

ΥΠΟΥΡΓΕΙΟ ΠΕΡΙΒΑΛΛΟΝΤΟΣ, ΧΩΡΟΤΑΞΙΑΣ &
ΔΗΜΟΣΙΩΝ ΕΡΓΩΝ
ΓΕΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑ ΔΗΜΟΣΙΩΝ ΕΡΓΩΝ
Δ/ΝΣΗ ΕΡΓΩΝ ΥΔΡΕΥΣΗΣ & ΑΠΟΧΕΤΕΥΣΗΣ

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΟΜΕΑΣ ΥΔΑΤΙΚΩΝ ΠΟΡΩΝ - ΥΔΡΑΥΛΙΚΩΝ
& ΘΑΛΑΣΣΙΩΝ ΕΡΓΩΝ

MINISTRY OF ENVIRONMENT, REGIONAL
PLANNING & PUBLIC WORKS
GENERAL SECRETARIAT OF PUBLIC WORKS
SECRETARIAT OF WATER SUPPLY & SEWAGE

NATIONAL TECHNICAL UNIVERSITY OF ATHENS
DIVISION OF WATER RESOURCES - HYDRAULIC
& MARITIME ENGINEERING

ΕΡΕΥΝΗΤΙΚΟ ΕΡΓΟ
ΕΚΤΙΜΗΣΗ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗ ΤΩΝ ΥΔΑΤΙΚΩΝ
ΠΟΡΩΝ ΤΗΣ ΣΤΕΡΕΑΣ ΕΛΛΑΔΑΣ

ΦΑΣΗ Β

ΤΕΥΧΟΣ 11
ΑΝΑΒΑΘΜΙΣΗ ΤΟΥ ΥΠΟΛΟΓΙΣΤΙΚΟΥ
ΠΕΡΙΒΑΛΛΟΝΤΟΣ ΓΙΑ ΤΗΝ
ΕΠΕΞΕΡΓΑΣΙΑ ΥΔΡΟΛΟΓΙΚΩΝ
ΔΕΔΟΜΕΝΩΝ

RESEARCH PROJECT
EVALUATION AND MANAGEMENT OF THE
WATER RESOURCES OF STEREA HELLAS

PHASE B

VOLUME 11
UPGRADE OF THE COMPUTATIONAL
ENVIRONMENT FOR THE
HYDROLOGICAL DATA PROCESSING

ΣΥΝΤΑΞΗ: Α. ΜΑΝΕΤΑΣ, Δ. ΚΟΥΤΣΟΓΙΑΝΝΗΣ
ΕΠΙΣΤΗΜΟΝΙΚΟΣ ΥΠΕΥΘΥΝΟΣ: Θ. ΞΑΝΘΟΠΟΥΛΟΣ
ΚΥΡΙΟΣ ΕΡΕΥΝΗΤΗΣ: Δ. ΚΟΥΤΣΟΓΙΑΝΝΗΣ

BY: A. MANETAS, D. KOUTSOYANNIS
SCIENTIFIC DIRECTOR: TH. XANTHOPOULOS
PRINCIPAL INVESTIGATOR: D. KOUTSOYANNIS

ΑΘΗΝΑ - ΣΕΠΤΕΜΒΡΙΟΣ 1995

ATHENS - SEPTEMBER 1995

Αναβάθμιση του υπολογιστικού περιβάλλοντος για την επεξεργασία υδρολογικών δεδομένων

Περιεχόμενα

1. Εισαγωγή	1
1.1 Αντικείμενο των τεύχους - Ιστορικό	1
1.2 Διάρθρωση των τεύχους	2
2. Γενική περιγραφή του νέου υπολογιστικού περιβάλλοντος	3
2.1 Το σύστημα ΥΔΡΟΣΚΟΠΙΟ	3
2.1.1 Το δίκτυο συνεργασίας του ΥΔΡΟΣΚΟΠΙΟΥ	3
2.1.2 Το επικοινωνιακό δίκτυο του ΥΔΡΟΣΚΟΠΙΟΥ	5
2.1.3 Τεχνικά χαρακτηριστικά του ΥΔΡΟΣΚΟΠΙΟΥ	5
2.1.4 Σύνολα δεδομένων	8
2.2 Βασικές αρχές της μετάβασης στο νέο υπολογιστικό περιβάλλον	10
2.3 Εργασίες υποδομής για την αναβάθμιση του υπολογιστικού περιβάλλοντος	11
2.4 Πρόσβαση στο σύστημα ΥΔΡΟΣΚΟΠΙΟ μέσω του λογιστικού πακέτου EXCEL	11
2.5 Σύντομη περιγραφή τυπικών στατιστικών υπολογισμών με το λογιστικό πακέτο EXCEL	12
3. Ανάπτυξη πρόσθετων στατιστικών υπολογιστικών διαδικασιών για το λογιστικό πακέτο EXCEL	13
3.1 Θεωρητική περιγραφή των διαδικασιών	13
3.1.1 Δοκιμή Kendall	13
3.1.2 Δοκιμή τοπικών ακροτάτων (Turning Point)	13
3.1.3 Δοκιμή Kruskal Wallis	14
3.1.4 Δοκιμή Φασματικής Ανάλυσης (Spectral Analysis)	15
3.2 Οδηγίες χρήσης της νέας στατιστικής βιβλιοθήκης	18
3.2.1 Kruskal-Wallis test	18
3.2.2 Turning point test	18
3.2.3 Kendall test	19
3.2.4 Spectral analysis	19
3.2.5 Gamma generator	20
4. Αναφορές	21

1. Εισαγωγή

1.1 Αντικείμενο του τεύχους - Ιστορικό

Το τεύχος αυτό αναφέρεται στην αναβάθμιση του υπολογιστικού περιβάλλοντος για την αρχειοθέτηση και επεξεργασία υδρομετεωρολογικών δεδομένων και συνοδεύεται από βιβλιοθήκη προγραμμάτων ηλεκτρονικού υπολογιστή για τη στατιστική επεξεργασία δεδομένων. Συγκεκριμένα παρέχει γενική περιγραφή του νέου υπολογιστικού περιβάλλοντος και των εργασιών που απαιτήθηκαν για τη μετάβαση σε αυτό, καθώς και οδηγίες χρήσης και θεωρητική και τεχνική τεκμηρίωση ορισμένων πρόσθετων υπολογιστικών διαδικασιών που χρειάστηκε να αναπτυχθούν. Η αναβάθμιση του υπολογιστικού περιβάλλοντος αναφέρεται στη μετάβαση από το περιβάλλον DOS, στο οποίο αναπτύχθηκαν όλα τα προγράμματα της φάσης Α του ερευνητικού έργου *Εκτίμηση και διαχείριση των υδατικών πόρων της Στερεάς Ελλάδας* στο συνδυασμό των συστημάτων UNIX και WINDOWS στα οποία αναπτύχθηκαν όλα τα προγράμματα της φάσης Β του ίδιου ερευνητικού έργου.

Το εν λόγω ερευνητικό έργο ανατέθηκε και χρηματοδοτήθηκε από τη Διεύθυνση Υδρευσης και Αποχέτευσης του ΥΠΕΧΩΔΕ (απόφαση Δ6/21609/8-9-1993) σε ερευνητική ομάδα του Τομέα Υδατικών Πόρων, Υδραυλικών και Θαλάσσιων Έργων του ΕΜΠ με επιστημονικό υπεύθυνο τον καθηγητή Θ. Ξανθόπουλο και συντονιστή τον επίκουρο καθηγητή Δ. Κουτσογιάννη. Η συγκεκριμένη εργασία την οποία καλύπτει το τεύχος αυτό προδιαγράφεται στο Παράρτημα της απόφασης ανάθεσης (άρθρα 2.2.1 και 2.3.1: Αναβάθμιση των βάσεων δεδομένων από λειτουργικό σύστημα MS-DOS σε UNIX).

Το νέο υπολογιστικό περιβάλλον καθορίστηκε κατά βάση από το σύστημα ΥΔΡΟΣΚΟΠΙΟ το οποίο αποτελεί τη βάση της Εθνικής Τράπεζας Υδρολογικής και Μετεωρολογικής Πληροφορίας της χώρας και αναπτύχθηκε στα πλαίσια του κοινοτικού ερευνητικού προγράμματος STRIDE ΕΛΛΑΣ 1992-94. Ανάδοχος του ερευνητικού έργου αυτού ήταν και πάλι ο Τομέας Υδατικών Πόρων, Υδραυλικών και Θαλάσσιων Έργων του ΕΜΠ με επιστημονικό υπεύθυνο τον Δ. Κουτσογιάννη, ενώ συμμετείχε σε αυτό και η διεύθυνση Δ7 του ΥΠΕΧΩΔΕ, καθώς και το σύνολο σχεδόν των δημόσιων φορέων της χώρας που ασχολούνται με την καταγραφή και αξιοποίηση της υδρολογικής και μετεωρολογικής πληροφορίας της χώρας. Λόγω της ουσιαστικής σύνδεσης των εργασιών που περιγράφονται στο τεύχος αυτό με τη λειτουργία του συστήματος ΥΔΡΟΣΚΟΠΙΟ, η γνώση της τελευταίας θεωρείται προϋπόθεση για την κατανόηση του νέου υπολογιστικού περιβάλλοντος. Για το λόγο αυτό έχει συμπεριληφθεί στα παραδοτέα του παρόντος ερευνητικού προγράμματος και το *Εγχειρίδιο Χρήσης* του ΥΔΡΟΣΚΟΠΙΟΥ.

1.2 Διάρθρωση του τεύχους

Το τεύχος περιλαμβάνει, εκτός από αυτό το εισαγωγικό κεφάλαιο, άλλα δύο κεφάλαια. Στο Κεφάλαιο 2 γίνεται η γενική περιγραφή του νέου υπολογιστικού συστήματος και των εργασιών που απαιτήθηκαν για τη μετάβαση σε αυτό. Στο Κεφάλαιο 3 περιγράφεται η ανάπτυξη πρόσθετου λογισμικού για τη στατιστική επεξεργασία υδρολογικών δεδομένων που έγινε στα πλαίσια του παρόντος ερευνητικού έργου, και δίνεται η κατάλληλη θεωρητική τεκμηρίωση.

