



## ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Εργαστήριο Τεχνητής Νοημοσύνης και Συστημάτων Μάθησης

ΔΠΜΣ: Επιστήμη Δεδομένων και Μηχανική Μάθηση

---

### Πρόβλεψη Χρονοσειρών Πλημμυρικών Φαινομένων με Χρήση Βαθέων Νευρωνικών Δικτύων

---



### Μεταπτυχιακή Διπλωματική Εργασία

από

Νικόλαος Α. Τεπετίδης

Πολιτικός Μηχανικός BSc, MEng

**Επιβλέπων:** Αθανάσιος Βουλόδημος, Επίκουρος Καθηγητής ΕΜΠ

Αθήνα, Ιούλιος 2023





## ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

ΔΠΜΣ: Επιστήμη Δεδομένων και Μηχανική Μάθηση

---

### Πρόβλεψη Χρονοσειρών Πλημμυρικών Φαινομένων με Χρήση Βαθέων Νευρωνικών Δικτύων

---

#### Μεταπτυχιακή Διπλωματική Εργασία

από

Νικόλαος Α. Τεπετίδης

Πολιτικός Μηχανικός BSc, MEng

**Επιβλέπων:** Αθανάσιος Βουλόδημος

Επίκουρος Καθηγητής ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 12η Ιουλίου 2023.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....  
Αθανάσιος Βουλόδημος

.....  
Κωνσταντίνος Καραντζαλος

.....  
Βησσαρίων Παπαδόπουλος

Επίκουρος Καθηγητής ΕΜΠ

Καθηγητής ΕΜΠ

Καθηγητής ΕΜΠ

Αθήνα, Ιούλιος 2023

Εικόνα εξωφύλλου δημιουργήθηκε από: Leonardo.ai  
Cover Page Illustration Generated by: Leonardo.ai

**Νικόλαος Α. Τεπετίδης**

Πολιτικός Μηχανικός Ε.Μ.Π., BSc, MEng

Copyright © Νικόλαος Τεπετίδης, 2023

Με επιφύλαξη παντός δικαιώματος

Απαγορεύεται η αντιγραφή, αποθήκευση σε αρχείο πληροφοριών, διανομή, αναπαραγωγή, μετάφραση ή μετάδοση της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό, υπό οποιαδήποτε μορφή και με οποιοδήποτε μέσο επικοινωνίας, ηλεκτρονικό ή μηχανικό, χωρίς την προηγούμενη έγγραφη άδεια του συγγραφέα. Επιτρέπεται η αναπαραγωγή, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν στη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις απόψεις του Τμήματος, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Μεταπτυχιακής Διπλωματικής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή.

Copyright © Nikolaos Tepetidis, 2023

All Rights Reserved

Neither the whole nor any part of this diploma thesis may be copied, stored in a retrieval system, distributed, reproduced, translated, or transmitted for commercial purposes, in any form or by any means now or hereafter known, electronic or mechanical, without the written permission from the author. Reproducing, storing and distributing this thesis for non-profitable, educational or research purposes is allowed, without prejudice to reference to its source and to inclusion of the present text. Any queries in relation to the use of the present thesis for commercial purposes must be addressed to its author. The content of this paper does not necessarily reflect the views of the Department, the Supervisor, or the committee that approved it.

In full awareness of the implications of copyright law, I hereby declare under my signature that I am the sole author of this Master's Thesis, for the completion of which all assistance is fully acknowledged and detailed in this thesis.

(Υπογραφή)

.....  
Νικόλαος Τεπετίδης





---

Νικόλαος Α. Τεπετίδης  
Μεταπτυχιακή Διπλωματική Εργασία  
*Πρόβλεψη Χρονοσειρών Πλημμυρικών Φαινομένων με Χρήση Βαθέων  
Νευρωνικών Δικτύων*  
Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  
Εργαστήριο Τεχνητής Νοημοσύνης και Συστημάτων Μάθησης  
ΔΠΜΣ: Επιστήμη Δεδομένων και Μηχανική Μάθηση  
Αθήνα 2023

---

Nikolaos L. Tepetidis  
Master Thesis  
*Time Series Forecasting of Flood Events Using Deep Neural Networks*  
National Technical University of Athens  
School of Electrical and Computer Engineering  
Artificial Intelligence and Learning Systems Laboratory  
Data Science and Machine Learning  
Athens 2023





*The best qualification of a prophet is to have a good memory.*

George Savile, 8th Baronet



*στην οικογένειά μου,*



# Ευχαριστίες

Το πέρας της παρούσας μεταπτυχιακής διπλωματικής εργασίας σηματοδοτεί την ολοκλήρωση των μεταπτυχιακών μου σπουδών στο Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών (ΔΠΜΣ) στο επιστημονικό πεδίο «Επιστήμη Δεδομένων και Μηχανική Μάθηση» του Εθνικού Μετσόβιου Πολυτεχνείου. Θα ήθελα σε αυτό το σημείο να ευχαριστήσω τους ανθρώπους που συνέβαλλαν στην εκπόνησή της ολοκληρώνοντας μία αρκετά απαιτητική χρονιά.

Καταρχάς θα ήθελα να ευχαριστήσω ιδιαίτερώς τον επιβλέποντα καθηγητή της εργασίας μου κ. Αθανάσιο Βουλόδημο, Επίκουρο Καθηγητή Ε.Μ.Π, ο οποίος με εμπιστεύτηκε από την πρώτη στιγμή, με ενθάρρυνε και με βοήθησε να ασχοληθώ με το θέμα της διπλωματικής. Με την καθοριστική καθοδήγησή του δημιουργήθηκε το πλαίσιο της εργασίας, το οποίο ήταν κομβικό στην επίτευξη του στόχου μου. Αισθάνομαι ιδιαίτερα τυχερός για τη συνεργασία και ευελπιστώ σε μελλοντική συνεργασία.

Η εκπόνηση της εργασίας, εκτός από μεταπτυχιακό φοιτητή, με βρήκε και υποψήφιο διδάκτορα της σχολής Πολιτικών Μηχανικών του ΕΜΠ. Σε αυτήν την ιδίομορφη κατάσταση με βοήθησε η ερευνητική ομάδα με την οποία συνεργάζομαι στα πλαίσια του διδακτορικού μου. Πιο συγκεκριμένα θα ήθελα να ευχαριστήσω τον κ. Δημήτρη Κουτσογιάννη, Καθηγητής ΕΜΠ, ο οποίος είναι και ο επιβλέπων του διδακτορικού μου, τη Θεανώ (Ανυ) Ηλιοπούλου, Διδάκτορα της σχολής Π.Μ. και ΕΔΙΠ ΕΜΠ, τον Παναγιώτη Δημητριάδη, Διδάκτορα της σχολής Π.Μ. και τον Γιάννη Καλογερά, Διδάκτορα της σχολής Π.Μ. Η σημαντική συνεισφορά τους σε όλα τα επίπεδα, οδήγησε και στην παρουσίαση ενός μέρους της παρούσας εργασίας και στο επιστημονικό συνέδριο της EGU στη Βιέννη τον Απρίλιο του 2023. Ήταν μία ωραία εμπειρία που δίχως τη βοήθειά τους ίσως να μην είχα την ευκαιρία να περάσω.

Δεν θα μπορούσα να παραλείψω τους συμφοιτητές που απέκτησα κατά τη διάρκεια των συγκεκριμένων σπουδών και έκαναν πιο εύκολο και ευχάριστο αυτό το ταξίδι. Επιπλέον ευχαριστώ τους φίλους μου για την πολύ καλή παρέα και την υποστήριξη τους, και ειδικότερα τον Γιώργο και τον Δημήτρη οι οποίοι συνέβαλαν άμεσα στην τελική μορφή του παρόντος.

Τέλος θα ήθελα να ευχαριστήσω την οικογένειά μου, τους γονείς μου, Ελένη και Λάζαρο και τον αδελφό μου Κώστα, για την συνεχή υποστήριξη που μου παρείχαν με κάθε τρόπο, σε ότι έχω καταφέρει μέχρι σήμερα.

Νικόλαος Τεπετίδης

Αθήνα,  
Ιούλιος 2023



# Περίληψη

Τα πλημμυρικά φαινόμενα αποτελούν σημαντική απειλή για τις κοινότητες παγκοσμίως, απαιτώντας ακριβή και έγκαιρη πρόβλεψη για τον αποτελεσματικό μετριασμό και την αντιμετώπισή τους. Πρόσφατες μελέτες έχουν αναδείξει τις δυνατότητες των δικτύων μετασχηματιστών (γνωστά ως transformers) στην αύξηση της ικανότητας πρόβλεψης. Ωστόσο, οι κλασικοί μετασχηματιστές παρουσιάζουν διάφορα προβλήματα όπως η χρονική υπολογιστική πολυπλοκότητα και οι υψηλές απαιτήσεις μνήμης, που καθιστούν δύσκολη τη μακροπρόθεσμη πρόβλεψη. Εξετάζεται η συνεισφορά στην πρόβλεψη χρονοσειρών πλημμυρικών φαινομένων με χρήση τεχνικών βαθιάς μάθησης, με ιδιαίτερη έμφαση στην αξιολόγηση της απόδοσης του μοντέλου πληροφόρησης Informer, το οποίο προσπαθεί να ξεπεράσει τις προαναφερθέντες αδυναμίες. Η μελέτη διερευνά τις δυνατότητες πρόβλεψης του μοντέλου Informer σε σύγκριση με παραδοσιακές στατιστικές μεθόδους, στοχαστικά μοντέλα και παραδοσιακά βαθιά νευρωνικά δίκτυα.

Το μοντέλο πληροφόρησης αποτελεί ένα αποδοτικό μοντέλο βασισμένο στην αρχιτεκτονική των μετασχηματιστών που έχει σχεδιαστεί για το πρόβλημα της μακροπρόθεσμης πρόβλεψης. Εισάγει καινοτόμα χαρακτηριστικά για την ενίσχυση της αποδοτικότητας και της ακρίβειας. Αρχικά, εισάγει ένα είδος πιθανοτικού μηχανισμού προσοχής (Probabilistic Sparsity attention mechanism) μειώνοντας την πολυπλοκότητα του υπολογιστικού χρόνου. Επιπλέον χρησιμοποιεί μια τεχνική εξαγωγής χρήσιμης πληροφορίας, που αναδεικνύει την κυρίαρχη προσοχή επιτρέποντας τον αποτελεσματικό χειρισμό εξαιρετικά μεγάλων ακολουθιών εισόδου. Πέρα από αυτό το μοντέλο ενσωματώνει ένα είδος αποκωδικοποιητή (decoder) το οποίο προβλέπει μακροπρόθεσμες ακολουθίες χρονοσειρών με ένα προς-τα-εμπρός πέρασμα. Η ακρίβεια, η αποδοτικότητα, καθώς και τα όρια των προσεγγίσεων αναδεικνύονται μέσα από εφαρμογές που αφορούν σε πραγματικούς ποταμούς. Αξιοποιώντας ημερήσια δεδομένα παροχής του ποταμού River Test στην Αγγλία ως τη βασική εφαρμογή μελέτης, η παρούσα εργασία διεξάγει μια αυστηρή αξιολόγηση της αποτελεσματικότητας των μοντέλων στην καταγραφή των πολύπλοκων χρονικών εξαρτήσεων που υπάρχουν στις χρονοσειρές πλημμυρών. Συμπληρωματικά το δίκτυο πληροφόρησης εφαρμόζεται σε ένα μεγάλο πλήθος ποταμών (>100), προκειμένου να συμπεριλάβει διαφορετικά σύνολα δεδομένων χρονοσειρών από διάφορες τοποθεσίες στο Ηνωμένο Βασίλειο, παρέχοντας πληροφορίες σχετικά με τη δυνατότητα γενίκευσης του μοντέλου.

Τα αποτελέσματα αναδεικνύουν τη σημαντική υπεροχή του μοντέλου Informer έναντι των παραδοσιακών μεθόδων, ειδικά στο πρόβλημα της μακροπρόθεσμης πρόβλεψης. Παρατηρείται ακόμη πως η υψηλή ασυμμετρία χρονοσειρών επηρεάζει αρνητικά την απόδοση του μοντέλου.

**Λέξεις κλειδιά:** πλημμύρες, παροχή ποταμών, πρόβλεψη χρονοσειρών, μακροπρόθεσμη πρόβλεψη, παραδοσιακά μοντέλα, βαθιά νευρωνικά δίκτυα, μετασχηματιστές, μηχανισμός προσοχής, μοντέλο πληροφόρησης.





# Abstract

Flood events pose a significant threat to communities worldwide, demanding accurate and timely forecasting for effective mitigation and response. Recent studies have shown the potential of Transformer networks in increasing prediction capacity. However, classical transformers present several problems such as computational time complexity and high memory requirements, which make Long Sequence Time-Series Forecasting (LSTF) intractable. The contribution to the prediction of time series of flood events using deep learning techniques is examined, with a particular focus on evaluating the performance of the Informer model, which attempts to address the previous issues. The study explores the predictive capabilities of the Informer model compared to statistical methods, stochastic models and traditional deep neural networks.

The Informer model is an efficient transformer-based model designed for LSTF. It introduces several key innovations to enhance efficiency and accuracy. Firstly, it introduces a Probabilistic Sparsity attention mechanism to reduce the computational time complexity. Additionally, it employs a self-attention distilling approach that highlights the most influential attention weights enabling efficient processing of extreme long input sequences. Moreover the model incorporates a generative style decoder that predicts the long time-series sequences at one forward operation. The accuracy, efficiency as well as the limits of the approaches are demonstrated via numerical benchmarks relating to real river applications. Using daily flow data from the River Test in England as the main case study, the thesis conducts a rigorous evaluation of the Informer efficacy in capturing the complex temporal dependencies inherent in flood time series. Among other things, the present work extends its analysis to encompass diverse time series datasets from various locations (>100) in the United Kingdom, providing insights into the generalizability of the Informer.

The results highlight the significant superiority of the Informer model over established forecasting methods, especially regarding the LSTF problem. Furthermore, it is observed that the high asymmetry of time series negatively affects the performance of the Informer network.

**Keywords:** flood, river streamflow, time series forecasting, long sequence time series forecasting (LSTF), traditional models, deep neural networks, transformers, attention mechanism, Informer model.



# Περιεχόμενα

<b>Ευχαριστίες</b>	<b>iii</b>
<b>Περίληψη</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>1 Εισαγωγή</b>	<b>1</b>
1.1 Πλημμυρικά φαινόμενα . . . . .	1
1.2 Αντικείμενο της εργασίας και κίνητρα . . . . .	4
1.3 Σχετικές έρευνες . . . . .	4
1.4 Διάρθρωση της εργασίας . . . . .	7
<b>2 Τεχνητή Νοημοσύνη (Artificial Intelligence - AI)</b>	<b>9</b>
2.1 Εισαγωγή . . . . .	9
2.2 Μηχανική Μάθηση (Machine Learning - ML) . . . . .	11
2.2.1 Είδη Μηχανικής Μάθησης . . . . .	12
2.3 Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks - ANN) . . . . .	15
2.3.1 Από τους βιολογικούς στους τεχνητούς νευρώνες . . . . .	15
2.3.2 Συνάρτηση ενεργοποίησης (Activation function) . . . . .	18
2.3.3 Πολυεπίπεδα τεχνητά νευρωνικά δίκτυα (Multilayer Perceptron) . . . . .	20
2.3.4 Εκπαίδευση τεχνητών νευρωνικών δικτύων (Training) . . . . .	21
2.3.5 Αλγόριθμοι Βελτιστοποίησης . . . . .	22
2.4 Βαθιά Μάθηση (Deep Learning - DL) . . . . .	27
<b>3 Επαναληπτικά Νευρωνικά Δίκτυα</b>	<b>31</b>
3.1 Απλά Επαναληπτικά Νευρωνικά Δίκτυα - RNNs . . . . .	31
3.2 Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης - LSTMs . . . . .	34
3.3 Δίκτυα Μετασχηματιστή - Transformers . . . . .	36
3.3.1 Μηχανισμός Προσοχής - Attention mechanism . . . . .	36
3.3.2 Αρχιτεκτονική Transformer . . . . .	38
3.4 Μοντέλο Πληροφόρησης - Informer model . . . . .	41

3.4.1	Αρχιτεκτονική Informer	41
<b>4</b>	<b>Στοχαστική προσέγγιση για πρόβλεψη χρονοσειρών</b>	<b>47</b>
4.1	Στοχαστικές διεργασίες - Γενικά	47
4.2	Κλιμακόγραμμα	51
4.3	Διεργασίες Hurst-Kolmogorov	52
4.4	Εκτίμηση πρόβλεψης χρονοσειρών	53
<b>5</b>	<b>Εφαρμογή στον ποταμό River Test</b>	<b>57</b>
5.1	Περιοχή μελέτης	57
5.2	Περιγραφή δεδομένων	59
5.3	Αρχιτεκτονικές μοντέλων	61
5.4	Μετρικές αξιολόγησης	68
5.5	Αποτελέσματα και σύγκριση μοντέλων	69
<b>6</b>	<b>Διερεύνηση μοντέλου Informer σε μεγάλο πλήθος ποταμών</b>	<b>79</b>
6.1	Περιοχή μελέτης	79
6.2	Περιγραφή δεδομένων	79
6.3	Αρχιτεκτονική μοντέλου Informer	83
6.4	Αποτελέσματα	83
<b>7</b>	<b>Σύνοψη και συμπεράσματα</b>	<b>87</b>
7.1	Συμπεράσματα	87
7.2	Μελλοντικές επεκτάσεις	88
	<b>Βιβλιογραφία</b>	<b>91</b>

# Κατάλογος Σχημάτων

1.1	Ο αριθμός των παγκόσμιων αναφερόμενων φυσικών καταστροφών από πλημύρες ανά δεδομένο έτος. (Πηγή: Our World in Data [3]). . . . .	3
1.2	Αριθμός των ανά έτος αναφορών (από το Google Scholar) που προσελκύουν δημοσιεύσεις που σχετίζονται με Transformers (Πηγή: [46]). . . . .	7
2.1	Διάγραμμα Euler για την σχηματική απεικόνιση των υποσυνόλων της τεχνητής νοημοσύνης (Πηγή: [49]). . . . .	10
2.2	Διάγραμμα Euler με τις εφαρμογές και τους αλγορίθμους των υποσυνόλων της τεχνητής νοημοσύνης. . . . .	11
2.3	Παράδειγμα υποπροσαρμογής (αριστερά), υπερπροσαρμογής (κέντρο) και ιδανικής προσαρμογής (δεξιά) (Πηγή: [53]). . . . .	13
2.4	Βιολογικός νευρώνας/ Από User:Dhp1080, translated and modified by Badseed. - Originally Neuron.jpg taken from the US Federal (public domain) (Nerve Tissue, retrieved March 2007), redrawn by User:Dhp1080 in Illustrator. Source: "Anatomy and Physiology" by the US National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) Program., CC BY-SA 3.0, <a href="https://commons.wikimedia.org/w/index.php?curid=24139135">https://commons.wikimedia.org/w/index.php?curid=24139135</a> . . . . .	16
2.5	Πολλαπλά επίπεδα σε ένα βιολογικό νευρωνικό δίκτυο. . . . .	17
2.6	Μορφή τεχνητού νευρωνικού δικτύου Perceptron. . . . .	17
2.7	Συναρτήσεις ενεργοποίησης και οι παράγωγοι τους (Πηγή: Aurélien Géron, 2019 [57]). . . . .	19
2.8	Η συνάρτηση ενεργοποίησης GELU και η παράγωγος της. (Πηγή: [58]) . . . . .	20
2.9	Πολυεπίπεδο ΤΝΔ (Multi-layer Perceptron). . . . .	21
2.10	Σχηματική απεικόνιση αλγορίθμου κατάβασης κλίσης (Gradient descent). . . . .	23
2.11	Παράδειγμα επιλογής μεγάλου (αριστερά) και μικρού (δεξιά) ρυθμού μάθησης. . . . .	24
2.12	Η απόδοση της βαθιάς μάθησης και των κλασικών αλγορίθμων μάθησης σε σχέση με τα διαθέσιμα δεδομένα (Πηγή: [65]). . . . .	28
2.13	Παράδειγμα δικτύου βαθιάς μάθησης (Πηγή: [66]). . . . .	29
3.1	Αρχιτεκτονική απλού επαναληπτικού νευρωνικού δικτύου RNN (Πηγή: Wikimedia [67, 68].) . . . . .	32

3.2	Αρχιτεκτονική ενός κελιού του νευρωνικού δικτύου LSTM (Πηγή: [73]). . . . .	35
3.3	Αναπαράσταση του νευρωνικού δικτύου LSTM, για τις χρονικές στιγμές $t - 1, t, t + 1$ , με τις αντίστοιχες πύλες (Πηγή: [74]). . . . .	36
3.4	Αρχιτεκτονική του μηχανισμού προσοχής (Πηγή: ‘Attention Is All You Need’ από τους Vaswani et al. [22]). . . . .	37
3.5	Επισκόπηση της αρχιτεκτονικής του μοντέλου μετασχηματιστή - Transformer (Πηγή: ‘Attention Is All You Need’ από τους Vaswani et al. [22]). . . . .	38
3.6	Αρχιτεκτονική μοντέλου Informer (Πηγή: ‘Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting’ από τους Zhou et al. [40]). . . . .	42
3.7	Απεικόνιση του μηχανισμού προσοχής ProbSparse (Πηγή: ‘Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting’ από τους Zhou et al. [40]). . . . .	44
3.8	Αρχιτεκτονική κωδικοποιητή (Πηγή: ‘Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting’ από τους Zhou et al. [40]). . . . .	44
4.1	Παράδειγμα στοχαστικής διεργασίας Wiener (Πηγή: [81]). . . . .	51
4.2	Ένα παράδειγμα μιας διαδικασίας συνεχούς χρόνου με δειγματοληψία σε χρονικά διαστήματα $D$ για μια συνολική περίοδο $T$ και με χρόνο απόκρισης του οργάνου $\Delta$ (Πηγή: [80]). . . . .	51
4.3	Γραφική απεικόνιση της εξίσωσης (4.22) (Πηγή: [79]). . . . .	55
5.1	Χάρτης με την τοποθεσία του ποταμού River Test (Πηγή: Google Earth). . . . .	57
5.2	Χάρτης του ποταμού River Test (με κόκκινο χρώμα), εντός της Νοτιοανατολικής Αγγλίας (Πηγή: [84]). . . . .	58
5.3	Μορφή του σετ των δεδομένων του ποταμού River Test. . . . .	59
5.4	Χάρτης με τις τοποθεσίες των σταθμών των μετρήσεων (Πηγή: [85]). . . . .	60
5.5	Η περιοχή μελέτης και η θέση του σταθμού παροχής Broadlands. . . . .	60
5.6	Χρονοσειρά παροχής σταθμού Broadlands. . . . .	61
5.7	Κλιμακόγραμμα της παροχής στο σταθμό Broadlands. . . . .	62
5.8	Αποτελέσματα δοκιμών για διαφορετικές παραμέτρους μοντέλου ARIMA. . . . .	64
5.9	Απόδοση του μοντέλου ARIMA στα δεδομένα εκπαίδευσης. . . . .	64
5.10	Αποτελέσματα πειραματικής διαδικασίας για συνδυασμούς ακολουθιών εισόδου και ορίζοντα πρόβλεψης μοντέλου Informer. . . . .	67
5.11	Αποτελέσματα μετρικών αξιολόγησης σε μορφή ραβδογράμματος. . . . .	71
5.12	Σύγκριση προβλεπόμενων τιμών μοντέλων και πραγματικών τιμών για ορίζοντα πρόβλεψης 2 ημερών. . . . .	72
5.13	Σύγκριση προβλεπόμενων τιμών μοντέλων και πραγματικών τιμών για ορίζοντα πρόβλεψης 10 ημερών. . . . .	72
5.14	Σύγκριση προβλεπόμενων τιμών μοντέλων και πραγματικών τιμών για ορίζοντα πρόβλεψης 20 ημερών. . . . .	73

5.15	Σύγκριση προβλεπόμενων τιμών μοντέλων και πραγματικών τιμών για ορίζοντα πρόβλεψης 40 ημερών. . . . .	73
5.16	Σύγκριση προβλεπόμενων τιμών μοντέλων και πραγματικών τιμών για ορίζοντα πρόβλεψης 60 ημερών. . . . .	74
5.17	Σύγκριση προβλεπόμενων τιμών μοντέλων και πραγματικών τιμών για ορίζοντα πρόβλεψης 80 ημερών. . . . .	74
5.18	Σύγκριση προβλεπόμενων τιμών μοντέλων και πραγματικών τιμών για ορίζοντα πρόβλεψης 100 ημερών. . . . .	75
5.19	Σύγκριση προβλεπόμενων τιμών μοντέλων και πραγματικών τιμών για ορίζοντα πρόβλεψης 168 ημερών. . . . .	75
5.20	Διάγραμμα σφάλματος (MSE) για κάθε μοντέλο συναρτήσει του χρονικού ορίζοντα πρόβλεψης. . . . .	76
5.21	Διάγραμμα σφάλματος (MSE) μοντέλων υπολογισμένα σε κάθε χρονικό βήμα για χρονικό ορίζοντα πρόβλεψης 100 ημερών. . . . .	76
5.22	Διάγραμμα διασποράς προβλεπόμενων τιμών μοντέλου Informer. . . . .	77
6.1	Χάρτης με της τοποθεσίες των διαθέσιμων σταθμών στο Ηνωμένο Βασίλειο της βάσης NRFA (Πηγή: [90]). . . . .	80
6.2	Γεωγραφική κατανομή των 112 σταθμών που χρησιμοποιήθηκαν στην εφαρμογή διερεύνησης του μοντέλου Informer. . . . .	82
6.3	Διάγραμμα διασποράς προβλεπόμενων τιμών μοντέλου Informer στο σύνολο των 112 σταθμών. . . . .	84
6.4	Θηκόγραμμα (box plot) της μετρικής σφάλματος MAPE μεταξύ εκτιμώμενων και πραγματικών παροχών. . . . .	85
6.5	Η απόδοση του μοντέλου Informer σε σύγκριση με την ασυμμετρία της χρονοσειράς. . . . .	85
6.6	Παράδειγμα χρονοσειράς με μικρό δείκτη ασυμμετρίας ( $skew = 1.2$ ). . . . .	86
6.7	Παράδειγμα χρονοσειράς με μεγάλο δείκτη ασυμμετρίας ( $skew = 6.6$ ). . . . .	86





# Κατάλογος Πινάκων

2.1	Συναρτήσεις ενεργοποίησης (Πηγή: Sebastian Raschka, 2016). . . . .	19
5.1	Περιγραφή χαρακτηριστικών δεδομένων. . . . .	59
5.2	Αναλογία των όρων (σε μέρες) του παρελθόντος $\nu$ και του μέλλοντος $k$ . . . . .	62
5.3	Υπερπαράμετροι μοντέλου LSTM. . . . .	65
5.4	Υπερπαράμετροι μοντέλου Informer. . . . .	66
5.5	Συνδυασμός ακολουθιών εισόδου (sequence length) και πρόβλεψης (prediction length) μοντέλου Informer. . . . .	66
5.6	Αποτελέσματα μοντέλων πρόβλεψης χρονοσειράς παροχών River Test. . . . .	70
6.1	Μεταδεδομένα σταθμών που χρησιμοποιήθηκαν για την εφαρμογή διερεύνησης του μοντέλου Informer (ενδεικτικά οι πρώτοι 20). . . . .	81
6.2	Υπερπαράμετροι μοντέλου Informer. . . . .	83

# Κεφάλαιο 1

## Εισαγωγή

### 1.1 Πλημμυρικά φαινόμενα

Τα *πλημμυρικά φαινόμενα* χρήζουν τεράστιας σημασίας παγκοσμίως ως φυσικές καταστροφές που πλήττουν διάφορες περιοχές, αφήνοντας βαθιές επιπτώσεις στις κοινωνίες. Ξεπερνούν τα γεωγραφικά όρια και επηρεάζουν εκατομμύρια ζώες σε όλο τον κόσμο, καθιστώντας τα κρίσιμο τομέα μελέτης και έρευνας. Ως υδρολογία πλημμυρών (flood hydrology) νοείται ο κλάδος της τεχνικής (engineering) υδρολογίας που εστιάζει στις διεργασίες που λαμβάνουν χώρα από την έναρξη μιας ισχυρής καταιγίδας μέχρι το πέρας της απορροής που οφείλεται στο συγκεκριμένο επεισόδιο βροχής. Η κατανόηση των επιπτώσεων και ο μετριασμός των κινδύνων που συνδέονται με τα πλημμυρικά φαινόμενα είναι υψίστης σημασίας για τη διαφύλαξη της ανθρωπίνης ζωής και την ελαχιστοποίηση των κοινωνικοοικονομικών και περιβαλλοντικών ζημιών. Στο πλαίσιο αυτό η Ευρωπαϊκή Ένωση μέσω του Ευρωπαϊκού Κοινοβουλίου εξέδωσε στις 23 Οκτωβρίου του 2007 την Οδηγία 2007/60/EK [1] για την αξιολόγηση και τη διαχείριση των κινδύνων πλημμύρας (assessment and management of flood risk). Σύμφωνα με το άρθρο 1 [1] των γενικών διατάξεων, σκοπός της οδηγίας αυτής είναι η θέσπιση ενός πλαισίου για την αξιολόγηση και τη διαχείριση των κινδύνων πλημμύρας, με στόχο τη μείωση των αρνητικών συνεπειών στην ανθρώπινη υγεία, το περιβάλλον, την πολιτιστική κληρονομιά και τις οικονομικές δραστηριότητες που συνδέονται με τις πλημμύρες στην Κοινότητα.

Τα πλημμυρικά φαινόμενα έχουν εκτεταμένες συνέπειες που υπερβαίνουν τις άμεσες φυσικές ζημιές. Η απώλεια ανθρώπινων ζώων και ο εκτοπισμός κοινοτήτων είναι από τα πιο τραγικά αποτελέσματα, αφήνοντας μόνιμες συναισθηματικές και κοινωνικές επιπτώσεις στις πληγείσες περιοχές. Επιπλέον, η καταστροφή των υποδομών, συμπεριλαμβανομένων σπιτιών, δρόμων, γεφυρών και υπηρεσιών κοινής ωφέλειας, διαταράσσει την καθημερινή ζωή και επιβάλλει σημαντικές οικονομικές επιβαρύνσεις. Οι γεωργικές εκτάσεις καταστρέφονται, οδηγώντας σε ελλείψεις τροφίμων και οικονομική αστάθεια. Ο συνολικός κοινωνικοοικονομικός αντίκτυπος

---

των πλημμυρικών φαινομένων απαιτεί προληπτικά μέτρα για τη βελτίωση της ακρίβειας των προβλέψεων και την ενίσχυση της ετοιμότητας για καταστροφές. Η Οδηγία 2007/60/EK διακρίνει τρία στάδια για την διαδικασία διαχείρισης του κινδύνου πλημμύρας. Αρχικά εφαρμόζεται η προκαταρκτική αξιολόγηση των κινδύνων πλημμύρας, στην οποία τα κράτη μέλη υποχρεούνται να προσδιορίσουν τις περιοχές εκείνες για τις οποίες συμπεραίνουν ότι υπάρχουν δυνητικοί σοβαροί κίνδυνοι πλημμύρας. Στη συνέχεια, εκπονούνται οι χάρτες επικινδυνότητας πλημμύρας και οι χάρτες κινδύνων πλημμύρας σε επίπεδο λεκάνης απορροής ποταμού. Οι χάρτες επικινδυνότητας πλημμύρας καλύπτουν τις γεωγραφικές περιοχές που θα μπορούσαν να πλημμυρίσουν σύμφωνα με τα ακόλουθα σενάρια: (α) πλημμύρες χαμηλής πιθανότητας ή σενάρια ακραίων φαινομένων, (β) πλημμύρες μέσης πιθανότητας (με περίοδο επαναληπτικότητας  $\geq 100$  χρόνια) και (γ) πλημμύρες υψηλής πιθανότητας ανάλογα με την περίπτωση.

Για κάθε ένα από τα παραπάνω σενάρια παρατίθενται και η έκταση της πλημμύρας, το βάθος νερού και η ταχύτητα ροής των υδάτων. Οι χάρτες κινδύνου πλημμύρας περιγράφουν τις δυνητικές αρνητικές συνέπειες που συνδέονται με τις πλημμύρες και εκφράζονται μέσω του αριθμού του πληθυσμού που ενδέχεται να πληγεί, της οικονομικής δραστηριότητας στην περιοχή που ενδέχεται να πληγεί, των εγκαταστάσεων που ενδέχεται να προκαλέσουν τυχαία ρύπανση σε περίπτωση πλημμύρας και όποια άλλη πληροφορία το κράτος μέλος θεωρεί χρήσιμη. Στο τρίτο και τελευταίο στάδιο, τα κράτη μέλη καταρτίζουν σχέδια διαχείρισης των κινδύνων πλημμύρας σε επίπεδο λεκάνης απορροής. Τα σχέδια αυτά περιλαμβάνουν μέτρα που εστιάζουν στη μείωση των δυσμενών συνεπειών που οι πλημμύρες έχουν για την ανθρώπινη υγεία, το περιβάλλον, την πολιτιστική κληρονομιά και την οικονομική δραστηριότητα, όπως επίσης και στη μείωση της πιθανότητας πλημμύρας.

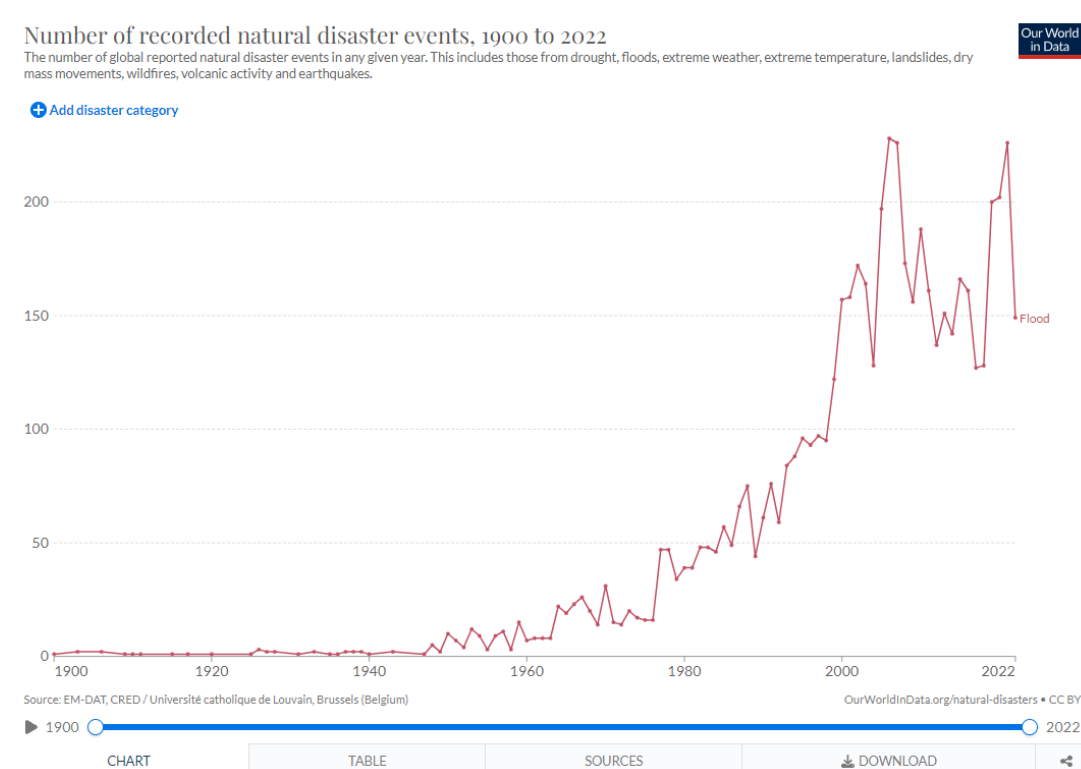
Σύμφωνα με το άρθρο 2 των γενικών διατάξεων της οδηγίας 2007/60/EK [1] ως *πλημμύρα* ορίζεται «η προσωρινή κάλυψη από νερό εδάφους το οποίο, υπό φυσιολογικές συνθήκες, δεν καλύπτεται από νερό». Αυτό περιλαμβάνει πλημμύρες από ποτάμια, ορεινούς χείμαρρους, εφήμερα ρεύματα και πλημμύρες από τη θάλασσα σε παράκτιες περιοχές. Οι ροές που παρατηρούνται στη φύση είναι εν γένει μη μόνιμες. Ένα σημαντικό πρόβλημα μη μόνιμης ροής που καλείται να αντιμετωπίσει ένας μηχανικός είναι η κίνηση ενός πλημμυρικού κύματος κατά μήκος ενός ανοικτού αγωγού, συνήθως φυσικού ποταμού. Τα πλημμυρικά κύματα αποτελούν φαινόμενα μη μόνιμης ροής. Ένα σύνηθες πρόβλημα στις πλημμύρες είναι αυτό της διόδευσης μιας πλημμύρας, δηλαδή το πρόβλημα της μαθηματικής αναπαράστασης της εξέλιξης ενός πλημμυρικού φαινομένου στο χώρο και στο χρόνο [2]. Τα υδρολογικά και υδραυλικά μεγέθη που ενδιαφέρουν σε τέτοια φαινόμενα είναι:

- η παροχή
- το βάθος ροής
- η βροχόπτωση

- η απορροή
- η θερμοκρασία

Ένα από τα πιο σημαντικά μεγέθη που ενδιαφέρουν στις πλημμύρες είναι η παροχή. Η παροχή αναφέρεται στον όγκο του νερού που ρέει μέσα από έναν ποταμό κατά τη διάρκεια μιας συγκεκριμένης χρονικής περιόδου. Αντιπροσωπεύει την ποσότητα του νερού που κινείται μέσα σε ένα κανάλι και συνήθως έχει μονάδες κυβικών μέτρων ανά δευτερόλεπτο ( $m^3/s$ ).

Στο Σχήμα 1.1 φαίνεται η εξέλιξη των αναφερόμενων πλημμυρικών φαινομένων από το 1900 μέχρι το 2022.



**Σχήμα 1.1:** Ο αριθμός των παγκόσμιων αναφερόμενων φυσικών καταστροφών από πλημμύρες ανά δεδομένο έτος. (Πηγή: Our World in Data [3]).

Η αποτελεσματική πρόβλεψη πλημμυρών διαδραματίζει κρίσιμο ρόλο στην ελαχιστοποίηση των κινδύνων και τη διαχείριση των πλημμυρικών φαινομένων. Οι έγκαιρες και ακριβείς προβλέψεις επιτρέπουν την εφαρμογή προληπτικών μέτρων, όπως συστήματα έγκαιρης προειδοποίησης, σχέδια εκκένωσης και κατανομή πόρων, τα οποία σώζουν ζωές και μειώνουν τις ζημιές. Παρέχοντας στους υπεύθυνους λήψης αποφάσεων αξιόπιστες πληροφορίες, η πρόγνωση πλημμυρών υποστηρίζει τις προσπάθειες διαχείρισης καταστροφών, διευκολύνει το σχεδιασμό αντιμετώπισης έκτακτης ανάγκης και ενισχύει την κατανομή των πόρων για την ανάκαμψη και την ανοικοδόμηση. Επομένως η πρόβλεψη χρονοσειρών ενός μεγέθους που σχετίζεται με τα πλημμυρικά φαινόμενα μπορεί να ωφελήσει σημαντικά πολλές πτυχές της ανθρωπότητας.

---

## 1.2 Αντικείμενο της εργασίας και κίνητρα

Η παρούσα εργασία ασχολείται με την χρήση αλγορίθμων βαθιάς μάθησης για την επίλυση του προβλήματος της πρόβλεψης χρονοσειρών πλημμυρικών φαινομένων. Ως *πρόβλεψη χρονοσειρών* (*time series forecasting*) νοείται η διαδικασία στην οποία γίνεται χρήση ενός μοντέλου ή μιας μεθόδου για την πραγματοποίηση εκτιμήσεων μελλοντικών τιμών με βάση τα ιστορικά δεδομένα που συλλέγονται με την πάροδο του χρόνου. Όπως έγινε σαφές η ακριβής εκτίμηση των πλημμυρικών γεγονότων αποτελεί μείζον θέμα και η βελτίωση της ακρίβειας της πρόβλεψης παρέχει πολύτιμες πληροφορίες για στρατηγικές ετοιμότητας και μετριασμού των καταστροφών. Η εργασία αναλύει κυρίως χρονοσειρές παροχής, ενώ τα μοντέλα αναπτύχθηκαν και δοκιμάστηκαν κάνοντας χρήση πραγματικών συνόλων δεδομένων πλημμυρών από διάφορες περιοχές στο Ηνωμένο Βασίλειο, με κύρια εφαρμογή τον ποταμό River Test στην Αγγλία, εξασφαλίζοντας τη δυνατότητα εφαρμογής και γενίκευσής τους.

Παραδοσιακές τεχνικές πρόβλεψης πλημμυρών, συμπεριλαμβανομένης της υδρολογικής μοντελοποίησης και της στατιστικής ανάλυσης, χρησιμοποιούνται εδώ και δεκαετίες. Ωστόσο, οι μέθοδοι αυτές βασίζονται συχνά σε ιστορικά δεδομένα και παραδοχές που ενδέχεται να μην αποτυπώνουν πλήρως την πολύπλοκη και δυναμική φύση των πλημμυρικών φαινομένων. Για το λόγο αυτό η παρούσα εργασία προσπαθεί να ξεπεράσει τους περιορισμούς των παραδοσιακών προσεγγίσεων, αφού εξετάζονται και αυτές, αναπτύσσοντας μεθόδους βαθιάς μάθησης. Πιο συγκεκριμένα δίνεται ιδιαίτερη έμφαση σε σύγχρονα μοντέλα βαθιάς μάθησης, γνωστά ως Transformer-based μοντέλα, τα οποία τυγχάνουν μεγάλης επιστημονικής προσοχής και κερδίζουν όλο και περισσότερο δημοτικότητα με υποσχόμενα αποτελέσματα στην πρόβλεψη χρονοσειρών. Τα μοντέλα αυτά προσφέρουν αποτελεσματικότερη καταγραφή εξαρτήσεων μεγάλης εμβέλειας (long-range dependencies) σε δεδομένα χρονοσειρών με την αξιοποίηση μηχανισμών προσοχής (attention mechanism) που επιτρέπουν γενική εποπτεία και αλληλεξάρτηση σε όλο το μήκος της χρονοσειράς προς ανάλυση. Η εργασία υλοποιεί ένα παίβιο στοχαστικό μοντέλο (SB), ένα στατιστικό μοντέλο (ARIMA), ένα επαναληπτικό δίκτυο βαθιάς μάθησης (LSTM) και ένα transformer-based μοντέλο βαθιάς μάθησης (Informer).

## 1.3 Σχετικές έρευνες

Η πρόβλεψη χρονοσειρών έχει προσελκύσει μεγάλο επιστημονικό ενδιαφέρον στους ερευνητές πολλών ειδικοτήτων και για την ανάλυση διαφόρων εργασιών. Έχουν χρησιμοποιηθεί κατά καιρούς πάρα πολλοί αλγόριθμοι για την επίλυση του προβλήματος της πρόβλεψης, από κλασικές μεθόδους στατιστικής μέχρι προχωρημένους αλγορίθμους βαθιάς μάθησης.

Έχουν χρησιμοποιηθεί διάφορες καθιερωμένες τεχνικές για την αποτύπωση διαφορετικών χαρακτηριστικών χρονοσειρών. Στο βιβλίο των Box et al. [4] παρέχονται αρκετές στατιστικές μέθοδοι και τεχνικές που χρησιμοποιούνται για την ανάλυση και την πρόβλεψη δεδομένων χρο-

---

νοσειρών, συμπεριλαμβανομένου και του Αυτοπαλίνδρομου Ολοκληρωμένου Κινητού Μέσου μοντέλου ARIMA (Autoregressive Integrated Moving Average) [5]. Το μοντέλο αυτό ακολουθεί τη διαδικασία Markov (Markov process) και δημιουργεί ένα αυτοπαλινδρομικό μοντέλο για αναδρομική διαδοχική πρόβλεψη. Στο [6] παρουσιάζεται μια ολοκληρωμένη επισκόπηση των μεθόδων εκθετικής εξομάλυνσης (exponential smoothing) για την πρόβλεψη χρονοσειρών. Εξετάζουν την πρακτική εφαρμογή διαφόρων τεχνικών εκθετικής εξομάλυνσης καθώς και τις επεκτάσεις τους για τον χειρισμό της τάσης, της εποχικότητας και άλλων συνιστωσών των δεδομένων χρονοσειρών. Μερικοί περιορισμοί των παραδοσιακών αυτών τεχνικών είναι πως συχνά υποθέτουν ότι τα δεδομένα των χρονοσειρών είναι γραμμικά και στάσιμα. Μοντέλα όπως το ARIMA και η εκθετική εξομάλυνση έχουν περιορισμένη δυνατότητα στην προσέγγιση πολύπλοκων μοτίβων των δεδομένων χρονοσειρών, όπως οι ανωμαλίες, οι απότομες αλλαγές κλίσης και οι μη γραμμικές σχέσεις. Ακόμη μία πρόκληση των παραδοσιακών μοντέλων είναι ο περιορισμός που ενδεχομένως να έχουν όταν επεξεργάζονται δεδομένα χρονοσειρών μεγάλης κλίμακας ή υψηλών διαστάσεων. Οι υπολογιστικές απαιτήσεις και η πολυπλοκότητα του μοντέλου μπορεί να αυξηθούν σημαντικά, καθιστώντας έτσι τις μεθόδους αυτές μη εφαρμόσιμες σε τέτοιες περιπτώσεις.

