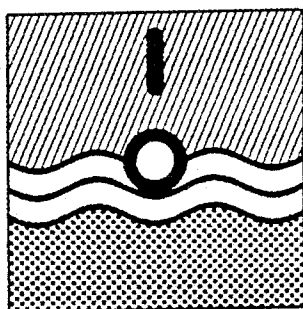


ΥΔΡΟΣΚΟΠΙΟ

ΠΡΟΓΡΑΜΜΑ STRIDE ΕΛΛΑΣ

ΔΗΜΙΟΥΡΓΙΑ ΕΘΝΙΚΗΣ ΤΡΑΠΕΖΑΣ
ΥΔΡΟΛΟΓΙΚΗΣ ΚΑΙ ΜΕΤΕΩΡΟΛΟΓΙΚΗΣ
ΠΛΗΡΟΦΟΡΙΑΣ



HYDROSCOPE

STRIDE HELLAS PROGRAMME

DEVELOPMENT OF A NATIONAL
DATA BANK FOR HYDROLOGICAL
AND METEOROLOGICAL
INFORMATION

ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΟΜΕΑΣ ΥΔΡΑΥΛΙΚΗΣ ΚΑΙ ΤΕΧΝΙΚΗΣ ΠΕΡΙΒΑΛΛΟΝΤΟΣ

ARISTOTLE UNIVERSITY OF THESSALONIKI
FACULTY OF TECHNOLOGY
DIVISION OF HYDRAULICS AND ENVIRONMENTAL
ENGINEERING

ΜΑΜΑΖΗΣ

ΑΝΑΠΤΥΞΗ ΜΕΘΟΔΩΝ ΣΥΜΠΛΗΡΩΣΗΣ
ΕΛΛΕΙΠΟΥΣΩΝ ΤΙΜΩΝ ΧΡΟΝΟΣΕΙΡΩΝ
ΔΕΔΟΜΕΝΩΝ

DEVELOPMENT OF METHODS FOR THE
SUBSTITUTION OF MISSING DATA IN TIME
SERIES

*Μ. Βαφειάδης, Κ. Κατσιφάρκης, Π. Αναστασιάδης,
Ν. Μυλόπουλος*

M. Vafiadis, K. Katsifarakis, P. Anastasiadis, N. Mylopoulos

Αριθμός τεύχους 2/12
Report number

ΘΕΣΣΑΛΟΝΙΚΗ - ΜΑΡΤΙΟΣ 1993
THESSALONIKI - MARCH 1993

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

Περίληψη	ii
1. ΕΙΣΑΓΩΓΗ	1
2. ΤΥΠΟΙ ΔΕΔΟΜΕΝΩΝ	3
3. ΤΥΠΟΙ ΕΛΛΕΙΨΕΩΝ	4
4. ΤΑΞΙΝΟΜΗΣΗ ΤΩΝ ΜΕΘΟΔΩΝ ΑΠΟΚΑΤΑΣΤΑΣΕΩΝ ΕΛΛΕΙΨΕΩΝ ΣΕ ΧΡΟΝΟΣΕΙΡΕΣ	4
5. ΕΠΙΛΟΓΗ ΜΕΘΟΔΟΥ ΑΠΟΚΑΤΑΣΤΑΣΕΩΣ ΕΛΛΕΙΨΕΩΝ ΣΕ ΧΡΟΝΟΣΕΙΡΕΣ	6
6. ΣΥΝΤΟΜΗ ΘΕΩΡΗΣΗ ΤΩΝ ΒΑΣΙΚΩΝ ΘΕΩΡΗΤΙΚΩΝ ΜΕΘΟΔΩΝ	8
7. ΣΥΜΠΕΡΑΣΜΑΤΑ	13
ΟΡΙΣΜΟΙ	14
ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ	15

Περίληψη

Οι χρονοσειρές δεδομένων περιέχουν πέρα από τις γενικές πληροφορίες των συνήθων τυχαίων μεταβλητών και τις επιπλέον πληροφορίες από τη διαδοχή των τιμών στο χρόνο, που προσδιορίζει τη δομή της σειράς στον χρόνο. Αντί για σειρά στο χρόνο σε ορισμένες περιπτώσεις εξετάζουμε και σειρές στο χώρο.

Η απουσία καποιων τιμών σε μία σειρά δημιουργεί προβλήματα στην αξιοποίηση της σειράς, αφού καταστρέφει την διαδοχή. Από την άλλη πλευρά, το γεγονός αυτό της δομής στον χρόνο, επιτρέπει πολλές φορές την εκτίμηση των δεδομένων που λείπουν ή την προσομοίωση των σειρών σε μελλοντικούς χρόνους. Πολλές μέθοδοι έχουν αναπτυχθεί για την συμπλήρωση των ελλειπουσών τιμών και χρησιμοποιούνται κατά περίπτωση ανάλογα με τον τύπο και την κλίμακα των δεδομένων. Οι βασικές από αυτές τις μεθόδους παρουσιάζονται αναλυτικά, αλλά δεν προτείνονται συγκεκριμένες λύσεις για την υπόγεια υδραυλική και υδρολογία, διότι όπως προκύπτει από την εξέταση των υπαρχόντων αρχείων, οι διατιθέμενες σειρές τέτοιων δεδομένων στον ελληνικό χώρο δεν αποτελούν "χρονοσειρές" με τη μαθηματική έννοια του όρου, καθώς τα διαστήματα δειγματολήψιας είναι μεγάλα και γενικά ακανόνιστα, και οι διατιθέμενες τιμές γενικά λίγες.

