



ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΣΤΡΑΤΗΓΙΚΕΣ ΔΙΑΧΕΙΡΙΣΗΣ ΠΕΡΙΒΑΛΛΟΝΤΟΣ, ΚΑΤΑΣΤΡΟΦΩΝ & ΚΡΙΣΕΩΝ

POST GRADUATE PROGRAM
ENVIRONMENTAL, DISASTER & CRISES MANAGEMENT STRATEGIES

Μεταπτυχιακή Διατριβή Ειδίκευσης
Master Thesis

Φυσικοί κίνδυνοι: Η χρήση data mining σε συστήματα έγκαιρης προειδοποίησης για την πρόβλεψη πλημμυρών. Ένα παράδειγμα από την Πολιτεία της Νέας Υόρκης

Natural hazards response: Data mining and online warning system for flooding prediction. An example from New York State

ΑΝΤΩΝΙΟΣ ΚΑΒΑΛΙΕΡΟΣ / ANTONIOS KAVALIEROS
Α.Μ. / R.N. : 14029

Ειδικές Εκδόσεις / Special Publications:

No. 2016007

Αθήνα, Οκτώβριος 2016
Athens, October 2016



ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΣΤΡΑΤΗΓΙΚΕΣ ΔΙΑΧΕΙΡΙΣΗΣ ΠΕΡΙΒΑΛΛΟΝΤΟΣ, ΚΑΤΑΣΤΡΟΦΩΝ & ΚΡΙΣΕΩΝ

POST GRADUATE PROGRAM
ENVIRONMENTAL, DISASTER & CRISES MANAGEMENT STRATEGIES

Μεταπτυχιακή Διατριβή Ειδίκευσης

Master Thesis

Φυσικοί κίνδυνοι: Η χρήση data mining σε συστήματα έγκαιρης προειδοποίησης για την πρόβλεψη πλημμυρών. Ένα παράδειγμα από την Πολιτεία της Νέας Υόρκης

Natural hazards response: Data mining and online warning system for flooding prediction. An example from New York State

ΑΝΤΩΝΙΟΣ ΚΑΒΑΛΙΕΡΟΣ / ANTONIOS KAVALIEROS

A.M. / R.N. : 14029

Τριμελής Εξεταστική Επιτροπή:

Δρ. Κ. Κυριακόπουλος,
Καθηγ. ΕΚΠΑ

Δρ. Σ. Κίλιας,
Καθηγ. ΕΚΠΑ

Δρ. Α. Μερτζάνης,
Αναπλ. Καθηγ. ΤΕΙ Στερεάς Ελλάδας



**NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS
FACULTY OF GEOLOGY AND GEOENVIRONMENT
SCHOOL OF SCIENCES
POSTGRADUATE PROGRAM
“STRATEGIES FOR ENVIRONMENTAL, CRISIS AND DISASTER
MANAGEMENT”**

POSTGRADUATE DESSERTATION TOPIC

**«Natural hazards response: Data mining and online
warning system for flooding prediction. An example
from New York State»**

**KAVALIEROS ANTONIOS
GEOLOGIST**

ATHENS 2016

**NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS
FACULTY OF GEOLOGY AND GEOENVIRONMENT
SCHOOL OF SCIENCES
POSTGRADUATE PROGRAM
“STRATEGIES FOR ENVIRONMENTAL, CRISIS AND DISASTER
MANAGEMENT”**

POSTGRADUATE DESSERTATION TOPIC

**«Natural hazards response: Data mining and online
warning system for flooding prediction. An example
from New York State»**

**KAVALIEROS ANTONIOS
GEOLOGIST**

SUPERVISORS:

KIRIAKOPOULOS KONTANTINOS, PROFESSOR, UOA, GR

MARSELLOS ANTONIOS, ASSISTANCE PROFESSOR, HOFSTRA UNIVERSITY, NY

TSAKIRI KATERINA, SENIOR LECTURER, UNIVERSITY OF BRIGHTON, UK

EXAMINING BOARD:

KIRIAKOPOULOS KONTANTINOS, PROFESSOR, UOA

KILIAS STEFANOS, PROFESSOR, UOA

MERTZANIS ARISTIDIS, ASSISTANT PROFESSOR, TEI OF LAMIA

AUGUST 2016

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΤΜΗΜΑ ΓΕΩΛΟΓΙΑΣ ΚΑΙ ΓΕΩΠΕΡΙΒΑΛΛΟΝΤΟΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
“ΣΤΡΑΤΗΓΙΚΕΣ ΔΙΑΧΕΙΡΙΣΗΣ ΠΕΡΙΒΑΛΛΟΝΤΟΣ, ΚΑΤΑΣΤΡΟΦΩΝ
ΚΑΙ ΚΡΙΣΕΩΝ”

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**«Φυσικοί κίνδυνοι: Η χρήση data mining σε συστήματα
έγκαιρης προειδοποίησης για την πρόβλεψη πλημμυρών.
Ένα παράδειγμα από την Πολιτεία της Νέας Υόρκης»**

ΚΑΒΑΛΙΕΡΟΣ ΑΝΤΩΝΙΟΣ
ΓΕΩΛΟΓΟΣ

ΕΠΙΒΛΕΠΟΝΤΕΣ:

ΚΥΡΙΑΚΟΠΟΥΛΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ, ΚΑΘΗΓΗΤΗΣ, ΥΟΑ, GR
ΜΑΡΣΕΛΛΟΣ ΑΝΤΩΝΙΟΣ, ΑΝΑΠΛΗΡΩΤΗΣ ΚΑΘΗΓΗΤΗΣ, HOFSTRA UNIVERSITY, NY
ΤΣΑΚΙΡΗ ΑΙΚΑΤΕΡΙΝΗ, ΛΕΚΤΟΡΑΣ, UNIVERSITY OF BRIGHTON, UK

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:

ΚΥΡΙΑΚΟΠΟΥΛΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ, ΚΑΘΗΓΗΤΗΣ ΕΚΠΑ
ΚΙΛΙΑΣ ΣΤΕΦΑΝΟΣ, ΚΑΘΗΓΗΤΗΣ ΕΚΠΑ
ΜΕΡΤΖΑΝΗΣ ΑΡΙΣΤΕΙΔΗΣ, ΕΠΙΚΟΥΡΟΣ ΚΑΘΗΓΗΤΗΣ ΤΕΙ ΛΑΜΙΑΣ

ΑΥΓΟΥΣΤΟΣ 2016

Περιεχόμενα

Περιεχόμενα.....	v
Περίληψη	vii
Abstract	viii
Πνευματικά δικαιώματα.....	viii
Κατάλογος Σχημάτων	ix
Κατάλογος Πινάκων	x
Κατάλογος Εικόνων	xi

Κεφάλαιο 1. Εισαγωγή

1.1. Εισαγωγή.....	12
1.2. Θεωρητικό μέρος	13

Κεφάλαιο 2. Μεθοδολογία

2.1. Εισαγωγή.....	17
2.2. Δεδομένα	18
2.3. Αποσύνθεση των χρονοσειρών.....	21
2.4. Ανάλυση ακατέργαστων δεδομένων των χρονοσειρών της απορροής του νερού	24
2.5. Πρόβλεψη των μακροπρόθεσμων συστατικών της απορροής του νερού	30
2.6. Πρόβλεψη των εποχιακών συστατικών της απορροής του νερού	33
2.7. Πρόβλεψη των βραχυπρόθεσμων συστατικών της απορροής του νερού	36
2.7.1. Μοντέλο θερινής περιόδου.....	37
2.7.2. Μοντέλο χειμερινής περιόδου	38

Κεφάλαιο 3. Αποτελέσματα

40

Κεφάλαιο 4. Συζήτηση-Συμπεράσματα.....

45

Βιβλιογραφία

47

Περίληψη

Πλημμύρες συμβαίνουν συνήθως κατά τη διάρκεια περιόδων με υπερβολική βροχόπτωση ή τήξη πάγου κατά τη χειμερινή περίοδο και μπορεί να επιφέρουν τεράστιες επιπτώσεις. Αυτές μπορεί είναι είτε οικονομικές – υλικές (όπως απώλεια περιουσίας, καταστροφές σε δημόσια περιουσία), είτε άυλες (όπως τραυματισμοί, απώλεια ζωής). Η πρόβλεψη τους λοιπόν είναι αναγκαία, αλλά αποτελεί και πρόκληση. Ωστόσο, υπάρχουν ενδείξεις. Οι πλημμύρες συνήθως συνοδεύονται από μια αύξηση στην απορροή του νερού του ποταμού. Το μοντέλο των Tsakiri et al., 2014, που χρησιμοποιείται στην παρούσα εργασία, είναι ικανό να προβλέψει την αύξηση αυτή σε μεγάλο ποσοστό. Υπάρχουν πολλά διαθέσιμα δεδομένα προς διαχείριση, από διάφορους φορείς, που μπορούν να χρησιμοποιηθούν από το μοντέλο αυτό. Ωστόσο, τα δεδομένα αυτά έχουν τεράστιο όγκο και γι' αυτό οι περισσότεροι επιστήμονες περιορίζονται σε μελέτες περίπτωσης. Την λύση έρχεται να δώσει το data mining. Τα λογισμικά data mining μπορούν να διαχειριστούν αυτόν τον όγκο και μάλιστα να λειτουργούν σε πραγματικό χρόνο. Η εισαγωγή δεδομένων και η εξαγωγή συμπερασμάτων σε πραγματικό χρόνο, μας δίνει την δυνατότητα να εξετάσουμε το ενδεχόμενο της χρήσης τους σε συστήματα έγκαιρης προειδοποίησης για την πρόβλεψη πλημμύρων. Το Ktime, που χρησιμοποιείται στην παρούσα εργασία, αποτελεί ένα καλό παράδειγμα τέτοιου λογισμικού. Παρουσιάζεται, λοιπόν, μια ροή εργασίας που χρησιμοποιεί ένα στατιστικό μοντέλο για την πρόβλεψη και την εξήγηση των χρονοσειρών της απορροής του νερού και μπορεί να χρησιμοποιηθεί σε πραγματικό χρόνο, χρησιμοποιώντας ένα παράδειγμα από την Νέα Υόρκη. Το μοντέλο, για την πρόβλεψη της απορροής του νερού, κάνει αποσύνθεση των χρονοσειρών των υδρολογικών και κλιματικών μεταβλητών σε εποχιακές, μακροπρόθεσμες και βραχυπρόθεσμες συνιστώσες. Επίσης, αναλύει την απορροή του νερού χρησιμοποιώντας ένα θερινό και ένα χειμερινό μοντέλο. Με αυτόν τον τρόπο βελτιώνεται η εξήγηση της απορροής του νερού έως και 81%. Το μοντέλο δείχνει ότι όσο αυξάνεται η απορροή του νερού, μακροπρόθεσμα ο υδροφόρος ορίζοντας αναπληρώνεται ενώ, εποχιακά εξαντλείται. Βραχυπρόθεσμα, το επίπεδο των υπόγειων υδάτων πέφτει κατά τη διάρκεια της χειμερινής περιόδου και αυξάνεται κατά τη θερινή περίοδο.

Abstract

Floods usually occur during periods of excessive precipitation or thawing in the winter period (ice jam) and have a huge impact. The effects of natural disasters are both tangible – financial (such as loss of property and damage to public property) and intangible (such as injuries and loss of life). Their prediction is therefore needed and constitutes a challenge for conducting research. However, there are indications. Floods are usually accompanied by an increase in river discharge. The model devised by Tsakiri et al., 2014, which is used in the present work, is able to predict river's water discharge to a great extent. There are many data available to be processed by various bodies, using this model. However, because the volume of data is daunting, most scientists use case studies. Data mining techniques are used that aid to solve this problem. Data mining software can handle this volume of data (big data) efficiently and can even operate in real time. The advantage of resulting conclusions in real time makes it possible to envisage their use in early warning systems for predicting floods. Knime, which is used in this project, is a good example of such software. In this study it is presented a workflow that uses a statistical model for the prediction and explanation of the timeseries of water discharge in real time, using an example from New York State. This model for predicting water discharge does decomposition of the timeseries of hydrologic and climatic variables in seasonal, long-term and short-term components. It also analyzes runoff using a summer and a winter model. This improves the explanation of water discharge up to 81%. We also observe that with increasing water discharge in the long term, the water table is replenished, while in the seasonal term it depletes. In the short term, the groundwater level falls during the winter season and increases during the summer season.

Keywords

Real time data mining, flood warning system, water discharge prediction, time series decomposition, prediction model

Πνευματικά δικαιώματα

- **KNIME Analytics Platform**, (<https://www.knime.org/>)

The KNIME® trademark and logo is used by KNIME.com AG under license from KNIME GmbH, and are registered in the United States. KNIME® is also registered in Germany.

KNIME.com AG, Technoparkstr. 1, 8005, Zurich, Switzerland

- **R, version 3.2.3**, (<https://www.r-project.org/>)

The R logo is © 2016 The R Foundation.

The logo is under the terms of the Creative Commons Attribution-ShareAlike 4.0 International license (CC-BY-SA 4.0) or the GNU General Public License version 2 (GPL-2).

Κατάλογος Σχημάτων

Σχήμα 1.	Η ροή εργασίας.....	18
Σχήμα 2.	Η αποσύνθεση των χρονοσειρών με την χρήση κινούμενου μέσου.....	23
Σχήμα 3.	Σύνθεση κόμβων για δίσεκτα έτη	35
Σχήμα 4.	Τα ακατέργαστα δεδομένα των χρονοσειρών της απορροής του νερού για την θερινή περίοδο του 2006.....	42
Σχήμα 5.	Τα ακατέργαστα δεδομένα των χρονοσειρών της απορροής του νερού για την χειμερινή περίοδο του 2010	43
Σχήμα 6.	Η ροή εργασίας για την προβολή των σταθμών σε παγκόσμιο χάρτη	45

Κατάλογος Πινάκων

Πίνακας 1.	Η συσχέτιση μεταξύ των ακατέργαστων δεδομένων και των μεταβλητών του αρχικού μοντέλου.....	26
Πίνακας 2.	Ο διαχωρισμός των περιόδων	28
Πίνακας 3.	Η συσχέτιση μεταξύ των μακροπρόθεσμων συστατικών της απορροής του νερού και των υπόλοιπων μεταβλητών του αρχικού μοντέλου.....	31
Πίνακας 4.	Η συσχέτιση μεταξύ του εποχιακού συστατικού της απορροής του νερού και των υπόλοιπων μεταβλητών του αρχικού μοντέλου	34
Πίνακας 5.	Η συσχέτιση μεταξύ του βραχυπρόθεσμου συστατικού της απορροής του νερού και των υπόλοιπων μεταβλητών για την θερινή περίοδο	38
Πίνακας 6.	Η συσχέτιση μεταξύ του βραχυπρόθεσμου συστατικού της απορροής του νερού και των υπόλοιπων μεταβλητών για την χειμερινή περίοδο ...	39
Πίνακας 7.	Η περίληψη των διακυμάνσεων και του συντελεστή συσχέτισης των εποχιακών, μακροπρόθεσμων και βραχυπρόθεσμων συστατικών για την θερινή και την χειμερινή περίοδο	42

Κατάλογος Εικόνων

Εικόνα 1.	Το πολυγωνικό φίλτρο	20
Εικόνα 2.	Οι διαθέσιμοι σταθμοί	21
Εικόνα 3.	Η ομαλοποίηση των ακατέργαστων δεδομένων με την χρήση κινούμενου μέσου	25
Εικόνα 4.	Ο λογάριθμος της χρονοσειράς της απορροής του νερού κατά την περίοδο 2005-2013.....	26
Εικόνα 5 & 6.	Η συσχέτιση μεταξύ των ακατέργαστων δεδομένων και των μεταβλητών για το Long Island	27
Εικόνα 7.	Περιοδόγραμμα κατακρημνισμάτων.....	29
Εικόνα 8.	Περιοδόγραμμα θερμοκρασίας	29
Εικόνα 9.	Χάρτης της λεκάνης απορροής του Long Island.....	31
Εικόνα 10 & 11.	Η συσχέτιση μεταξύ των μακροπρόθεσμων συστατικών και των μεταβλητών για το Long Island	32
Εικόνα 12.	Οι επιλογές του κόμβου γραμμικής παλινδρόμησης και η μορφή των αποτελεσμάτων	37

Κεφάλαιο 1.

1.1. Εισαγωγή

Στις επιστήμες της γης, ο πλανήτης που κατοικούμε χαρακτηρίζεται ως ένα ενεργό σύστημα. Αυτό εξηγείται σε μεγάλο βαθμό από την θεωρία της τεκτονικής των λιθοσφαιρικών πλακών. Η σύγκλιση, η απόκλιση, η σύγκρουση και οι πλευρικές ολισθήσεις των πλακών ερμηνεύουν τους σεισμούς, τα ηφαίστεια, τη δημιουργία των οροσειρών, τη γένεση των ωκεανών, καθώς και φαινόμενα όπως η κλιματική αλλαγή και οι πλημμύρες. Φυσικοί κίνδυνοι ή αλλιώς φυσικά φαινόμενα, όπως οι σεισμοί, οι πλημμύρες, οι ηφαιστειακές εκρήξεις και η παγκόσμια μεταβολή του κλίματος, παρέχουν στους επιστήμονες παρατηρήσεις που μπορεί να χρησιμοποιηθούν για να προσδιορίσουν τις αιτίες αυτών των φαινομένων και πιθανά, για να προβλέψουν επακόλουθα φαινόμενα (Κράνης Χ., 2015). Για τον λόγο αυτό, εδώ και δεκαετίες η παρακολούθηση της γης αποτελεί ένα καθημερινό φαινόμενο. Καθημερινό φαινόμενο όμως για λίγους, καθώς το κόστος είναι απαγορευτικό.

Ωστόσο, αυτό στις μέρες μας έχει αρχίσει να αλλάζει με την βοήθεια της τεχνολογίας. Ειδικά η ραγδαία ανάπτυξη στον τομέα της παρακολούθησης (δημιουργία παντός τύπου αισθητήρων) τα τελευταία χρόνια, έχει δώσει την δυνατότητα σε ολοένα και περισσότερες χώρες να λαμβάνουν μετρήσεις σε καθημερινή βάση (daily data) για διάφορες παραμέτρους. Οι παράμετροι αυτοί μπορεί να αφορούν θερμοκρασίες, πιέσεις, κινήσεις στον χώρο (μέσω GPS) και πλήθος άλλων. Ωστόσο, ο όγκος των δεδομένων, καθώς και ο χρόνος που απαιτείται κάθε φορά για να εξαχθούν νέα αποτελέσματα είναι μεγάλος με αποτέλεσμα οι ερευνητές να αναγκάζονται, στις περισσότερες περιπτώσεις, να περιοριστούν σε μελέτες περίπτωσης. Η ανάκτηση των δεδομένων αυτών μπορεί να γίνει με πολλούς τρόπους, κυρίως επιλέγοντας κάποιες παραμέτρους για ένα χρονικό διάστημα από κάποιους σταθμούς παρακολούθησης. Στην συνέχεια, τα δεδομένα εισάγονται σε λογισμικά, επεξεργάζονται και εξάγονται συμπεράσματα. Η διαδικασία αυτή πέρα από πολύπλοκη είναι και χρονοβόρα, καθώς εάν για παράδειγμα, επιθυμούμε να αλλάξουμε περιοχή μελέτης ή χρονικό διάστημα παρατήρησης, είναι απαραίτητη η επανάληψη μέρους αυτής. Για τον λόγο αυτό, τεράστιος όγκος δεδομένων παραμένει ανεπεξεργαστος – αχρησιμοποίητος γεγονός που έχει σαν αποτέλεσμα τον περιορισμένο αριθμό εξαγόμενων συμπερασμάτων. Γίνεται λοιπόν αντιληπτό ότι το πρόβλημα παρουσιάζεται στην μεθοδολογία.

Την λύση έρχεται να δώσει το data mining. Η χρήση data mining έχει πολλά πλεονεκτήματα. Αρχικώς, μας δίνει την δυνατότητα να επεξεργαζόμαστε τεράστιο όγκο δεδομένων για την διεξαγωγή αποτελεσμάτων. Επίσης, μας παρέχει την δυνατότητα, αφού πρώτα έχουμε δημιουργήσει μια ροή εργασίας (workflow), να αλλάζουμε άμεσα τα δεδομένα που εισάγουμε, όπως για παράδειγμα χρονικό διάστημα, και να εξάγουμε ταχύτητα αποτελέσματα. Επιπροσθέτως, μερικά προγράμματα data mining, παρέχουν την δυνατότητα

εισαγωγής δεδομένων σε πραγματικό χρόνο (real-time data mining). Όπως είναι αντιληπτό, η χρήση αυτής της μεθόδου αποτελεί 'μονόδρομο' για την άμεση διεξαγωγή συμπερασμάτων και κατά συνέπεια κύριο εργαλείο στην δημιουργία συστημάτων έγκαιρης προειδοποίησης.