2. Γενική περιγραφή του νέου υπολογιστικού περιβάλλοντος

2.1 Το σύστημα ΥΔΡΟΣΚΟΠΙΟ

Το ερευνητικό έργο ΥΔΡΟΣΚΟΠΙΟ αποτέλεσε μια κοινή προσπάθεια πολλών φορέων που ασχολούνται με τη συλλογή, την επεξεργασία και αξιοποίηση σε ερευνητικούς ή επιχειρησιακούς σκοπούς των μετεωρολογικών, υδρολογικών και υδρογεωλογικών πληροφοριών. Ο στόχος της προσπάθειας αυτής ήταν η δημιουργία σύγχρονης πληροφοριακής υποδομής για τον υδρολογικό κύκλο στην Ελλάδα. Πιο συγκεκριμένα το έργο αποσκοπούσε στην οργάνωση και συστηματοποίηση της μετεωρολογικής, υδρολογικής και υδρογεωλογικής πληροφορίας με αξιοποίηση των δυνατοτήτων που παρέχουν οι σύγχρονες μέθοδοι της πληροφορικής και των τηλεπικοινωνιών. Το έργο εντάχθηκε στα πλαίσια του κοινοτικού προγράμματος STRIDE ΕΛΛΑΣ με βασικό ανάδοχο το ΕΜΠ.

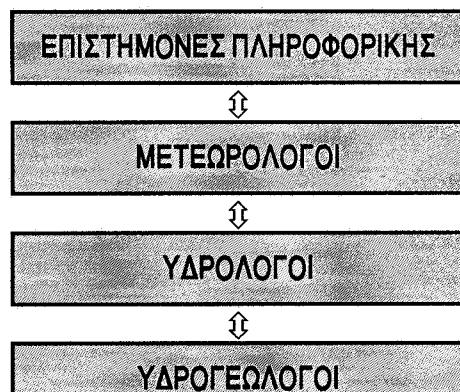
2.1.1 Το δίκτυο συνεργασίας του ΥΔΡΟΣΚΟΠΙΟΥ

Βασική στρατηγική του ΥΔΡΟΣΚΟΠΙΟΥ είναι η δημιουργία συνεργιστικού δικτύου υπηρεσιών και επιστημόνων σε δύο επίπεδα. Στο οριζόντιο επίπεδο δημιουργείται δίκτυο συνεργασίας ανάμεσα σε όλες τις ομάδες φορέων που ασχολούνται υπό διάφορες οπτικές γωνίες με τα ατμοσφαιρικά και υδρολογικά φαινόμενα: Πανεπιστήμια-Πολυτεχνεία, Μετεωρολογικές και Υδρολογικές Υπηρεσίες, Υπουργεία, Ερευνητικά Κέντρα και χρήστες των δεδομένων. Στο κατακόρυφο επίπεδο πραγματοποιείται συνεργασία επιστημόνων που ασχολούνται με τους επιμέρους επιστημονικούς κλάδους: Επιστήμονες Πληροφορικής, Μετεωρολόγοι, Υδρολόγοι, Υδρογεωλόγοι. (Σχήμα 1). Οι βασικοί κανόνες που τέθηκαν για το συνεργιστικό δίκτυο είναι δύο, η αυτονομία και η ανταλλαγή. Συγκεκριμένα, κάθε φορέας διατηρεί την αυτονομία του ως προς την αποθήκευση και διαχείριση των δικών του δεδομένων και την εξυπηρέτηση των δικών του επιχειρησιακών και ερευνητικών σκοπών. Παράλληλα, όλοι φορείς συνεργάζονται στο μέτρο των δυνατοτήτων τους, των αναγκών τους και των ιδιαιτεροτήτων τους ανταλλάσσοντας μεταξύ τους δεδομένα, εμπειρία και τεχνογνωσία.

Η παραπάνω στρατηγική επιδίωξη επιτυγχάνεται με δύο ειδών μέτρα: οργανωτικά και τεχνολογικά. Από οργανωτική άποψη το οριζόντιο επίπεδο υλοποιείται από την Εκτελεστική Επιτροπή που αποτελείται από τους υπεύθυνους των 14 Ερευνητικών Ομάδων (μια από κάθε φορέα που συμμετέχει, βλ. Σχ. 1α). Αντίστοιχα, το κατακόρυφο επίπεδο υλοποιείται από τέσσερις Τομεακές Επιστημονικές Επιτροπές, μια για κάθε επιστημονικό κλάδο του Σχήματος 1β, σε καθεμιά από τις οποίες συμμετέχουν επιστήμονες της αντίστοιχης ειδικότητας αδιακρίτως του φορέα προέλευσής τους.

ΠΑΝΕΠΙΣΤΗΜΙΑΚΟΙ ΦΟΡΕΙΣ	↔	ΜΕΤΕΩΡΟΛΟΓΙΚΕΣ & ΥΔΡΟΛΟΓΙΚΕΣ ΥΠΗΡΕΣΙΕΣ	↔	ΥΠΟΥΡΓΕΙΑ	↔	ΕΠΙΧΕΙΡΗΣΙΑΚΟΙ & ΕΡΕΥΝΗΤΙΚΟΙ ΦΟΡΕΙΣ
Εθνικό Μετσόβιο Πολυτεχνείο, Τομέας Υδατικών Πόρων, Υδραυλικών & Θαλάσσιων Έργων (ΕΜΠ/ΤΥΠΥΘΕ)		Εθνική Μετεωρολογική Υπηρεσία (ΕΜΥ)		Υπουργείο Βιομηχανίας, Ενέργειας & Τεχνολογίας, Διεύθυνση Υδατικού Δυναμικού & Φυσικών Πόρων (ΥΒΕΤ/ΔΥΔΦΠ)		Επιχείρηση Ύδρευσης & Αποχέτευσης Πρωτεύουσας (ΕΥΔΑΠ)
Πολυτεχνική Σχολή Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, Τομέας Υδραυλικής & Τεχνικής Περιβάλλοντος (ΠΣΑΠΘ/ΤΥΤΠ)		Δημόσια Επιχείρηση Ηλεκτρισμού, Διεύθυνση Ανάπτυξης Υδροηλεκτρικών Έργων (ΔΕΗ/ΔΑΥΕ)		Υπουργείο Γεωργίας, Γενική Διεύθυνση Εγγειοβελτιωτικών Έργων & Γεωργικών Διαρθρώσεων (ΥΠΓΕ/ΓΔΕΕΓΔ)		Εθνικό Κέντρο Έρευνας Φυσικών Επιστημών "Δημόκριτος" (ΕΚΕΦΕ "Δ")
Εθνικό & Καποδιστριακό Πανεπιστήμιο Αθηνών, Τομέας Φυσικής Εφαρμογών (ΕΚΠΑ/ΤΦΕ)		Εθνικό Αστεροσκοπείο Αθηνών, Ινστιτούτο Μετεωρολογίας & Φυσικής Ατμοσφαιρικού Περιβάλλοντος (ΕΑΑ/ΙΜΦΑΠ)		Υπουργείο Περιβάλλοντος, Χωροταξίας & Δημόσιων Έργων, Διεύθυνση Εγγειοβελτιωτικών Έργων (ΥΠΕΧΩΔΕ/ΔΕΕ)		Κέντρο Ανανεώσιμων Πηγών Ενέργειας (ΚΑΠΕ)
Πολυτεχνική Σχολή Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, Ενεργειακός Τομέας (ΠΣΑΠΘ/ΕΤ)						Ελληνική Εταιρεία Τοπικής Ανάπτυξης & Αυτοδιοίκησης (ΕΕΤΑΑ)

(α)



(β)

Σχήμα 1: Οριζόντιο (α) και κατακόρυφο (β) επίπεδο του συνεργιστικού δικτύου στο ΥΔΡΟΣΚΟΠΙΟ. Σε κάθε βαθμίδα του κατακόρυφου επιπέδου συμμετέχουν επιστήμονες από όλους τους φορείς του έργου.

2.1.2 Το επικοινωνιακό δίκτυο του ΥΔΡΟΣΚΟΠΙΟΥ

Από τεχνολογική άποψη το συνεργιστικό δίκτυο υλοποιείται με την ανάπτυξη ενός δικτύου από ανεξάρτητους υπολογιστές τύπου Σταθμού Εργασίας (Workstation) οι οποίοι συνδέονται μέσω τηλεπικοινωνιακών γραμμών υψηλής ταχύτητας. Πάνω στο δίκτυο αυτό λειτουργεί σύστημα σχεσιακής κατανεμημένης βάσης δεδομένων που επιτρέπει τη διαφανή ως προς τη θέση των δεδομένων προσπέλαση. Με τον τρόπο αυτό εξυπηρετούνται κατά τον καλύτερο δυνατό τρόπο οι κανόνες της αυτονομίας και ανταλλαγής. Το όλο σύστημα συνοδεύεται από ειδικό λογισμικό που αναπτύχθηκε στα πλαίσια του έργου.

Η πρώτη φάση του έργου, που ξεκίνησε το πρώτο εξάμηνο του 1992 και ολοκληρώθηκε στις αρχές του 1994 αποσκοπούσε στην δημιουργία του συνεργιστικού δικτύου, την προμήθεια και εγκατάσταση του υλικού εξοπλισμού (hardware) και του λογισμικού υποδομής (commercial system software), την ανάπτυξη του λογισμικού εφαρμογής (application software), και την πιλοτική εφαρμογή. Σε δεύτερη φάση του έργου θα εισαχθούν τα δεδομένα στην βάση, ενώ προβλέπεται και περαιτέρω ανάπτυξη του λογισμικού εφαρμογής.

