



**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ**

**ΔΙΕΡΕΥΝΗΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ ΜΕΣΟΠΡΟΘΕΣΜΗΣ
ΠΡΟΓΝΩΣΗΣ ΤΩΝ ΕΙΣΡΟΩΝ
ΤΑΜΙΕΥΤΗΡΩΝ ΜΕ ΧΡΗΣΗ ΣΤΟΧΑΣΤΙΚΩΝ ΤΕΧΝΙΚΩΝ
ΚΑΙ ΕΡΓΑΛΕΙΩΝ ΜΗΧΑΝΙΚΗΣ
ΜΑΘΗΣΗΣ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
του
Παναγιώτη Παγωτέλη

Επιβλέπων: Επίκουρος Καθηγητής Ανδρέας Ευστρατιάδης

Αθήνα, Νοέμβριος 2024

Περίληψη

Η διαχείριση των υδάτινων πόρων είναι ένα θέμα που έχει απασχολήσει την ανθρωπότητα για πολλούς αιώνες. Στη σύγχρονη εποχή, αυτό συνεπάγεται σε ανάγκη διαχείρισης νερού σε ταμιευτήρες, τεχνητούς ή φυσικούς, με στόχο την πλήρη και αδιάλειπτη κάλυψη των αναγκών ύδρευσης και άρδευσης, αλλά και γενικότερο όλων των ανθρώπινων δραστηριοτήτων που απαιτούν νερό. Ειδικότερα, στην Αττική, τις οποίες οι ανάγκες καλύπτονται από ένα σύστημα τριών ταμιευτήρων, του Ευήνου, του Μόρνου και της Υλίκης, θα ήταν ιδιαίτερα χρήσιμη η δυνατότητα πρόβλεψης των εισροών στους ταμιευτήρες για σύντομα διαστήματα στο μέλλον, προκειμένου να μπορεί να γίνεται καλύτερος προγραμματισμός της διάθεσης του νερού από αυτούς. Στην παρούσα εργασία γίνεται διερεύνηση του κατά πόσο μπορούν να χρησιμοποιηθούν τεχνητά ανατροφοδοτούμενα νευρωνικά δίκτυα, και συγκεκριμένα LSTM (Long Short-Term Memory, Δίκτυα Μακράς Βραχύχρονης Μνήμης), προκειμένου να προβλεφθούν οι εισροές στους ταμιευτήρες για την περίοδο Μαρτίου – Σεπτεμβρίου του υδρολογικού έτους, χρησιμοποιώντας σα δεδομένα για την πρόβλεψη τις βροχοπτώσεις και τις εισροές στις λεκάνες για την περίοδο Οκτωβρίου – Φεβρουαρίου. Τα δεδομένα που χρησιμοποιούνται για τη διαδικασία της εκπαίδευσης των μοντέλων (εκπαίδευση και επικύρωση όσον αφορά τα LSTM) έχουν παραχθεί με συνθετικό τρόπο από ιστορικά δεδομένα για τις λεκάνες, ενώ για την αξιολόγηση των μοντέλων χρησιμοποιούνται τα αυτούσια ιστορικά δεδομένα. Γίνεται σύγκριση των αποτελεσμάτων των προβλέψεων με το απλό γραμμικό μοντέλο, χρησιμοποιώντας μετρικές κατάλληλες για προβλήματα υδρολογικής φύσης, όπως είναι η NSE (Nash-Sutcliffe Efficiency) και η KGE (Kling-Gupta Efficiency). Ταυτόχρονα, σκοπός είναι και η βελτιστοποίηση των υπερπαραμέτρων των μοντέλων για την επίτευξη των καλύτερων αποτελεσμάτων.

Τελικά, προκύπτει πως η χρήση νευρωνικών δικτύων LSTM είναι υποσχόμενη προσέγγιση για την πρόβλεψη των εισροών σε ταμιευτήρες. Μπορούν να χρησιμοποιηθούν είτε ανεξάρτητα είτε ως μέρος υβριδικών μοντέλων, λειτουργώντας ως εργαλεία υποστήριξης αποφάσεων για τη διαχείριση υδατικών πόρων από αρμόδιους φορείς.

Abstract

The management of water resources is an issue that has been of concern to humanity for many centuries. In modern times, this implies the need to manage water in reservoirs, artificial or natural, in order to ensure full and uninterrupted coverage of the needs of water supply and irrigation, but more generally of all human activities that require water. In particular, in Attica, whose needs are covered by a system of three reservoirs, the Evinos, Mornos and Yliki, it would be particularly useful to be able to forecast the inflows into the reservoirs for short periods in the future, in order to be able to better plan the allocation of water from them. In this thesis, we investigate whether artificial feedback neural networks, namely LSTMs (Long Short-Term Memory Networks), can be used to predict reservoir inflows for the March – September period of the hydrological year, using rainfall and basin inflows for the October – February period as forecast data. The data used for the model training process (training and validation in terms of LSTMs) are synthetically generated from historical basin data, while the original historical data are used for model evaluation. The results of the predictions are compared with the simple linear model, using metrics suitable for hydrological problems, such as NSE (Nash-Sutcliffe Efficiency) and KGE (Kling-Gupta Efficiency). At the same time, the aim is to optimize the hyperparameters of the models to achieve the best results.

Ultimately, it is concluded that LSTM neural networks hold promise for predicting reservoir inflows. They can be used either independently or as part of hybrid models to support decision-making systems for water resource management by relevant agencies.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον κ. Ανδρέα Ευστρατιάδη, Επίκουρο Καθηγητή Ε.Μ.Π, που με εμπιστεύτηκε και με καθοδήγησε κατά τη διάρκεια της εκπόνησης της παρούσας εργασίας. Ήταν πάντα διαθέσιμος για να με συμβουλέψει και να μου προσφέρει κατευθύνσεις, ιδέες και εποικοδομητική κριτική, πράγματα απαραίτητα για τη βελτίωση ενός ανθρώπου, μίας προσπάθειας, αλλά και ενός νευρωνικού δικτύου. Καθ' όλη τη διάρκεια της εκπόνησης της εργασίας, είχε μία θετική αντίληψη και έναν ενθουσιασμό, τα οποία με τη σειρά τους αποτέλεσαν για εμένα μεγάλο κίνητρο. Είμαι ευγνώμων για την ευαισθησία και την άμεση ανταπόκριση που έδειξε σε κάθε πρόβλημα που προέκυψε και την καθαρή οπτική που μπορούσε να προσφέρει για την επίλυση κάθε ζητήματος.

Επίσης, θα ήθελα να ευχαριστήσω τον Δρ. Ιωάννη Τσουκαλά, ο οποίος μου προσέφερε σημαντικές κατευθύνσεις για την χρήση νευρωνικών δικτύων σε τέτοια προβλήματα και μου έδωσε κίνητρο να πειραματιστώ περισσότερο με τα διαθέσιμα εργαλεία της Python και συγκεκριμένα της βιβλιοθήκης Tensorflow / Keras.

Τέλος, θέλω να ευχαριστήσω την οικογένεια και τους φίλους μου, που με στήριξαν σε κάθε βήμα της φοιτητικής μου πορείας και ήταν πάντα εκεί όταν τους χρειαζόμουν.

Παγωτέλης Παναγιώτης

Αθήνα, Νοέμβριος 2024

Πίνακας Περιεχομένων

Περίληψη.....	2
Abstract.....	3
Ευχαριστίες.....	4
Πίνακας Περιεχομένων.....	5
1. Εισαγωγή.....	7
1.1. Αντικείμενο και Στόχοι.....	7
1.2. Διάρθρωση της Εργασίας.....	8
2. Το πρόβλημα της πρόβλεψης.....	9
2.1. Εισαγωγή στο πρόβλημα πρόβλεψης.....	9
2.2. Χρονικός ορίζοντας της πρόβλεψης.....	10
2.3. Προκλήσεις και περιορισμοί στην πρόβλεψη απορροής.....	11
3. Παραγωγή συνθετικών δεδομένων.....	13
3.1. Εισαγωγή.....	13
3.2. Λεκάνη απορροής Ευήνου.....	13
3.3. Λεκάνη απορροής Μόρνου.....	15
3.4. Λεκάνη απορροής Βοιωτικού Κηφισού (Υλίκης).....	18
3.5. Σύνοψη μεθοδολογίας γέννησης συνθετικών χρονοσειρών πρόγνωσης.....	20
4. Θεωρητικό πλαίσιο για τα μοντέλα πρόβλεψης.....	23
4.1. Εισαγωγή στη Μηχανική Μάθηση.....	23
4.2. Τεχνητά Νευρωνικά Δίκτυα.....	23
4.3. Δίκτυα LSTM.....	26
4.4. Εκπαίδευση ενός LSTM.....	27
4.5. Βιβλιοθήκη Keras και υλοποίηση του LSTM στην Python.....	29
5. Προβλέψεις.....	30
5.1. Γραμμικό μοντέλο.....	30
5.2. Γενική αρχιτεκτονική LSTM.....	32
5.3. Προβλέψεις για τη λεκάνη του Ευήνου.....	32
5.4. Προβλέψεις για τη λεκάνη του Μόρνου.....	34
5.5. Προβλέψεις για τη λεκάνη του Βοιωτικού Κηφισού.....	35
5.6. Συνολικές προβλέψεις.....	37
6. Αξιολόγηση των Μοντέλων.....	41

6.1. Μετρικές για την αξιολόγηση των μοντέλων	41
6.2. Αξιολόγηση των μοντέλων για τον Εύηνο	43
6.3. Αξιολόγηση των μοντέλων για τον Μόρνο	44
6.4. Αξιολόγηση των μοντέλων για την Υλίκη	46
6.5. Συνολική αποτίμηση των επιδόσεων των μοντέλων	50
7. Συμπεράσματα και προτάσεις	51
7.1. Συμπεράσματα	51
7.2. Ερευνητικές προτάσεις	52
Βιβλιογραφικές Αναφορές	54
Παράρτημα	56

1. Εισαγωγή

1.1. Αντικείμενο και Στόχοι

Το αντικείμενο της παρούσας εργασίας είναι η ανάπτυξη και αξιολόγηση ενός μοντέλου πρόβλεψης εισροών σε υδροηλεκτρικούς ταμιευτήρες χρησιμοποιώντας Δίκτυα Μακράς Βραχύχρονης Μνήμης (LSTM) και συνθετικά δεδομένα, τα οποία έχουν παραχθεί με τη χρήση στοχαστικών μεθόδων. Η προσέγγιση αυτή βασίζεται στην ικανότητα των LSTM να «μαθαίνουν» πολύπλοκα πρότυπα και εξαρτήσεις σε χρονοσειρές, καθιστώντας τα κατάλληλα δίκτυα για παροχή προβλέψεων σε υδρολογικά προβλήματα. Για την εκπαίδευση των μοντέλων επιλέχθηκαν στοχαστικά δεδομένα προκειμένου να γίνει εμπλουτισμός του πλήθους τους, καθώς τα διαθέσιμα ιστορικά δεδομένα είναι ελάχιστα για την εκπαίδευση τέτοιων μοντέλων. Ταυτόχρονα, η χρήση των συνθετικών δεδομένων προσθέτει μία παραπάνω αβεβαιότητα στα δεδομένα, η οποία δύναται να «μεταφραστεί» σε ένα μοντέλο με μεγαλύτερη ευρωστία, το οποίο ανταποκρίνεται καλύτερα στην πρόκληση της πρόβλεψης δεδομένων του πραγματικού κόσμου. Το πρακτικό ενδιαφέρον της μελέτης έγκειται στη σύγκριση της απόδοσης του LSTM μοντέλου έναντι του απλού γραμμικού μοντέλου, αλλά περισσότερο στη γενικότερη διερεύνηση της χρήσης τέτοιων μοντέλων στην πρόβλεψη εισροών σε ταμιευτήρες. Πρόκειται για προσέγγιση αιχμής στη διαχείριση υδατικών πόρων, πολλά υποσχόμενη για το μέλλον, καθώς η χρήση LSTM σε συνδυασμό με συνθετικά δεδομένα δεν έχει αξιοποιηθεί ευρέως έως σήμερα, ανοίγοντας νέους ορίζοντες στην πρόβλεψη εισροών και στη στρατηγική διαχείρισης ταμιευτήρων, ειδικότερα σε ξηρότερες εποχές, κατά τις οποίες παρατηρούνται φαινόμενα λειψυδρίας.

1.2. Διάρθρωση της Εργασίας

Η παρούσα εργασία διαρθρώνεται σε 7 κεφάλαια.

Στο πρώτο κεφάλαιο γίνεται αναφορά στο αντικείμενο της εργασίας, τους αρχικούς προβληματισμούς που αποτέλεσαν κίνητρο, και στους στόχους που θέλουμε να επιτύχουμε.

Το δεύτερο κεφάλαιο αποτελεί μία εισαγωγή στο πρόβλημα της πρόβλεψης απορροών και παρουσιάζει τις δυσκολίες που προκύπτουν.

Στο τρίτο κεφάλαιο παρουσιάζεται συνοπτικά η διαδικασία της παραγωγής των συνθετικών χρονοσειρών που χρησιμοποιήθηκαν για τις προβλέψεις.

Στο τέταρτο κεφάλαιο γίνεται μία εισαγωγή στην έννοια της Μηχανικής Μάθησης, των νευρωνικών δικτύων αλλά και πιο συγκεκριμένα στα δίκτυα LSTM.

Στο πέμπτο κεφάλαιο παρουσιάζονται τα αποτελέσματα των προβλέψεων που έγιναν με τα διάφορα μοντέλα, και ο τρόπος με τον οποίο αυτά προέκυψαν.

Στο έκτο κεφάλαιο αξιολογούνται με διάφορες μετρικές τα αποτελέσματα των μοντέλων, επιλέγονται τα βέλτιστα μοντέλα και γίνεται προσπάθεια για την εξήγηση των αποτελεσμάτων.

Στο έβδομο κεφάλαιο παρουσιάζονται τα συμπεράσματα της εργασίας και γίνονται κάποιες προτάσεις για μελλοντική επέκταση του ερευνητικού αντικειμένου.

2. Το πρόβλημα της πρόβλεψης

2.1. Εισαγωγή στο πρόβλημα πρόβλεψης

Η πρόβλεψη των απορροών στις υδρολογικές λεκάνες αποτελεί έναν από τους θεμελιώδεις τομείς μελέτης στην Επιστήμη της Διαχείρισης Υδατικών Πόρων, καθορίζοντας σε μεγάλο βαθμό τη στρατηγική διαχείρισης των αποθεμάτων νερού σε μία περιοχή. Το υδάτινο απόθεμα εξαρτάται από εποχιακά φαινόμενα και παρουσιάζει σημαντικές αυξομειώσεις, οπότε η δυνατότητα πρόγνωσης των μελλοντικών απορροών μπορεί να συμβάλει καθοριστικά στη βέλτιστη διαχείριση ενός συστήματος ταμιευτήρων, μέσω συστημάτων υποστήριξης αποφάσεων.

Οι εισροές στους ταμιευτήρες επηρεάζονται από ένα σύνολο παραγόντων, μεταξύ των οποίων είναι οι βροχοπτώσεις, η εποχιακή τήξη του χιονιού (όπου αυτό υπάρχει) και οι κλιματικές μεταβολές, ενώ παρουσιάζουν έντονη εποχικότητα και τυχαιότητα. Οι εισροές και οι βροχοπτώσεις είναι φαινόμενα που δεν ακολουθούν σταθερούς ρυθμούς και επηρεάζονται τόσο από τοπικά όσο και από μακροπρόθεσμα περιβαλλοντικά και κλιματικά μοτίβα, ειδικά σε γεωμορφολογικά / υδρολογικά σύνθετες περιοχές, όπως η Ηπειρωτική Ελλάδα. Στην πράξη, αυτό σημαίνει ότι το υδάτινο δυναμικό που προσφέρεται σε ένα ταμιευτήρα μπορεί να διαφέρει αισθητά από έτος σε έτος, με περιόδους αυξημένης παροχής να διαδέχονται περιόδους ξηρασίας.

Στο πρόβλημα διαχείρισης του νερού για κάλυψη αναγκών ύδρευσης και άρδευσης οι περίοδοι αυξημένης παροχής δεν επηρεάζουν σε τόσο μεγάλο βαθμό, όμως οι περίοδοι ξηρασίας, στις οποίες συγκαταλέγεται και η τωρινή περίοδος που διανύουμε, δύνανται να προκαλέσουν σοβαρά προβλήματα. Ως αποτέλεσμα, οι φορείς διαχείρισης των υδατικών πόρων καλούνται να προβλέπουν τα μοτίβα εισροών, ώστε να μπορεί να εξισορροπηθεί η ποσότητα που αποθηκεύεται με την ποσότητα που χρησιμοποιείται, εξασφαλίζοντας τη βιώσιμη παροχή νερού.

Στο πλαίσιο της πρόβλεψης των εισροών, η ραγδαία ανάπτυξη των νευρωνικών δικτύων έχει δημιουργήσει νέες προοπτικές. Συγκεκριμένα, τα δίκτυα LSTM έχουν δημιουργηθεί ειδικά για την αντιμετώπιση τέτοιων προβλημάτων. Τα LSTM διακρίνονται για την ικανότητά τους να «θυμούνται» πληροφορίες για μεγάλες χρονικές περιόδους, αναγνωρίζοντας σχέσεις που

διαφεύγουν από τις κλασικές γραμμικές μεθόδους. Σε αντίθεση με τα παραδοσιακά νευρωνικά δίκτυα, τα οποία περιορίζονται στη συνειρμική επεξεργασία των δεδομένων εισόδου – εξόδου, τα LSTM είναι εξοπλισμένα με μονάδες / νευρώνες μνήμης που διατηρούν χρονικές εξαρτήσεις και έχουν αποδειχθεί κατάλληλα για εφαρμογές πρόβλεψης χρονοσειρών, ειδικά σε σύνθετα υδρολογικά συστήματα με δυναμική συμπεριφορά και συχνές διακυμάνσεις.

2.2. Χρονικός ορίζοντας της πρόβλεψης

Ο χρονικός ορίζοντας της πρόβλεψης των εισροών που εξετάζεται σε αυτή την εργασία επικεντρώνεται στην περίοδο μεταξύ του Μαρτίου και του Σεπτεμβρίου, δηλαδή στο δεύτερο μισό του υδρολογικού έτους. Η επιλογή της συγκεκριμένης περιόδου έχει ιδιαίτερη σημασία για τους υδατικούς πόρους, καθώς καλύπτει τη θερινή περίοδο, κατά την οποία παρατηρούνται οι υψηλότερες ανάγκες χρήσης νερού και ταυτόχρονα περιορίζονται οι διαθέσιμες φυσικές εισροές. Παράλληλα, ως δεδομένα πρόβλεψης χρησιμοποιούνται τα δεδομένα που συλλέγονται κατά την περίοδο Οκτωβρίου – Φεβρουαρίου του ίδιου υδρολογικού έτους, τα οποία είναι αρκετά ώστε να δίνουν ικανές προβλέψεις. Υπάρχει η δυνατότητα να γίνει πρόβλεψη για ακόμα περισσότερους μήνες, έχοντας λιγότερα δεδομένα εισόδου, όμως αυτό μειώνει σημαντικά την ποιότητα των αποτελεσμάτων, για αυτό και επιλέγεται ο διαχωρισμός 5 προς 7 μηνών. Ειδικότερα, η πρόβλεψη των εισροών για αυτούς τους μήνες βοηθά στην καλύτερη προετοιμασία για τις αυξημένες ανάγκες άρδευσης, ύδρευσης και άλλων απαιτήσεων, διασφαλίζοντας ταυτόχρονα ότι οι ταμειυτήρες δε θα αντιμετωπίσουν σοβαρά προβλήματα σε κάποια περίοδο παρατεταμένης ξηρασίας.

