



**ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ
ΚΡΗΤΗΣ**

**Σχολή Τεχνολογικών Εφαρμογών
Τμήμα Εφαρμοσμένης Πληροφορικής & Πολυμέσων**

Πτυχιακή εργασία

**“Μελέτη χρονοσειρών και κατασκευή μοντέλου πρόβλεψης
ωριαίων τιμών, χρησιμοποιώντας τεχνικές νευρωνικών
δικτύων”**

Επιμέλεια:

Κυριακίδης Ιωάννης Α.Μ.: 969

Επιβλέπων:

κ. Παπαδουράκης Γεώργιος

Τριμελής επιτροπή:

**κ. Παπαδουράκης Γεώργιος
κ. Ακουμιανάκης Δημοσθένης
κ. Βασιλάκης Κωνσταντίνος**

Ηράκλειο Ιούνιος 2007

Περιεχόμενα:

1. Εισαγωγή	4
1.1 Κατηγορίες μεθόδων πρόγνωσης.....	4
1.2 Σκοπός μελέτης.....	5
2. Περιγραφή Δεδομένων	5
2.1 Σταθμοί μέτρησης.....	5
2.2 Περιγραφή χρονοσειρών.....	8
2.3 Επιλογή δεδομένων.....	9
3. Μελέτη χρονοσειράς βενζολίου (σταθμός Πατησίων)[Α]	12
3.1 Μελέτη μέσης τιμής.....	12
3.2 Μελέτη τυπικής απόκλισης (σ).....	17
3.3 Μελέτη συντελεστή συσχέτισης (correlation coefficient).....	23
3.4 Μελέτη κανονικής κατανομής.....	26
3.5 Μελέτη εκτροπής (skew).....	30
3.6 Μελέτη κύρτωσης (kurtosis).....	32
3.7 Μελέτη Συνδιακύμανσης – Συμμεταβλητότητας (covariance).....	34
3.8 Principal Component Analysis (P.C.A.).....	35
3.8.1 Εισαγωγή.....	35
3.8.2 Εφαρμογή της μεθόδου PCA στα δεδομένα μας.....	42
3.9 Υπολογισμός “περιοδογράμματος” της χρονοσειράς βενζολίου.....	44
3.9.1 Μελέτη στασιμότητας.....	44
3.9.2 Μελέτη γραμμής παλινδρόμησης.....	45
3.9.3 Μελέτη κινούμενου μέσου όρου.....	45
3.9.4 Άρση στασιμότητας χρονοσειράς βενζολίου.....	47
3.9.5 Χρήση FFT για τον υπολογισμό και μελέτη του περιοδογράμματος.....	48
4. Πρόγνωση ωριαίων τιμών βενζολίου	50
4.1 Εισαγωγή.....	50
4.2 Συναρτήσεις μεταφοράς.....	52
4.3 Είδη τεχνητών νευρωνικών δικτύων.....	53
4.4 Μέθοδος εκμάθησης Back Propagation.....	54
4.5 Δημιουργία & Μελέτη αποτελεσμάτων νευρωνικών δικτύων.....	56
5. Μελέτη χρονοσειρών όζοντος (σταθμός Πατησίων & Λιοσίων)[Β]	65
5.1 Ελλειπής τιμές.....	65
5.2 Υπολογισμός χαρακτηριστικών.....	67
5.3 Υπολογισμός συντελεστή συσχέτισης (correlation coefficient).....	77
6. Πρόγνωση μέγιστης μέσης τιμής κυλιόμενου οκτάωρου τιμών όζοντος	79
6.1 Εισαγωγή.....	79
6.2 Δημιουργία & Μελέτη αποτελεσμάτων νευρωνικών δικτύων.....	79
7. Συμπεράσματα	84
Παράρτημα Α’	86
Παράρτημα Β’	91
Βιβλιογραφία	100

Ευχαριστίες

Η παρούσα εργασία είναι αποτέλεσμα εντατικής εργασίας κατά το τελευταίο εξάμηνο των σπουδών μου στο τμήμα Εφαρμοσμένης Πληροφορικής και Πολυμέσων του Α.Τ.Ε.Ι. Κρήτης. Για την ολοκλήρωση της συνέβαλαν πολλοί άνθρωποι με πολλούς τρόπους.

Αρχικά θα ήθελα να εκφράσω την ευγνωμοσύνη μου στους γονείς μου και την οικογένεια μου, για την αμέριστη συμπαράσταση και ηθική τους υποστήριξη καθ' όλη τη διάρκεια των σπουδών μου.

Ιδιαίτερα όμως θα ήθελα να ευχαριστήσω τους δύο επιβλέπων καθηγητές της πτυχιακής εργασίας μου, τον κ. Παπαδουράκη Γεώργιο και τον κ. Καρατζά Κωνσταντίνο για την υποστήριξη, την συνεχή καθοδήγηση και τις χρήσιμες συμβουλές τους. Η εμπειρία, η υπομονή και κυρίως το ερευνητικό τους ενδιαφέρον βοήθησαν ουσιαστικά στην ολοκλήρωση της εργασίας. Επίσης θα ήθελα να τους ευχαριστήσω για την προσοχή που έδειξαν έτσι ώστε να υποβληθεί προς δημοσίευση (Παράρτημα Β').

Επίσης θέλω να ευχαριστήσω τους συμφοιτητές και φίλους μου (Γαϊτάνη Ηλία και Παυλίδη Θεόδωρο) που με τις παρατηρήσεις τους συνέβαλαν στο ποιοτικό επίπεδο της μελέτης στο σύνολο της. Αλλά και την φίλη μου Βέργου Ιωάννα που μου παρείχε σημειώσεις και βιβλιογραφία στατιστικής.

Σε όλους αυτούς λοιπόν, εκφράζω την απέραντη ευγνωμοσύνη μου...

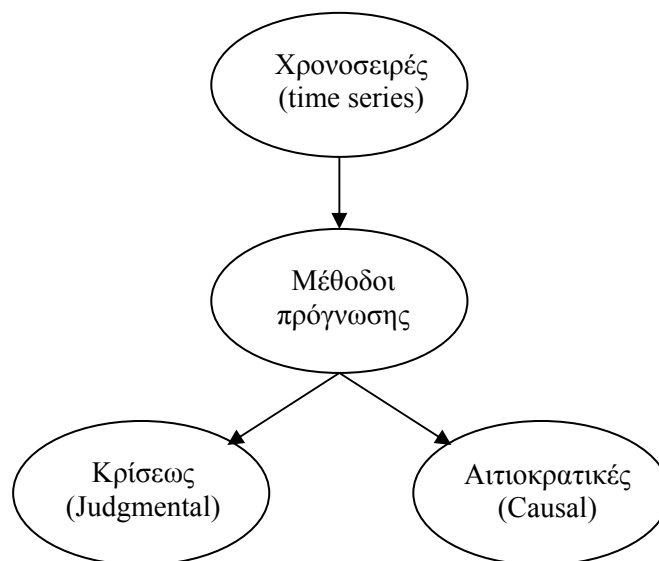
Κυριακίδης Ιωάννης
Ηράκλειο
Ιούνιος 2007

1. Εισαγωγή

Τι είναι πρόβλεψη; Η λέξη προέρχεται από την λέξη πρόγνωση < προγιγνώσκω και είναι η προαίσθηση | πρόνοια ή έγκαιρη φροντίδα για το μέλλον. Διάφορα έχουν ειπωθεί για την πρόγνωση και το μέλλον. “Το παρόν κυοφορεί το μέλλον” είχε πει ο Voltaire. “Ποτέ δεν σκέπτομαι το μέλλον, πλησιάζει πολύ γρήγορα”, Albert Einstein. “Η προσπάθεια να προβλέψεις το μέλλον μελετώντας μόνο το παρελθόν μοιάζει με την προσπάθεια να οδηγήσεις κοιτώντας μόνο από τον καθρέπτη”, Γεώργιος Κοσμετάτος [3]. Σ’ αυτή τη μελέτη θα επιχειρήσουμε να οδηγήσουμε κοιτώντας μόνο από τον καθρέπτη.

1.1. Κατηγορίες μεθόδων πρόγνωσης χρονοσειρών

Γενικά υπάρχουν δύο κατηγορίες μεθόδων πρόγνωσης, οι κρίσεως (judgmental) και οι αιτιοκρατικές (causal):



Σχήμα 1. Γραφική αναπαράσταση μεθόδων πρόγνωσης

Οι μέθοδοι Κρίσεως (Judgmental):

- Βασίζονται σε προσωπικές εκτιμήσεις.
- Αυτές συχνά συνδυάζονται με χρήση τυποποιημένων μεθοδολογιών.
- Είναι εύκολα κατανοητές και ευρύτατα χρησιμοποιούμενες.

Οι Αιτιοκρατικές (Causal) μέθοδοι:

- Αναγνώριση παρελθοντικών συσχετίσεων στα δεδομένα και χρήση τους για πρόβλεψη.
- Υπόθεση εργασίας: οι συσχετίσεις αυτές συνεχίζουν να ισχύουν και στο μέλλον.

1.2 Σκοπός μελέτης

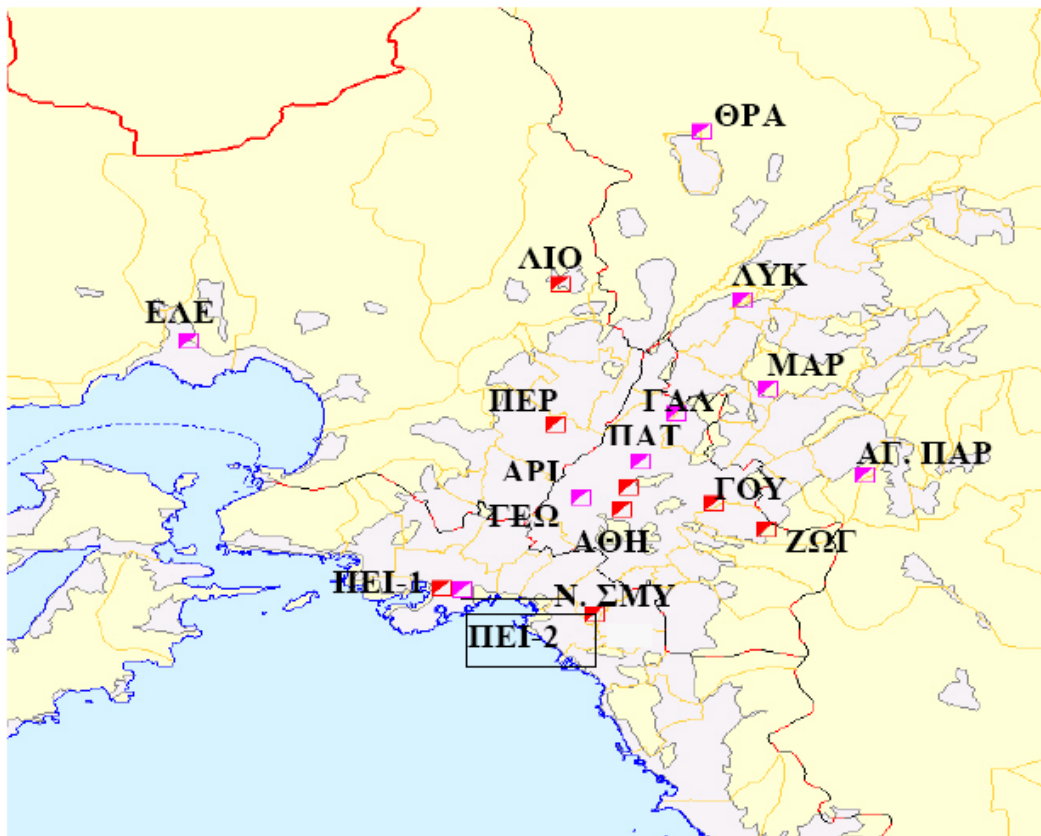
Σε αυτή την μελέτη θα γίνει μια προσπάθεια πρόβλεψης, των τιμών συγκεντρώσεων βενζολίου και όζοντος για την αστική περιοχή Αθηνών. Για την επίτευξη αυτού θα μελετήσουμε χρονοσειρές συγκεντρώσεων ποιότητας αέρα αλλά και χρονοσειρές μετεωρολογικών παραμέτρων, χρησιμοποιώντας αιτιοκρατικές μεθόδους. Όταν αναφερόμαστε στην πρόβλεψη τιμών βενζολίου θα χρησιμοποιούμε το συμβολισμό “[Α]”, και όταν αναφερόμαστε στην πρόβλεψη τιμών όζοντος θα χρησιμοποιούμε το συμβολισμό “[Β]”.

2. Περιγραφή Δεδομένων

2.1 Σταθμοί μέτρησης

Το Τμήμα Ποιότητας Ατμόσφαιρας, που ανήκει στη Δ/ση Ελέγχου Ατμοσφαιρικής Ρύπανσης και Θορύβου (ΕΑΡΘ) του ΥΠΕΧΩΔΕ είναι υπεύθυνο για τη λειτουργία του δικτύου μέτρησης ατμοσφαιρικής ρύπανσης της περιοχής Αθηνών [15].

Το 2005 η Δ/ση ΕΑΡΘ (Τμήμα Ποιότητας Ατμόσφαιρας), λειτούργησε δεκαεπτά σταθμούς μέτρησης ατμοσφαιρικής ρύπανσης στην περιοχή Αττικής (βλ. Σχήμα 2.) καθώς και ένα σταθμό στην Αλίαρτο-Βοιωτίας για τις ανάγκες του Προγράμματος Διασυνοριακής Μεταφοράς της Ρύπανσης (EMEP).



Σχήμα 2. Χάρτης της περιοχής Αττικής που εμφανίζονται οι θέσεις μέτρησης της ατμοσφαιρικής ρύπανσης.

Οι χρονοσειρές που θα χρησιμοποιήσουμε κατά της διάρκεια αυτής της μελέτης, προέρχονται από μετρήσεις που γίνονται σε δύο σημεία στην Αθήνα, στον “σταθμό” Πατησίων και στον “σταθμό” Λιοσίων. Παρακάτω παραθέτουμε ένα πίνακα με όλες τις μετρήσεις που έγιναν σε αυτούς τους σταθμούς από το 2003 έως και το 2005.

Σταθμός Πατησίων			
2002	2003	2004	2005
CO#	CO#	BENZ	BENZ
NO2	NO2	CO#	CO#
NO#	NO#	SMOKE	NO2
NOX	O3#	SO2	NO#
O3#	SMOKE	-	O3#
SMOKE	SO2	-	SO2
SO2	-	-	-
+ Μετεωρολογικά στοιχεία	+ Μετεωρολογικά στοιχεία	+ Μετεωρολογικά στοιχεία	+ Μετεωρολογικά στοιχεία

Πίνακας 1. Πίνακας συνολικών δεδομένων (ποιότητας αέρα και μετεωρολογικών) από τον σταθμό Πατησίων.

Σταθμός Λιοσίων			
2002	2003	2004	2005
NO2	NO2	NO2	NO2
NO#	NO#	NO#	NO#
O3#	O3#	O3#	O3#
SO2	SO2	SO2	SO2
+ Μετεωρολογικά στοιχεία	+ Μετεωρολογικά στοιχεία	+ Μετεωρολογικά στοιχεία	+ Μετεωρολογικά στοιχεία

Πίνακας 2. Πίνακας συνολικών δεδομένων (ποιότητας αέρα και μετεωρολογικών) από τον σταθμό Λιοσίων.

2.2 Περιγραφή χρονοσειρών

Air Quality Λιοσίων (02-05) [B]

- NO# : Nitric Oxide (Μονοξείδιο του αζώτου)
- NO₂ : Nitrogen Dioxide (Διοξείδιο Αζώτου)
- O₃# : Ozone (Οζον) [μg/m³]
- SO₂ : Sulfur Dioxide (Διοξείδιο Θείου)

Air Quality Πατησίων (02-05) [A & B]

- CO# : Carbon Monoxide (Μονοξείδιο Άνθρακα) [mg/m³]
- NO₂ : Nitrogen Dioxide (Διοξείδιο Αζώτου)
- NO# : Nitric Oxide (Μονοξείδιο του αζώτου)
- NO_x : Nitrogen Oxide (Οξείδιο Αζώτου)
- O₃# : Ozone (Οζον) [μg/m³]
- Smoke : Καπνός
- SO₂ : Sulfur Dioxide (Διοξείδιο Θείου)
- C₆H₆ : Βελζόλιο

Μετεωρολογικά Λιοσίων (02-05) [B]

- RH# : Relative Humidity (Σχετική Υγρασία)
- Ta# : Air Temperature (°C)
- WD# : Wind Direction (Χρήση τύπου: $WD=1+\sin(\theta+\pi/4)$)
- WS# : Wind Speed (Ταχύτητα Ανέμου) [ms⁻¹]

Μετεωρολογικά Πατησίων (02-05) [A & B]

- RH# : Relative Humidity (Σχετική Υγρασία)
- Ta# : Air Temperature (°C)
- WD# : Wind Direction (Χρήση τύπου: $WD=1+\sin(\theta+\pi/4)$)
- WS# : Wind Speed (Ταχύτητα Ανέμου) [ms⁻¹]

Περισσότερες πληροφορίες για τις συγκεντρώσεις ποιότητας αέρας υπάρχουν στο Παράρτημα Α'.

2.3 Επιλογή δεδομένων

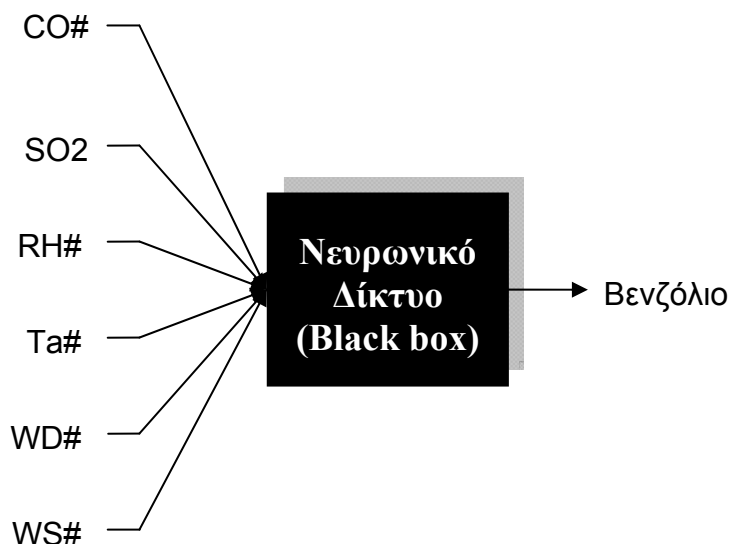
[A]:

Από τα δεδομένα αποστολής αλλά και τον στόχο της εργασίας (πρόβλεψη και μελέτη βενζολίου για τον σταθμό Πατησίων), τα δεδομένα που μπορούμε να χρησιμοποιήσουμε είναι:

2004	2005
BENZ	BENZ
CO#	CO#
SO2	SO2
+ ΜΕΤΕΩ...	+ ΜΕΤΕΩ...

Πίνακας 3. Πίνακας δεδομένων που μπορούν να χρησιμοποιηθούν από τον σταθμό Πατησίων

Η επιλογή έγινε βάση των κοινών χαρακτηριστικών, που διαθέτουμε για τα έτη 2004-2005 στα οποία υπάρχουν τιμές βενζολίου. Παρακάτω δείχνουμε την βασική δομή του νευρωνικού δικτύου μας που θα χρησιμοποιήσουμε για κάνουμε πρόβλεψη.



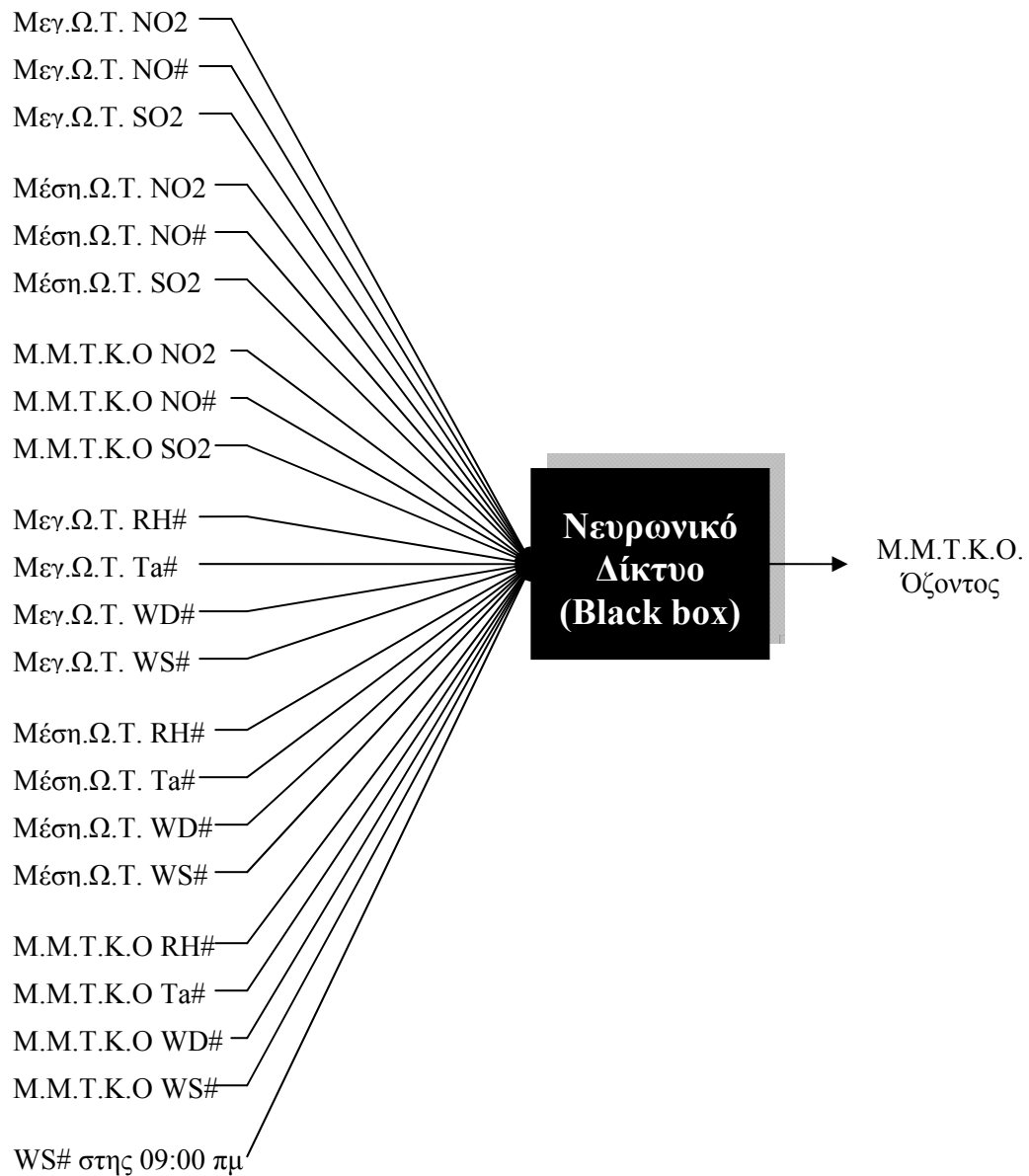
Παρατήρηση: Οι παραπάνω τιμές είναι ωριαίες.

[B]:

Από τα δεδομένα αποστολής αλλά και τον στόχο της εργασίας (πρόβλεψη μέγιστων μέσων τιμών κυλιόμενου οκταώρου όζοντος για τον σταθμό Πατησίων και Λιοσίων), τα δεδομένα που μπορούμε να χρησιμοποιήσουμε είναι:

Σταθμός Λιοσίων				Σταθμός Πατησίων		
2002	2003	2004	2005	2002	2003	2005
NO2	NO2	NO2	NO2	NO2	NO2	NO2
NO#	NO#	NO#	NO#	NO#	NO#	NO#
O3#	O3#	O3#	O3#	O3#	O3#	O3#
SO2	SO2	SO2	SO2	SO2	SO2	SO2
+ ΜΕΤΕΩ...	+ ΜΕΤΕΩ...	+ ΜΕΤΕΩ...	+ ΜΕΤΕΩ...	+ ΜΕΤΕΩ...	+ ΜΕΤΕΩ...	+ ΜΕΤΕΩ...

Πίνακας 4. Πίνακας δεδομένων που μπορούν να χρησιμοποιηθούν από τον σταθμό Λιοσίων και Πατησίων

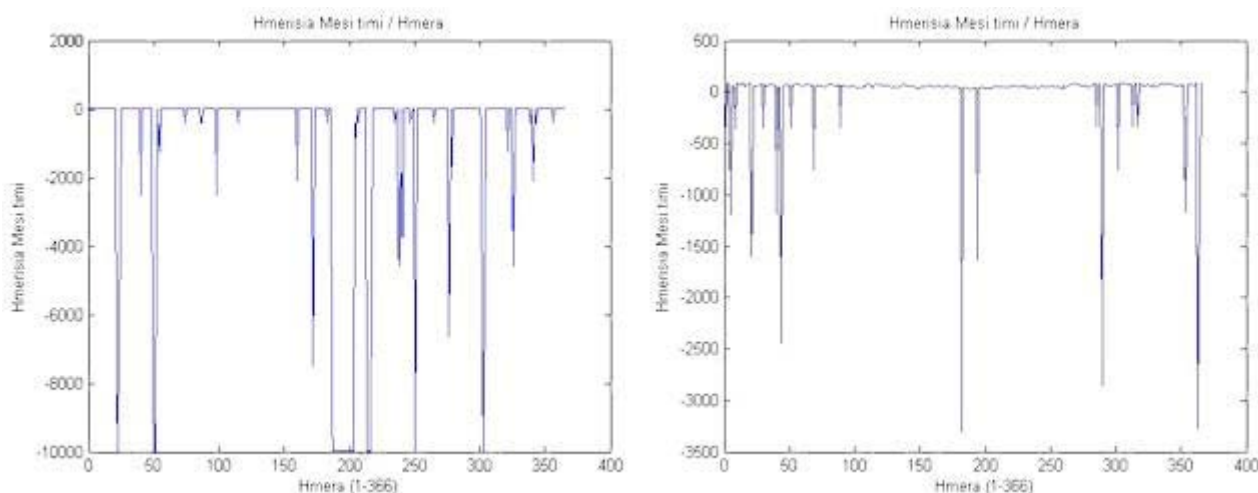


Μεγ.Ω.Τ.: Μέγιστη Ωριαία Τιμή
Μέση.Ω.Τ.: Μέση Ωριαία Τιμή
Μ.Μ.Τ.Κ.Ο.: Μέγιστη Μέση Τιμή Κυλιόμενου Οκταώρου

3. Μελέτη χρονοσειράς βενζολίου (σταθμός Πατησίων) [Α]

3.1 Μελέτη μέσης τιμής

Προσπαθώντας να βρούμε την μέση τιμή για κάθε ημέρα παρατηρούμε ότι υπάρχουν ελλείψεις στα δεδομένα. Οι ωριαίες τιμές για τις οποίες δεν υπάρχουν μετρήσεις, τις συμβολίζουμε με την τιμή -9999 (συμβολισμός μηχανήματος μέτρησης), αυτό φαίνεται και στα παρακάτω σχήματα με τις μέσες τιμές κάποιων χαρακτηριστικών, οι οποίες έχουν αλλοιωθεί. Έτσι στην συνέχεια εντοπίζουμε και διαγράφουμε αυτές τις τιμές.



Σχήμα 3. Αριστερά (α) βλέπουμε τον μέσο όρο τιμών βενζολίου για το έτος 2005, και δεξιά (β) τον μέσο όρο των τιμών των μετεωρολογικών στοιχείων σχετικής υγρασίας (Relative Humidity) για το έτος 2004.

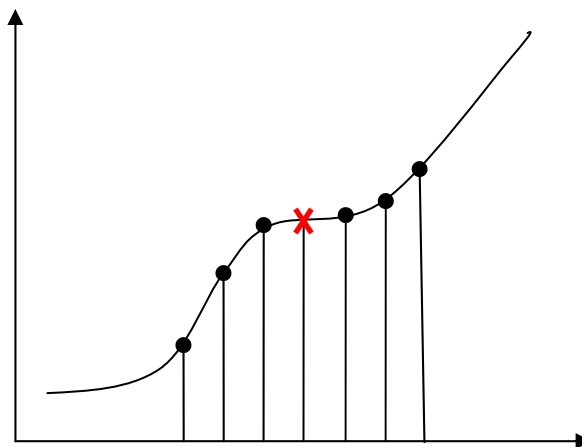
Παρατηρώντας τα δεδομένα μας είδαμε επίσης ότι συνήθως όταν είχαμε πολλές ελλείψεις (περισσότερες του 30%) σε μία ημέρα, οι ελλείψεις ήταν συνεχόμενες, έτσι θα ήταν δύσκολο να βρούμε τις τιμές αυτές χρησιμοποιώντας μεθόδους όπως την παρεμβολή Newton ή Lagrange. Έτσι οι ημέρες οι οποίες είχαν ελλείψεις περισσότερες ή ίσες του 30% (δηλαδή πάνω από 7 ελλείψεις), δεν τις περνούμε υπόψη, και τις διαγράψαμε από τα συνολικά δεδομένα. Οι περισσότερες ελλείψεις υπήρχαν στα δεδομένα βενζολίου, για τα οποία θέλουμε να κάνουμε πρόβλεψη. Παρακάτω υπάρχει πίνακας με τις ημέρες οι οποίες έχουν διαγραφεί ανά έτος από όλα τα δεδομένα (μετεωρολογικά και ποιότητας αέρα από τον σταθμό Πατησίων).

Έτος 2004		Έτος 2005	
Ημέρες	Ημερομηνίες	Ημέρες	Ημερομηνίες
1 - 84	1/1/2004 – 24/3/2004	22 - 24	22/1/2005 – 24/1/2005
180 - 182	28/6/2004 – 30/6/2004	49 - 52	18/2/2005 – 21/2/2005
214	1/8/2004	83 - 89	23/3/2005 – 30/3/2005
216-218	3/8/2004 – 5/8/2004	172 - 173	21/6/2005 – 22/6/2005
245 - 259	1/9/2004 – 15/9/2004	187 - 203	6/7/2005 – 22/7/2005
269 - 272	25/9/2004 – 28/9/2004	213 - 217	1/8/2005 – 5/8/2005

290	16/10/2004	237 - 238	25/8/2005 – 26/8/2005
297 - 300	23/10/2004 – 26/10/2004	240 - 241	28/8/2005 – 29/8/2005
327 - 329	22/11/2004 – 24/11/2004	250 - 251	7/9/2005 – 8/9/2005
336	1/12/2004	276-277	3/10/2005 – 4/10/2005
363	28/12/2004	302 - 304	29/10/2005 – 31/10/2005
		321	17/11/2005
		326 - 329	22/11/2005 – 25/11/2005

Συνολικά από το έτος 2004 διαγράφηκαν 110 ημέρες και από το έτος 2005 διαγράφηκαν 51 ημέρες.

Για τις υπόλοιπες μέρες οι οποίες είχαν απώλειες μικρότερες του 30%, δημιουργήσαμε και εφαρμόσαμε την μέθοδο παρεμβολής Lagrange. Έτσι για κάθε ελλιπή ωριαία τιμή βρήκαμε μια συνάρτηση $f(x)=y$, χρησιμοποιώντας ως γνωστά σημεία, τρία σημεία αριστερά της ελλιπής ωριαίας τιμής και τρία δεξιά αυτής όπως φαίνεται στο παρακάτω σχήμα.



Σχήμα 4. Παράδειγμα χρήσης πολυωνύμου Lagrange

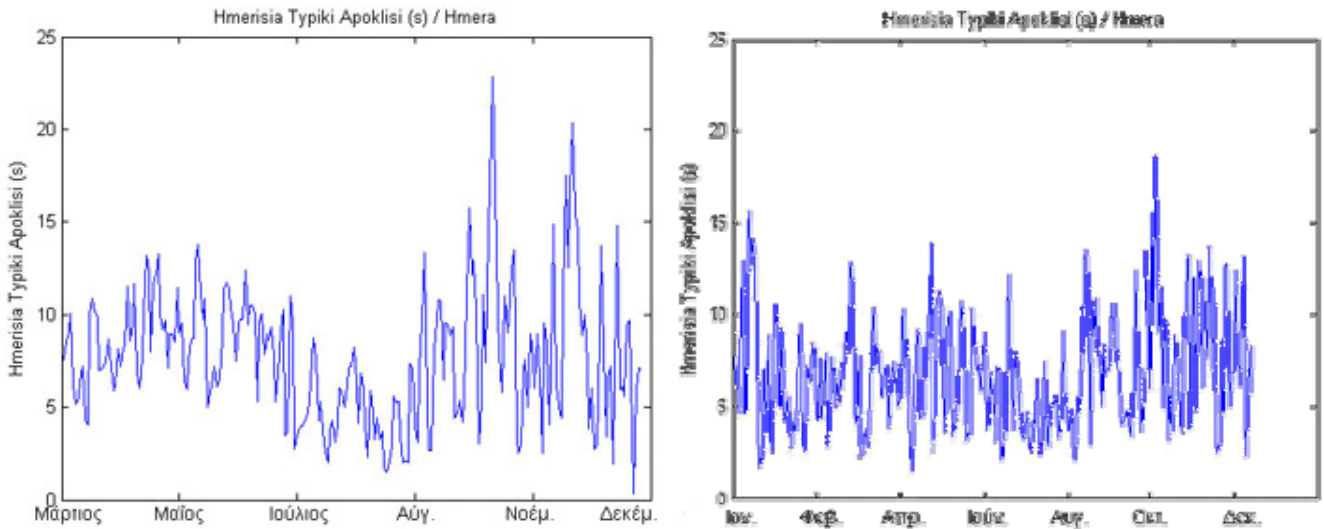
Γενικά ένα πολυώνυμο Lagrange δίνεται από τη παρακάτω σχέση:

$$y(x) = \sum_{i=0}^N \frac{y_i}{L_i(x_i)} L_i(x)$$

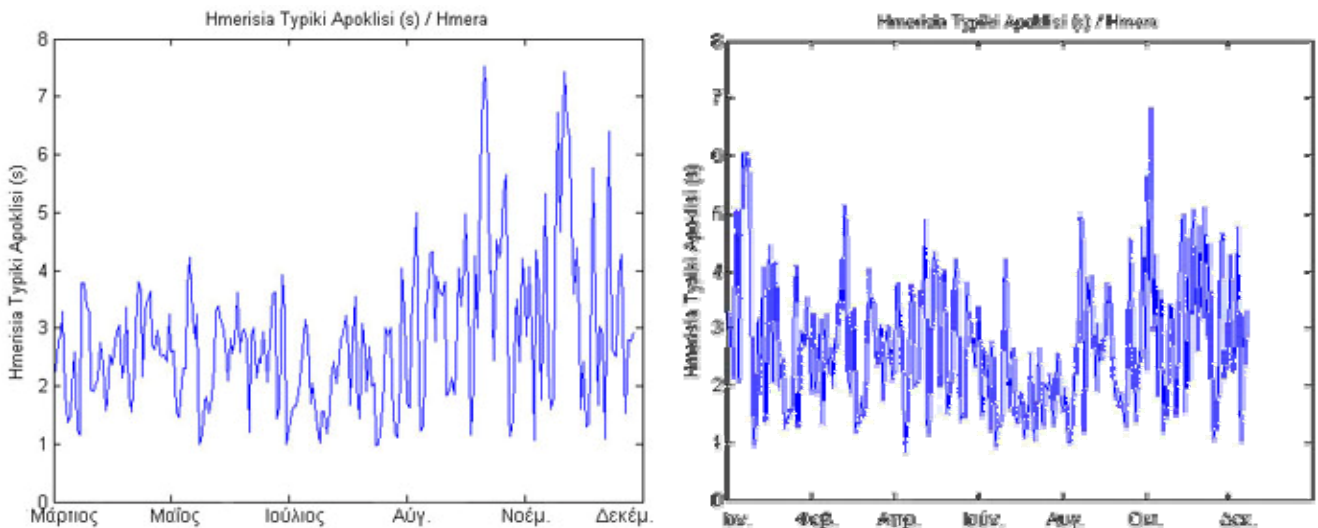
όπου $L_i(x_i)$:

$$L_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^N (x - x_j)$$

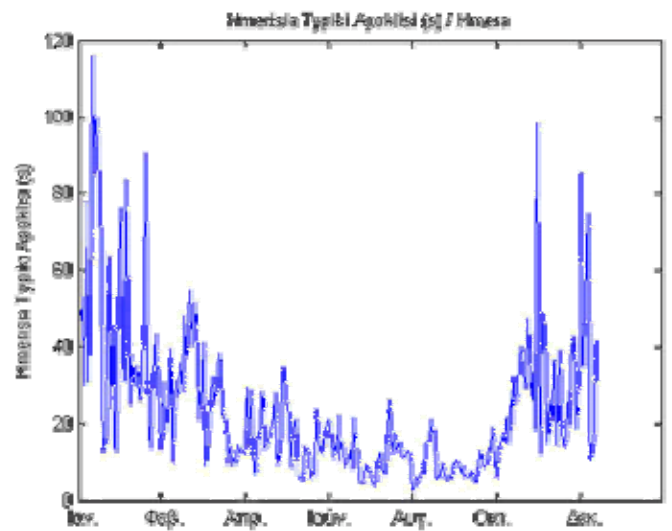
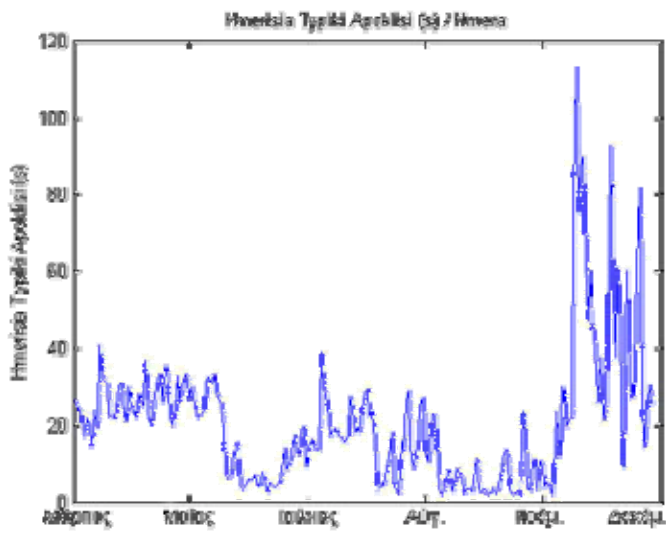
Μετά την ολοκλήρωση της διαδικασίας ανεύρεσης των ελλείπων τιμών, ξανά υπολογίσαμε τους ημερησίους μέσους όρους των συγκεντρώσεων ποιότητας αέρα και των μετεωρολογικών παραμέτρων. Επίσης αντί της διεύθυνσης αέρα WD κάναμε χρήση του τύπου $WDn=1+\sin[h+(\pi/4)]$.



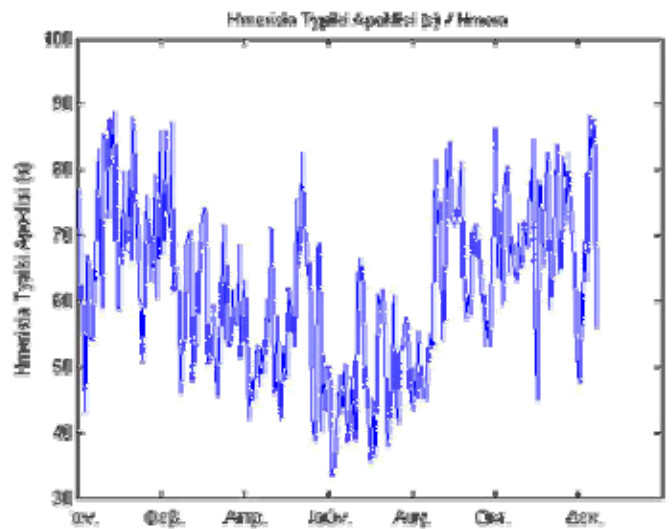
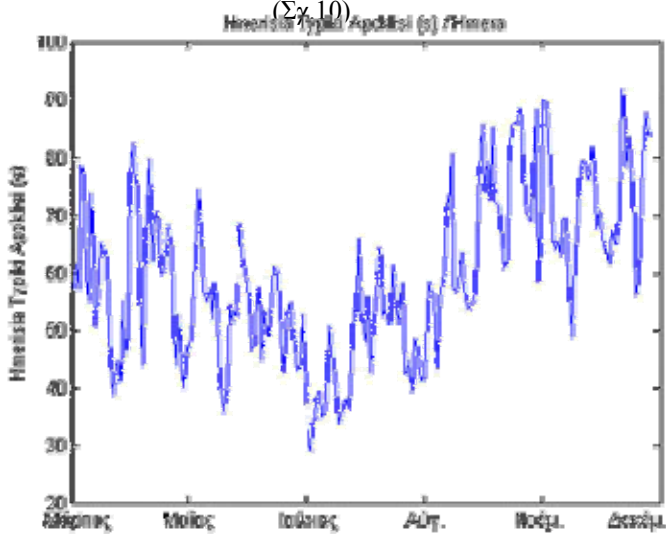
Σχήμα 5 & 6: Αριστερά φαίνεται ο μέσος όρος των τιμών βενζολίου για το έτος 2004 (Σχ.5), και δεξιά ο μέσος όρος των τιμών βενζολίου για το έτος 2005 (Σχ.6).