Οι συγγραφείς στο [7] προσφέρουν μια ολοκληρωμένη ανάλυση διαφορετικών μοντέλων μηχανικής μάθησης για την πρόβλεψη χρονοσειρών. Μεταξύ άλλων παρουσιάζεται η μελέτη σύγκρισης διαφόρων μοντέλων, συμπεριλαμβανομένων των τεχνητών νευρωνικών δικτύων (artificial neural networks), της παλινδρόμησης διανυσμάτων υποστήριξης (support vector regression) και των τυχαίων δασών (random forests), για διάφορα σετ δεδομένων. Η μελέτη παρέχει πληροφορίες σχετικά με τα πλεονεκτήματα και τους περιορισμούς κάθε μοντέλου. Στο [8] οι συγγραφείς εφαρμόζουν μεθόδους μηχανικής μάθησης για πρόβλεψη δεδομένων χρηματοοικονομικών χρονοσειρών. Το κύριο συμπέρασμα τους είναι η ύπαρξη έλλειψης βιβλιογραφίας που να εξετάζει αν και πως οι τεχνικές μηχανικής μάθησης μπορούν να βελτιώσουν τη σχέση κινδύνου-απόδοσης των επενδυτών υπό πραγματικές συνθήκες και περιορισμούς. Στο [9] εισάγεται ο αλγόριθμος XGBoost, ο οποίος αποτελεί μια βελτιστοποίηση της μεθόδου ενίσχυσης κλίσης (gradient boosting) για μοντέλα δέντρων απόφασης (decision tree models), ενώ στο [10] χρησιμοποιείται μία παραλλαγή για πιθανοτική πρόβλεψη.

Μεγάλο είναι και το ενδιαφέρον για τις μεθόδους βαθιάς μάθησης στην πρόβλεψη χρονοσειρών. Στο [11] αναλύεται το μοντέλο LSTM για διάφορες εργασίες μεταξύ αυτών και η πρόβλεψη χρονοσειρών, ενώ δίνεται έμφαση σε διάφορες αρχιτεκτονικές του μοντέλου για τον εντοπισμό των δυνατοτήτων και των περιορισμών του. Στην οικογένεια αυτή ανήκει και τα GRU [12] που χρησιμοποιούν δομή πύλης (gated structure) για τον έλεγχο της ροής πληροφοριών για την αντιμετώπιση του προβλήματος εξαφάνισης της κλίσης (vanishing gradient). Το [13] παρέχει μια ολοκληρωμένη επισκόπηση των τεχνικών βαθιάς μάθησης για την εξόρυξη δεδομένων και για την πρόβλεψη χρονοσειρών, χρησιμοποιώντας μοντέλα όπως RNN, LSTM και CNN. Στα [14, 15, 16] γίνεται η πρόβλεψη για δεδομένα πλημμυρικών φαινομένων όπως η παροχή και η

---

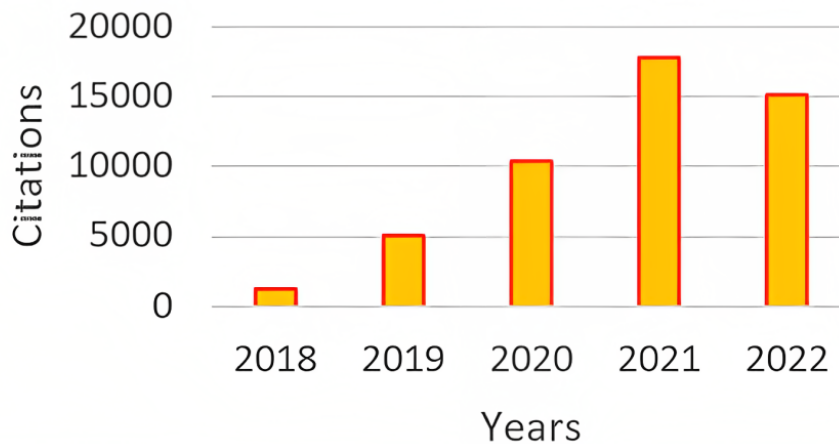
βροχή. Πιο συγκεκριμένα στο [14] οι συγγραφείς εξετάζουν ένα δίκτυο LSTM σε συνδυασμό με ένα επίπεδο συνελίξεων (convolutions) για να αντιμετωπίσουν τόσο τις χωρικές όσο και τις χρονικές εξαρτήσεις στα δεδομένα βροχόπτωσης. Όμοια και στο [15] εξετάζεται ο συνδυασμός των CNN και LSTM για την ωριαία πρόβλεψη παροχής σε ποταμούς. Η μελέτη στο [16] επικεντρώνεται στην ανάπτυξη πιθανοτικών μοντέλων πρόβλεψης για την καταγραφή της αβεβαιότητας που σχετίζεται με τις προβλέψεις της ροής των ποταμών. Οι συγγραφείς προτείνουν ένα Bayesian δίκτυο LSTM που μπορεί να υπολογίζει την εγγενή αβεβαιότητα των προβλεπόμενων τιμών. Τα αποτελέσματα καταδεικνύουν την αποτελεσματικότητα της προσέγγισης αυτής στην παροχή ακριβών προβλέψεων για πολλαπλά χρονικά βήματα μπροστά. Ερευνητές από την Amazon [17] δημιούργησαν το DeepAR, μία μεθοδολογία που βασίζεται σε επαναληπτικά δίκτυα όπως τα LSTMs παράγοντας πιθανοτική πρόβλεψη, ενώ εκπαιδεύεται σε πολλές χρονοσειρές (σχετικές με το πρόβλημα στο οποίο εφαρμόζεται). Στο [18] προτείνεται μία αρχιτεκτονική βαθιού νευρωνικού δικτύου βασισμένη σε προς τα πίσω και προς τα εμπρός σύνδεση του σφάλματος και σε μια πολύ βαθιά στοίβα πλήρως συνδεδεμένων επιπέδων. Στα [19, 20] οι συγγραφείς δοκιμάζουν τις επιδόσεις του μοντέλου Prophet, το οποίο συστήθηκε από ερευνητές του Facebook [21] και προσπαθεί να προσαρμόσει την πιο κατάλληλη καμπύλη, για την πρόβλεψη ημερήσιων και μηνιαίων παροχών ποταμών τόσο σε μικρό όσο και σε μεγάλο χρονικό ορίζοντα.

Το 2017 αποτελεί ένα ορόσημο για το πεδίο της βαθιάς μάθησης, καθώς παρουσιάστηκε το μοντέλο μετασχηματιστή ή Transformer [22], όπου εισήχθη η έννοια του μηχανισμού προσοχής και επιτρέπεται στο μοντέλο η εστίαση σε διαφορετικά κομμάτια της ακολουθίας εισόδου για την παραγωγή της εξόδου. Η εφαρμογή του μοντέλου ήταν καταρχάς το πεδίο της επεξεργασίας φυσικής γλώσσας (NLP) και ειδικότερα της μετάφρασης. Τέτοιου είδους μοντέλα δεν άργησαν να γίνουν ολοένα και πιο δημοφιλή σε πολλούς τομείς και εργασίες, με την πρόβλεψη χρονοσειρών να μην αποτελεί εξαίρεση σε αυτές. Στο Σχήμα 1.2 απεικονίζεται αυτή η εξέλιξη σε ένα διάγραμμα με τον αριθμό των αναφορών (από το Google Scholar) ανά έτος των δημοσιεύσεων που σχετίζονται με τα δίκτυα των μετασχηματιστών. Τα κύρια χαρακτηριστικά που αναδεικνύουν τα μοντέλα Transformer έναντι των λοιπών μοντέλων βαθιάς μάθησης (LSTM, CNN κτλ.) είναι η δυνατότητα ανίχνευσης εξαρτήσεων μεγάλης εμβέλειας, η δυνατότητα παραλληλοποίησης της εργασίας και η διαχείριση δεδομένων μεγάλης κλίμακας. Στο [23] οι συγγραφείς κάνουν μια εκτενή βιβλιογραφική επισκόπηση των εφαρμογών των Transformers σε εργασίες που έχουν να κάνουν με χρονοσειρές, τονίζοντας τα πλεονεκτήματα, τους περιορισμούς και πιθανές μελλοντικές ερευνητικές κατευθύνσεις. Από τις πρώτες εφαρμογές των Transformers σε προβλήματα πρόβλεψης χρονοσειρών [24, 25, 26], έγινε φανερή η υπεροχή τους σε σύγκριση με τα μέχρι στιγμής πιο διαδεδομένα μοντέλα. Η μεγάλη αυτή άνθηση στη χρήση των μοντέλων τέτοιας φύσης, έκανε τους ερευνητές γρήγορα να αναπτύξουν μοντέλα βασισμένα στην ιδέα των Transformers. Το αποτέλεσμα είναι να υπάρχουν στη βιβλιογραφία αρκετά τέτοια μοντέλα με συγκεκριμένες σχεδιαστικές επιλογές από την κάθε ερευνητική ομάδα για την αντιμετώπιση διαφορετικών προκλήσεων και απαιτήσεων σε εργασίες πρόβλεψης

χρονοσειρών. Για την ακρίβεια μοντέλα όπως τα Linformer (NLP)[27], Reformer (NLP) [28], Longformer (NLP) [29], SSDNet [30], TFT [31], Autoformer [32], Triformer [33], TCCT [34], Pyraformer [35], TACTiS [36], FEDformer [37], Scaleformer [38], Crossformer [39] μοιράζονται όλα κοινά σημεία αναφοράς αλλά έχουν και μοναδικά χαρακτηριστικά και βελτιώσεις για την εργασία της πρόβλεψης χρονοσειρών σε πολλά σενάρια δεδομένων όπως δεδομένα ασθενών, ζήτησης ενέργειας, οικονομίας, πωλήσεων, κίνησης αυτοκινήτων, θερμοκρασίας κ.ά.

Ένα μοντέλο το οποίο δείχνει να ξεχωρίζει από τα προαναφερθέντα, καθώς συγκεντρώνει αρκετά πλεονεκτικά χαρακτηριστικά, και με το οποίο θα ασχοληθεί κατά κόρον η παρούσα εργασία, είναι το Informer [40]. Βασική επιδίωξή του είναι η επίλυση του προβλήματος της πρόβλεψης ακολουθιών μεγάλου μήκους σε δεδομένα χρονοσειρών. Η αξιολόγηση διενεργήθηκε σε δεδομένα όπως η θερμοκρασία ηλεκτρικού μετασχηματιστή (Electricity Transformer Temperature-ETT), η κατανάλωση ηλεκτρικής ενέργειας και δεδομένα σχετικά με το κλίμα. Έχει χρησιμοποιηθεί από ερευνητές σε αρκετές εφαρμογές όπως η πρόβλεψη της ξηρασίας [41], των δονήσεων του εδράνου του κινητήρα [42], της παραγωγής αιολικής ενέργειας [43], της διαθεσιμότητας σταθμών φόρτισης ηλεκτρικών οχημάτων [44] και της ταχύτητας κυκλοφορίας [45]. Μεταξύ άλλων το μοντέλο Informer χρησιμοποιεί μια μέθοδο βασισμένη στην απόκλιση Kullback–Leibler (KL-divergence) για την απλοποίηση του πίνακα προσοχής. Αυτός ο αραιότερος πίνακας μειώνει δραστικά την πολυπλοκότητα, κοστίζοντας στο σύστημα μόνο  $N \log N$  για τον υπολογισμό.

### (a) Citations of Transformer papers in recent years



**Σχήμα 1.2:** Αριθμός των ανά έτος αναφορών (από το Google Scholar) που προσελκύουν δημοσιεύσεις που σχετίζονται με Transformers (Πηγή: [46]).

## 1.4 Διάρθρωση της εργασίας

Πέραν του παρόντος (Κεφάλαιο 1), όπου γίνεται μία εισαγωγή στα πλημμυρικά φαινόμενα, τη σπουδαιότητα ανάλυσης αυτών αλλά και το γενικό πλαίσιο στο οποίο κινείται η μεταπτυχιακή



---

διπλωματική, το τεύχος αποτελείται από έξι επιπλέον κεφάλαια.

Στο Κεφάλαιο 2 περιγράφονται ο κλάδος της Τεχνητής Νοημοσύνης, οι τομείς της Μηχανικής Μάθησης και της Βαθιάς Μάθησης με ιδιαίτερη έμφαση στα τεχνητά νευρωνικά δίκτυα και τον τρόπο λειτουργίας τους.

Στο Κεφάλαιο 3 αναλύονται τα κύρια μοντέλα βαθιάς μάθησης που επιλέχθηκαν να εφαρμοστούν, συμπεριλαμβανομένων των επαναληπτικών νευρωνικών δικτύων και των μετασχηματιστών.

Στο Κεφάλαιο 4 αναπτύσσεται η μεθοδολογία της στοχαστικής προσέγγισης για την πρόβλεψη χρονοσειρών, η οποία επιλέχθηκε να αποτελεί ένα αφελές μοντέλο αναφοράς.

Στο Κεφάλαιο 5 παρουσιάζονται αναλυτικά τα αποτελέσματα και η αξιολόγηση των μοντέλων, μέσω μίας εφαρμογής για την εκτίμηση της παροχής του ποταμού River Test στην Αγγλία.

Στο Κεφάλαιο 6 αξιολογείται το μοντέλο μετασχηματιστή Informer, εφαρμόζοντάς το σε ένα μεγάλο πλήθος ποταμών.

Τέλος στο Κεφάλαιο 7 συνοψίζονται τα τελικά συμπεράσματα που προκύπτουν από την παρούσα εργασία και δίνονται ενδεικτικές προτάσεις για μελλοντική έρευνα.

## Κεφάλαιο 2

# Τεχνητή Νοημοσύνη (Artificial Intelligence - AI)

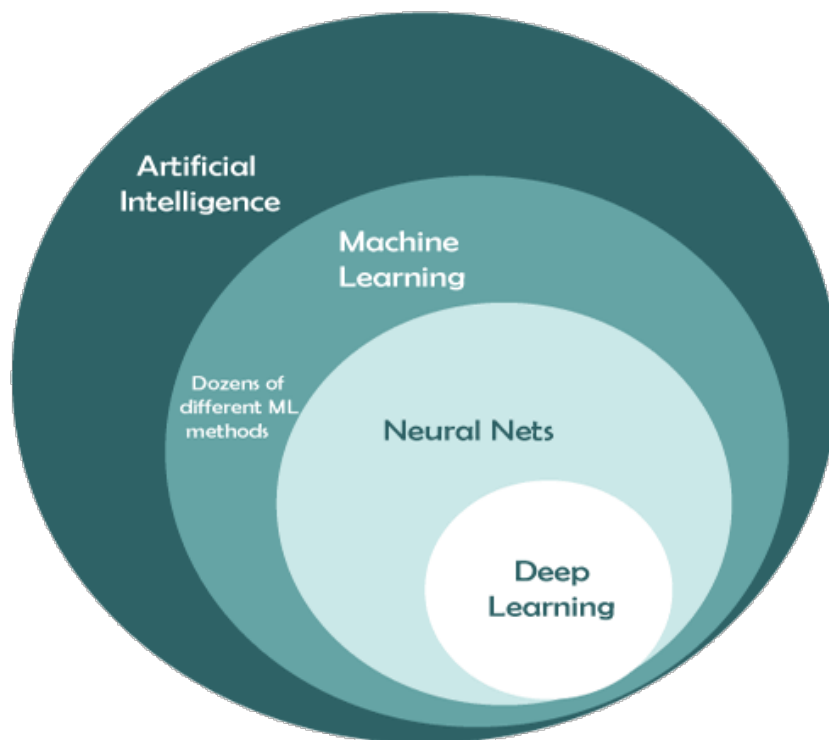
### 2.1 Εισαγωγή

Ο όρος *Τεχνητή Νοημοσύνη* (TN) αναφέρεται στον κλάδο της πληροφορικής ο οποίος ασχολείται με τη σχεδίαση και την υλοποίηση υπολογιστικών συστημάτων που μιμούνται στοιχεία της ανθρώπινης συμπεριφοράς τα οποία υπονοούν έστω και στοιχειώδη ευφυΐα [47]. Το κύριο πλεονέκτημα της τεχνητής νοημοσύνης σε σχέση με την ανθρώπινη συμπεριφορά είναι η ικανότητά της να αναλύει τεράστιες ποσότητες δεδομένων και εμπειριών πολύ πιο γρήγορα από ό,τι οι άνθρωποι.

Η τεχνητή νοημοσύνη συνδέει γνώσεις από πολλούς τομείς. Χρησιμοποιεί τεχνικές από τη θεωρία πιθανοτήτων, τα οικονομικά και το σχεδιασμό αλγορίθμων για την επίλυση πρακτικών προβλημάτων. Επιπλέον, ο τομέας της TN αντλεί στοιχεία από την επιστήμη των υπολογιστών, τα μαθηματικά, την ψυχολογία και τη γλωσσολογία. Η επιστήμη των υπολογιστών παρέχει εργαλεία για το σχεδιασμό και την κατασκευή αλγορίθμων, ενώ τα μαθηματικά προσφέρουν εργαλεία για τη μοντελοποίηση και την επίλυση των προβλημάτων βελτιστοποίησης που προκύπτουν. Παρότι η έννοια της TN υπάρχει από τον 19ο αιώνα, με τον μαθηματικό Άλαν Τούρινγκ να προτείνει το 1950 τη δοκιμή Τούρινγκ [48] - μία απλή δοκιμασία που θα μπορούσε να εξακριβώσει αν μία μηχανή διαθέτει ευφυΐα, η ικανότητα και αποτελεσματικότητά της δεν βοήθησε την ανάπτυξη της. Τις τελευταίες όμως δεκαετίες λόγω της ολοένα αυξανόμενης τεχνολογικής ανάπτυξης, διαθεσιμότητας υπολογιστικής ισχύος και του μεγάλου όγκου δεδομένων για την εκπαίδευση των συστημάτων τεχνητής νοημοσύνης ωφελήθηκε η ανάπτυξη και η αξιοπιστία τέτοιων συστημάτων.

Στην TN συμπεριλαμβάνονται και άλλοι τομείς, καθένας από τους οποίους εφαρμόζεται σε διαφορετικά προβλήματα και έχει ανεπτυγμένους δικούς του αλγορίθμους. Στο Σχήμα 2.1 α-

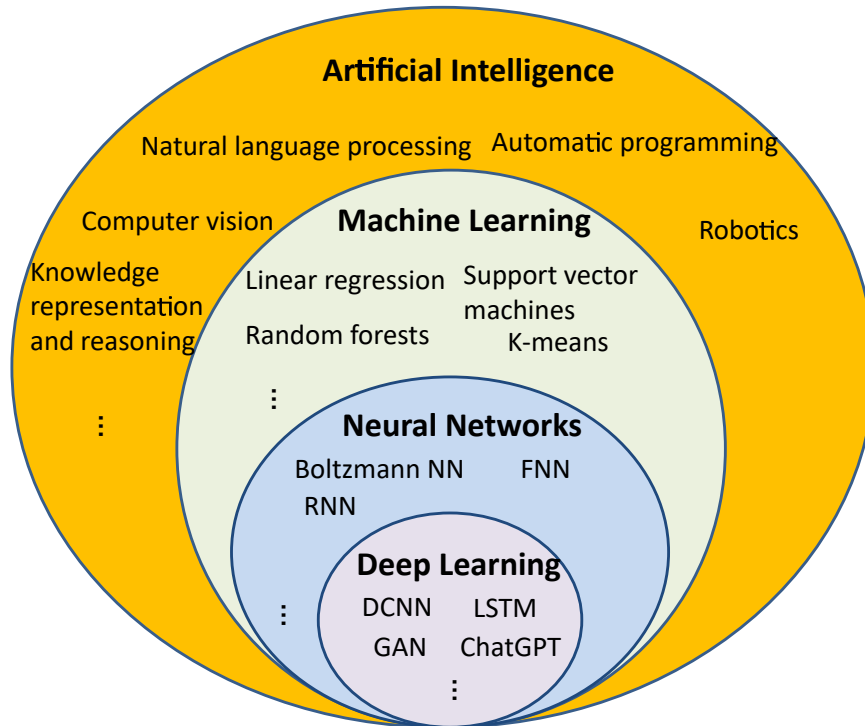
πεικονίζονται τα υποσύνολα της ΤΝ και πιο συγκεκριμένα η Μηχανική Μάθηση, τα Τεχνητά Νευρωνικά Δίκτυα και η Βαθιά μάθηση. Η Τεχνητή Νοημοσύνη είναι ο γενικός όρος για τη δημιουργία ευφών συστημάτων ικανών να εκτελούν εργασίες που σε άλλη περίπτωση θα απαιτούσαν ανθρώπινη νοημοσύνη. Από την άλλη η Μηχανική Μάθηση αποτελεί ένα υποπεδίο της ΤΝ το οποίο χρησιμοποιεί στατιστικές μεθόδους για να καταστήσει ικανό τον υπολογιστή να μάθει μέσα από τα δεδομένα χωρίς το ρητό προγραμματισμό του. Τα Τεχνητά Νευρωνικά Δίκτυα είναι υπολογιστικά μοντέλα εμπνευσμένα από τη λειτουργία του ανθρώπινου εγκεφάλου που χρησιμοποιούνται στην Μηχανική Μάθηση. Τέλος η Βαθιά Μάθηση είναι ένα υποσύνολο της Μηχανικής Μάθησης το οποίο χρησιμοποιεί τεχνητά νευρωνικά δίκτυα με πολλαπλά επίπεδα για την επεξεργασία σύνθετων μοτίβων και την εξαγωγή αναπαραστάσεων υψηλού επιπέδου από μεγάλους όγκους δεδομένων.



**Σχήμα 2.1:** Διάγραμμα Euler για την σχηματική απεικόνιση των υποσυνόλων της τεχνητής νοημοσύνης (Πηγή: [49]).

Στο Σχήμα 2.2 φαίνονται οι διάφοροι κλάδοι της τεχνητής νοημοσύνης μαζί με μερικές εφαρμογές για κάθε ένα υποσύνολο.

Στα επόμενα υποκεφάλαια θα ακολουθήσει η ανάλυση των υποπεδίων της ΤΝ έχοντας ως στόχο την καλύτερη κατανόηση των εννοιών και του τρόπου με τον οποίο οι αλγόριθμοι αυτοί λειτουργούν.



**Σχήμα 2.2:** Διάγραμμα Euler με τις εφαρμογές και τους αλγορίθμους των υποσυνόλων της τεχνητής νοημοσύνης.

## 2.2 Μηχανική Μάθηση (Machine Learning - ML)

Η *Μηχανική Μάθηση* (MM) ή *Machine Learning* (ML) αποτελεί αναμφισβήτητα μία από τις πιο επιδραστικές και ισχυρές τεχνολογίες στις μέρες μας. Μπορεί να χαρακτηριστεί ως η επιστήμη του προγραμματισμού των υπολογιστών ώστε να μπορούν να μαθαίνουν από τα δεδομένα. Ένας πιο γενικός ορισμός είναι πως η MM είναι το πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν χωρίς το ρητό προγραμματισμό τους (Arthur Samuel, 1959 [50]). Ο Tom Mitchell το 1997 [51] έδωσε έναν ορισμό προσανατολισμένο πιο πολύ στην επιστήμη της μηχανικής, με ένα πρόγραμμα υπολογιστή να θεωρείται ότι μαθαίνει από την εμπειρία  $E$  (Experience) σε σχέση με κάποια εργασία  $T$  (Task) και κάποιο μέτρο απόδοσης  $P$  (Performance), αν η απόδοσή του στην εργασία  $T$ , όπως μετράται από το μέτρο απόδοσης  $P$ , βελτιώνεται με την εμπειρία  $E$ .

Ο λόγος ύπαρξης της MM είναι διότι βρίσκει εφαρμογή και δίνει λύση σε πολλούς τομείς και σε σύνθετα προβλήματα. Πιο συγκεκριμένα είναι σημαντικό καθώς επιλύει: προβλήματα για τα οποία οι υπάρχουσες λύσεις απαιτούν επαναλαμβανόμενες κινήσεις ή μεγάλους καταλόγους κανόνων, πολύπλοκα προβλήματα για τα οποία δεν υπάρχουν καλές λύσεις από τις παραδοσιακές προσεγγίσεις, εξαγωγή συμπερασμάτων σχετικά με πολύπλοκα προβλήματα και διαχείριση μεγάλου όγκου δεδομένων.

---

## 2.2.1 Είδη Μηχανικής Μάθησης

Τα συστήματα μηχανικής μάθησης μπορούν να ταξινομηθούν ανάλογα με την ποσότητα και τον τύπο της εποπτείας που λαμβάνουν κατά την εκπαίδευση. Χωρίζονται σε τέσσερις μεγάλες κατηγορίες: επιβλεπόμενη μάθηση, μη επιβλεπόμενη μάθηση, ημιεπιβλεπόμενη μάθηση και ενισχυτική μάθηση. Η επιβλεπόμενη και η μη επιβλεπόμενη μάθηση είναι οι πιο συχνά χρησιμοποιούμενες προσεγγίσεις. Από την άλλη η ημιεπιβλεπόμενη και η ενισχυτική μάθηση είναι πιο πρόσφατες και πολυπλοκότερες προσεγγίσεις, αλλά με εντυπωσιακά αποτελέσματα σε διαφορετικά προβλήματα. Στη MM είναι πολύ διαδεδομένο το θεώρημα No Free Lunch (NFL) [52], το οποίο δηλώνει τους περιορισμούς της βελτιστοποίησης και της γενίκευσης. Σύμφωνα με το θεώρημα δεν υπάρχει ένας μόνο αλγόριθμος μάθησης που να αποδίδει το ίδιο καλά ή και βέλτιστα για όλα τα πιθανά προβλήματα που μπορεί να υπάρξουν. Κάθε εργασία έχει τις δικές της ιδιαιτερότητες, με πολλούς αλγορίθμους και προσεγγίσεις να ταιριάζουν στις ιδιαιτερότητες κάθε εργασίας.

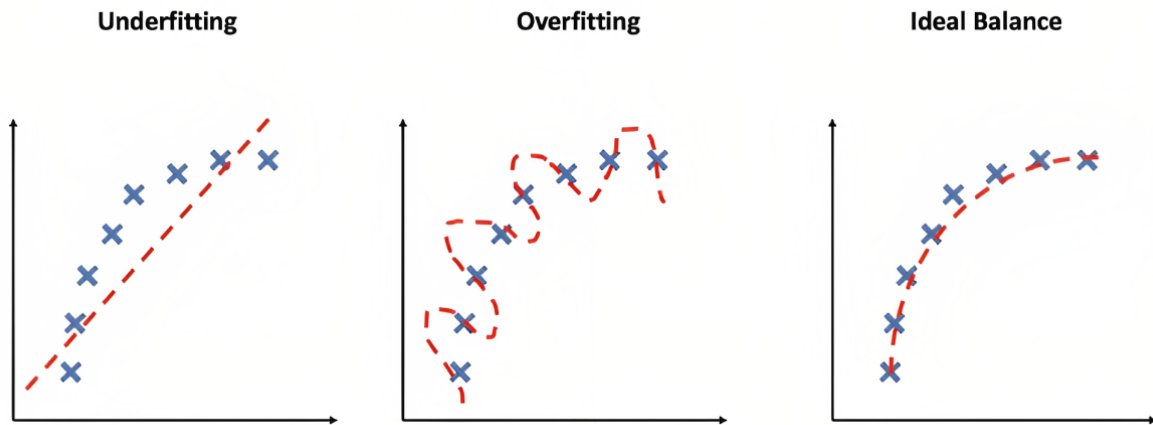
### Επιβλεπόμενη Μάθηση (Supervised Learning)

Στην *επιβλεπόμενη μάθηση*, τα δεδομένα εκπαίδευσης που τροφοδοτούνται στον αλγόριθμο περιλαμβάνουν τις επιθυμητές λύσεις, που ονομάζονται ετικέτες (labels). Αποτελεί τον πλέον πιο συχνό τύπο προβλημάτων. Ο σκοπός σε αυτό το είδος εργασιών είναι η μάθηση των αντιστοιχιών μεταξύ ενός συνόλου εισόδων  $X$  και εξόδων  $Y$ , δηλαδή η σχέση  $f(X) = Y$ . Ένα παράδειγμα ενός τέτοιου προβλήματος θα μπορούσε να είναι η ταξινόμηση ανεπιθύμητης ή και κακόβουλης αλληλογραφίας (spam emails). Η διαδικασία της εκπαίδευσης γίνεται με πολλά παραδείγματα δεδομένων email τα οποία παρέχονται μαζί με την κλάση (class) τους (είτε είναι ανεπιθύμητο είτε αποδεκτό), για να μάθει το μοντέλο πώς πρέπει να συμπεριφέρεται, εξ ου και ο όρος επιβλεπόμενη μάθηση. Η ικανότητα προσαρμογής σε νέες εισόδους και η πραγματοποίηση αξιόπιστων προβλέψεων είναι ίσως το κρισιμότερο κομμάτι της γενίκευσης της μηχανικής μάθησης. Στην εκπαίδευση η γενίκευση (generalisation) είναι αυτή που πρέπει να μεγιστοποιηθεί. Εάν το μοντέλο εκπαιδευτεί σε υπερβολικό βαθμό, δημιουργείται το πρόβλημα της υπερπροσαρμογής (overfitting), αποτυγχάνοντας έτσι να γενικεύσει σε νέα δεδομένα που δεν έχει δει. Με άλλα λόγια, το μοντέλο προσαρμόζεται πολύ καλά ή "υπερπροσαρμόζεται" στα δεδομένα εκπαίδευσης, καταγράφοντας θόρυβο και ιδιομορφίες που μπορεί να μην υπάρχουν στα μελλοντικά δεδομένα. Στο Σχήμα 2.3 απεικονίζονται κάποια παραδείγματα για τις περιπτώσεις της υποπροσαρμογής (underfitting), της υπερπροσαρμογής (overfitting) και της ιδανικής προσαρμογής (ideal fit).

Κάποιοι από τους πιο σημαντικούς αλγορίθμους στην επιβλεπόμενη μάθηση είναι οι εξής:

- Αφελής Bayes (Naive Bayes).
- k-κοντινότεροι γείτονες (k-Nearest Neighbors).
- Γραμμική παλινδρόμηση (Linear Regression).

- Λογιστική παλινδρόμηση (Logistic Regression).
- Μηχανές διανυσμάτων υποστήριξης (Support Vector Machines - SVMs).
- Δέντρα αποφάσεων και τυχαία δάση (Decision Trees and Random Forests).



**Σχήμα 2.3:** Παράδειγμα υποπροσαρμογής (αριστερά), υπερπροσαρμογής (κέντρο) και ιδανικής προσαρμογής (δεξιά) (Πηγή: [53]).

### Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning)

Στη μάθηση χωρίς επίβλεψη τα δεδομένα εκπαίδευσης δεν έχουν ετικέτες και το σύστημα προσπαθεί να μάθει χωρίς «δάσκαλο». Μόνο τα δεδομένα εισόδου παρέχονται στο σύστημα και εκείνο προσπαθεί να εντοπίσει πολύπλοκα μοτίβα που κρύβονται μέσα στα δεδομένα. Ένα παράδειγμα μη επιβλεπόμενης μάθησης θα μπορούσε να είναι η τμηματοποίηση πελατών για σκοπούς μάρκετινγκ (customer segmentation). Εστιάζοντας σε διάφορα χαρακτηριστικά μπορεί να γίνει η σωστή ομαδοποίηση των πελατών, απλώς από τα δεδομένα εισόδου.

Κάποιοι από τους πιο σημαντικούς αλγορίθμους στη μη επιβλεπόμενη μάθηση είναι οι εξής:

- Αλγόριθμοι ομαδοποίησης (Clustering):  
K-Means, DBSCAN, Hierarchical Cluster Analysis (HCA).
- Αλγόριθμοι ανίχνευσης ανωμαλιών (Anomaly detection):  
One-class SVM, Isolation Forest.
- Αλγόριθμοι οπτικοποίησης και μείωσης της διαστατικότητας (Visualization and dimensionality reduction):  
Principal Component Analysis (PCA), Kernel PCA.
- Αλγόριθμοι μάθησης κανόνων συσχέτισης (Association rule learning):  
Apriori, Eclat.

---

## Ημιεπιβλεπόμενη Μάθηση (Supervised Learning)

Ορισμένοι αλγόριθμοι μπορούν να διαχειριστούν δεδομένα εκπαίδευσης με μερική επισήμανση, συνήθως πολλά δεδομένα χωρίς ετικέτες και λίγα δεδομένα με ετικέτες. Αυτή η προσέγγιση καλείται *ημιεπιβλεπόμενη μάθηση* και αποτελεί ουσιαστικά μία μίξη των δύο προηγούμενων ειδών μάθησης. Ένα χαρακτηριστικό παράδειγμα αυτού του είδους είναι η αναγνώριση εικόνων από υπηρεσίες όπως η Google Photos. Η συγκεκριμένη υπηρεσία μπορεί να αναγνωρίσει τα άτομα που βρίσκονται σε διαφορετικές φωτογραφίες και να τα ομαδοποιεί (μη επιβλεπόμενη μάθηση). Εάν ο χρήστης του υποδείξει τα ονόματα αυτών των ατόμων σε πολύ λίγες φωτογραφίες (επιβλεπόμενη μάθηση), η υπηρεσία είναι σε θέση να αναγνωρίσει όλα τα πρόσωπα των φωτογραφιών. Σε πολλά προβλήματα μηχανικής μάθησης η επισήμανση των δεδομένων μπορεί να είναι δαπανηρή ή και χρονοβόρα (όπως η επισήμανση χειροκίνητα πολλών εικόνων) ενώ από την άλλη τα δεδομένα χωρίς ετικέτες είναι φθηνά και εύκολα στη συλλογή και την αποθήκευση. Έτσι, αυτό το είδος μάθησης είναι αρκετά σημαντικό.

Κάποιοι από τους πιο σημαντικούς αλγορίθμους στην ημιεπιβλεπόμενη μάθηση είναι οι εξής:

- Deep Belief Networks (DBNs).
- Restricted Boltzmann Machines (RBMs).
- Generative Adversarial Networks (GANs).

## Ενισχυτική Μάθηση (Reinforcement Learning)

Η *ενισχυτική μάθηση* (reinforcement learning) είναι ένας γενικός όρος που έχει δοθεί σε μια οικογένεια τεχνικών στις οποίες το σύστημα μάθησης προσπαθεί να μάθει μέσα από την άμεση αλληλεπίδραση με το περιβάλλον. Το σύστημα μάθησης που ονομάζεται πράκτορας (agent) μπορεί να παρατηρεί το περιβάλλον, να επιλέγει και να εκτελεί ενέργειες και να λαμβάνει αποφάσεις οι οποίες του αποφέρουν είτε ανταμοιβές είτε ποινές. Στη συνέχεια, πρέπει να μάθει μόνο του ποια είναι η καλύτερη στρατηγική, που ονομάζεται πολιτική (policy), για να λάβει τη μεγαλύτερη ανταμοιβή με την πάροδο του χρόνου. Μια πολιτική ορίζει ποια ενέργεια θα πρέπει να επιλέξει ο πράκτορας όταν βρίσκεται σε μια δεδομένη κατάσταση (state). Στόχος ενός αλγορίθμου ενισχυτικής μάθησης είναι η μεγιστοποίηση των ανταμοιβών του παίζοντας το παιχνίδι συνέχεια. Εφαρμόζεται στον έλεγχο κίνησης ρομπότ, στη βελτιστοποίηση εργασιών σε εργοστάσια, στη μάθηση επιτραπέζιων παιχνιδιών, στη βελτιστοποίηση τροχιάς, σχεδιασμό κίνησης κτλ.

Κάποιοι από τους πιο σημαντικούς αλγορίθμους στην ενισχυτική μάθηση είναι οι εξής:

- Q-Learning.
- Deep Q-Networks (DQN).
- Policy Gradient Methods

- 
- AlphaGo.

## 2.3 Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks - ANN)

*Νευρωνικό δίκτυο (Neural network)* ονομάζεται ένα σύστημα διασυνδεδεμένων νευρώνων. Στην περίπτωση βιολογικών νευρώνων, πρόκειται για ένα τμήμα νευρικού ιστού. Στην περίπτωση τεχνητών νευρώνων, πρόκειται για ένα αφηρημένο αλγοριθμικό κατασκεύασμα αποτελούμενο από κόμβους το οποίο εμπίπτει στον τομέα της υπολογιστικής νοημοσύνης, όταν στόχος του νευρωνικού δικτύου είναι η επίλυση κάποιου υπολογιστικού προβλήματος.

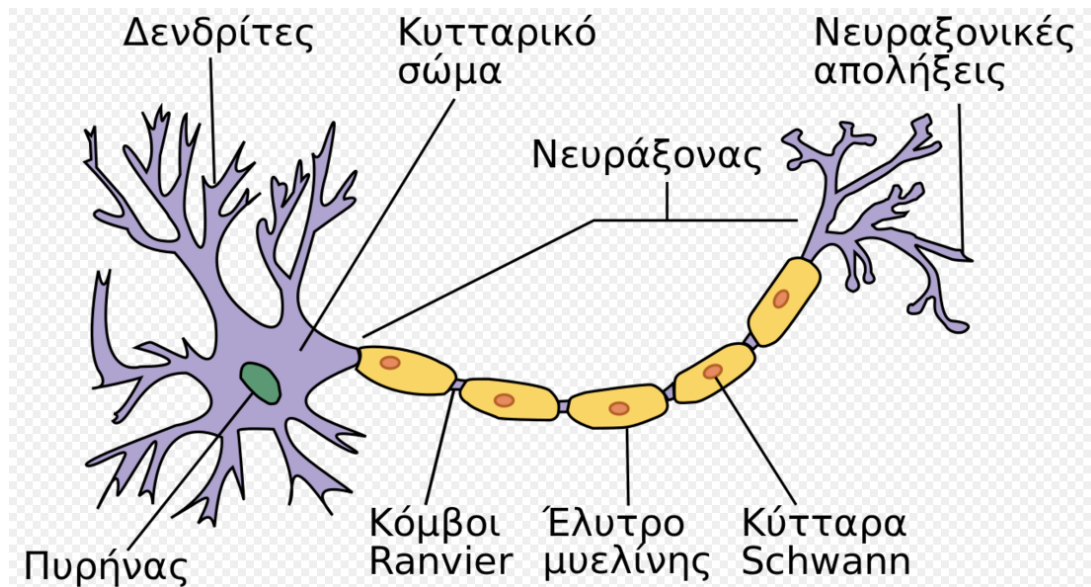
Η κύρια ιδέα που κρύβεται πίσω από την ανάπτυξη των τεχνητών νευρωνικών δικτύων (ΤΝΔ) ή artificial neural networks (ANN), είναι η πιστή αντιγραφή της δομής με την οποία είναι φτιαγμένος ο ανθρώπινος εγκέφαλος. Η υλοποίηση ενός «εγκεφάλου» από το μηδέν, όπως ο άνθρωπος, είναι ένα μεγάλο στοίχημα της ανθρωπότητας. Κάτι τέτοιο όμως είναι πολύ δύσκολο λόγω της πολυπλοκότητάς του. Εκτιμάται ότι ο ανθρώπινος εγκέφαλος αποτελείται από περισσότερα από 100 τρισεκατομμύρια (περίπου) νευρικά κύτταρα, νευρογλοιακά κύτταρα και νευρώνες ως επί το πλείστον. Η προσέγγιση λοιπόν αυτή, της αντιγραφής, θα πάρει αρκετό καιρό με εκτιμήσεις να δείχνουν πως χρειάζονται πάνω από 100 χρόνια για να αποκρυπτογραφηθεί η λειτουργία του ανθρώπινου εγκεφάλου. Είναι εφικτό ωστόσο η δημιουργία εναλλακτικών για την αντιμετώπιση του προβλήματος αυτού. Έτσι τα ΤΝΔ, αφού πήραν τα βασικά χαρακτηριστικά του βιολογικού διδύμου τους, σταδιακά διαφοροποιήθηκαν από αυτά. Η οργάνωση του υπκεφαλαίου αυτού αποτελεί αναπαραγωγή του αντίστοιχου κεφαλαίου στη προπτυχιακή μου διπλωματική εργασία [54].

### 2.3.1 Από τους βιολογικούς στους τεχνητούς νευρώνες

Ο *βιολογικός νευρώνας* [55] είναι το κύτταρο που αποτελεί δομικό μέρος και λειτουργική μονάδα του νευρικού συστήματος. Κάθε νευρώνας, όπως φαίνεται στο Σχήμα 2.4, αποτελείται από το κυτταρικό σώμα, τον άξονα, και τους δενδρίτες. Ο κυρίως κορμός του νευρώνα είναι το σώμα μέσα στο οποίο βρίσκεται όλο το γενετικό υλικό του οργανισμού. Ο άξονας είναι μια μεγάλη επέκταση, που το μέγεθος του μπορεί να είναι από μερικές έως και δεκάδες φορές μεγαλύτερο από το σώμα και εφάπτεται με άλλους νευρώνες. Οι άξονες σε μερικούς νευρώνες είναι καλυμμένοι με μια ουσία, που λέγεται μυελίνη, ενώ άλλοι είναι τελείως ακάλυπτοι. Κάθε νευρώνας έχει ένα μόνο άξονα, ο οποίος μεταδίδει σήματα σε άλλους νευρώνες. Υπάρχουν ακόμα οι λεπτές επεκτάσεις που μοιάζουν με διακλαδώσεις δέντρου και ονομάζονται δενδρίτες. Οι δενδρίτες συλλέγουν τα σήματα που στέλνονται στο κύτταρο. Οι νευρώνες επικοινωνούν μεταξύ τους και με άλλους νευρώνες μέσω συνάψεων, όπου η άκρη του νευράξονα καταλήγει στους δενδρίτες, στο σώμα ή στο νευράξονα άλλων νευρώνων. Στα σημεία που εφάπτονται οι δενδρίτες δημιουργείται η σύναψη. Ο άξονας έχει συνήθως πάρα πολλές διακλαδώσεις και



μπορεί να στέλνει σήματα σε διαφορετικά σημεία, πράγμα που τους επιτρέπει να επικοινωνούν ταυτόχρονα με δεκάδες χιλιάδες νευρικά κύτταρα. Το Σχήμα 2.4 συνοψίζει τα παραπάνω.



**Σχήμα 2.4:** Βιολογικός νευρώνας / Από User:Dhp1080, translated and modified by Badseed. - Originally Neuron.jpg taken from the US Federal (public domain) (Nerve Tissue, retrieved March 2007), redrawn by User:Dhp1080 in Illustrator. Source: "Anatomy and Physiology" by the US National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) Program., CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=24139135>.

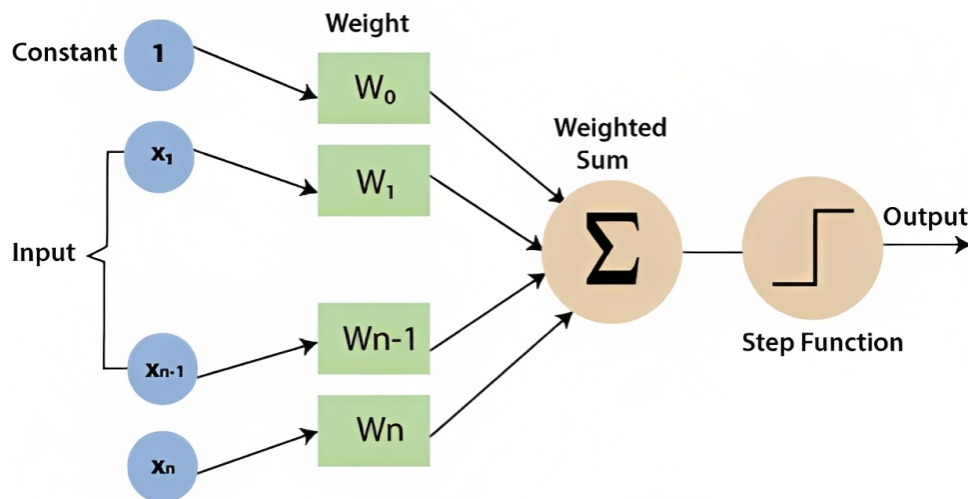
Οι βιολογικοί νευρώνες μεμονωμένα φαίνεται να συμπεριφέρονται με σχετικά απλό τρόπο, όμως είναι οργανωμένοι σε τεράστια δίκτυα αποτελούμενα από δισεκατομμύρια νευρώνες, με τον καθένα από αυτούς να συνδέεται με χιλιάδες άλλους. Είναι προφανές λοιπόν το πόσο περίπλοκοι μπορούν να γίνουν οι υπολογισμοί από ένα τέτοιο τεράστιο δίκτυο. Τα βιολογικά νευρωνικά δίκτυα είναι ένα αντικείμενο έρευνας αλλά κάποια κομμάτια έχουν χαρτογραφηθεί. Όπως μπορούμε να διακρίνουμε και στο Σχήμα 2.5, οι νευρώνες είναι οργανωμένοι σε διαδοχικά επίπεδα (layers). Το τεχνητό νευρωνικό δίκτυο όπως αναφέρθηκε φτιάχτηκε για να μιμηθεί το βιολογικό. Έτσι λοιπόν περιλαμβάνει πολλούς κόμβους (νευρώνες) διασυνδεδεμένους μεταξύ τους, που περνάνε πληροφορίες μέσα από τις συνδέσεις τους και ενεργοποιούνται όταν ξεπεράσουν ένα κατώφλι (threshold), βάρη (weights) και μια εξίσωση ενεργοποίησης (activation function). Οι συνδέσεις μεταξύ των κόμβων αντιπροσωπεύουν τον άξονα και τους δενδρίτες, τα βάρη αντιπροσωπεύουν τις συνάψεις ενώ η συνάρτηση ενεργοποίησης προσεγγίζει την δραστηριότητα στο σώμα.