Abstract

The data timeseries contain besides the general information on the usual random variables the additional information on the succession of data values in time, which determines the structure of the timeseries in time. Sometimes instead of series in time, we deal with series in space.

The lack of some values of a series, causes problems to the exploitation of this series, because it destroys the succession. On the other hand, the time structure, allows us to estimate missing data or to simulate the series for future times. Many methods have been developed for the substitution of missing data in time series, and they are used according to the type and scale of the data. The most fundamental of these methods are presented at length, but no proposition for suitable solutions on groundwater hydraulics and hydrogeology is made, because after the examination of the existing data archives, the available series can not be characterized as "timeseries", because the sampling intervals are large and irregular, and the available values few.

ΑΝΑΠΤΥΞΗ ΜΕΘΟΔΩΝ ΣΥΜΠΛΗΡΩΣΗΣ ΕΛΛΕΙΠΟΥΣΩΝ ΤΙΜΩΝ ΧΡΟΝΟΣΕΙΡΩΝ

ΔΕΔΟΜΕΝΩΝ

(ΥΔΡΟΓΕΩΛΟΓΙΚΑ ΔΕΔΟΜΕΝΑ)

1. ΕΙΣΑΓΩΓΗ

Τα περισσότερα φυσικά μεγέθη (μεταβλητές) που απασχολούν την Υδρολογία και την Υδρογεωλογία μεταβάλλονται μέσα στο χρόνο και μέσα στο χώρο. Δεν αρκούν συνεπώς κάποιες μεμονωμένες τιμές τους που μετρήθηκαν κάποια συγκεκριμένη ή ακόμη χειρότερα κάποια τυχαία στιγμή. Για να έχει κανείς μία πλήρη εικόνα του φυσικού μεγέθους και του σχετικού φαινομένου ή διεργασίας που παριστά η μεταβλητή, και να μπορεί να εκτιμήσει με ακρίβεια τα χαρακτηριστικά της, είναι αναγκαία η μέτρηση της για μεγάλο χρονικό διάστημα ή σε πολλά σημεία του χώρου, και σε καθορισμένα και σταθερά βήματα του χρόνου ή του χώρου.

Με τον τρόπο αυτό σχηματίζεται μία σειρά τιμών που καλείται γενικά **χρονοσειρά**. Στην περίπτωση που αναφερόμαστε σε σειρά δεδομένων στο χώρο μπορούμε να χρησιμοποιήσουμε τον όρο **χωροσειρά** που όμως είναι νεώτερος και δεν είναι τόσο καθιερωμένος όσο ο όρος **χρονοσειρά**.

Το γεγονός ότι οι τιμές μίας μεταβλητής παρουσιάζονται σαν σειρά, σημαίνει ότι πέρα από τα οποιαδήποτε προσδιοριστικά ή στατιστικά χαρακτηριστικά της τυχαιότητας, που εμφανίζει η μεταβλητή, αυτή παρουσιάζει και επιπλέον χαρακτηριστικά που προκύπτουν από τη διάταξη των τιμών σε σειρά (στο χρόνο ή τον χώρο αντίστοιχα). Τα χαρακτηριστικά αυτά προσδιορίζουν τη **χρονική** ή τη **χωρική δομή** της μεταβλητής, και επιτρέπουν εκτιμήσεις για μη μετρημένες ή μελλοντικές τιμές της μεταβλητής. Τα κύρια χαρακτηριστικά μίας χρονοσειράς (της μεταβλητής που εκφράζεται σαν χρονοσειρά) είναι η τάση, η συνάρτηση αυτοσυσχετίσεως και οι περιοδικότητες που εμφανίζει.

Όταν σε μία χρονοσειρά παρουσιάζονται κενά στις μετρήσεις κάποιων βημάτων ή ολόκληρων περιόδων, τότε υπάρχει πρόβλημα στην ανάλυση, μελέτη και τελικά αξιοποίηση της σειράς. Ανάλογα με τις ελλείψεις, τον τύπο της σειράς και τα διατιθέμενα άλλα στοιχεία από μετρήσεις ή ανάλογες σειρές, που σχετίζονται με την προβληματική σειρά, έχουν αναπτυχθεί διάφορες θεωρητικές και εμπειρικές μέθοδοι που επιτρέπουν την "συμπλήρωση" ή "αποκατάσταση" των σειρών. Αν και για τη διαδικασία αυτή υπάρχει σήμερα σημαντική βιβλιογραφία, η έρευνα συνεχίζεται και η επιλογή των πλέον κατάλληλων μεθόδων είναι θέμα που εξετάζεται γενικά κατά περίπτωση (τύπος δεδομένων, ελλείψεις, βοηθητικά στοιχεία).

Η εργασία αυτή όταν προγραμματίστηκε, είχε στόχο την επιλογή και προσδιορισμό των πλέον καταλλήλων μεθόδων για τα δεδομένα της υπόγειας υδραυλικής και υδρογεωλογίας, καθώς και την ανάπτυξη νέων μεθόδων που θα ανταποκρίνονταν καλύτερα στα δεδομένα αυτά και τις ιδιαίτερες ελληνικές συνθήκες. Ο σκοπός ήταν, να γίνει εφικτός ο πρωτογενής έλεγχος και κατ' αρχήν συμπλήρωση των σειρών, πριν από την εισαγωγή τους στη βάση δεδομένων, και κατόπιν η πλήρης αποκατάσταση των σειρών που θα αποτελέσουν τις σειρές αναφοράς της βάσεως δεδομένων. Από την έρευνα όμως τόσο πάνω στον τύπο των μεταβλητών που απασχολούν την υπόγεια υδραυλική και την υδρογεωλογία, όσο και των διατιθέμενων σήμερα δεδομένων στα αρχεία των αρμόδιων για την πραγματοποίηση μετρήσεων και τη συλλογή πληροφοριών υπηρεσιών, διαπιστώθηκε ότι μόνο σε εξαιρετικές περιπτώσεις υπήρχαν πραγματικά σειρές δεδομένων. Στις περισσότερες περιπτώσεις και για τα πλέον κοινά δεδομένα, όπως για παράδειγμα οι στάθμες νερού σε

