UNIVERSITY OF THESSALY
SCHOOL OF AGRICULTURAL SCIENCES
DEPARTMENT OF ICHTHYOLOGY AND AQUATIC ENVIRONMENT
LABORATORY OF ECOHYDRAULICS AND INLAND WATER MANAGEMENT

Ph.D. Thesis

Investigation of the Water Quality Parameters of Lake Kastoria from Time-Series Monitoring Data Using Machine Learning Techniques for Simulation and Prediction

Lamprini V. Karamoutsou
Chemical Engineer
MSc Water Resources Engineering

Volos, 2020

Investigation of the Water Quality Parameters of Lake Kastoria from Time-Series Monitoring Data Using Machine Learning Techniques for Simulation and Prediction

Advisory Committee:

1. Aris Psilovikos, Professor of University of Thessaly, Department of Ichthyology and Aquatic Environment, Sustainable Water Resources Management, Head of Laboratory of Ecohydraulics & Inland Water Management

2. Evangelia Farsirotou, Professor of University of Thessaly, Department of Ichthyology and Aquatic Environment, Hydrodynamic and Flood Control Works

3. Evangelos Keramaris, Assistant Professor of University of Thessaly, Department of Civil Engineering, Experimental Hydraulics

Examination Committee:

1. Aris Psilovikos, Professor of University of Thessaly, Department of Ichthyology and Aquatic Environment, Sustainable Water Resources Management, Head of Laboratory of Ecohydraulics & Inland Water Management

2. Evangelia Farsirotou, Professor of University of Thessaly, Department of Ichthyology and Aquatic Environment, Hydrodynamic and Flood Control Works

3. Evangelos Keramaris, Assistant Professor of University of Thessaly, Department of Civil Engineering, Experimental Hydraulics

4. Konstantinos Kormas, Professor of University of Thessaly, Department of Ichthyology and Aquatic Environment, Microbial Ecology of the Aquatic Environment

5. Maria Papadopoulou, Professor of National Technical University of Greece, School of Rural & Surveying Engineering, Natural Resources Management & Environmental Planning

6.Athanasios Loukas, Associate Professor of Aristotle University, Department of Rural and Surveying Engineering, Technical Hydrology & Water Resources Management

7. Csilla Farkas, Research Scientist, Norwegian Institute of Bio-economy Research (NIBIO), Norway, Division of Environment and Natural Resources

L. Karamoutsou

Investigation of the Water Quality Parameters of Lake Kastoria from Time-Series Monitoring Data Using Machine Learning Techniques for Simulation and Prediction

ISBN

- *Lamprini Karamoutsou, 2020, Investigation of the Water Quality Parameters of Lake Kastoria from Time-Series Monitoring Data Using Machine Learning Techniques for Simulation and Prediction*
- *Number of preliminary pages: 38*
- *Number of pages of PhD thesis: 219*
- *Number of Tables: 12*
- *Number of Figures: 116*
- *Number of References: 157*

L. Karamoutsou

Investigation of the Water Quality Parameters of Lake Kastoria from Time-Series Monitoring Data Using Machine Learning Techniques for Simulation and Prediction

Part of this PhD thesis was funded by the scholarships' project Integrated Marine and Inland Water Management of the **Greek State Scholarships Foundation (IKY)** and co-funded by 85% of resources **FM EEA 2009-2014** and by 15% from the **Public Investment Program.**

Διερεύνηση των Παραμέτρων της Ποιότητας του Νερού της Λίμνης Καστοριάς από Χρονοσειρές Παρακολούθησης με Χρήση Τεχνικών Μηχανικής Μάθησης για Προσομοίωση και Πρόβλεψη

Τριμελής Συμβουλευτική Επιτροπή:

1. Άρης Ψιλοβίκος, Καθηγητής Πανεπιστημίου Θεσσαλίας, Τμήμα Γεωπονίας Ιχθυολογίας και Υδάτινου Περιβάλλοντος, Αειφορική Διαχείριση Υδατικών Πόρων, Διευθυντής Εργαστηρίου Οικοϋδραυλικής και Διαχείρισης Εσωτερικών Υδάτων

2. Ευαγγελία Φαρσιρώτου, Καθηγήτρια Πανεπιστημίου Θεσσαλίας, Τμήμα Γεωπονίας Ιχθυολογίας και Υδάτινου Περιβάλλοντος, Υδροδυναμικά και Αντιπλημμυρικά Έργα

3. Ευάγγελος Κεραμάρης, Επίκουρος Καθηγητής Πανεπιστημίου Θεσσαλίας, Τμήμα Πολιτικών Μηχανικών, Πειραματική Υδραυλική

Επταμελής Εξεταστική Επιτροπή:

1. Άρης Ψιλοβίκος, Καθηγητής Πανεπιστημίου Θεσσαλίας, Τμήμα Γεωπονίας Ιχθυολογίας και Υδάτινου Περιβάλλοντος, Αειφορική Διαχείριση Υδατικών Πόρων, Διευθυντής Εργαστηρίου Οικοϋδραυλικής και Διαχείρισης Εσωτερικών Υδάτων

2. Ευαγγελία Φαρσιρώτου, Καθηγήτρια Πανεπιστημίου Θεσσαλίας, Τμήμα Γεωπονίας Ιχθυολογίας και Υδάτινου Περιβάλλοντος, Υδροδυναμικά και Αντιπλημμυρικά Έργα

3. Ευάγγελος Κεραμάρης, Επίκουρος Καθηγητής Πανεπιστημίου Θεσσαλίας, Τμήμα Πολιτικών Μηχανικών, Πειραματική Υδραυλική

4. Κωνσταντίνος Κορμάς, Καθηγητής Πανεπιστημίου Θεσσαλίας, Τμήμα Γεωπονίας Ιχθυολογίας και Υδάτινου Περιβάλλοντος, Μικροβιακή Οικολογία Υδάτινου Περιβάλλοντος

5. Μαρία Παπαδοπούλου, Καθηγήτρια Εθνικού Μετσόβιου Πολυτεχνείου, Σχολή Αγρονόμων & Τοπογράφων Μηχανικών, Διαχείριση Φυσικών Πόρων και Περιβαλλοντικός Σχεδιασμός

6. Αθανάσιος Λουκάς, Αναπληρωτής Καθηγητής Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, Τμήμα Αγρονόμων και Τοπογράφων Μηχανικών, Τεχνική Υδρολογία & Διαχείριση Υδατικών Πόρων

7. Csilla Farkas, Ερευνήτρια, Τομέας: Περιβάλλον και Φυσικοί Πόροι, Νορβηγικό Ινστιτούτο Βιο-Οικονομικών Ερευνών (NIBIO), Νορβηγία

Διερεύνηση των Παραμέτρων της Ποιότητας του Νερού της Λίμνης Καστοριάς από Χρονοσειρές Παρακολούθησης με Χρήση Τεχνικών Μηχανικής Μάθησης για Προσομοίωση και Πρόβλεψη

ISBN

- *Λαμπρινή Καραμούτσου, 2020, Διερεύνηση των Παραμέτρων της Ποιότητας του Νερού της Λίμνης Καστοριάς από Χρονοσειρές Παρακολούθησης με Χρήση Τεχνικών Μηχανικής Μάθησης για Προσομοίωση και Πρόβλεψη*
- *Αριθμός προκαταρκτικών σελίδων: 38*
- *Συνολικός αριθμός σελίδων Διδακτορικής Διατριβής: 219*
- *Αριθμός Πινάκων: 12*
- *Αριθμός Σχημάτων: 116*
- *Αριθμός Βιβλιογραφικών αναφορών: 157*

Λ. Καραμούτσου

Διερεύνηση των Παραμέτρων της Ποιότητας του Νερού της Λίμνης Καστοριάς από Χρονοσειρές Παρακολούθησης με Χρήση Τεχνικών Μηχανικής Μάθησης για Προσομοίωση και Πρόβλεψη

Μέρος της παρούσας διδακτορικής διατριβής πραγματοποιήθηκε μέσω του έργου Υποτροφίες ΙΚΥ στην ολοκληρωμένη Διαχείριση Θαλάσσιων και Εσωτερικών Υδάτων και συγχρηματοδοτήθηκε κατά 85% από τους πόρους του ΧΜ ΕΟΧ περιόδου 2009-2014 και κατά 15% από το πρόγραμμα Δημοσίων Επενδύσεων.

To the runaway,
To the greenhorn,
To the one who fights when a menace is born.

Acknowledgements

I would like to express my sincere gratitude to my advisor Prof. Aris Psilovikos for his continuous support, motivation and immense knowledge. His guidance, patience and invaluable contribution of time and ideas helped me a lot duringthe whole process of research and writing of this thesis.

Besides my advisor, I would like to acknowledge the other two members of my advisory committee, Prof. Evaggelia Farsirotou and Ass. Prof. Evaggelos Keramaris, for their support and the fruitful exchange of ideas.

Finally, I would like to thank the other four members of the examination committee, Prof. Athanasios Loukas, Prof. Konstantinos Kormas, Prof. Maria Papadopoulou and Researcher Csilla Farkas, for their insightful comments and suggestions.

My sincere thanks also go to Dr. Per Stålnacke, Director of Research at NIBIO, Norwegian Institute of Bio-economy Research, who provided me with the opportunity to join their team and Dr. Farkas, for her help and guidance during my stay in Norway. Moreover, I would also like to recognize the assistance ofPhD studentsThodoris Papathanasiou, expert on Geographical Information Systems and Konstantinos Liakos, expert on Machine Learning, for their advices and precious help while processing my PhD dissertation.

I am also grateful to the friends I have made during my stay at the city of Volos and especially my Georgia for supporting our good and crazy times. Thanks also go to Angeliki, Rymenidi – Ntokouzi family, Fanouria and my “mexican” friend for their support. The last year has not been an easy ride. I would like to thank Akis, Konstantinos and Rania, my friends who have become family. Without their precious support it would not be possible to finish this research. Thanks should also go to my lifelong friends Markella, Giannis and Maria.

Last but not least, I truly thank my parents Vasileia and Vasileios, who unconditionally support and love me during my good and bad times. Their discreet and catalytic presence without hints or criticism has taught me to live freely and responsibly. Finally, I am grateful to Vasilis, my wonderful niece Chrysanthi and my brother-in-law Lampros; my beloved sister Marina, my greatest supporter, for sticking by my side, helping me by any means.

L. V. Karamoutsou

Abstract

Lakes constitute important indicators of climate change. Geographically, they comprise the lowest points of an internal catchment's area, collecting the surface runoff water from the hydrographic network. Their physical, chemical and biological responses to the climate give a variety of priceless information. In recent years, water bodies have undergone extensive change as a result of widespread qualitative and quantitative degradation. In terms of quality deterioration, the first signs of pollution were identified in the surface waters of lakes and rivers. So, the need for collecting, studying and evaluating the information given from the lake is crucial.

Regarding the issue of bad data quality and scarcity with measurements on monthly basis, even if the measurements are made at a specific time with a very strict and specific monitoring schedule, the risk to loose data on peak events, such as point or non-point pollution phenomena due to natural or anthropogenic factors is hard to answer. Thus, these peak phenomena, cannot be recorded if the measurement of the specific quality parameter is conducted on a precise date and time and the peak phenomenon occurs a few hours or days later and before the next measurement. This risk can only be prevented by continuous, telemetric and integrated monitoring on a 24-hour schedule, with the state of the art protocols for rapid recording, tele-transmission and economy in power and monetary resources.

In a world where water is extensively polluted, many models have been applied to predict water quality parameters. Machine Learning (ML), an application of artificial intelligence, is a rapidly developing field. Artificial Neural Networks (ANNs), that mimic human brains, are able to map the non-linear relationships among variables that are characteristic of ecosystems. Nowadays, specifically, Deep Neural Networks (DNNs), as an upgraded and complex version of Artificial Neural Network, have dramatically improved the state-of-the-art in many scientific fields such as environmental studies. As a DNN needs a large amount of data in order to complete its learning process comparatively to ANN, Monitoring plays a decisive role in lake management by providing systematic information of the qualitative parameters.

The Study Area of the present PhD Thesis, comprises the catchment area of Lake Kastoria, which is locatedin the Region of Western Macedonia in Greece, protected by Directives, Regulations and International Conventions. Lake Kastoria is a shallow polymictic lake, with intense agricultural activities in its catchment area loading both point and nonpoint source pollutants. In order to record water quality characteristics (at lake's surface), four (4) telemetric

stations were installed at specific locations in Lake Kastoria, namely: Gkiole, Stavros, Psaradika and Toichio. The above telemetric stations were installed based on the natural, hydrological and geomorphological characteristics of the lake and the needs that existed in such a way as to obtain a representative estimation of its water quality parameters. A dataset from November 11, 2015 to March 15, 2018, on an hourly basis from the aforementioned four telemetric stations, is used. The available dataset consists of 1) Chlorophyl-a – Chl-a (μg/L), 2) pH, 3) Water Temperature – Tw (°C), 4) Electrical Conductivity of Water – ECw (μS/cm), 5) Turbidity (NTU), 6) Ammonia Nitrogen – N-NH_4 (ppm) (Not Available (NA) data for Stavros Station), 7) Nitrate Nitrogen – N-NO_3, (mg/L) and 8) Dissolved Oxygen – DO (mg/L).

Geographic Information Systems(GIS) is used, in order to approximate and understand the state of the study area. Especially: a) the Land Use Land Cover (LULC) changes from 1945 to 2000 and b) the recordings of the available water quality parameters and their spatial integration and simulation, by using the Inverse Distance Weighting method (IDW), are presented and discussed. GIS is used as a powerful tool in creating thematic maps, with any geographical information that can be quantified. Here, with the support of IDW method and the available water quality parameters measurements from four different monitoring locations, the spatial information isintegrated to the whole surface of Lake Kastoria. Useful information is extracted that allow us to understand: a) The consequences of the LULC changes, in the Water Quality and b) the correlation of each one of the water quality parameters, comparing with the other similar ones, in different locations and to assess the environmental and ecological status of the lake. It is concluded that, the IDW spatial interpolation models and the representative maps that have been derived, are in agreement with the recordings of the telemetric stations and the water quality processes, which take place in the catchment area of Lake Kastoria.

Furthermore, the ability of Deep Neural Networks to predict in real-time the quality parameter of DO, Chl-a and Turbidity in Lake Kastoria, is tested. The examined Feed Forward Deep Neural Networks (FF-DNN) of DO, are tested for four different structures based on the number of nodes, used in the hidden layers and the number of input variables. More specific, DO model with structures:

a) 7-64-64-1 for Gkiole, Psaradika and Toichio stationsand 6-64-64-1 for Stavros station where NH_4 parameter is not available,

b) 4-64-64-1 for all stations,

c) 7-32-32-1 for Gkiole, Psaradika and Toichio stations and 6-32-32-1 for Stavros station where NH_4 parameter is not availableand

d) 4-32-32-1 for all stations, are tested.

Moreover, Turbidity and Chl-a models, with structures7-64-64-1 for Gkiole, Psaradika and Toichio stations and 6-64-64-1 for Stavros station, are investigated.

In order to develop **Deep Neural Networks**– based on **Machine Learning** methods – the following have been used:

1) the **Python**Programming Language,
2) the **Spyder**Scientific Environment,
3) the **Tensorflow**Open Source Machine Learning Platform and
4) the **Nvidia**'**Scuda** (Compute Unified Device Architecture) Parallel Platform (Graphics Card) are used.

All of the above have been integrated into **Anaconda**, an Open Source Programming Language. The Optimal Network Architecture of each structure for each station is selected, based on MAE, MSE and NSE Statistical Descriptors.

The investigated models of Turbidity and Chl-a, with structures 7-64-64-1 for Gkiole, Toichio and Psaradika stations and 6-64-64-1 for Stavros Station, couldn't show a good performance. The inability of models could be explained due to their lack of the capability to handle such a large variance in the data. The well-trained DNN, with structure 7-32-32-1, produces results with the MAE of 0,54; 0,55, the MSE of 0,65; 0,68 and the NSE of 0,89; 0,86 for the training and the testing process respectively, with a good predictive ability for Gkiole station. For Toichio station, the structure 7-64-1 prevails with the MAE of 0,48; 0,49, the MSE of 0,52; 0,51 and the NSE of 0,92; 0,93 for the training and the testing process respectively and it constitutes the structure with the best performance compared with all the structures for all stations. For Psaradika station, the structure 4-64-64-1 produces the best results in relation to the other structures of this station with the MAE of 0,57; 0,58, the MSE of 0,65; 0,68 and the NSE of 0,91; 0,92 for the training and the testing process respectively. Finally, Stavros station with structure 4-32-32-1 presents results with the MAE of 0,69; 0,70, the MSE of 0,98; 1,01 and the NSE of 0,91; 0,90 for the training and the testing process respectively.

The optimal selected Feed Forward Deep Neural Networks (DNN) of DO for each station provide information in Real Time and comprise a powerful Decision Support System (DSS), for preventing accidental and emergency conditions that may arise from both Natural and

Anthropogenic Hazards.Based on the literature and the, so far, knowledge of the author, **the importance of this work also lies within the fact that there is no other published work using the same platforms, tools and methodology of Deep Learning, as in the present PhD Thesis, in order to investigate the predictive capacity of water quality parameters of lakes in real time, at both national and international level**.**Moreover, the fact that Lake Kastoria is monitored by four telemetric stations, offering continuous and uninterrupted data sets (a few hundred thousand), support the use of Deep Neural Networks, that need large amount of data in order to complete their learning process, and enhance the originality of the present PhD dissertation.**

Keywords: Machine Learning, DO model, Deep Neural Network, Geographic Information System, Water Quality, Lake Kastoria

Περίληψη

Οι λίμνες αποτελούν σημαντικούς δείκτες της κλιματικής αλλαγής. Από γεωγραφικής άποψης, αποτελούν τα χαμηλότερα σημεία της λεκάνης απορροής μιας εσωτερικής λεκάνης, συγκεντρώνοντας τα νερά από την επιφανειακή απορροή του υδρογραφικού δικτύου. Οι φυσικές, χημικές και βιολογικές δραστηριότητες που λαμβάνουν χώρα,παρέχουν μια ποικιλία ανεκτίμητων πληροφοριών. Τα τελευταία χρόνια, τα υδάτινα σώματα έχουν υποστεί εκτεταμένες αλλαγές ως αποτέλεσμα της ευρείας ποιοτικής και ποσοτικής τους υποβάθμισης. Όσον αφορά την υποβάθμιση της ποιότητας, τα πρώτα σημάδια ρύπανσης εντοπίστηκαν στα επιφανειακά ύδατα των λιμνών και των ποταμών. Επομένως, η ανάγκη συλλογής, μελέτης και αξιολόγησης των πληροφοριών που παρέχονται από τη λίμνη είναι υψίστης σημασίας.

Ωστόσο, σε ένα μηνιαίοσύστημα παρακολούθησης, ακόμη και αν οι μετρήσεις γίνονται σε συγκεκριμένη ώρα με πολύ αυστηρό και συγκεκριμένο χρονοδιάγραμμα, ο κίνδυνος απώλειας δεδομένων σε περιπτώσεις αιχμής, όπως τα σημειακά ή μη φαινόμενα ρύπανσης που οφείλονται σε φυσικούς ή ανθρωπογενείς παράγοντες, είναι σχεδόν αναπόφευκτος. Συνεπώς, τα συγκεκριμένα φαινόμενα (peakevents), είναι αδύνατον να καταγραφούν, στην περίπτωση που η μέτρηση της εκάστοτε παραμέτρου πραγματοποιείται σε μια συγκεκριμένη ημερομηνία και ώρα, ενώ το φαινόμενο αιχμής, εμφανίζεται μερικές ώρες ή ημέρες αργότερα και πριν από την επόμενη μέτρηση. Αυτός ο κίνδυνος μπορεί να αποφευχθεί μόνο με μια συνεχή, τηλεμετρική και ολοκληρωμένη παρακολούθηση σε 24ωρη βάση, με σύγχρονα πρωτόκολλα ταχείας εγγραφής και τηλεμετάδοσης.

Τα τελευταία χρόνια, πολλά μοντέλα έχουν εφαρμοστεί για την πρόβλεψη παραμέτρων ποιότητας νερού με Μηχανική Μάθηση (ML). Η Μηχανική Μάθηση, μια εφαρμογή τεχνητής νοημοσύνης, είναι ένας ταχέως αναπτυσσόμενος τομέας. Τα Τεχνητά Νευρωνικά Δίκτυα (ANN), που μιμούνται τον ανθρώπινο εγκέφαλο, είναι σε θέση να καθορίσουν τις μη γραμμικές σχέσεις μεταξύ των μεταβλητών που είναι χαρακτηριστικές των οικοσυστημάτων. Σήμερα, τα Βαθιά Νευρωνικά Δίκτυα (DNN), ως μια αναβαθμισμένη έκδοση των Τεχνητών Νευρωνικών Δικτύων (ANN), έχουν βελτιώσει εντυπωσιακά τη σύγχρονη τεχνολογία σε πολλούς επιστημονικούς τομείς των περιβαλλοντικών ερευνών. Δεδομένου ότι τα ΒαθιάΝευρωνικά Δίκτυα(DNN) απαιτούν έναnμεγάλο όγκο δεδομένων προκειμένου να ολοκληρώσουν τη διαδικασία εκμάθησής τους, σε

σχέση με τα Τεχνητά Νευρωνικά Δίκτυα (ANN), είναι σαφές ότι η συνεχής παρακολούθηση των ποιοτικών παραμέτρων της λίμνης διαδραματίζει καθοριστικό ρόλο στη διαχείρισήτης.

Στην παρούσα Διδακτορική Διατριβή, η περιοχή μελέτης περιλαμβάνει την λεκάνη απορροής της λίμνης Καστοριάς, η οποία εντοπίζεται στην Περιφέρεια Δυτικής Μακεδονίας στην Ελλάδα και προστατεύεται από Οδηγίες, Κανονισμούς και Διεθνείς Συμβάσεις. Η λίμνη της Καστοριάς,αποτελεί μια ρηχή, πολυμικτική λίμνη, με έντονες γεωργοκτηνοτροφικές δραστηριότητες στην λεκάνη απορροής της. Προκειμένου να καταγραφούν τα χαρακτηριστικά ποιότητας του νερού (στην επιφάνεια της λίμνης), εγκαταστάθηκαν τέσσερις (4) Τηλεμετρικοί Σταθμοί σε συγκεκριμένες τοποθεσίες στη Λίμνη Καστοριά και συγκεκριμένα οι σταθμοί: Γκιόλε, Σταυρός, Ψαράδικα και Τοιχίο. Για τους σκοπούς τις Διδακτορικής Διατριβής χρησιμοποιήθηκαν δεδομένα παραμέτρων ποιότητας για το χρονικό διάστημα από τις 11 Νοεμβρίου 2015 έως τις 15 Μαρτίου 2018 σε ωριαία βάση από τους τέσσερις τηλεμετρικούς σταθμούς (περίπου 20000 μετρήσεις ανά σταθμό) που βρίσκονται στην περιοχή μελέτης. Η διαθέσιμη βάση δεδομένων αποτελείται από τις ακόλουθες ποιοτικές παραμέτρους: 1) Χλωροφύλλη-α – Chl-a (μg/L), 2) Ενεργός Οξύτητα – pH, 3) Θερμοκρασία Νερού – Tw (°C), 4) Ηλεκτρική Αγωγιμότητα του Νερού – ECw(μS/cm), 5) Θολερότητα – Turb(NTU), 6) Νιτρικό Άζωτο – N-NO_3(mg/L),7) Αμμωνιακό Άζωτο – N-NH_4(ppm) (Μη διαθέσιμη πληροφορία για το σταθμό του Σταυρού) και 8) Διαλυμένο Οξυγόνο – DO(mg/L).

Για να προσεγγιστεί και να κατανοηθεί καλύτερα η κατάσταση της περιοχής μελέτης, παρουσιάζονται με τη βοήθεια των Γεωγραφικών Συστημάτων Πληροφοριών (GIS): α) οι χρήσεις γης και οι μεταβολές τους για τα έτη 1945 και 2000, και β) η Ντετερμινιστική Μέθοδος Σταθμισμένης Παρεμβολής Αντίστροφης Απόστασης (IDW). Τα Γεωγραφικά Συστήματα Πληροφοριών αποτελούν ένα ισχυρό εργαλείο για τη δημιουργία θεματικών χαρτών με οποιεσδήποτε γεωγραφικές πληροφορίες που μπορούν να ποσοτικοποιηθούν. Εδώ, με την υποστήριξη ντετερμινιστικών αλγορίθμων χωρικής παρεμβολής (IDW) και τις διαθέσιμες μετρήσεις παραμέτρων ποιότητας νερού από τέσσερις διαφορετικές θέσεις παρακολούθησης, οι χωρικές πληροφορίες προσομοιώνονται σεολόκληρη την επιφάνεια της λίμνης Καστοριάς, για τους 3 αντιπροσωπευτικότερους μήνες (Ιανουάριος, Μάιος και Σεπτέμβριος). Εξάγονται,έτσι, χρήσιμες πληροφορίες, μέσω θεματικών χαρτώναπεικόνισης διαβάθμισης χρώματος, που μας επιτρέπουν να κατανοήσουμε την κατανομή και τη δυναμική της εκάστοτε παραμέτρου. Με τον τρόπο αυτό, δύναται να εκτιμήσουμε την περιβαλλοντική και οικολογική κατάσταση της λίμνης και να προτείνουμε προληπτικά και διορθωτικά μέτρα.

Βάσει της χρωματικής απεικόνισης των χαρτών που εξήχθησαν από τη χρήση της στάθμισης της Αντίστροφης Απόστασης (Inverse Distance Weighted – IDW) συμπεραίνουμε ότι οι συγκεκριμένοι θεματικοί χάρτες ακολουθούν τη λογική και την ορθότητα των φυσικών μηχανισμών και βιολογικών διεργασιών, και συμβαδίζουν κατά περίπτωση με τις καταγεγραμμένες τιμές των τηλεμετρικών σταθμών.

Συγκεκριμένα, η χρωματική απεικόνιση του χάρτη της παραμέτρου της Θερμοκρασίας του Νερού – Τwυποδηλώνει ότι οι χαμηλότερες τιμές θερμοκρασίας παρατηρούνται στο βόρειο και δυτικό τμήμα της λίμνης Καστοριάς και στους τρεις μήνες αναφοράς, όπως και αναμένονταν γιατί στη θέση αυτή είναι το βαθύτερο σημείο της λίμνης. Παρόλα αυτά, η λίμνη δεν εμφανίζει στρωμάτωση. Αντίθετα, οι υψηλότερες τιμές Τwεμφανίζονται στο νοτιοανατολικό τμήμα της λίμνης. Η χαμηλότερη τιμή θερμοκρασίας εντοπίζεται τον μήνα Ιανουάριο, ενώ η υψηλότερη τον μήνα Σεπτέμβριο, εννοείται σε ότι αφορά τη σύγκριση μεταξύ των τριών παραπάνω μηνών.

Σε ό,τι αφορά την παράμετρο Διαλυμένου Οξυγόνου, τον μήνα Ιανουάριο, οι υψηλότερες τιμές παρατηρούνται στο βορειοανατολικό και νοτιοδυτικό τμήμα της λίμνης, όπου η κατάσταση αυτή εδραιώνεται επιπλέον στο νοτιοδυτικό τμήμα, τον μήνα Μάιο. Η κατάσταση αυτή στηρίζεται στο γεγονός της ύπαρξης χειμάρρων που εκβάλλουν στη λίμνη από τα βόρεια, ανατολικά και λιγότερο από τα νότια και εμπλουτίζουν τη λίμνη με νερό τη χειμερινή και ανοιξιάτικη περίοδο. Τον μήνα Σεπτέμβριο, οι υψηλότερες τιμές Διαλυμένου Οξυγόνου περιορίζονται στο νοτιοανατολικό τμήμα της λίμνης, και οφείλεται κυρίως στην κίνηση των υδάτων από το συχνή λειτουργία (άνοιγμα) των θυρίδων του καναλιού Γκιόλε, όπου παρατηρείται εκροή υδάτων από τη λίμνη για λόγους που έχουν να κάνουν με τη διαχείριση του φαινομένου του ευτροφισμού καθώς και σε καταστάσεις αιχμής εξαιτίας αυξημένης στάθμης νερού, πάνω από το επιτρεπτό όριο (629,80 μ). Γενικότερα, οι τιμές του Διαλυμένου Οξυγόνου, παραμένουν σταθερές κατά την διάρκεια του εξεταζόμενου έτους.

Αναφορικά με την παράμετρο της Χλωροφύλλης, τον μήνα Ιανουάριο, η υψηλότερη τιμή εμφανίζεται νοτιοανατολικά, όπου είναι η περίοδος που οι θυρίδες του καναλιού Γκιόλε δεν λειτουργούν, ενώ τα υπόλοιπα τμήματα της λίμνης φέρουν πολύ χαμηλότερες τιμές. Στον μήνα Μάιο, οι χαμηλές τιμές της χλωροφύλλης-α περιορίζονται στο βορειοδυτικό τμήμα της λίμνης, ενώ οι μεγαλύτερες τιμές παρατηρούνται στο βορειοανατολικό και νότιο τμήμα αυτής, λόγω της ύπαρξης χειμάρρων που εκβάλλουν στη λίμνη και της εντατικοποίησης των γεωργοκτηνοτροφικών δραστηριοτήτων της ευρύτερης περιοχής. Τον μήνα Σεπτέμβριο οι υψηλότερες τιμές χλωροφύλλης-α συνεχίζουν να εμφανίζονται στο βορειοανατολικό τμήμα της λίμνης, κατάσταση που δεν παρατηρείται τόσο έντονα στο νότιο τμήμα, λόγω της συχνής λειτουργίας (άνοιγμα) των θυρίδων του καναλιού Γκιόλε, οπότε και αποφορτίζεται η λίμνη από μια σημαντική ποσότητα ανόργανου και οργανικού φορτίου

Αναφορικά με την αγωγιμότητα, οι υψηλότερες τιμές εμφανίζονται στο νοτιοδυτικό τμήμα της λίμνης. Αντιθέτως, οι χαμηλότερες τιμές αγωγιμότητας παρατηρούνται στο νοτιοανατολικό τμήμα της λίμνης. Τους μήνες αναφοράς, Ιανουάριο και Σεπτέμβριο, το βόρειο τμήμα της λίμνης φέρει ένα μέτριο και χαμηλό προφίλ, όσον αφορά την παράμετρο της αγωγιμότητας, σε αντίθεση τον μήνα Μάιο που παρατηρείται μια τάση αύξησης της αγωγιμότητας στο συγκεκριμένο τμήμα.

Σύμφωνα με την παράμετρο του pH, η χρωματική απεικόνιση των χαρτών για τους τρεις μήνες αναφοράς, παρουσιάζεται παρόμοια. Οι υψηλότερες τιμές pH, και για τους τρεις μήνες αναφοράς, εντοπίζονται στο νοτιοανατολικό τμήμα της λίμνης, ενώ στο βόρειο και βορειοανατολικό τμήμα της λίμνης παρατηρούνται τιμές που τείνουν να προσεγγίσουν τις τιμές του νοτιοανατολικού τμήματος. Οι χαμηλότερες τιμές pH, εμφανίζονται συνήθως στο νοτιοδυτικό τμήμα της λίμνης, αλλά οι διαφορές με τους υπόλοιπους σταθμούς είναι τόσο μικρές που δε χρήζουν κάποιας ερμηνείας.

Σε ό,τι αφορά την θολερότητα, όπως συμβαίνει και στην παράμετρο του pH, η χρωματική απεικόνιση των χαρτών για τους τρεις μήνες αναφοράς, παρουσιάζεται παρόμοια. Οι υψηλότερες τιμές θολερότητας, και για τους τρεις μήνες αναφοράς, εμφανίζονται στο δυτικό τμήμα της λίμνης και συγκεκριμένα στη Βόρεια και Νότια Παραλία, λόγω του αστικού φορτίου που δέχονται τα συγκεκριμένα τμήματα και του ευτροφισμού. Αντιθέτως, οι χαμηλότερες τιμές θολερότητας, και για τους τρεις μήνες αναφοράς, παρουσιάζονται στο βορειοανατολικό, ανατολικό και νοτιοανατολικό τμήμα της λίμνης.

Τέλος, αναφορικά με τις παραμέτρους των Νιτρικών και Αμμωνιακών Ιόντων και πιο συγκεκριμένα του Νιτρικού και Αμμωνιακού Αζώτου, οι διαφοροποιήσεις οφείλονται στη λειτουργία (άνοιγμα) των θυρίδων του καναλιού Γκιόλε, αλλά και στις πιέσεις που δέχεται λιμναίο οικοσύστημα από το αστικό φορτίο και τις γεωργοκτηνοτροφικές δραστηριότητες.

Επίσης, η χρήση των Γεωγραφικών Συστημάτων Πληροφοριών,συμβάλλειστη δημιουργία μιας βάσης δεδομένων με όλα τα γεωχωρικά και υδρολογικά χαρακτηριστικά που αφορούν:α) τις εκθέσεις επιφανειών, β) τις ισοϋψείς καμπύλες με ισοδιάσταση 50 μέτρων, γ) τα υψόμετρα, δ) τις κλίσεις, ε) τηνακτογραμμή της λίμνης, στ) το υδρογραφικό δίκτυο ζ) τα όρια λεκάνης απορροής και η)τις θέσεις τηλεμετρικών σταθμών. Ακολουθεί η κατασκευή θεματικών χαρτών για την περιοχή μελέτης κάνοντας χρήση συγκεκριμένων χαρτών του Υπουργείου Γεωργίας όπως: α) ο γεωλογικός χάρτης του Ι.Γ.Μ.Ε του έτους 1983, β) ο βιοκλιματικός χάρτης και γ) ο χάρτης βλάστησης του Ιδρύματος Δασικών Ερευνών Αθηνών του έτους 1978. Επιπλέον, κατασκευάζονται χάρτες του αναθεωρημένου δικτύου προστατευόμενων περιοχών Natura 2000 και των χρήσεων γης (CorineLandCover) πρώτου και τρίτου επιπέδου, χρησιμοποιώντας γεωχωρικά δεδομένα από τον εθνικό κατάλογο "geodata.gov.gr". Από το χάρτη της μεταβολής των χρήσεων γης και τη σύγκριση των δεδομένων του 1945 με το 2000, έχουν προκύψει μεταβολές στις χρήσεις γης. Ωστόσο, παρά την αύξηση των γεωργικών και αστικών πηγών ρύπανσης, η λειτουργία του βιολογικού καθαρισμού από το 1995, έχει μειώσει σημαντικά το φαινόμενο του ευτροφισμού.

Εν συνεχεία, εξετάζεται η ικανότητα των Βαθιών Νευρωνικών Δικτύων (DNN) να προβλέπουν σε πραγματικό χρόνο τις παραμέτρους DO, Chl-a και Turbστη Λίμνη Καστοριά. Τα εξεταζόμενα Εμπρόσθιας τροφοδοσίας Νευρωνικά Δίκτυα (Feed-ForwardDNN) του DO, ελέγχονται για τέσσερις δομές με βάση τον αριθμό των κόμβων που χρησιμοποιούνται στα κρυμμένα επίπεδα και τον αριθμό των μεταβλητών εισόδου. Πιο συγκεκριμένα, το μοντέλοDO με δομές:

α) 7-64-64-1 για τους σταθμούς Γκιόλε, Ψαράδικα και Τοιχίο και 6-64-64-1 για το σταθμό Σταυρό, όπου δεν είναι διαθέσιμη η παράμετρος N-NH_4,

β) 4-64-64-1 για όλους τους σταθμούς,

γ) 7-32-32-1 για σταθμούς Γκιόλε, Ψαράδικα και Τοιχίοκαι 6-32-1 για το σταθμό Σταυρό, όπου η παράμετρος N-NH_4 δεν είναι διαθέσιμη και

δ) 4-32-32-1 για όλους τους σταθμούς.

Διερευνώνται, επίσης, τα μοντέλα

ε) Turbμε δομές 7-64-64-1 για τους σταθμούς Γκιόλε, Ψαράδικα και Τοιχίο και 6-64-64-1 για το σταθμό Σταυρό και

στ) Chl-aμε τις ίδιες ακριβώς δομές όπως οι παραπάνω.

Για την ανάπτυξη των **Βαθιών Νευρωνικών Δικτύων**– με βάση μεθόδους **Μηχανικής Μάθησης** – χρησιμοποιούνται τα εξής:

1) η γλώσσα προγραμματισμού **Python**,
2) το επιστημονικό περιβάλλον **Spyder**,
3) η πλατφόρμα εκμάθησης μηχανών ανοικτού κώδικα **Tensorflow** και
4) η παράλληλη πλατφόρμα **CUDA** (Compute Unified Device Architecture) της **Nvidia**

Όλα τα παραπάνω έχουν ενσωματωθεί στην **Anaconda**, μια γλώσσα προγραμματισμού ανοιχτού κώδικα.Η βέλτιστη αρχιτεκτονική δικτύου για κάθε σταθμό, επιλέγεται με βάση τους ελάχιστους στατιστικούς δείκτες MAE, MSE και NSE. Τα εξεταζόμενα μοντέλα Turbidityκαι Chl-a, με δομές 7-64-64-1 για τους σταθμούς Γκιόλε, Τοιχίο και Ψαράδικα, καθώςκαι 6-64-64-1 για το σταθμό Σταυρός, δεν έδειξαν καλή ικανότητα πρόβλεψης σε πραγματικό χρόνο. Η αδυναμία των μοντέλων δύναται να εξηγηθεί λόγω της πιθανής έλλειψης ικανότητας του συγκεκριμένου δικτύου να χειριστεί μια τόσο μεγάλη διακύμανση στα δεδομένα. Αντιθέτως, το καλά εκπαιδευμένο μοντέλο DO, με δομή 7-32-32-1 δίνει καλά αποτελέσματα με MAE 0,54; 0,55, MSE 0,65; 0,68 και NSE 0,89; 0,86 για το στάδιο της εκπαίδευσης και της δοκιμής αντίστοιχα, με καλή ικανότητα πρόβλεψης για το σταθμό Γκιόλε. Για το σταθμό Τοιχίο, η δομή 7-64-64-1 επικρατεί με MAE 0,48; 0,49, MSE 0,52; 0,51 και NSE 0,92; 0,93 στο στάδιο της εκπαίδευσης και της δοκιμής αντίστοιχα και αποτελεί τη δομή με τις καλύτερες επιδόσεις συγκριτικά με τις υπόλοιπες εξεταζόμενες δομές για όλους τους σταθμούς. Για τον σταθμό Ψαράδικα, η δομή 4-64-64-1 δίνει τα καλύτερα αποτελέσματα με MAE 0,57; 0,58, MSE 0,65; 0,68 και NSE 0,91; 0,92 στο στάδιο της εκπαίδευσης και της δοκιμής αντίστοιχα. Τέλος, ο σταθμόςΣταυρός,δίνει τα βέλτιστα αποτελέσματα στη δομή 6-32-1 MAE 0,69; 0,70, MSE 0,98; 1,01 και NSE 0,91; 0,90 για την διαδικασία της εκπαίδευσης και τη δοκιμής αντίστοιχα.

Η παρούσα διατριβή αποσκοπεί στον εντοπισμό προβλημάτων που ενδέχεται να προκύψουν, παρέχοντας ένα ισχυρό εργαλείο άμεσης λήψης αποφάσεων (DSS) για τη διαχείριση της ποιότητας των υδάτων της λίμνης, την πρόβλεψη των παραμέτρων ποιότητας σε πραγματικό χρόνο και την πρόληψη ατυχημάτων και καταστάσεων έκτακτης ανάγκης, που αφορούν τόσο Φυσικούς όσο και Ανθρωπογενείς Κινδύνους. Σύμφωνα με την βιβλιογραφία και τη γνώμη του συγγραφέα, η καινοτομία της παρούσας Διδακτορικής Διατριβής έγκειται στο γεγονός ότι **δεν παρουσιάζεται άλλη δημοσιευμένη εργασία που να χρησιμοποιεί τις ίδιες πλατφόρμες, εργαλεία και μεθοδολογία της Βαθιάς Μάθησης, προκειμένου να διερευνηθεί η ικανότητα πρόβλεψης παραμέτρων ποιότητας νερού λιμνών,τόσο σε εθνικό όσο και σε διεθνές επίπεδο.Επίσης, η πλήρης κάλυψη μιας λίμνης με τέσσερις τηλεμετρικούς σταθμούς σε συνεχή λειτουργία και πλήρη και αδιάλειπτη παροχή δεδομένων, της τάξης των μερικών εκατοντάδων χιλιάδων σε αριθμό, υποστηρίζουν τη χρήση των μοντέλων Βαθιάς Μάθησης,ενισχύουν και ενδυναμώνουν τη διατριβή και την πρωτοτυπία της.**

Λέξεις-κλειδιά: Μηχανική Μάθηση, Μοντέλο Διαλυμένου Οξυγόνου, Βαθύ Νευρωνικό Δίκτυο, Γεωγραφικά Συστήματα Πληροφοριών, Ποιότητα Νερού, Λίμνη Καστοριάς

List of Contents

List of Figures

CHAPTER 4: LITERATURE REVIEW

CHAPTER 5: GEOGRAPHIC INFORMATION SYSTEMS

CHAPTER 6: THE STRUCTURE OF DEEP NEURAL NETWORKS

CHAPTER 7: RESULTS AND DISCUSSION

List of Tables

List of Abbreviations

ANN:	Artificial Neural Network
BM:	Bayesian Model
BPNN:	Back Propagation Neural Network
BP:	Back Propagation
DL:	Deep Learning
DON:	Dissolved Organic Nitrogen
DSS:	Decision Support System
DT:	Decision Tree
DNN:	Deep Neural Network
EL:	Ensemble Learning
GHG:	Greenhouse gas
GIS:	Geographic Information System
GRNN:	General Regression Neural Network
IBM:	Instance Based Models
IDW:	Inverse Distance Weighting
IR:	Infrared
JMD:	Joint Ministerial Decision
LRM:	Linear Regression Model
MAE:	Mean Absolute Error
ML:	Machine Learning
MLP:	Multi-Layer Perceptron
MSE:	Mean Square Error
NSE:	Nash-Sutcliffe Efficiency
PAR:	Photosynthetically Active Radiation
PON:	Particulate Organic Nitrogen
RBNN:	Radial Basis Neural Network
ReLU:	Rectified Linear Unit
RMSprop:	Root Mean Square propagation
RNN:	Recurrent Neural Network
SVM:	Support Vector Model
SVR:	Support Vector Regression
UV:	Ultraviolet
VLSI:	Very Large Scale Integration

CHAPTER 1: INTRODUCTION

"The principle of all things is water;
All comes from water, and to water all returns".
Thales of Miletus, (635-543 BC)

1.1 Lake as an indicator of climate change

Lakes constitute important indicators of climate changeand geographically they comprise a network of the lowest points of an internal catchment's area, collecting water from the surrounding hydrographic network. Their physical, chemical and biological responses to the climate give a variety of priceless information (Figure 1.1) (Williamson et al., 2009).

Lakes are affected directly by changes in climate:

a) due to changes in mixing regime, including lake stratification, oxygen saturation by increase in temperature, and the frequency of extreme wind events,

b) by changes in trophic structure determined by temperature and

c) by complex interactions between temperature, nutrients, and physical forces (Jeppesen et al., 2009).

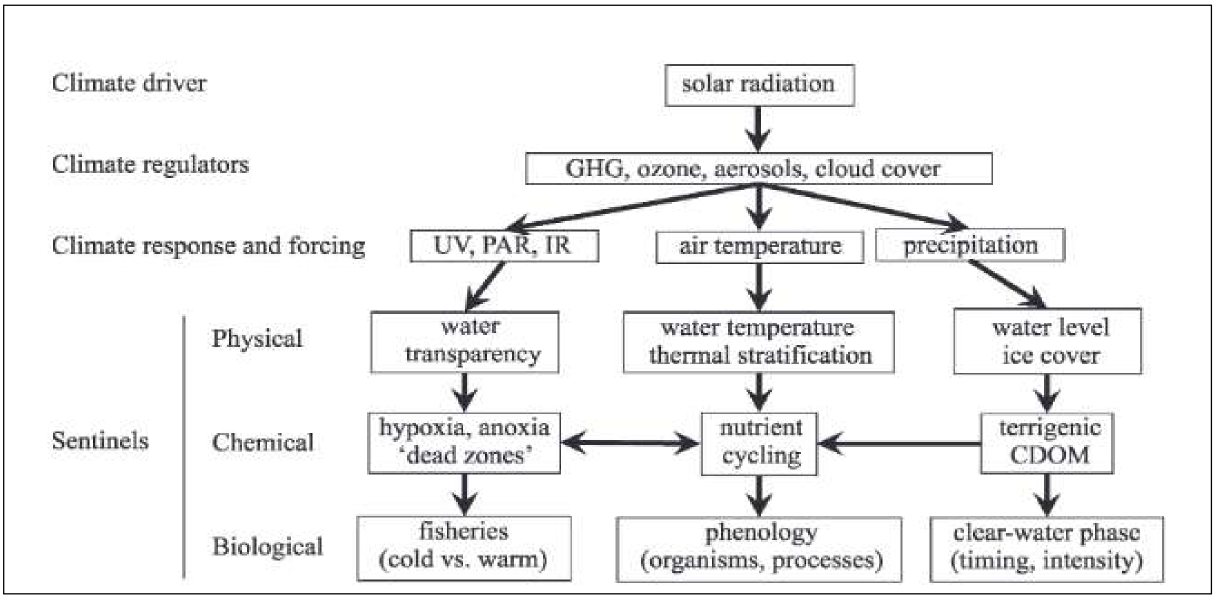

Figure 1.1: Flow chart of lakes as sentinels of climate change (Williamson et al., 2009)

Lakes can be considered as sentinels, integrators and regulators of climate change (Figure 1.2). Lakes are referenced as sentinelsas (Adrian et al., 2009):

- ✓ Lake ecosystems are well defined and are studied in a sustained fashion

✓ Respond directly to climate change and also incorporate the effects of climatedriven changes occurring within the catchment

✓ Integrate responses over time, which can filter out random noise

✓ Are distributed worldwide and, as such, can act as sentinels in many different geographic locations and climatic regions, capturing different aspects of climate change such as rising temperature.

Moreover, lakes are considered as integrators as they store signals of change in their sediments—integrating changes not only within theaquatic ecosystem, but also changes in the surrounding terrestrial ecosystem.

Also, lakes are referenced as regulators as they:

✓ Receive, process, and store large amounts of carbon from the surrounding terrestrial watershed as well as from the aquaticproductivity within their shorelines,

✓ Are involved in active exchange of greenhouse gases with the overlying atmosphere, and

✓ Alter regional climate by changing radiative forcing, cloud formation, precipitation, and evaporation (Williamson et al., 2009).

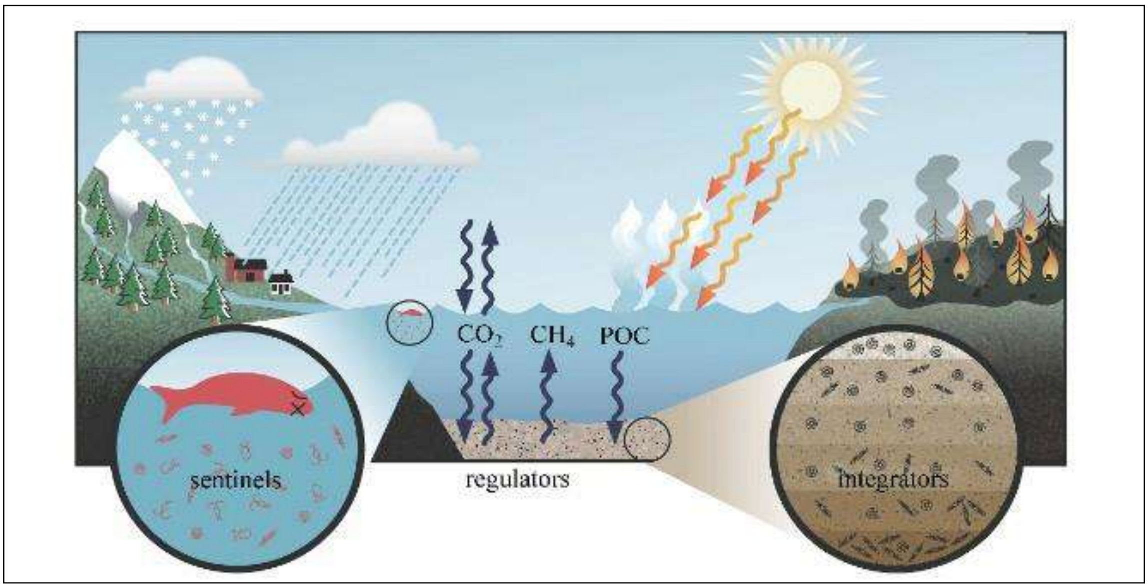

Figure 1.2: Lake as sentinel, integrator and regulator of climate change (Williamson et al., 2009).

1.2 Thesis background

In recent years, waterbodies have undergone extensive change as a result of widespread qualitative and quantitative degradation. In terms of quantitative degradation, water is wasted unreasonably for its various uses such as agriculture, industry, energy and domestic use.

Regarding quality deterioration, the first signs of pollution were identified in the surface waters of lakes and rivers, during the decade of ‘20s and the first models of Dissolved Oxygen depletion have been formulated, such as Streeter and Phelps (1925). Afterwards, the groundwater became gradually contaminated with organic and nutrients compounds. It should be mentioned that in case of rivers there is sufficient natural buffering capacity, but less for lakes and wetlands, provided the presence of pollutants that are biodegradable and non-preservatives. On the other hand, decontamination of an aquifer requires special treatment during which the aquifer becomes inactive and its full restoration is doubtful (Psilovikos, 1996).

It becomes clear that the need for collecting, studying and evaluating the information given from the lake is of paramount importance. Therefore, the need for monitoring and modeling lakes’ physical, chemical and biological responses is crucial (Williamson et al., 2009).

Machine Learning (ML), an application of Artificial Intelligence, is a rapidly developing field. To perform a task, Machine Learning comprises a learning process with an objective to learn and improve from experience (training data). A datasetin ML (also known as features or variables) can be qualitative (nominal, ordinal, binary) or quantitative (discrete or continuous). In order to calculate the performance of ML models, statistical and mathematical models are used. At the end of the learning process, the trained model can be used to predict new examples from the testing data with the help of the experience gained during the training procedure (Liakos et al., 2018). One clearly understands the importance of data sets (Monitoring) in order to apply Machine Learning techniques.

In a developing world where water is extensively polluted, many models have been applied to predict water quality parameters with the help of Machine Learning. Artificial Neural Networks (ANN), that mimic human brains, are able to map the non-linear relationships among variables that are characteristic of ecosystems (Lek et al., 1996). That is the reason why Artificial Neural Network Techniques have been applied in many case studies (Chen and Wang, 2020; Mellios et. al., 2020) to predict water quality parameters such as Dissolved Oxygen (Ranković et al., 2010; Akkoyunluet al., 2011; Ay and Kisi, 2012; Han et al., 2012; Wen et al., 2013), Turbidity (Fan et al., 2016; Khairi et al., 2016; Cao et al., 2019) and Chl-a (Muttil and Chau, 2006; Kuo et al., 2007; Cho et al., 2014) in rivers and lakes. Nowadays, specifically, Deep Neural Networks, as an upgraded and complex version of Artificial Neural Network, have dramatically improved the state-of-the-art in many scientific fields such as environmental studies.

1.3 Objectives and scope of thesis

Regarding how much Machine Learning has expanded nowadays, monitoring systems play a decisive role in lake management by providing systematic information of the qualitative parameters. Moreover, as a Deep Neural Network needs a large amount of data in order to complete its learning process comparatively to Artificial Neural Networks, it is clear that Monitoring is of paramount importance for applying Machine Learning techniques and especially Deep Learning.

The present dissertation provides a scientific insight into:

- ✓ The importance of Monitoring in lake management
- ✓ The implementation of Machine Learning techniques and especially Deep Neural Networks in Water Quality Management.

Based on the scientific research, the ability of Deep Neural Networks to predict the quality parameter of Dissolved Oxygen, Chlorophyll-a and Turbidity in Lake Kastoria, in Greece, is tested. In order to reach the aforementioned goal, quality time series from November 11, 2015 to March 15, 2018, on an hourly basis, from four telemetric stations located in the study area, are used (around 20.000 measurements per parameter per station).

Geographic Information Systems (GIS) is a very powerful tool in creating thematic maps with any geographical information that can be quantified. So, if water quality parameters measurements are available froma deteriorated number of different monitoring locations, with the support of Spatial Interpolation Algorithms, the limited spatial information can be integrated to the whole surface of a water body. Useful spatial information is extracted, providing with thematic color gradient maps that allow us to:

a) Geographically simulate the spatial distribution and variability of parameters on the surface of a water body – lake.

b) Understand the correlation of this particular parameter with other similar ones, in different locations

c) Assess the environmental and ecological status of the lake

d) Propose preventive and remedial measures.

With the help of GIS, in order to approximate and understand the state of the study area, a) the land use changes observed from 1945 to 2000 and b) the recordings of the available water quality parameters and their spatial distribution all over the surface of Lake Kastoria, are presented and discussed. In the case of Lake Kastoria, a deterministic interpolation algorithm is applied for the simulation of the geographic distributeion of the qualitative parameters of the water**,** namely Inverse Distance Weighted method (IDW). The reason why it is preferred to use deterministic method, instead of geostatistic ones, is that it can lead to successful results with very few monitoring points (Matzafleri et al., 2009; Zisou & Psilovikos, 2012).Moreover, useful information is extracted that allow us to understand, the consequences of the Land Use Land Cover (LULC) changes from 1945 to 2000, in the water quality.Despite the fact that agricultural and urban sources of pollution have increased, the operation of biological treatment plant since 1995, has significantly reduced the phenomenon of eutrophication.

The present PhD thesis is intended to:

- ✓ Identify problems that may arise
- ✓ Provide a useful supportive tool for Water Quality Management of the Lake
- ✓ Provide real-time prediction of water parameters
- ✓ Comprise a powerful Decision Support System, for preventing accidental and emergency conditions that may arise from both Natural and Anthropogenic Hazards.

1.4 Thesis innovation

Based on the literature and the knowledge of the author, the importance of this work lies within the fact that there is no other published work using the same platforms, tools and methodology of Deep Learning, as in the present PhD thesis, in order to investigate the predictive capacity of water quality parameters of lakes, at both national and international level. Moreover, the fact that Lake Kastoria is monitoredby four telemetric stations, offering continuous and uninterrupted data sets (a few hundred thousand measurements), support the use of Deep Neural Networks, that need large amount of data in order to complete their learning process, and enhance the originality of the present PhD dissertation.

1.5 Thesis outline

This section describes the structure of this PhD Thesis. In the introduction it is explained briefly why lakes constitute an important indicator of climate change. Chapter 1 gives the background, the objectives and scope of the thesis underlying the importance of collecting, studying and evaluating the information given from the lake via data sets (Monitoring) in order to apply Machine Learning techniques.

The following chapters (Chapter 2 and Chapter 3) provide the main body of the Thesis background. Specifically, Chapter 2 describes in detail the classifications of Machine Learning focusing in Artificial Neural Networks and Deep Neural Networks, while Chapter 3 clarifies the importance of Monitoring, the types of the existing Monitoring and introduces the case study of the PhD thesis.

Chapter 4 refers to a) the recording and evaluation of the existing investigations and studies with respect to the monitoring and the sustainable management of Lake Kastoria (case study) b) the recording of the existing policy background that has been implemented in Greece for water resources management, water quality conservation and restoration, and the protection of aquatic ecosystems and c) the recording implementation of technical neural networks in order to evaluate the water quality of aquatic systems.

Chapter 5 introduces the case study (Lake Kastoria) in relation to the land use changes observed from 1945 to 2000 and presents the recordings of the available water quality parameters by using theInverse Distance Weighting method (IDW) with the help of Geographic Information System.

Chapter 6 gives in detail information about the structure of the models, the tools and platforms, the statistical descriptors and in general the whole procedure used in this thesis.

Chapter 7 contains the results and discussion of this PhD thesis, while the closing chapter comprises the conclusion of this thesis and its main contributions followed by suggestions for future work.

References

Adrian, R., O'Reilly, C.M., Zagarese, H., Baines, S.B., Hessen, D.O., Keller, W., Livingstone, D.M., Sommaruga, R., Straile, D., Van Donk, E. and Weyhenmeyer, G.A., 2009. Lakes as sentinels of climate change. *Limnology and oceanography*, 54, pp.2283-2297.

Akkoyunlu, A., Altun, H. and Cigizoglu, H.K., 2011. Depth-integrated estimation of Dissolved Oxygen in a lake. *Journal of EnvironmentalEngineering*, 137(10), pp.961-967.

Ay, M. and Kisi, O., 2012. Modeling of dissolved oxygen concentration using different neural network techniques in Foundation Creek, El Paso County, Colorado. *Journal of Environmental Engineering*, 138(6), pp.654-662.

Cao, P., Zhao, W., Liu, S., Shi, L. and Gao, H., 2019. Using a Digital Camera Combined with Fitting Algorithm and TS Fuzzy Neural Network to Determine the Turbidity in Water. *IEEE Access*, 7, pp.83589-83599.

Chen, J. and Wang, Y., 2020. Comparing Activation Functions in Modeling Shoreline Variation Using Multilayer Perceptron Neural Network. *Water*, 12(5), p.1281.

Cho, S., Lim, B., Jung, J., Kim, S., Chae, H., Park, J., Park, S. and Park, J.K., 2014. Factors affecting algal blooms in a man-made lake and prediction using an Artificial Neural Network. *Measurement*, 53, pp.224-233.

Fan, S.K.S., Kuan, W.H., Fan, C. and Chen, C.Y., 2016. Rainfall threshold assessment corresponding to the maximum allowable turbidity for source water. *Water Environment Research*, 88(12), pp.2285-2291.

Han, H.G., Qiao, J.F. and Chen, Q.L., 2012. Model predictive control of dissolved oxygen concentration based on a self-organizing RBF neural network. *Control Engineering Practice,*20(4), pp.465 – 476.

Jeppesen, E., Kronvang, B., Meerhoff, M., Søndergaard, M., Hansen, K.M., Andersen, H.E., Lauridsen, T.L., Liboriussen, L., Beklioglu, M., Özen, A. and Olesen, J.E., 2009. Climate change effects on runoff, catchment phosphorus loading and lake ecological state, and potential adaptations. *Journal of Environmental Quality*, 38(5), pp.1930-1941.

Khairi, M.T.M., Ibrahim, S., Yunus, M.A.M., Faramarzi, M. and Yusuf, Z., 2016. Artificial neural network approach for predicting the water turbidity level using optical tomography. *Arabian Journal for Science and Engineering*, 41(9), pp.3369-3379.

Kuo, J.T., Hsieh, M.H., Lung, W.S. and She, N., 2007. Using artificial neural network for reservoir eutrophication prediction. *Ecological modelling*, 200(1-2), pp.171-177.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996. Application of Neural Networks to modelling nonlinear relationships in ecology. *Ecological Modelling,*90, pp.36–52.

Liakos, K.G., Busato, P., Moshou, D., Pearson, S. and Bochtis, D., 2018. Machine Learning in agriculture: A review. *Sensors*, 18(8), pp.2674.

Matzafleri, N., Psilovikos, A. and Mplanta, A., 2009. Water Quality Monitoring and Modeling in Lake Kastoria, using GIS. Assessment and Management of Pollution Sources. *Water Resources Management*, 23(15), pp.3221 – 3254.

Mellios, N., Moe, S. and Laspidou, C., 2020. Machine Learning Approaches for Predicting Health Risk of Cyanobacterial Blooms in Northern European Lakes. *Water*, 12(4), p.1191.

Muttil, N. and Chau, K.W., 2006. Neural network and genetic programming for modelling coastal algal blooms. *International Journal of Environment and Pollution*, 28, pp.223 – 238.

Psilovikos, A., 1996. Optimum management in aquifer studies using the Linear Programming (LP) method. An application to Eidomeni–Evzones aquifer. MSc Thesis, Aristotle University of Thessaloniki, School of Agriculture, 154 p.

Ranković, V., Radulović, J., Radojević, I., Ostojić, A. andČomić, L., 2010. Neural network modeling of dissolved oxygen in the Gruža reservoir, Serbia. *Ecological Modelling*, 221(8), pp.1239-1244.

Streeter, H.W. and Phelps, E.B., 1925. A Study of the Pollution and Natural Purification of the Ohio River, III, Factors Concerned in the Phenomena of Oxidation and Re-aeration. US Public Health Service.*Public Health Bulletin*,146, p.75.

Wen, X., Fang, J., Diao, M. and Zhang, C., 2013. Artificial neural network modeling of dissolved oxygen in the Heihe River, Northwestern China. *EnvironmentalMonitoring and Assessment*, 185(5), pp.4361-4371.

Williamson, C.E., Saros, J.E., Vincent, W.F. and Smol, J.P., 2009. Lakes and reservoirs as sentinels, integrators, and regulators of climate change. *Limnology and Oceanography*, 54(6part2), pp.2273-2282.

Zisou, C. and Psilovikos, A., 2012. Monitoring and Modeling of the Physicochemical Parameters Geographic Distribution in Lake Karla using GIS. Proceedings of the 14th Conference of the Hellenic Hydrotechnical Association, pp.1250 – 1261, Patra, Greece.

CHAPTER 2: MACHINE LEARNING APPROACHES

"Much learning does not teach understanding".
Heraclitus, (535 – 475 BC)

2.1 Machine learning background

Machine Learning (ML) is the phenomenon that a system improves its performance when performing a specific task, without the need to re-program it. According to this definition, Machine Learning is designed to create machines capable of learning, improving their performance in certain areas through the use of prior knowledge and experience. A relevant general definition of Machine Learning is given by Mitchell (1997): "A computer program says that it learns from experience E with some class of tasks T and measure of performance P, if its performance on tasks from T, as measured by P, is improved through experience E."

Machine Learning, as a part of Artificial Intelligence, studies algorithms that improve their behavior in a task assigned to them using their experience. As for the design of Machine Learning systems, the ability to identify is defined as the ability to acquire additional knowledge, which alters existing registered knowledge either changing its features or increasing its position. In the case of Artificial Neural Networks, learning is the ability of systems to transform their internal structure, rather than altering the knowledge that is incorporated therein their design.

Although we are far from creating machines that learn as well as humans, algorithms have been developed for specific areas of learning that have enabled modern commercial applications to emerge with considerable success.

Machine Learning is classified into broad categories depending on the learning type (supervised/unsupervised) and the learning models (regression/classification) (Liakos et al., 2018). Based on the learning type (Figure 2.1), ML is classified in:

a) Supervised and

b) Unsupervised learning.

Specifically, in supervised learning, there is a training set consisting of data points x (from some set X) and their labels y (from some set Y) and the goal is to learn a function f: $X \rightarrow Y$, that will accurately predict the labels of data points arising in the future (Sammut and Webb, 2017).In other words, supervised learning is the process where the algorithm constructs a function that represents given inputs (training set) at known desired outputs, with the ultimate goal of generalizing this function to inputs of unknown output. It is commonly used in regression, classification and interpretation problems. In his review, Solé (2007) underlines that"Applications in which the training data comprises examples of the input vectors along with their corresponding target vectors are known as supervised learning problems".

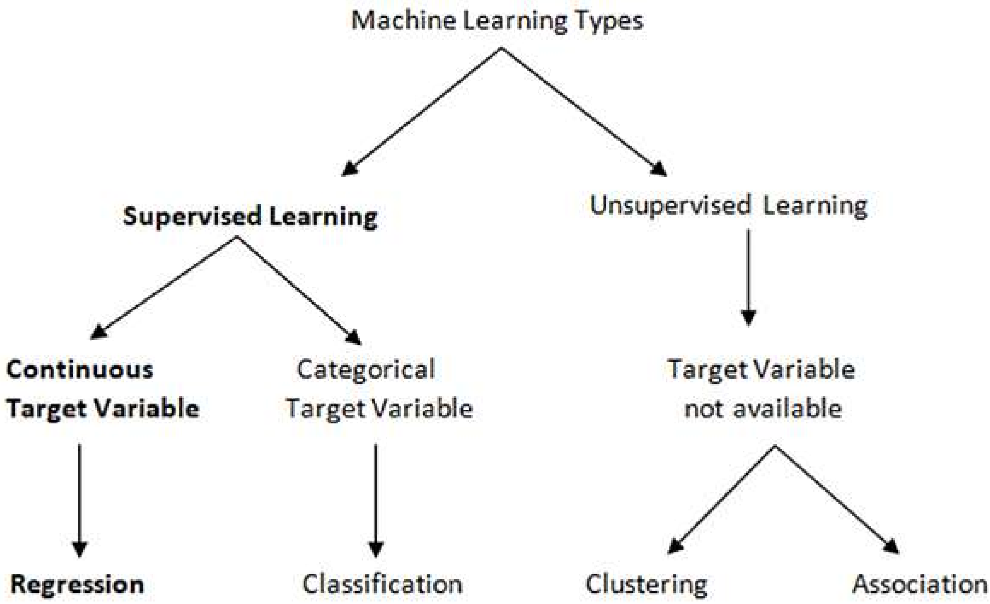

Figure 2.1: Machine Learning Types

However, in unsupervised learning, the target variable is not available so there is no distinction between training and test sets with data being unlabeled. The learner processes input data with the goal of discovering hidden patterns (Liakos et al., 2018). Otherwise, in unsupervised learning, the algorithm constructs a model for a set of inputs in the form of observations without knowing the desired outputs. It is used in clustering and association analysis problems. "In unsupervised learning, there is no instructor or teacher, and the algorithm must learn to make sense of the data without this guide" (Goodfellow et al., 2016).

Regression and classification are supervised learning approaches while clustering is an unsupervised learning approach. Regression predicts continuous valued output while classification predicts discrete number of values. In classification the data is categorized under different labels according to some parameters and then the labels are predicted for the data. Clustering separatesthe dataset into groups, called clusters. For each problem to be solved in the field of Machine Learning, there is an appropriate way of learning and for each way of learning there is at least one appropriate algorithm that can be used (Figure 2.10).

The most important phase of each algorithm is the Training phase, where the algorithm uses as an input a training set to achieve its purpose which is the creation of new knowledge. In addition, it can either use more or less existing knowledge or not at all. Training is followed by the certification phase of new knowledge produced. Usually, the authentication is performed first by the algorithm itself through recall procedures with the help of test data and then by the user's critique. Finally, new knowledge is given to use in applications where it is needed to solve real problems. Based on learning models, artificial neural networks are described in detail. In addition to the ANNs, the present dissertation refers to other models used to predict water quality (Georgouli, 2005).

2.2 Regression

Regression is a supervised learning model, which predicts an output according to known inputs. The most known algorithms of regression models are linear regression, logistic regression, stepwise regression and ordinary least squares regression (Figure 2.10).

2.3 Clustering

Clustering is a typical application of unsupervised learning model, that groups the dataset into clusters. The scope is to partition the unlabeled data in such a way that points within single cluster are very similar and points in different clusters are different. It determines grouping among unlabeled data.

2.4 Bayesian Models

Bayesian models (BM) are probabilistic graphical models in which the analysis is undertaken within the context of Bayesian inference (Hobbs and Hooten, 2015). This type of model belongs to the supervised learning category and can be used for either classification or regression problems. The Bayesian inference can be applied to regression problems. It combines one’s prior knowledge with the information contained in the data (Elster and Wübbeler, 2015).

When Bayesian inference is applied to classification problems it is assumed that all items of evidence are conditionally independent. Such models are commonly referred to as 'naive' Bayes, because the assumption is considered naïve. In bio-monitoring, the use of 'naive' Bayes model fits well asthe assumption that the occurrence of each taxon is independent of that of the others, given the water quality class, holds for the vast majority of taxa (Walley and Džeroski, 1996; Fox and Roberts, 2012).

2.5 Instance Based Models

Instance based models (IBM) are memory-based models. They store the training database and when a new input vector is presented, a set of similar related instances is retrieved from memory. Afterwards, their corresponding outputs are used to predict the output for the new query vector (instance). The disadvantage of these models is that their complexity grows with data (Solomatine et al., 2008). The k-nearest neighbor algorithm is the most used algorithm in this category.

2.6 Decision Trees

Decision Trees (DT) is the most well-known supervised model and has been successfully applied in many areas. Decision Trees are classification or regression models formulated in a tree-like architecture. The DT algorithm leads to the creation of a tree whose leaves are classes. This tree form can also be read as a set of rules called classification rules and provide a convincing answer to the question: How can a machine make general rules from specific observations and how reliable are these rules in practice?

The set of data is organized into a smaller homogeneous subset while producing a relevant tree graph. Each internal node represents a different pairwise comparison of a selected feature, while each branch represents the result of this comparison. The leaf nodes represent the final decision/prediction made after the root-to-leaf path (Liakos et al., 2018). The central focus of the decision tree growing algorithm is selecting which attribute (also known as features or input variables) to test at each node in the tree (Saghebian et al., 2014).

2.7 Support Vector Machines

Support vector machine (SVM) is a binary classifier that makes a linear separating hyperplane to classify data instances (Liakos et al., 2018). A hyperplane is a line that splits the input variable space and is selected to best separate the points in the input variable space by their class, either class 0 or class 1.

The distance between the hyperplane and the closest data points is referred to as the margin. The best or optimal hyperplane that can separate the two classes is the line that has the largest margin. Only these points are relevant in defining the hyperplane and in the construction of the classifier. These points are called the support vectors. They support or define the hyperplane (Figure 2.2) (Ng, 2000).

2.8 Ensemble Learning

Ensemble learning (EL) models scope to improve the predictive performance of a given statistical technique by creating a linear combination of simpler base learner. Considering that each trained ensemble represents a single hypothesis, these multiple-classifier systems enable hybridization of hypotheses not caused by the same base learner, providing better results in the case of significant diversity among the single models (Liakos et al., 2018). Random forest algorithm is a well-known algorithm in this category.

