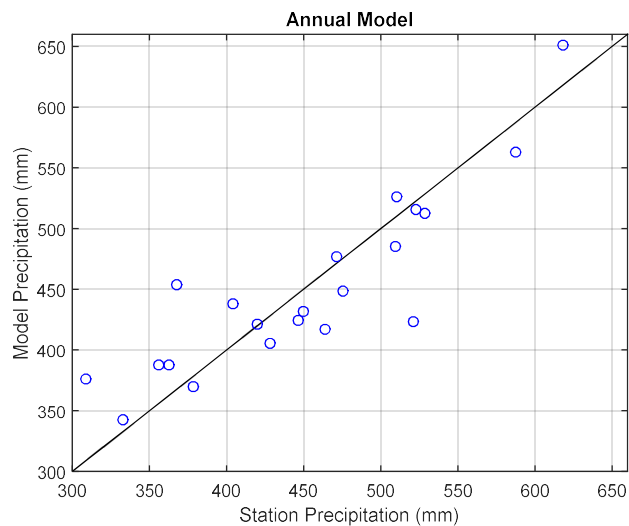
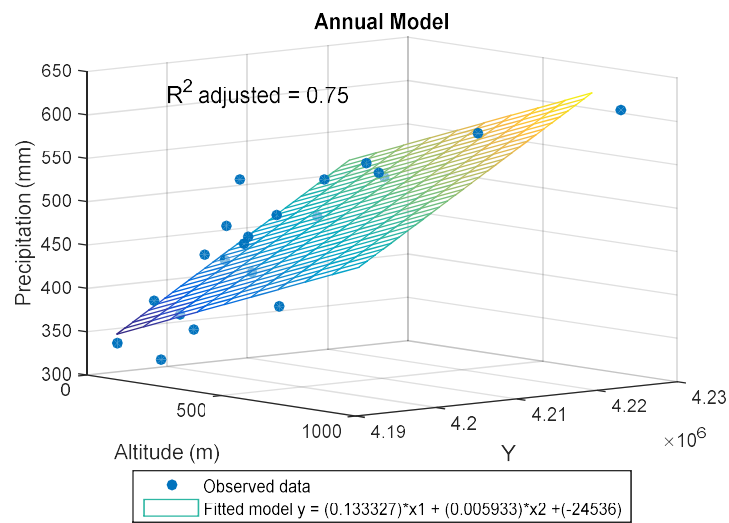


Figure 11 3-D regression model, scatter plot, boxplot of observed and estimated values, bar plot of percentage change between forecasted and observed values for the model of Annual precipitation

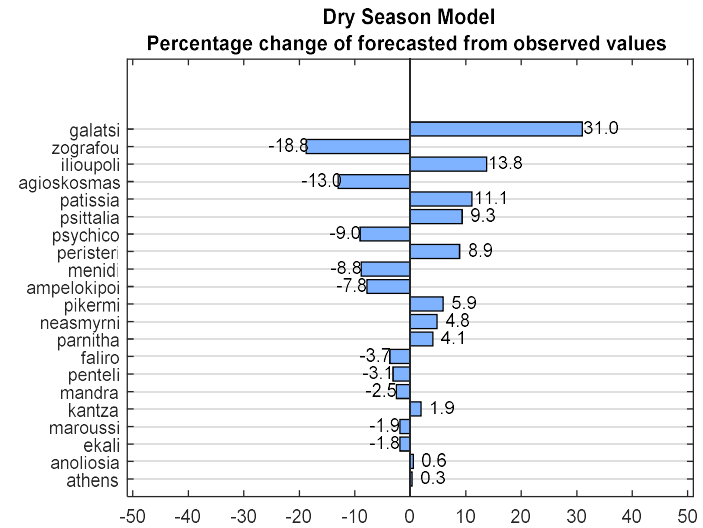
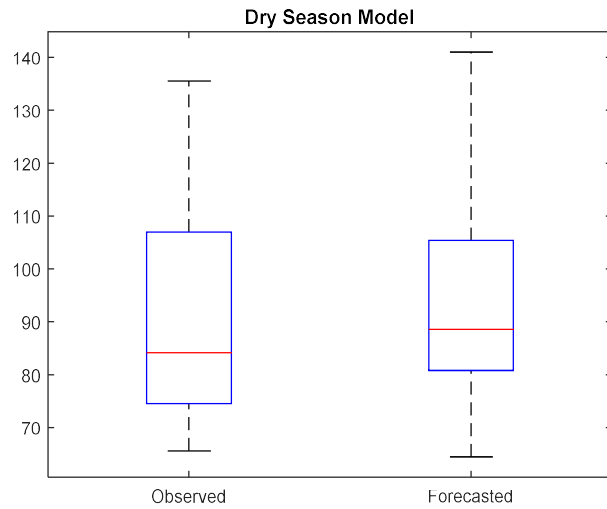
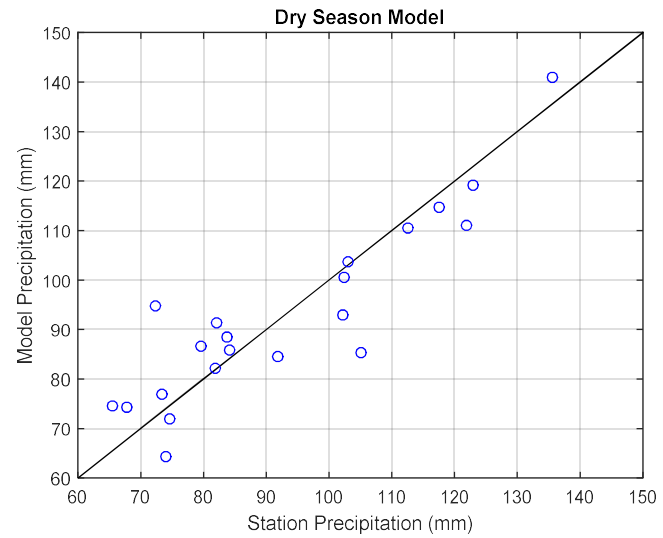
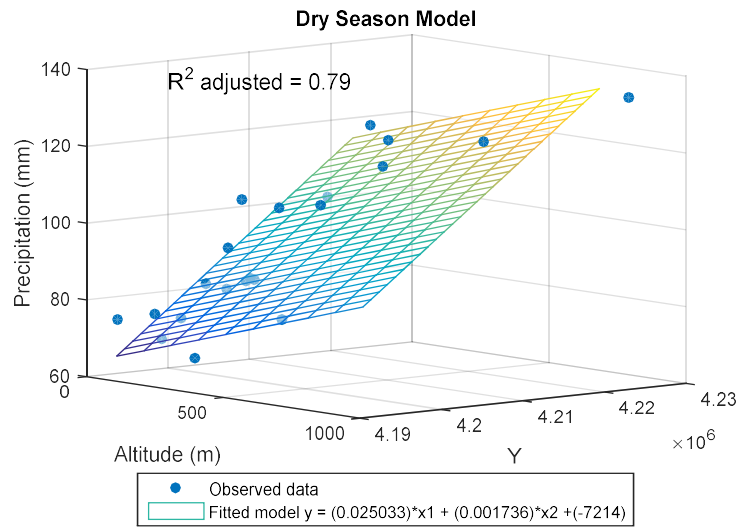


Figure 12 3-D regression model, scatter plot, boxplot of observed and estimated values, bar plot of percentage change between forecasted and observed values for the model of Dry Season precipitation

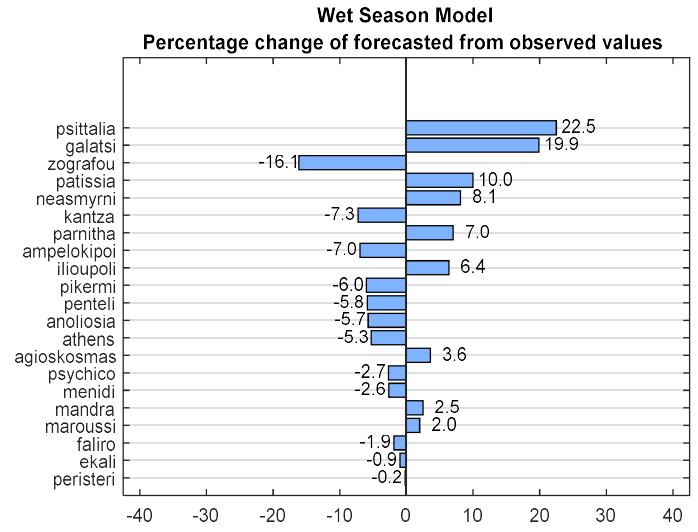
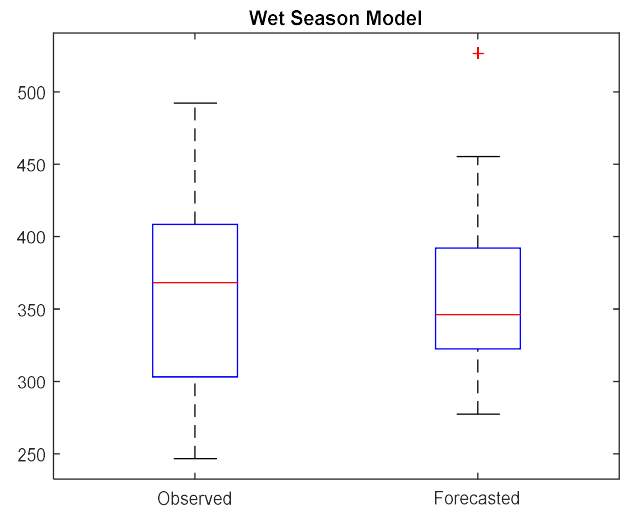
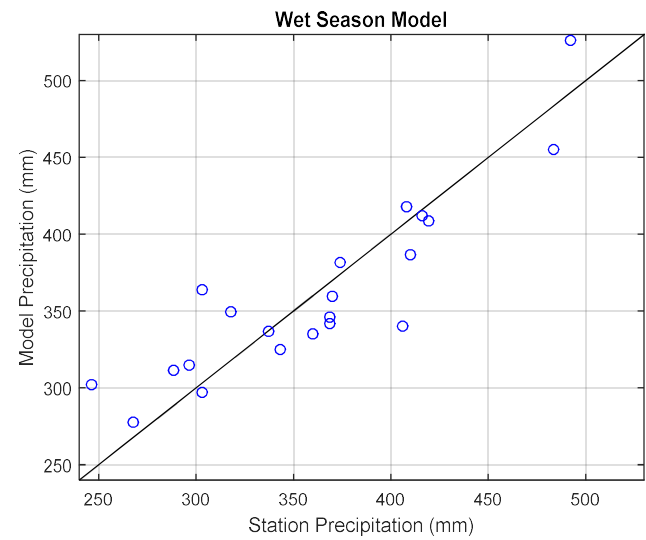
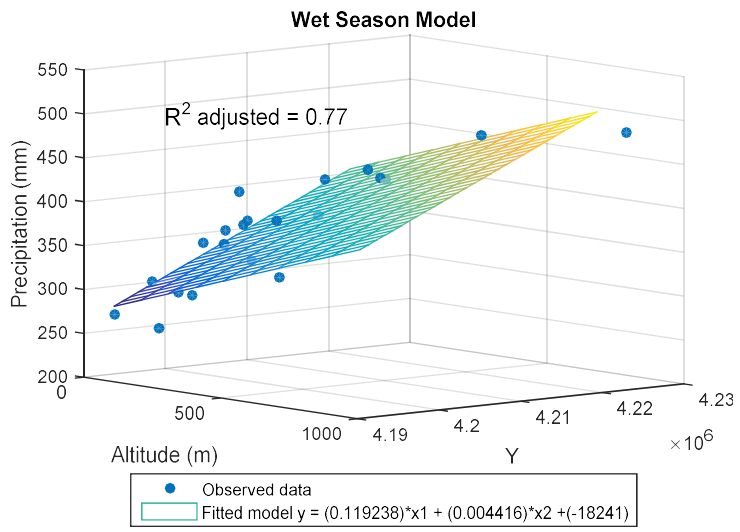


Figure 13 3-D regression model, scatter plot, boxplot of observed and estimated values, bar plot of percentage change between forecasted and observed values for the model of Wet Season precipitation

In the following table the resulted values for the scalar accuracy measures of Mean Error, Mean Absolute Error, Mean Squared Error and its root (RMSE), for each of the statistically significant models, are listed. According to the values of ME, all models, but the annual, are positive thus overestimating the actual precipitation values.

Table 3 Models evaluation statistics

<i>Model</i>	<i>ME</i>	<i>MAE</i>	<i>MSE</i>	<i>RMSE</i>
October	0.000	3	18	4
December	0.008	5	44	7
Annual	-0.172	30	1505	39
Dry Season	0.018	7	76	9
Wet Season	0.528	23	863	29

As it is obtained from the ranges of the above values, for each of the measures, a comparative conclusion among the five models cannot be conducted. This happens due to the different temporal scale of precipitation values for the majority of the models. Consequently, the differences between observed and estimated values vary. The above conclusion is shown through the following boxplot (Figure 14).

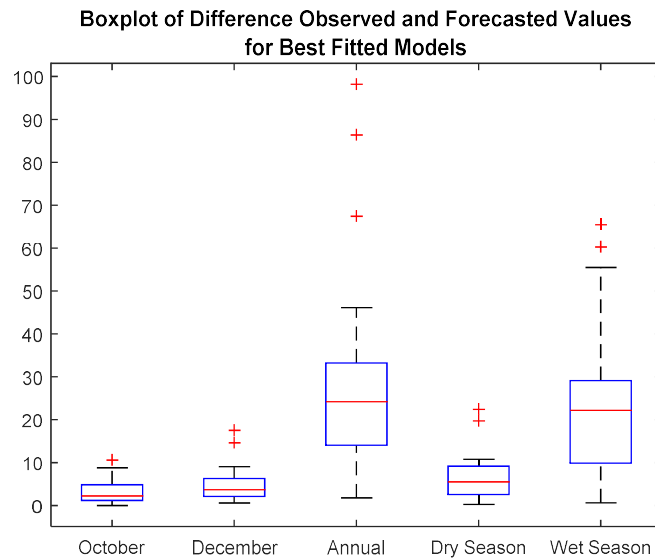


Figure 14 Boxplot for the set of differences between the pairs of observed and estimated values, for each of the statistically significant models

Hence, in order to facilitate the comparison of data in different scales, the RMSE values are normalized. This normalization can be done by numerous methods. The most common of these uses the average of the observed data. The normalized RMSE values for each model are shown in the next bar chart (Figure 15). For the cases where outliers are detected, which is the current case as shown in the previous boxplots corresponding to each model, normalization by means of the difference between the actual values corresponding to 25% and 75% of the observations is preferred.

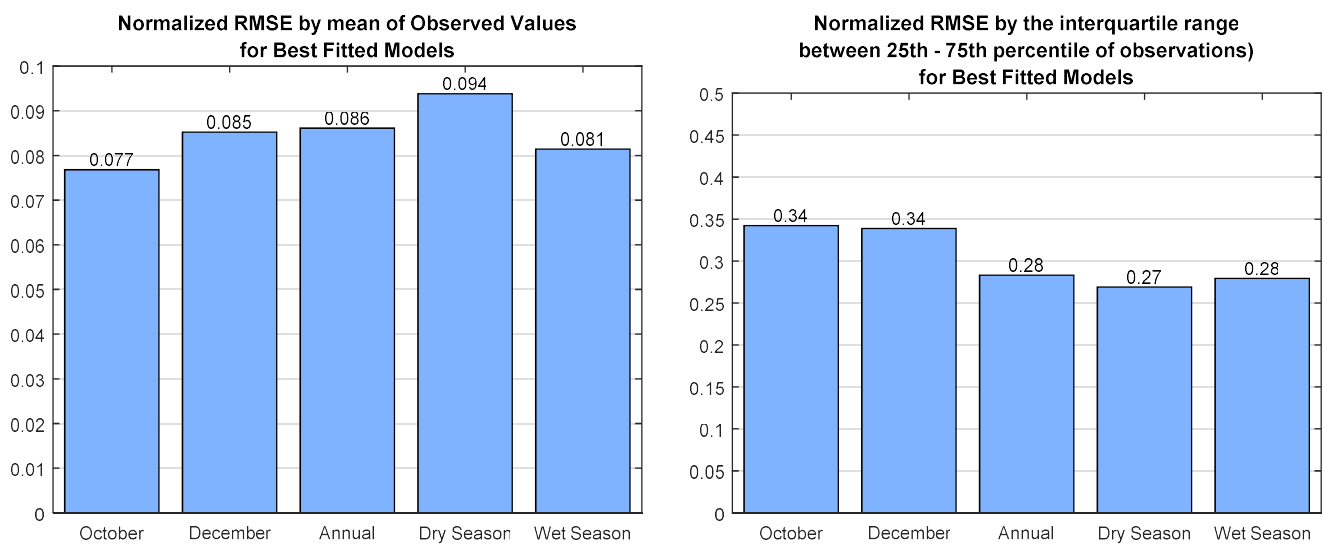


Figure 15 Normalized RMSE by the average of the observed values, as well as by the interquartile range between 25% and 75% of the observations

A lower value of NRMSE is not necessarily associated with better fitting results and therefore the next graph plots the values of the normalized NMSE, with a new form (suggested by Hanna et al., 1985), which ameliorates the potential problem of this indicator; the bias-preference towards models that underestimate the observed values.

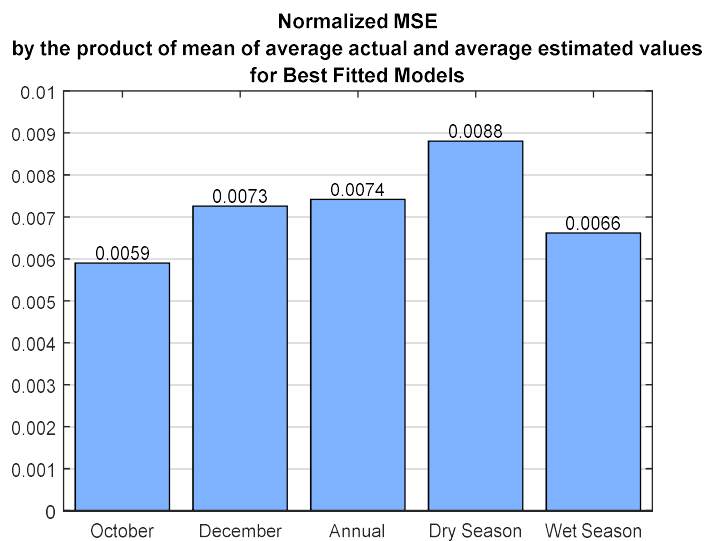


Figure 16 Normalized MSE, by the product of mean observed values and mean of estimated values

Depending on the statistical tool used each time, for the normalization of RMSE or MSE, the results and therefore the evaluation - ranking of the models may differ significantly. However, it is clear from all three charts that the indices slightly differ, examining one chart at a time.

Most notable and frequently applied evaluation approaches in similar scientific fields are the incorporation of POD, FAR and CSI evaluation indices, which values are generally derived from the frequencies a, b, c, d of discrete events (here, case d is not applicable), which consist a contingency table. Below, the relevant tables for each of the five statistically significant models are listed.

Table 4 Contingency tables for each of the five models – frequencies for discrete events

2x2 Contingency Table		Observed	
October		Yes	No
Forecast	Yes	16	4
	No	1	

2x2 Contingency Table		Observed	
December		Yes	No
Forecast	Yes	17	3
	No	1	

2x2 Contingency Table		Observed	
Annual		Yes	No
Forecast	Yes	18	2
	No	1	

2x2 Contingency Table		Observed	
Dry Season		Yes	No
Forecast	Yes	16	3
	No	2	

2x2 Contingency Table		Observed	
Wet Season		Yes	No
Forecast	Yes	18	2
	No	1	

Through the frequencies of a,b,c the verification measures of POD, CSI and FAR are calculated. The closer to 1 the POD and CSI are, the more reliable the model is. The reverse is true for the FAR. The highest value for the POD index is 0.95 and for the CSI index 0.86, which belong to the annual model and the wet season model. While the lowest value for the FAR index is 0.10, which also belongs to the annual model and the wet season model.

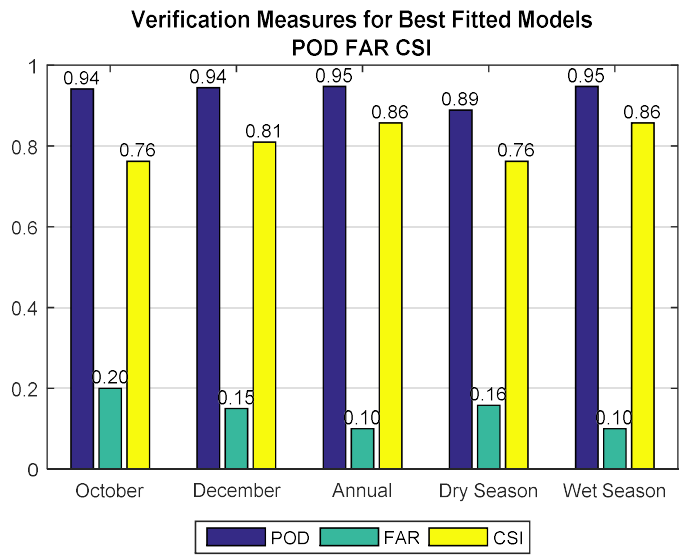


Figure 17 Verification measures – coefficients for the significant models.

Conclusions and Recommendations for Future Research

The main findings of this study are summarized as follows:

- The independent variables of terrain slope and aspect, as well as the longitude present the lowest correlation coefficients regarding the precipitation, and thus discarded from the very first stage of the model formulation. High correlations are found, in descending order, for the variables of: latitude, altitude and distance from the coastline.
- The results and consequently the conclusions drawn are characterized by spatial and statistical uncertainty, given the small sample of available fully-operating stations, as well as the short period of records (equal to or less than a decade of recorded rainfall values), while a climate period of 30 years is usually required for the study of these variables' models.
- According to the chart of average monthly rainfall, unexpectedly high rainfall depth is observed for the month of February. However, other monthly values are within more reasonable levels.
- Due to the small number of stations with complete time series that are used to fill in data for the remaining stations, the formation of a proper spatially distributed network of gauges was not always possible. This resulted in a biased filling for stations distant to those with records. Consequently, the fill-in procedure is more reliable for the case of intermediate gaps in the time series, since usually in this case most stations had a recorded value and therefore the fill-in is based on more available values. In contrast, for the first years of the decade considered, NOAA network stations have no records. This inevitably leads to more uncertain results, as only the values of the HOA network stations (10 stations) are taken into account.
- High rainfall variations with large spatial differences from the southern to the northern parts of the basin are observed, given the relatively small geographical area studied.
- There is a high correlation between the independent variables of latitude and altitude, which would statistically prevent the use of these two variables in the model. However, conceptually these variables are not related, but a fictitious correlation is appeared for the Attica basin, since its main and highest mountain ranges are located in the northernmost part of the study area.
- Lack-of-fit exhibit mostly models for the months with the lowest rainfall (July, August, September). Similar adjustment effects are also found at a spatial level. To illustrate this, as shown in the bar charts of the percentage change between estimated and observed rainfall values, it is mainly the stations that recorded the highest rainfall depths that had the lowest deviations.
- For the models where the time scale is coarser (i.e., the annual model, wet and dry season models), the fit to the actual data is remarkably improved compared to the monthly models (apart the October and December cases). This was quite predictable, as the amount of precipitation is significantly higher for those time periods, while data aggregation on larger time scales ensures that the abrupt variation in the time series data is reduced.

Based on the previous findings, the following suggestions for future research are drawn up:

- The same methodological framework can be applied, using rainfall data obtained by a denser network of rain gauges, which can adequately cover the whole Region of Attica. At the same time, it is considered necessary for the available stations to have a time series range close to a 30-year period, as an integrated climate reference period. According to the relevant scientific literature, this time range is the minimum period to be considered as sufficient and reliable for an accurate characterization of the climate prevailing in a region.
- Given the availability of a denser network, this methodology can be performed in a smaller spatial resolution and also may lead to results of higher spatial accuracy. At the same time, these results can be compared with those obtained in the frame of present work (where a 100m x 100m grid is used).
- The application of this methodology across Greece is also purposed, as the results of the present study are not representative for the whole Greek territory. The great heterogeneity of Greek climate and the simultaneously repetitive episodes of extreme and intense rainfall (with a higher frequency, especially in recent years due to climate change) makes this potential study a highly pressing issue and at the same time challenging.
- Within the same framework, many combinations of non-linear regression models can be also examined and compared to the present results, regarding their goodness-of-fit. At the same time, more independent variables and their correlation to the precipitation should be studied (e.g. incident solar radiation, temperature, wind intensity and direction).
- A comparative analysis is also recommended to determine the most suitable method for filling in missing rainfall data for the same study area. Kriging technique, applied in this thesis, is commonly used in problems involving the geospatial factor, however it would be very useful to compare the results of this method with simpler approaches (e.g. least squares completion, by neighboring stations data, that have a high correlation with the station to be completed) or more complex ones with stochastic models and algorithms derived from neural network theory.
- Finally, the methodology verification can be performed in the future based on accurate remote sensing data of higher spatial resolution, such as Radar or satellite rainfall estimates.

Selected References

Apostolopoulos, I. (2012). MSc Thesis in Greek. *Available as:* Αποστολόπουλος, Ι. (2014) *Εκτίμηση υδρομετεωρολογικών παραμέτρων καταιγίδων κατακόρυφης ανάπτυξης*, Μεταπτυχιακή Διπλωματική Εργασία ΔΠΜΣ «Επιστήμη και Τεχνολογία Υδατικών Πόρων» ΕΜΠ

- Baltas, E. A., Panagos, D. S., & Mimikou, M. A. (2015). An approach for the estimation of hydrometeorological variables towards the determination of ZR coefficients. *Environmental Processes*, 2(4), 751-759.
- Earls, J.; Dixon, B. Spatial Interpolation of Rainfall Data Using ArcGIS: A Comparative Study. In *Proceedings of the Proceedings of the 27th Annual ESRI International User Conference, San Diego, CA, USA, 4–8 August 2007*; Volume 31, pp. 1–9.
- Feloni, E., Mousadis, I., & Baltas, E. (2020). Flood vulnerability assessment using a GIS-based multi-criteria approach—The case of Attica region. *Journal of Flood Risk Management*, 13, e12563.
- Grammatikogiannis, A., Mamassis, N., Baltas, E., & Mimikou, M. (2005, September). A meteorological telemetric network for monitoring of the Athens wider area (METEONET). A real time approach from point to areal measurements. In *Proceedings of the Ninth International Conference on Environmental Science and Technology (9CEST), Rhodes Island, Greece* (pp. 1-3).
- Hession, S. L., & Moore, N. (2011). A spatial regression analysis of the influence of topography on monthly rainfall in East Africa. *International journal of climatology*, 31(10), 1440-1456.
- Jolliffe, I. T., & Stephenson, D. B. (2003). *Forecast Verification: A Practitioner's Guide in Atmospheric Science*.
- Koutsogiannis, et al. (2010) Technical Report in Greek. Available as: Κουτσογιάννης Δ., Μαρκόνης Ι., Κουκουβίνος Α., Παπαλεξίου Σ.Μ., Μαμάσης Ν., και Δημητριάδης Π. (2010) Υδρολογική μελέτη ισχυρών βροχοπτώσεων στη λεκάνη του Κηφισού, *Μελέτη διαχείρισης Κηφισού*, Ανάθεση: Γενική Γραμματεία Δημοσίων Έργων – Υπουργείο Περιβάλλοντος, Χωροταξίας και Δημόσιων Έργων.
- Lagouvardos, K., Kotroni, V., Bezes, A., Koletsis, I., Kopania, T., Lykoudis, S., Mazarakis, N., Papagiannaki, K. and Vougioukas, S. (2017). The automatic weather stations NOANN network of the National Observatory of Athens: Operation and database. *Geoscience Data Journal*, 4(1), 4-16.
- Maheras, P., & Anagnostopoulou, C. (2003). Circulation types and their influence on the interannual variability and precipitation changes in Greece. In *Mediterranean climate* (pp. 215-239). Springer, Berlin, Heidelberg.
- Marquínez, J., Lastra, J., & García, P. (2003). Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis. *Journal of hydrology*, 270(1-2), 1-11.

Nastos, P. T., Kapsomenakis, J., & Philandras, K. M. (2016). Evaluation of the TRMM 3B43 gridded precipitation estimates over Greece. *Atmospheric Research*, 169, 497-514.

Ranhao, S., Baiping, Z., & Jing, T. (2008). A multivariate regression model for predicting precipitation in the Daqing Mountains. *Mountain Research and Development*, 28(3), 318-325.

Rossa, A., Nurmi, P., & Ebert, E. (2008). Overview of methods for the verification of quantitative precipitation forecasts. In *Precipitation: Advances in measurement, estimation and prediction* (pp. 419-452). Springer, Berlin, Heidelberg.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic press.
geodata.gov.gr [last accessed on September 20, 2022]

Collaboration for Australian Weather and Climate Research – Australian Government, Bureau of Meteorology https://www.cawcr.gov.au/projects/verification/#Roebber_diagram [last accessed on June 20, 2022]

ESRI – Kriging in Geostatistical Analyst – How Kriging works <https://desktop.arcgis.com/> [last accessed on October 1, 2022]

1 ΕΙΣΑΓΩΓΗ

1.1 Αντικείμενο Εργασίας και Εισαγωγικά Στοιχεία

Αντικείμενο της παρούσας μεταπτυχιακής εργασίας αποτελεί η διαμόρφωση και εφαρμογή ενός μεθοδολογικού πλαισίου για την ανάπτυξη και διερεύνηση μοντέλων πολλαπλής γραμμικής παλινδρόμησης με στόχο την χωρική εκτίμηση του ύψους βροχόπτωσης για διαφορετικές χρονικές περιόδους (μηνιαία, ετήσια, κλπ.), με περιοχή εφαρμογής το λεκανοπέδιο της Αττικής.

Η περιοχή της Αττικής χαρακτηρίζεται από ποικίλο ανάγλυφο, με το μεγαλύτερο τμήμα της να παρουσιάζει πεδινό έως ημιορεινό ανάγλυφο, με ασυνεχείς εξάρσεις ορεινών όγκων. Όσον αφορά τον κλιματολογικό τύπο της Αττικής, η ίδια χαρακτηρίζεται από μεγάλης διάρκειας διαστρωματικές βροχοπτώσεις την ψυχρή περίοδο του έτους και από έντονα μεταγωγικά επεισόδια την θερμή. (Αποστολόπουλος, 2014). Η συγκεκριμένη γεωμορφολογία σε συνδυασμό με το προαναφερθέν κλιματολογικό προφίλ δημιουργεί ένα σύνολο από ποτάμια υδατικά συστήματα με χειμαρρώδη χαρακτηριστικά, τα οποία λόγω της έντονης αστικοποίησης συχνά οδηγούν σε αυξημένη πιθανότητα πρόκλησης πλημμυρικών φαινομένων (Feroni et al., 2020). Ειδικότερα για την λεκάνη του Κηφισού εκτιμάται ότι η δομημένη έκταση ξεπερνά το 70% και αυξάνεται διαρκώς, ενώ οι ορεινές και ημιορεινές περιοχές της λεκάνης στην Πάρνηθα και την Πεντέλη έχουν κατά καιρούς πληγεί από πυρκαγιές, με την καταστρεπτικότερη το καλοκαίρι του 2007 (Κουτσογιάννης κ.ά., 2010).

Η πρόβλεψη και η προστασία από πλημμυρικά επεισόδια, καθώς και η εξασφάλιση του πόσιμου και του αρδευτικού νερού για την προστασία από περιόδους ξηρασίας, μέσω της ορθολογικής διαχείρισης των υδατικών πόρων των λεκανών απορροής, είναι ζητήματα που κάνουν επιτακτική την ανάγκη ανάπτυξης δραστηριοτήτων και νέων τεχνικών με σκοπό την όσο το δυνατόν ακριβέστερη εκτίμηση και πρόγνωση των υδρολογικών διαδικασιών. Ιδιαίτερα σε μία περίοδο όπου η κλιματική αλλαγή είναι πλέον αισθητή, η γνώση των χρονικών και χωρικών μοτίβων της βροχόπτωσης σε κλιματική βάση είναι σημαντική και αναγκαία.

Υπό αυτό το πλαίσιο αναπτύσσονται μαθηματικά μοντέλα, μέσω της Ανάλυσης Παλινδρόμησης, που στόχο τους έχουν την χωρική εκτίμηση της βροχόπτωσης. Με τον όρο Ανάλυση Παλινδρόμησης (Regression Analysis) αναφερόμαστε σε ένα σύνολο στατιστικών τεχνικών, μέσω των οποίων προσδιορίζεται η πιθανή υπάρχουσα σχέση μεταξύ μεταβλητών με απώτερο σκοπό την πρόβλεψη μιας από αυτές μέσω των άλλων. Έτσι προσδιορίζεται ένα μοντέλο για αυτές τις μεταβλητές, το οποίο ονομάζεται Μοντέλο Παλινδρόμησης (Regression Model). Η απλούστερη μορφή παλινδρόμησης είναι η Απλή Γραμμική Παλινδρόμηση (simple linear regression), κατά την οποία υπάρχει γραμμική σχέση εξάρτησης μεταξύ της εξαρτημένης μεταβλητής Y και μιας ανεξάρτητης μεταβλητής X . Αντίστοιχα το μοντέλο που προκύπτει ονομάζεται Απλό Γραμμικό Μοντέλο Παλινδρόμησης. Η Πολλαπλή Γραμμική Παλινδρόμηση (multiple linear regression) είναι μία επέκταση της απλής παλινδρόμησης, καθώς τώρα στο μοντέλο αντί για μία ανεξάρτητη μεταβλητή X , υπάρχουν k ανεξάρτητες μεταβλητές X_1, X_2, \dots, X_k . Το πρόβλημα της παλινδρόμησης εντοπίζεται στον προσδιορισμό των άγνωστων συντελεστών των ανεξάρτητων μεταβλητών, η

εκτίμηση των οποίων πραγματοποιείται συνηθέστερα, μέσω της μεθόδου Ελαχίστων Τετραγώνων (Least Squares).

Συγκεκριμένα για την παρούσα διπλωματική εργασία, οι εξεταζόμενες ανεξάρτητες μεταβλητές που διερευνώνται, ως προς την συμμετοχή τους στο τελικό μοντέλο παλινδρόμησης, είναι μεταβλητές γεωχωρικής φύσεως (υψόμετρο, κλίση εδάφους κλπ.), ενώ η εξαρτημένη μεταβλητή είναι αυτή του ύψους βροχόπτωσης σε μηνιαία χρονική κλίμακα και άνω.

Η τρέχουσα πρακτική εξασφάλισης των απαιτούμενων βροχομετρικών δεδομένων για την διαμόρφωση των μοντέλων παλινδρόμησης, στηρίζεται σε μετρήσεις καταγεγραμμένες από βροχογράφους, από τα διαθέσιμα δίκτυα βροχογράφων της Αττικής. Ένας βροχογράφος είναι σε θέση να προσφέρει δεδομένα σημειακής βροχόπτωσης υψηλής ακρίβειας, ωστόσο τα δεδομένα αυτά αναφέρονται -και συνεπώς περιορίζονται- στη στενή περιοχή του κάθε βροχογράφου.

Σε ένα καλά οριοθετημένο δίκτυο βροχογράφων οι μετρήσεις θα μπορούσαν να δώσουν δεδομένα μεγαλύτερης αξιοπιστίας, παρ' όλα αυτά το πρόβλημα προσέγγισης των τιμών βροχόπτωσης που αντιστοιχούν σε ενδιάμεσες περιοχές, δηλαδή μεταξύ των σταθμών, παραμένει, καθώς η χωρική μεταβλητότητα δεν μπορεί να αποτυπωθεί επαρκώς. Μέσω της «εκπαίδευσης» των μοντέλων παλινδρόμησης με μετρήσεις από ένα τέτοιο σωστά χωρικά καταμεμημένο δίκτυο σταθμών, μπορούν να εξαχθούν εκτιμήσεις βροχόπτωσης σε μορφή γεωχωρικής πληροφορίας, καλύπτοντας έτσι όλη την περιοχή μελέτης στην οποία εμπίπτουν οι σταθμοί.

1.2 Πρωτότυπα σημεία

Η μελέτη, η κατανόηση και κατ' επέκταση η εκτίμηση των χρονικών και χωρικών μοτίβων των βροχοπτώσεων σε κλιματική κλίμακα είναι απαραίτητη για την πρόγνωση και προστασία από ακραία πλημμυρικά φαινόμενα, καθώς και την ορθή διαχείριση των διαθέσιμων υδατικών πόρων των λεκανών απορροής. Η Αττική αποτελεί μια περιοχή που χαρακτηρίζεται από έντονη αστικοποίηση και μεταβλητότητα, ιδίως ως προς τη γεωμορφολογία της και τα μετεωρολογικά και κλιματολογικά χαρακτηριστικά της.

Η ανάπτυξη ενός μεθοδολογικού πλαισίου για την ολοκληρωμένη διερεύνηση των χωρικών και χρονικών μοτίβων σε διαφορετικές περιόδους για το λεκανοπέδιο της Αττικής, με στόχο την εκτίμηση της βροχόπτωσης, είναι ένα θέμα το οποίο δεν έχει διερευνηθεί επαρκώς. Παράλληλα, για πρώτη φορά, λαμβάνεται υπόψη ένα τέτοιο δίκτυο βροχομετρικών σταθμών, το οποίο συνδυάζει πρωτογενή δεδομένα από δύο ανεξάρτητα δίκτυα, για το χωρικό εύρος του λεκανοπεδίου Αττικής.

Η εν λόγω διπλωματική εργασία αναπτύσσει και εφαρμόζει αφενός μία ολοκληρωμένη μεθοδολογία για την χωρική εκτίμηση της βροχόπτωσης σε διαφορετικές χρονικές βάσεις (μηνιαία, εποχική και ετήσια) και αφετέρου δίνει ιδιαίτερη έμφαση στην διαδικασία αξιολόγησης των επιμέρους στατιστικών προσεγγίσεων, και τελικά μπορεί να αξιοποιηθεί ως ένα εγχειρίδιο για ανάλογες εφαρμογές σε περιοχές που διαθέτουν ένα πυκνό δίκτυο σταθμών και διαθέσιμα δεδομένα για μεγαλύτερη χρονική περίοδο.

1.3 Διάρθρωση εργασίας

Το τεύχος της παρούσας διαρθρώνεται στα εξής 5 κεφάλαια:

Στο κεφάλαιο 1 (παρόν κεφάλαιο), όπου γίνεται μία πρώτη εισαγωγή στην αναγκαιότητα διερεύνησης μοντέλων παλινδρόμησης στη περιοχή της Αττικής, ενώ στο ίδιο κεφάλαιο αναλύεται το αντικείμενο της εργασίας και παρουσιάζεται η διάρθρωσή της.