Εάν απομονωθεί ένας από αυτούς τους κόμβους όπως απεικονίζεται στο Σχήμα 2.6 μπορούμε να δούμε την λειτουργία αυτού του απλού νευρωνικού δικτύου. Το μοντέλο αυτό ονομάζεται *Perceptron* και εφευρέθηκε από τον Frank Rosenblatt το 1957 [56]. Οι είσοδοι (inputs) είναι νούμερα όπως και η έξοδος και κάθε είσοδος συνδέεται με ένα βάρος (weight). Υπολογίζεται το σταθμισμένο άθροισμα ή *weighted sum* ( $z = w_1x_1 + w_2x_2 + \dots + w_nx_n = \mathbf{x}^T \mathbf{w}$ ), και



**Σχήμα 2.5:** Πολλαπλά επίπεδα σε ένα βιολογικό νευρωνικό δίκτυο.

έπειτα εφαρμόζεται σε αυτό μια συνάρτηση ενεργοποίησης (όπου στο σχήμα φαίνεται η βηματική συνάρτηση - step function). Τέλος η έξοδος από το νευρωνικό είναι το αποτέλεσμα  $h_w(\mathbf{x}) = \text{step}(z)$ , όπου  $z = \mathbf{x}^T \mathbf{w}$ . Όπως βλέπουμε στην Σχήμα 2.6 υπάρχει ένας νευρώνας με αρχική τιμή εισόδου  $x_0 = 1$  και βάρος  $w_0$ . Αυτό το βάρος καλείται πόλωση ή κατώφλι (bias, threshold) και μας δείχνει πόσο μεγάλο πρέπει να είναι το σταθμισμένο άθροισμα ώστε να αρχίζει ο νευρώνας να ενεργοποιείται. Εάν είναι μικρότερο από τη συγκεκριμένη τιμή τότε ο νευρώνας είναι ανενεργός. Αυτό επιτυγχάνεται με πρόσθεση του κατάλληλου αριθμού στο σταθμισμένο άθροισμα πριν περάσει από αυτήν η συνάρτηση ενεργοποίησης. Με αυτόν τον τρόπο προσομοιώνεται η λειτουργία των βιολογικών νευρικών κυττάρων, που κάποια μένουν ανενεργά ενώ άλλα ενεργοποιούνται.



**Σχήμα 2.6:** Μορφή τεχνητού νευρωνικού δικτύου Perceptron.

Το παραπάνω αποτελεί ένα μόνο κόμβο. Όπως αναφέρθηκε όμως κάθε κόμβος συνδέεται με άλλους και εκείνος με την σειρά του με άλλους φτιάχνοντας έτσι ένα δίκτυο, περιπλέκοντας ως ένα βαθμό την υπολογιστική διαδικασία. Χάρη στην γραμμική άλγεβρα είναι εφικτό να υπολογιστούν αποτελεσματικά οι έξοδοι ενός τέτοιου νευρωνικού δικτύου. Η εξίσωση που

---

χρησιμοποιείται είναι η εξής:

$$h_{W,b}(\mathbf{X}) = \varphi(\mathbf{XW} + \mathbf{b}) \quad (2.1)$$

, όπου:

- Το  $\mathbf{X}$  αντιπροσωπεύει τον πίνακα των εισόδων.
- Το  $\mathbf{W}$  έχει όλα τα συνδεδεμένα βάρη εκτός από αυτά της πόλωσης.
- Το  $\mathbf{b}$  που αποτελεί τον πίνακα με τους αριθμούς της πόλωσης.
- Το  $\varphi$  είναι η συνάρτηση ενεργοποίησης που θα αναλυθεί παρακάτω.

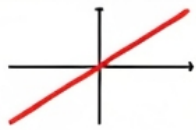
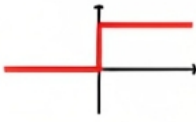
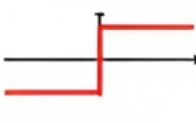




### 2.3.2 Συνάρτηση ενεργοποίησης (Activation function)

Το αποτέλεσμα που προκύπτει από το σταθμισμένο άθροισμα μετά και την προσθήκη της πόλωσης ( $z = w_1x_1 + w_2x_2 + \dots + w_nx_n + b$ ) μπορεί να είναι οποιοσδήποτε αριθμός στο διάστημα  $(-\infty, +\infty)$ . Ο νευρώνας δεν γνωρίζει τα όρια της μεταβλητής, οπότε θα πρέπει να αποφασιστεί τότε ο νευρώνας θα είναι ενεργός και πότε όχι, για να ενσωματωθεί η λειτουργία του ανθρώπινου εγκεφάλου. Αυτό το σκοπό έρχεται να καλύψει η *συνάρτηση ενεργοποίησης*. Ουσιαστικά συμπίεζει το σταθμισμένο άθροισμα σε ένα διάστημα ανάλογα με την συνάρτηση που θα επιλέξουμε, για παράδειγμα στο  $[0, 1]$  για την βηματική συνάρτηση (step function) και δηλώνει ότι για την τιμή 0 ο νευρώνας είναι ανενεργός ενώ για την τιμή 1 ενεργοποιείται. Άλλες συναρτήσεις που δεν έχουν μόνο δύο εξόδους, όπως η βηματική, αλλά αριθμούς σε ένα διάστημα δηλώνουν το πόσο ενεργός είναι ένας νευρώνας ανάλογα με την τιμή που έχει. Για παράδειγμα ένας νευρώνας με τιμή 0.2 είναι προφανώς λιγότερο ενεργός σε σχέση με εκείνον με τιμή 0.8.

Ένας ακόμη πολύ σημαντικός λόγος για τον οποίο χρησιμοποιούνται οι συναρτήσεις ενεργοποίησης είναι η μη γραμμικότητα που θέλουμε να προσδώσουμε στο πρόβλημα. Εάν πραγματοποιηθεί μια ζεύξη γραμμικών μετασχηματισμών, όπως γίνεται στα διάφορα επίπεδα των νευρώνων, το μόνο που λαμβάνεται ως αποτέλεσμα είναι ένας γραμμικός μετασχηματισμός. Άρα εάν δεν υπάρξει κάτι για να προσδώσει μη γραμμικότητα ανάμεσα στα επίπεδα των νευρώνων, τότε ακόμα και οι πολλές στρώσεις νευρώνων θα είναι ισοδύναμες με μία. Κάτι τέτοιο καθιστά αδύνατη την επίλυση περίπλοκων προβλημάτων. Συνεπώς η χρήση της συνάρτησης ενεργοποίησης καθίσταται απαραίτητη. Στον Πίνακα 2.1 περιγράφονται συνοπτικά οι πιο διαδεδομένες συναρτήσεις ενεργοποίησης και στο Σχήμα 2.7 φαίνονται τα γραφήματα των δημοφιλέστερων συναρτήσεων (step, logistic, tanh, ReLU), όπως και τα γραφήματα των παραγώγων τους.

Ακόμη μία αρκετά διαδεδομένη συνάρτηση ενεργοποίησης είναι η GELU. Η GELU (Gaussian Error Linear Unit) είναι μια μη γραμμική συνάρτηση ενεργοποίησης που χρησιμοποιείται συ-

**Πίνακας 2.1:** Συναρτήσεις ενεργοποίησης (Πηγή: Sebastian Raschka, 2016).

Activation Function	Equation	Example	1D Graph
Linear	$\phi(z) = z$	Adaline, linear regression	
Unit Step (Heaviside Function)	$\phi(z) = \begin{cases} 0 & z < 0 \\ 0.5 & z = 0 \\ 1 & z > 0 \end{cases}$	Perceptron variant	
Sign (signum)	$\phi(z) = \begin{cases} -1 & z < 0 \\ 0 & z = 0 \\ 1 & z > 0 \end{cases}$	Perceptron variant	
Piece-wise Linear	$\phi(z) = \begin{cases} 0 & z \leq -1/2 \\ z + 1/2 & -1/2 \leq z \leq 1/2 \\ 1 & z \geq 1/2 \end{cases}$	Support vector machine	
Logistic (sigmoid)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Logistic regression, Multilayer NN	
Hyperbolic Tangent (tanh)	$\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	Multilayer NN, RNNs	
ReLU	$\phi(z) = \begin{cases} 0 & z < 0 \\ z & z > 0 \end{cases}$	Multilayer NN, CNNs	

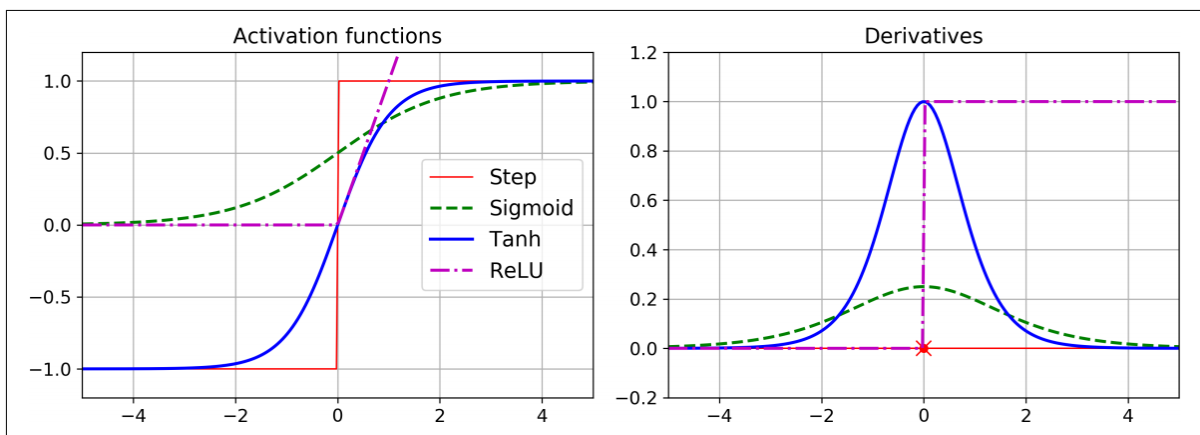
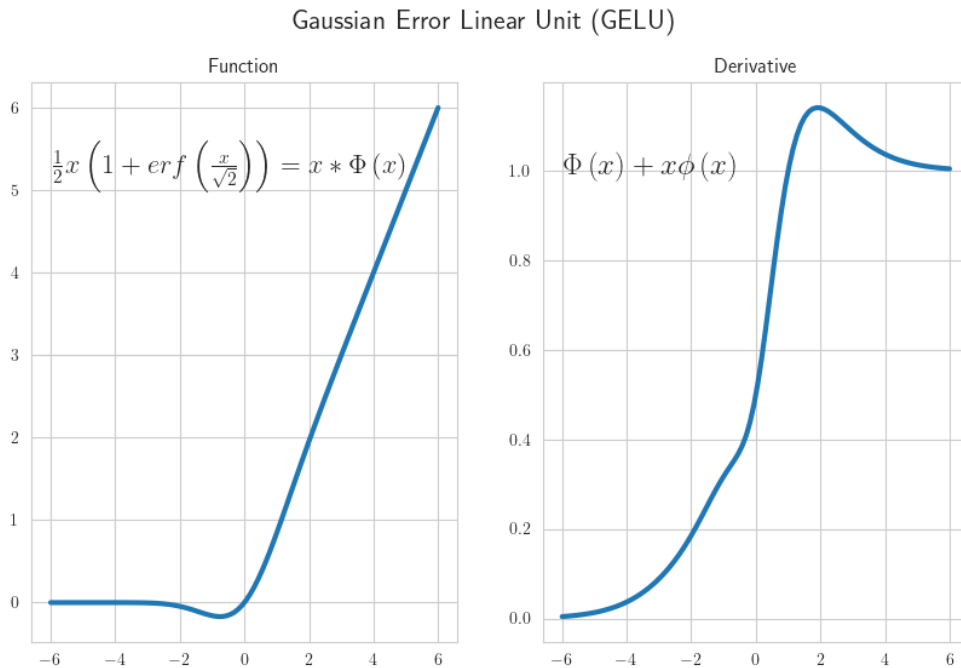


Figure 10-8. Activation functions and their derivatives

**Σχήμα 2.7:** Συναρτήσεις ενεργοποίησης και οι παράγωγοι τους (Πηγή: Aurélien Géron, 2019 [57]).

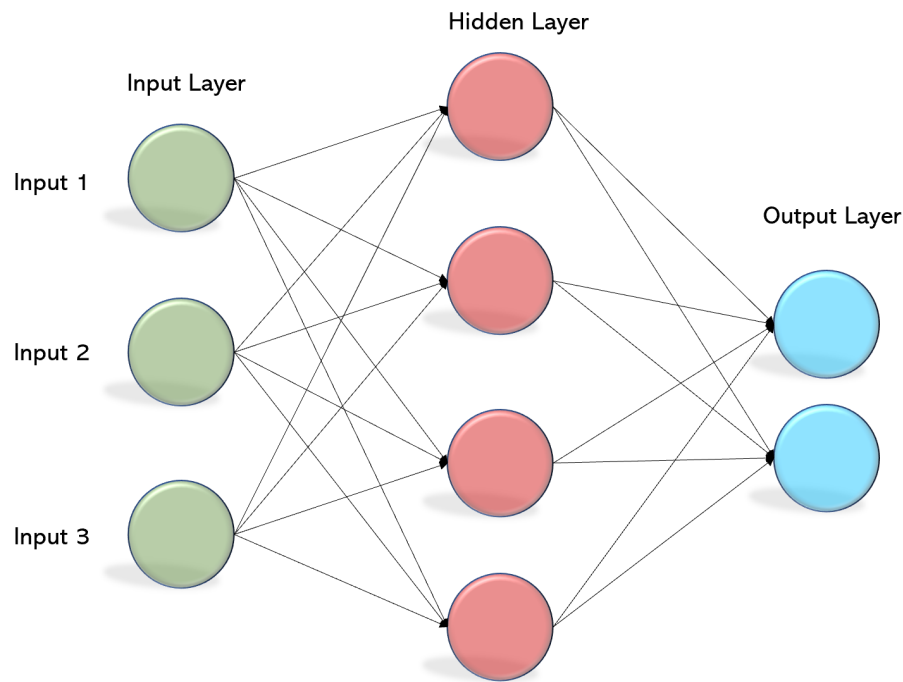
νήθως σε μοντέλα βαθιάς μάθησης. Εισήχθη ως εναλλακτική λύση σε άλλες συναρτήσεις ενεργοποίησης όπως η ReLU (Rectified Linear Unit) για να μετριάσει ορισμένους περιορισμούς της. Βοηθά το μοντέλο να μάθει πιο σύνθετα μοτίβα στα δεδομένα, ενώ παρουσιάζει ορισμένες επιθυμητές ιδιότητες όπως ότι είναι παντού διαφορίσιμη, σε αντίθεση με την ReLU. Στο Σχήμα 2.8 απεικονίζεται η συνάρτηση ενεργοποίησης GELU και η παράγωγός της.



**Σχήμα 2.8:** Η συνάρτηση ενεργοποίησης GELU και η παράγωγός της. (Πηγή: [58])

### 2.3.3 Πολυεπίπεδα τεχνητά νευρωνικά δίκτυα (Multilayer Perceptron)

Τα *πολυεπίπεδα νευρωνικά δίκτυα*, γνωστά με τον όρο MLP (Multilayer Perceptron), είναι το πιο διαδεδομένο είδος νευρωνικού δικτύου και αποτελεί συνέχεια του απλού νευρωνικού που αναλύθηκε προηγουμένως. Αυτό το είδος νευρωνικού περιέχει, εκτός από τα *επίπεδα εισόδου (input layer)* και *εξόδου (output layer)*, ένα ή περισσότερα επίπεδα νευρώνων ακόμα, τα λεγόμενα *κρυφά επίπεδα (hidden layers)*. Κάθε επίπεδο περιέχει ένα *νευρώνα πόλωσης (bias neuron)*, εκτός από το τελευταίο, και είναι πλήρως συνδεδεμένο με το επόμενο, μέσω των *βάρων (weights)*. Όταν ένα νευρωνικό δίκτυο περιέχει παραπάνω από δύο κρυφά επίπεδα τότε ονομάζεται *βαθύ νευρωνικό δίκτυο (deep neural network)*. Όπως απεικονίζεται στο Σχήμα 2.9, οι συνδέσεις μεταξύ των κόμβων δεν σχηματίζουν βρόχους και έτσι η πληροφορία κατευθύνεται προς μία κατεύθυνση από την είσοδο προς την έξοδο. Η μορφή αυτή των δικτύων συχνά αναφέρεται ως *νευρωνικά δίκτυα τα οποία τροφοδοτούνται προς τα εμπρός (feedforward neural networks)* και αποτελούν μία πολύ διαδεδομένη μορφή δικτύων.



Σχήμα 2.9: Πολυεπίπεδο ΤΝΔ (Multi-layer Perceptron).

### 2.3.4 Εκπαίδευση τεχνητών νευρωνικών δικτύων (Training)

Το κύριο χαρακτηριστικό των νευρωνικών δικτύων είναι η ικανότητα της μάθησης. Με τον όρο *μάθηση (learning)* στα ΤΝΔ, αναφερόμαστε στο να κατορθώσει ο υπολογιστής να βρει όλους αυτούς τους αριθμούς, τα βάρη και τις πολώσεις (weights and biases), ώστε τελικά να λύσουν το πρόβλημα που του δόθηκε. Κάτι τέτοιο επιτυγχάνεται με την εκπαίδευση του νευρωνικού, μια επαναληπτική διαδικασία σταδιακής προσαρμογής των παραμέτρων του δικτύου. Στη φάση της εκπαίδευσης στο νευρωνικό δίκτυο δίνεται ένα σετ δεδομένων, που αφορά την είσοδο και την έξοδο του συγκεκριμένου προβλήματος που θέλουμε να επιλυθεί, και εκείνο ψάχνει να βρει τις κατάλληλες παραμέτρους (βάρη και πολώσεις) ώστε με τις εισόδους που του δόθηκαν να επιστρέφει τις εξόδους για τις οποίες τροφοδοτήθηκε. Αφού εκπαιδευτεί το ΤΝΔ από εκεί και πέρα είναι σε λειτουργική κατάσταση για μελλοντικές εφαρμογές, με άλλα λόγια για καινούργια δεδομένα (εισόδους), με γνωστές πλέον παραμέτρους, θα παραγάγει τις εξόδους του προβλήματος.

Πως όμως το ΤΝΔ γνωρίζει ποιες παράμετροι είναι οι κατάλληλες;

Θα πρέπει να καθοριστεί ένα μέτρο απόδοσης. Ένας τρόπος να γίνει αυτό είναι μέσω μιας *συνάρτησης κόστους* όπου η ελαχιστοποίηση της είναι ο επιθυμητός στόχος. Επομένως με τη μάθηση νοείται η εύρεση εκείνων των βαρών και των πολώσεων που ελαχιστοποιούν μια συγκεκριμένη συνάρτηση κόστους. Η συνηθέστερη συνάρτηση κόστους είναι η αυτή του μέσου τετραγωνικού σφάλματος (Mean Square Error – MSE). Η MSE ουσιαστικά παίρνει τα τετράγωνα των διαφορών ανάμεσα σε κάθε έξοδο του νευρωνικού και κάθε πραγματική τιμή (η έξοδος που θέλουμε από τα δεδομένα μας) και τελικά παίρνει τον μέσο όρο αυτών των όρων.

$$C(y, o) = \frac{1}{N} \sum_{i=1}^N (y_i - o_i)^2 \quad (2.2)$$

,όπου  $y_i$  είναι οι προβλέψεις από το τεχνητό νευρωνικό δίκτυο και  $o_i$  είναι οι πραγματικές τιμές που θα έπρεπε να έχουμε.

### 2.3.5 Αλγόριθμοι Βελτιστοποίησης

Υπάρχουν αρκετοί αλγόριθμοι βελτιστοποίησης και στη συνέχεια περιγράφονται οι πιο διαδεδομένοι [59, 60].

#### Κατάβαση κλίσης (Gradient Descent)

Αρχικά ο προσδιορισμός και η κατανόηση του αλγορίθμου *κατάβασης κλίσης* (*gradient descent*) αποτελεί κρίσιμο σημείο για τους αλγορίθμους βελτιστοποίησης των ΤΝΔ, καθώς οι περισσότεροι μετέπειτα αλγόριθμοι βασίζονται σε αυτόν. Είναι ένας πολύ γενικός αλγόριθμος βελτιστοποίησης ικανός να βρει την βέλτιστη λύση σε ένα ευρύ φάσμα προβλημάτων. Η γενική ιδέα της κατάβασης κλίσης είναι να τροποποιήσει τις παραμέτρους της εξίσωσης προκειμένου να ελαχιστοποιήσει μια συνάρτηση κόστους. Ας υποθέσουμε ότι βρισκόμαστε στην κορυφή κάποιου βουνού (υψηλό κόστος) και θέλουμε να κατέβουμε στο κάτω μέρος της κοιλάδας (χαμηλό κόστος) (Σχήμα 2.10). Μια καλή ιδέα για να γίνει αυτό είναι να πάμε κατηφορικά προς την κατεύθυνση της απότομης πλαγιάς, με κάποιο μέγεθος βημάτων είτε μικρό είτε μεγάλο. Αυτό ακριβώς είναι που κάνει ο αλγόριθμος κατάβασης κλίσης, υπολογίζει την αρνητική μερική παράγωγο της συνάρτησης κόστους ως προς το διάνυσμα των παραμέτρων της και πηγαίνει προς την κατεύθυνση της φθίνουσας κλίσης. Έπειτα υπολογίζεται ξανά η αρνητική μερική παράγωγος στο σημείο με τις καινούργιες συντεταγμένες αυτή τη φορά. Η διαδικασία αυτή συνεχίζεται μέχρι να επιτευχθεί ένα ολικό ή τοπικό ελάχιστο της συνάρτησης, δηλαδή να μην είναι εφικτή η περαιτέρω κατάβαση προς τα κάτω.

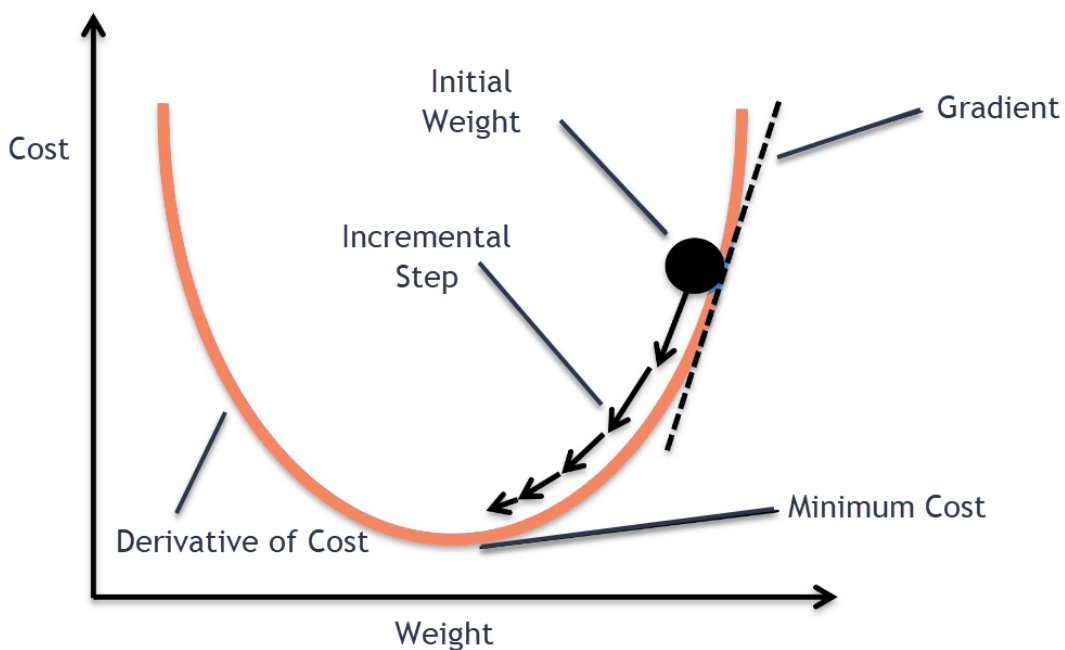
Το μέγεθος αυτών των βημάτων ονομάζεται *ρυθμός μάθησης* (*learning rate*). Εάν ο ρυθμός αυτός είναι μικρός θα πρέπει να κάνουμε περισσότερες επαναλήψεις για να συγκλίνει, κάτι που θα χρειαστεί αρκετό χρόνο. Από την άλλη εάν ο ρυθμός μάθησης είναι μεγάλος υπάρχει κίνδυνος να υπερπηδήσει το χαμηλότερο σημείο και να πάμε από την άλλη πλευρά της κοιλάδας αφού η κλίση αλλάζει συνεχώς. Αυτό μπορεί να οδηγήσει τον αλγόριθμο να αποκλίνει με αποτέλεσμα να αποτύχει να βρει μια καλή λύση. Κάτι ανάλογο παρουσιάζεται στο Σχήμα 2.11, όπου συγκρίνονται οι δύο ακραίες τιμές του ρυθμού μάθησης. Ο αλγόριθμος περιγράφεται συνοπτικά ως εξής:

$$\mathbf{g} = \nabla_{\theta} \sum_i C(f(\mathbf{x}^{(i)}; \theta), y^{(i)}) \quad (2.3)$$

$$\theta = \theta - e_k \times g \quad (2.4)$$

Αναζητείται η παράμετρος για την οποία ελαχιστοποιείται η συνάρτηση κόστους  $C$ , με τον ρυθμό μάθησης  $e_k$  να παίζει σημαντικό ρόλο.

Η ολοκλήρωση ενός πλήρη κύκλου με το σύνολο των δεδομένων εκπαίδευσης του τεχνητού νευρωνικού δικτύου καλείται *εποχή* (*epoch*). Συνήθως για την εκπαίδευση των ΤΝΔ χρειάζονται παραπάνω από μία εποχές (*epochs*). Ο αριθμός των εποχών ποικίλει ανάλογα με το πρόβλημα που αντιμετωπίζουμε και μπορεί να φτάνει τις μερικές εκατοντάδες ή και χιλιάδες για πολύπλοκα προβλήματα.



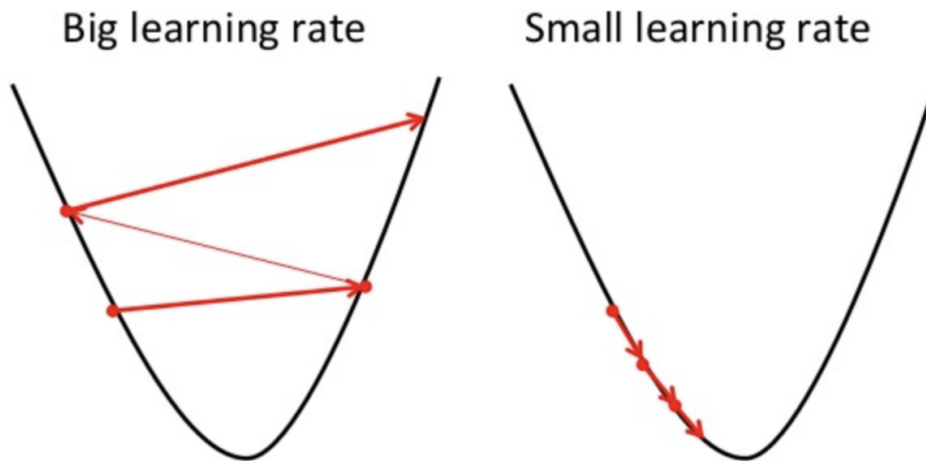
**Σχήμα 2.10:** Σχηματική απεικόνιση αλγορίθμου κατάβασης κλίσης (Gradient descent).

### Οπισθοδιάδοση (backpropagation)

Ο αλγόριθμος οπισθοδιάδοσης (*backpropagation*) [61] είναι ένας ευρέως διαδεδομένος αλγόριθμος και καθίσταται πολύ σημαντικός για την εκπαίδευση νευρωνικών δικτύων, καθώς υπολογίζει τις μερικές παραγώγους με έναν αποτελεσματικό τρόπο. Εφόσον οι παράγωγοι υπολογιστούν γίνεται απλή και εφικτή η εφαρμογή αλγορίθμων βελτιστοποίησης όπως αυτός της κατάβασης κλίσης (*gradient descent*), κάνοντας χρήση αυτών των παραγώγων.

Για ένα δεδομένο τεχνητό νευρωνικό δίκτυο και για μια συνάρτηση κόστους, η μέθοδος της οπισθοδιάδοσης (*backpropagation*) υπολογίζει με μια αποτελεσματική τεχνική τις μερικές παραγώγους ως προς τα βάρη, από την συνάρτηση κόστους. Με άλλα λόγια μπορεί να βρει πως πρέπει να τροποποιηθεί το κάθε βάρος και η κάθε πόλωση ώστε να μειωθεί το σφάλμα. Η λέξη «όπισθεν» που εμπεριέχεται στην ονομασία της μεθόδου αναφέρεται στο γεγονός ότι οι παρά-





**Σχήμα 2.11:** Παράδειγμα επιλογής μεγάλου (αριστερά) και μικρού (δεξιά) ρυθμού μάθησης.

γωγοί υπολογίζονται προς τα πίσω μέσα στο δίκτυο, με τις παραγώγους του τελευταίου επιπέδου νευρώνων να υπολογίζονται πρώτες και του πρώτου επιπέδου να υπολογίζονται τελευταίες. Οι υπολογισμένες μερικές παράγωγοι από το ένα επίπεδο (layer) ξαναχρησιμοποιούνται για τον υπολογισμό των παραγώγων του προηγούμενου επιπέδου, επηρεάζοντας έτσι η μία την άλλη. Αυτή η ροή προς τα πίσω της πληροφορίας για την συνάρτηση κόστους επιτρέπει ένα πιο αποτελεσματικό υπολογισμό των παραγώγων σε κάθε επίπεδο, έναντι της απλής προσέγγισης για τον υπολογισμό των παραγώγων σε κάθε επίπεδο ξεχωριστά.

Συνοψίζοντας λοιπόν, για κάθε εκπαίδευση που συμβαίνει ο αλγόριθμος οπισθοδιάδοσης πρώτα κάνει μια πρόβλεψη (προς τα εμπρός πέρασμα), μετράει το σφάλμα της συνάρτησης κόστους, έπειτα πηγαίνει σε κάθε επίπεδο αντίστροφα για να υπολογίσει την συνεισφορά που έχει κάθε σύνδεση νευρώνων στο σφάλμα και τελικά τροποποιεί ελαφρώς τα βάρη και τις πολώσεις για να ελαχιστοποιήσει το τελικό σφάλμα.

Στην παραπάνω διαδικασία κεντρικός είναι ο ρόλος του κανόνα της αλυσίδας (Chain rule) που έγινε γνωστός όταν οι Newton και Leibniz ανακάλυψαν τον απειροστικό λογισμό στα τέλη του 17ου αιώνα. Αν μια μεταβλητή  $z$  εξαρτάται από μια  $y$  και εκείνη με την σειρά της εξαρτάται από μια άλλη μεταβλητή  $x$  τότε ο κανόνας της αλυσίδας διατυπώνεται όπως φαίνεται στην εξίσωση (2.5):

$$\frac{\partial z}{\partial x_i} = \sum_j \frac{\partial z}{\partial y_j} \frac{\partial y_j}{\partial x_i} \quad (2.5)$$

Στο πλαίσιο των νευρωνικών δικτύων είναι σημαντικό να υπολογιστεί το πόσο επηρεάζει η κάθε παράμετρος την τελική συνάρτηση κόστους για να μπορέσει να την ελαχιστοποιήσει. Ξεκινώντας από το τέλος του νευρωνικού δικτύου, η συνάρτηση κόστους για ένα μόνο νευρώνα εξαρτάται από την τιμή που έχει σαν έξοδο και την πραγματική τιμή, η οποία είναι σταθερή. Η

τιμή εξόδου  $y_i$  όμως εξαρτάται από τα βάρη, τις τιμές των προηγούμενων νευρώνων και των πολώσεων, όπου συνολικά μπορεί να οριστεί αυτό το σταθμισμένο άθροισμα ως  $z$ . Στη συνέχεια εφαρμόζεται στο  $z$  μια συνάρτηση ενεργοποίησης για να προσδιοριστεί η τελική τιμή του νευρώνα. Για να υπολογιστεί λοιπόν η επιρροή που θα έχει μια μικρή μεταβολή των βαρών στην συνάρτηση κόστους, θα πρέπει να βρεθεί η μερική παράγωγος της συνάρτησης κόστους ως προς το αντίστοιχο βάρος που μελετάται. Σε αυτό το σημείο είναι που εφαρμόζεται ο κανόνας της αλυσίδας για να υπολογίσει αυτή την παράγωγο. Συγκεκριμένα για τα νευρωνικά δίκτυα ο κανόνας της αλυσίδας διατυπώνεται:

$$\frac{\partial C}{\partial w_{ji}} = \frac{\partial z_i}{\partial w_{ji}} \frac{\partial y_i}{\partial z_i} \frac{\partial C}{\partial y_i} \quad (2.6)$$

με τις μερικές παραγώγους τους δεξιού μέλους να είναι υπολογίσιμες. Με τον ίδιο τρόπο υπολογίζεται και η συνεισφορά των υπόλοιπων παραμέτρων στη συνάρτηση κόστους. Η ίδια λογική μπορεί να εφαρμοστεί προς τα πίσω για τα βάρη και τις πολώσεις των προηγούμενων νευρώνων μέχρι την αρχή του δικτύου και έτσι τελικά υπολογίζεται το ανάδελτα  $\nabla C$  της συνάρτησης κόστους.

### Στοχαστική κατάβαση κλίσης (Stochastic gradient descent)

Το πρόβλημα με τον αλγόριθμο κατάβασης κλίσης (gradient descent) έγκειται στο κομμάτι των δεδομένων που χρησιμοποιεί. Πιο συγκεκριμένα κάνει χρήση όλων των δεδομένων εκπαίδευσης που διαθέτει για να υπολογίσει τις παραγώγους σε κάθε βήμα, πράγμα που καθιστά τον αλγόριθμο αρκετά αργό όταν το σύνολο των δεδομένων είναι πολύ μεγάλο. Από την άλλη, ο αλγόριθμος της *στοχαστικής κατάβασης κλίσης* (Stochastic Gradient Descent ή SGD) επιλέγει με τυχαίο τρόπο ένα κομμάτι από αυτά τα δεδομένα εκπαίδευσης και υπολογίζει τις παραγώγους μόνο σε αυτό το κομμάτι. Αυτή η παραλλαγή της κλασικής περίπτωσης κάνει τον αλγόριθμο πολύ πιο γρήγορο από την στιγμή που μειώνονται τα δεδομένα που διαχειρίζεται σε κάθε επανάληψη. Ωστόσο λόγω της στοχαστικότητας που διατρέχει το σύστημα θα υπάρξουν μικρές και μεγάλες μεταβολές στην συνάρτηση κόστους με το τελικό αποτέλεσμα των τιμών των παραμέτρων να είναι καλό αλλά όχι το βέλτιστο. Επιπρόσθετα η στοχαστικότητα βοηθάει τον αλγόριθμο να εξέλθει από τα τοπικά ελάχιστα δίνοντας έτσι περισσότερες πιθανότητες στον αλγόριθμο να καταλήξει σε ολικό ελάχιστο. Η περιγραφή του αλγορίθμου δίνεται από τις παρακάτω σχέσεις:

$$\mathbf{g} = \frac{1}{m} \nabla_{\boldsymbol{\theta}} \sum_i C(f(\mathbf{x}^{(i)}; \boldsymbol{\theta}), y^{(i)}) \quad (2.7)$$

$$\boldsymbol{\theta} = \boldsymbol{\theta} - e_k \times \mathbf{g} \quad (2.8)$$

Το τεχνητό νευρωνικό δίκτυο αναπαρίσταται από το  $f(\mathbf{x}^{(i)}; \boldsymbol{\theta})$ , όπου το  $x^{(i)}, y^{(i)}$  είναι το σετ των

δεδομένων εκπαίδευσης, ενώ η παράγωγος της συνάρτησης κόστους  $C$  υπολογίζεται ως προς τις παραμέτρους . Το  $m$  αντιπροσωπεύει το μέγεθος των δεδομένων που χρησιμοποιούνται ενώ το  $g$  δηλώνει την παράγωγο (gradient). Ο ρυθμός μάθησης (learning rate)  $e_k$  καθορίζει το μέγεθος των βημάτων που υλοποιείται από τον αλγόριθμο.

### Αλγόριθμος AdaGrad

Πρόκειται για έναν προσαρμοστικό αλγόριθμο που ρυθμίζει το ρυθμό μάθησης (learning rate), κάτι που δεν συμβαίνει στους άλλους αλγορίθμους. Ουσιαστικά ο αλγόριθμος *AdaGrad* προσθέτει μια εξάρτηση του ποσοστού μάθησης με τις παραγώγους. Αρχικά συλλέγει τα τετράγωνα των παραγώγων, που έχουν υπάρξει μέχρι εκείνη την στιγμή, σε ένα διάνυσμα  $s$  και έπειτα διαιρεί το ρυθμό μάθησης με την τετραγωνική ρίζα αυτού του αθροίσματος  $s$ . Πιο συγκεκριμένα έχουμε:

$$\mathbf{g} = \frac{1}{m} \nabla_{\boldsymbol{\theta}} \sum_i C(f(\mathbf{x}^{(i)}; \boldsymbol{\theta}), y^{(i)}) \quad (2.9)$$

$$\mathbf{s} = \mathbf{s} + \mathbf{g}^T \mathbf{g} \quad (2.10)$$

$$\boldsymbol{\theta} = \boldsymbol{\theta} - e_k \times \mathbf{g} / \sqrt{\mathbf{s} + \varepsilon} \quad (2.11)$$

Όπου  $\varepsilon$  είναι ένας πολύ μικρός αριθμός (συνήθως  $10^{-10}$ ) για να αποφευχθεί η διαίρεση με το μηδέν. Το αποτέλεσμα θα είναι οι παράμετροι που λαμβάνουν μεγάλες παραγώγους θα δουν το ρυθμό μάθησης να μικραίνει και το αντίθετο.

### Αλγόριθμος Adam

Ο αλγόριθμος *Adam* (*Adaptive moments*) [62] αποτελεί ένα συνδυασμό διαφόρων άλλων αλγορίθμων και λειτουργεί με τις ροπές πρώτης και δεύτερης τάξης. Η κύρια ιδέα πίσω από αυτόν τον αλγόριθμο είναι να μειωθεί η ταχύτητα πραγματοποίησης του αλγορίθμου για μια καλύτερη ανίχνευση του ολικού ελαχίστου. Ο αλγόριθμος Adam δημιουργεί έναν εκθετικά εξασθενημένο μέσο όρο των προηγούμενων παραγώγων και των τετραγώνων των παραγώγων. Αναλυτικά ο αλγόριθμος περιγράφεται από τις σχέσεις (2.12)-(2.17):

$$\mathbf{g} = \frac{1}{m'} \nabla_{\boldsymbol{\theta}} \sum_i C(f(\mathbf{x}^{(i)}; \boldsymbol{\theta}), y^{(i)}) \quad (2.12)$$

$$\mathbf{m} = \beta_1 \mathbf{m} + (1 - \beta_1) \mathbf{g} \quad (2.13)$$

$$\mathbf{s} = \beta_2 \mathbf{s} + (1 - \beta_2) \mathbf{g}^T \mathbf{g} \quad (2.14)$$

$$\hat{\mathbf{m}} = \frac{\mathbf{m}}{1 - \beta_1^t} \quad (2.15)$$

$$\hat{\mathbf{s}} = \frac{\mathbf{s}}{1 - \beta_2^t} \quad (2.16)$$

$$\boldsymbol{\theta} = \boldsymbol{\theta} - e_k \times \hat{\mathbf{m}} / \sqrt{\hat{\mathbf{s}} + \varepsilon} \quad (2.17)$$

Όπου το  $t$  αντιπροσωπεύει τον αριθμό της επανάληψης (ξεκινώντας από 1). Οι τιμές που προτείνονται για τις παραμέτρους είναι:

$$\beta_1 = 0.9, \beta_2 = 0.999 \text{ και } \varepsilon = 10^{-8}.$$

### Αλγόριθμος L-BFGS

Αποτελεί έναν αλγόριθμο βελτιστοποίησης που βασίζεται στη μέθοδο του Νεύτωνα (γνωστή και ως Newton-Raphson) για την εύρεση των ριζών μιας πραγματικής εξίσωσης. Είναι μια παραλλαγή του αλγορίθμου BFGS (Broyden-Fletcher-Goldfarb-Shanno) διότι χρησιμοποιεί περιορισμένη ποσότητα υπολογιστικής μνήμης (Limited-memory) [63], όπως υποδηλώνει το αρχικό γράμμα του ονόματος. Και οι δύο αυτοί αλγόριθμοι κάνουν χρήση του αντίστροφου Εσσιανού (Hessian) πίνακα εκτίμησης για τον έλεγχο των μεταβλητών στον χώρο. Ο Εσσιανός (Hessian) πίνακας έχει σαν στοιχεία του τις δεύτερες μερικές παραγώγους της υπό ανάλυση συνάρτησης. Σε αντίθεση με τον BFGS που αποθηκεύει έναν  $n \times n$  προσεγγιστικό Εσσιανό πίνακα (όπου  $n$  είναι ο αριθμός των παραμέτρων του προβλήματος), ο L-BFGS [64] αποθηκεύει μόνο μερικά διανύσματα που αντιπροσωπεύουν την προσέγγιση. Αυτή η διαφορά είναι που ελευθερώνει μνήμη από τον υπολογιστή. Ο αλγόριθμος L-BFGS είναι κατάλληλος για προβλήματα βελτιστοποίησης με μεγάλο αριθμό μεταβλητών.

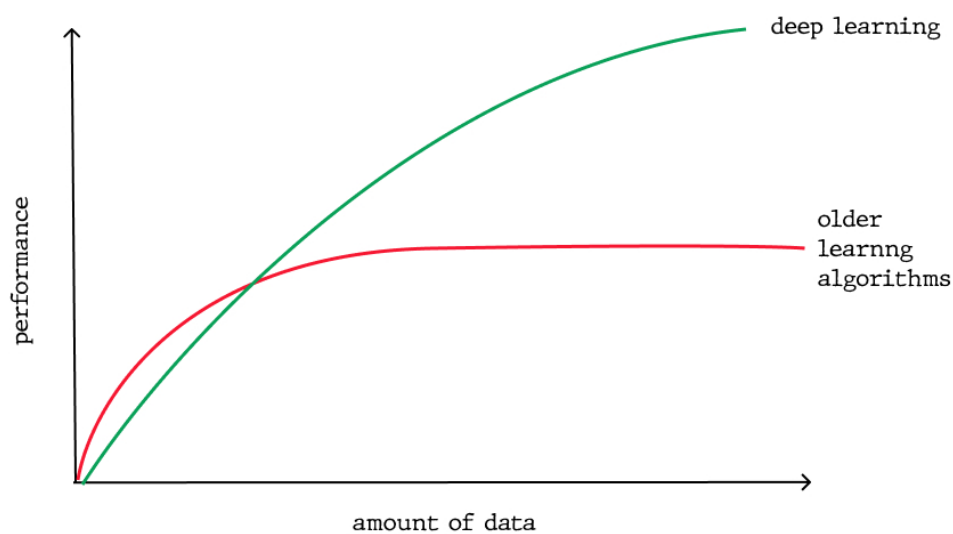
Περισσότερες πληροφορίες περί τεχνητών νευρωνικών δικτύων και μηχανικής μάθησης ο αναγνώστης παραπέμπεται στο βιβλίο του Aurélien Géron, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2019 [57].

## 2.4 Βαθιά Μάθηση (Deep Learning - DL)

Η *Βαθιά Μάθηση (BM)* ή *Deep Learning (DL)* αποτελεί ένα υποπεδίο της MM το οποίο δίνει έμφαση στην εκπαίδευση ΤΝΔ με πολλαπλά επίπεδα (συνήθως πάνω από δύο κρυφά επίπεδα), για να προβλέψει ή και να αποφασίσει ανάλογα με τη φύση του προβλήματος (εξού και ο όρος «βαθιά»). Καθώς αυξάνεται ο όγκος των δεδομένων, η πρόβλεψη σε κάποιο συγκεκριμένο πρόβλημα γίνεται ολοένα και πιο δύσκολη για αυτό χρειάζονται μοντέλα πιο περίπλοκα με μεγαλύτερες δυνατότητες.

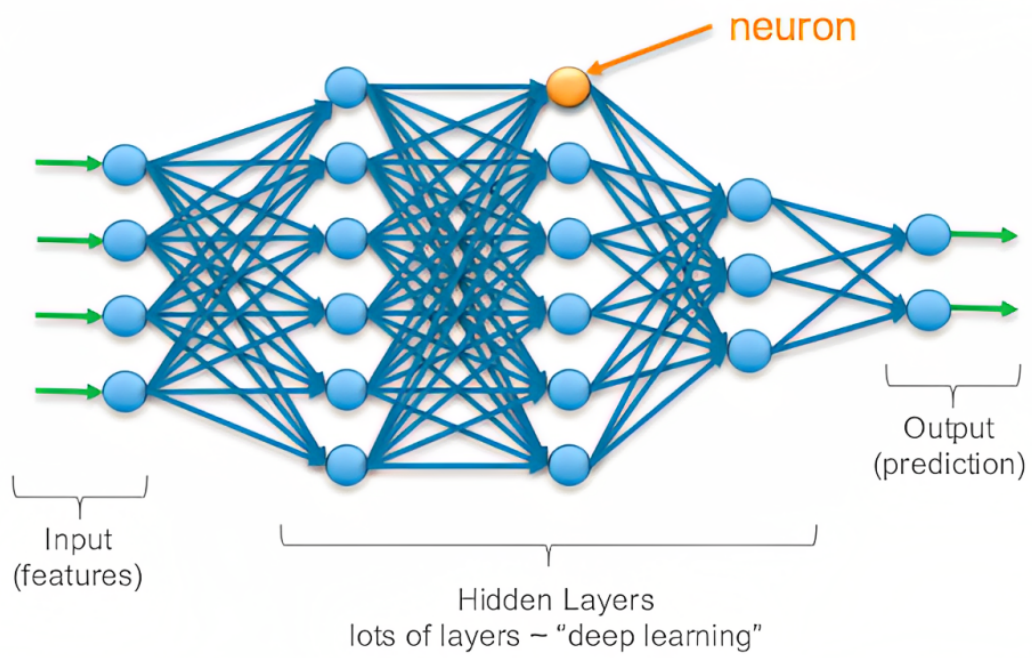
Η διαφορά μεταξύ της μηχανικής μάθησης και της βαθιάς μάθησης είναι η ικανότητα ανίχνευσης πολλαπλών μοτίβων. Το βασικό χαρακτηριστικό της BM είναι η ικανότητά της να μαθαίνει ιεραρχικές αναπαραστάσεις των δεδομένων εισόδου. Κάθε επίπεδο σε ένα βαθύ νευρωνικό δίκτυο μαθαίνει να εξάγει όλο και πιο σύνθετα χαρακτηριστικά από τα δεδομένα εισόδου. Τα αρχικά στρώματα μαθαίνουν βασικά χαρακτηριστικά χαμηλού επιπέδου, όπως ακμές (π.χ. σε ένα πρόβλημα αναγνώρισης εικόνων), ενώ τα επόμενα στρώματα μαθαίνουν πιο υψηλού επιπέ-

δου χαρακτηριστικά που είναι συνδυασμοί των χαρακτηριστικών χαμηλότερου επιπέδου. Αυτή η εκμάθηση ιεραρχικής αναπαράστασης επιτρέπει στα μοντέλα βαθιάς μάθησης να κατανοούν και να αναγνωρίζουν σύνθετα μοτίβα, κάνοντάς τα έτσι πολύ πιο αποδοτικά σε εργασίες όπως η αναγνώριση εικόνων (image recognition), η επεξεργασία φυσικής γλώσσας (natural language processing) και η αναγνώριση ομιλίας (speech recognition). Χαρακτηριστικό είναι το διάγραμμα του Σχήματος 2.12 όπου φαίνεται απόδοση των κλασικών μεθόδων μάθησης και της βαθιάς μάθησης σε σχέση με τον όγκο των διαθέσιμων δεδομένων. Είναι εμφανές πως η απόδοση βελτιώνεται σημαντικά όσο τα διαθέσιμα δεδομένα αυξάνονται, σε αντίθεση με τις κλασικές μεθόδους μάθησης όπου η απόδοση φτάνει σε ένα πλαφόν και όσο τα διαθέσιμα δεδομένα αυξάνονται δεν υπάρχει βελτίωση ως προς την απόδοση του μοντέλου.