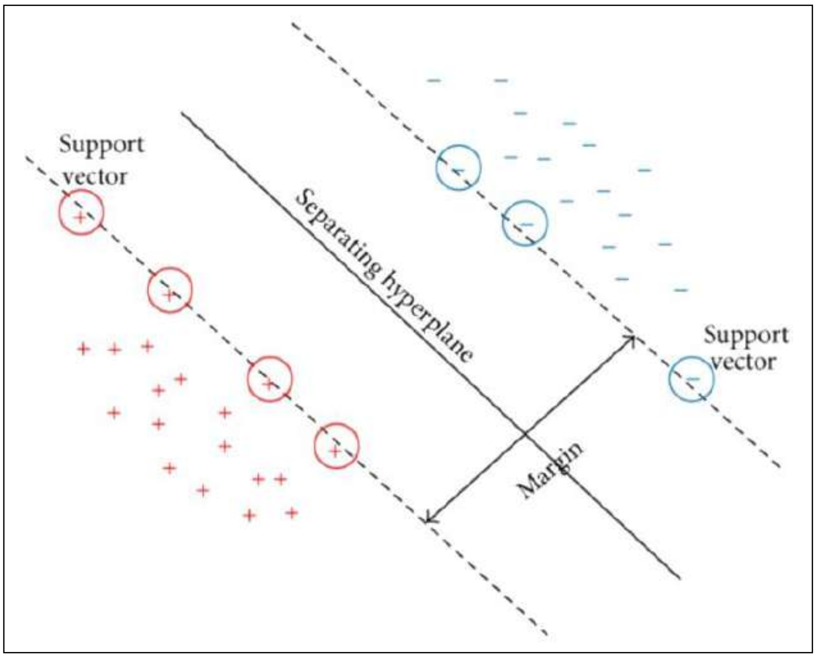

Figure 2.2: Definition of hyperplane via supports vectors (Li and Heap, 2008)

2.9 Artificial Neural Networks

The term Neural Networks (Connectionist Networks, Parallel Distributed Processing Models) describes a number of different mathematical models, inspired by corresponding biological models, that is, models that attempt to mimic the behavior of neurons in the human brain (Figure 2.3).

From 19thcentury, scientists believe that the brain is made up of distinct elements, neurons that communicate to each other. Neurons are the basic building block of the human brain. It is estimated that the brain contains about 10 billion neurons grouped into groups, each of which constitutes a natural neural network. Thus, the human brain contains hundreds of natural neural networks, each containing thousands of interconnected neurons with an average number of interconnections per neuron between 1000 and 10000 (Georgouli, 2005).

A neuron is separated from the rest of the cells by a membrane and has the ability to transmit electrical signals from that neuron to the other neurons it communicates with. Each neuron consists of 3 main parts (Kotopouli et al., 2009):

1. Dendrites, which act as input channels for the neuron,
2. The main cell body,
3. The axon, which connects one neuron to other neurons.

The axon of a single neuron transmits signals to the dendrites of neighboring neurons via the point of union called synapse. A neuron can receive signals from a set of neighboring neurons through the dendrites, process them, and power its output through the axon to another set of adjacent neurons. The signals coming through the dendrites are 'weighed' and the results are summed. When the sum exceeds the threshold level (threshold value), the neuron generates an output on its axon, which will then be transmitted to the neighboring neurons via synapses (Georgouli, 2005; Psilovikos, 2020).

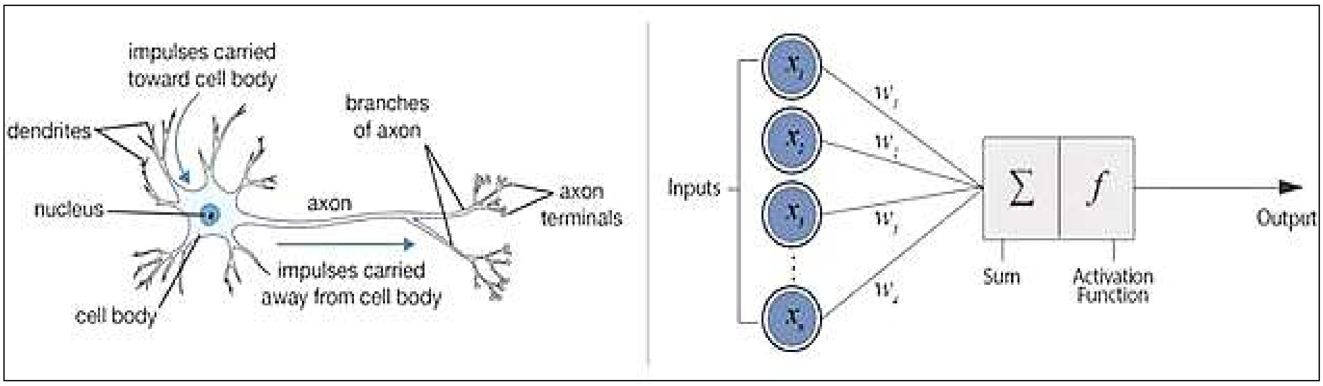

Figure 2.3: Biological Neuron versus Artificial Neural Network (Njogholo, 2018)

The mathematical models of Artificial Neural Networks, in correspondence with biological ones, consist of a number of simple and highly interconnected layers. Each artificial neuron consists of several inputs xi and only one output y. Each input x_i is 'weighed' with a weight w_i and the results are summed by the summation function F (Equation 2.1):

$$F = \sum_{\mathrm{i}}^{\mathrm{n}} \mathrm{xi} * \mathrm{wi} \quad (2.1)$$

The artificial neuron gives outputs through the transfer function only when the weighted sum of inputs is greater than a certain threshold value θ, that is, when (Equation 2.2):

$$\sum_i^n xi * wi - \theta > 0 \qquad (2.2)$$

Artificial Neural Networks (ANNs) are divided into Traditional ANNs and Deep ANNs and are typically used for both regression and classification problems (Sammut and Webb, 2017). Deep ANNs are most widely referred to as Deep Learning (DL) or Deep Neural Networks (DNNs). It is important to note here that **DNN's are simply an ANN with more than one hidden layer between the input and output layers.**

A brief description of ANNs:

- ✓ ANNs are usually organized into layers, where the learning procedure takes place.
- ✓ Layers consist of a number of units or nodes that are interconnected in such way that one unit is linked to many other units of the same or another layer.
- ✓ The units influence other units by stimulating or inhibiting their activation. To achieve this, the unit receives the weighted sum of all inputs through the links leading to it and produces a single output through the transition function if the sum exceeds a threshold value.
- ✓ Inputs are presented to the network via the input layer that communicates with one or more hidden layers The hidden layers are associated with the output layer from where the decision /prediction is given.

The key elements of the architecture of the ANNs that need to be defined when creating them are:

- ✓ The number of intermediate hidden layers,
- ✓ The number of units (or nodes) per layer,
- ✓ How units connect to each other,
- ✓ The activation function,
- ✓ The algorithms (optimizers) used to strengthen the links between units during the training procedure (Figure 2.10).

2.9.1 Architecture of Artificial Neural Networks

In terms of how the nodes are connectedto each other, there are two main categories:

1. Feed forward and
2. Feed backward

In a Feed Forward Neural Network (single or multi – layer) (Figure 2.4), the nodes are organized at different layers so that the nodes of a layer feed the next-layernodes until the last-layer nodes are fed. So, if there is no feedback from the outputs towards the inputs, the network is called Feed forward neural network. Such ANNs arereferred as Back-Propagation (BP) networks. According to Sazli (2006), Back-Propagation algorithm is the most used algorithm in the training of Feed-Forward Neural Networks. However, it should be clarified here that in the context of machine learning by referring to an algorithm, we mean a ''Learning Algorithm'' including the Optimization. So, a Back-Propagation is used as a sub-routine in the Optimization Algorithm (for example a Gradient Descent Learning Algorithm) to find the parameters of the model. However, in some cases, one can use Gradient Descent (Psilovikos & Tzimopoulos, 2004) without using Back propagation such as in logistic or linear regression, where the gradients can be calculated directly from the inputs and cost function value(Sazli, 2006).

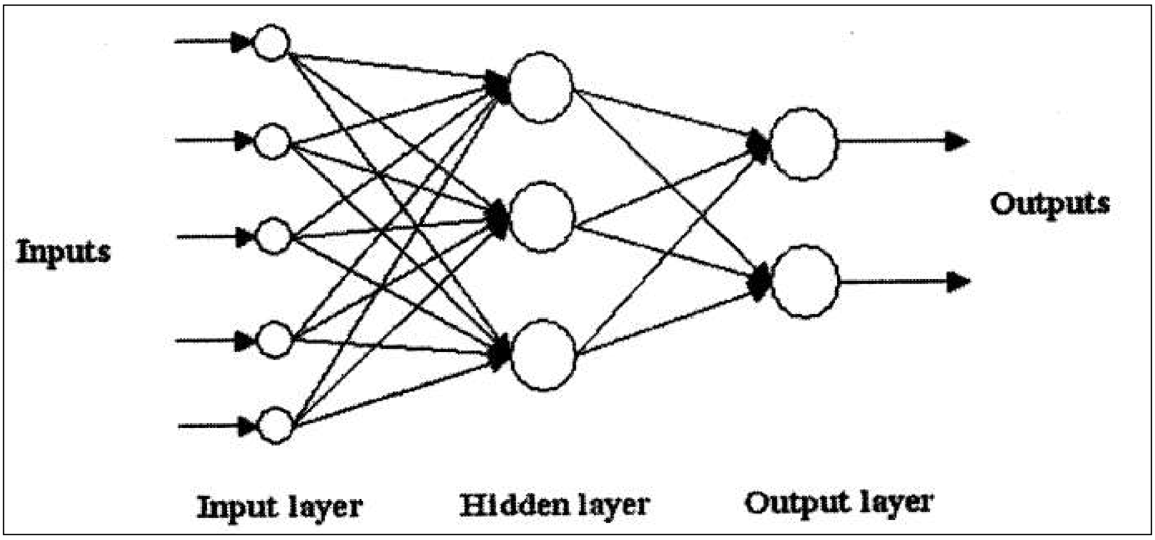

Figure 2.4: A multi – layer feed forward neural network (Sazli, 2006)

In the case of Feed Backward (or recurrent) networks, the nodes of a layer feed the nodes of the same layer or previous layer. If feedback relates to nodes at the same layer, then the networks are called auto-associated memories, differently they are called hetero-associated memories (Figure 2.5). Moreover, there is usually only one hidden layer. Although Feed Backward networks are very useful, Feed Forward networks are most commonly used(Network, 2018).

2.9.2 Advantages of Artificial Neural Networks

ANNs are significantly different from other computing techniques as (Karamoutsou and Psilovikos, 2019):

✓ They do not use symbols to represent model concepts

✓ They do not program the behavior of the model

✓ They are programmed to learn to produce specific outputs with given inputs.

The advantages of ANNs are summarized below:

Design:

A Neural Network is designed according to the human brain. Engineers derive from neurobiology new ideas for solving complex problems. ANNs are also designed to provide information not only on the particular set of models with which it is trained but also on any new model shows up. ANNs are uniform in analysis and design, meaning thatthe same symbolism is used in all scientific fields thereby permitting the diffusion of relevant know-how.

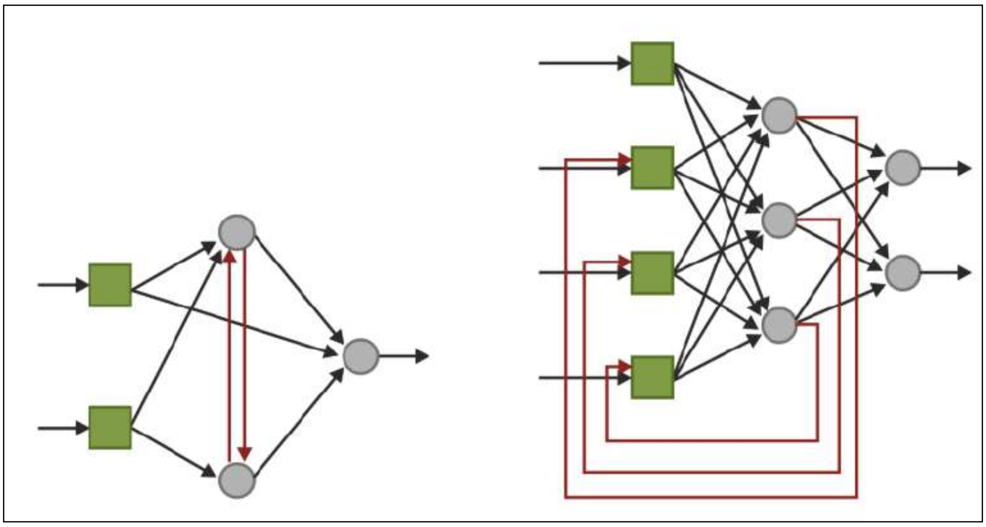

Figure 2.5: Feed backward neural networks: auto-associated memories (left) and hetero-associated memories (right) (Georgouli, 2005)

Implementation:

An ANN is constructed by the link of neurons, which are considered non-linear devices. Non-linearity is a very important property if the physical mechanism for producing the input signal is non-linear. Moreover, knowledge, in ANNs, is represented by a scattered format rather than a technically structured symbolic format.

Structure:

ANNs have the ability to adapt their weights to changes in their environment by simply repeating their training with a new set of examples from the changing environment. A hardware-implemented ANN has the important property of being "fault-tolerant". For example, if an intermediate level connection is deleted or a weight is corrupted, the network operation will not be affected because the overall average error will not change significantly. The layer structure of ANNs enables parallel functions and enables them to be implemented in VLSI (Very-Large-Scale Integration) technology that makes them suitable, among other technologies, to be incorporated into real-time applications (Georgouli, 2005).

2.9.3Deep Neural Networks (DNNs)

Deep ANNs are most widely referred to as Deep Learning (DL) or Deep Neural Networks (DNNs). DNN's are simply an ANN with at least two hidden layers between the input and output layers. Depending on the nature of the problem, it can be supervised, partially supervised, or even unsupervised.

A deep learning system is self-teaching. The computer trains itself to learn from datain a similar way to humans. That is one of its main advantages and has dramatically improved the state-of-the-art in many scientific fields (Liakos et al., 2018). It should be mentioned here that this is possible due to the layers of ANNs.

As it is described above, the simplest Artificial Neural Network consists of one hidden layer. In case that an ANN has more than three layers, including input and output layer, is called Deep Neural Network. In other words, without Neural Networks, there would not be Deep Learning.

Differences between Artificial Neural Networks and Deep Neural Networks are listed below (Georgouli, 2005; Koutsoukas et al., 2017):

- ✓ The structure of the networks comprises the primary difference between Artificial Neural Networks and Deep Neural Network as described above (Figure 2.6)
- ✓ Deep Neural Networks need large amount of data in order to complete their learning process comparatively to Artificial Neural Networks.In Deep Neural Network learning process is time consuming
- ✓ Deep Neural Network is also called theBlack Box, because the learning process in Deep Learning is very hard to explain
- ✓ Deep Neural network is more accurate

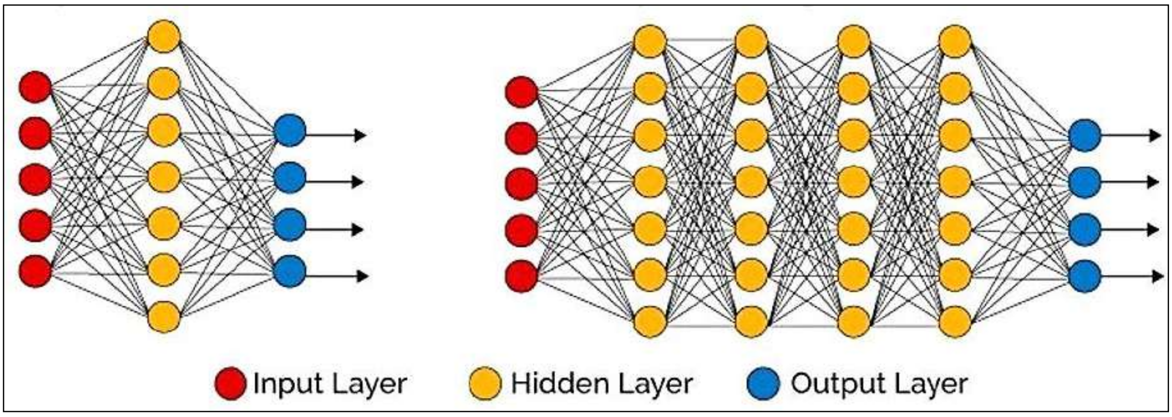

Figure 2.6: Artificial Neural Network (left) vs Deep Neural Network (right)

2.9.4 Activation function

Activation functions play an important role in neural networks. Their goal is to determine (Sentas, 2013; Psilovikos, 2020):

a) The output of a learning model,

b) Its accuracy, and

c) The computational efficiency of training a model.

The function is attached to each neuron in the network, and determines whether it should be activated or not, based on whether each neuron's input is relevant for the model's prediction. Depending on their type, activation functions help to normalize the output of each neuron to a range between 1 and 0 or -1 and 1. Activation functions also have a major effect on the neural network's ability to converge and the convergence speed.

As mentioned below, one of the goals of activation functions is that they must be computationally efficient as they are calculated across thousands or even millions of neurons for each data sample. Modern neural networks use a technique called back-propagation to train the model, which places an increased computational strain on the activation function, and its derivative function. The inputs, are fed into the neurons in the input layer. Each neuron has a weight, and multiplying the input number with the weight gives the output of the neuron, which is transferred to the next layer (https://missinglink.ai/guides/neural-network-concepts/7-types-neural-network-activation-functions-right/).

Activation function should be differentiable. It should be like this as to perform Back-Propagation optimization strategy while propagating backwards in the network to compute gradients of Error(loss) with respect to Weights and then accordingly optimize weights using

Gradient descend or any other Optimization technique to reduce Error (https://towards datascience.com/activation-functions-and-its-types-which-is-better a9a5310cc8f).

The simplest activation functions are linear (Figure 2.7), such as:

- ✓ Threshold functions,
- ✓ Sign functions,
- ✓ Hard limiter functions and
- ✓ Ramping functions.

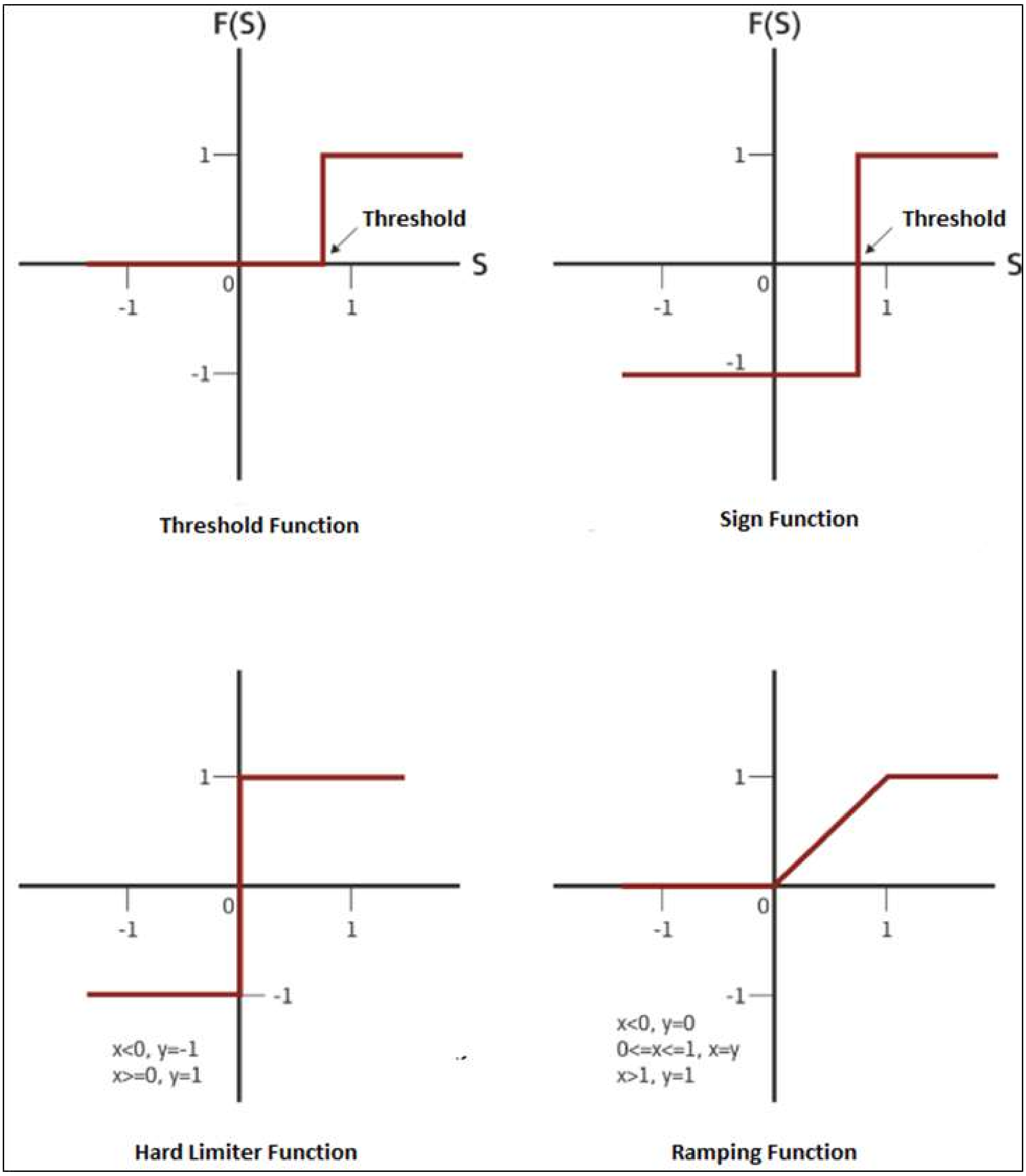

Figure 2.7: Linear activation functions (Georgouli, 2005)

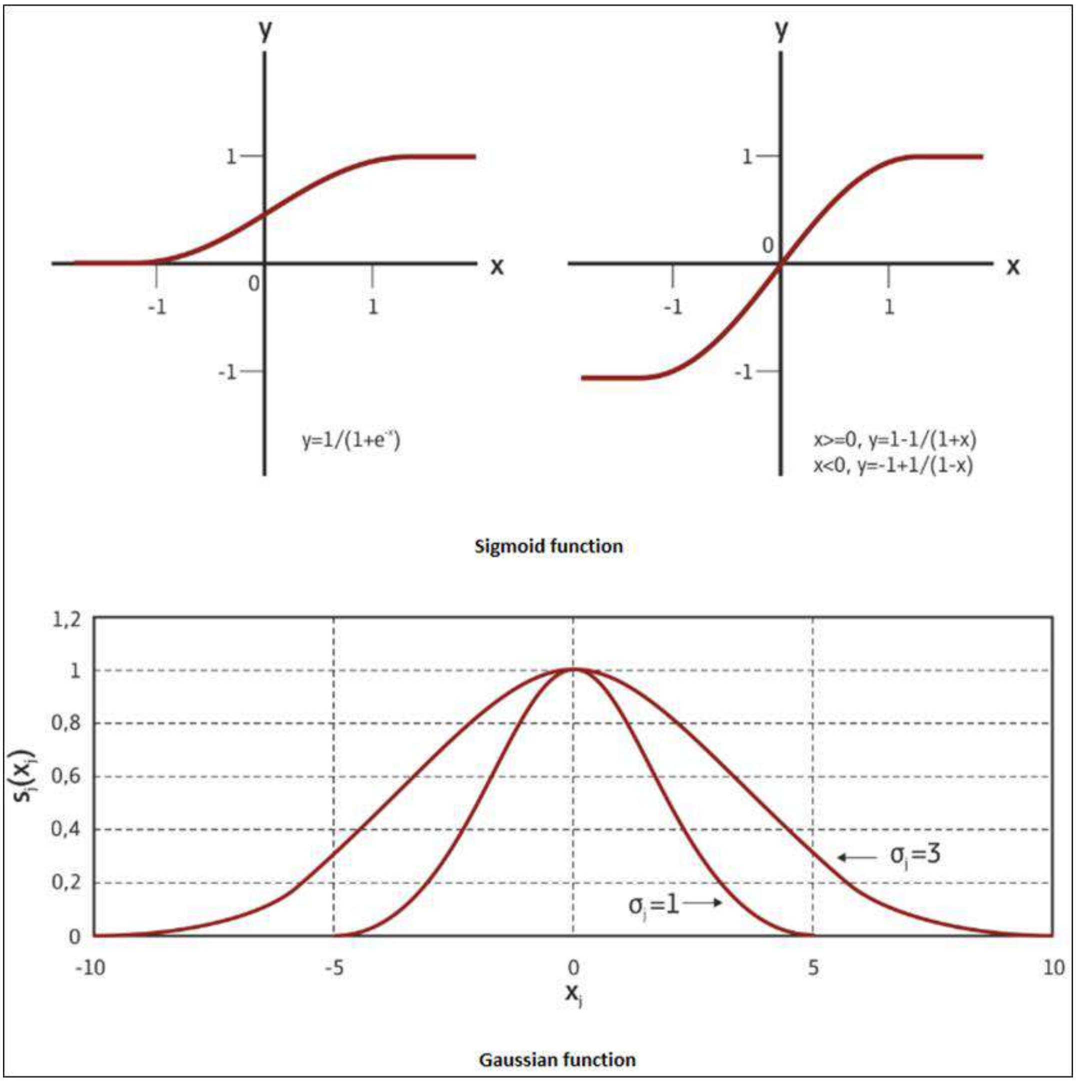

Figure 2.8: Sigmoid and Gaussian activation functions (Georgouli, 2005)

However, nonlinear functions are commonly used, such as sigmoid functions, Gaussian functions (Figure 2.8). The need for speed has led to the development of new functions such as ReLU (Rectified Linear Unit) functions. Non-linear functions address the problems of a linear activation function:

1. They allow Back-Propagation because they have a derivative function which is related to the inputs.
2. They allow "stacking" of multiple layers of neurons to create a Deep Neural Network. Multiple hidden layers of neurons are needed to learn complex data sets with high levels of accuracy.

Nowadays, Sigmoid and ReLU are the most popular types of activation functions used in Machine Learning. Despite its popularity Sigmoid Function causes problems during the training phase and degrades the accuracy of aDeep Neural Network Modeldue to the vanishing gradient problem. On the other hand, ReLUActivation Function avoids and rectifiesvanishing gradientproblem. Nowadays, the majority of Deep Learning Models use ReLU (Dureja and Pahwa, 2019).

However, a disadvantage of using ReLU is that some gradients can be fragile during training and this could result in Dead Neurons. It can cause a weight update which will makes it never activate on any data point again (Epelbaum, 2017).

2.9.5 Regularization techniques

It also should be mentioned that one of the main difficulties when dealing with Deep Learning techniques is to get the Deep Neural Network to train efficiently. A predictive model can be under or over fit if too little or too much data is used to train a Neural Network model. A solution in order to deal with overfitting and under fitting in a proper way is to use the regularization technique https://towards datascience.com/regularization-in-machine-learning 76441ddcf99a).

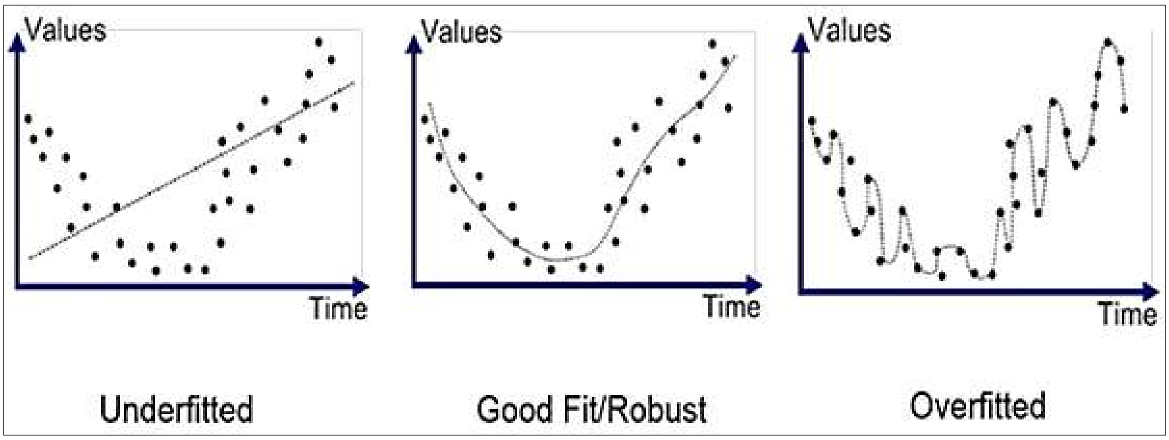

Figure 2.9: Visual representation of how a predictive model can be under or over fit if too little or too much data is used to train a Neural Network model (https://medium.com/)

It should be mentioned that it is also important to study our datasets, before even starting modeling (Figure 2.9). Some commonly used Regularization techniques include (Epelbaum, 2017):

- Early Stopping
- L1 Regularization
- L2 Regularization
- Dropout Regularization
- Training Data Augmentation
- Batch Normalization

- Bayesian regularization BP algorithm

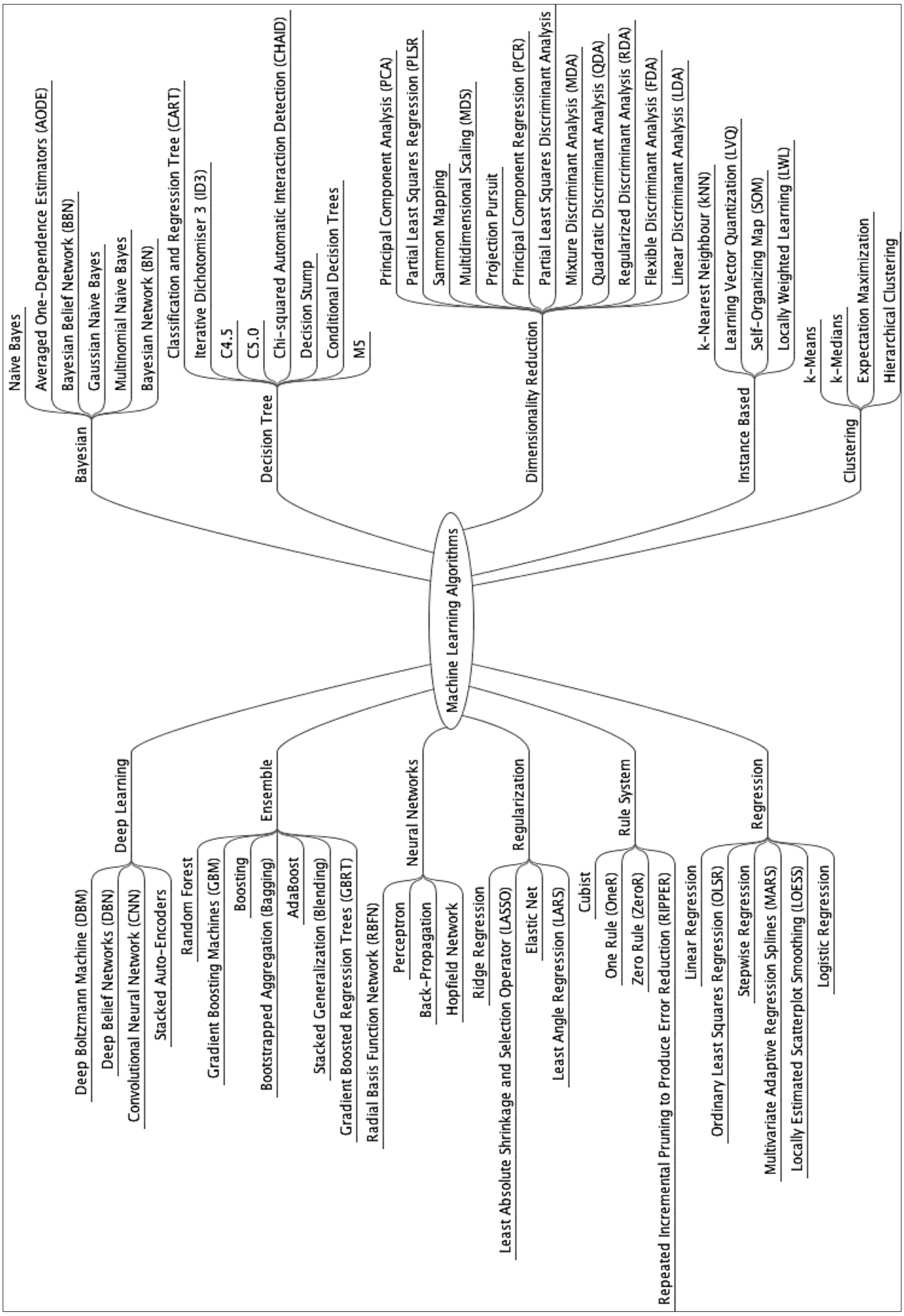

Figure 2.10: Machine Learning algorithmscommonly used in ML models

(https://machinelearningmastery.com/

References

Dureja, A. and Pahwa, P., 2019. Analysis of Non-Linear Activation Functions for Classification Tasks Using Convolutional Neural Networks. *Recent Patents on Computer Science*, 12(3), pp.156-161.

Elster, C. and Wübbeler, G., 2015. Bayesian regression versus application of least squares—an example. *Metrologia*, 53(1), p.10.

Epelbaum, T., 2017. Deep learning: Technical introduction. *ArXiv preprint arXiv:1709.01412*.

Fox, C.W. and Roberts, S.J., 2012. A tutorial on variational Bayesian inference. *Artificial intelligence review*, 38(2), pp.85-95.

Georgouli, C., 2005. Artificial Intelligence - An Introductory Approach. Association of Greek Academic Libraries. ISBN 978-960-603-031-4.

Goodfellow, I., Bengio, Y. and Courville, A., 2016. *Deep learning*. MIT Press, p.102.

Hobbs, N.T. and Hooten, M.B., 2015. Bayesian models: A statistical primer for ecologists. PrincetonUniversityPress.

Karamoutsou, L. and Psilovikos, A., 2019. The Use of Artificial Neural Networks in Predicting the Water Quality of Lake Kastoria. Proceedings of the 14th Conference of the Hellenic Hydrotechnical Association (EYE), pp.882 - 889, Volos.

Kotopouli, M., Psilovikos, A., Gkitsakis, N., Sapountzis, M. and Dimos, G., 2009. Time Series Analysis of Rain Amount at Rainfall Station Nestorio Kastoria using ANN. *Hydrotechnica*, 18 – 19, pp.49 – 64.

Koutsoukas, A., Monaghan, K.J., Li, X. and Huan, J., 2017. Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data. *Journal of cheminformatics*, 9(1), p.42.

Li, H., Leng, W., Zhou, Y., Chen, F., Xiu, Z. and Yang, D., 2014. Evaluation models for soil nutrient based on support vector machine and artificial neural networks. *The Scientific World Journal*, pp.

Liakos, K.G., Busato, P., Moshou, D., Pearson, S. and Bochtis, D., 2018. Machinelearning in agriculture: A review. *Sensors*, 18(8), p.2674.

Mitchell, T.M., 1997. Does machine learning really work? *AI magazine*, 18(3), pp.11-11.

Network, B.A.N., 2018. Botnet Detection Using a Feed-Forward Backpropagation Artificial Neural Network. In *Computational Intelligence in Information Systems: Proceedings of the Computational Intelligence in Information Systems Conference (CIIS 2018)*, Springer, 888, p.24.

Ng, A., 2000. CS229 Lecture notes. *CS229 Lecturenotes*, 1(1), pp.1-3.

Njogholo, B.M., 2018. Doctoral Dissertation in Investigating keystroke dynamics as a two-factor biometric security, Strathmore University.

Psilovikos, A. and Tzimopoulos, C., 2004. Comparison of Quadratic and Non-Linear programming (QP and NLP) Optimization Models (QP and NLP) in Groundwater Management. *Journal of Hydroinformatics*, 6(3), pp.175 – 185.

Psilovikos, A., 2020. Water Resources. Tziolas Pub, Thessaloniki, Greece, 700 p. ISBN: 978-960-418-602-0, (in Greek).

Saghebian, S.M., Sattari, M.T., Mirabbasi, R. and Pal, M., 2014. Ground water quality classification by decision tree method in Ardebil region, Iran. *Arabian journal of geosciences*, 7(11), pp.4767-4777.

Sammut, C. and Webb, G.I., 2017. Encyclopedia of machine learning and data mining. *Springer Publishing Company*, Incorporated, p.32 and 83.

Sazli, M.H., 2006. A brief review of feed-forward neural networks. *Communications, Faculty of Science, University of Ankara*, 50(1), pp.11-17.

Sentas, A., 2013. Stochastic Simulation Models in Surface Water Quality in River Nestos. Doctoral Dissertation in Sustainable Management of Aquatic Environment, Department of Ichthyology and Aquatic Environment, 215 pages.

Solé, J.L., 2007. Book review: Pattern recognition and machine learning. Cristopher M. Bishop. Information Science and Statistics. SORT-Statistics and Operations Research Transactions, *Springer 2006*, 31(2), 738 pages.

Solomatine, D.P., Maskey, M. and Shrestha, D.L., 2008. Instance-based learning compared to other data-driven methods in hydrological forecasting. *Hydrological Processes: An International Journal*,22(2), pp.275-287.

Walley, W.J. and Džeroski, S., 1996. Biological monitoring: a comparison between Bayesian, neural and machine learning methods of water quality classification. In *Environmental Software Systems, Springer*, Boston, MA, pp.229-240.

Internet sources

https://machine learning mastery.com/

https://medium.com/

https://missinglink.ai/guides/neural-network-concepts/7-types-neural-network-activation-functions-right/

https://towardsdatascience.com/activation-functions-and-its-types-which-is-better-a9a5310cc8f

CHAPTER 3: THE IMPORTANCE OF MONITORING

"Quality is not an act, it is a habit."
Aristotle, (384-322 BC)

Regarding water quality deterioration, the first signs of pollution were identified in the surface waters of lakes and rivers. Afterwards, the groundwater became gradually contaminated. It should be mentioned that in case of rivers, there is sufficient natural buffering capacity but less for lakes and wetlands, provided the presence of non-persistent pollutants that are biodegradable. On the other hand, decontamination of an aquifer requires special treatment during which the aquifer becomes inactive and its full restoration is doubtful (Psilovikos, 2020).

The need for collecting, studying and evaluating the information given from the lake and the need for monitoring and modeling their physical, chemical and biological responses is of significantimportance (Williamson et al., 2009).

Both the quantitative and qualitative status of water resources requires a systematic long-term monitoring. Significant efforts have been made in Greece in recent years by various agencies to create databases of surface and groundwater resources measurements. However, these efforts were discontinued after several years (Psilovikos, 2005).

Stakeholders that have previously dealt with water monitoring issues are (Psilovikos, 2020):

✓ Ministries:

a) Greek Ministry of Rural Development and Food

b) Greek Ministry of Economy & Development

c) Greek Ministry of Foreign Affairs

d) Greek Ministry of Health and Social Security

✓ Services:

a) Institute of Geological and Mining Exploration (IGME)

b) Public Power Company (PPC) and

c) Municipal WaterSupply andSewerage Companyof Larissa (DEYAL)

✓ Research institutes:

a) Hellenic Center for Marine Research (HCMR)

b) GreekBiotope/Wetland Centre (EKBY by itsGreekinitials)

c) Fisheries Research Institute (FRI)

d) General Chemical State Laboratory (GCSL)

- ✓ Universities and Technological Educational Institutions, in the context of research and education
- ✓ Individuals, in the context of private researches.

The above efforts were fragmentary in the sense that regular, systematic and organized measurements were not taken, in the context of a) a well-organized protocol, b) a predetermined time schedule with the character of ‘’time-series’’ and c) the points of measurement are not documented due to the real needs of the aquatic ecosystem. The issues raised are two (Psilovikos, 2020):

1. The issue of keeping a specific time and many repetitions during the day and
2. The issue of data scarcity with a frequency of measurement once a month.

During the Monitoring procedure, it is very important that the variation of parameters such as Dissolved Oxygen (DO) and Chlorophyll-a (Chl-a), can be wide-ranging during 24 hours (Psilovikos, 2005; Psilovikos, 2012).

When high concentrations of DO are observed, they mainly occur:

a) At shallow eutrophic lake systems.

b) At late spring - early summer,

c) In the morning and at noon, when high concentrations of DO are observed due to the photosynthetic productivity of algae and/or cyanobacteria which are also associated with correspondingly high concentrations of Chl-a.

d) When they are associated with low values of water temperature, which favors high values of DO of saturation, except in cases that the lake has an ice cap, which favors DO consumption and inability to replenish.

A typical example is the lake of Kastoria, where measured values of DO, higher than those of Oxygen Saturation (DOs), have been recorded during the spring, mainly in May and June.Therefore, due to the intense activity of the photosynthesis from phytoplankton, from all primary producers of the system, high DO values are observed, which are also associated with correspondingly high concentrations of Chl-a.This phenomenon is reversing in the autumn, where the respiration favors the photosynthesis and DO depletion occurs and stress to the biota (Watt, 2000; Matzafleri et al., 2009; Psilovikos, 2014).

When low concentrations of DO are observed, they mainly occur:

a) In the evening and night, where no photosynthetic activities are taken place and DO is consumed for the respiration of micro-organisms and the degradation of non-preservatives;

b) At water depths near the bed, where DO values are expected to be very low and may approach hypoxic conditions (1 - 2 mg/L) or even anoxic conditions (~ 0 mg/L). Anoxia could also enhance phosphorus release from the sediment to the water column, so increased eutrophic conditions could occur.

Apart from the above reasons, low DO values may occur due to the sedimentsOxygen Demand at the bed of a lake, known as Sediment Oxygen Demand (SOD). These low DO values can harm aquatic life leading to stress and affect water quality. For this reason, it is necessary to have intensive measurements with many repetitions on a daily basis, so that one can monitor the evolution and the durationof such a phenomenon. During summer, before the advanced phase of eutrophication, where the consumption of DO, due to respiration activities from benthic biota and invertebrates, is higher than the production of DO, due to photosynthetic activities, concentrations of DO values varywidely (Antonopoulos, 2010; Psilovikos, 2014).

Regarding the issue of data scarcity with measurements on monthly basis, even if the measurements are made at a specific time with a very strict and specific schedule, the risk to loose data on peak events, such as point or non-point pollution phenomena due to natural or anthropogenic factors or flooding phenomena is unquestionable. So, it is clear that these peak phenomena, cannot be recorded if the measurement of the quality parameter is conducted on a specific date and the peak phenomenon occurs a few days later. Till the next monthly measurement, the impacts of the peak phenomena could not be recorded. This risk can only be prevented by continuous automated and organized 24-hour monitoring.

The state of water quality is a complex result of natural and anthropogenic activities and their interactions with time and space. The objectives of the Monitoring are defined and guided by a number of Laws, Ministerial Decisions and Presidential Decrees. Monitoring Networks concern aquatic ecosystems such as springs, streams, rivers, lakes, fjords, estuaries, coastal waters and seawater. There are two main types of Monitoring (Wiersma, 2004; Psilovikos, 2020):

- ✓ The **Extended Monitoring**, which includes many sampling stations but receives a few samples per year, which analyzes a few parameters and lasts from one year to 3-4 years and
- ✓ The **Intensive Monitoring**, which includes more stations, depending on the size of the study area, many annual samples (at least 12, one each month) that are conducted for over 4 years.

The parameters that are measured are:

- ✓ Water Temperature (Tw - ^oC)
- ✓ pH

- ✓ Electrical Conductivity of Water (EC_w - μS/cm),
- ✓ Dissolved oxygen (DO - mg/L or ppm)
- ✓ Saturation Oxygen, as the fraction of Dissolved Oxygen / Saturation Oxygen (DOs - %)
- ✓ Redox potential (Eh or REDOX – mV)
- ✓ Turbidity (T - NTU)
- ✓ Organic Pollution Indicators such as DO, BOD, COD and SOD
- ✓ Eutrophication indices, such asTotal Phosphorus (TP – mg/L or ppm), Secchi Disk Depth (SD - m), Chlorophyll-a (Chl-a – mg/L) andTSI Indicators by Carlson (1977)
- ✓ Nitrate ions (NO_3^-) or Nitrate Nitrogen ($N\text{-}NO_3$ – mg/L)
- ✓ Nitrite ions (NO_2^-) or Nitrite Nitrogen ($N\text{-}NO_2$ – mg/L)
- ✓ Ammonium ions (NH_4^+) or Ammonium Nitrogen ($N\text{-}NH_4$ – mg/L)
- ✓ Orthophosphate ions (PO_4^{-3}) or Orthophosphate Phosphorus ($PO_4^{-3}\text{-}P$ – mg/L)
- ✓ The ratio of Nitrogen / Phosphorus (N / P)
- ✓ Alkalinity indices (pH, alkalinity, nitrates etc.)
- ✓ Ion Detection (Chlorine, Sulfur, Magnesium) (Cl, S, Mg - mg/L or ppm)
- ✓ Detection of Heavy Metals (Cadmium, Mercury, Copper, Zinc) (Cd, Hg, Cu, Zn - mg/L or ppm)
- ✓ Organic Pollutants (PCB, HCH)
- ✓ Radioactive Indicators
- ✓ Microbiological indices (coliforms, streptococci)
- ✓ Biological Indicators (phytoplankton, zooplankton, zoobenthos, fish, macrophytes).

3.1 Monitoring by Systematic Sampling

Existing technology is based on the measurement of water quality control parameters by field measurements, either by collecting samples and measuring the parameters in the laboratory. Sampling in lakes to determine water quality is a complex process. The methodologies used in the sampling and interpretation of the data are intended to determine the concentration of abiotic resources and the abundance of living resources, as well as other important features, such as morphometric characteristics in fish (Psilovikos, 2020).

Sampling is necessary to provide information from a characteristic part of the body of water. The most critical factors that need to be identified are:

- Sampling points,
- The frequency of sampling and

• The way to preserve the integrity of the samples prior to analysis procedure

Monthly sampling might skip a whole phytoplankton bloom. Also, sampling demands might require different frequency, if the ecology of the lake is known. Sampling in lakes must be carried out at least once a month, or more often on a daily or hourly basis. In lake monitoring, the key factors that determine the exact number of monitoring sites and the frequency of sampling are:

• Ecological importance

• Economic importance

• Morphological characteristics

• Hydrological characteristics

• Residence time

• Retention time

• Pressures from Agricultural, Industrial, Urban and Natural Sources of Pollution

• Time and Spatial Variations of Quality Parameters,

• Water disposal from Biological Cleaning

• Technical and Financial Capabilities

According to the WFD 2000/60 (2000), there are three kinds of Monitoring (Directive, W.F., 2003):

✓ The **Surveillance** Monitoring:

Parameters indicative of all the biological, hydromorphological and all general and specific physico-chemical quality elements are required to be monitored. The objectives of surveillance monitoring of surface waters are to provide information for: a) The efficient and effective design of future monitoring programs; b) The assessment of long term changes in natural conditions; and c) The assessment of long term changes resulting from widespread anthropogenic activity.

✓ The **Operational** Monitoring:

The objectives of operational monitoring are to: a) Establish the status of those bodies identified as being at risk of failing to meet their environmental objectives; and b) Assess any changes in the status of such bodies resulting from the programs of measures. Operational monitoring (or in some cases investigative monitoring) will be used to establish or confirm the status of bodies thought to be at risk. Therefore, it is operational monitoring that will produce the environmental quality ratios used for status classification for those water bodies included in operational monitoring.

✓ The **Investigative** Monitoring:

It may also be required in cases: a) Where the reason for any exceedances' (of Environmental Objectives) is unknown; b) Where surveillance monitoring indicates that the objectives set for a body of water are not likely to be achieved and operational monitoring has not already been established, in order to ascertain the causes of a water body; or water bodies failing to achieve the environmental objectives; or c) To ascertain the magnitude and impacts of accidental pollution. Investigative monitoring might also include alarm or early warning monitoring, for example, for the protection of drinking water intakes against accidental pollution.

Furthermore, many studies (Papastergiadou et al., 2007; Haidary et al., 2013; Karamoutsou et al., 2016; Yu et al., 2016; Shi et al., 2017;) have noticed the impact of catchment changes in land uses to water quality. In order to address this fact, the Water Framework Directive (WFD) assists Management Plans at river basin level. The Greek National Water Monitoring Network, as defined in JMD 140384/2011, involves fifty lakes. In total, fifty-three stations have been established, as it is shown in Figure (3.1) (Katsiapi et al., 2012; Mavromati et al., 2018).

3.2 Automated Telemetric Monitoring

In order to record peak phenomena, monitoring should be conducted telemetrically on a continuous basis, otherwise monthly data is not sufficient. In recent years, the importance of on - line measurement of various real - time physicochemical parameters in lakes, rivers, groundwater, coastal and marine waters, as well as in wastewater treatment plants has increased significantly.

For this reason, techniques have been developed to continuously monitor the parameters of water quality and waste, which differs from conventional electrochemical methods. New types of sensors have been developed, which essentially act as spectrophotometers, which a) have direct contact with the sample, b) have short response times, c) perform a large number of measurements, d) do not use chemical reagents and e) do not require sample preparation. In addition to the basic parameters (Dissolved Oxygen, Temperature, pH, Conductivity, Salinity,) they are also applied to Organic Load (BOD, COD or TOC) measurements of nutrient and toxicity (Psilovikos, 2020).

In case of Lake Kastoria, four Telemetric Stations, assist the main monitoring station, according to Directive 2000/60, Law 3199/2003 and JM 140384/2011, both spatially (more stations in different locations) and temporally (continuous monitoring). These results are collected by the Ministry of Environment & Energy of Greece.

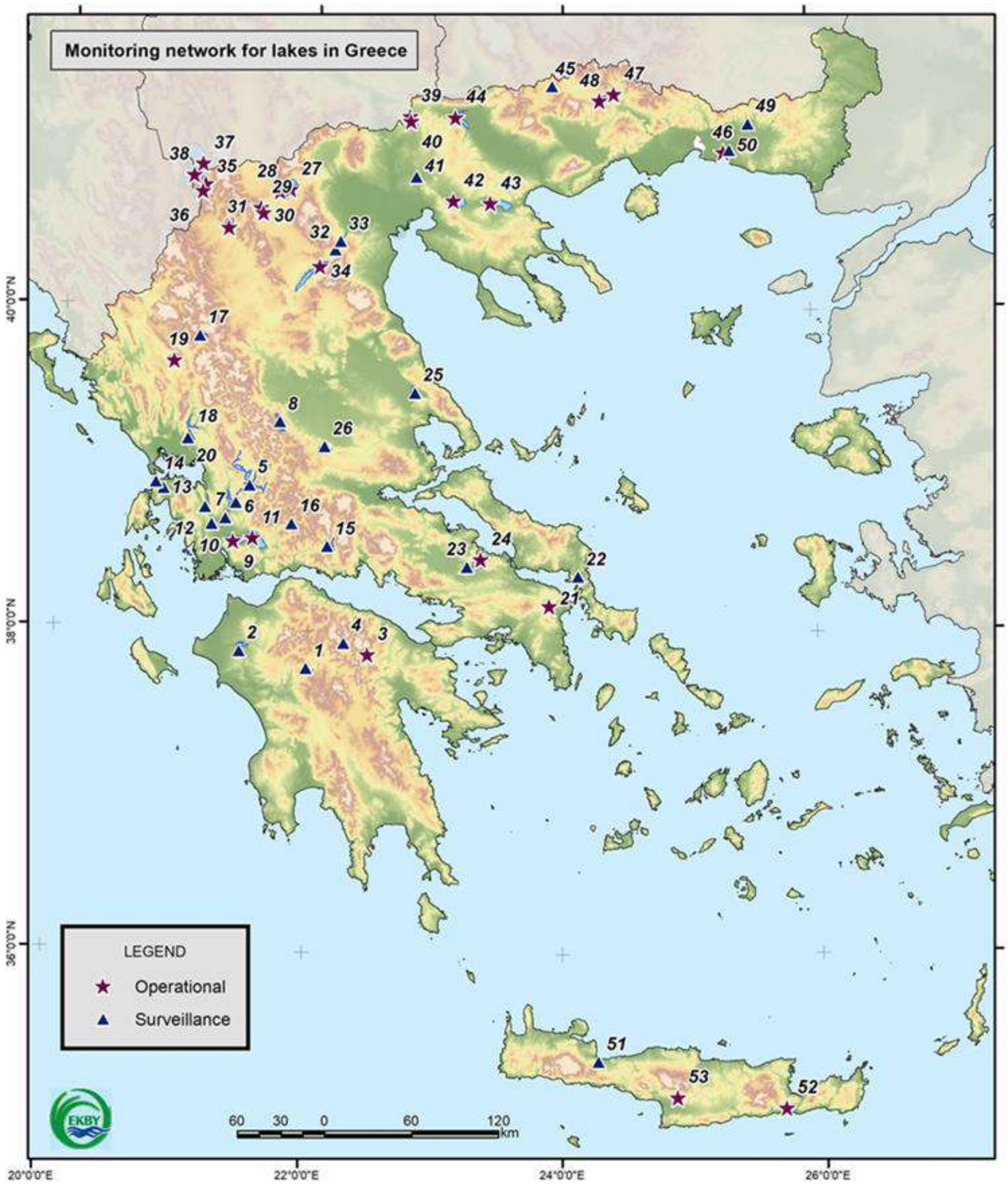

Figure 3.1: Monitoring network for lakes under operational monitoring (stars) or surveillance monitoring (triangles) in Greece Reservoirs:(1)T.L.Ladona,(2)T.L.Pineiou,(4)T.L.Feneou,(5) T.L. Kremaston, (6)T.L. Kastrakiou, (7)T.L.Stratou,(8)T.L. Tavropou, (15)T.L. Mornou, (16) T.L.Evinou, (17) T.L.PigonAoou, (18)T.L. Pournariou, (20)T.L. Pournariou II, (21) T.L. Marathona, (25)T.L. Karlas, (26)T.L. Smokovou, (32)T.L. Sfikias, (33)T.L. Asomaton, (34)T.L. Polyfytou, (44)T.L. Kerkini, (45)T.L. Leukogeion, (47)T.L. Platanovrysis, (48)T.L. Thisavrou, (49)T.L. Gratinis, (50)T.L. N. Adrianis, (52)T.L. Bramianou, (53) T.L. Faneromenis Natural Lakes: (3)Stymfalia, (9)Lysimacheia, (10)Ozeros, (11) Trichonida, (12)Amvrakia, (13)Voulkaria, (14)Saltini, (19)Pamvotida, (22)Dystos, (23)Yliki, (24)Paralimni, (27)Vegoritida, (28)Petron, (29)Zazari, (30) Cheimaditida, (31)Kastorias, (35) MikriPrespa A, (36)Mikri Prespa B, (37)Megali Prespa A, (38)Megali Prespa B, (39)Doirani 1, (40)Doirani 2, (41)Pikrolimni, (42)Koroneia, (43)Volvi, (46)Ismarida, (51)Kourna (Mavromati et al., 2018)

By considering how much Machine Learning has expanded nowadays, it is clear that monitoring systems, by providing systematic information, play a crucial role in lakes management.

Monitoring in real time of the qualitative parameters contributes in Management Practices, the most important of which are summarized below: (Heinonnen et al., 2004; Psilovikos, 2005; Psilovikos, 2020):

- ✓ Maintaining the ecological balance of the ecosystems and water resources of the region,
- ✓ Controlling water quality for irrigation
- ✓ Monitoring atmospheric conditions
- ✓ Determining microclimate indicators
- ✓ Warning in cases of crises caused by extreme events
- ✓ Continuous knowledge of the state of the water bodies

3.3 Satellite Monitoring Using Remote Sensing

Monitoring of water quality is based on the concentrations of various water quality parameters at certain stations, which are determined by sampling and analysis in the laboratory or by direct telemetry. For efficient management of surface water, additional information is needed on the spatial dispersion and time variations of these concentrations. Sensors location, which enables remote measurement of a range of parameters, is in many cases the only technique for observing and monitoring spatial variations in water quality. Monitoring with telemetric systems, enables the measurement of a number of physicochemical parameters in the surface, but it is not able to measure dissolved oxygen in the hypolimnion, toxic substances and acidity. However, the parameters that could be identified such as Chlorophyll-aand Turbidity can play an important role in monitoring the ecological status of a lake.

Water quality monitoring with remote sensing systems is based on the use of immersed sensors operating in the visible electromagnetic spectrum at a wavelength of 380-760 nm. Additionally, the temperature can be measured with an infrared sensor (3-30 μm). Remote Sensing is a very useful tool, to monitor hydrodynamic phenomena important to aquatic ecosystems and to create thematic maps. For example, the use of Tw, Chl-a, Nutrients, DO monitoring data and the maps obtained, can interpret water interactions; for example, water that outflows from karst springs and thermal stratification in lakes, eutrophication indexes etc. A sensor consists of a spectrometer that measures the radiation of water at various wavelength ratios. To allow measurements to be made under various brightness conditions, the measurement is often transformed into reflection (Psilovikos, 2020). Remote Sensing is often used in shallow and muddy wetlands, that it's not easy to install Telemetric Stations, for example the Mesologgi Lagoon (Albanakis et al., 2004).

3.4 Monitoring in Lake Kastoria: The Study Area

3.4.1 General characteristics of the Study Area

The Study Area comprises the catchment area of Lake Kastoria (Figure 3.2) which is located in the northeast part of Kastoria Prefecture. The area of Lake Kastoria is approximately 281 km^2, with an average water depth of 4,40 m and a maximum water depth of 9,10 m, while the total volume of water is 110 x 10^6 m^3. The minimum altitude of the catchment is 629,80 m, which is the water level of the lake and the maximum altitude is 2.128,00 m, which is the top of Mt. Vernon. Lake Kastoria is characterized as a Natura 2000 site (Council Directive 92/43 / EEC of 21 May 1992). Regarding the flora and fauna of the wider area of Lake Kastoria, it has a rich birdlife, as more than 200 species of birds are found in the area. In addition, it is particularly rich in fish and is considered the second richest lake in Greece's fish catches.

The catchment area of Lake Kastoria is delimited:

- ✓ North from the Mountain Verno, with the highest top "Vitsi", which is also the highest point of the lake catchment and from "Spyridaki", "Sikavitsa " and "AgiaParaskevi".
- ✓ East from "Falakron", "Kronos", "Doukas", "Mavrovouni", "Stena Klisouras" and "Pyrgos".
- ✓ South, from "Petrodes", "MikroVouno" and "Korisos".
- ✓ West, from "Korifi", "AgiaTrias", "Kazani" and "Perceli"
- ✓ The basin of Lake Kastoria is subdivided into 11 sub-basins of which:
- ✓ Nine (9) sub-basins refer to the main watercourses, which are the streams "Fountouklis", "Aposkepou", "Vyssinias", "Agios Athanasios", "Toichiou", "Metamorphosis", "Fotini", "Xiropotamos", "Istakos", "Kastoria-Dispilio"and "Ampelokipoi".
- ✓ One concerns the sub-basin of the city of Kastoria, whose runoff enters the lake through drainage pipes
- ✓ One concerns the sub-basin of the Ambelokipi area, southwest of the Kastoria basin, whose surface water outflows from Lake Kastoria to Gkiole stream which consists part of the hydrographic network of Aliakmon River. Water from Lake Kastoria outflows artificially from a weir in Gkiole stream and finally outflows to Aliakmon River. It is noted that the Gkiole's channel constitutes the only surface discharge of Lake Kastoria's basin, as well as the connection between Lake Kastoria and River Aliakmon. The presence of the weir in Gkiole's channel area enables the

optimization of the lake's hydrological functions, ensuring that the level of the lake will remain under the higher level of 629,80 m, otherwise Kastoria town will be flooded. This is why the geomorphological catchment of Lake Kastoria is different than the hydrological one. The latter does not contain the sub-catchment of Gkiole because it is discharging as an outflow of Lake Kastoria to River Aliakmon (Figure 3.3).

Figure 3.2: The geographical location of Lake Kastoria

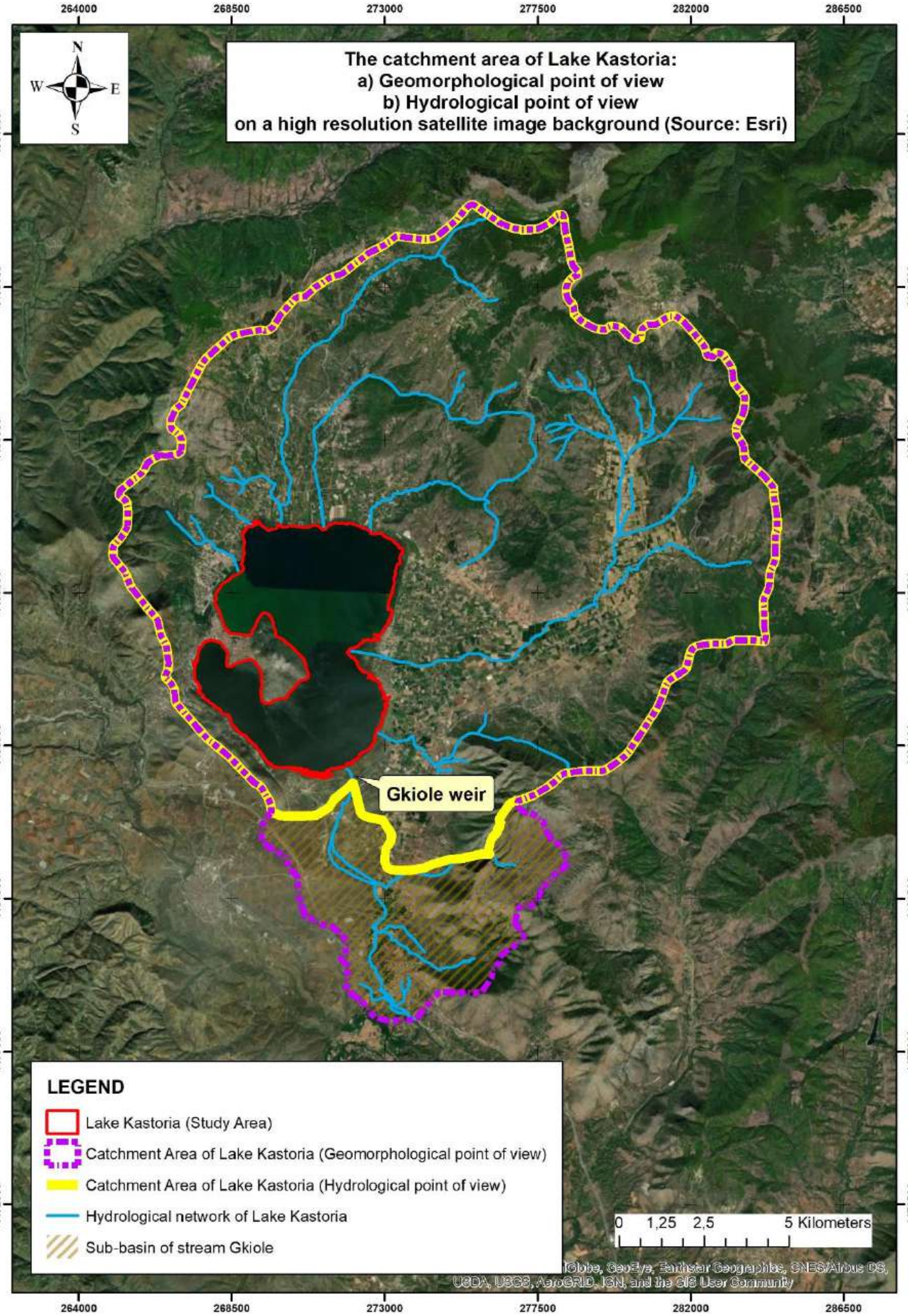

Figure 3.3: The catchment area of Lake Kastoria: a) Geomorphological and b) Hydrological point of view on a high resolution satellite image background(Psilovikos, 2020)

3.4.1.1 The Geological - Geomorphologic Characteristics

The region of Kastoria has a great variety of geological formations. According to the Geological Map of the Institute of Geological and Mining Research (IGME, 1983) (Figure 3.4), the wider area of Lake Kastoria includes the following:

- ✓ Post-tectonic and late-tectonic sediments
- ✓ Prealpine series
- ✓ Pelagonian zone
- ✓ Metamorphic rocks
- ✓ Igneous rocks-intrusive

Regarding the geomorphologic features of Lake Kastoria’s Basin, the morphology of the wider area is the result of its geological structure, its tectonic evolution and erosion. The products of the erosion, such as suspended materials and sediments, are transported in lower regions.

The following Figures (Figure 3.5, Figure. 3.6, Figure 3.7 and Figure 3.8) show the:

a) Triangulated Irregular Network (TIN), which shows a pseudo-three dimensional mapping of the catchment area (Figure 3.5)

b) Topographic Map, with the contour lines of the altitude (Figure 3.6). The contour lines have an interval of 50 m.

c) Aspect Map, with the eight directions (N, NE, E, SE, S, SW, W, NW) (Figure 3.7) and

d) Slope Map, which is divided in five classes (Figure 3.8)

All the digital maps have been designed with the help of Geographic Information Systems (GIS).The map backgrounds are based on:

a) the orthophotograph of Ktimatologio S.A. of the historical year (1945),

b) the various thematic maps of the Ministry of Agriculture, such as the Geological map of I.G.M.E. of the year 1983, the Bioclimatic map of I.F.R. Athens of the year 1978 and the Vegetation map of I.F.R. Athens of the year 1978,

c) the World Imagery base-map metadata by Esri, which presents low-resolution satellite imagery for the world and high-resolution satellite and aerial imagery, typically within 3-5 years of currency, for most of the world and

d) the «http://www.geodata.gov.gr», the open national list of geospatial data and services, which enables the construction ofupdated Natura 2000 protected areas network map and the Land Uses map (Corine Land Cover – year2000).

e) the aspect map, the topographic map with 50 m intervals, the elevation map, the slope map, are created by USGS DEM standard, which is a geospatial file format developed by the United States Geological Survey for storing a raster-based digital elevation model.

Many of the backgrounds of the maps that are created are digital images, whichare scanned by an analog map or an aerial photo. In order to digitally process the above scanned images, with a resolution of 1200 dpi, real geographic coordinates of scanned images based on the desired Geodetic Reference System, are given (Georeferenced). In our case, the Greek Geodetic Reference System 1987 (GRS ’87), is selected, which has been the Geodetic Reference System used in Greece, since the 1990s. The implementation of the Georeference, which is carried out using ArcGIS software, requires the identification of Photostables and Control Points. The selected number determines, the accuracy of the Georeference process. The detection of Photostables and Control Points is a crucial point, due to the existing “noise”and the deformation the scanned images. The selectedPhotostables and Control Points, are objects that show a strong difference with the surrounding area, (Characteristic – Crucial Points) in which they are located, in terms of their altitude and construction material (Geodetic points in the , and emerged from careful examination and identification of distinct elements, such as the data of a) vegetation (individually or groups of trees and shrubs), b) geomorphological characteristics (such as peaks or steep outbreaks of the parent material, stream bed), and c) technical constructions (such as farmhouses, crossroads). It is pointed out that in the case of Lake Kastoria, the detection of Photostables and Control Points near the riparian zone of the lake isavoided, due to Spatial and Temporal fluctuations of these zones, mainly due to the water level and flooding phenomena. Finally, the Control Points of the scanned image are then used as a reference.