Στο κεφάλαιο 2 προσδιορίζονται τα χαρακτηριστικά της περιοχής μελέτης που επιλέχθηκε και παρατίθενται τα διαθέσιμα πρωτογενή δεδομένα που χρησιμοποιήθηκαν στην ανάλυση.

Στο κεφάλαιο 3 παρουσιάζεται η μεθοδολογία που ακολουθήθηκε σχετικά με την ανάλυση, συμπλήρωση και επεξεργασία των πρωτογενών βροχομετρικών και γεωχωρικών δεδομένων, καθώς και η διαδικασία ανάπτυξης και τελικώς αξιολόγησης των μοντέλων παλινδρόμησης.

Στο κεφάλαιο 4 παρατίθενται τα αποτελέσματα που αφορούν την συμπλήρωση των βροχομετρικών χρονοσειρών, των εξαγόμενων μοντέλων παλινδρόμησης και την συγκριτική αξιολόγηση αυτών.

Στο κεφάλαιο 5 συνοψίζονται τα συμπεράσματα της εργασίας και προτείνονται θέματα για μελλοντική έρευνα.

Τέλος ακολουθεί η επιστημονική ξενόγλωσση και ελληνική βιβλιογραφία που αξιοποιήθηκε για την τεκμηρίωση της προσέγγισης κατά την εκπόνηση της παρούσας μεταπτυχιακής εργασίας.

2 ΠΕΡΙΟΧΗ ΜΕΛΕΤΗΣ ΚΑΙ ΠΡΩΤΟΓΕΝΗ ΔΕΔΟΜΕΝΑ

2.1 Εισαγωγή

Στο συγκεκριμένο κεφάλαιο γίνεται μία εκτενής περιγραφή της περιοχής μελέτης, του λεκανοπεδίου της Αττικής. Ειδικότερα παρατίθενται πληροφορίες και δεδομένα σχετικά με την διοικητική διάρθρωση και χωροταξία της περιοχής της Αττικής, τα γεωμορφολογικά χαρακτηριστικά της, αλλά και μία σειρά ειδικότερων χαρακτηριστικών που σχετίζονται με τις επικρατούσες κλιματολογικές συνθήκες. Πέρα από την περιοχή μελέτης στο παρόν κεφάλαιο απαριθμούνται τα πρωτογενή δεδομένα που χρησιμοποιήθηκαν για την εκπόνηση της παρούσας διπλωματικής εργασίας.

2.2 Χαρακτηριστικά Περιοχής Μελέτης

2.2.1 Γεωγραφικά και Γεωμορφολογικά Χαρακτηριστικά της Περιοχής Μελέτης

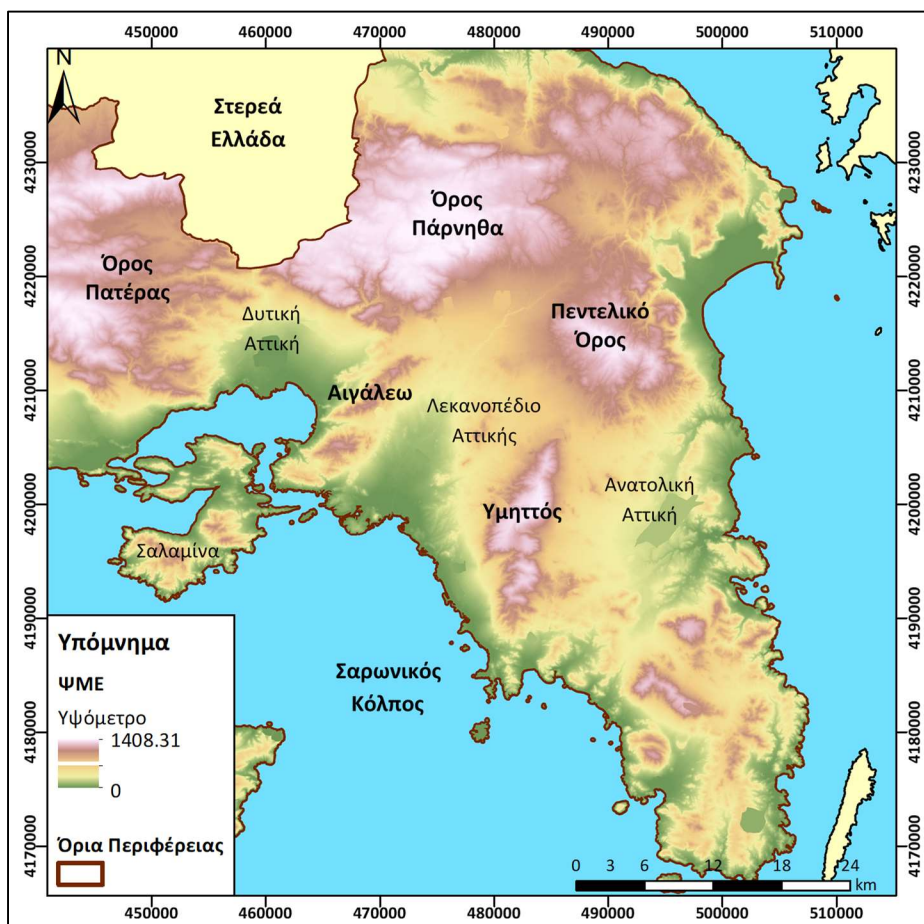
Η περιοχή μελέτης που επιλέχθηκε εμπερικλείει ένα κεντρικό ηπειρωτικό τμήμα της Περιφέρειας Αττικής όπως απεικονίζεται από το πλαίσιο στον επόμενο χάρτη (Σχήμα 2-1)



Σχήμα 2-1 Περιοχή μελέτης (μαύρο πλαίσιο), λεκάνη απορροής Κηφισού (μπλε περίγραμμα) και Π.Ε. Αττικής

Περιλαμβάνει τμήματα από όλες τις επιμέρους Περιφερειακές Ενότητες που απαρτίζουν την Περιφέρεια Αττικής, ενώ εμπεριέχει εξ' ολοκλήρου τις Π.Ε. του Βόρειου, Κεντρικού και Δυτικού Τομέα Αθηνών, καθώς και την Π.Ε. Πειραιώς. Το εύρος της περιοχής μελέτης επιλέχθηκε με στόχο να περιλαμβάνει το σύνολο της λεκάνης του Κηφισού, καθώς και ένα τμήμα περιφερειακά αυτού.

Το μεγαλύτερο τμήμα της Αττικής παρουσιάζει πεδινό έως ημιορεινό ανάγλυφο. Ειδικότερα όπως φαίνεται και στο Σχήμα 2-2 το ανάγλυφο του εδάφους στην μεγαλύτερη έκταση του είναι πεδινό, περιλαμβάνοντας ωστόσο ασυνεχείς μεταξύ τους εξάρσεις γης επιπέδου λόφου, όπως τα Τουρκοβούνια, του Φιλοπάππου, ο Λυκαβηττός και η Ακρόπολη. Αυτοί οι λόφοι αναπτύσσονται, με προσανατολισμό από τα βορειοανατολικά προς τα νοτιοδυτικά, ενώ το υψόμετρο σταδιακά μειώνεται προχωρώντας προς τα νοτιοδυτικά. Η περιοχή του Λεκανοπεδίου Αττικής περιβάλλεται από τέσσερις βασικούς ορεινούς όγκους. Στα ανατολικά βρίσκεται ο Υμηττός, στα βόρεια ο ορεινός σχηματισμός της Πεντέλης και η Πάρνηθα, και στα δυτικά το Όρος Αιγάλεω και αρκετές λοφοσειρές ή μεμονωμένοι λόφοι (Σχήμα 2-2). Η Πάρνηθα, με μέγιστο υψόμετρο 1413 m, είναι το ψηλότερο από τα βουνά του λεκανοπεδίου. Τέλος, το νότιο και δυτικό τμήμα του λεκανοπεδίου βρέχεται από τον Σαρωνικό κόλπο, ενώ ανατολικά βρέχεται από το Νότιο Ευβοϊκό κόλπο.



Σχήμα 2-2 Η ευρύτερη περιοχή μελέτης, στην Περιφέρεια Αττικής και το εδαφικό της ανάγλυφο

Οι κύριοι ποταμοί στο Λεκανοπέδιο είναι ο Κηφισός και ο Ιλισός, οι οποίοι διαρρέουν το Λεκανοπέδιο Αττικής. Ο Κηφισός εμφανίζει εκτεταμένο υδρογραφικό δίκτυο με πολλούς μικροχειμάρρους, οι κυριότεροι από τους οποίους είναι τα ρέματα της Χελιδονούς, του Κοκκιναρά, της Λαμπρινής και του Χαλανδρίου. Από τα 427 km² της συνολικής έκτασης του Λεκανοπεδίου Αττικής, τα 370 km² αποτελούν τη λεκάνη απορροής (Λ.Α.Π.) του ποταμού Κηφισού και τα 45 km² τη Λ.Α.Π. του Ιλισού (Gournelos and Maroukian, 1990). Σημειώνεται ότι το μεγαλύτερο μέρος του π. Ιλισού, ήδη από τις πηγές του, έχει υπογειοποιηθεί εδώ και αρκετά χρόνια, με στόχο την εύρυθμη λειτουργία του άσπεως. Η υπογειοποίηση του ποταμού ξεκίνησε επί Μεταξά στα τέλη της δεκαετίας του '30, από το ύψος της παλαιάς Σχολής Χωροφυλακής μέχρι την άλλοτε γέφυρα του Σταδίου, για να δημιουργηθεί η σημερινή λεωφόρος Μιχαλακοπούλου. Οι εργασίες επικάλυψης του Ιλισού ολοκληρώθηκαν στις αρχές της δεκαετίας του 1960. Από την αρχαιότητα και μέχρι τον 20ό αιώνα ο Ιλισός δεν εξέβαλε στη θάλασσα, αλλά ήταν παραπόταμος του Κηφισού, με τον οποίο συνέβαλε βόρεια του σημερινού Μοσχάτου. Κατά τη διάρκεια των έργων υπογειοποίησης του, ο Ιλισός εξετράπη και δημιουργήθηκε νέα κοίτη, κάτω από την Οδό Παναγή Τσαλδάρη στα όρια των Δήμων Μοσχάτου και Καλλιθέας, η οποία εκβάλλει στο μέσο του φαληρικού όρμου. Η αστικοποιημένη περιοχή κατά μήκος του ποταμού Κηφισού, καθώς και η περιοχή γύρω από το παλαιότερο μονοπάτι του ποταμού Ιλισού, έχουν συχνά υποστεί εκτεταμένες ζημιές κατά τη διάρκεια ακραίων γεγονότων καταιγίδας.

2.2.2 Κλίμα

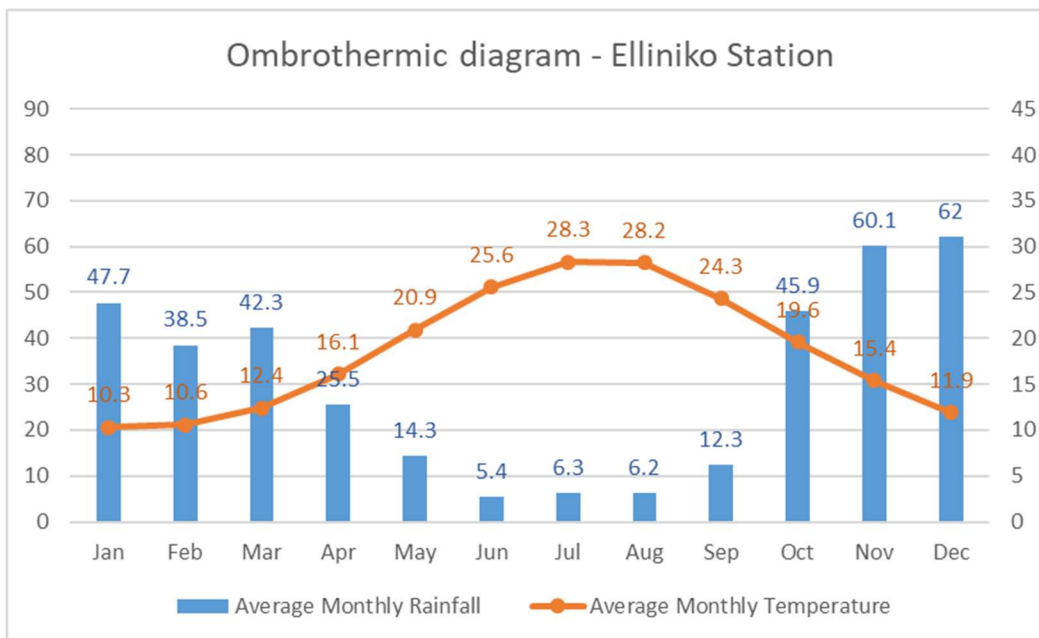
Το κλίμα της περιοχής χαρακτηρίζεται ως Μεσογειακό, σύμφωνα με την κλιματική ταξινόμηση Köppen, με εξαίρεση τα υψηλά σημεία, όπου προσομοιάζει περισσότερο σε ορεινό. Το Μεσογειακό κλίμα είναι ιδιαίτερο στον ευρύτερο γεωγραφικό χώρο της Ευρώπης και διαφέρει σημαντικά από τους άλλους δύο τύπους κλιμάτων που εντοπίζονται, το Ωκεάνιο (της δυτικής-Ατλαντικής Ευρώπης) και το Ηπειρωτικό (της ανατολικής Ευρώπης). Κατά τη διάρκεια της θερμής περιόδου (Απρίλιος-Σεπτέμβριος), επικρατούν συνθήκες ατμοσφαιρικού «εμποδισμού» με απόρροια το γνώριμο, σταθερό καιρό του καλοκαιριού στη Μεσόγειο, ενώ κατά την ψυχρή περίοδο έχει περισσότερες διακυμάνσεις, με απότομες αλλαγές και μεταβολές (Maheras & Anagnostopoulou, 2003). Σε γενικές γραμμές, θα μπορούσε κανείς να υποστηρίξει ότι καιρικά η Μεσόγειος έχει χαρακτηριστικά βορείου Αφρικής τη θερμή περίοδο και Κεντρικής Ευρώπης την ψυχρή.

Η Αττική είναι μία από τις περιοχές της Ελλάδας, στην οποία το φαινόμενο της έντονης αστικοποίησης, ιδιαίτερα κατά τις τελευταίες δεκαετίες είχε σαν άμεσο αποτέλεσμα τη διαμόρφωση ενός μικροκλίματος με διαφορετικά μικροκλιματολογικά χαρακτηριστικά από τα αντίστοιχα της αρχαιότητας (Καρύμπαλης και Παυλόπουλος, 2002). Η αστικοποίηση συνδυαστικά με την τοπογραφία της λεκάνης των Αθηνών είναι τα χαρακτηριστικά εκείνα τα οποία είναι υπεύθυνα για την δημιουργία των στοιχείων μικροκλίματος της περιοχής. Στην Αττική επικρατούν γενικά ήπιες θερμοκρασίες, χωρίς έντονες και απότομες διακυμάνσεις.

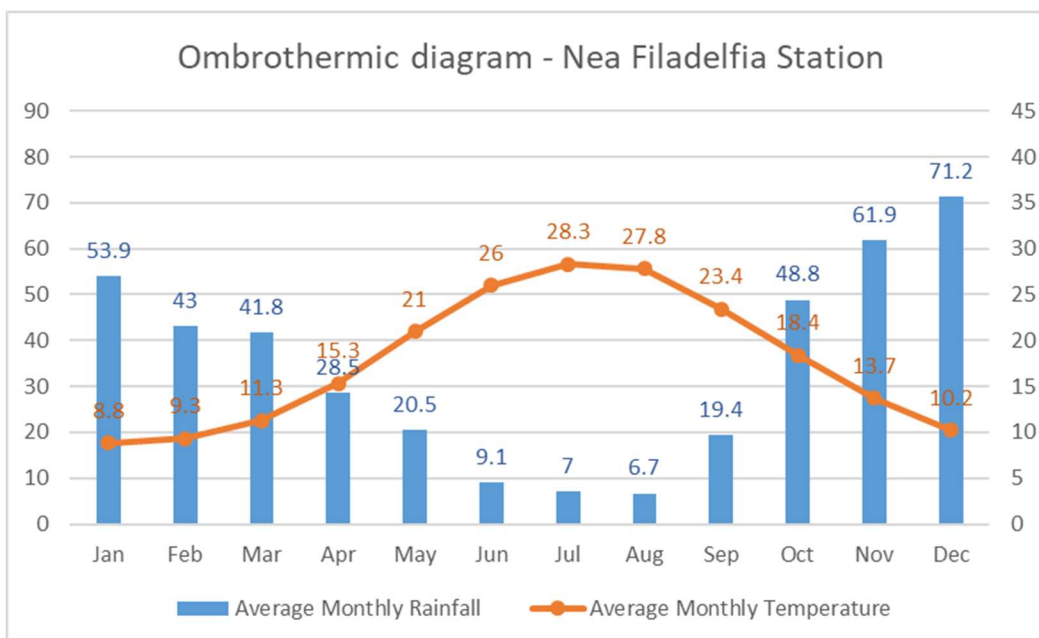
Από δεδομένα του Εθνικού Αστεροσκοπείου Αθηνών, για τον σταθμό του Θησείου, η μέση βροχόπτωση για χρονοσειρά δεδομένων 114 ετών (1899 – 2012) κυμαίνεται περίπου στα

400mm. Ωστόσο για ορεινά τμήματα, όπως την Πάρνηθα μπορεί να φτάσει και τα 1000mm. Η περίοδος των βροχοπτώσεων ξεκινά από τον Οκτώβριο και τελειώνει τον Μάιο. Σύμφωνα με χρονοσειρά δεδομένων 116 ετών (1897 – 2012), η μέση ετήσια θερμοκρασία φτάνει τους 18.3 βαθμούς. Ο θερμότερος μήνας είναι ο Ιούλιος με μέση μέγιστη θερμοκρασία 33°C, ενώ ο ψυχρότερος ο Ιανουάριος με μέση ελάχιστη θερμοκρασία 6.5°C. Η εκτεταμένη ηλιοφάνεια είναι ένα από τα σημαντικότερα κλιματολογικά χαρακτηριστικά της περιοχής. Η διάρκεια αυτής, σε ώρες, βρίσκεται σε υψηλά επίπεδα και κυμαίνεται μεταξύ 100-150 ωρών μηνιαίως τους χειμερινούς μήνες και μεταξύ 320-370 τους θερινούς μήνες. Οι άνεμοι που επικρατούν στο λεκανοπέδιο της Αττικής είναι σχεδόν όλο το χρόνο κυρίως βόρειοι και βορειοανατολικοί και σε μικρότερα ποσοστά από νότιες διευθύνσεις. Η ταχύτητα του επιφανειακού ανέμου ελαττώνεται τα τελευταία χρόνια. Σύμφωνα με τα δεδομένα του Εθνικού Αστεροσκοπείου Αθηνών για το λεκανοπέδιο προκύπτει η μέση ταχύτητα του ανέμου να κυμαίνεται μεταξύ των 2 και 4 m/s. Η σχετική υγρασία εκφράζεται επί τις εκατό και κυμαίνεται από 71,40% ανώτερη τιμή κατά τους χειμερινούς μήνες έως 45,10% κατά τους καλοκαιρινούς (<https://www.meteo.gr/climate/>).

Για τον χαρακτηρισμό του κλίματος χρησιμοποιούνται συνήθως οι παράγοντες της θερμοκρασίας και της κατακρήμνισης είτε για τον υπολογισμό συγκεκριμένων δεικτών, είτε για την απεικόνιση σχετικών κλιματικών διαγραμμάτων. Ένα χαρακτηριστικό μέγεθος, που αφορά αυτούς τους δύο κλιματικούς παράγοντες, και μπορεί να αποτυπωθεί μέσω διαγράμματος είναι η ξηρή και υγρή περίοδος του έτους. Οι Bagnouls and Gaussens (1957) απεικονίζουν γραφικά το κλίμα της περιοχής μέσω του ομβροθερμικού διαγράμματος. Το διάγραμμα καταρτίζεται σύμφωνα με την μηνιαία πορεία του μέσου μηνιαίου ύψους βροχής P σε mm και της μέσης μηνιαίας θερμοκρασίας T σε °C με κλίμακα όμως θερμοκρασιών διπλάσια της βροχόπτωσης $P = 2T$. Σύμφωνα με τους ίδιους συγγραφείς ένα μήνας χαρακτηρίζεται «ξηρός» όταν το σύνολο των κατακρημνισμάτων κατά την διάρκεια του είναι ίσο ή μικρότερο από το διπλάσιο της μέσης θερμοκρασίας του, δηλαδή αν $P \text{ mm} < 2 T \text{ }^{\circ}\text{C}$ (αυτή η εμπειρική σχέση έχει υιοθετηθεί και από τον UNESCO–FAO). Η κλειστή περιοχή μεταξύ των δύο καμπύλων του διαγράμματος δείχνει τη διάρκεια της περιόδου και την ένταση της ξηρασίας. Έτσι συγκεκριμένα για την περιοχή μελέτης και με σύμφωνα με τα ανοιχτά δεδομένα στον ιστότοπο της ΕΜΥ, καταρτίστηκαν τα ομβροθερμικά διαγράμματα των σταθμών του Ελληνικού και της Νέας Φιλαδέλφειας, όπως απεικονίζονται παρακάτω (Σχήμα 2-3, Σχήμα 2-4). Οι συγκεκριμένοι σταθμοί επιλέχθηκαν με γνώμονα τον έλεγχο του νότιου αλλά και του βόρειου τμήματος του λεκανοπεδίου για τυχόν διαφοροποιήσεις.



Σχήμα 2-3 Ομβροθερμικό διάγραμμα – Σταθμός Ελληνικού (Πηγή δεδομένων: Κλιματικός Άτλαντας ΕΜΥ – Ιδία επεξεργασία)



Σχήμα 2-4 Ομβροθερμικό διάγραμμα – Σταθμός Ελληνικού (Πηγή δεδομένων: Κλιματικός Άτλαντας ΕΜΥ – Ιδία επεξεργασία)

Είναι εμφανές και από τα δύο διαγράμματα, ότι η ξηρή περίοδος του έτους αντιστοιχεί στο εύρος μηνών Απριλίου έως και Σεπτεμβρίου. Ενώ για μια συνεχόμενη υγρή περίοδο ξεκινά από τον Οκτώβριο και κλείνει με το πέρας του Μαρτίου του επόμενου ημερολογιακού έτους. Οι διάρκειες αυτές ξηρής και υγρής περιόδου αντίστοιχα θα χρησιμοποιηθούν για την κατάταξη των δεδομένων βροχομετρικών δεδομένων στις δύο αυτές περιόδους με στόχο την προσαρμογή

μοντέλου με αποκρίσεις αυτές τις χρονικές κλίμακες, όπως θα αναλυθεί στη συνέχεια στο κεφάλαιο της Μεθοδολογίας.

Σε ότι αφορά την τάση εξέλιξης του καθεστώτος των βροχοπτώσεων στην περιοχή της Αττικής, αυτή συνήθως αποτυπώνεται μέσω του δείκτη ραγδαιότητας (RI), ο οποίος προκύπτει από την μέση ετήσια ένταση της βροχόπτωσης (mm/h). Σύμφωνα με τους Nastos et al. (2010), για την περίοδο 1962-2002 παρατηρούνται θετικές τάσεις στη ραγδαιότητα της βροχής (+0.43 mm/h ανά δεκαετία), της αυτές προέκυψαν από την επεξεργασία των μηνιαίων υψών βροχόπτωσης που καταγράφηκαν από μετεωρολογικό σταθμό του Εθνικού Αστεροσκοπείου Αθηνών. Οι θετικές τάσεις της ραγδαιότητας οφείλονται στο φαινόμενο της αστικής θερμικής νησίδας, η οποία είναι υπεύθυνη για ισχυρές καταιγίδες μεταγωγικής φύσεως (Nastos & Zerefos, 2007). Συγκεκριμένα, σε εποχική βάση, της έντασης της βροχής είναι μεγαλύτερη το καλοκαίρι και το φθινόπωρο απ' ό,τι την άνοιξη και το χειμώνα, γεγονός που υποδηλώνει περισσότερα επεισόδια σφοδρών καταιγίδων το καλοκαίρι και το φθινόπωρο (Nastos et al., 2010).

Συμπερασματικά το κλίμα της Αττικής έχει οδηγήσει σε ένα καθεστώς βροχοπτώσεων, το οποίο θα μπορούσε να χαρακτηριστεί διαταραγμένο, καθώς παράγει μικρό σχετικά ετήσιο ύψος βροχής αλλά αρκετά επεισόδια μεγάλων εντάσεων. Η συμπεριφορά αυτή έχει ως αποτέλεσμα να μην ευνοείται η κατείσδυση, να αυξάνεται επομένως η επιφανειακή απορροή και συνακόλουθα η πιθανότητα εκδήλωσης πλημμυρών εξαιτίας έντονων βροχοπτώσεων μικρής διάρκειας. Τα τελευταία 117 έτη έχουν σημειωθεί πάνω από 25 σοβαρά πλημμυρικά επεισόδια (Maroukian et al., 2005) με τα περισσότερα να καταγράφονται ανοδικά τις τελευταίες δεκαετίες σύμφωνα με τα στοιχεία της Εθνικής Γραμματείας Υδάτων (Πίνακας 2-1) Η έντονη και άναρχη αστική ανάπτυξη καθώς και ο κακός σχεδιασμός και συντήρηση των αποστραγγιστικών δικτύων συμβάλλουν καθοριστικά στην ενίσχυση τέτοιων επεισοδίων.

Πίνακας 2-1 Ιστορικά πλημμυρικά φαινόμενα και σημαντικές ιστορικές πλημμύρες στο ΥΔ Αττικής την περίοδο 1930-2009 (Εθνική Γραμματεία Υδάτων, 2017)

<i>Χρονικό διάστημα</i>	<i>Ιστορικές Πλημμύρες</i>	<i>Σημαντικές Ιστορικές Πλημμύρες</i>
<1930	6	6
1930-1960	5	5
1961-1980	29	19
1981-2000	47	18
2001-2009	66	6
Σύνολο	153	54

2.3 Διαθέσιμα Πρωτογενή Δεδομένα

2.3.1 Βροχομετρικά Δεδομένα

Το πιο καθοριστικό ρόλο για την ανάπτυξη των μοντέλων εκτίμησης βροχόπτωσης εντός της περιοχής μελέτης, που περιεγράφηκε προηγουμένως, αποτελούν τα πρωτογενή δεδομένα βροχόπτωσης. Στόχος της παρούσας διπλωματικής είναι η διερεύνηση μοντέλων, τα οποία ανταποκρίνονται επαρκώς, όχι μόνο στην ποσοτική εκτίμηση της βροχόπτωσης, αλλά λαμβάνουν υπόψη και την χωρική μεταβλητότητα αυτής. Συνεπώς και υπό αυτό το πρίσμα επιλέχθηκε ένα δίκτυο μετρητών βροχόπτωσης ομοιόμορφης χωρικής κάλυψης εντός της περιοχής μελέτης. Έτσι συλλέχθηκαν και στην συνέχεια επεξεργάστηκαν (όπως περιγράφεται εκτενώς στο επόμενο κεφάλαιο της Μεθοδολογίας) διαθέσιμες χρονοσειρές 10λεπτης βροχόπτωσης από ένα αρχικό σύνολο 30 σταθμών. Μέρος αυτών των σταθμών ανήκουν στο δίκτυο σταθμών του Εθνικού Αστεροσκοπείου Αθηνών NOA (Lagouvardos et al., 2017) , ενώ οι υπόλοιποι ανήκουν στο δίκτυο σταθμών του Υδρολογικού Παρατηρητήριου Αθήνας ΗΟΑ, το οποίο αποτελεί μία εξέλιξη του δικτύου METEONET (Grammatikogiannis et al., 2005).

Πίνακας 2-2 Σταθμοί δικτύου ΝΟΑ και ΗΟΑ και οι συντεταγμένες τους Χ, Υ σύμφωνα με το σύστημα συντεταγμένων ΕΓΣΑ '87

α/α	ΝΟΑ	Χ	Υ	α/α	ΗΟΑ	Χ	Υ
1	Αμπελόκηποι	478631	4203530	1	Άγιος Κοσμάς	475060	4193690
2	Ανάβυσσος	491500	4175991	2	Άνω Λιόσια	470968	4213918
3	Αθήνα	474861	4203170	3	Γαλάτσι	478564	4208803
4	Διόνυσος	490196	4217492	4	Ηλιούπολη	478838	4196512
5	Εκάλη	484311	4215491	5	Μάνδρα	461615	4219262
6	Φάληρο	472875	4197735	6	Μενίδι	476080	4217305
7	Ιπποκράτειος	481516	4227284	7	Πεντέλη	487813	4215150
8	Κάντζα	488048	4203239	8	Πικέρμι	493583	4205666
9	Λαύριο	504963	4174086	9	Ψυττάλεια	464231	4199302
10	Μαρκόπουλο	494311	4191882	10	Ζωγράφου	480493	4203306
11	Μαρούσι	483450	4211276				
12	Νέα Μάκρη	497811	4212060				
13	Νέα Σμύρνη	474805	4200123				
14	Πάρνηθα	475189	4223648				
15	Πατήσια	476134	4207998				
16	Περιστέρι	473845	4205768				
17	Πόρτο Ράφτη	501228	4192916				
18	Ψυχικό	480593	4207503				
19	Σπάτα	493449	4203828				
20	Βάρη	482516	4187017				

2.3.2 Γεωχωρικά Δεδομένα

Στο πλαίσιο της παρούσας εργασίας, πέραν των βροχομετρικών στοιχείων, συλλέχθηκε ένα σύνολο από γεωχωρικά δεδομένα. Τα δεδομένα αυτά χρησιμοποιήθηκαν σε διάφορα στάδια της μεθοδολογίας, όπως αυτά αναλύονται στο επόμενο κεφάλαιο της Μεθοδολογίας.

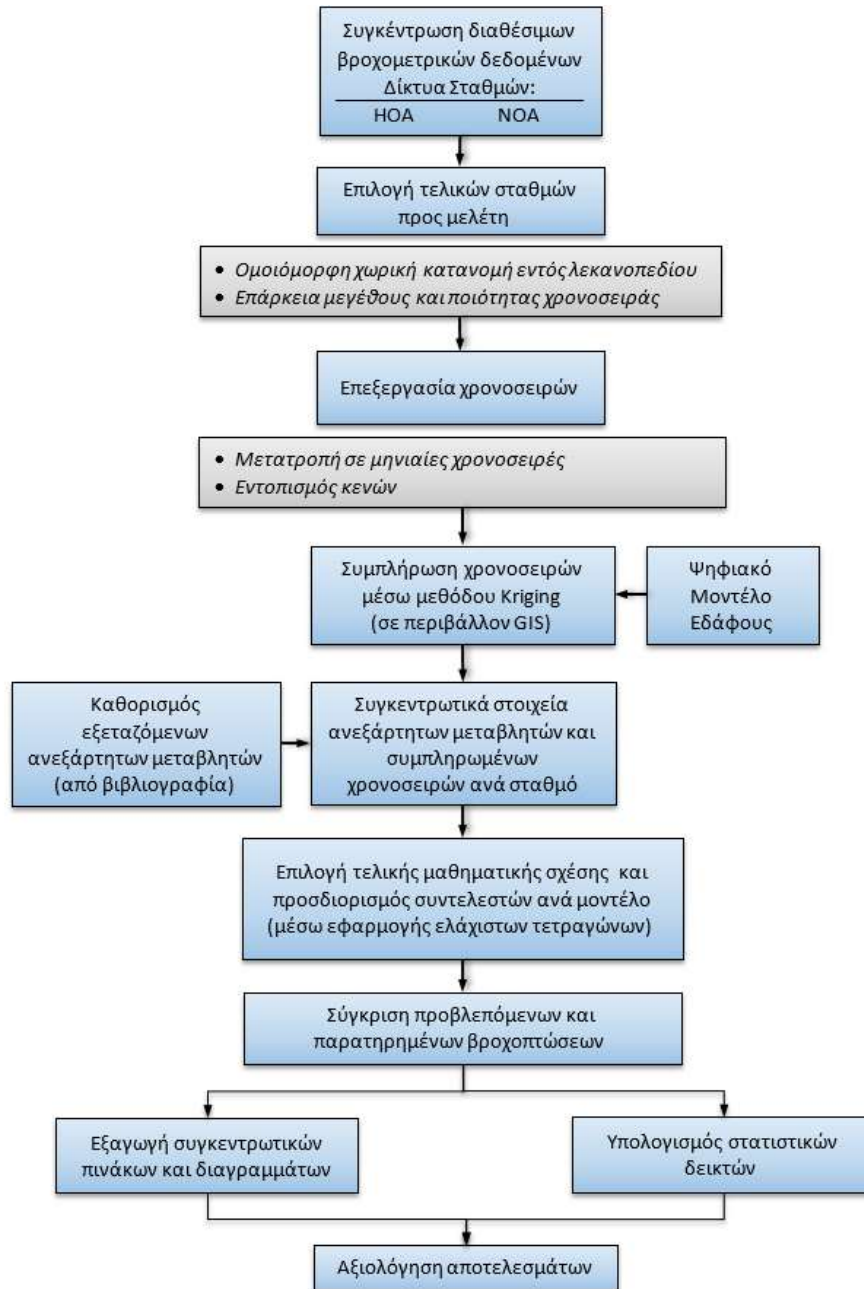
Το βασικότερο σύνολο δεδομένων που χρησιμοποιήθηκε για πολλαπλούς σκοπούς είναι το ψηφιακό μοντέλο εδάφους (DEM), το οποίο παρέχεται από την εταιρία «Κτηματολόγιο Α.Ε.», (Copyright © 2012, National Agency for Cadastre & Mapping S.A.), σε υψηλή χωρική ανάλυση με μέγεθος φατνίου 5m x 5m.

Πέρα από το ψηφιακό μοντέλο εδάφους χρησιμοποιήθηκε το γεωχωρικό πολύγωνο που εμπερικλείει την υδρολογική Λεκάνη του Κηφισού (Baltas et al. 2015). Ακόμη χρησιμοποιήθηκαν τα πολυγωνικά αρχεία της ακτογραμμής και των ορίων του Νομού Αττικής όπως αυτά παρέχονται από τον ιστότοπο του εθνικού καταλόγου ανοικτών γεωχωρικών δεδομένων (geodata.gov).

3 ΜΕΘΟΔΟΛΟΓΙΑ

3.1 Γενική Επισκόπηση

Στο κεφάλαιο αυτό γίνεται η περιγραφή της μεθοδολογίας που ακολουθήθηκε στα πλαίσια της διπλωματικής εργασίας. Ειδικότερα στο επόμενο διάγραμμα ροής παρουσιάζεται συνοπτικά η διαδικασία που ακολουθήθηκε για την προετοιμασία – επεξεργασία των βροχομετρικών δεδομένων, καθώς και των γεωχωρικών δεδομένων με στόχο την ανάπτυξη μοντέλων πρόβλεψης βροχοπτώσης, για την εξεταζόμενη περιοχή μελέτης.



Σχήμα 3-1 Γενικό διάγραμμα ροής μεθοδολογίας

3.2 Αναλυτική Περιγραφή

3.2.1 Επιλογή βροχομετρικών σταθμών

Αρχικά συλλέγονται τα διαθέσιμα βροχομετρικά δεδομένα για σταθμούς που εντοπίζονται εντός του Νομού Αττικής. Πιο συγκεκριμένα σε πρώτο στάδιο, επιλέχθηκαν συνολικά 30 μετεωρολογικοί σταθμοί, 10 από το δίκτυο ΗΟΑ και 20 από το Εθνικό Αστεροσκοπείο Αθηνών (ΝΟΑ), οι οποίοι αναγράφονται στον επόμενο πίνακα. Από το σύνολο των 30 σταθμών επιλέχθηκαν προς επεξεργασία οι χρονοσειρές των 21 σταθμών. Οι απορριπτέοι σταθμοί επισημαίνονται με διαφορετικό χρώμα στον επόμενο πίνακα.

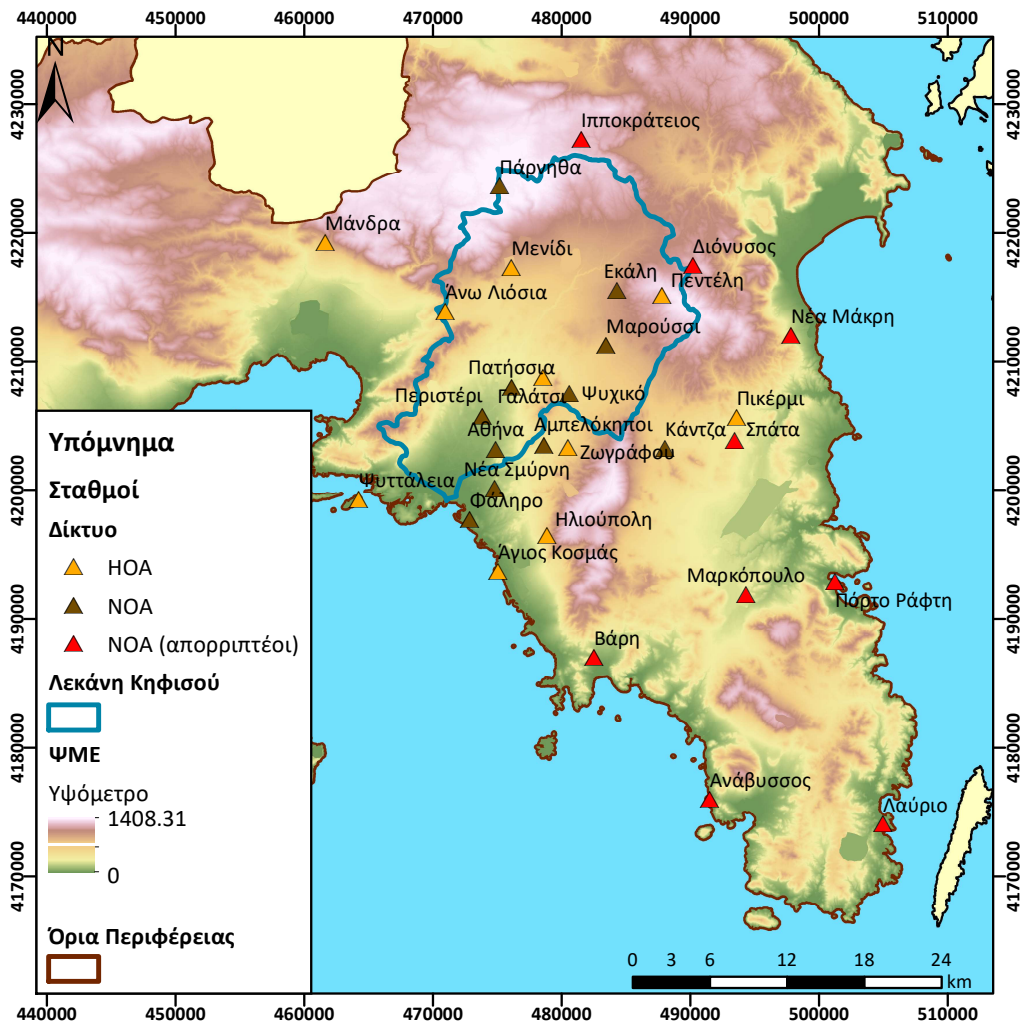
Πίνακας 3-1 Δίκτυο διαθέσιμων μετεωρολογικών σταθμών (ΝΟΑ & ΗΟΑ)

α/α	ΝΟΑ	α/α	ΗΟΑ
1	Αμπελόκηποι	1	Άγιος Κοσμάς
2	Ανάβυσσος	2	Άνω Λιόσια
3	Αθήνα	3	Γαλάτσι
4	Διόνυσος	4	Ηλιούπολη
5	Εκάλη	5	Μάνδρα
6	Φάληρο	6	Μενίδι
7	Ιπποκράτειος	7	Πεντέλη
8	Κάντζα	8	Πικέρμι
9	Λαύριο	9	Ψυττάλεια
10	Μαρκόπουλο	10	Ζωγράφου
11	Μαρούσι		
12	Νέα Μάκρη		
13	Νέα Σμύρνη		
14	Πάρνηθα		
15	Πατήσια		
16	Περιστέρι		
17	Πόρτο Ράφτη		
18	Ψυχικό		
19	Σπάτα		
20	Βάρη		

Η τελική εκλογή των χρονοσειρών – σταθμών προς επεξεργασία έγινε βάσει του χρονικού εύρους των διαθέσιμων καταγραφών κάθε σταθμού, καθώς και της αρτιότητας αυτών. Σταθμοί με τμηματικές καταγραφές και κενά στις χρονοσειρές κρίθηκαν ακατάλληλοι, εφόσον ακόμη και η μετέπειτα συμπλήρωση αυτών θα οδηγούσε σε χρονοσειρές με ισχυρή αβεβαιότητα, απρόσφορων για την δημιουργία των μετ' έπειτα μοντέλων βροχόπτωσης.