γεωτρήσεις, υπήρχαν μερικές μόνο μετρήσεις ανά έτος, που δεν μπορούν κατ' αρχήν να θεωρηθούν χρονοσειρά, με την αυστηρή θεωρητική έννοια του όρου, αφού ούτε το χρονικό βήμα μεταξύ των μετρήσεων είναι σταθερό, ούτε και είναι αρκετά πυκνές οι μετρήσεις ώστε να μπορεί κανείς να τις θεωρήσει "γειτονικές", υπό την έννοια ότι ενδιάμεσες μετρήσεις που δεν έγιναν δεν θα προσέθεταν επιπλέον πληροφορίες για τη χρονική (ή χωρική) εξέλιξη της μεταβλητής.

Το αποτέλεσμα ήταν να μην έχει πραγματικά αντικείμενο αυτή η εργασία, αφού δεν υπήρχαν χρονοσειρές. Επιπλέον, η ανάπτυξη οποιωνδήποτε μεθόδων προϋποθέτει και τη δοκιμαστική τουλάχιστον εφαρμογή και έλεγχο αυτών των μεθόδων σε επιλεγμένες σειρές δεδομένων, ώστε ακόμη και εάν δεν είναι δυνατόν να αποδειχθεί η αξιοπιστία τους και να τεκμηριωθεί πλήρως η εφαρμογή τους, τουλάχιστον να υπάρξει κάποια ιδέα της αποτελεσματικότητάς τους. Αυτό όμως μπορεί να γίνει μόνο εάν ήδη υπάρχουν αρκετές σειρές σε μορφή αρχείου πληροφορικής ή καλύτερα σε βάση δεδομένων.

Για τους παραπάνω λόγους η μελέτη αυτή τελικά περιορίζεται στην έκθεση γενικών αρχών και μεθόδων που εφαρμόζονται στην αποκατάσταση χρονοσειρών και σε κάποιες προτάσεις που αφορούν την παραπέρα οργάνωση της συλλογής δεδομένων στην υπόγεια υδραυλική και υδρογεωλογία.

2. ΤΥΠΟΙ ΔΕΔΟΜΕΝΩΝ

Από τα δεδομένα που θα καταχωρηθούν στη βάση δεδομένων μπορούν να θεωρηθούν ότι αποτελούν χρονοσειρές οι ακόλουθες κατηγορίες:

- α. Χρονοσειρές μετρήσεων παροχών (Δοκιμαστικές αντλήσεις)
- β. Στάθμες γεωτρήσεων
- δ. Φυσικές και χημικές παράμετροι ποιότητας νερού

Όλοι οι παραπάνω τύποι δεδομένων είναι ευαίσθητοι σε σφάλματα από την αρχική στιγμή της δημιουργίας τους (καταγραφή, μέτρηση, υπολογισμός, σχεδίαση) και μέχρι την τελική έκδοση τους.

Οι εξηρημένες από τον χρόνο αριθμητικές πληροφορίες αποτελούν τον κατ'έξοχή τύπο πληροφοριών που επιδέχονται ελέγχους και αυστηρές αντικειμενικές αναλύσεις της ποιότητας τους, αλλά συνήθως είναι αυτές που δεν μπορούν να επαναπροσδιοριστούν πρωτογενώς, εάν εξαρτώνται από απόλυτους χρόνους (ημερομηνίες). Σε ορισμένες περιπτώσεις ωστόσο είναι δυνατός ο επαναπροσδιορισμός τους, όπως για παραδειγμα στις δοκιμαστικές αντλήσεις. Όταν μιλούμε για χρονοσειρές, πρέπει να έχουμε υπόψη μας ότι γενικά το σημείο μετρήσεων στο χώρο είναι σταθερό και καλά προσδιορισμένο, και ότι το φυσικό μέγεθος και το βήμα χρόνου που χρησιμοποιούμε σχετίζονται άμεσα. Χρονικό βήμα μεγαλύτερο από τη συχνότητα των διακυμάνσεων του φυσικού μεγέθους, όχι μόνο θα οδηγήσει σε απώλεια πληροφορίας αλλά και θα αλλοιώσει την πραγματική εικόνα της εξελίξεως του φαινομένου. Για τον σκοπό αυτό πάντα στην αρχή των μετρήσεων και για μία περίοδο που θεωρείται δοκιμαστική, οι μετρήσεις γίνονται σε πιο πυκνά διαστήματα και εάν είναι δυνατόν και σε πιο πολλά σημεία. Κατόπιν, και εφόσον αποδειχθεί ότι η απαιτούμενη πληροφορία μπορεί να αποκτηθεί με ικανοποιητική ακρίβεια, αντιπροσωπευτικότητα και αξιοπιστία από μετρήσεις σε πιο αραιά διαστήματα, συστηματοποιούνται οι μετρήσεις με μεγαλύτερο χρονικό βήμα.