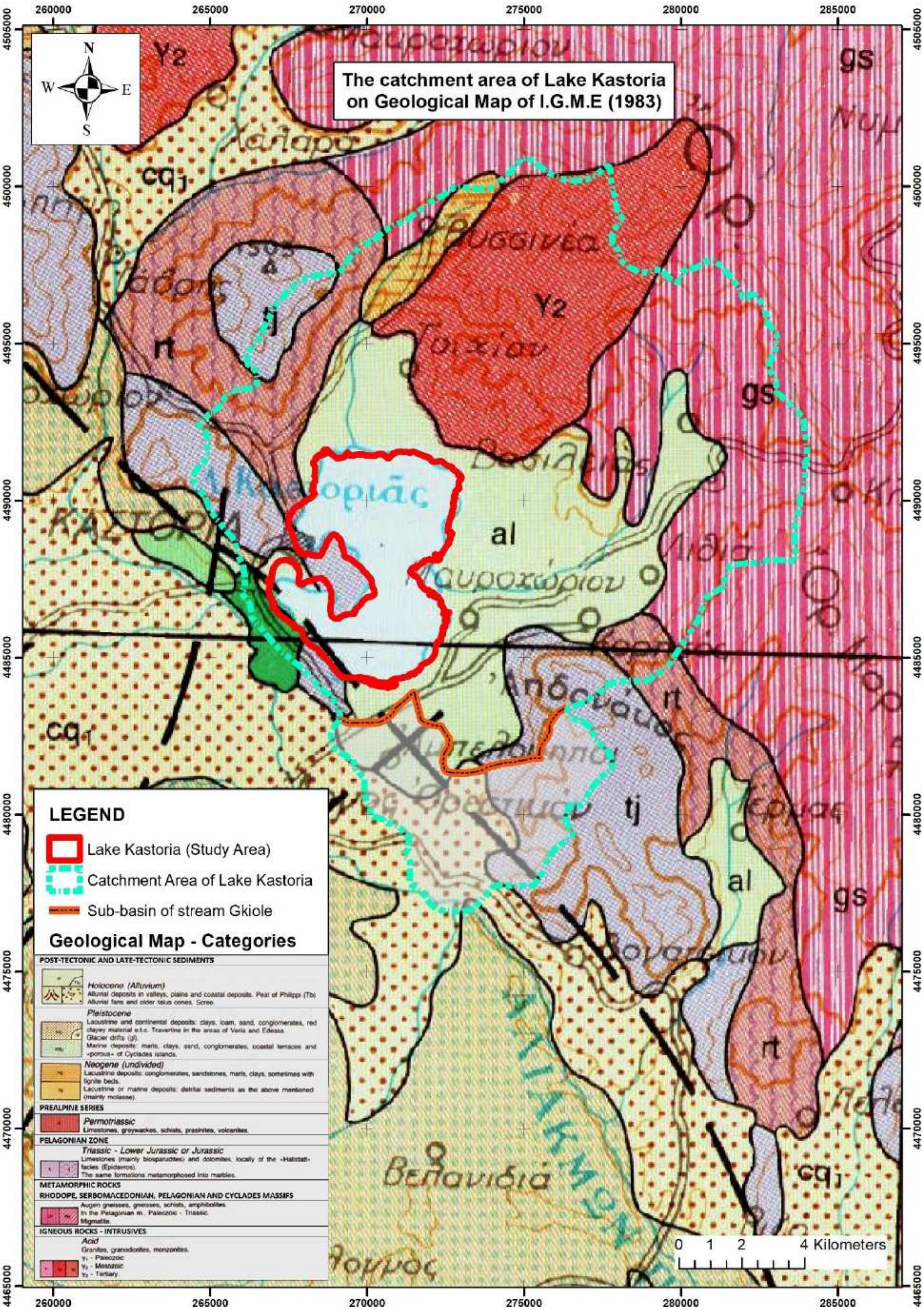

Figure 3.4: Geological map of the catchment area of Lake Kastoria (IGME, 1983)

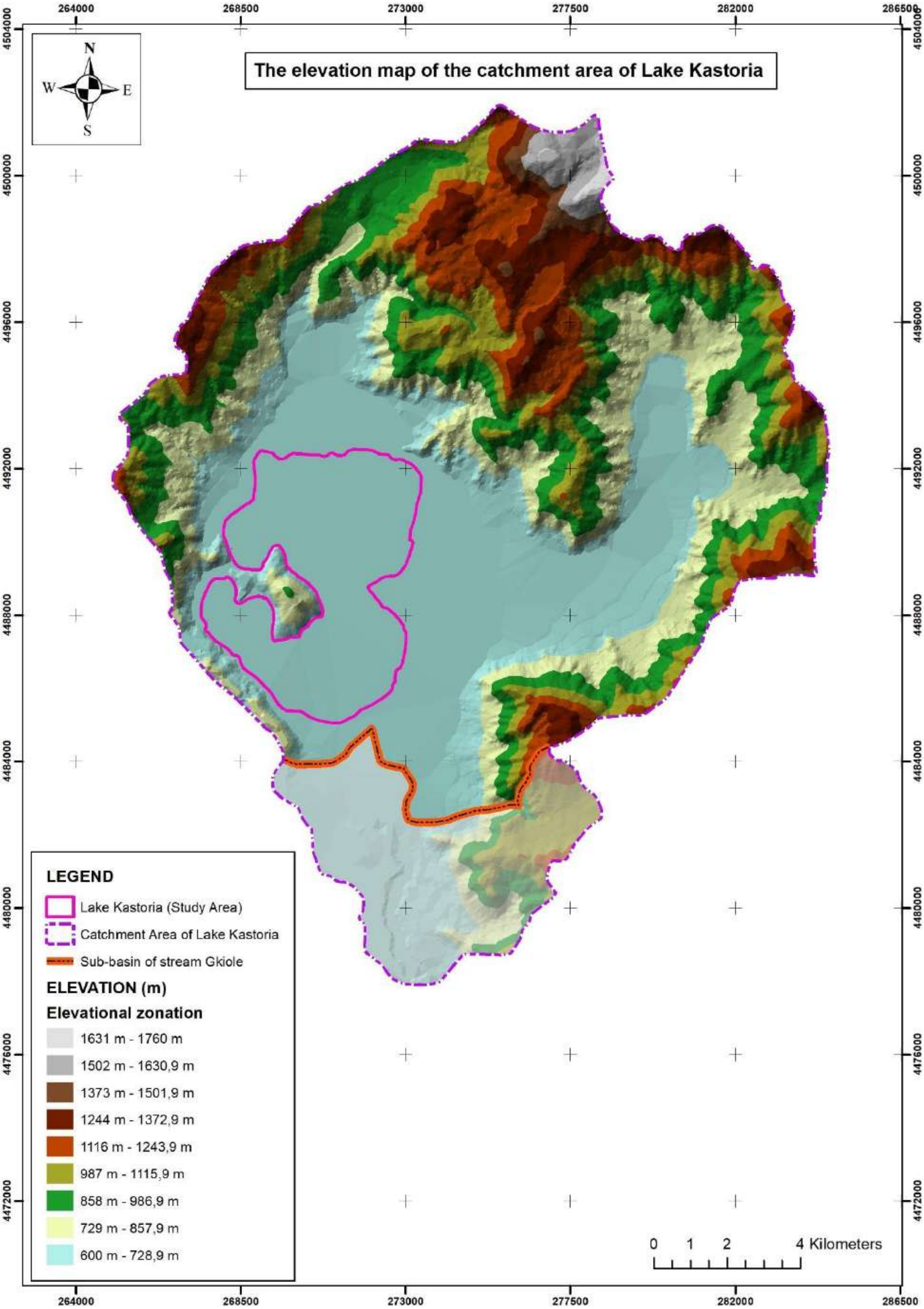

Figure 3.5: The elevation map of the catchment area of Lake Kastoria

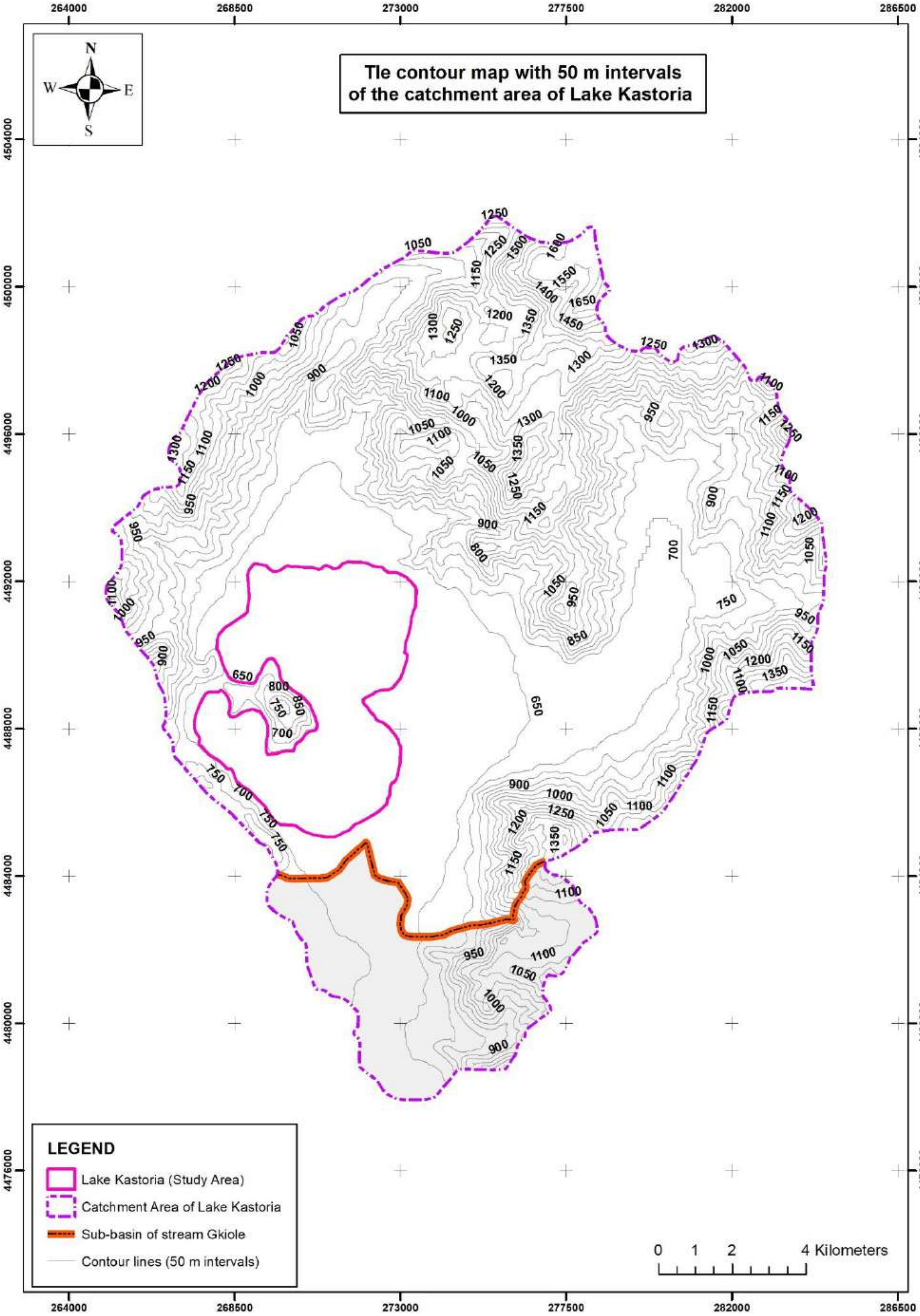

Figure 3.6: The topographic map of the catchment area of Lake Kastoria

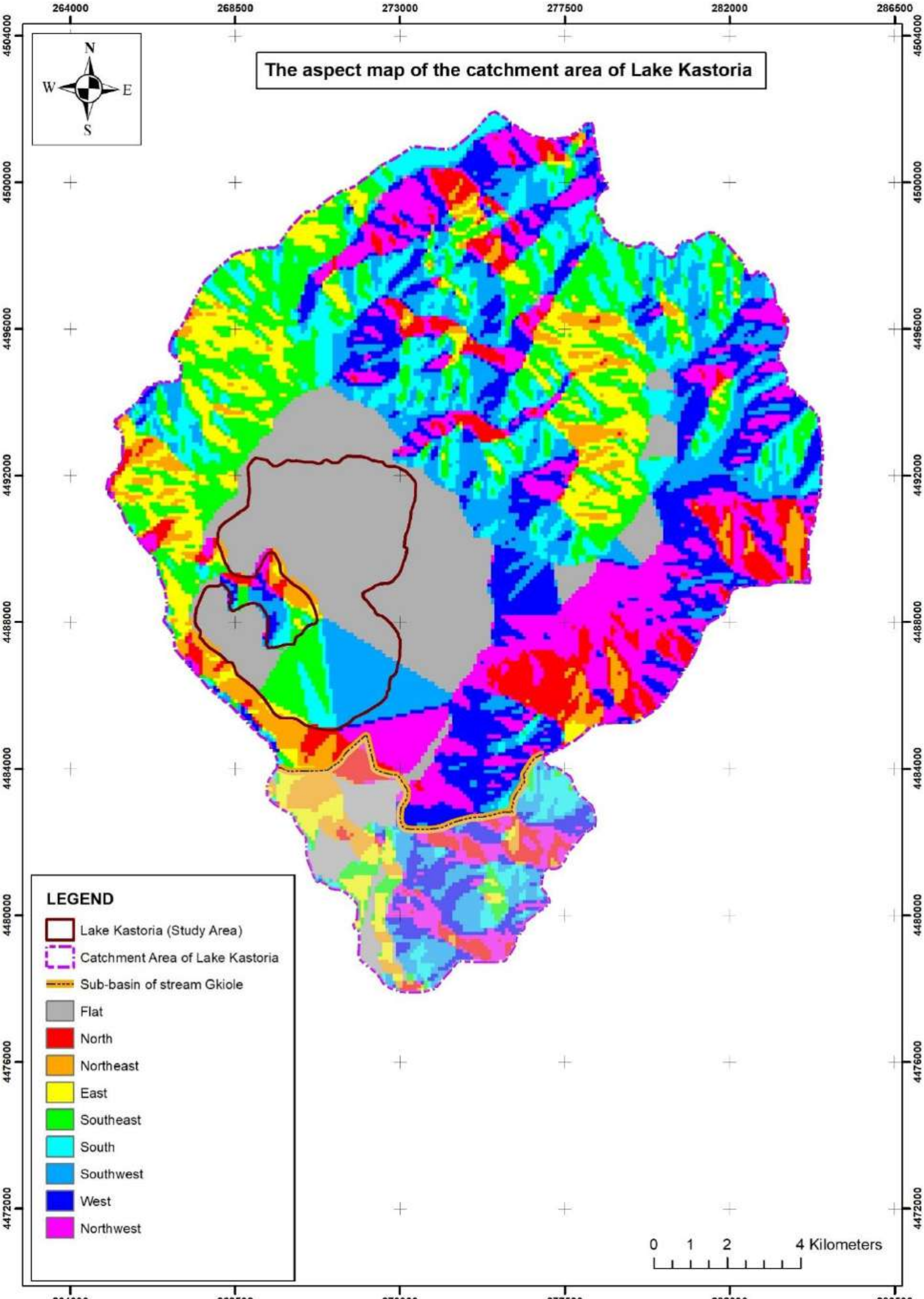

Figure 3.7: The aspect map of the catchment area of Lake Kastoria

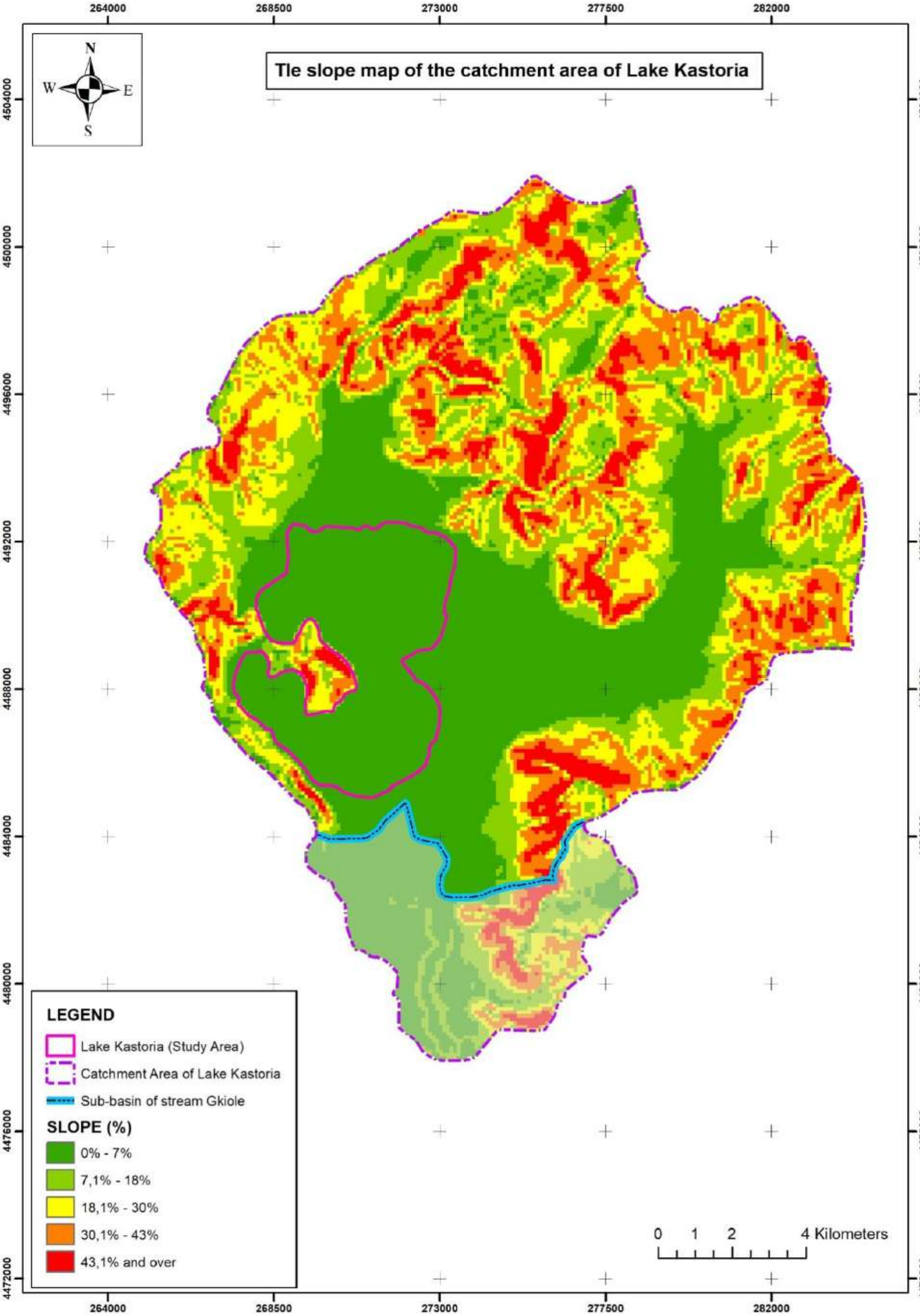

Figure 3.8: The slope map of the catchment area of the Lake Kastoria

3.4.1.2 Bioclimatic characteristics

The Bioclimatic map is based on the information of the geographical location of a region, its terrain morphology and its meteorological data.It is the key tool for the interpretation, analysis and study of the climatic and bioclimatic conditions of an area.Although the microclimate in an area varies and depends on more "sensitive" factors, such as air transport, the resulting bioclimatic map clearly shows whether climate conditions create very humid or semi-humid climates.Based on the studyof the Bioclimatic map of the Ministry of Agriculture in 1978 for LakeKastoria catchment area (Figure 3.9), a section of the mountainous catchment area shows humid climatic conditions, while the rest of the catchment shows semi-humid climatic conditions.

3.4.1.3 Vegetation characteristics

It is a fact that the natural environment in Greece impresses, not only with the naturalness that characterizes it, but also with the diversity it encompasses. The wider area of Lake Kastoria is undoubtedly of great ecological and aesthetic value, and it is no coincidence that it is characterized as a place of exceptional natural beauty. The ecosystems that develop in the wider area of its watershed are mountainous.

Based on the study of the Vegetation map of the Ministry of Agriculture (1978) for LakeKastoria’s catchment area (Figure 3.10), three (3) vegetation categories are presented:

- ✓ Thermophilous sub-continental deciduous of Quercus spp.
- ✓ Mediterranean formations of Fagus sp., Abiesborisii-regis and
- ✓ Mediterranean formations of Pinus Sylvestris, Picea Abies

The above-mentioned bioclimatic conditions are reflected in the present vegetation situation, taking into account the anthropogenic pressures that have operated from the past to the present.

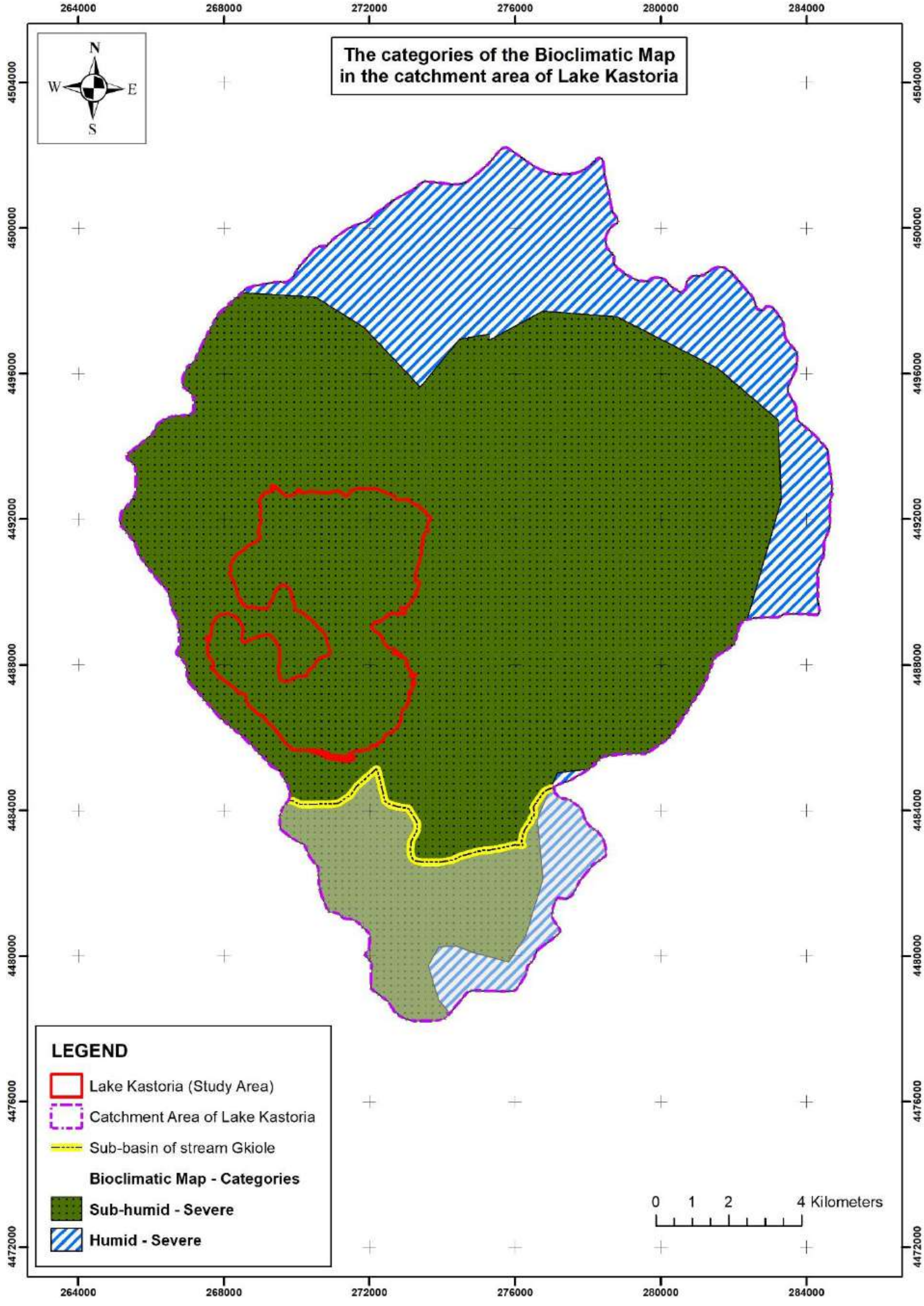

Figure 3.9: Bioclimatic map of the catchment area of Lake Kastoria

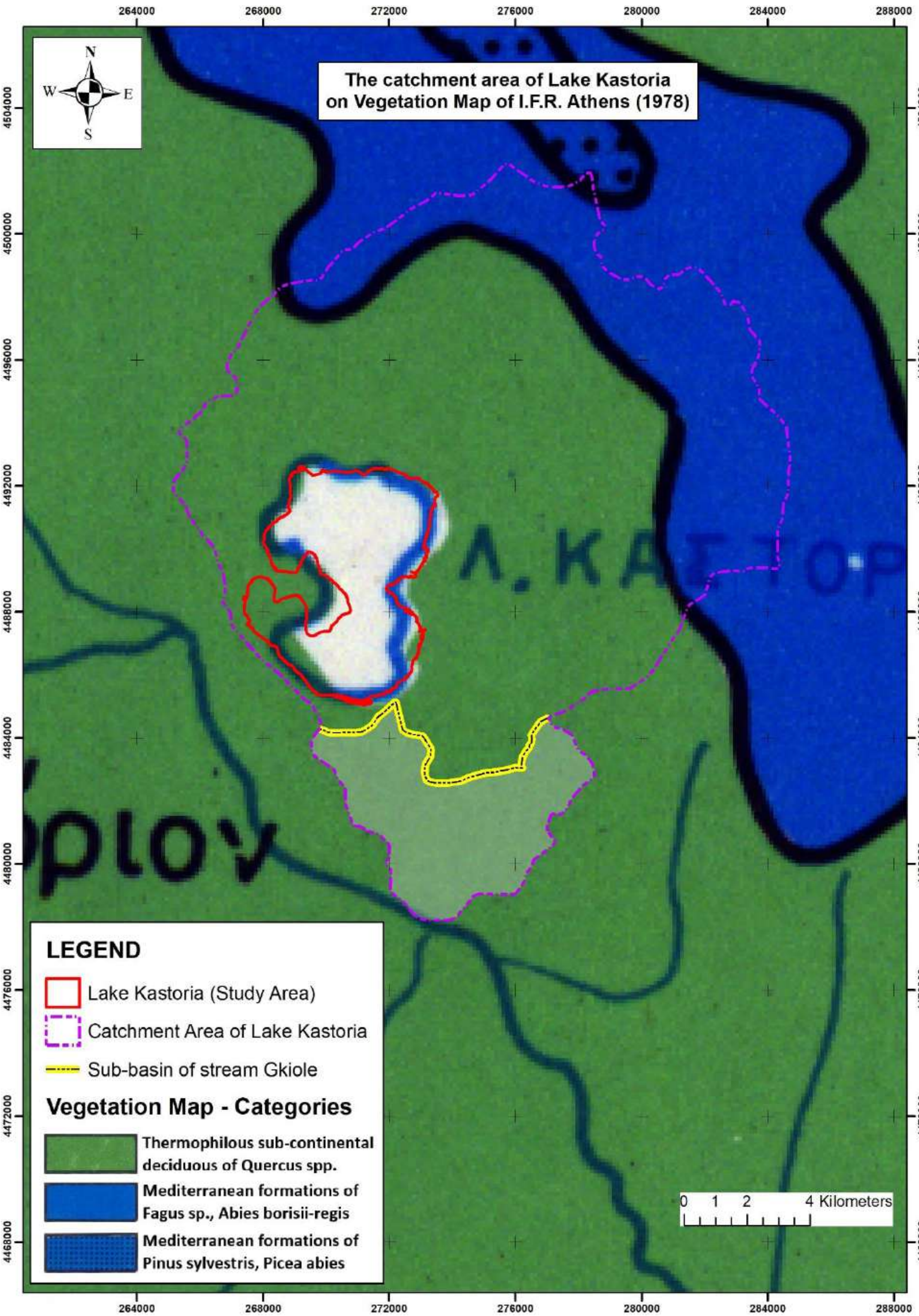

Figure 3.10: Vegetation map of the catchment area of Lake Kastoria (Source: Athens Forestry Research Foundation, Ministry of Agriculture – 1978)

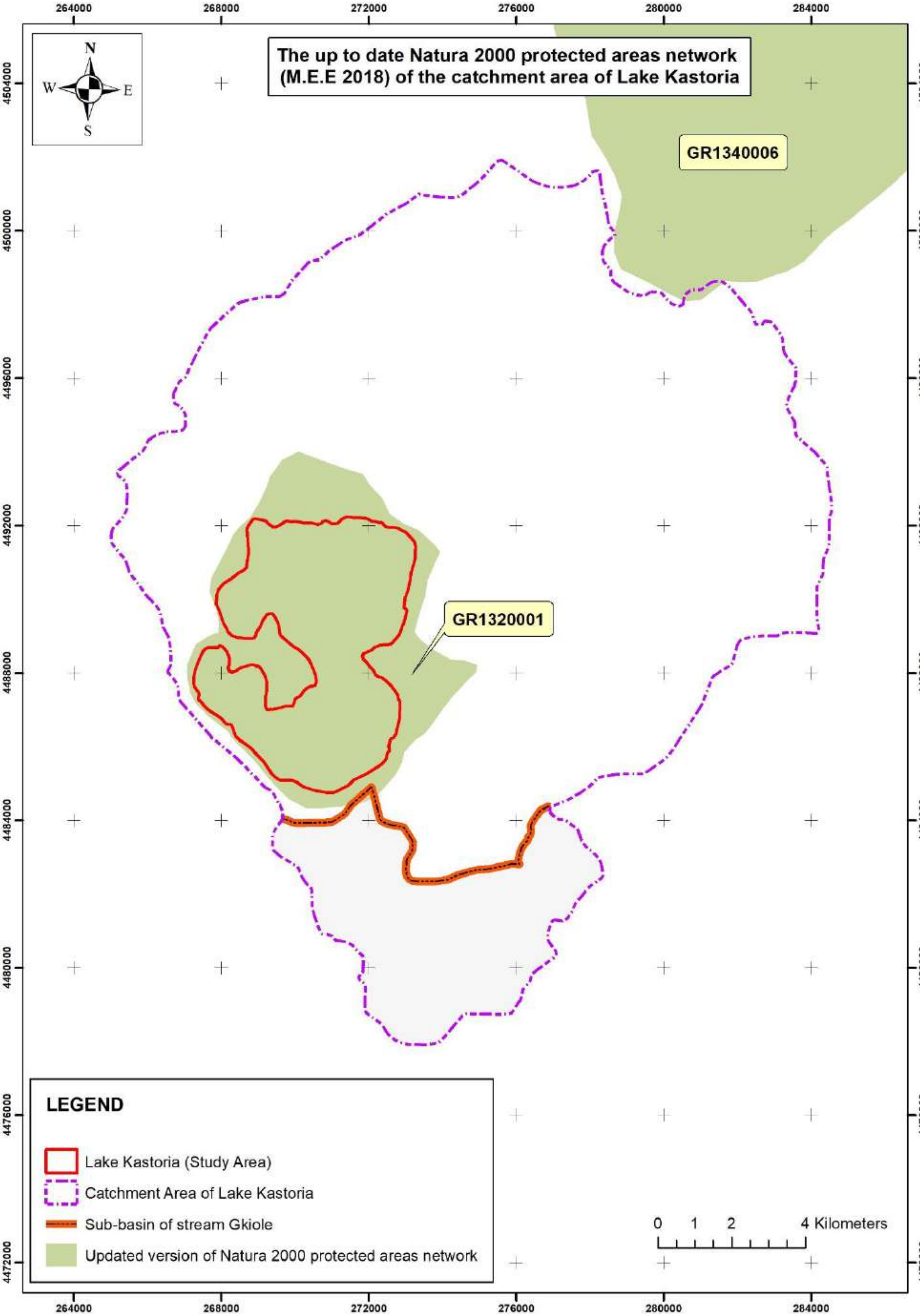

Figure3.11: The updated Natura 2000 of the catchment area of Lake Kastoria

3.4.1.4 Areas of the Natura 2000 network

The ecological importance of the lake's basin is enormous, due to itsbiodiversity. The lake also functions as a resting and feeding station for migratory birds. Moreover, it is a shield to protect soil from climate change and desertification. Lake Kastoria is characterized as a Natura 2000 site (Figure 3.11) (Council Directive 92/43 / EEC of 21 May 1992) on the conservation of natural habitats and of wild fauna and flora. The Natura 2000 network also includes Special Protection Areas designated under Directive 79/409 on the Conservation of Wild Birds.

3.4.2 Monitoring in Lake Kastoria

Despite the fact that since 1995 there is a biological treatmentplant for domestic waste, considerable amounts of point and non-point pollutants are deposited in the lake. These pollutants include fertilizers, pesticides, organic substances, nutrients and toxic substances. The increased nutrients and toxic substances that occur in the lake led to eutrophication, less water transparency, less dissolved oxygen and adverse conditions for aquatic organisms, especially fishes.

According to Naselli-Flores(2008), urban lakes, such as lake Kastoria, tend to receive high nutrient loads and show higher trophic status than non-urban ones. There is no doubt that highly eutrophic lakes are primarily affected by waste water inputs and intensive agriculture activities (Skoulikidis et al., 1998).

Since October 2002, a significant effort has been made by the Municipality of Kastoria with systematic measurements - sampling - at certain points of the lake, in order to record the following water quality parameters:

1. Temperature (T, ^{o}C),
2. Dissolved oxygen (DO, mg/L),
3. Ammonia nitrogen ($N-NH_4$, mg/L);
4. Nitrate nitrogen ($N - NO_3$, mg/L); and
5. Phosphate phosphorus ($P-PO_4$, mg/L).

Additionally, the conductivity (ECw - μS/cm), BOD (mg/L), COD (mg/L), nitrite nitrogen ($N - NO_2$, mg/L), total phosphorus (P, mg/L), pH, transparency (SecchiDisk Transparency -m) and Redox were also measured (Matzafleri et al., 2007).

These measurements took place once a month in order to:

✓ Create reliable monthly data time series,

✓ Control specific water parameters and

✓ Assess the quality status of Lake Kastoria.

The aforementioned attempts (on a monthly basis) are not organized at the same date and at the same time per month, so the comparison of the qualitative parameters of the water is not possible.

It should be mentioned here that the time step of each measurement is very important. For example, in a shallow, eutrophic lake as Lake Kastoria, the variance of dissolved oxygen is very wide during the day. Thus, large concentrations of dissolved oxygen are observed in the morning and afternoon hours due to photosynthetic algae productivity (cyanide, chlorophyll) and very low values are observed overnight because of its consumption by the microorganisms for the degradation of non-preservatives (Matzafleri et al., 2009).

Moreover, even if the measurements are taken at a specific time (once a month), they are inadequate, because "peak" data are "lost" due to any kind of pollution that might occur, just a few times after the pollution incidence. Till the next one measurement, the influence of the peak pollution incidence could not be recorded anymore. Therefore, a continuous and an organized monitoring throughout 24-hour period, is required. This is only possible with the automated monitoring - "Monitoring".

The advantages of such a system include:

- Installation in any position within the lake and possibility of moving to another.
- Transmitting of measurements anywhere via the Internet at predetermined intervals.
- Providing on-line information on the environmental status of the lake
- Ability to implement spatial simulation models - dynamic 2D maps - through a GIS system.
- Ability to act as a Decision Support System (DSS) taking immediate decisions and measures if environmental risks are threatened for exampleincidence of human pollution from domestic and agriculture activities.
- Ability to apply Artificial Neural Network, in order to build the appropriate models, make predictions and avoid extreme episodes
- Sustainable management of the aquatic ecosystem (Lake Kastoria – Natura Area) through the implementation of management plans which is the final goal.

As we mentioned, last monitoring (2002 - 2013) took place on a monthly basis, with 5 Stations and two Depths (surface and bed) per station. Since 2015, Lake Kastoriaismonitored telemetrically on a continuous daily basis, by 4 Stations at one location on its surface. This telemetric network was firstly proposed and documented by Psilovikos (2012).The telemetric stations record on a daily basis (Psilovikos, 2012; Karamoutsou and Psilovikos, 2019):

1. Chlorophyll a (Chl-a - μg/L),
2. pH,
3. Water temperature (Tw - ° C),
4. Conductivity (Cond - μS/cm),
5. Turbidity (Turb - NTU),
6. Dissolved Oxygen (DO - mg/L),
7. Ammonia nitrogen ($N\text{-}NH_4$, mg/L);
8. Nitrate nitrogen ($N - NO_3$, mg/L); and
9. Water level (m)

The determination of optimum locations of the telemetric stations in Lake Kastoria Lake, was conducted by Western Macedonia Development “AN.KO” and the University of Thessaly, under the supervision of Professor ArisPsilovikos (2012).The positions of current telemetric stations are shown in Figure (3.12).It is worth to notice that the delineation of the coastline and its performance was based on the use of high resolution satellite images (ESRI). Moreover, it is noted that its performance indicates the separation of the water body of Lake Kastoria from the terrestrial environment at the given time. Its orientation varies spatially and seasonally on the basis of hydrological characteristics and aquatic vegetation.

The criteria of the choice of the monitoring stations, in the present positions were: a) the environmental pressures occurred from the land uses of the catchment area b) the hydromorphological characteristics of the lake, c) the presence of corrosion, transport and deposition phenomena d) the inflows and outflows and f) issues related to the accessibility and the costs of stations’ installation and maintenance.

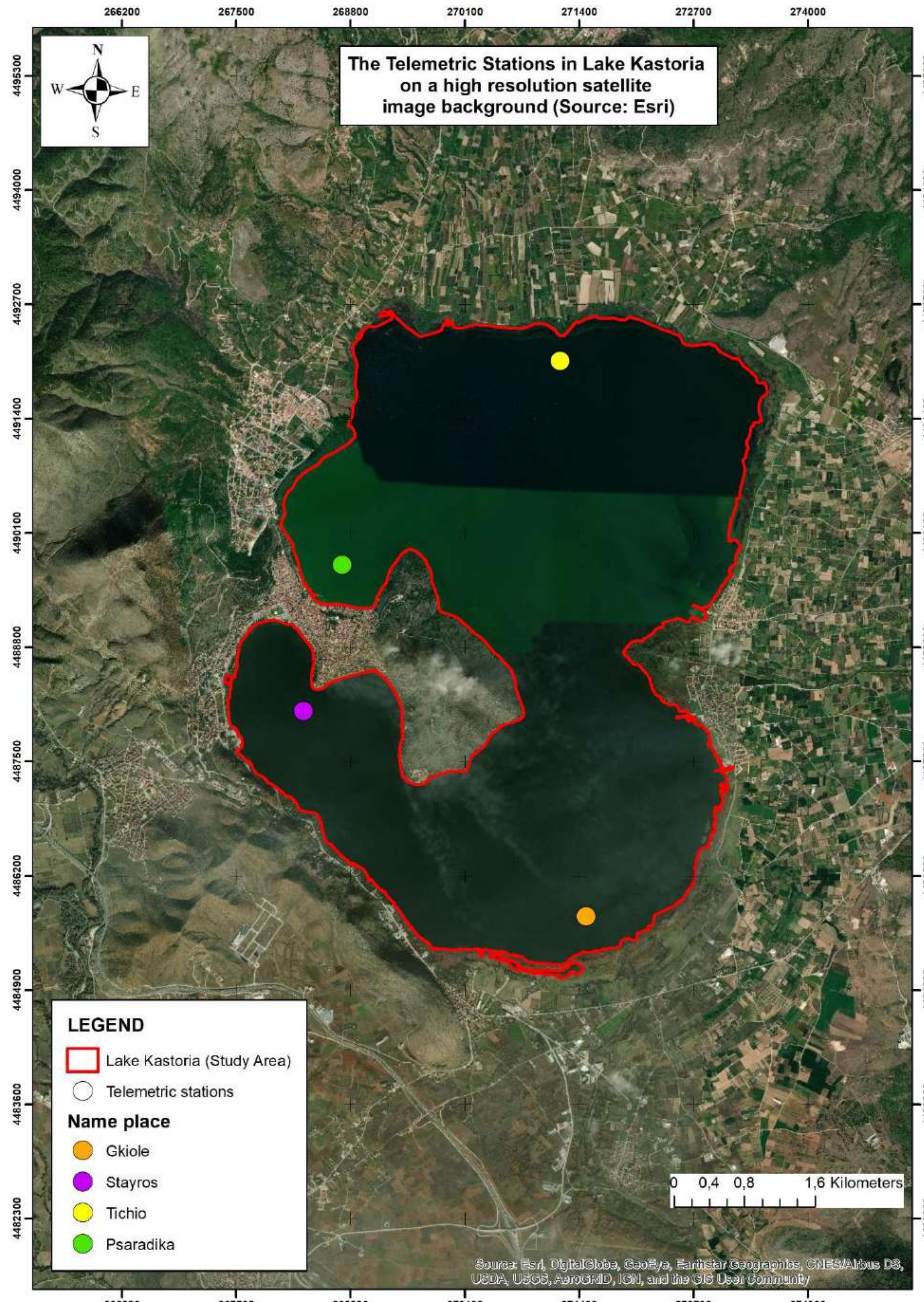

Figure 3.12: The position of telemetric stations in Lake Kastoria (Karamoutsou and Psilovikos, 2019)

References

Albanakis K., Vouvalidis K., Spanou P., Kommata S., Psilovikos A., 2004. Morphological mapping of the bottom of Mesologgi Lagoon with the use of astronaut photography and GIS. Bulletin of the Geological Society of Greece, Vol 36, Issue 2, pp. 950 – 957.

Antonopoulos, V., 2010. Environmental Hydraulics and Surface Water Quality, Tziolas Pub, Thessaloniki, Greece, 473 pages (in Greek).

Council Directive 79/409/EEC. 1979. The conservation of w*ild birds.* OJ L 103, 25.4.1979, pp. 1 - 18.

Council Directive 92/43/EEC. 1992. The conservation of natural habitats and of wild fauna and flora. OJ L 206, 22.7.1992, pp. 7 - 50.

Directive, W.F., 2003. Common implementation strategy for the water framework directive (2000/60/EC). *Guidance document*, (7).

Haidary, A., Amiri, B.J., Adamowski, J., Fohrer, N. and Nakane, K., 2013. Assessing the impacts of four land use types on the water quality of wetlands in Japan.Water Resources Management,27(7), pp.2217-2229.

Heinonen, P., Pietiläinen, O.P., Airaksinen, O., Haapala A., Hammar T., Holopainen, A.L., Hämäläinen H., Issakainen J., Kanninen, A., Karttunen, K. and Leka J., 2004. Monitoring and assessment of the ecological status of lakes. *A pilot procedure developed and tested in the Life Vuoksi project. The Finnish Environment*,719, p.106.

Karamoutsou L., Psilovikos A., Stålnacke P., Farkas C., 2016. Modifications in the Vegoritida's Lake Catchment and Water Level During the Last Sixty Years. European Water, 56, pp. 3 – 12.

Karamoutsou L. and Psilovikos A., 2019. The Use of Artificial Neural Networks in Predicting the Water Quality of Lake Kastoria. Proceedings of the 14thConference of the Hellenic Hydrotechnical Association (EYE), pp. 882 - 889, Volos.

Katsiapi, M., Mazaris, A. D., Charalampous, E., and Moustaka-Gouni, M., 2012. Watershed land use types as drivers of freshwater phytoplankton structure. In *Phytoplankton responses to human impacts at different scales*, Springer, Dordrecht, pp. 121-131.

Matzafleri N., 2007. Spatial Simulationof the Water Quality Parameters of Lake Kastoria. MSc in Sustainable Management of Aquatic Environment, Department of Ichthyology & Aquatic Environment, 254 p., Volos, Greece.

Mantzafleri N., Psilovikos A. and Mplanta A. 2009. Water quality monitoring and modeling in Lake Kastoria, using GIS. Assessment and management of pollution sources. *Water resources management*, 23(15), pp. 3221-3254.

Mavromati, E., Kagalou, I., Kemitzoglou, D., Apostolakis, A., Seferlis, M. and Tsiaoussi, V., 2018. Relationships Among Land Use Patterns, Hydromorphological Features and Physicochemical Parameters of Surface Waters: WFD Lake Monitoring in Greece. *Environmental Processes*, 5(1), pp.139-151.

Naselli-Flores, L.U.I.G.I., 2008. Urban lakes: ecosystems at risk, worthy of the best care. In Proceedings of Taal2007: the 12th world lake conference (Vol. 1333, p. 1337).

Papastergiadou, E.S., Retalis, A., Kalliris, P. and Georgiadis, T., 2007. Land use changes and associated environmental impacts on the Mediterranean shallow Lake Stymfalia, Greece. In Shallow Lakes in a Changing World (pp. 361-372). Springer, Dordrecht.

Psilovikos A., 2005. The Contribution of Monitoring to the Rational Management and Prevention of Environmental Risks on the Nestos Transboundary River. The perspective for the implementation of Directive 2000/60. *Hydrotechnica*, Vol15, pp. 87 – 102 (inGreek).

Psilovikos A., 2012. Monitoring System Documentation and Management of Kastoria Lake. Research Program, University of Thessaly. Phase A, B, C. ANKO, Volos, 2011 – 2015 (inGreek).

Psilovikos A., 2014. Eco-Hydraulics, Tziolas Pub, Thessaloniki, Greece, 536 p., ISBN:978-960-418-447-7 (in Greek).

Psilovikos A., 2020. Water Resources, Tziolas Pub, Thessaloniki, Greece, 700 p., ISBN:978-960-418-602-0 (in Greek).

Shi, P., Zhang, Y., Li, Z., Li, P. and Xu, G., 2017. Influence of land use and land cover patterns on seasonal water quality at multi-spatial scales. Catena, 151, pp.182-190.

Skoulikidis, N.T., Bertahas, I. and Koussouris, T., 1998. The environmental state of freshwater resources in Greece (rivers and lakes). Environmental Geology, 36(1-2), pp.1-17.

Watt, M.K., 2000. A hydrologic primer for New Jersey watershed management (No. 4140). US Department of the Interior, US Geological Survey.

Wiersma G.B., 2004. Environmental Monitoring. CRC Press, NY.

Williamson, C.E., Saros, J.E., Vincent, W.F. and Smol, J.P., 2009. Lakes and reservoirs as sentinels, integrators, and regulators of climate change. Limnology and Oceanography, 54(6part2), pp.2273-2282.

Yu, S., Xu, Z., Wu, W. and Zuo, D., 2016. Effect of land use types on stream water quality under seasonal variation and topographic characteristics in the Wei River basin, China. Ecological Indicators, 60, pp.202-212.

CHAPTER 4: LITERATURE REVIEW

"Become acquainted with …
the minds of many people"
Omiros (8th century, BC)

The present chapter refers to: a) the recording and evaluation of the existing researches and studies with respect to the monitoring and the sustainable management of Lake Kastoria (case study) b) the recording of the existing policy background that has been implemented in Greece for water resources management, water quality conservation and restoration, and the protection of aquatic ecosystems and c) the recording implementation of Machine Learning in order to evaluate the water quality of aquatic systems.

4.1 Studies with respect to the Sustainable Management of Lake Kastoria

During the last 50 years, many studies and research projects have been carried out for Lake Kastoria focusing on its:

a) Ecological status (biological parameters - ecological quality) (Gerasimou et al., 1970; Kousouris et al., 1991; Gkelis et al., 2005; Moustaka-Gouni et al., 2006a; Moustaka-Gouni et al., 2006b; Matzafleri et al., 2009; Psilovikos, 2020),

b) Surface and Groundwater balance (Kousouris et al., 1991; Psilovikos and Zarkadas, 2006),

d) Sediment deposition (Panagos et al., 1989),

e) Spatial simulation of the pollutants and water quality (Psilovikos et al., 2006; Matzafleri et al., 2009; Psilovikos, 2020),

f) Modelling (Hrissanthou et al., 2003; Psilovikos and Zarkadas, 2006; Matzafleri et al., 2009); and

g) Legislation (Psilovikos, 2008; Mavromati et al., 2017; Mavromati et al., 2018).

The study of Gerasimou et al. (1970) is an important work not only from a historical but also from a research point of view, since it records the problems of the lake that appear more than fifty years ago but still remain the same. It refers to the occurrence of hydrogen sulphide and methane, gases which occur in conditions of lack of dissolved oxygen due to the reductive - anaerobic activity of specific bacteria in areas close to the bed. It also refers to the existence of phytoplankton and classifies Lake Kastoria as a eutrophic lake. It works as an inspiration for ecological awareness not only for local authorities but also for the inhabitantsas the protection of the ecosystem is a duty of all. Comparing this study with current data, it has been observed that eutrophication phenomena remain the same. These phenomena seasonally show peaks due to the pollution of the lake with nutrient loads. It is remarkable that the fact that the lake has been in a steady state for fifty years is a proof that the buffering capacity of the lake kept it alive, denying pessimistic predictions and fears that the lake is heading to disaster. However, apart from natural processes, human interventions are crucial in order to contribute to its sustainable management.

Twenty years later, Kousouris et al. (1991), address the problem of eutrophication of Lake Kastoria with all its negative consequences on the quality of water. According to their calculations and field analyses of the main surface runoff, the annual load of total phosphorus entering the lake was estimated about 15000kg. There is a large discrepancy between the natural loads of total phosphorus (26 μg/L) and the total actual natural and anthropogenic loads (63 μg/L). Satellite images show that in the wetland of Lake Kastoria there is a 1,3 km^2 area covered by reeds (Phragmitesaustralis) and water chestnut (Trapanatans). From the authors' research, it has been found that each gram of reed dry weight contains 2 mg of phosphorus, while there are 800 g of reed dry weight per m^2. It is therefore proposed the periodic cutting 0,8 km^2 of reeds per year, out of a total of 1,3 km^2, so that aquatic vegetation management will not impair the biological balance of the lake. With this type of management, 8,5% of total phosphorus it is expected to be removedfrom the lake on a yearly basis.

Moreover, Professor Moustaka – Gouni's scientific team characterizes Lake Kastoria as a eutrophic - hypertrophic lake with seasonal peaks of phytoplankton and cyanobacteria. Their work is summarized as follows:

A) The microbiology of Lake Kastoria and the recording of three colonies of the cyanobacterium Limnothrix redekei (Gkelis et al., 2005) and Raphidiopsis mediterranea (Moustaka et al., 2009)

B) The seasonal dynamics of different phytoplankton colonies in the hypertrophic lake which shows seasonal peaks of toxic cyanobacteria (Moustaka-Gouni et al., 2006a) and

C) The composition of phytoplankton colonies in Lake Kastoria on a yearly basis. (Moustaka-Gouni et al., 2006b)

Awater balance model is used in order to estimate runoff, from the catchment to the water body of Lake Kastoria, by Psilovikos and Zarkadas (2006). The model requires data series of rainfall and potential evapotranspiration on a monthly basis. Rainfall has been estimated using the rainfall gradient method from a number of stations at different altitudes. With the help of GIS, the rainfall dataset is extracted to the weighted average elevation of the basin. The monthly potential evapotranspiration was calculated by the empirical methods of Turc, Thornthwaite and Blaney-Griddle. By introducing the values of these three methods into the hydrological model, it is observed that the monthly runoff obtained when using Thornthwaite and Mather (1955) method fits better the data than the Turc and Blaney-Griddle methods which underestimate the values of total runoff. The model is calibratedbased on the coefficient k, and a sensitivity analysis took place, for various valuesfrom 0,18 to 0,541. The different values of the coefficient k result from the estimations of the geological map of Greece on a scale of 1:500.000 (IGME, 1985).

A sedimentary research of the bed of Lake Kastoria, was conducted by Panagos et al., (1989) resulting in the creation of a map illustrating the fine and coarse sediment zones. According to IGME (2011), it was found that a much smaller part of the lake bed was occupied by fine sediments compared to 2011, which is verified from the testimonies of the inhabitants who swam thirty years ago in areas that are now covered by silt and clay. Almost all of the lake has been covered with fine material. The fact that these fine materials are deposited on pre-existing coarse depositions is because of the existence of low kinetic activities, that enables only the fine material movement.

With the help of Geographical Information Systems (GIS), two deterministic spatial simulation algorithms of the water quality parameters of Kastoria Lake were applied (Psilovikos et al., 2006; Matzafleri et al., 2009; Psilovikos, 2020): a) "Splines - RBF" and b) "IDW - Inverse Distance Weighting" methods. Water samples of lake's surfacewere taken on a monthly basis from three or five sampling points. Comparing the aforementioned deterministic methods, it could be concluded that a) "Splines – RBF" method cannot create closed curves by using 3 points, while b) the "IDW - Inverse Distance Weighting" method gives better results with "Bulleyes" closed curves around the station.

The importance of an automated monitoring that could take measurements on a 24-hour basis and could provide a powerful tool on stakeholders' hands is underlined by Psilovikos (2008 and 2020). In the case of Lake Kastoria, this tool has the following advantages:

- ✓ Continuous monitoring of the quantitative and qualitative parameters of water
- ✓ Direct identification of sources of pollution
- ✓ Sustainable conservation of the aquatic ecosystem (Lake Kastoria - Natura)
- ✓ Prevention of environmental and hydrological hazards (floods, pollution)
- ✓ Creating reliable time series data, to support the implementation of simulation and optimization models, for various management plans

Matzafleri et al., (2009) conducted an important work focusing on:

- ✓ Monitoring: Lake Kastoria is monitored on a monthly basis, from October 2002 to July 2006, at 5 sampling positions. A variety of parameters are monitored, namely Water Temperature (Tw), Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), pH, Electrical Conductivity of Water (ECw), Redox Potential (REDOX), AmmoniaNitrogen (NH_4-N), Nitrite Nitrogen(NO_2-N), NitrateNitrogen(NO_3-N) and Orthophosphates (SRP – PO_3-P). This work, focuses onthe assessment of the following parameters: Tw, DO, NH_4-N, NO_2-N and NO_3-N from October 2005 to July 2006.
- ✓ Modeling, by using the above spatial interpolation and geographical distribution algorithms of the aforementioned quality parameters for four periods-seasons: February 2005 & 2006, July 2005 & 2006 and their comparison.
- ✓ Pollution assessment and proposals – management practices – to reduce pollution and determine more appropriated monitoring positions.

What results from the analysis of the measurements is that Lake Kastoria shows signs of a gradual improvement in water quality. This leads to the conclusion that the buffering capacity of the systemstill operates. However, it also requires a systematic measurement of data at regular intervals, with integrated Real-Time Monitoring Systems to prevent natural and anthropogenic hazards and disasters. In addition, the preparation of these time series will help in the integrated and sustainable management of the lake and its catchment area.

Hrissanthou et al. (2003), simulate the natural processes of the Kastoria Lake Basin with the help of numerical models. The resulting complex numerical model requires data derived both from time series recordings from the existing meteorological stations and from in-situ measurements. For the erosion model, the Schmidt model is used. This sub-model of erosion is combined with a simplified rainfall-runoff model and a sub-model of sediment transport in the hydrographic network. In this way, a complex mathematical model is applied, which calculates both the total volume of water and the sediment transport and deposition – as a percentage of total erosion – at the sub-catchments. All of the above are introduced on a digital mapping background with the help of GIS. Finally, a separate numerical simulation model for groundwater flows, namely MODFLOW, is used to calculate the volume of water transferred from the lake to the aquifer.

Based on the national monitoring network of waters in Greece, in the context of Water Framework Directive 2000/60/EU, the goals of Mavromati et al., (2017; 2018) studies are i) to cite the main features of Greek lakes; ii) to classify them according to hydromorphological and physicochemical data from the first period of monitoring and iii) to investigate how the hydromorphological features influence water quality, iv) to present the pressures resulting from land cover and population density at river basin level and v) to link catchment area features with physicochemical results from the first period of WFD monitoring. The influence of catchment area's size and lake morphological parameters to eutrophication and to water quality in general, is indicated by the strong relationship between Schindler's ratio and Mean Depth with TP. Moreover, land cover and population density of the catchments play an important role in the physicochemical features of lakes in Greece, as boxplots revealed a pattern of high population densities and intensive agriculture related to high TP values.

4.2 Policy background

Joint Ministerial Decision 46399/1352/86 (Government Gazzete 438/B/03.07.1986) adopted national standards for water quality appropriate for drinking, swimming, living conditions of fish in freshwaters, sampling methods and frequency and analysis of the surface waters keeping the orders of the EC legislations 75/440/EC, 76/160/EC, 78/659/ EC, 79/923/EC and 79/869/EC, 1986 (OECD, 2000).

Directive 98/83/EC of 3 November 1998 concerns the quality of water intended for human consumption. The objective of this Directive shall be to protect human health from the adverse effects of any contamination of water intended for human consumption by ensuring that it is wholesome and clean. Moreover, Joint Ministerial Decision No. Y2/2600/2001 is an adaptation of Greek legislation to Council Directive 98/83/EC. According to this, drinking water is the water used for human consumption, either in its natural state or after being treated, whether it is supplied by a distribution network or packaged in bottles or containers. The main objective of this decision was to protect human health from the effects of pollution and contamination of water intended for human consumption ensuring that it is under the norms of the present legislation.

Water Framework Directive (WFD) 2000/60/EC (Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy) establishes the framework for the biological, quantitative and qualitative status of water resources in the countries of the European Union. The above directive addresses the serious problems that have arisen at European level due to the deterioration of water quality and the increased demand for good quality water. The Directive is an innovative legislative framework, aimed at the integrated management and protection of surface, ground, transitional and coastal waters whose quality is determined by biological, chemical, morphological and hydrological criteria (http://wfdver.ypeka.gr/).

The Directive combinesquality, ecological and quantitative targets for the protection of aquatic ecosystems and the good status of all water resources and sets the concept of integrated management at the geographic scale ofRiver Basins. In addition, it redefines the concept of the River Basin, which includes the inland surface (rivers, lakes), groundwater, transitional (delta, estuary) and coastal ecosystems.

The new water policy is innovating in four areas (http://wfdver.ypeka.gr/):

✓ It aims at integrated water management, in order to achieve sustainability

✓ It takes, as a basis, the integrated water management of the ”hydrological basins”

✓ It recognizes the human needs and the importance of water, in the functions of ecosystems

✓ It provides for public participation in planning, decision making and monitoring of the implementation of the water policy

The Water Framework Directive 2000/60/EC (WFD) is widely accepted as the most substantial and ambitious piece of European environmental legislation to date. It has been referred to as a once in a generation opportunity to restore Europe's waters and a potential template for future environmental regulations. However, seventeen years since it was adopted (Vourvoulis et al., 2017) and with many problems and delays in its implementation, the WFD has not delivered yet its main objectives of non-deterioration of water status and the achievement of good status for all EU waters.

The harmonization of Greek legislation with the Community Framework Directive 2000/60 /EC was carried out byLaw 3199/2003 (Government Gazette 280/A/09.12.2003) and the Presidential Decree 51/2007 (Government Gazzete 54/A/08.03.2007). The requirements of the above national legislation incorporate the basic concepts of the Water Resources Directive.Moreover, the new administrative structure is established, as well as the responsibilities of the individual parties at both national and regional level.The main priority of the Special Secretariat for Water is to draw up the River Basin Management Plans of the country’s 14 River BasinDistricts in accordance with community and national legislation (http://wfdver.ypeka.gr/).

Greece has fully transposed the Directive into its national legislation by Law 3199/2003 “Water Protection and Management – Harmonization with Directive 2000/60 / EC of the European Parliament and of the Council of 23 October 2000” and the Presidential Decree 51/2007. By virtue of the provisions ofLaw 3199/2003, individual Decisions have been issued, such as theJoint

Ministerial Decision 140384 / 19.8.2011 (Government Gazette B' 2017)"Definition of National Network for the Quality and Water Quality Monitoring by Determination of Measurement Positions and the bodies that are obliged to operate pursuant to article 4, paragraph 4 of Law 3199/2003", theDecision 706 /16.07.2010(Government Gazette B' 1383/2010)of the National Water Committee"Determination of River Basin country and the definition of the relevant Regions for the management and protection", pursuant to article 3 of the Presidential Decree 51/2007, following theapproval of River Basin Management Plansof the 14 River Basin Districts of the country for the 1stCycle of Management (2009-2015), and others (http://wfdver.ypeka.gr/).

Directive 2006/118/EC of the European Parliament and of the Council of 12 December 2006 on the protection of groundwater against pollution and deterioration concerns the protection of groundwater against pollution and degradation. Its aim is to adopt appropriate measures to prevent groundwater pollution and the deterioration of their status, thus ensuring the protection of public health and the environment. The groundwater Directive 2006/118/EU establishes measures in order to prevent and control pollution. The measures include: a) criteria for the assessment of good groundwater chemical status; and b) criteria for the identification and reversal of significant and sustained upward trends and for the definition of starting points for trend reversals. This Directive also complements the provisions preventing or limiting inputs of pollutants into groundwater already contained in Directive 2000/60/EC, and aims to prevent the deterioration of the status of all bodies of groundwater(https://eur-lex.europa.eu/).

According to Joint Ministerial Decision No 39626/2208/2009 (Government Gazzete B' 2075/25.09.2009), the measures for the protection of groundwater against pollution and deterioration in accordance with the provisions of Directive 2006/118/EC and the provisions relating to the prevention and restriction of groundwater pollution by the Law 3199/2003 and Presidential Decree 51/2007are determined.

Moreover, the goal of Ministerial Decision 51354/2641/E103/2010 (Government Gazzete 1909/B/ 08.12.2010). is to set environmental quality standards for the concentrations of certain pollutants in surface water in accordance with the requirements of Directive 2006/105/EC of the European Parliament and of the Council of 16thDecember 2018, on environmental quality standards in the field of environmental policy, on the abolition of Council Directives 82/176/EEC, 83/513/EEC, 84/156/EEC, 84/491/EEC and 86/280/EEC and the amendment of Directive 2000/60/ EC of the European Parliament and of the Council, as well as on concentrations of pollutants in surface water.

4.3 Machine Learning in Water QualityManagement

In such a developing world, as water is extensively polluted, many models have been applied to predict water quality parameters in order to protect aquatic ecosystems. The concern of scientists, in the applications of Artificial Neural Networks (ANN) in Hydrology and Water Resources Management, has been increasing in recent years. The application of Artificial Neural Networks (Traditional or Deep) has been introduced in the field of hydrological studies and water balance parameters, linear and non-linear, such as precipitation, evaporation, runoff and various water quality parameters and it has been concluded that the results produced are quite satisfactory in that concerns the field of forecasting (Diamantopoulou et al., 2005; Emamgholizadeh et al., 2014; Alizadeh et al., 2015; Antonopoulos et al., 2016; Sentas et al., 2016; Sentas et al., 2018; Karamoutsou and Psilovikos, 2019; Lu et al., 2019; Zhai et al., 2019; Chenand Wang, 2020; Mellios et. al., 2020)

French and el. (1994) were first applied a Feed-Forward ANN to make predictions of abundance of species of phytoplankton in Saidenbach Reservoir in Germany. Karul et al. (2000) used a three-layer Levenberg–Marquardt Feed Forward learning algorithm to model the eutrophication process in three water bodies of Turkey (Keban Dam Reservoir, Mogan and Eymir Lakes). Also, Cho et al. (2014) usea Feed Forward Neural Network with the help of Levenberg – Marquardt method in order to predict chlorophyll-a (Chl-a) levels at Lake Juam, with satisfactory results.

Lek and Guean (1999) introduced two kinds of ANNs (Back Propagation and Kohonen Self-Organizing Network) in ecology. Yabunaka et al. (1997) also applied a Back Propagation Artificial Neural Network model, to predict algal bloom by simulating the future growth of five phytoplankton species and the Chlorophyll-a (Chl-a) concentrations. Also, a Feed-Forward Back-Propagation Artificial Neural Network model, including one hidden layer, was built by Lu et al. (2019) to predict the Total Phosphorus (TP) concentration in the Lake Champlain, by using the following input variables for TP prediction:

a) Dissolved Inorganic Phosphorus (DIP), b) Secchi Depth (SD), c) Total Nitrogen (TN), d) Dissolved Oxygen (DO), e) Water Temperature (Tw) and f) Total Phosphorus inflow rate (TP inflow). Kuo et al. (2007), used four Back Propagation Neural Networks, in order to relate the key factors that influence Dissolved Oxygen (DO), Total Phosphorus (TP), Chlorophyll-a (Chl-a), and Secchi Disk Depth (SD) in a reservoir in central Taiwan. Zhai et al. (2019) predict the pH values in order to evaluate the water quality by using three types of Artificial Neural Networks namely a) Back Propagation, b) Radial Basis Function, c) Generalized Regression in the Taihu Lake in China.

As it has already mentioned, Neural Networks are popular tools for modeling highly complicated relationships, processes, and phenomena. Few studies (Khuan et al., 2002; Juahir et al., 2004; Gazzaz et al., 2012; Kadam el al., 2019) have been conducted in order to develop predictive models of the Water Quality Index for rivers by using more complex Neural Networks.Gazzaz et al. (2012), design a Feed-Forward, three-layer Perceptron Neural Network model for computing the Water Quality Index (WQI) for Kinta River in (Malaysia), while Kadam et al. (2019), use a three-layer Back Propagation algorithm, to predict WQI by using Levenberg–Marquardt method.

Nowadays, Deep Neural Networks have dramatically improved the state-of-the-art in many scientific fields such as environmental studies, as they are able to map the Non – Linear relationships among variables, which are characteristic of ecosystems (Lek et al., 1996). That is the reason why Neural Network techniques have been applied in many case studies to predict Dissolved Oxygen (Ranković et al., 2010; Akkoyunlu et al., 2011; Ay and Kisi, 2012; Han et al., 2012; Wen et al., 2013; Supian et al., 2018), Turbidity (Fan et al., 2016; Khairi et al., 2016; Cao et al., 2019) and Chl-a (Muttil and Chau, 2006; Kuo et al., 2007; Cho et al., 2014;) in rivers and lakes.

Zhang et al. (2019), applied a multi-layer Neural Network to approach complex regression functions and predict trends in Dissolved Oxygen. The aforementioned model gave accurate and better results for predicting the trend of DO, than the typical ANN model, Support Vector Regression (SVR) and Linear Regression Model (LRM).

Also, Ay and Kisi (2012), in order to estimate also DO concentration in two stations, examine the accuracy of two different Artificial Neural Network (ANN) techniques:

a) the Multi-Layer Perceptron (MLP), consisted of 4 variables as inputs, 1 hidden layer of 3 neurons and DO as an output layer and

b) the Radial Basis Neural Network (RBNN), consisted of two layers whose output nodes form a linear combination of the basic functions. The input variables used for the ANN models are pH, Tw, ECw, and Discharge (around 2.000 measurements for the first investigated station and around 5.000 for the second one). Comparative results, indicate that the RBNN model performs better than the other investigated models.

Moreover, Kuo et al. (2007), give useful information for an effective water quality management by predicting concentrations of water quality and eutrophication problems in the Te-

Chi Reservoir in Tawain, by developing four Back Propagation Neural Network models (DO / TP / Chl-a / SD model). The correlation coefficients between predicted values and measured data are well above 0,7 for these four models. The ANN models established by this research, perform well to address the water quality and eutrophication problems of the Te-Chi Reservoir.

Gazzaz et al. (2012), design a Feed-Forward, three-layer Perceptron Neural Network model for computing the water quality index (WQI) for Kinta River in (Malaysia). The best WQI predictions are associated with the Quick Propagation (QP) training algorithm resulted in a very high positive correlation ($r = 0{,}977$, $p < 0{,}01$) with the measured WQI values.

Cho et al. (2014), predict Chlorophyll-a (Chl-a) levels at Lake Juam, using a Feed Forward Network, with the help of Levenberg – Marquardt Method/ Algorithm. For this reason, water quality monitoring data, weather data and hydrologic data are used. Pearson's Correlation Analysis is used, to find the input data that has the best correlation with the value of Chl-a. During the training test, the input data is sequentially excluded one by one, according to the descending order of its correlation. Regarding to Pearson's correlation TOC, pH and atmospheric and water temperatures are the most important parameters of Chl-a prediction. The statistic variances for evaluating the applicability of the ANN, were the coefficient of determination, R^2, Root Mean Square Error (RMSE) and Nash–Sutcliffe Efficiency Coefficient. The ANN trained with the time series data successfully, predicted the Chl-a concentration and provided information regarding the principal factors affecting algal bloom at Lake Juam.

According to the literature, Muhammad et al. (2015), use rare and less used Machine Learning Classifiers in order to classify the water quality class of River Kinta in Malaysia. More specific, in this paper a dataset of Kinta River is used to find the best classification model in determining the quality class of the water. The classifiers tested area) Bayes model/Naive Bayes algorithm, b) Rules Model/Conjunctive Rule Algorithm, c) Trees Model/J48 algorithms, d) Lazy Model/Kstar Algorithms and e) Meta Model/Bagging Algorithms. The experimental phase concludes that, the Lazy model using KStar algorithm has the most outstanding accuracy of 86,67%, which is the best algorithm that can be used for this classification problem, while Meta model using Bagging algorithms, has the worst accuracy (71,85%).

Vijayashanthar et al. (2018), give useful information on the building of a model by focusing on the hyperparameters of the ANN model, in order to forecast the Fecal Indicator Bacteria (FIB) concentrations of the Chicago River. Concerning the optimizers, a) the RootMeanSquare propagation (RMSprop) b) the Stochastic Gradient Descent (SGD) and c) Adaptive moment estimation (Adam) are tested. Concerning the activation functions a) ELU, b) ReLU, and c) Tanh function are tested. The investigated range of the number of neurons is between 5 and 60, while the batch size among 8, 16, 32, 64, and 128. In order to select the optimum input variables (around 2.000 samples) from the proposed 10 (Water Temperature; Turbidity; Daily, 2-day, and 7-day Cumulative Rainfall; River Flow Discharge; Distance from the upstream water reclamation plant; and Number of upstream combined sewer outfalls), the Coefficient of Determination, R^2value, is chosen as the performance metric to indicate the Relative Goodness of fit of the Neural Network. Finally, eight input variables were eventually selected from the ten proposed ones. Concerning the optimal hyper-parameters, the RootMeanSquare propagation (RMSprop) Optimizer performs better than the rest optimizers in this study. Moreover, the number of neurons is set to 30; the batch size is set to 3, while ReLU is used as activation function.

Besides Neural Networks, there are other type models in Machine Learning to predict water quality. Rosly et al. (2006), proposed a multi classifier model to improve accuracy of water quality among different classifiers such as Naïve Bayes (NB), Decision Tree (J48), Sequential Minimal Optimization (SMO), Multi-Layer Perception (MLP), and Instance Based for K-Nearest neighbor (IBK) on water quality for datasets of Kinta River, Perak, Malaysia. Borsuk et al. (2004), proposed a Bayesian network as a promising method for performing integrated ecological modelling. Tan et al. (2012) proposed Least Squares Support Vector Machine in order to predict water quality time series data. Singh et al. (2011) applied both Support Vector Classification (SVC) and Support Vector Regression (SVR) models to the surface water quality data to optimize the monitoring program by predicting BOD values.

Zhang et al. (2019), approach complex regression functions and predict trends in Dissolved Oxygen. For that reason, a) a Multi-Layer Neural Network b) a typical ANN model c) a Support Vector Regression Model (SVM) and d) a Linear Regression Model (LRM) are built. In order to avoid capturing only linear dependencies (for example by using Pearson Correlation (Cho et al., 2014)), Mutual information (MI) for predicting the trend of Dissolved Oxygen is used. MI takes into account the Non – Linear relationships between the variables. The paper results that the Multi – Layer Neural Network, gives more accurate results for predicting the trend of DO, than the rest investigated models.