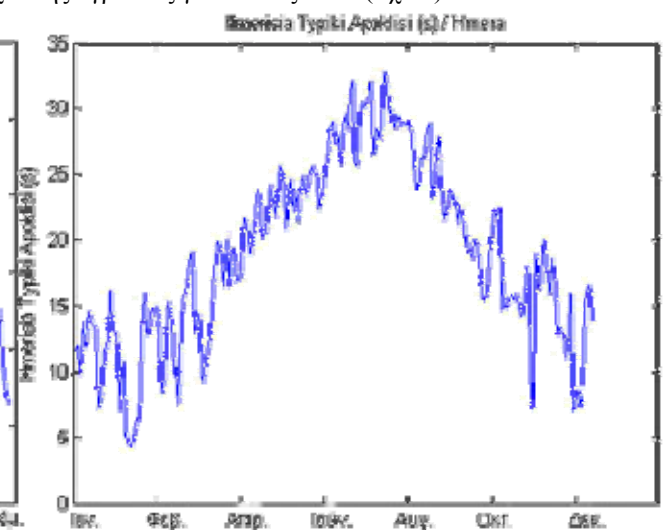
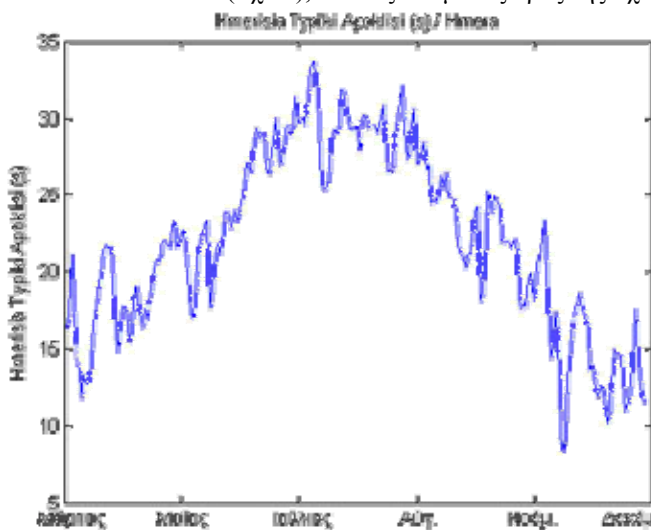
Σχήμα 7 & 8: Αριστερά φαίνεται ο μέσος όρος των τιμών μονοξειδίου του άνθρακα (CO#) για το έτος 2004 (Σχ.7), και δεξιά ο μέσος όρος των τιμών μονοξειδίου του άνθρακα για το έτος 2005 (Σχ.8).



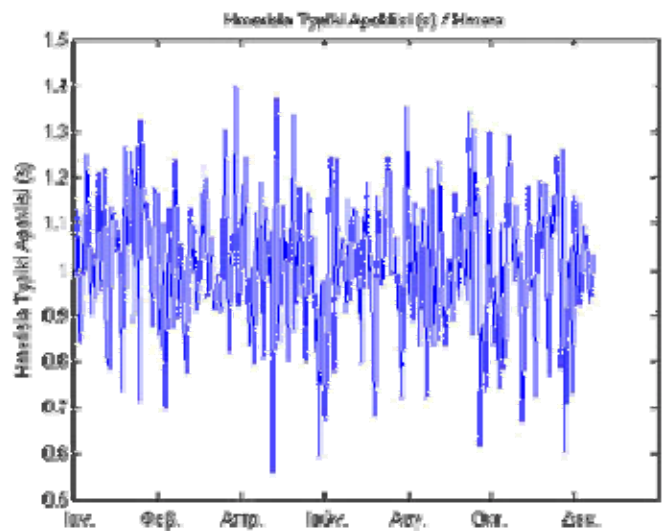
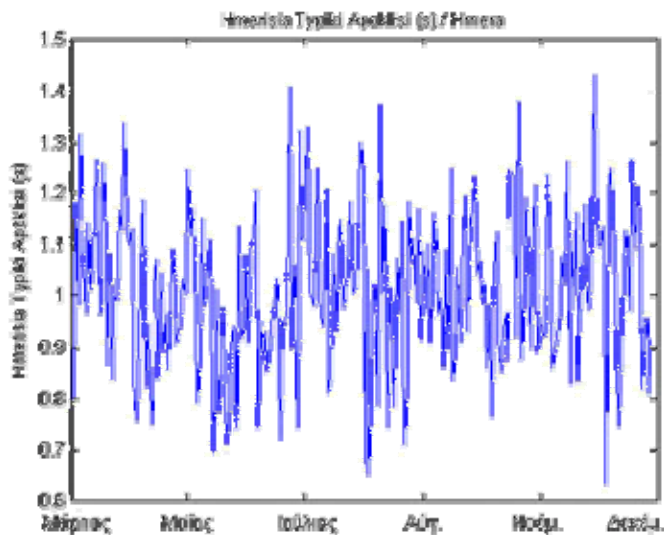
Σχήμα 9 & 10: Αριστερά φαίνεται ο μέσος όρος των τιμών διοξειδίου του θείου (SO₂) για το έτος 2004 (Σχ.9), και δεξιά ο μέσος όρος των τιμών διοξειδίου του θείου για το έτος 2005 (Σχ.10).



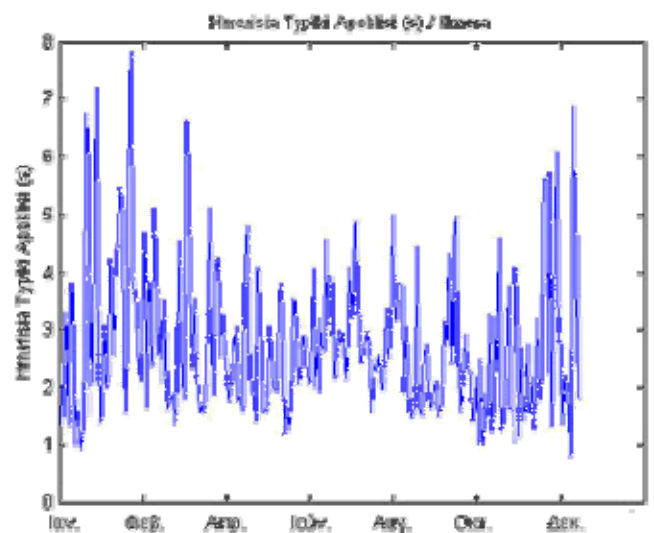
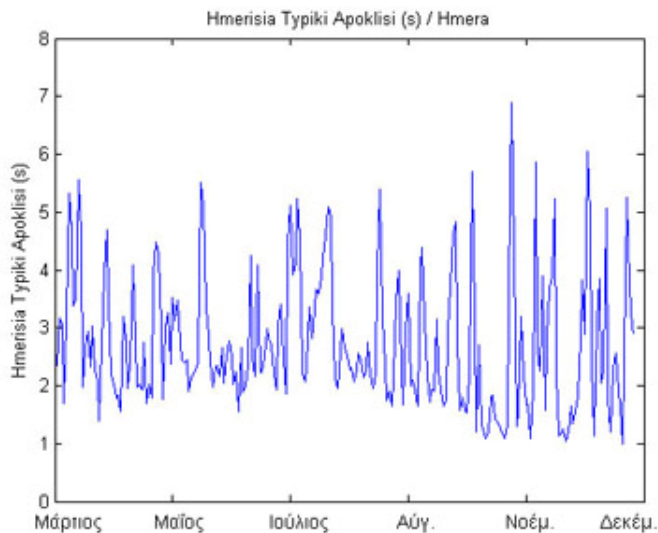
Σχήμα 11 & 12: Αριστερά φαίνεται ο μέσος όρος της σχετικής υγρασίας (RH#) για το έτος 2004 (Σχ.11), και δεξιά ο μέσος όρος της σχετικής υγρασίας για το έτος 2005 (Σχ.12).



Σχήμα 13 & 14: Αριστερά φαίνεται ο μέσος όρος της θερμοκρασίας (Ta) για το έτος 2004 (Σχ. 13), και δεξιά ο μέσος όρος της θερμοκρασίας (Ta) για το έτος 2005 (Σχ. 14).



Σχήμα 15 & 16: Αριστερά φαίνεται ο μέσος όρος των τιμών WDn (τροποποιημένες τιμές της διεύθυνσης του ανέμου) για το έτος 2004 (Σχ. 15), και δεξιά ο μέσος όρος των τιμών WDn για το έτος 2005 (Σχ. 16).



Σχήμα 17 & 18: Αριστερά φαίνεται ο μέσος όρος της ταχύτητας ανέμου (WS) για το έτος 2004 (Σχ.17), και δεξιά ο μέσος όρος της ταχύτητας ανέμου για το έτος 2005 (Σχ. 18).

3.2 Μελέτη τυπικής απόκλισης (σ):

Για τον υπολογισμό της τυπικής απόκλισης (σ), δημιουργήσαμε μια συνάρτηση υπολογισμού, και την χρησιμοποιήσαμε για τον υπολογισμό της τυπικής απόκλισης για κάθε ημέρα του έτους. Αυτό θα μας δώσει πληροφορίες για την διασπορά των δεδομένων μας ανά ημέρα. Γενικά η τυπική απόκλιση (του συνολικού πληθυσμού) υπολογίζεται ως εξής:

$$\sigma_x = \sqrt{\frac{D_x}{N - 1}}$$

όπου:

$$D_x = E(x - \mu)^2$$

$$\mu = E_x$$

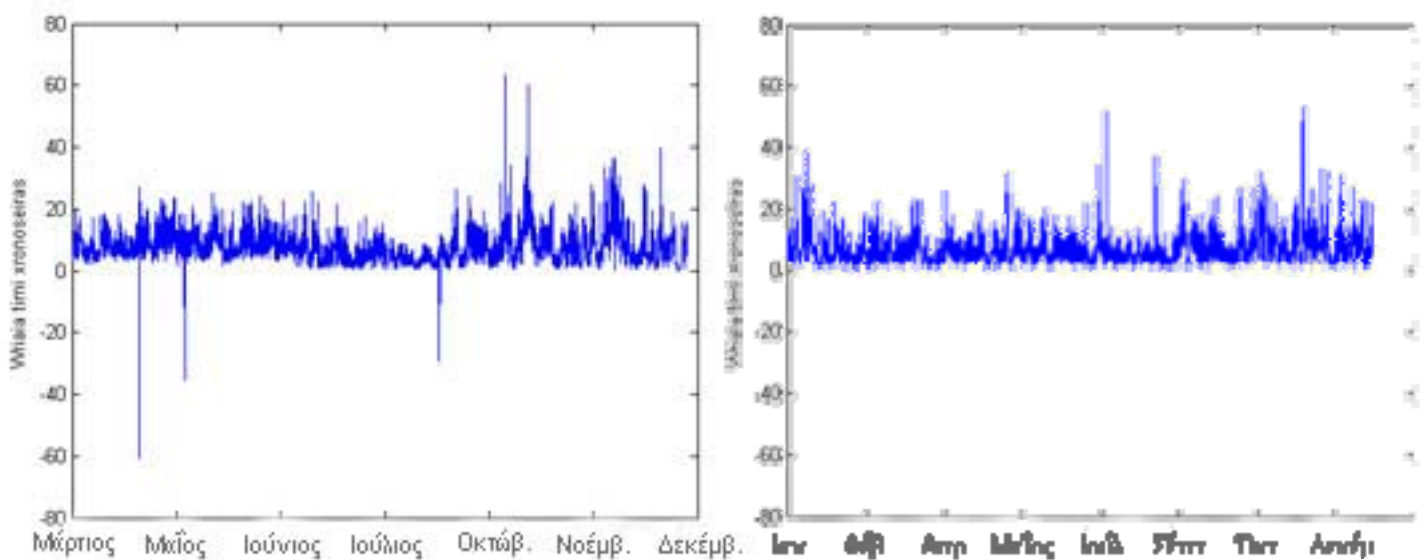
* x: τυχαία διακριτή μεταβλητή.

Παρακάτω παραθέτουμε τον πίνακα με τις τιμές της τυπικής απόκλισης ανά ρύπο και έτος. Επίσης παραθέτουμε τις γραφικές παραστάσεις των χρονοσειρών μας ανά ρύπο και έτος.

Πίνακας: Τυπικής Απόκλισης		
	2004	2005
BENZ	5.2877	4.7237
CO#	1.7448	1.6602
S02	23.33	27.503
METEO RH	16.642	15.841
METEO Ta	6.1737	7.0478
METEO WDn	0.70404	0.70807
METEO WS	1.7188	1.7897

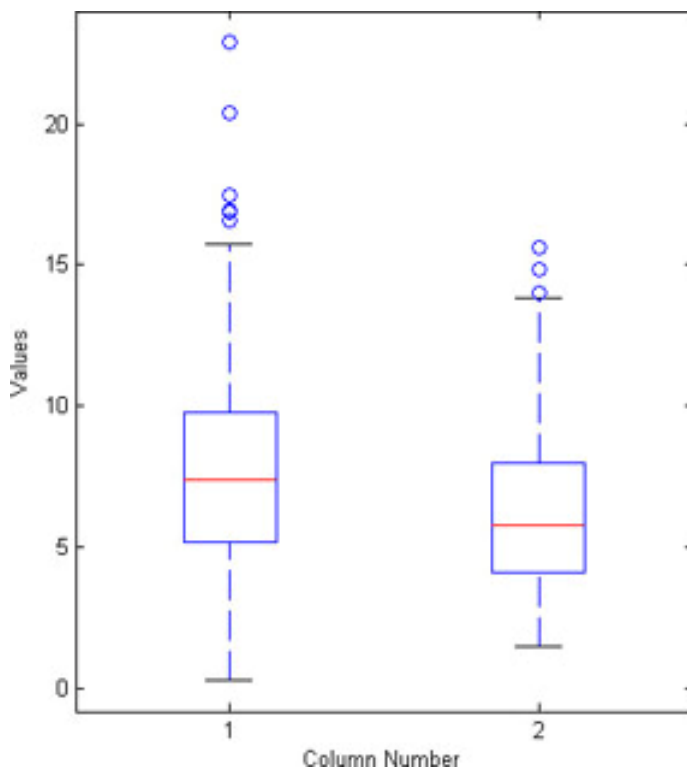
Πίνακας 5. Πίνακας τυπικής απόκλισης για τα έτη 2004-2005.

Η σημαντικότερη αδυναμία της τυπικής απόκλισης είναι η ευαισθησία της στις ακραίες τιμές της κατανομής. Οι ακραίες τιμές συμβάλλουν δυσανάλογα περισσότερο από τις υπόλοιπες τιμές στον υπολογισμό της τυπικής απόκλισης.



Σχήμα 19 & 20: Αριστερά φαίνεται η χρονοσειρά του των τιμών βενζολίου για το έτος 2004 (Σχ.19), και δεξιά η χρονοσειρά των τιμών βενζολίου για το έτος 2005.

Για την μελέτη της διασποράς μπορούμε να μελετήσουμε επίσης και το διάγραμμα κουτιού (box plot) για τα παραπάνω έτη [1]. Για την δημιουργία του παρακάτω διαγράμματος κουτιού χρησιμοποιήσαμε την συνάρτηση boxplot() του Matlab. Η στήλη1 αντιπροσωπεύει το έτος 2004 και η στήλη 2 το έτος 2005 από τους μέσους όρους των τιμές βενζολίου.

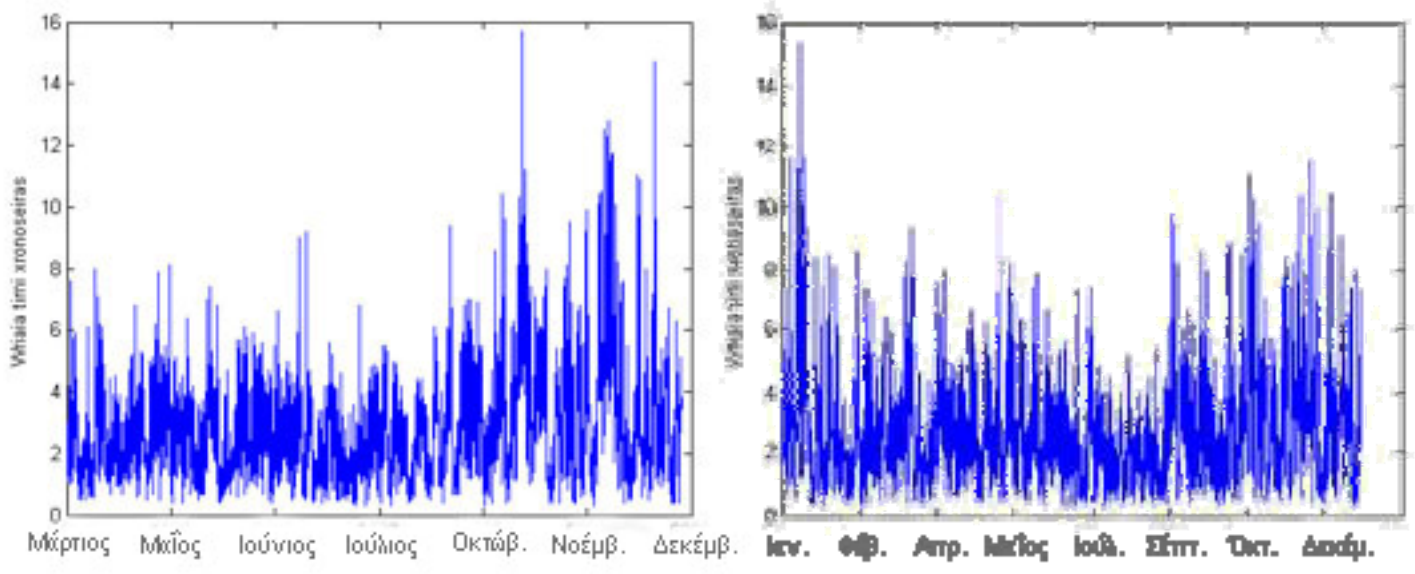


1. Από την θέση της διαμέσου στο σχήμα συμπεραίνουμε ότι στο έτος 2004 το κεντρικό τμήμα της κατανομής είναι συμμετρικό ενώ στο έτος 2005 δεν είναι και τόσο συμμετρικό. Γενικά όσο πιο πολύ η διάμεσος (κόκκινη γραμμή) απομακρύνεται από το κέντρο του, τόσο πιο ασύμμετρο είναι το κεντρικό 50% της κατανομής.

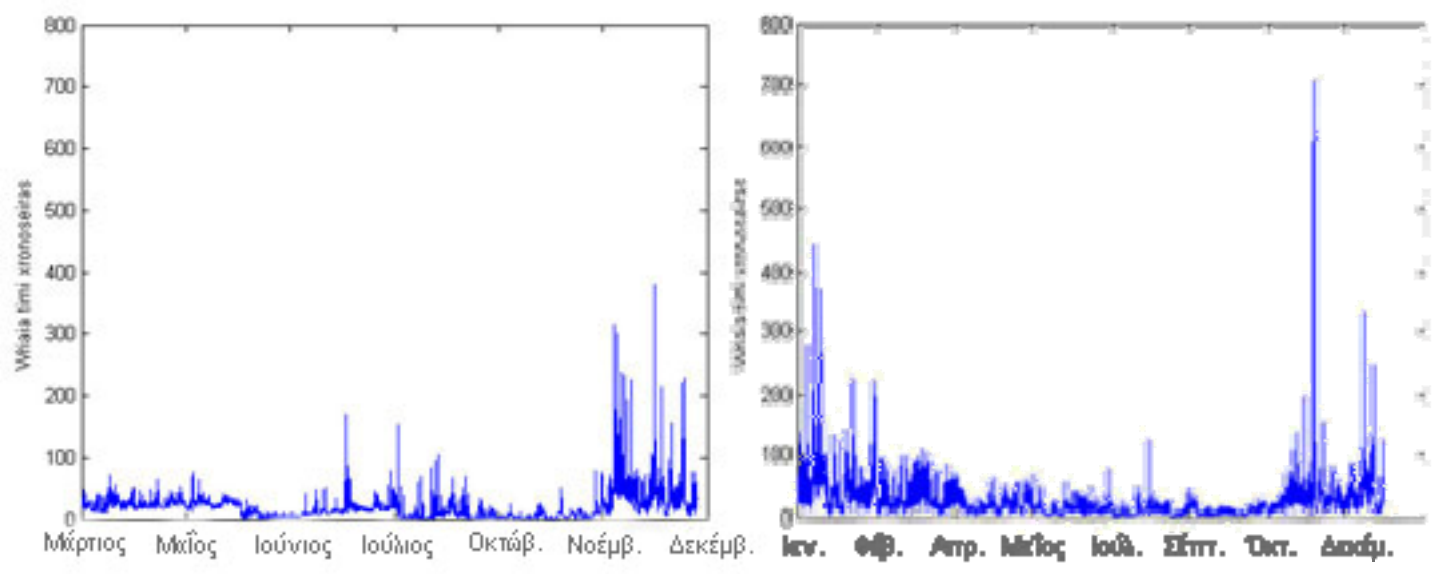
2. Από τα μεγέθη των “μουστακιών” (αριστερού και δεξιού) συμπεραίνουμε ότι δεν υπάρχει ασυμμετρία στο ανώτερο και στο κατώτερο 25% των τιμών για το έτος 2004. Αυτό συμβαίνει επειδή το μήκος του αριστερού “μουστακιού” είναι σχεδόν ίσο

με του δεξιού. Ενώ για στο έτος 2005 βλέπουμε ακριβώς το αντίθετο.

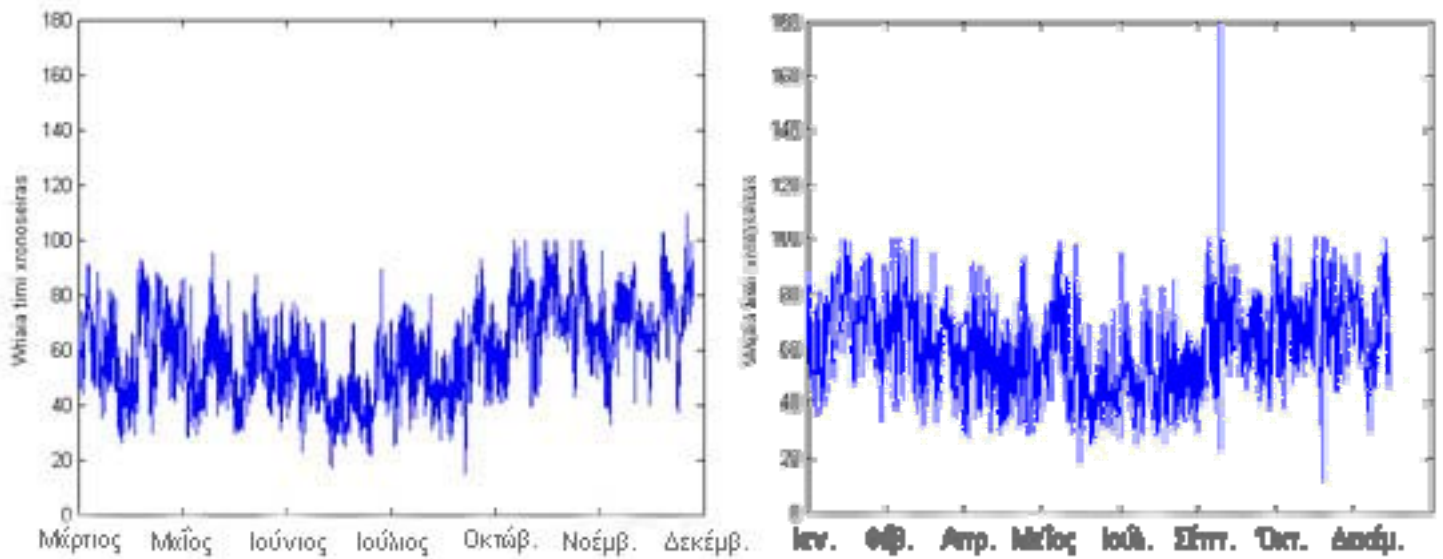
3. Επίσης παρατηρούμε ότι στο έτος 2004 έχουμε περισσότερες ακραίες τιμές σε σχέση με το έτος 2005. Από αυτό συμπεραίνουμε ότι ο μέσος όρος άλλα και η διάμεσος του έτους 2004 έχουν επηρεαστεί περισσότερο από ότι έχει επηρεαστεί ο μέσος όρος και η διάμεσος του έτους 2005.



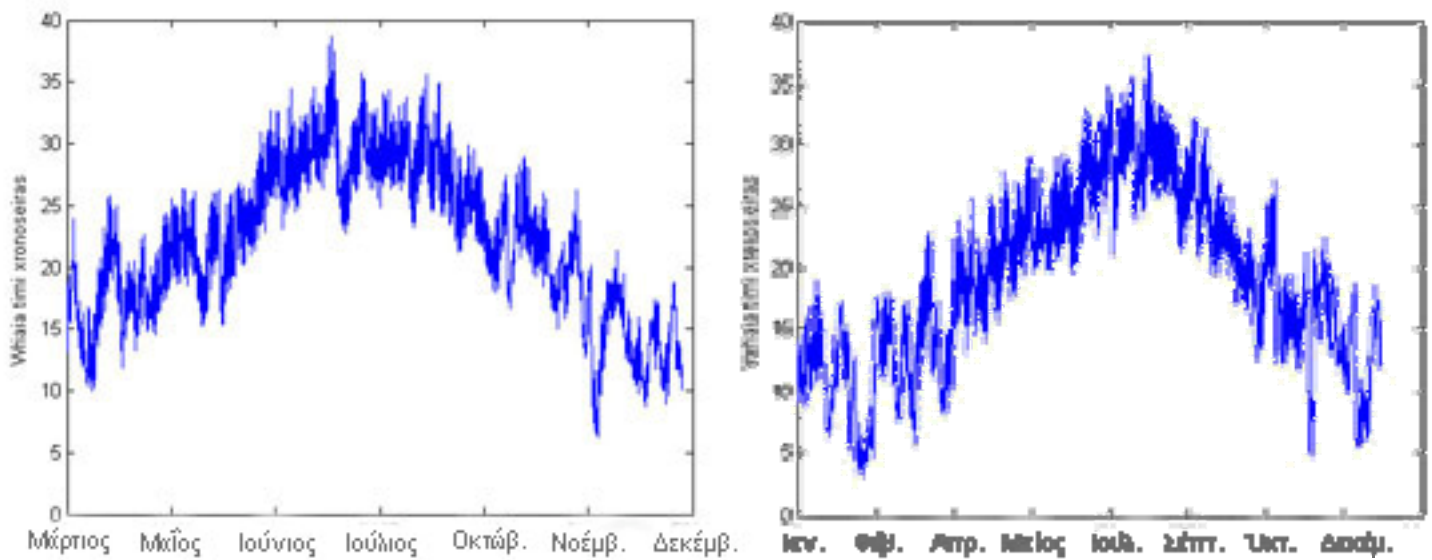
Σχήμα 20 & 21: Αριστερά φαίνεται η χρονοσειρά του μονοξειδίου του άνθρακα (CO#) για το έτος 2004 (Σχ.20), και δεξιά η χρονοσειρά του μονοξειδίου του άνθρακα για το έτος 2005.



Σχήμα 22 & 23: Αριστερά φαίνεται η χρονοσειρά του διοξειδίου του θείου (SO2) για το έτος 2004 (Σχ.22), και δεξιά η χρονοσειρά του διοξειδίου του θείου για το έτος 2005.



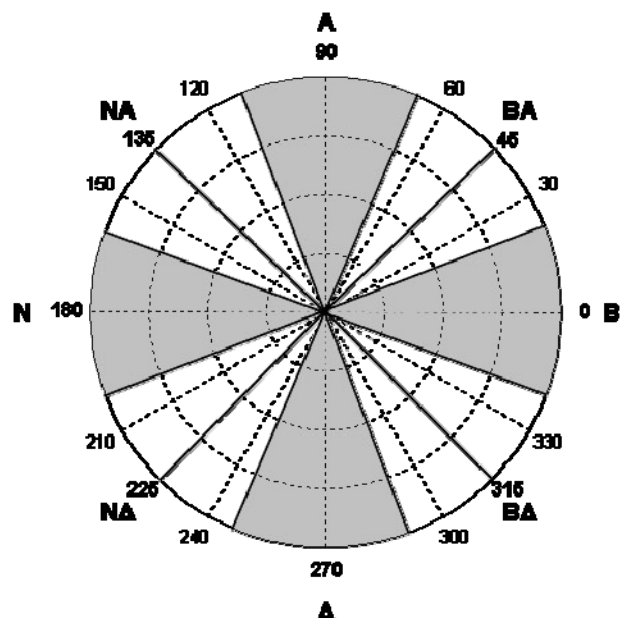
Σχήμα 24 & 25: Αριστερά φαίνεται η χρονοσειρά των τιμών της σχετικής υγρασίας (RH#) για το έτος 2004 (Σχ.24), και δεξιά η χρονοσειρά των τιμών της σχετικής υγρασίας για το έτος 2005.



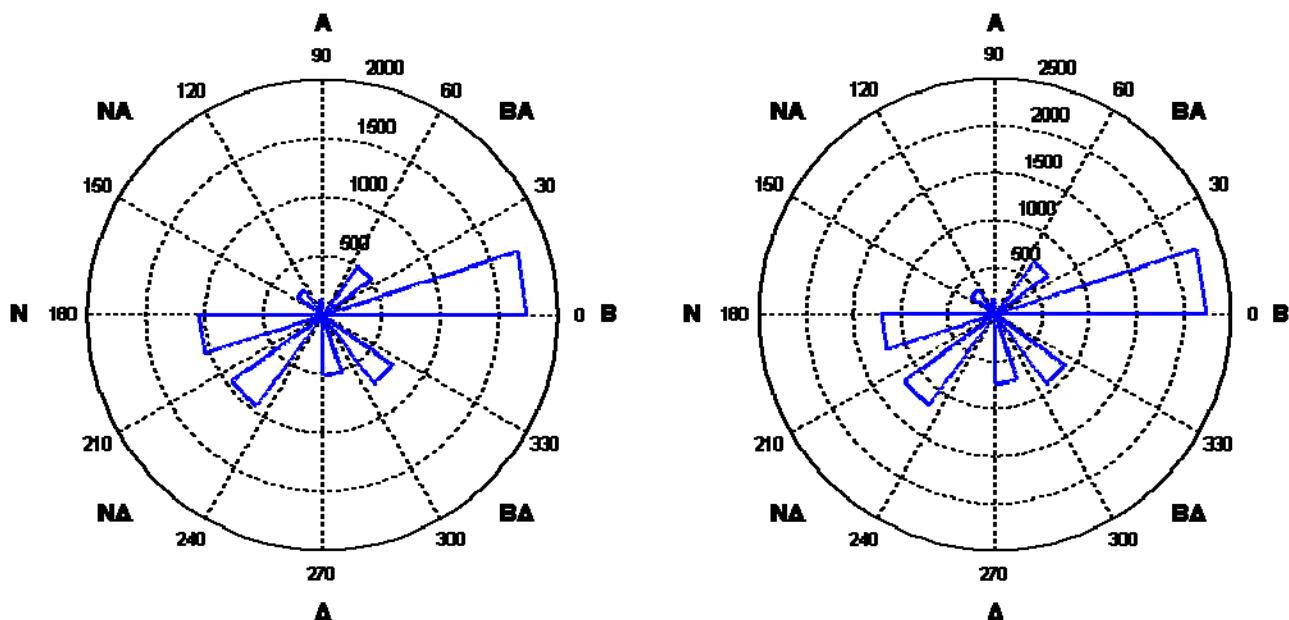
Σχήμα 26 & 27: Αριστερά φαίνεται η χρονοσειρά των τιμών της θερμοκρασίας για το έτος 2004 (Σχ.26), και δεξιά η χρονοσειρά των τιμών της θερμοκρασίας για το έτος 2005.

Για την απεικόνιση της χρονοσειράς διεύθυνσης ανέμου (Wind Direction), χρησιμοποιήσαμε διάγραμμα τύπου “Polar”, χρησιμοποιώντας συναρτήσεις του Matlab. Παρακάτω παραθέτουμε ένα παράδειγμα με την μορφή αυτών αλλά και τον τρόπο με τον οποίο ομαδοποιήσαμε τα δεδομένα μας. Οι ομάδες που δημιουργήσαμε αλλά και οι μοίρες που αντιστοιχούν σε κάθε μία από αυτές βρίσκεται στην παρακάτω λίστα:

- Βόρειος άνεμος: $0-22,5^\circ$ και $337,6-360^\circ$
- Βορειοανατολικός άνεμος: $22,6-67,5^\circ$
- Ανατολικός άνεμος: $67,6-112,5^\circ$
- Νοτιοανατολικός άνεμος: $112,6-157,5^\circ$
- Νότιος άνεμος: $157,6-202,5^\circ$
- Νοτιοδυτικός άνεμος: $202,6-247,5^\circ$
- Δυτικός άνεμος: $247,6-292,5^\circ$
- Βορειοδυτικός άνεμος: $292,6-337,5^\circ$

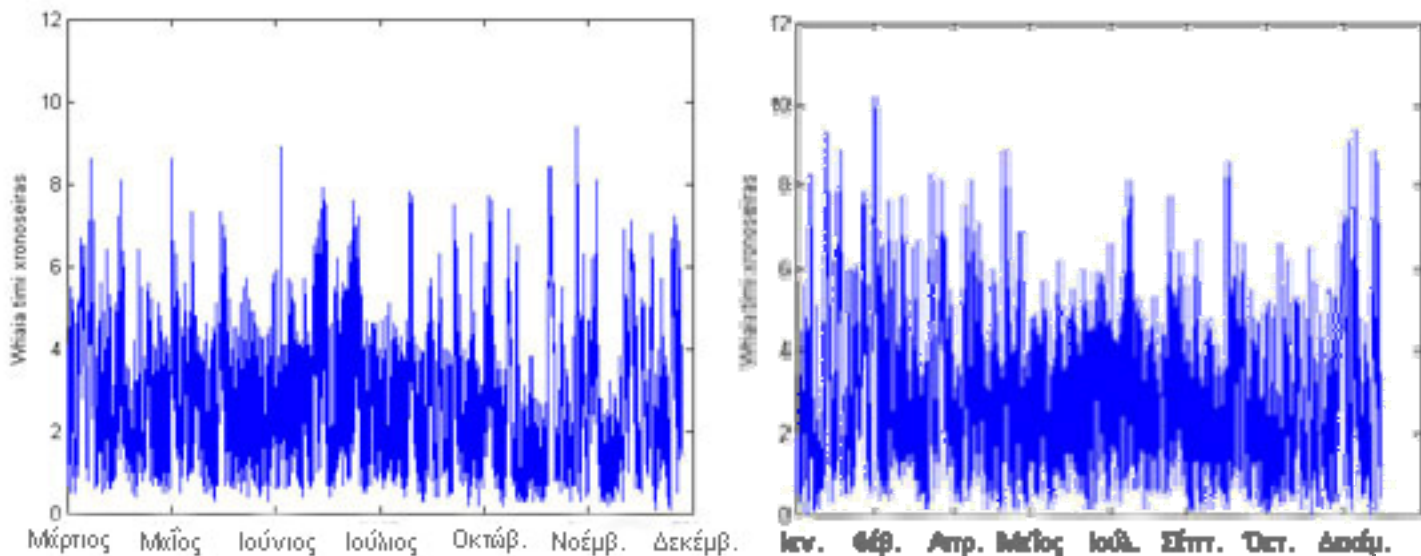


Στα διαγράμματα polar μπορούμε να διακρίνουμε το πλήθος εμφάνισης της κάθε ομάδας. Παρακάτω παραθέτουμε τα διαγράμματα polar για τα έτη 2004 και 2005 της διεύθυνσης ανέμου.



Σχήμα 28 & 29: Αριστερά φαίνεται η χρονοσειρά των τιμών WD για το έτος 2004 (Σχ.28), και δεξιά η χρονοσειρά των τιμών WD για το έτος 2005.

Από τις παραπάνω γραφικές παραστάσεις παρατηρούμε ότι δεν υπήρξαν αλλαγές στις διευθύνσεις των ανέμων για τα έτη 04-05.



Σχήμα 30 & 31: Αριστερά φαίνεται η χρονοσειρά των τιμών ταχύτητας ανέμου για το έτος 2004 (Σχ.30), και δεξιά η χρονοσειρά των τιμών ταχύτητας ανέμου για το έτος 2005.

Παρακάτω παραθέτουμε τον πίνακα με τα συμπεράσματα που βγάλαμε μετά από μελέτη των διαγραμμάτων κουτιού και των υπολοίπων χαρακτηριστικών.

	2004			2005		
	Θέση της διαμέσου (κέντρο;)	Ισομεγέθη “μουστάκια”	Υπαρξη ακραίων τιμών	Θέση της διαμέσου (κέντρο;)	Ισομεγέθη “μουστάκια”	Υπαρξη ακραίων τιμών
BENZ	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΝΑΙ
CO#	ΝΑΙ	ΟΧΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΝΑΙ
S02	ΟΧΙ	ΟΧΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΝΑΙ
METEO RH	ΟΧΙ	ΝΑΙ	ΟΧΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ
METEO Ta	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ	ΟΧΙ
METEO Wd _n	ΝΑΙ	ΝΑΙ	ΝΑΙ	ΟΧΙ	ΝΑΙ	ΝΑΙ
METEO WS	ΟΧΙ	ΟΧΙ	ΝΑΙ	ΟΧΙ	ΟΧΙ	ΝΑΙ

Πίνακας 6. Αποτελέσματα και συμπεράσματα διαγράμματος κουτιού.

3.3 Μελέτη συντελεστή συσχέτισης (correlation coefficient)

Ο συντελεστής συσχέτισης ή δείκτης συσχέτισης (correlation coefficient) είναι μια αριθμητική τιμή που κυμαίνεται μεταξύ του -1 (αρνητική συσχέτιση) και του +1 (θετική συσχέτιση). Αυτή η τιμή δείχνει το πόσο ισχυρή είναι η γραμμική εξάρτηση δύο μεταβλητών [1].

Ο υπολογισμός του δείκτη διαμόρφωσης μεταξύ δύο μεταβλητών γίνεται ως εξής [2]:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y},$$

Ο δείκτης συσχέτισης είναι ένας δείκτης ο οποίος αποτελείται από δύο στοιχεία [1]:

1. Από το θετικό ή το αρνητικό πρόσημο της αριθμητικής τιμής.
2. Από την ίδια την αριθμητική τιμή.

Το πρόσημο μας δίνει πληροφορίες για την κατεύθυνση της συσχέτισης. Πιο συγκεκριμένα, μας πληροφορεί για την κατεύθυνση που έχει η γραμμή παλινδρόμησης (regression line), δηλαδή η γραμμή που περνά όσο το δυνατόν πλησιέστερα από τα περισσότερα σημεία του διαγράμματος.

1. Αν το πρόσημο είναι θετικό, τότε η κλίση της γραμμής έχει μια πορεία από το κάτω αριστερό τμήμα προς το πάνω δεξιό τμήμα του διαγράμματος. Επίσης μας δείχνει ότι οι μεταβλητές μεταβάλλονται ως προς την ίδια κατεύθυνση (ομόρροπα).
2. Αν το πρόσημο είναι αρνητικό, η κλίση της γραμμής έχει μια πορεία από το πάνω αριστερό τμήμα προς το κάτω δεξιό τμήμα του διαγράμματος. Επίσης μας δείχνει ότι οι μεταβλητές μεταβάλλονται ως προς την αντίθετη κατεύθυνση (αντίρροπα).

