

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΥΔΑΤΙΚΩΝ ΠΟΡΩΝ & ΠΕΡΙΒΑΛΛΟΝΤΟΣ



Διερεύνηση μεθόδων τεχνητής νοημοσύνης -
μηχανικής μάθησης σε υδρολογικές λεκάνες
των Η.Π.Α.

Διπλωματική εργασία

Γκουρνέλος Νικόλαος Αρτέμιος

Επιβλέπων: Μπαλτάς Ευάγγελος, Καθηγητής ΕΜΠ

Αθήνα, Μάρτιος 2024

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΥΔΑΤΙΚΩΝ ΠΟΡΩΝ & ΠΕΡΙΒΑΛΛΟΝΤΟΣ

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον καθηγητή μου Μπαλτά Ευάγγελο για την καθοδήγηση του. Τον υποψήφιο διδάκτορα Γουργουλέτη Νικόλαο για την πολύτιμη βοήθειά του αλλά και τους γονείς μου και φίλους για τη βοήθεια και την στήριξη τους

Περίληψη

Ο στόχος της παρούσας διπλωματικής είναι να μελετήσει την πρόγνωση των απορροών των λεκανών. Οι μεθοδολογίες που ακολουθήθηκαν ήταν από τη μία η πρόγνωση απορροής με δεδομένα απορροής, δηλαδή η πρόγνωση με μια μεταβλητή και από την άλλη η πρόγνωση της απορροής με δεδομένα όπως βροχή, θερμοκρασία και άλλα. Η δεύτερη μεθοδολογία καλείται στη διπλωματική ως προσομοίωση της λεκάνης για να μπορέσει να διακριθεί από την πρόγνωση της πρώτης μεθοδολογίας. Στη μέθοδο πρόγνωσης αναπτύχθηκαν μοντέλα που αφορούν την κλασσική ανάλυση χρονοσειρών. Δημιουργήθηκαν έτσι 20 μοντέλα πρόγνωσης εκ των οποίων τα 4 ήταν εκθετικής ομαλοποίησης και τα 16 ήταν ARIMA. Ο ορίζοντας πρόγνωσης για την πρώτη μεθοδολογία ορίστηκε να είναι οι 10, 5 και 1 ημέρες. Στον ορίζοντα των 10 και 5 ημερών χρησιμοποιήθηκαν τα μοντέλα που αναφέρθηκαν παραπάνω ενώ στον ορίζοντα της 1 ημέρας όπως και στο μοντέλο προσομοίωσης της λεκάνης εφαρμόστηκαν εργαλεία από τη μηχανική μάθηση και ειδικότερα από τη βαθιά μηχανική μάθηση. Τα δεδομένα είναι ημερήσια και έχουν αποκτηθεί από τη βάση δεδομένων CAMELS όπου αντλούνται από 10 υδρογραφικές λεκάνες των ΗΠΑ. Το μοντέλο βαθιάς μηχανικής μάθησης που αναπτύχθηκε για τον ορίζοντα πρόγνωσης 1 ημέρας χρησιμοποίησε τη προσέγγιση του κυλιόμενου παραθύρου. Το μοντέλο βαθιάς μηχανικής μάθησης για την προσομοίωση της λεκάνης έχει ως σκοπό να μιμηθεί ως ένα βαθμό υδρολογικά μοντέλα διαχείρισης νερού όπως το SWAT (Soil Water Assessment Tool) αλλά περιορίζεται μόνο στην εκτίμηση απορροής. Όλα τα μοντέλα αναπτύχθηκαν σε περιβάλλον Python. Κάθε μοντέλο εκπαιδεύτηκε ξεχωριστά για κάθε μια από τις 10 λεκάνες και υπολογίστηκαν δείκτες όπως ο Nash-Sutcliffe Efficiency και Absolute Nash-Sutcliffe Efficiency. Στον ορίζοντα των 10 και 5 ημερών τα μοντέλα κρίθηκαν μη ικανοποιητικά, στον ορίζοντα της 1 ημέρας τα μοντέλα κρίθηκαν από ικανοποιητικά έως εξαιρετικά με τη βαθιά μηχανική μάθηση να ξεπερνά τα κλασσικά μοντέλα. Τέλος, το μοντέλο προσομοίωσης λεκάνης πετυχαίνει ικανοποιητικά αποτελέσματα και τροποποιείται έτσι ώστε να μπορεί να μοντελοποιήσει λεκάνες με λιγότερα δεδομένα.

Abstract

The goal of this research is to perform time series analysis to study streamflow forecasting. To do so two kinds of methods have been used. The first one is a univariate method which builds models that use streamflow as input to return streamflow as output. The second one is a method that simulates the basin and predicts the streamflow taking as input hydrology and meteorology variables such as precipitation, minimum and maximum temperature, pressure, solar radiation and time that sun is available each day, this is a multi-variate model. To construct the univariate models we built classical time series analysis like exponential smoothing and ARIMA with the help of statsmodels but also we used machine learning methods and more specifically we built a deep learning multi layer perceptron neural network that operates under the rolling window structure. To construct the multi-variate model a multi layer perceptron neural network was used with the help of tensorflow. The analysis used daily data that were retrieved from CAMELS a hydrology dataset with US and Canada basins for large scale models. The time steps that were used for univariate forecasting were 10, 5 and 1 day. The deep learning model was applied only to the 1

day time step forecasting. To measure the performance of each model in each time step we used the Nash Sutcliffe Efficiency and a variation of this metric, the Absolute Nash Sutcliffe Efficiency. In order to evaluate a model as satisfying and thus to classify our models we took under consideration fellow research papers that came up with results after testing hydrology models like SWAT. For the 10 and 5 days time steps the models were not satisfactory. For the time step of 1 day the results depending on the basin were satisfactory, good or excellent. Through the results ARIMA was superior than exponential smoothing and deep learning was superior than ARIMA in the cases that the comparison was available. Furthermore, the basin simulation model gave satisfactory results if we compute the average NSE value from the 10 basins, in this point it is made clear that a neural network was used for each one of the basins training 10 MLPs in this section of the research. Finally, we used an approach to train a single neural network on many different basins to create a model that can predict the streamflow of nearly ungauged basins, using 2 years of daily data to perform forecasting.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

Περίληψη.....	3
1 Εισαγωγή.....	15
1.1 Οι χρονοσειρές.....	15
1.2 Χρονοσειρές και μοντέλα προσομοίωσης.....	15
1.3 Χρονικό διάστημα καταμέτρησης της χρονοσειράς.....	16
1.4 Σημασία πρόγνωσης σε βραχυπρόθεσμο ορίζοντα.....	17
1.4.1 Στον αγροτικό τομέα.....	17
1.4.2 Στον τομέα της ενέργειας.....	17
1.4.3 Η έρευνα για τη σημασία βραχυχρόνιας πρόγνωσης.....	20
1.5 Αντικείμενο της εργασίας.....	20
2 Δεδομένα.....	21
3 Θεωρητικό υπόβαθρο μεθόδων.....	32
3.1 Εκθετική ομαλοποίηση.....	32
3.1.1 Απλή εκθετική ομαλοποίηση.....	32
3.1.2 Διπλή εκθετική ομαλοποίηση.....	32
3.1.3 Τριπλή εκθετική ομαλοποίηση.....	33
3.2 Τα μοντέλα ARIMA.....	33
3.2.1 Τα επιμέρους AR, I, MA.....	34
3.2.2 Τέστ για την επιλογή των παραμέτρων p, d, q	36
3.2.3 Τελική επιλογή παραμέτρων.....	41
3.2.4 Όλα τα μοντέλα ARIMA που χρησιμοποιήθηκαν.....	42
3.2.5 Μέθοδος μεγίστης πιθανοφάνειας.....	46
3.3 Μέθοδοι βαθιάς μηχανικής μάθησης.....	47
3.3.1 Χρήσιμοι ορισμοί.....	47
3.3.2 Νευρωνικά δίκτυα.....	49
3.3.3 Διαδικασία ανανέωσης των τιμών των βαρών (gradient descent).....	54
3.3.4 Οι εξισώσεις του νευρωνικού δικτύου.....	57
3.3.5 Συναρτήσεις ενεργοποίησης.....	59

3.3.6	Ρυθμός μάθησης	61
3.3.7	Μεγεθος πακέτου.....	62
3.4	Οι δείκτες Nash-Sutcliffe Efficiency και Absolute Nash-Sutcliffe Efficiency 62	
4	Εφαρμογή σε προγραμματιστικό περιβάλλον Python	64
4.1	Εκθετική Ομαλοποίηση.....	64
4.1.1	Δεδομένα	64
4.1.2	Πρόγνωση απορροής.....	64
4.1.3	Μοντέλα που δημιουργήθηκαν.....	66
4.1.4	Εξαγωγή αποτελεσμάτων σχεδίαση διαγραμμάτων	66
4.2	ARIMA.....	66
4.2.1	Δεδομένα	66
4.2.2	Η επιλογή των παραμέτρων p , d , q και τα αντίστοιχα τεστ.	66
4.2.3	Πρόγνωση απορροής.....	77
4.2.4	Αρχικοποίηση παραμέτρων και δημιουργία μοντέλων	78
4.2.5	Υπολογισμός σφάλματος και γραφικές παραστάσεις.....	79
4.3	Μέθοδοι Deep Learning στην Python.....	79
4.3.1	Πρόγνωση με τη μέθοδο του ‘κυλιόμενου παραθύρου’	79
4.3.2	Προσομοίωση-μοντελοποίηση λεκάνης με τη μέθοδο νευρωνικού δικτύου	81
5	Αποτελέσματα	84
5.1	Ορίζοντας πρόγνωσης 10 ημερών	84
5.2	Ορίζοντας πρόγνωσης 5 ημερών	96
5.3	Ορίζοντας πρόγνωσης 1 ημέρας	107
5.4	Μοντέλο deep learning προσομοίωσης λεκάνης	121
5.5	Συζήτηση επί των αποτελεσμάτων	126
5.5.1	Id λεκάνης 02014000, Potts Creek	126
5.5.2	Id λεκάνης 01411300, Tuckahoe river	127
5.5.3	Id λεκάνης 01440000, Flatbrook river	127
5.5.4	Id λεκάνης 01543500, Sinnemahoning Creek.....	128
5.5.5	Id λεκάνης 02016000, Cowpasture river.....	128

5.5.6	Id λεκάνης 02053800, South Fork Roanoke river	129
5.5.7	Id λεκάνης 02059500, Goose creek	129
5.5.8	Id λεκάνης 02069700, South Mayo river	130
5.5.9	Id λεκάνης 02143000, Henry Fork, near Henry river	130
5.5.10	Id λεκάνης 02350900, Kinchafoone creek	130
6	Συμπεράσματα	132
6.1	Σύγκριση ARIMA και εκθετικής ομαλοποίησης.....	132
6.2	Τα αξιοσημείωτα μοντέλα κλασσικών μεθόδων	132
6.3	Σύγκριση ARIMA και deep learning στον ορίζοντα πρόγνωσης 1 ημέρας.....	132
6.4	Το μοντέλο deep learning προσομοίωσης λεκάνης	134
7	Προέκταση των μεθόδων deep learning και προτάσεις για μελλοντική έρευνα	136
7.1	Επέκταση της μεθόδου deep learning προσομοίωσης λεκάνης	136
7.2	Μελλοντικές εναλλακτικές προτάσεις για νευρωνικά δίκτυα.....	137
8	Βιβλιογραφία	137

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1 Χάρτης Potts creek από USGS	22
Εικόνα 2 Χρονοσειρά Ημερήσιας Απορροής Potts creek.....	22
Εικόνα 3 Χάρτης Tuckahoe river από USGS.....	23
Εικόνα 4 Χρονοσειρά Ημερήσιας Απορροής Tuckahoe river	23
Εικόνα 5 Χάρτης Flatbrook river από USGS	24
Εικόνα 6 Χρονοσειρά Ημερήσιας Απορροής Flatbrook river.....	24
Εικόνα 7 Χάρτης Sinnemahoning Creek από USGS.....	25
Εικόνα 8 Χρονοσειρά Ημερήσιας Απορροής Sinnemahoning Creek.....	25
Εικόνα 9 Χάρτης Cowpasture river από USGS.....	26
Εικόνα 10 Χρονοσειρά Ημερήσιας Απορροής Cowpasture river	26
Εικόνα 11 Χάρτης South Fork Roanoke river από USGS	27
Εικόνα 12 Χρονοσειρά Ημερήσιας Απορροής South Fork Roanoke river.....	27
Εικόνα 13 Χάρτης Goose creek από USGS	28
Εικόνα 14 Χρονοσειρά Ημερήσιας Απορροής Goose creek.....	28
Εικόνα 15 Χάρτης South Mayo River από USGS.....	29
Εικόνα 16 Χρονοσειρά Ημερήσιας Απορροής South Mayo River	29
Εικόνα 17 Χάρτης Henry fork από USGS	30
Εικόνα 18 Χρονοσειρά Ημερήσιας Απορροής Henry fork.....	30
Εικόνα 19 Χάρτης Kinchafoonee creek από USGS	31
Εικόνα 20 Χρονοσειρά Ημερήσιας Απορροής Kinchafoonee creek	31
Εικόνα 21 Διάγραμμα αυτοσυσχέτισης για την λεκάνη 1.....	68
Εικόνα 22 Διάγραμμα αυτοσυσχέτισης για τη λεκάνη 2	68
Εικόνα 23 Διάγραμμα αυτοσυσχέτισης για τη λεκάνη 3	69
Εικόνα 24 Διάγραμμα αυτοσυσχέτισης για τη λεκάνη 4	69
Εικόνα 25 Διάγραμμα αυτοσυσχέτισης για τη λεκάνη 5	70
Εικόνα 26 Διάγραμμα αυτοσυσχέτισης για τη λεκάνη 6	70
Εικόνα 27 Διάγραμμα αυτοσυσχέτισης για τη λεκάνη 7	71

Εικόνα 28 Διάγραμμα αυτοσυσχέτισης για τη λεκάνη 8	71
Εικόνα 29 Διάγραμμα αυτοσυσχέτισης για τη λεκάνη 9	72
Εικόνα 30 Διάγραμμα αυτοσυσχέτισης για τη λεκάνη 10	72
Εικόνα 31 Διάγραμμα μερικής αυτοσυσχέτισης για τη λεκάνη 1	73
Εικόνα 32 Διάγραμμα μερικής αυτοσυσχέτισης για τη λεκάνη 2	73
Εικόνα 33 Διάγραμμα μερικής αυτοσυσχέτισης για τη λεκάνη 3	74
Εικόνα 34 Διάγραμμα μερικής αυτοσυσχέτισης για τη λεκάνη 4	74
Εικόνα 35 Διάγραμμα μερικής αυτοσυσχέτισης για τη λεκάνη 5	75
Εικόνα 36 Διάγραμμα μερικής αυτοσυσχέτισης για τη λεκάνη 6	75
Εικόνα 37 Διάγραμμα μερικής αυτοσυσχέτισης για τη λεκάνη 7	76
Εικόνα 38 Διάγραμμα μερικής αυτοσυσχέτισης για τη λεκάνη 8	76
Εικόνα 39 Διάγραμμα μερικής αυτοσυσχέτισης για τη λεκάνη 9	77
Εικόνα 40 Διάγραμμα μερικής αυτοσυσχέτισης για τη λεκάνη 10	77
Εικόνα 41 Column chart, ορίζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων.....	85
Εικόνα 42 Κοινό διάγραμμα μοντέλου ARIMA(1,1,0) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 10 ημερών, λεκάνης 1	86
Εικόνα 43 Κοινό διάγραμμα μοντέλου ARIMA(2,1,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 10 ημερών λεκάνης 1	86
Εικόνα 44 Column chart, ορίζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 2	87
Εικόνα 45 Κοινό διάγραμμα μοντέλου ARIMA(1,1,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 10 ημερών λεκάνης 2	87
Εικόνα 46 Column chart, ορίζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 3	88
Εικόνα 47 Κοινό διάγραμμα μοντέλου ARIMA(1,1,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 10 ημερών λεκάνης 3	88
Εικόνα 48 Column chart, ορίζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 4	89
Εικόνα 49 Column chart, ορίζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 5	90

Εικόνα 50 Κοινό διάγραμμα μοντέλου ARMA(2,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 10 ημερών λεκάνης 5	90
Εικόνα 51 Column chart, ορίζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 6	91
Εικόνα 52 Κοινό διάγραμμα μοντέλου ARMA(2,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 10 ημερών λεκάνης 6	91
Εικόνα 53 Column chart, ορίζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 7	92
Εικόνα 54 Κοινό διάγραμμα μοντέλου ARMA(2,1) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 10 ημερών λεκάνης 7	92
Εικόνα 55 Column chart, ορίζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 8	93
Εικόνα 56 Column chart, ορίζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 9	94
Εικόνα 57 Κοινό διάγραμμα μοντέλου 'no no' και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 10 ημερών λεκάνης 9	94
Εικόνα 58 Column chart, ορίζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 10	95
Εικόνα 59 Κοινό διάγραμμα μοντέλου ARIMA(2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 10 ημερών λεκάνης 10	95
Εικόνα 60 Column chart, ορίζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 1	96
Εικόνα 61 Κοινό διάγραμμα μοντέλου ARIMA(1,1,1) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 1	97
Εικόνα 62 Κοινό διάγραμμα μοντέλου 'no no' και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 1	97
Εικόνα 63 Column chart, ορίζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 2	98
Εικόνα 64 Κοινό διάγραμμα μοντέλου ARMA(2,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 2	98
Εικόνα 65 Column chart, ορίζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 3	99
Εικόνα 66 Κοινό διάγραμμα μοντέλου ARIMA(2,1,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 3	99

Εικόνα 67 Column chart, ορίζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 4	100
Εικόνα 68 Κοινό διάγραμμα μοντέλου ARMA(2,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 4	100
Εικόνα 69 Column chart, ορίζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 5	101
Εικόνα 70 Κοινό διάγραμμα μοντέλου ARIMA(2,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 5	101
Εικόνα 71 Column chart, ορίζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 6	102
Εικόνα 72 Κοινό διάγραμμα μοντέλου ARMA(2,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 6	102
Εικόνα 73 Column chart, ορίζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 7	103
Εικόνα 74 Κοινό διάγραμμα μοντέλου ARIMA(1,1,1) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 7	103
Εικόνα 75 Column chart, ορίζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 8	104
Εικόνα 76 Κοινό διάγραμμα μοντέλου ARMA(2,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 8	104
Εικόνα 77 Column chart, ορίζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 9	105
Εικόνα 78 Κοινό διάγραμμα μοντέλου ‘ad no’ και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 9	105
Εικόνα 79 Column chart, ορίζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 10	106
Εικόνα 80 Κοινό διάγραμμα μοντέλου ARIMA(1,1,0) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 10	106
Εικόνα 81 Column chart, ορίζοντας πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 1	107
Εικόνα 82 Κοινό διάγραμμα μοντέλου deep learning και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 1	108
Εικόνα 83 Κοινό διάγραμμα μοντέλου ARMA(2,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 1	108

Εικόνα 84 Column chart, ορίζοντας πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 2	109
Εικόνα 85 Κοινό διάγραμμα μοντέλου deep learning και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 2	109
Εικόνα 86 Column chart, ορίζοντας πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 3	110
Εικόνα 87 Κοινό διάγραμμα μοντέλου deep learning και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 3	110
Εικόνα 88 Κοινό διάγραμμα μοντέλου ARIMA(2,1,1) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 3	111
Εικόνα 89 Column chart, ορίζοντας πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 4	112
Εικόνα 90 Κοινό διάγραμμα μοντέλου deep learning και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 4	112
Εικόνα 91 Column chart, ορίζοντας πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 5	113
Εικόνα 92 Κοινό διάγραμμα μοντέλου ARMA(2,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 5	113
Εικόνα 93 Κοινό διάγραμμα μοντέλου deep learning και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 5	114
Εικόνα 94 Column chart, ορίζοντας πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 6	115
Εικόνα 95 Κοινό διάγραμμα μοντέλου deep learning και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 5	115
Εικόνα 96 Column chart, ορίζοντας πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 7	116
Εικόνα 97 Κοινό διάγραμμα μοντέλου 'no no' και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 7	116
Εικόνα 98 Κοινό διάγραμμα μοντέλου deep learning και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 7	117
Εικόνα 99 Column chart, ορίζοντας πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 8	118
Εικόνα 100 Κοινό διάγραμμα μοντέλου deep learning και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 8	118

Εικόνα 101 Column chart, ορίζοντας πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 9	119
Εικόνα 102 Column chart, ορίζοντας πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 10	120
Εικόνα 103 Κοινό διάγραμμα μοντέλου ARMA(2,1) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 10	120
Εικόνα 104 Κοινό διάγραμμα μοντέλου deep learning και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 10	121
Εικόνα 105 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 1 deep learning και πραγματικών τιμών.....	121
Εικόνα 106 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 2 deep learning και πραγματικών τιμών.....	122
Εικόνα 107 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 3 deep learning και πραγματικών τιμών.....	122
Εικόνα 108 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 4 deep learning και πραγματικών τιμών.....	123
Εικόνα 109 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 5 deep learning και πραγματικών τιμών.....	123
Εικόνα 110 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 6 deep learning και πραγματικών τιμών.....	124
Εικόνα 111 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 7 deep learning και πραγματικών τιμών.....	124
Εικόνα 112 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 8 deep learning και πραγματικών τιμών.....	125
Εικόνα 113 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 9 deep learning και πραγματικών τιμών.....	125
Εικόνα 114 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 10 deep learning και πραγματικών τιμών	126
Εικόνα 115 Scatter plot σκορ NSE του καλύτερου μοντέλου ARIMA και deep learning	133
Εικόνα 116 Scatter plot που απεικονίζει την αύξηση του δείκτη NSE που προσφέρει το deep learning σε κάθε λεκάνη., προέκυψε από την αλγεβρική διαφορά NSE deep learning και NSE ARIMA.....	133
Εικόνα 117 Scatter plot σκορ ANSE του καλύτερου μοντέλου ARIMA, εκθετικής ομαλοποίησης και deep learning.....	134

Εικόνα 118 Scatter plot που απεικονίζει την αύξηση του δείκτη NSE που προσφέρει το deep learning σε κάθε λεκάνη, προέκυψε από την αλγεβρική διαφορά NSE deep learning και NSE κλασσικών μεθόδων.....134

Εικόνα 119 Scatter plot των τιμών NSE για κάθε λεκάνη στο μοντέλο προσομοίωσης λεκάνης135

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1. Τιμές p-value μετά από Augmented Dickey Fuller test για όλες τις λεκάνες67

Πίνακας 2 Υποτυπώδης πίνακας όπου μια χρονοσειρά μετατρέπεται σε δεδομένα μορφής πίνακα80

1 Εισαγωγή

1.1 Οι χρονοσειρές

Η παρούσα εργασία έχει ως σκοπό την ανάλυση υδρολογικών δεδομένων. Τα δεδομένα αυτά συνήθως έχουν τη μορφή χρονοσειρών. Μια χρονοσειρά (D, 1994; Brockwell, 2009) αποτελείται από ένα σύνολο παρατηρήσεων που αναφέρονται σε ένα ορισμένο χρόνο. Οι χρονοσειρές ανάλογα με το σε ποιά διαστήματα καταγράφονται οι παρατηρήσεις μπορούν να χωριστούν σε δυο κατηγορίες στις διακριτές και στις συνεχείς. Οι διακριτές χρονοσειρές έχουν παρατηρήσεις που γίνονται σε διαφορετικά χρονικά διαστήματα ενώ οι συνεχείς έχουν παρατηρήσεις που καταγράφονται συνεχώς σε ένα χρονικό διάστημα. Αξίζει να σημειωθεί ότι ενώ οι συνεχείς έχουν θεωρητικά περισσότερη πληροφορία από τις διακριτές το γεγονός ότι δύσκολα μπορεί να τις επεξεργαστεί ο ηλεκτρονικός υπολογιστής τις καθιστά λιγότερο χρήσιμες (Gershenfeld, 1999), έτσι συνήθως χρησιμοποιούνται διακριτές χρονοσειρές στην ανάλυση. Στόχος της ανάλυσης είναι αφού μελετηθεί η συμπεριφορά των ιστορικών παρατηρήσεων να κατανοηθεί η δομή τους ώστε να δημιουργηθεί ένα μοντέλο που θα χρησιμεύσει στην πρόγνωση και την κατανόηση της μελλοντικής εξέλιξης των συγκεκριμένων παρατηρήσεων.

Πρέπει να τονιστεί ότι ενώ στη στατιστική θεωρία τα δείγματα που μελετώνται είναι ανεξάρτητες μεταξύ τους παρατηρήσεις τις περισσότερες φορές, σε αρκετές χρονοσειρές οι κοντινές χρονικά παρατηρήσεις παρουσιάζουν συσχέτιση (Fuller, 1995). Αυτό το γεγονός υποδεικνύει ότι ένας τρόπος πρόγνωσης είναι να λαμβάνονται υπόψη οι άμεσα παρελθοντικές παρατηρήσεις.

1.2 Χρονοσειρές και μοντέλα προσομοίωσης

Υπάρχουν δύο είδη προγνωστικών μοντέλων τα αιτιοκρατικά και τα πιθανοτικά (Lee, 2009). Στην πρώτη περίπτωση το μοντέλο εκτιμά ακριβώς την μελλοντική τιμή της μεταβλητής της χρονοσειράς. Το μοντέλο αυτό ονομάζεται αλλιώς και ντετερμινιστικό. Η δεύτερη περίπτωση είναι τα πιθανοτικά που ονομάζονται και στοχαστικά (Aris, 1978). Σε αυτά τα μοντέλα υπάρχει συνήθως ένα ντετερμινιστικό μέλος και επιπρόσθετα υπάρχει και ένα στοχαστικό μέλος που κάνει τις εκτιμήσεις να προσδιορίζονται με βάση πιθανοτικές κατανομές.

Μια χρονοσειρά μπορεί να αναφέρεται σε μια μόνο μεταβλητή παραδείγματος χάρη απορροή ποταμού ή σε περισσότερες μεταβλητές (πολυμεταβλητή ανάλυση).

Σε μια χρονοσειρά (George E. P. Box, 2015) μπορούν να διακριθούν 4 δομικές συνιστώσες:

1. Η τάση (trend)
2. Η κυκλικότητα
3. Η εποχικότητα
4. Η μη κανονική συνιστώσα, ο λευκός θόρυβος

Η τάση μιας χρονοσειράς μπορεί να είναι αυξητική μειούμενη ή σχεδόν μηδενικής κλίσης σε ένα ορισμένο χρονικό διάστημα.

Η κυκλική συνιστώσα περιγράφει μεσαίας κλίμακας αλλαγές που δημιουργούνται από επαναλαμβανόμενους κύκλους συνήθως σε χρονοσειρές που αφορούν οικονομικά δεδομένα.

Η εποχική συνιστώσα χαρακτηρίζει μεταβολές των χρονοσειρών κατά τη διάρκεια του έτους, όσον αφορά την υδρολογία. Συμπεριλαμβάνει την περιοδο βροχοπτώσεων και ανυδρίας.

Η τυχαία συνιστώσα αποτελείται από απρόβλεπτους παράγοντες που έχουν μη κανονική συμπεριφορά

Με βάση τα παραπάνω τα μοντέλα μπορούν να διακριθούν σε δύο κατηγορίες ανάλογα με το πως συνδέονται οι όροι που περιγράφηκαν μεταξύ τους (Pankratz, 1983; K.W. Hipel, 1994; Costa, 2019)

Έτσι, υπάρχει το πολλαπλασιαστικό μοντέλο:

$$Y(t) = T(t)*S(t)*C(t)*R(t)$$

Και το αθροιστικό μοντέλο:

$$Y(t) = T(t)+S(t)+C(t)+R(t)$$

Όπου T η τάση, S η εποχικότητα, C η κυκλικότητα, R η τυχαία συνιστώσα.

1.3 Χρονικό διάστημα καταμέτρησης της χρονοσειράς

Εφόσον τα περισσότερα δεδομένα είναι σε διακριτή μορφή η παράμετρος του πόσο συχνά καταμετρώνται, ονομάζεται 'χρονικό διάστημα καταμέτρησης' (time interval). Ανάλογα με το ανωτέρω στοιχείο οι χρονοσειρές χρησιμοποιούνται σε διαφορετικού είδους μελέτες. Σειρές με πολλές καταμετρήσεις στην ημέρα ή και στην ώρα χρησιμοποιούνται για την καθημερινή παρακολούθηση και πρόγνωση του καιρού, τέτοια δεδομένα είναι όλα τα μετεωρολογικά όπως η βροχή, η χιονόπτωση, η θερμοκρασία (Jinbo Qin1, 2017). Αντίστοιχα, σειρές με μεγάλα χρονικά διαστήματα καταμέτρησης χρησιμοποιούνται για να κατασκευαστούν έργα όπως φράγματα ή για να αποφασιστούν οι κατάλληλες μακροχρόνιες στρατηγικές διαχείρισης νερού. Πιο συγκεκριμένα, για να κατασκευαστεί ένα φράγμα πρέπει να υπάρχουν δεδομένα πολλών ετών (με μηνιαίο συνήθως χρονικό διάστημα). Αυτό συμβαίνει γιατί το φράγμα είναι ένα κοστοβόρο έργο που είναι δεδομένο ότι θα επηρεάσει το περιβάλλον οπότε είναι απαραίτητο να γνωρίζουμε πως θα είναι λειτουργικό και χρήσιμο πολλές δεκαετίες μετά την κατασκευή του. Ένα χαρακτηριστικό παράδειγμα αστοχίας στη μελέτη ενός φράγματος είναι η υποτίμηση της στερεοπαροχής του υδατορεύματος με αποτέλεσμα τα φερτά που συγκεντρώνονται στον ταμιευτήρα να είναι πολύ περισσότερα από τα εκτιμώμενα και να στοιχίζουν στον ωφέλιμο όγκο της κατασκευής μειώνοντας την λειτουργικότητα του έργου. Αντίστοιχα, όσον αφορά τη διαχείριση του νερού η ζήτηση για ύδρευση και άρδευση γίνεται με ορίζοντα μελέτης το μήνα βραχυπρόθεσμα ή και το χρόνο μακροπρόθεσμα.

1.4 Σημασία πρόγνωσης σε βραχυπρόθεσμο ορίζοντα

1.4.1 Στον αγροτικό τομέα

Στην γεωργία χάνεται το μεγαλύτερο ποσοστό του νερού που διατίθεται προς εκμετάλλευση. Συνολικά υπολογίζεται ότι κατά μέσο όρο αξιοποιείται μόλις το 55% του αρχικού όγκου νερού (Υπουργείο Περιβάλλοντος και Ενέργειας, 2023) με τις απώλειες νερού σε ορισμένες καλλιέργειες να φτάνουν και έως 80%. Αυτό είναι αποτέλεσμα της ελλιπούς συντήρησης των εγχειοβελτιωτικών έργων που προκαλούν διαρροές στο δίκτυο της χρήσης συστημάτων άρδευσης που συνοδεύονται από μεγάλες εξατμίσεις νερού και της άστοχης σε ορισμένες περιπτώσεις διαχείρισης του διαθέσιμου υδάτινου όγκου. Το τελευταίο πρόβλημα είναι απόρροια του γεγονότος ότι η ευρύτερη πρόγνωση και διαχείριση του νερού δεν εφαρμόζεται στις περισσότερες λεκάνες απορροής με ταμειυτήρες.

Χαρακτηριστικά παραδείγματα είναι όταν οι καλλιέργειες αναγκάζονται να ποτιστούν σε ημέρες με υψηλή εξάτμιση ή σε ημέρες με ισχυρούς ανέμους σε τέτοιες περιπτώσεις οι απώλειες νερού είναι δεδομένο πως θα είναι αυξημένες. Επιπλέον, η έλλειψη πρόγνωσης φέρνει τους αγρότες όλο και περισσότερο αντιμέτωπους με προβλήματα ξηρασίας και τους αναγκάζει να στραφούν στις γεωτρήσεις (Υπουργείο Περιβάλλοντος και Ενέργειας, 2014) που έχουν δημιουργήσει αρκετά προβλήματα στην υφαιμύρωση των υπογείων υδάτων καταστρέφοντας αγροτικές καλλιέργειες. Η προσπάθεια λοιπόν πρόγνωσης θα μπορούσε να διευκολύνει την ορθότερη διαχείριση του νερού ώστε να αποφευχθούν όσο είναι δυνατόν τέτοια προβλήματα. Αξίζει εδώ να σημειωθεί πως οι περισσότερες καλλιέργειες έχουν απαιτήσεις για άρδευση περίπου κάθε 7 με 14 ημέρες οπότε οι εκτιμήσεις των 5 και 10 ημερών σε συνδιασμό με το ότι κάθε αγρότης θα έχει τις δικές του καλλιέργειες και επομένως ανάγκες και ωράριο καθίστανται αρκετά βοηθητικές ειδικά αν γίνεται λόγος για μικρής κλίμακας φράγματα που δε συγκρατούν τεράστιους υδάτινους όγκους. Είναι εμφανές ότι ο επαναπροσδιορισμός των εγχειοβελτιωτικών έργων στον πλανήτη είναι ένα μείζον ζήτημα σε πολλές χώρες (Chris Seijger, 2023) και αρκετοί ερευνητές είναι στραμμένοι στη προσπάθεια πρόγνωσης σε βραχυπρόθεσμη κλίμακα όπως των 7 ημερών (*, Wu, Wang, & Hakala, 2022).

1.4.1.1 Στον τομέα της ενέργειας

Στο καθεστώς ενεργειακής κρίσης στο οποίο έχει περιέλθει ο πλανήτης είναι πλέον απαραίτητο να επενδύσει η ανθρωπότητα σε έργα που αφορούν την καλύτερη εκμετάλλευση κι διαχείριση των ανανεώσιμων πηγών ενέργειας για να μπορέσουν οι ΑΠΕ να καταλάβουν όλο και μεγαλύτερο ποσοστό στο ενεργειακό ισοζύγιο.

Η Ευρωπαϊκή Ένωση έχοντας αναγνωρίσει αυτή την ανάγκη έχει θέσει κάποιους στόχους / οδηγίες για τα κράτη μέλη της με δεσμευτικό χαρακτήρα. (EUR-Lex, 2018).

1.4.1.2 Ανανεώσιμες πηγές ενέργειας

Ανανεώσιμη ενέργεια είναι η ενέργεια που προέρχεται από φυσικούς πόρους ή διαδικασίες που συνεχώς ανανεώνονται. Παραδείγματος χάρη ο άνεμος και η ηλιοφάνεια δεν εξαντλούνται. (Boyle, 2012).

Οι πιο διαδεδομένες ανανεώσιμες πηγές ενέργειας είναι:

- Η αιολική, είναι μια μορφή που κατέχει κυρίαρχα ποσοστά στα ενεργειακά ισοζύγια πολλών χωρών όπως οι Ηνωμένες Πολιτείες Αμερικής και το Ηνωμένο Βασίλειο. Αιολικές εγκαταστάσεις στην ξηρά ή στη θάλασσα παράγουν ηλεκτρική ενέργεια καθώς τα πτερύγια των ανεμογεννητριών στρέφονται.
- Ηλιακή, το φως του ηλίου ως ένα από τα αγαθά που βρίσκεται σε αφθονία στον πλανήτη μας είναι σαφές ότι αποτελεί σημαντική πηγή ενέργειας, βέβαια η απόδοση των φωτοβολταϊκών δεν είναι πάντα σταθερή γιατί το ίδιο και η ηλιακή ακτινοβολία είναι κυμαινόμενη. Η ενέργεια παράγεται όταν το φως του ήλιου πέφτει πάνω σε ηλιακά πάνελ και ακολουθούν φυσικές και χημικές αντιδράσεις που συνιστούν το φωτοβολταϊκό φαινόμενο.
- Υδροηλεκτρική, αυτή είναι και η ενέργεια την διαχείριση της οποίας επιδιώκει να βελτιστοποιήσει η παρούσα διπλωματική εργασία. Κατασκευάζονται φράγματα ή έργα εκτροπής τα οποία δημιουργούν μια τεχνητή υψομετρική διαφορά και η δυναμική ενέργεια μετατρέπεται σε κινητική η οποία χρησιμοποιείται για να παραχθεί ηλεκτρική ενέργεια, η διαδικασία θα εξηγηθεί αναλυτικότερα και στη συνέχεια.
- Βιοενέργεια, πρόκειται για την οργανική ύλη που χρησιμοποιείται ως καύσιμο.

1.4.1.3 Υδροηλεκτρική ενέργεια

Αναλυτικότερα η Υδροηλεκτρική ενέργεια είναι μια από τις παλαιότερες και σημαντικότερες μορφές ενέργειας που εκμεταλλεύεται τον κύκλο του νερού και συγκεκριμένα την επιφανειακή απορροή των υδάτων. Δύο κατασκευές χρησιμοποιούνται για την ανωτέρω εκμετάλλευση:

- Φράγματα. Σε θέσεις που η γεωλογία και η μορφολογία του εδάφους ευνοεί δημιουργούνται κατασκευές είτε από χώμα είτε από σκυρόδεμα που δεσμεύουν το νερό ενός υδατορεύματος δημιουργώντας έτσι μια τεχνητή λίμνη. Έχοντας συγκεντρώσει πολλά κυβικά νερού δημιουργείται μια υψομετρική διαφορά μεταξύ της ελεύθερης επιφάνειας του νερού και του εδάφους. Το νερό πέφτει από το τεχνητό υψόμετρο και η δυναμική ενέργεια που μετατρέπεται σε κινητική ενέργεια των υδροστροβίλων οι οποίοι παράγουν ηλεκτρική ενέργεια. Τα φράγματα έχουν το πλεονέκτημα πως λειτουργούν ως αποθηκευτικός χώρος ενέργειας καθώς τα αποθέματα νερού μπορούν να δεσμευτούν και να αξιοποιηθούν μόνο όταν το επιβάλλει η ζήτηση. Σε μερικά φράγματα μάλιστα υπάρχουν και εγκαταστάσεις αντλησοταμίευσης, αυτές παρέχουν τη δυνατότητα αντλησης του νερού και μεταφοράς του από το χαμηλό υψόμετρο σε υψηλότερο. Αξίζει να σημειωθεί ότι η ενέργεια που χρειάζεται για να αντληθεί το νερό είναι περίπου διπλάσια από την ενέργεια που παράγει το νερό όταν χρησιμοποιείται, αυτό το γεγονός είναι ένδειξη πως η ενέργεια ανάλογα τη χρονική στιγμή που παράγεται έχει άλλη

αξία και επομένως η σωστή διαχείριση του νερού είναι πολύ σημαντική για να είναι το υδροηλεκτρικό όσο το δυνατόν πιο αποδοτικό. Επίσης πρέπει να σημειωθεί ότι όσο πιο πλήρης είναι ο ταμιευτήρας ενός έργου τόσο μεγαλύτερο ύψος πτώσης υπάρχει και επομένως τόσο περισσότερη ενέργεια θα παράγεται, όσο μεγαλύτερο το ύψος πτώσης τόσο το καλύτερο.

- Έργο εκτροπής. Όταν δεν υπάρχουν τα κατάλληλα στοιχεία για τη δημιουργία φράγματος είτε λόγω μορφολογίας εδάφους είτε λόγω γεωλογίας (τα πετρώματα έχουν μεγάλη υδατοπερατότητα και η αποθήκευση νερού καθίσταται ανέφικτη λόγω μεγάλων υπόγειων διαφυγών) κατασκευάζονται συχνά έργα εκτροπής. Τα έργα αυτά εκτρέπουν το υδατόρευμα, ή ένα μέρος του, από τον αρχικό ρού του και το κατευθύνουν σε σημεία με φυσική υψομετρική διαφορά, από εκεί παράγουν ενέργεια με τον τρόπο που περιγράφηκε προηγουμένως.

1.4.1.4 Χρηματιστήριο ενέργειας

Σε αυτό το σημείο αξίζει να αναφερθεί το χρηματιστήριο ενέργειας και θα εξηγηθεί πως λειτουργεί. (Enexgroup.gr, 2023).

Το χρηματιστήριο ενέργειας χωρίζεται σε 3 βασικές αγορές ανάλογα με το πότε αγοράζεται η ενέργεια και πότε είναι συμφωνημένο να παραχθεί. Αυτές οι αγορές είναι:

a. Η προθεσμιακή αγορά ενέργειας:

Στην παρούσα αγορά οι αγοραστές και οι πωλητές συμφωνούν για ενέργεια που θα παραχθεί τις επόμενες 2 ή περισσότερες ημέρες. Η παρούσα αγορά αναφέρεται κυρίως σε ορυκτά καύσιμα και δεν είναι η κύρια αγορά του χρηματιστηρίου ενέργειας. Στη συμφωνία αυτή περιλαμβάνεται ο τρόπος παραγωγής και η ρήτρα σε περίπτωση που ο παραγωγός δεν μπορεί να φέρει εις πέρας την ηλεκτροπαραγωγή.

b. Η day ahead αγορά:

Αυτή είναι η αγορά ‘της επόμενης μέρας’ και υφίσταται διότι όπως αναφέρθηκε πιο πάνω σε περίπτωση που η απαιτούμενη ενέργεια δεν μπορέσει να παραχθεί (επειδή παραδείγματος χάρη ήταν αιολική ενέργεια και ο άνεμος δεν είχε την κατάλληλη ένταση για την απαιτούμενη διάρκεια) ή επειδή οι ανάγκες του αγοραστή δεν αντικατοπτρίζουν ακριβώς το ποσό ενέργειας που αγοράστηκε στην γενική αγορά. Η αγορά αυτή λοιπόν είναι κατά κάποιον τρόπο διορθωτική,αγοραστές και πωλητές υποβάλουν τις εντολές προσφοράς και ζήτησης προκειμένου να εξισορροπήσουν το ισοζύγιο. Σε αυτή την αγορά η δυνατότητα των υδροηλεκτρικών που μπορούν να προσφέρουν ενέργεια στο δίκτυο εντός λεπτών είναι σημαντική.

c. Η intraday αγορά:

Με παρεμφερή λογική με την day ahead αγορά υπάρχει για να διορθώσει οποιαδήποτε δυσκολία μπορεί να εμφανιστεί. Εδώ και πάλι οι παραγωγοί υδροηλεκτρικής ενέργειας αποτελούν έναν από τους σημαντικούς παράγοντες στη συνεισφορά ενέργειας (Σειμανίδη, 2022).

1.4.2 Η έρευνα για τη σημασία βραχυχρόνιας πρόγνωσης

Όπως λοιπόν είναι κατανοητό οι παραγωγοί υδροηλεκτρικής ενέργειας έχουν μεγάλη δραστηριοποίηση στην day ahead και intraday αγορά για αυτό και μια πρόγνωση με ορίζοντα 5 και 10 ημερών είναι αρκετά χρήσιμη για να τοποθετήσει ο ηλεκτροπαραγωγός τις κατάλληλες εντολές πώλησης. Κάτι ακόμη που αξίζει να προστεθεί είναι πως με την αύξηση του ύψους πτώσης, το οποίο αυξάνεται όταν υψώνεται η στάθμη του νερού παράγεται περισσότερη ενέργεια με λιγότερο νερό που και αυτό μπορεί να συνυπολογιστεί στην καλύτερη αξιοποίηση των υδάτινων πόρων. Είναι πολλοί οι ερευνητές που έχουν συγκεντρώσει την προσοχή τους είτε στην βραχυχρόνια πρόγνωση της ζήτησης ενέργειας που μπορεί να καλυφθεί από υδροηλεκτρικές μονάδες (Dmitrieva, 2015) (Safaralieva, 2022) (Kavya, 2023) ή που απλώς προσπαθούν να προβλέψουν την απορροή ενός ποταμού με τη χρήση πολλαπλών μοντέλων (Jinbo Qin1, 2017)(multi-step-hydro, simulating and predicting).

1.5 Αντικείμενο της εργασίας

Αντικείμενο της **παρούσας διπλωματικής εργασίας** είναι η έρευνα πάνω στις μεθόδους που χρησιμοποιούνται για την πρόγνωση απορροής. Αρχικά, συγκεντρώθηκαν ημερήσια δεδομένα σε 10 λεκάνες των ΗΠΑ. Στη συνέχεια αναπτύχθηκαν σε προγραμματιστικό περιβάλλον Python μέθοδοι για την εκτίμηση της απορροής. Σε αυτές τις διαδικασίες συμπεριλαμβάνονται κλασσικές μέθοδοι όπως η ARIMA και η εκθετική ομαλοποίηση αλλά συμπεριλαμβάνονται και μέθοδοι από τον τομέα της τεχνητής νοημοσύνης και πιο συγκεκριμένα νευρωνικά δίκτυα από τον τομέα της βαθιάς μηχανικής μάθησης. Σκοπός είναι να φανεί πως συμπεριφέρεται το εκάστοτε μοντέλο σε διαφορετικά υδρολογικά μοτίβα κάθε λεκάνης και διαφορετικούς ορίζοντες πρόγνωσης για 1 , 5 και 10 ημέρες. Τέλος, αναπτύχθηκε και μια μέθοδος προσομοίωσης της λεκάνης με τη χρήση νευρωνικών δικτύων.

2 Δεδομένα

Για να μπορέσει η πρόγνωση να είναι στοχευμένη χρειάστηκαν ημερήσια δεδομένα. Η πηγή ήταν από το απόθεμα των Αμερικανών λεκανών ‘CAMELS dataset: Catchment Attributes and MEteorology for Large-sample Studies’ (NCAR UCAR, 2022). Σε αυτό το απόθεμα έχουν συγκεντρωθεί μετεωρολογικά δεδομένα και δεδομένα απορροής για 671 λεκάνες στις Ηνωμένες Πολιτείες Αμερικής. Για τη συγκεκριμένη μελέτη ανάλογα με τις εκάστοτε μεθόδους και τη δομή της εργασίας χρησιμοποιήθηκαν διαφορετικοί τύποι δεδομένων. Στις περισσότερες μεθόδους (ARIMA και εκθετική ομαλοποίηση) χρησιμοποιήθηκαν δεδομένα απορροής από 10 λεκάνες από το USGS:United States Geological Survey (USGS, 2023) η απορροή μετρήθηκε σε κυβικά μέτρα το δευτερόλεπτο (m^3/s). Ανάλογα με την εκάστοτε λεκάνη, χρησιμοποιήθηκαν οι διαθέσιμες χρονολογίες σε γενικές γραμμές τα δεδομένα που ελήφθησαν ήταν από 1 Ιανουαρίου 1980 έως 31 Δεκεμβρίου 2014.

Στη μέθοδο του νευρωνικού δικτύου χρησιμοποιήθηκαν και τα δεδομένα από το USGS αλλά και μετεωρολογικά δεδομένα από το DAYMET:Daily Surface Weather and Climatological Summaries (DAYMET, 2023). Από το DAYMET ελήφθησαν τα παρακάτω δεδομένα:

- Ημερομηνία μέσω των Year / Month / Day
- Ηλιοφάνεια μέσω του dayl σε δευτερόλεπτα (s)
- Βροχόπτωση μέσω του prcp σε millimeters ανά ημέρα (mm/day)
- Ηλιακή ακτινοβολία μέσω του srad (short waved radiation) σε watt ανά τετραγωνικό μέτρο (w/m^2)
- Μέγιστη και ελάχιστη θερμοκρασία μέσω των Tmax και Tmin σε βαθμούς Κελσίου (C)
- Πίεση μέσω του vp που μετράται σε Pascal (Pa).
- Ημερήσια απορροή

Σημειώνεται πως αναπτύχθηκαν δύο αλγόριθμοι με χρήση νευρωνικών δικτύων οπότε ανωτέρω έχουν παρουσιαστεί τα δεδομένα που χρησιμοποιήθηκαν και από τους δύο αλγορίθμους.

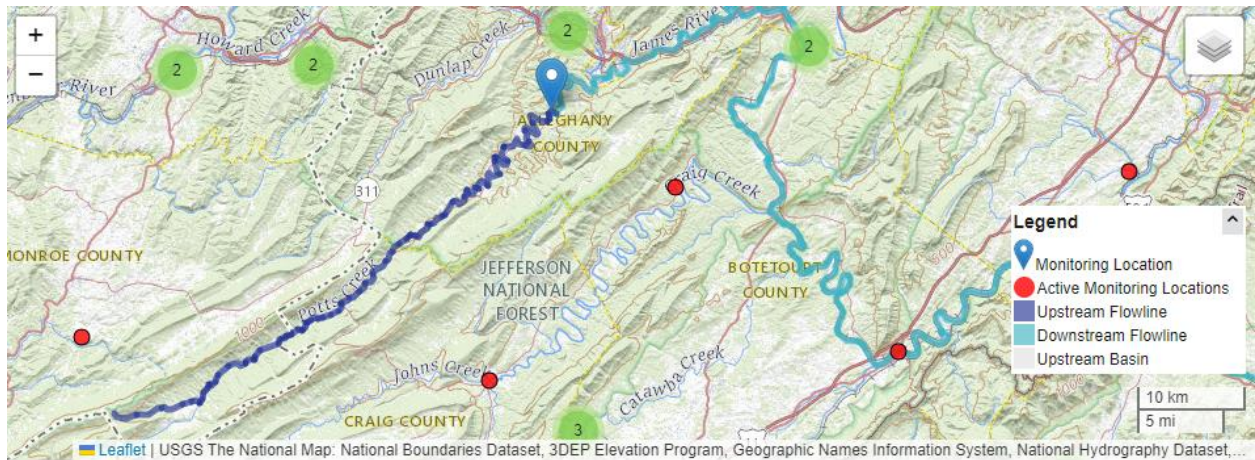
Οι λεκάνες διακρίνονται από έναν μοναδικό κωδικό id που τους έχει δοθεί και θα αναφερθούν με βάση αυτόν:

1. 02014000
2. 01411300
3. 01440000
4. 01543500
5. 02016000
6. 02053800

7. 02059500
8. 02069700
9. 02143000
10. 02350900

Παρακάτω παρατίθενται χάρτες για τον γεωγραφικό προσδιορισμό των κοιτών των ποταμών μελέτης αλλά και για κάποια χαρακτηριστικά των ευρύτερων λεκανών και γίνεται γραφική απεικόνιση των χρονοσειρών απορροής (USGS, 2023).

1. Id: 02014000 Potts creek

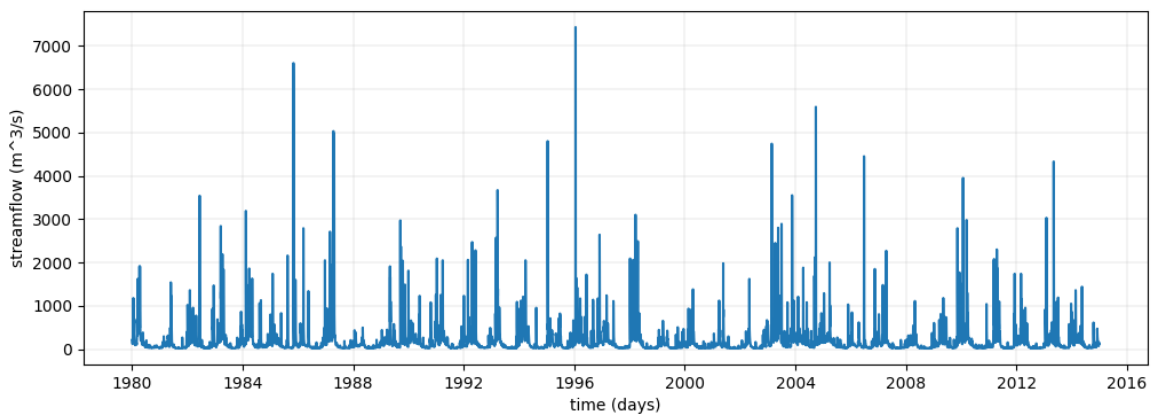


Εικόνα 1 Χάρτης Potts creek από USGS

Γεωγραφικό πλάτος = 37.67

Μέση ανύψωση του εδάφους = 582 μέτρα

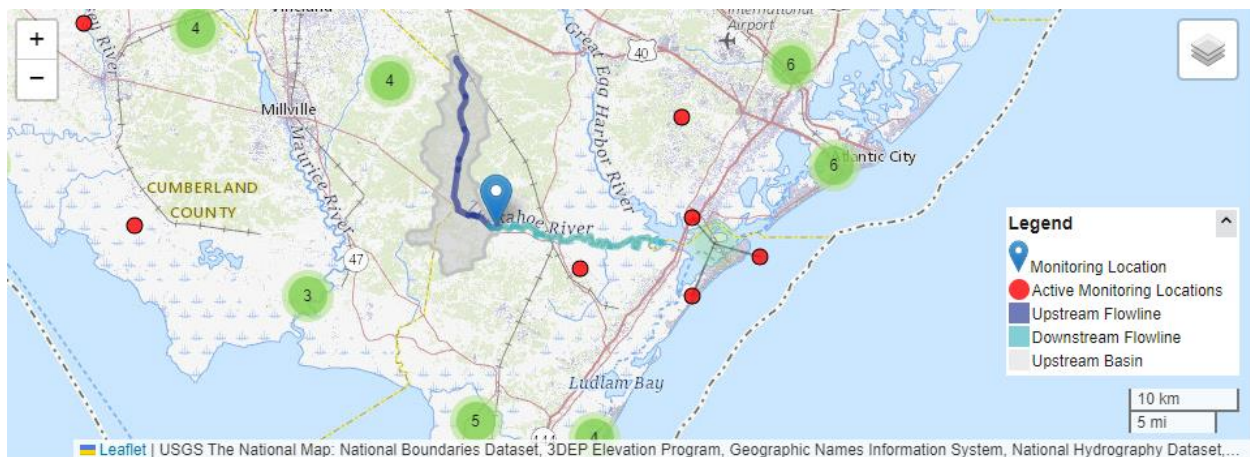
Επιφάνεια λεκάνης απορροής = 396 τετραγωνικά χιλιόμετρα.



Εικόνα 2 Χρονοσειρά Ημερήσιας Απορροής Potts creek

Το Potts creek πηγάζει από το ομώνυμο βουνό και είναι η πηγή του ποταμού Jackson που στη συνέχεια γίνεται ο ποταμός James. Το μήκος το είναι περίπου 52 χιλιόμετρα. Περνά παράλληλα από το βουνό Peters και ενισχύεται από όλα τα ρέματα στους νοτιοανατολικούς πρόποδες του. Διασχίζει την κομητεία Alleghany. Η λεκάνη του Potts creek αποτελείται από πυριτοκλαστικά πετρώματα στο 80% της έκτασης του και από μικτά ιζηματογενή σε ποσοστό κάλυψης 18%. Όσον αφορά την κάλυψη γης του περιλαμβάνει εύκρατα πλατύφυλλα δάση σε ποσοστό 90%, δάση που κυριαρχούνται από φυλλοβόλα δέντρα τα οποία ρίχνουν τα φύλλα τους το χειμώνα (MODIS Land Cover Type/Dynamics.nasa, 2022).

1. 2. Id: 01411300 Tuckahoe river

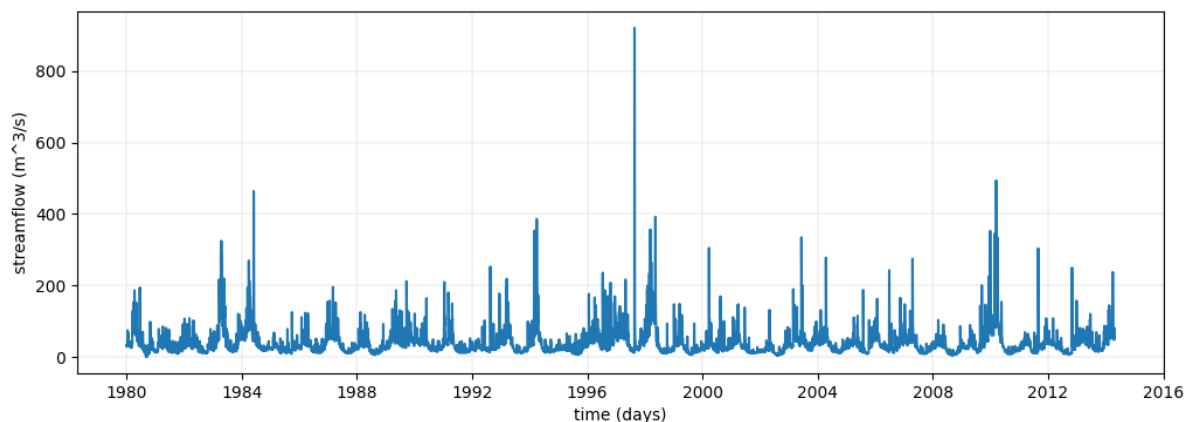


Εικόνα 3 Χάρτης Tuckahoe river από USGS

Γεωγραφικό πλάτος = 39.28

Μέση ανύψωση του εδάφους = 16 μέτρα

Επιφάνεια λεκάνης απορροής = 80 τετραγωνικά χιλιόμετρα



Εικόνα 4 Χρονοσειρά Ημερήσιας Απορροής Tuckahoe river

Ο ποταμός Tuckahoe είναι ένας ποταμός 43 χιλιομέτρων στο New Jersey ο ποταμός πηγάζει από την περιοχή των νοτίων Pinelands και εκβαλλει στον Ατλαντικό Ωκεανό. Η λεκάνη αποτελείται εξολοκλήρου από μη συγκεκολλημένα ιζήματα και καλύπτεται στο 85% της από μικτά δάση (MODIS Land Cover Type/Dynamics.nasa, 2022).

2. 3. Id: 01440000 Flatbrook river

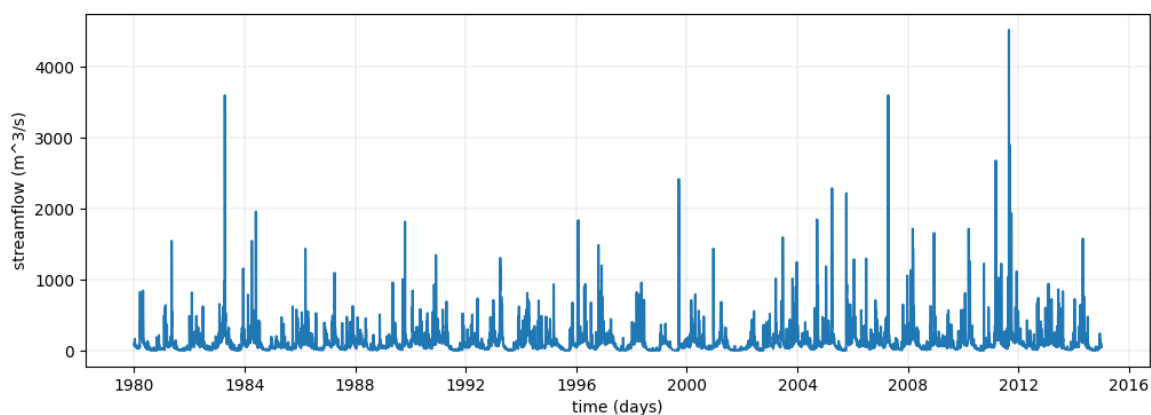


Εικόνα 5 Χάρτης Flatbrook river από USGS

Γεωγραφικό πλάτος = 41.16

Μέση ανύψωση του εδάφους = 130 μέτρα

Επιφάνεια λεκάνης απορροής = 165 τετραγωνικά χιλιόμετρα



Εικόνα 6 Χρονοσειρά Ημερήσιας Απορροής Flatbrook river

Ο Flatbrook είναι ένας παραπόταμος του Delaware ενός μεγάλου ποταμού που διανύει το New Jersey έχει σχετικά μικρό μήκος 26.6 χιλιόμετρα. Η γεωλογία της λεκάνης του παραποτάμου αποτελείται από πυριτοκλαστικά ιζηματογενή πετρώματα στο 84% και από ανθρακικά ιζηματογενή πετρώματα στο 15%. Στην περιοχή κυριαρχούν τα εύκρατα πλατύφυλλα δάση σε ποσοστό 90% (MODIS Land Cover Type/Dynamics.nasa, 2022).

4. Id: 01543500 Sinnemahoning Creek

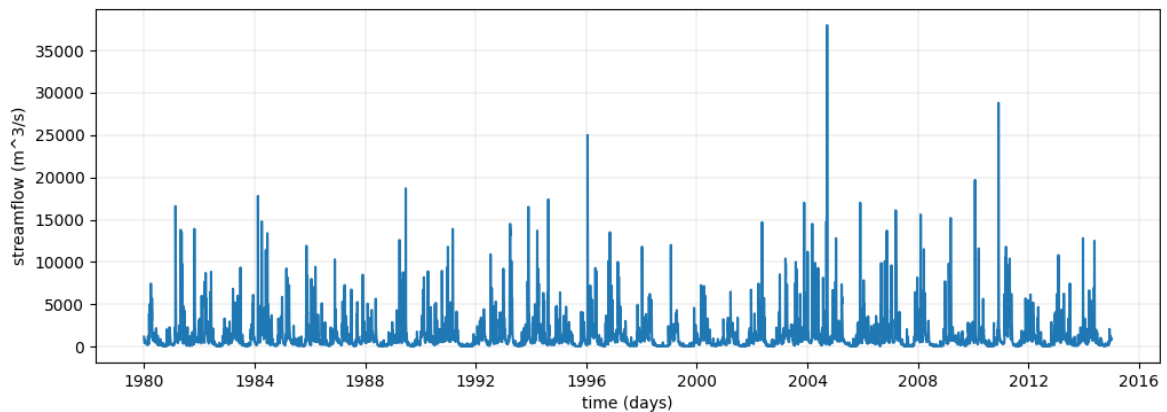


Εικόνα 7 Χάρτης Sinnemahoning Creek από USGS

Γεωγραφικό πλάτος = 41.14

Μέση ανύψωση του εδάφους = 281 μέτρα

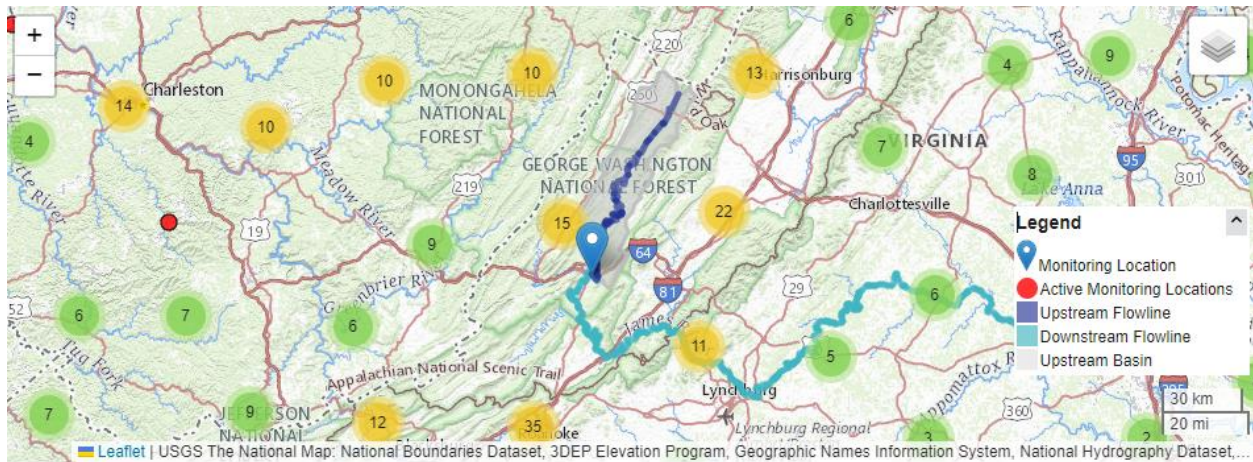
Επιφάνεια λεκάνης απορροής = 165 τετραγωνικά χιλιόμετρα



Εικόνα 8 Χρονοσειρά Ημερήσιας Απορροής Sinnemahoning Creek

Ο Sinnemahoning Creek είναι παραπόταμος του δυτικού σκέλους του Susquehanna River δημιουργείται με τη συμβολή δυο μικρότερων ποταμών του Bennet και του Driftwood. Το μήκος του είναι περίπου 50 χιλιόμετρα. Στη λεκάνη συναντώνται πυριτοκλασικά ιζηματογενή πετρώματα σε όλη την περιοχή και η καλυψη είναι εύκρατα φυλλοβόλα δάση στο 98% (MODIS Land Cover Type/Dynamics.nasa, 2022).

