

Προσομοίωση βροχόπτωσης – απορροής με τη χρήση της παλινδρόμησης των μηχανών διανυσμάτων υποστήριξης

Δημήτριος Μπότσης¹, Περικλής Λατινόπουλος², Κωνσταντίνος Διαμαντάρας³
^{1&2} Τμήμα Πολιτικών Μηχανικών, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης,
TK-54124 Θεσσαλονίκη, ³ Τμήμα Πληροφορικής, Αλεξάνδρειο Τεχνολογικό
Εκπαιδευτικό Ίδρυμα Θεσσαλονίκης, TK-57400 Σίνδος
e-mail: jimbotsis@civil.auth.gr, latin@civil.autn.gr, kdiamant@it.teithe.gr

Περίληψη

Στην παρούσα ερευνητική εργασία ένα ομοίωμα παλινδρόμησης των μηχανών διανυσμάτων υποστήριξης (ΜΔΥ) σχεδιάστηκε για τη διερεύνηση της σχέσης βροχόπτωσης-απορροής που αποτελεί ένα από τα περισσότερο πολύπλοκα υδρολογικά φαινόμενα. Η σχέση αυτή είναι μη-γραμμική και ενδείκνυται η χρήση μη-γραμμικών ομοιωμάτων όπως είναι η παλινδρόμηση των ΜΔΥ, τα οποία μπορεί να παρουσιάσουν σημαντικά πλεονεκτήματα στη διερεύνηση της. Είναι σημαντικό ότι πολλές έρευνες στη βιβλιογραφία επισημαίνουν ότι η παλινδρόμηση των ΜΔΥ είναι ένα πανίσχυρο εργαλείο στην προσομοίωση της μη-γραμμικότητας. Στην συγκεκριμένη εφαρμογή χρησιμοποιήθηκαν η ημερήσια βροχόπτωση, η βροχόπτωση των δύο προηγούμενων ημερών και η μετρημένη παροχή μίας ορεινής δασωμένης λεκάνης απορροής, ως δεδομένα για την προσομοίωση του ομοιώματος. Η λεκάνη απορροής βρίσκεται στη βόρεια Καλιφόρνια σε μία περιοχή που έχει το ίδιο γεωγραφικό πλάτος με τη βόρεια Ελλάδα. Τα δεδομένα βροχόπτωσης και απορροής διαιρέθηκαν σε τρία υποσύνολα, ένα για εκπαίδευση, ένα για επικύρωση και ένα για έλεγχο του ομοιώματος. Η απόδοση του ομοιώματος παλινδρόμησης των ΜΔΥ εξαρτάται από την επιλογή των συναρτήσεων πυρήνα. Στην παρούσα εφαρμογή χρησιμοποιήθηκαν η γραμμική, η πολυωνυμική και η Γκαουσιανή συνάρτηση πυρήνα. Η αξιοπιστία και η απόδοση του ομοιώματος ελέγχθηκε με τη χρησιμοποίηση τριών διαφορετικών κριτηρίων αξιολόγησης (συντελεστής συσχέτισης, μέσο τετραγωνικό σφάλμα και τετραγωνική ρίζα μέσου τετραγωνικού σφάλματος). Το ομοίωμα αποδείχθηκε ότι είναι αρκετά ικανό στην προσομοίωση της σχέσης βροχόπτωσης-απορροής επιβεβαιώνοντας τα αποτελέσματα αντίστοιχων άλλων ερευνητικών εργασιών.

Λέξεις κλειδιά: μηχανές διανυσμάτων υποστήριξης, παλινδρόμηση, βροχόπτωση, απορροή, λεκάνη απορροής

Εισαγωγή

Η σχέση μεταξύ βροχόπτωσης και απορροής είναι ένα από τα περισσότερο πολύπλοκα υδρολογικά φαινόμενα και η πρόβλεψη της απορροής εξαρτάται από ένα μεγάλο αριθμό παραμέτρων. Η δημιουργία της μέγιστης απορροής εξαρτάται πρωταρχικά από τη βροχόπτωση και ιδιαίτερα από τη ραγδαιότητα της. Ωστόσο η τελική απορροή που προκύπτει σε μία λεκάνη, εκτός από τη βροχόπτωση, είναι συνάρτηση των γεωμετρικών χαρακτηριστικών της, της δασοκάλυψης, των χρήσεων γης, της εδαφικής υγρασίας και του γεωλογικού υποβάθρου (Λατινόπουλος 2007). Το υδρολογικό υδατικό ισοζύγιο είναι ακόμη περισσότερο πολύπλοκο στις ορεινές δασωμένες λεκάνες απορροής.

Η κομοστέγη του δάσους στέκεται στον αέρα αποτελώντας φυσικό εμπόδιο στο νερό της βροχής και στη χιονόπτωση, ώστε να φτάσουν στο έδαφος. Μια ποσότητα

του νερού συγκρατείται από την κόμη των δέντρων και ονομάζεται υδατοσυγκράτηση, ενώ το νερό που διαπερνά την κόμη χαρακτηρίζεται ως διαπερώσα βροχή. Ταυτόχρονα νερό της βροχής ρέει κατά μήκος των κορμών των δέντρων αποτελώντας την κορμοαπορροή (Chang 2003, Κωτούλας 1995). Από το νερό που τελικά φτάνει στο έδαφος μία ποσότητα διηθείται στα βαθύτερα εδαφικά στρώματα, ενώ το υπόλοιπο αποτελεί την απορροή ή διηθείται σε επιφανειακά εδαφικά στρώματα αποτελώντας την υπεδάφεια (υποδερμική) απορροή η οποία τελικά συμμετέχει στην επιφανειακή απορροή (Κουτσογιάννης και Ξανθόπουλος 1999).