Τα συστήματα έγκαιρης προειδοποίησης αποτελούν σημαντικό στοιχείο για την μείωση του κινδύνου των καταστροφών. Αποτρέπουν την απώλεια της ζωής και μειώνουν την οικονομική και υλική επίπτωση των καταστροφών. Σύμφωνα με την United Nations International Strategy for Disaster Reduction (εφεξής UNISDR), ένα ολοκληρωμένο και αποτελεσματικό σύστημα έγκαιρης προειδοποίησης πρέπει να στηρίζεται σε τέσσερις βασικές λειτουργίες:

1. την ανάλυση κινδύνου
2. την παρακολούθηση και προειδοποίηση
3. την διάδοση και επικοινωνία
4. την ικανότητα αντίδρασης (UNISDR, <http://www.unisdr.org/2006/ppew/whats-ew/basics-ew.htm>)

Λίγο πιο αναλυτικά, η πρώτη βασική λειτουργία ενός ολοκληρωμένου συστήματος έγκαιρης προειδοποίησης, δηλαδή η ανάλυση κινδύνου, περιλαμβάνει τη συστηματική συλλογή δεδομένων και την αξιολόγηση τους σε σχέση με προκαθορισμένους κινδύνους και τρωτά σημεία. Η δεύτερη, δηλαδή η παρακολούθηση και προειδοποίηση, περιλαμβάνει την μελέτη των παραγόντων που υποδηλώνουν αν μια καταστροφή είναι επικείμενη, καθώς και τις μεθόδους που χρησιμοποιούνται για την ανίχνευση αυτών των παραγόντων.

Οι δυο πρώτες βασικές λειτουργίες ενός συστήματος έγκαιρης προειδοποίησης είναι και εκείνες που σχετίζονται με την χρήση data mining. Είναι αντιληπτό ότι, όσο πιο γρήγορα πραγματοποιείται κάθε λειτουργία τόσο πιο αποτελεσματικό ένα σύστημα έγκαιρης προειδοποίησης είναι. Αυτό αντανακλά την πραγματική σημασία της χρήσης data mining στα συστήματα έγκαιρης προειδοποίησης.

Η παρούσα εργασία θα βασιστεί σε ένα στατιστικό μοντέλο των Tsakiri et al. (2014), για την πρόβλεψη και την εξήγηση των χρονοσειρών επιφανειακής απορροής νερού, με την χρήση data mining χρησιμοποιώντας ένα παράδειγμα από την Νέα Υόρκη. Σκοπός της εργασίας αυτής είναι να επιτύχει τον προηγούμενο στόχο αυτού του μοντέλου, δηλαδή την ευρύτερη εφαρμογή του σε παρόμοιες λεκάνες απορροής και ως εκ τούτου την εφαρμογή του σε συστήματα έγκαιρης προειδοποίησης.

1.2. Θεωρητικό μέρος

Οι φυσικοί κίνδυνοι έχουν τεράστιο αντίκτυπο στον άνθρωπο. Απώλεια ζωής, απώλεια ιδιοκτησίας και φυσικά τεράστιες οικονομικές επιπτώσεις. Σεισμοί, τυφώνες, πλημμύρες, κύματα καύσωνα και κρύου, εκρήξεις ηφαιστείων και φωτιές είναι μερικές από τις πιο συχνές φυσικές καταστροφές. Όπως είναι αντιληπτό, η πρόβλεψη τέτοιων κινδύνων αποτελεί πρόκληση καθώς η εκδήλωση τους είναι 'απρόβλεπτη'. Η πίεση που ασκεί το ανθρώπινο είδος στον πλανήτη τις

τελευταίες δεκαετίες, σε συνδυασμό με την φύση εκδήλωσης των κινδύνων αυτών, έχει αυξήσει την αναγκαιότητα για πρόγνωση.

Οι πλημμύρες αποτελούν ένα φυσικό κίνδυνο που έχει τεράστιες επιπτώσεις και έκταση. Σύμφωνα με την EM-DAT η πλημμύρα του 2008 στην Αμερική κατέχει την πρώτη θέση ως προς το μέγεθος, αφού επηρέασε πάνω από 11.000.000 ανθρώπους. Οι πλημμύρες λαμβάνουν χώρα συνήθως κατά τη διάρκεια περιόδων υπερβολικής βροχόπτωσης ή τήξης πάγου κατά τη χειμερινή περίοδο και συνήθως συνοδεύονται από μια αύξηση στην παροχή του ποταμού. Επίσης, μπορεί να αποτελούν υποπροϊόν άλλης καταστροφής. Χαρακτηριστικό παράδειγμα αποτελούν οι τεράστιες καταιγίδες που μαστίζουν την Αμερική επιφέροντας τυφώνες και πλημμύρες αποτελώντας, σύμφωνα με την EM-DAT, τις πιο σημαντικές καταστροφές στην ήπειρο τα τελευταία πενήντα χρόνια.

Η κλιματική αλλαγή, που πιστεύεται από πολλούς ότι αποτελεί προϊόν της πίεσης που ασκεί το ανθρώπινο είδος στον πλανήτη, έχει άμεση σχέση με τις πλημμύρες. Έχει αποδειχθεί ότι υπάρχει σχέση μεταξύ της αύξησης των πλημμυρών του ποταμού και την κλιματική αλλαγή (Bronstert A., 1995, Hartmann H., 2013, Kotlarski et al., 2012). Εκτός από αυτή την αυξημένη δραστηριότητα των πλημμυρών, έχει επίσης αποδειχθεί ότι η παγκόσμια θερμοκρασία έχει αυξηθεί τόσο στην ατμόσφαιρα και την επιφάνεια της Γής, όσο και στην τροπόσφαιρα (Brohan et al., 2006, Smith et al., 2005), γεγονός που συμπίπτει με αυξανόμενες ποσότητες υδρατμών στην ατμόσφαιρα (Trenberth et al., 2005). Κατά συνέπεια, λόγω των αυξημένων επιπέδων υγρασίας οι καταιγίδες παράγουν εντονότερες βροχοπτώσεις, με αποτέλεσμα να αυξάνεται και ο κίνδυνος πλημμυρών (Trenberth K., 2011). Οι αυξημένες θερμοκρασίες κατά τη χειμερινή περίοδο είναι γνωστό πως διαλύουν την παγωμένη επιφάνεια των ποταμών και ως εκ τούτου έχουμε τήξη αυτού του πάγου τα λεγόμενα ice jams." Τα ice jams είναι μια εντοπισμένη συσσώρευση πάγου που μπορεί να δημιουργήσει μπλοκαρίσματα της ροής του ποταμού και να αυξήσει την πιθανότητα πλημμύρας στα ανάντη (Beltaos S., 1997). Κατά την τελευταία δεκαετία, παρατηρείται σημαντική αύξηση στην εκροή νερού των ποταμών (Kern A., 2008, Marsellos et al., 2010). Αυτή η αύξηση μπορεί να εξηγηθεί από την αύξηση της θερμοκρασίας κατά την χειμερινή περίοδο που επιφέρει λιώσιμο των πάγων (Kern A., 2008). Η απαλλαγή αυτή του νερού εξαρτάται από την περιοχή της λεκάνης απορροής και των παραποτάμων που εκβάλλουν στο ρεύμα.

Για παράδειγμα, το Schoharie Creek, μια υδρολογική μονάδα με μια λεκάνη απορροής περίπου 886 Mi² (2300 km²) στη Νέα Υόρκη των ΗΠΑ, ρέει 93 μίλια (150 χιλιόμετρα) βόρεια από τους πρόποδες του βουνού Indian Head, στα βουνά Catskill, μέσω της κοιλάδας Schoharie, στον ποταμό Mohawk. Είναι ένας από τους δύο κύριους παραπόταμους του ποταμού Mohawk, με τον άλλο να είναι δυτικά του Καναδά. Οκτώ περαιτέρω παραπόταμοι εκβάλλουν στον Schoharie Creek.

Είναι ζωτικής σημασίας να μοντελοποιήσουμε τις πλημμύρες, προκειμένου να βοηθήσουμε τις τοπικές κυβερνήσεις και τις κοινότητες ώστε να λάβουν προληπτικά μέτρα για την ελαχιστοποίηση ή την πρόληψη των ζημιών στην ιδιοκτησίες. Όπως είναι αντιληπτό, η μοντελοποίηση αυτή μπορεί να εφαρμοστεί σε παγκόσμιο επίπεδο, αρκεί να υπάρχουν τα αντίστοιχα δεδομένα προς επεξεργασία.

Αρκετές μελέτες σχετικά με τις πλημμύρες (Bronstert A., 1995, Hartmann H., 2013, Kotlarski et al., 2012, Johnston S., 2010) δίνουν μια μερική εξήγηση της απορροής του νερού σε σχέση με τις κλιματολογικές μεταβλητές. Ο κύριος λόγος είναι η παρουσία αβεβαιοτήτων σε όλες τις μεταβλητές των χρονολογικών σειρών (π.χ. missing values). Ειδικότερα, έχει δειχθεί ότι οι αβεβαιότητες σε βασικές μεταβλητές ή από τις επιπτώσεις των κλιμάκων που σχετίζονται με διαφορετικές συχνότητες των μεταβλητών μπορεί να αλλάξουν τις συνέπειες της ανάλυσης μεταξύ των μεταβλητών (Zurbenko I., 1999, Tsakiri et al., 2008-09). Επιπλέον, έχει δειχθεί ότι ο διαχωρισμός των χρονοσειρών σε διαφορετικά συστατικά είναι απαραίτητος για την αποφυγή παρεμβολών από διαφορετικές δομές συνδιακύμανσης που υπάρχουν μεταξύ των συνιστωσών των χρονοσειρών. Η απουσία διαχωρισμού των κλιμάκων μπορεί να οδηγήσει σε λανθασμένες εκτιμώμενες παραμέτρους του γραμμικού μοντέλου ή άλλων πολυπαραγοντικών μοντέλων (κύρια συστατικά - μέρη, κανονικά ζεύγη). Για το λόγο αυτό, η αποσύνθεση χρονοσειρών είναι απαραίτητη πριν από την πραγματοποίηση της ανάλυσης αφού η απουσία διαχωρισμού των κλιμάκων μπορεί να οδηγήσει σε παρερμηνείες (Tsakiri K., 2011). Σε προηγούμενες μελέτες σχετικά με την απορροή του νερού, η χρονοσειρά της μεταβλητής απορροής του νερού σπάνια έχει αναλυθεί σε διαφορετικά συστατικά (Bronstert A., 1995, Hartmann H., 2013, Kotlarski et al., 2012, Johnston S., 2010). Συγκεκριμένα, οι Kotlarski et al. ερεύνησαν τις πλημμύρες στον Έλβα ποταμό τον Αύγουστο του 2002, χρησιμοποιώντας τα κατακρημνίσματα ως μεταβλητή (Kotlarski et al., 2012). Στο μοντέλο τους, οι συγγραφείς καθορίζουν ένα πρότυπο συσχέτισης με τιμή 0,75 αποδίδοντας μια τιμή R^2 ίση με 0,56. Μια εν μέρει ερμηνεία της απορροής του νερού από τις κλιματολογικές μεταβλητές μπορεί να οφείλεται σε μια μικτή παρέμβαση στην κλίμακα των χρονοσειρών. Η εξήγηση της απορροής του νερού μπορεί να βελτιωθεί με τη χρήση της αποσύνθεσης των χρονοσειρών. Ο κύριος σκοπός του μοντέλου των Tsakiri et al. (2014), είναι η παρουσίαση μιας νέας μεθοδολογίας για την εξήγηση και την βελτίωση της ικανότητας πρόβλεψης των χρονοσειρών της απορροής των ποτάμιων υδάτων με τη χρήση κλιματικών και υδρολογικών μεταβλητών. Το μοντέλο ενσωματώνει διαφορετική διαχείριση στην συχνότητα των κλιμάκων. Οι εν λόγω κλίμακες έχουν καθοριστεί με βάση την φασματική και συνασματική ανάλυση και παρέχουν μια φυσική εξήγηση για την υδρολογία της περιοχής. Ειδικότερα, τα δεδομένα χωρίζονται σε εποχικά (seasonal), μακροπρόθεσμα (long term) και βραχυπρόθεσμα (short term) συστατικά με τη χρήση φίλτρου [το αρχικό φίλτρο Kolmogorov-Zurbenko (εφεξής KZ) θα ενσωματωθεί σε έναν κόμβο (node) στο λογισμικό data mining] (Zurbenko I., 1986). Το φίλτρο αυτό παρέχει αποτελεσματικό διαχωρισμό των συχνοτήτων και η περιγραφή του θα γίνει στην επόμενη ενότητα.

Στην παρούσα εργασία χρησιμοποιούνται δύο μοντέλα για την πρόβλεψη των χρονοσειρών απορροής του νερού για να αποφευχθούν μεικτές παρεμβολές. Αυτό είναι ιδιαίτερα αναγκαίο σε υψηλά γεωγραφικά πλάτη όπου παρατηρούνται παρατεταμένοι χειμώνες και καλοκαίρια. Αυτές οι παρατεταμένες εποχές, όπως ο χειμώνας (ο οποίος έχει παραταθεί κατά ένα μήνα ή περισσότερο) σε αυτό το γεωγραφικό πλάτος, μπορεί να παρέχουν παρεμβολές λόγω της διαφορετικής συσχέτισης στη δομή τους και πρέπει να διαχωριστούν, όπως αποδεικνύεται σε πρότυπες στατιστικές προσεγγίσεις. Ένα μοντέλο είναι σχεδιασμένο για τη χειμερινή περίοδο και ένα άλλο μοντέλο για τη θερινή περίοδο. Αυτό οφείλεται στο γεγονός ότι οι πλημμύρες στην Schoharie Creek που αφορούν το μοντέλο, μπορεί να προκαλούνται από ισχυρές βροχοπτώσεις κατά τη διάρκεια του καλοκαιριού ή ταχύ λιώσιμο του χιονιού (ice jam) κατά την παρατεταμένη διάρκεια της χειμερινής περιόδου (Garver J., 2009). Και τα δύο μοντέλα, διαχωρίζουν τις χρονοσειρές της απορροής του νερού, του βάθους της στάθμης του νερού κάτω από την επιφάνεια της γης (που αναφέρεται ως στάθμη των υπόγειων υδάτων, groundwater), καθώς και τις κλιματικές μεταβλητές. Στη συνέχεια σχεδιάζουμε τα διαφορετικά μοντέλα

πολλαπλών μεταβλητών, για κάθε κλίμακα, των χρονοσειρών απορροής στη ροής εργασίας μας (workflow).

Οι διαφορετικές συχνότητες είναι πάντα ασυσχέτιστες και ως αποτέλεσμα τα μοντέλα πολλαπλών μεταβλητών σχεδιάζονται για κάθε κλίμακα των χρονοσειρών της απορροής του νερού, ξεχωριστά. Η μακροπρόθεσμη συνιστώσα και η εποχική συνιστώσα είναι οι ίδιες, τόσο για το χειμερινό όσο και για το θερινό μοντέλο. Ωστόσο, η βραχυπρόθεσμη συνιστώσα έχει σχεδιαστεί με τη χρήση δύο διαφορετικών μοντέλων για την πρόβλεψη της χρονοσειράς της απορροής του νερού. Τα φυσικά φαινόμενα (όπως το λιώσιμο του πάγου ή οι πλημμύρες) είναι αποτέλεσμα όλων των συνιστωσών (εποχικές, μακροπρόθεσμες και βραχυπρόθεσμες) και όχι απλώς βραχυπρόθεσμων διακυμάνσεων. Η χρήση του φίλτρου (KZ), έχει αποδεχτεί ότι βελτιώνει την εξήγηση, όσον αφορά τις χρονοσειρές απορροής του νερού, έως 81 % (Tsakiri et al., 2014).

Κεφάλαιο 2.

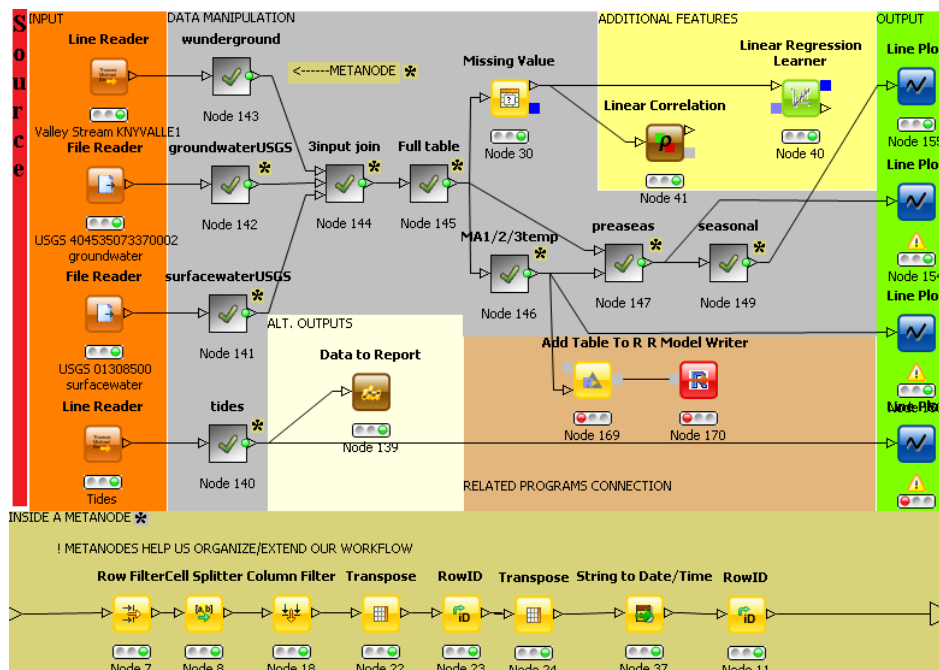
Μεθοδολογία

2.1. Εισαγωγή

Η παρούσα ροή εργασίας (workflow) ακολουθεί μια τυπική ροή σήματος με την δομή «είσοδος – διαμόρφωση – έξοδος». Η δομή αυτή στην συγκεκριμένη περίπτωση, αν και παραμένει ένα ψηφιακό σήμα, μεταφράζεται σε «δεδομένα – διαχείριση δεδομένων – εξαγωγή συμπερασμάτων». Λίγο πιο συγκεκριμένα, σε αυτήν την γενική δομή μπορεί να προστεθούν στην έξοδο – εξαγωγή συμπερασμάτων και πρόσθετα χαρακτηριστικά, όπως η γραμμική συνοχή (linear correlation, R^2), η γραμμική παλινδρόμηση (linear regression) και οι πρόσθετοι έξοδοι. Ωστόσο, όπως φαίνεται και στο Σχήμα 1, εμφανίζονται ξεχωριστά καθώς αποτελούν προϋπόθεση ώστε να έχουμε στην έξοδο μας σωστά αποτελέσματα. Επίσης, ένα ακόμα κομμάτι που ανήκει στην διαμόρφωση – διαχείριση δεδομένων είναι τα πρόσθετα λογισμικά (στην περίπτωση μας η R i386 3.2.3). Τα πρόσθετα λογισμικά διακρίνονται σαν κατηγορία για να αναδειχθούν οι πολύπλευρες δυνατότητες του παρόντος προγράμματος data mining.

Το πρόγραμμα που χρησιμοποιούμε είναι το Knime. Πρόκειται για ένα λογισμικό data mining με πολλές δυνατότητες. Η επιλογή του παρόντος λογισμικού έγινε για αρκετούς λόγους, κάποιοι από αυτούς ενδεικτικά είναι:

- Πρόκειται για ένα λογισμικό που έχει την δυνατότητα ανάκτησης δεδομένων σε πραγματικό χρόνο [real time data mining, με τον όρο εννοούμε ότι παίρνουμε δεδομένα από URL και όχι μόνο μπορούμε να επιλέγουμε ημερομηνίες από το σύνδεσμο (link), αλλά και να το προγραμματίζουμε να κάνει βρόγχο (loop), ώστε να ανακτώνται όσα καινούργια δεδομένα είναι διαθέσιμα από τον κάθε πάροχο)
- Έχει την επιλογή προσωρινής αποθήκευσης καθιστώντας το ιδανικό για μεγάλα δεδομένα
- Έχει μια πολύ καλή διαδικτυακή κοινότητα (forum για την επίλυση προβλημάτων και μελλοντικών προτάσεων για αναβάθμιση)
- Έχει μεγάλη δυνατότητα συνεργασίας με πρόσθετα λογισμικά (R, - SQL, MS ACCESS, GIS, MAPS, κ.α.)
- Διανέμεται δωρεάν



Σχήμα 1: Η ροή εργασίας. Διακρίνονται τα επιμέρους μέρη με ονόματα και χρώματα (είσοδος – input, διαχείριση δεδομένων – data manipulation, έξοδος και εναλλακτικές έξοδοι – output and alt. outputs, πρόσθετα χαρακτηριστικά - additional features, συνεργαζόμενα προγράμματα – related programs connection). Στο κάτω μέρος του σχήματος, διακρίνεται και η εσωτερική δομή ενός μετακομβίου (metanode) (Τροποποιημένο από Καβαλιέρος Α., 2016, Knime).

