

ΚΕΦΑΛΑΙΟ 4^ο

ΜΗ ΠΑΡΑΜΕΤΡΙΚΕΣ ΣΤΑΤΙΣΤΙΚΕΣ ΤΕΧΝΙΚΕΣ ΓΙΑ ΠΕΡΙΒΑΛΛΟΝΤΙΚΑ ΔΕΔΟΜΕΝΑ

Εισαγωγή

Εξ' αιτίας των ιδιαιτεροτήτων και των προβλημάτων των δεδομένων τα συμπεράσματα που προκύπτουν από την εφαρμογή της κλασσικής στατιστικής σε ημερήσιους μέσους όρους, όπως τα γενικά δεδομένα που χρησιμοποιήθηκαν για την εύρεση συσχετίσεων ή για την εφαρμογή ελέγχου υποθέσεων είναι ενδεικτικά και χρησιμεύουν για την εύρεση των κατευθύνσεων προς τις οποίες μπορούν να εφαρμοστούν ειδικές στατιστικές αναλύσεις για την εξαγωγή επιστημονικά τεκμηριωμένων συμπερασμάτων. Επίσης γίνεται αντιληπτή η αναγκαιότητα των στατιστικών τεχνικών που υπερβαίνουν τα προβλήματα και τις ιδιαιτερότητες των περιβαλλοντικών δεδομένων.

Τα περιβαλλοντικά δεδομένα, παρουσιάζουν ιδιαιτερότητες που δυσχεραίνουν τη στατιστική τους ανάλυση και την εξαγωγή ουσιαστικών και αξιόπιστων συμπερασμάτων με τις κλασσικές στατιστικές μεθόδους (έλεγχοι υποθέσεων με προαπαιτούμενη γνώση της κατανομής, μοντέλα χρονολογικών σειρών κλπ). Το γεγονός αυτό οφείλεται στις missing values, τις ακραίες τιμές και στις τιμές σε όρια που δεν ανιχνεύονται πάντα με ακρίβεια. Τα παραπάνω προβλήματα είναι ιδιαίτερα συχνά στα περιβαλλοντικά δεδομένα εξ' αιτίας της δυσκολίας που παρουσιάζει η μέτρηση και συλλογή τους, και οφείλεται σε διάφορους παράγοντες όπως, η ευαισθησία των μηχανημάτων μέτρησης, η ασυνέχεια στις μετρήσεις λόγω βλαβών ή τη δημιουργία νέων σταθμών μέτρησης. Ιδιαίτερα στο Λεκανοπέδιο της Αττικής από το 1985 έως σήμερα, καθώς τις τελευταίες δύο δεκαετίες στην Ελλάδα παρουσιάζεται έντονη αύξηση του ενδιαφέροντος για τη μόλυνση του περιβάλλοντος στις αστικές περιοχές.

Η απαίτηση για την εφαρμογή νέων τεχνικών οδήγησε σε προτάσεις νέων μεθόδων ή προσαρμογές παλαιότερων που βασίζονται κυρίως σε μη παραμετρικές μεθόδους οι οποίες είναι οι κατάλληλες να εφαρμόζονται και σε μικρά σετ δεδομένων (π.χ. όπως οι ετήσιες τιμές ρύπων) οδηγώντας σε αξιόπιστα αποτελέσματα. Οι παραπάνω τεχνικές είναι πρακτικές, εύκολα εφαρμόσιμες καθώς δεν απαιτούν τη γνώση της συγκεκριμένης κατανομής των δεδομένων. Επίσης δεν επηρεάζονται από τα συνήθη κενά των μετρήσεων (missing values), λάθος μετρήσεις και ακραίες τιμές που πολύ συχνά εμφανίζονται κατά τις μετρήσεις των ρύπων αλλά και των παραμέτρων που τους επηρεάζουν όπως μετεωρολογικές, γεωγραφικές κλπ. Ο κυριότερος και πιο διαδεδομένος στατιστικός έλεγχος που εφαρμόζεται στα δεδομένα που αφορούν την ατμόσφαιρα είναι ο έλεγχος Mann-Kendall. Στην παρούσα εργασία παρουσιάζεται και εφαρμόζεται και με την αρχική του μορφή αλλά και στις πιο σύγχρονες εκδοχές του. Επίσης παρουσιάζεται και εφαρμόζεται η μέθοδος του Sen για την εκτίμηση της κλίσης της τάσης καθώς και οι συσχετίσεις και συνδιακυμάνσεις των εξεταζόμενων μεταβλητών.

4.1 Έλεγχος Mann-Kendall για την εύρεση της τάσης

Ο μη παραμετρικός έλεγχος Mann-Kendall εφαρμόζεται όταν υπάρχει η πεποίθηση ότι τα δεδομένα μπορεί να αποτελούν χρονολογική σειρά με τάση αύξουσα ή φθίνουσα, όπως στην περίπτωση των ατμοσφαιρικών ρύπων. Έτσι γίνεται η υπόθεση ότι οι παρατηρήσεις X_i προκύπτουν από μία συνεχή μονότονη -αύξουσα ή φθίνουσα- συνάρτηση του χρόνου $f(t_i)$ ακολουθώντας το μοντέλο (I):

$$X_i = f(t_i) + \varepsilon_i \quad (I)$$

όπου τα ε_i είναι τα κατάλοιπα τα οποία προέρχονται από την ίδια κατανομή με μέσο 0.