Η σχέση μεταξύ βροχόπτωσης και απορροής είναι ένα φυσικό φαινόμενο που παίζει πολύ σημαντικό ρόλο στη διαχείριση των υδατικών πόρων και στην πρόβλεψη των πλημμυρικών παροχών, για το λόγο αυτό έχει αναπτυχθεί ένας μεγάλος αριθμός ομοιωμάτων για την προσομοίωση της, με διαφορετικό βαθμό πολυπλοκότητας (Solaimani 2009). Τα ομοιώματα αυτά κατηγοριοποιούνται ως εμπειρικά, μαύρα-κουτιά (black-box) και στοχαστικά (Rajurkar et al 2002).

Στην παρούσα ερευνητική εργασία έγινε μία εφαρμογή της παλινδρόμησης των μηχανών διανυσμάτων υποστήριξης (ΜΔΥ) στην προσομοίωση της σχέσης βροχόπτωσης-απορροής σε μία ορεινή λεκάνη με σημαντική δασοκάλυψη. Η εν λόγω εφαρμογή καταδεικνύει κάποια από τα σημαντικότερα χαρακτηριστικά και προτερήματα των μη γραμμικών μοντέλων στην πρόβλεψη της απορροής.

Ο αλγόριθμος των μηχανών διανυσμάτων είναι μία μη-γραμμική γενίκευση του αλγόριθμου που αναπτύχθηκε από τον Varnik στις αρχές του 1960 (Varnik and Lerner 1963, Varnik and Chervonenkis 1964). Η ανάπτυξη των μηχανών διανυσμάτων υποστήριξης στηρίχθηκε στη δομή της αρχής της ελαχιστοποίησης του κινδύνου που προέρχεται από τη στατιστική. Η μέθοδος αυτή έγινε διάσημη και χρησιμοποιήθηκε σε μεγάλο βαθμό στην προσομοίωση μεγάλης και μικρής κλίμακας φυσικών φαινομένων και μη-γραμμικών προβλημάτων.

Τα τελευταία χρόνια έχουν πραγματοποιηθεί αρκετές εφαρμογές της παλινδρόμησης των ΜΔΥ στον τομέα της υδρολογίας και συγκεκριμένα στη διερεύνηση της σχέσης βροχόπτωσης-απορροής και πρόβλεψής της απορροής. Για παράδειγμα οι Gray και Han το 2004 εφάρμοσαν τις ΜΔΥ στην πρόβλεψη της απορροής εστιάζοντας στην επιλογή της κατάλληλης δομής του ομοιώματος και των παραμέτρων που χρησιμοποίησαν. Οι Sivapragasam και Liong το 2004 χρησιμοποίησαν την παλινδρόμηση των ΜΔΥ για την πρόβλεψη της απορροής. Οι Sivapragasam και Liong το 2005 διαίρεσαν το πεδίο απορροής σε τρεις υποπεριοχές και εφάρμοσαν διαφορετικά ομοιώματα παλινδρόμησης των ΜΔΥ για να προσδιορίσουν ημερήσιες απορροές σε περιοχές με μεγάλο, μέσο και χαμηλό υψόμετρο. Οι Lin et al το 2006 παρουσίασαν ένα ομοίωμα παλινδρόμησης ΜΔΥ για την πρόβλεψη μακροπρόθεσμης μηνιαίας απορροής και προχώρησαν σε μία σύγκριση των αποτελεσμάτων με στοχαστικά ομοιώματα και με τεχνητά νευρωνικά δίκτυα, διαπιστώνοντας ότι το ομοίωμα της παλινδρόμησης των ΜΔΥ έδωσε καλύτερα αποτελέσματα.

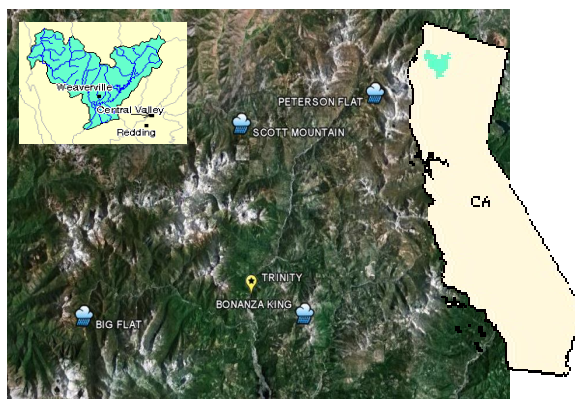
Υλικά και μέθοδοι

Περιοχή έρευνας

Το πεδίο εφαρμογής της παρούσας ερευνητικής εργασίας είναι μία ορεινή λεκάνη απορροής στη βόρεια Καλιφόρνια σε μία περιοχή που έχει το ίδιο γεωγραφικό πλάτος με τη βόρεια Ελλάδα. Η περιοχή έρευνας παρουσιάζει παρόμοια χαρακτηριστικά με αυτά της βόρειας Ελλάδας καθώς χαρακτηρίζεται από

ανάλογες κλιματολογικές συνθήκες και βλάστηση. Η έκταση της λεκάνης ανέρχεται σε 385.9 km² και η μεγαλύτερη επιφάνεια της καλύπτεται από δάσος πεύκης.