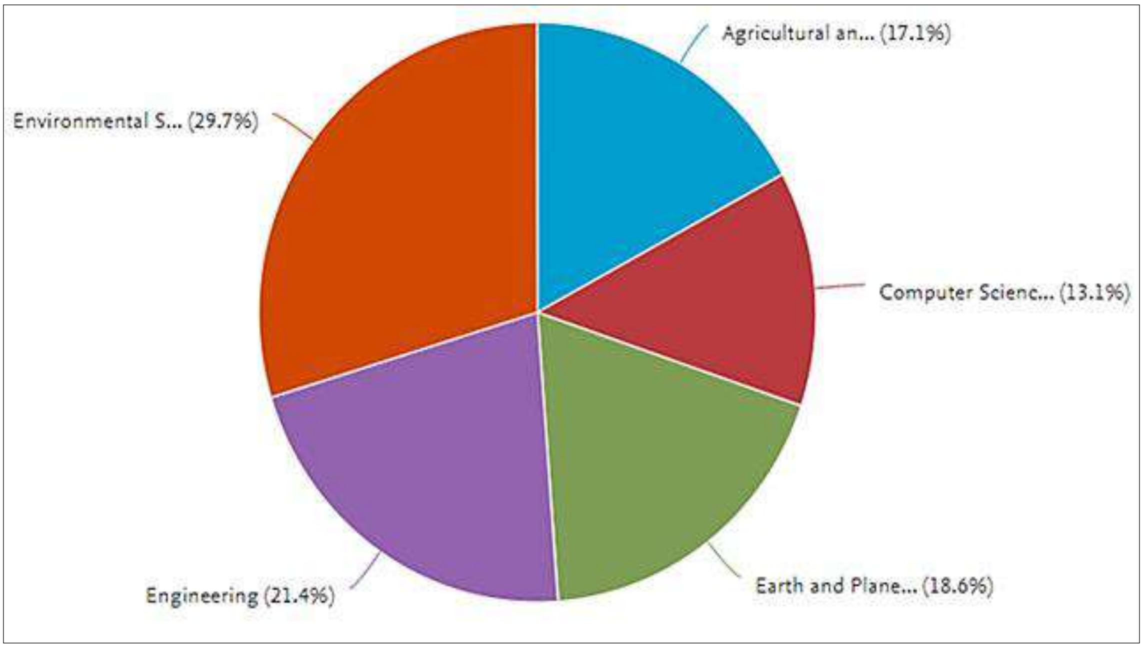

Figure 4.1: The percentage of documents found in Scopus related to "Neural Networks" or "Machine Learning" and "Water Quality" or "Eutrophication" for different subjects' areas

An increasing trend in the number of the articles studying Machine Learning to predict Water Quality parameters in aquatic systems is observed, during the last fifteen years (Figure 4.2). Especially in the field of our interest, Environmental Subject Area, the percentage of documents found in Scopus reaches almost the percentage of 30% (Figure 4.1).

Concerning the models used in order to predict Water Quality parameters, ANNs show the most impressive rise, from 1995 until 2019, as their use has increased dramatically (Figure 4.3). The other investigated Machine Learning Models namely Bayesian, Ensemble and Support Vector Machines, show also an increasing trend but significantly smaller than those of ANNs. More specific, Support Vector Machines are used more than Ensemble models, while Bayesian models are used less in order to predict water quality parameters, worldwide. However, ANNs have been used almost twice compared with the aforementioned models.

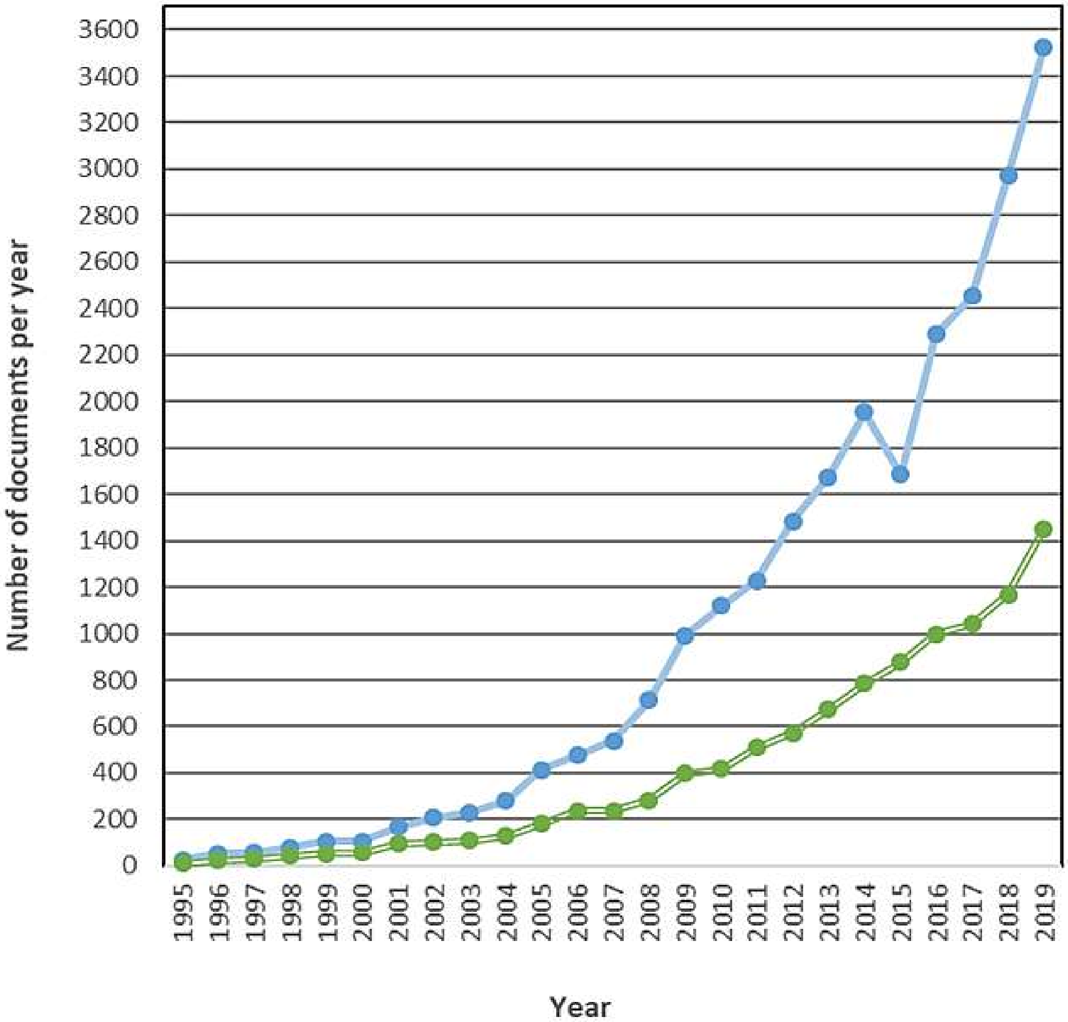

Figure 4.2: Evolution of documents found in Scopus related to “Neural Networks” or “Machine Learning” and “Water Quality” or “Eutrophication” a) for all subject areas (blue line) and b) Environmental Sciences subject area (green line).

Finally, DNN, an upgraded and more complex version of ANN, shows an upward trend mainly from 2010 onwards.The remarkable rise,from 2017 and on,occurs mainly due to the appearance of new platforms and tools (such as Tensorflow, Nvidia’s GPU) that facilitate their use in terms of accuracy and speed.

According to the knowledge of the author and the existing studies, **the importance of this PhD thesis also lies within the fact that there is no other published work using the same structure, platforms and methodology of Deep Learning, as described in Chapter 6, in order to predict water quality parameters of lakes in real time.**

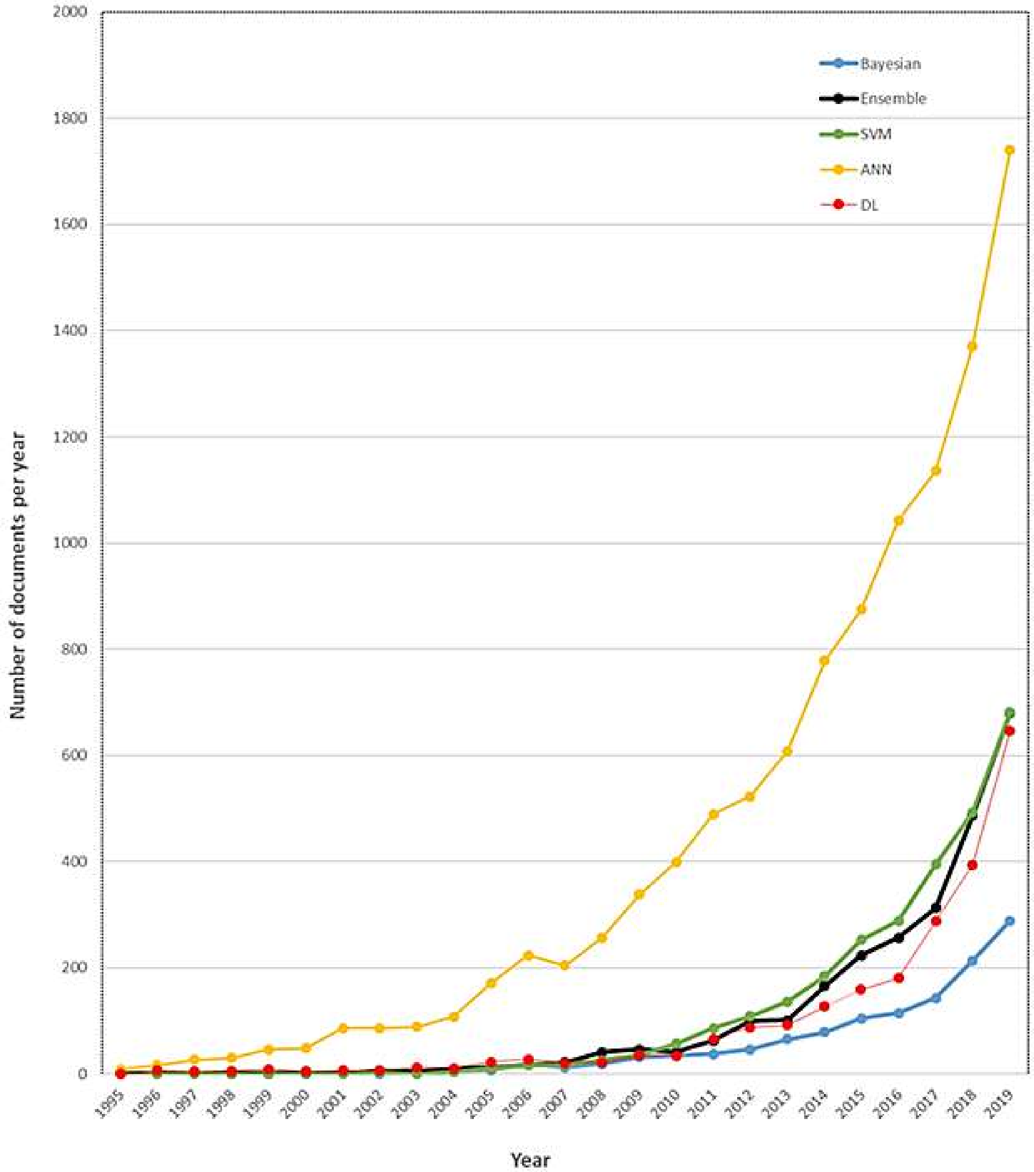

Figure 4.3: Evolution of documents found in Scopus related to a) "Artificial Neural Networks" and "Water Quality" or "Eutrophication" (yellow line) b) "Support Vector Machine" and "Water Quality" or "Eutrophication" (green line) c) "Ensemble Learning" and "Water Quality" or "Eutrophication" (black line) d) "Deep learning" and "Water Quality" or "Eutrophication" (red line) e) "Bayesian models" and "Water Quality" or "Eutrophication" (blue line) for Environmental Sciences subject area

References

Akkoyunlu, A., Altun, H. and Cigizoglu, H.K., 2011. Depth-integrated estimation of dissolved oxygen in a lake. *Journal of Environmental Engineering*,137(10), pp.961-967.

Alizadeh, M.J. and Kavianpour, M.R., 2015. Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean. *Marine pollution bulletin*,98(1-2), pp.171-178.

Antonopoulos, V.Z., Gianniou S.K., Antonopoulos, A.V., 2016. Artificial Neural Networks and empirical equations to estimate daily evaporation: application to lake Vegoritis, Greece. *Hydrological Sciences Journal*,61(14), pp.2590-2599.

Ay, M. and Kisi, O., 2012. Modeling of dissolved oxygen concentration using different neural network techniques in Foundation Creek, El Paso County, Colorado.*Journal of Environmental Engineering*,138(6), pp.654-662.

Borsuk, M.E., Stow, C.A. and Reckhow, K.H., 2004. A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecological Modelling*,173(2-3), pp.219-239.

Cao, P., Zhao, W., Liu, S., Shi, L. and Gao, H., 2019. Using a Digital Camera Combined with Fitting Algorithm and TS Fuzzy Neural Network to Determine the Turbidity in Water. *IEEE Access*,7, pp.83589-83599.

Chen, J. & Wang, Y., 2020. Comparing Activation Functions in Modeling Shoreline Variation Using Multilayer Perceptron Neural Network. *Water*, Vol 12(5), p. 1281.

Cho, S., Lim, B., Jung, J., Kim, S., Chae, H., Park, J., Park, S. and Park, J.K., 2014. Factors affecting algal blooms in a man-made lake and prediction using an Artificial Neural Network.*Measurement*,53, pp.224-233.

Council Directive 98/83/EC. 1998. The quality of water intended for human consumption. OJ L 330, 5.12.1998, pp. 32 - 54.

Diamantopoulou, M.J., Papamichail, D.M., Antonopoulos, V.Z., 2005. The use of a Neural Network technique for the prediction of Water Quality parameters. *Operational Research*, 5(1), pp.115-125.

Directive, W.F., 2003. Common implementation strategy for the water framework directive (2000/60/EC). *Guidance document*, (7).

Directive 2006/118/EC. 2006. The protection of groundwater against pollution and deterioration. European Parliament and of the Council of 12December 2006.OJ L 372, 27.12.2006, pp. 19 – 31.

Emamgholizadeh, S., Kashi H., Marofpoor, I., Zalaghi, E., 2014. Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. *International Journal of Environmental Science and Technology*, 11(3), pp.645-656.

Fan, S.K.S., Kuan, W.H., Fan, C. and Chen, C.Y., 2016. Rainfall threshold assessment corresponding to the maximum allowable turbidity for source water. *Water Environment Research*, 88(12), pp.2285-2291.

French, M., Recknagel, F., 1994. Modeling algal blooms in freshwaters using artificial neural networks. In: Zanetti, P. (Ed.), Computer Techniques in Environmental Studies V, vol. II: Environment Systems. *Computational Mechanics Publications*, Boston, pp. 87–94.

Gazzaz, N.M., Yusoff, M.K., Aris, A.Z., Juahir, H. and Ramli, M.F., 2012. Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Marine pollution bulletin*,64(11), pp.2409-2420.

Gerasimou, I., Simotas, A. andKolangis, S., 1970. Research on the exploitation potential of Lake Kastoria. Ministry of Northern Greece, Directorate for Economic Affairs, Working Group on the Aquatic Environment (in Greek).

Gkelis, S., Rajaniemi, P., Vardaka, E., Moustaka-Gouni, M., Lakaras, T. and Sivonen, K., 2005. Limnothrixredekei (Van Goor) Meffert (cyanobacteria) strains from Lake Kastoria, Greece form a separate phylogenetic group. *Microbial Ecology*, 49:176–182. doi:10.1007/s00248-003-2030-7.

Han, H.G., Qiao, J.F. and Chen, Q.L., 2012. Model predictive control of dissolved oxygen concentration based on a self-organizing RBF neural network. Control

Hrissanthou, V., Mylopoulos N., Tolikas, D. and Mylopoulos, Y., 2003. Simulation Modeling of Runoff, Groundwater Flow and Sediment Trasport into Kastoria Lake, Greece. *Water Resources Management*, 17, 223–242.

Juahir, H., Zain, S.M., Toriman, M.E., Mokhtar, M., Man, H.S., 2004. Application of artificial neural network models for predicting water quality index. *JurnalKejuruteraanAwam* 16 (2), 42–55.

Kadam, A.K., Wagh, V.M., Muley, A.A., Umrikar, B.N. and Sankhua, R.N., 2019. Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India.*Modeling Earth Systems and Environment*, pp.1-12.

Karamoutsou, L. and Psilovikos, A., 2019. The Use of Artificial Neural Networks in Predicting the Water Quality of Lake Kastoria. *Proceedings of the 14th Conference of the Hellenic Hydrotechnical Association (EYE)*, pp.882 - 889, Volos.

Karul, C., Soyupak, S., Cilesiz, A.F., Akbay, N., Germen, E., 2000. Case studies on the use of neural networks in eutrophication modelling. *Ecological Modelling*, 134, 145–152.

Khairi, M.T.M., Ibrahim, S., Yunus, M.A.M., Faramarzi, M. and Yusuf, Z., 2016. Artificial neural network approach for predicting the water turbidity level using optical tomography. *Arabian Journal for Science and Engineering*,41(9), pp.3369-3379

Khuan, L.Y., Hamzah, N., Jailani, R., 2002. Prediction of Water Quality Index (WQI) Based on Artificial Neural Network (ANN). In: Proceedings of the Student Conference on Research and Development, Shah Alam, Malaysia.

Kousouris, Th., Fotis, G., Diapoulis, I., Bertachas, N., Nikolaidis,N. and Gritzalis,K., 1991. The restoration of Lake Kastoria. Theoretical assessment and practical application. Geotechnical Scientific Issues, Volume 2/1991, 49 – 67. GEOTEE. Thessaloniki (in Greek).

Kuo, J.T., Hsieh, M.H., Lung, W.S. and She, N., 2007. Using Artificial Neural Network for reservoir eutrophication prediction. *Ecological modelling*,200(1-2), pp.171-177.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996. Application of neural networks to modelling nonlinear relationships in ecology. *Ecological modelling,* 90, 36–52.

Lek, S., Guean, J.F., 1999. Artificial Neural Networks as a tool in ecological modeling, an introduction. *Ecological modelling,* 120, 65–73.

Lu, F., Zhang, H. and Liu, W., 2019. Development and application of a GIS-based artificial neural network system for water quality prediction: a case study at the Lake Champlain area. *Journal of Oceanology and Limnology*, pp.1-11.

Mavromati, E., Kagalou, I., Kemitzoglou, D., Apostolakis, A. and Tsiaoussi, V., 2017. Linkages between physicochemical status and hydromorphology in Greek lakes under WFD policy. *European Water*,58, pp.273-279.

Mavromati, E., Kagalou, I., Kemitzoglou, D., Apostolakis, A., Seferlis, M. and Tsiaoussi, V., 2018. Relationships among Land Use Patterns, Hydromorphological Features and Physicochemical Parameters of Surface Waters: WFD Lake Monitoring in Greece.*Environmental Processes*,5(1), pp.139-151.

Matzafleri, N., Psilovikos, A. & Mplanta, A., 2009. Water Quality Monitoring and Modeling in Lake Kastoria, using GIS. Assessment and Management of Pollution Sources. *Water Resources Management*, 23(15), pp. 3221 – 3254.

Mellios, N., Moe, S. and Laspidou, C., 2020. Machine Learning Approaches for Predicting Health Risk of Cyanobacterial Blooms in Northern European Lakes. *Water*, 12(4), p.1191.

Moustaka-Gouni, M., Vardaka, E., Michaloudi, E., Kormas, K., Tryfon, E., Mihalatou, H., Gkelis, S., Lanaras, T., 2006. Plankton food web structure in a eutrophic polymictic lake with a history of toxic cyanobacterial blooms. *Limnology &Oceanography, 51* (1, part 2), pp.715–727

Moustaka-Gouni, M., Vardaka, E., Tryfon, E., 2006. Phytoplankton species succession in a shallow Mediterranean lake (L. Kastoria, Greece): steady-state dominance of LimnothrixRedekei, Microcystis aeruginosa and Cylindrospermopsisraciborskii.*Hydrobiologia*, 575(1), pp.129–140. Doi:10.1007/s10750-006-0360-4.

Moustaka-Gouni, M., Kormas, K. A., Vardaka, E., Katsiapi, M., and Gkelis, S., 2009. Raphidiopsis mediterranea Skuja represents non-heterocytous life-cycle stages of Cylindrospermopsis raciborskii (Woloszynska) Seenayya et Subba Raju in Lake Kastoria (Greece), its type locality: Evidence by morphological and phylogenetic analysis. *Harmful Algae*, *8*(6), pp. 864-872.

Muhammad, S.Y., Makhtar, M., Rozaimee, A., Aziz, A.A. and Jamal, A.A., 2015. Classification model for water quality using machine learning techniques. *International Journal of Software Engineering and its Applications*,9(6), pp.45-52.

Muttil, N. and Chau, K.W., 2006. Neural network and genetic programming for modelling coastal algal blooms. International Journal of Environment and Pollution.

OECD (Organization for Economic Cooperation and Development), 2000. Environmental Performance Reviews–Greece.

Panagos, A., Conisoliatis, N. and Varnavas, S., 1989. Texture and composition of the Lake Kastoria sediments, Ann. Geol. PaysHell, XXXIV/1, pp.105 – 122.

Psilovikos, Ar., 2008. The water cycle in Kastoria Lake, the ways of monitoring agricultural pollution and the prospect for the implementation of Directive 2000/60. Kastoria Lake Volunteer Team Meeting, Municipality of Kastoria, Greece, November 2008 (in Greek).

Psilovikos, A., 2020. Water Resources. Tziolas Pub, Thessaloniki, Greece, ISBN: 978-960-418-602-0, 700 pages (in Greek).

Psilovikos, A., Filippou, Chr., Margoni, S., Mantzafleri, N. and Kougolos, A., 2006. Comparison of Two Spatial Simulation Algorithms of the Kastoria Lake Water Quality Parameters Using Geographic Information Systems (GIS*). Proceedings of the 1st Panhellenic Conference entitled: "Small-Scale Wastewater Treatment Plants"*, pp. 25 - 32, Portaria, Pelion, (in Greek).

Psilovikos, A., Zarkadas P., 2006. Water Balance Simulation Model in the Watershed of Kastoria Lake*, Proceedings of the 10th Conference of the Hellenic Hydrotechnical Association (EYE)*. Vol I, pp.63 – 70, Xanthi, Greece, December 2006 (in Greek).

Ranković, V., Radulović, J., Radojević, I., Ostojić, A. and Čomić, L., 2010. Neural network modeling of dissolved oxygen in the Gruža reservoir, Serbia. *Ecological Modelling*,221(8), pp.1239-1244.

Rosly, R., Makhtar, M., Awang, M.K., Rahman, M.N.A. and Deris, M.M., 2006. Multi-classifier models to improve accuracy of water quality application. *Asian Research Publishing Network Journal of Engineering and Applied Sciences*, 11, pp.3208-3211.

Sentas A., Psilovikos A., Psilovikos T., Matzafleri, N., 2016. Comparison of the performance of stochastic models in forecasting daily dissolved oxygen data in Dam-Lake Thesaurus. *DesalinationandWaterTreatment*,57(25), pp.11660-11674.

Sentas A., Karamoutsou L., Charizopoulos N., Psilovikos T., Psilovikos A. and Loukas A., 2018. The use of Stochastic Models for short term predictions of Water Parameters of the Thesaurus Dam, Nestos River, Greece. *3rd EWaS International Conference 2018*, 2 (634), pp.1-6, Lefkada, Greece

Singh, K.P., Basant, N. and Gupta, S., 2011. Support Vector Machines in Water Quality Management.*Analytica Chimica Acta*, Vol 703, Issue 2, pp.152 – 162.

Supian, S., Achmad, K.T.B., Riyadhi, I., Subiyanto, S., Adiana, G., Yusra, A.I. and Mamat, M., 2018. Mathematical model for dissolved oxygen prediction in CirataReservoir, West Java by using artificial neural network*Journal of Fundamental and Applied Sciences*, 10(1), pp.66-78.

Tan, G., Yan, J., Gao, C. and Yang, S., 2012. Prediction of water quality time series data based on least squares support vector machine. *Procedia Engineering*,31, pp.1194-1199.

Vijayashanthar, V., Qiao, J., Zhu, Z., Entwistle, P. and Yu, G., 2018. Modeling fecal indicator bacteria in urban waterways using artificial neural networks. *Journal of Environmental Engineering*,144(6), p.05018003.

Voulvoulis, N., Arpon, K.D. and Giakoumis, T., 2017. The EU Water Framework Directive: From great expectations to problems with implementation. *Science of the Total Environment*,575, pp.358-366.

Wen, X., Fang, J., Diao, M. and Zhang, C., 2013. Artificial neural network modeling of dissolved oxygen in the Heihe River, Northwestern China.*Environmental Monitoring and Assessment*,85(5), pp.4361-4371.

Yabunaka, K.-I., Hosomi, M., Murakami, A., 1997. Novel application of a backpropagation artificial neural network model formulated to predict algal bloom. *Water Sci. Technol*. 36 (5), 89–97

Zhai, W., Zhou, X., Man, J., Xu, Q., Jiang, Q., Yang, Z., Jiang, L., Gao, Z., Yuan, Y. and Gao, W., 2019. Prediction of water quality based on Artificial Neural Network with grey theory. In*IOP Conference Series: Earth and Environmental Science*, IOP Publishing, 295(4), p. 042009

Zhang, Y., Fitch, P., Thorburn, P. and Vilas, M.D.L.P., 2019. Applying multi-layer artificial neural network and mutual information to the prediction of trends in dissolved oxygen. *Frontiers in Environmental Science*,7, p.46.

Internet sources:

http://wfdver.ypeka.gr/

https://eur-lex.europa.eu/

CHAPTER 5: GEOGRAPHIC INFORMATION SYSTEMS AND ECOLOGICAL QUALITY SIMULATION AND ASSESSMENT

"Nothing endures but change"
Heraclitus, (535 – 475 BC)

5.1 Land use changeof Lake Kastoria

Each lake is the recipient of substances due to human activities or natural processes. The main components of these substances are organic loads, phosphorus compounds, nitrogen compounds, various toxic compounds and trace metals. Also, sediment transfer and deposition, that affects both biotic (organisms) and abiotic (geomorphology, hydrology) characteristics, is of paramount importance.Lake Kastoria is a shallow polymictic lake, which means that it is too shallow to developthermal stratification. It has high water temperatures during summer (natural processes) and it is an urban lake with intense agricultural activities in its catchment area (anthropogenic processes) loading both point and nonpoint source pollutants. The presence of high concentrations of nutrients, such as nitrogen and phosphorus salts, indicates the intense both natural processes and agricultural activity in the catchment's area.

Table 5.1: Land uses (1st Level) of the catchment area of Lake Kastoria using Corine Land Cover (2000)

Land Uses of the cathment area of Lake Kastoria using Corine Land Cover (2000)				
A/A	**LAND USES (1st Level)**	**POLYGONS**	**SURFACE AREA (m2)**	**PERCENTAGE (%)**
1	1. ARTIFICIAL SURFACES	**11**	**5115920,182**	**1,824**
2	2. AGRICULTURAL AREAS	**30**	**111178896,200**	**39,629**
3	3. FORESTS AND SEMI-NATURAL AREAS	**86**	**133166163,300**	**47,467**
4	4. WETLANDS	**4**	**2099967,712**	**0,749**
5	5. AQUATIC SURFACES	**1**	**28984848,618**	**10,332**
	Total:	**132**	**280545796**	**100**

Lake Kastoria has a catchment area of 281 km^2. Land uses of Lake Kastoria's basin are determined by its soil morphology, the existing water potential and the development of the area in general. The catchment area of Lake Kastoria is mainly covered by forests and semi-natural areas (47,47%). Agriculture occupies the most of the remaining area corresponding to 39,62% that results in considerable amounts of pollutants being deposited on the ground. The causes of pollution are point and non-point sources, such as urban waste water (despite the fact that since 1995 there is a biological treatment plant), diffuse sources, rainfall in urban and outlying areas, direct rainfall, landfills and general human activities. Lake Kastoria itself occupies 10,33% of the catchment area while its wetlands, a rate of 0,75%.

Tables 5.1, 5.2, and 5.3 present the Land Uses of the catchment area of Lake Kastoria using Corine Land Cover (2000) at 1st, 2nd and 3rd Level (https: //geodata.gov.gr/dataset/corine-2000). Moreover, the Land Uses of the catchment area are illustrated graphically in Figures 5.1 and 5.2 at the 1st and 3rd Level respectively. It should be mentioned that in order to process the maps of the wider area of Lake Kastoria, the software ArcGIS (v.10.2) is used. A Geographic Information System (GIS) is a system which is designed to allow capturing, visualization, analysis, management and interpretation of data in order to understand relationships, patterns, and trends. It constitutes a production mechanism of high quality maps that facilitates the implementation of complex analyses (Karamoutsou et al., 2016; Psilovikos, 2020).

Table 5.2: Land uses (2nd Level) of the catchment area of Lake Kastoria using Corine Land Cover (2000)

Land Uses of the cathment area of Lake Kastoria using Corine Land Cover (2000)				
A/A	**LAND USES (2rd Level)**	**POLYGONS**	**SURFACE AREA (m2)**	**PERCENTAGE (%)**
1	1.1 -URBAN NETWORK	**8**	**4312658,274**	**1,537**
2	1.2 - INDUSTRIAL-COMMERCIAL ZONES AND TRANSPORT NETWORKS	**1**	**195900,679**	**0,070**
3	1.3 - WASTE, WASTE DISPOSAL PLANTS AND BUILDING PLACES	**1**	**256242,017**	**0,091**
4	1.4 - ARTIFICIAL NON-AGRICULTURAL ZONES	**1**	**351119,212**	**0,125**
5	2.1 - ARABLE LAND	**9**	**28842908,435**	**10,281**
6	2.2 - PERMANENT CULTIVATION	**2**	**13879545,680**	**4,947**
7	2.3 - MEADOWS	**2**	**712883,469**	**0,254**
8	2.4 - REPRESENTATIVES AGRICULTURAL AREAS	**17**	**67743558,570**	**24,147**
9	3.1 - FORESTS	**33**	**41026754,450**	**14,624**
10	3.2 - COMBINATIONS OF BUSH OR/AND VEGETATION HERBS	**53**	**92139408,840**	**32,843**
11	4.1 - INLAND WETLANDS	**4**	**2099967,712**	**0,749**
12	4.2 - LAND WATERS	**1**	**28984848,618**	**10,332**
	Total:	**132**	**280545796**	**100**

Table 5.3: Land uses (3rd Level) of the catchment area of Lake Kastoria using Corine Land Cover (2000)

Land Uses of the cathment area of Lake Kastoria using Corine Land Cover (2000)				
A/A	**LAND USES (3rd Level)**	**POLYGONS**	**SURFACE AREA (m2)**	**PERCENTAGE (%)**
1	1.1.2 - Uninterrupted urban tissue	**8**	**4312658,274**	**1,537**
2	1.2.1 - Industrial and commercial zones	**1**	**195900,679**	**0,070**
3	1.3.1 - Mineral extraction sites	**1**	**256242,017**	**0,091**
4	1.4.2 - Sports and Leisure Facilities	**1**	**351119,212**	**0,125**
5	2.1.1 - Irrigated arable land	**9**	**28842908,435**	**10,281**
6	2.2.2 - Fruit trees and fruit-bearing plantations	**2**	**13879545,680**	**4,947**
7	2.3.1 - Meadows	**2**	**712883,469**	**0,254**
8	2.4.2 - Complex crops	**4**	**37885617,873**	**13,504**
9	2.4.3 - Land mainly used for agriculture along with important natural vegetation	**13**	**29857940,698**	**10,643**
10	3.1.1 - Broad-leaved forest	**25**	**35989475,607**	**12,828**
11	3.1.2 - Coniferous forest	**5**	**3621230,396**	**1,291**
12	3.1.3 - Mixed forest	**3**	**1416048,445**	**0,505**
13	3.2.1 - Natural pastures	**23**	**55380991,069**	**19,740**
14	3.2.3 - Sclerophyllous vegetation	**6**	**6842700,034**	**2,439**
15	3.2.4 - Transitional woodland and shrubby areas	**24**	**29915717,732**	**10,663**
16	4.1.1 - Inland swamps	**4**	**2099967,712**	**0,749**
17	5.1.2 - Stagnant water surfaces	**1**	**28984848,618**	**10,332**
	TOTAL:	**132**	**280545796**	**100**

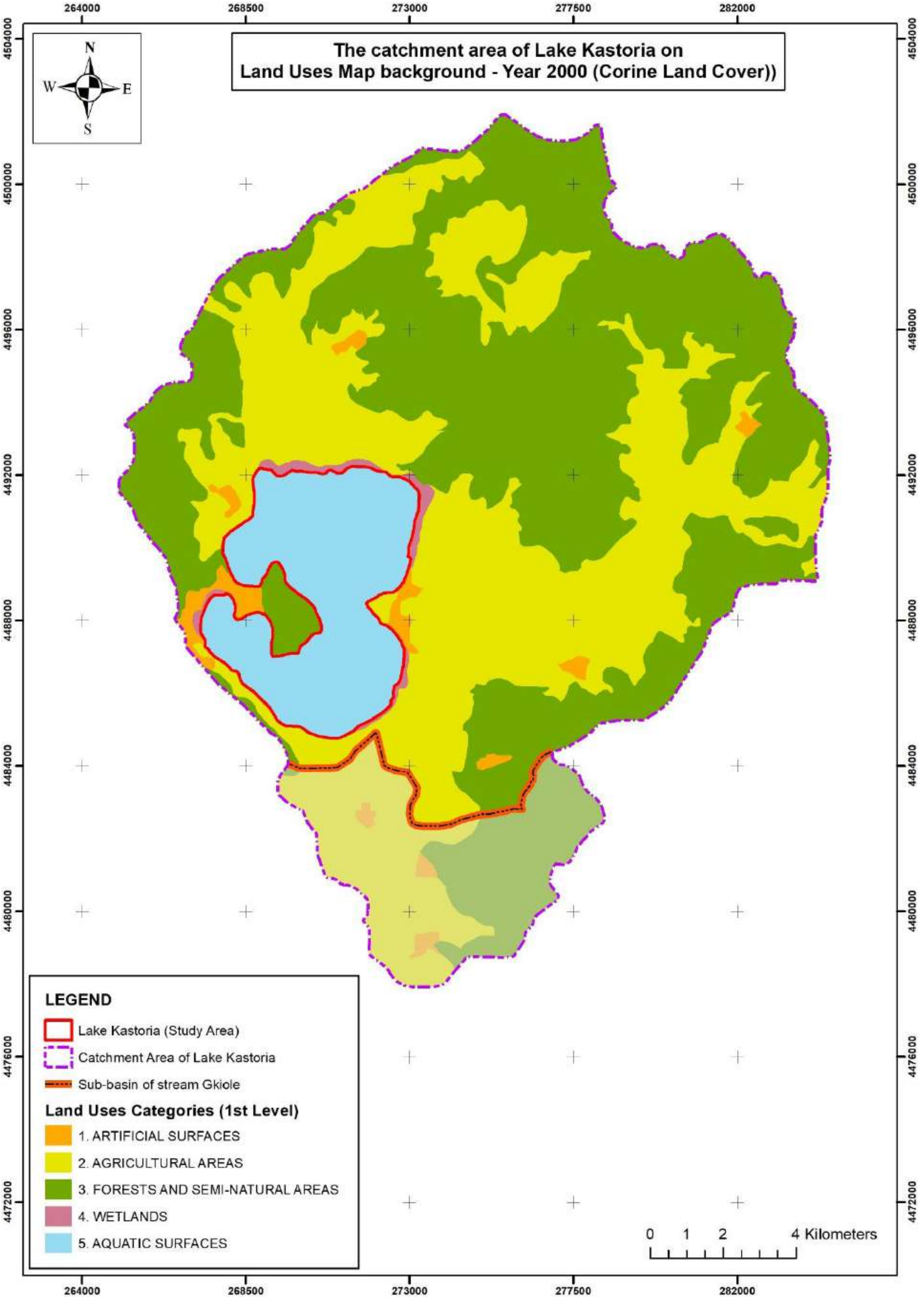

Figure5.1: Land Use Map (1st level) of the catchmentareaofLakeKastoria (CorineLandCover, 2000)

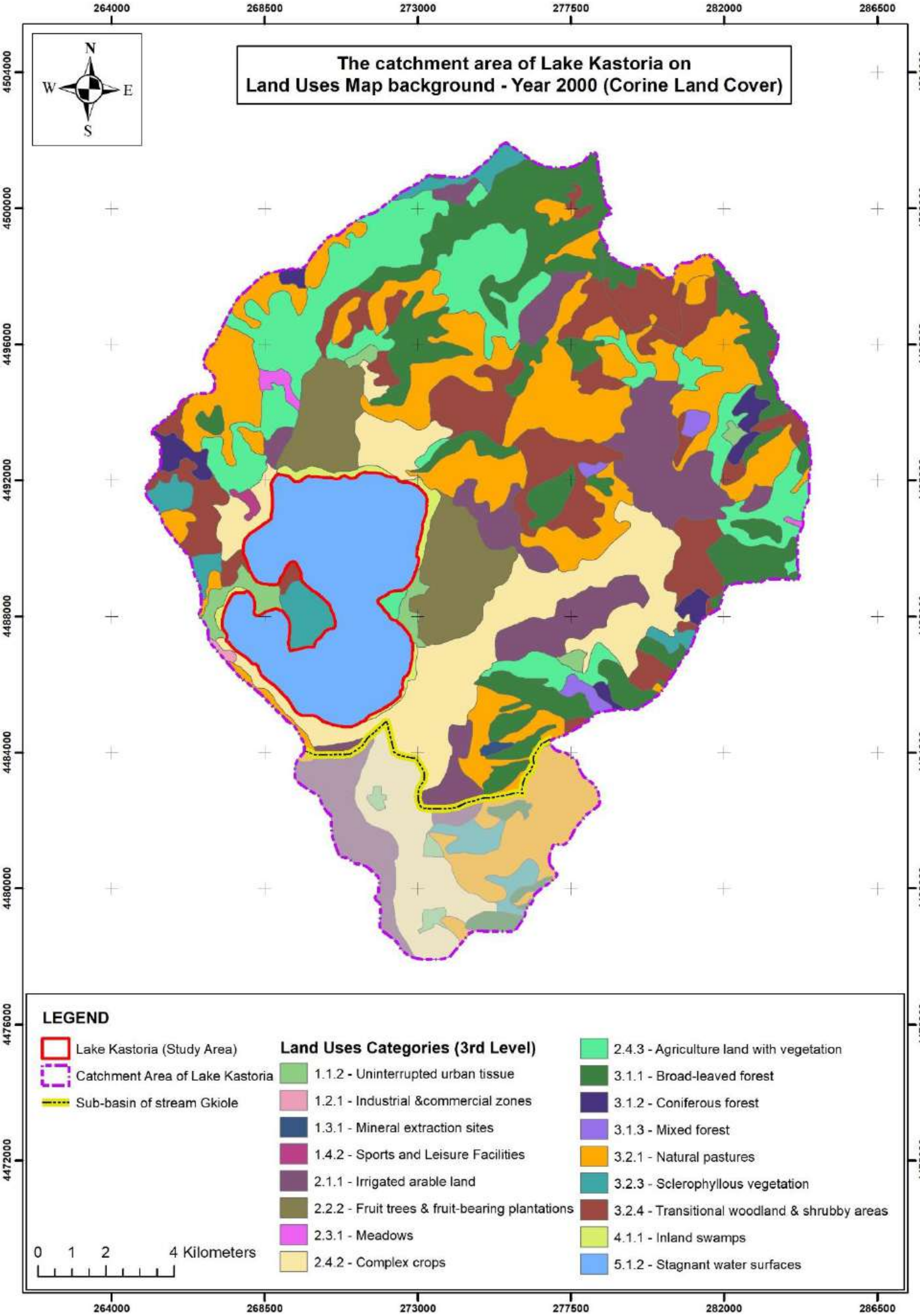

Figure5.2: Land Use Map (3rd level) of the catchmentareaofLakeKastoria (CorineLandCover, 2000)

In order to draw useful conclusions about land use changes in the study area, by comparing the historical year (1945) to the recent year (2000), an orthographic photography of the historical year is compared with the Corine Land Cover (2000). This comparison can be seen schematically in Figure 5.3, where the background of the recent year is 2015-2016 (the most recently available one), in which there are no significant land use changes compared to the year 2000.

For the reliability of the results, Lake Kastoria's basinis demarcated in the historical year at 1^{st}Land Use Level. The boundaries of the catchment areafor the year 1945 are shown in Figure 5.4, while the absolute value and the percentage rates for each land cover and their comparison with the year 2000 are shown in Table 5.4. It is noted that the land use map of the year 1945 iscreated by monoscopic photo interpretation, so the comparison of the land uses between the historical year and the recent one, refers only to the 1^{st}Land Use Level.

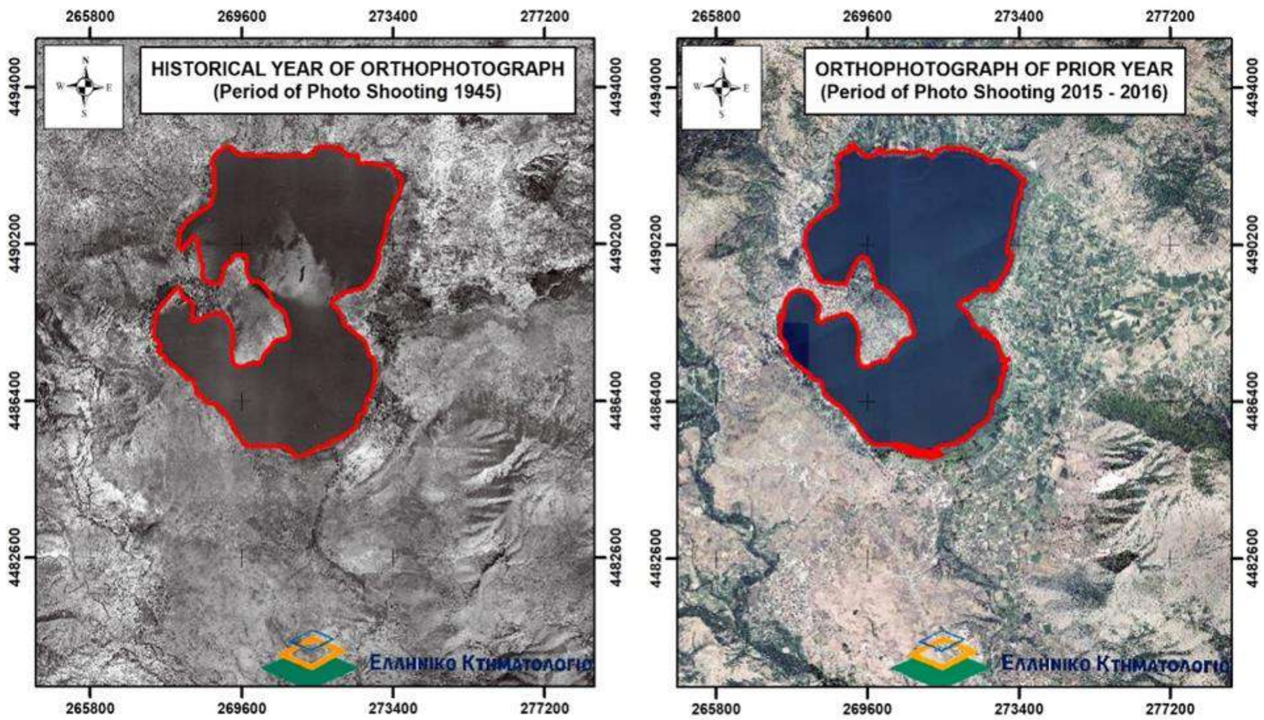

Figure 5.3: Land uses of the catchment area of Lake Kastoria: 1945 (left) and 2015-2016 (right)

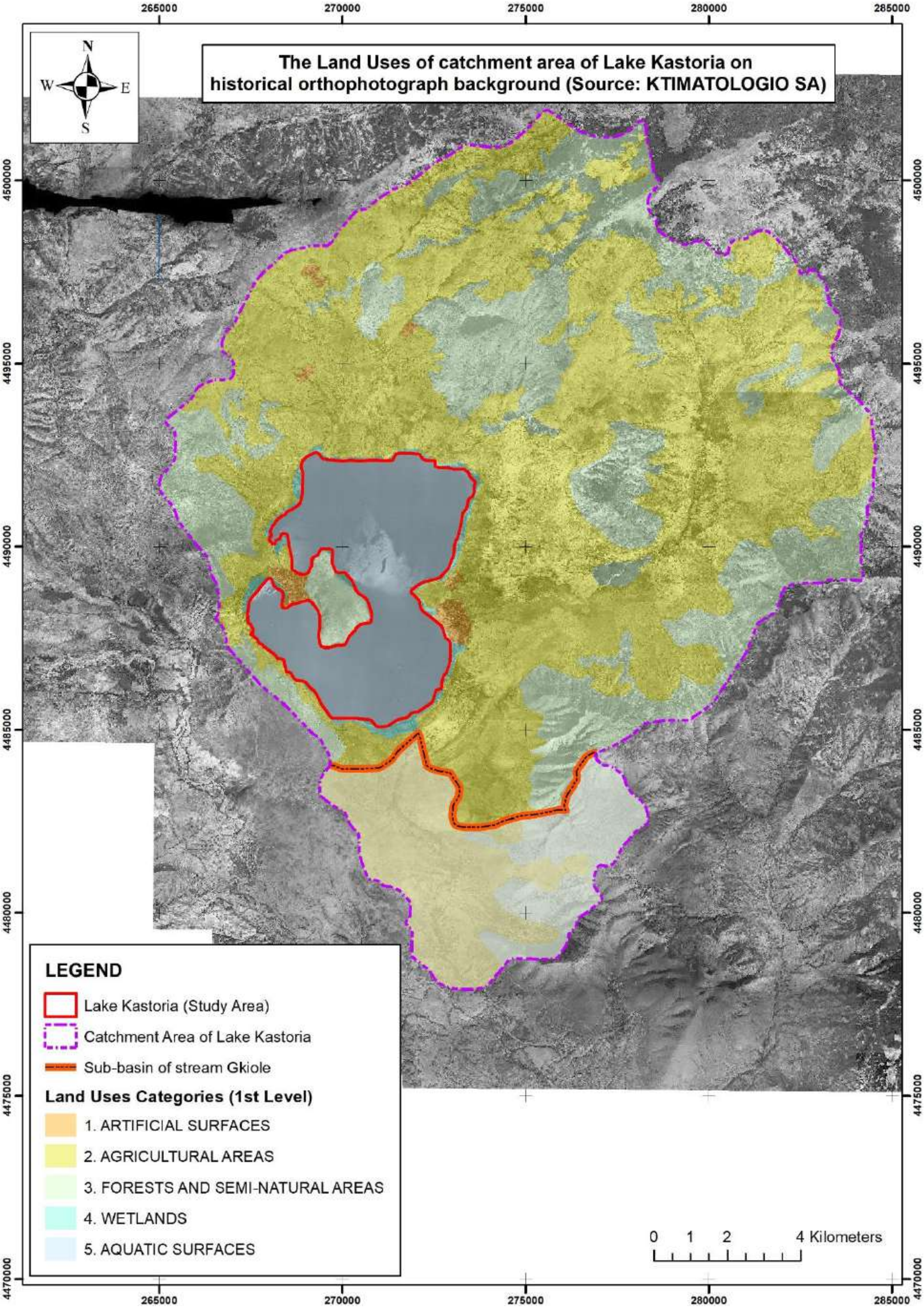

Figure 5.4: LandUse Map (1stlevel) of the catchmentareaofLakeKastoriaatthehistoricalyear (1945)

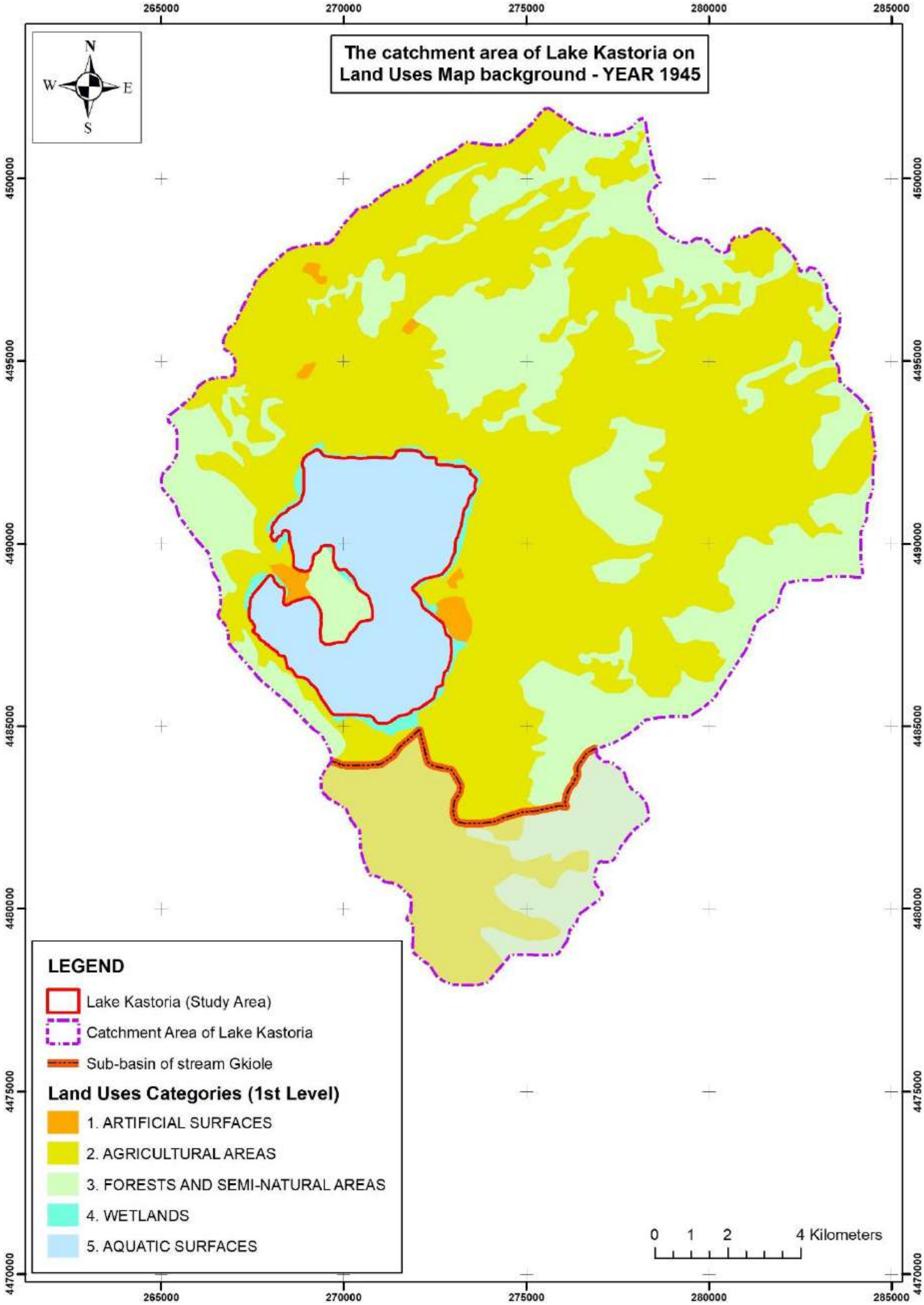

Figure 5.5: Land uses map (1st level) of the catchment area of Lake Kastoria at the historical year (1945)

Table 5.4: Land Useof Lake Kastoria's Basin for the years 1945 and 2000 at 1st Land Cover Level

A/A	LAND USES	YEAR 1945		YEAR 2000	
		METHOD OF ANALYSIS (PHOTOINTERPRETATION)		METHOD OF ANALYSIS (PROCESSING CORINE)	
	1st Level	SURFACE AREA (m2)	PERCENTAGE (%)	SURFACE AREA (m2)	PERCENTAGE (%)
1	1. ARTIFICIAL SURFACES	2153520,189	0,768	5115920,182	1,824
2	2. AGRICULTURAL AREAS	165369272,364	58,946	111178896,200	39,629
3	3. FORESTS AND SEMI-NATURAL AREAS	81759881,391	29,143	133166163,300	47,467
4	4. WETLANDS	3308731,316	1,179	2099967,712	0,749
5	5. AQUATIC SURFACES	27954390,740	9,964	28984848,618	10,332
TOTAL:		280545796	100	280545796	100

As it can be seen from Table 5.4, from 1945 to 2000, the forest and semi-natural areas of Lake Kastoria's catchment area show a rise of 60%, agriculture decreased by 35%, aquatic areas show stability, while wetlands have a remarkable decrease of 60%. Finally, the artificial surfaces doubled their area and in particular they increased by 140% as expected (Figure 5.6).Despite the fact that agricultural and urban sources of pollution have increased, the presence of the biological treatment system since 1995, has significantly reduced the phenomenon of eutrophication.

It is therefore concluded, that the classification of land cover at the 1st reference level of the Kastoria Lake Basin remains the same, although their percentages have changed significantly. These data could probably support further research that has to do with post – classification of the Land Use – Land Changes and can be used as a tool for estimating also the changes in the pollution in the water body of the Lake, according the study of Elhag et al. (2013).

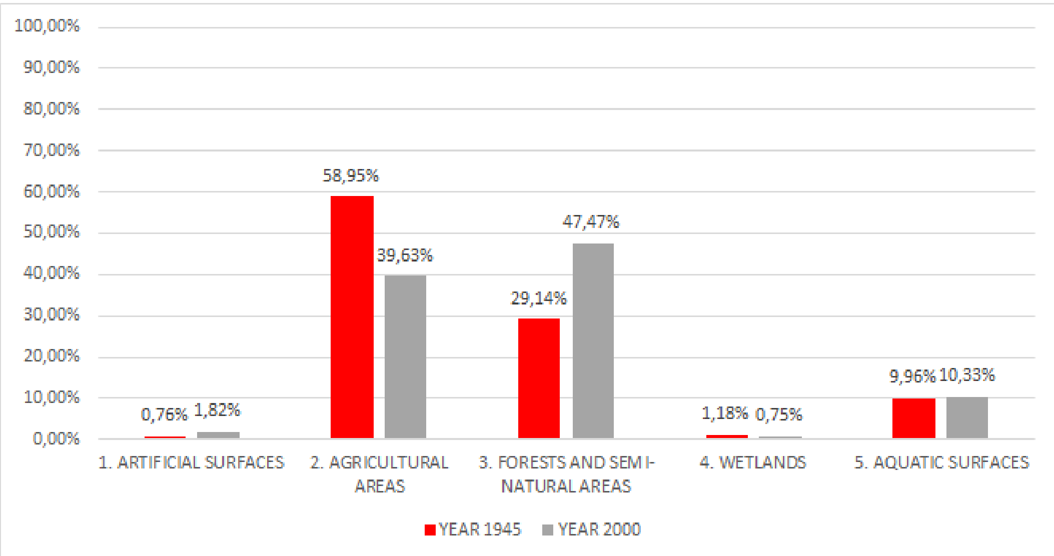

Figure 5.6: Land Cover changes at 1st Level of the Kastoria Lake Basin in 1945 and 2000

5.2 Data collection

For the goal of this dissertation the records of water quality parameters from four stations located in Lake Kastoria are used. In order to record water quality characteristics (at lake's surface), four (4) telemetric stations were installed at specific locations in Lake Kastoria, namely: Gkiole, Stavros, Psaradika and Toichio (Figure 5.7). As it has already mentioned, the above telemetric stations were installed based on the characteristics of the lake and the needs that existed in such a way as to obtain a representative estimate of its water quality parameters. Specifically, the records, on an hourly basis, contain the following parameters:

1. Temperature (Tw – °C),
2. Dissolved oxygen (DO – mg/L),
3. Ammonia nitrogen ($N – NH_4$, mg/L) (NA data for Stavros Station).
4. Nitrate nitrogen ($N – NO_3$, mg/L)
5. Conductivity (Cond – μS/cm),
6. Turbidity (Turb – NTU),
7. pH
8. Chlorophyll-a (Chl-a – μg/L).

Gkiole station is located nearGkiole's stream which is actually an open channel. It is noted that the Gkiole's channelconstitutes the only surface discharge of Lake Kastoria'sbasin, as well as the connection between Lake Kastoria and River Aliakmon. The presence of the weir in the Gkiole's channel, enables the optimization of the lake's hydrological functions, ensuring that the level of the lake will remain under the higher level of 629,80 m, otherwise Kastoria town will be flooded. This is why the geomorphological cathcment of Lake Kastoria is different than the geomorphological one. The latter, does not contain the sub-catchment of Gkiole because it is opening as an outflow of Lake Kastoria to River Aliakmon.

Stavros and Psaradika stations are located on the South and North coastline, respectively,that defines the characteristic coastal places of the city of Kastoria. They constitute crucial stations for measuring the ecological sensitivity of the aquatic ecosystem by the pressures exerted by the city itself.

Finally, Toichiostation is located in the northern part of Lake Kastoria and specifically inthe estuary of a complex Delta group formed by the torrent Toichio and the torrents "Vyssinia" (Lakkou) and "Metamorphosis". The areas at the lowest part of the torrent Toichio are intended for agricultural use.

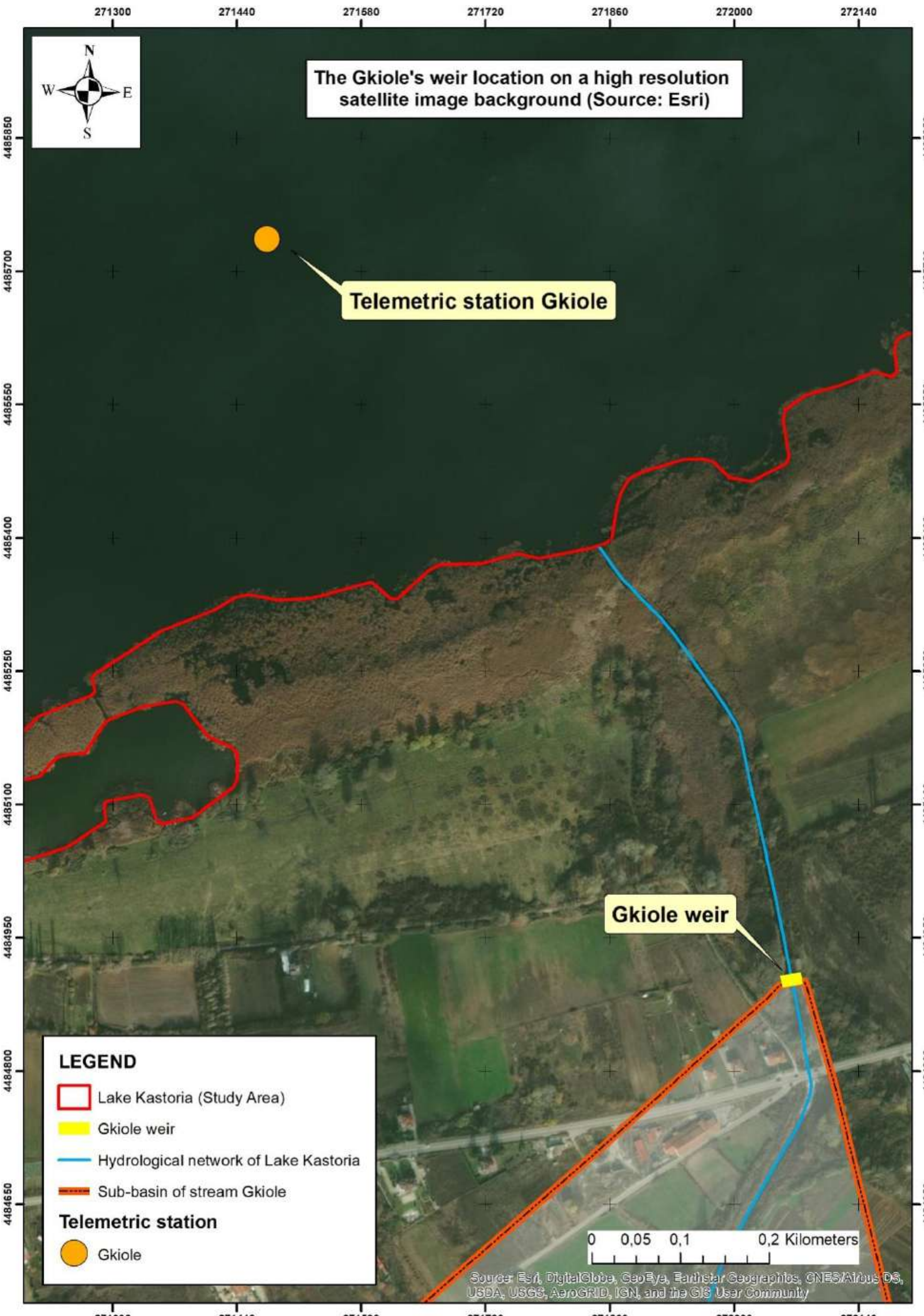

Figure 5.7: The Gkiole's weir location on a high resolution satellite image background (Source: Esri)

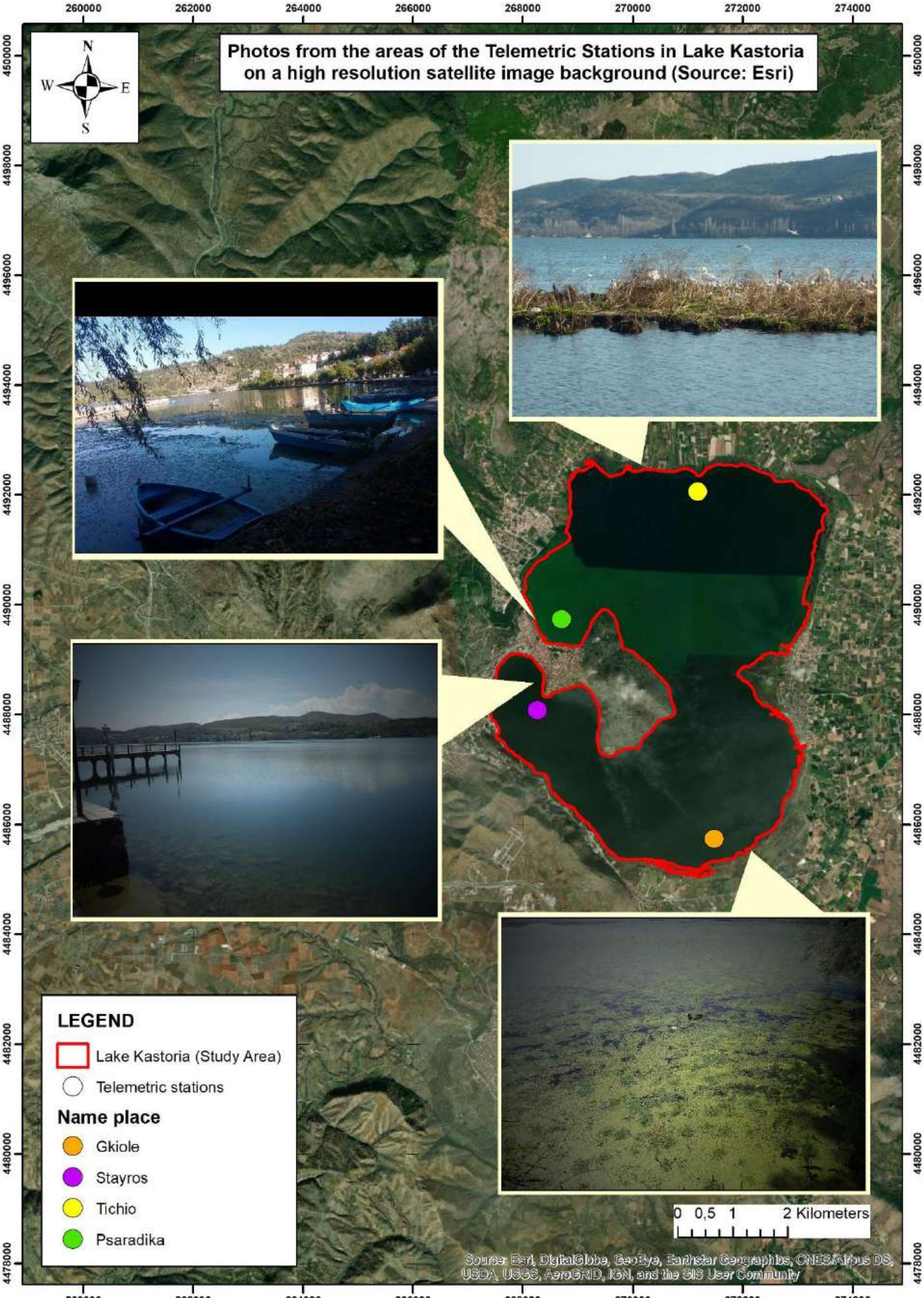

Figure 5.8: Telemetric Stations located in Lake Kastoria

5.2.1 Water Temperature (Tw – °C)

Water Temperature, is an important parameter as it affects the solubility and the reaction time of the substances and it is the major factor that influences the concentration of Dissolved Oxygen in the natural aquatic systems. Temperature controls the rate of biochemical reactions, such as the degradation of organic matter by aquatic micro-organisms.

Most aquatic species require a certain temperature range to survive. If the Water Temperature exceeds these limits, the species will become stressed in order to survive and some may be killed. As the water temperature approaches the optimum for aquatic organisms, they become active, they consume more food and more oxygen. Thus, the Water Temperature has also an impact on the behavior and the growth of the organisms. Regarding the maximum permitted values of Water Temperature for aquatic life (Joint Ministerial Decision 46300 (1986)) for salmonids the temperature equals to 21,5 °C and for cyprinid equals to 28°C (Psilovikos, 2014).

A strong linear correlation is observed between water and atmospheric temperatures (Smith, 1981). Variations from this linear correlation often take place, as the water temperature rarely drops below 0 °C, despite the fact that extreme climatic conditions may occur. Variations also occur during spring, due to the melting of snow and can maintain water temperatures below the atmospheric temperature for weeks or even longer. Finally, rainfallmay lead to different water and atmospheric temperatures (Ward, 1985).

The main source of heat in the lakes is solar radiation, most of which is absorbed directly by the lake. In a shallow lake, the sediments can absorb significant amounts of solar radiation and this heat can, in part, be transferred to the waters (Wetzel, 2006).

5.2.2 Dissolved oxygen (DO – mg/L) and Dissolved Oxygen of Saturation (DOs – mg/L)

Dissolved Oxygen (DO) is essential for the metabolism of all aerobic organisms that live in the water. Aquatic organisms, do not use the oxygen of water molecules, as theycannot extract it in order to make it available. In fact, a small amount of oxygen, with a concentration of about 10 molecules of oxygen (O_2) per 106 molecules of water (H_2O), is dissolved in water. The aforementionedDissolved Oxygen is actually used by aquatic biota.

The concentration of Dissolved Oxygen in the water, reflects the balance between the processes that add oxygen, such as photosynthesis and reaeration and those that remove it, such as respiration from the biota. More specifically, Dissolved Oxygen sources include atmosphere, photosynthetic activity and hydro-mechanical oxygen distribution. The system's balance, is achieved by the metabolism of the organisms and by the abiotic chemical reactions (Wetzel 2006).

Temperature fluctuations, daily and seasonal changes in photosynthetic activity and re-aereation supply and deoxygenation process, make the Dissolved Oxygen concentration unstable.It is noted that deoxygenationis the reduction of the oxygen content, due to human activities a result ofanthropogenic emissions of carbon dioxide andeutrophicationdriven excess production. On the other hand, the low Water Temperature, favors the water body's ability to store larger amounts of Dissolved Oxygen in its volume. According to Chapra et al. (1997), the lower the Water Temperature is, the higher the Oxygen of Saturation (Psilovikos, 2014) (Equation 5.1):

$$\mathrm{DO_s} = 14{,}652 - 0{,}4102 * \mathrm{T} + 0{,}007991 * \mathrm{T}^2 - 0{,}000077794 * \mathrm{T}^3 \quad (5.1)$$

where:

DO_s: Oxygen saturation concentration (1 atm) (mg/L)

T: Water Temperature (°C)

Chemical oxidation of inorganic compounds, such as ammonia (by nitrification) and biochemical oxidation of organic compounds, are the main processes leading to a decrease in the concentration of Dissolved Oxygen. Large quantities of organic residues, such as manure, vegetable residues and high concentrations of ammonia and other inorganic compounds from fertilizers, can cause a rapid decrease of Dissolved Oxygen in water. The increase ofthe algae growth, due to the large amount of nutrients reaching the water, results in a high concentration of organic matter which, after decomposition, contributes to the reduction of Dissolved Oxygen (Wetzel 2006).

The range of Dissolved Oxygen concentration, is usually between 0 mg/L and 14 mg/L at a temperature at which water begins to turn into ice (about 0 °C). In literature, higher concentrations such as 16 mg/L or 18 mg/L are also recorded, especially in late spring, due to photosynthesis, which prevails the respiration processes (Matzafleri, 2007; Psilovikos, 2020). Despite the fact that at concentrations below 3 mg/L, fish deaths can occur, in Scandinavian lakes fish survive at lower dissolved oxygen values (Antonopoulos 2010; Psilovikos, 2014).