Ο έλεγχος είναι μεταξύ της ισχύς της μηδενικής υπόθεσης H_0 και της εναλλακτικής της H_1 όπως παρακάτω:

H_0 : οι παρατηρήσεις είναι τυχαία καταναμευμένες στο χρόνο

H_1 : υπάρχει τάση στα δεδομένα (αύξουσα ή φθίνουσα)

Η στατιστική συνάρτηση που χρησιμοποιείται για τον έλεγχο υποθέσεων εξαρτάται από το πλήθος των δεδομένων και από την τυχόν έντονη αλληλεξάρτηση τους (δηλαδή πολύ κοντινές τιμές). Έτσι έχουμε τις παρακάτω περιπτώσεις (βλέπε **Timo Salmi, Anu Maatta, Pia Anttila, Tuija Ruoho-Airola, Toni Amnell, (2002)**. Detecting Trends of Annual Values of Atmospheric Pollutants by the Mann-Kendall Test and Sen's Slope Estimates – The Excel Template Application MAKESENS, Publications on Air Quality No 31, *Finnish Meteorological Institute*):

Πλήθος δεδομένων $n < 10$

Χρησιμοποιείται η παρακάτω στατιστική συνάρτηση

$n-1 \quad n$

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(x_j - x_k) \quad (2)$$

$k=1 \quad j=k+1$

όπου οι x_j, x_k παρατηρήσεις αντιπροσωπεύουν ετήσιες τιμές και $j, k > k$ και

$$\text{sgn}(x_j - x_k) = \begin{cases} 1 & \text{εάν } (x_j - x_k) > 0 \\ 0 & \text{εάν } (x_j - x_k) = 0 \\ -1 & \text{εάν } (x_j - x_k) < 0 \end{cases} \quad (3)$$

Οι τιμές που προκύπτουν από την εξίσωση (2) συγκρίνονται με την τιμή της θεωρητικής κατανομής S που αντιστοιχεί σε επίπεδο σημαντικότητας α . Θετικές τιμές της στατιστικής συνάρτησης S υποδηλώνουν την ύπαρξη αύξουσας τάσης ενώ αρνητικές τιμές την ύπαρξη φθίνουσας τάσης. Όσο πιο μεγάλο είναι το πλήθος των παρατηρήσεων n τόσο πιο ακριβής μπορεί να είναι ο έλεγχος κατά τον παρακάτω πίνακα:

Επίπεδο σημαντικότητας α	Πλήθος παρατηρήσεων n
0.1	≥ 4
0.05	≥ 5
0.01	≥ 6
0.001	≥ 7

Πλήθος δεδομένων $n \geq 10$

Χρησιμοποιείται η παρακάτω στατιστική συνάρτηση

$$Z = \begin{cases} (S-1)/[VAR(S)]^{1/2} & \text{εάν } S > 0 \\ 0 & \text{εάν } S = 0 \\ (S+1)/[VAR(S)]^{1/2} & \text{εάν } S < 0 \end{cases} \quad (4)$$

Όπου η διακύμανση της S υπολογίζεται από τον παρακάτω τύπο

$$VAR(S) = 1/18 \left[1(n-1)(2n+5) - \sum_{p=1}^q tp(t_p-1)(2t_p+5) \right] \quad (5)$$

Οι τιμές που προκύπτουν από την εξίσωση (4) συγκρίνονται με την τιμή της τυποποιημένης κανονικής κατανομής Z που αντιστοιχεί σε επίπεδο σημαντικότητας α . Θετικές τιμές της στατιστικής συνάρτησης Z υποδηλώνουν την ύπαρξη αύξουσας τάσης ενώ αρνητικές τιμές την ύπαρξη φθίνουσας τάσης.

4.2 Μέθοδος τους Sen για την εύρεση της κλίσης της τάσης

Η μέθοδος του Sen (βλέπε **Sirois, Allan, (1998)**. A Brief and Biased Overview of Time Series Analysis or How to Find that Evasive Trend. In WMO report No. 133: WMO/EMEP workshop on Advanced Statistical methods and their Application to Air Quality Data sets Helsinki, 14-18 September 1998) εφαρμόζεται στις περιπτώσεις δεδομένων που μπορεί να γίνει η υπόθεση της ύπαρξης γραμμικής τάσης. Έτσι γίνεται η θεώρηση ότι τα δεδομένα ακολουθούν την παρακάτω εξίσωση και προκύπτουν από την συνάρτηση f η οποία έχει ως ανεξάρτητη μεταβλητή το χρόνο t :

$$f(t) = Qt + B_i \quad (6)$$

Η μεταβλητή Q εκφράζει την κλίση της τάσης και η B είναι η σταθερά της εξίσωσης. Η εκτίμηση της κλίσης της τάσης προκύπτει από τη διάμεσο των N κλίσεων Q_i των ζευγαριών των n παρατηρήσεων των δεδομένων, δηλαδή:

$$Q_i = (x_j - x_k) / (j - k), \quad \text{όπου } j > k \quad (7)$$

Αρα το πλήθος N των εκτιμώμενων κλίσεων Q_i είναι $N = n(n-1)/2$

Η εκτιμήτρια της κλίσης της τάσης του Sen ορίζεται ως η διάμεσος των Q_i , που τίθενται σε διάταξη από τη μικρότερη στη μεγαλύτερη:

$$Q = Q_{[(N+1)/2]}, \text{ εάν } N \text{ περιττός αριθμός} \quad (8)$$

$$Q = \frac{1}{2} (Q_{[N/2]} + Q_{[(N+1)/2]}), \text{ εάν } N \text{ άρτιος αριθμός}$$

