

**ΜΠΣ ΒΙΟΣΤΑΤΙΣΤΙΚΗ**



---

**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ**  
*ΙΑΤΡΙΚΗ ΣΧΟΛΗ*  
*ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ*

---

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ**  
*ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ*

---

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**  
**ΤΖΙΜΑ ΑΙΚΑΤΕΡΙΝΗΣ**

ΔΙΑΧΡΟΝΙΚΕΣ ΤΑΣΕΙΣ ΤΩΝ ΕΠΙΔΡΑΣΕΩΝ ΤΩΝ ΑΙΩΡΟΥΜΕΝΩΝ  
ΑΤΜΟΣΦΑΙΡΙΚΩΝ ΣΩΜΑΤΙΔΙΩΝ ΣΤΗΝ ΑΘΗΝΑ ΤΗΝ ΠΕΡΙΟΔΟ  
1992-2012

ΑΘΗΝΑ, 2017

## Ευχαριστίες

---

Για την εκπόνηση της παρούσας διπλωματικής εργασίας θα ήθελα να ευχαριστήσω από καρδιάς την κα Σαμόλη Ευαγγελία, όχι μόνο για την προθυμία και τη συνεχή καθοδήγησή της αλλά και για το ότι μου έδωσε την ευκαιρία να συνεργαστώ μαζί της και να σταθώ δίπλα σε έναν άνθρωπο με καλή διάθεση και αγάπη για τη διδασκαλία και τη διαπαιδαγώγηση.

Επίσης, θα ήθελα να ευχαριστήσω την κα Κατσουγιάννη Κλέα και τον κο Γρυπάρη Αλέξανδρο για τη συμμετοχή τους στην τριμελή επιτροπή εξέτασης της διπλωματικής εργασίας, για το χρόνο που αφιέρωσαν, αλλά και για τα πολύτιμα σχόλιά τους.

Τέλος, θα ήθελα να ευχαριστήσω τον κο Αναλυτή Αντώνη για την προθυμία του και το χρόνο που μου διέθεσε.

## ΠΕΡΙΕΧΟΜΕΝΑ

---

<b>1</b>	<b>Εισαγωγή</b> .....	<b>1</b>
<b>ΓΕΝΙΚΟ ΜΕΡΟΣ</b>		
<b>2</b>	<b>Ατμοσφαιρική Ρύπανση</b> .....	<b>2</b>
2.1	Ατμοσφαιρική Ρύπανση.....	2
2.2	Αιωρούμενα Σωματίδια .....	2
<b>3</b>	<b>Επιδράσεις των PM<sub>10</sub> στην Υγεία</b> .....	<b>5</b>
<b>4</b>	<b>Σκοπός της Μελέτης</b> .....	<b>7</b>
<b>5</b>	<b>Μεθοδολογική Προσέγγιση</b> .....	<b>7</b>
5.1	Γραμμική Παλινδρόμηση.....	7
5.2	Ανάλυση Χρονοσειρών .....	11
5.3	Παλινδρόμηση Poisson.....	11
5.4	Ημι-Πιθανοφάνεια.....	12
5.5	Αυτοσυσχέτιση .....	14
5.6	Έλεγχος Πιθανών Συγχυτικών Παραγόντων .....	15
5.7	Γενικευμένα Προσθετικά Μοντέλα .....	16
5.8	Εξομάλυνση.....	18
5.8.1	Παραμετρικές Συναρτήσεις Εξομάλυνσης.....	18
5.8.2	Πολυώνυμα Παρεμβολής με Περιορισμό .....	19
5.9	Κριτήρια Επιλογής Μοντέλου .....	21
5.9.1	Κριτήριο Μερικής Αυτοσυσχέτισης (PACF) .....	21
5.9.2	Κριτήριο Πληροφορίας Akaike (AIC).....	22
<b>ΕΙΔΙΚΟ ΜΕΡΟΣ</b>		
<b>6</b>	<b>Πρόβλεψη των Συγκεντρώσεων των PM<sub>10</sub> την περίοδο 1992-2000</b> .....	<b>23</b>
6.1	Μοντέλο Πολλαπλής Γραμμικής Εξάρτησης .....	23
6.2	Πρόβλεψη των Συγκεντρώσεων των PM <sub>10</sub> την περίοδο 1992-2000 ...	27
<b>7</b>	<b>Επιδημιολογική Ανάλυση για τη διερεύνηση των επιδράσεων των σωματιδίων στην ημερήσια θνησιμότητα</b> .....	<b>29</b>
7.1	Δεδομένα.....	29
7.2	Περιγραφική Ανάλυση .....	31
7.3	Στατιστική Ανάλυση .....	37
<b>8</b>	<b>Αποτελέσματα</b> .....	<b>40</b>
<b>9</b>	<b>Συζήτηση</b> .....	<b>46</b>
<b>10</b>	<b>Συμπεράσματα</b> .....	<b>48</b>
	<i>Περίληψη</i> .....	49

<i>Abstract</i> .....	51
<i>Βιβλιογραφία</i> .....	52

# 1 ΕΙΣΑΓΩΓΗ

---

Τις τελευταίες δεκαετίες έχει πραγματοποιηθεί μία σειρά επιδημιολογικών μελετών, τόσο στην Αμερική όσο και την Ευρώπη για τη διερεύνηση των επιπτώσεων των ατμοσφαιρικών ρύπων, με επίκεντρο ενδιαφέροντος τις συγκεντρώσεις των αιωρούμενων σωματιδίων, στην υγεία και την ποιότητα ζωής. Οι μελέτες αυτές έχουν τεκμηριώσει τις επιβλαβείς επιδράσεις τους στην ανθρώπινη υγεία και τα αιωρούμενα σωματίδια έχουν αναγνωριστεί ως ένας από τους πιο σημαντικούς ρύπους και χαρακτηρίζεται ως οι «αόρατοι εχθροί» της υγείας μας.

Η Αθήνα έχει πληθυσμό πάνω από 4.000.000 κατοίκους και περιέχει περισσότερα από 1.600.000 οχήματα, καθώς και το ήμισυ των βιομηχανικών και εμπορικών δραστηριοτήτων της Ελλάδας (Eurostat, 2004; Chaloulakou et al., 2003). Η μεγάλη κλίμακα των ρυπογόνων εκπομπών σε συνδυασμό με τη κλειστή τοπογραφία του λεκανοπεδίου της δυσχεραίνει την ατμοσφαιρική κυκλοφορία, τον αερισμό, τη διάχυση των ρύπων και έχει ως αποτέλεσμα τη ρύπανση του αέρα σε σημαντικό βαθμό.

Ο ρόλος της πολιτείας για την αντιμετώπιση του προβλήματος τόσο σε επίπεδο ελέγχων των ρυπογόνων πηγών (βιομηχανίες, αυτοκίνητα, καυστήρες) όσο και σε επίπεδο αντιμετώπισης με τη λήψη μέτρων ήταν και είναι ουσιάδης. Οι διάφορες πολιτικές που εφαρμόστηκαν και εφαρμόζονται για τη μείωση των επιπέδων της ατμοσφαιρικής ρύπανσης, αλλά και οι κοινωνικο-οικονομικές αλλαγές των τελευταίων χρόνων έχουν αλλάξει το χημικό προφίλ της ατμοσφαιρικής ρύπανσης ενώ τα επίπεδα των σωματιδίων έχουν μειωθεί (Karanasiou et al., 2014).

Στα πλαίσια της διπλωματικής εργασίας εφαρμόστηκαν γενικευμένα προσθετικά μοντέλα Poisson προσαρμοσμένα για την ύπαρξη υπερ-διασποράς, ώστε να διερευνηθεί η επίδραση των αιωρούμενων σωματιδίων διαμέτρου μικρότερης των 10 $\mu$ m (PM<sub>10</sub>) στη θνησιμότητα των κατοίκων της Αθήνας και η πιθανή διαχρονική τους τάση για την περίοδο 1992-2012. Μετρήσεις για τις συγκεντρώσεις των PM<sub>10</sub> δεν υπήρχαν για τη χρονική περίοδο πριν το 2001, οπότε εκτιμήθηκαν από ένα μοντέλο πολλαπλής γραμμικής παλινδρόμησης βασισμένο σε μετρήσεις του μαύρου καπνού.

Το μοντέλο πρόβλεψης κατασκευάστηκε από τη συλλογή δεδομένων της περιόδου 2001-2004, κατά την οποία υπήρχαν ταυτόχρονες μετρήσεις των δύο ρύπων, των PM<sub>10</sub> και του μαύρου καπνού, και εφαρμόστηκε στα έτη 1992-2000. Για το λόγο αυτό η επιδημιολογική ανάλυση πραγματοποιήθηκε για το σύνολο της χρονικής περιόδου 1992-2012, αλλά και χωριστά για την περίοδο με πραγματικές μετρήσεις 2001-2012 και την περίοδο με προβλέψεις των συγκεντρώσεων 1992-2000. Για τη διερεύνηση των χρονικών τάσεων χρησιμοποιήθηκε η μεθοδολογία που προτείνεται από τους Dominici et al (2007).

# ΓΕΝΙΚΟ ΜΕΡΟΣ

## 2 ΑΤΜΟΣΦΑΙΡΙΚΗ ΡΥΠΑΝΣΗ

---

### 2.1 ΑΤΜΟΣΦΑΙΡΙΚΗ ΡΥΠΑΝΣΗ

Ατμοσφαιρική ρύπανση ονομάζεται η παρουσία ρύπων στην ατμόσφαιρα, δηλαδή η συγκέντρωση κάθε είδους ουσιών, θορύβου, ακτινοβολίας ή άλλων μορφών ενέργειας σε ποσότητα ή διάρκεια, που μπορεί να προκαλέσουν αρνητικές επιπτώσεις στην υγεία των ζωντανών οργανισμών και γενικότερα να διαταράξουν την οικολογική ισορροπία σε μεγάλη ή μικρή γεωγραφική κλίμακα (Μελάς Δ., 2000).

Υπάρχουν διάφορες πηγές ατμοσφαιρικής ρύπανσης, ανθρωπογενούς και φυσικής προέλευσης. Παραδείγματα φυσικών πηγών είναι οι ηφαιστιογενείς εκρήξεις, οι πυρκαγιές των δασών, ο κονιορτός και η εκνέφωση θαλάσσιου άλατος. Η ανθρωπογενής ατμοσφαιρική ρύπανση προκαλείται κυρίως από τη βιομηχανία, τις μεταφορές και τα νοικοκυριά. Σε μία τυπική πόλη, η βιομηχανία ευθύνεται για το 50% της ατμοσφαιρικής ρύπανσης, τα μέσα μεταφοράς για το 35%, ενώ τα νοικοκυριά για το 15% (Βικιπαίδεια, 2016).

Η ατμοσφαιρική ρύπανση βλάπτει σοβαρά την υγεία του ανθρώπου και το περιβάλλον. Ο Παγκόσμιος Οργανισμός Υγείας (Π.Ο.Υ., World Health Organization, WHO) εξετάζει κατά περιόδους τα αποτελέσματα των επιστημονικών ερευνών ώστε να αναπροσαρμόζει τις οδηγίες του για την ατμοσφαιρική ρύπανση. Σκοπός των συστάσεων είναι η απόκτηση του βέλτιστου επιπέδου Υγείας μέσω της επίτευξης ενός ασφαλούς επιπέδου στις συγκεντρώσεις των ατμοσφαιρικών ρύπων. Στην Ευρώπη υπάρχει ο Ευρωπαϊκός Οργανισμός Περιβάλλοντος (ΕΠΟ, European Environment Agency (EEA)) που επίσης ελέγχει τα επίπεδα ποιότητας αέρα, ώστε να μην κινδυνεύει η υγεία του ανθρώπου και του περιβάλλοντος.

Σύμφωνα με τις μέχρι σήμερα έρευνες ο πιο επικίνδυνος ρύπος για τη δημόσια υγεία είναι τα αιωρούμενα σωματίδια καθώς επιβαρύνουν σε σημαντικό βαθμό την υγεία του ανθρώπου.

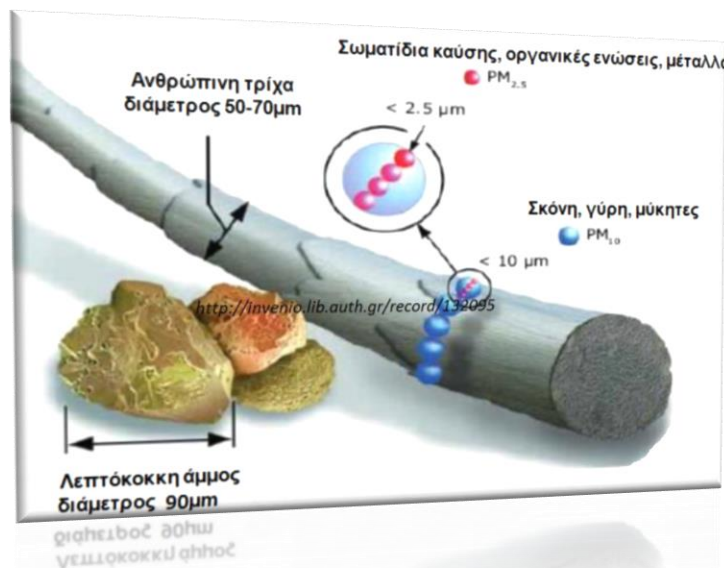
### 2.2 ΑΙΩΡΟΥΜΕΝΑ ΣΩΜΑΤΙΔΙΑ

Ως αιωρούμενα ατμοσφαιρικά σωματίδια (PM) χαρακτηρίζουμε ένα μείγμα στερεών και υγρών σωματιδίων, διαφορετικής προέλευσης, μεγέθους και σύστασης. Προέρχονται είτε από φυσικές πηγές είτε από ανθρωπογενείς δραστηριότητες. Λειτουργούν ως καταλυτικές επιφάνειες και ως μέσα μεταφοράς ενώσεων, ενδεχομένως τοξικών ή οξειδωτικών. Ανάλογα με τον τρόπο σχηματισμού τους, τα PM διακρίνονται σε πρωτογενή, όταν εκπέμπονται κατευθείαν στον ατμοσφαιρικό αέρα και δευτερογενή, όταν σχηματίζονται μέσω χημικών αντιδράσεων στην ατμόσφαιρα από αέρια (ατμοσφαιρικό οξυγόνο, υδρατμί), αντιδρώντα μόρια (Ο<sub>3</sub>),

ελεύθερες ρίζες), ρύπους (Διοξείδιο του Θείου (SO<sub>2</sub>), Διοξείδιο του Αζώτου (NO<sub>2</sub>)), καθώς και οργανικά αέρια (VOCs προερχόμενα από φυσικές ή ανθρωπογενείς πηγές).

Τα αιωρούμενα σωματίδια ταξινομούνται με βάση την αεροδυναμική τους διάμετρο. Με τον όρο αεροδυναμική ορίζεται η διάμετρος που πρέπει να έχει ένα σφαιρικό σώμα πυκνότητας 1g/cm<sup>3</sup>, ώστε κάτω από τις ίδιες συνθήκες να έχει την ίδια ταχύτητα καθίζησης με το υπό εξέταση σωματίδιο.

Αρχικά, οι μετρήσεις των αιωρούμενων σωματιδίων αναφερόταν στα ολικά αιωρούμενα σωματίδια (Total Suspended Particulates, TSP), δίχως να γίνεται διαφοροποίηση αυτών ανάλογα με το μέγεθός τους. Η προσέγγιση αυτή ωστόσο εξελίχθηκε με την ανάπτυξη της τεχνολογίας και την ανακάλυψη των διαφορετικών επιπτώσεων των σωματιδίων ανάλογα με τη διάμετρό τους.



Γράφημα 1 Διάμετρος των αιωρούμενων σωματιδίων PM<sub>10</sub> και PM<sub>2.5</sub>

Η μέτρηση των TSP αντικαταστάθηκε με αυτή των PM<sub>10</sub>, δηλαδή με τα αιωρούμενα σωματίδια διαμέτρου μικρότερης ή ίσης των 10 μm. Στη συνέχεια προτάθηκαν οι επιπλέον διαχωρισμοί μεταξύ των 2.5 με 10 μm (PM<sub>2.5-10</sub>), μικρότερης των 2.5 μm (PM<sub>2.5</sub>) και μικρότερης των 0.1 μm (PM<sub>0.1</sub>).

Κοινή ιδιότητα των αιωρούμενων σωματιδίων είναι η οπτική τους συμπεριφορά, δηλαδή η ελάττωση της ορατότητας, καθώς και της ηλιακής ακτινοβολίας που φτάνει στη γη εξαιτίας της παρουσίας των αιωρούμενων σωματιδίων στην ατμόσφαιρα. Τα αιωρούμενα σωματίδια ανάλογα με το μέγεθός τους και τη συγκέντρωσή τους στην ατμόσφαιρα, διαθλούν, περιθλούν, διαχέουν ή ακόμα αποκόπτουν το φως.

Τα αιωρούμενα σωματίδια διεισδύουν στις πνευμονικές κυψελίδες με βαθμό αντιστρόφως ανάλογο του μεγέθους τους. Ιδιαίτερο ενδιαφέρον παρουσιάζουν τα εισπνεόμενα PM, δηλαδή τα σωματίδια με αεροδυναμική διάμετρο < 10 μm λόγω της ικανότητάς τους να εισέρχονται και να αποτίθενται στο κατώτερο αναπνευστικό σύστημα. Στην παρούσα εργασία θα μελετήσουμε τις επιδράσεις των PM<sub>10</sub> στην ολική, καρδιαγγειακή και αναπνευστική θνησιμότητα.



*Γράφημα 2: Αιθαλομίχλη στην Αθήνα*

Τα αιωρούμενα σωματίδια παρακολουθούνται σε δίκτυα ποιότητας του ατμοσφαιρικού αέρα. Η καταγραφή των  $PM_{10}$  γίνεται από ένα δίκτυο που αποτελείται από οχτώ μετρητικές διατάξεις, οι οποίες έχουν τοποθετηθεί σε διάφορες περιοχές της Αθήνας. Η μέση ανώτατη ετήσια τιμή που έχει καθορίσει η Ευρωπαϊκή Ένωση για τις συγκεντρώσεις των  $PM_{10}$  είναι  $40 \mu\text{g}/\text{m}^3$  (μικρογραμμάρια ανά κυβικό μέτρο), ενώ η μέση ανώτατη ημερήσια τιμή είναι  $50 \mu\text{g}/\text{m}^3$  ενώ αυτή επιτρέπεται να ξεπερνιέται 35 φορές το χρόνο. Στο κέντρο της Αθήνας οι μέσες ημερήσιες τιμές ξεπερνιούνται πάνω από 150 φορές το χρόνο ([www.ypeka.gr](http://www.ypeka.gr)).



### 3 ΕΠΙΔΡΑΣΕΙΣ ΤΩΝ PM<sub>10</sub> ΣΤΗΝ ΥΓΕΙΑ

Η διερεύνηση των επιδράσεων των ρύπων στην υγεία ταξινομείται ανάλογα με τη βραχυχρόνια ή μακροχρόνια έκθεση σε αυτούς. Ενώ και τα δύο είδη εκθέσεων έχουν συσχετιστεί με επιβαρυντικές επιδράσεις στην ανθρώπινη υγεία, λόγω της εστίασης της διπλωματικής στη διερεύνηση των επιδράσεων μετά από βραχυχρόνια έκθεση, η βιβλιογραφική ανασκόπηση επικεντρώθηκε σε αυτό το είδος επιδράσεων.

Πολυάριθμες μελέτες έχουν διερευνήσει τις βραχυπρόθεσμες επιδράσεις των PM<sub>10</sub> στην ανθρώπινη υγεία με τελικό στόχο την αντιμετώπιση του προβλήματος της ατμοσφαιρικής ρύπανσης και τον καθορισμό ορίων και στοχοθετημένων τιμών για την ποιότητα του αέρα. Ένα γενικό πλαίσιο των αποτελεσμάτων δείχνει ότι η βραχυπρόθεσμη έκθεση σε PM<sub>10</sub> συνδέεται με την αύξηση της ολικής, καρδιαγγειακής και αναπνευστικής ημερήσιας θνησιμότητας, όσο και με την αύξηση των εισαγωγών στα νοσοκομεία από καρδιαγγειακές και αναπνευστικές αιτίες (WHO, 2013 (REVIHAAP)).

Το 1993 ξεκίνησε το πρόγραμμα APHEA (Air Pollution and Health: A European Approach), στο οποίο συμμετείχαν 22 ερευνητικές ομάδες από 20 ευρωπαϊκές χώρες με στόχο τη διερεύνηση των βραχυπρόθεσμων επιπτώσεων της ατμοσφαιρικής ρύπανσης στην υγεία (Katsouyanni et al., 1996). Έτσι, δημιουργήθηκε ένα ευρωπαϊκό δίκτυο επιστημόνων, το οποίο ανέλυσε ένα μεγάλο όγκο δεδομένων με ενιαίο και τυποποιημένο τρόπο, τα αποτελέσματα των οποίων χρησιμοποιήθηκαν για τη λήψη ρυθμιστικών μέτρων.

Τα αποτελέσματα αφορούσαν τις βραχυπρόθεσμες επιπτώσεις των ατμοσφαιρικών σωματιδίων και των αέριων ρύπων (Οζον (O<sub>3</sub>), Διοξείδιο του Αζώτου (NO<sub>2</sub>) και Διοξείδιο του Θείου (SO<sub>2</sub>)), διάφορες χρονικές υστερήσεις, τις καμπύλες δόσης-απόκρισης, το θέμα της μετατόπισης της θνησιμότητας ή μεθοδολογικά ζητήματα. Σχετικά με τα PM<sub>10</sub> διαπιστώθηκε πως η αύξηση κατά 10 μg/m<sup>3</sup> σχετίζεται με αύξηση στην ημερήσια ολική θνησιμότητα κατά 0.6%, στην καρδιαγγειακή κατά 0.8% και στην αναπνευστική κατά 1% (Katsouyanni et al., 2001). Η επίδραση των PM<sub>10</sub> στην ολική και καρδιαγγειακή θνησιμότητα είναι άμεση (μετά την πρώτη και τη δεύτερη ημέρα έκθεσης), ενώ στην αναπνευστική υπάρχει μία παρατεταμένη χρονοϋστέρηση στην επίδραση. Η σχέση μεταξύ της επίδρασης των PM<sub>10</sub> και της θνησιμότητας είναι γραμμική, δηλαδή είτε τα επίπεδα αυξηθούν από 30 στα 40 μg/m<sup>3</sup> είτε από τα 75 στα 85 μg/m<sup>3</sup> η επίδρασή τους στη θνησιμότητα είναι η ίδια (Samoli et al., 2005).

Αργότερα, πραγματοποιήθηκε η πολυκεντρική μελέτη APHENA (Katsouyanni et al., 2009), στην οποία αναλύθηκαν δεδομένα χρονολογικών σειρών από την Ευρώπη, τις ΗΠΑ (δεδομένα από τα πρόγραμμα APHEA και NMMAPS<sup>1</sup> αντίστοιχα) και τον Καναδά, για τη διερεύνηση της επίδρασης της ατμοσφαιρικής ρύπανσης στην υγεία του πληθυσμού. Αντικείμενο της έρευνας ήταν η ανάλυση των δεδομένων με τη χρήση ενός κοινού πρωτοκόλλου, η σύγκριση των αποτελεσμάτων και η εύρεση των πηγών της πιθανής ετερογένειάς τους. Τα αποτελέσματα των εκτιμήσεων της επίδρασης της ατμοσφαιρικής ρύπανσης στην υγεία των πληθυσμών της Ευρώπης και των ΗΠΑ ήταν παρόμοια, ενώ του Καναδά υψηλότερα (Samoli et al., 2008). Μετά από το συνδυασμό των αποτελεσμάτων βρέθηκε ότι αν τα επίπεδα της συγκέντρωσης των PM<sub>10</sub> αυξάνονταν κατά 10 μg/m<sup>3</sup> τότε το ποσοστό των ημερήσιων θανάτων θα αυξανόταν από 0.2 έως 0.6%. Βρέθηκε επίσης ότι πηγές ετερογένειας των επιδράσεων μεταξύ

<sup>1</sup> NMMAPS επιδημιολογική μελέτη των ΗΠΑ για τη διερεύνηση των επιπτώσεων της ατμοσφαιρικής ρύπανσης στην υγεία

Ευρώπης και ΗΠΑ αποτελούν οι κλιματολογικοί παράγοντες (θερμοκρασία, υγρασία) και οι υπόλοιποι ατμοσφαιρικοί ρύποι.

Η πολυκεντρική μελέτη, MED-PARTICLES, οργανώθηκε για την ανάλυση δεδομένων χρονολογικών σειρών, από 10 χώρες της Μεσογείου, οι οποίες διαφέρουν από τις υπόλοιπες ευρωπαϊκές στα επίπεδα ατμοσφαιρικής ρύπανσης λόγω κλίματος, συχνότητας των επεισοδίων μεταφοράς σκόνης από τη Σαχάρα και χημικής σύνθεσης των σωματιδίων. Η κατανομή των σωματιδίων βρέθηκε να διαφέρει στις διάφορες περιοχές κυρίως λόγω των μετεωρολογικών παραγόντων (Karanasidou et al., 2014). Ορισμένες περιοχές είχαν υψηλά επίπεδα συγκεντρώσεων των PM λόγω της κυκλοφορίας οχημάτων, ενώ άλλες λόγω της βιομηχανίας ή της οικιακής θέρμανσης. Βρέθηκαν βραχυπρόθεσμες επιπτώσεις των PM στη θνησιμότητα του πληθυσμού της Μεσογείου, κυρίως από τα λεπτόκοκκα σωματίδια, δηλαδή διαμέτρου μικρότερης από 2.5 μm (Samoli et al. 2013).

Τα επιστημονικά στοιχεία των παραπάνω επιδημιολογικών μελετών σε συνδυασμό με τα στοιχεία κλινικών και τοξολογικών ερευνών συνέβαλαν σε περιβαλλοντικές ρυθμίσεις για την προστασία της υγείας του ανθρώπου (WHO, 2013). Σκοπός τους ήταν η μείωση της έκθεσης σε παράγοντες που επηρεάζουν δυσμενώς την υγεία και η εφαρμογή πολιτικών που να οδηγούν σε μείωση της επιβάρυνσης των επιπτώσεων.

Στη συνέχεια των προηγούμενων ευρημάτων, πραγματοποιήθηκε μία μελέτη στις ΗΠΑ με στόχο τη διερεύνηση των χρονικών τάσεων των επιδράσεων των PM<sub>10</sub> στη θνησιμότητα (Dominici et al. 2007). Συγκεκριμένα, τέθηκε το ερώτημα αν ο κίνδυνος έχει αλλάξει κατά τη διάρκεια της περιόδου από το 1987-2000, στην οποία είχαν γίνει ουσιαστικές αλλαγές και ληφθεί μέτρα μείωσης των επιπέδων των ρύπων με σκοπό την προστασία της υγείας του ανθρώπου. Από την ανάλυση των δεδομένων χρονολογικών σειρών του προγράμματος NMMAPS των ΗΠΑ βρέθηκε μία ασθενής ένδειξη ότι οι άμεσες επιπτώσεις των PM<sub>10</sub> μειώθηκαν κατά τη διάρκεια των χρόνων ωστόσο αυτές συνέχιζαν να συσχετίζονται ισχυρά με αυξημένη θνησιμότητα.

Η ατμοσφαιρική ρύπανση αποτελεί πρόβλημα όλης της Ευρώπης και της Ελλάδας. Οι ατμοσφαιρικοί ρύποι που εκλύονται σε μία χώρα ενδέχεται να μεταφερθούν στην ατμόσφαιρα και να επιδεινώσουν ή να καταστήσουν κακή την ποιότητα του αέρα σε μία άλλη περιοχή. Η χώρα μας βρίσκεται σταθερά στις πρώτες θέσεις των ευρωπαϊκών χωρών με τις υψηλότερες συγκεντρώσεις PM<sub>10</sub> στην ατμόσφαιρα. Το πρόβλημα αυτό είναι ιδιαίτερα εμφανές στην ευρύτερη αστική περιοχή της Αθήνας.