**Σχήμα 2.12:** Η απόδοση της βαθιάς μάθησης και των κλασικών αλγορίθμων μάθησης σε σχέση με τα διαθέσιμα δεδομένα (Πηγή: [65]).

Συνήθως τα δίκτυα βαθιάς μάθησης αποτελούνται από ένα επίπεδο εισόδου, πολλαπλά κρυφά επίπεδα και ένα επίπεδο εξόδου. Ένα τέτοιο δίκτυο παρουσιάζεται στο Σχήμα 2.13. Η βαθιά μάθηση έχει φέρει επανάσταση σε διάφορους τομείς, όπως η όραση υπολογιστών, η επεξεργασία φυσικής γλώσσας, η αναγνώριση ομιλίας και τα συστήματα συστάσεων. Έχει επιτύχει αξιοσημείωτα αποτελέσματα σε εργασίες όπως η ταξινόμηση εικόνων, η ανίχνευση αντικειμένων, η μετάφραση, η ανάλυση συναισθήματος κ.ά. Οι συνεχείς εξελίξεις στις αρχιτεκτονικές βαθιάς μάθησης, στους αλγορίθμους βελτιστοποίησης και στην υπολογιστική ισχύ έχουν δημιουργήσει σημαντικές ανακαλύψεις στην έρευνα και τις εφαρμογές της ΤΝ.



**Σχήμα 2.13:** Παράδειγμα δικτύου βαθιάς μάθησης (Πηγή: [66]).

---

## Κεφάλαιο 3

# Επαναληπτικά Νευρωνικά Δίκτυα

### 3.1 Απλά Επαναληπτικά Νευρωνικά Δίκτυα - RNNs

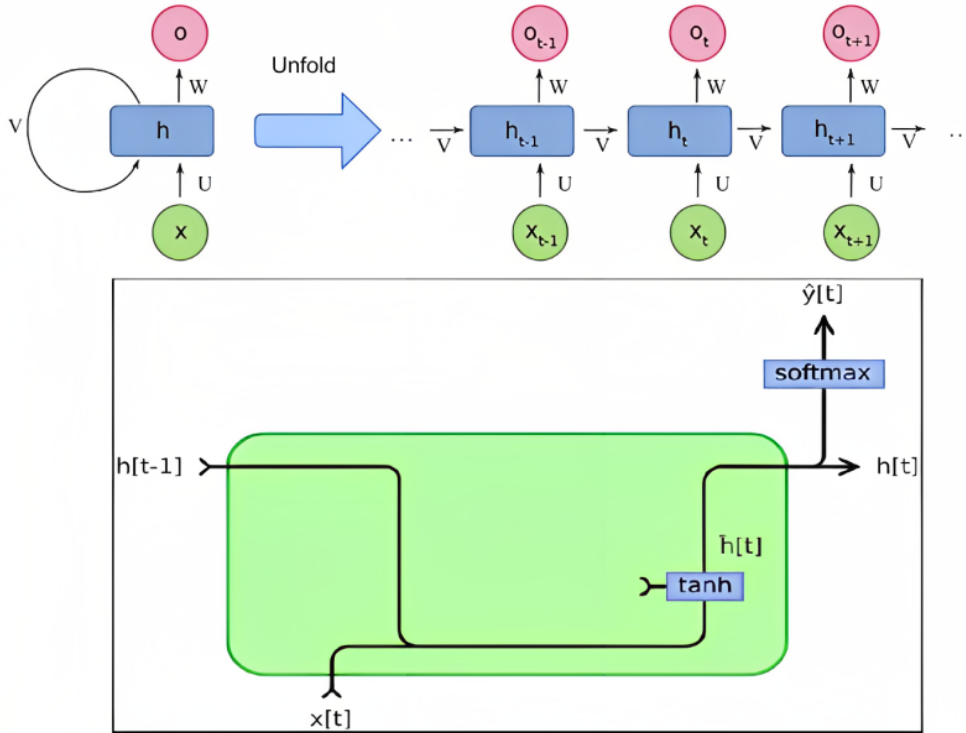
Τα απλά επαναληπτικά νευρωνικά δίκτυα (*Recurrent Neural Networks - RNNs*) αποτελούν μία ειδική κατηγορία τεχνητών νευρωνικών δικτύων που επιτρέπει στις πληροφορίες να ρέουν και στις δύο κατευθύνσεις, κάνοντας τα έτσι να είναι λειτουργικά για δεδομένα χρονοσειρών ή δεδομένα που περιλαμβάνουν ακολουθίες. Για παράδειγμα μια τέτοια ακολουθία θα μπορούσε να είναι η  $x(t) = x(1), \dots, x(\tau)$ , με το δείκτη  $t$  να παίρνει τιμές από 1 μέχρι  $\tau$ , όπως σε διάφορες εφαρμογές επεξεργασίας φωνής και γλώσσας, όπου για να προβλεφθεί η επόμενη λέξη μίας πρότασης είναι σημαντική η γνώση των προηγούμενων λέξεων. Η λέξη «επαναληπτικά» αντιπροσωπεύει την εκτέλεση της ίδιας εργασίας για κάθε στοιχείο μιας ακολουθίας, με την έξοδο να εξαρτάται από τους προηγούμενους υπολογισμούς. Στα μοντέλα RNN εισάγεται η έννοια της μνήμης και με αυτόν τον τρόπο αποτυπώνεται η πληροφορία για το τι έχει υπολογιστεί μέχρι εκείνη τη χρονική στιγμή.

Η αρχιτεκτονική των δικτύων RNN φαίνεται στο Σχήμα 3.1, όπου διακρίνεται το δίκτυο τόσο σε συμπαγή μορφή ενός συμβολισμού του RNN (αριστερό μέρος), όσο και ξεδιπλωμένο (δεξί μέρος). Το ξεδιπλωμένο δίκτυο προκύπτει από την ένωση των αντιγράφων του RNN για κάθε χρονικό βήμα  $t$ . Για παράδειγμα εάν η ακολουθία που μας ενδιαφέρει είναι μια πρόταση 3 λέξεων, το δίκτυο θα ξετυλιχτεί σε ένα νευρωνικό δίκτυο 3 επιπέδων, ένα επίπεδο (layer) για κάθε λέξη.

Οι παράμετροι του RNN είναι οι ακόλουθοι:

- είσοδος  $x(t)$ , λαμβάνεται ως είσοδος στο δίκτυο στο χρονικό βήμα  $t$ .
- κρυφή κατάσταση  $h(t)$ , αντιπροσωπεύει τη κρυφή κατάσταση τη χρονική στιγμή  $t$  και έχει το ρόλο της εισαγωγής μνήμης στο δίκτυο. Ο υπολογισμός του εξαρτάται από την





**Σχήμα 3.1:** Αρχιτεκτονική απλού επαναληπτικού νευρωνικού δικτύου RNN (Πηγή: Wikimedia [67, 68].)

είσοδο στο τρέχον χρονικό βήμα και από την τιμή της κρυφής κατάστασης στο προηγούμενο χρονικό βήμα, όπως φαίνεται στην εξίσωση (3.1).

$$h_t = f(Ux_t + Vh_{t-1}) \quad (3.1)$$

Η συνάρτηση  $f$  μπορεί να είναι οποιοσδήποτε μη γραμμικός μετασχηματισμός, όπως για παράδειγμα η συνάρτηση υπερβολικής εφαπτομένης  $\tanh$ .

- βάρη  $U, W, V$  συνδέουν τα διάφορα κομμάτια του δικτύου. Η είσοδος συνδέεται με τις κρυφές καταστάσεις μέσω του πίνακα βαρών  $U$ , η έξοδος συνδέεται με τις κρυφές καταστάσεις μέσω του πίνακα βαρών  $W$ , ενώ τέλος οι κρυφές καταστάσεις συνδέονται μεταξύ τους με τον πίνακα  $V$ .
- έξοδος  $o(t)$ , αποτελεί την έξοδο του δικτύου.

Η λειτουργία του δικτύου RNN ξεκινάει με το προς τα εμπρός πέρασμα (forward pass), το οποίο μπορεί να περιγραφεί από τις εξισώσεις (3.2) έως (3.5).

$$a_t = b + Vh_{t-1} + Ux_t \quad (3.2)$$

$$h_t = \tanh(a_t) \quad (3.3)$$

$$o_t = c + Wh_t \quad (3.4)$$

---

$$\hat{y}_t = \text{softmax}(o_t) \quad (3.5)$$

Για το παραπάνω σετ εξισώσεων έγινε η υπόθεση ότι η έξοδος είναι διακριτή ενώ ως συνάρτηση ενεργοποίησης χρησιμοποιήθηκε η υπερβολική εφαπτομένη. Τα  $b, c$  είναι κατώφλια (biases) ενώ για την τελική έξοδο  $\hat{y}_t$  εφαρμόζεται η λειτουργία softmax πάνω στην έξοδο  $o_t$ , για να ληφθεί ένα διάνυσμα κανονικοποιημένων πιθανοτήτων.

Στη συνέχεια για την εκπαίδευση των επαναληπτικών νευρωνικών δικτύων απαιτείται η υλοποίηση ενός αλγορίθμου οπισθοδιάδοσης backpropagation. Στα RNN χρησιμοποιείται ο αλγόριθμος Back Propagation Through Time (BPTT), ο οποίος αποτελεί μία ειδική κατηγορία του αλγορίθμου backpropagation. Πιο συγκεκριμένα, όπως περιγράφεται στο [69] και στο [70], το RNN ξετυλίγεται στο χρόνο, δημιουργώντας μια αλυσίδα συνδεδεμένων αντιγράφων του δικτύου, ένα για κάθε χρονικό βήμα. Το σφάλμα διαδίδεται ξανά μέσω αυτής της αλυσίδας αντιγράφων δικτύου, για αυτό και η ονομασία διάδοση προς τα πίσω μέσω του χρόνου. Κατά την διάρκεια της εκπαίδευσης ο αλγόριθμος BPTT ενημερώνει τα βάρη του RNN ελαχιστοποιώντας τη διαφορά μεταξύ των προβλεπόμενων και των πραγματικών εξόδων για μια δεδομένη ακολουθία εισόδων. Αυτό επιτυγχάνεται υπολογίζοντας την μερική παράγωγο ως προς τα βάρη του δικτύου της συνάρτησης απώλειας ή σφάλματος (loss function), η οποία στη συνέχεια χρησιμοποιείται για την ενημέρωση των βαρών μέσω του αλγορίθμου gradient descent. Ωστόσο ο υπολογισμός της παραγώγου για κάθε χρονική στιγμή μπορεί να είναι υπολογιστικά ακριβό και να οδηγήσει σε προβλήματα όπως αυτό της "εξαφάνισης των κλίσεων" (vanishing gradients) και αυτό της "εκρηκτικής αύξησης των κλίσεων" (exploding gradients). Ωστόσο ο αλγόριθμος BPTT παραμένει ένας αρκετά δημοφιλής αλγόριθμος για την εκπαίδευση των RNNs.

Τα απλά επαναληπτικά νευρωνικά δίκτυα διακρίνονται σε τέσσερις κύριους τύπους δικτύων:

- ένα προς ένα (one on one), τα οποία έχουν μία είσοδο και μία έξοδο.
- ένα προς πολλά (one to many), τα οποία έχουν μία είσοδο και πολλαπλές εξόδους.
- πολλά προς ένα (many to one), τα οποία έχουν πολλαπλές εισόδους και μία μόνο έξοδο.
- πολλά προς πολλά (many to many), τα οποία έχουν πολλές εισόδους και πολλές εξόδους.

Το πρόβλημα της εξαφάνισης των κλίσεων εμφανίζεται όταν οι παράγωγοι που χρησιμοποιούνται για τον υπολογισμό των βαρών αρχίζουν να εξαλείφονται, δηλαδή να παίρνουν τιμές πολύ κοντά στο μηδέν. Αποτέλεσμα αυτού είναι το μοντέλο να μην μαθαίνει, ενώ το ακριβώς αντίθετο οδηγεί στο πρόβλημα της εκρηκτικής αύξησης των κλίσεων. Τα δύο αυτά προβλήματα προσπαθούν να επιλυθούν με διάφορους μεθόδους από συγκεκριμένους τύπους RNN τα οποία μπορούν να χειριστούν μακροπρόθεσμες εξαρτήσεις (long-term dependencies), όπως για παράδειγμα τα λεγόμενα Long Short-Term Memory (LSTMs) ή τα Gated Recurrent Unit (GRUs). Στο υποκεφάλαιο 3.2 αναλύονται τα δίκτυα LSTMs.

## 3.2 Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης - LSTMs

Τα δίκτυα μακράς βραχυπρόθεσμης μνήμης (*Long Short-Term Memory - LSTMs*) αποτελούν μία τροποποιημένη έκδοση των RNN, στα οποία διευκολύνεται η μνήμη προηγούμενων δεδομένων για μεγάλες χρονικές περιόδους. Εισήχθησαν για πρώτη φορά από τους Hochreiter και Schmidhuber το 1997 [71]. Σε πολλές περιπτώσεις είναι πολύ σημαντική η γνώση των προηγούμενων χρονικών βημάτων για μεγάλο μήκος, ώστε να προκύψει σωστή πρόβλεψη του αποτελέσματος στο εκάστοτε πρόβλημα. Για παράδειγμα σε μία μακροσκελή κριτική μίας ταινίας ή μίας παράστασης, οι λέξεις στην αρχή της κριτικής μπορεί να φανούν χρήσιμες για την τελική κατάταξη της κριτικής σε "καλή" ή "κακή".

Το LSTM είναι κατάλληλο για ταξινόμηση, επεξεργασία και πρόβλεψη χρονοσειρών. Σε ένα δίκτυο LSTM υπάρχουν τρεις πύλες (gates) όπως φαίνεται στο Σχήμα 3.3, ενώ για την εκπαίδευση του μοντέλου χρησιμοποιείται ο αλγόριθμος back-propagation. Οι τρεις αυτές πύλες είναι οι εξής:

1. **Πύλη εισόδου (input gate):** Η πύλη εισόδου αποφασίζει πόσο πολύτιμη είναι η τρέχουσα είσοδος για την επίλυση της εργασίας, επιλέγοντας έτσι ποια τιμή από την είσοδο πρέπει να χρησιμοποιηθεί για την τροποποίηση της μνήμης. Η σιγμοειδής συνάρτηση (sigmoid function) αποφασίζει για το ποιες τιμές θα μετατραπούν σε 0 και σε 1, ενώ η συνάρτηση υπερβολικής εφαπτομένης καθορίζει το επίπεδο σπουδαιότητας των τιμών δίνοντας μία τιμή βαρύτητας από -1 έως 1. Η έξοδος της πύλης εισόδου μπορεί να περιγραφεί από τις εξισώσεις (3.6) και (3.7).

$$i_t = (W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.6)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3.7)$$

2. **Πύλη λήθης (forget gate):** Η πύλη λήθης ελέγχει ποιες πληροφορίες του κελιού πρέπει να "ξεχαστούν", με τις δεδομένες νέες πληροφορίες που εισήχθησαν στο δίκτυο. Αυτό καθορίζεται από τη σιγμοειδή συνάρτηση. Πιο συγκεκριμένα δέχεται την προηγούμενη κατάσταση ( $h_{t-1}$ ) καθώς και την είσοδο ( $x_t$ ) και εξάγει έναν αριθμό μεταξύ 0 (που σημαίνει πως αυτό θα παραλειφθεί) και 1 (που σημαίνει πως αυτό θα αξιοποιηθεί), για κάθε αριθμό του κελιού  $C_{t-1}$ . Η έξοδος της πύλης λήθης δίνεται από τη σχέση (3.8):

$$f_t = (W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3.8)$$

3. **Πύλη εξόδου (output gate):** Η πύλη εξόδου ελέγχει ποιες πληροφορίες που κωδικοποιούνται στην κατάσταση κελιού, αποστέλλονται στο δίκτυο ως είσοδος στο επόμενο χρονικό βήμα. Αυτό επιτυγχάνεται μέσω του διανύσματος εξόδου  $h(t)$ . Επομένως η είσοδος και η μνήμη χρησιμοποιούνται για να αποφασιστεί η έξοδος του κελιού. Η έξοδος της πύλης εξόδου περιγράφεται από την εξίσωση (3.9) ενώ η έξοδος του κελιού δίνεται

από την εξίσωση (3.10).

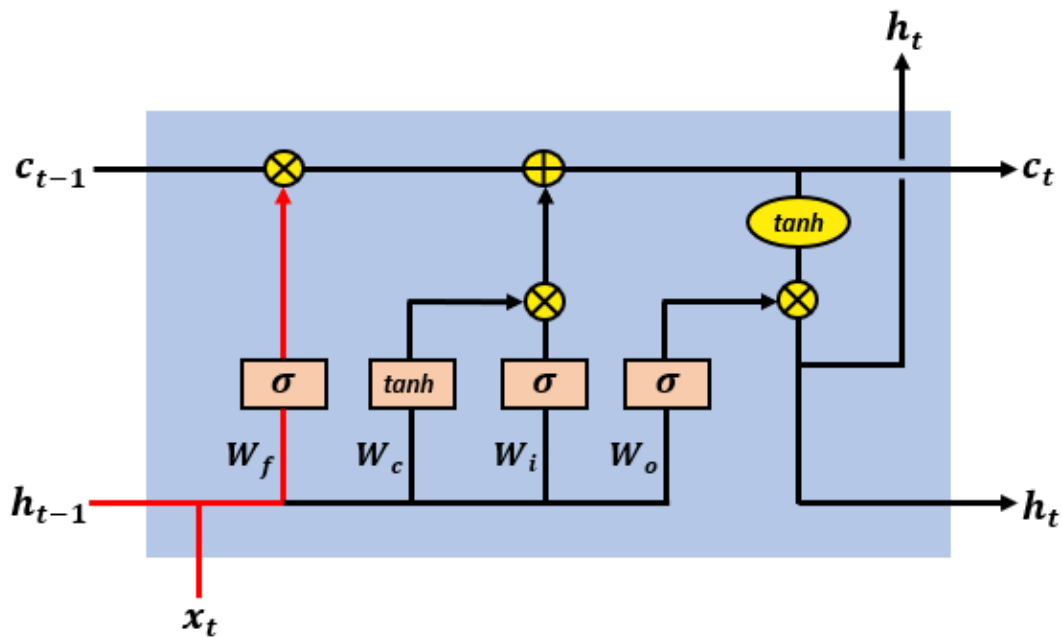
$$o_t = (W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3.9)$$

$$h_t = o_t \otimes \tanh(c_t) \quad (3.10)$$

Οι μακροπρόθεσμες εξαρτήσεις κωδικοποιούνται στα διανύσματα των κελιών κατάστασης (ή αλλιώς cells states) και με αυτόν τον τρόπο μπορεί να αποφευχθεί το πρόβλημα της εξαφάνισης των κλίσεων. Η κατάσταση του κελιού LSTM έχει τη μορφή:

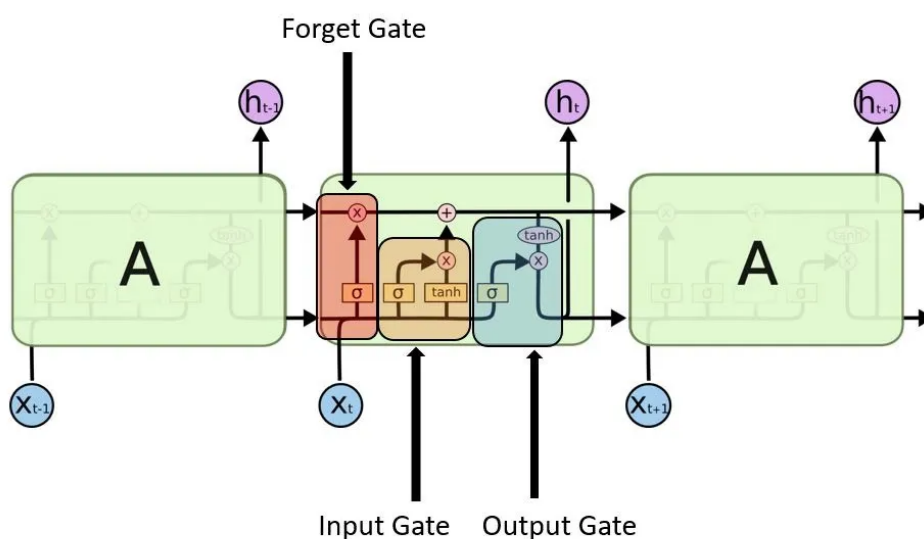
$$c_t = c_{t-1} \otimes f_t \oplus \tilde{c}_t \otimes i_t \quad (3.11)$$

Συνοπτικά μπορεί να αναφερθεί πως τα δίκτυα LSTM έχουν κάποια εσωτερικά κελιά κατάστασης που λειτουργούν ως κελιά μακροπρόθεσμης ή βραχυπρόθεσμης μνήμης. Η έξοδος του δικτύου LSTM διαμορφώνεται από την κατάσταση αυτών των κελιών. Κάτι τέτοιο είναι πολύ σημαντικό όταν στην πρόβλεψη του νευρωνικού δικτύου χρειάζεται η συνεισφορά των ιστορικών τιμών εισόδου και όχι μόνο της τελευταίας εισόδου [72].



Σχήμα 3.2: Αρχιτεκτονική ενός κελιού του νευρωνικού δικτύου LSTM (Πηγή: [73]).

Στη γενική περίπτωση όμως το πρόβλημα που αντιμετωπίζουν τα RNNs συνεχίζεται να εμφανίζεται και στα LSTMs, όπως για παράδειγμα σε προτάσεις μεγάλου μήκους φαίνεται να μην συμπεριφέρονται καλά. Ο λόγος για τον οποίο συμβαίνει κάτι τέτοιο είναι διότι η πιθανότητα να διατηρηθεί η επιρροή μίας λέξης που βρίσκεται πολύ μακριά από την τρέχουσα λέξη είναι εκθετικά μειούμενη με την απόσταση αυτής της λέξης. Αυτό σημαίνει πως όταν οι προτάσεις είναι αρκετά μεγάλες το μοντέλο συχνά ξεχνά το περιεχόμενο των απομακρυσμένων θέσεων στην ακολουθία. Ακόμα ένα πρόβλημα που αντιμετωπίζουν αυτά τα δίκτυα είναι η παραλληλοποίηση της εργασίας για την επεξεργασία προτάσεων είναι δύσκολη, αφού πρέπει οι λέξεις να



**Σχήμα 3.3:** Αναπαράσταση του νευρωνικού δικτύου LSTM, για τις χρονικές στιγμές  $t - 1, t, t + 1$ , με τις αντίστοιχες πύλες (Πηγή: [74]).

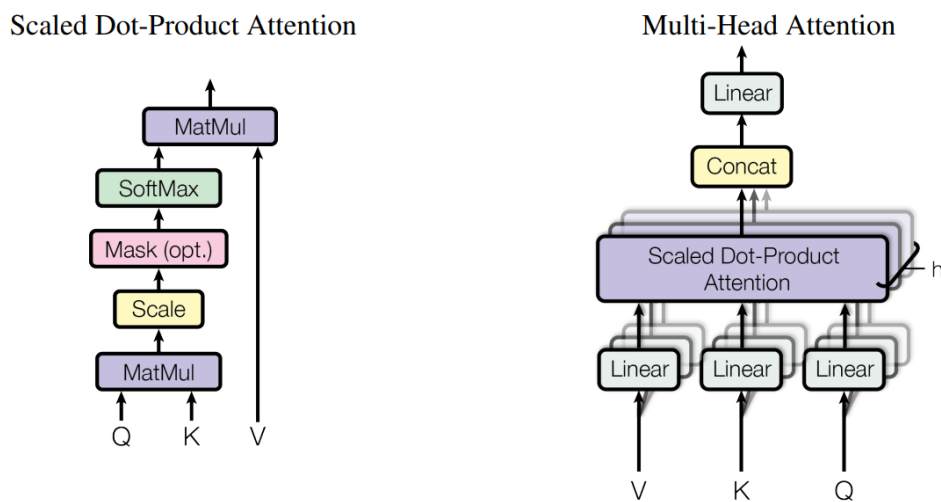
επεξεργάζονται μία προς μία. Συνοψίζοντας λοιπόν τα δίκτυα LSTMs και RNNs έχουν κυρίως τα εξής μειονεκτήματα: (α) η παραλληλοποίηση αποτρέπεται από τον διαδοχικό υπολογισμό, (β) η περιορισμένη μνήμη και (γ) δεν υπάρχει ρητή μοντελοποίηση εξαρτήσεων μεγάλου και μικρού μήκους. Τα προαναφερθέντα μειονεκτήματα έρχονται να επιλύσουν τα δίκτυα μετασχηματιστή, τα οποία εξετάζονται στη συνέχεια.

### 3.3 Δίκτυα Μετασχηματιστή - Transformers

#### 3.3.1 Μηχανισμός Προσοχής - Attention mechanism

Το Δίκτυο Μετασχηματιστή (*Transformer*) αποτελεί ένα τεχνητό νευρωνικό δίκτυο το οποίο προτάθηκε για πρώτη φορά στην επιστημονική δημοσίευση "Attention is all you need" του 2017 [22] για να λύσει ορισμένα από τα ζητήματα των RNN και LSTM. Για την επίλυση μερικών από τα προβλήματα που αναφέρθηκαν στα υποκεφάλαια 3.1 και 3.2 προτάθηκε η δημιουργία μίας τεχνικής για να δίνεται προσοχή (attention) σε συγκεκριμένα κομμάτια της ακολουθίας, π.χ. συγκεκριμένες λέξεις μιας πρότασης. Ο μηχανισμός προσοχής (*attention mechanism*) επιτρέπει στο μοντέλο Transformer να έχει επαρκή μακροπρόθεσμη μνήμη. Το μοντέλο μπορεί να εστιάζει σε όλα τα προηγούμενα στοιχεία που έχουν δημιουργηθεί. Η προσοχή είναι πολύ σημαντική έννοια σε δεδομένα ακολουθίας καθώς αντιπροσωπεύει και ένα κομμάτι του τρόπου με τον οποίο ο άνθρωπος λειτουργεί. Για παράδειγμα στην μετάφραση δίνεται ιδιαίτερη προσοχή στη λέξη που μεταφράζεται εκείνη τη στιγμή, όταν γράφεται μια ηχογράφηση δίνεται προσοχή σε σχετικά κομμάτια του ηχητικού ή ομοίως όταν περιγράφεται μια εικόνα.

Η κύρια ιδέα του μηχανισμού προσοχής βασίζεται στον υπολογισμό ενός *βάρους προσοχής* (*attention weight*) για κάθε στοιχείο στην ακολουθία εισόδου, το οποίο αντανακλά τη σημασία ή τη συνάφειά του με την τρέχουσα έξοδο. Λειτουργεί υπολογίζοντας πρώτα ένα σύνολο διανυσμάτων ερωτήματος (query), κλειδιού (key) και τιμής (value) για κάθε στοιχείο της ακολουθίας εισόδου. Το διάνυσμα ερωτήματος αντιπροσωπεύει την τρέχουσα κατάσταση εξόδου του δικτύου, ενώ τα διανύσματα κλειδιού και τιμής αντιπροσωπεύουν το στοιχείο της ακολουθίας εισόδου και το σχετικό διάνυσμα των χαρακτηριστικών του, αντίστοιχα. Στη συνέχεια τα διανύσματα αυτά χρησιμοποιούνται για τον υπολογισμό ενός συνόλου βαρών προσοχής μέσω μίας σχέσης εσωτερικού γινομένου, η οποία μετρά την ομοιότητα μεταξύ των διανυσμάτων. Στο Σχήμα 3.4 μπορεί να συνοψιστεί η αρχιτεκτονική του μηχανισμού προσοχής.



**Σχήμα 3.4:** Αρχιτεκτονική του μηχανισμού προσοχής (Πηγή: ‘Attention Is All You Need’ από τους Vaswani et al. [22]).

Η εξίσωση που περιγράφει την σχέση του εσωτερικού γινομένου δίνεται από την (3.12).

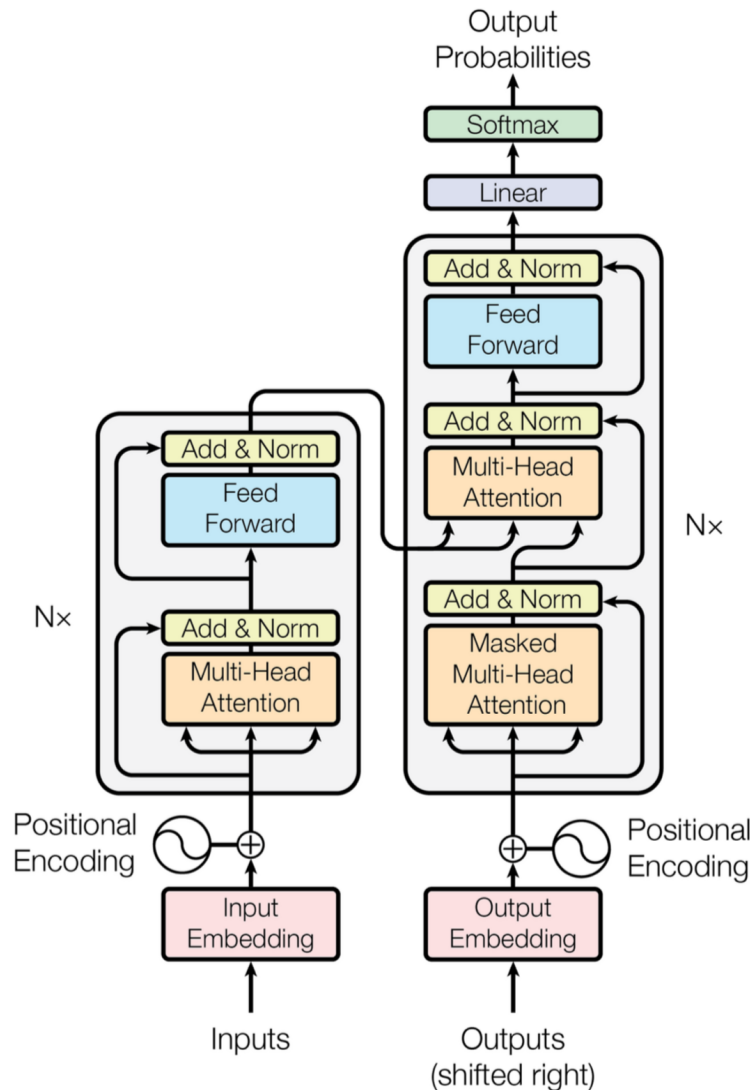
$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3.12)$$

, όπου  $Q$  (Query),  $K$  (Key) και  $V$  (Value) αποτελούν τις εισόδους του μηχανισμού προσοχής. Το εσωτερικό γινόμενο των ερωτημάτων  $Q$  και κλειδιών  $K$ , διαιρείται με έναν συντελεστή κλίμακας  $\frac{1}{\sqrt{d_k}}$ , όπου  $d_k$  η διάσταση των κλειδιών και ερωτημάτων.

Τα βάρη προσοχής χρησιμοποιούνται για τον υπολογισμό ενός σταθμισμένου αθροίσματος των διανυσμάτων τιμών, το οποίο συνοψίζει τις σχετικές πληροφορίες στην ακολουθία εισόδου. Το διάνυσμα αυτό στη συνέχεια ενώνεται με την τρέχουσα κατάσταση εξόδου και περνά μέσα από ένα προς τα εμπρός τροφοδοτούμενο νευρωνικό δίκτυο (feedforward neural network) για να δημιουργηθεί η επόμενη κατάσταση εξόδου.

### 3.3.2 Αρχιτεκτονική Transformer

Όπως προαναφέρθηκε οι Vaswani et al. [22] παρουσίασαν ένα νέο νευρωνικό δίκτυο που ονόμασαν Transformer, του οποίου η αρχιτεκτονική είναι τύπου κωδικοποιητή-αποκωδικοποιητή (encoder-decoder) και βασίζεται στο μηχανισμό προσοχής (attention mechanism). Εφαρμογή του νευρωνικού δικτύου της συγκεκριμένης δημοσίευσης αποτελεί η εργασία της μετάφρασης.



**Σχήμα 3.5:** Επισκόπηση της αρχιτεκτονικής του μοντέλου μετασχηματιστή - Transformer (Πηγή: ‘Attention Is All You Need’ από τους Vaswani et al. [22]).

Μία βασική διαφορά είναι πως η ακολουθία εισόδου μπορεί να παραλληλοποιηθεί έτσι ώστε η GPU να μπορεί να χρησιμοποιηθεί αποτελεσματικά και να αυξηθεί η ταχύτητα της εκπαίδευσης σημαντικά. Επιπλέον χρησιμοποιώντας το επίπεδο πολλαπλών κεφαλών προσοχής (*multi-headed attention layer*) συμβάλει στην λύση του προβλήματος της εξαφάνισης των κλίσεων. Επομένως τα βασικά μειονεκτήματα που επισημάνθηκαν για τα προηγούμενα δύο νευρωνικά δίκτυα (RNN και LSTM) επιλύονται σε κάποιο βαθμό με την αρχιτεκτονική των μετασχηματιστών.

---

## Encoder Block

Στην εργασία της μετάφρασης, για τη λειτουργία του δικτύου, κάθε λέξη που πρέπει να μεταφραστεί θα πρέπει να μετατραπεί σε διάνυσμα αριθμών. Σε αυτό βοηθάει η έννοια της *ενσωμάτωσης (Embedding)*. Θα μπορούσε να παρομοιαστεί με ένα λεξικό όπου οι λέξεις με παρόμοια σημασία ομαδοποιούνται μαζί στο χώρο ενσωμάτωσης (embedding space) και αποδίδονται συγκεκριμένες τιμές σε αυτές. Οι λέξεις, με αυτόν τον τρόπο, μετατρέπονται σε διανύσματα. Η κάθε λέξη μπορεί να επιδέχεται διαφορετικής ερμηνείας ανάλογα με τη θέση της στην πρόταση και για τον λόγο αυτόν χρησιμοποιείται ο *κωδικοποιητής θέσης (Positional Encoding)*, ο οποίος δίνει νόημα στη λέξη σύμφωνα με τη θέση της σε μια πρόταση, μέσω διανυσμάτων. Η επεξεργασμένη αυτή είσοδος εισέρχεται στο μπλοκ του κωδικοποιητή.

## Multi-head attention

Μία από τις κυριότερες λειτουργίες των μετασχηματιστών είναι η *αυτοπροσοχή (self attention)*. Εστιάζει στο πόσο σχετική είναι μία συγκεκριμένη λέξη της πρότασης σε σχέση με τις υπόλοιπες, ενώ έχει τη μορφή ενός διανύσματος (attention vector). Το διάνυσμα αυτό δίνει, όπως είναι φυσικό, μεγάλο βάρος στην ίδια τη λέξη. Για την αντιμετώπιση του προβλήματος αυτού προσδιορίζονται πολλαπλά διανύσματα προσοχής ανά λέξη και υπολογίζεται ένας σταθμισμένος μέσος όρος, για να υπολογιστεί το τελικό διάνυσμα προσοχής κάθε λέξης. Η διαδικασία αυτή ονομάζεται *multi-head self attention*. Στη συνέχεια τα διανύσματα προσοχής προωθούνται σε ένα απλό *προς-τα-εμπρός τροφοδοτούμενο νευρωνικό δίκτυο (feedforward neural network)*, με κύριο σκοπό να τα μετατρέψει σε αποδεκτή μορφή για τα επόμενα επίπεδα είτε αυτό είναι κωδικοποιητής (encoder) είτε αποκωδικοποιητής (decoder). Το νευρωνικό δίκτυο δέχεται κάθε φορά ένα διάνυσμα προσοχής και τα οποία είναι ανεξάρτητα μεταξύ τους. Σε αντίθεση με τα RNN, εδώ μπορεί να εφαρμοστεί η παραλληλοποίηση, ξεπερνώντας ένα σημαντικό πρόβλημα των επαναληπτικών δικτύων.

## Decoder Block

Όπως περιγράφεται και στην αρχική δημοσίευση [22], στη περίπτωση της μετάφρασης συνήθως χρειάζεται η μετάφραση ενός κειμένου από μία γλώσσα (π.χ Αγγλικά) σε μία άλλη (π.χ. Ελληνικά). Στη διαδικασία της εκπαίδευσης του μοντέλου είναι απαραίτητο να δοθούν στο δίκτυο η αγγλική πρόταση μαζί με την αντίστοιχη μετάφρασή της στα ελληνικά. Επομένως ο *κωδικοποιητής (encoder)* δέχεται την αγγλική πρόταση ενώ ο *αποκωδικοποιητής (decoder)* δέχεται την ελληνική πρόταση. Όπως και στον κωδικοποιητή έτσι και στον αποκωδικοποιητή υπάρχει το επίπεδο ενσωμάτωσης (embedding layer) και στη συνέχεια ο κωδικοποιητής θέσης (positional encoder), που έχει στόχο να μετατρέψει τις λέξεις σε διανύσματα.

Στη συνέχεια υπάρχει ένα κομμάτι για να προσφέρει τον μηχανισμό προσοχής, όπως και στον κωδικοποιητή, με την ονομασία *masked multi-head attention*. Προσπαθεί να αναγνωρίσει τις σχέσεις μεταξύ των διαφορετικών μελών σε μια ακολουθία. Ο όρος μάσκα ή κάλυμμα (masked) στην ονομασία αναφέρεται στο γεγονός ότι κατά τη διάρκεια της διαδικασίας αποκωδικοποίησης, ο μηχανισμός προσοχής είναι καλυμμένος για να αποτρέψει τον αποκωδικοποιητή από το



---

να παρακολουθεί μελλοντικές τιμές της ακολουθίας εξόδου. Αυτό επιτυγχάνεται με τον μηδενισμό των βαρών προσοχής για τα μελλοντικά κομμάτια κατά τη διάρκεια του υπολογισμού της συνάρτησης softmax. Ο όρος multi-head αναφέρεται στο γεγονός ότι ο μηχανισμός προσοχής εφαρμόζεται πολλές φορές παράλληλα, επιτρέποντας στο μοντέλο την καταγραφή διαφορετικών τύπων σχέσεων και εξαρτήσεων μεταξύ των εισόδων. Στο παράδειγμα της μετάφρασης, αφού δοθεί μία αγγλική λέξη θα μεταφραστεί στα ελληνικά και θα συγκριθεί με την πραγματική ελληνική λέξη (η οποία τροφοδοτήθηκε στον αποκωδικοποιητή). Μετά τη σύγκριση των δύο, θα ενημερωθεί η τιμή του πίνακα και με αυτόν τον τρόπο έπειτα από αρκετές επαναλήψεις θα επιτευχθεί η μάθηση του μοντέλου. Το μοντέλο δεν πρέπει να έχει πρόσβαση στην λέξη που καλείται να μεταφράσει διότι χάνεται έτσι το νόημα της μάθησης. Επομένως ο ρόλος του masked multi-head attention είναι η απόκρυψη της πληροφορίας της πραγματικής λέξης.

Τα διανύσματα προσοχής που προκύπτουν από το προηγούμενο στάδιο μαζί με αυτά που προκύπτουν από τον κωδικοποιητή εισέρχονται σε ακόμη επίπεδο προσοχής πολλαπλών κεφαλών (multi-head attention block), το οποίο έχει την ονομασία Encoder-Decoder Attention Block. Σε αυτό το στάδιο επιτυγχάνεται αντιστοίχιση των αγγλικών με τις ελληνικές λέξεις και εντοπίζονται οι σχέσεις μεταξύ τους. Από την έξοδο παράγονται διανύσματα προσοχής για κάθε λέξη των αγγλικών και των ελληνικών στις αντίστοιχες προτάσεις. Κάθε διάνυσμα αντιπροσωπεύει τη σχέση των λέξεων με τις υπόλοιπες και στις δύο γλώσσες. Τα διανύσματα προσοχής προωθούνται σε ένα feed forward δίκτυο διαμορφώνοντας τα διανύσματα εξόδου σε κάτι εύκολα αποδεκτό από ένα άλλο επίπεδο αποκωδικοποιητή ή από ένα γραμμικό επίπεδο. Το γραμμικό επίπεδο αποτελεί ένα feed forward επίπεδο το οποίο στο τέλος τροφοδοτεί σε ένα Softmax επίπεδο, μετατρέποντας την είσοδο σε μια κατανομή πιθανοτήτων, η οποία είναι ερμηνεύσιμη από τον άνθρωπο. Το τελικό αποτέλεσμα είναι η λέξη με τη μεγαλύτερη πιθανότητα μετά την μετάφραση.

### **Transformers in time series forecasting**

Σε αντιστοιχία με τον τρόπο επίλυσης του προβλήματος της μετάφρασης, τα δίκτυα των μετασχηματιστών (transformers) επιλύουν και το πρόβλημα της πρόβλεψης χρονοσειρών. Για την ακρίβεια στην διαδικασία της κωδικοποίησης τα δεδομένα της χρονοσειράς εισόδου περνούν από τον κωδικοποιητή του μοντέλου Transformer, ο οποίος αποτελείται από πολλαπλά στρώματα, καθένα από τα οποία περιέχει έναν μηχανισμό αυτοπροσοχής (self-attention) και ένα νευρωνικό δίκτυο τροφοδότησης προς τα εμπρός. Ο μηχανισμός αυτοπροσοχής επιτρέπει στο μοντέλο να καταγράφει εξαρτήσεις μεταξύ διαφορετικών χρονικών βημάτων στην ακολουθία εισόδου. Μετά την ολοκλήρωση της κωδικοποίησης, ο αποκωδικοποιητής χρησιμοποιείται για να παράγει την ακολουθία εξόδου. Η διαδικασία της αποκωδικοποίησης λειτουργεί βήμα-προς-βήμα (step-by-step), παράγοντας μία έξοδο κάθε φορά. Στον αποκωδικοποιητή είναι επίσης ενσωματωμένος ένας μηχανισμός αυτοπροσοχής για να παρακολουθείται το περιεχόμενο των παραγόμενων προηγούμενων βημάτων.

Κατά τη διαδικασία της εκπαίδευσης, το μοντέλο βελτιστοποιείται ώστε να ελαχιστοποιηθεί η

---

διαφορά μεταξύ της προβλεπόμενης και της πραγματικής ακολουθίας εξόδου. Μόλις το μοντέλο εκπαιδευτεί, μπορεί να χρησιμοποιηθεί για να κάνει προβλέψεις σε νέες ακολουθίες εισόδου, δημιουργώντας αναδρομικά κάθε βήμα της ακολουθίας εξόδου με βάση τα προηγούμενα βήματα.

### 3.4 Μοντέλο Πληροφόρησης - Informer model

Το μοντέλο Πληροφόρησης - *Informer* το οποίο προτάθηκε από τους Zhou et al. το 2021 [40], αποτελεί ένα βαθύ τεχνητό νευρωνικό δίκτυο που στόχο έχει την πρόβλεψη χρονοσειρών μεγάλης ακολουθίας (long sequence time-series). Εστιάζει στο πρόβλημα πρόβλεψης χρονοσειρών για μεγάλη ακολουθία μέλλοντος (*Long Sequence Time Series Forecasting - LSTF*). Το συγκεκριμένο πρόβλημα απαιτεί υψηλή ικανότητα πρόβλεψης για την καταγραφή μακροπρόθεσμης εξάρτησης (*long-range dependence*) μεταξύ εισόδων και εξόδων. Η άνοδος της εφαρμογής των δικτύων Transformer έχουν τονίσει την δυνατότητά τους στην αύξηση της ικανότητας πρόβλεψης. Παρόλα αυτά μερικά ζητήματα που αφορούν τη λειτουργία των μετασχηματιστών, όπως η τετραγωνική χρονική πολυπλοκότητα ( $\mathcal{O}(n^2)$ ), η υψηλή χρήση της μνήμης και ο περιορισμός της αρχιτεκτονικής του κωδικοποιητή-αποκωδικοποιητή συνθέτουν τη μη παραγωγικότητα των μετασχηματιστών στο πρόβλημα αυτό.

Για την αντιμετώπιση τέτοιων ζητημάτων σχεδιάστηκε ο *Informer*, ένα αποδοτικό μοντέλο βασισμένο στο δίκτυο μετασχηματιστή (transformer-based model) με τρία βασικά χαρακτηριστικά:

1. **ProbSparse self-attention mechanism**: ένα είδος μηχανισμού προσοχής το οποίο πετυχαίνει  $\mathcal{O}(L \log L)$  στη χρονική πολυπλοκότητα και τη χρήση μνήμης, όπου  $L$  το μήκος της ακολουθίας εισόδου.
2. **Self-attention distilling**: μια τεχνική που αναδεικνύει την κυρίαρχη προσοχή κάνοντας αποτελεσματικό τον χειρισμό εξαιρετικά μεγάλων ακολουθιών εισόδου.
3. **Generative style decoder**: ένα είδος αποκωδικοποιητή το οποίο προβλέπει μακροπρόθεσμες ακολουθίες χρονοσειρών με ένα προς τα εμπρός πέρασμα (one forward pass) και όχι βήμα προς βήμα (step-by-step), γεγονός που βελτιώνει δραστικά την ταχύτητα παραγωγής της πρόβλεψης.