Το διάστημα εμπιστοσύνης της εκτίμησης Q της κλίσης της τάσης υπολογίζεται με τη βοήθεια μη παραμετρικής τεχνικής βασισμένης στην κανονική κατανομή. Τα όρια του διαστήματος είναι το $Q_{\min}$ (κάτω όριο) που ορίζεται ως η M_1^{η} μεγαλύτερη από τις διατεταγμένες κλίσεις Q_i και αντίστοιχα το $Q_{\max}$ (άνω όριο) είναι η $(M_2+1)^{\eta}$ μεγαλύτερη από τις διατεταγμένες κλίσεις Q_i . Τα M_1 και M_2 είναι συναρτήσεις του πλήθους N των εκτιμώμενων κλίσεων Q_i και της συνάρτησης C_a που εξαρτάται από τη διακύμανση S . Έτσι έχουμε:

$$C_a = Z_{1-\alpha/2} [\text{VAR}(S)]^{1/2} \quad (9)$$

$$M_1 = (N - C_a) / 2 \quad (10)$$

$$M_2 = (N + C_a) / 2$$

Τα M_1 και M_2 , όταν δεν είναι ακέραιοι στην πράξη στρογγυλοποιούνται προς τα κάτω και προς τα πάνω αντίστοιχα.

4.3 Εφαρμογή του Ελέγχου Mann-Kendall και της μεθόδου του Sen Χρήση του προγραμματιστικού εργαλείου MAKESENS.

Η εφαρμογή του ελέγχου Mann-Kendall για την εύρεση της τάσης και της μεθόδου Sen για την εύρεση της κλίσης της τάσης έγινε με τη χρήση μίας εφαρμογής του EXCEL με τη χρήση μακροεντολών της Visual Basic. Το εργαλείο ονομάζεται MAKESENS και έχει εκπονηθεί από το Φιλανδικό Μετεωρολογικό Ινστιτούτο. Αποτελεί μέρος project, που επιδοτήθηκε από το Συμβούλιο Υπουργών των Σκανδιναβικών χωρών, για την έρευνα της

ατμοσφαιρικής ρύπανσης ως συνέπειας του πληθυσμού και της μεταφοράς ρύπων μεταξύ των Σκανδιναβικών και Βαλτικών χωρών. Έχουν γίνει οι κατάλληλες προσαρμογές για την εφαρμογή του στους ρύπους του Λεκανοπεδίου της Αττικής για τα έτη 1987-2000.

Η εφαρμογή του αφορά *ετήσια* δεδομένα για την εύρεση της τάσης και εάν υπάρχει την εύρεση της κλίσης της. Δεν είναι κατάλληλο για την ανίχνευση της εποχικότητας. Είναι όμως κατάλληλο και για μικρό αριθμό δεδομένων, πρόβλημα που παρουσιάζεται στο Λεκανοπέδιο λόγω της έναρξης των μετρήσεων από το 1984 και τη συνεχή έναρξη καινούριων σταθμών ή την αναβάθμιση των παλαιότερων όπου προσθέτονται οι μετρήσεις περισσότερων ρύπων. Επιτρέπει την ύπαρξη missing values, συχνού προβλήματος στα δεδομένα. Η εφαρμογή του γίνεται σε κάθε σταθμό μέτρησης και τις αντίστοιχες μέτρησης του κάθε ρύπου μεμονωμένα. Έτσι επιτυγχάνεται να δοθεί η εικόνα της ατμοσφαιρικής μόλυνσης σε κάθε περιοχή του Λεκανοπεδίου χωριστά. Πληροφορία πολύ χρήσιμη για τη σωστή λήψη και εφαρμογή των αναγκαίων μέτρων από τους αρμόδιους φορείς. Όπως φαίνεται στη συνέχεια, στην πράξη η ακριβής εφαρμογή της θεωρίας και η τήρηση όλων των προϋποθέσεων δεν είναι εφικτή. Όμως σε πολλές περιπτώσεις που έστω και προσεγγιστικά τηρούνται οι προϋποθέσεις είναι δυνατή η εξαγωγή χρήσιμων συμπερασμάτων.

Στη συνέχεια παρουσιάζονται τα αποτελέσματα για κάθε σταθμό και ρύπο χωριστά, με τη βοήθεια ενός διαγράμματος και των αποτελεσμάτων των ελέγχων Mann-Kendall και Sen σε πίνακα με τα αντίστοιχα συμπεράσματα. Έχουν υπολογιστεί το μοντέλο της χρονολογικής σειράς καθώς και τα διαστήματα εμπιστοσύνης 95% ($\alpha=0.05$) και 99% ($\alpha=0.01$). Τα διαστήματα εμπιστοσύνης εμφανίζονται στο διάγραμμα καθώς και το διάγραμμα των καταλοίπων. Με την παρατήρηση του διαγράμματος προκύπτει η εποπτική εικόνα του ρύπου και πληροφορία που δε δίνει μόνο η αποδοχή ή απόρριψη του ελέγχου βάση του υπολογισμένου S ή Z του ελέγχου. Τα αποτελέσματα παραθέτονται στο Παράρτημα Ι.