Η Αθήνα κατοικείται από το ένα τρίτο του πληθυσμού της Ελλάδας, περιέχει περισσότερα από 1.600.000 οχήματα καθώς και το ήμισυ των βιομηχανικών και εμπορικών δραστηριοτήτων. Η μεγάλη κλίμακα των εκπομπών των ρύπων σε συνδυασμό με την κλειστή τοπογραφία του λεκανοπεδίου της δυσχεραίνει την ατμοσφαιρική κυκλοφορία και έχει ως αποτέλεσμα τη ρύπανση του αέρα σε σημαντικό βαθμό.

Λαμβάνοντας υπ' όψιν την υιοθέτηση διάφορων πολιτικών για τη μείωση των επιπέδων της ατμοσφαιρικής ρύπανσης αλλά και τις κοινωνικο-οικονομικές αλλαγές των τελευταίων χρόνων πραγματοποιήθηκε, η παρούσα μελέτη για τη διερεύνηση των διαχρονικών αλλαγών από το 1992 έως το 2012 στις επιδράσεις των αιωρούμενων σωματιδίων για την περιοχή της Αθήνας, έχοντας ως αφορμή τη μελέτη των Dominici et al. (2007) που μελέτησαν το ίδιο ερώτημα στις ΗΠΑ.

## 4 ΣΚΟΠΟΣ ΤΗΣ ΜΕΛΕΤΗΣ

---

Σκοπός της μελέτης είναι η διερεύνηση των χρονικών τάσεων της επίδρασης των αιωρούμενων σωματιδίων διαμέτρου μικρότερης των 10 $\mu$ m (PM<sub>10</sub>) στην ημερήσια θνησιμότητα των κατοίκων της Αθήνας για την περίοδο 1992-2012.

## 5 ΜΕΘΟΔΟΛΟΓΙΚΗ ΠΡΟΣΕΓΓΙΣΗ

---

Στο παρόν κεφάλαιο αναπτύσσονται το μοντέλο πολλαπλής γραμμικής παλινδρόμησης που χρησιμοποιήθηκε για την πρόβλεψη των επιπέδων των PM<sub>10</sub> και το μοντέλο ανάλυσης επιδημιολογικών χρονοσειρών για την εκτίμηση της επίδρασης των PM<sub>10</sub> και των διαχρονικών τάσεων αυτής στη θνησιμότητα.

### 5.1 ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Η ανάλυση παλινδρόμησης αποτελεί τη μεθοδολογία για ανάλυση μεταβλητών και μοντελοποίηση αυτών όταν μας ενδιαφέρει η σχέση μεταξύ μίας μεταβλητής (Y) και μίας (X) ή πολλών άλλων μεταβλητών ( $X_i$ ,  $i=1, 2, \dots, n$ ). Η Y ονομάζεται εξαρτημένη μεταβλητή και οι  $X_i$  ανεξάρτητες ή επεξηγηματικές.

Με την ανάλυση παλινδρόμησης έχουμε τη δυνατότητα να απαντήσουμε σε μία πληθώρα ερωτημάτων που αφορούν τη σχέση μεταξύ εξαρτημένης μεταβλητής και επεξηγηματικών μεταβλητών όπως «τί συμβαίνει στην εξαρτημένη μεταβλητή κατά μέση τιμή όταν κάποια επεξηγηματική μεταβλητή αλλάζει, ενώ όλες οι υπόλοιπες παραμένουν σταθερές;» Όταν υπάρχει γραμμική σχέση μεταξύ της αναμενόμενης τιμής της Y και των  $X_i$  τότε αυτή μπορεί να περιγραφεί από ένα μοντέλο γραμμικής παλινδρόμησης. Δεν είναι συνηθισμένο η πραγματική σχέση μεταξύ δύο ή περισσότερων μεταβλητών να είναι πλήρως γραμμική αλλά υποθέτουμε ότι το μοντέλο αποτελεί μια περιγραφή της σχέσης μεταξύ των μεταβλητών και δε συμμορφώνεται αναγκαστικά με την πραγματική τους σχέση. Την προσεγγίζει όμως τόσο ώστε να δικαιολογείται η χρήση του μοντέλου. Αποκλίσεις των δεδομένων από το εφαρμοζόμενο μαθηματικό μοντέλο μπορεί να απεικονίζουν την ασυμφωνία μεταξύ του μοντέλου και της φύσης, καθώς επίσης και πηγές ανακρίβειας στη συλλογή των πληροφοριών. Γενικά, επιλέγουμε το μοντέλο που εφαρμόζει όσο το δυνατόν καλύτερα με τα παρατηρούμενα δεδομένα έτσι ώστε το μεγαλύτερο μέρος της απόκλισης των δεδομένων από το μοντέλο να οφείλεται στην τυχαία μεταβλητότητα.

Στη περίπτωση μίας επεξηγηματικής μεταβλητής έχουμε απλή γραμμική παλινδρόμηση (simple linear regression) ενώ όταν έχουμε παραπάνω, ονομάζεται πολλαπλή γραμμική παλινδρόμηση (multiple linear regression). Στις επόμενες παραγράφους παρουσιάζουμε λίγο πιο αναλυτικά το μοντέλο πολλαπλής γραμμικής παλινδρόμησης.

### ΜΟΝΤΕΛΟ ΓΡΑΜΜΙΚΗΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Έστω ότι από ένα πληθυσμό λαμβάνουμε ένα δείγμα μεγέθους n και για κάθε παρατήρηση καταγράφουμε τις τιμές της εξαρτημένης μεταβλητής και των ανεξάρτητων. Με βάση αυτά τα δεδομένα, θα διερευνήσουμε τη σχέση μεταξύ των μεταβλητών. Έστω ότι έχουμε ένα δείγμα  $\{Y_i, X_{i1}, \dots, X_{ik-1}\}$ ,  $i=1, \dots, n$ , από n παρατηρήσεις με k-1 επεξηγηματικές μεταβλητές. Το μοντέλο της πολλαπλής

παλινδρόμησης υποθέτει ότι η εξαρτημένη μεταβλητή συνδέεται με τις ανεξάρτητες με τη σχέση

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{k-1} X_{ik-1} + \varepsilon_i, i = 1, 2, \dots, n \quad [1]$$

Όπου  $\beta_0$  ο σταθερός όρος της παλινδρόμησης και δείχνει την τιμή της εξαρτημένης  $Y$  όταν όλες οι επεξηγηματικές έχουν την τιμή 0 και  $\beta_1, \beta_2, \dots, \beta_{k-1}$  είναι οι παράμετροι κλίσης και δείχνουν τη μεταβολή της μέσης τιμής της  $Y$  για μεταβολή μιας συγκεκριμένης επεξηγηματικής μεταβλητής κατά μία μονάδα, ενώ όλες οι υπόλοιπες παραμείνουν σταθερές. Στη σχέση [1] περιλαμβάνεται και ο στοχαστικός όρος  $\varepsilon_i$  (σφάλμα παλινδρόμησης), ο οποίος είναι μία τυχαία μεταβλητή με τιμές μη παρατηρήσιμες. Αυτές οφείλονται σε τυχαίους μη σταθερούς παράγοντες που επηρεάζουν τις μεταβολές στις τιμές των ανεξάρτητων μεταβλητών. Για αυτό το λόγο οι τιμές θεωρούνται ότι προέρχονται από κατανομές με  $E(\varepsilon_i)=0$ , είναι ανεξάρτητες μεταξύ τους καθώς και με τις τιμές των ανεξάρτητων μεταβλητών. Υποθέτουμε επίσης την ύπαρξη ομοσκεδαστικότητας (όλοι οι όροι  $\varepsilon_i$  έχουν την ίδια διακύμανση για κάθε  $i=1, 2, \dots, n$ ).

Οι συντελεστές της πολλαπλής παλινδρόμησης  $\beta_0, \beta_1, \dots, \beta_{k-1}$ , καθώς και η διασπορά του στοχαστικού όρου είναι άγνωστες παράμετροι και στόχος μας είναι να εκτιμήσουμε όλες τις παραπάνω παραμέτρους όσο το δυνατόν καλύτερα (δηλαδή πιο κοντά στις πραγματικές τους τιμές).

Εδώ θα χρειαστεί να τονιστούν δύο πράγματα. Αρχικά, ότι η γραμμικότητα του μοντέλου αναφέρεται στις παραμέτρους του και όχι στις επεξηγηματικές μεταβλητές του (π.χ.  $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{k-1} X_{ik-1} + \varepsilon_i$  είναι ένα γραμμικό μοντέλο, αντιθέτως με το  $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2^2 X_{i2} + \dots + \beta_{k-1} X_{ik-1} + \varepsilon_i$ ). Επίσης, κάθε παράμετρος επηρεάζεται από όλες τις επεξηγηματικές μεταβλητές που περιλαμβάνονται στο μοντέλο. Έτσι, για παράδειγμα, η παράμετρος  $\beta_1$  δεν εξηγεί τη συνολική επίδραση της  $X_1$  πάνω στην  $Y$  αλλά την «επιπλέον» επίδραση όταν προσθέσουμε τη συγκεκριμένη μεταβλητή στο μοντέλο, αν όλες οι άλλες παραμείνουν ίδιες. Άρα ως αποτέλεσμα των παραπάνω έχουμε ότι οι εκτιμήσεις των παραμέτρων μπορεί να αλλάζουν κάθε φορά που είτε προσθέτουμε καινούριες μεταβλητές στο μοντέλο είτε αφαιρούμε ήδη υπάρχουσες.

Το μοντέλο πολλαπλής παλινδρόμησης

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{k-1} X_{ik-1} + \varepsilon_i, i = 1, 2, \dots, n$$

μπορεί να γραφτεί σε μορφή πινάκων ως εξής. Αν θέσουμε

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & X_{11} & \dots & X_{1,k-1} \\ 1 & X_{21} & \dots & X_{2,k-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{n,k-1} \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{k-1} \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

θα έχουμε

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon}$$

Με τις υποθέσεις ότι  $E(\varepsilon_i) = 0$  και  $Var(\varepsilon_i) = \sigma^2 I$  όπου  $\sigma^2$  η διακύμανση του διαταρακτικού όρου και  $I$  ο μοναδιαίος πίνακας  $n \times n$ .

### ΕΚΤΙΜΗΣΗ ΤΩΝ ΣΥΝΤΕΛΕΣΤΩΝ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Μία από τις πιο ευρέως εφαρμόσιμες μεθόδους εκτίμησης των παραμέτρων του γραμμικού μοντέλου είναι η μέθοδος των ελαχίστων τετραγώνων, η οποία είναι εννοιολογικά απλή και υπολογιστικά ξεκάθαρη. Ένα άλλο σημαντικό πλεονέκτημα της μεθόδου είναι ότι έγκυρα αποτελέσματα από ένα σχετικά μικρό σύνολο παρατηρήσεων. Σε αυτή τη μέθοδο η εκτίμηση των συντελεστών του γραμμικού υποδείγματος  $\beta_0, \beta_1, \dots, \beta_{k-1}$  στηρίζεται στην αρχή ότι οι εκτιμήσεις της εξαρτημένης μεταβλητής  $Y$  (που θα προκύψουν από την εκτίμηση των συντελεστών) θα πρέπει να έχουν την ελάχιστη δυνατή απόσταση από τις παρατηρούμενες τιμές τους. Αυτό θα συμβεί αν ελαχιστοποιήσουμε το άθροισμα των τετραγώνων των καταλοίπων  $\hat{\epsilon}_i$  όπου ως κατάλοιπα ορίζουμε τη διαφορά μεταξύ των παρατηρούμενων τιμών των δεδομένων και των προβλεπόμενων ή προσαρμοσμένων τιμών που προκύπτουν από το γραμμικό μοντέλο. Άρα, αν θέσουμε

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_{k-1} \end{bmatrix}$$

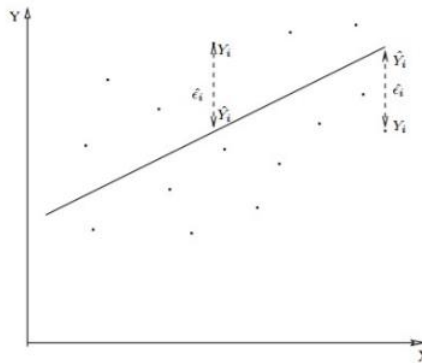
το διάνυσμα των εκτιμώμενων παραμέτρων, η παλινδρόμηση εκτιμάται από την ευθεία

$$\hat{Y} = X\hat{\beta} \quad [2]$$

και τα κατάλοιπα ορίζονται ως

$$\epsilon_i = Y - X\hat{\beta} \quad [3]$$

Όπου η τιμή  $\hat{Y}_i$  καλείται προσαρμοσμένη τιμή (fitted value) ενώ η  $Y_i$  παρατηρούμενη τιμή (observed value).



*Γράφημα 3: Γραμμή παλινδρόμησης και κατάλοιπα*

Στο **Γράφημα 3** ο οριζόντιος άξονας αντιπροσωπεύει τις τιμές μιας ανεξάρτητης μεταβλητής και ο κάθετος τις τιμές της εξαρτημένης μεταβλητής  $Y$ . Τα σημεία του σχήματος αντιστοιχούν σε ζεύγη παρατηρήσεων  $(Y_i, X_i)$  και η ευθεία που βρίσκεται στο σχήμα είναι η ευθεία της παλινδρόμησης η οποία όπως είπαμε και πριν, εκτιμάται από την ευθεία  $\hat{Y} = X\hat{\beta}$ . Τα κατάλοιπα τα οποία βρίσκονται είναι οι αποκλίσεις των προσαρμοσμένων τιμών από τις παρατηρούμενες  $\hat{\epsilon}_i = Y_i - \hat{Y}_i$ . Η μέθοδος των ελαχίστων τετραγώνων δεν ελαχιστοποιεί το άθροισμα των καταλοίπων, το οποίο είναι σταθερά 0, αλλά το άθροισμα των τετραγώνων τους.

Θέλουμε να ελαχιστοποιήσουμε το άθροισμα

$$\begin{aligned} \sum \hat{\epsilon}_i^2 &= \hat{\epsilon}_i' \epsilon_i = (Y - X\hat{\beta})'(Y - X\hat{\beta}) = Y'Y - Y'X\hat{\beta} - \hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \\ &= Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \end{aligned}$$

Για να ελαχιστοποιήσουμε ως προς  $\hat{\beta}$  πρέπει

$$\frac{\partial(Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta})}{\partial\hat{\beta}} = 0$$

Χρησιμοποιώντας διαφορικό λογισμό έχουμε ότι το παραπάνω ισχύει όταν

$$X'X\hat{\beta} = X'Y$$

Αν ο πίνακας  $X'X$  είναι αντιστρέψιμος ο εκτιμητής ελαχίστων τετραγώνων που προκύπτει δίνεται από τον τύπο

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Μπορεί να αποδειχθεί ότι  $E(\hat{\beta}) = \beta$  και  $Var(\hat{\beta}) = \sigma^2(X'X)^{-1}$  δηλαδή ότι ο εκτιμητής είναι αμερόληπτος και αποτελεσματικός.

Ένας πίνακας  $X'X$  για να είναι αντιστρέψιμος πρέπει οι στήλες του να αποτελούν ανεξάρτητα διανύσματα. Αυτό στην πραγματικότητα σημαίνει ότι οι ανεξάρτητες μεταβλητές του μοντέλου αποτελούν διαφορετικές πηγές ερμηνευτικότητας της εξαρτημένης μεταβλητής  $Y$ . Αν δε συμβαίνει αυτό, τότε κάποιες από τις ανεξάρτητες μεταβλητές έχουν υψηλό βαθμό συσχέτισης. Αυτό το πρόβλημα μπορούμε να το λύσουμε με τον «αποκλεισμό» από το μοντέλο αυτών των μεταβλητών ή με αύξηση του αριθμού των παρατηρήσεων, ελπίζοντας ότι αυτό θα έχει ως αποτέλεσμα τη μείωση της συσχέτισης μεταξύ των ανεξάρτητων μεταβλητών, οι οποίες προηγουμένως παρουσίαζαν υψηλή συσχέτιση. (Χανιαλίδης  $X$ ., διπλωματική εργασία)

#### ΕΡΜΗΝΕΥΟΝΤΑΣ ΤΗ ΣΥΝΟΛΙΚΗ ΜΕΤΑΒΛΗΤΟΤΗΤΑ ΤΟΥ ΜΟΝΤΕΛΟΥ

Η δειγματική διασπορά των παρατηρήσεων  $Y_i$  χωρίζεται σε δύο αθροίσματα:

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \quad [4]$$

τα οποία συμβολίζονται με SST, SSE και SSR αντίστοιχα. Το SST εκφράζει τη συνολική παρατηρούμενη μεταβλητότητα των  $Y_i$ , το SSR εκφράζει τη μεταβλητότητα των προσαρμοσμένων τιμών ενώ το SSE εκφράζει τη μεταβλητότητα των  $Y_i$  σε σχέση με τις αντίστοιχες προσαρμοσμένες τιμές. Η μεταβλητότητα του SSR ερμηνεύεται από το μοντέλο ενώ του SSE όχι. Το ηλίκο ή συντελεστής προσδιορισμού

$$R^2 = \frac{SSR}{SSE} = \frac{SST - SSE}{SST} \quad [5]$$

Εκφράζει το ποσοστό της μεταβλητότητας των  $Y_i$  που ερμηνεύεται από το μοντέλο.

#### ΣΥΓΚΡΙΣΗ ΜΟΝΤΕΛΩΝ

Όταν θέλουμε να εξετάσουμε ποιες από τις ανεξάρτητες μεταβλητές επηρεάζουν τη μεταβλητή απόκρισης  $Y$  εφαρμόζουμε όλα τα δυνατά μοντέλα και επιλέγουμε αυτό που δίνει τα «καλύτερα» αποτελέσματα ως προς κάποιο κριτήριο καλής εφαρμογής του μοντέλου στα εμπειρικά δεδομένα. Αν για παράδειγμα έχουμε καταγράψει τρεις μεταβλητές  $X_1, X_2, X_3$  τότε μπορούμε να εξετάσουμε τα μοντέλα:

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\varepsilon}, \quad Y = \hat{\beta}_0 + \hat{\beta}_1 X_2 + \hat{\varepsilon}, \quad Y = \hat{\beta}_0 + \hat{\beta}_1 X_3 + \hat{\varepsilon},$$

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\varepsilon}, \quad Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_3 + \hat{\varepsilon},$$

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_2 + \hat{\beta}_2 X_3 + \hat{\varepsilon}, \quad Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\varepsilon}$$

Και να θεωρήσουμε καλύτερο αυτό που δίνει το μεγαλύτερο συντελεστή προσδιορισμού  $R^2$ . Θα διαπιστώσουμε ότι το μοντέλο με το μεγαλύτερο  $R^2$  δεν είναι πάντοτε το καλύτερο. Αυτό συμβαίνει γιατί όσο προσθέτουμε ανεξάρτητες μεταβλητές

στο μοντέλο το  $R^2$  αυξάνεται ή παραμένει σταθερό. Όταν προσθέτουμε ανεξάρτητες μεταβλητές το SSE μειώνεται ή μένει σταθερό, αφού

$$SSE = \min_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{k-1}} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \dots - \hat{\beta}_{k-1} X_{ik-1})^2 \quad [6]$$

Και η ελαχιστοποίηση γίνεται σε μεγαλύτερο χώρο (περισσότερα  $\hat{\beta}_i$ ) ενώ το SST παραμένει πάντοτε σταθερό. Αντί του  $R^2$  προτείνεται η χρήση του προσαρμοσμένου  $R^2$  (adjusted). Έτσι, καλύτερο θα θεωρείται το μοντέλο με το μεγαλύτερο

$$R^2_{adj} = 1 - \frac{SSE/(n-k)}{SST/(n-1)} \quad [7]$$

Το προσαρμοσμένο  $R^2$  δείχνει ως καλύτερο το μοντέλο που έχει το μικρότερο  $SSE/(n-k)$  (το SST είναι σταθερό σε όλα τα μοντέλα), ενώ δεν αυξάνεται πάντοτε όταν αυξάνεται ο αριθμός των ανεξάρτητων μεταβλητών. (Μπούτσικας Μ., 2004)

## 5.2 ΑΝΑΛΥΣΗ ΧΡΟΝΟΣΕΙΡΩΝ

Το σύνολο των δεδομένων, τα οποία συλλέγονται διαχρονικά και εκφράζουν την εξέλιξη των τιμών μιας μεταβλητής κατά τη διάρκεια ίσων διαδοχικών χρονικών περιόδων, ονομάζεται χρονοσειρά ή χρονολογική σειρά (time series). Η χρονοσειρά αποτελείται από ένα σύνολο παρατηρήσεων μιας μεταβλητής, οι τιμές της οποίας είναι ιεραρχημένες με βάση τη χρονική περίοδο στην οποία αναφέρονται (π.χ. έτος, τρίμηνο, μήνας κ.α.)

Μαθηματικά μία χρονοσειρά ορίζεται από τις τιμές  $Y_1, Y_2, \dots$  κάποιας μεταβλητής  $Y$  κατά τις χρονικές στιγμές  $t_1, t_2, \dots$ . Επομένως το  $Y$  είναι μία συνάρτηση του  $t$  και συμβολίζεται  $Y=F(t)$ . Η γραφική παράσταση της συνάρτησης παρουσιάζει την εξέλιξη της μεταβλητής  $Y$  στο χρόνο.

Η ανάλυση χρονοσειρών εφαρμόζεται σε επιδημιολογικές μελέτες που διαθέτουν τακτικές σειρές περιβαλλοντικών και ιατρικών δεδομένων, όπως ο ημερήσιος αριθμός θανάτων, ο αριθμός εισαγωγών στα νοσοκομεία, οι συγκεντρώσεις των ρύπων και οι τιμές της θερμοκρασίας. Βασικός στόχος είναι συνήθως η προσέγγιση της σχέσης μεταξύ της έκβασης και μίας ή περισσότερων σειρών έκθεσης (Touloumi et al., 2006). Εκτιμάται δηλαδή πόσο θα μεταβληθεί η έκβαση (για παράδειγμα ο ημερήσιος αριθμός θανάτων) αν μεταβληθούν τα επίπεδα της έκθεσης (όπως η συγκέντρωση των  $PM_{10}$ ) στην ατμόσφαιρα.

## 5.3 ΠΑΛΙΝΔΡΟΜΗΣΗ POISSON

Οι εκβάσεις ( $Y$  εξαρτημένη μεταβλητή) στις μελέτες της περιβαλλοντικής επιδημιολογίας εκφράζουν αριθμό γεγονότων (counts) κάποιου δείκτη υγείας, συνεπώς περιγράφονται από τη στοχαστική διαδικασία Poisson και αναλύονται χρησιμοποιώντας τα αντίστοιχα γενικευμένα μοντέλα παλινδρόμησης.

Η εφαρμογή μιας στοχαστικής διαδικασίας Poisson υποθέτει ότι ο κίνδυνος της υπό μελέτη έκβασης για τον υποκείμενο πληθυσμό είναι ομοιογενής (Nurminen, 1997), ενώ η πιθανότητα εμφάνισης  $y$  συμβάντων σε μία ημέρα δίνεται από τον τύπο

$$\Pr(y|\lambda) = \frac{e^{-\lambda} \lambda^y}{y!} \quad [8]$$

Όπου  $\lambda$  ο αναμενόμενος κίνδυνος εμφάνισης της έκβασης. Ισχύει η υπόθεση ότι  $E(Y) = \mu = \lambda$  και  $Var(Y) = \lambda$ , δηλαδή ότι η διακύμανση είναι ίση με την αναμενόμενη τιμή και δεν είναι σταθερή. Αυτό σημαίνει ότι όταν ο αριθμός των γεγονότων θα είναι μεγάλος τότε οι παρατηρήσεις θα παρουσιάζουν μεγάλη διακύμανση ενώ όταν ο αριθμός των γεγονότων θα είναι μικρός τότε και η διακύμανση των παρατηρήσεων αντίστοιχα θα μικραίνει.

Η γενική μορφή των μοντέλων παλινδρόμησης Poisson είναι:

$$\ln \mu_t = \ln E(Y_t) = \beta_1 X_{1t} + \dots + \beta_k X_{kt} \quad [9]$$

Όπου  $Y_t$  ο παρατηρούμενος αριθμός συμβάντων την ημέρα  $t$ ,  $E(Y_t)$  η αναμενόμενη ή μέση τιμή,  $X_1, X_2, \dots, X_k$  οι ανεξάρτητες ή επεξηγηματικές μεταβλητές και  $\beta_1, \beta_2, \dots, \beta_k$  οι αντίστοιχοι συντελεστές εξάρτησης.

Τα μοντέλα παλινδρόμησης Poisson είναι πολλαπλασιαστικά. Αυτό σημαίνει ότι αν αυξηθούν τα επίπεδα της μίας, έστω  $X_i$ , εκ των  $k$  ανεξάρτητων μεταβλητών,  $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_k$ , υπό την απουσία αλληλεπίδρασης, κατά  $c$  μονάδες, τότε θα μεταβληθεί η εξαρτημένη μεταβλητή κατά  $e^{c\beta_i}$ . Αναλυτικά, αν

$$\text{τη χρονική στιγμή } t_1 \text{ έχουμε, } \ln[E(Y_{t_1})] = \beta_0 + \beta_1 X_1 + \dots + \beta_i X_i + \dots + \beta_k X_k \quad [10]$$

ενώ

$$\text{τη χρονική στιγμή } t_2 \text{ έχουμε, } \ln[E(Y_{t_2})] = \beta_0 + \beta_1 X_1 + \dots + \beta_i (X_i + c) + \beta_k X_k \quad [11]$$

Δηλαδή τα επίπεδα της  $X_i$  τη χρονική στιγμή  $t_2$  αυξήθηκαν κατά  $c$  μονάδες, τότε αφαιρώντας τις σχέσεις [10],[11] έχουμε ότι,

$$\ln[E(Y_{t_2})] - \ln[E(Y_{t_1})] =$$

$$(\beta_0 + \beta_1 X_{1t} + \dots + \beta_i (X_i + c) + \beta_k X_{kt}) - (\beta_0 + \beta_1 X_{1t} + \dots + \beta_i X_i + \dots + \beta_k X_{kt})$$

$$\Rightarrow \ln[E(Y_{t_2})] - \ln[E(Y_{t_1})] = \beta_i (X_i + c - X_i) \Rightarrow \ln[E(Y_{t_1})] - \ln[E(Y_{t_2})] = \beta_i$$

$$\Rightarrow \ln \frac{E(Y_{t_2})}{E(Y_{t_1})} = c\beta_i \Rightarrow \frac{E(Y_{t_2})}{E(Y_{t_1})} = e^{c\beta_i} \quad [12]$$

Ένα πρόβλημα που συνήθως συναντάται στην παλινδρόμηση Poisson είναι η μεγάλη διακύμανση, δηλαδή όταν η διακύμανση της εξαρτημένης είναι μεγαλύτερη από τη μέση τιμή, σύμφωνα με τις ιδιότητες της κατανομής Poisson. Το πρόβλημα αυτό αντιμετωπίζεται εκτιμώντας τη διακύμανση της διαδικασίας Poisson ως  $Var(Y) = \varphi \cdot \lambda$ , όπου  $\varphi$  η παράμετρος υπερδιασποράς. Στα μοντέλα παλινδρόμησης η σταθερά  $\varphi$  εκτιμάται από τα κατάλοιπα του μοντέλου κατά Pearson  $\chi^2$  μέσω του τύπου

$$\hat{\varphi} = \frac{\chi^2}{n - k} = \frac{\sum_i \frac{y_t - \hat{y}_t}{\hat{y}_t}}{n - k} \quad [13]$$

Όπου  $y_i$  οι παρατηρούμενες τιμές της εξαρτημένης μεταβλητής,  $\hat{y}_i$  οι αναμενόμενες τιμές,  $n$  ο αριθμός των παρατηρήσεων και  $k$  ο αριθμός των παραμέτρων. Στη συνέχεια γίνεται πολλαπλασιασμός της σταθεράς με τον πίνακα μεταβλητότητας.