Το μοντέλο αυτό έχει δείξει αρκετά ικανοποιητικά και ενθαρρυντικά αποτελέσματα σε διάφορα σύνολα δεδομένων αναφοράς.

#### 3.4.1 Αρχιτεκτονική Informer

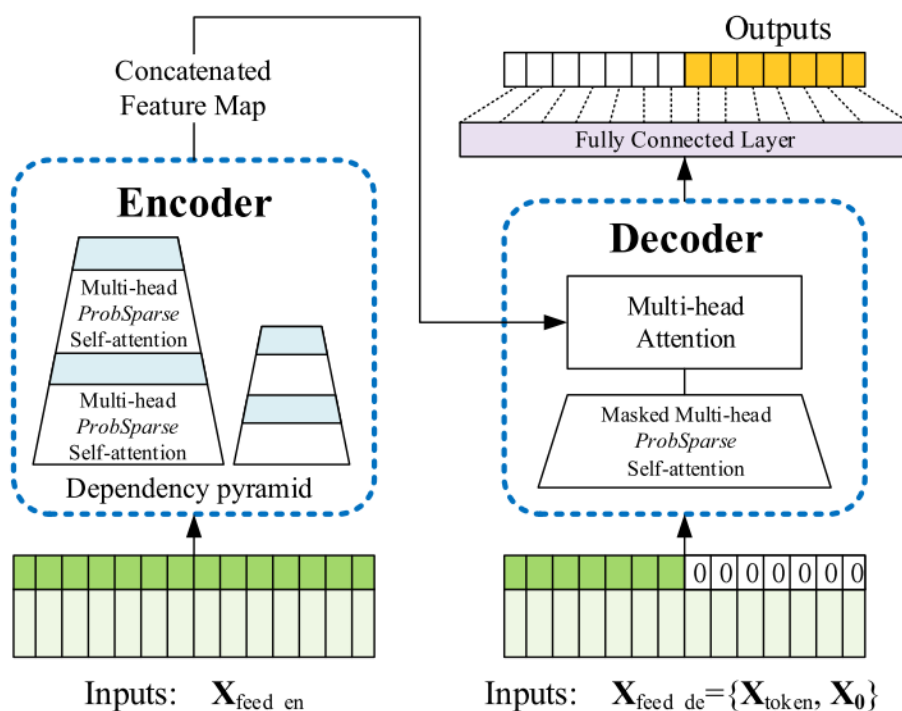
Το μοντέλο *Informer* διατηρεί την κλασική αρχιτεκτονική των μετασχηματιστών, δηλαδή αυτή του κωδικοποιητή-αποκωδικοποιητή (encoder-decoder architecture) ενώ παράλληλα στοχεύει

στην επίλυση του προβλήματος πρόβλεψης χρονοσειρών μεγάλων ακολουθιών. Η αρχιτεκτονική του μοντέλου φαίνεται στο Σχήμα 3.6. Απαρτίζεται από τρία βασικά συστατικά: το επίπεδο του κωδικοποιητή (*encoder layer*), το επίπεδο του αποκωδικοποιητή (*decoder layer*) και το επίπεδο όπου γίνεται η πρόβλεψη (*prediction layer*).

Ο κωδικοποιητής λαμβάνει την χρονοσειρά εισόδου και εφαρμόζει μια σειρά από μπλοκ κωδικοποίησης για την εξαγωγή των χαρακτηριστικών του. Κάθε μπλοκ κωδικοποίησης αποτελείται από τον μηχανισμό αυτοπροσοχής *ProbSparse* (*Probabilistic Sparsity self-attention mechanism*), ένα προς τα εμπρός τροφοδοτούμενο νευρωνικό δίκτυο και ένα επίπεδο κανονικοποίησης. Ο μηχανισμός αυτοπροσοχής *ProbSparse* μειώνει το υπολογιστικό κόστος του μηχανισμού αυτοπροσοχής, διατηρώντας παράλληλα την αποτελεσματικότητά του. Στο συγκεκριμένο μηχανισμό αντί για την εξίσωση (3.12) χρησιμοποιείται η παρακάτω εξίσωση:

$$Attention(Q, K, V) = softmax\left(\frac{\bar{Q}K^T}{\sqrt{d}}\right)V \quad (3.13)$$

Η κύρια διαφορά μεταξύ των σχέσεων 3.12 και 3.13 είναι πως στη δεύτερη, τα κλειδιά  $K$  δεν συνδέονται με όλα τα δυνατά ερωτήματα  $Q$  αλλά με τα  $u$  πιο κυρίαρχα ερωτήματα  $\bar{Q}$ .



**Σχήμα 3.6:** Αρχιτεκτονική μοντέλου Informer (Πηγή: 'Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting' από τους Zhou et al. [40]).

Οι συγγραφείς του [40] προτείνουν μια εμπειρική προσέγγιση για την αποτελεσματική απόκτηση της μέτρησης της αραιότητας του ερωτήματος (*query sparsity measurement*), όπως περιγράφεται

φεται στο Λήμμα 1.

**Λήμμα 1.** Για κάθε ερώτημα (query)  $\mathbf{q}_i \in R^d$  και  $\mathbf{k}_j \in R^d$  στο σύνολο των κλειδιών  $\mathbf{K}$  υπάρχει το όριο  $\ln L_K \leq M(\mathbf{q}_i, \mathbf{K}) \leq \max_j \{\mathbf{q}_i \mathbf{k}_j^T / \sqrt{d}\} - \frac{1}{L_K} \sum_{j=1}^{L_k} \mathbf{q}_i \mathbf{k}_j^T / \sqrt{d} + \ln L_K$ . Επομένως προκύπτει η προτεινόμενη μέγιστη-μέση μέτρηση ως εξής:

$$\bar{M}(\mathbf{q}_i, \mathbf{K}) = \max_j \left\{ \frac{\mathbf{q}_i \mathbf{k}_j^T}{\sqrt{d}} \right\} - \frac{1}{L_K} \sum_{j=1}^{L_k} \frac{\mathbf{q}_i \mathbf{k}_j^T}{\sqrt{d}} \quad (3.14)$$

Ο ψευδοκώδικας του προτεινόμενου μηχανισμού προσοχής ProbSparse που επιτυγχάνει μικρότερη χρήση μνήμης περιγράφεται στον Αλγόριθμο 1 [75]. Οι βαθμολογίες-σκορ της αυτοπροσοχής (self-attention scores) σχηματίζουν μια κατανομή μακράς ουράς όπου τα «ενεργά» ερωτήματα (active queries) βρίσκονται στο πάνω κομμάτι του χάρτη χαρακτηριστικών, στις βαθμολογίες «κεφαλής», ενώ τα «τεμπέλικα» ερωτήματα (lazy queries) βρίσκονται στο κάτω κομμάτι του χάρτη χαρακτηριστικών, στις βαθμολογίες «ουράς». Ο μηχανισμός ProbSparse σχεδιάστηκε με τέτοιο τρόπο ώστε να επιλέγει τα «ενεργά» ερωτήματα και όχι τα «τεμπέλικα» ερωτήματα. Στο Σχήμα 3.7 φαίνονται αυτά τα δύο είδη ερωτημάτων. Ο μηχανισμός αυτός με τα ερωτήματα Top-u σχηματίζει έναν αραιό μετασχηματιστή (sparse Transformer) από την κατανομή πιθανοτήτων. Η επιλογή των ερωτημάτων αντί για τα κλειδιά γίνεται διότι επιτυγχάνεται καλύτερη αναπαράσταση της εισόδου. Εάν γινόταν η επιλογή των κορυφαίων Top-u κλειδιών θα καταστρεφόταν η αναπαράσταση των κορυφαίων συστατικών της εισόδου, καθώς τα πλήρη κλειδιά διατηρούν απλώς το τετριμμένο άθροισμα των τιμών εντός των βαθμολογιών της κατανομής.

---

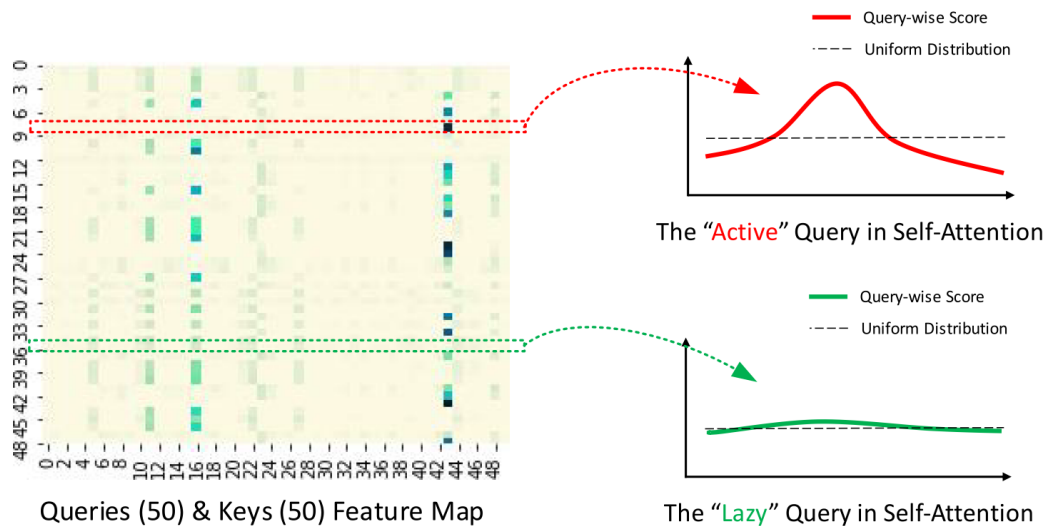
**Algorithm 1 - Function** forward ( $\mathbf{Q}, \mathbf{K}, \mathbf{V}, c$ ):

---

**Input:** tensor  $\mathbf{Q} \in \mathbb{R}^{L_Q \times d}$ ,  $\mathbf{K} \in \mathbb{R}^{L_K \times d}$ ,  $\mathbf{V} \in \mathbb{R}^{L_V \times d}$ ; sampling factor  $c$

**Output:** the self-attention feature map  $\mathbf{S}$

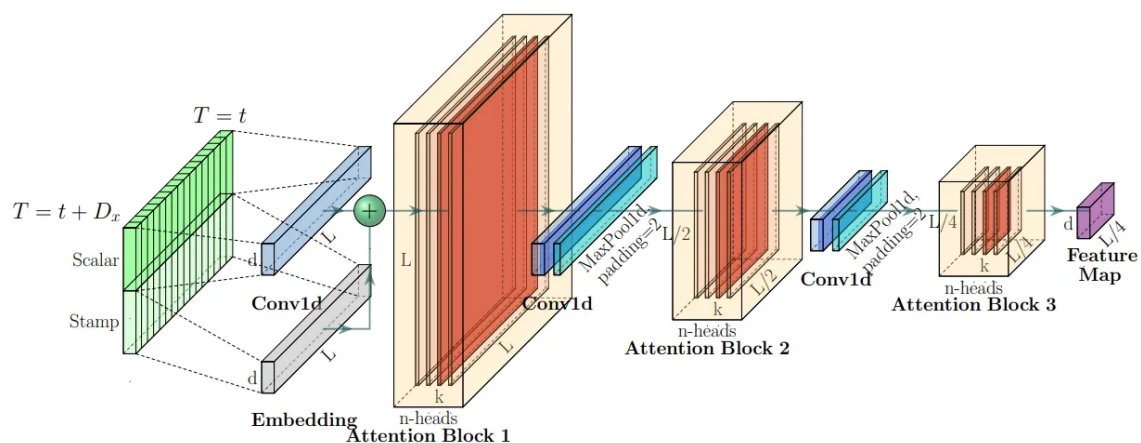
- 1: Hyperparameter  $\rightarrow$  set  $u = c \ln L_Q$  and  $U = L_Q \ln L_K$
  - 2: randomly select  $U$  dot-product pairs as the cropping operator  $[\cdot]$
  - 3: set the sample matrix  $\bar{\mathbf{S}} = [\mathbf{Q}\mathbf{K}^T]_U$
  - 4: compute the measurement  $\bar{M} = \max(\bar{\mathbf{S}}) - \text{mean}(\bar{\mathbf{S}})$  for each query
  - 5: select Top- $u$  queries under  $\bar{M}$  as  $\bar{\mathbf{Q}}$
  - 6: Header field  $\rightarrow$  set  $\mathbf{S}_h = \text{Softmax}(\bar{\mathbf{Q}}\mathbf{K}^T \sqrt{d}) \cdot \mathbf{V}$
  - 7: Tail field  $\rightarrow$  set  $\mathbf{S}_t = \text{mean}(\mathbf{V})$
  - 8: build  $\mathbf{S} = \{\mathbf{S}_h, \mathbf{S}_t\}$  by the original row arrangement accordingly
-



**Σχήμα 3.7:** Απεικόνιση του μηχανισμού προσοχής ProbSparse (Πηγή: 'Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting' από τους Zhou et al. [40]).

### Encoder

Ο κωδικοποιητής σχεδιάστηκε με τέτοιο τρόπο ώστε να γίνεται εφικτή η εξαγωγή ισχυρών εξαρτήσεων μεγάλης κλίμακας των μεγάλων διαδοχικών εισόδων υπό τον περιορισμό της χρήσης μνήμης. Στο Σχήμα 3.8 απεικονίζεται η αρχιτεκτονική του κωδικοποιητή.



**Σχήμα 3.8:** Αρχιτεκτονική κωδικοποιητή (Πηγή: 'Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting' από τους Zhou et al. [40]).

Ο χάρτης των χαρακτηριστικών (feature map) περιέχει επιπλέον συνδυασμούς τιμών  $V$  (Values) ως αποτέλεσμα του μηχανισμού αυτοπροσοχής ProbSparse. Η λειτουργία της αυτοπροσοχής με απόσταση (self-attention distilling) χρησιμοποιείται για τη βαθμολόγηση των κυρίαρχων χαρακτηριστικών και τη δημιουργία ενός συγκεντρωμένου χάρτη χαρακτηριστικών αυτοπροσοχής στο επόμενο επίπεδο. Στην ουσία πρόκειται για μια διαδικασία συμπίεσης του μεγάλου και πολύπλοκου μηχανισμού αυτοπροσοχής μοντέλου μετασχηματιστή (transformer model), σε μια

---

απλούστερη και μικρότερη μορφή που μπορεί να χρησιμοποιηθεί στο μοντέλο Informer.

Όπως διακρίνεται από το Σχήμα 3.8 ο κωδικοποιητής αποτελείται από αρκετά μπλοκ προσοχής (attention blocks), συνελκτικά επίπεδα Conv1d και επίπεδα maxpooling, για να κωδικοποιηθούν τα δεδομένα εισόδου. Τα αντίγραφα της κύριας στοίβας (Attention Block 1) με συνεχώς μειούμενες εισροές στο μισό, αυξάνουν την αξιοπιστία της λειτουργίας απόσταξης (distilling operation). Στο τέλος του κωδικοποιητή, συνδυάζεται ο χάρτης χαρακτηριστικών (Feature Map) για να κατευθυνθεί η έξοδος του κωδικοποιητή απευθείας στον αποκωδικοποιητή.

### Decoder

Ο δομή του αποκωδικοποιητή δεν διαφέρει σε μεγάλο βαθμό με εκείνου που δημοσιεύτηκε από τους Vaswani et al. [22]. Είναι ικανός για τη δημιουργία μεγάλων διαδοχικών εξόδων μέσω μόνο μίας διαδικασίας προώθησης (one forward pass). Όπως φαίνεται και στο Σχήμα 3.6 περιλαμβάνει μια στοίβα από πανομοιότυπα στρώματα προσοχής πολλαπλών κεφαλών (multi-head attention layers). Η κύρια διαφορά είναι ο τρόπος με τον οποίο παράγονται οι προβλέψεις μέσω μίας διαδικασίας που ονομάζεται generative inference και επιταχύνει σημαντικά τη διαδικασία μακροχρόνιας πρόβλεψης (long-range prediction). Ο αποκωδικοποιητής τροφοδοτείται με τα παρακάτω διανύσματα:

$$\mathbf{X}_{de}^t = \text{Concat}(\mathbf{X}_{token}^t, \mathbf{X}_0^t) \in \mathbb{R}^{(L_{token}+L_y) \times d_{model}} \quad (3.15)$$

,όπου το  $\mathbf{X}_{token}^t \in \mathbb{R}^{L_{token} \times d_{model}}$  είναι το αρχικό token, το  $\mathbf{X}_0^t \in \mathbb{R}^{L_y \times d_{model}}$  είναι κρυφή θέση (placeholder) για την ακολουθία στόχο. Στη συνέχεια εφαρμόζεται ο μηχανισμός masked multi-head attention στον υπολογισμό της αυτοπροσοχής ProbSparse θέτοντας τα εσωτερικά γινόμενα στο  $-\infty$ . Αποτρέπεται, με αυτόν τον τρόπο, η κάθε θέση να παρακολουθεί τις επόμενες θέσεις, γεγονός που αποτρέπει την αυτόματη παλινδρόμηση (auto-regressive). Τέλος ένα πλήρως συνδεδεμένο στρώμα παράγει την τελική έξοδο, ενώ η διάσταση της εξόδου  $d_y$  εξαρτάται από το αν υλοποιείται μονομεταβλητή ή πολυμεταβλητή πρόβλεψη.

---

## Κεφάλαιο 4

# Στοχαστική προσέγγιση για πρόβλεψη χρονοσειρών

### 4.1 Στοχαστικές διεργασίες - Γενικά

Οι *στοχαστικές διεργασίες* είναι οικογένειες τυχαίων μεταβλητών (συμβολίζεται ως  $x(t)$ , όπου τα υπογραμμισμένα σύμβολα υποδηλώνουν τυχαίες μεταβλητές και το  $t$  υποδηλώνει το χρόνο). Συχνά χρησιμοποιούνται για την αναπαράσταση της χρονικής εξέλιξης των φυσικών διεργασιών. Οι φυσικές διεργασίες καθώς και η μαθηματική αναπαράστασή τους ως στοχαστικές διεργασίες εξελίσσονται σε συνεχή χρόνο. Κάθε τυχαία μεταβλητή αντιπροσωπεύει την τιμή ή την κατάσταση του συστήματος σε μια συγκεκριμένη χρονική στιγμή και η συλλογή αυτών των μεταβλητών αποτελεί τη στοχαστική διαδικασία. Ένα παράδειγμα στοχαστικής διεργασίας φαίνεται στο Σχήμα 4.1.

Ενώ η στοχαστική μοντελοποίηση διαβάζει το παρελθόν με παρόμοιο τρόπο με το αντίστοιχο ντετερμινιστικό μοντέλο, όταν αναφέρεται στο μέλλον, η στοχαστική μέθοδος δεν παράγει μια προβολή, αλλά μάλλον μια πιθανολογική εκτίμησή του, η οποία εξαρτάται από τις παρατηρήσεις του παρελθόντος. Με αυτόν τον τρόπο λαμβάνεται υπόψη και εκτιμάται ο παράγοντας αβεβαιότητας. Μια άλλη σημαντική διάκριση της στοχαστικής μεθοδολογίας στη μοντελοποίηση και στην πρόβλεψη είναι η λιτότητα που τη διακρίνει. Γενικά η αρχή της απλότητας στην επιστήμη και στη μοντελοποίηση φυσικών φαινομένων αποτελεί βάση εφαρμογής, ενώ παράλληλα η συνεχής εξέλιξη στην υπολογιστική ισχύ και στις σύγχρονες μηχανές έχει οδηγήσει σε παραμέληση της απλότητας που οδηγεί στην πολυπλοκότητα και στην αναπαραγωγή κάθε λεπτομέρειας της φυσικής πολυπλοκότητας [76].

Στην οικογένεια αυτή των τυχαίων μεταβλητών  $X(t)$ , όπου  $t \in T$ , αν το  $T$  είναι το σύνολο των πραγματικών αριθμών τότε πρόκειται για ανέλιξη συνεχούς χρόνου, ενώ αν το  $T$  είναι σύνολο ακέραιων αριθμών τότε η ανέλιξη ονομάζεται διακριτού χρόνου [77]. Για ένα συγκεκριμένο



πεδίο χρόνου  $t$ , η  $\underline{x}(t)$  είναι μία τυχαία μεταβλητή με κατανομή:

$$F(x, t) = P(\underline{x}(t) < x) \quad (4.1)$$

,όπου η συνάρτηση  $F(x, t)$  ονομάζεται συνάρτηση κατανομής πρώτης τάξης της διεργασίας  $\underline{x}(t)$ , ενώ η μερική παράγωγος ως προς  $x$  ονομάζεται συνάρτηση πυκνότητας πιθανότητας ενώ η παράγωγος ως προς  $x$ , ονομάζεται συνάρτηση πυκνότητας πιθανότητας  $f(x)$  ή αλλιώς συνάρτηση πυκνότητας πρώτης τάξης:

$$f(x, t) := \frac{dF(x)}{dx} \quad (4.2)$$

Συνήθως η γνώση της συνάρτησης κατανομής δεν είναι απαραίτητη, καθώς απαιτούνται μόνο ορισμένες ιδιότητες της διεργασίας. Μερικές από αυτές οι ιδιότητες είναι η μέση τιμή (mean), η αυτοσυσχέτιση (autocorrelation) και η αυτοσυνδιακύμανση (autocovariance) ή αυτοσυνδιασπορά, οι οποίες χαρακτηρίζονται και ιδιότητες δεύτερης τάξης της στοχαστικής διεργασίας  $\underline{x}(t)$ . Στη συνέχεια παρουσιάζονται οι ορισμοί αυτών των ιδιοτήτων (εξισώσεις (4.3)-(4.7)) όπως αναφέρονται στο [77]:

- *Μέση τιμή*: Η μέση τιμή  $\mu(t)$  της  $\underline{x}(t)$  ορίζεται ως η αναμενόμενη τιμή της τυχαίας μεταβλητής  $\underline{x}(t)$ .

$$\mu(t) := E[\underline{x}(t)] = \int_{-\infty}^{\infty} x f(x, t) dx \quad (4.3)$$

- *Αυτοσυσχέτιση*: Η αυτοσυσχέτιση  $R(t_1, t_2)$  της  $\underline{x}(t)$  ορίζεται ως η αναμενόμενη τιμή του εσωτερικού γινομένου των  $\underline{x}(t_1)\underline{x}(t_2)$ :

$$R(t_1, t_2) := E[\underline{x}(t_1)\underline{x}(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2; t_1, t_2) dx_1 dx_2 \quad (4.4)$$

Για  $t_1 = t_2 = t$  η τιμή της  $R(t_1, t_2)$  παίρνει την μορφή:

$$R(t, t) = E[\underline{x}^2(t)] \quad (4.5)$$

- *Αυτοσυνδιακύμανση*: Η αυτοσυνδιακύμανση  $C(t_1, t_2)$  της  $\underline{x}(t)$  ορίζεται ως η συνδιακύμανση των τυχαίων μεταβλητών  $\underline{x}(t_1)$  και  $\underline{x}(t_2)$ :

$$C(t_1, t_2) = R(t_1, t_2) - \mu(t_1)\mu(t_2) \quad (4.6)$$

Για  $t_1 = t_2 = t$  η τιμή της  $C(t_1, t_2)$  παίρνει την μορφή:

$$C(t, t) = \text{Var}(\underline{x}(t)) = \text{E} \left[ \left( \underline{x}(t) - \mu_x(t) \right)^2 \right] \quad (4.7)$$

Κανονικοποιώντας την συνάρτηση αυτοσυσχέτισης  $R(t_1, t_2)$  με αφαίρεση του μέσου όρου και διαίρεση με τη διακύμανση λαμβάνουμε τον συντελεστή αυτοσυσχέτισης [78]:

$$\rho(t_1, t_2) = \frac{C(t_1, t_2)}{\sqrt{C(t_1, t_1)}\sqrt{C(t_2, t_2)}}, |\rho(t_1, t_2)| \leq 1 \quad (4.8)$$

Δύο βασικές έννοιες μιας στοχαστικής διεργασίας είναι η στασιμότητα (stationarity) και η εργοδικότητα (ergodicity). Στάσιμη ονομάζεται μία στοχαστική διεργασία στην οποία η στατιστικές ιδιότητες παραμένουν σταθερές με την πάροδο του χρόνου. Με άλλα λόγια, έχει την ίδια κατανομή πιθανότητας για τις τιμές της σε κάθε χρονική στιγμή, ανεξάρτητα από το πότε γίνονται οι παρατηρήσεις. Για την ακρίβεια μία διεργασία μπορεί να οριστεί στάσιμη με την αυστηρή έννοια (strict-sense) και με την ευρεία έννοια (wide-sense) [77, 79]:

- *Στασιμότητα με την αυστηρή έννοια:* Όταν η από κοινού κατανομή της δεν επηρεάζεται από μια αλλαγή στο χρόνο. Δηλαδή, η από κοινού κατανομή (joint probability distribution) για τις χρονικές στιγμές  $t_1, t_2, \dots, t_n$  είναι ίδια με την από κοινού κατανομή για τις χρονικές στιγμές  $t_{1+\tau}, t_{2+\tau}, \dots, t_{n+\tau}$ , όπου  $\tau$  θετικός ακέραιος.
- *Στασιμότητα με την ευρεία έννοια:* Όταν η μέση τιμή είναι σταθερή, εξίσωση (4.9) και η αυτοσυνδιασπορά της εξαρτάται μόνο από τη διαφορά του χρόνου, εξίσωση (4.10), δηλαδή:

$$\text{E}[x(t)] = \mu \quad (4.9)$$

$$\text{E}[(\underline{x}(t) - \mu)(\underline{x}(t + \tau) - \mu)] = C(\tau) \quad (4.10)$$

Μια στάσιμη στοχαστική διεργασία είναι εργοδική, αν κάθε παράμετρος της κατανομής μπορεί να προσδιοριστεί από μια απλή δειγματοσυνάρτηση της διεργασίας. Δεδομένου ότι οι παράμετροι υπολογίζονται ως χρονικές μέσες τιμές, ο παραπάνω ορισμός εκφράζεται και με τον εξής τρόπο: Μια διεργασία είναι εργοδική αν οι χρονικοί μέσοι είναι ίσοι με τους θεωρητικούς μέσους. Για παράδειγμα, στην εξίσωση (4.11) φαίνεται μια στοχαστική διεργασία που είναι εργοδική ως προς τη μέση τιμή αν για διακριτό χρόνο, ενώ στην εξίσωση (4.12) για συνεχή χρόνο.

$$E[\underline{x}(t)] = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=0}^N \underline{x}(t) \quad (4.11)$$

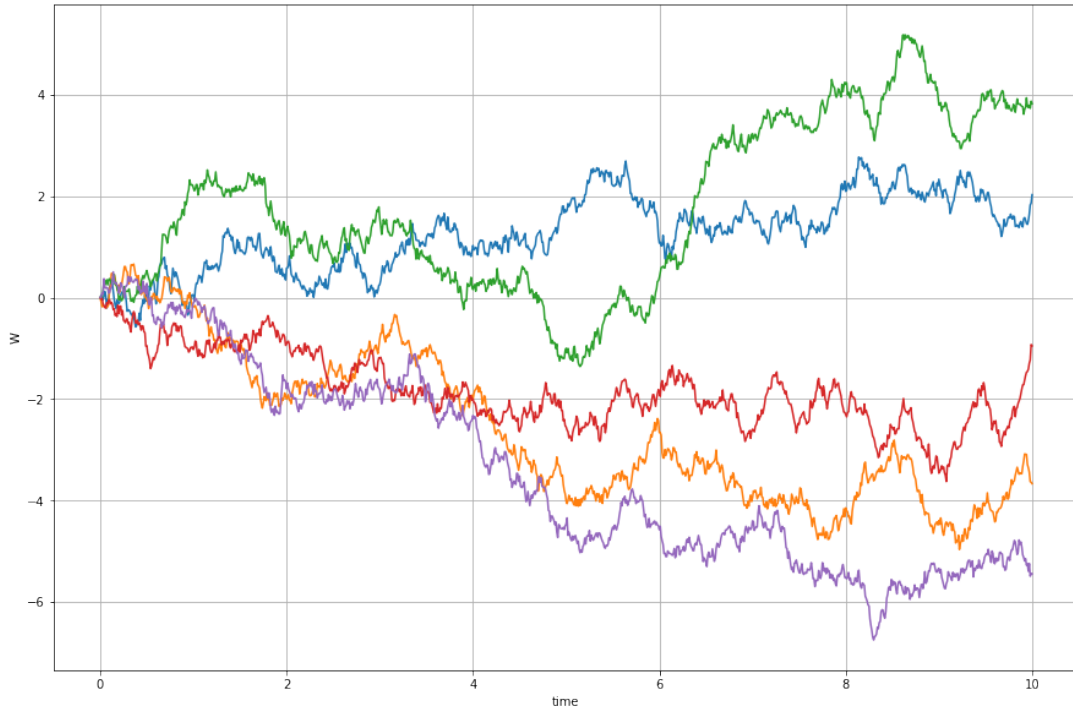
$$E[\underline{x}(t)] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \underline{x}(t) dx \quad (4.12)$$

Υπάρχουν ορισμένες ευρέως χρησιμοποιούμενες στοχαστικές διεργασίες όπως Λευκός Θόρυβος (White Noise), Wiener, Markov, Poisson κ.ά. Ο λευκός θόρυβος είναι μια βασική και απλή στοχαστική διαδικασία που χαρακτηρίζεται από ασυσχέτιστες και πανομοιότυπα κατανεμημένες τυχαίες μεταβλητές. Έχει σταθερή μέση τιμή και διακύμανση και κάθε παρατήρηση είναι ανεξάρτητη από τις άλλες. Ο λευκός θόρυβος χρησιμοποιείται συχνά ως διαδικασία αναφοράς ή ως βασική διαδικασία στη στατιστική ανάλυση. Η διεργασία Wiener (Σχήμα 4.1), γνωστή και ως κίνηση Brown (Brownian motion), είναι μια στοχαστική διεργασία συνεχούς χρόνου που μοντελοποιεί την τυχαία κίνηση σωματιδίων σε ένα ρευστό. Έχει συνεχείς διαδρομές και χαρακτηρίζεται από γκαουσιανές προσανξήσεις. Η διαδικασία Wiener χρησιμοποιείται ευρέως σε διάφορους τομείς, συμπεριλαμβανομένης της φυσικής, της χρηματοοικονομικής και της μηχανικής, για τη μοντελοποίηση τυχαίων διακυμάνσεων και συμπεριφορών που μοιάζουν με διάχυση. Μια διαδικασία Markov, επίσης γνωστή ως αλυσίδα Markov (Markov chain), είναι μια στοχαστική διαδικασία που ακολουθεί την ιδιότητα Markov, η οποία δηλώνει ότι η μελλοντική κατάσταση εξαρτάται μόνο από την παρούσα κατάσταση και είναι ανεξάρτητη από τις προηγούμενες καταστάσεις. Οι διαδικασίες Markov χρησιμοποιούνται ευρέως για τη μοντελοποίηση συστημάτων με μεταβάσεις χωρίς μνήμη, όπως οι τυχαίοι περίπατοι (Random walk). Τέλος η διεργασία Poisson είναι μια διαδικασία καταμέτρησης που μοντελοποιεί την εμφάνιση διακριτών γεγονότων σε συνεχή χρόνο. Χρησιμοποιείται συχνά για τη μοντελοποίηση αφίξεων, όπως η άφιξη πελατών σε έναν πάγκο εξυπηρέτησης.

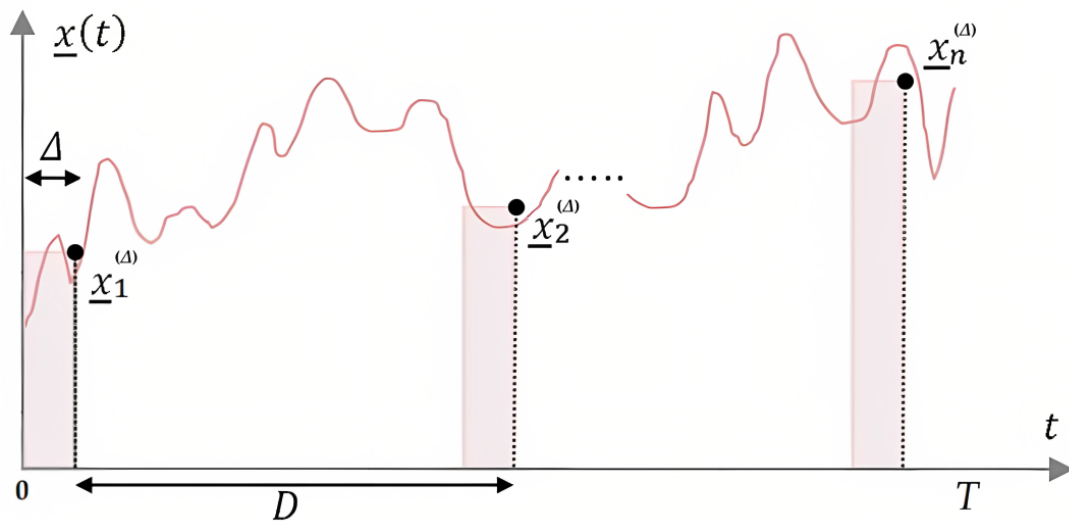
Οι χρονοσειρές που προέρχονται από φυσικές διεργασίες χαρακτηρίζονται από ένα χρονικό διάστημα δειγματοληψίας  $\Delta$ , συχνά καθορισμένο από τον παρατηρητή και από ένα χρόνο απόκρισης  $\Delta$  του οργάνου (Σχήμα 4.2). Οι χρονικές σταθερές  $D$  και  $\Delta$  επηρεάζουν την εκτίμηση των στατιστικών ιδιοτήτων διεργασίας συνεχούς χρόνου [80]. Η στοχαστική διεργασία διακριτού χρόνου  $\underline{x}_i^{(\Delta)}$ , για  $D = \Delta > 0$  μπορεί να υπολογιστεί από την  $\underline{x}(t)$  ως:

$$\underline{x}_i^{(\Delta)} = \frac{\int_{(i-1)\Delta}^i \underline{x}(\xi) d\xi}{\Delta} \quad (4.13)$$

,όπου  $i \in [1, n]$  είναι ένας δείκτης που αναπαριστά το διακριτό χρόνο,  $n = [T/\Delta]$  ο συνολικός αριθμός των παρατηρήσεων και  $T \in [0, \infty)$  η χρονική διάρκεια των παρατηρήσεων.



Σχήμα 4.1: Παράδειγμα στοχαστικής διεργασίας Wiener (Πηγή: [81]).



Σχήμα 4.2: Ένα παράδειγμα μιας διαδικασίας συνεχούς χρόνου με δειγματοληψία σε χρονικά διαστήματα  $D$  για μια συνολική περίοδο  $T$  και με χρόνο απόκρισης του οργάνου  $\Delta$  (Πηγή: [80]).

## 4.2 Κλιμακόγραμμα

Το κλιμακόγραμμα [82] αποτελεί ένα βασικό εργαλείο που χρησιμοποιείται σε στοχαστικές αναλύσεις χρονοσειρών φυσικών φαινομένων. Όπως αναφέρεται στο [80] ο όρος προέρχεται από την ελληνική λέξη κλίμα και ορίζεται ως το γράφημα της διασποράς μιας στοχαστικής διεργασίας  $\underline{x}(t)$  (θεωρώντας στασιμότητα) ως προς την χρονική κλίμακα  $k$  και συμβολίζεται με τον όρο  $\gamma(k)$ . Το κλιμακόγραμμα είναι αρκετά χρήσιμο για την ανίχνευση της μακροχρόνιας με-

ταβολής ή αλλιώς εξάρτησης, εμμονής, ομαδοποίησης μιας διεργασίας, κάτι που συμβαίνει και με τα μοντέλα Transformers (βλ. Κεφάλαιο 3). Αυτό μπορεί να συμβεί μέσω του συντελεστή Hurst  $H$ , που ισούται με το μισό της κλίσης του κλιμακογράμματος σε ένα διπλό λογαριθμικό γράφημα.

Το εμπειρικό κλιμακόγραμμα μπορεί να απεικονιστεί υπολογίζοντας την διασπορά (ή την τυπική απόκλιση) μιας χρονοσειράς, συναθροισμένης σε πολλαπλές κλίμακες  $k$ , για κάθε μια κλίμακα  $k$ , και εμφανίζοντας το διπλό λογαριθμικό διάγραμμα διασποράς-κλίμακας (ή τυπικής απόκλισης-κλίμακας). Το θεωρητικό κλιμακόγραμμα μιας στοχαστικής διεργασίας υπολογίζεται μέσω της εξίσωσης κλιμακογράμματος για την εκάστοτε διεργασία. Στη συνέχεια αναλύονται δύο βασικές κατηγορίες διεργασιών, οι διεργασίες Markov και οι διεργασίες Hurst-Kolmogorov, και ορίζονται οι εξισώσεις των θεωρητικών τους κλιμακογραμμάτων.

Η εξίσωση υπολογισμού του εμπειρικού κλιμακογράμματος μιας διεργασίας συναθροισμένης σε κλίμακα  $k$ , για συνεχή χρόνο (εξίσωση (4.14)) και για διακριτό χρόνο (εξίσωση (4.15)) δίνονται ως [80]:

$$\gamma(m) := \frac{Var[\sum_{l=k(i-1)+1}^{ki} x_l^{(\Delta)}]}{k^2} = \frac{Var[\sum_{l=1}^k x_l^{(\Delta)}]}{k^2} \quad (4.14)$$

$$\gamma_k^{(\Delta)}(k) := \frac{Var[\sum_{l=k(i-1)+1}^{ki} x_l^{(\Delta)}]}{k^2} = \frac{Var[\sum_{l=1}^k x_l^{(\Delta)}]}{k^2} \quad (4.15)$$

Όπως αναφέρθηκε και προηγουμένως οι διεργασίες Markov βασίζονται στην απλή παραδοχή πως η μελλοντική κατάσταση εξαρτάται μόνο από την παρούσα κατάσταση και είναι ανεξάρτητη από τις προηγούμενες καταστάσεις. Στοχαστικά μοντέλα που αντιστοιχούν στις διεργασίες Markov, όπως το μοντέλο AR(1) δεν μοντελοποιούν τις φυσικές διεργασίες αποτελεσματικά. Επομένως για την επίτευξη της μοντελοποίησης της μακροπρόθεσμης δομής που παρουσιάζουν οι φυσικές διεργασίες, ορίζονται παρακάτω οι στοχαστικές διεργασίες Hurst-Kolmogorov.

### 4.3 Διεργασίες Hurst-Kolmogorov

Εν αντιθέσει με τις διεργασίες Markov, στις διεργασίες Hurst-Kolmogorov (HK), το μέλλον εξαρτάται από το καταγεγραμμένο παρελθόν [83]. Στην εξίσωση (4.16) ορίζεται μαθηματικά το κλιμακόγραμμα αυτών των διεργασιών [76].

$$\gamma(k) = \lambda(\alpha/k)^{2-2H} \quad (4.16)$$

,όπου οι συντελεστές  $\lambda$  και  $\alpha$  αποτελούν αδιάστατους συντελεστές του κλιμακογράμματος,

ενώ ο συντελεστής  $H$  είναι ο συντελεστής Hurst, και περιγράφει τη μακροπρόθεσμη συμπεριφορά της δομής της αυτοσυσχέτισης, ή αλλιώς την ύπαρξη εμμονής σε μια διεργασία, όπως περιγράφεται από το φαινόμενο Hurst. Πιο συγκεκριμένα με βάση τον συντελεστή Hurst διακρίνονται τρεις περιπτώσεις ως εξής:

- $H = 0.5$ : τυχαία μεταβολή (λευκός θόρυβος). Σε αυτή τη περίπτωση δεν υπάρχει συσχέτιση μεταξύ των στοιχείων στο χρόνο  $t$  και  $t + 1$ .
- $0.5 < H < 1$ : η χρονοσειρά εμφανίζει μακροπρόθεσμη εμμονή ή αλλιώς μακροπρόθεσμη εξάρτηση. Μια μεγάλη τιμή έχει μεγαλύτερη πιθανότητα να ακολουθείται από μια μεγάλη τιμή και το αντίστοιχο για τις μικρές τιμές. Πρόκειται για την συνηθέστερη περίπτωση γεωφυσικών διεργασιών.
- $0 < H < 0.5$ : η χρονοσειρά παρουσιάζει συμπεριφορά αντίθετη της εμμονής, δηλαδή μια αύξηση της τιμής ακολουθείται από μια μείωση της τιμής.

Η δυναμική Hurst-Kolmogorov προέρχεται από τους Harold Edwin Hurst (1880-1978) και Andrey Kolmogorov (1903-1987) ορίζει πως σε τυχαία γεγονότα εμφανίζονται υψηλές και χαμηλές τιμές, ενώ σε μια φυσική διεργασία η τάση να συμβαίνουν τέτοια γεγονότα είναι μεγαλύτερη και τείνουν να συμβαίνουν σε ομάδες. Οι τιμές του συντελεστή  $H$ , αναπαρίστανται από την κλίση του κλιμακογράμματος μιας χρονοσειράς, η οποία είναι  $2 - 2H$ , στην περίπτωση του κλιμακογράμματος διασποράς και  $1 - H$  στην περίπτωση του κλιμακογράμματος τυπικής απόκλισης.

## 4.4 Εκτίμηση πρόβλεψης χρονοσειρών

Όπως διατυπώνεται στο [79] στην κλασική στατιστική τα δείγματα είναι εξ ορισμού σύνολα IID στοχαστικών μεταβλητών. Για την εκτίμηση μελλοντικών καταστάσεων οι κλασικές στατιστικές μέθοδοι χρησιμοποιούν όλες τις παρατηρήσεις του δείγματος. Μπορεί να διαμορφωθεί το εξής ερώτημα: πόσοι όροι του παρελθόντος χρειάζεται να ληφθούν υπόψη για την εκτίμηση ενός μέσου όρου ο οποίος να είναι αντιπροσωπευτικός για τον μέσο όρο του μέλλοντος για μία περίοδο μήκους  $\kappa$ ;

Ο μέσος όρος της μελλοντικής περιόδου μήκους  $\kappa$  ως προς το παρόν και το παρελθόν ορίζεται:

$$\underline{\mu}_\kappa := \mathbb{E} \left[ \frac{1}{\kappa} (x_1 + \dots + x_\kappa) \mid \underline{x}_0, \underline{x}_{-1}, \dots \right] \quad (4.17)$$

Υποθέτοντας ότι υπάρχει ένας μεγάλος αριθμός  $n$  παρατηρήσεων από το παρόν και το παρελθόν, αλλά επιλέγονται μόνο  $\nu \leq n$  για την εκτίμηση, τότε έχουμε:

$$\hat{\underline{\mu}}_{\nu} := \frac{1}{\nu}(\underline{x}_0 + \underline{x}_{-1} + \dots + \underline{x}_{-\nu+1}) \quad (4.18)$$

Για να απαντηθεί το παραπάνω ερώτημα αρκεί να βρεθεί το  $\nu$  που ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα:

$$A(\kappa, \nu) := \mathbb{E} \left[ (\hat{\underline{\mu}}_{\nu} - \hat{\underline{\mu}}_{\kappa})^2 \right] \quad (4.19)$$

, η οποία μπορεί να γραφτεί ως:

$$\begin{aligned} A(\kappa, \nu) &= \mathbb{E} \left[ \left( \frac{1}{\kappa}(\underline{x}_1 + \dots + \underline{x}_{\kappa}) - \frac{1}{\nu}(\underline{x}_0 + \underline{x}_{-1} + \dots + \underline{x}_{-\nu+1}) \right)^2 \right] \\ &= \mathbb{E} \left[ \left( -\frac{\underline{x}_1}{\nu} - \dots - \frac{\underline{x}_{\nu}}{\nu} + \frac{\underline{x}_{\nu+1}}{\kappa} + \dots + \frac{\underline{x}_{\nu+\kappa}}{\kappa} \right)^2 \right] \end{aligned} \quad (4.20)$$

Όπως αποδεικνύεται στο [79] η (4.20) γράφεται σε όρους κλιμακογράμματος ως:

$$A(\kappa, \nu) = \left( \frac{1}{\kappa} + \frac{1}{\nu} \right) (\kappa\gamma(\kappa) + \nu\gamma(\nu) - (\nu + \kappa)\gamma(\nu + \kappa)) \quad (4.21)$$

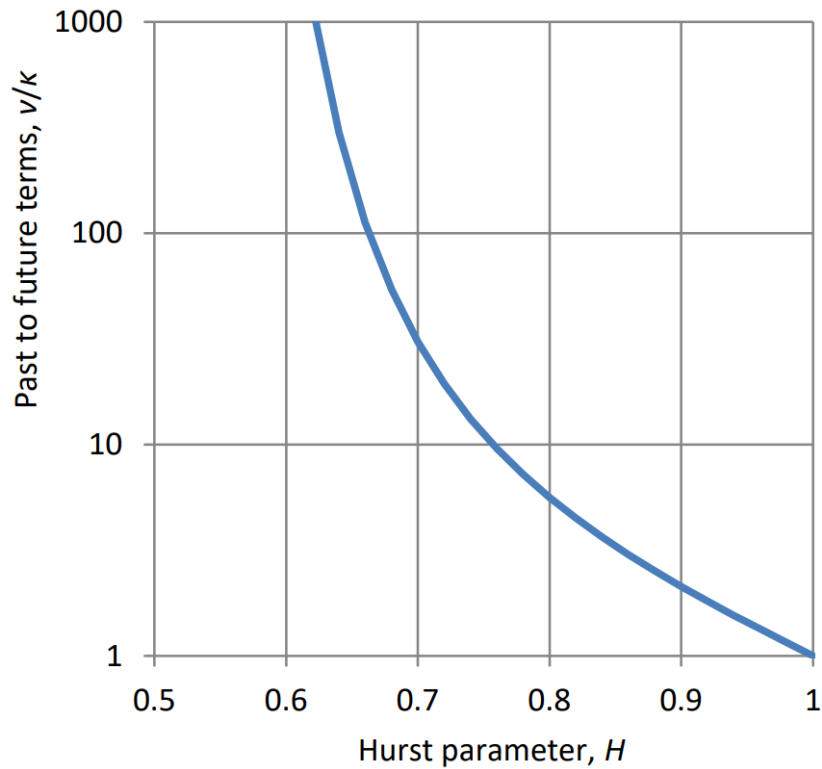
Για την περίπτωση της διεργασίας Hurst-Kolmogorov για την οποία η συνάρτηση μοιάζει με  $\gamma(\kappa) = \lambda^2(\kappa/\alpha)^{2H-2}$ , η τιμή του  $\nu$  που ελαχιστοποιεί το σφάλμα  $A$  είναι:

$$\nu = \frac{\kappa}{(\max(0, 2.5H - 1.5))^{2.5}} \quad (4.22)$$

Η εξίσωση (4.21) διακρίνει δύο περιπτώσεις ανάλογα με την τιμή του  $H$ :

- Για  $H \leq 0.6$ , οδηγεί σε  $\nu = \infty$ , το οποίο σημαίνει πως μελλοντική μέση εκτίμηση είναι ο μέσος όρος του συνόλου των  $n$  παρατηρήσεων.
- Για  $H > 0.6$ , μπορούν να χρησιμοποιηθούν  $\nu < n$  όροι για την εκτίμηση του μελλοντικού μέσου.