5. Id: 02016000 Cowpasture river

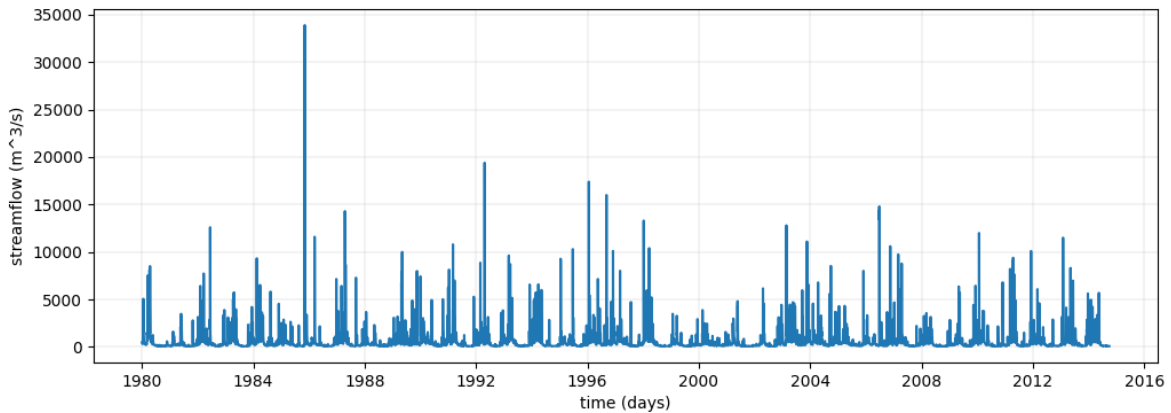


Εικόνα 9 Χάρτης Cowpasture river από USGS

Γεωγραφικό πλάτος = 38.23

Μέση ανύψωση του εδάφους = 487 μέτρα

Επιφάνεια λεκάνης απορροής = 1773 τετραγωνικά χιλιόμετρα

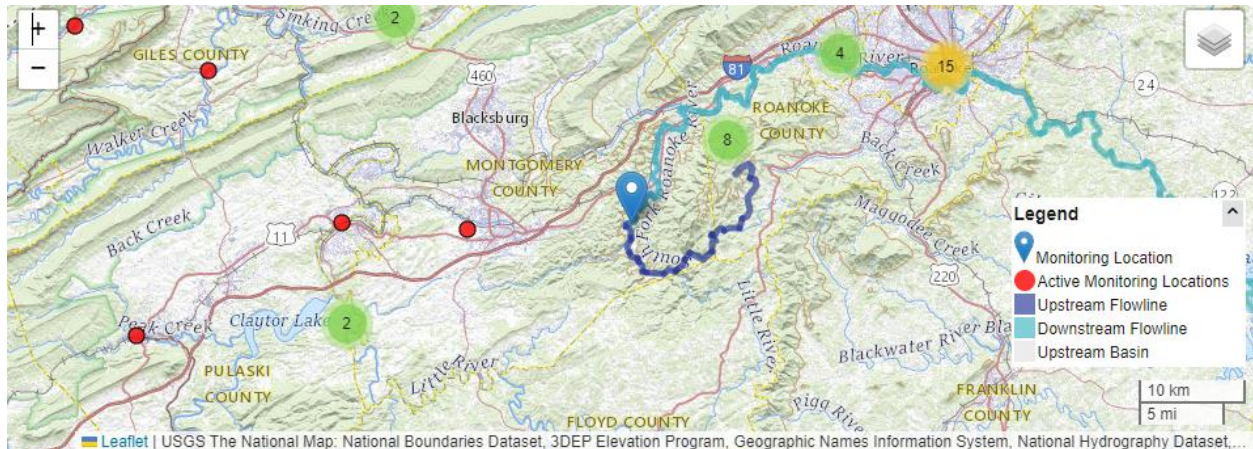


Εικόνα 10 Χρονοσειρά Ημερήσιας Απορροής Cowpasture river

Ο ποταμός Cowpasture είναι ο κύριος παραπόταμος του James river της δυτικής Virginia έχει μήκος 135.5 χιλιόμετρα. Πηγάζει από τη βορειοανατολική κομητεία Highland κοντά στο βουνό Bullpasture διανύει μια στενή κοιλάδα κοντά στον Εθνικό Δρυμό George Washington. Στη συνέχεια ενώνεται με τον παραπόταμο Shaws Fork που του δίνει εύρος αλλά τον διατηρεί ακόμη σε μικρό βάθος. Ύστερα, περνά από ένα στενό φαράγγι και ενώνεται με τον Bullpasture ποταμό και αποκτά μεγάλο υδάτινο όγκο. Από εκεί και έπειτα ενώνεται με πολλούς μικρούς παραποτάμους και αυξάνει το βάθος αρκετά. Τέλος, ενώνεται με τον Jackson για να δημιουργήσουν τον ποταμό James. Γεωλογικά τα πετρώματα που συναντώνται είναι τα πυριτοκλαστικά ιζηματογενή σε βαθμό 75% και τα υπόλοιπα είναι μικτά ιζηματογενή. Οι χρήσεις

γης είναι και πάλι εύκρατα πλατύφυλλα φυλοβόλλα δάση στο 93% της λεκάνης (MODIS Land Cover Type/Dynamics.nasa, 2022).

6. Id: 02053800 South Fork Roanoke river

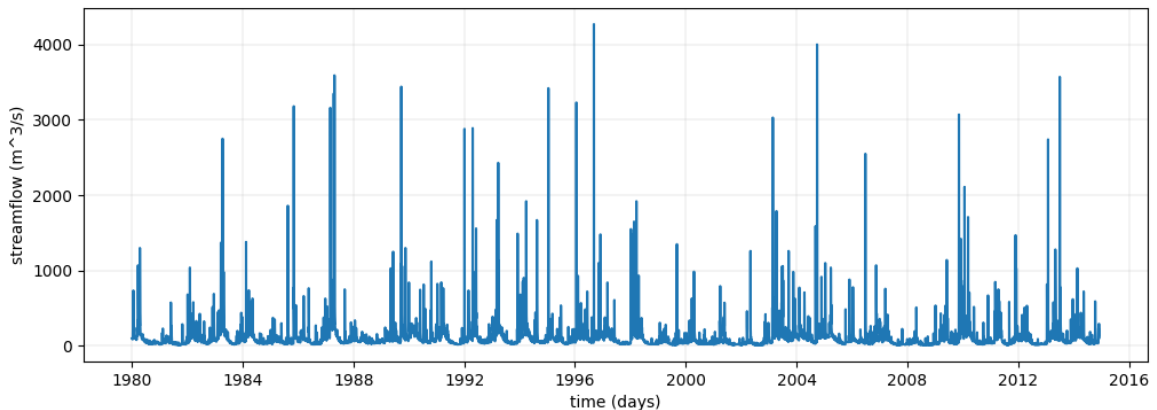


Εικόνα 11 Χάρτης South Fork Roanoke river από USGS

Γεωγραφικό πλάτος = 37.1

Μέση ανύψωση του εδάφους = 924 μέτρα (NCAR UCAR, 2022) – 1364 μέτρα (NCAR UCAR, 2022)

Επιφάνεια λεκάνης απορροής = 282 τετραγωνικά χιλιόμετρα (DAYMET, 2023) – 109 τετραγωνικά χιλιόμετρα (NCAR UCAR, 2022).



Εικόνα 12 Χρονοσειρά Ημερήσιας Απορροής South Fork Roanoke river

Να σημειωθεί πως τα δεδομένα των μοντέλων έχουν ληφθεί από τη βάση του CAMELS και το Daymet αλλά οι χάρτες έχουν ληφθεί από το Usgs και υπάρχει μια μικρή διαφορά όσον αφορά τη μέση ανύψωση και την επιφάνεια της λεκάνης αλλά τα μοντέλα έχουν εκπαιδευτεί σύμφωνα με το Daymet.

Πρόκειται για τον ποταμό που αποτελεί τη νότια πηγή του Roanoke έχει μήκος 32 χιλιόμετρα. Ενώνεται με τον North Fork Roanoke και έτσι σχηματίζεται ο κύριος ποταμός. Ο ποταμός έχει

μήκος 660 χιλιόμετρα και αποτελεί έναν από τους μεγαλύτερους ποταμούς της Βόρειας Καρολίνα σχηματίζοντας πολλές λίμνες στο πέρασμά του. Όσον αφορά τη γεωλογία αποτελείται από μεταμορφωσιγενή πετρώματα στο 48% και από πυριτοκλαστικά ιζηματογενή στο 32 %. Η κάλυψη γης έχει τα φυλλοβόλα εύκρατα δάση να κυριαρχούν σε ποσοστό 88%.

7. Id: 02059500 Goose creek

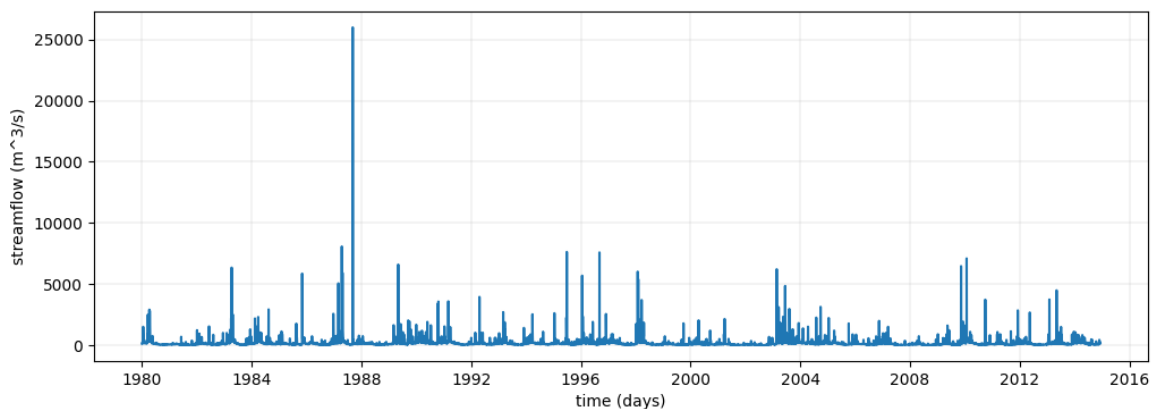


Εικόνα 13 Χάρτης Goose creek από USGS

Γεωγραφικό πλάτος = 37.34

Μέση ανύψωση του εδάφους = 360 μέτρα

Επιφάνεια λεκάνης απορροής = 486 τετραγωνικά χιλιόμετρα



Εικόνα 14 Χρονοσειρά Ημερήσιας Απορροής Goose creek

Ο ποταμός Goose Creek πηγάζει κοντά στο Εθνικό μονοπάτι των Απαλλάχιων Όρεων και ενισχύεται από παραποτάμους όπως ο Wolf και ο Bore Auger και διανύει το νοτιοδυτικό τμήμα της κομητείας Bredford. Το μήκος του είναι περίπου 40 χιλιόμετρα, αποτελεί παραπόταμο του Roanoke που προαναφέραμε στη λεκάνη 5 ενός μεγάλου ποταμού που διανύει τη Virginia και τη North Carolina. Τα πετρώματα που απαντώνται στην παρούσα λεκάνη είναι μεταμορφωσιγενή

στο 74% της και πυριτοκλαστικά ιζηματογενή στο 20%. Οι χρήσεις γης είναι καλλιέργειες και φυσική βλάστηση στο 85% της λεκάνης (MODIS Land Cover Type/Dynamics.nasa, 2022).

8. Id: 02069700 South Mayo River

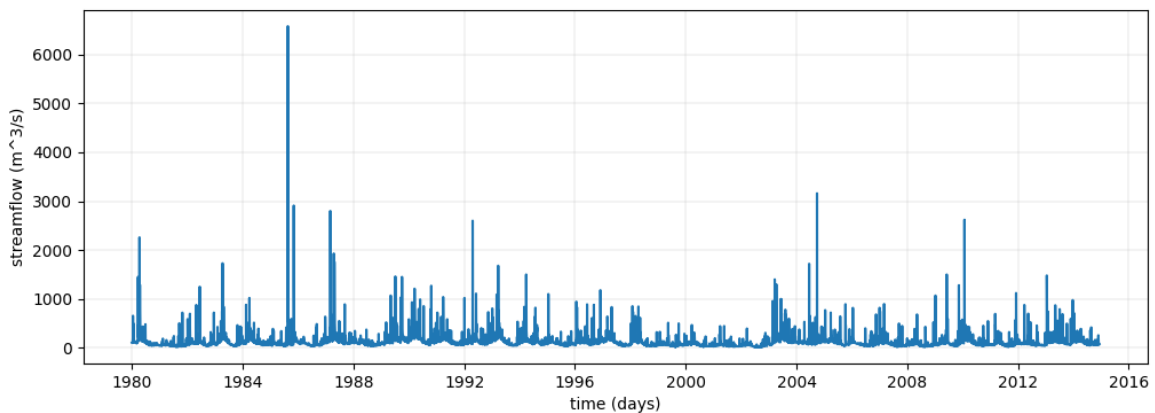


Εικόνα 15 Χάρτης South Mayo River από USGS

Γεωγραφικό πλάτος = 36.61

Μέση ανύψωση του εδάφους = 493 μέτρα

Επιφάνεια λεκάνης απορροής = 221 τετραγωνικά χιλιόμετρα



Εικόνα 16 Χρονοσειρά Ημερήσιας Απορροής South Mayo River

Ο ποταμός πηγάζει κοντά στην κομητεία Patrick μαζί με τον North Mayo και στη βορειοδυτική περιοχή της κομητείας Rockingham συνδέονται. Ο Mayo είναι ένας ποταμός που διασχίζει την πολιτεία της North Carolina για να ενωθεί στο τέλος με τον Dan. Ο κύριος γεωλογικός σχηματισμός είναι μεταμορφωσιγενή σε ποσοστό 64% και εκρηξιγενή σε ποσοστό 18%. Η κάλυψη γης είναι εύκρατα φυλλοβόλα δάση στο 96% (MODIS Land Cover Type/Dynamics.nasa, 2022).

9. Id: 02143000 Henry fork, near Henry river

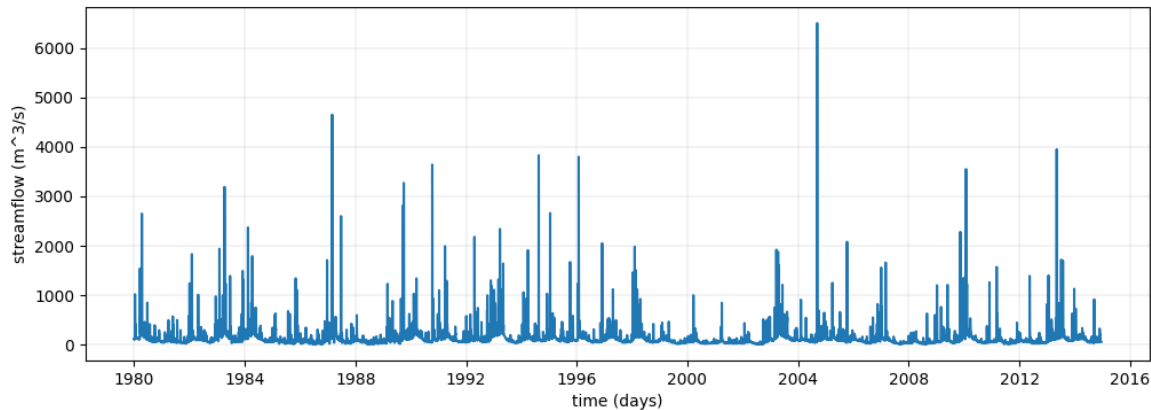


Εικόνα 17 Χάρτης Henry fork από USGS

Γεωγραφικό πλάτος = 35.59

Μέση ανύψωση του εδάφους = 497 μέτρα

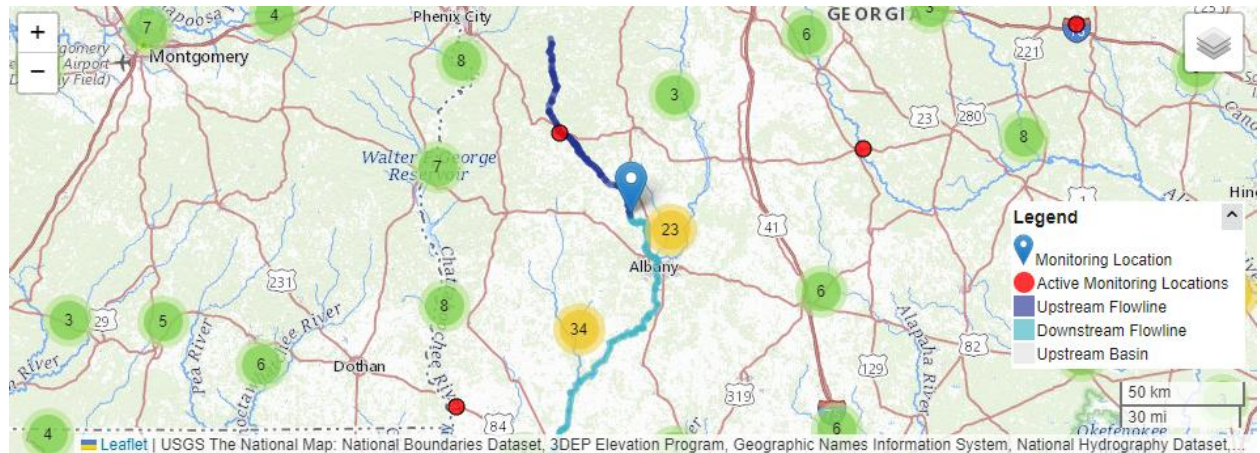
Επιφάνεια λεκάνης απορροής = 215 τετραγωνικά χιλιόμετρα



Εικόνα 18 Χρονοσειρά Ημερήσιας Απορροής Henry fork

Ο Henry fork είναι ένας μικρός σχετικά ποταμός μήκους 64 χιλιομέτρων και 4^{ος} παραπόταμος του South Fork Catwaba. Πηγάζει περίπου μισό χιλιόμετρο νοτιοδυτικά της κομητείας Burke. Αποτελείται εξ'ολοκλήρου από μεταμορφωσιγενή πετρώματα και η κάλυψη γης του είναι κατά 64% από μικτά δάση (MODIS Land Cover Type/Dynamics.nasa, 2022).

10. Id: 02350900 Kinchafoonee creek

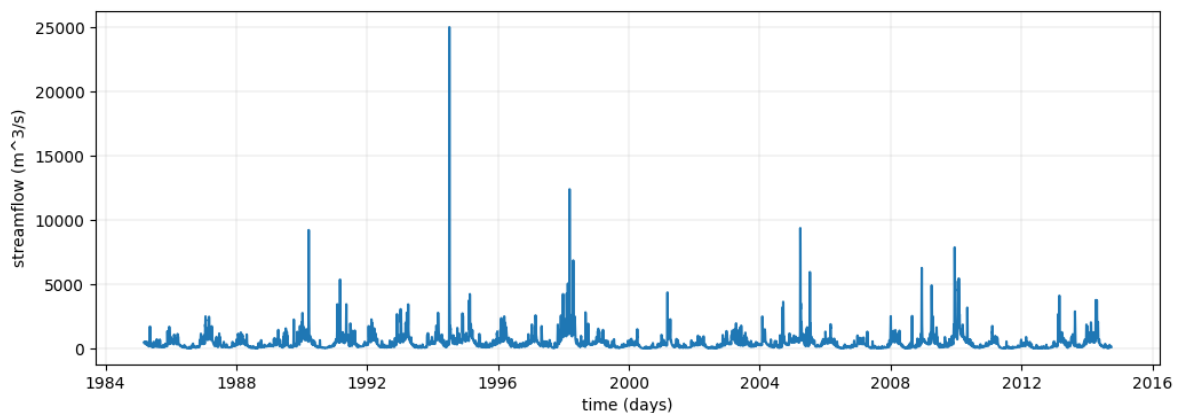


Εικόνα 19 Χάρτης Kinchafoonee creek από USGS

Γεωγραφικό πλάτος = 37.1

Μέση ανύψωση του εδάφους = 151 μέτρα

Επιφάνεια λεκάνης απορροής = 1364 τετραγωνικά χιλιόμετρα.



Εικόνα 20 Χρονοσειρά Ημερήσιας Απορροής Kinchafoonee creek

Είναι ένας ποταμός μήκους 147.7 χιλιομέτρων που πηγάζει κοντά στην πόλη Buena Vista στη πολιτεία Georgia και καταλήγει στον ποταμό Flint με την ένωση να γίνεται στο Albany. Όσον αφορά τη γεωλογία αποτελείται από μη συγκεκολλημένα ιζήματα στο 90% του και από πυριτοκλαστικά ιζηματογενή στο 8%. Η κυριότερη χρήση γης είναι σαβάνα με αρκετά δέντρα σε ποσοστό 50% (MODIS Land Cover Type/Dynamics.nasa, 2022).

3 Θεωρητικό υπόβαθρο μεθόδων

3.1 Εκθετική ομαλοποίηση

Η εκθετική ομαλοποίηση χρησιμοποιείται συχνά για την εξομάλυνση διακριτών χρονοσειρών με στόχο τη μελλοντική τους πρόγνωση. Στην εκθετική ομαλοποίηση γίνεται η παραδοχή ότι οι πιο πρόσφατες τιμές της χρονοσειράς είναι και αυτές που έχουν μεγαλύτερη επίδραση στην πρόγνωση σε σχέση με τις παλαιότερες. Στην ουσία είναι μια διαδικασία κατασκευής ενός εκθετικού μέσου όρου με φθίνοντες συντελεστές βάρους για τις παρελθοντικές παρατηρήσεις ενώ ο μεγαλύτερος συντελεστής βάρους δίνεται στην πιο πρόσφατη παρατήρηση.

3.1.1 Απλή εκθετική ομαλοποίηση

Στην περίπτωση της απλής εκθετικής ομαλοποίησης σε μια δεδομένη χρονοσειρά y_t η εξίσωση ομαλοποίησης έχει τη μορφή:

$$\hat{y}_{t+1} = a * y_t + (1-a) * \hat{y}_t \quad (3.1)$$

Στην απλή εκθετική ομαλοποίηση τα δεδομένα δεν πρέπει να παρουσιάζουν ανοδική ή καθοδική τάση (Rob J. Hyndman, 2008). Οι τιμές \hat{y}_{t+1} , \hat{y}_t σημαίνουν ότι υπάρχει πρόγνωση των τιμών. Αν ο όρος \hat{y}_t αντικατασταθεί με την αντίστοιχη εξίσωση μέσω της αναδρομής φαίνεται πως η πρόγνωση του \hat{y}_{t+1} είναι οι παρελθοντικές τιμές της χρονοσειράς μειωμένες με βήμα a κάθε φορά. Η εξίσωση αυτή ονομάζεται και 'επίπεδο' καθώς προσπαθεί να εντοπίσει μια τιμή γύρω από την οποία κυμαίνονται οι τιμές της χρονοσειράς, σαν κινητός μέσος όρος. Συχνά στη βιβλιογραφία υπάρχει αυτός ο συμβολισμός $l_t = \hat{y}_{t+1}$ με το l να προκύπτει από το αγγλικό level(=επίπεδο).

3.1.2 Διπλή εκθετική ομαλοποίηση

Η βασική ιδέα της διπλής εκθετικής ομαλοποίησης είναι να λαμβάνει υπόψη της την πιθανότητα οι χρονοσειρές να εμφανίζουν ανοδικές ή καθοδικές τάσεις (Rob J. Hyndman, 2008). Η διπλή εκθετική ομαλοποίηση ονομάζεται αλλιώς 'Holt's Linear Trend Model' από τον Holt που αναβάθμισε την απλή εκθετική ομαλοποίηση το 1957. Η εξίσωση είναι:

$$\hat{y}_{t+1} = l_t + b_t \quad (3.2)$$

Ο όρος l_t είναι ένας το 'επίπεδο' που αναφέρθηκε στην απλή εκθετική ομαλοποίηση και ο όρος b_t είναι η τάση. Ο όρος l_t πλέον πρέπει να τροποποιηθεί ελαφρώς για να αφαιρεθεί η παράμετρος της τάσης (να γίνει detrending όπως αναφέρεται στην αγγλική βιβλιογραφία). Ο όρος b_t είναι πρακτικά μια νέα εκθετική ομαλοποίηση στην οποία αφαιρείται η παράμετρος του παρελθοντικού 'επιπέδου' από το υπάρχον 'επίπεδο'. Αυτές οι αφαιρέσεις έχουν ως σκοπό το να μη μοντελοποιηθεί καμία παράμετρος δύο φορές. Έτσι συγκεντρωτικά οι εξισώσεις είναι:

- $l_t = a * y_t + (1-a) * (l_{t-1} - b_{t-1})$ (a)
- $b_t = \beta * (l_t - l_{t-1}) + (1-\beta)*b_{t-1}$ (b)

3.1.3 Τριπλή εκθετική ομαλοποίηση

Έπειτα από τον Holt, ο Winters πήρε το μοντέλο της διπλής εκθετικής ομαλοποίησης και του προσέθεσε τον όρο της εποχικότητας (Rob J. Hyndman, 2008). Η εξίσωση συνοπτικά είναι:

$$\hat{y}_{t+1} = l_t + b_t + s_{t-mk} \quad (3.3)$$

Όπου m είναι η περίοδος της εποχικότητας στην περίπτωση της παρούσας εργασίας είναι 365 ημέρες. Το k είναι μια παράμετρος που δείχνει το πόσο πολύ πρέπει να 'αλλάξει' η περίοδος για να υπολογιστεί η νέα τιμή. Η εξίσωση του k είναι $k = \text{floor}(h-1/m) - 1$.

Ο όρος είναι s_{t-mk} ορίζεται ως την εποχικότητα και η εξίσωσή της δίνεται στη συνέχεια.

$$s_t = \gamma * (y_t - l_{t-1} - b_{t-1}) + (1-\gamma)*s_{t-m} \quad (c)$$

Η εξίσωση c έρχεται να συμπληρώσει τις a , b και να αποτελέσουν την τριπλή εκθετική ομαλοποίηση.

3.2 Τα μοντέλα ARIMA

Τα μοντέλα ARIMA είναι μια ευρύτερη οικογένεια μοντέλων και τα αρχικά τους προέρχονται από το Autoregressive (AR) Integrated (I) Moving average (MA) (George E. P. Box, 2015). Στην ουσία αποτελούν συνδυασμό των ανωτέρω μεθόδων οι οποίες θα εξηγηθούν αναλυτικότερα παρακάτω. Τα μοντέλα ARIMA είναι ιδιαίτερα αποτελεσματικά στο να εκτιμούν μεταβλητές που έχουν μια εξάρτηση από παρελθοντικά χρονικά δεδομένα. Μια τέτοια μεταβλητή είναι και στην περίπτωση μας η απορροή στην οποία χρησιμοποιούνται ημερήσια δεδομένα παλαιότερων απορροών για να εκτιμηθούν ανάλογα με τον ορίζοντα πρόγνωσης οι επόμενες 1, 5 ή 10 μέρες. Τα μοντέλα έχουν σαν βασική παραδοχή (K.W. Hipel, 1994) ότι κάθε χρονοσειρά μπορεί να εκφραστεί από:

- Έναν συνδυασμό από τις προηγούμενες τιμές(AR)
- Τη διαφορά μεταξύ τους (I)
- Έναν όρο τυχαίου σφάλματος (MA)

3.2.1 Τα επιμέρους AR, I, MA

AR(p) Μοντέλο Αυτοπαλινδρόμησης τάξης p

Το AR(autoregressive) πρώτο σκέλος του ARIMA είναι το μοντέλο αυτοπαλινδρόμησης. Τα μοντέλα αυτοπαλινδρόμησης είναι στατιστικά μοντέλα που χρησιμοποιούνται σε πολλούς κλάδους όπως στα οικονομικά, σε συμπεριφορές καταναλωτών και σε υδρολογικά προβλήματα που είναι και το αντικείμενο της παρούσας εργασίας. Το μοντέλο δηλώνει την εξάρτηση της παρούσας τιμής της χρονοσειράς από τις παρελθοντικές τιμές της. Είναι πρακτικά σαν ένα μοντέλο πολυωνυμικής παλινδρόμησης που υπάρχει η εξαρτημένη μεταβλητή $y(t)$ και υπάρχουν και αντί για τις ανεξάρτητες μεταβλητές υπάρχουν παρελθοντικές τιμές της ίδιας εξαρτημένης μεταβλητής. Πιο αναλυτικά η εξίσωση είναι:

$$y(t) = b + \varphi_1 * y(t-1) + \varphi_2 * y(t-2) + \varphi_3 * y(t-3) + \dots \varphi_p * y(t-p) + \varepsilon(t) \quad (3.4)$$

(Pankratz, 1983)

$$y(t) = b + \varphi_1 * y(t-1) + \varphi_2 * y(t-2) + \varphi_3 * y(t-3) + \dots \varphi_p * y(t-p) + \varepsilon(t) \quad (3.5)$$

(Pankratz, 1983)

Για να εκτιμηθεί η τιμή της απορροής τη χρονική στιγμή t χρησιμοποιούνται οι τιμές όσων προγενέστερων χρονικών στιγμών κριθεί απαραίτητο και για αυτό έχει χρησιμοποιηθεί ο όρος p πρακτικά η ανωτέρω εξίσωση προσδιορίζεται από αυτό το p , συμβολίζεται ως AR(p) και ονομάζεται 'μοντέλο αυτοπαλινδρόμησης τάξης p '. Πέραν των όρων απορροής υπάρχει και ο σταθερός όρος b όπως σε κάθε παλινδρόμηση που είναι ανεξάρτητος και μπορεί να λάβει μια συγκεκριμένη τιμή ανάλογα με τα δεδομένα. Τέλος, υπάρχει ο όρος $\varepsilon(t)$ που συμβολίζει το τυχαίο σφάλμα, ονομάζεται λευκός θόρυβος και λαμβάνει τυχαίες τιμές από μία κατανομή και στην περίπτωση της εργασίας από την κανονική κατανομή με μέση τιμή το 0 και τυπική απόκλιση την σ^2 . Στην παρούσα εργασία οι διαδικασίες ARIMA υλοποιούνται μέσα από το πακέτο statsmodels της Python. Οι τιμές των b και $\varphi_{1..p}$ καθορίζονται από τα ίδια τα δεδομένα που έχουν ήδη δοθεί στο μοντέλο και για να βρεθεί η καταλληλότερη τιμή το statsmodels χρησιμοποιεί την Εκτίμηση Μέγιστης Πιθανοφάνειας (EMΠ) (statsmodels, 2023) γίνεται δε αναλυτική επεξήγηση της μεθόδου παρακάτω.

I(d). Μοντέλο διαφοροποίησης τάξης (d)

Όπως αναφέρθηκε στο κεφάλαιο των χρονοσειρών ένα από τα χαρακτηριστικά μιας χρονοσειράς είναι η στασιμότητα, με τις μεθόδους AR και MA δεν είναι δυνατό να υπάρξει πρόγνωση αν η χρονοσειρά είναι μη στάσιμη καθώς υπάρχει κινητή τυπική απόκλιση και μέση τιμή. Ο σκοπός της διαφοροποίησης είναι να μετατραπεί μια χρονοσειρά από μη στάσιμη σε στάσιμη. Αυτό επιτυγχάνεται με τη διαφοροποίηση.

Πρακτικά η διαφοροποίηση είναι μια απλή αφαίρεση της προηγούμενης τιμής από την παρούσα τιμή έτσι δημιουργείται μια νέα διαφοροποιημένη χρονοσειρά που πλέον μπορούν να εφαρμοστούν μέθοδοι πρόγνωσης. Η μαθηματική διατύπωση είναι:

$$Y(t) = X(t) - X(t-1)$$

Αν μετά τη διαφοροποίηση η χρονοσειρά συνεχίσει να είναι μη στάσιμη τότε επαναλαμβάνεται η διαφοροποίηση μέχρι η χρονοσειρά να είναι στάσιμη. Στη συγκεκριμένη εργασία χρησιμοποιήθηκε μόνο μια διαφοροποίηση.

Αξίζει να σημειωθεί πως υπάρχει και εποχική διαφοροποίηση.

Η τάξη της διαφοροποίησης συμβολίζεται με το γράμμα d εξ' ου και I(d) προκειμένου να φανεί πόσες διαφοροποιήσεις χρειάζονται.

MA(q). Μοντέλο κινητού μέσου όρου

Το μοντέλο κινητού μέσου όρου καθορίζει ότι η τρέχουσα τιμή μιας χρονοσειράς εξαρτάται από μια σειρά παρελθοντικών τυχαίων σφαλμάτων και τον μέσο όρο της χρονοσειράς. Σε μαθηματική διατύπωση έχει ως εξής:

$$Y_t = c + \varepsilon_t + \theta_1 * \varepsilon_{t-1} + \theta_2 * \varepsilon_{t-2} + \dots + \theta_q * \varepsilon_{t-q} \quad (3.6)$$

Σε αυτή την εξίσωση υπάρχουν μέχρι q τυχαία σφάλματα και επομένως είναι q τάξης, συμβολίζεται ως MA(q). Ο όρος c είναι στην πραγματικότητα η αναμενόμενη τιμή της χρονοσειράς και αυτό φαίνεται καλύτερα αν στην ανωτέρω εξίσωση πάρουμε και στα δυο μέλη τις αναμενόμενες τιμές τότε θα έχουμε:

$$E(Y_t) = E(c + \varepsilon_t + \theta_1 * \varepsilon_{t-1} + \theta_2 * \varepsilon_{t-2} + \dots + \theta_q * \varepsilon_{t-q}) \quad (3.7)$$

Επειδή οι όροι των τυχαίων σφαλμάτων ακολουθούν κανονική κατανομή με μέση τιμή 0 και τυπική απόκλιση σ^2 , $\varepsilon \sim N(0, \sigma^2)$ η μέση τιμή των σφαλμάτων είναι 0 και επομένως η εξίσωση παίρνει τη μορφή

$$E(Y_t) = E(c) = c \quad (3.8)$$

και επομένως μπορεί να υποστηριχθεί πως ο όρος c που είναι μια σταθερά είναι στην πραγματικότητα ο μέσος όρος της ίδιας της χρονοσειράς. Έτσι το μοντέλο κινητού μέσου όρου είναι πρακτικά ένας μέσος όρος ακολουθούμενος από μια σειρά με τυχαία σφάλματα των προηγούμενων τιμών. Το μοντέλο αυτό επικεντρώνεται στη αναπαράσταση των σχέσεων των κοντινών τιμών μεταξύ τους καθώς εμπλέκει τους παρελθοντικούς λευκούς θορύβους και μεταβάλλοντας το μέσο όρο από τη μια τιμή στην αμέσως επόμενη, είναι ιδιαίτερα καλό στο να αναγνωρίζει μεγάλες μεταβολές στις χρονοσειρές και να τις ενσωματώνει στην πρόγνωση.

3.2.2 Τέστ για την επιλογή των παραμέτρων ρ , d , q

Όπως έχει ήδη αναφερθεί ανωτέρω οι παράμετροι αυτοί δηλώνουν την επιμέρους τάξη του κάθε μοντέλου. Για να επιλεγούν πρέπει να προσδιοριστούν κάποια βασικά χαρακτηριστικά της χρονοσειράς για αυτό το λόγο εφαρμόστηκαν τέστ που προσδιόρισαν για κάθε χρονοσειρά για όλες τις λεκάνες:

- τη στασιμότητα
- την αυτοσυσχέτιση
- τη μερική αυτοσυσχέτιση

Στη συνέχεια παρατίθενται περισσότερες πληροφορίες για τα προαναφερθέντα αυτά στοιχεία.

3.2.2.1 Στασιμότητα

Μια σημαντική έννοια για την κατανόηση της ανάλυσης των χρονοσειρών είναι η στασιμότητα (D, 1994). Η στασιμότητα διακρίνεται σε δύο κατηγορίες:

- Αυστηρή στασιμότητα. Μπορούμε να διακρίνουμε την αυστηρή στασιμότητα όταν η συνδυασμένη πιθανοτική κατανομή των τιμών της $y_t, y_{t+1}, + \dots y_{t+n}$ είναι σταθερή και ανεξάρτητη από το χρόνο. Αυτό σημαίνει πως σε οποιαδήποτε χρονική στιγμή όλα τα στατιστικά χαρακτηριστικά της χρονοσειράς είναι σταθερά. Πιο συγκεκριμένα, η μέση τιμή, η διασπορά, η ασυμετρία (skewness) και η κύρτωση δεν αλλάζουν σε οποιοδήποτε διάστημα και αν εξετάσουμε τη χρονοσειρά. Μια χρονοσειρά είναι δύσκολο να παρουσιάσει αυστηρή στασιμότητα και δεν συναντάται συχνά.

- Μη αυστηρή στασιμότητα. Είναι ο πιο συνηθισμένος τύπος στασιμότητας και καλείται και στασιμότητα δευτέρας τάξης. Σε αυτές τις περιπτώσεις η χρονοσειρά έχει σταθερά μόνο τα δύο πρώτα στατιστικά χαρακτηριστικά δηλαδή τη μέση τιμή και την διασπορά.

Στην παρούσα εργασία για να εξεταστεί αν η χρονοσειρά είναι στάσιμη σε κάθε μια από τις 10 λεκάνες έγινε χρήση του augmented Dickey Fuller(ADF) τεστ.

3.2.2.1.1 Dickey fuller test

Το Dickey Fuller τεστ είναι ένα πολύ σημαντικό τέστ για τις χρονοσειρές γιατί αναγνωρίζει αν μια χρονοσειρά είναι στάσιμη (Dickey & Fuller, 1979). Αυτό είναι χρήσιμο γιατί πολλά μοντέλα ανάλυσης χρονοσειρών θεωρούν τους ότι η χρονοσειρά είναι στάσιμη.

Τα τέστ αυτά έχουν συνήθως μια αρχική υπόθεση που θεωρείται ως αληθής και ονομάζεται null hypothesis H_0 . Για το dickey fuller η υπόθεση είναι ότι το μοντέλο AR(1)

$$y(t) = b + \varphi * y(t-1) + \varepsilon(t) \quad (3.9)$$

Έχει μοναδιαίες ρίζες επομένως το $\varphi = 1$ ή $\varphi = -1$ και επομένως όπως εξηγείται και στο κεφάλαιο μοναδιαίες ρίζες η χρονοσειρά δεν είναι στάσιμη.

Η εναλλακτική υπόθεση γνωστή στη στατιστική ως alternative hypothesis είναι ότι το μοντέλο δεν έχει μοναδιαίες ρίζες και οι απόλυτες τιμές του φ είναι μικρότερες του 1, $|\varphi| < 1$ και επομένως η χρονοσειρά είναι στάσιμη.

Στη συνέχεια το τέστ αφαιρεί και από τα δύο μέλη της εξίσωσης τον όρο $y(t-1)$. Έτσι προκύπτει:

$$y(t)-y(t-1) = b + (\varphi-1) * y(t-1) + \varepsilon(t) \quad (3.10)$$

Το οποίο διαφορετικά γράφεται:

$$\Delta y(t) = b + \delta * y(t-1) + \varepsilon(t) \quad (3.11)$$

Βλέποντας πως εκφράζεται η null hypothesis με τη εξίσωση 2 είναι ευκολα αντιληπτό πως πρέπει το $\delta = 0$ και αντίστοιχα στην εναλλακτική πρέπει το $\delta < 0$.

Ο λόγος για τον οποίο κάνουμε αυτή την αφαίρεση είναι γιατί στα πλαίσια της null hypothesis έχουμε δημιουργήσει μια στάσιμη χρονοσειρά καθώς το δ είναι μηδέν και επομένως ο όρος $y(t-1)$ που είναι και ο μη στάσιμος όρος έχει πλέον μηδενιστεί. Δεν θα μπορούσαμε να εφαρμόσουμε ένα T τεστ για το δ έτσι όπως είναι η χρονοσειρά γιατί όπως αναφέρθηκε και πάνω το δ αποτελεί συντελεστή ενός μη στάσιμου όρου αντί αυτού εφαρμόζουμε ένα ειδικό τεστ που είναι παρόμοιο με το T τεστ αλλά αντί για την T κατανομή το συγκρίνουμε με μια τοποποιημένη κατανομή που ονομάζεται dickey fuller κατανομή. Επομένως υπολογίζουμε το:

$$t_{\delta'} = \delta' / s.e.(\delta')$$

Αφού υπολογιστεί ο όρος αυτός θα το συγκριθεί με την κρίσιμη τιμή της dickey fuller κατανομής και αν $t_{\delta'} < DF_{\text{Critical}}$ τότε θα απορριφθεί η null hypothesis που σημαίνει ότι απορρίπτεται το γεγονός ότι η χρονοσειρά έχει μοναδιαία ρίζα και επομένως γίνεται δεκτό ότι η χρονοσειρά είναι στάσιμη. Αν από την άλλη $t_{\delta'} > DF_{\text{Critical}}$ τότε γίνεται δεκτή η null hypothesis ότι οι δύο κατανομές είναι στατιστικά σημαντικές και επομένως η χρονοσειρά δεν μπορεί να θεωρηθεί στάσιμη καθώς το $\phi = 1$ ή $\phi = -1$ και όπως θα αναφερθεί στις μοναδιαίες ρίζες δεν υπάρχει στασιμότητα.

3.2.2.1.2 Μοναδιαίες Ρίζες

Στην ανάλυση χρονοσειρών οι μοναδιαίες ρίζες είναι ένα χαρακτηριστικό που βοηθά στον εντοπισμό του αν μια χρονοσειρά είναι στάσιμη ή όχι (Dolado, Jenkinson, & Sosvilla-Rivero, 1990).

Έστω ότι έχουμε ένα AR(1) μοντέλο, που εξηγείται στο κεφάλαιο που αναπτύσσεται η ARIMA οικογένεια μοντέλων. Το μοντέλο αυτό έχει εξίσωση:

$$y(t) = \phi * y(t-1) + \varepsilon(t) \quad (3.12)$$

Αν λάβουμε την αντιστοιχη εξίσωση για το $y(t-1)$ και την αντικαταστήσουμε στην ανωτέρω εξίσωση χρησιμοποιώντας την αναδρομή θα καταλήξουμε να μπορούμε να εκφράσουμε ένα AR μοντέλο σαν ένα μοντέλο MA με όρους ε_i όσοι είναι και οι όροι της χρονοσειράς.

$$y(t) = \phi * [\phi * y(t-2) + \varepsilon(t-1)] + \varepsilon(t) \quad (3.13)$$

Πολλαπλασιάζοντας τον όρο φ με τους όρους εντός της αγκύλης προκύπτει η εξίσωση:

$$y(t) = (\varphi)^2 * y(t-2) + \varphi * \varepsilon(t-1) + \varepsilon(t) \quad (3.14)$$

Δημιουργείται λοιπόν ύστερα από πολλές αντικαταστάσεις το μοντέλο MA με τη συμπυκνωμένη μορφή του να είναι:

(3.15)

$$y(t) = \varphi^t * y_0 + \sum_{k=[0,t-1]}(\varphi^k * \varepsilon_{t-k}) \quad (3.15)$$

Εξετάζοντας τα δύο βασικά χαρακτηριστικά της στασιμότητας την αναμενόμενη τιμή $E(x)$ και διασπορά $Var(x)$ προκύπτουν τα παρακάτω σενάρια:

Για την αναμενόμενη τιμή $E(x)$ έχουμε από τη σχέση (3.15) αν εφαρμόσουμε την αναμενόμενη τιμή και στα δύο μέλη της εξίσωσης θα καταλήξουμε στο:

$$E[y(t)] = E[\varphi^t * y_0 + \sum_{k=[0,t-1]}(\varphi^k * \varepsilon_{t-k})]$$

Είναι όμως γνωστό(αναφέρεται στο κεφάλαιο που επεξηγείται το ARIMA) ότι τα τυχαία σφάλματα ε ακολουθούν κανονική κατανομή με μέση τιμή 0 και τυπική απόκλιση σ (Pankratz, 1983). Έτσι, η παραπάνω εξίσωση παίρνει τη μορφή καθώς $E[\varepsilon] = 0$:

$$E[y(t)] = E[\varphi^t * y_0]$$

Και γνωρίζοντας ότι ο όρος φ είναι μια σταθερά υψωμένη στη δύναμη t η σχέση γίνεται:

$$E[y(t)] = \varphi^t * E[y_0] = \varphi^t * y_0 \quad (A)$$

Στη συνέχεια για την εύρεση της διασποράς $Var(x)$ της χρονοσειράς στην εξίσωση (3.15) πάλι θα πάρουμε τη διασπορά και στα δύο μέλη.

$$Var[y(t)] = Var[\varphi^t * y_0 + \sum_{k=[0,t-1]}(\varphi^k * \varepsilon_{t-k})] \quad (3.16)$$

Το $\varphi^t * y_0$ όπως αναφέρθηκε ήδη και παραπάνω είναι σταθερός όρος οπότε η διασπορά της χρονοσειράς θα προκύψει μόνο από τον όρο που βρίσκεται μέσα στο άθροισμα Σ . Επομένως, θα μπορούσαμε να βγάλουμε ως κοινό παράγοντα το σ^2 και εντός αγκύλης θα έμεναν οι διαφορετικές δυνάμεις του φ που στην ουσία αποτελούν τους συντελεστές (και επομένως σταθερούς όρους με $\text{Var}(\varphi) = 0$) των εκάστοτε ε .

Πλέον η εξίσωση (5) θα λάβει τη μορφή:

$$\text{Var}[y(t)] = \sigma^2 * [\varphi^0 + \varphi^2 + \varphi^4 \dots + \varphi^{2(t-1)}] \quad (\text{B})$$

Με τα παραπάνω βήματα έχει υπολογιστεί για κάθε χρονική στιγμή t για τη χρονοσειρά η αναμενόμενη τιμή και η διασπορά. Στη συνέχεια το πρόβλημα θα χωριστεί σε 3 περιπτώσεις και θα κριθεί αν η χρονοσειρά είναι στάσιμη ή όχι.

- Περίπτωση 1 $|\varphi| < 1$:
 - Η αναμενόμενη τιμή βρίσκεται από την εξίσωση Α. Αν υπολογίσουμε το φ^t με το $t \rightarrow \infty$ τότε είναι γνωστό πως το όριο αυτό θα γίνει στο 0 καθώς έτσι συμπεριφέρεται οποιοσδήποτε αριθμός μεταξύ $[-1, 1]$ που υψώνεται σε δύναμη που τείνει στο ∞ . Άρα $E(x) = 0$.
 - Η διασπορά είναι όπως φαίνεται στην εξίσωση Β. Διακρίνεται ο όρος εντός παρένθεσης που είναι μια γεωμετρική ακολουθία

$$\varphi^0 + \varphi^2 + \varphi^4 \dots + \varphi^{2(t-1)}$$

είναι γνωστό πως όταν σε μια τέτοια ακολουθία ο όρος φ^2 είναι μικρότερος από 1 τότε μπορούμε να αθροίσουμε όλη την ακολουθία. Στην περίπτωση μας είναι μικρότερη από 1 καθώς οποιοσδήποτε αριθμός μικρότερος της μονάδας υψωμένος στο δύναμη του 2 δίνει πάλι αριθμό μικρότερο της μονάδας. Επομένως, η διασπορά είναι:

$$\text{Var}[y(t)] = \sigma^2 / (1 - \varphi^2) \quad (\text{Γ})$$

Έτσι, φαίνεται πως και ο όρος αυτός είναι σταθερός, επομένως η διασπορά είναι σταθερή και επομένως η χρονοσειρά είναι στάσιμη.

- Περίπτωση 2 $|\varphi| > 1$:

Σε αυτή την περίπτωση προκύπτει πολύ εύκολα από την εξίσωση 1 ότι κάθε επόμενος όρος της χρονοσειράς θα είναι επαυξημένος (κατ' απόλυτη τιμή) από τον προηγούμενο όρο επομένως είναι σαφές ότι η χρονοσειρά αυτή είτε θα αυξάνεται είτε θα μειώνεται συνέχεια, άρα η χρονοσειρά δεν είναι στάσιμη.

Επίσης, αν λάβουμε την αναμενόμενη τιμή από την εξίσωση Α διαπιστώνουμε εύκολα πως όταν το $t \rightarrow \infty$ τότε και ο όρος αυτός τείνει στο ∞ .

- Περίπτωση 3 $|\varphi| = 1$:

Όταν σε ένα AR μοντέλο ικανοποιείται η περίπτωση 3 το ϕ δηλαδή παίρνει τιμές $\phi = -1$ ή $\phi = 1$ τότε λέμε ότι το μοντέλο έχει μοναδιαίες ρίζες. Αυτό είναι ένα πολύ βασικό χαρακτηριστικό με βάση το οποίο λειτουργεί και το augmented dickey fuller τεστ που χρησιμοποιήσαμε.

- Η αναμενόμενη τιμή από την εξίσωση A με $\phi = 1$ προκύπτει ίση με το y_0 και επομένως ικανοποιεί το κριτήριο στασιμότητας.
- Η διασπορά από την εξίσωση B αθροίζοντας όλες τις μονάδες, καταλήγουμε στην απλοποιημένη μορφή:

$$Var[y(t)] = t * \sigma^2 (\Delta)$$

Όπως διακρίνεται από εδώ η διασπορά δεν είναι σταθερή καθώς όσο πλησιάζουμε σε μεγαλύτερα t συνέχεια πολλαπλασιάζεται και ο όρος σ^2 με αποτέλεσμα να μεγαλώνει η διασπορά συνεχώς. Πρακτικά ενώ η αναμενόμενη τιμή είναι ίδια η αύξηση της διασποράς κάνει δύσκολη την πρόγνωση. Οι χρονοσειρές με μοναδιαίες ρίζες είναι αρκετά δύσκολο να χαρακτηριστούν ως στάσιμες αλλά μέσω διαγράμματος και για αυτό το λόγο στην παρούσα εργασία εφαρμόστηκε το augmented dickey fuller τεστ που αναφέρθηκε στο προηγούμενο υποκεφάλαιο.

3.2.2.2 Συναρτήσεις Αυτοσυσχέτισης και Μερικής Αυτοσυσχέτισης

Για να μπορέσει να εκτιμηθεί το εύρος στο οποίο θα κυμαίνονται οι τιμές των παραμέτρων p και q χρησιμοποιήθηκαν διαγράμματα για τις συναρτήσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης. Η συνάρτηση αυτοσυσχέτισης ορίζει τη σχέση που έχουν τα σημεία μιας χρονοσειράς όσον αφορά την πρόγνωση κάθε νέας τιμής της χρονοσειράς (George E. P. Box, 2015). Για να αναπαρασταθεί αυτό κατασκευάστηκαν διαγράμματα που παρατίθενται στο κεφάλαιο επεξήγησης του κώδικα. Η συνάρτηση μερικής αυτοσυσχέτισης υπολογίζει τη συσχέτιση που έχουν δύο όροι μιας χρονοσειράς όταν αφαιρεθεί η επίδραση των ενδιάμεσων όρων. Ομοίως με τη συνάρτηση αυτοσυσχέτισης τα διαγράμματα μερικής αυτοσυσχέτισης παρατίθενται στο κεφάλαιο επεξήγησης του κώδικα. Η συνάρτηση αυτοσυσχέτισης υποδεικνύει τον όρο p , ενώ η συνάρτηση μερικής αυτοσυσχέτισης υποδεικνύει τον όρο q .

3.2.3 Τελική επιλογή παραμέτρων

Στη συνέχεια λαμβάνοντας υπόψη τα τεστ που πραγματοποιήθηκαν αλλά και την υπολογιστική δύναμη και γνωρίζοντας ότι ένα μοντέλο που προωθεί την πολυπλοκότητα δεν είναι απαραίτητα καλό ορίστηκαν κάποιες μέγιστες τιμές q , d , p και σε όλο το εύρος των τιμών αυτών πραγματοποιήθηκε αναζήτηση πλέγματος για όλους τους δυνατούς συνδυασμούς που δημιουργούνται.

Το εύρος λοιπόν που ορίστηκε να λάβει η κάθε υπερπαραμέτρος είναι:

- p , 0-2
- d , 0-1

- q, 0-2

Παρακάτω αναφέρονται όλα τα μοντέλα που προέκυψαν από αυτή την αναζήτηση πλέγματος.

3.2.4 Όλα τα μοντέλα ARIMA που χρησιμοποιήθηκαν

1. MA(1) Κινητού μέσου όρου, $p = 0, d = 0, q = 1$

Το μοντέλο έχει εξίσωση:

$$Y_t = c + \varepsilon_t + \theta_1 * \varepsilon_{t-1} \quad (3.17)$$

2. MA(2) Κινητού μέσου όρου, $p = 0, d = 0, q = 2$

Το μοντέλο έχει την εξίσωση:

$$Y_t = c + \varepsilon_t + \theta_1 * \varepsilon_{t-1} + \theta_2 * \varepsilon_{t-2} \quad (3.18)$$

3. I(d) Μοντέλο διαφοροποίησης, $p = 0, d = 1, q = 0$

Το μοντέλο προκύπτει από διαφοροποίηση μια φορά των όρων των χρονοσειρών, ενώ δεν προέκυψε από το augmented dickey fuller test (το οποίο όπως έχει αναφερθεί είναι το τεστ που χρησιμοποιηθεί για να ελέγξει τη στασιμότητα) κάποια τάση (trend) το μοντέλο διαφοροποίησης συμπεριλήφθηκε γιατί ένα seasonal component είναι πολύ πιθανό να υπάρχει και επίσης συμπεριλήφθηκε για λόγους πληρότητας.

Με εξίσωση:

$$Y(t) = X(t) - X(t-1) \quad (3.19)$$

4. ARIMA(0,1,1) Μοντέλο διαφοροποίησης και κινητού μέσου όρου, $p = 0, d = 1, q = 1$

Πρώτα λαμβάνει χώρα η διαφοροποίηση και έπειτα εφαρμόζεται το μοντέλο κινητού μέσου όρου όπως αναφέρθηκε και πιο πάνω.

Βήμα 1:

$$Y(t) = X(t) - X(t-1)$$

Βήμα 2:

$$Y_t = c + \varepsilon_t + \theta_1 * \varepsilon_{t-1}$$

5. ARIMA(0,1,2) Μοντέλο διαφοροποίησης και κινητού μέσου όρου, $p = 0$, $d = 1$, $q = 2$

Παρόμοιο με το προηγούμενο απλά το μοντέλο του κινητού μέσου όρου που εφαρμόζεται στο Βήμα 2 είναι δεύτερης τάξης.

Βήμα 1:

$$Y(t) = X(t) - X(t-1)$$

Βήμα 2:

$$Y_t = c + \varepsilon_t + \theta_1 * \varepsilon_{t-1} + \theta_2 * \varepsilon_{t-2}$$

6. AR(1) Μοντέλο αυτοπαλινδρόμησης, $p = 1$, $d = 0$, $q = 0$

Είναι το πρώτο μοντέλο που έχει εξηγηθεί και πρώτης τάξης

Με εξίσωση:

$$y(t) = b + \varphi_1 * y(t-1) + \varepsilon(t) \quad (3.20)$$

7. ARMA(1,1) Μοντέλο αυτοπαλινδρόμησης και κινητού μέσου όρου, $p = 1$, $d = 0$, $q = 1$

Είναι ένας συνδυασμός των επιμέρους μοντέλων.

Με εξίσωση:

$$y(t) = b + \varphi_1 * y(t-1) + \varepsilon_t + \theta_1 * \varepsilon_{t-1} \quad (3.21)$$

Εδώ αξίζει να σημειωθεί πως πιο πάνω φαίνονται 2 όροι τυχαίου σφάλματος ε_t επειδή όμως το νοήμα του τυχαίου σφάλματος για τη χρονική στιγμή t είναι να προσφέρει μια στοχαστικότητα δε χρειάζεται αυτός ο όρος να προστεθεί 2 φορές και έτσι το ε_t αναγράφεται

8. ARMA(1,2) Μοντέλο αυτοπαλινδρόμησης και κινητού μέσου όρου, $p = 1, d = 0, q = 2$
 Ομοίως με το πάνω παράδειγμα με διαφορά στην τάξη του μοντέλου κινητού μέσου όρου.
 Με εξίσωση:

$$y(t) = b + \varphi_1 * y(t-1) + \varepsilon_t + \theta_1 * \varepsilon_{t-1} + \theta_2 * \varepsilon_{t-2} \quad (3.22)$$

9. ARIMA(1,1,0) Μοντέλο διαφοροποίησης και αυτοπαλινδρόμησης, $p = 1, d = 1, q = 0$

Το μοντέλο αυτό συνδυάζει τη διαφοροποίηση η οποία χρονικά σαν διαδικασία προηγείται και επί της διαφοροποιημένης χρονοσειράς εφαρμόζεται η αυτοπαλινδρόμηση πρώτης τάξης.

Βήμα 1:

$$Y(t) = X(t) - X(t-1)$$

Βήμα 2:

$$y(t) = b + \varphi_1 * y(t-1) + \varepsilon_t$$

10. ARIMA(1,1,1) Μοντέλο διαφοροποίησης, αυτόπαλινδρόμησης και κινητού μέσου όρου, $p = 1, d = 1, q = 1$.

Στο ολοκληρωμένο πλέον μοντέλο ARIMA στην αρχή γίνεται η διαφοροποίηση της χρονοσειράς και στη συνέχεια γίνεται η εφαρμογή του ARMA δηλαδή της αυτοπαλινδρόμησης και του μεταβλητού μέσου όρου.

Βήμα 1:

$$Y(t) = X(t) - X(t-1)$$

Βήμα 2:

$$y(t) = b + \varphi_1 * y(t-1) + \varepsilon_t + \theta_1 * \varepsilon_{t-1}$$

11. ARIMA(1,1,2) Μοντέλο διαφοροποίησης, αυτόπαλινδρόμησης και κινητού μέσου όρου, $p = 1, d = 1, q = 2$.

Όμοιο με το μοντέλο που αναφέρθηκε ανωτέρω με τη διαφορά ότι το μοντέλο κινητού μέσου όρου είναι δεύτερης τάξης.

Βήμα 1:

$$Y(t) = X(t) - X(t-1)$$

Βήμα 2:

$$y(t) = b + \varphi_1 * y(t-1) + \varepsilon_t + \theta_1 * \varepsilon_{t-1} + \theta_2 * \varepsilon_{t-2}$$

12. AR(2) Μοντέλο αυτοπαλινδρόμησης, $p = 2$, $d = 0$, $q = 0$

Με εξίσωση:

$$y(t) = b + \varphi_1 * y(t-1) + \varphi_2 * y(t-2) \quad (3.23)$$

13. ARMA(2,1) Μοντέλο αυτοπαλινδρόμησης και κινητού μέσου όρου, $p = 2$, $d = 0$, $q = 1$

Ο συνδυασμός των δυο μοντέλων έχει αναφερθεί ανωτέρω με τη διαφορά τώρα ότι το μοντέλο αυτοπαλινδρόμησης είναι δεύτερης τάξης.

Με εξίσωση:

$$y(t) = b + \varphi_1 * y(t-1) + \varphi_2 * y(t-2) + \varepsilon_t + \theta_1 * \varepsilon_{t-1} \quad (3.24)$$

14. ARMA(2,2) Μοντέλο αυτοπαλινδρόμησης και κινητού μέσου όρου, $p = 2$, $d = 0$, $q = 2$

Όπως αναφέρθηκε και παραπάνω με τη διαφορά ότι πλέον το μοντέλο κινητού μέσου όρου είναι δεύτερης τάξης.

Με εξίσωση:

$$y(t) = b + \varphi_1 * y(t-1) + \varphi_2 * y(t-2) + \varepsilon_t + \theta_1 * \varepsilon_{t-1} + \theta_2 * \varepsilon_{t-2} \quad (3.25)$$

15. ARIMA(2,1,0) Μοντέλο διαφοροποίησης και αυτοπαλινδρόμησης, $p = 2$, $d = 1$, $q = 0$

Όπως έχει προαναφερθεί πρώτα λαμβάνει χώρα η διαφοροποίηση και έπειτα η αυτοπαλινδρόμηση στη διαφοροποιημένη χρονοσειρά.