Αν και δεν υπάρχουν κάποιοι γενικοί αποδεκτοί κανόνες σχετικά με το πότε ένας δείκτης συσχέτισης θεωρείται υψηλός, μέτριος ή χαμηλός, θα μπορούσαμε εμπειρικά να ερμηνεύσουμε το δείκτη συσχέτισης σύμφωνα με τον παρακάτω πίνακα:

Αν ο δείκτης είναι μικρότερος ή ίσος του $\pm 0,29$	Δεν υπάρχει συσχέτιση
Αν ο δείκτης κυμαίνεται μεταξύ $\pm 0,30$ και $\pm 0,49$	Χαμηλή συσχέτιση
Αν ο δείκτης κυμαίνεται μεταξύ $\pm 0,50$ και $\pm 0,69$	Μέτρια συσχέτιση
Αν ο δείκτης κυμαίνεται μεταξύ $\pm 0,70$ και $\pm 0,79$	Υψηλή συσχέτιση
Αν ο δείκτης κυμαίνεται μεταξύ $\pm 0,80$ και $\pm 0,99$	Πολύ υψηλή συσχέτιση

Για τον υπολογισμό του δείκτη συσχέτισης της χρονοσειράς Βενζολίου με τις αντίστοιχες τιμές όζοντος αλλά και με τις μετεωρολογικές παραμέτρους που είναι διαθέσιμες για τις ίδιες χρονικές περιόδους, χρησιμοποιήσαμε την συνάρτηση

coorcoef η οποία υπάρχει έτοιμη στο Matlab. Παρακάτω υπάρχει ο πίνακας με τα αποτελέσματα συσχέτισης:

Σταθμός Πατησίων	2004		2005	
	ρ: (δείκτης συσχέτισης)	Βαθμός συσχέτισης	ρ: (δείκτης συσχέτισης)	Βαθμός συσχέτισης
Βενζόλιο με όζον	-	-	-0.54373	Μέτρια συσχέτιση
Βενζόλιο με RH	0.15652	Δεν υπάρχει συσχέτιση	0.055649	Δεν υπάρχει συσχέτιση
Βενζόλιο με Ta	-0.091353	Δεν υπάρχει συσχέτιση	-0.018866	Δεν υπάρχει συσχέτιση
Βενζόλιο με WDi	-0.0010646	Δεν υπάρχει συσχέτιση	-0.0028954	Δεν υπάρχει συσχέτιση
Βενζόλιο με WS	-0.38766	Χαμηλή συσχέτιση	-0.44203	Χαμηλή συσχέτιση
Βενζόλιο με CO	0.80017	Πολύ υψηλή συσχέτιση	0.9114	Πολύ υψηλή συσχέτιση
Βενζόλιο με SO2	0.3285	Χαμηλή συσχέτιση	0.47364	Χαμηλή συσχέτιση

Πίνακας 7. Αποτελέσματα δείκτη συσχέτισης για έτη 2004-2005 (σταθμός Πατησίων).

Σύμφωνα με katroulis, 1996, η σχέση μεταξύ CO και WS στην περιοχή Πατησίων είναι:

$$CO = 10,7(WS)^{-0,99}$$

Επειδή από τον παραπάνω πίνακα βλέπουμε ότι υπάρχει “Πολύ υψηλή συσχέτιση” στα δεδομένα Βενζολίου και CO, θα χρησιμοποιήσουμε τον παραπάνω τύπο για πρόβλεψη τιμών Βενζολίου και CO.

Για τον έλεγχο των αποτελεσμάτων πρόβλεψης, θα χρησιμοποιήσουμε στατιστικές ποσότητες όπως τον δείκτη προσδιορισμού (R-Square) και το τυπικό σφάλμα εκτίμησης (Standard error of estimation).

- Ο δείκτης προσδιορισμού μας πληροφορεί για το κοινό ποσοστό της διακύμανσης που ερμηνεύουν οι δύο μεταβλητές X και Y τις οποίες μελετάμε. Για τον υπολογισμό του δείκτη προσδιορισμού (r^2) αρχικά μπορούμε να βρούμε τον δείκτη συσχέτισης (r) και στην συνέχεια να τον υψώσουμε εις το τετράγωνο [1].
- Κάθε προσπάθεια να προβλέψουμε το Y για μια συγκεκριμένη τιμή του X εμπεριέχει κάποιο σφάλμα [1]. Για το λόγο αυτό, επειδή δεν μπορούμε να υπολογίσουμε με ακρίβεια το Y, καταλήγουμε να υπολογίσουμε το \hat{Y} , το οποίο είναι η καλύτερη δυνατή εκτίμηση που μπορούμε να κάνουμε. Ένας εύκολος τρόπος για να βρούμε την τιμή του συνολικού σφάλματος που προκύπτει στην προσπάθεια μας να προβλέψουμε το Y από το X θα ήταν να προσθέσουμε όλες τις διαφορές ($Y - \hat{Y}$) και στη συνέχεια να διαιρέσουμε το

άθροισμα τους με τον αριθμό των διαφορών, ώστε να υπολογίσουμε το “μέσο όρο διαφορών” της συγκεκριμένης κατανομής. Επειδή όμως, όπως και στην περίπτωση του υπολογισμού της τυπικής απόκλισης, οι θετικές τιμές αλληλοακυρώνονται με τις αρνητικές τιμές (οπότε το αλγεβρικό τους αποτέλεσμα είναι 0), ο τρόπος αυτός δεν είναι σωστός. Η καλύτερη λύση είναι να προσθέσουμε αυτές τις αποστάσεις, αφού όμως πρώτα τις υψώσουμε στο τετράγωνο $(Y - \hat{Y})^2$ για να εξαλείψουμε τις αρνητικές τιμές, και στη συνέχεια να διαιρέσουμε του άθροισμα τους με το N (πιο συγκεκριμένα με το N-2, επειδή προσπαθούμε να υπολογίσουμε μια παράμετρο). Τέλος, υπολογίζοντας την τετραγωνική ρίζα αυτού του πηλίκου, μπορούμε να μετατρέψουμε την συγκεκριμένη μέση τιμή στην αρχική κλίμακα μέτρησης. Αυτή η εκτίμηση του σφάλματος που προκύπτει στην προσπάθεια μας να προβλέψουμε το Y από το X ονομάζεται τυπικό σφάλμα εκτίμησης και συμβολίζεται με $S_{X,Y}$ για να υποδηλώσει ότι ουσιαστικά πρόκειται για την τυπική απόκλιση του Y που έχει προβλεφθεί από το X. Ο τύπος με τον οποίο υπολογίζουμε το τυπικό σφάλμα εκτίμησης ($S_{X,Y}$) είναι:

$$S_{X,Y} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{N - 2}}$$

Παρακάτω παραθέτουμε τον πίνακα με τα αποτελέσματα του δείκτη προσδιορισμού αλλά και του τυπικού σφάλματος εκτίμησης, ως προς τις προβλεπόμενες τιμές (σχέση katroulis 1996) και τις πραγματικές τιμές.

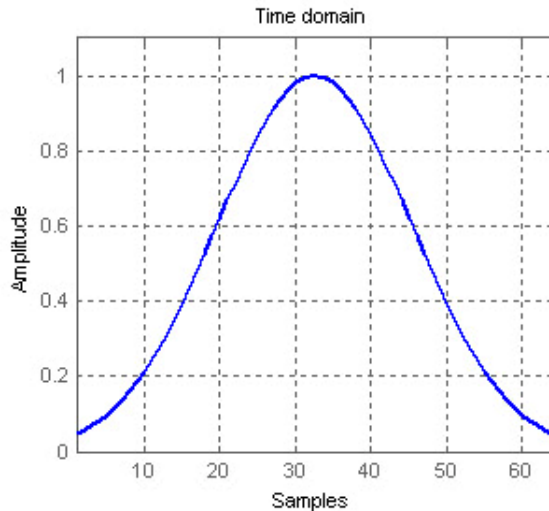
		Δείκτης προσδιορισμού (R-Square)	Τυπικό σφάλμα εκτίμησης
2004	Προβλ. Τιμές CO – Πραγμ. Τιμές CO	0,1858	6,1418
	Προβλ. Τιμές Βενζ. – Πραγμ. Τιμές Βενζ.	0,1056	6,6
2005	Προβλ. Τιμές CO – Πραγμ. Τιμές CO	0,1677	6,793
	Προβλ. Τιμές Βενζ. – Πραγμ. Τιμές Βενζ.	0,1192	5,339

Από τον παραπάνω πίνακα βλέπουμε ότι η σχέση katroulis 1996 δεν μας δίνει “καλά” αποτελέσματα για τα έτη 2004-2005.

Η πολύ υψηλή συσχέτιση μεταξύ βενζολίου και CO καταδεικνύει, την ομοιότητα που υπάρχει μεταξύ αυτών σε σχέση με τις εκπομπές τους. Έτσι, αναμένουμε ότι η κυκλοφορία (κύρια πηγή CO) είναι επίσης και η κύρια πηγή βενζολίου, για ένα σημείο με υψηλή πυκνότητα κυκλοφορίας όπως ο σταθμός Πατησίων. Εφόσον μάλιστα ληφθούν υπόψη και τα αποτελέσματα της εκτίμησης εκπομπών των Kassomenos et.al 2006, τότε συνάγεται ότι τα στοιχεία traffic time και traffic speed (επηρεάζουν CO) και τα στοιχεία (motorcycles, μη καταλυτικά, traffic speed) οδηγούν στο συμπέρασμα πως η ταχύτητα κίνησης ενδέχεται να είναι ο κρίσιμος μηχανισμός που επηρεάζει τις εκπομπές βενζολίου και CO στο κέντρο της πόλης. Επομένως, μια πολιτική κυκλοφοριακών ρυθμίσεων που θα αποσκοπούσε στην μεταβολή της ταχύτητας κίνησης, αναμένεται ότι θα ανακούφιζε, τοπικά την ατμοσφαιρική φόρτιση.

3.4 Μελέτη κανονικής κατανομής

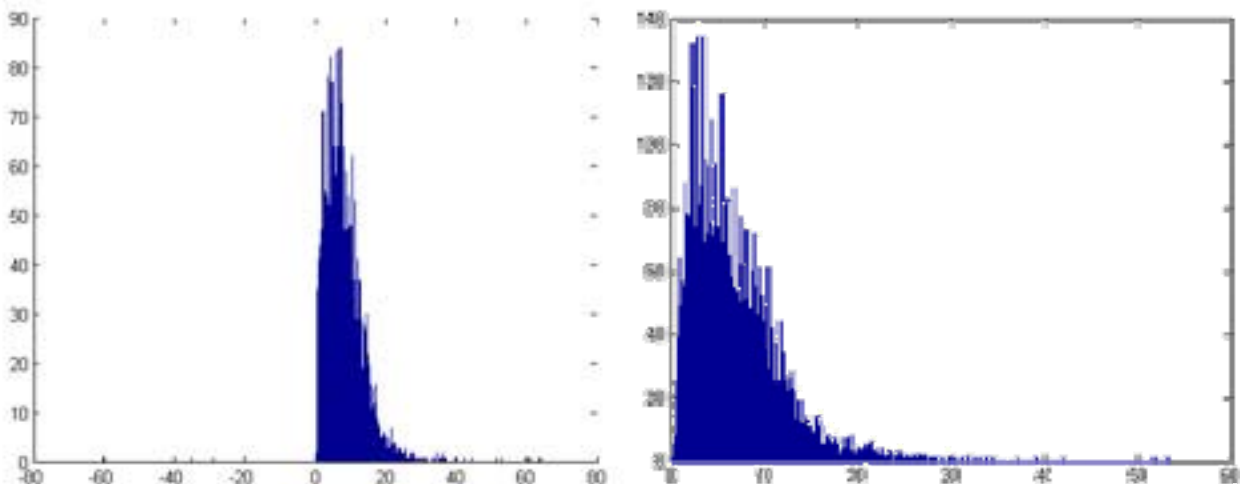
Η μελέτη της κανονικής κατανομής έχει μια μακρά ιστορία, η οποία ξεκινά από το 18^ο αιώνα. Η εν λόγω κατανομή ορίστηκε με ακρίβεια από τον Pierre-Simon Laplace και πήρε την πιο συνήθη μορφή της από τον Carl Friedrich Gauss (πολύ συχνά άλλωστε ονομάζεται και καμπύλη του Gauss).



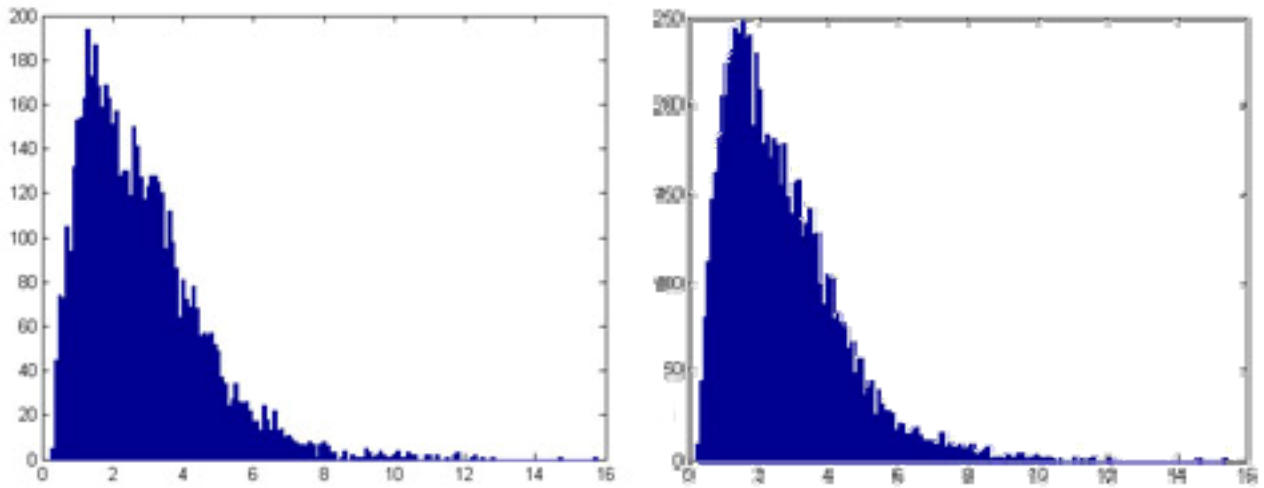
Αριστερά δείχνουμε ένα παράδειγμα μιας κανονικής κατανομής (Σχήμα 26). Ο όρος “κανονική κατανομή” είναι ένας τεχνικός όρος και με κανέναν τρόπο δεν θα πρέπει να θεωρούμε όλες τις άλλες κατανομές μεταβλητών μη κανονικές. Πρόκειται για μία κωδωνοειδή συμμετρική κατανομή, όπου η τετμημένη (οριζόντιος άξονας – X-axis) αναπαριστά διαφορετικές πιθανές τιμές την μεταβλητής που μετρήθηκε (X), ενώ η τεταγμένη (κάθετος άξονας – Y-axis)

αναπαριστά τη συχνότητα ή την πιθανότητα εμφάνισης των τιμών του X.

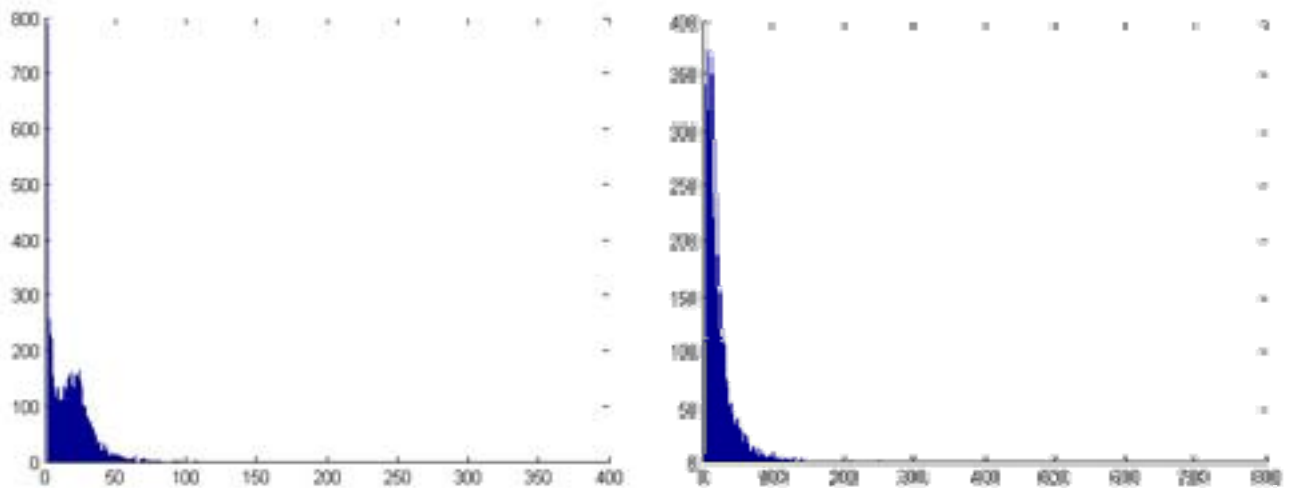
Με την βοήθεια του Matlab και χρησιμοποιώντας την συνάρτηση hist(), δημιουργήσαμε τις παρακάτω γραφικές παραστάσεις, στις οποίες μπορούμε να δούμε εάν τα δεδομένα μας έχουν κάποια “κανονική κατανομή”.



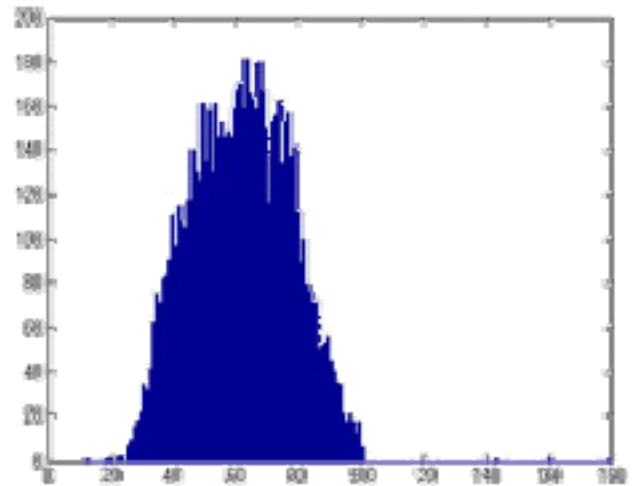
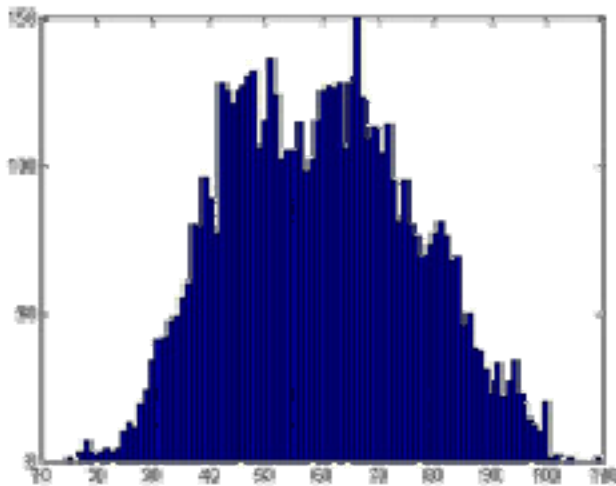
Σχήμα 32 & 33. Αριστερά η γραφική παράσταση για την μελέτη της “κανονικής κατανομής” των τιμών βενζολίου για το έτος 2004 και δεξιά (Σχήμα 33) για το έτος 2005.



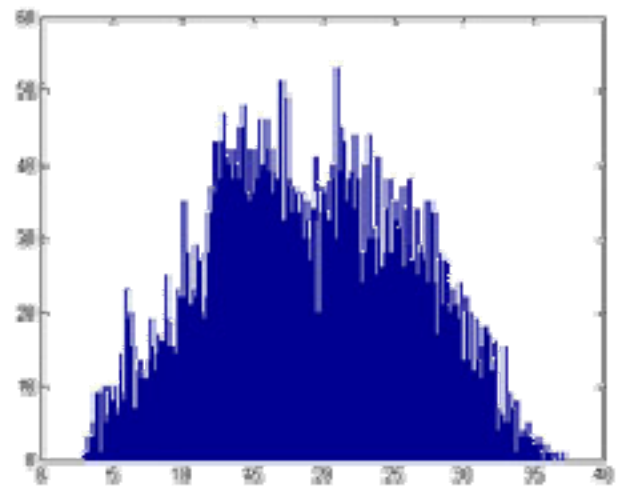
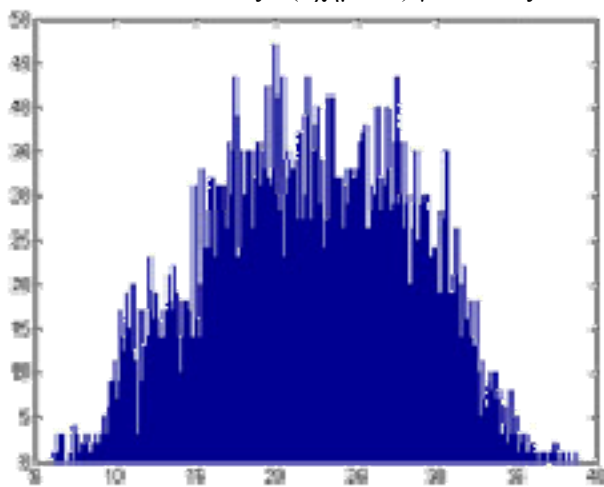
Σχήμα 34 & 35. Αριστερά η γραφική παράσταση για την μελέτη της “κανονικής κατανομής” των τιμών μονοξειδίου του άνθρακα (CO#) για το έτος 2004 και δεξιά (Σχημά 35) για το έτος 2005.



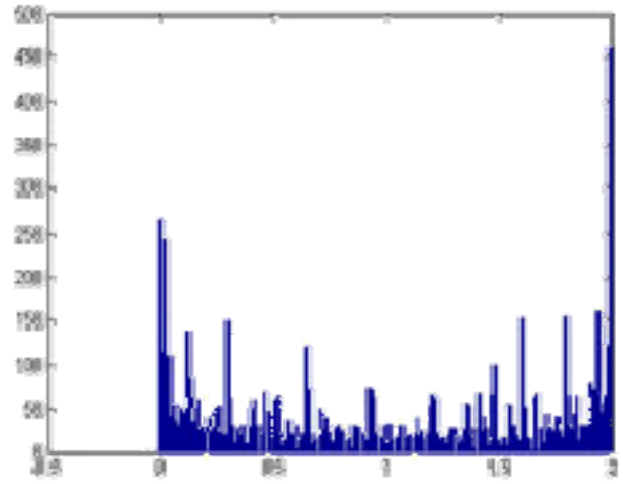
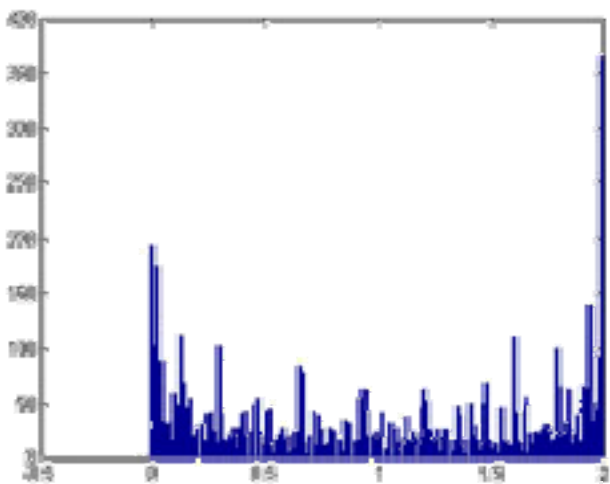
Σχήμα 36 & 37. Αριστερά η γραφική παράσταση για την μελέτη της “κανονικής κατανομής” των τιμών διοξειδίου του θείου (SO2) για το έτος 2004 και δεξιά (Σχημά 337) για το έτος 2005.



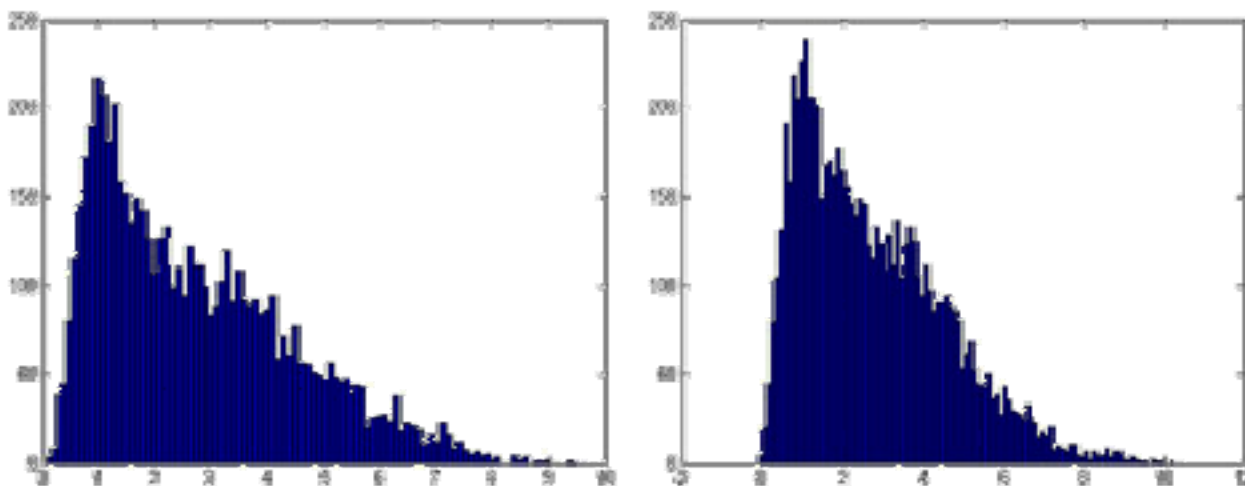
Σχήμα 38 & 39. Αριστερά η γραφική παράσταση για την μελέτη της “κανονικής κατανομής” των τιμών των μετεωρολογικών στοιχείων, σχετικής υγρασίας (RH) για το έτος 2004 και δεξιά (Σχημά 39) για το έτος 2005.



Σχήμα 40 & 41. Αριστερά η γραφική παράσταση για την μελέτη της “κανονικής κατανομής” των τιμών των μετεωρολογικών στοιχείων, θερμοκρασίας (Ta) για το έτος 2004 και δεξιά (Σχημά 41) για το έτος 2005.



Σχήμα 42 & 43. Αριστερά η γραφική παράσταση για την μελέτη της “κανονικής κατανομής” των τιμών των μετεωρολογικών στοιχείων, WdN για το έτος 2004 και δεξιά (Σχημά 43) για το έτος 2005.



Σχήμα 44 & 45. Αριστερά η γραφική παράσταση για την μελέτη της “κανονικής κατανομής” των τιμών των μετεωρολογικών στοιχείων, ταχύτητας ανέμου (WS) για το έτος 2004 και δεξιά (Σχημά 45) για το έτος 2005.

Επίσης για την μελέτη της κανονικής κατανομής των δεδομένων μας χρησιμοποιήσαμε τις παρακάτω δοκιμές, οι συναρτήσεις των οποίων υπάρχουν στο Matlab:

- Kolmogorov-Smirnov: kstest()
- Lilliefors: lillietest()
- Jarque-Bera: jbttest()

Οι παραπάνω συναρτήσεις παίρνουν ως όρισμα, τις τιμές των δεδομένων μας και τις συγκρίνουν με μία πρότυπη κανονική κατανομή (η οποία έχει μέση τιμή 0 και διασπορά 1). Η έξοδοι των συναρτήσεων παίρνουν την τιμή 1 εάν μπορούμε να απορρίψουμε την υπόθεση ότι τα δεδομένα εισόδου έχουν κανονική κατανομή, και τιμή 0 εάν δεν μπορούμε να απορρίψουμε αυτή την υπόθεση. Απορρίπτουμε την υπόθεση εάν η δοκιμή είναι ενδεικτική σε βαθμό 5%.

	2004			2005		
	Kolmogorov-Smirnov	Lilliefors	Jarque-Bera	Kolmogorov-Smirnov	Lilliefors	Jarque-Bera
BENZ	1	1	1	1	1	1
CO#	1	1	1	1	1	1
S02	1	1	1	1	1	1
METEO RH	1	1	1	1	1	1
METEO Ta	1	1	1	1	1	1
METEO WDe	1	1	1	1	1	1
METEO WS	1	1	1	1	1	1

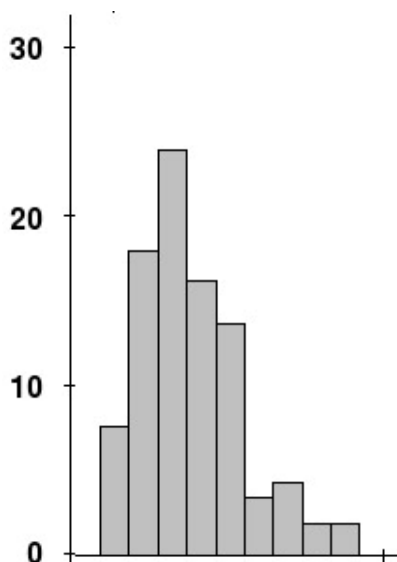
Από τον παραπάνω πίνακα βλέπουμε ότι τα δεδομένα μας δεν έχουν κανονική κατανομή.

3.5 Μελέτη εκτροπής (Skew)

Η εκτροπή (Skew) έχει οφέλη σε πολλές περιοχές [4]. Πολλά απλοϊκά πρότυπα υποθέτουν ότι υπάρχει κανονική κατανομή, δηλαδή ότι τα στοιχεία είναι συμμετρικά ως προς το μέσο όρο. Αλλά στην πραγματικότητα, τα σημεία των στοιχείων δεν είναι τέλεια συμμετρικά. Η κατανόηση της εκτροπής του συνόλου των δεδομένων μας δείχνει εάν οι αποκλίσεις από το μέσο όρο πρόκειται να είναι θετικές ή αρνητικές.

- Θετική εκτροπή (Positive Skew): Όταν έχουμε θετική εκτροπή, στα δεξιά θα έχουμε μεγαλύτερη “ουρά”.
- Αρνητική εκτροπή (Negative Skew): Όταν έχουμε αρνητική εκτροπή, στα αριστερά θα έχουμε μεγαλύτερη “ουρά”.
- Το μέγεθος της εκτροπής μας δείχνει το μέγεθος της “ουράς”.

Για παράδειγμα στην παρακάτω γραφική παράσταση μελέτης της κανονικής κατανομής, βλέπουμε ότι στα δεξιά έχουμε μεγαλύτερη ουρά (θετική εκτροπή).



Ο τύπος για την εύρεση της εκτροπής είναι ο παρακάτω:

$$skew(x) = \frac{\mu_3}{\sigma^3} = \frac{m_3}{(m_2)^{3/2}} = \frac{\sqrt{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^3}{\sum_{i=1}^n ((x_i - \bar{x})^2)^{3/2}}$$

Στον παρακάτω πίνακα εμφανίζουμε τα αποτελέσματα χρήσης της συνάρτησης skewness() του Matlab στις χρονοσειρές των δεδομένων μας για τα έτη 2004 και 2005.

	2004		2005	
	Τιμή Εκτροπής	Μεγαλύτερη “ουρά”	Τιμή Εκτροπής	Μεγαλύτερη “ουρά”
BENZ	1.2774	δεξιά	1.7040	δεξιά
CO#	1.4595	δεξιά	1.3899	δεξιά
S02	4.9934	δεξιά	6.6167	δεξιά
METEO RH	0.1572	δεξιά	0.1136	δεξιά
METEO Ta	-0.0784	αριστερά	0.0294	δεξιά
METEO WDn	-0.0243	αριστερά	-0.0375	αριστερά
METEO WS	0.7825	δεξιά	0.8195	δεξιά

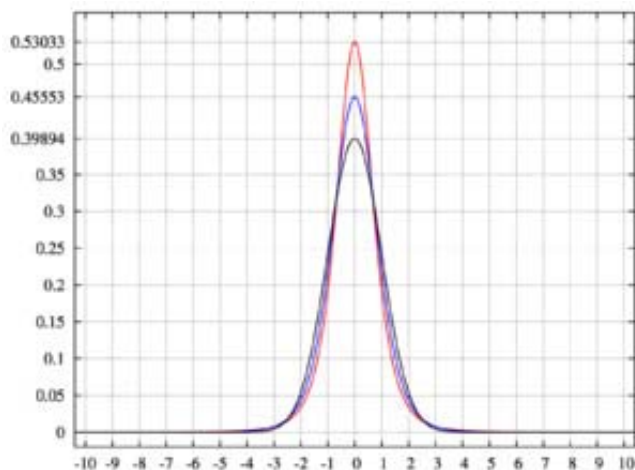
Πίνακας 8. Αποτελέσματα εκτροπής των δεδομένων για τα έτη 2004-2005 (σταθμός Πατησίων).

3.6 Μελέτη κύρτωσης (Kurtosis)

Όταν έχουμε υψηλή κύρτωση, έχουμε “αιχμηρότερες κορυφές”, και “παχύτερες ουρές”, ενώ με χαμηλή κύρτωση έχουμε πιο στρογγυλοποιημένες κορυφές και ευρύτερους “ώμους” [5].

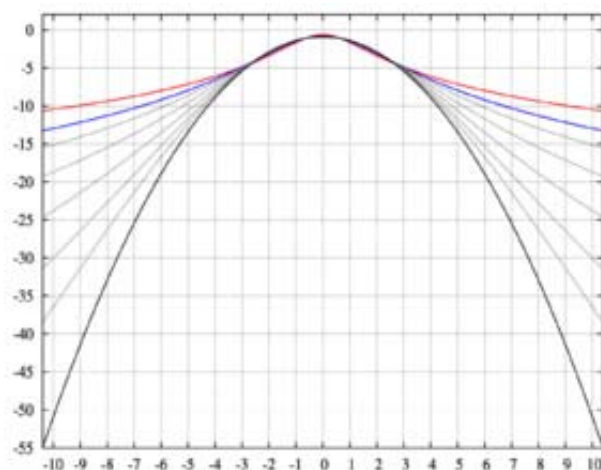
- Η θετική κύρτωση ονομάζεται leptokurtic: στην οποία έχουμε οξύτερες “κορυφές” γύρω από το μέσο όρο (υπάρχει μεγάλη πιθανότητα ότι μία κανονικά κατανομημένη μεταβλητή να παίρνει τιμές κοντά στο μέσο όρο) και “παχιές ουρές” (υπάρχει μεγάλη πιθανότητα μια κανονικά κατανομημένη μεταβλητή να παίρνει ακραίες τιμές).

Π.χ.



- Η αρνητική κύρτωση ονομάζεται platykurtic: στην οποία έχουμε μικρότερες “κορυφές” γύρω από τον μέσο όρο (υπάρχει μικρότερη πιθανότητα ότι μία κανονικά κατανομημένη μεταβλητή να παίρνει τιμές γύρω από το μέσο όρο) και λεπτές “ουρές” (υπάρχει μικρότερη πιθανότητα μια κανονικά κατανομημένη μεταβλητή να παίρνει ακραίες τιμές).

Π.χ.



Ο τύπος με τον οποίο υπολογίζουμε την κύρτωση είναι ο παρακάτω:

$$kurt(x) = \frac{m_4}{m_2^2} - 3 = \frac{n \cdot \sum_{i=1}^n (x_i - \bar{x})^4}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^2} - 3$$

Στον παρακάτω πίνακα εμφανίζουμε τα αποτελέσματα χρήσης της συνάρτησης *kurtosis* () του Matlab στις χρονοσειρές των δεδομένων μας για τα έτη 2004 και 2005.

	2004	2005
BENZ	17.7147	8.8319
CO#	6.6853	6.0278
S02	45.2707	97.2379
METEO RH	2.3631	2.9179
METEO Ta	2.2576	2.2286
METEO WDn	1.5126	1.4929
METEO WS	2.9418	3.2519

Πίνακας 9. Αποτελέσματα κύρτωσης των δεδομένων για τα έτη 2004-2005 (σταθμός Πατησίων).

Από το παραπάνω πίνακα μπορούμε να διακρίνουμε ότι όλα μας τα δεδομένα έχουν θετική κύρτωση (*leptokurtic*).

3.7 Συνδιακύμανση – Συμμεταβλητότητα (covariance)

Ο δείκτης συσχέτισης χρησιμοποιείται στον υπολογισμό του βαθμού στον οποίο δύο μεταβλητές συμμεταβάλλονται βασίζεται στη στατιστική έννοια της συνδιακύμανσης (covariance). Η συνδιακύμανση είναι ουσιαστικά ένας αριθμός που αντικατοπτρίζει το βαθμό στον οποίο δύο μεταβλητές μεταβάλλονται επηρεάζοντας η μία την άλλη (συμμεταβάλλονται).

Όταν έχουμε υψηλή θετική συνδιακύμανση, τότε έχουμε υψηλή θετική συσχέτιση των δύο μεταβλητών (μεταβάλλονται με τον ίδιο τρόπο). Όταν έχουμε υψηλή αρνητική συνδιακύμανση, τότε έχουμε υψηλή αρνητική συσχέτιση των δύο μεταβλητών (μεταβάλλονται με τον αντίστροφο τρόπο). Αν δεν υπάρχει συσχέτιση της μίας μεταβλητής με την άλλη, τότε η συνδιακύμανση είναι 0.

Ο τύπος μέσω του οποίου υπολογίζουμε την συνδιακύμανση δυο μεταβλητών X και Y είναι ο εξής:

$$Cov_{XY} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{N}$$

Στον παρακάτω πίνακα φαίνονται τα αποτελέσματα της συνδιακύμανσης των τιμών βενζολίου με τα υπόλοιπα χαρακτηριστικά ανά έτος. Για τον υπολογισμό των παρακάτω τιμών χρησιμοποιήθηκε η συνάρτηση cov() του Matlab.

	2004	2004 (%)	2005	2005(%)
Βενζόλιο με CO	7.3824	11	7.1475	9,3
Βενζόλιο με SO2	40.5245	60	61.5329	79,7
Βενζόλιο με RH	13.7737	20	4.1642	5,4
Βενζόλιο με Ta	-2.9822	4	-0.6281	0,8
Βενζόλιο με WDn	-0.0040	0	-0.0097	0,0
Βενζόλιο με WS	-3.5233	5	-3.7368	4,8

Πίνακας 10. Αποτελέσματα συνδιακύμανσης των δεδομένων για τα έτη 2004-2005 (σταθμός Πατησίων).

3.8 Principal Component Analysis (P.C.A)

3.8.1 Εισαγωγή

Η Principal Component Analysis (P.C.A.) [6] είναι μια μέθοδος αναγνώρισης κάποιων προτύπων μέσα στα δεδομένα μας, και έκφρασης τους με τέτοιο τρόπο έτσι ώστε να τονίσουμε τις ομοιότητες και τις διαφορές τους. Λόγω της δυσκολίας εύρεσης προτύπων, σε δεδομένα πολλών διαστάσεων, (όπου δεν έχουμε την πολυτέλεια των γραφικών αναπαραστάσεων των δεδομένων μας) η μέθοδος PCA είναι ένα πολύ χρήσιμο εργαλείο για ανάλυση δεδομένων.

Ένα άλλο βασικό πλεονέκτημα της μεθόδου PCA είναι ότι από την στιγμή που βρεθούν αυτά τα πρότυπα στα δεδομένα μας, μπορούμε να συμπίεσουμε τα δεδομένα μας, απλά μειώνοντας τον αριθμό των διαστάσεων χωρίς να έχουμε μεγάλη απώλεια πληροφοριών.

Παρακάτω δείχνουμε τα βήματα της μεθόδου PCA:

1. Αφαιρούμε τον μέσο όσο από κάθε διάσταση των δεδομένων (Data Adjust), έτσι ώστε τώρα ο μέσος όρος των δεδομένων της διάστασης να είναι μηδέν.

Για παράδειγμα:

	x	y		x	y
	2.5	2.4		.69	.49
	0.5	0.7		-1.31	-1.21
	2.2	2.9		.39	.99
	1.9	2.2		.09	.29
Data =	3.1	3.0	DataAdjust =	1.29	1.09
	2.3	2.7		.49	.79
	2	1.6		.19	-.31
	1	1.1		-.81	-.81
	1.5	1.6		-.31	-.31
	1.1	0.9		-.71	-1.01
	$\bar{x}=1,81$			$\bar{x}=0$	
	$\bar{y}=1,91$			$\bar{y}=0$	

2. Υπολογίζουμε τον πίνακα συνδιακύμανσης (covariance matrix) των δεδομένων μας, όπως έχει προαναφερθεί στο προηγούμενο κεφάλαιο. Για τα παραπάνω δεδομένα περιμένουμε ένα πίνακα συνδιακύμανσης 2 x 2 επειδή έχουν δύο διαστάσεις (x και y). Αυτό συμβαίνει γιατί υπολογίζουμε την συνδιακύμανση για κάθε συνδυασμό των διαστάσεων μας.

$$\text{COV} = \begin{pmatrix} \text{COV}(x, x) & \text{COV}(x, y) \\ \text{COV}(y, x) & \text{COV}(y, y) \end{pmatrix}$$

Παρακάτω παραθέτουμε τον πίνακα συνδιακύμανσης ο οποίος έχει υπολογιστεί από την συνάρτηση `cov()` του Matlab για τα παραπάνω δεδομένα.

$$\text{cov} = \begin{pmatrix} 0,61656 & 0,61544 \\ 0,61544 & 0,71656 \end{pmatrix}$$

Από τον παραπάνω πίνακα συνδιακύμανσης βλέπουμε ότι $\text{cov}(x,y) = \text{cov}(y,x)$.

