Assessment of proper wastewater treatment level according to marine ecosystem state

Volf, Goran

Assessment of Proper Wastewater Treatment Level according to Marine Ecosystem State

Ph.D Thesis

Goran Volf

Faculty of Civil Engineering at the University of Rijeka, Croatia

Doctoral Degree Programme

Hydrotehnics

Supervisor

Prof.dr.sc. Boris Kompare

Co-supervisor

Prof.dr.sc. Nevenka Ožanić

Rijeka, 2012

"Any entity or natural unit that includes living and nonliving parts interacting to produce a stable system in which the exchange of materials between the living and nonliving parts follows circular paths is an ecological system or ecosystem. The ecosystem is the largest functional unit in ecology, since it includes both organisms (biotic communities) and abiotic environment, each influencing the properties of the other and both necessary for maintenance of life as we have it on the earth. A lake is an example of an ecosystem."

Odum. E.P., 1953,

Fundamentals of Ecology.

BLANK PAGE

Acknowledgements

The foremost acknowledgements certainly go to the scientific project „Hydrology of sensitive water resources in karst“ (114-0982709-2549) funded by the Ministry of Science, Education and Sports of the Republic of Croatia, under leadership of Prof.dr.sc. Nevenka Ožanić, Faculty of Civil Engineering at the University of Rijeka which enabled a great part of my Ph.D research.

I would like to express my deep gratitude and special thanks to my supervisor, Prof.dr.sc. Boris Kompare for his irreplaceable guidance and support during the past few years of our collaborations that resulted in the accomplishment of the Ph.D Thesis. I would also like to thank Prof. Kompare that I was given the opportunity to place my Ph.D Thesis in the field of sanitary engineering which I find very exciting and fascinating.

Also I would like to express my sincere gratitude to my co-supervisor and project manager, Prof.dr.sc. Nevenka Ožanić for her help, guidance and remarkable support during the emergence of this Ph.D Thesis.

Special thanks to Dr.sc. Nataša Atanasova who, together with my supervisor and co-supervisor greatly helped me in guidance and gave me tremendous support during research and writing scientific papers and reviewing drafts of the Ph.D Thesis. I would also like to thank her for always finding the time to answer every question related to the domain of machine learning and any others questions and doubts that I had.

Next, I would like to thank to Dr.sc. Robert Precali and Dr.sc. Danilo Degobbis from the Centre for Marine Research in Rovinj for providing the data which were used in Ph.D Thesis. I would like to thank them for ongoing support and valuable comments during writing and reviewing of the scientific papers as a partial fulfilment of this Ph.D Thesis. Also, to Dr.sc. Robert Precali I would like to thank for the guidance and support during reviewing drafts of the Ph.D Thesis.

To Dr.sc. Francesco Tornatore I would like to thank for the data, valuable information and studies related to the Po River. My thanks go also to Dr.sc. Barry M. Evans for his valuable comments and his professional help with the AVGWLF software application.

Finally, I would like to thank all of my friends and colleges at the Faculty of Civil Engineering in Rijeka, as well as my friends and colleges in Ljubljana who offered me their help and support during the stay at the Faculty of Civil and Geodetic Engineering in Ljubljana.

Above all, I would like to thank my parents, my brother and my girlfriend for their support.

BLANK PAGE

Ph.D Thesis title: Assessment of Proper Wastewater Treatment Level according to Marine Ecosystem State

Author: Goran Volf

Supervisor: Prof.dr.sc. Boris Kompare

Co-supervisor: Prof.dr.sc. Nevenka Ožanić

Key words: northern Adriatic, machine learning, water quality, ecosystem state, watershed management, wastewater treatment level

Abstract

One of the most important sources of pollution in marine ecosystems are those produced by human activities in the associated watersheds. Understanding the linkage between water quality of marine ecosystems and surrounding watersheds is important in order to better assess processes in marine ecosystems and to evaluate different management options aimed at improving the marine ecosystem state. The goal of this Ph.D Thesis is to contribute to achieve an operational ecosystem-based management of marine ecosystems, which has been called for in the Marine Strategy Framework Directive (2008/56/EC), by improving the scientific knowledge of the functioning of marine ecosystems, especially under the actions of different pressures (such as nutrient enrichment) from surrounding watersheds, analyzed alone or together, both anthropogenic and natural. Attention is paid to the assessment of proper wastewater treatment level according to the desired marine ecosystem state.

Northern Adriatic (NA) is chosen for the case study area. NA is the shallowest and subsequently one of the most productive parts in the Adriatic, as well as in the whole Mediterranean Sea. Surrounding watershed area of NA is spread over four neighboring countries, e.g. Italy, Slovenia, Croatia and Switzerland. This watershed is characterized by different natural as well as anthropogenic pressures, e.g. agriculture and urbanization.

To achieve the goal of this PhD Thesis the principles of Driving forces-Pressures-States-Impacts-Responses (DPSIR) framework were used, in particular on defining Pressures, how Pressures exerted on large-scale marine ecosystems translate into State changes, how State changes act to Impacts and to define corresponding Responses.

First, ArcView Generalized Watershed Loading Function (AVGWLF) was used to calculate nutrient loadings e.g. Pressures (nitrogen and phosphorus) from surrounding NA watershed. Second, using machine learning (ML) tools (Weka and Cubist) State of marine

ecosystem was defined trough development of phytoplankton and mucilage descriptive models in a form of regression trees. In the third step the State of NA marine ecosystem was linked to Pressures from surrounding watershed also with the use of ML tools. The link between Pressures and States is a fundamental prerequisite to achieve a real ecosystem-based approach of marine ecosystems. For this task was used Multi Target Stepwise Model Tree Induction (MTSMOTI). MTSMOTI simulates chlorophyll-a (Chl-a), dissolved inorganic nitrogen (DIN), total phosphorus (P_{tot}) and oxygen saturation (Osat) in NA from which trophic conditions of the ecosystem are being evaluated using the TRIX. Through evaluation of various scenarios which present different management options e.g. responses in the watershed (e.g. different wastewater treatment level, fertilizers with less nutrients in agriculture etc.) and through their influence on marine ecosystem (TRIX) the proper level of wastewater treatment was determined which is significant response in reduction of nutrients.

Naslov teme doktorskog rada: "Određivanje odgovarajućeg stupnja pročišćavanja otpadnih voda s obzirom na stanje morskog ekosustava"

Autor: Goran Volf dipl. ing. građ.

Mentor: Prof.dr.sc. Boris Kompare

Ko-mentor: Prof.dr.sc. Nevenka Ožanić

Ključne riječi: sjeverni Jadran, strojno učenje, kvaliteta vode, stanje ekosustava, upravljanje slivom, tretman otpadne vode.

Sažetak

Jedan od najvažnijih izvora zagađenja morskih ekosustava je onaj proizveden od strane ljudskih aktivnosti u pripadajućim slivovima. Razumijevanje povezanosti između kakvoće morskih ekosustava i okolnih slivova važno je kako bi se bolje procijenili procesi u morskim ekosustavima te evaluirale različite mogućnosti upravljanja s ciljem poboljšanja stanja morskog ekosustava. Cilj ovog doktorskog rada je pridonijeti postizanju operativnog, na ekosustavu baziranog upravljanja morskim ekosustavima, koji je dan u okviru Marine Strategy Framework Directive (2008/56/EC), podizanju znanstvene spoznaje o funkcioniranju morskih ekosustava, osobito pod djelovanjem različitih pritisaka iz okolnih slivova, analiza pojedinačno ili zajedno, kako antropogenih tako i prirodnih pritisaka, kao što je obogaćenje nutrijentima. Pozornost se posvećuje procjeni odgovarajućeg stupnja pročišćavanja otpadnih voda u skladu sa stanjem morskog ekosustava.

Za područje istraživanja izabran je sjeverni Jadran. Sjeverni Jadran najpliće je, a također i jedno od najproduktivnijih područja u Jadranu, kao i na cijelom Mediteranu. Okolno slivno područje sjevernog Jadrana rasprostire se preko četiriju susjednih zemalja, Italije, Slovenije, Hrvatske i Švicarske. Ovo slivno područje odlikuje se različitim pritiscima iz okoline, kako prirodnim tako i antropogenim kao što su poljoprivreda i urbanizacija.

Da bi se postigao cilj ovog doktorskog rada korišteni se principi bazirani na DPSIR modelu (Driving forces-Pressures-States-Impacts-Responses, odnosno Pokretne sile-Pritisci-Stanja-Utjecaji-Odgovori), posebno definiranje Pritisaka, kako se ti Pritisci velikih morskih ekosustava prenose na promjene Stanja, kako promjene Stanja djeluju na Utjecaje, te definiranje odgovarajućih Odgovora.

Prvo su korištenjem Arc View Generalized Watershed Loading Function (AVGWLF) izračunati nutrijenti (dušik-N i fosfor-P), tj. Pritisci na cijelom slivnom području sjevernog

Jadrana. Drugo, upotrebom alata strojnog učenja (Weka i Cubist) definirano je Stanje morskog ekosustava izradom modela dinamike koncentracije fitoplanktona i modela cvjetanja mora (TIN/PO_4 model) danih u obliku regresijskih stabala. U trećem koraku povezna su Stanja morskog ekosustava sjevernog Jadrana sa Pritiscima iz slivnog područja korištenjem alata strojnog učenja. Veza između Pritisaka i Stanja temeljni je preduvjet za postizanje objektivnog, na ekosustavu baziranog principa upravljanja morskim ekosustavima. Za ovaj zadatak korišten je Multi Target Stepwise Model Tree Induction algoritam (MTSMOTI). MTSMOTI koristeći vrijednosti o količinama ukupnog fosfora i ukupnog dušika u slivu simulira, tj. računa klorofil, otopljeni anorganski dušik, ukupni fosfor i zasićenje kisikom u sjevernom Jadranu preko kojih se zatim ocjenjuju trofična stanja ekosustava korištenjem TRIX-a. Provedbom evaluacije različitih scenarija koji predstavljaju različite opcije upravljanja tj. Odgovora u slivnom području (npr. različiti stupnjevi pročišćavanja otpadnih voda, gnojivo s manjom koncentracijom nutrijenta poljoprivredi i sl.) te njihovim utjecajem na morski ekosustav (TRIX) definirani su odgovarajući stupnjevi pročišćavanja otpadnih voda, što predstavlja značajan Odgovor kod smanjenja količine nutrijenata.

Table of Contents

BLANK PAGE

List of figures

List of tables

BLANK PAGE

List of mostly used abbreviations

AVGWLF	ArcView Generalized Watershed Loading Function
DEM	Digital Elevation Model
DPSIR	Driving forces-Pressures-States-Impacts-Responses
EPA	Environmental Protection Agency
EEA	European Environment Agency
EU	European Union
EQ	Ecological Quality
GES	Good Environmental Status
GIS	Geographic Information System
GWLF	Generalized Watershed Loading Function model
ML	Machine Learning
MSFD	Marine Strategy Framework Directive
MTSMOTI	Multi Target Stepwise Model Tree Induction
N	Nitrogen
NA	Northern Adriatic
OECD	Organization for Economic Co-Operation and Development
P	Phosphorus
UNEP	United Nations Environment Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
UWWTD	Urban Waste Water Treatment Directive
WFD	Water Framework Directive
WWT	Wastewater Treatment
WWTP	Wastewater Treatment Plant

Chapter 1

Introduction

1.1 Motivation

The subject of this Ph.D Thesis is dealing with the issue of preserving marine ecosystems so that they can have normal functioning. One of the most important sources of pollution in marine ecosystems is the one related to human activities in the associated river watersheds (e.g. wastewater treatment plants, agriculture etc.). Understanding the relationship between water quality and pressures from river watersheds is important in order to better assess the processes in marine ecosystems and to evaluate different management options in watersheds aimed at improving the state of marine ecosystems. Introduction of suitable management options in watersheds, such as agricultural practices and wastewater treatment, it is expected that the state of marine ecosystems will improve.

1.2 Problem definition-state of the art

Marine ecosystems are home to a host of different species ranging from tiny planktonic organisms that comprise the base of the marine food web (i.e., phytoplankton and zooplankton) to large marine fishes and mammals. In marine ecosystems nutrients and light are required to produce food and energy. However, both nutrients and light are limiting factors in ecosystem productivity (Barnes and Hughes, 1999).

Marine ecosystems are very important in to the overall health of both marine and terrestrial environments. According to the World Resources Center (http://earthtrends.wri.org/searchable_db/index.php?theme=1), coastal habitats alone account for approximately one third of all marine biological productivity, and estuarine ecosystems are among the most productive regions on the planet. In addition, other marine ecosystems such as coral reefs, provide food and shelter to the highest levels of marine diversity in the world.

To preserve marine ecosystems it is of crucial importance to focus on limiting human-caused damage and on restoring the damaged ones.

1.2.1 Management of marine ecosystems

Many different pressures (e.g. urbanization, agriculture, etc.) from surrounding watersheds are influencing European marine ecosystems, providing at the same time many benefits, so that their sustainable management appears as a complex issue, requiring the integration of knowledge on the functioning of ecological, economic and social systems.

European marine ecosystems are being degraded as a consequence of continuously increasing pressure from anthropogenic activities, which include nutrient enrichment. This nutrient enrichment (industrial pollution, urban growth and tourism) may lead to eutrophication, intensification of maritime activities, fishery and aquaculture. Many European Union (EU) reports (e.g. European Environment Agency, EEA, 2001b, 2003a, 2003b, 2006) have highlighted the state of the marine environment, its increasing vulnerability and the need for further and stricter regulation of nutrient release. Eutrophication, a product of anthropogenic nutrient enrichment of water bodies, is an important impact affecting the integrity of European seas (EEA, 2001a, 2003a). The EU has adopted several directives and policies intended, directly or indirectly, to struggle with eutrophication (e.g. Nitrates Directive, Urban Waste Water Treatment Directive, Water Framework Directive, Common Agricultural Policy and Marine Strategy Framework Directive). The EU's Water Framework Directive (WFD, 2000/60/EC) is coordinating much of this action within national and international watershed-scale ("River Basin District") boundaries. Special measures and interventions (policy and legal reforms, investments in nutrient reduction technology at source) have been planned and partially implemented through coordinated international and national actions by the regional seas conventions and their secretariats (Barcelona Convention in the Mediterranean, HELCOM in the Baltic, OSPAR in the North Sea, and Bucharest Convention in the Black Sea), together with associated projects and programmes. The Urban Waste Water Treatment Directive (UWWTD, 91/271/EEC) which has strong relationship with WFD gives objective to protect the environment from the adverse effects of urban waste water discharges and discharges from certain industrial sectors and concerns the collection, treatment and discharge of: (1) Domestic waste water, (2) Mixture of waste water and (3) Waste water from certain industrial sectors. Marine Strategy Framework Directive (MSFD, 2008/56/EC) adopted in June 2008 provides environmental quality targets (Good Environmental Status; GES) that cannot be achieved without tackling eutrophication. Task Group 5 Report adopted in April 2010 (part of MSFD) is designed to provide guidance for the interpretation and application of the Eutrophication Quality Descriptor (QD5), one of eleven quality descriptors (1) Biological diversity, (2) Non-indigenous species, (3) Population of commercial/shell fish, (4) Elements of marine food webs, **(5) Eutrophication**, (6) Sea floor integrity, (7) Alteration of hydrographical conditions, (8) Contaminants, (9) Contaminants in fish and seafood for human consumption, (10) Marine litter and (11) Introduction of energy, including underwater noise) required for evaluation of GES in the MSFD.

The solution proposed by the MSFD is an operational "Ecosystem-Based Approach", in its broadest meaning, to the management of marine ecosystems, coupling sustainable use, conservation and socio-economic issues. The goal of the Directive is to achieve a GES for European seas by year 2020, achieving the full economic potential that society can obtain from marine ecosystem services, in a way which is sustainable and in harmony with the environment.

According to the MSFD, which has established European Marine Regions, Member States must develop strategies to define and then achieve a GES for their marine regions, through cooperation with other Member States or non-EU countries whom they share the regions with. Clear environmental targets must be set, and monitoring programmes should be established in order to assess regularly the progressions made towards such goals. Such progressions should be made by State Members by means of technically feasible, cost-effective measures, based on prior impact assessments and cost-benefit analyses in order to evaluate their effectiveness and consequences.

In order to "apply an Ecosystem-Based Approach to the management of human activities, ensuring that the collective pressure of such activities is kept within levels compatible with the achievement of GES and that the capacity of marine ecosystems to respond to human-induced changes is not compromised, while enabling the sustainable use of marine goods and services by present and future generations" (MSFD, 2008/56/EC), it is extremely important to understand how ecological, economic and social systems are interconnected. A prerequisite to it is the knowledge of how marine ecosystems function.

A contribution aimed at increasing of understanding of functioning marine ecosystems, in order to control and manage the activities in the contributing areas especially for determining the proper wastewater treatment (WWT) level according to its state is the goal of this Ph.D Thesis.

One of the specific European marine ecosystems which will be presented in this Ph.D Thesis is the northern Adriatic (NA). The NA is a shallow enclosed basin located between Italy, Slovenia and Croatia. It receives large fresh water discharges particularly from Italian rivers (e.g. Po River, Adige etc.) that drain intensely developed watersheds. The NA is one of the most productive areas in the Mediterranean, and in the 1980s suffered severe eutrophication as evidenced by hypoxia and fish kills (Vollenweider *et al.*, 1992; Degobbis *et al.*, 2000).

1.2.2 Wastewater Treatment Plants (WWTP) and their influence on marine ecosystems

In this Ph.D Thesis the main goal is focused on determining of proper WWT level (with respect to nutrient removal only) according to marine ecosystem state. WWTP are facilities designed to speed up the natural purification process that occurs in natural waters and to remove contaminants in wastewater that might otherwise interfere with the natural process in the receiving waters. This section presents the WWT levels, Directives concerning WWT and the influences of WWTP on marine ecosystems, negative as well as positive ones.

The alternative methods for municipal WWTP are classified into three major categories (Shun Dar Lin, 2007):

1. **Primary** (physical process) treatment,
2. **Secondary** (biological process) treatment and
3. **Tertiary/Advanced** (combination of physical, chemical and biological process).

As presented in Figure 1.1 each category should include previous treatment devices (preliminary), disinfection (if necessary) and sludge management (treatment and disposal). The treatment devices in the preliminary treatment are not necessarily to be included depending on the wastewater characteristics and regulatory requirements.

Briefly description of wastewater treatment levels is given in text bellow:

Preliminary; Removal of wastewater constituents such as rags, sticks, floatables, grit, and grease that may cause maintenance or operational problems with the treatment operations, processes, and ancillary systems.

Primary (1); Removal of portion of the suspended solids and organic matter from wastewater.

Secondary (2); Removal of biodegradable organic matter (in solution or suspension) and suspended solids.

Tertiary/Advanced (3); Removal of specific wastewater constituents which cannot be removed by secondary treatment individual treatment processes are necessary to remove nutrients (nitrogen and phosphorus), additional suspended solids, refractory organics, heavy metals and dissolved solids.

Disinfection; the purpose of disinfection in the treatment of wastewater is to substantially reduce the number of microorganisms in the water to be discharged back into the environment for the later use of drinking, bathing, irrigation, etc.

In table 1.1 are presented percents of pollutant removal by each WWT level.

Table 1.1 Removal of pollutants by WWT level in percents [%] (Shun Dar Lin, 2007)

WWT Level / Pollutant	Preliminary	Primary	Secondary	Tertiary
Biological Oxygen Demand	30	50-70	90-95	>95
Total Suspended Solids	60	80-90	90-95	>95
Total Nitrogen	15	25	40	>90
Total Phosphorus	5	10	30	>90

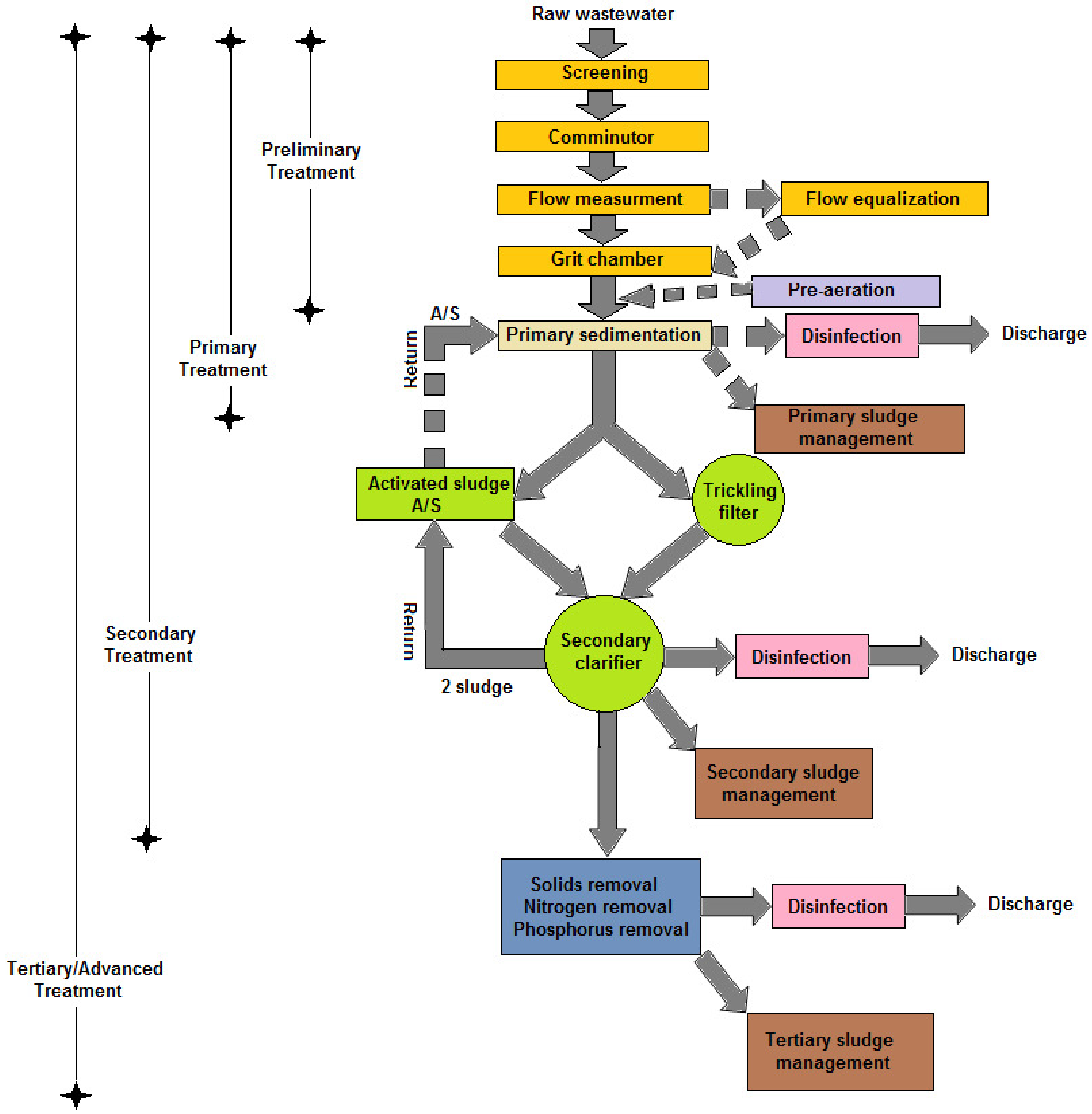

Figure 1.1 Flow chart of most common wastewater treatment processes (Shun Dar Lin, 2007)

WWTP influences on marine ecosystems can be positive and/or negative. When speaking about influences the focus is put on nutrients (nitrogen and phosphorus) influences which are one of the most important factors for marine ecosystem productivity. Under negative influences can be placed direct responses such as changes in primary production, algal biomass, sedimentation of organic matter, altered nutrient ratios, harmful algal blooms, and indirect responses such as changes in benthos biomass, benthos community structure, benthic macrophytes, habitat quality, water transparency, sediment biogeochemistry, mortality of aquatic organisms, food web structure etc. Increase in phytoplankton biomass and the resultant decrease in transparency and light intensity can become an indirect response that limits growth of submerged vascular plants (Cloern, 2011). Positive influences are not frequently mentioned at the present time. They are mainly visible in stimulating natural productivity, e.g. in oligotrophic seas. Among desirable changes in phytoplankton-based systems may be put an increase in benthic animals and the production of harvestable fish, at least up to some point at which hypoxia or anoxia may outweigh the positive influence of a grater food supply (Nixon and Fulweiler, 2009). For instance, a small input of nutrients to the oligotrophic southern Adriatic would increase the production of harvestable fish, while this wouldn't be the case for its north-western part which is considered eutrophic.

It is the occurrence of hypoxia and anoxia that is best documented and understood and, perhaps, most severe impact of eutrophication. The link between nutrient inputs and accelerated organic production resulting in low oxygen is the most common concern for managers and marine ecologists.

As mentioned in text above (see sub-section 1.2.1.) the directive dealing with wastewater treatment is UWWTD (91/271/EEC). Objectives of this directive are to **protect the environment from the adverse effects of urban waste water discharges and discharges from waste water from certain industrial (agro-food) sectors.** This concerns the **collection, treatment and discharge of urban waste water, treatment and discharge of waste water from industrial sectors.** UWWTD proposes the following WWT:

- Secondary treatment (i.e. biological treatment involving organic carbon removal).
- Additional nitrogen (N) and phosphorus (P) removal ("advanced treatment") in sensitive areas, i.e. basically water bodies being eutrophic or tending to be eutrophic.
- Exceptions possible in less-sensitive areas, i.e. certain marine areas, and in high mountain areas.

Water body is defined as a sensitive area if it falls into one of the following groups:

- Natural freshwater lakes and other freshwater bodies, estuaries and coastal waters that are eutrophic or which in the near future may become eutrophic if protective actions are not taken.
  - For freshwater bodies-removal of P, and for big agglomerations removal of N, unless it can be demonstrated that the removal will have no effect on the level of eutrophication.
  - For estuaries, bays and other coastal waters-removal of P and/or N, unless it can be demonstrated that the removal will have no effect on the level of eutrophication.
- Surface freshwaters intended for abstraction of drinking water which could contain more than 50 mg/l concentration of nitrate (Drinking Water Abstraction Directive; 75/440/EEC).
- Areas where further treatment than secondary is necessary to fulfil Council Directives.

Under less sensitive areas by UWWTD can be put a marine water body or area if the discharge of wastewater does not adversely affect the environment as a result of morphology, hydrology or specific hydraulic conditions which exist in that area. The following elements shall be taken into consideration when identifying less sensitive areas: open bays, estuaries and other coastal waters with a good water exchange and not subject to eutrophication or oxygen depletion or which are considered unlikely to become eutrophic or to develop oxygen depletion due to the discharge of urban waste water.

In this Ph.D Thesis WWT level will be determined according to marine ecosystem state through defining the trophic conditions (from Ultra-oligotrophic to Hypereutrophic conditions). Proposal for WWT level focuses only on nutrient loads, but not on the effect of organic (carbon) load on marine ecosystems, namely, the effect of organic load was not observed in this research. Because of this a Secondary WWT level is needed for removal this load.

1.2.3 The Driving forces-Pressures-States-Impacts-Responses (DPSIR) framework

For the purpose of this Ph.D Thesis the interactions between pressures from watershed, state of the marine ecosystems, and responses will be identified and quantified through descriptive mathematical models. For the identification of processes, interactions and system functions, the well known DPSIR framework (from EEA, Gabrielsen and Bosch, 2003) will be used.

The EEA assesses the "State" (**S**) of the environment using the DPSIR methodology. Namely, the State (**S**) is the result of specific Drivers (**D**) and Pressures (**P**), positive or negative, which Impact (**I**) the environment. The Responses (**R**) represent the solutions (e.g. policies, investments) that should then be done to improve or maintain that state. The EEA report also looks at "Outlooks" (**O**) for the state of the environment-namely, what will happen to that state over time based on various scenarios.

The DPSIR framework is in some way a conceptual model (see Figure 1.2) representing direct interactions through a loop in the way that human being interacts with the environment. The conceptual model starts from the main Drivers which act on society and reflect social needs and economic demands: using an example from agriculture, such needs would be to maintain a high level of profits or of employment in that economical sector, or to satisfy the market demands for agricultural products. Human activities, i.e. agriculture in the example, are performed to satisfy the needs and the demands, resulting in Pressures on marine ecosystems. In a sense, Drivers cause Pressures. Pressures lead to modifications in the environment, so that its State is affected and changes. For example, pressures from agriculture (nutrient enrichment) can lead to an increase of the phytoplankton biomass which could lead to phytoplankton blooms and mucilage or to eutrophication of marine ecosystem. As told, Pressures from agriculture affect such States, which change, and this leads to Impacts.

Impacts are not only modifications of the environmental State, but they need to be defined with respect of the use that human being makes of the environment, i.e. Impacts are the consequences, caused by changes in ecosystem quality and State, for human welfare and for the social or economic benefits that human being obtain from ecosystems. From above example (agriculture) nutrient enrichment is a Pressure due to driving forces such as population increase or the need to make agricultural practices more efficient through fertilizers, and it can cause the onset of eutrophication (Impact). Eutrophication can lead to economically negative consequences for tourism due to the degradation of water quality (State change), but also to economically positive consequences for fisheries targeting small pelagic fish, whose stock could increase following the greater availability of food (more primary production could mean more zooplankton, which small pelagic feed upon).

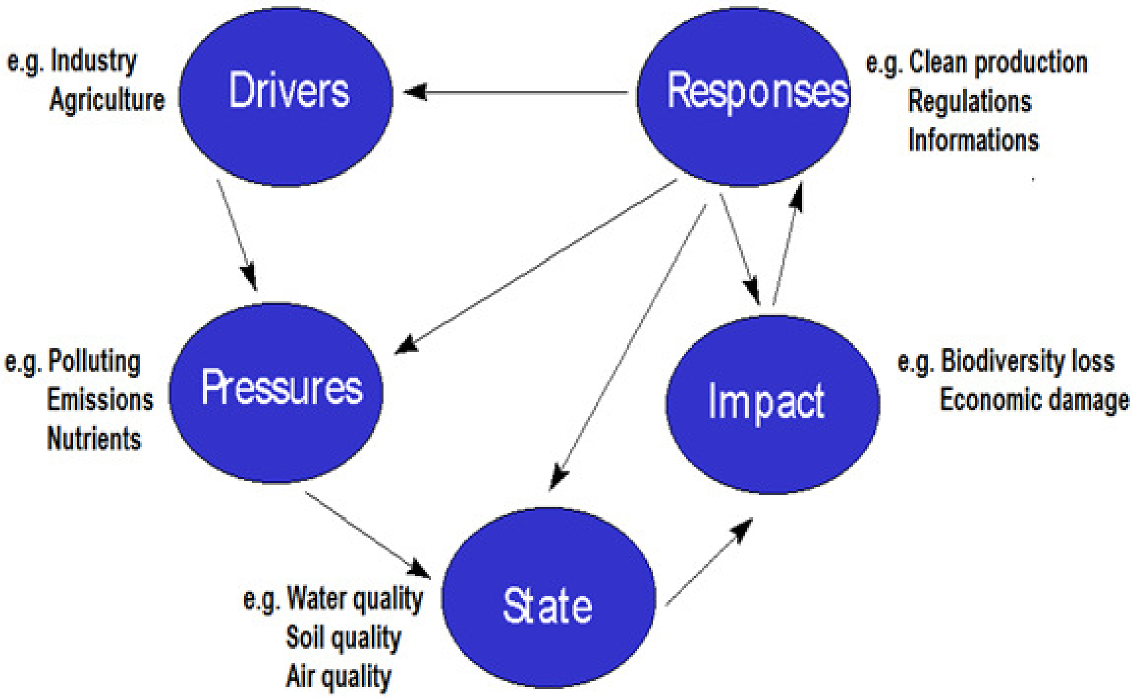

Figure 1.2 The DPSIR framework (Source: EEA)

The final step of the DPSIR framework is Responses that society or policy makers adopt in order to counteract the negative Impacts for human welfare. Such Responses may address: (1) Drivers, e.g. agriculture example, the need to maintain a high level of employment or production in the agriculture could be lowered by policies providing incentives to encourage the employment of farmers in other economic sectors, thus leading to a decrease of the agriculture effort, (2) Pressures, e.g. in the agriculture example, policies to enforce of use fertilizer with lower share of nutrients, (3) States e.g. agriculture example, eutrophication of the marine ecosystem because of anthropogenic nutrient enrichment, this step could be the dredging and removal of polluted sediments from the marine bottom, (4) Impacts, e.g. in the agriculture example, reduced economic incomes for farmers due to reduced harvesting may be compensated by government subsidies.

The presence of feedbacks going from the responses to the other steps of DPSIR conceptualization (Figure 1.2) highlights that the management of marine ecosystems is necessarily an adaptive process, where efficient solutions must be iteratively searched for, because of the contrasting issues that policy makers must solve, of the inherent variability of ecosystem dynamics and, not least, of the incomplete empirical and theoretical knowledge we have about the functioning of marine ecosystems. In this light, a computerized analytical tool that successfully simulates different managerial practices and subsequent marine ecosystem responses, may spare lot of time and enormous money compared to searching solutions by experimenting (iterating solutions) in full scale world.

1.3 Purpose of the thesis

The main purpose of this Ph.D Thesis is to define the proper WWT level (with respect to nutrient removal only) according to the desired marine ecosystem state. The work will be based on the parts of the DPSIR framework, in particular on defining Pressures (with an emphasis on WWT), how Pressures exerted on large-scale marine ecosystems translate into State changes, how State changes act to Impacts and to define corresponding Responses. Also Outlooks for the State of the environment will be done. Namely, what will happen to that state over time based on various scenarios. The goal is to contribute to achieving an operational ecosystem-based management of marine ecosystems, which has been called for in the MSFD (2008/56/EC), by improving the scientific knowledge of the functioning of marine ecosystems, especially under the actions of different pressures from surrounding watersheds, analyzed alone or together, and both anthropogenic and natural, such as, nutrient enrichment. Attention will be paid to defining the proper WWT level according to marine ecosystem state. The link between Pressures and States which is a fundamental prerequisite to achieve a real ecosystem-based approach will be done using machine learning (ML) tools. Relationship between State changes and Impacts will be taken into account only partially through conceptual model linking Pressures and States, and through providing management advices or evaluations for the case study. Corresponding Responses will be defined, like proper municipal WWT level and proper level of fertilizer application/washout in agricultural areas.

The main goal of this Ph.D Thesis is thus to conceptualize a managerial strategy for defining appropriate nutrient reduction measures in the watershed to sustain the desired marine ecosystem state. To accomplish the main purpose of this Ph.D Thesis several supporting tasks have to been done:

1. Calculate nutrient loads (Pressures) from watershed using a Geographic Information System (GIS) watershed model ArcView Generalized Watershed Loading Function (AVGWLF). The results from the model will further be used for defining the state of marine ecosystem.
2. Assessing the State of marine ecosystem by developing descriptive models. To understand the functioning of ecosystem it is of crucial importance to understand the ecosystem’s main biogeochemical and hydrological characteristic and process. Two descriptive models will be elaborated: (1) descriptive model for phytoplankton and (2) descriptive model for mucilage events. For managing purposes a predictive model for phytoplankton will be constructed.

3. Linking the Pressures to State. Using the calculated nutrient loads data and linking them to measured marine data, a model defining marine ecosystem state will be developed.
4. Create and evaluate scenarios to successfully control and manage the activities in the contributing watersheds especially for determining the proper WWT level according to marine ecosystem state (e.g. Responses).

A conceptual model which describes the tasks to accomplish main purpose of this Ph.D Thesis is presented on Figure 1.3. Pressures (nutrients) from surrounding watershed affect the State of the marine ecosystem. With proper management of Pressures in watershed (agriculture, WWTP, etc.) State of marine ecosystem can be controlled and thus achieved the desired functioning.

In each separate chapter short introduction is presented giving the description of the specific problem and review of literature dealing with that specific problem. The contributions to the World's science are given in Chapter 8 at the end of this Ph.D Thesis.