Στα επόμενα κεφάλαια θα γίνει αναφορά της μεθοδολογίας του μοντέλου των Tsakiri et al., 2014 και αντιστοιχία της μεθοδολογίας αυτής στο Knime. Σημειώνεται ότι πρόκειται για την ίδια μεθοδολογία που πλέον όμως έχει αυτοματοποιηθεί. Η ροή εργασίας αυτή μπορεί να χρησιμοποιηθεί άμεσα και σε άλλες περιοχές χωρίς πρόσθετη παρέμβαση. Επιπροσθέτως, στο τελευταίο κεφάλαιο θα αναφερθούμε και σε μελλοντικά σενάρια επέκτασης, όπως η επιλογή της περιοχής μελέτης (σε πολύγωνο) απευθείας στον χάρτη.

2.2. Δεδομένα

Στο μοντέλο των Tsakiri et al. (2014), τα δεδομένα που χρησιμοποιήθηκαν ήταν οι καθημερινές χρονοσειρές απορροής του νερού (water discharge), το επίπεδο των υπόγειων υδάτων (groundwater level), οι παλίρροιες (tides), και οι κλιματικές μεταβλητές (climatic variables). Αντίστοιχα ήταν και για την παρούσα εργασία. Όλα τα δεδομένα είναι ημερήσια (daily data) και όλοι οι σταθμοί παρακολούθησης βρίσκονταν σε μια περιοχή κοντά στο Schoharie Watershed, όπου είναι η περιοχή μελέτης του αρχικού μοντέλου (άρα και η ίδια για την παρούσα εργασία για λόγους επαλήθευσης). Επίσης, έναν ακόμα παράγοντα επιλογής των σταθμών, πέραν της αποστάσεως, αποτελεί και η πληρότητα των δεδομένων αυτών (λιγοστά επίπεδα χαμένων δεδομένων (missing values)). Στο αρχικό μοντέλο τα δεδομένα αυτά είχαν ληφθεί μέσα από την ηλεκτρονική δημόσια βάση δεδομένων της Εθνικής Υπηρεσίας Ωκεανών και Ατμόσφαιρας (NOAA, <http://www.noaa.gov/>), όσον αφορά τα κλιματολογικά δεδομένα και το Γεωλογικό Ινστιτούτο των ΗΠΑ (USGS, <https://www.usgs.gov/>), όσον αφορά την απορροή του

νερού και το βάθος της στάθμης του νερού κάτω από την επιφάνεια) ωστόσο, στην παρούσα εργασία επιλέχθηκε η λήψη των καθημερινών κλιματολογικών δεδομένων (εκτός από τις παλίρροιες) από το Weather Underground (<https://www.wunderground.com/>), καθώς παρείχε περισσότερες μεταβλητές και λιγότερα χαμένα δεδομένα (missing values). Σημειώνεται ότι, η παρούσα ροή εργασίας λειτούργησε και με τα ίδια δεδομένα από τις ίδιες πηγές και τους ίδιους σταθμούς με το αρχικό μοντέλο για επαλήθευση.

Λίγο πιο αναλυτικά, οι μεταβλητές που χρησιμοποιήθηκαν (όπως φαίνεται και στο Σχήμα 1, δεξιά από την είσοδο (input) με το πορτοκαλί χρώμα ακολουθούν οι αντίστοιχοι κόμβοι (nodes) wunderground (a), water discharge USGS (b), groundwater USGS (c) και tides (d)) είναι:

- a) Οι κλιματικές μεταβλητές (από το NOAA στο αρχικό και από το Weather Underground στο παρόν μοντέλο)
- b) Οι καθημερινές χρονοσειρές απορροής του νερού, water discharge (σε κυβικά πόδια το δευτερόλεπτο, ft^3/sec), από το USGS (τόσο στο αρχικό, όσο και στο παρόν μοντέλο)
- c) Η στάθμη των υπόγειων υδάτων, groundwater level (σε πόδια, ft., elevation above NGVD 1929) από το USGS (τόσο στο αρχικό, όσο και στο παρόν μοντέλο)
- d) Οι παλίρροιες, tides (σε πόδια, ft.) από το NOAA (τόσο στο αρχικό, όσο και στο παρόν μοντέλο)

Αναλυτικότερα, οι κύριες κλιματικές μεταβλητές που χρησιμοποιήθηκαν στο παρόν, καθώς και η αντιστοιχία αυτών με το αρχικό μοντέλο είναι:

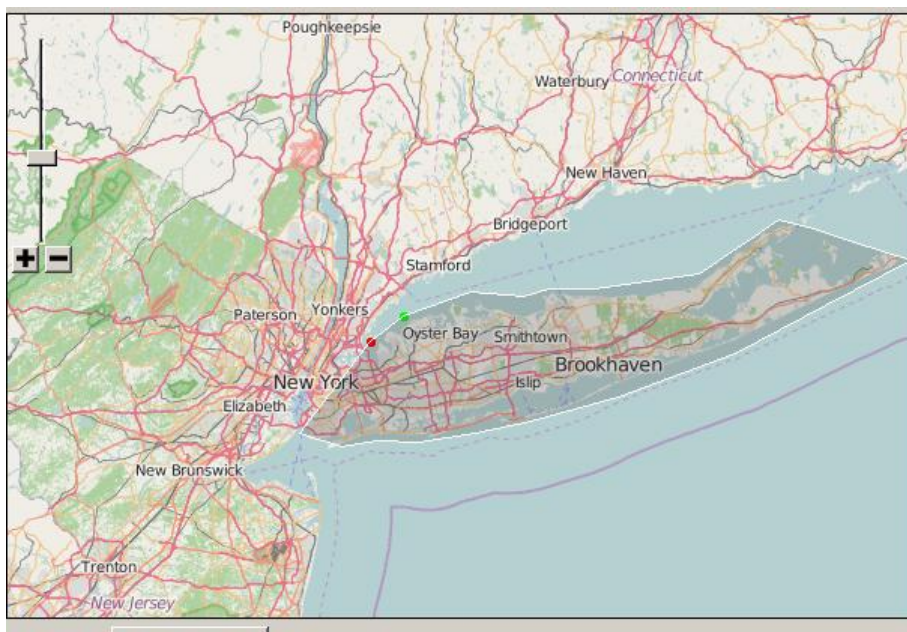
- Οι καθημερινές χρονοσειρές της υψηλής, χαμηλής και μέσης θερμοκρασίας του αέρα (High, Avg, Low temperature) σε βαθμούς Κελσίου ($^{\circ}\text{C}$), από το Weather Underground, (εν αντιθέσει με τις καθημερινές χρονοσειρές στο αρχικό μοντέλο, που προέρχονταν από το NOAA και αφορούσαν την θερμοκρασία του αέρα σε Φαρενάιτ, $^{\circ}\text{F}$)
- Η μέγιστη και μέση ταχύτητα του ανέμου (Wind speed Max, Avg) σε χιλιόμετρα ανά ώρα (Km/hr), από το Weather Underground (εν αντιθέσει με το αρχικό μοντέλο, που προέρχονταν από το NOAA και αφορούσαν την ταχύτητα του ανέμου σε μετρά ανά δευτερόλεπτο (m / sec))
- Η συνολική βροχόπτωση και η χιονόπτωση σε χιλιοστά ανά ώρα (mm/hr), (τόσο στο αρχικό, όσο και στο παρόν μοντέλο αλλά από άλλη υπηρεσία)

Σε αυτές τις μεταβλητές μπορούν να προστεθούν και άλλες όπως:

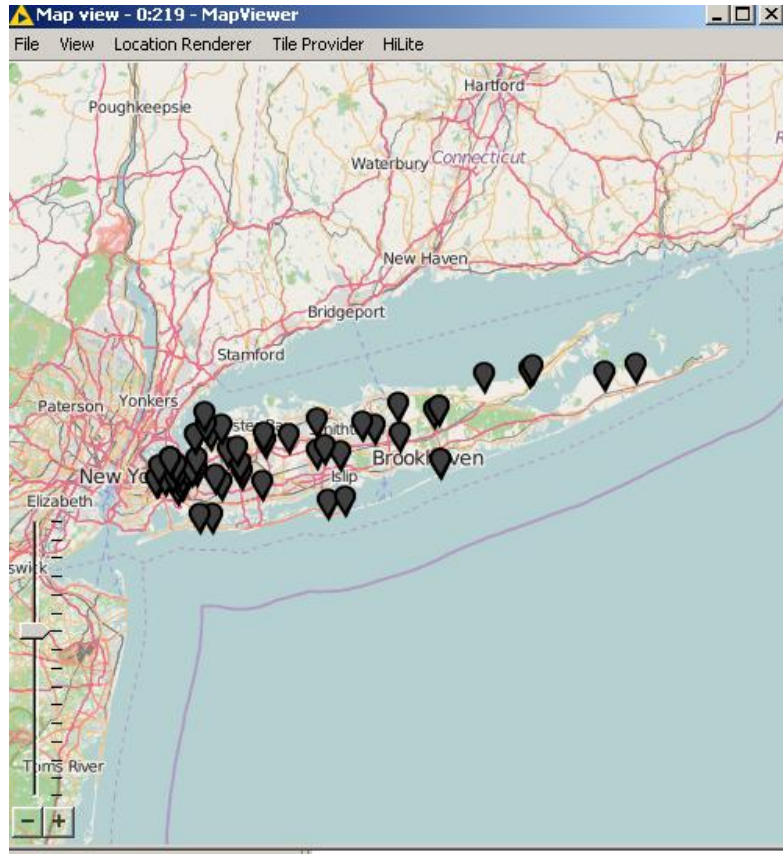
- Το υψηλό, χαμηλό και μέσο σημείο δρόσου (High, Avg, Low dew point) σε βαθμούς Κελσίου ($^{\circ}\text{C}$)
- Η υψηλή, χαμηλή και μέση απόλυτη υγρασία (High, Avg, Low humidity) σε γραμμάρια υδρατμών ανά κυβικό μέτρο αέρα

- Η υψηλή και χαμηλή πίεση (max, min pressure) σε Εκτοπασκάλ (hPa). Οι οποίες βοηθούν στην πιο ολοκληρωμένη κατανόηση και καλύτερη διεξαγωγή συμπερασμάτων.

Όλα τα δεδομένα των μεταβλητών έχουν αναλυθεί για την ίδια περιοχή και χρονική περίοδο. Προέρχονται από κοντινούς σταθμούς στο Schoharie Creek και αφορούν την περίοδο 1^η Ιανουαρίου του 2005 με 31^η Δεκεμβρίου του 2012 ώστε να περιέχεται και δίσεκτο έτος (2012). Σημειώνεται ότι η παρούσα ροή εργασίας έχει δοκιμαστεί, από συμβατικό ηλεκτρονικό υπολογιστή, για πολύ μεγαλύτερη περίοδο με καθημερινή αναβάθμιση δεδομένων μέσω βρόγχου (loop). Επίσης, έχει δοκιμαστεί και για άλλες περιοχές (πέραν αυτής του αρχικού μοντέλου) με ευκολία, είτε αλλάζοντας κομμάτι της ηλεκτρονικής διεύθυνσης (URL) στην είσοδο (input) με αυτό άλλων σταθμών, είτε επιλέγοντας επιμέρους περιοχή από μια ευρύτερη περιοχή μέσω φίλτρου πολυγώνων (π.χ. έχοντας εισάγει σαν ευρύτερη περιοχή την Νέα Υόρκη να επιλέξουμε το Λονγκ Άιλαντ), όπως φαίνεται στην Εικόνα 1 και στην συνέχεια έχουμε τους σταθμούς (Εικόνα 2), που θα δώσουμε στην βασική ροή εργασίας μας. Οι περιπτώσεις αυτές θα αναπτυχθούν περεταίρω στο τελευταίο κεφάλαιο.



Εικόνα 1: Το πολυγωνικό φίλτρο. Διακρίνεται το επιλεγμένο πολύγωνο το οποίο στον επόμενο κόμβο (node) θα μας εμφανίσει του σταθμούς στην επιλεγμένη επιμέρους περιοχή (Εικόνα 2) (Ktime, geo coordinates filter).



Εικόνα 2: Η προβολή των διαθέσιμων σταθμών στην επιμέρους επιλεγμένη περιοχή (Knime, map viewer).

2.3. Αποσύνθεση των χρονοσειρών

Έχει δειχθεί σε αρκετές μελέτες, ότι ο διαχωρισμός των χρονοσειρών σε διαφορετικά συστατικά (components) είναι απαραίτητος προκειμένου να αποφευχθούν τυχόν συμβολές από διαφορετικές συνδιακυμάνσεις στις δομές μεταξύ των συστατικών μιας χρονοσειράς (Zurbenko I., 1999, Tsakiri et al., 2008-09-11-13). Έτσι και στην παρούσα ροή εργασίας ακολουθείται η ίδια δομή, όπως φαίνεται και στο Σχήμα 1. Η χρονική σειρά μιας μεταβλητής μπορεί να εκφράζεται με τον τύπο:

$$X(t) = L(t) + Se(t) + Sh(t), \quad (1)$$

όπου $X(t)$ αντιπροσωπεύει την αρχική χρονοσειρά μιας μεταβλητής,

$L(t)$ είναι η μακροπρόθεσμη συνιστώσα (long term),

$Se(t)$ είναι η εποχική συνιστώσα (seasonal),

και $Sh(t)$ είναι η βραχυπρόθεσμη συνιστώσα (long term).

Λίγο πιο αναλυτικά, η μακροπρόθεσμη συνιστώσα μας περιγράφει τις διακυμάνσεις μιας χρονοσειράς σε σχέση με ένα συγκεκριμένο όριο (threshold) το οποίο έχουμε καθορίσει, η εποχική συνιστώσα περιγράφει τις διακυμάνσεις από έτος σε έτος και η βραχυπρόθεσμη συνιστώσα περιγράφει τις βραχυπρόθεσμες διακυμάνσεις. Η τάση της μακροπρόθεσμης συνιστώσας περιλαμβάνει πληροφορίες σχετικά ,τόσο για τη τάση, όσο και για την κυκλική συνιστώσα (OECD, 2005). Η κυκλική συνιστώσα μιας χρονοσειράς είναι αυτή που μας περιγράφει τις διακυμάνσεις γύρω από την τάση. Η κυκλική συνιστώσα λοιπόν (ή περιοδικότητα) θεωρείται η συνιστώσα που μας δείχνει τις διακυμάνσεις μιας χρονοσειράς στον χρόνο, έχοντας πρώτα ορίσει ένα συγκεκριμένο κατώτατο όριο, για παράδειγμα, για ενάμιση χρόνο, μικρότερη τιμή από αυτή αποδίδεται στην τάση.

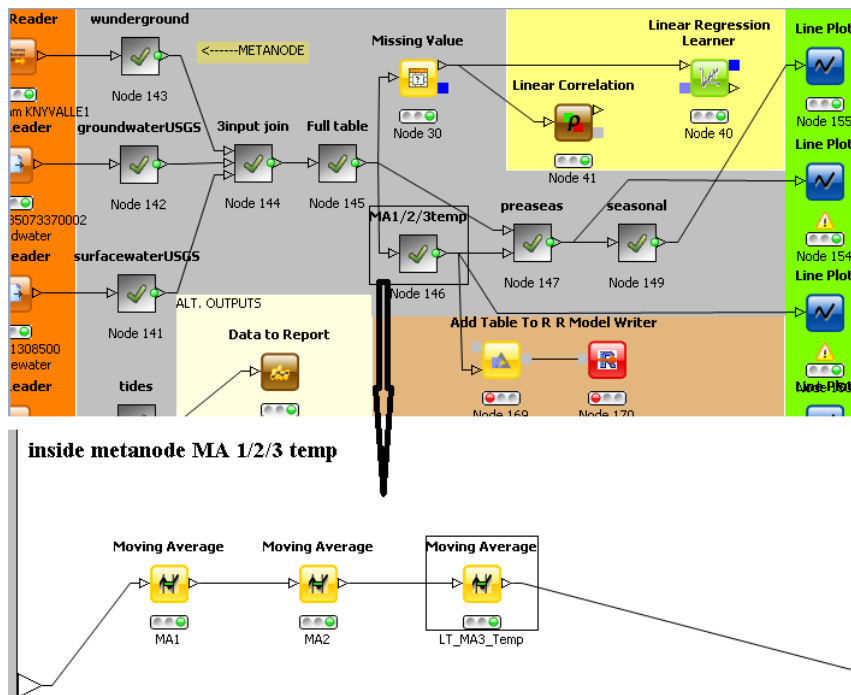
Εν συνεχεία, το αρχικό μοντέλο χρησιμοποιεί το KZ φίλτρο για να ξεχωρίσει τις μακροπρόθεσμες από τις βραχυπρόθεσμες διακυμάνσεις σε μια χρονοσειρά (Zurbenko I., 1986). Στην παρούσα ροή εργασίας χρησιμοποιείται ένας ή περισσότεροι κόμβοι που αντικαθιστούν το φίλτρο. Η επιλογή του φίλτρου αυτού έγινε καθώς μας παρέχει το μικρότερο επίπεδο παρεμβολών μεταξύ των κλιμάκων (εποχική, μακροπρόθεσμη και βραχυπρόθεσμη συνιστώσα) μιας χρονοσειράς σε συνδυασμό με την ευκολία χρήσης του. Επίσης μας επιτρέπει την πιο φυσική ερμηνεία των κλιμάκων (Yang W., 2010) και παρέχει αποτελεσματικό διαχωρισμό συχνοτήτων για εφαρμογή απευθείας στα σύνολα δεδομένων (datasets) τα οποία μπορεί να έχουν ελλιπή δεδομένα (missing values) (Eskridge et al., 1997, Rao et al., 1997, Yang W., 2010). Ο τρόπος που φιλτράρει το KZ είναι επίσης γνωστό ότι παρέχει τα καλύτερα και πιο ακριβή αποτελέσματα στο βέλτιστο μέσο τετραγωνικό σφάλμα (Zurbenko I., 1986, Yang W., 2010, Close B., 2013). Το φίλτρο KZ έχει εφαρμοστεί αποτελεσματικά για την εξήγηση και πρόβλεψη του προβλήματος του όζοντος και για την επεξήγηση των χρονοσειρών της χρήσης των υδάτων στη Gainesville, Φλόριντα (Tsakiri et al., 2010-11-13, Rao et al., 1994-97, Hogrefe et al., 2000). Συγκεκριμένα, το φίλτρο KZ είναι ένα φίλτρο χαμηλής διέλευσης (low pass), που ορίζεται από p επαναλήψεις (iterations) ενός απλού κινούμενου μέσου όρου (moving average) m σημείων. Ο κινητός μέσος όρος του KZ εκφράζεται με τον τύπο:

$$Y_t = \left(\frac{1}{m}\right) \sum_{j=-k}^k, \quad (2)$$

όπου $m = 2k + 1$.

Η έξοδος από την πρώτη επανάληψη γίνεται η είσοδος για τη δεύτερη επανάληψη και ούτω καθεξής. Η χρονοσειρά που παράγεται από τις p επαναλήψεις του φίλτρου που περιγράφεται στον τύπο (2) συμβολίζεται με την σχέση:

$$Y_t = KZ_{m,p}(X_t). \quad (3)$$



Σχήμα 2: Η αποσύνθεση των χρονοσειρών με την χρήση του κινούμενου μέσου. Διακρίνεται ότι μέσα στον μετα-κόμβο 146 φαίνονται οι 3 (όπου $p=3$ στην σχέση 3) συνεχόμενες επαναλήψεις αντίστοιχες με του KZ φίλτρου (Τροποποιημένο από Καβαλιέρος Α., 2016, Knime, moving average).

Η παράμετρος m στην σχέση (3) έχει καθοριστεί για να παρέχει την μέγιστη εξήγηση της χρονοσειράς απορροής του νερού (water discharge) σε σχέση με τις κλιματολογικές μεταβλητές.