Οι διαθέσιμες καταγραφές των σταθμών που ανήκουν στο δίκτυο σταθμών του Εθνικού Αστεροσκοπείου (ΝΟΑ) υπήρξαν ελλιπείς τόσο σε ότι αφορά το εύρος, όσο και στην ύπαρξη ενδιάμεσων κενών. Επομένως, ένα δεύτερο κριτήριο επιλογής μεταξύ σταθμών με παρόμοια

προβλήματα μεταξύ τους, ήταν η επίτευξη ενός καλά οριοθετημένου δικτύου σταθμών με όσο το δυνατόν περισσότερο ομοιόμορφη χωρική κατανομή, εντός και πέριξ της λεκάνης απορροής του Κηφισού.



Σχήμα 3-2 Χάρτης με δίκτυο σταθμών HOA & NOA και απορριπτέων σταθμών και τα όρια της λεκάνης απορροής Κηφισού

3.2.2 Επεξεργασία πρωτογενών χρονοσειρών

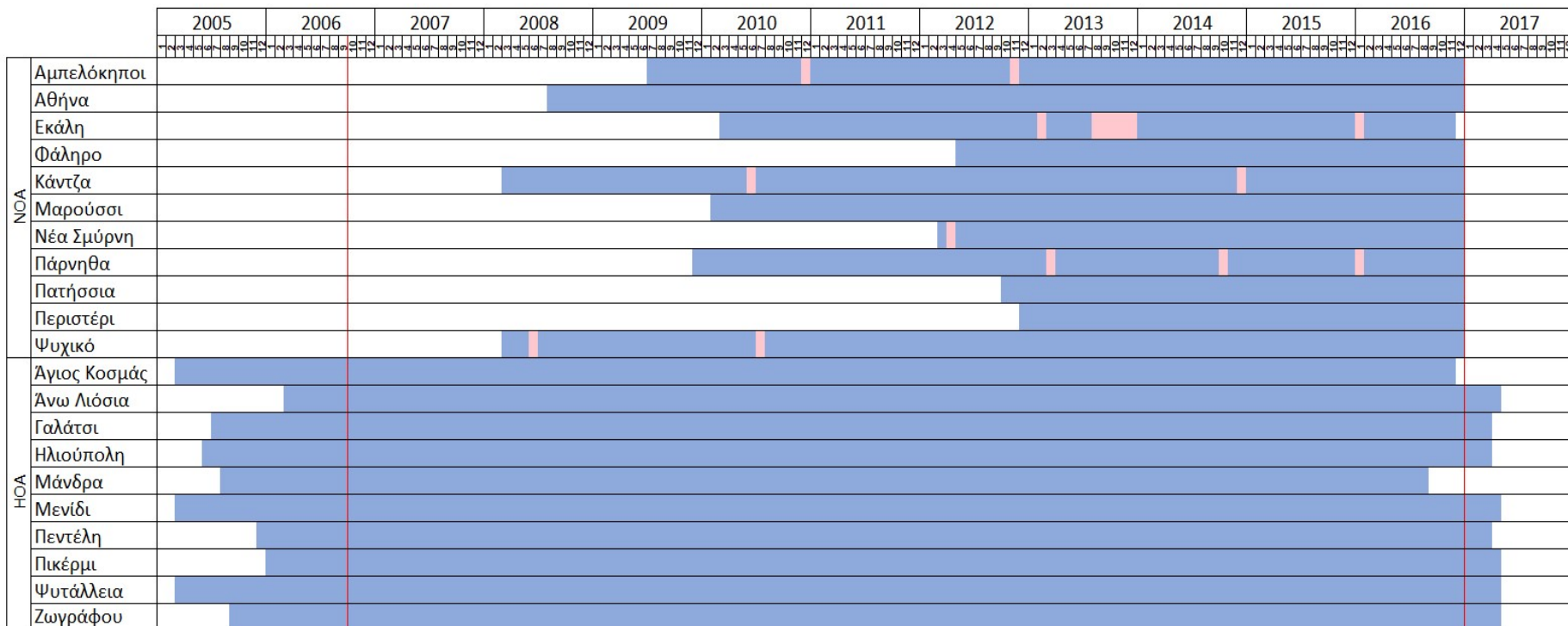
Μετά την τελική επιλογή των σταθμών ακολουθεί η επεξεργασία των χρονοσειρών τους. Η επεξεργασία των πρωτογενών βροχομετρικών δεδομένων, με τη μορφή χρονοσειρών, έγινε εξ' ολοκλήρου μέσω της σύνταξης και εκτέλεσης κώδικα στο προγραμματιστικό περιβάλλον της MATLAB.

Οι διαθέσιμες χρονοσειρές, ανεξάρτητα προέλευσης δικτύου, έχουν για κάθε σταθμό, χρονικό βήμα 10 λεπτών. Λόγω του υψηλού αριθμού ενδιάμεσων κενών που εντοπίστηκαν στις χρονοσειρές του δικτύου NOA, κρίθηκε σκόπιμο τα μοντέλα που θα αναπτυχθούν να αναφέρονται σε μηνιαία (και μεγαλύτερη) χρονική κλίμακα με στόχο την εξαγωγή ασφαλών

συμπερασμάτων, εξαιτίας της μεγαλύτερης συνοχής των προκυπτουσών χρονοσειρών μηνιαίου βήματος και της μείωσης του αριθμού των ενδιάμεσων κενών. Η μεγάλη μεταβλητότητα σε συνδυασμό με τα πολυάριθμα κενά καταγραφών εντός της χρονοσειράς ενδέχεται να οδηγήσουν σε πιο επισφαλή δεδομένα για την στατιστική ανάλυση στην εξαγωγή των απαιτούμενων μοντέλων πρόβλεψης, στο σενάριο όπου τα μοντέλα αναφέρονταν σε ημερήσιο βήμα. Συνεπώς στόχος είναι η εξαγωγή συνολικά 15 μοντέλων: 12 διαφορετικών μοντέλων ανά μήνα, καθώς επίσης ένα μοντέλο με ετήσια χρονική κλίμακα και ακόμη δύο που αναφέρονται στην υγρή και ξηρή περίοδο του έτους.

Αρχικά προσδιορίζεται η ευρύτερη συνολική περίοδος διαθέσιμων καταγραφών, έναρξη και λήξη ανά σταθμό. Στη συνέχεια, εντοπίζονται τα χρονικά σημεία των ενδιάμεσων κενών (μη καταγεγραμμένων τιμών). Δεδομένου του μήκους των διαθέσιμων χρονοσειρών για κάθε σταθμό, το εύρος του χρονικού διαστήματος προς μελέτη οριοθετείται από τον Οκτώβρη του 2006³ έως και τον Δεκέμβρη του 2016, με στόχο την εκμετάλλευση ενός συνόλου δεδομένων βροχής δεκαετίας. Το μήκος των χρονοσειρών κάθε σταθμού, τα καταγεγραμμένα κενά εντός αυτών και το καθολικά επιλεγμένο χρονικό διάστημα προς μελέτη παρουσιάζονται στο Σχήμα 3-3. Το συγκεκριμένο διάστημα θεωρείται η προσφορότερη λύση για την δεδομένη περίπτωση, καθώς για την περίπτωση μεγαλύτερου εύρους συναντάται ακόμη μικρότερη διαθεσιμότητα δεδομένων, ακόμα και για τους σταθμούς του δικτύου ΗΟΑ, ενώ η περίπτωση μικρότερου εύρους, όπου δεν συμπληρώνει την δεκαετία, κρίνεται ελλιπές για την ανάπτυξη μοντέλων με στόχο την μηνιαία και πλέον πρόβλεψη ύψους βροχής.

³ Ανάμεσα στο σύνολο των 15 μοντέλων που μελετώνται, αναπτύσσονται δύο εποχιακά που αφορούν την ξηρή και υγρή περίοδο του έτους. Για το μοντέλο της υγρής περιόδου και την συμπλήρωση δειγμάτων δεκαετίας υγρών περιόδων απαιτούνται και οι υγροί μήνες από το έτος του 2016. Έτσι, διευκρινίζεται ότι η τελικώς επιλεγόμενη χρονοσειρά δεν περιορίζεται στην δεκαετία 2007-2016, καθώς συμπεριλαμβάνει και τους συγκεκριμένους μήνες από το 2016.



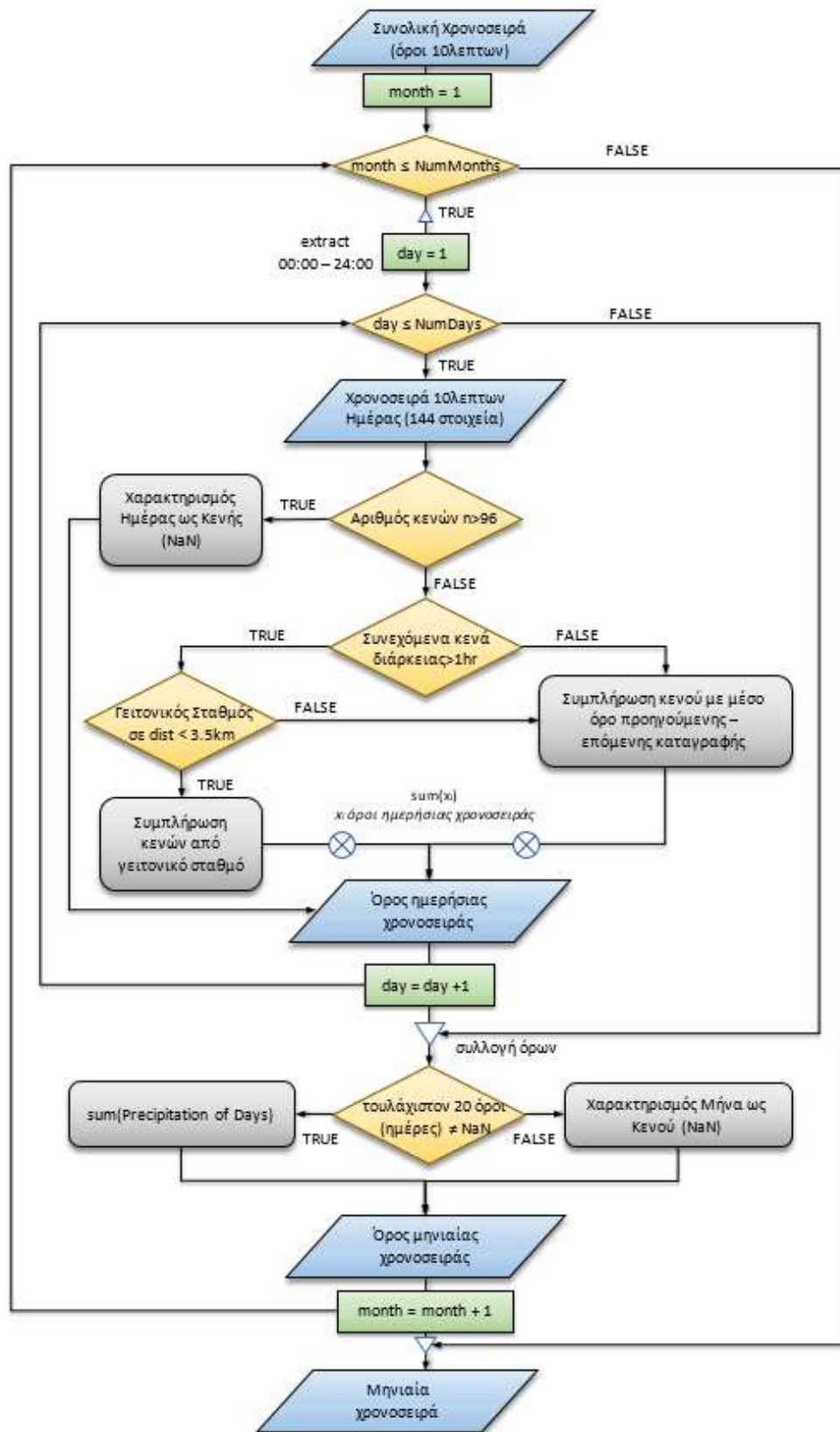
Σχήμα 3-3 Χρονικά εύρη διαθέσιμων πρωτογενών χρονοσειρών ανά σταθμό και δίκτυο. Τα κελιά με ροζ χρωματική απόχρωση αφορούν τα ενδιάμεσα κενά της χρονοσειράς, ενώ οι κάθετες κόκκινες γραμμές σηματοδοτούν την έναρξη και την λήξη του επιλεγμένου χρονικού διαστήματος (10ετίας), εντός του οποίου θα διενεργηθεί η προσομοίωση

Οι διαθέσιμες χρονοσειρές, ανεξάρτητα προέλευσης δικτύου, έχουν για κάθε σταθμό, χρονικό βήμα 10 λεπτών. Έτσι για την δημιουργία μοντέλων με μηνιαία απόκριση βροχής, τα διαθέσιμα πρωτογενή δεδομένα πρέπει να συγχωνευθούν αρχικά σε επίπεδο ημέρας και στη συνέχεια αθροιστικά σε επίπεδο μήνα. Ωστόσο η συγκεκριμένη διαδικασία δεν μπορεί να γίνει σειριακά και αδιάκοπα, λόγω της ύπαρξης κενών στις καταγραφές.

Τα κενά αξιολογούνται ως προς τον συνολικό αριθμό τους σε επίπεδο ημέρας, με στόχο τον χαρακτηρισμό ή όχι της εκάστοτε ημέρας ως κενή καταγραφών. Εάν ο αριθμός των κενών είναι μικρός, αυτά συμπληρώνονται ως ο μέσος όρος της καταγραφής ακριβώς πριν το κενό και της αμέσως επόμενης καταγραφής μετά την λήξη της κενής περιόδου. Σε κάποιες περιπτώσεις για την συμπλήρωση, λήφθηκαν υπόψη οι καταγραφές γειτονικών σταθμών για τις ίδιες χρονικές στιγμές, ιδιαίτερα όταν αυτοί οι σταθμοί είχαν σχετικά μικρή απόσταση μεταξύ τους (π.χ. Αθήνα, Αμπελόκηποι, Πατησιά, Περιστέρι).

Οι καταγεγραμμένες τιμές γειτονικών σταθμών χρησιμοποιήθηκαν όχι μόνο για την κάλυψη ενδεχόμενων κενών στην χρονοσειρά, αλλά και την ποιοτική αξιολόγηση των καταγραφών και την αποφυγή τεχνικών σφαλμάτων. Έτσι ελέγχθηκε εάν οι καταγεγραμμένες τιμές ενός σταθμού διέφεραν κατά μεγάλο ποσοστό από αυτές γειτονικού σταθμού για τις ίδιες χρονικές στιγμές.

Μετά την ποιοτική αξιολόγηση των τιμών, την μερική συμπλήρωση κενών (όπου αυτό ήταν δυνατό – μικρές περίοδοι) και τον συνεπακόλουθο χαρακτηρισμό συγκεκριμένων ημερών ως κενών (εξαιτίας πολυάριθμων κενών), οι χρονοσειρές απομονώνονται σε τμήματα μεταξύ ελλειπουσών καταγραφών και στη συνέχεια ανάγονται (άθροισμα βροχόπτωσης ημερών ενός μήνα) σε χρονοσειρές μηνιαίου βήματος. Στο επόμενο διάγραμμα ροής (Σχήμα 3-4) περιγράφεται σχηματικά και αναλυτικά η διαδικασία σύνταξης κώδικα που ακολουθήθηκε για την τελική διαμόρφωση των μηνιαίων χρονοσειρών και τον εντοπισμό των τελικών κενών που επισημαίνονται στο Σχήμα 3-3.



Σχήμα 3-4 Αναλυτικό διάγραμμα ροής επεξεργασία πρωτογενών δεδομένων για τη διαμόρφωση μηνιαίων χρονοσειρών

3.2.3 Επιλογή ανεξάρτητων μεταβλητών και επεξεργασία γεωχωρικών δεδομένων

Στο πλαίσιο της ανάπτυξη μοντέλων εκτίμησης βροχόπτωσης, για την περιοχή μελέτης, καθορίζεται μία σειρά από γεωμορφολογικούς παράγοντες, οι οποίοι συνδέονται (ή ενδέχεται να συνδέονται) με το ύψος της βροχόπτωσης. Στη συνέχεια διερευνάται ποιοι από αυτούς θα συμπεριλαμβάνονται ως ανεξάρτητες μεταβλητές στο εκάστοτε μοντέλο. Κατά συνέπεια συλλέχθηκε ένα σύνολο από γεωχωρικά δεδομένα, όπως αυτά παρατίθενται στο σχετικό κεφάλαιο των διαθέσιμων Πρωτογενών Δεδομένων. Η πρωταρχική εννοιολογική επιλογή αυτών των γεωχωρικών μεταβλητών έγινε βάσει της σχετικής επιστημονικής βιβλιογραφίας (Gofa et al., 2019; Sevruk, 1997; Gouvas et al., 2009). Έτσι οι τελικές προς εξέταση μεταβλητές είναι το υψόμετρο, οι γεωγραφικές συντεταγμένες X και Y κατά ΕΓΣΑ '87, η κλίση εδάφους, ο προσανατολισμός, η απόσταση από ποτάμιο υδατικό σύστημα και η απόσταση από την ακτογραμμή.

Το σύνολο της γεωχωρικής πληροφορίας από το πρωτογενές επίπεδο έως και την σταδιακή επεξεργασία τους, καθώς και την παραγωγή των απαιτούμενων μεταδεδομένων έγινε μέσω της χρήσης συστήματος γεωγραφικών πληροφοριών (GIS) και πιο συγκεκριμένα μέσω του λογισμικού ArcGIS (έκδοσης 10.5.1), όπου όλες οι αρχικές και οι παραγώμενες οντότητες αποθηκεύτηκαν σε κατάλληλα σχεδιασμένη γεωβάση.

Αρχικά ορίζεται το σύστημα συντεταγμένων, το ΕΓΣΑ '87 (Greek Grid '87), σύμφωνα με το οποίο θα ορίζονται πλέον όλα τα αρχικά και στη συνέχεια τα εξαγόμενα γεωχωρικά επίπεδα. Στη συνέχεια εισάγονται στον χάρτη τα γεωγραφικά στοιχεία των σταθμών (όνομα, προέλευση δικτύου NOA ή HOA και συντεταγμένες X και Y) και δημιουργώντας έτσι ένα σημειακό επίπεδο με τις θέσεις των σταθμών (point layer, με στοιχεία αποθηκευμένα στον Attribute Table). Εισάγεται ακόμη το πολυγωνικό επίπεδο που εμπερικλείει τη λεκάνη απορροής του Κηφισού. Βάσει αυτών των δύο αυτών επιπέδων δημιουργείται ένα νέο βοηθητικό πολυγωνικό επίπεδο, εντός του οποίου θα στηριχθεί η μεθοδολογία εξαγωγής των επόμενων απαιτούμενων χωρικών επιπέδων. Το συγκεκριμένο επίπεδο οφείλει να περιλαμβάνει κάθε σταθμό και να περικλείει εντός του την λεκάνη απορροής του Κηφισού και δημιουργείται με στόχο την ελαχιστοποίηση του υπολογιστικού φόρτου.

Στη συνέχεια σειρά έχει η εισαγωγή του ψηφιακού μοντέλου εδάφους (DEM) και η αποκοπή του σύμφωνα με το βοηθητικό πολυγωνικό επίπεδο εργασίας. Για λόγους ελαχιστοποίησης υπολογιστικού φόρτου και χωρικής αβεβαιότητας, επιλέχθηκε το σύνολο των τιμών να αντιστοιχούν σε χωρικό φάτνιο διαστάσεων 100m x 100m. Εξαιτίας αυτού η χωρική ακρίβεια του DEM μετατράπηκε από 5x5 σε 100x100 m.

Ύστερα μέσω του πλέον επεξεργασμένου ψηφιακού μοντέλου εδάφους εξάχθηκαν τα χωρικά επίπεδα της κλίσης εδάφους και προσανατολισμού (εντολές slope και aspect, από το πακέτο Spatial Analyst).

Έπειτα, με στόχο την εύρεση της απόστασης κάθε σταθμού από ποτάμια υδατικά συστήματα και με δεδομένο το ψηφιακό μοντέλο εδάφους, δημιουργείται νέο επίπεδο με τα ποτάμια στην

ευρύτερη περιοχή μελέτης, μέσω της χρήσης της διαθέσιμης εργαλειοθήκης Hydrology. Στη συνέχεια, μέσω υπολογισμού των ευκλείδειων αποστάσεων παράγεται νέο χωρικό επίπεδο με την κοντινότερη απόσταση κάθε σημείου της περιοχής μελέτης από ποτάμιο υδατικό σύστημα.

Με χρήση και πάλι του εργαλείου της ευκλείδειας απόστασης και δεδομένου του πολυγωνικού επιπέδου της ακτογραμμής, εξάγεται νέο επίπεδο με την κοντινότερη απόσταση κάθε σημείου πλέον από την θάλασσα.

Παρότι οι συντεταγμένες για κάθε έναν από τους επιλεγμένους σταθμούς είναι γνωστές και ακριβείς, για λόγους ομοιογένειας ως προς την χωρική ακρίβεια του συνόλου των δεδομένων που θα εξεταστούν ως ανεξάρτητες μεταβλητές, εξάγονται και τα επίπεδα με τη χωρική πληροφορία συντεταγμένων X και Y σε ακρίβεια φαντίου 100x100.

Τέλος, για την εύρεση των τιμών που αντιστοιχούν σε κάθε σταθμό, σε κάθε περίπτωση μεταβλητής, αυτές εξάγονται από τα αντίστοιχα επίπεδα μέσω της εντολής Extract Multi Values to Points, της εργαλειοθήκης Spatial Analyst Tools.

3.2.4 Συμπλήρωση χρονοσειρών με την μέθοδο Kriging

Μετά την διαμόρφωση των χρονοσειρών για κάθε σταθμό με μηνιαίο βήμα, όπως επεξηγήθηκε διεξοδικά στο προηγούμενο υποκεφάλαιο, σειρά έχει η συμπλήρωση των τιμών που απαιτούνται για την συμπλήρωση του χρονικού διαστήματος της δεκαετίας, καθώς και των ενδιάμεσων κενών μηνιαίων τιμών, λόγω έλλειψης καταγραφών.

Η συμπλήρωση των ελλειπουσών τιμών της μηνιαίας χρονοσειράς κάθε σταθμού έγινε με τη βοήθεια του λογισμικού ArcGIS, ενός συστήματος γεωγραφικών πληροφοριών, ενώ υποβοηθητικά χρησιμοποιήθηκε και η διαμόρφωση λογιστικών φύλλων σε περιβάλλον Excel.

Η μέθοδος Kriging, κρίθηκε ως η καταλληλότερη για την συμπλήρωση των συγκεκριμένων τιμών. Η ίδια αποτελεί μια γεωστατιστική τεχνική παρεμβολής, η οποία εφαρμόζεται συχνά σε προβλήματα χωρικής ανάλυσης κατακρήμνισης.(Μάρης κ.ά., 2014). Με τον όρο γεωστατιστική ορίζεται ένα σύνολο στατιστικών τεχνικών που σχετίζονται με μεταβλητές που μεταβάλλονται στο χώρο. Οι τεχνικές αυτές βασίζονται στην υπόθεση ότι η χωρική διακύμανση της μεταβλητής είναι τυχαία. Η μέθοδος ονομάστηκε Kriging από τον G. Matheron, που χρησιμοποίησε το όνομα του Νοτιοαφρικανού μηχανικού μεταλλείων D. G. Krige, ο οποίος είχε αναπτύξει στη δεκαετία του '50 μεθόδους για την εκτίμηση της κατανομής μεταλλευμάτων, βασιζόμενος στις τιμές των δειγμάτων.

Η μέθοδος Kriging απαιτεί την ανάπτυξη ενός στατιστικού μοντέλου, που να περιγράφει τη διακύμανση, σαν λειτουργία διαχωρισμού της απόστασης. Μόλις αναπτυχθεί το βασικό στατιστικό μοντέλο, η επιφάνεια δημιουργείται με τη στάθμιση των γειτονικών υψομέτρων, σύμφωνα με την απόσταση διαχωρισμού. Αυτή είναι ίσως ο πιο κρίσιμος και δύσκολος στόχος, δεδομένου ότι η επιλογή είναι κάπως αυθαίρετη. Ένα από τα σημαντικότερα πλεονεκτήματα των μεθόδων γεωχωρικής ανάλυσης, όπως η μέθοδος Kriging (Oliver and Webster, 1990) είναι η δυνατότητά τους να ποσοτικοποιήσουν την αβεβαιότητα στις παραγόμενες εκτιμήσεις. Σε αυτές

τις μεθόδους, οι στατιστικές υποθέσεις γίνονται για να προσδιορίσουν και να αξιολογήσουν την πολυδιάστατη χωρική δομή των υδρολογικών διαδικασιών. Η Kriging είναι μια βέλτιστη χωρική μέθοδος εκτίμησης, βασισμένη σε ένα μοντέλο της εξαρτημένης χωρικά, διαφοράς. Το μοντέλο της χωρικής ανάλυσης καλείται βαριόγραμμα (variogram) ή εναλλακτικά χρησιμοποιείται ο όρος, ημι-βαριόγραμμα (semi-variogram).

Η παρεμβολή επιφάνειας με την μέθοδο Kriging στηρίζεται στην επιλογή ενός μοντέλου βαριογράμματος. Οι ανακρίβειες λόγω των δεδομένων με «θόρυβο» ή λόγω της ανισοτροπίας προκαλούν λάθη στην παρεμβαλλόμενη επιφάνεια, που μπορεί να μην είναι προφανή ή να μην μπορούν να προσδιοριστούν εύκολα παρά μόνο μέσω της επιβεβαίωσης με δεδομένα που προέκυψαν από τον αλγόριθμο δημιουργίας μιας επιφάνειας. Η αβεβαιότητα στις παραμέτρους του βαριογράμματος, θα μπορούσε να οδηγήσει σε τρεις σημαντικές ασυνέπειες:

- Συστηματικό σφάλμα (απόκλιση) στην εκτίμηση του σημείου της πολυδιάστατης μεταβλητής
- Διαφορετική χωρική κατανομή του μέτρου της αβεβαιότητας
- Λάθος επιλογή του κατάλληλου μοντέλου βαριογράμματος

Στην συνέχεια παρουσιάζονται συνοπτικά οι διάφορες παραλλαγές της μεθόδου (Χριστοφίδης, 1994):

- **Ordinary – simple Kriging.** Η πλέον διαδεδομένη, έχει τις παρακάτω παραδοχές:
 - α) η μεταβλητή ακολουθεί κανονική κατανομή,
 - β) η εκτίμηση είναι αμερόληπτη,
 - γ) μονιμότητα δευτέρου βαθμού,
 - δ1) ο τοπικός μέσος όρος είναι γνωστός (simple), ή
 - δ2) ο τοπικός μέσος είναι άγνωστος (ordinary)
- **Neighbourhood Kriging.** Αν και η τοπική μέση τιμή και διασπορά είναι σταθερές σε όλη την περιοχή, στις περισσότερες εφαρμογές τα δεδομένα περιέχουν τοπικές διακυμάνσεις. Για το λόγο αυτό στην εκτίμηση της άγνωστης τιμής συμμετέχουν τα κοντινότερα σημεία ή αυτά που περιλαμβάνονται στη γύρω περιοχή.
- **Block Kriging.** Αντιμετωπίζει την ολοκλήρωση των εκτιμημένων τιμών σε μεγαλύτερες περιοχές
- **Universal Kriging.** Εφαρμόζεται στην περίπτωση που τα δεδομένα περιέχουν τάση (trend)
- **Disjunctive Kriging** Υπολογίζει για κάθε εκτίμηση και την πιθανότητα η αληθινή τιμή να υπερβαίνει ένα συγκεκριμένο κατώφλι.
- **Cokriging.** Η εκτίμηση με το κανονικό Kriging βελτιώνεται σημαντικά όταν η μεταβλητή που εξετάζεται συνδέεται με κάποια άλλη μεταβλητή για την οποία υπάρχουν μετρήσεις
- **Space time Kriging** Σχετίζεται με την εισαγωγή της χρονικής διάστασης των δεδομένων.

Η μέθοδος παρεμβολής Kriging περιλαμβάνει πέντε μαθηματικά μοντέλα για την μοντελοποίηση των εμπειρικών semi-variograms (Μάρης κ.ά., 2014):

- Gaussian
- Γραμμικός

- Σφαιρικός
- Εγκύκλιος
- Εκθετικός

Η μέθοδος Kriging είναι παρόμοια με την μέθοδο IDW (Inverse Distance Weighting – Μέθοδος Αντιστρόφων Αποστάσεων), καθώς σταθμίζει τις περιβάλλουσες καταγεγραμμένες τιμές με στόχο την πρόβλεψη μίας μη καταγεγραμμένης τιμής σε συγκεκριμένη θέση χωρικά.

Η γενική εξίσωση που περιγράφει και τις δύο μεθόδους παρεμβολής διαμορφώνεται ως το σταθμισμένο άθροισμα των διαθέσιμων δεδομένων:

$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i Z(s_i) \quad (3.1)$$

όπου:

$Z(s_i)$: η καταγεγραμμένη τιμή για την θέση i .

λ_i : το άγνωστο βάρος της τιμής στη θέση i .

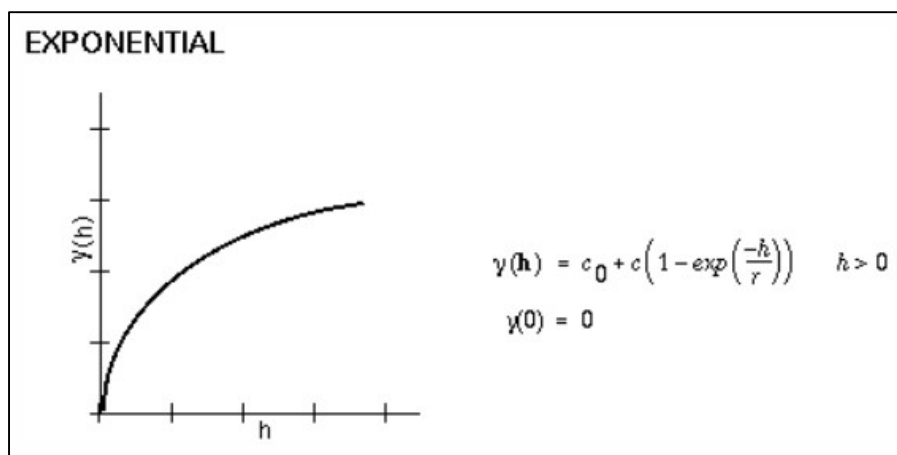
s_0 : η θέση, όπου ζητείται η πρόβλεψη

N : ο αριθμός των καταγεγραμμένων τιμών

Στην μέθοδο IDW, το βάρος λ_i , εξαρτάται μόνο από την απόσταση της θέσης όπου ζητείται η πρόβλεψη. Ωστόσο, σύμφωνα με την μέθοδο Kriging, για τον υπολογισμό των βαρών δεν απαιτείται μόνο η απόσταση μεταξύ των σημείων με τις δεδομένες τιμές και της θέσης πρόβλεψης, αλλά εξαρτάται και από την συνολική χωρική διάταξη των δεδομένων σημείων. Για να χρησιμοποιηθεί η χωρική διάταξη στα βάρη, πρέπει να ποσοτικοποιηθεί η χωρική αυτοσυσχέτιση. Έτσι, στη συνηθέστερη παραλλαγή Kriging, αυτή του ordinary, το βάρος, λ_i , εξαρτάται από ένα προσαρμοσμένο μοντέλο, σύμφωνα με τα δεδομένα σημεία, με μεταβλητές την απόσταση από τη θέση πρόβλεψης και τις χωρικές σχέσεις μεταξύ των μετρούμενων τιμών γύρω από τη θέση πρόβλεψης. (ESRI - Kriging in Geostatistical Analyst⁴)

Στην παρούσα διπλωματική εργασία χρησιμοποιήθηκε η παραλλαγή Ordinary Kriging και επιλέχθηκε το εκθετικό μοντέλο ημι-βαριογράμματος (exponential semi-variogram model – βλ. Σχήμα 3-5), βασιζόμενοι στην επιστημονική δημοσίευση των Earl and Dixon (2007).

⁴ <https://desktop.arcgis.com/>



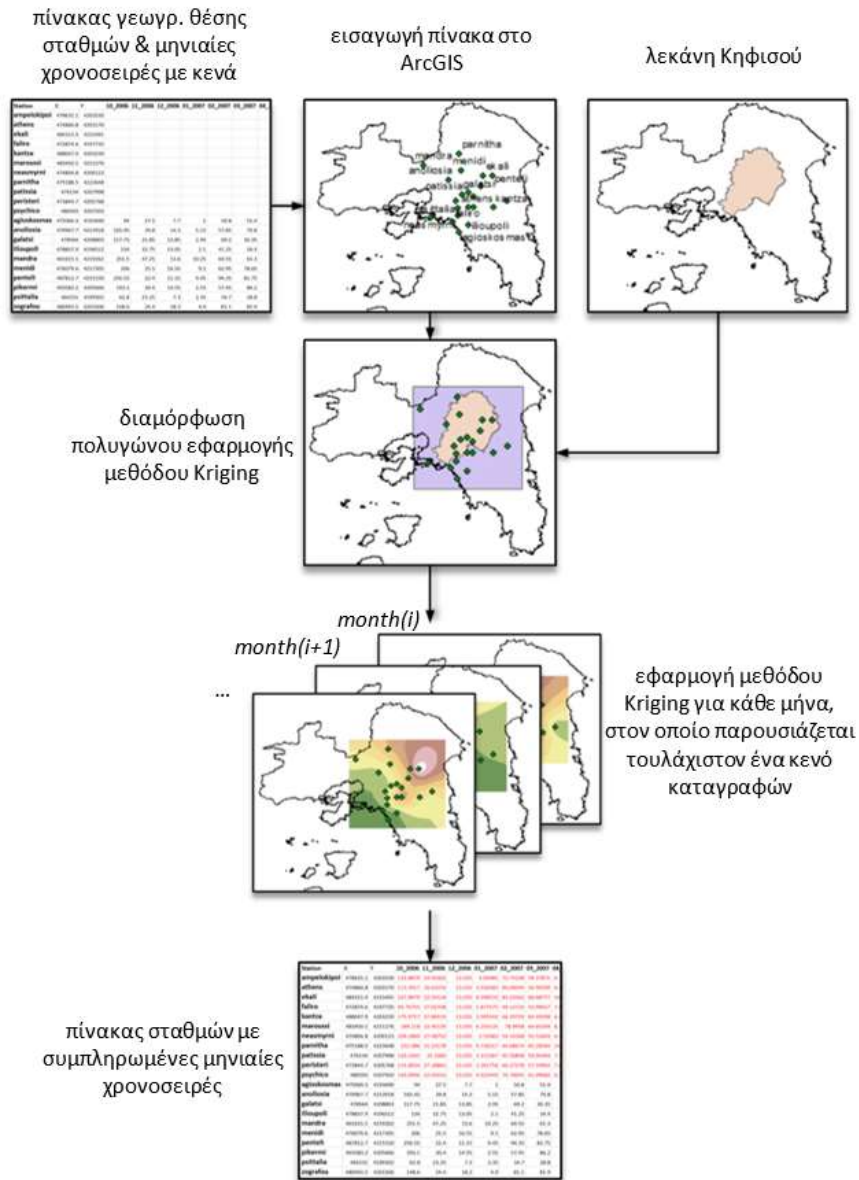
Σχήμα 3-5 Απεικόνιση μοντέλου εκθετικής ημιμεταβλητότητας (Exponential semivariance model) (ESRI – How Kriging works⁵)

Για την εφαρμογή της συγκεκριμένης μεθόδου, αρχικά εισάγονται τα μηνιαία βροχομετρικά δεδομένα για κάθε σταθμό, από λογιστικό φύλλο του excel, όπως αυτά έχουν διαμορφωθεί από την επεξεργασία των δεδομένων που περιγράφηκε προηγουμένως. Στον ίδιο πίνακα ανά σταθμό περιλαμβάνονται και οι γεωγραφικές συντεταγμένες αυτών X,Y σύμφωνα με το γεωγραφικό σύστημα ΕΓΣΑ '87, για την τοποθέτηση αυτών ως σημείων στο χάρτη. Στη συνέχεια ορίζεται το γεωχωρικό επίπεδο, εντός του οποίου θα εφαρμοστεί η συμπλήρωση κατά τη μέθοδο Kriging, το οποίο είναι ένα υποτήμημα του λεκανοπεδίου της Αττικής, εντός του οποίου περιλαμβάνονται οι εξεταζόμενοι σταθμοί και η λεκάνη απορροής του Κηφισού (βλ.2.2.1). Έπειτα από τα πακέτα εργαλείων του ArcGIS επιλέγεται το Spatial Analyst Tools και πιο συγκεκριμένα από τις επιλογές στην λίστα εργαλείων Interpolation, η μέθοδος Kriging.