Salinity also effects the concentration of Dissolved Oxygen of Saturation, according toChapra et al. (1997) (Psilovikos, 2014) (Equation 5.2):

$$\mathrm{DO_{ss}} = [14{,}652 - 0{,}4102 * T + 0{,}007991 * T^2 - 0{,}000077794 * T^3]\left[1 - \frac{Chlor}{100000}\right] \quad (5.2)$$

where:

Tw: Water temperature (°C)

Chlor: Concentrationofchlorine anions (Salinity = 1,80655*Chlor) (ppm)

DO_{ss}: Dissolved Oxygen of Saturation (1 atm and S salinity) (mg/L)

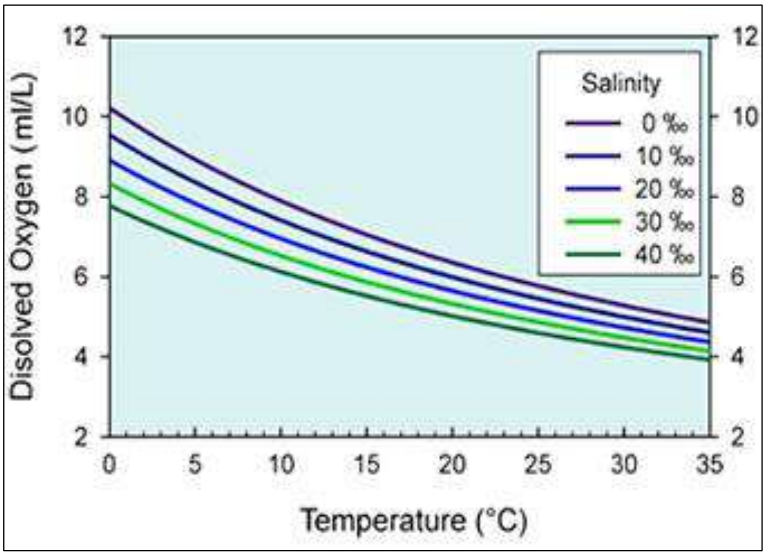

Figure 5.9: Dissolved Oxygen of Saturation – Temperature – Salinity (Available at: https://images.app.goo.gl/3szxMRZaThd7w9Kx8)

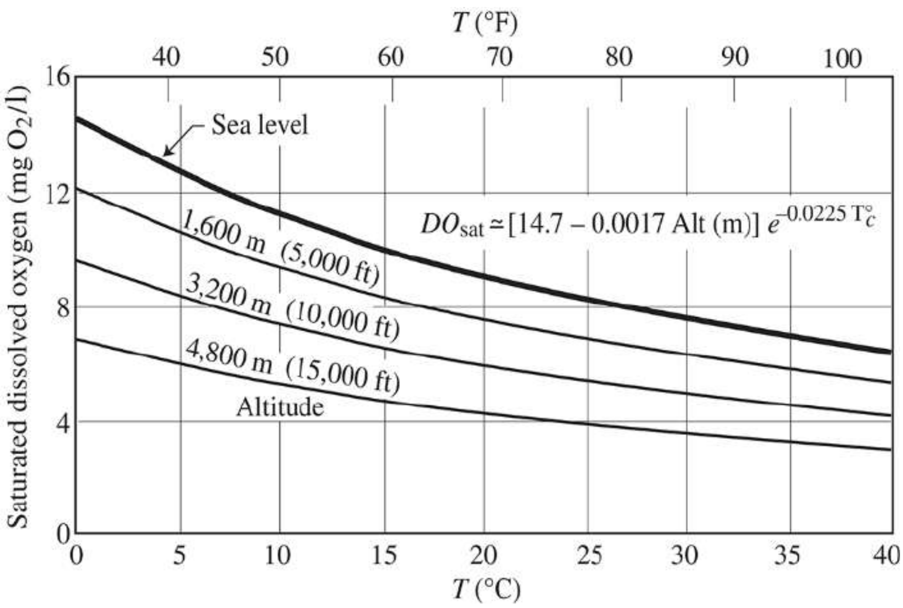

Figure 5.10: Dissolved Oxygen of Saturation – Temperature – Altitude (Julien, 2018)

Moreover, DOs, is affected by altitude according to Figure 5.10. According to JMD 46399 (1986) values between 80 and 120% of DOs are considered to be excellent, while below 65% and more than 125% are considered harmful to aquatic organisms. In addition, according to WHO (1996), values above 75% are considered acceptable for domestic use (Psilovikos, 2014; Sentas et al., 2019).

5.2.3Nitrogen forms and sources

Nitrogen can reach water bodies through: a) atmospheric deposition on the catchment or directly on the water body, b) leaching from diffuse sources within the catchment (fertilizers and manure), c) sediment erosion of rich soils in Nitrogen and surface applications of manure in catchmentsand d) direct input from point sources.

A variety of N forms are commonly found in all aquatic ecosystems such as inorganic N forms (NO^{2-}, NO^{3-}, NH^{4+}, NH_3) and organic N fractions (Dissolved Organic Nitrogen – DON, PON, etc). Nitrite, Nitrate, Ammonium and DON are directly available for plant uptake, whileN_2O and NO (gaseous forms of N) are exchanged with the atmosphere. Nitrate dominates in highly enriched rivers while DON dominates in less enriched rivers.

Figure 5.11, shows nitrogen species concentrations in different European streams, ranging from oligotrophic to hypertrophic. Durand et al., (2011) concluded thatboth nitrate and DON concentrations increase in absolute terms. Moreover, nitrate increasesas a proportion of total nitrogenwhile DON decreases as a proportion of total nitrogen (Figure 5.11).

Nitrogen is needed, to support primary production by aquatic flora and algae. Nitrogen also plays a role in determining the food web structure and relative productivity of any aquaticsystem through microbial, algal and plant uptake of N in the form of both inorganic N species and DON (Durand et al., 2011).

In aquatic systems, NO_3–N is reduced due to its absorption from algae and plants and is increased due to nitrification processes. The rate of reduction of ammonia in aquatic systems is a combination of reducing ammonium nitrate and escaping to the atmosphere. The second of these processes is considered negligible, compared to the amounts oxidized to nitrogen and nitrate nitrogen. In general, the ammonium nitrogen (NH_4-N) rate of concentration change due to: a) mineralization of organic nitrogen, b) nitrification (NH_4-N is converted to NO_2-N), c) NH_4-N assimilation from algae and plants and d) its escape into the atmosphere (Psilovikos, 2014).

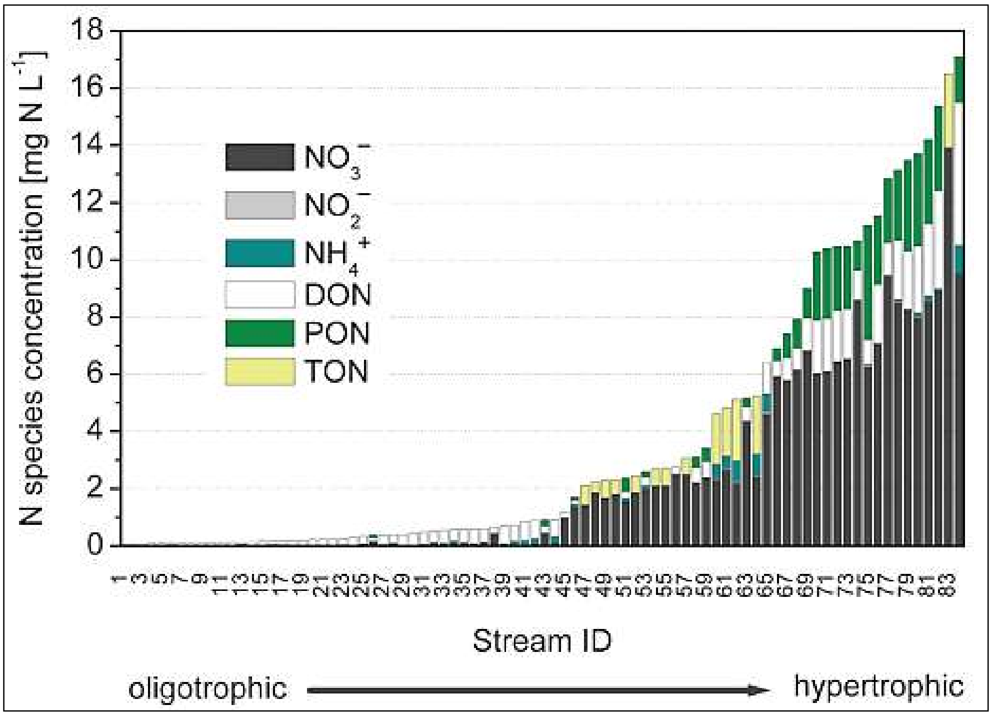

Figure 5.11: Concentrations of N species in 57 European streams and rivers (Durand et al., 2011)

It should be mentioned here that, according to WHO (1996), a maximum concentration of 45,7 mg/L NO^{3-} or 10 mg/L NO_3-N is considered as the maximum concentration for drinking water. In Europe, the maximum values are 50 or 11,3 mg/L respectively. These concentrations are in accordance with JMC 2600 (2001) and JMC 46399 (1986). The last decision mentions the same maximum values for living aquatic organisms. According to Antonopoulos (2010), NO_3-N values exceeding 0,5 mg/L show eutrophic conditions, while values above 1 mg/L show anthropogenic pollution interventions from urban and agricultural sources (Psilovikos, 2014).

5.2.4Electrical Conductivity of Water (ECw – μS/cm)

Electrical ConductivityofWater, is its ability to conduct electricity, a property associated with the concentration of ionized substances, their mobility and Water Temperature at the time of its measurement. Such substances are Fluoride, Chlorine, Sodium, Calcium and Magnesium salts. A sharp increase in ECw, is a sign of pollution. The higher the conductivity in freshwater, the greater the possible sign in organic productivity.

ECw for irrigation water, greater than 700μS/cm, can cause soil salinity problems, depending on the type of soil, the type of crop, the climate and the way the water is applied. ECwgreater than 3000 μS/cm, is generally unsuitable for irrigation. According to JMD 46399 (1986), the acceptable values for both drinking water and living aquatic organisms are less than 1000 μS/cm. Moreover, according to JMC 2600 (2001) and WHO (1996), the maximum permitted values for drinking water are 2500 and 400 μS/cm, respectively (Psilovikos, 2014; Sentas et al., 2019).

5.2.5Turbidity (Turb – NTU)

Turbidityis considered as an indicator of the quality of water. It is a measure of the degree to which the water loses its transparency, due to the presence of suspended solids. The more Total Suspended Solids (TSS – mg/L) in the water, the higher theTurbidity is. It is well known that DO dissolves better in water, as the Tw decreases. The TSS absorb heat from the sunlight, thus making thewaterwarmerand so, reducing the DO concentration in the water, causing stress and possibledeath to aquatic organisms. The suspended solids have a large specific surface area, due to their small size, which often absorbs toxic substances. In accordance with Joint Ministerial Decision46399 (1986), the desired values of TSS for aquatic organisms are 25 mg/L while no maximum values are provided (Psilovikos, 2014).

The TSS scatter the light, thus decreasing the photosynthetic activity of plants and algae, which contributes to lowering more DO concentration. In lakes, the turbiditynowadays is measured with electronic probes in Nefelometric Turbidity Units (NTU). A traditional, cheap and very effective way, in case we don't have something else available, is indirectly estimation with the Secchi Disk. This is a black and white disk attached to a rope, which measures the depth that the disk reaches, before it disappears from sight (Johnson et al., 2007). There is an empirical formula, combining the Color, the Secchi Disk Depth and the Nefelometric Turbidity Units (Psilovikos, 2014) (Equation 5.3):

$$\frac{1}{SD} = 0{,}003\ C_{OL} + 0{,}152\ T_{UR} \qquad (5.3)$$

where:

S_D: Secchi Disk Depth (m)

C_{OL}: Color (mg/l)

T_{UR}: Nefelometric Turbidity Units (NTU).

5.2.6 Active acidity (pH)

The active acidity (pH) of water, is a crucial factor in the functioning of aquatic ecosystems, and an important tool for assessing other water quality parameters, such as alkalinity (Stednick, 1991). pH is used to determine the acidity or alkalinity of water and its units are measured on a log scale. Its values range from 0 to 14, with 7 being considered the level of neutrality. Values below 7 characterize water as acidic and above 7 as alkalic. Active acidity is a measure of the free H^+ anions and OH^- hydroxyl cations of water. If the concentration of the former is greater than that of the latter, then the water is characterized as acidic. As the active acidity is influenced by chemicals added to the water is an important and essential indicator of water quality. Acidification affects aquatic systems at all levels and has obvious effects on plant and animal communities. Aquatic organisms are affected both directly by toxic conditions and indirectly by the reduction of acid-sensitive species that consist food for other species. This degrades the biodiversity of aquatic systems. As depth increases, pH values decrease due to the increase of CO_2 that dissolves in water. pH is also related to photosynthesis, as it increases phytoplankton take up carbonate ions and so pH increases. In other words, eutrophic conditions with high algal biomass have high pH.

According to JMD 46399 (1986), JMD 2600 (2001) and WHO (1996),t he acceptable values for drinking water range from 6,5 to 8,5. Moreover, according to JMC 46399 (1986) the acceptable values for aquatic organisms range from 5,5 to 9. In waterswith a pH below 5, generally not many fish species are observed. ApH value less than 4, causes the death of all vertebrates, some invertebrates, and many microorganisms. In terms of chemical erosion, generally low pH water (<7) causes erosion of many materials. Some erosion products are toxic to aquatic organisms. In the case of flora, pH value less than 4,5, may harm plant organisms either directly or indirectly.

5.2.7 Chlorophyll-a (Chl-a – μg/L)

Chlorophyll-a (Chl-a) is the major photosynthetic phytoplankton pigment and a trophic indicator in aquatic ecosystems. Chlorophyll-a is often used as an estimate of algal biomass, with blooms being estimated, when Chl-a concentrations exceeds 40 μg/L. Because eutrophication is defined as an aquatic ecosystems response to nutrient loading, the ability to identify important factors and predict subsequent algal blooms with the use of a Chl-a equation could be a key lake water management tool. Biological, chemical and physical controlscan be used to prevent or remove algae from the water (Balali et al.,2013).

Biological controls involve introducing species that prey on algae. Because there is no practical way to increase the populations of organisms that feed on planktonic algae, biological controls only work on filamentous algae and macroalgae.

Physical control such as raking or seining is another option for filamentous or macroalgae, but it is not a viable approach for planktonic algae due to their microscopic nature. Physical management uses a rake, seine, wire screen, or other similar device, to pull or cut the algae and remove it from water. Methods like this are generally short-lived as the algae will recolonize very quickly afterward. Another method of physical control involves non-toxic dyes and colorants to shade the water and to limit the sunlight penetration on which algal growth is dependent. However, this kind of control method may suppress the natural food chain of the pond by reducing the planktonic algae that acts as the basis of the ecosystem. Shading the bottom of the pond can also be achieved by using physical barriers such as shade mats to prevent sunlight penetration (Sink et al., 2014).

The most effective and commonly utilized chemical controls for algae are algaecides containing copper such as copper sulfate or chelated copper complexes. Copper sulfate and chelated copper complexes have demonstrated excellent control (90% or greater) of all three types of algae when applied correctly at labelled rates. Copper sulfate and chelated copper complexes are contact herbicides that act as cell toxicants, and only kill algal cells that the copper contacts directly. Therefore, they must be sprayed or broadcasted over the entire area where algae are growing to provide adequate control. In most cases, an aqueous form of copper is recommended because it is less complicated and easier to apply correctly. Effects of copper on algae can be observed in 3 to 10 days with the full effects of the treatment demonstrated in 4 to 6 weeks (Sink et al., 2014).

An empirical treatment in physical control of algal removal, took place in Kastoria Lake, with the rational use of the Gkiole's Stream weir, so as to relief the Lake from algae, by discharging amounts of water to Gkiole's Stream, as explained above.

5.3 Spatial Simulation- Inverse Distance Weighting method (IDW)

Geographic Information Systems (GIS) is a very powerful tool in creating thematic maps with any geographical information that can be quantified. So, if water parameter measurements are available from different measuring locations, like in the present PhD, with the help of Monitoring, thematic color gradient maps can be constructed that allow us to (Psilovikos, 2020):

a) Geographically simulate the variability of parameters on the surface of a lake.

b) Understand the correlation of this particular parameter with others in different locations

c) Assess the environmental status of the lake

d) Propose preventive and remedial measures.

Depending on the method used, the geographical distribution of the parameters and the resulting thematic maps, may not correspond to reality. One should be very careful,selecting the appropriate algorithm. In order to calculate the water quality parameters asthere are no measurements on the entire surface of Lake Kastoria, a Deterministic Spatial Simulation Algorithm of its water quality parameters is applied to the data of the four telemetric stations. With the help of the Spatial Analyst Tool (Interpolation) of ArcGis software 10.2, two algorithms have been used: a) the "RBF – Radial Basis Function" and the b) the "IDW – Inverse Distance Weighting" in previous research for Lake Kastoria, showing that the first one could not simulate well the spatial distribution of the contour curves, when the latter one, achieved it successfully (Matzafleri et al., 2009; Mantzafleri, unpublished material). Just the opposite, occurred in the case of Lake Karla (Zisou & Psilovikos, 2012), where the better algorithm was the RBF.

Radial Basis Function—RBF (splines): this is an algorithm, by which a surface with a minimum curvature is interpolated, by the sampling–interpolation points. This is achieved by fitting a number of low order local polynomial interpolators, to a specific number of neighbor interpolation points, which pass through the sampling points. This method provides better results for the surfaces with mild gradients of the value they refer to, such as water level or pollution concentrations. It is not an appropriate method if there are significant variances in the gradient of these values, because it cannot elevate successfully the calculated values. The governing equation is given by the formula (Bishop 1995; Matzafleri et al., 2009)(Equation 5.4):

$$z(s_0) = \sum_{i=1}^{n} \omega_i \varphi(||s_i - s_{j0}||) + \omega_{n+1} \qquad (5.4)$$

where:

$z(s_0)$ is the unknown value of the water quality parameter in prediction location s0.

φ(r) is a radial basis function.

$r = ||s_i - s_0||$ is the Euclidean distance between the prediction location s0 and each data location si and

ω_i (i = 1, 2,..., n + 1) are weights to be estimated.

The IDW is an accurate interpolation method, where the maximum and minimum values on the simulated surface, do not exceed those of the field values, while it is possible to map the variation either locally, or throughout the area that the sampling stations cover. It is a method where, it is able to approximate the magnitude of the variation better than its structure in space (Lietal., 2008).

The IDW method gives a weight (degree of significance) to the points for which there are measurements, in order to obtain prediction values for the rest of the study area. However, weights depend not only on the distance between the points for which there are measurements and the forecasting points, but also on the overall spatial distribution among the available measurements. The weights decrease as the distance increases, so the neighboring known points gain more weight and have a greater influence on estimation.

In general, the IDW method has the following advantages:

- ✓ It interpolates values even with three sampling points that are the minimum points that are required for this method
- ✓ It does not give values outside of the range ''minimum – maximum'' values, but always within them and produces "bull eyes" around data locations.
- ✓ It is the most suitable in the formulation of the interpolation isodynamic contours
- ✓ It results in smooth and continuous surface changes between observations.

The formula of IDW method is finally obtained as follows (Dale and Fortin, 2014; Psilovikos, 2020) (Equation 5.5):

$$\hat{z}(x_0) = \frac{\sum_{j=1}^{m} z(x_j) {d_{ij}}^{-k}}{\sum_{j=1}^{m} {d_{ij}}^{-k}} \qquad (5.5)$$

where:

k: the distance influence coefficient, which is usually 1 or 2.

d_{ij}: the distances between the unsampling location i (x_0) and the sampling locations j (x_j).

According to Charizopoulos et al. (2018), the power of*k*can be optimized. The optimization process calculates several models and chooses the power value that leads to the model with the minimum Root Mean Square Prediction Error (RMSPE). IDW assumes that the simulated surface is being driven by the local variation which can be captured through the neighborhood. Changing neighborhood options (shape, sectors and the number of neighbors) may lead to a better model. Further research, is proposed on the optimization process of IDW method.

Moreover, the reason why it is preferred to use deterministic method, instead of geostatistic ones, is that it can lead to successful results with very few monitoring points (Matzafleri et al., 2009; Zisou & Psilovikos, 2012). It should be noted that the simulation of the eight (8) parameters across the surface of Lake Kastoria is carried out for fixed periods of time, namely: a) January 2016-2018, b) May 2017-2018 and c) September 2017-2018. The significance of this particular choice of months, lies in the fact that for those time periods of interest parameters show their representative values close to their range. These representative values are undoubtedly indicators of the status and viability of an ecosystem, so it is very important to analyze them during theaforementioned months they occur and evolve. The periods, are according to previous studies in Lake Kastoria, occurred by Matzafleri (2007), Matzafleri et al., (2009) and are about to validate their results ten years later. For the period 2005-2015, useful results are about to be published, based on the traditional sampling procedure that took place in Lake Kastoria, on a monthly basis (unpublished material by Matzafleri).

The North, East and South partsof the catchment of Lake Kastoria, are characterized by extensive arable landand, in contrast to itsWest part, it is occupied by the city of Kastoria, forest and semi-natural areas. The Gkiole Stream, constitutes the only surface discharge of Lake Kastoria’s basin, as well as the connection between the Lake and the Aliakmonas River. The Stavros and Psaradika stations, are located South and North of the coastline respectively, that are the most significant spots of the urban part of the Lake Kastoria. They constitute stations of significant importance, for measuring the ecological sensitivity of the aquatic ecosystem by the pressures exerted by the city itself. Finally, the Toichiostation is located in theNorthern part of Lake Kastoria. The areas at the lowest part of the Torrent Toichio are intended for agricultural use. Concerning the North-Western side of the lake, the waters are characterized by a stationary water status.

In an aquatic ecosystem, such as a lake ecosystem, in January, the lowest Tw values are presented; large quantities of precipitation are detected, strong winds occur, while the concentration of nutrients are poor. Under prevailing environmental conditions, therefore, one expects high levels of Dissolved Oxygen inside the lake, mainly in surface water, although photosynthesis activity is poor.

In May, the temperature rises, the regeneration of aquatic and riparian vegetation take place, photosynthesis activity show its peak, as abundance of photosynthetic organisms and of Dissolved Oxygen in the waters are at seasonally satisfactory levels. In addition to surface water, excess nutrients are present, while concentrations of Nitrate, Nitrite, Ammonium and Phosphate, which play a key role in the lake's nutritional status, are increasingly detected. Generally, in the given period of time, the so-called “good” side of the eutrophication is presented, where the DO balance is positive and the re-aeration level is greater than the respiration one.

On the contrary, in September the conditions of the “bad” side of eutrophication are prevailed. Aquatic and riparian flora start to decompose, aerobic microorganisms and especially bacteria, which consume excessive amounts of DO, through respiration processes, are rapidly increased. Consequently, the reduction in DO is likely to cause stress to fish populations, even sometimes leading to fish deaths. Moreover, nutrients that flow into the lake are increased dramatically, causing hypoxic conditions, especially in shallow lakes and lakes in both urban and agricultural catchments, such as Lake Kastoria and Pamvotida (Ioannina). Subsequently, anaerobic bacteria create gaseous decomposition products, such as the toxic Hydrogen Sulfide (H_2S) and Ammonia in both ionized (NH_4^+) and non-ionized (NH_3) forms or in the form of Nitrogen Ammonia ($N - NH_4^+$).

Considering on the other water quality parameters below, which assessment is going to follow, it is a little bit risky to try to interpret, because we have monitoring data only from the abiotic parameters and the ecosystem of Lake Kastoria is so quite complicated, that it is unsafe to assess successfully. So, we try to explain, based only to the abiotic monitoring data, by quoting the results of the research. Another point is that, we take for granted that the electronic sensors of the monitoring probes, are frequently maintained, calibrated and replaced when a damage occurs. There is much uncertainty of the assessment of these parameters and only for the parameters of Tw, DO, Chl-a and Turb, we try to obtain a satisfactory interpretation, considering that Tw is a very important input independed parameter and the three latter, are the output parameters, obtained by the structure of the Deep Learning Machine Models that follow. Further Research could give us the opportunity to evolute this study with Carlson’s Indexes for Total Phosphorus, Chl-a and Secchi Disk, N/P ratio, phycocyanin etc, which are parameters that are not monitored yet, in a continuous base. Finally, the design of the contour deterministic maps contains uncertainty too, because from a number of four point values, we spatially integrate and assume that the curves which prevail, gives us representative values of the reality. Unfortunately, we cannot use geostatistical methods, because of the low number of monitoring points, but we do our best, using the most appropriate deterministic method, which is the IDW.

Regarding the concentration of Chlorophyll-a (Figure 5.12), it is noted that the highest levels of Chlorophyll-a are detected peripherally of Gkiole’s telemetric station, in January, while for the rest telemetric stations its concentration is low showing the minimum value (4,42 μg/L) near Stavros station. In May, the lowest values of Chlorophyll-a concentration are restricted to the perimeter of the Psaradika station (NorthWest), while its highest values are detected in the northeastern and southern parts of the lake. In September, the highest chlorophyll-a values continue to appear in the Northeastern part of the lake. On the contrary, Stavros station (SouthWest), show its minimum value.

In all aquatic systems, there is a strong annual cycle at Water Temperature. Regarding the correlation of Tw with the surface of Lake Kastoria and the season, it is observed that the lowest Tw values are observed in the Northern and Western part of Lake Kastoria during all the reference months, as expected. In contrast, the highest Tw values, appear in the Southeastern part of the lake. The lowest Tw is in January, while the highest is in September (Figure 5.13), considering the periods – months that we take into account.

Regarding the parameter of Dissolved Oxygen, in January, the highest values are observed in the Northeastern and Southwestern part of the lake, where this situation is further established in the Southwestern part, in May. This is explained mainly bythe torrents that flow into the lake from the north and east enriching the lake with water during the winter and spring. In September, the concentration of Dissolved Oxygen isdecreased in the Southeastern part of the lake, due to the opening of Gkiole's weir. According to Figure 5.14, DO values remain constant throughout the year. However, in the past, based on the literature (Matzafleri et al., 2009; Psilovikos, 2020), low values of Dissolved Oxygen wereobserved in January, due to the sharp drop of Tw in the Lake's surface where it got frozen for a couple of weeks, making the enrichment of the water with DO, difficult.

Regarding pH, the highest pH values are found around Gkiolestation, while the lowest pH value is observed at Stavrosstation. These differences are not significantbut they are an indicator showing that the lowest pH values occur near the urban side of Lake Kastoria. Figure 5.15, shows that pH values are constant for the three representative months.

ConcerningElectrical Conductivity of Water (ECw), it is observed that, in January and May, its range is small, indicating that there are no significant fluctuations in ECw values during this month. In contrast to the month of September, a wider range of ECw values is observed (around 60 μS/cm), which is not enough though to proceed to its assessment. Accordingto Figure 5.16, the minimum ECw values are detected in the Southern part of the lake and in particular around Gkiole station. The maximum values are found to the Southwest part at "Stavros" station. In the reference months, January and September, the Northern part of the lake shows a moderate and low profile, in contrast to May, where ECwis higher.

In May, the turbidity range is very high (exceeding 300 NTUs), indicating significant fluctuations. The highest turbidity value in January, May and September reaches 128,9; 309,21 and 55,7 NTUs, respectively (Figure 5.17). The highest turbidity values, for the three months of reference, appear in the Western part of the lake, mainly due to the urban load. In contrast, the lowest values of turbidity, for the three months of reference, occur in the Northeastern, Eastern and Southeastern parts of the lake, due to the presence of torrents and Gkiole’s weir. It should be mentioned here that rainfall events, can increase erosion and result in greater amounts of sediment washing into Lake Kastoria, and finally cause increased Turbidity values. Moreover, the increased Turbidity values appear in the city's Northern part, due to the presence of a small jetty that collects phytoplankton.

Bacteria play an important role in the Nitrogen Cycle. Apart from the fact that they oxidize Ammonium to Nitrite and Nitrate, they also have the ability to bind atmospheric nitrogen and reduce nitrate and nitrite to ammonium or even nitrogen. Also, the presence of an agricultural and/or urban area near the lake enhances the presence of Nitrate, Nitrite and Ammonium ions. From Figure 5.18, it is concluded that the minimum values of Nitrate are detected in the Southern part of the lake, and in particular at Gkiolestation, in January and May. The maximum values are found to the North and Southwest side, in May and September.The maximum Nitrate concentration occurs in September, with a value of 1,72mg/L on the Southwest side of the lake, while the minimum one appears in May, with a value of 0,78mg /L on the South side of the lake and around Gkiole’s station. It is noted that in September, in the interior of the lake where thelowest concentration of Chlorophyll-a is detected, the concentrations of nitrate are enhanced and vice versa.

Concerning Ammonium concentration, its range is appeared small (0,5 mg/L), indicating that there were no significant fluctuations during May. On the other hand, in January, the concentration of Ammonium is 2,6 mg/L.From Figure 5.19, it is observed that the concentrations of ammonium in May and September show similar patterns. In January, the lowest concentration of ammonium is found in the southern part of the lake and in particular around Gkiole station, while the highest concentration of ammonium is detectedin the western part of the lake.

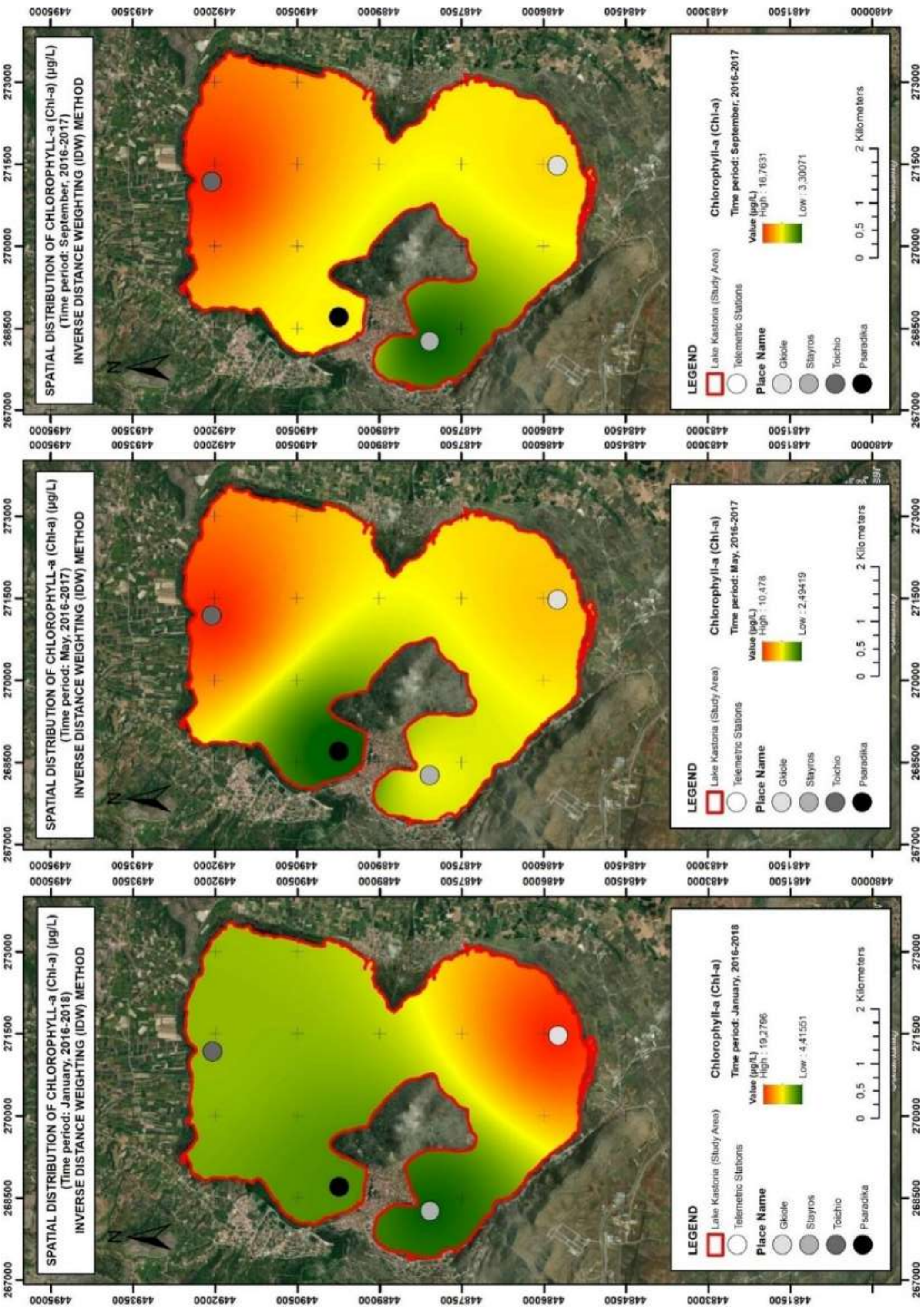

Figure 5.12: Spatial distribution of Chl-a. Time period: January 2016-2018 (left), May 2016-2017 (middle) and September 2016-2017 (right)

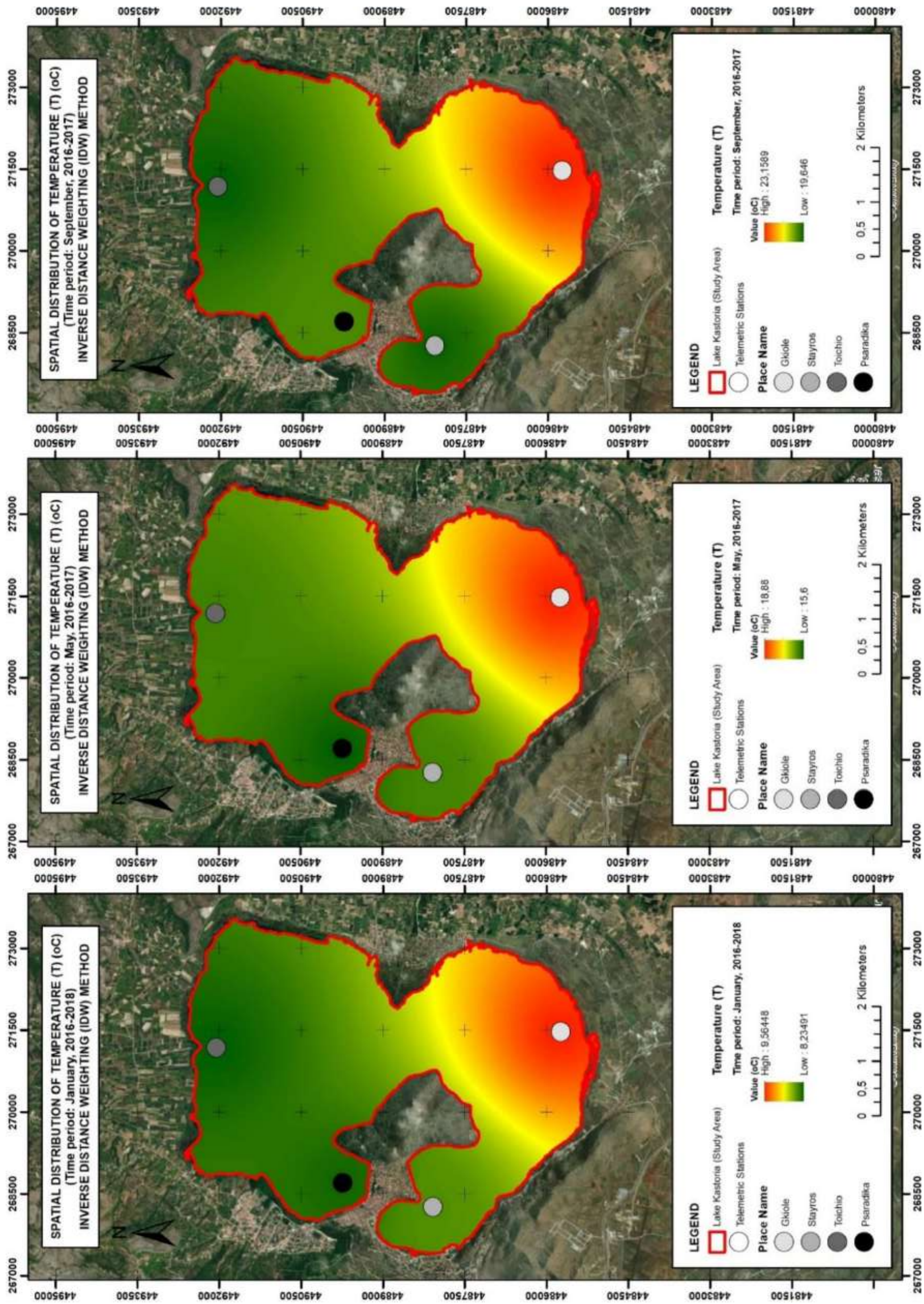

Figure 5.13: Spatial distribution of Temperature. Time period: January 2016-2018 (left), May (middle) and September (right)

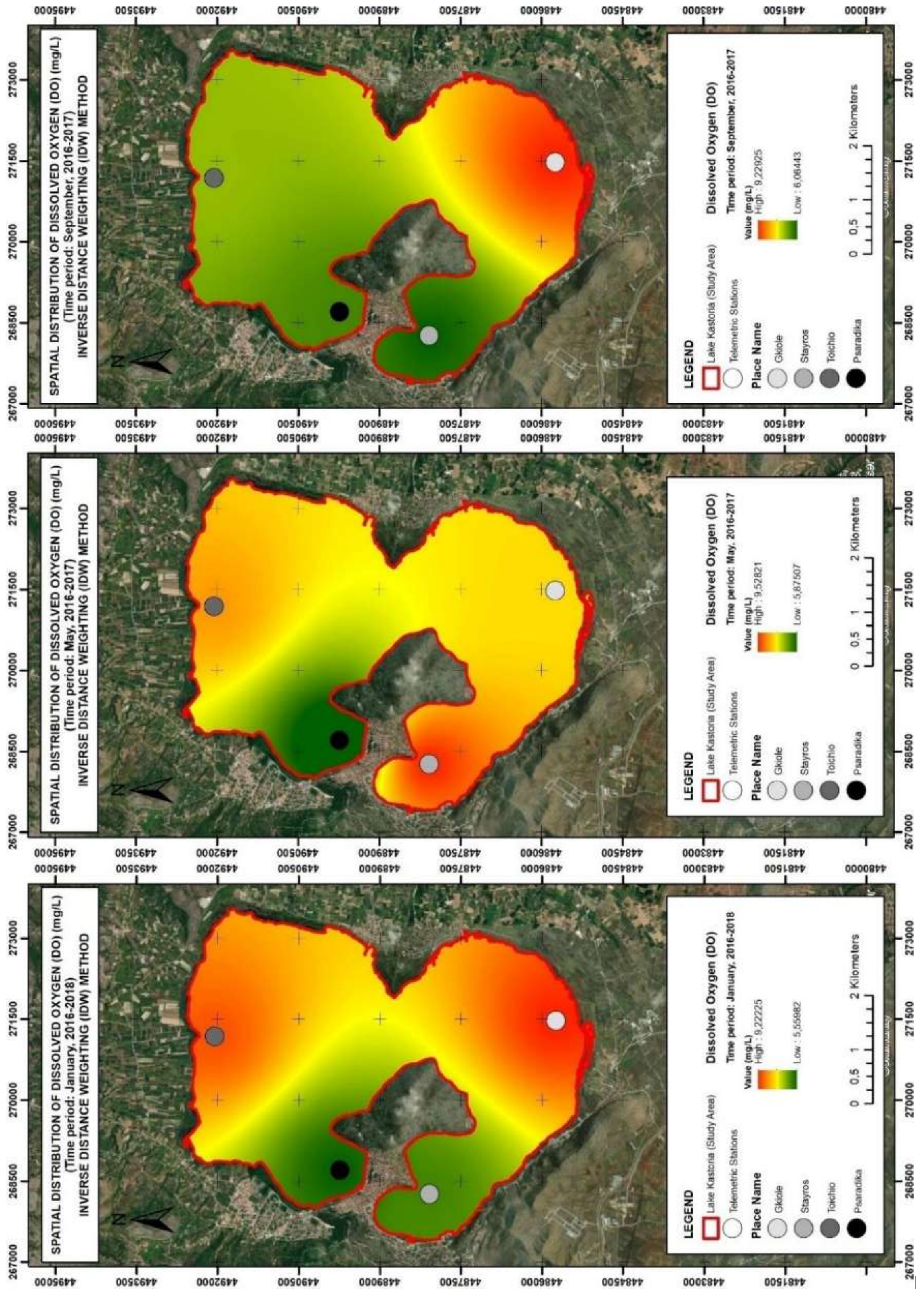

Figure 5.14: Spatial distribution of Dissolved Oxygen. Timeperiod: January 2016-2018 (left), May 2016-2017 (middle) andSeptember 2016-2017 (right)

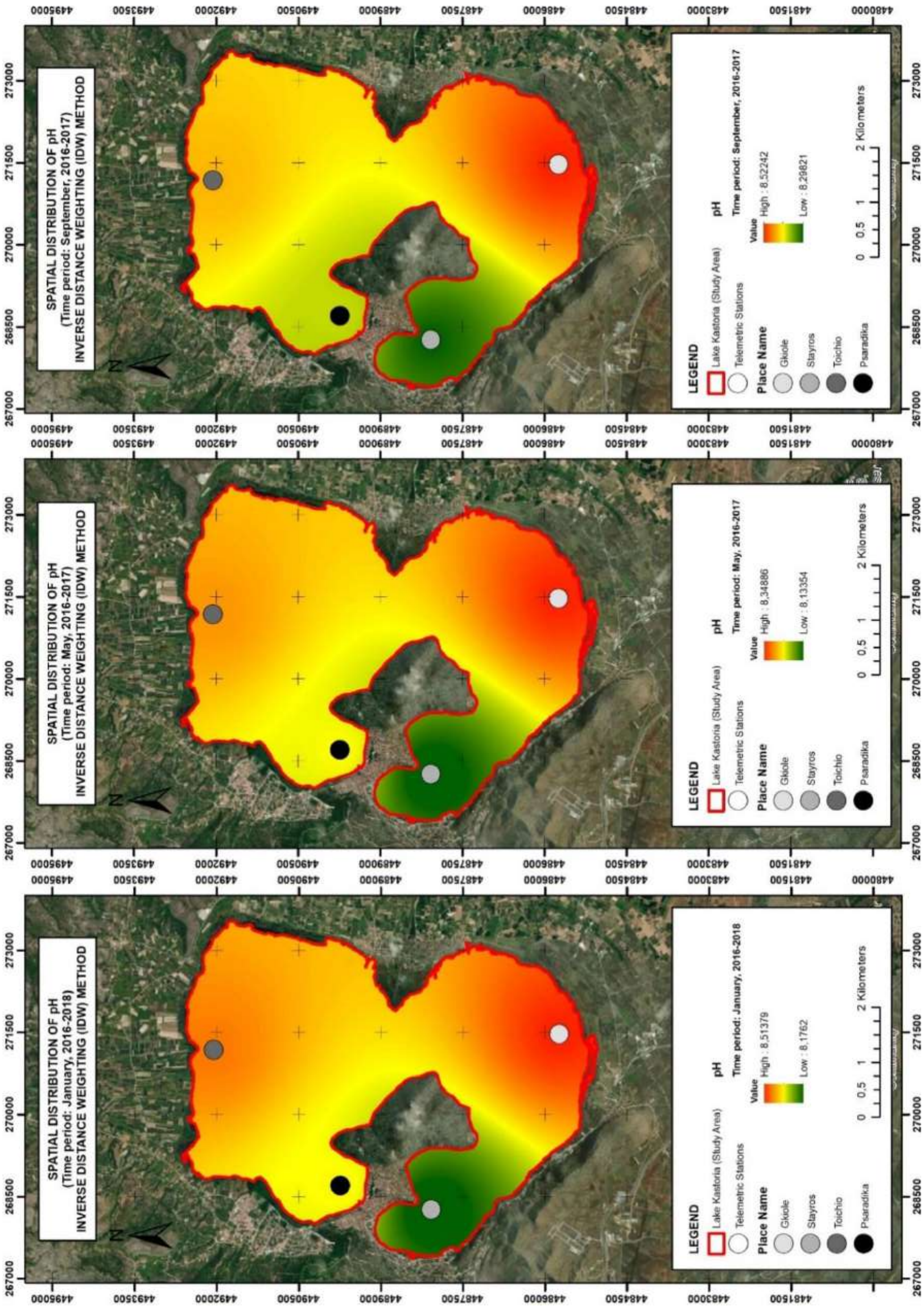

Figure 5.15: Spatial distribution of pH. Time period: January 2016-2018 (left) May 2016-2017 (middle) and September 2016-2017 (right)

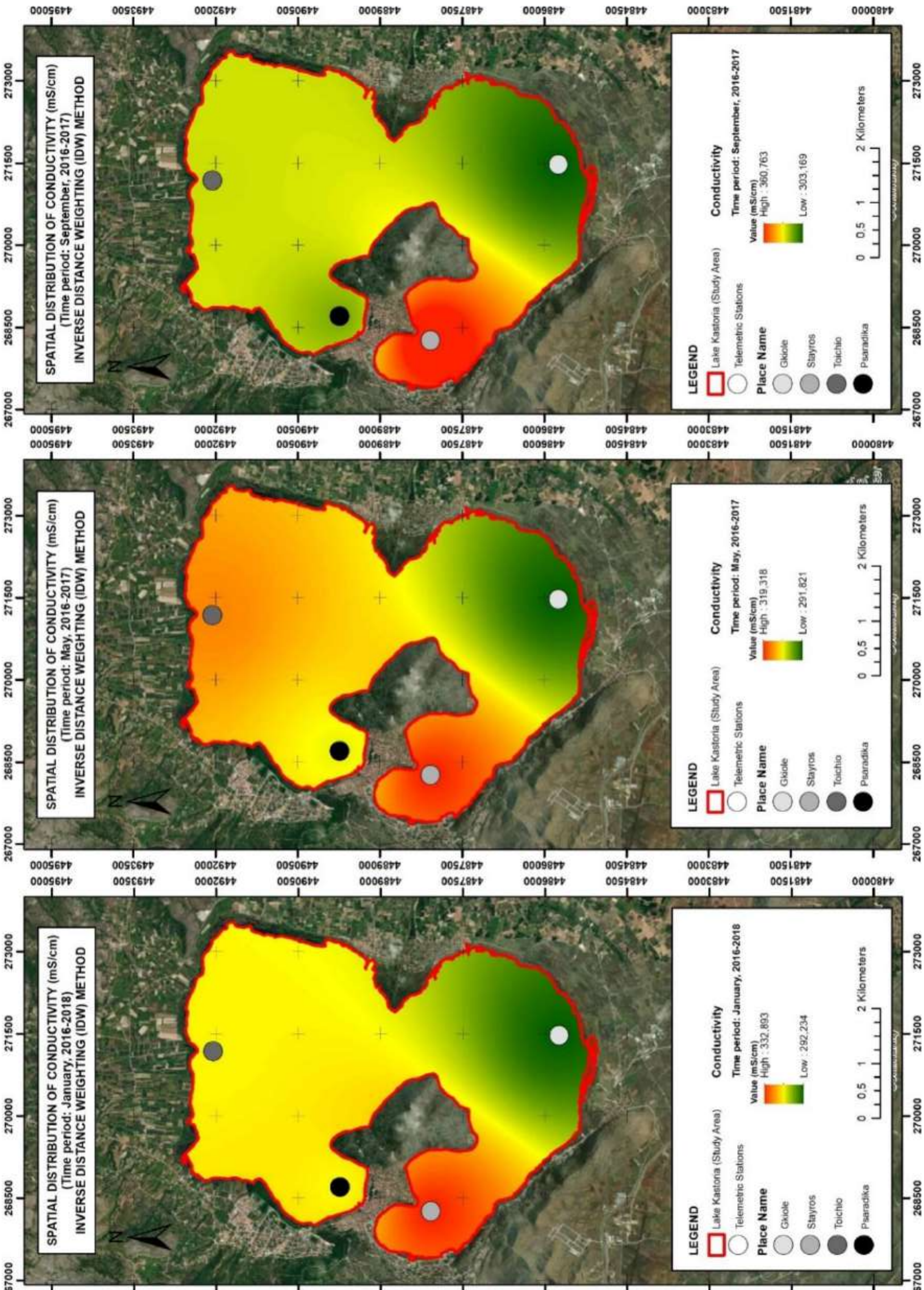

Figure 5.16: Spatial distribution of Conductivity. Time period: January 2016-2018 (left), May 2016-2017 (middle) and September 2016-2017 (right)

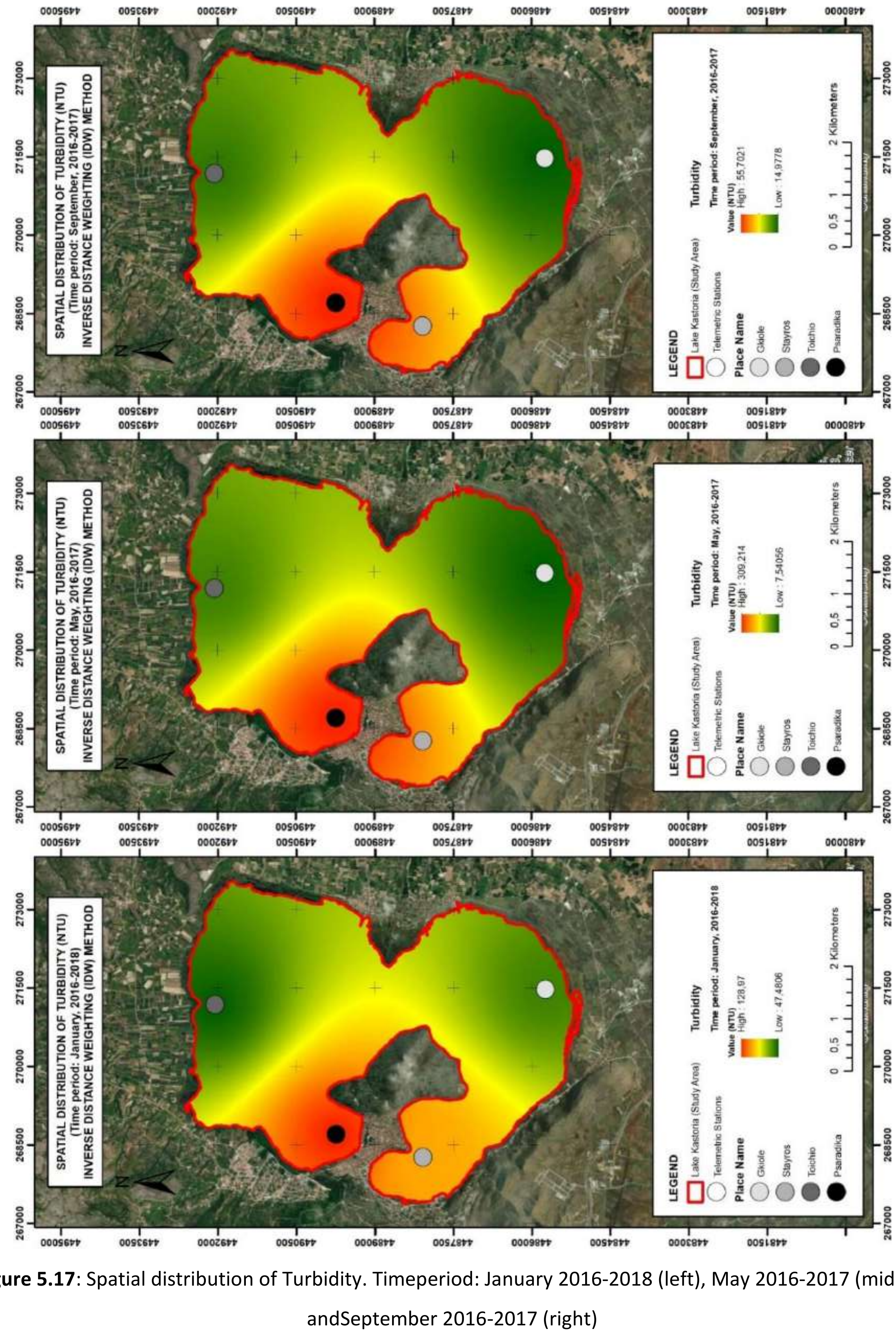

Figure 5.17: Spatial distribution of Turbidity. Timeperiod: January 2016-2018 (left), May 2016-2017 (middle) andSeptember 2016-2017 (right)

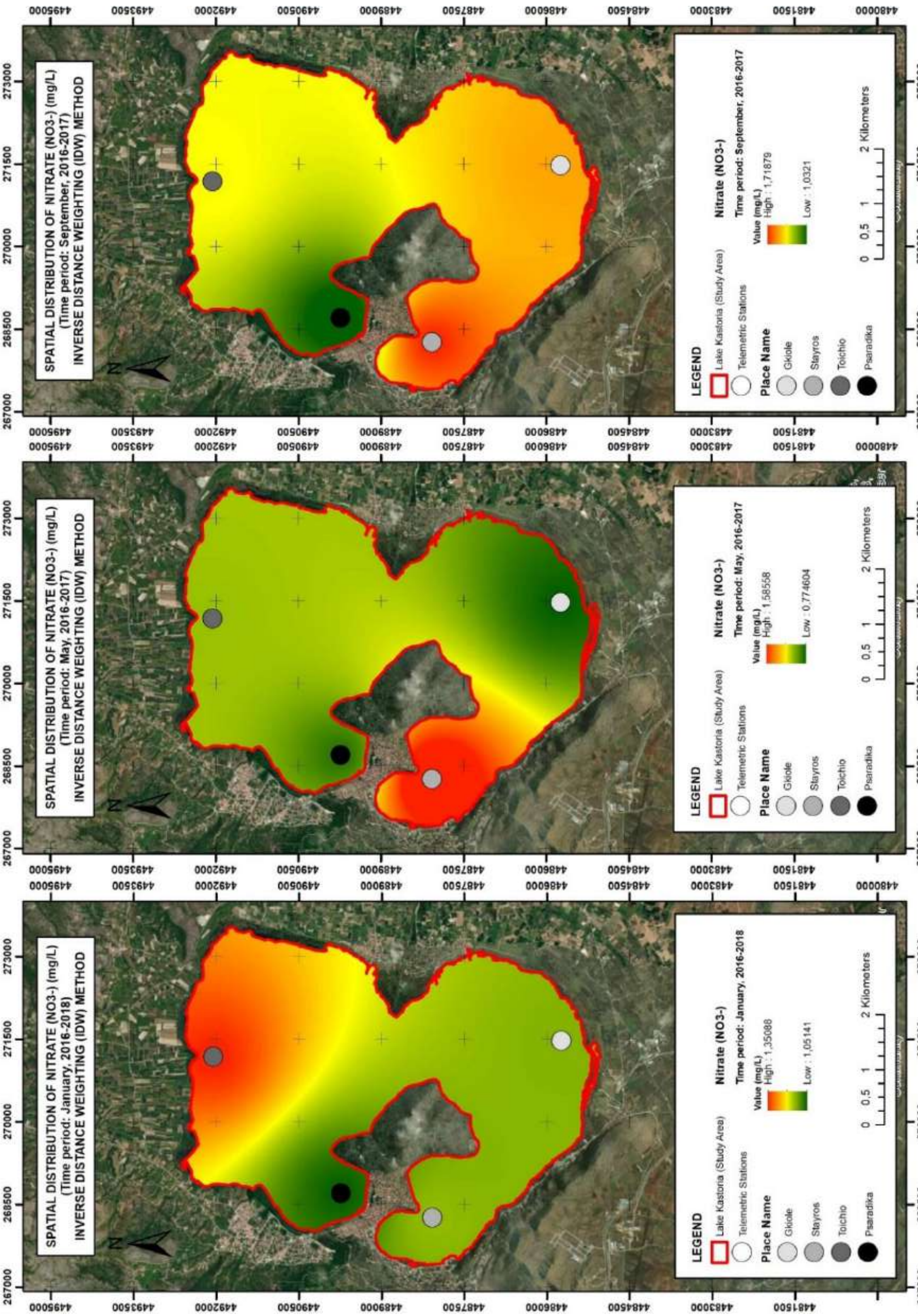

Figure 5.18: Spatial distribution of Nitrate. Time period: January 2016-2018 (left), May 2016-2017 (middle) and September 2016-2017 (right)

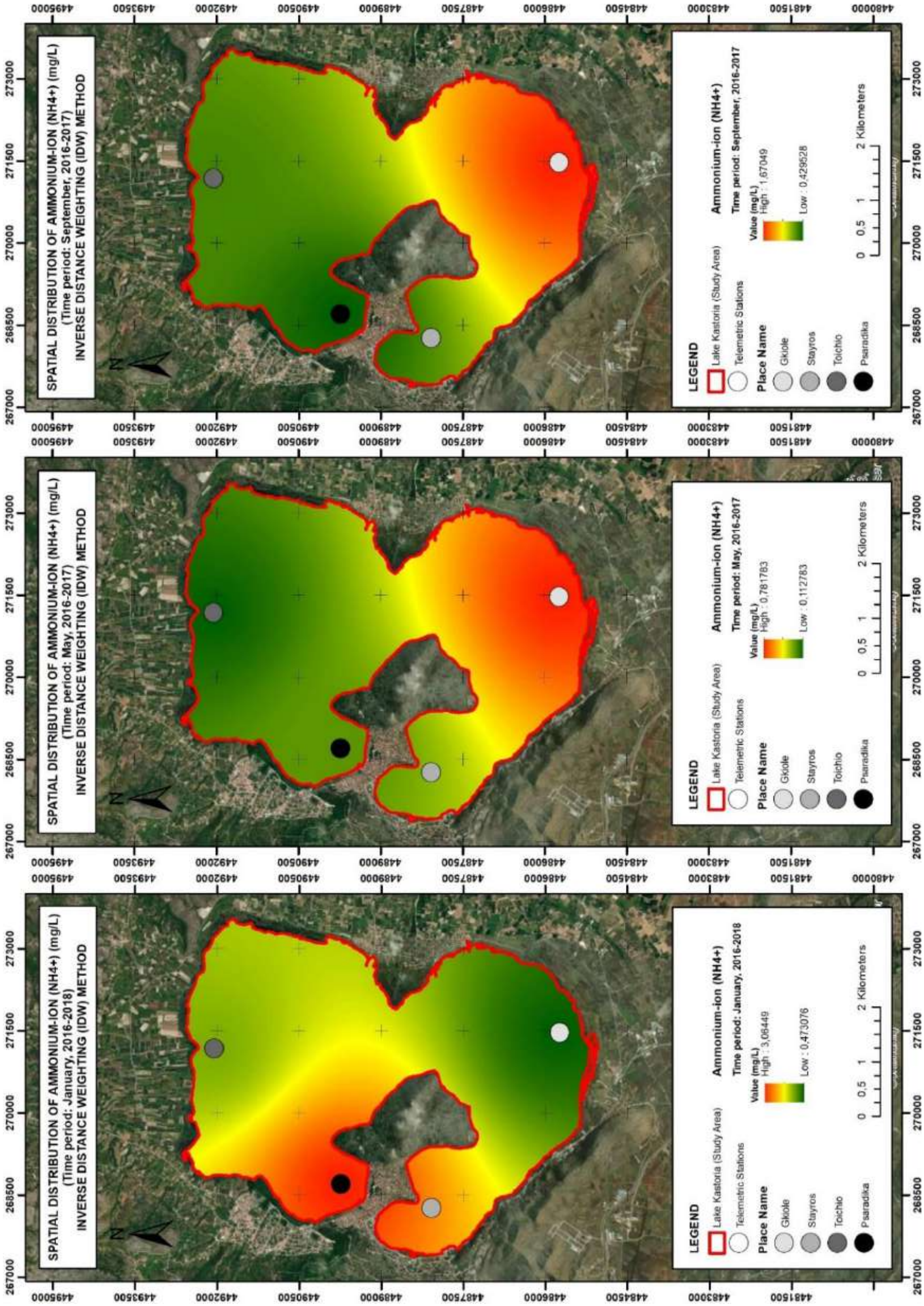

Figure 5.19: Spatial distribution of Ammonium. Time period: January 2016-2018 (left), May 2016-2017 (middle) and September 2016-2017 (right)

References

Antonopoulos, V., 2010. Environmental Hydraulics and Surface Water Quality, Tziolas Pub, Thessaloniki, Greece, 473 pages (in Greek).

Balali, S., Hoseini, A., Ghorbnia, R., Kordi, H. and Khozani, E.A., 2013. Relationships between nutrients and chlorophyll a concentration in the international Alma Gol Wetland, Iran. *International Journal of Aquatic Biology*, 1(2), pp.68-75.

Bishop, C. M., 1995. *Neural networks for pattern recognition*. Oxford university press.

Chapra, S. C., Canale, R. P., & Amy, G. L., 1997. Empirical models for disinfection by-products in lakes and reservoirs. Journal of Environmental Engineering, *123*(7), pp. 714 - 715.

Charizopoulos, N., Zagana, E. and Psilovikos, A., 2018. Assessment of natural and anthropogenic impacts in groundwater, utilizing multivariate statistical analysis and inverse distance weighted interpolation modeling: the case of a Scopia basin (Central Greece). *Environmental Earth Sciences*,*77*(10), p.380.

Dale, M.R. and Fortin, M.J., 2014. *Spatial analysis: a guide for ecologists*. Cambridge University Press.

Durand, P., Breuer, L., Johnes, P.J., Billen, G., Butturini, A., Pinay, G., Van Grinsven, H., Garnier, J., Rivett, M., Reay, D.S. and Curtis, C., 2011. Nitrogen processes in aquatic ecosystems.

Elhag, M., Psilovikos, A. and Sakellariou-Makrantonaki, M. 2013. Land use changes and its impacts on water resources in Nile Delta region using remote sensing techniques. *Environment, Development and Sustainability*, 15(5), 1189-1204.

Johnson, G., Gervino, N., Gunderson, L., Hotka, L., MacGregor, M., Vavricka, M., Thompson, B., Ganske, L., Leach, M., Schaub, T. and Zadak, C., 2007. Turbidity TMDL protocols and submittal requirements. St. Paul, MN: Minnesota Pollution Control Agency, USA.

Julien, P.Y., 2018. *Stream Restoration. In River Mechanics*. Cambridge University Press. pp. 348-378. doi:10.1017/9781316107072.014

Karamoutsou L., Psilovikos A., Stålnacke P., Farkas C., 2016. Modifications in the Vegoritida's Lake Catchment and Water Level During the Last Sixty Years. European Water, Vol 56, pp. 3 – 12.

Matzafleri N., 2007. Spatial Simulation of the Water Quality Parameters of Lake Kastoria. MSc in Sustainable Management of Aquatic Environment, Department of Ichthyology & Aquatic Environment.

Mantzafleri, N., Psilovikos, A. and Blanta, A., 2009. Water quality monitoring and modeling in Lake Kastoria, using GIS. Assessment and management of pollution sources. *Water resources management*, 23(15), pp.3221-3254.

Matzafleri N. (Unpublished Material). Technical Report of the PhD Thesis.

Psilovikos A., 2014. Ecohydraulics. Tziolas Pub, Thessaloniki, Greece, 536 p. ISBN:978-960-418-447-7 40 (in Greek).

Psilovikos, A., 2020. Water Resources. Tziolas Pub, Thessaloniki, Greece, 700 p.ISBN: 978-960-418-602-0, (in Greek).

Sink, T., Gwinn, J., Gerke, H. and House, M., 2014. Managing and Controlling Algae in Ponds. *Texas A&M AgriLife Extension Service: College Station, TX, USA*.

Smith, K., 1981. The prediction of river water temperatures/prédiction des températures des eaux de rivière. *Hydrological Sciences Journal*, 26(1), pp.19-32.

Stednick, J.D., 1991. *Wildland water quality sampling and analysis*. Elsevier.

Ward, J.V., 1985. Thermal characteristics of running waters. In *Perspectives in southern hemisphere limnology, Springer*, Dordrecht, pp. 31-46

Wetzel, R.G. and Likens, G.E., 2013. *Limnological analyses*. Springer Science & Business Media.

World Health Organization (WHO), 1996. Guidelines for Drinking-Water Quality, Vol.2: Health Criteria and Other Supporting Information, 2nd Edition, World Health Organization, Geneva, 1996.

Zisou, C. and Psilovikos, A., 2012. Monitoring and Modeling of the Physicochemical Parameters Geographic Distribution in Lake Karla using GIS. Proceedings of the 14th Conference of the Hellenic Hydrotechnical Association, pp.1250 – 1261, Patra, Greece.

Internet sources:

https: //geodata.gov.gr/dataset/corine-2000

CHAPTER 6: DEEP NEURAL NETWORKS

"Analyze the unknown based on the known"
Solon, (630-560 BC)

Deep Learning (DL) is the state-of-art paradigm of Artificial Neural Network (ANN) computing. It is a new breakthrough in Machine Learning, compared to shallow learning algorithms. DL is capable of self-learning data features creating a model with the given dataset. This PhD thesis reports the preliminary investigation of deep learning for simulation and real-time prediction. An application of deep learning, for efficiently and effectively extracting the intelligence from data, with potential of applications in simulation and prediction analysis is described in this chapter. Moreover, in order to achieve the aforementioned goal, the platforms and tools used for the models creation are described in detail.

6.1 Machine Learning tools

6.1.1 Programming language

The **Python** programming language is a scripting language, which enables the user to integrate several pieces of code by allowing controlling one or more applications (https://www.python.org/).Its main feature is its flexibility, as the same piece of code can be used with little or no change over a wide range of devices with different architectural and computing capabilities (Abadi et al., 2016).Another advantage is the use and interconnection of many libraries, which makes it an ideal programming language for developing machine learning models that use many different libraries and platforms. One of its applications is to developDeep Neural Networks.

6.1.2 Tools and Platforms

In order to develop deep neural networks -based on Machine Learning methods- the **Python** programming language, the **Spyder** scientific environment, the **Tensorflow** open source machine learning platform and **Nvidia's CUDA** (Compute Unified Device Architecture) parallel platform (graphics card) are used. All of the above have been integrated into **Anaconda**, an Open Source Programming Language.

The **CUDA** platform is a parallel computing platform developed by **Nvidia** that enables users to exploit the computing power of graphics card kernels for training and testing machine learning models. This results in the verification and creation of many models in a very fast time up to 30 times faster than the speed offered by the computing power of the processor (https://developer.nvidia.com/cuda-downloads).

Moreover, the goal of the **Anaconda** tool is to simplify the management and deployment of the packages needed in order to create Machine Learning Models.More specific, Anaconda provides the tools needed to (https://www.anaconda.com/):

✓ Collect data from files and databases,

✓ Manage environments with **Conda**. It should be noticed here that **Conda** is an Open Source Package Management and Environment Management System that installs, runs and updates packages and their dependenciesquickly and easily. It creates, saves, loads and switches between environments on a local computer and it is included in all versions of **Anaconda**(https://docs.conda.io/en/latest/).

✓ Share, reproduce projects and

✓ Deploy projects into production with the single click of a button.

The main reason for using **Anaconda** and **CUDA** is that the model can be created in a variety of virtual environments and with the possibility of using parallel programming through the computing power of the graphics card. The above procedure gives the user the advantage of using large databases and complex learning algorithms, without having to install and re-install various versions of libraries, offering training, verification and testing of the model in a very short time. Moreover, there is no need of buying computing systems at high costs with lower computing power.

6.1.3 Graphical user interface

The **Spyder** scientific environment, written in **Python**, is used as it offers a unique combination of the advanced editing, analysis, debugging, and profiling functionality with an integrated deployment tool with interactive execution, data exploration and visualization capabilities of a scientific package (https://www.spyder-ide.org/).**Spyder** is also an open source code processing environment but only for the **Python** programming language. The **Spyder** environment enables the user to execute line-by-line the code while displaying the results for each execution step. This allows the user to see if each step of the code is able to print individual diagrams or sections of code without having to re-execute the entire code again.

6.1.4 Libraries

Various libraries have been used for **Python**programming (Figure 6.1). The libraries needed are not loaded in default. For that reason, a good practice to begin the code is by importing all the libraries needed. With the help of **Pathlib**, file paths can be represented by proper path objects which make code dealing with file paths. The **Matplotlib** mathematical library is used for importing and editing complex mathematical operations, as well as for editing and visualizing a variety of diagrams (complex or not). Moreover, **Pandas**provide some powerful objects which are very useful for cleaning, transforming, manipulating and analyzing data. Furthermore, seaborn is a library for making statistical graphics in **Python**. A functionality that it offers for example is the option for visualizing univariate distributions.It is built on top of matplotlib and closely integrated withpandasdata structures.**Seaborn**aims to make visualization a central part of exploring and understanding data.

```
IPython 7.8.0 -- An enhanced Interactive Python.

In [1]: from __future__ import absolute_import, division, print_function,
unicode_literals
   ...: import pathlib
   ...: import matplotlib.pyplot as plt
   ...: import pandas as pd
   ...: import seaborn as sns
   ...: import tensorflow as tf

   ...: from tensorflow.keras import layers
   ...:
   ...: print(tf.__version__)
1.9.0
```

Figure 6.1: Libraries for **Machine Learning**

The most important of all libraries is the open source **Tensorflow** machine learning library. The **Tensorflow** platform is an end-to-end open source platform for developing machine learning models, developed by Google. It consists of a comprehensive, flexible ecosystem set of tools and libraries that allow users to easily use and develop Machine Learning Models (https://www.tensorflow.org/).This library provides State-of-the-Art Machine Learning Methods allowing the users to create their own Machine Learning Models in a user-friendly, multi-pack and add-on environment. In this PhD thesis, the **Tensorflow GPU library** is used, which enables the user with the **CUDA** tool to train and test the models they have developed through the computing power of the graphics card kernels.

6.2Deep Neural Networks Development

6.2.1 Input and output data

The predictive ability of networks is based on their training, which in turn depends on the amount of information available on the network. For the purpose of this thesis, time series of quality parameters from 11 November 2015 to 15 March of 2018 were used on an hourly basis (20,228 measurements/per parameter/per station) from 4 telemetric stations located in the study area.More specific, the available data consist of 1) **Chl-a** (μg/L), 2) **pH**, 3) **Tw** (°C), 4) **ECw**(μS/cm), 5) **Turb** (NTU), 6) **Ammonia Nitrogen**($N - NH_4$, ppm) (Not Available (NA) data for Stavros Station), 7) **Nitrate Nitrogen** ($N - NO_3$, mg/L) and 8) **Dissolved Oxygen** (mg/L).Three different models are investigated in the present thesis and more specific:

- ✓ DO model
- ✓ Turbidity model and
- ✓ Chl-a model

An important step in Deep Learning, is the selection of appropriate input variables. Based on the literature, as it has already mentioned, Neural Network studies reported that most important water quality parameters, for the modelling of Dissolved Oxygen, are pH and Water Temperature (Ranković et al. 2010; Wen et al. 2013). In addition to those two inputs, many other water quality parameters have also been used. Wen et al. (2013), Basant et al. (2010), Singh et al. (2009) and Kuo et al. (2007), reported the importance of NO_3–N and NH_4–N as input variables for DO models. In addition, Ranković et al. (2010), proposed Turbidity as an input variable, while Kuo et al. (2007), reported the importance of Chl-a. ECw, is also proposed as an input variable for DO modelling (Ay and Kisi, 2012; Wen et al. 2013). Fewer studies found in literature concerning the modelling of Chl-a. The most representative one, is the study of Kuo et al. (2007), reporting the importance of T, NO_3–N, DO, pH and Turb, as input variables in a Chl-a model. Concerning the modelling of Turbidity, studies reported the importance of pH, conductivity, NH_4–N, DO and T, as input variables for the model (Iglesias et al., 2014).