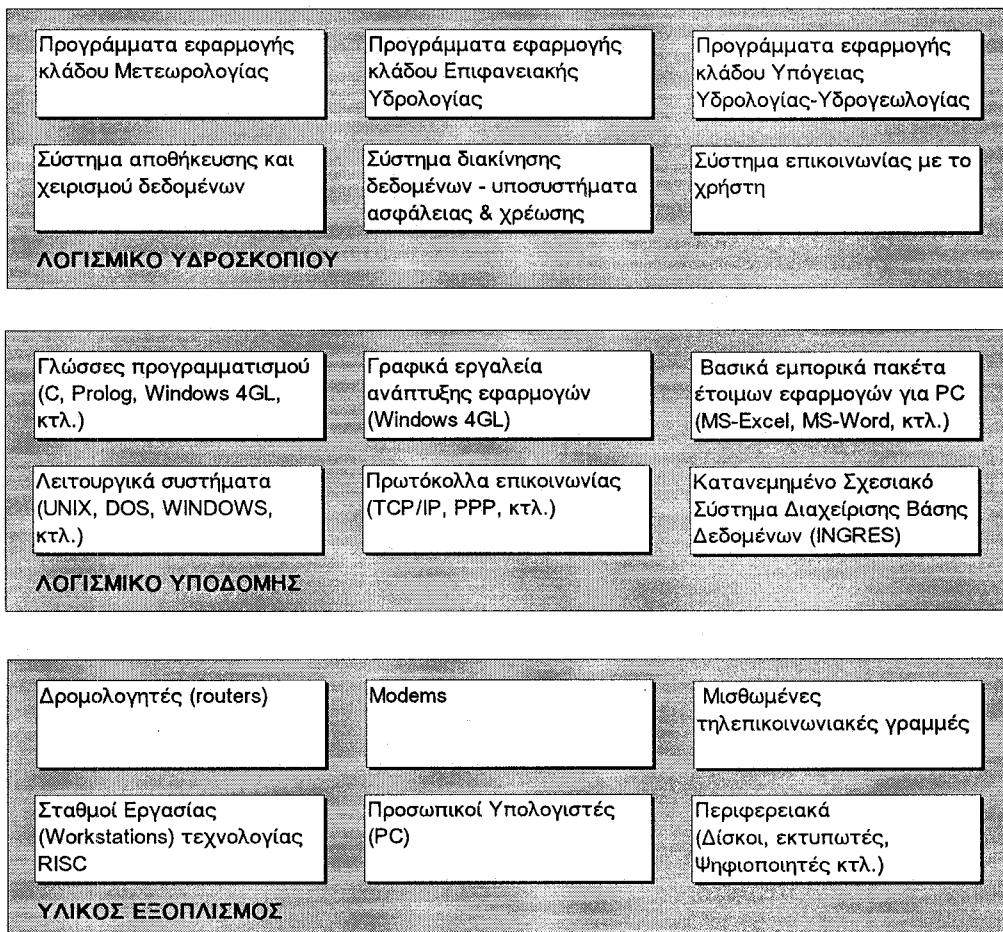
2.1.3 Τεχνικά χαρακτηριστικά του ΥΔΡΟΣΚΟΠΙΟΥ

Από τεχνική άποψη το ΥΔΡΟΣΚΟΠΙΟ αναπτύσσεται σε τρία στρώματα που φαίνονται στο Σχήμα 2. Το κατώτερο στρώμα είναι ο υλικός εξοπλισμός του συστήματος και περιλαμβάνει δύο υποστρώματα, τα υπολογιστικά συστήματα και τη δικτυακή υποδομή. Το ενδιάμεσο στρώμα περιλαμβάνει το λογισμικό υποδομής του συστήματος το οποίο έχει εμπορική προέλευση. Τέλος, το ανώτερο στρώμα είναι το λογισμικό που αναπτύσσεται για την υλοποίηση του ΥΔΡΟΣΚΟΠΙΟΥ και περιλαμβάνει δύο υποστρώματα, το εξειδικευμένο λογισμικό του συστήματος και τα προγράμματα εφαρμογών.

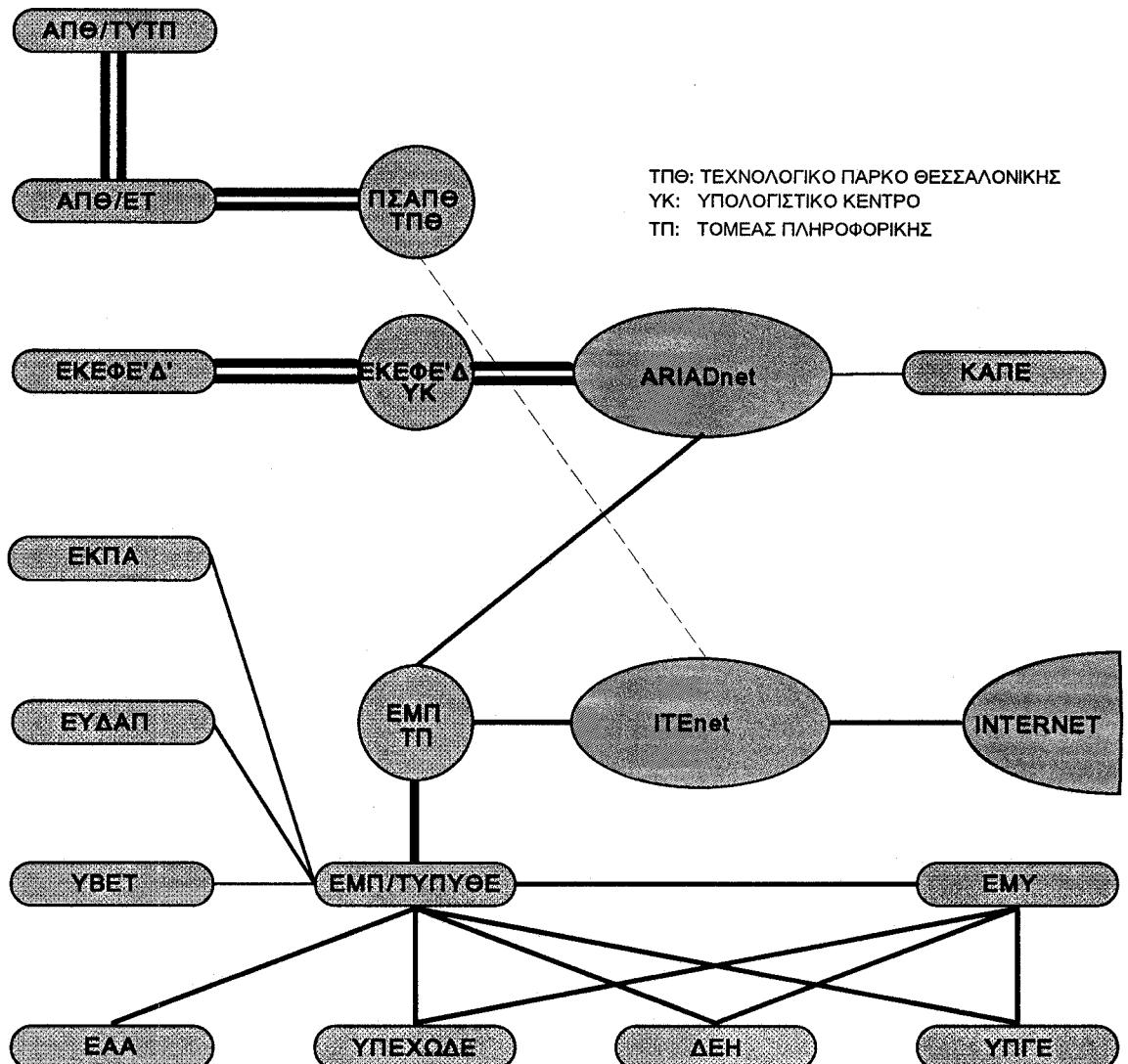
Το ΥΔΡΟΣΚΟΠΙΟ περιέλαβε στην πρώτη του φάση 13 κύριους κόμβους, όπου κάθε κόμβος υλοποιείται από ένα ή περισσότερους Σταθμούς Εργασίας (Workstations). Σε κάθε κόμβο υπάρχει ένα τοπικό δίκτυο (LAN) από Προσωπικούς Υπολογιστές (PC) που διασυνδέονται με τον Σταθμό Εργασίας μέσω ethernet. Οι κύριοι κόμβοι διασυνδέονται με ιδιωτικό δίκτυο ευρείας περιοχής (WAN), η τοπολογία του οποίου προέκυψε μετά από τεχνικο-οικονομική βελτιστοποίηση και φαίνεται στο Σχήμα 3. Κατά βάση χρησιμοποιούνται για τη διασύνδεση μισθωμένες δημόσιες γραμμές είτε ψηφιακές ταχύτητας 64 kbit/s είτε αναλογικές ταχύτητας 14.4 kbit/s. Για μικρές αποστάσεις η επικοινωνία γίνεται είτε με ethernet είτε με ιδιωτικές τηλεφωνικές γραμμές ταχύτητας 144 kbit/s. Τέλος για πολύ μεγάλες αποστάσεις χρησιμοποιούνται τα υπάρχοντα ακαδημαϊκά δίκτυα μεταγωγής δεδομένων. Σε κάθε κόμβο που περιλαμβάνει τηλεφωνική σύνδεση, η επικοινωνία εξυπηρετείται με δρομολογητή (router).

Τον κορμό του λογισμικού υποδομής αποτελεί το Σχεσιακό Σύστημα Διαχείρισης Βάσης Δεδομένων (Relational Data Base Management System - RDBMS) INGRES που λειτουργεί πάνω από λειτουργικό σύστημα UNIX. Το UNIX διαθέτει αρκετά πλεονεκτήματα, όπως τη διαθεσιμότητα του για μεγάλο εύρος τύπων υπολογιστών, την εξασφάλιση μεταφερσιμότητας και την ενσωμάτωση λειτουργιών δικτύωσης. Παράλληλα, το σύστημα RDBMS εξασφαλίζει την κατανεμημένη λειτουργία της Τράπεζας Δεδομένων και τη διαφανή αλλά και ασφαλή και ελεγχόμενη πρόσβαση προς οποιοδήποτε κόμβο της Τράπεζας.

Το εξειδικευμένο λογισμικό του ΥΔΡΟΣΚΟΠΙΟΥ εξασφαλίζει τις βασικές εσωτερικές λειτουργίες της Τράπεζας δεδομένων, τη διακίνηση των δεδομένων μεταξύ των διάφορων κόμβων και την επικοινωνία του συστήματος με το χρήστη. Η επικοινωνία αυτή γίνεται κατά βάση σε γραφικό περιβάλλον (Graphical User Interface), φιλικό προς το χρήστη, και περιλαμβάνει σύστημα οπτικοποίησης των δεδομένων. Τέλος, τα προγράμματα εφαρμογής, τα οποία επίσης λειτουργούν σε γραφικό περιβάλλον, εξυπηρετούν την εισαγωγή και ανάκτηση, τον έλεγχο, τη συμπλήρωση και την επεξεργασία των δεδομένων, καθώς και την παραγωγή δευτερογενών δεδομένων.