Ως δεδομένα εισόδου επιλέγεται το θεματικό επίπεδο που εμπεριέχει την πληροφορία του αρχικών εισαχθέντος πίνακα σταθμών και βροχομετρικών δεδομένων, ενώ στην κατηγορία των Z-values, επιλέγεται ο εκάστοτε μήνας (στήλη του αρχικού πίνακα), ο οποίος παρουσιάζει τουλάχιστον ένα κενό ως προς τις καταγραφές. Επιλέγεται η παραλλαγή Ordinary Kriging με εκθετικό μοντέλο ημι-βαριογράμματος (Exponential semi-variogram model). Ύστερα ορίζεται η επιφάνεια εντός της οποίας θα εφαρμοστεί η μέθοδος Kriging καθώς και το μέγεθος του χωρικού κελιού, με διαστάσεις 100x100, εφόσον και τα υπόλοιπα διαθέσιμα γεωχωρικά επίπεδα έχουν την ίδια ακρίβεια. Η μέθοδος Kriging συμπληρώνει την τιμή του σταθμού που παρουσιάζεται ως κενή στον πίνακα, μέσω των καταγεγραμμένων τιμών που αντιστοιχούν στους υπόλοιπους σταθμούς, σύμφωνα με το θεωρητικό υπόβαθρο που παρατέθηκε προηγουμένως. Η εφαρμογή του συγκεκριμένου εργαλείου εντός του περιβάλλοντος ArcGIS επαναλαμβάνεται για όλους τους μήνες, εντός των οποίων καταγράφεται τουλάχιστον ένα κενό, για το επιλεγθέν χρονικό διάστημα, δηλαδή συνολικά 87 φορές. Έπειτα οι νέες τιμές εξάγονται, συνολικά από όλα τα επίπεδα που προέκυψαν από την εφαρμογή του πακέτου Kriging, στο αρχικό θεματικό επίπεδο με τα στοιχεία των σταθμών, οι πληροφορίες του οποίου τελικώς μεταφέρονται, σε έναν

⁵ <https://desktop.arcgis.com/>

συνολικά συμπληρωμένο πλέον πίνακα σε λογιστικό φύλλο excel. Η συνολικά περιγραφόμενη μεθοδολογία αποτυπώνεται στο Σχήμα 3-6.



Σχήμα 3-6 Διαδικασία συμπλήρωσης κενών στις μηνιαίες χρονοσειρές των σταθμών, μέσω μεθόδου Kriging

3.2.5 Ανάπτυξη μοντέλων παλινδρόμησης

Εφόσον οι μηνιαίες χρονοσειρές είναι πλέον πλήρεις και συμπληρωμένες για κάθε σταθμό, για το προκαθορισμένο χρονικό διάστημα, ενώ παράλληλα έχουν επιλεγεί και υπολογιστεί για κάθε σταθμό, οι ανεξάρτητες μεταβλητές προς εξέταση για την εισαγωγή τους στην τελική σχέση του μοντέλου, ακολουθεί η διαδικασία για τον τελικό καθορισμό της μορφής των μοντέλων απόκρισης μηνιαίας βροχόπτωσης.

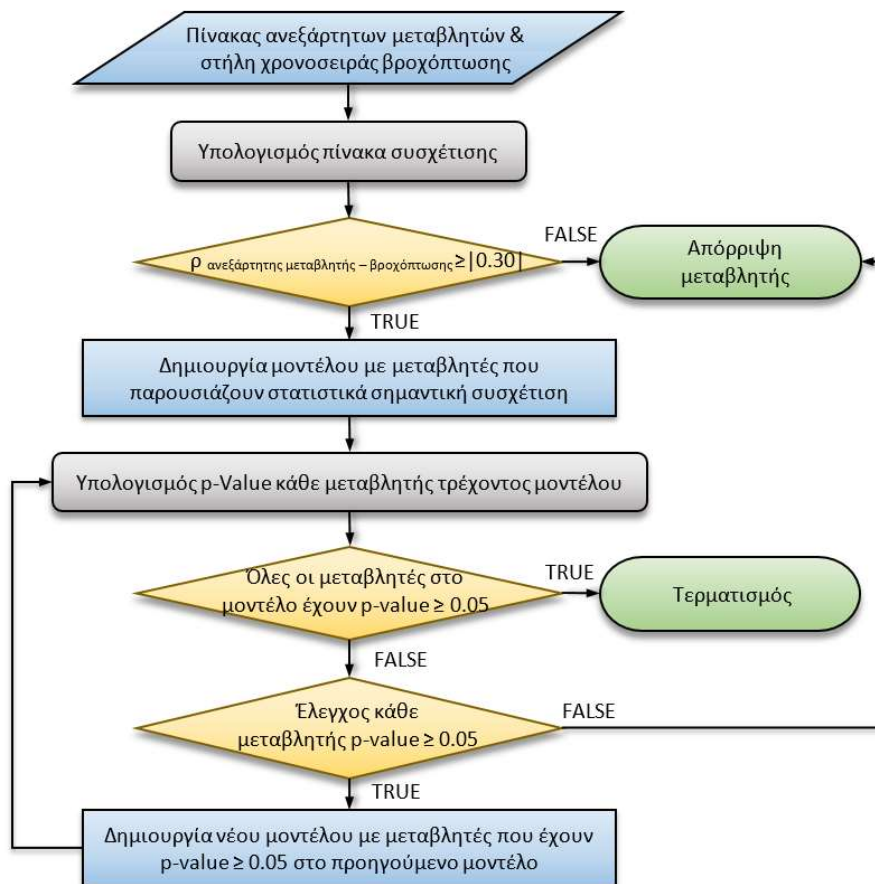
Όπως αναλύεται και στο υποκεφάλαιο που αφορά την επεξεργασία των πρωτογενών δεδομένων, ο υψηλός αριθμός ενδιάμεσων κενών που διαπιστώθηκε στις χρονοσειρές δεκάλεπτου βήματος, οδήγησε στην επιλογή μηνιαίας χρονικής κλίμακας απόκρισης για τα μελετώμενα μοντέλα. Σε ανάλογες περιπτώσεις στη διεθνή επιστημονική βιβλιογραφία εξάγεται ένα διαφορετικό μοντέλο ανά περίπτωση μήνα, λόγω της εποχικότητας των δεδομένων. Σε συγκεκριμένες περιπτώσεις μηνιαίων βροχομετρικών δεδομένων και με στόχο την ανάπτυξη ενός συνολικού μοντέλου που θα λαμβάνει υπόψη την περιοδικότητα των μηνών, έχουν αναπτυχθεί σχέσεις που περιλαμβάνουν ημιτονοειδείς συναρτήσεις (Hession et al., 2011). Ωστόσο οι προσπάθειες που έγιναν για να ενταχθεί ένας τέτοιος όρος στην παρούσα περίπτωση δεδομένων είχαν ως αποτέλεσμα στατιστικούς δείκτες που υποδείκνυαν κακή προσαρμογή του μοντέλου στα δεδομένα. Επομένως στόχος είναι η ανάπτυξη 12 διαφορετικών μοντέλων, για κάθε μήνα του έτους, καθώς επίσης ένα μοντέλο με ετήσια χρονική κλίμακα και ακόμη δύο που αναφέρονται στην υγρή και ξηρή περίοδο του έτους, συνολικά 15 μοντέλα.

Για το ετήσιο μοντέλο, τα τελικά δεδομένα που εξάχθηκαν μέσα από την μεθοδολογία που περιγράφηκε στο προηγούμενο υποκεφάλαιο, συγχωνεύθηκαν με αποτέλεσμα μια χρονοσειρά 10 τιμών για κάθε σταθμό, με κάθε τιμή να αποτυπώνει την βροχόπτωση του εκάστοτε έτους. Στη συνέχεια για κάθε σταθμό εξάχθηκε η μέση τιμή των 10 ετών, οδηγώντας έτσι σε 21 τιμές (21 σταθμοί) με μέση ετήσια βροχόπτωση δεκαετίας. Αντίστοιχα για τα μοντέλα που αναφέρονται στις υγρές και ξηρές περιόδους του έτους, τα μηνιαία δεδομένα ομαδοποιήθηκαν και αθροίστηκαν σύμφωνα με τους μήνες που αντιστοιχούν στην κάθε περίοδο. Από τα στοιχεία των ομβροθερμικών διαγραμμάτων που παρατίθενται στο υποκεφάλαιο 2.2.2, καθώς και από την γενικότερη επιστημονική βιβλιογραφία, σχετική με το κλίμα στην περιοχή της Αττικής, η ξηρή περίοδος του έτους διαρκεί από Απρίλιο έως και Σεπτέμβρη, ενώ η υγρή περίοδος από τον Οκτώβρη έως και τον Μάρτιο του επόμενου έτους. Συνεπώς, για την ξηρή περίοδο προέκυψαν με την ίδια διαδικασία 21 τιμές με την μέση συνολική βροχόπτωση των ξηρών μηνών της δεκαετίας, από το 2007 έως το 2016, ενώ για την υγρή προέκυψαν οι τιμές για την μέση συνολική βροχόπτωση ανά σταθμό, των υγρών μηνών της δεκαετίας, από την υγρή περίοδο του 2006-2007 έως και την αυτή του 2015-2016.

Η μορφή των μοντέλων, δηλαδή η επιλογή των ανεξάρτητων μεταβλητών που πρόκειται να μετέχουν στο μοντέλο κρίθηκε σκόπιμο να είναι ίδια, για κάθε περίπτωση από τις 15. Ο κύριος γνώμονας για την τελική διαμόρφωση της σχέσης είναι η καλή προσαρμογή της εκάστοτε εξεταζόμενης σχέσης στη χρονοσειρά των ετήσιων τιμών. Ωστόσο για λόγους επαλήθευσης των αποτελεσμάτων, σχέσεις οι οποίες πληρούσαν τα όρια έστω και μέρους των στατιστικά σημαντικών δεικτών, εφαρμόστηκαν εκ νέου στις μηνιαίες χρονοσειρές, προκειμένου να διεξαχθεί ένα ασφαλέστερο συμπέρασμα για την καταλληλότητα της εξεταζόμενης σχέσης συνολικότερα.

Στο επόμενο Σχήμα 3-7, παρουσιάζεται η διαδικασία που ακολουθήθηκε για την τελική διαμόρφωση του τελικού μοντέλου παλινδρόμησης, μέσω διαγράμματος ροής. Η συγκεκριμένη μεθοδολογία αποτελεί μία παραλλαγή των Stepwise Procedures που εφαρμόζονται συχνά σε ανάλογα προβλήματα και ονομάζεται Backward Elimination, λόγω του αλγορίθμου σταδιακής

απόρριψης μεταβλητών από το μοντέλο (Hession, S. L. et al., 2011 & Marquinez et al., 2003). Αναλυτικότερα, μετά την σύνταξη του συνολικού πίνακα τιμών ανεξάρτητων μεταβλητών και αυτών της εξαρτημένης μεταβλητής βροχόπτωσης υπολογίζεται ο πίνακας συσχετίσεων. Μέσω αυτού του πίνακα ελέγχονται οι ανεξάρτητες μεταβλητές που φέρονται να έχουν κάποια συσχέτιση με το εξαρτώμενο μέγεθος της βροχόπτωσης. Εάν η εκάστοτε τιμή συσχέτισης ανεξάρτητης και εξαρτημένης μεταβλητής ισούται ή υπερβαίνει την τιμή 0.30 ($\rho \geq 0.30$)⁶, τότε η συγκεκριμένη μεταβλητή θα μετέχει στο μοντέλο παλινδρόμησης που θα σχηματιστεί στη συνέχεια, αλλιώς απορρίπτεται ως μη συσχετιζόμενη. Αμέσως μετά διαμορφώνεται το μοντέλο που περιλαμβάνει τις παραπάνω μεταβλητές και υπολογίζονται οι αντίστοιχες τιμές p-value, για κάθε μια από αυτές. Οι μεταβλητές για τις οποίες, η p-value τιμή υπερβαίνει το 0.05⁷, εξαλείφονται από το μοντέλο και δημιουργείται νέο με τις εναπομένουσες ανεξάρτητες μεταβλητές. Η ίδια διαδικασία ακολουθείται έως ότου το τελικώς διαμορφωμένο μοντέλο εμπεριέχει μεταβλητές για τις οποίες το p-value είναι ίσο ή βρίσκεται κάτω από το όριο του 0.05.



Σχήμα 3-7 Διάγραμμα ροής μεθοδολογίας επιλογής συμμετοχής ανεξάρτητων μεταβλητών στο μοντέλο

⁶ Το παραπάνω αποτελεί ένα συχνό κριτήριο

⁷ Εφόσον επιλέχθηκε διάστημα εμπιστοσύνης 95%

Από τις διάφορες μορφές μοντέλων που εξετάστηκαν δεν ήταν εφικτό να επιτευχθεί το κριτήριο του ορίου p-value για όλες τις μεταβλητές κάθε μοντέλου από τα 15. Συνεπώς η τελική σχέση (οι ανεξάρτητες μεταβλητές που συμμετέχουν) που επιλέχθηκε είναι εκείνη που έδωσε τα βέλτιστα αποτελέσματα συγκριτικά με τις υπόλοιπες εξεταζόμενες σχέσεις, ως προς τα κριτήρια που τέθηκαν και αναλύθηκαν προηγουμένως.

Μετά την οριστικοποίηση των 15 μοντέλων παλινδρόμησης, τα θεματικά επίπεδα της προβλεπόμενης βροχόπτωσης αποτυπώθηκαν χωρικά για κάθε περίπτωση μοντέλου, μέσω της εργαλειοθήκης του Raster Calculator σε περιβάλλον ArcGIS. Από τα παραγόμενα αυτά επίπεδα, εξάγονται οι τιμές βροχόπτωσης που αντιστοιχούν στο φατνίο, εντός του οποίου βρίσκεται κάθε ένας από τους εξεταζόμενους σταθμούς. Το σύνολο αυτών των τιμών αποτελούν τις εκτιμώμενες – προβλεπόμενες βροχοπτώσεις κάθε μοντέλου, οι οποίες σε επόμενο στάδιο θα συγκριθούν με τις πραγματικές – παρατηρούμενες, με στόχο την αξιολόγηση της ορθότητας κάθε μοντέλου, μέσω συγκεκριμένων στατιστικών δεικτών και συγκριτικών διαγραμμάτων.

3.2.6 Καθορισμός κριτηρίων αξιολόγησης των μοντέλων

Τόσο κατά τα στάδια ανάπτυξης των μοντέλων παλινδρόμησης, όσο και κατά το στάδιο της μετέπειτα αξιολόγησης αυτών για την τελική εξαγωγή καθολικών συμπερασμάτων υπολογίστηκαν μία σειρά από στατιστικούς δείκτες, ενώ παράλληλα για τον ίδιο σκοπό διαμορφώθηκαν κατάλληλα διαγράμματα. Τα παραπάνω κριτήρια (στατιστικοί δείκτες και διαγράμματα) αναλύονται εννοιολογικά στη συνέχεια.

Απλός και Προσαρμοσμένος Συντελεστής Προσδιορισμού (R^2 , R_{adj}^2)

Στα μοντέλα παλινδρόμησης, υπάρχει ένας ορισμένος βαθμός επιπέδου συσχέτισης μεταξύ των ανεξάρτητων και εξαρτημένων μεταβλητών στο σύνολο δεδομένων που επιτρέπουν την πρόβλεψη της εξαρτημένης μεταβλητής. Στη στατιστική ανάλυση, αυτή η συσχέτιση μπορεί να υπολογισθεί χρησιμοποιώντας τον συντελεστή προσδιορισμού (R^2) και του προσαρμοσμένου συντελεστή προσδιορισμού (R_{adj}^2). Μέσω των συγκεκριμένων στατιστικών δεικτών μπορεί να προσδιοριστεί το πόσο μεγάλο μέρος της διακύμανσης στην τιμή μιας εξαρτημένης μεταβλητής (y) εξηγείται από τις τιμές των ανεξάρτητων μεταβλητών ($X_1, X_2, X_3 \dots$).

Ο συντελεστής προσδιορισμού (R^2) χρησιμοποιείται για τον προσδιορισμό – ποσοτικοποίησης της συσχέτισης μεταξύ των ανεξάρτητων μεταβλητών (προγνωστικών) και του στόχου. Με άλλα λόγια, αποτελεί έναν δείκτη που εκφράζει την αρτιότητα ενός μοντέλου παλινδρόμησης σε σύγκριση με τον μέσο όρο.

Ειδικότερα για τον συντελεστή προσδιορισμού ισχύει:

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (3.2)$$

όπου,

SSR: Άθροισμα τετραγώνων υπολοίπων (Sum Squared Regression), είναι το άθροισμα των τετραγώνων της διαφοράς μεταξύ της πραγματικής παρατηρούμενης τιμής (y) και της προβλεπόμενης τιμής (\hat{y}).

SST: Συνολικό άθροισμα τετραγώνων (Total Sum of Squares), είναι το άθροισμα των τετραγώνων της διαφοράς μεταξύ της πραγματικής παρατηρούμενης τιμής (y) και του μέσου όρου των παρατηρούμενων τιμών (\bar{y}).

Πέρα από τον κλασικό συντελεστή προσδιορισμού, συχνά χρησιμοποιείται και ο προσαρμοσμένος συντελεστής προσδιορισμού (R_{adj}^2) με στόχο την επιλογή του καταλληλότερου μοντέλου και περιγράφεται από την επόμενη σχέση:

$$R_{adj}^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1} \quad (3.3)$$

όπου,

n : το πλήθος του δείγματος

p : ο αριθμός των ανεξάρτητων μεταβλητών του μοντέλου παλινδρόμησης

R^2 : ο συντελεστής προσδιορισμός του μοντέλου παλινδρόμησης

Σύμφωνα με την σχέση του συντελεστή προσδιορισμού, η εξίσωση πλέον που περιγράφει τον προσαρμοσμένο συντελεστή προσδιορισμού διαμορφώνεται ως εξής:

$$R_{adj}^2 = 1 - \frac{SSR/(n - p)}{SST/(n - 1)} \quad (3.4)$$

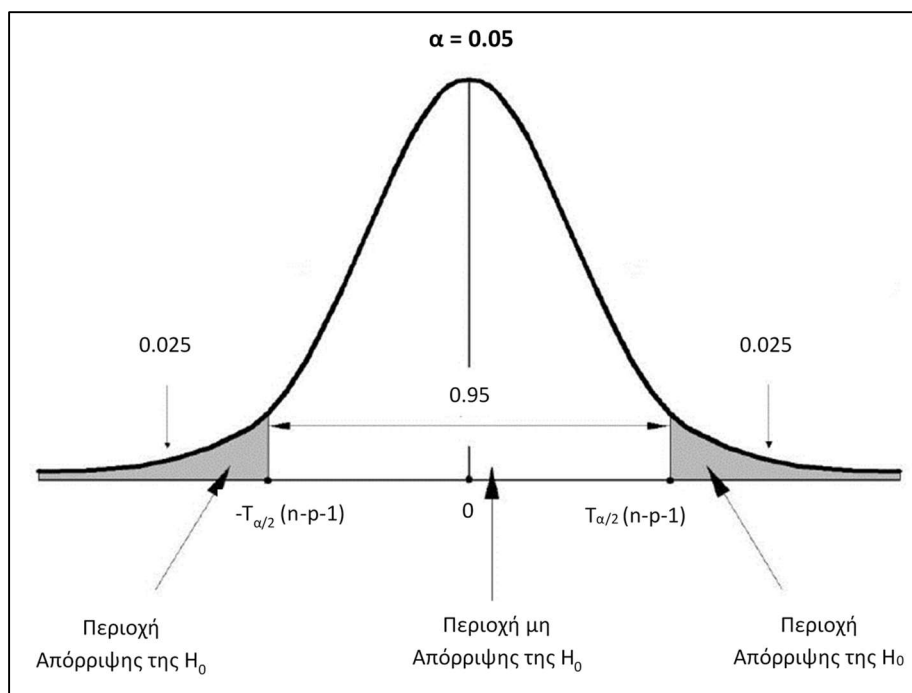
Συγκρίνοντας τους δύο αυτούς δείκτες διαπιστώνεται ότι, ο στατιστικός δείκτης του προσαρμοσμένου συντελεστή προσδιορισμού (R_{adj}^2) είναι μια τροποποιημένη εκδοχή του κλασικού συντελεστή προσδιορισμού (R^2), ο οποίος λαμβάνει υπόψη του τον αριθμό των ανεξάρτητων μεταβλητών πρόβλεψης, που συμμετέχουν στο μοντέλο παλινδρόμησης. Ο προσαρμοσμένος συντελεστής (R_{adj}^2) αυξάνεται όταν η νέα εισερχόμενη στο μοντέλο μεταβλητή, βελτιώνει το μοντέλο σε μεγαλύτερο βαθμό απ' ότι θα μπορούσε αυτό να συμβεί τυχαία και αντίστροφα. Γενικά οι τιμές που λαμβάνει ο προσαρμοσμένος συντελεστής προσδιορισμού είναι συνηθέστερα μικρότερες από αυτές του κλασικού συντελεστή προσδιορισμού.

Ειδικότερα, ο προσαρμοσμένος συντελεστής προσδιορισμού (R_{adj}^2) αποδίδει μεγαλύτερη ποινή (penalty) σε μοντέλα τα οποία έχουν μεγάλο αριθμό παραμέτρων, με άλλα λόγια ως συγκεκριμένος συντελεστής είναι περισσότερο «φειδωλός» σε πολυπαραμετρικά μοντέλα και κατά συνέπεια ένας πιο έμπιστος δείκτης για την επιλογή μεταξύ μοντέλων με διαφορετικό αριθμό παραμέτρων.

Παρατηρούμενο επίπεδο σημαντικότητας (p -value)

Για την ένταξη των ανεξάρτητων μεταβλητών στο μοντέλο πολλαπλής γραμμικής παλινδρόμησης, πραγματοποιείται έλεγχος υποθέσεων. Ο έλεγχος υποθέσεων αποτελείται από τη **μηδενική υπόθεση** (null hypothesis) και την **εναλλακτική υπόθεση** (alternative hypothesis). Η μηδενική υπόθεση είναι η υπόθεση προς εξέταση. Με τη διαδικασία ελέγχου υποθέσεων διαπιστώνεται εάν η μηδενική υπόθεση μπορεί να απορριφθεί ή όχι. Η μηδενική υπόθεση απορρίπτεται αν τα δεδομένα του δείγματος δείξουν ότι έχει πολύ μικρή πιθανότητα να είναι αληθινή. Στην παρούσα εργασία ως μηδενική υπόθεση θεωρείται η ένταξη ή μη ένταξη μίας ανεξάρτητης μεταβλητής στο μοντέλο πολλαπλής γραμμικής παλινδρόμησης και συνεπώς ο έλεγχος υποθέσεων αφορά στην απόρριψη ή μη απόρριψη της υπόθεσης αυτής.

Για να απορριφθεί η μηδενική υπόθεση πρέπει να προκύψει τιμή για έναν στατιστικό έλεγχο η οποία να είναι ακραία, δηλαδή να έχει πολύ μικρή πιθανότητα εμφάνισης. Η πιθανότητα να προκύψει μια ακραία τιμή σύμφωνα με τη μηδενική υπόθεση συμβολίζεται με το λατινικό γράμμα p (p -value) και ονομάζεται παρατηρούμενο επίπεδο σημαντικότητας. Η τιμή του p συγκρίνεται κάθε φορά με το προκαθορισμένο επίπεδο σημαντικότητας α και αν είναι μικρότερη ή ίση του α η μηδενική υπόθεση H_0 απορρίπτεται. Στα πλαίσια της παρούσας διπλωματικής εργασίας το προκαθορισμένο επίπεδο σημαντικότητας επιλέγεται το $\alpha=0.05$, τότε για $p<0.05$, απορρίπτεται η μηδενική υπόθεση. Η τιμή p είναι το μικρότερο επίπεδο σημαντικότητας στο οποίο η αρχική υπόθεση μπορεί να απορριφθεί. Τα αποτελέσματα του δείγματος που υποδεικνύουν ότι μια μηδενική υπόθεση πρέπει να απορριφθεί σε ένα επίπεδο σημαντικότητας α ονομάζονται στατιστικά σημαντικά (statistically significant) αποτελέσματα και συχνά ο έλεγχος αναφέρεται ως έλεγχος σημαντικότητας. Οι έννοιες του επιπέδου σημαντικότητας και της τιμής p -value απεικονίζονται στο ακόλουθο σχήμα (Σχήμα 3-8). (Ηλιοπούλου, 2016)



Σχήμα 3-8 Περιοχές απόρριψης και μη απόρριψης της μηδενικής υπόθεσης για έλεγχο με την *t-Student* κατανομή (αμφίπλευρος έλεγχος και επίπεδο σημαντικότητας $\alpha=0,05$), όπου n το πλήθος των παρατηρήσεων και p ο αριθμός των ανεξάρτητων μεταβλητών που συμμετέχουν στο μοντέλο (Πηγή: Ηλιοπούλου, 2016)

Μέτρα Αξιολόγησης μοντέλων πρόγνωσης

Σύμφωνα με την σχετική επιστημονική βιβλιογραφία⁸, η οποία αφορά την αξιολόγηση βροχομετρικών (ή άλλων κλιματικών) δεδομένων πρόγνωσης, η ίδια μπορεί να βασιστεί σε διάφορους δείκτες επαλήθευσης ή βαθμολόγησης που βασίζονται σε τέσσερις βασικές παραμέτρους (a,b,c,d), όπως φαίνεται στον επόμενο Πίνακα.

Πίνακας 3-2 Παράμετροι αξιολόγησης

<i>2x2 Contingency Table</i>		Observed	
		Yes	No
Forecast	Yes	a	b
	No	c	d

Εξετάζοντας μια περίπτωση πρόγνωσης βροχής που είτε συμβαίνει, είτε όχι, το γεγονός αυτό είναι κατηγορηματικό, μη πιθανολογικό και διακριτό. Παρόμοια πρόγνωση μπορεί επίσης να αφορά σε προειδοποίηση κάποιου ακραίου καιρικού φαινομένου. Έχοντας ένα σύνολο

⁸ Jolliffe and Stephenson, 2003
 Wilks, 2011
 Rossa et al., 2008
 Nastos et al., 2016

δεδομένων πρόγνωσης και των αντίστοιχων επίγειων μετρήσεων, η αξιολόγηση συνοψίζεται σε έναν πίνακα συχνοτήτων 2x2, ο οποίος περιλαμβάνει τέσσερα πιθανά σενάρια:

- i. Ένα γεγονός να προγνωσθεί και να συμβεί (a)
- ii. Ένα γεγονός να προγνωσθεί και να μη συμβεί (b)
- iii. Ένα γεγονός να μη προγνωσθεί και να συμβεί (c)
- iv. Ένα γεγονός να μη προγνωσθεί και να μη συμβεί (d).

Στα πλαίσια της συγκεκριμένης μεταπτυχιακής εργασίας και λόγω της εξεταζόμενης χρονικής κλίμακας της εξαρτημένης μεταβλητής (ύψη μηνιαίας, ετήσιας ή ξηρής και υγρής περιόδου βροχόπτωσης), το σενάριο που περιγράφεται από την παράμετρο d, δεν είναι δυνατό να πραγματοποιηθεί. Ειδικότερα, δεν υπάρχουν παρατηρήσεις εντός των χρονοσειρών, στις οποίες καταγράφεται μηδενική βροχόπτωση για τις συγκεκριμένες χρονικές κλίμακες και επομένως δεν μπορούν να συγκριθούν με τις αντίστοιχες τιμές απόκρισης του μοντέλου.

Έτσι για την περίπτωση σεναρίου (i) καταμετρώνται οι περιπτώσεις για τις οποίες οι τιμές πρόβλεψης μοντέλου (forecasted) εμπεριέχονται στο διάστημα $\pm 10\%$ της καταγεγραμμένης τιμής (observed). Για περιπτώσεις όπου η τιμή πρόγνωσης βρίσκεται πάνω από το 10% της παρατηρούμενης τιμής πραγματοποιείται το σενάριο (ii), ενώ για τις τιμές που βρίσκονται κάτω από το -10% της παρατηρούμενης τιμής, θεωρείται ότι πραγματοποιείται το σενάριο (iii).

Λαμβάνοντας υπόψη τα προηγούμενα και αναφορικά με τις λοιπές παραπάνω παραμέτρους, εισάγονται διάφοροι δείκτες αξιολόγησης. Στη συνέχεια περιγράφονται εκείνοι που χρησιμοποιήθηκαν στην παρούσα εργασία.

Hit Rate (H) ή Probability of Detection (POD).

Το «ποσοστό επιτυχία» είναι το κλάσμα των παρατηρούμενων γεγονότων που προβλέπονται σωστά. Υπολογίζεται ως εξής:

$$H = POD = \frac{a}{a + c} \quad (3.5)$$

Είναι επίσης γνωστή ως Πιθανότητα Ανίχνευσης (POD). Κυμαίνεται από μηδέν (0) για την ανεπαρκή πρόγνωση μέχρι ένα (1) για την ιδανική.

False Alarm Ratio (FAR)

Ο «λόγος εσφαλμένου συναγερμού» είναι το κλάσμα των «ναι» προγνώσεων που ήταν λάθος (δηλ. βροχή βάσει πρόγνωσης που όμως δε συνέβη):

$$FAR = \frac{b}{a + b} \quad (3.6)$$

Οι τιμές του κυμαίνονται από μηδέν (0) για περιπτώσεις ιδανικής πρόβλεψης μέχρι και ένα (1) για κακές προβλέψεις.

Threat Score(TS) ή Critical Success Index (CSI)

Η «βαθμολογία κινδύνου», ή όπως είναι πιο γνωστός ο κρίσιμος δείκτης επιτυχίας συνδυάζει το ποσοστό επιτυχίας και το λόγο εσφαλμένου συναγερμού σε μια βαθμολόγηση για τα γεγονότα χαμηλής συχνότητας. Υπολογίζεται με βάση τον τύπο:

$$TS = CSI = \frac{a}{a + b + c} \quad (3.7)$$

Επίσης λαμβάνει τιμές μεταξύ [0,1] και δε λαμβάνει υπόψη την παράμετρο d («δεν προβλέπεται/δε συμβαίνει»)⁹. Ο δείκτης CSI λαμβάνει τιμές κοντά στο μηδέν για περιπτώσεις κακών προβλέψεων ενώ για τιμές κοντά στη μονάδα, οι προβλέψεις θεωρούνται ιδανικές.

Οι δείκτες CSI, POD και FAR χρησιμοποιούνται ευρέως από την Εθνική Μετεωρολογική Υπηρεσία των ΗΠΑ για την επαλήθευση προγνώσεων/προειδοποιήσεων σε περιπτώσεις ισχυρών καταιγίδων και ανεμοστρόβιλων.

Mean Absolute Error (MAE)

Το «μέσο απόλυτο σφάλμα» είναι ένα βαθμωτό μέτρο ακρίβειας, το οποίο υπολογίζεται σύμφωνα με την ακόλουθη σχέση:

$$MAE = \frac{\sum_i^n |y_i - o_i|}{n} \quad (3.8)$$

όπου,

y_i : η i-οστή τιμή πρόβλεψης (forecasted value)

o_i : η i-οστή παρατηρούμενη τιμή (observed value)

n : το πλήθος των ζευγών παρατηρούμενων – προβλεπόμενων από το μοντέλο τιμών

Mean Square Error (MSE)

Το «μέσο τετραγωνικό σφάλμα» αποτελεί επίσης ένα βαθμωτό μέτρο ακρίβειας, παρόμοιο με το μέσο απόλυτο σφάλμα, το οποίο υπολογίζεται σύμφωνα με την ακόλουθη σχέση:

$$MSE = \frac{\sum_i^n (y_i - o_i)^2}{n} \quad (3.9)$$

όπου,

⁹ Πέραν του συγκεκριμένου δείκτη (CSI), συχνά υπολογίζεται και ο δείκτης του «ποσοστού των σωστών προβλέψεων – PC (Percent Correct), όπου ισχύει $PC = \frac{a+d}{n}$. Στην παρούσα εργασία λόγω του ότι δεν υφίσταται η παράμετρος d, οι δείκτες CSI και PC, ταυτίζονται.

y_i : η *i*-οστή τιμή πρόβλεψης (forecasted value)

o_i : η *i*-οστή παρατηρούμενη τιμή (observed value)

n : το πλήθος των ζευγών παρατηρούμενων – προβλεπόμενων από το μοντέλο τιμών

Σε αυτή την περίπτωση τα σφάλματα που προκύπτουν από τα ζεύγη τιμών πρόβλεψης και παρατήρησης υψώνονται αρχικά στο τετράγωνο και στη συνέχεια υπολογίζεται ο μέσος τους. Το μέτρο MSE είναι περισσότερο ευαίσθητος σε υψηλά σφάλματα (outliers), συγκριτικά με τον δείκτη MAE.

Root Mean Square Error (RMSE)

Η τετραγωνική ρίζα του δείκτη MSE είναι το μέγεθος RMSE (Root Mean Square Error).

$$RMSE = \sqrt{\frac{\sum_i^n (y_i - o_i)^2}{n}} \quad (3.10)$$

Normalized Root Mean Square Error (NRMSE) – Normalized Mean Square Error (NMSE)

Το κανονικοποιημένο μέσο τετραγωνικό σφάλμα και η κανονικοποιημένη ρίζα του τετραγωνικού σφάλματος είναι επί της ουσίας μεγέθη που χρησιμοποιούνται με στόχο την ευκολότερη σύγκριση (των αντιστοιχών μη κανονικοποιημένων μεγεθών) σε περιπτώσεις συνόλων δεδομένων με διαφορετικές κλίμακες.

Σύμφωνα με την διαθέσιμη βιβλιογραφία παρατίθενται αρκετές σχέσεις που χρησιμοποιούν διαφορετικά στατιστικά μέτρα ως μέσο για την κανονικοποίηση του RMSE. Οι συνηθέστερες σχέσεις είναι οι εξής:

- $NRMSE = \frac{RMSE}{\bar{y}}$, όπου \bar{y} είναι η μέση τιμή των παρατηρήσεων
- $NRMSE = \frac{RMSE}{y_{max} - y_{min}}$, όπου y_{max} και y_{min} η μέγιστη και η ελάχιστη τιμή από τις διαθέσιμες παρατηρήσεις
- $NRMSE = \frac{RMSE}{\sigma}$, όπου σ η τυπική απόκλιση των παρατηρήσεων
- $NRMSE = \frac{RMSE}{Q_1 - Q_2}$, όπου ο παρονομαστής αποτελεί την διαφορά μεταξύ 25% και 75% αντίστοιχα τιμών από το εύρος του δείγματος των παρατηρήσεων.

Η τελευταία σχέση αποτελεί μία ιδανικότερη επιλογή κανονικοποίησης, στην περίπτωση όπου στο δείγμα των πραγματικών τιμών, παρατηρούνται ακραίες τιμές – outliers. (Gareth et al., 2013; Mehdizadeh, 2020).

Σε ότι αφορά τη κανονικοποίηση ως προς τον μέσο όρο, από την σχετική βιβλιογραφία τιμές μικρότερες του 0.1 υποδεικνύουν καλή προσαρμογή του μοντέλου.

Ωστόσο από την ανάλυση των Mentaschi et al. (2013), το μέγεθος NRMSE δεν είναι απολύτως αξιόπιστο για την αξιολόγηση των εκτιμήσεων ενός μοντέλου. Καθώς ο συγκεκριμένος δείκτης σχετίζεται με την μεροληψία και την διασπορά των προσομοιωμένων τιμών. Οι τελευταίοι δείκτες εμφανίζονται στατιστικά εξαρτώμενες, μάλιστα σχεδόν γραμμικά. Αυτό συμβαίνει επειδή οι θετικές ή αρνητικές συνεισφορές του μέσου όρου της προσομοίωσης συνεπάγονται γενικότερα μία αντίστοιχη ενίσχυση της διασποράς των προσομοιωμένων τιμών. Ο δείκτης NRMSE τείνει να ελαχιστοποιείται για τιμές προσομοίωσης που υποεκτιμούν τις παρατηρήσεις και έτσι χαμηλότερη τιμή του NRMSE δεν συνδέεται πάντα με το βέλτιστο μοντέλο.

Για την παράκαμψη αυτού του προβλήματος οι Hanna, S. et al. (1985), έχουν προτείνει μία άλλη μορφή κανονικοποίησης, για το μέγεθος MSE:

$$NMSE = \frac{\overline{(y_i - o_i)^2}}{\bar{y} \cdot \bar{o}} \quad (3.11)$$

Η κανονικοποίηση ως προς το γινόμενο των μέσων όρων παρατηρούμενων και εκτιμώμενων τιμών εξασφαλίζει ότι το NMSE δεν θα μεροληπτεί ως προς μοντέλα που υποεκτιμούν ή υπερεκτιμούν.

Mean Error (ME)

Το «μέσο σφάλμα» είναι ένα βαθμωτό μέτρο ακρίβειας το οποίο υπολογίζεται από την σχέση που ακολουθεί:

$$ME = \frac{\sum_i^n (y_i - o_i)}{n} \quad (3.12)$$

Ο συγκεκριμένος δείκτης είναι παρόμοιος με αυτούς του μέσου απολύτου σφάλματος (MAE) και του μέσο τετραγωνικού σφάλματος (MSE), όπου:

y_i : η i -οστή τιμή πρόβλεψης (forecasted value)

o_i : η i -οστή παρατηρούμενη τιμή (observed value)

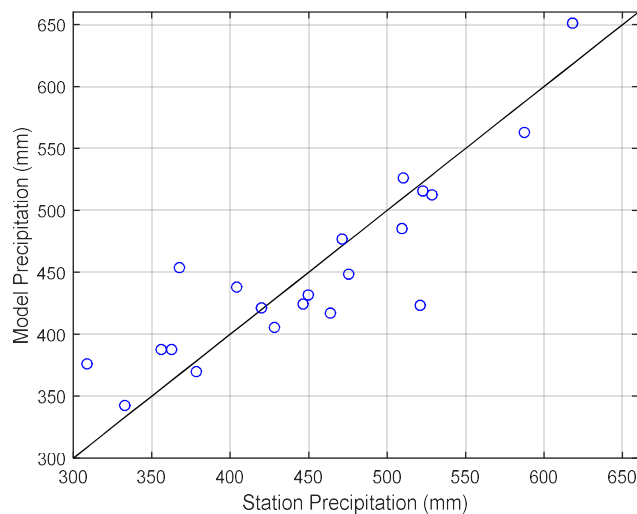
n : το πλήθος των ζευγών παρατηρούμενων – προβλεπόμενων από το μοντέλο τιμών

Το μέσο σφάλμα (ME) λαμβάνει υπόψη στον υπολογισμό του μέσου όρου, τόσο θετικά όσο και αρνητικά σε πρόσημο σφάλματα. Αυτό έχει σαν αποτέλεσμα ο συγκεκριμένος δείκτης να είναι επίσης γνωστός ως «μεροληψία» (bias).