Η πρόγνωση των εισροών απαιτεί τη χρήση ιστορικών δεδομένων που προέρχονται από τις περιόδους Οκτωβρίου έως Φεβρουαρίου, κατά τις οποίες οι βροχοπτώσεις και οι εισροές είναι συνήθως υψηλότερες. Αυτά τα δεδομένα προσφέρουν πληροφορίες σχετικά με την υδρολογική κατάσταση του εξαμήνου που προηγείται της περιόδου πρόβλεψης, βοηθώντας το μοντέλο να εκτιμήσει πόσο πιθανό είναι να συνεχιστούν ή να περιοριστούν οι υδάτινες εισροές κατά τη θερινή περίοδο. Η χρήση των προηγούμενων εξαμήνων αποτελεί βασικό στοιχείο για τα LSTM, καθώς τα δεδομένα αυτά είναι πιθανό να εμπεριέχουν τις απαραίτητες πληροφορίες που χρειάζονται τα μοντέλα για να εντοπίσουν μοτίβα που σχετίζονται με τις μελλοντικές εισροές.

2.3. Προκλήσεις και περιορισμοί στην πρόβλεψη απορροής

Η πρόβλεψη των απορροών λεκανών, ιδιαίτερα μέσω μεθόδων τεχνητής νοημοσύνης όπως τα LSTM (Long Short-Term Memory) δίκτυα, συνοδεύεται από αρκετές δυσκολίες, που σχετίζονται κυρίως με την αβεβαιότητα των υδρολογικών και κλιματικών παραγόντων, τα διαθέσιμα δεδομένα, την ακρίβεια των προβλέψεων, και τη δυνατότητα γενίκευσης των αποτελεσμάτων. Οι προκλήσεις αυτές επηρεάζουν την αποτελεσματικότητα της πρόβλεψης και την αξιοπιστία των μοντέλων, καθιστώντας σημαντική την κατανόηση και διαχείριση αυτών των περιορισμών για την επιτυχή εφαρμογή των LSTM δικτύων.

Τα υδρολογικά συστήματα είναι εξαιρετικά δυναμικά και επηρεάζονται από πολλαπλούς παράγοντες, όπως οι καιρικές συνθήκες, η γεωμορφολογία, και η κλιματική περίοδος. Τα συστήματα αυτά παρουσιάζουν πολύπλοκη και έντονα μη γραμμική συμπεριφορά, με σημαντικές αυξομειώσεις στις εισροές στους ταμιευτήρες από περίοδο σε περίοδο. Επομένως, ακόμα και μικρές αλλαγές στις κλιματικές συνθήκες, όπως η διακύμανση της θερμοκρασίας και η ένταση των βροχοπτώσεων, μπορούν να έχουν μεγάλο αντίκτυπο στις εισροές, ενώ οι μεγάλες αποκλίσεις από το συνηθισμένο μοτίβο προκαλούν πρόσθετη δυσκολία στην πρόβλεψη. Τα LSTM δίκτυα είναι κατάλληλα για τη διαχείριση χρονοσειρών με μη γραμμική συμπεριφορά, αλλά δυσκολεύονται όταν οι συνθήκες αποκλίνουν σημαντικά από τα μοτίβα στα δεδομένα εκπαίδευσης. Για παράδειγμα, ακραία φαινόμενα όπως πλημμύρες ή περίοδοι παρατεταμένης ξηρασίας μπορεί να διαταράξουν την προβλεψιμότητα των εισροών.

Ένα από τα πιο κρίσιμα ζητήματα στην πρόβλεψη εισροών αφορά τη διαθεσιμότητα και την ποιότητα των ιστορικών δεδομένων βροχοπτώσεων και εισροών. Οι χρονοσειρές που χρησιμοποιούνται για την εκπαίδευση των LSTM μοντέλων συνήθως παρουσιάζουν ασυνέπειες στις μετρήσεις ή πολύ μικρό μήκος, γεγονός που έχει έντονα αρνητικό αντίκτυπο στις επιδόσεις των μοντέλων. Το ίδιο ισχύει και για τα υδρολογικά δεδομένα. Η συλλογή δεδομένων για τα υδρολογικά συστήματα εξαρτάται συχνά από παραδοσιακές μεθόδους μέτρησης, που δεν καλύπτουν πάντα με ακρίβεια τη χωρική και χρονική μεταβλητότητα των φαινομένων. Ιδανικά, τα δεδομένα θα πρέπει να είναι συνεχώς διαθέσιμα και να καλύπτουν μια ευρεία χρονική περίοδο

για να αποτυπώνουν τα μοτίβα και τις εποχιακές τάσεις. Ωστόσο, αυτό είναι αδύνατο να επιτευχθεί πρακτικά, αφού η μοναδική λεκάνη για την υπάρχουν δεδομένα για μία αρκετά μεγάλη χρονική περίοδο είναι αυτή του ποταμού Νείλου στην Αίγυπτο. Στην Ελλάδα, υπάρχουν λίγα περισσότερα δεδομένα για τη λεκάνη του Βοιωτικού Κηφισού, σε καμία περίπτωση, όμως, δεν είναι αρκετά για την εκπαίδευση σύνθετων μοντέλων.

Για αυτόν τον λόγο, η χρήση συνθετικών δεδομένων για την εκπαίδευση αποτελεί σχεδόν μονόδρομο, καθώς υπάρχει ανάγκη για αρκετά μεγαλύτερο όγκο πληροφορίας προκειμένου να μπορούν να εκπαιδευτούν τα μοντέλα. Το γεγονός ότι τα συνθετικά δεδομένα έχουν παραχθεί με εξελιγμένες στατιστικές μεθόδους επιτρέπει στα μοντέλα να εντοπίζουν αρκετά πιο σύνθετα μοτίβα, με την έλλειψη πραγματικών δεδομένων να είναι εν τέλει μερικώς αντιμετωπίσιμη.

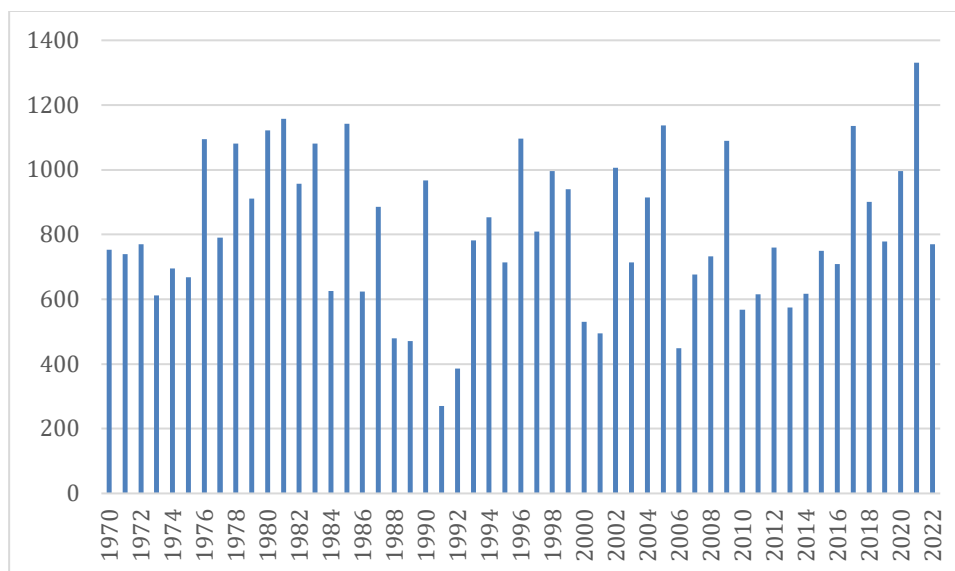
3. Παραγωγή συνθετικών δεδομένων

3.1. Εισαγωγή

Τα συνθετικά δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευση των μοντέλων στο πλαίσιο αυτής της εργασίας έχουν παραχθεί στο πλαίσιο της ερευνητικής εργασίας (Α. Ευστρατιάδης, και Γ.-Κ. Σακκή, Αναλύσεις υδατικών ισοζυγίων και απολογιστική έκθεση υδρολογικού έτους 2022-23, Εκσυγχρονισμός της διαχείρισης του συστήματος των υδατικών πόρων ύδρευσης της Αθήνας – Αναθεώρηση, 30 σελίδες, Τομέας Υδατικών Πόρων και Περιβάλλοντος – Εθνικό Μετσόβιο Πολυτεχνείο, Ιανουάριος 2024). Η παραγωγή των συνθετικών δεδομένων δεν αποτελεί σκοπό της παρούσας εργασίας, αλλά μόνο η αξιοποίησή τους.

3.2. Λεκάνη απορροής Ευήνου

Στον **Πίνακα 3.1** δίνονται τα στατιστικά χαρακτηριστικά του δείγματος βροχόπτωσης, από τον Οκτώβριο του 1970 έως τον Φεβρουάριο του 2023, σε διάφορες χρονικές κλίμακες ενδιαφέροντος (πρώτο πεντάμηνο, επιμέρους και συνολικά, περίοδος Μαρτίου – Σεπτεμβρίου, σύνολο υδρολογικού έτους).

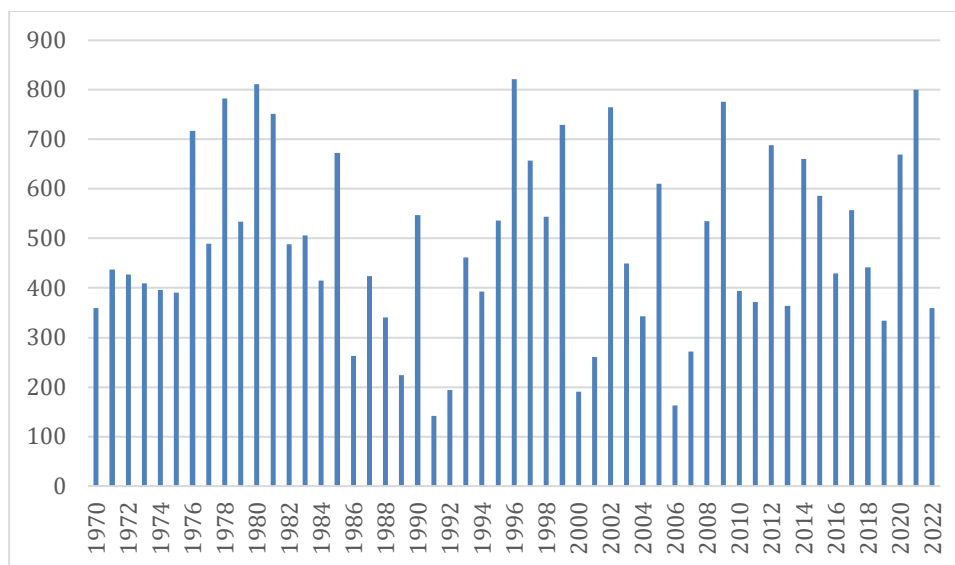


Εικόνα 3.1: Αθροιστική βροχόπτωση λεκάνης Ευήνου κατά το πεντάμηνο Οκτωβρίου – Φεβρουαρίου (mm).

Πίνακας 3.1: Στατιστικά μεγέθη χρονοσειράς βροχόπτωσης λεκάνης Ευήνου για τα υδρολογικά έτη 1970-71 έως 2022-23, για διάφορες χρονικές κλίμακες (mm).

	Οκτ.	Νοε.	Δεκ.	Ιαν.	Φεβ.	Οκτ.- Φεβ.	Μαρ.- Σεπ.	Έτος
Μέση τιμή	102.6	197.4	199.8	154.0	152.1	806.0	403.6	1209.6
Τυπική Απόκλιση	68.8	96.4	122.8	105.6	96.3	230.5	136.0	267.7
Ελάχιστη τιμή	0.3	2.0	2.4	0.0	3.6	270.4	175.8	703.9
Μέγιστη τιμή	296.2	423.6	554.6	485.6	390.7	1331.6	778.0	1783.1

Στον **Πίνακα 3.2** δίνονται τα στατιστικά χαρακτηριστικά του δείγματος απορροών, από τον Οκτώβριο του 1970 έως τον Φεβρουάριο του 2023, σε διάφορες χρονικές κλίμακες ενδιαφέροντος (πρώτο πεντάμηνο, επιμέρους και συνολικά, περίοδος Μαρτίου – Σεπτεμβρίου, σύνολο υδρολογικού έτους).



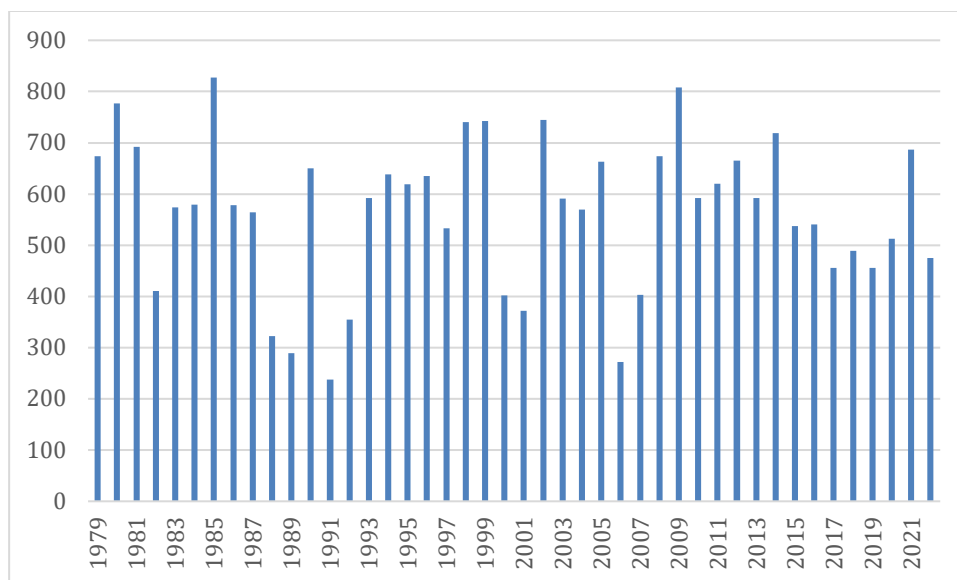
Εικόνα 3.2: Αθροιστική απορροή λεκάνης Ευήνου κατά το πεντάμηνο Οκτωβρίου – Φεβρουαρίου (mm).

Πίνακας 3.2: Στατιστικά μεγέθη χρονοσειράς απορροής λεκάνης Ευήνου για τα υδρολογικά έτη 1970-71 έως 2022-23, για διάφορες χρονικές κλίμακες (mm).

	Οκτ.	Νοε.	Δεκ.	Ιαν.	Φεβ.	Οκτ.- Φεβ.	Μαρ.- Σεπ.	Έτος
Μέση τιμή	24.1	72.0	139.0	124.7	128.6	488.4	318.0	806.4
Τυπική Απόκλιση	18.5	48.6	99.6	77.1	68.4	182.3	93.6	219.6
Ελάχιστη τιμή	5.4	9.6	21.0	16.2	12.3	141.6	97.9	321.7
Μέγιστη τιμή	77.0	220.7	475.0	344.0	284.9	821.8	497.6	1182.6

3.3. Λεκάνη απορροής Μόρνου

Στον **Πίνακα 3.3** δίνονται τα στατιστικά χαρακτηριστικά του δείγματος βροχόπτωσης, από τον Οκτώβριο του 1979 έως τον Φεβρουάριο του 2023, σε διάφορες χρονικές κλίμακες ενδιαφέροντος (πρώτο πεντάμηνο, επιμέρους και συνολικά, περίοδος Μαρτίου – Σεπτεμβρίου, σύνολο υδρολογικού έτους).

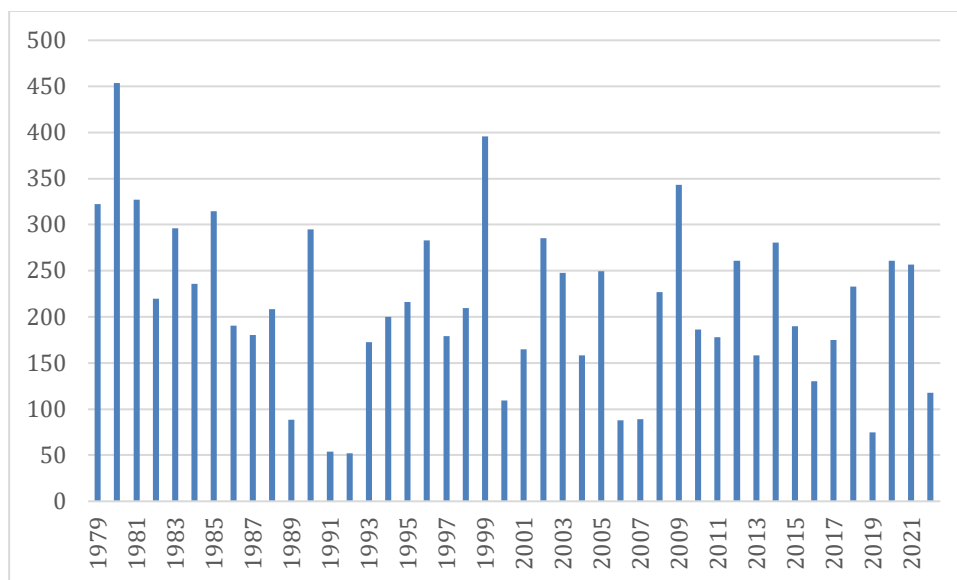


Εικόνα 3.3: Αθροιστική βροχόπτωση λεκάνης Μόρνου κατά το πεντάμηνο Οκτωβρίου – Φεβρουαρίου (mm).

Πίνακας 3.3: Στατιστικά μεγέθη χρονοσειράς βροχόπτωσης λεκάνης Μόρνου για τα υδρολογικά έτη 1979-80 έως 2022-23, για διάφορες χρονικές κλίμακες (mm).