4.4 Πολυμεταβλητός Έλεγχος Mann-Kendall για την εύρεση της τάσης

Ο Έλεγχος - ή Μονομεταβλητός - Έλεγχος Mann-Kendall (Univariate Mann-Kendall Test) που περιγράφεται και εφαρμόζεται παραπάνω είναι η πρόταση του ελέγχου όπως αρχικά παρουσιάστηκε. Στην πορεία προκειμένου να επιλυθεί το πρόβλημα της ανίχνευσης της τάσης στους ατμοσφαιρικούς ρύπους διεξοδικότερα, ο έλεγχος επεκτάθηκε με καινούριες προτάσεις προκειμένου να περιληφθούν:

1. η εποχικότητα,
2. οι σταθμοί μέτρησης,
3. φυσικοί παράγοντες όπως τα μετεωρολογικά ή υδρολογικά δεδομένα.

Ο σκοπός αυτών των προτάσεων είναι η ανίχνευση της τάσης ενσωματώνοντας στον έλεγχο με μορφή μεταβλητών, όλους τους παράγοντες που επηρεάζουν τις τιμές των ατμοσφαιρικών ρύπων. Έτσι επιτυγχάνεται αφ' ενός η ανίχνευση της ύπαρξης τάσης και αφ' ετέρου ο προσδιορισμός των *φυσικών* αιτιών που την επηρεάζουν με αποτέλεσμα να είναι εφικτή η απομόνωση της τάσης που οφείλεται σε *ανθρωπογενείς* παράγοντες από την τάση που οφείλεται σε διακυμάνσεις των *φυσικών* αιτιών.

4.5 Εποχικός Έλεγχος Mann-Kendall ή Έλεγχος Hirsch-Slack

(The seasonal Mann-Kendall test or Hirsch-Slack test)

Στην περίπτωση που τα δεδομένα μέτρησης του ατμοσφαιρικού ρύπου, που είναι η αιτιατή μεταβλητή (response variable), είναι εποχικά – δηλαδή είναι διαχωρισμένα ποιοτικά σε (ω) εποχές, – ο έλεγχος εφαρμόζεται στα δεδομένα κάθε (ω) εποχής, διαχωρίζοντας τα ουσιαστικά σε (ω) υπό – σειρές (sub series), όπου η κάθε υπό – σειρά αντιπροσωπεύει μία εποχή (βλέπε **Claudia Libiseller, (2002)**. Trend Testing in the presence of covariates, IMPACT 12, *Linköping University, Dept of Mathematics*).

$$T_j = \sum_{k < l} \text{sign}(Z_{lj} - Z_{kj}) \quad j = 1, \dots, \omega \quad (11)$$

Η T_j είναι η στατιστική συνάρτηση ελέγχου Mann-Kendall για κάθε εποχή j , όπου στη συνέχεια η άθροιση των αντίστοιχων T_j για κάθε (ω) εποχή καταλήγει στη εποχική συνάρτηση ελέγχου S όπως παρακάτω.

$$S = \sum_{j=1}^{\omega} T_j, \quad (12)$$

Η S ακολουθεί ασυμπτωτικά την κανονική κατανομή με μηδενική μέση τιμή και διακύμανση $\text{Var}[S]$ που προκύπτει από τον παρακάτω τύπο.

$$\text{Var}[S] = \sum_{j=1}^{\omega} \text{Var}[T_j] + \sum_{\substack{j,g=1 \\ g \neq j}}^{\omega} \text{Cov}(T_j, T_g), \quad (13)$$

όπου η διακύμανση $\text{Var}[T_j]$ της στατιστικής συνάρτησης Mann-Kendall (ανά εποχή) καθώς και η συνδιακύμανση $\text{Cov}(T_j, T_g)$ τους (ανά δύο) υπολογίζονται όπως παρακάτω.

$$\text{Var}[T_j] = \frac{n_j(n_j - 1)(2n_j + 5) - \sum_{i=1}^m t_i(t_i - 1)(2t_i + 5)}{18}, \quad (14)$$

- n_j είναι ο αριθμός των παρατηρήσεων (μετά την αφαίρεση των missing values) της εποχής j ,
- m είναι ο αριθμός των ομάδων δεδομένων που υπάρχει αλληλεξάρτηση,
- t_i εκφράζει το αριθμητικό μέγεθος της κάθε ομάδας i^{th} που υπάρχει αλληλεξάρτηση.

$$Cov(T_j T_g) = \left[S_{jg} + 4 \sum_{m=1}^n R_{mj} R_{mg} - n(n_j + 1)(n_g + 1) \right] / 3, \quad (15)$$

- n_j, n_g είναι ο αριθμός των παρατηρήσεων (μετά την αφαίρεση των missing values) της εποχής και i και j αντίστοιχα,
- $S_{jg} = \sum_{m < n} \text{sign}[(Z_{nj} - Z_{mj})(Z_{ng} - Z_{mg})]$. (16)
- R είναι ο πίνακας όπου οι παρατηρήσεις (εξαιρουμένων των missing values) αντιπροσωπεύονται διατεταγμένες ανά εποχή, ως στοιχεία. Έτσι η διάταξη R_{mj} , του m^{th} στοιχείου της εποχής i^{th} , υπολογίζεται όπως παρακάτω,

$$R_{mj} = \left[n_j + 1 + \sum_{k=1}^n \text{sign}(x_{mj} - x_{kj}) \right] / 2 \quad (17)$$

όπου $\text{sign}(x_{mj} - x_{kj})$ ορίζεται ως μηδέν εάν οι παρατηρήσεις x_{mj} ή x_{kj} δεν υπάρχουν (missing values), ώστε η διάμεσος $(n_j + 1)/2$ να ορίσει τις μη υπάρχουσες παρατηρήσεις (missing values).