3. Σε αυτό το βήμα θα πρέπει να υπολογίσουμε τα *eigenvectors* και τα *eigenvalues* (βαρύτητα των *eigenvectors*) του πίνακα συνδιακύμανσης. Πριν όμως από αυτό, θα κάνουμε μια μικρή εισαγωγή στις έννοιες αυτών.

Όπως ξέρουμε μπορούμε να πολλαπλασιάσουμε δύο πίνακες, υπό τον όρο ότι έχουν κατάλληλα μεγέθη. Τα *eigenvectors* είναι μια ειδική κατάσταση ενός τέτοιου πολλαπλασιασμού. Παρακάτω δείχνουμε δύο πολλαπλασιασμούς μεταξύ ενός πίνακα και ενός διανύσματος.

a. $\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 11 \\ 5 \end{pmatrix}$

b. $\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} = 4 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix}$

Στο πρώτο παράδειγμα, το αποτέλεσμα είναι ένα διάνυσμα. Ενώ στο δεύτερο παράδειγμα, το αποτέλεσμα είναι ακριβώς τέσσερις φορές μεγαλύτερο από το διάνυσμα που ξεκινήσαμε. Το διάνυσμα $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$ του δεύτερου παραδείγματος

αντιπροσωπεύει μια γραμμή από την αρχή των αξόνων (0,0) στο σημείο (3,2). Τον τετραγωνικό πίνακα μπορούμε να τον θεωρήσουμε ως ένα πίνακα μετασχηματισμού. Εάν πολλαπλασιάσουμε αυτόν τον πίνακα με ένα διάνυσμα, θα πάρουμε ένα διάνυσμα το οποίο έχει μετασχηματιστεί από την αρχική του θέση.

Τα *eigenvectors* προέρχονται από την φύση του μετασχηματισμού. Ας φανταστούμε έναν πίνακα μετασχηματισμού που όποτε τον πολλαπλασιάζουμε, δημιουργούνται διανύσματα στην γραμμή $y=x$. Αυτό το διάνυσμα (και όλα τα πολλαπλάσια του, γιατί δεν μας ενδιαφέρει το μήκος που θα έχει), θα είναι ένα *eigenvector* του πίνακα μετασχηματισμού.

Τα *eigenvectors* έχουν κάποιες ιδιότητες, τις οποίες αναφέρουμε παρακάτω:

- *Eigenvectors* μπορούμε να υπολογίσουμε για τετραγωνικούς πίνακες.
- Δεν έχουν όλοι οι τετραγωνικοί πίνακες *eigenvectors*.
- Εάν ένας τετραγωνικός πίνακας ($n \times n$) έχει *eigenvectors*, τότε θα έχει n *eigenvectors*. Για παράδειγμα εάν μας είχε δοθεί ένας πίνακας 3×3 θα είχαμε 3 *eigenvectors*.
- Εάν αλλάξω τις αναλογίες του διανύσματος πριν τον πολλαπλασιασμό, πάλι θα έχω τα ίδια αποτελέσματα. Για παράδειγμα:

$$2 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 6 \\ 4 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 6 \\ 4 \end{pmatrix} = \begin{pmatrix} 24 \\ 16 \end{pmatrix} = 4 \times \begin{pmatrix} 6 \\ 4 \end{pmatrix}$$

Αυτό συμβαίνει γιατί αλλάζοντας την αναλογία του διανύσματος, το μόνο που καταφέραμε ήταν να αυξήσουμε τον μήκος του.

- Όλα τα eigenvectors ενός πίνακα, είναι κάθετα μεταξύ τους. Αυτό είναι σημαντικό γιατί μπορούμε αντί να εκφράσουμε τα δεδομένα μας στους άξονες x και y, να τα εκφράσουμε ως προς τα eigenvectors.
- Επίσης όταν υπολογίζουμε τα eigenvectors θέλουμε να έχουν μήκος ακριβώς ένα. Αυτό συμβαίνει γιατί το μήκος ενός διανύσματος ή ενός eigenvector δεν επηρεάζει την διεύθυνση. Έτσι όλα τα eigenvectors θα έχουν το ίδιο μήκος. Για το δικό μας παράδειγμα, είχαμε το εξής eigenvector:

$$\begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

Το οποίο έχει μήκος:

$$\sqrt{3^2 + 2^2} = \sqrt{13}$$

Έτσι για να έχει μήκος ένα, διαιρούμε το διάνυσμα με το $\sqrt{13}$:

$$\begin{pmatrix} 3 \\ 2 \end{pmatrix} \div \sqrt{13} = \begin{pmatrix} 3/\sqrt{13} \\ 2/\sqrt{13} \end{pmatrix}$$

Για τον υπολογισμό των eigenvectors που έχουν μικρούς πίνακες (μέχρι 3 x 3) μπορούμε σχεδόν εύκολα να τα υπολογίσουμε. Όμως για μεγαλύτερους πίνακες θα πρέπει να χρησιμοποιήσουμε κάποιες βιβλιοθήκες οι οποίες θα υπολογίζουν τα eigenvalues.

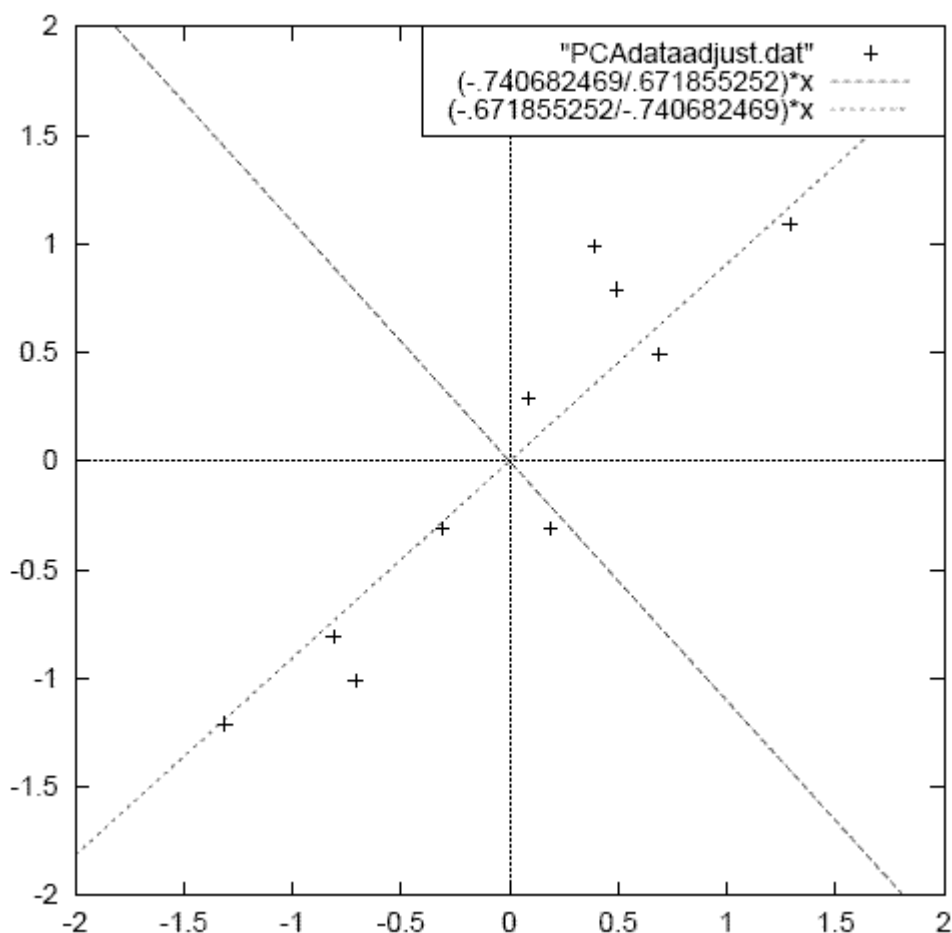
Τα eigenvalues είναι άμεσα συνδεδεμένα με τα eigenvectors. Στην πραγματικότητα τα eigenvalues και τα eigenvectors τα υπολογίζουμε μαζί. Στα παραπάνω παραδείγματα το eigenvalue ήταν τέσσερα, το οποίο το αναφέρουμε και ως Principal Component (P.C). Επίσης μπορούμε να διακρίνουμε ότι και στα δύο παραδείγματα το eigenvalue ήταν τέσσερα (ακόμα και όταν αλλάξαμε την αναλογία του διανύσματος μας).

Παρακάτω παραθέτουμε τα eigenvectors και τα eigenvalues του παραδείγματος μας.

$$eigenvalues = \begin{pmatrix} 0,490833989 \\ 1,28402771 \end{pmatrix}$$

$$eigenvectors = \begin{pmatrix} -0,735178656 & -0,677873399 \\ 0,677873399 & -0,735178656 \end{pmatrix}$$

Αρχικά παρατηρούμε ότι τα eigenvectors έχουν μήκος ένα. Στη συνέχεια παρατηρώντας την παρακάτω γραφική παράσταση βλέπουμε ότι τα δεδομένα έχουν ισχυρή σχέση μεταξύ τους. Όπως περιμέναμε και από τον πίνακα συνδιακύμανσης, οι δύο μεταβλητές όντως αυξάνονται μαζί. Στην παρακάτω γραφική παράσταση απεικονίζονται επίσης και τα eigenvectors (με διακεκομμένες γραμμές). Όπως αναφέραμε και προηγουμένως είναι κάθετες μεταξύ τους. Το πιο σημαντικό είναι ότι μας παρέχουν πληροφορίες για κάποιο πρότυπο που υπάρχει στα δεδομένα. Παρατηρούμε ότι το ένα eigenvector περνάει από την μέση των σημείων. Το οποίο μας δείχνει το πώς τα δύο δεδομένα μας (x και y) συνδέονται μέσω της γραμμής. Το δεύτερο eigenvector μας δίνει άλλες πληροφορίες (όχι τόσο σημαντικές) για κάποιο πρότυπο στα δεδομένα μας. Μας πληροφορεί ότι όλα τα σημεία ακολουθούν την βασική γραμμή αλλά με κάποια απόσταση από αυτή.



Σχήμα 46. Τα δεδομένα μας, στα οποία έχει αφαιρεθεί ο μέσος όρος (Adjusted Data), και επίσης τα eigenvectors και eigenvalues (με διακεκομμένες γραμμές).

Με αυτή την διαδικασία της ανεύρεσης των eigenvectors από τον πίνακα συνδιακύμανσης, ήμαστε σε θέση να δημιουργήσουμε γραμμές οι οποίες χαρακτηρίζουν τα δεδομένα μας.

4. Σε αυτό το βήμα πετυχαίνουμε συμπίεση των δεδομένων μας και την μείωση των διαστάσεων τους. Παρατηρώντας τα eigenvectors και τα eigenvalues του προηγούμενου βήματος βλέπουμε ότι τα eigenvalues έχουν πολύ διαφορετικές τιμές. Στην πραγματικότητα τα eigenvalues με τις υψηλότερες τιμές είναι οι κύριες συνιστώσες (Principal Components) των δεδομένων μας.

Στο παράδειγμα μας τα eigenvector με το υψηλότερο eigenvalue είναι η γραμμή που περνά από τη μέση των σημείων. Είναι η πιο σημαντική σχέση μεταξύ των δεδομένων μας. Συνήθως ταξινομούμε τα eigenvectors βάση των eigenvalues με αύξουσα σειρά (από τις μεγαλύτερες προς τις μικρότερες τιμές) έτσι ώστε να είναι ταξινομημένα σε σειρά “σημαντικότητας”.

Τώρα μπορούμε να απορρίψουμε τις συνιστώσες με χαμηλή “σημαντικότητα” (χαμηλό eigenvalue). Βεβαίως θα χάσουμε κάποιες πληροφορίες αλλά εάν τα eigenvalues είναι μικρές τιμές δεν θα έχουμε σημαντικές απώλειες. Εάν απορρίψουμε κάποιες συνιστώσες, τα τελικά δεδομένα θα έχουν λιγότερες διαστάσεις από τα αρχικά. Συγκεκριμένα, εάν αρχικά είχαμε n διαστάσεις και έτσι έχουμε υπολογίσει n eigenvectors και eigenvalues, εάν στην συνέχεια αποφασίζαμε να κρατήσουμε τα πρώτα p eigenvectors, τα τελικά μας δεδομένα θα έχουν p διαστάσεις.

Έτσι αρχικά θα πρέπει να δημιουργήσουμε το feature vector. Το οποίο είναι στην ουσία ένας πίνακας με τα eigenvectors τα οποία θέλουμε να κρατήσουμε από την λίστα των eigenvectors. Παρακάτω δείχνουμε πως γίνεται αυτό:

$$\text{FeatureVector} = (\text{eig}_1 \text{ eig}_2 \text{ eig}_3 \dots \text{eig}_n)$$

Στο δικό μας παράδειγμα έχουμε δεδομένα δύο διαστάσεων, δηλαδή με δύο eigenvectors. Έτσι είτε θα δημιουργήσουμε ένα feature vector και με τα δύο eigenvectors:

$$\begin{pmatrix} -0,677873399 & -0,735178656 \\ -0,735178656 & 0,677873399 \end{pmatrix}$$

ή θα επιλέξουμε να κρατήσουμε αυτά με την μεγαλύτερη “σημαντικότητα”.

$$\begin{pmatrix} -0,677873399 \\ -0,735178656 \end{pmatrix}$$

5. Δημιουργούμε την μορφή των δεδομένων έτσι ώστε να εκφράζονται βάση των γραμμών που δείχνουν τα eigenvectors (Principal Components). Χρησιμοποιώντας την παρακάτω σχέση:

$$\text{Final Data} = \text{RowFeatureVector} \times \text{RowDataAdjust}$$

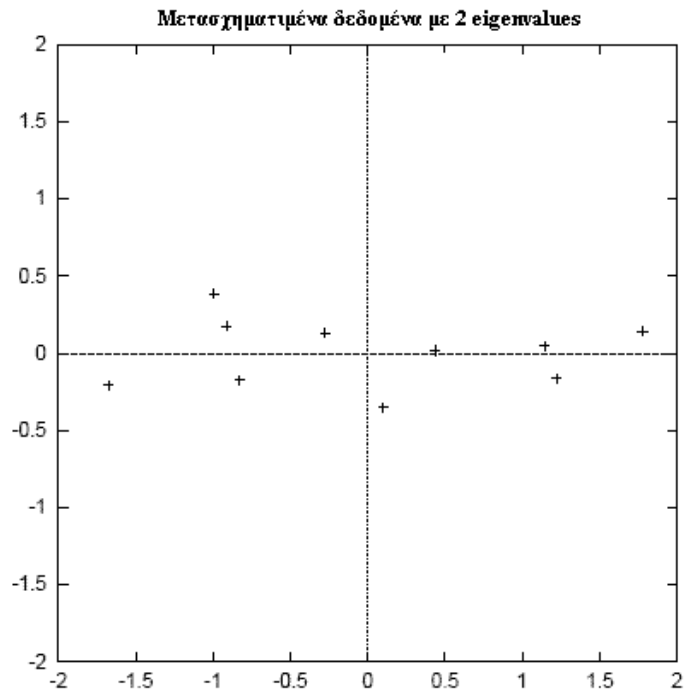
Τα αρχικά μας δεδομένα είχαν δύο άξονες (x και y), έτσι τα εκφράζαμε βάση αυτών. Είναι δυνατόν να εκφράσουμε τα δεδομένα μας σε δύο οποιουσδήποτε άξονες επιθυμούμε (εάν φυσικά είναι κάθετοι μεταξύ τους). Αυτός ήταν και ο λόγος που τα eigenvectors έπρεπε να ήταν κάθετα μεταξύ τους. Έτσι μετατρέψαμε τα δεδομένα μας, και τώρα δεν εκφράζονται βάση των αξόνων x και y αλλά από τους άξονες των eigenvectors. Στην περίπτωση που τα νέα δεδομένα είχαν μειωμένες διαστάσεις (είχαμε απορρίψει κάποια eigenvectors), τα νέα δεδομένα θα εκφράζονται βάση των διανυσμάτων που επιλέξαμε να κρατήσουμε. Στην συνέχεια χρησιμοποιούμε τον παραπάνω τύπο για να βρούμε τα νέα δεδομένα την μία με ολόκληρο το feature vector, και την άλλη με το feature vector στο οποίο κρατήσαμε μόνο τα σημαντικά eigenvectors.

Στη περίπτωση όπου χρησιμοποιήσαμε ολόκληρο το feature vector (και τα δύο eigenvectors) πήραμε τα αρχικά μας δεδομένα τα οποία έχουν

περιστραφεί στους άξονες των eigenvectors, όπως φαίνεται και στο παρακάτω σχήμα.

Μετασχηματισμένα
Δεδομένα =

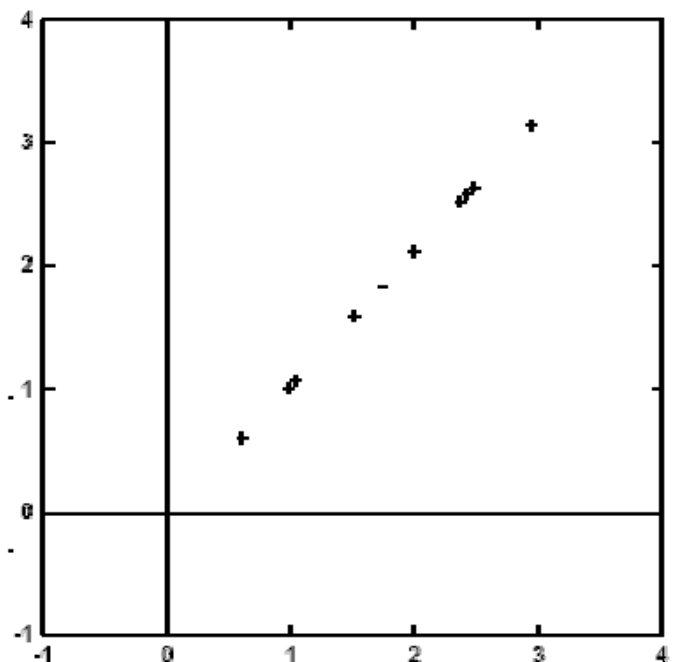
x	y
-827970186	-175115307
1.77758033	.142857227
-.992197494	.384374989
-.274210416	.130417207
-1.67580142	-.209498461
-.912949103	.175282444
.0991094375	-.349824698
1.14457216	.0464172582
.438046137	.0177646297
1.22382056	-.162675287



Στην περίπτωση όπου κρατήσαμε τα eigenvectors με τα υψηλότερα eigenvalues, είχαμε δεδομένα μίας διάστασης, το οποίο ήταν και αναμενόμενο. Εάν συγκρίνουμε τα παρακάτω νέα δεδομένα και στις δύο περιπτώσεις, θα δούμε ότι τα δεδομένα την αυτής, είναι ακριβώς η στήλη x της πρώτης. Αυτό συμβαίνει επειδή έχουμε απορρίψει την συνεισφορά του μικρότερου eigenvector και έτσι εκφράζουμε τα δεδομένα μας μόνο με βάση του άλλου. Η γραφική παράσταση των δεδομένων θα είναι μίας διάστασης, και θα είναι σημεία πάνω στον άξονα x του σχήματος την προηγούμενης περίπτωσης.

Μετασχηματισμένα δεδομένα
(ανά eigenvector)

x
-827970186
1.77758033
-.992197494
-.274210416
-1.67580142
-.912949103
.0991094375
1.14457216
.438046137
1.22382056



Τελικά έχουμε καταφέρει να μετατρέψουμε τα δεδομένα μας έτσι ώστε να εκφράζονται βάση των μεταξύ τους προτύπων. Αυτά τα πρότυπα είναι γραμμές που περιγράφουν όσο το δυνατόν καλύτερα τη σχέση μεταξύ τους.

6. Η τελευταία φάση είναι να πάρουμε πίσω τα δεδομένα μας, ειδικά εάν χρησιμοποιούμε την μέθοδο για συμπίεση. Στην παρακάτω σχέση δείχνουμε το πώς επιτυγχάνεται αυτό:

$$\mathbf{RowOriginalData} = (\mathbf{RowFeatureVector}^T \times \mathbf{FinalData}) + \mathbf{OriginalMean}$$

T: Transformed (Μετασχηματισμένο)

3.8.2 Εφαρμογή της μεθόδου PCA στα δεδομένα μας

Για να εφαρμόσουμε την μέθοδο PCA στα δεδομένα εισόδου μας χρησιμοποιήσαμε συναρτήσεις του Matlab, όπως την princomp() και την pcacov(). Από αυτές βρήκαμε τα eigenvalues. Παρακάτω παραθέτουμε τον πίνακα των τιμών και ποσοστών eigenvalues.

Παράμετροι	Eigenvalues (%)	Eigenvalues
SO2	68.518	672.62
CO	28.096	275.81
RH	2.8455	27.933
Ta	0.34162	3.3536
WS	0.14718	1.4448
WD	0.050801	0.49869

Υπάρχουν πολλά κριτήρια για την επιλογή των eigenvalues [7], παρακάτω εμφανίζουμε ονομαστικά κάποια από αυτά.

- **Kaiser criterion:** Μία κοινή εμπειρική μέθοδος για την απόρριψη των λιγότερο σημαντικών συνιστωσών, είναι ο K1. Εντούτοις δημιουργήθηκε νωρίτερα από τον Guttman το 1954, αλλά το κριτήριο έχει συνδεθεί με την εργασία του Kaiser (1960), η οποία στηρίχθηκε πάνω του. Ο κανόνας του Kaiser είναι να απορρίπτονται όλες οι συνιστώσες όπου οι τιμές των eigenvalues είναι μικρότερες της μονάδας. Μπορεί να υπερεκτιμήσει ή να υποτιμήσει τον αληθινό αριθμό συνιστωσών. Η υπεροχή των στοιχείων μελέτης προσομοίωσης μας υποδεικνύει ότι είναι ένα συντηρητικό κριτήριο που υπερεκτιμά συνήθως τον αληθινό αριθμό συνιστωσών, μερικές φορές σε μεγάλο βαθμό (Lance, Butts, and Michels 2006).
- **Scree plot:** Στο scree test του Catel δημιουργούμε μια γραφική παράσταση όπου ο άξονας x αντιπροσωπεύει τον αριθμό των συνιστωσών και ο άξονας y τις τιμές των eigenvalues (όπως φαίνεται και στην παρακάτω γραφική παράσταση). Οι τιμές των eigenvalues έχουν πτωτική πορεία καθ' όλη την διάρκεια της γραφικής παράστασης. Όταν η πτώση παύει και η καμπύλη κάνει έναν αγκώνα προς τη λιγότερο απότομη πτώση, ο κανόνας του Catel λέει ότι θα πρέπει να απορρίψουμε όλες τις συνιστώσες μετά από αυτή που αρχίζει ο αγκώνας. Αυτός ο κανόνας μερικές φορές επικρίνεται γιατί είναι δυνατόν να επηρεαστεί από την αντίληψη του ερευνητή. Δηλαδή η επιλογή του “αγκώνα” μπορεί να είναι υποκειμενική επειδή η καμπύλη ενδέχεται να έχει πολλούς αγκώνες ή να υπάρχει μια ομαλή καμπύλη. Το κριτήριο scree plot μπορεί να οδηγήσει σε λιγότερους ή περισσότερους συντελεστές από το κριτήριο Kaiser.
- **Parallel analysis (PA):** επίσης γνωστό και ως Humphrey-Ilgen parallel analysis. Η PA συχνά προτείνεται ως η καλύτερη μέθοδος για να αξιολογήσει τον πραγματικό αριθμό συνιστωσών (Velicer, Gatton, and Fava, 2000:67; Lance, Butts, and Michels, 2006). Η PA επιλέγει τις συνιστώσες που είναι μεγαλύτεροι παρά τυχαίοι. Τα πραγματικά δεδομένα αναλύονται σε συνιστώσες (factor analysis), και επίσης ξεχωριστά γίνεται ανάλυση συνιστωσών ενός πίνακα με τυχαίους αριθμούς οι οποίοι αντιπροσωπεύουν τον ίδιο αριθμό περιπτώσεων και μεταβλητών. Και για τις πραγματικές και τυχαίες λύσεις, δημιουργούμε μια γραφική παράσταση όπου ο άξονας x αντιπροσωπεύει τον αριθμό των συνιστωσών και ο άξονας y, συσσωρευτικά eigenvalues. Όπου οι δύο γραμμές ενώνονται, καθορίζουν τον αριθμό των συνιστωσών που θα εξαχθούν.
- **Minimum average partial (MAP) criterion:** Το κριτήριο αυτό αναπτύχθηκε από τον Velicer, είναι παρόμοιο με το PA, με καλά αποτελέσματα, αλλά είναι πιο πολύπλοκο, O'Connor (2000).

- **Variance explained criteria:** Μερικοί ερευνητές χρησιμοποιούν τον κανόνα όπου απλά κρατούν τόσες συνιστώσες έτσι ώστε να αντιπροσωπεύουν το 90% (μερικές φορές 80%) της συνολικής συνεισφοράς. Ο στόχος του ερευνητή επικεντρώνεται στην φειδωλότητα (όπου ερμηνεύει την συνεισφορά με όσο το δυνατόν λιγότερους συντελεστές), έτσι το κριτήριο μπορεί να φτάσει και μέχρι το 50%.
- **Joliffe criterion:** Ένα λιγότερο χρησιμοποιημένο κριτήριο, το οποίο είναι ένα πιο ανεκτικό εμπειρικό κριτήριο που μπορεί να οδηγήσει και σε δύο φορές περισσότερους συντελεστές από το κριτήριο Kaiser.
- **Mean eigenvalue:** Σε αυτόν τον κανόνα χρησιμοποιούμε μόνο τις συνιστώσες στις οποίες οι τιμές των eigenvalues είναι μεγαλύτερες από τον μέσο όρο των eigenvalues. Αυτός ο απόλυτος κανόνας μπορεί να οδηγήσει σε πολύ λίγους παράγοντες.

Γενικά πριν απορριφθεί ένας συντελεστής θα πρέπει ο ερευνητής να ελέγξει την συσχέτιση με την εξαρτημένη μεταβλητή. Ένας πολύ μικρός συντελεστής μπορεί να έχει έναν μεγάλο συσχετισμό με την εξαρτημένη μεταβλητή, οπότε και δεν θα πρέπει να απορριφθεί. Επίσης εμπειρικά, θα πρέπει να υπάρχουν τουλάχιστον τρεις υψηλοί συντελεστές, λιγότεροι μπορεί να υποδηλώνουν ότι ο ερευνητής έχει χρησιμοποιήσει πάρα πολλούς συντελεστές.

Εμείς για την επιλογή των eigenvalues χρησιμοποιήσαμε το κριτήριο του Kaiser, κατά το οποίο απορρίπτουμε τα eigenvalues των οποίων οι τιμές είναι μικρότερες της μονάδας. Έτσι απορρίπτουμε την τελευταία τιμή (6^η) από τον πίνακα των eigenvalues. Αυτό μπορούμε να το δούμε και στον πίνακα συνδιακύμανσης (σε ποσοστά) όπου έχουμε ένα χαρακτηριστικό (WDn) το οποίο δεν έχει καμία σχέση με τις τιμές βενζολίου για τις οποίες θέλουμε να κάνουμε πρόβλεψη.

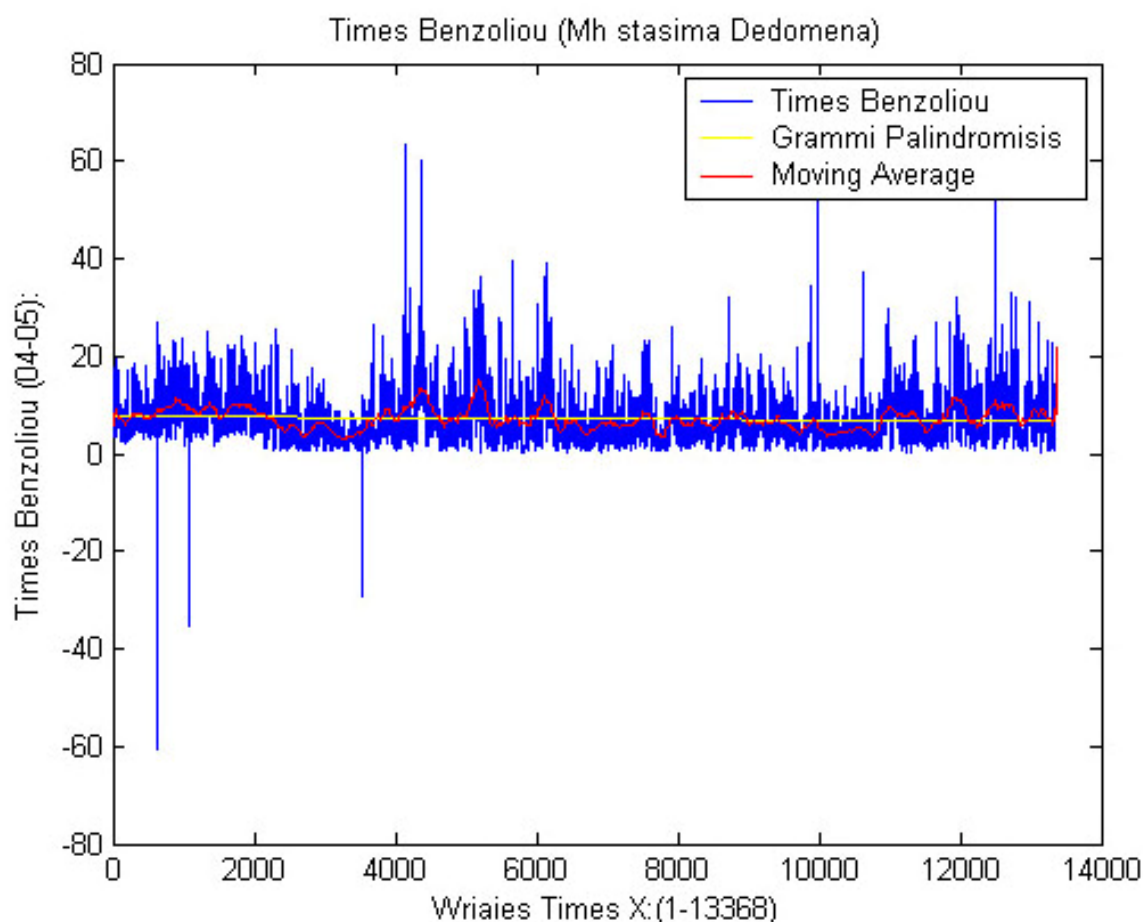
Τελικά επιλέξαμε πέντε principal components των οποίων η συμβολή φτάνει στο 99,95%.

Στην συνέχεια θα χρησιμοποιήσουμε τα διαμορφωμένα δεδομένα, τα οποία έχουν μειωμένες διαστάσεις ως είσοδο στον νευρωνικό δίκτυο για να κάνουμε πρόβλεψη των τιμών βενζολίου (ο αρχικός μας στόχος).

3.9 Υπολογισμός “περιοδογράμματος” της χρονοσειράς βενζολίου

3.9.1 Μελέτη στασιμότητας

Για τον υπολογισμό του περιοδογράμματος θα πρέπει η χρονοσειρά βενζολίου να είναι στάσιμη. Για να το δούμε αυτό, δημιουργήσαμε την παρακάτω γραφική παράσταση, στην οποία φαίνονται οι ωριαίες εναλλαγές των τιμών βενζολίου, η γραμμή παλινδρόμησης αλλά και ο κινούμενος μέσος όρος (Moving Average) από τον οποίο βλέπουμε ότι τα δεδομένα μας δεν είναι στάσιμα.



Σχήμα 47. Μη στάσιμη χρονοσειρά βενζολίου (γραμμή παλινδρόμησης, κινούμενος μέσος όρος).

3.9.2 Μελέτη γραμμής παλινδρόμησης

Η γραμμή παλινδρόμησης του παραπάνω σχήματος είναι η εξής:

$$y = a + bx = 7,6809 - 7,4315 \cdot 10^{-5} \cdot x \approx 7,6809$$

Για να βρούμε την γραμμή παλινδρόμησης ($y=a+bx$), αρχικά υπολογίσαμε τους συντελεστές παλινδρόμησης a και b (regression coefficients), από τους τύπους που αναγράφονται παρακάτω.

$$b = \frac{N \left(\sum_{i=1}^N x_i y_i \right) - \left(\sum_{i=1}^N x_i \right) \left(\sum_{i=1}^N y_i \right)}{N \left(\sum_{i=1}^N x_i^2 \right) - \left(\sum_{i=1}^N x_i \right)^2}$$
$$a = \frac{\sum_{i=1}^N y_i}{N} - b \frac{\sum_{i=1}^N x_i}{N}$$

3.9.3 Μελέτη κινούμενου μέσου όρου

Πριν συνεχίσουμε θα πρέπει να αναφερθούμε στον κινούμενο μέσο όρο (Moving Average) και στον τρόπο υπολογισμού του. Στην στατιστική ο κινούμενος μέσος όρος είναι μια τεχνική που χρησιμοποιείται στην ανάλυση χρονοσειρών [16]. Εφαρμόζεται στα οικονομικά και ειδικότερα στην τεχνική ανάλυση. Μπορεί επίσης να χρησιμοποιηθεί ως μια γενική μέθοδος εξομάλυνσης (Smoothing), οπότε σ' αυτή την περίπτωση τα ακατέργαστα δεδομένα δεν χρειάζονται να είναι χρονοσειρές.

Ο κινούμενος μέσος όρος μπορεί να υπολογιστεί για κάθε χρονοσειρά. Στα οικονομικά συνήθως εφαρμόζεται σε κάποιο πλήθος πληροφοριών όπως στις τιμές αποθεμάτων, στις επιστροφές ή στις εμπορικές συναλλαγές. Οι κινούμενοι μέσοι όροι χρησιμοποιούνται στην εξομάλυνση βραχυπρόθεσμων διακυμάνσεων, δίνοντας έμφαση κατά συνέπεια στις πιο μακροπρόθεσμες ροπές ή κύκλους. Το κατώτερο όριο (threshold) μεταξύ του βραχυπρόθεσμου και του μακροπρόθεσμου εξαρτάται από την εφαρμογή, και οι παράμετροι του κινούμενου μέσου όρου θα τεθούν αναλόγως.

Από μαθηματικής άποψης, κάθε ένας από αυτούς τους κινούμενους μέσους όρους είναι ένα παράδειγμα μίας συνέλιξης (convolution). Επίσης αυτοί οι μέσοι όροι είναι παρόμοιοι με τα χαμηλοπερατά φίλτρα (low-pass filters) που χρησιμοποιούνται στην επεξεργασία σήματος.

Στα πλαίσια αυτής της εργασίας θα μελετήσουμε τον απλό κινούμενο μέσο όρο (Simple Moving Average: SMA) τον οποίο έχουμε χρησιμοποιήσει παραπάνω. Αυτός ο μέσος όρος δεν χρησιμοποιεί την έννοια του “βάρους” στα δεδομένα, επίσης υπολογίζεται από τα n προηγούμενα σημεία. Για παράδειγμα, ένας απλός κινούμενος

μέσος όρος δέκα ημερών τιμών κλεισίματος (χρηματιστηρίου), είναι ο μέσος όρος των τιμών κλεισίματος των προηγούμενων δέκα ημερών. Εάν αυτές οι τιμές είναι $PM, PM_{-1} + \dots + PM_{-9}$ τότε ο τύπος είναι ο εξής:

$$SMA = \frac{PM + PM_{-1} + \dots + PM_{-9}}{10}$$

Καθώς υπολογίζουμε διαδοχικές τιμές, μία νέα τιμή εισάγεται στο άθροισμα, και μια παλιά τιμή απορρίπτεται, έτσι είναι περιττό να υπολογίζουμε κάθε φορά το πλήρες άθροισμα:

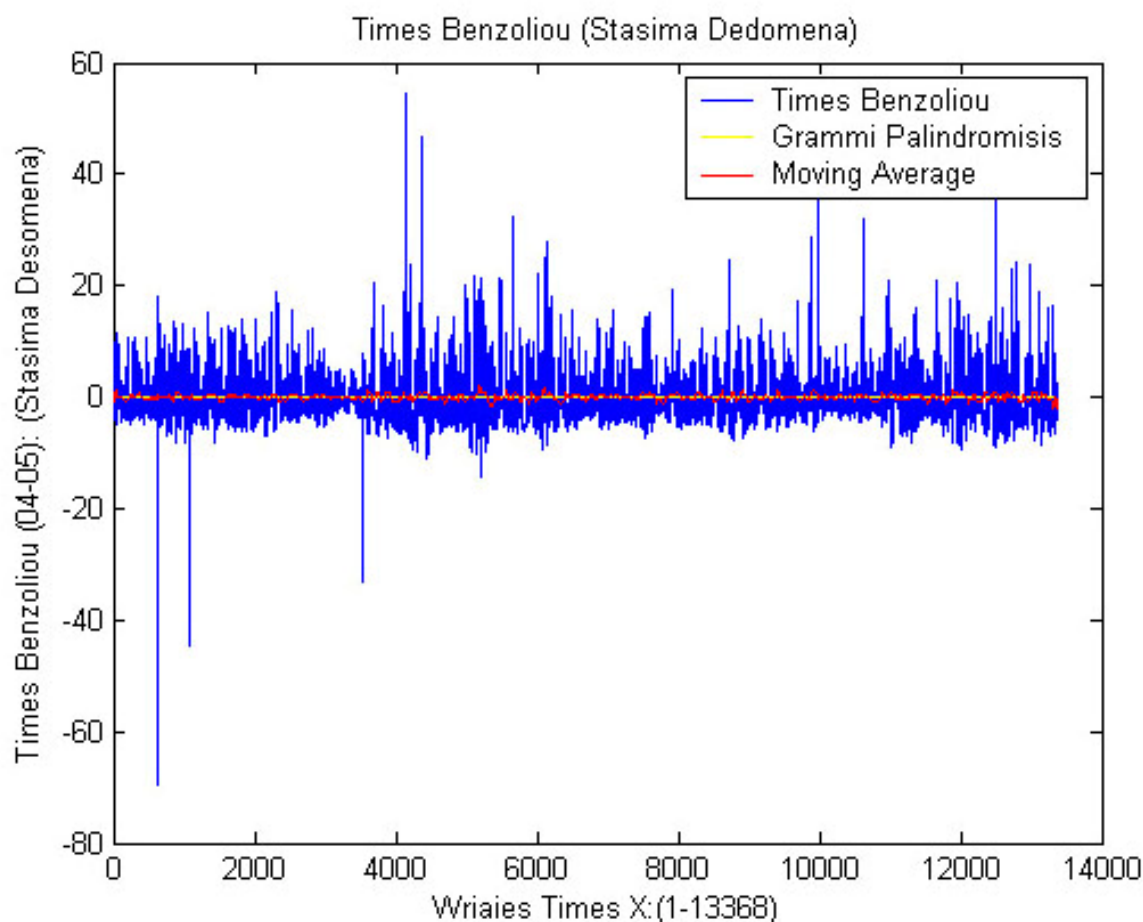
$$SMA_{Today} = SMA_{Yesterday} - \frac{PM_{-n+1}}{n} + \frac{PM_{+1}}{n}$$

Στην τεχνική ανάλυση υπάρχουν διάφορες δημοφιλείς τιμές για το n , όπως 10 ημέρες, 40 ημέρες, ή 200 ημέρες. Η περίοδος που επιλέγεται, εξαρτάται από το είδος της μεταβολής που κάποιος επικεντρώνεται, όπως σύντομη, ενδιάμεση, ή μακροπρόθεσμη. Τα επίπεδα του μέσου όρου σε κάθε περίπτωση ερμηνεύονται ως υποστήριξη σε μία αυξανόμενη αγορά ή αντίσταση σε μία φθίνουσα.

Σε όλες τις περιπτώσεις ένας κινούμενος μέσος όρος καθυστερεί πίσω από την πιο πρόσφατη δράση τιμών, από την ίδια την φύση της εξομάλυνσης. Ο κινούμενος μέσος όρος μπορεί να καθυστερήσει σε μία ανεπιθύμητη έκταση, και μπορεί να επηρεαστεί πάρα πολύ από τις παλιές τιμές που χάνονται από το μέσο όρο. Σε αυτές τις περιπτώσεις μπορούμε να στραφούμε σε τεχνικές που δίνουν πρόσθετη βαρύτητα στις πρόσφατες τιμές, όπως στην τεχνική WMA (Weighted Moving Average) ή στην EMA (Exponential Moving Average).

3.9.4 Άρση στασιμότητας χρονοσειράς βενζολίου

Στην συνέχεια αφαιρέσαμε από την χρονοσειρά βενζολίου τον κινούμενο μέσο όρο, έτσι ώστε να καταστεί στάσιμη και με μηδενική μέση τιμή. Στην ουσία η γραμμή παλινδρόμησης είναι ο μέσος όρος για κάθε σημείο. Παρακάτω παραθέτουμε την γραφική παράσταση με τα στάσιμα δεδομένα την γραμμή παλινδρόμησης η οποία βρίσκεται στο μηδέν αλλά και τον κινούμενο μέσο όρο ο οποίος κινείται γύρω από το μηδέν.