Βασική προϋπόθεση για τον έλεγχο της αξιοπιστίας ενός ομοιώματος είναι η ύπαρξη ιστορικών στοιχείων. Έτσι ο λόγος για τον οποίο επιλέχθηκε μία λεκάνη απορροής στη συγκεκριμένη γεωγραφική περιοχή είναι διότι υπάρχουν διαθέσιμα δεδομένα βροχόπτωσης, χιονόπτωσης και το κυριότερο μετρημένες παροχές στην έξοδο της λεκάνης, ενώ όπως αναφέρθηκε νωρίτερα προσομοιώνει αρκετά τις συνθήκες των λεκανών απορροής στη βόρεια Ελλάδα. Στην εικόνα 1 παρουσιάζεται η θέση της λεκάνης και των σταθμών μέτρησης μετεωρολογικών και υδρολογικών δεδομένων.



Εικόνα 1: Περιοχή έρευνας
Figure 1: Study area

Παλινδρόμηση μηχανών διανυσμάτων υποστήριξης (ΜΔΥ)

Στην υδρολογία τα περισσότερα φυσικά φαινόμενα είναι πολύπλοκα και τείνουν προς τη μη-γραμμικότητα. Στην παρούσα ερευνητική εργασία η βασική αρχή της παλινδρόμησης των ΜΔΥ είναι να προσδιορίσουν τη μη-γραμμικότητα στη σχέση μεταξύ της βροχόπτωσης και απορροής. Η παλινδρόμηση των ΜΔΥ μπορεί να εκτιμήσει τη μη-γραμμική συνάρτηση μεταξύ ενός διανύσματος εισόδου x και μίας μεταβλητής εξόδου. Χρησιμοποιώντας τη μη-γραμμική συνάρτηση $x \rightarrow \Phi(x)$ η σχέση μεταξύ x και y μπορεί να μετατραπεί σε ένα πρόβλημα γραμμικής παλινδρόμησης μεταξύ $\Phi(x)$ και y . Η μορφή της εξίσωσης που περιγράφει τη σχέση της γραμμικής παλινδρόμησης μεταξύ $\Phi(x)$ και y είναι (εξίσωση 1):

$$f(x, \omega) = \omega \cdot \Phi(x) + b \quad (1)$$

Ο σκοπός της παλινδρόμησης των ΜΔΥ είναι να βρεθούν τα βέλτιστα ω , b και οι κατάλληλες παράμετροι στη συνάρτηση $\Phi(x)$ ώστε να κατασκευαστεί μία προσέγγιση της συνάρτησης $f(x, \omega)$. Επιπλέον στόχος της παλινδρόμησης των ΜΔΥ είναι να βρεθεί μία συνάρτηση f η οποία έχει το πολύ απόκλιση ϵ από τους στόχους, για όλα τα δεδομένα εκπαίδευσης, και ταυτόχρονα να είναι όσο το δυνατόν περισσότερο επίπεδη (Smola and Scholkopf 2003). Σφάλματα μεγαλύτερα από ϵ δεν είναι αποδεκτά. Το να είναι η συνάρτηση $f(x, \omega)$ επίπεδη σημαίνει ότι επιδιώκεται ένα πολύ μικρό ω και αυτό επιτυγχάνεται με την ελαχιστοποίηση του όρου $1/2\|\omega\|^2$. Το πρόβλημα αυτό μπορεί να γραφτεί ως ένα πρόβλημα κυρτής βελτιστοποίησης (εξίσωση 2):

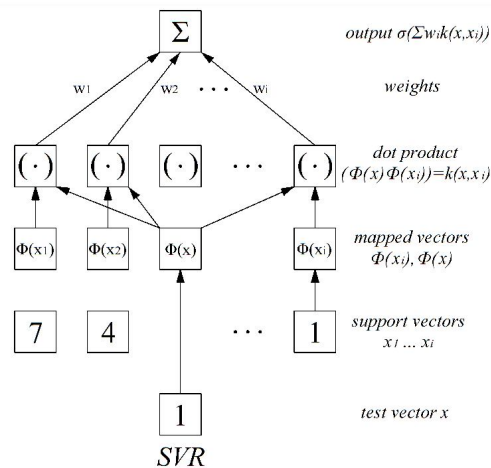
$$\text{minimize } \frac{1}{2} \|\omega\|^2 \quad \text{subject to } \begin{cases} y - (\omega \cdot \Phi(x)) - b \leq \epsilon \\ (\omega \cdot \Phi(x)) + b - y \leq \epsilon \end{cases} \quad (2)$$

Η σιωπηρή υπόθεση στην εξίσωση (2) ήταν ότι μια τέτοια συνάρτηση f πραγματικά υπάρχει προσεγγίζοντας όλα τα ζευγάρια των εισόδων και στόχων (x,y) με ακρίβεια ϵ , ή με άλλα λόγια, ότι το πρόβλημα κυρτής βελτιστοποίησης είναι εφικτό (Smola and Scholkopf 2003). Μερικές φορές, ωστόσο, αυτό μπορεί να μην συμβεί ή το πρόβλημα μπορεί να αναγνωρίσει μεγαλύτερα σφάλματα.