4.6 Εφαρμογή του Ελέγχου Mann-Kendall σε πολλαπλούς σταθμούς μέτρησης

Η αντιμετώπιση σετ δεδομένων που προέρχονται από διαφορετικούς σταθμούς μέτρησης, που ανήκουν ακόμη και σε διαφορετική ευρύτερη περιοχή είναι ο «κανόνας» στην αντιμετώπιση προβλημάτων που αφορούν ατμοσφαιρικούς ρύπους. Η στατιστική ανάλυση που βασίζεται στους μέσους όρους των σταθμών είναι επισφαλής, ενώ η παράλειψη της ενσωμάτωσης μιας ποιοτικής μεταβλητής που να αντιπροσωπεύει τους διαφορετικούς σταθμούς μέτρησης στη στατιστική συνάρτηση ελέγχου οδηγεί στην απώλεια πολύτιμης πληροφορίας. Ιδιαίτερα εάν ληφθεί υπόψη το φαινόμενο της μεταφοράς των ρύπων σε μεγάλες αποστάσεις, γίνεται εύκολα αντιληπτή η χρησιμότητα και η αναγκαιότητα του *Ελέγχου Mann-Kendall* σε πολλαπλούς σταθμούς μέτρησης, που βασίζεται στον Έλεγχο

Αθροισμάτων Συνδιακύμανσης (*Covariance-Sum test*) (βλέπε **Claudia Libiseller, (2002)**. Trend Testing in the presence of covariates, IMPACT 12, *Linköeping University, Dept of Mathematics*).

Ο έλεγχος βασίζεται στην ίδια λογική και θεωρία με τον Εποχικό Έλεγχο Mann-Kendall. Η στατιστική συνάρτηση ελέγχου είναι η V και εκφράζεται από το άθροισμα των (εποχικών) στατιστικών συναρτήσεων Mann-Kendall διαιρεμένο με το άθροισμα των αντιστοίχων διακυμάνσεων και των συνδιακυμάνσεων μεταξύ τους. Η V ορίζεται και υπολογίζεται από τον παρακάτω τύπο:

$$V = \mathbf{1}^T \mathbf{S} / \mathbf{1}^T \mathbf{\Gamma} \mathbf{1}, \quad (18)$$

- $\mathbf{\Gamma}$ συμβολίζει τον πίνακα Διακύμανσης-Συνδιακύμανσης των (εποχικών) συναρτήσεων ελέγχου Mann-Kendall,
- $\mathbf{1}$ συμβολίζει ένα διάνυσμα με όλα τα στοιχεία του μοναδιαία.

4.7 Έλεγχος Mann-Kendall παρουσία συνδιακυμαινόμενων μεταβλητών ή Μερικός Έλεγχος (partial test) Mann-Kendall

Όπως έχει ήδη αναφερθεί επανειλημμένα οι μετεωρολογικές συνθήκες και η χημεία της ατμόσφαιρας πολύ συχνά μπορούν να επηρεάσουν έντονα τις τιμές των ατμοσφαιρικών ρύπων. Όταν αυτό συμβαίνει η πληρέστερη στατιστική ανάλυση για την εύρεση της τάσης των ρύπων, επιτυγχάνεται με τη εφαρμογή του *Μερικού Ελέγχου Mann-Kendall*. Ο προαναφερόμενος έλεγχος έχει προταθεί έτσι ώστε η συνάρτηση ελέγχου για την εύρεση της τάσης να υπολογίζεται «κάτω» από την υπόθεση της συνάρτησης ελέγχου Mann-Kendall των συνδιακυμαινόμενων μεταβλητών. Στη συνέχεια παρουσιάζεται αναλυτικά ο *Μερικός έλεγχος Mann-Kendall* (βλέπε **Claudia Libiseller, Anders Grimvall (2002)**. Performance of Partial Mann-Kendall Tests for Trend Detection in the Presence of Covariates, *Linköeping University, Dept of Mathematics*).

1. Το πρώτο βήμα είναι ο διαχωρισμός των μεταβλητών (απόκρισης και επεξηγηματικών) με βάση την ποιοτική μεταβλητή της εποχής, εφόσον βέβαια είναι επιθυμητό να συμπεριληφθεί η εποχικότητα στην ανάλυση.
2. Για κάθε αλληλουχία δεδομένων, όπως προκύπτουν από το πρώτο βήμα υπολογίζεται το διάνυσμα $\mathbf{T}$,

$$\mathbf{T} = [\mathbf{T}_\alpha, \mathbf{T}_\beta]^T \quad (18)$$

- το διάνυσμα $\mathbf{T}$ ακολουθεί ασυμπτωτικά πολυμεταβλητή κανονική κατανομή με μέσο $\boldsymbol{\mu}$ και διακύμανση $\boldsymbol{\Gamma}$ όπως παρακάτω:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_\alpha \\ \boldsymbol{\mu}_\beta \end{bmatrix}_{\delta \times 1} \quad \text{και} \quad \boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\Gamma}_{\alpha\alpha} & \boldsymbol{\Gamma}_{\alpha\beta} \\ \boldsymbol{\Gamma}_{\beta\alpha} & \boldsymbol{\Gamma}_{\beta\beta} \end{bmatrix}_{\delta \times \delta}. \quad (19)$$

- όπου τα α και β αντιπροσωπεύουν τη μεταβλητή απόκρισης και την ομάδα των επεξηγηματικών μεταβλητών αντίστοιχα,
- Ο πίνακας $\boldsymbol{\Gamma}$ είναι ο πίνακας των διακυμάνσεων και των συνδιακυμάνσεων μεταξύ των μεταβλητών