- Εάν το $ME = 0$, δεν υπάρχει μεροληψία
- Εάν $ME > 0$, οι προβλέψεις, κατά μέσο όρο, είναι υπερεκτιμημένες
- Εάν $ME < 0$, οι προβλέψεις, κατά μέσο όρο, είναι υποεκτιμημένες

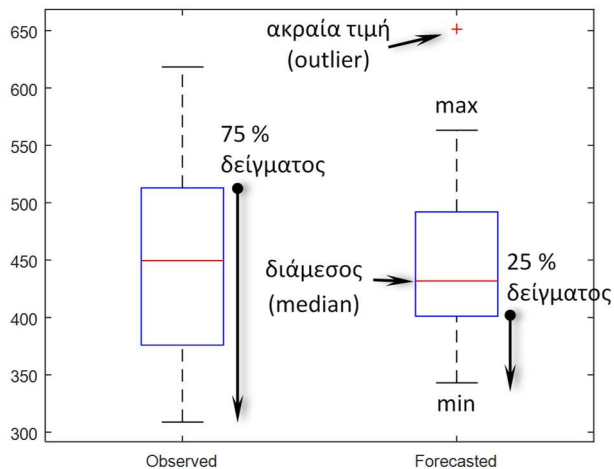
Διαγράμματα

Η επαλήθευση μοντέλων πρόβλεψης συνεχών μεταβλητών, όπως στην συγκεκριμένη περίπτωση (ύψη βροχόπτωσης), συχνά περιλαμβάνει ορισμένα διερευνητικά διαγράμματα, όπως προτείνονται (Collaboration for Australian Weather and Climate Research – Australian Government, Bureau of Meteorology). Έτσι διαμορφώνονται συγκριτικά ραβδογράμματα (bar plots) που συνοψίζουν τις τιμές των δεικτών, που αναλύθηκαν εκτενώς προηγουμένως και έλαβε κάθε μοντέλο. Πέραν αυτών και για κάθε μοντέλο σχεδιάζεται ένα διάγραμμα διασποράς (scatter plot), όπου οι παρατηρούμενες τιμές συγκρίνονται με αυτές που εκτιμήθηκαν απ' το μοντέλο (τιμές πρόβλεψης), όπως απεικονίζεται στο Σχήμα 3-9. Είναι εμφανές ότι όσο πλησιέστερα τα σημεία βρίσκονται στην διαγώνιο – ευθεία $y=x$, τόσο καλύτερη είναι και η προσαρμογή του μοντέλου παλινδρόμησης στα μοντέλα.



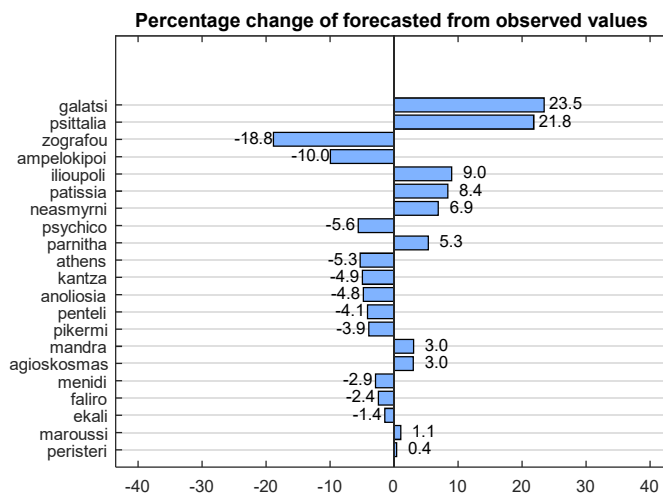
Σχήμα 3-9 Διάγραμμα διασποράς ζευγών παρατηρούμενων και εκτιμώμενων τιμών από το μοντέλο παλινδρόμησης

Ακόμη, σχεδιάζονται τα θηκογράμματα (box plots) κάθε μοντέλου, όπου για κάθε ομάδα τιμών, παρατηρούμενων και εκτιμώμενων, αντλούνται πληροφορίες σχετικά με το εύρος του δείγματος από μέγιστη έως και ελάχιστη τιμή, το 50% των παρατηρήσεων που εμπερικλείονται στο εύρος του διαστήματος 25% έως και 75% του συνόλου των τιμών, την διάμεσο του δείγματος, καθώς και τις ακραίες τιμές (outliers), όπως απεικονίζεται στο Σχήμα 3-10.



Σχήμα 3-10 Θηκογράμματα (box plots) για το δείγμα των παρατηρούμενων και εκτιμώμενων, αντίστοιχα, τιμών

Με στόχο την εξαγωγή συμπερασμάτων που αφορούν αξιολόγηση της προσαρμογής των μοντέλων σε χωρικό επίπεδο, δηλαδή ανά σταθμό, δημιουργήθηκαν ραβδογράμματα με τις ποσοστιαίες διαφορές μεταξύ εκτιμώμενων και παρατηρούμενων τιμών, όπως απεικονίζεται στο Σχήμα 3-11.



Σχήμα 3-11 Ραβδόγραμμα ποσοστιαίας διαφοράς εκτιμώμενης και παρατηρούμενης τιμής ανά σταθμό

Οι στατιστικοί δείκτες που αναλύθηκαν σε συνδυασμό με τα προηγούμενα διαγράμματα συνεκτιμώνται για το σύνολο των μοντέλων και των σταθμών με στόχο την εξαγωγή συμπερασμάτων.

4 ΑΠΟΤΕΛΕΣΜΑΤΑ

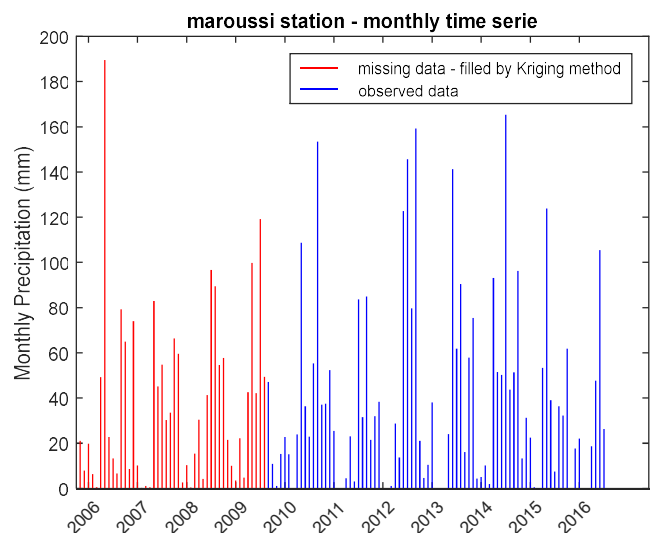
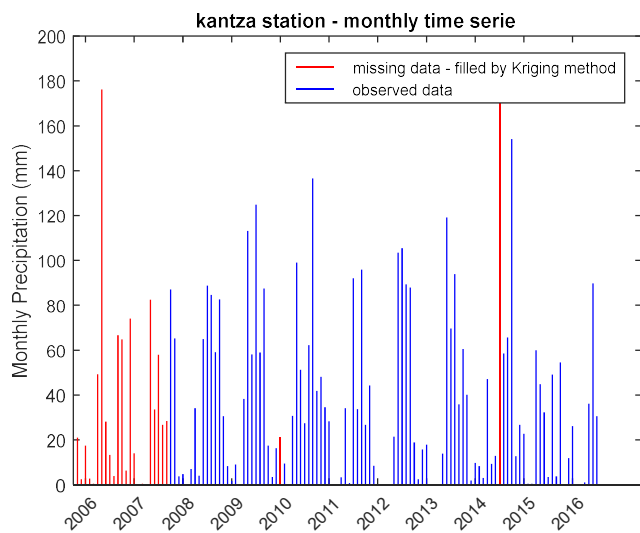
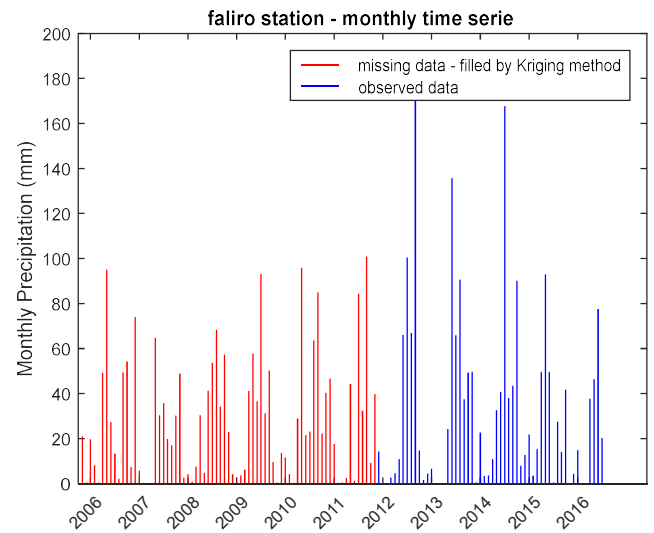
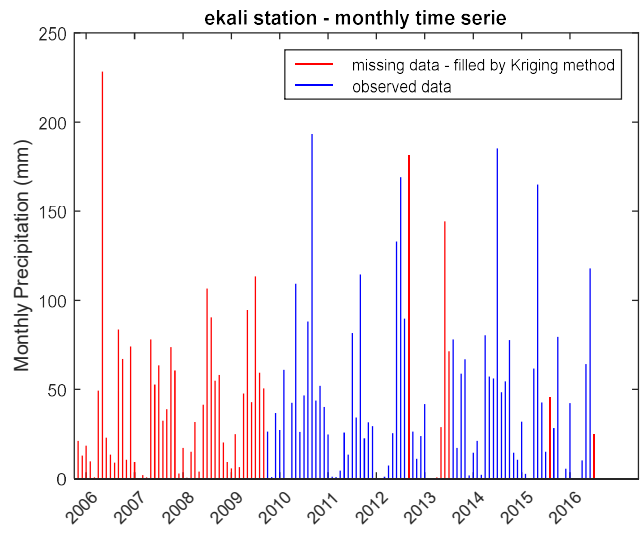
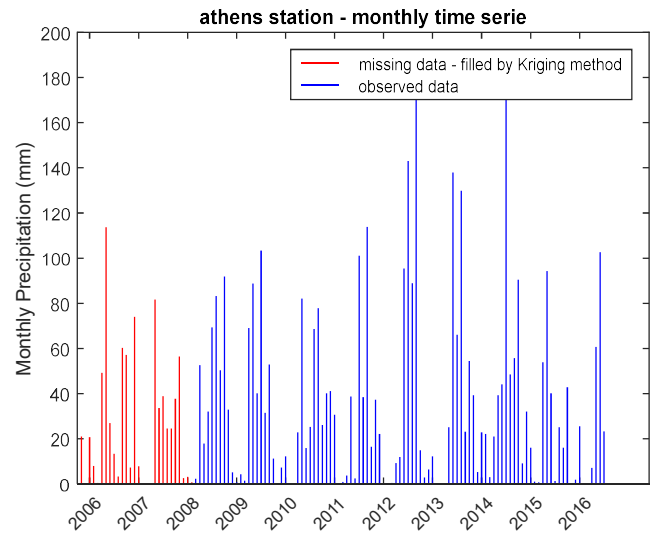
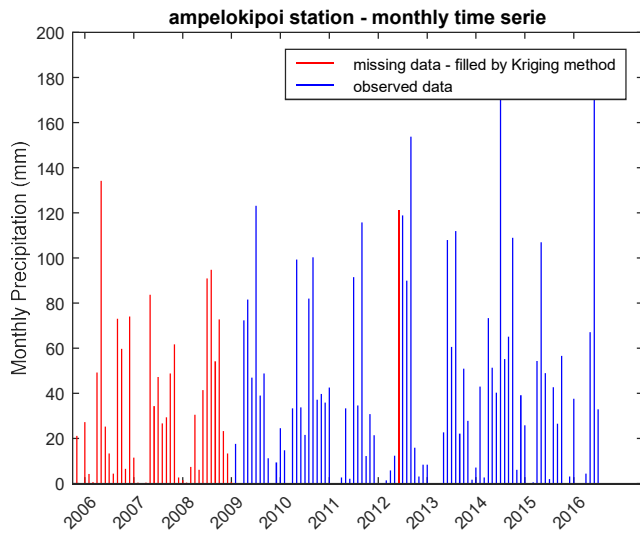
Στο παρόν κεφάλαιο παρατίθενται και εξηγούνται το σύνολο των αποτελεσμάτων που εξάχθηκαν στα πλαίσια αυτής της διπλωματικής εργασίας.

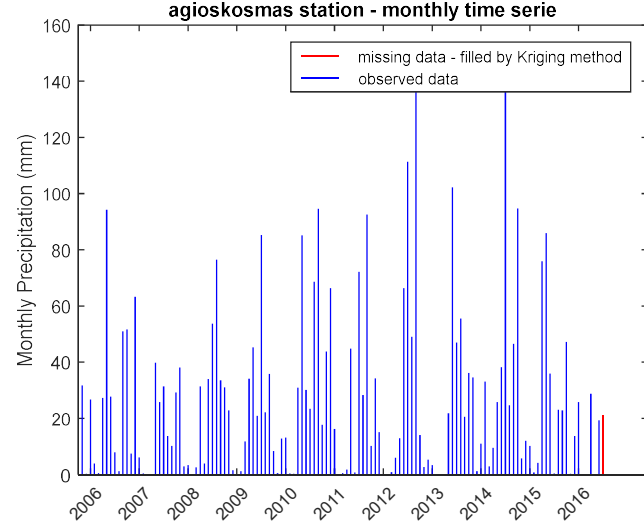
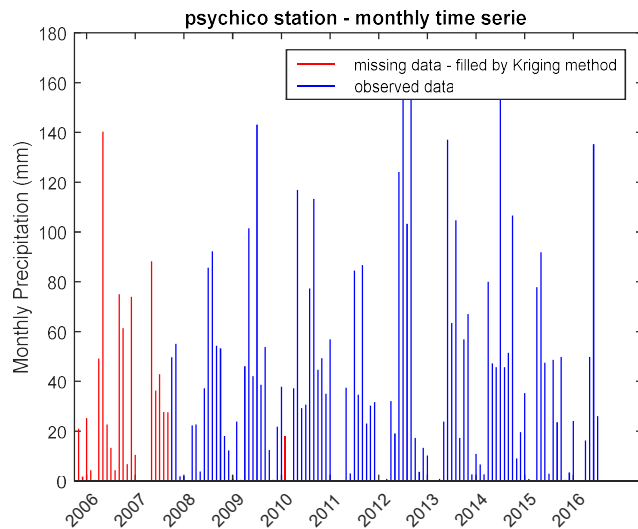
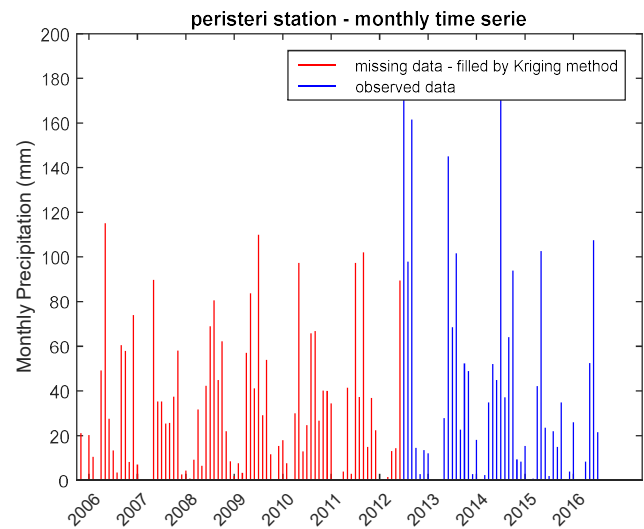
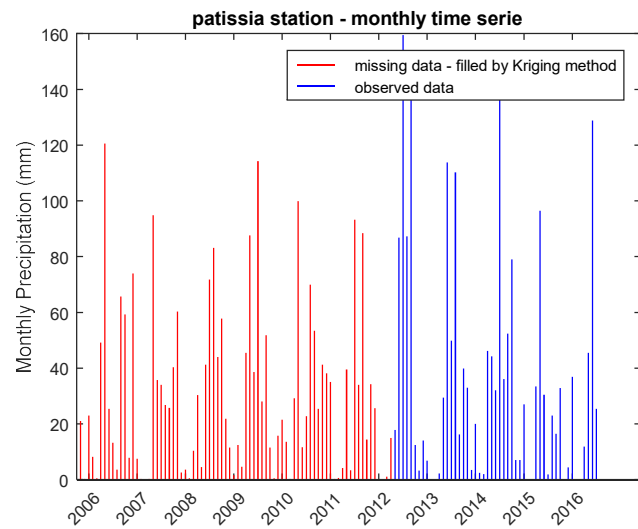
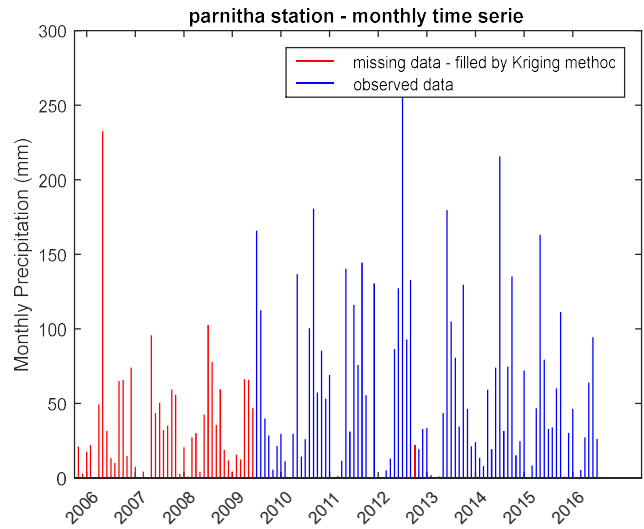
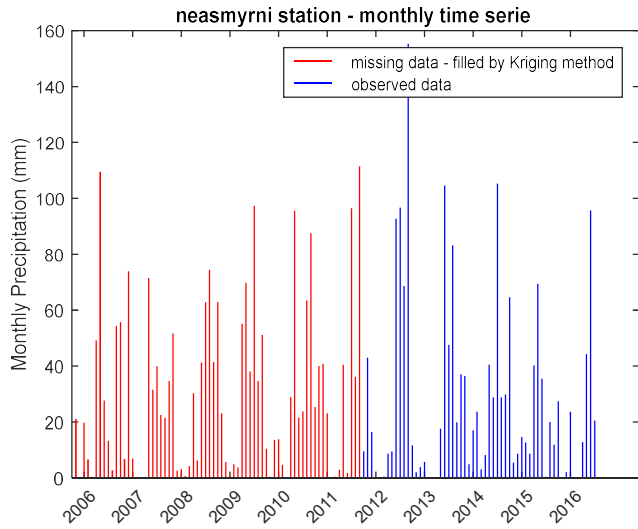
4.1 Συμπλήρωση και ομαδοποίηση βροχομετρικών δεδομένων

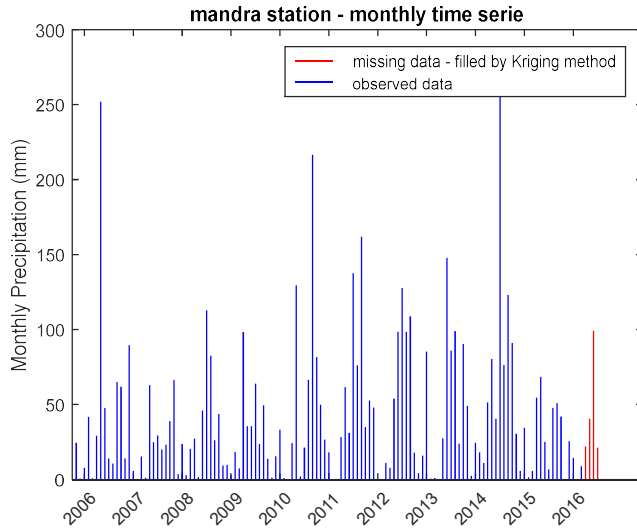
Από το σύνολο των 21 σταθμών του λεκανοπεδίου της Αττικής που επιλέχθηκαν και τελικώς μελετήθηκαν, για 13 από αυτούς χρειάστηκε να γίνει συμπλήρωση των αρχικά διαθέσιμων χρονοσειρών: Αμπελόκηποι, Αθήνα, Εκάλη, Φάληρο, Κάντζα, Μαρούσι, Νέα Σμύρνη, Πάρνηθα, Πατήσια, Περιστερί, Ψυχικό, Άγιος Κοσμάς και Μάνδρα. Οι δύο (2) τελευταίοι προαναφερθέντες σταθμοί ανήκουν στο δίκτυο ΗΟΑ, ενώ οι λοιποί 11 σταθμοί ανήκουν στο δίκτυο σταθμών του Εθνικού Αστεροσκοπείου (ΝΟΑ) και η συμπλήρωση έγινε για το εκάστοτε χρονικό εύρος που απαιτείτο για να συμπληρωθεί η μηνιαία χρονοσειρά δεκαετίας. Αναφορικά για αυτούς τους προς συμπλήρωση σταθμούς, το Περιστερί είναι ο σταθμός με το μικρότερο εύρος χρονοσειράς, ενώ οι σταθμοί της Κάντζας και του Ψυχικού είναι αυτοί με το μεγαλύτερο, από τους υπόλοιπους που ανήκουν στο δίκτυο ΝΟΑ.

Σε κάποιους σταθμούς πέρα από την συμπλήρωση αυτού του εύρους, έπρεπε να συμπληρωθούν και ενδιάμεσα κενά που προέκυψαν στην χρονοσειρά λόγω ανεπαρκών καταγραφών. Η μέθοδος που χρησιμοποιήθηκε για την συμπλήρωση είναι η Kriging, όπως περιεγράφηκε αναλυτικά στο σχετικό υποκεφάλαιο στην Μεθοδολογία. Από τους 13 σταθμούς, όπου εφαρμόστηκε η συγκεκριμένη μεθοδολογία, συνολικά σε 6 η συμπλήρωση αφορούσε και ενδιάμεσα κενά καταγραφών: Αμπελόκηποι, Εκάλη, Κάντζα, Πάρνηθα, Ψυχικό και Νέα Σμύρνη. Η Εκάλη είναι ο σταθμός με τα περισσότερα και μεγαλύτερα σε εύρος ενδιάμεσα κενά καταγραφών.

Στο επόμενο Σχήμα 4-1, απεικονίζονται σε μορφή γραφήματος οι χρονοσειρές των μηνιαίων υψών βροχόπτωσης, για κάθε σταθμό όπου εφαρμόστηκε η συμπλήρωση με την μέθοδο Kriging. Με κόκκινο χρώμα διακρίνονται οι τιμές βροχόπτωσης που προέκυψαν μετά την μεθοδολογία συμπλήρωσης.

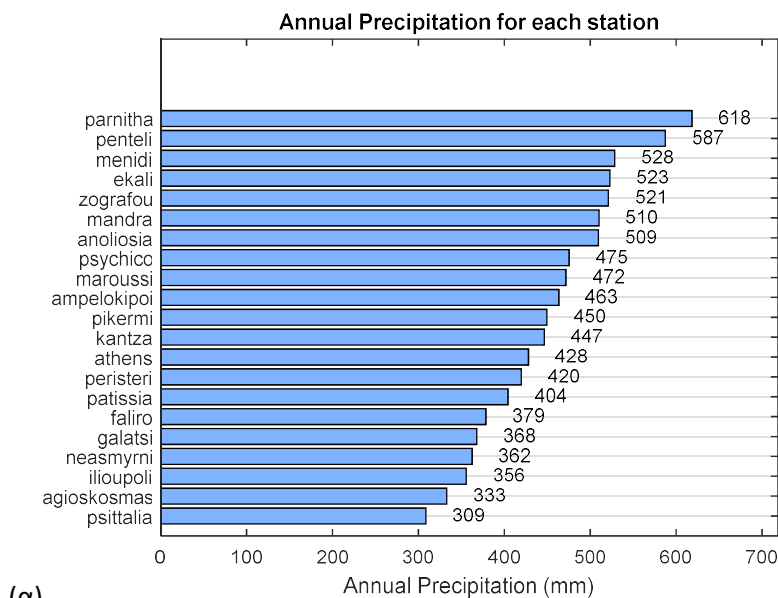




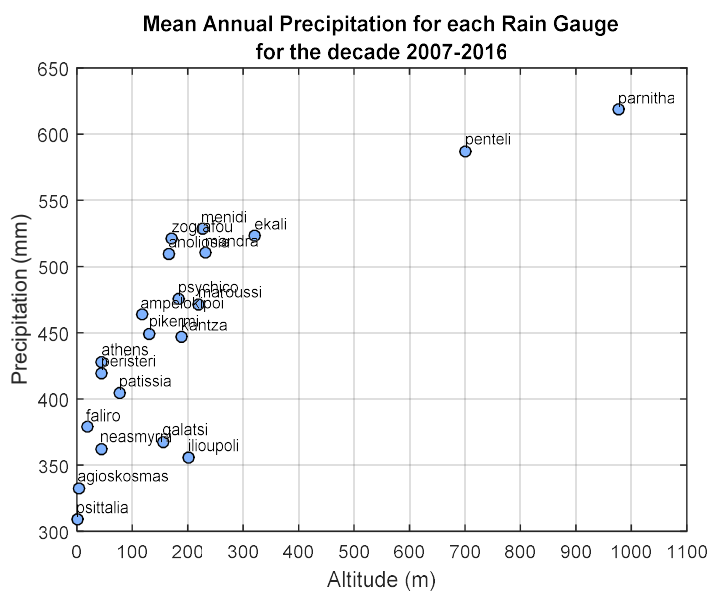


Σχήμα 4-1 Χρονοσειρά μηνιαίων υψών βροχόπτωσης για κάθε σταθμό όπου εφαρμόστηκε συμπλήρωση με μέθοδο Kriging, χρωματική διακριτοποίηση συμπληρωμένων και διαθέσιμων τιμών χρονοσειράς

Μετά την διαδικασία συμπλήρωσης, οι χρονοσειρές, που πλέον έχουν όμοιο χρονικό εύρος, βήμα και σύσταση, ομαδοποιούνται είτε ανά σταθμό, είτε ανά μήνα ή ανά κλιματική περίοδο (ξηρή ή υγρή) είτε ανά έτος, ανάλογα την σκοπιά θεώρησης. Έτσι διαμορφώθηκε το Σχήμα 4-2, όπου απεικονίζεται ραβδόγραμμα (barplot) των μέσων ετησίων βροχοπτώσεων, ανά σταθμό για την δεκαετία 2007-2016, σε φθίνουσα ταξινόμηση. Όπως και αναμενόταν οι σταθμοί με τα υψηλότερα ύψη βροχόπτωσης βρίσκονται γεωχωρικά πλησίον ή εντός ορεινών όγκων και βόρεια εντός του λεκανοπεδίου Αττικής, ενώ οι χαμηλότερες τιμές σημειώνονται σε σταθμούς που βρίσκονται στο νότιο παράκτιο της περιοχής μελέτης. Συνολικά για όλη την περιοχή μελέτης και σύμφωνα με τους διαθέσιμους 21 σταθμούς, το μέσο ετήσιο ύψος βροχόπτωσης για την δεκαετία 2007-2016 είναι 451 mm.



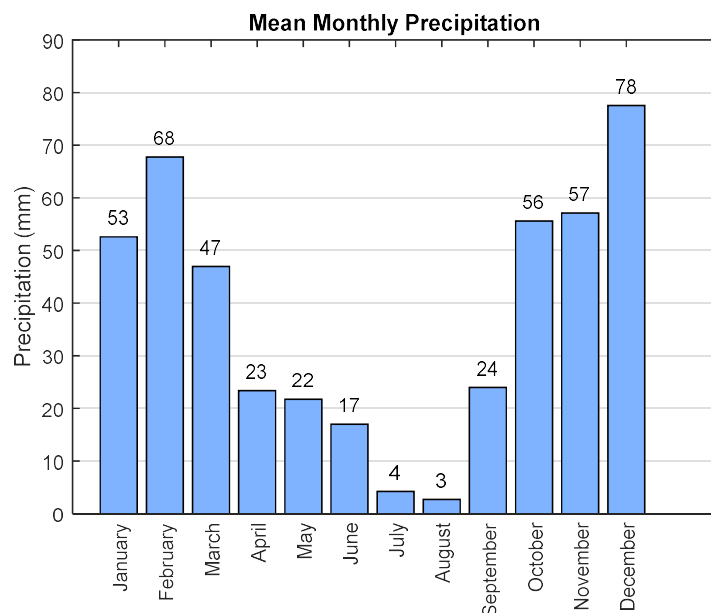
(α)



(β)

Σχήμα 4-2 Μέση ετήσια βροχόπτωση δεκαετίας 2007-2016, ανά σταθμό (α) φθίνουσα ταξινόμηση (β) ως προς το υψόμετρο κάθε σταθμού

Στο επόμενο ραβδόγραμμα (Σχήμα 4-3), οι μηνιαίες χρονοσειρές ομαδοποιήθηκαν ανά μήνα για το σύνολο των σταθμών και έτσι παρουσιάζεται η μέση βροχόπτωση ανά μήνα για την περιοχή του λεκανοπεδίου Αττικής. Ο μήνας με τις υψηλότερες βροχοπτώσεις είναι ο Δεκέμβρης με μέσο ύψος βροχής τα 78mm για την δεκαετία 2007-2016, ενώ ακολουθεί ο Φεβρουάριος με 68mm. Το χαμηλότερο ύψος βροχόπτωσης καταγράφεται για τον μήνα Ιούλιο, όπως αναμενόταν, εφόσον αποτελεί τον ξηρότερο μήνα του έτους, για την περιοχή της Αττικής, με μόλις 3mm βροχής για όλη την διάρκεια του μήνα.



Σχήμα 4-3 Μέση μηνιαία βροχόπτωση δεκαετίας 2007-2016, για το σύνολο των σταθμών

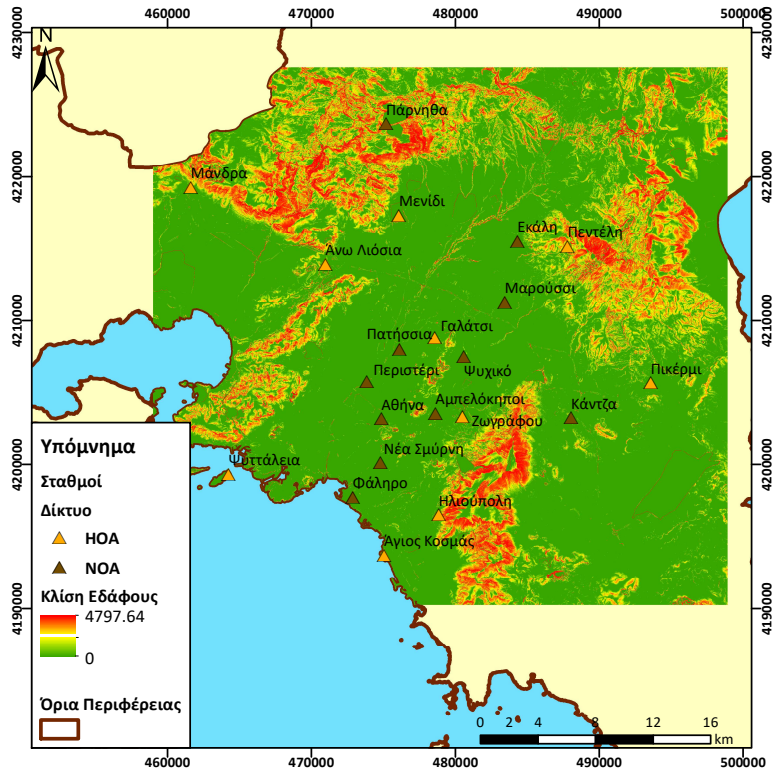
4.2 Τιμές Ανεξάρτητων Μεταβλητών

Εφόσον τα δεδομένα βροχόπτωσης είναι πλέον ομαδοποιημένα και έτοιμα να χρησιμοποιηθούν ως εξαρτημένες μεταβλητές των μοντέλων παλινδρόμησης που πρόκειται να αναπτυχθούν, σειρά έχει η διαμόρφωση των απαιτούμενων τιμών εξαρτημένων μεταβλητών.

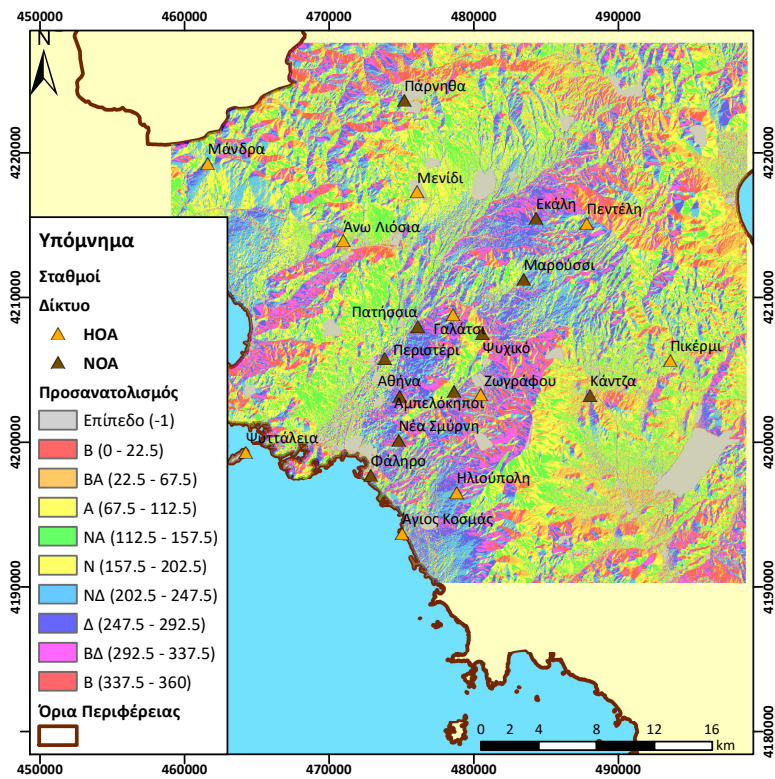
Καταρχάς καθορίζεται το είδος των ανεξάρτητων μεταβλητών που θα εξεταστούν, όπως περιεγράφηκε αναλυτικά στο κεφάλαιο της Μεθοδολογίας. Οι τελικές προς εξέταση μεταβλητές είναι το υψόμετρο, οι γεωγραφικές συντεταγμένες Χ και Υ κατά ΕΓΣΑ '87, η κλίση εδάφους, ο προσανατολισμός, η απόσταση από ποτάμιο υδατικό σύστημα και η απόσταση από την ακτογραμμή.

Για την εύρεση των τιμών κάθε μιας από τις παραπάνω ανεξάρτητες μεταβλητές εξάχθηκαν τα αντίστοιχα χωρικά επίπεδα, μέσω της χρήσης συστήματος γεωγραφικών πληροφοριών (GIS) και πιο συγκεκριμένα μέσω του λογισμικού ArcGIS (έκδοσης 10.5.1). Από το επεξεργασμένο ΨΜΕ (αύξηση μεγέθους φατίου σε 100x100 και περικοπή στο πολύγωνο περιοχής μελέτης) εξάγονται τα χωρικά επίπεδα της κλίσης εδάφους και του προσανατολισμού. Ενώ μέσω του εργαλείου υπολογισμού των ευκλείδειων αποστάσεων, των επιπέδων ποτάμιων υδατικών συστημάτων και ακτογραμμή εξάγονται τα χωρικά επίπεδα με την κοντινότερη απόσταση κάθε σημείου από ποτάμι και από την θάλασσα, αντίστοιχα.

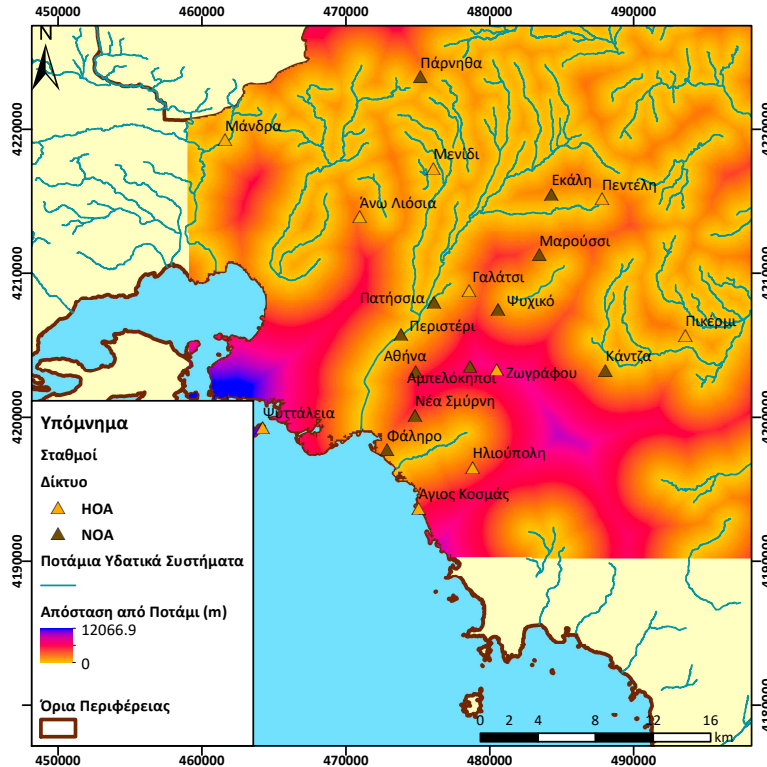
Παρότι οι συντεταγμένες για κάθε έναν από τους επιλεγμένους σταθμούς είναι γνωστές και ακριβείς, για λόγους ομοιογένειας ως προς την χωρική ακρίβεια του συνόλου των δεδομένων που θα εξεταστούν ως ανεξάρτητες μεταβλητές, εξάγονται και τα επίπεδα με τη χωρική πληροφορία συντεταγμένων Χ και Υ σε ακρίβεια φατίου 100x100. Παρακάτω παρατίθενται οι χάρτες με τα θεματικά επίπεδα των προαναφερθέντων μεταβλητών.