	Οκτ.	Νοε.	Δεκ.	Ιαν.	Φεβ.	Οκτ.- Φεβ.	Μαρ.- Σεπ.	Έτος
Μέση τιμή	87.2	143.8	131.2	110.4	92.7	565.2	249.1	814.3
Τυπική Απόκλιση	58.5	66.3	67.3	75.6	58.3	144.6	84.4	154.1
Ελάχιστη τιμή	1.5	2.1	1.1	0.7	1.7	237.9	56.8	449.3
Μέγιστη τιμή	258.7	368.3	281.5	302.7	232.8	827.3	465.9	1049.8

Στον **Πίνακα 3.4** δίνονται τα στατιστικά χαρακτηριστικά του δείγματος απορροών, από τον Οκτώβριο του 1979 έως τον Φεβρουάριο του 2023, σε διάφορες χρονικές κλίμακες ενδιαφέροντος (πρώτο πεντάμηνο, επιμέρους και συνολικά, περίοδος Μαρτίου – Σεπτεμβρίου, σύνολο υδρολογικού έτους).



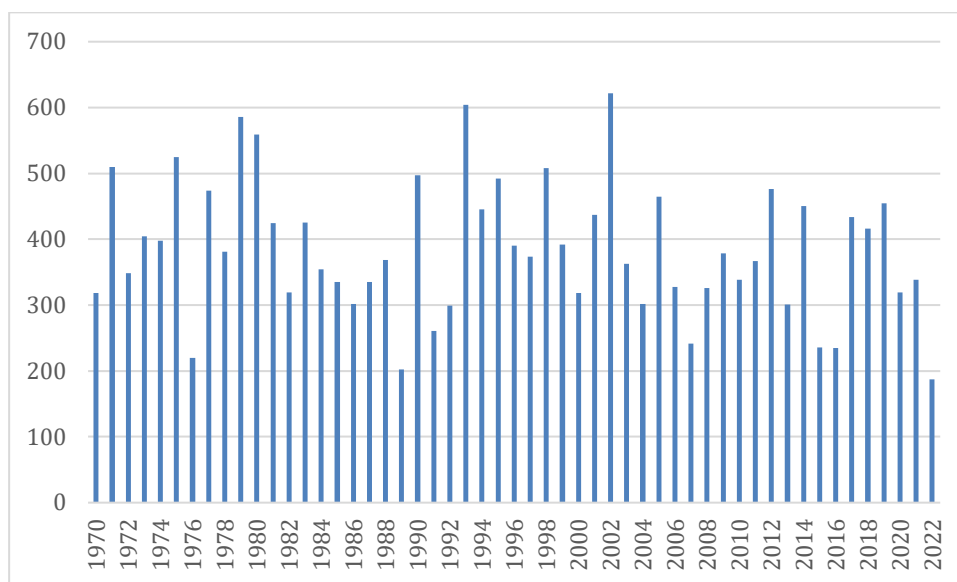
Εικόνα 3.4: Αθροιστική απορροή λεκάνης Μόρνου κατά το πεντάμηνο Οκτωβρίου – Φεβρουαρίου (mm).

Πίνακας 3.4: Στατιστικά μεγέθη χρονοσειράς απορροής λεκάνης Μόρνου για τα υδρολογικά έτη 1979-80 έως 2022-23, για διάφορες χρονικές κλίμακες (mm).

	Οκτ.	Νοε.	Δεκ.	Ιαν.	Φεβ.	Οκτ.- Φεβ.	Μαρ.- Σεπ.	Έτος
Μέση τιμή	14.4	34.6	53.3	54.2	56.2	212.6	197.2	409.9
Τυπική Απόκλιση	9.1	24.4	36.2	34.7	31.1	88.2	56.0	125.6
Ελάχιστη τιμή	1.1	5.0	7.0	5.4	7.1	52.0	66.7	155.3
Μέγιστη τιμή	42.4	108.0	167.6	142.8	141.7	453.6	339.4	710.8

3.4. Λεκάνη απορροής Βοιωτικού Κηφισού (Υλίκης)

Στον Πίνακα 3.5 δίνονται τα στατιστικά χαρακτηριστικά του δείγματος βροχόπτωσης, από τον Οκτώβριο του 1970 έως τον Φεβρουάριο του 2023, σε διάφορες χρονικές κλίμακες ενδιαφέροντος (πρώτο πεντάμηνο, επιμέρους και συνολικά, περίοδος Μαρτίου – Σεπτεμβρίου, σύνολο υδρολογικού έτους).

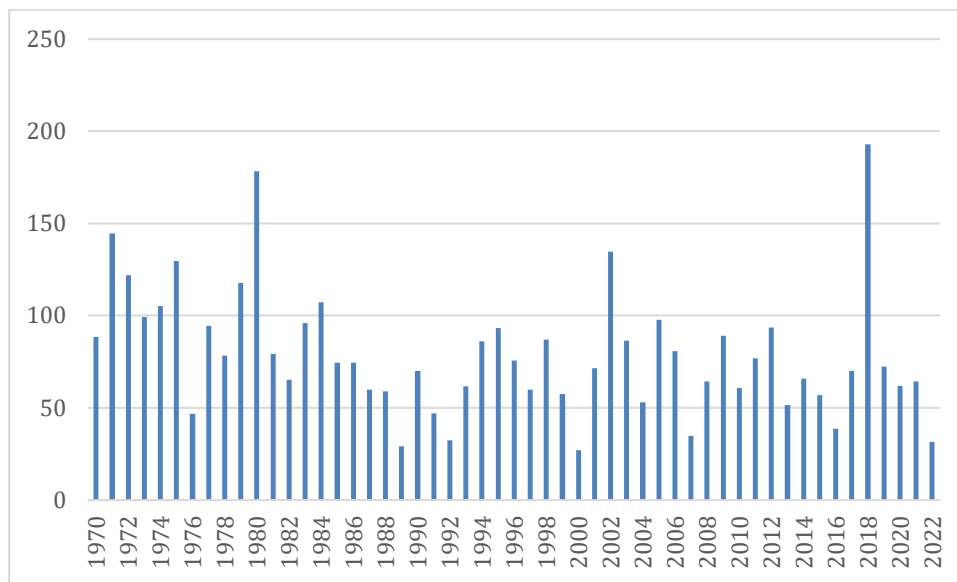


Εικόνα 3.5: Αθροιστική βροχόπτωση λεκάνης Υλίκης κατά το πεντάμηνο Οκτωβρίου-Φεβρουαρίου (mm).

Πίνακας 3.5: Στατιστικά μεγέθη χρονοσειράς βροχόπτωσης λεκάνης Υλίκης για τα υδρολογικά έτη 1970-71 έως 2022-23, για διάφορες χρονικές κλίμακες (mm).

	Οκτ.	Νοε.	Δεκ.	Ιαν.	Φεβ.	Οκτ.- Φεβ.	Μαρ.- Σεπ.	Έτος
Μέση τιμή	59.3	78.8	91.0	80.1	75.4	384.5	198.7	583.2
Τυπική Απόκλιση	44.2	51.7	48.2	47.8	43.4	101.0	65.6	125.5
Ελάχιστη τιμή	0.0	1.2	4.5	3.0	6.4	186.9	62.6	335.9
Μέγιστη τιμή	224.2	249.7	214.2	212.4	187.0	621.8	347.7	855.5

Στον **Πίνακα 3.6** δίνονται τα στατιστικά χαρακτηριστικά του δείγματος απορροών, από τον Οκτώβριο του 1970 έως τον Φεβρουάριο του 2023, σε διάφορες χρονικές κλίμακες ενδιαφέροντος (πρώτο πεντάμηνο, επιμέρους και συνολικά, περίοδος Μαρτίου – Σεπτεμβρίου, σύνολο υδρολογικού έτους).



Εικόνα 3.6: Αθροιστική απορροή λεκάνης Υλίκης κατά το πεντάμηνο Οκτωβρίου-Φεβρουαρίου (mm).

Πίνακας 3.6: Στατιστικά μεγέθη χρονοσειράς απορροής λεκάνης Υλίκης για τα υδρολογικά έτη 1970-71 έως 2022-23, για διάφορες χρονικές κλίμακες (mm).

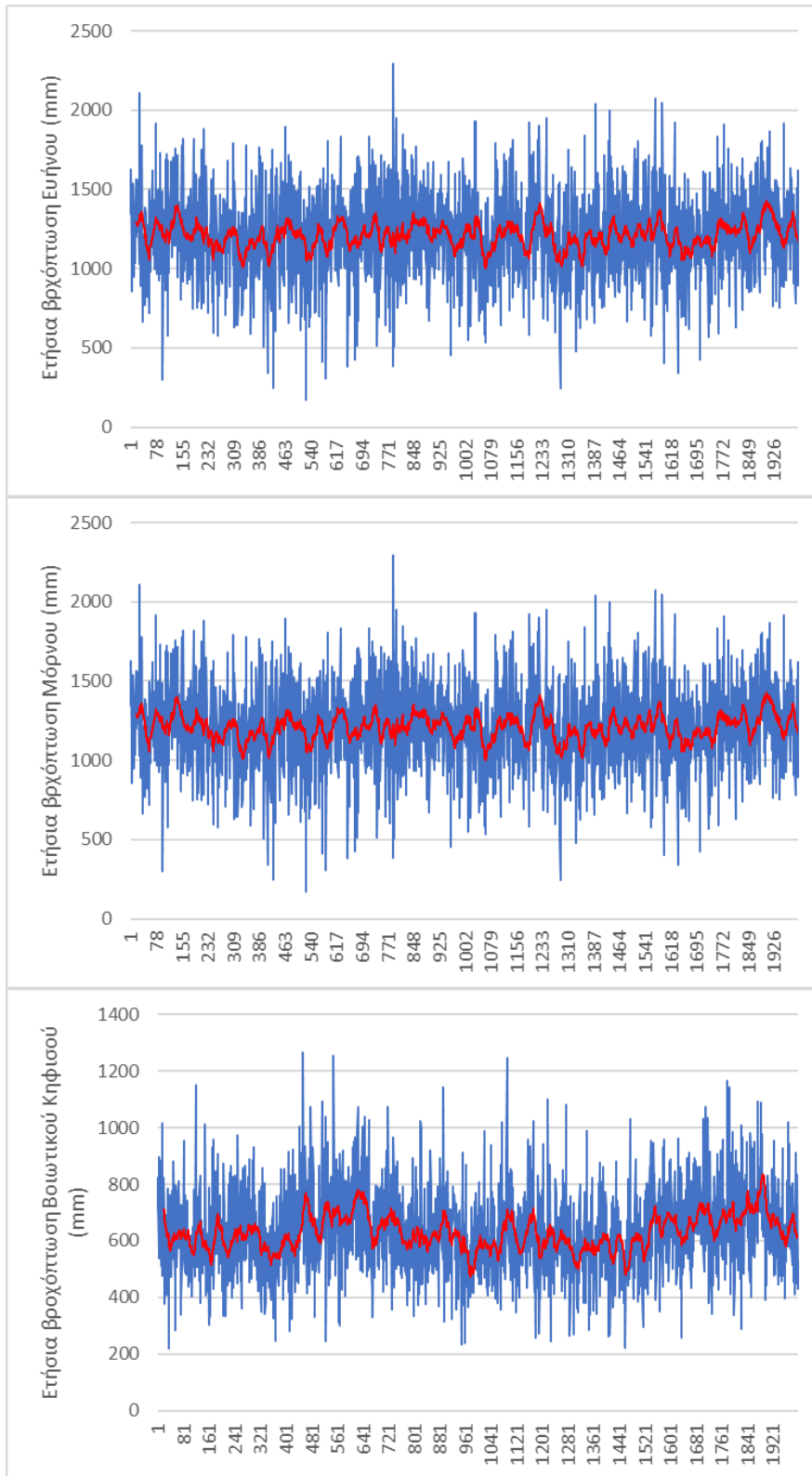
	Οκτ.	Νοε.	Δεκ.	Ιαν.	Φεβ.	Οκτ.- Φεβ.	Μαρ.- Σεπ.	Έτος
Μέση τιμή	7.5	10.2	15.2	22.1	24.2	79.2	62.7	141.9
Τυπική Απόκλιση	5.0	4.7	6.4	13.4	13.5	33.8	36.1	63.4
Ελάχιστη τιμή	0.0	3.4	4.7	7.4	1.8	26.9	5.4	33.5
Μέγιστη τιμή	21.8	25.8	31.5	72.5	64.8	192.9	148.0	340.9

3.5. Σύνοψη μεθοδολογίας γέννησης συνθετικών χρονοσειρών πρόγνωσης

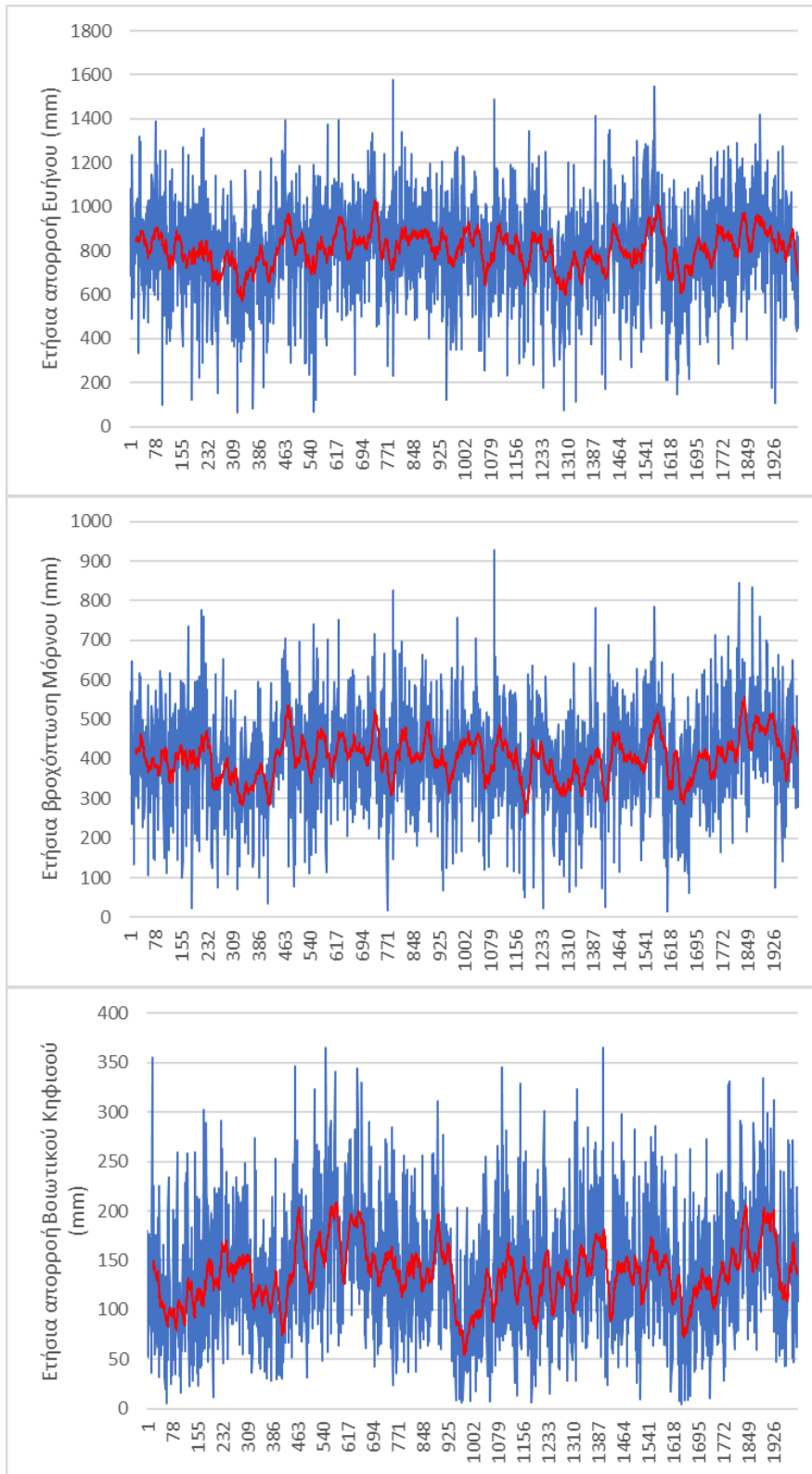
Επιγραμματικά, η παραγωγή των συνθετικών χρονοσειρών πρόγνωσης των εισροών και απωλειών των ταμιευτήρων του ΕΥΣ γίνεται σε δύο χρονικές κλίμακες, λαμβάνοντας υπόψη τις τιμές του πλέον πρόσφατου μήνα (στην προκειμένη περίπτωση, Φεβρουάριος 2024), και τις τιμές όλων των προηγούμενων υδρολογικών ετών.

Η γέννηση των συνθετικών χρονοσειρών έγινε μέσω πολυμεταβλητής στοχαστικής προσομοίωσης δύο χρονικών κλιμάκων, προκειμένου να αναπαράγονται σε αυτές τα στατιστικά χαρακτηριστικά και χρονικές και χωρικές συσχετίσεις των ιστορικών δειγμάτων. Σημαντικό ζητούμενο ήταν η αναπαραγωγή της μακροπρόθεσμης εμμονής, που αναφέρεται και ως δυναμική Hurst-Kolmogorov (Koutsoyiannis, 2011· Dimitriadis and Koutsoyiannis, 2015), προκειμένου να περιγραφούν οι μεγάλης κλίμακας υδροκλιματικές μεταβολές, με συσσώρευση μακρών περιόδων ξηρών και υγρών ετών. Η απαίτηση αυτή επιτεύχθηκε με την εισαγωγή μιας θεωρητικής συνάρτησης αυτοσυνδιασποράς, με κατάλληλη προσαρμογή των παραμέτρων της, για την περιγραφή της δομής αυτοσυσχέτισης των διεργασιών στην ετήσια κλίμακα (Koutsoyiannis, 2000· Efstratiadis *et al.*, 2014· Tsoukalas *et al.*, 2018). Με τη διαδικασία αυτή, παρήχθησαν χρονοσειρές με έντονη εμμονή, με συντελεστές Hurst της τάξης του 0.75 έως 0.80 για τις απορροές (πιο υψηλοί στην καρστική λεκάνη του Βοιωτικού Κηφισού, όπως υποδεικνύουν και τα ιστορικά δεδομένα), και 0.65 έως 0.70 για τις βροχοπτώσεις.

Ακολουθούν οι **Εικόνες 3.7** και **3.8**, όπου φαίνονται οι παραχθείσες συνθετικές χρονοσειρές για βροχόπτωση και απορροή για τις τρεις λεκάνες.



Εικόνα 3.7: Συνθετικές χρονοσειρές υψών βροχόπτωσης λεκανών Ευήνου, Μόρνου και Βοιωτικού Κηφισού, μήκους 2000 ετών, και κινούμενοι μέσοι όροι 20 ετών (mm).



Εικόνα 3.8: Συνθετικές χρονοσειρές υψών απορροής λεκανών Ευήνου, Μόρνου και Βοιωτικού Κηφισού, μήκους 2000 ετών, και κινούμενοι μέσοι όροι 20 ετών (mm).

4. Θεωρητικό πλαίσιο για τα μοντέλα πρόβλεψης

4.1. Εισαγωγή στη Μηχανική Μάθηση

Η μηχανική μάθηση (MM) είναι το πεδίο της τεχνητής νοημοσύνης που στοχεύει στην ανάπτυξη αλγορίθμων και μοντέλων για την αυτοματοποιημένη βελτίωση της απόδοσης συστημάτων μέσω εμπειρίας. Στηρίζεται στη στατιστική και την αλγοριθμική ανάλυση και έχει τη δυνατότητα να μαθαίνει από δεδομένα για να επιλύει προβλήματα όπως η κατηγοριοποίηση, η πρόβλεψη και η αναγνώριση μοτίβων.