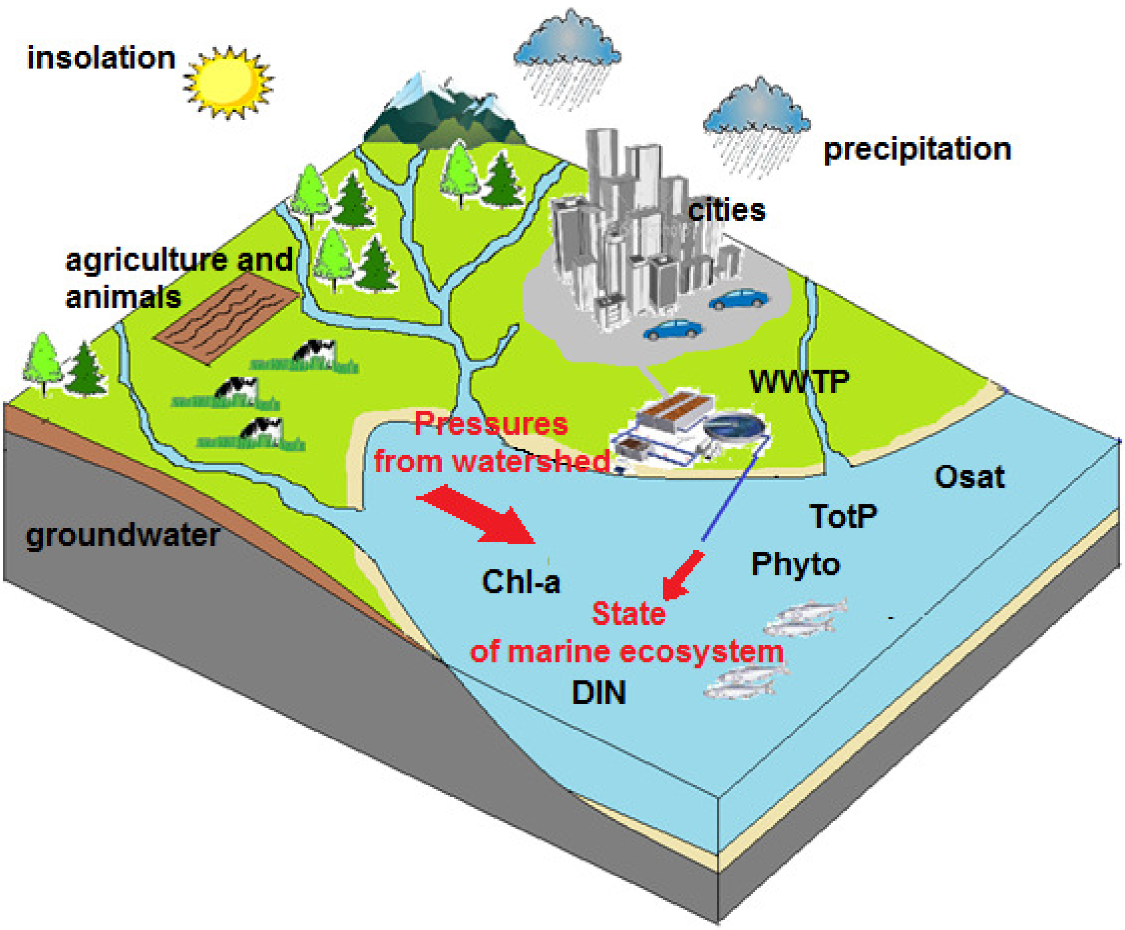

Figure 1.3 A conceptual model for defining the state of the marine ecosystem

1.4 Outline of the thesis

Most of the following chapters are based upon or literally represent papers which are already published (Volf *et al*., 2011: Volf, G., Kompare, B., Atanasova, N., Precali, R., Ožanic, N., Descriptive and predictive models of phytoplankton in northern Adriatic, Chapter 5), submitted for publication (Volf, G., Kompare, B., Atanasova, N., Precali, R., Ožanic, N., Relating mucilage events in northern Adriatic to nutrients ratios, Chapter 5; Volf, G., Kompare, B., Atanasova, N., Ožanic, N., Modelling nutrient loads to northern Adriatic, Chapter 4) or in final preparation for submitting (Volf, G., Kompare, B., Atanasova, N., Ožanic, N., Assessing the proper wastewater treatment level according to marine ecosystem state (northern Adriatic), Chapters 6 and 7).

The Ph.D Thesis is structured as follows:

Chapter 2: Study area and problem description

The chapter describes the case study area (NA basin with related watershed) with definition of the main problems of the case study area.

Chapter 3: Materials and methods

In the chapter are described methods, ML tools and GIS applications used in this Ph.D Thesis. Here are also presented data description used for modelling, their sampling and analysis.

Chapter 4: Modelling pressures (nutrients) to northern Adriatic

This chapter describes initial assumptions, boundary conditions and other relevant data leading to the results of modelling the nutrient loads e.g. pressures to NA using ArcView Generalized Watershed Loading Function watershed model.

Chapter 5: Assessing the state of northern Adriatic

Models describing the state of the NA ecosystem are presented in this chapter: (1) model describing phytoplankton dynamics in NA in period 1972 to 2007, (2) model describing and explaining mucilage events and (3) prediction model for phytoplankton 14 days in advance. The models are developed using ML tools.

Chapter 6: Linking the state of the northern Adriatic marine ecosystem to the pressures from surrounding watershed

The model defining the state of marine ecosystem for the NA has been developed here. The model contribute to the understanding the linkage between the activities in the surrounding watershed to the water quality in the marine ecosystem.

Chapter 7: Summary of the results and discussion

All experiments summary results done in this Ph.D Thesis are presented in this chapter. Chapter also presents scenarios evaluation and proposal for optimal watershed management together with determining the proper WWT according to marine ecosystem state. It is important to be able to predict the alternative outcomes of ecosystem state with changes in nutrient pressure as We aim for GES of the European seas by 2020.

Chapter 8: Conclusions and further work

Final conclusions and guidelines for further research, as original contributions of this Ph.D Thesis are given here.

Chapter 2

Study area and problem description

2.1 Why northern Adriatic (NA) for study area?

To be able to understand the functioning of an ecosystem it is of crucial importance to understand its main biogeochemical and hydrological characteristics and processes. The NA was chosen for this research because of several reasons:

1. NA is close to My working place, and has been studied already for a long time at the Center for Marine Research (CMR), Rovinj from which the data were obtained. Also for the NA watershed, like Po River one large amount of measured data is available.
2. Surrounding countries (Italy, Croatia and Slovenia) in the watershed are characterized by different anthropogenic pressures and levels of urbanization and agriculture, ranging from the strongly inhabited and agriculturally developed Po River watershed in Italy to the Croatian and Slovenian mostly natural areas.
3. NA is one of the most productive basins in the Mediterranean, shallow and characterized by wide inter-seasonal and inter-annual variations of environmental parameters (e.g. temperature, salinity) and circulation.
4. There is a variety of anthropogenic pressures which include coastal pollution, nutrient enrichment and commercial fishing. Huge nutrient loads discharged by the surrounding rivers cause eutrophication which is one of the main problems of marine ecosystems nowadays.

Therefore, from reasons above which constitute the NA as an unique ecosystem this appears to be a suitable case study area for advancing the ecosystem approach to the management of marine ecosystems and surrounding watersheds.

2.2 Study area description

2.2.1 Northern Adriatic basin

The Adriatic Sea (Figure 2.1) is subdivided into three regional basins (northern, central and southern), differing in bathymetry, physiography and biogeochemical features. The NA is the shallowest area, while its north-western part in particular is one of the most productive areas in the Adriatic, as well as in the Mediterranean (e.g. Sournia, 1973, Mozetič *et al.*, 2009). The NA is a semi-enclosed basin of about 32 000 km^2. The basin is narrow (210 km wide at maximum) and shallow (depth up to 100 m, 29 m on average).

Numerous rivers and streams discharge nutrient rich freshwaters into the NA shallow waters (Raicich, 1996). Semi-enclosed circulation, characterized by cyclonic and anticyclonic atmospheric eddies prevails during spring and summer, significantly reducing the water exchange rate with the remainder of the Adriatic Sea (Supić *et al.*, 2000, Grilli *et al.*, 2005). The Po River, with an average flow rate of 1 500 m^3/s, is the most important source of nutrients in the region with a load of about 12 10^9 mol/a (168 kt/a) of total nitrogen and 0.5 10^9 mol/a (15.5 kt/a) of total phosphorus, (Degobbis and Gilmartin, 1990). These discharges generated strong trophic gradients that were refined by dividing the NA into eutrophic, mesotrophic, and oligotrophic regions, where only the western coastal waters are considered as eutrophic (Hopkins *et al.*, 1999).

Circulation of the water masses is primarily driven by air-sea interactions and freshwater discharge (Artegiani *et al.*, 1997). Rivers exert a strong influence on the system, affecting circulation, leading to a very short residence time (less than 3.3 months on average, Artioli *et al.*, 2008), influencing the biological dynamics of low-medium trophic levels (Santojanni *et al.*, 2006) and causing eutrophication and related phenomena of anoxia (Justic *et al.*, 1987; Caddy, 2000; Degobbis *et al.*, 2000; Artioli *et al.*, 2008) through the high nutrient loads discharged, particularly by the Po River. Phytoplankton productivity in the NA is most likely to be P-limited as confirmed with bioassay studies and analyses of dissolved inorganic nutrients (Pojed and Kveder, 1977; Degobbis and Gilmartin, 1990). The changes in nutrient ratios in the surface layer of the NA, influenced by the Po River discharges, coincided with an increased frequency of mucilage events (formation of macroaggregates up to several meters long in the upper water column and surface or subsurface organic layers; Stachowitsch *et al.*, 1990, Precali *et al.*, 2005).

Many studies have been undertaken, resulting in a substantial amount of knowledge about the NA ecosystem and its productivity. Just a few decades ago, parts of the NA were eutrophic for most of the time during the year, but environmental protection measures put in force since that time are now giving noticeable results. Trends towards oligotrophication of the basin, particularly evident from mid winter to late summer, have been documented (Harding *et al.*, 1999, Degobbis *et al.*, 2000, Mozetič *et al.*, 2009) as a consequence of the reduction of the phosphorus load during the late 1980s (de Wit and Bendoricchio, 2001) and because of the Po River flow reduction in the last past years (Cozzi and Gianni, 2011). The reduction during the late 1980s was mainly the result of a gradual reduction of polyphosphate content in detergents (Provini *et al.*, 1992, Pagnotta *et al.*, 1995).

Figure 2.1 Adriatic Sea-northern, central and southern basins

The latest studies performed on long term data give strong evidence that the still common perception of the northern Adriatic as a very eutrophic basin is no longer appropriate, at least for its northern part and especially in recent years due to reduced nutrients' loads (Mozetič *et al.*, 2009). However, episodes of algal blooms and anoxia were still noted in the last two decades (Degobbis *et al.*, 2000, Precali *et al.*, 2005), indicating that eutrophic episodes may still prevail for shorter time in a long run of relatively stable mesotrophic or even oligotrophic conditions. Such events may then be rather to eutrophication attributed to sudden changes in nutrients' ratios (Mozetič *et al.*, 2009).

2.2.2. Northern Adriatic watershed

NA watershed (Figure 2.2) measures approximately 110 600 km^2 and is spread over four neighbouring countries, Italy, Slovenia, Croatia and Switzerland. These surrounding countries are characterized by different anthropogenic pressures and levels of urbanization, ranging from the strongly inhabited and intensively agriculturally exploited Po River watershed in Italy to the Croatian and Slovenian mostly natural areas. NA watershed consists of following sub-watersheds: Po River, Adige, Piave, Livenza, Tagliamento, Isonzo, Dragonja, Mirna, Brenta-Bacchiglione, tributary of lagune Marana-Grado, SW Istrian tributary and other smaller watersheds.

The biggest watershed belongs to the Po River and embraces an area of approximately 71 000 km^2 (64 % of whole NA watershed). Population in watershed is approximately 16 000 000 inhabitants. The river is 652 km long. The average annual rainfall in the area is about 980 mm/a with minimal values of 700 to 900 mm/a and maximal of 1 400 to 1 600 mm/a. The average annual temperature in this area is 13.9 ^{o}C. Typical mean values are 3 ^{o}C for winter and 25 ^{o}C for summer (Palmeri *et al.*, 2005). The climate may be classified as temperate sub-oceanic (warm temperate oceanic and sub-oceanic, partially sub-Mediterranean, in coastal areas) (Thornthwaite, 1948).

Some details about the major sub-watersheds of the NA watershed can be found in Table 2.1.

Table 2.1 Some details about the major sub-watersheds of the NA watershed

River (watershed) name	Watershed area [km^2]	River length [km]	Population [inhabitants]	Average discharge [m^3/s]
Po River[1]	71 000	652	16 000 000	1 600[4]
Adige[2]	12 100	409	1 700 000	200[4]
Piave[3]	4 100	220	381 000	132[5]
Tagliamento[3]	2 580	175	165 000	100[4]
Brenta-Bacchiglione[3]	5 840	174	1 450 000	71 (for Brenta)[5]
Livenza[3]	2 221	113	365 359	100[4]
Isonzo[3]	3 400	137	295 790	200[4]

Source: http://www.adbpo.it[1], http://www.bacino-adige.it[2], http://www.adbve.it[3], Raicich, 1994[4], Smith *et al.*, 2006[5]

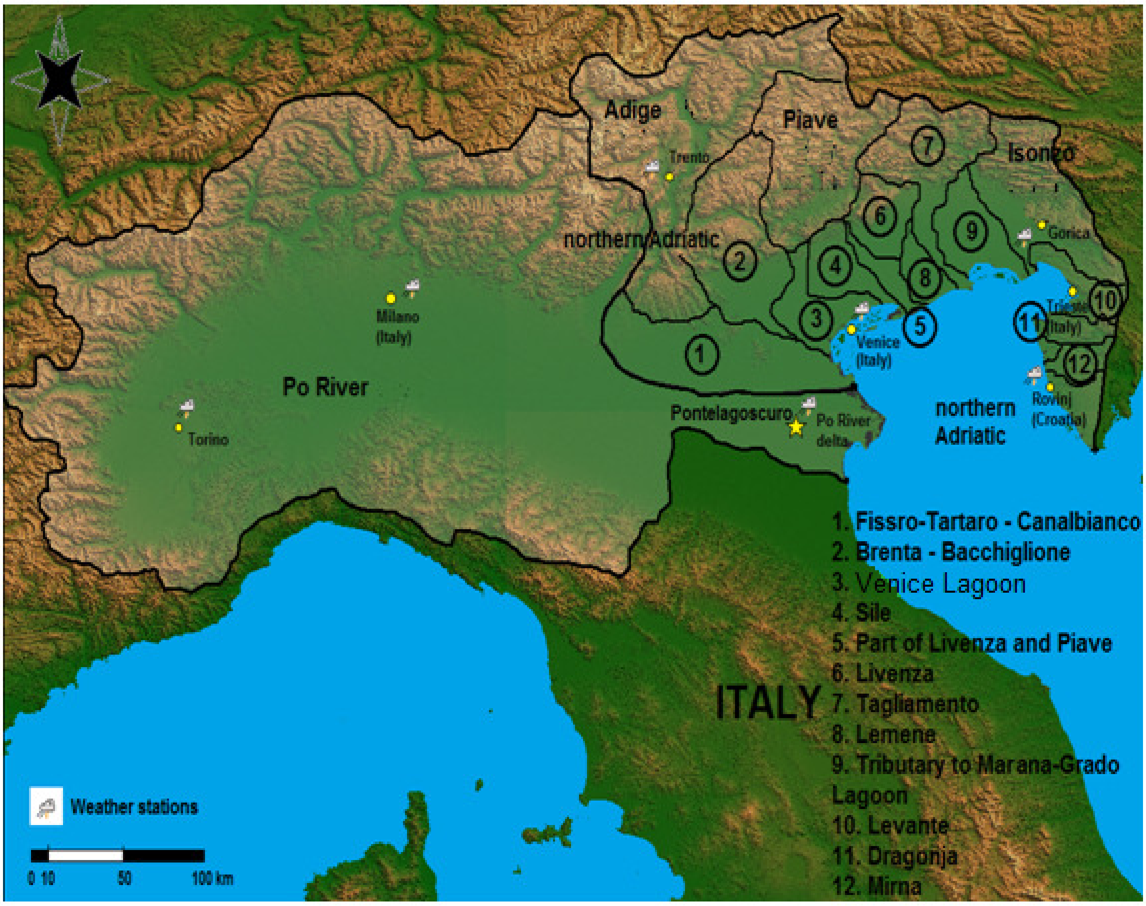

Figure 2.2 Northern Adriatic watershed

BLANK PAGE

Chapter 3

Modelling tools and data description

In the chapter are described methods, machine learning (ML) tools and Geographic Information System (GIS) applications used in this Ph.D Thesis. Here are also presented data description used for modelling, their sampling and analysis.

3.1 Modelling tools

3.1.1 Watershed simulation model

Modelling the impact of non-point source pollution in watersheds is a complex problem, and one that has concerned natural resource managers for many years. The development of spatially distributed hydrologic models has led to improved model forecasting at the cost of requiring more detailed spatial information. In addition, the analysis is much more sensitive to errors in the data, or lack of data. Incorporation of watershed models into a GIS was improved by streamlining data input and provided better interpretation of model outputs. Integration with GIS has shown to be necessary for the efficient and proper operation of models in resource management situations (Singh and Frevert, 2006).

Watershed simulation models, in fact, are commonly considered to be essential tools for evaluating the sources and controls of sediment and nutrient loading to surface waters. Such models provide a framework for integrating the data that describe the processes and land-surface characteristics that determine pollutant loads transported to nearby water bodies. Excellent historical overviews on the utility of computer models for quantifying and analysing pollution problems within watersheds throughout the country over the past three decades are provided by Moore (1991), Wilson (1996), Deliman *et al.* (1999), and Arnold *et al.* (2000).

3.1.1.1 ArcView Generalized Watershed Loading Function (AVGWLF)

Data manipulation and subsequent simulation modelling in this Ph.D Thesis is managed via an interface called AVGWLF (Evans *et al.*, 2002) between the GIS software package (ArcView) and the Generalized Watershed Loading Function model (GWLF; Haith and Shoemaker, 1987). The model is easy to use and relies on data input that is generally less exotic and easier to compile than other watershed-oriented water quality models such as Soil and Water Assessment Tool, Storm Water Management Model and Hydrologic Simulation Program Fortran (Deliman *et al.*, 1999). The model has also been endorsed by the United States Environmental Protection Agency as a good "mid-level" model that contains algorithms for simulating most of the key mechanisms controlling nutrient fluxes within a watershed (U.S. EPA, 1999).

The core watershed simulation model for AVGWLF GIS-based application is the GWLF model (see Figure 3.1). GWLF is a lumped, non-point source nutrient loading model in which the loading functions provide a practical compromise between simple empirical export coefficients that predict annual losses of nutrients to water and complex chemical simulation models that require unrealistically large amounts of detailed data for most practical applications at the catchment scale. GWLF was originally developed by Haith and Tubbs (1981) and validated by Haith and Shoemaker (1987) to simulate sediment, dissolved and total phosphorus and nitrogen loads from a watershed given variable-size source areas (e.g., agricultural, forested, and developed land). It also has algorithms for calculating septic system loads, and allows for the inclusion of point source discharge like wastewater treatment plants. It is a continuous simulation model which uses daily time steps for weather data and water balance calculations. Monthly calculations are made for sediment and nutrient loads, based on the daily water balance accumulated to monthly values.

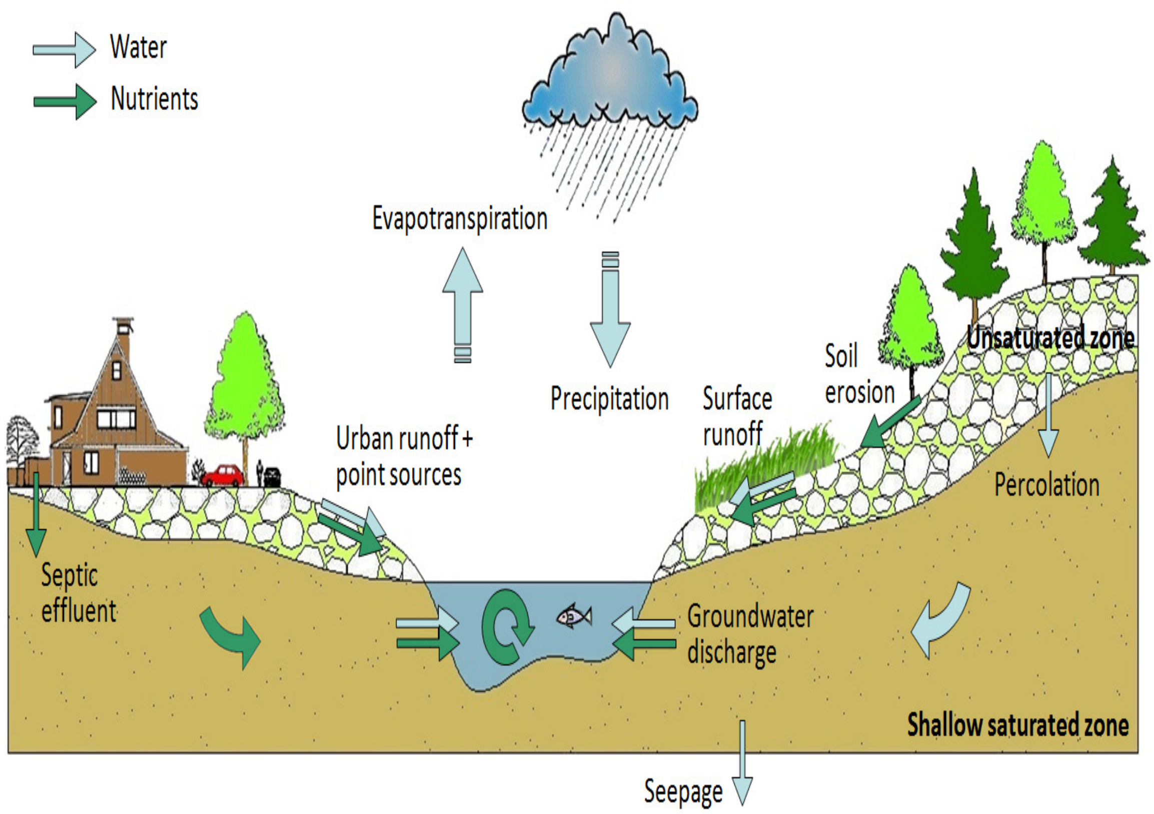

Figure 3.1 Schematic representation of the hydrologic and nutrient cycle in GWLF model

GWLF calculates dissolved liquid and solid phase nitrogen and phosphorous in stream flow using equations 3.1 and 3.2. Dissolved nutrient load is transported by runoff and eroded soil from various source areas, each of which is considered uniform with respect to soil and land cover.

$$LD_n = DP_n + DR_n + DG_n + DS_n \qquad (3.1)$$

$$LS_n = SP_n + SR_n + SU_n \qquad (3.2)$$

Where, LD_n and LS_n are the dissolved and solid phase nutrient load respectively (kg), DP_n and SP_n are the point source dissolved and solid phase nutrient load respectively (kg), DR_n and SR_n are the rural runoff dissolved and solid phase nutrient load respectively (kg), DG_n is the ground water dissolved nutrient load (kg), DS_n is the septic system dissolved nutrient load (kg), SU_n is the urban runoff nutrient load (kg).

For execution, the AVGWLF model requires three separate input files containing transport, nutrient, and weather-related data. The transport file defines the necessary parameters for each source area to be considered (e.g., area size, curve number, etc.) as well as global parameters (e.g., initial storage, sediment delivery ratio, etc.). The nutrient file specifies the various loading parameters for the different source areas identified (e.g., number of septic systems, urban source area accumulation rates, manure concentrations, etc.). The weather file contains daily average temperature and total precipitation values for each year simulated. Additionally we use retention file for nutrient retention in the watershed. The retention file allows users to account for the pollutant-attenuating effect of lakes, ponds and wetlands within the watershed being simulated (Evans *et al.*, 2002).

3.1.2 Machine learning tools

Branch of artificial intelligence concerned with the design and development of algorithms that allow computers to evolve behaviours based on empirical data, such as from sensor data or databases is ML. A learner can take advantage of examples (data) to capture characteristics of interest of their unknown underlying probability distribution. Data can be seen as examples that illustrate relations between observed variables. A major focus of machine learning research is to automatically learn to recognize complex patterns and make intelligent decisions based on data; the difficulty lies in the fact that the set of all possible behaviours given all possible inputs is too large to be covered by the set of observed examples (training data). Hence the learner must generalize from the given examples, so as to be able to produce a useful output in new cases (Witten and Frank, 2000). Kompare (1995) in his Ph.D Thesis gives some advantages of ML tools:

1. ML generalize the data and present their knowledge in a more compact, easier to understand,
2. Build new knowledge about the observed domain,
3. Identify the system structure and parameter values, and with it automatically build the model,
4. Search space for possible model behaviour with the use of qualitative modelling.

Tools of artificial intelligence build models independently, or help experts from certain areas in a way to mediate him information in a more compact form. With these new "views" expert can easily build a better model.

3.1.2.1 Weka

Weka (Witten and Frank, 2000) is a collection of machine learning algorithms for data mining tasks and contains tools for data pre-processing, classification, regression, clustering, association rules and visualization. It is also well-suited for developing new machine learning schemes. In this research Weka was used for developing the descriptive types of models (phytoplankton concentration model and mucilage model). Models were built in form of regression trees which put a single value of the target variable in the leaves what gives them more descriptive character or capabilities. Before going to explaining piecewise or *tree-structured regression*, something about linear regression will be told.

Linear regression is a method, which aims to express the dependent variable (also called a target variable, or class) as a linear combination of the independent variables (also called attributes or descriptors) from the given measurements (examples). Examples can be represented in a form of a table where each row (example) has the form ($a_1, a_2,\ldots,a_n, x$), where a_i are values of the N attributes (also independent variables or descriptors) and x is the value of the class. The task of the simple linear regression is to express the class value in form of given in equation 3.3.

$$x = a_1 * w_1 + a_2 * w_2 + \ldots + a_n * w_n = \Sigma a_i * w_i \qquad (3.3)$$

where w_i are weights, which are learned (calculated) from the *training set.*

While the simple linear regression calculates one equation (one weighing vector) for the entire data set, piecewise or *tree-structured regression* divides the data set into several subsets on which *uniform class value* or *linear equation* can be applied. The division to subsets is based on tests of the values of the input attributes which are put as nodes in a regression tree. Thus, regression trees are hierarchical structures composed of nodes and branches, where the internal nodes contain tests on the input attributes. Each branch of an internal test corresponds to an outcome of the test and the predictions for the values of the target variable (the class) are stored in the leaves which are the terminal nodes in the tree. If the leafs contain a single value for the class prediction, then we are talking about simple regression trees, while if a linear equation is used for prediction in the leaf, we are talking of model trees (Quinlan, 1992, Witten and Frank, 2000). Figure 3.2 illustrates the procedure of constructing regression and model trees.

One of the mostly used algorithm for induction of regression trees is the M5 algorithm (Quinlan, 1992), based on the top-down induction of decision trees algorithm (Quinlan, 1986). In Ph.D Thesis was used M5P algorithm implemented in Weka software for building regression trees. M5P (Wang and Witten, 1997) is a reconstruction of Quinlan's M5 algorithm (Quinlan, 1992) for inducing trees of regression models. M5P combines a conventional decision tree with the possibility of linear regression functions at the nodes. Decision-tree induction algorithm is used to build a tree, but instead of maximizing the information gain at each inner node, a splitting criterion is used that minimizes the intra-subset variation in the class values down each branch. The splitting procedure in M5P stops if the class values of all instances that reach a node vary very slightly, or only a few instances remain. The tree is pruned back from each leaf. When pruning an inner node is turned into a leaf with a regression plane. To avoid sharp discontinuities between the sub-trees a smoothing procedure is applied that combines the leaf model prediction with each node along the path back to the root, smoothing it at each of these nodes, by combining it with the value predicted by the linear model for that node. Techniques devised by Breiman *et al.* (1984) for their Classification and Regression Trees system are adapted in order to deal with, enumerated attributes and missing values. All enumerated attributes are turned into binary variables so that all splits in M5P are binary. As to missing values, M5P uses a technique called “surrogate splitting” that finds another attribute to split on in place of the original one and uses it instead. During training, M5P uses as surrogate attribute the class value in the belief that this is the attribute most likely to be correlated with the one used for splitting. When the splitting procedure ends all missing values are replaced by the average values of the corresponding attributes of the training examples reaching the leaves. During testing an unknown attribute value is replaced by the average value of that attribute for all training instances that reach the node, with the effect of choosing always the most populous sub-node. M5P generates models that are compact and relatively comprehensible. More about instance-based learning can be found in (e.g. Stanfill and Waltz, 1986; Aha *et al.*, 1991).

After the tree is constructed from the training (learning) set of data, it is necessary to assess the model quality, e.g., the accuracy of prediction. This can be done by simulating the model on a testing set of data and comparing the predicted values of the target with the actual values. Another option is to employ cross-validation. The given (training) data set is partitioned on a chosen number of folds (n). In turn, each fold is used for testing, while the remainder (n-1 folds) is used for training. The final error is the averaged error of all the models throughout the procedure.

DATA SET (EXAMPLES)

AT1	AT2	AT3	...	TARGET (CLASS)
3.67	8.500	0.005	...	2.133
4.15	7.207	0.005	...	2.601
5.32	8.357	0.011	...	3.718
7.80	7.929	0.005	...	3.481
8.11	7.096	0.005	...	1.791
9.36	7.804	0.005	...	1.128
10.87	6.018	0.005	...	1.471
11.10	7.400	0.006	...	1.521
10.23	5.457	0.011	...	0.869
8.39	5.486	0.014	...	0.535
7.42	5.486	0.013	...	1.034
4.06	8.307	0.005	...	1.636
...	...	...	...	...

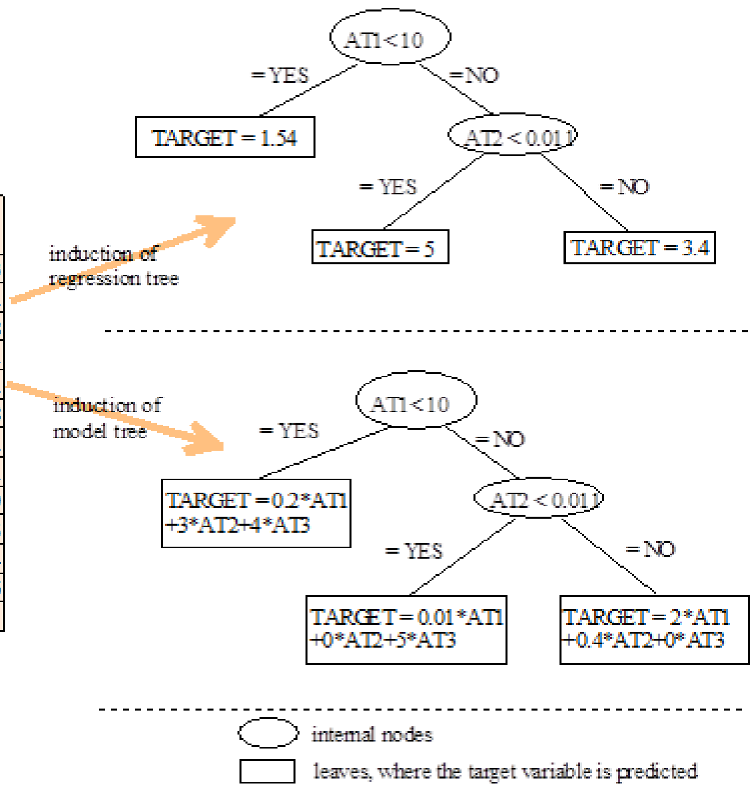

Figure 3.2 Induction of regression and model trees from given data set (examples)

The size of the error between the actual and the predicted values can be calculated by several measures to evaluate the model accuracy: root mean-squared error, mean absolute error, root relative squared error, relative absolute error, and correlation coefficient (R). In experiments of Ph.D Thesis the accuracy of the models is evaluated through the correlation coefficient.

3.1.2.2 Cubist

Cubist is a powerful tool for generating rule-based models that balance the need for accurate prediction against the requirements of intelligibility. Cubist models generally give better results than those produced by simple techniques such as multivariate linear regression, while also being easier to understand than neural networks.

Unlike the regression tree, the rule based regression models use regression equation in the terminal nodes which allow a more accurate prediction of the class value, but on the other hand they are less interpretable. This method was applied for prediction of phytoplankton concentration model.

Rule-based regression models for numeric prediction are yet another model representation, which is similar to the regression tree models. The models are interpreted as a set of IF-THEN rules where each rule is associated with a multivariate linear model. A rule indicates that, whenever a case satisfies all the conditions, the linear model is appropriate for predicting the value of the target attribute. The algorithms for rule induction mostly represent different variations of the M5 algorithm. The algorithm implemented in a software package Cubist (http://www.rulequest.com/cubist-info.html) was applied for modelling, in which the basic M5 algorithm was enhanced by combining the model-based and instance-based learning (Quinlan, 1992).

The accuracy of predictions can be done by simulating the model on a testing set of data and comparing the predicted values of the target with the actual values. Another option is to employ cross-validation. The given (training) data set is partitioned on a chosen number of folds (n). In turn, each fold is used for testing, while the remainder (n-1 folds) is used for training. The final error is the averaged error of all the models throughout the procedure.

The size of the error between the actual and the predicted values can be calculated by average error, relative error and correlation coefficient (R). The average error magnitude is straightforward enough. The relative error magnitude is the ratio of the average error magnitude to the error magnitude that would result from always predicting the mean value; for useful models, this is less than 1. The correlation coefficient measures the agreement between the cases' actual values of the target attribute and those values predicted by the model.

3.1.2.3 Multi Target Stepwise Model Tree Induction (MTSMOTI)

Prediction of several targets associated with a case is involved in many problems encountered in ecology, for example in this research was used for linking Pressures from surrounding watershed and State of the marine ecosystem.

The problem of predicting several target variables simultaneously has been approached in the predictive clustering framework by Blockeel *et al.* (1998), where now methods exist to construct clusters of examples which are similar to each other and simultaneously associate a predictive model (classification or regression) with each constructed cluster. Several systems have been developed to induce decision and regression trees (Blockeel *et al.*, 1998; Mehta *et al.*, 1996; Draper and Smith, 1982) or rules (Malerba *et al.*, 2004) within the predictive clustering framework, but only MTSMOTI (Appice and Džeroski, 2007) can induce a model tree to predict the values of several continuous target variables simultaneously.

If it is given a set of observed data in a form (a_1, a_2,…,a_n, x_1, x_2,…,x_n) where a_i are independent variables, the goal is to predict several target or dependent variables x_1,…,x_n, where the range of each dependent variable can be either a finite set of unordered category labels for classification or a subset of real number for regression.

As mentioned before (see sub-section 3.1.2.1.) model trees (Niblett and Bratko, 1986; Cestnik and Bratko, 1991; Karalic, 1992; Ceci *et al.*, 2003; Džeroski *et al.*, 2006) are decision trees whose leaves contain linear regression models which predict the value of a single continuous target variable. MTSMOTI is an algorithm, which is an extension of SMOTI (Malerba *et al.*, 2004) system that builds regression models. MTSMOTI unlike the SMOTI builds single multi target model trees, which are much smaller than the total size of the individual trees and they preserve accuracy in prediction. This single tree is induced much faster than the set of individual trees (Appice and Džeroski, 2007). Figure 3.3 illustrates the procedure of constructing multi target model trees.

DATA SET (EXAMPLES)

AT1	AT2	AT3	...	TARGET1 (CLASS)	TARGET2 (CLASS)	...
3.67	8.500	0.005	...	2.133	1.453	...
4.15	7.207	0.005	...	2.601	0.887	...
5.32	8.357	0.011	...	3.718	5.889	...
7.80	7.929	0.005	...	3.481	3.954	...
8.11	7.096	0.005	...	1.791	2.335	...
9.36	7.804	0.005	...	1.128	8.775	...
10.87	6.018	0.005	...	1.471	1.553	...
11.10	7.400	0.006	...	1.521	3.348	...
10.23	5.457	0.011	...	0.869	0.213	...
8.39	5.486	0.014	...	0.535	0.332	...
7.42	5.486	0.013	...	1.034	1.984	...
4.06	8.307	0.005	...	1.636	7.445	...
...	...	...	...	...	...	...