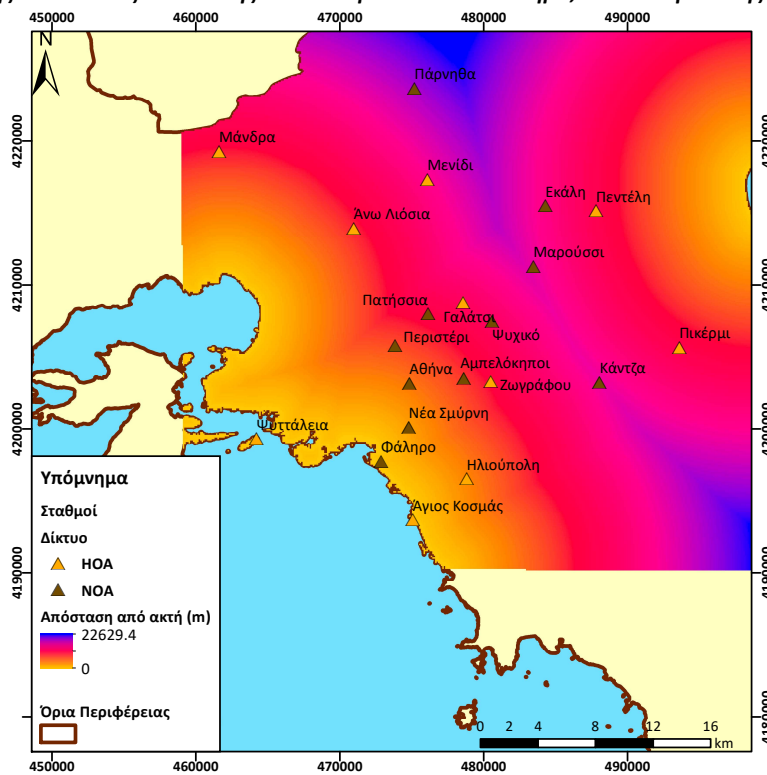
Σχήμα 4-4 Χάρτης κλίσης εδάφους (ποσοστιαία %) στο πολύγωνο της περιοχής μελέτης



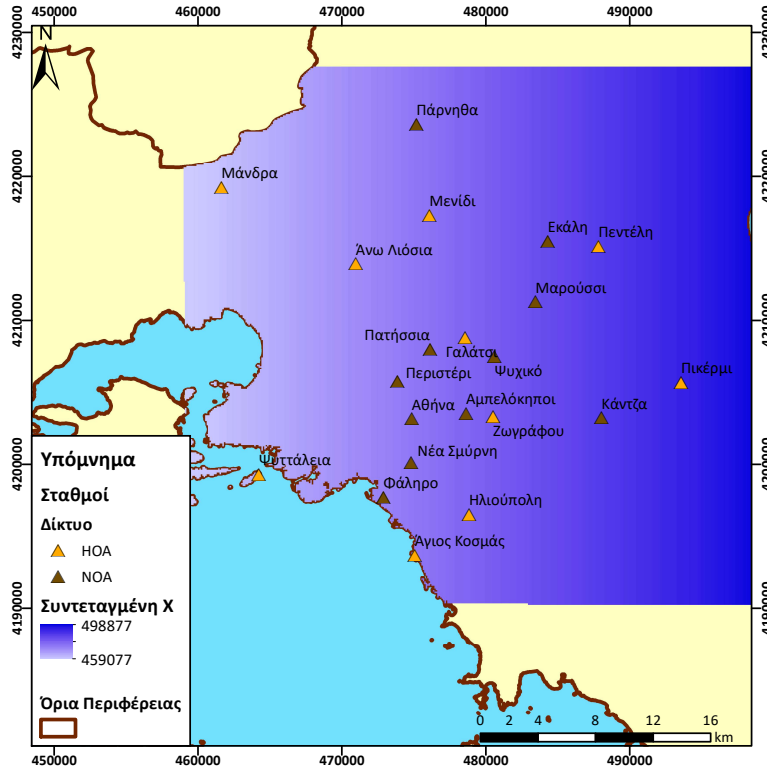
Σχήμα 4-5 Χάρτης προσανατολισμού (μοίρες °) και οι κλάσεις τιμών, στο πολύγωνο της περιοχής μελέτης



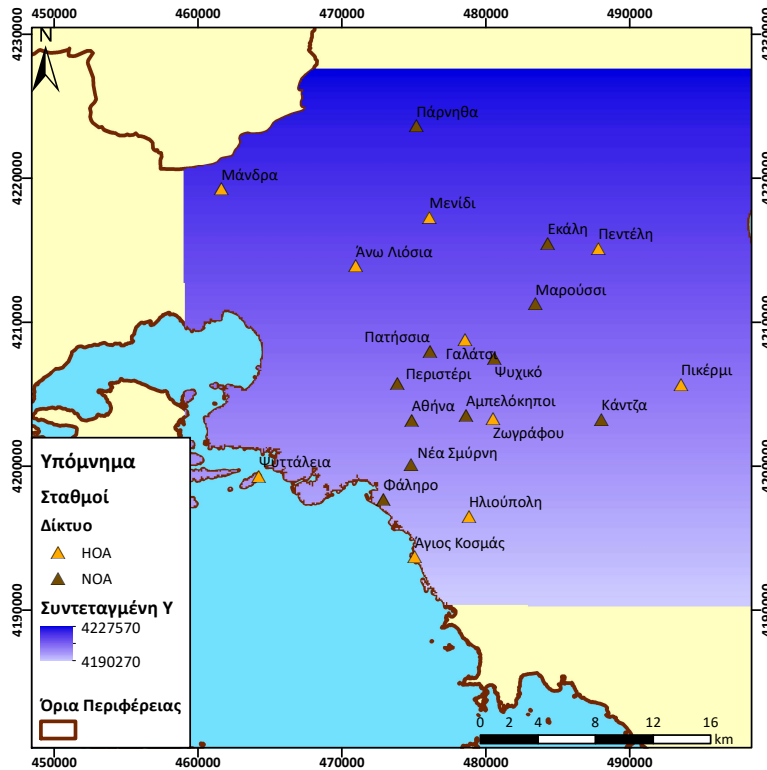
Σχήμα 4-6 Χάρτης ευκλείδειας απόστασης από ποτάμι υδατικό σύστημα, στο πολύγωνο της περιοχής μελέτης



Σχήμα 4-7 Χάρτης ευκλείδειας απόστασης από την ακτογραμμή, στο πολύγωνο περιοχής μελέτης



Σχήμα 4-8 Χάρτης συντεταγμένων Χ κατά ΕΓΣΑ '87, στο πολύγωνο περιοχής μελέτης



Σχήμα 4-9 Χάρτης συντεταγμένων Υ κατά ΕΓΣΑ '87, στο πολύγωνο περιοχής μελέτης

Μετά την παραγωγή των παραπάνω χωρικών επιπέδων, εξάγεται η τιμή που αντιστοιχεί σε κάθε σταθμό για κάθε ανεξάρτητη μεταβλητή. Οι τιμές αυτές παρουσιάζονται στον επόμενο πίνακα.

Πίνακας 4-1 Τιμές εξεταζόμενων ανεξάρτητων μεταβλητών, ανά σταθμό

α/α	Σταθμοί	Υψόμετρο (m)	Κλίση Εδάφους (%)	Προσανατολισμός (°)	Απόσταση από Ποτάμι (m)	Απόσταση από Ακτή (m)	X	Y
1	Αμπελόκηποι	117	0.13	225.0	4710	7720	478677	4203569
2	Αθήνα	45	6.03	240.2	2642	4802	474877	4203169
3	Εκάλη	321	13.47	65.1	949	13913	484277	4215469
4	Φάληρο	18	7.49	135.0	1273	361	472877	4197769
5	Κάντζα	188	2.06	151.4	894	15370	488077	4203269
6	Μαρούσι	220	3.37	268.8	1556	15412	483477	4211269
7	Νέα Σμύρνη	43	5.77	339.4	2600	2561	474777	4200169
8	Πάρνηθα	978	0.00	-1	1063	17974	475177	4223669
9	Πατήσια	76	2.62	240.5	224	9661	476177	4207969
10	Περιστερί	45	2.80	194.0	0	6772	473877	4205769
11	Ψυχικό	185	9.43	167.3	1360	11739	480577	4207469
12	Άγιος Κοσμάς	3	6.68	208.9	3130	0	475077	4193669
13	Άνω Λιόσια	167	0.00	-1	1649	8161	470977	4213869
14	Γαλάτσι	156	16.00	14.0	1393	11512	478577	4208769
15	Ηλιούπολη	201	8.91	296.0	2220	4031	478877	4196469
16	Μάνδρα	232	0.00	-1	141	8502	461577	4219269
17	Μενίδι	226	0.00	-1	412	14287	476077	4217269
18	Πεντέλη	701	3.75	243.6	539	10439	487777	4215169
19	Πικέρμι	131	11.93	52.5	800	10222	493577	4205669
20	Ψυττάλεια	0	0.00	-1	6800	0	464277	4199269
21	Ζωγράφου	172	9.78	309.1	4875	9021	480477	4203269

4.3 Μοντέλα Παλινδρόμησης

Η δημιουργία των μοντέλων παλινδρόμησης λαμβάνει χώρα εντός του προγραμματιστικού περιβάλλοντος της MATLAB, όπου και ακολουθείται η μεθοδολογία που αναπτύχθηκε εκτενώς στο προηγούμενο κεφάλαιο. Ως δεδομένα εισόδου οφείλουν να εισαχθούν το σύνολο των τιμών ανεξάρτητων και εξαρτημένων μεταβλητών. Για τον σκοπό αυτό και για κάθε διαφορετικό μοντέλο παλινδρόμησης που ερευνάται δημιουργείται ένας ενιαίος πίνακας, ο οποίος περιλαμβάνει τα στοιχεία του Πίνακας 4-1 και κάθε φορά μία επιπλέον στήλη της εξαρτημένης μεταβλητής – ύψους βροχόπτωσης για την εκάστοτε χρονικά εξεταζόμενη κλίμακα (μήνα, έτος ή ξηρή και υγρή περίοδος έτους).

Πέρα από μοντέλα πολλαπλής γραμμικής παλινδρόμησης, εξετάστηκαν και μοντέλα πολυωνυμικής ή λογαριθμικής παλινδρόμησης, όπως αντλήθηκαν από σχετική επιστημονική

βιβλιογραφία¹⁰, τα οποία ωστόσο δεν προσαρμόστηκαν ικανοποιητικά στις παρατηρούμενες τιμές (συγκριτικά με τα μοντέλα πολλαπλής παλινδρόμησης) , όπως φάνηκε μέσα από τους στατιστικούς δείκτες των συντελεστών συσχέτισης, προσδιορισμού και προσαρμοσμένου συντελεστή προσδιορισμού και έτσι τελικώς απορρίφθηκαν. Οι απορριφθείσες σχέσεις παρουσιάζονται ενδεικτικά παρακάτω.

$$\log_{10}(P) = b_0 + b_1 \cdot h + b_2 \cdot slp + b_3 \cdot asp + b_4 \cdot dRiv + b_5 \cdot dCst + b_6 \cdot X + b_7 \cdot Y \quad (4.1)$$

$$P = b_0 + b_1 \cdot h + b_2 \cdot h^2 + b_3 \cdot h^3 + b_4 \cdot slp + b_5 \cdot slp^2 + b_6 \cdot slp^3 + b_7 \cdot asp + b_8 \cdot asp^2 + b_9 \cdot asp^3 + b_{10} \cdot X + b_{11} \cdot X^2 + b_{12} \cdot X^3 + b_{13} \cdot Y + b_{14} \cdot Y^2 + b_{15} \cdot Y^3 \quad (4.2)$$

όπου,

P : ύψος βροχόπτωσης σε mm

b_i : συντελεστής (σταθερός ή ανεξάρτητης μεταβλητής) αδιάστατος

h : υψόμετρο εδάφους σε m

slp : κλίση εδάφους σε ποσοστιαία μορφή %

asp : προσανατολισμός εδάφους σε μοίρες (°)

X : συντεταγμένη x , κατά ΕΓΣΑ '87

Y : συντεταγμένη y , κατά ΕΓΣΑ '87

Έτσι για την διαμόρφωση της σχέσης του τελικού μοντέλου πολλαπλής γραμμικής παλινδρόμησης, εξετάστηκε σε πρώτο στάδιο ο πίνακας συντελεστών συσχέτισης, με σκοπό τον καθορισμό των αρχικών ανεξάρτητων μεταβλητών που θα συμμετέχουν στο μοντέλο. Στη συνέχεια μέσω της παραλλαγής Backward Elimination της μεθόδου Stepwise Procedure, ενός αλγορίθμου σταδιακής απόρριψης μεταβλητών, όπως αναλύθηκε στο κεφάλαιο της Μεθοδολογίας, καθορίζεται η τελική σχέση. Ενδεικτικά ο πίνακας συσχέτισης των ανεξάρτητων μεταβλητών με τις τιμές ετήσιας βροχόπτωσης κάθε σταθμού παρουσιάζεται ακολούθως.

¹⁰ *Ranhao et al., 2008*

Πίνακας 4-2 Πίνακας συντελεστών συσχέτισης μεταξύ εξεταζόμενων ανεξάρτητων μεταβλητών και εξαρτημένης μεταβλητής, για τις ετήσιες τιμές βροχόπτωσης των διαθέσιμων σταθμών

	Υψόμετρο	Κλίση Εδάφους	Προσανατολισμός	Απόσταση από Ποτάμι	Απόσταση από Ακτή	X	Y	Βροχόπτωση
Υψόμετρο	1							
Κλίση Εδάφους	-0.166	1						
Προσανατολισμός	-0.218	0.137	1					
Απόσταση από Ποτάμι	-0.313	-0.054	0.184	1				
Απόσταση από Ακτή	0.627	-0.029	-0.269	-0.501	1			
X	0.230	0.468	0.285	-0.253	0.481	1		
Y	0.713	-0.269	-0.526	-0.527	0.763	-0.003	1	
Βροχόπτωση	0.782	-0.232	-0.225	-0.385	0.738	0.258	0.839	1

Είναι εμφανές ότι μεταβλητές της κλίσης εδάφους, προσανατολισμού και συντεταγμένης X απορρίπτονται από το πρώτο κιάλας στάδιο διαμόρφωσης του μοντέλου, καθώς οι αντίστοιχοι συντελεστές συσχέτισης με την βροχόπτωση είναι ιδιαίτερα χαμηλοί. Σχετικά υψηλές συσχετίσεις παρουσιάζουν με φθίνουσα κατάταξη οι μεταβλητές της γεωγραφικής συντεταγμένης Y, του υψομέτρου και της απόστασης από την ακτογραμμή. Παρόμοια εικόνα παρουσιάζεται και στους συντελεστές συσχέτισης για μηνιαία ύψη βροχόπτωσης όσο και για ύψη βροχής που αφορούν τις υγρές και ξηρές περιόδους του έτους.

Η τελική σχέση που προσαρμόζεται καλύτερα από τις εξεταζόμενες και για το σύνολο των 15 μοντέλων που εξετάζονται είναι η ακόλουθη:

$$Precipitation = b_0 + b_1 \cdot Altitude + b_2 \cdot Y_{coordinate} \quad (4.3)$$

Η παραπάνω σχέση συσχετίζει γραμμικά την εξαρτημένη μεταβλητή βροχόπτωσης με τις ανεξάρτητες μεταβλητές του υψομέτρου εδάφους και της συντεταγμένης Y (γεωγραφικό πλάτος, κατά σύστημα ΕΓΣΑ '87).

Η συγκεκριμένη σχέση εφαρμόζεται και για τα 15 διαφορετικά μοντέλα, για κάθε ένα από τα οποία διαφοροποιούνται οι εισαχθείσες τιμές της εξαρτημένης μεταβλητής (βροχόπτωση), εξάγοντας έτσι τις τιμές των συντελεστών b_0, b_1, b_2 που αντιστοιχούν σε κάθε μοντέλο, όπως επίσης και ένα σύνολο από στατιστικούς δείκτες, βάσει των οποίων τα μοντέλα αξιολογούνται. Στον επόμενο πίνακα παρατίθενται οι τιμές των συντελεστών κάθε μοντέλου, καθώς και οι στατιστικοί δείκτες R_{adj}^2 και p-value που αποτελούν και τα βασικότερα κριτήρια αξιολόγησης των μοντέλων.

Πίνακας 4-3 Συντελεστές, $R^2_{adjusted}$, p -values, μοντέλων παλινδρόμησης

α/α	Μοντέλο	Συντελεστές			$R^2_{adjusted}$	p -values		
		Σταθερά	Υψόμετρο	Συντετ. Υ		Σταθερά	Υψόμετρο	Συντετ. Υ
1	Ιανουάριος	-2249	1.29E-02	5.46E-04	0.52	0.048	0.153	0.044
2	Φεβρουάριος	-3378	1.25E-02	8.18E-04	0.57	0.014	0.232	0.013
3	Μάρτιος	-1408	3.01E-02	3.44E-04	0.55	0.310	0.014	0.297
4	Απρίλιος	-1051	3.15E-03	2.55E-04	0.48	0.026	0.382	0.023
5	Μάιος	-846	1.54E-02	2.06E-04	0.78	0.068	0.000	0.063
6	Ιούνιος	-2656	6.75E-03	6.35E-04	0.79	0.000	0.139	0.000
7	Ιούλιος	-83	3.74E-03	2.05E-05	0.08	0.833	0.251	0.826
8	Αύγουστος	-794	-1.32E-03	1.89E-04	0.32	0.015	0.587	0.014
9	Σεπτέμβρης	-1785	-2.70E-03	4.30E-04	0.31	0.017	0.631	0.016
10	Οκτώβρης	-3555	1.35E-02	8.58E-04	0.80	0.000	0.045	0.000
11	Νοέμβρης	-2760	1.17E-02	6.69E-04	0.39	0.074	0.337	0.070
12	Δεκέμβρης	-3971	2.76E-02	9.61E-04	0.77	0.004	0.011	0.003
13	Ετήσιο	-24536	1.33E-01	5.93E-03	0.75	0.002	0.031	0.002
14	Ξηρή Περίοδος	-7214	2.50E-02	1.74E-03	0.79	0.000	0.046	0.000
15	Υγρή Περίοδος	-18241	1.19E-01	4.42E-03	0.77	0.003	0.013	0.003

Στον προηγούμενο πίνακα, επισημαίνονται χρωματικά οι τιμές του συντελεστή R^2_{adj} που βρίσκονται κάτω από το όριο του 0.70, δηλαδή μοντέλα τα οποία δεν επιδεικνύουν ικανοποιητική προσαρμογή στις παρατηρούμενες τιμές. Παράλληλα χρωματική επισήμανση εφαρμόζεται και στις τιμές των παραμέτρων (b_i) των ανεξάρτητων μεταβλητών που δεν είναι στατιστικά σημαντικές, δηλαδή για τιμές p -value μεγαλύτερες από το όριο του 0.05¹¹. Το τελευταίο κριτήριο είναι καθοριστικό για τον διαχωρισμό των τελικά στατιστικά σημαντικών μοντέλων που αξίζει να σχολιασθούν στα επόμενα. Αναλυτικότερα, παρότι κάποια μοντέλα (π.χ. αυτά που αντιστοιχούν στον μήνα Μάιο και Ιούνιο) παρουσιάζουν υψηλό προσαρμοσμένο συντελεστή προσδιορισμού (0.78 και 0.79 αντίστοιχα), απορρίπτονται εφόσον δεν είναι στατιστικά σημαντικά και επομένως άξια περαιτέρω σχολιασμού, αφού οι τιμές p -value δεν πληρούν το όριο που προαναφέρθηκε. Συνεπώς στο επόμενο κεφάλαιο θα παρουσιαστούν περισσότερα στοιχεία αποτελεσμάτων για τα μοντέλα των: Οκτωβρίου, Δεκεμβρίου, ετησίου, ξηρής και υγρής περιόδου αντίστοιχα.

4.4 Στατιστικοί δείκτες και γραφήματα στατιστικά σημαντικών μοντέλων

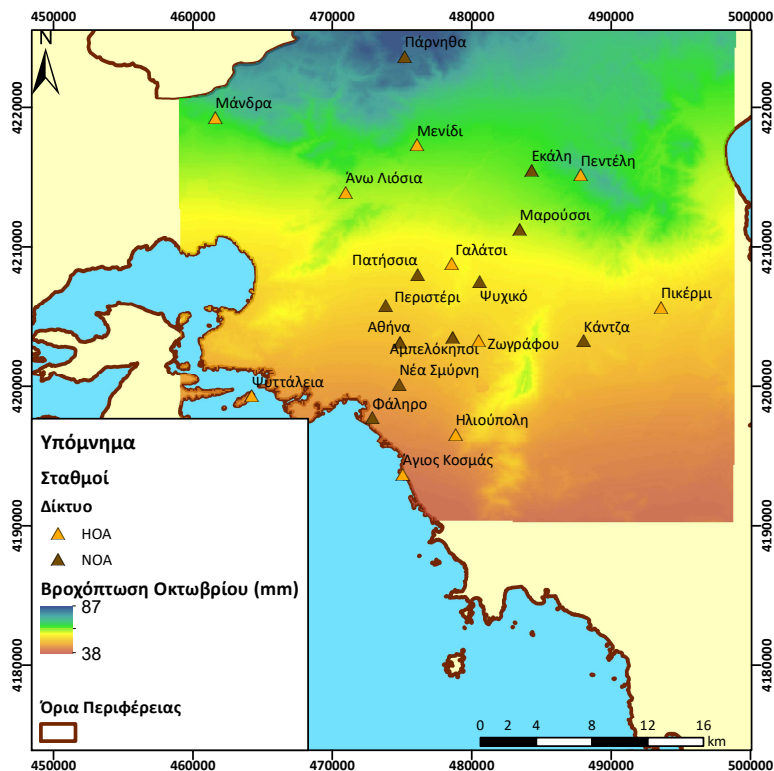
Για κάθε ένα από τα στατιστικά σημαντικά μοντέλα παρατίθενται ένα χάρτης με την προβλεπόμενη, από το μαθηματικό μοντέλο, βροχόπτωση στην περιοχή μελέτης και ένα σύνολο διαγραμμάτων. Τα διαγράμματα αυτά σχετίζονται με την προσαρμογή του κάθε μοντέλου και την αξιολόγηση αυτής, στα πραγματικά δεδομένα καθώς και με πληροφορίες για την κατανομή παρατηρούμενου και εκτιμώμενου δείγματος αλλά και την διερεύνηση της προσαρμογής του μοντέλου υπό το γεωχωρικό πρίσμα. Ειδικότερα τα διαγράμματα αυτά είναι τα εξής:

¹¹ για επίπεδο σημαντικότητας $\alpha=5\%$ (διάστημα εμπιστοσύνης 95%)

- τρισδιάστατο διάγραμμα αποτύπωσης του επιπέδου, όπως καθορίζεται από την μαθηματική σχέση του μοντέλου που το περιγράφει
- διάγραμμα διασποράς σημείων με τετμημένη την πραγματικά καταγεγραμμένη τιμή και τεταγμένη την προβλεπόμενη από το μοντέλο τιμή
- θηκόγραμμα (boxplot) συνόλου δείγματος παρατηρούμενων και προβλεπόμενων τιμών
- ραβδόγραμμα (barplot) με φθίνουσα κατάταξη, ποσοστιαίων διαφορών μεταξύ ζεύγους τιμών (παρατηρούμενων – προβλεπόμενων από το μοντέλο βροχοπτώσεων), για κάθε εξεταζόμενο σταθμό

4.4.1 Μοντέλο Παλινδρόμησης Οκτωβρίου

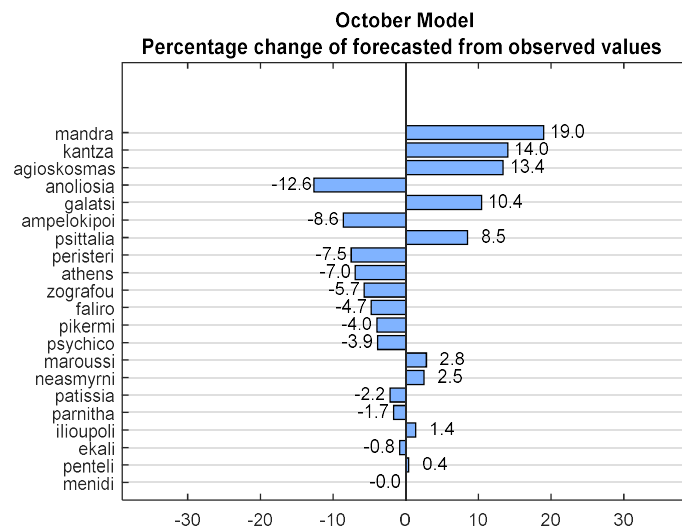
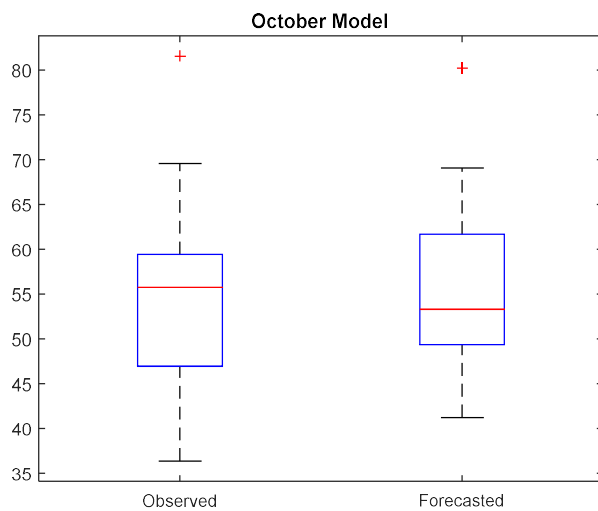
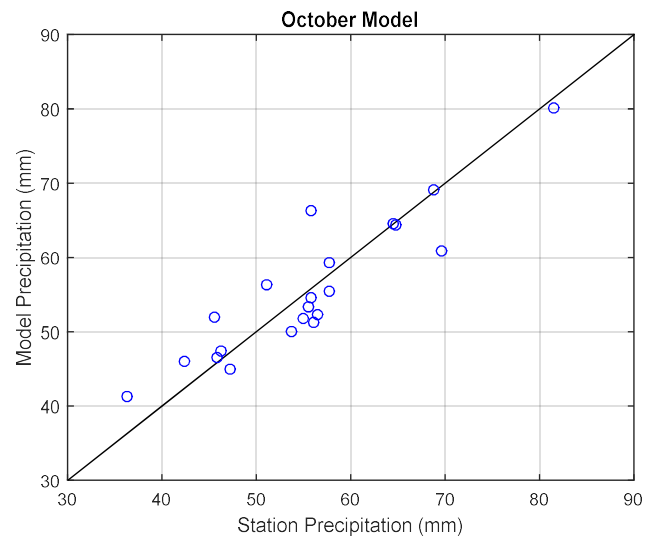
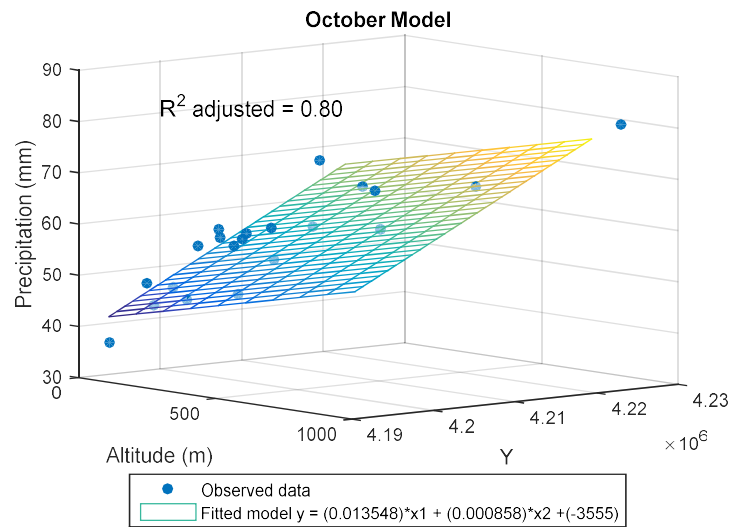
Στον επόμενο χάρτη (Σχήμα 4-10) αποτυπώνεται η ποσοτικό-χωρική κατανομή του προβλεπόμενου συνόλου της βροχόπτωσης για τον μήνα Οκτώβριο. Το εύρος κυμαίνεται περί των 40-90 mm βροχής, με τα μεγαλύτερα ύψη να προβλέπονται στο βόρειο τμήμα της περιοχής μελέτης και ειδικότερα στον ορεινό όγκο της Πάρνηθας και ελαφρώς χαμηλότερα στη Πεντέλη. Στο νότιο τμήμα και ιδιαίτερα στην παράκτια ζώνη καταγράφονται τα χαμηλότερα ύψη.



Σχήμα 4-10 Χάρτης χωρικά προβλεπόμενης μηνιαίας βροχόπτωσης σε mm για τον Οκτώβριο, σύμφωνα με το μοντέλο που αναπτύχθηκε, στο πολύγωνο περιοχής μελέτης

Στο επόμενο τρισδιάστατο διάγραμμα είναι εμφανής η καλή προσαρμογή του μοντέλου, καθώς το επίπεδο που το αποτυπώνει βρίσκεται κοντά στα σημεία των καταγεγραμμένων πραγματικών

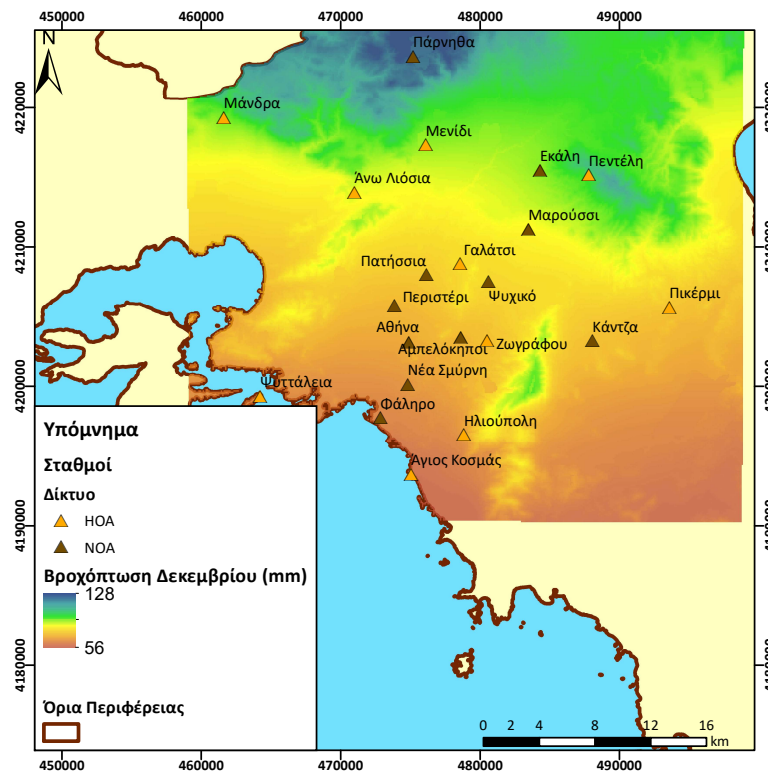
τιμών, ενώ ο προσαρμοσμένος συντελεστής προσδιορισμού είναι αρκετά υψηλός (0.80). Η καλή προσαρμογή του μοντέλου είναι εμφανής και από το διάγραμμα διασποράς, όπου τα περισσότερα σημεία εντοπίζονται επάνω στην διαγώνιο $y = x$ ή είναι αρκετά κοντά σε αυτήν. Στα θηκογράμματα τόσο του δείγματος παρατηρούμενων τιμών όσο και σε αυτό των τιμών πρόβλεψης απεικονίζεται μία ακραία τιμή (outlier) στο καθένα. Το εύρος των δύο δειγμάτων είναι παρόμοιο, με ελαφρώς μεγαλύτερο αυτό των πραγματικών τιμών, ενώ η αντίστοιχη διάμεσος παρατηρείται σε ελαφρώς υψηλότερα από την διάμεσο των τιμών του μοντέλου. Στο ραβδόγραμμα οι σταθμοί που εμφανίζουν αποκλίσεις μεγαλύτερες του $\pm 10\%$ είναι οι εξής: Μάνδρα, Κάντζα, Άγιος Κοσμάς, Άνω Λιόσια και Γαλάτσι. Ενώ κάτω του $\pm 1\%$ βρίσκονται οι σταθμοί της Εκάλης, της Πεντέλης και ο σταθμός στο Μενίδι.



Σχήμα 4-11 Τρισδιάστατο διάγραμμα μοντέλου παλινδρόμησης, διάγραμμα διασποράς, θεκόγραμμα δειγμάτων πραγματικών και εκτιμώμενων τιμών, ραβδόγραμμα ποσοστιαίων αποκλίσεων πραγματικών και εκτιμώμενων τιμών ανά σταθμό, για το μοντέλο Οκτωβρίου (με σειρά κατάταξης από αριστερά προς τα δεξιά και από πάνω προς τα κάτω)

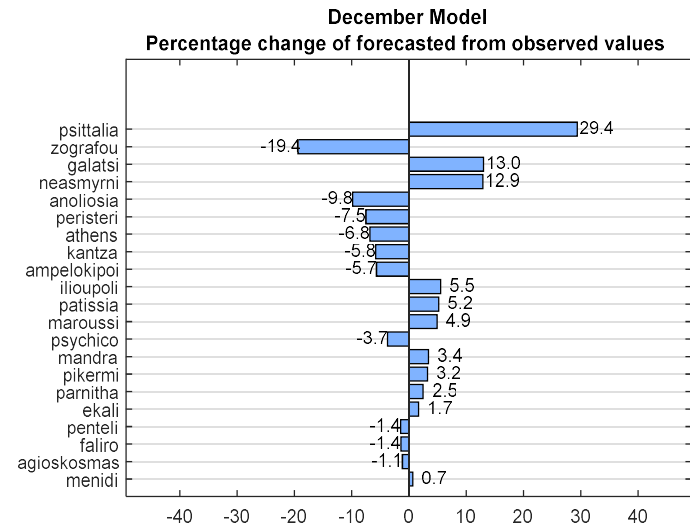
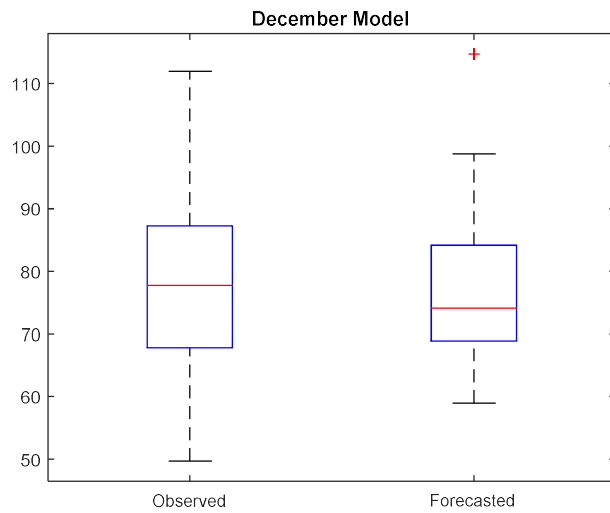
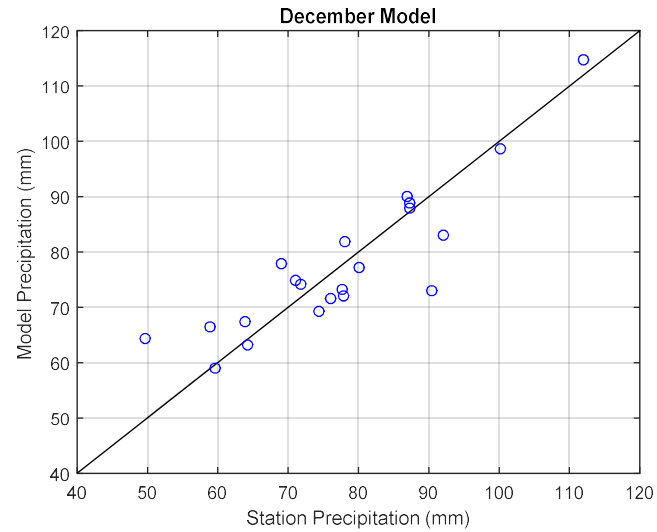
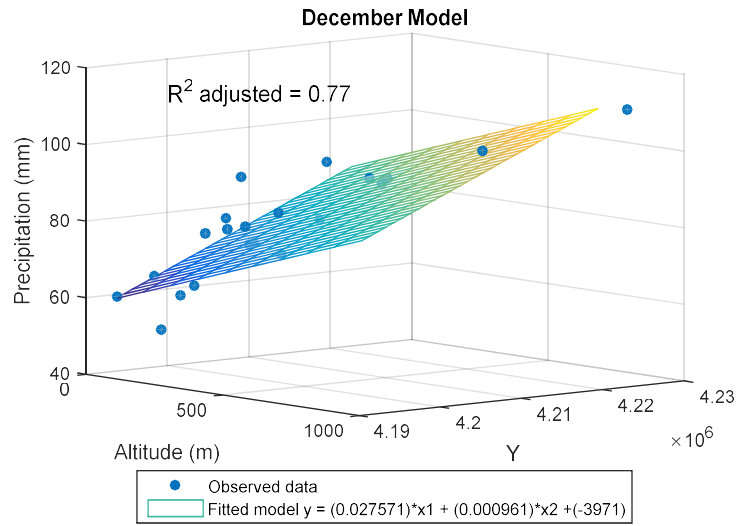
4.4.2 Μοντέλο Παλινδρόμησης Δεκεμβρίου

Στον επόμενο χάρτη (Σχήμα 4-12) αποτυπώνεται η ποσοτικό-χωρική κατανομή του προβλεπόμενου συνόλου της βροχόπτωσης για τον μήνα Δεκεμβρίου. Το εύρος κυμαίνεται περί των 55-130 mm βροχής, με τα μεγαλύτερα ύψη να προβλέπονται και εδώ όμοια με το προηγούμενο μοντέλο, δηλαδή κυρίως στο βόρειο τμήμα της περιοχής μελέτης στους ορεινούς όγκους της Πάρνηθας και της Πεντέλης. Στο νότιο τμήμα και ιδιαίτερα στην παράκτια ζώνη καταγράφονται τα χαμηλότερα ύψη.