Για τον τύπο των μεταβλητών της υπόγειας υδρολογίας και της υδρογεωλογίας και τη μορφή με την οποία απαντούν στα αρχεία των διαφόρων υπηρεσιών, έχει γίνει μία εμπεριστατωμένη μελέτη στα πλαίσια του Υδροσκοπίου από την ομάδα του Υπουργείου Γεωργίας : "Αξιολόγηση πλήθους, μορφής και αξιοπιστίας διαθέσιμων δεδομένων υπόγειας υδρολογίας και υδρογεωλογίας", ΥΠΓΕ/Διεύθυνση Γεωλογίας και Υδρολογίας, Αθήνα 1992, αρ.τεύχους 7/1.

3. ΤΥΠΟΙ ΕΛΛΕΙΨΕΩΝ

3.1. Μία μεμονωμένη τιμή

Είναι η πλέον συνηθισμένη περίπτωση που απαντά σε χρονοσειρές μετρήσεων από παρατηρητή (συνήθως ημερησίων δεδομένων) οπότε η απουσία του παρατηρητού για κάποιο λόγο δεν επιτρέπει την πραγματοποίηση και καταγραφή της μετρήσεως

3.2. Μερικές μεμονωμένες τιμές

Πρόκειται για την προηγούμενη περίπτωση, που επαναλαμβάνεται όμως στο χρόνο και προκύπτει σαν οργανωτικό πρόβλημα της διαδικασίας των μετρήσεων ή αξιοπιστίας του παρατηρητού.

3.3. Σειρά τιμών

Αυτή η περίπτωση παρουσιάζεται στη περίπτωση που οι μετρήσεις γίνονται αυτόματα με καταγραφικά όργανα και παρουσιασθεί βλάβη στα όργανα, μεταξύ δύο διαδοχικών επισκέψεων του επιφορτισμένου με τον έλεγχο τους προσώπου, ή και στην περίπτωση που για ποικίλους λόγους διακοπούν για κάποιο διάστημα οι μετρήσεις από όργανο ή παρατηρητή.

3.4. Συστηματική έλλειψη ολόκληρων περιόδων

Πρόκειται για την περίπτωση που οι μετρήσεις γίνονται με "αποστολές" ή "εκστρατείες", οι οποίες προγραμματίζονται μόνο σε ορισμένες περιόδους του έτους και όχι σε μόνιμη βάση. Ακόμη αυτή η έλλειψη παρατηρείται είτε σε θερινές περιόδους άδειας του προσωπικού, είτε σε χειμερινές περιόδους, οπότε η πρόσβαση στα σημεία των μετρήσεων είναι δυσχερής.

3.5. Ελλείψεις ετών

Είναι η ίδια πρακτικά περίπτωση όπως και στην προηγούμενη κατηγορία. Επιπλέον μπορεί να οφείλεται σε απώλειες στα αρχεία.

3.6. Πολλές σειρές τιμών μικρής διάρκειας που αναφέρονται στην ίδια μεταβλητή

Είναι μία συνηθισμένη περίπτωση όταν ο αρχικός στόχος των μετρήσεων δεν ήταν μία συνεχής καταγραφή του φαινομένου, αλλά η κατά περίπτωση συγκέντρωση στοιχείων για ανάγκες μικρής κλίμακας (Π.χ. προγραμματισμός μικρής κλίμακας έργων ή επεμβάσεων σε έργα που ήδη λειτουργούν).

4. ΤΑΞΙΝΟΜΗΣΗ ΤΩΝ ΜΕΘΟΔΩΝ ΑΠΟΚΑΤΑΣΤΑΣΕΩΝ ΕΛΛΕΙΨΕΩΝ ΣΕ ΧΡΟΝΟΣΕΙΡΕΣ

Οι μέθοδοι που απαντούν στη διεθνή βιβλιογραφία και αφορούν την αποκατάσταση χρονοσειρών μπορούν να ταξινομηθούν ως προς τον τρόπο λειτουργίας τους στις ακόλουθες κατηγορίες:

1. Καθαρά εμπειρικές
2. Μέθοδοι παρεμβολής
3. Μέθοδοι παλινδρομήσεως
4. Χωρικές (απλές ή γεωστατιστικές)
5. Δημιουργία μαθηματικού μοντέλου της χρονοσειράς, μετά από ανάλυση χρονοσειράς.

5. ΕΠΙΛΟΓΗ ΜΕΘΟΔΟΥ ΑΠΟΚΑΤΑΣΤΑΣΕΩΣ ΕΛΛΕΙΨΕΩΝ ΣΕ ΧΡΟΝΟΣΕΙΡΕΣ

5.1. Για μεμονωμένες τιμές

1. Μέσος όρος αντίστοιχων τιμών άλλων ετών.
2. Μέσος όρος αντίστοιχων τιμών άλλων ετών, διορθωμένος στη βάση των μέσων ετησίων τιμών έτους ή περιόδου.
3. Μέσος όρος τιμών γειτονικών σημείων μετρήσεως.
4. Μέσος όρος τιμών γειτονικών σημείων μετρήσεως, διορθωμένος στη βάση των μέσων τιμών των δειγμάτων.
5. Μέσος όρος τιμών γειτονικών σημείων μετρήσεως, διορθωμένος στη βάση των μέσων τιμών των δειγμάτων και της τυπικής αποκλίσεως των δειγμάτων.
6. Μέσος όρος τιμών γειτονικών σημείων μετρήσεως, διορθωμένος στη βάση των μέσων ετησίων τιμών έτους ή περιόδου.
7. Μέσος όρος τιμών γειτονικών σημείων μετρήσεως, διορθωμένος στη βάση των μέσων ετησίων τιμών έτους ή περιόδου και της τυπικής αποκλίσεως των τιμών έτους ή περιόδου.