## 5.4 ΗΜΙ-ΠΙΘΑΝΟΦΑΝΕΙΑ

Οι παράμετροι ενός μοντέλου Poisson εκτιμώνται με τη μέθοδο της μέγιστης Πιθανοφάνειας. Η μέθοδος αυτή επιλέγει το σύνολο των παραμέτρων του μοντέλου που μεγιστοποιεί τη συνάρτηση Πιθανοφάνειας. Διαισθητικά, αυτό μεγιστοποιεί τη «συμφωνία» από το επιλεγμένο μοντέλο με τα παρατηρούμενα δεδομένα. Συγκεκριμένα, έστω  $Y_1, Y_2, \dots, Y_n$  τυχαίο δείγμα από την κατανομή Poisson και



θέλουμε να βρούμε την εκτιμήτρια μέγιστης Πιθανοφάνειας, τότε θα ακολουθήσουμε τα τρία βήματα:

1. Εύρεση της συνάρτησης Πιθανοφάνειας

$$L(\theta) = L(\theta, y_1, y_2, \dots, y_n) = \prod_{i=1}^n P(y_i, \theta)$$

2. Μεγιστοποίηση της  $L(\theta)$  ως προς  $\theta$

$$\hat{\theta} \text{ θα είναι η τιμή για την οποία: } \frac{\partial L(\theta)}{\partial \theta} = 0 \text{ ή } \frac{\partial l(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \log L(\theta) = 0$$

3. Για να είναι μέγιστο, πρέπει  $\frac{\partial^2 L(\theta)}{\partial \theta^2} < 0$  ή  $\frac{\partial^2 l(\theta)}{\partial \theta^2} < 0$

Συνεπώς, έχουμε:

$$L(\lambda) = \prod_{j=1}^n P(y_j, \lambda) = \prod_{j=1}^n e^{-\lambda} \frac{\lambda^{y_j}}{y_j!} = e^{-n\lambda} \frac{\lambda^{\sum_{j=1}^n y_j}}{y_1! y_2! \dots y_n!} \quad [14]$$

Όπου  $Y_1, Y_2, \dots, Y_n$  ο αριθμός των γεγονότων. Τότε η συνάρτηση πιθανοφάνειας είναι:

$$L(\lambda) = e^{-n\lambda} \frac{\lambda^{\sum y_j}}{y_1! y_2! \dots y_n!}, \text{ για } \lambda \geq 0 \quad [15]$$

Η λογαριθμοσυνάρτηση Πιθανοφάνειας είναι:

$$l(\lambda) = \log L(\lambda, Y) = \sum_{j=1}^n \log \left( \frac{e^{-\lambda} \lambda^{y_j}}{y_j!} \right) = -n\lambda + \sum_{j=1}^n y_j \log \lambda - \log \prod_{j=1}^n y_j$$

Παραγωγίζοντας ως προς  $\lambda$  την παραπάνω σχέση προκύπτει ότι

$$\frac{\partial}{\partial \lambda} l(\lambda) = -n + \frac{1}{\lambda} \sum y_j$$

Για να βρούμε το μέγιστο εξισώνουμε την παραπάνω σχέση με το 0 και λύνουμε ως προς  $\lambda$  για να βρούμε το ζητούμενό μας, δηλαδή,

$$\frac{\partial}{\partial \lambda} l(\lambda) = 0 \Rightarrow -n + \frac{1}{\lambda} \sum y_j = 0 \Rightarrow \hat{\lambda} = \frac{\sum y_j}{n} = \bar{y} \quad [16]$$

Στη συνέχεια βρίσκουμε τη δεύτερη παράγωγο της λογαριθμοσυνάρτησης Πιθανοφάνειας ως προς  $\lambda$

$$\frac{\partial^2}{\partial \lambda^2} l(\lambda, Y)_{\lambda=\hat{\lambda}} = -\frac{1}{\hat{\lambda}^2} \sum y_j < 0$$

Και αφού αυτή είναι αρνητική, το  $\hat{\lambda}$  αποτελεί την εκτιμήτρια μέγιστης Πιθανοφάνειας για την κατανομή Poisson (Παπαγεωργίου Χαράλαμπος).

Όταν υπάρχει το πρόβλημα της υπερ-διασποράς τότε δεν υπολογίζουμε την Πιθανοφάνεια αλλά την Ημι-Πιθανοφάνεια. Για την κατασκευή της ημι-Πιθανοφάνειας ορίζουμε τη συνάρτηση ως εξής:

$$U = u(\mu, Y) = \frac{y - \mu}{\phi \text{Var}(\mu)} \quad [17]$$

Όπου  $\mu$  η μέση τιμή και  $\phi$  η παράμετρος υπερ-διασποράς. Ισχύει ότι  $E(U) = 0$ ,  $\text{Var}(U) = \frac{1}{\phi \text{Var}(\mu)}$  και  $-E \frac{\partial Q}{\partial \mu} = \frac{1}{\phi \text{Var}(\mu)}$

Η συνάρτηση λογαριθμικής Πιθανοφάνειας είναι :

$$Q(\mu, Y) = \int_y^\mu \frac{y-t}{\phi \text{Var}(t)} dt \quad [18]$$

Και για να βρούμε την εκτιμήτρια μέγιστης Ημι-Πιθανοφάνειας  $\hat{\theta}$  έχουμε:

$$\frac{\vartheta}{\vartheta\theta} \sum_{i=1}^n Q_i = \sum_{i=1}^n \frac{y_i - \mu_i}{Var(\mu_i)} \frac{\vartheta \lambda_i}{\vartheta\theta} = \sum_{i=1}^n \frac{y_i - \mu_i}{Var(\mu_i)} \frac{x_i}{g(\mu_i)} = 0 \quad [19]$$

Όπου  $\mu_i$  η μέση τιμή και  $g$  η συνάρτηση δεσμού της κατανομής Poisson (Jeff Jill).

## 5.5 ΑΥΤΟΣΥΣΧΕΤΙΣΗ

Η αυτοσυσχέτιση που παρατηρείται συνήθως στην ανάλυση στοιχείων χρονολογικών σειρών και αναφέρεται συχνά ως αυτοπαλινδρόμηση ή ως σειριακή συσχέτιση. Αυτό συμβαίνει καθώς οι ανεξάρτητες και οι εξαρτημένες μεταβλητές αποτελούν διαδοχικές μετρήσεις στο χρόνο, συνεπώς συσχετίζονται και όσο πιο κοντά χρονικά είναι μεταξύ τους τόσο μεγαλύτερη είναι και η συσχέτισή τους. Σε ένα μοντέλο παλινδρόμησης Poisson

$$\ln E(Y_i) = \sum_{j=1}^k \hat{\beta}_j X_{ij} + e_i \quad [20]$$

Όπου  $Y_i$  η εξαρτημένη μεταβλητή,  $X_{ij}$  οι ανεξάρτητες μεταβλητές,  $\hat{\beta}_j$  οι αντίστοιχοι συντελεστές μερικής εξάρτησης και  $e_i$  τα κατάλοιπα του μοντέλου, οι εκτιμώμενες τιμές προκύπτουν από τον τύπο:

$$\hat{Y}_i = e^{\sum_{j=1}^k \hat{\beta}_j X_{ij}} \quad [21]$$

Και τα κατάλοιπα υπολογίζονται ως οι διαφορές των εκτιμώμενων από τις πραγματικές τιμές:

$$e_i = Y_i - \hat{Y}_i \quad [22]$$

Προϋπόθεση του μοντέλου αποτελεί η ανεξαρτησία και η τυχαιότητα των καταλοίπων. Όταν αυτά δεν πληρούνται τότε τα αποτελέσματα είναι μεροληπτικά. Συγκεκριμένα, εισάγεται σφάλμα στην εκτίμηση των τυπικών σφαλμάτων των συντελεστών μερικής εξάρτησης  $\hat{\beta}_j$  αλλά όχι στους ίδιους τους συντελεστές. Έτσι, δεν μπορούμε να έχουμε έγκυρη αξιολόγηση της σημαντικότητας των  $\hat{\beta}_j$ .

Ο βαθμός εξάρτησης των τιμών του διαταρακτικού όρου εκτιμάται με το συντελεστή αυτοσυσχέτισης :

$$\rho_{\varepsilon_t, \varepsilon_{t-s}} = \frac{\text{cov}(\varepsilon_t, \varepsilon_{t-s})}{\sqrt{\text{Var}(\varepsilon_t)} \sqrt{\text{Var}(\varepsilon_{t-s})}} \quad [23]$$

με  $-1 \leq \rho \leq 1, s \neq 0$

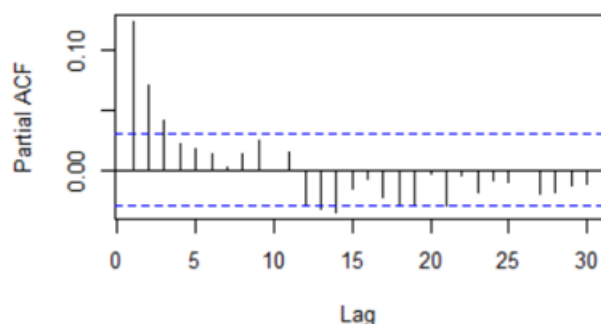
όπου  $s$  μια χρονική μετακίνηση είτε προς τα εμπρός είτε προς τα πίσω, δηλαδή μπορεί να είναι είτε θετικό είτε αρνητικό.

Ο συντελεστής συσχέτισης δεν εξαρτάται από το  $t$  αλλά μόνο από την υστέρηση  $s$ . Η σχέση που υπάρχει ανάμεσα στο συντελεστή συσχέτισης  $\rho$  και στην χρονική υστέρηση  $s$ , ονομάζεται συνάρτηση αυτοσυσχέτισης (autocorrelation function) και η γραφική της απεικόνιση ονομάζεται διάγραμμα αυτοσυσχέτισης (correlogram). Στην ανάλυση χρονολογικών σειρών η σημασία της συνάρτησης αυτοσυσχέτισης είναι πολύ μεγάλη, γιατί δείχνει τόσο το βαθμό όσο και το μήκος ή τη χρονική διάρκεια της μνήμης της στοχαστικής διαδικασίας.

Ένα διάγραμμα αυτοσυσχέτισης εμφανίζει γραφικά και αριθμητικά τη συνάρτηση αυτοσυσχέτισης (ACF), που είναι το σύνολο των συντελεστών συσχέτισης μαζί με τις τυπικές αποκλίσεις των συντελεστών για διαδοχικές υστερήσεις σε ένα συγκεκριμένο φάσμα υστερήσεων (παράδειγμα από 1 έως 30). Για παράδειγμα αν η πρώτη

παρατήρηση συσχετίζεται ισχυρά με τη δεύτερη, και η δεύτερη με την τρίτη τότε η πρώτη σχετίζεται σχετικά με την τρίτη. Αυτό σημαίνει ότι το πρότυπο της σειράς εξαρτήσεων μπορεί να μεταβληθεί μετά την απομάκρυνση του πρώτου βαθμού αυτοσυσχέτισης, δηλαδή με τη χρονική υστέρηση της σειράς κατά 1. Έτσι, αντί του πρώτου στοιχείου  $t$  της σειράς θεωρούμε το στοιχείο  $s-1$ .

Η συνάρτηση μερικής αυτοσυσχέτισης (PACF) αποτελεί μία επέκταση της ACF όπου παρουσιάζεται η εξάρτηση από τα ενδιάμεσα στοιχεία όσο απομακρύνεται η υστέρηση. Δηλαδή, η μερική αυτοσυσχέτιση σε χρονική υστέρηση  $s$  είναι η αυτοσυσχέτιση μεταξύ του  $X_t$  και  $X_{t-s}$  που δεν εξηγείται από τις υστερήσεις 1 έως  $s-1$  (Box & Jenkins). Στην Εικόνα 4 απεικονίζεται ένα γράφημα PACF, στο οποίο ο  $x$ -άξονας περιγράφει τις χρονικές υστερήσεις (lags) 1, 2, ..., 30, ο  $y$ -άξονας το συντελεστή μερικής αυτοσυσχέτισης για κάθε χρονική υστέρηση 1, 2, ..., 30 και η διακεκομμένη γραμμή προσδιορίζει το 95% Διάστημα Εμπιστοσύνης.



*Γράφημα 4: Μερική συσχέτιση καταλοίπων*

Το 95% ΔΕ για το συντελεστή μερικής αυτοσυσχέτισης, για μεγάλο αριθμό παρατηρήσεων προσεγγίζεται από τον τύπο  $\pm 2/\sqrt{n}$ .

Στο **Γράφημα 4** παρατηρούμε ότι ο συντελεστής αυτοσυσχέτισης των χρονικών υστερήσεων 1 και 2 είναι στατιστικά σημαντικοί αφού οι διακεκομμένες γραμμές συμβολίζουν το 95% ΔΕ για τη σημαντικότητα, ενώ οι υπόλοιποι ή δεν είναι ή είναι στο μεταίχμιο της στατιστικής σημαντικότητας.

## 5.6 ΈΛΕΓΧΟΣ ΠΙΘΑΝΩΝ ΣΥΓΧΥΤΙΚΩΝ ΠΑΡΑΓΟΝΤΩΝ

Η ορθή εκτίμηση της συσχέτισης μεταξύ της έκθεσης και της έκβασης προϋποθέτει το συνυπολογισμό των επιδράσεων των συγχυτικών παραγόντων. Βασικοί συγχυτικοί παράγοντες στην ανάλυση χρονοσειρών όπου μονάδα παρατήρησης είναι η μονάδα του χρόνου είναι η μακροχρόνια τάση και η εποχικότητα.

Μακροχρόνια τάση είναι η ομαλή διαχρονική κίνηση, την οποία ακολουθεί μία χρονολογική σειρά κατά μία μακρά χρονική περίοδο και μπορεί να είναι ανοδική, καθοδική, σύνθετη ή και μηδενική σε ορισμένες περιπτώσεις. Η τάση δεν επαναλαμβάνεται ή τουλάχιστον δεν επαναλαμβάνεται μέσα στα όρια του χρόνου που καλύπτουν τα στοιχεία μας. Έτσι λοιπόν σε ένα διάγραμμα η μακροχρόνια τάση περιγράφεται από μία ευθεία γραμμή (ανοδική ή καθοδική) ή μία καμπύλη. Μπορεί να δείξει δηλαδή την ανοδική ή την καθοδική εξέλιξη των συγκεντρώσεων των  $PM_{10}$  στην ατμόσφαιρα.

Εποχικότητα είναι η περιοδική μεταβολή που παρατηρείται σε μία μεταβλητή μέσα σε ένα χρονικό διάστημα (χρόνος, εξάμηνο) και επαναλαμβάνεται με την ίδια περίοδο και στα επόμενα χρονικά διαστήματα. Έτσι για παράδειγμα ο ημερήσιος αριθμός

θανάτων παρουσιάζει αυξομειώσεις μέσα στο έτος, κατά τους χειμερινούς και θερινούς μήνες παρουσιάζει αύξηση και μείωση στους υπόλοιπους. Αυτό το φαινόμενο επαναλαμβάνεται κάθε χρόνο.

Για τον έλεγχο των συγχυτικών παραγόντων στην ανάλυση επιδημιολογικών χρονοσειρών έχουν προταθεί διάφορες μέθοδοι όπως είναι:

- Η χρήση τμηματικών συναρτήσεων με την εφαρμογή διπλής ή τριπλής αλληλεπίδρασης μεταξύ του χρόνου, του μήνα και ακόμη της ημέρας της εβδομάδας. Οι συναρτήσεις αυτές προσφέρουν ευελιξία αλλά δεν αποτελούν την καταλληλότερη προσέγγιση για την περιγραφή της σχέσης. Γραφικά αυτή η σχέση απεικονίζεται σε μία κλιμακωτή γραμμή.
- Η χρήση τριγωνομετρικών συναρτήσεων (ζευγάρια ημιτονοειδών και συνημιτονοειδών) της ημέρας μέσα στο χρόνο, η οποία πρέπει να συνδυάζεται με μία γραμμική ή μη συνάρτηση της χρονικής περιόδου για την εξάλειψη της μακροχρόνιας τάσης. Το μειονέκτημα αυτής της προσέγγισης είναι ότι υποθέτει κορυφές στο ίδιο ύψος και την ίδια χρονική στιγμή κάθε έτος.
- Η χρήση συναρτήσεων εξομάλυνσης του χρόνου, όπως είναι τα πολυώνυμα παρεμβολής (spline functions) τα οποία παρουσιάζουν περισσότερη ευελιξία από τα προηγούμενα. Ωστόσο η χρήση συναρτήσεων εξομάλυνσης του χρόνου χρήζει προσοχής. Η σχέση μεταξύ της εποχικότητας και της μακροχρόνιας τάσης με τη θνησιμότητα ή με τις συγκεντρώσεις των  $PM_{10}$  δεν είναι γνωστή, έτσι ο βαθμός της εξομάλυνσης του χρόνου πρέπει να επιλεγθεί κατάλληλα, ώστε να μην υπάρξει υπερ-εξομάλυνση ή υπο-εξομάλυνση. Η υπερ-εξομάλυνση μπορεί να αφήσει υπολειπόμενη συγχυτική επίδραση στα κατάλοιπα και στην κατανομή τους να παρατηρηθούν κυκλικά πρότυπα. Η υπο-εξομάλυνση μπορεί να οδηγήσει στην υποεκτίμηση της επίδρασης των  $PM_{10}$ , αφού ένα μέρος της επίδρασης μπορεί να μεταφέρεται στην επίδραση των συγχυτικών παραγόντων. Τα κριτήρια για την επιλογή του βαθμού εξομάλυνσης μπορεί να βασίζονται στην ελαχιστοποίηση της συνάρτησης μερικής αυτοσυσχέτισης (minimized PACF), σε μεθόδους καλής προσαρμογής (GCV) ή σε προηγούμενη γνώση.

## 5.7 ΓΕΝΙΚΕΥΜΕΝΑ ΠΡΟΣΘΕΤΙΚΑ ΜΟΝΤΕΛΑ

Η χρήση μεθόδων εξομάλυνσης για τον έλεγχο των συγχυτικών παραγόντων, όπως η εποχικότητα, καθιστά απαραίτητη την εφαρμογή των Γενικευμένων Προσθετικών Μοντέλων (Generalized Additive Models, GAM) που αποτελούν επέκταση των Γενικευμένων Γραμμικών Μοντέλων (GLM) και παρέχουν τη δυνατότητα να περιγραφεί η σχέση μεταξύ της εξαρτημένης μεταβλητής και των ανεξάρτητων τιμών με παραμετρικές και μη παραμετρικές συναρτήσεις εξομάλυνσης.

Η μορφή των Γενικευμένων Προσθετικών Μοντέλων είναι η εξής:

$$g(\mu) = g[E(Y)] = a_o + \sum_{j=1}^q f_j(X_j) = \eta(x) \quad [24]$$

όπου  $g$  η συνάρτηση δεσμού (link function),  $a_o$  μία σταθερά και  $f_j$  συναρτήσεις εξομάλυνσης οι οποίες περιγράφουν τη σχέση μεταξύ της μετασχηματισμένης μέσης τιμής της εξαρτημένη μεταβλητής  $Y$  και των ανεξάρτητων μεταβλητών  $X_i$ . Η  $\eta(x)$  ονομάζεται προσθετικός ερευνητής και για τη διασπορά της  $Y$  ισχύει ότι  $Var(Y) = \varphi \cdot$

$Var(\mu)$ . Η μορφή των Γενικευμένων Προσθετικών Μοντέλων όταν η  $Y$  ακολουθεί την κατανομή Poisson γράφεται:

$$\eta(x) = \ln(\mu_t) = \ln[E(Y_t)] = a_o + \sum_{j=1}^q f_j(X_{tj}, \lambda_j) + \sum_{i=1}^p \beta_i X_i \quad [25]$$

όπου  $Y_t$  ο παρατηρούμενος αριθμός των γεγονότων την ημέρα  $t$ ,  $X_{tj}$  οι μεταβλητές που μεταβάλλονται στο χρόνο και  $f_j$  οι συναρτήσεις εξομάλυνσης με  $\lambda_j$  βαθμούς ελευθερίας. Το  $\sum_{j=1}^q f_j(X_{tj}, \lambda_j)$  αποτελεί το κομμάτι που περιέχει τις συναρτήσεις εξομάλυνσης και το  $\sum_{i=1}^p \beta_i X_i$  το γραμμικό κομμάτι.

Η διαδικασία εκτίμησης των παραμέτρων εξαρτάται από τη χρήση παραμετρικών ή μη παραμετρικών συναρτήσεων εξομάλυνσης. Στην πρώτη περίπτωση οι παράμετροι εκτιμώνται με τη μέθοδο των σταθμισμένων ελαχίστων τετραγώνων (weighted least square). Στη δεύτερη περίπτωση η εκτίμηση των παραμέτρων γίνεται σε δύο στάδια με τη χρήση δύο αλγορίθμων, του αλγορίθμου τοπικής εκτίμησης και του αλγορίθμου backfitting (Hastie & Tibshirani, 1990). Η σύγκλιση των δύο αλγορίθμων ρυθμίζεται από παραμέτρους, η μία δηλώνει το μέγιστο επιτρεπτό αριθμό επαναλήψεων ( $M$ ) και η δεύτερη την ακρίβεια της σύγκλισης ( $\epsilon$ ).

Ο αλγόριθμος τοπικής εκτίμησης είναι μία γενίκευση της μεθόδου εκτίμησης του Fisher (Fisher scoring) που χρησιμοποιείται στα Γενικευμένα Γραμμικά Μοντέλα για την εύρεση των εκτιμητών Μέγιστης Πιθανοφάνειας. Σε κάθε επανάληψη του αλγορίθμου διαμορφώνεται η αντίστοιχη προσαρμοσμένη εξαρτημένη μεταβλητή και στη συνέχεια εφαρμόζεται το προσθετικό μοντέλο. Στον αλγόριθμο backfitting, οι μεταβλητές προσεγγίζονται διαδοχικά ενώ κάθε συνάρτηση εξομάλυνσης εκτιμάται μέσω της εξομάλυνσης των μερικών καταλοίπων (Partial Residuals) (Hastie & Tibshirani, 1986).

Στη χρήση των Γενικευμένων Προσθετικών Μοντέλων έχουν εντοπιστεί ορισμένα προβλήματα τα οποία σχετίζονται με την εφαρμογή των αλγορίθμων για τη μη παραμετρική εξομάλυνση. Συγκεκριμένα, οι Dominici et al. (2002) βρήκαν ότι οι εκτιμώμενες παράμετροι του μοντέλου είναι μικρές και όταν χρησιμοποιούνται περισσότερες από μία μη παραμετρικές συναρτήσεις εξομάλυνσης για τον έλεγχο των συγχυτικών παραγόντων, τότε τα κριτήρια σύγκλισης που εφαρμόζονται δεν οδηγούν πάντοτε σε σύγκλιση και έτσι οι εκτιμήσεις των παραμέτρων του παραμετρικού μέρους του μοντέλου είναι εσφαλμένες. Ένα άλλο πρόβλημα εντοπίστηκε από τους Ramsay et al. (2002) είναι η υπο- εκτίμηση της διασποράς των παραμέτρων η οποία οφείλεται στην παρουσία του μη παραμετρικού ανάλογου της πολυσυγγραμμικότητας. Στα ΓΓΜ η συγγραμμικότητα λαμβάνεται υπόψιν στον υπολογισμό των τυπικών σφαλμάτων, έτσι όσο αυξάνεται η συγγραμμικότητα, τόσο αυξάνονται οι τιμές των τυπικών σφαλμάτων. Στα ΓΠΜ δεν ισχύει αυτό, το τυπικό σφάλμα του παραμετρικού μέρους υπολογίζεται προσεγγίζοντας το μη παραμετρικό μέρος ως γραμμικό. Η αδυναμία των ΓΠΜ να ανιχνεύσουν τη μη παραμετρική πολυσυγγραμμικότητα οδηγεί σε υποεκτίμηση της διασποράς των παραμέτρων του μοντέλου. Ο βαθμός της υποεκτίμησης αυξάνεται με την αύξηση του αριθμού των βαθμών ελευθερίας δεδομένου ότι αγνοείται μεγαλύτερος αριθμός μη γραμμικών όρων κατά τον υπολογισμό του τυπικού σφάλματος του παραμετρικού μοντέλου (Ramsay et al. 2002; Chambers & Hastie, 1992), (Αναστασίου Ζ., διπλωματική εργασία).

Σχετικά με το πρόβλημα της υποεκτίμησης των τυπικών σφαλμάτων έχουν προταθεί από τους Dominici et al. (2004) διορθωμένοι αλγόριθμοι για την προσέγγιση αξιόπιστων επιδράσεων. Για την αποφυγή αυτών των προβλημάτων ωστόσο

συνίσταται η εφαρμογή παραμετρικών συναρτήσεων εξομάλυνσης κι όταν αυτή είναι ανεπαρκής, η εφαρμογή πολωνύμων παρεμβολής με περιορισμό (penalized regression splines) (Marx & Eilers, 1998).

## 5.8 ΕΞΟΜΑΛΥΝΣΗ

Οι συναρτήσεις εξομάλυνσης χρησιμοποιούνται για την περιγραφή της μη γραμμικής σχέσης μεταξύ της έκβασης  $Y$  και μίας ή περισσότερων χρονο-μεταβαλλόμενων μεταβλητών  $X_1, X_2, \dots, X_q$ . Οι συναρτήσεις αυτές μπορεί να είναι παραμετρικές, ημι-παραμετρικές και μη παραμετρικές. Λόγω των προβλημάτων στην εφαρμογή μη παραμετρικών συναρτήσεων που αναπτύχθηκαν στην ενότητα 5.7 έχει επικρατήσει η χρήση των άλλων δύο μεθόδων. Στη συνέχεια αναπτύσσονται οι παραμετρικές και ημι-παραμετρικές.

### 5.8.1 ΠΑΡΑΜΕΤΡΙΚΕΣ ΣΥΝΑΡΤΗΣΕΙΣ ΕΞΟΜΑΛΥΝΣΗΣ

Μία παραμετρική συνάρτηση, η οποία αποτελεί την πιο συνηθισμένη τεχνική εξομάλυνσης είναι η πολωνυμική  $k^{\text{ου}}$  βαθμού:

$$Y = a_0 + a_1x + a_2x^2 + \dots + a_kx^k + \varepsilon_i \quad [26]$$

Όσο πιο μεγάλος είναι ο βαθμός του πολωνύμου τόσο πιο καλά ακολουθεί τα δεδομένα, δηλαδή αποτυπώνονται καλύτερα οι διακυμάνσεις της σχέσης μεταξύ της  $Y$  και της  $X$ . Όμως ταυτοχρόνως το συνολικό αποτέλεσμα μοιάζει σε μία έντονα κυματιστή εικόνα, η οποία ίσως να μην αποδίδει στο βέλτιστο την πραγματική σχέση που θέλουμε να εκτιμήσουμε. Επιπροσθέτως, για το μοντέλο ο βαθμός του πολωνύμου μεταφράζεται σε βαθμούς ελευθερίας, δηλαδή όσο μεγαλύτερος είναι τόσοι περισσότεροι βαθμοί ελευθερίας χρεώνονται στο μοντέλο. Έτσι, τα πολυώνυμα δεν είναι επαρκή για την αποτύπωση μίας σύνθετης σχέσης.

Μία ευέλικτη τεχνική εξομάλυνσης είναι τα παραμετρικά πολυώνυμα παρεμβολής. Τα παραμετρικά πολυώνυμα παρεμβολής αποτελούν τμηματικά πολυώνυμα που ενώνονται ομαλά σε ορισμένα σημεία, τους κόμβους. Στα σημεία αυτά τα πολυώνυμα παρεμβολής, η  $1^{\text{η}}$  και η  $2^{\text{η}}$  παράγωγός τους πρέπει είναι συνεχή. Ο βαθμός εξομάλυνσης εξαρτάται από τους βαθμούς ελευθερίας ( $\beta_e$ ) που καθορίζονται από τον αριθμό των κόμβων ( $\beta_e = \text{κόμβοι} - 1$ ). Η επιλογή του αριθμού των κόμβων και η θέση τους είναι πολύ σημαντικές γιατί επηρεάζουν τις εκτιμήσεις.