Σχήμα 2: Λειτουργικές συνιστώσες ΥΔΡΟΣΚΟΠΙΟΥ



ΥΠΟΜΝΗΜΑ

- — — — — ΜΙΣΘΩΜΕΝΗ ΑΝΑΛΟΓΙΚΗ ΓΡΑΜΜΗ 9.6 kbit/s
- — — — — ΜΙΣΘΩΜΕΝΗ ΑΝΑΛΟΓΙΚΗ ΓΡΑΜΜΗ 14.4 kbit/s
- — — — — ΜΙΣΘΩΜΕΝΗ ΨΗΦΙΑΚΗ ΓΡΑΜΜΗ 64 kbit/s
- — — — — ΙΔΙΩΤΙΚΗ ΓΡΑΜΜΗ 144 kbit/s
- — — — — ΓΡΑΜΜΗ ETHERNET
- — — — — ΚΟΜΒΟΣ ΥΔΡΟΣΚΟΠΙΟΥ
- — — — — ΚΟΜΒΟΣ ΆΛΛΟΥ ΔΙΚΤΥΟΥ
- — — — — ΑΚΑΔΗΜΑΪΚΟ ΔΙΚΤΥΟ ΔΕΔΟΜΕΝΩΝ

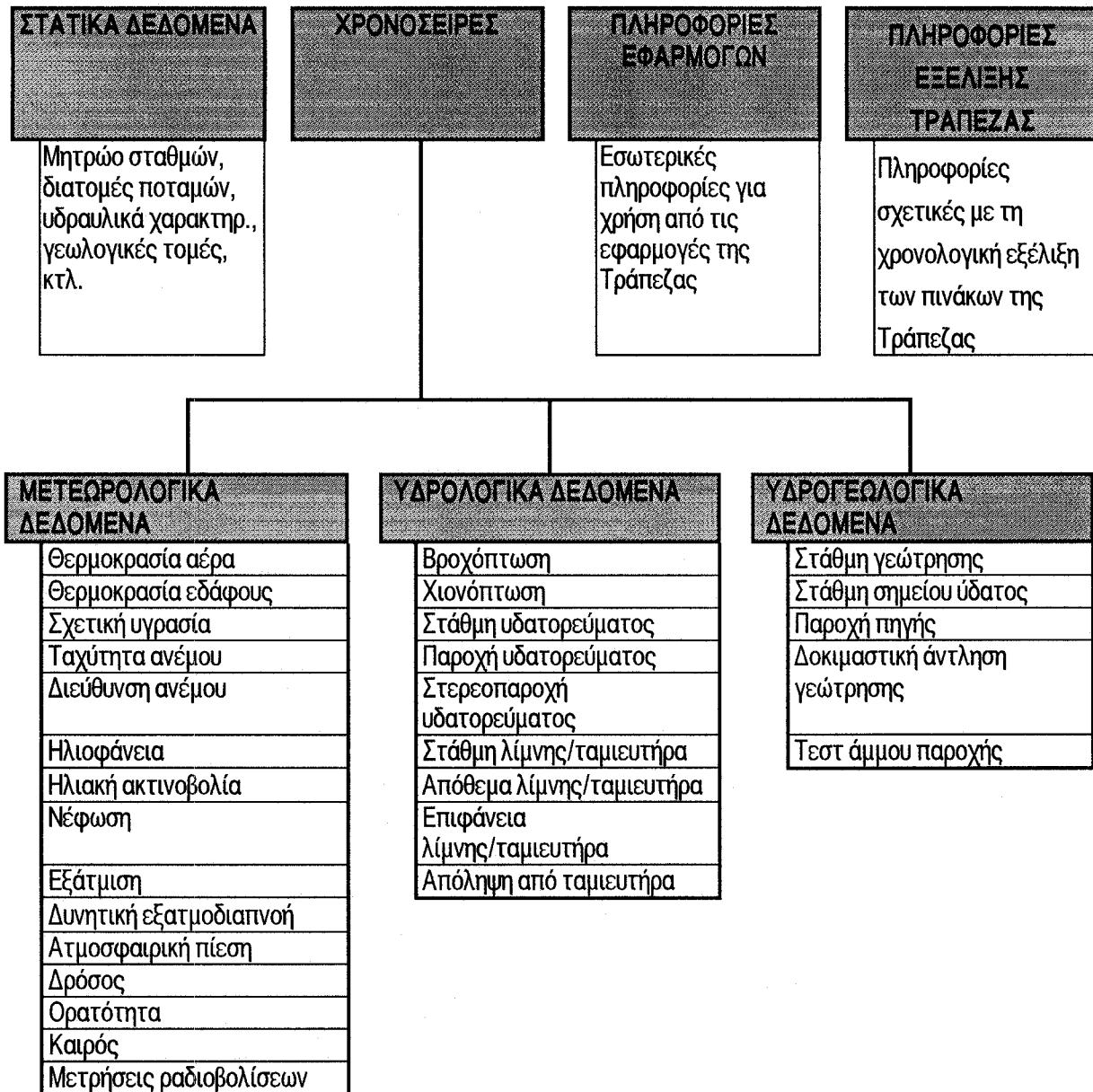
Σχήμα 3: Τοπολογία δικτύου μεταγωγής δεδομένων

2.1.4 Σύνολα δεδομένων

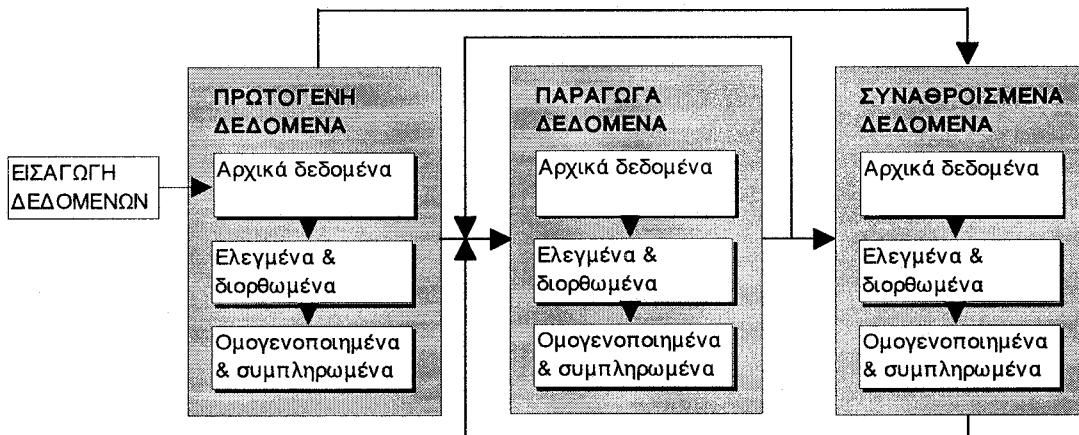
Στην πρώτη του φάση το ΥΔΡΟΣΚΟΠΙΟ μπορεί να διαχειριστεί δεδομένα από κάθε είδους μετεωρολογικές μετρήσεις και παρατηρήσεις, δεδομένα μετρήσεων υδρολογικών σταθμών επιφάνειας καθώς και δεδομένα υπογείων υδάτων. Διαχειρίζεται επίσης παράγωγα δεδομένα που προκύπτουν από τα πρωτογενή δεδομένα μετρήσεων με βάση απλούς και καθιερωμένους μετασχηματισμούς. Παραδείγματα τέτοιων παράγωγων δεδομένων είναι η εξάτμιση ή εξατμοδιαπνοή κατά Penman που υπολογίζεται βάσει μιας σειράς μετεωρολογικών δεδομένων και η παροχή υδατορεύματος που υπολογίζεται με βάση τη στάθμη μέσω των καμπυλών στάθμης-παροχής.

Τα δεδομένα της Τράπεζας ανήκουν κατά βάση σε δύο κύριες κατηγορίες: τα στατικά δεδομένα και τις χρονοσειρές. Στα στατικά δεδομένα ανήκουν π.χ. χαρακτηριστικά μετρητικών σταθμών, οι διατομές υδατορευμάτων οι γεωλογικές τομές κτλ. Παρά την ονομασία τους, ορισμένα από τα στατικά δεδομένα αλλάζουν με την πάροδο του χρόνου. Οι αλλαγές αυτές αποτυπώνονται στην Τράπεζα. Οι χρονοσειρές αποθηκεύονται κατ' αρχήν ως πρωτογενή δεδομένα στο χρονικό βήμα (σταθερό ή μεταβλητό) στο οποίο είναι διαθέσιμες, το οποίο ξεκινά από 1 min και φτάνει τον 1 μήνα ή το 1 έτος. Το σύνολο των πρωτογενών δεδομένων που υπάρχουν μέχρι και σήμερα και πρόκειται να εισαχθούν στην Τράπεζα είναι περίπου 450 000 000. Στην Τράπεζα διατίθενται και συναθροισμένες χρονοσειρές με μεγαλύτερο από το αρχικό χρονικό βήμα. Στο Σχήμα 4 φαίνονται συνοπτικά οι κατηγορίες δεδομένων της Τράπεζας. Ας σημειωθεί ότι, εκτός από τις δύο κύριες κατηγορίες δεδομένων τις Τράπεζας, υπάρχουν και δύο δευτερεύουσες κατηγορίες πληροφοριών που δεν περιλαμβάνουν υδρομετεωρολογικά δεδομένα αλλά εσωτερικές πληροφορίες της Τράπεζας και χρονολογικές πληροφορίες για την εξέλιξη των άλλων αντικειμένων της Τράπεζας.