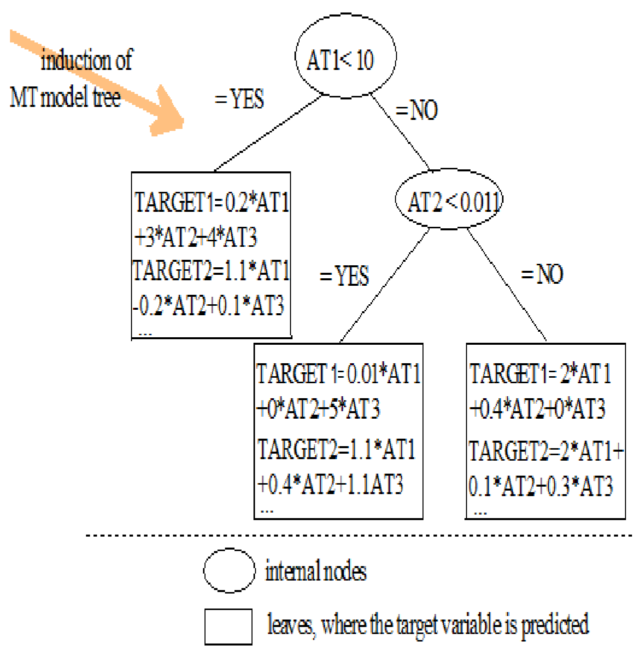

Figure 3.3 Induction of multi target model trees from given data set (examples)

After the tree is constructed, it is necessary to assess the model quality, i.e., the accuracy of prediction. This can be done by simulating the model on a testing set of data and comparing the predicted values of the target with the actual values or by employing cross-validation procedure.

The size of the error between the actual and the predicted values can be calculated by average relative mean square error, relative mean square error and correlation coefficient.

3.2 Data sampling, analysis and sources

3.2.1 Dataset used for modelling nutrient loads e.g. pressures

The dataset used for modelling the nutrient loads in the northern Adriatic (NA) watershed using AVGWLF model is composed of: (1) land use/cover data, (2) digital topographic data and soil maps (3) hydro-meteorological data, (4) population and wastewater data and (5) water quality and quantity data. Part of the watershed which belongs to Switzerland was not used in model analysis due to lack of data.

3.2.1.1 Land use/cover data

Watershed land use data is one of the most critical layers, since the pollutant loads emanating from a watershed are largely dictated by land surface conditions. AVGWLF use following 17 land use categories: water, low and high development, hay/pasture, row crops, other crops, coniferous forest, mixed forest, deciduous forest, wooded wetland, emergent wetland, quarry, coal mines, beaches, transitional, grass. Figure 3.4 presents the distribution of land use categories in NA watershed. The major categories in the NA sub-watersheds, together with their export coefficients are listed in Table 4.1, Chapter 4, Section 4.2. The export coefficient for each land-use type is expressed in mg/l and kg/ha/d. The land use layer and land use data are obtained from EEA (http://www.eea.europa.eu/publications/COR0-landcover) for years 2000 and 2006.

Agricultural land (cropland, hay/pasture) and urbanized areas are main diffuse sources of nutrients in NA watershed. As evident from Figure 3.4 approximately 47 % of the Po River watershed area is covered with agricultural land, whereas in Adige watershed only 17 % of all watershed area.

3.2.1.2 Topographic and soil data

Elevation layer (Figure 3.5) consists of topographic data and it is used to calculate land slope-related data for use within AVGWLF. The elevation layer was obtained from the Consortium for Spatial Information (http://srtm.csi.cgiar.org), where the spatial resolution is 1x1 km.

Soil maps are used to hold information pertaining to various soils-related properties important for calculating the nutrients' wash-off. The soil data were obtained from European Soil Portal (http://eusoils.jrc.ec.europa.eu). Specific fields required for this layer include:

1. available water-holding capacity of the soil in cm,
2. soil erodibility (or "K" factor) value for each soil unit,
3. dominant soil hydrologic group class for each soil unit ("A", "B", "C", or "D"), (Table 3.1),
4. soil organic matter content (%) and
5. soil N and P mass friction (mg/kg).

Table 3.1 Soil hydrological groups used in the GWLF model (Evans *et al.*, 2008)

Soil hydrologic group	Soil permeability (and runoff potential) characteristics	Soil texture
A	Soil exhibiting low surface runoff potential	Sand, loamy sand, sandy loam
B	Moderately course soil with intermediate rates of water transmission	Silty loam, loam
C	Moderately fine texture soils with slow rates of water transmission	Sandy clay loam
D	Soils with high surface runoff potential	Clay loam, silty loam, sandy clay, silty clay, clay

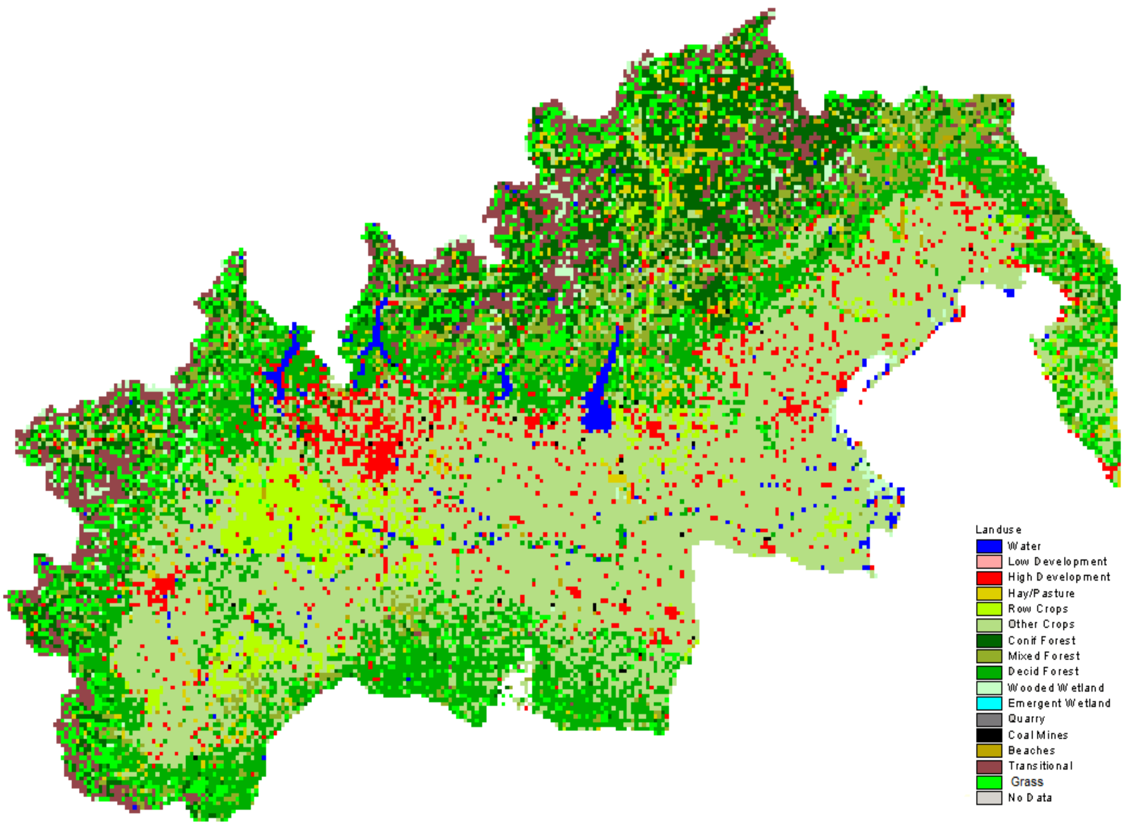

Figure 3.4 Land use/cover layer for NA watershed

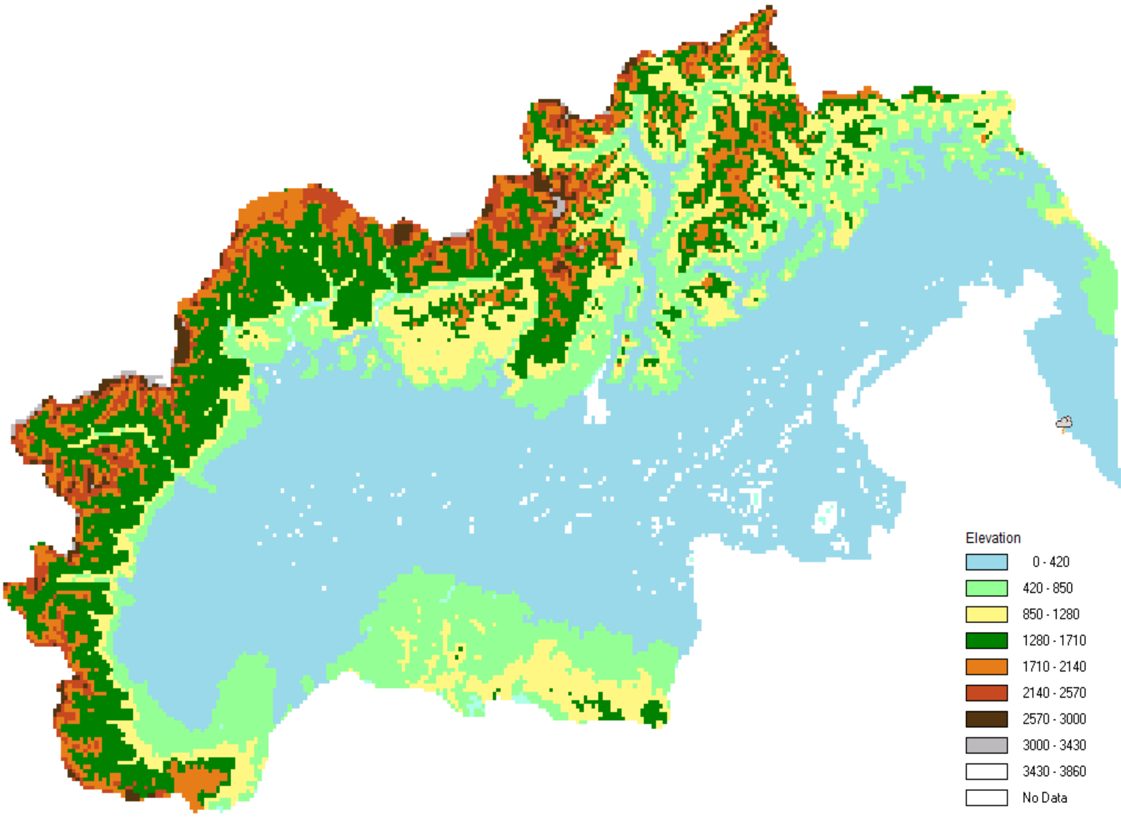

Figure 3.5 Elevation layer for NA watershed

3.2.1.3 Hydro-meteorological data

Hydro-meteorological data consists of daily precipitation in mm, air temperature in ^{o}C and daylight hours in h. The daily hydro meteorological data of the last 9 years (1999 to 2007) were provided by agencies Meteorological and Hydrological Service (Croatia), Slovenian Environment Agency (Slovenia) and Water Research Institute (Italy) and include the stations near Rovinj, Gorica, Venice, Pontelagoscuro, Trento, Milano and Torino. Geographical positions of the weather stations are shown on Figure 2.2 in Chapter 2, Section 2.2, Sub-section 2.2.2.

3.2.1.4 Population and wastewater generation data

To calculate the nutrient loadings from generated wastewater in the watershed following data were used (see Table 3.2):

1. total number of inhabitants,
2. inhabitants connected to sewers and
3. inhabitants connected and treated with WWTP.

The specific daily load per inhabitant for nitrogen (N) of 12.0 g/d and phosphorus (P) of 1.5 g/d (Tchobanoglous *et al.*, 2003) was taken into account to calculate the nutrients' loads in generated wastewater. This load is reduced for the inhabitants connected and treated with WWTP (see Chapter 4, Section 4.2).

Table 3.2 Number of inhabitants, inhabitants connected to sewers and inhabitants connected and treated with WWTP in some Italy regions in NA watershed

Region name	Population [inhabitants]	Inhabitants connected to sewers [%]	Inhabitants connected and treated with WWTP [%]
Piemonte	4 432 571	89.9	82.5
Lombardia	9 742 676	93.9	77.8
Veneto	4 885 548	78.1	78.7
Emillia-Romagna	4 337 979	84.8	78.8

Also for each urban area nutrient runoff coefficients were set up in the land use/cover layer (see Chapter 4, Section 4.2).

The data were provided by „Autorita di bacino del fiume Po" (www.adbpo.it/on-multi/ADBPO/Home.html), „Bacini idrografici delle alpi orientali" (www.alpiorientali.it), The National Institute of Statistics (www.istat.it), Slovenian Environment Agency (www.arso.gov.si) and Istrian Region (www.istra-istria.hr).

3.2.1.5 Water quality and quantity data

The measured data include flow values in m^3/s and loads of dissolved N and P and total N and P in t/a and t/mo for Po River at station Pontelagoscuro (Figure 3.7, Chapter 3, Sub-section 3.2.2.). Annual flow rates (Q_a) in m^3/s, dissolved inorganic nitrogen (*DIN*) and total phosphorus (P_{tot}) loads in t/a for Po River in period 1999 to 2007 are presented in Figure 3.6.

Data were provided by "Autorita di bacino del fiume Po" (http://www.adbpo.it) and collected from studies UNEP (1995) and" Autorita di bacino del fiume Po" (2008).

As the model calculates the flow rates and nutrients' loads in the rivers, these data were used to calibrate and validate the model. Measured flow rates and nutrients' loads in Po River from 1999 to 2002 were used for calibration, while the data from 2003 to 2007 for validating the model performance.

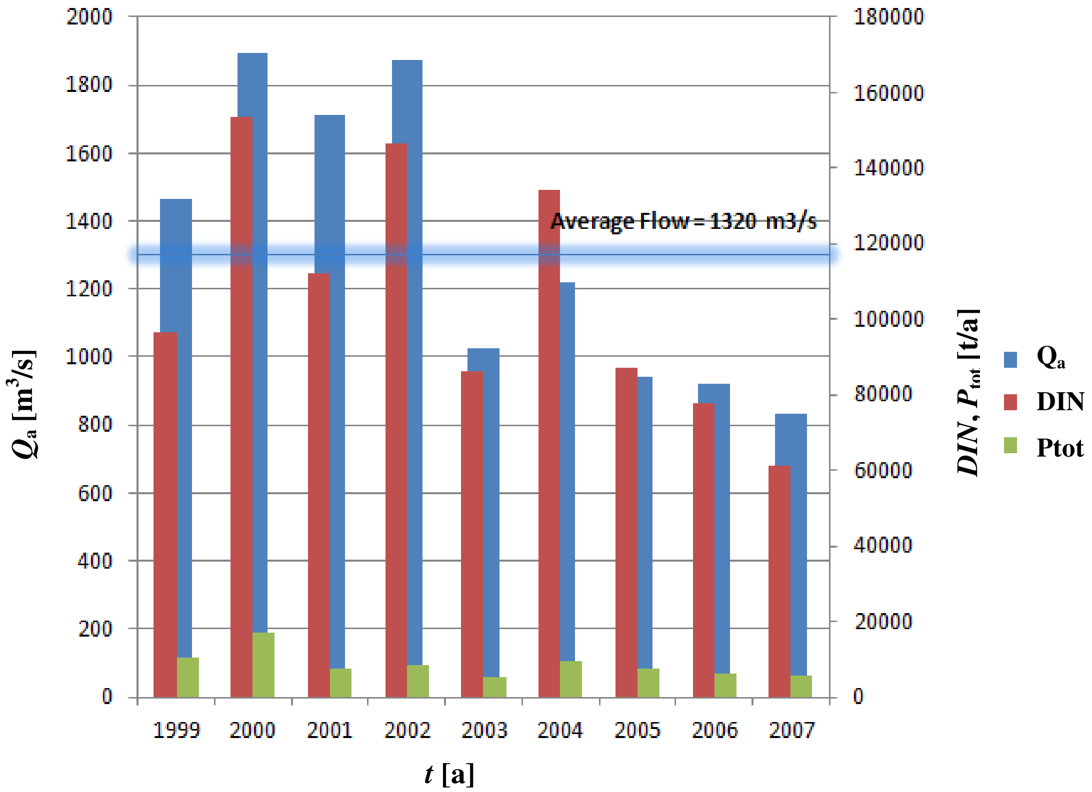

Figure 3.6 Annual flow rates (Q_a), dissolved inorganic nitrogen (DIN) and total phosphorus (P_{tot}) loads for Po River at Pontelagoscuro in period from 1999 to 2007

3.2.2 Data used for modelling the state of northern Adriatic

The data set used for modelling and interpretation comprises physical, chemical and biological parameters. Data were collected at six stations (SJ108, SJ101, SJ103, SJ105, SJ107 and RV001) on the profile from 12 Nm off the Po River delta to 1 Nm off Rovinj on the western Istrian coast by the Center for Marine Research (CMR) in Rovinj (Figure 3.5). This transect is 92 km in length, with station depths of 37 m and is considered representative for the shallowest part of the NA delimited by the line Cape Kamenjak-Rimini, (Revelante and Gilmartin, 1983, Degobbis *et al.*, 2000), approximately down to the 50 m isobaths with a surface area od about 19 000 km^2 and a volume of 635 km^3 (Degobbis and Gilmartin, 1990). Marked eutrophication gradients are often established between the predominantly mesotrophic north-western part of this region with its south-eastern part which is under the influence of oligotrophic waters originated in the central Adriatic.

The water column was sampled with 5 l Niskin samplers at 0.3, 5, 10 and 20 meters, and at 2 meters above the bottom from 1972 to 2007 with near monthly frequency.

Nutrient analyses (ammonium-NH_4, nitrite-NO_2, nitrate-NO_3, orthophosphate-PO_4, orthosilicate-SiO_4 and total phosphorus-TotP) were performed aboard the research vessel immediately after sample collection. The analyses were performed by methods used in oceanographic research defined by Strickland and Parsons (1972), using Beckman DU and Shimadzu UV mini-1240 and UV-1800 spectrophotometers with 10 cm cells. Method accuracies for NO_3, NO_2, NH_4, PO_4 and SiO_4 are 73 %, 73 %, 75 %, 73 % and 76 %, respectively, and detection limits are 0.05, 0.01, 0.1, 0.02 and 0.05 µmol/l, respectively. Total inorganic nitrogen (*TIN*) was calculated as the sum of NH_4, NO_2, and NO_3. Temperature (Temp) was measured with reversing thermometers, salinity (SAL) by Beckman RS 7c or Yeo-Kal MKII high precision salinometers in the ashore laboratory. Analysis of pH was performed also aboard the research vessel using Radiometer pH meters. The samples for total phytoplankton counts (Phyto; micro and nano fractions) were preserved with lugol solution and counted according to Utermöhl (1958) using Carl Zeiss inverse microscopes. Chlorophyll *a* (Chl-*a*) was determined fluorometrically after extraction from filters (GF/F) with acetone (Parsons *et al.*, 1985). The saturation percent of dissolved oxygen (Osat) in each water sample was calculated (from the quotient between the measured oxygen concentration and oxygen solubility) using the Benson and Krause equation (UNESCO, 1986).

Daily Po River flow (Q_{Po}) data measured at Pontelagoscuro, 90 km from the outlet (Figure 3.7) from January 1966 to December 2007 were obtained from the Agenzia Regionale Prevenzione e Ambiente dell'Emilia Romagna, Servizio Idrometeorologico, Parma. Data used for modelling the state of NA are presented in Table 3.3.

Table 3.3 Data used for modelling the state of NA

Parameter	Description	Unit
Q_{Po}	Po River flow	m^3/s
Temp	Temperature	°C
SAL	Salinity	
Dene	Density excess	kg/m^3
pH	pH	
NO_3	Moles of Nitrate as N	µmol/l
NO_2	Moles of Nitrite as N	µmol/l
NH_4	Moles of Ammonium as N	µmol/l
TotP	Total phosphorus	µmol/l
TIN	Total inorganic nitrogen (NO_3+ NO_2+ NH_4)	µmol/l
TIN/PO_4	Total Inorganic Nitrogen/Orthophosphate as P	mol/mol
TIN/SiO_4	Total Inorganic Nitrogen/Orthosilicates as Si	mol/mol
Chl-*a*	Chlorophyll *a*	µg/l
Phyto	Total phytoplankton	l^{-1}
Osat	Oxygen saturation	

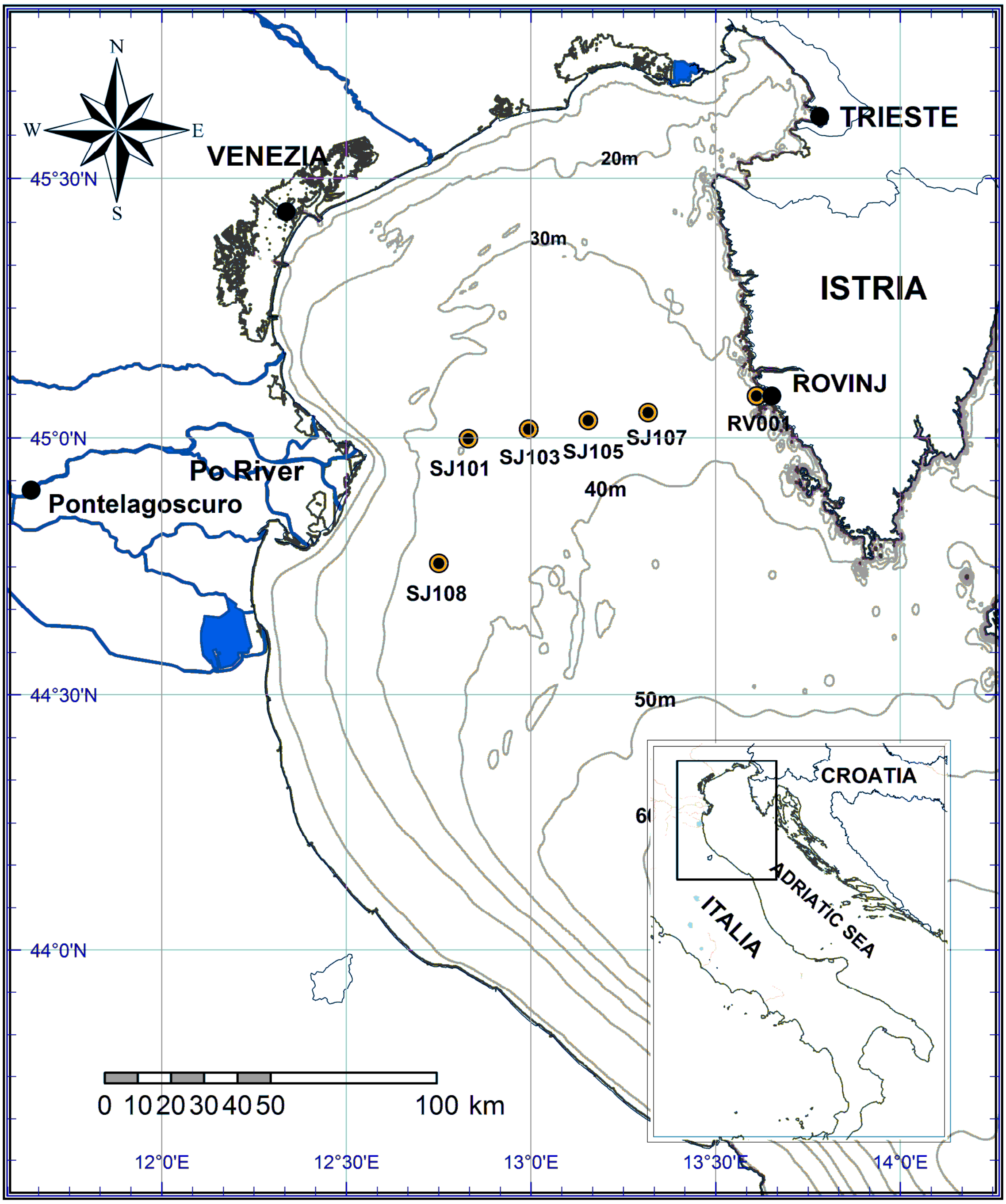

Figure 3.7 Geographic positions of measurement stations

Chapter 4

Modelling pressures (nutrients) from watershed to northern Adriatic

This chapter goes under part **P** e.g. Pressures of Driving forces-Pressures-States-Impacts-Responses (DPSIR) framework and is focused on a model-based quantitative assessment of the freshwater and associated nutrients fluxes to the northern Adriatic (NA). For this task the Geographic Information System (GIS) based nutrient loading model, ArcView Generalized Watershed Loading Function (AVGWLF, Evans *et al.*, 2002) to simulate the nutrient loadings in the given period from diffuse (different land uses) and point sources (wastewater treatment plants and urban drainage systems) was used. Compared to other watershed-oriented water quality models such as Soil and Water Assessment Tool, Storm Water Management Model, Modelling Nutrient Emissions in River Systems and Hydrologic Simulation Program Fortran, AVGWLF model is relatively easy to use due to its 'modest' requirements on data input and at the same time complex enough for our research goals, as it is capable of simulating most of the key mechanisms controlling nutrient fluxes within a watershed.

Increased nutrient concentrations in rivers, lakes and coastal seas as a consequence of various human activities such as agriculture or wastewater discharges have several undesirable effects, most of which are related to the increased growth of phytoplankton and other aquatic plants. The so-called eutrophication which presents the link between nutrients and increased organic production leads to a shift in the biological structure, and in severe cases even to oxygen depletion, production of toxins, and the collapse of entire aquatic ecosystems (OECD, 1982). These effects are more emphasized in shallow water bodies with poor water exchange; such is the northern part of the Adriatic Sea. Additionally, its north-western part is one of the most productive areas in the Adriatic Sea, as well as in the Mediterranean (e.g. Sournia, 1973; Mozetič *et al.*, 2009). Numerous rivers and streams discharge nutrient rich freshwaters into the NA shallow waters (Raicich, 1996). These rivers play important role in sustaining the marine productivity in the NA. Changes in NA riverine inputs are therefore potential drivers for long-term changes in the marine ecosystem.

The majority of the nutrients undoubtedly come with the Po River, which is the biggest contributing watershed to the NA. Thus, most of the latest studies are focused to the Po River watershed (de Wit and Bendoricchio, 2001; Palmeri *et al.*, 2005; Spillman *et al.*, 2007), while very few (Degobbis, 1988; UNEP, 1995; Cozzi and Giani, 2011) have been done for the entire

NA watershed. Degobbis (1988) in his Ph.D Thesis presents cycle and balance of nutrients for NA. UNEP (1995) deals with eutrophication, nutrient loads, source types, load assessment and effects on marine life. Nutrient loads and source types have been taken from several older papers and studies, and cannot be reliable for present state. Cozzi and Giani (2011) present the analysis of the runoff and nutrient loads by NA rivers, in order to point out their current impact on marine ecosystem. Nutrient loads were calculated using measured data. Not every river in NA watershed is monitored, so there could be some minor deviations in the results.

Previous authors (see above) who were dealing with the NA ecosystem dynamics have mainly taken into account only the contribution of the Po River. Standpoint of this Ph.D Thesis is that all nutrient loads shall be taken into account, i.e. all other watersheds may and really do significantly contribute. Moreover, such quantification of nutrients loads for all sub-watersheds is very important for successful and adaptive management of the NA water quality. This work has also been done to present detailed analysis of nutrient loads and source types for the entire NA watershed in period 1999 to 2007.

Nine year period from 1999 to 2007 of measured data, e.g., physical, chemical and biological parameters collected in the NA and its surrounding rivers, was used to calibrate and validate the model. The model was calibrated on period from 1999 to 2002 and validated on the rest of the data set (2003 to 2007). Indicating good validation results, the model was used to estimate (1) the quantities of nutrients released from each sub-watershed and thus providing an estimate of the importance of all watersheds compared to the biggest nutrient contributor, e.g. the Po River watershed, (2) the major sources of nitrogen and phosphorus in the NA watershed regarding the type of anthropogenic activity, enabling their control for in watershed management.

4.1 Model setup and calibration

The NA watershed area is modelled by dividing it into 17 sub-watersheds (Chapter 2, Section 2.2, Sub-section 2.2.2, Figure 2.2, Table 2.1). Following assumptions were made:

1. Average treatment efficiency has been adopted for all the watersheds by applying reduction coefficients (around 40 % for nitrogen (N) and around 30 % for phosphorus (P); secondary treatment; Shun Dar Lin, 2007) to emissions from collected and treated inhabitants with Wastewater Treatment Plant (WWTP). These reduction coefficients were estimated considering that WWTPs in the NA watershed show a great variability in efficiencies. Hence average treatment efficiency has been adopted for all sub-watersheds. In Po River watershed 80 % of WWTP have Secondary, 16 % have Tertiary and 4 % have Primary treatment of wastewater (www.adbpo.it/on-multi/ADBPO/Home.html)

2. For agricultural areas with flat landscapes (slope less than 0.5 %-using Digital Elevation Model-DEM) it was assumed that they are tile drained.
3. Atmospheric deposition of nutrients was not taken into account.
4. Nutrient concentrations in groundwater (GW) were estimated using land use/cover and soil (geomorphic conditions; highly or less porous soils) layer. For example, intensively-fertilized areas (e.g., cropland in row crops) underlain by highly porous material (e.g., fractured limestone or sandy soils) often exhibit sub-surface water concentrations of around 10 mg/l of nitrogen (Evans and Corradini, 2007). In AVGWLF, GW P is estimated using the groundwater N as a “surrogate” for identifying areas where levels of dissolved P may be high due to agricultural activities.

From the data described in Chapter 3, Section 3.2, the AVGWLF model calculates the flow rates in m^3/mo, sediment yield in kg/mo and nutrients’ (N and P) loads in kg/mo per sub-watershed. The model includes number of parameters that can be grouped as transport, sediment and nutrient parameters. Transport parameters influence the movement of the runoff and sediments from any given area in the catchment down to the NA. Transport parameters include: soil erodibility factor (K), slope length and steepness factor (LS), cover factor (C), management factor (P), weighted curve number values (WCN), weighted average growing season evapotranspiration (WGET), weighted average dormant season evapotranspiration (WDET). The values of the transport parameters were taken as default or calculated values from the AVGWLF model.

Parameters for sediment yield estimation include the slope length and slope steepness parameters, together designated as LS factor. This factor determines the effect of topography on soil erosion and was estimated from the DEM (Arhounditsis *et al*., 2002).

Nutrient parameters include the export coefficients (expressed as concentrations in mg/l) from various land uses. These values were mostly taken from studies and literature for the observed area. Some values were taken from Evans *et al.* (2008), Haith and Shoemaker (1987), Adeka *et al.* (2007), Jennings *et al.* (2009) and George (2010) for different source areas which are more or less representative of the study area (Table 4.1).

The calibration of the model’s parameters was performed so that optimal fit is obtained between the modelled and the measured annual and monthly values of the flow and nutrient concentrations in the Po River measured at Pontelagoscuro (Figure 3.6). Of all parameters listed above, only export coefficients were calibrated using the data from 1999 to 2002. The calibrated values of the parameters are presented in Table 4.1.

During the calibration process the effect of lakes, ponds and wetlands for nutrient retention in the watershed was taken into account. The percentage of lakes, ponds and wetlands area was calculated using ArcView application (using retention file).

Table 4.1 Export coefficients (*K*) for each land use type in NA before and after the calibration process

Land cover type	Values from literature		Values after calibration	
	KN_{dis} [mg/l]	KP_{dis} [mg/l]	KN_{dis} [mg/l]	KP_{dis} [mg/l]
Hay/pasture	1-3 [1]	0.1-0.5 [1], 0.010-0.015 [4,5]	2.1	0.014
Row crops	0.6-4.1 [2]	0.11-0.95 [2], 0.013 [5]	2.9	0.015
Other crops	0.6-4.1 [2]	0.11-0.95 [2], 0.013 [5]	2.5	0.011
Coniferous forest	0.1-0.2 [1]	0.006-0.012 [1], 0.004 [4]	0.19	0.006
Deciduous forest	0.1-0.2 [1]	0.006-0.012 [1]	0.21	0.008
Mixed forest	0.1-0.2 [1]	0.006-0.012 [1]	0.20	0.007
Emergent wetland	0.19 [1]	0.006 [1]	0.19	0.006
Wooded wetland	0,19 [1]	0.006 [1]	0.19	0.006
Natural grassland	1.8 [3], 0.75 [1]	0.29 [1], 0.3 [3], 0.001 [4]	0.80	0.02
Transitional/non vegetated land	0.23 [2]	0.07 [2]	0.20	0.08
Mineral extraction sites	0.012 [1]	0.002 [1], 0.001 [4], 0.179 [5]	0.012	0.002
	N [kg/ha/d]	**P [kg/ha/d]**	**N [kg/ha/d]**	**P [kg/ha/d]**
High development	0.101 [1]	0.011 [1]	0.101	0.011
Low development	0.01 [1]	0.002 [1]	0.012	0.002

Source: Evans *et al.*, 2008 [1]; Haith and Shoemaker, 1987 [2]; Adeka *et al.*, 2007 [3]; Jennings *et al.*, 2009 [4]; George, 2010.

4.2 Results and discussion

Validation of the model was performed by simulating it on the data from the period 2003 to 2007. The simulations indicate good fit between the modelled and measured values of the flow rate at Pontelagoscuro (Po River, Chapter 3, Section 3.2, Figure 3.6). The correlation between measured and the modelled data is presented in Table 4.2.

Table 4.2 Coefficient of determination R^2 between modelled and measured data

Parameter	Description	Units	R^2
$Q_{avg.mo}$	Average monthly flow	m^3/s	0.91
$Q_{avg.a}$	Average annual flow	m^3/s	0.95
$N_{tot.mo}$	Total monthly nitrogen	kg/mo	0.82
$N_{tot.a}$	Total annual nitrogen	kg/a	0.91
$N_{tot.DIS.mo}$	Total monthly dissolved nitrogen	kg/mo	0.81
$N_{tot.DIS.a}$	Total annual dissolved nitrogen	kg/a	0.93
$P_{tot.mo}$	Total monthly phosphorus	kg/mo	0.90
$P_{tot.a}$	Total annual phosphorus	kg/a	0.92
$P_{tot.DIS.mo}$	Total monthly dissolved phosphorus	kg/mo	0.85
$P_{tot.DIS.a}$	Total annual dissolved phosphorus	kg/a	0.83

Apart from the variables in Table 4.2, it can be seen high match for average annual flow in the validation period 2003 to 2007 ($Q_{measured.avg.a}$ = 988 m^3/s, $Q_{model.avg.a}$ = 989 m^3/s).

4.2.1 Nutrient loads and major sources of nutrients in NA watershed

The loads in t/a of $N_{tot.a}$, $N_{tot.DIS.a}$, $P_{tot.a}$ and $P_{tot.DIS.a}$ together with the annual precipitations ($Prec_a$) in mm in the NA watershed are presented in Figures 4.1 and 4.2 respectively. The retention of nutrients in the watershed is around 25 % for $N_{tot.a}$ and 20 % for $P_{tot.a}$ (from AVGWLF retention file). The difference between $N_{tot.DIS.a}$ and $N_{tot.a}$ is approximately 55 % and between $P_{tot.DIS.a}$ and $P_{tot.a}$ is 48 %. Average load of $P_{tot.a}$ is 12 568 t/a (min. 10 647 t/a and max. 28 439 t/a), while the average load of $N_{tot.a}$ is 305 795 t/a (min. 211 618 t/a and max. 545 971 t/a). These average loads are very similar to the values found by Artioli *et al.* (2008), i.e. 351 000 t/a of N_{tot} and 12 000 t/a of P_{tot}. As suggested by “Autorita di bacino del fiume Po” (2008) phosphorus largely depends on erosion by flood events, while nitrogen directly on the discharged water volume due to its higher solubility.

A slight decrease of nutrient loads can be observed in the last 4 years of the simulation period (Figures 4.1 and 4.2). Similarly, Cozzi and Gianni (2011) claim that there is a decrease of nutrient loads after 2004. However, their estimations indicate smaller values than those obtained in this research. This is probably because Cozzi and Gianni (2011) used monitoring data, where not every river in NA watershed is monitored, or the discrepancy may also be, because of insufficient number of weather stations were used in the model of this research.

Monthly loads of $N_{tot.mo}$ and $P_{tot.mo}$ in t/mo averaged through the simulation period together with monthly precipitations ($Prec_{mo}$) in mm are presented in Figure 4.3. In average, months with highest nutrient loads are between August and January, mainly associated with higher precipitations that washout fertilizers from agricultural areas.