Η στατιστική συνάρτηση ελέγχου εναλλακτικά μπορεί να υπολογιστεί ως άθροισμα ανά εποχή, σταθμό μέτρησης είτε μόνο για την μεταβλητή απόκρισης είτε και για την μεταβλητή απόκρισης και τις επεξηγηματικές μεταβλητές

Η - υπό την υπόθεση $\mathbf{S}_\beta = \mathbf{s}_\beta$ - , κατανομή της $\mathbf{S}_\alpha$ είναι πολυμεταβλητή κανονική με μέσο και διακύμανση όπως παρακάτω:

$$E(\mathbf{S}_\alpha | \mathbf{S}_\beta = \mathbf{s}_\beta) = \boldsymbol{\mu}_\alpha + \boldsymbol{\Gamma}_{\alpha\beta} \boldsymbol{\Gamma}_{\beta\beta}^{-1} (\mathbf{s}_\beta - \boldsymbol{\mu}_\beta) \quad (20)$$

$$Var(\mathbf{S}_\alpha | \mathbf{S}_\beta = \mathbf{s}_\beta) = \boldsymbol{\Gamma}_{\alpha\alpha} - \boldsymbol{\Gamma}_{\alpha\beta} \boldsymbol{\Gamma}_{\beta\beta}^{-1} \boldsymbol{\Gamma}_{\beta\alpha} \quad (21)$$

Εάν η μηδενική υπόθεση είναι αληθής τότε δεν υπάρχει τάση στις μεταβλητές και επομένως ισχύει,

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_\alpha \\ \boldsymbol{\mu}_\beta \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad (22)$$

κατά συνέπεια ο παραπάνω τύπος για την αναμενόμενη τιμή απλοποιείται και γίνεται,

$$E(\mathbf{S}_\alpha | \mathbf{S}_\beta = \mathbf{s}_\beta) = \boldsymbol{\Gamma}_{\alpha\beta} \boldsymbol{\Gamma}_{\beta\beta}^{-1} \mathbf{s}_\beta \quad (23)$$

4.8 Εφαρμογή του Ελέγχου Mann-Kendall - Χρήση του προγραμματιστικού εργαλείου IMPACT «MULTIMK/PARTIALMK»

Το πρόγραμμα (project) με την επωνυμία IMPACT εκπονήθηκε από το Τμήμα Μαθηματικών του Πανεπιστημίου του Linköping (LIU) στη Σουηδία. Η επιδότηση έγινε από την Ευρωπαϊκή Ένωση στα πλαίσια του ερευνητικού προγράμματος IST. Το IMPACT περιλαμβάνει δύο προγραμματιστικά εργαλεία, τα οποία εξειδικεύονται στον προσδιορισμό της τάσης σε ρυπαντές του φυσικού περιβάλλοντος σε διαφορετικές περιπτώσεις. Στην παρούσα εργασία εφαρμόζεται το MULTIMK/PARTIALMK που αφορά τη μελέτη των ρυπαντών της ατμόσφαιρας. Το δεύτερο προγραμματιστικό εργαλείο το FLOWNORM αφορά τη μελέτη της των ρυπαντών σε ποταμούς. Τα δύο προγραμματιστικά εργαλεία του IMPACT βασίζονται σε μακροεντολές της Visual Basic εφαρμοσμένες στο Excel.

Τελικός στόχος των παραπάνω προγραμματιστικών εργαλείων είναι ο προσδιορισμός της τάσης των φυσικών διακυμάνσεων (natural fluctuations) – τα επίπεδα των ρύπων στην ατμόσφαιρα– διαχωρίζοντας την επίδραση των ανθρωπογενών παραγόντων (human impact). Το προγραμματιστικό εργαλείο MULTIMK/PARTIALMK που αφορά την έρευνα της ατμοσφαιρικής ρύπανσης βασίζεται στην εφαρμογή του μη παραμετρικού ελέγχου Mann-Kendall για την εύρεση της τάσης (*Πολυμεταβλητός Έλεγχος Mann-Kendall για την εύρεση της τάσης και Μερικός Έλεγχος Mann-Kendall*).

Η χρησιμότητα του MULTIMK/PARTIALMK, ως εργαλείου στήριξης, για τον προσδιορισμό της ποιότητας του ατμοσφαιρικού φυσικού περιβάλλοντος είναι μεγάλη καθώς δίνει τη δυνατότητα του ποιοτικού ελέγχου των επιπέδων της ρύπανσης στην ατμόσφαιρα. Έτσι μπορεί να γίνει η σωστή εκτίμηση των ορίων-στόχων σε κάθε περιοχή και η σωστή αξιολόγηση των μετρήσεων των ρύπων από τους σταθμούς ελέγχου. Δίνεται η δυνατότητα της ενσωμάτωσης πληροφορίας που αφορά την εποχικότητα, την εισαγωγή γεωγραφικής ποιοτικής μεταβλητής – σταθμός μέτρησης – κλπ. Επίσης δίνεται η δυνατότητα του ποσοτικού

προσδιορισμού της επίδρασης των μετεωρολογικών παραμέτρων που αποτελούν πολύ σημαντική παράμετρο στα επίπεδα των τιμών ρύπων.