Σχήμα 48. Στάσιμη χρονοσειρά βενζολίου (γραμμή παλινδρόμησης, κινούμενος μέσος όρος).

Η γραμμή παλινδρόμησης του παραπάνω σχήματος είναι η εξής:

$$y = a + bx = 0,0109 - 2,3791 \cdot 10^{-6} \cdot x \approx 0$$

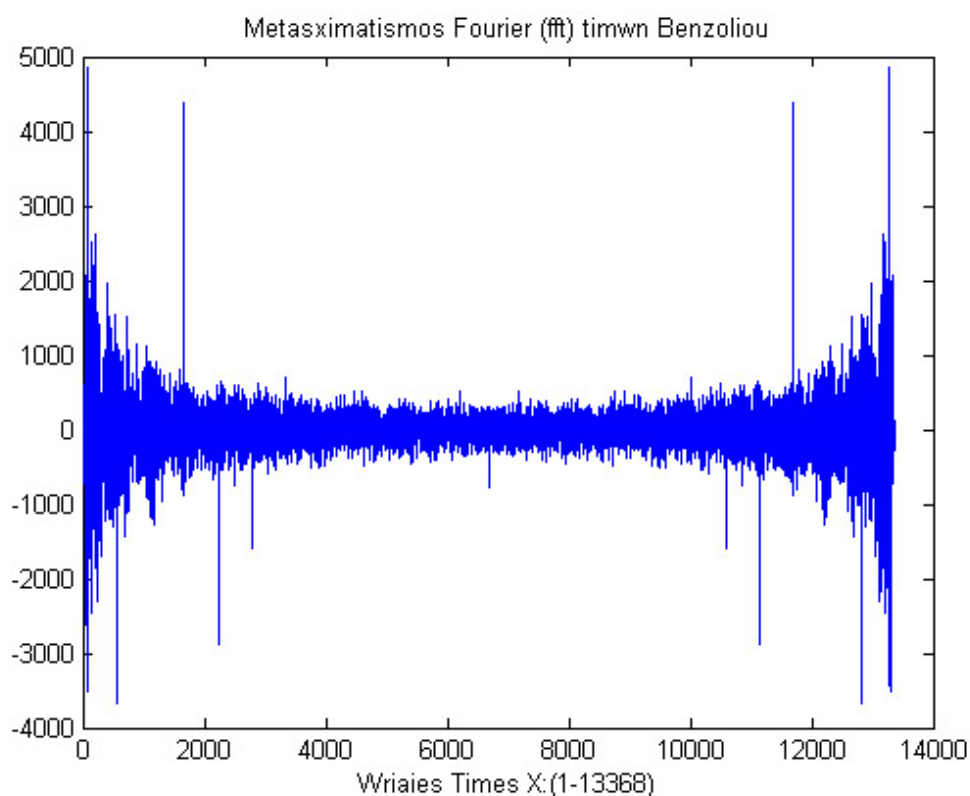
3.9.5 Χρήση FFT για υπολογισμό και μελέτη του περιοδογράμματος

Μια περιοδική κυματομορφή μπορεί να εκφραστεί, όπως έχουμε δει, ως ένα άθροισμα φασματικών συνιστωσών. Αυτές οι συνιστώσες έχουν πεπερασμένα πλάτη και χωρίζονται από πεπερασμένα διαστήματα συχνότητας $f_0=1/T_0$ [12]. Για τον υπολογισμό του αθροίσματος των φασματικών συνιστωσών χρησιμοποιούμε τον Μετασχηματισμό Fourier και ο αλγόριθμος για την πραγματοποίησή του είναι ο FFT (Fast Fourier Transform)

Για μία χρονοσειρά X , N στοιχείων, ο FFT παράγει ένα N -διάστατο διάνυσμα, εφαρμόζοντας τη σχέση [13]:

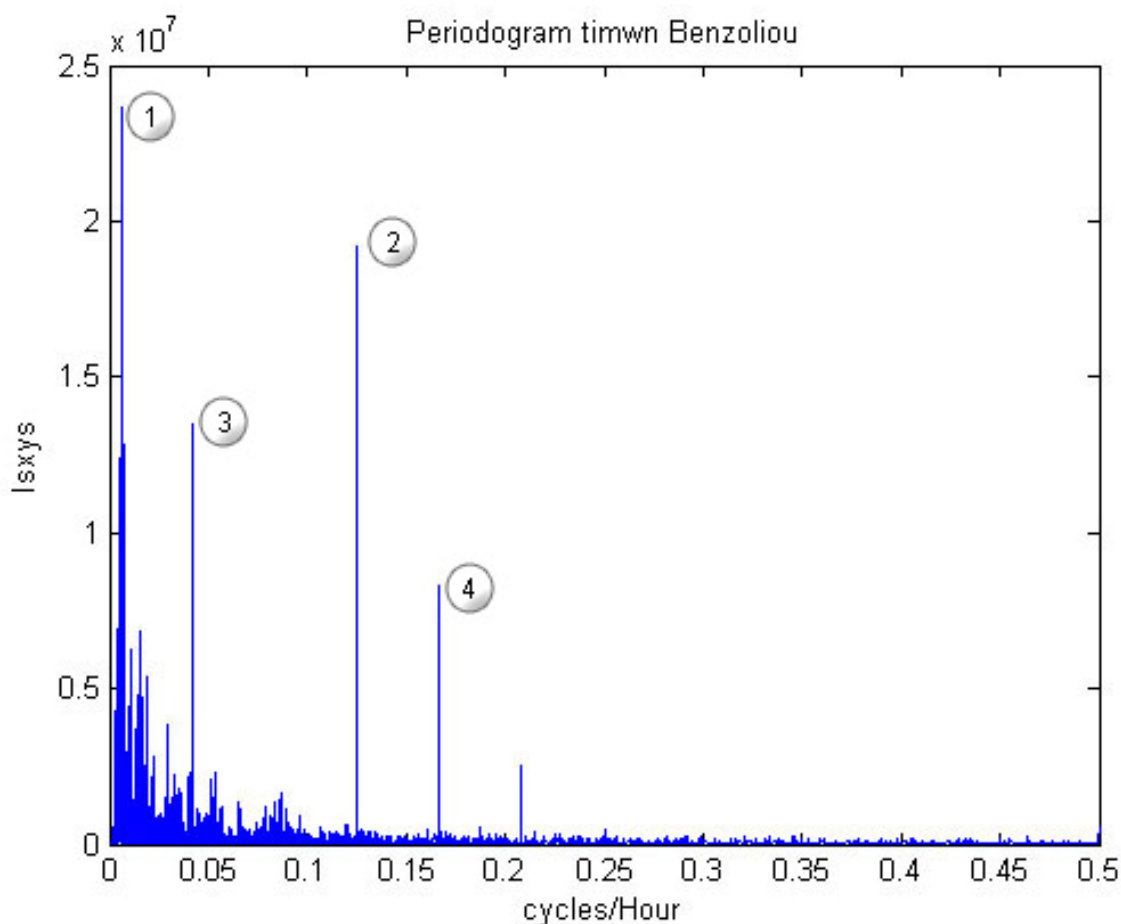
$$x(n) = \frac{1}{N} \sum_{k=1}^N \alpha(k) \cos\left(\frac{2\pi(k-1)(n-1)}{N}\right) + b(k) \sin\left(\frac{2\pi(k-1)(n-1)}{N}\right)$$

Το μέγεθος (μέτρο) του τετραγώνου του Y ονομάζεται “ισχύς” και ένα διάγραμμα της “ισχύος” προς την συχνότητα ονομάζεται περιοδογράμμο. Για την δημιουργία του περιοδογράμματος χρειαζόμαστε το μισό πλήθος τιμών της ισχύος. Γιατί, τα αποτελέσματα του Μ.Φ. είναι άρτια συμμετρικά ως προς την μέση του άξονα x . Όπως φαίνεται και στην παρακάτω γραφική παράσταση.



Σχήμα 49. Γραφική παράσταση μετασχηματισμού Fourier, χρονοσειράς βενζολίου (σταθμός Πατησίων)

Η μεγαλύτερες τιμές της ισχύος θα μας δείξουν την περιοδικότητα της χρονοσειράς μας. Στην παρακάτω γραφική παράσταση εμφανίζεται το περιοδογράμμο της χρονοσειράς βενζολίου.



Σχήμα 50. Γραφική παράσταση περιοδογράμματος, χρονοσειράς βενζολίου (σταθμός Πατησίων).

Στον παρακάτω πίνακα παρουσιάζουμε την περιοδικότητα των τεσσάρων πιο μεγάλων τιμών τις ισχύος (πιο ισχυρές), όπως τις έχουμε αριθμήσει στην γραφική παράσταση.

A/A	Κύκλοι / Ωρα	Περιοδικότητα
1	0,0059844	7 ημερών
2	0,125	8 ωρών
3	0,041667	24 ωρών
4	0,16667	6 ωρών

Η περιοδικότητα του παραπάνω πίνακα ενισχύει τον αρχικό, συλλογισμό μας ότι η κύρια πηγή βενζολίου είναι η κυκλοφορία. Αυτό το συμπεραίνουμε γιατί υπάρχει περιοδικότητα της κυκλοφορίας ανά εβδομάδα (περισσότερη κίνηση τις καθημερινές, και λιγότερη το σαββατοκύριακο). Επίσης η περιοδικότητα το οκταώρου είναι πολύ σημαντική καθώς υποδηλώνει την αρχή και το τέλος των εργασιών, με αποτέλεσμα επιπρόσθετη κυκλοφορία για αυτές τις περιόδους. Με το ίδιο σκεπτικό, και στη περιοδικότητα ανά ημέρα (24 ωρών), υπάρχει περισσότερη κίνηση κατά την διάρκεια της ημέρας και λιγότερη τις απογευματινές με βραδινές ώρες.

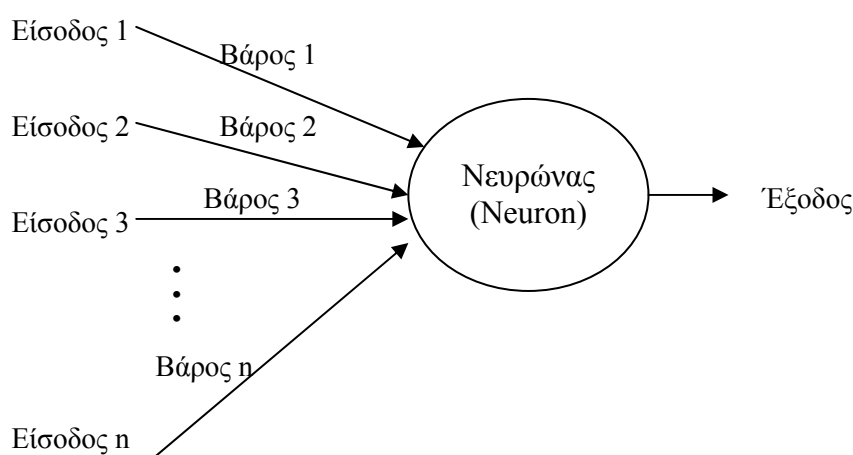
4. Πρόγνωση ωριαίων τιμών βενζολίου

4.1 Εισαγωγή

Για την πρόγνωση των τιμών βενζολίου χρησιμοποιήσαμε τεχνητά νευρωνικά δίκτυα. Τα οποία αποτελούν μια σχετικά νέα περιοχή στις φυσικές επιστήμες [8], καθ' όσον έχουν γίνει γνωστά και έχουν αναπτυχθεί κατά τα τελευταία σαράντα χρόνια. Η λειτουργία τους προσπαθεί να συνδυάσει τον τρόπο σκέψης του ανθρώπινου εγκεφάλου με τον αφηρημένο μαθηματικό τρόπο σκέψης. Έτσι στα νευρωνικά δίκτυα χρησιμοποιούμε ιδέες όπως, π.χ. ένα δίκτυο μαθαίνει και εκπαιδεύεται, θυμάται ή ξεχνά μια αριθμητική τιμή, κτλ. Πράγμα που μέχρι τώρα τα αποδίδουμε μόνο στην ανθρώπινη σκέψη.

Ο εγκέφαλος αποτελείται από νευρωνικά δίκτυα (βιολογικά νευρωνικά δίκτυα). Κάθε νευρωνικό δίκτυο αποτελείται από ένα μεγάλο αριθμό μονάδων, που ονομάζονται νευρώνες, ή νευρώνια (neurons). Ο νευρώνας είναι η πιο μικρή ανεξάρτητη μονάδα του δικτύου. Οι νευρώνες συνεχώς και ασταμάτητα επεξεργάζονται πληροφορίες, παίρνοντας και στέλνοντας ηλεκτρικά σήματα σε άλλους νευρώνες. Έτσι προσπαθώντας να προσομοιώσουμε την λειτουργία του εγκεφάλου τα τεχνικά νευρωνικά δίκτυα έχουν την ίδια δομή με τα βιολογικά νευρωνικά δίκτυα.

Κάθε νευρώνας μπορεί να έχει ένα n πλήθος εισόδων αλλά μία μόνο έξοδο. Κάθε είσοδος έχει μια δική της τιμή βάρους, και η οποία υποδηλώνει πόσο στενά είναι συνδεδεμένοι οι δύο νευρώνες που συνδέονται με το βάρος αυτό. Η τιμή των βαρών κυμαίνεται στο διάστημα -1 ως 1 . Παρακάτω δείχνουμε την δομή ενός νευρώνα.



Η τιμή εξόδου του παραπάνω νευρώνα υπολογίζεται ως εξής:

1. Αρχικά υπολογίζουμε το άθροισμα d

$$d = \sum_{i=1}^n x_i * w_i$$

Όπου:

x_i : Η είσοδος στην θέση i
 w_i : Το βάρος της θέσης i

2. Στην συνέχεια χρησιμοποιούμε κάποια συνάρτηση μεταφοράς, όπως την συνάρτηση Sigmoid για να υπολογίζουμε την έξοδο. Παρακάτω θα δώσουμε περισσότερες πληροφορίες για τις συναρτήσεις μεταφοράς.

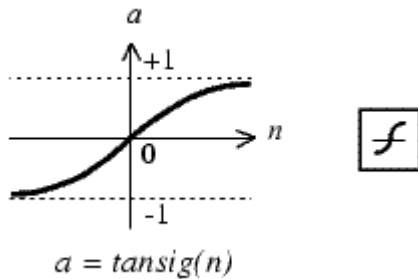
Το θ είναι μια τιμή κατωφλιού (threshold). Όπου όταν το $S > \theta$ ο νευρώνας ενεργοποιείται, ενώ όταν το $S < \theta$ τότε ο νευρώνας παραμένει αδρανής [8]. Η εμπειρία έχει δείξει ότι καλύτερες τιμές για το θ είναι μεταξύ του 0,25-1 [9].

$$Eξοδος = S(d + \theta)$$

4.2 Συναρτήσεις μεταφοράς

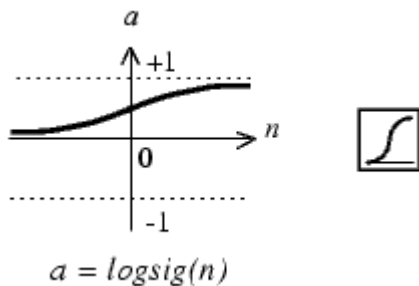
Τα νευρωνικά δίκτυα δουλεύουν με τιμές από -1 ως 1, έτσι θα πρέπει να μετατρέψουμε τα δεδομένα μας έτσι ώστε να βρίσκονται σε αυτό το εύρος τιμών. Στα νευρωνικά δίκτυα χρησιμοποιούμε συνήθως τις τρεις παρακάτω συναρτήσεις μεταφοράς, για τις οποίες παραθέτουμε το εκάστοτε γράφημα, σύμβολο αλλά και τον αλγόριθμο με τον οποίο υπολογίζεται [10].

- Hyperbolic Tangent Sigmoid Transfer Function:



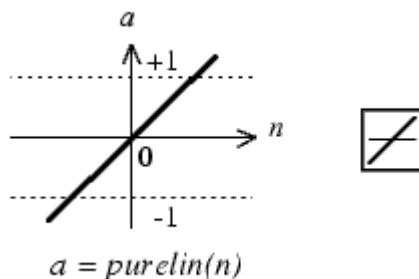
$$\text{tan sig}(n) = \frac{2}{(1 + e^{-2 * n})} - 1$$

- Log Sigmoid Transfer Function"



$$\text{log sig}(n) = \frac{1}{1 + e^{-n}}$$

- Linear Function:

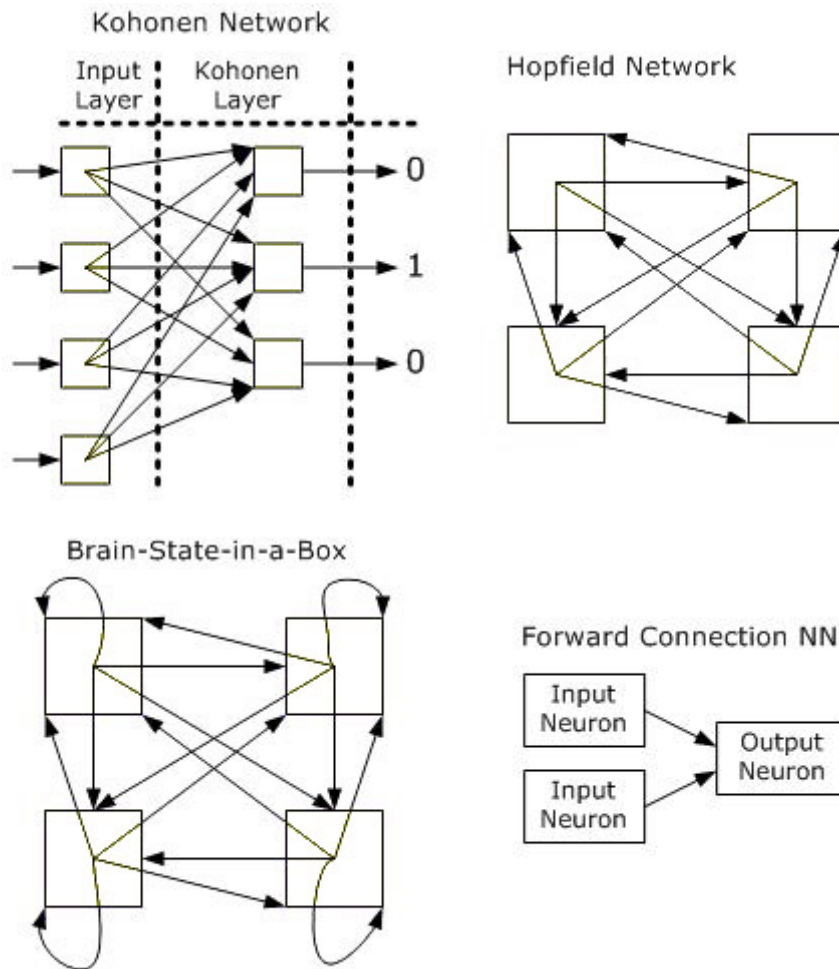


$$\text{purepin}(n) = n$$

4.3 Είδη τεχνητών νευρωνικών δικτύων

Υπάρχουν πολλά είδη τεχνητών νευρικών δικτύων. Παρακάτω υπάρχει μία λίστα μερικών σημαντικών νευρωνικών δικτύων αλλά και τρόπος με τον οποίο γίνονται οι συνδέσεις σε αυτά [9]:

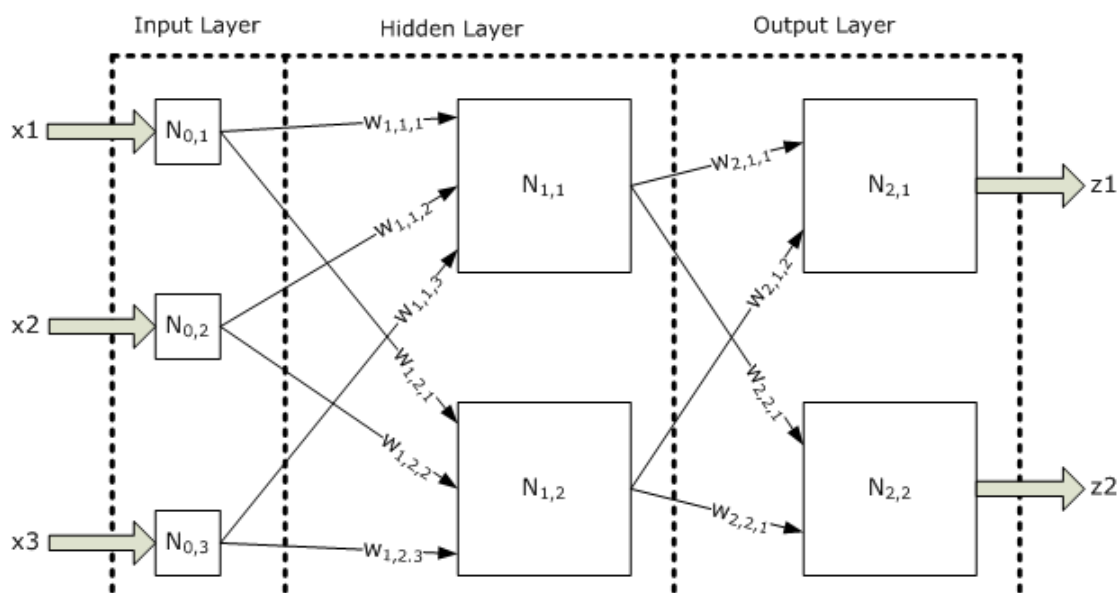
- Kohonen Networks
- Hopfield Networks
- Brain-State-in-a-Box,
- Forward Connection Neural Network



Σχήμα 51. Σχηματική αναπαράσταση κατηγοριών τεχνητών νευρωνικών δικτύων.

4.4 Μέθοδος εκμάθησης Back Propagation

Εμείς χρησιμοποιήσαμε νευρωνικό δίκτυο Forward Connection και για την εκπαίδευσή του, χρησιμοποιούμε την μέθοδο Back Propagation. Παρακάτω εμφανίζουμε ένα διάγραμμα στο οποίο φαίνεται ο τρόπος με τον οποίο γίνονται οι συνδέσεις σε αυτό.



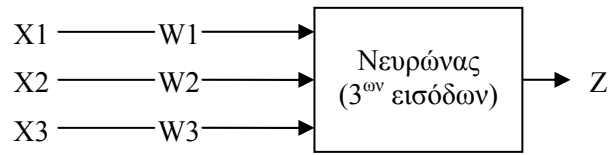
Σχήμα 52. Διάγραμμα τρόπου σύνδεσης των νευρώνων στην μέθοδο εκμάθησης Back Propagation.

Συνοπτικά η διαδικασία εκμάθησης στα νευρωνικά δίκτυα με την χρήση της μεθόδου Back Propagation είναι η εξής [9]:

- Αρχικά θέτουμε τιμές στα βάρη των εισόδων για όλους τους νευρώνες με κάποια τυχαία τιμή ανάλογα με την συνάρτηση μεταφοράς που έχει επιλεγεί.
- Εισάγουμε τις τιμές εισόδου.
- Υπολογίζουμε τις τιμές εξόδου.
- Συγκρίνουμε το αποτέλεσμα που υπολογίσαμε, με το επιθυμητό αποτέλεσμα (υπολογίζουμε το “σφάλμα”).
- Αλλάζουμε τις τιμές των βαρών και τιμής καταφλιού θ για όλους τους νευρώνες χρησιμοποιώντας το σφάλμα (που υπολογίστηκε στο προηγούμενο βήμα).
- Επαναλαμβάνουμε αυτή τη διαδικασία μέχρι το σφάλμα να φτάσει σε κάποια επιθυμητή τιμή (π.χ. σφάλμα $< 1\%$). Το οποίο και θα σημαίνει ότι το νευρωνικό μας δίκτυο έχει εκπαιδευτεί σωστά ενώ εάν φτάσουμε στο όριο των επαναλήψεων που έχουμε ορίσει το νευρωνικό μας δίκτυο δεν έχει εκπαιδευτεί σωστά.

Η εκμάθηση (δηλ. οι αλλαγές των βαρών και των τιμών θ) γίνεται διαφορετικά στα Output Layer (εξωτερικό επίπεδο) απ' ό τι στα Hidden Layers (κρυφά επίπεδα).

- Η εκμάθηση στο εξωτερικό επίπεδο γίνεται ως εξής:



Για κάθε νευρώνα στο επίπεδο εξόδου, υπολογίζουμε το σφάλμα e , και στη συνέχεια αλλάζουμε την τιμή θ αλλά το βάρος του.

$$e = Z * (1 - Z) * (y - Z)$$

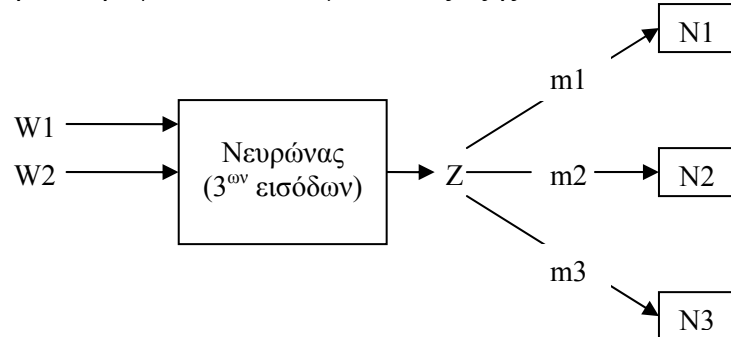
$$\Delta\theta = \lambda * e \quad \dots\eta \text{ αλλαγή στο } \theta$$

$$\Delta w_i = \Delta\theta * x_i \quad \dots\eta \text{ αλλαγή στο βάρος της μονάδας εισόδου } i$$

Όπου:

- Z = η έξοδος του επιπέδου εξόδου
- y = το επιθυμητό αποτέλεσμα
- λ = ο ρυθμός εκμάθησης (συνήθων 0,2)

- Η εκμάθηση των κρυφών επιπέδων γίνεται ως εξής:



$$g = \sum_{i=1}^n m_i * e_i$$

$$e = Z * (1 - Z) * g \quad \dots\text{το σφάλμα του νευρώνα κρυφού επιπέδου}$$

$$\Delta\theta = \lambda * e \quad \dots\eta \text{ αλλαγή στο } \theta$$

$$\Delta w_i = \Delta\theta * x_i \quad \dots\eta \text{ αλλαγή στο βάρος της μονάδας εισόδου } i$$

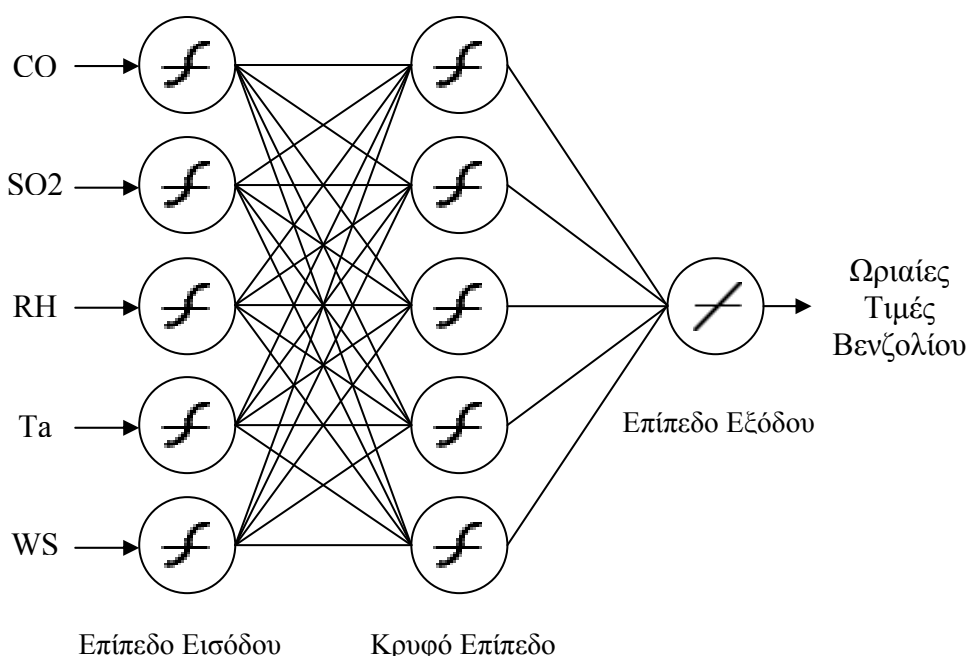
Όπου:

- Z = η έξοδος του κρυφού επιπέδου
- m_i = τα βάρη των νευρώνων N_i
- e_i = το σφάλμα (e) για τον νευρώνα N_i
- n = το πλήθος των νευρώνων του επόμενου επιπέδου (στο παραπάνω παράδειγμα είναι 3)

4.5 Δημιουργία & Μελέτη αποτελεσμάτων νευρωνικών δικτύων

Για την πρόβλεψη τιμών βενζολίου δημιουργήσαμε πολλές δομές νευρωνικών δικτύων, στις οποίες κάναμε πολλές φορές εκμάθηση έτσι ώστε να μην “ξεγελαστούμε” από την περίπτωση που η διαδικασία εκμάθησης βρει κάποιο “local minimum”. Οι αλλαγές στις δομές έγιναν αυξάνοντας τα επίπεδα αλλά και τους νευρώνες σε κάθε επίπεδο. Το αρχικό μας νευρωνικό δίκτυο ήταν, δύο επιπέδων. Ένα κρυφό επίπεδο με πέντε νευρώνες (επειδή έχουμε πέντε εισόδους) και το επίπεδο εξόδου με ένα νευρώνα (επειδή κάνουμε πρόβλεψη ενός χαρακτηριστικού).

Παρακάτω εμφανίζουμε την γραφική αναπαράσταση, του αρχικού μας νευρωνικού δικτύου:



Σχήμα 53. Γραφική αναπαράσταση του αρχικού τεχνητού νευρωνικού δικτύου.

Αρχικά κάναμε ομαλοποίηση (normalize) των δεδομένων μας, πριν τα χρησιμοποιήσουμε στο νευρωνικό δίκτυο. Η ομαλοποίηση έγινε με την συνάρτηση μεταφοράς “Hyperbolic Tangent Sigmoid”. Στο κρυφό επίπεδο χρησιμοποιήσαμε την ίδια συνάρτηση μεταφοράς αλλά στο επίπεδο εξόδου χρησιμοποιήσαμε την γραμμική συνάρτηση μεταφοράς. Αυτή είναι μια χρήσιμη δομή για προβλήματα προσέγγισης (ή παλινδρόμησης) της συνάρτησης [10] αλλά επίσης έχει δώσει τα πιο ικανοποιητικά αποτελέσματα σε παλιότερες μελέτες [14]. Για την φάση της εκμάθησης (training) χρησιμοποιήσαμε τον αλγόριθμο Levenberg-Marquardt, όπου χρησιμοποιήσαμε το 50% των δεδομένων. Για την επικύρωση (validation) χρησιμοποιήσαμε το 25% και το υπόλοιπο 25% για την φάση του ελέγχου (testing).

Χρησιμοποιήσαμε δύο τρόπους για να χωρίζουμε τα δεδομένα. Στον έναν τα δεδομένα εισάχθηκαν σειριακά, δηλαδή το πρώτο 50% χρησιμοποιήθηκε για εκμάθηση, το επόμενο 25% για επικύρωση και το υπόλοιπο για έλεγχο. Παρακάτω εμφανίζουμε σχηματικά αυτόν τον τρόπο.

Δεδομένα εκμάθησης (50%)	Δεδ. Επικύρωσης (25%)	Δεδ. Ελέγχου (25%)
--------------------------	-----------------------	--------------------

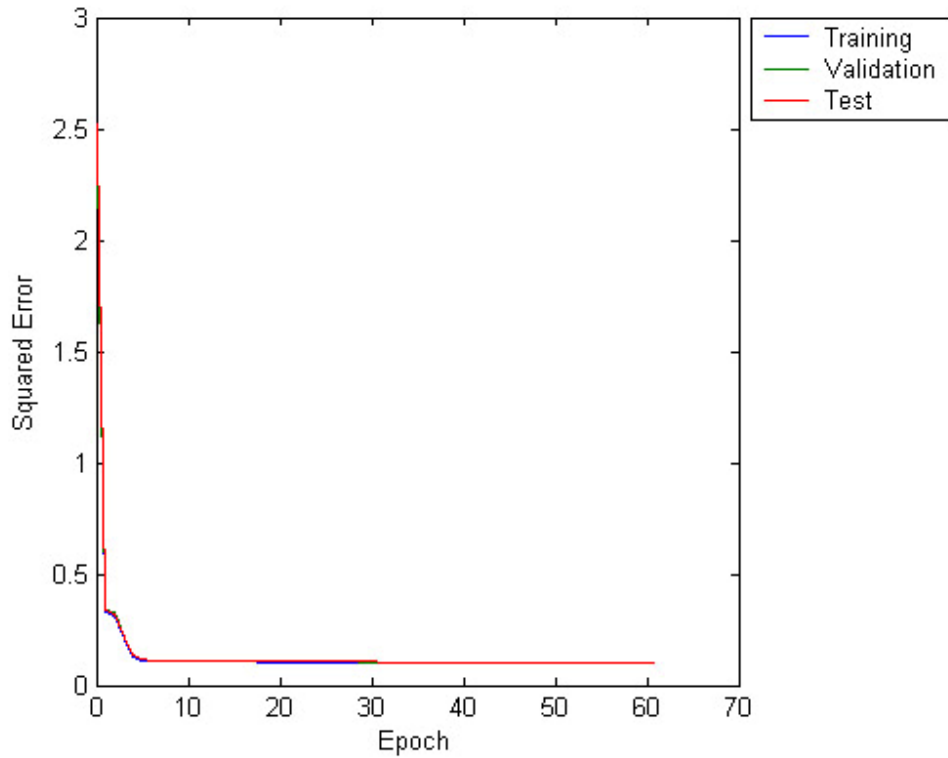
Στον δεύτερο τρόπο, τα δεδομένα δεν εισάχθηκαν σειριακά στο νευρωνικό δίκτυο αλλά εναλλάσσονταν διαδοχικά. Συγκεκριμένα ανά τέσσερις τιμές οι δύο χρησιμοποιήθηκαν για εκμάθηση, και από μία για την επικύρωση και έλεγχο. Αυτό έγινε έτσι ώστε το νευρωνικό μας δίκτυο να μην εκπαιδευτεί μόνο για μία συγκεκριμένη χρονική περίοδο (του συνόλου των δεδομένων που διαθέτουμε), αλλά γενικά από όλη την χρονική περίοδο. Παρακάτω εμφανίζουμε σχηματικά αυτόν τον τρόπο.

Tr.	Test	Tr.	Val.	Tr.	Test	Tr.	Val.	...	Tr.	Test	Tr.	Val.
-----	------	-----	------	-----	------	-----	------	-----	-----	------	-----	------

Tr.: Εκμάθηση
 Test: Έλεγχος
 Val.: Επικύρωση

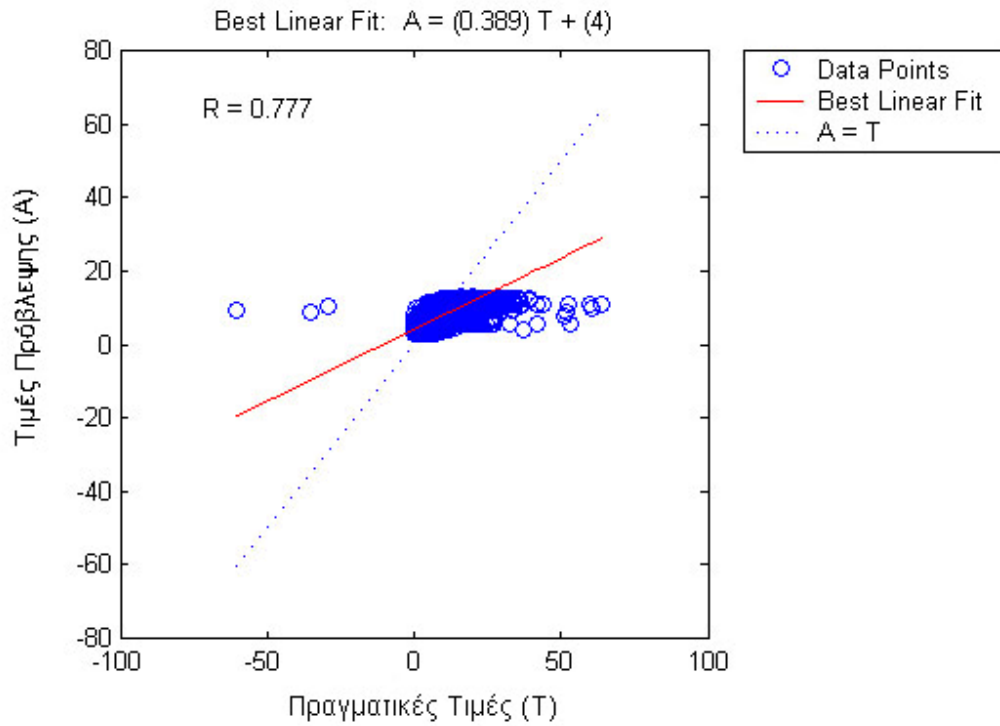
Παρακάτω δείχνουμε την διαδικασία που πραγματοποιήθηκε για κάθε ξεχωριστή δοκιμή, αλλά επίσης και συγκεντρωτικά στοιχεία από όλες τις δοκιμές που πραγματοποιήθηκαν.

Σε αυτή την δοκιμή χρησιμοποιήσαμε τον δεύτερο τρόπο επιλογής δεδομένων (εναλλάξ), όπου η διαδικασία εκμάθησης σταμάτησε σε 61 επαναλήψεις γιατί υπήρξε αύξηση του σφάλματος (validation error). Παρακάτω παραθέτουμε την γραφική παράσταση του σφάλματος.



Σχήμα 54. Γραφική αναπαράσταση του σφάλματος εκμάθησης.

Στην συνέχεια πραγματοποιήσαμε μια ανάλυση για να δούμε το πώς ανταποκρίνεται το νευρωνικό μας δίκτυο. Έτσι δημιουργήσαμε την παρακάτω γραφική παράσταση για να συγκρίνουμε την έξοδο του νευρωνικού δικτύου (A) μας με τις πραγματικές τιμές (T). Η τιμή προσέγγισης R (Regression Value) βλέπουμε ότι είναι 0,777.



Best Linear Fit είναι η γραμμή παλινδρόμησης των πραγματικών και των προβλεπόμενων τιμών. Επίσης για τον έλεγχο των αποτελεσμάτων πρόβλεψης θα υπολογίσαμε τον δείκτη προσδιορισμού (R-Square) και το τυπικό σφάλμα εκτίμησης (Standard error of estimation), όπως και στο Κεφ. 3.3. Ο δείκτης προσδιορισμού ισούται με 0,61 και το τυπικό σφάλμα εκτίμησης ισούται με 3,339.

Όμως δοκιμάσαμε και άλλες δομές νευρικών δικτύων προσπαθώντας να αυξήσουμε την απόδοση πρόβλεψης. Οι δοκιμές μας έγιναν αυξάνοντας το πλήθος των νευρώνων αλλά και των κρυφών επιπέδων. Παρακάτω εμφανίζουμε τον πίνακα των αποτελεσμάτων για κάθε δοκιμή και με τους δύο τρόπους επιλογής δεδομένων.