Ας υποθέσουμε ότι η  $S(x)$  είναι ένα κυβικό πολυώνυμο και το ορίζουμε σε ένα διάστημα  $[\alpha, \beta]$ . Θεωρούμε διαδοχικά τα σημεία  $\alpha = t_0 < t_1 < \dots < t_n = \beta$  και τις τμηματικές συναρτήσεις από τις οποίες αποτελείται:

$$S(x) = \begin{cases} S_1(x) & \text{αν } t_0 \leq x < t_1 \\ S_2(x) & \text{αν } t_1 \leq x < t_2 \\ \vdots & \\ S_{n-1}(x) & \text{αν } t_{n-1} \leq x < t_n \end{cases} \quad [27]$$

όπου

$$S(x) = \begin{cases} S_1(x) = \alpha_1(x - x_1)^3 + b_1(x - x_1)^2 + c_1(x - x_1) + d_1 \\ S_2(x) = \alpha_2(x - x_2)^3 + b_2(x - x_2)^2 + c_2(x - x_2) + d_2 \\ \vdots \\ S_{n-1}(x) = \alpha_{n-1}(x - x_n)^3 + b_{n-1}(x - x_n)^2 + c_{n-1}(x - x_n) + d_{n-1} \end{cases} \quad [28]$$

Η  $S(x)$  παρεμβάλλεται από τα σημεία  $t_0, t_1, \dots, t_n$  και είναι συνεχής στο κλειστό διάστημα  $[t_0, t_n]$  ενώ πρώτη και η δεύτερη παράγωγός της  $S'(x), S''(x)$  είναι συνεχής στο ανοιχτό  $(t_0, t_n)$ . Έτσι προκύπτει ένα σύστημα  $4n-2$  εξισώσεων με  $4n$  αγνώστους.

Αν θεωρήσουμε επιπλέον ότι είναι φυσικό πολυώνυμο παρεμβολής τότε θα ισχύει ότι  $S''(t_0) = S''(t_n) = 0$ , συνεπώς το διάστημα εκτός του  $[t_0, t_n]$  ορίζεται με μία ευθεία. (Arne Morten Kvarving)

### 5.8.2 ΠΟΛΥΩΝΥΜΑ ΠΑΡΕΜΒΟΛΗΣ ΜΕ ΠΕΡΙΟΡΙΣΜΟ

Αρκετές φορές είναι απαραίτητο να χρησιμοποιήσουμε μία συνάρτηση εξομάλυνσης της οποίας η μορφή δεν είναι αυστηρά προκαθορισμένη, όπως στον έλεγχο της εποχικότητας και των μακροχρόνιων τάσεων στις χρονοσειρές. Έστω ένα σύνολο  $m$  συναρτήσεων με συγκεκριμένες ιδιότητες και γνωστές παραμέτρους  $b_j, j = 1, 2, \dots, m$ . Το σύνολο αυτό ονομάζεται βάση και τα στοιχεία που το αποτελούν συναρτήσεις βάσης. Κάθε συνάρτηση εξομάλυνσης μπορούμε να την εκφράσουμε ως ένα άθροισμα κάποιων συναρτήσεων βάσης, δηλαδή

$$f(x) = \sum_{j=1}^m b_j(x)a_j \quad [29]$$

όπου  $a_j, j = 1, 2, \dots, m$  άγνωστοι συντελεστές. Οι συναρτήσεις βάση επιλέγονται ανάλογα με τις θεωρητικές τους ιδιότητες και τους πρακτικούς σκοπούς που εξυπηρετούν. Ας υποθέσουμε για παράδειγμα ως συναρτήσεις βάσεις τα κυβικά πολυώνυμα παρεμβολής, τότε η συνάρτηση  $f$  γράφεται:

$$f(x) = \sum_{j=1}^{m+2} b_j(x)a_j \quad [30]$$

$$\text{με} \quad b_j(x) = \begin{cases} |x - x_j^*|^3, & j = 1, 2, \dots, m \\ 1, & j = m + 1 \\ x, & j = m + 2 \end{cases} \quad [31]$$

οι συναρτήσεις βάσης και  $x_j^*$  τα σημεία των κόμβων. Οι  $f(x), f'(x), f''(x)$  είναι συνεχείς στους κόμβους, όπου  $f'(x), f''(x)$  η πρώτη και η δεύτερη παράγωγος της  $f$ . Ισχύει ότι

$$\sum_{j=1}^m a_j = 0 \quad \text{και} \quad \sum_{j=1}^m a_j x_j = 0 \quad [32, 33]$$

Αυτό σημαίνει ότι  $f''(x) = 0$ , αν  $x$  δεν ανήκει στο διάστημα  $[x_1, x_2]$ .

Στα πολυώνυμα παρεμβολής με περιορισμό είναι δυνατόν να χρησιμοποιηθεί ένας μεγάλος αριθμός κόμβων και κατά συνέπεια ένας μεγάλος αριθμός συναρτήσεων βάσης, ο οποίος όμως δεν επηρεάζει τις εκτιμήσεις των μοντέλων. Αυτό επιτυγχάνεται θέτοντας περιορισμούς (penalties) στη συνάρτηση Πιθανοφάνειας, διαδικασία στην οποία εκτιμώνται οι παράμετροι του μοντέλου. Αυτό έχει σαν αποτέλεσμα το μοντέλο να ακολουθεί καλά τα δεδομένα και να είναι ομαλό.

Η εφαρμογή μεγάλου αριθμού κόμβων κάνει τη γραφική παράσταση της  $f$  να είναι ιδιαίτερα κυματιστή. Ένα μέτρο της καμπυλότητας είναι

$$J(f) = \int [f''(x)]^2 dx \quad [34]$$

Θεωρώντας μία δεδομένη συνάρτηση  $f$  με συγκεκριμένη συνάρτηση βάσης ισχύει ότι

$$f''(x) = \sum_{j=1}^m a_j b''_j(x) = \mathbf{b}''(x)^T \mathbf{a} \quad [35]$$

και

$$f''(x) = \mathbf{a}^T \mathbf{b}''(x)^T \mathbf{b}''(x) \mathbf{a} = \mathbf{a}^T \mathbf{S}(x) \mathbf{a} \quad [36]$$

όπου

$$\mathbf{S}(x) = \begin{bmatrix} b''_1(x) & b''_1(x)b''_2(x) & b''_1(x)b''_3(x) & \dots & b''_1(x)b''_m(x) \\ b''_2(x)b''_1(x) & b''_2(x) & b''_2(x)b''_3(x) & \dots & b''_2(x)b''_m(x) \\ b''_3(x)b''_1(x) & b''_3(x)b''_2(x) & b''_3(x) & \dots & b''_3(x)b''_m(x) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b''_m(x)b''_1(x) & b''_m(x)b''_2(x) & b''_m(x)b''_3(x) & \dots & b''_m(x) \end{bmatrix} \quad [37]$$

Έτσι

$$J(f) = \mathbf{a}^T \int \mathbf{S}(x) dx \mathbf{a} = \mathbf{a}^T \mathbf{H} \mathbf{a} \quad [38]$$

Επομένως, δεδομένης μίας συνάρτησης βάσης μπορεί να υπολογιστεί ο  $m \times m$  πίνακας  $\mathbf{H}$  που είναι ανεξάρτητος από τις παραμέτρους  $a_j$  και με βάση αυτόν μπορούμε να εκφράσουμε το  $J(f)$  ως τετραγωνική μορφή στο διάνυσμα των παραμέτρων  $\mathbf{a}$  ( $a_j, j = 1, 2, \dots, m$ ). Με αυτόν τον τρόπο εκτιμάται ο περιορισμός της «κύμανσης» (wiggleness) (Wood 2009; Αναστασίου, διπλωματική εργασία).

Στα πολυώνυμα παρεμβολής με περιορισμό χρησιμοποιείται στην αρχή ένας μεγάλος αριθμός κόμβων τότε προστίθεται ένας περιορισμός και στη συνέχεια το ομαλοποιεί. Η μέθοδος αυτή χρησιμοποιεί μία παράμετρο  $\lambda$ , την παράμετρο εξομάλυνσης που καθορίζει το πόσο θα περιοριστεί η επιρροή των κόμβων. Αυτό σημαίνει ότι η θέση και ο αριθμός των κόμβων δεν έχουν μεγάλη σημασία. Έτσι, από την παράμετρο εξομάλυνσης καθορίζεται πόσο εξομαλυμένο θα είναι το τελικό αποτέλεσμα. Όσο μεγαλύτερη είναι η τιμή του  $\lambda$ , τόσο πιο εξομαλυμένη είναι η τελική καμπύλη. Η παράμετρος εξομάλυνσης συνδέεται άμεσα με τους βαθμούς ελευθερίας της εξομαλυμένης καμπύλης γραμμής. Όσο αυξάνεται το  $\lambda$ , οι βαθμοί ελευθερίας ελαττώνονται.

Το μοντέλο  $E(y_i) = f(x_i)$  όπου η  $f$  είναι μία συνάρτηση εξομάλυνσης μπορεί να εκτιμηθεί ελαχιστοποιώντας την ποσότητα

$$\sum_{i=1}^n [f(x_i) - y_i]^2 + \lambda \int [f''(x)]^2 dx \quad [39]$$

όπου  $\lambda$  η παράμετρος εξομάλυνσης. Επιλέγοντας μία συνάρτηση βάσης για την  $f$ , μπορούν να καθοριστούν ο πίνακας σχεδιασμού (design matrix), ο οποίος έχει τη μορφή:

$$\mathbf{X} = \begin{bmatrix} b_1(x_1) & b_2(x_1) & \dots & b_m(x_1) \\ b_1(x_2) & b_2(x_2) & \dots & b_m(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ b_1(x_n) & b_2(x_n) & \dots & b_m(x_n) \end{bmatrix} = \begin{bmatrix} b(x_1)^T \\ b(x_2)^T \\ \vdots \\ b(x_n)^T \end{bmatrix} \quad [40]$$

και ο πίνακας των περιορισμών (penalty matrix)  $\mathbf{H}$ . Έτσι η παραπάνω σχέση γράφεται:

$$\min \|\mathbf{X}\mathbf{a} - \mathbf{y}\|^2 + \lambda \mathbf{a}^T \mathbf{H} \mathbf{a} \quad [41]$$

όπου  $\|\cdot\|^2$  η Ευκλείδεια νόρμα. Αν η παράμετρος  $\lambda$  είναι γνωστή τότε,



$$\hat{\mathbf{a}} = [\mathbf{X}^T \mathbf{X} + \lambda \mathbf{H}]^{-1} \mathbf{X}^T \mathbf{y} \quad [42]$$

Η παράμετρος εξομάλυνσης μπορεί να εκτιμηθεί χρησιμοποιώντας διάφορες μεθόδους, όπως

- Τη Γενίκευση Διασταυρωμένης Επικύρωσης (GCV)
- Τη Συνάρτηση Μερικής Αυτοσυσχέτισης (PACF)
- Το κριτήριο BIC/AIC
- Τη Συνάρτηση Πιθανοφάνειας (ML/REML)
- Να προκαθοριστεί βάση της βιβλιογραφίας

Ανάλογα με τη μέθοδο που θα ακολουθήσουμε το τελικό αποτέλεσμα διαφέρει.

Η παραπάνω διαδικασία εκτίμησης είναι αρκετά γενική και μπορεί να εφαρμοστεί σε πληθώρα συναρτήσεων βάσης και περιορισμών. Ωστόσο, η συνάρτηση που ελαχιστοποιεί την [40], ανάμεσα σε τις συναρτήσεις και όχι μόνο αυτές που μπορούν να αναπαρασταθούν χρησιμοποιώντας συγκεκριμένα σύνολα συναρτήσεων βάσης, είναι ένα φυσικό κυβικό πολυώνυμο παρεμβολής με ένα κόμβο για κάθε τιμή  $X_i$ . Αυτό το αποτέλεσμα μπορεί να γενικευθεί για περιορισμούς διαφορετικής τάξης και για συναρτήσεις εξομάλυνσης με οποιοδήποτε αριθμό μεταβλητών και είναι αυτή η ιδιότητα που αποδεικνύει τις βάσεις με πολυώνυμα παρεμβολής ως μία φυσική επιλογή για την αναπαράσταση συναρτήσεων εξομάλυνσης.

Όταν χρησιμοποιούνται πολυώνυμα παρεμβολής με περιορισμό, οι παράμετροι του μοντέλου εκτιμώνται από το Δεσμευμένο Άθροισμα των Τετραγώνων των Καταλοίπων όταν τα μοντέλα είναι προσθετικά (additive models). Στην περίπτωση των επιδημιολογικών μελετών θνησιμότητας, όπου τα δεδομένα ακολουθούν κατανομή Poisson και τα μοντέλα είναι γενικευμένα προσθετικά (general additive models), τότε εκτιμώνται από τη δεσμευμένη λογαριθμική συνάρτηση Πιθανοφάνειας, έστω  $l(\mathbf{a})$ . Τότε η αρνητική δεσμευμένη λογαριθμική συνάρτηση Πιθανοφάνειας ελαχιστοποιείται και η [40] παίρνει τη μορφή:

$$-l(\mathbf{a}) - \frac{1}{2} \lambda \mathbf{a}^T \mathbf{H} \mathbf{a} \quad [43]$$

## 5.9 ΚΡΙΤΗΡΙΑ ΕΠΙΛΟΓΗΣ ΜΟΝΤΕΛΟΥ

Όταν υποθέτουμε κάποιο στατιστικό μοντέλο, σκοπός μας είναι να περιγράψουμε τη σχέση μεταξύ των μεταβλητών με τον καλύτερο δυνατό τρόπο. Για την επιλογή του βέλτιστου μοντέλου υπάρχουν κριτήρια καλής προσαρμογής, όπως το κριτήριο συνάρτησης μερικής αυτοσυσχέτισης (PACF) και το κριτήριο πληροφορίας του AIC, τα οποία θα αναπτυχθούν στη συνέχεια.

### 5.9.1 ΚΡΙΤΗΡΙΟ ΣΥΝΑΡΤΗΣΗΣ ΜΕΡΙΚΗΣ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ (PACF)

Σύμφωνα με το κριτήριο της συνάρτησης μερικής αυτοσυσχέτισης το καλύτερο μοντέλο είναι εκείνο που ελαχιστοποιεί κατά απόλυτη τιμή το άθροισμα των μερικών αυτοσυσχετίσεων των καταλοίπων του μοντέλου για χρονική υστέρηση ορισμένων ημερών. Με αυτόν τον τρόπο αποφεύγονται θέματα υπερ-εξομάλυνσης σε περίπτωση που εφαρμοστούν συναρτήσεις εξομάλυνσης στο μοντέλο και ανεπαρκούς ελέγχου πιθανών συγχυτικών παραγόντων, ενώ ταυτόχρονα ελαχιστοποιείται η αυτοσυσχέτιση των υπολοίπων που είναι εγγενής στις χρονοσειρές και παραβιάζει τις προϋποθέσεις εφαρμογής του μοντέλου.

### 5.9.2 ΚΡΙΤΗΡΙΟ ΠΛΗΡΟΦΟΡΙΑΣ ΑΚΑΙΚΕ (AIC)

Το AIC αποτελεί ένα μέτρο καλής προσαρμογής ενός στατιστικού μοντέλου. Τα στατιστικά μοντέλα κατατάσσονται βάσει του AIC τους, με βέλτιστο αυτό που έχει την μικρότερη τιμή. Η τιμή του AIC για κάθε μοντέλο δίνεται από τον τύπο:

$$AIC = -2\log L_{max} + 2k \quad [44]$$

Όπου  $k$  ο αριθμός των παραμέτρων και  $L_{max}$  η μέγιστη τιμή της συνάρτησης Πιθανοφάνειας του μοντέλου. Ο πρώτος όρος ερμηνεύεται σαν ένα μέτρο καλής προσαρμογής και ο δεύτερος ως μία ποινή (penalty) για κάθε μία παράμετρο που χρειάζεται να εκτιμηθεί.

Στα μοντέλα που είναι προσαρμοσμένα για την ύπαρξη υπερ-διασποράς η σχέση [43] γίνεται

$$AIC = -2\log U_{max} + 2k\hat{\phi} \quad [45]$$

όπου  $U_{max}$  η μέγιστη τιμή της Ημι-Πιθανοφάνειας και  $\hat{\phi}$  η εκτίμηση της παραμέτρου υπερ-διασποράς.

## ΕΙΔΙΚΟ ΜΕΡΟΣ

### 6 ΠΡΟΒΛΕΨΗ ΤΩΝ ΣΥΓΚΕΝΤΡΩΣΕΩΝ ΤΩΝ $PM_{10}$ ΤΗΝ ΠΕΡΙΟΔΟ 1992-2000

---

Οι μετρήσεις των συγκεντρώσεων των  $PM_{10}$  για τη χρονική περίοδο πριν το 2001 δεν υπήρχαν, οπότε για να αυξήσουμε το χρονικό διάστημα μελέτης των διαχρονικών τάσεων των επιδράσεων εκτιμήσαμε τις συγκεντρώσεις για την περίοδο 1992-2000. Για την εκτίμηση χρησιμοποιήθηκε ένα μοντέλο πολλαπλής γραμμικής παλινδρόμησης βασισμένο σε μετρήσεις των επιπέδων του μαύρου καπνού, ο οποίος αποτελεί μέρος των αιωρούμενων σωματιδίων και μετριόταν από το ΥΠΕΚΑ από το 1992-2004.

Αρχικά συλλέχθηκαν δεδομένα για το 2001-2004, περίοδο κατά την οποία οι σταθμοί κατέγραφαν ταυτοχρόνως τις συγκεντρώσεις των  $PM_{10}$  και του μαύρου καπνού, τα οποία χρησιμοποιήθηκαν για την κατασκευή ενός αξιόπιστου μοντέλου γραμμικής παλινδρόμησης. Στη συνέχεια βάση του επιλεγμένου στατιστικού αυτού μοντέλου εκτιμήθηκαν οι μετρήσεις των  $PM_{10}$  για τα έτη 1992-2000.

Η ανάλυση περιορίστηκε στις ημέρες με συγκεντρώσεις του μαύρου καπνού και των  $PM_{10}$  κάτω από  $150\mu\text{g}/\text{m}^3$ .

#### 6.1 ΜΟΝΤΕΛΟ ΠΟΛΛΑΠΛΗΣ ΓΡΑΜΜΙΚΗΣ ΕΞΑΡΤΗΣΗΣ

Στο μοντέλο για την πρόβλεψη αρχικά χρησιμοποιήθηκαν τα επίπεδα του μαύρου καπνού (Katsougianni et al. (2009); David Muir & Duncan P.H. Laxen (1995)), ενώ σταδιακά προστέθηκαν οι μετεωρολογικοί παράγοντες και συγκεκριμένα η θερμοκρασία, η σχετική υγρασία, η βαρομετρική πίεση, η ταχύτητα και η διεύθυνση του ανέμου, ενώ τέλος προστέθηκε το Διοξείδιο του Αζώτου ( $NO_2$ ). Στο μοντέλο πρόγνωσης προστέθηκε έλεγχος για την εποχικότητα και την ημέρα της εβδομάδας (καθημερινή-Σαββατοκύριακο) (Shahraiyini, Sodoudi, (2015); Chaloulakou, Grivas, Spyrellis (2012))

Η επιλογή όλων των όρων έγινε με βάση την τιμή του  $R_{adj}^2$ , που αποτελεί τη μεταβλητότητα της συγκέντρωσης των αιωρούμενων σωματιδίων που ερμηνεύεται από το μοντέλο. Όσο πιο υψηλή είναι η τιμή του τόσο πιο μεγάλο μέρος της μεταβλητότητας ερμηνεύεται από το μοντέλο.

Τα βήματα για τη διαμόρφωση του μοντέλου πρόβλεψης παρουσιάζονται στον Πίνακα 6.1 και αναλύονται στη συνέχεια.

**Πίνακας 6.1:** Προσαρμοσμένο  $R^2$  για τα μοντέλα πρόγνωσης των επιπέδων  $PM_{10}$

Πολλαπλή Γραμμική Παλινδρόμηση		
Βήματα	Μεταβλητές	$R^2_{adj}$
1	Μαύρος Καπνός	26.22%
2	+ εποχικότητα	47.28%
3	+ θερμοκρασία	55.97%
4	+ σχετική υγρασία	57.52%
5	+ ταχύτητα του ανέμου	58.97%
6	+ βαρομετρική πίεση, διεύθυνση του ανέμου, ημέρα της εβδομάδας	60.07%
7	+ Διοξείδιο του Αζώτου	62.80%

Αναλυτικότερα,

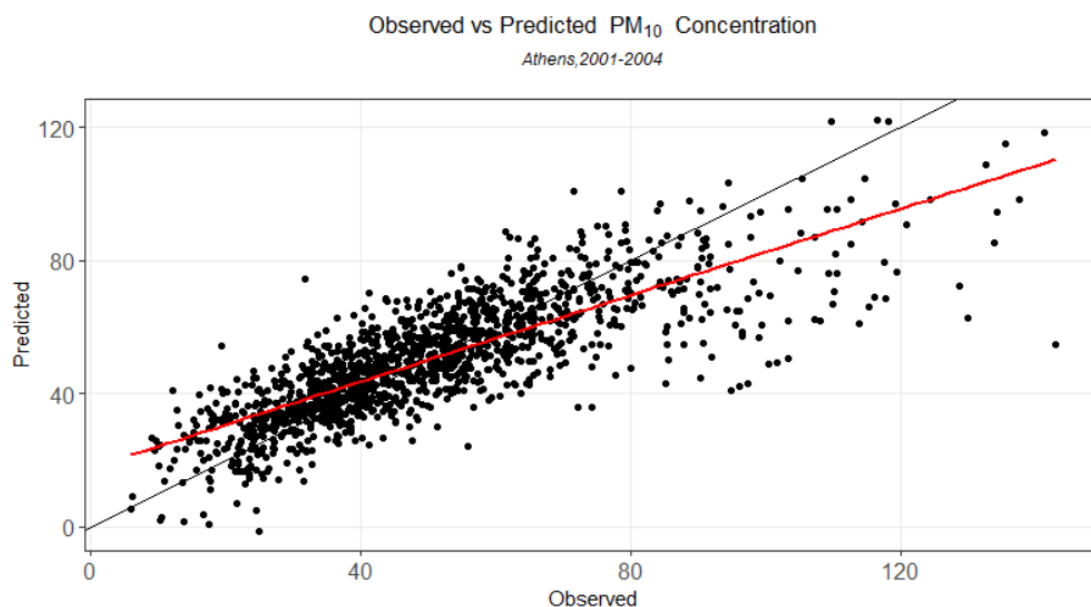
- Ο Μαύρος Καπνός εισάχθηκε ως γραμμικός όρος λόγω της γραμμικής του σχέσης με τη συγκέντρωση των  $PM_{10}$ .
- Για τον έλεγχο της εποχικότητας χρησιμοποιήθηκαν πολυώνυμα παρεμβολής (natural splines) με 6 βαθμούς ελευθερίας σε κάθε έτος. Οι επιλογή των βαθμών ελευθερίας έγινε βάση των προβλέψεων των  $PM_{10}$  ώστε να μην αποκλίνουν από το εύρος των πραγματικών τιμών. Τα πολυώνυμα παρεμβολής προτιμήθηκαν έναντι τριγωνομετρικών συναρτήσεων και διπλής ή τριπλής αλληλεπίδρασης του έτους, του μήνα και της ημέρας της εβδομάδας λόγω της ευελιξίας τους αλλά κυρίως γιατί κατέστησαν δυνατή την πρόβλεψη των τιμών  $PM_{10}$  εκτός της χρονικής περιόδου 2001-2004.
- Για τον έλεγχο της θερμοκρασίας χρησιμοποιήθηκαν πολυώνυμα παρεμβολής με 6 βαθμούς ελευθερίας. Η επιλογή των βαθμών ελευθερίας έγινε κι αυτή βάση των προβλέψεων. Τα πολυώνυμα παρεμβολής επιλέχθηκαν αντί των πολυωνύμων 3<sup>ου</sup> βαθμού.
- Η σχετική υγρασία εισάχθηκε στο μοντέλο σαν τετραγωνικός όρος έναντι του γραμμικού.
- Η ταχύτητα του ανέμου προστέθηκε σαν κυβικός όρος (πολυώνυμο 3<sup>ου</sup>) έναντι του αντίστροφου μετασχηματισμού του.
- Η βαρομετρική πίεση, η διεύθυνση του ανέμου και το Διοξείδιο του Αζώτου εισάχθηκαν ως γραμμικοί όροι.

Επίσης, πραγματοποιήθηκαν έλεγχοι χρησιμοποιώντας τις μετρήσεις του Μαύρου Καπνού και του  $NO_2$  με χρονοϋστέρηση μίας ημέρα.

Η μορφή της συνάρτησης είναι:

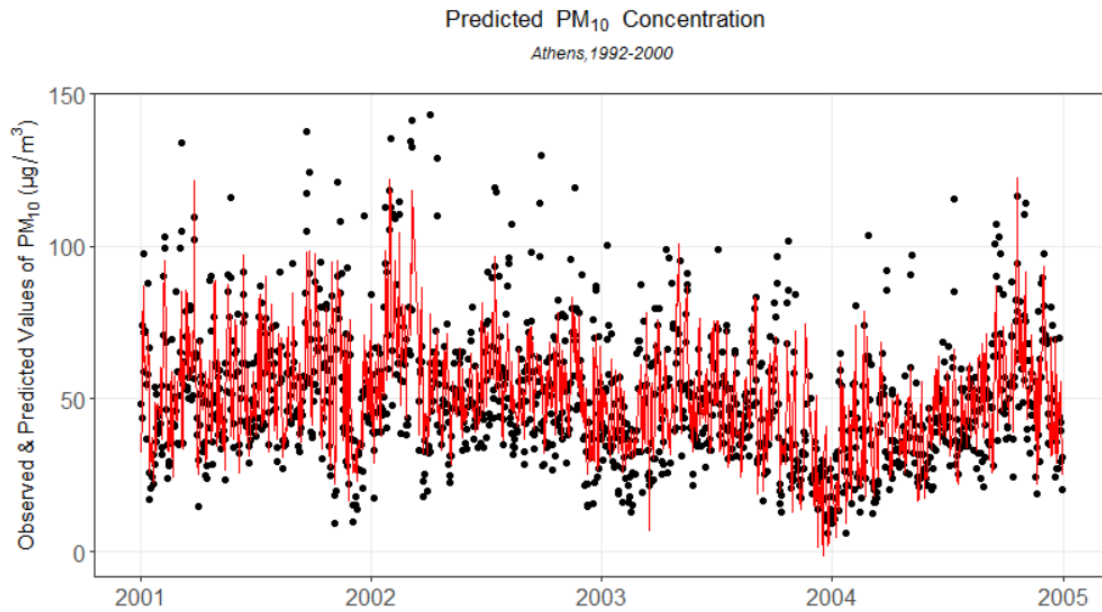
$$\begin{aligned}
 E[Y] = & b_0 + b_1 BS + ns \left( period_{2001-2004}, \frac{length\ period}{365} * 6 \right) \\
 & + ns(temperature, df = 6) \\
 & + \sum_{i=1}^2 c_i relative\ humidity + \sum_{j=1}^3 d_j wind\ speed + b_2 dow \\
 & + b_3 wind\ direction + b_4 barometric\ pressure \\
 & + b_5 NO_2
 \end{aligned}
 \tag{46}$$

Τα αποτελέσματα του μοντέλου παρουσιάζονται γραφικά στα παρακάτω διαγράμματα 1 και 2. Παρατηρούμε στο **Γράφημα 5** ότι η σχέση μεταξύ των πραγματικών και των προβλεπόμενων τιμών των  $PM_{10}$  είναι γραμμική και οι τιμές παρουσιάζουν μία καλή συμφωνία. Στα επίπεδα  $0 - 80 \mu\text{g}/\text{m}^3$  συγκεντρώνεται το μεγαλύτερο πλήθος των τιμών ενώ άνω των  $80 \mu\text{g}/\text{m}^3$  οι τιμές είναι διάσπαρτες και λίγες.