Μία γραφική αναπαράσταση της εξίσωσης (4.22) φαίνεται στο Σχήμα 4.3. Στον κατακόρυφο άξονα απεικονίζεται ο λόγος των όρων του παρελθόντος με του μέλλοντος  $\nu/\kappa$  και στον οριζόντιο η τιμές της παραμέτρου Hurst  $H$ . Έτσι, όσο μεγαλύτερη γίνεται η παράμετρος  $H$ , τόσο λιγότερες τιμές του παρελθόντος απαιτούνται για την αντιπροσωπευτική πρόβλεψη του μέλλοντος.



**Σχήμα 4.3:** Γραφική απεικόνιση της εξίσωσης (4.22) (Πηγή: [79]).

Στη γενική περίπτωση στην οποία το κλιμακόγραμμα της χρονοσειράς είναι  $\gamma(\kappa) = \Gamma(\kappa)/\kappa^2$  θα πρέπει να βελτιστοποιηθεί η σχέση (4.23) για την εύρεση του ελάχιστου σφάλματος.

$$A(\kappa, \nu) = \left( \frac{1}{\kappa} + \frac{1}{\nu} \right) \left( \frac{\Gamma(\kappa)}{\kappa} + \frac{\Gamma(\nu)}{\nu} - \frac{\Gamma(\nu + \kappa)}{\nu + \kappa} \right) \quad (4.23)$$

Περισσότερες πληροφορίες για την ανάλυση και τις αποδείξεις της παρούσας μεθοδολογίας ο αναγνώστης παραπέμπεται στο βιβλίο του Koutsoyiannis [79].

Στη παρούσα εργασία η μεθοδολογία αυτού του Κεφαλαίου χρησιμοποιείται για την δημιουργία ενός απλού μοντέλου αναφοράς για την σύγκριση με τα υπόλοιπα μοντέλα.



---

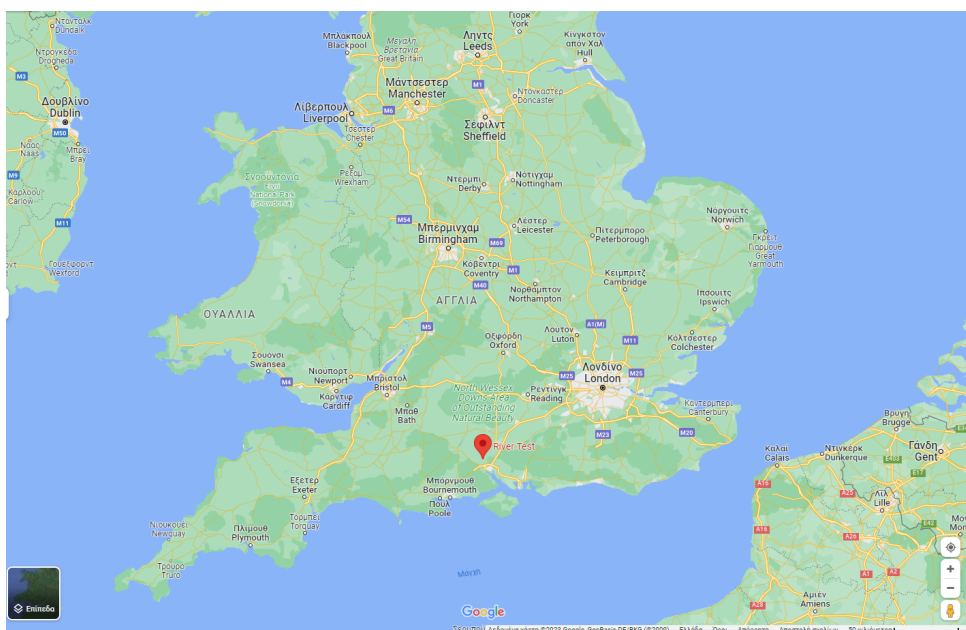
# Κεφάλαιο 5

## Εφαρμογή στον ποταμό River Test

### 5.1 Περιοχή μελέτης

Η περιοχή μελέτης επικεντρώνεται στον ποταμό River Test [84], ένα ποτάμι στο Hampshire στη νότια Αγγλία. Αναβλύζει στο Ashe κοντά στο Basingstoke και ρέει προς τα νότια για 64 km<sup>2</sup> μέχρι το Southampton Water. Στους οικισμούς του Test περιλαμβάνονται οι πόλεις Stockbridge και Romsey.

Μεγάλο μέρος του ποταμού είναι μια βιολογική περιοχή ειδικού επιστημονικού ενδιαφέροντος 438 ha (εκτάρια) ή 4 380 στρέμματα ή 4 380 000 km<sup>2</sup>. Αποτελεί μέρος της περιοχής Solent, Southampton Water Ramsar και της ζώνης Ειδικής Προστασίας. Ο ποταμός χρησιμοποιείται κατά κύριο λόγο για ψάρεμα από τις πηγές του μέχρι το παλιρροιακό του όριο.



Σχήμα 5.1: Χάρτης με την τοποθεσία του ποταμού River Test (Πηγή: Google Earth).

Ο ποταμός πηγάζει κοντά στο χωριό Ashe, 11 km δυτικά του Basingstoke, και ρέει δυτικά μέσα από τα χωριά Overton, Laverstoke και την πόλη Whitchurch, πριν ενωθεί με τον ποταμό Bourne Rivulet στο Testbourne. Στη συνέχεια διασχίζει τα χωριά Longparish και Middleton προς το Wherwell και το Chilbolton, όπου ενώνονται οι ποταμοί Dever και Anton. Η διαδρομή και του ποταμού φαίνεται στο Σχήμα 5.2 με κόκκινο χρώμα.



**Σχήμα 5.2:** Χάρτης του ποταμού River Test (με κόκκινο χρώμα), εντός της Νοτιοανατολικής Αγγλίας (Πηγή: [84]).

Ο ποταμός River Test είναι ένα σημαντικό υδατόρευμα γνωστό για την οικολογική του ποικιλομορφία, την ιστορική του και την υδρολογική του σημασία. Το ρέμα αυτό έχει πολύ πλούσια πανίδα και χλωρίδα, με κανένα άλλο ποτάμι πεδινής περιοχής στην Αγγλία να μην έχει πλουσιότερη. Έχουν καταγραφεί περισσότερα από 100 είδη ανθοφόρων φυτών στις όχθες του και 232 είδη ασπόνδυλων στο ποτάμι. Είναι επίσης σημαντικός υγρότοπος για τα πτηνά, με αρκετά είδη όπως παπαγάλοι κ.ά. Επιπλέον ο ποταμός φημίζεται για την ποιότητα των υδάτων του, η οποία ελέγχεται από τον Οργανισμό Περιβάλλοντος στην Αγγλία. Το οικοσύστημά του αποτελείται από διάφορα ενδιαίτηματα, παρέχοντας ένα ευρύ φάσμα μικροοικότοπων για τους υδρόβιους οργανισμούς. Ο ποταμός χρησιμεύει ως ζωτικός βιότοπος για πολλά είδη που εξαρτώνται από το νερό, ενώ έχει τόσο οικολογική όσο και πολιτιστική σημασία. Έχει αναγνωριστεί ως τοπίο ιδιαίτερου φυσικού κάλλους και τα γραφικά τοπία του προσελκύουν επισκέπτες και λάτρεις της φύσης.

Στη συνέχεια παρουσιάζεται μία επισκόπηση του σετ των δεδομένων του ποταμού που αξιοποιήθηκαν για την παρούσα εργασία.

## 5.2 Περιγραφή δεδομένων

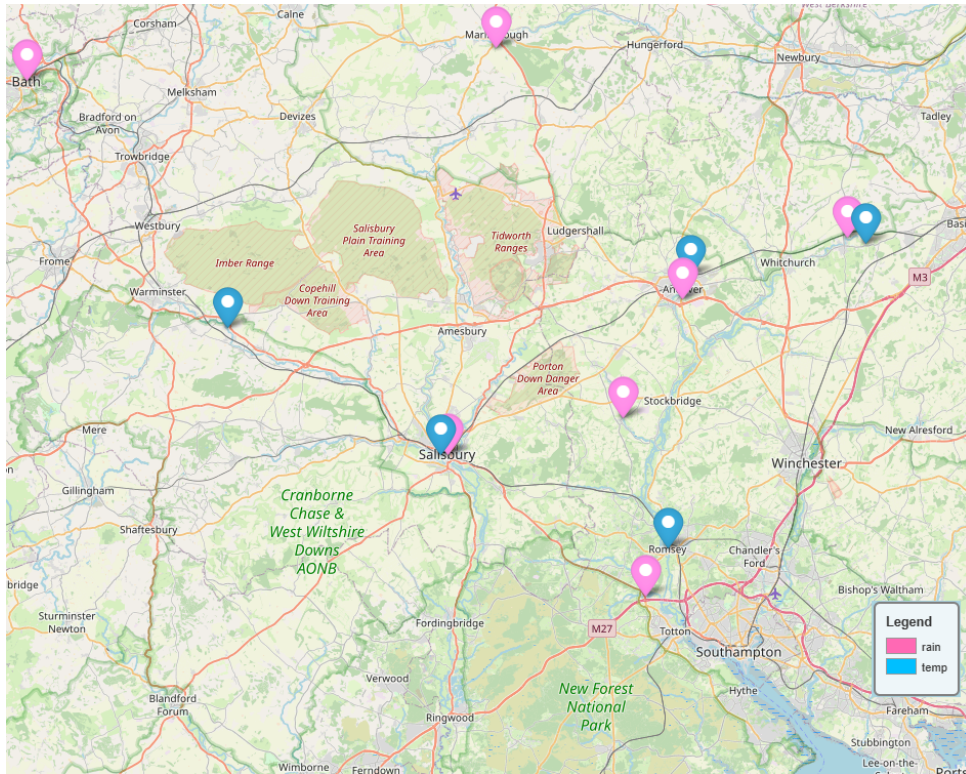
Το δεδομένα για τον ποταμό River Test είναι διαθέσιμα στο Kaggle [85], και τα οποία συλλέχθηκαν, αποθηκεύτηκαν, επεξεργάστηκαν και ενοποιήθηκαν σε ένα csv αρχείο από εθνικά κέντρα δεδομένων της Αγγλίας. Ειδικότερα αξιοποιήθηκαν τα κέντρα δεδομένων The CEDA Archive [86], Defra [87] και National River Flow Archive [88]. Συλλέχθηκαν 12 χαρακτηριστικά (features), μεταξύ των οποίων παροχή από σταθμούς του ποταμού και κλιματικά δεδομένα, όπως η βροχόπτωση και η θερμοκρασία, σε διαφορετικά σημεία της περιοχής Hampshire στην Αγγλία. Τα δεδομένα που παρέχονται είναι ημερήσια δεδομένα 41 ετών και καλύπτουν το χρονικό διάστημα μεταξύ 1980-2021. Πρόκειται λοιπόν για 15341 τιμές (σε κάποια χαρακτηριστικά λίγο λιγότερες), με το Σχήμα 5.3 να εμφανίζει τις 20 πρώτες γραμμές των δεδομένων σε μορφή πίνακα με όλα τα χαρακτηριστικά τους, ενώ στον Πίνακα 5.1 περιγράφονται τα χαρακτηριστικά αυτά. Επιπλέον αυτών, υπάρχει και αρχείο με τα μεταδεδομένα των σταθμών, συντεταγμένες και το μετρούμενο μέγεθος τους (βροχή, θερμοκρασία), όπως φαίνονται και στο χάρτη του Σχήματος 5.4.

	date	rainfall_Marborough	rainfall_Andover	rainfall_Bath	rainfall_Ower	rainfall_Quidhampton	rainfall_Salisbury	rainfall_Stockbridge	tasmax_Andover	tasmin_Andover	flow_Broadlands	flow_Ower
0	1/1/1980	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.44	-3.86	13.45	1.03
1	1/2/1980	0.85	0.25	1.16	0.36	1.05	0.68	0.59	3.26	-5.86	13.01	0.90
2	1/3/1980	12.03	11.89	5.86	10.27	11.27	10.04	12.34	10.49	-5.42	13.48	2.42
3	1/4/1980	0.92	0.19	1.18	0.47	0.28	0.54	0.30	8.69	2.71	16.15	2.73
4	1/5/1980	0.08	0.13	0.22	0.01	0.01	0.02	0.02	7.95	1.30	14.43	1.32
5	1/6/1980	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.56	2.51	13.63	1.08
6	1/7/1980	0.05	0.01	0.02	0.00	0.01	0.00	0.00	5.16	0.32	13.09	1.07
7	1/8/1980	0.00	0.00	0.00	0.00	0.01	0.00	0.00	4.74	0.53	12.69	1.06
8	1/9/1980	0.00	0.00	0.01	0.00	0.00	0.02	0.00	3.06	-1.52	12.52	0.82
9	1/10/1980	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.79	-0.44	12.31	0.55
10	1/11/1980	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.52	-3.44	11.98	0.51
11	1/12/1980	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.22	-3.93	11.88	0.48
12	1/13/1980	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.30	-4.57	11.80	0.47
13	1/14/1980	3.73	0.54	5.81	0.02	0.42	2.00	0.27	0.56	-5.13	11.71	0.47
14	1/15/1980	0.18	0.28	0.16	0.00	0.38	0.10	0.03	2.07	-4.36	11.24	0.46
15	1/16/1980	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.55	-0.41	11.61	0.46
16	1/17/1980	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.60	-1.70	11.58	0.48
17	1/18/1980	0.05	0.11	0.00	0.01	0.03	0.36	0.46	1.35	-1.73	11.49	0.55
18	1/19/1980	2.97	2.11	1.23	4.29	2.11	4.27	3.12	4.23	-1.99	11.41	0.67
19	1/20/1980	15.12	17.80	11.23	17.23	13.77	18.35	16.97	6.74	-0.26	11.53	1.12

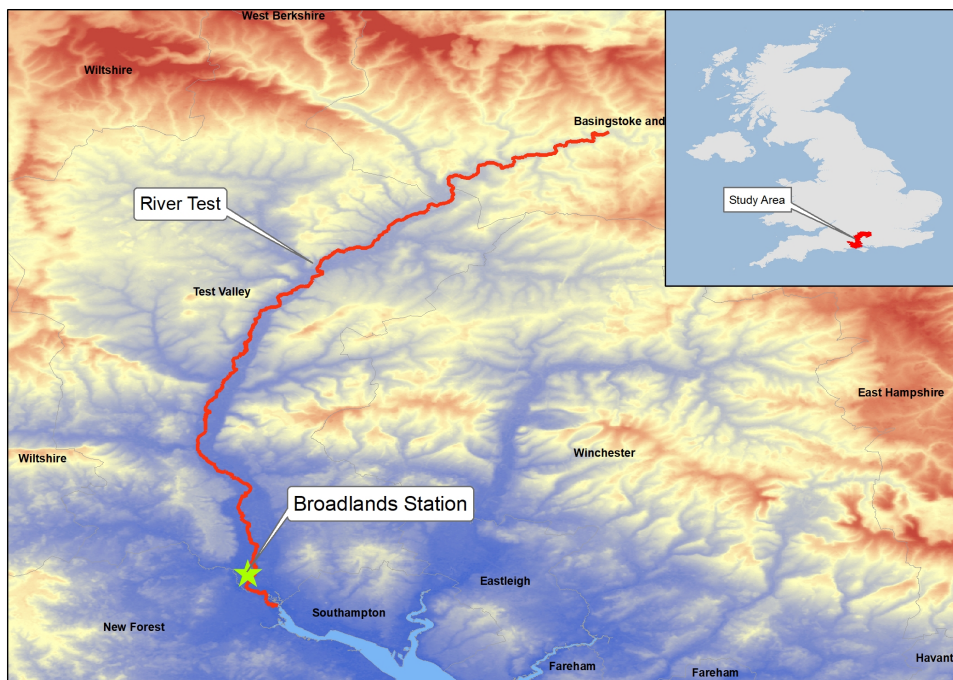
Σχήμα 5.3: Μορφή του σετ των δεδομένων του ποταμού River Test.

Πίνακας 5.1: Περιγραφή χαρακτηριστικών δεδομένων.

Χαρακτηριστικά	Περιγραφή	Μονάδες
date	η ημερομηνία της μέρας που πραγματοποιήθηκε η μετρούμενη μέτρηση	d
rainfall_{location}	το παρατηρούμενο ημερήσιο ύψος βροχής στη συγκεκριμένη τοποθεσία (location)	mm
tasmax	η μέγιστη ημερήσια παρατηρούμενη θερμοκρασία	°C
tasmin	η ελάχιστη ημερήσια παρατηρούμενη θερμοκρασία	°C
flow_{location}	η ημερήσια παροχή στη συγκεκριμένη τοποθεσία (location)	m <sup>3</sup> /sec



Σχήμα 5.4: Χάρτης με τις τοποθεσίες των σταθμών των μετρήσεων (Πηγή: [85]).

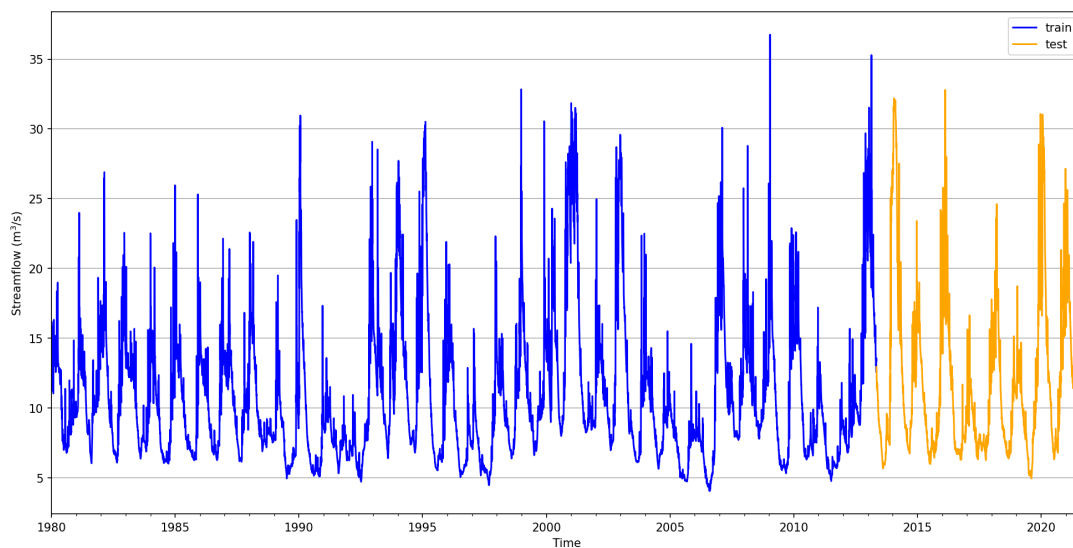


Σχήμα 5.5: Η περιοχή μελέτης και η θέση του σταθμού παροχής Broadlands.

Κατά τη διάρκεια της προεπεξεργασίας των δεδομένων δεν παρατηρήθηκαν λανθασμένες τιμές ή κενά διαστήματα μεταξύ χρονικών περιόδων. Για το πρόβλημα της πρόβλεψης χρονοσειρών, στην παρούσα εργασία επιλέχθηκε η στήλη `flow_Broadlands` για την εφαρμογή των μεθόδων. Πρόκειται για τη μετρούμενη παροχή στη περιοχή Broadlands, με το Σχήμα 5.5 να απεικονίζει

την περιοχή μελέτης και μαζί με τον επιλεγμένο σταθμό. Η διακύμανση της παροχής είναι άμεσα συνυφασμένη με τα πλημμυρικά φαινόμενα, κάτι που κάνει αυτό το χαρακτηριστικό των δεδομένων ιδιαίτερο.

Οι μετρήσεις της παροχής είναι 15 225 τιμές, από 1/1/1980 έως και 6/9/2021. Ο διαχωρισμός train:val:test έγινε με την αναλογία 0.7:0.1:0.2, ένα αρκετά σύνηθες σχήμα διαχωρισμού. Επομένως κατά προσέγγιση τα πρώτα 33 χρόνια χρησιμοποιήθηκαν για την εκπαίδευση του εκάστοτε μοντέλου και τα υπόλοιπα για δοκιμή. Στο Σχήμα 5.6 φαίνεται η χρονοσειρά της παροχής του σταθμού, μαζί με τον διαχωρισμό των δεδομένων εκπαίδευσης (train) και δοκιμής (test).



**Σχήμα 5.6:** Χρονοσειρά παροχής σταθμού Broadlands.

Στόχος της εφαρμογής αποτελεί η πρόβλεψη της χρονοσειράς παροχής του σταθμού Broadlands για τα δεδομένα δοκιμής, με διαφορετικό ορίζοντα πρόβλεψης. Ο ορίζοντας πρόβλεψης επιλέχθηκε να λαμβάνει τις τιμές 2, 10, 20, 40, 60, 80, 100 και 168 μέρες. Στην ανάλυση χρησιμοποιήθηκαν ένα στατιστικό μοντέλο (ARIMA), ένα στοχαστικό μοντέλο (SB) και δύο δίκτυα βαθιάς μάθησης (LSTM και Informer). Η ανάλυση και η δημιουργία του κώδικα πραγματοποιήθηκε στη γλώσσα προγραμματισμού Python αξιοποιώντας σε κάθε περίπτωση τις απαραίτητες βιβλιοθήκες που προσφέρονται. Αξιοποιήθηκε το περιβάλλον του Google Colab στο οποίο δίνεται η πρόσβαση σε υπολογιστικούς πόρους και πιο συγκεκριμένα σε NVIDIA Tesla K80 με 12GB VRAM, δημιουργώντας ιδανικές συνθήκες για εκπαίδευση βαθέων νευρωνικών δικτύων.

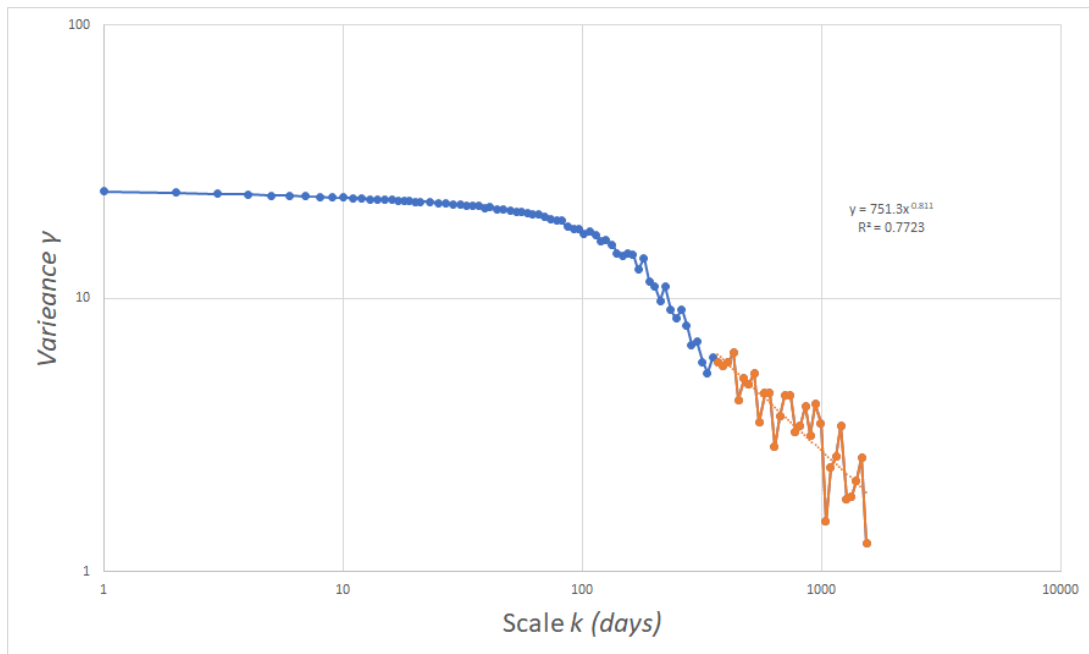
### 5.3 Αρχιτεκτονικές μοντέλων

#### SB

Στην παρούσα εφαρμογή αποφασίστηκε η εφαρμογή του στοχαστικού μοντέλου του υποκεφαλαίου 4.4, το οποίο παίρνει τη θέση ενός Αφελή ή Naive μοντέλου. Η προσέγγιση Naive θεωρεί πως η πρόβλεψη των μελλοντικών τιμών μιας χρονοσειράς θα είναι ίσες με την πιο πρόσφατη

παρατηρούμενη τιμή. Σε αντίθεση με το Naive, το SB μοντέλο υπολογίζει τον ακριβή αριθμό των παρελθοντικών όρων που πρέπει να ληφθούν υπόψιν και να υπολογιστεί μια αντιπροσωπευτική μέση τιμή για τον δεδομένο ορίζοντα πρόβλεψης. Το μοντέλο αυτό χρησιμοποιείται ως βάση για σύγκριση με πιο προηγμένα μοντέλα πρόβλεψης.

Για την εύρεση των απαιτούμενων όρων (στην περίπτωση μας μερών) που θα πρέπει να ληφθούν υπόψιν σε σχέση με το χρονικό ορίζοντα για την πρόβλεψη, είναι αναγκαίος ο υπολογισμός του συντελεστή Hurst  $H$ , ο οποίος προσεγγίζεται από το κλιμακόγραμμα. Στο Σχήμα 5.7 φαίνεται το κλιμακόγραμμα της παροχής του σταθμού Broadlands, δηλαδή της χρονοσειράς του Σχήματος 5.6. Για λόγους απλούστευσης λήφθηκε υπόψη μόνο η μακροχρόνια εξάρτηση (πορτοκαλί χρώμα Σχήματος 5.7) και όχι το πλήρες κλιμακόγραμμα, από όπου προέκυψε ένα συντελεστής  $H = 0.7$ .



**Σχήμα 5.7:** Κλιμακόγραμμα της παροχής στο σταθμό Broadlands.

Αξιοποιώντας την εξίσωση (4.22), για δεδομένο  $H$  και ορίζοντα πρόβλεψης  $\kappa$ , προκύπτει το απαιτούμενο  $\nu$ . Στον Πίνακα 5.2 συνοψίζονται οι συνδυασμοί των  $\kappa$  και  $\nu$ .

**Πίνακας 5.2:** Αναλογία των όρων (σε μέρες) του παρελθόντος  $\nu$  και του μέλλοντος  $\kappa$ .

$\kappa$ (days)	2	10	20	40	60	80	100	168
$\nu$ (days)	64	320	640	1280	1920	2560	3200	5376

---

## ARIMA

Το μοντέλο ARIMA αποτελεί ένα αρκετά διαδεδομένο στατιστικό μοντέλο εδώ και αρκετά χρόνια. Συνδυάζει αυτοπαλίνδρομες (autoregressive - AR), διαφορικές (differencing - I) και κινητού μέσου όρου (moving average - MA) συνιστώσες για να συλλάβει τα υποκείμενα μοτίβα και τη δυναμική σε μια χρονοσειρά δεδομένων. Το μοντέλο περιγράφεται από τρεις κύριες παραμέτρους:

- $p$  - τάξη των αυτοπαλίνδρομων όρων: Η αυτοπαλίνδρομη συνιστώσα (AR) αντιπροσωπεύει τη σχέση μεταξύ μιας παρατήρησης και του αριθμού των καθυστερημένων (lagged) παρατηρήσεων. Η παράμετρος  $p$  καθορίζει τον αριθμό των καθυστερημένων όρων που περιλαμβάνονται στο μοντέλο.
- $d$  - βαθμός διαφοροποίησης: Η συνιστώσα διαφοροποίησης (I) χρησιμοποιείται για να καταστήσει τη χρονοσειρά στάσιμη αφαιρώντας τις τάσεις ή την εποχικότητα. Η παράμετρος  $d$  υποδεικνύει τον αριθμό των διαφορών που πρέπει να γίνουν για να επιτευχθεί η στασιμότητα.
- $q$  - σειρά των όρων του κινητού μέσου όρου: Η συνιστώσα του κινητού μέσου όρου (MA) λαμβάνει υπόψη την επίδραση των παρελθοντικών όρων σφάλματος στην τρέχουσα τιμή.

Συνήθως ο προσδιορισμός του καλύτερου συνόλου παραμέτρων ( $p, d, q$ ) γίνεται με μία μετρική, στόχος της οποίας είναι να βρεθεί ο συνδυασμός των παραμέτρων που παρέχει την καλύτερη ισορροπία (trade-off) μεταξύ της πολυπλοκότητας του μοντέλου και της καλής προσαρμογής στα δεδομένα εκπαίδευσης. Στο μοντέλο ARIMA το ρόλο αυτό παίζει η μετρική AIC (Akaike Information Criterion), όπου συγκρίνει την καλή προσαρμογή διαφορετικών μοντέλων, με τις χαμηλότερες τιμές AIC να δηλώνουν καλύτερη ισορροπία μεταξύ προσαρμογής και πολυπλοκότητας του μοντέλου. Υπολογίζεται βάση της συνάρτησης λογαριθμικής πιθανοφάνειας και μπορεί να γραφτεί ως:

$$AIC = -2\log(L) + 2k \quad (5.1)$$

όπου το  $\log(L)$  είναι η μέγιστη λογαριθμική πιθανοφάνεια και  $k$  είναι ο αριθμός των παραμέτρων του μοντέλου. Στο Σχήμα 5.8 φαίνονται οι τιμές της μετρικής για διάφορες τιμές παραμέτρων. Από την ανάλυση αυτή προέκυψε πως το καλύτερο μοντέλο είναι το ARIMA(3,1,3), ενώ στο Σχήμα 5.9 παρουσιάζεται η προσαρμογή αυτού του μοντέλου στα δεδομένα εκπαίδευσης.



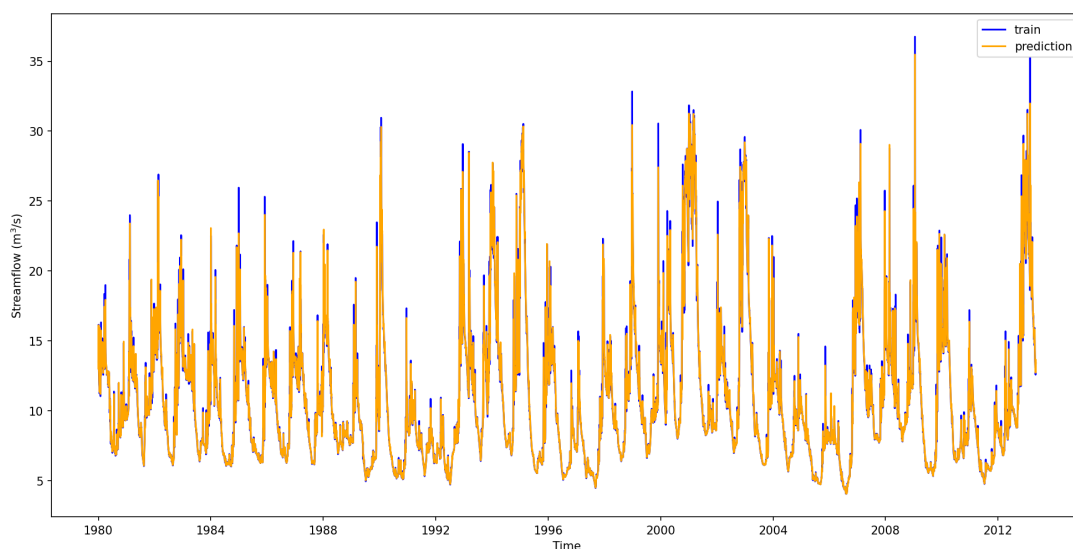
```

Performing stepwise search to minimize aic
ARIMA(2,1,2)(0,0,0)[0] intercept : AIC=34974.465, Time=10.03 sec
ARIMA(0,1,0)(0,0,0)[0] intercept : AIC=36154.416, Time=0.39 sec
ARIMA(1,1,0)(0,0,0)[0] intercept : AIC=36146.237, Time=0.99 sec
ARIMA(0,1,1)(0,0,0)[0] intercept : AIC=36134.299, Time=4.24 sec
ARIMA(0,1,0)(0,0,0)[0] : AIC=36152.416, Time=0.26 sec
ARIMA(1,1,2)(0,0,0)[0] intercept : AIC=34973.954, Time=14.67 sec
ARIMA(0,1,2)(0,0,0)[0] intercept : AIC=35094.962, Time=3.46 sec
ARIMA(1,1,1)(0,0,0)[0] intercept : AIC=35867.694, Time=8.44 sec
ARIMA(1,1,3)(0,0,0)[0] intercept : AIC=34972.633, Time=16.91 sec
ARIMA(0,1,3)(0,0,0)[0] intercept : AIC=34984.359, Time=5.40 sec
ARIMA(2,1,3)(0,0,0)[0] intercept : AIC=34972.381, Time=36.94 sec
ARIMA(3,1,3)(0,0,0)[0] intercept : AIC=34943.929, Time=40.12 sec
ARIMA(3,1,2)(0,0,0)[0] intercept : AIC=35115.820, Time=43.46 sec
ARIMA(4,1,3)(0,0,0)[0] intercept : AIC=34950.416, Time=23.63 sec
ARIMA(3,1,4)(0,0,0)[0] intercept : AIC=34945.769, Time=44.10 sec
ARIMA(2,1,4)(0,0,0)[0] intercept : AIC=34949.619, Time=33.29 sec
ARIMA(4,1,2)(0,0,0)[0] intercept : AIC=34951.542, Time=34.74 sec
ARIMA(4,1,4)(0,0,0)[0] intercept : AIC=34947.920, Time=59.05 sec
ARIMA(3,1,3)(0,0,0)[0] : AIC=34941.929, Time=9.70 sec
ARIMA(2,1,3)(0,0,0)[0] : AIC=34970.380, Time=9.48 sec
ARIMA(3,1,2)(0,0,0)[0] : AIC=34960.318, Time=12.94 sec
ARIMA(4,1,3)(0,0,0)[0] : AIC=34948.416, Time=6.11 sec
ARIMA(3,1,4)(0,0,0)[0] : AIC=34943.769, Time=14.01 sec
ARIMA(2,1,2)(0,0,0)[0] : AIC=34972.465, Time=2.52 sec
ARIMA(2,1,4)(0,0,0)[0] : AIC=34947.619, Time=7.99 sec
ARIMA(4,1,2)(0,0,0)[0] : AIC=34949.542, Time=6.65 sec
ARIMA(4,1,4)(0,0,0)[0] : AIC=34945.920, Time=11.43 sec

```

Best model: ARIMA(3,1,3)(0,0,0)[0]  
Total fit time: 461.031 seconds

**Σχήμα 5.8:** Αποτελέσματα δοκιμών για διαφορετικές παραμέτρους μοντέλου ARIMA.



**Σχήμα 5.9:** Απόδοση του μοντέλου ARIMA στα δεδομένα εκπαίδευσης.

## LSTM

Στη συνέχεια αναπτύχθηκε το μοντέλο LSTM, η υλοποίηση του οποίου πραγματοποιήθηκε με τη χρήση της βιβλιοθήκης PyTorch, μία από τις πιο διαδεδομένες και ευρέως χρησιμοποιούμενες βιβλιοθήκες για βαθιά μάθηση. Για την εκπαίδευση χρησιμοποιείται η χρονοσειρά και δημιουργούνται ακολουθίες σταθερού μήκους, αφού πρώτα έχουν κανονικοποιηθεί. Κατά την καταγραφή αυτών των ακολουθιών, καταγράφεται επίσης και η τιμή που προέκυψε αμέσως μετά την εν λόγω ακολουθία. Κάθε ακολουθία τροφοδοτείται στο μοντέλο LSTM, μία κάθε φορά, και προχωράει με αυτόν τον τρόπο η διαδικασία της εκπαίδευσης. Έπειτα από δοκιμές και συνδυασμό αρχιτεκτονικών, οι βέλτιστες υπερπαραμέτροι (hyperparameters) και η αρχιτεκτονική του δικτύου που χρησιμοποιήθηκε τελικά, διακρίνονται στον Πίνακα 5.3.

Πίνακας 5.3: Υπερπαραμέτροι μοντέλου LSTM.

Υπερπαραμέτροι	Hyperparameters	Τιμές
Νευρώνες	Neurons	100
Πυκνά Επίπεδα	Dense Layers (Fully Connected)	5
Απόρριψη	Dropout	0.2
Μέγεθος δέσμης	Batch size	16
Μήκος ακολουθίας	Sequence length	180
Αλγόριθμος βελτιστοποίησης	Optimizer	Adam
Συνάρτηση ενεργοποίησης	Activation function	ReLU
Συνάρτησης απώλειας	Loss function	mse
Ρυθμός μάθησης	Learning rate	$5 \cdot 10^{-4}$
Εποχές	Epochs	20

## Informer

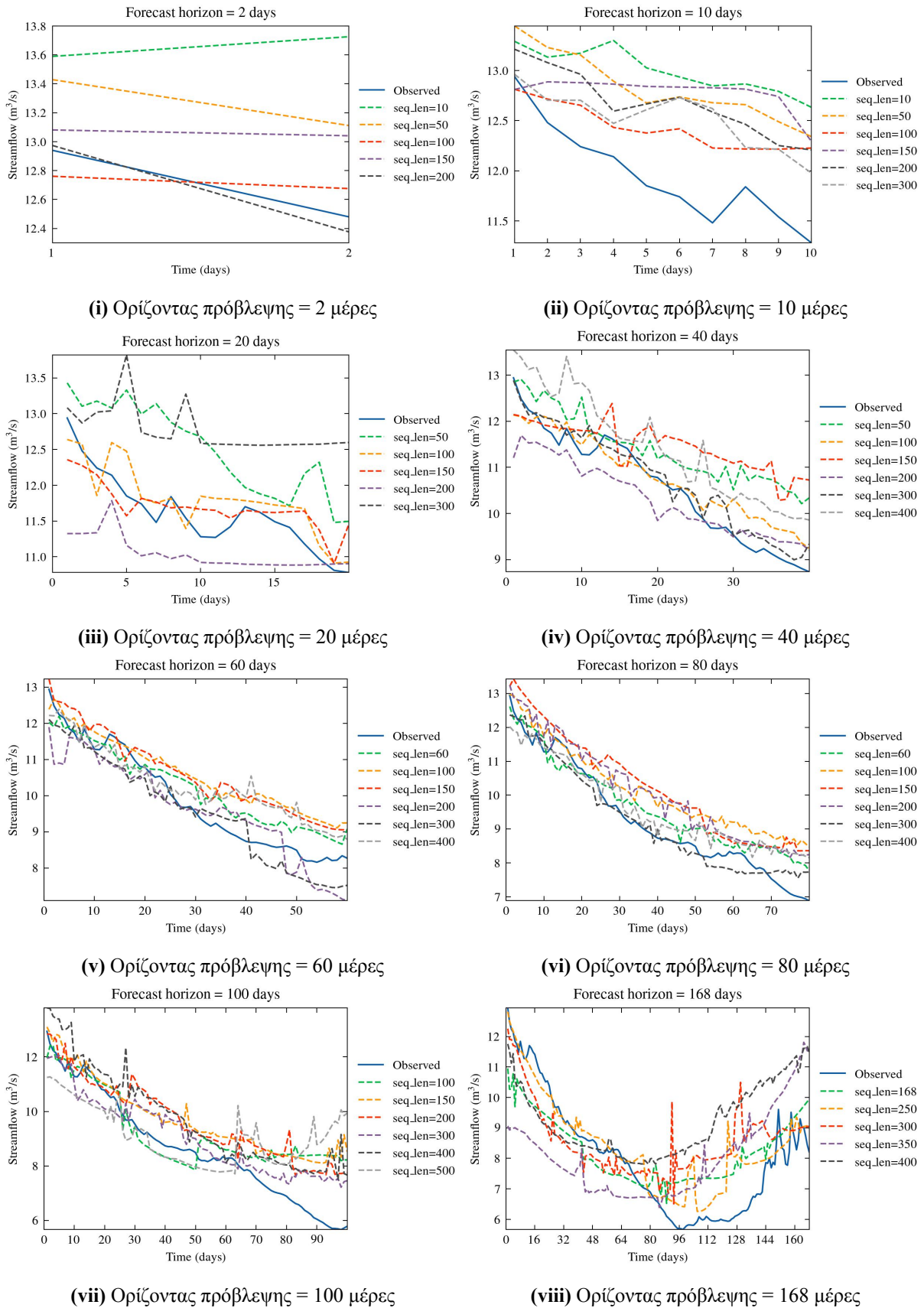
Το τελευταίο μοντέλο που αναλύθηκε στην παρούσα εργασία είναι το transformer-based μοντέλο Informer. Για την εκπαίδευση, αφού χωρίζονται τα δεδομένα σε δεδομένα εκπαίδευσης, επαλήθευσης και δοκιμής, κανονικοποιούνται και χωρίζονται σε υποακολουθίες εισόδου, στόχου. Έπειτα η είσοδος τροφοδοτείται στον κωδικοποιητή από όπου συνεχίζει στον αποκωδικοποιητή και εκεί παράγονται οι προβλεπόμενες μελλοντικές τιμές της χρονοσειράς. Η αρχιτεκτονική του μοντέλου που ακολουθήθηκε διακρίνεται στον Πίνακα 5.4, χωρίς να φαίνεται το μήκος ακολουθίας εισόδου του κωδικοποιητή και το μήκος ακολουθίας πρόβλεψης. Έπειτα από μια πειραματική διαδικασία στοχευμένης προσαρμογής (fine-tuning), με αρκετούς συνδυασμούς μεταξύ των ακολουθιών εισόδου και πρόβλεψης, προέκυψαν οι βέλτιστες τιμές των συνδυασμών που δίνονται στον Πίνακα 5.5. Ακόμη στο Σχήμα 5.10 παρουσιάζονται οι αποδόσεις για διάφορες τιμές της ακολουθίας εισόδου σε δεδομένο ορίζοντα πρόβλεψης του μοντέλου Informer.

**Πίνακας 5.4:** Υπερπαράμετροι μοντέλου Informer.

Υπερπαράμετροι	Hyperparameters	Τιμές
Νευρώνες	Encoder input size	2
Νευρώνες	Decoder input size	2
Νευρώνες	Dimension of model	512
Αριθμός παράλληλων κεφαλών	Number of heads	8
Νευρώνες	Encoder layers	512
Νευρώνες	Decoder layers	512
Απόρριψη	Dropout	0.05
Προσοχή	Attention	ProbSparse
Συνάρτηση ενεργοποίησης	Activation function	GELU
Μέγεθος δέσμης	Batch size	32
Αλγόριθμος βελτιστοποίησης	Optimizer	Adam
Συνάρτησης απώλειας	Loss function	mse
Ρυθμός μάθησης	Learning rate	$10^{-4}$
Εποχές	Epochs	8
Υπομονή για πρόωρη διακοπή	Early stopping patience	3

**Πίνακας 5.5:** Συνδυασμός ακολουθιών εισόδου (sequence length) και πρόβλεψης (prediction length) μοντέλου Informer.

prediciton length (days)	2	10	20	40	60	80	100	168
sequence length (days)	200	100	150	300	300	300	300	250



**Σχήμα 5.10:** Αποτελέσματα πειραματικής διαδικασίας για συνδυασμούς ακολουθιών εισόδου και ορίζοντα πρόβλεψης μοντέλου Informer.

---

## 5.4 Μετρικές αξιολόγησης

Η παρούσα εργασία εξέτασε ένα ευρύ πεδίο μετρικών για την αξιολόγηση της απόδοσης των μοντέλων. Για την ακρίβεια υπολογίστηκαν τα ακόλουθα σφάλματα, τα οποία περιγράφονται αναλυτικότερα στη συνέχεια [89]:

- Μέσο τετραγωνικό σφάλμα (MSE)
- Ρίζα μέσου τετραγωνικού σφάλματος (RMSE)
- Μέσο απόλυτο σφάλμα (MAE)
- Μέσο απόλυτο ποσοστιαίο σφάλμα (MAPE)

### MSE

Το Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error - MSE) είναι μια κοινή μετρική που χρησιμοποιείται για την αξιολόγηση της ακρίβειας των μοντέλων πρόβλεψης. Υπολογίζει τη μέση τετραγωνική διαφορά μεταξύ των προβλεπόμενων τιμών και των πραγματικών τιμών σε μια εργασία πρόβλεψης. Πρόκειται για μια θετική τιμή που αυξάνεται και η τιμή του καθώς αυξάνεται το σφάλμα του μοντέλου. Η μαθηματική σχέση που περιγράφει το σφάλμα MSE φαίνεται στην εξίσωση (5.2). Είναι σημαντικό να σημειωθεί ότι το MSE δίνει ίση βαρύτητα τόσο στις υπερεκτιμήσεις (overestimations) όσο και στις υποεκτιμήσεις (underestimations), γεγονός που μπορεί να μην είναι πάντα επιθυμητό σε ορισμένες εφαρμογές πρόβλεψης. Οι μονάδες μέτρησης είναι οι μονάδες του μεγέθους προς εκτίμηση υψωμένο στο τετράγωνο.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5.2)$$

όπου  $y_i$  είναι η  $i$ -οστή παρατηρούμενη τιμή,  $\hat{y}_i$  είναι η αντίστοιχη τιμή πρόβλεψης και  $n$  ο αριθμός των παρατηρήσεων.