5.2. Για μεμονωμένες τιμές σε σειρές τιμών

1. Μέσος όρος αντίστοιχων τιμών άλλων ετών.
2. Μέσος όρος αντίστοιχων τιμών άλλων ετών, διορθωμένος στη βάση των μέσων ετησίων τιμών έτους ή περιόδου.
3. Μέσος όρος τιμών γειτονικών σημείων μετρήσεως.
4. Μέσος όρος τιμών γειτονικών σημείων μετρήσεως, διορθωμένος στη βάση των μέσων τιμών των δειγμάτων.
5. Μέσος όρος τιμών γειτονικών σημείων μετρήσεως, διορθωμένος στη βάση των μέσων τιμών των δειγμάτων και της τυπικής αποκλίσεως των δειγμάτων.
6. Μέσος όρος τιμών γειτονικών σημείων μετρήσεως, διορθωμένος στη βάση των μέσων ετησίων τιμών έτους ή περιόδου.
7. Μέσος όρος τιμών γειτονικών σημείων μετρήσεως, διορθωμένος στη βάση των μέσων ετησίων τιμών έτους ή περιόδου και της τυπικής αποκλίσεως των τιμών έτους ή περιόδου.
8. Γραμμική παρεμβολή.
9. Πολυωνυμική ή μη γραμμική παρεμβολή.
10. Ανάλυση συσχέτισεως των σειρών και συμπλήρωση με τη μέθοδο της παλινδρομήςσεως.
11. Ανάλυση συσχέτισεως των σειρών και συμπλήρωση με τη μέθοδο της παλινδρομήςσεως, με εφαρμογή διορθωτικών συντελεστών με βάση τα μήκη των σειρών.

Για να είναι επιτυχημένη η συμπλήρωση μιάς χρονοσειράς θα πρέπει κάθε τιμή που υπολογίζουμε, να προσεγγίζει όσο το δυνατόν περισσότερο την άγνωστη τιμή που πρόκειται να υποκαταστήσει και να ικανοποιεί τα στατιστικά χαρακτηριστικά της καθώς και τα στοχαστικά χαρακτηριστικά της σειράς στην οποία ανήκει.

Αρα πριν από κάθε προσπάθεια για συμπλήρωση μιάς χρονοσειράς πρέπει να προηγηθούν τα εξής στάδια.

1. Αναζήτηση στα αρχεία αντιγράφων της σειράς χωρίς ελλείψεις
2. Αναζήτηση στα αρχεία αντιγράφων άλλης σειράς χωρίς ελλείψεις που θα μπορούσε να υποκαταστήσει αυτήν που διαθέτουμε και είναι ελλιπής.
3. Αναζήτηση στα αρχεία άλλων σειρών χωρίς ελλείψεις που δεν μπορούν κατ'άρχην να υποκαταστήσουν αυτήν που διαθέτουμε και είναι ελλιπής, αλλά μπορούν να βοηθήσουν μόνες τους ή σε συνδυασμό με άλλες, στην αποκατάσταση της σειράς που μας ενδιαφέρει.
4. Στατιστική ανάλυση, ανάλυση χρονοσειράς και χωρική ανάλυση του πεδίου για τις σειρές που έχουμε στη διάθεση μας, ώστε να εξάγουμε όσο το δυνατόν περισσότερες πληροφορίες που θα μας επιτρέψουν να υπολογίσουμε για τις τιμές που λείπουν όσο το δυνατόν πλησιέστερους αντικαταστάτες.

5.3. Για μεμονωμένες τιμές ή σειρά τιμών σε χρονοσειρές τιμών

1. Μέσος όρος αντίστοιχων τιμών άλλων ετών.
2. Μέσος όρος αντίστοιχων τιμών άλλων ετών, διορθωμένος στη βάση των μέσων ετησίων τιμών έτους ή περιόδου.
3. Μέσος όρος τιμών γειτονικών σημείων μετρήσεως.
4. Μέσος όρος τιμών γειτονικών σημείων μετρήσεως, διορθωμένος στη βάση των μέσων τιμών των δειγμάτων.
5. Μέσος όρος τιμών γειτονικών σημείων μετρήσεως, διορθωμένος στη βάση των μέσων τιμών των δειγμάτων και της τυπικής αποκλίσεως των δειγμάτων.
6. Μέσος όρος τιμών γειτονικών σημείων μετρήσεως, διορθωμένος στη βάση των μέσων ετησίων τιμών έτους ή περιόδου.
7. Μέσος όρος τιμών γειτονικών σημείων μετρήσεως, διορθωμένος στη βάση των μέσων ετησίων τιμών έτους ή περιόδου και της τυπικής αποκλίσεως των τιμών έτους ή περιόδου.
8. Γραμμική παρεμβολή.
9. Πολυωνυμική ή μη γραμμική παρεμβολή.
10. Ανάλυση συσχέτισεως των σειρών και συμπλήρωση με τη μέθοδο της παλινδρομήσεως.
11. Ανάλυση συσχέτισεως των σειρών και συμπλήρωση με τη μέθοδο της παλινδρομήσεως, με εφαρμογή διορθωτικών συντελεστών με βάση τα μήκη των σειρών.
12. Ανάλυση περιοδικότητας (φασματική ανάλυση). Εάν υπάρχουν, συμπλήρωση των σειρών με εφαρμογή μοντέλου χρονοσειράς.