Το μοντέλο των Tsakiri et al. (2014), εξέτασε αρχικώς τα περιοδογράμματα όλων των χρονοσειρών. Η χρήση περιοδογραμμάτων επαλήθευσε ότι η αποσύνθεση των χρονοσειρών είναι απαραίτητη για την πρόβλεψη της χρονοσειράς της απορροής του νερού χρησιμοποιώντας τις κλιματικές μεταβλητές. Αρχικά, σχεδιαστήκαν πολλά διαφορετικά πολυπαραγοντικά μοντέλα που να αντιστοιχούν σε κάθε κλίμακα των χρονοσειρών. Ωστόσο, για την αποσύνθεση της χρονοσειράς, χρησιμοποιήθηκε το φίλτρο KZ. Τα πολυπαραγοντικά μοντέλα περιγράφουν την προβολή της χρονοσειράς της απορροής του νερού (water discharge) στο χώρο που ορίζεται από τη μεταβλητή του υπόγειου νερού (ground water) και των κλιματικών μεταβλητών (climatic variables). Η προβολή αυτή μπορεί να εκφραστεί μέσω της γραμμικής παλινδρόμησης (linear regression) μεταξύ της χρονοσειράς της επιφανειακής απορροής του νερού και των υδρολογικών και κλιματικών μεταβλητών (όπως στο Σχήμα 1, κόμβος 40, με κίτρινο φόντο, Knime, node 40, linear regression, additional features, yellow box). Ως εκ τούτου, το μοντέλο αποσυνθέτει όλες τις χρονοσειρές των μεταβλητών, χρησιμοποιώντας το φίλτρο KZ. Μετά την εφαρμογή του φίλτρου KZ, πρέπει να προβλέψουμε κάθε συστατικό των χρονοσειρών της απορροής του νερού, ξεχωριστά. Για το λόγο αυτό, χρησιμοποιήθηκαν πολυπαραγοντικά μοντέλα για την πρόβλεψη κάθε συστατικού των χρονοσειρών της απορροής του νερού. Για το πολυπαραγοντικό μοντέλο, επιλεχθήκαν διαφορετικές κλιματικές μεταβλητές για να εξηγηθούν τα συστατικά. Τέλος, εκτιμήθηκε η χρονοσειρά της απορροής του νερού χρησιμοποιώντας τις κλιματικές μεταβλητές για τη χειμερινή και θερινή περίοδο.

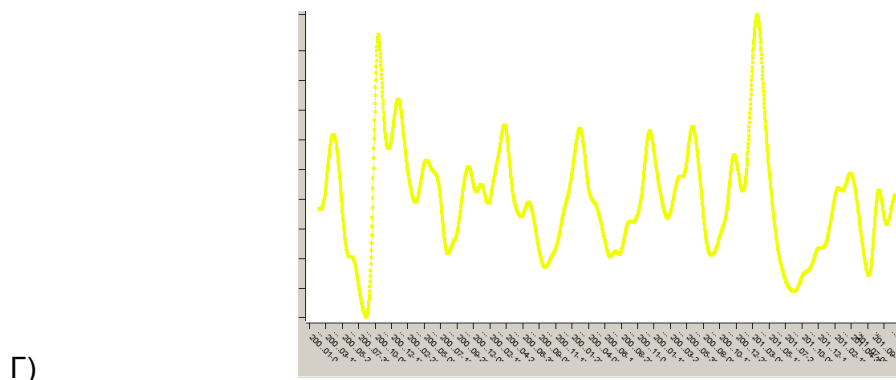
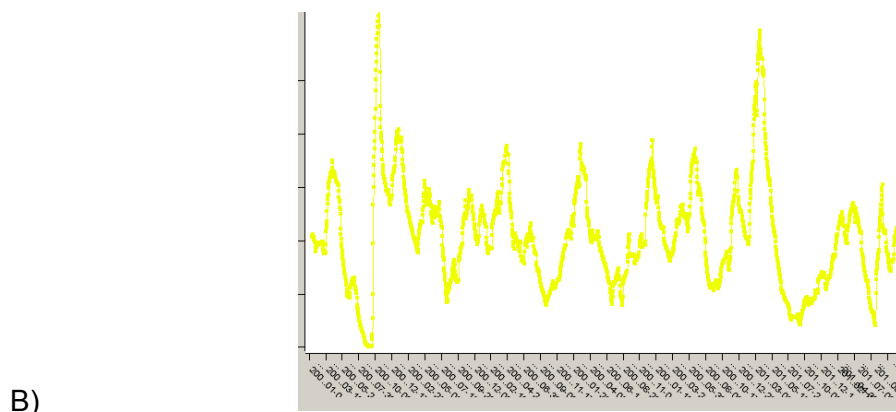
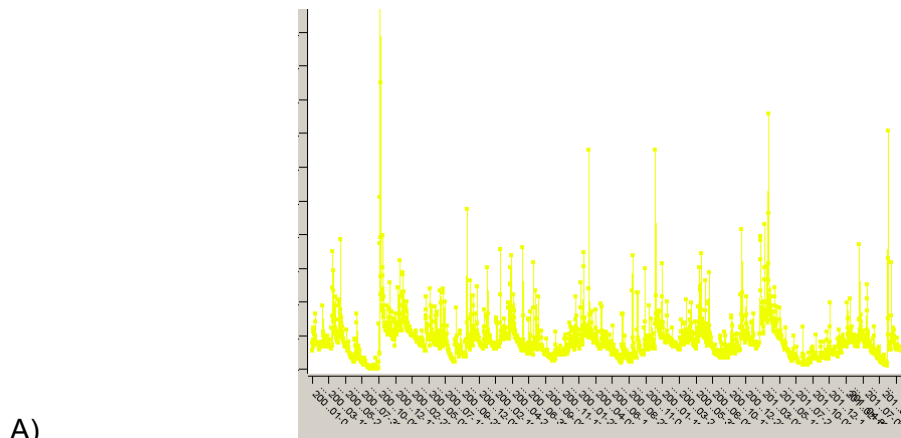
Στην παρούσα ροη εργασίας ακολουθήθηκε η ίδια μεθοδολογία με αυτή του αρχικού μοντέλου η οποία πραγματοποιήθηκε με μεγαλύτερη ακρίβεια (περισσότερες κλιματικές μεταβλητές) και πλέον αυτοματοποιημένα. Το KZ φίλτρο έδωσε την θέση του στους κόμβους κινητού μέσου (moving average nodes) όπως φαίνεται στο Σχήμα 2. Η γραμμική παλινδρόμηση έγινε με την χρήση ενός κόμβου γραμμικής παλινδρόμησης (linear regression node).

Στις ενότητες που ακολουθούν, περιγράφουμε τη μέθοδο πρόβλεψης των εποχιακών (seasonal), των μακροπρόθεσμων (short term) και βραχυπρόθεσμων (long term) συστατικών των χρονοσειρών της απορροής του νερού (water discharge) για το Schoharie Creek, στη Νέα Υόρκη ως παράδειγμα περιοχής εφαρμογής αλλά και επαλήθευσης για την ροη εργασίας μας. Όπως προαναφέρθηκε, έχουν αναλυθεί δεδομένα και για άλλες περιοχές στη Νέα Υόρκη και για μεγαλύτερα χρονικά διαστήματα.

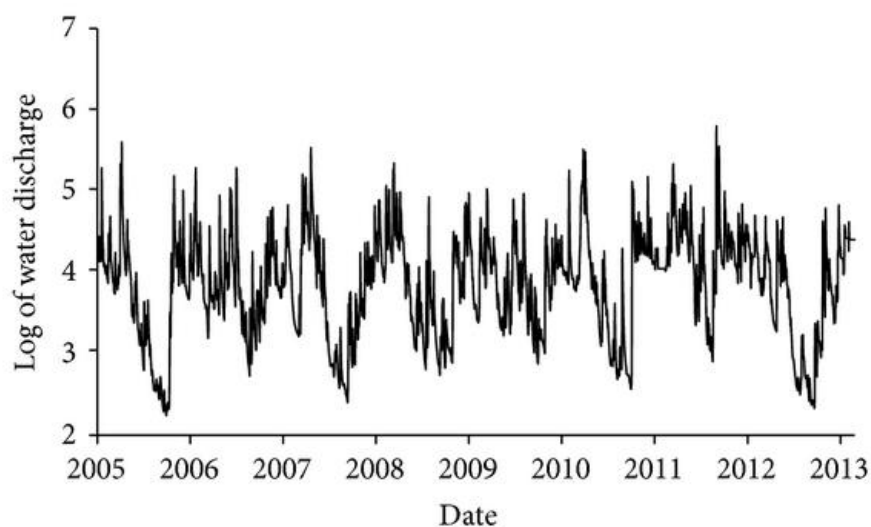
2.4. Ανάλυση ακατέργαστων δεδομένων των χρονοσειρών της απορροής του νερού

Στην παρούσα ενότητα θα περιγράψουμε και θα συγκρίνουμε την ανάλυση των ακατέργαστων δεδομένων των χρονοσειρών της απορροής του νερού, τόσο του αρχικού μοντέλου, όσο και με την χρήση του data mining. Καθώς η ροη εργασίας μας αποτελεί την ενσωμάτωση του αρχικού μοντέλου στο λογισμικό μας, όπως είναι αντιληπτό, η σύγκριση αφορά στην προσαρμοστικότητα, την ευελιξία, την αποδοτικότητα και φυσικά την ταχύτητα, όπου είναι και ζητούμενο της εργασίας. Η σύγκριση φαντάζει άδικη καθώς η λέξη κλειδί «σε πραγματικό χρόνο» (όπως απαιτεί ένα σύστημα έγκαιρης προειδοποίησης) δεν μπορεί να πραγματοποιηθεί από την αρχική απόπειρα.

Αρχικώς εφαρμόσαμε την προηγούμενη μεθοδολογία για την αποσύνθεση της χρονοσειράς για να εξηγήσουμε και να προβλέψουμε τις χρονοσειρές της απαλλαγής του νερού για το Schoharie Creek. Για την εξήγηση και πρόβλεψη της απαλλαγής του νερού (water discharge), χρησιμοποιούμε τις κλιματικές μεταβλητές (climatic variables) και το επίπεδο των υπόγειων υδάτων (groundwater level). Ο λογάριθμος της απαλλαγής νερού από την περιοχή μελέτης μετρήθηκε για το διάστημα Ιανουάριος 2005 - Φεβρουάριος 2013 για το αρχικό μοντέλο. Στην παρούσα ροη εργασίας μετρήθηκε η απαλλαγή του νερού, τόσο για το παρόν διάστημα, για λογούς επαλήθευσης, όσο και για αλλά χρονικά διαστήματα για λογούς προσαρμοστικότητας. Σημειώνεται ότι χρησιμοποιούμε τον λογάριθμο προκειμένου να σταθεροποιηθεί η διακύμανση της χρονοσειράς της απορροής του νερού. Στην Εικόνα 3 φαίνονται τα αρχικά δεδομένα και στην συνέχεια η σταδιακή ομαλοποίησή τους με τα κινούμενα μέσα (moving average). Αξίζει επίσης να αναφερθεί ότι η διαδικασία αυτή είναι πλέον αυτοματοποιημένη στην ροη εργασίας μας και μπορεί να χρησιμοποιηθεί και σε πραγματικό χρόνο (real time). Ενδεικτικά στην Εικόνα 4 φαίνεται η ομαλοποίηση από το αρχικό μοντέλο.



Εικόνα 3: Η ομαλοποίηση των ακατέργαστων δεδομένων με την χρήση κινούμενου μέσου. Ακατέργαστα δεδομένα, row data (α), αρχικό στάδιο ομαλοποίησης, moving average 1 (β) και τελικό στάδιο ομαλοποίησης, moving average 3 – Longterm (γ). Σημειώνεται ότι οι παρούσες εικόνες δεν αφορούν το ίδιο διάστημα και μέρος με την Εικόνα 4 (Ktime, moving average node).



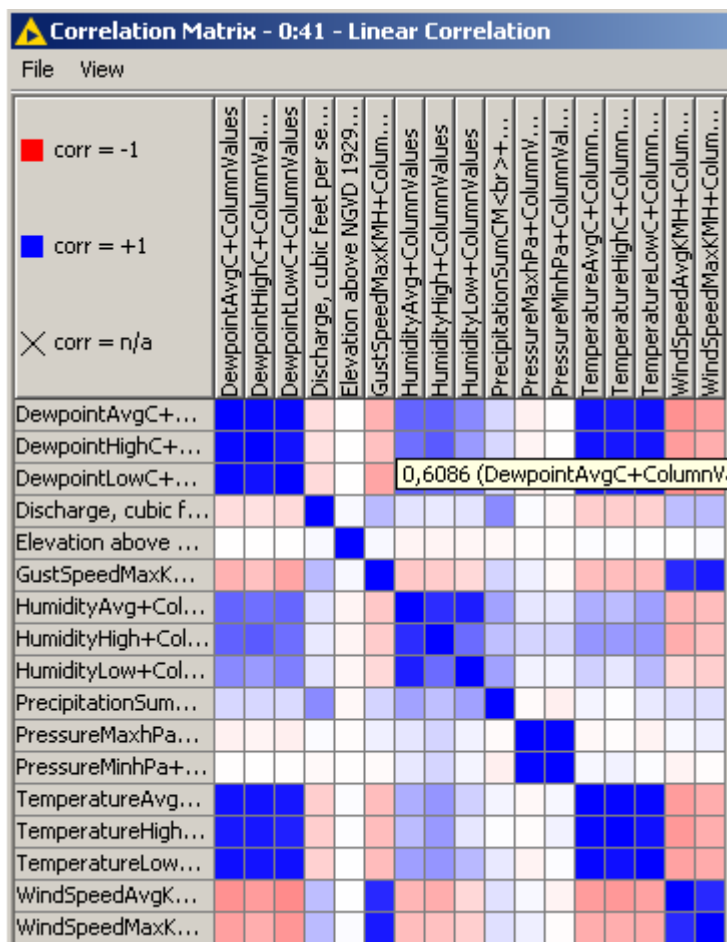
Εικόνα 4: Ο λογάριθμος της χρονοσειράς της απορροής του νερού σε κυβικά πόδια (ft³/sec) κατά την περίοδο Ιανουάριου 2005- Φεβρουάριου 2013. (Tsakiri et al., 2014)

Εν συνεχεία του αρχικού μοντέλου, παρουσιάζεται ο Πίνακας 1 όπου φαίνεται η συσχέτιση μεταξύ των ακατέργαστων δεδομένων και των μεταβλητών. Αντίστοιχα, στην Εικόνα 5 παρουσιάζεται το μοντέλο συσχέτισης των ακατέργαστων δεδομένων και των νέων μεταβλητών και στην Εικόνα 6 οι μετρήσεις αυτών. Το μοντέλο συσχέτισης του προγράμματος data mining, Ktime, που φαίνεται στην Εικόνα 6 είναι ιδιαίτερα χρήσιμο καθώς χρησιμοποιείται για μια γρήγορη συσχέτιση, παρέχοντας την επιλογή να φιλτράρουμε από πιο σημείο συσχέτισης (R^2) και πάνω θέλουμε να προβάλει (απολυτό αριθμό).

Πίνακας 1: Η συσχέτιση μεταξύ των ακατέργαστων δεδομένων και των μεταβλητών του αρχικού μοντέλου των Tsakiri et al. (2014)

Table 1: Correlation matrix of the raw data between the water discharge, the climatic variables, and the groundwater level.

Variables	Water discharge	Groundwater level	Temperature	Tide	Wind speed	Precipitation
Water discharge	1	-0.65	-0.481	0.254	0.203	0.109
Groundwater level	-0.65	1	0.178	-0.277	-0.094	0.033
Temperature	-0.481	0.178	1	0.041	-0.193	-0.007
Tide	0.254	-0.277	0.041	1	0.105	-0.008
Wind speed	0.203	-0.094	-0.193	0.105	1	0.139
Precipitation	0.109	0.033	-0.007	-0.008	0.139	1



Εικόνα 5: Το μοντέλο συσχέτισης του Knime που δείχνει τη συσχέτιση μεταξύ των ακατέργαστων δεδομένων και των μεταβλητών για το Long Island (Knime, linear correlation matrix node).

Row ID	D Dewpoi...	D Dewpoi...	D Dewpoi...	D Dischar...	D Elevati...	D GustSp...	D Humidi...	D Humidi...	D Humidi...	D Precip...	D Pressur...	D Pressur...	D Temper...	D Temper...	D Temper...	D WindSp...	D WindS
DewpointAvg...	1	0.978	0.98	-0.133	-0.004	-0.303	0.609	0.62	0.462	0.157	-0.056	-0.007	0.942	0.908	0.949	-0.431	-0.371
DewpointHigh...	0.978	1	0.933	-0.117	-0.005	-0.244	0.563	0.645	0.394	0.156	-0.042	-0.003	0.938	0.914	0.931	-0.387	-0.318
DewpointLow...	0.98	0.933	1	-0.146	-0.006	-0.352	0.6	0.567	0.503	0.145	-0.064	-0.006	0.917	0.877	0.94	-0.454	-0.414
Discharge, cu...	-0.133	-0.117	-0.146	1	0.025	0.269	0.108	0.287	0.106	0.458	0.017	-0.024	-0.192	-0.193	-0.179	0.253	0.268
Elevation abo...	-0.004	-0.005	-0.006	0.025	1	0.026	-0.044	-0.042	-0.035	-0.026	-0.011	-0.008	0.01	0.008	0.006	0.004	0.016
GustSpeedMa...	-0.303	-0.244	-0.352	0.269	0.026	1	-0.216	-0.2	-0.151	0.17	0.06	-0.018	-0.257	-0.258	-0.261	0.841	0.907
HumidityAvg...	0.609	0.563	0.6	0.108	-0.044	-0.216	1	0.828	0.891	0.365	0.097	0.09	0.319	0.265	0.377	-0.287	-0.26
HumidityHigh...	0.62	0.645	0.567	0.087	-0.042	-0.2	0.828	1	0.58	0.25	0.17	0.165	0.416	0.402	0.417	-0.322	-0.243
HumidityLow...	0.462	0.394	0.503	0.106	-0.035	-0.151	0.891	0.58	1	0.37	0.052	0.041	0.182	0.094	0.273	-0.153	-0.189
Precipitation...	0.157	0.156	0.145	0.458	-0.026	0.17	0.365	0.25	0.37	1	-0.02	-0.061	0.045	0.007	0.085	0.119	0.123
PressureMax...	-0.056	-0.042	-0.064	0.017	-0.011	0.06	0.097	0.17	0.062	-0.02	1	0.988	-0.029	-0.012	-0.047	0.028	0.065
PressureMinh...	-0.007	-0.003	-0.006	-0.024	-0.008	-0.018	0.09	0.165	0.041	-0.061	0.988	1	0.032	0.05	0.012	-0.047	-0.011
Temperature...	0.942	0.938	0.917	-0.192	0.01	-0.257	0.319	0.416	0.182	0.045	-0.029	0.032	1	0.985	0.981	-0.387	-0.322
Temperature...	0.908	0.914	0.877	-0.193	0.008	-0.258	0.265	0.402	0.094	0.007	-0.012	0.05	0.985	1	0.943	-0.403	-0.319
Temperature...	0.949	0.931	0.94	-0.179	0.006	-0.261	0.377	0.417	0.273	0.085	-0.047	0.012	0.981	0.943	1	-0.365	-0.328
WindSpeedAv...	-0.431	-0.387	-0.454	0.253	0.004	0.841	-0.287	-0.322	-0.153	0.119	0.028	-0.047	-0.387	-0.403	-0.365	1	0.845
WindSpeedM...	-0.371	-0.318	-0.414	0.268	0.016	0.907	-0.26	-0.243	-0.189	0.123	0.065	-0.011	-0.322	-0.319	-0.328	0.845	1

Εικόνα 6: Η συσχέτιση μεταξύ των ακατέργαστων δεδομένων και των μεταβλητών στο Knime για το Long Island (Knime, linear correlation matrix node / linear correlation measures).