Τα αποτελέσματα (output) που παρέχει το εργαλείο MULTIMK/PARTIALMK προσφέρονται για την εξαγωγή συμπερασμάτων για την ύπαρξη ή μη τάσης σε κάθε ρύπο. Επίσης δίνεται η δυνατότητα της ανίχνευσης της ύπαρξης εποχικότητας (season) καθώς και διαφοροποίησης ανά περιοχή μικρότερη (site) ή ευρύτερη (plot) στην τάση του κάθε ρύπου. Επίσης με το μερικό έλεγχο MK (partial test MK) μπορεί να προσδιοριστεί η τάση ενός ρύπου σε σχέση με κάποια άλλη μεταβλητή που υπάρχει εξάρτηση, η μεταβλητή αυτή μπορεί να είναι είτε ρύπος είτε κάποια μετεωρολογική μεταβλητή. Επίσης δίνεται η δυνατότητα για την ανίχνευση των εξαρτήσεων μεταξύ των μεταβλητών από τους πίνακες με τις συνδιακυμάνσεις και τις συσχετίσεις μεταξύ των ρύπων του ελέγχου Mann-Kendall.

Στους πίνακες που παρουσιάζονται στη συνέχεια παραθέτονται τα αποτελέσματα του ελέγχου Mann-Kendall όπως εφαρμόστηκε στα μηνιαία δεδομένα για τα έτη 1997 έως 2001, με τη χρήση του MULTIMK/PARTIALMK. Τα αποτελέσματα έχουν προκύψει από την εφαρμογή του ελέγχου Mann-Kendall με 4 διαφορετικούς συνδυασμούς προκειμένου να εξαχθούν κατά το δυνατό πληρέστερα συμπεράσματα. Έτσι σε όλους τους υπό εξέταση ρύπους εφαρμόστηκε ο έλεγχος Mann-Kendall για την εύρεση της τάσης εισάγοντας την ποιοτική μεταβλητή της εποχής (μήνα). Ιδιαίτερα για το όζον και τα οξείδια του αζώτου εφαρμόστηκε ο μερικός έλεγχος Mann-Kendall (partial test), αφού υπάρχει σχέση επεξηγηματικής μεταβλητής (explanatory variable) και μεταβλητής απόκρισης (response variable). Η εφαρμογή των δύο παραπάνω ελέγχων έγινε, τόσο με την άθροιση των εποχικών αποτελεσμάτων όσο και χωρίς αυτήν.

Στον πίνακα 6 παρουσιάζονται τα αποτελέσματα για την ύπαρξη ή μη τάσης (αύξουσας ή φθίνουσας) όπως αυτά προκύπτουν από την εφαρμογή του

μονομεταβλητού ελέγχου Mann-Kendall. Στον πίνακα 7 δίδονται τα αντίστοιχα αποτελέσματα από την εφαρμογή του εποχικού ελέγχου Mann-Kendall στα ίδια δεδομένα, δηλαδή με την εισαγωγή της εποχικότητας ως ποιοτικής μεταβλητής. Θετική τιμή της τυποποιημένης στατιστικής συνάρτησης ελέγχου **MK-Stat** – ακολουθεί την $N(0,1)$ – υποδηλώνει την ύπαρξη αύξουσας τάσης ενώ αρνητική τιμή της συνάρτησης υποδηλώνει την ύπαρξη φθίνουσας τάσης. Επίσης στους πίνακες εμφανίζονται ο αριθμός των δεδομένων ανά έλεγχο (Nonmiss), η τιμή της στατιστικής συνάρτησης (Test Stat), η τυποποιημένη διακύμανση (Std. Dev.), επίσης – όπου υπάρχει – και η ποιοτική μεταβλητή που εκφράζει την εποχικότητα.

Στους πίνακες 8 και 9 παρουσιάζονται με αντίστοιχο τρόπο με τους 6 και 7 πίνακες, τα αποτελέσματα από την εφαρμογή του μερικού ελέγχου Mann-Kendall με το όζον (O_3) ως μεταβλητή απόκρισης και τα οξείδια του αζώτου NO και NO_2 ως επεξηγηματικές μεταβλητές. Η ερμηνεία των αποτελεσμάτων διαφέρει από την περίπτωση εφαρμογής του μονομεταβλητού ή εποχικού ελέγχου Mann-Kendall στην επιπρόσθετη πληροφορία για την μεταβλητή απόκρισης που δίνει το p-value και είναι η πιθανότητα να παρατηρηθεί η συγκεκριμένη τιμή της MK-Stat ή μεγαλύτερη της εάν είναι θετική και μικρότερη της εάν είναι αρνητική.

Πολυμεταβλητός Έλεγχος (Multivariate) Mann-Kendall

(Μηνιαία δεδομένα για όλους τους σταθμούς μέτρησης του Λεκανοπεδίου της Αττικής 1997-2001)

Πίνακας 5(α)

Συνδυαστικοί Έλεγχοι (Combined Tests)

o3				
	Nonmiss	Test Stat	Std.Dev.	MK-Stat
All	60	-8	22.15	-0.36
co				
	Nonmiss	Test Stat	Std.Dev.	MK-Stat
All	60	-58	34.38	-1.69
no				
	Nonmiss	Test Stat	Std.Dev.	MK-Stat
All	60	-18	12.81	-1.41

no2				
	Nonmiss	Test Stat	Std.Dev.	MK-Stat
All	60	-47	25.53	-1.84
nox				
	Nonmiss	Test Stat	Std.Dev.	MK-Stat
All	48	-13	9.43	-1.38
so2				
	Nonmiss	Test Stat	Std.Dev.	MK-Stat
All	60	-99	44.79	-2.21

Πίνακας 5(β)