5.4. Για μεμονωμένες τιμές σε πεδίο, για το οποίο είναι γνωστή η χωρική του δομή

1. Εφαρμογή γεωστατιστικών μεθόδων.

6. ΣΥΝΤΟΜΗ ΘΕΩΡΗΣΗ ΤΩΝ ΒΑΣΙΚΩΝ ΘΕΩΡΗΤΙΚΩΝ ΜΕΘΟΔΩΝ

Οι εμπειρικές μέθοδοι αφορούν απλούς μέσους όρους αντιστοιχών τιμών σειρών από "γειτονικά" σημεία μετρήσεων, με ή χωρίς σταθμικούς συντελεστές, χωρίς ιδιαίτερη ανάλυση και εξασφάλιση των στατιστικών χαρακτηριστικών των παραγόμενων δεδομένων

Σε σειρές όπου το βήμα μετρήσεως είναι πολύ πυκνό σε σχέση με τη μεταβλητότητα του φυσικού μεγέθους στον χρόνο, κάποιες τιμές που τυχόν λείπουν μπορούν να υπολογισθούν από τις προηγούμενες και τις επόμενες στη σειρά, με εφαρμογή απλών μεθόδων παρεμβολής.

Στην περίπτωση που διατίθενται πλήρεις και αξιόπιστες σειρές με καλή συσχέτιση με την προβληματική σειρά, μπορούν να εφαρμοσθούν διάφορες μέθοδοι που αξιοποιούν τα στατιστικά χαρακτηριστικά και τη συσχέτιση των σειρών αυτών μεταξύ των. Η πλέον προφανής προσέγγιση είναι να χρησιμοποιήσει κανείς μία εξίσωση παλινδρομήσεως με εξηρημένη μεταβλητή την υπο εξέταση σειρά και ανεξάρτητες μία ή περισσότερες σειρές μετρήσεων σε "γειτονικά" σημεία, δηλαδή σειρές για τις οποίες έχει βρεθεί ότι έχουν μεγάλο βαθμό συσχέτισεως προς την προς συμπλήρωση σειρά.

$$y_i = \frac{1}{n} \left(\frac{m_y}{m_{x_1}} x_1 + \frac{m_y}{m_{x_2}} x_2 + \dots + \frac{m_y}{m_{x_n}} x_n \right)$$

Η παραπάνω σχέση εξασφαλίζει ότι μέση τιμή των εκτιμώμενων τιμών θα προσεγγίζει τη μέση τιμή της μεταβλητής, δεν εξασφαλίζει όμως καθόλου την τυπική της απόκλιση.

Ενα άλλο μοντέλο παλινδρομήσεως μεταξύ δύο μόνο σειρών, όπου συνεκτιμάται τόσο η μέση τιμή τους όσο και η τυπική τους απόκλιση είναι το ακόλουθο:

$$\hat{y} = \alpha + \beta x = m_y + r \frac{s_y}{s_x} (x - m_x)$$

Και στις δύο περιπτώσεις οι στατιστικές παράμετροι που χρησιμοποιούνται σαν σταθμικοί παράγοντες έχουν προηγούμενα εκτιμηθεί από την κοινή, πλήρη περίοδο καταγραφής.

Για τη μέση τιμή, την τυπική απόκλιση και τα σφάλματα στην εκτίμηση τους προκύπτει ότι:

$$\begin{aligned} \mu_{\hat{y}} &= \mu_y & \sigma_{\hat{y}} &= \hat{\rho} \sigma_y \\ \mu_e &= 0 & \sigma_e &= \sigma_y \sqrt{1 - \hat{\rho}^2} \end{aligned}$$

Επειδή η εκτίμηση της y είναι αβίαστη ως προς τη μέση τιμή της, αλλά όχι και προς την μεταβλητότητα της, μπορούμε να συμπεριλάβουμε τον ακόλουθο διορθωτικό όρο στην εξίσωση

$$z \sigma_{\hat{y}} \sqrt{1 - \hat{\rho}^2}$$

που γίνεται

$$\hat{y} = \mu_y + \hat{\rho} \frac{\sigma_{\hat{y}}}{\sigma_{\hat{x}}} (x - \mu_x) + z \sigma_{\hat{y}} \sqrt{1 - \hat{\rho}^2}$$

με z μία κανονική τυχαία μεταβλητή με μέση τιμή 0 και τυπική απόκλιση 1. $[N(0,1)]$

Η μέση τιμή και η τυπική απόκλιση είναι τώρα αβίαστες, αλλά το σφάλμα στην εκτίμηση της τυπικής αποκλίσεως διπλασιάστηκε:

$$\begin{aligned} \mu_{\hat{y}} &= \mu_y & \sigma_{\hat{y}} &= \sigma_y \\ \mu_e &= 0 & \sigma_e &= \sigma_y \sqrt{2(1 - \hat{\rho}^2)} \end{aligned}$$

Μειονέκτημα όμως για το μοντέλο αυτό είναι και η επιλογή του z , που φαίνεται αυθαίρετη.