Οι συσχετίσεις μεταξύ της απορροής του νερού και των κλιματολογικών μεταβλητών, σύμφωνα με το αρχικό μοντέλο, είναι στατιστικά σημαντικές. Για την εξήγηση, της χρονοσειράς της απορροής του νερού με τη χρήση των ακατέργαστων δεδομένων από τις κλιματικές μεταβλητές, πραγματοποιείται μια γραμμική παλινδρόμηση με συντελεστή παλινδρόμησης (R^2) ίσο με 0.590. Ωστόσο, στην παρούσα ροή εργασίας ο συντελεστής παλινδρόμησης (R^2) διαφέρει ελαφρώς εξαιτίας των περισσότερων μεταβλητών. Το μοντέλο αυτό δίνει ένα μέτριο R^2 για την χρονοσειρά της απορροής του νερού (σχέση 4). Συγκεκριμένα, η χρονοσειρά των

ακατέργαστων δεδομένων της απορροής του νερού, μπορεί να εκφραστεί μέσα από μια γραμμικής παλινδρόμησης ως εξής:

$$\begin{aligned} WD(t) = & -0.378T(t) + 0.113TD(t) + 0.049WS(t) \\ & -0.551GWL(t) + 0.118PR(t) + \varepsilon(t), \end{aligned} \quad (4)$$

$$R^2 = 0.590,$$

όπου $WD(t)$, $T(t)$, $TD(t)$, $WS(t)$, $GWL(t)$ και $PR(t)$ δηλώνουν τα ανεπεξέργαστα δεδομένα του λογαρίθμου της απορροής του νερού, τη θερμοκρασία, την παλίρροια, την ταχύτητα του ανέμου, το επίπεδο των υπόγειων υδάτων και τα κατακρημνίσματα αντίστοιχα. Επιπλέον, στην σχέση 4, το $\varepsilon(t)$ αντιπροσωπεύει τα κατάλοιπα της σχέσης και R^2 είναι το τετράγωνο του συντελεστή συσχέτισης. Η σχέση της απορροής του νερού με τις κλιματικές μεταβλητές ενισχύεται με τον διαχωρισμό των εποχιακών (seasonal), των μακροπρόθεσμων (long term) και βραχυπρόθεσμων (short term) μεταβολών σε όλες τις χρονοσειρές. Ο διαχωρισμός των κλιμάκων της χρονοσειράς επαληθεύεται μέσω της εξέτασης της συνοχής μεταξύ των μεταβλητών. Οι βασικές περιόδους όλων των μεταβλητών που προέρχονται από τα περιοδογράμματα είναι συγκεκριμένες. Μπορεί να παρατηρηθεί ότι σχεδόν όλες οι μεταβλητές αποτελούνται από ένα χρονικό διάστημα 365 ημερών που περιγράφει την εποχική συνιστώσα των μεταβλητών. Ορισμένες μεταβλητές αποτελούνται από σύντομες περιόδους (7 ημέρες, 4 ημέρες, κλπ), οι οποίες συνδέονται με τα βραχυπρόθεσμα συστατικά των μεταβλητών (Πίνακας 2). Ειδικότερα, τα κατακρημνίσματα αποτελούν μια μεταβλητή που παρουσιάζεται κυρίως σε μικρές περιόδους και για το λόγο αυτό συμβάλλει τα μέγιστα στα βραχυπρόθεσμα συστατικά της χρονοσειράς της απορροής του νερού.

Πίνακας 2: Ο διαχωρισμός των περιόδων (Tsakiri et al., 2014)

Table 2: Periods derived by the graph of periodograms of the water discharge variable and the climatic variables.

Variable	Period (days)		
	First peak of period	Second peak of period	Third peak of period
Water discharge	365	594	
Groundwater level	365	991	198
Tide	365	185	14
Precipitation	27	7	4
Wind speed	365		
Temperature	365		

Στις Εικόνες 7 και 8 γεννώνται δύο περιοδογράμματα για τη θερμοκρασία και τα κατακρημνίσματα χρησιμοποιώντας τον DZ (DiRienzo A., Zurbenko I., 1999) αλγόριθμο στο λογισμικό R. Από τον Πίνακα 2 και τις Εικόνες 7 και 8, μπορούμε να συμπεράνουμε ότι είναι απαραίτητο να αποσυντεθεί η χρονοσειρά σε διαφορετικά συστατικά για να αποφευχθεί η συνεισφορά των διαφόρων συχνοτήτων μεταξύ των κλιμάκων των χρονοσειρών. Σημειώνεται ότι, το λογισμικό R συνεργάζεται άψογα με το λογισμικό Ktime (όπως για παράδειγμα στο

Σχήμα 1), το μονό που απαιτείται είναι το λογισμικό R, ο αλγόριθμος που θα τρέξει το R και μερικοί κόμβοι στο Knime (R nodes pack).

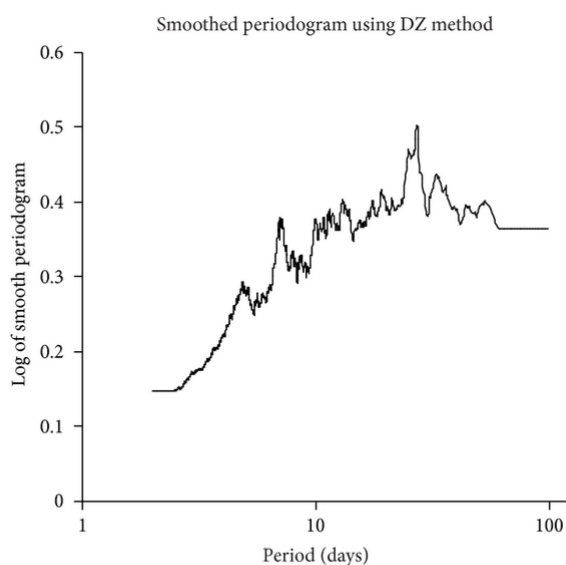


Figure 3: Periodogram of the precipitation variable. The y -axis describes the spectral density using the KZFT (Kolmogorov Zurbenko Fourier Transform) method.

Εικόνα 7: Περιοδόγραμμα κατακρημνισμάτων. Ο κάθετος άξονας περιγράφει τη φασματική πυκνότητα χρησιμοποιώντας τη μέθοδο KZFT (Kolmogorov Zurbenko Fourier Transform) (Tsakiri et al., 2014).

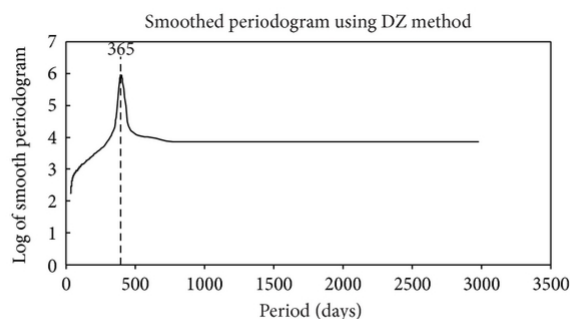


Figure 2: Periodogram of the temperature variable. The y -axis describes the spectral density using the KZFT (Kolmogorov Zurbenko Fourier Transform) method.

Εικόνα 8: Περιοδόγραμμα θερμοκρασίας. Ο κάθετος άξονας περιγράφει τη φασματική πυκνότητα χρησιμοποιώντας τη μέθοδο KZFT (Kolmogorov Zurbenko Fourier Transform) (Tsakiri et al., 2014).

Για την αποσύνθεση της χρονοσειράς, στο αρχικό μοντέλο χρησιμοποιήθηκε το φίλτρο $KZ_{33,3}$ (μήκος 33 με 3 επαναλήψεις) για να παρέχει μια φυσική εξήγηση της βάσης των χρονοσειρών της απορροής του νερού. Σε συνέχεια των Rao et al. [24], οι Tsakiri et al., επέλεξαν τις παραμέτρους του φίλτρου KZ για την παροχή της βέλτιστης λύσης της μελέτης τους. Συγκεκριμένα, οι παράμετροι επιλέχθηκαν για να διατηρήσουν τις ιδιότητες του κάθε συστατικού των χρονοσειρών και να μειώσουν τις διακυμάνσεις των βραχυπρόθεσμων συνιστωσών στις μακροπρόθεσμες και εποχιακές συνιστώσες των χρονοσειρών.

Το φίλτρο KZ εφαρμόστηκε στον λογάριθμο της καθημερινής απορροής νερού και παράγει μια χρονοσειρά που στερείται των βραχυπρόθεσμων διακυμάνσεων και αποτελείται μόνο από τις μακροπρόθεσμες διακυμάνσεις των χρονοσειρών ($L(t)$). Το ίδιο φίλτρο εφαρμόστηκε και στις μεταβλητές της ημερήσιας θερμοκρασίας, της παλίρροιας, της ταχύτητας του ανέμου, των κατακρημνισμάτων και του επιπέδου των υπόγειων υδάτων. Αντίστοιχα, στην παρούσα ροή εργασίας το φίλτρο $KZ_{33,3}$ αντικαταστάθηκε από τρία κινούμενα μέσα με αντίστοιχο μήκος. Ωστόσο μεταγενέστερα επιλέχθηκε μήκος 29 για την καλύτερη αποτύπωση, τόσο των διακυμάνσεων των χρονοσειρών, όσο και των δίσεκτων ετών.

2.5. Πρόβλεψη των μακροπρόθεσμων συστατικών της απορροής του νερού

Για να εξηγήσουμε τα μακροπρόθεσμα συστατικά της απορροής του νερού και τη σχέση τους με το επίπεδο των υπόγειων υδάτων και τις κλιματικές μεταβλητές, σύμφωνα με το αρχικό μοντέλο, εξετάστηκε το άθροισμα των φιλτραρισμένων δεδομένων της ημερήσιας θερμοκρασίας, της παλίρροιας, της ταχύτητας του ανέμου και των κατακρημνισμάτων τεσσάρων συνεχών ημερών. Για την ανάλυση, στο αρχικό μοντέλο, ορίστηκαν οι χρονοσειρές των μακροπρόθεσμων συνιστωσών, δηλαδή της απορροής του νερού, της θερμοκρασίας, της παλίρροιας, της ταχύτητας του ανέμου, των κατακρημνισμάτων και του επιπέδου των υπόγειων υδάτων με $WDKZ(t)$, $TKZ(t)$, $TDKZ(t)$, $WSKZ(t)$, $PRKZ(t)$, και $GWLKZ(t)$, αντίστοιχα. Στην παρούσα ροή εργασίας όλες οι μεταβλητές έχουν, είτε την αρχική ονοματολογία (όπως την ορίζει η πηγή), είτε μπορούμε να ορίσουμε εμείς μια. Ωστόσο, προτείνεται η πρώτη περίπτωση, τόσο γιατί δεν χρειάζεται να δημιουργούμε στην ροή εργασίας επιπρόσθετη επεξεργασία, όσο και για την αποφυγή επιπλοκών. Ένα παραδείγματα επιπλοκής μπορεί να είναι το ότι η πηγή μπορεί να αλλάξει ονοματολογία στα δεδομένα της και έτσι τα νέα δεδομένα που θα εισέρχονται, μετά τον βρόγχο (loop), θα κάνουν την ροή εργασίας μας να διακοπεί.

Η συσχέτιση μεταξύ της μακροπρόθεσμης συνιστώσας της απορροής του νερού και των κατακρημνισμάτων για το Schoharie Creek είναι αδύναμη (Πίνακας 3). Αυτό μπορεί να ελεγχθεί από τις κορυφές του περιοδογράμματος της μεταβλητής των κατακρημνισμάτων που συμβαίνουν σε σύντομες περιόδους (Πίνακας 2). Ως εκ τούτου, η μεταβλητή κατακρημνίσματα συμβάλλει κυρίως στη βραχυπρόθεσμη συνιστώσα της απορροής του νερού, ενώ το επίπεδο των υπόγειων υδάτων συμβάλλει κατά κύριο λόγο στη μακροπρόθεσμη συνιστώσα της απορροής του νερού (Πίνακας 3). Αντίστοιχα, στην ροή εργασίας μας η συσχέτιση μεταξύ της μακροπρόθεσμης συνιστώσας της απορροής του νερού και των κατακρημνισμάτων, τόσο για το Schoharie Creek (ίδια αποτελέσματα με το αρχικό μοντέλο με μικρή διάφορα του R^2 λόγω των περισσότερων μεταβλητών), όσο και για το Long Island (Atlantic Ocean/Long Island Sound Watershed, Εικόνα 9, <http://www.dec.ny.gov/lands/48375.html>, Department of Environmental Conservation, New York State) (Εικόνες 10 και 11) είναι αδύναμη.



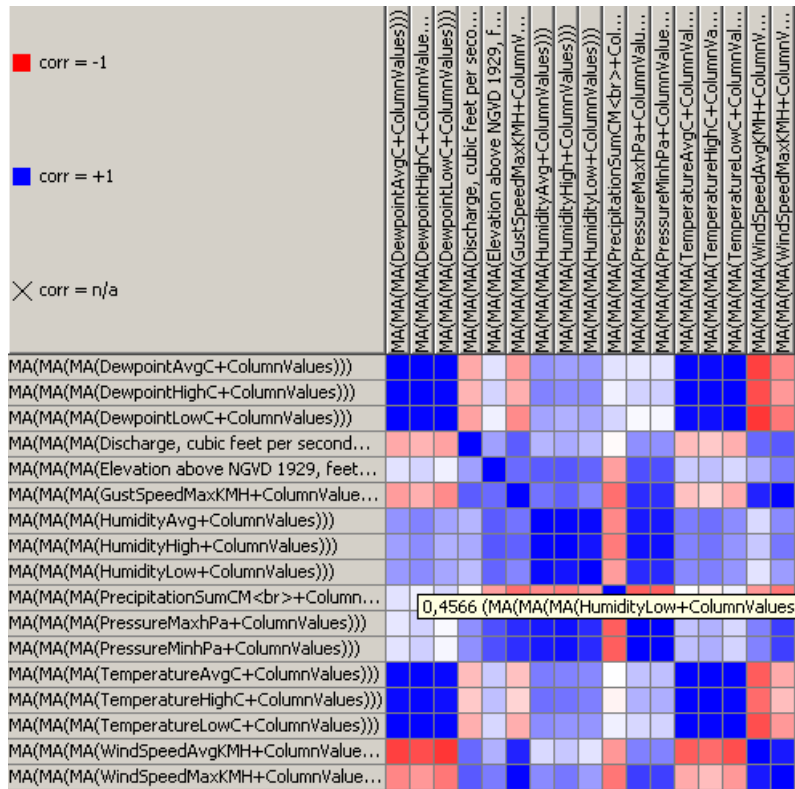
Εικόνα 9: Λεπτομερής χάρτης της λεκάνης απορροής του Long Island (Department of Environmental Conservation, New York State, <http://www.dec.ny.gov/lands/53733.html>).

Η στάθμη των υπόγειων υδάτων είναι η μόνη μεταβλητή που αποτελείται από χρονικές περιόδους μεγαλύτερες από ένα έτος (991 ημέρες, Πίνακας 2). Η περίοδος αυτή μπορεί να εξηγηθεί από τα μεγαλύτερης χρονικής κλίμακας ατμοσφαιρικά επεισόδια ή από τις μεταβολές της ροής του ποταμού λόγω ανθρωπογενών επιπτώσεων (όπως φράγματα, έργα ηλεκτρικής ενέργειας, γέφυρες, κλπ).

Πίνακας 3: Η συσχέτιση μεταξύ των μακροπρόθεσμων συστατικών της απορροής του νερού και των κατακρημισμάτων καθώς και των υπόλοιπων μεταβλητών του αρχικού μοντέλου των Tsakiri et al. (2014).

Table 3: Correlation matrix between the long term components of the water discharge, the climatic variables, and the groundwater level.

Variable	Water discharge	Groundwater level	Temperature	Tide	Wind speed	Precipitation (4 days)
Water discharge	1	-0.744	-0.641	0.39	0.596	0.175
Groundwater level	-0.744	1	0.26	-0.412	-0.43	-0.224
Temperature	-0.641	0.26	1	0.033	-0.691	0.103
Tide	0.39	-0.412	0.033	1	0.462	-0.114
Wind speed	0.596	-0.43	-0.691	0.462	1	-0.131
Precipitation (4 days)	0.175	-0.224	0.103	-0.114	-0.131	1



Εικόνα 10: Το μοντέλο συσχέτισης του Knime μεταξύ των μακροπρόθεσμων συστατικών και των μεταβλητών για το Long Island (Knime, linear correlation matrix node).

Row ID	D MA(MA...	D MA(MA...	D MA(MA...	D MA(MA(Discharge...	D MA(MA(Elevation...	D MA(MA...	D MA(MA...	D MA(MA...	D MA(MA...	D MA(MA...	D MA(MA...	D MA(MA...	D MA(MA...	D MA(MA...
MA(MA(MA(DewpointAvgC+ColumnValues)))	1	0.995	0.996	-0.336	0.112	-0.39	0.424	0.378	0.401	0.113	0.096	0.108	0.984	
MA(MA(MA(DewpointHighC+ColumnValues)))	0.995	1	0.984	-0.294	0.168	-0.314	0.489	0.45	0.457	0.059	0.176	0.188	0.992	
MA(MA(MA(DewpointLowC+ColumnValues)))	0.996	0.984	1	-0.366	0.06	-0.455	0.362	0.31	0.35	0.164	0.023	0.035	0.968	
MA(MA(MA(Discharge, cubic feet per second...))	-0.336	-0.294	-0.366	1	0.372	0.638	0.292	0.334	0.262	-0.019	0.441	0.436	-0.263	
MA(MA(MA(Elevation above NGVD 1929, feet...))	0.112	0.168	0.06	0.372	1	0.582	0.641	0.66	0.608	-0.377	0.689	0.689	-0.209	
MA(MA(MA(GustSpeedMaxKMh+ColumnValue...))	-0.39	-0.314	-0.455	0.638	0.582	1	0.55	0.62	0.479	-0.564	0.826	0.82	-0.242	
MA(MA(MA(HumidityAvg+ColumnValues)))	0.424	0.489	0.362	0.292	0.641	0.55	1	0.966	0.97	-0.466	0.905	0.909	0.516	
MA(MA(MA(HumidityHigh+ColumnValues)))	0.378	0.45	0.31	0.334	0.66	0.62	0.966	1	0.923	-0.519	0.941	0.944	0.489	
MA(MA(MA(HumidityLow+ColumnValues)))	0.401	0.457	0.35	0.262	0.608	0.479	0.97	0.923	1	-0.388	0.837	0.84	0.466	
MA(MA(MA(PrecipitationSumCM +Column...))	0.113	0.059	0.164	0.019	-0.377	-0.564	-0.466	-0.519	-0.388	1	-0.632	-0.631	0.004	
MA(MA(MA(PressureMaxhPa+ColumnValues)))	0.096	0.176	0.023	0.441	0.689	0.826	0.905	0.941	0.837	-0.632	1	1	0.238	
MA(MA(MA(PressureMinhPa+ColumnValues)))	0.108	0.188	0.035	0.436	0.689	0.82	0.909	0.944	0.84	-0.631	1	1	0.25	
MA(MA(MA(TemperatureAvgC+ColumnValue...))	0.984	0.992	0.968	-0.263	0.209	-0.242	0.516	0.489	0.466	0.004	0.238	0.25	1	
MA(MA(MA(TemperatureHighC+ColumnValue...))	0.966	0.981	0.945	-0.21	0.252	-0.169	0.566	0.549	0.503	-0.047	0.308	0.319	0.996	
MA(MA(MA(TemperatureLowC+ColumnValue...))	0.993	0.994	0.984	-0.312	0.156	-0.319	0.453	0.419	0.415	0.058	0.159	0.171	0.996	
MA(MA(MA(WindSpeedAvgKMh+ColumnValue...))	-0.749	-0.699	-0.789	0.591	0.31	0.865	0.148	0.219	0.111	-0.407	0.499	0.49	-0.639	
MA(MA(MA(WindSpeedMaxKMh+ColumnValue...))	-0.475	-0.406	-0.534	0.649	0.518	0.985	0.46	0.534	0.388	-0.536	0.762	0.755	-0.333	

Εικόνα 11: Η συσχέτιση μεταξύ των μακροπρόθεσμων συστατικών και των μεταβλητών στο Knime για το Long Island (Knime, linear correlation matrix node / linear correlation measures).

Η πρόβλεψη, στο αρχικό μοντέλο, της μακροπρόθεσμης συνιστώσας της απορροής του νερού πραγματοποιείται με μια γραμμική παλινδρόμηση, χρησιμοποιώντας τον φιλτραρισμένο λογάριθμο των χρονοσειρών της απορροής του νερού, των κλιματικών μεταβλητών και του επίπεδου των υπόγειων υδάτων με τιμή R^2 ίση με 0.833. Η μακροπρόθεσμη συνιστώσα του λογαρίθμου της χρονοσειράς της απορροής του νερού μπορεί να εκφραστεί ως εξής:

$$WDKZ(t) = -0.704TKZ(t) + 0.337TDKZ(t) - 0.241WSKZ(t) - 0.494GWLKZ(t) + 0.143PRKZ(t) + \varepsilon(t), \quad (5)$$

$$R^2 = 0.833$$

Η τιμή P που αφορά τους συντελεστές των μεταβλητών του γραμμικού μοντέλου παλινδρόμησης είναι ίση με το μηδέν. Έτσι, ο συντελεστής της γραμμικής παλινδρόμησης είναι στατιστικά σημαντικός. Η κλίμακα του $\varepsilon(t)$ αντιστοιχεί σε ποσοστιαία αλλαγή της μακροπρόθεσμης συνιστώσας των δεδομένων της απορροής του νερού λόγω άλλων επιπτώσεων εκτός από τις κλιματικές μεταβλητές.