**Γράφημα 5** Στικτόγραμμα των πραγματικών (*observed*) και των προβλεπόμενων (*predicted*) τιμών της συγκέντρωσης των επιπέδων  $PM_{10}$ . Ο *x*-άξονας συμβολίζει τις πραγματικές τιμές και ο *y*-άξονας τις προβλεπόμενες. Η κόκκινη γραμμή ορίζει την ευθεία παλινδρόμησης.

Στο **Γράφημα 6** απεικονίζονται οι παρατηρούμενες και οι προβλεπόμενες τιμές των  $PM_{10}$  στη διάρκεια των τεσσάρων χρόνων 2001-2004. Παρατηρούμε ότι το στατιστικό μοντέλο γραμμικής παλινδρόμησης ακολουθεί ικανοποιητικά την εποχικότητα και τη μεταβλητότητα των πραγματικών τιμών αλλά καθίσταται ασθενές στις ακραίες τιμές.



**Γράφημα 6** Στικτόγραμμα και γραμμικό διάγραμμα των πραγματικών και των προβλεπόμενων τιμών της συγκέντρωσης των  $PM_{10}$  ( $\mu\text{g}/\text{m}^3$ ) αντίστοιχα, για την περίοδο 2001-2004.

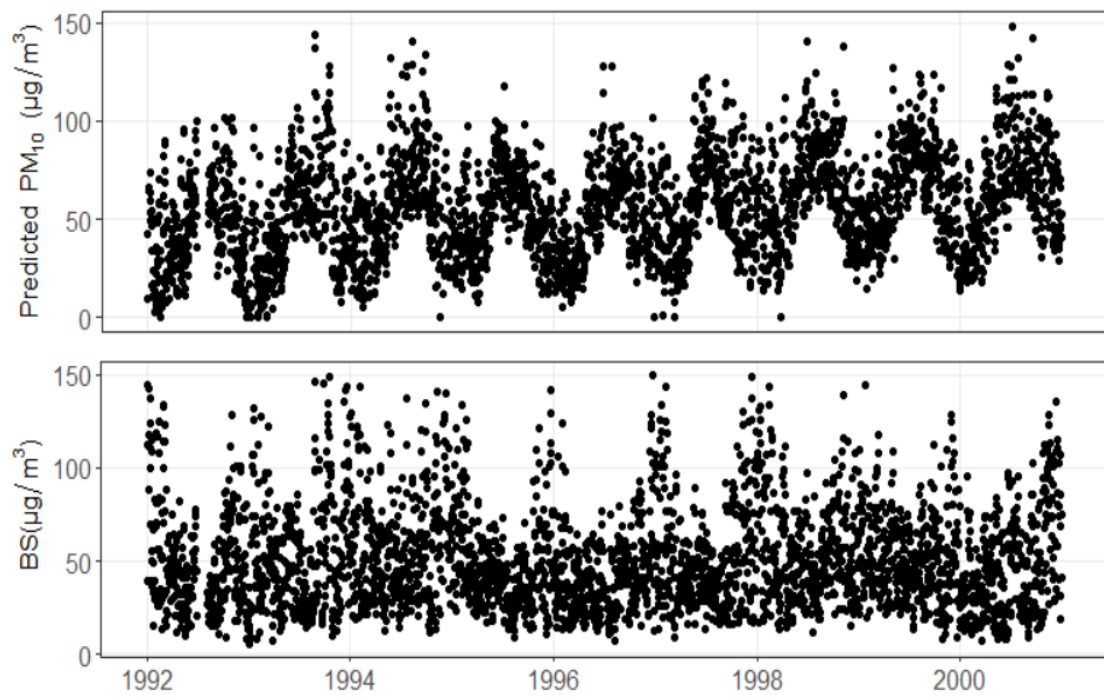
## 6.2 ΠΡΟΒΛΕΨΗ ΤΩΝ ΣΥΓΚΕΝΤΡΩΣΕΩΝ ΤΩΝ PM<sub>10</sub> ΤΗΝ ΠΕΡΙΟΔΟ 1992-2000

Το επόμενο βήμα ύστερα από την κατασκευή του στατιστικού μοντέλου είναι η πρόβλεψη των τιμών της συγκέντρωσης των PM<sub>10</sub> για την περίοδο που δεν υπήρχαν μετρήσεις, δηλαδή πριν το 2001. Συλλέχθηκαν τα δεδομένα από το 1992-2000 των παραμέτρων του μοντέλου και συγκεκριμένα των μετεωρολογικών μεταβλητών (θερμοκρασία, σχετική υγρασία, βαρομετρική πίεση, ταχύτητα και διεύθυνση του ανέμου), των ρύπων (μαύρος καπνός, NO<sub>2</sub>) και εφαρμόστηκε σε αυτά το στατιστικό μοντέλο. Ο Πίνακας 6.2 περιλαμβάνει τα περιγραφικά μέτρα θέσης και διασποράς των προβλεπόμενων τιμών PM<sub>10</sub> και του μαύρου καπνού που αποτελεί το βασικό προγνωστικό παράγοντα.

**Πίνακας 6.2:** Περιγραφικά μέτρα θέσης και διασποράς των προβλεπόμενων τιμών των PM<sub>10</sub>, 1992-2000

Μεταβλητές	Μέσος ± Τυπική Απόκλιση	Ελάχιστη τιμή	Εκατοστημόρια			Μέγιστη τιμή
			25 <sup>ο</sup>	50 <sup>ο</sup>	75 <sup>ο</sup>	
Προβλεπόμενες τιμές (PM <sub>10</sub> )	56.23 ± 25.97	-45.27	36.68	55.68	73.55	177.90
Μαύρος Καπνός (BS)	50.52 ± 26.87	5.96	30.25	45	65	149.25

Η μέση τιμή και η διάμεσος των προβλεπόμενων τιμών των PM<sub>10</sub> είναι 56.23 μg/m<sup>3</sup> και 55.68 μg/m<sup>3</sup> αντίστοιχα. Παρατηρούμε ότι η ελάχιστη τιμή τους είναι αρνητική, το στατιστικό μοντέλο έδωσε εκτός αυτής, άλλες 15 αρνητικές τιμές, τις οποίες μετατρέψαμε σε 0.1 μg/m<sup>3</sup>. Στο **Γράφημα 7** απεικονίζονται γραφικά οι ημερήσιες μετρήσεις των δύο ρύπων. Συγκρίνοντας τα δύο γραφήματα παρατηρούμε ότι η μεταβλητότητα των εκτιμήσεων αντικατοπτρίζει τη μεταβλητότητα των τιμών του μαύρου καπνού. Το 1992 παρουσιάζεται, όπως φαίνεται, μία πτωτική τάση των επιπέδων της συγκέντρωσής τους στην ατμόσφαιρα, αλλά το 1993-1995 και από το 1997 και μετά παρουσιάζεται μία αυξητική τάση.



*Γράφημα 7* Στικτόγραμμα των ημερήσιων προβλεπόμενων επιπέδων της συγκέντρωσης των  $PM_{10}$  ( $\mu\text{g}/\text{m}^3$ ) (πάνω) και των πραγματικών τιμών του μαύρου καπνού  $BS$  ( $\mu\text{g}/\text{m}^3$ ) (κάτω) για τη χρονική περίοδο 1992-2000.



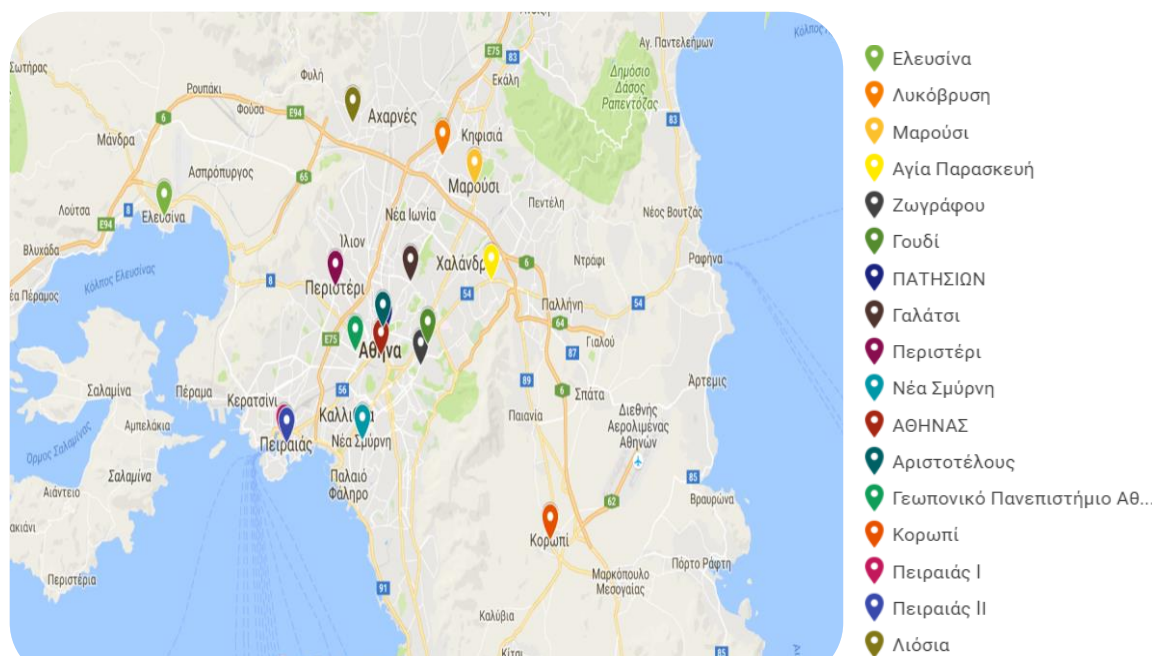
## 7 ΕΠΙΔΗΜΙΟΛΟΓΙΚΗ ΑΝΑΛΥΣΗ ΓΙΑ ΤΗ ΔΙΕΡΕΥΝΗΣΗ ΤΩΝ ΕΠΙΔΡΑΣΕΩΝ ΤΩΝ ΣΩΜΑΤΙΔΙΩΝ ΣΤΗΝ ΗΜΕΡΗΣΙΑ ΘΝΗΣΙΜΟΤΗΤΑ

Εφαρμόστηκαν μοντέλα Poisson προσαρμοσμένα για ύπαρξη υπερ-διασποράς και με έλεγχο για πιθανούς συγχυτικούς παράγοντες. Η επιδημιολογική ανάλυση περιορίστηκε στις ημέρες με συγκεντρώσεις των  $PM_{10}$  κάτω από  $150\mu g/m^3$  και πραγματοποιήθηκε για το σύνολο της χρονικής περιόδου 1992-2012, αλλά και χωριστά για την περίοδο με μετρήσεις 2001-2012 και την περίοδο με προβλέψεις των συγκεντρώσεων των  $PM_{10}$  1992-2000. Στη συνέχεια διερευνήθηκε η διαχρονική τάση των επιδράσεων με την προσθήκη όρου αλληλεπίδρασης μεταξύ του χρόνου και του ρύπου.

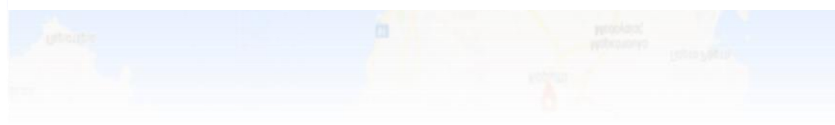
### 7.1 ΔΕΔΟΜΕΝΑ

#### ΑΤΜΟΣΦΑΙΡΙΚΟΙ ΡΥΠΟΙ- ΜΕΤΕΩΡΟΛΟΓΙΚΟΙ ΠΑΡΑΓΟΝΤΕΣ

Για την επιδημιολογική ανάλυση των δεδομένων χρονοσειρών χρησιμοποιήθηκαν οι μετρήσεις των ρύπων που καταγράφονται από ειδικούς σταθμούς του ΥΠΕΚΑ (Γράφημα 8). Το δίκτυο αποτελείται από μετρητικές διατάξεις οι οποίες έχουν τοποθετηθεί σε διάφορες περιοχές της πρωτεύουσας σύμφωνα με τις οδηγίες της Ευρωπαϊκής Ένωσης (Πίνακας 7.1).



Γράφημα 8: Σταθμοί Μέτρησης των Ατμοσφαιρικών Ρύπων υπό την ευθύνη του ΥΠΕΚΑ



**Πίνακας 7.1:** Σταθμοί Μέτρησης των Ατμοσφαιρικών Ρύπων υπό την ευθύνη του ΥΠΕΚΑ

Θέση	Σταθμοί Χαρακτηρισμός	Μετρούμενοι Ρύπο				
		PM <sub>10</sub>	SO <sub>2</sub>	NO <sub>2</sub>	O <sub>3</sub>	CO
Αθηνάς	Κυκλοφορίας Αστικός		✓	✓	✓	✓
Αριστοτέλους	Κυκλοφορίας Αστικός	✓		✓		
Γεωπονική	Περιαστικός- Βιομηχανικός			✓	✓	✓
Λιόσια	Υποβάθρου Περιαστικός			✓	✓	
Λυκόβρυση	Περιαστικός	✓		✓	✓	
Μαρούσι	Κυκλοφορίας Αστικός	✓		✓	✓	✓
Νέα Σμύρνη	Υποβάθρου Αστικός			✓	✓	✓
Πατησίων	Κυκλοφορίας Αστικός		✓	✓	✓	✓
Πειραιάς I	Κυκλοφορίας Αστικός	✓	✓	✓	✓	✓
Περιστερί	Υποβάθρου Αστικός		✓	✓	✓	✓
Αγ. Παρασκευή	Υποβάθρου Περιαστικός	✓		✓	✓	
Ελευσίνα	Περιαστικός Βιομηχανικός	✓	✓	✓	✓	
Κορωπί	Υποβάθρου Περιαστικός	✓		✓	✓	
<b>Παλαιότεροι Σταθμοί</b>						
Ζωγράφου	Υποβάθρου Περιαστικός	✓		✓	✓	
Γουδί	Κυκλοφορίας Αστικός	✓		✓		
Γαλάτσι	Υποβάθρου Περιαστικός		✓	✓	✓	
Πειραιάς II	Υποβάθρου Αστικός		✓	✓	✓	

Για την επιδημιολογική ανάλυση επιλέχθηκαν οι σταθμοί που χαρακτηρίζονται ως αστικοί ή κυκλοφορίας ώστε να είναι αντιπροσωπευτικοί της έκθεσης του αστικού πληθυσμού.

Η ημερήσια συγκέντρωση κάθε ρύπου υπολογίστηκε ως μέσος όρος των συγκεντρώσεων από τις μετρήσεις των σταθμών που ικανοποιούσαν το κριτήριο πληρότητας των 75% καταγεγραμμένων τιμών για όλη την περίοδο της μελέτης. Οι ελλείπουσες τιμές των χρονοσειρών αντικαταστάθηκαν σύμφωνα με τον ακόλουθο τύπο:

$$\hat{x}_{ijk} = \bar{x}_{i.k} \frac{\bar{x}_{.jk}}{\bar{x}_{..k}}$$

Μία ελλείπουσα τιμή την ημέρα  $i$ , του έτους  $k$  από το σταθμό  $j$  αντικαθίσταται από ένα σταθμισμένο μέσο όρο των τιμών των άλλων σταθμών, όπου  $\bar{x}_{i.k}$  η μέση τιμή την ημέρα  $i$  του έτους  $k$  όλων των σταθμών,  $\bar{x}_{.jk}$  η μέση τιμή του σταθμού  $j$  για το έτος  $k$  και  $\bar{x}_{..k}$  η συνολική μέση τιμή του έτους  $k$ .

Συλλέχθηκαν δεδομένα για τους ακόλουθους ρύπους:

- Αιωρούμενα σωματίδια διαμέτρου μικρότερης ή ίσης των 10  $\mu\text{m}$  (PM<sub>10</sub>) : ημερήσια μέση
- Διοξείδιο του αζώτου (NO<sub>2</sub>) : ημερήσια μέγιστη τιμή του 1-ωρου κυλιόμενου μέσου
- Όζον (O<sub>3</sub>) : ημερήσια μέγιστη τιμή του 8-ωρου κυλιόμενου μέσου
- Διοξείδιο του θείου (SO<sub>2</sub>) : ημερήσια μέση τιμή

- Μονοξείδιο του άνθρακα (CO) : ημερήσια μέγιστη τιμή του 8-ωρου κυλιόμενου μέσου
- Θερμοκρασία : ημερήσια μέση τιμή
- Σχετική Υγρασία : ημερήσια μέση τιμή

### ΘΝΗΣΙΜΟΤΗΤΑ

Η ημερήσια θνησιμότητα από όλες τις αιτίες, από καρδιαγγειακά ή αναπνευστικά αίτια συλλέχθηκαν από την Ελληνική Στατιστική Υπηρεσία (ΕΛ. ΣΤΑΤ) για τη χρονική περίοδο 1992-2012.

## 7.2 ΠΕΡΙΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ

Οι Πίνακες 7.2-7.4 παρουσιάζουν τα περιγραφικά μέτρα θέσης και διασποράς των ατμοσφαιρικών ρύπων, των μετεωρολογικών παραγόντων και της θνησιμότητας στις περιόδους 1992-2000, 2001-2012 και 1992-2012 αντίστοιχα. Οι περίοδοι διακρίθηκαν με βάση τη διαθεσιμότητα των μετρήσεων των PM<sub>10</sub>. Η πρώτη περίοδος περιλαμβάνει τις προβλεπόμενες από το στατιστικό μοντέλο τιμές, η δεύτερη τις πραγματικές μετρήσεις στους σταθερούς σταθμούς ενώ η τρίτη και τις δύο. Συγκρίνοντας τα διάφορα περιγραφικά μέτρα θέσης και διασποράς των ατμοσφαιρικών ρύπων μεταξύ των δύο περιόδων παρατηρούμε ότι οι τιμές έχουν μειωθεί. Για παράδειγμα η διάμεσος των PM<sub>10</sub> στην πρώτη περίοδο έχει τιμή 55.64 μg/m<sup>3</sup> ενώ στην άλλη 40.25 μg/m<sup>3</sup>. Την ίδια μείωση παρατηρούμε στους υπόλοιπους ατμοσφαιρικούς ρύπους. Κάτι το οποίο πρέπει να τονιστεί είναι η μείωση της διασποράς των συγκεντρώσεών τους. Αυτό μπορούμε να το εντοπίσουμε καλύτερα στο **Γράφημα 9** με τα θηκογράμματα που ακολουθεί.

**Πίνακας 7.2:** Περιγραφικά μέτρα θέσης και διασποράς ατμοσφαιρικών ρύπων, μετεωρολογικών παραγόντων και θνησιμότητας, 2001-2012

Μεταβλητή	Μέρες (n)	Μέσος ± Τυπική Απόκλιση	Ελάχιστη τιμή	Εκατοστημόρια			Μέγιστη τιμή
				25°	50°	75°	
<b>Ρύποι</b>							
PM <sub>10</sub> (μg/m <sup>3</sup> )	4344	44.11 ± 19.46	5.97	30.50	40.25	53.41	146.16
NO <sub>2</sub> (μg/m <sup>3</sup> )	4358	87.35 ± 30.08	23.11	66.22	83.56	103.96	261.71
O <sub>3</sub> (μg/m <sup>3</sup> )	4358	64.49 ± 26.33	4.67	43.20	65.13	85.58	144.03
CO (μg/m <sup>3</sup> )	4358	1.71 ± 0.99	0.32	1.02	1.42	2.10	7.55
SO <sub>2</sub> (μg/m <sup>3</sup> )	4347	14.35 ± 11.02	1.71	6.84	11.06	18.22	80.41
<b>Μετεωρολογικά</b>							
Θερμοκρασία (°C)	4358	18.86 ± 7.60	-6.94	12.9	18.26	25.46	36.38
Σχετική Υγρασία (%)	4358	63.13 ± 15.10	20.25	51.63	64.25	74.25	100
<b>Αριθμός ημερήσιων θανάτων</b>							
Όλες οι αιτίες	4358	80 ± 13	40	71	79	88	128
Καρδιαγγειακή	4358	37 ± 8	14	32	36	42	74
Αναπνευστική	4358	8 ± 3	0	6	8	10	25

**Πίνακας 7.3:** Περιγραφικά μέτρα θέσης και διασποράς ατμοσφαιρικών ρύπων, μετεωρολογικών παραγόντων και θνησιμότητας στην Αθήνα την περίοδο 1992-2000

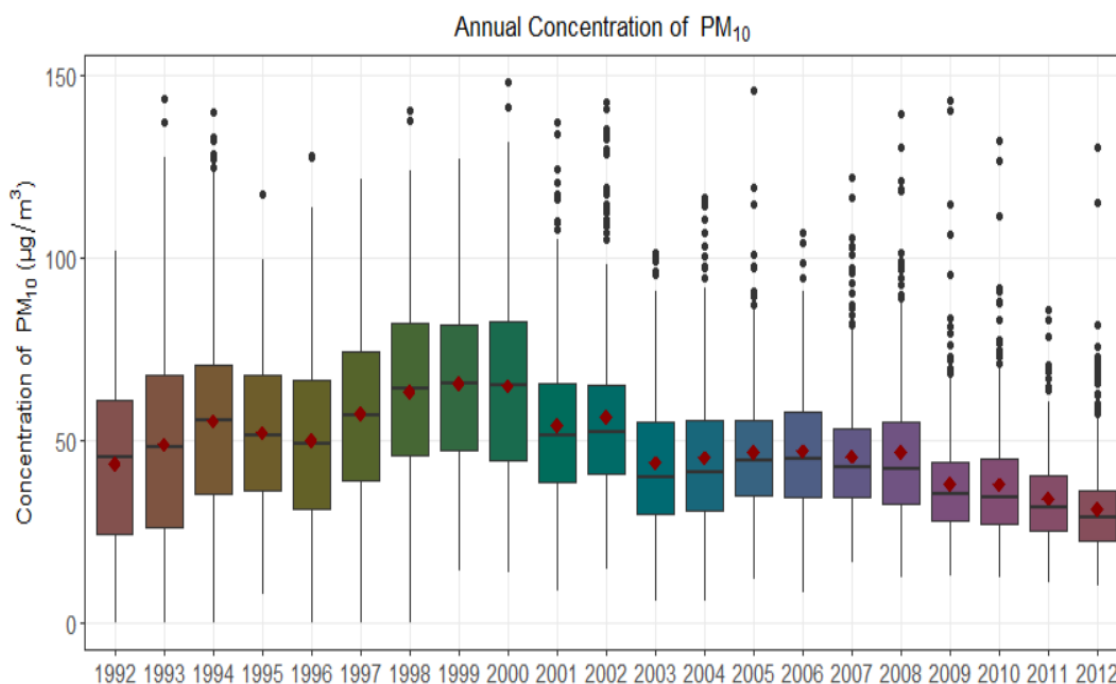
Μεταβλητή	Μέρες (n)	Μέσος ± Τυπική Απόκλιση	Ελάχιστη τιμή	Εκατοστημόρια			Μέγιστη τιμή
				25°	50°	75°	
<b>Ρύποι</b>							
Εκτιμώμενες							
PM <sub>10</sub> (μg/m <sup>3</sup> )	3194	56.05 ± 25.48	0.1	36.66	55.64	73.42	148.25
NO <sub>2</sub> (μg/m <sup>3</sup> )	3278	112.11 ± 42.27	9.86	83	104.54	133.33	314.12
O <sub>3</sub> (μg/m <sup>3</sup> )	3280	66.98 ± 27.85	8.82	43.46	65.22	87.27	178.87
CO (μg/m <sup>3</sup> )	3280	3.79 ± 1.83	0.65	2.51	3.31	4.64	18.41
SO <sub>2</sub> (μg/m <sup>3</sup> )	3275	34.33 ± 23.14	6.18	19.43	29.58	44.02	243.19
<b>Μετεωρολογικά</b>							
Θερμοκρασία (°C)	3268	18.18 ± 7.44	0.3	12.01	17.72	24.93	36.51
Υγρασία (%)	3268	61.85 ± 13.58	21.38	51.63	62.00	72.38	97.88
<b>Αριθμός ημερήσιων θανάτων</b>							
Όλες οι αιτίες	3280	75 ± 14	34	65	74	84	191
Καρδιαγγειακή	3280	37 ± 9	14	31	37	43	111
Αναπνευστική	3280	5 ± 3	0	3	5	7	25

**Πίνακας 7.4:** Περιγραφικά μέτρα θέσης και διασποράς ατμοσφαιρικών ρύπων, μετεωρολογικών παραγόντων και θνησιμότητας στην Αθήνα την περίοδο 1992-2012

Μεταβλητή	Μέρες (n)	Μέσος ± Τυπική Απόκλιση	Ελάχιστη	Εκατοστημόρια			Μέγιστη
				25°	50°	75°	
<b>Ρύποι</b>							
PM <sub>10</sub> (μg/m <sup>3</sup> )	7538	49.17 ± 22.98	0.1	32.13	45.28	62.81	148.25
NO <sub>2</sub> (μg/m <sup>3</sup> )	7638	97.98 ± 37.86	9.86	71.75	91.67	115.78	314.12
O <sub>3</sub> (μg/m <sup>3</sup> )	7638	65.56 ± 27.01	4.67	43.29	65.16	86.19	178.87
CO (μg/m <sup>3</sup> )	7638	2.61 ± 1.75	0.32	1.30	2.19	3.38	18.41
SO <sub>2</sub> (μg/m <sup>3</sup> )	7622	23.22 ± 20.09	1.71	9.42	17.27	30.89	243.19
<b>Μετεωρολογικά</b>							
Θερμοκρασία (°C)	7626	18.57 ± 7.54	-6.94	12.53	18.06	25.16	36.51
Υγρασία (%)	7626	62.58 ± 14.48	20.25	51.63	63.25	73.38	100
<b>Αριθμός ημερήσιων θανάτων</b>							
Όλες οι αιτίες	7638	78 ± 14	34	69	77	87	191
Καρδιαγγειακή	7638	37 ± 9	14	31	37	43	111
Αναπνευστική	7638	7 ± 4	0	4	6	9	25

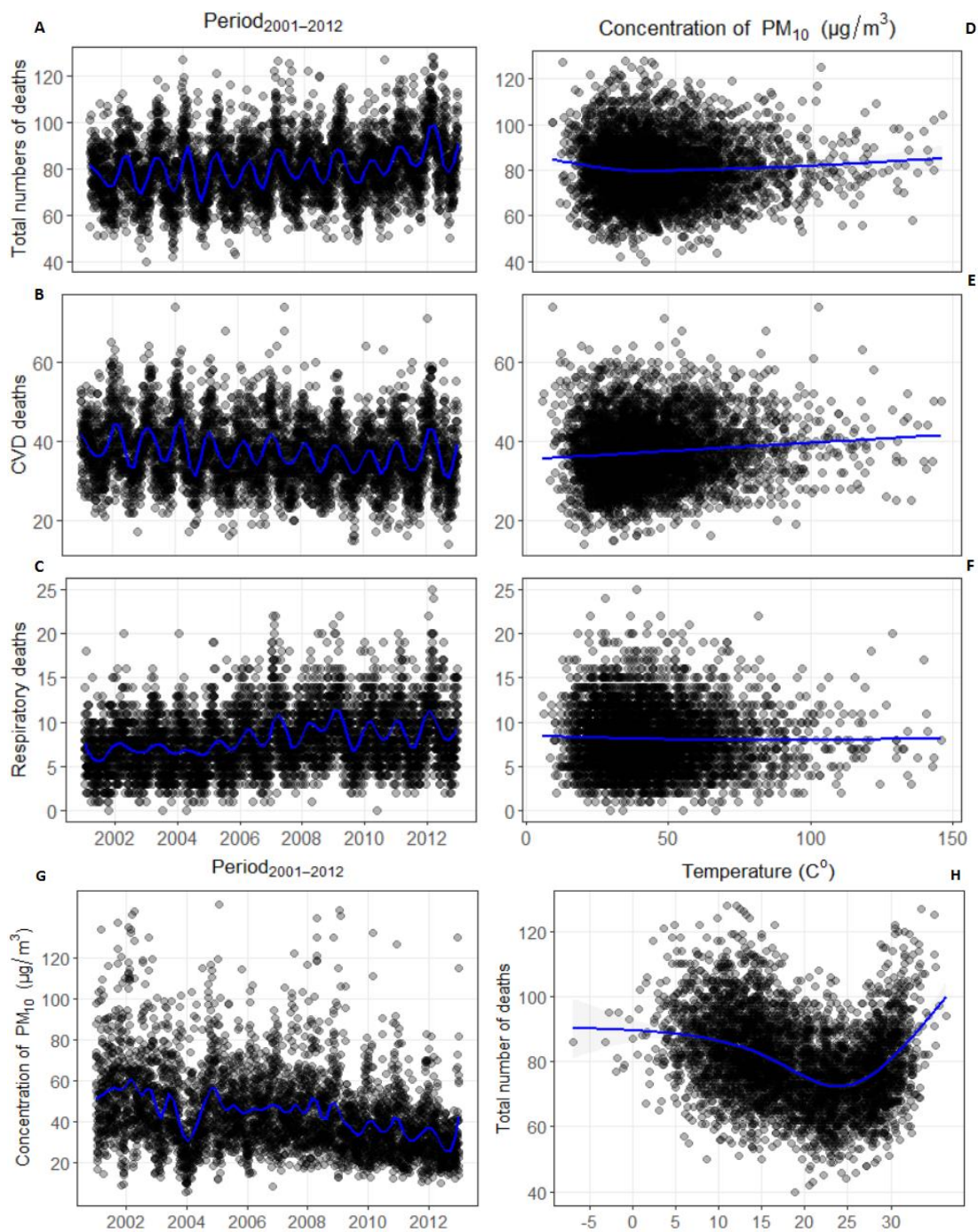
## ΓΡΑΦΗΜΑΤΑ

Στο **Γράφημα 9** το κάθε θηκόγραμμα συνοψίζει τα βασικά περιγραφικά μέτρα των ημερήσιων συγκεντρώσεων των  $PM_{10}$  ανά έτος, όπως τη διάμεσο, τα εκατοστημόρια, το ενδοτεταρτομοριακό εύρος και τις ακραίες τιμές. Η κάτω οριζόντια πλευρά του ορθογωνίου παραλληλογράμμου αποτελεί το 25<sup>ο</sup> εκατοστημόριο, η πάνω οριζόντια πλευρά το 75<sup>ο</sup>, στο εσωτερικό του η οριζόντια μαύρη γραμμή τη διάμεσο και ο κόκκινος ρόμβος τη μέση τιμή. Τέλος, οι τιμές που βρίσκονται εκτός των φρακτών είναι οι ακραίες τιμές. Κατά την πάροδο των χρόνων παρατηρούμε ότι το ενδοτεταρτομοριακό εύρος, η διαφορά δηλαδή του 75<sup>ο</sup> εκατοστημορίου με το 25<sup>ο</sup>, και αντίστοιχα η διάμεσος μειώνονται.