Σε αυτήν την περίπτωση το αρχικό πρόβλημα της εξίσωσης (2) τροποποιείται με την εισαγωγή των μεταβλητών χαλαρότητας ξ_i, ξ_i' για τις μετρήσεις “πάνω” και “κάτω” από το πεδίο σφάλματος ϵ , ενώ και οι δύο παίρνουν θετικές τιμές (Διαμαντάρας 2007). Οι μεταβλητές χαλαρότητας εισάγονται προκειμένου να αντιμετωπιστούν οι ανέφικτοι περιορισμοί του προβλήματος βελτιστοποίησης της εξίσωσης (2). Για το λόγο αυτό προκύπτει η μορφή της εξίσωσης (3) (Vapnik 1995).

$$\text{minimize } \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i') \quad \text{subject to } \begin{cases} y - (\omega, \Phi(x)) - b \leq \epsilon \\ (\omega, \Phi(x)) + b - y \leq \epsilon \\ \xi_i, \xi_i' \end{cases} \quad (3)$$

Στην εξίσωση (3) ο πρώτος όρος $1/2\|\omega\|^2$ παριστάνει τη γενίκευση και ο δεύτερος όρος $C \cdot \Sigma(\xi_i + \xi_i')$ αφορά το εμπειρικό ρίσκο. Το C είναι μία θετική παράμετρος η οποία ορίζει το βαθμό σφάλματος όταν προκύψει κάποιο λάθος στην εκπαίδευση (Wu et al 2008). Η αρχιτεκτονική της μηχανής παλινδρόμησης απεικονίζεται στην εικόνα 2.



Εικόνα 2: Αρχιτεκτονική παλινδρόμησης μηχανών διανυσμάτων υποστήριξης
Figure 2: Architecture of support vector regression

Η απόδοση του ομοιώματος παλινδρόμησης των ΜΔΥ εξαρτάται από την επιλογή της συνάρτησης πυρήνα και την επιλογή των υπερ-παραμέτρων. Στην παρούσα εφαρμογή προσομοιώθηκαν ο γραμμικός πυρήνας (linear kernel), ο πολυωνυμικός (polynomial kernel) και ο Γκαουσιανός (Gaussian or radial basis function - RBF) μεταβάλλοντας την παράμετρο που ελέγχει το εύρος της Γκαουσιανής συνάρτησης. Συγκριμένα για την Γκαουσιανή συνάρτηση πυρήνα χρησιμοποιήθηκαν τρεις διαφορετικές τιμές της παραμέτρου $g=1,2\&3$ που αντιπροσωπεύει το εύρος της συνάρτησης. Στη συνέχεια παρουσιάζονται οι εξισώσεις των τριών συναρτήσεων πυρήνα (4), (5) και (6).

Γραμμική: $k(x, y) = \langle x, y \rangle$ (4)

Πολυωνυμική: $k(x, y) = [\langle x, y \rangle + \theta]^p$ (5)

Γκαουσιανή RBF: $k(x, y) = e^{-g\|x-y\|^2}$ (6)

Για την προσομοίωση του ομοιώματος της παλινδρόμησης των ΜΔΥ χρησιμοποιήθηκε το λογισμικό SVMlight το οποίο είναι πολύ ισχυρό και ικανό εργαλείο στην επίλυση πολύπλοκων προβλημάτων.

Επεξεργασία δεδομένων

Στην παρούσα εφαρμογή χρησιμοποιήθηκαν μετεωρολογικά στοιχεία από τέσσερεις σταθμούς μέτρησης στην λεκάνη απορροής. Συγκεκριμένα οι εισοδοί του ομοιώματος ήταν το ύψος βροχής σε χιλιοστά (mm) και το ύψος χιονιού σε ισοδύναμο νερό επίσης σε χιλιοστά (mm). Αντίστοιχα οι στόχοι του ομοιώματος ήταν οι μετρήσεις της παροχής του χειμάρρου στην έξοδο της λεκάνης όπου υπάρχει εγκατεστημένος σταθμός μέτρησης σε κυβικά μέτρα ανά δευτερόλεπτο (m^3/sec). Τα μετεωρολογικά δεδομένα προέρχονται από το California Data Exchange Center (CDEC – Department of Water Resources) και οι μετρήσεις της παροχής από το U.S. Geological Survey (USGS).

Τόσο οι μετρήσεις της βροχόπτωσης και του χιονιού όσο και οι μετρήσεις της παροχής ήταν καθημερινές για μία χρονική περίοδο 21 ετών από τις 31-01-1989 έως 31-01-2011. Δεδομένου ότι οι συνθήκες υγρασίας του εδάφους, πριν από την εκδήλωση ενός ακραίου φαινομένου, επηρεάζουν σημαντικά την υδρολογική συμπεριφορά των λεκανών (Findell and Eltahir 1997), στις εισόδους του ομοιώματος χρησιμοποιήθηκε και η ιστορία της βροχόπτωσης και χιονόπτωσης για τις δύο προηγούμενες ημέρες. Έτσι στο ομοίωμα διαμορφώθηκαν 8036 ζευγάρια εισόδων-εξόδων, κάθε ένα από τα οποία αποτελούνταν από 12 εισόδους και μία έξοδο, έχοντας την παρακάτω μορφή:

$$\text{εισοδοί:} \begin{cases} x_1, x_1 - 1, x_1 - 2 \\ x_2, x_2 - 1, x_2 - 2 \\ x_3, x_3 - 1, x_3 - 2 \\ x_4, x_4 - 1, x_4 - 2 \end{cases} \rightarrow \text{στόχος ή έξοδος: } y$$