Η Τράπεζα διαθέτει ακόμη λειτουργίες ελέγχου και συμπλήρωσης δεδομένων. Έτσι δημιουργούνται τρία επίπεδα δεδομένων. Το πρώτο επίπεδο περιλαμβάνει τα αρχικά δεδομένα, το δεύτερο τα ελεγμένα και διορθωμένα δεδομένα και το τρίτο τα ομογενοποιημένα και συμπληρωμένα (χωρίς κενά) δεδομένα. Όλες οι διαδικασίες μετάβασης από χαμηλότερο επίπεδο προς υψηλότερο περιέχονται στο λογισμικό της Τράπεζας. Ο χρήστης των δεδομένων έχει την ευχέρεια να επιλέξει ένα από τα τρία επίπεδα δεδομένων, ανάλογα με το σκοπό της εφαρμογής του. Οι τυπικές μορφές επεξεργασίας των δεδομένων δίνονται στο Σχήμα 5.



Σχήμα 4: Κατηγορίες δεδομένων του ΥΔΡΟΣΚΟΠΙΟΥ



Σχήμα 5: Τυπικές μορφές επεξεργασίας δεδομένων

2.2 Βασικές αρχές της μετάβασης στο νέο υπολογιστικό περιβάλλον

Στην πρώτη φάση του παρόντος ερευνητικού έργου αναπτύχθηκαν προγράμματα αρχειοθέτησης και επεξεργασίας υδρολογικών δεδομένων σε μηνιαία βάση, τα οποία λειτουργούσαν σε προσωπικούς υπολογιστές και σε υπολογιστικό περιβάλλον DOS. Θεωρήθηκε όμως απαραίτητο και προβλέφθηκε κατάλληλα στην απόφαση ανάθεσης του ερευνητικού έργου να γίνει η μετάβαση στο νέο υπολογιστικό περιβάλλον UNIX/INGRES του ΥΔΡΟΣΚΟΠΙΟΥ. Αυτό προϋπέθετε την προμήθεια και εγκατάσταση του απαραίτητου υπολογιστικού εξοπλισμού (Σταθμός εργασίας και λογισμικό υποδομής) στη Διεύθυνση Έργων Ύδρευσης και Αποχέτευσης του ΥΠΕΧΩΔΕ, καθώς και την κατάλληλη σύνδεση με το δίκτυο του ΥΔΡΟΣΚΟΠΙΟΥ. Με τη σύνδεση αυτή εξασφαλίζεται η δυνατότητα ταυτόχρονης προσπέλασης και ενημέρωσης της βάσης δεδομένων της Στερεάς Ελλάδας και από τον κόμβο του ΕΜΠ και από τον κόμβο του ΥΠΕΧΩΔΕ, ενώ το ίδιο ισχύει για την ανάπτυξη και εκτέλεση των προγραμμάτων. Εξ άλλου είναι δυνατή η πρόσβαση και σε δεδομένα άλλων φορέων που συμμετέχουν στο ΥΔΡΟΣΚΟΠΙΟ, στο βαθμό που αυτά έχουν εισαχθεί στην κατανεμημένη βάση.

Η χρήση του νέου υπολογιστικού εξοπλισμού δεν αποκλείει και την παράλληλη χρήση προσωπικών υπολογιστών σε συνεργασία με τον τοπικό σταθμό εργασίας αλλά και το γενικό δίκτυο του ΥΔΡΟΣΚΟΠΙΟΥ. Οι προσωπικοί υπολογιστές είναι μάλιστα πιο εύχρηστοι αλλά και οικονομικότεροι σε πολλές περιπτώσεις επεξεργασίας δεδομένων, λόγω της δυνατότητας χρήσης καθιερωμένων εμπορικών πακέτων λογισμικού. Τα τελευταία διαρκώς τελειοποιούνται και διατίθενται σε φτηνές τιμές λόγω της μεγάλης αγοράς που υπάρχει σε προγράμματα προσωπικών υπολογιστών.

Πολλές από τις απαραίτητες υπολογιστικές διαδικασίες επεξεργασίας υδρολογικών δεδομένων υπάρχουν ενσωματωμένες στο σύστημα του ΥΔΡΟΣΚΟΠΙΟΥ. Αρκετές όμως δεν είχαν προβλεφθεί στην πρώτη φάση του ΥΔΡΟΣΚΟΠΙΟΥ και έτσι έπρεπε να γίνει εκ νέου προγραμματισμός τους λαμβάνοντας υπόψη τις απαιτήσεις του νέου υπολογιστικού περιβάλλοντος. Συγκεκριμένα οι διαδικασίες αυτές αναφέρονται στη στατιστική επεξεργασία των υδρολογικών χρονοσειρών.

Παίρνοντας υπόψη τα παραπάνω γενικά δεδομένα, αποφασίστηκε (με τη σύμφωνη γνώμη των μελών της ομάδας εργασίας του ΥΠΕΧΩΔΕ) να αναπτυχθούν οι απαιτούμενες συμπληρωματικές υπολογιστικές εφαρμογές σε προσωπικούς υπολογιστές με υπολογιστικό περιβάλλον WINDOWS. Πιο συγκεκριμένα επελέγη ως πλατφόρμα ανάπτυξης των εφαρμογών το λογιστικό πακέτο EXCEL της Microsoft, το οποίο αποτελεί ένα από τα πιο εξελιγμένα και διαδεδομένα πακέτα του είδους του και προσφέρει πολλές δυνατότητες επεξεργασίας και γραφικής απεικόνισης

δεδομένων. Οι εφαρμογές αναπτύχθηκαν με χρήση των γλωσσών προγραμματισμού C και Visual Basic και τελικά ενσωματώθηκαν υπό μορφή βιβλιοθήκης στην πλατφόρμα του EXCEL. Προηγουμένως εξασφαλίστηκε η δυνατότητα συνεργασίας του EXCEL με την κατανεμημένη βάση δεδομένων του ΥΔΡΟΣΚΟΠΙΟΥ, όπως περιγράφεται στην ενότητα 2.4.

2.3 Εργασίες υποδομής για την αναβάθμιση του υπολογιστικού περιβάλλοντος

Συνοπτικά πραγματοποιήθηκαν οι ακόλουθες εργασίες υποδομής για τη μετάβαση στο νέο υπολογιστικό περιβάλλον, οι οποίες και προβλέπονταν στην απόφαση ανάθεσης του ερευνητικού έργου:[†]

- α. Έγινε προμήθεια υπολογιστικού εξοπλισμού και λογισμικού υποδομής, και εγκατάστασή τους στη Διεύθυνση Έργων Ύδρευσης και Αποχέτευσης του ΥΠΕΧΩΔΕ. Συγκεκριμένα αγοράστηκε σταθμός εργασίας τύπου Hewlett-Packard 9000/710 συνοδευμόμενος από λειτουργικό σύστημα UNIX (HP-UX). Αγοράστηκε επίσης και εγκαταστάθηκε το Σύστημα Διαχείρισης Βάσης Δεδομένων INGRES και το Σύστημα Γεωγραφικών Πληροφοριών ARC/INFO.
- β. Έγινε on-line σύνδεση του Τομέα Υδατικών Πόρων του ΕΜΠ και της Διεύθυνσης Έργων Ύδρευσης και Αποχέτευσης του ΥΠΕΧΩΔΕ, μέσω του δικτύου ΥΔΡΟΣΚΟΠΙΟΥ και ειδικότερα της μισθωμένης ψηφιακής γραμμής ΕΜΠ-ΥΠΕΧΩΔΕ.
- γ. Μεταφέρθηκε το σύνολο των πρωτογενών δεδομένων της Α' φάσης στο νέο υπολογιστικό περιβάλλον.
- δ. Ολοκληρώθηκε η εκμάθηση και προσαρμογή της ομάδας εργασίας του ΥΠΕΧΩΔΕ στο νέο υπολογιστικό περιβάλλον.
- ε. Έγιναν διάφορες προσαρμογές των προγραμμάτων του ΥΔΡΟΣΚΟΠΙΟΥ για τις ανάγκες του παρόντος ερευνητικού έργου.

2.4 Πρόσβαση στο σύστημα ΥΔΡΟΣΚΟΠΙΟ μέσω του λογιστικού πακέτου EXCEL

Η πρόσβαση στη βάση δεδομένων του ΥΔΡΟΣΚΟΠΙΟΥ για την ανάκτηση δεδομένων γίνεται με τη χρήση του προγράμματος Microsoft Query. Πρέπει βεβαίως να έχει προηγηθεί η εγκατάσταση των κατάλληλων οδηγών (drivers) για την επικοινωνία του Microsoft Query με την Ingres. Η διαδικασία της ανάκτησης δεδομένων είναι πολύ απλή και περιλαμβάνει την εκτέλεση της εντολής *Data-Get External Data...* μέσα από το EXCEL. Μόλις εκτελεστεί η εντολή αυτή

[†] Σημειώνεται ότι καμία από τις εργασίες αυτές δεν κοστολογήθηκε στο παρόν ερευνητικό έργο. Το έργο επιβαρύνθηκε μόνο από τη δαπάνη προμήθειας εξοπλισμού.

ενεργοποιείται το Microsoft Query και ζητάει το όνομα της βάσης με την οποία θέλουμε να συνδεθούμε. Μόλις συμπληρωθεί το όνομα εμφανίζονται όλοι οι διαθέσιμοι πίνακες της βάσης και αφού οριστεί κάποιος πίνακας εμφανίζονται τα πεδία του. Αφού επιλεγούν τα επιθυμητά δεδομένα του πίνακα τότε με την εκτέλεση της εντολής *File-Return Data to Microsoft Excel* τα δεδομένα από τη βάση του ΥΔΡΟΣΚΟΠΙΟΥ τοποθετούνται στο φύλλο εργασίας του EXCEL.