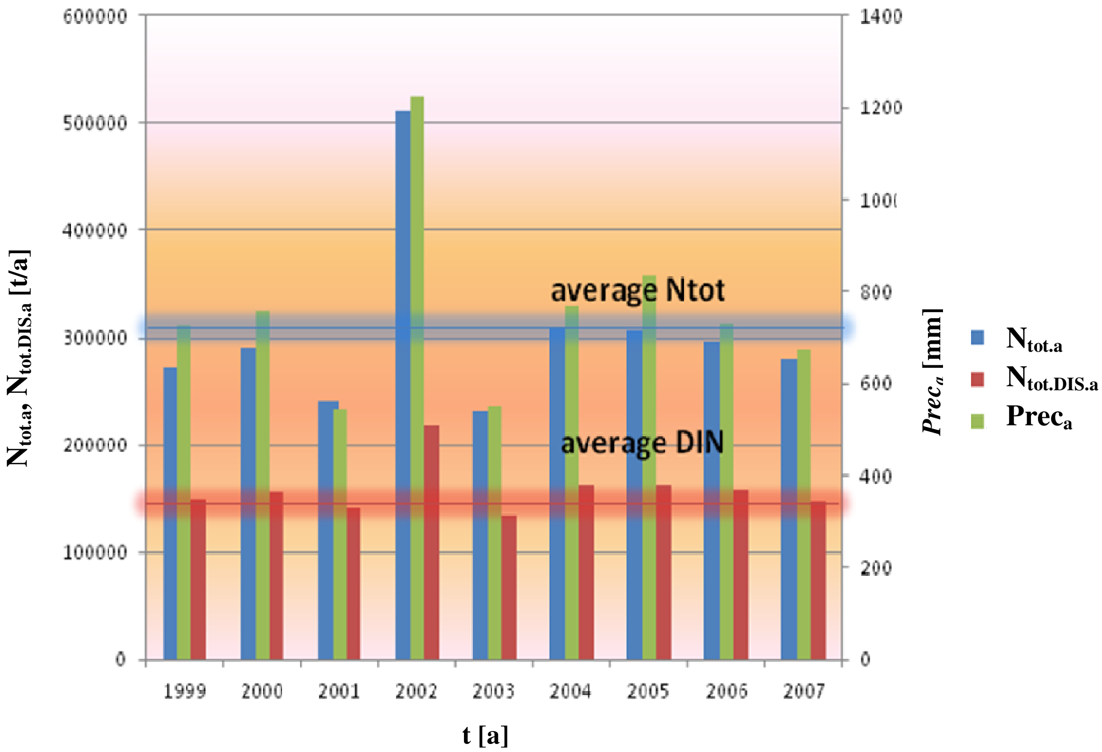

Figure 4.1 Simulated total annual nitrogen ($N_{tot.a}$) and dissolved annual nitrogen ($N_{tot.DIS.a}$) with annual precipitations (*Prec*$_a$) for the whole NA watershed

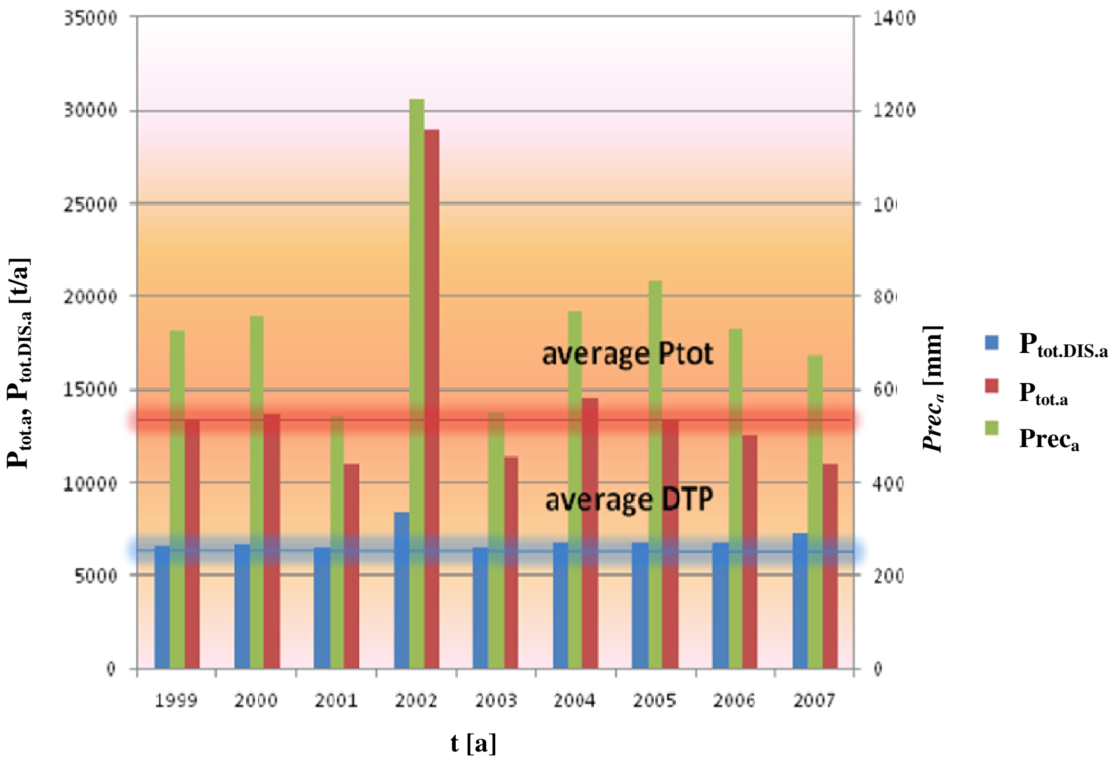

Figure 4.2 Simulated total annual phosphorus ($P_{tot.a}$) and dissolved annual phosphorus ($P_{tot.DIS.a}$) in with annual precipitations (*Prec*$_a$) for the whole NA watershed

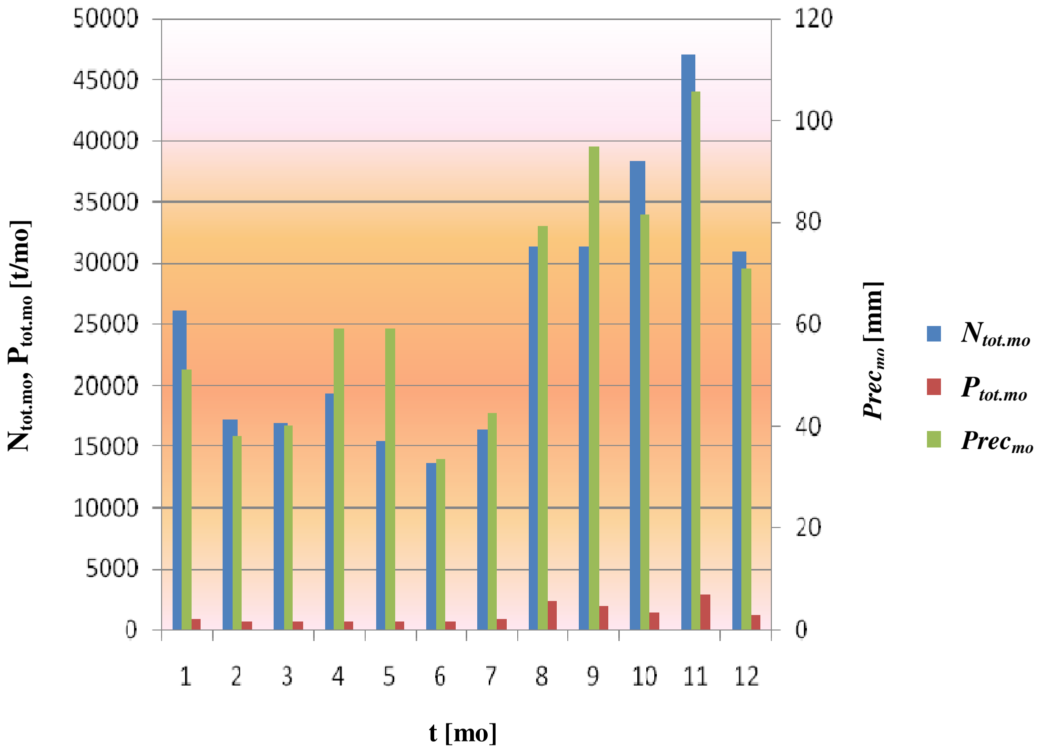

Figure 4.3 Simulated total monthly nitrogen ($N_{tot.mo}$) and total monthly phosphorus ($P_{tot.mo}$) with monthly precipitations ($Prec_{mo}$; average values for the whole period of modeled years)

Major sources of N in NA watershed are WWTP plus urban systems-US (27 %), tile drainage (agricultural terrains with tile drainage or drained by ditches)-TD (29 %) and groundwater-GW (29 %) (Figure 4.4), while for P major sources are WWTP plus urban systems-US (43 %) and cropland surfaces (30 %) (Figure 4.5). Cropland surfaces include row crops and other crops from the land use layer. This is slightly different from the UNEP (1995) study, where human activities contribute with 47 % for P and 20 % for N, while 44 % for P and 64 % for N comes from agriculture activities.

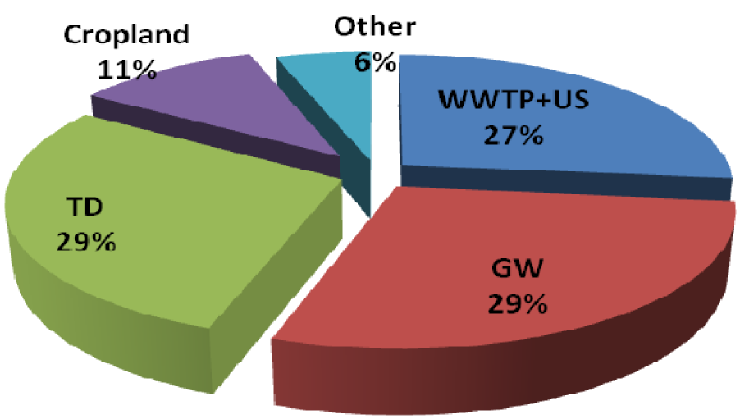

Figure 4.4 Major loading sources for nitrogen

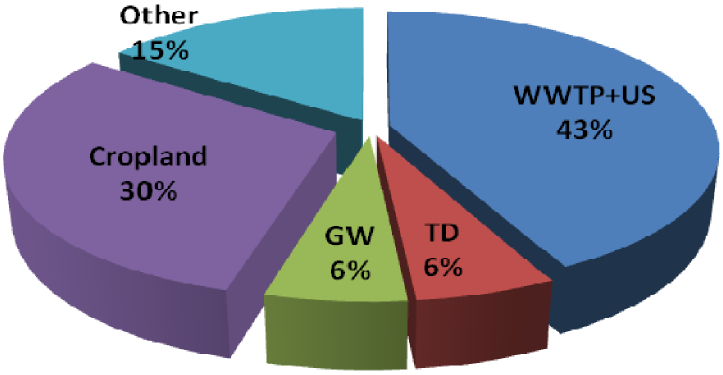

Figure 4.5 Major loading sources for phosphorus

4.2.2 Contribution of nutrient loads by watersheds

The results of this study confirm the predominant nutrient contribution of the Po River watershed in the NA with approximately 68 % of the entire load. However the 32 % of the other watersheds are not negligible and should be taken into account when managing the water quality of NA (Figure 4.6).

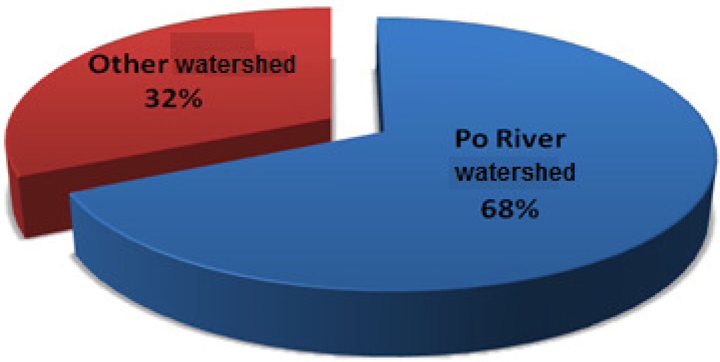

Figure 4.6 Percentage ratio of nutrient loading for NA (Po River vs. other watersheds)

Table 4.3 presents the average N_{tot} and P_{tot} from each sub-watershed in t/a, t/km^2 and in percents (%) of total load. Comparing the loads in percents of total load with the watershed area in percents we can conclude that some watershed area and nutrient loads have similar percents (like watersheds no. 1, 2, 6, 13) while other have higher or lower loads compared to watershed area.

Table 4.3 Total nitrogen (N_{tot}) and total phosphorus (P_{tot}) in t/a, t/km^2 and % compared with watershed areas (average values for the whole period of modelled years)

Watershed No.	Name of Watershed	Area [km²]	Area [%] of total area	N_{tot} [t/a]	N_{tot} [t/km²]	N_{tot} [%] of total load	P_{tot} [t/a]	P_{tot} [t/km²]	P_{tot} [%] of total load
1	Fissro-Tartaro-Canalbianco	2 885	2.6	11 610	4.0	3.8	321	0.1	2.1
2	Brenta-Bacchiglione	5 840	5.2	24 926	4.2	8.3	1 000	0.1	6.8
3	Venice Lagoon	550	0.5	9 146	16.6	3.0	568	1.0	3.8
4	Sile	846	0.7	931	1.1	0.3	42	0.04	0.2
5	Part of Livenza and Piave	546	0.4	886	1.6	0.3	39	0.07	0.2
6	Livenza	1 700	1.5	4 100	2.4	1.3	145	0.08	0.9
7	Tagliamento	2 900	2.6	5 273	1.8	1.7	133	0.04	0.9
8	Lemene	1 241	1.1	2 342	1.8	0.7	75	0.06	0.5
9	Marano - Grado Lagoon	1 275	1.1	2 844	2.2	0.9	78	0.06	0.5
10	Levante *	1 149	1.0	2 112	1.8	0.7	66	0.05	0.4
11	Dragonja **	96	0.09	477	4.9	0.1	18	0.1	0.1
12	Mirna	458	0.4	862	1.8	0.2	34	0.07	0.2
13	Po River	71 000	64.2	203 030	2.8	67.8	8 965	0.1	61.0
14	Adige	12 100	10.9	26 108	2.1	8.7	758	0.06	5.1
15	Piave	4 100	3.7	6 521	1.5	2.1	163	0.04	1.1
16	Isonzo	3 400	3.0	4 862	1.4	1.6	125	0.03	0.8
17	Part of NW Istria	495	0.4	886	1.7	0.3	38	0.07	0.2

* Levante watershed include also Timavo

** Under Dragonja are included also some smaller rivers in Slovenia (Rižana, Badaševica, Drnica)

Note that Results of the model for nutrient loads have been compared to other work much as possible (e.g. Cozzi and Giani, 2011 for Adige, Isonzo etc.; Kennish and Paerl, 2010 for Venice Lagoon)

When discussing specific loads in t/km^2, the watersheds with the biggest specific loads for N_{tot} are Fissro-Tartaro Canalbianco, Brenta-Bacchiglione, Venice Lagoon and Dragonja, not Po River. So, maybe the loads in Po River are difficult to be reduced more, but they can be reduced in other watersheds and contribute to the general reduction of nutrient loads in NA. The same goes for P_{tot}. Note that watershed with the biggest specific loads and most sensitive water body in NA is Venice Lagoon with 16.6 t/km^2 for N_{tot} and 1.0 t/km^2 for P_{tot}. Reduction efforts should be redirected there and not predominantly to Po River watershed like it is typically suggested. Po River watershed deserves special attention because of its size and should be efficiently controlled. But the yearly specific loads for N_{tot} of 2.8 t/km^2 and for P_{tot} of 0.1 t/km^2 are even lower then the values of some other sub-watersheds in NA, and quite possibly those high values could be easier reduced then the nutrient loads values of the Po River watershed.

In figures 4.7 and 4.8 are presented N_{tot} and P_{tot} in kg for each watershed area (average values for the whole period of modelled years). It can be concluded that loads of nutriets decrease from west (Po River) to east (Istran peninsula). Figures 4.9 and 4.10 presents specific N_{tot} and P_{tot} in kg/ha for each watershed area like it is presented and discussed in Table 4.3.

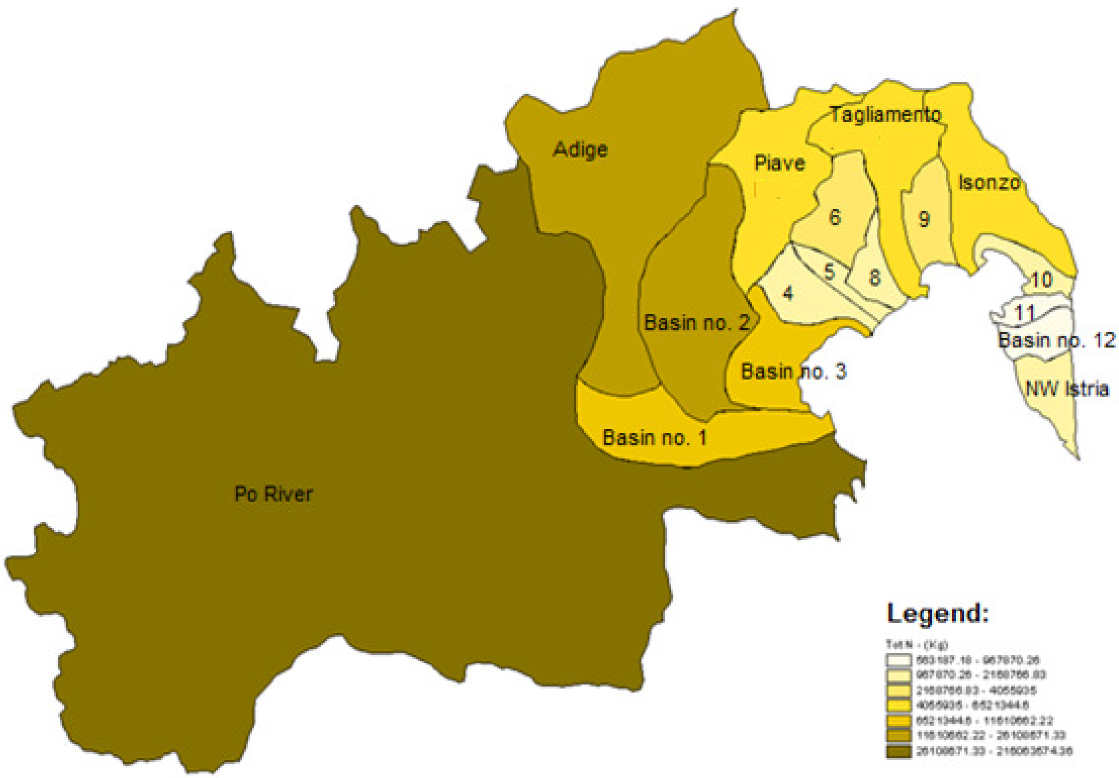

Figure 4.7 Total nitrogen (N_{tot}) for each watershed area (average values for the whole period of modelled years)

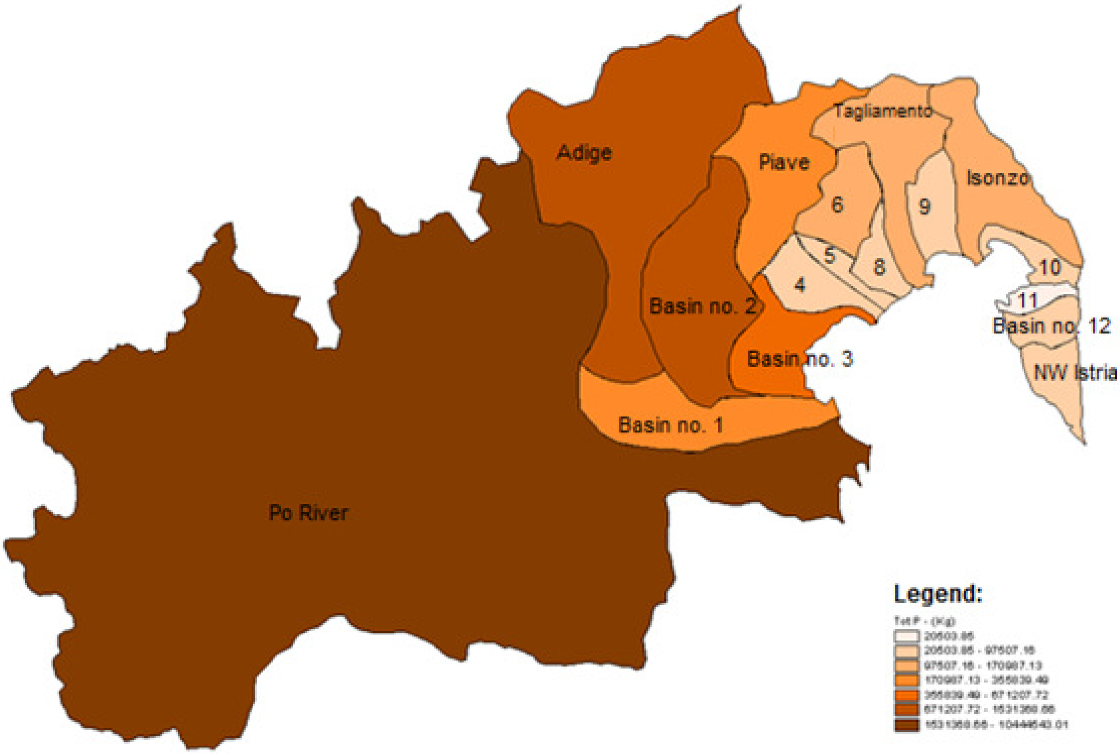

Figure 4.8 Total phosphorus (P_{tot}) for each watershed area (average values for the whole period of modelled years)

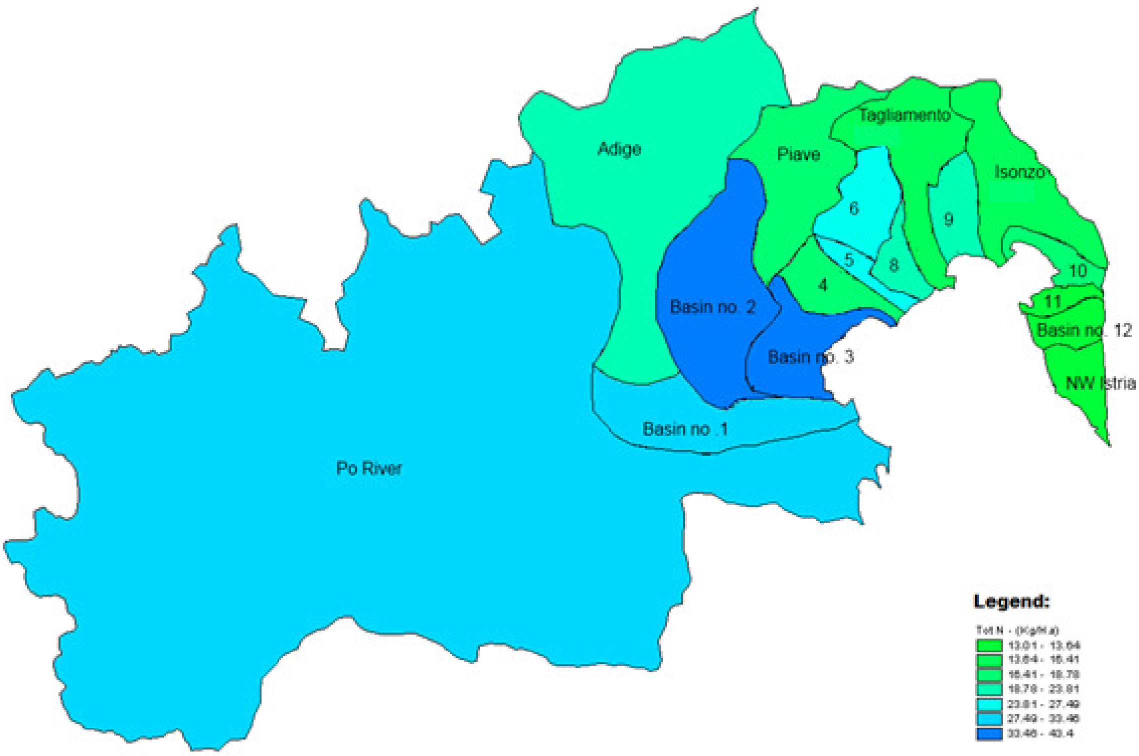

Figure 4.9 Specific loads of total nitrogen (N_{tot}) for each watershed area (average values for the whole period of modelled years)

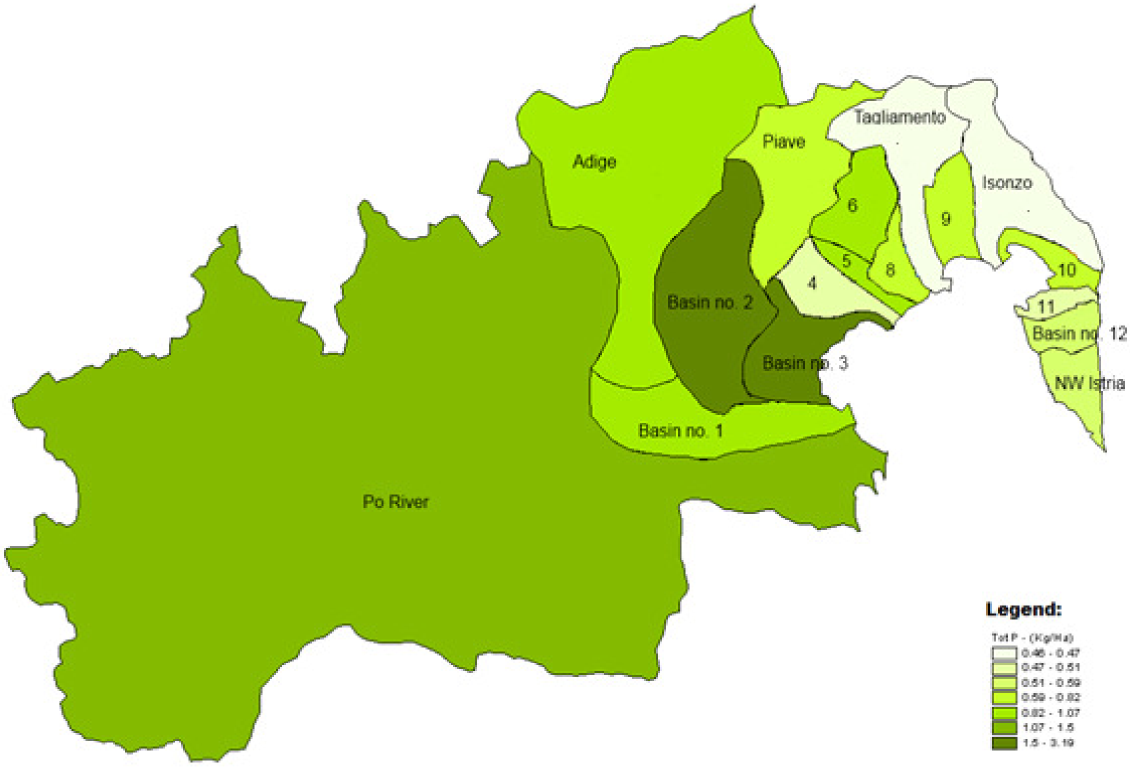

Figure 4.10 Specific loads of total phosphorus (P_{tot}) for each watershed area (average values for the whole period of modelled years)

4.3 Summary

Po River is the main contributor in nutrient loadings for NA (around 68 %). However, other contributing areas are not negligible and have to be considered as well. As mentioned in Section 4.2 agricultural areas and WWTP are major sources of nutrient loadings for NA. Proper management of these areas, such as use of fertilizers with lower share of nutrients and introducing suitable wastewater treatment, may reduce the nutrient loadings to NA. But, if we look at NA the latest study performed on long term data carried strong evidence that the still common perception of the NA as a very eutrophic basin is no longer appropriate, at least for its northern part and in recent years (Mozetič *et al.*, 2009). However, episodes of algal blooms, anoxia, and mucilage events were still noted in the last two decades (Degobbis *et al.*, 2000; Precali *et al.*, 2005), indicating that eutrophic episodes may still prevail for shorter time in a long run of relatively stable mesotrophic or even oligotrophic conditions. Taking all this into account it must be considered all this fact and decide what is the best to be done.

The model is also possible to use for controlling nutrient loads and proper management options in the NA watershed as will be presented in Chapter 6.

Chapter 5

Assessing the state of northern Adriatic

To understand the functioning of ecosystem it is of crucial importance to understand the ecosystem's main biogeochemical and hydrological characteristics and processes. State description (part **S** of Driving forces-Pressures-States-Impacts-Responses (DPSIR) framework) and data analysis of northern Adriatic (NA) ecosystem using machine learning (ML) tools is given in this chapter.

NA as it has been mentioned in Chapter 2 is the most productive region in the Mediterranean which results in many problems. Eutrophication is widely recognised as a major problem affecting Europe's seas, like here in the case of NA. Many attempts have been made to determine the inter-annual and seasonal variability of different environmental and biological parameters by using datasets of varying time spans, consistency of sampling, and spatial coverage. Most studies, however, were based on data referring to the specific site, such as the Gulf of Trieste (Cataletto *et al*., 1995, Mozetič *et al.*, 1998, Fonda Umani *et al.*, 2004, Kamburska and Fonda Umani, 2006, Solidoro *et al*., 2007, Conversi *et al.*, 2009), the coastal area in front of Lagoon of Venice (Bernardi Aubry *et al.*, 2004), profiles in the basin (Degobbis *et al.*, 2000; Tedesco et al., 2007), while only a very few studies attempted to address the regional basin scale (Zavatarelli *et al.*, 1998, Mozetič *et al.*, 2009, Solidoro *et al.*, 2009).

The latest study performed on long term data carried strong evidence that the still common perception of the NA as a very eutrophic basin is no longer appropriate, at least for its northern part and in recent years (Mozetič *et al.*, 2009). However, episodes of algal blooms, anoxia, and mucilage events were still noted in the last two decades (Degobbis *et al.*, 2000, Precali *et al.*, 2005), indicating that eutrophic episodes may still prevail for shorter time in a long run of relatively stable mesotrophic or even oligotrophic conditions. This indicates that not only nutrient enrichment (eutrophication) gives rise to excessive algal growth and mucilage formation, but also stressing conditions, e.g. change in nutrients ratio may be the cause (Degobbis *et al.*, 2000).

Mucilage events, as another significant problem, have been documented several times during the past two centuries in the NA, while their frequency has significantly increased since 1988 (Russo *et al.*, 2005). In fact, these events occurred in the past in intervals of approximately 10 to 50 years, but in the last eighteen years the phenomenon has recurred with a higher frequency with variable intensities and durations (1988, 1989, 1991, 1997, 2000 to 2004, 2007, 2010; Stachowitsch *et al.*, 1990; Degobbis *et al.*, 1995, 1999; Vollenweider *et al.*, 1995; Cozzi *et al.*, 2004; Precali *et al.*, 2005; DeLazzari *et al.*, 2008; CMR, Rovinj, unpub. data).

For the development of the mucilage phenomenon large variations of nutrient fluxes and their ratios, during freshets in the NA and water column stratification, rather than the absolute amount of the nutrient inputs are more important (Grilli *et al.,* 2005). But, mucilage formation can also form in areas not directly affected by freshwater inputs (Precali *et al.,* 2005). Orthosilicate (SiO_4) does not play a significant limiting role in phytoplankton growth so it is probably not essential for the development of the mucilage phenomenon (Degobbis *et al.*, 2005).

The changes in nutrient ratios in the surface layer of the NA, influenced by Po River discharges, coincided with an increased frequency of mucilage events (formation of macro aggregates up to several meters long in the upper water column and surface or subsurface organic layers; Stachowitsch *et al.,* 1990, Precali *et al.,* 2005).

There is a large consensus that the mucilage phenomenon is generated by synergic combinations of several factors (e.g. Degobbis *et al.*, 1999). One of the most significant is the change of TIN/PO_4 ratio which can increase the phytoplankton excretion of polysaccharide mucus, the matrix of the mucilaginous material (e.g. Myklestad, 1995; Grilli *et al.*, 2005). Phytoplankton productivity in the NA is most likely to be P-limited and confirmed with bioassay studies and analyses of dissolved inorganic nutrients (Pojed and Kveder, 1977; Degobbis and Gilmartin, 1990; Pečar *et.al.*, 2004). This limitation is confirmed by the high inorganic TIN/PO_4 ratios (>25) of riverine inflows, much larger than the Redfield ratio of 16, considered optimal for phytoplankton growth (Redfield *et al.*, 1963). Consequently, the particulate matter is P depleted with highly variable particulate-N/particulate-P ratios between 6 and 49 (Giani *et al.*, 2003, Ogrinc and Faganeli, 2006). Along with phosphorus, silica, or orthosilicate (SiO_4) may be a limiting nutrient and thus triggering stressing behaviour of the phytoplankton, the mucuous excretion, known as mucilage.

Deserti *et al.*, (2005) investigate the connections between the mucilage and the change of the climatic conditions. The descriptive analysis pointed out that the mucilage events can be grouped in three main clusters: (1) 1920 to 1930; (2) 1983 to 1991 and (3) 1997 to 2002.

Mucilage scavenging plankton and detrital particles, settling on the bottom, can determine hypoxic and/or anoxic conditions particularly in the bottom waters and at the sediment-mucilage interface. The suffocation of benthic and epibenthic (including nekton) organisms poses serious fishery and sanitary problems, with important socio-economical implications.

In this chapter will be shown the advantages of ML methods to build understandable and interpretable models of phytoplankton dynamics and description of appearance of mucilage events. Most commonly, data analyses were performed with only classical and just recently with advanced statistical approaches such as principal component analysis (PCA, Bernardi Aubry *et al.*, 2004; Tedesco *et al.*, 2007). Although these techniques provide very useful insights in the data, they are sometimes limited in terms of interpretability due to their black-box nature. On the other hand, a branch of ML methods and tools were proven to produce descriptive, e.g. transparent-box models, which generally allow much easier interpretation (Kompare, 1995, Kompare *et al.*, 2001; Atanasova *et al.*, 2008; Džeroski, 2009; Volf *et al.*, 2011).

Two regression-based ML methods were applied to extract knowledge from a long term data set (1972 to 2007) taken from six stations capturing the profile from the Po River delta (Italy) to Rovinj on the Istrian coast (Croatia). Specific objectives of this research are: (1) the automatical reconstruction of knowledge from the data about the phytoplankton dynamics in the NA that have been assembled over the past decades of research in the area, (2) confirmation or rejection of some known patterns about ecosystem behaviour, (3) presentation of the knowledge in a descriptive tree-like model, (4) relate mucilage events to nutrient ratios, (5) revel which environmental variables in best way indicate mucilage events and (7) construction of a short-term predictive model of phytoplankton concentration.

5.1 Modelling experiments

For the modelling experiments the entire span of the historic data were used. At each station the measured parameters for the top 10 m of the water column were averaged (more related to eutrophication; mucilage phenomenon breaks out primarily in the upper water column). Additionally, information about the temporal occurrence of the mucilage events ware obtained (CMR, Rovinj, unpub. data).

The experiments were designed to elaborate models, following the objectives presented in introduction part of this chapter, e.g.:

(1) Descriptive model for phytoplankton concentration using a wide span of historical data,
(2) TIN/PO_4 ratio model describing appearance of mucilage events, and
(3) Prediction model for the phytoplankton concentration 14 days in advance, given the present values of the measured parameters.

5.1.1 Description of the experiments

For the first experiment the ML algorithm M5P for regression trees integrated in the Weka modelling software was used. The total phytoplankton (Phyto) was set as a target (dependant) variable, whereas date (year, month), Po River flow, temperature, salinity, density, pH, NO_3, NO_2, NH_4, molar ratios TIN/PO_4 and TIN/SiO_4 (see Table 5.1) were independent variables (descriptors) from which phytoplankton is modelled. The above parameters were mainly used because they best represent the parts of the ecosystem on top of which the target variable relays. Total phytoplankton counts were preferred as the target variable instead of chlorophyll *a* or oxygen super saturation, as parameters of phytoplankton biomass and photosynthetic activity, respectively. The determinations of these parameters are easier, but phytoplankton counts represent a more direct quantitative measure of biomass changes. Po River flow rates were used as a rough measure of the eutrophication pressure to the investigated ecosystem combined with nutrient concentrations in the sea as a measure of the eutrophication degree.

For the second experiment the machine learning algorithm M5P for regression trees integrated in the Weka modelling software was also used. The experiment was designed to first elaborate a model for TIN/PO_4 ratio. The ratio TIN/PO_4 was set as dependant variable, while Po River flow rate, year, month, sea water temperature, salinity, pH and density were given as independent variables (Table 5.2).

The third experiment is aimed to build up a model for prediction of phytoplankton concentration for 14 days in advance. To obtain more reliable and accurate model the data set was pre-processed by performing cubic spline interpolation between the measured data points and sampling these splines at daily frequency. In this way, by transforming the measurements with monthly frequency to daily data, a larger data set was constructed which resulted in more accurate models. For this task the software Cubist was used, which is an algorithm for building rule-based predictive models. In this case the descriptors or independent variables are month, temperature, flow, salinity, density, pH, NO_3, NO_2, NH_4, molar ratios TIN/PO_4 and TIN/SiO_4 and total phytoplankton (see Table 5.1) measured at time *t*, whereas the target or the dependant variable is interpolated value for the phytoplankton concentration at time *t+14 days*.