Σταθμός Πατησίων	Κρυφά Επίπεδα (\mathcal{F})			Δεδομένα Εναλλάξ			Σειριακά Δεδομένα		
				Επιπ. Εξοδ. (\mathcal{Z})			Επιπ. Εξοδ. (\mathcal{Z})		
	Κ.Ε. 1	Κ.Ε. 2	Κ.Ε. 3	n	R	R ²	n	R	R ²
Δομή 1	5	-	-	61	0,777	0,60	25	0,764	0,58
Δομή 2	5	5	-	59	0,789	0,62	45	0,775	0,60
Δομή 3	5	5	5	97	0,788	0,62	24	0,777	0,60
Δομή 4	10	-	-	66	0,785	0,62	16	0,776	0,60
Δομή 5	10	10	-	52	0,793	0,63	27	0,78	0,61
Δομή 6	10	10	10	33	0,786	0,62	27	0,78	0,61
Δομή 7	15	-	-	51	0,79	0,62	43	0,785	0,62
Δομή 8	15	15	-	25	0,789	0,62	21	0,777	0,60
Δομή 9	15	15	15	19	0,791	0,63	19	0,778	0,61
Δομή 10	20	-	-	66	0,794	0,63	63	0,785	0,62
Δομή 11	20	20	-	16	0,79	0,62	14	0,78	0,61
Δομή 12	20	20	20	14	0,794	0,63	16	0,779	0,61

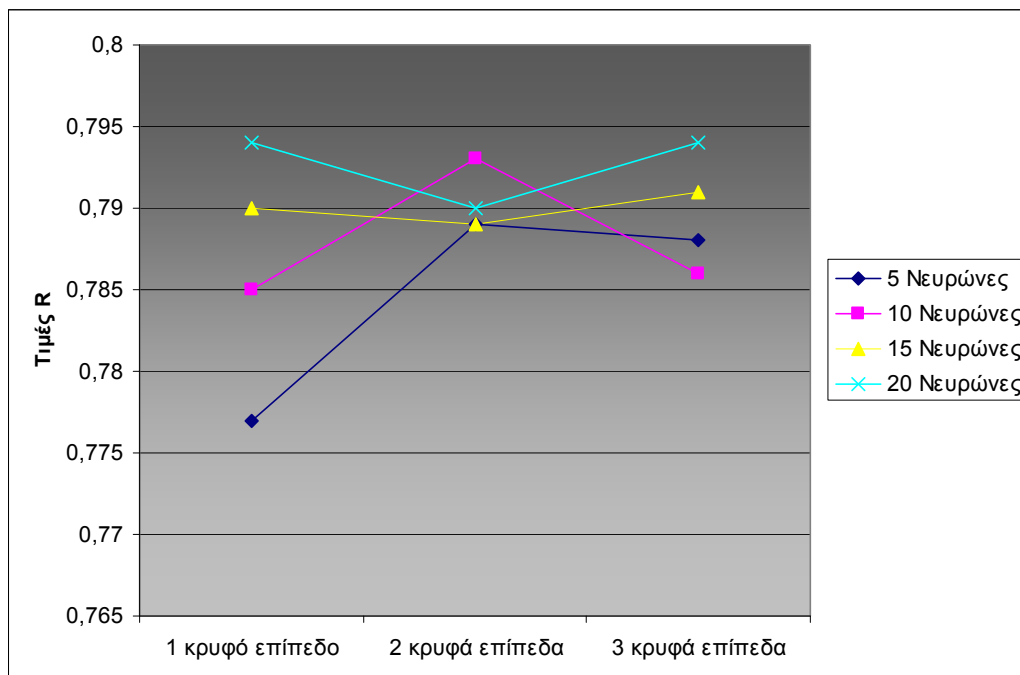
Πίνακας 11. Πίνακας αποτελεσμάτων απόδοσης της πρόβλεψης χρησιμοποιώντας διάφορες δομές νευρωνικών δικτύων.

Επίσης δημιουργήσαμε τις παρακάτω γραφικές παραστάσεις για να μπορέσουμε να μελετήσουμε και να αξιολογήσουμε τα αποτελέσματα κατά την αύξηση των επιπέδων σε σχέση με την αύξηση των νευρώνων.

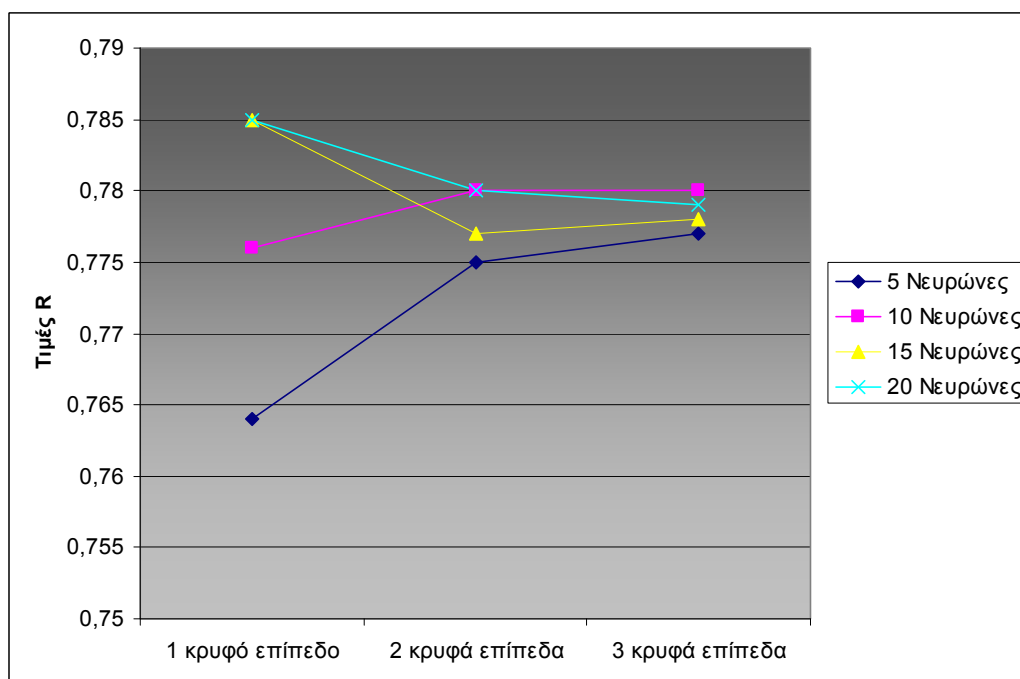
Από τον παραπάνω πίνακα αρχικά βλέπουμε ότι υπήρξε μια μικρή μείωση στην απόδοση της πρόβλεψης μας, χρησιμοποιώντας σειριακά τα δεδομένα για την φάση της εκμάθησης.

Στο νευρωνικό δίκτυο χρησιμοποιώντας εναλλάξ τα δεδομένα, βλέπουμε ότι όταν χρησιμοποιούμε ένα επίπεδο κατά την αύξηση των νευρώνων θα έχουμε καλύτερα αποτελέσματα, αλλά με δύο ή τρία επίπεδα δεν ισχύει απαραίτητα κάτι τέτοιο. Καλύτερα αποτελέσματα είχαμε κατά τις δοκιμές 10 και 12 με τιμή προσέγγισης 0,794. Αλλά είναι προτιμότερο το νευρωνικό δίκτυο της δοκιμής 10 (μοντέλο 1), γιατί χρησιμοποιεί λιγότερα επίπεδα, έτσι θα είναι ταχύτερο και πιο απλό κατά την διαδικασία εκμάθησης.

Στο νευρωνικό δίκτυο όπου χρησιμοποιούμε σειριακά τα δεδομένα, βλέπουμε ότι γενικά χρησιμοποιώντας ένα επίπεδο είχαμε τα καλύτερα αποτελέσματα. Η δοκιμή 7 (μοντέλο 2, με 1 επίπεδο 15 νευρώνων) είχε την καλύτερη απόδοση με την χρήση λιγότερων νευρώνων.



Σχήμα 55. Γραφική παράσταση απόδοσης ως προς τον αριθμό επιπέδων και νευρώνων για το νευρωνικό δίκτυο όπου χρησιμοποιούμε εναλλάξ τα δεδομένα για εκμάθηση.



Σχήμα 56. Γραφική παράσταση απόδοσης ως προς τον αριθμό επιπέδων και νευρώνων για το νευρωνικό δίκτυο όπου χρησιμοποιούμε σειριακά τα δεδομένα για εκμάθηση.

Για να μπορέσουμε να συγκρίνουμε αλλά και να αξιολογήσουμε τα αποτελέσματα των δύο μοντέλων (1 & 2) νευρωνικών δικτύων (με τα καλύτερα αποτελέσματα), δημιουργήσαμε τους παρακάτω πίνακες, με στατιστικά στοιχεία τα οποία συνιστώνται από Wilmott (1982, 1985) [17].

Σταθμός Πατησίων	Οριακή τιμή βενζολίου 5 $\mu\text{g}/\text{m}^3$		Οριακή τιμή βενζολίου 10 $\mu\text{g}/\text{m}^3$	
	Μοντέλο 1	Μοντέλο 2	Μοντέλο 1	Μοντέλο 2
Observed mean	7,184	7,184	7,184	7,184
Predicted mean	6,802	6,930	6,802	6,930
Observed standard deviation	5,003	5,003	5,003	5,003
Predicted standard deviation	2,507	2,520	2,507	2,520
Normalised mean difference (NMD)	0,053	0,035	0,053	0,035
Root mean square error (RMSE)	3,449	3,412	3,449	3,412
RMSE systematic	0,030	0,030	0,030	0,030
RMSE unsystematic	0,038	0,038	0,038	0,038
Correlation coefficient	0,78	0,785	0,780	0,785
Index of agreement (IA)	0,772	0,777	0,772	0,777
Mean absolute error (MAE)	2,176	2,163	2,176	2,163
Mean percentage error (MPE)	0,510	0,523	0,510	0,523
Mean bias error (MBE)	0,382	0,255	0,382	0,255
Number of observed alarms	8164	8164	3103	3103
Number of predicted alarms	9178	9425	1972	2125
Number of hours with correct alarm (A)	7559	7635	1682	1794
Correct Alarm Rate(%)	82,4%	81%	85,3%	84,4%
Number of hours with false alarm (C)	1619	1790	290	331
False Alarm Rate(%)	17,6 %	19%	14,7%	15,6%
B	605	523	1421	1309
Critical success index (CSI)	0,773	0,767	0,496	0,522

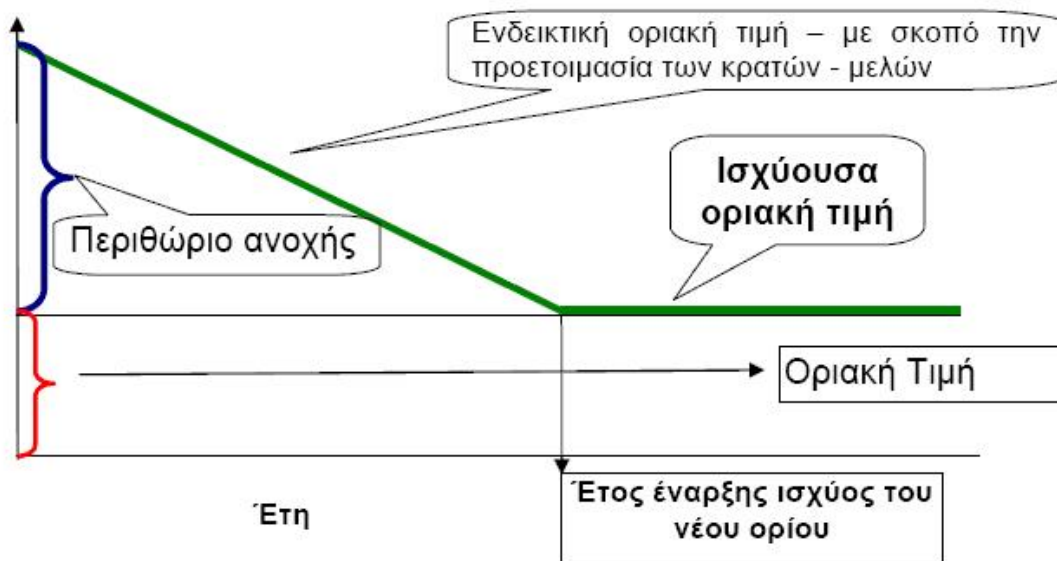
Πίνακας 12. Πίνακας στατιστικών στοιχείων, για την μελέτη της απόδοσης των τεχνητών νευρωνικών δικτύων.

Στον παραπάνω πίνακα μελετάμε τα δύο μοντέλα μας για οριακή τιμή βενζολίου 5 και 10 $\mu\text{g}/\text{m}^3$. Λόγω του ότι στην χώρα μας ισχύουν νομοθετικά όρια για τους ρύπους βενζολίου αλλά επίσης και τους ρύπους διοξειδίου του θείου, καπνού, αιωρούμενων σωματιδίων, διοξειδίου του αζώτου, μόλυβδου, όζοντος, μονοξειδίου του άνθρακα, σύμφωνα με τα όρια ποιότητας ατμόσφαιρας που έχουν καθιερωθεί στην Ευρωπαϊκή Ένωση [18]. Με μια σειρά από νέες οδηγίες σχετικά με την ατμοσφαιρική ρύπανση, θεσπίστηκαν από την Ευρωπαϊκή Ένωση, πέραν των άλλων, νέα όρια για τους διάφορους ατμοσφαιρικούς ρύπους. Τα όρια αυτά αναφέρονται τόσο στην προστασία της ανθρώπινης υγείας όσο και των οικοσυστημάτων. Παρακάτω παρουσιάζομαι τον πίνακα ορίων για το βενζόλιο, σύμφωνα με την οδηγία 2000/69/ΕΚ (το όριο θα ισχύσει από 1.1.2010).

Όρια βενζολίου	Ενδεικτική οριακή τιμή σε $\mu\text{g}/\text{m}^3$					Οριακή τιμή, $\mu\text{g}/\text{m}^3$
	2005	2006	2006	2008	2009	
Μέση ετήσια τιμή	10	9	8	7	6	5

Με τις οδηγίες αυτές για κάθε ρύπο ορίζεται μία οριακή τιμή για την προστασία της ανθρώπινης υγείας, με το αντίστοιχο έτος έναρξης ισχύος της. Παράλληλα δίνεται και ένα περιθώριο ανοχής, το οποίο αθροίζεται στην οριακή τιμή, δίνοντας έτσι την ενδεικτική οριακή τιμή, η οποία ισχύει στο μεσοδιάστημα έως την θέση σε ισχύ της οριακής τιμής. Το περιθώριο ανοχής κάθε χρόνο μειώνεται, έτσι ώστε στην ημερομηνία ισχύος του νέου ορίου να μηδενιστεί.

Επιπρόσθετα, τα κράτη μέλη πρέπει να εκπονούν και να υλοποιούν σχέδια δράσης για την προετοιμασία τους όσον αφορά στην επίτευξη και τήρηση των ορίων.



Σχήμα 56. Επεξήγηση της εφαρμογής της τιμής στόχου και οριακής τιμής με βάση τις οδηγίες της Ε.Ε.

Συγκεκριμένα η σταθερή απόκλιση (standard deviation) των πραγματικών και των προβλεπόμενων τιμών, προσδιορίζουν το ποσό της διαφοράς που το μοντέλο καταλαμβάνει συγκρίνοντας το με τα πραγματικά δεδομένα.

Η μέση τιμή του απόλυτου σφάλματος (mean absolute error: MAE) και η ρίζα του μέσου τετραγωνικού σφάλματος (RMSE) συνοψίζουν την διαφορά μεταξύ των πραγματικών και προβλεπόμενων τιμών.

Στη συνέχεια υπολογίσαμε το Systematic RMSE που μας δείχνει την διαφορά μεταξύ των πραγματικών και προβλεπόμενων τιμών, και το Unsystematic RMSE το οποίο υπολογίζεται για τους υπόλοιπους παράγοντες που δεν μπορούν να ελεγχθούν. Ένα καλό μοντέλο θα πρέπει να έχει unsystematic RMSE πολύ μεγαλύτερο από το systematic. Παρατηρούμε ότι τα επιλεγόμενα μοντέλα παρουσιάζουν μικρή διαφορά.

Ένα πιο χρήσιμο μέτρο για την απόδοση μοντέλων είναι το Index Of Agreement (IA), το οποίο παίρνει τιμές από μηδέν έως ένα., όπου στο μηδέν έχουμε την χειρότερη πρόβλεψη και στο ένα την καλύτερη. Έτσι εάν έχουμε παρόμοια τιμή IA σε δύο μοντέλα, συμπεραίνουμε ότι θα έχουν και παρόμοια ικανότητα πρόβλεψης. Όπως συμβαίνει στα δικά μας δύο μοντέλα.

Το False alarm rate είναι το ποσοστό των περιπτώσεων, για τις οποίες η πρόβλεψη έδειξε υπέρβαση του καθορισμένου ορίου (ύπαρξη alarm), αλλά στην πραγματικότητα δεν είχε συμβεί κάτι τέτοιο. Ένα αξιόπιστο μοντέλο δεν πρέπει να έχει μόνο μεγάλο ποσοστό επιτυχών προβλέψεων υπέρβασης ορίου, αλλά και όσο το δυνατόν πιο μικρό ποσοστό false alarm (Jorquera et.al.,1998) [17].

Το μέτρο Critical Success Index (CSI) καθορίζει το πόσο καλά προβλεύτηκαν οι τιμές υψηλής ρύπανσης, και δεν επηρεάζεται από τον αριθμό των σωστά προβλεπόμενων υπερβάσεων ορίου.

Γενικά και τα δύο μοντέλα έχουν πολύ καλά ποσοστά correct και false alarm. Θα πρέπει να σημειωθεί πάντως ότι το δεύτερο μοντέλο έχει λίγο καλύτερα ποσοστά και στο correct και στο false alarms, όπου στο ποσοστό correct alarms έχει λίγο μεγαλύτερες τιμές και στο ποσοστό false alarms έχει λίγο μικρότερες. Όμως στο δεύτερο μοντέλο οι τιμές B (υπερβάσεις του ορίου που έχουν πραγματοποιηθεί αλλά δεν έχουν προβλευτεί), είναι πολύ μεγαλύτερες από αυτές του πρώτου. Αυτός είναι και ο λόγος που το δεύτερο μοντέλο έχει μικρότερες τιμές CSI από το πρώτο.

5. Μελέτη χρονοσειρών όζοντος (σταθμός Πατ. & Λιο.) [B]

5.1 Ελλιπής τιμές

Σε αυτό το κεφάλαιο θα μελετήσουμε την χρονοσειρά όζοντος και κάνουμε πρόβλεψη ημερησίας μέγιστης μέσης τιμής κυλιόμενου οκταώρου (MMTK) για τον “σταθμό” Πατησίων και Λιοσίων. Αρχικά βλέπουμε ότι και σε αυτές τις ομάδες δεδομένων υπάρχουν ελλείψεις τιμών (χρήση τιμής -9999), από τα συστήματα καταγραφής. Όπως και στο Κεφ. 3.1 οι ημέρες οι οποίες είχαν ελλείψεις περισσότερες η ίσες του 30%, διαγράφονται από τα δεδομένα μας, γιατί δεν θα ήταν δυνατή η ανάκτηση τους χρησιμοποιώντας μεθόδους όπως την παρεμβολή Newton ή Lagrange. Για τις υπόλοιπες ημέρες χρησιμοποιήσαμε την παρεμβολή Lagrange για να υπολογίσουμε την ελλιπείς τιμές. Στον παρακάτω πίνακα φαίνονται οι ημέρες που είχαν ελλείψεις περισσότερες του 30% και έτσι διαγράφηκαν από το σύνολο των δεδομένων. Η διαγραφή έγινε ανά σταθμό και έτος.

Σταθμός Πατησίων					
Έτος 2002		Έτος 2003		Έτος 2005	
Ημέρες	Ημερομηνίες	Ημέρες	Ημερομηνίες	Ημέρες	Ημερομηνίες
8 - 14	8/1/02 – 14/1/02	123 – 125	3/5/03 – 5/5/03	83 – 89	24/3/05 – 30/3/05
19 - 21	19/1/02 – 21/2/02	209 - 365	28/7/03 – 31/12/03	240 – 241	28/8/05 – 29/8/05
157 - 158	6/6/02 – 7/6/02			321	17/11/05
189	8/7/02			326 – 329	22/11/05 – 25/11/05
210 - 252	29/7/02 – 9/9/02				
Συνολικά: 56		Συνολικά: 160		Συνολικά: 14	

Πίνακας 13. Ημέρες οι οποίες έχουν διαγραφεί από τον σταθμό Πατησίων λόγω ελλείψεων.

Σταθμός Λιοσίων			
Έτος 2002		Έτος 2003	
Ημέρες	Ημερομηνίες	Ημέρες	Ημερομηνίες
148 – 149	28/5/02 – 29/5/02	37 – 42	6/2/03 – 11/2/03
247	4/9/02	44 – 52	13/2/03 – 21/2/03
		102 – 104	12/4/03 – 14/4/03
		306 – 346	2/11/03 – 12/12/03
		363 - 364	29/12/03 – 30/12/03
Συνολικά: 3		Συνολικά: 61	

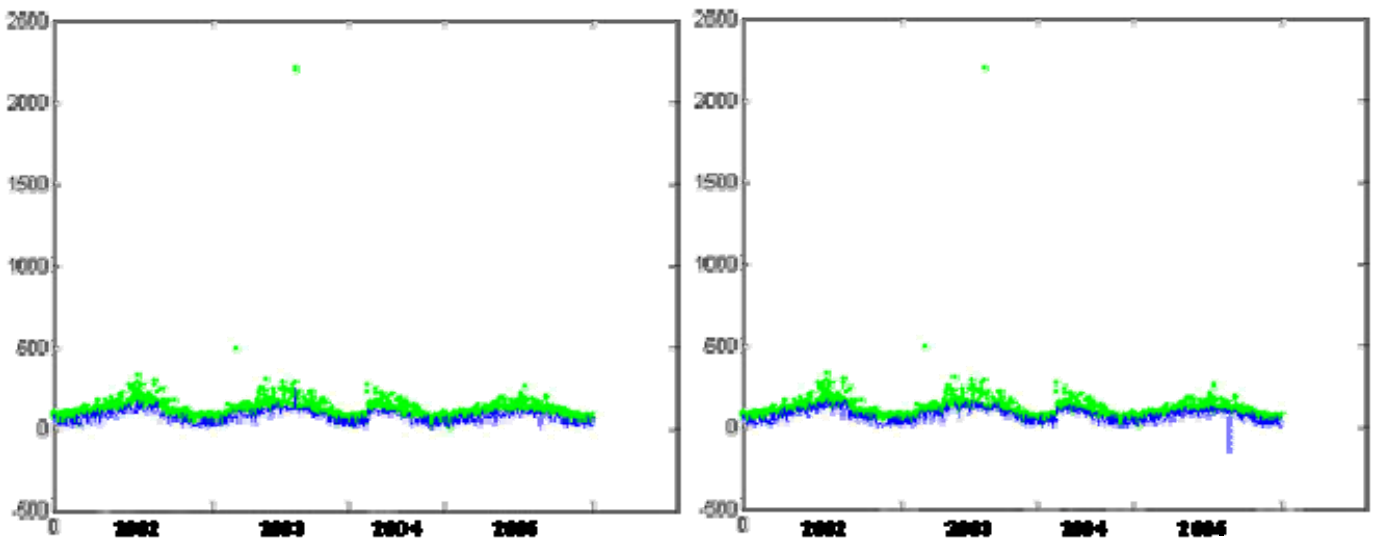
Πίνακας 14. Ημέρες οι οποίες έχουν διαγραφεί από τον σταθμό Λιοσίων για τα έτη 2002-2003 λόγω ελλείψεων.

Σταθμός Λιοσίων			
Έτος 2004		Έτος 2005	
Ημέρες	Ημερομηνίες	Ημέρες	Ημερομηνίες
45 – 183	14/2/04 – 1/7/04	37 – 38	6/2/05 – 7/2/05
188 – 191	6/7/04 – 9/7/04	104 – 105	14/4/05 – 15/4/05
281	7/10/04	173 – 182	22/6/05 – 1/7/05
283 – 285	9/10/04 – 11/10/04	185 – 186	4/7/05 – 5/7/05
287	13/10/04	225 – 228	13/8/05 – 16/8/05
313	8/11/04	240 - 241	28/8/05 – 29/8/05
		266	23/9/05
		282 – 283	9/10/05 – 10/10/05
		302 – 304	29/10/05 – 31/10/05
		321	17/11/05
		328	24/11/05
		363	29/12/05
Συνολικά: 149		Συνολικά: 31	

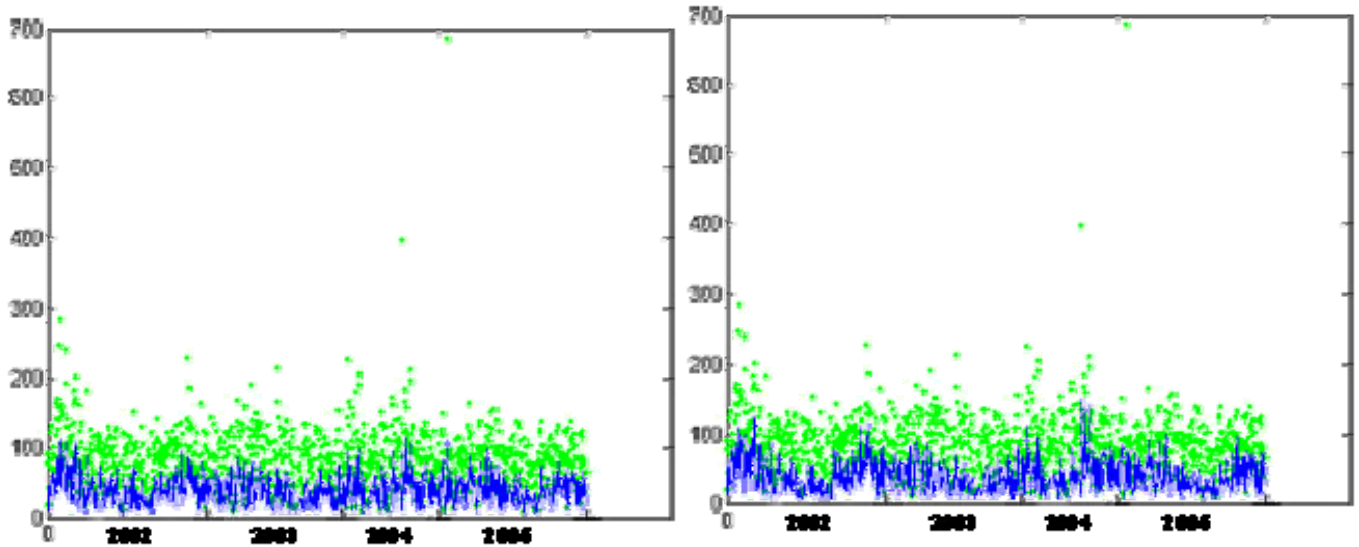
Πίνακας 15. Ημέρες οι οποίες έχουν διαγραφεί από τον σταθμό Λιοσίων για τα έτη 2004-2005 λόγω ελλείψεων.

5.2 Υπολογισμός χαρακτηριστικών

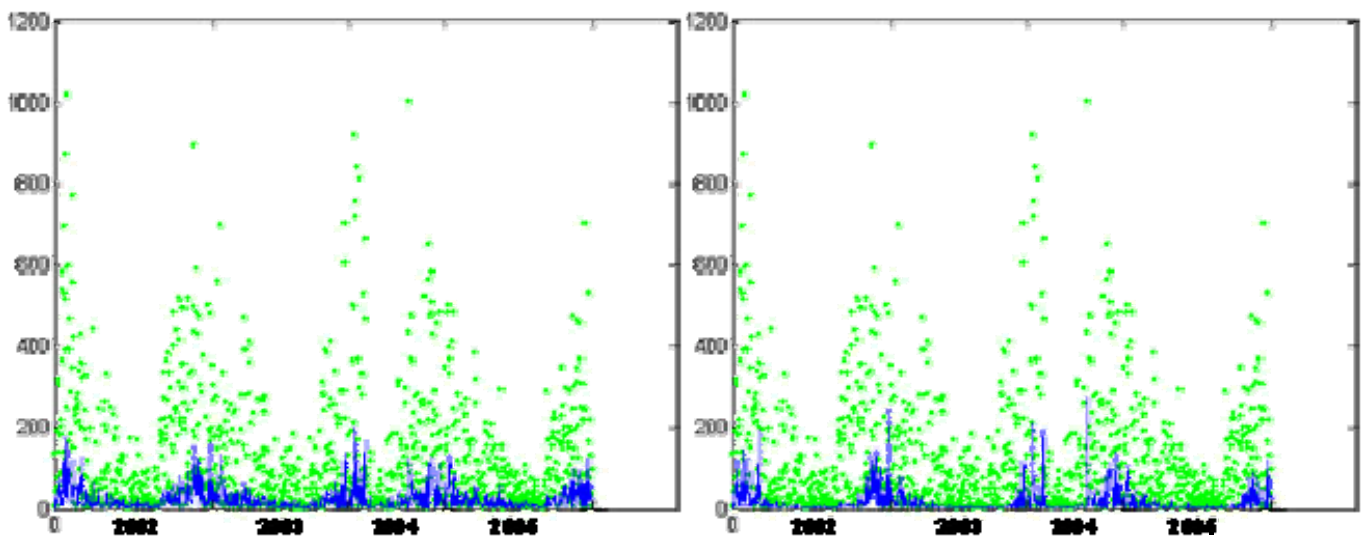
Στην συνέχεια για κάθε ημέρα αλλά και σταθμό, υπολογίσαμε την μέγιστη και μέση ωριαία τιμή. Επίσης υπολογίσαμε την μέγιστη μέση τιμή κυλιόμενου οκταώρου (MMTKO) για κάθε ημέρα. Για τον υπολογισμό της, υπολογίσαμε 24 μέσες τιμές οκταώρου για κάθε ημέρα και επιλέξαμε την μεγαλύτερη. Για την πρώτη τιμή της κάθε ημέρας χρησιμοποιήσαμε τις ωριαίες τιμές από 17:00 (προηγούμενης Ημέρας) - 01:00 (τρέχουσας) συνεχίζοντας διαδοχικά μέχρι το τελευταίο οκτάωρο από 16:00 (τρέχουσας) - 24:00. Την MMTKO την υπολογίσαμε διότι είναι το κριτήριο που θέτει η σχετική περιβαλλοντική νομοθεσία για την προστασία της ανθρώπινης υγείας. Παρακάτω εμφανίζουμε τις γραφικές παραστάσεις αυτών αλλά και μέσες ημερήσιες τιμές της ταχύτητας του ανέμου στις 9:00 π.μ, αρχικά για τον σταθμό Λιοσίων και στην συνέχεια για τον σταθμό Πατησίων.



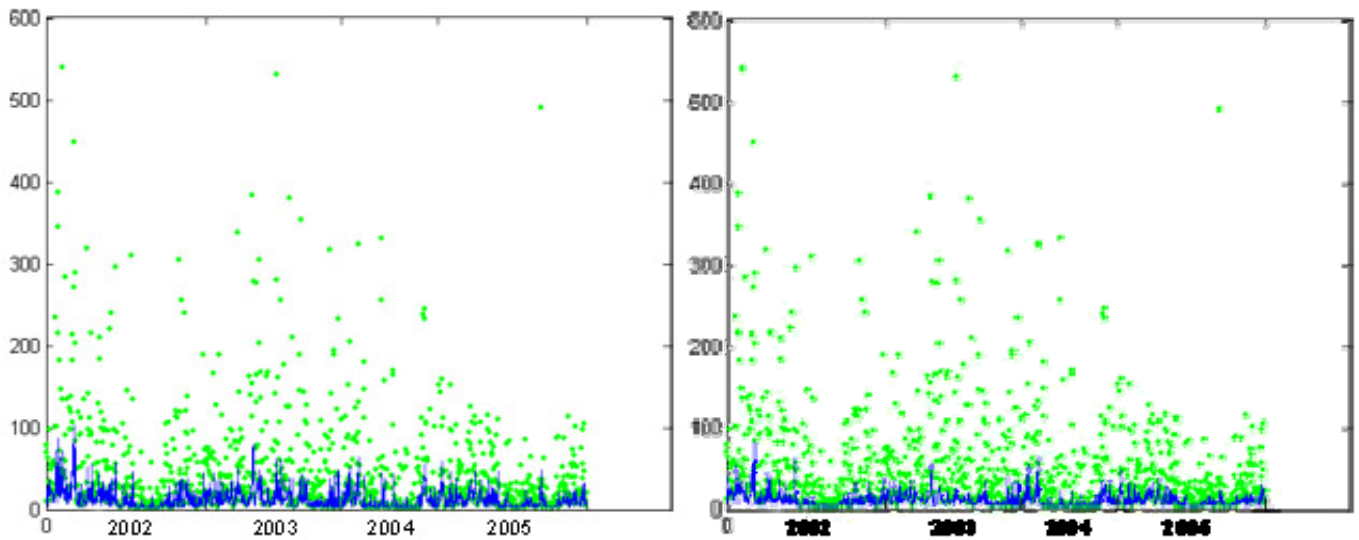
Σχήμα 57 & 58. Αριστερά φαίνονται οι μέσες (μπλε γραμμή) και μέγιστες τιμές (πράσινες κουκίδες) όζοντος (O₃) για κάθε ημέρα για τα έτη 2002-2005. Δεξιά φαίνονται οι MMTKO, πάλι σε σχέση με τις μέγιστες τιμές.



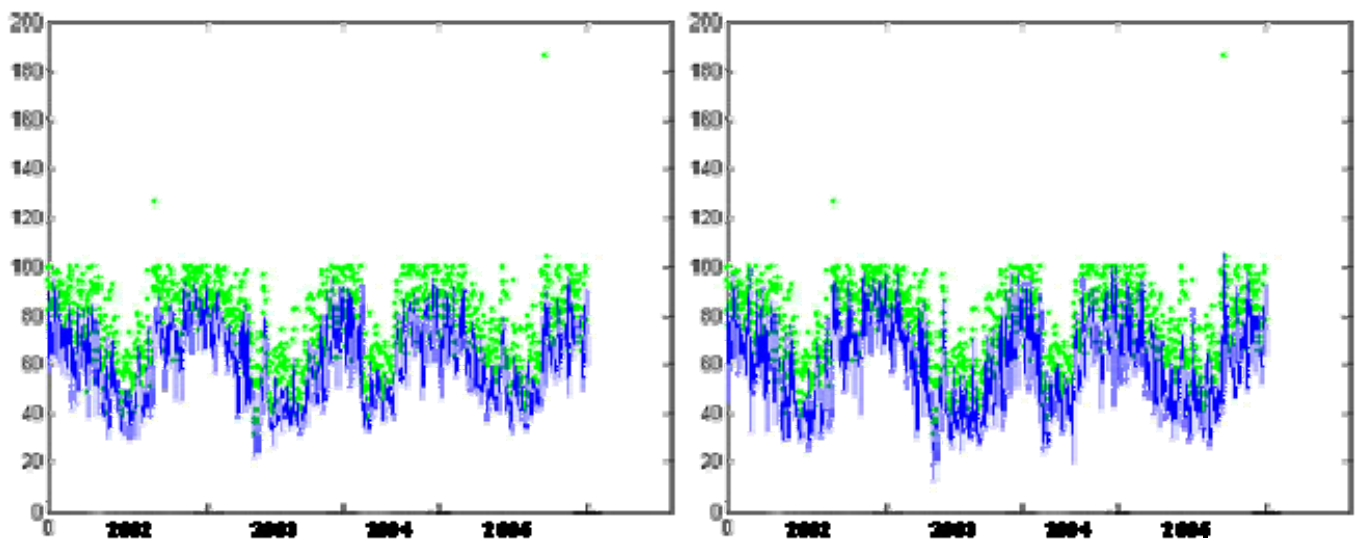
Σχήμα 59 & 60. Αριστερά φαίνονται οι μέσες (μπλε γραμμή) και μέγιστες τιμές (πράσινες κουκίδες) διοξειδίου του αζώτου (NO₂) για κάθε ημέρα για τα έτη 2002-2005 από τον σταθμό Λιοσίων. Δεξιά φαίνονται οι ΜΜΤΚΟ, πάλι σε σχέση με τις μέγιστες τιμές.



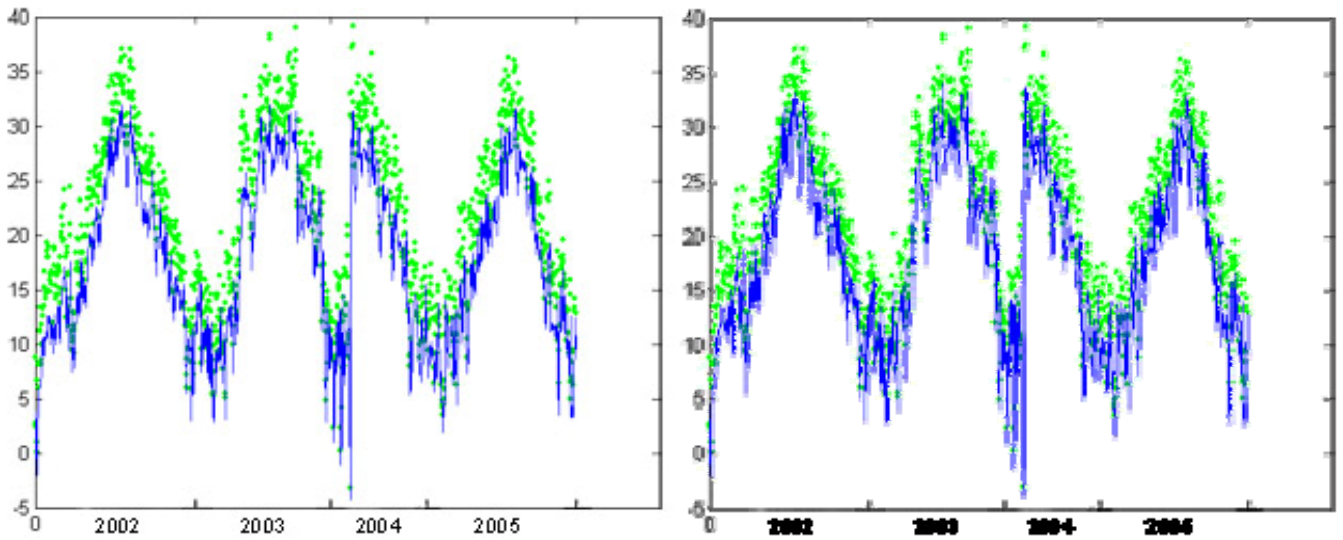
Σχήμα 60 & 61. Αριστερά φαίνονται οι μέσες (μπλε γραμμή) και μέγιστες τιμές (πράσινες κουκίδες) μονοξειδίου του αζώτου (NO) για κάθε ημέρα για τα έτη 2002-2005 από τον σταθμό Λιοσίων. Δεξιά φαίνονται οι ΜΜΤΚΟ, πάλι σε σχέση με τις μέγιστες τιμές.



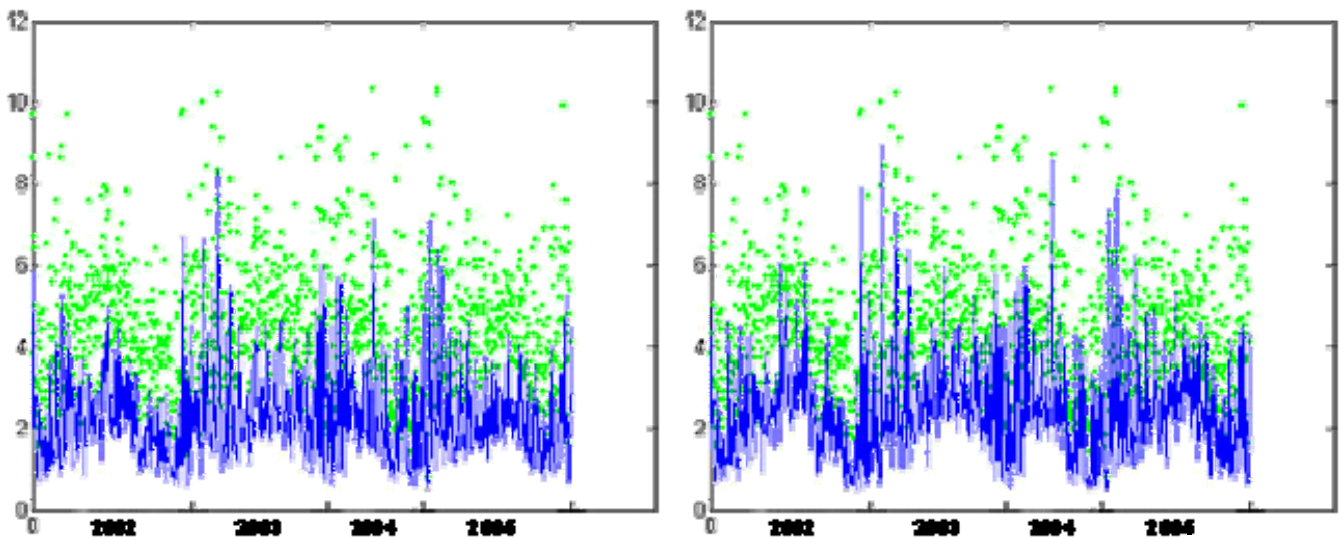
Σχήμα 62 & 63. Αριστερά φαίνονται οι μέσες (μπλε γραμμή) και μέγιστες τιμές (πράσινες κουκίδες) διοξειδίου του θείου (SO₂) για κάθε ημέρα για τα έτη 2002-2005 από τον σταθμό Λιοσίων. Δεξιά φαίνονται οι MMTKO, πάλι σε σχέση με τις μέγιστες τιμές.