Βήμα 1:

$$Y(t) = X(t) - X(t-1)$$

Βήμα 2:

$$y(t) = b + \varphi_1 * y(t-1) + \varphi_2 * y(t-2)$$

16. ARIMA(2,1,1) Μοντέλο διαφοροποίησης, αυτόπαλινδρόμησης και κινητού μέσου όρου, $p = 2$, $d = 1$, $q = 1$.

Πρώτα εφαρμόζεται διαφοροποίηση και μετά στη διαφοροποιημένη χρονοσειρά εφαρμόζεται το μοντέλο ARMA(2,1).

Βήμα 1:

$$Y(t) = X(t) - X(t-1)$$

Βήμα 2:

$$y(t) = b + \varphi_1 * y(t-1) + \varphi_2 * y(t-2) + \varepsilon_t + \theta_1 * \varepsilon_{t-1}$$

17. ARIMA(2,1,2) Μοντέλο διαφοροποίησης, αυτόπαλινδρόμησης και κινητού μέσου όρου, $p = 2$, $d = 1$, $q = 2$.

Το τελευταίο και πιο πολύπλοκο μοντέλο ARIMA, είναι ίδιας λογικής με το ARIMA(2,1,1) που αναφέρθηκε ανωτέρω απλά το μοντέλο κινητού μέσου όρου είναι δεύτερης τάξης.

Βήμα 1:

$$Y(t) = X(t) - X(t-1)$$

Βήμα 2:

$$y(t) = b + \varphi_1 * y(t-1) + \varphi_2 * y(t-2) + \varepsilon_t + \theta_1 * \varepsilon_{t-1} + \theta_2 * \varepsilon_{t-2}$$

3.2.5 Μέθοδος μέγιστης πιθανοφάνειας

Η μέθοδος της μέγιστης πιθανοφάνειας χρησιμοποιείται για τον υπολογισμό θ μιας παραμέτρου από μια συνάρτηση κατανομής πιθανότητας. Αν υποθέσουμε ότι έχουμε X_n τιμές από μια τυχαία μεταβλητή δηλαδή είναι τιμές που προκύπτουν από ανεξάρτητα πειράματα της ίδιας κατανομής και επιθυμούμε να προβλέψουμε μια μελλοντική τιμή της X θα επιλέξουμε την τιμή που μεγιστοποιεί την πιθανότητα αυτή. Για να το επιτύχουμε δημιουργούμε τη γραφική παράσταση της συνάρτησης κατανομής της x , $f(X_\theta)$ η καμπύλη αυτή ονομάζεται συνάρτηση πιθανοφάνειας του θ . Η μέγιστη τιμή του θ στη συνάρτηση καλείται μέγιστη πιθανοφάνεια. Αυτή είναι και η τιμή που θέλουμε να βρεθεί.

Στην περίπτωση των τιμών μιας χρονοσειράς δεν μπορεί να εφαρμοστεί η ανωτέρω διαδικασία διότι οι τιμές δεν θεωρούνται ανεξάρτητες. Έτσι μέσω της δεσμευμένης πιθανότητας δημιουργείται ένα σύνολο δεδομένων που είναι ανεξάρτητα. Πρακτικά, θεωρείται το X της

αρχικής κατανομής δεδομένο και στη συνέχεια εκφράζεται συναρτήσει του w . Αυτό δημιουργεί w που είναι ανεξάρτητα και επομένως μπορεί να εφαρμοστεί η μέγιστη πιθανοφάνεια.

Ο υπολογισμός των παραμέτρων μιας χρονοσειράς γίνεται αφού βρούμε τις τιμές των παραμέτρων που μεγιστοποιεί τη συνάρτηση πιθανοφάνειας των δεδομένων τιμών. Για να υλοποιηθεί αυτό πρέπει αρχικά να επιλεγεί ένα μοντέλο, να ελεγχθεί ως προς τη στασιμότητα του και στη συνέχεια να υπολογιστούν οι παράμετροι της μέγιστης πιθανοφάνειας. Πολλές φορές όταν το μοντέλο περιλαμβάνει πολλές παραμέτρους είναι δύσκολο να βρεθούν αναλυτικές λύσεις της συνάρτησης μέγιστης πιθανοφάνειας τότε καταφεύγουμε σε αριθμητικές μεθόδους επίλυσης της συνάρτησης, μια τέτοια είναι η επαναληπτική μέθοδος Newton- Raphson που χρησιμοποιεί τη λογαριθμική συνάρτηση μέγιστης πιθανοφάνειας και την πρώτη και δεύτερη παράγωγο της.

3.3 Μέθοδοι βαθιάς μηχανικής μάθησης

3.3.1 Χρήσιμοι ορισμοί

Τεχνητή νοημοσύνη

Η τεχνητή νοημοσύνη εμφανίστηκε περί τα 1950 από την τότε αναπτυσσόμενη επιστήμη των ηλεκτρονικών υπολογιστών. Είχε ως στόχο να αυτοματοποιήσει λογικές διαδικασίες και να μιμηθεί την ανθρώπινη νοημοσύνη (Oxford Reference, n.d.). Γενικά, η τεχνητή νοημοσύνη θεωρείται ένα υπερσύνολο που περιέχει τη μηχανική μάθηση και τη βαθιά μηχανική μάθηση έννοιες που θα εξεταστούν στην παρούσα διπλωματική εργασία. Ένα χαρακτηριστικό παράδειγμα τεχνητής νοημοσύνης είναι ο προγραμματισμός ενός υπολογιστή να μάθει να παίζει σκάκι. Ο υπολογιστής μαθαίνει από μόνος πώς να πραγματοποιήσει μια συγκεκριμένη διαδικασία.

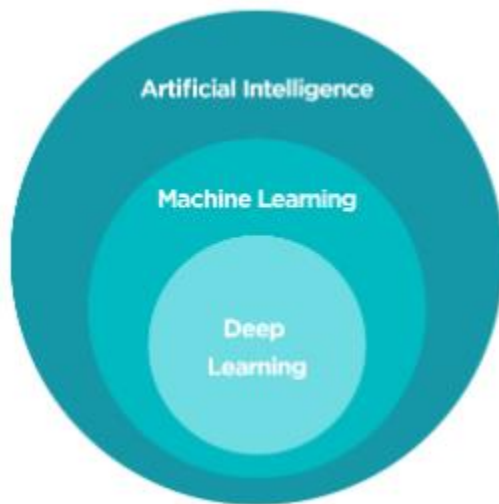
Μηχανική μάθηση

Η μηχανική μάθηση είναι ένας τομέας της τεχνητής νοημοσύνης στον οποίο οι υπολογιστές με χρήση μεγάλης ποσότητας δεδομένων μαθαίνουν πώς να εκτελούν οι ίδιοι διαδικασίες αντί να τους προγραμματίζουμε να τα εκτελέσουν (W.Trask, 2019). Εν ολίγοις οι υπολογιστές παρατηρούν ένα μοτίβο και προσπαθούν να το μιμηθούν είτε με άμεσο είτε με έμμεσο τρόπο. Η μηχανική μάθηση βασίστηκε στην επιστήμη της στατιστικής, λαμβάνοντας από αυτήν τους θεμελιώδεις κανόνες όμως στην εξέλιξη της διαφοροποιήθηκε με κύριο πλεονέκτημα να επεξεργάζεται σύνθετες και πολύ μεγάλες βάσεις δεδομένων εκεί που η απλή εφαρμογή μεθόδων στατιστικής καθίσταται αδύνατη. Περιληπτικά, η διαδικασία εκπαίδευσης περιλαμβάνει τα δεδομένα εισόδου και τα γνωστά αποτελέσματα, με έναν αλγόριθμο που σταδιακά εκτελείται και μετρά την απόκλιση μεταξύ εκτίμησης και πραγματικού αποτελέσματος επιτυγχάνεται ελαχιστοποίηση αυτής της απόκλισης και επομένως όσο είναι εφικτό ταύτιση των τιμών εκτίμησης με τις πραγματικές τιμές.

Βαθιά μηχανική μάθηση

Η βαθιά μηχανική μάθηση είναι ένα υποσύνολο της μηχανικής μάθησης και οφείλει το όνομά της στην ύπαρξη διαδοχικών στρωμάτων απεικόνισης και εκπαίδευσης των δεδομένων εισόδου (W.Trask, 2019). Η βαθιά μηχανική μάθηση χρησιμοποιείται για να λύσει προβλήματα κυρίως σε

τομείς όπως η ανάλυση εικόνας, η ανάλυση και αναγνώριση κειμένου και πολλές άλλες εφαρμογές. Γενικά είναι ένα σύνολο εργαλείων από την εργαλειοθήκη της μηχανικής μάθησης που δομείται κυρίως με τη χρήση νευρωνικών δικτύων που είναι μια οικογένεια αλγορίθμων οι οποίοι έχουν σε πολύ γενικό βαθμό εμπνευστεί από τον ανθρώπινο εγκέφαλο.



Εικόνα 21 Η τεχνητή νοημοσύνη, η μηχανική μάθηση και η βαθιά μηχανική μάθηση ως σύνολα.

Είναι χρήσιμο να παρατηρηθεί πως δε χρησιμοποιείται όλο το φάσμα της βαθιάς μηχανικής μάθησης στην τεχνητή νοημοσύνη αλλά γενικά η χρήση των νευρωνικών δικτύων μπορεί να χρησιμοποιηθεί ως τρόπος επίλυσης πολλών προβλημάτων.

Ο ευρύτερος τομέας της μηχανικής και επομένως της βαθιάς μηχανικής μάθησης μπορεί να διακριθεί σε δύο υποκατηγορίες την επιτηρούμενη (ή αλλιώς μάθηση με επίβλεψη) (supervised learning) και την μη επιτηρούμενη (unsupervised learning).

Επιτηρούμενη μάθηση

Στην επιτηρούμενη μάθηση χρησιμοποιείται ένα σύνολο ως δεδομένα εισόδου και μετά από υπολογιστική διαδικασία εξάγεται ένα άλλο σύνολο ως δεδομένα εξόδου (Ng, 2018).

Χαρακτηριστικές εφαρμογές είναι:

- Να λαμβάνει το μοντέλο ως δεδομένα εισόδου πίξελ μιας εικόνας για να βλέπει την παρουσία ή απουσία ενός συγκεκριμένου ζώου ή αντικειμένου στην εικόνα
- Να λαμβάνει το μοντέλο μετεωρολογικά δεδομένα για να εκτιμά την πιθανότητα βροχής ή άλλου καιρικού φαινομένου
- Να λαμβάνει ένα μοντέλο δεδομένα βροχής και χαρακτηριστικών από λεκάνες για να εκτιμά πόσο νερό θα συγκεντρώσει ο ταμιευτήρας της λεκάνης αυτό είναι και το μοντέλο που χρησιμοποιήθηκε στην παρούσα διπλωματική

Όλα αυτά είναι μοντέλα τα οποία προσπαθούν να ‘μάθουν’ ένα μοτίβο το οποίο συνδέει τα δεδομένα εισόδου και εξόδου.

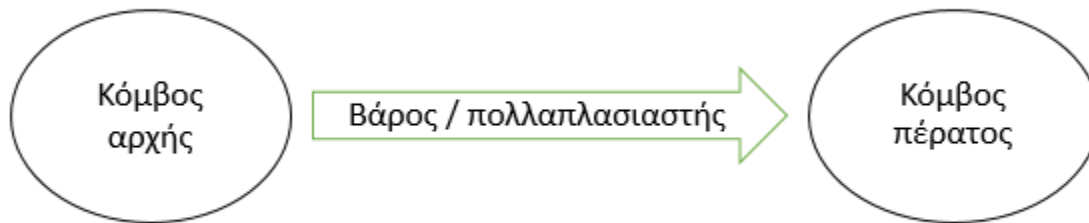
Μη επιτηρούμενη μάθηση

Η μη επιτηρούμενη μάθηση μοιράζεται ένα κοινό στοιχείο με την επιτηρούμενη και αυτή είναι η μετατροπή ενός συνόλου δεδομένων σε ένα άλλο σύνολο δεδομένων (Ng, 2018). Η διαφορά είναι ότι το σύνολο δεδομένων δεν είναι γνωστό δεν υπάρχει ‘σωστή απάντηση’ αντιθέτως η φιλοσοφία είναι να μπορέσει το μοντέλο να ανιχνεύσει κάποιο μοτίβο και να το παρουσιάσει ή να το μιμηθεί. Στην πραγματικότητα ο τρόπος με τον οποίο λειτουργεί αυτό το σύνολο αλγορίθμων είναι να προσπαθούν να κατηγοριοποιήσουν σε διάφορες ομάδες (clusters) τα διάφορα δεδομένα. Στην παρούσα διπλωματική θα ασχοληθούμε μόνο με την επιτηρούμενη μάθηση και για αυτό δε θα δοθεί συνέχεια στην μη επιτηρούμενη.

3.3.2 Νευρωνικά δίκτυα

Ένα νευρωνικό δίκτυο είναι ο βασικός πυλώνας γύρω από τον οποίο δομείται μια μέθοδος βαθιάς μηχανικής μάθησης (W.Trask, 2019). Η απαρχή των νευρωνικών δικτύων έγινε το 1943 από τον νευρολόγο Warren McCulloch του πανεπιστημίου του Illinois και τον μαθηματικό Walter Pitts του πανεπιστημίου του Σικάγο. Το 1954 οι Belmont Farley και Wesley Clark κατάφεραν να εκτελέσουν το πρώτο νευρωνικό δίκτυο στο MIT (Massachusetts Institute of Technology).

Η επεξήγηση της δομής των δικτύων αυτών θα ξεκινήσει με ένα απλό νευρωνικό δίκτυο. Το νευρωνικό δίκτυο αποτελείται από δομές όπως αυτή που φαίνεται στην παρακάτω εικόνα που θα αναλυθεί.



Εικόνα 22 Δύο κόμβοι που συνδέονται με έναν νευρώνα, ένα απλοποιημένο νευρωνικό δίκτυο

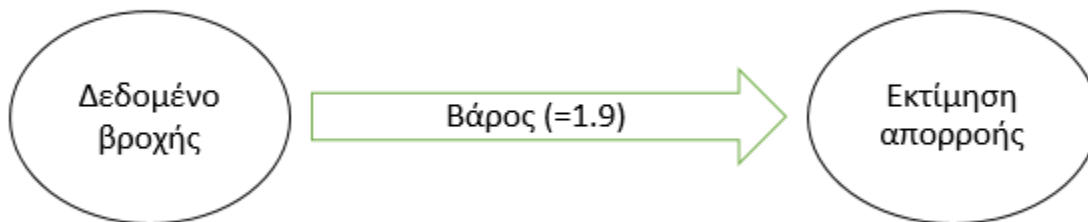
Αυτό το σχήμα αποτελεί τον δομικό λίθο των νευρωνικών δικτύων. Συνίσταται από έναν κόμβο αρχής, έναν κόμβο πέρατος και έναν πολλαπλασιαστή που ονομάζεται βάρος και συνδέει τους δύο κόμβους. Πρακτικά η σχέση που συνδέει αυτές τις μονάδες είναι:

$$(\text{κόμβος αρχής}) * (\text{βάρος}) = \text{κόμβος πέρατος}$$

Στο απλοποιημένο αυτό νευρωνικό δίκτυο ο κόμβος αρχής αντιπροσωπεύει τα δεδομένα εισόδου και ο κόμβος πέρατος αντιστοιχεί στα δεδομένα εξόδου, το βάρος είναι ο πολλαπλασιαστής που μετατρέπει τα δεδομένα εισόδου σε δεδομένα εξόδου.

Για να είναι πιο κατανοητή η επεξήγηση του νευρωνικού δικτύου θα χρησιμοποιηθεί ως παράδειγμα και θα αναλυθεί το πρόβλημα που αντιμετωπίστηκε στην παρούσα διπλωματική.

Οπότε το ανωτέρω σχήμα μπορεί να μετατραπεί σε:



Εικόνα 23 Απλοποιημένο νευρωνικό δίκτυο με παραδείγματα από την υδρολογία

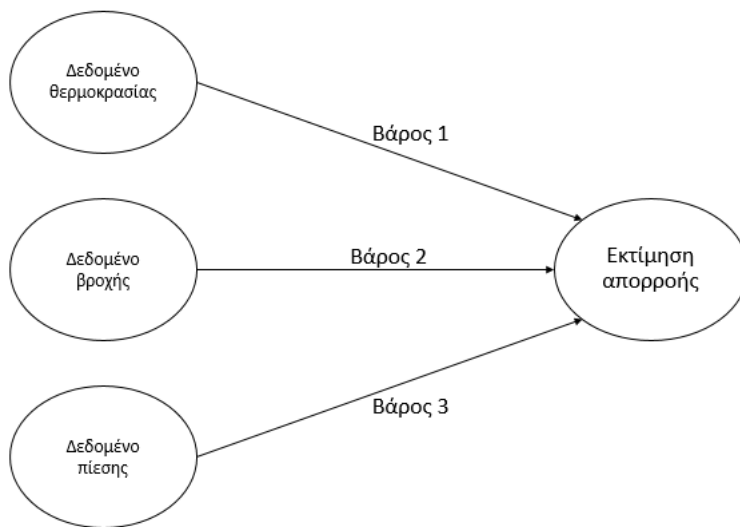
Έτσι θα μπορούσε να εκτιμηθεί η κάθε απορροή με βάση το πόσο έχει βρέξει. Για παράδειγμα αν είχαμε:

- Δεδομένο βροχής = 10 (mm)
- Με το γνωστό στο σχήμα βάρος = 1.9 ($*10^9 \text{ mm}^2 / \text{s}$)

Προκύπτει πως η εκτιμώμενη απορροή είναι $10 * 1.9 = 19 \text{ (m}^3 / \text{s)}$.

Αξίζει να σημειωθεί σε αυτό το σημείο ότι οι μονάδες που χρησιμοποιούνται είναι ενδεικτικές. Πράγματι στα δεδομένα βροχής εισάγεται η βροχή σε mm και στα δεδομένα εξόδου εξάγεται η απορροή σε m^3 / s αλλά αυτό δε σημαίνει ότι οι μονάδες του βάρους είναι αυτές που αναγράφονται. Στη συνέχεια που η μορφή του νευρωνικού δικτύου θα γίνει πιο πολύπλοκη θα είναι αρκετά πιο ξεκάθαρο για ποιό λόγο δεν υπάρχει νόημα στο να εκφραστεί με μονάδες το βάρος.

Η αρχιτεκτονική των νευρωνικών δικτύων επιτρέπει τώρα να υπάρχουν πολλοί κόμβοι αρχής που καταλήγουν σε έναν κόμβο πέρατος όπως φαίνεται στο παρακάτω σχήμα επεκτείνοντας το παράδειγμα.



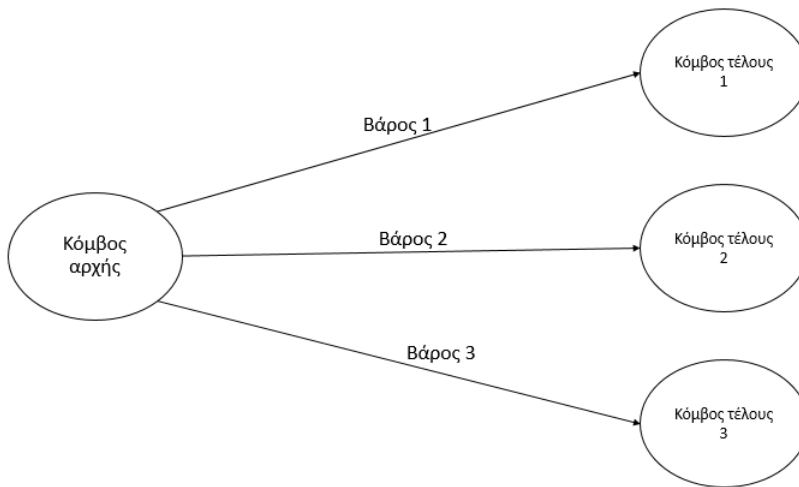
Εικόνα 24 Δίκτυο με 3 κόμβους να καταλήγουν σε 1 κόμβο

Για λόγους απλότητας των πράξεων θα συμβολίζονται τα βάρη με w από το αγγλικό weight έτσι πλέον το βάρος 1 θα συμβολίζεται με w_1 .

Το παραπάνω νευρωνικό για να εκτιμήσει την απορροή λαμβάνει και τα 3 δεδομένα (θερμοκρασία, βροχή, πίεση):

$$\text{Θερμοκρασία} * w_1 + \text{Βροχή} * w_2 + \text{Πίεση} * w_3 = \text{Απορροή}$$

Επιπρόσθετα μια δομή που απαντάται στα νευρωνικά είναι από έναν κόμβο να εξέρχονται 3 ή περισσότερες συνάψεις και να καταλήγουν σε 3 κόμβους. Αυτή η δομή υπάρχει συνήθως στα ενδιάμεσα στρώματα ενός νευρωνικού δικτύου.

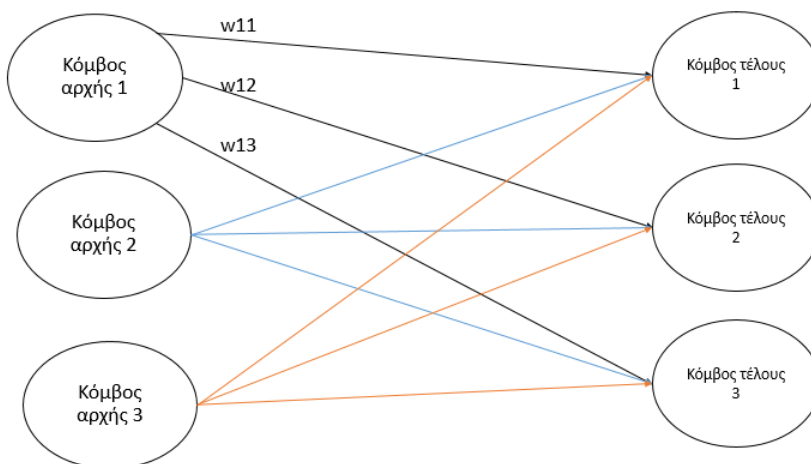


Εικόνα 25 Δίκτυο με 1 κόμβο να καταλήγει σε 3 κόμβους

Από εδώ δεν προκύπτει μια μόνο εξίσωση, αντι αυτού υποδηλώνεται πως στον υπολογισμό του εκάστοτε κόμβου τέλους συμμετέχει (με ένα συγκεκριμένο βάρος) ο κόμβος αρχής. Προκύπτουν δηλαδή τα 3 παρακάτω σεν εξισώσεων:

- Κόμβος τέλους 1 = w_1 * κομβος αρχής 1
- Κόμβος τέλους 2 = w_2 * κομβος αρχής 2
- Κόμβος τέλους 3 = w_3 * κομβος αρχής 3

Τέλος υπάρχει και ο συνδυασμός των 2 τελευταίων παραδειγμάτων που δημιουργεί την παρακάτω δομή:



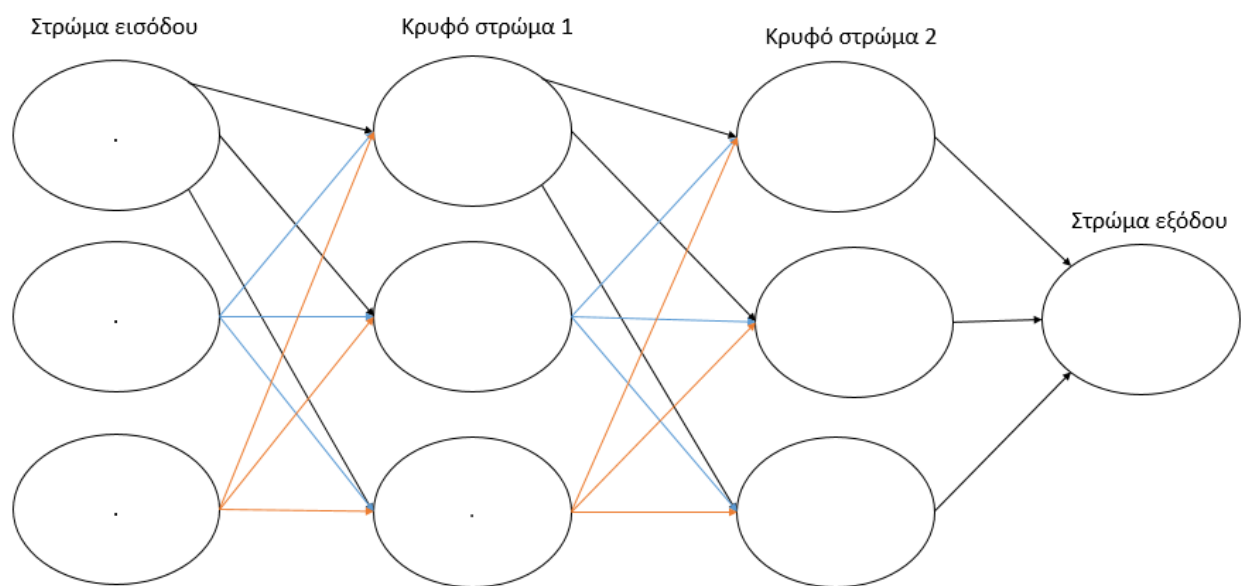
Εικόνα 26 Δίκτυο με 3 κόμβους να καταλήγουν σε 3 κόμβους

Οι εξισώσεις που προκύπτουν από την παραπάνω δομή είναι ο συνδυασμός των εξισώσεων των παραπάνω 2 περιπτώσεων:

- Κόμβος τέλους 1 = $w_{11} * \text{κομβος αρχής 1} + w_{21} * \text{κόμβος αρχής 2} + w_{31} * \text{κόμβος αρχής 3}$
 - Κόμβος τέλους 2 = $w_{12} * \text{κομβος αρχής 1} + w_{22} * \text{κόμβος αρχής 2} + w_{32} * \text{κόμβος αρχής 3}$
- Κόμβος τέλους 3 = $w_{13} * \text{κομβος αρχής 1} + w_{23} * \text{κόμβος αρχής 2} + w_{33} * \text{κόμβος αρχής 3}$

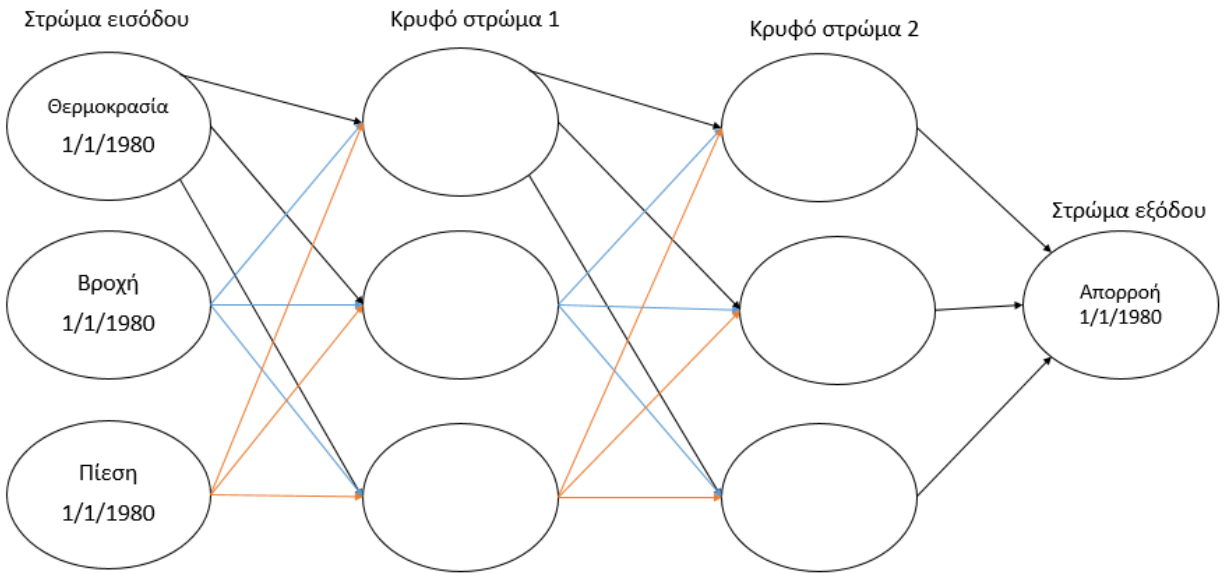
Οι συμβολισμοί των βαρών πέραν του κόμβου αρχής 1 δεν αναγράφονται για να αποφευχθεί επιπλέον πολυπλοκότητα, αλλά η φιλοσοφία της ονομασίας είναι παρόμοια με αυτή των πινάκων. Στον όνομα w_{12} ο όρος 1 δηλώνει ότι ο κόμβος αρχής είναι ο νούμερο 1 και αντίστοιχα ο όρος 2 δηλώνει ότι ο όρος τέλους είναι ο όρος 2. Επιπλέον, τα βέλη ανάλογα με τον κόμβο από τον οποίο εκκινούν χρωματίζονται και με άλλο χρώμα για να είναι πιο σαφή.

Τώρα που κατανοήθηκαν οι βασικές δομές του νευρωνικού δικτύου φαίνεται πως όλο το νευρωνικό δίκτυο αποτελείται από ένα σύνολο κόμβων και συνάψεις που τους συνδέουν. Έτσι μπορούμε να συνδυάσουμε την τελευταία δομή και να προκύψει ένα ολοκληρωμένο νευρωνικό δίκτυο.



Εικόνα 27 Ολοκληρωμένο 'ρηχό' νευρωνικό δίκτυο (shallow neural network)

Στο ανωτέρω σχήμα έχουμε φτάσει πλέον στην τελική μορφή ενός συνηθισμένου νευρωνικού δικτύου. Αξίζει να σημειωθεί πως ένα νευρωνικό δίκτυο μπορεί να πάρει όποια μορφή του δώσει ο κατασκευαστής του ανάλογα με το πρόβλημα που έχει να επιλύσει οπότε αυτή η αρχιτεκτονική δεν είναι πανάκεια αλλά συναντάται πολύ συχνά. Ένας χώρος που εφαρμόζεται είναι και στην παρούσα διπλωματική που λαμβάνοντας ημερήσια δεδομένα διαφορετικής φύσης (βροχή, θερμοκρασία, πίεση) προσπαθεί να εκτιμηθεί η ημερήσια απορροή. Όπως στο παρακάτω σχήμα:



Εικόνα 28 Ρηχό νευρωνικό δίκτυο με παραδείγματα από την υδρολογία

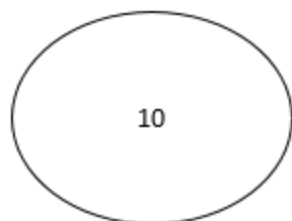
Επειδή οι όροι αυτοί προέρχονται από την αγγλική βιβλιογραφία και στα ελληνικά είναι μεταφρασμένοι αξίζει να αναφερθούν και οι αγγλικοί όροι. Επομένως, στρώμα εισόδου (= input layer), κρυφό στρώμα 1 (= hidden layer 1), κρυφό στρώμα 2 (= hidden layer 2), στρώμα εξόδου (= output layer).

Πρακτικά λοιπόν εισάγονται αριθμοί στο στρώμα εισόδου και εκτελώντας τα κατάλληλα αθροίσματα με βάρη (weighted summations) καταλήγουμε στον κόμβο / στρώμα εξόδου που είναι η απορροή.

3.3.3 Διαδικασία ανανέωσης των τιμών των βαρών (gradient descent)

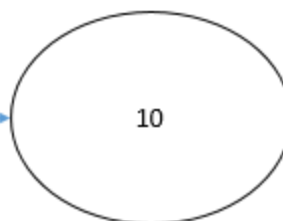
Η gradient descent μεταφρασμένη στα ελληνικά ως καθοδική κλίση είναι ο τρόπος με τον οποίο τα περισσότερα νευρωνικά δίκτυα ανανεώνουν τα βάρη τους. Αυτό χρειάζεται διότι για να εκτιμήσει στοχευμένα ένα νευρωνικό δίκτυο θα πρέπει τα βάρη του δηλαδή οι πολλαπλασιαστές των αρχικών δεδομένων να είναι κατάλληλα (W.Trask, 2019) (Ng, 2018). Πιο συγκεκριμένα παρακάτω παρατίθεται ένα παράδειγμα για το πως ανανεώνεται το βάρος σε ένα απλό νευρωνικό δίκτυο όπως αυτό που παρουσιάστηκε στην αρχή του κεφαλαίου.

Δεδομένο βροχής



$w=1$

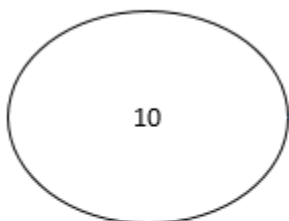
Δεδομένο απορροής



Εικόνα 29 Κόμβοι δικτύου με αναπαράσταση τιμών

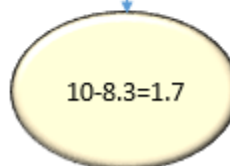
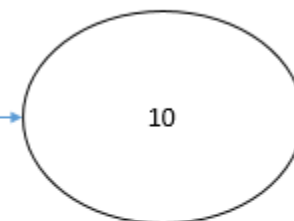
Σε αυτή την περίπτωση ένα δεδομένο βροχής πολλαπλασιάζεται με το βάρος μας δίνει την εκτίμηση του νευρωνικού δικτύου για την απορροή. Κοινώς, $10 * 1 = 10$. Στη συνέχεια αυτό το αποτέλεσμα συγκρίνεται με το κανονικό αποτέλεσμα που είναι ήδη γνωστό και φαίνεται κατά πόσο το νευρωνικό δίκτυο έσφαλλε στην πρόγνωση του. Αν υποθέσουμε ότι η πραγματική τιμή της απορροής ήταν $8.3 \text{ m}^3/\text{s}$ τότε το νευρωνικό δίκτυο έχει αστοχήσει κατά 1.7 μονάδες και πρέπει να ανανεώσει το βάρος του. Αυτό γίνεται με τη χρήση ενός παρένθετου κόμβου που συνήθως ονομάζεται ‘κόμβος δέλτα’. Αυτός ο κόμβος υπολογίζει σε αλγεβρική τιμή τη διαφορά πρόγνωσης και πραγματικής τιμής.

Δεδομένο βροχής



$w=1$

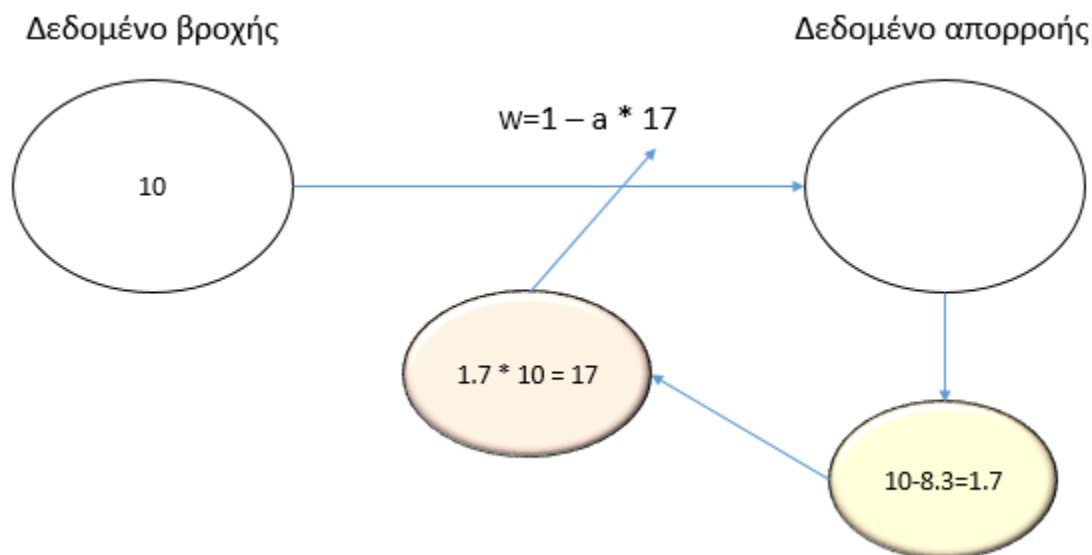
Δεδομένο απορροής



Εικόνα 30 Σύγκριση εκτιμώμενου αποτελέσματος με πραγματική τιμή και υπολογισμός διαφοράς

Η τιμή δέλτα = + 1.7 υποδηλώνει ότι το νευρωνικό δίκτυο αστόχησε προβλέποντας κατά 1.7 πιο αυξημένη τιμή από ότι θα έπρεπε. Στη συνέχεια, γίνεται ο υπολογισμός του όρου ‘δέλτα του βάρους’ (weight_delta) πολλαπλασιάζοντας το ‘δέλτα’ με την τιμή εισόδου (10 για τη βροχή)

Έτσι, προκύπτει η τιμή $1.7 * 10 = 17$. Αυτή η τιμή πολλαπλασιάζεται στη συνέχεια με έναν μειωτικό συντελεστή που ονομάζεται ‘ρυθμός μάθησης’ (learning rate) και συμβολίζεται με a αυτός ο όρος αφαιρείται από την ήδη υπάρχουσα τιμή του βάρους. Πιο αναλυτικά στο σχήμα:



Εικόνα 31 Υπολογισμός όρου 'δέλτα' και ανανέωση βάρους

Μια συνήθης τιμή του $a = 0.01$ και έτσι το νέο βάρος που θα προκύψει θα είναι

$$W = 1 - 0.01 * 17 = 0.83$$

Με αυτό τον τρόπο το βάρος έχει πλέον ανανεωθεί και προσαρμόσκει στη νέα πληροφορία που θέλει όταν υπάρχει δεδομένο βροχής 10 να εκτιμάται απορροή ίση με 8.3. Αυτός ήταν ένας ολοκληρωμένος κύκλος του gradient descent. Η ανωτέρω διαδικασία περιγράφει το θεωρητικό υπόβαθρο το οποίο λειτουργεί πίσω από την ανανέωση των βαρών και τη μάθηση. Στην εφαρμογή των νευρωνικών δικτύων υπολογίζεται και ένας όρος που ονομάζεται ‘σφάλμα’ αυτός ο όρος αποτελεί τον σημαντικότερο δείκτη που φανερώνει αν το νευρωνικό δίκτυο λειτουργεί ορθά και δοκιμάζει την απόδοσή του. Η φιλοσοφία πίσω από την ανανέωση βαρών είναι ότι κάθε φορά για το σφάλμα ‘ευθύνεται’ ένα βάρος το οποίο δεν είχε σωστή τιμή και μέσω της μεθόδου που εξηγήθηκε κάθε βάρος που συμβάλλει στη δημιουργία σφάλματος ανανεώνεται. Η διαδικασία της ‘μάθησης’ του μοντέλου πρακτικά ταυτίζεται με τη διαδικασία ελαχιστοποίησης του σφάλματος, σε ένα άριστο νευρωνικό δίκτυο το σφάλμα θα έπρεπε να είναι 0. Αυτό θα έδειχνε πως το μοντέλο εκτιμά σωστά την κάθε τιμή πρόγνωσης.

Τα σφάλματα μπορούν να έχουν πολλές μορφές ανάλογα με το τι θέλει να πετύχει κάθε φορά ο σχεδιαστής του νευρωνικού δικτύου (Ng, 2018). Το δέλτα είναι και αυτό ένα είδος σφάλματος και η αγγλική του ονομασία είναι ‘raw error’. Δεν αποτελεί όμως ουσιαστικά σφάλμα καθώς τότε δε θα μπορούσε να είναι αρνητικό. Ένα άλλο πολύ γνωστό σφάλμα είναι το μέσο τετραγωνικό σφάλμα (mean square error). Είναι η διαφορά της τιμής πρόγνωσης και της πραγματικής τιμής υψωμένη στο τετράγωνο για να εξασφαλιστεί πως είναι θετική.

$$\sigma = (x_{pred} - x_{real})^2 \quad (3.26)$$

Στην παραπάνω εξίσωση όπου x_{pred} = η τιμή πρόγνωσης του μοντέλου, x_{real} = η πραγματική τιμή. Άλλο ένα σφάλμα που είναι κοινό και μάλιστα χρησιμοποιήθηκε στην παρούσα εργασία είναι το ‘απόλυτο σφάλμα’ (absolute error). Η εξίσωση του απόλυτου σφάλματος είναι:

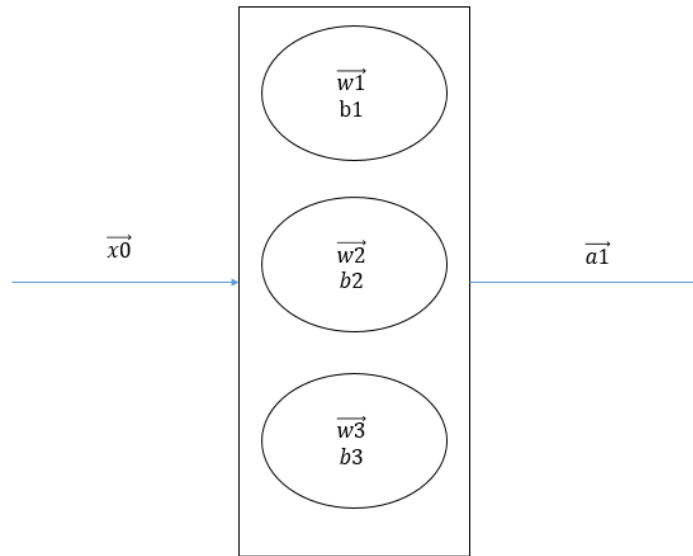
$$\sigma = |x_{pred} - x_{real}| \quad (3.27)$$

Στην παρούσα διπλωματική εργασία δοκιμάστηκαν και τα δυο σφάλματα και φάνηκε πως το μέσο τετραγωνικό σφάλμα βοηθά το νευρωνικό δίκτυο να εκτιμά καλύτερα την τιμή για αυτό και επιλέχθηκε. Το σφάλμα αυτό όταν εφαρμοστεί σε ένα σύνολο παραδειγμάτων εκπαίδευσης συνιστά τη συνάρτηση κόστους.

3.3.4 Οι εξισώσεις του νευρωνικού δικτύου

Αφού η διαδικασία για τη διόρθωση του σφάλματος και την προσπάθεια καλύτερης πρόγνωσης έγινε σαφής είναι κατάλληλο να αναφερθεί η πλήρης δομή του νευρωνικού καθώς χάρη η απλούστευσης προηγουμένως είχε γίνει παράλειψη κάποιον στοιχείων.

Ένα ολοκληρωμένο στρώμα νευρώνων δέχεται ένα διάνυσμα δεδομένων και εξάγει έναν άλλο διάνυσμα δεδομένων, πρακτικά κάθε νευρώνας δέχεται το αρχικό διάνυσμα δεδομένων (μια τιμή από κάθε νευρώνα προηγούμενου στρώματος) και εξάγει μία τιμή (W.Trask, 2019). Αυτή η διαδικασία επαναλαμβάνεται συνεχώς μεχρι να καταλήξει στο στρώμα εξόδου και στην τελική εκτίμηση της ζητούμενης τιμής. Για καλύτερη κατανόηση θα αναλυθούν τα παρακάτω σχήματα.



Εικόνα 32 Αναπαράσταση στρώματος ενός νευρωνικού δικτύου

Εδώ φαίνεται ότι κάθε στρώμα δέχεται ένα διάνυσμα x_0 και εξάγει ένα διάνυσμα a_1 . Οι νευρώνες στο εσωτερικό τους λαμβάνουν το διάνυσμα x και εκτελούν τα εξής 2 βήματα λαμβάνοντας ως παράδειγμα το νευρώνα 1:

$$z_1 = \vec{x}_0 * \vec{w}_1 + b_1 \quad (3.28)$$

$$a_1 = g(z_1) \quad (3.29)$$

Στο βήμα 1 εκτελούνται οι πράξεις με τα βάρη που ήδη αναφέρθηκαν και ταυτόχρονα προστίθεται και μια σταθερά b . Έτσι η εξίσωση αν το διάνυσμα x ήταν ένα διάνυσμα $x(1,3)$ θα ήταν:

$$z_1 = x_{01} * w_{11} + x_{02} * w_{12} + x_{03} * w_{13} + b_1 \quad (3.30)$$

Γίνεται δηλαδή ένα μοντέλο πολυωνυμικής παλινδρόμησης.

Στο βήμα 2 εφαρμόζεται η συνάρτηση ενεργοποίησης. Πρακτικά, το αποτέλεσμα του βήματος 1 χρησιμοποιείται σαν είσοδος στη συνάρτηση ενεργοποίησης και από εκεί προκύπτουν οι τελικές τιμές του διανύσματος a , κάθε νευρώνας συμβάλλει στο διάνυσμα a με μια τιμή.

Σε συνέχεια του προηγούμενου κεφαλαίου η ολοκληρωμένη διαδικασία της καθοδικής κλίσης με εξισώσεις που επιτυγχάνει την ανανέωση των βαρών ενός κόμβου είναι:

$$w_j = w_j - a \frac{d}{dw_j} J(\vec{w}, b) \quad (3.31)$$

$$b_j = b_j - a \frac{d}{db_j} J(\vec{w}, b) \quad (3.32)$$

Όπου j (μικρό) είναι ο αριθμός του εκάστοτε βάρους όπως φαίνεται και στην εξίσωση 1 πιο πάνω.

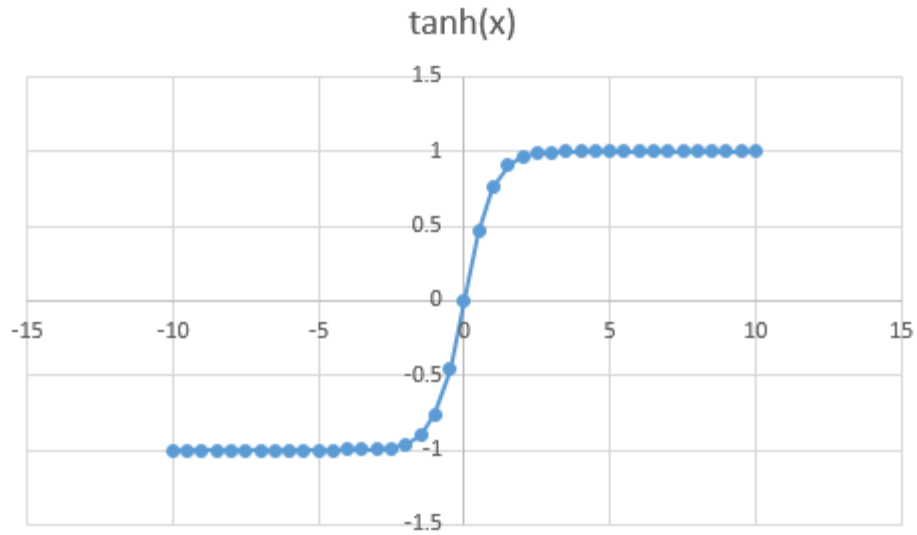
Όπου J είναι η συνάρτηση κόστους, προκύπτει από το σφάλμα που έχει επιλεγεί. Οι ανανεώσεις αυτές εκτελούνται ταυτόχρονα. Πιο συγκεκριμένα, οι ανανεώσεις που γίνονται στο διάνυσμα w λαμβάνουν υπόψη τους το μη ανανεωμένο J (που περιλαμβάνει το προηγούμενο b) αντίστοιχα και για τις ανανεώσεις του b λαμβάνεται το μη ανανεωμένο J .

3.3.5 Συναρτήσεις ενεργοποίησης

Όπως είναι πλέον σαφές τα βάρη w και οι τιμές b καθορίζουν την συσχέτιση μεταξύ δεδομένων και αποτελεσμάτων. Η συνάρτηση ενεργοποίησης προσδίδει στο νευρωνικό το χαρακτηριστικό της μερικής συσχέτισης ένα στοιχείο που είναι απαραίτητο για να μπορεί το μοντέλο να προσαρμόζει τις παραμέτρους w , b (Ng, 2018). Κάποιες από τις συναρτήσεις ενεργοποίησης είναι οι:

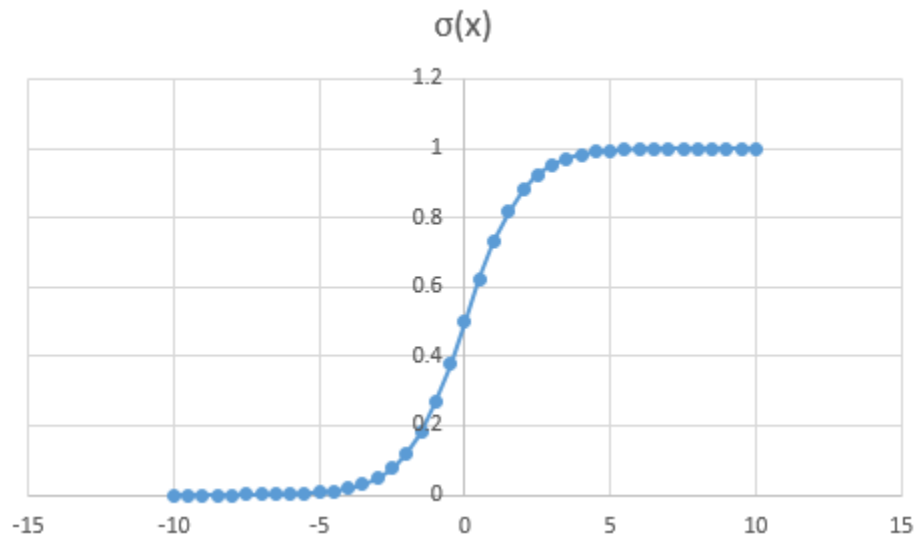
- Υπερβολική εφαπτομένη

$$\tanh(x) = \frac{2}{1+e^{-2x}} - 1 \quad (3.33)$$

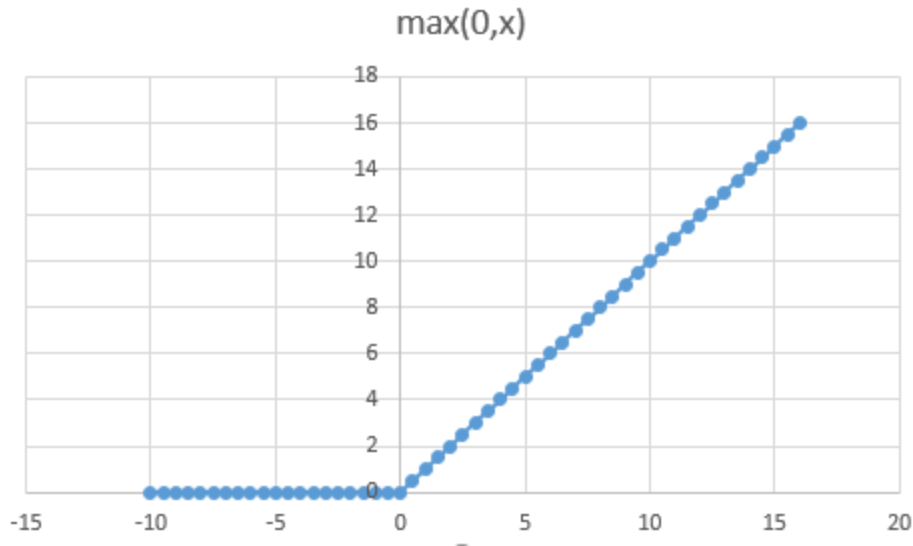


- Σιγμοειδής συνάρτηση

$$\sigma(z) = \frac{1}{1+e^{-z}} \quad (3.34)$$



- Συνάρτηση ReLu (Rectified Linear Unit)



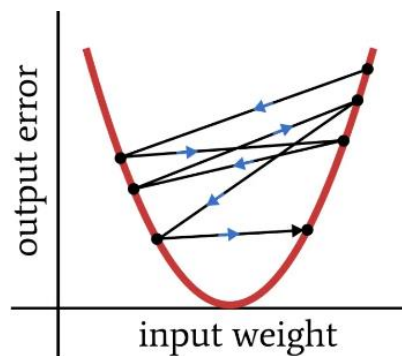
Στην παρούσα εργασία ως συνάρτηση ενεργοποίησης χρησιμοποιείται η ReLu. Αυτή δίνει στις τιμές με αρνητική επιρροή τιμή 0 με αυτό τον τρόπο τα βάρη που συμβάλλουν στην τελική εκτίμηση είναι συγκεκριμένα και έτσι στην επερχόμενη ανανέωση αλλάζουν μόνο αυτά, έτσι επιτυγχάνεται η μερική συσχέτιση.

3.3.6 Ρυθμός μάθησης

Η διαδικασία της καθοδικής κλίσης που συμβάλλει στη μάθηση έχει ως σκοπό οι τιμές των βαρών να είναι τέτοιες ώστε να ελαχιστοποιηθεί το σφάλμα. Υπάρχει όμως το σενάριο οι τιμές να μειωθούν ή να αυξηθούν πάρα πολύ με αποτέλεσμα να βρεθούμε στην άλλη πλευρά της καμπύλης. Αυτό μπορεί να συμβαίνει επειδή ένα δεδομένο έχει πολύ μεγάλη τιμή και δημιουργεί μεγάλο 'δέλτα του βάρους' καθώς όπως αναφέρθηκε και πάνω

$$\text{'δέλτα βάρους'} = \text{τιμή εισόδου} * \text{δέλτα}$$

Έτσι ενδέχεται να μεταπηδήσει το επιθυμητό σημείο που η κλίση είναι μηδέν όπως φαίνεται στο σχήμα.



Εικόνα 33 Αποτυχία σύγκλισης στην ελάχιστη τιμή λόγω υψηλού ρυθμού μάθησης (overshooting learning rate)

Έτσι ο πιο απλός τρόπος να αποφευχθεί αυτό το φαινόμενο είναι να εισαχθεί ο ρυθμός μάθησης και να προσαρμοστεί το βήμα με το οποίο μεταβάλλονται οι τιμές των βαρών. Στη συνέχεια, στην

ανάλυση του κώδικα θα εξηγηθεί ότι ο ρυθμός μάθησης μπορεί να μεταβάλλεται ανάλογα με το σε ποίο σημείο είναι η διαδικασία μάθησης ανάλογα με την κλίση.

3.3.7 Μεγεθος πακέτου

Μια άλλη παράμετρος που αξίζει να αναφερθεί είναι το ποια χρονική στιγμή πρέπει να ανανεωθούν τα βάρη. Μια μέθοδος είναι να ανανεώνονται μετά από κάθε ένα παράδειγμα και ονομάζεται στοχαστική καθοδική κλίση (stochastic gradient descent). Ένα σύνολο δεδομένων εισόδου και εξόδου αποτελεί ένα ‘παράδειγμα εκπαίδευσης’. Συνήθως, η ανανέωση των βαρών γίνεται μετά από ένα συγκεκριμένο αριθμό παραδειγμάτων εκπαίδευσης που είναι σε πολλαπλάσια του 16, αυτά λογίζονται ως ένα πακέτο. Όταν υπάρχουν πολλά παραδείγματα εκπαίδευσης ο αριθμός των παραδειγμάτων που περιλαμβάνει κάθε πακέτο ποικίλει και εξαρτάται από τη διαδικασία σύγκλισης του αλγορίθμου. Συνήθως όταν τα δεδομένα είναι πάρα πολλά τα πακέτα μπορεί να περιλαμβάνουν 128 ή και 256 παραδείγματα εκπαίδευσης, αντίθετα αν αρχίζουν να μειώνονται το πακέτο μπορεί να περιλαμβάνει 64, 32 μέχρι και 8. Για να βρεθεί το ιδανικό μέγεθος για κάθε πρόβλημα πρέπει να τεσταριστούν πολλοί συνδιασμοί και να φανεί ποιο οδηγεί στην καλύτερη λειτουργία του νευρωνικού.

3.4 Οι δείκτες Nash-Sutcliffe Efficiency και Absolute Nash-Sutcliffe Efficiency

Για να μπορέσουν τα μοντέλα να συγκριθούν τόσο μεταξύ διαφορετικών λεκανών όσο και μεταξύ διαφορετικών οριζόντων πρόγνωσης αλλά και το ένα με το άλλο έπρεπε να χρησιμοποιηθεί ένας αδιάστατος δείκτης για τη μέτρηση σφάλματος. Ο δείκτης NSE λαμβάνει τιμές από το $-\infty$ έως το $+1$ (Nash, 1970). Όταν η τιμή ισούται με τη μονάδα σημαίνει πως υπάρχει πλήρης ταύτιση μεταξύ παρατηρούμενων και εκτιμώμενων τιμών από το μοντέλο. Όταν ο δείκτης είναι μικρότερος του 0 σημαίνει πως η μέση τιμή της χρονοσειράς είναι καλύτερη εκτιμήτρια από το μοντέλο που αναπτύχθηκε.

Ο δείκτης NSE χρησιμοποιείται ευρύτατα στην υδρολογία για να καλιμπραριστούν μοντέλα (αλλαγή παραμέτρων που αποσκοπούν στη μείωση της απόκλισης μεταξύ πραγματικών και εκτιμώμενων τιμών).

Δίνεται από τον τύπο:

$$NSE = 1 - \frac{\sum(\theta_i - s_i)^2}{\sum(\theta_i - \mu_0)^2} \quad (3.35)$$

Όπου S_i και θ_i είναι οι προσομοιωμένες και οι πραγματικές τιμές αντίστοιχα, μ_0 είναι η παρατηρηθείσα μέση τιμή (P.Kraus, 2005)

Ένα μειονέκτημα του δείκτη NSE είναι πως υπερεκτιμά τις υψηλές τιμές απορροής σε μια χρονοσειρά (καθώς έχει τους όρους τετραγώνων) αντίστοιχα οι χαμηλότερες τιμές μειώνονται

σε βαρύτητα. Για τον παραπάνω λόγο χρησιμοποιήθηκε και μια τροποποιημένη μορφή του αρχικού δείκτη NSE. Είναι η απόλυτη τιμή του δείκτη, συμβολίζεται με ANSE και προκύπτει από την αντικατάσταση των τετραγώνων του αριθμητή και του παρανομαστή της αρχικής εξίσωσης με απόλυτες τιμές.

$$NSE = 1 - \frac{\sum |\theta_i - s_i|}{\sum |\theta_i - \mu_0|} \quad (3.36)$$

Σε αυτό το σημείο αξίζει να σημειωθεί πως έχουν γίνει αρκετές έρευνες για τις τιμές του NSE που μπορούν να επικυρώσουν ένα μοντέλο ως αποδεκτό προς χρήση. Δεν αρκεί μόνο να ξεπερνά τη μονάδα αλλά πρέπει να έχει και ένα σχετικά ικανοποιητικό σκορ. Ύστερα από βιβλιογραφική έρευνα βρέθηκαν και παρουσιάζονται κάποιες διαβαθμίσεις των NSE σκορ.

Πρόταση 1 Ο Moriasi και άλλοι ερευνητές (Daniel Moriasi, 2007) έχοντας συγκρίνει πολλά μοντέλα για μηνιαίο βήμα πρόγνωσης προτείνονται τα εξής εύρη:

- Μη ικανοποιητικό $NSE < 0.50$
- Ικανοποιητικό $0.50 < NSE < 0.65$
- Καλό $0.65 < NSE < 0.75$
- Πολύ καλό $0.75 < NSE$

Πρόταση 2: Ο Skaggs στο μοντέλο 'DRAINMOD' (Skaggs, 2012) που χρησιμοποιεί για να μελετήσει τις απορροές σε λεκάνες για άρδευση προτείνει ως εύρος:

- Μη ικανοποιητικό $NSE < 0.40$
- Ικανοποιητικό $0.40 < NSE < 0.60$
- Καλό $0.60 < NSE < 0.75$
- Εξαιρετικό $0.75 < NSE$

Πρόταση 3: Τέλος στο paper 'Comparison of WEAP and SWAT models for streamflow prediction in the Hadejia-Nguru Wetlands, Nigeria' (U. A. Ibrahim, 2022) προτείνεται η εξής κατάταξη:

- Μη ικανοποιητικό $NSE < 0.36$
- Ικανοποιητικό $0.36 < NSE < 0.75$
- Εξαιρετικό $0.75 < NSE$

4 Εφαρμογή σε προγραμματιστικό περιβάλλον Python

Για να υλοποιηθεί η ανάλυση των χρονοσειρών χρησιμοποιήθηκαν οι μέθοδοι των οποίων η θεωρία εξηγήθηκε στα προηγούμενα κεφάλαια. Στο κεφάλαιο αυτό θα εξηγηθεί ακριβώς πως λειτουργεί η κάθε μέθοδος τι σκοπό επιτυγχάνει και θα επεξηγηθούν κάποια πακέτα της Python τα οποία χρησιμοποιήθηκαν για να εκτελεστεί ο κώδικας. Η ανάλυση μπορεί να διακριθεί σε δύο μέρη:

1. Μέθοδοι πρόγνωσης. Η πλειονότητα των μεθόδων που αναπτύχθηκαν στην παρούσα εργασία έχουν ως σκοπό την πρόγνωση της απορροής των ποταμών που μελετήθηκαν. Η πρόγνωση αυτή έγινε σε ημερήσια βάση για ορίζοντες 1, 5 και 10 ημερών.
2. Μέθοδος μοντελοποίησης. Εφαρμόστηκε και μία μέθοδος μοντελοποίησης λεκανών με χρήση των νευρωνικών δικτύων που αναφέρθηκαν στα προηγούμενα κεφάλαια.

4.1 Εκθετική Ομαλοποίηση

Η Μέθοδος αυτή χρησιμοποιήθηκε για να εκτιμήσει την απορροή σε κάθε μια από τις 10 λεκάνες μελέτης σε ημερήσια βάση για ορίζοντες πρόγνωσης 1, 5 και 10 ημέρες. Ο κώδικας κατασκευάστηκε με τέτοιο τρόπο ώστε να είναι εφικτό ο ορίζοντας πρόγνωσης να αλλάζει αλλάζοντας μόνο μια μεταβλητή. Στη συνέχεια εκτελέστηκε για κάθε έναν από τους τρεις διαφορετικούς ορίζοντες πρόγνωσης και για κάθε μια από τις 10 λεκάνες. Η εκτέλεση αναλύεται και επεξηγείται στα βήματα που ακολουθούν.