**Γράφημα 9** Θηκογράμματα της ετήσιας συγκέντρωσης των  $PM_{10}$  για τη χρονική περίοδο 1992-2012.

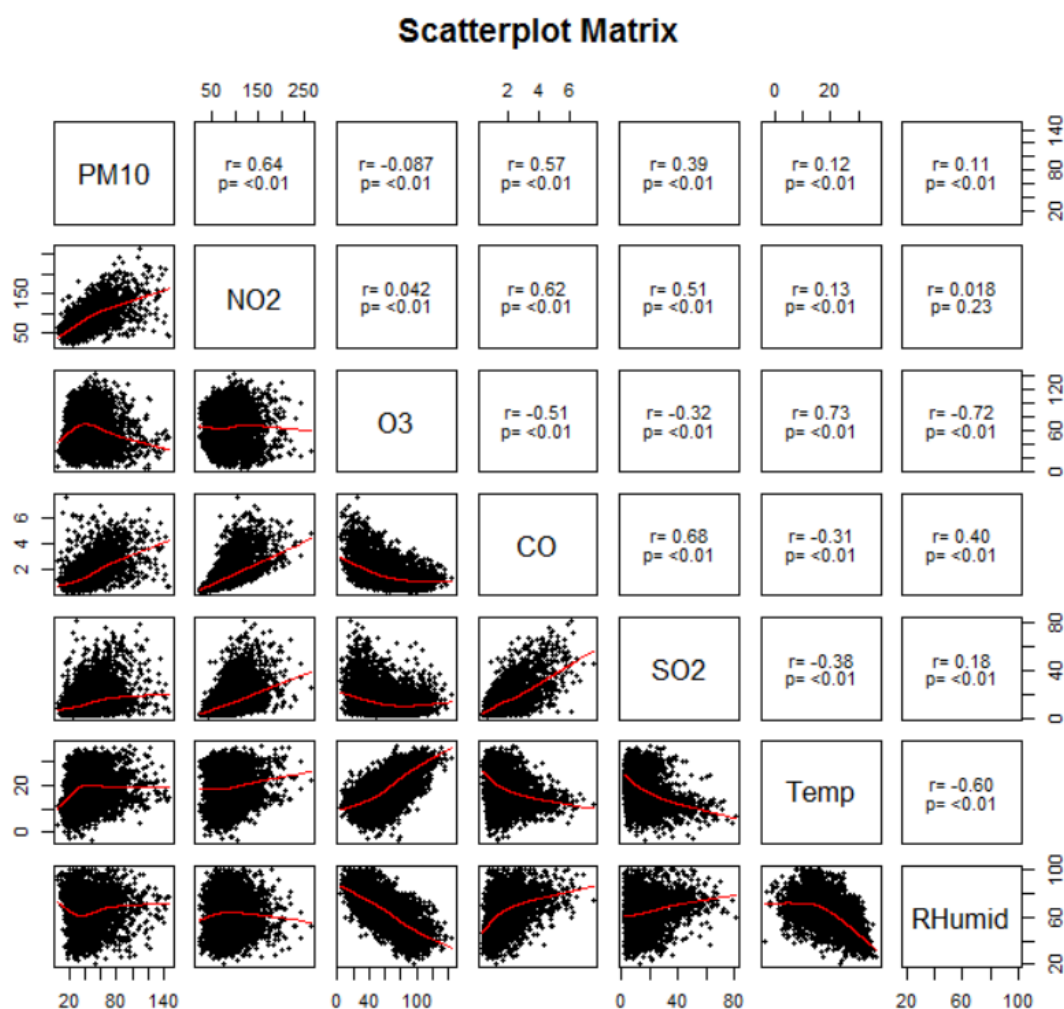
Για την πλήρη περιγραφή των δεδομένων δίνονται τα ακόλουθα στικτογράμματα (**Γράφημα 10**). Τα στικτογράμματα **A**, **B**, **C** απεικονίζουν τον ολικό, καρδιαγγειακό και αναπνευστικό ημερήσιο αριθμό θανάτων και το στικτόγραμμα **G** απεικονίζει την ημερήσια συγκέντρωση των  $PM_{10}$  την χρονική περίοδο 2001-2012. Τα στικτογράμματα **D**, **E** και **F** αντικατοπτρίζουν τη σχέση της ολικής και κατά αιτίας θνησιμότητας με τη συγκέντρωση των  $PM_{10}$  και το γράφημα **H** απεικονίζει τη σχέση μεταξύ της ημερήσιας ολικής θνησιμότητας και της ημερήσιας μέσης θερμοκρασίας. Παρατηρούμε, για παράδειγμα, ισχυρή εποχικότητα και μακροχρόνια αυξητική τάση στον ολικό αριθμό θανάτων (**A**). Λιγότεροι θάνατοι παρουσιάζονται την περίοδο του φθινοπώρου ή της άνοιξης σε σύγκριση με το χειμώνα και το καλοκαίρι (**A**), (**H**). Η συγκέντρωση των  $PM_{10}$  χαρακτηρίζεται από την εποχικότητα, μεγαλύτερες τιμές το χειμώνα και το καλοκαίρι παρά την άνοιξη και το φθινόπωρο, και από μία μακροχρόνια πτωτική τάση (**G**). Η σχέση μεταξύ της θνησιμότητας και των  $PM_{10}$  φαίνεται να είναι γραμμική με μία ελαφρά αυξητική τάση (**D**, **E**, **F**).



**Γράφημα 10** Στικτογράμματα του ολικού ημερήσιου αριθμού θανάτων (A), του καρδιαγγειακού (B) του αναπνευστικού (C) και της ημερήσιας συγκέντρωσης των PM<sub>10</sub> την περίοδο 2001-2012 (G), της σχέσης μεταξύ της ολικής και κατά αιτία θνησιμότητας με τη συγκέντρωση των PM<sub>10</sub> αντίστοιχα (D,E,F) για την ίδια περίοδο στην Αθήνα και της σχέσης της ολικής θνησιμότητας με τη θερμοκρασία (H).

## ΣΥΣΧΕΤΙΣΗ ΜΕΤΑΞΥ ΤΩΝ ΡΥΠΩΝ ΚΑΙ ΤΩΝ ΜΕΤΕΩΡΟΛΟΓΙΚΩΝ ΠΑΡΑΓΟΝΤΩΝ

Στο **Γράφημα 11** απεικονίζονται οι σχέσεις μεταξύ των ατμοσφαιρικών ρύπων και των μετεωρολογικών παραγόντων, και παρατηρούμε ότι προσεγγιστικά είναι γραμμικές. Στο γράφημα επίσης παρουσιάζονται οι συντελεστές συσχέτισης Pearson μεταξύ των μεταβλητών για την περίοδο 2001-2012. Στους **Πίνακες 7.5-7.6** δίνονται οι συντελεστές συσχέτισης Pearson ξεχωριστά για τις περιόδους 1992-2000 και 1992-2012. Οι μεταβλητές που έχουν συντελεστή συσχέτισης μεγαλύτερο του 0.7 κατά απόλυτη τιμή συσχετίζονται μεταξύ τους ισχυρά. Στον **Πίνακα 7.5** παρατηρούμε ότι οι μεταβλητές  $PM_{10}$  και θερμοκρασία συσχετίζονται ισχυρά μεταξύ τους καθώς έχουν συντελεστή συσχέτισης 0.73. Αυτός ο υψηλός συντελεστής μπορεί να αιτιολογηθεί από το γεγονός ότι στο μοντέλο πρόβλεψης των  $PM_{10}$  της περιόδου 1992-2000 η θερμοκρασία ήταν προβλεπτικός παράγοντάς του. Οι συντελεστές συσχέτισης των  $PM_{10}$  και  $NO_2$  στις χρονικές περιόδους 1992-2000, 2001-2012 και 1992-2012 έχουν τιμές 0.66, 0.64 και 0.68 αντίστοιχα.



**Γράφημα 11** Γραφική απεικόνιση της σχέσης μεταξύ των ατμοσφαιρικών ρύπων ( $PM_{10}$ ,  $NO_2$ ,  $O_3$ ,  $CO$ ,  $SO_2$ ), της θερμοκρασίας ( $Temp$ ) και της σχετικής υγρασίας ( $RHumid$ ), και τιμές των συντελεστών συσχέτισης Pearson αντίστοιχα, δεδομένα περιόδου 2001-2012.

**Πίνακας 7.5:** Συντελεστές Συσχέτισης Pearson μεταξύ ατμοσφαιρικών ρύπων και μετεωρολογικών παραγόντων στην Αθήνα, 1992-2000

Μεταβλητές	Προβλεπόμενα					Θερμο- κρασία	Σχετική Υγρασία
	PM <sub>10</sub>	SO <sub>2</sub>	NO <sub>2</sub>	O <sub>3</sub>	CO		
Προβλεπόμενα							
PM <sub>10</sub>	1						
SO <sub>2</sub>	0.02	1					
NO <sub>2</sub>	0.66	0.49	1				
O <sub>3</sub>	0.52	-0.24	0.17	1			
CO	0.22	0.66	0.59	-0.40	1		
Θερμοκρασία	0.73	-0.31	0.16	0.76	-0.29	1	
Σχετική Υγρασία	-0.29	0.14	0.01	-0.69	0.43	-0.65	1

**Πίνακας 7.6:** Συντελεστές Συσχέτισης Pearson των ατμοσφαιρικών ρύπων και των μετεωρολογικών παραγόντων στην Αθήνα, 1992-2012

Μεταβλητές	PM <sub>10</sub>	SO <sub>2</sub>	NO <sub>2</sub>	O <sub>3</sub>	CO	Θερμοκρασία	Σχετική Υγρασία
PM <sub>10</sub>	1						
SO <sub>2</sub>	0.25	1					
NO <sub>2</sub>	0.68	0.56	1				
O <sub>3</sub>	0.22	-0.20	0.12	1			
CO	0.42	0.76	0.64	-0.32	1		
Θερμοκρασία	0.39	-0.30	0.12	0.74	-0.26	1	
Σχετική Υγρασία	-0.08	0.10	0.002	-0.71	0.29	-0.62	1



## 7.3 ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ

### ΜΟΝΤΕΛΟ POISSON

Η σχέση μεταξύ των  $PM_{10}$  και της θνησιμότητας διερευνήθηκε εφαρμόζοντας γενικευμένα προσθετικά μοντέλα Poisson προσαρμοσμένα για την ύπαρξη υπερ-διασποράς. Η μορφή του μοντέλου είναι η εξής:

$$\log E[Y] = b_0 + b * lag01PM_{10} + s(trend, k) + s(temp, df) + s(temp_{lag1}, df) + [others] \quad [46]$$

όπου  $E[Y^t]$  η αναμενόμενη τιμή του ημερήσιου αριθμού θανάτων την t-ημέρα ( $Y^t$ ) με διασπορά  $Var[Y^t] = \varphi E[Y^t]$ ,  $\varphi$  είναι η παράμετρος υπερ-διασποράς,  $PM_{10}$  τα μέσα επίπεδα συγκέντρωσης των αιωρούμενων σωματιδίων της ίδιας και της προηγούμενης ημέρας θανάτου, confounders η λίστα των συγχυτικών παραγόντων, την οποία θα περιγράψουμε εκτενέστερα παρακάτω, και  $b$  ο συντελεστής που μας ενδιαφέρει, είναι ο προσαρμοσμένος λογαριθμικός σχετικός κίνδυνος της θνησιμότητας για μία μονάδα αύξησης της συγκέντρωσης των  $PM_{10}$ .

Οι συγχυτικοί παράγοντες επιλέχθηκαν εκ των προτέρων με βάση την προηγούμενη γνώση, τη βιβλιογραφία και τις προκαταρκτικές διερευνητικές αναλύσεις. Η λίστα των συγχυτικών παραγόντων στο μοντέλο [46] είναι:

- i. *Εποχικότητα*: Προκειμένου να ληφθούν υπόψιν οι μακροπρόθεσμες και οι εποχιακές τάσεις εφαρμόστηκαν πολυώνυμα παρεμβολής με περιορισμό (penalized regression splines) με φυσικά κυβικά πολυώνυμα παρεμβολής (natural cubic splines) ως συναρτήσεις βάσης της μορφής  $s(trend, df)$  (Wood, 2000). Οι συναρτήσεις εξομάλυνσης  $s$  περιγράφουν τη μη γραμμική σχέση μεταξύ των χρονικά μεταβαλλόμενων μεταβλητών, του ημερολογιακού έτους και της ημερήσιας θνησιμότητας χρησιμοποιώντας τους αντίστοιχους βαθμούς ελευθερίας. Αρχικά επιλέχθηκαν,  $k=50$  συναρτήσεις βάσης, ενώ ο τελικός έλεγχος για την εποχικότητα βασίστηκε στην επιλογή των βαθμών ελευθερίας, οι οποίοι ελαχιστοποιούσαν την απόλυτη τιμή του αθροίσματος των μερικών αυτοσυσχετίσεων των καταλοίπων του μοντέλου για τη χρονική υστέρηση από 1 έως 30 ημέρες με την προϋπόθεση ενός ελάχιστου ορίου των 3 βαθμών ελευθερίας ανά έτος (Katsouyanni et al., 2009; Samoli et al., 2008).
- ii. *Θερμοκρασία*: Για τον έλεγχο τόσο των υψηλών όσο και των χαμηλών θερμοκρασιών εφαρμόστηκαν φυσικά κυβικά πολυώνυμα παρεμβολής (natural cubic splines) με 3 βαθμούς ελευθερίας με διαφορετικές χρονικές υστερήσεις (lags): θερμοκρασίες της ίδιας ημέρας προστέθηκαν για τον έλεγχο των υψηλών θερμοκρασιών (lag 0) και θερμοκρασίες της προηγούμενης ημέρας (lag 1) για τον έλεγχο των χαμηλών θερμοκρασιών (Katsouyanni et al., 2009)
- iii. *Σχετική Υγρασία*: Ο όρος για τον έλεγχο της υγρασίας είναι γραμμικός (linear term).
- iv. *Γρίπη*: Λόγω των μη διαθέσιμων στοιχείων, χρησιμοποιήθηκε η μέθοδος του APHEA-2 για τον έλεγχο της γρίπης (Katsouyanni et al. 2001). Δημιουργήσαμε μία ψευδομεταβλητή η οποία λαμβάνει την τιμή 1 όταν ο 7-ημερών κυλιόμενος μέσος της αναπνευστικής θνησιμότητας είναι μεγαλύτερος από το 90<sup>ο</sup> εκατοστημόριο της και την τιμή 0 διαφορετικά. Λόγω ορισμού της μεταβλητής αυτής χρησιμοποιήθηκε μόνο στην ανάλυση των συνολικών και καρδιαγγειακών θανάτων.

- v. *Μείωση του πληθυσμού κατά την περίοδο των θερινών διακοπών*: Ορίσαμε μία διατάξιμη  $3^{ov}$  επιπέδων μεταβλητή, που λαμβάνει την τιμή 2 για την περίοδο δύο εβδομάδων στα μέσα Αυγούστου, την τιμή 1 για την περίοδο 16 Ιουλίου έως 31 Αυγούστου, με εξαίρεση τις δύο εβδομάδες που αντιπροσωπεύει η κατηγορία 2, και την τιμή 0 (κατηγορία αναφοράς) για τις υπόλοιπες ημέρες.
- vi. *Αργίες*: Δημιουργήσαμε μία ψευδομεταβλητή η οποία παίρνει την τιμή 1 στις επίσημες αργίες εκτός της Κυριακής και 0 διαφορετικά.
- vii. *Ημέρα της εβδομάδας*: Ορίσαμε κατηγορική μεταβλητή με 6 επίπεδα, ένα για κάθε μία ημέρα της εβδομάδας, Κυριακή=0, Δευτέρα=1, Τρίτη=2, Τετάρτη=3, Πέμπτη=4, Παρασκευή=5 και Σάββατο=6.

Για να εξετάσουμε τις άμεσες επιπτώσεις της έκθεσης στη θνησιμότητα χρησιμοποιήσαμε το μέσο όρο των επιπέδων συγκέντρωσης των αιωρούμενων σωματιδίων ( $PM_{10}$ ) της ημέρας θανάτου και της προηγούμενης (lag 0-1). Η δομή της χρονικής υστέρησης των δευτερευουσών μεταβλητών (π.χ. μετεωρολογικοί παράγοντες) εξαρτήθηκε από τη χρονική υστέρηση της έκθεσης. Ο αριθμός των ημερών έκθεσης που γίνεται έλεγχος για τις δευτερεύουσες μεταβλητές δεν μπορεί να είναι κατά πολύ μεγαλύτερος από τον αριθμό των ημερών έκθεσης της βασικής μεταβλητής, των  $PM_{10}$ , καθώς ένα μέρος της επίδρασης που μελετάται πιθανώς να αποδοθεί λανθασμένα σε αυτές.

Εφαρμόστηκαν μοντέλα δύο ρύπων για να ελεγχθεί η πιθανή συγχυτική επίδραση άλλων ρύπων στα σωματίδια. Στο μοντέλο συμπεριλήφθηκαν εναλλάξ τέσσερις ατμοσφαιρικοί ρύποι: το  $NO_2$ ,  $O_3$ ,  $CO$ ,  $SO_2$ .

Η επιδημιολογική ανάλυση έγινε ξεχωριστά για τις περιόδους 1992-2000, 2001-2012 και 1992-2012. Αρχικά χρησιμοποιήθηκαν τα δεδομένα της δωδεκαετίας 2001-2012, στην οποία οι μετρήσεις των  $PM_{10}$  ήταν διαθέσιμες, για να κατασκευάσουμε το επιδημιολογικό μοντέλο. Στη συνέχεια ερευνήθηκαν οι επιδράσεις κατά την περίοδο 1992-2000 χρησιμοποιώντας τις εκτιμώμενες συγκεντρώσεις των  $PM_{10}$  ενώ τέλος, εκτιμήθηκαν οι επιδράσεις για ολόκληρη τη χρονοσειρά 1992-2012. Η μελέτη χωρίστηκε στα εξής στάδια:

**1<sup>ο</sup> Στάδιο** Εκτιμήσαμε την επίδραση των αιωρούμενων σωματιδίων στην ολική, την καρδιαγγειακή και αναπνευστική θνησιμότητα στην Αθήνα την περίοδο 2001-2012.

**2<sup>ο</sup> Στάδιο** Ερευνήσαμε αν διαφοροποιήθηκε η επίδραση των  $PM_{10}$  στην ολική θνησιμότητα μετά το ξέσπασμα της οικονομικής κρίσης στην Ελλάδα συγκρίνοντας τις επιδράσεις μεταξύ των περιόδων 2001-2007 και 2008-2012 με την προσθήκη ενός όρου αλληλεπίδρασης μεταξύ του ρύπου και μιας ψευδομεταβλητής που παίρνει την τιμή 1 για τα έτη 2008-2012 και 0 διαφορετικά.

**3<sup>ο</sup> Στάδιο** Εξετάσαμε τη χρονική τάση της επίδρασης των  $PM_{10}$  χρησιμοποιώντας δεδομένα από το 2001-2012. Η γραμμική συνάρτηση θα έχει τη μορφή  $\beta(t) = a_0 + a_1t$ , όπου  $a_1$  είναι η μέση κλίση που ερμηνεύεται ως η μεταβολή της ποσοστιαίας αύξησης στη θνησιμότητα για μία αύξηση  $10\mu g/m^3$  στα επίπεδα της μέσης ημερήσιας συγκέντρωσης όταν ο χρόνος αυξάνεται κατά ένα έτος. Για να την υπολογίσουμε εισαγάγαμε έναν όρο αλληλεπίδρασης μεταξύ του ρύπου και των χρόνων σε συνεχή μεταβλητή.

**4<sup>ο</sup> Στάδιο** Υπολογίσαμε την επίδραση των  $PM_{10}$  ξεχωριστά για κάθε έτος, με την προσθήκη ενός όρου αλληλεπίδρασης μεταξύ του ρύπου και της χρονικής περιόδου σαν κατηγορική μεταβλητή για τα χρόνια 2001-2012.

**5<sup>ο</sup> Στάδιο** Εξετάσαμε την επίδραση των PM<sub>10</sub> στη ολική θνησιμότητα για την περίοδο 1992 έως 2012. Στη συνέχεια ερευνήσαμε τη χρονική τάση της επίδρασης αυτής με τον τρόπο που αναφέραμε στο 3<sup>ο</sup> στάδιο.

**6<sup>ο</sup> Στάδιο** Για την περίοδο 1992-2012 χρησιμοποιήσαμε όρους αλληλεπίδρασης μεταξύ του ρύπου και των παρακάτω ψευδομεταβλητών για να διερευνήσουμε αν η επίδραση των PM<sub>10</sub> στην ολική θνησιμότητα διαφοροποιείται στις περιόδους που ορίζουν οι ψευδομεταβλητές αυτές. Οι ψευδομεταβλητές ορίστηκαν ως εξής:

- Η πρώτη ψευδομεταβλητή έχει 5 επίπεδα, λαμβάνει την τιμή 0 για τα χρόνια 1992-1996 (κατηγορία αναφοράς), την τιμή 1 για τα 1997-2000, την τιμή 2 για τα 2001-2004, την τιμή 3 για τα 2005-2008 και την τιμή 4 για τα 2009-2012.
- Η δεύτερη ψευδομεταβλητή έχει 3 επίπεδα, λαμβάνει την τιμή 0 για τα έτη 1992-2000 (κατηγορία αναφοράς), την τιμή 1 για τα 2001-2007 και την τιμή 2 για τα 2008-2012.
- Η τρίτη ψευδομεταβλητή έχει 12 επίπεδα και ορίστηκε ως 0 για το μήνα Ιανουάριο, 1 για το Φεβρουάριο, 2 για το Μάρτιο, ... και 12 για το μήνα Δεκέμβριο.

Τα αποτελέσματα παρουσιάζονται ως η ποσοστιαία μεταβολή της θνησιμότητας που σχετίζεται με αύξηση 10 μg/m<sup>3</sup> στη συγκέντρωση των PM<sub>10</sub>. Στο 1<sup>ο</sup> στάδιο για τον υπολογισμό αυτής της ποσοστιαίας μεταβολής και του διαστήματος εμπιστοσύνης χρησιμοποιήσαμε τους τύπους

$$\text{Ποσοστιαία Μεταβολή} = (e^{10*b} - 1) * 100\% \quad [47]$$

$$\Delta E = [e^{(b \pm 1.96*se)*10} - 1] * 100\% \quad [48]$$

όπου *b* ο συντελεστής και *se* το τυπικό σφάλμα των PM<sub>10</sub> από το μοντέλο [1]. Στο 2<sup>ο</sup> στάδιο το μοντέλο αλληλεπίδρασης είναι της μορφής:

$$\log E[Y] = b_0 + b_1 * lag01PM_{10} + b_2 t_{2008-2012} + confounders \quad [49]$$

Για να υπολογίσουμε τη συνολική επίδραση των PM<sub>10</sub> στη θνησιμότητα για την περίοδο 2001-2007 χρησιμοποιήσαμε το συντελεστή και το τυπικό σφάλμα των PM<sub>10</sub>, και τα αντικαταστήσαμε στους τύπους [7.2] και [7.3] αντίστοιχα. Για να υπολογίσουμε τη συνολική επίδραση στην περίοδο 2008-2012 όμως χρειάστηκε να κάνουμε ένα ακόμη βήμα, να υπολογίσουμε το συντελεστή και το τυπικό σφάλμα που αντιστοιχούν σε αυτήν την περίοδο με τους τύπους:

$$b_{2008-2012} = b_1 + b_2 \quad [50]$$

$$se_{2008-2012} = \sqrt{Var(b_1, b_1) + Var(b_2, b_2) + 2cov(b_1, b_2)} \quad [51]$$

Καθώς το μοντέλο μας δίνει την επίδραση της κατηγορίας αναφοράς (*b*<sub>1</sub>) και τη διαφορά της επίδρασης (*b*<sub>2</sub>) μεταξύ των περιόδων που εξετάζουμε και εισάγουμε ως αλληλεπίδραση. Ομοίως εργαστήκαμε και στα υπόλοιπα ζητούμενα.

Η στατιστική ανάλυση πραγματοποιήθηκε με τη χρήση του στατιστικού πακέτου R (έκδοση 3.2.2). Η στατιστική σημαντικότητα θεωρήθηκε στο α<0.05.