Σχήμα 4-12 Χάρτης χωρικά προβλεπόμενης μηνιαίας βροχόπτωσης σε mm για το Δεκέμβριο, σύμφωνα με το μοντέλο που αναπτύχθηκε, στο πολύγωνο περιοχής μελέτης

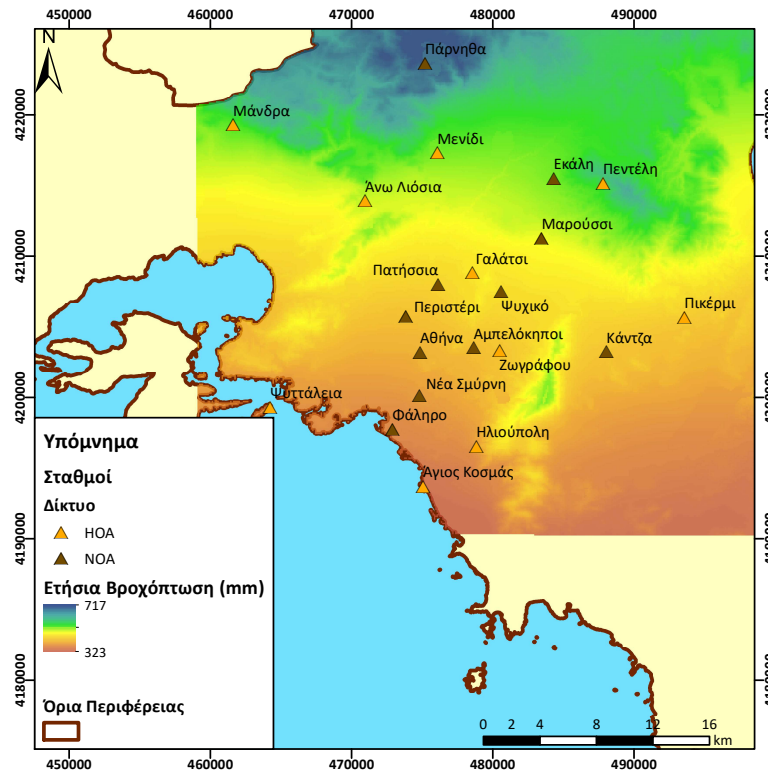
Στο επόμενο τρισδιάστατο διάγραμμα είναι εμφανής η καλή προσαρμογή του μοντέλου, εφόσον το επίπεδο της σχέσης, βρίσκεται κοντά στα σημεία των καταγεγραμμένων τιμών, ενώ ο προσαρμοσμένος συντελεστής προσδιορισμού είναι αρκετά υψηλός (0.77). Η καλή προσαρμογή του μοντέλου είναι εμφανής και από το διάγραμμα διασποράς, όπου τα περισσότερα σημεία εντοπίζονται επάνω στην διαγώνιο $y = x$ ή είναι αρκετά κοντά σε αυτήν. Στο θηκογράμμα του δείγματος των τιμών πρόβλεψης απεικονίζεται μία ακραία τιμή (outlier), που αντιστοιχεί στην τιμή του σταθμού της Πάρνηθας. Το εύρος του δείγματος των πραγματικών τιμών είναι μεγαλύτερο από αυτό των εκτιμώμενων. Η διάμεσος βρίσκεται στο κέντρο του 50% των παρατηρήσεων για το σύνολο των πραγματικών τιμών, ενώ στις εκτιμώμενες τιμές παρατηρείται ήπια θετική ασυμμετρία. Στο ραβδόγραμμα ο σταθμός της Ψυττάλειας σημειώνει την μεγαλύτερη απόκλιση (~30%). Λοιποί σταθμοί με απόλυτη απόκλιση άνω του 10% είναι οι: Ζωγράφου, Γαλάτσι και Νέα Σμύρνη. Και σε αυτό το μοντέλο ο σταθμός στο Μενίδι σημειώνει την μικρότερη απόκλιση της τάξεως του 0.7%



Σχήμα 4-13 Τρισδιάστατο διάγραμμα μοντέλου παλινδρόμησης, διάγραμμα διασποράς, θηκόγραμμα δειγμάτων πραγματικών και εκτιμώντων τιμών, ραβδόγραμμα ποσοστιαίων αποκλίσεων πραγματικών και εκτιμώντων τιμών ανά σταθμό, για το μοντέλο Δεκεμβρίου (με σειρά κατάταξης από αριστερά προς τα δεξιά και από πάνω προς τα κάτω)

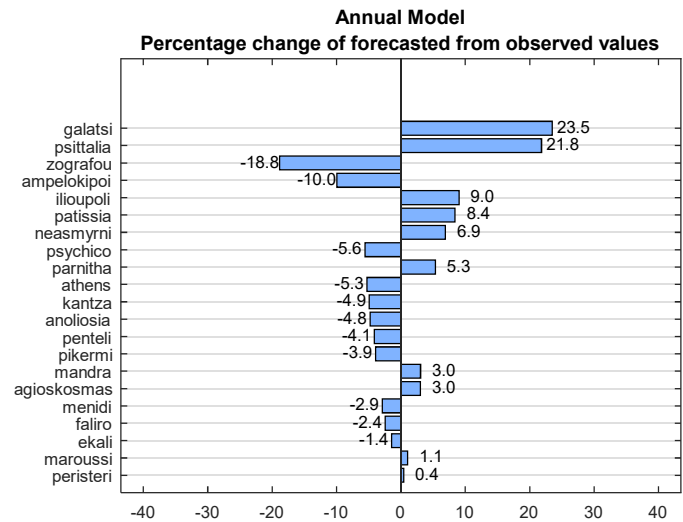
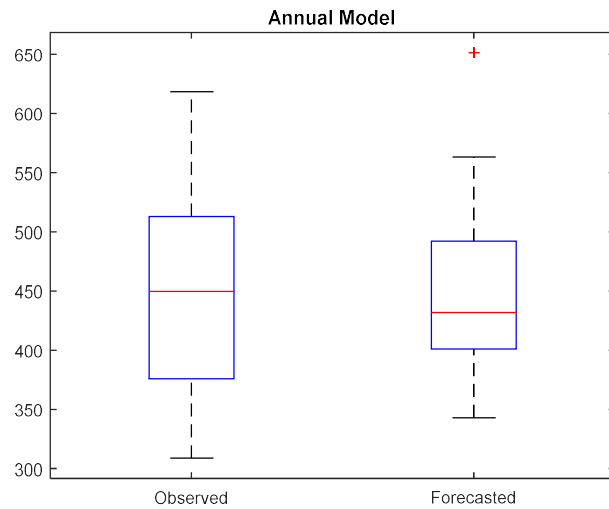
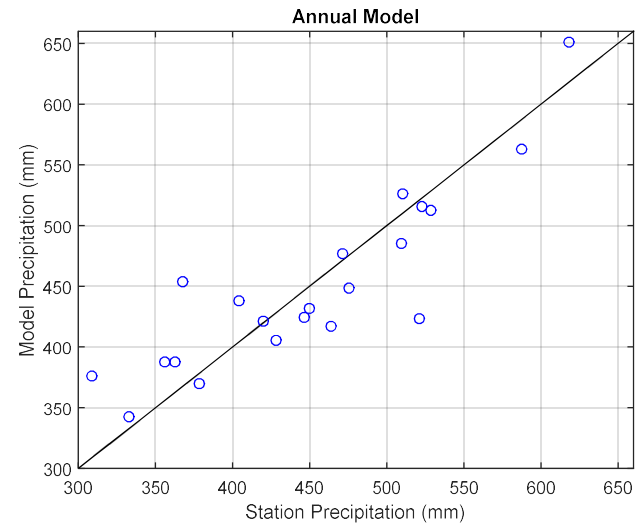
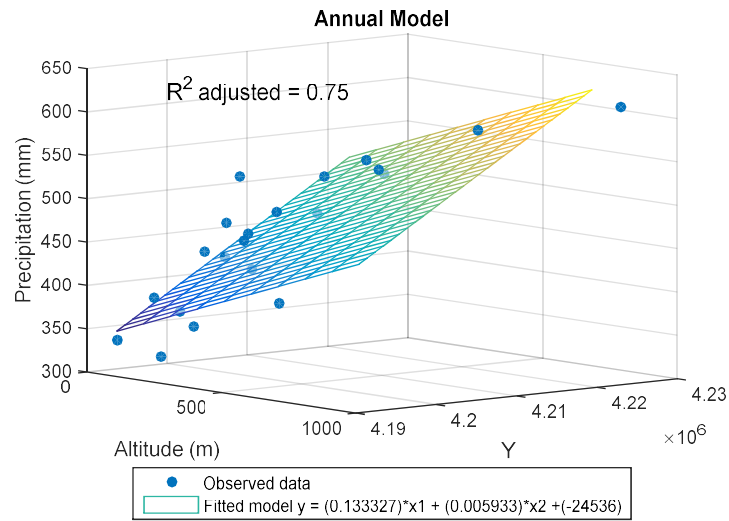
4.4.3 Μοντέλο Παλινδρόμησης Έτους

Σύμφωνα με τον ακόλουθο χάρτη (Σχήμα 4-14), ο οποίος αντιστοίχως αναφέρεται στην εκτιμώμενη ετήσια βροχόπτωση για την περιοχή μελέτης, το εύρος κυμαίνεται περί των 320-720 mm βροχής. Στο μεγαλύτερο μέρος της έκτασης, της περιοχής μελέτης κυριαρχούν χαμηλά ύψη βροχόπτωσης. Οι υψηλότερες τιμές εντοπίζονται ξανά στους ορεινούς όγκους της Πάρνηθας και της Πεντέλης.



Σχήμα 4-14 Χάρτης χωρικά προβλεπόμενης ετήσιας βροχόπτωσης σε mm, σύμφωνα με το μοντέλο που αναπτύχθηκε, στο πολύγωνο περιοχής μελέτης

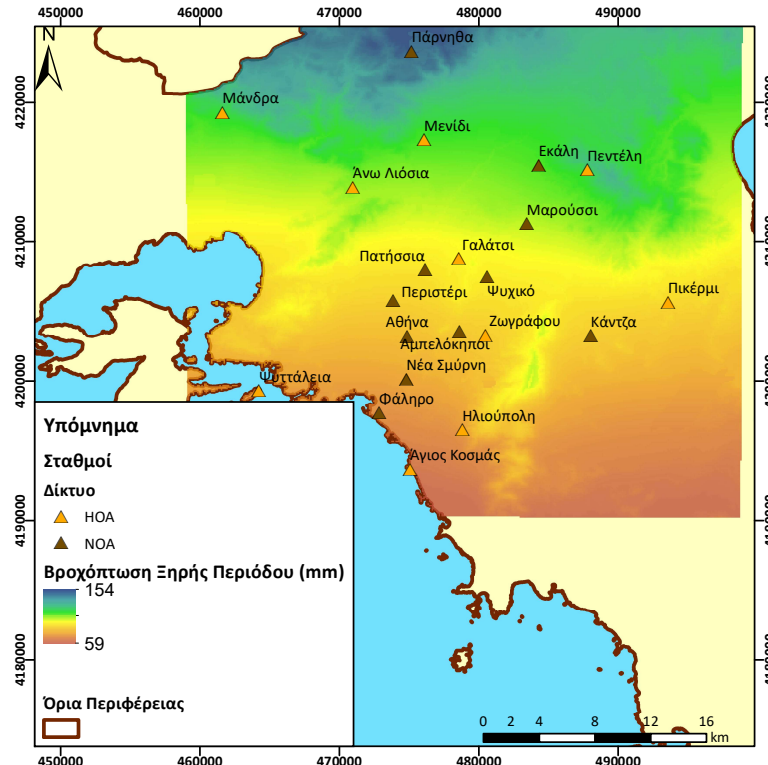
Στο τρισδιάστατο διάγραμμα που ακολουθεί η προσαρμογή του μοντέλου είναι αρκετά ικανοποιητική, όπως και η τιμή του προσαρμοσμένου συντελεστή προσδιορισμού (0.75). Η ικανοποιητική προσαρμογή του μοντέλου απεικονίζεται ομοίως στο διάγραμμα διασποράς. Τα θηκογράμματα παρουσιάζουν παρόμοια μορφή με αυτά του προηγούμενου μοντέλου (μηνός Δεκεμβρίου). Στο θηκογράμμα του δείγματος των τιμών πρόβλεψης απεικονίζεται ακραία τιμή (outlier), που δεν λαμβάνεται υπόψη. Το εύρος του δείγματος των πραγματικών τιμών είναι μεγαλύτερο από αυτό των εκτιμώμενων, ακόμη για τις εκτιμώμενες τιμές παρατηρείται και εδώ ήπια θετική ασυμμετρία. Στο ραβδόγραμμα οι σταθμοί του Γαλατσίου, της Ψυττάλειας και Ζωγράφου, σε φθίνουσα κατάταξη, απεικονίζονται με απόλυτη απόκλιση άνω του 10%. Ο σταθμός στο Περιστερί σημειώνει την μικρότερη απόκλιση της τάξεως του 0.4%.



Σχήμα 4-15 Τρισδιάστατο διάγραμμα μοντέλου παλινδρόμησης, διάγραμμα διασποράς, θηκόγραμματα δειγμάτων πραγματικών και εκτιμώνων τιμών, ραβδόγραμμα ποσοστιαίων αποκλίσεων πραγματικών και εκτιμώνων τιμών ανά σταθμό, για το ετήσιο μοντέλο (με σειρά κατάταξης από αριστερά προς τα δεξιά και από πάνω προς τα κάτω)

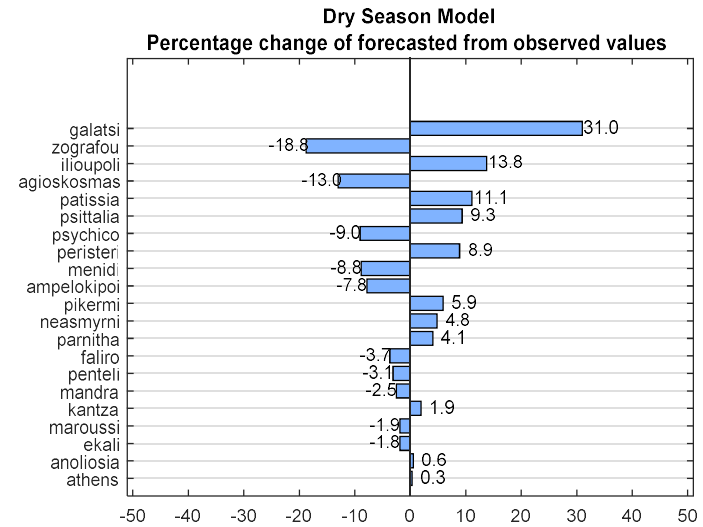
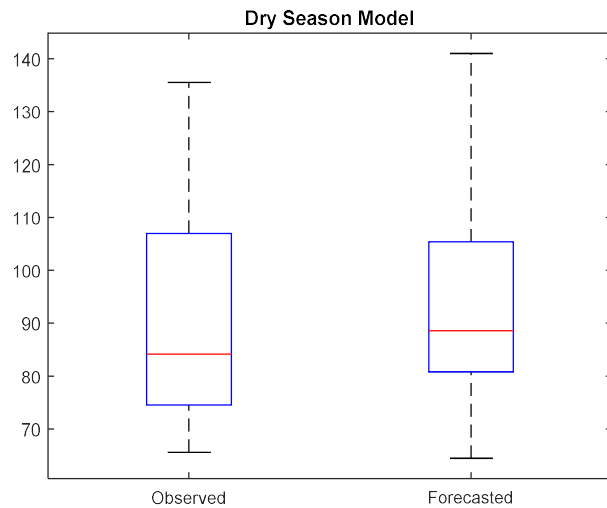
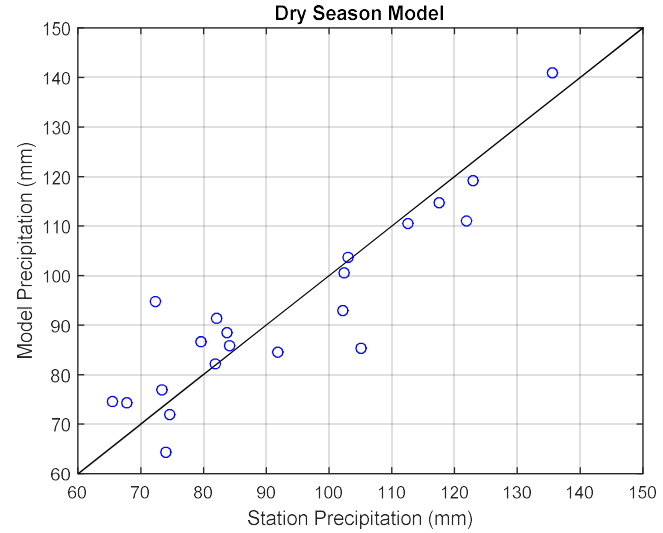
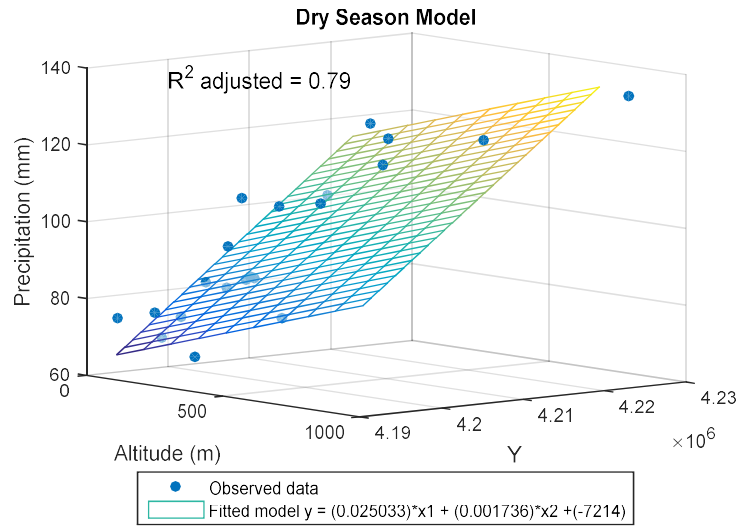
4.4.4 Μοντέλο Παλινδρόμησης Ξηρής Περιόδου

Ο επόμενος χάρτης (Σχήμα 4-16), παρουσιάζει χωρικά παρόμοια εικόνα με αυτόν του ετήσιου μοντέλου, με διαφορετικό εύρος, το οποίο κυμαίνεται περί των 60-155 mm βροχής. Στο μεγαλύτερο μέρος της έκτασης, της περιοχή μελέτης κυριαρχούν χαμηλά ύψη βροχόπτωσης, για το προαναφερθέν εύρος. Μέσα ύψη βροχόπτωσης (όπως στην περιοχή του Υμηττού) φαίνονται περισσότερο εξασθενημένα συγκριτικά με την εικόνα του ετήσιου μοντέλου. Οι υψηλότερες τιμές εντοπίζονται ξανά στους ορεινούς όγκους της Πάρνηθας και της Πεντέλης.



Σχήμα 4-16 Χάρτης χωρικά προβλεπόμενης βροχόπτωσης σε mm για την ξηρή περίοδο του έτους, σύμφωνα με το μοντέλο που αναπτύχθηκε, στο πολύγωνο περιοχής μελέτης

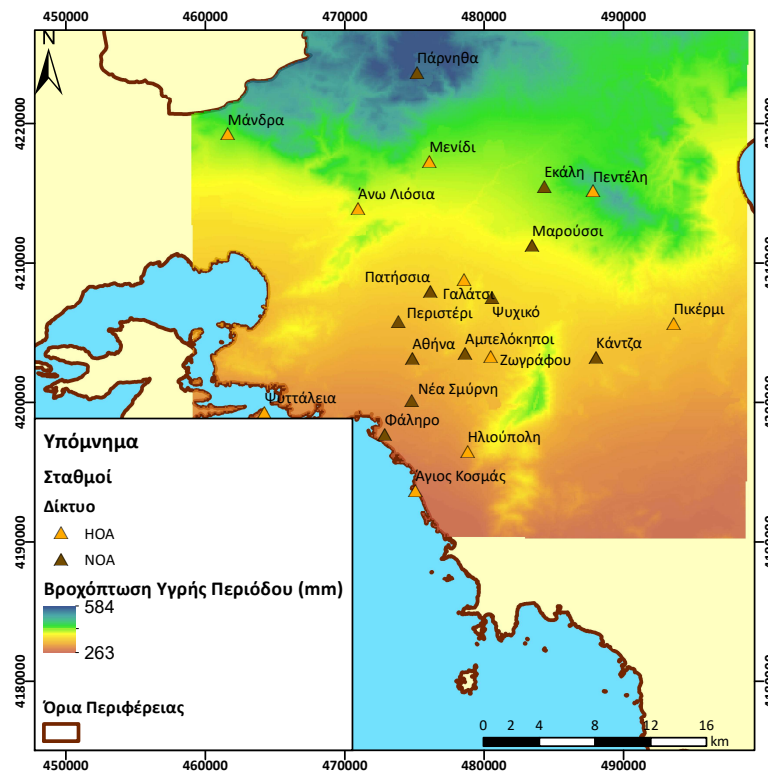
Παρά την πολύ καλή προσαρμογή του μοντέλου, όπως τεκμηριώνεται από το ακόλουθο τρισδιάστατο διάγραμμα καθώς και από την υψηλή τιμή του προσαρμοσμένου συντελεστή προσδιορισμού (0.79), στο διάγραμμα διασποράς εμφανίζεται μία ήπια συγκέντρωση τιμών στο εύρος των 0-90mm των αξόνων, η οποία δεν παρουσιάζει την αναμενόμενη γραμμική μορφή. Σε κανένα από τα δύο θηκογράμματα δεν παρουσιάζονται outliers, ενώ τα εύρη είναι παρόμοια με ελαφρώς μεγαλύτερο αυτό που αντιστοιχεί στο σύνολο των εκτιμώμενων τιμών. Και τα δύο σύνολα τιμών παρουσιάζουν θετική ασυμμετρία. Στο ακόλουθο ραβδόγραμμα εμφανίζονται σχετικά μεγαλύτερες αποκλίσεις από τα προηγούμενα μοντέλα. Ειδικότερα οι σταθμοί στο Γαλάτσι, Ζωγράφου, Ηλιούπολη, Αγ. Κοσμά και Πατήσια, παρουσιάζουν (φθίνουσα κατάταξη) απόλυτη απόκλιση άνω του 10%. Ο σταθμός της Αθήνας σημειώνει την μικρότερη απόκλιση της τάξεως του 0.3%.



Σχήμα 4-17 Τρισδιάστατο διάγραμμα μοντέλου παλινδρόμησης, διάγραμμα διασποράς, θηκόγραμμα δειγμάτων πραγματικών και εκτιμώμενων τιμών, ραβδόγραμμα ποσοστιαίων αποκλίσεων πραγματικών και εκτιμώμενων τιμών ανά σταθμό, για το μοντέλο ξηρής περιόδου του έτους (με σειρά κατάταξης από αριστερά προς τα δεξιά και από πάνω προς τα κάτω)

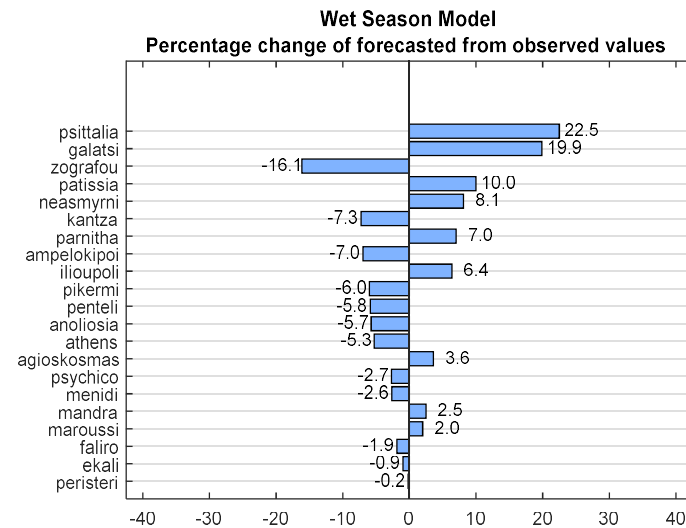
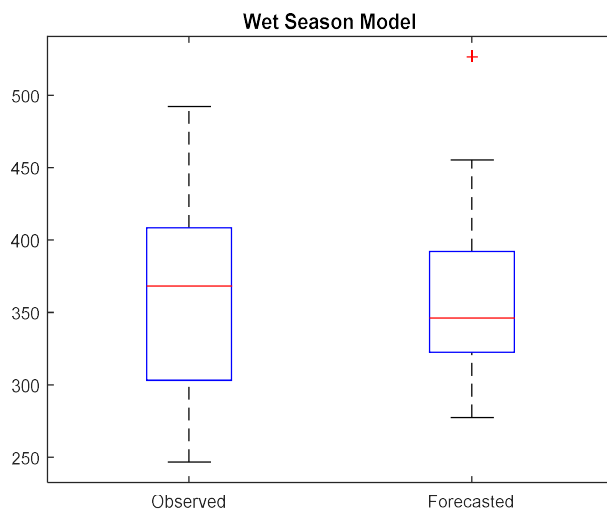
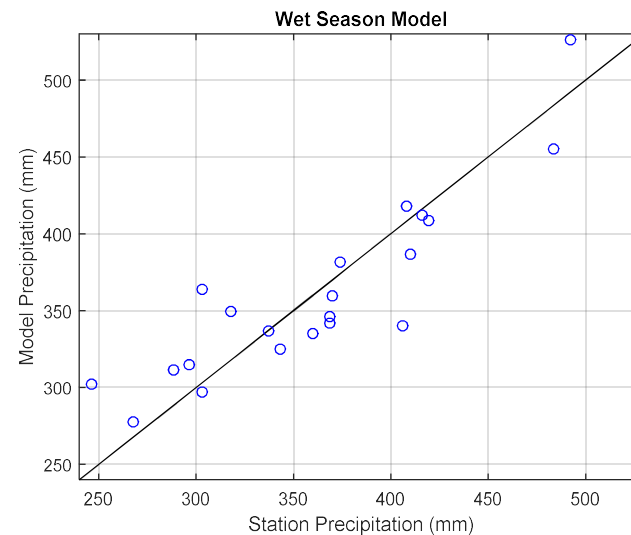
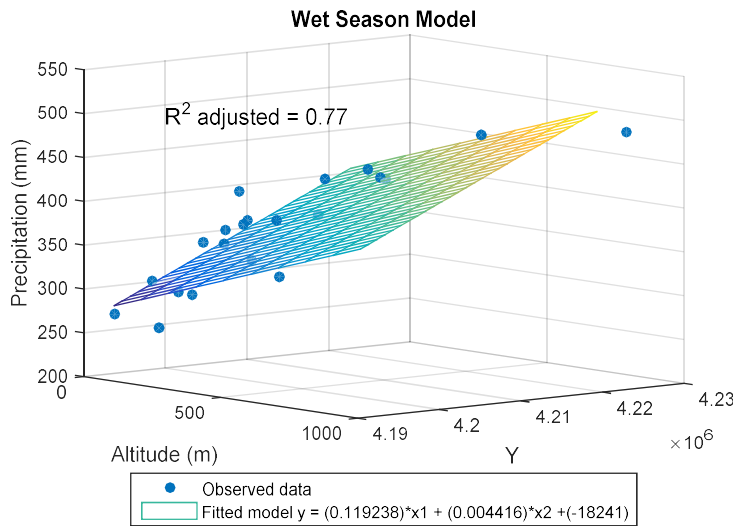
4.4.5 Μοντέλο Παλινδρόμησης Υγρής Περιόδου

Ο επόμενος χάρτης (Σχήμα 4-18), προσομοιάζει περισσότερο με αυτόν του ετήσιου μοντέλου, ποιοτικά, συγκριτικά με όλους τους προηγούμενους, ενώ παρουσιάζει εύρος, το οποίο κυμαίνεται περί των 260-585 mm βροχής. Στο μεγαλύτερο μέρος της έκτασης, της περιοχή μελέτης κυριαρχούν χαμηλά ύψη βροχόπτωσης, για το προαναφερθέν εύρος. Οι υψηλότερες τιμές περιορίζονται στην Πάρνηθα και στην Πεντέλη.



Σχήμα 4-18 Χάρτης χωρικά προβλεπόμενης βροχόπτωσης σε mm για την υγρή περίοδο του έτους, σύμφωνα με το μοντέλο που αναπτύχθηκε, στο πολύγωνο περιοχής μελέτης

Η ικανοποιητική προσαρμογή του μοντέλου διαπιστώνεται τόσο στο τρισδιάστατο γράφημα (με τιμή συντελεστή προσδιορισμού στο 0.77), όσο και στο γράφημα διασποράς. Ακραία τιμή (outlier) παρουσιάζεται στο θηκόγραμμα που αντιστοιχεί στο σύνολο των προβλεπόμενων από το μοντέλο τιμών, ενώ το εύρος των πραγματικών τιμών είναι μεγαλύτερο από το δεύτερο σύνολο τιμών. Η διάμεσος των πραγματικών τιμών εμφανίζεται υψηλότερα από αυτήν των εκτιμώμενων τιμών. Σύμφωνα με το ραβδόγραμμα απόλυτες αποκλίσεις μεγαλύτερες του 10% παρουσιάζουν οι εξής σταθμοί: Ψυττάλεια, Γαλάτσι και Ζωγράφου. Ο σταθμός στο Περιστερί εμφανίζει την χαμηλότερη απόκλιση στο -0.2%.



Σχήμα 4-19 Τρισδιάστατο διάγραμμα μοντέλου παλινδρόμησης, διάγραμμα διασποράς, θηκόγραμματα δειγμάτων πραγματικών και εκτιμώμενων τιμών, ραβδόγραμμα ποσοστιαίων αποκλίσεων πραγματικών και εκτιμώμενων τιμών ανά σταθμό, για το μοντέλο υγρής περιόδου του έτους (με σειρά κατάταξης από αριστερά προς τα δεξιά και από πάνω προς τα κάτω)

4.4.6 Συγκριτικά Αποτελέσματα

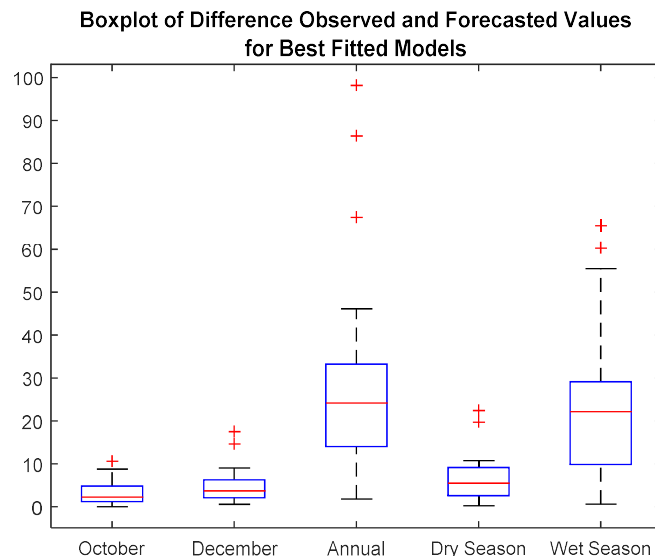
Στο παρόν υποκεφάλαιο εξετάζονται συγκριτικά, τα διάφορα μέτρα ακρίβειας και οι δείκτες αξιολόγησης, όπως υπολογίζονται για τα καλύτερα προσαρμοζόμενα μοντέλα που αναλύθηκαν προηγουμένως.

Στον επόμενο πίνακα παρατίθενται τα βαθμωτά μέτρα ακρίβειας μέσου σφάλματος (Mean (Bias) Error-ME), μέσου απόλυτου σφάλματος (Mean Absolute Error - MAE), μέσου τετραγωνικού σφάλματος (Mean Square Error - MSE), καθώς και η ρίζα αυτού (Root Mean Square Error - RMSE). Από τις τιμές του μεγέθους ME, όλα τα μοντέλα, με εξαίρεση το ετήσιο εμφανίζουν θετικές τιμές και συνεπώς φαίνεται να υπερεκτιμούν τις πραγματικές τιμές.

Πίνακας 4-4 Βαθμωτά μέτρα ακριβείας για τα μοντέλα καλύτερης προσαρμογής

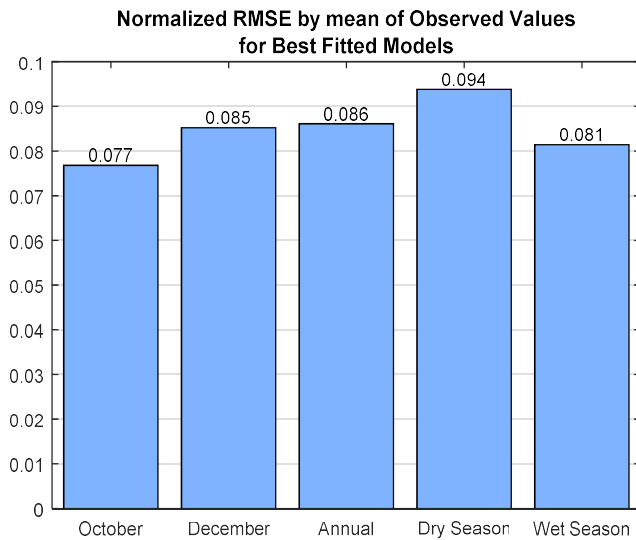
<i>Μοντέλο</i>	<i>ME</i>	<i>MAE</i>	<i>MSE</i>	<i>RMSE</i>
Οκτώβριος	0.000	3	18	4
Δεκέμβριος	0.008	5	44	7
Ετήσιο	-0.172	30	1505	39
Ξηρή Περίοδος	0.018	7	76	9
Υγρή Περίοδος	0.528	23	863	29

Όπως είναι εμφανές από τα εύρη των παραπάνω τιμών, η σύγκριση μεταξύ των μοντέλων, δεν είναι εφικτή. Αυτό είναι άλλωστε αναμενόμενο, εφόσον τα εξεταζόμενα μοντέλα αντιστοιχούν σε διαφορετικές χρονικές κλίμακες δεδομένων και επομένως σε άλλες τάξεις μεγέθους. Έτσι τα εύρη των αποκλίσεων μεταξύ πραγματικών και εκτιμώμενων από το εκάστοτε μοντέλο τιμών διαφέρουν σε τάξη μεγέθους, καθώς και οι διασπορές αυτών. Αυτό αποτυπώνεται γραφικά στο επόμενο θηκογράμματα του κάθε μοντέλου (Σχήμα 4-20).

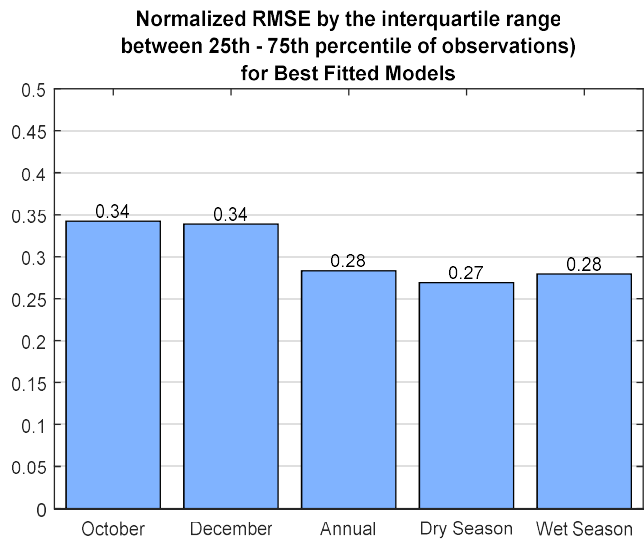


Σχήμα 4-20 Θηκογράμματα διαφορών μεταξύ πραγματικών και εκτιμώμενων τιμών, για τα καλύτερα προσαρμοζόμενα μοντέλα

Συνεπώς για την διευκόλυνση της σύγκρισης δεδομένων διαφορετικής κλίμακας οι τιμές των RMSE κανονικοποιούνται. Όπως ήδη έχει αναφερθεί στο κεφάλαιο της Μεθοδολογίας, η κανονικοποίηση γίνεται με διάφορες μεθόδους. Η συνηθέστερη αυτών χρησιμοποιεί τον μέσο όρο των πραγματικών μέσων. Οι κανονικοποιημένες τιμές RMSE για κάθε μοντέλο φαίνονται στο επόμενο ραβδόγραμμα (Σχήμα 4-21). Τιμές του NRMSE πλησιέστερα στο μηδέν (0) υποδεικνύουν καλύτερη «συγκέντρωση» των πραγματικών τιμών γύρω από την γραμμή του μοντέλου που αναπτύχθηκε. Γενικότερα τιμές του NRMSE (με κανονικοποίηση ως προς την μέση τιμή των παρατηρήσεων) που βρίσκονται κάτω του 0.1, όπως και εδώ συνιστούν αρκετά υψηλή – ικανοποιητική προσαρμογή.



Σχήμα 4-21 Κανονικοποιημένες RMSE τιμές, μέσω μέσου όρου των πραγματικών καταγραφών, για τα καλύτερα προσαρμοζόμενα μοντέλα

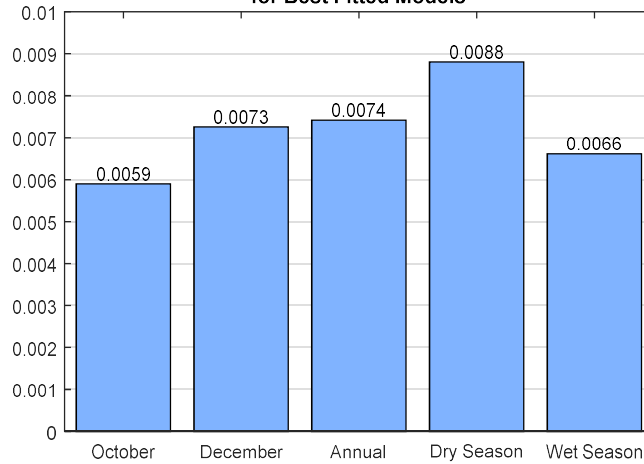


Σχήμα 4-22 Κανονικοποιημένες RMSE τιμές, μέσω του εύρους μεταξύ του 25 % έως 75 % των παρατηρήσεων

Ωστόσο για περιπτώσεις δεδομένων, όπου παρουσιάζονται ακραίες τιμές – outliers, κάτι που στην περίπτωση των εξεταζόμενων μοντέλων ισχύει, όπως αποτυπώνεται στα προηγούμενα θηκογράμματα των προηγούμενων υποκεφαλαίων, που αντιστοιχούν σε κάθε μοντέλο, προτιμάται η κανονικοποίηση μέσω της διαφοράς των πραγματικών τιμών που αντιστοιχούν στο 25% και στο 75% των παρατηρήσεων.