4.1.1 Δεδομένα

Τα δεδομένα που χρησιμοποιήθηκαν για την εκθετική ομαλοποίηση ήταν όπως έχει ήδη αναφερθεί από τη βάση δεδομένων των Ηνωμένων Πολιτειών Αμερικής, CAMELS (NCAR UCAR, 2022). Ήταν ημερήσια δεδομένα απορροής που βρίσκονταν σε μορφή κειμένου. Στη συνέχεια μετατράπηκαν σε μορφή excel και εισήχθησαν στον κώδικα με τη μορφή Pandas DataFrame που είναι μια κλάση της Python η οποία επιτρέπει την επεξεργασία δεδομένων και χρησιμοποιείται πολύ συχνά στον κλάδο της ανάλυσης δεδομένων στην Python. Σε όλες τις λεκάνες και σε όλους τους ορίζοντες πρόγνωσης χρησιμοποιήθηκαν ως δεδομένα εκπαίδευσης όλα τα έτη εκτός από το τελευταίο. Στο τελευταίο έτος εφαρμόστηκε ένας walk forward αλγόριθμος που θα εξηγηθεί στη συνέχεια.

4.1.2 Πρόγνωση απορροής

Η διαδικασία της πρόγνωσης της απορροής των επόμενων ημερών έγινε με την αλληλεπίδραση 2 συναρτήσεων που κατασκευάστηκαν (Brownlee, 2018). Η πρώτη εξυπηρετούσε την πρόγνωση της αμέσως επόμενης ημέρας οπότε θα μπορούσε να θεωρηθεί ο βασικός πυλώνας της πρόγνωσης και η δεύτερη χρησιμοποιούσε τον βασικό πυλώνα επαναληπτικά για να εκτιμήσει όλη τη χρονοσειρά που έχει ζητηθεί (τις τελευταίες 365 ημέρες).

Στην πρώτη συνάρτηση που ονομάζεται 'exp_smoothing_forecast' γίνεται η σύνθεση του μοντέλου της εκθετικής ομαλοποίησης και αρχικοποιούνται οι υπερ-παράμετροι εισόδου ώστε να

είναι εφικτό να κληθεί η παρούσα συνάρτηση πολλές φορές έχοντας κατασκευάσει πολλά διαφορετικά μοντέλα ανάλογα με το ανάγκη κατασκευής απλής διπλής ή τριπλής εκθετικής ομαλοποίησης. Η συνάρτηση αυτή αφού κατασκευάσει το μοντέλο λαμβάνει μια χρονοσειρά και με τη μέθοδο της μεγίστης πιθανοφάνειας προσαρμόζει τις παραμέτρους και εκτιμά την αμέσως επόμενη τιμή της χρονοσειράς με βάση το υπάρχον μοντέλο.

Στη δεύτερη συνάρτηση που ονομάζεται ‘holt_winters_walk_forward_validation’ λαμβάνεται υπόψιν ο ορίζοντας πρόγνωσης που έχει δοθεί ως υπερπαραμέτρος στην μέθοδο. Παραδείγματος χάρη όταν ο ορίζοντας είναι 5 ημέρες η συνάρτηση forecast εκτελείται την για την ημέρα 1 στη συνέχεια κατασκευάζεται μια προσωρινή χρονοσειρά με τις πραγματικές τιμές (μέχρι πριν την ημέρα 1, και στη θέση της ημέρας 1 υπάρχει η πρόγνωση του μοντέλου) αυτή η χρονοσειρά ξαναδίνεται ως δεδομένο στη συνάρτηση forecast και εκτιμάται η ημέρα 2, αυτή αποθηκεύεται ξανά σε μια προσωρινή χρονοσειρά και η διαδικασία επαναλαμβάνεται μέχρι να εκτιμηθεί η ημέρα 5. Έπειτα από την πρόγνωση της ημέρας 5, οι προσωρινές χρονοσειρές διαγράφονται και στη θέση των εκτιμήσεων πλέον τοποθετούνται οι πραγματικές τιμές απορροής (Brownlee, 2018). Πιο συγκεκριμένα, με τα παρακάτω βήματα διευκρινίζεται η πορεία εκτέλεσης της πρόγνωσης:

- Βήμα 1: Πρόγνωση της ημέρας 1

$$\hat{Y}_t = f(y_{t-n}, \dots, y_{t-1})$$

- Βήμα 2: Πρόγνωση της ημέρας 2

$$\hat{Y}_{t+1} = f(y_{t-n}, \dots, y_{t-1}, \hat{Y}_t)$$

όπου \hat{Y}_t η πρόγνωση της απορροής στο βήμα 1

- Βήμα 3: Πρόγνωση της ημέρας 3

$$\hat{Y}_{t+2} = f(y_{t-n}, \dots, y_{t-1}, \hat{Y}_t, \hat{Y}_{t+1})$$

- Βήμα 4: Πρόγνωση της ημέρας 4

$$\hat{Y}_{t+3} = f(y_{t-n}, \dots, y_{t-1}, \hat{Y}_t, \hat{Y}_{t+1}, \hat{Y}_{t+2})$$

- Βήμα 5: Πρόγνωση της ημέρας 5

$$\hat{Y}_{t+4} = f(y_{t-n}, \dots, y_{t-1}, \hat{Y}_t, \hat{Y}_{t+1}, \hat{Y}_{t+2}, \hat{Y}_{t+3})$$

Ύστερα από το βήμα 5 οι εκτιμήσεις διαγράφονται από τη χρονοσειρά και στη θέση τους τοποθετούνται οι πραγματικές τιμές. Αντίστοιχα, όταν ο ορίζοντας πρόγνωσης αλλάζει σε 1 τότε εκτελείται μόνο το πρώτο βήμα και όταν ο ορίζοντας αλλάζει σε 10 εκτελούνται οι αντίστοιχες ενέργειες μέχρι το βήμα 10.

4.1.3 Μοντέλα που δημιουργήθηκαν

Χρησιμοποιώντας μια συνάρτηση για την επιλογή παραμέτρων δημιουργήθηκαν 4 μοντέλα, αυτά με βάση το αν χρησιμοποιούνταν απλή διπλή ή τριπλή εκθετική ομαλοποίηση και ο τρόπος που συνδέονται μεταξύ τους (Brownlee, 2018). Πιο αναλυτικά τα μοντέλα ήταν:

Μοντέλο no no: Ένα μοντέλο που υπάρχει μόνο ο όρος του επιπέδου, είναι δηλαδή απλή εκθετική ομαλοποίηση, οι όροι προστίθενται μεταξύ τους.

Μοντέλο ad no: Ένα μοντέλο διπλής εκθετικής ομαλοποίησης (Holt's linear model) με τους όρους επιπέδου και τάσης δεν υπάρχει ο όρος της εποχικότητας, οι όροι προστίθενται μεταξύ τους. Μοντέλο ad ad: Ένα μοντέλο τριπλής εκθετικής ομαλοποίησης (Holt Winters model) με τους όρους επίπεδο, τάση και εποχικότητα να προστίθενται μεταξύ τους.

Μοντέλο ad m: Ένα μοντέλο τριπλής εκθετικής ομαλοποίησης με τους όρους επιπέδου και τάσης να προστίθενται μεταξύ τους και ύστερα ο όρος εποχικότητας να πολλαπλασιάζεται με την πρόσθεση αυτών.

4.1.4 Εξαγωγή αποτελεσμάτων σχεδίαση διαγραμμάτων

Για να μπορέσει η απόδοση των μοντέλων να συγκριθεί με άλλα μοντέλα και σε όλες τις λεκάνες έπρεπε να αδιαστατοποιηθεί το σφάλμα οπότε χρησιμοποιήθηκε ο συντελεστής απόδοσης Nash-Sutcliffe και μια μετατροπή του με την προσθήκη απόλυτης τιμής. Αυτά τα 2 μεγέθη ποσοτικοποίησαν το σφάλμα του εκάστοτε μοντέλου για όλες τις περιπτώσεις και υλοποιήθηκαν με καλώντας κάθε φορά μια συνάρτηση. Έτσι, στις τελευταίες 365 ημέρες μιας λεκάνης συγκρίθηκαν οι 'εκτιμώμενες από το μοντέλο' τιμές με τις πραγματικές τιμές. Επιπρόσθετα, σχεδιάστηκαν οι δύο αυτές χρονοσειρές σε κοινό διάγραμμα.

4.2 ARIMA

Η μέθοδος ARIMA όπως και η μέθοδος της εκθετικής ομαλοποίησης χρησιμοποιήθηκαν για να προβλέψουν την απορροή σε ημερίσια βάση με ορίζοντα πρόγνωσης 1, 5 και 10 ημέρες. Στην περίπτωση των μοντέλων ARIMA όπως και στην εκθετική ομαλοποίηση ο ορίζοντας πρόγνωσης μπορεί να αλλάξει στον κώδικα με χρήση μιας μόνο μεταβλητής οπότε οι ίδιες γραμμές κώδικα θα εκτελεστούν και για τις τρεις περιπτώσεις του ορίζοντα πρόγνωσης (Brownlee, 2018).

4.2.1 Δεδομένα

Τα δεδομένα που χρησιμοποιήθηκαν ήταν όπως και της εκθετικής ομαλοποίησης τα δεδομένα απορροής από τη βάση δεδομένων των ποταμών των ΗΠΑ, CAMELS. Από εκεί επιλέχθηκαν 10 λεκάνες και εφαρμόστηκαν τα διαφορετικά μοντέλα της μεθόδου ARIMA.

4.2.2 Η επιλογή των παραμέτρων p , d , q και τα αντίστοιχα τεστ.

Τα μοντέλα ARIMA είναι μια μεγάλη οικογένεια μοντέλων με παραμέτρους p, d, q . Αποφασίστηκε να χρησιμοποιηθούν κάποια τεστ τα οποία θα ορίσουν έναν ελάχιστο και έναν μέγιστο αριθμό για τις τιμές p, q, d . Αυτές θα οριοθετήσουν το εύρος της αναζήτησης πλέγματος που θα γίνει στη

συνέχεια. Με τον όρο ‘αναζήτηση πλέγματος’ εννοείται ότι θα δημιουργηθούν όλοι οι δυνατοί συνδυασμοί μοντέλων που επιτρέπουν οι ελάχιστες και οι μέγιστες των παραμέτρων p , d , q .

Για τον προσδιορισμό της τιμής d που ορίζει την τάξη διαφοροποίησης εφαρμόστηκε το Augmented Dickey Fuller test για να εξεταστεί αν η χρονοσειρά είναι στάσιμη ή όχι. Στη συνέχεια, παρατίθεται ο πίνακας που δείχνει τον κωδικό της κάθε λεκάνης και την τιμή της p -value που αναφέρεται στη στασιμότητα ή μη της χρονοσειράς της λεκάνης. Αυτός ο πίνακας κατασκευάστηκε αφού εκτελέστηκε επαναληπτικά το υποπρόγραμμα ‘adf_stationarity_test’ για κάθε λεκάνη. Σημειώνεται πως αν η p -value είναι μικρότερη από 0.05 η χρονοσειρά εύλογα μπορεί να θεωρηθεί στάσιμη (Dickey & Fuller, 1979).

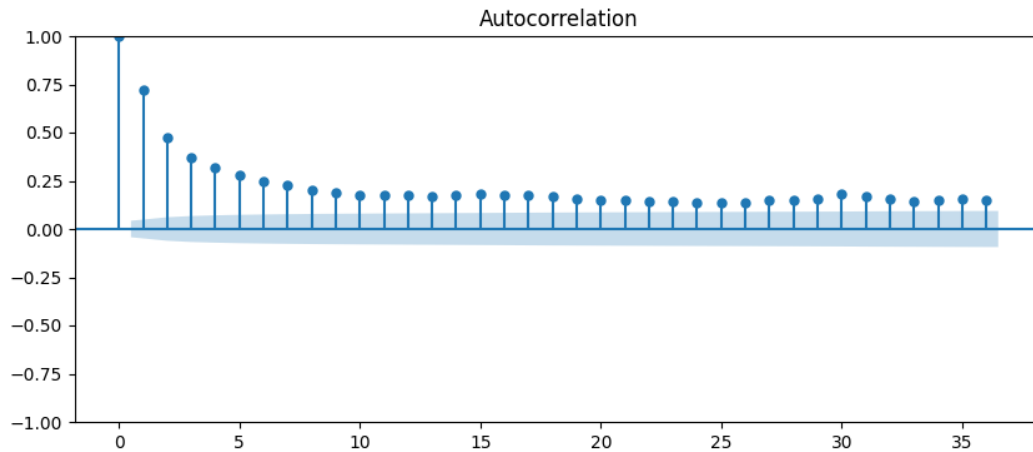
Πίνακας 1. Τιμές p -value μετά από Augmented Dickey Fuller test για όλες τις λεκάνες

ID λεκάνης	p -value
02014000	$1.51 \cdot 10^{-17}$
01411300	$5.38 \cdot 10^{-7}$
01440000	$1.14 \cdot 10^{-18}$
01543500	$1.7 \cdot 10^{-23}$
02016000	$1.24 \cdot 10^{-18}$
02053800	$9.54 \cdot 10^{-9}$
02059500	$2.4 \cdot 10^{-7}$
02069700	$6.45 \cdot 10^{-16}$
02143000	$1.5 \cdot 10^{-6}$
02350900	$1.75 \cdot 10^{-14}$

Από τον ανωτέρω πίνακα είναι κατανοητό πως όλες οι χρονοσειρές είναι στάσιμες καθώς όλες οι τιμές είναι μικρότερες από 0.01. Αυτό σημαίνει ότι η διαφοροποίηση θα έχει αμελητέο ρόλο στην εκτίμηση πιο στοχευμένης απορροής. Παρά το γεγονός αυτό αποφασίστηκε το εύρος της τιμής d να είναι από 0 έως 1 και επομένως να αναπτυχθούν μοντέλα με μιας τάξης διαφοροποίησης.

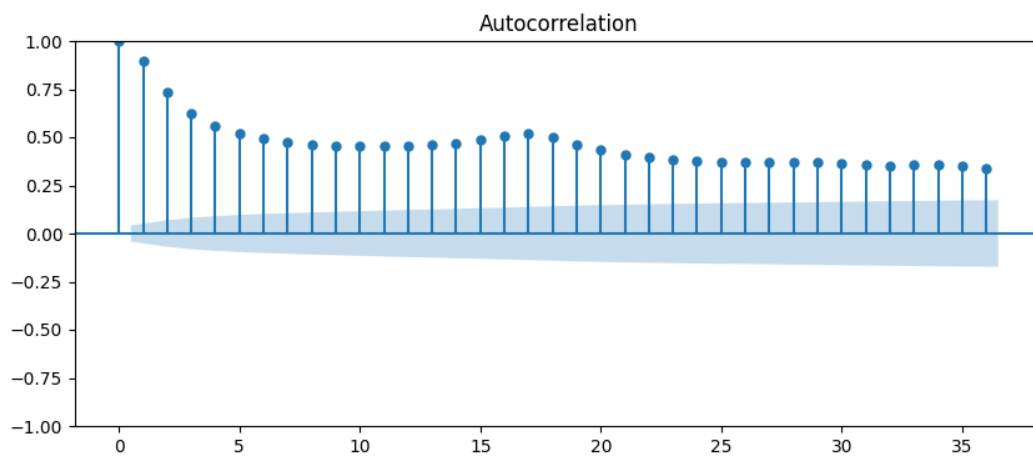
Για τον προσδιορισμό της τιμής p που ορίζει την τάξη του μοντέλου αυτοπαλινδρόμησης χρησιμοποιήθηκε η συνάρτηση αυτοσυσχέτισης (ACF, autocorrelation function). Το γράφημα της έγινε για κάθε λεκάνη. Στα παρακάτω διαγράμματα παρατηρείται πως η τιμή της αυτοσυσχέτισης βαίνει μειούμενη όσο ο χρονικός ορίζοντας αυξάνει και σταθεροποιείται μετά. Η συμπεριφορά αυτή ποικίλει βεβαίως ανάλογα και με τα χαρακτηριστικά της κάθε χρονοσειράς. Στη συνέχεια παρατίθεται ο κωδικός κάθε λεκάνης και το διάγραμμα αυτοσυσχέτισης.

1. Id λεκάνης: 02014000



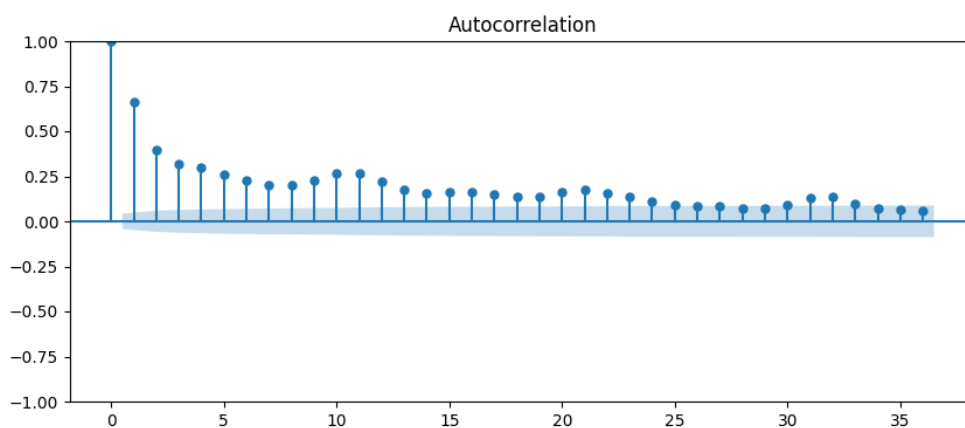
Εικόνα 34 Διάγραμμα αυτοσυσχέτισης χρονοσειράς απορροής για την λεκάνη 1

2. Id λεκάνης: 01411300



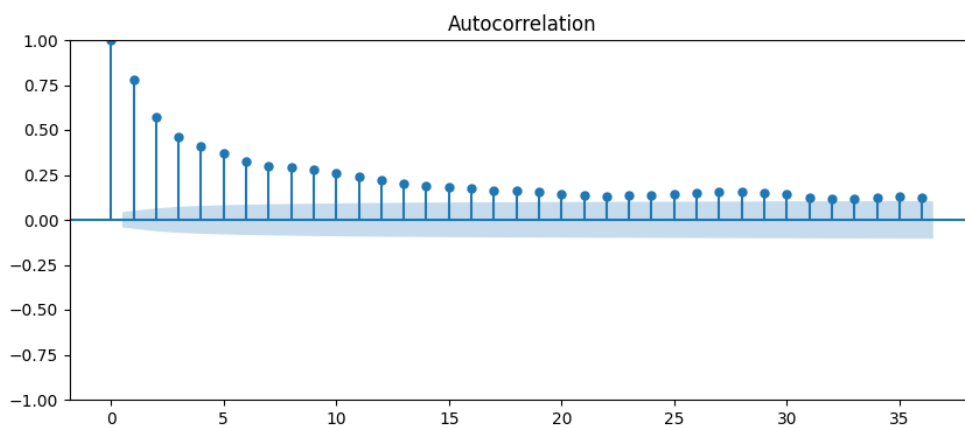
Εικόνα 35 Διάγραμμα αυτοσυσχέτισης χρονοσειράς απορροής για τη λεκάνη 2

3. Id λεκάνης: 01440000



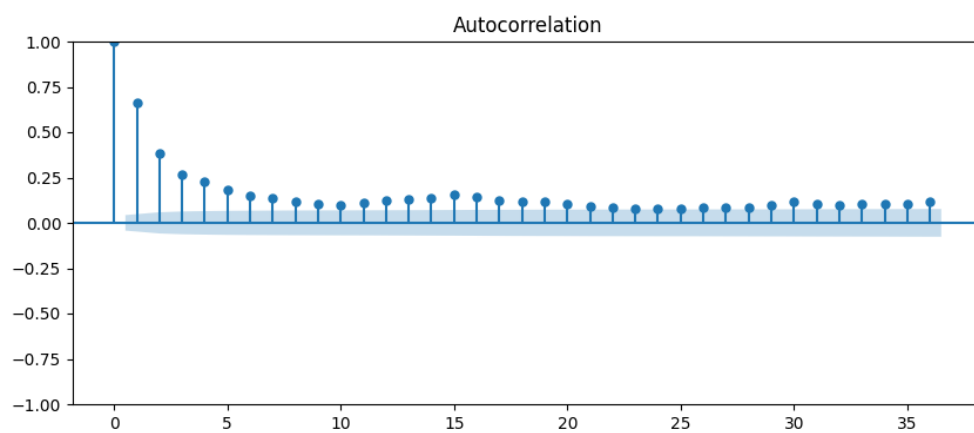
Εικόνα 36 Διάγραμμα αυτοσυσχέτισης χρονοσειράς απορροής για τη λεκάνη 3

4. Id λεκάνης: 01543500



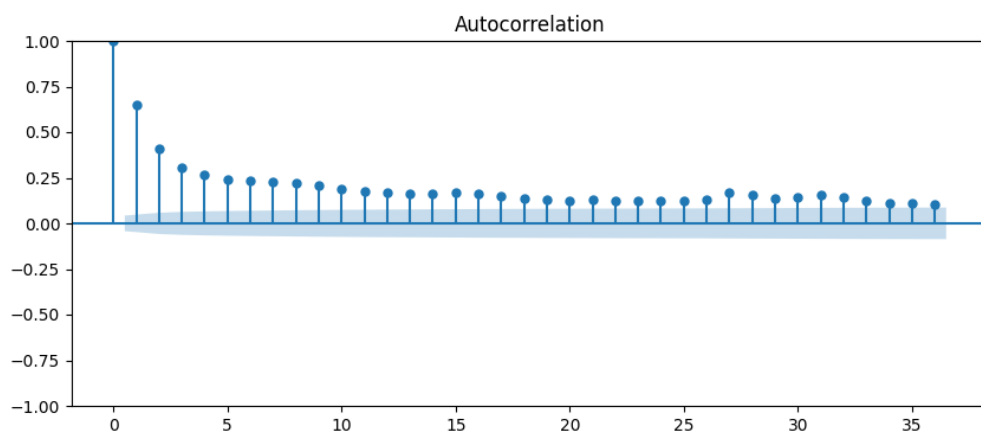
Εικόνα 37 Διάγραμμα αυτοσυσχέτισης χρονοσειράς απορροής για τη λεκάνη 4

5. Id λεκάνης: 02016000



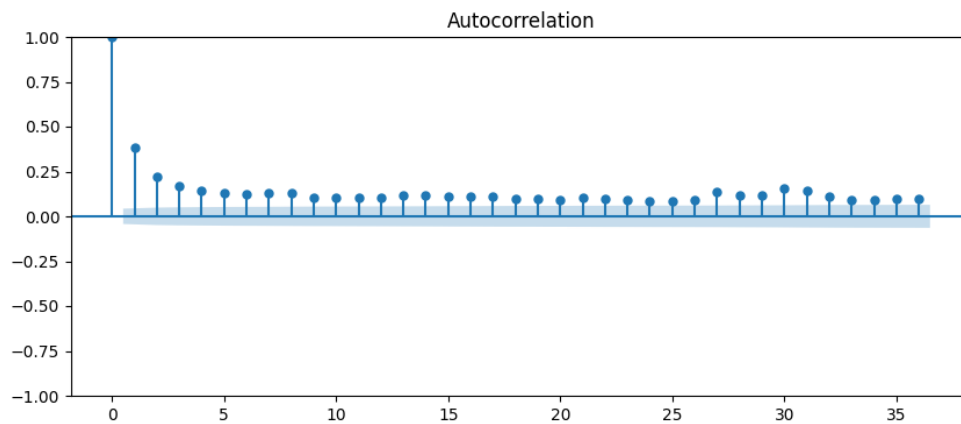
Εικόνα 38 Διάγραμμα αυτοσυσχέτισης χρονοσειράς απορροής για τη λεκάνη 5

6. Id λεκάνης: 02053800



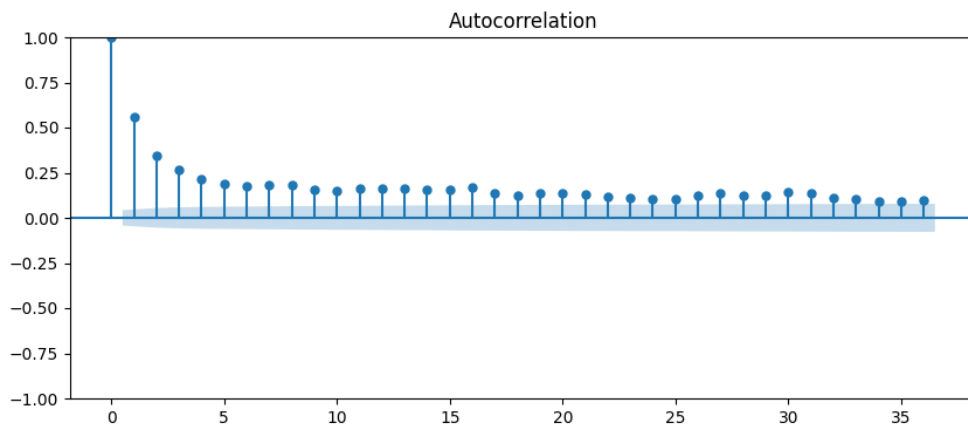
Εικόνα 39 Διάγραμμα αυτοσυσχέτισης χρονοσειράς απορροής για τη λεκάνη 6

7. Id λεκάνης: 02059500



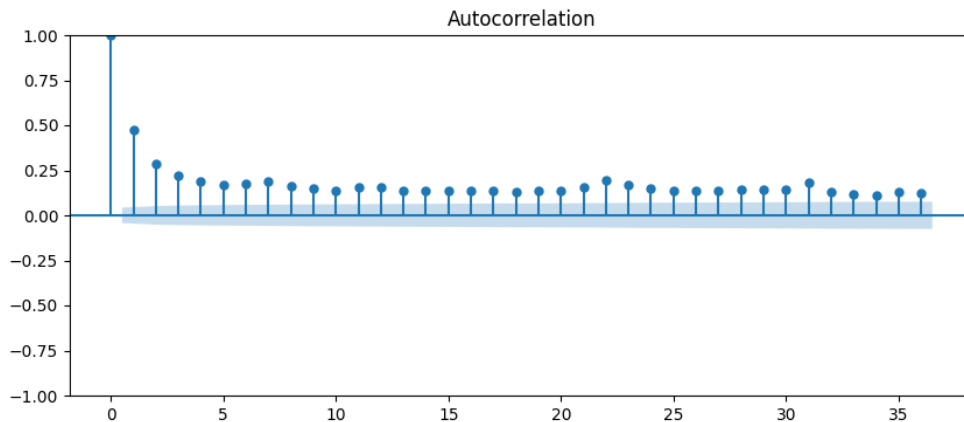
Εικόνα 40 Διάγραμμα αυτοσυσχέτισης χρονοσειράς απορροής για τη λεκάνη 7

8. Id λεκάνης: 02069700



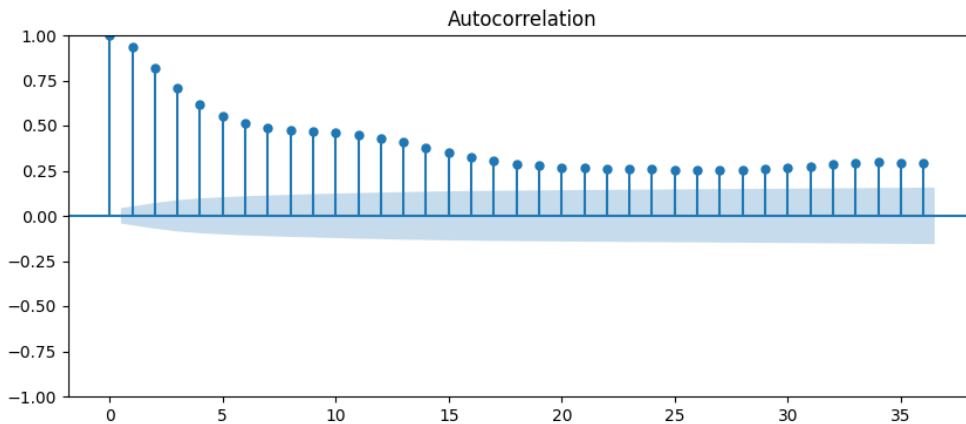
Εικόνα 41 Διάγραμμα αυτοσυσχέτισης χρονοσειράς απορροής για τη λεκάνη 8

9. Id λεκάνης: 02143000



Εικόνα 42 Διάγραμμα αυτοσυσχέτισης χρονοσειράς απορροής για τη λεκάνη 9

10. Id λεκάνης: 0235090

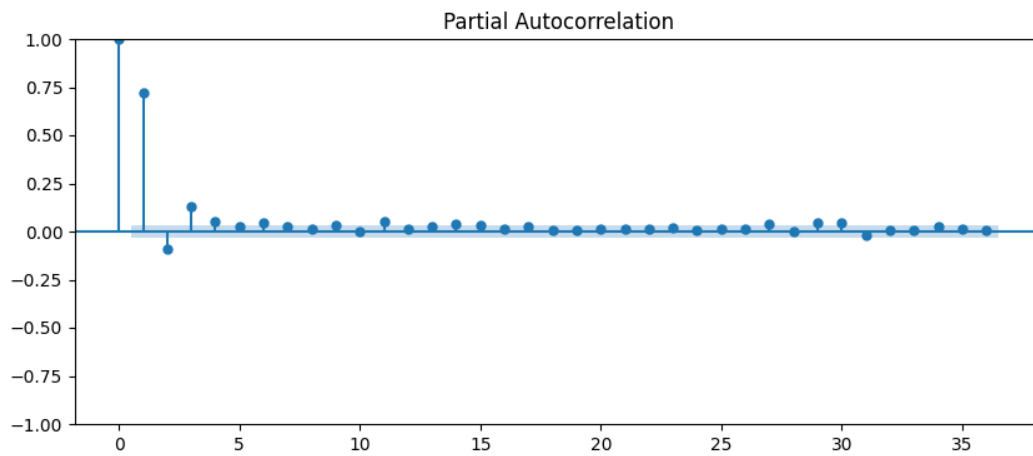


Εικόνα 43 Διάγραμμα αυτοσυσχέτισης χρονοσειράς απορροής για τη λεκάνη 10

Έχοντας υπόψη τα διαγράμματα αλλά και το γεγονός ότι η συνάρτηση αυτοσυσχέτισης υποδεικνύει τον αριθμό του όρου p του μοντέλου αυτοπαλινδρόμησης αποφασίστηκε το εύρος του p να είναι από 0 έως 2. Είναι προφανές ότι θα ήταν αδύνατο ή τουλάχιστον υπερβολικά κοστοβόρο να χρησιμοποιηθούν 35 όροι στο μοντέλο για αυτό και αποφασίστηκαν οι δύο πρώτοι όροι ως οι πιο σημαντικοί σε κάθε λεκάνη.

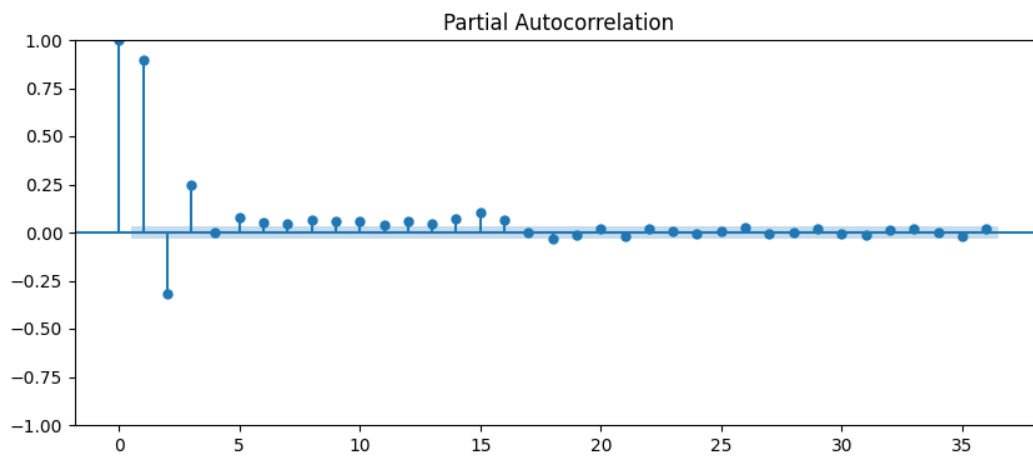
Για τον προσδιορισμό της παραμέτρου d που χρησιμοποιείται ώστε να βρεθεί η τάξη των μοντέλων του κινητού μέσου όρου κατασκευάστηκε το διάγραμμα της συνάρτησης μερικής αυτοσυσχέτισης για κάθε λεκάνη. Τα διαγράμματα και τα id των λεκανών παρατίθενται στη συνέχεια.

1. Id λεκάνης: 02014000



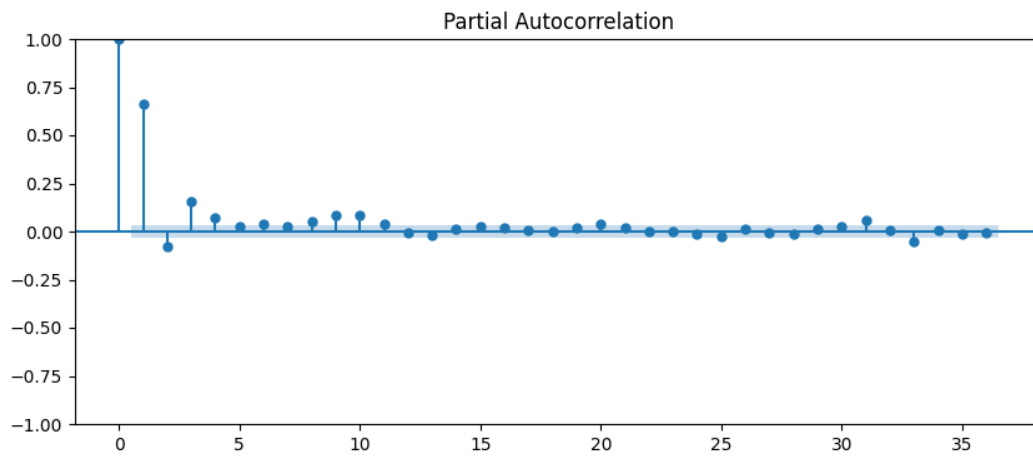
Εικόνα 44 Διάγραμμα μερικής αυτοσυσχέτισης χρονοσειράς απορροής για τη λεκάνη 1

2. Id λεκάνης: 01411300



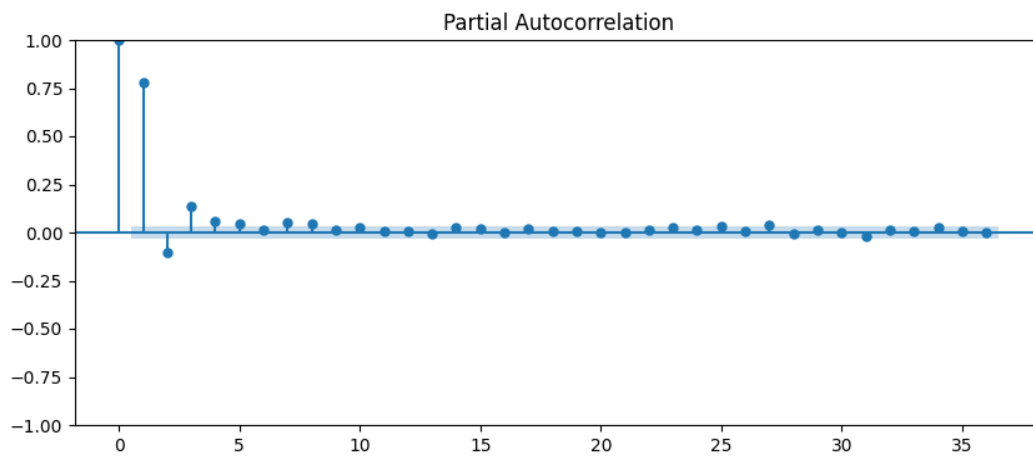
Εικόνα 45 Διάγραμμα μερικής αυτοσυσχέτισης χρονοσειράς απορροής για τη λεκάνη 2

3. Id λεκάνης: 01440000



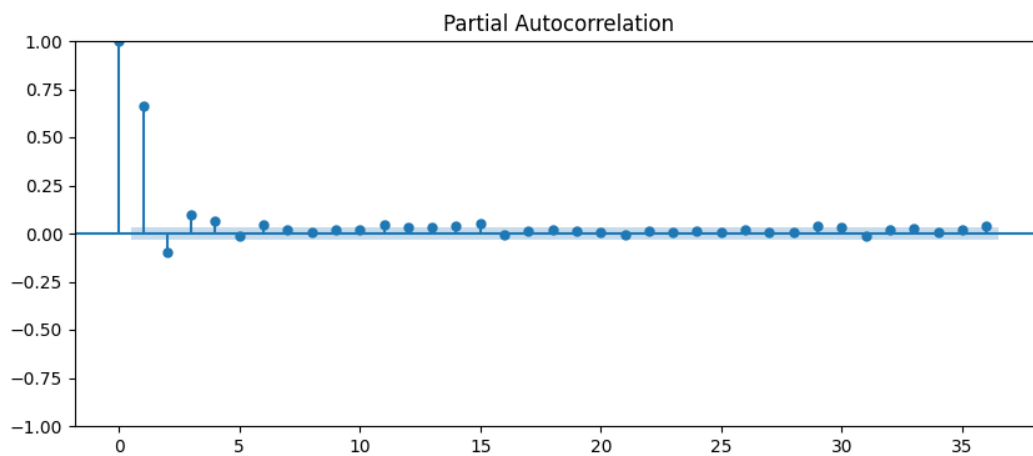
Εικόνα 46 Διάγραμμα μερικής αυτοσσχέτισης χρονοσειράς απορροής για τη λεκάνη 3

4. Id λεκάνης: 01543500



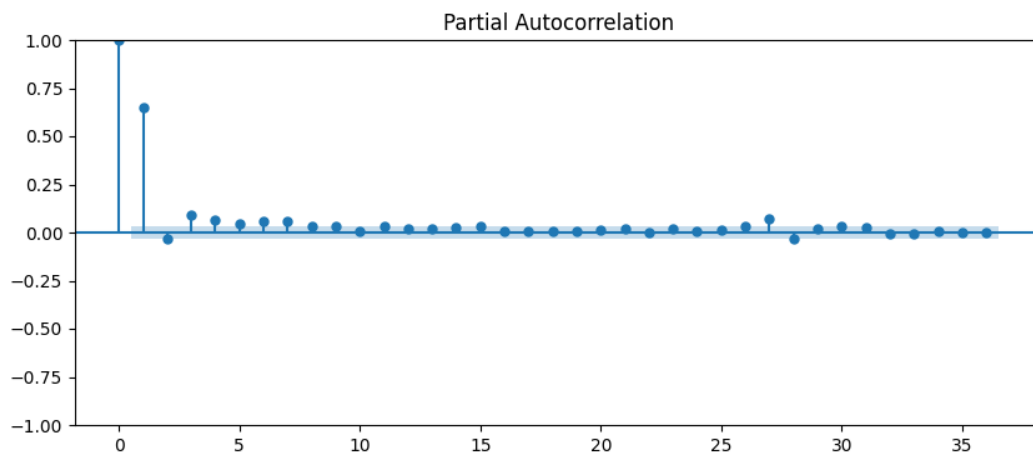
Εικόνα 47 Διάγραμμα μερικής αυτοσσχέτισης χρονοσειράς απορροής για τη λεκάνη 4

5. Id λεκάνης: 02016000



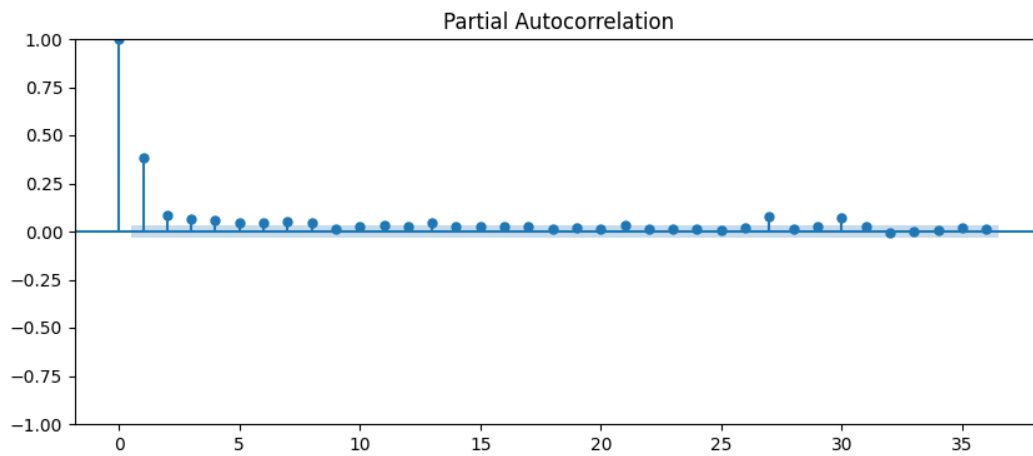
Εικόνα 48 Διάγραμμα μερικής αυτοσσχέτισης χρονοσειράς απορροής για τη λεκάνη 5

6. Id λεκάνης: 02053800



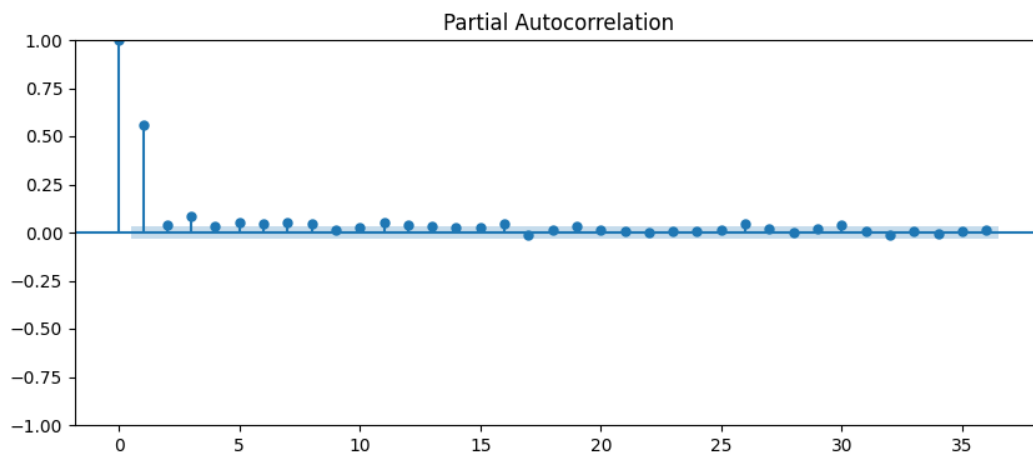
Εικόνα 49 Διάγραμμα μερικής αυτοσσχέτισης χρονοσειράς απορροής για τη λεκάνη 6

7. Id λεκάνης: 02059500



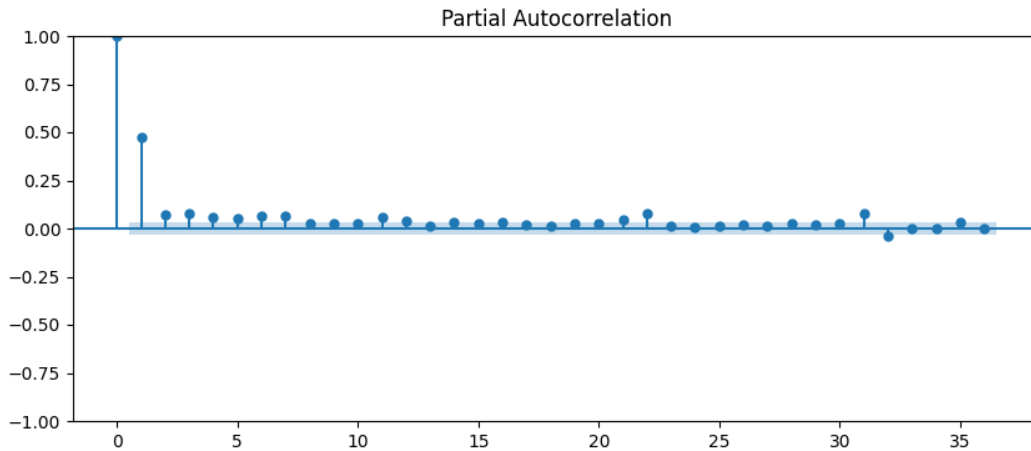
Εικόνα 50 Διάγραμμα μερικής αυτοσυσχέτισης χρονοσειράς απορροής για τη λεκάνη 7

8. Id λεκάνης: 02069700



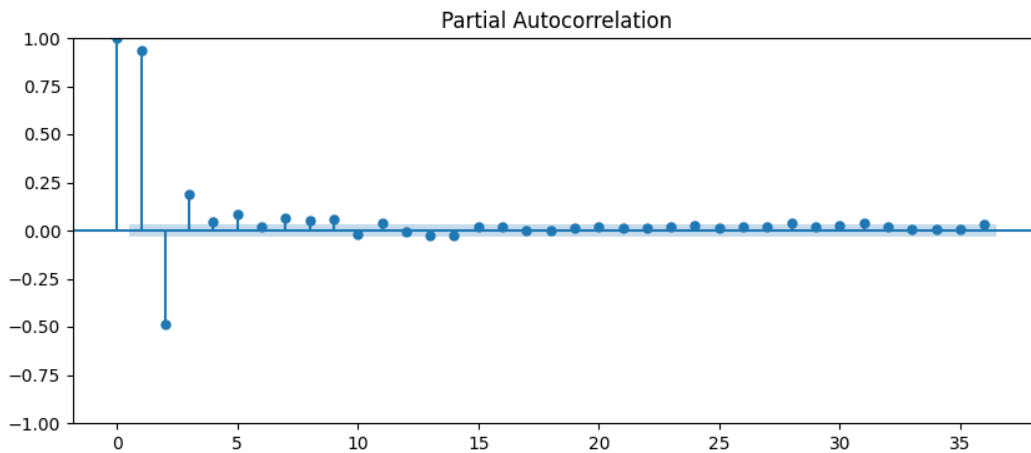
Εικόνα 51 Διάγραμμα μερικής αυτοσυσχέτισης χρονοσειράς απορροής για τη λεκάνη 8

9. Id λεκάνης: 02143000



Εικόνα 52 Διάγραμμα μερικής αυτοσυσχέτισης χρονοσειράς απορροής για τη λεκάνη 9

10. Id λεκάνης: 02350900



Εικόνα 53 Διάγραμμα μερικής αυτοσυσχέτισης χρονοσειράς απορροής για τη λεκάνη 10

Σε αυτή την περίπτωση είναι πλέον εμφανές πως οι όροι που υποδεικνύονται από με τη μερική συσχέτιση είναι οι πρώτοι δυο οπότε και επιλέχθηκε το εύρος των τιμών του q να είναι από 0 έως 2.

4.2.3 Πρόγνωση απορροής

Στην παρούσα περίπτωση όπως και στην εκθετική ομαλοποίηση η πρόγνωση γίνεται με τη συνδρομή 2 συναρτήσεων. Οι συναρτήσεις έχουν εκτελεστεί για τις 10 λεκάνες για τους 3 ορίζοντες πρόγνωσης 1, 5 και 10 ημέρες. Η πρώτη συνάρτηση ονομάζεται 'arima_forecast' εκτιμά την απορροή της αμέσως επόμενης ημέρας λαμβάνοντας υπόψη τις απορροές των προηγούμενων ημερών. Αυτή η συνάρτηση είναι ο βασικός πυλώνας της πρόγνωσης και χρησιμοποιείται επαναλαμβανόμενα όπως είχε γίνει και με την εκθετική ομαλοποίηση. Η δεύτερη συνάρτηση

ονομάζεται 'arima_walk_forward_validation' και πραγματοποιεί την ίδια επαναληπτική διαδικασία που εκτελεί η 'holt_winters_walk_forward_validation'. Έχοντας την πρόγνωση της επόμενης ημέρας από την πρώτη συνάρτηση, η walk forward δημιουργεί μια προσωρινή χρονοσειρά που έχει αποθηκευμένη την εκτιμώμενη τιμή απορροής. Η προσωρινή χρονοσειρά χρησιμοποιείται ως δεδομένο για να εκτιμηθεί η απορροή της επόμενης μέρας (στην περίπτωση του ορίζοντα πρόγνωσης των 5, 10 ημερών) μέσω την πρώτης συνάρτησης 'arima_forecast'. Η προσωρινή χρονοσειρά προσαυξάνεται με τις επιπλέον εκτιμήσεις μέχρις ότου να καλυφθεί ο ορίζοντας πρόγνωσης. Έτσι, μετά την πρόγνωση της 5^{ης} ημέρας οι εκτιμώμενες τιμές της χρονοσειράς (δηλαδή των τελευταίων 5 ημερών) αντικαθίστανται από τις πραγματικές. Αυτή η διαδικασία επαναλαμβάνεται μέχρι να ολοκληρωθούν και οι 365 ημέρες του τέστ.

4.2.4 Αρχικοποίηση παραμέτρων και δημιουργία μοντέλων

Έχοντας το εύρος τιμών για κάθε μια από τις παραμέτρους p , d , q δημιουργήθηκε μια συνάρτηση για να γίνει η αναζήτηση πλέγματος. Τα μοντέλα που δημιουργήθηκαν στην αναζήτηση αυτή δεν είναι μόνο ARIMA αλλά και για τα υποσύνολα του που είναι τα ARMA, AR, MA. Αναφέρονται επιγραμματικά στη συνέχεια.

1. MA(1) Κινητού μέσου όρου, $p = 0$, $d = 0$, $q = 1$
2. MA(2) Κινητού μέσου όρου, $p = 0$, $d = 0$, $q = 2$
3. I(d) Μοντέλο διαφοροποίησης, $p = 0$, $d = 1$, $q = 0$
4. ARIMA(0,1,1) Μοντέλο διαφοροποίησης και κινητού μέσου όρου, $p = 0$, $d = 1$, $q = 1$
5. ARIMA(0,1,2) Μοντέλο διαφοροποίησης και κινητού μέσου όρου, $p = 0$, $d = 1$, $q = 2$
6. AR(1) Μοντέλο αυτοπαλινδρόμησης, $p = 1$, $d = 0$, $q = 0$
7. ARMA(1,1) Μοντέλο αυτοπαλινδρόμησης και κινητού μέσου όρου, $p = 1$, $d = 0$, $q = 1$
8. ARMA(1,2) Μοντέλο αυτοπαλινδρόμησης και κινητού μέσου όρου, $p = 1$, $d = 0$, $q = 2$
9. ARIMA(1,1,0) Μοντέλο διαφοροποίησης και αυτοπαλινδρόμησης, $p = 1$, $d = 1$, $q = 0$
10. ARIMA(1,1,1) Μοντέλο διαφοροποίησης, αυτοπαλινδρόμησης και κινητού μέσου όρου, $p = 1$, $d = 1$, $q = 1$
11. ARIMA (1,1,2) Μοντέλο διαφοροποίησης, αυτοπαλινδρόμησης και κινητού μέσου όρου, $p = 1$, $d = 1$, $q = 2$
12. AR(2) Μοντέλο αυτοπαλινδρόμησης, $p = 2$, $d = 0$, $q = 0$
13. ARMA(2,1) Μοντέλο αυτοπαλινδρόμησης και κινητού μέσου όρου, $p = 2$, $d = 0$, $q = 1$
14. ARMA(2,2) Μοντέλο αυτοπαλινδρόμησης και κινητού μέσου όρου, $p = 2$, $d = 0$, $q = 2$
15. ARIMA(2,1,0) Μοντέλο διαφοροποίησης και αυτοπαλινδρόμησης, $p = 2$, $d = 1$, $q = 0$

16. ARIMA(2,1,1) Μοντέλο διαφοροποίησης, αυτόπαλινδρόμησης και κινητού μέσου όρου, $p = 2$, $d = 1$, $q = 1$

17. ARIMA(2,1,2) Μοντέλο διαφοροποίησης, αυτόπαλινδρόμησης και κινητού μέσου όρου, $p = 2$, $d = 1$, $q = 2$

4.2.5 Υπολογισμός σφάλματος και γραφικές παραστάσεις

Χρησιμοποιήθηκαν οι συναρτήσεις που υπολογίζουν τους συντελεστές Nash-Sutcliffe Efficiency και Absolute Nash-Sutcliffe Efficiency για τον προσδιορισμό του σφάλματος για να είναι εφικτό να συγκριθούν τα διαφορετικά μοντέλα μεταξύ τους αλλά και μεταξύ τους και με διαφορετικό ορίζοντα πρόγνωσης καθώς οι δείκτες αυτοί μπορούν να μας δώσουν πληροφορίες για το πόσο βελτιώθηκε η δυνατότητα εκτίμησης του εκάστοτε μοντέλου από τις 10 ημέρες χωρίς ανανέωση, στις 5 και ύστερα στη 1. Έχοντας την χρονοσειρά που εκτιμήθηκε από το εκάστοτε μοντέλο και την πραγματική χρονοσειρά της λεκάνης, για κάθε μοντέλο έγινε η γραφική αναπαράσταση των δύο χρονοσειρών σε κοινό διάγραμμα, όπως και στην περίπτωση της εκθετικής ομαλοποίησης.

4.3 Μέθοδοι Deep Learning στην Python

Την ανάλυση της εργασίας αυτής ολοκληρώνει η μέθοδος βαθιάς μηχανικής μάθησης. Εδώ αξίζει να επισημανθεί πως δημιουργήθηκαν δύο προγράμματα βασισμένα στη μέθοδο του deep learning το ένα είχε ως σκοπό την πρόγνωση όπως εξυπηρετούν τα ARIMA και εκθετική ομαλοποίηση που έχουν ήδη αναφερθεί και το άλλο είχε ως σκοπό την μοντελοποίηση-προσομοίωση της λεκάνης εκτιμά και αυτό απορροή. Τα δύο προγράμματα αν και έχουν παρεμφερή αρχιτεκτονική και χρησιμοποιούν τον τύπο νευρωνικού δικτύου πολλαπλών στρωμάτων (multi layer perceptron), χρησιμοποιούν άλλα δεδομένα και η εφαρμογή τους έχει διαφορετικές προεκτάσεις για αυτούς τους λόγους θα αναφερθούν σε δύο ξεχωριστά υποκεφάλαια εδώ.

4.3.1 Πρόγνωση με τη μέθοδο του ‘κυλιόμενου παραθύρου’

Η μέθοδος ‘κινούμενου παραθύρου’ στην αγγλική βιβλιογραφία γνωστή ως rolling window είναι μια μέθοδος της ανάλυσης δεδομένων που χρησιμοποιεί έναν ορισμένο αριθμό από προηγούμενα δεδομένα (στην περίπτωσή της εργασίας 5). Εκτιμά με την ίδια έννοια που εκτιμούν τα μοντέλα ARIMA, εκθετικής ομαλοποίησης αλλά για ορίζοντα πρόγνωσης μόνο την 1 ημέρα οπότε μπορεί να συγκριθεί με τα μοντέλα που προαναφέρθηκαν αλλά μόνο στην περίπτωση της ημερίσιας ανανέωσης δεδομένων (Li Shen1, 2021).

4.3.1.1 Δεδομένα και προεπεξεργασία

Τα δεδομένα ήταν οι απορροές από τις προαναφερθείσες 10 λεκάνες. Αφού εισήχθησαν από το αρχείο κειμένου σε μορφή Pandas DataFrame χρησιμοποιήθηκε μια συνάρτηση για να δημιουργήσει την επιθυμητή μορφή. Η συνάρτηση αυτή ονομάζεται ‘split_for_rolling’ και έχει ως σκοπό να μετατρέψει τη χρονοσειρά σε πίνακα. Η μορφή του πίνακα είναι τέτοια ώστε 5 τιμές να είναι τα δεδομένα εισόδου και η έκτη τιμή να είναι η πρόγνωση. Πιο αναλυτικά, έστω ότι υπάρχει μια χρονοσειρά από το 1 έως το 10, αυτή σε μορφή πίνακα που προαναφέρθηκε θα γίνει ως εξής

Πίνακας 2 Υποτυπώδης πίνακας όπου μια χρονοσειρά από το 1-10 μετατρέπεται σε δεδομένα μορφής πίνακα

Δεδομένα εισόδου	Δεδομένα εξόδου
1 2 3 4 5	6
2 3 4 5 6	7
3 4 5 6 7	8
4 5 6 7 8	9
5 6 7 8 9	10

Καλώντας λοιπόν τη συνάρτηση αυτή όλη η χρονοσειρά λαμβάνει αυτή τη μορφή πίνακα και είναι έτοιμη να εισαχθεί στο νευρωνικό δίκτυο. Σημειώνεται ότι οι τελευταίες 365 ημέρες που έχουν ορισθεί σαν τιμές για αξιολόγηση (test set) έχουν διαχωριστεί από πριν και έχουν λάβει και αυτές τη μορφή πίνακα με χρήση της ίδιας συνάρτησης.

4.3.1.2 Νευρωνικό δίκτυο

Το νευρωνικό δίκτυο που κατασκευάστηκε έχει 6 κρυφά στρώματα και το κάθε ένα έχει τους εξής νευρώνες:

1. 128 νευρώνες
2. 64 νευρώνες
3. 32 νευρώνες
4. 32 νευρώνες
5. 16 νευρώνες
6. 8 νευρώνες

Είναι προφανές πως το στρώμα εισόδου έχει 5 νευρώνες (όσα είναι και τα δεδομένα εισόδου) και το στρώμα εξόδου έχει 1 νευρώνα (το δεδομένο εξόδου, η τιμή που καλείται το νευρωνικό δίκτυο να εκτιμήσει).

Η συνάρτηση κόστους υπολογίζει το μέσο τετραγωνικό σφάλμα για την καθοδική κλίση (gradient descend). Χρησιμοποιήθηκε ο αλγόριθμος βελτιστοποίησης ADAM που επιτρέπει στο ρυθμό μάθησης να προσαρμόζεται σε κάθε βήμα έτσι επιτυγχάνει πιο αποδοτικά την εύρεση ελάχιστης τιμής. Η αρχική τιμή του ρυθμού μάθησης όπως και το πλήθος των επαναληπτικών εκπαιδεύσεων στα δεδομένα (epochs) άλλαξε χειροκίνητα κάθε φορά που εκτελούνταν οι κώδικες για άλλη λεκάνη. Η ανωτέρω αλλαγή κρίνεται απαραίτητη για να επιτευχθεί ο περιορισμός των σφαλμάτων στις εκτιμήσεις. Η συνάρτηση ενεργοποίησης (activation function) σε κάθε νευρώνα ήταν η ReLu εκτός από τον τελευταίο νευρώνα που δεν χρησιμοποιήθηκε κάποια συνάρτησης ενεργοποίησης.

Έχοντας εκπαιδεύσει το νευρωνικό δίκτυο, αφού προσδιοριστούν τα βάρη στο μοντέλο από εκεί και έπειτα ο πίνακας των τελευταίων 365 ημερών αφού εισαχθεί στο δίκτυο δείχνει πόσο καλά το μοντέλο έχει εκτιμήσει τις απορροές.

4.3.1.3 Υπολογισμός σφάλματος και γραφικές παραστάσεις

Για τον υπολογισμό του σφάλματος χρησιμοποιήθηκαν οι συναρτήσεις που υπολογίζουν το Absolute Nash-Sutcliffe efficiency και το Nash-Sutcliffe efficiency. Οι ίδιοι δείκτες που χρησιμοποιήθηκαν στα μοντέλα ARIMA και εκθετικής ομαλοποίησης. Σημειώνεται ότι οι δείκτες αυτοί μπορούν να συγκριθούν με τα μοντέλα που έχουν αναφερθεί είδη. Τέλος έγιναν οι γραφικές παραστάσεις των δύο χρονοσειρών σε κοινό διάγραμμα, δηλαδή της χρονοσειράς που εκτιμήθηκε από το νευρωνικό και της πραγματικής χρονοσειράς της λεκάνης.

4.3.2 Προσομοίωση-μοντελοποίηση λεκάνης με τη μέθοδο νευρωνικού δικτύου

Η παρούσα μέθοδος διαφέρει από τις προηγούμενες διότι για αυτές τα δεδομένα εισόδου είναι παρελθοντικές τιμές απορροής (για πρόγνωση τη χρονική στιγμή t , δίνεται απορροή της $t-1$) ενώ στην παρούσα περίπτωση τα δεδομένα εισόδου είναι τιμές που δεν περιλαμβάνουν απορροή και που αναφέρονται στην ίδια χρονική στιγμή με την πρόγνωση (για πρόγνωση τη χρονική στιγμή t , δίνονται δεδομένα τη χρονική στιγμή t) (W.Trask, 2019).

4.3.2.1 Δεδομένα και προεπεξεργασία

Τα δεδομένα ελήφθησαν και αυτά από το CAMELS που είναι μια βάση δεδομένων των ΗΠΑ. Ως δεδομένα εισόδου χρησιμοποιήθηκαν για τις 10 λεκάνες τα αδιάστατα δεδομένα:

1. Βροχής
2. Ηλιοφάνειας
3. Ημερομηνία
4. Ηλιακή ακτινοβολία
5. Μέγιστη θερμοκρασία
6. Ελάχιστη θερμοκρασία
7. Πίεση
8. Επιφάνεια λεκάνης

Τα δεδομένα πέρασαν από μια διαδικασία κανονικοποίησης κατά την οποία όλα μεταφράστηκαν σε μέγιστη τιμή 1 και ελάχιστη τιμή 0. Αυτό έγινε γιατί λόγω των διαφορετικών μονάδων μέτρησης μπορεί λιγότερο σημαντικές μεταβλητές εισόδου να φαίνεται ότι επηρεάζουν πιο πολύ το αποτέλεσμα από ότι πραγματικά συμβαίνει. Παραδείγματος χάρη αν η ηλιοφάνεια αυξηθεί κατά 5000 s μία μέρα και αντίστοιχα η βροχή αυξηθεί κατά 15 mm την ίδια μέρα το νευρωνικό είναι πιθανό να 'θεωρήσει' ότι για μια ενδεχόμενη αύξηση απορροής ευθύνεται η άνοδος 5000 s του κόμβου ηλιοφάνειας και όχι 15 mm του κόμβου βροχόπτωσης. Η κανονικοποίηση έγινε με τη συνάρτηση 'StandardScaler'.