Μία τετάρτη μέθοδος για την εκτίμηση ελλειπουσών τιμών, είναι αυτή που χρησιμοποιεί ίσα ποσοστιαία σημεία (ανηγμένες μεταβλητές):

$$\frac{\hat{y} - \mu_{\hat{y}}}{\sigma_y} = \frac{x - \mu_x}{\sigma_x}$$

ή

$$\hat{y} = m_y + \frac{\sigma_{\hat{y}}}{\sigma_x} (x - \mu_x)$$

Οι παράμετροι κατανομής της y είναι:

$$\begin{aligned} \mu_{\hat{y}} &= \mu_y & \sigma_{\hat{y}} &= \sigma_y \\ \mu_e &= 0 & \sigma_e &= \sigma_y \sqrt{2(1 - \hat{\rho})} \end{aligned}$$

Μπορούμε να παρατηρήσουμε ότι:

$$\sigma_y \sqrt{2(1 - \hat{\rho})} \leq \sigma_y \sqrt{2(1 - \hat{\rho}^2)}$$

Παρατηρούμε ότι οι διάφορες λύσεις δεν καλύπτουν το ίδιο αποτελεσματικά τα διάφορα σφάλματα και συστηματικές εκτροπές της εκτιμήσεως. Ανάλογα με το είδος των δεδομένων και τη σημασία τους, θα πρέπει να προαποφασισθεί ποιο από τα στατιστικά χαρακτηριστικά είναι σπουδαιότερο γαι τη συγκεκριμένη αποκατάσταση της σειράς.

Σε κάθε περίπτωση η αποκατάσταση μίας σειράς y με βάση μία σειρά x , θα έχει νόημα, μόνον αν το ισοδύναμο μήκος της σειράς μετά από τη συμπλήρωση των ελλειπουσών τιμών ικανοποιεί τη σχέση:

$$n^* = \frac{n_x}{1 + \frac{n_x - n_y}{n_y - 2} (1 - \hat{\rho}^2)} > n_y$$

όπου n^* το ισοδύναμο μήκος της σειράς y μετά την αποκατάσταση.

Από την ανάλυση ορισμένων σειρών δεδομένων προκύπτει ότι, εκτός από τον χαρακτήρα της τυχαίας μεταβλητής, αυτές παρουσιάζουν και στοχαστικό χαρακτήρα και ορισμένα προσδιοριστικά χαρακτηριστικά, όπως τάση και περιοδικότητα. Η προσομοίωση μίας τέτοιας σειράς μπορεί να βασισθεί σ'αυτά ακριβώς τα προσδιοριστικά χαρακτηριστικά της, προκειμένου να επιτύχουμε μία καλή εκτίμηση της χρονικής της πορείας, αλλά και για να συμπληρώσουμε κάποιες τιμές που λείπουν ή να επεκτείνουμε τη σειρά.

Για την εκτίμηση των περιοδικοτήτων, η μέθοδος MESA (Maximum Entropy Spectral Analysis) έχει αποδειχθεί ότι είναι ένα πολύ ισχυρό και αξιόπιστο εργαλείο για τον εντοπισμό των κρυμμένων περιοδικοτήτων σε μία χρονοσειρά. Η εκτίμηση όμως του εύρους της κάθε περιοδικότητας και η σχετική συνεισφορά της στη διαμόρφωση της σειράς δεν μπορούν να προκύψουν από τη MESA κατά τρόπο ακριβή και αξιόπιστο, λόγω της φύσεως της μεθόδου και των φίλτρων που επιβάλλει στα δεδομένα. Μπορεί να εφαρμοσθεί όμως σε συνδυασμό με κάποια άλλη μέθοδο, όπως πχ. παλινδρόμηση και μέθοδο ελαχίστων τετραγώνων με την οποία θα προσδιοριστούν τα αντίστοιχα πλάτη α_n και β_n των αρμονικών συνιστωσών.

Αν αποδεχθούμε την παράσταση μίας σειράς ως περιοδικής, αυτό σημαίνει ότι μαθηματικά θα έχουμε μία παράσταση της μορφής:

$$\begin{aligned} F(t) &= \sum_n \left(\alpha_n \sin 2\pi \frac{t}{T_n} + \beta_n \cos 2\pi \frac{t}{T_n} \right) \\ &= \sum_n r_n \sin \left(2\pi \frac{t}{T_n} + \varphi_n \right) \end{aligned}$$

$$\begin{aligned} r_n^2 &= \alpha_n^2 + \beta_n^2 \\ \varphi_n &= \tan^{-1}(\alpha_n / \beta_n) \end{aligned}$$

με

$$\begin{aligned} \sigma_m &= \frac{1}{n} [\alpha_n^2 \sigma_{\alpha n}^2 + \beta_n^2 \sigma_{\beta n}^2]^{\frac{1}{2}} \\ &\cong \sigma \end{aligned}$$

εάν

$$\sigma_{\alpha n} \equiv \sigma_{\beta n} \equiv \sigma$$

όπου $F(t)$ = το ύψος βροχής που παρατηρήθηκε κατά το έτος t (σαν έκφραση των αποκλίσεων από τη μέση τιμή, T_n οι διάφορες περιодικότητες που εκτιμήθηκαν με κάποια μέθοδο φασματικής ανάλυσεως.