At the present PhD thesis, based on the literature, all the available parameters are taken into account as input parameters for each investigated model. Subsequently, based on the literature and taking into account the resulting tables, concerning the impact of each input parameter to the output parameter (DO), Deep Neural Networks consisting of 4 input parameters namely:

1) Nitrate nitrogen ($N - NO_3$, mg/L),

2) pH,

3) Tw (° C) and

4) ECw (μS/cm)

are also tested.

6.2.2 Number of Hidden Layers

The generalization capability of an Artificial Neural Network is linked to its hidden layer. Too many hidden layers in a network, increases computational burden and causes over-fitting which results in poor prediction. Several studies, show that one or two hidden layers mostly produce better performance(Smithson et al., 2016;Gaya et al., 2017).It should be reminded, that if an ANN has more than three layers, including input and output layer, it is called **Deep Neural Network**. As the scope of this thesis is to apply a Deep Neural Network, for modelling and prediction of quality parameter, two hidden layers are used in the present thesis for different structures of the investigated models.

6.2.3 Number of nodes in the Hidden Layer

Deciding the appropriate number of nodes to use, is crucial for effective learning and performance of the network. Nevertheless, there is no systematic approach to determine the optimal number of nodes to utilize for a problem (Smithson et al., 2016; Gaya et al., 2017).At the present PhD thesis,the preliminary investigation of Deep Neural Networks for simulation and real-time prediction, uses as default 64 neurons per hidden layer. The predictive ability of a Deep Neural Network, consisting of 32 nodes per hidden layer, is also tested.

6.2.4 Training Epochs

In order to train a Neural Network, many epochs are needed. Several studies indicated that convergence could be achieved with training epochs from 85 to 5000 epochs (Smithsonet al., 2016; Gaya et al., 2017). In this PhD thesis, 1000 epochs are used to train the investigated models. In casesthat convergence could not achieved (Turbidity and Chl-a models), the number of epochs reaches 5000.

6.2.5 Activation function

Activation functions are an essential part of neural networks as they provide non-linearity. The absence of non-linearity, turns the neural network to a simple logistic regression model. An activation function is a function which is applied to the output of a neural network layer, which is then passed as the input to the next layer. The most widely used activation function is the **Rectified Linear Unit** (**ReLU**) (Figure 6.2) that is defined as (Equation 6.1):

$$f(x) = max(0, x) \qquad (6.1)$$

where: x is the input to the neuron.

An advantage of **ReLU** is that it is a highly simplified and easy to calculate function. It is also very quick to use and train when compared with other activation functions. Moreover, in Machine Learning, updating a parameter is proportional to the partial derivative of the error function with respect to these parameters. If the gradient becomes too small, the updates will not work and the network may stop the training procedure. **ReLU** doesn't saturate in the positive direction, while other activation functions like sigmoid and hyperbolic tangent saturate in both directions. Therefore, it has fewer vanishing gradients, resulting in better training (https://www.tensorflow.org/).The limitation of **ReLU** is that its mean output is not zero (Hossen et al., 2017).

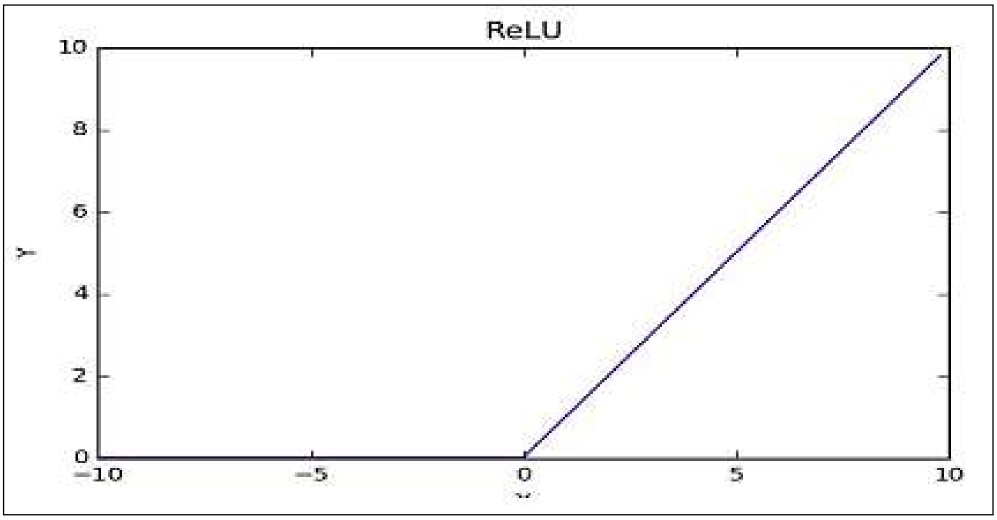

Figure 6.2: ReLU Activation Function

6.2.6 Early stopping technique

Early Stopping is one of the most popular, and also effective, techniques that are used to prevent over-fitting. Specifically, the validation data set is used to compute the loss function at the end of each training epoch. Once the loss stops decreasing, the training stopsand the test data is used in order to compute the final accuracy. The point, at which the validation loss starts to increase, is when the model starts to over fit the training data (Figure 6.3) (Epelbaum, 2017).

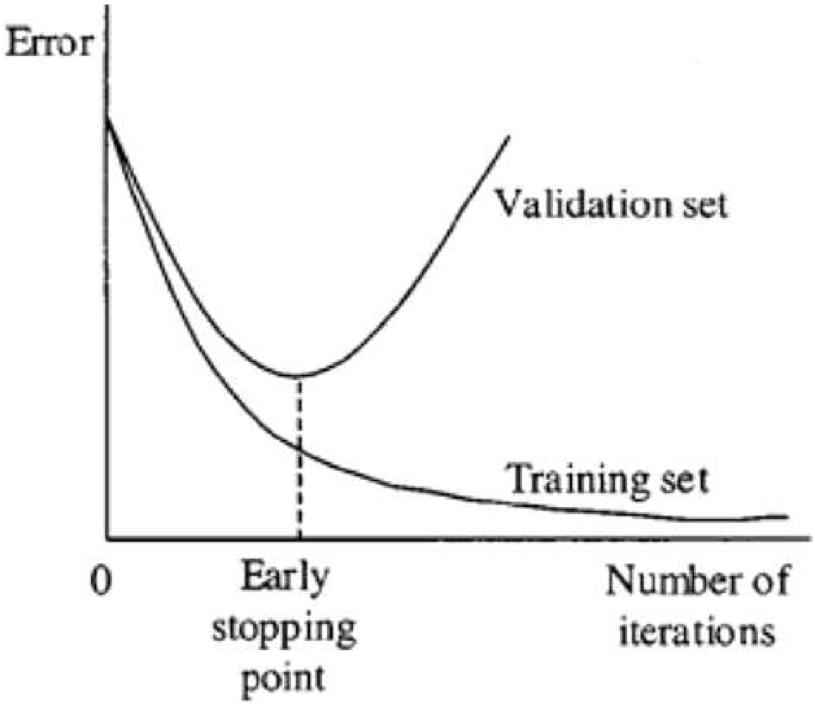

Figure 6.3: Early stopping technique

It should be mentioned here, that the Optimal Stopping Point can be considered as a hyper-parameter. This makes Early Stopping more efficient, than other hyper-parameter optimization techniques, which typically require a complete run of the model to test out a single hyper-parameter value. Moreover, it does not require any changes to the model, that could change the learning dynamics of the system (Epelbaum, 2017).

6.2.7 Learning rate

The learning rate is a hyper-parameter that controls how much to change the model in response to the estimated error each time the model weights are updated. If the learning rate is too large, the network fails to converge with allowable error over training set. Choosing a too small learning rate, resultsin slow training process. Most of the Neural Networks utilizea value from 0,01 to 0,3 of the learning rate. In the present PhD thesis, taking advantage of the computing power of the graphic card, a learning rate of 0,001 is used.

6.2.8 Optimization algorithm

With a strong ability to find the most optimistic result, the **RMSprop** is chosen as the optimization algorithm in this PhD thesis.**RMSprop** is one of the most popular optimization algorithms used in Deep Learning, as it is a fast and good optimizer (Karparthy, 2017).It uses a moving average of squared gradients to normalize the gradient itself.That has an effect of balancing the step size — decreases the step for large gradient to avoid exploding, and increase the step for small gradient to avoid vanishing. RMSprop avoids the decay of the learning rate to zero.

6.2.9 The structures

Deep Learning based techniqueis used in order to develop the (Karamoutsou and Psilovikos, 2019):

a) DO model,

b) Chl-a model and

c) Turb model

which are applied to the data of the four stations separately. To achieve the goal of this dissertation a Deep Neural Network (DNN) was built for each case consisting of:

1) An input layer of quality parameters, depending on the investigated structure

2) Two densely connected hidden layers, consisted of 64 or 32 units/nodes each, depending on the investigated structure and

3) An output layer of one quality parameter (DO/Chl-a/Turbidity), depending on the investigated model.

The examined Deep Neural Networks structures are listed below:

DO model with structures:

✓ 7-64-64-1 for Gkiole, Psaradika and Toichio stations (Figure 6.4) and6-64-1 for Stavros station, where NH_4 parameter is not available (Figure 6.5)

✓ 4-64-64-1 for all stations (Figure 6.6)

✓ 7-32-32-1 for Gkiole, Psaradika and Toichio stations (Figure 6.7) and 6-32-1 for Stavros station where NH_4 parameter is not available (Figure 6.8)

✓ 4-32-32-1 for all stations (Figure 6.9)

Turbidity and Chl-a model with structures:

✓ 7-64-64-1 for Gkiole, Psaradika and Toichio stations (Figure 6.4) and (6-64-1) for Stavros station where NH_4 parameter is not available (Figure 6.5).

Here, a DNN with structurefor example (7-64-64-1), indicates a model comprising 7 inputs, 64 nodes per hidden layer and one output node. In terms of how the nodes are connected to each other, a Feed Forward Neural Network is used for each case, as there is no feedback from the outputs towards the inputs. The examined Deep Feed Forward Neural Networks (DNN) structures of DO, Chl-a and Turbidity models for Lake Kastoria are shown below.

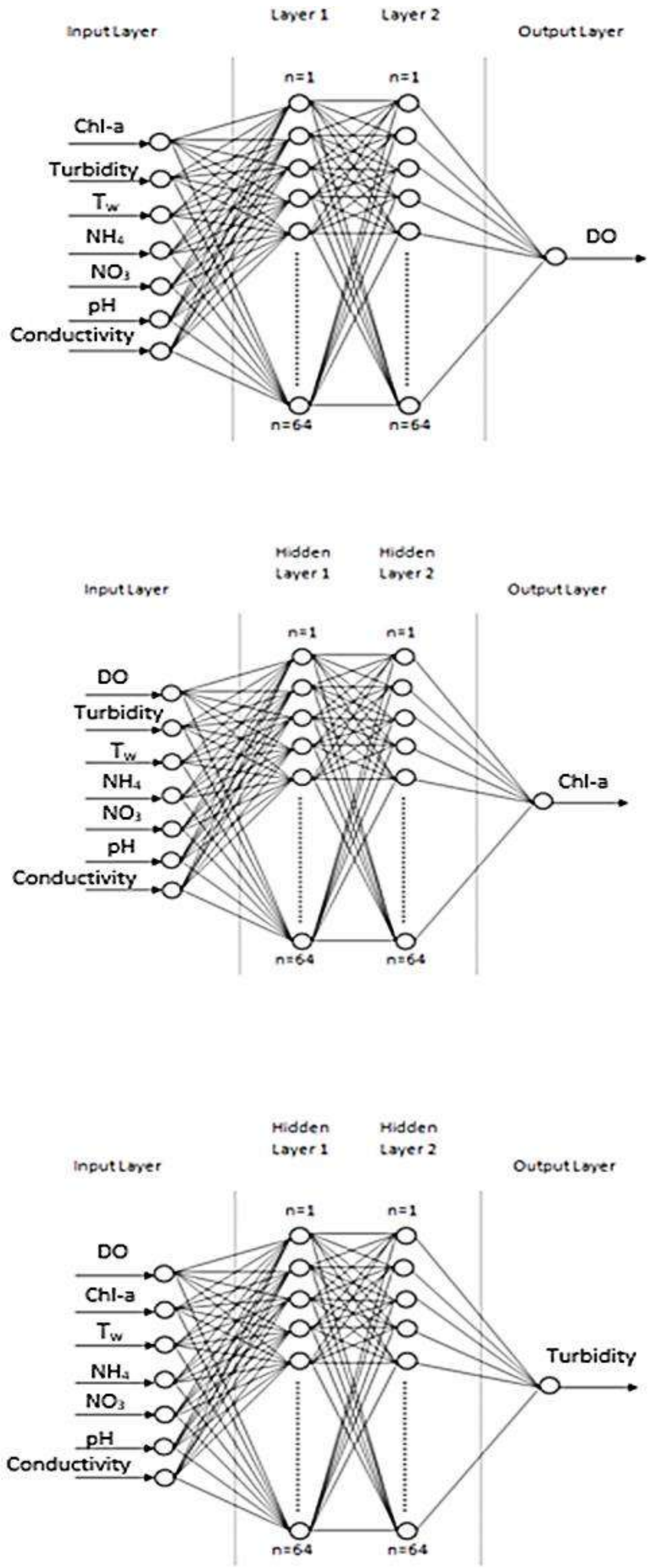

Figure 6.4: DO, Turbidity and Chl-a model with structure 7-64-64-1 for Gkiole, Toichio andPsaradika Station

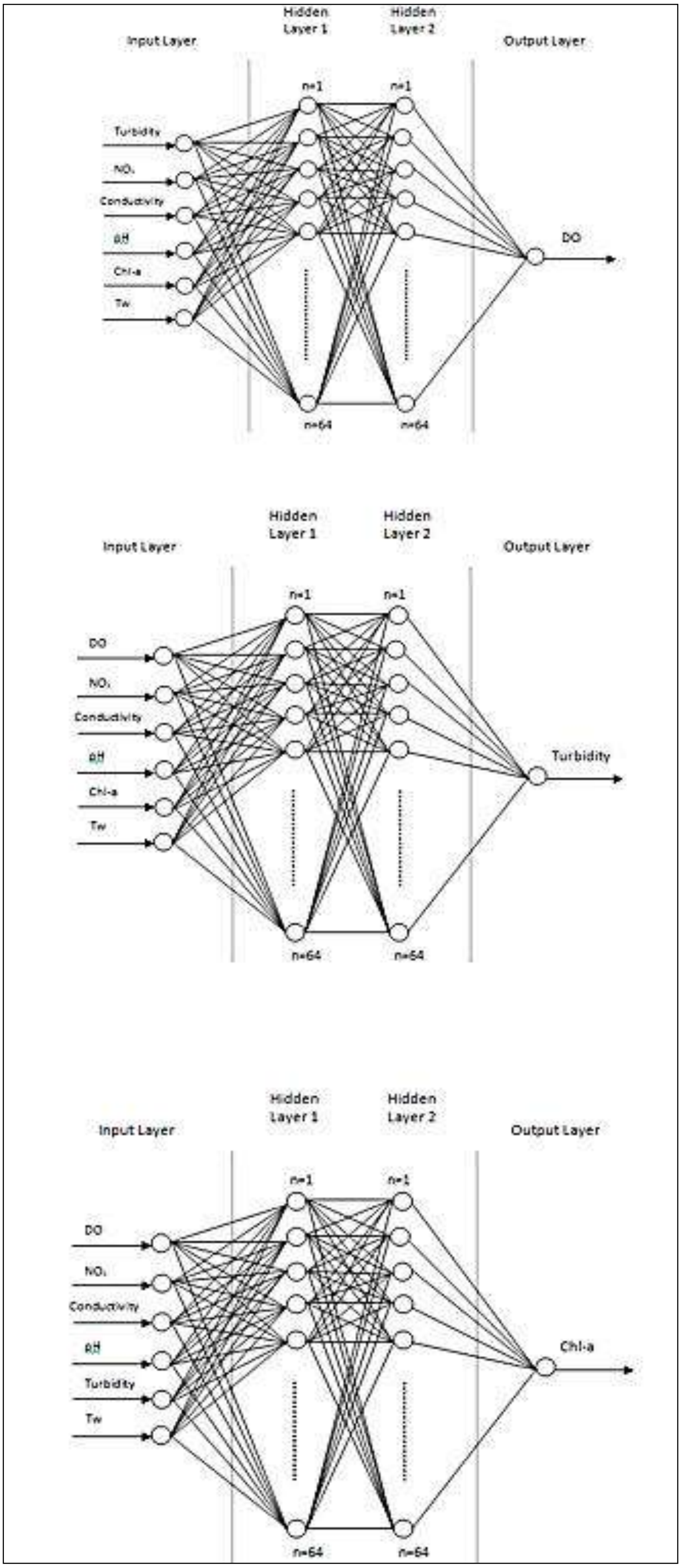

Figure 6.5: DO, Turbidity and Chl-a modelDO models with structure 6-64-64-1 for Stavros Station

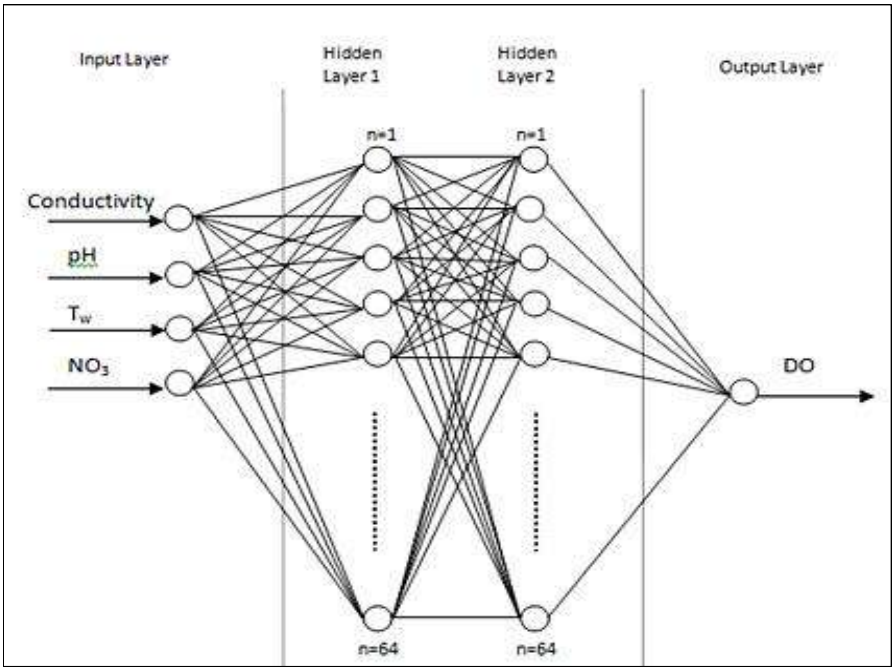

Figure 6.6: DO models with structure 4-64-64-1 for all stations

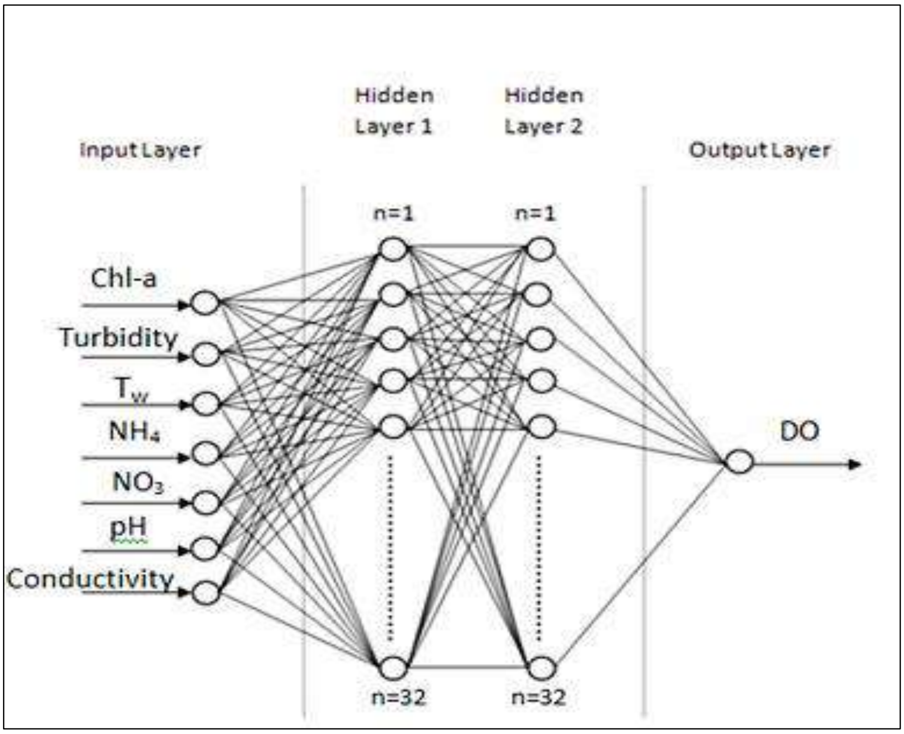

Figure 6.7: DO models with structure 7-32-32-1 for Gkiole, Toichio and Psaradika Station

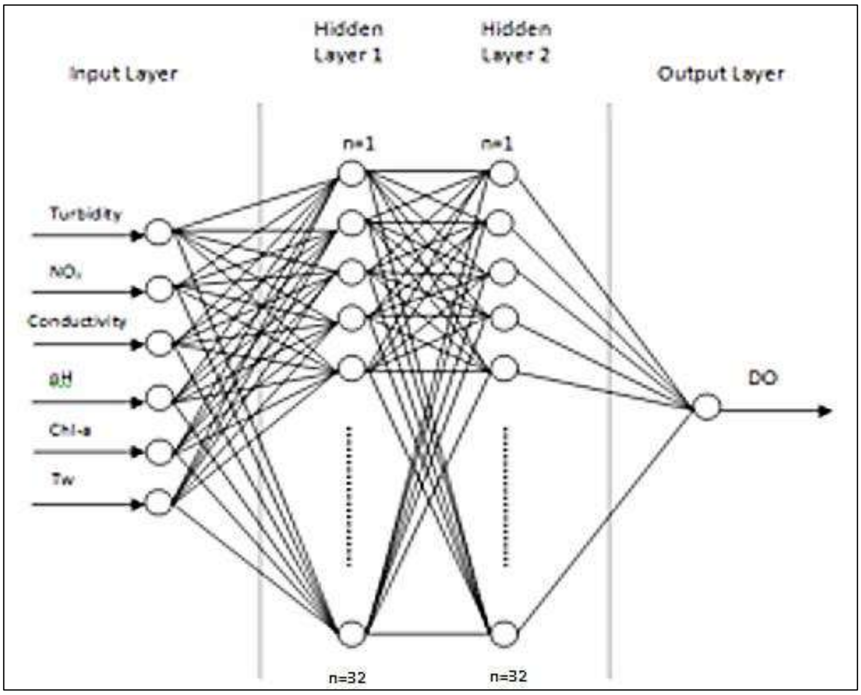

Figure 6.8: DO models with structure 6-32-32-1 for Stavros Station

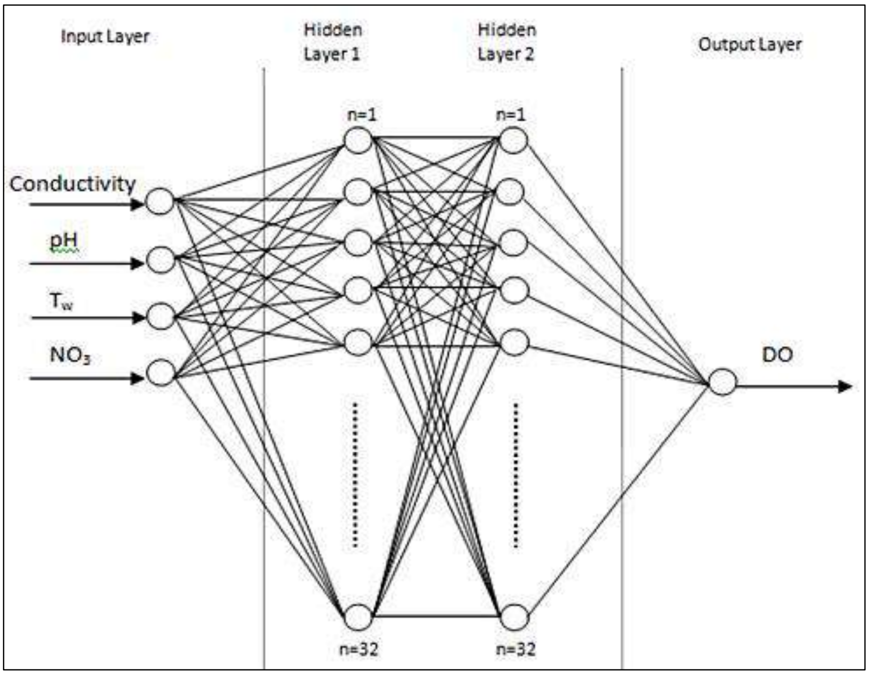

Figure 6.9: DO models with structure 4-32-32-1 for all stations

6.3 Steps of the process

The procedure that takes place, is described as follows:

6.3.1 Preparation of the Dataset

Step 1: The dataset is imported.

Step 2: The dataset is cleaned from unknown values. Because the data of some variables were unrecorded, these data were removed from the dataset. Therefore, the last dataset for each station includes (Table 6.1)

Table 6.1: Number of data/parameter

Station	Number of data/parameter
Gkiole	16922
Toichio	19763
Psaradika	16574
Stavros	19154

Step 3: In this context, there were three phases, the training, testing, and validation for each station. 80% of the data was used for the training process, and 20% for testing. A percentage of 80% of training data was re-divided into 60% for the training process and the remaining20% for validation.

Step 4: The dataset is inspected and the overall statistics are taken into account. The detailed result tables are shown in the next chapter.

Step 5: Split features (inputs) from labels (outputs) (*In Machine Learning feature is input; label is output. This applies to regression problems. Afeatureis one column of the data in the input set*).

Step 6: The normalization of the dataset takes place. This normalized dataset is that it will be used to train the models. This is helpful in order to keep the fitting parameters on a scale that the computer can handle without a damaging loss of accuracy. Normalizing original data boosts the performance of Neural Network Models significantly.

6.3.2 The model

Step 1: Build the model: Based on different structures, twenty models in total with two densely connected hidden layers and an input layer of DO /Turbidity/Chl-a, are conducted. This dissertation approaches building a **Keras**Neural Network, using the **Sequential model API**.The **Rectified Linear Unit (ReLU)**activation functionand the **RMSprop** as the Optimization Algorithmareused, as described above.

Step 2: Inspect the model: In this step *"the summary "*method is used in order to print the description of the model. The detailed result tables are shown in the next chapter. Here, a batch of ten samples is taken from the training data, in order to try out the model.

Step 3: Train the model: The models, are firstly trained for 1000 epochs. When the results are not satisfactory, the epochs are increased. In cases of Turbidity and Chl-a model, where the validation and training processes couldn't converge, the number of epochs is increased to 5000.At this point, it should be mentioned that Early Stopping Technique is used as it is useful, in order to avoid over-fitting (Epelbaum, 2017). For that reason, the "*model.fit"* is updated to stop training when the validation score is not improved.

Step 4: Make predictions: In this step the test set is used, to see how well the model is expected to make predictions. The result, Tables and Figures, are shown in the next chapter.

6.4Statistical descriptors

The following statistical measures were used to evaluate the predictive ability of the Neural Network Models:

a) The **Mean Absolute Error (Mean Absolute Error, MAE)**

b) The **Mean Square Error (Mean Square Error, MSE)** and

c) The**Nash–Sutcliffe model Efficiency coefficient (NSE)**

After calculating the predicted values, it is of significant importance to know how far the "*predicted yvalue"*is away from the "*generated y value"*. To do this, a method is designed to calculate the "*gap*". This design is known as theloss function. The loss function evaluates the error performed by the Feed Forward Neural Network when it tries to estimate the data to be predicted. For a regression problem, this is simply a mean square error (MSE) evaluation.

a) The **Mean Square Error** is defined as (Equation 6.2):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (6.2)$$

It is a statistic to quantify the difference between observed and estimated values of a parameter and proportional to the measure of dispersion of a random variable. In proportion to the **Standard Deviation**, the square root of the **MSE** is also obtained.

b) The **Mean Absolute Error** is defined as (Equation 6.3):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \qquad (6.3)$$

It is another statistic that measures the accuracy of a model's adaptation and prediction. Lower values of the above criteria, indicate model estimators closer to actual values, thus better adaptive or predictive ability of the model (Benyahya et al., 2007).

c) The**Nash–Sutcliffe Model Efficiency Coefficient (NSE)**is used to evaluate the performance of HydrologicModels. Itisdefinedas (NashandSutcliffe, 1970) (Equation 6.4):

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{z})^2} \qquad (6.4)$$

where:

$y_i$and $\hat{y}_i$ denote the observedmeasurement and the model predicted values from the ith element;

$\overline{z}$ the average of the observedmeasurements and n represents the number of observations.

Nash–Sutcliffe efficiency can range from −∞ to 1. An efficiency of 1 (*NSE*=1), corresponds to a perfect match of modeled and observed data. An efficiency of 0 (*NSE*=0), indicates that the model predictions are as accurate as the average of the observed data, whereas an efficiency less than zero (*NSE*<0) occurs when the observed mean is a better predictor than the model. The closer the model efficiency is to 1, the more accurate the model is.

References

Abadi M., Barham P., Chen J., Chen Z., Davis A., Dean J., Devin M., Ghemawat S., Irving G., Isard M., Kudlur M., 2016. Tensorflow: A system for large-scale machine learning. In 12th {USENIX} Symposium on Operating Systems Design and Implementation, 16, pp. 265-283.

Ay, M. and Kisi, O., 2012. Modeling of dissolved oxygen concentration using different neural network techniques in Foundation Creek, El Paso County, Colorado. *Journal of Environmental Engineering*, 138(6), pp.654-662.

Basant, N., Gupta, S., Malik, A. and Singh, K.P., 2010. Linear and nonlinear modeling for simultaneous prediction of dissolved oxygen and biochemical oxygen demand of the surface water—a case study. *Chemometrics and Intelligent Laboratory Systems*, 104(2), pp.172-180.

Benyahya L., St-Hilaire A., Quarda T.B., Bobée B., Ahmadi-Nedushan B., 2007. Modeling of water temperatures based on stochastic approaches: case study of the Deschutes River. *Journal of Environmental Engineering and Science*, 6(4), pp.437-448.

Epelbaum, T., 2017. Deep learning: Technical introduction. *arXiv preprint arXiv:1709.01412*.

Gaya, M.S., Zango, M.U., Yusuf, L.A., Mustapha, M., Muhammad, B., Sani, A., Tijjani, A., Wahab, N.A. and Khairi, M.T.M., 2017. Estimation of turbidity in water treatment plant using Hammerstein-Wiener and neural network technique. *Indonesian Journal of Electrical Engineering and Computer Science*, 5(3), pp.666-672.

Hossen, T., Plathottam, S.J., Angamuthu, R.K., Ranganathan, P. and Salehfar, H., 2017, September. Short-term load forecasting using deep neural networks (DNN). In *2017 North American Power Symposium (NAPS)*, IEEE, pp. 1-6.

Iglesias, C., Torres, J.M., Nieto, P.G., Fernández, J.A., Muñiz, C.D., Piñeiro, J.I. and Taboada, J., 2014. Turbidity prediction in a river basin by using artificial neural networks: a case study in northern Spain. *Water resources management*, 28(2), pp.319-331.

Karamoutsou, L. and Psilovikos, A., 2019. The Use of Artificial Neural Networks in Predicting the Water Quality of Lake Kastoria. *Proceedings of the 14th Conference of the Hellenic Hydrotechnical Association (EYE)*, pp.882 - 889, Volos.

Karparthy, A., 2017. A peek at trends in machine learning. Available at: https://medium.com/@karpathy/a-peek-at-trends-in-machine-learning-ab8a1085a106

Kuo, J.T., Hsieh, M.H., Lung, W.S. and She, N., 2007. Using artificial neural network for reservoir eutrophication prediction. *Ecological modelling*, 200(1-2), pp.171-177.

Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology*, 10(3), pp.282-290.

Ranković, V., Radulović, J., Radojević, I., Ostojić, A. and Čomić, L., 2010. Neural network modeling of dissolved oxygen in the Gruža reservoir, Serbia.*Ecological Modelling*,221(8), pp.1239-1244.

Singh, K.P., Basant, A., Malik, A. and Jain, G., 2009. Artificial neural network modeling of the river water quality—a case study.*Ecological Modelling*,220(6), pp.888-895.

Smithson, S.C., Yang, G., Gross, W.J. and Meyer, B.H., 2016, November. Neural networks designing neural networks: multi-objective hyper-parameter optimization. In *Proceedings of the 35th International Conference on Computer-Aided Design*, pp. 1-8.

Wen, X., Fang, J., Diao, M. and Zhang, C., 2013. Artificial neural network modeling of dissolved oxygen in the Heihe River, Northwestern China.*EnvironmentalMonitoring and Assessment*,185(5), pp.4361-4371.

Internet sources

https://www.tensorflow.org/

https://docs.conda.io/en/latest/

https://www.anaconda.com/

https://www.spyder-ide.org/

https://www.python.org/

https://developer.nvidia.com/cuda-downloads

https://www.tensorflow.org/tutorials/keras/regression

CHAPTER 7: RESULTS AND DISCUSSION

"Your future will be bright only
if you take care of your present"
Isocrates, (436-338 BC)

7.1 Assessment of the state of Lake Kastoria

The Gkiole's channel constitutes the only surface discharge of Lake Kastoria's catchment, as well as the connection between Lake Kastoria and River Aliakmon. The presence of the weir vertically in Gkiole's channel flow, enables the optimization of the Lake Kastoria'sWater Balance and Eutrophication Control. For this reason, it is an important Monitoring Station for eutrophication control, as it is essential to know Specific Water Quality Parameters, in order to control of the possible potential pollution to Aliakmonas River. Today, the stream presents very smooth inclines which, along with the lush vegetation of the area, inhibit the discharge of water. Moreover, Toichio Station was the first point in time in the lake, that experienced eutrophication phenomena, due to Algal Blooms ("greening" of water). The Stavros and Psaradika stations are located South and North of the coastline, respectively, that defines the urban front of the City of Kastoria. They constitute Monitoring Stations, very important for measuring Water Quality Parameters and Chl-a, as a significant Ecological Parameter, assessing the sensitivity of the aquatic ecosystem by the Urban Pollution pressures. Concerning the Western side of the lake, the waters are characterized by a stationary water status, in the urban part of the lake where Kastoria city is settled.

Table 7.1: Descriptive Statistics of the Water Quality Datafor Gkiole Station

Variables	Unit	Mean	Std dev	Min	25%	50%	75%	Max
Chlorophyll-a	(μg/L)	12,44	19,65	0,80	5,00	7,80	14,20	441,10
Conductivity	(μS/cm)	294,74	27,80	225,00	276,00	301,00	318,00	338,00
pH	-	8,42	0,37	7,50	8,10	8,40	8,70	9,60
Turbidity	NTU	21,28	66,11	0,10	0,60	5,10	21,80	3.000,00
Temperature	^oC	14,32	7,58	0,40	7,00	13,90	21,30	30,10
NH_4-N	ppm	1,53	2,47	0,10	0,20	0,80	1,80	20,00
NO_3-N	mg/L	1,43	3,06	0,01	0,30	0,41	1,98	63,94
DO	mg/L	8,76	2,39	0,40	7,18	8,96	10,37	17,65

It is crucial, for environmental and public health concern, that the concentration of each quality parameter must be within the allowable limits defined by the Legislation.For Gkiole, Toichio, Psaradika and Stavros station Tw has a mean value of 14,32°C (Table 7.1), 14,33°C (Table 7.2), 12,5°C (Table 7.3) and 14,58°C (Table 7.4) respectively. Regarding the maximum permitted values of Tw for aquatic life (Joint Ministerial Decision 46300 (1986)) for salmonids the Tw equals to 21,50°C and for cyprinid equals to 28°C. The maximum recorded values of Twfor all stations exceed the permitted ones. More specific, for Gkiole, Toichio, Psaradika and Stavros station the maximum water temperatures equals to 30,10°C, 30,40°C, 32,40°C and 31,80°C respectively.

For Dissolved Oxygen (DO – mg/L), the minimum values observed in the dataset for Gkiole, Toichio, Psaradika and Stavros station equals to 0,4 mg/L,0,01 mg/L,0,01 mg/L and 0,01 mg/L respectively, in all cases less than 4 mg/L, which is the minimum absolute permitted value for aquatic life. The range of DO concentration, is usually between 5 mg/L and 9 mg/L, but both extreme values close to 0 mg/L minimum and 19 mg/L maximum – or even more – have been observed. According to DOsat formula, the highest DOsat value is equal to 14,652 mg/L and it is associated with the Tw value of 0°C. In literature and in the Monitoring data, higher concentrations than 14,652 mg/L, are also found. For Gkiole station the maximum value for DO is 17,65 mg/L, for Toichio station equals to 18,27 mg/L, while for Psaradika and Stavros station the maximum values of DO are 16,00mg/Land 19,28 mg/L respectively. These values are so high, due to the intense photosynthesis phenomenon, during the period of May and June, which favors the respiration phenomenon, yet.

Table 7.2: Descriptive statisticsof the water quality data for Toichio Station

Variables	Unit	Mean	Std dev	Min	25%	50%	75%	Max
Chlorophyll-a	μg/L	11,74	11,48	0,30	5,40	8,40	14,10	453,30
Conductivity	μS/cm	316,18	25,26	239,00	307,00	318,00	331,00	369,00
pH	-	8,41	0,35	5,40	8,10	8,40	8,60	9,80
Turbidity	NTU	20,87	52,28	0,10	0,00	1,80	49,00	3000,00
Temperature	°C	14,33	7,83	0,80	6,80	13,80	21,80	30,40
NH_4	mg/L	0,56	1,14	0,10	0,00	0,10	0,60	19,70
NO_3	mg/L	1,30	0,74	0,01	0,83	1,24	1,70	3,72
DO	mg/L	8,10	2,60	0,01	6,34	8,29	9,73	18,27

In case of $NO_3 - N$, the mean concentrations for Gkiole, Toichio, Psaradika and Stavros station are equal to 1,43, 1,30, 1,07 and 1,48 mg/L respectively. According to the literature, values for $NO_3 - N$ that exceed 0,3 mg/L, favors the development of eutrophic conditions. Moreover, it should be mentioned here that, according to WHO (1996), a maximum concentration of 10 mg/L NO_3-N is considered as the maximum concentration for drinking water. In Europe, the maximum values are 11,3 mg/L respectively. These concentrations are in accordance with JMC 2600 (2001) and JMC 46399 (1986). The last Decision mentions the same maximum values for living aquatic organisms. The maximum concentrations for Toichio, Psaradika and Stavros station are 3,72 mg/L, 7,89 mg/L and 19,28 mg/L respectively. Gkiole station high values (max value=63,94 mg/L) can be considered as an accidental and occasional event. According to Antonopoulos (2010), NO_3-N values exceeding 0,5 mg/L, show eutrophic conditions, while values above 1 mg/L, show anthropogenic pollution interventions from urban and agricultural sources (Psilovikos, 2014).

Table 7.3: Descriptive statisticsof the water quality data for Psaradika Station

Variables	Unit	Mean	Std dev	Min	25%	50%	75%	Max
Chlorophyll-a	(μg/L)	5,61	11,63	0,10	0,20	2,10	6,00	445,70
Conductivity	(μS/cm)	315,00	25,32	229,00	300,00	320,00	329,00	375,00
pH	-	8,33	0,29	7,70	8,10	8,30	8,50	9,80
Turbidity	NTU	60,85	346,86	0,10	0,10	4,30	9,50	3.000,00
Temperature	°C	12,85	7,78	0,10	6,10	11,10	18,70	32,40
NH_4	ppm	0,40	1,47	0,10	0,10	0,10	0,70	20,00
NO_3	mg/L	1,07	0,67	0,01	0,78	0,95	1,30	7,89
DO	mg/L	6,85	2,95	0,01	5,00	7,38	9,07	16,00

Electrical Conductivity of Water, for irrigation purposes, greater than 700 μS/cm, can cause soil salinity problems, depending on the type of soil, the type of crop, the climate and the way the water is applied. ECw greater than 3000 μS/cm, like in Lake Karla (Mellios, 2020), is generally unsuitable for irrigation. According to JMD 46399 (1986), the acceptable values for both drinking water and living aquatic organisms are less than 1000 μS/cm. Moreover, according to JMC 2600 (2001) and WHO (1996), the maximum permitted values for drinking water are 2500 and 400 μS/cm respectively (Psilovikos, 2014). None of the four stations recorded values above the permitted limits.

Gkiole, Toichio, Psaradika and Stavros station recorded mean values of pH 8,42; 8,41; 8,33 and 8,22 respectively. According to JMD 46399 (1986), JMD 2600 (2001) and WHO (1996) the acceptable values for drinking water range from 6,5 to 8,5. Moreover, according to JMC 46399 (1986) the acceptable values for aquatic organisms range from 5,5 to 9 (Psilovikos, 2014). In waters with a pH below 5, generally not many fish species are observed. Moreover, an active acidity below 4 causes the death of all vertebrates, some invertebrates, and many microorganisms. In the case of flora, water below pH 4,5 may harm plant organisms either directly or indirectly. None of the four stations recorded values below the permitted limits.

Table7.4: Descriptive statisticsof the water quality data for Stavros Station

Variables	Unit	Mean	Std dev	Min	25%	50%	75%	Max
Chlorophyll-a	μg/L	6,52	9,78	0,10	1,50	3,00	6,80	125,00
Conductivity	μS/cm	341,66	56,50	234,00	304,00	331,00	390,00	457,00
pH	-	8,22	0,37	5,60	7,95	8,20	8,50	9,50
Turbidity	NTU	9,75	86,28	0,10	0,10	0,10	0,10	2512,00
Temperature	^oC	14,58	8,20	0,10	6,80	13,70	21,90	31,80
NH_4	mg/L	NA	NA	NA	NA	NA	NA	NA
NO_3	mg/L	1,48	1,36	0,01	5,27	7,60	9,60	19,28
DO	mg/L	7,27	3,18	0,01	5,27	7,60	9,60	19,28

Regarding Chl-a, Gkiole and Toichio station recorded mean values of 12,44 μg/L and 11,74 μg/L respectively. The areas at the lowest part of the torrent Toichio are intended for agricultural use and, as it has already mentioned, Toichio station was the first point in time in the lake, that experienced eutrophication phenomena. North (Stavros) and South(Psaradika) of the coastline that defines the city of Kastoria, the stations recorded lower meanChl-a values of 6,52 μg/L and5,61 μg/L respectively. According to Thomann and Mueller (1987), lakes with Chl-a values higher than 10 μg/L, are considered as eutrophic. Nürnberg (1996), considers the trophic status of a lake as eutrophic when Chl-a values range from 9,1 to 25 μg/L. Moreover, according to Nurnmerg (1996) and OECD (1982), a lake is considered as hypertrophic when Chl-a values exceed 25 μg/L.

Finally, for Gkiole, Toichio, Psaradika and Stavros station the recorded mean values of Turbidity are 21,28; 20,87; 60,85 and 9,75 NTU respectively. In general, Turbidity levels can range from less than 1 NTU to more than 1.000 NTU. Water at 5 NTU is visibly cloudy, while water at 25 NTU is murky. According to WHO (1996), the turbidity of drinking water should not be more than

5 NTU, and should ideally be below 1 NTU.Figures 7.1-7.4 demonstratethe data distribution of water quality variables (diagonally), as well as, how variables affect each other for Gkiole, Toichio, Psaradika and Stavros station respectively. It is obvious that the distribution of most water quality variables is not symmetric. Artificial Neural Network, can be a suitable method in modeling water quality variables, as it supports the nonlinear relationships between variables.

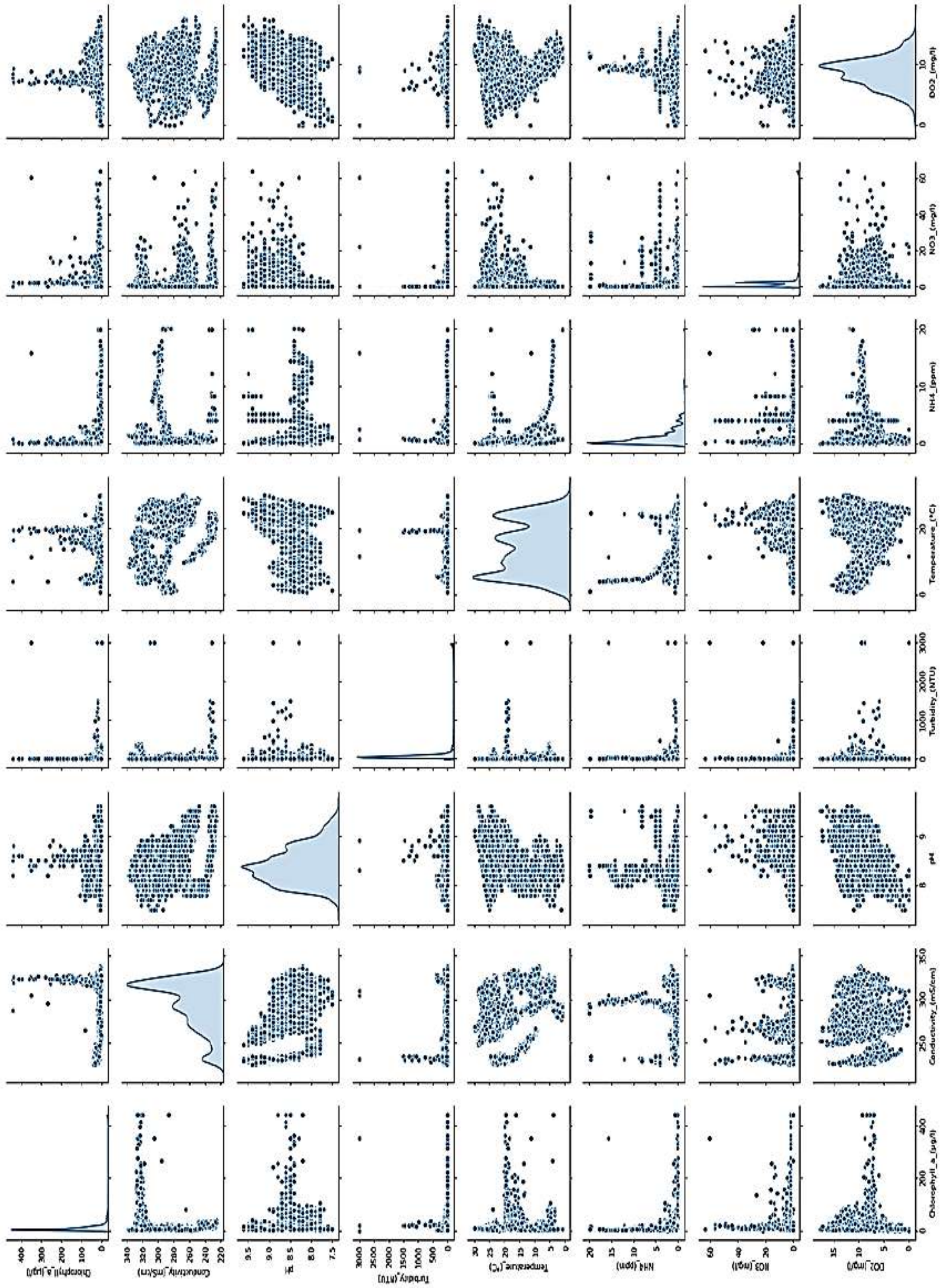

Figure 7.1: Joint distribution of pairs of parameters from the training set for Gkiole station

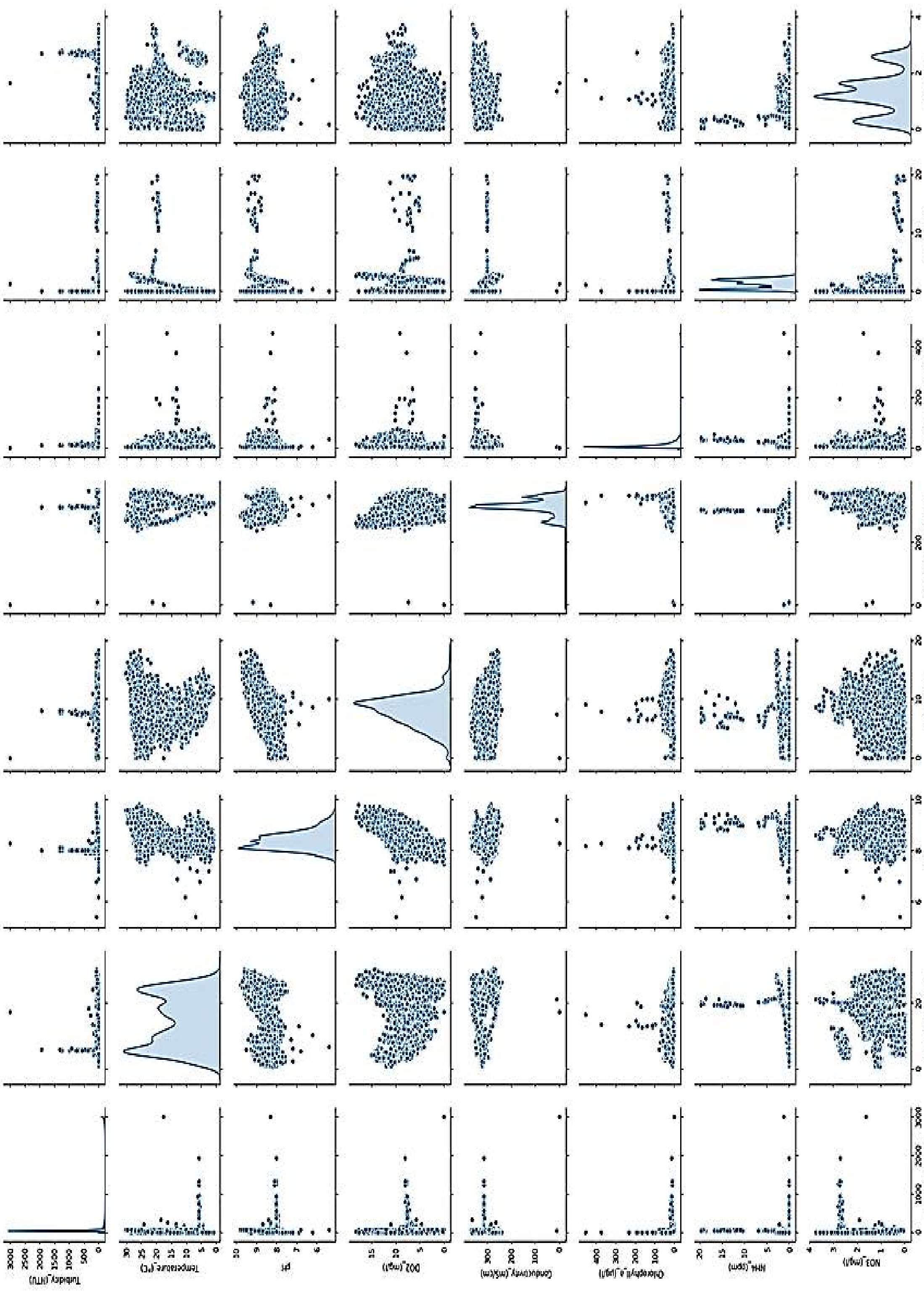

Figure 7.2: Joint distribution of pairs of parameters from the training set for Toichio station

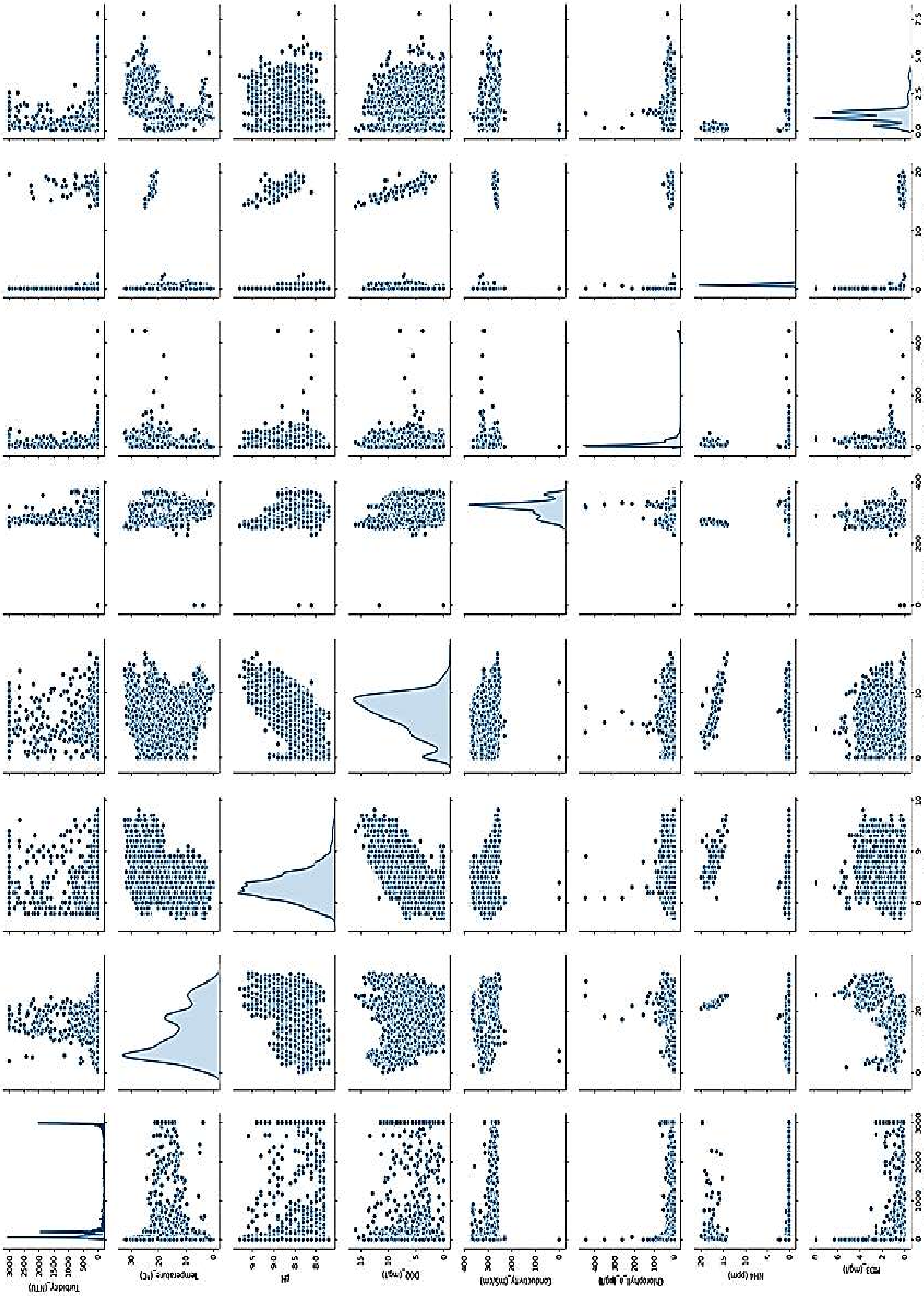

Figure 7.3: Joint distribution of pairs of parameters from the training set for Psaradika station

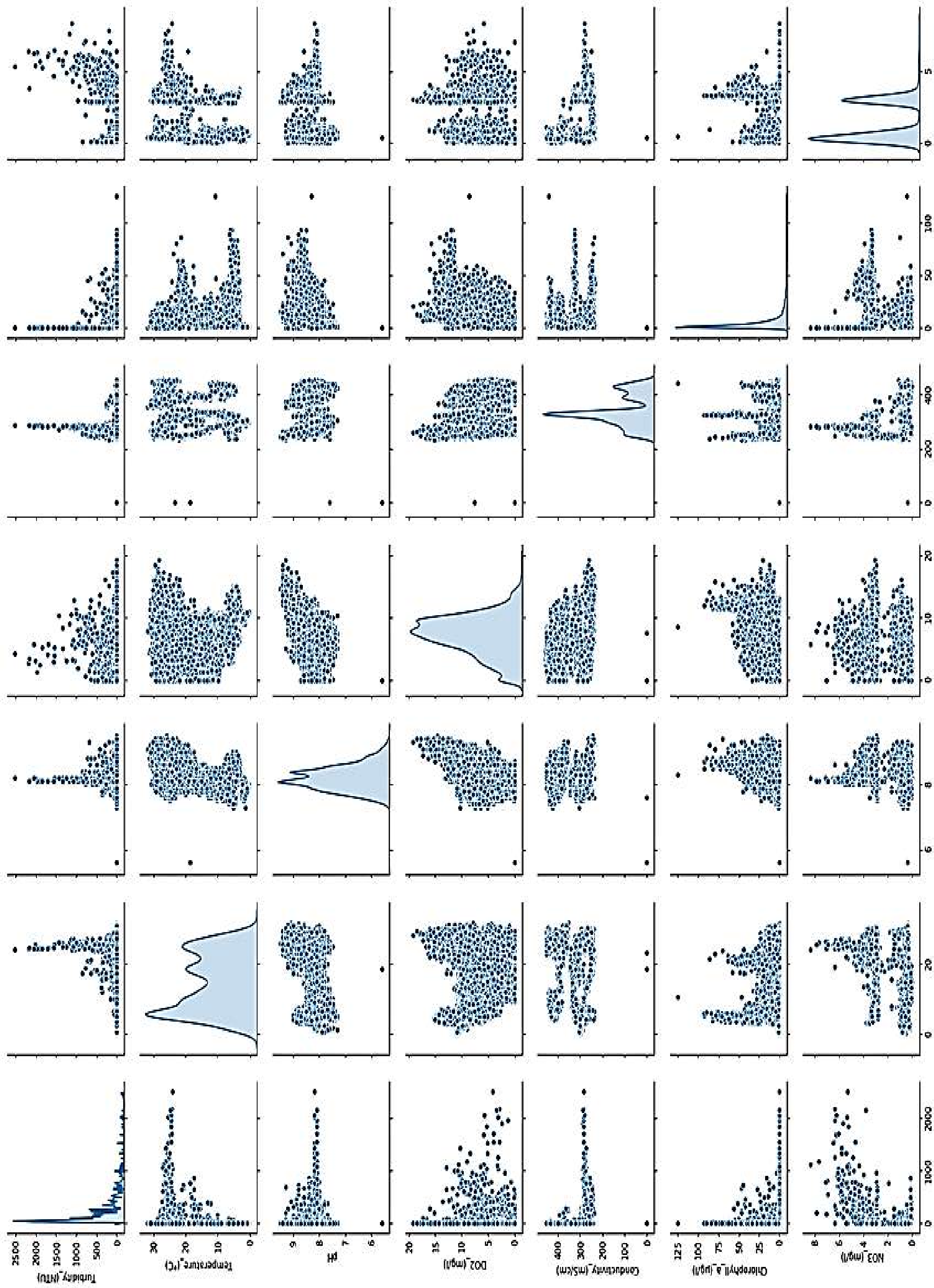

Figure 7.4: Joint distribution of pairs of parameters from the training set for Stavros station

Increased pH values in a lake causes an increase to phytoplankton's population. Subsequently, the Oxygen Demand for respiration processes, causes a decrease in Dissolved Oxygen concentration. According to Wetzel (2001), the increaseof the Nitrate concentration, reduces the Dissolved Oxygen, because it will also increase the population of phytoplankton, that requires DO for the respiratory process. In addition, the decomposition process of phytoplankton, both aerobic and anaerobic, causes a decrease in DO. Therefore, it can be concluded that the low DO content in Lake Kastoria, is caused by the change of pH (from low – acidic to high – alkaline conditions) and by the high concentration of nutrient levels of nitrates, which increase the population of phytoplankton, that requires more DO oxygen for their growth, respiration and decomposition of dead phytoplankton.

The Lake of Kastoria is characterized as a eutrophic lake, mainly during the summer months and early autumn, as expected. This is further aggravated by the fact that Lake Kastoria is a shallow urban lake. The fact that the state of Lake Kastoria is eutrophic, does not mean that the water quality status is poor for its use and the water quality inadequate. In general, the good or bad quality of water in an aquatic ecosystem, should not be linked directly to whether the ecosystem is eutrophic or not (Kagalou and Psilovikos, 2013).

In terms of assessment of lake Kastoria's state, it is understood that in an aquatic ecosystem, even lower values than the maximum permitted values, for uses such as drinking water and aquatic life, are able to cause eutrophic phenomena. Most months of the year, the lake is classified as eutrophic-hypereutrophic. Lake Kastoria, as a eutrophic lake, is characterized by high primary activity, high nutrient content, the ability to support growth of aquatic flora, abundance of macrophytes and phytoplankton as indicators of clean and murky water respectively.

The intense photosynthetic processes, taking place in Lake Kastoria, result in high values of dissolved oxygen even higher than those of oxygen saturation. This phenomenon occurs from mid-spring to early summer and is beneficial to aquatic organisms.High values of DO are also associated with correspondingly high concentrations of Chl-a, without the adverse effects of eutrophication yet commencing (Matzafleri et al., 2009; Psilovikos, 2014; Psilovikos, 2020).In the evening hours, respiration is favored, where due to eutrophication and the intense needs of the organisms, oxygen is consumed, thus causing stress to the aquatic organism when it falls to low levels. Daily fluctuations that occur, require continuous monitoring of dissolved oxygen on a daily basis. Lake Kastoria presents all the above mentioned characteristics as a eutrophic lake.

From time to time, however, lake Kastoria is characterized as hypertrophic. Hypertrophic lakes are characterized by high concentrations of nutrients and algal blooms phenomena. Moreover, they exhibit strong odor effects. In addition, high chlorophyll values (> 40 μg/L) and low dissolved oxygen valuesapproaching anoxia and creating dead zones, mainly during the evening hours, occur. Also, hypertrophic lakes are characterized by the presence of toxic gases such as H_2S, NH_3 and CH_4.

7.2 Structure: 7-64-64-1

Structure 7-64-64-1 indicates that each investigated Deep Neural Network is composed of one input layer with seven input variables, two hidden layer with 64 neurons each, and one output layer with one output variable. In case of "Stavros" station the investigated structure is 6-64-64-1, as NH_4parameter is not available for this station. It should be noted here that for Gkiole, Toichio and Psaradika station, the structure 6-64-64-1 is also tested. The differences between 7-64-64-1 and 6-64-64-1 structures are considered negligible, and the models with structure 7-64-64-1 are prevailed in order to preserve in the input layer all the available information.

7.2.1 Gkiole Station

7.2.1.1 DO model

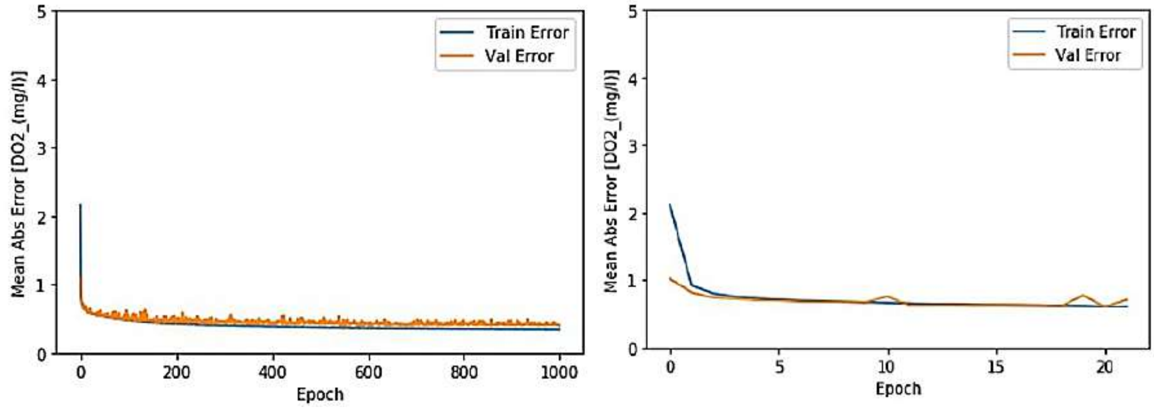

Figure 7.5: Mean Absolute Error (DO model, Gkiole station, Structure 7-64-64-1)

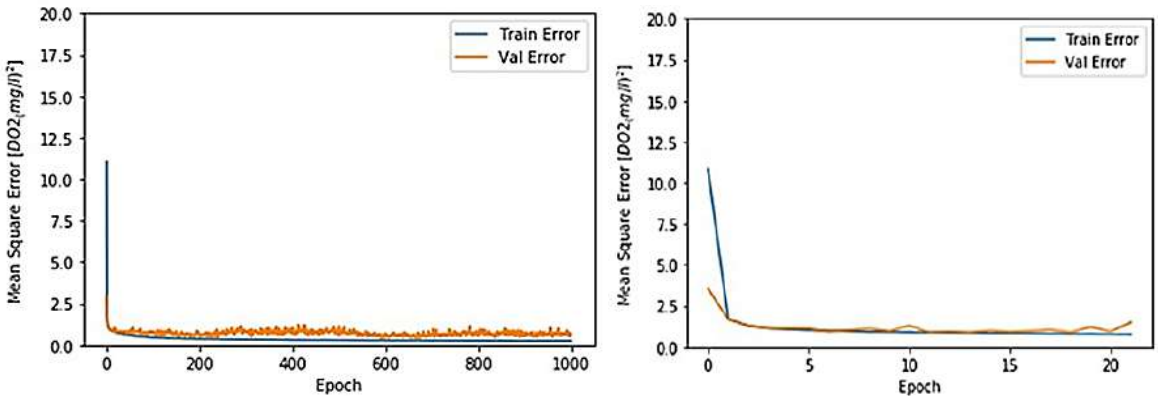

Figure 7.6: Mean Square Error (DO model, Gkiole station, Structure 7-64-64-1)

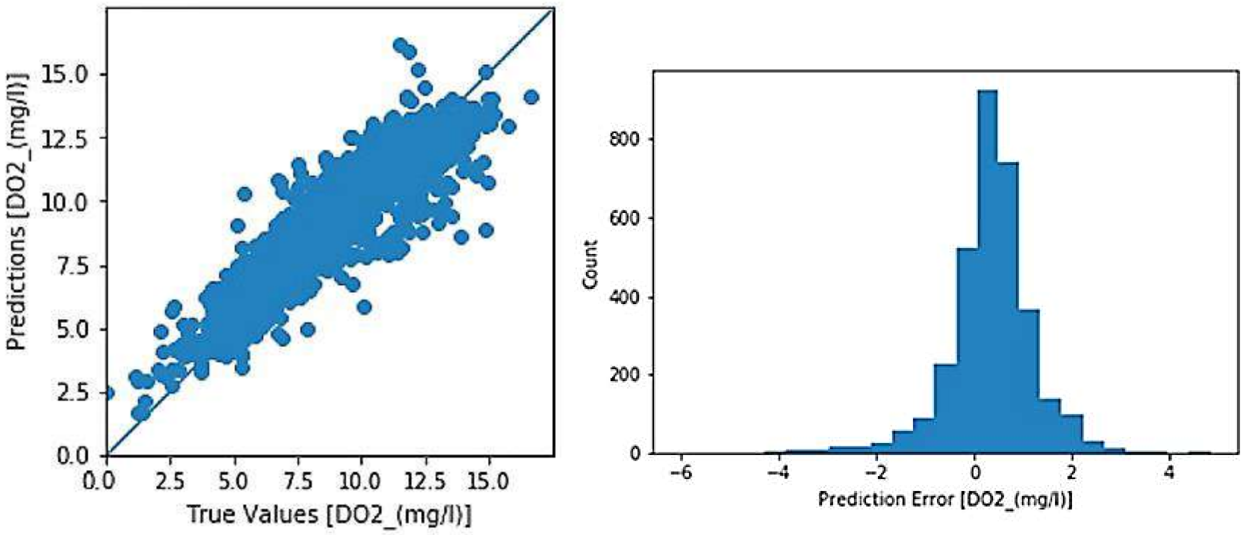

Figure 7.7: a) True vs Predicted values (left), b) Predicted Error (right) (DO model, Gkiole station, Structure 7-64-64-1)

Figures 7.5 and 7.6, illustrate the Mean Absolute Error and the Mean Square Error of DO model for Gkiole station, during the training process of 1000 epochs. The model seems to "learn" from the dataset after the 20th epoch. The Training Mean Absolute Error equals to 0,69, while the Training Mean Square Error equals to 0,89. In order to examine how well the model generalizes after the training procedure, the test set is used. The resulting values of Mean Absolute Error and Mean Square Error equal to 0,71 and 0,92 respectively. The fact thatthe obtained training errors were slightly lower than the tested ones, indicates that over-fitting is avoided. Figure 7.7a shows the true vs the predicted values achieved, after the training procedure, showing how well one can expect the model to predict when it is used in the real world. It looks like the DO model with structure 7-64-64-1, predicts reasonably well. Finally, Figure 7.7b illustrates the prediction error distribution.