### RMSE

Η Ρίζα Μέσου Τετραγωνικού Σφάλματος (Root Mean Squared Error - RMSE) προκύπτει από μέσο τετραγωνικό σφάλμα αν λάβουμε την τετραγωνική ρίζα του, όπως προκύπτει και στην εξίσωση (5.3). Παρέχει ένα μέτρο του μέσου μεγέθους των σφαλμάτων του μοντέλου πρόβλεψης. Το RMSE εκφράζεται στις ίδιες μονάδες με τα αρχικά δεδομένα.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5.3)$$

## MAE

Το Μέσο Απόλυτο Σφάλμα (Mean Absolute Error - MAE) μετρά τη μέση απόλυτη διαφορά μεταξύ των τιμών πρόβλεψης και των πραγματικών τιμών, χωρίς να λαμβάνει υπόψη την κατεύθυνσή τους. Σε αντίθεση με το MSE και RMSE, το σφάλμα MAE δεν τιμωρεί σημαντικά τα μεγάλα σφάλματα καθώς υπολογίζει την απόλυτη διαφορά και όχι το τετράγωνό της. Ακόμη έχει μεγάλη εξάρτηση από την κλίμακα των δεδομένων, όπως και οι προηγούμενες δύο μετρικές. Μικρό MAE σημαίνει καλύτερη απόδοση του μοντέλου. Εκφράζεται στις ίδιες μονάδες με τα αρχικά δεδομένα και η σχέση υπολογισμού του δίνεται στην εξίσωση (5.4).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5.4)$$

## MAPE

Το Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Mean Absolute Percentage Error - MAPE) είναι η ποσοστιαία διαφορά μεταξύ της πραγματικής τιμής και της τιμής πρόβλεψης. Συχνά αποτελεί τη βασική μετρική για τη μέτρηση των περισσότερων μοντέλων πρόβλεψης, καθώς είναι ανεξάρτητη της κλίμακας και μπορεί να γίνει σύγκριση μεταξύ διαφορετικών χρονοσειρών. Ουσιαστικά πρόκειται για το σφάλμα MAE εκφρασμένο σε ποσοστό για την εύκολη κατανόηση και ερμηνεία του. Είναι ιδιαίτερα ευαίσθητο σε μηδενικές ή κοντά στο μηδέν πραγματικές τιμές, καθώς η διαίρεση με τέτοιες τιμές μπορεί να οδηγήσει σε απροσδιόριστες ή άπειρες τιμές. Ορίζεται ως εξής:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \cdot 100\% \quad (5.5)$$

## 5.5 Αποτελέσματα και σύγκριση μοντέλων

Όπως προαναφέρθηκε ο χρονικός ορίζοντας για την πρόβλεψη κυμαίνεται από 2 μέχρι και 168 μέρες. Στον Πίνακα 5.6 φαίνονται αναλυτικά η τιμές όλων των μοντέλων ως προς τις τέσσερις μετρικές MSE, RMSE, MAE και MAPE για κάθε ένα από τους χρονικούς ορίζοντες πρόβλεψης. Επισημαίνεται πως η μετρική MAPE δεν εμφανίζεται ως ποσοστό. Σε κάθε ορίζοντα πρόβλεψης τα καλύτερα σκορ τονίζονται με έντονη γραφή (bold). Επιπλέον στα Σχήματα 5.12-5.19 φαίνονται τα διαγράμματα προβλεπόμενων και πραγματικών τιμών.

Από τον Πίνακα 5.6 μπορούν να εξαχθούν κάποια συμπεράσματα. Τα μοντέλα βαθιάς μάθησης δείχνουν να πηγαίνουν αρκετά καλά σε σύγκριση με το naïve στοχαστικό και το στατιστικό μοντέλο ARIMA, που αδυνατεί να ακολουθήσει τα σύνθετα μοτίβα της χρονοσειράς. Το LSTM δίνει λίγο καλύτερα αποτελέσματα στο χρονικό ορίζοντα 10 ημερών, ενώ ο Informer φαίνεται να έχει την καλύτερη απόδοση γενικότερα. Το μοντέλο Informer βελτιώνει σημαντικά τα αποτελέσματα και το σφάλμα πρόβλεψης αυξάνεται ομαλά και αργά όσο αυξάνεται ο ορίζοντας

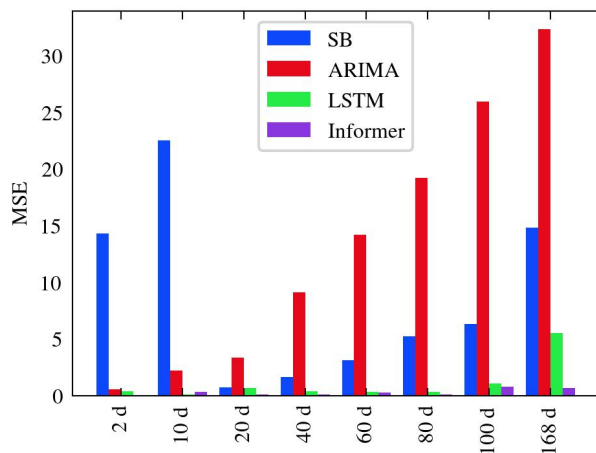
πρόβλεψης, γεγονός που αποδεικνύει την επιτυχία του Informer στην ενίσχυση της ικανότητας πρόβλεψης στο πρόβλημα πρόβλεψης σε μακρινό ορίζοντα (LSTF). Η απόδοση του Informer σε σχέση τη μετρική MAPE δεν ξεπερνά το 10%, κάτι που δηλώνει την πολύ καλή ακρίβειά του. Όπως φαίνεται και στο Σχήμα 5.20, όπου παρουσιάζεται γραφικά το σφάλμα (MSE) των μοντέλων συναρτήσει του μήκους περιόδου πρόβλεψης, το σφάλμα αυξάνεται όσο μεγαλύτερο είναι το μήκος πρόβλεψης στα περισσότερα μοντέλα. Αξίζει να σημειωθεί ότι κάτι τέτοιο δεν είναι τόσο εμφανές στο μοντέλο Informer, που φαίνεται να έχει μια πιο σταθερή μεταβλητότητα του σφάλματος ανεξάρτητα από τον χρονικό ορίζοντα πρόβλεψης.

Στο Σχήμα 5.21 μπορεί να διακριθεί το σφάλμα (MSE) των μοντέλων υπολογισμένα σε κάθε χρονικό βήμα για το χρονικό ορίζοντα των 100 ημερών, δηλαδή το σφάλμα της πρώτης μέρας, της δεύτερης μέρας κ.ο.κ. Κοινή συμπεριφορά των μοντέλων είναι πως το σφάλμα αυξάνεται με την πάροδο των ημερών, με το μοντέλο Informer να πετυχαίνει τη λιγότερη αύξηση αυτού του σφάλματος.

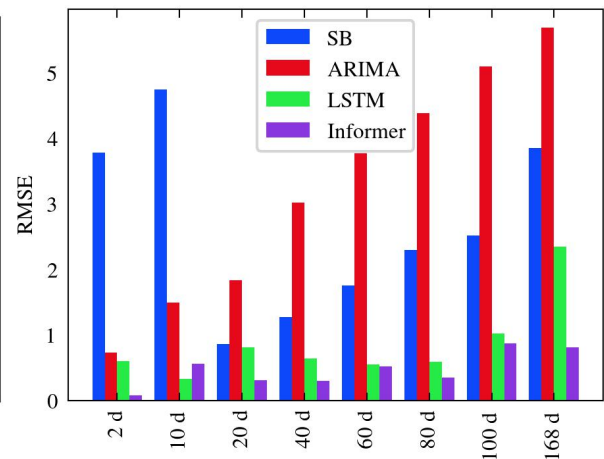
**Πίνακας 5.6:** Αποτελέσματα μοντέλων πρόβλεψης χρονοσειράς παροχών River Test.

Method	Metric	2	10	20	40	60	80	100	168
SB	MSE	14.302	22.519	0.748	1.627	3.091	5.252	6.314	14.798
	RMSE	3.782	4.745	0.865	1.276	1.758	2.292	2.513	3.847
	MAE	3.775	4.721	0.767	1.062	1.527	1.994	2.179	3.459
	MAPE	0.297	0.397	0.068	0.107	0.168	0.240	0.296	0.502
ARIMA	MSE	0.527	2.224	3.356	9.092	14.197	19.196	25.933	32.290
	RMSE	0.726	1.491	1.832	3.015	3.768	4.381	5.092	5.682
	MAE	0.689	1.417	1.756	2.781	3.499	4.082	4.718	5.376
	MAPE	0.055	0.120	0.154	0.279	0.381	0.481	0.625	0.758
LSTM	MSE	0.362	<b>0.105</b>	0.656	0.407	0.304	0.349	1.047	5.520
	RMSE	0.601	<b>0.325</b>	0.810	0.638	0.551	0.591	1.023	2.349
	MAE	0.589	<b>0.261</b>	0.637	0.482	<b>0.413</b>	0.452	0.746	1.802
	MAPE	0.046	<b>0.021</b>	0.056	0.045	<b>0.041</b>	0.050	0.104	0.255
Informer	MSE	<b>0.006</b>	0.311	<b>0.094</b>	<b>0.091</b>	<b>0.269</b>	<b>0.124</b>	<b>0.760</b>	<b>0.648</b>
	RMSE	<b>0.077</b>	0.558	<b>0.307</b>	<b>0.301</b>	<b>0.519</b>	<b>0.352</b>	<b>0.872</b>	<b>0.805</b>
	MAE	<b>0.069</b>	0.502	<b>0.251</b>	<b>0.241</b>	0.465	<b>0.278</b>	<b>0.720</b>	<b>0.564</b>
	MAPE	<b>0.005</b>	0.043	<b>0.022</b>	<b>0.024</b>	0.049	<b>0.032</b>	<b>0.095</b>	<b>0.083</b>

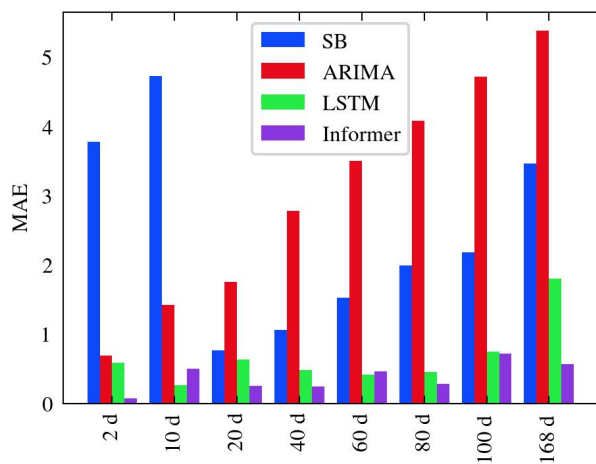
Μια γενικότερη και πιο άμεση εποπτεία του Πίνακα 5.6 απεικονίζεται στο Σχήμα 5.11 σε μορφή ραβδογράμματος ανά μετρική αξιολόγησης.



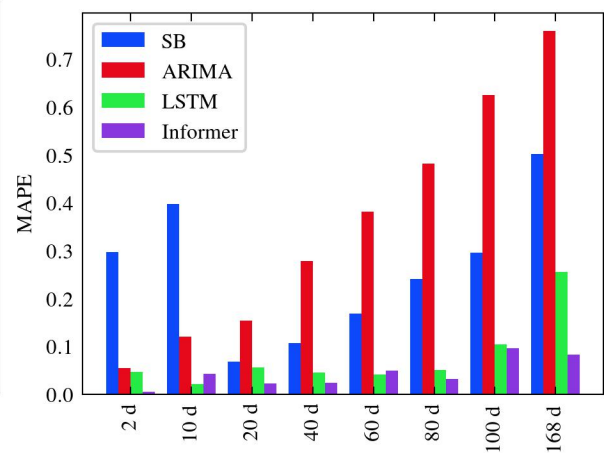
(i) Μετρική αξιολόγησης MSE.



(ii) Μετρική αξιολόγησης RMSE.



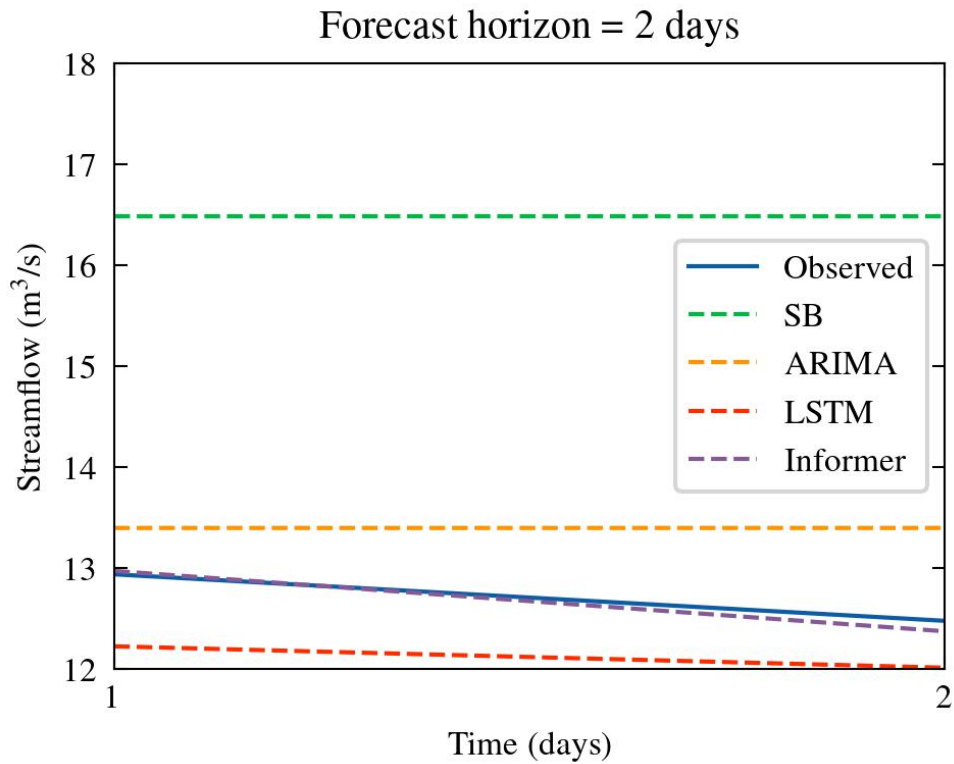
(iii) Μετρική αξιολόγησης MAE.



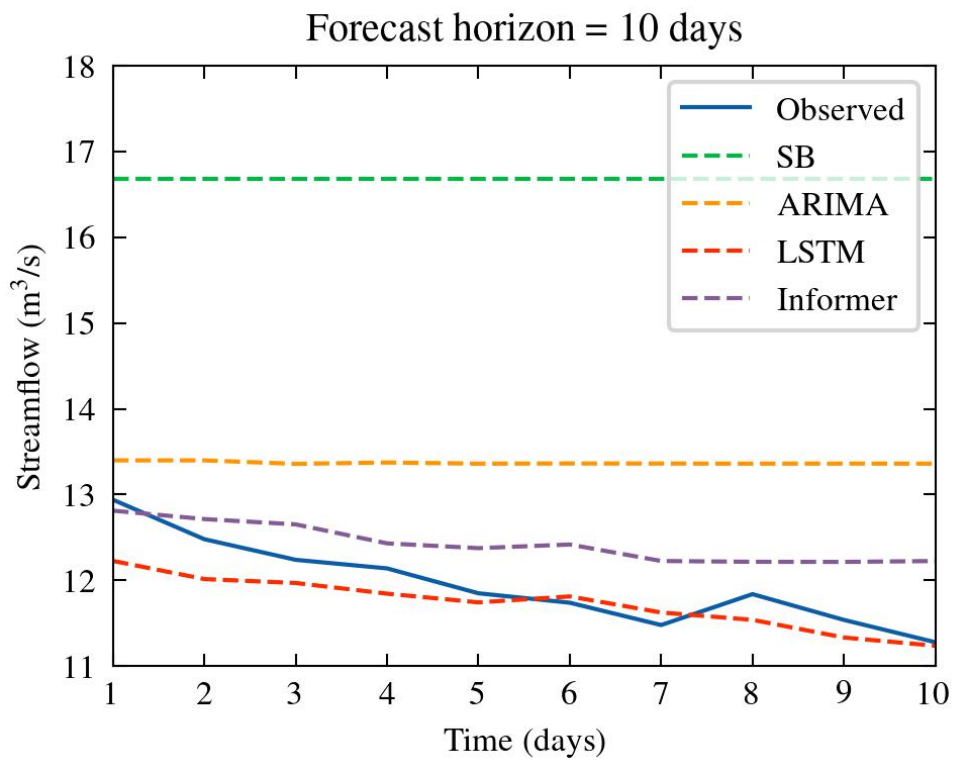
(iv) Μετρική αξιολόγησης MAPE.

Σχήμα 5.11: Αποτελέσματα μετρικών αξιολόγησης σε μορφή ραβδογράμματος.

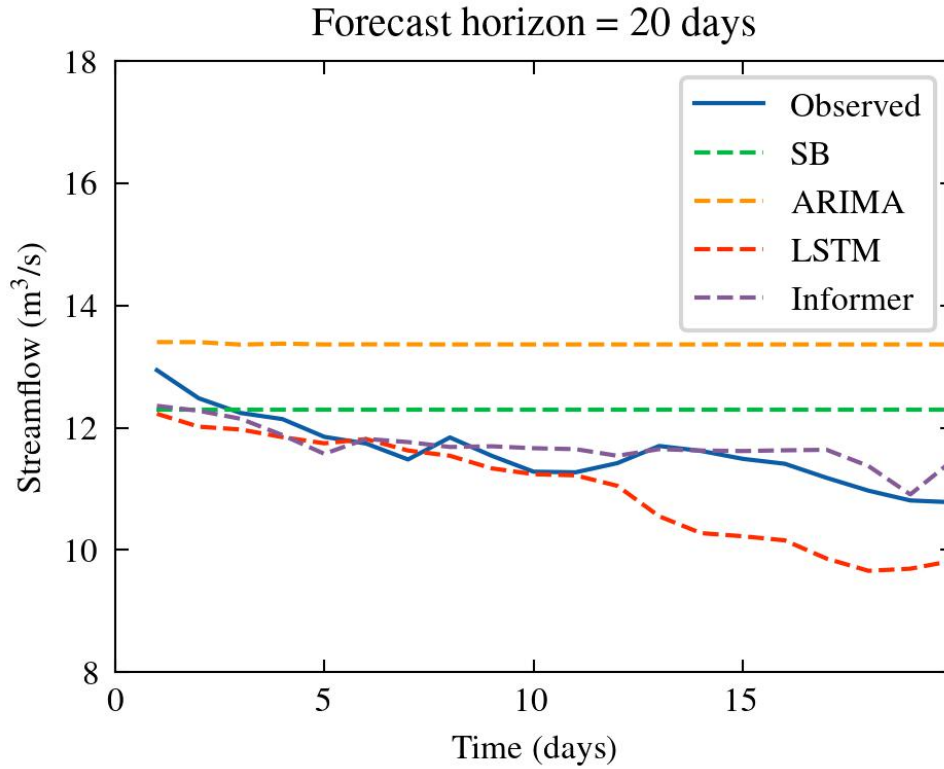




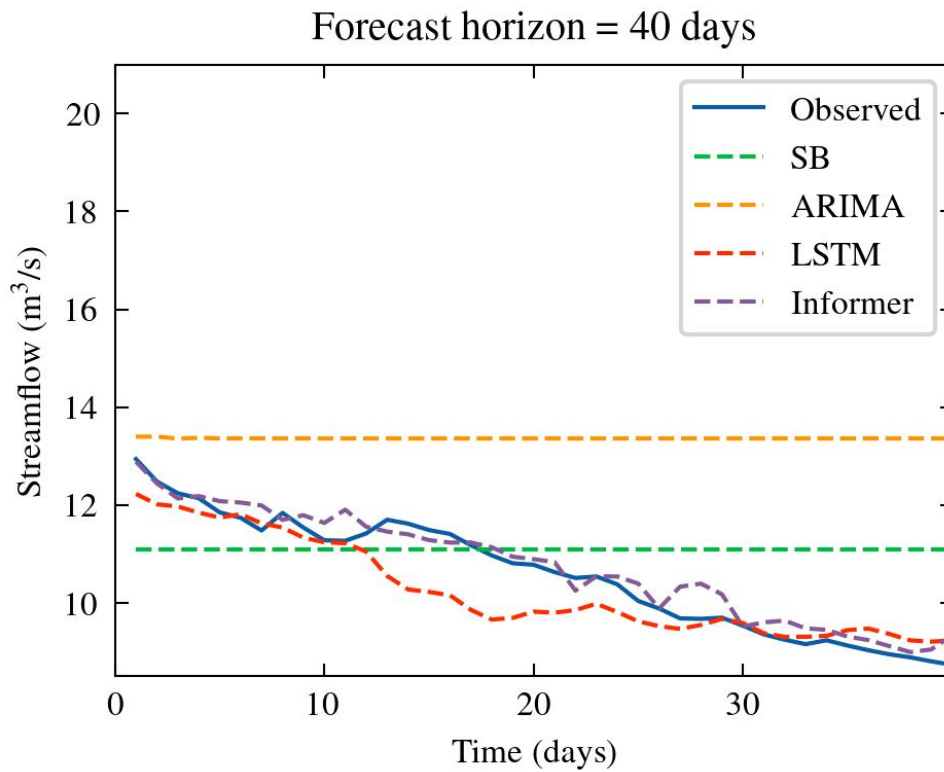
**Σχήμα 5.12:** Σύγκριση προβλεπόμενων τιμών μοντέλων και πραγματικών τιμών για ορίζοντα πρόβλεψης 2 ημερών.



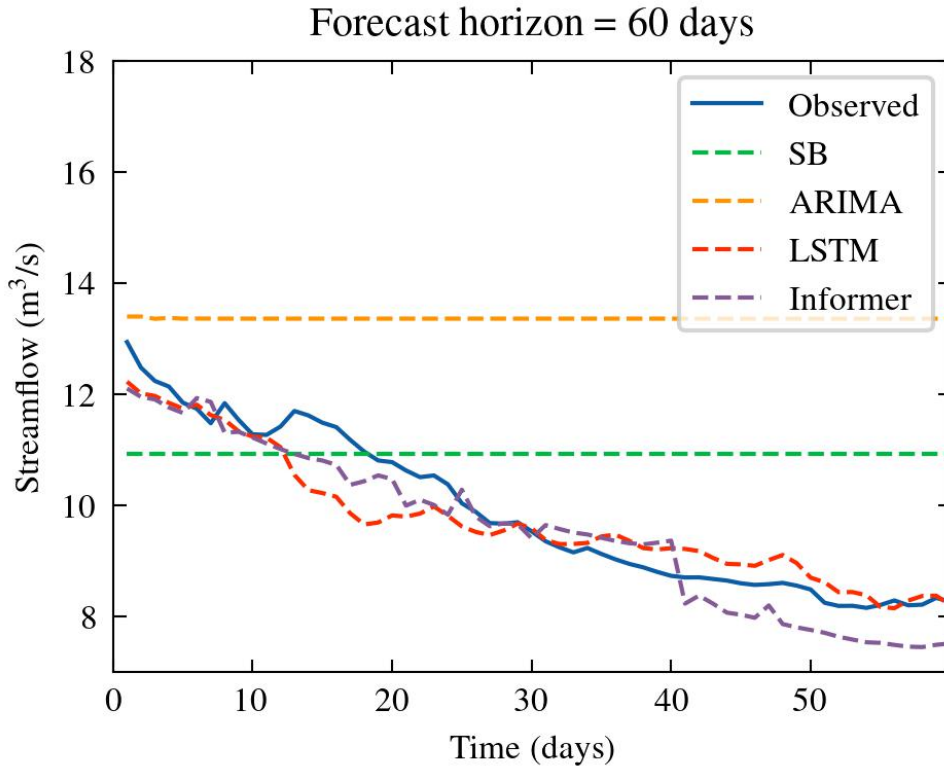
**Σχήμα 5.13:** Σύγκριση προβλεπόμενων τιμών μοντέλων και πραγματικών τιμών για ορίζοντα πρόβλεψης 10 ημερών.



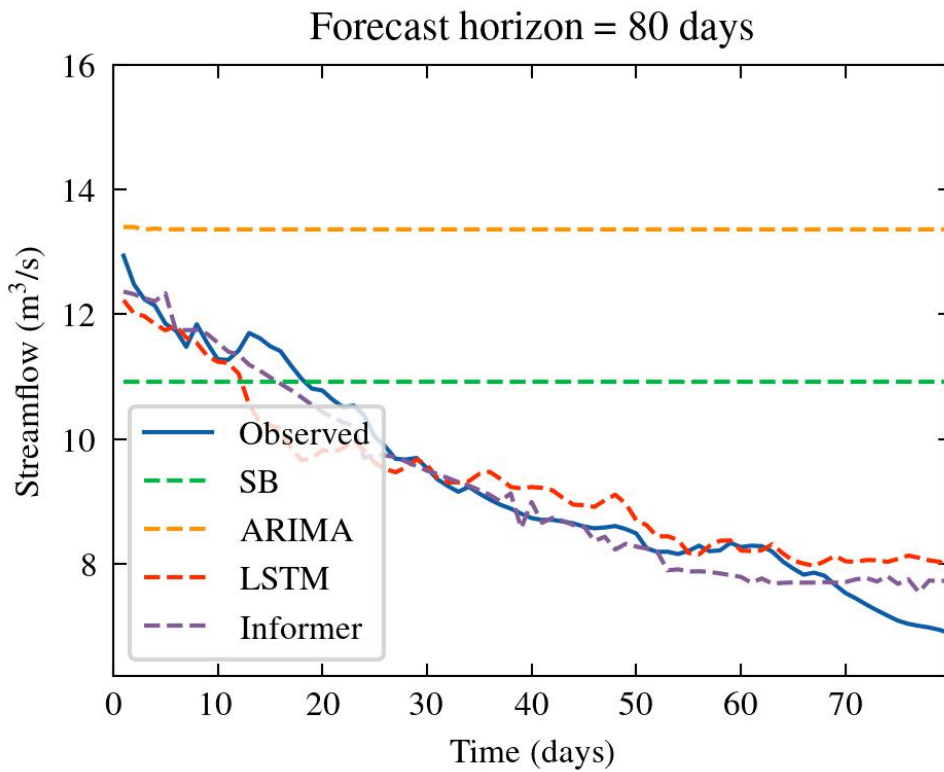
**Σχήμα 5.14:** Σύγκριση προβλεπόμενων τιμών μοντέλων και πραγματικών τιμών για ορίζοντα πρόβλεψης 20 ημερών.



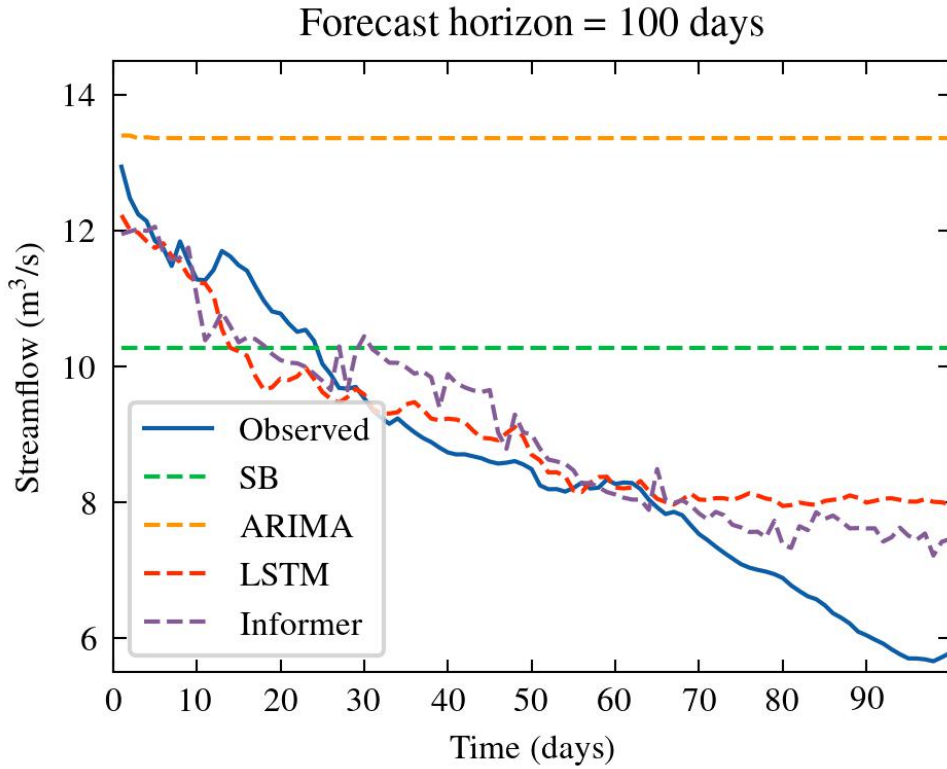
**Σχήμα 5.15:** Σύγκριση προβλεπόμενων τιμών μοντέλων και πραγματικών τιμών για ορίζοντα πρόβλεψης 40 ημερών.



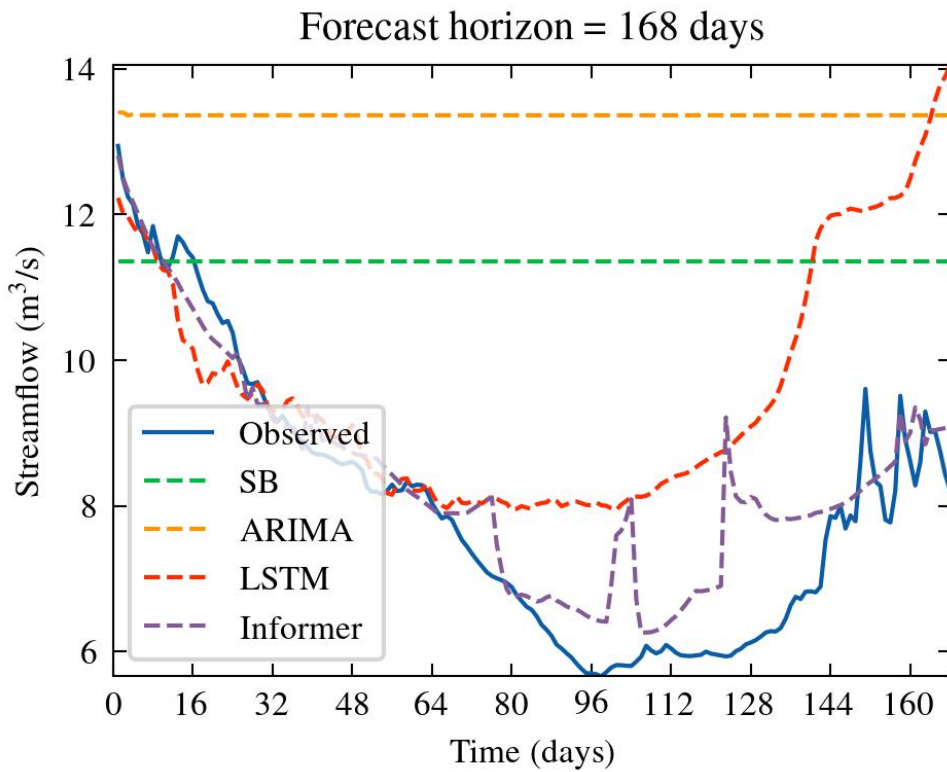
**Σχήμα 5.16:** Σύγκριση προβλεπόμενων τιμών μοντέλων και πραγματικών τιμών για ορίζοντα πρόβλεψης 60 ημερών.



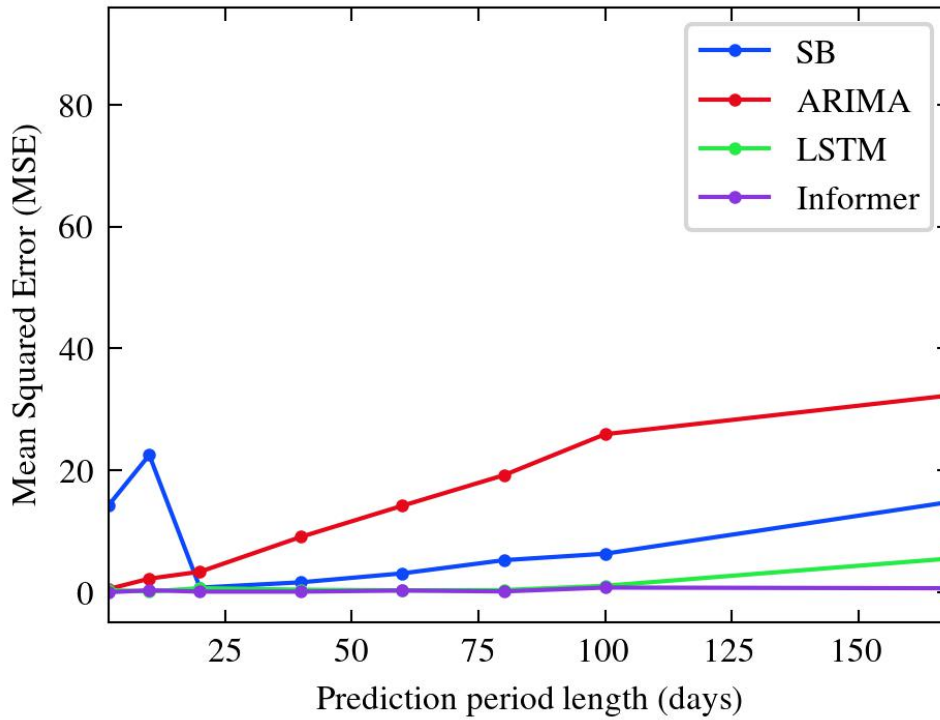
**Σχήμα 5.17:** Σύγκριση προβλεπόμενων τιμών μοντέλων και πραγματικών τιμών για ορίζοντα πρόβλεψης 80 ημερών.



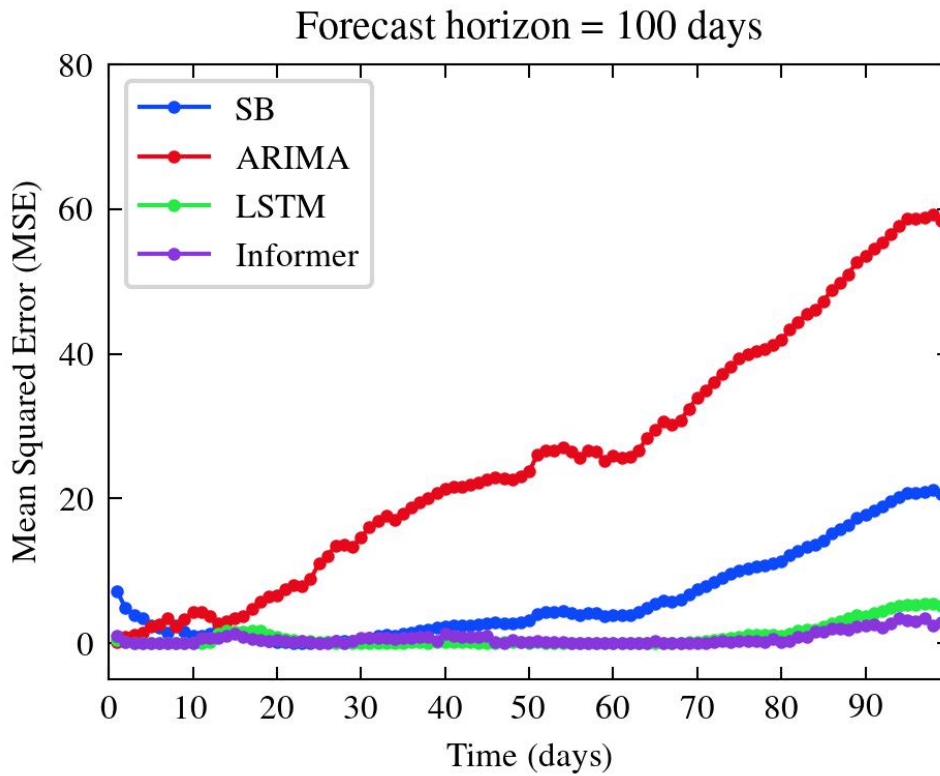
**Σχήμα 5.18:** Σύγκριση προβλεπόμενων τιμών μοντέλων και πραγματικών τιμών για ορίζοντα πρόβλεψης 100 ημερών.



**Σχήμα 5.19:** Σύγκριση προβλεπόμενων τιμών μοντέλων και πραγματικών τιμών για ορίζοντα πρόβλεψης 168 ημερών.

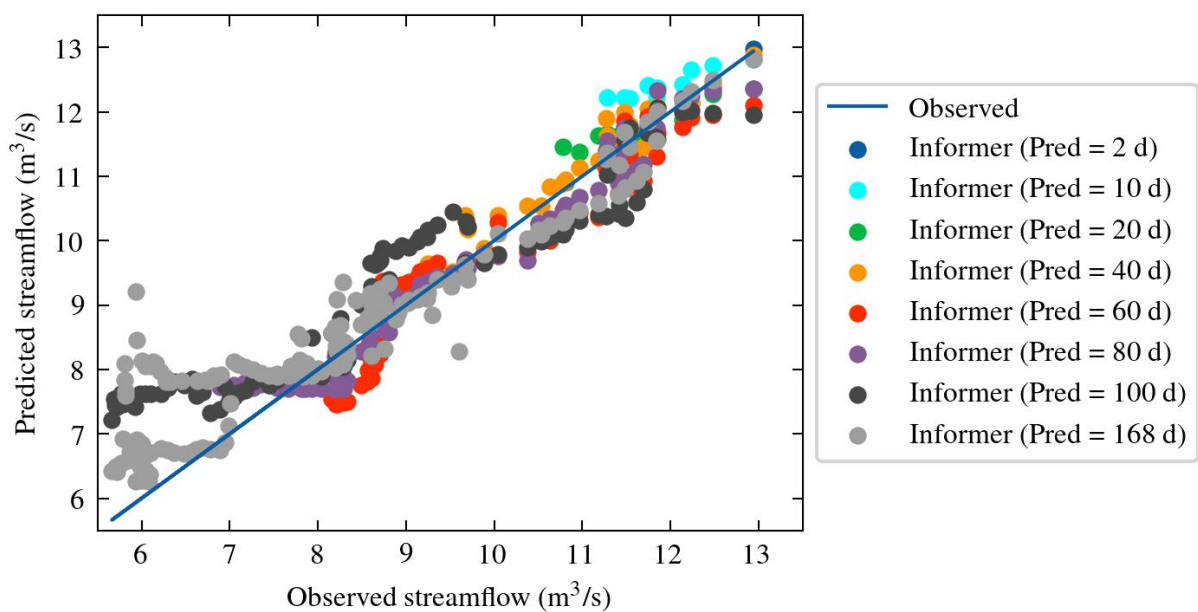


**Σχήμα 5.20:** Διάγραμμα σφάλματος (MSE) για κάθε μοντέλο συναρτήσει του χρονικού ορίζοντα πρόβλεψης.



**Σχήμα 5.21:** Διάγραμμα σφάλματος (MSE) μοντέλων υπολογισμένα σε κάθε χρονικό βήμα για χρονικό ορίζοντα πρόβλεψης 100 ημερών.

Από τη σύγκριση των μοντέλων εκείνο που φαίνεται να έχει την καλύτερη συμπεριφορά είναι το μοντέλο Informer. Το Σχήμα 5.22 συνοψίζει τα αποτελέσματα του Informer σε ένα διάγραμμα διασποράς παρατηρούμενων και τιμών πρόβλεψης παροχής. Η μπλε συνεχής γραμμή του διαγράμματος αντιστοιχεί στην βέλτιστη συμπεριφορά, αν κάθε πρόβλεψη του μοντέλου ταυτιζόταν με την πραγματική παρατηρούμενη τιμή. Η γενική συμπεριφορά είναι καλή με τη γραμμή τάσης να έχει συντελεστή προσδιορισμού (coefficient of determination)  $R^2 = 0.91$ . Η χειρότερη απόδοση εμφανίζεται στις χαμηλές τιμές, που στην περίπτωση μας ανταποκρίνονται σε το τοπικό ακρότατο (ελάχιστο). Ολοκληρώνοντας είναι ιδιαίτερα σημαντικό να τονιστεί πως ο Informer υποστηρίζει και πολυμεταβλητή πρόβλεψη, πράγμα που σημαίνει πως θα μπορούσαν να αξιοποιηθούν και τα υπόλοιπα μετρούμενα μεγέθη του σετ δεδομένων βελτιώνοντας την απόδοση, αλλά εδώ δόθηκε έμφαση στη μονομεταβλητή διαδικασία πρόβλεψης.



**Σχήμα 5.22:** Διάγραμμα διασποράς προβλεπόμενων τιμών μοντέλου Informer.

---

## Κεφάλαιο 6

# Διερεύνηση μοντέλου Informer σε μεγάλο πλήθος ποταμών

### 6.1 Περιοχή μελέτης

Η περιοχή μελέτης που επιλέχθηκε για την διερεύνηση της απόδοσης του μοντέλου Informer σε ένα μεγάλο πλήθος χρονοσειρών (>100) είναι η ευρύτερη περιοχή του Ηνωμένου Βασιλείου, όπου ανήκει και ο ποταμός της προηγούμενης εφαρμογής.

Οι πλημμύρες αποτελούν σημαντική απειλή για πολλές περιοχές παγκοσμίως, και το Ηνωμένο Βασίλειο δεν αποτελεί εξαίρεση. Με την ποικιλόμορφη τοπογραφία της και τα εκτεταμένα δίκτυα ποταμών το Ηνωμένο Βασίλειο βιώνει μια σειρά από πλημμυρικά φαινόμενα που έχουν σημαντικές επιπτώσεις. Η μελέτη αυτή εστιάζει στους ποταμούς του Ηνωμένου Βασιλείου και στοχεύει στην ανάπτυξη και αξιολόγηση μιας προσέγγισης που βασίζεται στη βαθιά μάθηση για την πρόβλεψη πλημμυρικών συμβάντων και πιο συγκεκριμένα στο μοντέλο Informer. Ένας πολύ σημαντικός λόγος επιλογής της συγκεκριμένης περιοχής μελέτης είναι πως το Ηνωμένο Βασίλειο διαθέτει μια ολοκληρωμένη βάση δεδομένων που συλλέγει και διατηρεί δεδομένα ροής ποταμών από διάφορους μετρητικούς σταθμούς σε ολόκληρη τη χώρα. Πρόκειται για τη βάση του National River Flow Archive (NRFA) [88] η οποία και αξιοποιείται στη παρούσα εφαρμογή. Ο χάρτης του Σχήματος 6.1 απεικονίζει τη πληθώρα σταθμών της βάσης NRFA όπως επίσης και των ποταμών.

### 6.2 Περιγραφή δεδομένων

Σύμφωνα με όσα προαναφέρθηκαν τα δεδομένα λήφθηκαν από τη βάση NRFA [88] και είναι διαθέσιμα δωρεάν. Όπως και στην προηγούμενη εφαρμογή, η ανάλυση αφορά τα δεδομένα χρονοσειρών παροχών ποταμών με σκοπό την πρόβλεψη της χρονοσειράς για ένα συγκεκρι-





**Σχήμα 6.1:** Χάρτης με της τοποθεσίες των διαθέσιμων σταθμών στο Ηνωμένο Βασίλειο της βάσης NRFA (Πηγή: [90]).

μένο χρονικό ορίζοντα. Η αναζήτηση μπορεί να γίνει πολύ εύκολα από τη βάση αυτή [90] αξιοποιώντας και φιλτράροντας μέσω διάφορων κριτηρίων που ενδέχεται να θέλει κάποιος για τη δικιά του εργασία (π.χ. η ύπαρξη δεδομένων σε πραγματικό χρόνο).