Τα δεδομένα επίσης πέρασαν από μια διαδικασία διαχωρισμού σε σετ εκπαίδευσης και σετ αξιολόγησης. Το τελευταίο έτος έμεινε για την αξιολόγηση όπως συνέβει με τα μοντέλα ARIMA και εκθετικής ομαλοποίησης

4.3.2.2 Νευρωνικό δίκτυο

Η αρχιτεκτονική του νευρωνικού δικτύου είναι παρόμοια με την αρχιτεκτονική του δικτύου του κυλιόμενου παραθύρου. Πιο συγκεκριμένα αποτελείται από 6 κρυφά στρώματα και οι νευρώνες στο κάθε ένα είναι:

1. 128 νευρώνες
2. 64 νευρώνες
3. 32 νευρώνες
4. 32 νευρώνες
5. 16 νευρώνες
6. 8 νευρώνες

Το πρώτο στρώμα (στρώμα εισόδου) έχει 8 νευρώνες, όσα και τα δεδομένα εισόδου. Το τελευταίο στρώμα (στρώμα εξόδου) έχει 1 νευρώνα την απορροή που καλείται να εκτιμήσει το μοντέλο.

Η συνάρτηση κόστους υπολογίζει και εδώ το μέσο τετραγωνικό σφάλμα και χρησιμοποιείται ο αλγόριθμος ADAM (Ng, 2018) για την καθοδική κλίση. Οι συναρτήσεις ενεργοποίησης είναι οι ReLu σε όλο το νευρωνικό εκτός του τελευταίου νευρώνα όπου δεν χρησιμοποιήθηκε συνάρτηση ενεργοποίησης. Οι επαναλήψεις του train set (epochs) παρέμειναν 50 και οι ανανεώσεις των βαρών έγιναν με πακέτα που το κάθε ένα είχε 32 παραδείγματα εκπαίδευσης (batches = 32).

Έχοντας εκπαιδεύσει το νευρωνικό δίκτυο το τεστ σετ διοχετεύεται στα βάρη που έχουν δημιουργηθεί και κατασκευάζεται έτσι η εκτιμώμενη χρονοσειρά.

4.3.2.3 Υπολογισμός σφαλμάτων και γραφικές παραστάσεις

Χρησιμοποιήθηκαν και πάλι οι συναρτήσεις που υπολογίζουν την απόδοση μέσω του Absolute Nash-Sutcliffe efficiency και του Nash-Sutcliffe efficiency. Σημειώνεται εδώ πως αυτό έγινε για τους λόγους που αναφέρθηκαν στην εισαγωγή της μεθόδου και στο κεφάλαιο επεξήγησης της θεωρίας της βαθιάς μηχανικής μάθησης. Τέλος, έγιναν οι γραφικές παραστάσεις των δύο χρονοσειρών (της εκτιμώμενης και της πραγματικής) σε κοινό γράφημα.

4.3.2.4 Προέκταση της μεθόδου

Είναι γνωστό από την αρχή της εργασίας πως τα δεδομένα που χρησιμοποιήθηκαν ήταν από το 1980 έως το 2014, 34 έτη που δημιουργούν περίπου 12410 ημερήσια παραδείγματα εκπαίδευσης. Με την χρήση του προγράμματος ‘mlp_transfer_learning’ κατασκευάστηκε ένα νευρωνικό δίκτυο που εκπαιδεύτηκε σε δεδομένα και από τις 10 λεκάνες αλλά και από νέα δεδομένα από νέες λεκάνες με αποτέλεσμα το σετ εκπαίδευσης να αριθμεί 163.510 δεδομένα. Αφού το νευρωνικό εκπαιδεύτηκε τα βάρη του αποθηκεύτηκαν σε ένα ξεχωριστό αρχείο ‘expe_model.h5’ στη συνέχεια κατασκευάστηκε ένα άλλο πρόγραμμα το ‘transfer_test’. Εκεί δόθηκαν νέες λεκάνες με ελάχιστα δεδομένα προς εκπαίδευση (2 έτη με 730 ημερίσια δεδομένα) και ζητήθηκε να εκτιμηθεί το επόμενο έτος της ‘άγνωστης λεκάνης’.

5 Αποτελέσματα

Συνολικά στην παρούσα εργασία πραγματοποιήθηκαν οι εξής διαδικασίες:

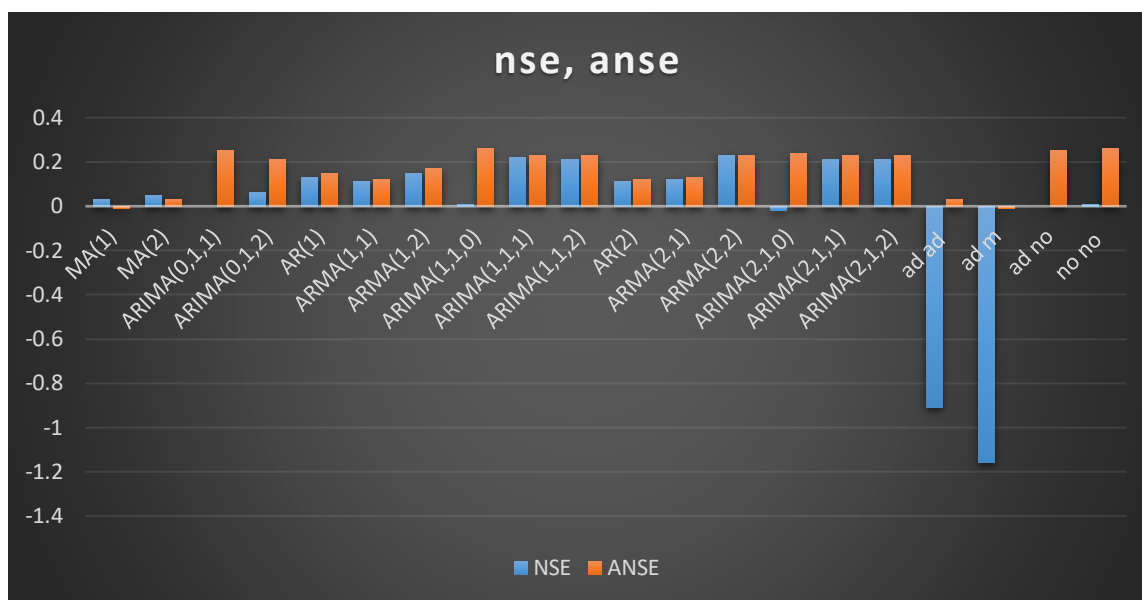
1. Ορίζοντας πρόγνωσης 10 ημερών:
 - a. Μοντέλα ARIMA
 - b. Μοντέλα εκθετικής ομαλοποίησης
2. Ορίζοντας πρόγνωσης 5 ημερών:
 - a. Μοντέλα ARIMA
 - b. Μοντέλα εκθετικής ομαλοποίησης
3. Ορίζοντας πρόγνωσης 1 ημέρας:
 - a. Μοντέλα ARIMA
 - b. Μοντέλα εκθετικής ομαλοποίησης
 - c. Μοντέλο deep learning rolling window
4. Μοντέλο deep learning προσομοίωσης λεκάνης

Τα αποτελέσματα θα παρουσιαστούν με τον τρόπο που αναφέρεται παραπάνω. Για τις 10 λεκάνες κατασκευάστηκαν διαγράμματα στηλών, με τους δείκτες Nash-Sutcliffe efficiency και Absolute Nash-Sutcliffe efficiency. Παρουσιάστηκαν όλα τα μοντέλα που έλαβαν μέρος σε κάθε ορίζοντα πρόγνωσης οι συμβολισμοί τους είναι οι ίδιοι με αυτούς που εξηγήθηκαν και στα προηγούμενα κεφάλαια για τις 10 λεκάνες.

5.1 Ορίζοντας πρόγνωσης 10 ημερών

Ο ορίζοντας πρόγνωσης των 10 ημερών σημαίνει πως η ανανέωση δεδομένων γίνεται κάθε 10 ημέρες. Το μοντέλο για την πρόγνωση των 10 ημερών προσδιορίζει επαναληπτικά τις διαδοχικές τιμές δηλαδή κάθε εκτιμώμενη τιμή απορροής καθίσταται τιμή εισαγωγής στο αρχείο δεδομένων, αυτό κάνει τους όρους NSE και ANSE να είναι πολύ δύσκολο να φτάσουν τιμές κοντά στη μονάδα (ιδανικό μοντέλο). Οι αποκλίσεις τους στα διαγράμματα είναι πολύ μεγάλες.

1) Id λεκάνης 02014000

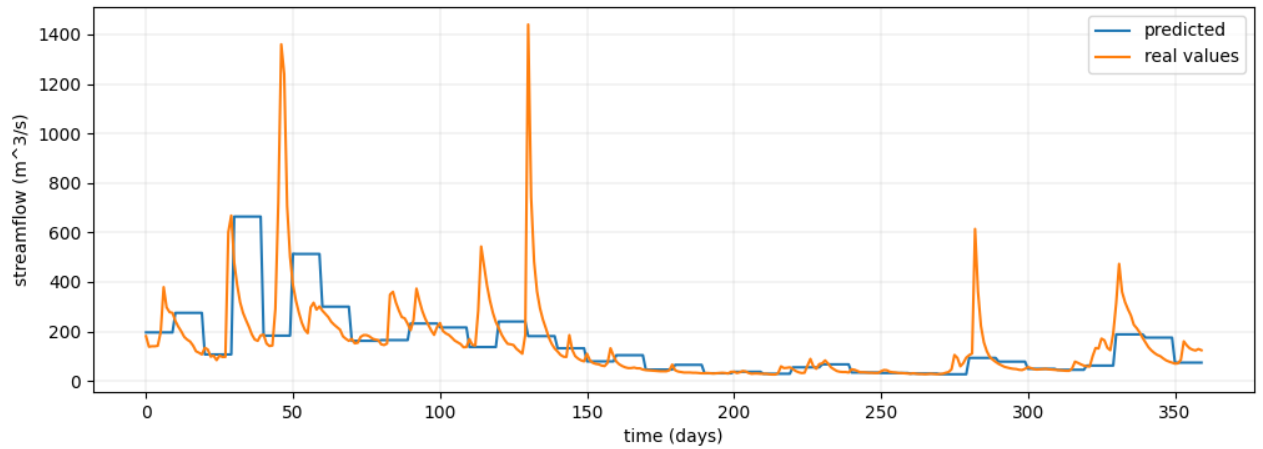


Εικόνα 54 Column chart, ορίζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων

Όπως έχει προαναφερθεί το ANSE είναι ένας δείκτης που δεν ‘τιμωρεί’ τις ακραίες καταστάσεις φαίνεται ότι αυτός είναι πιο αυξημένος στα περισσότερα μοντέλα. Τις ψηλότερες τιμές τις σημειώνει στα μοντέλα ARIMA(1,1,0), no no (απλή εκθετική ομαλοποίηση) με τιμή ANSE = 0.26. Αντίστοιχα μοντέλα όπως το ARIMA(1,1,1), ARIMA(2,1,1), ARIMA(2,1,2), ARMA(2,2) ενώ είναι χαμηλότερα στον δείκτη ANSE είναι αρκετά ψηλά στον δείκτη NSE και επομένως θεωρούνται καλύτερα μοντέλα εν γένει.

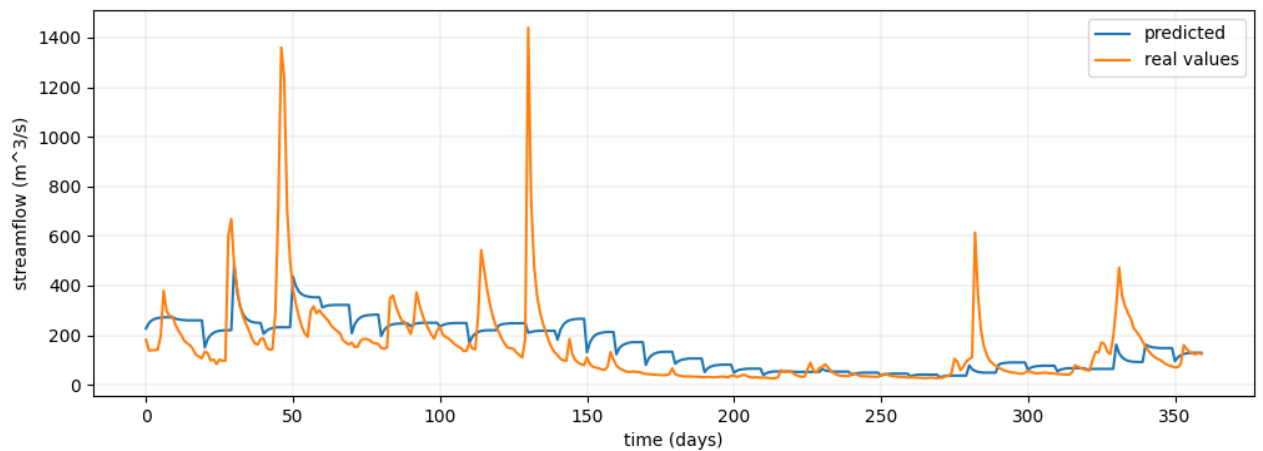
Παρακάτω παρατίθενται δύο διαγράμματα από τις αντίστοιχες ομάδες μοντέλων που σχολιάστηκαν για να φανεί η συμπεριφορά τους.

- ARIMA(1,1,0)



Εικόνα 55 Κοινό διάγραμμα μοντέλου ARIMA(1,1,0) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 10 ημερών, λεκάνης 1

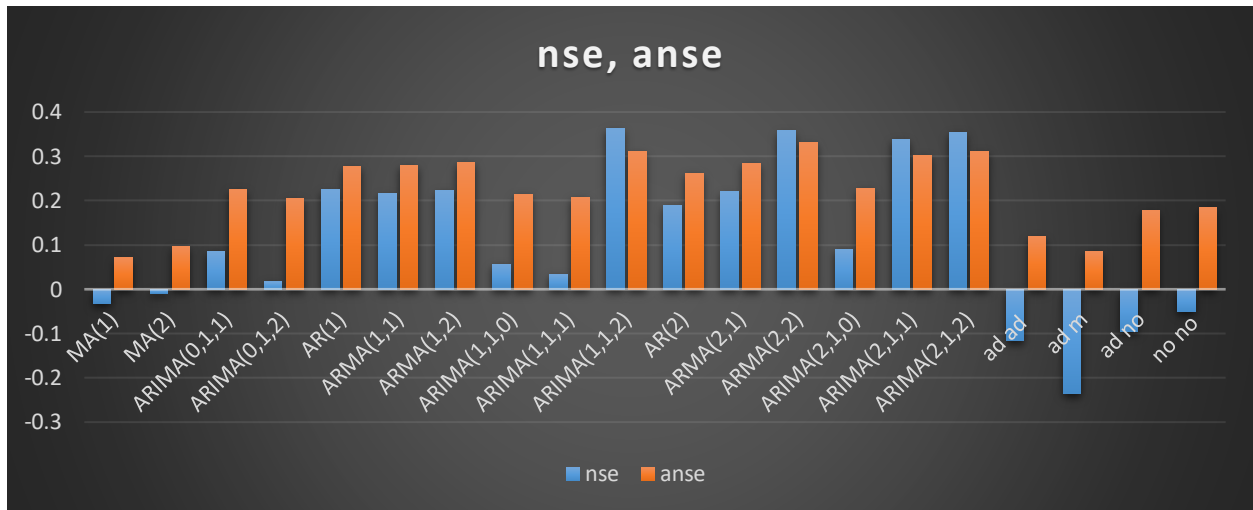
- ARIMA(2,1,2)



Εικόνα 56 Κοινό διάγραμμα μοντέλου ARIMA(2,1,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 10 ημερών λεκάνης 1

Όπως φαίνεται και τα δύο μοντέλα απέχουν πολύ από την πραγματική χρονοσειρά.

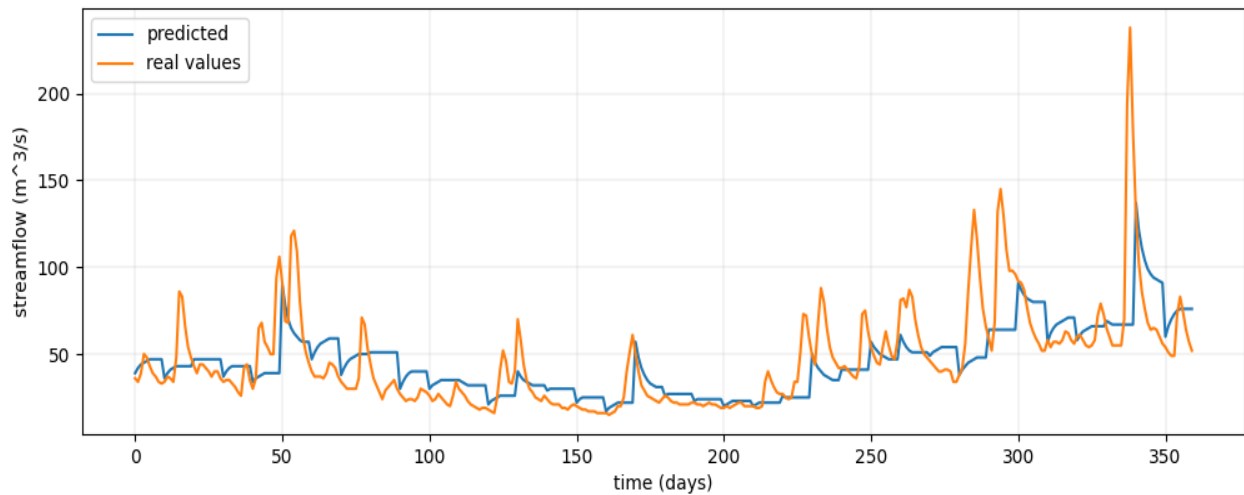
2) Id λεκάνης 01411300



Εικόνα 57 Column chart, ορίζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 2

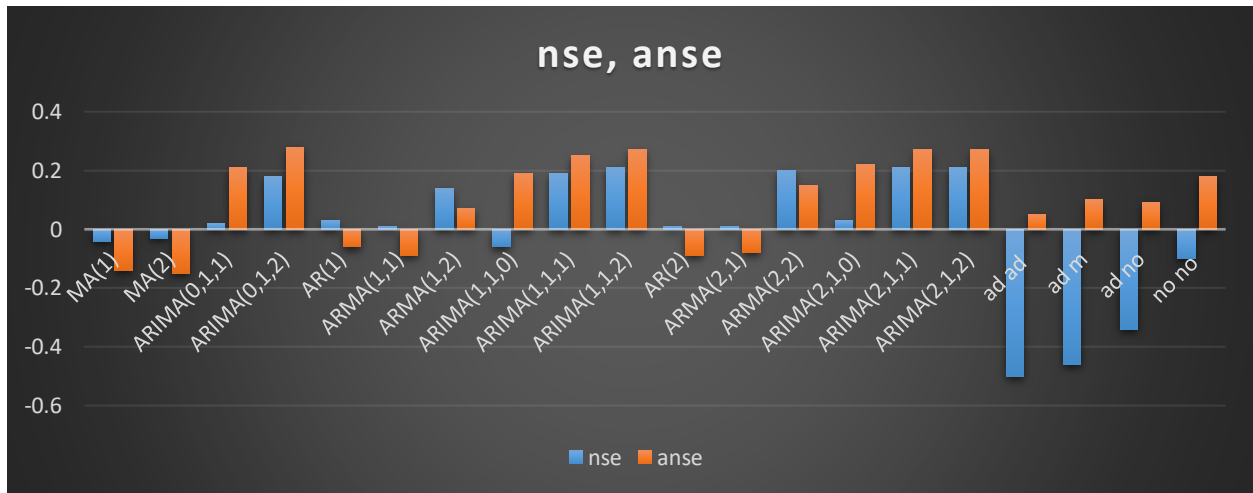
Τα μοντέλα εκθετικής ομαλοποίησης φαίνεται πως είναι αρκετά χειρότερα από τα μοντέλα ARIMA. Στη λεκάνη αυτή παρατηρούνται άλλα μοντέλα που έχουν επαυξημένο τον NSE όπως το ARIMA(1,1,2) και άλλα που έχουν τον ANSE όπως το ARMA(1,2).

Παρατίθεται η γραφική παράσταση του ARIMA(1,1,2) .



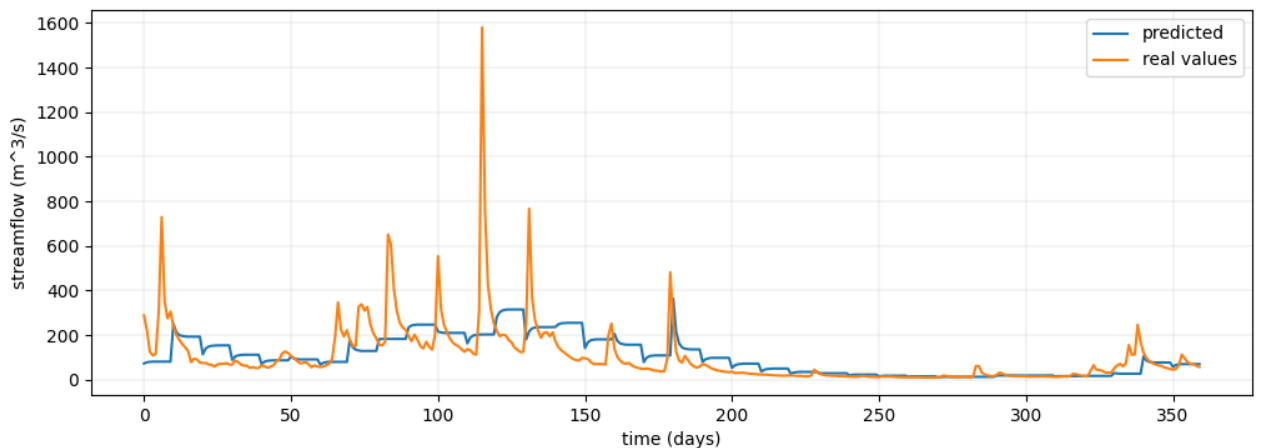
Εικόνα 58 Κοινό διάγραμμα μοντέλου ARIMA(1,1,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 10 ημερών λεκάνης 2

3) Id λεκάνης 01440000



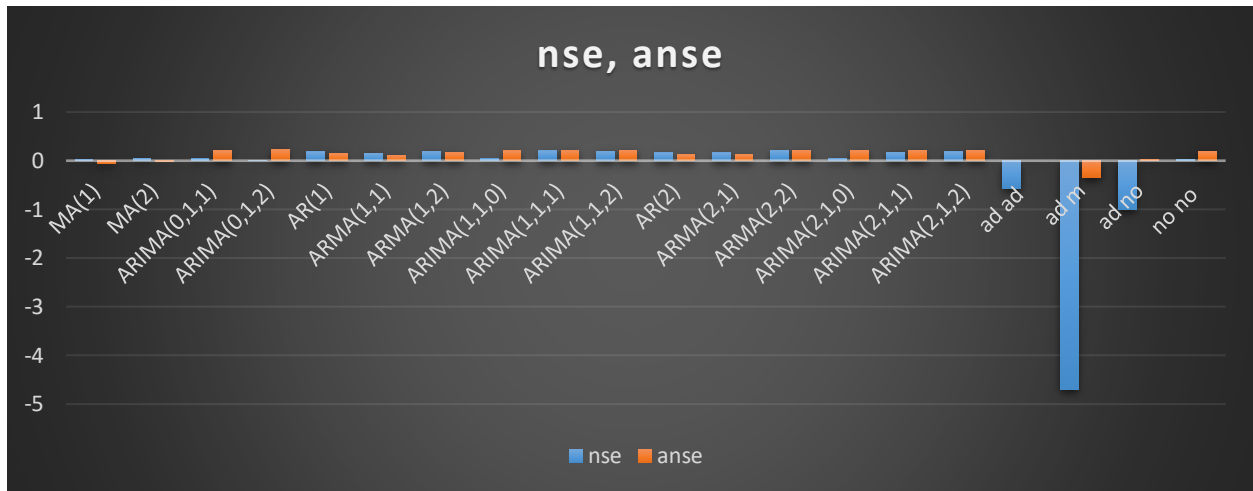
Εικόνα 59 Column chart, ορίζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 3

Όπως φαίνεται και από το διάγραμμα κανένα μοντέλο δεν ξεπερνά το 0.3 των NSE και ANSE. Τα μοντέλα ARIMA φαίνεται πως είναι εμφανώς ανώτερα από τα μοντέλα εκθετικής ομαλοποίησης. Τα καλύτερα σκορ τα συγκεντρώνουν τα μοντέλα ARIMA(1,1,2), ARIMA(2,1,1), ARIMA(2,1,2) με NSE = 0.21, ANSE=0.27. Παρατίθεται η γραφική παράσταση του ARIMA(1,1,2)



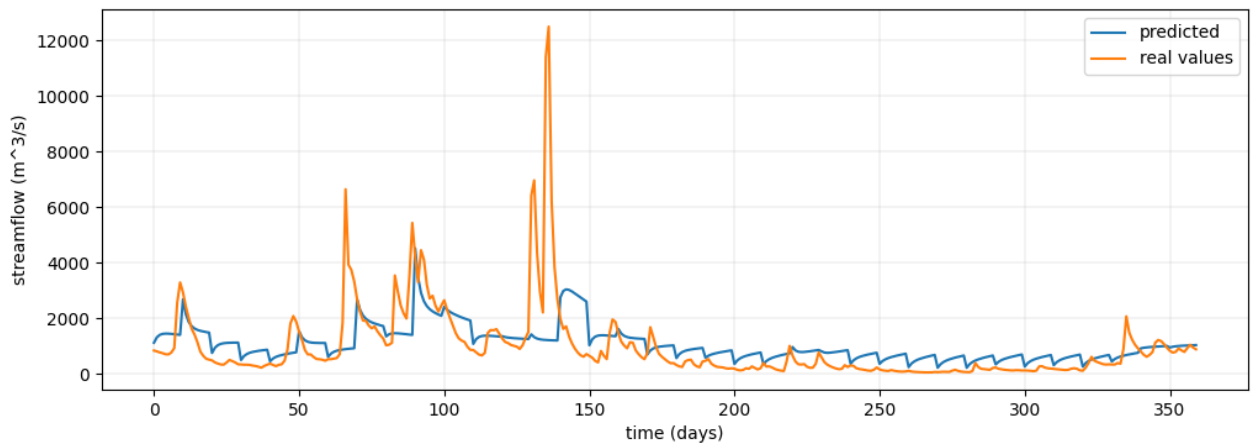
Εικόνα 60 Κοινό διάγραμμα μοντέλου ARIMA(1,1,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 10 ημερών λεκάνης 3

4) Id λεκάνης 01543500

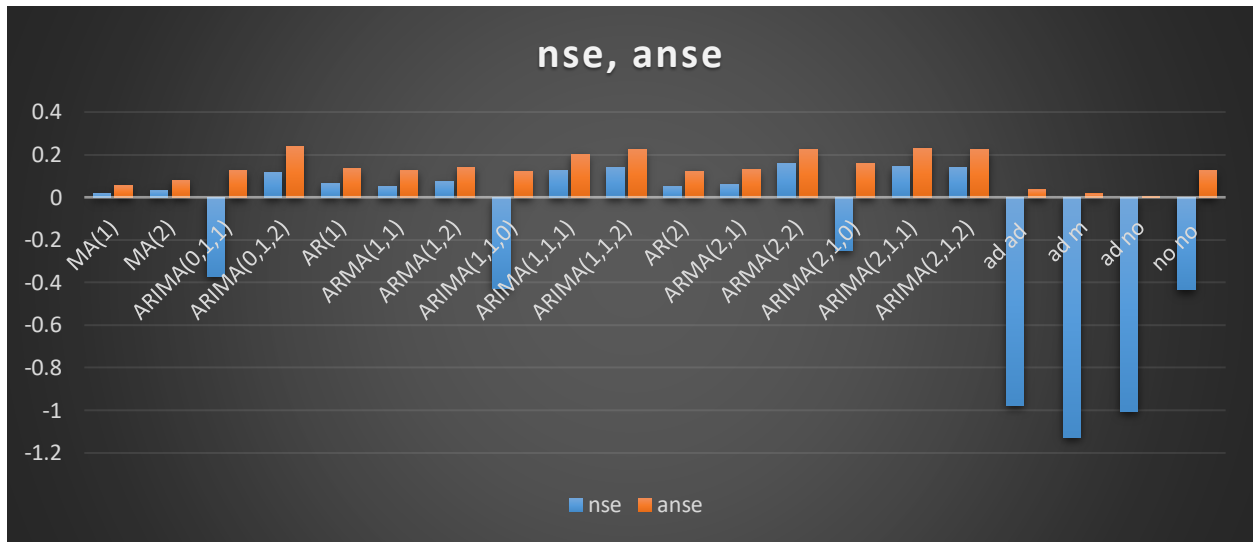


Εικόνα 61 Column chart, ορίζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 4

Αξίζει να σημειωθεί ότι από τα μοντέλα εκθετικής ομαλοποίησης το απλούστερο (no no) λειτουργεί καλύτερα από τα πιο σύνθετα. Γενικά κανένα μοντέλο δεν ξεπερνά το σκόρ 0.22 και δε θα παρουσιαστεί εδώ ενδεικτικό γράφημα. Στη συνέχεια παρατίθεται το γράφημα του ARMA(2,2)

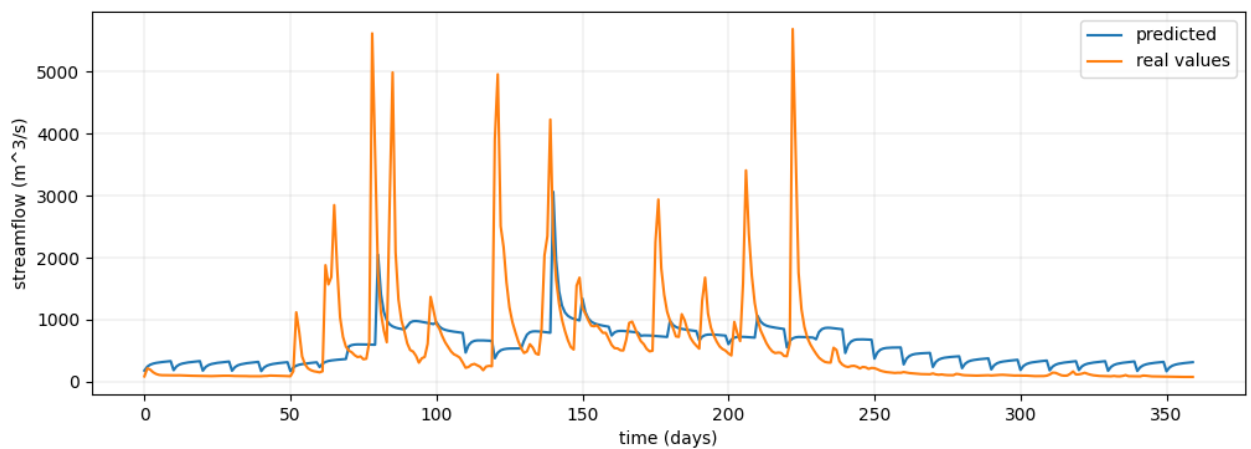


5) Id λεκάνης 02016000



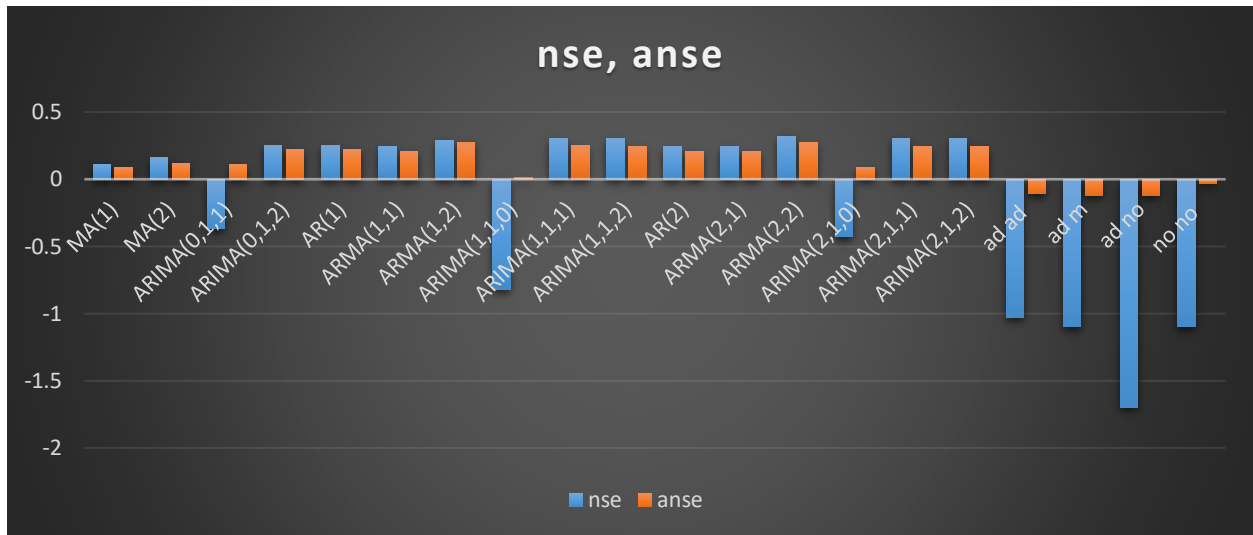
Εικόνα 62 Column chart, ορίζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 5

Ομοίως με την προηγούμενη λεκάνη τα μοντέλα ARIMA φαίνεται να πετυχαίνουν καλύτερα σκορ από τα μοντέλα εκθετικής ομαλοποίησης αλλά σε γενικές γραμμές πολύ χαμηλά που δεν ξεπερνούν το 0.23. Στη συνέχεια παρατίθεται το γράφημα του ARMA(2,2) που σημείωσε το ψηλότερο NSE = 0.16.



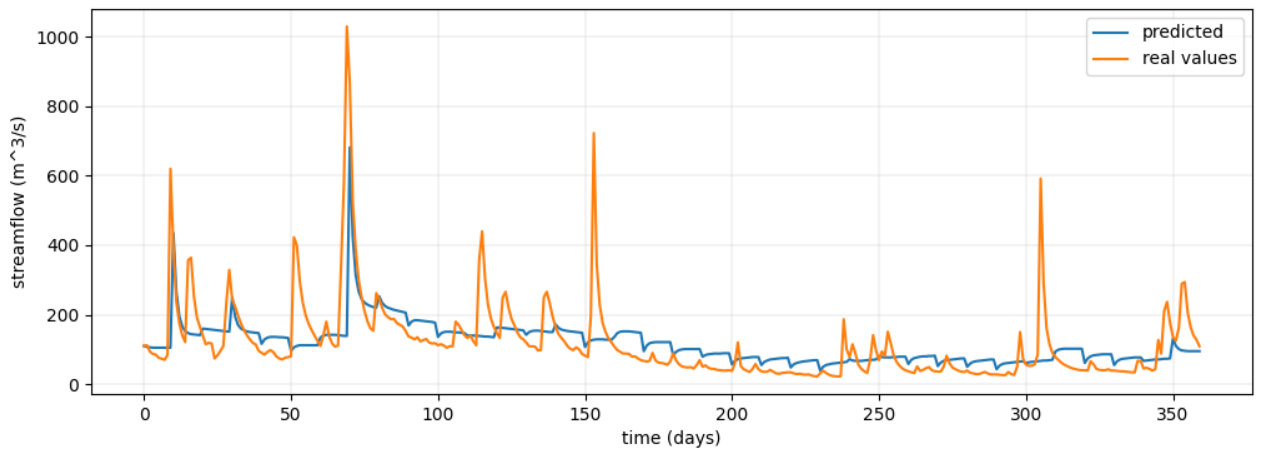
Εικόνα 63 Κοινό διάγραμμα μοντέλου ARMA(2,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 10 ημερών λεκάνης 5

6) Id λεκάνης 02053800



Εικόνα 64 Column chart, ορίζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 6

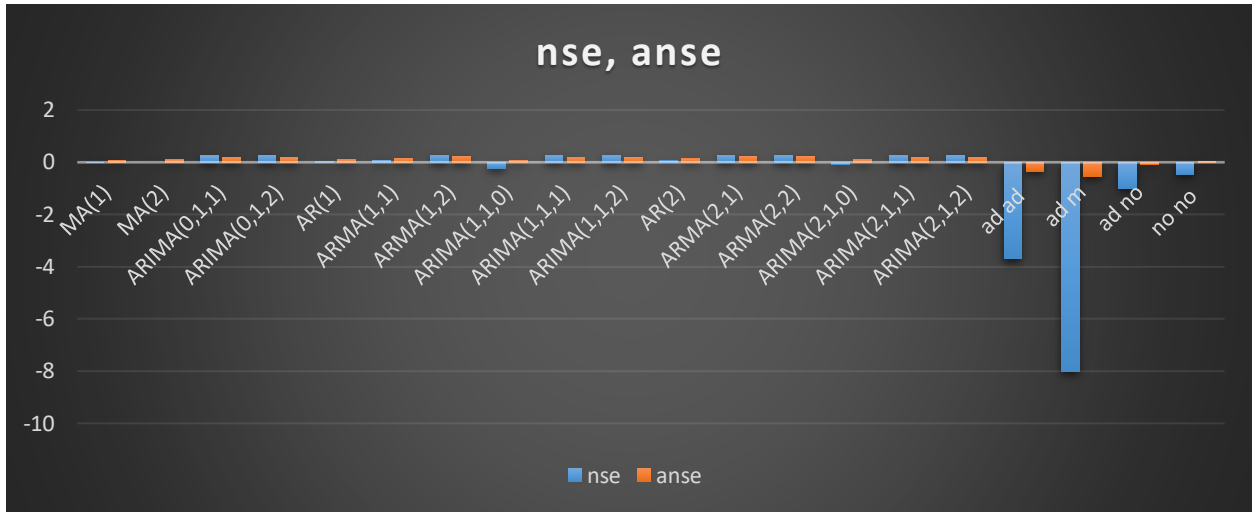
Στην παρούσα λεκάνη παρατηρείται καλύτερη απόδοση στα μοντέλα ARIMA σε σχέση με τις προηγούμενες και το μοντέλο που συγκεντρώνει το καλύτερο σκορ είναι το ARMA(2,2) με NSE = 0,32 και ANSE = 0.27. Στη συνέχεια παρατίθεται η γραφική του παράσταση.



Εικόνα 65 Κοινό διάγραμμα μοντέλου ARMA(2,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 10 ημερών λεκάνης 6

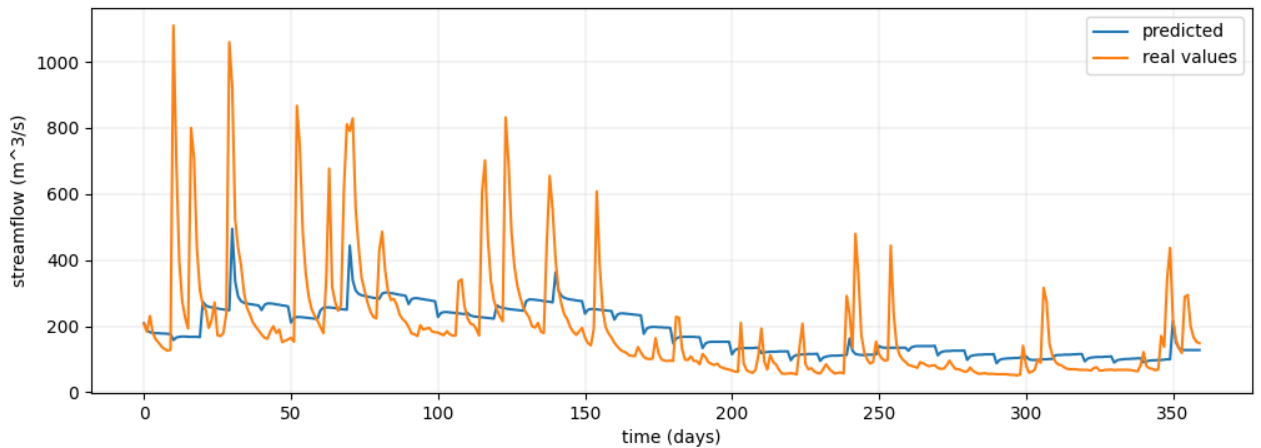
Είναι και η πρώτη φορά μέχρι στιγμής που μοντέλο μπορεί να εκτιμήσει σωστά ένα υψηλό ακρότατο (μεταξύ ημέρας 50 και 100).

7) Id λεκάνης 02059500



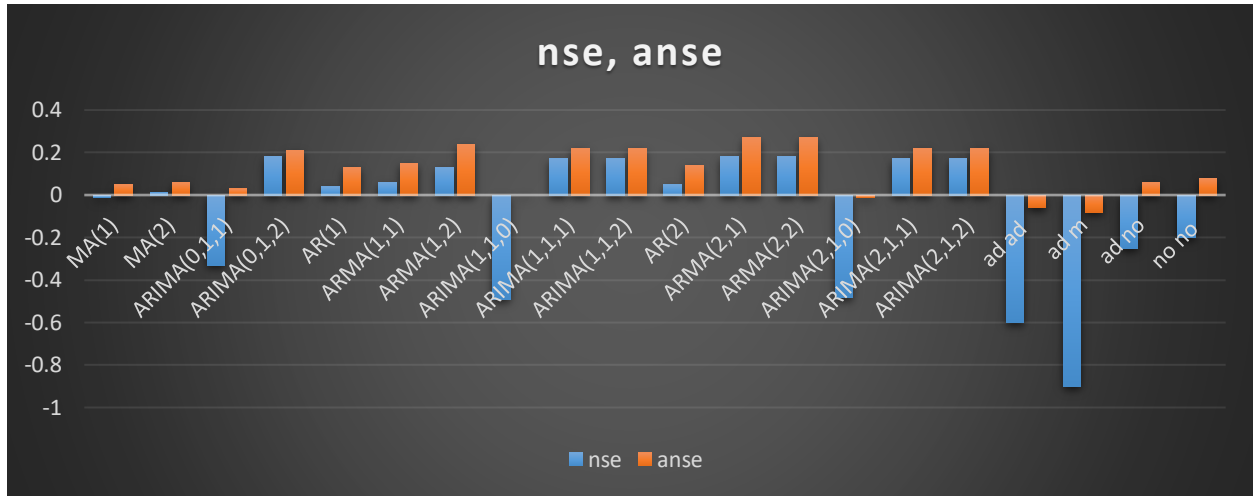
Εικόνα 66 Column chart, οριζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 7

Στην παρούσα περίπτωση τα μοντέλα εκθετικής ομαλοποίησης έχουν μεγάλες αρνητικές τιμές και επηρεάζουν την κλίμακα. Τα καλύτερα σκορ τα συγκεντρώνουν τα ARMA(1,2), ARMA(2,1), ARMA(2,2) με τιμές NSE = 0.27 και ANSE = 0.24. Στη συνέχεια παρατίθεται η γραφική παράσταση του ARMA(2,1).



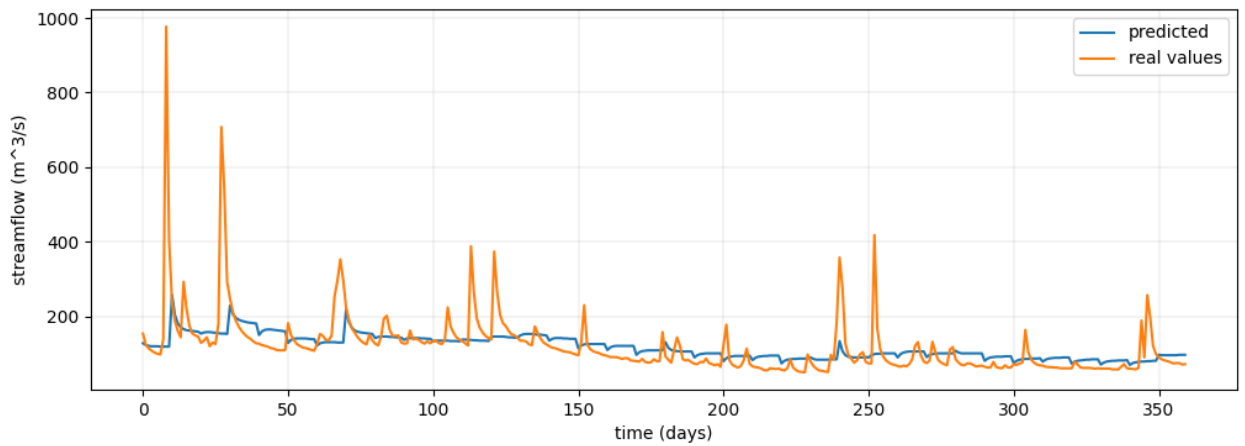
Εικόνα 67 Κοινό διάγραμμα μοντέλου ARMA(2,1) και πραγματικών τιμών απορροής για οριζοντα πρόγνωσης 10 ημερών λεκάνης 7

8) Id λεκάνης 02069700

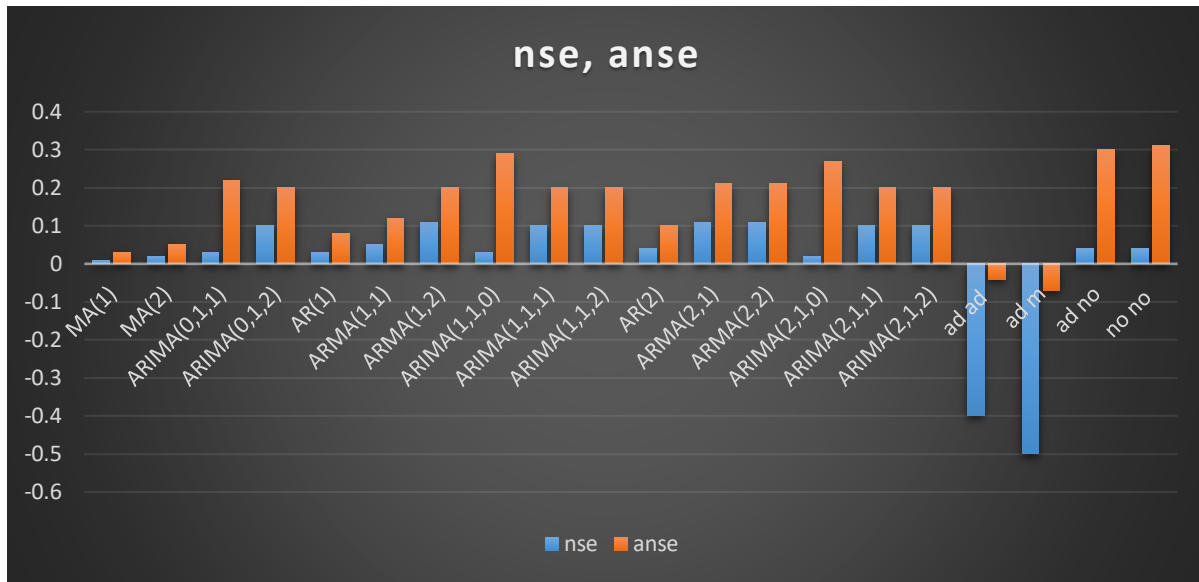


Εικόνα 68 Column chart, οριζόντια πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 8

Τα μοντέλα ούτε εδώ σημειώνουν κάποια σημαντική άνοδο στα σκορ με τα ARMA(1,2) και ARMA(2,2) να είναι τα καλύτερα με σκορ NSE=0.18 και ANSE=0.27. Στη συνέχεια παρατίθεται το ARMA(2,2).

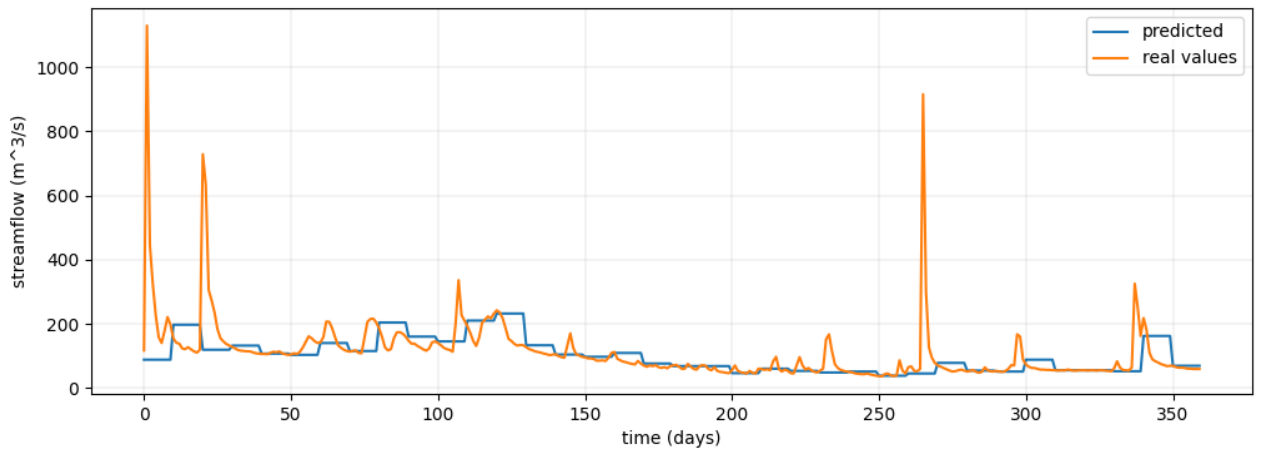


9) Id λεκάνης 02143000



Εικόνα 69 Column chart, ορίζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 9

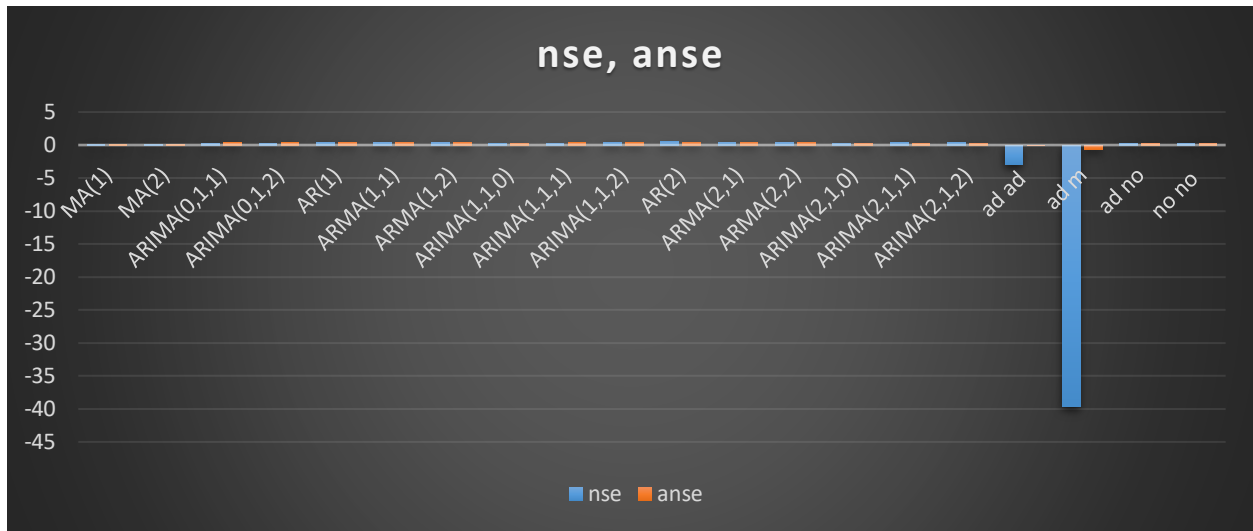
Η παρούσα λεκάνη αποτελεί την μοναδική περίπτωση στη συνθήκη του ορίζοντα πρόγνωσης 10 ημερών που ένα μοντέλο εκθετικής ομαλοποίησης πετυχαίνει καλύτερο σκορ σε σχέση με τα υπόλοιπα μοντέλα και αυτό είναι το ANSE=0.31. Στη συνέχεια παρατίθεται το διάγραμμα του.



Εικόνα 70 Κοινό διάγραμμα μοντέλου 'no no' και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 10 ημερών λεκάνης 9

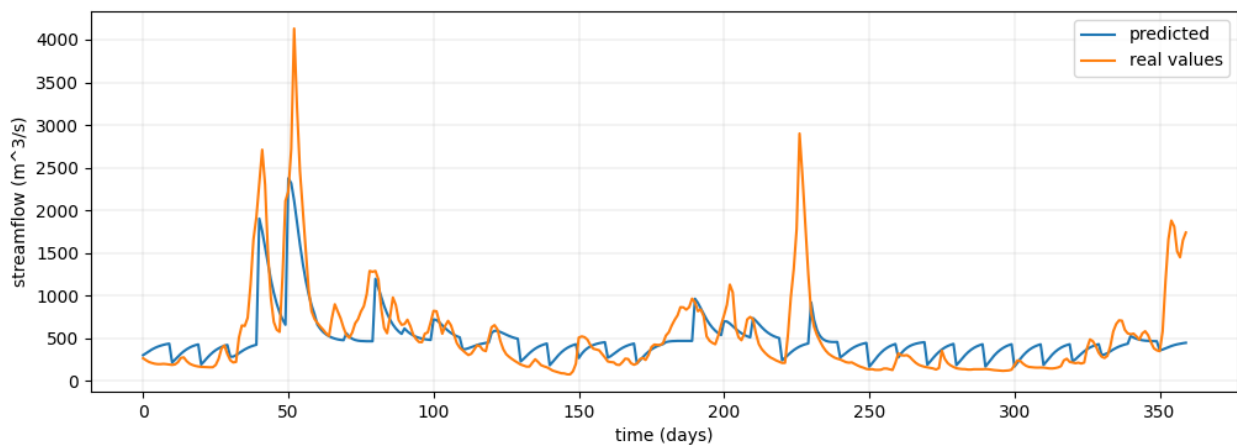
Όπως φαίνεται επειδή ο δείκτης ANSE δεν δίνει εξέχουσα βαρύτητα στις αποκλίσεις όταν οι τιμές είναι ακραίες το μοντέλο έχει καλό σκορ παρόλο που δεν εκτιμά σωστά καμία ακραία τιμή απορροής.

10) Id λεκάνης 02350900



Εικόνα 71 Column chart, ορίζοντας πρόγνωσης 10 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 10

Σε αυτή την περίπτωση παρατηρείται το μοντέλο AR(2) να πετυχαίνει αρκετά υψηλό σκορ σε σχέση με τις λεκάνες που παρουσιάστηκαν ανωτέρω. Γενικά τα περισσότερα μοντέλα ARIMA φαίνεται να πετυχαίνουν ικανοποιητικό σκορ δεδομένου του ορίζοντα πρόγνωσης 10 ημερών. Πιο συγκεκριμένα στο AR(2), σημειώθηκε για $NSE = 0.47$ και $ANSE = 0.35$. Στη συνέχεια παρατίθεται η γραφική αναπαράστασή του.



Εικόνα 72 Κοινό διάγραμμα μοντέλου AR(2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 10 ημερών λεκάνης 10

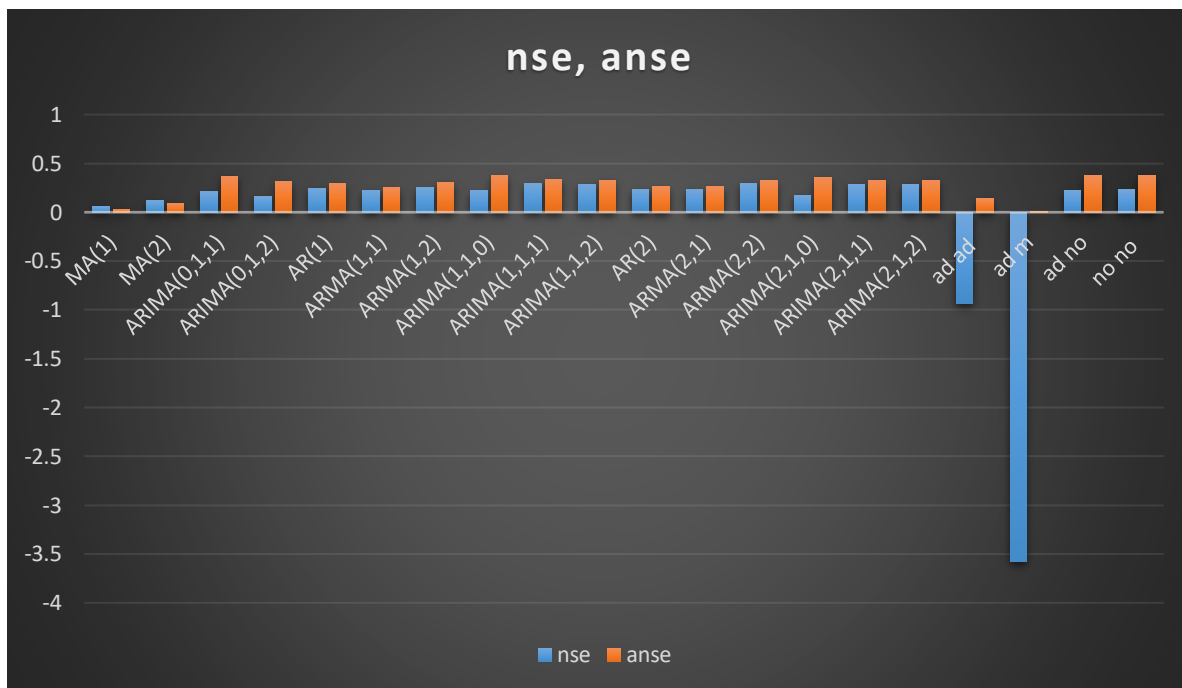
Το μοντέλο καταφέρει να εκτιμήσει αρκετά ικανοποιητικά τις περισσότερες υψηλές τιμές απορροής ενώ στις χαμηλές τιμές μπορεί και αναγνωρίζει το επίπεδο στο οποίο βρίσκονται με αποτέλεσμα να μην αστοχεί πολύ. Το γεγονός ότι ο δείκτης ANSE είναι μικρότερος από τον NSE υποδηλώνει πως η μεγαλύτερη αστοχία συγκεντρώνεται στις χαμηλότερες τιμές.

Συνοψίζοντας τα αποτελέσματα του ορίζοντα 10 ημερών παρατηρείται ότι τα μοντέλα ακόμη δεν μπορούν να προβλέψουν ικανοποιητικά τις χρονοσειρές απορροής. Έχουν σίγουρα δημιουργηθεί μοντέλα που πετυχαίνουν σκορ άνω του μηδενός και επομένως έχουν μια χρησιμότητα αλλά τα σκορ δεν είναι τόσο υψηλά ώστε τα μοντέλα να θεωρούνται αξιόπιστα. Αξίζει βέβαια να σημειωθεί πως η ανανέωση κάθε 10 ημέρες προσδίδει αρκετή δυσκολία καθώς τα μοντέλα καλούνται να προβλέψουν τα επόμενα 10 βήματα μόνα τους.

5.2 Ορίζοντας πρόγνωσης 5 ημερών

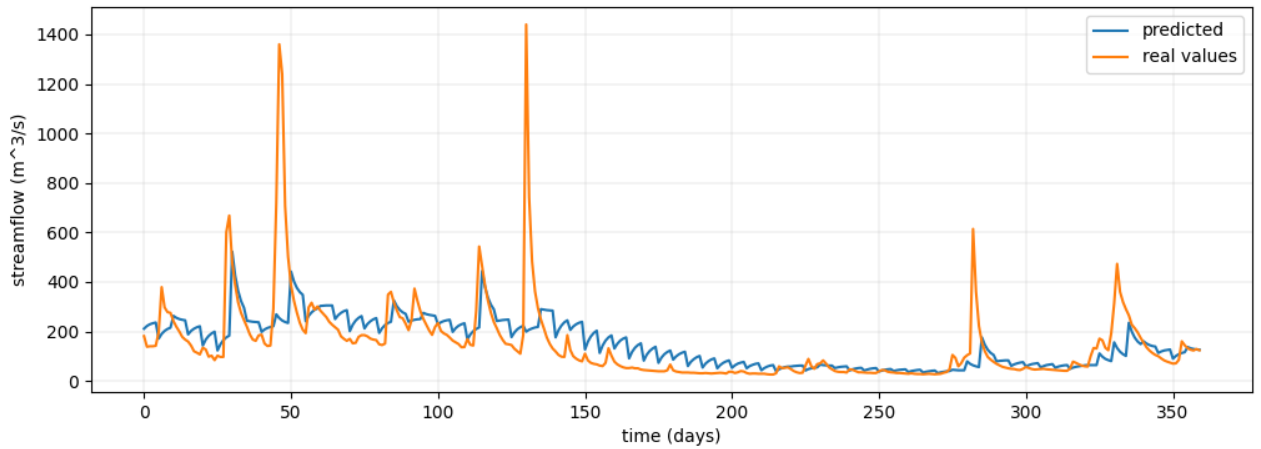
Στην περίπτωση αυτή τα μοντέλα ανανεώνονται κάθε πέντε ημέρες με τα πραγματικά δεδομένα, αυτό σημαίνει πως καλούνται να προβλέψουν μόνο τις επόμενες 5 ημέρες που και πάλι είναι ένα δύσκολο έργο αλλά αρκετά πιο βατό από τον ορίζοντα πρόγνωσης 10 ημερών. Θα παρουσιαστούν τα id των λεκανών και όπου κριθεί χρήσιμο και οι γραφικές παραστάσεις ορισμένων μοντέλων.