Table 5.1 Measured data used for modelling

Parameter	Description	Unit	Descriptive and predictive phytoplankton models	Mucilage events (TIN/PO_4 and TIN/SiO_4) model
Month	Month of sampling		●	●
Year	Year of sampling		●	●
Q_{Po}	Po River flow	m^3/s	●	●
Temp	Temperature	°C	●	●
SAL	Salinity		●	●
Dene	Density excess	kg/m^3	●	●
pH	pH		●	●
NO_3	Moles of Nitrate as N	μmol/l	●	
NO_2	Moles of Nitrite as N	μmol/l	●	
NH_4	Moles of Ammonium as N	μmol/l	●	
TIN/PO_4	$\frac{\text{Total Inorganic Nitrogen}}{\text{Orthophosphate as P}}$	$\frac{mol}{mol}$	●	●
TIN/SiO_4	$\frac{\text{Total Inorganic Nitrogen}}{\text{Orthosilicates as Si}}$	$\frac{mol}{mol}$	●	●
Phyto	Total phytoplankton	l^{-1}	●	●
Phyto_pred	Total phytoplankton moved for 14 days	l^{-1}	●	

5.1.2 Selection of training and testing data sets

The aim of models is to be applicable and valid to the entire observed area, meaning that they should perform as accurately as possible on the data sets from all stations. To achieve this, two most commonly used procedures of building and testing models were applied: the entire data set (all stations) is taken for training while validating with cross-validation and one portion of the data set is selected for training and the other for testing. The reason for the second validation procedure is the draw-back of the cross validation procedure for our case, i.e. some of the stations may be strongly affected by some variables that were not included in the data set. Such stations (or data subsets) may present a big noise in the entire data set and prevent building of the acceptable model. To select the best model, each one of them was simulated on the data from the rest of the data (stations) and the error between the simulated and measured data was used.

The model performing most accurately according to the validation method and having good descriptive power was selected as a representative phytoplankton model and TIN/PO_4 model for the observed area.

5.2 Results

5.2.1 Descriptive model for phytoplankton (1st model)

From the historical data from six stations (Table 5.1) and using the model selection criteria, a descriptive phytoplankton model was constructed (Figure 5.1). For some stations not acceptable results by training a model on the entire data set and validating by cross-validation were obtained, and it is supposed that data on this stations are noisy (e.g. station SJ108 is more influenced by the Po River inflow than other stations). To avoid this, second procedure for validating was applied, e.g., training on each station's data and validating on others. According to the selection criteria the best model was induced from the data measured on the station SJ107. The model was validated by simulating it on testing data sets, e.g., the data from the remaining stations. The accuracy of the model is given by the correlation coefficient (R) between the modelled and the measured values of the phytoplankton concentration. The correlation coefficient for selected model (Figure 5.1) on the training data set is 0.60, while the correlation coefficients between the phytoplankton values simulated with the model and the values in the testing data sets, e.g., measured at the stations RV001, SJ101, SJ103, SJ105, and SJ108 are, 0.51, 0.36, 0.44, 0.45 and 0.26 respectively.

The correlation coefficients are relatively low. Still, knowing the complexity of the tackled domain and the lack of enough data of appropriate accuracy and frequency, the obtained results are more than satisfactory. Further, comparing the models among themselves, another valuable piece of knowledge about the functioning of the ecosystem in NA was obtained. For example, it is indicated very low correlation coefficient between the simulated and the measured values on the station SJ108. Similarly, low accuracies were observed for all models induced from the data at the stations RV001, SJ107, SJ105, SJ103, and SJ101, respectively, when tested on the data from SJ108 (accuracies not presented here). This indicates that the ecosystem behaviour at this station differs (e.g. is more influenced by the Po River, as it is closer to the river inflow) when compared with the rest of the investigated profile and cannot be modelled sufficiently well with the available data set. Thus, the generality of the model cannot be extended to this part of the area.

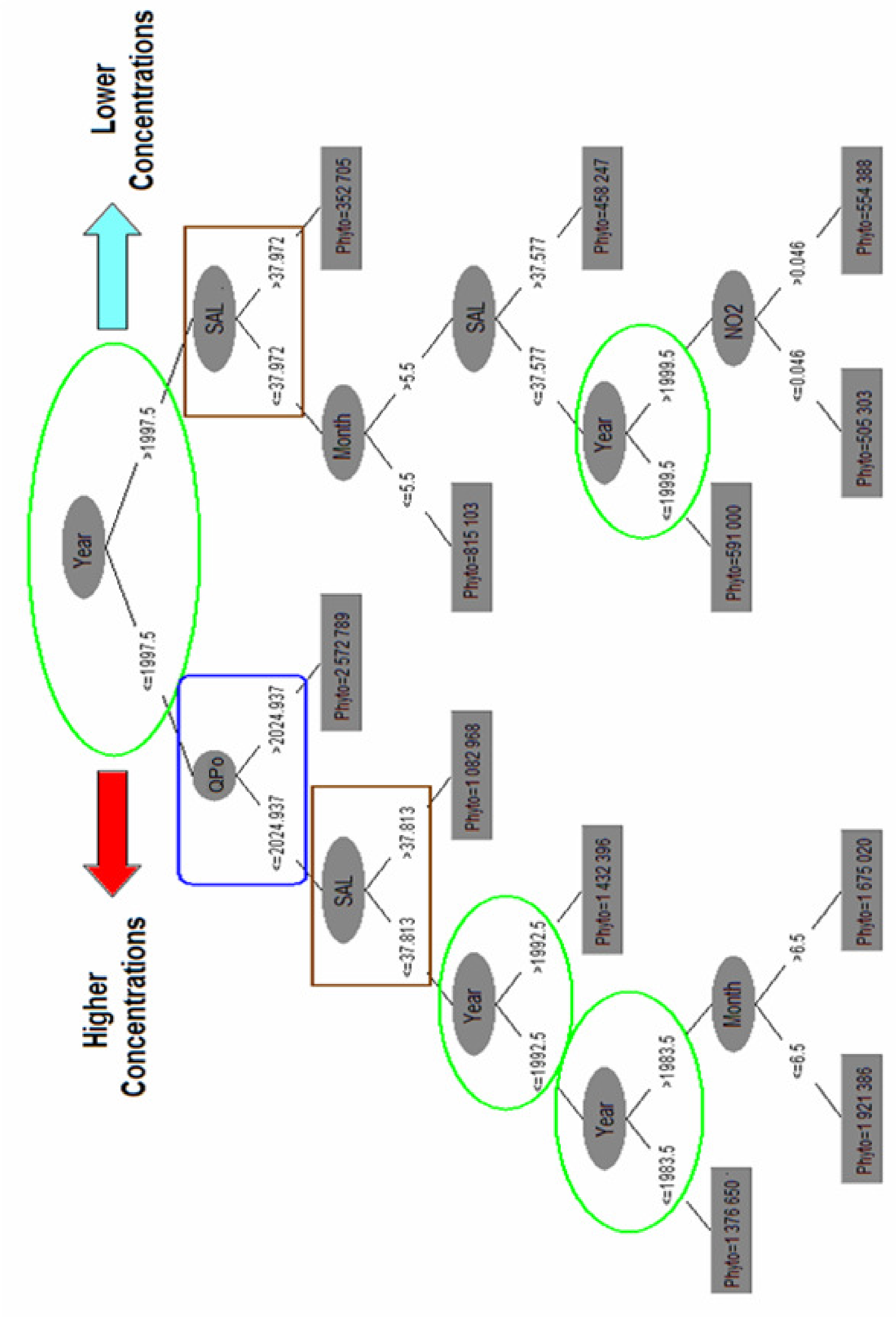

Figure 5.1 The model for dynamics of phytoplankton concentration (1st model) for station SJ107 (units for the threshold values for the parameters used are reported in Table 5.1)

The goal of this (1st) model (Figure 5.1) is to explain how the phytoplankton concentration has been changing in the NA and to identify the most influential factors of this dynamics. The model is constructed from five variables (descriptors) from which the phytoplankton concentration in the given observed period of time (from 1972 to 2007) can be determined. These variables are year, month, salinity, Po River flow and, surprisingly, NO_2, which, at the first glance, would not be taken as very significant for phytoplankton growth. Instead, the ML tool discovered this parameter as more informative than the others which were then (automatically) omitted during the model construction procedure. Similarly, the molar ratio TIN/PO_4, which is typically very informative variable for phytoplankton concentration, does not appear in the model. Namely, the ratio in this case is generally higher than 16 which indicate that the limiting nutrient for phytoplankton growth is always phosphorus (Redfield, 1934), e.g. no additional information can be obtained from the ratio values.

Evidently, this model confirms some of the conclusions from previous research in the phytoplankton dynamics in the NA and gives an easy to read structured knowledge representation. As indicated previously, the salinity and Po River flow appear to be the most important indicators of trophic changes in the observed ecosystem.

5.2.2 TIN/PO_4 model describing appearance of mucilage events (2nd model)

The goal of this (2nd) model is to give an insight in how TIN/PO_4 ratio is changing in the NA marine ecosystem, and what are the most influential factors for this change, as this ratio is recognized as one of the more important and necessary factors if not a trigger for the mucilage production (Herndl, 1992; Degobbis *et al.*, 1999).

From the historical data from six stations (Table 5.1, Figure 3.5) a TIN/PO_4 ratio model was constructed for whole NA area (Figure 5.2). Parts A, B and C of model tree on Figure 5.2 are shown as sub-trees on Figures 5.3, 5.4 and 5.5. The accuracy of the model is given by the correlation coefficient (R) between the modelled and measured values of the TIN/PO_4 ratio. The correlation coefficient for selected model (Figure 5.2) using cross-validation method is 0.55.

The regression tree model in Figure 5.2 shows the different average values of the TIN/PO_4 ratio in given time periods and under different conditions. The model is read like IF-THEN rules, starting from the top node, e.g. *IF (Year ≤ 1999 and Temp > 19.6 and* $Q_{Po} \leq 1015$ *and Year > 1995) THEN (*$TIN/PO_4 = 18.2$*)*. Analysing the tree two characteristic periods for the value of the TIN/PO_4 ratio can be distinguished, where the environmental factors differently influenced the value of the ratio, e.g. with different threshold values. These periods are before and after year 1999 (see the top node of the tree in Figure 5.2), e.g., periods from 1972 to 1999 and from 2000 to 2007 (see Figures 5.3, 5.4 and 5.5).

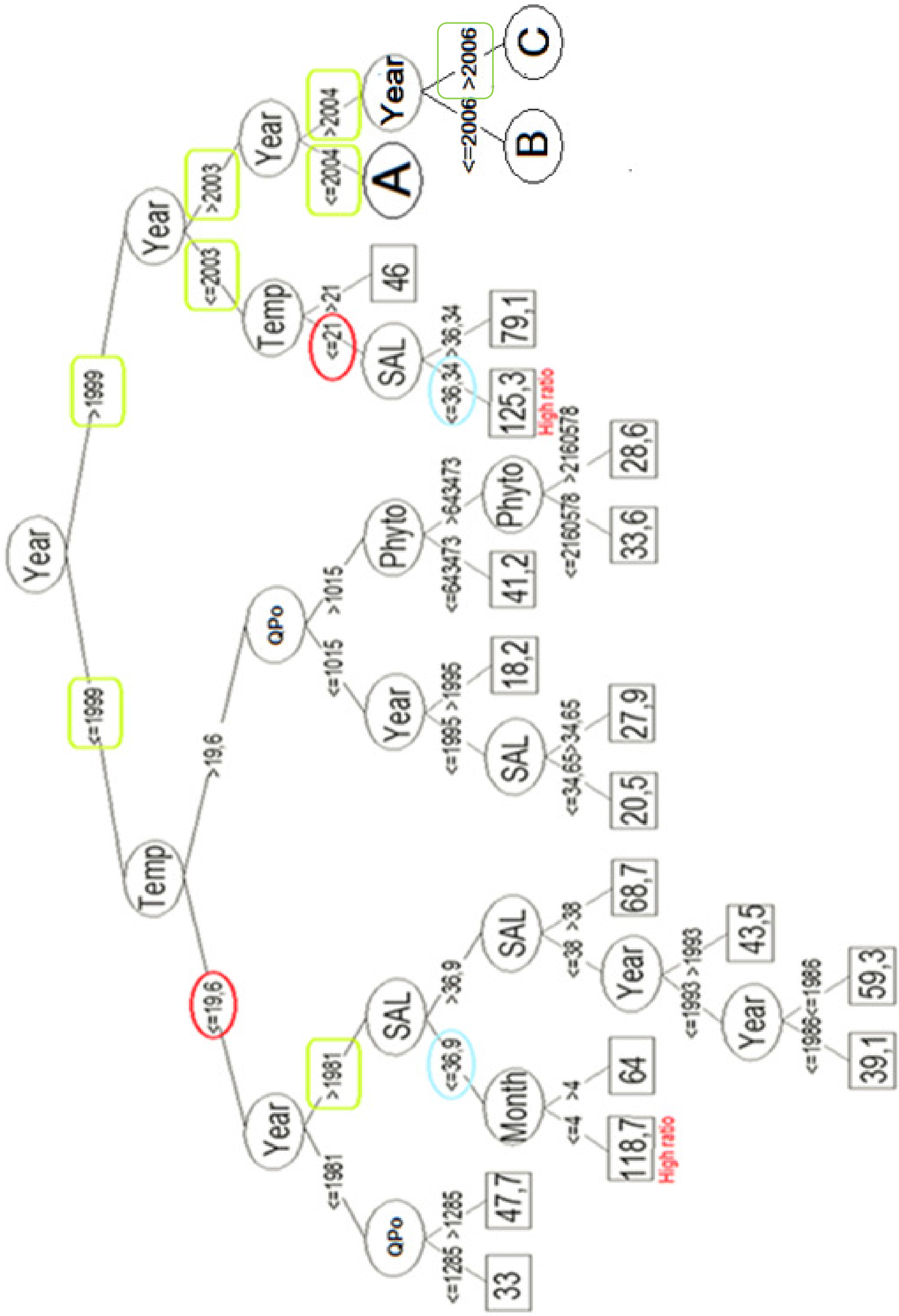

Figure 5.2 The model for TIN/PO_4 ratio (2nd model). High nutrient ratios coincide with observed mucilage events (units for the threshold values for the parameters used are reported in Table 5.1)

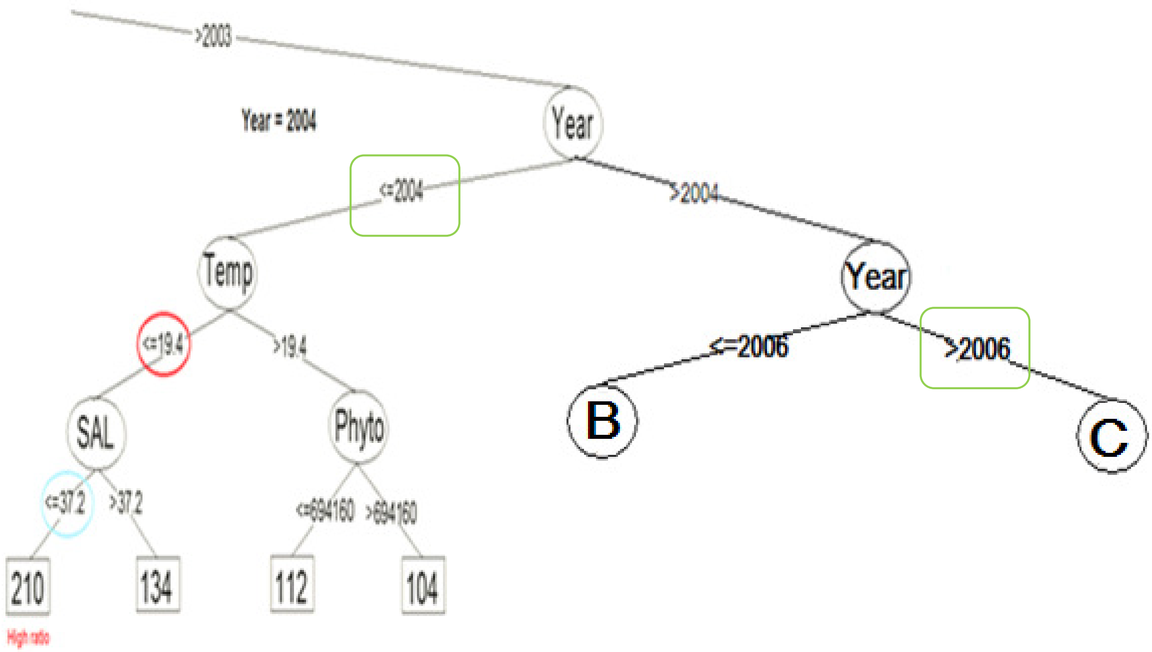

Figure 5.3 Part of sub-tree (part A) for the model tree presented in Figure 5.2 for year 2004

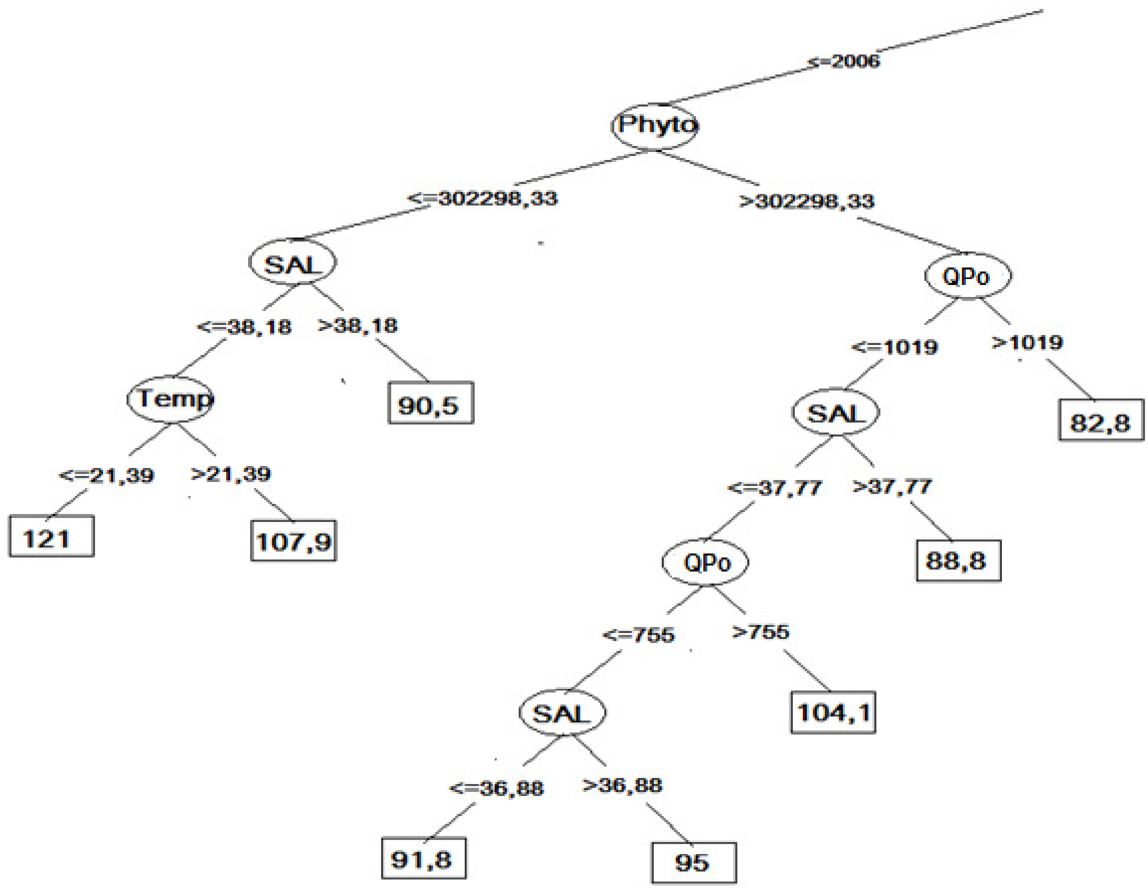

Figure 5.4 Part of sub-tree (part B) for sub-tree presented in Figure 5.3 for years 2005 and 2006

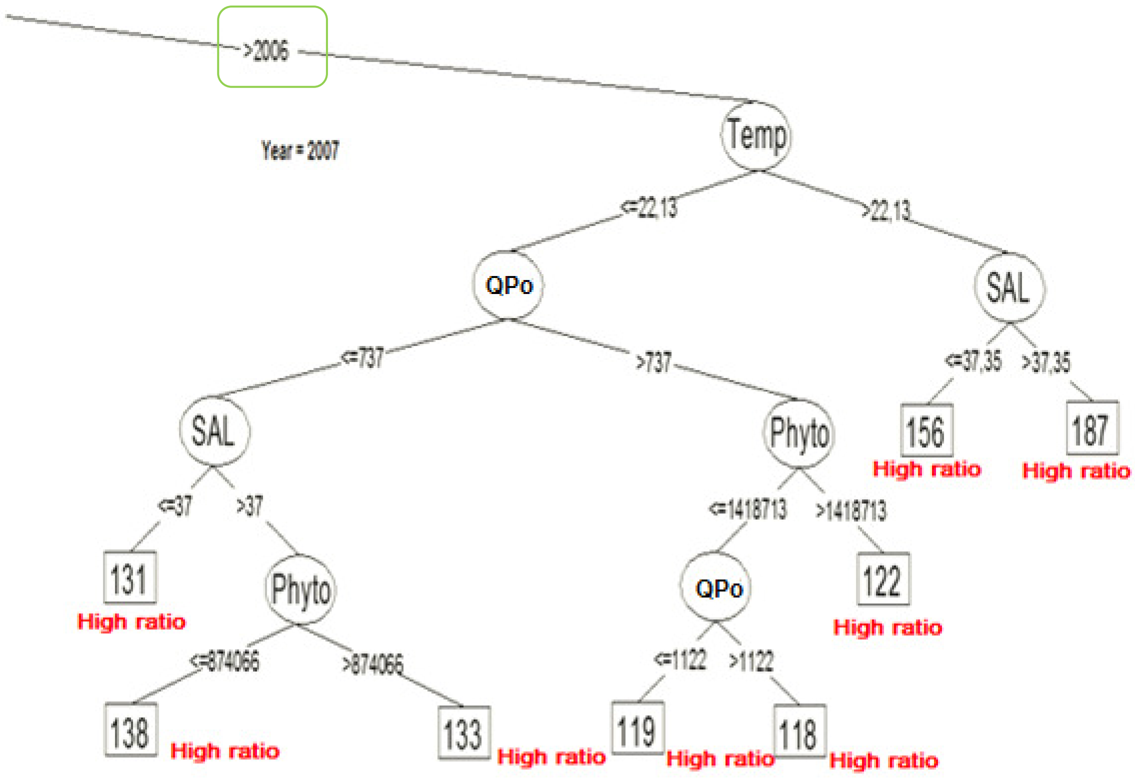

Figure 5.5 Part of sub-tree (part C) for sub-tree presented in Figure 5.4 for year 2007

In the first period the water temperature (Temp) has an important influence on the ratio with a threshold value of 19.6 ^{o}C. If the temperature was higher than the average value of the TIN/PO_4 ratio was relatively low, e.g. between 20.5 and 33.6. Otherwise, at lower temperatures we observe different influences on the ratio values at two main sub-periods, e.g.:

The first period is before 1981 (1972 to 1981), when only the Po River flow (Q_{Po}) influenced the value of the ratio. But, according to the model the average value of the ratio was always relatively low (33 and 47.7). No mucilage events were noted during this period. The rules that confirm this are read as follows:

1. *IF (Year $\leq$ 1999 and Temp $\leq$ 19.6 and Year $\leq$ 1981 and $Q_{Po} \leq 1285$) THEN (TIN/PO_4 = 33),* and
2. *IF (Year $\leq$ 1999 and Temp $\leq$ 19.6 and YEAR $\leq$ 1981 and $Q_{Po} > 1285$) THEN ($TIN/PO_4 = 47.7$)*

The second characteristic sub-period is between 1982 and 1999, when salinity (SAL) and month of the year (Month) influenced the average value of the ratio, with threshold values of 36.9 and 4 respectively. At lower salinity value (SAL $\leq$ *36.9)* and months before April (Month $\leq$ 4) is observed higher average TIN/PO_4 ratio value of 118.7, otherwise for months after April the average value of the ratio is 64. Also, at higher salinity values (SAL > 36.9) the average value of the ratio is between 39.1 and 68.7, much lower than the ratio of 118.7. The rules that confirm this are evident from the tree model on Figure 5.2.

Interestingly, in this period mucilage events appeared during the second sub-period between 1982 and 1999 (events in 1988, 1989, 1991 and 1997), coinciding with highest average values of the TIN/PO_4 ratio (118.7). These values were triggered by temperatures lower than 19.6 ^oC, salinity lower than 36.9 and months of year before April.

In the second characteristic period, between 2000 and 2007, three main characteristic sub-periods can be distinguished. In the first, between 2000 and 2003 (see Figure 5.2) the model reveals high average TIN/PO_4 ratio value of 125.3 at temperature lower than 21 ^oC and salinity lower than 36.34. Mucilage events appeared frequently in this period, i.e. in 2000, 2001, 2002 and in 2003. In the second sub-period for year 2004 when mucilage event is observed (Figure 5.3, Year ≤ 2004) for high ratio value (TIN/PO_4 = 210) are responsible like in other periods when mucilage was observed temperature and salinity values lower than 19.4 ^oC and 37.2, respectively. Other values of ratio for this sub-period are between 104 and 134. In the third sub period, between 2005 and 2007 (Figures 5.4 and 5.5) the model also reveals high average TIN/PO_4 ratio values (range between 82.8-121) influenced by Phytoplankton (Figure 5.4). In 2005 and 2006 (Figure 5.4) there were no mucilage events and when compared to 2007 when mucilage was observed ratio values are not so high. Probably these high values are a result of a decrease of Po River inflow, in the last 5 years a clear signal indicating a reduction of the available orthophosphate in the ecosystem was identified with an accumulation of inorganic nitrogen (Mozetič *et al.*, 2009). Year 2007 (see Figure 5.5) is characteristic because mucilage events were observed in Spring-Summer period and in Autumn-Winter period (CMR, Rovinj, unpub. data) so the ratio values in whole period are high (TIN/PO_4 = 118-187).

Similarly to the first characteristic period (1972 to 1999), in the second period, the mucilage events coincide with the high TIN/PO_4 ratio values. However, compared to the first period in the second one, for his first sub-period (2000 to 2003) are observed different threshold values of the temperature and salinity, e.g. 21 ^oC and 36.34, respectively, while for second sub-period (year 2004) threshold values of the temperature and salinity are similar to the first period between 1982 and 1999 (19.6 and 19.4 ^oC, 36.9 and 37.2, respectively). In 2007 mucilage events were observed in Spring-Summer period and in Autumn-Winter period so the ratio values in whole period are high (TIN/PO_4 = 118-187).

The second part of the modelling experiment was designed to test possible correlation between the TIN/SiO_4 ratio and the mucilage events. As already presumed the obtained models in this case have very low correlation coefficients, thus they confirm the hypothesis that the orthosilicate (SiO_4) in general does not trigger the mucilage events (Degobbis *et al.*, 2005; results of second experiment are not presented).

5.2.3 Predictive model for phytoplankton concentration (3rd model)

The model for predicting the phytoplankton concentration for 14 days in advance was built with the Cubist algorithm for rules induction, using the data from 1990 to 2007. To select the best model the model selection procedure described in Section 5.2 was used. As indicated in the previous section the ecosystem behaviour at the point SJ108 differs from the rest of the area, e.g. the available data only allow a generic model for the area from RV001 to SJ101. The most accurate predictive phytoplankton model for this area was induced from the joint data sets from the stations RV001 to SJ101 and validated by cross-validation. The model achieves high accuracy when tested with 10-fold cross validation, with correlation coefficient 0.91. Expectedly, the accuracy of this model is much higher compared to the accuracy of the descriptive model (1st model). This is due to: (1) the different learning algorithm, i.e. in this case the algorithm uses linear equations to predict the dependant variable while a simple regression tree (1st model) puts a single value in the terminal nodes of the tree, and (2) the data set for learning is much bigger as interpolated values were used.

The model is composed of ten rules, each of which is related to a linear equation predicting the phytoplankton concentration for 14 days in advance (Table 5.2). To predict this concentration the following data are needed at present: Po River flow, month, temperature, salinity, density, pH, NO_3, NO_2, NH_4, molar ratios TIN/PO_4 and TIN/SiO_4 and phytoplankton concentration (Table 5.1). The rule selection depends on the values of the variables in the rule. When a rule is selected, a corresponding equation is applied to calculate the phytoplankton concentration 14 days in advance.

Table 5.2 Predictive model for phytoplankton concentration, rule based model (3rd model, units of the parameters used are reported in Table 5.1)

Rule no.	**Rules:**	**Equations:**
Rule 1	If Phyto $\leq$ 804 620	Phyto_pred = 732 300 + 0.917 Phyto – 233 140 Dene – 58 205 Temp + 179 186 SAL – 7 782 Month
Rule 2	If Temp > 9.65 Temp $\leq$ 20.31 Phyto > 804 620.5 Phyto $\leq$ 2 807 349	Phyto_pred = 5.24147e+006 + 0.931 Phyto – 84 220 Temp – 57 364 Dene – 199 934 NO_2 – 20 263 NO_3 – 7 368 TIN/SiO_4 + 29 Flow - 368 417 pH + 88 810 NH_4 + 19 504 SAL
Rule 3	If Temp > 20.31 Phyto > 804 620.5	Phyto_pred = -2.76291e+006 + 0.716 Phyto + 359 934 Dene + 103 143 Temp + 135 511 Month – 256 581 SAL – 1 125 TIN/PO_4 + 7 567 TIN/SiO_4
Rule 4	If Temp $\leq$ 9.65 Phyto > 804 620 Phyto $\leq$ 2 807 349	Phyto_pred = -6.68528e+006 + 0.87 Phyto – 390 486 Dene + 386 746 SAL – 80 868 Temp + 218 220 NH_4 - 852 TIN/PO_4 + 21 132 NO_3 – 12 263 Month + 559 168 pH
Rule 5	If TIN/PO_4 $\leq$ 62.35 Phyto > 2 807 349	Phyto_pred = 1.16131e+007 + 0.664 Phyto – 141 666 Temp – 6 625 TIN/PO_4 – 147 108 NO_3 – 97 863 Dene+ 406 748 NO_2 + 17 884 TIN/SiO_4 – 660 243 pH + 87 424 NH_4
Rule 6	If Month > 4 TIN/PO_4 > 62.35 Phyto > 2 807 349 Phyto $\leq$ 1.13e+007	Phyto_pred = 7.14367e+007 – 1.18988e+006 Temp – 3.28797e+006 Dene + 2.00528e+006 SAL + 0.831 Phyto – 4.86272e+006 NO_2 – 278 807 Month + 2.30991e+006 NH_4 – 258 224 NO_3 – 55 637 TIN/SiO_4 – 4.16418e+006 pH + 2 917 TIN/PO_4 - 57 Q_{Po}
Rule 7	If Temp $\leq$ 12.28 NH_4 $\leq$ 0.31 TIN/PO_4 > 62.35 Phyto > 2 807 349	Phyto_pred = 3.97828e+007 + 8.67119e+006 Dene + 1.92162e+006 Temp – 5.92294e+006 SAL – 8.71454e+006 NH_4 + 797 894 NO_3 + 1 566 Flow + 0.8 Phyto – 1.05622e+007 pH – 10 620 TIN/PO_4 + 2.05592e+006 NO_2
Rule 8	If Temp $\leq$ 12.28 NH_4 > 0.31 TIN/PO_4 > 62.35 Phyto > 2 807 349	Phyto_pred = 1.26724e+008 - 3.25732e+007 Dene – 7.30295e+006 Temp + 2.64069e+007 SAL – 7.87193e+006 NO_2 + 0.78 Phyto + 610 748 NO_3 – 1.16366e+007 pH – 2.89067e+006 NH_4
Rule 9	If Month <= 4 Temp > 12.28 TIN/PO_4 > 62.35 Phyto > 2 807 349	Phyto_pred = 1.3374e+008 + 4.30887e+006 Dene – 4.10462e+006 SAL + 1.756 Phyto – 815 924 NO_3 – 4.89689e+006 NO_2 – 1.19144e+007 pH – 4 543 Temp
Rule 10	If Month > 4 TIN/PO_4 > 62.35 Phyto > 1.13e+007	Phyto_pred = -4.05262e+008 – 4.50557e+006 Temp + 19 140 Q_{Po} + 5.66287e+007 pH – 1.05564e+007 NH_4 + 36 997 TIN/SiO_4

The performance of the predictive model (e.g. 3rd model) is presented in Figures 5.6 to 5.11. Each figure represents the modelled vs. the measured values of the phytoplankton concentration for the given station. The Figures indicate high accuracy of the model compared to the measured data, with also good prediction of the peak values.