Οι αλγόριθμοι της μηχανικής μάθησης διακρίνονται κυρίως σε τρεις κατηγορίες:

- **Επιβλεπόμενη Μάθηση (Supervised Learning):** Περιλαμβάνει εκπαίδευση με δεδομένα που περιέχουν ετικέτες ή κατηγορίες, π.χ., για την κατηγοριοποίηση (classification) ή την παλινδρόμηση (regression).
- **Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning):** Αναζητά κρυφά μοτίβα και σχέσεις στα δεδομένα χωρίς να απαιτείται σήμανση ή κατηγοριοποίηση των δεδομένων.
- **Ενισχυτική Μάθηση (Reinforcement Learning):** Το μοντέλο μαθαίνει μέσω αλληλεπίδρασης με το περιβάλλον του, βελτιστοποιώντας τις αποφάσεις του μέσω ανταμοιβών ή ποινών.

Οι τεχνικές που εφαρμόζονται στη συγκεκριμένη εργασία ανήκουν στην κατηγορία της Επιβλεπόμενης Μάθησης.

4.2. Τεχνητά Νευρωνικά Δίκτυα

Τα τεχνητά νευρωνικά δίκτυα (artificial neural networks) αποτελούν ένα μοντέλο υπολογισμού εμπνευσμένο από τη λειτουργία των βιολογικών νευρώνων και συγκεκριμένα του ανθρώπινου εγκεφάλου. Βασικό ζητούμενο για την χρήση των ΤΝΔ, ή κοινώς νευρωνικών δικτύων, αποτελεί η ανάγκη αποφυγής των περιορισμών των απλών γραμμικών μοντέλων, προκειμένου να μπορούν

να αναγνωριστούν και να μοντελοποιηθούν αρκετά πιο σύνθετες σχέσεις μεταξύ δεδομένων. Βασικά τους χαρακτηριστικά αποτελούν το γεγονός ότι αποκτούν γνώση μέσω μίας διαδικασίας μάθησης και αυτή η γνώση «αποθηκεύεται» σε κάποιες τιμές που ονομάζονται «συνοπτικά βάρη» και χαρακτηρίζουν τις συνδέσεις μεταξύ των νευρώνων του δικτύου. Ένα ΤΝΔ αποτελείται από:

- Το στρώμα εισόδου (input layer),
- Ένα ή περισσότερα κρυφά στρώματα (hidden layers), και
- Το στρώμα εξόδου (output layer).

Κάθε νευρώνας σε ένα ΤΝΔ λαμβάνει πολλαπλές εισόδους, τις οποίες πολλαπλασιάζει με αντίστοιχα βάρη και προσθέτει ένα όρο μεροληψίας (bias). Η συνολική τιμή περνάει από μια συνάρτηση ενεργοποίησης που καθορίζει την έξοδο του νευρώνα. Κατά τη διαδικασία της εκπαίδευσης, ένας αλγόριθμος τροποποιεί τα βάρη με σκοπό την επίτευξη του επιθυμητού αποτελέσματος.

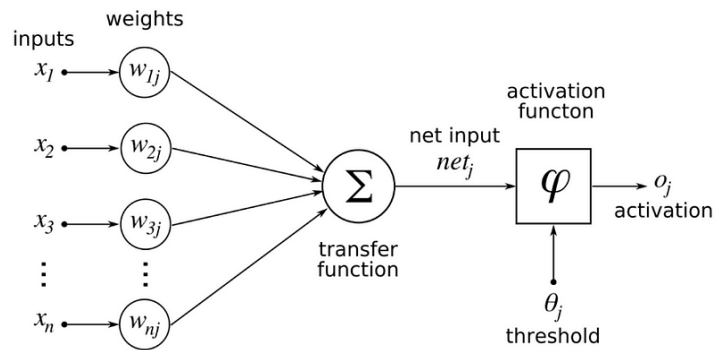
Η έξοδος κάθε νευρώνα y υπολογίζεται μέσω της εξίσωσης:

$$y = f(\sum_{i=1}^n w_i x_i + b)$$

όπου:

- x_i : οι είσοδοι του νευρώνα,
- w_i : τα βάρη που συνδέονται με τις εισόδους,
- b : ο όρος μεροληψίας, και
- f : η συνάρτηση ενεργοποίησης, που είναι συνήθως μη γραμμική (π.χ., sigmoid ή ReLU).

Η «ανατομία» ενός στρώματος ενός νευρωνικού δικτύου είναι συνήθως η εξής:



Εικόνα 4.1: «Ανατομία» ενός στρώματος νευρωνικού δικτύου.

Υπάρχουν τρεις βασικές αρχιτεκτονικές νευρωνικών δικτύων:

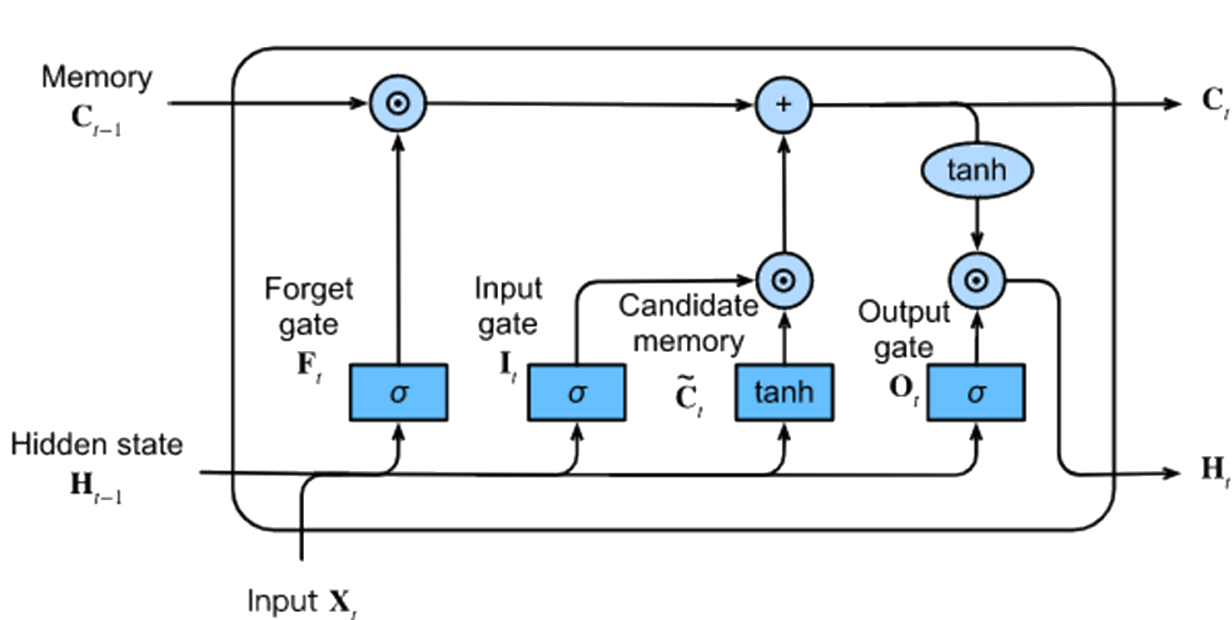
- Το **Single Feedforward Neural Network** είναι το πιο απλό είδος τεχνητού νευρωνικού δικτύου. Αποτελείται από ένα επίπεδο εισόδου (input layer), ένα επίπεδο εξόδου (output layer) και, ενδεχομένως, ένα ή περισσότερα κρυφά επίπεδα (hidden layers). Η ροή της πληροφορίας είναι μονοκατευθυντική, δηλαδή από την είσοδο προς την έξοδο, χωρίς αναδρομικές συνδέσεις. Το μοντέλο αυτό είναι κατάλληλο για βασικές εφαρμογές, όπως η απλή ταξινόμηση και παλινδρόμηση.
- Το **Multi-Layer Feedforward Neural Network (MFFNN)** επεκτείνει την αρχιτεκτονική του Single Feedforward, περιλαμβάνοντας πολλαπλά κρυφά επίπεδα. Αυτά τα επίπεδα αυξάνουν τη δυνατότητα του δικτύου να μαθαίνει περίπλοκα μοτίβα και μη εξαρτήσιες που μπορεί να υπάρχουν στα δεδομένα. Συνήθως, χρησιμοποιούν αλγόριθμο backpropagation για τη βελτιστοποίηση των βαρών τους. Οι περισσότερες εφαρμογές μηχανικής μάθησης βασίζονται σε τέτοιες πολυεπίπεδες δομές.
- Τα **Αναδρομικά Νευρωνικά Δίκτυα (RNNs)** είναι μοναδικά επειδή διαθέτουν συνδέσεις αναδρομής, επιτρέποντας στα δεδομένα να ρέουν και προς τα πίσω. Αυτό τους επιτρέπει να διατηρούν πληροφορίες από προηγούμενα χρονικά βήματα, κάνοντάς τα ιδανικά για ακολουθιακά δεδομένα. Παρόλο που τα RNNs έχουν ισχυρές δυνατότητες, παρουσιάζουν δυσκολίες εκπαίδευσης. Γι' αυτό συχνά χρησιμοποιούνται παραλλαγές, όπως τα LSTMs και GRUs, όπως γίνεται και στην παρούσα εργασία.

4.3. Δίκτυα LSTM

Τα LSTMs (Long Short-Term Memory Networks) είναι μια ιδιαίτερη κατηγορία αναδρομικών νευρωνικών δικτύων (Recurrent Neural Networks - RNNs) που σχεδιάστηκαν για να αντιμετωπίσουν τα προβλήματα των τυπικών RNNs, όπως το **vanishing gradient** και η αδυναμία διατήρησης μακροπρόθεσμων εξαρτήσεων. Εισήχθησαν από τους Hochreiter και Schmidhuber το 1997 και έχουν αποδειχθεί εξαιρετικά αποτελεσματικά σε εφαρμογές όπως η ανάλυση χρονοσειρών, η αναγνώριση ομιλίας και η μηχανική μετάφραση.

Η βασική αρχή λειτουργίας των LSTMs είναι η ενσωμάτωση μηχανισμών **πύλης (gates)** που ρυθμίζουν τη ροή της πληροφορίας εντός του μοντέλου. Οι πύλες αυτές περιλαμβάνουν:

- **Πύλη εισόδου (Input gate):** Ελέγχει ποια νέα πληροφορία θα αποθηκευτεί στη μνήμη.
- **Πύλη διαγραφής (Forget gate):** Αποφασίζει ποια πληροφορία από τη μνήμη θα διαγραφεί.
- **Πύλη εξόδου (Output gate):** Καθορίζει ποια πληροφορία από τη μνήμη θα χρησιμοποιηθεί στο τρέχον χρονικό βήμα.



Εικόνα 4.2: Διάγραμμα λειτουργίας ενός νευρώνα τύπου LSTM.

Αυτή η δομή επιτρέπει στις LSTMs να διατηρούν πληροφορίες για μεγάλες ακολουθίες δεδομένων, καθιστώντας τες ιδανικές για προβλήματα όπου οι εξαρτήσεις μεταξύ των δεδομένων είναι μακροπρόθεσμες. Η καινοτομία των LSTMs έγκειται στο γεγονός ότι αντιμετωπίζουν τις αδυναμίες των παραδοσιακών RNNs, παρέχοντας παράλληλα μεγαλύτερη σταθερότητα και ακρίβεια σε δύσκολες εφαρμογές, χάρη στη λειτουργία μνήμης τους.

Τα **Bidirectional LSTMs (BiLSTMs)** επεκτείνουν τη λειτουργικότητα των LSTMs προσθέτοντας τη δυνατότητα επεξεργασίας δεδομένων και προς τις δύο κατευθύνσεις: από την αρχή προς το τέλος (**forward pass**) και από το τέλος προς την αρχή (**backward pass**). Αυτή η διπλή ροή πληροφορίας επιτρέπει την πλήρη κατανόηση του συμφραζομένου σε χρονοσειρές ή ακολουθίες, όπου το μέλλον μπορεί να παρέχει εξίσου σημαντικά δεδομένα με το παρελθόν.

Τα BiLSTMs αποτελούνται από δύο ξεχωριστά επίπεδα LSTMs:

1. Το **πρώτο επίπεδο** επεξεργάζεται την είσοδο από την αρχή προς το τέλος.
2. Το **δεύτερο επίπεδο** επεξεργάζεται την ίδια είσοδο, αλλά αντίστροφα.

Τα αποτελέσματα από τα δύο αυτά επίπεδα συνδυάζονται για να παράγουν μια πιο πλούσια αναπαράσταση της εισόδου. Έτσι, αξιοποιούνται καλύτερα τα διαθέσιμα δεδομένα.

4.4. Εκπαίδευση ενός LSTM

Η διαδικασία εκπαίδευσης ξεκινά με την είσοδο δεδομένων στο δίκτυο και τον υπολογισμό της εξόδου σε κάθε κατεύθυνση. Στη συνέχεια, η έξοδος των δύο κατευθύνσεων συνδυάζεται (π.χ. μέσω μέσου όρου) για να παραχθεί το τελικό αποτέλεσμα. Η παραγόμενη έξοδος συγκρίνεται με την αναμενόμενη έξοδο, και υπολογίζεται το σφάλμα. Στη συγκεκριμένη περίπτωση χρησιμοποιήθηκε το μέσο τετραγωνικό σφάλμα ΜΤΣ (MSE), με τύπο:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Αυτό το σφάλμα επιστρέφεται στο δίκτυο μέσω της διαδικασίας της πίσω διάδοσης (backpropagation) και χρησιμοποιείται για την ενημέρωση των **βαρών** και των **παραμέτρων** του δικτύου. Το BiLSTM εκπαιδεύεται με χρήση αλγορίθμων βελτιστοποίησης, όπως:

- **Στοχαστική κατάβαση κλίσης (Stochastic Gradient Descent, SGD):** Ενημερώνει τα βάρη με βάση το σφάλμα σε μικρά δείγματα δεδομένων.
- **Adam ή RMSProp:** Βελτιωμένες μέθοδοι που προσαρμόζουν δυναμικά το ρυθμό μάθησης για ταχύτερη και πιο σταθερή σύγκλιση.

Στη συγκεκριμένη περίπτωση έγινε χρήση του αλγορίθμου Adam, ο οποίος χρησιμοποιείται πολύ συχνά για την επίλυση πληθώρας προβλημάτων. Η διπλής κατεύθυνσης δομή του LSTM επιτρέπει στο μοντέλο να αξιοποιήσει πλήρως τα χρονικά μοτίβα και τις σχέσεις στα δεδομένα, προσφέροντας μεγαλύτερη ακρίβεια και αποδοτικότητα σε εφαρμογές που απαιτούν ακριβή ανάλυση ακολουθιών, όπως πρόβλεψη χρονοσειρών ή ανάλυση φυσικής γλώσσας.

Παράλληλα, μέρος της εκπαίδευσης αποτελεί και η εύρεση των καλύτερων υπερπαραμέτρων των μοντέλων, με σκοπό την εύρεση του μοντέλου που θα μπορέσει να προσφέρει τις καλύτερες δυνατές προβλέψεις, με το μικρότερο δυνατό σφάλμα. Για τους σκοπούς της συγκεκριμένης εργασίας δοκιμάστηκαν δύο διαφορετικοί τρόποι τυποποίησης των δεδομένων (Κανονικοποίηση και τυποποίηση με log), διαφορετικοί ρυθμοί μάθησης learning rate, διαφορετικά batch sizes για την εκπαίδευση, αλλά και διαφορετικές τιμές dropout για το 1^ο LSTM layer.

Το dropout είναι μια τεχνική κανονικοποίησης (regularization) που χρησιμοποιείται στα νευρωνικά δίκτυα για την αποφυγή της υπερπροσαρμογής (overfitting), κάτι ιδιαίτερα σημαντικό στη συγκεκριμένη περίπτωση, αφού χρησιμοποιούνται συνθετικά δεδομένα. Κατά την εκπαίδευση του δικτύου, «απενεργοποιούνται» τυχαία (δηλαδή, μηδενίζονται) ορισμένοι νευρώνες. Αυτό σημαίνει ότι αυτοί οι νευρώνες δεν συμβάλλουν στη διάδοση των δεδομένων και στην εκτέλεση των ενημερώσεων βαρών κατά τη διάρκεια του συγκεκριμένου βήματος. Ο στόχος αυτής της διαδικασίας είναι να μειωθεί η εξάρτηση του μοντέλου από συγκεκριμένους νευρώνες και να ενισχυθεί η γενίκευση (generalization) του δικτύου σε νέα δεδομένα.

4.5. Βιβλιοθήκη Keras και υλοποίηση του LSTM στην Python

Η βιβλιοθήκη Keras αποτελεί ένα από τα πιο δημοφιλή εργαλεία για την ανάπτυξη και εκπαίδευση νευρωνικών δικτύων στην Python. Αρχικά, αναπτύχθηκε ως υψηλού επιπέδου API πάνω στο TensorFlow και άλλες βιβλιοθήκες, με σκοπό να διευκολύνει τη διαδικασία δημιουργίας μοντέλων μηχανικής μάθησης. Η Keras παρέχει έναν ευέλικτο τρόπο ανάπτυξης βαθιών νευρωνικών δικτύων μέσω μιας καθαρής και κατανοητής συντακτικής μορφής. Προσφέρει τη δυνατότητα κατασκευής και εκπαίδευσης νευρωνικών δικτύων με ελάχιστο κώδικα, διευκολύνοντας έτσι την χρήση της σε σχέση με παλαιότερες υλοποιήσεις αντίστοιχων βιβλιοθηκών.

Μερικά από τα κύρια χαρακτηριστικά της βιβλιοθήκης Keras περιλαμβάνουν:

- Εύκολη και γρήγορη δημιουργία μοντέλων νευρωνικών δικτύων με τη χρήση έτοιμων layers και εργαλείων,
- Υποστήριξη πολλών τύπων δικτύων, όπως Convolutional Neural Networks (CNNs) και Recurrent Neural Networks (RNNs), στην κατηγορία των οποίων ανήκουν και τα LSTM,
- Ευκολία στην ενσωμάτωση με άλλες βιβλιοθήκες μηχανικής μάθησης και ανάλυσης δεδομένων, όπως η TensorFlow.

Η βιβλιοθήκη έχει αποδειχθεί ιδιαίτερα χρήσιμη για έρευνα στη μηχανική μάθηση και χρησιμοποιείται ευρέως για την υλοποίηση μοντέλων που αποσκοπούν στην επίλυση προβλημάτων πολλών διαφορετικών επιστημονικών τομέων.