## 8 ΑΠΟΤΕΛΕΣΜΑΤΑ

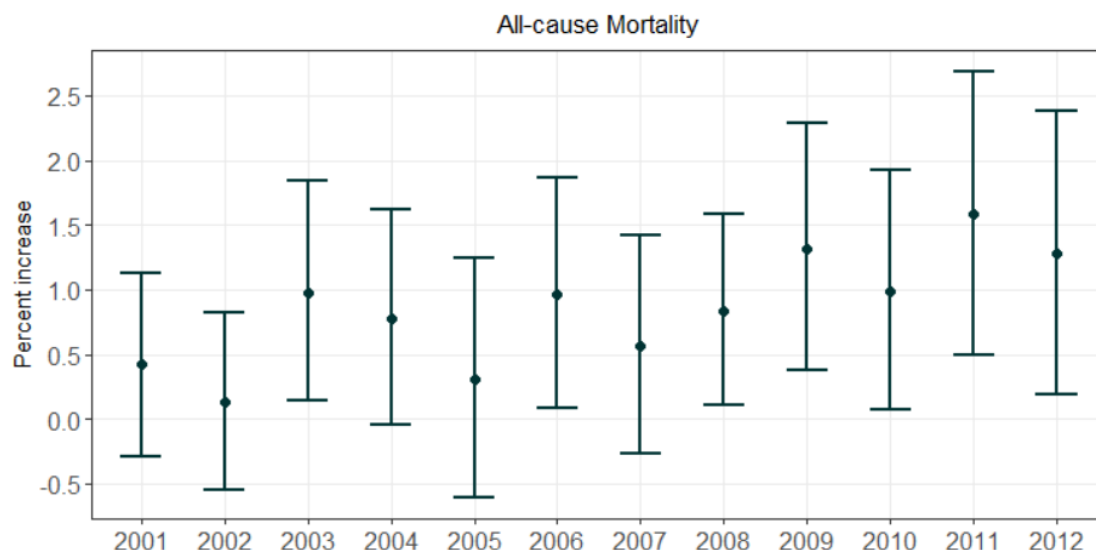
### ΑΝΑΛΥΣΗ ΠΕΡΙΟΔΟΥ 2001-2012

Ο Πίνακας 8.1 παρουσιάζει τις εκτιμήσεις της ποσοστιαίας μεταβολής στην ολική, καρδιαγγειακή και αναπνευστική θνησιμότητα που σχετίζονται με μία αύξηση των  $10\mu\text{g}/\text{m}^3$  στη συγκέντρωση των  $\text{PM}_{10}$  (lag 0-1) για την περίοδο 2001-2012 στην Αθήνα. Τα επίπεδα της συγκέντρωσης των  $\text{PM}_{10}$  σχετίζονται σε στατιστικά σημαντικό βαθμό με την ολική και κατά αιτία θνησιμότητα. Συγκεκριμένα  $10\mu\text{g}/\text{m}^3$  αύξηση στα  $\text{PM}_{10}$  σχετίζεται με 0.73% αύξηση στο συνολικό αριθμό ημερήσιων θανάτων, 0.86% στον αριθμό καρδιαγγειακών θανάτων και 1.76% στον αριθμό από αναπνευστικά αίτια. Οι επιδράσεις των  $\text{PM}_{10}$  στη θνησιμότητα από αναπνευστικά αίτια είναι μεγαλύτερες από αυτές στην ολική ή καρδιαγγειακή θνησιμότητα. Οι εκτιμήσεις από τα μοντέλα δύο ρύπων δείχνουν ότι δεν υπάρχει συγχυτική επίδραση από άλλους ρύπους.

**Πίνακας 8.1:** Ποσοστιαία μεταβολή (95% ΔΕ) στη θνησιμότητα που σχετίζεται με μία αύξηση  $10\mu\text{g}/\text{m}^3$  των lag 0-1  $\text{PM}_{10}$ , στην Αθήνα το 2001-2012

	Θνησιμότητα		
	Ολική	Καρδιαγγειακή	Αναπνευστική
$\text{PM}_{10}$	0.73 (0.45, 1.01)	0.86 (0.47, 1.26)	1.76 (0.94, 2.58)
Ελέγχοντας για			
$\text{SO}_2$	0.68 (0.37, 0.98)	0.87 (0.44, 1.30)	1.50 (0.63, 2.39)
$\text{NO}_2$	0.77 (0.46, 1.08)	0.97 (0.53, 1.42)	1.55 (0.64, 2.47)
$\text{O}_3$	0.70 (0.41, 0.98)	0.85 (0.45, 1.25)	1.69 (0.88, 2.52)
CO	0.83 (0.51, 1.15)	0.94 (0.48, 1.39)	1.77 (0.84, 2.71)

Υπολογίσαμε επίσης τη γραμμική τάση στην πάροδο των χρόνων από το 2001 έως το 2012 της επίδρασης των  $\text{PM}_{10}$  (lag 0-1) στην ολική, την καρδιαγγειακή και την αναπνευστική θνησιμότητα. Οι εκτιμώμενες κλίσεις για την ολική, καρδιαγγειακή και αναπνευστική θνησιμότητα είναι  $\hat{\alpha}_1 = 0.09$  (95%ΔΕ: 0.02, 1.16),  $\hat{\alpha}_1 = 0.11$  (95% ΔΕ : 0.005, 0.21) και  $\hat{\alpha}_1 = 0.16$ , (95%ΔΕ: - 0.05, 0.37) αντίστοιχα με την αλλαγή ενός έτους. Οι χρονικές τάσεις είναι σημαντικές σε στατιστικά σημαντικό βαθμό για την ολική και καρδιαγγειακή θνησιμότητα αλλά όχι για την αναπνευστική. Το Γράφημα 12 απεικονίζει την ποσοστιαία αύξηση (με το 95% ΔΕ) στη θνησιμότητα για μία αύξηση  $10\mu\text{g}/\text{m}^3$  στα επίπεδα των  $\text{PM}_{10}$  ανά έτος. Στο γράφημα εντοπίζουμε μία έντονη αυξητική τάση των επιπέδων των  $\text{PM}_{10}$  μετά το 2008. Η ποσοστιαία αύξηση της οποίας το διάστημα εμπιστοσύνης δεν περιέχει το 0 είναι στατιστικά σημαντική.

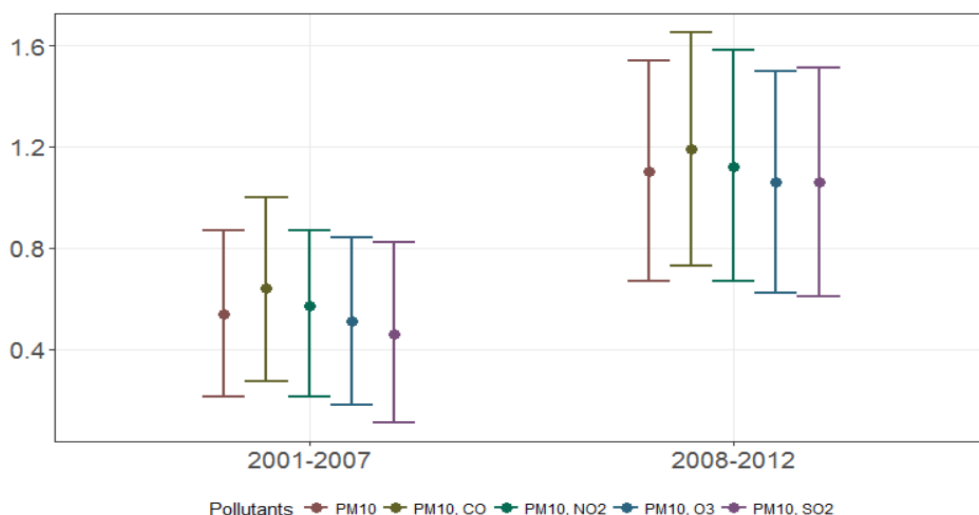


**Γράφημα 12** Η ποσοστιαία μεταβολή στον ημερήσιο αριθμό των θανάτων που σχετίζεται με την αύξηση  $10\mu\text{g}/\text{m}^3$  στα επίπεδα των lag 0-1  $\text{PM}_{10}$  για κάθε έτος από το 2001-2012.

Ο Πίνακας 8.2 περιλαμβάνει τις ποσοστιαίες αυξήσεις στην ολική θνησιμότητα που σχετίζονται με την αύξηση  $10\mu\text{g}/\text{m}^3$  στα επίπεδα των  $\text{PM}_{10}$  ξεχωριστά για τις περιόδους 2001-2007 και 2008-2012. Η επίδραση των  $\text{PM}_{10}$  στη θνησιμότητα διαφέρει σε στατιστικά σημαντικό βαθμό στις δύο περιόδους (0.54%, ΔΕ: 0.21, 0.87% σε σύγκριση με 1.10%, ΔΕ: 0.67, 1.54%, p-value=0.03). Επίσης, η διαφορά των επιδράσεων των  $\text{PM}_{10}$  στη θνησιμότητα που εκτιμήθηκαν από τα μοντέλα ανά δύο ρύπων συνεχίζει να είναι στατιστικά σημαντική. Το Γράφημα 13 απεικονίζει τα αποτελέσματα του Πίνακα 8.2. Ο y-άξονας συμβολίζει την ποσοστιαία αύξηση στην ολική θνησιμότητα που σχετίζεται με μία αύξηση  $10\mu\text{g}/\text{m}^3$  στα επίπεδα των  $\text{PM}_{10}$  και ο x-άξονας συμβολίζει τις δύο συγκρινόμενες περιόδους 2001-2007 και 2008-2012.

**Πίνακας 8.2:** Ποσοστιαία μεταβολή (95% ΔΕ) στην ολική θνησιμότητα που συσχετίζεται με μία αύξηση  $10 \mu\text{g}/\text{m}^3$  των lag 0-1  $\text{PM}_{10}$ , στην Αθήνα το 2001-2012

Ολική Θνησιμότητα			
Ρύποι	2001-2007	2008-2012	p-value αλληλεπίδρασης
$\text{PM}_{10}$	0.54 (0.21, 0.87)	1.10 (0.67, 1.54)	0.03
Ελέγχοντας για			
$\text{SO}_2$	0.46 (0.11, 0.82)	1.06 (0.61, 1.51)	0.02
$\text{NO}_2$	0.57 (0.21, 0.87)	1.12 (0.67, 1.58)	0.03
$\text{O}_3$	0.51 (0.18, 0.84)	1.06 (0.62, 1.50)	0.03
CO	0.64 (0.27, 1.00)	1.19 (0.73, 1.65)	0.03

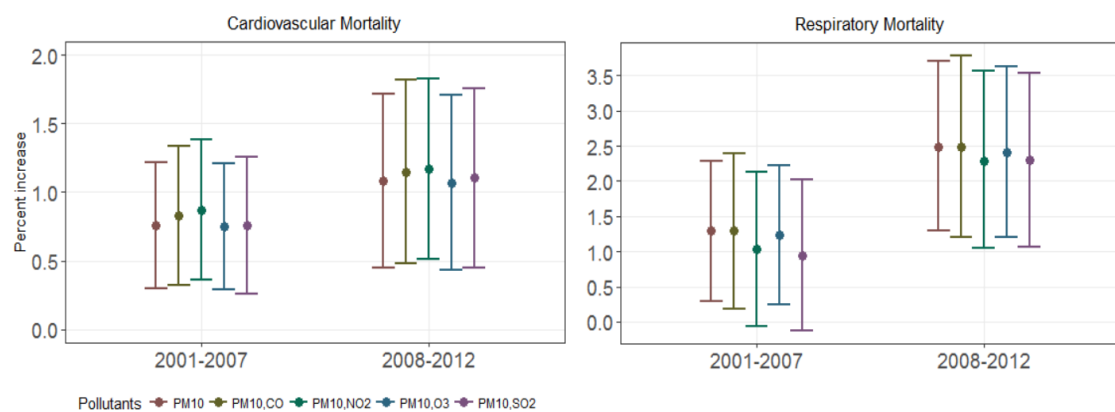


**Γράφημα 13** Η ποσοστιαία μεταβολή στον ημερήσιο ολικό αριθμό θανάτων που σχετίζεται με αύξηση  $10\mu\text{g}/\text{m}^3$  στα επίπεδα των lag 0-1  $\text{PM}_{10}$  για τις χρονικές περιόδους 2001-2007 και 2008-2012. Η κουκίδα απεικονίζει την ποσοστιαία μεταβολή και οι πάνω και κάτω οριζόντιες γραμμές ορίζουν το 95% ΔΕ.

Ομοίως, ο Πίνακας 8.3 παρουσιάζει τις ποσοστιαίες μεταβολές στην καρδιαγγειακή και αναπνευστική θνησιμότητα αντίστοιχα για κάθε αύξηση  $10\mu\text{g}/\text{m}^3$  στα επίπεδα των  $\text{PM}_{10}$  στις περιόδους 2001-2007 και 2008-2012. Οι ποσοστιαίες μεταβολές της περιόδου 2008-2012 έχουν σημειώσει αύξηση σε σύγκριση με τις ποσοστιαίες μεταβολές της περιόδου 2001-2007 αλλά αυτή η αύξηση δεν είναι στατιστικά σημαντική. Οι εκτιμήσεις από τα μοντέλα ανά δύο ρύπων, δεν αλλάζουν την εικόνα αυτή. Στο Γράφημα 14 απεικονίζονται τα αποτελέσματα του Πίνακα 8.3.

**Πίνακας 8.3:** Ποσοστιαία μεταβολή (95% ΔΕ) στην καρδιαγγειακή και αναπνευστική θνησιμότητα που συσχετίζεται με μία αύξηση  $10\mu\text{g}/\text{m}^3$  στα lag 0-1 των  $\text{PM}_{10}$ , στην Αθήνα το 2001-2012

Ρύποι	Καρδιαγγειακή θνησιμότητα		p-value for interaction
	2001-2007	2008-2012	
$\text{PM}_{10}$	0.76 (0.30, 1.22)	1.08 (0.45, 1.72)	0.381
Ελέγχοντας για			
$\text{SO}_2$	0.87 (0.36, 1.38)	1.17 (0.51, 1.83)	0.425
$\text{NO}_2$	0.75 (0.29, 1.21)	1.07 (0.43, 1.71)	0.385
$\text{O}_3$	0.76 (0.26, 1.26)	1.11 (0.45, 1.76)	0.378
CO	0.83 (0.32, 1.34)	1.15 (0.48, 1.82)	0.395
Αναπνευστική θνησιμότητα			
$\text{PM}_{10}$	1.29 (0.30, 2.28)	2.49 (1.29, 3.70)	0.107
Ελέγχοντας για			
$\text{SO}_2$	1.03 (-0.06, 2.13)	2.29 (1.05, 3.56)	0.090
$\text{NO}_2$	1.24 (0.25, 2.23)	2.41 (1.21, 3.63)	0.113
$\text{O}_3$	0.94 (-0.12, 2.02)	2.30 (1.07, 3.54)	0.069
CO	1.29 (0.19, 2.39)	2.49 (1.21, 3.78)	0.107



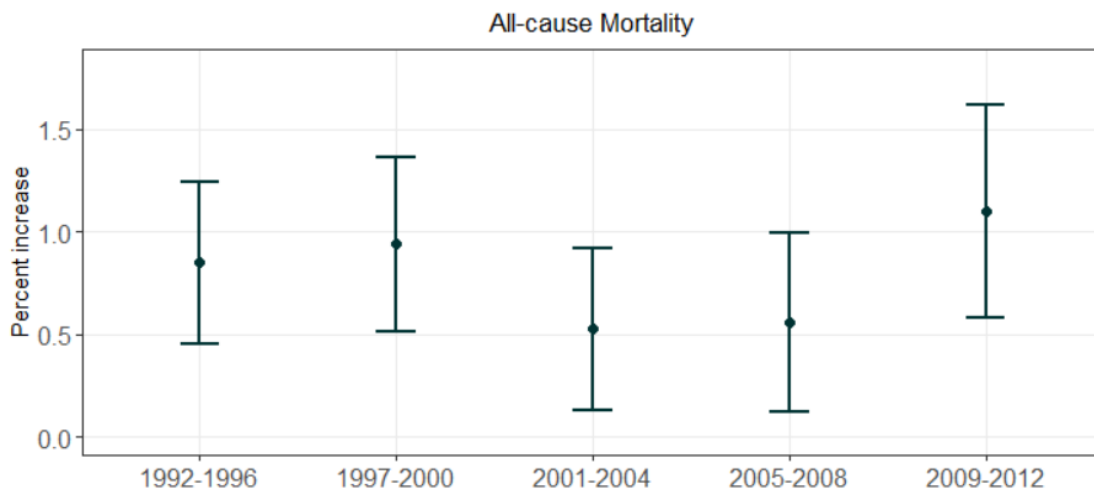
**Γράφημα 14** Η ποσοστιαία μεταβολή της καρδιαγγειακής (δεξιά) και αναπνευστικής (αριστερά) θνησιμότητας για αύξηση  $10 \mu\text{g}/\text{m}^3$  στα επίπεδα των lag 0-1  $\text{PM}_{10}$ , στις χρονικές περιόδους 2001-2007 και 2008-2012.

## ΑΝΑΛΥΣΗ ΠΕΡΙΟΔΟΥ 1992-2012

Για την περίοδο 1992-2012 βρήκαμε ισχυρές συσχετίσεις μεταξύ της επίδρασης των  $PM_{10}$  στην ολική, καρδιαγγειακή και αναπνευστική θνησιμότητα. Η ποσοστιαία αύξηση στην ολική θνησιμότητα υπολογίστηκε στα 0.77% (95% ΔΕ:0.54, 0.99%) για αύξηση των  $10 \mu\text{g}/\text{m}^3$  στα επίπεδα συγκέντρωσης των  $PM_{10}$ , στα 1.05% (95% ΔΕ:0.73, 1.36%) και 1.28% (95% ΔΕ:0.58, 1.97%) αντίστοιχα για τα καρδιακά και αναπνευστικά.

Υπολογίσαμε τη γραμμική τάση για την περίοδο 1992-2012, ομοίως όπως στην ανάλυση της περιόδου 2001-2012. Η μέση ποσοστιαία αύξηση στην ολική, καρδιαγγειακή και αναπνευστική θνησιμότητα που σχετίζεται με μία αύξηση  $10\mu\text{g}/\text{m}^3$  στα επίπεδα των  $PM_{10}$  είναι 0.79% (95%ΔΕ: 0.42, 1.16%) , 1.01% (95% ΔΕ: 0.49, 1.52) και 0.37% (95% ΔΕ: -0.8, 1.56) αντίστοιχα. Ισχυρά στατιστικά σημαντικές είναι οι επιδράσεις των  $PM_{10}$  στην ολική και καρδιαγγειακή θνησιμότητα ενώ δεν είναι στην αναπνευστική. Οι εκτιμώμενες κλίσεις είναι μη στατιστικά σημαντικές με τη μεταβολή ανά έτος.

Το **Γράφημα 15** παρουσιάζει τη ποσοστιαία μεταβολή στη συνολική ημερήσια θνησιμότητα για  $10\mu\text{g}/\text{m}^3$  στα  $PM_{10}$  ανά 4-ετία στην Αθήνα για το σύνολο της περιόδου 1992-2012. Παρατηρούμε ότι η επίδραση των αιωρούμενων σωματιδίων στην ολική θνησιμότητα είναι στατιστικά σημαντική σε όλες τις περιόδους με τη μέγιστη τιμή της να παρουσιάζεται την περίοδο 2009-2012. Οι επιδράσεις που διαφέρουν μεταξύ τους σε στατιστικά σημαντικό βαθμό είναι της περιόδου 2009-2012 με 2001-2004 και με 2005-2008, κάτι το οποίο παρατηρήσαμε και στην ανάλυση της περιόδου 2001-2012.



**Γράφημα 15** Οι ποσοστιαίες μεταβολές της ολικής θνησιμότητας που σχετίζονται με την αύξηση  $10\mu\text{g}/\text{m}^3$  των επιπέδων των lag 0-1  $PM_{10}$  ανά 4-ετία για την περίοδο 1992-2012.

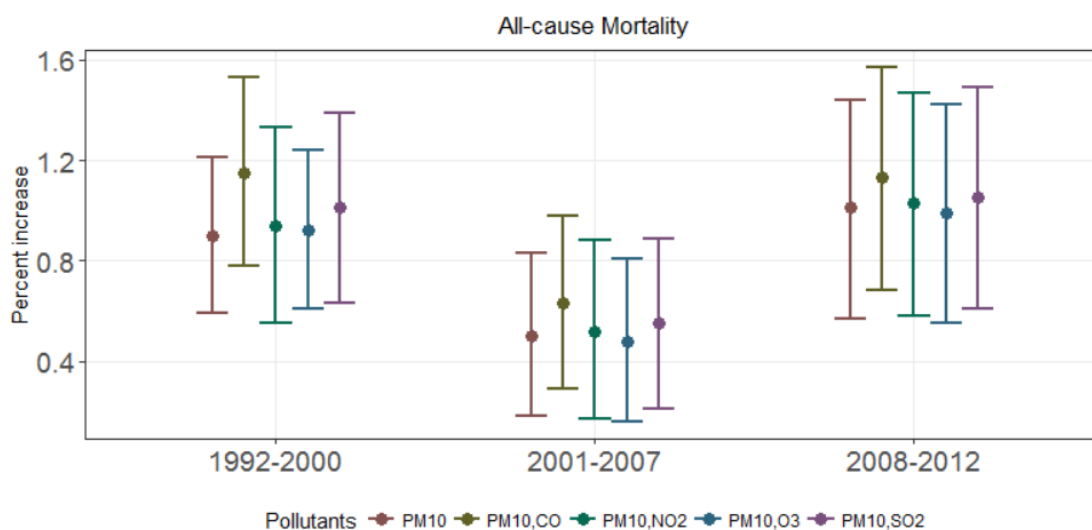
Ο **Πίνακας 8.4** και το **Γράφημα 16** συνοψίζουν τις εκτιμήσεις της ποσοστιαίας μεταβολής στην ολική θνησιμότητα που σχετίζεται με μία αύξηση  $10\mu\text{g}/\text{m}^3$  στα επίπεδα των  $PM_{10}$  ξεχωριστά για τρεις περιόδους, 1992-2000, 2001-2007 και 2008-2012 τόσο από τα μοντέλα ενός όσο και δύο ρύπων. Η ανάλυση έδειξε ότι η επίδραση των  $PM_{10}$  είναι στατιστικά σημαντική σε κάθε περίοδο, ενώ όπως και στην προηγούμενη ανάλυση δε φαίνεται να υπάρχουν συγχυτικές επιδράσεις από άλλους ρύπους. Συγκρίνοντας τις δύο πρώτες περιόδους παρατηρούμε ότι η διαφορά τους δεν



είναι σημαντική σε επίπεδο στατιστικής σημαντικότητας 5% ( $p=0.054$ ) αλλά είναι στο 10%. Οι επιδράσεις των περιόδων 1992-2000 και 2008-2012 δε διαφέρουν σημαντικά μεταξύ τους αλλά όπως απεικονίζεται στο γράφημα κυμαίνονται στα ίδια ή και λίγο πιο υψηλά επίπεδα.

**Πίνακας 8.4:** Ποσοστιαία μεταβολή (95% ΔΕ) στην ολική θνησιμότητα που συσχετίζεται με μία αύξηση  $10 \mu\text{g}/\text{m}^3$  των lag 0-1  $\text{PM}_{10}$ , στην Αθήνα το 1992-2012

Ρύποι	Ολική θνησιμότητα			p-value αλληλεπίδρασης	
	1992-2000	2001-2007	2008-2012	92-00 vs 01-07	92-00 vs 08-12
$\text{PM}_{10}$	0.90 (0.59, 1.21)	0.50 (0.18, 0.83)	1.01 (0.57, 1.44)	0.05	0.66
Ελέγχοντας για					
$\text{SO}_2$	1.01 (0.63, 1.39)	0.55 (0.21, 0.89)	1.05 (0.61, 1.49)	0.05	0.88
$\text{NO}_2$	0.94 (0.55, 1.33)	0.52 (0.17, 0.88)	1.03 (0.58, 1.47)	0.03	0.73
$\text{O}_3$	0.92 (0.61, 1.24)	0.48 (0.16, 0.81)	0.99 (0.55, 1.42)	0.04	0.81
CO	1.15 (0.78, 1.53)	0.63 (0.29, 0.98)	1.13 (0.68, 1.57)	0.01	0.91



**Γράφημα 16** Η ποσοστιαία μεταβολή στον ημερήσιο ολικό αριθμό θανάτων που σχετίζεται με αύξηση  $10 \mu\text{g}/\text{m}^3$  στα επίπεδα των  $\text{PM}_{10}$  για τις χρονικές περιόδους 1992-2000, 2001-2007 και 2008-2012.

## 9 ΣΥΖΗΤΗΣΗ

---

Η ατμοσφαιρική ρύπανση αποτελεί ένα σοβαρό πρόβλημα που επηρεάζει την ποιότητα του αέρα και την ποιότητα της ζωής. Τα επίπεδα ατμοσφαιρικής ρύπανσης στην Ελλάδα βρίσκονται κατά κανόνα πάνω από το μέσο όρο των 27 χωρών της Ευρωπαϊκής Ένωσης, ακόμη κι όταν υπάρχει συμμόρφωση με τα όρια. Το 2011, η Ελλάδα κατείχε την 3<sup>η</sup> υψηλότερη θέση ως προς το CO, την 5<sup>η</sup> ως προς το NO<sub>2</sub>, την 6<sup>η</sup> ως προς SO<sub>2</sub> και τα PM<sub>10</sub>, την 7<sup>η</sup> ως προς τα PM<sub>2.5</sub> και την 8<sup>η</sup> ως προς το O<sub>3</sub> (EEA, 2013).

Το μεγαλύτερο πρόβλημα για την ποιότητα του αέρα στα αστικά κέντρα της Ελλάδας αποτελούν τα αιωρούμενα σωματίδια με σχεδόν μόνιμες υπερβάσεις των ορίων σε σταθμούς αστικού-κυκλοφορίας. Η κυκλοφορία των αυτοκινήτων θεωρείται η σημαντικότερη πηγή των ρύπων αυτών (Karagulian et al. (2015)).

Η διαχείριση της ποιότητας του ατμοσφαιρικού αέρα απαιτεί τη συστηματική παρακολούθηση της ατμοσφαιρικής ρύπανσης και τη λήψη δραστικών μέτρων για τη βελτίωσή του. Η εφαρμογή των οδηγιών από την Ευρωπαϊκή Ένωση, η βελτίωση των χαρακτηριστικών του κινητήρα στα αυτοκίνητα καθώς και η μακρά απαγόρευση των πετρελαιοκίνητων αυτοκινήτων στα αστικά κέντρα οδήγησε στη διαχρονική μείωση των συγκεντρώσεων των αερίων και σωματιδιακών ρύπων.

Στη μελέτη Karanasiou et al. (2014) αναφέρεται ότι ανάμεσα στις Ευρωπαϊκές Μεσογειακές πόλεις η μεγαλύτερη μείωση της συγκέντρωσης των PM<sub>10</sub> παρατηρήθηκε στην Αθήνα (-4μg/m<sup>3</sup> ανά χρόνο, για την περίοδο 2007-2009). Αντίθετα αυξήθηκαν σημαντικά οι εκπομπές από την οικιακή θέρμανση, καθώς η αύξηση της τιμής του πετρελαίου θέρμανσης οδήγησε πολλά νοικοκυριά στην καύση ξύλου και βιομάζας (Karamanolis (2013)). Το 2011 επιστήμονες μελετώντας τη χημική σύσταση των PM<sub>10</sub> στην ατμόσφαιρα της Αθήνας εντόπισαν αρσενικό, που αποτελεί την πιο τοξική χημική ουσία σύμφωνα με την Υπηρεσία Περιβαλλοντικής Προστασίας των ΗΠΑ. Το αρσενικό κατά πάσα πιθανότητα προέρχεται από τα αυτοκίνητα, τα λεωφορεία ή από τους καυστήρες θέρμανσης (Ανεξάρτητη Ενημέρωση, 2011).

Έτσι, προκύπτει το ερώτημα, αν οι επιπτώσεις στην υγεία έχουν αλλάξει ύστερα από τις διάφορες πολιτικές για τη μείωση των ατμοσφαιρικών ρύπων, τις κοινωνικο-οικονομικές αλλαγές των τελευταίων χρόνων, τη διαχρονική μείωση των επιπέδων των σωματιδίων και της πιθανής αλλαγής του χημικού τους προφίλ. Στα πλαίσια της διπλωματικής εργασίας διερευνήθηκαν οι διαχρονικές τάσεις το 1992-2012 στις επιδράσεις των PM<sub>10</sub> στη θνησιμότητα των κατοίκων της Αθήνας.

Τα αποτελέσματα έδειξαν ότι οι επιδράσεις των PM<sub>10</sub> για το διάστημα 1992-2012 στη θνησιμότητα δε διαφοροποιήθηκαν σημαντικά διαχρονικά αλλά παρέμειναν σημαντικές. Ωστόσο υπήρξε αύξηση των επιδράσεων των PM<sub>10</sub> στην ολική και καρδιαγγειακή θνησιμότητα την περίοδο 2001-2012.