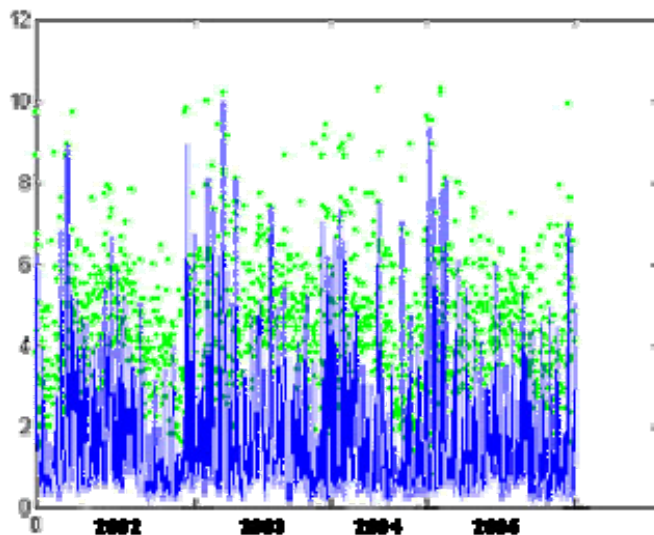
Σχήμα 64 & 65. Αριστερά φαίνονται οι μέσες (μπλε γραμμή) και μέγιστες τιμές (πράσινες κουκίδες) σχετικής υγρασίας (RH) για κάθε ημέρα για τα έτη 2002-2005 από τον σταθμό Λιοσίων. Δεξιά φαίνονται οι MMTKO, πάλι σε σχέση με τις μέγιστες τιμές.



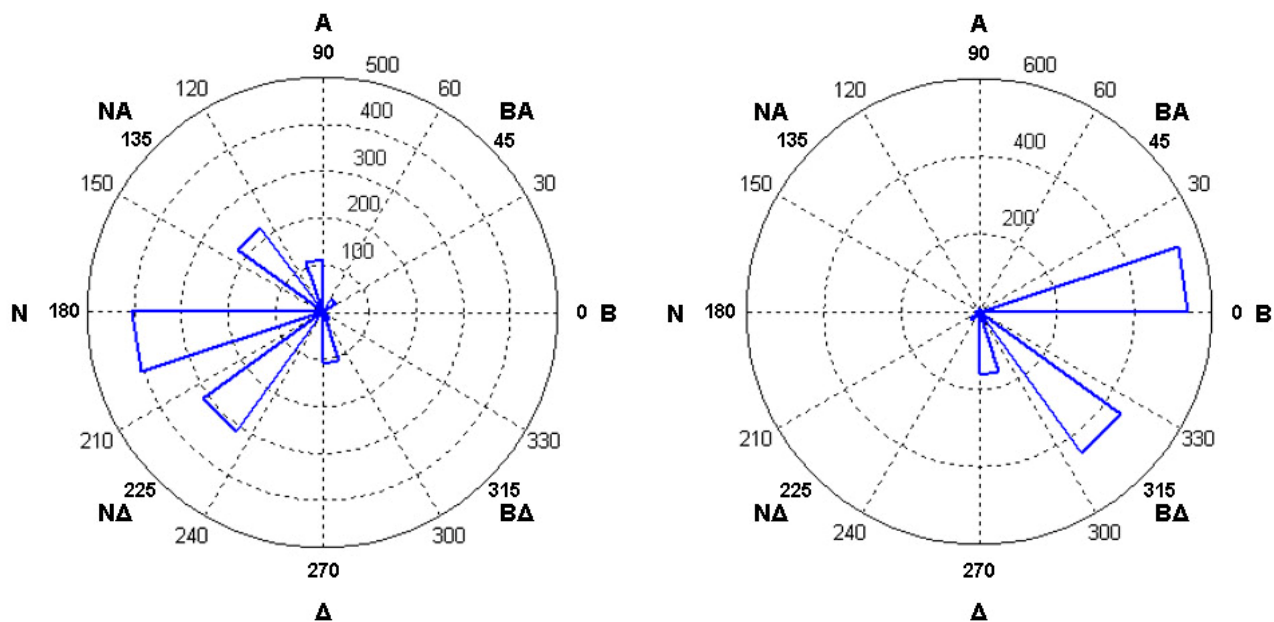
Σχήμα 66 & 67. Αριστερά φαίνονται οι μέσες (μπλε γραμμή) και μέγιστες τιμές (πράσινες κουκίδες) θερμοκρασίας (Ta) για κάθε ημέρα για τα έτη 2002-2005 από τον σταθμό Λιοσίων. Δεξιά φαίνονται οι MMTKO, πάλι σε σχέση με τις μέγιστες τιμές.



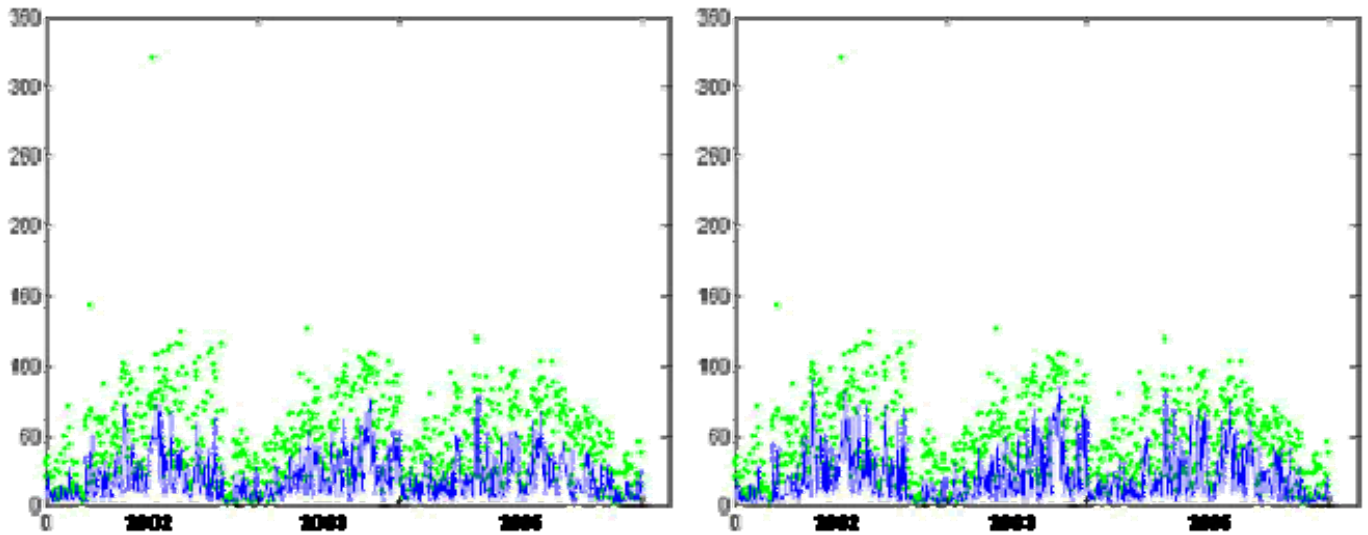
Σχήμα 68 & 69. Αριστερά φαίνονται οι μέσες (μπλε γραμμή) και μέγιστες τιμές (πράσινες κουκίδες) ταχύτητας του ανέμου (WS) για κάθε ημέρα για τα έτη 2002-2005 από τον σταθμό Λιοσίων. Δεξιά φαίνονται οι MMTKO, πάλι σε σχέση με τις μέγιστες τιμές.



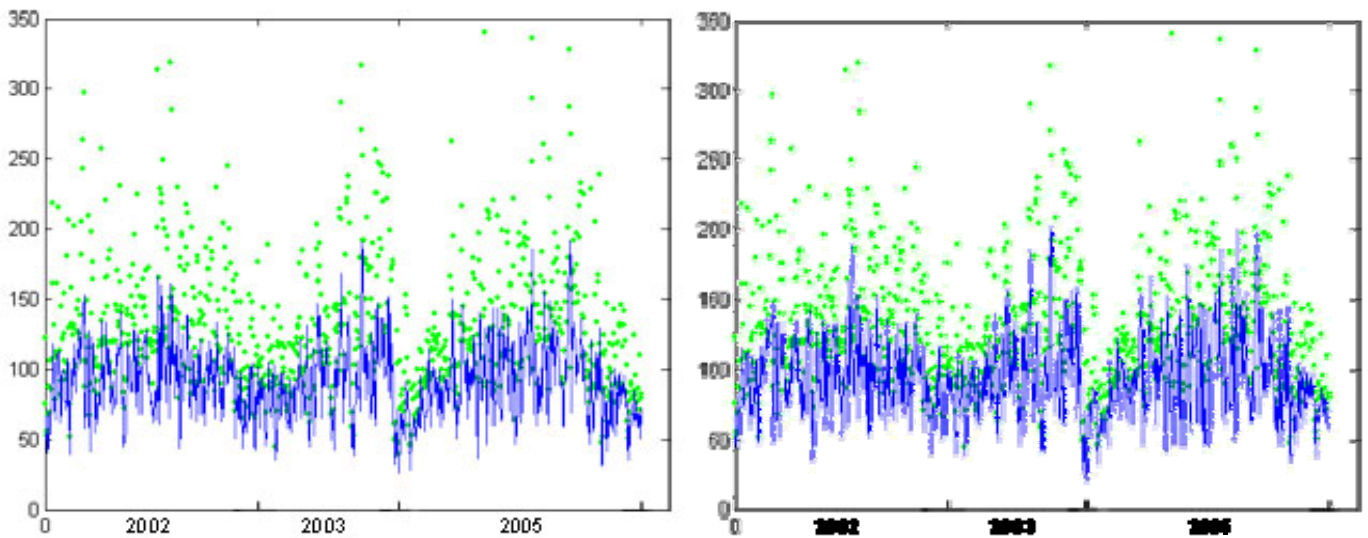
Σχήμα 70. Μέσες (μπλε γραμμή) και μέγιστες τιμές (πράσινες κουκίδες) ταχύτητας του ανέμου (WS) στις 9:00π.μ. για κάθε ημέρα για τα έτη 2002-2005 από τον σταθμό Λιοσίων.



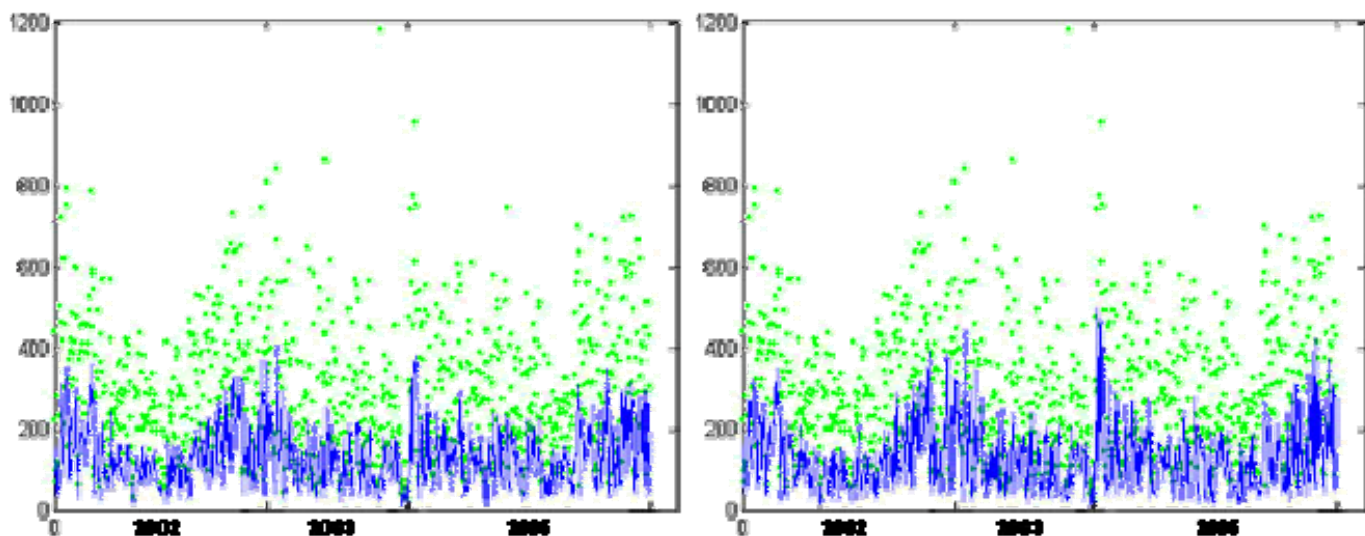
Σχήμα 71 & 72. Αριστερά φαίνεται ο μέσος όρος των ανέμων για τα έτη 2002-2005 για τον σταθμό Λιοσίων, και δεξιά οι μέγιστες τιμές για την ίδια περίοδο.



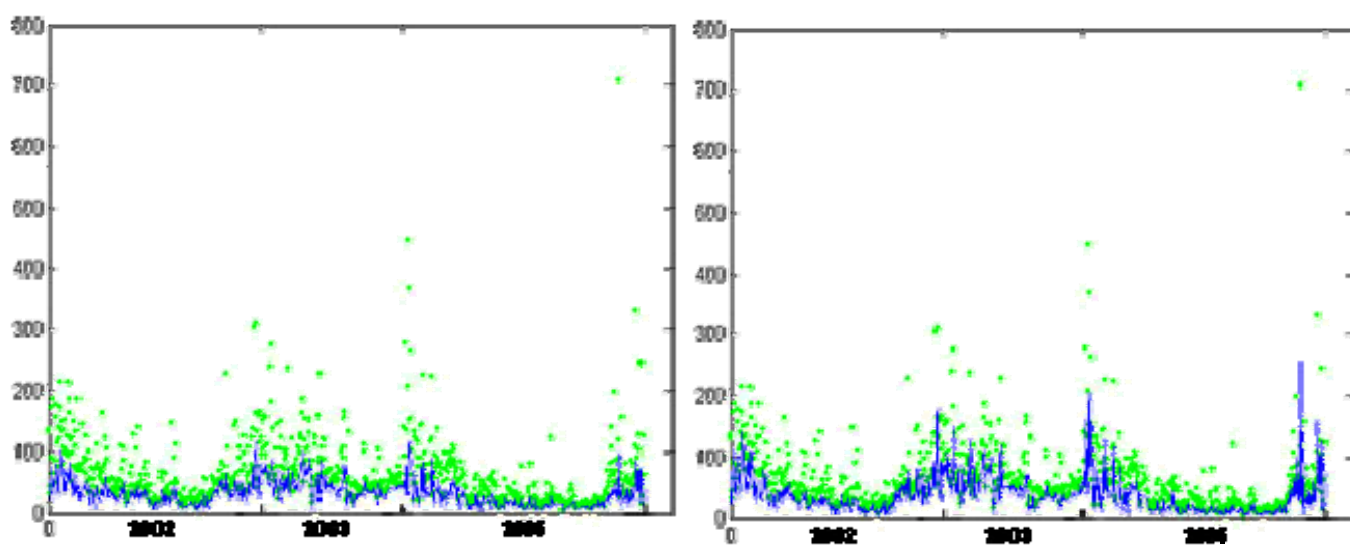
Σχήμα 73 & 74. Αριστερά φαίνονται οι μέσες (μπλε γραμμή) και μέγιστες τιμές (πράσινες κουκίδες) όζοντος (O₃) για κάθε ημέρα για τα έτη 2002,2003 και 2005 από τον σταθμό Πατησίων. Δεξιά φαίνονται οι ΜΜΤΚΟ, πάλι σε σχέση με τις μέγιστες τιμές.



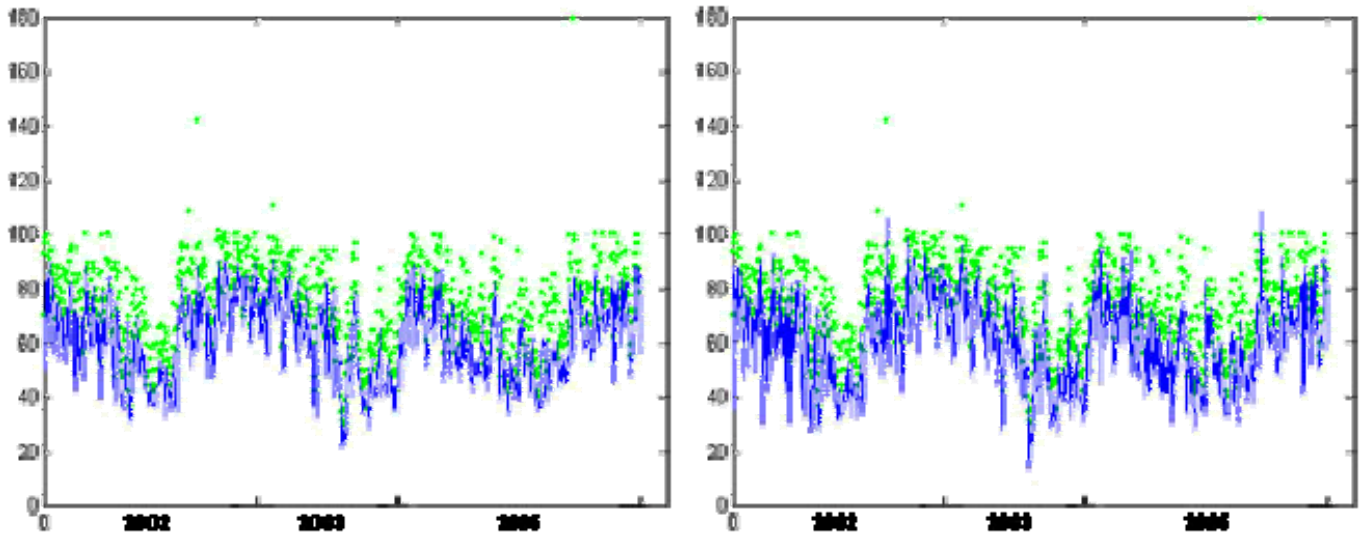
Σχήμα 75 & 76. Αριστερά φαίνονται οι μέσες (μπλε γραμμή) και μέγιστες τιμές (πράσινες κουκίδες) διοξειδίου του αζώτου (NO₂) για κάθε ημέρα για τα έτη 2002,2003 και 2005 από τον σταθμό Πατησίων. Δεξιά φαίνονται οι ΜΜΤΚΟ, πάλι σε σχέση με τις μέγιστες τιμές.



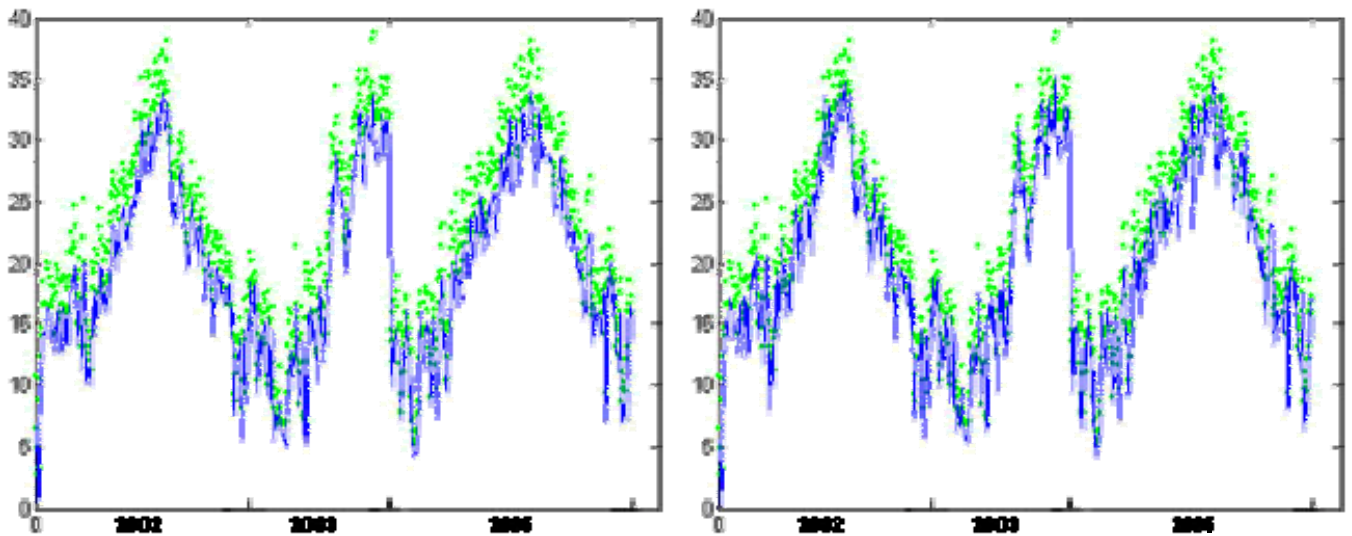
Σχήμα 77 & 78. Αριστερά φαίνονται οι μέσες (μπλε γραμμή) και μέγιστες τιμές (πράσινες κουκίδες) μονοξειδίου του αζώτου (NO) για κάθε ημέρα για τα έτη 2002,2003 και 2005 από τον σταθμό Πατησίων. Δεξιά φαίνονται οι ΜΜΤΚΟ, πάλι σε σχέση με τις μέγιστες τιμές.



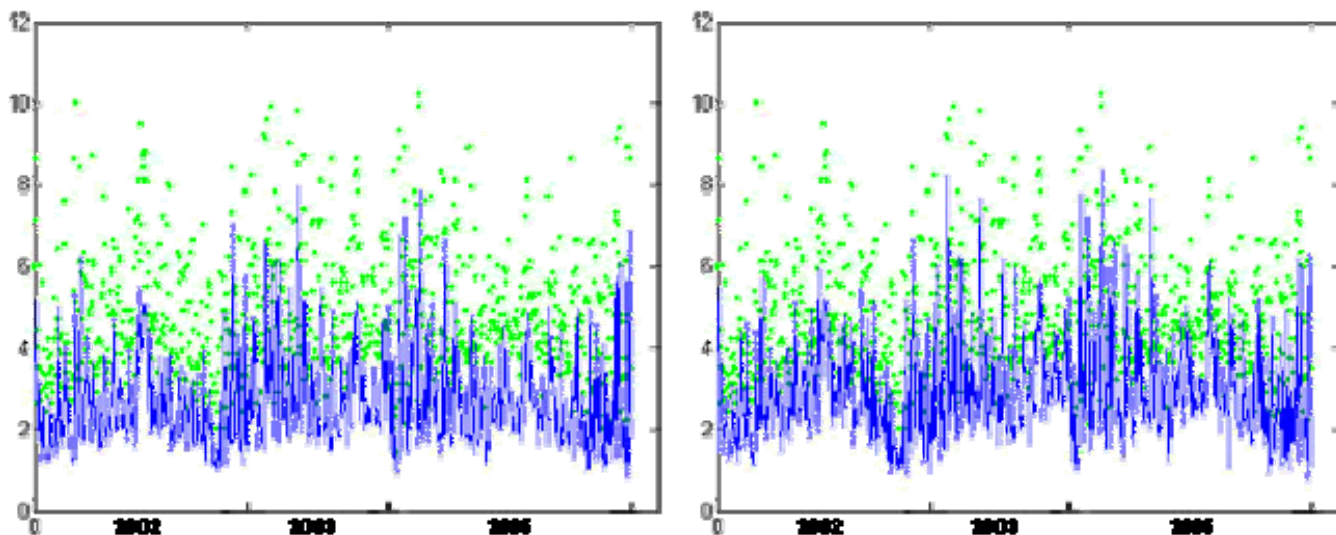
Σχήμα 79 & 80. Αριστερά φαίνονται οι μέσες (μπλε γραμμή) και μέγιστες τιμές (πράσινες κουκίδες) διοξειδίου του θείου (SO₂) για κάθε ημέρα για τα έτη 2002,2003 και 2005 από τον σταθμό Πατησίων. Δεξιά φαίνονται οι ΜΜΤΚΟ, πάλι σε σχέση με τις μέγιστες τιμές.



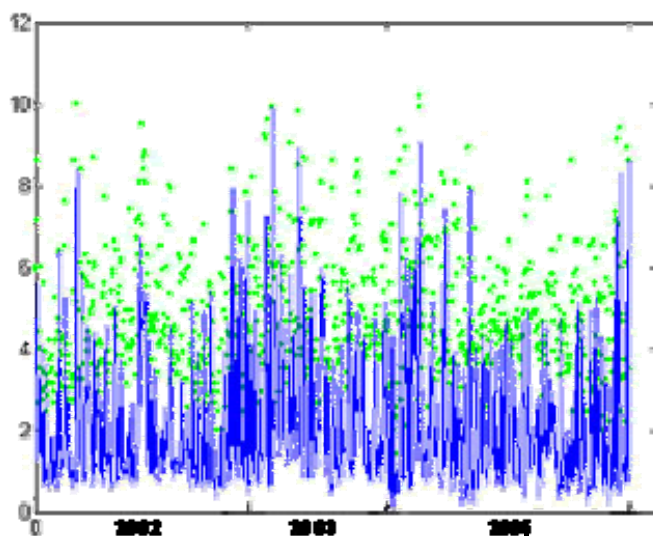
Σχήμα 81 & 82. Αριστερά φαίνονται οι μέσες (μπλε γραμμή) και μέγιστες τιμές (πράσινες κουκίδες) σχετικής υγρασίας (RH) για κάθε ημέρα για τα έτη 2002,2003 και 2005 από τον σταθμό Πατησίων. Δεξιά φαίνονται οι MMTKO, πάλι σε σχέση με τις μέγιστες τιμές.



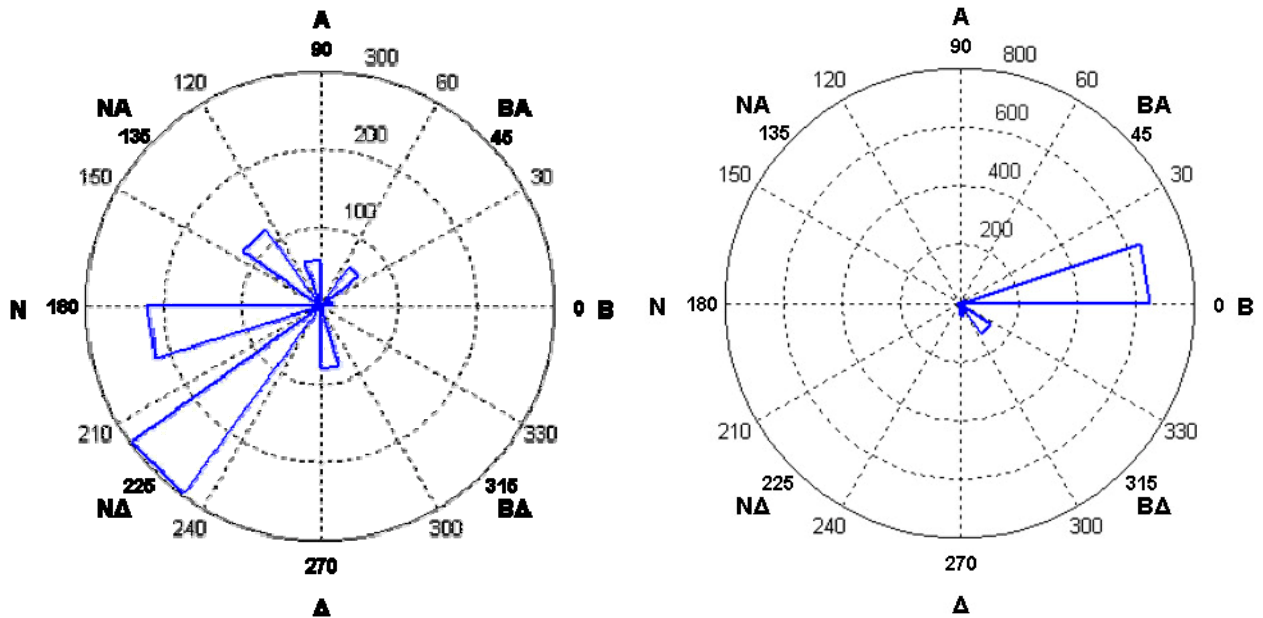
Σχήμα 83 & 84. Αριστερά φαίνονται οι μέσες (μπλε γραμμή) και μέγιστες τιμές (πράσινες κουκίδες) θερμοκρασίας (Ta) για κάθε ημέρα για τα έτη 2002,2003 και 2005 από τον σταθμό Πατησίων. Δεξιά φαίνονται οι MMTKO, πάλι σε σχέση με τις μέγιστες τιμές.



Σχήμα 85 & 86. Αριστερά φαίνονται οι μέσες (μπλε γραμμή) και μέγιστες τιμές (πράσινες κουκίδες) ταχύτητας του ανέμου (WS) για κάθε ημέρα για τα έτη 2002,2003 και 2005 από τον σταθμό Πατησίων. Δεξιά φαίνονται οι MMTKO, πάλι σε σχέση με τις μέγιστες τιμές.



Σχήμα 87. Μέσες (μπλε γραμμή) και μέγιστες τιμές (πράσινες κουκίδες) ταχύτητας του ανέμου (WS) στις 9:00π.μ. για κάθε ημέρα για τα έτη 2002,2003 και 2005 από τον σταθμό Πατησίων.



Σχήμα 88 & 89. Αριστερά φαίνεται ο μέσος όρος των ανέμων για τα έτη 2002,2003 και 2005 για τον σταθμό Πατησίων, και δεξιά οι μέγιστες τιμές για την ίδια περίοδο.

Από τις παραπάνω γραφικές παραστάσεις και των δύο σταθμών, βλέπουμε ότι οι μέσες τιμές και οι ΜΜΤΚΟ έχουν σχεδόν ίδιες τιμές. Επίσης βλέπουμε ότι σε κάποια χαρακτηριστικά (όπως στο ΝΟ, Ο3, Τα, RH) υπάρχει επανάληψη ανά έτος. Τέλος θα πρέπει να σημειωθεί ότι στο σταθμό Πατησίων οι πιο ισχυροί άνεμοι ήταν βόρειοι αλλά κυρίως ήταν νοτιοδυτικοί. Στο σταθμό Λιοσίων οι πιο ισχυροί άνεμοι ήταν οι βόρειοι και οι βορειοδυτικοί αλλά κυρίως ήταν νοτιοδυτικοί.

5.3 Υπολογισμός συντελεστή συσχέτισης (correlation coefficient)

Για να μελετήσουμε το πόσο ισχυρή είναι η συσχέτιση της εξαρτημένης μεταβλητής (ΜΜΤΚΟ όζοντος) με τα υπόλοιπα χαρακτηριστικά χρησιμοποιήσαμε τον συντελεστή συσχέτισης, όπως αναφέρεται και στο Κεφ. 3.3. Παρακάτω παραθέτουμε τον πίνακα με τα αποτελέσματα συσχέτισης.

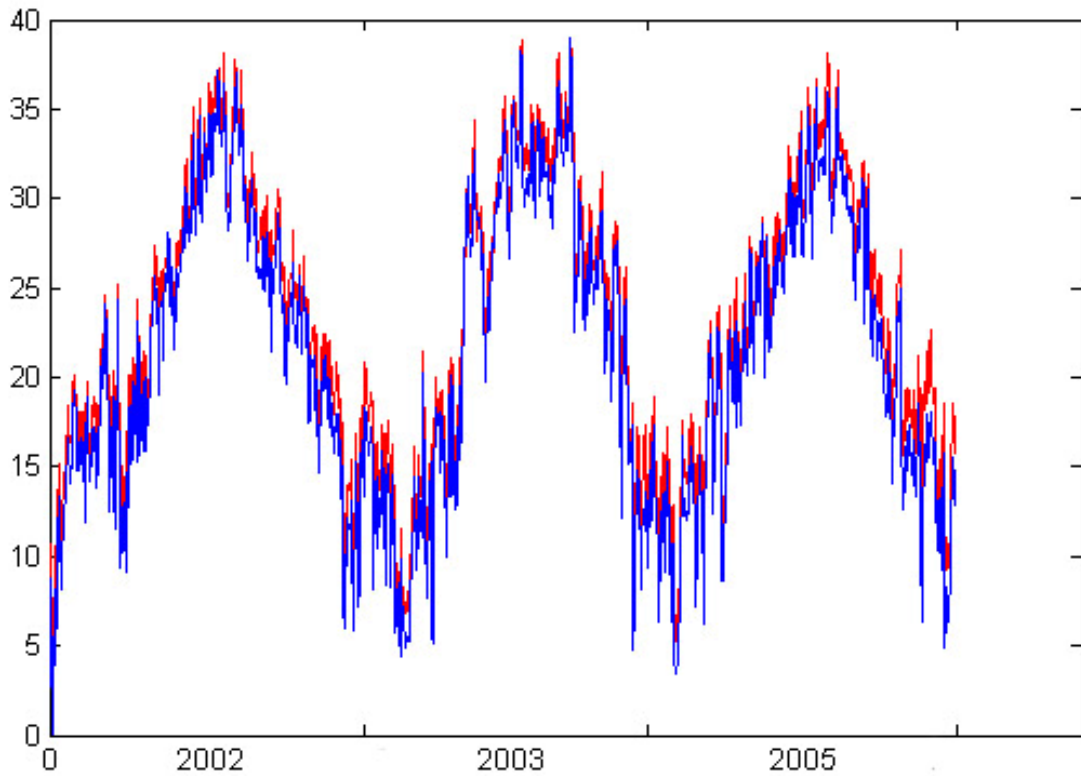
Συσχέτιση με ΜΜΤΚΟ Όζοντος		Σταθμός Πατησίων 02, 03, 05		Σταθμός Λιοσίων 02-05	
		ρ: δείκτης συσχέτισης	βαθμός συσχέτισης	ρ: δείκτης συσχέτισης	βαθμός συσχέτισης
RH	Μέγιστες Τιμ.	-0,50855	Μέτρια συσχ.	-0,67793	Μέτρια συσχ.
	Μέσες Τιμ.	-0,52141	Μέτρια συσχ.	-0,72501	Υψηλή συσχ.
	ΜΜΤΚΟ	-0,53021	Μέτρια συσχ.	-0,71448	Υψηλή συσχ.
Ta	Μέγιστες Τιμ.	0,40438	Χαμηλή συσχ.	0,6687	Μέτρια συσχ.
	Μέσες Τιμ.	0,4163	Χαμηλή συσχ.	0,69497	Μέτρια συσχ.
	ΜΜΤΚΟ	0,41149	Χαμηλή συσχ.	0,69676	Μέτρια συσχ.
WD	Μέγιστες Τιμ.	0,01109	Δεν υπάρχει	0,11173	Δεν υπάρχει
	Μέσες Τιμ.	-0,49096	Χαμηλή συσχ.	0,046919	Δεν υπάρχει
	ΜΜΤΚΟ	-0,53754	Μέτρια συσχ.	0,00257	Δεν υπάρχει
WS	Μέγιστες Τιμ.	0,38215	Χαμηλή συσχ.	0,34762	Χαμηλή συσχ.
	Μέσες Τιμ.	0,3991	Χαμηλή συσχ.	0,34727	Χαμηλή συσχ.
	ΜΜΤΚΟ	0,51859	Μέτρια συσχ.	0,53059	Μέτρια συσχ.
	9:00π.μ.	0,22374	Δεν υπάρχει	0,11532	Δεν υπάρχει
NO ₂	Μέγιστες Τιμ.	-0,12598	Δεν υπάρχει	0,15532	Δεν υπάρχει
	Μέσες Τιμ.	-0,28962	Δεν υπάρχει	-0,20123	Χαμηλή συσχ.
	ΜΜΤΚΟ	-0,42172	Μέτρια συσχ.	-0,4139	Μέτρια συσχ.
NO	Μέγιστες Τιμ.	-0,46657	Μέτρια συσχ.	-0,62867	Χαμηλή συσχ.
	Μέσες Τιμ.	-0,63858	Μέτρια συσχ.	-0,46563	Μέτρια συσχ.
	ΜΜΤΚΟ	-0,7129	Υψηλή συσχ.	-0,56106	Μέτρια συσχ.
SO ₂	Μέγιστες Τιμ.	-0,35786	Χαμηλή συσχ.	-0,080317	Δεν υπάρχει
	Μέσες Τιμ.	-0,36627	Χαμηλή συσχ.	-0,20084	Δεν υπάρχει
	ΜΜΤΚΟ	-0,38947	Χαμηλή συσχ.	-0,31586	Χαμηλή συσχ.

Πίνακας 15. Αποτελέσματα συντελεστή συσχέτισης ΜΜΤΚΟ όζοντος με τα υπόλοιπα χαρακτηριστικά.

Στον παραπάνω πίνακα παρατηρούμε ότι στα Πατήσια υπάρχει υψηλή συσχέτιση της ΜΜΤΚΟ του όζοντος με το μονοξείδιο του αζώτου (NO). Αυτή η σχέση μπορεί να υπάρξει κατά την διάρκεια έντονης ηλιακής ακτινοβολίας (αυξημένη θερμοκρασία) [Παράρτημα], γιατί υπό αυτές τις συνθήκες μπορεί να σχηματιστεί το όζον από οξειδία του αζώτου.

Αυτό δεν συνέβη στα Λιόσια γιατί όπως φαίνεται και στο παρακάτω σχήμα κατά την διάρκεια των μελετών μας (02,03,05) στα Πατήσια παρατηρήθηκαν μεγαλύτερες θερμοκρασίες απ' ότι στα Λιόσια. Συγκεκριμένα παρατηρήθηκε μέση θερμοκρασία υψηλότερη κατά 1,65 βαθμούς C°.

Επίσης παρατηρούμε υψηλή συσχέτιση των τιμών σχετικής υγρασίας (RH) με την ΜΜΤΚΟ του όζοντος. Συγκεκριμένα οι διακυμάνσεις της σχετικής υγρασίας επηρεάζουν αντιστρόφως ανάλογα τις τιμές του όζοντος.



Σχήμα 90. Γραφική παράσταση μέγιστων ημερησίων τιμών θερμοκρασίας για τα έτη 2002, 2003 και 2005 για το σταθμό Πατησίων (κόκκινη) σε σχέση με του σταθμό Λιοσίων (μπλε).

6. Πρόγνωση ΜΜΤΚΟ τιμών όζοντος

6.1 Εισαγωγή

Για την πρόγνωση τιμών όζοντος χρησιμοποιήθηκε νευρωνικό δίκτυο Forward Connection και για την εκπαίδευση του, χρησιμοποιήσαμε την μέθοδο Back Propagation, (βλέπε Κεφ. 4.3 και 4.4). Δημιουργήσαμε πολλές δομές νευρωνικών δικτύων, στις οποίες κάναμε πολλές φορές εκμάθηση έτσι ώστε να μην “ξεγελαστούμε” από την περίπτωση που η διαδικασία εκμάθησης βρει κάποιο “local minimum”. Οι αλλαγές στις δομές έγιναν αυξάνοντας τα επίπεδα αλλά και τους νευρώνες σε κάθε επίπεδο. Το αρχικό μας νευρωνικό δίκτυο ήταν, δύο επιπέδων. Ένα κρυφό επίπεδο με είκοσι δύο νευρώνες (επειδή έχουμε είκοσι δύο εισόδους) και το επίπεδο εξόδου με ένα νευρώνα (επειδή κάνουμε πρόβλεψη ενός χαρακτηριστικού).

6.2 Δημιουργία & Μελέτη αποτελεσμάτων νευρωνικών δικτύων

Όπως και στην πρόβλεψη τιμών βενζολίου, κάναμε ομαλοποίηση (normalize) των δεδομένων μας, πριν τα εισάγουμε στο νευρωνικό δίκτυο. Η ομαλοποίηση έγινε με την συνάρτηση μεταφοράς “Hyperbolic Tangent Sigmoid” (βλέπε Κεφ. 4.2). Στο κρυφό επίπεδο χρησιμοποιήσαμε την ίδια συνάρτηση μεταφοράς αλλά στο επίπεδο εξόδου χρησιμοποιήσαμε την γραμμική συνάρτηση μεταφοράς. Για την φάση της εκμάθησης (training) χρησιμοποιήσαμε τον αλγόριθμο Levenberg-Marquardt, όπου χρησιμοποιήσαμε το 50% των δεδομένων. Για την επικύρωση (validation) χρησιμοποιήσαμε το 25% και το υπόλοιπο 25% για την φάση του ελέγχου (testing). Τα δεδομένα τα εισήχθησαν σειριακά (βλέπε Κεφ. 4.5). Παρακάτω παραθέτουμε τον πίνακα με τα αποτελέσματα πρόγνωσης των δομών των τεχνητών νευρωνικών δικτύων που δημιουργήσαμε για τους δύο σταθμούς.