Ήδη, από το σχετικό τμήμα από το κεφάλαιο της Μεθοδολογίας υποδηλώνεται ότι μια χαμηλότερη τιμή του NRMSE δεν συνδέεται απαραίτητα με καλύτερα αποτελέσματα προσαρμογής και για αυτό τον λόγο στο επόμενο γράφημα αποτυπώνονται οι τιμές του κανονικοποιημένου NMSE, σύμφωνα με την σχέση (3.11), η οποία βελτιώνει το πιθανό πρόβλημα μεροληψίας του δείκτη ως προς μοντέλα που υποεκτιμούν συγκριτικά με τις πραγματικές τιμές.

Normalized MSE
by the product of mean of average actual and average estimated values
for Best Fitted Models



Σχήμα 4-23 Κανονικοποιημένες MSE τιμές, ως προς το γινόμενο μέσης τιμής πραγματικών και μέσης τιμής εκτιμώμενων αντίστοιχα τιμών, για τα καλύτερα προσαρμοζόμενα μοντέλα

Ανάλογα με το στατιστικό εργαλείο που χρησιμοποιείται κάθε φορά, για την κανονικοποίηση των μεγεθών RMSE ή MSE, τα αποτελέσματα και συνεπώς η αξιολόγηση – κατάταξη των μοντέλων μπορεί να διαφέρει σημαντικά. Είναι ωστόσο σαφές και από τα τρία διαγράμματα ότι οι δείκτες διαφέρουν σε σχετικά μικρό βαθμό μεταξύ τους.

Περισσότερο αξιόλογοι και συχνά εφαρμοζόμενοι σε παρόμοια επιστημονικά πεδία, με τα οποία καταπιάνεται η παρούσα διπλωματική εργασία, είναι οι δείκτες αξιολόγησης POD, FAR και CSI, οι τιμές των οποίων προκύπτουν από τους πίνακες με την διαμόρφωση που υποδείχθηκε στο υπό-κεφάλαιο 3.2.6. Ακολουθώς, παρατίθενται οι σχετικοί πίνακες για κάθε ένα από τα πέντε στατιστικά σημαντικά μοντέλα.

Πίνακας 4-5 Τιμές των a,b,c για τα στατιστικά σημαντικά μοντέλα υπό μορφή κατηγοριοποίησης μεταβλητών

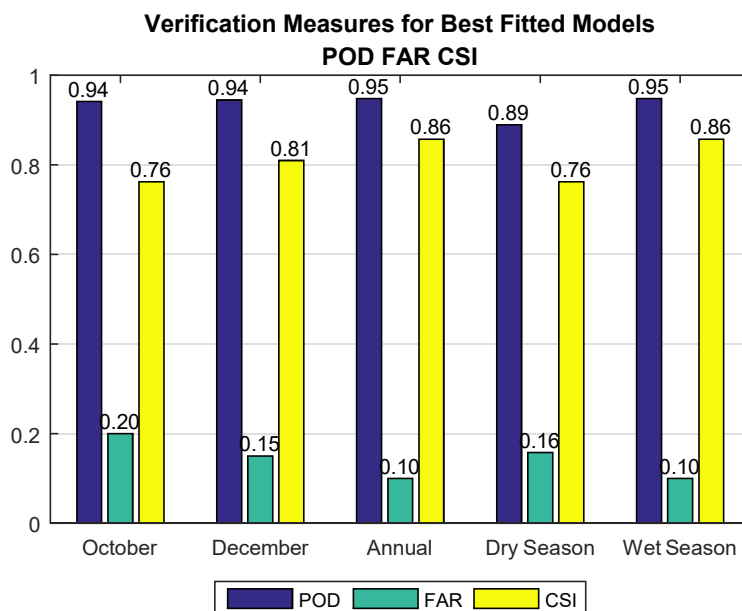
<i>2x2 Contingency Table</i> October		Observed	
		Yes	No
Forecast	Yes	16	4
	No	1	

<i>2x2 Contingency Table</i> December		Observed	
		Yes	No
Forecast	Yes	17	3
	No	1	

<i>2x2 Contingency Table</i> Annual		Observed	
		Yes	No
Forecast	Yes	18	2
	No	1	

<i>2x2 Contingency Table</i> Dry Season		Observed	
		Yes	No
Forecast	Yes	16	3
	No	2	

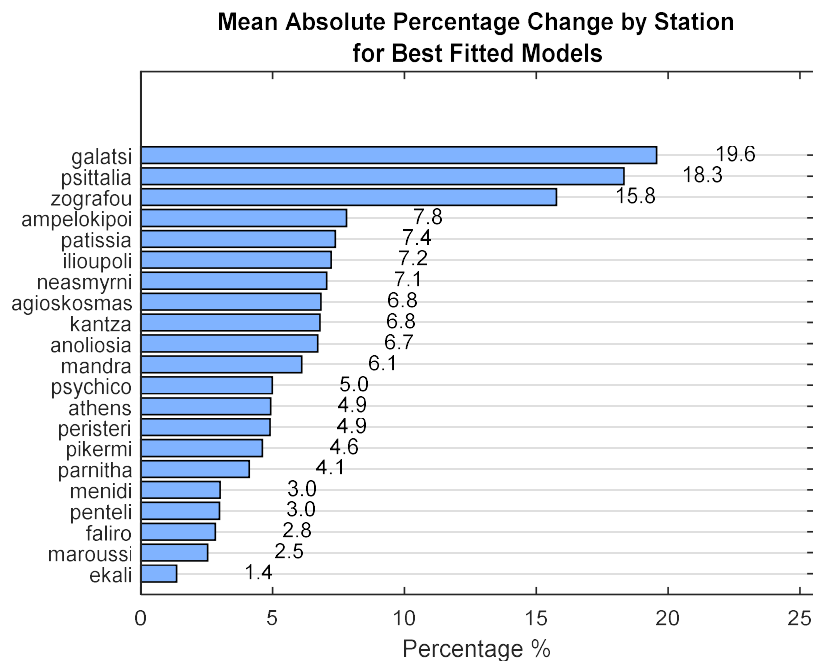
<i>2x2 Contingency Table</i>		Observed	
		Yes	No
Forecast	Wet Season	18	2
	No	1	



Σχήμα 4-24 Δείκτες αξιολόγησης καλύτερα προσαρμοσμένων μοντέλων POD, FAR, CSI

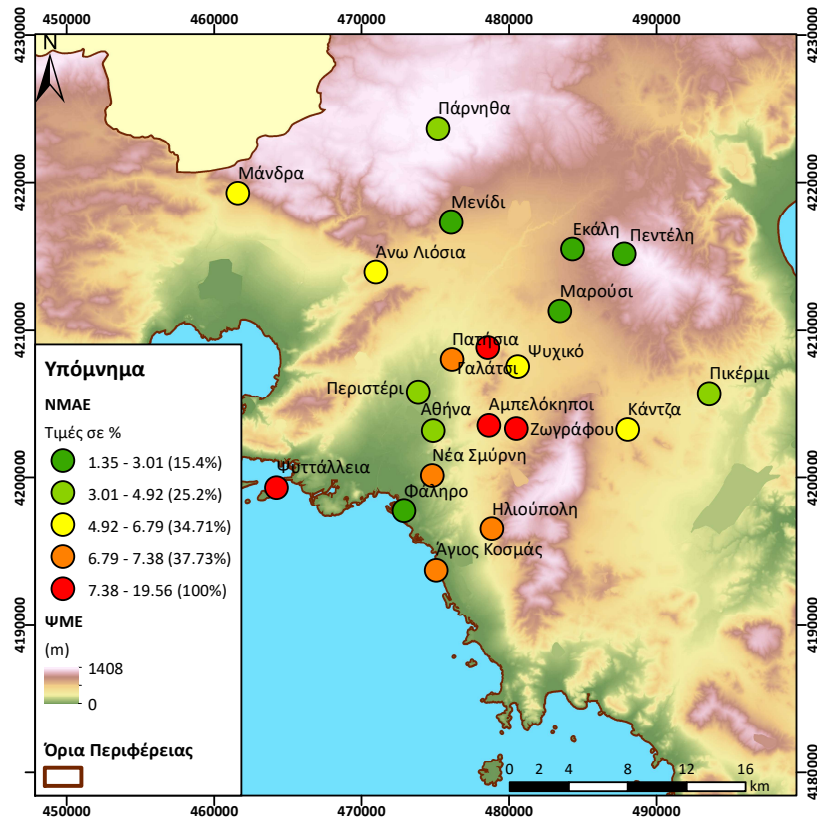
Υπενθυμίζεται από την ανάλυση αυτών των δεικτών στο κεφάλαιο της Μεθολογίας, ότι όσο πλησιέστερα ο δείκτης POD και ο CSI βρίσκονται στην μονάδα, τόσο πιο αξιόπιστο είναι το μοντέλο. Το αντίστροφο ισχύει για τον δείκτη FAR. Η υψηλότερη τιμή για τον δείκτη POD, είναι το 0.95 και για τον δείκτη CSI η 0.86, οι οποίες ανήκουν στο ετήσιο μοντέλο και σε αυτό της υγρής περιόδου. Ενώ η χαμηλότερη τιμή για τον δείκτη FAR, είναι το 0.10 και ανήκει επίσης στο ετήσιο και στο μοντέλο υγρής περιόδου.

Πέραν της συγκριτικής ανάλυσης στατιστικών δεικτών μεταξύ μοντέλων θεωρήθηκε σημαντικό να διερευνηθεί κατά πόσο ικανοποιητικά προσαρμόζονται σε χωρικό επίπεδο, συνολικά όλα τα μοντέλα πολλαπλής γραμμικής παλινδρόμησης με ίδιες ανεξάρτητες μεταβλητές. Έτσι υπολογίστηκε όπως απεικονίζεται στο επόμενο ραβδόγραμμα (Σχήμα 4-25), η μέση ποσοστιαία διαφορά (κατά απόλυτη τιμή) μεταξύ εκτιμώμενων και πραγματικών τιμών για κάθε σταθμό, λαμβάνοντας υπόψη το σύνολο των στατιστικά σημαντικών μοντέλων.



Σχήμα 4-25 Μέσο ποσοστό διαφοράς, κατά απόλυτη τιμή, ανά σταθμό, λαμβάνοντας υπόψη μόνο τα στατιστικά σημαντικά μοντέλα

Τα υψηλότερα ποσοστά και ειδικότερα αυτά που σημειώνουν ποσοστό απόκλισης άνω του 10%, ανήκουν στους σταθμούς του Γαλατσίου, της Ψυττάλειας και του Ζωγράφου. Ενώ το χαμηλότερο ποσοστό ανήκει στον σταθμό της Εκάλης με μόλις μέσο ποσοστό διαφοράς 1.4%. Συνολικά οι τιμές των ποσοστών είναι αρκετά ικανοποιητική, εφόσον 10 στους 21 σταθμούς (~48%) βρίσκονται σε επίπεδο ίσο ή μικρότερο του 5% μέσης απόλυτης απόκλισης. Για την καλύτερη χωρική κατανομή αυτών των ποσοστών παρατίθεται ο επόμενος χάρτης (Σχήμα 4-26), με τα ποσοστά που αναφέρθηκαν ανά σταθμό, κατανεμημένα σε κλάσεις.



Σχήμα 4-26 Χάρτης ποσοστιαίων απόλυτων αποκλίσεων (NMAE) για κάθε σταθμό, της περιοχής μελέτης¹²

Εξετάζοντας τον παραπάνω χάρτη είναι εμφανές ότι σταθμοί βορειότεροι, οι οποίοι παράλληλα βρίσκονται σε υψηλό υψόμετρο εδάφους, ανταποκρίνονται καλύτερα καθώς καταγράφουν τα μικρότερα ποσοστά απόλυτων αποκλίσεων. Πλησιέστερα προς την ακτογραμμή με εξαίρεση τον σταθμό του Φαλήρου σημειώνονται μεγαλύτερα ποσοστά.

Οι προαναφερθείσες τιμές αποκλίσεων υπολογίστηκαν κατά απόλυτη τιμή και επομένως οφείλει να εξεταστεί κατά πόσο τα στατιστικά σημαντικά μοντέλα που εξετάζονται, υπερεκτιμούν ή υποεκτιμούν αντίστοιχα τις πραγματικές παρατηρήσεις ανά σταθμό.

Στον επόμενο πίνακα (Πίνακας 4-6) παρατίθενται ανά σταθμό οι μέσες απόλυτες διαφορές, σε ποσοστιαία μορφή και αντίστοιχα οι φορές που κάθε ένα από τα εκάστοτε στατιστικά σημαντικά εξεταζόμενα μοντέλα (5) υπερεκτιμούν ή υποεκτιμούν την πραγματική βροχόπτωση που καταγράφηκε. Συνολικά από τους 21 σταθμούς, κατά κύριο λόγο, οι 9 υπερεκτιμούν την βροχόπτωση και 12 την υποεκτιμούν.

¹² τα όρια των κλάσεων αναφέρονται σε ποσοστά (%) σφαλμάτων NMAE, η τιμή δίπλα σε παρένθεση δίπλα από τα όρια κάθε κλάσης, αναφέρεται σε ποσοστημόριο των τιμών που βρίσκεται μέχρι και εντός αυτής της κλάσης

Πίνακας 4-6 Μέση απόλυτη ποσοστιαία διαφορά μεταξύ παρατηρημένων και εκτιμώμενων τιμών βροχόπτωσης, από τα στατιστικά σημαντικά μοντέλα, για κάθε σταθμό – Φορές Υπερεκτίμησης και φορές υποεκτίμησης τιμών συγκριτικά με τις πραγματικές

Σταθμοί	Μέση Απόλυτη Διαφορά (%)	Φορές Υπερεκτίμησης	Φορές Υποεκτίμησης
Αμπελόκηποι	7.80	0	5
Αθήνα	4.92	1	4
Εκάλη	1.35	1	4
Φάληρο	2.83	0	5
Κάντζα	6.79	2	3
Μαρούσι	2.53	4	1
Νέα Σμύρνη	7.05	5	0
Πάρνηθα	4.11	4	1
Πατήσια	7.38	4	1
Περιστέρι	4.90	2	3
Ψυχικό	4.99	0	5
Άγιος Κοσμάς	6.83	3	2
Άνω Λιόσια	6.71	1	4
Γαλάτσι	19.56	5	0
Ηλιούπολη	7.22	5	0
Μάνδρα	6.10	4	1
Μενίδι	3.01	1	4
Πεντέλη	2.98	1	4
Πικέρμι	4.61	2	3
Ψυττάλεια	18.32	5	0
Ζωγράφου	15.76	0	5

Πίνακας 4-7 Μέση απόλυτη ποσοστιαία διαφορά μεταξύ παρατηρημένων και εξαγόμενων από όλα τα μοντέλα τιμών βροχόπτωσης, ανά σταθμό – αύξουσα ταξινόμηση τιμών, φορές

Σταθμοί	Μέση Απόλυτη Διαφορά (%)	Φορές Υπερεκτίμησης	Φορές Υποεκτίμησης
Ψυχικό	4.94	3	12
Μενίδι	5.17	4	11
Μάνδρα	6.21	9	6
Μαρούσσι	6.40	8	7
Νέα Σμύρνη	6.58	13	2
Άνω Λιόσια	7.05	6	9
Εκάλη	7.87	6	9
Πάρνηθα	9.50	12	3
Ηλιούπολη	10.00	12	3
Πατήσια	10.69	14	1
Άγιος Κοσμάς	10.83	7	8
Φάληρο	11.61	4	11
Πεντέλη	11.80	5	10
Κάντζα	12.03	7	8
Περιστέρι	12.57	12	3
Αμπελόκηποι	13.31	2	13
Αθήνα	14.12	4	11
Ζωγράφου	15.72	1	14
Πικέρμι	21.60	6	9
Γαλάτσι	21.88	14	1
Ψυττάλεια	24.08	13	2

5 ΣΥΜΠΕΡΑΣΜΑΤΑ – ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ

5.1 Σύνοψη

Η παρούσα μεταπτυχιακή εργασία αποτελεί μία πρώτη προσπάθεια διαμόρφωσης και εφαρμογής ενός μεθοδολογικού πλαισίου για την ανάπτυξη μοντέλων χωροχρονικής εκτίμησης τιμών βροχόπτωσης¹³, στην περιοχή μελέτης του λεκανοπεδίου της Αττικής.

Η επιλογή των βροχομετρικών σταθμών προς εξέταση έγινε με κριτήριο την διαθεσιμότητα, την ποιότητα των δεδομένων και την ομοιόμορφη χωρική κατανομή εντός και πλησίον της περιοχής μελέτης του λεκανοπεδίου. Για τον σκοπό αυτό χρησιμοποιήθηκαν βροχομετρικά δεδομένα σταθμών του Υδρολογικού Παρατηρητήριου Αθήνας (δίκτυο ΗΟΑ) και του Εθνικού Αστεροσκοπείου (δίκτυο ΝΟΑ).

Τα διαθέσιμα πρωτογενή δεδομένα εισάχθηκαν, υπέστησαν επεξεργασία και ομαδοποιήθηκαν καταλλήλως, μέσω της σύνταξης και εκτέλεσης κώδικα στο προγραμματιστικό περιβάλλον MATLAB. Παράλληλα η απαραίτητη συμπλήρωση χρονοσειρών έγινε μέσω της χρήσης συστήματος γεωγραφικών πληροφοριών (GIS) και πιο συγκεκριμένα μέσω του λογισμικού ArcGIS.

Μέσω της ολοκληρωμένης βιβλιογραφικής επισκόπησης που προηγήθηκε, καθορίστηκαν οι προς εξέταση ανεξάρτητες μεταβλητές που δυνητικά συσχετίζονται με το μέγεθος της βροχόπτωσης και στη συνέχεια ακολουθείται αλγόριθμος για την εύρεση του βέλτιστου μοντέλου πολλαπλής γραμμικής παλινδρόμησης.

Από τα εξεταζόμενα μοντέλα επιλέγονται τα στατιστικώς σημαντικά τα οποία και τελικώς αξιολογούνται, βάσει στατιστικών δεικτών και διαγραμμάτων, ενώ παράγονται τα αντίστοιχα γεωχωρικά επίπεδα πρόβλεψης βροχόπτωσης για την περιοχή μελέτης.

5.2 Συμπεράσματα

Τα κύρια συμπεράσματα της εργασίας συνοψίζονται στα εξής:

1. Είναι εμφανές ότι μεταβλητές της κλίσης εδάφους, προσανατολισμού και τετμημένης X απορρίπτονται από το πρώτο κιάλο στάδιο διαμόρφωσης του μοντέλου, καθώς οι αντίστοιχοι συντελεστές συσχέτισης με την βροχόπτωση είναι ιδιαίτερα χαμηλοί. Υψηλές συσχετίσεις παρουσιάζουν – σε φθίνουσα κατάταξη – οι μεταβλητές της γεωγραφικής τεταγμένης Y , του υψομέτρου και της απόστασης από την ακτογραμμή.
2. Τα αποτελέσματα και κατά συνέπεια τα συμπεράσματα που εξάγονται χαρακτηρίζονται από χωρική και στατιστική αβεβαιότητα, δεδομένου του μικρού δείγματος των διαθέσιμων σταθμών, καθώς και της μικρής περιόδου (ίσης ή μικρότερης της δεκαετίας)

¹³ Μηνιαίας, εποχικής, ετήσιας

καταγεγραμμένων τιμών βροχόπτωσης, ενώ συνηθέστερα για την μελέτη τέτοιων μοντέλων απαιτείται κλιματολογική περίοδος 30 ετών.

3. Συγκρίνοντας τα ομβροθερμικά διαγράμματα του Ελληνικού και της Νέας Φιλαδέλφειας, όπως αυτά διαμορφώθηκαν από τα ανοιχτά δεδομένα της ΕΜΥ διάρκειας 55 ετών, με τη μέση μηνιαία βροχόπτωση δεκαετίας των επιλεχθέντων σταθμών παρατηρούνται μη αναμενόμενα υψηλά επίπεδα βροχής για τον μήνα Φεβρουάριο. Ωστόσο οι λοιπές μηνιαίες τιμές κυμαίνονται σε περισσότερο λογικά επίπεδα. Αυτό πιθανότατα οφείλεται στο μικρό χρονικό εύρος των διαθέσιμων χρονοσειρών (10 έτη) και στο γεγονός ότι η συμπλήρωση των ελλειπών τιμών χρονοσειράς (για τους μήνες του Φεβρουαρίου) έγινε βάσει μόνο σταθμών με υψηλές καταγεγραμμένες βροχοπτώσεις που επηρέασε τις εξαγόμενες τιμές.
4. Λόγω του μικρού αριθμού σταθμών με πλήρη διαθέσιμη χρονοσειρά, για την συμπλήρωση των λοιπών σταθμών, δεν ήταν εφικτή πάντα η διαμόρφωση ενός δικτύου σταθμών (με ήδη καταγεγραμμένες τιμές) καλά χωρικά κατανομημένων. Αυτό είχε σαν αποτέλεσμα την μεροληπτική συμπλήρωση των σταθμών που βρίσκονταν σε μεγάλη απόσταση από αυτούς με τις καταγραφές.
5. Η διαδικασία της συμπλήρωσης είναι πιο αξιόπιστη για την περίπτωση των ενδιάμεσων κενών στις χρονοσειρές, επειδή συνηθέστερα για αυτή την περίπτωση, οι περισσότεροι από τους σταθμούς είχαν για το δεδομένο χρονικό σημείο καταγεγραμμένη τιμή και επομένως η συμπλήρωση γίνεται βάσει περισσότερων τιμών. Αντίθετα για τα πρώτα έτη της δεκαετίας που εξετάζεται, οι σταθμοί του δικτύου ΝΟΑ δεν παρουσιάζουν καταγραφές. Κάτι που οδηγεί αναπόφευκτα σε περισσότερο αβέβαια αποτελέσματα, εφόσον λαμβάνονται υπόψη μόνο οι τιμές των σταθμών του δικτύου ΗΟΑ (10 σταθμοί).
6. Από την ανάλυση των διαθέσιμων βροχομετρικών δεδομένων καθώς και από την χωρική κατανομή τους, παρατηρούνται υψηλές διασπορές βροχόπτωσης με μεγάλες χωρικές διαφορές από το νότιο στο βόρειο λεκανοπέδιο, δεδομένης της μικρής σχετικά γεωγραφικής έκτασης που μελετάται.
7. Υψηλή συσχέτιση παρατηρείται μεταξύ των ανεξάρτητων μεταβλητών της γεωγραφικής τεταγμένης Y και του υψομέτρου, αυτό στατιστικά θα απέτρεπε την χρήση των δύο αυτών μεταβλητών στο μοντέλο. Ωστόσο εννοιολογικά οι συγκεκριμένες μεταβλητές δεν συνδέονται, αλλά η συσχέτιση εμφανίζεται πλασματικά για το λεκανοπέδιο της Αττικής, εφόσον οι βασικοί ορεινοί όγκοι της βρίσκονται στο βορειότερο τμήμα της περιοχής μελέτης.
8. Από τα μοντέλα πολλαπλής γραμμικής παλινδρόμησης που εξάχθηκαν, οι χαμηλής επίδοσης προσαρμογές παρατηρήθηκαν για τους μήνες με τα χαμηλότερα ύψη βροχόπτωσης (Ιούλιος, Αύγουστος, Σεπτέμβρης). Μάλιστα για τους συγκεκριμένους μήνες οι συντελεστές που αντιστοιχούν στην ανεξάρτητη μεταβλητή του ύψους

προκύπτουν αρνητικοί. Αυτό παρατηρείται συγκριτικά και για τα μοντέλα μεγαλύτερων χρονικών αποκρίσεων. Ειδικότερα οι στατιστικοί δείκτες προσαρμογής εμφανίζονται υψηλότεροι για το ετήσιο μοντέλο και για αυτό της υγρής περιόδου, συγκριτικά με το μοντέλο που αναφέρεται στην ξηρή περίοδο του έτους.

9. Ανάλογα αποτελέσματα προσαρμογής διαπιστώνονται και σε χωρικό επίπεδο. Επεξηγηματικότερα, όπως φαίνεται από τα ραβδογράμματα ποσοστιαίας απόκλισης εκτιμώμενων και παρατηρούμενων τιμών ανά μοντέλο, κατά κύριο λόγο σταθμοί οι οποίοι κατέγραφαν τα μεγαλύτερα ύψη βροχόπτωσης σημείωναν και τις μικρότερες αποκλίσεις.
10. Παρατηρείται επίσης ότι για τις περιπτώσεις των μοντέλων, όπου η χρονική κλίμακα είναι μεγαλύτερη της μηνιαίας (ετήσιο μοντέλο και μοντέλα υγρής και ξηρής περιόδου), η προσαρμογή των μοντέλων είναι αξιοσημείωτα βελτιωμένη συγκριτικά με τα μοντέλα μηνιαία κλίμακας (με εξαίρεση αυτά των μηνών του Οκτωβρίου και του Δεκεμβρίου που είναι επίσης ικανοποιητική). Το γεγονός αυτό είναι αναμενόμενο, καθώς και τα ύψη των βροχοπτώσεων είναι υψηλότερα, ενώ παράλληλα η ομαδοποίηση των δεδομένων σε μεγαλύτερες χρονικές κλίμακες εξασφαλίζει την μείωση της απότομης διακύμανσης των δεδομένων της χρονοσειράς (σ.σ., απότομη αύξηση, αποκαλούμενη και spike χρονοσειράς).

5.3 Μελλοντική Έρευνα

Βάσει των προηγούμενων συμπερασμάτων τα παρακάτω μπορούν να αποτελέσουν θέματα προς μελλοντική διερεύνηση:

1. Προτείνεται η εφαρμογή του παρόντος μεθοδολογικού πλαισίου να γίνει με τη χρήση βροχομετρικών δεδομένων από ένα πιο πυκνό δίκτυο σταθμών, το οποίο να καλύπτει επαρκώς όλο το εύρος του ΥΔ Αττικής. Παράλληλα κρίνεται αναγκαίο για τους υπό αξιοποίηση σταθμούς το εύρος των χρονοσειρών να πλησιάζει την περίοδο των 30 ετών, όπως ορίζεται μία ολοκληρωμένη κλιματική περίοδο αναφοράς. Το συγκεκριμένο χρονικό εύρος, σύμφωνα με την σχετική επιστημονική βιβλιογραφία, είναι το ελάχιστο για να μπορεί μια περίοδος να θεωρηθεί επαρκής και ασφαλής, για τον ορθό χαρακτηρισμό του κλίματος μίας περιοχής.
2. Δεδομένης της διαθεσιμότητας ενός πιο πυκνού δικτύου σταθμών, η συγκεκριμένη μεθοδολογία μπορεί να εκτελεσθεί σε μικρότερο χωρικό φατνίο και να οδηγήσει σε αποτελέσματα που θα χαρακτηρίζονται από μεγαλύτερη χωρική ανάλυση. Ταυτόχρονα, αυτά τα μελλοντικά αποτελέσματα μπορούν να συγκριθούν με αυτά που προκύπτουν από την παρούσα εργασία (με φατνίο 100m x 100m).
3. Κρίνεται σκόπιμο η ίδια ή παρόμοια μεθοδολογία ανάπτυξης μοντέλων παλινδρόμησης να διευρυνθεί για όλη την επικράτεια της χώρας, καθώς τα αποτελέσματα της παρούσας

μελέτης δεν είναι αντιπροσωπευτικά για ολόκληρη την έκταση της Ελλάδας. Η μεγάλη ανομοιογένεια του κλίματος της Ελλάδας και των ταυτόχρονα επαναλαμβανόμενων επεισοδίων ακραίων και έντονων βροχοπτώσεων, με μεγαλύτερη συχνότητα ιδιαίτερα τα τελευταία έτη λόγω της κλιματικής αλλαγής, το καθιστά ως ένα θέμα ιδιαίτερου ενδιαφέροντος και παράλληλα άκρως απαιτητικό.

4. Ακόμη προτείνεται η εξέταση άλλων μη γραμμικών μοντέλων παλινδρόμησης (που θα αντληθούν από σχετική βιβλιογραφία) και την σύγκριση των εξαγόμενων αποτελεσμάτων με αυτά των γραμμικών μοντέλων. Συγχρόνως μπορούν εξετασθούν περισσότερες ανεξάρτητες μεταβλητές και η συσχέτιση τους με το μέγεθος της βροχόπτωσης, όπως της προσπίπτουσας ηλιακής ακτινοβολίας, της θερμοκρασίας, την ένταση και την κατεύθυνση του ανέμου. Επιπλέον ιδιαίτερα ενδιαφέρουσα θα ήταν και η προσπάθεια ανάπτυξης μοντέλων παλινδρόμησης με αποκρίσεις ημερήσιου χρονικού βήματος. Για το τελευταίο αναγκαία προϋπόθεση είναι η διαθεσιμότητα χρονοσειρών τριαντακονταετίας και άνω.
5. Παράλληλα, συνίσταται η διενέργεια συγκριτικής ανάλυσης για τον εντοπισμό βέλτιστης μεθόδου συμπλήρωσης ελλιπών βροχομετρικών χρονοσειρών, για την ίδια περιοχή μελέτης. Η μεθοδολογία Kriging που εφαρμόστηκε στην παρούσα εργασία είναι ιδιαίτερα δημοφιλής για τέτοιου είδους προβλήματα που εμπεριέχουν μάλιστα τον χωρικό παράγοντα, ωστόσο θα ήταν ιδιαίτερα χρήσιμο, τα αποτελέσματα αυτής της μεθόδου να συγκριθούν με απλούστερες μεθοδολογίες (π.χ. συμπλήρωση μέσω ελαχίστων τετραγώνων με γειτονικούς σταθμούς που παρουσιάζουν υψηλή συσχέτιση με τον σταθμό προς συμπλήρωση) ή πιο σύνθετες, που βασίζονται σε στοχαστικά μοντέλα και αλγορίθμους που αντλούνται από την θεωρία νευρωνικών δικτύων.
6. Τέλος, η επαλήθευση δεδομένων ή/και αποτελεσμάτων της παρούσας μεθοδολογίας μπορεί να πραγματοποιηθεί για τις ίδιες χωρικές εκτάσεις και χρονικές περιόδους, με τηλεμετρικά δεδομένα βροχόπτωσης, έμμεσα μέσω της ανακλαστικότητας Radar ή από αξιόπιστα δορυφορικά προϊόντα βροχόπτωσης.

6 ΒΙΒΛΙΟΓΡΑΦΙΑ

Ξενόγλωσση

- Baltas, E. A., Panagos, D. S., & Mimikou, M. A. (2015). An approach for the estimation of hydrometeorological variables towards the determination of ZR coefficients. *Environmental Processes*, 2(4), 751-759.
- Bagnouls, F., & Gaussen, H. (1957, May). Les climats biologiques et leur classification. In *Annales de géographie* (Vol. 66, No. 355, pp. 193-220). Armand Colin.
- Earls, J.; Dixon, B. Spatial Interpolation of Rainfall Data Using ArcGIS: A Comparative Study. In *Proceedings of the Proceedings of the 27th Annual ESRI International User Conference, San Diego, CA, USA, 4–8 August 2007*; Volume 31, pp. 1–9.
- Feloni, E., Mousadis, I., & Baltas, E. (2020). Flood vulnerability assessment using a GIS-based multi-criteria approach—The case of Attica region. *Journal of Flood Risk Management*, 13, e12563.
- Gareth, J., Daniela, W., Trevor, H., & Robert, T. (2013). *An introduction to statistical learning: with applications in R*. Spinger.
- Georganta, C. (2022). Rainfall Intensity-Duration curves, for the assessment of flood risk in the Attica region.
- Gofa, F., Mamara, A., Anadranistakis, M., & Flocas, H. (2019). Developing gridded climate data sets of precipitation for Greece based on homogenized time series. *Climate*, 7(5), 68.
- Gournelos, T., & Maroukian, H. (1990). Geomorphological observations concerning the evolution of the Basin of Athens.
- Gouvas, M., Sakellariou, N., & Xystrakis, F. (2009). The relationship between altitude of meteorological stations and average monthly and annual precipitation. *Studia Geophysica et Geodaetica*, 53(4), 557-570.
- Grammatikogiannis, A., Mamassis, N., Baltas, E., & Mimikou, M. (2005, September). A meteorological telemetric network for monitoring of the Athens wider area (METEONET). A real time approach from point to areal measurements. In *Proceedings of the Ninth International Conference on Environmental Science and Technology (9CEST), Rhodes Island, Greece* (pp. 1-3).

- Hanna, S. R., & Heinold, D. W. (1985). *Development and application of a simple method for evaluating air quality models* (No. 4409). American Petroleum Institute.
- Hession, S. L., & Moore, N. (2011). A spatial regression analysis of the influence of topography on monthly rainfall in East Africa. *International journal of climatology*, 31(10), 1440-1456.
- Jolliffe, I. T., & Stephenson, D. B. (2003). *Forecast Verification: A Practitioner's Guide in Atmospheric Science*.
- Lagouvardos, K., Kotroni, V., Bezes, A., Koletsis, I., Kopania, T., Lykoudis, S., Mazarakis, N., Papagiannaki, K. and Vougioukas, S. (2017). The automatic weather stations NOANN network of the National Observatory of Athens: Operation and database. *Geoscience Data Journal*, 4(1), 4-16.
- Maheras, P., & Anagnostopoulou, C. (2003). Circulation types and their influence on the interannual variability and precipitation changes in Greece. In *Mediterranean climate* (pp. 215-239). Springer, Berlin, Heidelberg.
- Maroukian, V., Sermegas, E., Gaki-Papanastassiou, J., & Karymbalis, E. (2005). The role of morphometric parameters and human interference in floods at the lower reaches of Kifisos River delta (Attica). In *Proceedings of the 7th Panhellenic congress on hydrogeology* (Vol. 1, pp. 289-299).
- Marquínez, J., Lastra, J., & García, P. (2003). Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis. *Journal of hydrology*, 270(1-2), 1-11.
- Mehdizadeh, S. (2020). Using AR, MA, and ARMA time series models to improve the performance of MARS and KNN approaches in monthly precipitation modeling under limited climatic data. *Water Resources Management*, 34(1), 263-282.
- Mentaschi, L., Besio, G., Cassola, F., & Mazzino, A. (2013, April). Why NRMSE is not completely reliable for forecast/hindcast model test performances. In *Geophysical Research Abstracts* (Vol. 15).
- Nastos, P. T., Kapsomenakis, J., & Philandras, K. M. (2016). Evaluation of the TRMM 3B43 gridded precipitation estimates over Greece. *Atmospheric Research*, 169, 497-514.
- Nastos, P. T., & Zerefos, C. S. (2007). On extreme daily precipitation totals at Athens, Greece. *Advances in Geosciences*, 10, 59-66.

Nastos, P. T., & Zerefos, C. S. (2010). Cyclic modes of the intra-annual variability of precipitation in Greece. *Advances in Geosciences*, 25, 45-50.

Oliver, M. A., & Webster, R. (1990). Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information System*, 4(3), 313-332.

Ranhao, S., Baiping, Z., & Jing, T. (2008). A multivariate regression model for predicting precipitation in the Daqing Mountains. *Mountain Research and Development*, 28(3), 318-325.

Rossa, A., Nurmi, P., & Ebert, E. (2008). Overview of methods for the verification of quantitative precipitation forecasts. In *Precipitation: Advances in measurement, estimation and prediction* (pp. 419-452). Springer, Berlin, Heidelberg.

Sevruk, B. (1997). Regional dependency of precipitation-altitude relationship in the Swiss Alps. In *Climatic change at high elevation sites* (pp. 123-137). Springer, Dordrecht.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic press.

Ελληνική

Αποστολόπουλος, Ι. (2014) *Εκτίμηση υδρομετεωρολογικών παραμέτρων καταιγίδων κατακόρυφης ανάπτυξης*, Μεταπτυχιακή Διπλωματική Εργασία ΔΠΜΣ «Επιστήμη και Τεχνολογία Υδατικών Πόρων» ΕΜΠ

Ειδική Γραμματεία Υδάτων (ΕΓΥ), (2017). ΣΧΕΔΙΟ ΔΙΑΧΕΙΡΙΣΗΣ ΚΙΝΔΥΝΩΝ ΠΛΗΜΜΥΡΑΣ των Λεκανών Απορροής Ποταμών του Υδατικού Διαμερίσματος Αττικής. ΣΤΑΔΙΟ Ι, 3η φάση Παραδοτέο 6, ΧΑΡΤΕΣ ΕΠΙΚΙΝΔΥΝΟΤΗΤΑΣ ΠΛΗΜΜΥΡΑΣ –ΜΗ ΤΕΧΝΙΚΗ ΕΚΘΕΣΗ. Σελ. 47

Ηλιοπούλου, Π. (2016). Γεωγραφική ανάλυση.

Καρύμπαλης Ε., Παυλόπουλος Κ., (2002) «Παλαιογεωγραφική Αναπαράσταση της ροής του Ηριδανού και Ανθρώπινες επεμβάσεις», Γεωγραφίες, Νο 3.

Κουτσογιάννης Δ., Μαρκόνης Ι., Κουκουβίνος Α., Παπαλεξίου Σ.Μ., Μαμάσης Ν., και Δημητριάδης Π. (2010) Υδρολογική μελέτη ισχυρών βροχοπτώσεων στη λεκάνη του Κηφισού, *Μελέτη διαχείρισης Κηφισού*, Ανάθεση: Γενική Γραμματεία Δημοσίων Έργων – Υπουργείο Περιβάλλοντος, Χωροταξίας και Δημόσιων Έργων

Μάρης Φ., Παπαρρίζος Σπ., Καράτζιος Γ.(2014) Υδρογεωπληροφορική, Εκδόσεις ΔΙΣΙΓΜΑ

Χριστοφίδης Α., (1994) Συμπλήρωση Ελλιπών Υδρομετεωρολογικών Χρονοσειρών σε κατανεμημένες σχεσιακές βάσεις δεδομένων – Διπλωματική Εργασία, ΕΜΠ

Ιστότοποι

Εθνικός κατάλογος ανοικτών γεωχωρικών δεδομένων. <https://geodata.gov.gr/en/> (Τελευταία προσπέλαση: 20/09/2022).

Κλιματικός Άτλαντας ΕΜΥ. <http://climatlas.hnms.gr/> (Τελευταία προσπέλαση: 10/06/2022).

Κλίμα Αττικής. <https://www.meteo.gr/climate/> (Τελευταία προσπέλαση: 10/06/2022).

Collaboration for Australian Weather and Climate Research – Australian Government, Bureau of Meteorology. https://www.cawcr.gov.au/projects/verification/#Roebber_diagram (Τελευταία προσπέλαση: 20/06/2022).

ESRI – Kriging in Geostatistical Analyst, How Kriging works. <https://desktop.arcgis.com/> (Τελευταία προσπέλαση: 01/10/2022).