2.5 Σύντομη περιγραφή τυπικών στατιστικών υπολογισμών με το λογιστικό πακέτο EXCEL

Στο λογιστικό πακέτο EXCEL υπάρχουν ενσωματωμένες πολλές από τις βασικές λειτουργίες που χρησιμοποιούνται κατά τη στατιστική επεξεργασία υδρολογικών δεδομένων. Η χρήση των λειτουργιών αυτών είναι άμεση, δεδομένου ότι παρέχονται υπό μορφή συναρτήσεων. Παρακάτω αναφέρονται μερικές από τις πιο συνηθισμένες συναρτήσεις:

AVERAGE	Υπολογίζει τη μέση τιμή μιας σειράς δεδομένων.
CHITEST	Επιστρέφει το αποτέλεσμα του ελέγχου προσαρμογής χ^2 .
CORREL	Υπολογίζει το συντελεστή συσχέτισης μεταξύ δυο σειρών.
FDIST	Επιστρέφει την πιθανότητα της κατανομής F.
FINV	Είναι η αντίστροφη συνάρτηση της FDIST.
FTEST	Επιστρέφει το αποτέλεσμα του ελέγχου F.
GAMMADIST	Επιστρέφει την πιθανότητα της κατανομής Γάμα.
GAMMAINV	Είναι η αντίστροφη συνάρτηση της GAMMADIST.
GEOMEAN	Επιστρέφει το γεωμετρικό μέσο όρο.
NORMDIST	Επιστρέφει την πιθανότητα της κανονικής κατανομής.
NORMINV	Είναι η αντίστροφη συνάρτηση της NORMDIST.
SKEW	Επιστρέφει το συντελεστή ασυμμετρίας μιας σειράς δεδομένων.
SLOPE	Επιστρέφει την κλίση της γραμμής παλινδρόμησης.
STDEV, STDEVP	Υπολογίζει την τυπική απόκλιση μιας σειράς δεδομένων.
VAR	Υπολογίζει τη διασπορά μιας σειράς δεδομένων.

Αναλυτικότερη περιγραφή για τις συναρτήσεις καθώς και για το είδος των ορισμάτων τους υπάρχει στο εγχειρίδιο και στο αρχείο βοήθειας του EXCEL.

Όπως προαναφέρθηκε, αρκετές ακόμη λειτουργίες, απαραίτητες για τη στατιστική επεξεργασία υδρολογικών δεδομένων, οι οποίες δεν περιλαμβάνονται στις παραπάνω διαθέσιμες από το EXCEL ενσωματωμένες συναρτήσεις, αναπτύχθηκαν και ενσωματώθηκαν σε αυτό, όπως αναλυτικά περιγράφεται στο επόμενο κεφάλαιο.

3. Ανάπτυξη πρόσθετων στατιστικών υπολογιστικών διαδικασιών για το λογιστικό πακέτο EXCEL

3.1 Θεωρητική περιγραφή των διαδικασιών

3.1.1 Δοκιμή Kendall

Η δοκιμή Kendall ελέγχει την ύπαρξη ή όχι τάσεων, ανοδικών ή πτωτικών, σε μια δεδομένη χρονοσειρά. Σε χρονοσειρά X_1, X_2, \dots, X_v εξετάζονται όλα τα δυνατά ζεύγη παρατηρήσεων X_i, X_j με $j > i$ και έστω P ο συνολικός αριθμός των ζευγών που πληρούν τη σχέση $X_j > X_i$. Ο αριθμός P γίνεται μέγιστος όταν η χρονοσειρά είναι συνεχώς αύξουσα. Αποδεικνύεται ότι για μια χωρίς τάση χρονοσειρά, η αναμενόμενη τιμή της τυχαίας μεταβλητής P είναι

$$E(P) = v(v - 1) / 4 \quad (1)$$

Σύμφωνα με τη δοκιμή υπολογίζεται η ανηγμένη μεταβλητή

$$T = [4P / v(v - 1)] - 1 \quad (2)$$

που έχει αναμενόμενη τιμή $E(T) = 0$ για μια τυχαία διαδοχή της χρονοσειράς. Αποδεικνύεται ότι η διασπορά του T είναι

$$\text{Var}(T) = 2(2v + 5) / 9v(v - 1) \quad (3)$$

και ότι η κατανομή της μεταβλητής του Kendall T / $\text{Var}(T)^{1/2}$ συγκλίνει στην κανονική κατανομή όσο το v μεγαλώνει. Στην περίπτωση του αμφίπλευρου ελέγχου η μηδενική υπόθεση περί μη ύπαρξης τάσης δεν απορρίπτεται για κάποιο επίπεδο σημαντικότητας α , όταν ισχύει η σχέση $-z_{\alpha/2} \leq T / \text{Var}(T)^{1/2} \leq z_{\alpha/2}$ όπου τ η παρατηρημένη τιμή της μεταβλητής T , όπως προκύπτει από το δείγμα, και $z_{\alpha/2}$ η ανηγμένη μεταβλητή της κατανομής Gauss για πιθανότητα υπέρβασης $\alpha/2$. Στην περίπτωση του μονόπλευρου ελέγχου η μηδενική υπόθεση περί μη ύπαρξης πτωτικής ή ανοδικής τάσης δεν απορρίπτεται για κάποιο επίπεδο σημαντικότητας α , όταν ισχύει η σχέση $\tau / \text{Var}(\tau)^{1/2} \geq z_\alpha$ ή $\tau / \text{Var}(\tau)^{1/2} \leq z_\alpha$, αντίστοιχα.

3.1.2 Δοκιμή τοπικών ακροτάτων (Turning Point)

Η δοκιμή αυτή εξετάζει την τυχαιότητα μιας χρονοσειράς (βλ. πχ. Kotegoda, σ. 31). Αν συμβολίσουμε με X_i μια χρονοσειρά με v δεδομένα τότε μπορούμε να ορίσουμε ως τοπικό ακρότατο στη θέση i το σημείο εκείνο για το οποίο ισχύει $X_i > X_{i-1}$ και $X_i > X_{i+1}$ (τοπικό μέγιστο) ή $X_i < X_{i-1}$ και $X_i < X_{i+1}$ (τοπικό ελάχιστο). Γενικώς σε μια

χρονοσειρά τρία διαδοχικά σημεία μπορούν να βρίσκονται σε μια από τις ακόλουθες έξι διαφορετικές καταστάσεις:

- 1) $X_{i-1} > X_i > X_{i+1}$
- 2) $X_{i-1} > X_{i+1} > X_i$
- 3) $X_i > X_{i-1} > X_{i+1}$
- 4) $X_i > X_{i+1} > X_{i-1}$
- 5) $X_{i+1} > X_i > X_{i-1}$
- 6) $X_{i+1} > X_{i-1} > X_i$

Από τα παραπάνω είναι φανερό ότι, αν η διαδοχή είναι τυχαία, τότε η πιθανότητα εμφάνισης τοπικού ακρότατου είναι ίση με $2/3$. Πρέπει να τονιστεί ότι η εμφάνιση τοπικού ακροτάτου δεν είναι δυνατή για τις τιμές X_1 και X_v . Έτσι το αναμενόμενο πλήθος τοπικών ακροτάτων σε μια τυχαία χρονοσειρά είναι $E(P) = 2(v - 2) / 3$. Μπορεί επίσης να δειχτεί ότι $\text{Var}(P) = (16v - 29) / 90$ και συνεπώς η μεταβλητή p μπορεί να εκφραστεί σε κανονικοποιημένη μορφή ως $Z = [P - E(P)] / \sqrt{\text{Var}(P)}$. Επειδή η εμφάνιση πολλών ή ελάχιστων τοπικών ακροτάτων φανερώνει τη μη τυχαιότητα μιας χρονοσειράς, ο έλεγχος της μεταβλητής Z πρέπει να είναι αμφίπλευρος.

3.1.3 Δοκιμή Kruskal Wallis

Με τη δοκιμή αυτή ελέγχουμε αν k το πλήθος ομάδες δεδομένων προέρχονται από πληθυσμούς με ίδια κατανομή. Η εναλλακτική υπόθεση είναι ότι τουλάχιστον μια ομάδα δεδομένων προέρχεται από πληθυσμό με διαφορετική κατανομή. Η δοκιμή είναι μη παραμετρική γιατί δεν κάνει καμία παραδοχή για την κατανομή των πληθυσμών των δειγμάτων.

Συμβολίζουμε με X_{lj} τη μεταβλητή που αντιπροσωπεύει το υπ' αριθμόν l δεδομένο της ομάδας με αριθμό j , έτσι ώστε $1 \leq l \leq v_j$ και $1 \leq j \leq k$, όπου v_j είναι το πλήθος των δεδομένων της ομάδας με αριθμό j . Ενοποιώντας όλες τις ομάδες που περιέχουν τις τιμές x_{lj} αποκτούμε ένα συνολικό δείγμα μεγέθους

$$\mu = \sum_{j=1}^k v_j \quad (4)$$

Με βάση το δείγμα αυτό, καταταγμένο σε αύξουσα σειρά, μπορούμε να αντιστοιχίσουμε αύξοντες αριθμούς ή βαθμούς (ranks) R_{lj} σε όλες τις μ τιμές x_{lj} (Για την περίπτωση που έχουμε ταυτόσημες τιμές x_{lj} χρησιμοποιούμε το μέσο όρο των αντίστοιχων βαθμών). Επανερχόμενοι στα αρχικά επιμέρους δείγματα των ξεχωριστών διαρκειών υπολογίζουμε για κάθε τιμή το μέσο βαθμό

$$\bar{R}_j = \frac{1}{v_j} \sum_{l=1}^k R_{lj} \quad (5)$$

Αν όλα τα επιμέρους δείγματα έχουν την ίδια κατανομή τότε κάθε \bar{R}_j θα πρέπει να βρίσκεται πολύ κοντά στην τιμή $(\mu + 1) / 2$, διαφορετικά οι τιμές \bar{R}_j θα διαφέρουν

σημαντικά μεταξύ τους. Αυτό μας οδηγεί στη χρήση της ακόλουθης στατιστικής συνάρτησης των Kruskal-Wallis (βλ. πχ. Hirsch et al., σ. 17.24), η οποία συνδυάζει τους μέσους βαθμούς από όλες τις επιμέρους ομάδες:

$$H = \frac{12}{\mu(\mu+1)} \sum_{j=1}^k n_j \left(\bar{R}_j - \frac{\mu+1}{2} \right)^2 \quad (6)$$

Ισοδύναμα η παραπάνω γράφεται και με τη μορφή

$$H = \frac{12}{\mu(\mu+1)} \sum_{j=1}^k \frac{\tilde{R}_j^2}{n_j} - 3(\mu+1) \quad (7)$$

όπου \tilde{R}_j το άθροισμα των βαθμών της ομάδας j , δηλαδή $\tilde{R}_j = n_j \bar{R}_j$.