5. Προβλέψεις

5.1. Γραμμικό μοντέλο

Σα «μοντέλο αναφοράς» για την αξιολόγηση των επιδόσεων του LSTM χρησιμοποιήθηκε ένα απλό γραμμικό μοντέλο δύο μεταβλητών, με δεδομένα εισόδου την αθροιστική βροχόπτωση και την αθροιστική απορροή για την περίοδο Οκτωβρίου – Φεβρουαρίου, που είχε σκοπό να προβλέψει την αθροιστική απορροή για την περίοδο Μαρτίου – Σεπτεμβρίου. Όπως έγινε και με τα LSTM, η «εκπαίδευση» των γραμμικών μοντέλων έγινε με τα συνθετικά δεδομένα και η αξιολόγησή τους έγινε με βάση τις επιδόσεις τους στην πρόβλεψη των πραγματικών. Σαν όρος σφάλματος χρησιμοποιήθηκε και εδώ το Μέσο Τετραγωνικό Σφάλμα MSE.

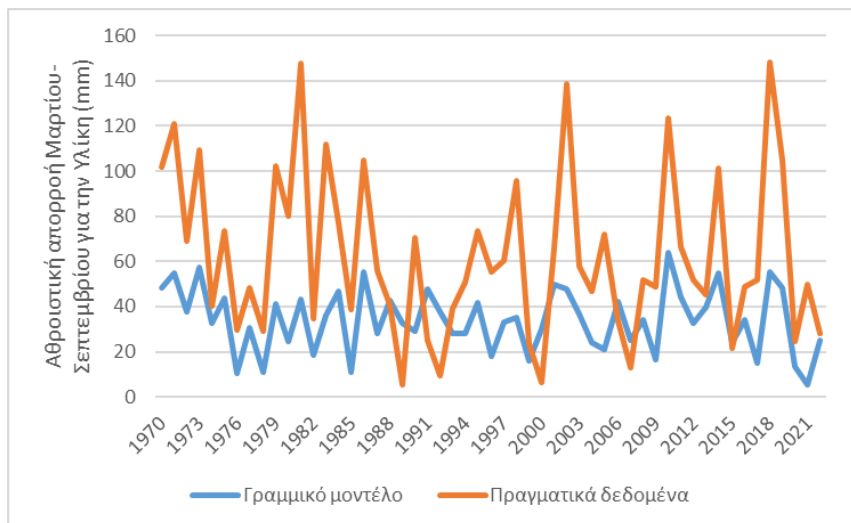
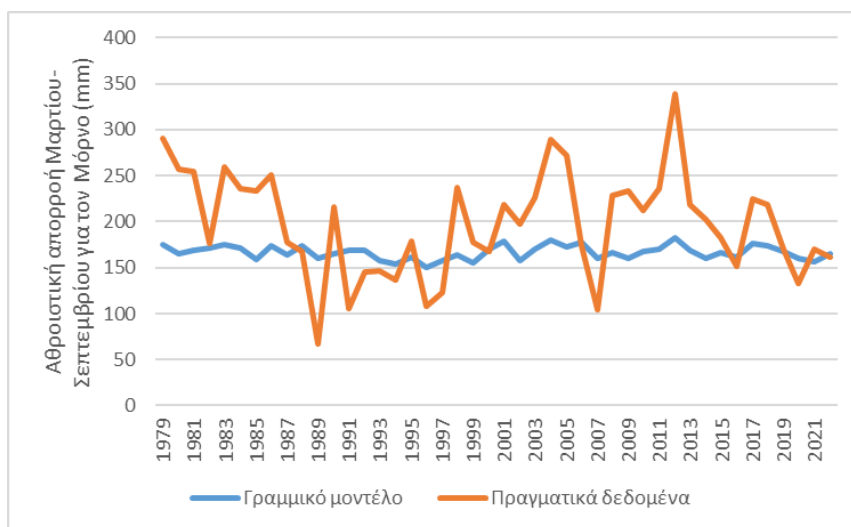
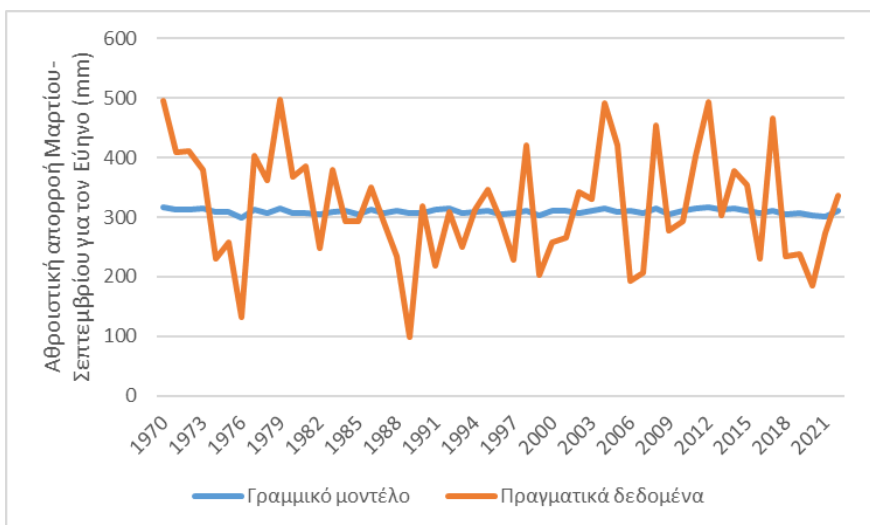
Χρησιμοποιείται ο τύπος:

$$\hat{Y} = w_1 x_1 + w_2 x_2 + b.$$

Μετά την παλινδρόμηση για τον υπολογισμό των συντελεστών στα συνθετικά δεδομένα προκύπτει:

- Εύηνος: $w_1 = -0.013$, $w_2 = 0.035$, $b = 309.01$,
- Μόρνος: $w_1 = -0.049$, $w_2 = 0.149$, $b = 165.00$,
- Υλίκη: $w_1 = -0.047$, $w_2 = 0.305$, $b = 53.01$.

Κατά την πρόβλεψη των πραγματικών δεδομένων με τα παραπάνω μοντέλα, προκύπτουν τα εξής διαγράμματα:



Εικόνα 5.1: Προβλέψεις της απορροής Μαρτίου – Σεπτεμβρίου από το γραμμικό μοντέλο και πραγματικά δεδομένα για τις τρεις λεκάνες (mm).

5.2. Γενική αρχιτεκτονική LSTM

Το μοντέλο LSTM που χρησιμοποιήθηκε για τις προβλέψεις σχεδιάστηκε με 2 μεταβλητές εισόδου (βροχόπτωση και απορροή), με την κάθε μία να έχει 5 χρονικά βήματα (Οκτώβριος – Φεβρουάριος). Η «ανατομία» του αποτελείται από 2 στρώματα LSTM (128 και 64 νευρώνων, με το στρώμα των 128 να είναι το στρώμα εισόδου του μοντέλου και το στρώμα των 64 να είναι «κρυφό»). Για την έξοδο χρησιμοποιήθηκε ένα layer τύπου Dense, που προσφέρεται από το Keras, με 7 νευρώνες, καθώς γίνεται πρόβλεψη για την απορροή των 7 μηνών ξεχωριστά. Για το πρώτο στρώμα, αυτό της εισόδου, έγινε και εφαρμογή της μεθόδου κανονικοποίησης dropout, όπως αναφέρθηκε στο **Κεφάλαιο 4.4**.

Για την εκπαίδευση χρησιμοποιήθηκαν 100 epochs, με ταυτόχρονη χρήση του early_stopping, καθώς ήταν ιδιαίτερα σημαντικό να γίνει αποφυγή της υπερπροσαρμογής (overfitting). Αυτό σημαίνει πως αν κατά τη διάρκεια της εκπαίδευσης δεν παρατηρηθεί βελτίωση του σφάλματος επικύρωσης για 10 συνεχόμενα βήματα, η εκπαίδευση σταματάει. Αν δε συμβεί αυτό, η εκπαίδευση του μοντέλου θα διαρκέσει για 100 επαναλήψεις.

Όσον αφορά τα δεδομένα της εκπαίδευσης, αυτά χωρίστηκαν σε δύο χρονοσειρές, μία για την εκπαίδευση (training) και άλλη μία για την επικύρωση (validation) του μοντέλου, χωρίς να γίνει κάποια ανακατάταξη στη σειρά τους. Χρησιμοποιήθηκε split σε ποσοστό 80% / 20%.

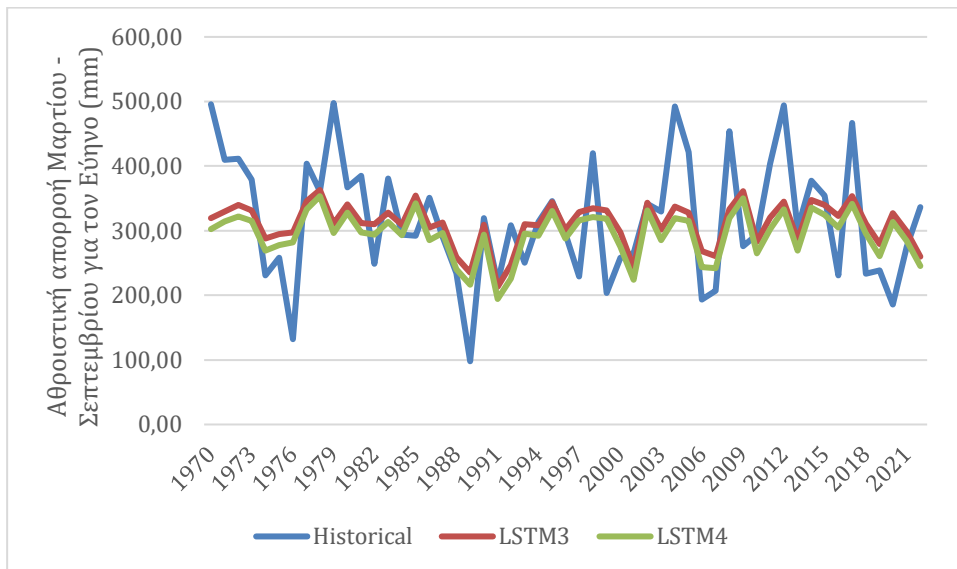
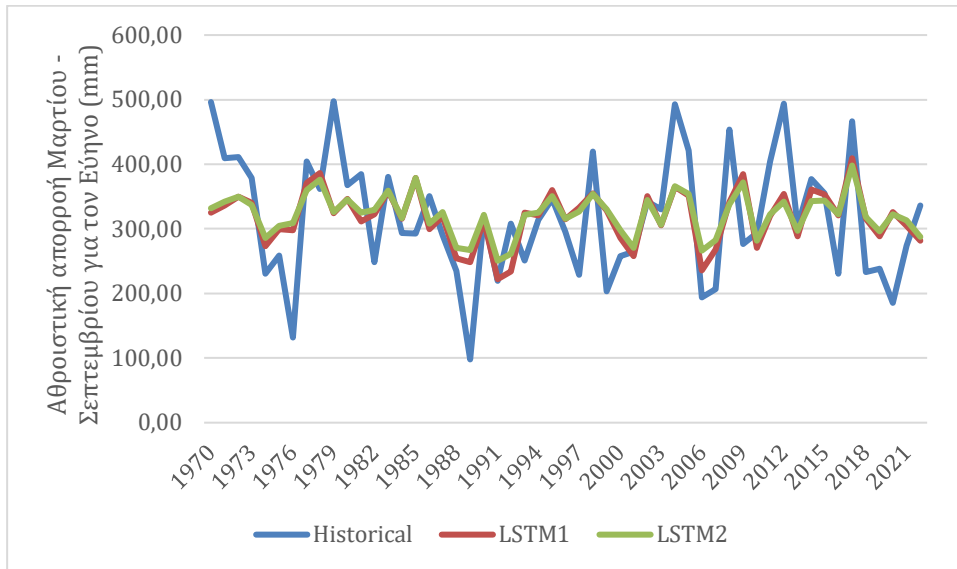
Όπως αναφέρθηκε και στο **Κεφάλαιο 4**, έγινε αναζήτηση πλέγματος (grid search) για την εύρεση των βέλτιστων υπερπαραμέτρων batch size, learning rate και dropout για κάθε λεκάνη και κάθε μέθοδο τυποποίησης (Κανονική $\frac{x-\mu}{\sigma}$, log).

5.3. Προβλέψεις για τη λεκάνη του Ευήνου

Για τον Ευήνο έγιναν συνολικά 4 διαδικασίες αναζήτησης πλέγματος, 2 για κάθε μέθοδο τυποποίησης των δεδομένων. Τα αποτελέσματα που προέκυψαν είναι τα εξής:

Πίνακας 5.1: Βέλτιστες υπερπαράμετροι για τον Εύηνο.

Τυποποίηση	Learning Rate	Batch Size	Dropout
Κανονική	0.001	4	0
Κατανομή	0.001	2	0
Log	0.001	8	0.1
	0.001	16	0.1



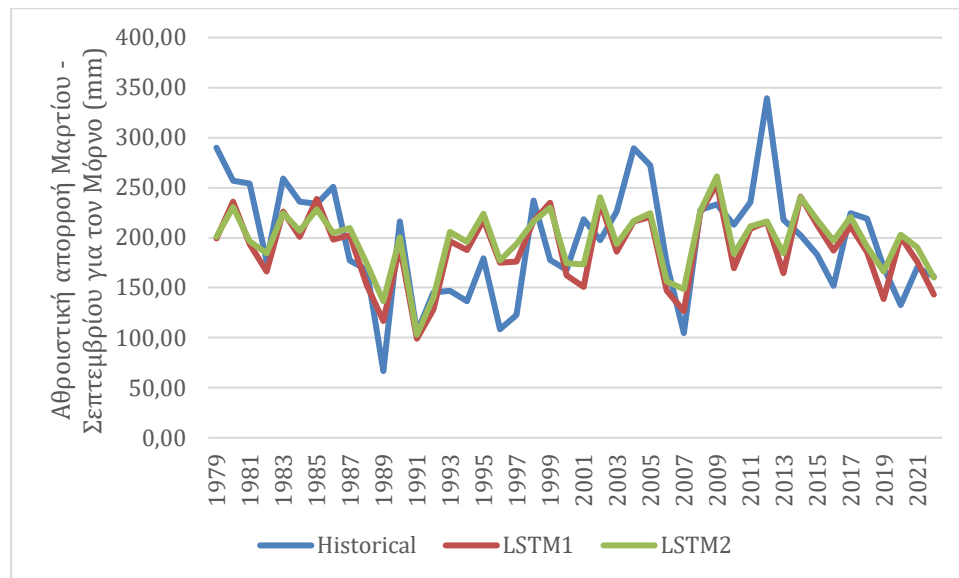
Εικόνα 5.2: Προβλέψεις της αποροής Μαρτίου – Σεπτεμβρίου από τα LSTM (1,2 Κανονική Κατανομή, 3, 4 Log) και πραγματικά δεδομένα για τον Εύηνο (mm).

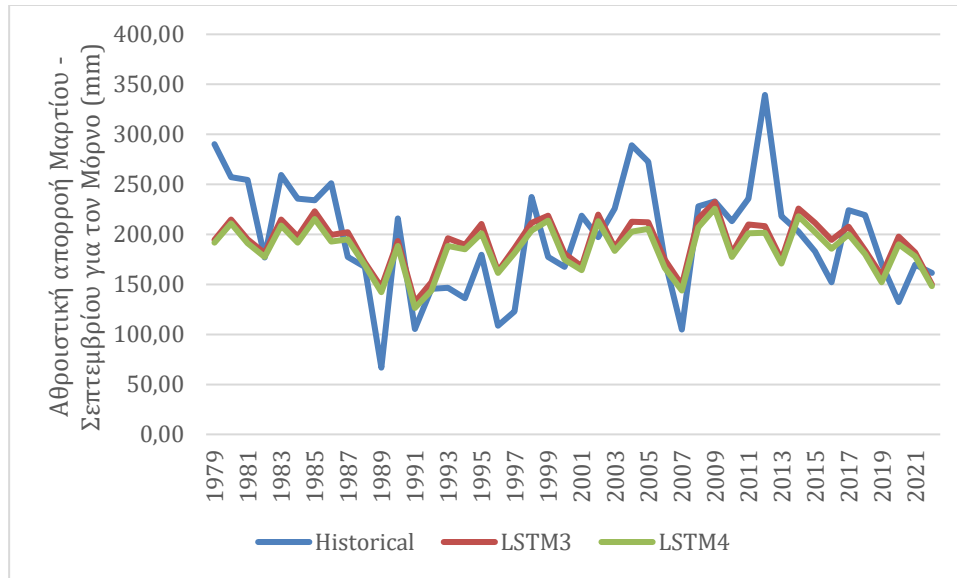
5.4. Προβλέψεις για τη λεκάνη του Μόρνου

Για τον Μόρνο, αντίστοιχα, έγιναν συνολικά 4 διαδικασίες αναζήτησης πλέγματος, 2 για κάθε μέθοδο τυποποίησης των δεδομένων. Τα αποτελέσματα που προέκυψαν είναι τα εξής:

Πίνακας 5.2: Βέλτιστες υπερπαραμέτροι για τον Μόρνο.

Τυποποίηση	Learning Rate	Batch Size	Dropout
Κανονική	0.01	32	0
Κατανομή	0.005	32	0
Log	0.001	16	0
	0.0005	8	0





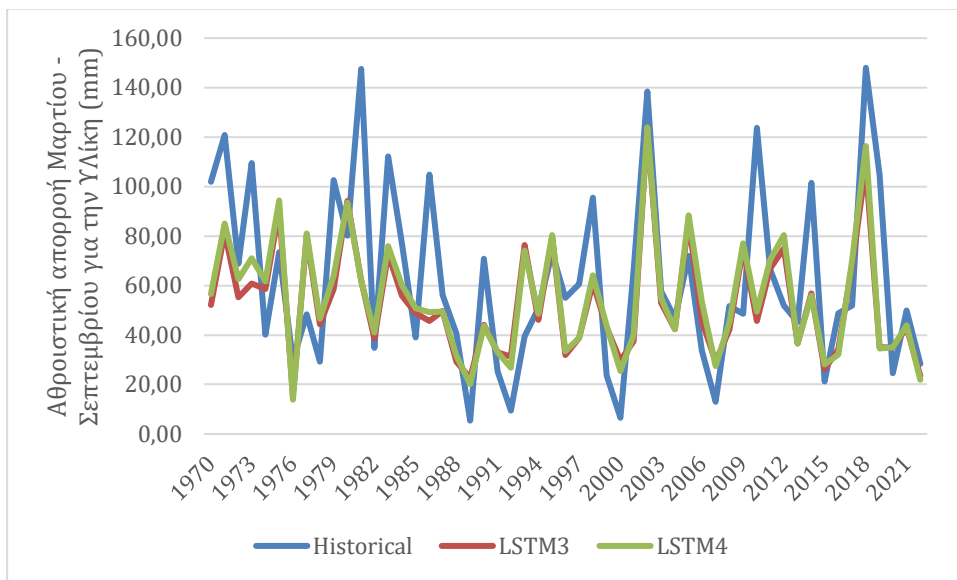
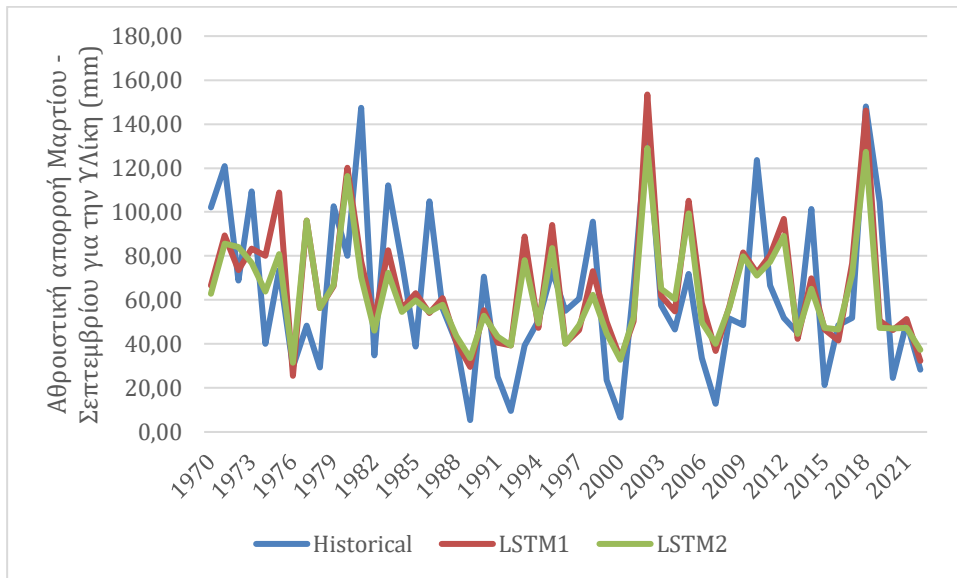
Εικόνα 5.3: Προβλέψεις της απορροής Μαρτίου – Σεπτεμβρίου από τα LSTM (1, 2 Κανονική Κατανομή, 3, 4 Log) και πραγματικά δεδομένα για τον Μόρνο (mm).