Πολυμεταβλητοί Στατιστικοί Έλεγχοι (Univariate Test Statistics)

o3				
	Nonmiss	Test Stat	Std.Dev.	MK-Stat
1	5	7	3.96	1.77
2	5	1	3.96	0.25
3	5	-9	3.96	-2.27
4	5	-3	3.96	-0.76
5	5	1	3.96	0.25
6	5	-5	3.96	-1.26
7	5	-1	3.96	-0.25
8	5	-1	3.96	-0.25
9	5	-3	3.96	-0.76
10	5	9	3.96	2.27
11	5	1	3.96	0.25
12	5	-5	3.96	-1.26
co				
	Nonmiss	Test Stat	Std.Dev.	MK-Stat
1	5	-8	4.08	-1.96
2	5	-9	3.96	-2.27
3	5	-8	4.08	-1.96
4	5	-2	4.08	-0.49
5	5	0	4.08	0.00
6	5	-6	4.08	-1.47
7	5	-7	3.96	-1.77
8	5	-5	3.96	-1.26
9	5	-3	3.96	-0.76
10	5	-1	3.96	-0.25
11	5	-3	3.96	-0.76
12	5	-6	4.08	-1.47

no				
	Nonmiss	Test Stat	Std.Dev.	MK-Stat
1	5	-7	3.96	-1.77
2	5	-7	3.96	-1.77
3	5	-1	3.96	-0.25
4	5	1	3.96	0.25
5	5	-1	3.96	-0.25
6	5	7	3.96	1.77
7	5	-3	3.96	-0.76
8	5	-1	3.96	-0.25
9	5	1	3.96	0.25
10	5	-9	3.96	-2.27
11	5	5	3.96	1.26
12	5	-3	3.96	-0.76

no2				
	Nonmiss	Test Stat	Std.Dev.	MK-Stat
1	5	-9	3.96	-2.27
2	5	-3	3.96	-0.76
3	5	9	3.96	2.27
4	5	1	3.96	0.25
5	5	-9	3.96	-2.27
6	5	-9	3.96	-2.27
7	5	-9	3.96	-2.27
8	5	3	3.96	0.76
9	5	-3	3.96	-0.76
10	5	-7	3.96	-1.77
11	5	-3	3.96	-0.76
12	5	-8	3.83	-2.09

nox				
	Nonmiss	Test Stat	Std.Dev.	MK-Stat
1	4	-4	2.94	-1.36
2	4	-2	2.94	-0.68
3	4	2	2.94	0.68
4	4	6	2.94	2.04
5	4	-5	2.77	-1.81
6	4	-2	2.94	-0.68
7	4	-4	2.94	-1.36
8	4	2	2.94	0.68
9	4	0	2.94	0.00
10	4	-4	2.94	-1.36
11	4	0	2.94	0.00
12	4	-2	2.94	-0.68

so2				
	Nonmiss	Test Stat	Std.Dev.	MK-Stat
1	5	-6	4.08	-1.47
2	5	-9	3.96	-2.27
3	5	-10	4.08	-2.45
4	5	-10	4.08	-2.45
5	5	-10	4.08	-2.45
6	5	-8	4.08	-1.96
7	5	-8	4.08	-1.96
8	5	-6	4.08	-1.47
9	5	-7	3.96	-1.77
10	5	-8	4.08	-1.96
11	5	-8	4.08	-1.96
12	5	-9	3.96	-2.27

Πίνακας 5(γ)

Μερικός Έλεγχος (Partial Test) Mann-Kendall

(Μηνιαία δεδομένα για όλους του σταθμούς μέτρησης του Λεκανοπεδίου της Αττικής 1997-2001)

no				
	Nonmiss	Test Stat	Std.Dev.	MK-Stat
All	60	-18	12.81	
no2				
	Nonmiss	Test Stat	Std.Dev.	MK-Stat
All	60	-47	25.53	-1.84
o3				
	Nonmiss	Test Stat	Std.Dev.	MK-Stat
All	60	-8	22.15	-0.36

Partial Mann-Kendall Test	
MK Stat of response variable	-8
Conditional Mean	-24.01
Conditional Std. Dev.	16.85
Partial MK Stat.	0.95
p-value	0.171

Πίνακας 6(α)

Πίνακας Διακυμάνσεων Συνδιακυμάνσεων του Στατιστικού Ελέγχου Mann-Kendall
(Var-Covar Matrix of the Mann-Kendall Statistics)

	o3	co	no	no2	nox	so2
	All	All	All	All	All	All
o3	490.67	300.00	178.67	354.67	88.00	293.33
co	300.00	1182.00	256.67	608.33	42.33	1245.67
no	178.67	256.67	164.00	286.00	71.33	365.33
no2	354.67	608.33	286.00	651.67	181.67	998.00
nox	88.00	42.33	71.33	181.67	89.00	259.33
so2	293.33	1245.67	365.33	998.00	259.33	2006.33

Πίνακας 6(β)

Πίνακας Συσχετίσεων του Στατιστικού Ελέγχου Mann-Kendall
Correlation Matrix of the Mann-Kendall Statistics

	o3	co	no	no2	nox	so2
	All	All	All	All	All	All
o3	1.000	0.394	0.630	0.627	0.421	0.296
co	0.394	1.000	0.583	0.693	0.131	0.809
no	0.630	0.583	1.000	0.875	0.590	0.637
no2	0.627	0.693	0.875	1.000	0.754	0.873
nox	0.421	0.131	0.590	0.754	1.000	0.614
so2	0.296	0.809	0.637	0.873	0.614	1.000