Μια εξαιρετική ιδιότητα των ΜΔΥ είναι η ικανότητα τους να διαχειρίζονται τη μη-γραμμικότητα, το θόρυβο και τα επισφαλή δεδομένα (Wu et al 2008). Επιπλέον με την κατάλληλη επεξεργασία των δεδομένων είναι πιθανό να βελτιωθεί η απόδοση των ομοιωμάτων και να προκύψουν καλύτερα αποτελέσματα (Maier and Dandy 2000). Αρχικά πραγματοποιήθηκε γραμμική παρεμβολή των δεδομένων με σκοπό να διορθωθούν πιθανά λάθη μετρήσεων. Συγκεκριμένα υλοποιήθηκαν διορθώσεις με τη μέθοδο αυτή σε εξαιρετικά μεγάλες τιμές, οι οποίες δεν θα μπορούσαν να καταγράφουν μία πραγματική μέτρηση, ενώ με την ίδια μέθοδο έγινε και συμπλήρωση των κενών στις χρονοσειρές. Στη συνέχεια δημιουργήθηκαν δέκα διαφορετικές τοπολογίες σε κάθε μία από τις οποίες τα δεδομένα διαιρέθηκαν με διαφορετικό και τυχαίο τρόπο σε τρία σύνολα. Έτσι σε κάθε τοπολογία το 70% των δεδομένων χρησιμοποιήθηκαν για εκπαίδευση (training), το 15% για επικύρωση (validation) και το υπόλοιπο 15% για έλεγχο (testing).

Κριτήρια αξιολόγησης

Για τον έλεγχο της απόδοσης και αξιοπιστίας των ομοιωμάτων χρησιμοποιήθηκαν τρία διαφορετικά κριτήρια αξιολόγησης τα οποία είναι το μέσο τετραγωνικό σφάλμα (MSE), η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (RMSE) και ο συντελεστής συσχέτισης (R-value). Κάθε κριτήριο αντιπροσωπεύεται από ένα δείκτη ο οποίος προσδιορίζεται από τη σύγκριση των τιμών πρόβλεψης του ομοιώματος και των στόχων που τέθηκαν στο ομοίωμα και αποτελούν πραγματικές μετρήσεις. Οι δείκτες αυτοί των κριτηρίων αξιολόγησης υπολογίζονται για κάθε υποσύνολο δεδομένων (εκπαίδευση, επικύρωση και έλεγχο) σε κάθε μία από τις δέκα τοπολογίες. Στη συνέχεια υπολογίζεται ο μέσος όρος των

δεικτών από τις δέκα τοπολογίες για τα τρία υποσύνολα, έτσι ώστε να ελαχιστοποιηθεί το σφάλμα της τυχαιότητας που μπορεί να εμπεριέχει μία πολύ καλή ή μία πολύ κακή εφαρμογή του ομοιώματος.

Το μέσο τετραγωνικό σφάλμα (MSE) μπορεί να χρησιμοποιηθεί για να ποσοτικοποιήσει την απόκλιση μίας τιμής που εκτιμήθηκε σε σχέση με την πραγματική τιμή της ποσότητας που ερευνάται. Στις ΜΔΥ το μέσο τετραγωνικό σφάλμα χρησιμοποιείται για να μετρήσει την απόδοση της διαδικασίας εκπαίδευσης της μηχανής. Το μέσο τετραγωνικό σφάλμα ορίζεται ως ο μέσος όρος των αθροισμάτων των τετραγώνων των διαφορών μεταξύ των στόχων και των τιμών πρόβλεψης (εξίσωση 7).

$$MSE = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2 \quad (7)$$

Η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (RMSE) είναι ένα μέτρο των διαφορών μεταξύ των τιμών που εκτιμήθηκαν από το ομοίωμα σε σχέση με τις πραγματικές τιμές των μετρήσεων (εξίσωση 8).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (8)$$

Ο συντελεστής συσχέτισης είναι (R-value) ένα μέτρο της γραμμικής παλινδρόμησης μεταξύ των τιμών που εκτιμήθηκαν και των στόχων του ομοιώματος (εξίσωση 9).

$$R = \frac{\sum_{i=1}^N (x_i - x)(y_i - y)}{\sqrt{\sum_{i=1}^N (x_i - x)^2} \sqrt{\sum_{i=1}^N (y_i - y)^2}} \quad (9)$$

Όπου N είναι ο αριθμός των δειγμάτων, x_i και y_i είναι οι στόχοι και οι εκτιμώμενες τιμές αντίστοιχα για $i=1, \dots, n$ και x και y είναι οι μέσες τιμές των στόχων και των εκτιμώμενων τιμών αντίστοιχα.

Όσο μικρότερο είναι το μέσο τετραγωνικό σφάλμα και η τετραγωνική ρίζα του και όσο ο συντελεστής συσχέτισης προσεγγίζει τη μονάδα τόσο καλύτερη είναι η συμπεριφορά του ομοιώματος και παράγει τα βέλτιστα αποτελέσματα.

Αποτελέσματα

Η απόδοση του ομοιώματος της παλινδρόμησης των ΜΔΥ διακρίνεται στον πίνακα 1. Στον πίνακα αυτό καταγράφονται οι μέσες τιμές των δεικτών που εκφράζουν τα κριτήρια αξιολόγησης του ομοιώματος για τους τρεις τύπους των συναρτήσεων πυρήνων που προσομοιώθηκαν.