5.5. Προβλέψεις για τη λεκάνη του Βοιωτικού Κηφισού

Για την Υλίκη, αντίστοιχα, έγιναν συνολικά 4 διαδικασίες αναζήτησης πλέγματος, 2 για κάθε μέθοδο τυποποίησης των δεδομένων. Τα αποτελέσματα που προέκυψαν είναι τα εξής:

Πίνακας 5.3: Βέλτιστες υπερπαράμετροι για την Υλίκη.

Τυποποίηση	Learning Rate	Batch Size	Dropout
Κανονική	0.01	32	0.5
Κατανομή	0.005	32	0.4
Log	0.001	16	0
	0.0005	8	0



Εικόνα 5.4: Προβλέψεις της απορροής Μαρτίου – Σεπτεμβρίου από τα LSTM (1, 2 Κανονική Κατανομή, 3, 4 Log) και πραγματικά δεδομένα για την Υλίκη (mm).

5.6. Συνολικές προβλέψεις

Πίνακας 5.4: Προβλέψεις της αθροιστικής απορροής Μαρτίου – Σεπτεμβρίου από όλα τα μοντέλα για τον Εύηνο (mm).

Έτος	Historical	Linear	LSTM1	LSTM2	LSTM3	LSTM4
1970	496.10	316.39	324.77	332.28	319.53	302.70
1971	409.30	313.52	336.36	342.09	329.81	314.17
1972	411.40	313.19	349.44	349.77	339.61	322.33
1973	378.80	314.16	340.84	337.13	331.28	315.54
1974	231.00	307.86	272.75	286.08	288.08	269.54
1975	258.30	309.17	299.30	304.88	294.69	277.71
1976	131.80	299.05	298.15	308.53	297.54	282.10
1977	403.90	312.67	371.33	360.35	345.52	332.68
1978	361.50	307.30	386.21	376.24	362.99	353.42
1979	497.60	314.34	323.93	326.20	311.34	296.64
1980	367.50	306.96	346.20	345.51	340.56	328.31
1981	384.90	307.11	311.25	325.01	312.02	297.29
1982	248.40	304.98	322.24	329.78	309.96	294.14
1983	380.40	307.97	358.59	358.82	328.16	313.80
1984	293.50	310.98	315.89	316.49	309.57	292.96
1985	292.50	304.05	379.13	378.77	354.05	342.07
1986	350.60	313.00	299.93	308.48	304.84	285.39
1987	290.30	307.41	319.42	326.09	312.93	296.16
1988	234.60	310.85	254.29	270.63	258.93	239.84
1989	97.90	306.17	248.31	267.20	235.50	216.61
1990	319.40	307.34	313.13	321.38	309.50	293.53
1991	219.30	313.11	222.07	250.60	213.07	194.27
1992	308.10	314.69	234.21	261.45	247.08	225.97
1993	250.60	307.40	324.95	322.09	310.17	295.61
1994	313.70	308.64	320.78	325.06	308.63	292.09
1995	346.10	311.66	360.27	350.79	343.75	330.76
1996	295.60	304.78	314.92	315.89	301.68	287.72
1997	228.90	306.27	332.53	326.82	328.97	315.82
1998	419.90	310.48	354.88	354.05	335.11	321.43
1999	203.30	303.61	328.57	330.50	331.59	318.60
2000	258.20	311.02	286.11	298.24	297.43	275.14
2001	264.80	311.73	257.78	270.66	242.36	224.35
2002	341.90	307.61	350.96	344.76	343.52	331.46
2003	329.80	311.07	305.90	306.30	301.70	285.60
2004	492.50	314.12	364.90	366.25	337.36	319.80

2005	421.70	308.67	351.80	353.96	328.32	314.86
2006	193.60	309.83	236.06	266.12	268.42	243.65
2007	207.10	307.27	268.40	282.14	261.03	241.94
2008	453.60	315.18	341.46	337.91	333.47	319.31
2009	276.40	304.20	384.86	371.69	360.76	350.07
2010	293.50	311.76	270.26	281.84	282.93	264.64
2011	403.90	314.99	320.10	322.30	320.58	302.74
2012	493.90	316.22	354.27	341.82	344.67	332.96
2013	303.60	312.02	288.54	296.92	286.22	269.13
2014	377.40	314.03	361.19	343.36	347.51	335.50
2015	354.50	311.47	352.89	344.11	339.77	324.43
2016	230.60	307.66	320.53	322.34	322.52	305.07
2017	466.80	310.27	409.38	398.09	353.53	341.59
2018	233.20	305.19	316.05	318.03	310.95	295.54
2019	238.40	307.00	288.30	296.47	279.09	260.47
2020	185.70	302.27	325.88	322.95	327.52	313.57
2021	274.20	300.90	304.72	313.13	299.38	284.22
2022	336.10	310.55	281.65	288.02	260.10	245.38

Πίνακας 5.5: Προβλέψεις της αθροιστικής απορροής Μαρτίου – Σεπτεμβρίου από όλα τα μοντέλα για τον Μόρνο (mm).

Έτος	Historical	Linear	LSTM1	LSTM2	LSTM3	LSTM4
1979	290.20	175.24	199.01	200.83	194.39	191.43
1980	257.20	165.21	236.15	230.55	214.64	210.92
1981	254.30	168.98	193.56	196.69	195.12	191.40
1982	176.90	171.23	166.40	184.23	181.04	177.49
1983	259.30	175.51	226.26	224.48	214.68	208.53
1984	235.90	171.75	201.00	207.07	197.86	191.66
1985	233.80	159.25	238.81	228.53	222.91	215.54
1986	251.00	174.08	197.96	204.52	199.33	192.93
1987	177.40	163.75	202.74	209.43	202.34	195.26
1988	167.70	174.21	152.54	173.92	172.17	169.39
1989	66.70	160.73	116.89	136.73	147.04	142.01
1990	216.00	165.30	191.38	200.23	193.41	188.45
1991	105.10	169.00	99.03	103.02	132.25	126.41
1992	145.40	169.26	128.34	140.40	151.88	144.07
1993	146.80	157.83	196.30	205.76	195.89	188.40
1994	136.40	153.98	187.38	195.79	189.75	184.95
1995	179.40	161.35	217.86	223.79	210.29	201.09
1996	108.40	149.95	174.98	177.19	162.90	161.28

1997	122.80	157.13	176.07	193.53	186.07	181.43
1998	237.20	164.03	216.56	215.98	211.73	203.79
1999	177.60	155.01	234.77	229.79	218.61	213.81
2000	167.70	170.28	162.22	174.31	180.66	174.51
2001	218.50	179.37	151.08	173.47	168.29	164.33
2002	197.40	157.87	235.63	240.67	219.63	213.24
2003	225.90	169.66	185.81	193.84	186.58	183.67
2004	289.30	180.18	216.15	217.04	212.53	202.79
2005	272.50	173.10	220.44	224.33	211.98	205.40
2006	174.00	177.62	147.20	156.17	174.87	166.34
2007	104.60	160.80	126.68	148.68	149.35	144.09
2008	228.00	165.96	224.86	226.59	216.01	207.28
2009	233.20	160.13	253.80	261.30	232.31	225.64
2010	212.90	167.70	169.45	183.79	181.41	177.36
2011	235.60	169.73	209.25	211.46	209.91	200.97
2012	339.40	183.00	215.93	216.00	207.95	201.38
2013	218.10	168.48	164.42	184.21	174.63	170.63
2014	202.50	159.91	240.77	240.64	225.60	218.05
2015	183.00	165.94	213.00	218.11	211.32	202.40
2016	151.90	161.13	187.00	197.15	194.18	185.73
2017	224.40	176.11	211.38	220.40	207.85	199.76
2018	219.20	173.69	185.46	190.84	183.78	180.13
2019	170.70	168.12	138.54	166.09	158.55	152.07
2020	132.40	159.59	200.69	203.02	197.58	190.20
2021	169.90	156.61	176.38	190.68	181.95	178.06
2022	161.20	165.70	143.08	160.43	149.68	148.29

Πίνακας 5.6: Προβλέψεις της αθροιστικής απορροής Μαρτίου – Σεπτεμβρίου από όλα τα μοντέλα για την Υλίκη (mm).

Έτος	Historical	Linear	LSTM1	LSTM2	LSTM3	LSTM4
1970	102.00	48.50	66.63	62.74	52.03	56.61
1971	120.80	54.85	89.45	85.53	81.42	85.06
1972	68.80	37.56	73.62	84.23	55.14	62.53
1973	109.40	57.44	83.27	76.94	60.81	70.96
1974	40.10	32.39	80.01	63.88	58.50	61.58
1975	73.50	43.82	108.84	80.97	90.57	94.46
1976	29.60	10.28	25.53	31.50	15.61	13.86
1977	48.30	30.42	96.08	96.11	79.85	81.12
1978	29.30	10.85	56.31	56.42	44.28	46.58
1979	102.50	41.17	66.41	66.86	58.56	63.99
1980	80.10	24.38	120.03	116.28	94.05	93.59

1981	147.50	43.26	77.71	70.44	62.23	61.81
1982	34.80	18.38	51.05	46.15	38.56	40.93
1983	112.10	36.10	82.70	72.24	72.85	75.94
1984	77.00	46.91	56.02	54.70	55.95	60.09
1985	38.90	10.85	63.10	59.91	48.89	50.74
1986	104.90	55.46	54.08	54.48	45.70	49.17
1987	56.10	28.29	60.78	57.79	49.71	49.49
1988	40.60	42.70	40.88	43.58	29.33	31.74
1989	5.40	32.39	29.51	33.55	22.50	20.16
1990	70.70	28.86	55.47	52.76	44.13	43.75
1991	25.20	47.91	40.71	43.35	32.95	33.06
1992	9.50	37.69	39.41	39.40	31.20	26.88
1993	39.40	28.09	88.74	78.05	76.37	74.10
1994	50.70	28.10	47.42	49.35	46.16	48.50
1995	73.30	41.64	94.08	83.61	79.72	80.46
1996	55.10	17.99	40.45	40.19	31.95	33.39
1997	60.50	33.20	46.28	48.50	38.68	38.84
1998	95.50	35.02	73.08	62.34	61.04	64.04
1999	23.70	15.78	50.03	44.88	43.00	43.77
2000	6.60	29.99	33.34	32.86	29.60	25.55
2001	66.10	49.82	50.33	51.98	37.39	39.99
2002	138.50	47.66	153.45	129.10	123.55	123.97
2003	57.70	36.84	61.75	65.17	53.02	54.79
2004	46.70	24.05	54.89	60.35	42.31	42.36
2005	71.90	21.20	105.12	99.50	86.08	88.30
2006	33.60	42.05	58.72	49.91	45.71	53.52
2007	12.90	25.00	36.90	40.08	29.30	27.37
2008	51.70	34.15	56.99	56.57	42.17	44.92
2009	48.60	16.34	81.68	79.92	75.27	77.04
2010	123.70	63.90	72.41	71.16	45.78	49.47
2011	66.60	44.21	80.48	77.45	66.91	70.31
2012	51.90	32.88	96.99	89.14	75.01	80.46
2013	45.20	39.66	42.46	44.50	36.49	36.50
2014	101.40	54.71	69.94	65.00	56.80	56.01
2015	21.30	24.10	46.50	47.44	26.08	28.10
2016	48.70	34.36	41.49	46.65	34.27	32.03
2017	51.90	15.18	76.69	71.56	68.07	70.13
2018	148.00	55.49	146.08	127.33	108.23	116.29
2019	104.60	48.05	50.60	47.27	35.07	34.54
2020	24.50	13.43	46.37	47.03	35.04	34.91
2021	49.90	5.29	51.22	47.39	43.11	43.84
2022	28.30	25.24	32.43	37.26	23.55	21.78

6. Αξιολόγηση των Μοντέλων

6.1. Μετρικές για την αξιολόγηση των μοντέλων

Για την αξιολόγηση των μοντέλων με βάση τις προβλέψεις τους για τα πραγματικά δεδομένα θα χρησιμοποιηθούν τρεις μετρικές που συναντώνται συχνά σε προβλήματα πρόβλεψης της υδρολογίας.

Αρχικά, θα χρησιμοποιηθεί η μετρική Nash – Sutcliffe Efficiency (NSE), η οποία υπολογίζεται με τον τύπο:

$$NSE = 1 - \frac{\sum_{i=1}^n (Q_{o,i} - Q_{p,i})^2}{\sum_{i=1}^n (Q_{o,i} - \bar{Q}_o)^2}$$

όπου:

- $Q_{o,i}$: η παρατηρημένη τιμή στο βήμα i ,
- $Q_{p,i}$: η τιμή της πρόβλεψης στο βήμα i ,
- \bar{Q}_o : η μέση τιμή των παρατηρήσεων.

Η τιμή του δείκτη NSE κυμαίνεται από $-\infty$ έως 1. Μια τιμή NSE ίση με 1 υποδεικνύει απόλυτη ταύτιση μεταξύ της παρατηρημένης και της προβλεπόμενης τιμής, ενώ αρνητικές τιμές υποδεικνύουν ότι το προγνωστικό μοντέλο είναι λιγότερο ακριβές από τη μέση τιμή των παρατηρήσεων. Έτσι, επιτυγχάνεται μία ποσοτικοποίηση της ακρίβειας των προβλέψεων.

Παράλληλα, γίνεται χρήση και μίας τροποποιημένης μετρικής LNSE, στην οποία τα δεδομένα των προβλέψεων και των παρατηρήσεων μετασχηματίζονται λογαριθμικά. Η συγκεκριμένη τροποποίηση επιτρέπει την καλύτερη αξιολόγηση των μοντέλων όσον αφορά την πρόβλεψη των χαμηλών ροών, κάτι ιδιαίτερα σημαντικό στη συγκεκριμένη περίπτωση. Ο τύπος της είναι:

$$\text{LNSE} = 1 - \frac{\sum_{i=1}^n (\log Q_{o,i} - \log Q_{p,i})^2}{\sum_{i=1}^n (\log Q_{o,i} - \overline{\log Q_o})^2}$$

Τέλος, χρησιμοποιείται άλλη μία συνήθης μετρική της υδρολογίας, η Kling – Gupta Efficiency (KGE), η οποία υπολογίζεται με τον τύπο:

$$\text{KGE} = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2}$$

όπου:

- r , ο συντελεστής συσχέτισης Pearson,
- α , ένας όρος που αντιπροσωπεύει τη μεταβλητότητα των λαθών των προβλέψεων και ισούται με τον λόγο των τυπικών αποκλίσεων των προβλέψεων και των παρατηρήσεων,
- β , ο λόγος των μέσων τιμών των προβλέψεων και των παρατηρήσεων.

Αντίστοιχα με τον δείκτη NSE, η τιμή του KGE κυμαίνεται από $-\infty$ έως 1, με την τιμή 1 να υποδεικνύει τέλεια συμφωνία μεταξύ των παρατηρήσεων και των προβλέψεων και τιμές μικρότερες από 0 να υποδεικνύουν κακή απόδοση του μοντέλου. Η μετρική αυτή είναι σημαντική καθώς συνδυάζει τρεις διαφορετικές παραμέτρους: τη συσχέτιση, τη μεροληψία και τη μεταβλητότητα μεταξύ των τιμών των προβλέψεων και των παρατηρήσεων.

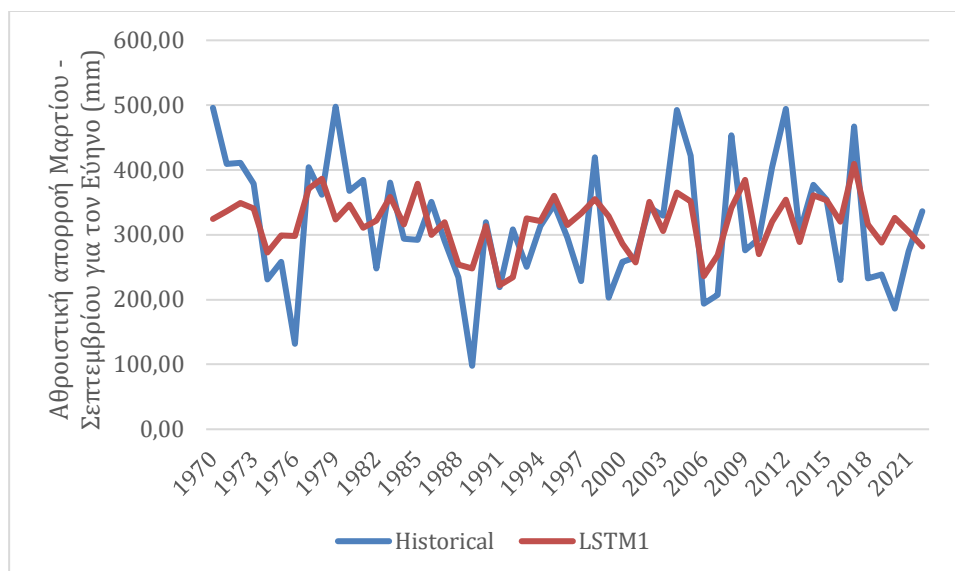
6.2. Αξιολόγηση των μοντέλων για τον Εύηνο

Πίνακας 6.1: Μετρικές επίδοσης των μοντέλων για τον Εύηνο.