Έτσι από την σχέση (5) συμπεραίνεται ότι, το 83,3% των μακροπρόθεσμων διακυμάνσεων της απορροής του νερού μπορούν να εξηγηθούν από τις μακροπρόθεσμες διακυμάνσεις της στάθμης των υπόγειων υδάτων και των σημαντικών κλιματικών μεταβλητών. Οι μεταβλητές αυτές (όπως φαίνεται στον Πίνακα 3) είχαν επιλεγεί στο αρχικό μοντέλο καθώς οι αρχικές πρόσθετες κλιματικές μεταβλητές, όπως το επίπεδο της θάλασσας, η κατεύθυνση του ανέμου ή η σχετική υγρασία δεν ήταν στατιστικά σημαντικές (η τιμή P που σχετίζεται με το t -test είναι μεγαλύτερη από το επίπεδο σημαντικότητας). Ωστόσο, στην παρούσα ροή εργασίας, έχοντας νέα πηγή με περισσότερες κλιματικές μεταβλητές, όπως φαίνεται και στην Εικόνα 11, υπάρχουν και άλλες με μεγάλο επίπεδο σημαντικότητας, όπως η ταχύτητα του ανέμου και η μέγιστη ταχύτητα ριπής. Προκύπτουν λοιπόν έτσι έχουμε διαφορετικές τιμές στον συντελεστή γραμμικής παλινδρόμησης, οι οποίες αντικατοπτρίζουν και μια πιο ακριβή αποτύπωση της πραγματικής κατάστασης.

Η σχέση (5) δίνεται για το φυσικό λογάριθμο της απορροής του νερού, ο πρόσθετος όρος (t) έχει μια πολλαπλασιαστική επίδραση, $\exp(\varepsilon(t))$, στα αρχικά δεδομένα. Δεδομένου ότι ο όρος (t) είναι αρκετά μικρός, μπορούμε να πάρουμε ότι:

$$\exp(\varepsilon(t)) \approx 1 + \varepsilon(t), \text{ δηλαδή,}$$

$$\ln(1 + \varepsilon(t)) \approx \varepsilon(t). \quad (6)$$

Ως εκ τούτου, το $\varepsilon(t) * 100$ αντιστοιχεί σε ποσοστιαίες αλλαγές μακροπρόθεσμων διακυμάνσεων της απορροής του νερού που δεν εξηγούνται από το επίπεδο των υπογείων υδάτων και τις κλιματικές μεταβλητές.

2.6. Πρόβλεψη των εποχιακών συστατικών της απορροής του νερού

Η εποχιακή (seasonal) συνιστώσα μιας χρονοσειράς αντιπροσωπεύει τις διακυμάνσεις από έτος σε έτος της αντίστοιχης μεταβλητής. Η πρόβλεψη της εποχιακής συνιστώσας των χρονοσειρών της απορροής του νερού γίνεται με την χρήση των εποχιακών συστατικών των κλιματικών μεταβλητών και της στάθμης των υπόγειων υδάτων. Ειδικότερα, η εποχιακή συνιστώσα της απορροής του νερού, $WD_{Se}(d)$, μπορεί να ορίζεται από την σχέση:

$$WD_{Se}(d) = \frac{1}{r} \sum_{(y=1)}^Y (WD(t) - WD) \quad (7)$$

$d = 1, \dots, 365$, $d=1$ αντιπροσωπεύει τον Ιανουάριο και ούτω καθεξής

$$y = 1, \dots, Y, \quad (8)$$

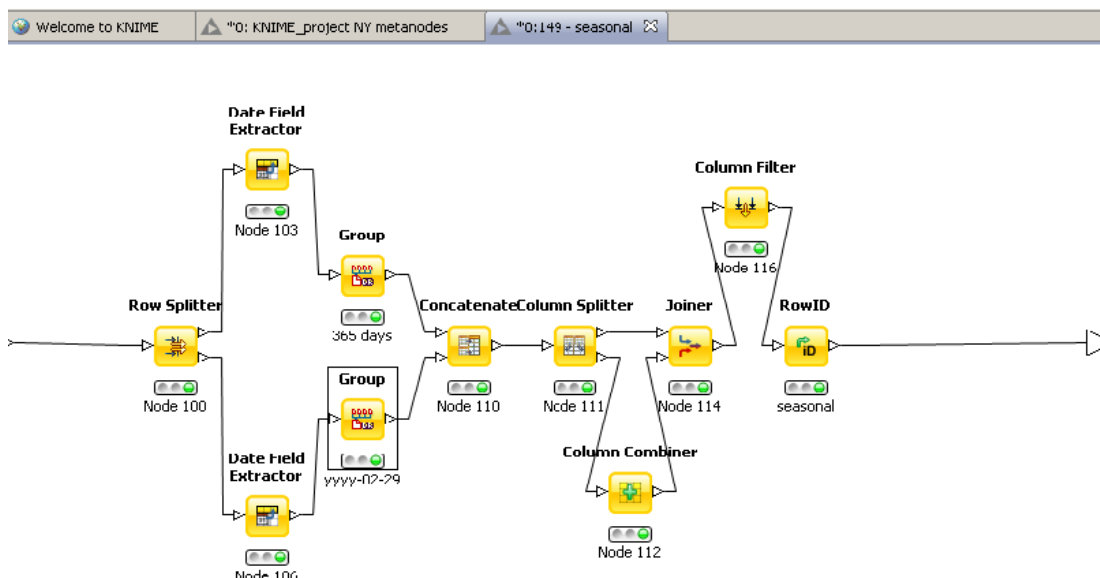
όπου το d αντιπροσωπεύει τις ημέρες του έτους, y είναι ο αριθμός των ετών των παρατηρούμενων τιμών, $WD(t)$ είναι τα ανεπεξέργαστα δεδομένα της χρονοσειράς της απορροής του νερού και $WD_{KZ}(t)$ είναι τα μακροπρόθεσμα συστατικά της απορροής του νερού που προέρχεται από το εφαρμογή του φίλτρου KZ. Ομοίως, έχουμε ορίσει τα εποχιακά συστατικά των κλιματικών μεταβλητών και του επιπέδου των υπόγειων υδάτων. Ειδικότερα, τα εποχιακά συστατικά των μεταβλητών της θερμοκρασίας, της παλίρροιας, της ταχύτητας του ανέμου, του αθροίσματος τεσσάρων συνεχόμενων ημερών με βροχοπτώσεις και του επιπέδου των υπόγειων υδάτων συμβολίζονται με $M_{Se}(d)$, $TD_{Se}(d)$, $WS_{Se}(d)$, $PR_{Se}(d)$ και $GWL_{Se}(d)$ αντίστοιχα. Η μέγιστη συσχέτιση μεταξύ της απορροής του νερού και των κατακρημνισμάτων εμφανίζεται όταν θεωρούμε την μεταβλητή ως το άθροισμα της για τέσσερις ημέρες.

Η συσχέτιση μεταξύ της εποχικής συνιστώσας της απορροής του νερού και των κλιματικών μεταβλητών, στο αρχικό μοντέλο, όπως φαίνεται και στον Πίνακα 4, είναι ισχυρή για τις μεταβλητές της θερμοκρασίας, της παλίρροιας, καθώς και της στάθμης των υπόγειων υδάτων. Επιπλέον, όλες οι μεταβλητές που εξετάστηκαν στο αρχικό μοντέλο έχουν μια περίοδο 365 ημερών (Πίνακας 2). Η μεταβλητή που σχετίζεται περισσότερο με την εποχική συνιστώσα της απορροής του νερού είναι αυτή της παλίρροιας (Πίνακας 4). Ωστόσο, στην παρούσα ροή εργασίας έχουμε κάποιες διαφοροποιήσεις. Βασικότερη αυτών, πέραν των περισσότερων μεταβλητών, είναι η ενσωμάτωση και δίσεκτων ετών που πραγματοποιήθηκε με την χρήση φίλτρων. Η δομή φαίνεται στο Σχήμα 3 που αποτελεί την εσωτερική διάταξη του μετακομβού 149 (metanode, node 149, seasonal) του Σχήματος 1. Λίγο πιο συγκεκριμένα, το αρχικό φίλτρο διαχωρίζει την 366^η μέρα, αν υπάρχει, και στην συνέχεια την προσθέτει στις 365.

Πίνακας 4: Η συσχέτιση μεταξύ του εποχιακού συστατικού της απορροής του νερού και των υπόλοιπων μεταβλητών του αρχικού μοντέλου (Tsakiri et al. 2014)

Table 4: Correlation matrix between the seasonal components of the water discharge, the climatic variables, and the groundwater level.

Variable	Water discharge	Groundwater level	Temperature	Tide	Wind speed	Precipitation (4 days)
Water discharge	1	0.63	0.579	0.893	0.418	0.27
Groundwater level	0.63	1	0.492	0.831	0.406	0.012
Temperature	0.579	0.492	1	0.627	0.3	-0.046
Tide	0.893	0.831	0.627	1	0.428	0.089
Wind speed	0.418	0.406	0.3	0.428	1	0.044
Precipitation (4 days)	0.27	0.012	-0.046	0.089	0.044	1



Σχήμα 3: Μέσα στον μετα-κόμβο 146 (seasonal) φαίνεται η σύνθεση των κόμβων ώστε να προστεθούν τα δίσεκτα έτη στην ροή εργασίας μας. Μόνος περιορισμός είναι να έχουμε τετραψήφιο αριθμό έτους. (Τροποποιημένο από Καβαλιέρος Α., 2016, Knime, metanode, node 149 seasonal).

Και εδώ η πρόβλεψη της εποχιακής συνιστώσας της απορροής του νερού έγινε με γραμμική παλινδρόμηση χρησιμοποιώντας τα εποχιακά συστατικά των κλιματικών μεταβλητών και του επιπέδου των υπόγειων υδάτων με συντελεστή παλινδρόμησης R^2 ίσο με 0.912. Ειδικότερα, η εποχιακή συνιστώσα του λογαρίθμου της απορροής του νερού μπορεί να εκφραστεί με την σχέση:

$$WD_{Se}(d) = 0.042T_{Se}(d) + 1.103TD_{Se}(d) + 0.061WS_{Se}(d) - 0.335GWL_{Se}(d) + 0.174PR_{Se}(d) + \varepsilon(t), \quad (9)$$

$$R^2 = 0.912$$

Ως εκ τούτου, το 91,2 % των διακυμάνσεων της εποχιακής συνιστώσας της απορροής του νερού μπορεί να εξηγηθεί από τις κλιματικές μεταβλητές και τη στάθμη των υπόγειων υδάτων, όπως περιγράφεται στη σχέση 9. Οι τιμές P για τους συντελεστές των μεταβλητών της γραμμικής παλινδρόμησης είναι ίσες με το μηδέν. Έτσι, οι συντελεστές της γραμμικής παλινδρόμησης είναι στατιστικά σημαντικοί. Σημειώνεται ότι η προσθήκη επιπλέον κλιματικών μεταβλητών, τόσο στο αρχικό μοντέλο, όπως το επίπεδο της θάλασσας, η κατεύθυνση του ανέμου και η σχετική υγρασία, όσο και στην παρούσα ροή εργασίας, όπως το σημείο δρόσου, η ταχύτητα του ανέμου και η μέγιστη ταχύτητα ριπής δεν αλλάζουν την τιμή του R^2 στη σχέση 9.

2.7. Πρόβλεψη των βραχυπρόθεσμων συστατικών της απορροής του νερού

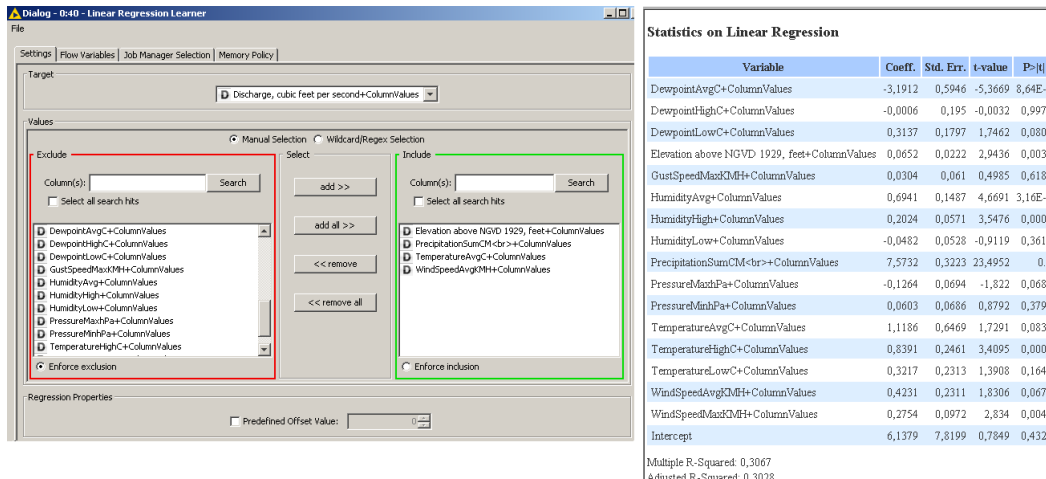
Για την πρόβλεψη της βραχυπρόθεσμης συνιστώσας της απορροής του νερού, λαμβάνονται υπόψη, οι βραχυπρόθεσμες συνιστώσες των κλιματικών μεταβλητών και του επίπεδου των υπόγειων υδάτων. Η βραχυπρόθεσμη συνιστώσα της χρονοσειράς της απορροής του νερού, $WD_{Sh}(t)$, μπορεί να οριστεί ως εξής:

$$WD_{Sh}(t) = WD(t) - WD_{KZ}(t) - WD_{Se}(d), \quad (10)$$

όπου το $WD(t)$ αντιπροσωπεύει τα ανεπεξέργαστα δεδομένα της χρονοσειράς της απορροής του νερού, το $WD_{KZ}(t)$ αντιπροσωπεύει τη μακροπρόθεσμη συνιστώσα της απορροής του νερού και $WD_{Se}(d)$ είναι η εποχιακή συνιστώσα της χρονοσειράς της εκκένωσης του νερού. Ομοίως, μπορούμε να ορίσουμε τα βραχυπρόθεσμα συστατικά των υπόλοιπων μεταβλητών. Σε αυτό το σημείο αξίζει να σημειωθεί ότι, πράξεις μέσα στην ροή εργασίας μας (έκτος αυτών που πραγματοποιούνται μέσω του λογισμικού R και αυτών για την γραμμική παλινδρόμηση), πραγματοποιούνται από προγραμματιζόμενους μαθηματικούς κόμβους (math formula node, Knime).

Για τη βραχυπρόθεσμη συνιστώσα της απορροής του νερού, εξετάστηκαν δυο διαφορετικά μοντέλα. Το πρώτο περιγράφει την πρόβλεψη της βραχυπρόθεσμης συνιστώσας κατά τη θερινή περίοδο (Μάιος - μέσα Σεπτεμβρίου) ενώ το δεύτερο, κατά τη διάρκεια του χειμώνα (Δεκέμβριος - μέσα Μαρτίου). Το μοντέλο των Tsakiri et al. δημιούργησε δύο μοντέλα καθώς πλημμύρες στα ποτάμια της Νέας Υόρκης προκαλούνται, είτε από εκτεταμένα κατακρημνίσματα (εκτεταμένες βροχοπτώσεις ή τροπικές καταιγίδες κατά τη διάρκεια της θερινής περιόδου), είτε από ταχεία τήξη του χιονιού (φαινόμενα ice jam κατά τη διάρκεια της παρατεταμένης περιόδου του χειμώνα). Το μοντέλο των βραχυπρόθεσμων συστατικών του χειμώνα αποδίδει τα μέγιστα, όπως είναι αντιληπτό, κατά τη διάρκεια της παρατεταμένης περιόδου του χειμώνα (Δεκέμβριος - Μάρτιος), ενώ αντίστοιχα το θερινό μοντέλο κατά τη διάρκεια της παρατεταμένης περιόδου του καλοκαιριού (Μάιος - Σεπτέμβριος).

Στην παρούσα εργασία, ο παραπάνω διαχωρισμός για την συσχέτιση της μακροπρόθεσμης συνιστώσας της απορροής του νερού των κλιματικών μεταβλητών και του επίπεδου των υπόγειων υδάτων πραγματοποιείται (και σε πραγματικό χρόνο) με την χρήση φίλτρων (όπως αντίστοιχα γίνεται και στο Σχήμα 3 για τα δίσεκτα έτη), τα οποία διαχωρίζουν χρονικά διαστήματα σε αντιστοιχία με το χειμερινό και το καλοκαιρινό μοντέλο. Εν συνεχεία, όπως και στο αρχικό μοντέλο, πραγματοποιείται το μοντέλο συσχέτισης (correlation matrix) για κάθε χρονικό διάστημα ξεχωριστά. Τέλος, η γραμμική παλινδρόμηση (linear regression) πραγματοποιείται από αντίστοιχους κόμβους (όπως φαίνεται στο Σχήμα 1 και 2), σε αντιστοιχία με το αρχικό μοντέλο, για τον υπολογισμό του συντελεστή συσχέτισης R^2 (Εικόνα 12).



Εικόνα 12: Οι επιλογές του κόμβου της γραμμικής παλινδρόμησης (αριστερά) και η μορφή των αποτελεσμάτων (δεξιά). Στις επιλογές (αριστερό μέρος εικόνας) διακρίνονται η επιλογή του στόχου της γραμμικής παλινδρόμησης, οι διαθέσιμες μεταβλητές (με κόκκινο περίγραμμα), οι επιλεγμένες μεταβλητές (με πράσινο περίγραμμα) και η επιλογή της προκαθορισμένης τιμής μετατόπισης (Knime, linear regression node).

Η παρούσα ροή εργασίας περιλαμβάνει περισσότερες μεταβλητές (και θα μπορούσε να περιλάβει και περισσότερες αν το επιλέγαμε) από το αρχικό μοντέλο. Σε κάθε περίπτωση, αυτά τα εργαλεία correlation matrix και linear regression είναι εκείνα που υποδεικνύουν ποιες μεταβλητές χρειαζόμαστε. Μέσα σε κάθε κόμβο μπορούμε, είτε να βλέπουμε όλες τις μεταβλητές, είτε να επιλεγούμε εμείς τις επιθυμητές. Η δυνατότητα της ενσωμάτωσης πολλών μεταβλητών, μπορεί μεν να δίνει πιο ακριβή αποτελέσματα σε σχέση με την πραγματική αναμενόμενη χρονοσειρά που θα προβλέψουμε, ωστόσο όμως, μπορεί να δημιουργηθούν επιπλοκές στους συντελεστές συσχέτισης. Μην ξεχνάμε πως το ζητούμενο, για ένα σύστημα έγκαιρης ποδοπάτησης, είναι οι κορυφές μιας χρονοσειράς που θα προβλέψουμε (picks, αυτά είναι που υποδηλώνουν πλημύρα όπως θα δούμε παρακάτω).

Είναι αντιληπτό, ότι η προσαρμοστικότητα και η ταχύτητα που προσφέρει ένα πρόγραμμα data mining, είναι βασικά συστατικά για την χρήση του σε συστήματα έγκαιρης προειδοποίησης.

2.7.1. Μοντέλο θερινής περιόδου

Η πρόβλεψη της βραχυπρόθεσμης συνιστώσας της χρονοσειράς της απορροής του νερού, κατά την θερινή περίοδο, έγινε μέσω μιας γραμμικής παλινδρόμησης (linear regression), χρησιμοποιώντας τις ημερήσιες βραχυπρόθεσμες συνιστώσες της θερμοκρασίας, των κατακρημνισμάτων (άθροισμα τεσσάρων συνεχόμενων ημερών) και του επίπεδου των υπόγειων υδάτων. Τα βραχυπρόθεσμα συστατικά των παραπάνω μεταβλητών συμβολίζονται με $T_{SH}(t)$, $PR_{SH}(t)$, $PR4_{SH}(t)$ και $GWL_{SH}(t)$ αντίστοιχα. Ο Πίνακας 5 δείχνει τη συσχέτιση μεταξύ των βραχυπρόθεσμων συστατικών των μεταβλητών για τη θερινή περίοδο (Μάιος έως Σεπτέμβριος).

Πίνακας 5: Η συσχέτιση μεταξύ του βραχυπρόθεσμου συστατικού της απορροής του νερού, των κλιματικών μεταβλητών και του επιπέδου των υπόγειων υδάτων του αρχικού μοντέλου για την θερινή περίοδο (Tsakiri et al. 2014).

Table 5: Correlation matrix between the short term components of the water discharge, the climatic variables, and the groundwater level for the summer period.