Στη συγκεκριμένη εφαρμογή η επιλογή έγινε με βάση κυρίως τριών κριτηρίων:

1. Η ύπαρξη ημερήσιων δεδομένων παροχών.
2. Το μήκος της χρονοσειράς να είναι μεγάλο (> 20 ετών).
3. Η ύπαρξη συνεχών δεδομένων χωρίς κενά διαστήματα (ποσοστό πληρότητας > 95%)

Από την ανάλυση αυτή προέκυψε ένα σύνολο 112 σταθμών δεδομένων των οποίων τα χαρακτηριστικά μεταδεδωμένα των πρώτων 20 σταθμών παρουσιάζονται στον Πίνακα 6.1, ενώ οι γεωγραφικές τους θέσεις φαίνονται στο Σχήμα 6.2. Τα δεδομένα αφορούν την ημερήσια παροχή στους ποταμούς. Στον πίνακα διακρίνεται και η ασυμμετρία της χρονοσειράς, που υπολογίστηκε για κάθε σταθμό. Ο λόγος υπολογισμού της ασυμμετρίας είναι για τη διερεύνηση της σχέσης που ενδεχομένως μπορεί να έχει ένα χαρακτηριστικό μέγεθος της χρονοσειράς με την απόδοση του μοντέλου. Στο πλαίσιο των δεδομένων χρονοσειρών ο όρος αναφέρεται στην ασυμμετρία της κατανομής των σημείων των δεδομένων. Παρέχει πληροφορίες σχετικά με το σχήμα της κατανομής και το βαθμό στον οποίο αποκλίνει από τη συμμετρία. Ένας συνηθισμένος τρόπος υπολογισμού της ασυμμετρίας (skewness) είναι με το συντελεστή Fisher-Pearson, που δίνεται

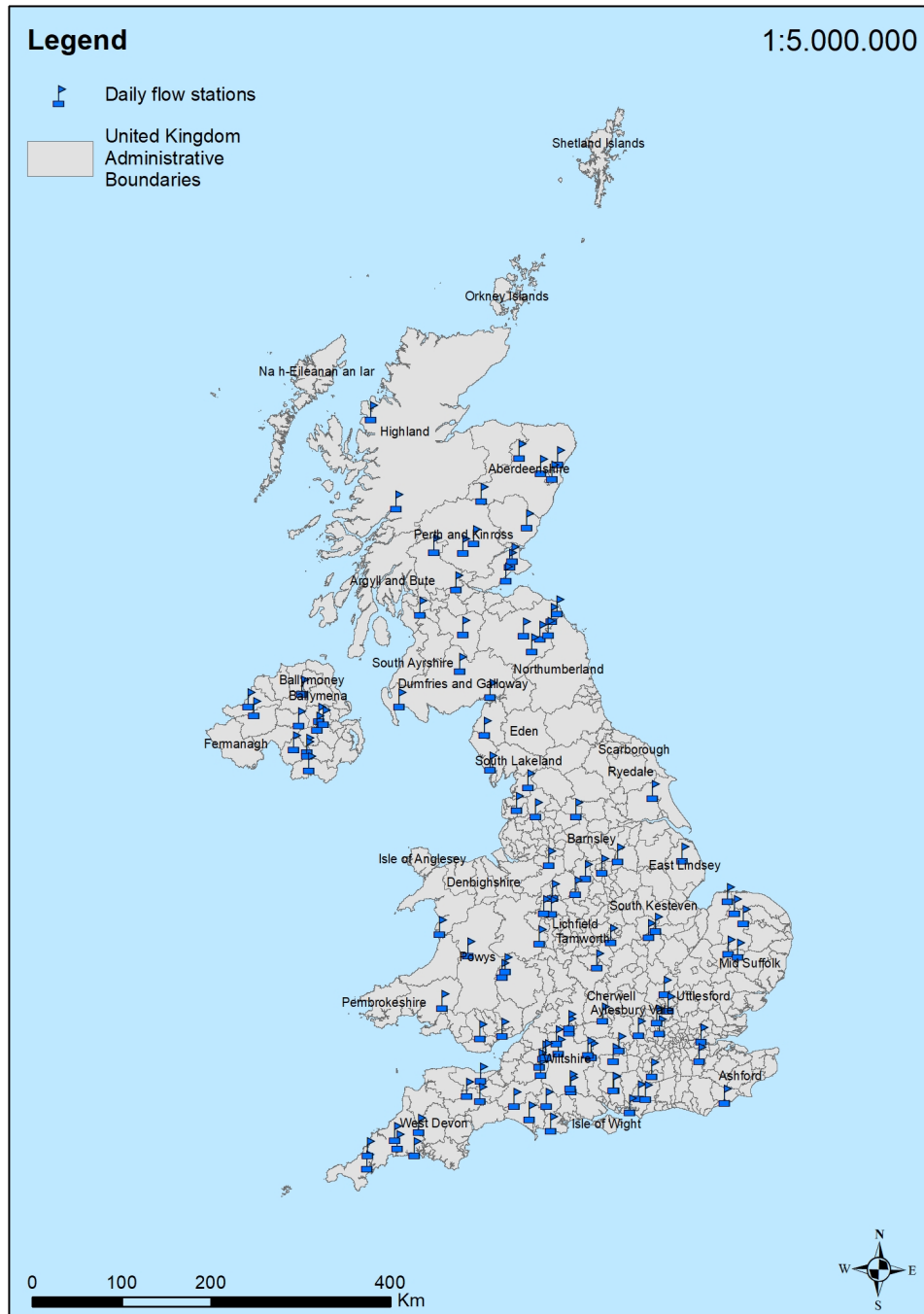
στην εξίσωση (6.1).

$$skew = \gamma_1 := \frac{\sum_{i=1}^n (y_i - \mu)^3 / n}{\sigma^3} \quad (6.1)$$

, όπου  $y_i$  η εκάστοτε τιμή της χρονοσειράς,  $\mu$  η μέση τιμή της χρονοσειράς,  $\sigma$  η τυπική απόκλιση και  $n$  το πλήθος δεδομένων. Πολλές φορές σε διάφορα λογισμικά χρησιμοποιείται ένας προσαρμοσμένος συντελεστής, ονομαζόμενος adjusted Fisher-Pearson, στον οποίο αντί το κλάσμα να διαιρείται με το πλήθος  $n$  πολλαπλασιάζεται με τον όρο  $\frac{n}{(n-1)(n-2)}$ . Η ασυμμετρία μπορεί να θετική όταν η αριστερή ουρά της κατανομής είναι μακρύτερη, αρνητική όταν η δεξιά ουρά της κατανομής είναι μακρύτερη και μηδενική όταν η κατανομή έχει συμμετρία.

**Πίνακας 6.1:** Μεταδεδομένα σταθμών που χρησιμοποιήθηκαν για την εφαρμογή διερεύνησης του μοντέλου Informer (ενδεικτικά οι πρώτοι 20).

id	Όνομα σταθμού	Πρώτη ημέρα	Τελευταία ημέρα	Μήκος	Ασυμμετρία
9004	Bogie at Redcraig	1980-12-01	2021-09-30	41	4.49
10003	Ythan at Ellon	1983-05-18	2021-09-30	38	3.61
11001	Don at Parkhill	1969-12-01	2021-09-30	52	4.41
11002	Don at Haughton	1969-07-01	2021-09-30	52	4.28
12007	Dee at Mar Lodge	1982-09-10	2021-09-30	39	3.58
13008	South Esk at Brechin	1983-01-01	2021-09-30	38	3.86
14001	Eden at Kemback	1967-10-01	2021-09-30	54	3.91
14005	Motray Water at St Michaels	1984-01-01	2021-09-30	37	5.24
15015	Almond at Newton Bridge	1986-01-01	2021-09-30	35	5.57
15023	Braan at Hermitage	1983-01-01	2021-09-30	38	3.97
15024	Dochart at Killin	1982-01-01	2021-09-30	39	3.03
17002	Leven at Leven	1969-08-01	2000-12-31	31	2.59
18014	Bannock Burn at Bannockburn	1979-01-01	2021-09-30	42	5.86
21015	Leader Water at Earlston	1966-10-01	2021-09-30	55	5.42
21016	Eye Water at Eyemouth Mill	1967-10-01	2021-09-30	54	6.92
21021	Tweed at Sprouston	1969-06-01	2021-09-30	52	3.39
21022	Whiteadder Water at Hutton Castle	1969-09-01	2021-09-30	52	6.84
21023	Leet Water at Coldstream	1970-11-01	2021-09-30	51	9.19
21024	Jed Water at Jedburgh	1971-08-01	2021-09-03	50	5.63
26006	Elmswell Beck at Little Driffield	1980-03-22	2021-09-30	41	1.93



**Σχήμα 6.2:** Γεωγραφική κατανομή των 112 σταθμών που χρησιμοποιήθηκαν στην εφαρμογή διερεύνησης του μοντέλου Informer.

### 6.3 Αρχιτεκτονική μοντέλου Informer

Η ανάλυση με το μοντέλο Informer πραγματοποιήθηκε για όλες τις χρονοσειρές με την ίδια αρχιτεκτονική και για το ίδιο δεδομένο χρονικό ορίζοντα (30 ημέρες). Ο συνδυασμός των ακολουθιών εισόδου και εξόδου επιλέχθηκε με βάση την προηγούμενη εφαρμογή του ποταμού River Test και κρατήθηκε σταθερός. Για την εκπαίδευση, αφού χωρίζονται τα δεδομένα σε δεδομένα εκπαίδευσης, επαλήθευσης και δοκιμής (ίδια αναλογία 70:10:20 %), κανονικοποιούνται και χωρίζονται σε υποακολουθίες εισόδου, στόχου, ακριβώς όπως και πριν. Η αρχιτεκτονική του μοντέλου που ακολουθήθηκε βρίσκεται στον Πίνακα 6.2.

Πίνακας 6.2: Υπερπαράμετροι μοντέλου Informer.

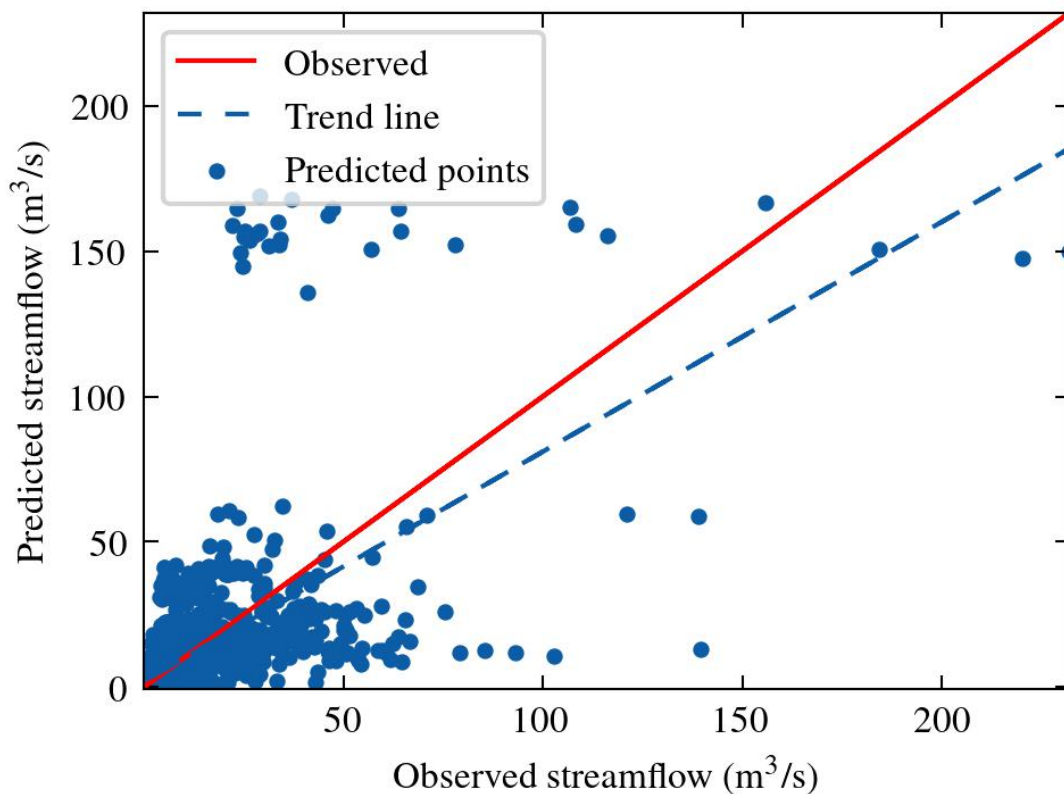
Υπερπαράμετροι	Hyperparameters	Τιμές
Ακολουθία εισόδου	sequence length of encoder	300
Αριθμός αρχικών κομματιών αποκωδικοποιητή	label length of decoder	48
Ακολουθία πρόβλεψης	prediction length	30
Νευρώνες	Encoder input size	2
Νευρώνες	Decoder input size	2
Νευρώνες	Dimension of model	512
Αριθμός παράλληλων κεφαλών	Number of heads	8
Νευρώνες	Encoder layers	512
Νευρώνες	Decoder layers	512
Απόρριψη	Dropout	0.05
Προσοχή	Attention	ProbSparse
Συνάρτηση ενεργοποίησης	Activation function	GELU
Μέγεθος δέσμης	Batch size	32
Αλγόριθμος βελτιστοποίησης	Optimizer	Adam
Συνάρτησης απώλειας	Loss function	mse
Ρυθμός μάθησης	Learning rate	$10^{-4}$
Εποχές	Epochs	8
Υπομονή για πρόωρη διακοπή	Early stopping patience	3

### 6.4 Αποτελέσματα

Για το χρονικό ορίζοντα των 30 ημερών και για την ίδια αρχιτεκτονική το μοντέλο Informer εκπαιδεύτηκε με στόχο την πρόβλεψη της παροχής του ποταμού για τον εκάστοτε σταθμό. Αξίζει να επισημανθεί πως η αντιμετώπιση αυτή δεν ενδείκνυται σε προβλήματα πρόβλεψης με χρήση βαθιάς μάθησης, καθώς κάθε χρονοσειρά θα πρέπει να αντιμετωπίζεται και να βελτιστοποιείται μεμονωμένα. Ωστόσο κάτι τέτοιο σε τόσο μεγάλο πλήθος χρονοσειρών απαιτεί μεγάλο υπολογιστικό κόστος και χρόνο. Επομένως στη συγκεκριμένη περίπτωση επιλέχθηκε η διερεύνηση της απόδοσης του μοντέλου βάσει της σταθερής αρχιτεκτονικής του δικτύου.

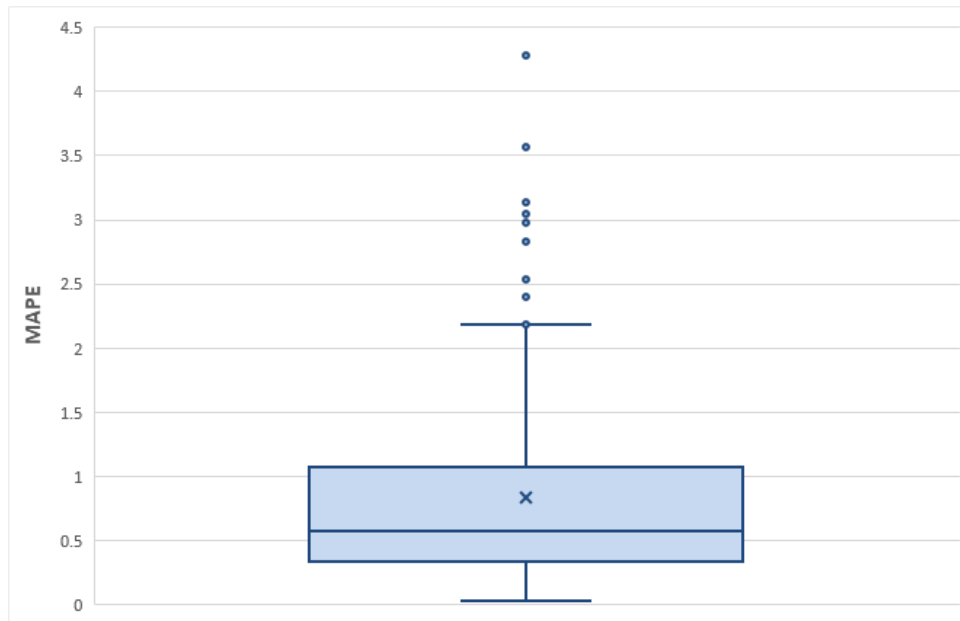
Το Σχήμα 6.3 συνοψίζει τα αποτελέσματα πρόβλεψης και των 112 χρονοσειρών σε ένα διάγραμμα διασποράς ανάμεσα στις παρατηρούμενες και τις εκτιμώμενες παροχές. Στο σχήμα φαίνεται και η ιδανική πρόβλεψη (με κόκκινη συνεχή γραμμή), όπου η πρόβλεψη συμπίπτει με την παρατήρηση. Η γραμμή τάσης των προβλέψεων (μπλε διακεκομμένη γραμμή) παρουσιάζει συντελεστή προσδιορισμού (coefficient of determination)  $R^2 = 0.4$ , υποδηλώνοντας την όχι και τόσο καλή απόδοση του μοντέλου. Παρατηρείται επίσης πως στις μικρές τιμές των παροχών το μοντέλο δείχνει καλύτερη συμπεριφορά, ενώ από την άλλη πλευρά οι μεγάλες τιμές των παροχών έχουν μεγαλύτερη αβεβαιότητα και χειρότερα αποτελέσματα.

Στο Σχήμα 6.4 παρουσιάζεται επιπλέον το θηκόγραμμα (box plot) των μετρικών απόδοσης MAPE. Πιο συγκεκριμένα για κάθε χρονοσειρά και για κάθε πρόβλεψη 30 ημερών, υπολογίστηκε η μετρική MAPE. Το σύνολο αυτών των σφαλμάτων απεικονίζεται μέσω του θηκογράμματος. Ο λόγος επιλογής αυτής της μετρικής είναι πως πρόκειται για τη μόνη αδιαστατοποιημένη μετρική που δεν εξαρτάται από την κλίμακα και το εύρος των αρχικών δεδομένων, κάνοντάς την κατάλληλη στην αξιολόγηση σφαλμάτων από διαφορετικές χρονοσειρές. Το διάγραμμα αποτυπώνει τη μέτρια συμπεριφορά του δικτύου Informer, έχοντας διάμεση τιμή της τάξης του 0.5 ή 50%.



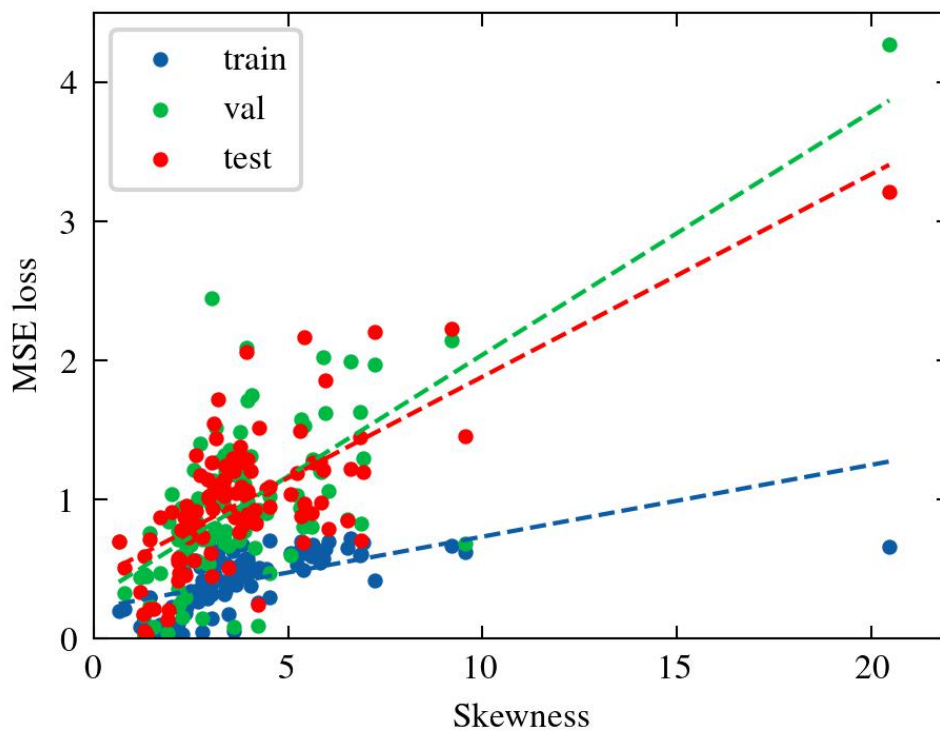
**Σχήμα 6.3:** Διάγραμμα διασποράς προβλεπόμενων τιμών μοντέλου Informer στο σύνολο των 112 σταθμών.

Για τη διερεύνηση της κακής συμπεριφοράς του μοντέλου επιλέχθηκε η σύγκριση των συναρτήσεων απώλειας των τριών υποδιαίρεσεων της χρονοσειράς (train, validation, test set) σε σχέση



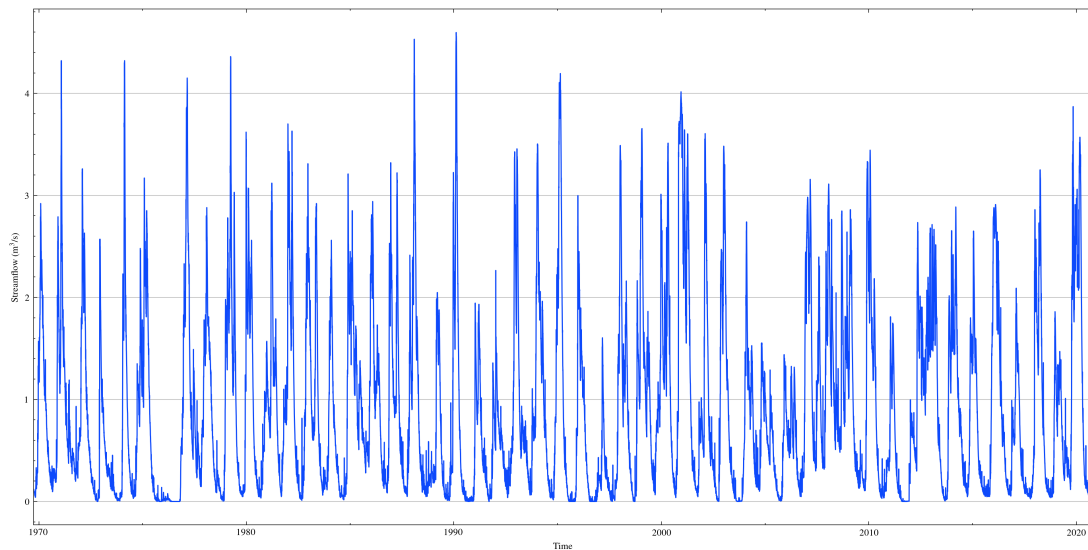
**Σχήμα 6.4:** Θηκόγραμμα (box plot) της μετρικής σφάλματος MAPE μεταξύ εκτιμώμενων και πραγματικών παροχών.

με ένα χαρακτηριστικό μέγεθος της χρονοσειράς όπως η ασυμμετρία. Τα αποτελέσματα αυτής της ανάλυσης φαίνονται στο Σχήμα 6.5 μαζί με τις αντίστοιχες γραμμές τάσης τους. Είναι φανερό πως υπάρχει μία ένδειξη συσχετισμού της ασυμμετρίας με την απόδοση του μοντέλου. Χρονοσειρές με πολύ μεγάλη ασυμμετρία φαίνεται να μην έχουν τα επιθυμητά αποτελέσματα.

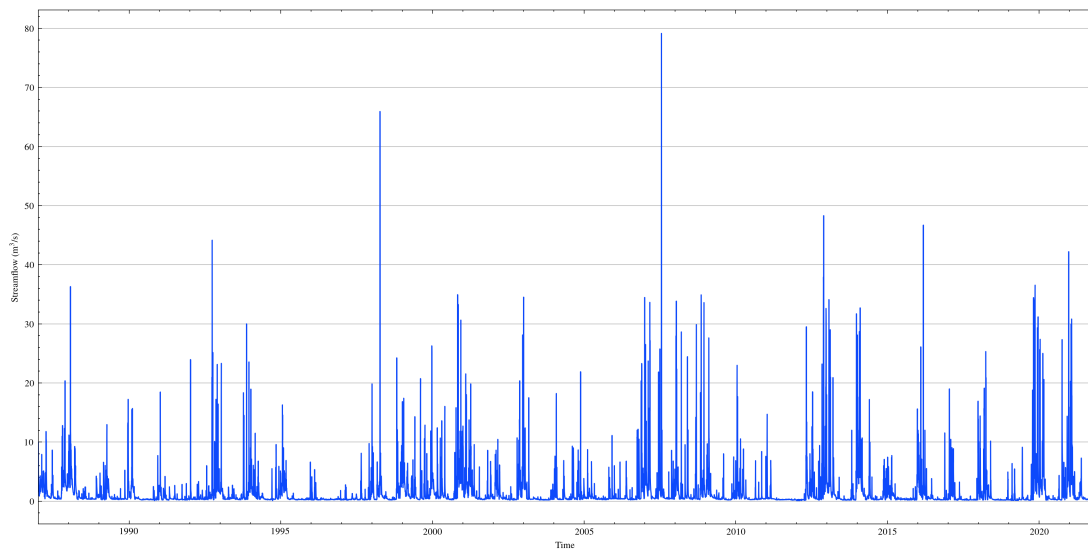


**Σχήμα 6.5:** Η απόδοση του μοντέλου Informer σε σύγκριση με την ασυμμετρία της χρονοσειράς.

Ένα παράδειγμα χρονοσειράς με μικρή ασυμμετρία εμφανίζεται στο Σχήμα 6.6, ενώ με μεγάλη στο Σχήμα 6.7. Η μεγάλη ασυμμετρία εμφανίζει πολλές ακραίες τιμές (outliers) κάνοντας πιο δύσκολη τη διαδικασία της πρόβλεψης.



**Σχήμα 6.6:** Παράδειγμα χρονοσειράς με μικρό δείκτη ασυμμετρίας ( $skew = 1.2$ ).



**Σχήμα 6.7:** Παράδειγμα χρονοσειράς με μεγάλο δείκτη ασυμμετρίας ( $skew = 6.6$ ).

# Κεφάλαιο 7

## Σύνοψη και συμπεράσματα

### 7.1 Συμπεράσματα

Η παρούσα εργασία είχε ως στόχο τη μελέτη της συνεισφοράς δικτύων βαθιάς μάθησης στην πρόβλεψη χρονοσειρών πλημμυρικών φαινομένων. Τα πλημμυρικά φαινόμενα αποτελούν μείζον πρόβλημα στις μέρες μας και η ακριβής πρόβλεψη συνδράμει στην αντιμετώπιση αυτών. Δόθηκε ιδιαίτερη έμφαση σε σύγχρονα εργαλεία βαθιάς μάθησης, όπως οι μετασχηματιστές (transformers) και πιο συγκεκριμένα στο δίκτυο Informer. Παρ' όλα αυτά δεν αμελήθηκε η σύγκριση με πιο παραδοσιακά δίκτυα βαθιάς μάθησης (LSTM) και στατιστικών μοντέλων (ARIMA), καθώς επίσης έγινε και η εισαγωγή σε μία αφελή στοχαστική προσέγγιση για πρόβλεψη ως μοντέλο αναφοράς. Τα μοντέλα αξιολογήθηκαν σε πραγματικά δεδομένα παροχής στον ποταμό River Test στην Αγγλία, ενώ πραγματοποιήθηκε και μία διερεύνηση του μοντέλου Informer μεμονωμένα, σε μεγάλο πλήθος ποταμών του Ηνωμένου Βασιλείου. Η ανάλυση και η επεξεργασία των εφαρμογών οδήγησε σε κάποια συμπεράσματα, τα οποία συνοψίζονται ακολούθως:

1. Η τεχνολογία των μοντέλων που βασίζονται στους μετασχηματιστές (transformer based) παρουσιάζουν μεγάλες δυνατότητες για τη βελτίωση της ακρίβειας των μοντέλων πρόβλεψης χρονοσειρών.
2. Το μοντέλο Informer μειώνει δραστικά τη χρονική πολυπλοκότητα και τις απαιτήσεις μνήμης των κλασικών μετασχηματιστών.
3. Το μοντέλο Informer δείχνει ικανότητα για αποτελεσματική πρόβλεψη, τόσο σε βραχυπρόθεσμη όσο και σε μακροπρόθεσμη ακολουθία πρόβλεψης.
4. Από την εφαρμογή του River Test το μοντέλο Informer φαίνεται να έχει πολύ καλή απόδοση και αρκετά καλύτερη από τα υπόλοιπα μοντέλα. Το στατιστικό μοντέλο ARIMA



---

δεν καταφέρνει να ακολουθήσει τα σύνθετα μοτίβα της χρονοσειράς. Από την άλλη το δίκτυο LSTM έχει καλή απόδοση στις βραχυπρόθεσμες προβλέψεις, χάνοντας έδαφος στην πρόβλεψη σε μακρινό χρονικό ορίζοντα εκεί που ο Informer δίνει σημαντικά καλύτερα αποτελέσματα.

5. Η στοχαστική προσέγγιση είναι απλή και θεωρητικά τεκμηριωμένη σε στοχαστικές διαδικασίες, κατάλληλη για χρήση ως μοντέλο αναφοράς.
6. Η διερεύνηση της εφαρμογής του μοντέλου Informer σε μεγάλο πλήθος ποταμών, αποτυπώνει μία ένδειξη πως η απόδοσή του μειώνεται σε χρονοσειρές με μεγάλη ασυμμετρία.
7. Ορισμένα μειονεκτήματα των μετασχηματιστών είναι η υπολογιστική πολυπλοκότητα, η περιορισμένη ερμηνευσιμότητα και η επεκτασιμότητα. Από την άποψη αυτή, κάθε χρονοσειρά θα πρέπει να προσεγγίζεται και να αξιολογείται ανεξάρτητα.

Τα συμπεράσματα που προέκυψαν από την παρούσα εργασία είναι ιδιαίτερα ενθαρρυντικά τόσο για τη δυνατότητα των δικτύων που βασίζονται σε μετασχηματιστές να επιλύουν προβλήματα μηχανικού, όπως αυτό της πρόβλεψης των πλημμυρών, αλλά και για τη μελλοντική τους ανάπτυξη.

## 7.2 Μελλοντικές επεκτάσεις

Η μελέτη που πραγματοποιήθηκε στα πλαίσια της παρούσας μεταπτυχιακής διπλωματικής εργασίας θα μπορούσε σίγουρα να επεκταθεί και ίσως να επιφέρει και βελτίωση των μοντέλων. Ακολουθούν ερευνητικές κατευθύνσεις που θα μπορούσαν να επεκτείνουν ή/και να βελτιώσουν σημαντικά την παρούσα εργασία:

1. Αξιοποίηση και αξιολόγηση και άλλων μοντέλων και δικτύων που βασίζονται σε μετασχηματιστές, για το πρόβλημα της πρόβλεψης χρονοσειρών πλημμυρικών δεδομένων.
2. Εκμετάλλευση και των άλλων μεγεθών που σχετίζονται με τα πλημμυρικά φαινόμενα (π.χ. θερμοκρασία, βροχή κ.ά.) στο πλαίσιο της πολυμεταβλητής πρόβλεψης (multivariate forecasting), με στόχο την βελτίωση των προβλέψεων.
3. Ποσοτικοποίηση της αβεβαιότητας και επίλυσή της μέσω πιθανοτικής πρόβλεψης.
4. Αξιολόγηση των μοντέλων υλοποιώντας προβλέψεις σε πραγματικό χρόνο, όπου η σύγκριση είναι άμεση.
5. Διερεύνηση της απόδοσης συνδυασμών μοντέλων (Ensemble Modeling) μετασχηματιστών με στατιστικά, στοχαστικά ή άλλα δίκτυα βαθιάς μάθησης.
6. Βελτιστοποίηση του μοντέλου Informer και της αρχιτεκτονικής του.

- 
7. Ανάλυση των επιπτώσεων που μπορεί να έχουν τα χαρακτηριστικά των χρονοσειρών, όπως η ασυμμετρία και η κύρτωση, στην απόδοση του μοντέλου Informer και εύρεση τεχνικών ή μετασχηματισμών για την αντιμετώπιση του πιθανού προβλήματος που επιφέρουν.
  8. Προσπάθεια ερμηνείας της λειτουργίας των μοντέλων βαθιάς μάθησης (Interpretable AI).

---

## Βιβλιογραφία

- [1] E.K. (2007). Οδηγία 2007/60/EK του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 23ης Οκτωβρίου 2007 για την αξιολόγηση και τη διαχείριση των κινδύνων πλημμύρας. *Επίσημη Εφημερίδα της Ευρωπαϊκής Ένωσης της 6.11.2007: L 288: 27 - 34.*
- [2] Κουτσογιάννης, Δ. (2011). *Σχεδιασμός Αστικών Δικτύων Αποχέτευσης*. Έκδοση 4, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα.
- [3] Ritchie, H., Rosado, P. & Roser, M. (2022). Natural Disasters. *Our World in Data*. <https://ourworldindata.org/grapher/number-of-natural-disaster-events?facet=none&country=~Flood>.
- [4] Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, Hoboken.
- [5] Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock price prediction using the ARIMA model. In *2014 UKSim-AMSS 16th international conference on computer modelling and simulation*, pp. 106-112, IEEE.
- [6] Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media.
- [7] Ahmed, N.K., Atiya, A.F., Gayar, N. E., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5-6), 594—621, doi: 10.1080/07474938.2010.481556.
- [8] Krollner, B., Vanstone, B.J., & Finnie, G.R.. (2010). Financial time series forecasting with machine learning techniques: a survey. In *The European Symposium on Artificial Neural Networks*.
- [9] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794.

- 
- [10] März, A. (2019). XGBoostLSS: An Extension of XGBoost to Probabilistic Forecasting. *arXiv:1907.03178*.
- [11] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2222-2232, Oct. 2017, doi: 10.1109/TNNLS.2016.2582924.
- [12] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. In *Proceedings of the Neural Information Processing Systems Workshop on Deep Learning*.
- [13] Wang, S., Cao, J., & Philip, S. Y. (2020). Deep learning for spatio-temporal data mining: A survey. *IEEE transactions on knowledge and data engineering*, 34(8), 3681-3700, doi: 10.1109/TKDE.2020.3025580.
- [14] Moraux, A., Dewitte, S., Cornelis, B., & Munteanu, A. (2019). Deep Learning for Precipitation Estimation from Satellite and Rain Gauges Measurements. *Remote Sensing*, 11(21), 2463, doi: <https://doi.org/10.3390/rs11212463>.
- [15] Ghimire, S., Yaseen, Z. M., Farooque, A. A., Deo, R. C., Zhang, J., & Tao, X. (2021). Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks. *Scientific Reports* 11(1), 17497, doi: <https://doi.org/10.1038/s41598-021-96751-4>.
- [16] Ghobadi, F., & Kang, D., Multi-Step Ahead Probabilistic Forecasting of Daily Streamflow Using Bayesian Deep Learning: A Multiple Case Study. *Water*, 14(22), 3672, doi: <https://doi.org/10.3390/w14223672>.
- [17] Flunkert, V., Salinas, D., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3), 1181–1191, doi: <https://doi.org/10.1016/j.ijforecast.2019.07.001>.
- [18] Oreshkin, B. N., Carpo, D., Chapados, N., & Bengio, Y. (2019). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. In *International Conference on Learning Representations*.
- [19] Tyralis, H. & Papacharalampous, G. A. (2018). Large-scale assessment of Prophet for multi-step ahead forecasting of monthly streamflow. *Advances in Geosciences*, 45, 147–153. <https://doi.org/10.5194/adgeo-45-147-2018>.
- [20] Papacharalampous, G. A. & Tyralis, H. (2018). Evaluation of random forests and Prophet for daily streamflow forecasting. *Advances in Geosciences*, 45, 201–208, doi: <https://doi.org/10.5194/adgeo-45-201-2018>.

- 
- [21] Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. *The American Statistician*, 72(1), 37–45 doi: <https://doi.org/10.1080/00031305.2017.1380080>.
- [22] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30, pp. 5998–6008.
- [23] Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., & Sun, L. (2022). Transformers in time series: A survey. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI'23)*.
- [24] Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y. X., & Yan, X. (2019). Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *Advances in neural information processing systems*, 32.
- [25] Wu, N., Green, B., Ben, X. & O'Banion, S. (2020). Deep transformer models for time series forecasting: The influenza prevalence case. *arXiv preprint arXiv:2001.08317*.
- [26] Wu, S., Xiao, X., Ding, Q., Zhao, P., Wei, Y., and Huang, J. (2020). Adversarial Sparse Transformer for Time Series Forecasting. *Advances in Neural Information Processing Systems*, Vol. 33, pp. 17105–17115.
- [27] Wang, S., Li, B. Z., Khabisa, M., Fang, H., & Ma, H. (2020). Linformer: Self-Attention with Linear Complexity. *arXiv preprint arXiv:2006.04768*.
- [28] Kitaev, N., Kaiser, L., & Levskaya, A. (2020). Reformer: The Efficient Transformer. In *International Conference on Learning Representations*.
- [29] Beltagy, I., Peters, M. E., and Cohan, A. (2020). Longformer: The Long-Document Transformer. *arXiv:2004.05150*.
- [30] Lin, Y., Koprinska, I., & Rana, M. (2021). SSDNet: State Space Decomposition Neural Network for Time Series Forecasting. In *2021 IEEE International Conference on Data Mining (ICDM)*, pp. 370-378, IEEE, doi: <https://doi.org/10.1109/icdm51629.2021.00048>
- [31] Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4), 1748–1764. doi: <https://doi.org/10.1016/j.ijforecast.2021.03.012>
- [32] Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 34, 22419-22430.

- 
- [33] Cirstea, R. G., Guo, C., Yang, B., Kieu, T., Dong, X., & Pan, S. (2022). Triformer: Triangular, Variable-Specific Attentions for Long Sequence Multivariate Time Series Forecasting. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, doi: <https://doi.org/10.24963/ijcai.2022/277>.
- [34] Shen, L., & Wang, Y. (2022). TCCT: Tightly-coupled convolutional transformer on time series forecasting. *Neurocomputing*, 480, 131–145, doi: <https://doi.org/10.1016/j.neucom.2022.01.039>.
- [35] Liu, S., Yu, H., Liao, C., Li, J., Lin, W., Liu, A. X., & Dustdar, S. (2022). Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In *International Conference on Learning Representations*.
- [36] Drouin, A., Marcotte, É., & Chapados, N. (2022). Tactis: Transformer-attentional copulas for time series. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 5447–5493, PMLR 162:5447-5493.
- [37] Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., & Jin, R. (2022). FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *39th International Conference on Machine Learning (ICML)*, pp. 27268-27286, PMLR.
- [38] Shabani, A., Abdi, A., Meng, L., & Sylvain, T. (2023). Scaleformer: iterative multi-scale refining transformers for time series forecasting. In *The Eleventh International Conference on Learning Representations*.
- [39] Zhang, Y., & Yan, J., (2023). Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The Eleventh International Conference on Learning Representations*.
- [40] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12), 11106-11115, doi: <https://doi.org/10.1609/aaai.v35i12.17325>.
- [41] Shang, J., Zhao, B., Hua, H., Wei, J., Qin, G., & Chen, G. (2023). Application of Informer Model Based on SPEI for Drought Forecasting. *Atmosphere*, 14(6), 951.
- [42] Yang, Z., Liu, L., Li, N., & Tian, J. (2022). Time Series Forecasting of Motor Bearing Vibration Based on Informer. *Sensors*, 22(15), 5858.
- [43] Huang, X., & Jiang, A. (2022). Wind Power Generation Forecast Based on Multi-Step Informer Network. *Energies*, 15(18), 6642.

- 
- [44] Luo, R., Song, Y., Huang, L., Zhang, Y., & Su, R. (2023). AST-GIN: Attribute-Augmented Spatiotemporal Graph Informer Network for Electric Vehicle Charging Station Availability Forecasting. *Sensors*, 23(4), 1975.
- [45] Luo, R., Song, Y., Huang, L., Zhang, Y., & Su, R. (2022). Stgin: A Spatial Temporal Graph-Informer Network for Long Sequence Traffic Speed Forecasting. *arXiv preprint arXiv:2210.01799*.
- [46] Ulhaq, A., Akhtar, N., Pogrebna, G., & Mian, A. (2022). Vision transformers for action recognition: A survey. *arXiv preprint arXiv:2209.05700*.
- [47] Τεχνητή νοημοσύνη (2023). Wikipedia. [https://en.wikipedia.org/wiki/Artificial\\_intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence) (Ημερομηνία Πρόσβασης: 20-5-2023).
- [48] Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, Vol. 59, No. 236, pp. 433-460.
- [49] Arbel, N. Deep learning vs. Machine learning vs. Artificial Intelligence. *Javatpoint*. <http://www.javatpoint.com/deep-learning-vs-machine-learning-vs-artificial-intelligence> (Accessed: 22-05-2023).
- [50] Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210-229.
- [51] Mitchell, T. M. (1997). *Machine learning* (Vol. 1). New York: McGraw-hill.
- [52] Wolpert D. H., & Macready W. G. (1997). No free lunch theorems for optimization. In *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67-82, doi: 10.1109/4235.585893.
- [53] Addagatla, A. (2021, April). Investigating Underfitting and Overfitting. *Medium*. <https://medium.com/geekculture/investigating-underfitting-and-overfitting-70382835e45c> (Accessed: 26-05-2023).
- [54] Τεπετίδης, Ν. (2020). *Διόδευση πλημμυρών με χρήση τεχνητών νευρωνικών δικτύων*. Διπλωματική εργασία, Σχολή Πολιτικών Μηχανικών ΕΜΠ, Αθήνα.
- [55] Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., Lamantia, A. S., McNamara, J. O., & Williams, S. M. (2001). *Neuroscience*. 2<sup>nd</sup> Edition, Sinauer Associates, Sutherland (MA).
- [56] Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. In *Psychological review*, 65(6), 386-408.



- 
- [57] Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 2<sup>nd</sup> Edition, O'Reilly Media, United States of America, pp. 510.
- [58] Tanner, G. (2021, September). Activation functions. *ML Explained*. <https://ml-explained.com/blog/activation-functions-explained> (Accessed: 22-06-2023).
- [59] Smolyakov, V. (2018, January). Neural Network Optimization Algorithms. *Medium*. <https://towardsdatascience.com/neural-network-optimization-algorithms-1a44c282f61d> (Accessed: 15-10-2022).
- [60] Doshi, S. (2019). Various Optimization Algorithms for Training Neural Network. *Medium*. <https://towardsdatascience.com/optimizers-for-training-neuralnetwork-59450d71caf6> (Accessed: 15-10-2022).
- [61] McGonagle, J., Shalkouski, G., Williams, C. et al. Backpropagation. *Brilliant*. <https://brilliant.org/wiki/backpropagation/> (Accessed: 15-10-2022).
- [62] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- [63] Pytlak, R. (2009). Limited Memory Quasi-Newton Algorithms. *Conjugate Gradient Algorithms in Nonconvex Optimization*, 159-190.
- [64] Krzaczynski, R. (2020, March). Limited-Memory-Broyden-Fletcher-Goldfarb-Shanno Algorithm. *Medium*. <https://towardsdatascience.com/limited-memorybroyden-fletcher-goldfarb-shanno-algorithm-in-ml-net-118dec066ba>.
- [65] Vorontsova, M. (2019, September). Deep Learning vs Machine Learning: Overview & Comparison, *Soshace*. <https://soshace.com/deep-learning-vs-machine-learning-overview-comparison/> (Accessed: 31-05-2023).
- [66] Ronaghan, S. (2018, July). Deep Learning: Overview of Neurons and Activation Functions. *Medium*. <https://srnghn.medium.com/deep-learning-overview-of-neurons-and-activation-functions-1d98286cf1e4> (Accessed: 31-05-2023).
- [67] Borges, P. (2018, October). Deep Learning: Recurrent Neural Networks. *Medium*. <https://medium.com/deeplearningbrasil/deep-learning-recurrent-neural-networks-f9482a24d010> (Accessed: 23-02-2023).
- [68] Wikimedia Commons contributors. Recurrent neural network unfold.svg. *Wikimedia Commons, the free media repository*. [https://commons.wikimedia.org/w/index.php?title=File:Recurrent\\_neural\\_network\\_unfold.svg&oldid=655242383](https://commons.wikimedia.org/w/index.php?title=File:Recurrent_neural_network_unfold.svg&oldid=655242383) (Accessed: 23-02-2023).

- 
- [69] Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560.
- [70] Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2021). Dive into Deep Learning. *arXiv preprint arXiv:2106.11342*.
- [71] Hochreiter, S. & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780, doi: <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [72] Ποντίκη, Μ. (2022). *Πρόβλεψη της τιμής του κρυπτονομίσματος με αλγόριθμους βαθιάς μηχανικής μάθησης*. Διπλωματική εργασία, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών ΕΜΠ, Αθήνα.
- [73] Arbel, N. (2018, December). How LSTM networks solve the problem of vanishing gradients. *Medium*. <https://medium.datadriveninvestor.com/how-do-lstm-nets-work-solve-the-problem-of-vanishing-gradients-a6784971a577> (Accessed: 28-02-2023).
- [74] Mittal, A. (2019, October). Understanding RNN and LSTM. *Medium*. <https://aditi-mittal.medium.com/understanding-rnn-and-lstm-f7cdf6dfc14e> (Accessed: 28-02-2023).
- [75] Zhou, H., Li, J., Zhang, S., Zhang, S., Yan, M., & Xiong, H. (2023). Expanding the prediction capacity in long sequence time-series forecasting. *Artificial Intelligence*, 318, 103886.
- [76] Koutsoyiannis, D. (2016). Generic and parsimonious stochastic modelling for hydrology and beyond. *Hydrological Sciences Journal*, 61(2), 225-244.
- [77] Papoulis, A., & Pillai, S. U. (1991). *Random Variables and Stochastic Processes*. McGraw Hill.
- [78] Papadopoulos, V., & Giovanis, D. G. (2017). *Stochastic Finite Element Methods: An Introduction*. Springer.
- [79] Koutsoyiannis, D. (2022). *Stochastics of Hydroclimatic Extremes - A Cool Look at Risk*. Edition 2, ISBN: 978-618-85370-0-2, 346 pages, doi:10.57713/kallipos-1, Kallipos Open Academic Editions, Athens.
- [80] Dimitriadis, P., & Koutsoyiannis, D. (2015). Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst–Kolmogorov processes. *Stochastic Environmental Research & Risk Assessment*, 29 (6), 1649–1669, doi:10.1007/s00477-015-1023-7.

- 
- [81] Wu, J. (2018, June). Stochastic Processes and Monte Carlo Method. *QuantConnect*. <https://www.quantconnect.com/learning/articles/introduction-to-options/stochastic-processes-and-monte-carlo-method> (Accessed: 10-06-2023).
- [82] Koutsoyiannis, D. (2013a). Hydrology and change. *Hydrological Sciences Journal*, 58(6), 1177-1197.
- [83] Moschos, E. (2018). *Natural variability and persistence in oceanic variables*, Diploma Thesis, School of Civil Engineering NTUA, Athens.
- [84] Wikipedia contributors. (2023). River Test. *Wikipedia*. [https://en.wikipedia.org/wiki/River\\_Test](https://en.wikipedia.org/wiki/River_Test) (Accessed: 13-06-2023).
- [85] Zhou, E. (2022, October). *River Test Flow* [Data Set]. Kaggle. <https://www.kaggle.com/datasets/jzhou3083/rivertest> (Accessed: 15-01-2023).
- [86] Met Office (2006): UK Daily Temperature Data, Part of the Met Office Integrated Data Archive System (MIDAS). NCAS British Atmospheric Data Centre. <https://data.ceda.ac.uk/badc/ukmo-hadobs/data/insitu/MOHC/HadOBS/HadUK-Grid/v1.1.0.0>.
- [87] Department for Environment Food & Rural Affairs. API documentation. <https://environment.data.gov.uk/hydrology/doc/reference>.
- [88] UK National River Flow Archive. <https://nrfa.ceh.ac.uk/>.
- [89] Howell, E. (2022, December). An overview of forecasting performance metrics. *Medium*. <https://towardsdatascience.com/an-overview-of-forecasting-performance-metrics-ef548dad0134> (Accessed: 25/06/2023).
- [90] UK National River Flow Archive. Search Data. <https://nrfa.ceh.ac.uk/data/search> (Accessed: 10/02/2023).