Από τα αποτελέσματα του πίνακα προκύπτει ότι το ομοίωμα με τη γραμμική συνάρτηση πυρήνα δίνει τα καλύτερα αποτελέσματα και προσομοιώνει ιδανικότερα τη σχέση μεταξύ βροχόπτωσης και απορροής στη συγκεκριμένη λεκάνη απορροής. Περισσότερο αναλυτικά προκύπτει ότι η γραμμική συνάρτηση πυρήνα παρουσιάζει τις μικρότερες μέσες τιμές της τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (RMSE), οι οποίες είναι αντίστοιχα για τα υποσύνολα εκπαίδευσης, επικύρωσης και ελέγχου των δεδομένων: 15.52, 16.30 και 14.58, γεγονός που τεκμηριώνει την υπεροχή της σε σχέση με τις άλλες δύο συναρτήσεις πυρήνα, δηλαδή την πολυωνυμική και την Γκαουσιανή.

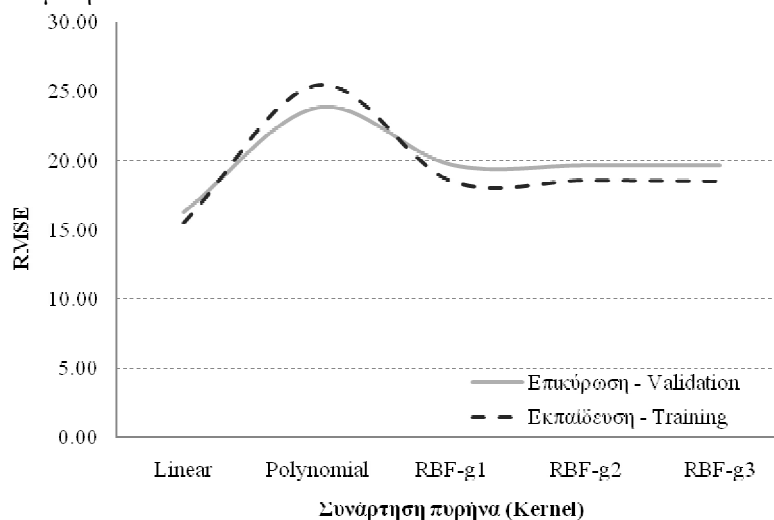
Συγκρίνοντας τις άλλες δύο συναρτήσεις πυρήνα μεταξύ τους, προκύπτει ότι το ομοίωμα με την Γκαουσιανή συνάρτηση πυρήνα επιδεικνύει μία καλύτερη συμπεριφορά σε σχέση με αυτό που προσομοιώνει την πολυωνυμική. Πιο

συγκεκριμένα οι τιμές της τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (RMSE), για τα υποσύνολα των δεδομένων εκπαίδευσης, επικύρωσης και ελέγχου είναι: 18.50, 19.71 και 17.64 για την Γκαουσιανή με πλάτος 1, 18.56, 19.68 και 17.70 για την Γκαουσιανή με πλάτος 2, 18.49, 19.67 και 17.67 για την Γκαουσιανή με πλάτος 3 και αντίστοιχα για την πολυωνμική είναι 25.45, 23.81 και 20.69. Από τη σύγκριση είναι εμφανές ότι η Γκαουσιανή συνάρτηση και με τα τρία διαφορετικά πλάτη παρουσιάζει μικρότερες τιμές σε σχέση με την πολυωνμική.

Πίνακας 1: Τιμές κριτηρίων αξιολόγησης ομοιώματος SVR
Table 1: Evaluation criteria values of SVR model

		SVR - models				
		Linear	P/mial	RBF-g1	RBF-g2	RBF-g3
Training	MSE	242.50	717.45	343.79	345.89	343.34
	RMSE	15.52	25.45	18.50	18.56	18.49
	R	0.5669	0.3541	0.4106	0.4126	0.4106
Validation	MSE	281.52	653.17	403.96	402.58	402.38
	RMSE	16.30	23.81	19.71	19.68	19.67
	R	0.5821	0.5574	0.4421	0.4382	0.4363
Testing	MSE	227.87	519.16	326.03	327.84	326.84
	RMSE	14.58	20.69	17.64	17.70	17.67
	R	0.5971	0.5153	0.4584	0.4670	0.4554

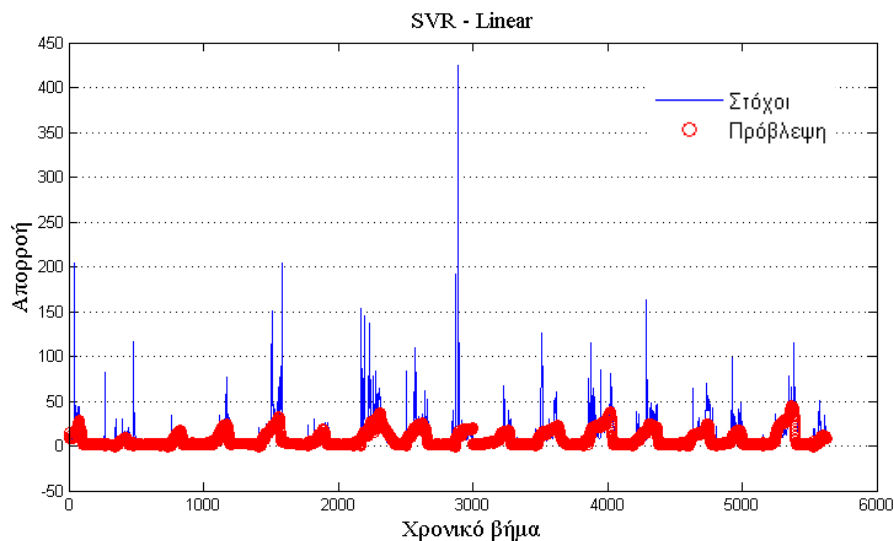
Η απόδοση του ομοιώματος παλινδρόμησης των ΜΔΥ παρουσιάζεται γραφικά στην εικόνα 3. Ανάλογο συμπέρασμα με τα παραπάνω προκύπτει και από τις καμπύλες εκπαίδευσης και επικύρωσης που καταδεικνύουν ότι τα καλύτερα αποτελέσματα παράγονται από το ομοίωμα που προσομοιώνει τη γραμμική συνάρτηση πυρήνα και τα χειρότερα από το ομοίωμα που προσομοιώνει την πολυωνμική.