1) Id λεκάνης 02014000

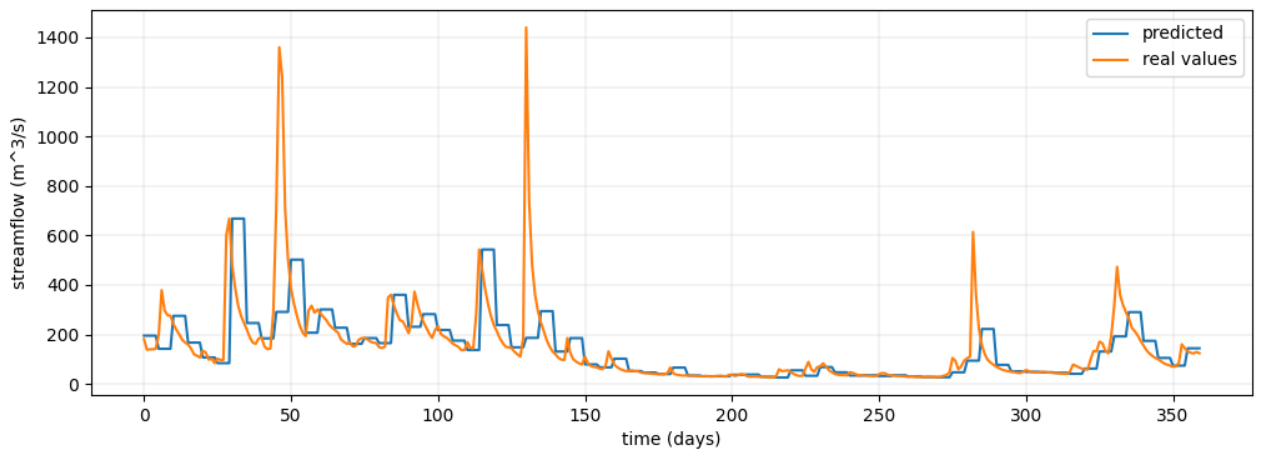


Εικόνα 73 Column chart, ορίζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 1

Στην παρούσα περίπτωση αν και το μοντέλο εκθετικής ομαλοποίησης ad m επηρεάζει την κλίμακα τα σκορ που πετυχαίνουν τα μοντέλα είναι αρκετά υψηλά με τιμές που κυμαίνονται κοντά στο 0.25 για NSE και 0.3 για ANSE. Το μοντέλο που έχει και τα δύο σκορ αρκετά υψηλά είναι το ARIMA(1,1,1) με NSE = 0.29 και ANSE = 0.34. Το μοντέλο που σημειώσει το υψηλότερο σκορ γενικά ήταν η απλή εκθετική ομαλοποίηση (no no) με NSE = 0.23 και ANSE = 0.38. Στη συνέχεια παρατίθενται οι γραφικές παραστάσεις ARIMA(1,1,1) και no no.



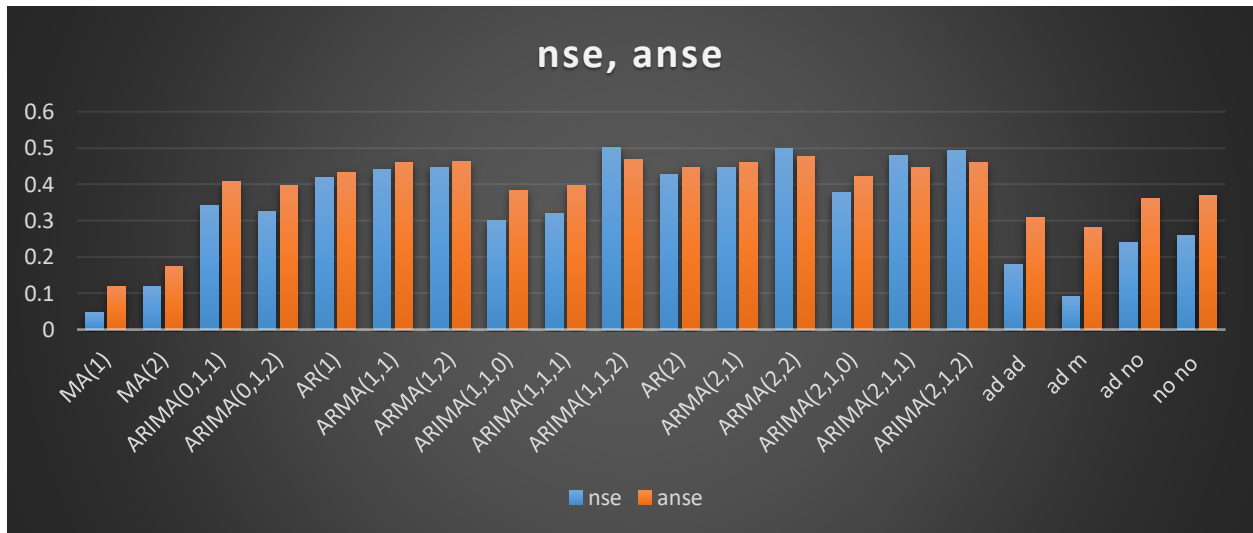
Εικόνα 74 Κοινό διάγραμμα μοντέλου ARIMA(1,1,1) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 1



Εικόνα 75 Κοινό διάγραμμα μοντέλου 'no no' και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 1

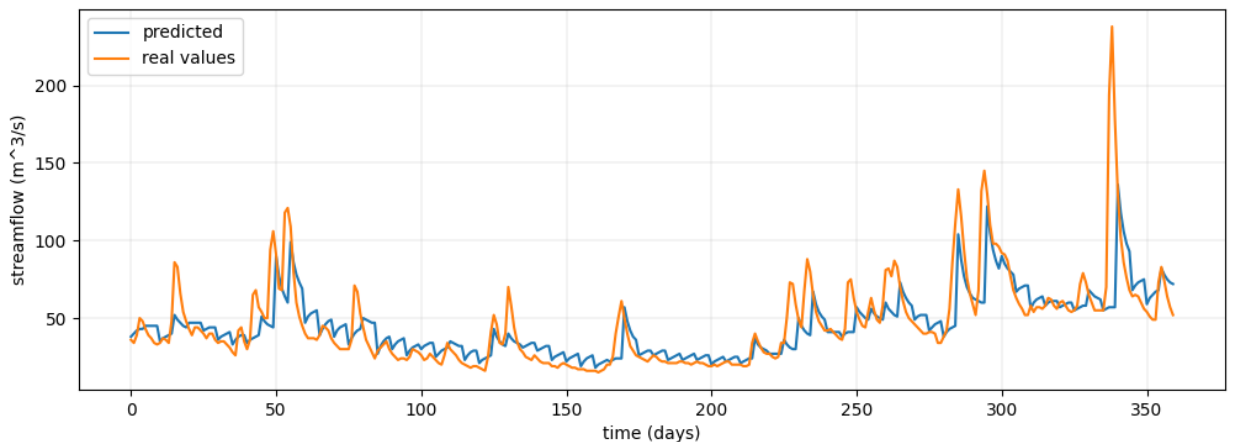
Από τη σύγκριση των δύο γραφικών παραστάσεων φαίνεται πως το μοντέλο ARIMA είναι πιο ικανοποιητικό στην εκτίμηση αλμάτων απορροής ενώ το μοντέλο εκθετικής ομαλοποίησης είναι καλύτερο στο να εκτιμά την απορροή όταν δεν σημειώνονται σημαντικές αιχμές. Σε φαινόμενα που η απορροή σημειώνει μεγάλη αύξηση σε μικρό χρονικό διάστημα (παραδείγματος χάρι ημέρες 50, 140) το μοντέλο εκθετικής ομαλοποίησης αστοχεί σε μεγάλο βαθμό.

2) Id λεκάνης 01411300



Εικόνα 76 Column chart, ορίζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 2

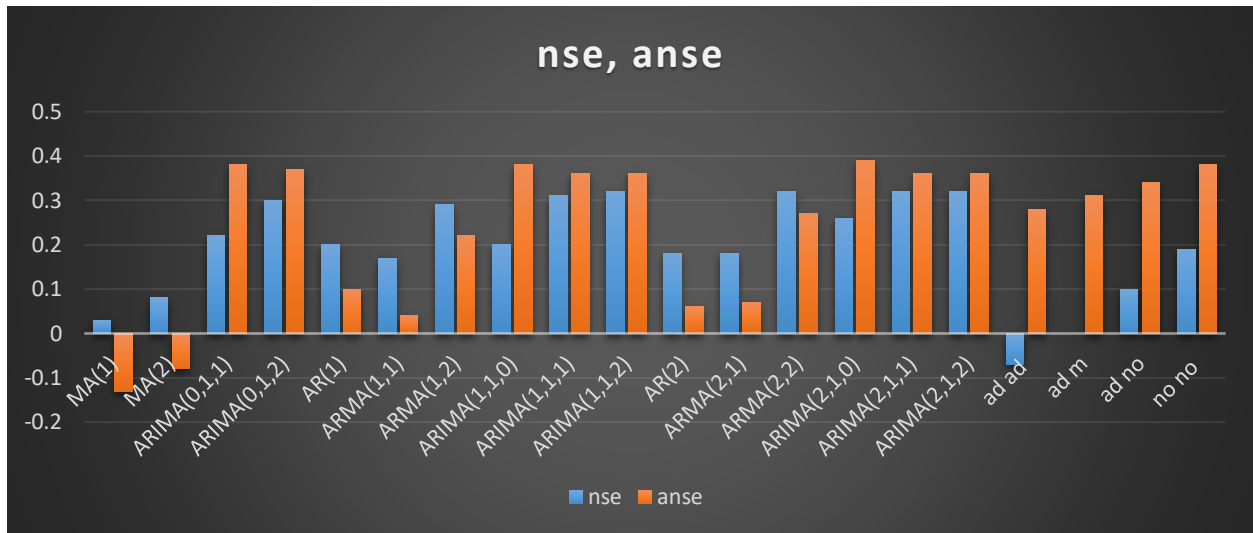
Το διάγραμμα στηλών της λεκάνης παρουσιάζει εμφανώς πολύ αυξημένα σκορ σε σχέση με όλα τα προηγούμενα σχεδόν σε όλα τα μοντέλα. Οι τιμές κυμαίνονται γύρω από το σκορ 0.45. Τα μοντέλα ARIMA με autoregressive όρο φαίνονται να λειτουργούν αρκετά καλύτερα από τα μοντέλα εκθετικής ομαλοποίησης. Καλύτερο μοντέλο και στα δύο σκορ είναι το ARMA(2,2) με NSE = 0.5 και ANSE = 0.48. Οι τιμές αυτές καθιστούν το μοντέλο αρκετά ικανοποιητικό που θα μπορούσε να χρησιμοποιηθεί στην ενεργά στην εκτίμηση απορροής. Στη συνέχεια παρατίθεται η γραφική παράσταση του ARMA(2,2).



Εικόνα 77 Κοινό διάγραμμα μοντέλου ARMA(2,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 2

Όπως φαίνεται και από τα διαγράμματα ανωτέρω το ARMA(2,2) προσεγγίζει αρκετά ικανοποιητικά την πραγματική χρονοσειρά καταφέροντας να εκτιμήσει και τις χαμηλές αλλά και τις υψηλές τιμές του.

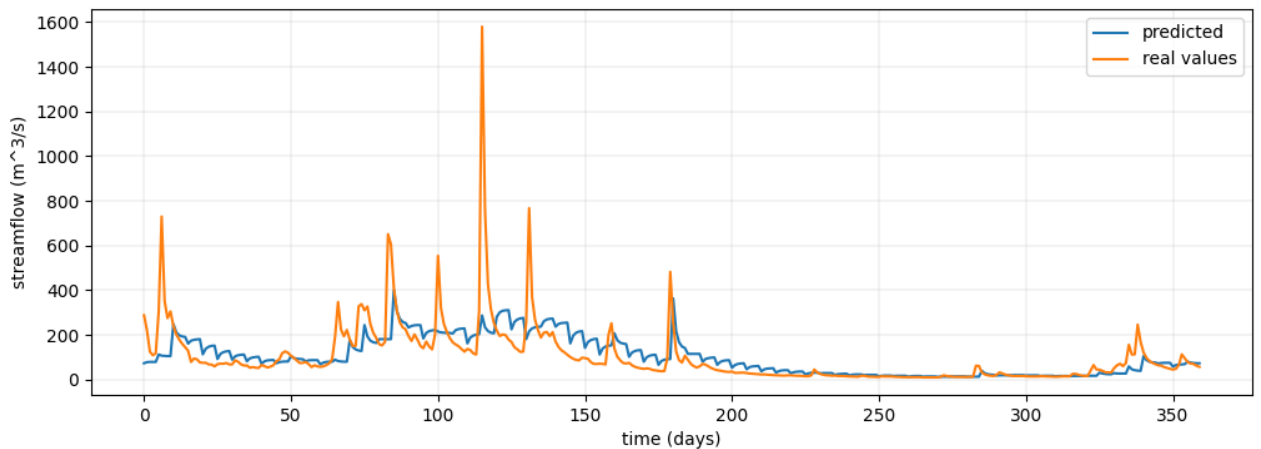
3) Id λεκάνης 01440000



Εικόνα 78 Column chart, ορίζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 3

Στην παρούσα λεκάνη παρατηρούνται αρκετά μοντέλα με σκορ πάνω από 0.3. Τα μοντέλα ARIMA παρουσιάζουν αρκετά κοντά τα σκορ NSE και ANSE, Αντιθέτως, τα NSE των μοντέλων εκθετικής ομαλοποίησης είναι πολύ κατώτερα γεγονός που τα κάνει μη ικανοποιητικά σε σύγκριση με τα υπόλοιπα μοντέλα. Συνδιαστικά καλά σκορ έχουν τα ARIMA(1,1,2), ARIMA(2,1,1), ARIMA(2,1,2). Με σκορ NSE = 0.32 και ANSE = 0.36.

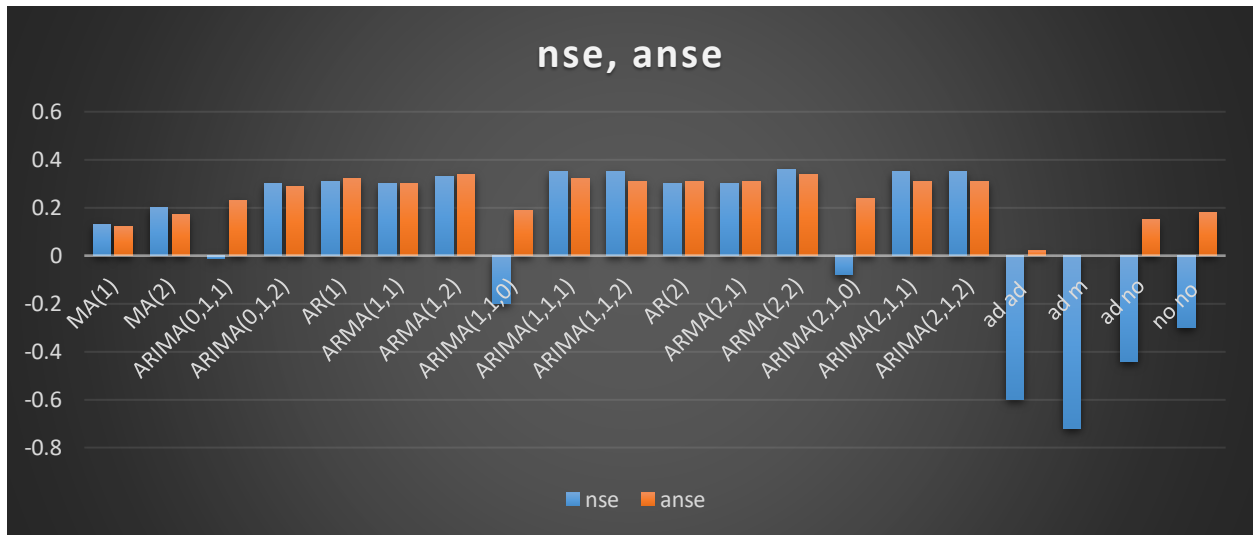
Στη συνέχεια παρατίθεται η γραφική παράσταση του ARIMA(2,1,2).



Εικόνα 79 Κοινό διάγραμμα μοντέλου ARIMA(2,1,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 3

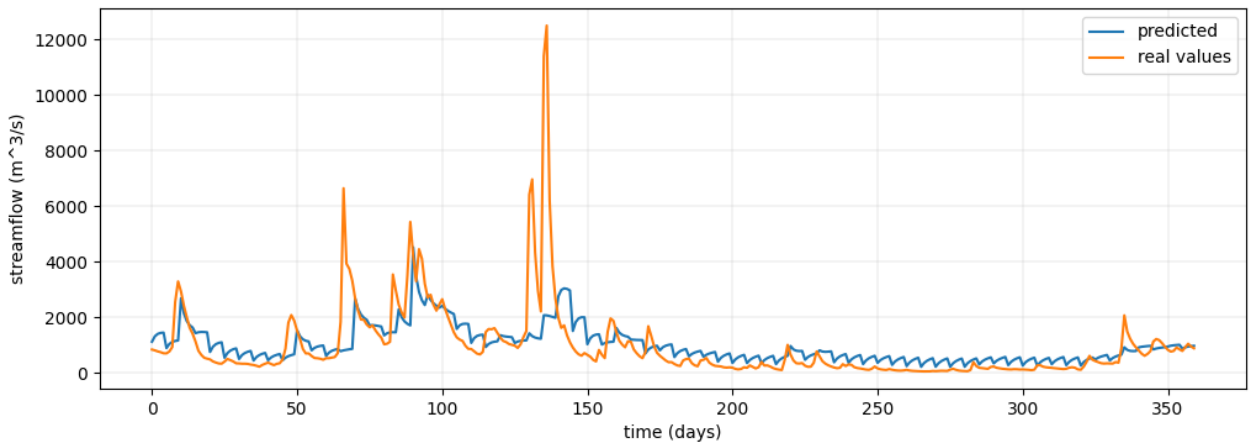
Φαίνεται πως το μοντέλο δεν εκτιμά στοχευμένα τις ακραίες τιμές απορροής παρουσιάζει μεν αύξηση απορροής αλλά μικρότερης κλίμακας από την πραγματική χρονοσειρά.

4) Id λεκάνης 01543500



Εικόνα 80 Column chart, οριζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 4

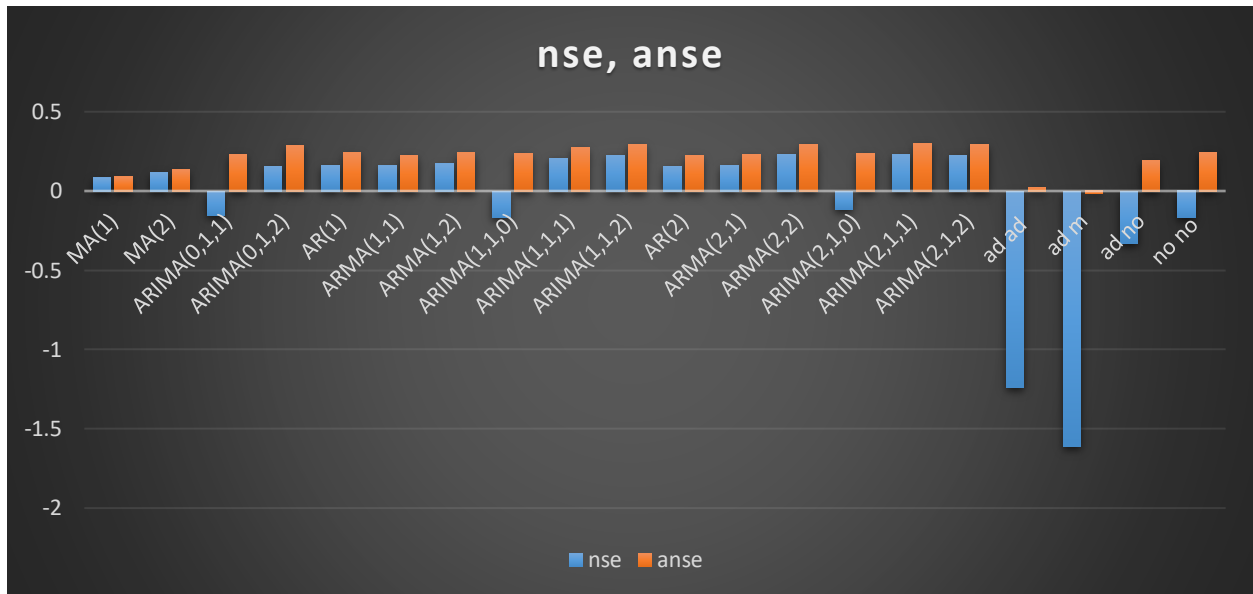
Στη λεκάνη αυτή τα μοντέλα ARIMA φαίνεται να πετυχαίνουν πολύ υψηλότερα σκορ από τα μοντέλα εκθετικής ομαλοποίησης . Υπάρχουν αρκετά μοντέλα με σκορ άνω του 0.3. Την καλύτερη εκτίμηση την σημειώνει το ARMA(2,2) με NSE = 0.36 και ANSE = 0.34 στη συνέχεια παρατίθεται το διάγραμμά του.



Εικόνα 81 Κοινό διάγραμμα μοντέλου ARMA(2,2) και πραγματικών τιμών απορροής για οριζοντα πρόγνωσης 5 ημερών λεκάνης 4

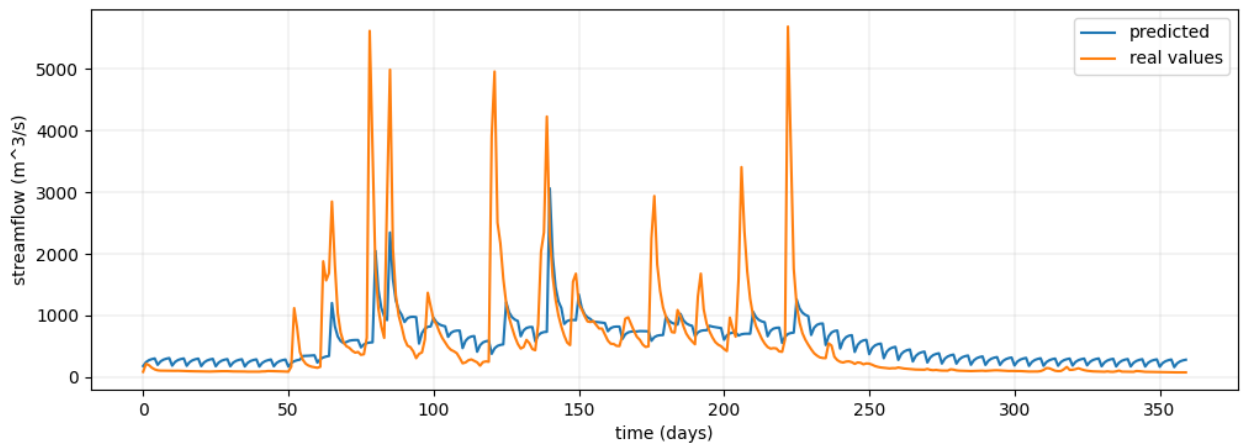
Στην ανωτέρω γραφική παράσταση φαίνεται το μοντέλο να προσαρμόζεται καλά στις χαμηλές τιμές απορροής και όσον αφορά τις υψηλές άλλες τιμές τις εκτιμά στοχευμένα όπως της ημέρας 90 και σε άλλες τιμές αστοχεί όπως στις ημέρες 130-135.

5) Id λεκάνης 02016000



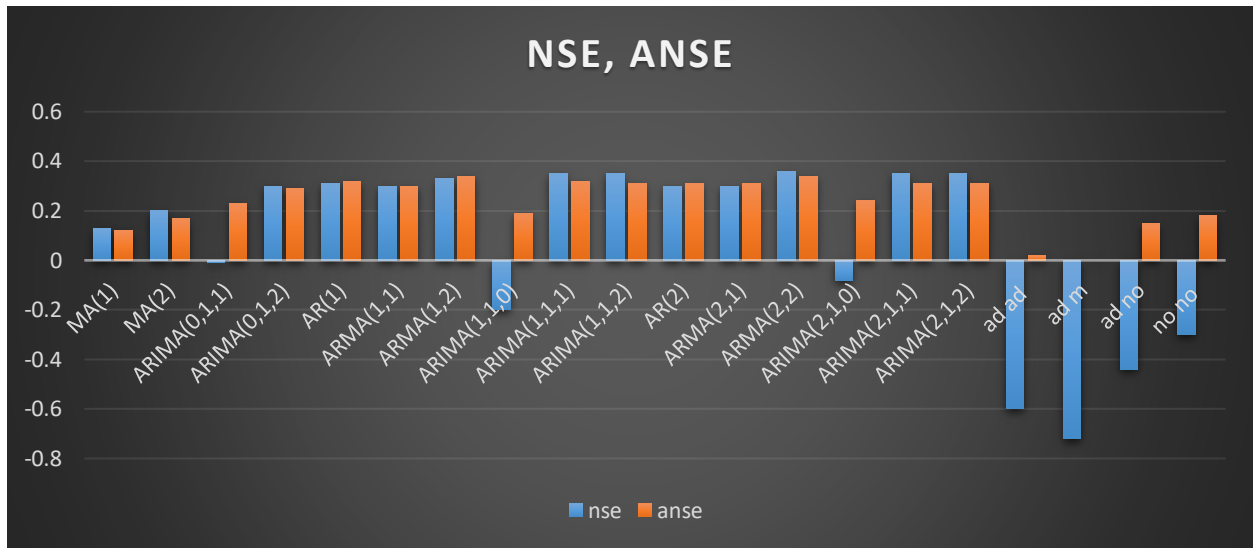
Εικόνα 82 Column chart, ορίζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 5

Στην παρούσα περίπτωση και πάλι τα μοντέλα εκθετικής ομαλοποίησης φαίνεται να υστερούν σημαντικά των ARIMA. Τα μοντέλα δεν παρουσιάζουν ιδιαίτερα υψηλά σκορ με καλύτερο να είναι το ARMA(2,2) με NSE = 0.23 και ANSE = 0.3. Στη συνέχεια παρατίθεται το γράφημα του ARMA(2,2).



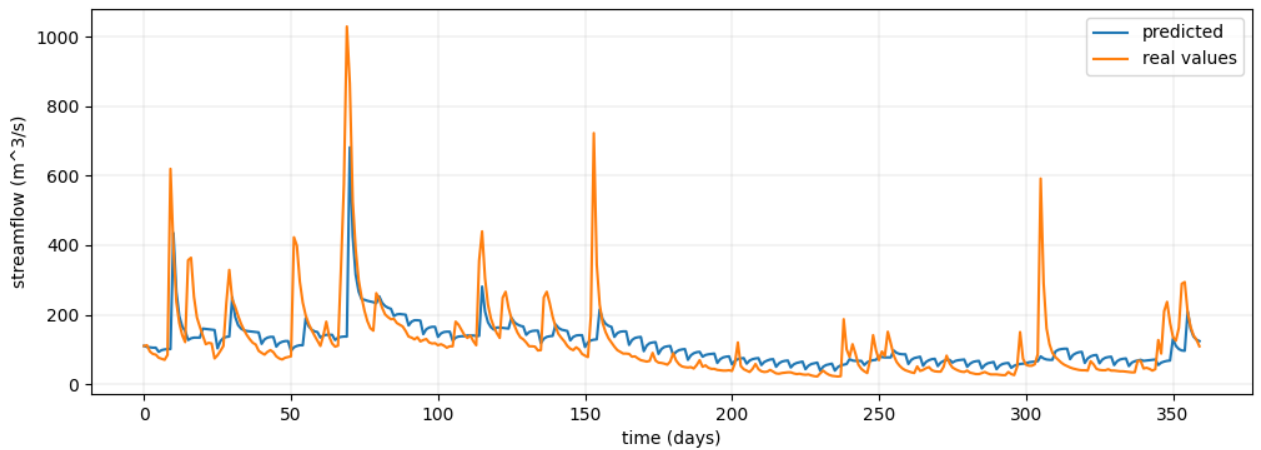
Εικόνα 83 Κοινό διάγραμμα μοντέλου ARIMA(2,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 5

6) Id λεκάνης 02053800



Εικόνα 84 Column chart, ορίζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 6

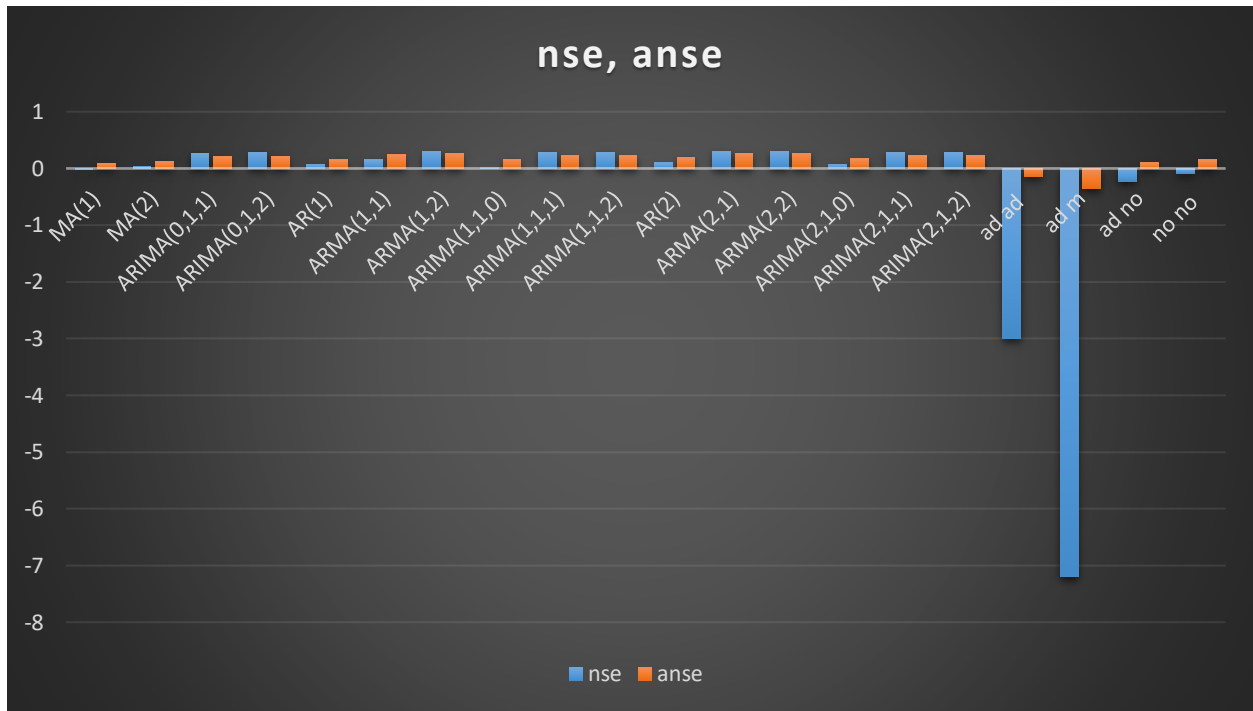
Εδώ παρουσιάζεται άλλη μια περίπτωση που τα μοντέλα εκθετικής ομαλοποίησης είναι εμφανώς κατώτερα σε απόδοση από τα μοντέλα ARIMA. Στα περισσότερα μοντέλα τα σκορ βρίσκονται πάνω από 0.3. Το καλύτερο σκορ και στους 2 δείκτες σημειώνεται από το ARMA(2,2) με NSE = 0.36 και ANSE = 0.34 στη συνέχεια παρατίθεται η γραφική παράσταση του μοντέλου.



Εικόνα 85 Κοινό διάγραμμα μοντέλου ARMA(2,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 6

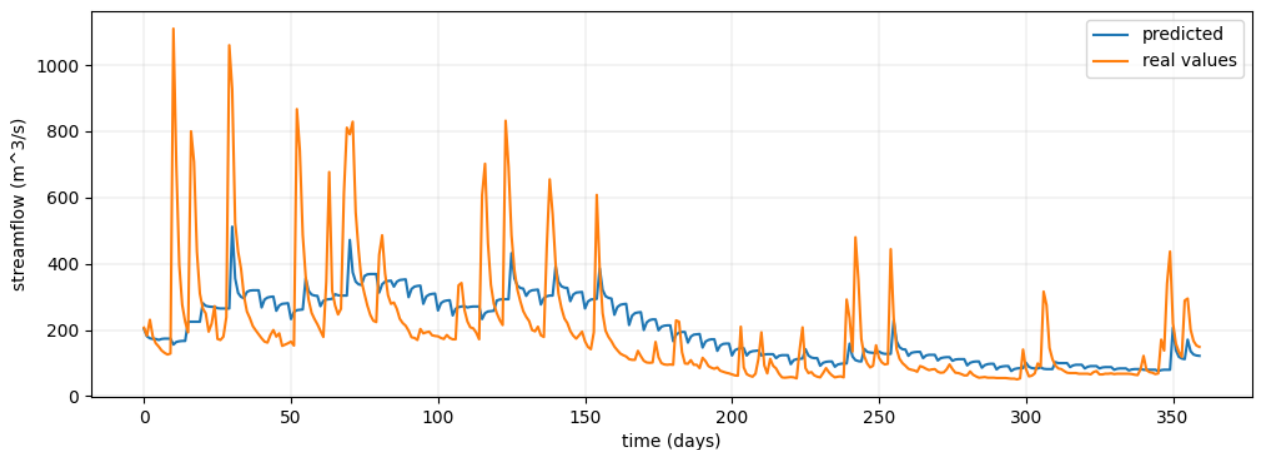
Όπως φαίνεται το μοντέλο εκτιμά αρκετά καλά κάποιες υψηλές τιμές απορροής ενώ υπάρχουν και τιμές που αστοχεί.

7) Id λεκάνης 02059500



Εικόνα 86 Column chart, ορίζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 7

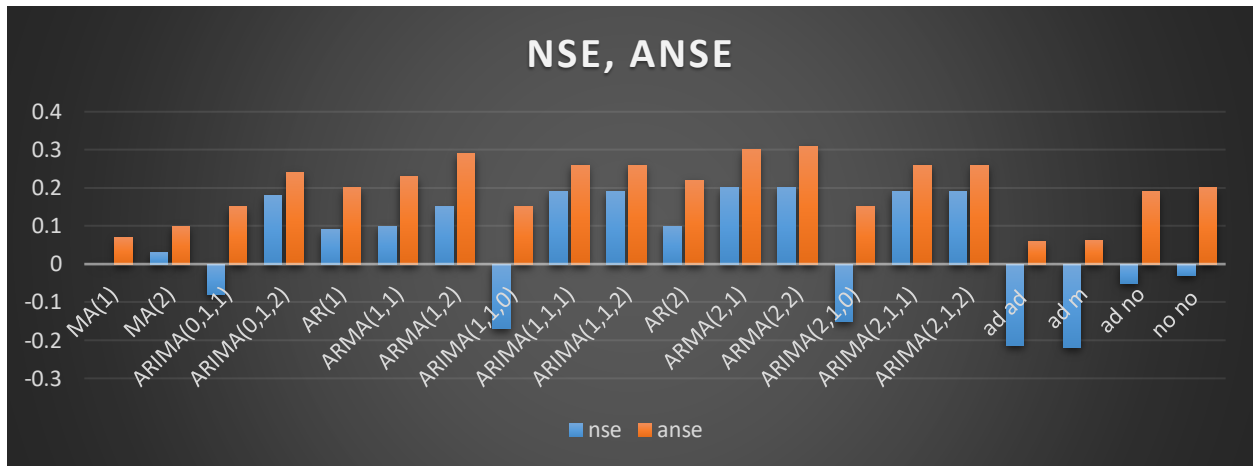
Σε αυτή τη λεκάνη τα μοντέλα δεν πετυχαίνουν καλά σκορ με το υψηλότερο να είναι των ARIMA(1,1,1), ARIMA(1,1,2), ARIMA(2,1,1), ARIMA(2,1,2) σημειώνοντας NSE = 0.29 και ANSE = 0.22. Στη συνέχεια παρατίθεται το γράφημα του ARIMA(1,1,1).



Εικόνα 87 Κοινό διάγραμμα μοντέλου ARIMA(1,1,1) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 7

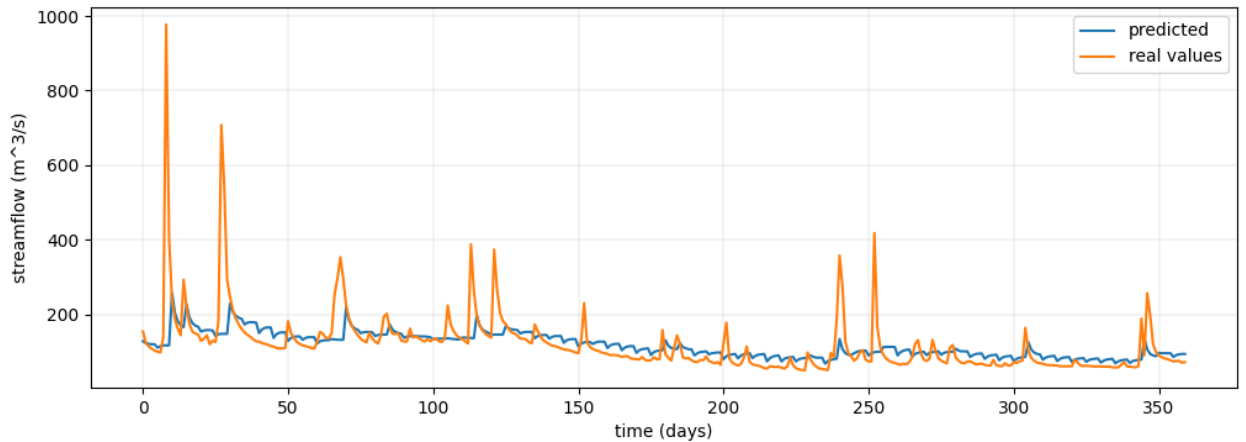
Παρατηρείται πως σε σχέση με τα προηγούμενα μοντέλα το συγκεκριμένο αδυνατεί να εκτιμήσει καλά και τις ακραίες τιμές αλλά και τις χαμηλότερες.

8) Id λεκάνης 02069700



Εικόνα 88 Column chart, ορίζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 8

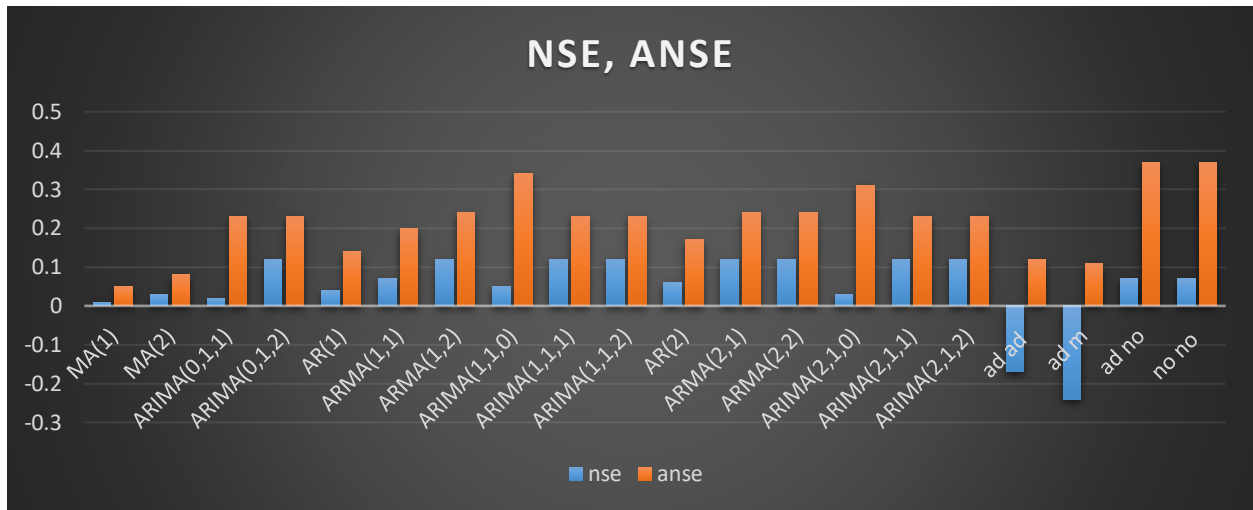
Στην παρούσα λεκάνη μόνο το μοντέλο ARMA(2,2) καταφέρνει σημειώσει στον δείκτη ANSE την τιμή 0.31. Ο δείκτης NSE είναι στο ίδιο επίπεδο σε αρκετά μοντέλα ARIMA. Τα μοντέλα εκθετικής ομαλοποίησης δεν παρουσιάζουν ικανοποιητικές τιμές δεικτών. Στη συνέχεια παρατίθεται η γραφική απεικόνιση του ARMA(2,2).



Εικόνα 89 Κοινό διάγραμμα μοντέλου ARMA(2,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 8

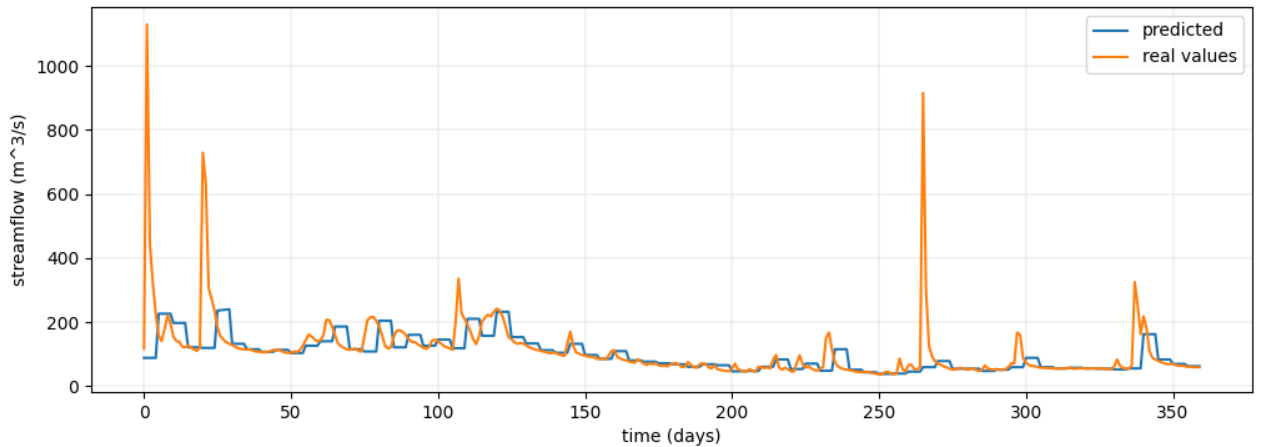
Το μοντέλο φαίνεται πως επιτυχημένα μπορεί να εκτιμήσει κάποιες ακραίες τιμές αλλά αστοχεί στις περισσότερες και για αυτό το σκορ του είναι αρκετά χαμηλό.

9) Id λεκάνης 02143000



Εικόνα 90 Column chart, ορίζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 9

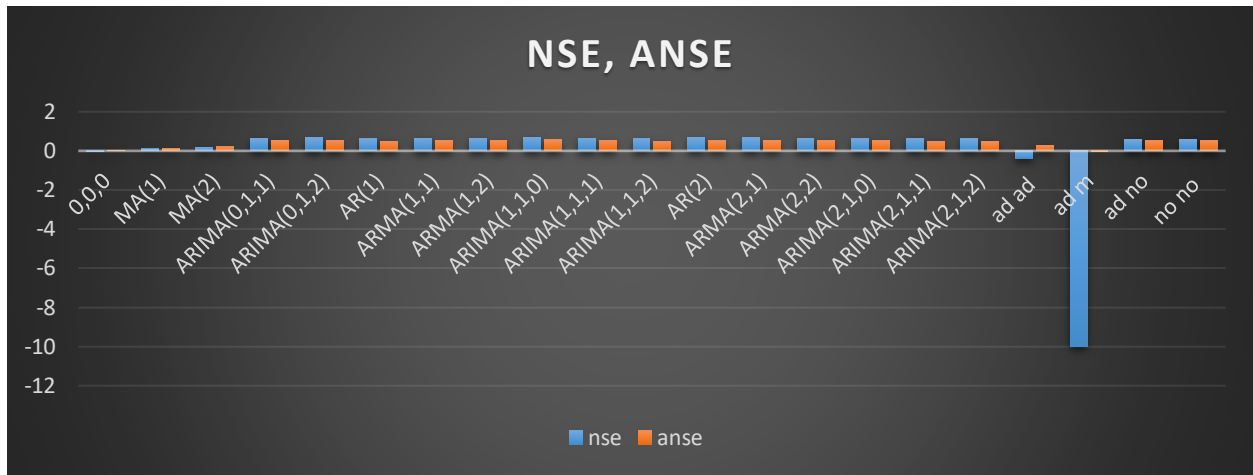
Στην παρούσα λεκάνη τα μοντέλα δεν πετυχαίνουν υψηλά σκορ στο NSE αλλά πετυχαίνουν αρκετά ικανοποιητικά σκορ στο ANSE. Σημειώνεται πως το υψηλότερο ANSE σημειώνεται από τα μοντέλα εκθετικής ομαλοποίησης (ad no, no no) με ANSE = 0.37 στη συνέχεια παρατίθεται το διάγραμμα εκθετικής ομαλοποίησης με πρόσθεση στον όρο τάσης(ad no).



Εικόνα 91 Κοινό διάγραμμα μοντέλου 'ad no' και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 5 ημερών λεκάνης 9

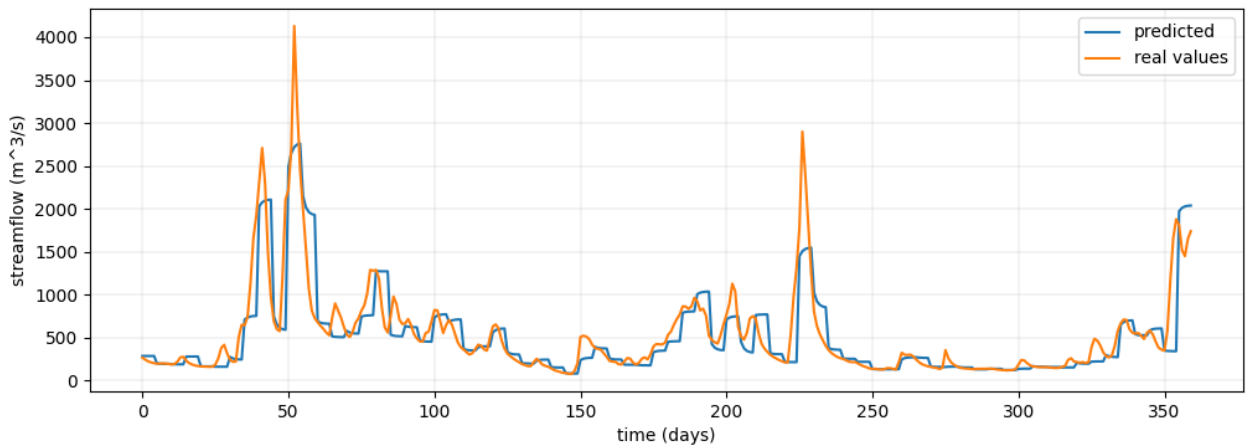
Όπως φαίνεται το μοντέλο αστοχεί πλήρως στην εκτίμηση ακραίων τιμών απορροής και δεν θεωρείται ικανοποιητικό.

10) Id λεκάνης 02350900



Εικόνα 92 Column chart, οριζοντας πρόγνωσης 5 ημερών, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 10

Σε αυτή τη λεκάνη τα μοντέλα ARIMA σημειώνουν πολύ καλά σκορ όπως και το μοντέλο απλής εκθετικής ομαλοποίησης. Οι τιμές των δεικτών κυμαίνονται πάνω από το 0.6. Καλύτερη επίδοση σημειώνει το ARIMA(1,1,0) με NSE = 0.70 και ANSE = 0.57. Στη συνέχεια παρατίθεται η γραφική παράσταση του μοντέλου.



Εικόνα 93 Κοινό διάγραμμα μοντέλου ARIMA(1,1,0) και πραγματικών τιμών απορροής για οριζοντα πρόγνωσης 5 ημερών λεκάνης 10

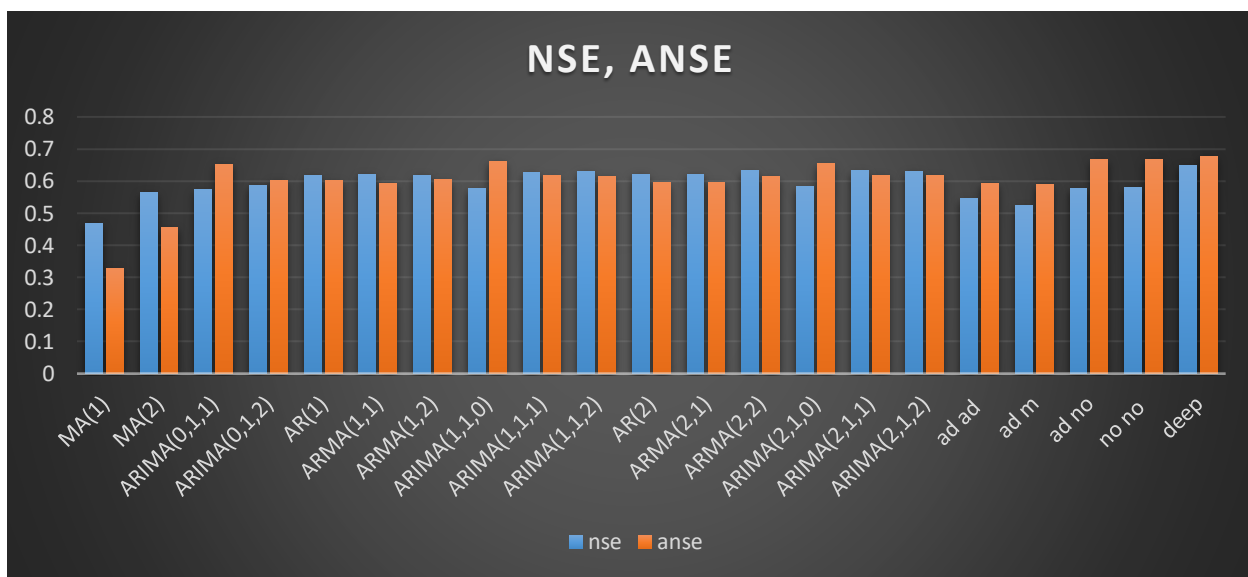
Όπως φαίνεται και από τα σκορ αλλά και από το γράφημα το μοντέλο εκτιμά αρκετά καλά τις περισσότερες ακραίες τιμές και ακολουθεί την απορροή όταν έχει χαμηλές τιμές. Είναι ένα αρκετά ικανοποιητικό μοντέλο και με NSE άνω του 0.7 είναι δεδομένο πως μπορεί να ληφθεί υπόψη στην εκτίμηση απορροής έχοντας υπόψη πως πρόκειται για εκτιμήσεις με οριζοντα πρόγνωσης 5 ημερών.

Συμπερασματικά, στον ορίζοντα πρόγνωσης 5 ημερών σε κάποιες λεκάνες σημειώνονται σημαντικές διαφορές με άνοδο της τιμής του NSE 0.28 μονάδες (λεκάνη με id 02350900) ενώ σε άλλες οι τιμές NSE δεν καταφέρνουν να ξεπεράσουν το 0.3 καθιστώντας το μοντέλο μη ικανοποιητικό. Παρουσιάστηκαν μοντέλα με δείκτες άνω του 0.5 που θα μπορούσαν να συμμετάσχουν στην πρόγνωση απορροής των αντίστοιχων λεκανών.

5.3 Ορίζοντας πρόγνωσης 1 ημέρας

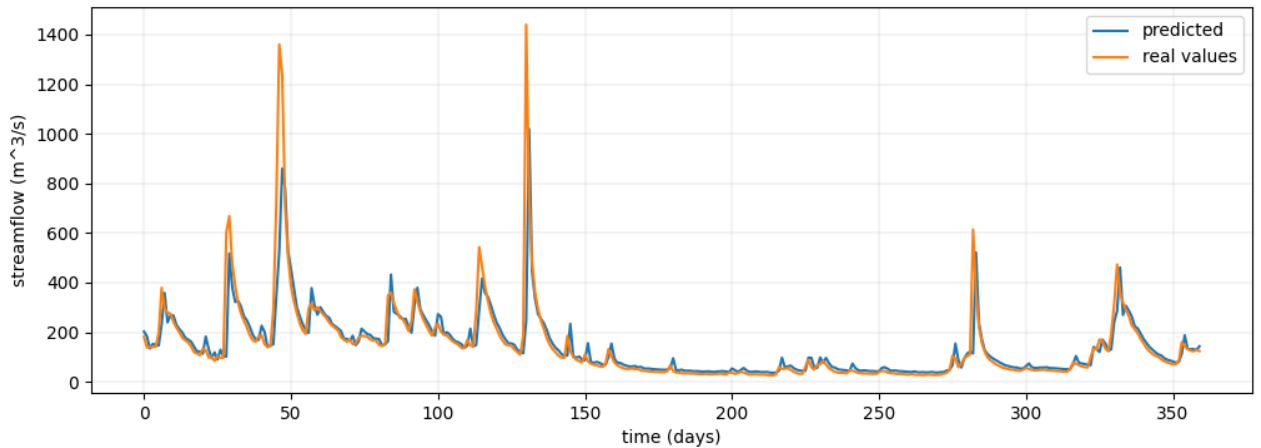
Ο ορίζοντας πρόγνωσης μιας ημέρας υποδεικνύει ότι το μοντέλο κάθε φορά εκτιμά μόνο την αμέσως επόμενη ημέρα και ύστερα ανανεώνεται με το πραγματικό δεδομένο απορροής. Με αυτό τον τρόπο δεν εκτιμά βασιζόμενο σε δεδομένα εισόδου που έχει το ίδιο εκτιμήσει αντί αυτού εκπαιδευτεί στα πραγματικά δεδομένα κάθε φορά. Αξίζει να σημειωθεί ότι πλέον στις διαδικασίες προστίθεται και η βαθιά μηχανική μάθηση με τη μέθοδο του κυλιόμενου παραθύρου. Είναι επίσης σημαντικό να αναφερθεί πως το όριο για να θεωρηθεί το NSE αρκετά επιτυχημένο

1) Id λεκάνης 02014000

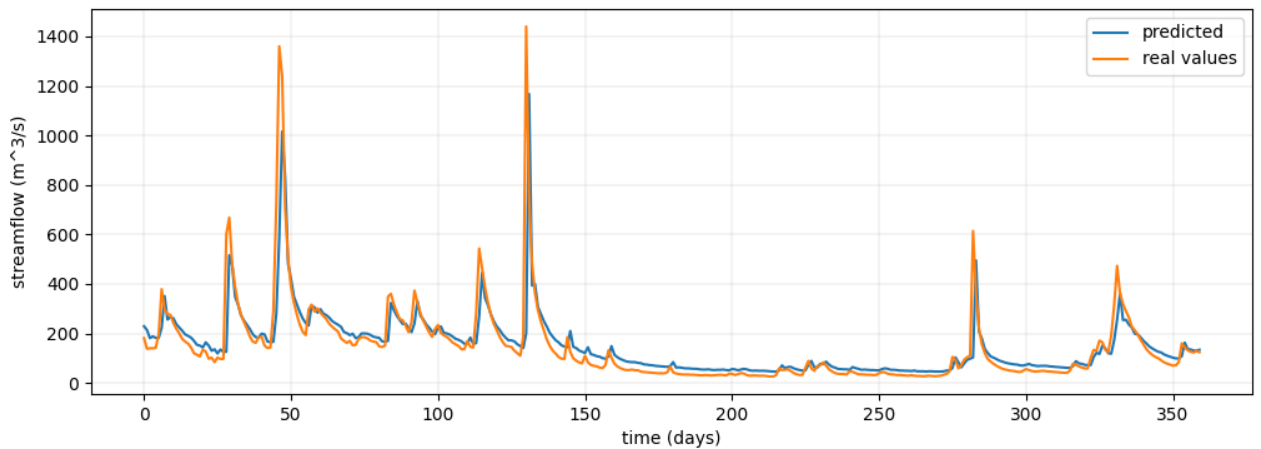


Εικόνα 94 Column chart, ορίζοντας πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 1

Από την πρώτη λεκάνη του ορίζοντα πρόγνωσης μιας ημέρας φαίνεται το μεγάλο αιχμή που σημειώνουν όλα τα μοντέλα στην απόδοση. Από το ψηλότερο NSE ορίζοντα πρόγνωσης 5 ημερών να είναι το 0.29 πλέον το ψηλότερο των μεθόδων ARIMA είναι 0.63. Αξίζει επίσης να σημειωθεί πως τα μοντέλα εκθετικής ομαλοποίησης σημειώνουν παρεμφερή σκορ με τα μοντέλα ARIMA. Τέλος σημειώνεται πως το μοντέλο βαθιάς μηχανικής μάθησης (deep) πετυχαίνει σκορ που ξεπερνά όλα τα υπόλοιπα μοντέλα με NSE = 0.65. Στη συνέχεια παρατίθεται η γραφική παράσταση του deep learning μοντέλου αλλά και του υψηλότερου εκ των μοντέλων ARIMA δηλαδή το ARMA(2,2) .



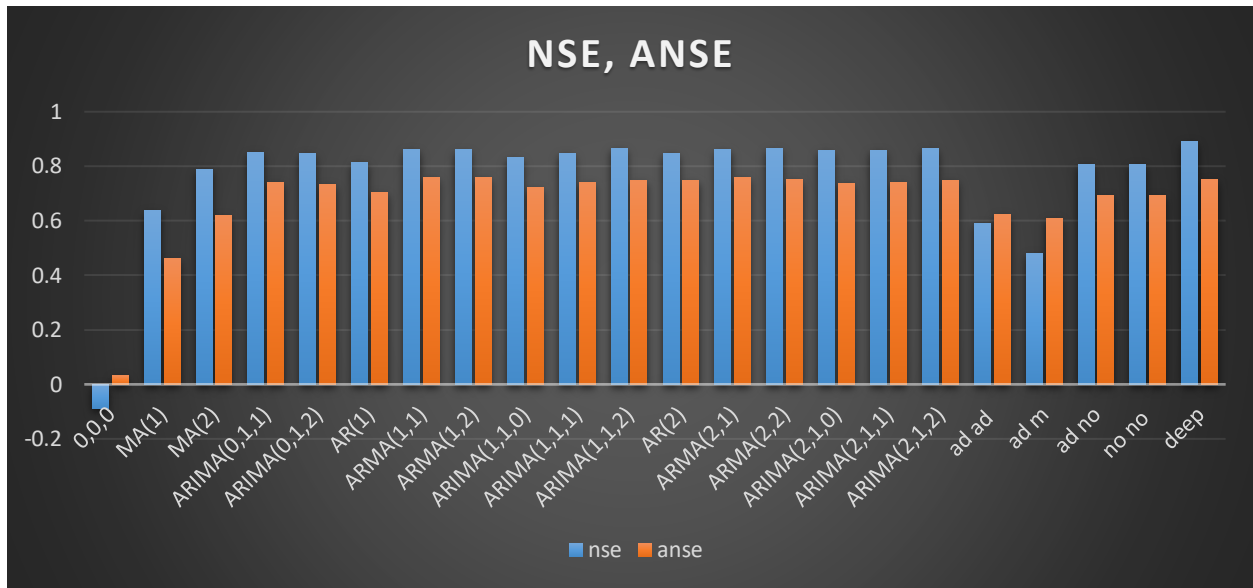
Εικόνα 95 Κοινό διάγραμμα μοντέλου deep learning και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 1



Εικόνα 96 Κοινό διάγραμμα μοντέλου ARMA(2,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 1

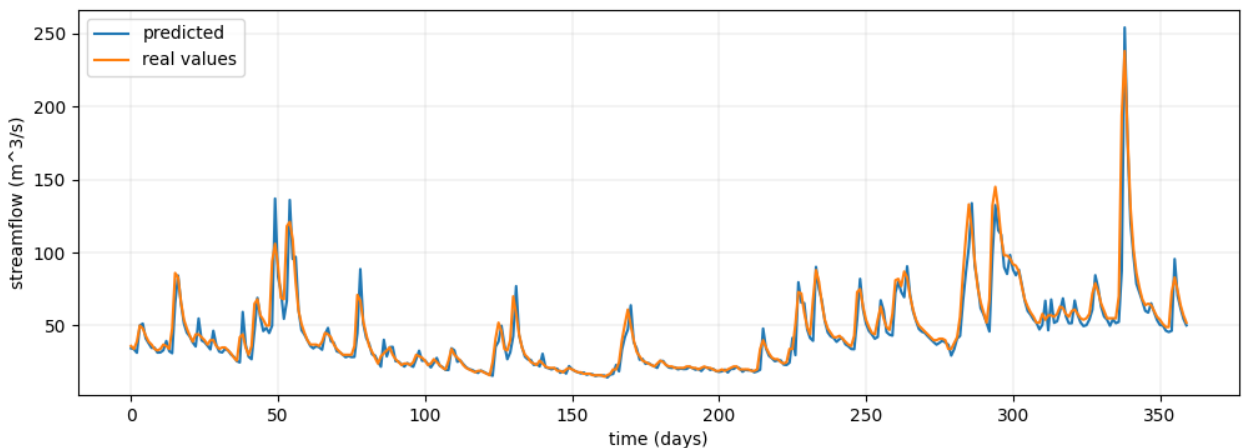
Συγκρίνοντας τα 2 γραφήματα παρατηρείται πως στις 2 πιο ακραίες τιμές απορροής το ARMA συμπεριφέρεται καλύτερα από το deep learning. Το deep learning όμως συμπεριφέρεται καλύτερα στις υπόλοιπες μικρότερης κλίμακας ακραίες τιμές και παράλληλα μπορεί να προσαρμοστεί εξαιρετικά όταν οι τιμές της απορροής είναι μικρής κλίμακας. Για αυτό το λόγο πετυχαίνει καλύτερο σκορ και στους δύο δείκτες.

2) Id λεκάνης 01411300



Εικόνα 97 Column chart, ορίζοντας πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 2

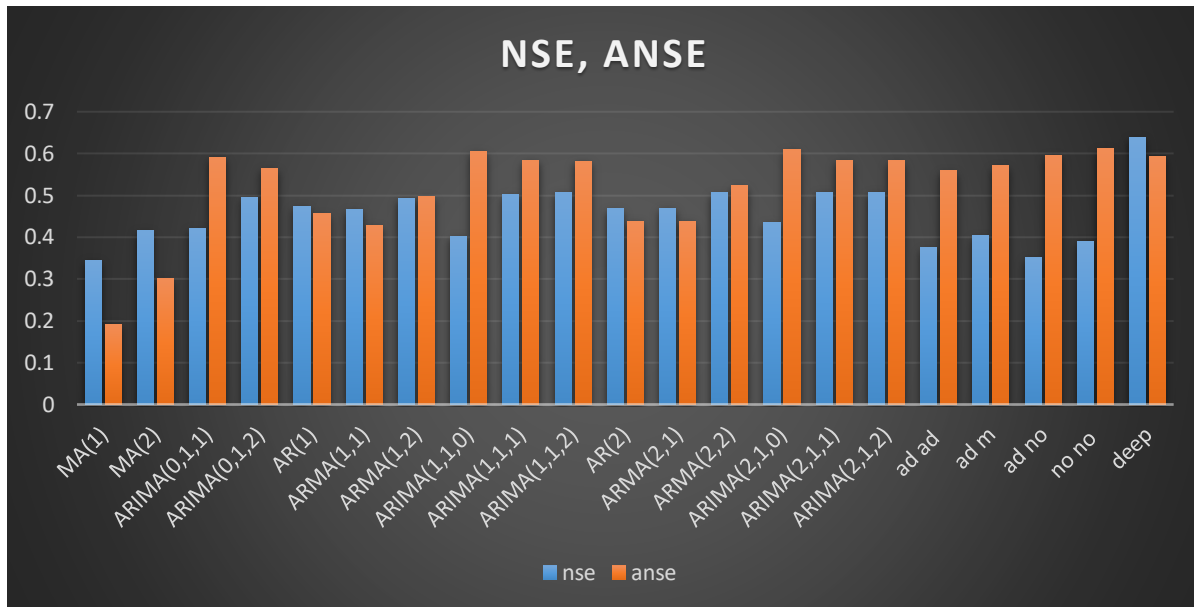
Στην παρούσα λεκάνη πάλι τα σκορ είναι αρκετά αυξημένα στα περισσότερα μοντέλα. Μεταξύ ARIMA και εκθετικής ομαλοποίησης τα ARIMA επιδεικνύουν μια υπεροχή σε αποτελέσματα. Το ARIMA μοντέλο με το καλύτερο σκορ είναι το ARIMA(2,1,2) με NSE = 0.87 και ANSE = 0.75. Το μοντέλο deep learning ξεπερνά και στους δύο δείκτες όλα τα μοντέλα με NSE = 0.88 και ANSE = 0.77. Παρατίθεται το μοντέλο του deep learning.