ΣΥΜΒΟΛΙΣΜΟΙ:

μ_x, μ_y	:	πραγματική μέση τιμή της τυχαίας μεταβλητής (μέση τιμή πληθυσμού)
σ_x, σ_y	:	πραγματική τυπική απόκλιση της τυχαίας μεταβλητής (τυπ. απόκλ. πληθυσμού)
m_x, m_y	:	μέση τιμή του δείγματος της τυχαίας μεταβλητής
s_x, s_y	:	τυπική απόκλιση του δείγματος της τυχαίας μεταβλητής
n_x, n_y	:	μήκος της σειράς της τυχαίας μεταβλητής
ρ_{xy}	:	πραγματικός συντελεστής συσχέτισης μεταξύ των τυχαίων μεταβλητών x και y (συντελεστής συσχέτισης μεταξύ των πληθυσμών των τυχαίων μεταβλητών x και y)
r_{xy}	:	συντελεστής συσχέτισης μεταξύ των δειγμάτων των τυχαίων μεταβλητών x και y (συντελεστής συσχέτισης μεταξύ των δειγμάτων των τυχαίων μεταβλητών x και y)
$\wedge$	:	σύμβολο που υποδηλώνει εκτίμηση
$\varepsilon = y - \hat{y}$	:	υπόλοιπο (σφάλμα) εκτιμήσεως

7. ΣΥΜΠΕΡΑΣΜΑΤΑ

Η συμπλήρωση των ελλειπουσών τιμών σε σειρές δεδομένων είναι μία λεπτή διαδικασία που εξαρτάται από το είδος των δεδομένων, των ελλείψεων και των διατιθέμενων αξιόπιστων και πλήρων σειρών δεδομένων που μπορούν να αποτελέσουν τη βάση αναφοράς και υπολογισμού για την αποκατάσταση των αντίστοιχων σειρών που έχουν ελλείψεις.

Για να γίνει δυνατή οποιαδήποτε τυποποίηση και αυτοματοποίηση της, θα πρέπει προηγουμένως να γίνουν τα ακόλουθα:

1. Πλήρης καταγραφή και ταξινόμηση των σειρών δεδομένων που χρήζουν αποκαταστάσεως.
2. Οι σειρές δεδομένων που θα επιλεγούν για αποκατάσταση και οι σειρές αναφοράς θα πρέπει να βρίσκονται σε υπόβαθρο πληροφορικής, για να είναι δυνατός ο χειρισμός τους, και ο πειραματισμός και σύγκριση με τις διάφορες μεθόδους για τη "συμπλήρωση" ή "αποκατάσταση" των σειρών.
3. Ειδικά στην περίπτωση πολλών καταγραφών της υπόγειας υδραυλικής και υδρογεωλογίας, ενώ αναφέρονται σαν "χρονοσειρές" δεν αποτελούν στην πραγματικότητα σειρές, αλλά "συλλογές" δεδομένων, διότι το χρονικό βήμα και η πυκνότητα των μετρήσεων είναι μεταβαλλόμενη. Σ'αυτήν την περίπτωση, που αποτελεί τον κανόνα, δεν είναι δυνατή οποιαδήποτε επέμβαση με ποσοτικό χαρακτήρα, στατιστικά ισχυρή, αλλά μόνο ενδεικτικές, ποιοτικές αποκαταστάσεις τιμών.

ΟΡΙΣΜΟΙ

- Δεδομένα : κάθε είδους πληροφορία που είναι προϊόν μετρήσεων (στη φύση, το εργαστήριο ή το χάρτη)
- Απόκλιση : η εκτροπή από την κανονική τιμή
- Επίπεδο : το επίπεδο, σε ποσοστό κάποιας βασικής τιμής, που σημαντικότητας καθορίζεται σαν όριο για την αποδοχή ή όχι της μιάς μεταβλητής
- Περιοδικότητα : η ρυθμική μεταβολή ενός μεγέθους (στο χρόνο ή το χώρο)
- Τυχαιότητα : η κατάσταση όπου δεν είναι δυνατόν να εκτιμηθεί με ακρίβεια η τιμή ενός μεγέθους που μεταβάλλεται.
- Συσχέτιση : η εκτίμηση του πόση σχέση υπάρχει μεταξύ δύο μεταβλητών με βάση τα ζεύγη των τιμών τους (γραμμική, μη γραμμική)
- Παλινδρόμηση : η εκτίμηση της τιμής μιάς μεταβλητής με τη βοήθεια της τιμής μιάς (απλή παλινδρόμηση) ή πολλών άλλων (πολλαπλή) μεταβλητών
- Χρονοσειρά : μία σειρά δεδομένων που μετρήθηκαν σε ίσα διαδοχικά χρονικά διαστήματα και στο αυτό σημείο του χώρου.
- Συμπλήρωση σειρών : η αποκατάσταση των ελλείψεων και των κενών στις σειρές δεδομένων
- Χρονικό βήμα : η χρονική απόσταση μεταξύ δύο διαδοχικών σημείων μιάς χρονοσειράς
- Χωροσειρά : μία σειρά δεδομένων που αντιστοιχεί σε κανονικά διατεταγμένα σημεία στο χώρο

ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ

Commission for Hydrology, WMO, 1975, Hydrological Forecasting Practices

Εκθεση ΥΔΡΟΣΚΟΠΙΟΥ αρ.τεύχους 7/1, ΥΠΓΕ/Διεύθυνση Γεωλογίας και Υδρολογίας, Αθήνα 1992, Αξιολόγηση πλήθους, μορφής και αξιοπιστίας διαθέσιμων δεδομένων υπόγειας υδρολογίας και υδρογεωλογίας

Cavadias G.S., 1989, Introduction to exploratory data analysis

Llamas J., G. Morin, 1985, Hydrologie generale: principes et applications

Sharma R.K., Dhanpat Rai & Sons, 1983, Hydrology & Water Resources

Shaw E.M., Van Nostrand Reinhold (UK), 1983, Hydrology in Practice

Ven Te Chow, D.R. Maidment, L.W. Mays, McGraw-Hill, 1988, Applied hydrology

Viessman W. et al., Harper & Row, 1977, Introduction to Hydrology

Walton, McGraw-Hill, 1970, Groundwater Resource Evaluation