	Water discharge	Groundwater level	Temperature average (4 days)	Precipitation	Precipitation (4 days)
Water discharge	1	-0.379	-0.247	0.124	0.502
Groundwater level	-0.379	1	0.099	0.194	-0.069
Temperature average (4 days)	-0.247	0.099	1	0.004	-0.249
Precipitation	0.124	0.194	0.004	1	-0.004
Precipitation (4 days)	0.502	-0.069	-0.249	-0.004	1

Η γραμμική παλινδρόμηση που πραγματοποιήθηκε για την πρόβλεψη της βραχυπρόθεσμης συνιστώσας της απορροής του νερού για την θερινή περίοδο, χρησιμοποιώντας τα βραχυπρόθεσμα συστατικά των κλιματικών μεταβλητών και της στάθμης των υπόγειων υδάτων, έγινε με συντελεστή συσχέτισης R^2 ίσο με 0.447. Συγκεκριμένα, η βραχυπρόθεσμη συνιστώσα της απορροής του νερού μπορεί να εκφραστεί, μέσω μιας γραμμικής παλινδρόμησης, ως εξής:

$$\begin{aligned} \text{WDS}_h(t) = & -0.097\text{TSh}(t) - 0.377\text{GWLS}_h(t) + 0.199\text{PRSh}(t) \\ & + 0.453\text{PR4Sh}(t) + \varepsilon(t), \end{aligned} \quad (11)$$

$$R^2 = 0.447.$$

Οι τιμές P , για τους συντελεστές των μεταβλητών του μοντέλου της γραμμικής παλινδρόμησης, είναι ίσες με το μηδέν.

2.7.2. Μοντέλο χειμερινής περιόδου

Όπως προαναφέρθηκε, πλημμύρες στα ποτάμια της Νέας Υόρκης, κατά τη διάρκεια του χειμώνα, συμβαίνουν κυρίως εξαιτίας της γρήγορης τήξης των πάγων, λόγω της αύξησης της θερμοκρασίας του αέρα. Γι' αυτό, τόσο στο αρχικό μοντέλο, όσο και στην παρούσα εργασία, χρησιμοποιείται η μέση θερμοκρασία στην γραμμική παλινδρόμηση του μοντέλου πρόβλεψης. Σύμφωνα με τους Tsakiri et al., μέγιστη συσχέτιση μεταξύ της απορροής του νερού και της μέσης θερμοκρασίας έχουμε όταν λαμβάνουμε την μέση θερμοκρασία για πάνω από τέσσερις ημέρες (όπως για τα κατακρημνίσματα). Επιπλέον αυτής, το μοντέλο της χειμερινής περιόδου χρησιμοποιεί τις ακόλουθες μεταβλητές: την παλίρροια, την ταχύτητα του ανέμου, το επίπεδο των υπόγειων υδάτων και το άθροισμα κατακρημνισμάτων τεσσάρων ημερών. Οι βραχυπρόθεσμες συνιστώσες που αναφέρθηκαν παραπάνω συμβολίζονται με $\text{T4}_{\text{Sh}}(t)$, $\text{WS}_{\text{Sh}}(t)$, $\text{PR4}_{\text{Sh}}(t)$ και $\text{GWL}_{\text{Sh}}(t)$ αντίστοιχα. Ο Πίνακας 6 δείχνει το μοντέλο συσχέτισης για τα βραχυπρόθεσμα συστατικά των μεταβλητών.

Πίνακας 6: Η συσχέτιση μεταξύ του βραχυπρόθεσμου συστατικού της απορροής του νερού, των κλιματικών μεταβλητών και του επιπέδου των υπόγειων υδάτων του αρχικού μοντέλου για την χειμερινή περίοδο (Tsakiri et al., 2014)

Table 6: Correlation matrix between the short term components of the water discharge, the climatic variables, and the groundwater level for the winter period.

	Water discharge	Groundwater level	Temperature average (4 days)	Tide	Wind speed	Precipitation (4 days)
Water discharge	1	0.398	0.57	0.732	0.25	0.229
Groundwater level	0.398	1	0.408	0.597	0.12	-0.005
Temperature average (4 days)	0.57	0.408	1	0.444	0.136	0.038
Tide	0.732	0.597	0.444	1	0.181	0.117
Wind speed	0.25	0.12	0.136	0.181	1	0.155
Precipitation (4 days)	0.229	-0.005	0.038	0.117	0.155	1

Η πρόβλεψη της βραχυπρόθεσμης συνιστώσας της απορροής του νερού για την χειμερινή περίοδο, πραγματοποιήθηκε με μια γραμμική παλινδρόμηση, χρησιμοποιώντας τα βραχυπρόθεσμα συστατικά των κλιματικών μεταβλητών και της στάθμης των υπόγειων υδάτων. Η γραμμική παλινδρόμηση που πραγματοποιήθηκε έδωσε συντελεστή συσχέτισης R^2 ίσο με 0.719. Συγκεκριμένα, η βραχυπρόθεσμη συνιστώσα της απορροής του νερού μπορεί να εκφραστεί ως εξής:

$$WD_{Sh}(t) = 0.323T4_{Sh}(t) + 0.629TD_{Sh}(t) + 0.087WS_{Sh}(t) - 0.117GWL_{Sh}(t) + 0.130PR4_{Sh}(t) + \varepsilon(t), \quad (12)$$

$$R^2 = 0.719.$$

Οι μεταβλητές που ελήφθησαν υπόψη για την πρόβλεψη της βραχυπρόθεσμης συνιστώσας της απορροής του νερού κατά τη χειμερινή περίοδο είναι διαφορετικές από εκείνες που χρησιμοποιήθηκαν την καλοκαιρινή περίοδο. Επομένως, ο διαχωρισμός των κλιμάκων των χρονοσειρών είναι απαραίτητος δεδομένου ότι, διαφορετικά συστατικά σχετίζονται με διαφορετικά φυσικά φαινόμενα (όπως το ταχύ λιώσιμο του χιονιού και φαινόμενα ice jam κατά τη χειμερινή περίοδο και εκτεταμένες βροχοπτώσεις ή τροπικές καταιγίδες κατά τη καλοκαιρινή περίοδο). Οι τιμές P για τους συντελεστές των μεταβλητών του μοντέλου γραμμικής παλινδρόμησης είναι ίσες με το μηδέν.

Κεφάλαιο 3.

Αποτελέσματα

Ένας ποταμός, όπως είναι αντιληπτό, μπορεί είτε να κερδίζει είτε να χάνει νερό από το σύστημα υπόγειων υδάτων. Το ύψος του υδροφόρου ορίζοντα αντανakλά μια ισορροπία μεταξύ του ρυθμού αναπλήρωσης μέσω των κατακρημνισμάτων, και της απομάκρυνσης μέσω της απορροής. Εν ολίγη, κάθε ανισορροπία αυξάνει ή μειώνει το επίπεδο του υδροφόρου ορίζοντα, ενεργώντας με το αντίθετο αποτέλεσμα στην τιμή στάθμης των υπόγειων υδάτων (*η στάθμη των υπόγειων υδάτων μετράται ως η απόσταση του βάθους της στάθμης του νερού κάτω από την επιφάνεια*).

Κατά την θερινή περίοδο, η απορροή του ποταμού του Schoharie Creek αυξάνει καθώς ρέει προς τα κατάντη, και οι παραπόταμοι και τα υπόγεια ύδατα του παρέχουν επιπλέον νερό (υπενθυμίζεται ότι οκτώ παραπόταμοι συμβάλλουν στον Schoharie Creek πριν φτάσει στο Mohawk River). Κατά τη διάρκεια του καλοκαιριού, οι υπόγειοι υδάτινοι πόροι φαίνεται να αυξάνονται με αφθονία (επαναφόρτιση). Η φυσική αυτή αναπλήρωση θα μειώσει την μετρούμενη τιμή της στάθμης των υπόγειων υδάτων και θα ενισχύσει την απορροή του νερού του ποταμού. Ως εκ τούτου, παράγεται μια αρνητική συσχέτιση.

Κατά την χειμερινή περίοδο, η αναπλήρωση του υδροφόρου ορίζοντα από τις βροχοπτώσεις είναι συχνά ελλιπής και το ποσοστό της είναι μικρότερο από την απορροή του νερού του ποταμού. Το επίπεδο του υδροφόρου ορίζοντα αναμένεται λοιπόν να πέσει και μπορεί να οδηγηθεί σε πολύ χαμηλή στάθμη νερού (εξάντληση). Η τιμή της στάθμης των υπόγειων υδάτων θα αυξηθεί (το επίπεδο του υδροφόρου μειώνεται) και η απορροή του νερού επίσης (κυρίως μέσα από το λιώσιμο των πάγων για την παρούσα περιοχή μελέτης). Αυτό παράγει μια θετική συσχέτιση.

Λόγω του διαχωρισμού των κλιμάκων των χρονοσειρών, η μακροπρόθεσμη συνιστώσα της απορροής του νερού παρουσιάζει αρνητική συσχέτιση με τη μακροπρόθεσμη συνιστώσα της στάθμης των υπόγειων υδάτων και τη θερμοκρασία (Πίνακας 3), ενώ η συσχέτιση μεταξύ των εποχιακών συστατικών αυτών των μεταβλητών είναι θετική (Πίνακας 4). Επίσης, έχει παρατηρηθεί μια θετική συσχέτιση μεταξύ των βραχυπρόθεσμων συστατικών αυτών των μεταβλητών κατά τη διάρκεια του χειμώνα (Πίνακας 6), ενώ υπάρχει μία αρνητική συσχέτιση κατά τη θερινή περίοδο (Πίνακας 5).

Η αυξημένη βροχόπτωση κατά τη διάρκεια της τελευταίας δεκαετίας στην περιοχή μελέτης του αρχικού μοντέλου (Schoharie Creek) εξηγεί την αρνητική συσχέτιση μεταξύ των μακροπρόθεσμων συστατικών της απορροής του νερού και του επιπέδου των υπόγειων υδάτων. Το φαινόμενο αυτό παρατηρείται και στα βραχυπρόθεσμα συστατικά, κατά τη διάρκεια της θερινής περιόδου, όταν οι βροχοπτώσεις είναι έντονες. Αντίθετα, θετική συσχέτιση μεταξύ

της απορροής του νερού και της στάθμης των υπόγειων υδάτων παρατηρείται κατά τη διάρκεια της παρατεταμένης χειμερινής περιόδου.

Η συσχέτιση μεταξύ της βραχυπρόθεσμης συνιστώσας της απορροής του νερού και των κατακρημνισμάτων, κατά τη διάρκεια του χειμώνα, είναι χαμηλότερη από ότι κατά τη διάρκεια του καλοκαιριού. Σύμφωνα με τους Tsakiri et al., αυτό οφείλεται στο γεγονός ότι στις περιοχές που παρατηρείται παρατεταμένη χειμερινή περίοδος, οι βροχοπτώσεις δεν συμβάλλουν στον υδροφόρο ορίζοντα με τον ίδιο ρυθμό που συμβάλλουν κατά τη διάρκεια της θερινής περιόδου.

Σύμφωνα με το αρχικό μοντέλο, η συνολική εξήγηση της εκτίμησης της απορροής του νερού, κατά την θερινή περίοδο, έγινε με τον συνδυασμό των σχέσεων 5, 9, και 11 για την παρουσίαση των εποχιακών, μακροπρόθεσμων και βραχυπρόθεσμων συστατικών (Πίνακας 7). Η συμβολή της μακροπρόθεσμης συνιστώσας των κλιματικών μεταβλητών και της στάθμης των υπόγειων υδάτων στην χρονοσειρά της απορροής του νερού είναι 49% (0.588×0.833) όπως εκφράζεται από την σχέση 5. Αντίστοιχα, η παρούσα ροή εργασίας, με την χρήση των καινούργιων μεταβλητών (αναπροσαρμογή της σχέσης 5, με την ίδια λογική με το αρχικό μοντέλο, για την ενσωμάτωση των νέων μεταβλητών) έδωσε 46,2%, για την ίδια περιοχή μελέτης. Σημειώνεται ότι, η αναπροσαρμογή των σχέσεων έγινε εκεί όπου διαπιστώθηκε συντελεστής συσχέτισης (R^2) μεταβλητών με αξιοσημείωτη τιμή ώστε η μεταβλητή να συμπεριληφθεί στη σχέση (όπως για παραδείγματα για την πρόβλεψη των μακροπρόθεσμων συστατικών της απορροής του νερού)

Επιπλέον, χρησιμοποιώντας την σχέση 9 η συνεισφορά της εποχιακής συνιστώσας είναι 16,2% (0.178×0.912), ενώ η συνεισφορά του βραχυπρόθεσμου συστατικού είναι 10,4% (0.232×0.447) όπως εκφράζεται από την σχέση 11. Στην παρούσα ροή εργασίας είχαμε συνεισφορά 15,4% και 9,8% αντίστοιχα.

Ο συνδυασμός των σχέσεων 5, 9 και 11 κατά τρόπο παρόμοιο με την σχέση 1, στο αρχικό μοντέλο, εξάγει το 75,6% της συνολικής διακύμανσης της χρονοσειράς της απορροής του νερού, με τη χρήση των κλιματικών μεταβλητών και το επίπεδο των υπόγειων υδάτων, κατά τη θερινή περίοδο. Αντίστοιχα, η παρούσα ροή εργασίας, ενώ έδωσε το ίδιο ποσοστό (με ακρίβεια πρώτου δεκαδικού) για την αρχική περιοχή μελέτης με τις ίδιες μεταβλητές (για λογούς επαλήθευσης), με τις καινούργιες μεταβλητές έδωσε 69,8%.

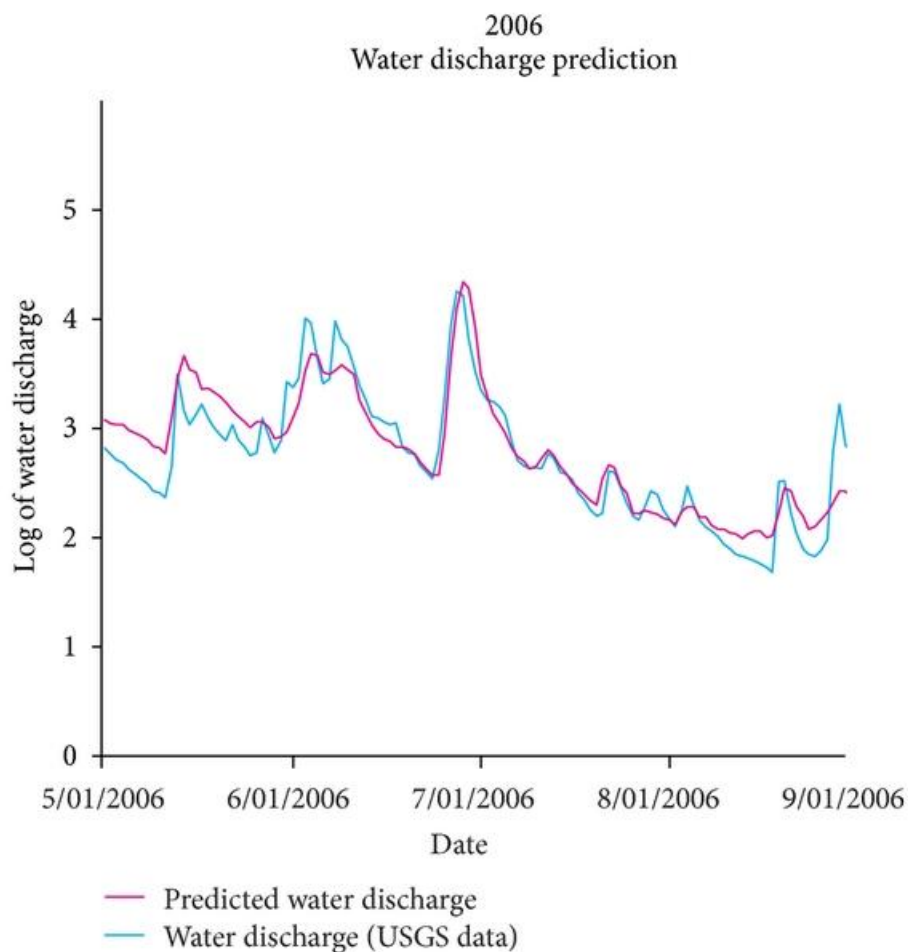
Το Σχήμα 4 δείχνει τα ακατέργαστα δεδομένα των χρονοσειρών της απορροής του νερού (μπλε γραμμή) μαζί με αυτά του μοντέλου πρόβλεψης (μωβ γραμμή), που προέρχονται από τις σχέσεις 5, 9 και 11 για το 2006 (όπου παρατηρήθηκαν καλοκαιρινές πλημμύρες στην 29^η Ιουνίου στις λεκάνες Mohawk, Hudson και Catskills). Αντίστοιχα, εξετάστηκαν παρόμοια γραφήματα για τα άλλα έτη. Σε αυτό το σημείο αξίζει να σημειωθεί πως η προβολή οποιοδήποτε δεδομένων σε οποιαδήποτε σημείο, της ροής εργασίας μας, καθώς και συνδυασμός αυτών, πραγματοποιείται πολύ εύκολα μέσα στο πρόγραμμα data mining, Knime. Επίσης, μπορούν να προβληθούν με οποιαδήποτε μορφή όπως ιστόγραμμα, γράφημα, διάγραμμα διασποράς, διάγραμμα πίτας και πολλά άλλα. Σε περίπτωση που θέλουμε να προβάλουμε αρχικά δεδομένα με τα αποτελέσματα ενός μοντέλου (όπως για σύγκριση) μπορούμε πολύ απλά να πάρουμε τα δεδομένα από τις εξόδους των εκάστου κόμβων (κάθε κόμβος έχει ένα τρίγωνο στο τέλος του (έξοδος, output), εκτός αυτών που είναι για προβολή ή σύνδεση με εξωτερική πηγή, που από

εκεί τραβώντας μια γραμμή συνδέουμε την έξοδο του με οποία είσοδο (input) θέλουμε) και να τα συνδυάσουμε με έναν κόμβο ζεύξης (joiner node). Μετά μπορούμε να τα προβάλουμε.

Πίνακας 7: Η περίληψη των διακυμάνσεων και του συντελεστή συσχέτισης των εποχιακών, μακροπρόθεσμων και βραχυπρόθεσμων συστατικών για την θερινή και την χειμερινή περίοδο του αρχικού μοντέλου (Tsakiri et al. 2014)

Table 7: Summary of variances and the coefficient of determinations for the seasonal and long and short term component for the summer and winter period.

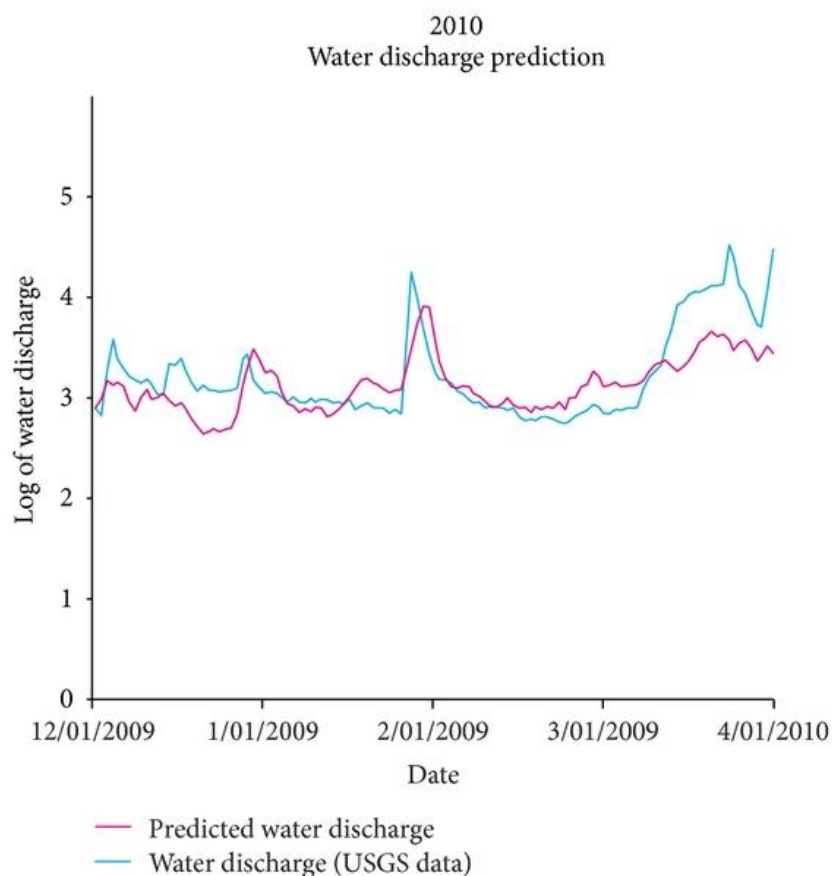
	Long term	Seasonal term	Short term
Summer period			
Variance	0.588	0.178	0.232
R^2	0.833	0.912	0.447
Winter period			
Variance	0.531	0.162	0.307
R^2	0.833	0.912	0.719



Σχήμα 4: Τα ακατέργαστα δεδομένα των χρονοσειρών της απορροής του νερού (μπλε γραμμή) μαζί με αυτά του αρχικού μοντέλου πρόβλεψης των Tsakiri et al. (μωβ γραμμή) για την θερινή περίοδο του 2006 (Tsakiri et al., 2014)

Το αρχικό μοντέλο, με το συνδυασμό των σχέσεων 5, 9, και 12, εκτίμησε τη συνολική εξήγηση της χρονοσειράς της απορροής του νερού από τις κλιματικές μεταβλητές για τη χειμερινή περίοδο (Πίνακας 7). Συγκεκριμένα, η συμβολή των μακροπρόθεσμων συστατικών των κλιματικών μεταβλητών και της στάθμης των υπογείων υδάτων με την χρονοσειρά της απορροής του νερού είναι 44,2% (0.531×0.833), όπως περιγράφεται στην σχέση 5. 14,8% (0.162×0.912) είναι η συμβολή των εποχιακών συστατικών των κλιματικών μεταβλητών, όπως περιγράφεται στην σχέση 9, ενώ 22,1% (0.307×0.719) είναι η συμβολή των βραχυπρόθεσμων συστατικών, όπως περιγράφεται από την σχέση 12. Κατά συνέπεια, η συνολική εξήγηση της χρονοσειράς της απορροής του νερού χρησιμοποιώντας τις κλιματικές μεταβλητές και την στάθμη των υπογείων υδάτων κατά τη χειμερινή περίοδο είναι 81,1%. Η παρούσα ροή εργασίας έδωσε 39.4%, 14.2% και 20.8% αντίστοιχα ποσοστά συμβολής, ως αποτέλεσμα να εξηγείται το 74,4% της χρονοσειράς της απορροής του νερού από τις επιλεγμένες μεταβλητές.