Εικόνα 98 Κοινό διάγραμμα μοντέλου deep learning και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 2

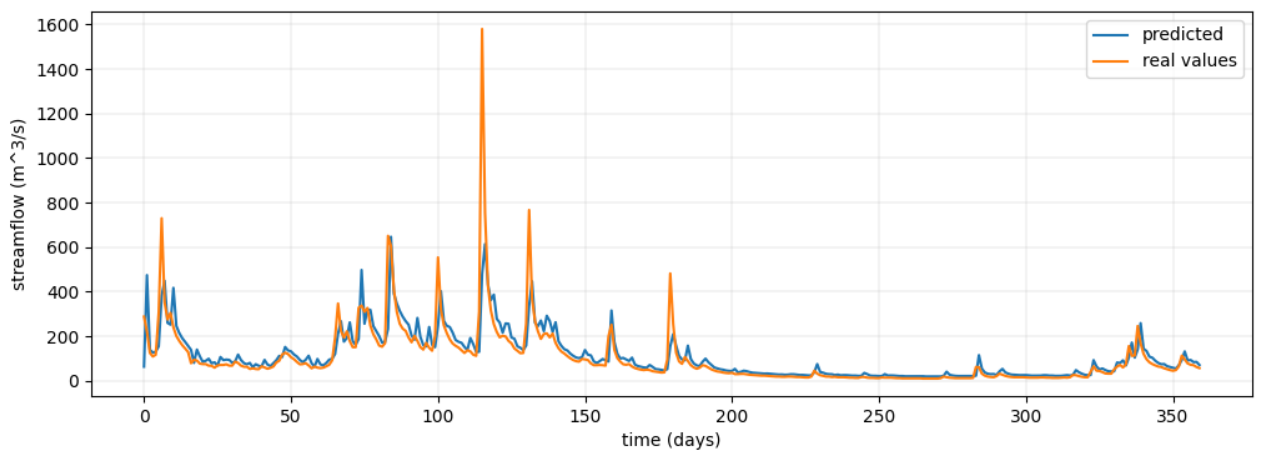
Όπως φαίνεται το νευρωνικό δίκτυο εκτιμά με μεγάλη ακρίβεια και τις ακραίες τιμές αλλά και τις χαμηλότερες.

3) Id λεκάνης 01440000

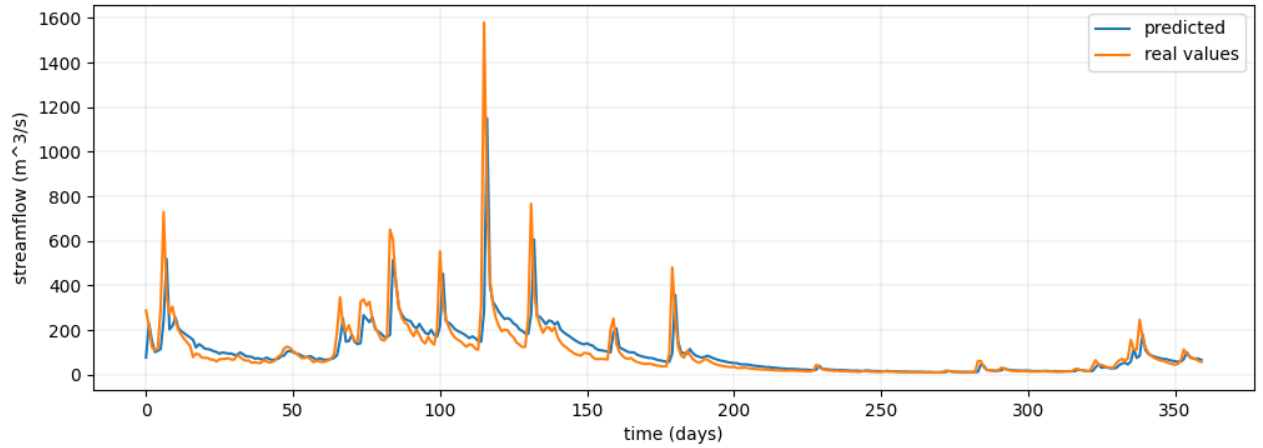


Εικόνα 99 Column chart, ορίζοντας πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 3

Εδώ τα υψηλότερα μοντέλα ARIMA δεν επιτυγχάνουν σκορ NSE αρκετά μεγάλο, το καλύτερο το συγκεντρώνει το ARIMA(2,1,1) με NSE=0.51 και ANSE = 0.58. Το μοντέλο του deep learning πετυχαίνει NSE=0.64 και ANSE = 0.59. Στη συνέχεια παρατίθενται οι γραφικές παραστάσεις των deep learning και ARIMA(2,1,1) αντίστοιχα.



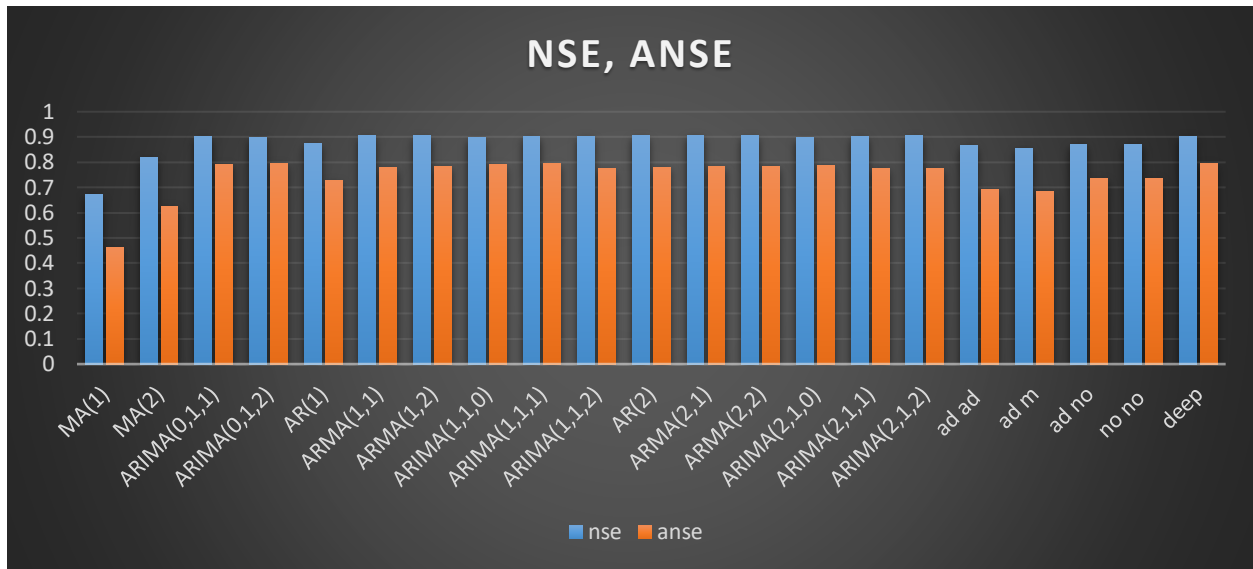
Εικόνα 100 Κοινό διάγραμμα μοντέλου deep learning και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 3



Εικόνα 101 Κοινό διάγραμμα μοντέλου $ARIMA(2,1,1)$ και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 3

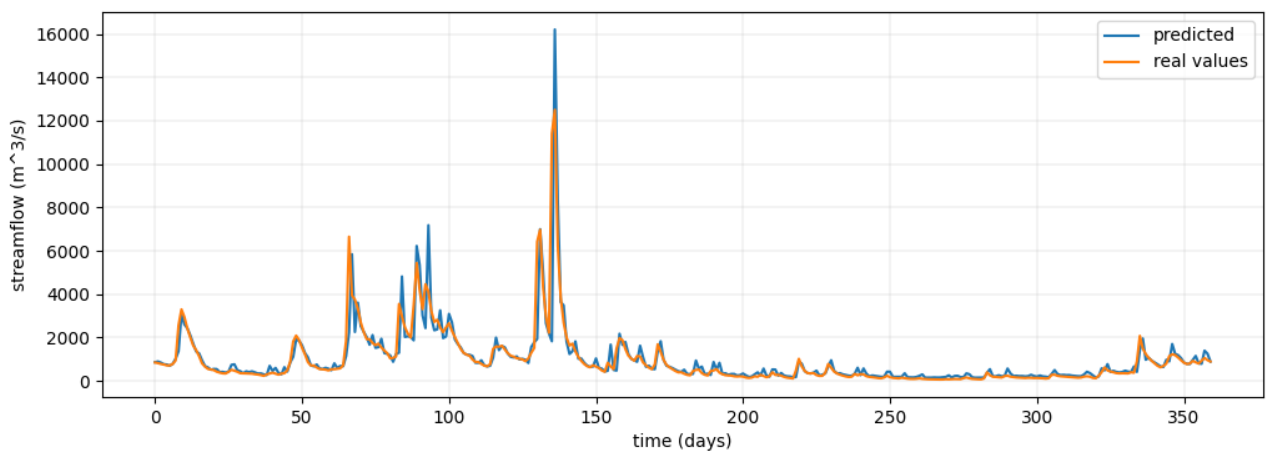
Ενώ το μοντέλο $ARIMA(2,1,1)$ εκτιμά καλύτερα τις ακραίες τιμές παρουσιάζει μια μικρή απόκλιση στις περισσότερες τιμές απορροής. Το deep learning από την άλλη ενώ σημειώνει μειωμένη ακρίβεια στην πρόγνωση των ακραίων τιμών εκτιμά εξαιρετικά όλες τις υπόλοιπες τιμές και για αυτό συγκεντρώνει αρκετά μεγαλύτερο NSE σκορ.

4) Id λεκάνης 01543500



Εικόνα 102 Column chart, οριζοντιας πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 4

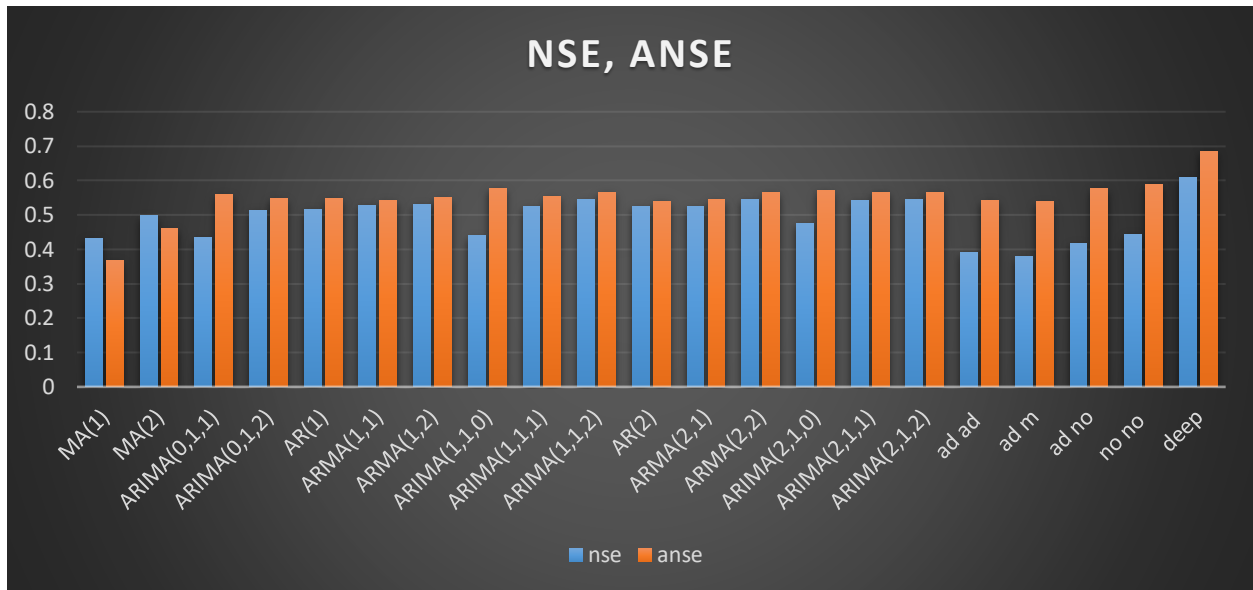
Στην παρούσα λεκάνη αρκετά μοντέλα επιτυγχάνουν σκορ $NSE = 0.9$. Το μοντέλο deep learning εδώ δεν καταφέρνει να ξεπεράσει το ARMA(2,1) στον δείκτη NSE. Για το ARMA(2,1) $NSE = 0.91$ έναντι 0.90 που είναι του deep learning. Στο ANSE το deep learning σημειώνει $ANSE = 0.80$ έναντι 0.78 που είναι του ARMA(2,1). Στη συνέχεια παρατίθεται η γραφική παράσταση του deep learning.



Εικόνα 103 Κοινό διάγραμμα μοντέλου deep learning και πραγματικών τιμών απορροής για οριζοντια πρόγνωσης 1 ημέρας λεκάνης 4

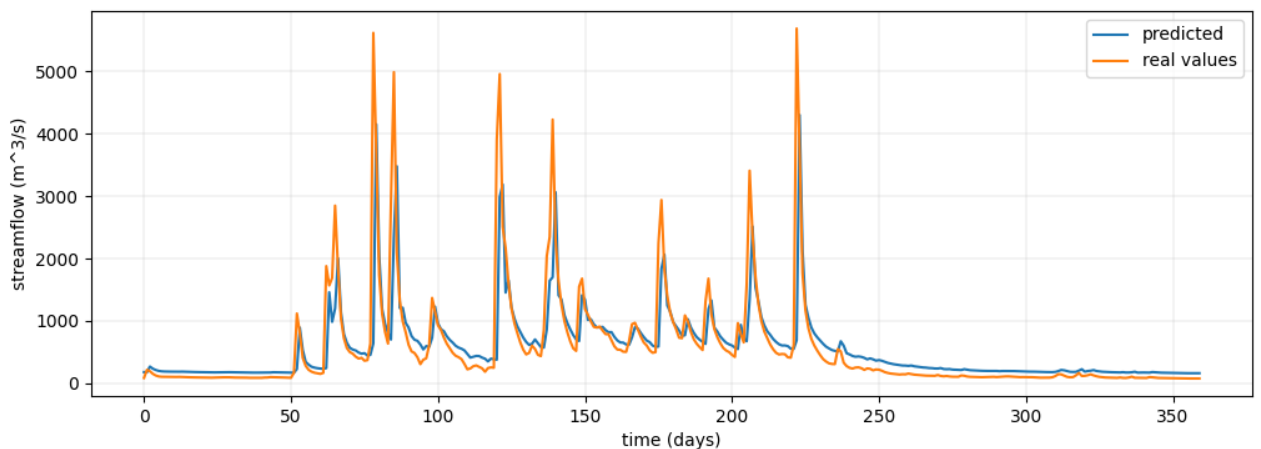
Φαίνεται πως το μοντέλο υπερεκτιμά κάποιες ακραίες τιμές αλλά εκτιμά με μεγάλη ακρίβεια τις υπόλοιπες.

5) Id λεκάνης 02016000

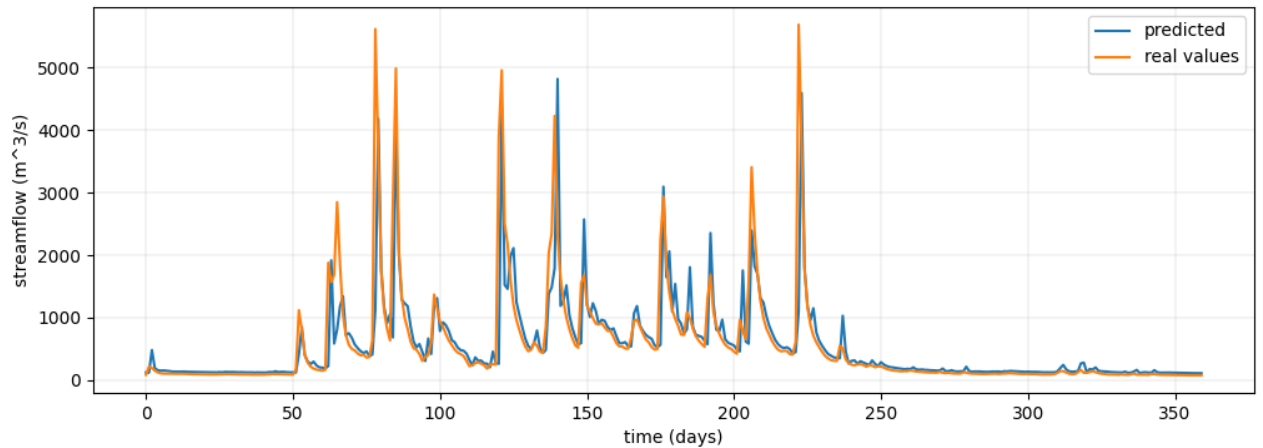


Εικόνα 104 Column chart, ορίζοντας πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 5

Σε αυτή την περίπτωση είναι φανερό πως κανένα από τα μοντέλα ARIMA και εκθετικής ομαλοποίησης δεν ξεπερνούν το όριο του 0.6 σε κανέναν από τους δείκτες. Το πιο επιτυχημένο μοντέλο αναμεσα τους είναι το ARMA(2,2) με NSE = 0.55 και ANSE = 0.57. Το μοντέλο deep learning πετυχαίνει NSE = 0.61 και ANSE = 0.69. Στη συνέχεια παρατίθενται οι γραφικές παραστάσεις των ARMA(2,2) και deep learning αντίστοιχα.



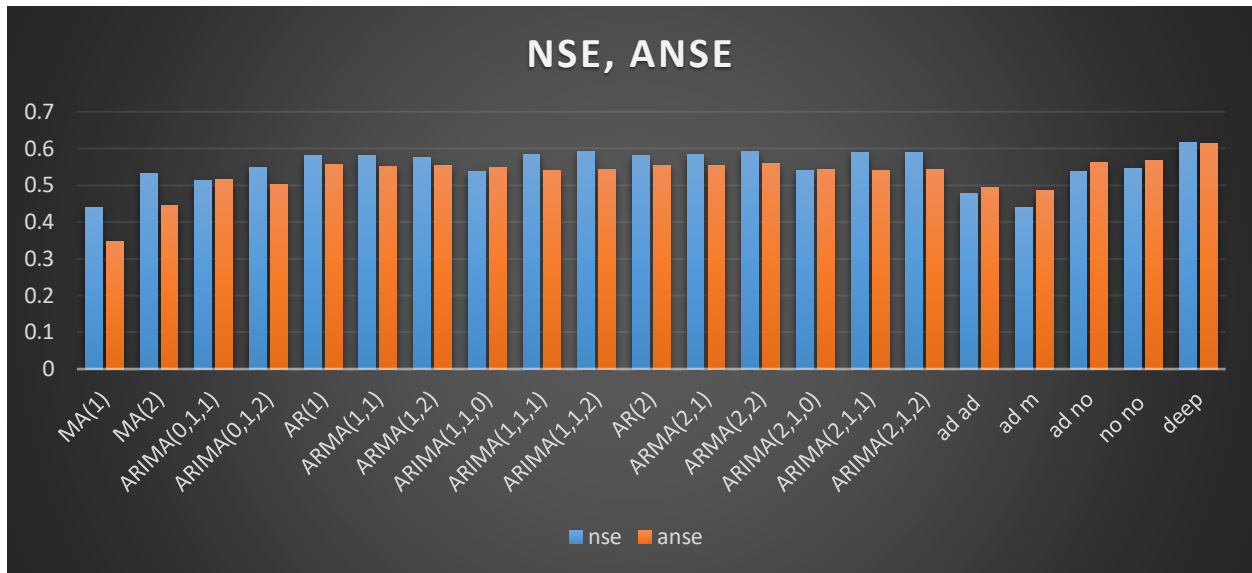
Εικόνα 105 Κοινό διάγραμμα μοντέλου ARMA(2,2) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 5



Εικόνα 106 Κοινό διάγραμμα μοντέλου deep learning και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 5

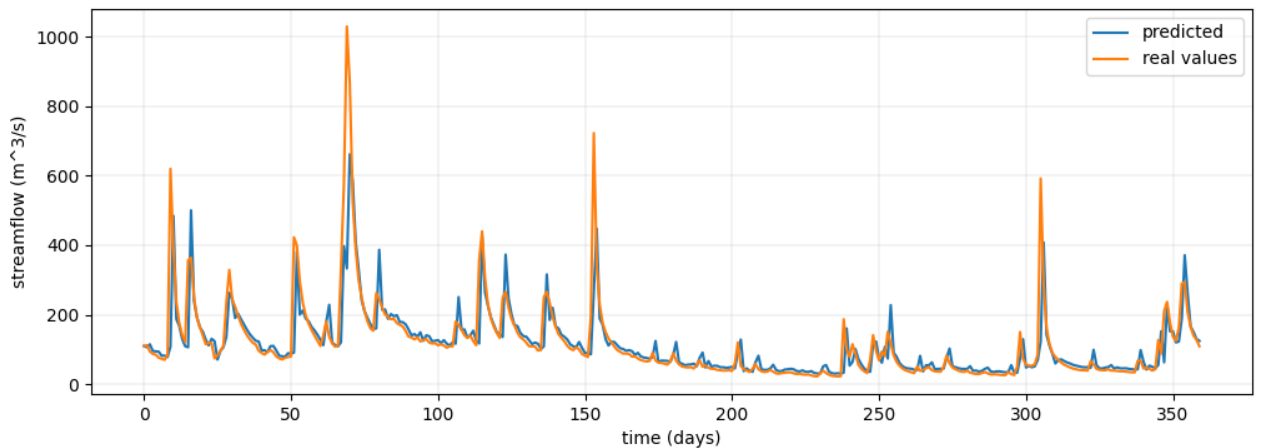
Το μοντέλο deep learning (κάτω) φαίνεται να εκτιμά καλύτερα τις πιο υψηλές τιμές απορροής αλλά και τις πιο χαμηλές, υπερεκτιμά τις ενδιάμεσες σε κάποιες περιπτώσεις. Το μοντέλο ARMA(2,2) ενώ φαίνεται να εκτιμά καλύτερα τις ενδιάμεσες τιμές (εκεί όπου το deep learning τις υπερεκτιμούσε) στις πολύ υψηλές τιμές εκτιμά λιγότερο καλά τις απορροές και στις χαμηλές εμφανίζει ξεκάθαρες αποκλίσεις.

6) Id λεκάνης 02053800



Εικόνα 107 Column chart, οριζόντια πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 6

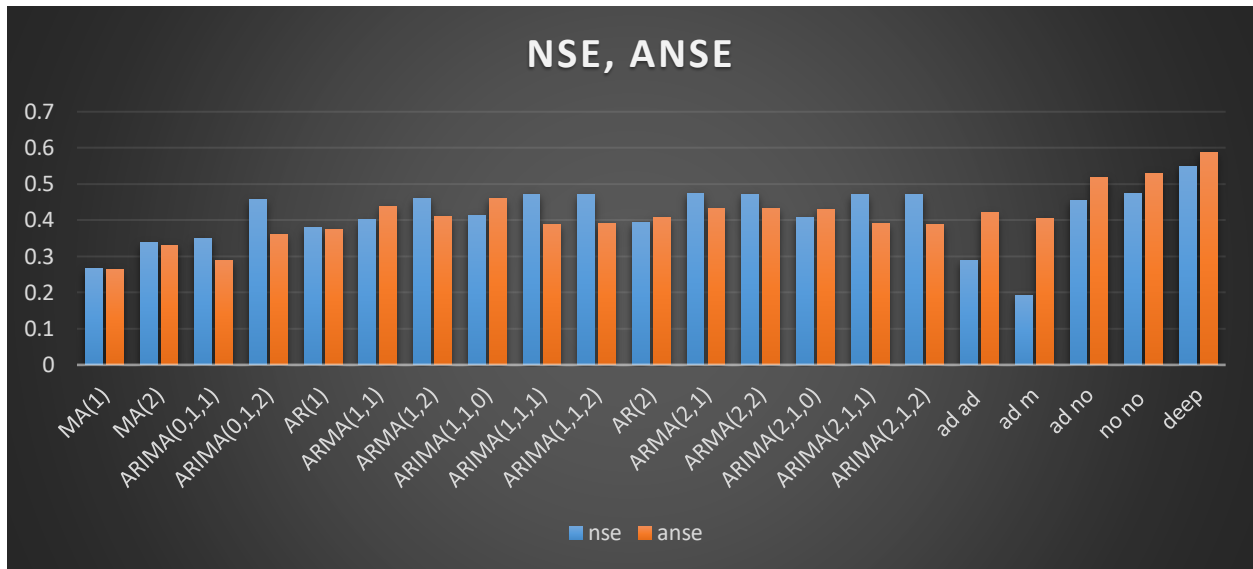
Στην παρούσα λεκάνη συναντώνται οι ίδιες συνθήκες με την ακριβώς προηγούμενη λεκάνη. Το μοντέλο ARMA(2,2) σημειώνει το καλύτερο σκορ μεταξύ των ARIMA και εκθετικής ομαλοποίησης με $NSE = 0.6$ και $ANSE = 0.56$. Το μοντέλο deep learning σημειώνει μεγαλύτερα σκορ με $NSE = 0.62$ και $ANSE = 0.61$. Παρατίθεται το γράφημα του deep learning μοντέλου.



Εικόνα 108 Κοινό διάγραμμα μοντέλου deep learning και πραγματικών τιμών απορροής για οριζόντια πρόγνωσης 1 ημέρας λεκάνης 5

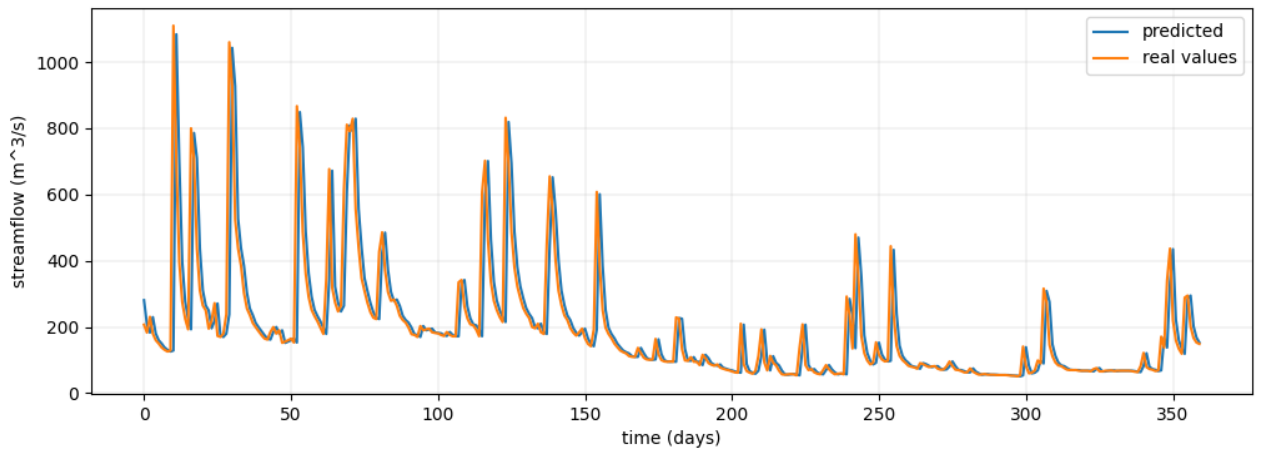
Παρατηρείται ότι το μοντέλο εκτιμά σχετικά καλά τις υψηλές τιμές απορροής ενώ εκτιμά εξαιρετικά τις χαμηλότερες τιμές.

7) Id λεκάνης 02059500

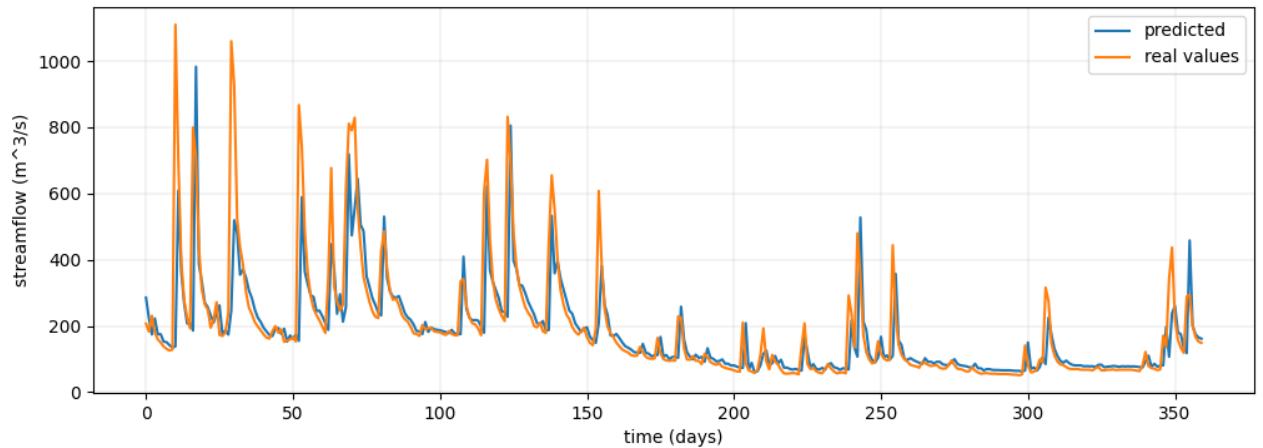


Εικόνα 109 Column chart, οριζόντια πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 7

Σε αυτή τη λεκάνη τα μοντέλα δείχνουν να μην πετυχαίνουν υψηλά σκορ όπως στις προηγούμενες λεκάνες. Από τα μοντέλα ARIMA και εκθετικής ομαλοποίησης το μοντέλο που σημειώνει τα μεγαλύτερα σκορ είναι η απλή εκθετική ομαλοποίηση με $NSE = 0.47$ και $ANSE = 0.53$. Το μοντέλο deep learning πετυχαίνει ακόμη υψηλότερα σκορ πετυχαίνοντας $NSE = 0.55$ και $ANSE = 0.59$. Στη συνέχεια παρατίθενται οι γραφικές παραστάσεις της εκθετικής ομαλοποίησης (no no) και του νευρωνικού δικτύου.



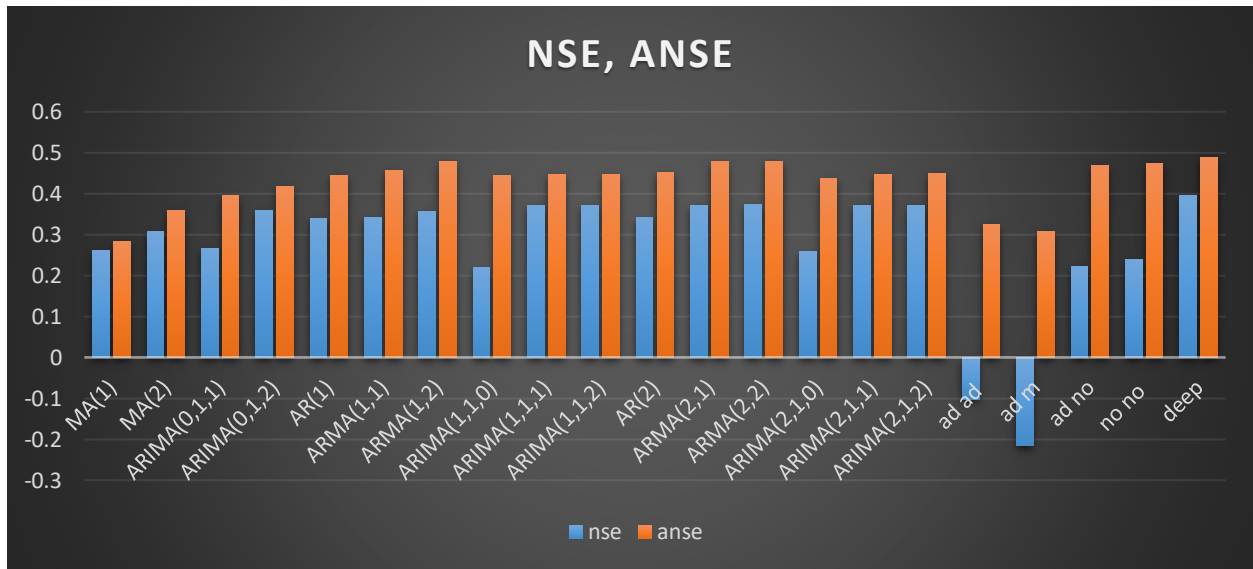
Εικόνα 110 Κοινό διάγραμμα μοντέλου 'no no' και πραγματικών τιμών απορροής για οριζόντια πρόγνωσης 1 ημέρας λεκάνης 7



Εικόνα 111 Κοινό διάγραμμα μοντέλου *deep learning* και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 7

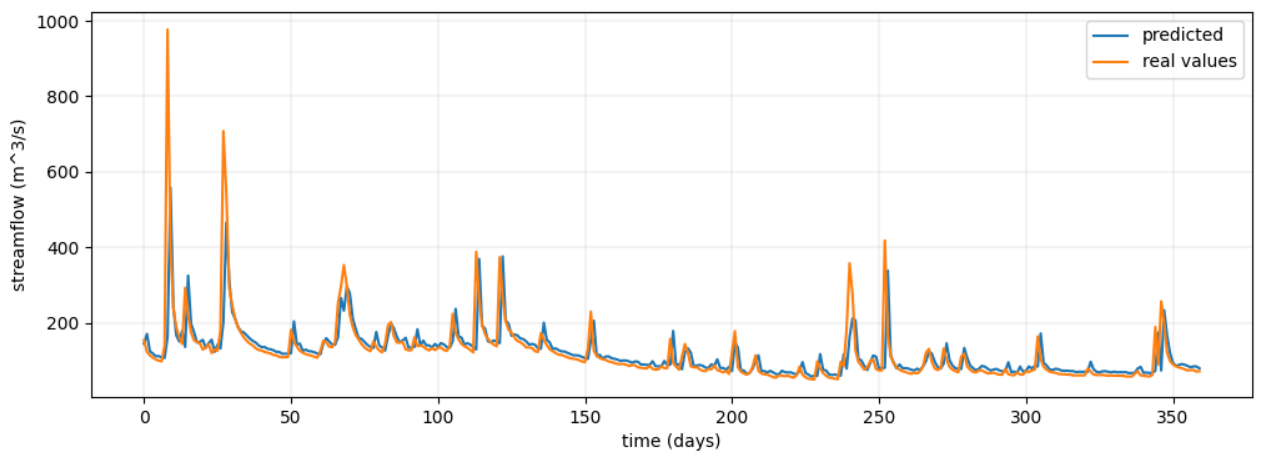
Είναι φανερό πως το μοντέλο απλής εκθετικής ομαλοποίησης οφείλει την επιτυχία του στις καθημερινές ανανεώσεις δεδομένων για αυτό και είναι ένα ελαφρώς μετατοπισμένο προς τα δεξιά αντίτυπο της πραγματικής χρονοσειράς. Εκλείπουν οι οριζόντιες γραμμές που έχουν τα μοντέλα απλής εκθετικής ομαλοποίησης στους ορίζοντες πρόγνωσης 5 και 10 ημερών. Από την άλλη το μοντέλο *deep learning* φαίνεται να παρουσιάζει αποκλίσεις στις εκτιμήσεις αλλά αυτές οι εκτιμήσεις γίνονται την ίδια χρονική στιγμή που πραγματοποιείται και η πραγματική απορροή για αυτό κατάφερε να συγκεντρώσει και μεγαλύτερα σκορ.

8) Id λεκάνης 02069700



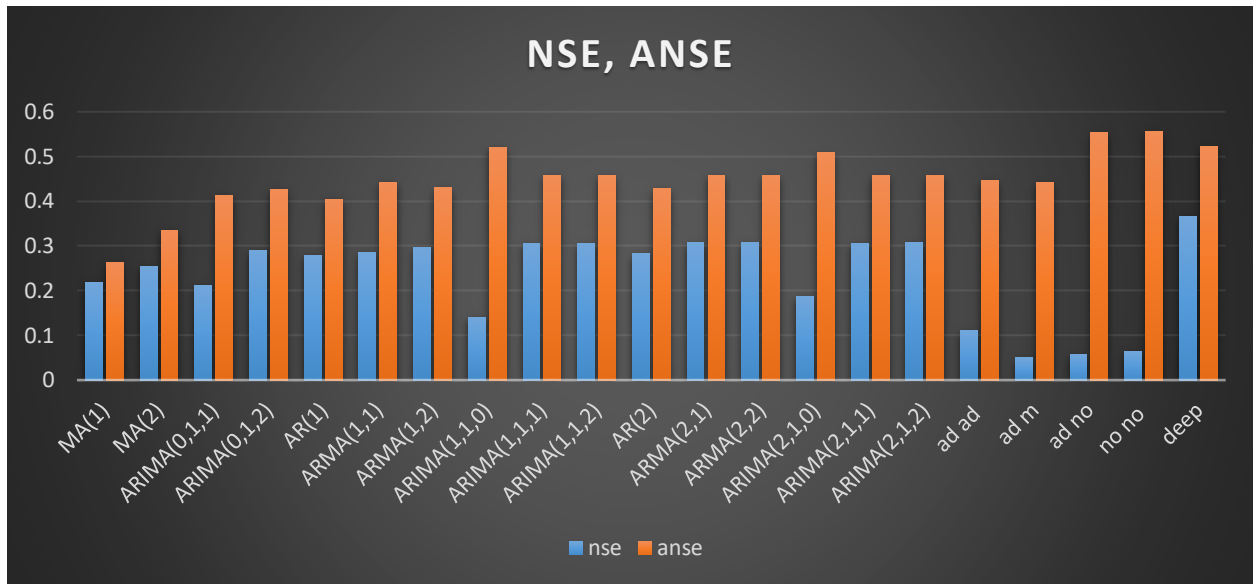
Εικόνα 112 Column chart, ορίζοντας πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 8

Στην παρούσα λεκάνη φαίνεται πως όλα τα μοντέλα δεν σημειώνουν αρκετά υψηλά σκορ. Τα ARIMA συμπεριφέρονται καλύτερα από της εκθετικής ομαλοποίησης απέχουν αρκετά όμως από την τιμή $NSE = 0.4$ με υψηλότερη τιμή την $NSE = 0.37$ που σημειώνει το ARMA(2,2). Το μοντέλο deep learning πετυχαίνει υψηλότερο σκορ και στους 2 δείκτες με $NSE = 0.39$ και $ANSE = 0.49$. Παρατίθεται το γράφημα του μοντέλου deep learning.



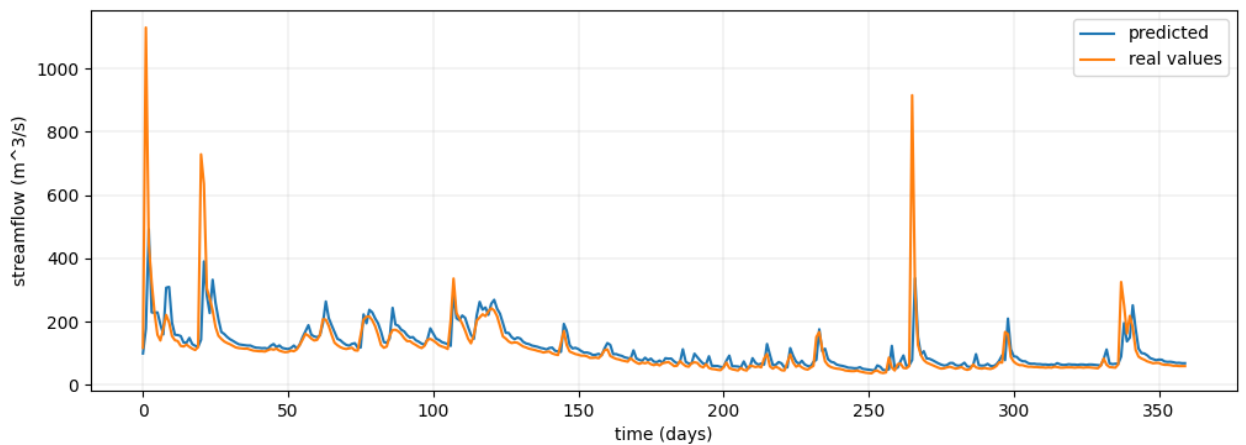
Εικόνα 113 Κοινό διάγραμμα μοντέλου deep learning και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 8

9) Id λεκάνης 02143000

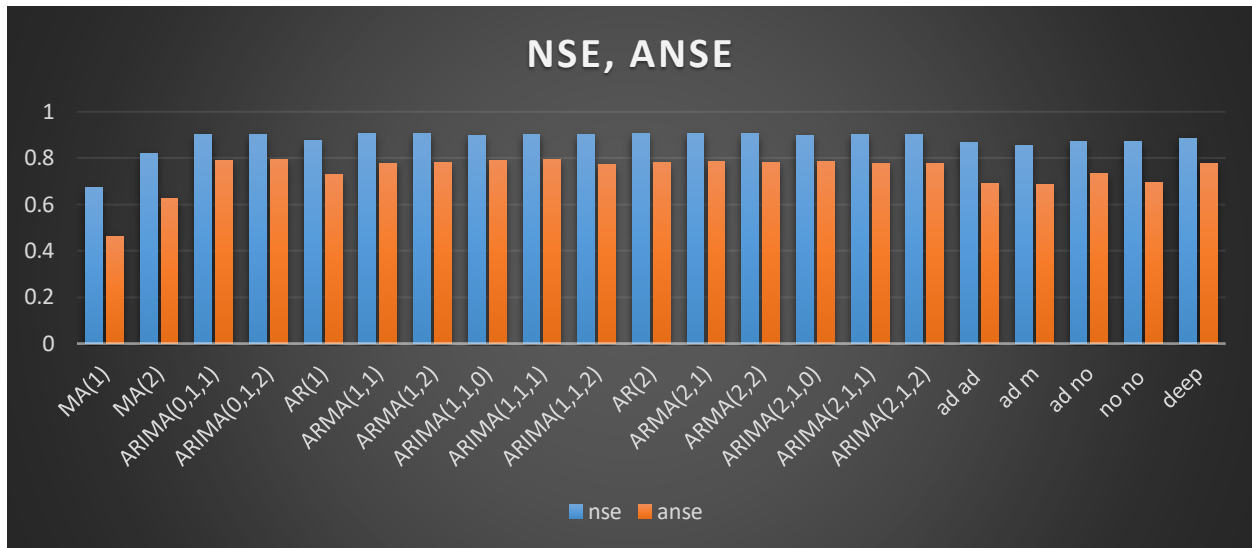


Εικόνα 114 Column chart, ορίζοντας πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 9

Στην παρούσα λεκάνη τα μοντέλα συμπεριλαμβανομένου του νευρωνικού δικτύου πετυχαίνουν αρκετά χαμηλό NSE, κάτω από 0.4. Πιο συγκεκριμένα, το νευρωνικό δίκτυο πετυχαίνει NSE = 0.37 και δεύτερα έρχονται τα ARMA(1,2), ARMA(2,2), ARIMA(2,1,2) με NSE = 0.31. Στη συνέχεια παρατίθεται το γράφημα του deep learning.

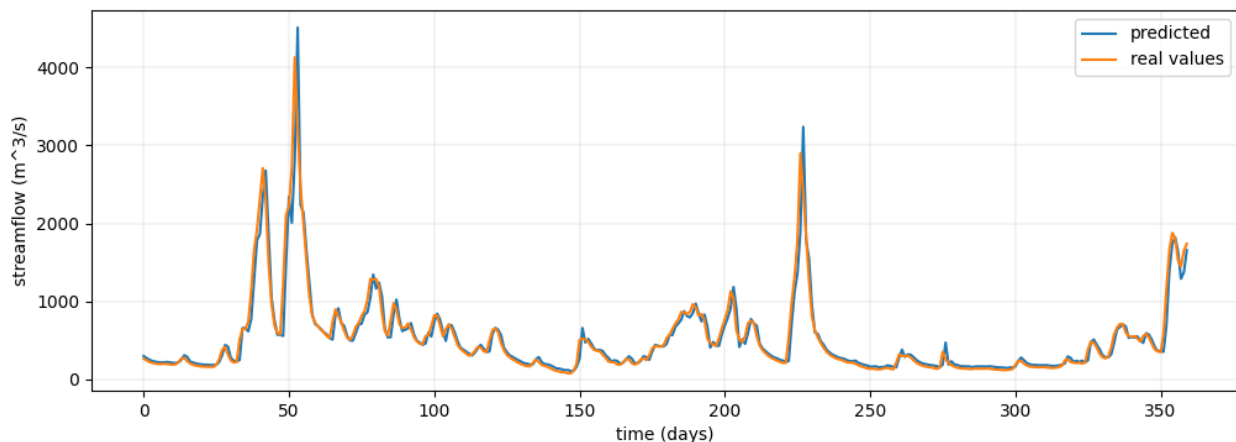


10) Id λεκάνης 02350900

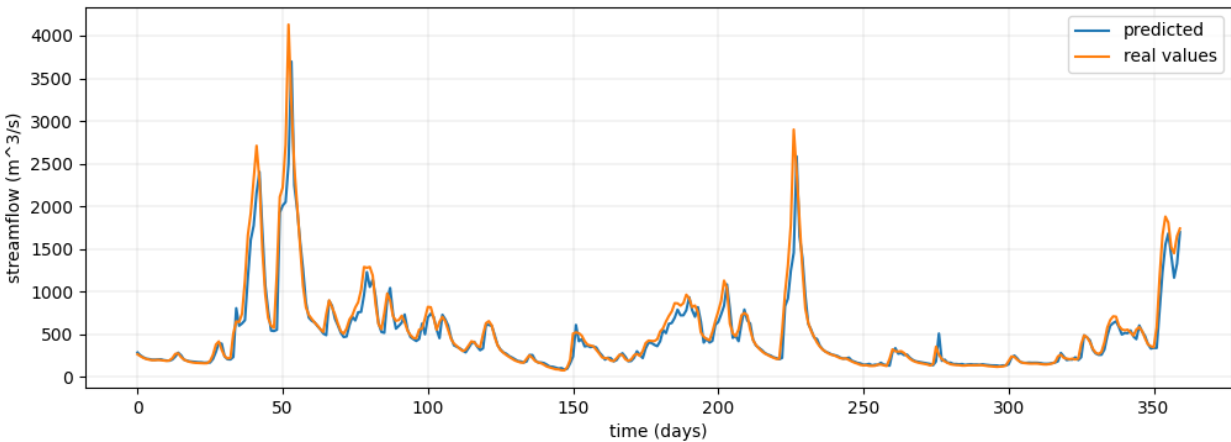


Εικόνα 115 Column chart, ορίζοντας πρόγνωσης 1 ημέρας, αναπαράσταση σκορ NSE και ANSE όλων των μοντέλων λεκάνης 10

Στην παρούσα λεκάνη όλα τα μοντέλα πετυχαίνουν αρκετά υψηλά σκορ. Αξίζει να σημειωθεί πως είναι η μοναδική λεκάνη που το μοντέλο deep learning δεν καταλαμβάνει την υψηλότερη θέση στο σκορ του NSE, τη θέση αυτή έχει το ARMA(2,1) με $NSE = 0.906$ (έναντι 0.902 του νευρωνικού) όσον αφορά το ANSE το ARMA(2,1) σημείωσε σκορ $ANSE = 0.78$ (έναντι 0.80 του νευρωνικού δικτύου). Είναι δεδομένο πως κάθε ένα από αυτά τα μοντέλα μπορεί να χρησιμοποιηθεί ικανοποιητικά για την πρόγνωση απορροής της παρούσας λεκάνης. Στη συνέχεια παρατίθεται η γραφική παράσταση του ARMA(2,1) και του deep learning.



Εικόνα 116 Κοινό διάγραμμα μοντέλου ARMA(2,1) και πραγματικών τιμών απορροής για ορίζοντα πρόγνωσης 1 ημέρας λεκάνης 10



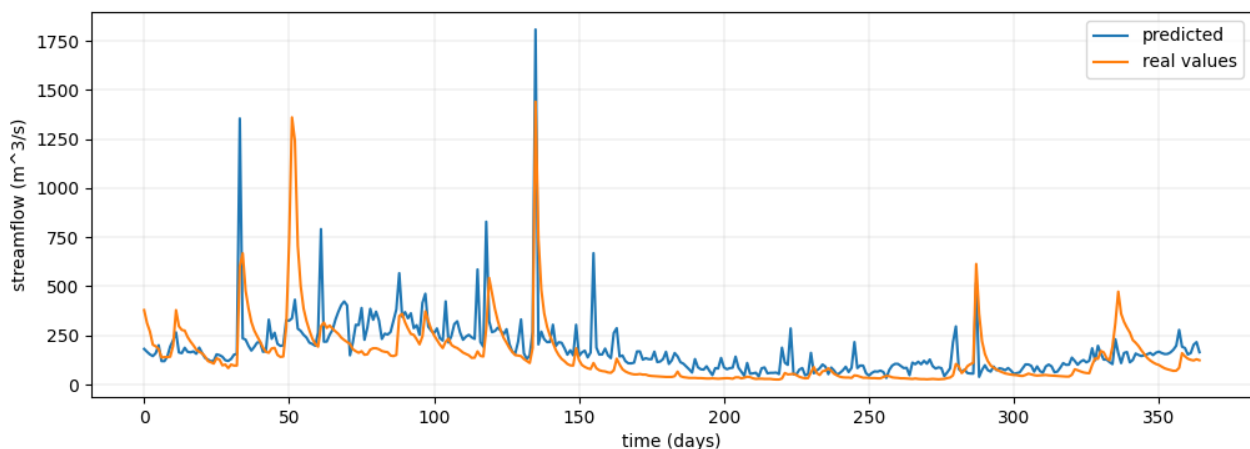
Εικόνα 117 Κοινό διάγραμμα μοντέλου *deep learning* και πραγματικών τιμών απορροής για οριζόντια πρόγνωσης 1 ημέρας λεκάνης 10

Συμπερασματικά το άλμα των τιμών NSE και ANSE είναι αισθητό. Για κάθε μια από τις λεκάνες υπήρχε τουλάχιστον ένα μοντέλο που ξεπερνά την τιμή $NSE = 0.36$ γεγονός που κάνει την πρόγνωση του αποδεκτή σύμφωνα με την Τρίτη έρευνα του υποκεφαλαίου 4.4. Ταυτόχρονα υπήρχαν και 7 από τις 10 λεκάνες που ο δείκτης NSE ήταν μεγαλύτερος ή ίσος του 0.6.

5.4 Μοντέλο *deep learning* προσομοίωσης λεκάνης

Όπως έχει προαναφερθεί το μοντέλο λαμβάνει δεδομένα εισόδου τιμές που δεν αναφέρονται σε παρελθοντικές χρονικές στιγμές της απορροής (όπως κάθε μοντέλο που έχει παρουσιαστεί μέχρι στιγμής) αντί αυτού λαμβάνει δεδομένα που αναφέρονται στην παρούσα χρονική στιγμή εκτός της απορροής.

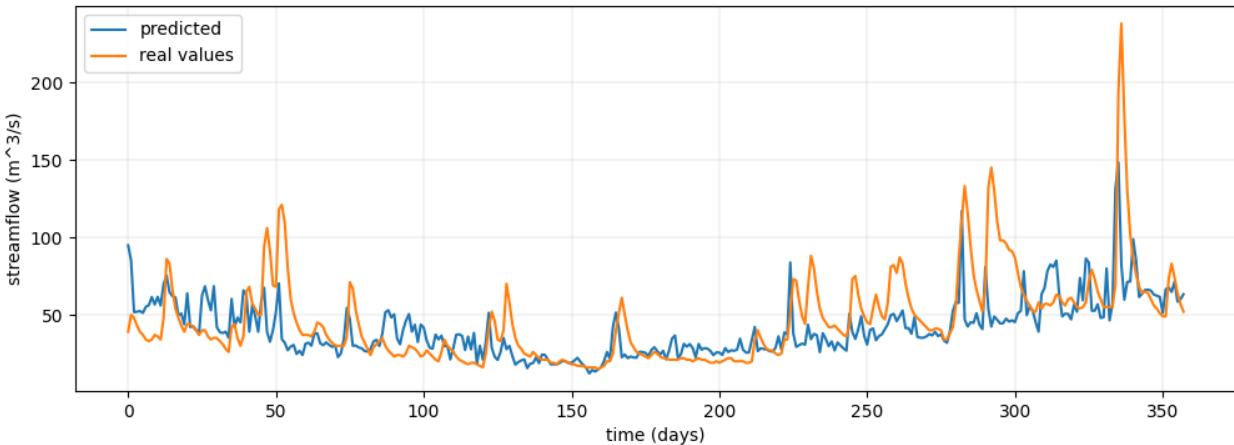
1. Id λεκάνης 02014000



Εικόνα 118 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 1 *deep learning* και πραγματικών τιμών

Το μοντέλο παρατηρείται πως εκτιμά ικανοποιητικά κάποιες από τις υψηλές τιμές της απορροής ενώ σε κάποιες άλλες αστοχεί. Επίσης, παρουσιάζει αποκλίσεις σε πιο χαμηλές τιμές. Τα σκορ που συγκέντρωσε ήταν $NSE = 0.27$ και $ANSE = 0.27$.

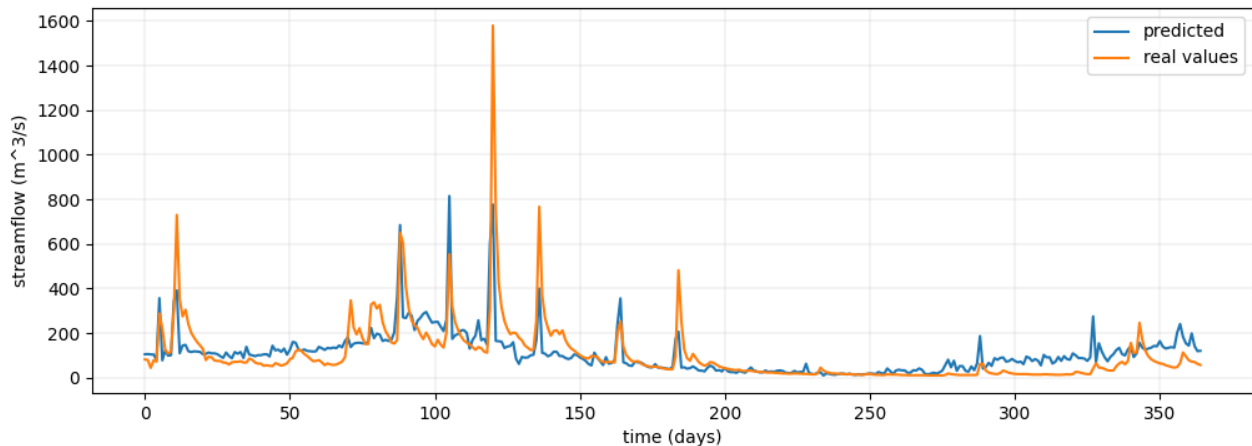
2. Id λεκάνης 01411300



Εικόνα 119 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 2 *deep learning* και πραγματικών τιμών

Στην παρούσα περίπτωση η απορροή έχει μια χρονοσειρά αρκετά ιδιαίτερη διότι οι τιμές είναι σχετικά χαμηλές. Δεν υπάρχουν κάποιες αιχμές υψηλών τιμών όπως στις περισσότερες περιπτώσεις. Το μοντέλο δεν εκτιμά ικανοποιητικά ούτε τις υψηλές αλλά ούτε και τις ενδιάμεσες τιμές απορροής. Σημειώνει $NSE = 0.28$ και $ANSE = 0.25$.

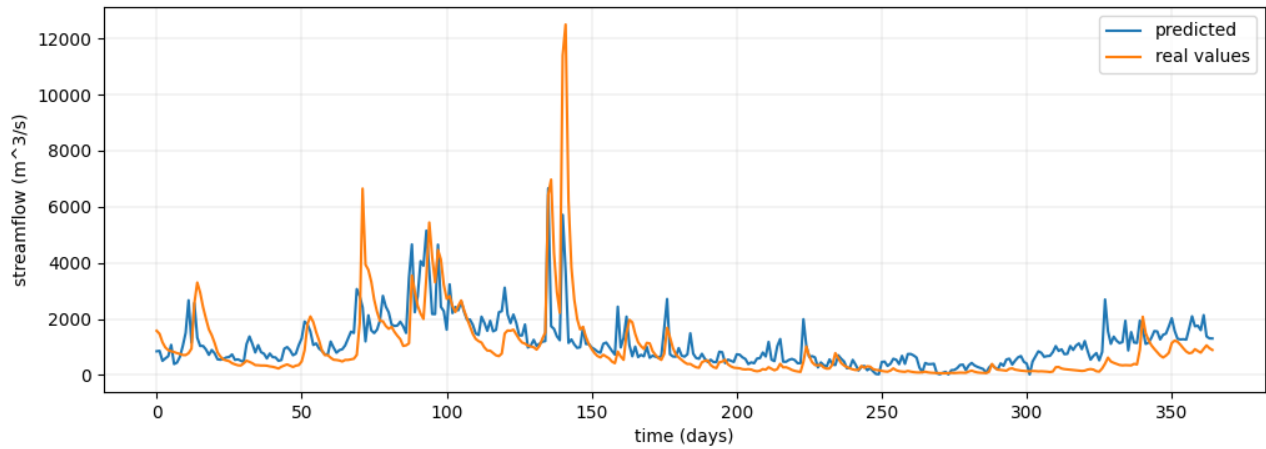
3. Id λεκάνης 01440000



Εικόνα 120 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 3 *deep learning* και πραγματικών τιμών

Το μοντέλο φαίνεται να εκτιμά αρκετά ικανοποιητικά τις αιχμές στις τιμές απορροής και έχει κάποιες αποκλίσεις στις χαμηλότερες τιμές. Τα σκορ που συγκέντρωσε ήταν $NSE = 0.56$, $ANSE = 0.38$.

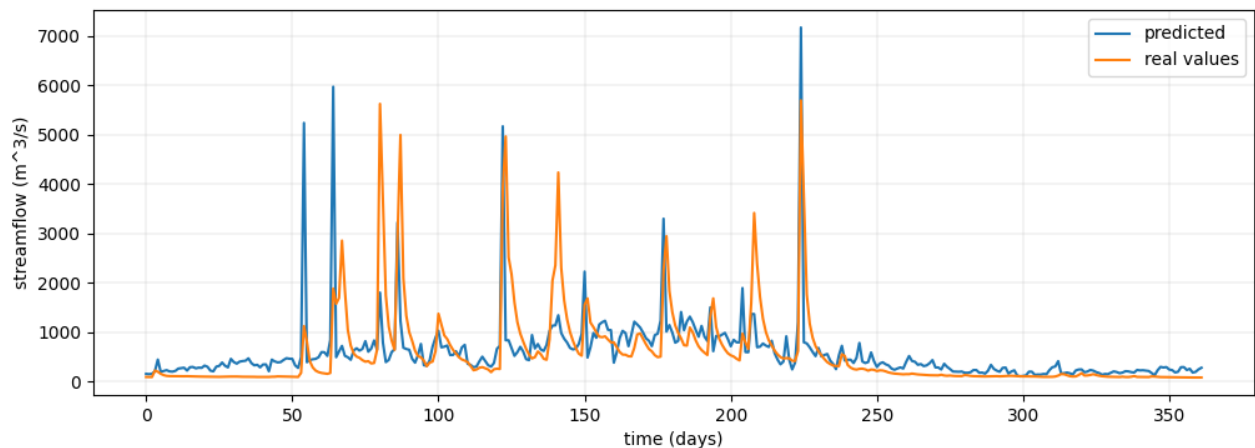
4. Id λεκάνης 01543500



Εικόνα 121 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 4 deep learning και πραγματικών τιμών

Το μοντέλο φαίνεται να εκτιμά αρκετά στοχευμένα τη διπλή αιχμή στηναπορροή των ημερών 135-140. Φαίνεται να αποκλίνει στις πιο χαμηλές τιμές αλλά εν γένει κυμαίνεται αρκετά κοντά στην πραγματική απορροή. Τα σκορ του είναι $NSE = 0.40$, $ANSE = 0.26$.