Συγκεκριμένα, εκτιμήθηκε ότι οι επιδράσεις των PM<sub>10</sub> στην ολική θνησιμότητα την περίοδο 2008-2012 σε σύγκριση με την περίοδο 2001-2007 σχεδόν διπλασιάστηκαν: αύξηση 10μg/m<sup>3</sup> στα επίπεδα των PM<sub>10</sub> συσχετίστηκε με αύξηση στην ολική θνησιμότητα κατά 1.10% (95%ΔΕ:0.67, 1.54%), ενώ πριν το 2008 η αντίστοιχη αύξηση εκτιμήθηκε στο 0.54% (95%ΔΕ: 0.21, 0.87%). Ενώ εκτιμήθηκε ότι οι επιδράσεις των PM<sub>10</sub> στην καρδιαγγειακή θνησιμότητα αυξήθηκαν την περίοδο 2001-

2012, (0.76% το 2001-2007 έναντι 1.08% το 2008-2012), η διαφορά των επιδράσεων δεν ήταν στατιστικά σημαντική. Αυτό ενδεχομένως να οφείλεται στο γεγονός ότι δεν υπήρξε περαιτέρω ανάλυση της επίδρασης στην καρδιαγγειακή θνησιμότητα της ηλικιακής κατηγορίας άνω των 75 ετών, η οποία αντιστοιχεί στο 30-40% περίπου της ολικής θνησιμότητας. Οι επιπτώσεις των  $PM_{10}$  στην αναπνευστική θνησιμότητα επίσης, δε φάνηκε να διαφοροποιούνται κατά τη διάρκεια της περιόδου 2001-2012 σε στατιστικά σημαντικό βαθμό, ωστόσο οι επιδράσεις μετά το 2008 σημείωσαν σημαντική αύξηση (1.29% έναντι 2.49%).

Οι επιδράσεις της βραχυχρόνιας έκθεσης στα  $PM_{10}$  στη θνησιμότητα στην Αθήνα που βρέθηκαν ήταν συμβατές με τα ευρήματα προηγούμενων αναφορών για τον ίδιο πληθυσμό. Συγκεκριμένα, η άμεση έκθεση στα  $PM_{10}$  (lag 0-1) σχετίστηκε θετικά και σε ισχυρά στατιστικά σημαντικό βαθμό με την ολική θνησιμότητα: 0.7% αύξηση στη θνησιμότητα για  $10\mu g/m^3$  αύξηση στα επίπεδα των  $PM_{10}$ . Στη διπλωματική εργασία του Αναστασίου Ζαχαρία (2014), η αύξηση  $10\mu g/m^3$  στα επίπεδα των  $PM_{10}$  σχετίστηκε με αύξηση 0.6% στην ολική θνησιμότητα, διερευνώντας την περίοδο 2001-2011. Ο έλεγχος που πραγματοποιήθηκε για την πιθανή συγχυτική επίδραση άλλων αέριων ρύπων στις εκτιμήσεις έδειξε ότι δεν υπάρχει συγχυτική επίδραση, γεγονός που επίσης συμφωνεί με την διπλωματική εργασία του Αναστασίου Ζαχαρία.

Στη μελέτη των Samoli et al. (2013) εκτιμήθηκε ότι η έκθεση σε  $PM_{10}$  (lag 0-1) σχετίζεται με αύξηση της ολικής θνησιμότητας κατά 0.4% (95%ΔΕ: -0.2, 0.8%) και ότι η παρατεταμένη έκθεση σε  $PM_{10}$  (lag 0-5) σχετίζεται με αύξηση της καρδιαγγειακής θνησιμότητας κατά 0.5% (95%ΔΕ: -0.5, 2%) και της αναπνευστικής κατά 1% (95%ΔΕ: -1, 3%) αλλά όχι σε στατιστικά σημαντικό βαθμό, για την περίοδο 2007-2009. Οι επιβαρυντικές επιδράσεις στην κατά αιτία θνησιμότητα από την έκθεση σε  $PM_{10}$  (lag 0-1) συνάδουν με τις επιβαρυντικές εκτιμήσεις της εργασίας για την περίοδο 2001-2012, οι οποίες βρέθηκαν να είναι στατιστικά σημαντικές και της τάξης του 0.9% (95%ΔΕ: 0.5, 1.3%) και 1.8% (95%ΔΕ: 0.9, 1.6%) αντίστοιχα.

Στη μελέτη Samoli et al. (2011) διερευνήθηκε η επίδραση της έκθεσης των lag 1  $PM_{10}$  για την περίοδο 2001-2006 και εκτιμήθηκε ότι για κάθε αύξηση  $10\mu g/m^3$  των επιπέδων των  $PM_{10}$  αναμένεται αύξηση της τάξης του 0.7% (95%ΔΕ: 0.4, 1%) στην ολική θνησιμότητα σε στατιστικά σημαντικό βαθμό και στην αναπνευστική θνησιμότητα στο 0.8% (95%ΔΕ: -0.1, 1.7%) αλλά όχι σε στατιστικά σημαντικό βαθμό. Τα αποτελέσματα των εκτιμήσεων για την περίοδο 2001-2007 της παρούσας εργασίας είναι συμβατά (0.5% (95%ΔΕ: 0.2, 0.8%) και 1.3% (95%ΔΕ: 0.3, 2.3%) αντίστοιχα). Επίσης, οι επιβαρυντικές εκτιμήσεις των  $PM_{10}$  στην καρδιαγγειακή θνησιμότητα (0.8% (95%ΔΕ: 0.3, 1.2%)) συμφωνούν με τις επιβαρυντικές εκτιμήσεις των Samoli et al. (2011) (1.1% (95%ΔΕ: 0.1, 2%) της καρδιαγγειακής θνησιμότητας άνω των 75 και 0.4% (95%ΔΕ: -0.3, 1.1% κάτω των 75 ετών).

Στη μελέτη Katsouyanni et al. (2001) εκτιμήθηκε η άμεση επίδραση της έκθεσης των  $PM_{10}$  για την περίοδο 1992-1996 και παρατηρήθηκε ότι για αύξηση  $10\mu g/m^3$  των  $PM_{10}$  αναμένεται αύξηση στην ολική θνησιμότητα 1.5% (95%ΔΕ: 1.0, 2.0%) σε στατιστικά σημαντικό βαθμό. Μετρήσεις των  $PM_{10}$  στην Αθήνα για την περίοδο 1992-1996 δεν υπήρχαν και εκτιμήθηκαν με ένα μοντέλο πολλαπλής παλινδρόμησης βασισμένο στις μετρήσεις του μαύρου καπνού. Στην παρούσα μελέτη ακολουθήσαμε παρόμοια μεθοδολογία, εκτιμώντας τα  $PM_{10}$  για την περίοδο 1992-2000 και βρήκαμε συμβατές σημαντικές επιβαρυντικές επιδράσεις αλλά μικρότερης κλίμακας, αφού για αύξηση  $10\mu g/m^3$  στα  $PM_{10}$  συνδέθηκε με αύξηση στην ολική θνησιμότητα κατά 0.8% (95%ΔΕ:)) σε στατιστικά σημαντικό βαθμό, στην ίδια περίοδο. Οι υψηλότερες επιδράσεις που εκτιμήθηκαν από τους Katsouyanni et al. (2001) πιθανώς οφείλονται στο διαφορετικό

επιδημιολογικό μοντέλο που εφαρμόστηκε, αφού για τον έλεγχο της εποχικότητας είχαν εφαρμοστεί μη παραμετρικές δοκιμασίες που όπως αναφέρθηκε στο Γενικό Μέρος, στην ενότητα 5.7 οδηγούσαν σε υπερ-εκτίμηση των επιδράσεων.

Ένας περιορισμός της μελέτης είναι ότι οι συγκεντρώσεις των  $PM_{10}$  για τη μισή σχεδόν χρονική περίοδο δεν υπήρχαν και εκτιμήθηκαν από ένα μοντέλο πολλαπλής παλινδρόμησης βασισμένο σε μετρήσεις του μαύρου καπνού, κάτι το οποίο εισάγει σφάλμα. Ένας δεύτερος περιορισμός είναι ότι στην ανάλυση χρησιμοποιήθηκαν δεδομένα μόνο από την περιοχή της Αθήνας, οπότε τα αποτελέσματα δεν μπορούν να γενικευθούν. Ωστόσο, η μεγάλη έκταση των χρονοσειρών από το 1992-2012 και ο μεγάλος όγκος δεδομένων κάνει τα αποτελέσματα πιο ισχυρά.

## 10 ΣΥΜΠΕΡΑΣΜΑΤΑ

---

Τα τελευταία χρόνια γίνεται συστηματική, επιστημονική προσπάθεια για την εκτίμηση των επιδράσεων στην υγεία από την έκθεση του πληθυσμού σε ατμοσφαιρικούς ρύπους με στόχο τον εφοδιασμό των αρμόδιων φορέων με κατάλληλα εργαλεία για το σχεδιασμό ιεραρχημένων περιβαλλοντικών στρατηγικών με σκοπό την προστασία της δημόσιας υγείας. Στην παρούσα μελέτη βρέθηκαν στατιστικά σημαντικές επιδράσεις από τη βραχυχρόνια έκθεση των  $PM_{10}$  στην ολική, καρδιαγγειακή και αναπνευστική θνησιμότητα για όλη την περίοδο της ανάλυσης 1992-2012. Επιπλέον υπάρχουν ενδείξεις ότι η κρίση έχει επιβαρύνει την κατάσταση καθώς βρέθηκε στατιστικά σημαντική αύξηση των επιδράσεων των  $PM_{10}$  στην ολική και καρδιαγγειακή θνησιμότητα με το έτος έναρξης της κρίσης το 2008. Έτσι διαπιστώνεται η αδυναμία της πολιτείας να αντιμετωπίσει την ατμοσφαιρική ρύπανση και τους κινδύνους που αυτή εγκυμονεί, ειδικά εν μέσω κρίσης.

## Περίληψη

**Εισαγωγή:** Οι επιδράσεις από τη βραχυχρόνια έκθεση σε αιωρούμενα σωματίδια διαμέτρου μικρότερης των 10  $\mu\text{m}$  ( $\text{PM}_{10}$ ) στην ολική, καρδιαγγειακή και αναπνευστική θνησιμότητα έχουν διερευνηθεί και τεκμηριωθεί από μία σειρά επιδημιολογικών ερευνών. Οι διάφορες πολιτικές που εφαρμόστηκαν για τη μείωση των επιπέδων της ατμοσφαιρικής ρύπανσης, αλλά και οι κοινωνικο-οικονομικές αλλαγές των τελευταίων χρόνων έχουν αλλάξει το χημικό προφίλ της ατμοσφαιρικής ρύπανσης ενώ τα επίπεδα των σωματιδίων έχουν μειωθεί. Στην παρούσα μελέτη διερευνήθηκαν οι χρονικές τάσεις της επίδρασης των  $\text{PM}_{10}$  στη θνησιμότητα των κατοίκων της Αθήνας για την περίοδο 1992-2012.

**Δεδομένα-Μέθοδοι:** Συλλέχθηκαν ημερήσια δεδομένα ολικής, καρδιαγγειακής και αναπνευστικής θνησιμότητας, ατμοσφαιρικών ρύπων και μετεωρολογικών παραγόντων για την περιοχή της Αθήνας, περιόδου 1992-2012. Μετρήσεις για τις συγκεντρώσεις των  $\text{PM}_{10}$  δεν υπήρχαν για τη χρονική περίοδο πριν το 2001, οπότε εκτιμήθηκαν από ένα μοντέλο πολλαπλής γραμμικής παλινδρόμησης βασισμένο στις μετρήσεις του μαύρου καπνού, ο οποίος αποτελεί μέρος των αιωρούμενων σωματιδίων. Το μοντέλο πρόβλεψης κατασκευάστηκε από τη συλλογή δεδομένων της περιόδου 2001-2004, κατά την οποία υπήρχαν ταυτόχρονες μετρήσεις των δύο ρύπων και εφαρμόστηκε στα έτη 1992-2000. Η επιδημιολογική ανάλυση πραγματοποιήθηκε για το σύνολο της χρονικής περιόδου 1992-2012, αλλά και χωριστά για την περίοδο με πραγματικές μετρήσεις 2001-2012 και την περίοδο με προβλέψεις 1992-2000, εφαρμόζοντας γενικευμένα προσθετικά μοντέλα Poisson προσαρμοσμένα για την ύπαρξη υπερ-διασποράς. Για τη διερεύνηση των χρονικών τάσεων χρησιμοποιήθηκε η μεθοδολογία που προτείνεται από τους Dominici et al. (2007).

**Αποτελέσματα:** Η επιδημιολογική ανάλυση για την περίοδο 2001-2012 έδειξε ότι αύξηση 10  $\mu\text{g}/\text{m}^3$  στα επίπεδα των lag 0-1  $\text{PM}_{10}$  συσχετίζεται με αύξηση 0.73% (95% $\Delta\text{E}$ : 0.45, 1.01%) στην ολική, 0.86% (95% $\Delta\text{E}$ :0.47,1.26%) στην καρδιαγγειακή και 1.76% (95% $\Delta\text{E}$ :0.94, 2.58%) στην αναπνευστική θνησιμότητα. Κατά τη διάρκεια αυτής της περιόδου σημειώθηκε αύξηση των επιπτώσεων των  $\text{PM}_{10}$  στην ολική και καρδιαγγειακή σε στατιστικά σημαντικό βαθμό. Θεωρώντας ως έτος έναρξης της οικονομικής κρίσης το 2008 στην Ελλάδα συγκρίθηκαν οι περίοδοι 2001-2007 και 2008-2012 και εντοπίστηκε στατιστικά σημαντική διαφορά στις επιδράσεις των  $\text{PM}_{10}$  στην ολική θνησιμότητα (0.54%, 95% $\Delta\text{E}$ : 0.21, 0.87% σε σύγκριση με 1.10%, 95% $\Delta\text{E}$ : 0.67, 1.54% αντίστοιχα, p-value-για τη διαφορά 0.03). Εφαρμόστηκαν μοντέλα δύο ρύπων ώστε να ελεγχθούν πιθανές συγχυτικές επιδράσεις των  $\text{SO}_2$ ,  $\text{NO}_2$ ,  $\text{O}_3$ ,  $\text{CO}$  από τα οποία δεν παρατηρήθηκαν διαφοροποιήσεις στις εκτιμήσεις μας.

Για το σύνολο της περιόδου 1992-2012 βρέθηκαν επίσης ισχυρές συσχετίσεις μεταξύ της επίδρασης των  $\text{PM}_{10}$  στην ολική, καρδιαγγειακή και αναπνευστική θνησιμότητα, αφού η αύξηση στη συγκέντρωση των  $\text{PM}_{10}$  κατά 10  $\mu\text{g}/\text{m}^3$  συσχετίστηκε με αύξηση 0.77% (95%  $\Delta\text{E}$ :0.54, 0.99%) στην ολική, 1.05% (95%  $\Delta\text{E}$ :0.73, 1.36%) στην καρδιαγγειακή και 1.28% (95%  $\Delta\text{E}$ :0.58, 1.97%) στην αναπνευστική θνησιμότητα.

Διερευνώντας τη χρονική τάση των επιδράσεων των  $PM_{10}$  αυτής της περιόδου στην ολική αλλά και κατά αιτία θνησιμότητα, βρέθηκε ότι δε διαφοροποιήθηκαν σημαντικά διαχρονικά αλλά οι επιδράσεις παραμείναν σημαντικές.

Συγκρίνοντας τις βραχυπρόθεσμες επιδράσεις των  $PM_{10}$  στη θνησιμότητα μεταξύ των περιόδων 1992-2000, 2001-2007 και 2008-2012 παρατηρήθηκε διαφορά σε επίπεδο στατιστικής σημαντικότητας 10% ανάμεσα στις δύο πρώτες περιόδους ( $p=0.054$ ) ενώ δεν υπήρξε διαφορά μεταξύ της πρώτης και της τρίτης περιόδου.

**Συμπεράσματα:** Βρέθηκαν στατιστικά σημαντικές επιδράσεις από τη βραχυπρόθεσμη έκθεση σε  $PM_{10}$  στην ολική, καρδιαγγειακή και αναπνευστική θνησιμότητα των κατοίκων της Αθήνας για την περίοδο 1992-2012, συμβατές με προηγούμενα ευρήματα. Υπήρχε στατιστικά σημαντική αύξηση των  $PM_{10}$  στην ολική και καρδιαγγειακή θνησιμότητα την περίοδο 2001-2012 που φαίνεται να διαφοροποιούνται σε σχέση με το έτος έναρξης της κρίσης το 2008.

## *Abstract*

---

**Introduction:** Various studies have investigated the short-term effects of particulate matter with aerodynamic diameter  $\leq 10$  (PM<sub>10</sub>) on all-cause, cardiovascular and respiratory mortality. The recent policy efforts and the social-economic situation at national level have not delivered the expected results, in spite of the improvements, as serious air pollution health impacts persist. In this study, we investigated the temporal trends in the short-term effect of PM<sub>10</sub> on mortality, by using data for Athens, Greece during the period 1992-2012.

**Data & Methods:** Data were selected on daily counts of all-cause and cause-specific mortality, time series of air pollution and meteorological parameters for Athens, for the period 1992-2012. PM<sub>10</sub> measurements were not available for the period 1992-2000, but were estimated using a regression model relating collocated PM<sub>10</sub> and black smoke measurements. Epidemiological analysis was performed separately for three periods — 1992-2000 using the predicted values of PM<sub>10</sub>, 2001-2012 using the measurements of PM<sub>10</sub> and 1992-2012 — and data were analyzed using Poisson regression models. To investigate temporal trends of short-term effects we used the methodology approach of Dominici et al. (2007) by including interaction terms in the model.

**Results:** A 10 $\mu\text{g}/\text{m}^3$  increase in PM<sub>10</sub> was associated with a 0.73% (95% CI: 0.45, 1.01%) increase in all-cause mortality (0-1 day cumulative lag), a 0.86% (95% CI: 0.47, 1.26%) in cardiovascular mortality and a 1.76% (95% CI: 0.94, 2.58%) in respiratory mortality, for 2001-2012. Associations between day-to-day variation in ambient PM<sub>10</sub> levels and all-cause and cardiovascular mortality counts became stronger over this period. The beginning of the year 2008 was selected as the starting year of the economic crisis in Greece and the difference of the effects of PM<sub>10</sub> in all-cause mortality between two periods, 2001-2007 and 2008-2012 was positively statistically significant (0.54%, 95% CI: 0.21, 0.87% vs 1.10%, 95% CI: 0.67%, 1.54% respectively,  $p=0.03$ ). Estimates from two-pollutant models did not provide evidence of confounding for associations between PM<sub>10</sub> and mortality.

For the whole period 1992-2012, a 10 $\mu\text{g}/\text{m}^3$  increase in PM<sub>10</sub> was associated with 0.77% (95% CI: 0.54, 0.99%) increase in all-cause mortality, a 1.05% increase (95% CI: 0.73, 1.36%) in cardiovascular mortality and a 1.28% (95% CI: 0.58, 1.97%) in respiratory mortality. The linear trend over time in the short-term effect of PM<sub>10</sub> on all-cause and cause-specific mortality from 1992 through 2012 was not statistically significant, although the associations between PM<sub>10</sub> and all-cause and cardiovascular mortality remained positively and statistically significant.

The effects for total mortality were greater for 2001-2007 than the period 1992-2000 but the difference was statistical significant at the level 10% ( $p=0.054$ ).

**Conclusions:** We found strong evidence of adverse effects of PM<sub>10</sub> on mortality in Athens, at the period 1992-2012. The effects of PM<sub>10</sub> on all-cause increased during 2001-2012, possibly as a consequence of the crisis in Greece, that started in 2008.

## Βιβλιογραφία

---

- AirDMS (2016) Αιωρούμενα Σωματίδια. Available: <http://www.air-quality.gr/pm.php>
- Box and Jenkins (1970) Available: <http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4463.htm>
- Chambers J., Hastie T. (1992) *Statistical Methods in S*. London, Chapman and Hall/CRC.
- Chaloulakou A., Grivas G., Spyrellis N. (2003) Neural Network and Multiple Regression Models for PM<sub>10</sub> Prediction in Athens: A Comparative Assessment. *Air & Waste Manage. Assoc.* 53(10): 1183-1190
- Dominici F., McDermott A., Zeger S.L., Samet J.M. (2002) On the use of generalized additive models in time-series studies of air pollution and health. *Am J Epidemiol* 156(3): 193-203.
- Dominici F., McDermott A., Hastie T. (2004) Improved semiparametric time series models of air pollution and mortality. *Journal of the ASA.* 99(468): 938-948
- Dominici F., Peng R.D., Zeger S L., et al. (2007) Particulate air pollution and mortality in the United States: Did the risks change from 1987 to 2000?. *Am J Epidemiol.* 146(8): 880-888.
- EEA, 2013a. Air Quality in Europe. Report No 9/2013 European Environment Agency Copenhagen, Denmark.
- Guardian (2016) Available: <https://www.theguardian.com/environment/2016/oct/30/air/quality-worsens-greece-recession-bites-world-pollutionwatch>
- Jill Jeff, Quasi-Likelihood Models in R. Available: [http://pages.wustl.edu/files/pages/imce/tpm\\_qmle.pdf](http://pages.wustl.edu/files/pages/imce/tpm_qmle.pdf).
- Hastie T., Tibshirani R. (1986) Generalized additive models. *Statistical Science* 1(3): 297-318.
- Hastie T., Tibshirani R. (1990) Generalized additive models. 1<sup>st</sup> edition. Chapman and Hall/CRC.
- Karagulian F. et al. (2015) Contributions to cities' ambient particulate matter (PM): A systematic review of local source contributions at global level. *Atmospheric Environment*, 120:475-483.
- Karamanoli E. (2013) Financial crisis harms respiratory health in Greece. *The Lancet Respir Med*, 1(7): 511-512.
- Karanasiou et al. (2014). Particulate matter and gaseous pollutants in the Mediterranean Basin: results from the MED-PARTICLES project. *Sci Total Environ*, 488-489: 297-315.
- Katsouyanni K., Schwartz J., Touloumi G., Spix C., Zanobetti A., Anderson H.R., et al. (1996). Short term effects of air pollution on health: a European approach using epidemiologic time series data: the APHEA protocol. *J Epidem Commun Health.* 50(S1): S12-S18.
- Katsouyanni K., Touloumi G., Samoli E., Gryparis A. et al. (2001) Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within APHEA2 project. *Epidemiology.* 12(5): 521-531.



- Katsouyanni K., Samet J., HR A., Atkinson R., et al. (2009) Air Pollution and health: a European and North American approach (APHENA). *Research Report 142*. Boston: Health Effects Institute.
- Kvarning A.M. (2008) Natural Cubic Splines. Department of Mathematical Sciences, Norwegian University of Science and Technology. Available: <https://www.math.ntnu.no/emner/TMA4215/2008h/cubicsplines.pdf>
- Marx B.D., Eilers P.H.C. (1998) Direct generalized additive modeling with penalized likelihood. *Computational Statistics and Data Analysis*. 28(2): 193-209.
- Nurminen M. (1997) The Use of Time Series Analysis in Environmental Epidemiology. In: C. Corvalan, M. Nurminen, H-Pastide (eds). Linkage Methods for Environment and Health Analysis. Geneva, World Health Organization. 73-95.
- Ramsay T., Burnett R., Krewski D. (2003) The effect of concurvity in generalized additive models linking mortality to ambient particulate matter. *Epidemiology* 14(1): 18-23.
- Shahraiyni H.T., Sodoudi S. (2016) Statistical Modeling Approaches for PM<sub>10</sub> Prediction in Urban Areas; A Review of 21<sup>st</sup>- Century Studies. *Atmos.* 7,15.
- Samoli E., et al. (2005). Estimating the exposure-response relationships between particulate matter and mortality within the APHEA multicity project. *Environ Health Perspect.* 113(1): 88-95.
- Samoli et al. (2008) Acute Effects of Ambient Particulate Matter on Mortality in Europe and North America: Results from the APHENA Study. *Environ Health Perspect.* 116(11): 1480-6.
- Samoli E., Kougea E., Kassomenos P., Analitis A., et al (2011) Does the presence of desert dust modify the effect of PM<sub>10</sub> on mortality in Athens, Greece? *Sci Total Environ.* 409(11): 2049-54.
- Samoli E., Stafoggia M., Rodopoulou S., et al. (2013). Associations between fine and coarse particles and mortality in Mediterranean cities: results from the MED-PARTICLES project. *Environ Health Perspect.* 121(8): 932-938
- Touloumi G., Samoli E., Pipikou M., Le Tertre A., et al. (2006) Seasonal confounding in air pollution and health time-series studies: effect on air pollution effect estimates. *Stat Med.* 25(24): 4164-4178.
- WHO (2016) Available: <http://www.who.int/mediacentre/factsheets/fs313/en/>
- Wood S.N. (2000) Modeling and smoothing parameter estimation with multiple quadratic penalties. *J. R. Statist. Soc. B.* 62(2): 413-428.
- Αναστασίου Ζ. (2014) Χαμένα έτη ζωής από την επίδραση των αιωρούμενων ατμοσφαιρικών σωματιδίων στην Αθήνα. Διπλωματική Εργασία, ΠΜΣ στη Βιοστατιστική, ΕΚΠΑ
- Ανεξάρτητη Ενημέρωση (2011) Available: [Tvxs.gr/news/περιβάλλον/αρσενικό-στην-ατμόσφαιρα-του-λεκανοπεδίου-αττικής](http://tvxs.gr/news/περιβάλλον/αρσενικό-στην-ατμόσφαιρα-του-λεκανοπεδίου-αττικής)
- Βικιπαίδεια (2016) Ατμοσφαιρική Ρύπανση. Available: [https://el.wikipedia.org/wiki/Ατμοσφαιρική\\_ρύπανση](https://el.wikipedia.org/wiki/Ατμοσφαιρική_ρύπανση)
- Βικιπαίδεια (2004) Ευρύτατες Αστικές Περιοχές. Available: [https://el.wikipedia.org/wiki/Ευρύτερες\\_αστικές\\_περιοχές#cite\\_note-4](https://el.wikipedia.org/wiki/Ευρύτερες_αστικές_περιοχές#cite_note-4)
- ΕΛ.ΣΤΑΤ. Available: [www.statistics.gr](http://www.statistics.gr)
- ΕΟΠ (2016) Available: <http://www.eea.europa.eu/el/themes/air/intro>
- Μάργια Γ. (2009) Ανάλυση και Πρόβλεψη Χρονοσειρών. Διπλωματική Εργασία, ΠΜΣ Υπολογιστική Φυσική, ΑΠΘ

- Μελάς Δ. (2000) Ατμοσφαιρική Ρύπανση, Περιβαλλοντική Εκπαίδευση, Τμήμα Φυσικής, ΑΠΘ. Available: [lap.physics.auth.gr/rms/upload/WEB\\_POLLUTION.DOC](http://lap.physics.auth.gr/rms/upload/WEB_POLLUTION.DOC)
- Μπούτσικας Μ. (2004) Σημειώσεις μαθήματος Στατιστικά Προγράμματα, Τμήμα Στατιστικής Και Ασφαλιστικής, ΠΑΠΕΙ. Available: [http://www.actuar.aegean.gr/notes/SPSS\\_lesson11.pdf](http://www.actuar.aegean.gr/notes/SPSS_lesson11.pdf)
- Παπαγεωργίου Χ. (2010) Σημειώσεις μαθήματος Στατιστικής Ι, Τμήμα Μαθηματικού, ΕΚΠΑ
- Χαλιανίδης Χ. (2010) Συμπερασματολογία και επιλογή μεταβλητών για μοντέλα ποσοστημοριακής παλινδρόμησης. Διπλωματική Εργασία, ΠΜΣ στη Στατιστική και Επιχειρησιακή Έρευνα, ΕΚΠΑ
- ΥΠΕΚΑ (2016) Available: <http://www.ypeka.gr/?tabid=492>