Εικόνα 3: Απόδοση ομοιώματος SVR
Figure 3: Performance of SVR

Σχετικά με την Γκαουσιανή συνάρτηση είναι χαρακτηριστικό ότι παρουσιάζει μία σταθερότητα και για τις τρεις τιμές της παραμέτρου που καθορίζει το εύρος της. Γενικά φαίνεται ότι κατά τη μεταβολή του πλάτους της Γκαουσιανής συνάρτησης από 1 σε 2 και στη συνέχεια σε 3 το ομοίωμα δεν αντιδρά και συνεχίζει δίνοντας παρόμοια αποτελέσματα.

Στην εικόνα 4 παρουσιάζεται γραφικά η εφαρμογή με τα καλύτερα αποτελέσματα από την προσομοίωση του ομοιώματος παλινδρόμησης των ΜΔΥ με τη γραμμική συνάρτηση πυρήνα. Με μπλέ χρώμα παριστάνονται οι στόχοι που τέθηκαν στο ομοίωμα και αποτελούν τις μετρημένες τιμές της παροχής και με κόκκινο χρώμα είναι οι τιμές πρόβλεψης του ομοιώματος. Διακρίνεται χαρακτηριστικά η ικανοποιητική προσομοίωση της σχέσης βροχοπτώσης-απορροής αλλά και η αδυναμία να ακολουθήσουν οι τιμές πρόβλεψης τις πολύ ακραίες τιμές γεγονός που προβληματίζει.



Εικόνα 4: Πρόβλεψη απορροής σε σύγκριση με τους στόχους του ομοιώματος
Figure 4: Predicted runoff compared with corresponding flows (targets)

Συμπεράσματα

Η παρούσα ερευνητική εργασία καταδεικνύει ότι τα ομοιώματα παλινδρόμησης των ΜΔΥ μπορούν να χρησιμοποιηθούν για τη διερεύνηση της σχέσης μεταξύ βροχοπτώσης-απορροής και την πρόβλεψη χρονοσειρών απορροής. Ωστόσο κατά την προσομοίωση των ομοιωμάτων προέκυψαν και κάποιες αδυναμίες οι οποίες χρήζουν βαθύτερης έρευνας, με κυριότερη την αδυναμία να ακολουθήσουν οι τιμές πρόβλεψης τις πολύ ακραίες τιμές.

Οι Han και Yang το 2001 επισήμαναν το ίδιο πρόβλημα κατά την προσομοίωση της παροχής χειμάρρου σε μία λεκάνη απορροής (Bird Creek) στις ΗΠΑ εφαρμόζοντας ένα ανάλογο ομοίωμα παλινδρόμησης ΜΔΥ. Επιπλέον οι Bray και Han το 2004 κατέδειξαν τις δυσκολίες των ομοιωμάτων παλινδρόμησης των ΜΔΥ στην πρόβλεψη πλημμυρικών παροχών.

Κύριο πλεονεκτήματα τους είναι η δυνατότητα επεξεργασίας μεγάλου όγκου δεδομένων σε σύντομο χρονικό διάστημα, γεγονός που τα κάνει να υπερέχουν σε σχέση με άλλα ομοιώματα. Επίσης διαπιστώθηκε ότι η επιλογή της συνάρτησης

πυρήνα έχει καίρια σημασία για τη συμπεριφορά του ομοιώματος και συγκεκριμένα από την παρούσα εφαρμογή προέκυψε ότι το ομοίωμα με τη γραμμική συνάρτηση πυρήνα έδωσε τα καλύτερα αποτελέσματα. Ο Bray το 2002 κατέληξε στο ίδιο συμπέρασμα διαπιστώνοντας ότι η γραμμική συνάρτηση πυρήνα επέδειξε την καλύτερη συμπεριφορά στην προσομοίωση της απορροής συγκρίνοντας την με άλλες συναρτήσεις πυρήνα (Γκαουσιανή, πολυωνυμική και σιγμοειδή).

Από την παρούσα εφαρμογή προέκυψε ότι η χρήση της βροχόπτωσης και χιονόπτωσης των προηγούμενων δύο ημερών στις εισόδους του ομοιώματος μαζί με την ημερήσια βροχόπτωση και χιονόπτωση βελτίωσαν αισθητά τη συμπεριφορά του ομοιώματος και τα εξαγόμενα αποτελέσματα. Το γεγονός αυτό επιβεβαιώνει το φυσικό φαινόμενο ότι στη δημιουργία της απορροής παίζει σημαντικό ρόλο η υγρασία του εδάφους και κατά πόσο αυτό είναι κορεσμένο σε νερό.