Η στατιστική συνάρτηση H ακολουθεί κατανομή χ^2 με $k-1$ βαθμούς ελευθερίας. Ετσι απορρίπτουμε τη μηδενική υπόθεση και δεχόμαστε την εναλλακτική, σε επίπεδο σημαντικότητας α , όταν ισχύει $H > \chi^2_{1-\alpha}(k-1)$. Κανονικά η δοκιμή είναι εφαρμόσιμη μόνο στην περίπτωση που οι ομάδες έχουν περισσότερα από πέντε στοιχεία.

3.1.4 Δοκιμή Φασματικής Ανάλυσης (Spectral Analysis)

Η δοκιμή της φασματικής ανάλυσης αποσκοπεί στην ανάδειξη κρυμμένων περιοδικοτήτων σε μια χρονοσειρά πεπερασμένου μήκους.

Αν με $x(n)$ συμβολίζουμε μια στάσιμη στοχαστική ανέλιξη σε διακριτό χρόνο n τότε η φασματική συνάρτηση (power spectrum) $P_{xx}(f)$ ορίζεται ως ο διακριτός μετασχηματισμός Fourier της συνάρτησης αυτοσυγχέτισης $r_{xx}(k)$ της $x(n)$, όπου με k συμβολίζουμε τη χρονική υστέρηση. Η φασματική συνάρτηση δίνεται από την εξίσωση

$$P_{xx}(f) = \sum_{k=-\infty}^{\infty} r_{xx}(k) \exp(-j2\pi fk) \quad -\frac{1}{2} \leq f \leq \frac{1}{2} \quad (8)$$

όπου $j = \sqrt{-1}$ και $\exp(ja) = \cos a + j \sin a$.

Δεδομένου ότι η συνάρτηση αυτοσυγχέτισης είναι άρτια (για πραγματική στοχαστική ανέλιξη), ήτοι $r_{xx}(k) = r_{xx}(-k)$, ενώ παράλληλα το ημίτονο είναι περιττή συνάρτηση, ήτοι $\sin(-2\pi fk) = -\sin(2\pi fk)$, οι όροι ημιτόνου στο παραπάνω ανάπτυγμα Fourier αλληλοαναρούνται, οπότε η (8) γράφεται

$$P_{xx}(f) = \sum_{k=-\infty}^{\infty} r_{xx}(k) \cos(2\pi fk) \quad -\frac{1}{2} \leq f \leq \frac{1}{2} \quad (9)$$

Η αυστηρή εφαρμογή της παραπάνω εξίσωσης θα απαιτούσε να είναι γνωστές άπειρες τιμές της συνάρτησης αυτοσυγχέτισης. Αυτό βεβαίως δεν είναι δυνατόν γιατί οι χρονοσειρές που μελετάμε είναι πεπερασμένου μήκους. Έτσι το πρόβλημα που τίθεται είναι ο προσδιορισμός της φασματικής συνάρτησης μιας τυχαίας χρονοσειράς μέσω ενός πεπερασμένου δείγματος. Το παραπάνω πρόβλημα έχει λύση στην περίπτωση όπου η χρονοσειρά θεωρείται εργοδική ως προς την αυτοσυγχέτιση. Βασιζόμενοι στην τελευταία παραδοχή μπορούμε να εκφράσουμε προσεγγιστικά τη συνάρτηση φάσματος ως εξής:

$$\hat{P}_{xx}(f) = \sum_{k=-(\nu-1)}^{\nu-1} \hat{r}_{xx}(k) \cos(2\pi fk) \quad (10)$$

όπου ν είναι το πλήθος των σημείων της χρονοσειράς, ενώ $\hat{r}_{xx}(k)$ είναι η δειγματική εκτίμηση της συνάρτησης αυτοσυγχέτισης που δίνεται από την εξίσωση

$$\begin{aligned} \hat{r}_{xx}(k) = & \\ & \frac{\frac{1}{\nu-k} \sum_{n=0}^{\nu-1-k} x(n)x(n+k) - \frac{1}{(\nu-k)^2} \sum_{n=0}^{\nu-1-k} x(n) \sum_{n=0}^{\nu-1-k} x(n+k)}{\sqrt{\frac{1}{\nu-k} \sum_{n=0}^{\nu-1-k} [x(n)]^2 - \frac{1}{(\nu-k)^2} \left[\sum_{n=0}^{\nu-1-k} x(n) \right]^2} \sqrt{\frac{1}{\nu-k} \sum_{n=0}^{\nu-1-k} [x(n+k)]^2 - \frac{1}{(\nu-k)^2} \left[\sum_{n=0}^{\nu-1-k} x(n+k) \right]^2}} \end{aligned} \quad (11)$$

όπου $k = 0, 1, \dots, \nu - 1$, ενώ για $k = -(v - 1), \dots, -1$ εφαρμόζεται η εξίσωση $\hat{r}_{xx}(k) = r_{xx}(-k)$.

Η παραπάνω εκτίμηση δεν προσεγγίζει ικανοποιητικά την πραγματική συνάρτηση αυτοσυγχέτισης, όταν η υστέρηση παίρνει μεγάλες τιμές, γιατί το πλήθος των στοιχείων της χρονοσειράς $x(n)$ που συμμετέχουν στον υπολογισμό είναι μικρό. Ένας τρόπος για να αντιμετωπιστεί το πρόβλημα αυτό είναι να μειώσουμε το βάρος των αυτοσυχετίσεων μεγάλης υστέρησης που συμμετέχουν στον προσδιορισμό της φασματικής συνάρτησης. Γι' αυτό χρησιμοποιούμε μια συνάρτηση βάρους, την $w(k)$, με την παρακάτω αναλυτική έκφραση [Kay, 1988, σ. 71]

$$w(k) = \begin{cases} 0.54 + 0.46 \cos \frac{\pi k}{\mu} & |k| \leq \mu \\ 0 & |k| > \mu \end{cases} \quad (12)$$

όπου μ ακέραιος που εκφράζει το πλήθος των συντελεστών αυτοσυγχέτισης που χρησιμοποιούμε για τον υπολογισμό των τιμών της φασματικής συνάρτησης. Ο αριθμός μ δεν μπορεί να υπερβαίνει την τιμή $\nu - 1$ (όπου ν το μήκος της χρονοσειράς). Στην πράξη όμως για υπολογιστικούς λόγους συνιστάται να μην υπερβαίνει την τιμή $\nu / 5$ γιατί τότε υπεισέρχονται σφάλματα στον υπολογισμό των

αυτοσυνσχετίσεων όπως αναφέρθηκε και παραπάνω. Έτσι αντί της σχέση (10), για την εκτίμηση $\hat{P}_{xx}(f)$ χρησιμοποιείται η ακόλουθη τροποποιημένη σχέση:

$$\hat{P}_{xx}(f) = \sum_{k=-\mu}^{\mu} w(k) \hat{r}_{xx}(k) \cos(2\pi fk) \quad (13)$$

Η τελευταία σχέση καλείται φασματική συνάρτηση Blackman-Tuckey [Kay, 1988, σ. 80] και είναι αυτή που χρησιμοποιούμε στο πρόγραμμα που εκπονήσαμε.

Ένα πρόβλημα που συχνά προκύπτει αφορά στο στατιστικό έλεγχο της φασματικής συνάρτησης. Έχει αποδειχτεί [Koopmans, 1974· βλ. και Kay, 1988, σ. 81, καθώς και Kottegoda, 1980, σ. 54] ότι για κατάλληλο αριθμό βαθμών ελευθερίας m που δίνεται από τη σχέση

$$m = \frac{2\nu}{\sum_{k=-\mu}^{\mu} w^2(k)} \quad (14)$$

η στατιστική συνάρτηση $m\hat{P}_{xx}(f)/P_{xx}(f)$, όπου η $\hat{P}_{xx}(f)$ είναι η εκτιμημένη φασματική συνάρτηση από την εξίσωση των Blackman-Tuckey (Εξ. (13)) και $P_{xx}(f)$ είναι η αντίστοιχη θεωρητική φασματική συνάρτηση (Εξ. (9)), ακολουθεί (κατά προσέγγιση) κατανομή $\chi^2(m)$. Έτσι, μπορούμε να γράψουμε την ακόλουθη έκφραση πιθανότητας

$$\Pr\left(\chi^2_{\alpha/2}(m) < \frac{m\hat{P}_{xx}(f)}{P_{xx}(f)} \leq \chi^2_{1-\alpha/2}(m)\right) = 1 - \alpha \quad (15)$$

όπου $\Pr(\)$ συμβολίζει πιθανότητα και $1 - \alpha$ είναι συντελεστής εμπιστοσύνης. Συνεπώς, τα όρια εμπιστοσύνης της $P_{xx}(f)$ για βαθμό εμπιστοσύνης $1 - \alpha$ είναι

$$\frac{m\hat{P}_{xx}(f)}{\chi^2_{1-\alpha/2}(m)} \quad \text{και} \quad \frac{m\hat{P}_{xx}(f)}{\chi^2_{\alpha/2}(m)} \quad (16)$$

Εξ άλλου, αυτό μας επιτρέπει να ελέγξουμε την υπόθεση αν η χρονοσειρά αποτελεί λευκό θόρυβο ή όχι, και στη δεύτερη περίπτωση να προσδιορίσουμε τις χαρακτηριστικές περιόδους της. Ο όρος λευκός θόρυβος περιγράφει μια τελείως τυχαία χρονοσειρά, στην οποία δεν υπάρχει καμιά περιοδικότητα. Στην περίπτωση του λευκού θορύβου ισχύει $r_{xx}(k) = 0$ για όλες τις τιμές του k εκτός από την τιμή 0 όπου ισχύει $r_{xx}(0) = 1$. Εύκολα μπορεί να διαπιστωθεί από την εξίσωση (9), ότι στην περίπτωση αυτή θα είναι $P_{xx}(f) = 1$, δηλαδή η φασματική συνάρτηση είναι σταθερή και ίση με 1 για όλες τις συχνότητες (από εκεί προέρχεται και ο όρος λευκός θόρυβος). Αν υποθέσουμε ότι μια χρονοσειρά αποτελεί λευκό θόρυβο, τότε, σύμφωνα με την (15), με πιθανότητα $1 - \alpha$ η εκτίμηση $\hat{P}_{xx}(f)$ θα βρίσκεται ανάμεσα στα όρια

$$\frac{\chi^2_{\alpha/2}(m)}{m} \quad \text{και} \quad \frac{\chi^2_{1-\alpha/2}(m)}{m} \quad (17)$$

Αν σε κάποια συχνότητα f η εκτίμηση $\hat{P}_{xx}(f)$ συμβεί να βρεθεί έξω από αυτά τα όρια, αυτό αποτελεί στατιστική ένδειξη ότι η χρονοσειρά δεν αποτελεί λευκό θόρυβο. Ειδικότερα, αν έχουμε υπέρβαση του άνω ορίου, αυτό δείχνει ότι χρονοσειρά εμφανίζει χαρακτηριστική περιοδικότητα ίση με $1/f$.