	Historical	LSTM1	LSTM2	LSTM3	LSTM4	Linear
Μέσος όρος	318.05	318.99	322.57	310.67	294.92	309.44
Τυπική απόκλιση	93.60	41.02	32.26	33.43	35.64	3.97
Συσχέτιση	1.00	0.58	0.60	0.54	0.54	0.66
MSE	0	5959	6219	6544	6978	8363
NSE	1.00	0.32	0.29	0.25	0.20	0.05
LNSE	1.00	0.29	0.24	0.26	0.26	0.04
KGE	1.00	0.30	0.23	0.21	0.22	-0.02

Όπως φαίνεται από τις παραπάνω μετρικές, τις καλύτερες επιδόσεις στην πρόβλεψη των πραγματικών δεδομένων τις έχει το μοντέλο LSTM1, στο οποίο η τυποποίηση των μεταβλητών έχει γίνει με χρήση της κανονικής κατανομής. Αυτό είναι ένα λογικό αποτέλεσμα, καθώς τα δεδομένα αυτής της λεκάνης φαίνεται είναι κατανεμημένα με τρόπο που παρομοιάζει μερικώς στην Κανονική Κατανομή. Εντύπωση προκαλεί το γεγονός πως το LSTM1 έχει καλύτερη επίδοση στα πραγματικά δεδομένα έναντι του LSTM2, καθώς οι υπερπαραμέτροι του LSTM2 είναι θεωρητικά καλύτερες, για την πρόβλεψη των συνθετικών δεδομένων όμως.

Όπως φαίνεται παρακάτω, το LSTM1 μπορεί να εντοπίζει ικανοποιητικά τις τάσεις της πραγματικής χρονοσειράς, αντιμετωπίζει όμως πρόβλημα με την πρόβλεψη της έντασης των πιο ακραίων φαινομένων, είτε αυτά είναι περίοδοι ξηρασίας, είτε περίοδοι περισσής απορροής.



Εικόνα 6.1: Προβλέψεις της απορροής Μαρτίου – Σεπτεμβρίου από το LSTM 1 (τυποποίηση με Κανονική Κατανομή) και πραγματικά δεδομένα για τον Εύρηνο (mm).

6.3. Αξιολόγηση των μοντέλων για τον Μόρνο

Πίνακας 6.2: Μετρικές επίδοσης των μοντέλων για τον Μόρνο.

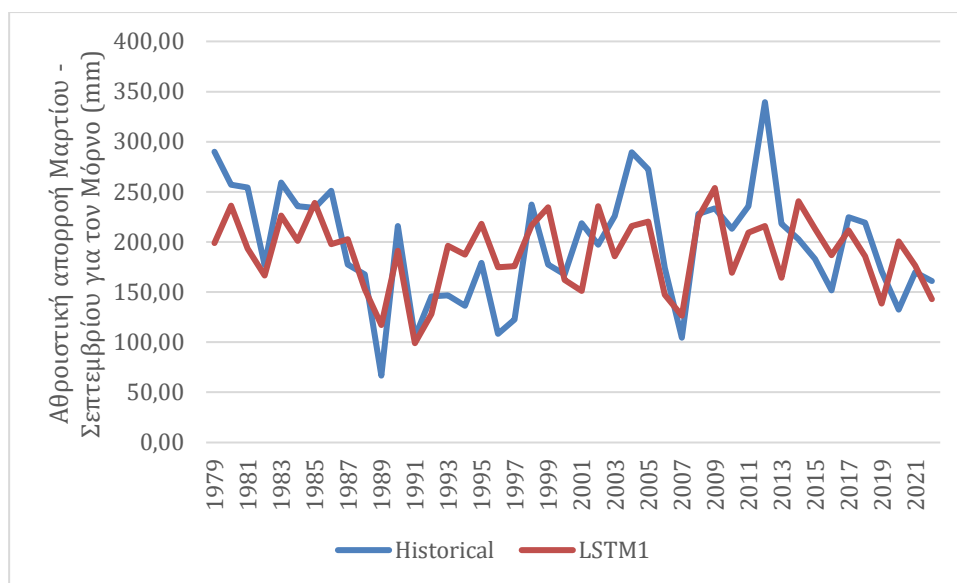
	Historical	LSTM1	LSTM2	LSTM3	LSTM4	Linear
Μέσος όρος	197.22	189.39	197.31	191.61	185.74	166.67
Τυπική απόκλιση	55.96	36.17	30.74	23.70	22.81	7.60
Συσχέτιση	1.00	0.61	0.60	0.62	0.63	0.61
MSE	0	2013	1675	1721	1796	2994
NSE	1.00	0.36	0.36	0.34	0.31	-0.15
LNSE	1.00	0.44	0.40	0.38	0.38	-0.01
KGE	1.00	0.48	0.40	0.31	0.30	0.04

Όπως φαίνεται από τις παραπάνω μετρικές, τις καλύτερες επιδόσεις στην πρόβλεψη των πραγματικών δεδομένων έχει και πάλι το μοντέλο LSTM1, στο οποίο η τυποποίηση των μεταβλητών έχει γίνει με χρήση της κανονικής κατανομής. Αυτό είναι ένα λογικό αποτέλεσμα,

καθώς τα δεδομένα αυτής της λεκάνης φαίνεται είναι κατανεμημένα με τρόπο που παρομοιάζει μερικώς στην Κανονική Κατανομή.

Όπως και στον Εύηνο, είναι παράξενο πως το LSTM1 έχει καλύτερη επίδοση στα πραγματικά δεδομένα έναντι του LSTM2, καθώς οι υπερπαραμέτροι του LSTM2 είναι θεωρητικά καλύτερες, για την πρόβλεψη των συνθετικών δεδομένων όμως. Αυτό ενισχύει την υπόθεση πως είναι μάλλον καλύτερα το μοντέλο να εκπαιδεύεται με λιγότερο πιστό τρόπο στα συνθετικά δεδομένα. Αυτό του δίνει μεγαλύτερη δυνατότητα να ανταποκρίνεται ορθά και να μπορεί να προσαρμόζεται καλύτερα σε νέα δεδομένα, χωρίς να προσκολλάται πολύ σε μοτίβα που τυχόν έχει δει κατά την εκπαίδευσή του.

Όπως φαίνεται παρακάτω, το LSTM1 μπορεί να εντοπίζει πολύ ικανοποιητικά τις τάσεις της πραγματικής χρονοσειράς (πετυχαίνοντας αρκετά υψηλές τιμές σε όλες τις μετρικές), αντιμετωπίζει όμως και αυτό μικρό πρόβλημα με την πρόβλεψη της έντασης των πιο ακραίων φαινομένων, είτε αυτά είναι περίοδοι ξηρασίας, είτε περίοδοι περισσής απορροής.



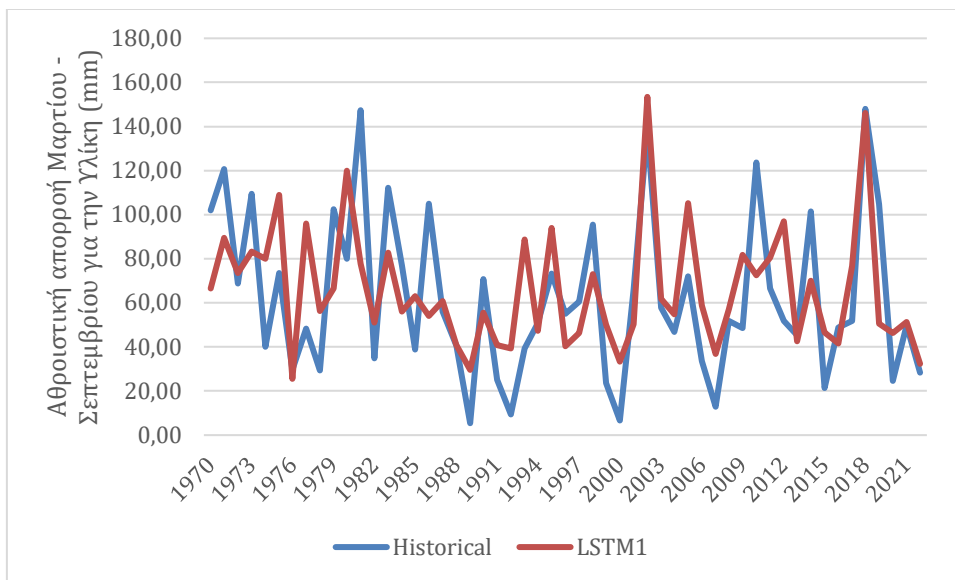
Εικόνα 6.2: Προβλέψεις της απορροής Μαρτίου – Σεπτεμβρίου από το LSTM 1 (τυποποίηση με Κανονική Κατανομή) και πραγματικά δεδομένα για τον Μόρνο (mm).

6.4. Αξιολόγηση των μοντέλων για την Υλίκη

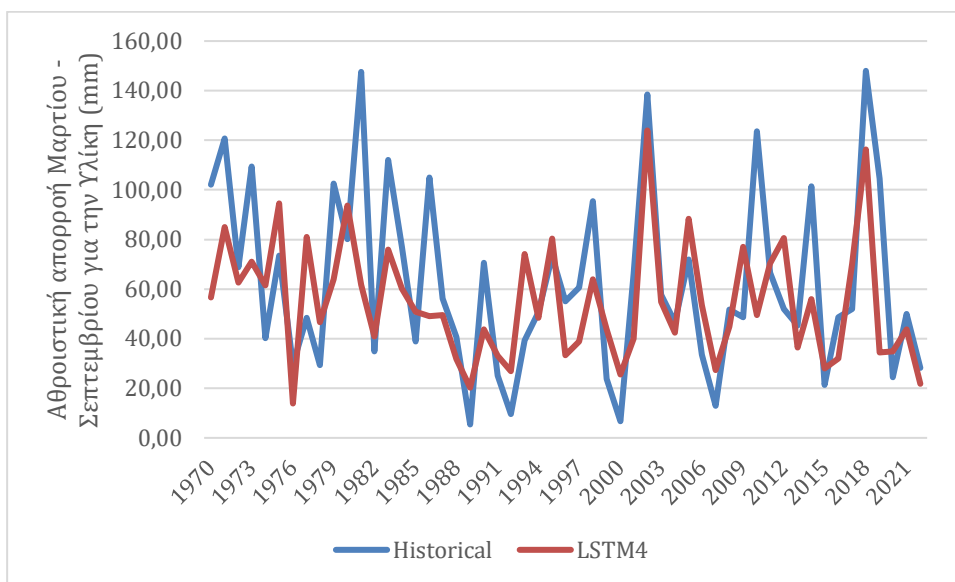
Πίνακας 6.3: Μετρικές επίδοσης των μοντέλων για την Υλίκη.

	Historical	LSTM1	LSTM2	LSTM3	LSTM4	Linear
Μέσος όρος	62.75	66.12	63.04	53.22	54.89	34.04
Τυπική απόκλιση	36.09	27.11	22.34	22.47	23.71	13.87
Συσχέτιση	1.00	0.63	0.62	0.61	0.63	0.64
MSE	0	812	798	908	840	1680
NSE	1.00	0.38	0.39	0.30	0.36	-0.29
LNSE	1.00	0.37	0.36	0.41	0.47	-0.43
KGE	1.00	0.55	0.46	0.44	0.48	0.15

Με βάση τις μετρικές, το καλύτερο μοντέλο είναι και πάλι το LSTM1, στο οποίο η τυποποίηση των μεταβλητών έχει γίνει με χρήση της κανονικής κατανομής. Όμως, το μοντέλο LSTM4 παρουσιάζει πολύ καλή επίδοση στη μετρική LNSE, και συνεπώς στην πρόβλεψη των χαμηλών ροών, κάτι που αποτελεί μία από τις βασικότερες προκλήσεις της πρόβλεψης απορροής. Γενικά, λόγω της γεωμορφολογίας της Υλίκης, όλα τα μοντέλα φαίνεται να πετυχαίνουν αρκετά καλύτερες επιδόσεις σε σχέση με τις άλλες λεκάνες.



Εικόνα 6.3: Προβλέψεις της απορροής Μαρτίου – Σεπτεμβρίου από το LSTM 1 (τυποποίηση με Κανονική Κατανομή) και πραγματικά δεδομένα για την Υλίκη (mm).



Εικόνα 6.4: Προβλέψεις της απορροής Μαρτίου – Σεπτεμβρίου από το LSTM 4 (τυποποίηση με log) και πραγματικά δεδομένα για την Υλίκη (mm).

Με βάση τις μετρικές, το καλύτερο μοντέλο είναι και πάλι το LSTM1, στο οποίο η τυποποίηση των μεταβλητών έχει γίνει με χρήση της κανονικής κατανομής. Όμως, το μοντέλο LSTM4 παρουσιάζει πολύ καλή επίδοση στη μετρική LNSE, και συνεπώς στην πρόβλεψη των χαμηλών ροών, κάτι που αποτελεί μία από τις βασικότερες προκλήσεις της πρόβλεψης απορροής. Γενικά, λόγω της γεωμορφολογίας της Υλίκης, όλα τα μοντέλα φαίνεται να πετυχαίνουν αρκετά καλύτερες επιδόσεις σε σχέση με τις άλλες λεκάνες.

Όπως επιβεβαιώνεται και οπτικά από τις **Εικόνες 6.3** και **6.4**, το μοντέλο LSTM1 έχει γενικά πολύ καλές επιδόσεις, εντοπίζοντας με ακρίβεια και τις περιόδους υψηλής απορροής και μπορώντας να προβλέψει σε πολύ καλό βαθμό την έντασή τους. Αντίθετα, στις ξηρές περιόδους, μπορεί να προβλέψει την τάση αλλά αντιμετωπίζει μερική δυσκολία στην πρόβλεψη της έντασης της χαμηλής απορροής.

Αυτή την αδυναμία έρχεται να καλύψει το μοντέλο LSTM4, το οποίο έχει γενικότερα λίγο χειρότερη επίδοση, εντοπίζοντας όμως με πολύ μεγάλη ακρίβεια και τις περιόδους κατά τις οποίες υπάρχει ξηρασία, αλλά και την ένταση αυτής, όπως επιβεβαιώνεται και από το συγκριτικά πολύ υψηλό LNSE.

Οι παραπάνω επιδόσεις ανοίγουν τον δρόμο για μία υβριδική προσέγγιση που να συνδυάζει αυτά τα δύο αυτά μοντέλα, ώστε να μπορούν να προβλεφθούν με πολλή μεγάλη ακρίβεια όλες οι περίοδοι, είτε αυτές έχουν χαμηλή, είτε μέση, είτε υψηλή απορροή. Βέβαια, η λεκάνη του Βοιωτικού Κηφισού είναι αρκετά ιδιαίτερη, οπότε είναι λογικό τα μοντέλα να παρουσιάζουν αρκετά διαφορετικές συμπεριφορές και επιδόσεις σε σχέση με άλλες λεκάνες.

Μία εύκολη προσέγγιση που μπορεί να εφαρμοστεί είναι η εξής:

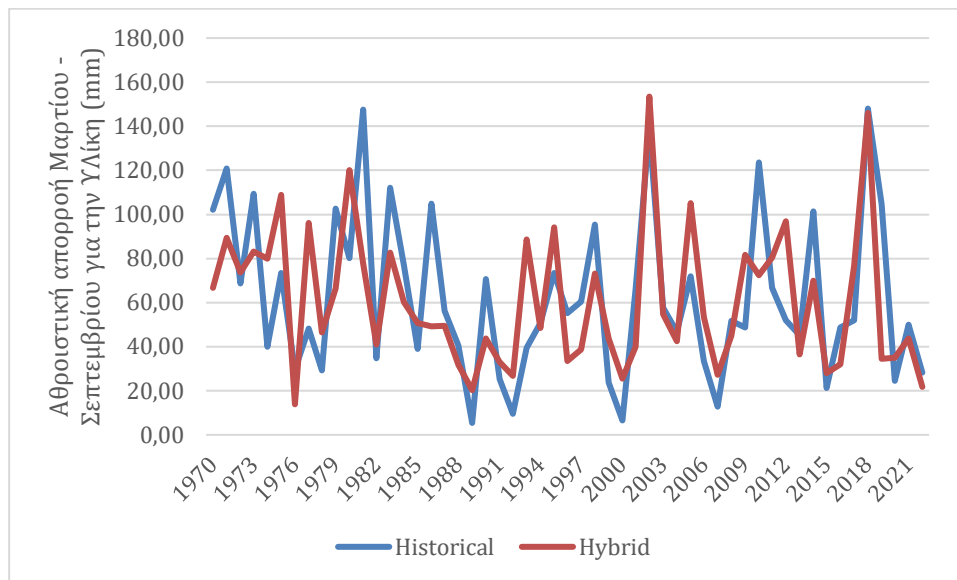
Σε κάθε βήμα γίνεται σύγκριση μεταξύ της τιμής του LSTM1 και του LSTM4 και αν $LSTM1 \geq \mu(LSTM1)$, τότε επιλέγεται η τιμή LSTM1, αλλιώς επιλέγεται η τιμή του LSTM4.

Τα αποτελέσματα που προκύπτουν από ένα τέτοιο απλό υβριδικό μοντέλο είναι πολύ ικανοποιητικά, καθώς η γενικότερη καλή επίδοση του LSTM1 συνδυάζεται με τις πολύ καλές επιδόσεις του LSTM4 στις χαμηλές απορροές.

Πίνακας 6.4: Μετρικές επίδοσης των μοντέλων για την Υλίκη, με την προσθήκη του υβριδικού μοντέλου.

	Historical	LSTM1	LSTM2	LSTM3	LSTM4	Linear	Hybrid
Μέσος όρος	62.75	66.12	63.04	53.22	54.89	34.04	61.03
Τυπική απόκλιση	36.09	27.11	22.34	22.47	23.71	13.87	30.99
Συσχέτιση	1.00	0.63	0.62	0.61	0.63	0.64	0.65
MSE	0	812	798	908	840	1680	813
NSE	1.00	0.38	0.39	0.30	0.36	-0.29	0.38
LNSE	1.00	0.37	0.36	0.41	0.47	-0.43	0.47
KGE	1.00	0.55	0.46	0.44	0.48	0.15	0.62

Επιτυγχάνεται ένα βελτιωμένο και αρκετά υψηλό KGE, διατηρώντας ταυτόχρονα το πολύ καλό LNSE του LSTM4.



Εικόνα 6.5: Προβλέψεις της απορροής Μαρτίου – Σεπτεμβρίου από το μοντέλο Hybrid και πραγματικά δεδομένα για την Υλίκη (mm).

Οι τιμές των μετρικών επιβεβαιώνονται και από την **Εικόνα 6.5**, όπου φαίνονται οι συνολικά πολύ καλές επιδόσεις, ιδιαίτερα στις πιο ξηρές περιόδους, που είναι και το μεγαλύτερο ζητούμενο.