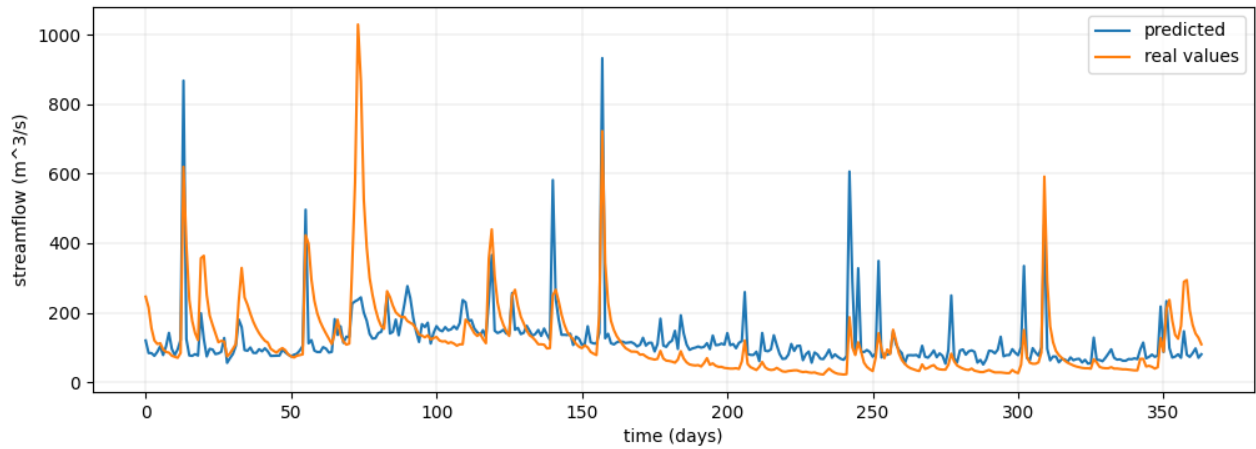
5. Id λεκάνης 02016000



Εικόνα 122 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 5 deep learning και πραγματικών τιμών

Το μοντέλο φαίνεται να υπερεκτιμά κάποιες αιχμές στην τιμή της απορροής και να εκτιμά αρκετά ικανοποιητικά κάποια άλλα. Οι χαμηλές τιμές απορροής εκτιμώνται με μικρές αποκλίσεις. Τα σκορ που συγκεντρώνει είναι $NSE = 0.31$, $ANSE = 0.39$.

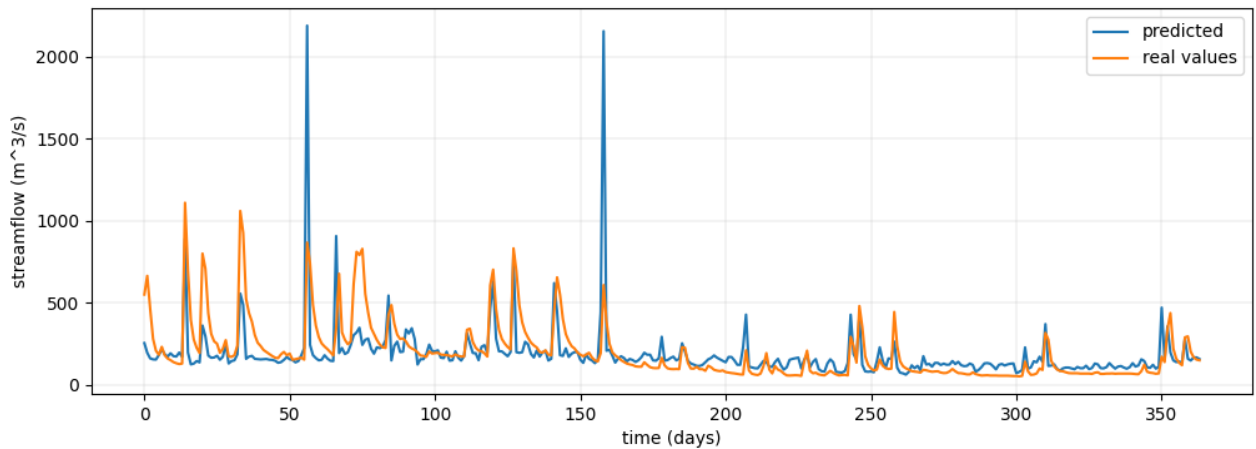
6. Id λεκάνης 02053800



Εικόνα 123 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 6 *deep learning* και πραγματικών τιμών

Στην παρούσα λεκάνη το μοντέλο φαίνεται να αστοχεί σε κάποιες υψηλές τιμές απορροής και να εκτιμά καλύτερα τις χαμηλές παρουσιάζουν όμως και αυτές μικρές αποκλίσεις. Τα σκορ που συγκεντρώνει είναι $NSE = 0.45$, $ANSE = 0.38$.

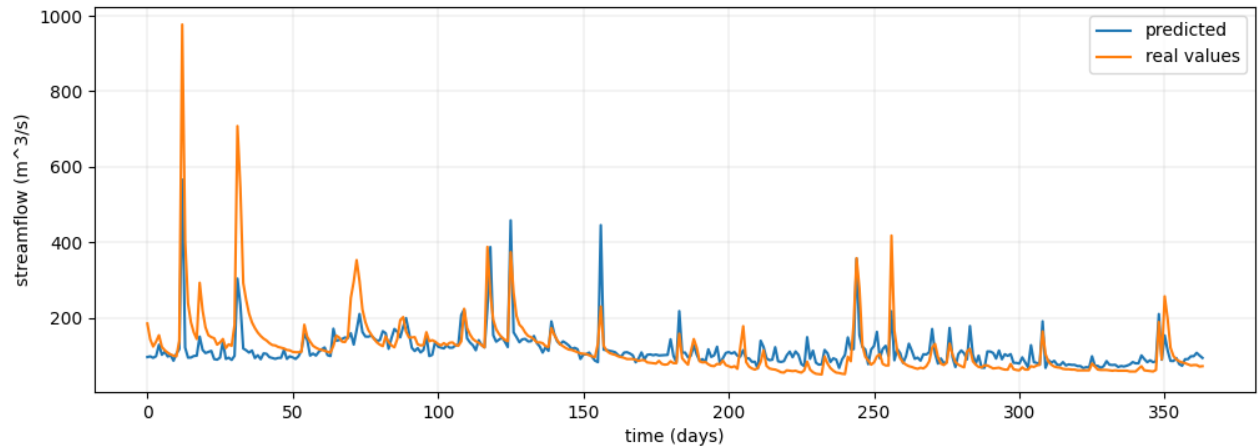
7. Id λεκάνης 02059500



Εικόνα 124 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 7 *deep learning* και πραγματικών τιμών

Στην παρούσα λεκάνη φαίνεται το μοντέλο να εκτιμά αρκετά επιτυχημένα τις αιχμές και τις χαμηλότερες τιμές, με 2 εξαιρέσεις όπου υπερεκτιμά τις απορροές και αστοχεί σε μεγάλο βαθμό. Τα σκορ που σημειώνει είναι $NSE = 0.32$, $ANSE = 0.38$.

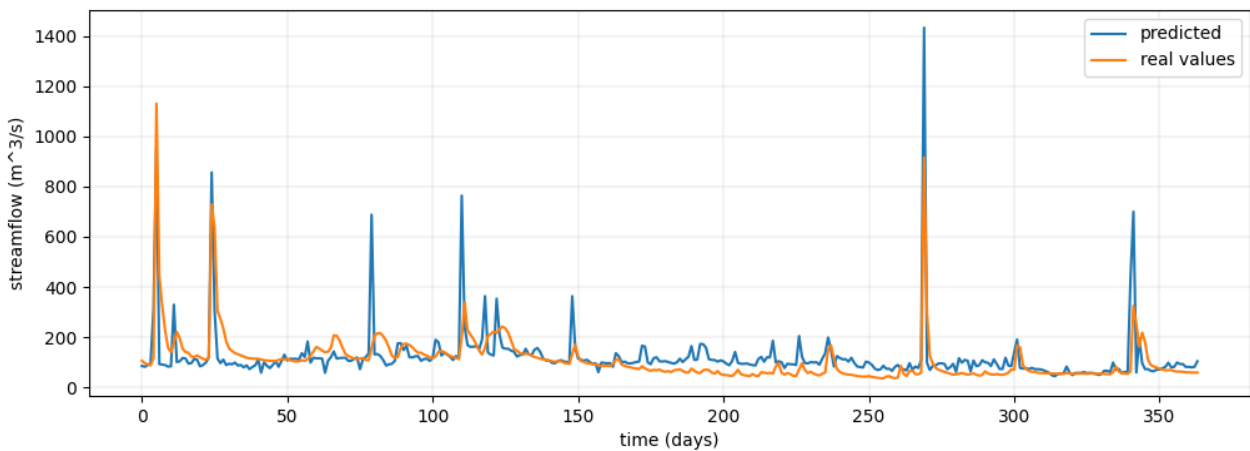
8. Id λεκάνης 02069700



Εικόνα 125 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 8 *deep learning* και πραγματικών τιμών

Στην παρούσα λεκάνη φαίνεται το μοντέλο να εκτιμά αρκετά επιτυχημένα τις υψηλότερες τιμές και να παρουσιάζει μικρές αστοχίες στις χαμηλότερες. Τα σκορ που συγκέντρωσε είναι $NSE = 0.54$, $ANSE = 0.34$.

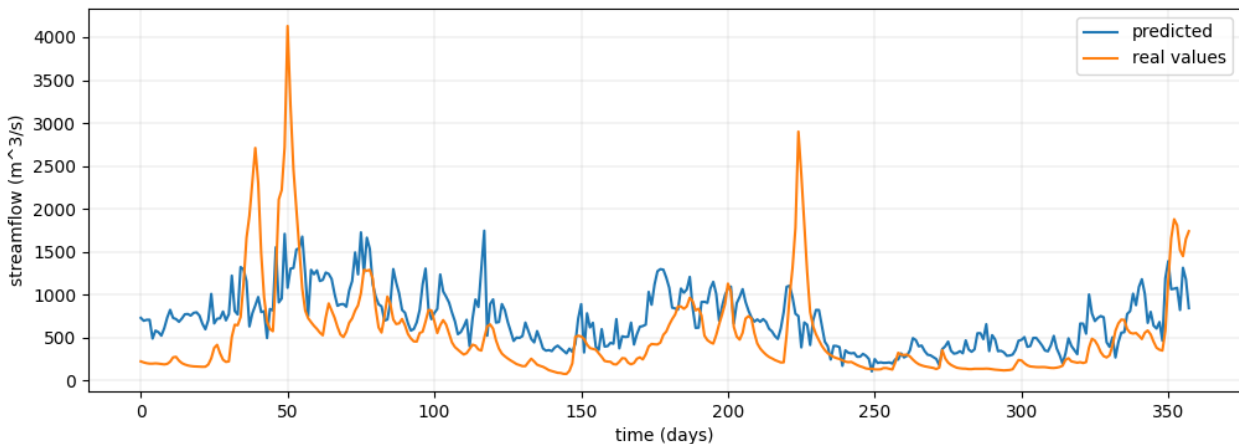
9. Id λεκάνης 0214300



Εικόνα 126 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 9 *deep learning* και πραγματικών τιμών

Το μοντέλο υπερεκτιμά κάποιες χαμηλές τιμές απορροής, από την άλλη εντοπίζει τις αιχμές όποτε συμβαίνουν, εν γένει αστοχεί στις περισσότερες χαμηλές τιμές. Τα σκορ που συγκέντρωσε είναι $NSE = 0.32$, $ANSE = 0.14$.

10. Id λεκάνης 02350900



Εικόνα 127 Κοινό διάγραμμα μοντέλου προσομοίωσης λεκάνης 10 deep learning και πραγματικών τιμών

Στην τελευταία λεκάνη μελέτης το μοντέλο φαίνεται να αστοχεί με αποκλίσεις τόσο στις ενδιάμεσες όσο και στις υψηλές τιμές απορροής. Τα σκορ που συγκεντρώνει είναι $NSE = 0.07$, $ANSE = -0.16$ είναι τα χαμηλότερα σκορ που έχει σημειώσει μέχρι στιγμής το μοντέλο προσομοίωσης λεκάνης.

Έτσι, αφού παρουσιάστηκαν τα αποτελέσματα και των τεσσάρων διαδικασιών ακολουθεί σχολιασμός και συμπεράσματα επί αυτών.

5.5 Συζήτηση επί των αποτελεσμάτων

Στη συνέχεια τα σκορ που σημείωσαν τα μοντέλα στον δείκτη NSE θα συγκριθούν με την κατάταξη που προτείνουν οι ερευνητικές εργασίες που αναφέρθηκαν στο υποκεφάλαιο 4.4. Θα γίνεται αναφορά στις έρευνες με τον όρο ‘πρόταση 1, 2 και 3’ αντίστοιχα

5.5.1 Id λεκάνης 02014000, Potts Creek

Όπως έχει αναφερθεί η λεκάνη του Potts Creek πρόκειται για έναν ποταμό σχετικά μικρού μήκους με αρκετές αιχμές στις τιμές απορροής και μικρής αλλά και μεγάλης κλίμακας.

Στον ορίζοντα των 10 ημερών το ARIMA(1,1,1) αναδείχθηκε καλύτερο με $NSE = 0.22$. Σύμφωνα με τις έρευνες που αναφέρθηκαν στο υποκεφάλαιο 4.4 δεν είναι αξιόπιστη τιμή για αυτο το λόγο δεν μπορεί να χρησιμοποιηθεί ακόμη και αν είναι καλύτερη εκτιμήτρια από τη μέση τιμή (καθώς είναι μεγαλύτερη από 0). Σημειώνεται επίσης ότι τα μοντέλα εκθετικής ομαλοποίησης αστόχησαν να προβλέψουν την απορροή με σκορ NSE τις περισσότερες φορές κάτω του μηδενός.

Στον ορίζοντα των 5 ημερών τα ARIMA(1,1,1) και ARMA(2,2) πετυχαίνουν το υψηλότερο $NSE = 0.29$, μια άνοδο της τάξης του 0.07 από τον ορίζοντα 10 ημερών. Επιπρόσθετα, τα μοντέλα ARIMA(1,1,0) και ‘ad no’, ‘no no’ σημειώνουν τιμές δείκτη $ANSE = 0.38$. Η τιμή αυτή αν

πρόκειται για διαχείριση νερού ίσως να είναι ικανοποιητική. Όσον αφορά το NSE τα μοντέλα κρίνονται μη ικανοποιητικά και από τις 3 προτεινόμενες κατατάξεις του υποκεφαλαίου 4.4.

Στον ορίζοντα της 1 ημέρας όλα τα μοντέλα ARIMA σημείωσαν τιμές κοντά στο $NSE = 0.60$ ενώ τα μοντέλα εκθετικής ομαλοποίησης τιμές κοντά στο $NSE = 0.55$, έτσι είναι προφανές πως τα ARIMA υπερτερούν των μοντέλων εκθετικής ομαλοποίησης. Το καλύτερο μοντέλο ήταν το ARMA(2,2) με $NSE = 0.63$ σημειώνοντας άνοδο της τάξης του 0.34 από το αντίστοιχο των 5 ημερών. Στο ANSE το καλύτερο σκορ από την οικογένεια ARIMA σημειώθηκε από το ARIMA(2,1,0) με $ANSE = 0.65$. Πρέπει να υπογραμμιστεί ότι το μοντέλο deep learning κατάφερε να ξεπεράσει τα καλύτερα σκορ και των δύο περιπτώσεων σημειώνοντας $NSE = 0.65$ και $ANSE = 0.68$. Σε σχέση με τα εύρη που έχουν αναφερθεί στο υποκεφάλαιο 4.4 το μοντέλο ARMA(2,2) θεωρείται ικανοποιητικό για τις προτάσεις 1,3 και καλό για την πρόταση 2. Το deep learning θεωρείται ικανοποιητικό για την πρόταση 3 και καλό για τις προτάσεις 1, 2. Είναι αισθητό λοιπόν πως πετυχαίνει καλύτερες αποδόσεις και μπορεί να ληφθεί υπόψη σε μελέτες.

5.5.2 Id λεκάνης 01411300, Tuckahoe river

Η λεκάνη του ποταμού αποτελείται από απορροές συνήθως κάτω των $200 \text{ m}^3/\text{s}$ αλλά έχει κάποιες αιχμές στις τιμές απορροής μεσαίου μεγέθους που φτάνουν λίγο χαμηλότερα από τα $400 \text{ m}^3/\text{s}$. Στον ορίζοντα πρόγνωσης 10 ημερών το ARIMA(1,1,2) είναι αυτό που σημειώνει το μεγαλύτερο σκορ με $NSE = 0.363$ είναι αρκετά υψηλό σκορ για ορίζοντα 10 ημερών και μαζί με τη λεκάνη 10 είναι τα μόνα σκορ που είναι ικανοποιητικά σύμφωνα με την πρόταση 3.

Στον ορίζοντα 5 ημερών το ARIMA(1,1,2) συνεχίζει να πετυχαίνει το μεγαλύτερο σκορ με $NSE = 0.50$ θεωρείται ικανοποιητικό από τις προτάσεις 1,3 και καλό από την πρόταση 2.

Στον ορίζοντα 1 ημέρας πολλά μοντέλα ARIMA σημειώνουν σκορ κοντά στο $NSE = 0.86$ ένα σκορ που καθιστά το μοντέλο εξαιρετικό και στις 3 προτάσεις. Να σημειωθεί ότι το νευρωνικό δίκτυο το υπερβαίνει με σκορ $NSE = 0.89$. Εδώ φαίνεται πως δεν είναι απαραίτητο όταν η λεκάνη δίνει εξαιρετικά μοντέλα τα ARIMA να υπερέχουν του deep learning.

5.5.3 Id λεκάνης 01440000, Flatbrook river

Ο Flatbrook είναι ο μικρότερος ποταμός από άποψη μήκους, οι απορροές του είναι σχετικά μικρές με ακρετές αιχμές με παροχές έως 1000 κυβικά μέτρα το δευτερόλεπτο και κάποια λιγότερες αιχμές άνω των 1000 κυβικών στη διάρκεια των ετών.

Στον ορίζοντα των 10 ημερών κανένα μοντέλο δεν ξεπερνά το $NSE = 0.21$ και επομένως κρίνονται μη ικανοποιητικά και από τις 3 προτάσεις. Αξίζει να σημειωθεί πως τα μοντέλα εκθετικής ομαλοποίησης σημειώνουν $NSE < 0$ γεγονός που δείχνει την αστοχία τους. Τα ARIMA λειτουργούν καλύτερα αλλά όχι αρκετά καλά για να ληφθούν υπόψη σε μια μελέτη.

Στον ορίζοντα των 5 ημερών το μοντέλο ARMA(2,2) πετυχαίνει σκορ $NSE = 0.32$ σημειώνοντας μια άνοδο της τάξης του 0.11 από τον ορίζοντα 10 ημερών το μοντέλο ARIMA(2,1,2) σημειώνει το ίδιο NSE και $ANSE = 0.36$ (έναντι $ANSE=0.27$ του ARMA(2,2)). Το γεγονός πως υπάρχουν πολλά μοντέλα με δείκτη ANSE μεγαλύτερο του δείκτη NSE υποδεικνύει πως πολλά μοντέλα

μπορούν να εκτιμούν καλά τις χαμηλότερες τιμές απορροής και αστοχούν στις υψηλότερες τιμές. Κανένα από τα μοντέλα δεν κρίνεται ικανοποιητικό.

Στον ορίζοντα της 1 ημέρας τα περισσότερα μοντέλα ARIMA σημειώνουν NSE που αγγίζει το 0.50. Τα μοντέλα εκθετικής ομαλοποίησης βρίσκονται αρκετά χαμηλότερα με $NSE = 0.40$. Το μοντέλο ARIMA(2,1,1) σημειώνει $NSE = 0.51$ και $ANSE = 0.58$. Το μοντέλο deep learning σημειώνει $NSE = 0.64$ και $ANSE = 0.59$, ενώ δηλαδή διατηρεί την ίδια ευστοχία με το ARIMA (2,1,1) (ελαφρώς καλύτερη) στις χαμηλές τιμές απορροής στις υψηλές βελτιώνεται κατά πολύ πετυχαίνοντας NSE αυξημένο κατά 0.13. Σύμφωνα με την πρόταση 1 το μοντέλο είναι ικανοποιητικό αλλά πολύ κοντά στο να χαρακτηριστεί καλο, σύμφωνα με την πρόταση 2 θεωρείται καλό και με την πρόταση 3 θεωρείται ικανοποιητικό. Είναι φανερό λοιπόν πως βελτιώνει σημαντικά την πρόγνωση σε σχέση με το καλύτερο μοντέλο ARIMA.

5.5.4 Id λεκάνης 01543500, Sinnemahoning Creek

Οι απορροές του Sinnemahoning creek είναι αρκετά υψηλές σε σχέση με άλλες λεκάνες που έχουν εξεταστεί στην παρούσα διπλωματική εργασία. Η πιο ψηλή τιμή απορροής είναι $35,000 \text{ m}^3 / \text{s}$.

Στον ορίζοντα πρόγνωσης των 10 ημερών το ARMA(2,2) που πετυχαίνει τα ψηλότερα σκορ σημειώνει $NSE = 0.21$ και $ANSE = 0.22$. Δεν μπορεί να θεωρηθεί ικανοποιητικό μοντέλο. Τα μοντέλα εκθετικής ομαλοποίησης σημειώνουν αρνητικές τιμές NSE σε 3 από τις 4 περιπτώσεις γεγονός που υποδεικνύει την αστοχία τους.

Στον ορίζοντα πρόγνωσης των 5 ημερών σημειώνεται σκορ $NSE = 0.36$ (αύξηση 0.14 από τον ορίζοντα πρόγνωσης 10 ημερών) και $ANSE = 0.34$ από το μοντέλο ARMA(2,2). Αστοχία στην εκτίμηση των υψηλών τιμών απορροής. Για τον παραπάνω λόγο δεν θεωρείται ικανοποιητικό από τις προτάσεις 1 και 2, από την πρόταση 3 θεωρείται οριακά ικανοποιητικό.

Στον ορίζοντα πρόγνωσης της 1 ημέρας σημειώνεται πολύ μεγάλο άλμα με τα σκορ NSE να φτάνουν το 0.9 (αύξηση κατά 0.54 από τον ορίζοντα πρόγνωσης 5 ημερών). Το ARMA(2,1) σημειώνει το υψηλότερο σκορ $NSE=0.91$ ενώ ακολουθεί το deep learning με 0.90. Τα σκορ αυτά καθιστούν τα μοντέλα εξαιρετικά και στις 3 προτάσεις. Παρατηρείται πως στις λεκάνες που τα μοντέλα σημειώνουν σκορ $NSE > 0.9$ το καλύτερο από τα μοντέλα ARIMA συμπεριφέρεται καλύτερα από το deep learning.

5.5.5 Id λεκάνης 02016000, Cowpasture river

Η χρονοσειρά λεκάνης απορροής του Cowpasture δείχνει έναν ποταμό με αρκετές αιχμές τιμών αλλά μεσαίου μεγέθους, εξάιρεση αποτελεί η αιχμή τιμών του 1986 με τιμή απορροής να φτάνει λίγο πιο κάτω από $35.000 \text{ m}^3/\text{s}$.

Στον ορίζοντα των 10 ημερών το ARMA(2,2) είναι αυτό που σημειώνει το υψηλότερο σκορ με $NSE = 0.16$ και $ANSE = 0.23$. Τα σκορ αυτά είναι αρκετά χαμηλά και το μοντέλο κρίνεται μη ικανοποιητικό και από τις 3 προτάσεις του υποκεφαλαίου 4.4. Αξίζει επίσης να σημειωθεί πως όλα τα μοντέλα εκθετικής ομαλοποίησης σημειώνουν σκορ με αρνητικές τιμές.

Στον ορίζοντα των 5 ημερών το ARIMA(2,2) σημειώνει τα καλύτερα σκορ με NSE = 0.23 και ANSE = 0.29. Σημειώνεται αύξηση του NSE σε σχέση με τον ορίζοντα 10 ημερών αλλά είναι αρκετά μικρή για να μπορέσει να θεωρηθεί ικανοποιητικό.

Τέλος, στον ορίζοντα 1 ημέρας το ARMA(2,2) βρίσκεται στην ίδια θέση με τα ARIMA(2,1,2) και ARIMA(2,1,1) σημειώνοντας NSE = 0.54 και ANSE = 0.57. Η αύξηση αυτή της τάξης του 0.32 στον δείκτη NSE το καθιστά ικανοποιητικό σύμφωνα με όλες τις προτάσεις. Τέλος να υπογραμμιστεί ότι το deep learning πετυχαίνει NSE = 0.61 και ANSE = 0.69. Αυτά τα σκορ το καθιστούν ικανοποιητικό σύμφωνα με τις προτάσεις 1, 3 και καλό σύμφωνα με την πρόταση 2.

5.5.6 Id λεκάνης 02053800, South Fork Roanoke river

Η συμπεριφορά του ποταμού αυτόν μοιάζει αρκετά με του ποταμού λεκάνης 1 (Potts creek) καθώς πρόκειται για απορροές σχετικά χαμηλές οι οποίες όμως παρουσιάζουν πολλές αιχμές. Από τη σύγκριση θα φανεί πως και η συμπεριφορά των μοντέλων παρουσιάζει μεγάλη ομοιότητα σε σχέση με τα μοντέλα του Potts creek.

Στον ορίζοντα 10 ημερών το μοντέλο ARMA(2,2) σημειώνει το υψηλότερο σκορ με NSE = 0.32 και ANSE = 0.27. Τα σκορ αυτά δεν γίνονται δεκτά από καμιά από τις 3 προτάσεις. Αν και σε σχέση με τη λεκάνη 1 είναι αρκετά πιο αυξημένα (0.10 μεγαλύτερα από το Potts creek). Στον ορίζοντα 5 ημερών τα μοντέλα δεν παρουσιάζουν σημαντική αύξηση των σκορ, το καλύτερο NSE = 0.36 από το ίδιο μοντέλο ARMA(2,2) καθιστά τις τιμές οριακά ικανοποιητικές σύμφωνα με την πρόταση 3 και μη ικανοποιητικές σύμφωνα με τις προτάσεις 1, 2. Παρόμοια μικρή άνοδος σημειώθηκε και στα μοντέλα της λεκάνης 1.

Στον ορίζοντα της 1 ημέρας το υψηλότερο ARIMA είναι το ARMA(2,2) με τιμή NSE = 0.59 και το deep learning πετυχαίνει ελαφρώς πιο αυξημένο NSE = 0.62. Παρεμφερή συμπεριφορά σημειώνεται και στη λεκάνη 1 γεγονός που επαληθεύει πως τα μοντέλα αδυνατούν να πετύχουν καλά NSE σε λεκάνες με τέτοια απορροή. Το σκορ του deep learning θεωρείται ικανοποιητικό σύμφωνα με τις προτάσεις 1,3 και καλό σύμφωνα με την πρόταση 2. Το σκορ του ARMA(2,2) θεωρείται ικανοποιητικό και στις 3 προτάσεις.

5.5.7 Id λεκάνης 02059500, Goose creek

Ο ποταμός αυτός χαρακτηρίζεται από μέσες τιμές απορροής με μικρά αιχμές που οριακά ξεπερνούν τα 5,000 m³/s με εξαίρεση ένα άλμα που ξεπερνά τα 25,000 m³/s.

Στον ορίζοντα των 10 ημερών το ARMA(2,2) σημειώνει NSE = 0.27 και ANSE = 0.24. Τα σκορ αυτά είναι υψηλά σε σχέση με άλλα αντίστοιχα σκορ ορίζοντα 10 ημερών αλλά δεν είναι ικανοποιητικά.

Στον ορίζοντα πρόγνωσης 5 ημερών το ARMA(2,2) συνέχισε να έχει τις καλύτερες επιδόσεις με NSE = 0.3 και ANSE = 0.27. Η άνοδος που σημειώνει είναι αρκετά μικρή δεν μπορεί να θεωρηθεί ικανοποιητικό μοντέλο.

Τέλος, στον ορίζοντα πρόγνωσης 1 ημέρας το ARMA(2,2) συνεχίζει να είναι πρώτο από τα κλασσικά μοντέλα με NSE = 0.55. Το deep learning σημειώνει NSE = 0.61 αυξημένο δηλαδή

κατά 0.06 από το ARMA(2,2). Θεωρείται καλό σύμφωνα με την πρόταση 2 και ικανοποιητικό σύμφωνα με τις προτάσεις 1,3.

5.5.8 Id λεκάνης 02069700, South Mayo river

Ο ποταμός αυτός χαρακτηρίζεται από χαμηλές σχετικά τιμές απορροής με μόνο ένα άλμα να σημειώνει τιμή άνω των 5000 m³/s. Οι αιχμές μεσαίου μεγέθους που έχουν τιμές άνω των 2000 m³/s είναι κάτω από 10.

Στον ορίζοντα πρόγνωσης 10 ημερών και πάλι το ARMA(2,2) σημειώνει το καλύτερο σκορ με NSE = 0.17 και ANSE = 0.22. Οι τιμές αυτές είναι αρκετά μακριά από τις ικανοποιητικές.

Στον ορίζοντα πρόγνωσης 5 ημερών και πάλι το ARMA(2,2) κατέχει την πρώτη θέση με NSE = 0.2 και ANSE = 0.31. Η μεγάλη αυτή διαφορά δείχνει πως το μοντέλο υστερεί στις υψηλές τιμές απορροής έναντι των χαμηλότερων.

Στον ορίζοντα πρόγνωσης 1 ημέρας το ARMA(2,2) συνεχίζει να πετυχαίνει τα υψηλότερα σκορ με NSE = 0.37 και ANSE = 0.48. Αυτή η αισθητή διαφορά τιμής δείχνει πως δεν μπορεί να εκτιμήσει σωστά τις υψηλές απορροές σε αντίθεση με τις χαμηλές όπου εκεί έχει σχετική επιτυχία. Το ARMA(2,2) κρίνεται ικανοποιητικό μόνο σε 1 από τις 3 προτάσεις. Το deep learning μοντέλο που αναπτύχθηκε σημείωσε NSE = 0.39 οριακά γίνεται αποδεκτό και από την 2^η πρόταση.

5.5.9 Id λεκάνης 02143000, Henry Fork, near Henry river

Ο ποταμός αυτός αν και μικρός παρουσιάζει συχνά αιχμές στην απορροή με αρκετές τιμές να ξεπερνούν τα 2000 m³/s. Στον ορίζοντα των 10 ημερών τα σκορ των μοντέλων είναι ιδιαίτερα χαμηλά στο NSE = 0.11 που σημειώνεται από τα ARMA(1,2), ARMA(2,2), ARMA(2,1) μοντέλα. Είναι προφανές πως δεν είναι ικανοποιητικά.

Στον ορίζοντα των 5 ημερών τα ίδια μοντέλα ARMA κατέχουν την πρώτη θέση με NSE = 0.12. Το σκορ αυτό είναι αρκετά χαμηλό, σημείωσε αύξηση μόλις 0.01 βαθμό από το προηγούμενο. Τα μοντέλα δεν μπορούν να προβλέψουν καθόλου τις αιχμές των τιμών πιθανότατα διότι η περίοδος στην οποία γίνονται είναι τόσο σύντομη που το μοντέλο δεν προλαβαίνει να προλάβει αύξηση στην ανανέωση του, παραλαμβάνει απευθείας αυξομείωση.

Στον ορίζοντα της 1 ημέρας το μοντέλο ARMA(2,1) σημειώνει ελάχιστα καλύτερα σκορ από το ARMA(2,2) με στρογγυλοποίηση και τα 2 μοντέλα δίνουν NSE = 0.31 και ANSE = 0.42. Σημειώνει μεν αύξηση αλλά όχι αρκετά μεγάλη για να θεωρηθεί ικανοποιητικό μοντέλο όσον αφορά τον NSE. Από την άλλη το μοντέλο deep learning σημειώνει NSE = 0.36 και ANSE = 0.52 τιμές που το καθιστούν ικανοποιητικό σε 1 από τις 3 προτάσεις.

5.5.10 Id λεκάνης 02350900, Kinchafoone creek

Ο ποταμός Kinchafoone creek είναι ο μεγαλύτερος σε μήκος ποταμός που μελετά η παρούσα εργασία. Σημειώνει κάποιες αιχμές μεσαίου μεγέθους αλλά σε γενικές γραμμές οι τιμές του είναι χαμηλές.

Στον ορίζοντα 10 ημερών την πρώτη θέση για τον δείκτη NSE έχει το μοντέλο AR(2) με $NSE = 0.47$ η τιμή αυτή είναι η μεγαλύτερη τιμή που έχει σημειωθεί σε ορίζοντα πρόγνωσης 10 ημερών σε όλες τις λεκάνες. Είναι ικανοποιητική σε 2 από τις 3 προτάσεις του υποκεφαλαίου 4.4. Η εξάρτηση της εκάστοτε τιμής της παρούσας χρονοσειράς υποδεικνύεται και από το διάγραμμα αυτοσυσχέτισης όπου και οι 2 πρώτες τιμές είχαν αυτο συσχέτιση μεγαλύτερη από 0.75. Στον ορίζοντα των 5 ημερών πάλι το AR(2) σημειώνει την μεγαλύτερη τιμή $NSE=0.68$. Το σκορ αυτό κρίνεται ως ικανοποιητικό σύμφωνα με τις προτάσεις 1,3 και καλό σύμφωνα με την πρόταση 2.

Στον ορίζοντα 1 ημέρας τα μοντέλα ARIMA σημειώνουν μεγαλύτερο σκορ από το μοντέλο deep learning και οι δύο μέθοδοι όμως έχουν δείκτες NSE που καθιστούν τα μοντέλα εξαιρετικά και στις 3 προτάσεις.

6 Συμπεράσματα

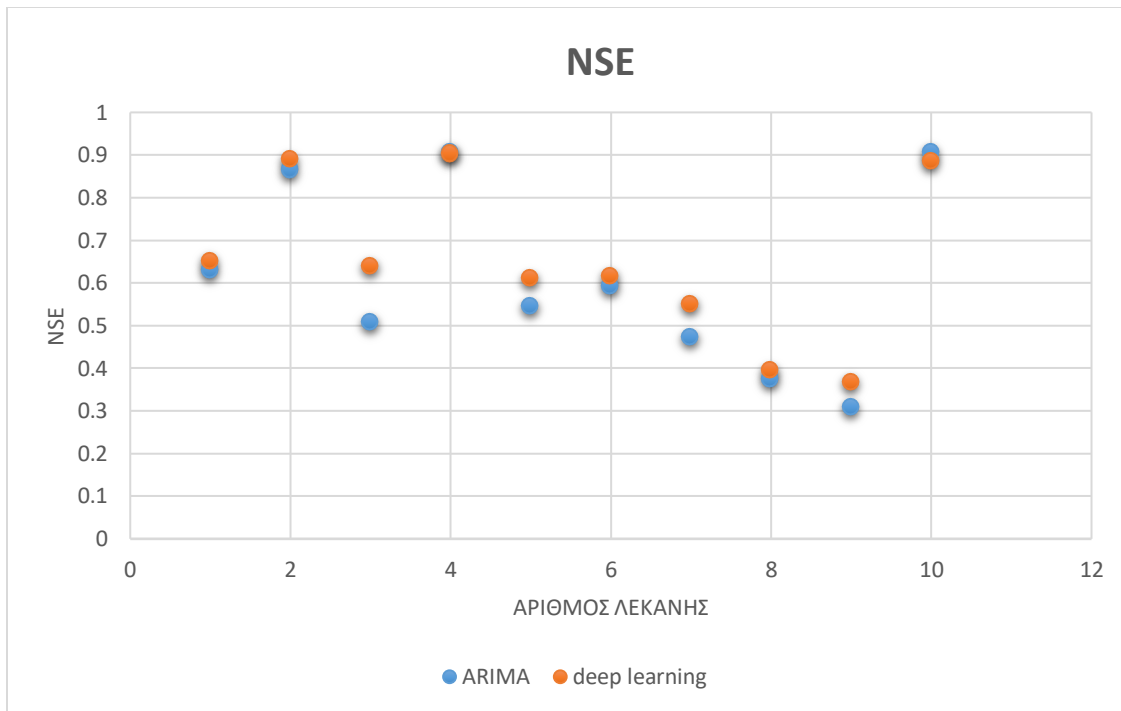
6.1 Σύγκριση ARIMA και εκθετικής ομαλοποίησης

Στις περιπτώσεις πρόγνωσης τα μοντέλα ARIMA ξεπέρασαν τα μοντέλα εκθετικής ομαλοποίησης. Στους ορίζοντες 10 και 5 ημερών η εκθετική ομαλοποίηση κρίθηκε μη ικανοποιητική και μάλιστα αρκετές φορές ήταν χειρότερη εκτιμήτρια και από τη μέση τιμή (σημείωνε $NSE < 0$). Στον ορίζοντα 1 ημέρας τα μοντέλα εκθετικής ομαλοποίησης ήταν και πάλι πιο άστοχα από τα μοντέλα ARIMA και τις περισσότερες φορές συνέχιζαν να μην θεωρούνται ικανοποιητικά από την πλειοψηφία των προτάσεων. Από τα παραπάνω είναι σαφές πως προτιμάται το ARIMA σε κάθε περίπτωση έναντι της εκθετικής ομαλοποίησης αν γίνεται λόγος για τον δείκτη NSE. Αξίζει να αναφερθεί όμως ότι σε ορισμένες περιπτώσεις τα μοντέλα εκθετικής ομαλοποίησης σημειώνουν υψηλότερο σκορ στον ANSE.

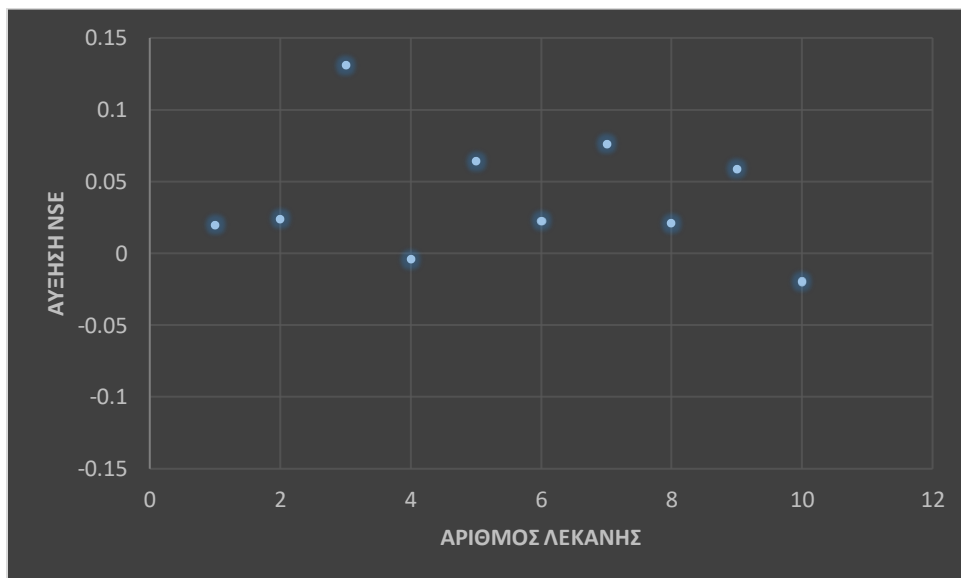
Αξίζει επίσης να γίνει αναφορά σε ορισμένα μοντέλα που διακρίθηκαν. Πρέπει αρχικά να αναφερθεί το μοντέλο ARMA(2,2) καθώς αυτό το μοντέλο σημείωσε τα καλύτερα σκορ στις περισσότερες περιπτώσεις. Στα τεστ επιλογής παραμέτρων p , q , d φάνηκε ότι οι χρονοσειρές είναι στάσιμες άρα δε χρειάζεται διαφοροποίηση και πως πολλοί όροι αξίζει να συμμετέχουν στο μοντέλο. Έτσι το ARMA(2,2) ήταν το μοντέλο με τα μέγιστα p και q χωρίς διαφοροποίηση. Αξίζει επίσης να αναφερθεί και το ARIMA(1,1,2) που στη λεκάνη 2, Tuckahoe river, είναι το μόνο μοντέλο της λεκάνης που πετυχαίνει σκορ ικανοποιητικό σύμφωνα με την πρόταση 1. Το μοντέλο AR(2) στην λεκάνη 10, Kinchafoone creek, ήταν αυτό που ενώ είναι απλό σημείωσε το μεγαλύτερο σκορ στους ορίζοντες πρόγνωσης 5 και 10 ημερών. Τέλος, το μοντέλο απλής εκθετικής ομαλοποίησης συμπεριφέρθηκε καλύτερα από τα υπόλοιπα μοντέλα εκθετικής ομαλοποίησης που χρησιμοποιήθηκαν.

6.2 Σύγκριση ARIMA και deep learning στον ορίζοντα πρόγνωσης 1 ημέρας

Το μοντέλο deep learning στις περισσότερες περιπτώσεις ξεπερνά το καλύτερο μοντέλο που προκύπτει από τις μεθόδους ARIMA. Στη συνέχεια παρατίθενται διαγράμματα που δείχνουν τη διαφορά στα σκορ NSE που σημείωσε το καλύτερο μοντέλο ARIMA και το deep learning.

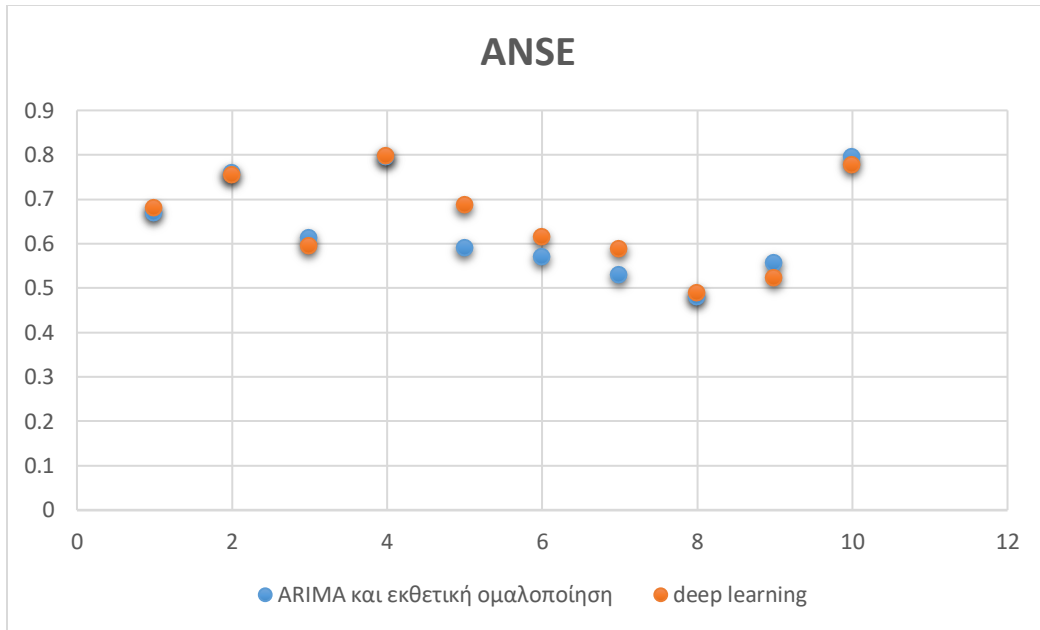


Εικόνα 128 Scatter plot σκορ NSE του καλύτερου μοντέλου ARIMA και deep learning

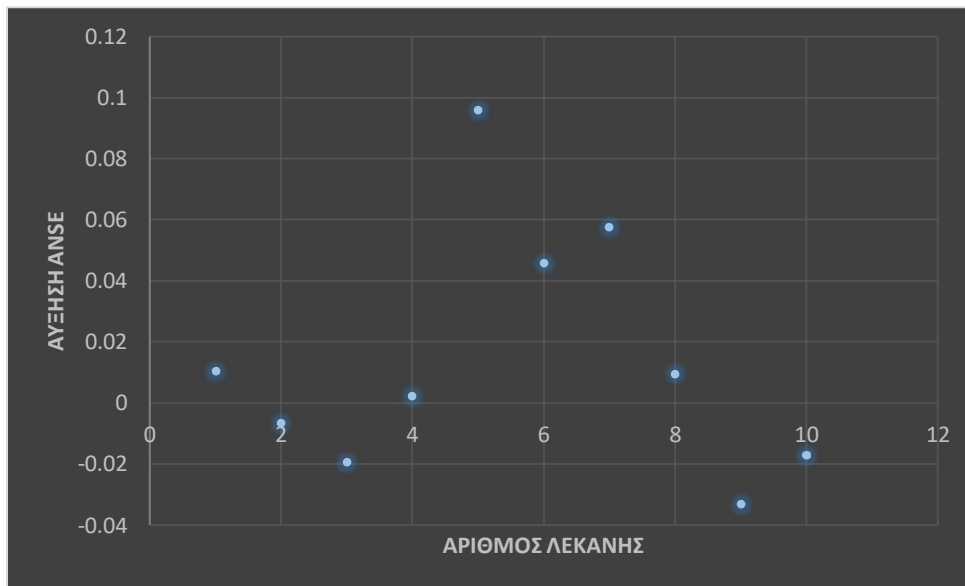


Εικόνα 129 Scatter plot που απεικονίζει την αύξηση του δείκτη NSE που προσφέρει το deep learning σε κάθε λεκάνη., προέκυψε από την αλγεβρική διαφορά NSE deep learning και NSE ARIMA

Στη συνέχεια παρατίθενται ακριβώς τα ίδια διαγράμματα για τον δείκτη ANSE και σε αυτή την περίπτωση στη σύγκριση θα συμμετέχουν και τα μοντέλα εκθετικής ομαλοποίησης.



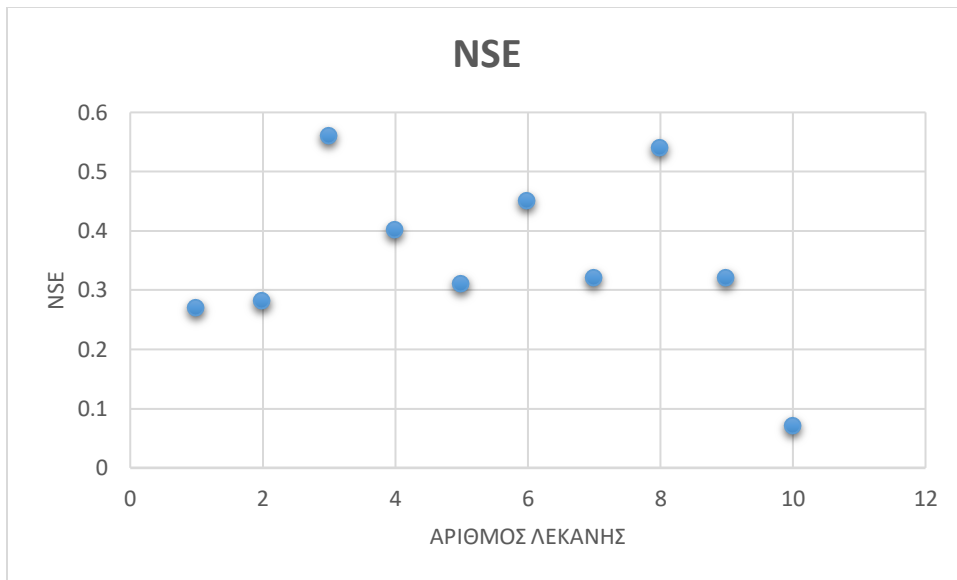
Εικόνα 130 Scatter plot σκορ ANSE του καλύτερου μοντέλου ARIMA, εκθετικής ομαλοποίησης και deep learning



Εικόνα 131 Scatter plot που απεικονίζει την αύξηση του δείκτη NSE που προσφέρει το deep learning σε κάθε λεκάνη, προέκυψε από την αλγεβρική διαφορά NSE deep learning και NSE κλασσικών μεθόδων

6.3 Το μοντέλο deep learning προσομοίωσης λεκάνης

Αυτο το μοντέλο παρουσιάζει τιμές NSE που το καθιστούν ικανοποιητικό έως καλό σε κάποιες περιπτώσεις και σε άλλες μη ικανοποιητικό. Σε 4 από τις 10 περιπτώσεις σημειώνει σκορ ικανοποιητικό σε τουλάχιστον μια από τις προτάσεις. Ο μέσος όρος είναι $NSE = 0.369$ γεγονός που το καθιστά ικανοποιητικό σύμφωνα με την πρόταση 3. Στη συνέχεια παρατίθεται το διάγραμμα με τα NSE σκορ για κάθε λεκάνη.



Εικόνα 132 Scatter plot των τιμών NSE για κάθε λεκάνη στο μοντέλο προσομοίωσης λεκάνης

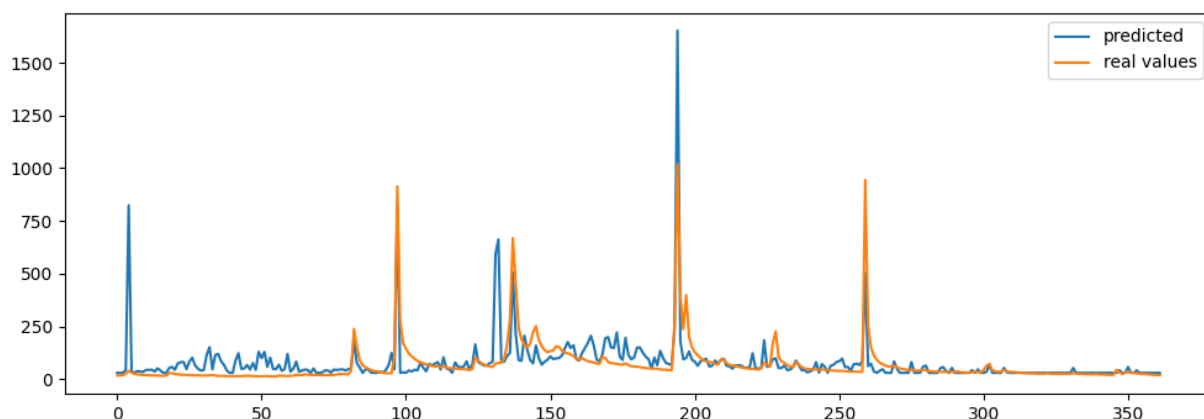
7 Προέκταση των μεθόδων deep learning και προτάσεις για μελλοντική έρευνα

Από τα παραπάνω αποτελέσματα προκύπτει ότι οι μέθοδοι βαθιάς μηχανικής μάθησης έχουν εφαρμογή και μπορούν να βοηθήσουν αρκετά σε μελέτες υδρολογίας. Στο κεφάλαιο αυτό θα γίνει αναφορά σε μια προέκταση του deep learning προσομοίωσης λεκάνης που έγινε στην παρούσα διπλωματική εργασία σε μικρή κλίμακα και θα προταθεί και μια άλλη αρχιτεκτονική νευρωνικού δικτύου που ίσως να σημειώνει μεγαλύτερα σκορ.

7.1 Επέκταση της μεθόδου deep learning προσομοίωσης λεκάνης

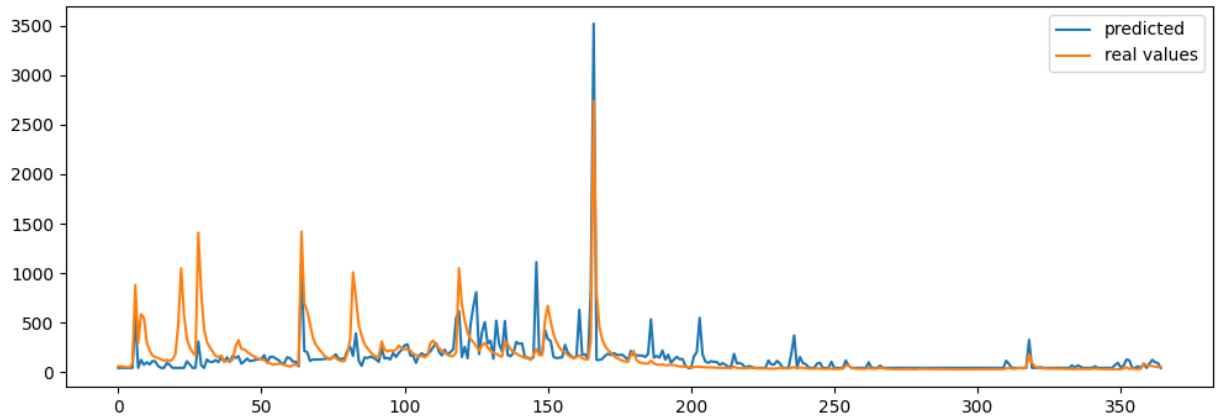
Κατασκευάστηκε ένα νευρωνικό δίκτυο που εκπαιδεύτηκε και στις 10 λεκάνες που χρησιμοποιήθηκαν στην παρούσα εργασία. Ο στόχος ήταν το μοντέλο να λάβει πληροφορίες από πολλές λεκάνες με διαφορετικά δεδομένα (έκταση, βροχόπτωση, απορροή, θερμοκρασία). Μετά την εκπαίδευση τα βάρη του μοντέλου αποθηκεύτηκαν. Δημιουργήθηκε ένα νέο νευρωνικό δίκτυο ίδιας αρχιτεκτονικής με το προαναφερθέν σε αυτό τοποθετήθηκαν τα βάρη που είχαν αποθηκευτεί. Αυτή η διαδικασία είναι γνωστή και ως transfer learning όταν βάρη από ένα άλλο νευρωνικό τοποθετούνται σε ένα νέο νευρωνικό. Το νέο νευρωνικό δίκτυο εκπαιδεύτηκε σε νέες λεκάνες όχι με δεδομένα 30 ετών όπως έγιναν στα προηγούμενα μοντέλα αλλά με δεδομένα μόλις 2 ετών. Με αυτό τον τρόπο κατασκευάστηκε ένα μοντέλο που δε θα χρειάζεται δεκαετίες δεδομένων αλλά λίγα δεδομένα, προσπαθώντας έτσι να λύσει το πρόβλημα της έλλειψης δεδομένων που υπάρχει σε αρκετές λεκάνες στη χώρα μας. Στη συνέχεια παρατίθενται τα διαγράμματα για όσες λεκάνες εφαρμόστηκε πιλοτικά και έστειλαν αυτό το μοντέλο.

Λεκάνη με ID 01632900, Smith Creek near new market



Το νευρωνικό δίκτυο σημείωσε $NSE = 0.29$ και $ANSE = 0.18$

Λεκάνη με ID 02015700, Bullpasture river



Το νευρωνικό δίκτυο σημείωσε $NSE = 0.42$ και $ANSE = 0.31$.

Τα αποτελέσματα είναι αρκετά ικανοποιητικά καθώς στην πρώτη περίπτωση είναι αρκετά κοντά στην οριακή τιμή του $NSE = 0.36$ και στην δεύτερη πετυχαίνει σκορ NSE που το καθιστά ικανοποιητικό σε 2 από τις 3 έρευνες. Προκύπτει από τα αποτελέσματα ότι πρέπει να ληφθούν υπόψη και άλλα δεδομένα καθώς σημειώνονται αστοχίες όπου το μοντέλο εκτιμά πλημμυρικές παροχές οι οποίες δεν συμβαίνουν. Παρά τις αστοχίες όμως μπορεί να εκτιμήσει αρκετά καλά τις αιχμές της παροχής τις περισσότερες φορές.

7.2 Μελλοντικές εναλλακτικές προτάσεις για νευρωνικά δίκτυα

Η πρώτη επιλογή είναι να χρησιμοποιηθούν μοντέλα Long Short Term Memory αυτά τα μοντέλα έχουν τη δυνατότητα να κρατούν τις πληροφορίες από παλαιότερα παραδείγματα εκπαίδευσης και να συμμετέχουν αυτές οι πληροφορίες ενεργά στο να αλλάζουν τα βάρη τους. Μια άλλη επιλογή είναι να ενωθούν τα δεδομένα εισόδου που χρησιμοποιήθηκαν στα δύο νευρωνικά δίκτυα της παρούσας εργασίας. Αυτό σημαίνει να κατασκευαστεί ένα νευρωνικό δίκτυο τύπου κυλιόμενο παράθυρο που θα λαμβάνει επιπλέον σαν δεδομένα εισόδου, θερμοκρασία, βροχή, ημερομηνία και άλλα. Θα μπορούσε τέλος να διανθηστεί με περισσότερα δεδομένα εισόδου το μοντέλο προσομοίωσης της λεκάνης όπως να εισαχθούν επιπλέον η γεωλογία και η κάλυψη της κάθε λεκάνης.

8 Βιβλιογραφία

- *, L. F., Wu, W., Wang, Q., & Hakala, K. (2022). A hybrid framework for short-term irrigation demand forecasting. doi:10.1016/j.agwat.2022.107861
- (2023). Ανάκτηση Φεβρουάριος 10, 2024, από Υπουργείο Περιβάλλοντος και Ενέργειας: <http://www.opengov.gr/minenv/>
- Aris, R. (1978). *Mathematical Modelling Techniques*. Dover Publications.
- Boyle, G. (2012). *Renewable Energy: Power for a Sustainable Future*. Sage.
- Brockwell, P. J. (2009). *Time Series: Theory and Methods*. Springer. Ανάκτηση από <https://www.jstor.org/stable/43754558>
- Brownlee, J. (2018). *Deep Learning for time series forecasting*.
- Chris Seijger, P. H. (2023). How do societies reform their agricultural water management towards new priorities for water , agriculture and the environment. *Agricultural Water Management*. doi:<https://doi.org/10.1016/j.agwat.2022.108104>
- Costa, M. (2019). Time series forecasting using Holt-Winters exponential smoothing: An application to economic data. doi:10.1063/1.5137999
- D, H. (1994). *Time series analysis*. Princeton University Press.
- Daniel Moriasi, J. A. (2007). MODEL EVALUATION GUIDELINES FOR SYSTEMATIC QUANTIFICATION OF ACCURACY IN WATERSHED SIMULATIONS' . *Transactions of the ASABE*.
- DAYMET. (2023, Μάρτιος). Ανάκτηση από <https://daymet.ornl.gov/>
- Dickey, D. A., & Fuller, W. A. (1979). "Distribution of the Estimators for Autoregressive Time Series with a Unit Root". *Journal of the American Statistical Association*. doi:doi:10.1080/01621459.1979.10482531
- Dmitrieva, K. (2015). Forecasting of a hydropower plant.
- Dolado, J. J., Jenkinson, T., & Sosvilla-Rivero, S. (1990). Cointegration and Unit Roots. doi:doi:10.1111/j.1467-6419.1990.tb00088
- EUR-Lex. (2018). Ανάκτηση Ιανουάριος 2024, από <https://eur-lex.europa.eu/EL/legal-content/summary/renewable-energy.html>
- Fuller, W. A. (1995). *Introduction to Statistical Time Series, 2nd Edition* (2nd εκδ.).
- George E. P. Box, G. M. (2015). *Time Series Analysis: Forecasting and Control, 5th Edition*. Wiley.
- Gershenfeld, N. A. (1999). *The Nature of mathematical Modeling*. Cambridge University Press.

- Jinbo Qin¹, J. L. (2017). Simulating and Predicting of Hydrological Time Series Based on TensorFlow Deep Learning. doi:DOI: 10.15244/pjoes/81557
- K.W. Hipel, A. M. (1994). *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier.
- Kavya, M. (2023). Short term water demand forecast modelling using artificial intelligence for. *science direct*, 22. doi:<https://doi.org/10.1016/j.scs.2023.104610>
- Kratzert, F. N. (2023, Ιανουάριος 31). Caravan - A global community dataset for large-sample hydrology. doi:<https://doi.org/10.1038/s41597-023-01975-w>
- Lazzeri, F. (2012). *Machine Learning for Time Series Forecasting with Python*. John Wiley & Sons Inc.
- Lee, E. A. (2009). The problem with Threads.
- Li Shen¹, Z. W. (2021). Determining the Rolling Window Size of Deep Neural Network Based Models on Time Series Forecasting. *Journal of Physics*. doi:DOI 10.1088/1742-6596/2078/1/012011
- Maximum Likelihood Estimation.statsmodels*. (2023). Ανάκτηση Ιουνιος 10, 2023, από Statsmodels: https://www.statsmodels.org/devel/examples/notebooks/generated/generic_mle.html
- MODIS Land Cover Type/Dynamics.nasa*. (2022). Ανάκτηση Δεκέμβρης 2023, από <https://modis.gsfc.nasa.gov/about/>
- Nash, J. a. (1970). River Flow Forecasting through Conceptual Model. *Journal of Hydrology*. doi:[http://dx.doi.org/10.1016/0022-1694\(70\)90255-6](http://dx.doi.org/10.1016/0022-1694(70)90255-6)
- NCAR UCAR*. (2022). Ανάκτηση Φεβρουάριος 2023, από <https://gdex.ucar.edu/dataset/camels/file.html>
- Ng, A. (2018). *Machine Learning Yearning*.
- Oxford Reference*. (n.d.). Ανάκτηση Ιανουάριος 2024, από <https://www.oxfordreference.com/display/10.1093/oi/authority.20110803095426960>
- P.Kraus. (2005). Advances in geosciences comparison of different efficient criteria for hydro model assessment. doi:doi.org/10.5194/adgeo-5-89-2005
- Pankratz, A. (1983). *Forecasting with Univariate Box - Jenkins Models: Concepts and Cases*. Wiley. doi:10.1002/9780470316566
- Rob J. Hyndman, A. B. (2008). *Forecasting with Exponential Smoothing*. Springer-Verlag Berlin and Heidelberg GmbH & Co. KG.
- Safaraliev, M. (2022). *Medium-term forecasting of power generation by hydropower plants*. Guilin, China. doi:<https://doi.org/10.1016/j.egyr.2022.09.164>

- Skaggs, R. (2012). DRAINMOD: model use, calibration, and validation. *Transactions of the ASABE*. doi:DOI:10.13031/2013.42259
- statsmodels. (2023). Ανάκτηση Φεβρουάριος 2023, από <https://www.statsmodels.org/dev/generated/statsmodels.tsa.arima.model.ARIMA.html>
- Tensorflow/time_series. (n.d.). Ανάκτηση Φεβρουάριος 10, 2023, από TensorFlowCore: https://www.tensorflow.org/tutorials/structured_data/time_series
- U. A. Ibrahim, S. D. (2022, November). Comparison of WEAP and SWAT models for streamflow prediction in the Hadejia-Nguru Wetlands, Nigeria. doi:<https://doi.org/10.1007/s40808-022-01435-1>
- USGS. (2023). Ανάκτηση Μαρτιος 10, 2023, από <https://www.usgs.gov/>
- W.Trask, A. (2019). *Grooking Deep Learning*. Manning.
- Enexgroup.gr. (2023). Ανάκτηση Μάρτιος 15, 2023, από <https://www.enexgroup.gr>
- Σειμανίδη. (2022, 6 25). *Τι είναι και πως λειτουργεί το Χρηματιστήριο Ενέργειας*. Ανάκτηση από ecopress.gr: <https://ecopress.gr/afieroma-chrimatistirio-energias-pr/>
- Υπουργείο Περιβάλλοντος και Ενέργειας. (2014). Ανάκτηση Φεβρουάριος 9, 2024, από <http://www.opengov.gr/minenv/?p=5684>