7.2.1.2 Turbidity model

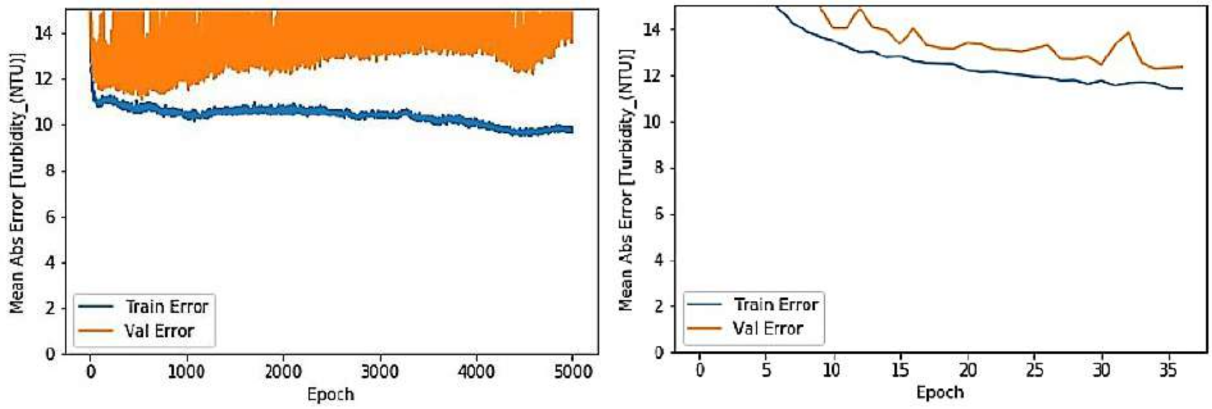

Figure 7.8: Mean Absolute Error (Turbidity model, Gkiole station, Structure 7-64-64-1)

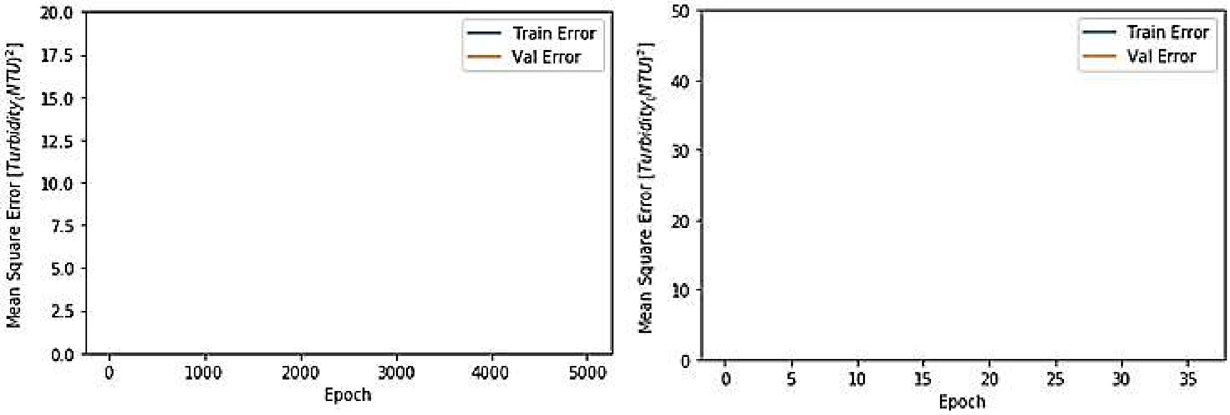

Figure 7.9: Mean Square Error (Turbidity model, Gkiole station, Structure 7-64-64-1)

For Turbidity model with structure 7-64-64-1, the training procedure takes place for 1000 epochs. As convergence could notbe achieved, the number of epochs is increased to 5000, but without any arithmetic improvement (Figures 7.8 and 7.9). The Testing set Mean Absolute Error equals to 10,53, while the Testing Mean Square Error equals to 778,88, indicating that the Turb modeldoes not show a good predictive capacity.

7.2.1.3 Chl-amodel

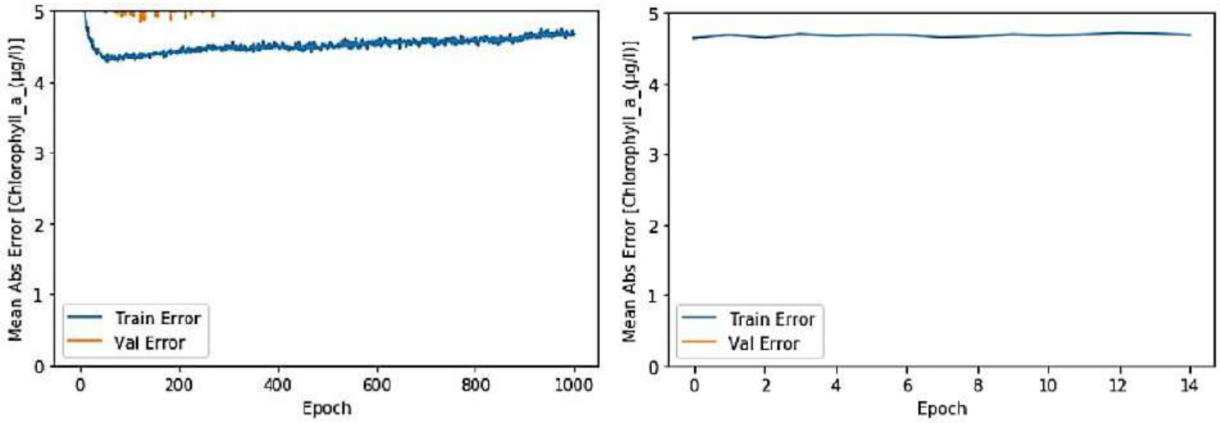

Figure 7.10: Mean Absolute Error (Chl-a model, Gkiole station, Structure 7-64-64-1)

Regarding Chl-a model with structure 7-64-64-1, the training process takes place for 1000 epochs (Figure 7.10). Convergence could notbe achieved, so the number of epochs is increased to 5000, but even in this case no decrease in the validationerror is achieved. The Testing set Mean Absolute Error equals to 6,34, while the Testing Mean Square Error equals to 375,95, indicating that Chl-a modeldoes not have any predictive ability.

7.2.2 Toichio Station

7.2.2.1 DO model

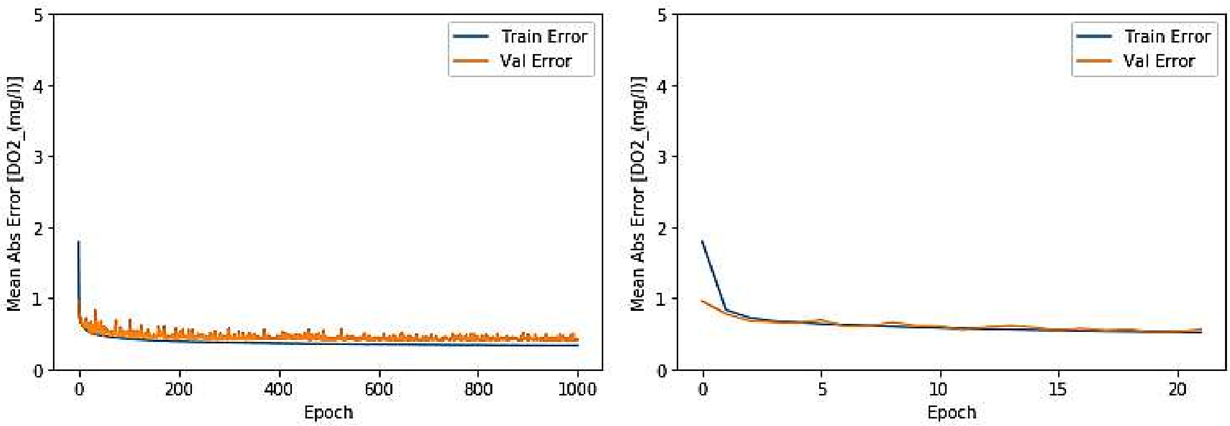

Figure 7.11: Mean Absolute Error (DO model, Toichio station, Structure 7-64-64-1)

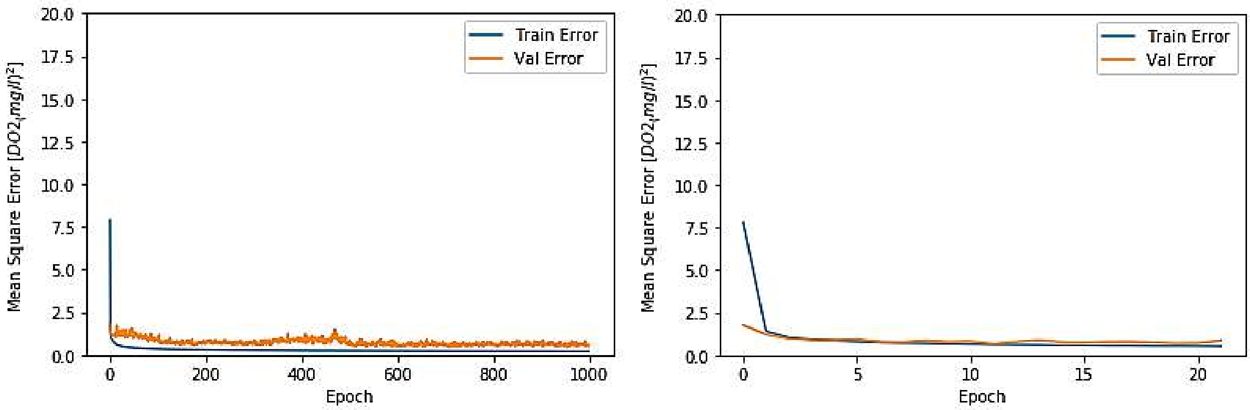

Figure 7.12: Mean Square Error (DO model, Toichio station, Structure 7-64-64-1)

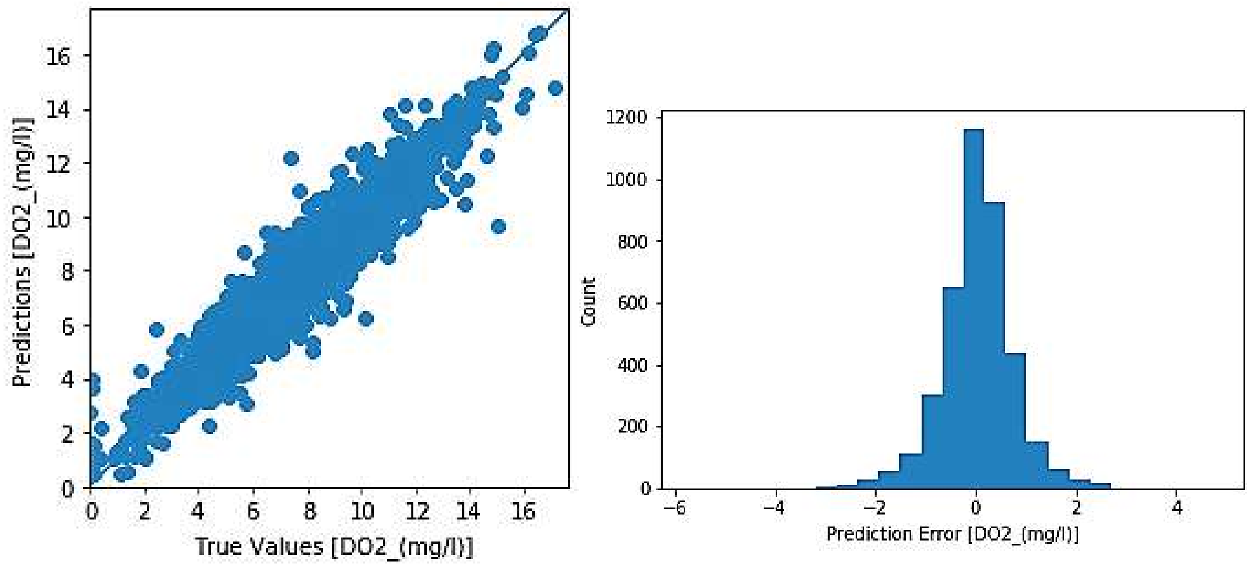

Figure 7.13: a) True vs Predicted values (left), b) Predicted Error (right)(DO model, Toichio station, Structure 7-64-64-1)

Figures 7.11 and 7.12 illustrate the Mean Absolute Error and the Mean Square Error of DO model for Toichio station during the training process of 1000 epochs. The model seems to "learn" from the dataset after the 20th epoch. The Training Mean Absolute Error equals to 0,48, while the Training Mean Square Error equals to 0,52. Moreover, the resulting values of Mean Absolute Error and Mean Square Errorfor test set, equal to 0,49 and 0,51, respectively. The fact thatthe Obtained Training Errors were slightly lower than the Tested Ones, indicates that a good fit is obtained. Figure 7.13a shows that DO model with structure 7-64-64-1 for Toichio station predicts very well. Finally, Figure 7.7b illustrates the prediction error distribution.

7.2.2.2 Turbidity model

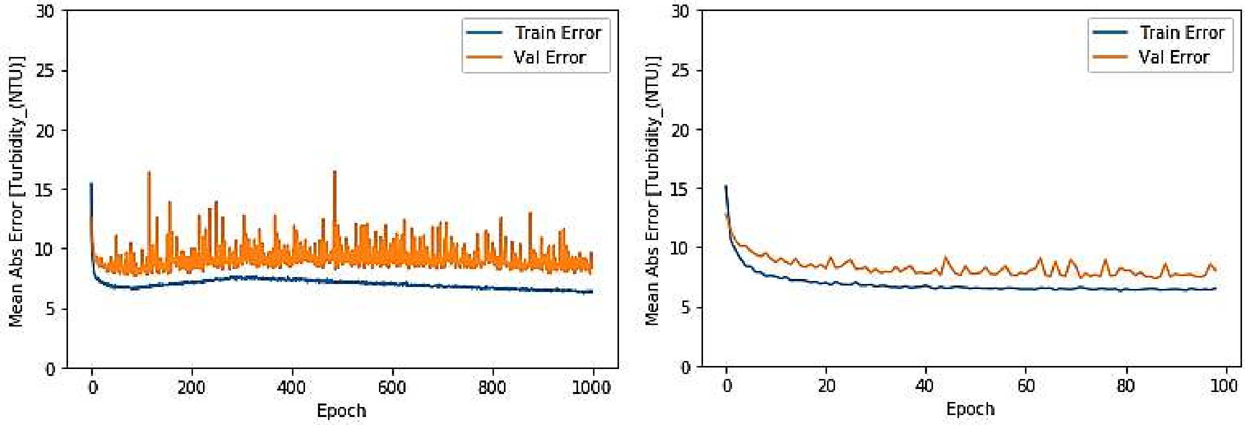

Figure 7.14: Mean Absolute Error (Turbidity model, Toichio station, Structure 7-64-64-1)

Turbidity modelhas a bad predictive ability asconvergence could notbe achieved (Figure 7.14). The Testing set Mean Absolute Error equals to 6,90, while the Testing Mean Square Error equals to 1121,15.

7.2.2.3 Chl-a model

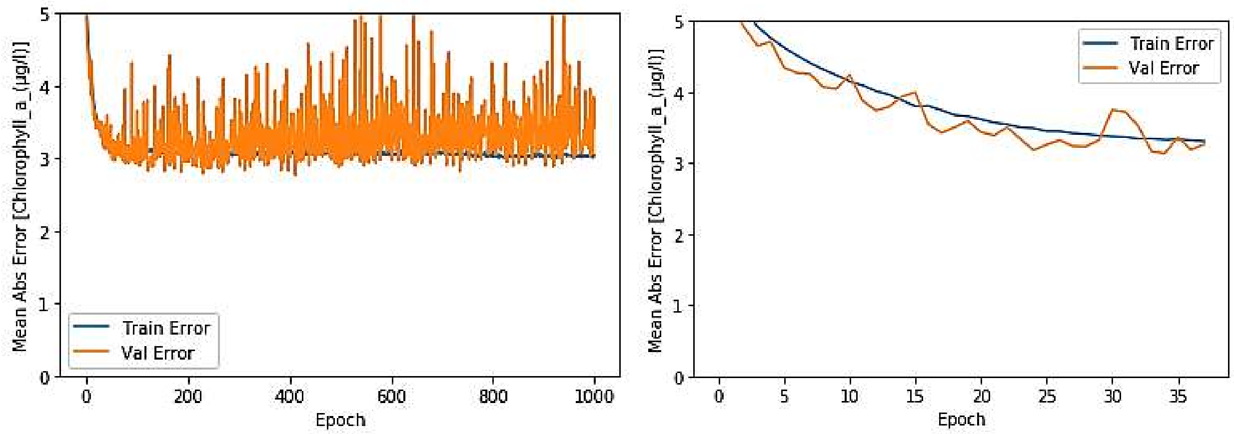

Figure 7.15: Mean Absolute Error (Chl-a model, Toichio station, Structure 7-64-64-1)

For Chl-a model with structure 7-64-64-1 (Toichio station), the training process takes place for 1000 epochs (Figure 7.15). No convergence is achieved for the aforementioned structure. The Testing set Mean Absolute Error equals to 3,39, while the Testing Mean Square Error equals to 66,12; showing no predictive capacity.

7.2.3 Psaradika Station

7.2.3.1 DO model

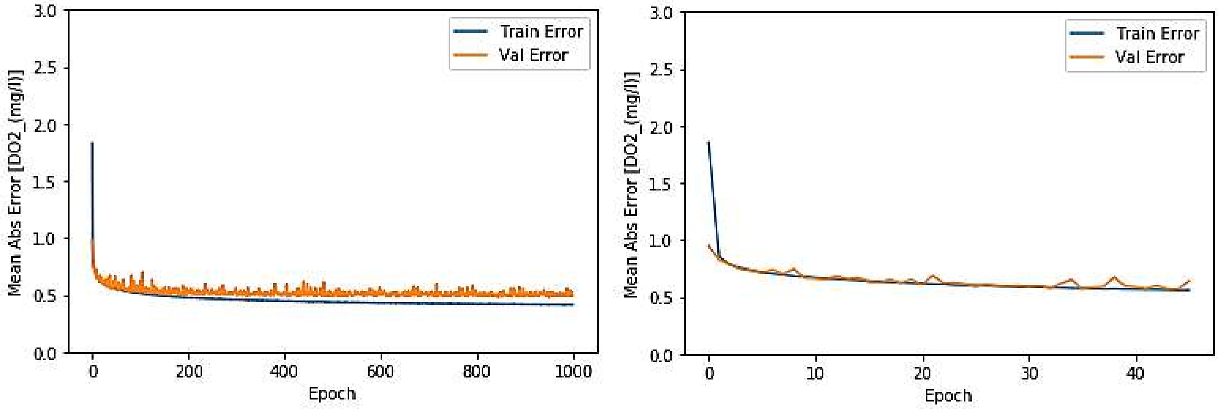

Figure 7.16: Mean Absolute Error (DOmodel, Psaradika station, Structure 7-64-64-1)

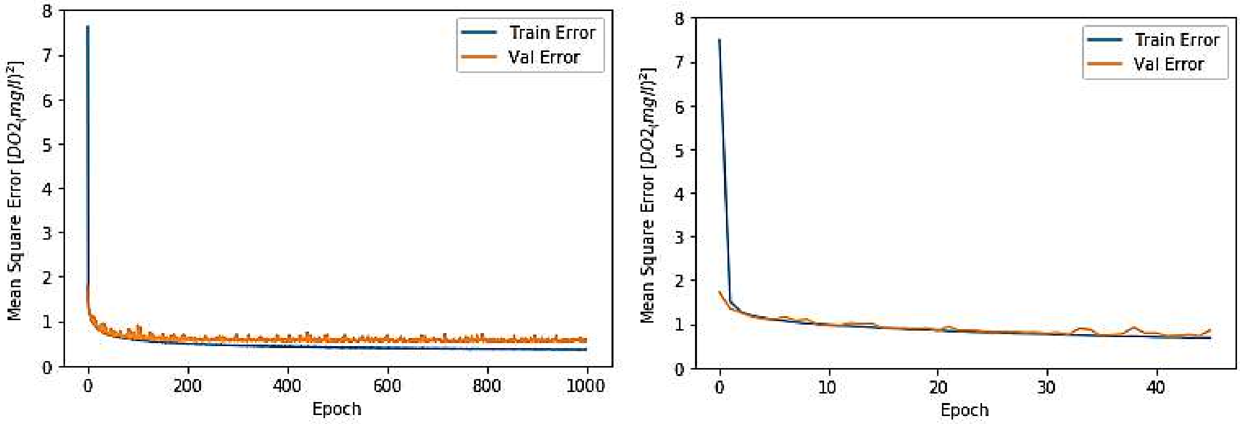

Figure 7.17: Mean Square Error (DOmodel, Psaradika station, Structure 7-64-64-1)

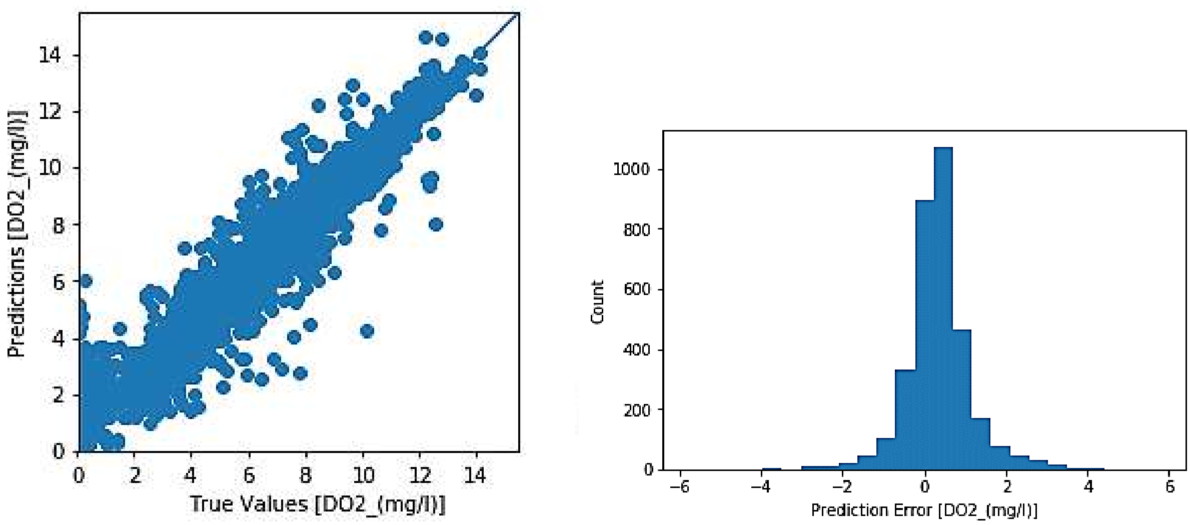

Figure 7.18: a) True vs Predicted values (left), b) Predicted Error (right) (DOmodel, Psaradika station, Structure 7-64-64-1)

Figures 7.16 and 7.17show the Mean Absolute Error and the Mean Square Error of DO model for Psaradika station, during the training process of 1000 epochs. The model seems to "learn" from the dataset after the 40th epoch, fact that also is confirmed manually, from the obtained sheets of Training and Validation statistical results, for each case. The Training Mean Absolute Error equals to 0,60, while the Training Mean Square Error equals to 0,77. The test set is used to examine the predictive capacity of the investigated model. The values obtained of Mean Absolute Error and Mean Square Error equal to 0,63 and 0,80 respectively, underlying that over-fitting is avoided. DO model with structure 7-64-64-1 for Psaradika station predicts reasonably well (Figure 7.18a). Finally, Figure 7.18b shows the prediction error distribution.

7.2.3.2 Turbidity model

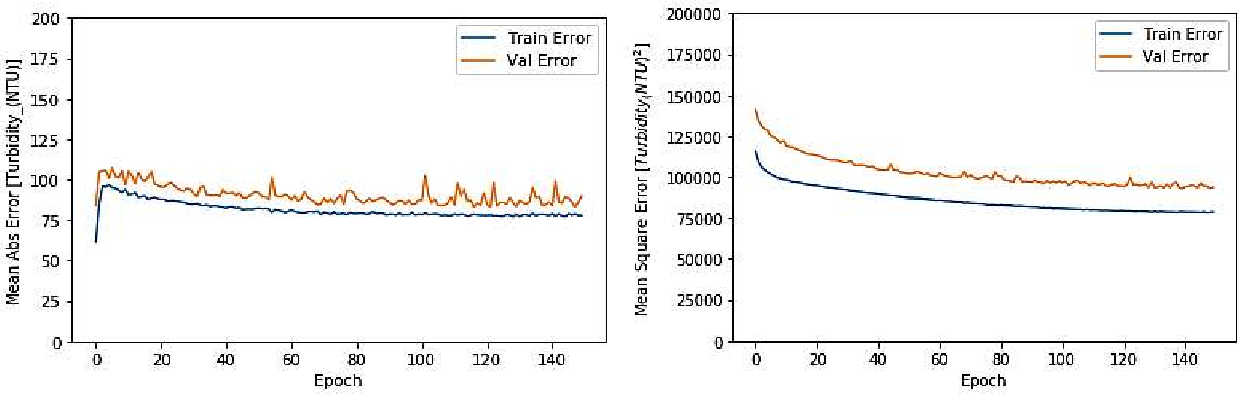

Figure 7.19: Mean Absolute Error (Turbidity model, Psaradika station, Structure 7-64-64-1)

Figure 7.19 shows that in case of Turbidity model for Psaradika station, training and validation phase could not converge even in case that epochs reach 5000.The Testing Mean Absolute Error equals to 84,73, while the Testing Mean Square Error equals to 92594,44.

7.2.3.3 Chl-a model

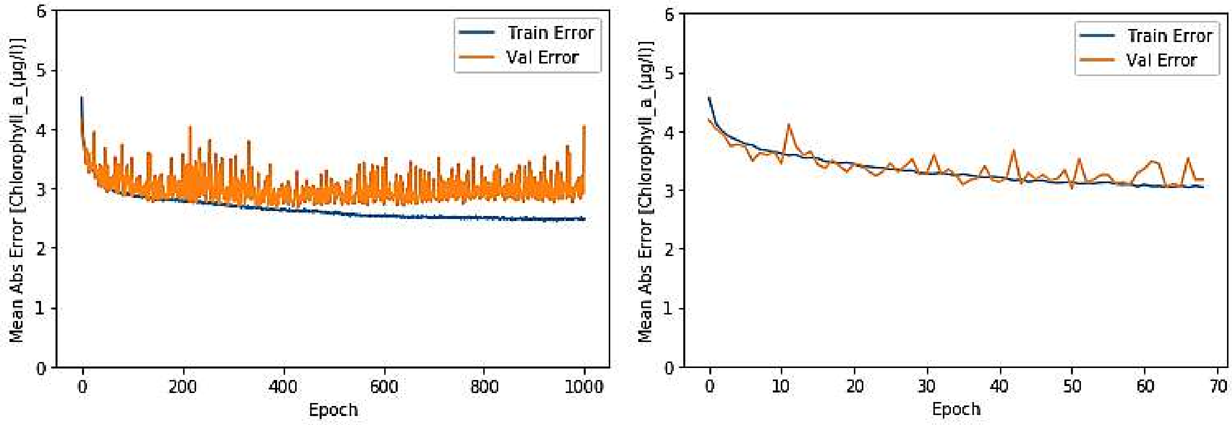

Figure 7.20: Mean Absolute Error (Chl-a model, Psaradika station, Structure 7-64-64-1)

For Chl-a model (structure 7-64-64-1 – Psaradika station), the Testing Mean Absolute Error equals to 3,23, while the Testing Mean Square Error equals to 127,32. As the number of epochs is increased, the difference between validation error values and train error values is also increased. No convergence is achieved for Chl-a model, with structure 7-64-64-1 for Psaradika station (Figure 7.20).

7.2.4 Stavros Station

7.2.4.1 DO model

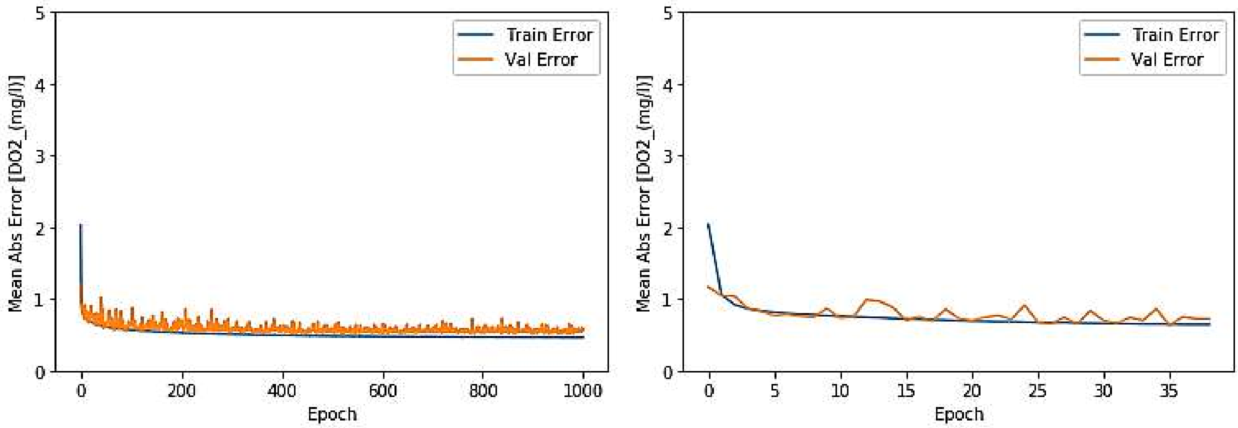

Figure 7.21: Mean Absolute Error (DO model, Stavros station, Structure 6-64-64-1)

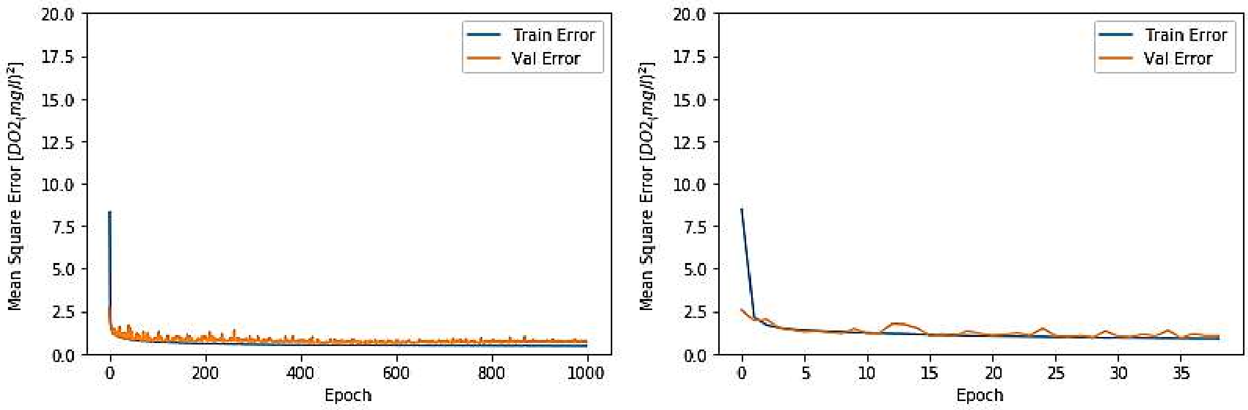

Figure 7.22: Mean Square Error (DOmodel, Stavros station, Structure 6-64-64-1)

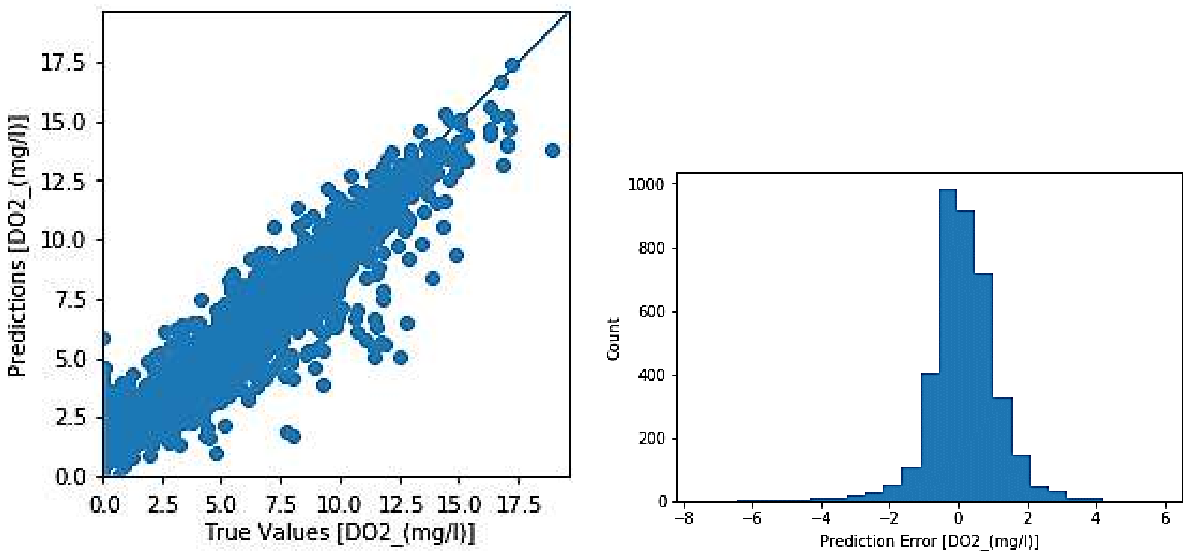

Figure 7.23: a) True vs Predicted values (left), b) Predicted Error (right)(DOmodel, Stavros station, Structure 6-64-64-1)

Figures 7.21 and 7.22, show the training process during 1000epochs of DO model, with structure 6-64-64-1 for Stavros station. The convergence is achieved and themodel seems to understand the dataset after the 35th epoch. The Training Mean Absolute Error equals to 0,73, while the Training Mean Square Error equals to 1,02. The values obtained of Mean Absolute Error and Mean Square Error for test process equal to 0,74 and 1,08 respectively. DO model with structure 6-64-64-1 for Stavros station, predicts reasonably well (Figure 7.23a). Finally, Figure 7.23b shows the prediction error distribution.

7.2.4.2 Turbidity model

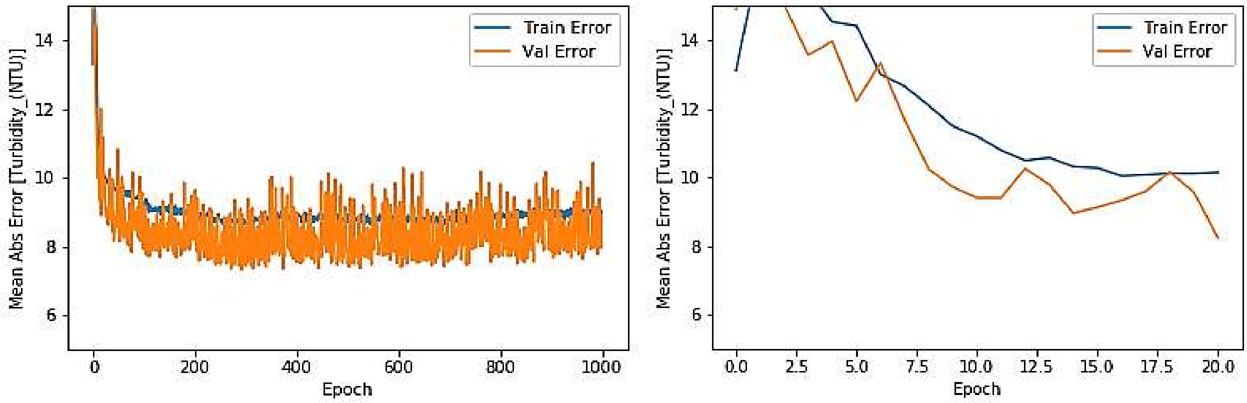

Figure 7.24: Mean Absolute Error (Turbiditymodel, Stavros station, Structure 6-64-64-1)

Turbidity model for Stavros station, could not converge during the training process (Figure 7.24). The Testing Mean Absolute Error equals to 7,32, while the Testing Mean Square Error equals to 1431,49, indicating that the model does not have a good performance.

7.2.4.3 Chl-a model

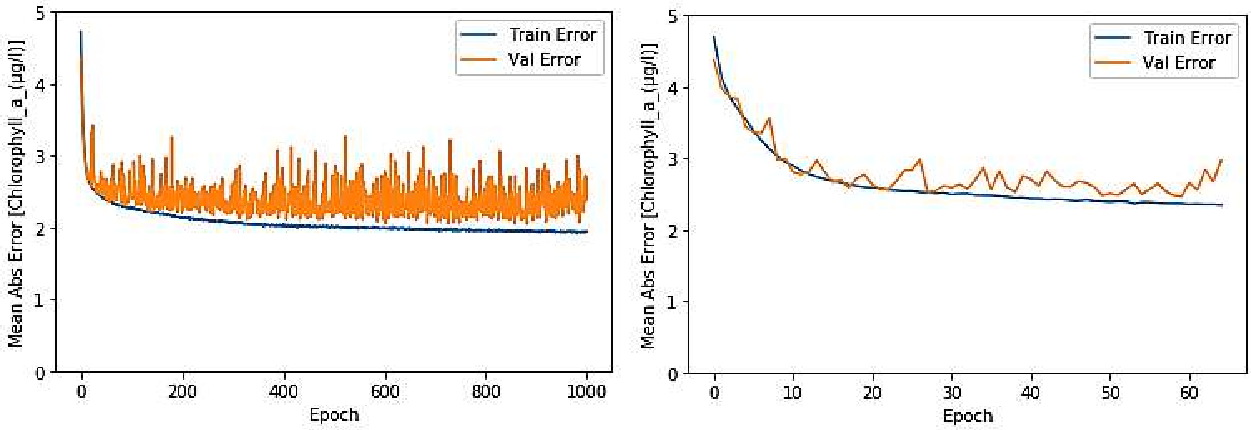

Figure 7.25: Mean Absolute Error (Chl-amodel, Stavros station, Structure 6-64-64-1)

Chl-a model, with structure 6-64-64-1 for Stavros station, show a better performance compared with the aforementioned models of Chl-a, but still the obtained values of Testing Mean Absolute Error (equals to 2,91) and the Testing Mean Square Error (equals to 21,05), are not satisfactory.

7.3 Structure: 4-64-64-1

Structure 4-64-64-1, indicates that each investigated Deep Neural Network is composed of one input layer with four input variables, two hidden layer with 64 neurons each, and one output layer with one output variable.

7.3.1 Gkiole Station

7.3.1.1 DO model

According to Figures 7.26 and 2.27, DO model with structure 4-64-64-1 seems to achieve its training procedure, from the 60th epoch and on. The Training Mean Absolute Error equals to 0,58, while the Training Mean Square Error equals to 0,83. In order to examine how well the model generalizes after the training procedure, thetestset is used. The resulting values of Mean Absolute Error and Mean Square Error, equal to 0,61 and 0,85 respectively. The fact that, the obtained training errors were slightly lower than the tested ones, indicates that over-fitting is avoided. Figure 7.28a, illustrates the true (recorded) versus the predicted values obtained. It looks like the DO model, with structure 4-64-64-1 for Gkiole station, has a good predictive capacity. Finally, Figure 7.28b shows the prediction error distribution.

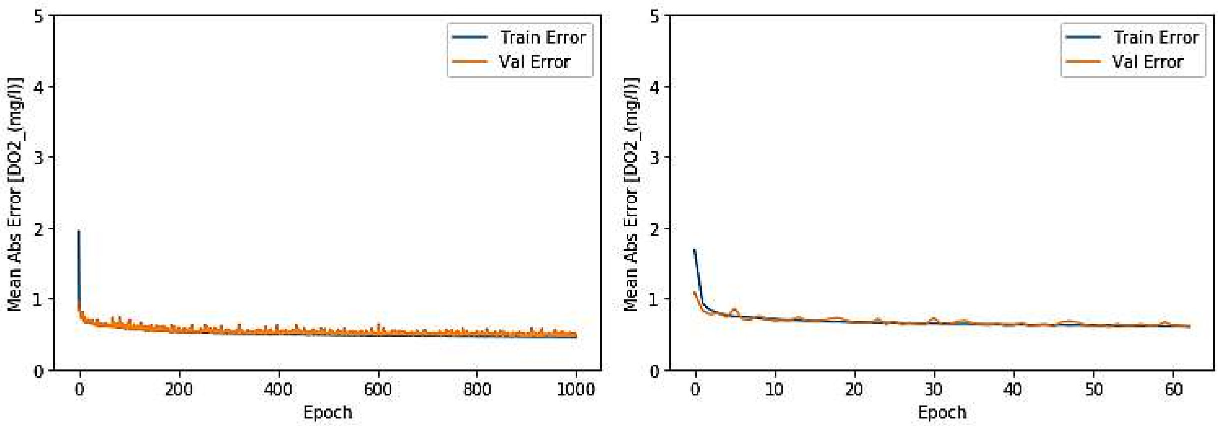

Figure 7.26: Mean Absolute Error (DO model, Gkiole station, Structure 4-64-64-1)

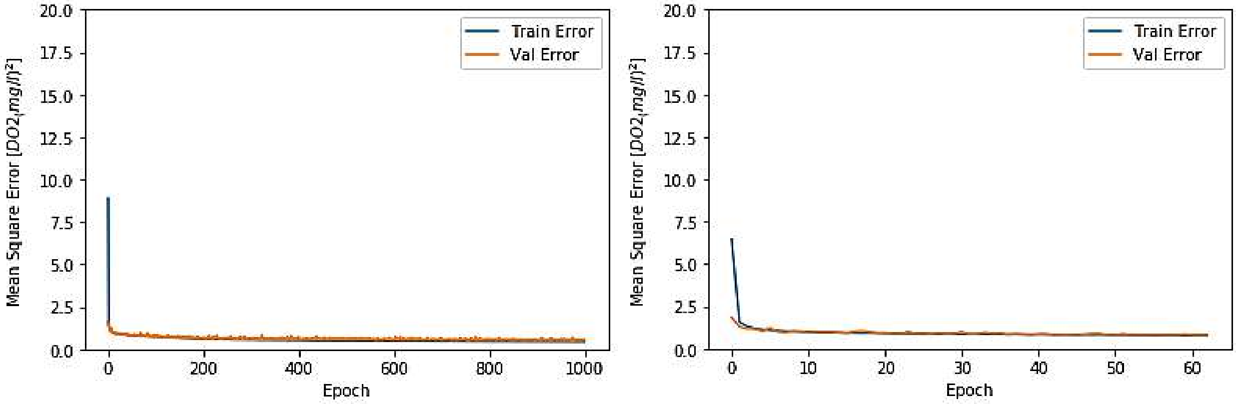

Figure 7.27: Mean Square Error (DO model, Gkiole station, Structure 4-64-64-1)

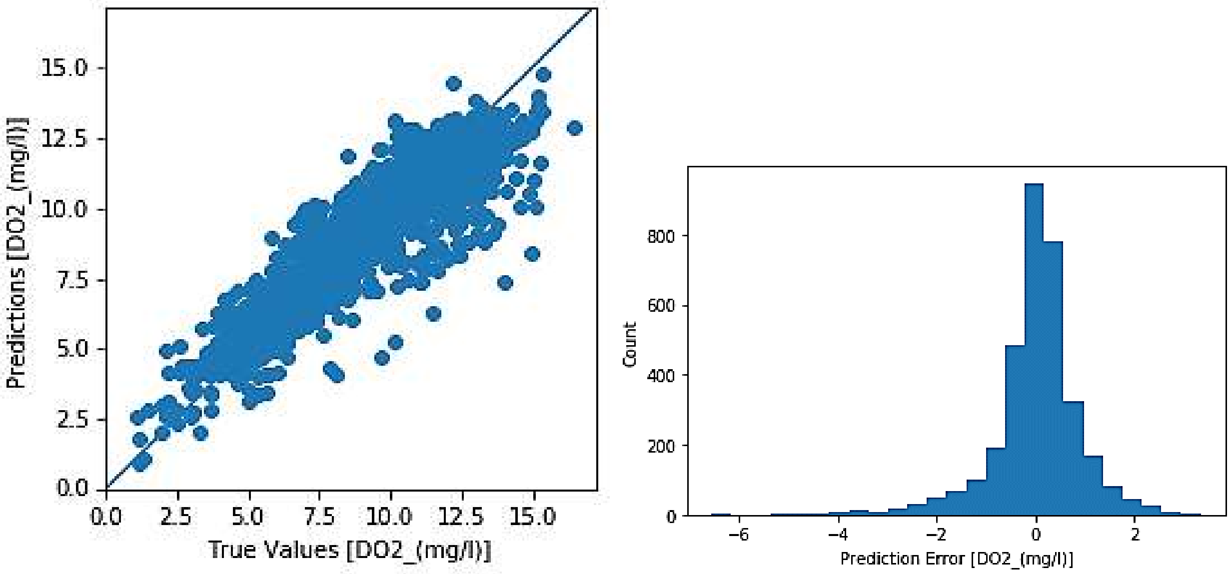

Figure 7.28: a) True vs Predicted values (left), b) Predicted Error (right)(DO model, Gkiole station, Structure 4-64-64-1)

7.3.2 ToichioStation

7.3.2.1 DO model

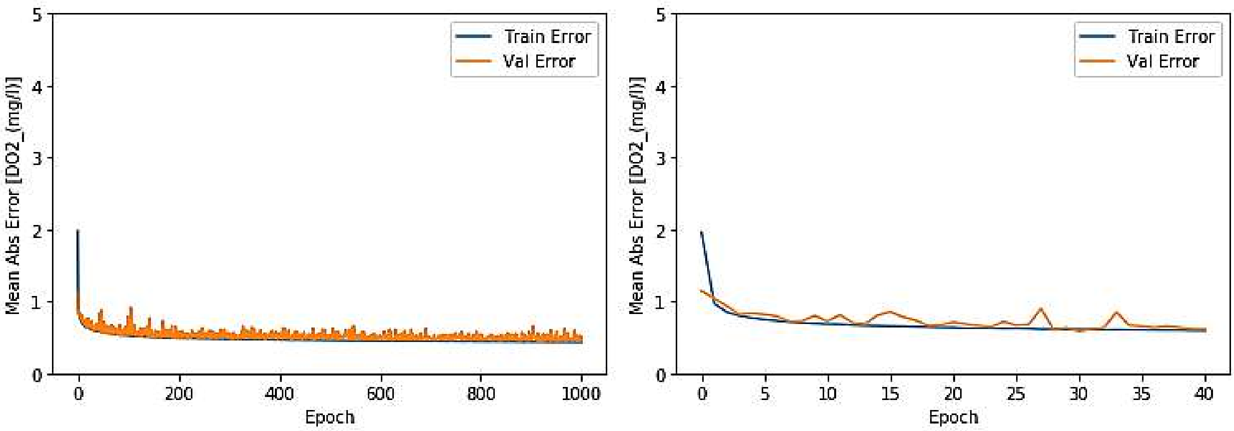

Figure 7.29: Mean Absolute Error (DOmodel, Toichio station, Structure 4-64-64-1)

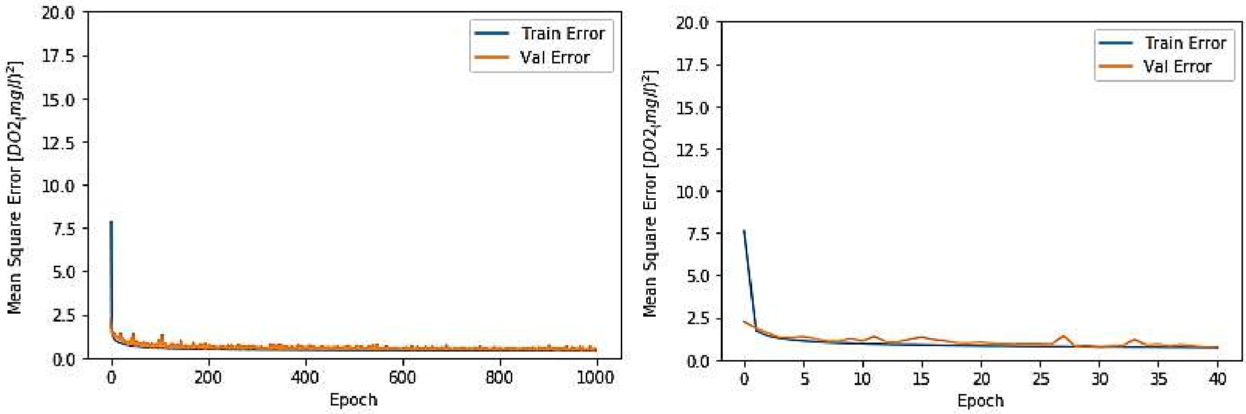

Figure 7.30: Mean Square Error (DOmodel, Toichio station, Structure 4-64-64-1)

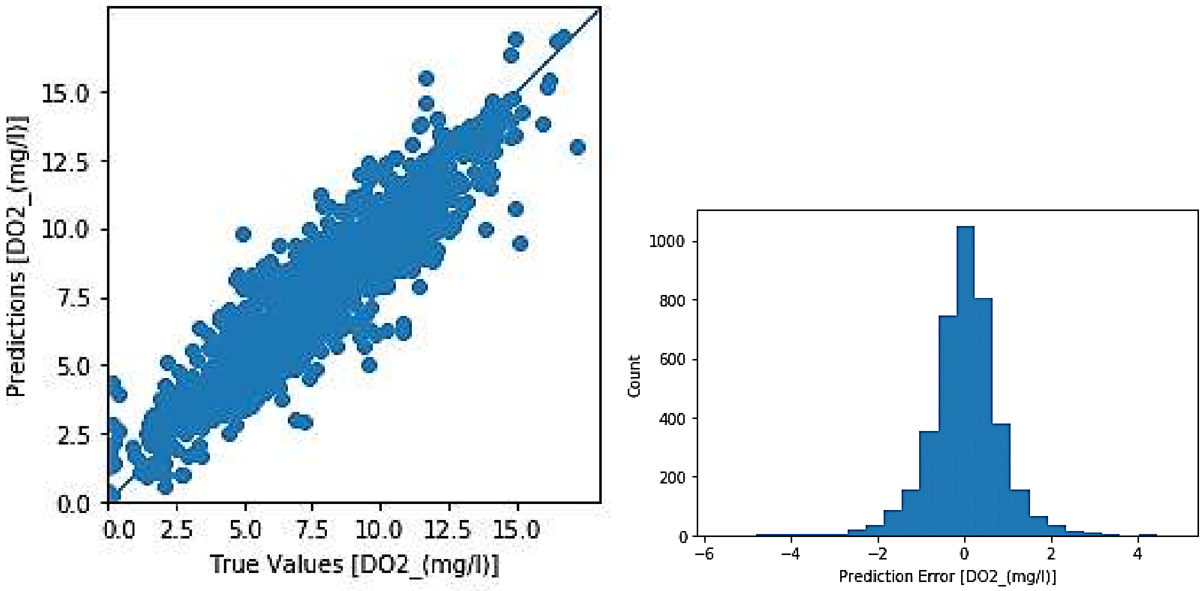

Figure 7.31: a) True vs Predicted values (left), b) Predicted Error (right) (DOmodel, Toichio station, Structure 4-64-64-1)

Figures 7.29 and 7.30 demonstrate the Mean Absolute Error and the Mean Square Error of DO model, with structure 4-64-64-1 for Toichio station, during the training process of 1000 epochs. The model ''understands" the dataset after the 40th epoch. The Training Mean Absolute Error, equals to 0,62, while the Training Mean Square Error equals to 0,70. Moreover, the Testing Mean Absolute Error and Mean Square Error, equals to 0,62 and 0,72 respectively, showing that the DO model predicts reasonably well. Finally, Figure 7.31 illustrates the true values versus the predicted ones and the prediction error distribution.

7.3.3 Psaradika Station

7.3.3.1 DO model

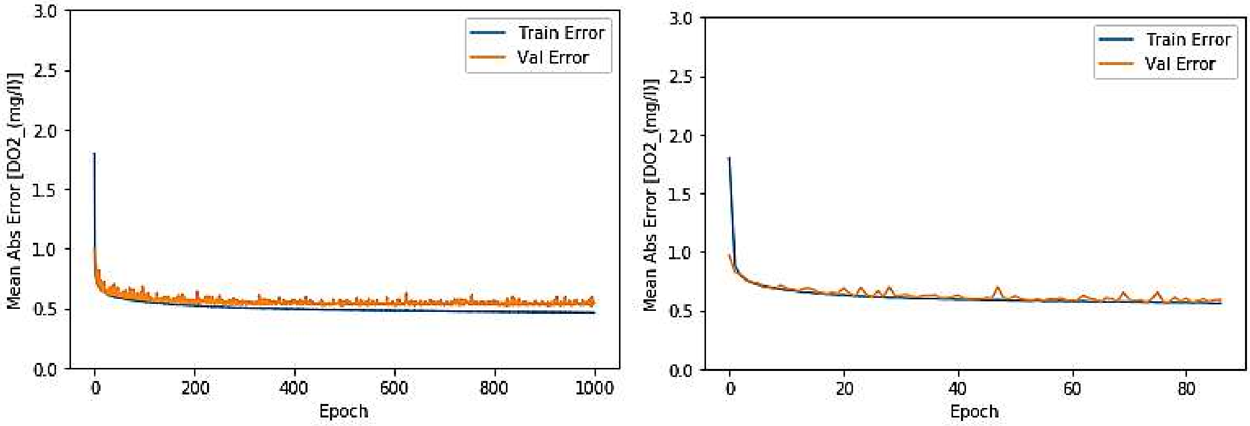

Figure 7.32: Mean Absolute Error (DOmodel, Psaradika station, Structure 4-64-64-1)

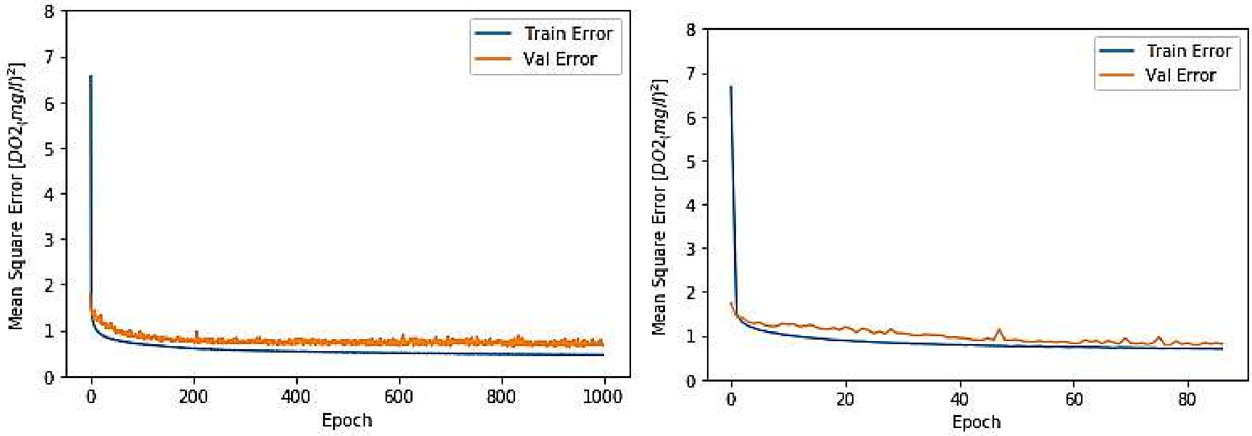

Figure 7.33: Mean Square Error (DOmodel, Psaradika station, Structure 4-64-64-1)

In case of Psaradika station, the DO model shows a good performance (Figures 7.32 and 7.33). The Training Mean Absolute Error equals to 0,57, while the Training Mean Square Error equals to 0,65. Furthermore, Figure 7.34a shows the true (recorded) versus the predicted values, achieved after the training procedure, while Figure 7.34b shows the prediction error distribution.

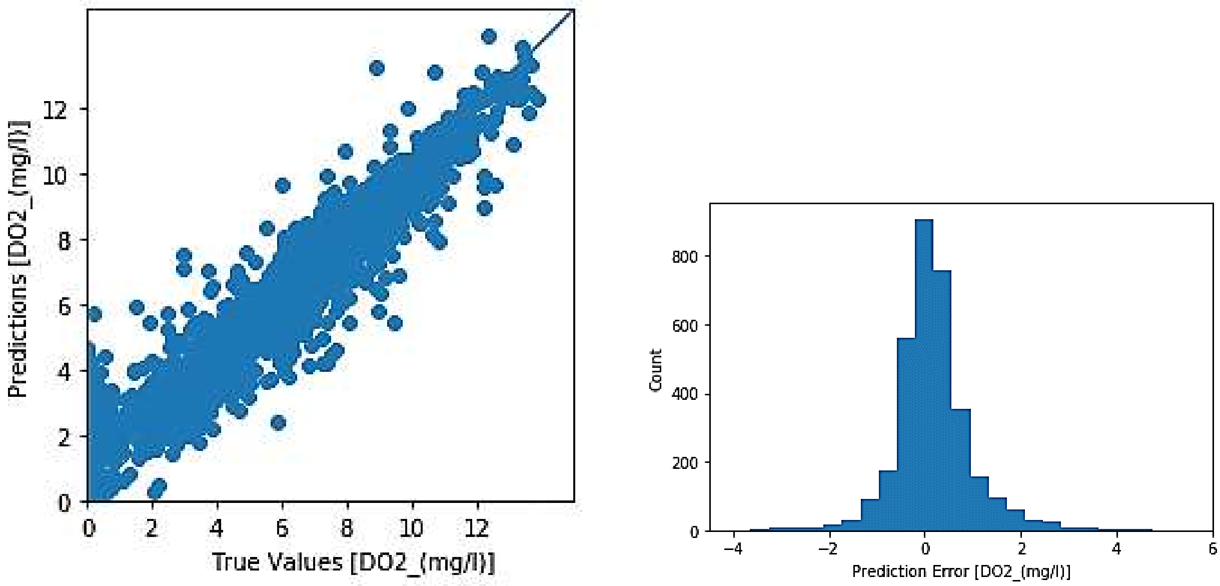

Figure7.34: a) True vs Predicted values (left), b) Predicted Error (right)(DOmodel, Psaradika station, Structure 4-64-64-1)

7.3.4 Stavros Station

7.3.4.1 DO model

Figures 7.35 and 7.36, illustrates that the DO model for Stavros station ‘’learns” the dataset after the 40th epoch. The Training Mean Absolute Error equals to 0,69, while the Training Mean Square Error equals to 0,98. Furthermore, the Testing Mean Absolute Error and Mean Square Error equal to 0,72 and 1,11 respectively, showing that the DO model predicts well. Figure 7.37 illustrates the true values (recorded) versus the predicted ones and the prediction error distribution.

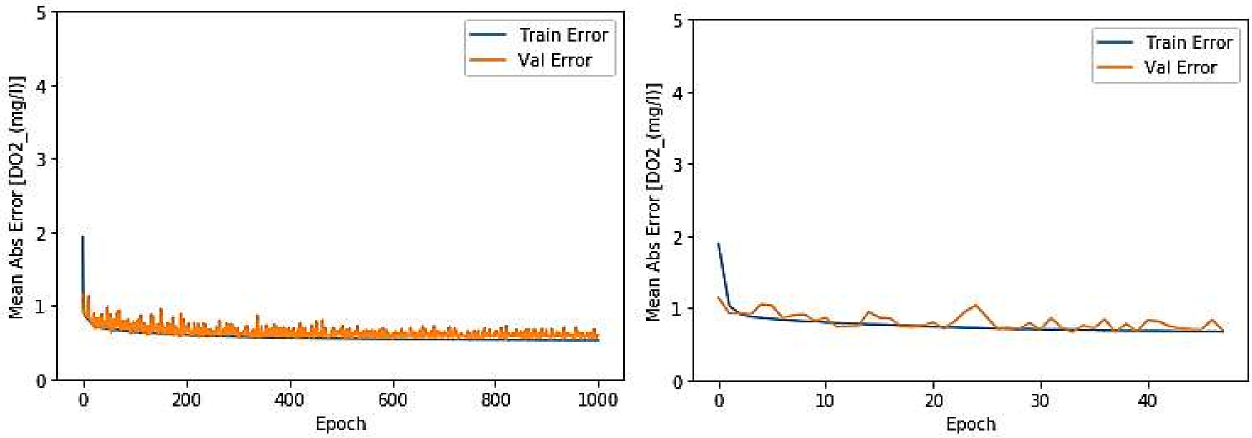

Figure 7.35: Mean Absolute Error (DOmodel, Stavros station, Structure 4-64-64-1)

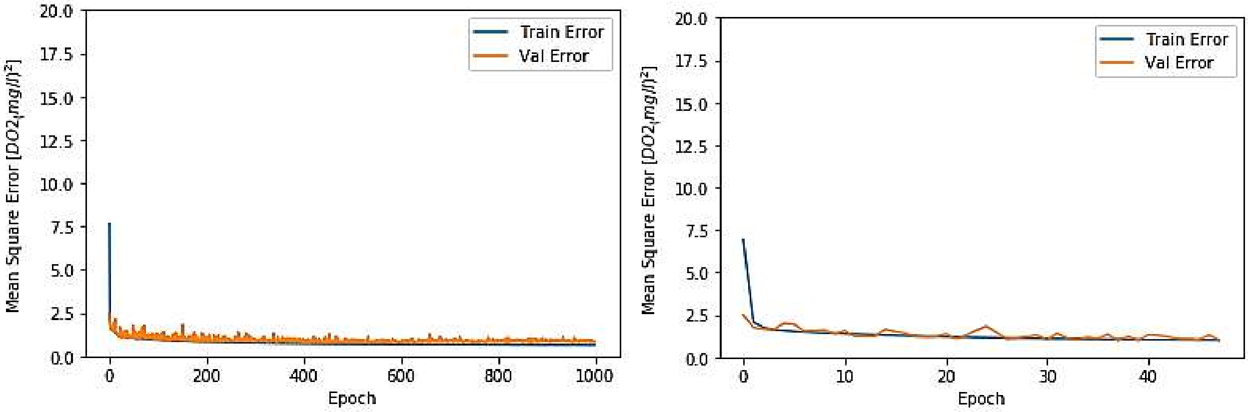

Figure 7.36: Mean Square Error (DOmodel, Stavros station, Structure 4-64-64-1)

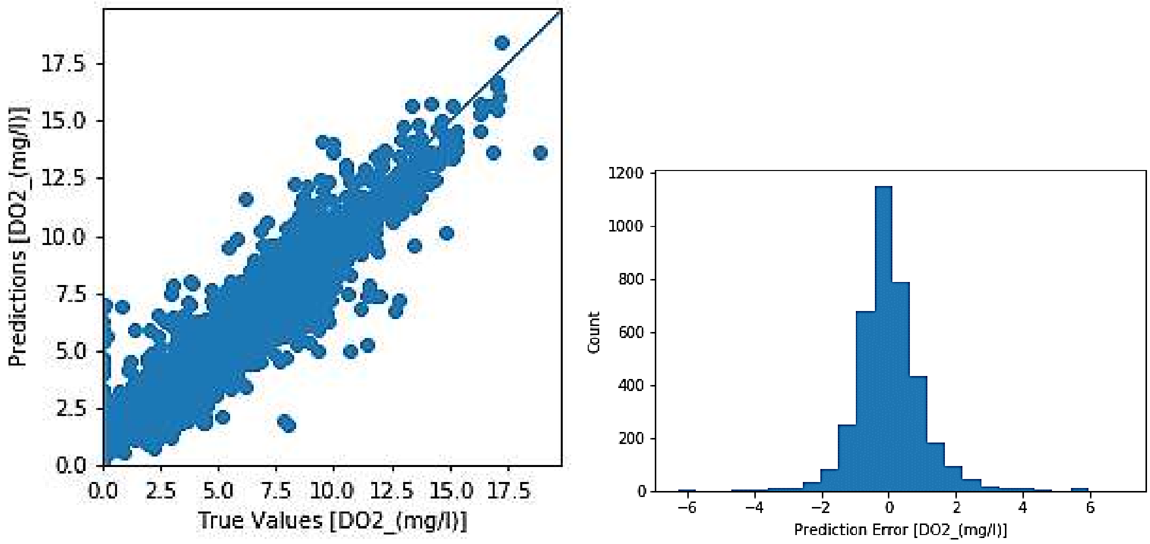

Figure 7.37: a) True vs Predicted values (left), b) Predicted Error (right) (DO model, Stavros station, Structure 4-64-64-1)

7.4 Structure: 7-32-32-1

Structure 7-32-32-1 indicates that each investigated Deep Neural Network, is composed of one input layer with seven input variables, two hidden layers with 32 neurons each, and one output layer with one output variable. In case of “Stavros” station the investigated structure is 6-32-32-1, as NH_4-N parameter is not available for this station.