3.2 Οδηγίες χρήσης της νέας στατιστικής βιβλιοθήκης

Η βιβλιοθήκη στατιστικών προσθέτει ένα καινούργιο μενού στο EXCEL με το όνομα Statistics. Κάτω απ' αυτό το μενού βρίσκονται οι εντολές:

Kruskal-Wallis test

Turning point test

Kendall test

Spectral analysis

Gamma generator

Παρακάτω ακολουθεί αναλυτική περιγραφή και παραδείγματα χρήσης.

3.2.1 Kruskal-Wallis test

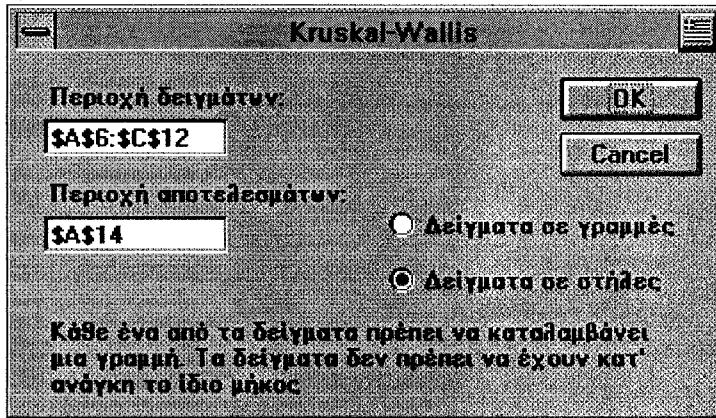
Με την επιλογή της εντολής εμφανίζεται το πλαίσιο διαλόγου που φαίνεται στο Σχήμα 6.

Στο πεδίο *Περιοχή δειγμάτων* πρέπει να συμπληρωθούν το άνω αριστερά και το κάτω δεξιά κελί μεταξύ των οποίων είναι τα δείγματα (ομάδες) για τα οποία μας ενδιαφέρει ο έλεγχος.

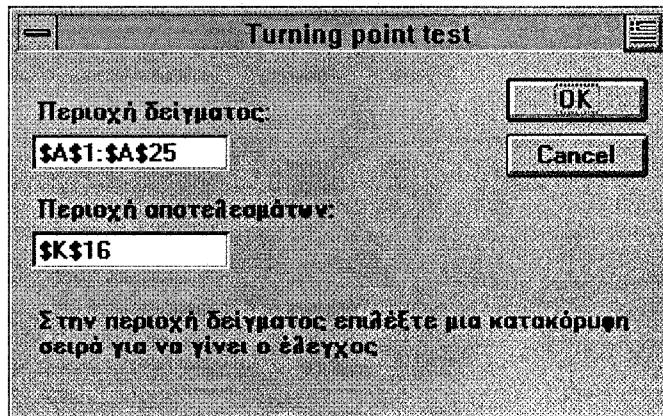
Στο πεδίο *Περιοχή αποτελεσμάτων* πρέπει να συμπληρωθεί το κελί κάτω από το οποίο θα αρχίσουν να εμφανίζονται τ' αποτελέσματα. Οι επιλογές *Δείγματα σε σειρές* και *Δείγματα σε στήλες* καθορίζουν τον τρόπο με τον οποίο έχουν εισαχθεί τα δείγματα στο φύλλο εργασίας. Για να λειτουργήσει ο έλεγχος Kruskal-Wallis πρέπει να υπάρχουν τουλάχιστον δύο δείγματα. Μόλις τελειώσει ο έλεγχος στην περιοχή των αποτελεσμάτων εμφανίζεται ένα κόκκινο κελί, αν ο έλεγχος αποτύχει, ή ένα πράσινο, αν ο έλεγχος πετύχει. Το νούμερο μέσα σε αυτό το κελί είναι το αποτέλεσμα του ελέγχου.

3.2.2 Turning point test

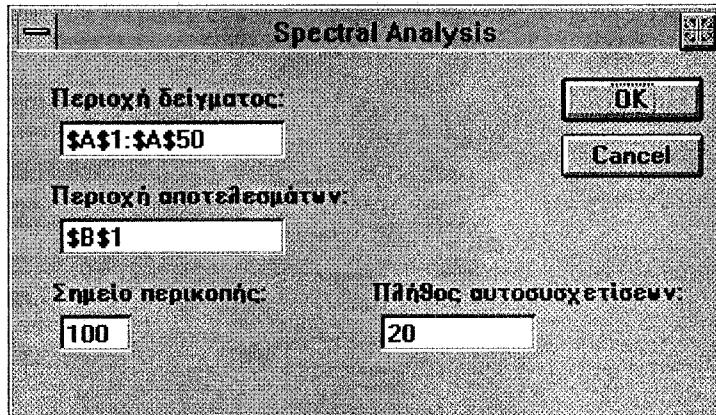
Με την επιλογή της εντολής εμφανίζεται το πλαίσιο διαλόγου του Σχήματος 7. Το πεδίο *Περιοχή δείγματος* περιέχει τα κελιά που καθορίζουν την αρχή και το τέλος του δείγματος προς έλεγχο. Το πεδίο *Περιοχή αποτελεσμάτων* καθορίζει τη θέση στην οποία θα εμφανιστούν τα αποτελέσματα. Μόλις τελειώσει ο έλεγχος στην περιοχή αποτελεσμάτων εμφανίζεται ένα κόκκινο ή πράσινο πεδίο ανάλογα με την αποτυχία ή την επιτυχία του ελέγχου.



Σχήμα 6. Πλαίσιο διαλόγου δοκιμής Kruscal-Wallis.



Σχήμα 7. Πλαίσιο διαλόγου δοκιμής τοπικών ακροτάτων.



Σχήμα 8. Πλαίσιο διαλόγου φασματικής ανάλυσης.

3.2.3 Kendall test

Ο τρόπος λειτουργίας είναι παρόμοιος με τον έλεγχο Turning point.

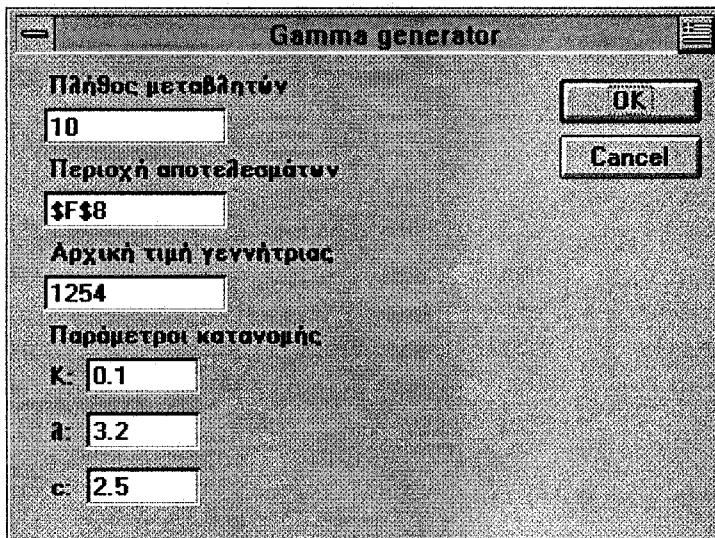
3.2.4 Spectral analysis

Οι παράμετροι του ελέγχου αυτού καθορίζονται από το πλαίσιο διαλόγου του Σχήματος 8. Για τα πεδία Περιοχή δείγματος και Περιοχή αποτελεσμάτων ισχύουν τα

ίδια όπως παραπάνω. Το πεδίο *Σημεία περικοπής* καθορίζει το πλήθος των συχνοτήτων μεταξύ του διαστήματος $[0, 0.5]$ που θα υπολογίσει η διαδικασία. Το πεδίο *Πλήθος αυτοσυγχετίσεων* καθορίζει τον αριθμό των αυτοσυγχετίσεων που θα χρησιμοποιηθούν για να προκύψει κάθε μια από τις συχνότητες.

3.2.5 Gamma generator

Με την επιλογή της εντολής εμφανίζεται το πλαίσιο διαλόγου του Σχήματος 9. Το πεδίο *Πλήθος μεταβλητών* καθορίζει το πλήθος των τυχαίων αριθμών που θα παραχθούν. Το πεδίο *Περιοχή αποτελεσμάτων* καθορίζει τη θέση του φύλλου εργασίας κάτω από την οποία θα εμφανιστούν τα αποτελέσματα. Το πεδίο *Αρχική τιμή γεννήτριας* καθορίζει την τιμή με την οποία θα αρχίσει η παραγωγή των τυχαίων αριθμών. Τα υπόλοιπα πεδία καθορίζουν τις μεταβλητές της τριπαραμετρικής κατανομής γάμα. Η παραγωγή τυχαίων μεταβλητών κατανομής γάμα μπορεί να γίνει και με την εισαγωγή της συνάρτησης $\text{grand}(κ, λ, c)$ σε κάποιο πεδίο του φύλλου εργασίας.



Σχήμα 9. Πλαίσιο διαλόγου γεννήτριας τυχαίων αριθμών από την κατανομή γάμα.

4. Αναφορές

Hirsch, R. M., D. L. Helsen, T. A. Cohn, and E. J. Gilroy, Statistical analysis of hydrologic data, in *Handbook of Hydrology*, edited by D. R. Maidment, McGraw-Hill, 1993.

Kay, S. M., *Modern Spectral Estimation, Theory and Application*, Prentice Hall, 1988

Kotegoda, N. T., *Stochastic Water Resources Technology*, Macmillan Press, 1980.