6.5. Συνολική αποτίμηση των επιδόσεων των μοντέλων

Σε όλες τις λεκάνες απορροής συνολικά καλύτερο μοντέλο είναι το LSTM1, γεγονός που δείχνει πως η τυποποίηση με βάση την κανονική κατανομή φαίνεται να είναι η ιδανική για το συγκεκριμένο πρόβλημα πρόβλεψης. Ταυτόχρονα, όπως αναφέρθηκε και στην ανάλυση των επιδόσεων των μοντέλων για τον Μόρνο, φαίνεται πως η βελτιστοποίηση των υπερπαραμέτρων ως το τα συνθετικά δεδομένα επηρεάζει αρνητικά τις επιδόσεις των μοντέλων. Συνεπώς, αυτά πρέπει να εκπαιδεύονται με λιγότερο αυστηρή επικύρωση, προκειμένου να μην γίνεται υπερπροσαρμογή και να μπορούν να γενικεύουν καλύτερα σε νέα δεδομένα.

Βασικός σκοπός της συγκεκριμένης έρευνας ήταν η αποφυγή της χρήσης οποιουδήποτε πραγματικού δεδομένου κατά τη διαδικασία της εκπαίδευσης – επικύρωσης, καθώς είναι πολύ σημαντικό να δημιουργηθούν και να εκπαιδευθούν μοντέλα που δε θα εξαρτώνται από τα λιγοστά, και πολλές φορές αναξιόπιστα, πραγματικά δεδομένα.

7. Συμπεράσματα και προτάσεις

7.1. Συμπεράσματα

Και στις τρεις λεκάνες, το μοντέλο LSTM κατάφερε να έχει αρκετά καλές επιδόσεις, παρά την πολύ δύσκολη φύση του προβλήματος πρόβλεψης με την χρήση συνθετικών δεδομένων. Αποδεικνύεται πως τα νευρωνικά δίκτυα γενικότερα αλλά και τα μοντέλα LSTM πιο συγκεκριμένα μπορούν να διαδραματίσουν σημαντικό ρόλο στο μέλλον στο πρόβλημα της πρόβλεψης απορροής σε υδρολογικές λεκάνες.

Στον Εύηνο και στον Μόρνο το LSTM μπορεί να εκτιμήσει ικανοποιητικά, ιδιαίτερα σε σύγκριση με κλασικά ντετερμινιστικά μοντέλα, τη συμπεριφορά της απορροής και τη μεταβλητότητά της. Καταφέρνει να μοντελοποιήσει τη σύνθετη και έντονα μη γραμμική (όπως φαίνεται από τις προβλέψεις του απλού γραμμικού μοντέλου) συμπεριφορά του υδρολογικού κύκλου και να εντοπίσει τις έντονες διακυμάνσεις, αυτές δηλαδή που επηρεάζουν σε μεγαλύτερο βαθμό τα συστήματα διαχείρισης υδατικών πόρων.

Στη λεκάνη του Βοιωτικού Κηφισού / Υλίκης, το LSTM αποδίδει εξαιρετικά, ειδικά με την χρήση του απλού υβριδικού μοντέλου για την αξιοποίηση δύο διαφορετικών μοντέλων και τρόπων τυποποίησης των δεδομένων εισόδου και εξόδου αυτών. Έτσι, αποδεικνύεται πως ο συνδυασμός πιο εξελιγμένων προβλεπτικών μεθόδων όπως τα LSTM με πιο κλασικές μεθόδους, όπως η δημιουργία υβριδικών μοντέλων, μπορεί να αποδώσει σε πολύ μεγάλο βαθμό και να ενισχύσει τις ήδη καλές επιδόσεις των LSTM. Μία σημαντική πληροφορία που πρέπει να ληφθεί υπόψη για τη συγκεκριμένη λεκάνη είναι ότι έχει ιδιαίτερα χαρακτηριστικά όσον αφορά το κομμάτι της καρστικότητας. Αυτό με τη σειρά του μπορεί να επηρεάσει την αποθηκευτικότητα της λεκάνης και τελικά τη συμπεριφορά της απορροής της, εξομαλύνοντάς τη μερικώς. Αντίθετα, οι άλλες δύο λεκάνες που μελετήθηκαν παρουσιάζουν αρκετά λιγότερο ομαλή συμπεριφορά, γεγονός που πιθανότατα σχετίζεται με την λιγότερο καρστική φύση τους.

7.2. Ερευνητικές προτάσεις

Όπως αποδείχθηκε από τις επιδόσεις τους, τα μοντέλα LSTM έχουν αρκετά μεγάλο μέλλον στην πρόβλεψη χρονοσειρών απορροής, ειδικότερα σε προβλήματα όπως αυτό που μελετήθηκε, η λύση των οποίων είναι σημαντική για την υποστήριξη των συστημάτων λήψης αποφάσεων για τη διαχείριση των υδατικών πόρων μίας περιοχής όπως είναι η Αττική. Σε συνδυασμό με την χρήση μεθόδων επαύξησης των υπαρχόντων δεδομένων, όπως είναι η παραγωγή συνθετικών δεδομένων που εφαρμόστηκε εδώ, αλλά και με τη συνεργασία τους με άλλα μοντέλα και στατιστικές μεθόδους, μπορούν να αποτελέσουν μέρος συστημάτων με εξαιρετικές επιδόσεις.

Είναι δυνατός ο συνδυασμός τους με πιο κλασικά μοντέλα πρόβλεψης, με πιο σύνθετα μοντέλα βροχής – απορροής, προκειμένου να αξιοποιηθούν με πιο ορθολογικό και αναλυτικό τρόπο τα γεωμορφολογικά και υδρολογικά χαρακτηριστικά των λεκανών, αλλά και με εξελιγμένες στατιστικές μεθόδους, όπως είναι οι συζεύξεις (copulas). Ταυτόχρονα, υπάρχει αρκετό περιθώριο για περαιτέρω βελτιστοποίηση των μοντέλων που χρησιμοποιούνται, με διαφορετικούς τρόπους τυποποίησης, εισαγωγής και επαύξησης των δεδομένων εκπαίδευσης, αλλά και για αλλαγές στην ανατομία και στις υπερπαραμέτρους των μοντέλων. Σε αυτό το πλαίσιο εντάσσεται και η δημιουργία υβριδικών μοντέλων που αποτελούνται από πολλά νευρωνικά δίκτυα, με την χρήση ταξινόμησης για αρχικό εντοπισμό υγρών και ξηρών περιόδων και τον συνδυασμό τους με μοντέλα παλινδρόμησης, όπως είναι τα LSTM για την πιο ακριβή πρόβλεψη χρονοσειρών. Επίσης, εφικτός είναι και ο συνδυασμός πολλαπλών LSTM, όπως αποδείχθηκε στην περίπτωση της Υλίκης, που μπορεί να επιφέρει σημαντικά βελτιωμένα αποτελέσματα χωρίς σημαντική αύξηση των απαιτήσεων σε δεδομένα ή πολύ μεγαλύτερο υπολογιστικό κόστος.

Σε συνέχεια της παρατήρησης που έγινε για τη λεκάνη του Βοιωτικού Κηφισού / Υλίκης όσον αφορά την καρστικότητα, διαφαίνεται η σημασία για περαιτέρω διερεύνηση χαρακτηριστικών όπως η καρστικότητα και η αποθηκευτικότητα λεκανών σε σχέση με την απόδοση προβλεπτικών μοντέλων απορροής αλλά και η αξιοποίησή τους στη διαδικασία της πρόβλεψης, είτε ως απλές παραμέτρους, είτε στο πλαίσιο της χρήσης νευρωνικών δικτύων τύπου PINN (Physics Informed Neural Network) για πιο αυστηρή μοντελοποίηση της απορροής.

Τέλος, αξίζει να μελετηθεί και η επιρροή της εποχιακής τήξης του χιονιού στην απορροή των λεκανών, καθώς μπορεί να επηρεάσει σημαντικά συγκεκριμένες χρονιές.

Βιβλιογραφικές Αναφορές

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Zheng, X., 2016. TensorFlow: A system for large-scale machine learning. Συνεδρίαση της 12ης Συνάντησης του Οργανισμού USENIX για Υπολογιστικά Συστήματα και Μηχανική Μάθηση (MLSys '16). Διαθέσιμο στο: <https://arxiv.org/abs/1605.08695>.
2. Bishop, C.M., 2006. Pattern Recognition and Machine Learning. New York: Springer.
3. Goodfellow, I., Bengio, Y. and Courville, A., 2017. Deep Learning. MIT Press, Cambridge, MA.
4. Gupta, H.V., Kling, H., Yilmaz, K.K., & Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1–2), σελ. 80–91. Διαθέσιμο στο: <https://doi.org/10.1016/j.jhydrol.2009.08.003>.
5. Hochreiter, S. & Schmidhuber, J., 1997. Long Short-Term Memory. *Neural Computation*, 9(8), σελ. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.
6. Kratzert, F., Klotz, D., Brenner, C., Schulz, K. and Herrnegger, M., 2018. Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, 22(11), σελ. 6005–6022. Διαθέσιμο στο: <https://hess.copernicus.org/articles/22/6005/2018/>. DOI: 10.5194/hess-22-6005-2018.
7. Lamontagne, J.R., Barber, C.A. & Vogel, R.M., 2020. Improved estimators of model performance efficiency for skewed hydrologic data. *Water Resources Research*, 56, e2020WR027101. <https://doi.org/10.1029/2020WR027101>.
8. Li, Y., Hu, D., Bai, J. και Qiu, Y., 2017. Multi-step Ahead Time Series Forecasting for Different Data Patterns Based on LSTM Recurrent Neural Network. Στα πρακτικά του συνεδρίου: 2017 14th Web Information Systems and Applications Conference (WISA), Liuzhou, China, σελ. 305-310. DOI: 10.1109/WISA.2017.25.
9. McKinney, W., 2010. Data Structures for Statistical Computing in Python. DOI: 10.25080/majora-92bf1922-00a.
10. Nash, J.E. and Sutcliffe, J.V., 1970. ‘River flow forecasting through conceptual models. Part I — A discussion of principles’, *Journal of Hydrology*, 10(3), σελ. 282–290.

11. Nielsen, M., 2015. Neural Networks and Deep Learning. Διαθέσιμο στο: <http://neuralnetworksanddeeplearning.com>.
12. Pedregosa, F. et al., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res..
13. Ruder, S., 2017. An overview of gradient descent optimization algorithms. arXiv. Διαθέσιμο στο: <https://arxiv.org/abs/1609.04747>.
14. Schuster, M. & Paliwal, K.K., 1997. Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11), σελ. 2673–2681. DOI: 10.1109/78.650093.
15. Schaefli, B. and Gupta, H.V., 2007. Do Nash values have value? Hydrol. Process., 21: 2075-2080. DOI: <https://doi.org/10.1002/hyp.6825>.
16. Van Rossum, G., 1995. Python 2.7 tutorial. Διαθέσιμο στο: <https://docs.python.org/2/tutorial/>.
17. Ευστρατιάδης, Α. και Σακκή, Γ.-Κ., 2024. Αναλύσεις υδατικών ισοζυγίων και απολογιστική έκθεση υδρολογικού έτους 2022-23: Εκσυγχρονισμός της διαχείρισης του συστήματος των υδατικών πόρων ύδρευσης της Αθήνας – Αναθεώρηση. Τομέας Υδατικών Πόρων και Περιβάλλοντος – Εθνικό Μετσόβιο Πολυτεχνείο, Ιανουάριος, 2024.

Παράρτημα

Παρακάτω παρατίθεται ο κώδικας που χρησιμοποιήθηκε σε Python 3.10. Για να γίνει σωστή φόρτωση των δεδομένων απαιτείται να βρίσκονται ήδη σε κατάλληλη μορφή σε αρχεία xlsx, όπως είναι εμφανές και από τον κώδικα. Η τροποποίηση του για αλλαγή της λεκάνης μελέτης είναι εύκολη, καθώς, όπως και η τροποποίηση του πλέγματος υπερπαραμέτρων που πρόκειται να δοκιμαστεί, απαιτείται να γίνει μόνο σε ένα σημείο, ακριβώς πριν από την έναρξη της διαδικασίας εκπαίδευσης του μοντέλου. Τέλος, η μέθοδος τυποποίησης των δεδομένων εισόδου επιλέγεται κατά τη φόρτωσή τους.

```
import numpy as np
import pandas as pd
from keras.models import Sequential
from keras.layers import LSTM, Dense, Dropout, Bidirectional
from keras.optimizers import Adam
from keras.callbacks import EarlyStopping
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, MinMaxScaler, RobustScaler
from sklearn.metrics import mean_squared_error

# Set random seed for reproducibility
seed = 13
np.random.seed(seed)

# Scalers dictionary for easy reference
scaler_options = {
    'standard': StandardScaler,
    'log': None # Custom log scaling
}

# Model creation function
def create_model(dropout, learning_rate):
    """
    Function to create a Bidirectional LSTM model.

    Args:
    - dropout: Dropout rate to avoid overfitting.
    - learning_rate: Learning rate for the optimizer.
```

```

Returns:
- Compiled Keras model.
"""
model = Sequential()
model.add(Bidirectional(LSTM(128, return_sequences=True, input_shape=(5,
2))))
model.add(Dropout(dropout))
model.add(Bidirectional(LSTM(64, return_sequences=False)))
model.add(Dense(7, activation='linear')) # Output layer for 7 targets
model.compile(optimizer=Adam(learning_rate=learning_rate), loss='mse',
metrics=['mae'])
return model

# Function to read and preprocess data
def read_data(basin, scaler_name):
    """
    Function to read, preprocess, and scale data for model training.

    Args:
    - basin: Name of the basin (e.g., 'Ylikh').
    - scaler_name: The type of scaler to apply (e.g., 'standard', 'log').

    Returns:
    - Processed and scaled features and targets.
    """
    # Read synthetic data for training
    rain_synth = pd.read_excel('Synthetic.xlsx', sheet_name=f'{basin} Rain',
usecols=range(1, 13))
    runoff_synth = pd.read_excel('Synthetic.xlsx', sheet_name=f'{basin} Runoff',
usecols=range(1, 13))

    # Combine first 5 columns as features (X) and last 7 columns as targets (y)
    X_train = np.hstack((rain_synth.iloc[:, :5].values, runoff_synth.iloc[:,
:5].values))
    y_train = runoff_synth.iloc[:, 5:].values

    # Initialize the target scaler (if required)
    scaler_y = None

    # Apply scaling to features
    if scaler_name == 'log':
        X_scaled = np.log1p(X_train)
        scaler_X = None
    else:
        scaler_X = scaler_options[scaler_name]()

```

```

X_scaled = scaler_X.fit_transform(X_train)

# Apply log scaling to targets
y_scaled = np.log1p(y_train)

# Reshape features for LSTM input (shape should be (batch_size, 5, 2))
feature_1 = X_scaled[:, :5] # First 5 columns for rain
feature_2 = X_scaled[:, 5:] # Next 5 columns for runoff

# Stack the two features along the last axis
reshaped_X = np.stack((feature_1, feature_2), axis=-1) # Shape becomes
(batch_size, 5, 2)

return reshaped_X, y_scaled, scaler_X, scaler_y

# Early stopping callback to prevent overfitting
early_stopping = EarlyStopping(monitor='val_loss', patience=10,
restore_best_weights=True)

# Custom grid search function to find best model parameters
def custom_grid_search(basin, param_grid):
    """
    Function to perform custom grid search to find the best model parameters.

    Args:
    - basin: Name of the basin.
    - param_grid: Dictionary containing hyperparameters for grid search.

    Returns:
    - Best model and corresponding parameters.
    """
    best_model = None
    best_params = None
    best_val_mse = np.inf # Initialize with a very large value to track best
model

# Load synthetic data for training and historical data for validation
for scaler in ['log']: #['log', 'normal']
    X, y, scaler_X, scaler_y = read_data(basin, scaler)
    X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.2,
shuffle=False)

    # Loop over hyperparameter grid
    for dropout in param_grid['dropout']:

```

```

for learning_rate in param_grid['learning_rate']:
    for batch_size in param_grid['batch_size']:
        # Create the model with current hyperparameters
        model = create_model(dropout, learning_rate)
        # Train the model with early stopping
        model.fit(
            X_train, y_train,
            validation_data=(X_val, y_val),
            epochs=100,
            batch_size=batch_size,
            callbacks=[early_stopping],
            verbose=0
        )

        # Evaluate on validation data
        val_preds = model.predict(X_val)
        val_mse = mean_squared_error(y_val, val_preds)

        # Check if this model is the best so far
        if val_mse < best_val_mse:
            best_val_mse = val_mse
            best_model = model
            best_params = {'dropout': dropout, 'learning_rate':
learning_rate, 'batch_size': batch_size, 'scaler': scaler}

    return best_model, best_params

# Define parameter grid for hyperparameter tuning / different combinations for
different basins
param_grid = {
    'dropout': [0], #[0, 0.1, 0.2, 0.4, 0.5]
    'learning_rate': [0.0005, 0.001, 0.005, 0.01],
    'batch_size': [2, 4, 8, 16, 32, 64]
}

# Train and evaluate for each basin
best_models = {}
for basin in ['Ylikh']: #['Evinos', 'Mornos', 'Ylikh']
    print(f"Training for basin: {basin}")

    # Perform custom grid search to find best model
    best_model, best_params = custom_grid_search(basin, param_grid)

    best_models[basin] = {'model': best_model, 'params': best_params}
    print(f"Best parameters for {basin}: {best_params}")

```

```

# Predict with the best model for each basin
for basin, model_info in best_models.items():
    model = model_info['model']
    best_params = model_info['params']
    scaler = model_info['params']['scaler']
    _, _, scaler_X, scaler_y = read_data(basin, scaler)

    # Load historical data for prediction (validation set)
    rain_hist = pd.read_excel('historical_data.xlsx', sheet_name=f'{basin} Rain',
usecols=range(1, 13))
    runoff_hist = pd.read_excel('historical_data.xlsx', sheet_name=f'{basin}
Runoff', usecols=range(1, 13))

    # Combine first 5 columns as features (X) and last 7 columns as targets (y)
    X_pred = np.hstack((rain_hist.iloc[:, :5].values, runoff_hist.iloc[:,
:5].values))
    y_pred = runoff_hist.iloc[:, 5:].values

    # Apply scaling to features using the same scaler as training data
    if 'log' in model_info['params']['scaler']:
        X_pred_scaled = np.log1p(X_pred)
    else:
        X_pred_scaled = scaler_X.transform(X_pred)

    # Reshape features for LSTM input
    feature_pred_1 = X_pred_scaled[:, :5]
    feature_pred_2 = X_pred_scaled[:, 5:]
    reshaped_pred_X = np.stack((feature_pred_1, feature_pred_2), axis=-1)

    # Make predictions
    predictions = model.predict(reshaped_pred_X)
    predictions_scaled = np.expm1(predictions)

    # Save predictions to Excel
    predictions_df = pd.DataFrame(predictions_scaled, columns=[f"Output_{i+1}"
for i in range(predictions_scaled.shape[1])])
    predictions_df.to_excel(f"{basin}_predictions_output.xlsx", index=False)
    print(best_params)
    print(f"Predictions for {basin} saved to {basin}_predictions_output.xlsx")

```