Ως παράδειγμα του αρχικού μοντέλου, το Σχήμα 5 δείχνει τα ακατέργαστα δεδομένα της απορροής του νερού (μπλε γραμμή) για το έτος 2010, μαζί με αυτά του μοντέλου πρόβλεψης (μωβ γραμμή), που προέρχονται από τις σχέσεις 5, 9 και 12. Η χρονοσειρά του μοντέλου πρόβλεψης αντιπροσωπεύει την εκδήλωση του φαινομένου του λιωσίματος πάγου (φαινόμενο ice jam) του 2010 (όπου σημειώθηκαν εκτεταμένες πλημμύρες στις 25 - 26 Ιανουαρίου σε ολόκληρη την ανατολική κεντρική Νέα Υόρκη και την γειτονική δυτική Νέα Αγγλία, από συνδυασμό βροχής, λιώσιμο χιονιού και παγωμένου εδάφους).



Σχήμα 5: Ακατέργαστα δεδομένα της απορροής του νερού (μπλε γραμμή) μαζί με αυτά του αρχικού μοντέλου πρόβλεψης των Tsakiri et al. (μωβ γραμμή) για την χειμερινή περίοδο του 2010 (Tsakiri et al., 2014)

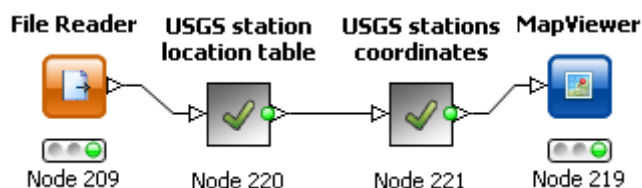
Με βάση τα αποτελέσματα του θερινού και του χειμερινού αρχικού μοντέλου, αποδεικνύεται ότι η αποσύνθεση της χρονοσειράς βελτιώνει την ικανότητά μας να περιγράψουμε και να προβλέψουμε τις μεταβολές στις χρονοσειρές της απορροής του νερού κατά περίπου δύο φορές. Ειδικότερα, η ανεξήγητη διακύμανση που προέρχεται από τα ανεπεξέργαστα δεδομένα που περιγράφονται στην σχέση 4 είναι 41 %, ενώ η ανεξήγητη διακύμανση που προέρχεται από την μέθοδο αποσύνθεσης της απορροής του νερού κατά την περίοδο του χειμώνα είναι 19 %. Παρόμοια η ανεξήγητη διακύμανση που μπορεί να προκύψει για τη θερινή περίοδο ανέρχεται στο 24,4 %. Αντίστοιχα, στην παρούσα εργασία, με την χρήση των νέων μεταβλητών, είχαμε ανεξήγητη διακύμανση από τα ανεπεξέργαστα δεδομένα 43,8%, 25,6% από την αποσύνθεση της απορροής του νερού κατά την περίοδο του χειμώνα και 30,2% κατά την περίοδο του καλοκαιριού.

Η ενσωμάτωση του αρχικού μοντέλου (όμοιοι σταθμοί με όμοιες μεταβλητές) πάνω στην παρούσα ροή εργασίας επαλήθευσε την ορθότητα λειτουργίας της. Εν συνεχεία, έγινε η χρήση των νέων μεταβλητών σαν μια αναβάθμιση του αρχικού μοντέλου. Από το σημείο αυτό και μετά η ροή εργασίας μας περιελάμβανε το μοντέλο και ήταν έτοιμη να δοκιμαστεί και σε άλλες περιοχές με ευκολία. Ο στόχος αυτός επετεύχθη, προβλέποντας χρονοσειρές οι οποίες επαληθευτήκαν με πραγματικά σημαντικά γεγονότα (πλημμυρικά συμβάντα). Μετά από αυτό το σημείο, θεωρώντας ότι η ροή εργασίας μας ενσωματώνει το μοντέλο των Tsakiri et al., δοκιμάστηκε σε βρόγχο (loop), για την επαλήθευση της ενσωμάτωσης των νέων καθημερινών δεδομένων των πηγών. Σημειώνεται ότι, η προσθήκη νέων δεδομένων γίνεται σε πραγματικό χρόνο από την στιγμή που οι πηγές ανεβάζουν στην βάση δεδομένων τους τα δεδομένα. Ωστόσο, υπήρχαν περιπτώσεις που η ροή εργασίας μας διακοπτόταν. Μια τέτοια, όπως προαναφέρθηκε, είναι η περίπτωση που οι πηγές παρέχουν χαμένα δεδομένα (missing values).

Κεφάλαιο 4.

Συζήτηση

Η παρούσα μελέτη επικεντρώθηκε στην πρόβλεψη της ημερησίας χρονοσειράς της απορροής του νερού με την χρήση λογισμικού data mining, για την διαχείριση των διαθέσιμων μεταβλητών σε πραγματικό χρόνο, για την πρόβλεψη πλημμύρων. Απώτερος σκοπός η χρήση σε συστήματα έγκαιρης προειδοποίησης. Σύμφωνα με την παρούσα εργασία όλα συνηγορούν πως κάτι τέτοιο είναι εφικτό. Επίσης, σαν μελλοντικός στόχος ορίζεται και η ενσωμάτωση σταθμών και κατά συνέπεια δεδομένων από μια μεγάλη περιοχή, όπως η Νέα Υόρκη. Αυτό μπορεί να επιτευχθεί αφού πρώτα γίνει χρήση των συντεταγμένων των σταθμών από το Knime ώστε να προβληθούν σε έναν παγκόσμιο χάρτη (παραδείγματα ροη εργασίας φαίνεται στο Σχήμα 6.



Σχήμα 6: Η ροη εργασίας για την προβολή των σταθμών σε παγκόσμιο χάρτη. Όπως και στην αρχική ροη εργασίας μας (Σχήμα 1), ο πρώτος κόμβος αποτελεί την είσοδο, οι δυο επόμενοι την διαχείριση δεδομένων και ο τέταρτος την έξοδο (τροποποιημένο από Καβαλιέρος Α., 2016, Knime).

Εν συνεχεία, πρέπει να συνδυαστεί η αρχική ροη εργασίας για κάθε έναν σταθμό με την ροη εργασίας της προβολής των σταθμών σε παγκόσμιο χάρτη. Μέσα στον τρίτο μετα-κόμβο (node 221) του Σχήματος 6, εμπεριέχεται το πολυγωνικό φίλτρο (Εικόνα 1).

Το μοντέλο στο οποίο βασίζεται η παρούσα εργασία έχει αποδείξει ότι η αποσύνθεση των χρονοσειρών είναι απαραίτητη λόγω της παρουσίας των βραχυπρόθεσμων διακυμάνσεων της χρονοσειράς. Επίσης, έχει αποδειχθεί ότι οι εποχιακές, μακροπρόθεσμες και βραχυπρόθεσμες συνιστώσες της απορροής του νερού μπορούν να εξηγηθούν μέσα από την εξέταση κλιματικών μεταβλητών και το επίπεδο των υπόγειων υδάτων.

Τα αποτελέσματα δείχνουν ότι καθώς αυξάνεται η απορροή του νερού μακροπρόθεσμα ο υδροφόρος ορίζοντας αναπληρώνεται, ενώ εποχιακά εξαντλείται. Σε βραχυπρόθεσμο επίπεδο, τα υπόγεια ύδατα μειώνονται κατά τη διάρκεια της χειμερινής περιόδου και αυξάνονται κατά τη θερινή περίοδο. Η βραχυπρόθεσμη συνιστώσα της απορροής του νερού σχετίζεται με καιρικές διακυμάνσεις και βραχυπρόθεσμα φαινόμενα όπως η βροχή, οι καταιγίδες και οι κυκλώνες κατά τη θερινή περίοδο και ταχεία αύξηση της θερμοκρασίας και καταιγίδες κατά τη διάρκεια της χειμερινής περιόδου.

Η παρούσα εργασία χρησιμοποιεί το μοντέλο των Tsakiri et al καθώς μπορεί να προβλέψει τα κρίσιμα επεισόδια της απορροής του νερού με την προϋπόθεση οι άλλες μεταβλητές να ληφθούν πριν από την εκδήλωση του γεγονότος (έστω και για λίγα λεπτά, όπως στα συστήματα έγκαιρης προειδοποίησης τσουνάμι).

Τα λογισμικά data mining και η πληθώρα διαθέσιμων δεδομένων, αποτέλεσαν πρωταρχικό λόγο για την σύνταξη της παρούσας εργασίας. Η δυνατότητα που παρέχεται από τα εν λόγω λογισμικά, για την διαχείριση τεραστίων όγκων δεδομένων σε συνδυασμό με την ανάκτηση τους σε πραγματικό χρόνο και η ευκολία προσαρμοστικότητας τους, αποτέλεσαν σημεία κλειδιά. Ωστόσο, το παρόν εγχείρημα αποτελεί μια θεωρητική προσέγγιση.

Συμπεράσματα

Η πρόβλεψη της ημερήσιας χρονοσειράς της απορροής του νερού, χρησιμοποιώντας τις κλιματικές μεταβλητές και το επίπεδο των υπόγειων υδάτων μπορεί να βελτιωθεί σημαντικά μέσω της αποσύνθεσης των χρονοσειρών. Η αποσύνθεση των διαφόρων συστατικών (κλίμακες), όπως τα εποχιακά, τα μακροπρόθεσμα και τα βραχυπρόθεσμα συστατικά, μπορεί να πραγματοποιηθεί με την χρήση λογισμικού data mining για την αποφυγή εσφαλμένων αποτελεσμάτων και τον διπλασιασμό της ακρίβειας πρόβλεψης των χρονοσειρών της απορροής του νερού σε σχέση με τα ανεπεξέργαστα δεδομένα. Το λογισμικό data mining, Knime, μπορεί να ενσωματώσει τα μοντέλα του καλοκαιριού και του χειμώνα ώστε να βελτιώσει την πρόβλεψη των πλημμυρών που προκλήθηκαν από τις καταιγίδες και το ταχύ λιώσιμο του χιονιού.

Η καλύτερη κατανόηση των πλημμυρών μπορεί να μειώσει την πιθανότητα δαπανηρών ζημιών που συνδέονται με τους εν λόγω κινδύνους. Για το λόγο αυτό, είναι σημαντικό να μπορούμε να σχεδιάσουμε ένα δυναμικό μοντέλο που μπορεί να αλλάξει σε πραγματικό χρόνο, χρησιμοποιώντας νέες παρατηρήσεις από διαφορετικούς σταθμούς. Ένα τέτοιο μοντέλο μπορεί να σχεδιαστεί με τη χρήση λογισμικού data mining και στατιστικών τεχνικών. Ένα τέτοιο δυναμικό μοντέλο παρέχει τη δυνατότητα της άμεσης λήψης αποφάσεων και την αποτελεσματική ανταπόκριση των τοπικών αρχών (π.χ. πυροσβεστική, αστυνομία, νοσοκομεία, κλπ) σε περίπτωση πλημμύρας. Πιο συγκεκριμένα, η ανάλυση της αποσύνθεσης χρονοσειρών και πολυμεταβλητών τεχνικών (μοντέλα πολλών μεταβλητών) μας επιτρέπουν να βελτιώσουμε την πρόβλεψη των πλημμυρών (Marsellos et al, 2016).

Και μην ξεχνάμε ότι τα δεδομένα βρίσκονται σε περίσσια, αρκεί να τα διαχειριστούμε. Η μαζικότητα στην σκέψη του επιστήμονα του 21^{ου} αιώνα μπορεί να φαντάζει υπερβολική αλλά δεν θα έπρεπε να αποτελεί εξαίρεση.

Βιβλιογραφία

1. Beltaos, S., 1997. "Onset of river ice breakup," *Cold Regions Science and Technology*, vol. 25, no. 3, pp. 183–196.
2. Brohan, P., Kennedy, J., Harris, I., Tett, S. F. B., and Jones, P. D., 2006. "Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850," *Journal of Geophysical Research*, vol. 11, no. 12.
3. Bronstert, A., 1995. "River flooding in Germany: influenced by climate change?" *Physics and Chemistry of the Earth*, vol. 20, no. 5-6, pp. 445–450.
4. Close, B., and Zurbenko, I., 2013. "Kolmogorov-Zurbenko adaptive filters," version 3, R-software, <http://cran.r-project.org/web/packages/kza/index.html>.
5. DiRienzo, A. G., and Zurbenko, I. G., 1999. "Semi-adaptive nonparametric spectral estimation," *Journal of Computational and Graphical Statistics*, vol. 8, no. 1, pp. 41–59.
6. Eskridge, R. E., Ku, J. Y., Trivikrama Rao, S., Steven Porter, P., and Zurbenko, I. G., 1997. "Separating different scales of motion in time series of meteorological variables," *Bulletin of the American Meteorological Society*, vol. 78, no. 7, pp. 1473–1483.
7. Garver, J. I., and Cockburn, J. M. H., 2009. "A historical perspective of ice jams on the lower Mohawk River," *Proceedings of the Mohawk Symposium*, pp. 25–29.
8. Hartmann, H., and Andresky, L., 2013. "Flooding in the Indus river basin-A spatiotemporal analysis of precipitation records," *Global and Planetary Changes*, vol. 107, pp. 25–35.
9. Hogrefe, C., Rao, S. T., Zurbenko, I. G., and Porter, P. S., 2000. "Interpreting the information in ozone observations and model predictions relevant to regulatory policies in the Eastern United States," *Bulletin of the American Meteorological Society*, vol. 81, no. 9, pp. 2083–2106.
10. Johnston, S. A., and Garver, J. I., 2010. "Record of flooding on the Mohawk River from 1634 to 2000 based on historical archives," *Geological Society of America Abstracts with Programs*, vol. 33, no. 1.
11. Kern, A. L., 2008. *Study of 20th century trends in stream flow for West Canada and schoharie creeks of the Mohawk-Hudson Rivers watershed [Ph.D. thesis]*, Department of Geology, Union College, Schenectady, NY, USA.
12. Kotlarski, S., Hagamann, S., Krahe, P., Podzun, R., and Jacob, D., 2012. "The Elbe river flooding 2002 as seen by an extended regional climate model," *Journal of Hydrology*, vol. 472-473, pp. 169–183.
13. Marsellos, A. E., Garver, J. I., Cockburn, J. M. H., and Tsakiri, K., 2010. "Determination of historical channel changes and meander cut-off points using LiDAR and GIS in Schoharie creek, NY," *Proceedings of the Mohawk Symposium*, pp. 28–33, Schenectady, NY, USA.

14. Marsellos, A. E., and Garver, J. I., 2010. "Channel incision and landslides identified by LiDAR in the lower reaches of Schoharie creek, New York," *Geological Society of America Abstracts with Programs*, vol. 42, no. 1, p. 73.
15. Marsellos, A. E., Tsakiri, K. G., Kavalieros, A., 2016. "Data Mining for Immediate Decision Making in Flood Hazard Events: An application at Mohawk Watershed in New York," *Mohawk Watershed Symposium, College Park, Union College, Schenectady NY*.
16. OECD, 2005. "Section 4: guidelines for the reporting of different forms of data," in *Data and Metadata Reporting and Presentation Handbook*, OECD, Paris, France.
17. Rao, S. T., and Zurbenko, I. G., 1994. "Detecting and tracking changes in ozone air quality," *Journal of the Air and Waste Management Association*, vol. 44, no. 9, pp. 1089–1092.
18. Rao, S. T., Zurbenko, I. G., Neagu, R., Porter, P. S., Ku, J. Y., and Henry, R. F., 1997. "Space and time scales in ambient ozone data," *Bulletin of the American Meteorological Society*, vol. 78, no. 10, pp. 2153–2166.
19. Smith, T. M., and Reynolds, R. W., 2005. "A global merged land-air-sea surface temperature reconstruction based on historical observations (1880–1997)," *Journal of Climate*, vol. 18, no. 12, pp. 2021–2036.
20. Trenberth, K. E., Fasullo, J., and Smith, L., 2005. "Trends and variability in column-integrated atmospheric water vapor," *Climate Dynamics*, vol. 24, no. 7-8, pp. 741–758.
21. Trenberth, K. E., 2011. "Changes in precipitation with climate change," *Climate Research*, vol. 47, no. 1-2, pp. 123–138.
22. Tsakiri, K. G., and Zurbenko, I. G., 2008. "Destructive effect of the noise in principal component analysis with application to ozone pollution," *Proceedings of the Joint Statistical Meetings*, pp. 3054–3068.
23. Tsakiri, K. G., and Zurbenko, I. G., 2009. "Model prediction of ambient ozone concentrations," *Proceedings of the Joint Statistical Meetings*, pp. 3054–3068.
24. Tsakiri, K., and Zurbenko, I. G., 2011. "Effect of noise in principal component analysis," *Journal of Statistics and Mathematics*, vol. 2, no. 2, pp. 40–48.
25. Tsakiri, K. G., and Zurbenko, I. G., 2010. "Determining the main atmospheric factor on ozone concentrations," *Journal of Meteorology and Atmospheric Physics*, vol. 109, no. 3-4, pp. 129–137.
26. Tsakiri, K. G., and Zurbenko, I. G., 2011. "Prediction of ozone concentrations using atmospheric variables," *Journal of Air Quality, Atmosphere, and Health*, vol. 4, no. 2, pp. 111–120.
27. Tsakiri, K., and Zurbenko, I. G., 2013. "Explanation of fluctuation in water use time series," *Journal of Environmental and Ecological Statistics*, vol. 20, no. 3, pp. 399–412.
28. Tsakiri, K. G., Marsellos, A. E., and Zurbenko, I. G., 2014. "An efficient prediction model for water discharge in Schoharie Creek, NY," *Journal of Climatology*, vol. 2014, Article ID 284137, 10 pages.
29. Yang, W., and Zurbenko, I., 2010. "Kolmogorov-Zurbenko filters," *Wiley Interdisciplinary Reviews*, vol. 2, no. 3, pp. 340–351.
30. Yang, W., and Zurbenko, I., 2010. "Nonstationarity," *Wiley Interdisciplinary Reviews*, vol. 2, no. 1, pp. 107–115.

31. Zurbenko, I. G., 1986. *The Spectral Analysis of Time Series, Statistics and Probability*, North Holland, Amsterdam, The Netherlands.
32. Zurbenko, I. G., and Sowizral, M., 1999. "Resolution of the destructive effect of noise on linear regression of two time series," *Far East Journal of Theoretical Statistics*, vol. 3, pp. 139–157.
33. Κράνης, Χ., Σκούρτσος, Ε., Σούκης, Κ., 2015. Συστήματα του Γήινου Περιβάλλοντος και Φυσικοί Πόροι.
34. EM-DAT | The international disasters database, 2016, <http://www.emdat.be/> (last viewed 26/8/2016)
35. National Oceanic and Atmospheric Administration U.S. Department of Commerce, 2016, NOAA, <http://www.noaa.gov/>
36. United Nations International Strategy for Disaster Reduction, UNISDR, 2016, Basics of early warning, <http://www.unisdr.org/> (last viewed 26/8/2016)
37. US Geological Survey, USGS, 2016, <https://www.usgs.gov>
38. Weather Underground, 2016, [https://www.wunderground.com/Department of Environmental Conservation, New York State, Atlantic Ocean/Long Island Sound Watershed](https://www.wunderground.com/Department%20of%20Environmental%20Conservation,%20New%20York%20State,%20Atlantic%20Ocean/Long%20Island%20Sound%20Watershed), 2016, <http://www.dec.ny.gov>
39. Wikipedia, "Kolmogorov-Zurbenko Filters," http://en.wikipedia.org/wiki/Kolmogorov%E2%80%93Zurbenko_filter