7.4.1 Gkiole Station

7.4.1.1 DO model

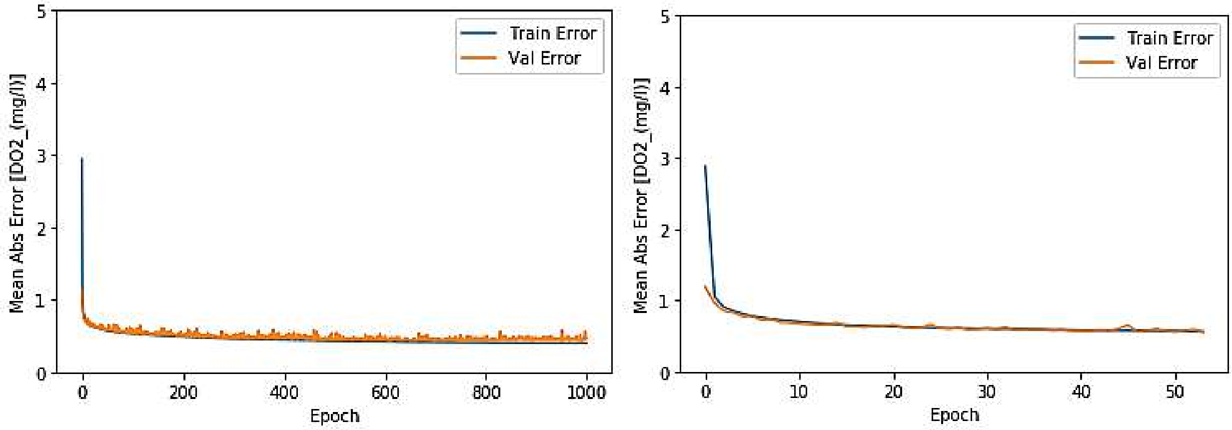

Figure 7.38: Mean Absolute Error (DO model, Gkiole station, Structure 7-32-32-1)

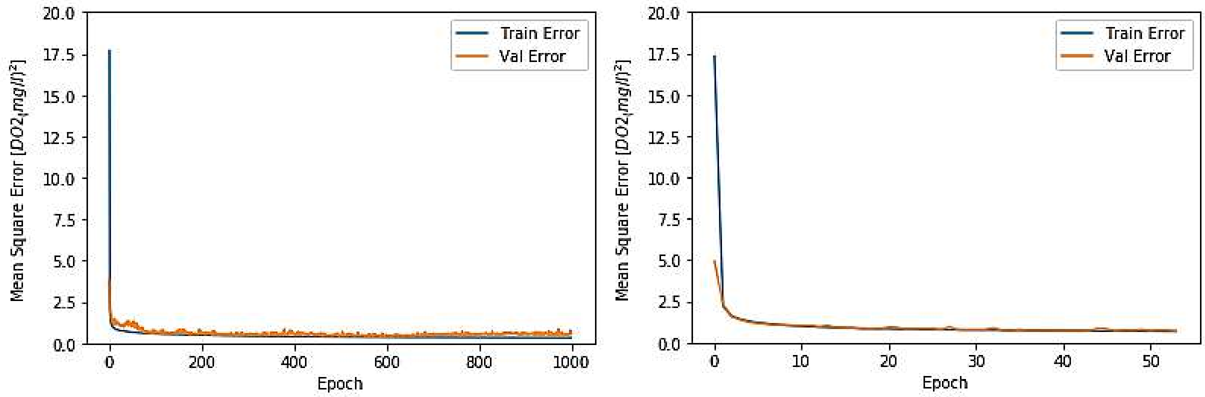

Figure 7.39: Mean Square Error (DO model, Gkiole station, Structure 7-32-32-1)

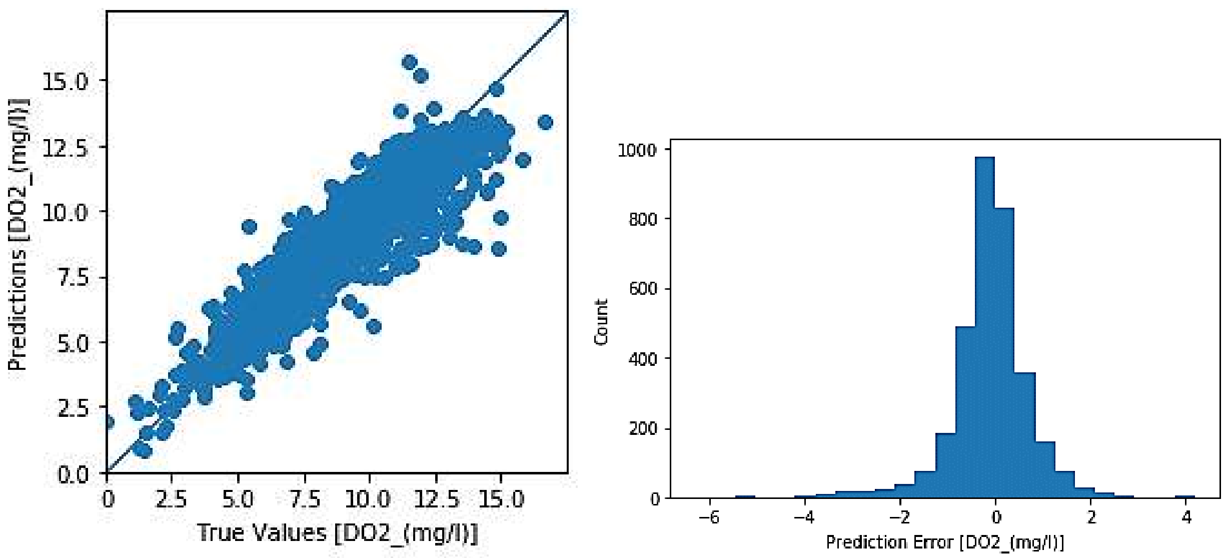

Figure 7.40: a) True vs Predicted values (left), b) Predicted Error (right)(DO model, Gkiole station, Structure 7-32-32-1)

7.4.2 Toichio Station

7.4.2.1 DO model

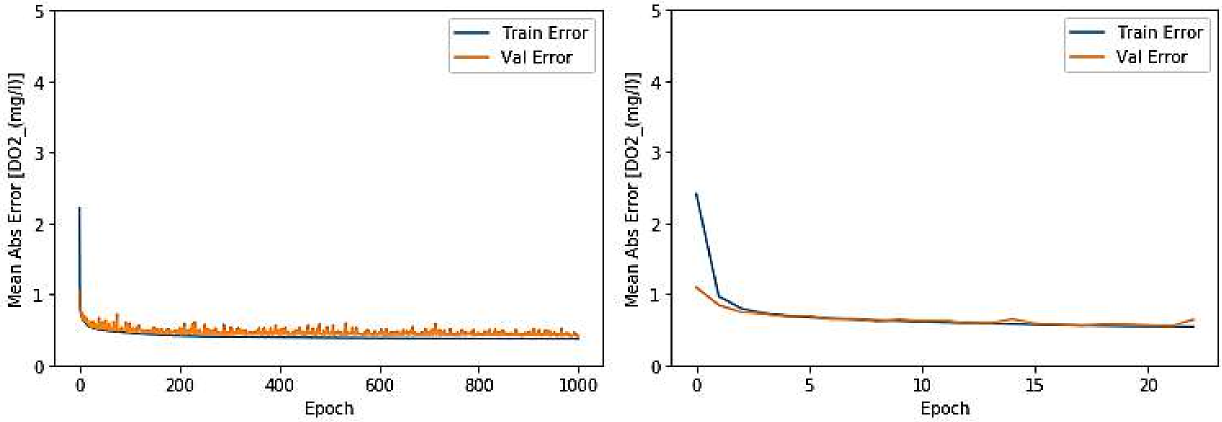

Figure 7.41: Mean Absolute Error (DOmodel, Toichio station, Structure 7-32-32-1)

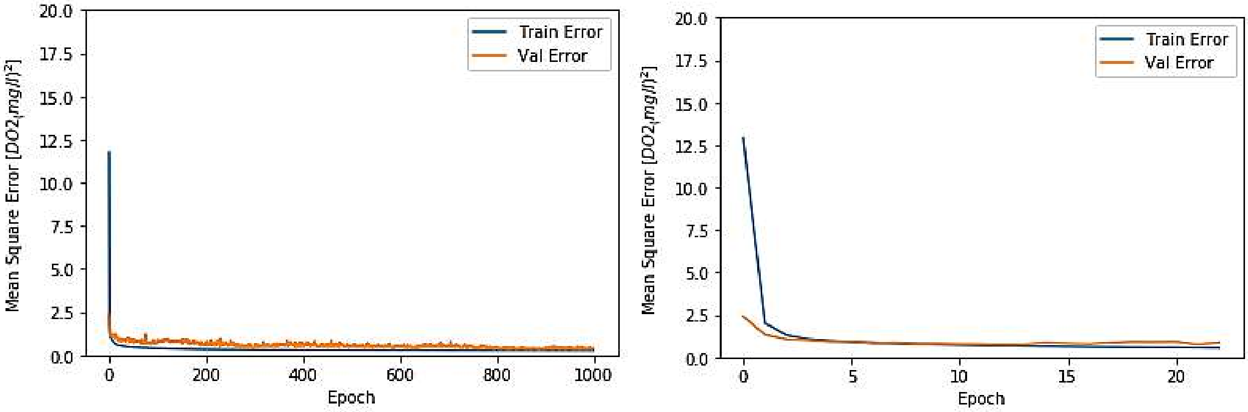

Figure 7.42: Mean Square Error (DOmodel, Toichio station, Structure 7-32-32-1)

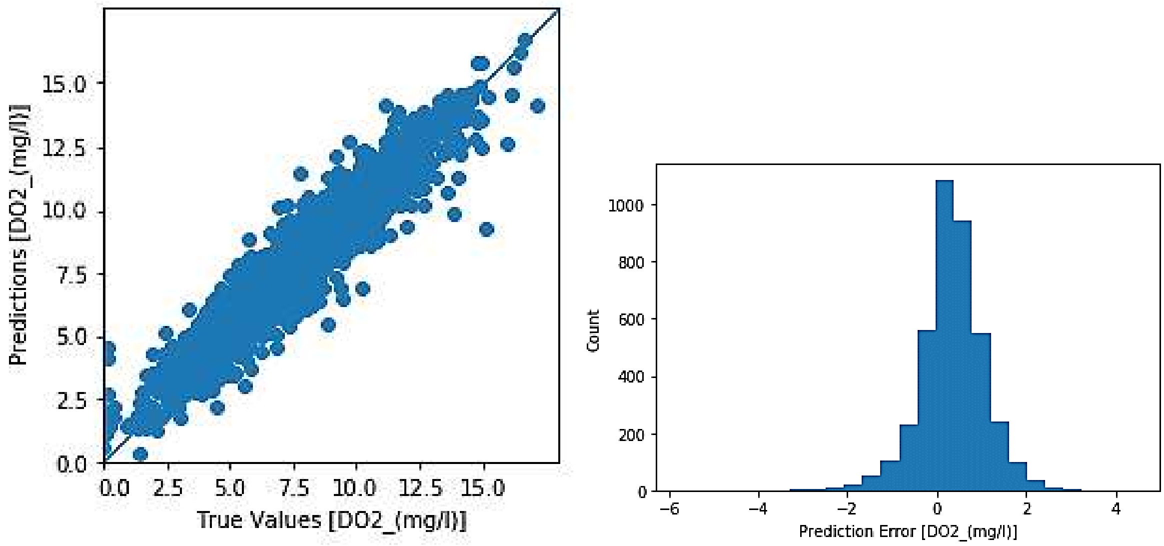

Figure 7.43: a) True vs Predicted values (left), b) Predicted Error (right) (DOmodel, Toichio station, Structure 7-32-32-1)

7.4.3 Psaradika Station

7.4.3.1 DO model

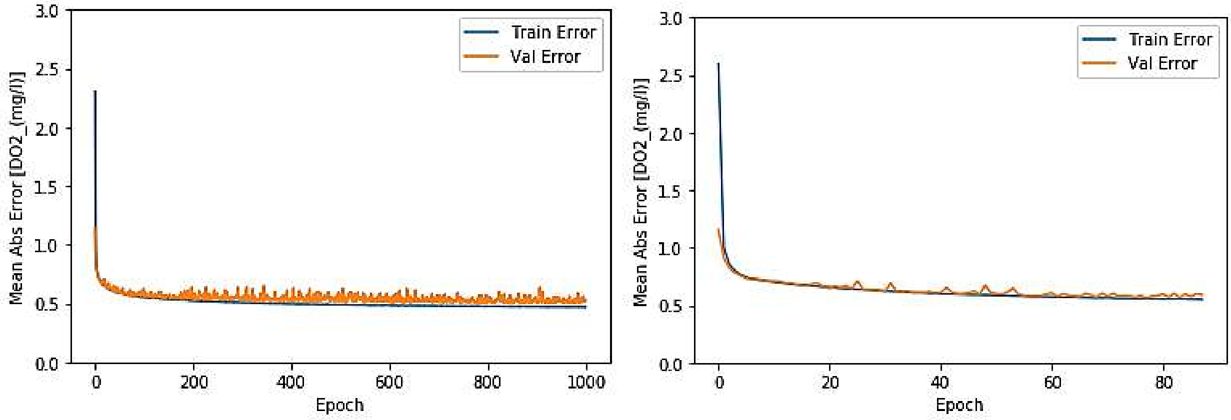

Figure 7.44: Mean Absolute Error (DOmodel, Psaradika station, Structure 7-32-32-1)

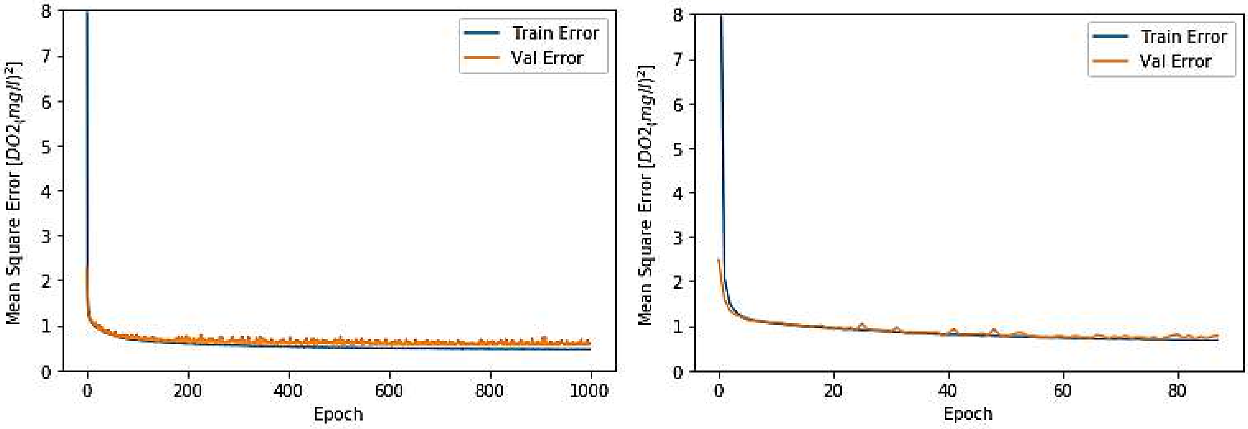

Figure 7.45: Mean Square Error (DOmodel, Psaradika station, Structure 7-32-32-1)

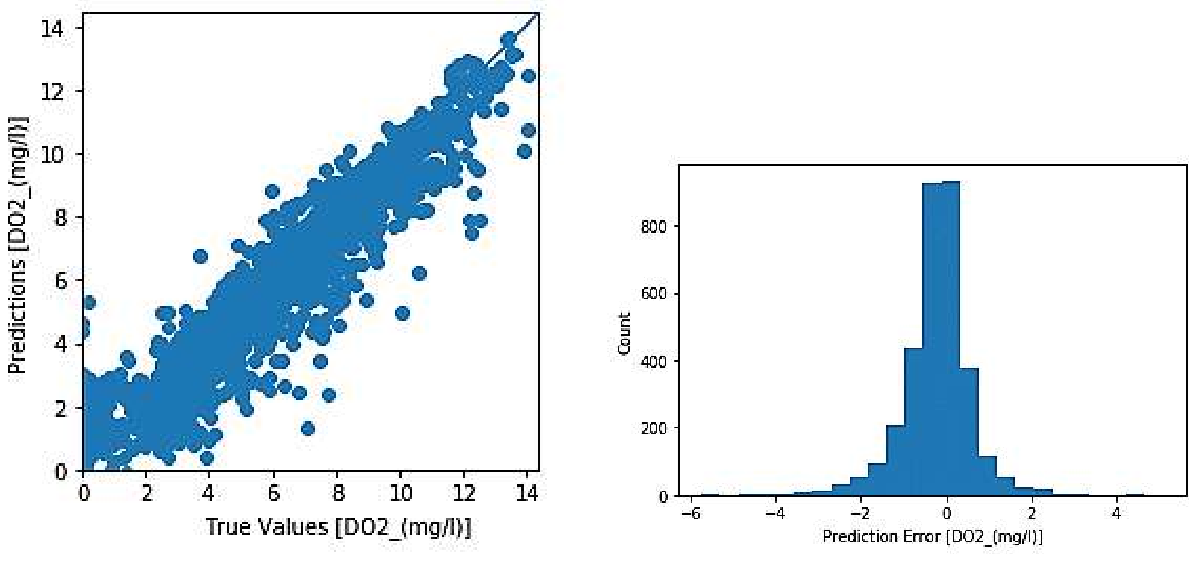

Figure 7.46: a) True vs Predicted values (left), b) Predicted Error (right)(DOmodel, Psaradika station, Structure 7-32-32-1)

7.4.4 Stavros Station

7.4.4.1 DO model

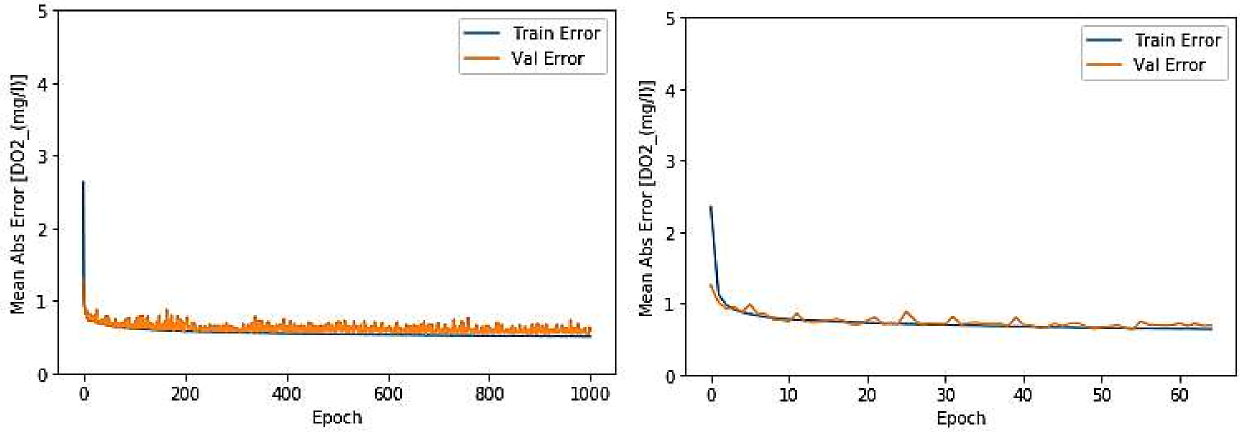

Figure 7.47: Mean Absolute Error (DOmodel, Stavros station, Structure 6-32-32-1)

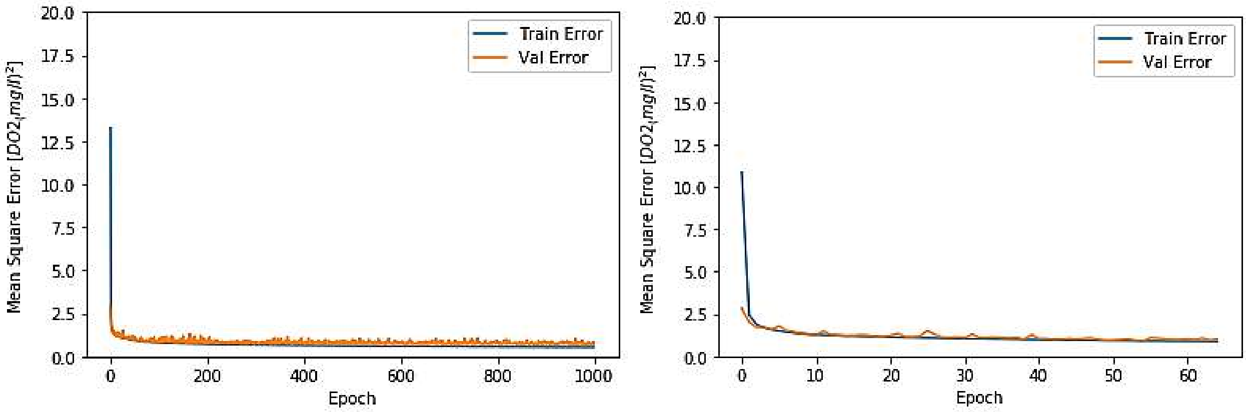

Figure 7.48: Mean Square Error (DOmodel, Stavros station, Structure 6-32-32-1)

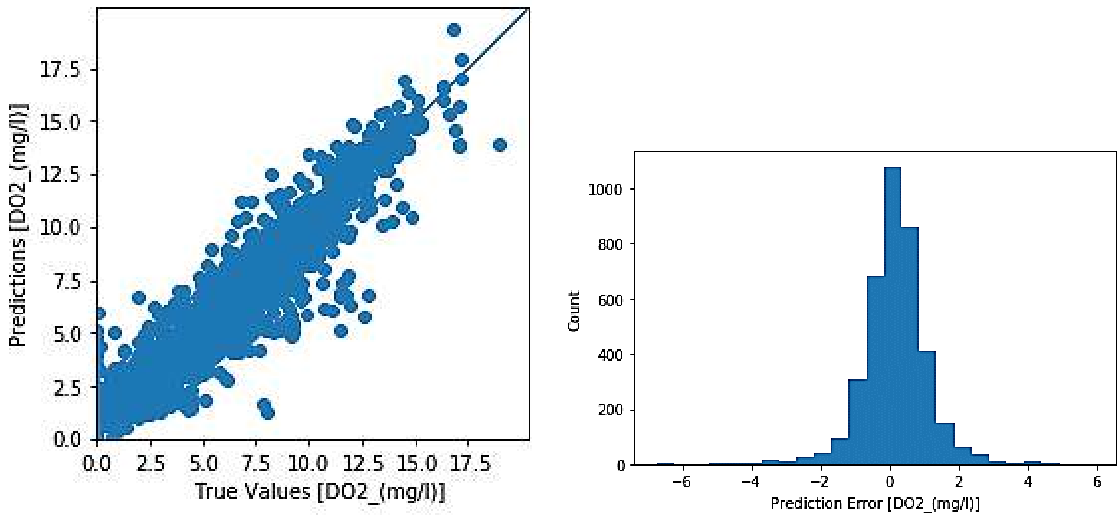

Figure 7.49: a) True vs Predicted values (left), b) Predicted Error (right) (DOmodel, Stavros station, Structure 6-32-32-1)

The DO models with structures 7-32-32-1, produce results with the MAE of 0,54; 0,55 and the MSE of 0,65; 0,68 for the training and the testing process respectively, for Gkiole station, the MAE of 0,50; 0,54 and the MSE of 0,54; 0,57 for the training and the testing process respectively, for Toichio station, the MAE of 0,57; 0,58 and the MSE of 0,70; 0,72 for the training and the testing process respectively, for Psaradika station, the MAE of 0,69; 0,70 and the MSE of 0,98; 1,01 for the training and the testing process respectively, for “Stavros” station, with structure 6-32-1. For all the investigated DO models, the training process takes place for 1000 epochs and convergence is achieved (Figures 7.38 and 7.39 for Gkiole station; 7.41 and 7.42 for Toichio station; 7.44 and 7.45 for Psaradika station; 7.47 and 7.48 for Stavros station). All DO models with structure 7-32-32-1 and 6-32-32-1, predicts reasonably well (Figures 7.40a; 7.43a; 7.46a and 7.49a). Finally, Figures 7.40b; 7.43b; 7.46b and 7.49b, shows the prediction error distribution of DO models for all stations.

7.5 Structure: 4-32-32-1

Structure 4-32-32-1, indicates that each investigated Deep Neural Network, is composed of one input layer with four input variables, two hidden layers with 32 neurons each, and one output layer with one output variable.

7.5.1 Gkiole Station

7.5.1.1 DO model

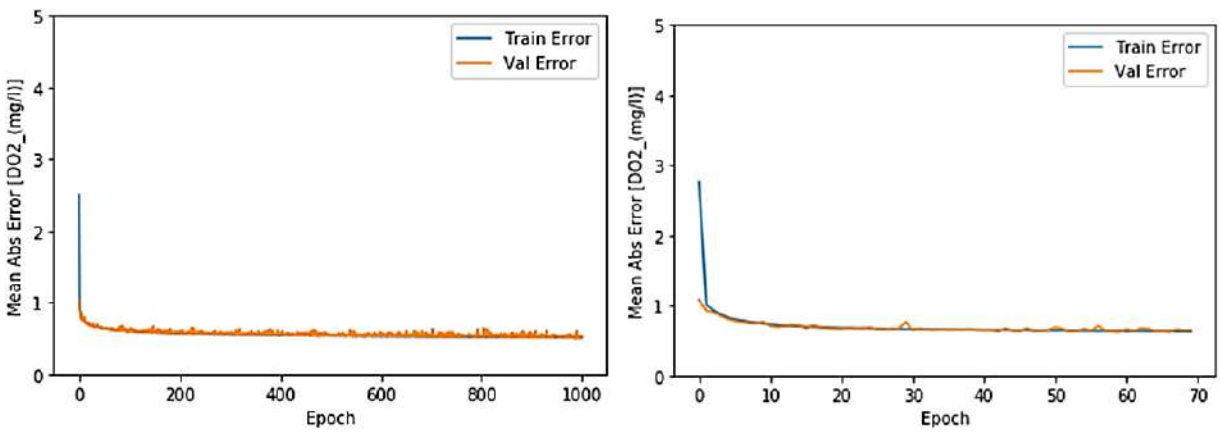

Figure 7.50: Mean Absolute Error (DO model, Gkiole station, Structure 4-32-32-1)

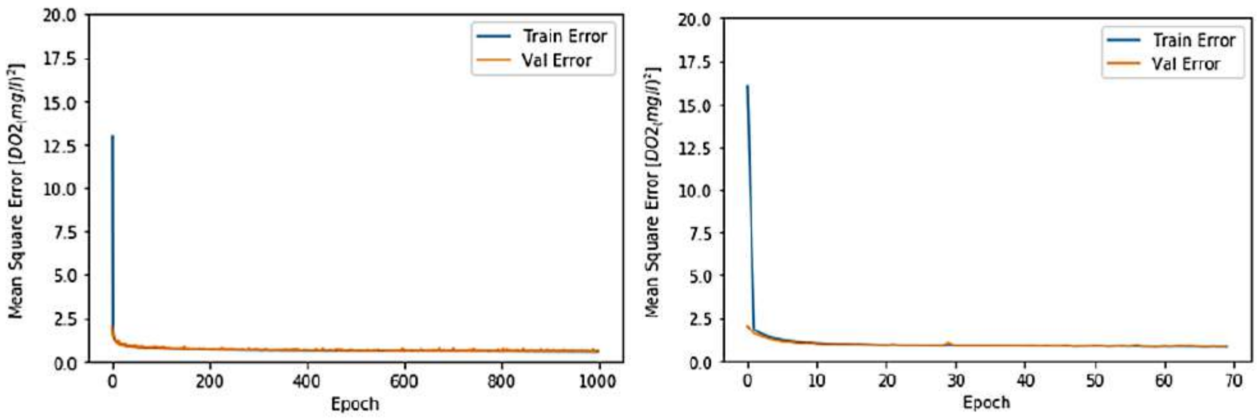

Figure 7.51: Mean Square Error (DO model, Gkiole station, Structure 4-32-32-1)

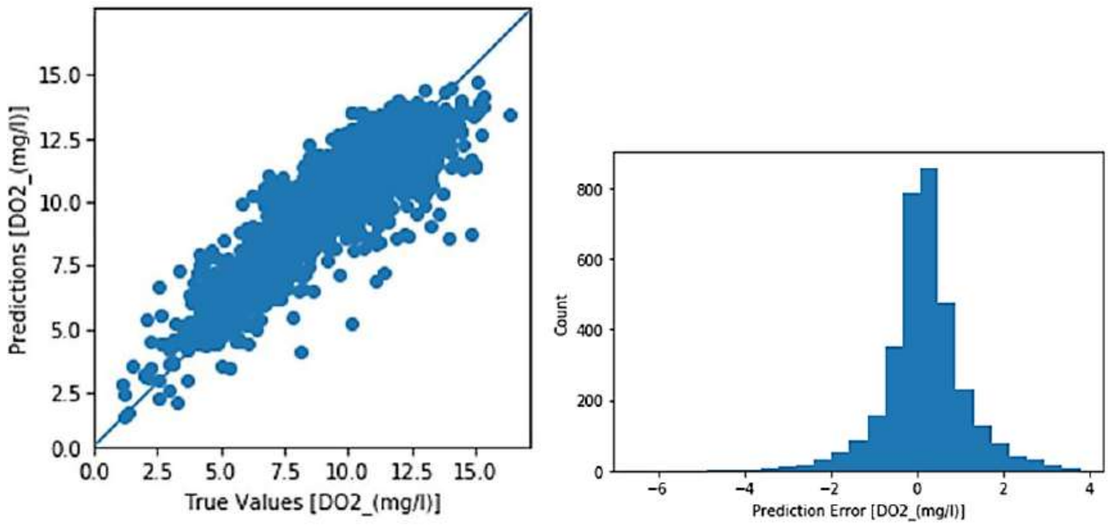

Figure 7.52: a) True vs Predicted values (left), b) Predicted Error (right) (DO model, Gkiole station, Structure 4-32-32-1)

7.5.2 Toichio Station

7.5.2.1 DO model

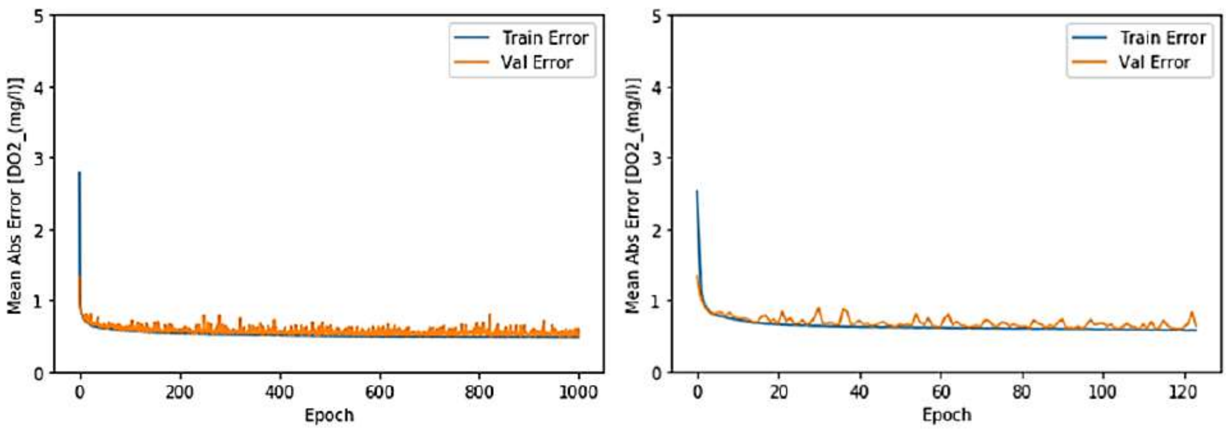

Figure 7.53: Mean Absolute Error (DOmodel, Toichio station, Structure 4-32-32-1)

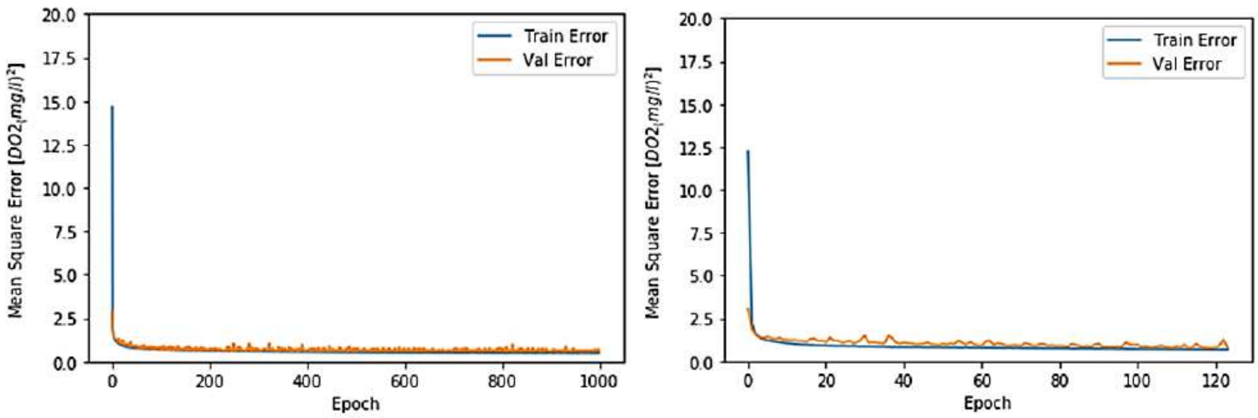

Figure 7.54: Mean Square Error (DOmodel, Toichio station, Structure 4-32-32-1)

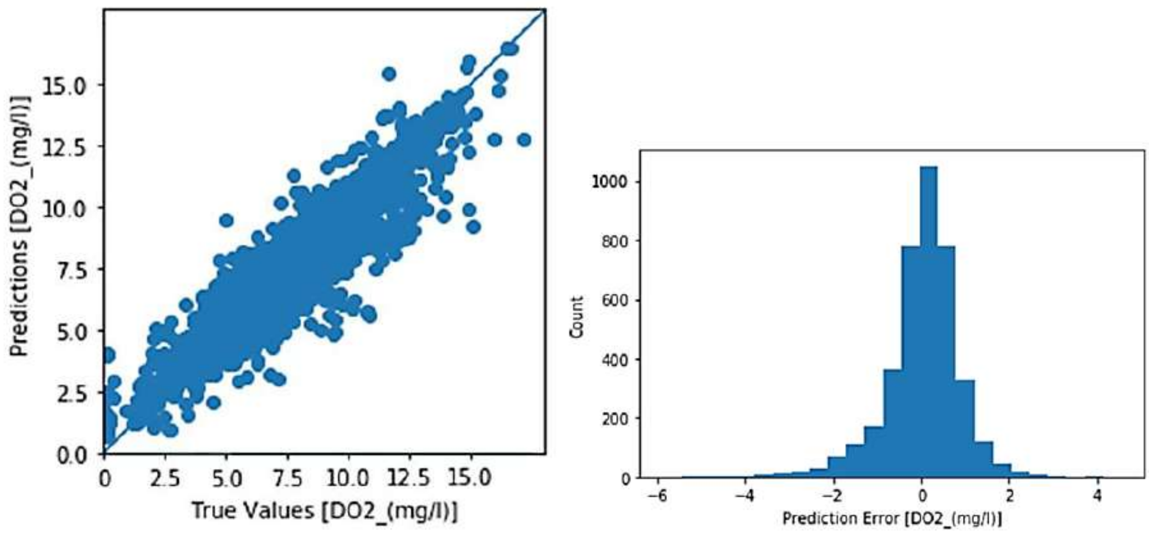

Figure 7.55: a) True vs Predicted values (left), b) Predicted Error (right) (DOmodel, Toichio station, Structure 4-32-32-1)

7.5.3 Psaradika Station

7.5.3.1 DO model

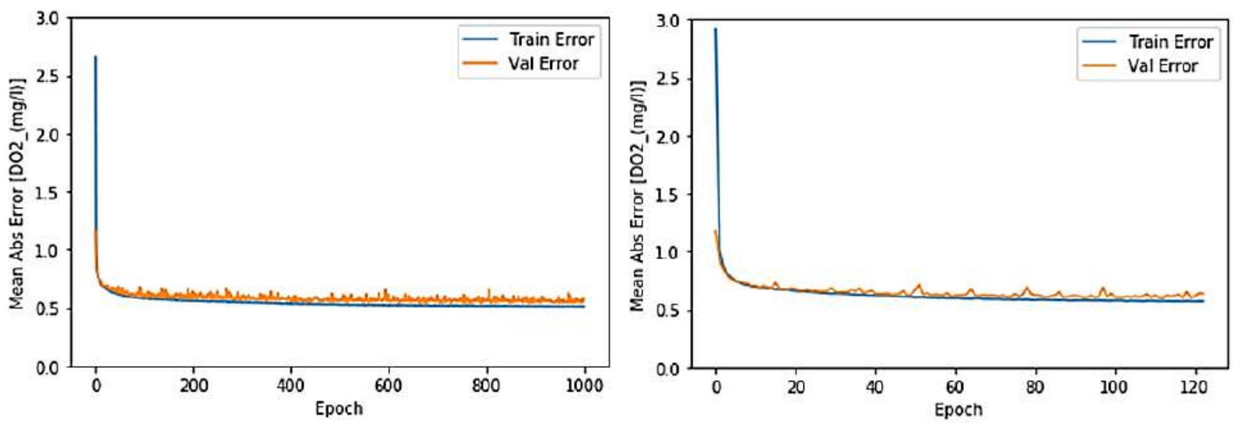

Figure 7.56: Mean Absolute Error (DOmodel, Psaradika station, Structure 4-32-32-1)

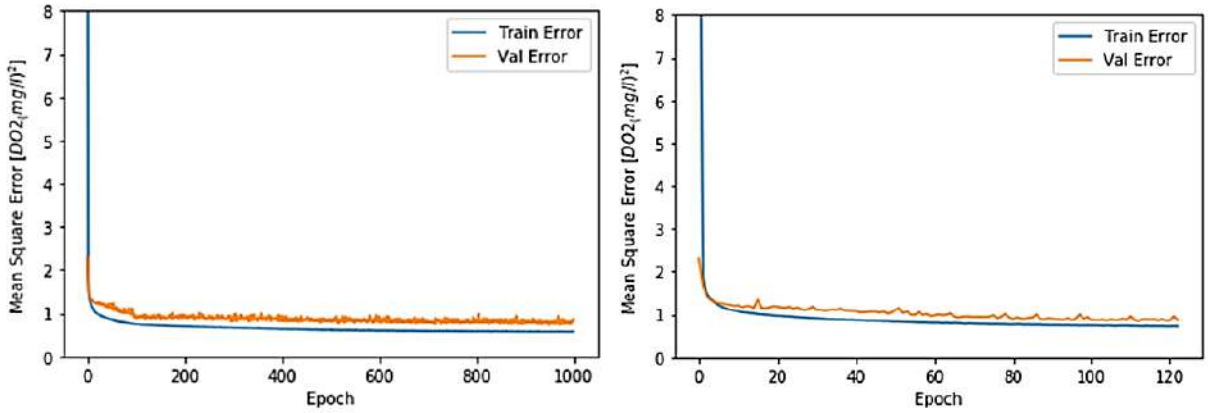

Figure 7.57: Mean Square Error (DOmodel, Psaradika station, Structure 4-32-32-1)

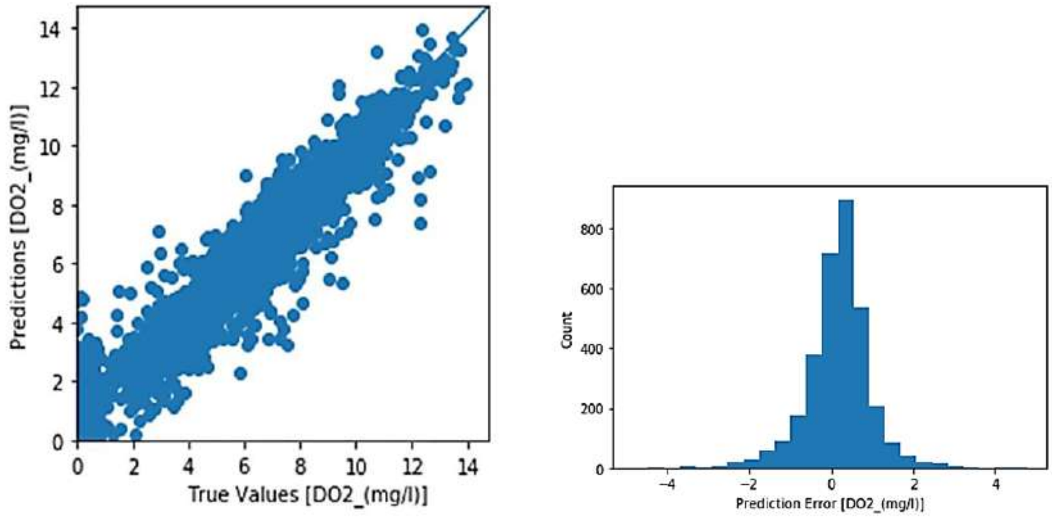

Figure 7.58: a) True vs Predicted values (left), b) Predicted Error (right)(DOmodel, Psaradika station, Structure 4-32-32-1)

7.5.4 Stavros Station

7.5.4.1 DO model

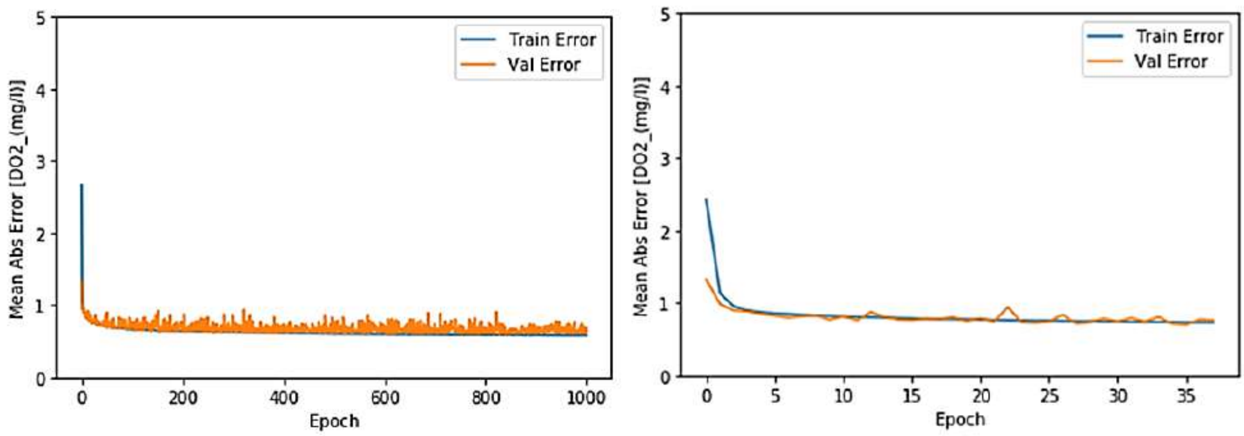

Figure 7.59: Mean Absolute Error (DOmodel, Stavros station, Structure 4-32-32-1)

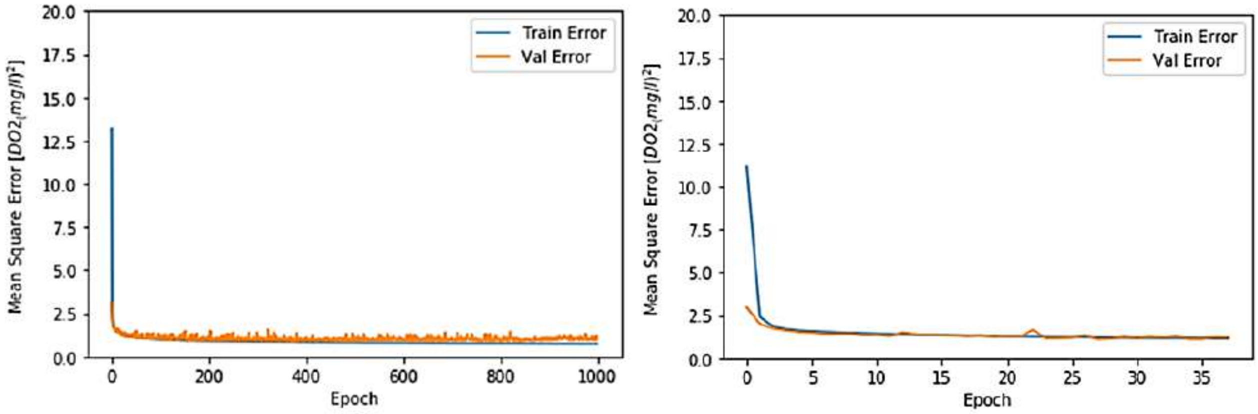

Figure 7.60: Mean Square Error (DOmodel, Stavros station, Structure 4-32-32-1)

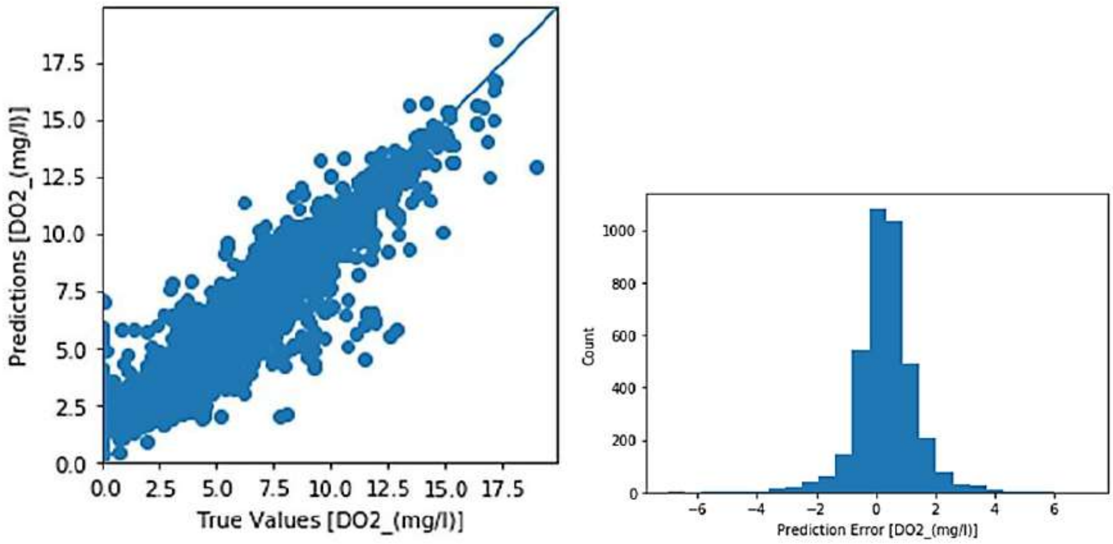

Figure 7.61: a) True vs Predicted values (left), b) Predicted Error (right)(DOmodel, Stavros station, Structure 4-32-32-1)

Dissolved Oxygen models with structures 4-32-1, produce results with the MAE of 0,64; 0,67 and the MSE of 0,84; 0,94 for the training and the testing process respectively, for Gkiole station, the MAE of 0,58; 0,62 and the MSE of 0,76; 0,77 for the training and the testing process respectively, for Toichio station, the MAE of 0,58; 0,60 and the MSE of 0,69; 0,71 for the training and the testing process respectively, for Psaradika station and the MAE of 0,73; 0,78 and the MSE of 1,23; 1,30 for the training and the testing process respectively, for Stavros station.For all the investigated DO models, the training process takes place for 1000 epochs and convergence is achieved(Figures 7.50 and 7.51 for Gkiole station; 7.53 and 7.54 for Toichio station; 7.56 and 7.57 for Psaradika station; 7.59 and 7.60 for Stavros station). All DO models with structure 7-32-32-1 and 6-32-32-1 predicts reasonably well (Figures 7.52a; 7.55a; 7.58a and 7.61a). Finally, Figures

7.52b; 7.55b; 7.58b and 7.61b, shows the prediction error distribution of DO models for all stations.

In general, the structure 7-64-1 is the most appropriate for Toichio station, while the structure 4-64-1 for Psaradika station. Moreover, the best performance for structure 7-32-32-1 shows the Gkiolestation, while Psaradika station is prevailed for models structure 4-32-32-1. Also, the investigated models ofTurbidity and Chl-a with structures 7-64-64-1 for Gkiole, Toichio and Psaradika stations and 6-64-64-1 for Stavros station, couldn't show a good performance. The models are firstly trained for 1000 epochs. When the results are not satisfactory, the epochs are increased. In cases of Turbidity and Chl-a model, where the validation and training processes couldn't converge, the number of epochs is increased to 5000, but still no improvement is recorded. The predictive abilityof the developed system has been investigated, according to statistical methods described in Chapter 5. The inability of models could be explained due to their lack of the capability to handle such a large variance in the data. For example, the minimum value of Turbidity for Psaradika station equals to 0,10 μS/cm, while its maximum value is 3000 μS/cm. Similarly, the minimum value of Chl-a for Gkiole station is 0,80 μg/L, while its maximum value reaches 441,10 μg/L. Hence, another type of Neural Network, that could handle the large variance in data, may be more suitable. Probably an evaluation both on the wet and dry seasons may guarantee a better generalization.

The Dissolved Oxygen models are trained for two different numbers of neurons, in the hidden layers and for two different input combinations. The optimal network architecture, of each structure for each station, is selected based on the one with the minimum statistical descriptors. Overall, four structures are compared as shown in Table 7.5 for each station. The well-trained DNN with structure 7-32-32-1, produces results with the MAE of 0,54; 0,55 and the MSE of 0,65; 0,68 for the training and the testing process respectively, with a good predictive ability for Gkiole station. For Toichio station, the structure 7-64-64-1 prevailswith the MAE of 0,48; 0,49 and the MSE of 0,52; 0,51 for the training and the testing process respectively and it constitutes the structure with the best performance compared with all the structures for all stations. For Psaradika station, the structure 4-64-64-1 produces the best results in relation to the others structures of this station with the MAE of 0,57; 0,58 and the MSE of 0,65; 0,68 for the training and the testing process respectively. Finally, Stavros station, presents results with the MAE of 0,69; 0,70 and the MSE of 0,98; 1,01 for the training and the testing process respectively (Structure 6-32-32-1). However, the selected structure 6-32-32-1 for Stavros station, is the less appropriate when compared to the aforementioned selected structures for Gkiole, Toichio and Psaradika station. However, it should be mentioned that, in all the investigated structures, the training errors are slightly lower than the tested ones, fact that indicates the good fit that has achieved.

Table 7.5: Statistical descriptors of the investigated structures for all stations

Station	Structure	Training			Test		
		MAE	MSE	NSE	MAE	MSE	NSE
Gkiole	7-64-64-1	0,69	0,89	0,81	0,71	0,92	0,79
	4-64-64-1	0,58	0,83	0,84	0,61	0,85	0,82
	7-32-32-1	0,54	0,65	0,89	0,55	0,68	0,86
	4-32-32-1	0,64	0,84	0,82	0,67	0,94	0,81
Toichio	7-64-64-1	0,48	0,52	0,92	0,49	0,51	0,93
	4-64-64-1	0,62	0,70	0,89	0,62	0,72	0,91
	7-32-32-1	0,50	0,54	0,91	0,54	0,57	0,92
	4-32-32-1	0,58	0,76	0,90	0,62	0,77	0,91
Psaradika	7-64-64-1	0,60	0,77	0,88	0,63	0,80	0,90
	4-64-64-1	0,57	0,65	0,91	0,58	0,68	0,92
	7-32-32-1	0,57	0,70	0,89	0,58	0,72	0,90
	4-32-32-1	0,58	0,69	0,89	0,60	0,71	0,90
Stavros	6-64-64-1	0,73	1,02	0,87	0,74	1,08	0,86
	4-64-64-1	0,69	1,07	0,88	0,72	1,11	0,86
	6-32-32-1	0,69	0,98	0,91	0,70	1,01	0,90
	4-32-32-1	0,73	1,23	0,86	0,78	1,30	0,85

Based on the investigated structures, the results demonstrate that the proposed DNN models (Table 7.6) constitute the best choice for modelling Dissolved Oxygen for each station. Table 7.6, also illustrates the efficiency of the selected structures. Moreover, it is concluded that the models that show the most appropriate predictive ability, based on their statistical descriptors for Gkiole, Toichio and Stavros station, are the models consisting of all the input parameters. Despite the fact that not all the input parameters have a high impact on the output, the use of all the available information seems to give a better performance in Deep Learning.In other words, all the parameters have a significant effect on the performance of the DNN model and cannot be excluded from the input variables. This inability to explain network behavior, may seem unacceptable to the scientist, but one should remember that when one is moving in the ambiguous contexts of stochastic phenomena and, in particular, Artificial Intelligence, this may not only be acceptable, but that's the scope. In other words, the structure and function of the network becomes “autonomous” achieving the idea of “Machine Learning”.

In case of Toichio station, the selected model does not use all the parameters to achieve a good performance, but it needs a Deep Network of 64 nodes per layer, instead of 32 in order to give the best results compared to other structures for this station.The fact that the model with structure 4-32-32-1, utilizing fewer inputs and nodes in hidden layers, is less appropriate in relation to all the investigated models for all stations,constitutes an evidence that complex Neural Networksare a promising field for improvement.

Table 7.6: The Optimal selected structures for each station

Station	Structure	Training			Test		
		MAE	MSE	NSE	MAE	MSE	NSE
Gkiole	7-32-32-1	0,54	0,65	0,89	0,55	0,68	0,86
Toichio	7-64-64-1	0,48	0,52	0,92	0,49	0,51	0,93
Psaradika	4-64-64-1	0,57	0,65	0,91	0,58	0,68	0,92
Stavros	6-32-32-1	0,69	0,98	0,91	0,70	1,05	0,90

When applying Deep Learning, one seeks to produce better results than the already existing shallow structures. Using Deep Learning has yielded results that are better than the previously existing techniques, indicates that this is an auspicious field for improvement.

Table 7.7 gives the hyper-parameters and statistical descriptors of investigated Neural Networks in literature for Dissolved Oxygen prediction. Rancovic et al. (2010), proposed a Back Propagation Neural Network in order to predict Dissolved Oxygen. Their shallow model with one hidden layer and structure 7-15-1, produces less appropriate results in relation to the proposed structures of this PhD thesis (Table 7.7). Antanasijević, et al. (2013), have also proposed a Back Propagation Network with better predictive ability than the aforementioned study, but quite similar to the results of this thesis, with a structure 4-10-1 and one hidden layer with the MAE of 0,55; 0,72 and the MSE of 0,54; 0,68 for the training and the testing process respectively. Wen et al. (2013) applied a Multilayer Perceptron model, to predict dissolved oxygen with very good predictive ability and structure 8-14-1. The learning rate that was used to achieve their goal, equals to 0,01; a value much higher than 0,001 used in this thesis. Ay and Kisi (2012), proposed both a Multi Layer Perceptron model and a RBNN model, with a large number of inputs compared with other studies (4765/parameter), but still fewer than the number of data used for this thesis. The statistical descriptors demonstrated a good predictive ability. However, the fact thatthe obtainedtraining errors were slightly higher than the tested ones, indicates that over-fitting was not avoided in their study such in case of Antanasijević, et al. (2013) with the Recurrent Neural

Network model and a number of samples per parameter around 4500 (comparatively high with other studies.

Table 7.7: Hyper-parameters and Statistical descriptors of investigated Neural Networks in literature

Model	Inputs	Number of Samples per parameter	Hidden Layers	Hidden Nodes	Activation Function	Learning Rate	Statistical Descriptors				Reference
							Train		Test		
							MAE	MSE	MAE	MSE	
BPNN	pH, NH_4^+, NO_3^- NO_2^-, ECw, Fe, Mn^{2+}	180	1	15	Sigmoid	NA	0,83	1,47	1,64	3,87	Ranković et al., 2010
MLP	pH, ECw, TA, Cl^-, Ca^{2+}, TH, NH_4^+, NO_3^-	164	1	14	Tan-Sigmoid	0,01	-	0,17	-	0,20	Wen et al., 2013
MLP	pH, T, ECw	4765	1	1	Purelin/logsig	NA	0,25	0,11	0,23	0,08	Ay et al., 2012
RBNN	T, pH, ECw	4765	2	12	Gaussian	NA	0,23	0,10	0,21	0,07	Ay et al., 2012
GRNN	Flow, pH, T, ECw	≈ 4500	1	61	NA	NA	0,07	0,03	0,60	0,60	Antanasijević et al., 2013
BPNN	Flow, pH, T, ECw	≃ 4500	1	10	NA	NA	0,55	0,54	0,72	0,68	Antanasijević et al., 2013
RNN	Flow, pH, T, ECw	≃ 4500	1	11	NA	NA	0,44	0,42	0,49	0,35	Antanasijević et al., 2013
MLP	BOD, COD, SS, TKN, TP Tcoliform, NH_4^+, NO_3^- NO_2^-, pH	384	1	16	Hypervolic tangent Sigmoid	NA	NA	NA	0,70	0,78	Areerachakul et al., 2011

It becomes clear, that in the present PhD thesis, the use of the selected tools and platforms described in Chapter 6, gives the advantage of using large databases, while the training and testing procedure is obtained in a very short time (the GPU card contains 3584 cores and use them all). As it has already mentioned, the learning rate is a crucial hyper-parameter in Deep Learning, controlling how much to change the model in response to the estimated error, each time the model weights are updated. If the learning rate is too large, the network fails to converge. Most of the Neural Networks utilizea value from 0,01 to 0,3 of the learning rate. In the present PhD thesis, taking advantage of the computing power of the graphic card, a learning rate of 0,001 is used. Moreover, Tensorflow provides state-of-the-art Machine Learning methods and the possibility of using parallel programming through the computing power of the graphics card makes the process of Machine Learning constantlygaining ground. Also, the Python programming language is free of charge, compared with Matlab that is widely used in Machine Learning.

References

Antanasijević, D., Pocajt, V., Povrenović, D., Perić-Grujić, A. and Ristić, M., 2013. Modelling of dissolved oxygen content using artificial neural networks: Danube River, North Serbia, case study. Environmental Science and Pollution Research, 20(12), pp.9006-9013.

Areerachakul, S., Junsawang, P. and Pomsathit, A., 2011. Prediction of dissolved oxygen using artificial neural network. In Int Conf Comput Commun Manage, 5, pp. 524-528.

Ay, M. and Kisi, O., 2012. Modeling of Dissolved Oxygen concentration using different neural network techniques in Foundation Creek, El Paso County, Colorado.*Journal of Environmental Engineering*,138(6), pp.654-662.

Kagalou, I. and Psilovikos, A., 2013. Assessment of the Typology and the Trophic Status of two Mediterranean Lake Ecosystems (NW Greece). *Water Resources*, 41(3), pp 335 – 343.

Mantzafleri N., Psilovikos A., and Mplanta A. 2009. Water quality monitoring and modeling in Lake Kastoria, using GIS. Assessment and management of pollution sources. *Water resources management*,23(15), pp.3221-3254.

Mellios, N., Moe, S. and Laspidou, C., 2020. Machine Learning Approaches for Predicting Health Risk of Cyanobacterial Blooms in Northern European Lakes. *Water*, 12(4), p.1191.

Nürnberg, G.K., 1996. Trophic state of clear and colored, soft-and hard water lakes with special consideration of nutrients, anoxia, phytoplankton and fish. Lake and Reservoir Management, 12(4), pp.432-447.

Organization of Economic Coorperation and Development (OECD), 1982. Eutrophication of waters. Monitoring, assessment and control, OECD cooperative programme on monitoring of inland waters (eutrophication control). Environmental Directorate OECD, France, Paris.

Psilovikos A., 2014. Ecohydraulics. Tziolas Pub, Thessaloniki, Greece, 536 p. ISBN:978-960-418-447-7 40 (in Greek).

Psilovikos, A., 2020. Water Resources. Tziolas Pub, Thessaloniki, Greece, 700 p. ISBN: 978-960-418-602-0, (in Greek).

Ranković, V., Radulović, J., Radojević, I., Ostojić, A. and Čomić, L., 2010. Neural network modeling of dissolved oxygen in the Gruža reservoir, Serbia. Ecological Modelling, 221(8), pp.1239-1244.

Thomann, R.V. and Mueller, J.A., 1987. Principles of surface water quality modeling and control. Harper & Row Publishers.

Wen, X., Fang, J., Diao, M. and Zhang, C., 2013. Artificial neural network modeling of dissolved oxygen in the Heihe River, Northwestern China. Environmentalmonitoring and assessment, 185(5), pp.4361-4371.

Wetzel, R.G., 2001. Limnology: Lake and River Ecosystems. Gulf professional publishing.

World Health Organization (WHO), 1996. Guidelines for Drinking-Water Quality, Vol.2: Health Criteria and Other Supporting Information, 2nd Edition, World Health Organization, Geneva, 1996.

CHAPTER 8: CONCLUSIONS

"Preventing is better than curing"
Hippocrates (460 – 370 BC)

Lakes constitute important indicators of climate change. Geographically, they comprise the lowest points of an internal catchment's area, collecting the surface runoff water from the hydrographic network. Their physical, chemical and biological responses to the climate give a variety of priceless information. That is the reason why monitoring systems play a decisive role in lake management by providing systematic information of the qualitative parameters.

The monitoring of the water quality parameters of Lake Kastoria, is based on a complete telemetric, self-powered, low-energy consumption system of electronic monitoring stations. Since 2015, Lake Kastoria is monitored telemetrically on a continuous daily basis, by 4 Stations at one location on its surface. The telemetric stations record: Chlorophyll a (Chl-a - μg/L), pH, Water temperature (Tw - ° C), Conductivity (Cond - μS/cm), Turbidity (Turb - NTU), Dissolved Oxygen (DO - mg/L), Ammonia nitrogen ($N\text{-}NH_4$, mg/L); Nitrate nitrogen ($N - NO_3$, mg/L); and Water level (m). Such integrated systems constitute powerful management tools for the sustainability of the aquatic ecosystems. The aforementioned Telemetric Stations, assist the main monitoring station of Lake Kastoria, (according to Directive 2000/60, Law 3199/2003 and JM 140384/2011), both spatially (more stations in different locations) and chronicle (continuous monitoring). These results are collected by the Ministry of Environment and Energy of Greece. After all, without the present monitoring system of Lake Kastoria, it would be impossible to apply Deep Neural Network models.

The Real – Time Monitoring of the Water Quality Parameters contributes in Management Practices, the most important of which are summarized below:

- ✓ Controlling water quality for irrigation
- ✓ Monitoring atmospheric conditions
- ✓ Determining microclimate indicators
- ✓ Warning in cases of crises caused by extreme events
- ✓ Continuous knowledge of the state of the water bodies
- ✓ Maintaining the Ecological Balance of the ecosystems and water resources of the region.

An increasing trend in the number of the articles studying Machine Learning to predict Water Quality parameters in aquatic systems is observed, during the last fifteen years. Concerning the models used in order to predict Water Quality parameters, ANNs show the most impressive rise, from 1995 until 2019, as their use has increased dramatically. DNN, an upgraded and more complex version of ANN, shows an upward trend mainly from 2010 onwards. The remarkable rise, from 2017 and on, occurs mainly due to the appearance of new platforms and tools that facilitate their use in terms of accuracy and speed.

At the present PhD Dissertation, the ability of Deep Neural Networks to predict in real-time the quality parameter of DO, Chl-a and Turbidity in Lake Kastoria, is tested. The examined Feed Forward Deep Neural Networks (FF-DNN) of DO, are tested for four different structures based on the number of nodes, used in the hidden layers and the number of input variables. The investigated models of Turbidity and Chl-a, with structures 7-64-64-1 for Gkiole, Toichio and Psaradika stations and 6-64-64-1 for Stavros Station, couldn't show a good performance. The inability of models could be explained due to their lack of the capability to handle such a large variance in the data.

The optimal selected Feed Forward Deep Neural Networks (DNN) of DO for each station provide information in Real Time and comprise a powerful Decision Support System (DSS), for preventing accidental and emergency conditions that may arise from both Natural and Anthropogenic Hazards**.** The well-trained DNN, with structure 7-32-32-1, produces results with the MAE of 0,54; 0,55, the MSE of 0,65; 0,68 and the NSE of 0,89; 0,86 for the training and the testing process respectively, with a good predictive ability for Gkiole station. For Toichio station, the structure 7-64-1 prevails with the MAE of 0,48; 0,49, the MSE of 0,52; 0,51 and the NSE of 0,92; 0,93 for the training and the testing process respectively and it constitutes the structure with the best performance compared with all the structures for all stations. For Psaradika station, the structure 4-64-1 produces the best results in relation to the other structures of this station with the MAE of 0,57; 0,58, the MSE of 0,65; 0,68 and the NSE of 0,91; 0,92 for the training and the testing process respectively. Finally, Stavros station with structure 4-32-1 presents results with the MAE of 0,69; 0,70, the MSE of 0,98; 1,01 and the NSE of 0,91; 0,90 for the training and the testing process respectively.

In practice, the use of Neural Networks in hydrology, tends to mimic hydrological processes, which science does not fully understand or can express with the help of a mathematical formula. However, it should be noted that, according to their structure and function, Neural Networks generally do not provide a better understanding of hydrological processes and natural phenomena, as they simplify physics and "degenerate" into weights and threshold values.

One has to understand the natural procedures that occur, in order to select the appropriate network structure and to apply the appropriate training algorithm. Moreover, the right choice of network architecture, activation functions, and learning methods, could be substantiated through a test and control process. Also the selection of training data is of paramount importance, as it requires proper preparation and normalization.

However, in a water management point of view, due to the user-friendly nature of the proposed Neural Networks, they can be implemented in real time by non-specialists, as no knowledge of the phenomenon is required after the end of training phase. When applying Deep Learning, one seeks to produce better results than the already existing shallow structures. Employing Deep Learning, has yielded results in these cases that are better, than the previously existing techniques, which is evidence that this is a promising field for improvement.

Moreover, useful information is extracted that allow us to understand the consequences of the LULC changes. From 1945 to 2000, the forest and semi-natural areas of Lake Kastoria's catchment area show a rise of 60%, agriculture decreased by 35%, aquatic areas show stability, while wetlands have a remarkable decrease of 60%. Finally, the artificial surfaces doubled their area and in particular they increased by 140% as expected. Despite the fact that agricultural and urban sources of pollution have increased, the presence of the biological treatment system since 1995, has significantly reduced the phenomenon of eutrophication.

Moreover, it is concluded, that the obtained IDW maps, of the three most representative months, are in agreement with the values of the monitored parameters of the telemetric stations, as well as the ecological processes, that take place in water body of Lake Kastoria.

The North, East and South parts of the catchment of Lake Kastoria, are characterized by extensive arable land and, in contrast to its West part, it is occupied by the city of Kastoria, forest and semi-natural areas. The Gkiole Stream, constitutes the only surface discharge of Lake Kastoria's basin, as well as the connection between the Lake and the Aliakmonas River. The Stavros and Psaradika stations, are located South and North of the coastline respectively, that are the most significant spots of the urban part of the Lake Kastoria. They constitute stations of significant importance, for measuring the ecological sensitivity of the aquatic ecosystem by the pressures exerted by the city itself. Finally, the Toichio station is located in the Northern part of Lake Kastoria. The areas at the lowest part of the Torrent Toichio are intended for agricultural use. Concerning the North-Western side of the lake, the waters are characterized by a stationary water status.

In an aquatic ecosystem, such as a lake ecosystem, in January, the lowest Tw values are presented; large quantities of precipitation are detected, strong winds occur, while the concentration of nutrients are poor. Under prevailing environmental conditions, therefore, one expects high levels of Dissolved Oxygen inside the lake, mainly in surface water, although photosynthesis activity is poor.

In May, the temperature rises, the regeneration of aquatic and riparian vegetation take place, photosynthesis activity show its peak, as abundance of photosynthetic organisms and of Dissolved Oxygen in the waters are at seasonally satisfactory levels. In addition to surface water, excess nutrients are present, while concentrations of Nitrate, Nitrite, Ammonium and Phosphate, which play a key role in the lake's nutritional status, are increasingly detected. Generally, in the given period of time, the so-called “good” side of the eutrophication is presented, where the DO balance is positive and the re-aeration level is greater than the respiration one.

On the contrary, in September the conditions of the “bad” side of eutrophication are prevailed. Aquatic and riparian flora start to decompose, aerobic microorganisms and especially bacteria, which consume excessive amounts of DO, through respiration processes, are rapidly increased. Consequently, the reduction in DO is likely to cause stress to fish populations, even sometimes leading to fish deaths. Moreover, nutrients that flow into the lake are increased dramatically, causing hypoxic conditions, especially in shallow lakes and lakes in both urban and agricultural catchments, such as Lake Kastoria. Subsequently, anaerobic bacteria create gaseous decomposition products, such as the toxic Hydrogen Sulfide (H_2S) and Ammonia in both ionized (NH_4^+) and non-ionized (NH_3) forms or in the form of Nitrogen Ammonia ($N - NH_4^+$), especially in areas very close to the bottom and the sediments of the lake.

However, it is a little bit risky to try to interpret water quality parameters, because we have monitoring data only from the abiotic parameters and the ecosystem of Lake Kastoria is so quite complicated, that it is unsafe to assess successfully. So, we try to explain, based only to the abiotic monitoring data, by quoting the results of the research. Another point is that, we take for granted that the electronic sensors of the monitoring probes, are frequently maintained, calibrated and replaced when a damage occurs. There is much uncertainty of the assessment of these parameters and only for the parameters of Tw, DO, Chl-a and Turbidity, we try to obtain a satisfactory interpretation, considering that Tw is a very important input independed parameter and the three latter, are the output parameters, obtained by the structure of the Deep Learning Machine Models.

Finally, the design of the contour deterministic maps contains uncertainty too, because from a number of four point values, we spatially integrate and assume that the equipotential curves which prevail, gives us representative values of the reality. Unfortunately, we cannot use geostatistical methods, because of the low number of monitoring points, but we do our best, using the most appropriate deterministic method (IDW).

The highest levels of Chlorophyll-a are detected peripherally of Gkiole's telemetric station, in January, while for the rest telemetric stations its concentration is low showing the minimum value (4,42 μg/L) near Stavros station. In May, the lowest values of Chlorophyll-a concentration are restricted to the perimeter of the Psaradika station (Northwest), while its highest values are detected in the Northeastern and Southern parts of the lake. In September, the highest chlorophyll-a values continue to appear in the Northeastern part of the lake. On the contrary, Stavros station (Southwest), show its minimum value.

Regarding the correlation of Tw with the surface of Lake Kastoria and the season, it is observed that the lowest Tw values are observed in the Northern and Western part of Lake Kastoria during all the reference months. In contrast, the highest Tw values, appear in the Southeastern part of the lake. The lowest Tw is in January, while the highest is in September, considering the periods – months that we take into account.

Concerning the parameter of Dissolved Oxygen, in January, the highest values are observed in the Northeastern and Southwestern part of the lake, where this situation is further established in the Southwestern part, in May. This is explained mainly by the torrents that flow into the lake from the north and east, enriching the lake with water during the winter and spring. In September, the concentration of Dissolved Oxygen is decreased in the Southeastern part of the lake, due to the opening of Gkiole's weir. In general, DO values remain constant throughout the year. However, in the past, based on the literature, low values of Dissolved Oxygen were observed in January, due to the sharp drop of Tw in the Lake's surface where it got frozen for a couple of weeks, making the enrichment of the water with DO, difficult.

The highest pH values are found around Gkiole station, while the lowest pH value is observed at Stavros station. These differences are not significant, but they are an indicator showing that the lowest pH values occur near the urban side of Lake Kastoria. pH values are constant for the three representative months.

Concerning Electrical Conductivity of Water (ECw), it is observed that, in January and May, its range is small, indicating that there are no significant fluctuations in ECw values during this month. In contrast to the month of September, a wider range of ECw values is observed (around 60 μS/cm), which is not enough though to proceed to its assessment. According to Figure 5.16, the minimum ECw values are detected in the Southern part of the lake and in particular around

Gkiole station. The maximum values are found to the Southwest part at Stavros station. In the reference months, January and September, the Northern part of the lake shows a moderate and low profile, in contrast to May, where ECw is higher.

In May, the turbidity range is very high (exceeding 300 NTUs), indicating significant fluctuations. The highest turbidity value in January, May and September reaches 128,9; 309,21 and 55,7 NTUs, respectively. The highest turbidity values, for the three months of reference, appear in the Western part of the lake, mainly due to the urban load. In contrast, the lowest values of turbidity, for the three months of reference, occur in the Northeastern, Eastern and Southeastern parts of the lake, due to the presence of torrents and Gkiole’s weir. It should be mentioned here that rainfall events, can increase erosion and result in greater amounts of sediment washing into Lake Kastoria, and finally cause increased Turbidity values. Moreover, the increased Turbidity values appear in the city's Northern part, due to the presence of a small jetty that collects phytoplankton.

Also, the presence of an agricultural and/or urban area near the lake enhances the presence of Nitrate, Nitrite and Ammonium ions. It is concluded that the minimum values of Nitrate are detected in the Southern part of the lake, and in particular at Gkiolestation, in January and May. The maximum values are found to the North and Southwest side, in May and September. The maximum Nitrate concentration occurs in September, with a value of 1,72mg/L on the Southwest side of the lake, while the minimum one appears in May, with a value of 0,78mg /L on the South side of the lake and around Gkiole’s station. It is noted that in September, in the interior of the lake where the lowest concentration of Chlorophyll-a is detected, the concentrations of nitrate are enhanced and vice versa.

Concerning Ammonium concentration, its range is appeared small (0,5 mg/L), indicating that there were no significant fluctuations during May. On the other hand, in January, the concentration of Ammonium is 2,6 mg/L. The concentrations of ammonium in May and September show similar patterns. In January, the lowest concentration of ammonium is found in the southern part of the lake and in particular around Gkiole station, while the highest concentration of ammonium is detected in the western part of the lake.

Based on the literature and the knowledge of the author, the importance of this work also lies within the fact that there is no other published work using the same platforms, tools and methodology of Deep Learning, as in the present PhD Dissertation, in order to investigate the predictive capacity of water quality parameters of lakes in real time, at both national and

international level. Moreover, the fact that Lake Kastoria is monitored by four telemetric stations, offering continuous and uninterrupted data sets (a few hundred thousand measurements), support the use of Deep Neural Networks, that need a large amount of data in order to complete their learning process, and enhance the originality of the present PhD Dissertation.

Possible future work regarding the present PhD Dissertation could include:

a. As it has mentioned, the classification of land cover at the 1st reference level of the Kastoria Lake Basin remains the same, although their percentages have changed significantly. These data could probably support further research that has to do with post – classification of the Land Use – Land Changes and can be used as a tool for estimating also the changes in the pollution in the water body of the Lake.
b. The optimization process of IDW method. Changing neighborhood options (shape, sectors and the number of neighbors) may lead to a better model.
c. Evolute this study with new water quality indexes for Total Phosphorus, Chl-a and Secchi Disk, N/P ratio, phycocyanin etc, which are parameters that are not monitored yet, in a continuous base.
d. Use additional information captured from a modern drone (Uncrewed Aerial Vehicle) equipped with a multispectral camera as ground-truth information to calibrate satellite imagery in order to improve quantification the specific quality parameters of the water from the study area. The additional information could also be enhanced by using data (ground data collection) derived from field work (targeted area samplings) in the study area. Existing operational algorithms could be tested or maybe new ones could be created in order to find the best fit of the band ratio.
e. The use of more complex Machine Learning methods (such as Convolutional Neural Networks - CNNs), not only for Lake Kastoria, but also for other national and international lakes, mainly in neighboring countries with cross-border water resources (such as Prespes, Doirani and Kerkini).