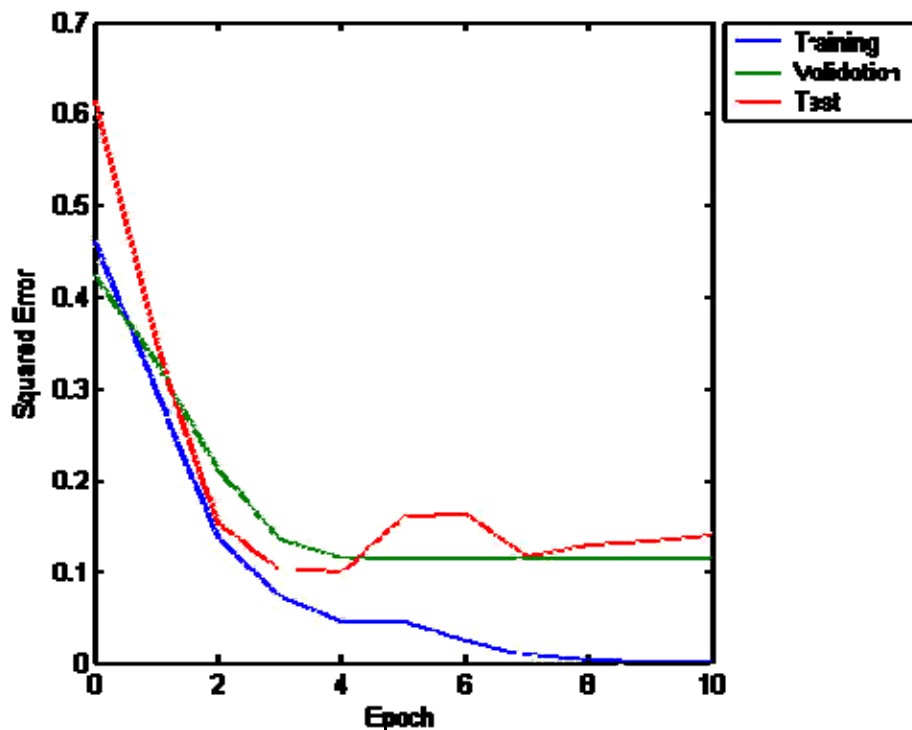
Σταθμός Πατησίων	Κρυφά Επίπεδα (\mathcal{F})			Επίπεδο Εξόδου (\mathcal{Z})			
	Κ.Ε. 1	Κ.Ε. 2	Κ.Ε. 3	n	R	R ²	S
Δομή 1	22	-	-	10	0,86	0,73	11,058
Δομή 2	22	10	-	19	0,85	0,73	10,999
Δομή 3	22	10	5	10	0,87	0,76	11,237
Δομή 4	22	22	-	12	0,85	0,72	10,15
Δομή 5	22	22	22	13	0,84	0,7	10,708

Πίνακας 16. Πίνακας αποτελεσμάτων απόδοσης πρόβλεψης χρησιμοποιώντας διάφορες δομές νευρωνικών δικτύων (σταθμός Πατησίων). Το S είναι το τυπικό σφάλμα εκτίμησης.

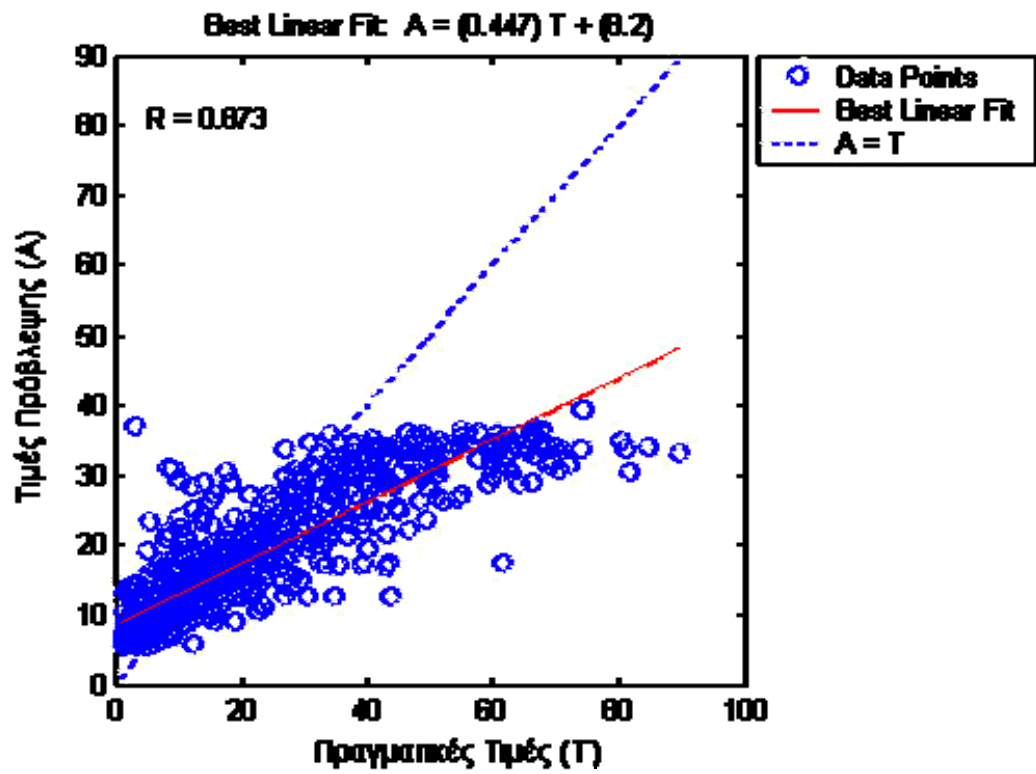
Σταθμός Λιοσίων	Κρυφά Επίπεδα (\mathcal{F})			Επίπεδο Εξόδου (\mathcal{S})			
	Κ.Ε. 1	Κ.Ε. 2	Κ.Ε. 3	n	R	R ²	S
Δομή 1	22	-	-	24	0,86	0,73	22,865
Δομή 2	22	10	-	13	0,88	0,77	21,752
Δομή 3	22	10	5	11	0,86	0,74	23,556
Δομή 4	22	22	-	11	0,85	0,72	24,059
Δομή 5	22	22	22	12	0,88	0,77	21,497

Πίνακας 17. Πίνακας αποτελεσμάτων απόδοσης πρόβλεψης χρησιμοποιώντας διάφορες δομές νευρωνικών δικτύων (σταθμός Λιοσίων). Το S είναι το τυπικό σφάλμα εκτίμησης.

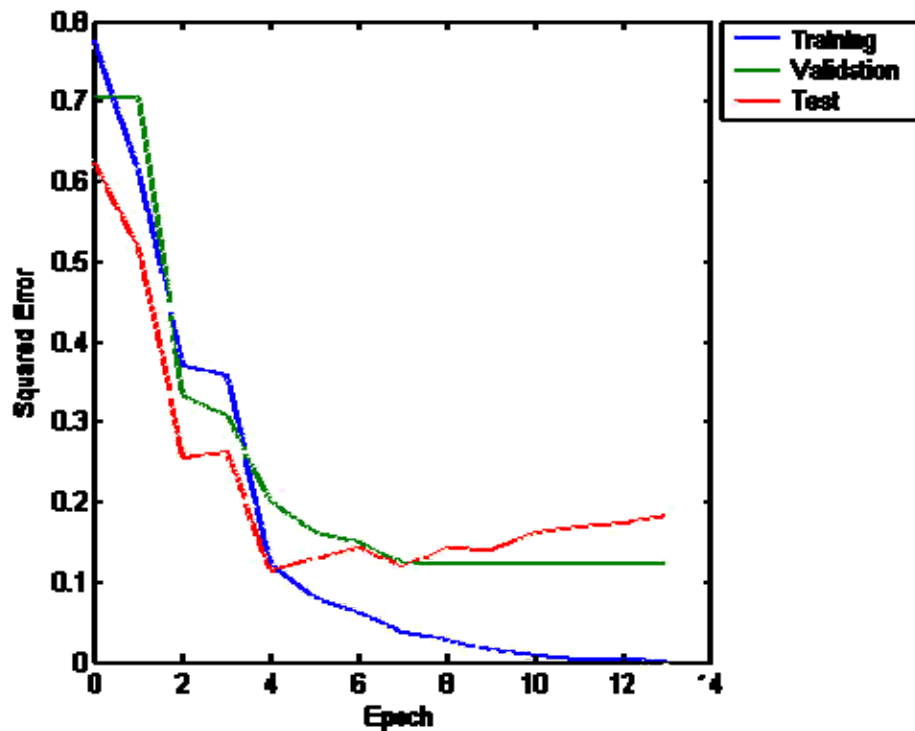
Παρατηρώντας τους δύο παραπάνω πίνακες βλέπουμε ότι και στους δύο σταθμούς είχαμε παρόμοια αποτελέσματα αυξάνοντας το πλήθος των κρυφών επιπέδων. Στον σταθμό Πατησίων η Δομή 3 είχε τα καλύτερα αποτελέσματα για αυτό το λόγω είναι και η δομή επιλογής μας. Στον σταθμό Λιοσίων η Δομή 2 και 5 είχε τα καλύτερα αποτελέσματα, αλλά για λόγους απλότητας διαλέξαμε την Δομή 2 η οποία έχει λιγότερα κρυφά επίπεδα και νευρώνες. Παρακάτω παραθέτουμε τις γραφικές παραστάσεις του σφάλματος εκμάθησης, και της σχέσης των προβλεπόμενων με των πραγματικών τιμών.



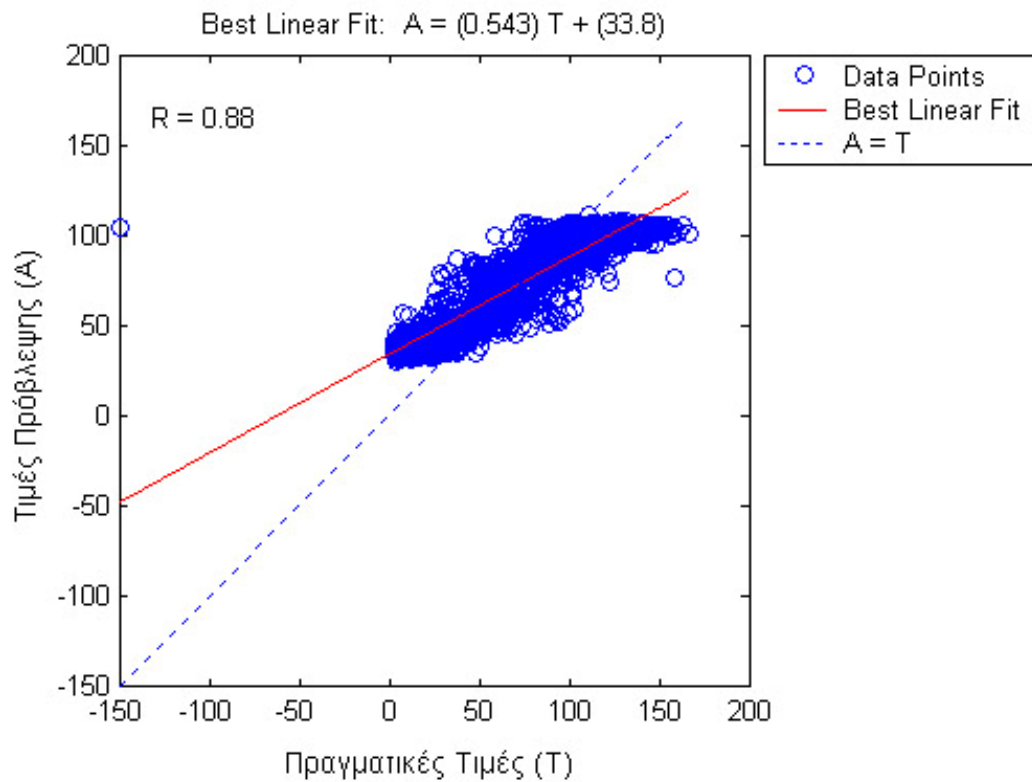
Σχήμα 61. Γραφική αναπαράσταση του σφάλματος εκμάθησης σταθμού Πατησίων κατά την πρόβλεψη ΜΜΤΚΟ όζοντος.



Σχήμα 62. Σχέση πραγματικών τιμών και προβλεπόμενων τιμών, σταθμού Πατησίων κατά την πρόβλεψη ΜΜΤΚΟ όζοντος.



Σχήμα 63. Γραφική αναπαράσταση του σφάλματος εκμάθησης σταθμού Λιοσίων κατά την πρόβλεψη ΜΜΤΚΟ όζοντος.



Σχήμα 64. Σχέση πραγματικών τιμών και προβλεπόμενων τιμών, σταθμού Λιοσίων κατά την πρόβλεψη MMTKO όζοντος.

Στην συνέχεια δημιουργήσαμε τον παρακάτω πίνακα στατιστικών στοιχείων, για να αξιολογήσουμε και να συγκρίνουμε τα νευρωνικά δίκτυα επιλογής μας. Στατιστικά στοιχεία τα οποία συνιστώνται από Wilmott (1982, 1985) [17].

Οριακή τιμή όζοντος 120 $\mu\text{g}/\text{m}^3$	Σταθμός Πατησίων (Δομή 3)	Σταθμός Λιοσίων (Δομή 2)
Observed mean	19,101	69,650
Predicted mean	16,739	71,629
Observed standard deviation	18,098	39,859
Predicted standard deviation	9,263	24,575
Normalised mean difference (NMD)	0,124	0,028
Root mean square error (RMSE)	11,224	21,734
RMSE systematic	0,382	0,624
RMSE unsystematic	0,395	1,448
Correlation coefficient	0,873	0,88
Index of agreement (IA)	0,825	0,882
Mean absolute error (MAE)	7,299	16,491
Mean percentage error (MPE)	0,781	0,646
Mean bias error (MBE)	2,362	-1,979
Number of observed alarms	0	121
Number of predicted alarms	0	0
Number of hours with correct alarm (A)	0	0
Correct Alarm Rate(%)	0	0
Number of hours with false alarm (C)	0	0
False Alarm Rate(%)	0	0
B	0	121
Critical success index (CSI)	0	0

Πίνακας 18. Πίνακας στατιστικών στοιχείων, για την μελέτη της απόδοσης των τεχνητών νευρωνικών δικτύων πρόβλεψης όζοντος (σταθμοί Πατησίων και Λιοσίων).

Στον παραπάνω πίνακα η τιμή ορίου επιλέγει σύμφωνα με την οδηγία 2002/2/ΕΚ [18]. Πληροφορίες για τα στατιστικά στοιχεία υπάρχουν και σε προηγούμενο κεφάλαιο (Κεφ. 4,5). Θα πρέπει όμως να σημειωθεί ότι δεν υπήρχαν “alarms”, έτσι δεν μπορούμε να αξιολογήσουμε το νευρωνικό δίκτυο ως προς την δυνατότητα πρόβλεψης υπέρβασης ορίου. Παρατηρούμε ότι στο σταθμό Λιοσίων υπήρξαν 121 υπερβάσεις ορίου σε διάρκεια τεσσάρων ετών, για τις οποίες το νευρωνικό μας δίκτυο δεν κατάφερε να προβλέψει κάποια από αυτές.

7. Συμπεράσματα

Οι βασικοί στόχοι αυτής της μελέτης ήταν η μελέτη χρονοσειρών και η δημιουργία μοντέλων πρόβλεψης ωριαίων τιμών βενζολίου και μέγιστων μέσων τιμών κυλιόμενου οκταώρου τιμών όζοντος.

Κατά την μελέτη χρονοσειρών, χρησιμοποιήσαμε μεθόδους και συναρτήσεις όπως (τυπική απόκλιση, συντελεστή συσχέτισης, συμμεταβλητότητα), οι οποίες μας βοήθησαν να κατανοήσουμε την φύση των δεδομένων μας.

Κατά την μελέτη του συντελεστή συσχέτισης, παρατηρήσαμε τις σχέσεις που έχουν τα δεδομένα μας με τις εξαρτημένες μεταβλητές (βενζολίου και όζοντος). Για την μεταβλητή βενζολίου, παρατηρήσαμε “πολύ υψηλή συσχέτιση” με την χρονοσειρά μονοξειδίου του άνθρακα (CO). Πράγμα που καταδεικνύει, την ομοιότητα που υπάρχει μεταξύ αυτών σε σχέση με τις εκπομπές τους. Έτσι, αναμένουμε ότι η κυκλοφορία (κύρια πηγή CO) είναι επίσης και η κύρια πηγή βενζολίου, για ένα σημείο με υψηλή πυκνότητα κυκλοφορίας όπως ο σταθμός Πατησίων. Εφόσον μάλιστα ληφθούν υπόψη και τα αποτελέσματα της εκτίμησης εκπομπών των Kassomenos et.al 2006, τότε συνάγεται ότι τα στοιχεία traffic time και traffic speed (επηρεάζουν CO) και τα στοιχεία (motorcycles, μη καταλυτικά, traffic speed) οδηγούν στο συμπέρασμα πως η ταχύτητα κίνησης ενδέχεται να είναι ο κρίσιμος μηχανισμός που επηρεάζει τις εκπομπές βενζολίου και CO στο κέντρο της πόλης. Επομένως, μια πολιτική κυκλοφοριακών ρυθμίσεων που θα αποσκοπούσε στην μεταβολή της ταχύτητας κίνησης, αναμένεται ότι θα ανακούφιζε, τοπικά την ατμοσφαιρική φόρτιση. Για την μεταβλητή όζοντος παρατηρούμε ότι στα Πατήσια υπάρχει υψηλή συσχέτιση της MMTKO του όζοντος με το μονοξείδιο του αζώτου (NO). Αυτή η σχέση μπορεί να υπάρξει κατά την διάρκεια έντονης ηλιακής ακτινοβολίας (αυξημένη θερμοκρασία) [Παράρτημα], γιατί υπό αυτές τις συνθήκες μπορεί να σχηματιστεί το όζον από οξείδια του αζώτου. Αυτό συνέβη στα Πατήσια και όχι στα Λιόσια γιατί όπως φαίνεται και στο σχήμα 90 κατά την διάρκεια των μελετών μας (02,03,05) στα Πατήσια παρατηρήθηκαν μεγαλύτερες θερμοκρασίες απ’ ότι στα Λιόσια. Συγκεκριμένα παρατηρήθηκε μέση θερμοκρασία υψηλότερη κατά 1,65 βαθμούς C°. Επίσης παρατηρούμε υψηλή συσχέτιση των τιμών σχετικής υγρασίας (RH) με την MMTKO του όζοντος. Συγκεκριμένα οι διακυμάνσεις της σχετικής υγρασίας επηρεάζουν αντιστρόφως ανάλογα τις τιμές του όζοντος.

Χρησιμοποιώντας την μέθοδο Principal Component Analysis (P.C.A) κατά την πρόβλεψη βενζολίου καταφέραμε να αναγνωρίσουμε κάποια πρότυπα μέσα στα δεδομένα μας, και να τα εκφράσουμε με τέτοιο τρόπο, έτσι ώστε να τονίσουμε τις ομοιότητες και τις διαφορές τους. Έτσι απορρίψαμε από τα δεδομένα μας το χαρακτηριστικό της ταχύτητας του ανέμου (WS) επειδή όπως φαίνεται και στον πίνακα συνδιακύμανσης (πίνακας 10) δεν είχε σχέση με τις μεταβολές της χρονοσειράς βενζολίου.

Η μελέτη του περιοδογράμματος της χρονοσειράς βενζολίου ενίσχυσε τον αρχικό, συλλογισμό μας ότι η κύρια πηγή βενζολίου είναι η κυκλοφορία. Αυτό το συμπεραίνουμε γιατί υπάρχει περιοδικότητα της κυκλοφορίας ανά εβδομάδα (περισσότερη κίνηση τις καθημερινές, και λιγότερη το σαββατοκύριακο). Επίσης η περιοδικότητα το οκταώρου είναι πολύ σημαντική καθώς υποδηλώνει την αρχή και το τέλος των εργασιών, με αποτέλεσμα επιπρόσθετη κυκλοφορία για αυτές τις περιόδους. Με το ίδιο σκεπτικό, και στη περιοδικότητα ανά ημέρα (24 ωρών), υπάρχει περισσότερη κίνηση κατά την διάρκεια της ημέρας και λιγότερη τις απογευματινές με βραδινές ώρες.

Χρησιμοποιώντας τεχνητά νευρωνικά δίκτυα forward connection όπου για την εκμάθηση τους χρησιμοποιήσαμε την μέθοδο Back Propagation είχαμε αρκετά καλά αποτελέσματα πρόβλεψης (με δείκτη προσδιορισμού 0,63 για πρόβλεψη βενζολίου και 0,76 για πρόβλεψη όζοντος). Επίσης ήταν αρκετά αξιόπιστα στην πρόβλεψη υπερβάσεων ορίων (με ευστοχία 81%-85,3% κατά την πρόβλεψη βενζολίου). Εκτός από το νευρωνικό δίκτυο πρόβλεψης τιμών όζοντος (για τον σταθμό Λιοσίων) όπου δεν καταφέραμε να προβλέψουμε κάποια υπέρβαση ορίων από τις 121 που πραγματοποιήθηκαν κατά τα έτη 02-05.

ΠΑΡΑΡΤΗΜΑ Α΄

ΚΥΡΙΟΙ ΑΤΜΟΣΦΑΙΡΙΚΟΙ ΡΥΠΟΙ

Ατμοσφαιρική ρύπανση:

Ατμοσφαιρική ρύπανση καλείται, η παρουσία στην ατμόσφαιρα κάθε είδους ουσιών σε συγκέντρωση ή διάρκεια που μπορούν να προκαλέσουν αρνητικές επιπτώσεις στην υγεία, στους ζωντανούς οργανισμούς και στα οικοσυστήματα και γενικά να καταστήσουν το περιβάλλον ακατάλληλο για τις επιθυμητές χρήσεις του [18]. Κάτω από ορισμένες συνθήκες, η ατμοσφαιρική ρύπανση μπορεί να φτάσει σε επίπεδα που μπορεί να δημιουργήσουν ανεπιθύμητες συνθήκες διαβίωσης. Σε αυτήν την περίπτωση έχει επικρατήσει να λέγεται ότι έχουμε “Νέφος”. Το “Νέφος” παρουσιάζεται σε δύο μορφές:

- **Νέφος καπνομίχλης**, σχηματίζεται όταν μετρώνται υψηλές συγκεντρώσεις ρύπων, όπως μονοξειδίου του άνθρακα, διοξειδίου του θείου και αιωρούμενων σωματιδίων, σε συνδυασμό με σχετικά χαμηλή θερμοκρασία και μεγάλη σχετική υγρασία.
- **Φωτοχημικό νέφος**, παρουσιάζεται όταν παρατηρούνται υψηλές θερμοκρασίες, μεγάλη ηλιοφάνεια σε ένταση και διάρκεια, μικρή σχετική υγρασία και υψηλή συγκέντρωση οξειδίων του αζώτου, υδρογονανθράκων, και δευτερογενών προϊόντων τους.

Για να αντιμετωπίσουμε αποτελεσματικά το πρόβλημα του νέφους πρέπει να γνωρίσουμε, πώς δημιουργείται, από τι αποτελείται, τι επιδράσεις δημιουργεί στο περιβάλλον, και τι μπορούν να κάνουν πολιτεία και κοινωνία για την καταπολέμηση του.

Βασικότεροι ατμοσφαιρικοί ρύποι:

Όζον (O₃): Αέριο, άχρωμο, με χαρακτηριστική οσμή, το κύριο συστατικό του φωτοχημικού νέφους στην επιφάνεια της γης (τροπόσφαιρα). Στην ατμόσφαιρα (στρατόσφαιρα), ωστόσο το όζον έχει ευεργετικό ρόλο απορροφώντας την επιβλαβή υπεριώδη ακτινοβολία του ήλιου.

- **Πηγές στο περιβάλλον:** Το όζον σχηματίζεται στην κατώτερη ατμόσφαιρα ως αποτέλεσμα αλυσίδας χημικών αντιδράσεων μεταξύ του οξυγόνου, πτητικών οργανικών ενώσεων (VOCs), και οξειδίων του αζώτου υπό συνθήκες έντονης ηλιακής ακτινοβολίας και υψηλών θερμοκρασιών. Πηγές των ρύπων που συντελούν στη δημιουργία του όζοντος είναι τα οχήματα, εργοστάσια, χωματερές, χημικά διαλυτικά και πολλές άλλες μικρές πηγές όπως βενζινάδικα, αγροτικός εξοπλισμός, κλπ.
- **Επιδράσεις:** Το όζον σε μεγάλες συγκεντρώσεις προκαλεί σημαντικά προβλήματα στην ανθρώπινη υγεία και το περιβάλλον όπου ζούμε. Προκαλεί ερεθισμό στην αναπνευστική οδό, διαταραχή της αναπνευστικής λειτουργίας, αίσθημα ξηρότητας στο λαιμό, πόνο στο στήθος, βήχα, άσθμα, φλεγμονή στους πνεύμονες, πιθανή επιδεικτικότητα σε μολύνσεις του αναπνευστικού και ερεθισμό των οφθαλμών. Το όζον έχει επίσης δυσμενείς επιδράσεις στα φυτά, μειώνει την παραγωγή στις αγροτικές καλλιέργειες και προκαλεί ζημιές στην δασική βλάστηση.

Μονοξείδιο του άνθρακα (CO): Αέριο, άοσμο και άχρωμο, εκπέμπεται από τις εξατμίσεις των μηχανών των βενζινοκίνητων αυτοκινήτων και πάσης φύσεως μηχανών όταν συντελείτε ατελής καύση της καύσιμης ύλης.

- **Πηγές στο περιβάλλον:** Κυρίως τα βενζινοκίνητα αυτοκίνητα. Υψηλές συγκεντρώσεις του μπορούν να βρεθούν σε κλειστά μέρη όπως χώροι στάθμευσης, ελλιπώς αεριζόμενες υπόγειες διαβάσεις, ή κατά μήκος των δρόμων σε περιόδους κυκλοφοριακής αιχμής.
- **Επιδράσεις:** Μειώνει την ικανότητα του αίματος να μεταφέρει οξυγόνο σε βασικούς ιστούς του οργανισμού, επιδρώντας κυρίως στο καρδιαγγειακό και νευρικό σύστημα. Χαμηλές συγκεντρώσεις του επηρεάζουν δυσμενώς άτομα με καρδιακά προβλήματα και μειώνουν τις σωματικές επιδόσεις νεαρών και υγιών ατόμων. Υψηλότερες συγκεντρώσεις προκαλούν συμπτώματα όπως ζαλάδα, πονοκεφάλους και κόπωση.

Διοξείδιο του αζώτου (NO₂): Είναι αέριο με καφέ-κίτρινο χρώμα και ιδιάζουσα οσμή. Σε υψηλές συγκεντρώσεις δίνει το χαρακτηριστικό χρώμα του στην όψη του ουρανού στις αστικές περιοχές.

- **Πηγές στο περιβάλλον:** Η χρήση καυσίμων κυρίως σε αυτοκίνητα αλλά και σε βιομηχανικούς καυστήρες ή σε σταθμούς ηλεκτροπαραγωγής παράγει μονοξείδιο του αζώτου. Αυτό με διάφορες χημικές αντιδράσεις που ενισχύονται με την παρουσία της ηλιακής ακτινοβολίας μετατρέπεται σε διοξείδιο του αζώτου.
- **Επιδράσεις:** Σημαντικός ρύπος για τη δημιουργία όξινης βροχής. Σε υψηλές συγκεντρώσεις βλάπτει ανθρώπους και βλάστηση. Στα παιδιά μπορεί να προκαλέσει αναπνευστικές ασθένειες. Στους ασθματικούς προκαλεί δυσκολία στην αναπνοή.

Αιωρούμενα Σωματίδια: Σωματίδια σε στερεή ή υγρή φάση που αιωρούνται στην ατμόσφαιρα για μεγαλύτερα ή μικρότερα χρονικά διαστήματα ανάλογα με το μέγεθος και τις άλλες φυσικοχημικές τους ιδιότητες.

- **Πηγές στο περιβάλλον:**
 - Φυσικές πηγές: ηφαιστειακή δραστηριότητα, θάλασσα, σκόνη από απογυμνωμένο έδαφος, μεταφορά από μεγάλες αποστάσεις κυρίως από περιοχές ερήμων.
 - Ανθρωπογενείς πηγές: βιομηχανικές δραστηριότητες, παραγωγή τσιμέντου, γύψου, χυτήρια μεταλλεύματος, εξορυκτικές δραστηριότητες, κατασκευαστικές/οικοδομικές δραστηριότητες, οχήματα (κυρίως πετρελαιοκίνητα οχήματα και δίκυκλα), πυρκαγιές, αγροτικές δραστηριότητες, άλλες καύσεις (καλοριφέρ, τζάκια, ψησταριές), και ιδιαίτερα στο εσωτερικό περιβάλλον κάπνισμα και μαγείρεμα. Η συμμετοχή του αυτοκινήτου οφείλεται στην καύση του καυσίμου, στη φθορά των ελαστικών και στην επαναιώρηση. Μικρότερα σε μέγεθος σωματίδια δημιουργούνται δευτερογενώς στην ατμόσφαιρα από αντιδράσεις αερίων ρύπων. Οι αντιδράσεις αυτές επιταχύνονται παρουσία ηλιακής ακτινοβολίας και σε υψηλές θερμοκρασίες.
- **Επιδράσεις:**
 - Οι επιδράσεις στην υγεία εξαρτώνται πολύ από το μέγεθος των σωματιδίων και τη σύστασή τους. Όσο μικρότερα σε μέγεθος είναι τα σωματίδια τόσο βαθύτερα εισχωρούν στο αναπνευστικό σύστημα του ανθρώπου. Γενικά σωματίδια με μέγεθος μεγαλύτερο από 10μm δεν

εισχωρούν στο κατώτερο αναπνευστικό σύστημα. Τα μικρότερα από 10 μm σωματίδια μπορούν να προκαλέσουν σοβαρές ασθένειες στο αναπνευστικό ή να επιδεινώσουν τα συμπτώματα ευαίσθητων ομάδων του πληθυσμού. Ομάδες υψηλού κινδύνου αποτελούν οι ηλικιωμένοι, τα παιδιά και τα άτομα που πάσχουν από άσθμα και άλλες καρδιαγγειακές παθήσεις.

- Τα αιωρούμενα σωματίδια προκαλούν φθορές στα υλικά και μειώνουν την ορατότητα, ιδιαίτερα σε συνθήκες αυξημένης υγρασίας.
- Τα αιωρούμενα σωματίδια επηρεάζουν τις οπτικές ιδιότητες της ατμόσφαιρας μεταβάλλοντας το ισοζύγιο ακτινοβολίας στην ατμόσφαιρα. Επίσης συνεισφέρουν στη δημιουργία νεφών ως πυρήνες συμπύκνωσης, αλλά και στην αλλαγή των ιδιοτήτων των νεφών και τη συχνότητα των βροχοπτώσεων.

Διοξείδιο του θείου (SO₂): Άχρωμο, αέριο, άοσμο σε χαμηλές συγκεντρώσεις αλλά με έντονη ερεθιστική οσμή σε πολύ υψηλές συγκεντρώσεις.

- **Πηγές στο περιβάλλον:** Εργοστάσια παραγωγής ενέργειας, βιομηχανίες, κεντρικές θερμάνσεις, διυλιστήρια πετρελαίου, χημικές βιομηχανίες, χαρτοβιομηχανίες.
- **Επιδράσεις:** Στην ατμόσφαιρα σχηματίζει θειικές ενώσεις μεταξύ των οποίων και θειικό οξύ και σωματίδια αποτελούμενα από θειικές ενώσεις. Επηρεάζει άτομα με αναπνευστικά προβλήματα από μόνο του ή σε συνέργεια με τα αιωρούμενα σωματίδια. Προκαλεί αλλοιώσεις σε βλάστηση και υλικά. Μειώνει την ορατότητα και αυξάνει την οξύτητα λιμνών και ποταμών.

Μόλυβδος, Αρσενικό, Κάδμιο και Νικέλιο (Pb, As, Cd, Ni): Είναι μέταλλα τα οποία βρίσκονται στην ατμόσφαιρα σε αιωρούμενα σωματίδια είτε υπό στοιχειακή μορφή είτε υπό μορφή ενώσεων (οξειδίων, θεικών ή θειούχων).

- **Πηγές στο περιβάλλον:**
 - **Φυσικές πηγές:** Ο μόλυβδος, βρίσκεται στο έδαφος ως αποτέλεσμα της απόσπασης βράχων, της ηφαιστειακής δραστηριότητας, τις πυρκαγιές δασών κ.α. Το αρσενικό βρίσκεται σε αφθονία στις ορεινές περιοχές της Ευρώπης με τη μορφή θειούχων ενώσεων. Άλλες φυσικές πηγές αρσενικού είναι η ηφαιστειακή δραστηριότητα, από την οποία εκπέμπεται με μορφή θειούχων αλάτων ή οξειδίων. Το κάδμιο βρίσκεται στη φύση σε μικρές ποσότητες κυρίως σε ορυκτά που περιέχουν θειούχες ενώσεις του ψευδάργυρου, μόλυβδου και χαλκού. Επίσης προέρχεται από τη βλάστηση, τις πυρκαγιές δασών και τα ηφαίστεια. Το νικέλιο, βρίσκεται σε μεγάλη αφθονία στους μετεωρίτες, στον γήινο πυρήνα και σε λιγότερη έκταση στην επιφάνεια της γης. Κυρίως βρίσκεται σε μορφή θειούχων αλάτων ή οξειδίων.
 - **Ανθρωπογενείς πηγές:** Ο μόλυβδος, εκπέμπεται κυρίως από τις διεργασίες παραγωγής του, από την απόρριψη στο περιβάλλον προϊόντων που περιέχουν μόλυβδο και από την καύση υγρών καυσίμων και ξύλων. Το αρσενικό εκπέμπεται κυρίως υπό μορφή οξειδίων, από χυτήρια αρσενικού και από την καύση καυσίμων. Παλαιότερα η χρήση ζιζανιοκτόνων ήταν ακόμη μια πηγή ρύπανσης. Το κάδμιο, εκπέμπεται από τις παραγωγικές διαδικασίες παραγωγής μόλυβδου, ψευδαργύρου, χαλκού, σιδήρου ή χάλυβα με τη μορφή

θειούχων ή θειικών αλάτων. Επίσης από την καύση καυσίμων υπό τη μορφή οξειδίων ή υπό στοιχειακή μορφή και από την καύση απορριμμάτων υπό τη μορφή χλωριούχων αλάτων. Το νικέλιο, εκπέμπεται από την καύση καυσίμων, από μεταλλουργικές εργασίες παραγωγής νικελίου ή χάλυβα. Το νικέλιο από τις διεργασίες αυτές εκπέμπεται ως θειικό άλας ή υπό τη μορφή οξειδίων. Χρησιμοποιείται ευρέως στη βιομηχανία ως καταλυτής.

- **Επιδράσεις:** Ο μόλυβδος προκαλεί αναιμία, αναπτυξιακές ανωμαλίες σε έμβρυα, βρέφη και παιδιά και βλάβες στο νευρικό σύστημα. Το αρσενικό επιδρά κυρίως στο ανώτερο αναπνευστικό και στο καρδιαγγειακό σύστημα και προκαλεί επίσης αύξηση της αρτηριακής πίεσεως. Είναι επίσης πιθανόν να προκαλεί καρκίνο στους πνεύμονες. Το κάδμιο επιδρά κυρίως στα νεφρά και στο αναπαραγωγικό σύστημα. Επίσης έχει χαρακτηριστεί ως καρκινογόνο. Το νικέλιο δεν θεωρείται καρκινογόνο. Πιθανόν να προκαλεί δερματικές παθήσεις.

Πρέπει να τονιστεί ότι τα μέταλλα αυτά επιδρούν στην υγεία κυρίως μέσω τροφικής αλυσίδας εάν έχει μολυνθεί και λιγότερο με την εισπνοή.

Πολυκυκλικοί Αρωματικοί Υδρογονάνθρακες (ΠΑΥ): Είναι οργανικές χημικές ενώσεις που περιέχουν άνθρακα και υδρογόνο. Αποτελούνται από τρεις ή περισσότερους συμπυκνωμένους βενζολικούς δακτυλίους και βρίσκονται κυρίως υπό μορφή ατμών ή σωματιδίων. Η χαρακτηριστικότερη ένωση της κατηγορίας αυτής είναι το βενζο(α)πυρένιο.

- **Πηγές στο περιβάλλον:**
 - Στις φυσικές πηγές περιλαμβάνονται πυρκαγιές και η ηφαιστειακή δραστηριότητα.
 - Στις ανθρωπογενείς πηγές περιλαμβάνονται η βιομηχανία (παραγωγής κωκ, αλουμινίου και επεξεργασίας ξύλου), η θέρμανση στις οικίες όταν χρησιμοποιούνται ξύλα και κάρβουνο, και τα οχήματα.
- **Επιδράσεις:** Ορισμένοι από τους Πολυκυκλικούς Αρωματικούς Υδρογονάνθρακες και κυρίως το βενζο(α)πυρένιο έχουν χαρακτηριστεί ως καρκινογόνες ενώσεις.

Βενζόλιο (C₆H₆): Χημική ένωση σε υγρή μορφή που αποτελείται από άνθρακα και υδρογόνο με χαρακτηριστική οσμή. Στην ατμόσφαιρα βρίσκεται σε μορφή ατμών επειδή το σημείο ζέσεως του είναι χαμηλό.

- **Πηγές στο περιβάλλον:** Το βενζόλιο εκπέμπεται στην ατμόσφαιρα κυρίως από ανθρωπογενείς δραστηριότητες. Η κύρια πηγή είναι τα βενζινοκίνητα οχήματα, ιδιαίτερα εκείνα στα οποία δεν λειτουργεί σωστά ο καταλύτης, ενώ άλλες πηγές είναι η βιομηχανία (δυλιστήρια, χημική βιομηχανία), η διακίνηση καυσίμων και η οικιακή θέρμανση.
- **Επιδράσεις:** Το βενζόλιο προκαλεί ασθένειες του αίματος και έχει χαρακτηριστεί ως καρκινογόνος ένωση.

ΠΑΡΑΡΤΗΜΑ Β΄

**ΥΠΟΒΛΗΘΗΚΕ ΠΡΟΣ ΔΗΜΟΣΙΕΥΣΗ ΣΤΟ ΔΙΕΘΝΕΣ
ΣΥΝΕΔΡΙΟ ΕΡΓΑΛΕΙΩΝ ΤΕΧΝΙΤΗΣ ΝΟΗΜΟΣΥΝΗΣ
(ICTAI 2007)**

Βιβλιογραφία:

1. Πέτρος Λ. Ρούσσοσ – Γιάννης Τσαούσης, Στατιστική εφαρμοσμένη στις κοινωνικές επιστήμες. Β' Αναθεωρημένη Έκδοση. Παρ. 5,5 & 5,6 (σελ.: 106-110), Παρ 10.3 (σελ: 213-215), Παρ 10.4 (σελ.: 218).
2. Wikipedia: The Free Encyclopedia:
http://en.wikipedia.org/wiki/Correlation_coefficient
3. Τζιραλής Γεώργιος: Doctoral Researcher, Dipl. Mechanical Engineer. Διαφάνειες διαλέξεων πρόγνωσης, Εθνικό Μετσόβιο Πολυτεχνείο ΜΜ ΒΔΕΕ 2006. (qtzi@central.ntua.gr)
4. Wikipedia: The Free Encyclopedia:
<http://en.wikipedia.org/wiki/Skewness>
5. Wikipedia: The Free Encyclopedia:
<http://en.wikipedia.org/wiki/Kurtosis>
6. Lindsay I Smith, A tutorial on Principal Components Analysis, (February 26, 2002)
7. G. David Garson, Professor of Public Administration Editor, Social Science Computer Review College of Humanities and Social Sciences.
<http://www2.chass.ncsu.edu/garson/pa765/factor.htm>
8. Αργυράκης Πάνος, Νευρωνικά Δίκτυα και Εφαρμογές (Τόμος Β'), Πάτρα 2001 (Κεφ. 1.1 & 4)
9. Abdul Habra, "Neural Networks – An Introduction"
<http://www.tek271.com/articles/neuralNet/IntoToNeuralNets.html>
10. Matlab Help. Keywords: tansig, logsig, purelin, Neural Network.
11. StatSoft, Inc. © Copyright, 1984-2003, "Neural Networks"
<http://www.statsoft.com/textbook/stneunet.html>
12. Herbert Taub – Donald L. Schilling, Αρχές Τηλεπικοινωνιακών Συστημάτων, 3^η βελτιωμένη Ελληνική Έκδοση. (Κεφ. 1.10 σελ. 35)
13. Επ. Καθ. Κωνσταντίνος Καρατζάς. Διαφάνειες παρουσίασης: "Στοιχεία Περιβαλλοντικών Χρονοσειρών"
14. Θεοδώρα Σλίνη, Κώστας Καρατζάς, Νικόλας Μουσιόπουλος, "Correlation of air pollution and meteorological data using neural networks", Int. J. Environment and Pollution, Vol. 20, Nos. 1–6, 2003
15. ΥΠΕΧΩΔΕ Γεν. Δ/ση Περιβάλλοντος Δ/νση ΕΑΡΘ. Η Ατμοσφαιρική ρύπανση στην αθήνα. Έκθεση 2005 (Τμήμα Ποιότητας Ατμόσφαιρας)
16. Wikipedia: The Free Encyclopedia:
http://en.wikipedia.org/wiki/Weighted_moving_average
17. Θεοδώρα Σλίνη, Κώστας Καρατζάς, Α. Παπαδόπουλος, "Regression Analysis And Urban Air Quality Forecasting: An Application For The City Of Athens". Copyright© 2002 Global Nest. (σελ.: 158-161).
18. ΥΠΕΧΩΔΕ Γεν. Δ/ση Περιβάλλοντος Δ/νση ΕΑΡΘ. Η Ατμοσφαιρική ρύπανση στην αθήνα. Έκθεση 2006 (Τμήμα Ποιότητας Ατμόσφαιρας). Μάρτιος 2007. (σελ.: 29-30, Παράρτημα Ι, Παράρτημα ΙV)