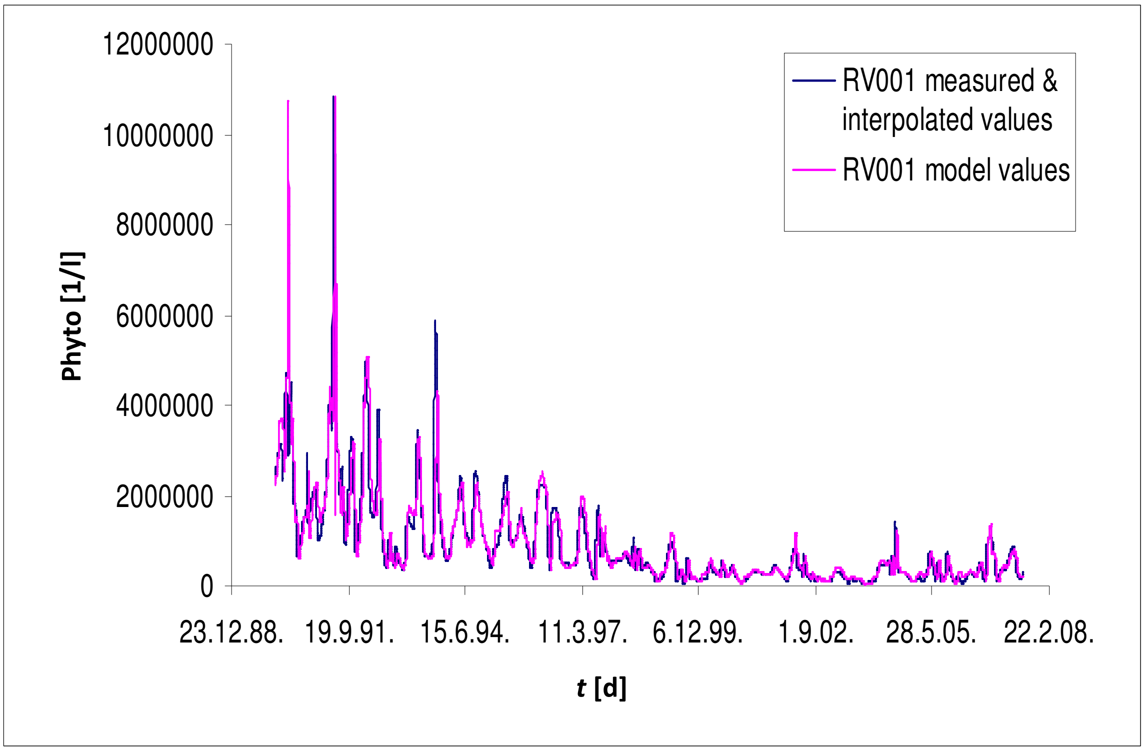

Figure 5.6 Results of (3rd) model for station RV001 vs. measured & interpolated values (R= 0.88)

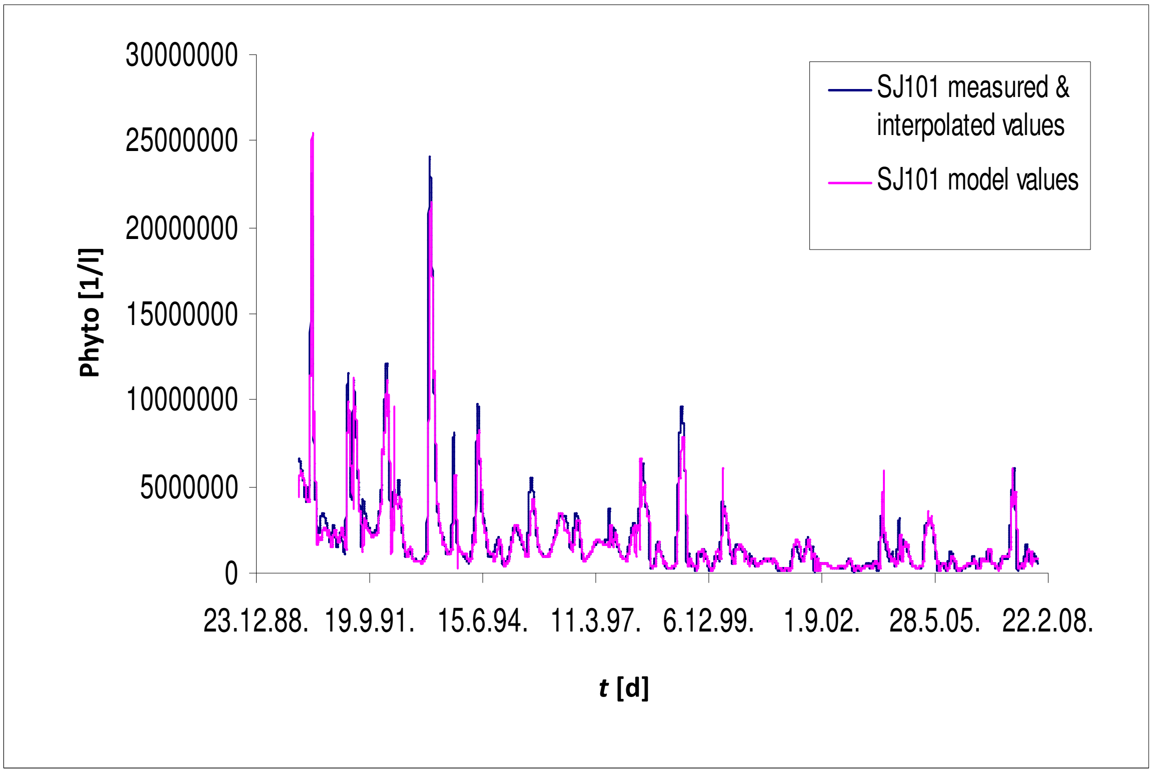

Figure 5.7 Results of (3rd) model for station SJ101 vs. measured & interpolated values (R= 0.95)

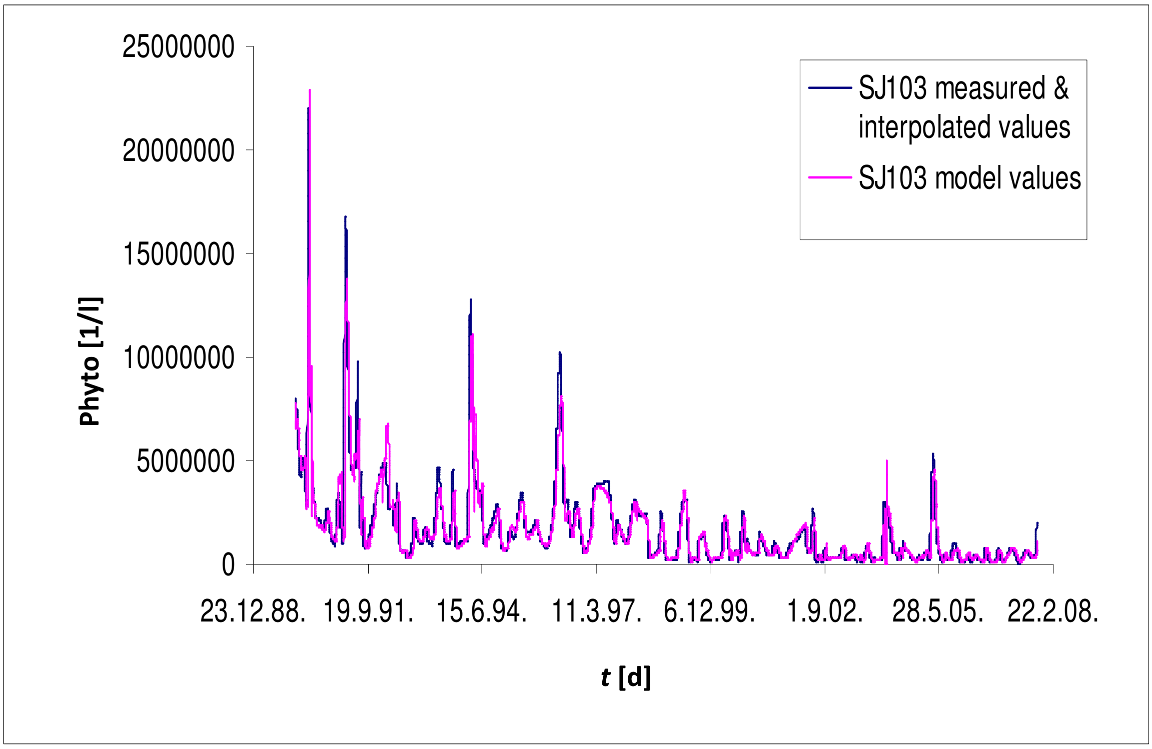

Figure 5.8 Results of (3rd) model for station SJ103 vs. measured & interpolated values (R= 0.91)

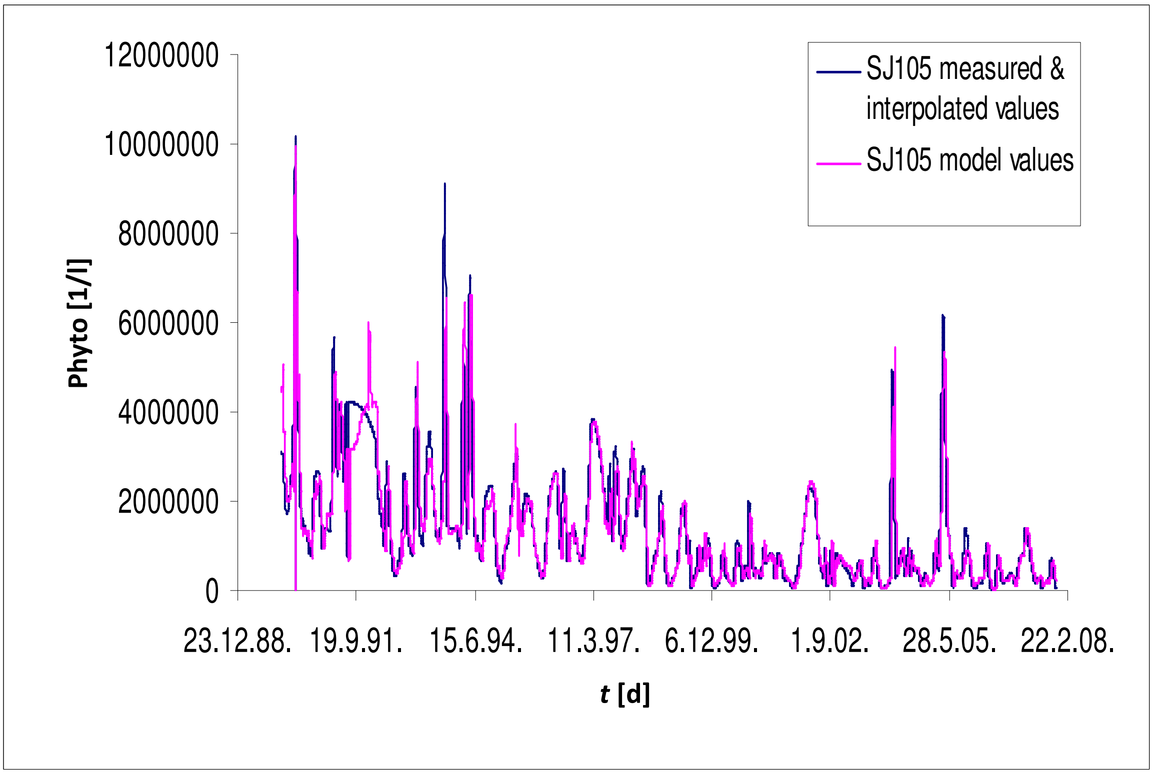

Figure 5.9 Results of (3rd) model for station SJ105 vs. measured & interpolated values (R= 0.87)

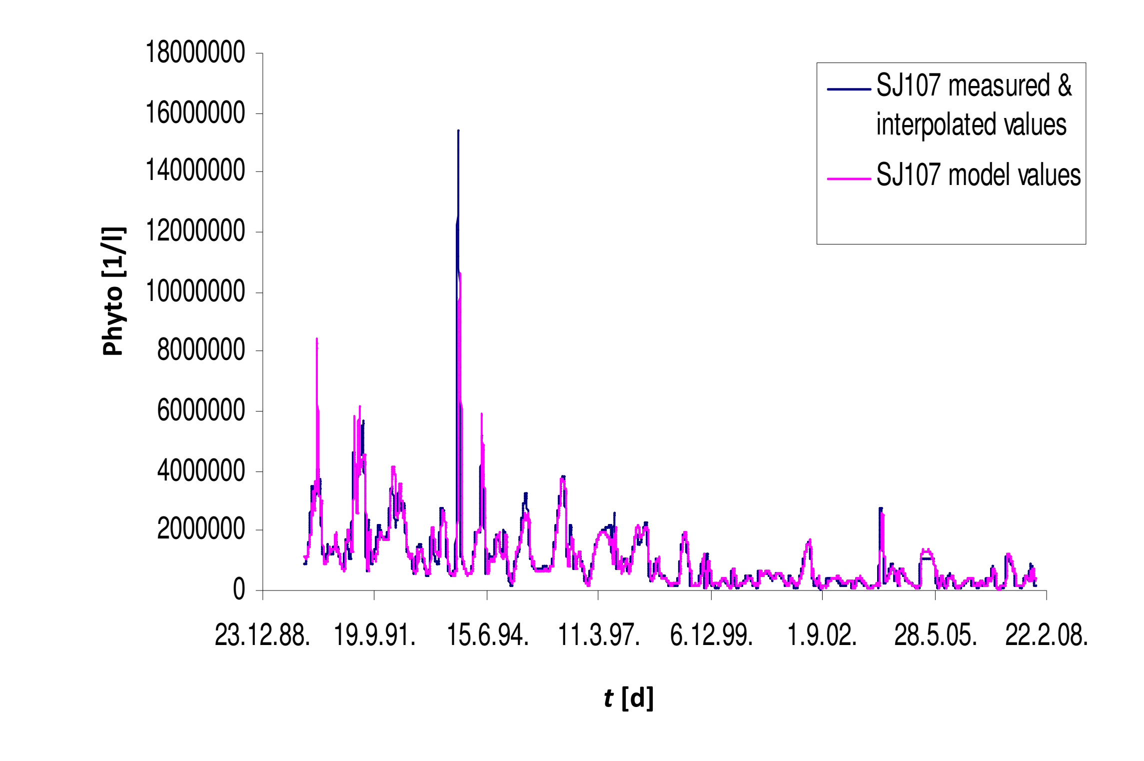

Figure 5.10 Results of (3rd) model for station SJ107 vs. measured & interpolated values (R= 0.88)

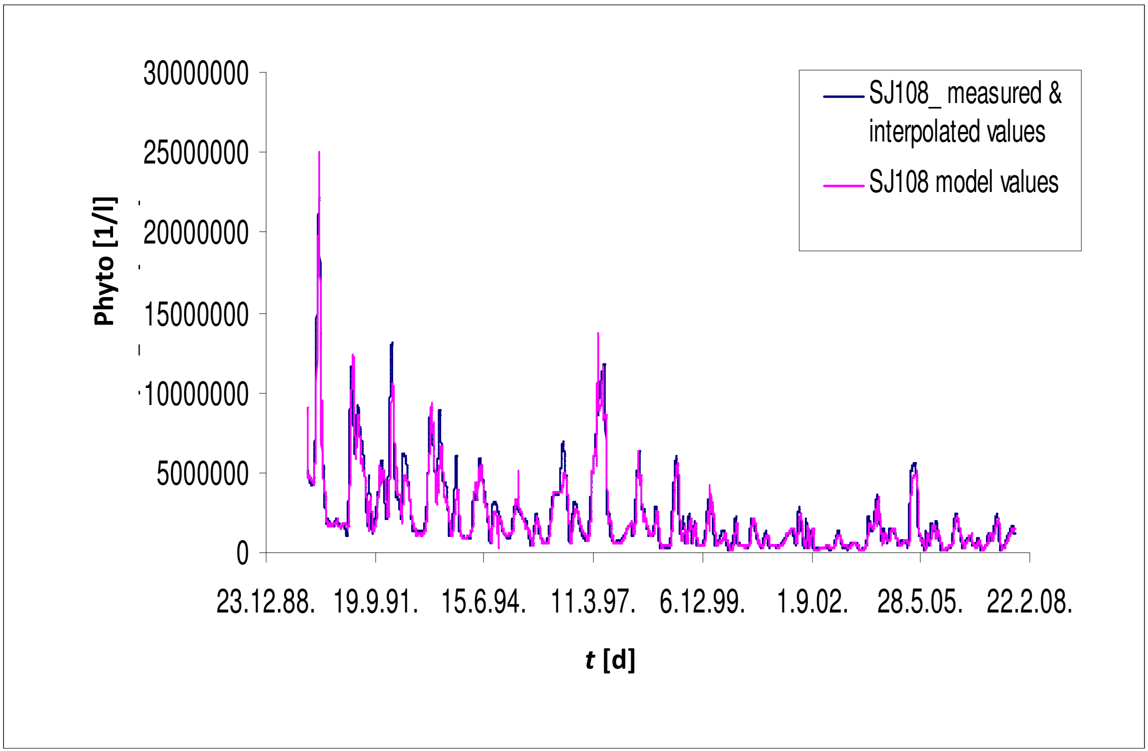

Figure 5.11 Results of (3rd) model for station SJ108 vs. measured & interpolated values (R= 0.72)

For better data visualization Ocean Data View (ODV) software suite was used (http://odv.awi.de/en/home). The phytoplankton data measured on the 15.May.1997 (Figure 5.12) were compared with the values for the same day simulated by the model (Figure 5.13). The errors (in brackets) between the measured and the predicted values (Figures 5.14 and 5.15) are as follows:

- station RV001: - 295 737 l^{-1}, (44 %),
- station SJ101: - 9 556 l^{-1}, (0.5 %),
- station SJ103: 231 392 l^{-1}, (6 %),
- station SJ105: - 185 900 l^{-1}, (6 %),
- station SJ107: 109 414 l^{-1}, (5 %),
- station SJ108: - 4 083 101 l^{-1}, (43 %).

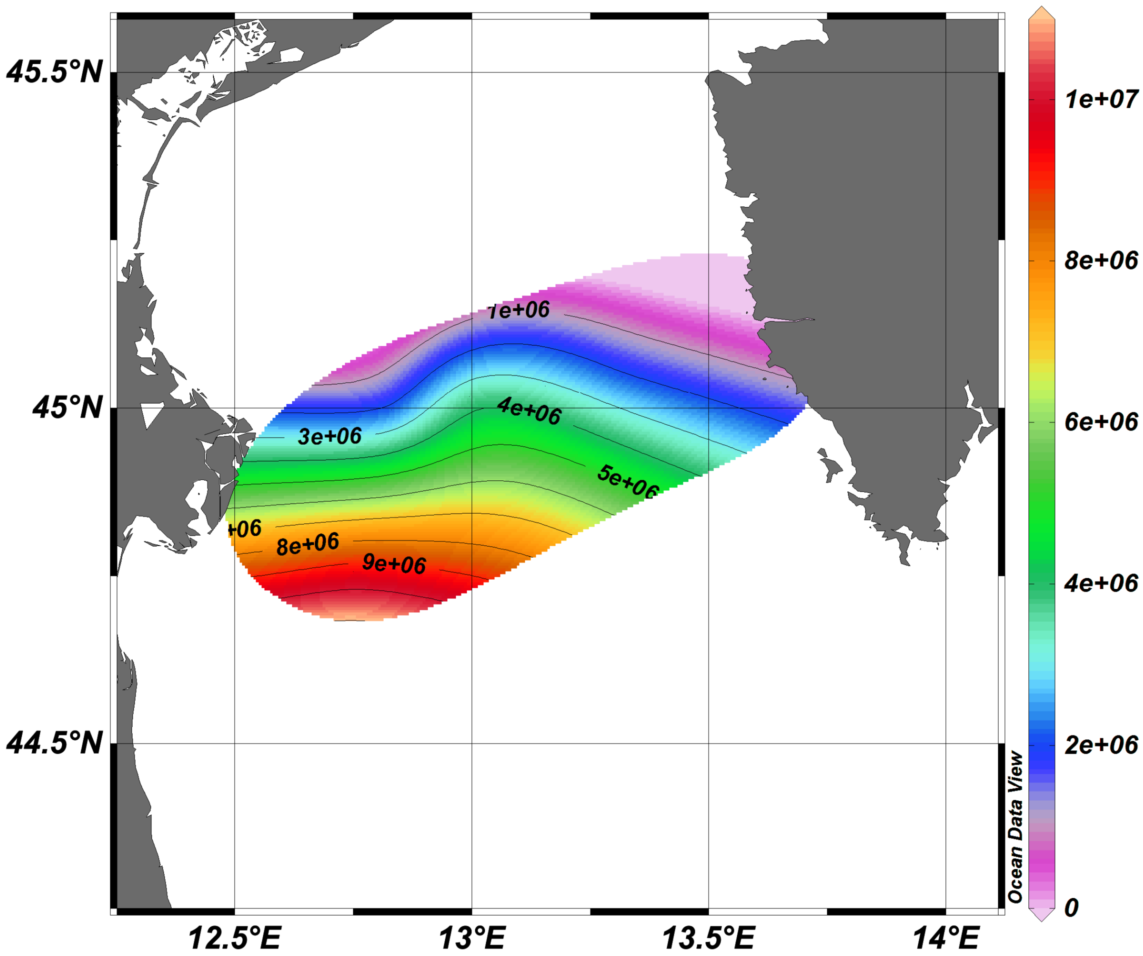

Figure 5.12 Measured (interpolated) values of phytoplankton concentration in cells per l on 15.5.1997.

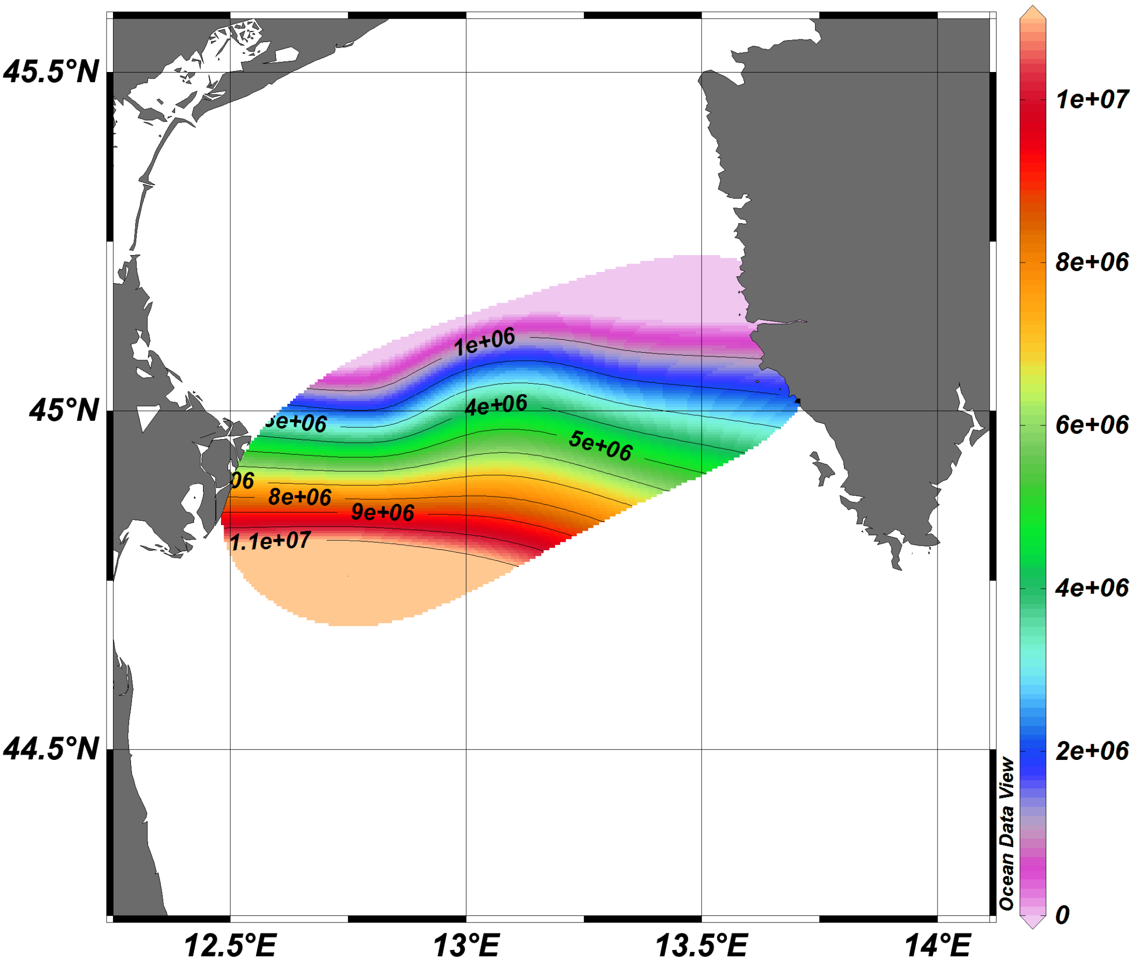

Figure 5.13 Predicted (3^{rd} model) values of phytoplankton concentration in cells per l on 15.5.1997. (calculated on data from 14 days before)

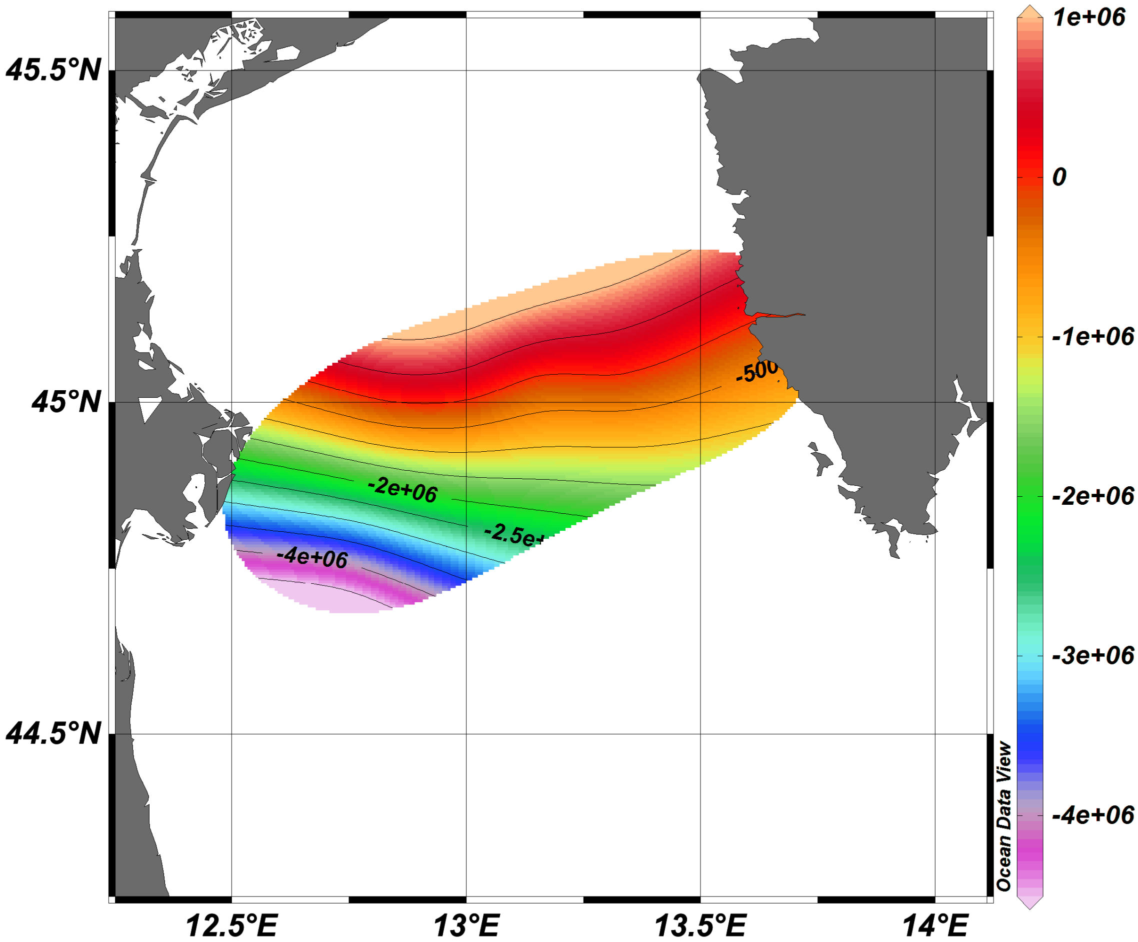

Figure 5.14 Differences between measured (interpolated) and predicted (3rd model) values of phytoplankton concentrations in cells per l on 15.5.1997.

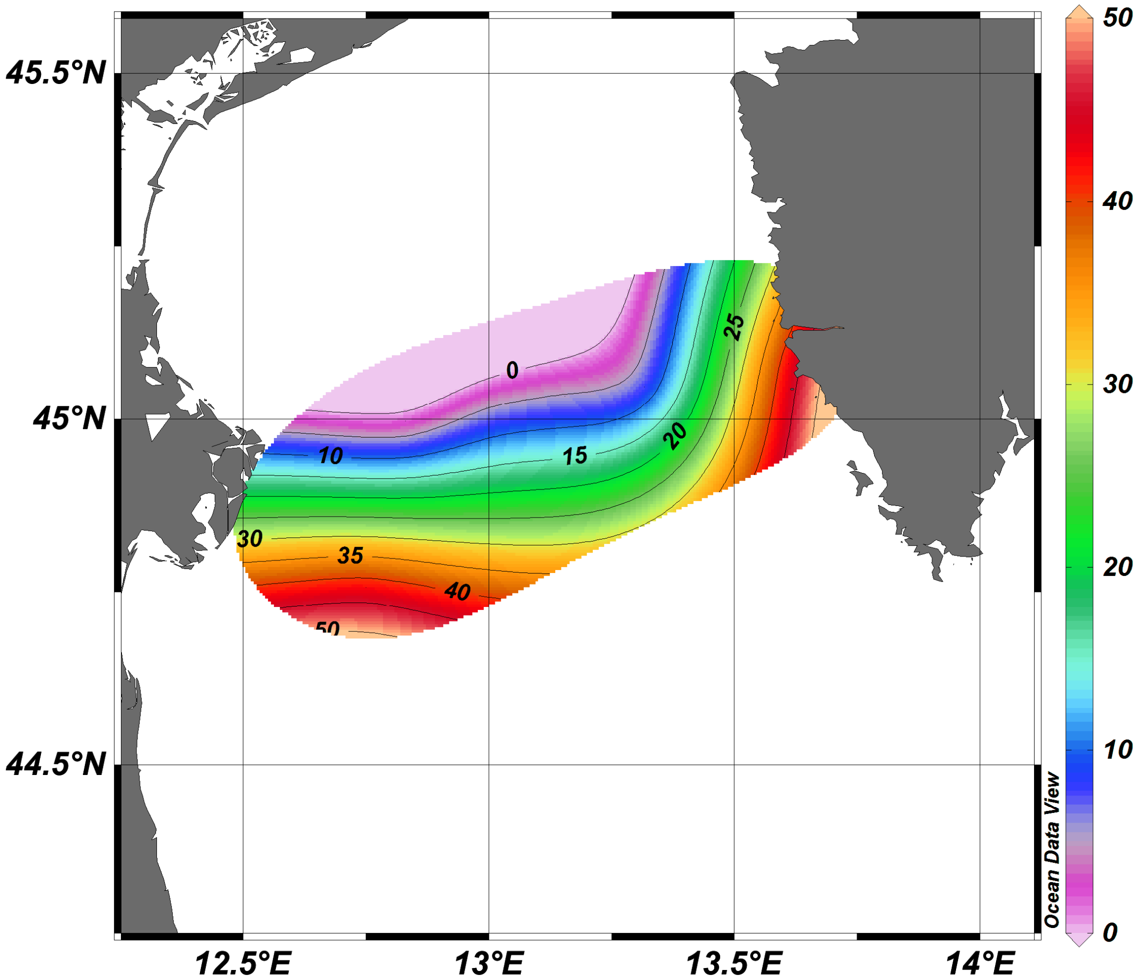

Figure 5.15 Differences between measured (interpolated) and predicted (3rd model) values of phytoplankton concentrations on 15.5.1997. in percent (%).

As expected, the high error occurs on station SJ108 (43 %) due to the reasons mentioned in the subsection 5.2.1. Nevertheless, the highest error is observed for the station RV001 (44 %), which can be related to more notable oligotrophic conditions in this part and, thus, very small phytoplankton concentrations. Oligotrophic conditions are due to the inflow of more oligotrophic water from southern Adriatic. At both stations the variance of the system can greatly influence the signal to noise ratio.

5.3 Discussion

5.3.1 Descriptive model for phytoplankton concentration (1st model)

The selected model for station SJ107 (Figure 5.1) indicates that during the 1997 a significant change in the phytoplankton dynamics occurred. The phytoplankton concentration was higher before 1997 (2 to 4 times), mainly due to changes in the Po River discharge rates. The calculated limiting flow rate amounts to 2 025 m^3/s. Higher rates than the limiting value imply an increase of the phytoplankton concentrations.

However, in both cases (before and after 1997) salinity is the main signal indicating changes of the impact of freshwater inputs to the area, but also of inflow of more saline waters from the central Adriatic. A reduction of riverine nutrients input and extended saline waters intrusion contributed to lower phytoplankton concentrations after 1997, most often throughout the investigated area of the NA. The reduction of the Po River discharges became drastic after 2002, with a consequent oligotrophication of the observed ecosystem (Mozetič *et al.*, 2009).

The changes in the 1993 and 2000 (Figure 5.1) are difficult to understand but coincide with unusually high freshwater discharges in the NA in autumn (Supić *et al.*, 2006). In October 1993 the Po River flow rates were markedly higher than any monthly averages through the year since 1917 when the measurements started. Exceptionally high flows occurred also in the second part of October and November 2000, respectively. Unusually marked stratification persisted also in December, due to the presence of a thick freshened surface layer. In these conditions an extended near-anoxia developed in the bottom layers, as never previously observed since 1972 in the investigated area (CMR, Rovinj data base).

In addition, the changes since 1984 can also be related to the reduction of polyphosphate contents in detergents, with a consequent marked decrease of phosphorus compound in river waters (Provini *et al.*, 1992, Pagnotta *et al.*, 1995).

5.3.2 TIN/PO_4 model describing mucilage events (2nd model)

The model (Figures 5.2, 5.3, 5.4 and 5.5) indicates that high average values of the TIN/PO_4 ratio coincide with observed mucilage events. Recall that four mucilage events occurred between 1982 and 1999 (1988, 1989, 1991 and 1997 (mucilage was observed, although less visible in the largest part of the NA; http://lepo.it.da.ut.ee/~olli/eutr/html/htmlBook_134.html), four between 2000 and 2003 (2000, 2001, 2002, 2003 (this event lasted less than one month), one in 2004 and one in 2007; Stachowitsch *et al.*, 1990; Degobbis *et al.*, 1995, 1999; Vollenweider *et al.*, 1995; Cozzi *et al.*, 2004; Precali *et al.*, 2005; De Lazzari *et al.*, 2008; CMR, Rovinj, unpub.

data). This indicates that at certain levels of P limitation (TIN/PO_4 signal clearly indicate) mucilage events frequency increase.

The complete mechanism of mucilage formation could not be revealed in detail by this model due to insufficient amount of descriptors in the data set. Undoubtedly, the mucilage formation is a result of multiple factors, where the TIN/PO_4 ratio is one of them but not the only one. In this research mucilage events are identified indirectly, through the values of the TIN/PO_4 ratio solely. Given that the TIN/PO_4 ratio is one of the needed conditions for mucilage appearance, other conditions not revealed in this study, are needed to be fulfilled for the mucilage appearance. The model is induced solely from measured data, and thus, if such information (pattern) does not exist in the training data it cannot appear in the model. Instead, the model indicates the favourable conditions for mucilage development, which could be revealed from the learning data set. The coincidence of the high TIN/PO_4 ratio with the mucilage events clearly indicates that the P limitation is one of the main triggering mechanisms. Still, from this ratio alone it cannot be reliably concluded if its effect on bacteria could affect the degradation of organic matter favouring accumulation of mucilage (e.g. Azam *et al.*, 1999; Pugnetti *et al.*, 2005).

Related to previous research of the mucilage phenomena in the NA, the model confirms some of the results, particularly those related to the effects of salinity and temperature on mucilage formation, e.g. that the mucilage phenomenon is primarily developed in lower salinity (32 to 37) and oxygenated surface waters (Degobbis *et al.*, 2005). Recall, threshold values of salinity in the model are 36.9, 36.34, etc. Precali *et al.* (2005) showed that a major number of aggregates accumulated in correspondence with strong pycnoclines with differences in density anomaly of 2 kg/m^3 or higher, due to temperature and salinity vertical changes. Observations of mucilage events in 2000, 2001, and 2002 suggest that increased air and sea temperature could play a role, even though secondary, in the mucilage phenomenon (Russo *et al.*, 2005). Salinity, temperature and other factors are also important for growth of planktonic algae in NA and can lead to intense blooms in marine coastal waters (Cucchiari *et al.*, 2008). Finally, Deserti *et al.* (2005) grouped the mucilage events in three main clusters. While they identify: (1) 1920 to 1930, (2) 1983 to 1991 and (3) 1997 to 2002, the model developed here groups them into: (1) 1982 to 1999, (2) 2000 to 2003, (3) 2004 and (4) 2005 to 2007.

Additionally to the previous research, the model developed here, reveals the threshold values of salinity (SAL) and temperature (Temp) in the entire observed period (1972 to 2007) that lead to high values of the TIN/PO_4 ratio as indicator for mucilage events.

To demonstrate the model results the graphically presentation is given for the mucilage events together with the environmental variables (temperature and salinity) in the period 1982 to 1999 (Figure 5.16), 2000 to 2003 (Figure 5.17) and 2004 to 2007 (Figure 5.18). Please note that this is not an exact representation, as the data are taken at different time scales. While temperature and salinity are measured monthly, the mucilage events are observed in approximate time periods of their appearance (CMR, Rovinj, unpub. data). Still the Figures clearly confirm the results from the model presented in Figures 5.2, 5.3, 5.4 and 5.5. On each figure salinity (SAL), temperature (Temp) and TIN/PO_4 ratio values are presented together with their threshold values according to the model presented in Figures 5.2, 5.3, 5.4 and 5.5.

The mucilage events occur when salinity (less than 36.9 for period 1982 to1999 and 36.34 for 2000 to 2003) and temperature (less than 19.6 °C for period 1982 to 1999 and 21 °C for 2000 to 2003) are below the marked thresholds and above the predicted average value of TIN/PO_4 ratio with the model (above 118.7 for period 1982 to 1999 and 125.3 for 2000 to 2003, see Figures 5.16 and 5.17).

The mucilage event in 2004 occurred when salinity and temperature (less than 37.2 and 19.4 °C respectively, see Figure 5.18), are below the marked thresholds and above the predicted average value of TIN/PO_4 ratio with the model (above 210, Figure 5.3). As said before, mucilage event in 2007 is specific because mucilage events were observed in Spring-Summer and in Autumn-Winter period (see Figure 5.18). Currently, mechanism that leads to the mucilage events in Autumn-Winter period is not clear. In 2007 the ratio values in whole period are high (TIN/PO_4 = 118-187)

Two important phenomena are also revealed on Figures 5.16, 5.17 and 5.18, which cannot be revealed by the model, for occurrence of mucilage, e.g.: (1) the mucilage events occur when sea-water temperature is rising (see also Degobbis *et al.*, 2005 and Russo *et al.*, 2005), and (2) an obvious increase of phytoplankton concentration (Phyto) before the mucilage events (see also Totti *et al.*, 2005).

The temperature increase during a mucilage event cannot be confirmed with the model, as it only reveals the threshold value of the temperature below which mucilage occur, e.g. the limit value of the temperature increasing, whereas phytoplankton does not appear in the model at all.