Γενικά προκύπτει ότι η παλινδρόμηση των ΜΔΥ μπορεί να χρησιμοποιηθεί στη διερεύνηση της σχέσης βροχόπτωσης-απορροής, ωστόσο υπάρχουν αρκετά πεδία που χρήσουν περισσότερης και βαθύτερης έρευνας. Τέλος μία αντικειμενική δυσκολία είναι η γενίκευση των ομοιωμάτων σε περισσότερες και με διαφορετικά χαρακτηριστικά λεκάνες απορροής.

Rainfall – runoff modeling using support vector regression

Dimitrios Botsis¹, Periklis Latinopoulos², Konstantinos Diamantaras³

^{1&2} Department of Civil Engineering, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece, ³ Department of Informatics, Alexander Technological Education Institute of Thessaloniki, GP-57400 Sindos, Greece
e-mail: jimbotsis@civil.auth.gr, latin@civil.autn.gr, kdiamant@it.teithe.gr

Abstract

In this study a support vector regression (SVR) model designed to estimate the relationship between rainfall and runoff, which describes one of the most complex hydrologic phenomena. This relationship is nonlinear and it is suggested that nonlinear models like SVR may have notable advantages in estimating rainfall–runoff mapping. As a matter of fact, studies in the literature have reported that SVR are very powerful in terms of nonlinear mapping. The daily rainfall, the rainfall of the last two days and streamflow data of a mountainous watershed were used as a case study for developing the rainfall-runoff model of this investigation. The watershed is located in northern California in a region with the same geographic latitude as Northern Greece. The dataset of rainfall and streamflow is divided into three parts; the first is used for training, the second for testing and the third for validating the model. The performance of the SVR model depends on the choice of the kernel function and the model parameters. In the present study, the linear, the polynomial kernel, the Gaussian (or RBF) kernel were used. Each kernel function has one parameter, which could influence the behavior of the SVR model. The performance and reliability of SVR was evaluated through three different criteria (correlation coefficient, mean square error and root mean square error). SVR has been found effective in rainfall-runoff simulation by many researchers in a variety of applications. The results of the study demonstrate that the SVR model is effective in simulating the rainfall–runoff relationship.

Βιβλιογραφία

- Bray M. and Han D., 2004. Identification of support vector machines for runoff modeling. *Journal of Hydroinformatics*, 6, 265–280.
- Bray M., 2002. Identification of Support Vector Machines for Runoff Modelling. MEng thesis, Department of Civil Engineering, University of Bristol.
- Chang M., 2003. *Forest hydrology: an introduction to water and forests*. CRC Press.
- Findell, K.L. and Eltahir E.A.B., 1997. An analysis of the soil moisture-rainfall feedback, based on direct observations from Illinois. *Water Resour. Res.* 33, 725–735.
- Han D. and Yang Z., 2001. River flow modelling using support vector machines. In XXIX IAHR Congress, Beijing, China, 17–21 September, pp. 494–499, Qinghua University Press, China.
- Lin J.Y., Cheng C.T. and Chau K.W., 2006. Using support vector machines for long-term discharge prediction. *Hydrol. Sci. – J.* 51 (4), 599–611.
- Maier H.R. and Dandy G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications, *Environ. Model Software* 15 (1), 101–123.
- Rajurkar M.P., Kothiyari U.C. and Chaube U.C., 2002. Artificial neural networks for daily rainfall-runoff modeling, *Hydrological Sciences Journal*, Vol. 47 (6), 865 – 877.
- Sivapragasam C. and Liong S.Y., 2004. Identifying optimal training data set – a new approach. In: Liong, S.Y., Phoon, K.K., Babovic, V. (Eds.), *Proceedings of the Sixth International Conference on Hydroinformatics*, Singapore, 21–24 June 2004. World Scientific Publishing Co., Singapore.
- Sivapragasam, C. and Liong, S.Y., 2005. Flow categorization model for improving forecasting. *Nordic Hydrol.* 36 (1), 37–48.
- Smola A.J. and Scholkopf B., 2003. *A tutorial of support vector regression*.
- Solaimani K., 2009. Rainfall-runoff Prediction Based on Artificial Neural Network (A Case Study: Jarahi Watershed). *American-Eurasian J. Agriculture & Environment*, Vol 5 (6), 856 – 865.
- Vapnik V. and Chervonenkis A., 1964. A note on one class of perceptrons. *Automation and Remote Control*, 25.
- Vapnik V. and Lerner A., 1963. Pattern recognition using generalized portrait method. *Automation and Remote Control*, 24, 774–780.
- Vapnik V., 1995. *The Nature of Statistical Learning Theory*. Springer, New York.
- Wu C.L., Chau, K.W. and Li Y.S., 2008. River stage prediction based on a distributed support vector regression. *Journal of hydrology*, 358, 96–111.
- Διαμαντάρας Κ., 2007. *Τεχνητά Νευρωνικά Δίκτυα*. Κλειδάριθμος. Αθήνα.
- Κουτσογιάννης Δ. και Ξανθόπουλος Θ., 1999. *Τεχνική Υδρολογία*. Εθνικό Μετσόβειο Πολυτεχνείο.
- Κωτούλας Δ., 1995. *Μαθήματα δασικής υδρολογίας*. Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης. Έκδοση: Υπηρεσία Δημοσιευμάτων ΑΠΘ. Θεσσαλονίκη.
- Λατινόπουλος Π., 2007. Προστασία και εξυγίανση υπόγειων νερών. Σημειώσεις, Πρόγραμμα Μεταπτυχιακών Σπουδών Προστασία Περιβάλλοντος και Βιώσιμη Ανάπτυξη, Α.Π.Θ. Θεσσαλονίκη.