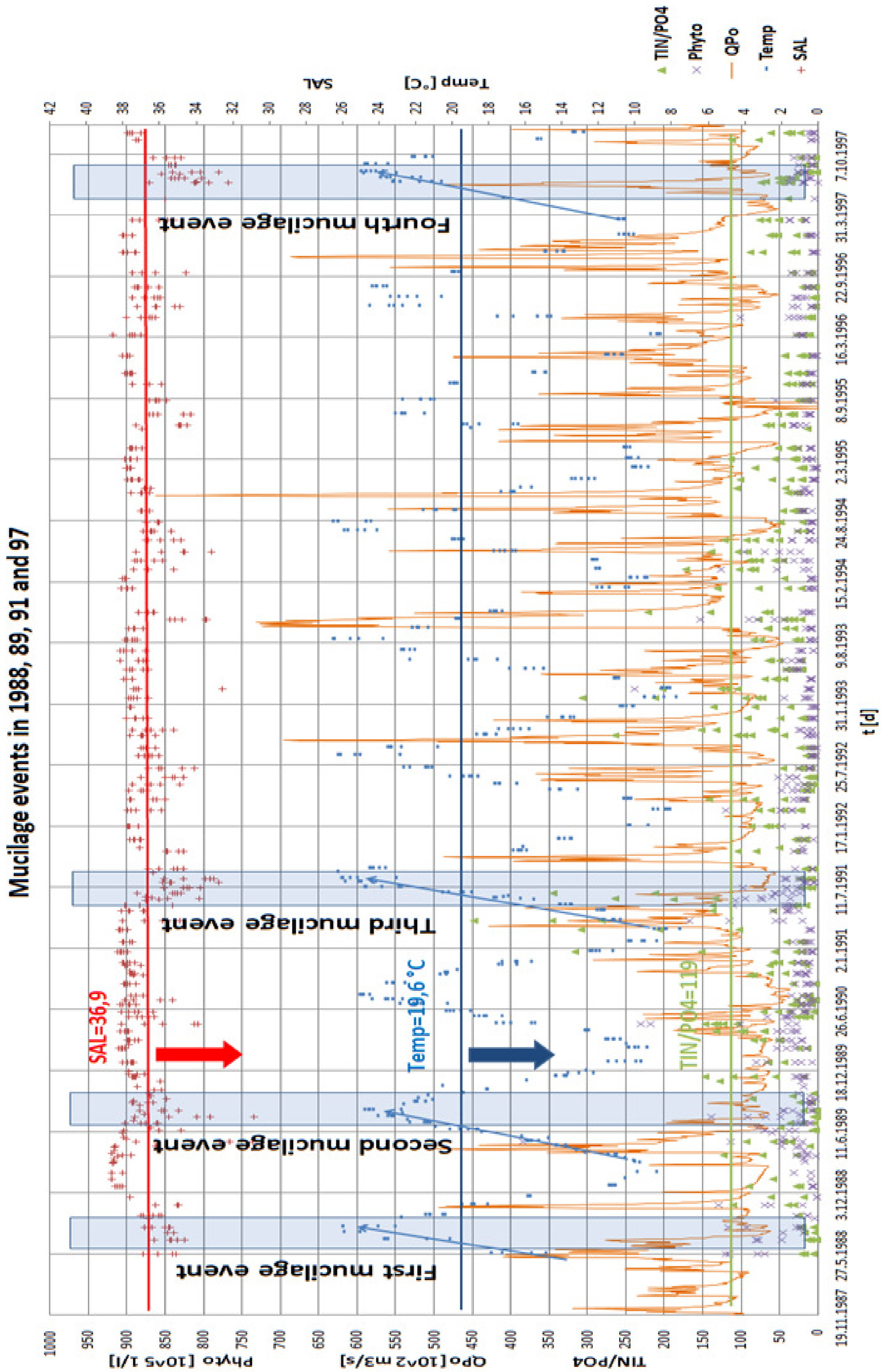

Figure 5.16 Mucilage events in period 1982 to 1999

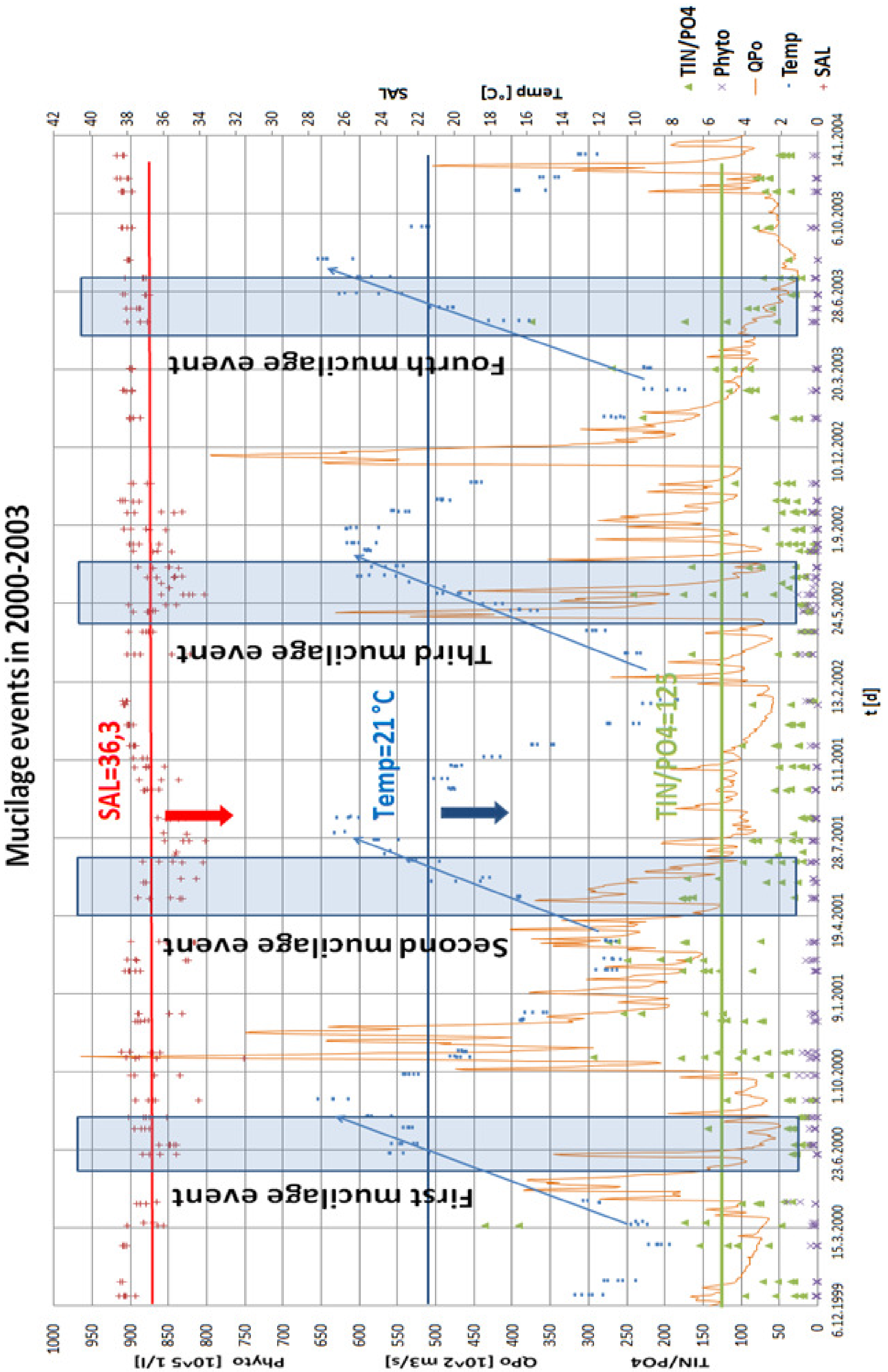

Figure 5.17 Mucilage events in period 2000 to 2003

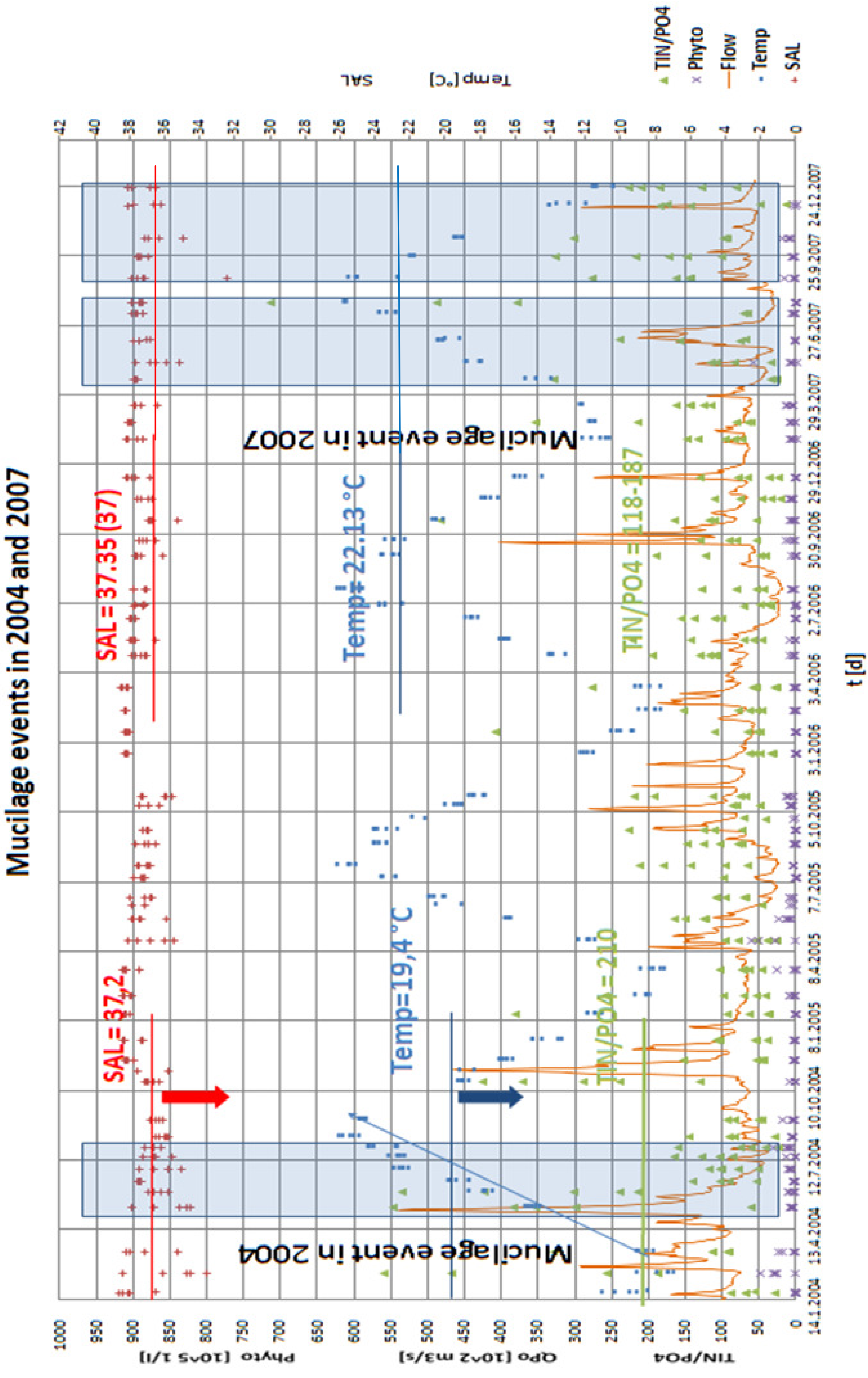

Figure 5.18 Mucilage events in 2004 and 2007

5.3.3 Predictive model for phytoplankton concentration (3rd model)

A simple and yet efficient phytoplankton prediction model (Table 5.2) was developed for the NA using ML tools. The model calculates the phytoplankton concentration 14 days in advance, given the present observed values of thirteen variables and can thus be used as a warning tool for water management purposes. Previous modelling attempts in this ecosystem mainly include more complex conceptual models describing and simulating the phytoplankton seasonal cycle and its horizontal distribution (Zavatarelli *et al.*, 2000), trophic networks (Barausse *et al.*, 2009) or sporadic algal blooms (Thornton *et al.*, 1999). While such approaches are very useful for understanding the ecosystem behaviour under different conditions, they usually lack the predictive power needed for the efficient management of water quality.

Unlike the mentioned conceptual models which are derived from the theoretical modelling knowledge, the model developed in this research is induced from measured data and therefore lacks the descriptive power of the conceptual models. Still, in its structure the discovered model includes variables (total phytoplankton, temperature, TIN/PO_4, month (April) and NH_4) which are very important for prediction of phytoplankton concentrations in this ecosystem, as also indicated by other researches. Namely, different species of phytoplankton grow in various temperature intervals (Cucchiari *et al.*, 2008); the TIN/PO_4 molar ratio can be assumed as an indicator of nutrients limitation of algal blooms (Redfield, 1934); the month (April) is also important because the phytoplankton blooms in spring or early summer are extended over larger areas, interesting the most part of the investigated stations (Degobbis *et al.*, 2000); and finally, NH_4 is the nitrogen compound preferred in the algal assimilation processes (e.g. Raymont, 1980).

Finally, the aim of this (3rd) model is prediction and, given the accuracy of the model performance on unseen data, it can be a highly useful water management tool as a self standing model predicting the phytoplankton concentrations or integrated in more complex watershed models which include nutrient generation watershed activities. Such integrated model can be used to control the nutrient loadings from the watershed.

5.4 Summary

Two ML techniques were applied on long-term measured data in NA to obtain: (1) a descriptive model for the dynamics of phytoplankton concentration (2) a model describing mucilage events and (3) a predictive model for phytoplankton concentration. The phytoplankton descriptive model, e.g. the 1st model, reveals the patterns and the important environmental variables for the phytoplankton dynamics over the period from 1972 to 2007. The model does not achieve high accuracy when tested on unseen data; however it successfully identifies some of the triggers of changes in the phytoplankton dynamics by confirming the hypothesis made in previous research.

The model describing mucilage events or the 2nd model strongly confirmed the assumption that the mucilage events are connected with the changes of TIN/PO_4 ratio in the system, e.g. the model says that mucilage events coincide with the change of the TIN/PO_4 ratio. Four distinctive periods of mucilage appearance were identified (1) 1982 to 1999, (2) 2000 to 2003, (3) 2004 and (4) 2005 to 2007. Additionally to the previous research, the model reveals the threshold values of salinity and water temperature during the entire observed period that lead to high values of the TIN/PO_4 ratio as indicator for mucilage events. Thus at certain levels of P limitation (seen through the TIN/PO_4 ratio) the frequency of mucilage events increases. The induced model for the TIN/SiO_4 ratio is very weak correlated with the mucilage events, thus confirm the experts' notion that orthosilicate is not triggering these events.

The predictive model, e.g. the 3rd model, gives accurate predictions of phytoplankton concentration for 14 days in advance correctly predicting the peak values of the phytoplankton concentration. As such, it can be efficiently used for water management purposes, e.g. as a phytoplankton concentration prediction supplement to watershed models that simulate nutrients loadings and concentrations in the aquatic environment as a consequence of human and natural activities in the watershed (e.g. land use, untreated wastewater etc.).

BLANK PAGE

Chapter 6.

Linking the state of northern Adriatic marine ecosystem to the pressures from surrounding watershed

The machine learning (ML) method Multi Target Stepwise Model Tree Induction (MTSMOTI, Appice and Džeroski, 2007, see Chapter 3, Section 3, Sub-section 3.1.2.3) was applied in this research to find the relationship between the nutrient loads from the watershed and the water quality parameters in northern Adriatic (NA). Estimated water quality parameters were used to calculate the trophic index (TRIX), commonly and lately used measure for evaluation of the trophic state of marine ecosystems.

TRIX (Vollenweider *et al.*, 1998) is calculated as a functional dependency of multiple water quality parameters, as presented in equation 6.1.

$$\mathbf{TRIX = (Log_{10}[Chl\text{-}a \cdot |\mathit{D}\%O| \cdot TIN \cdot P_{tot}] + k) / m} \quad (6.1)$$

where Chl-*a* is the concentration of Chlorophll *a* in µg/l, |*D*%O| is the oxygen as absolute % deviation from saturation, TIN is the concentration of total inorganic nitrogen in µg/l as sum of NH_4, NO_2, and NO_3, P_{tot} is the concentration of total phosphorus in µg/l.

The parameters k = 1.5 and m = 12/10 = 1.2, are the scale coefficients, introduced to fix the lower limit value of the Index and the extension of the related Trophic Scale, from 0 to 10 TRIX units covering a range of trophic conditions from oligotrophy to eutrophy (see Table 6.1). These conditions are tailored to area of Emilia-Romagna (NA), and can be used for whole Adriatic and Mediterranean Sea.

This index summarizes the different factors influencing the trophic state: (1) the productivity factors (biomass as concentration of chlorophyll *a* and dissolved oxygen as absolute deviance from saturation) and (2) the nutritional ones (concentration of total inorganic nitrogen and total phosphorus).

The limits among the rest of the Ecological Quality (EQ) classes (high to bad) were adjusted to the scale proposed by the Organization for Economic Co-operation and Development (Vollenweider and Kerekes, 1982; Navaro *et al.,* 2009) as follows: **ultra-oligotrophic = high**, **oligotrophic = good**, **mesotrophic = moderate**, **eutrophic = poor** and **hypereutrophic = bad**.

TRIX was used in many studies to evaluate trophic conditions of marine ecosystems. Vollenweider *et al.* (1998) on the case of NW Adriatic originally proposed the trophic scale, turbidity and generalized water quality index (TRIX). Giovanardi and Vollenweider (2004) applied TRIX to two areas, Emilia-Romagna in Adriatic and Tuscany in Tyrrhenian Sea which are two trophically different responding coastal systems. In research they illustrate some of the arising interpretative problems in using TRIX. Artioli *et al.* (2005) used TRIX in coastal waters of Po River. In their study they merged two models, one for nutrients loads carried by the Po River and the other for water quality in coastal zone. Ignatiades (2008) used TRIX for scaling the trophic status of the Aegean Sea. Pettine *et al.* (2007) also applied TRIX on Italian coastal waters. Salas *et al.* (2008) apply TRIX in transitional ecosystems, the Mar Menor lagoon in Spain and Mondego estuary in Portugal.

Table 6.1 Classification of the trophic state using TRIX (Navarro *et al.*, 2009)

TRIX	State	Trophic conditions
0-4	High	Ultra-oligotrophic
4-5	Good	Oligotrophic
5-6	Moderate	Mesotrophic
6-8	Poor	Eutrophic
8-10	Bad	Hypereutrophic

6.1 Modelling experiment

For this experiment the algorithm MTSMOTI was used. MTSMOTI induces a regression tree for simultaneous prediction of multiple target (dependant) variables out of a set of independent (descriptors) variables (see Chapter 3, Section 3). The P_{tot}, TIN, Chl-*a* in µg/l and Osat were target (dependant) variables (it must be noted that this data are measured with near monthly frequency), whereas the total monthly nitrogen (WATNtot) and the total monthly phosphorus (WATPtot) from the NA watershed in kg/month (see Table 6.2) were taken as independent variables (descriptors). The dependant variables are measured marine data and they represent the State (**S**) of NA marine ecosystem (see Section 3, Table 3.2, Figure 3.5), while the WATNtot and the WATPtot were simulated with the watershed model (AVGWLF) presented in Chapter 4 and they represent the Pressures (**P**) from the surrounding watershed. Predictive model for the dependant variables was built for each station, e.g., RV001, SJ101, SJ103, SJ105, SJ107 and SJ108. Subsequently, the TRIX, indicating the state of the marine ecosystem was calculated using the equation 6.1. The procedure of the data flow and model induction is presented in Figure 6.1. The model for calculating TRIX is further used for evaluating scenarios regarding the influence of different levels of wastewater treatment to the NA ecosystem and finally for suggesting the proper one in order to preserve good ecological state (Chapter 7).

To test the models' accuracy one of the most commonly used procedures of building and testing models is applied: one portion of the data set is selected for training (60 %) and the other for testing (40 %) from which correlation coefficient (R) between the measured and simulated values was calculated.

Table 6.2 Measured marine data in NA and simulated nutrient loads from watershed used for marine ecosystem state model

Parameter	Interpretation	Unit
Month	Month of sampling	
P_{tot}	Total phosphorus	μg/l
NO_3	Nitrate	μg/l
NO_2	Nitrite	μg/l
NH_4	Ammonium	μg/l
TIN	Total inorganic nitrogen ($NO_3 + NO_2 + NH_4$)	μg/l
Chl-*a*	Chlorophyll *a*	μg/l
Osat	Oxygen saturation	
WATNtot	Total nitrogen from watershed (simulated)	kg/mo
WATPtot	Total phosphorus from watershed (simulated)	kg/mo

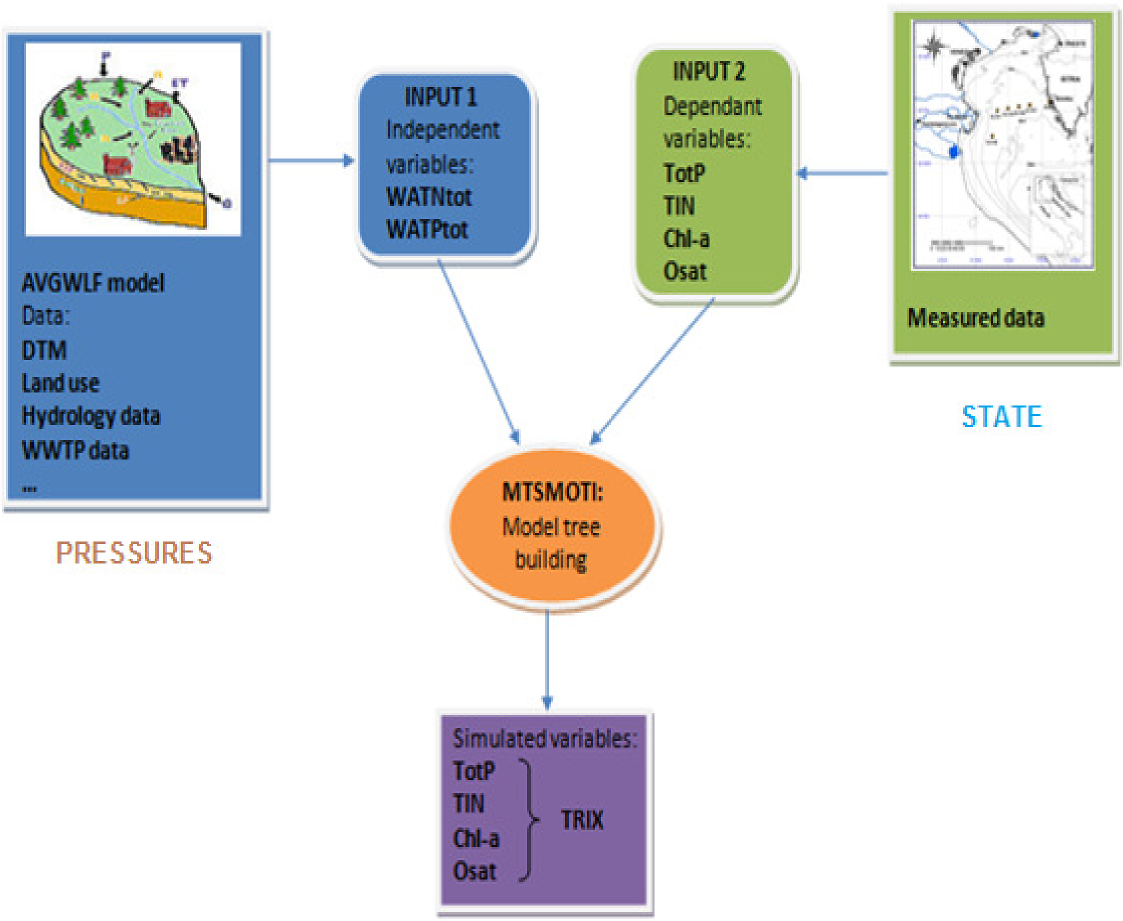

Figure 6.1 Schematically procedure of data flow and resulting models for linking the State (**S**) of marine ecosystem to the Pressures (**P**) from surrounding watershed

6.2 Results and discussion

Using MTSMOTI, a model tree was built for each station, containing a regression equation for each dependant variable in its leaves, e.g. for P_{tot}, TIN, Chl-*a* and Osat. Each model calculates monthly values of parameters from which monthly TRIX is calculated as explained in introduction part of this Chapter. Yearly TRIX values were calculated by averaging the monthly ones.

For further research only station SJ108 is taken into consideration because it is the most critical station where the highest TRIX values were observed (see Figure 6.10). High TRIX values are related to the direct influence of the Po River (see Figure 3.5 in Chapter 3).

Figure 6.2 represents the model tree for station SJ108. Model for this station contains six leaves which contain equations for calculating dependant variables (see Table 6.3) only from simulated nutrient loads from watershed, e.g. independent variables representing inner nodes of the model tree. Similar concept have models for other measurement stations have (not presented here).

Correlation coefficients (R) between measured and simulated values of parameters are represented in Table 6.4. Values of correlation coefficients are relatively acceptable, ranging from minimum value of 0.39 to maximum value of 0.86. The lowest correlation coefficients values for TRIX was observed for stations SJ105 (0.60) and SJ103 (0.61), while the highest values have stations RV001 (0.69), SJ108 (0.69) and SJ107 (0.71).

Table 6.3 Sets of equations for model tree from Figure 6.2

Set 1: Osat = 0.948 + 0.051 Month P_{tot} = 4.309 + 0.760 Month TIN = 71.937 + 0.318 Month Chl-*a* = 0.396 + 0.581 Month	**Set 4:** Osat = 1.372-2.934e-6 WATPtot P_{tot} = 9.411-3.574e-5 WATPtot TIN = 52.410-2.973e-4 WATPtot Chl-*a* = 5.281-1.178e-5 WATPtot
Set 2: Osat = 1.090+9.480e-8 WATNtot P_{tot} = 4.836+1.973e-6 WATNtot TIN = -8.110+7.684e-5 WATNtot Chl-*a* = 0.122+3.205e-6 WATNtot	**Set 5:** Osat = 1.279-0.025 Month P_{tot} = 9.522-0.196 Month TIN = 85.558-1.301 Month Chl-*a* = 3.627-0.197 Month
Set 3: Osat = 1.071+9,388e-8 WATNtot P_{tot} = -0,0865+9.004e-6 WATNtot TIN = 6.024+1.244e-5 WATNtot Chl-*a* = -1.143+3.376e-6 WATNtot	**Set 6:** Osat = 0.955+4.969e-8 WATNtot P_{tot} = 19.894-3.454e-6 WATNtot TIN = 503.618-1.244e-4 WATNtot Chl-*a* = 6.744-8.843e-7 WATNtot

Table 6.4 Correlation coefficients (R) between measured and simulated values of parameters for each station

Stations / Parameters	SJ108	SJ101	SJ103	SJ105	SJ107	RV001
Osat	0.73	0.53	0.39	0.52	0,76	0.71
Chl-a	0.61	0.51	0.41	0.58	0.86	0.61
TIN	0.72	0.71	0.68	0.63	0.62	0.79
P_{tot}	0.6	0.52	0,49	0.48	0,48	0.45
TRIX	**0.69**	**0.66**	**0.61**	**0.60**	**0.71**	**0.69**

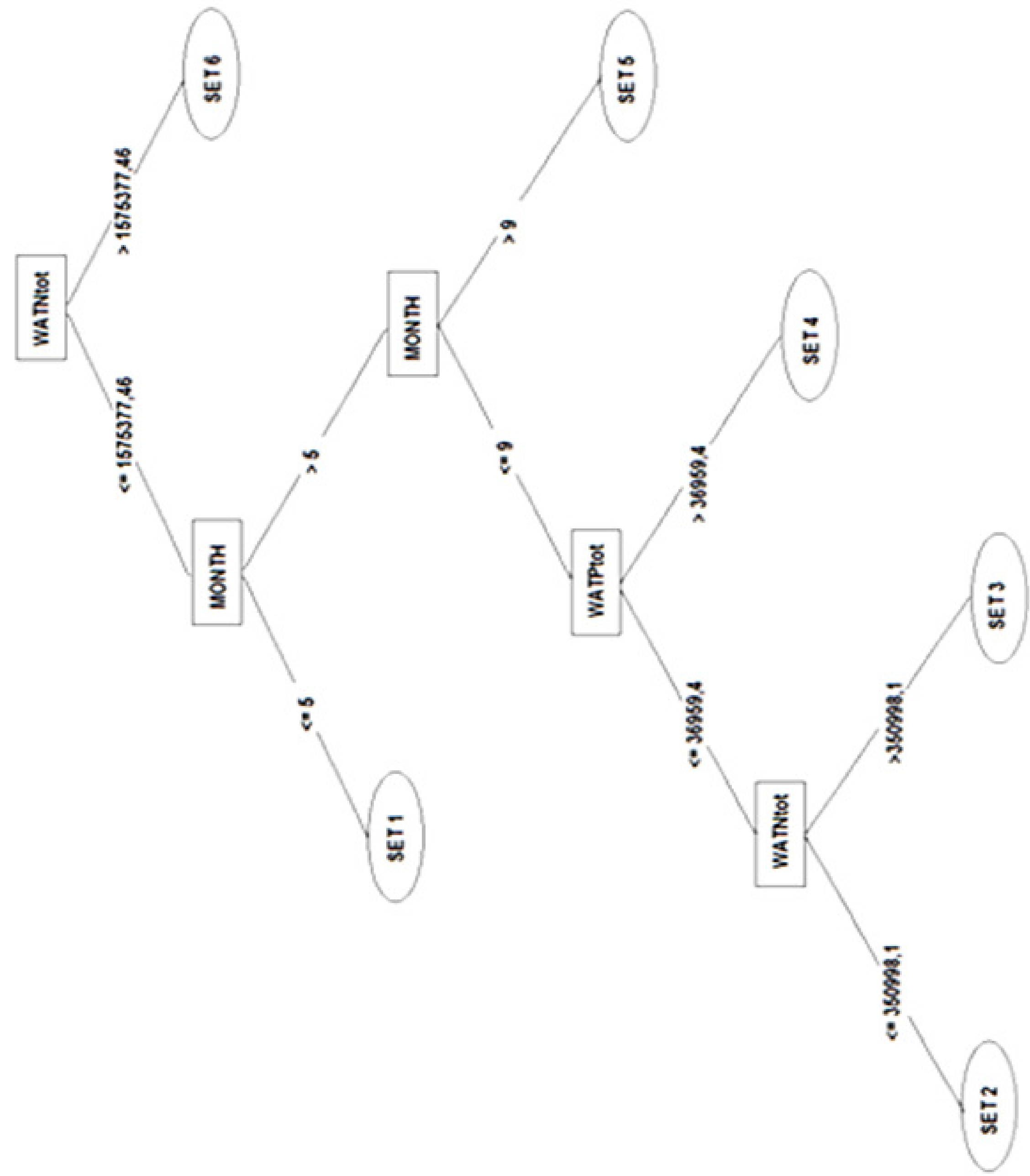

Figure 6.2 The model tree for station SJ108 (units for the threshold values for the parameters used are reported in Table 6.2)

Simulated and measured monthly values of parameters on station SJ108 which are used for calculating TRIX (Chl-*a*, Osat, P_{tot} and TIN) are presented on Figures 6.3 to 6.6. The good coincide of measured and simulated values is observed on figures, with very good matching of peak values for some parameters.

For Osat, P_{tot} and TIN almost all (small and high) peak values have very good matching between measured and simulated values. For Chl-*a* very good matching is only observed for small peak values, while for high peak values only position in time is accurate.

In addition to simulated and measured values of P_{tot} on Figure 6.5 simulated phosphorus loads from watershed (WATPtot) are also presented. Simulated and measured values of TIN and simulated nitrogen loads from the watershed (WATNtot) similarly as on Figure 6.5 are presented on Figure 6.6. Comparison of phosphorus and nitrogen loads from the watershed with measured/simulated values of P_{tot} and TIN respectively shows that nutrient loads from watershed almost immediately increase nutrient concentrations in sea-water.

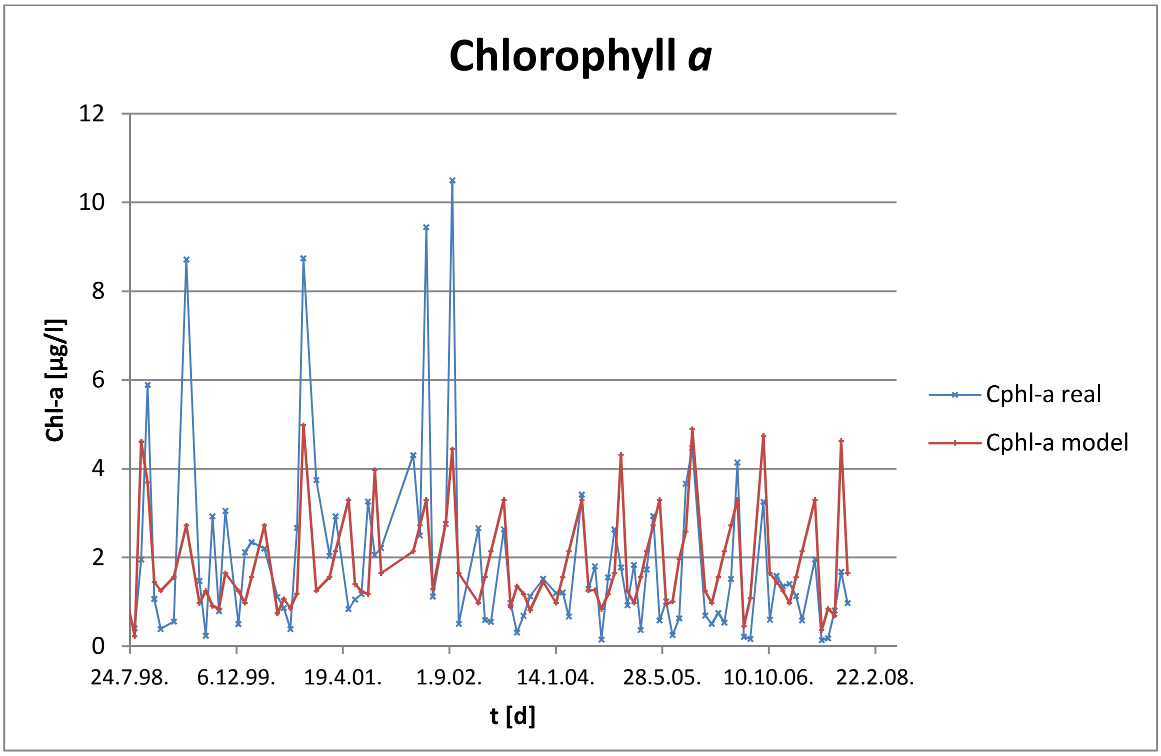

Figure 6.3 Simulated and measured values of Chlorophyll *a* (Chl-*a*) on SJ108 in period 1999 to 2007

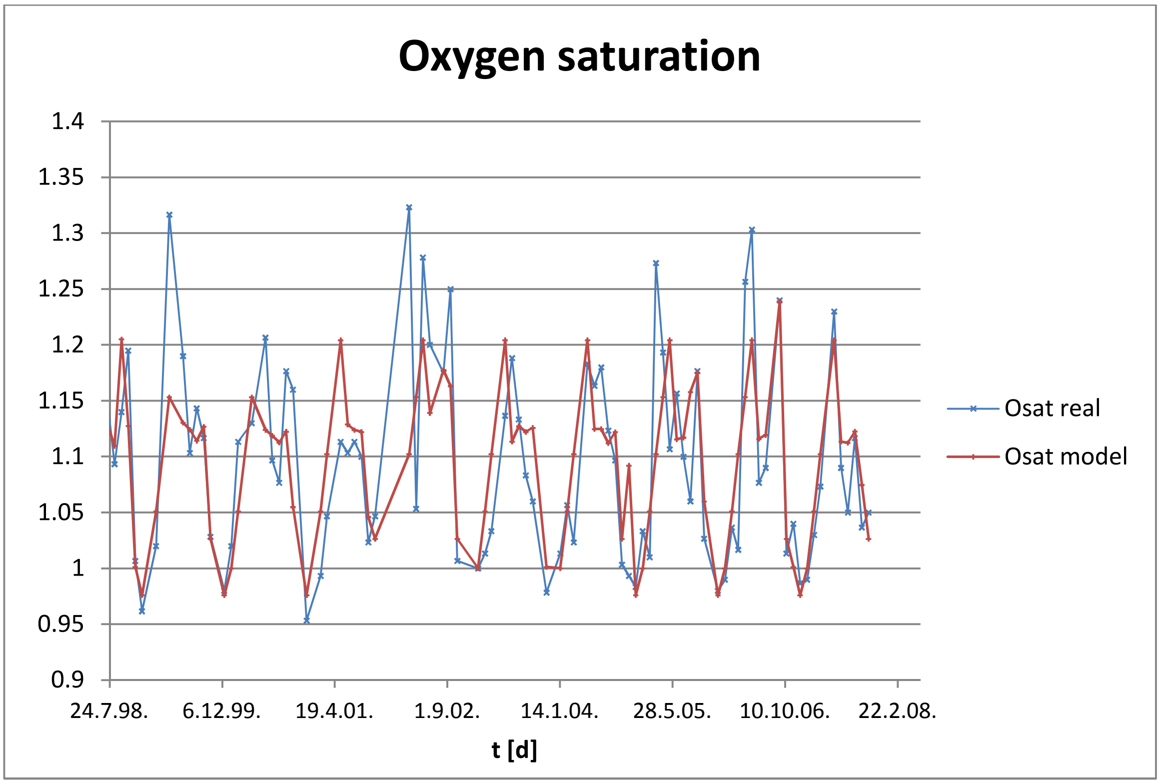

Figure 6.4 Simulated and measured values of oxygen saturation (Osat) on SJ108 in period 1999 to 2007

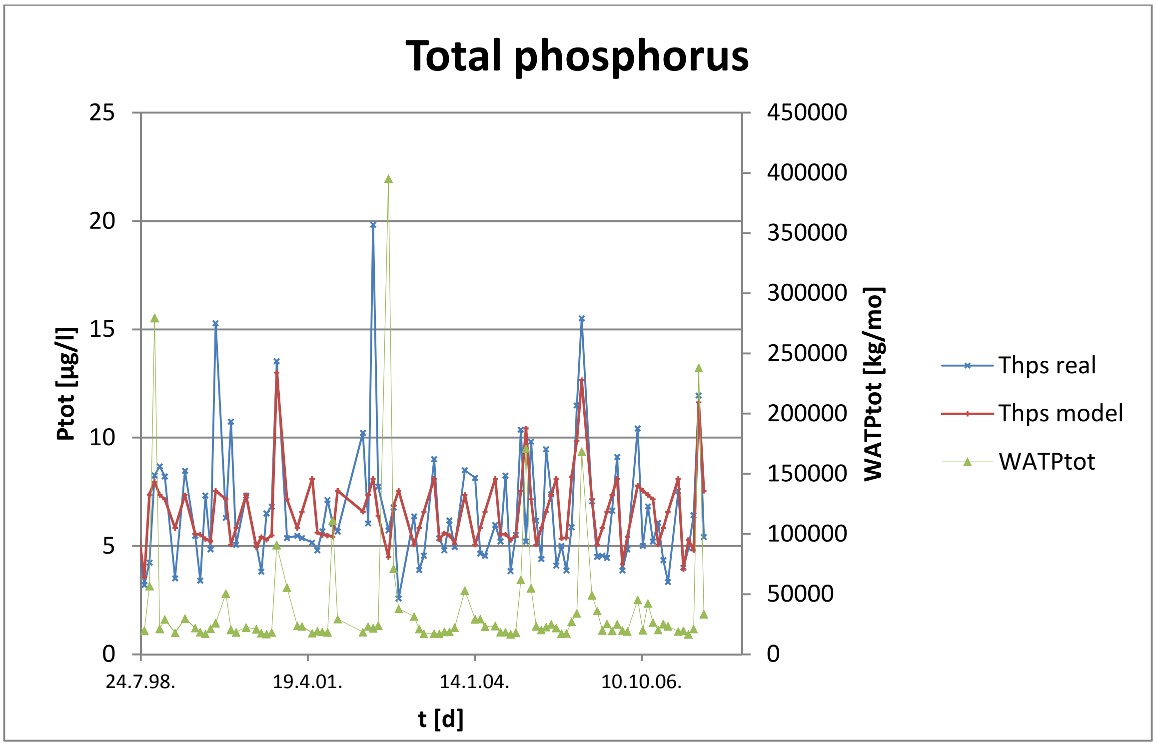

Figure 6.5 Simulated and measured values of total phosphorus (P_{tot}) on SJ108 and phosphorus load from watershed (WATPtot) in period 1999 to 2007

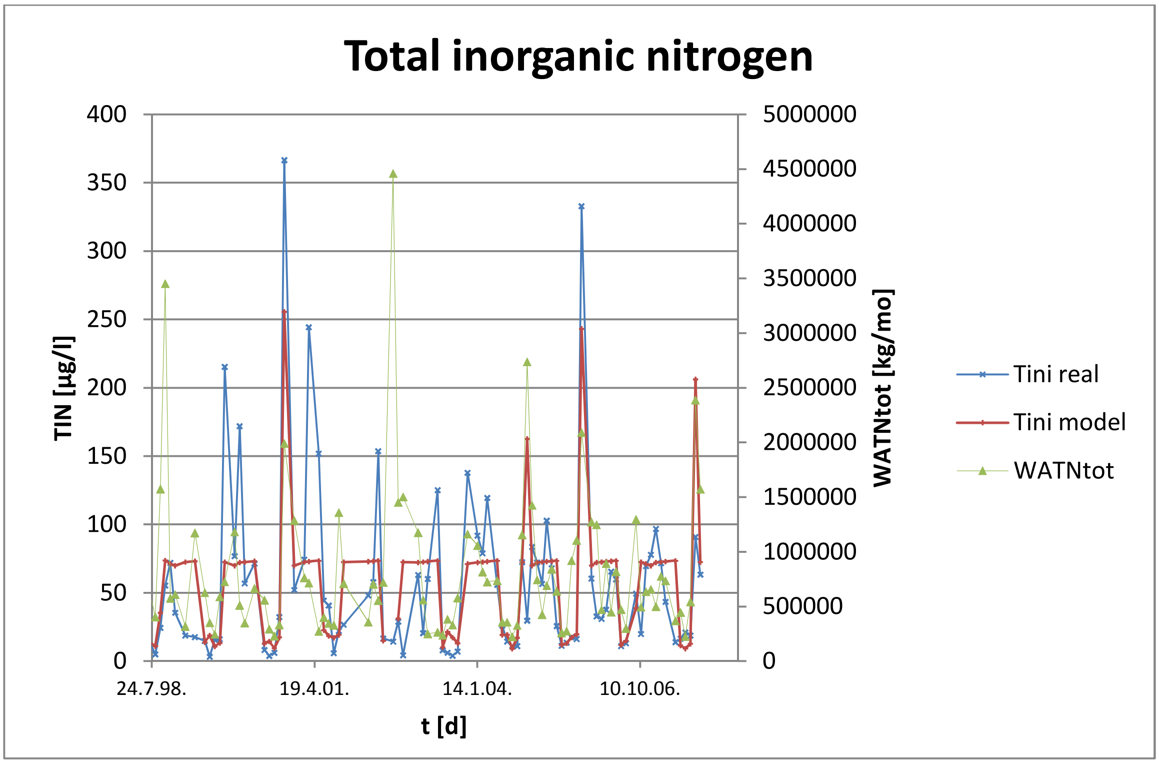

Figure 6.6 Simulated and measured values of total inorganic nitrogen (TIN) on SJ108 and nitrogen load from watershed (WATNtot) in period 1999 to 2007

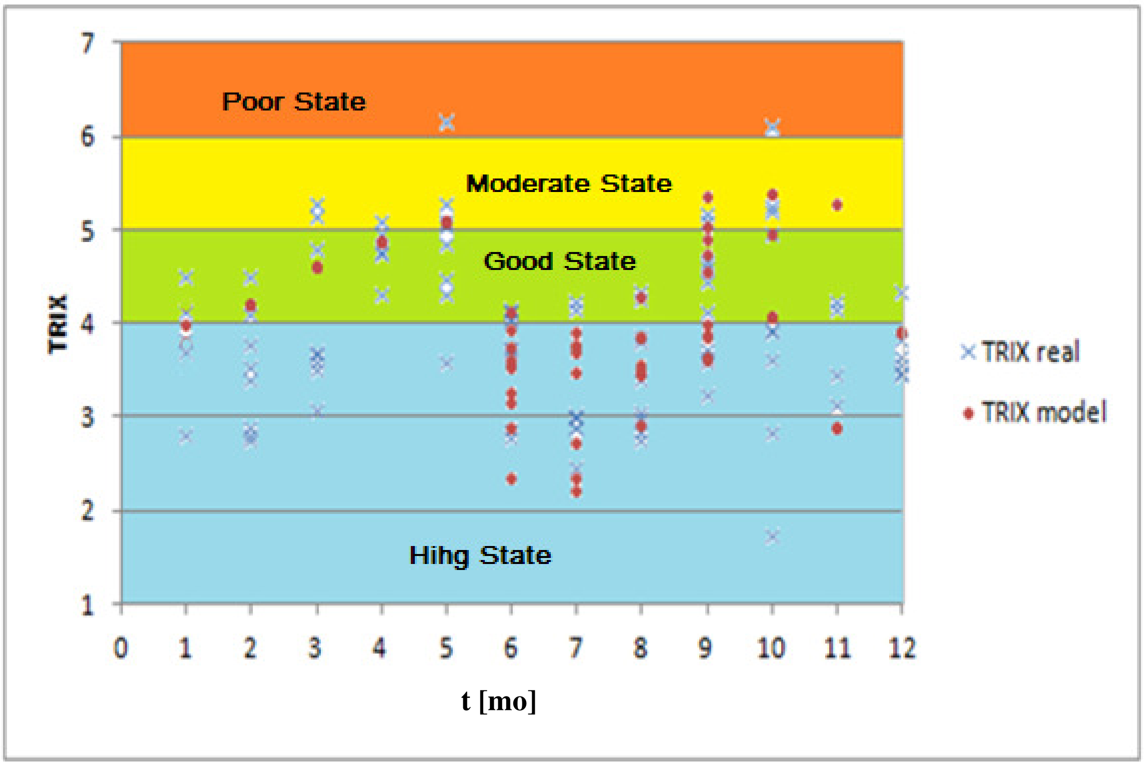

Figure 6.7 Monthly values of TRIX in period 1999 to 2007 for station SJ108 (from measured and simulated data)

Monthly TRIX values from simulated and measured parameters for station SJ108 are presented on Figure 6.7. Simulated values cover large part of measured ones, with best coverage from June to October. For other months simulated covers only one to two of measured values. In general, for those months there is a lot less of measured data available than for other months so the model (Figure 6.2) did not have enough input data.

Annual values of TRIX with simulated nutrient loads from watershed are presented on Figure 6.8, while monthly values of TRIX with simulated nutrient loads from watershed in period 1999 to 2007 are presented on Figure 6.9. TRIX values on both figures are in correlation to nutrient loads from watershed, e.g. higher nutrient loads give higher TRIX values which can best be seen on Figure 6.8 for year 2002.

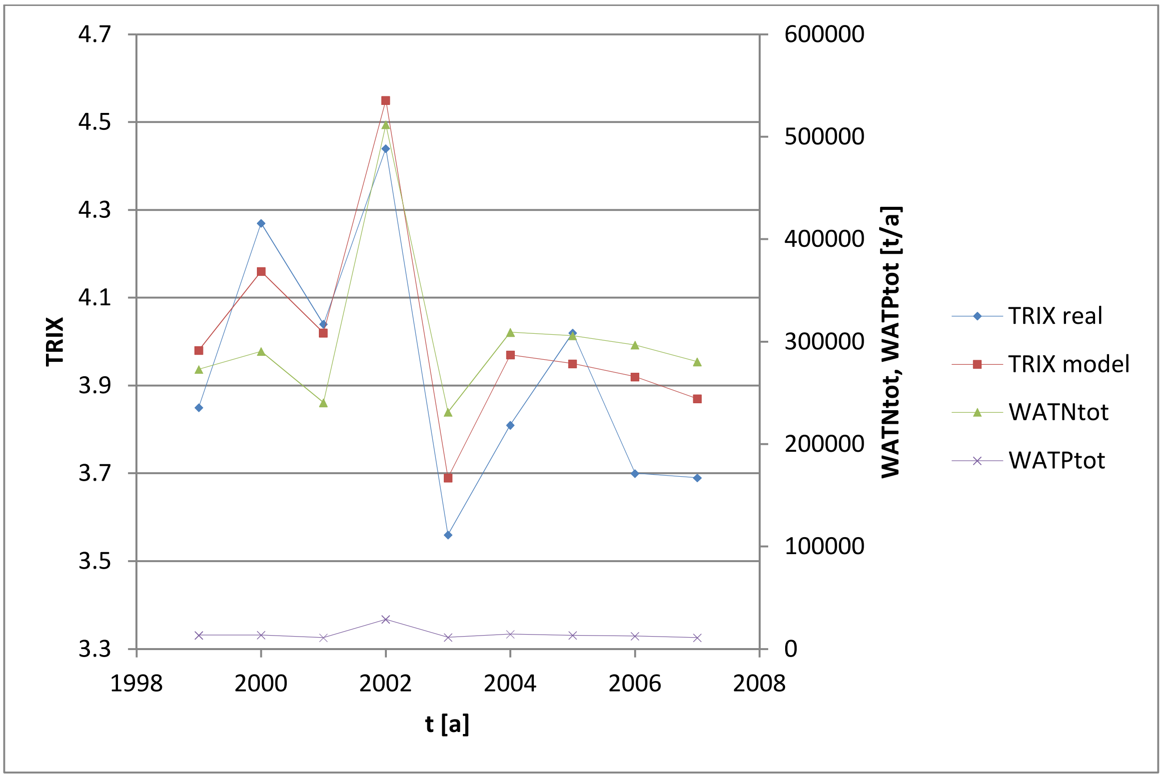

Figure 6.8 Annual TRIX values on station SJ108 and simulated nutrient loads from watershed (WATNtot and WATPtot) in period 1999 to 2007

Average TRIX values for all measurement stations are presented on Figure 6.10. TRIX values for each station calculated from measured and from simulated values have very good matching. Also, clear graduation of TRIX can be observed from station SJ108 (nearest to the Po River delta) to station RV001 (near Rovinj).

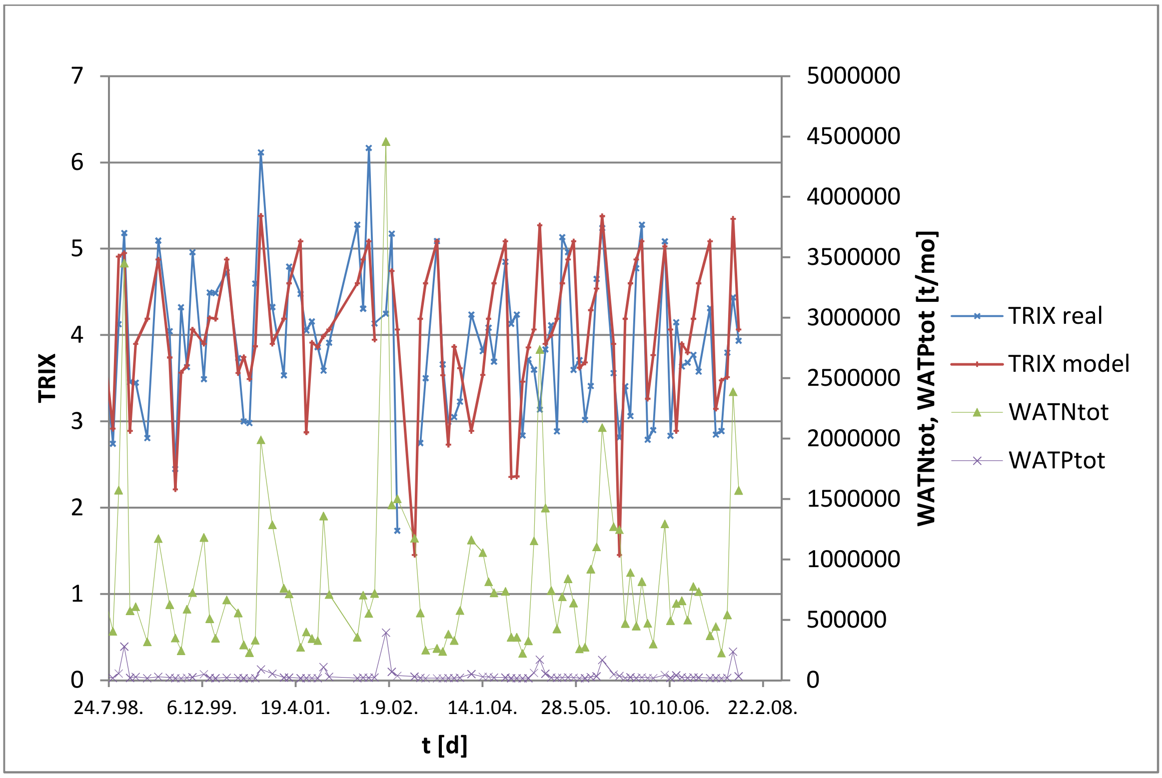

Figure 6.9 Monthly TRIX on station SJ108 and simulated nutrient loads from watershed (WATNtot and WATPtot) in period 1999 to 2007

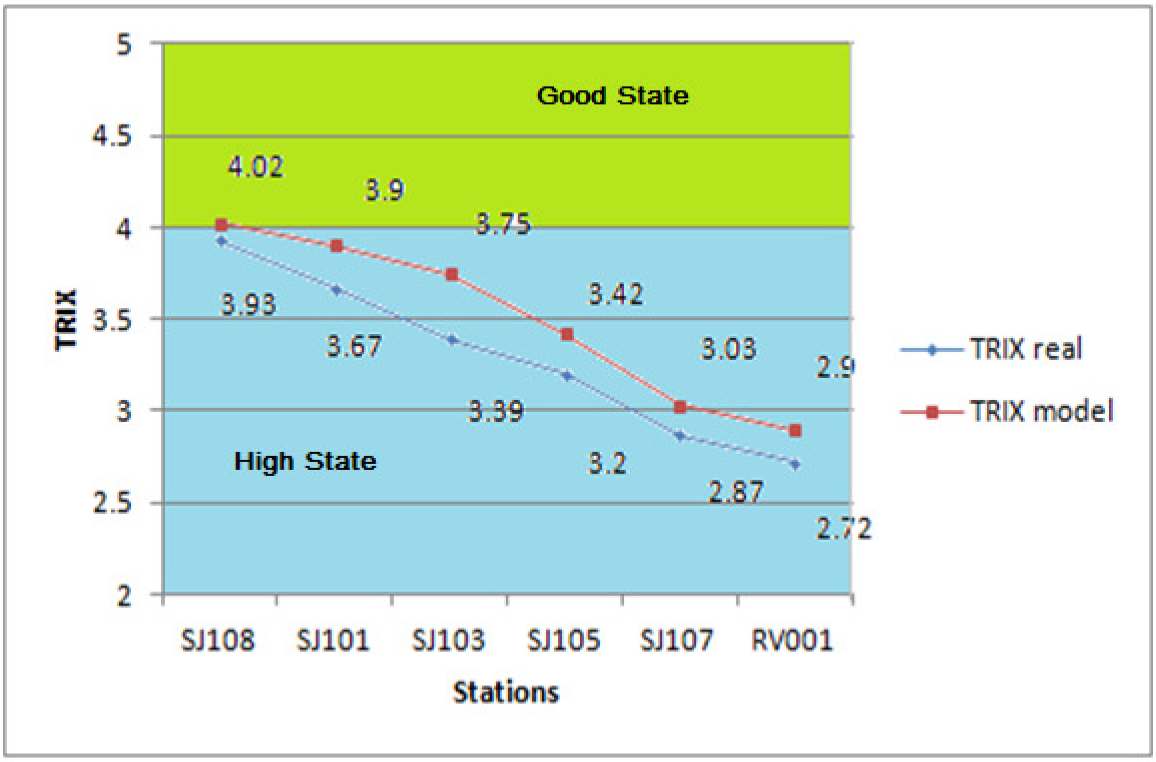

Figure 6.10 Average TRIX values for period 1999 to 2007 (from measured and simulated data) for all measurement stations

6.3 Model verification

To simulate future scenarios (see Chapter 8, Section 8.1) in order to evaluate the impact of variations in nutrient loads from the NA watershed on the marine ecosystem, as consequence of different watershed management strategies the model was tested. To test the model, for "Current State" monthly nutrient loads for station SJ108 calculated in AVGWLF model have been averaged for each month. These average monthly loads have then been used in the model to calculate monthly values of Chl-a, TIN, P_{tot}, Osat and TRIX. Annual TRIX value for "Current State" is obtained by averaging the monthly ones (see Figure 6.11, "Current State" = 0.0 %, TRIX = 4.0). Changing of TRIX for increasing or decreasing nutrient loads (both N and P for the same extent) is presented on Figure 6.11. Operating boundaries for the model are from -75 % to +60 % of average monthly nutrient loads.

To get from current state (TRIX value is 4.0), which in this case represents the boundary between **High** and **Good State** to **Moderate State** (TRIX value is between 5 and 6) nutrient loads have to been increased for about 60 %. To obtain TRIX value of 3.0 nutrient loads have to be decreased for more than 50 %.

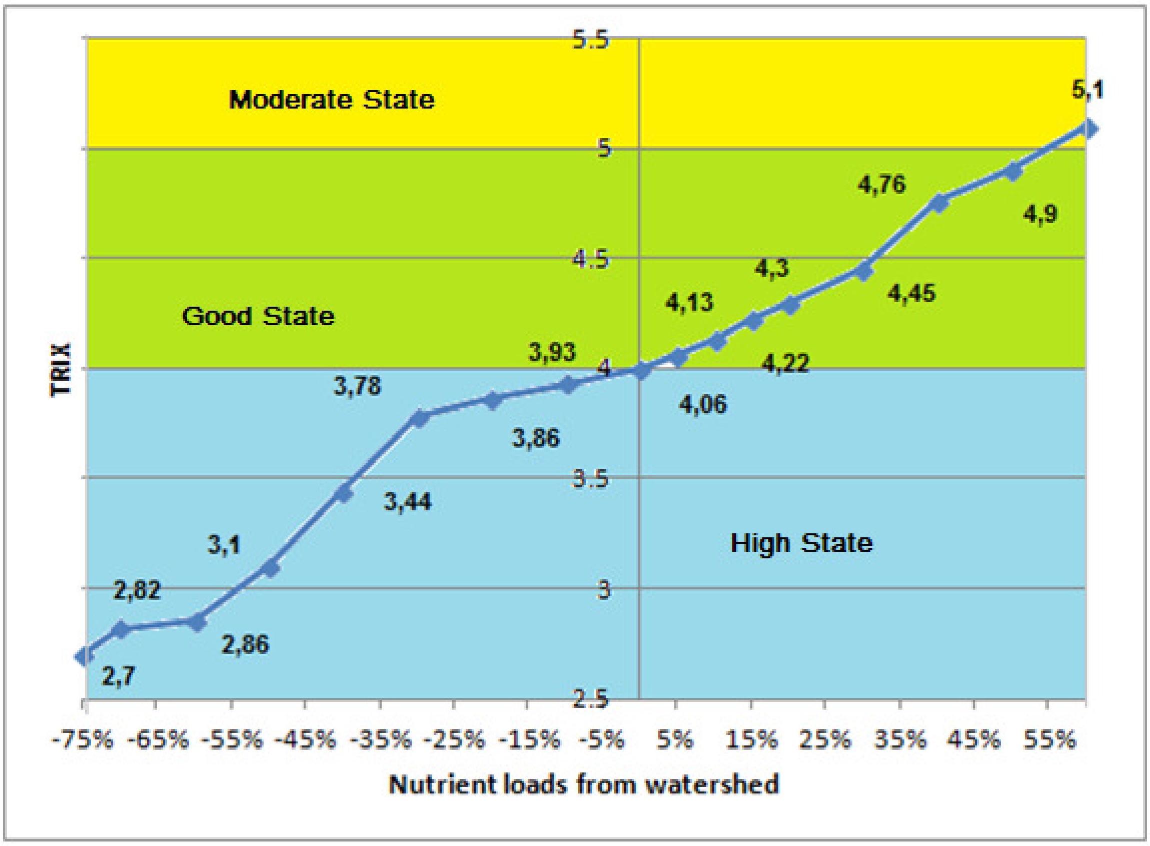

Figure 6.11 TRIX values for different variations in nutrient loads from watershed

6.4 Summary

Two modelling domains have been merged in this research using ML tools. One domain represents the State of marine ecosystem and other Pressures from surrounding watershed that are affecting that State. With ML tool MTSMOTI the model defining the State of NA was developed. Model simulates Chl-a, TIN, P_{tot} and Osat, variables that describe the State, using only nutrient loads (Pressures) from watershed (WATNtot and WATPtot). Simulated variables are used for calculation of trophic index (TRIX) through which the trophic state of marine ecosystem is evaluated. From previous remarks can be seen that the model has good agreement with observed data.

The model's operating boundaries were determined by simulating the state variables with variable nutrient loads from the watershed and calculating the TRIX. The model responds reasonable within the range of -75 % to +60 % of average monthly nutrient loads. Correspondingly the TRIX changes from High to Moderate State with respect to the input of nutrients' loads.

It would be important for further research to have deeper insight into nutrient dynamics and more data from surrounding watersheds in order to create more extensive database which can then be processed with ML tools. Also, a linkage to a hydrodynamic model would increase the understanding of the marine ecosystem functioning and improve the predictions of future scenarios.

Chapter 7.

Summary results and discussion

Part **R** e.g. Responses of Driving forces-Pressures-States-Impacts-Responses (DPSIR) framework will be presented in this chapter. This will be done through scenarios evaluation and proposal for optimal watershed management (defining the proper wastewater treatment level and agricultural fertilization intensity in watersheds, etc.). Also "Outlooks" (**O**) for the State of the marine environment will be done. Namely, what will happen to that state over time based on various scenarios. Before going further summary results of all tasks will be presented:

(1) Nutrient loads (e.g. Pressures) have been calculated for northern Adriatic (NA) watershed in period 1999 to 2007 using a Geographic Information System (GIS) watershed model ArcView Generalized Watershed Loading Function (AVGWLF, See Chapter 4). Also, major sources of nutrients and contribution of nutrient loads have been determinate here.

(2) The State of NA marine ecosystem was assessed by developing two descriptive models and one predictive model (see Chapter 5). Descriptive models have been done for dynamic of phytoplankton concentration and for TIN/PO_4 ratio which describes mucilage events in NA in period 1972 to 2007. Prediction model has been done for phytoplankton concentration which can calculate phytoplankton concentrations 14 days in advance. Machine learning (ML) tools Weka and Cubist have been used to build up the models.

(3) Simulated Pressures from watershed have been linked to the State of NA marine ecosystem (see Chapter 6). The link between Pressures and State was done using ML algorithm Multi Target Stepwise Model Tree Induction (MTSMOTI) for the most critical station SJ108 with the highest TRIX values.

To propose optimal watershed management (e.g. Responses) using scenarios evaluation, the procedure described in Chapter 6 will be followed. Briefly, the State of marine ecosystem will be linked to the Pressures from surrounding watershed using ML tool MTSMOTI which in a form of a model tree simulates variables for calculating TRIX which represents the State. The procedure for proposing optimal watershed management using the simulated variables for TRIX calculation is presented on Figure 7.1. In optimal watershed management are included: (1) defining the proper wastewater treatment level, (2) controlling the nutrients in agriculture, etc.

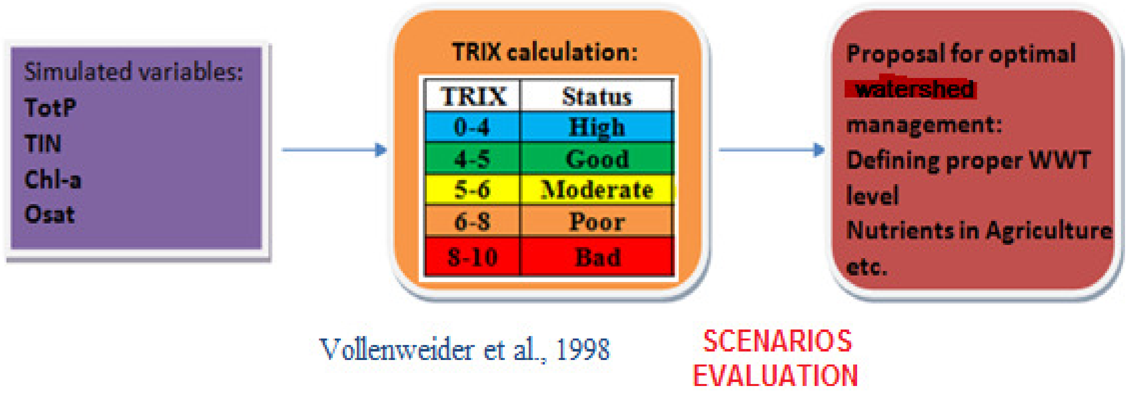

Figure 7.1 Use of simulated variables from built models to calculate TRIX and to propose optimal watershed management

7.1 Scenarios evaluation

For optimal watershed management several scenarios have been developed. In Chapter 4 when calculating nutrient loads average treatment efficiency has been adopted for all the watersheds by applying reduction coefficients (around 40 % for nitrogen (N) and around 30 % for phosphorus (P), e.g. secondary treatment of wastewater). This scenario has been taken as a base scenario which represents “Current State” of the marine ecosystem (see Chapter 6). Among this, ten scenarios have been developed which are described in Table 7.1. TRIX values for each scenario are presented on Figure 7.2 while discussion of the scenarios is given in Section 7.2.

Table 7.1 Description of the scenarios evaluation

Scenario	Description of the scenario
Base Scenario	Secondary treatment level of wastewater
First Scenario	No treatment of wastewater
Second Scenario	Primary treatment level of wastewater
Third Scenario	Tertiary treatment level of wastewater
Fourth Scenario	Case 3 + reduction in agriculture for 10 % (both N and P)
Fifth Scenario	Case 3 + reduction in agriculture for 20 % (both N and P)
Sixth Scenario	Case 1 + increase in agriculture for 10 % (both N and P)
Seventh Scenario	Secondary treatment + reduction in agriculture for 20 % (both N and P)
Eighth Scenario	Secondary treatment + increase in agriculture for 20 % (both N and P)
Ninth Scenario	Secondary treatment + increase in population for 5 %
Tenth Scenario	Secondary treatment + increase in population for 10 %

As mentioned above, for base scenario was taken secondary treatment of wastewater which removes 40 % of N and 30 % of P. This scenario has been presented in Chapters 4 and 6. Detailed description of other scenarios is given in text bellow.

In the first scenario there is no treatment of wastewater, e.g. wastewater is directly discharged into recipient. In the second one primary treatment of wastewater is selected. This treatment removes 25 % of N and 10 % of P. Third scenario represents tertiary treatment of wastewater in which is removed more than 90 % of N and P. In the fourth scenario among tertiary treatment of wastewater is also given reduction of nutrients (both N and P) in agriculture for 10 %. The fifth scenario is the same as fourth, except reduction of nutrients in agriculture is 20 %. In the sixth scenario with no treatment of wastewater (First scenario) is also given increase of nutrients in agriculture for 10 %. Scenarios from seven to ten represents base scenario (Secondary treatment) with reduction of nutrients (both N and P) in agriculture for 20 % (Seventh Scenario) increase of nutrients in agriculture for 20 % (Eighth Scenario), increase in population for 5 % (Ninth Scenario) and increase in population for 10 % (Tenth Scenario). Population that has been increased in Ninth and Tenth Scenarios is assumed that has been treated with secondary treatment of wastewater.

As stated in Chapter 4, Wastewater Treatment Plants (WWTP) make 27 % of load for N_{tot}, and 43 % for P_{tot} (see Figures 4.4 and 4.5 in Chapter 4).

Should be noted that in all developed scenarios with regards on WWTP into consideration is not taken the effect of organic (carbon) load to marine ecosystem, it is taken only effect of nutrient loads.

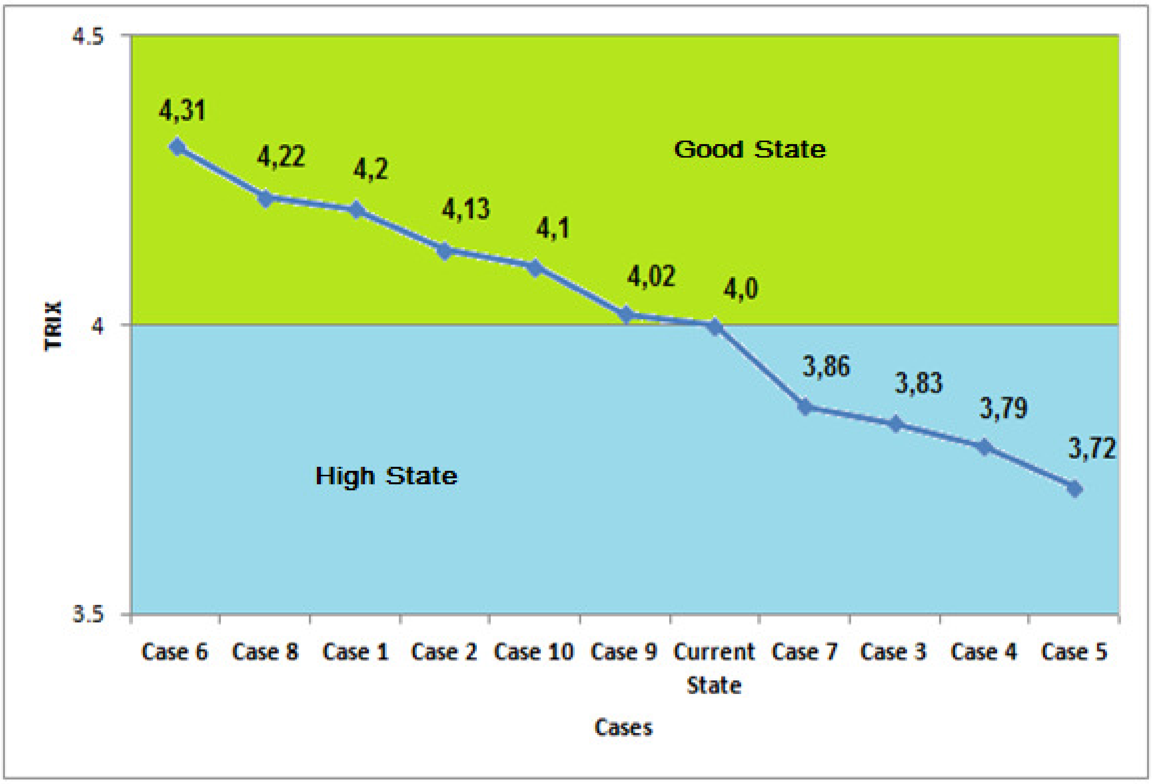

Figure 7.2 TRIX values for scenarios described in Table 7.1

7.2 Proposal for optimal watershed management

The first factor promoting eutrophication is nutrient enrichment. This explains why the main eutrophic areas are to be found primarily not far from the coast, mainly in areas receiving heavy nutrient loads. However, natural eutrophication can also occurred in upwelling areas.

An increase in the amount of nutrients in coastal areas leads to increased phytoplankton biomass during the spring bloom, but also to additional episodic blooms during summer and autumn. For Europe and adjacent seas, the primary production map computed in summer from satellite data shows the very heterogeneous distribution of highly productive areas along the European shores: while the whole shallow south and eastern North Sea, as well as a significant part of the Baltic Sea, and the Black Sea are highly productive, the Atlantic and Mediterranean shores exhibit only a strip of high production along the coast. In the Mediterranean area, only in the NA is noticed as eutrophic (Figure 7.3).

The current understanding of nutrient loading pressure and its consequences to the marine ecosystem, gaps in knowledge, and research needs are considered in relation to the conceptual framework for eutrophication. The research needs to fill gaps in understanding are grouped according to the Marine Strategy Framework Directive (MSFD, 2008/56/EC) as: (1) nutrient supply and enrichment; and (2) eutrophication symptoms (see Task Group 5 Report adopted in April, 2010).

It is important to be able to understand the mechanisms of eutrophication and to predict the alternative outcomes of ecosystem status with changes in nutrient pressure, as well as the uncertainty in the anticipated recovery pace and endpoints as a function of reductions in nutrient loading mandated by the MSFD (2008/56/EC) as the aim is Good Environmental Status (GES) of the European seas by 2020. It is important to set GES targets with safety margins for sustainable maintenance and fostering of marine ecosystems and services.

GES with regard to eutrophication (Task Group 5 Report adopted in April, 2010) has been achieved when the biological community remains well-balanced and retains all necessary functions in the absence of undesirable disturbance associated with eutrophication (e.g. excessive algal blooms, low dissolved oxygen, declines in sea grasses, kills of benthic organisms and/or fish) and/or where there are no nutrient-related impacts on sustainable use of ecosystem goods and services.

On an EU level, the importance of infrastructure improvements is highlighted, in order to provide long-term datasets and information to help avoid misdiagnosis of new events/changes, improve interpretation of trends, and facilitate development of management measures.

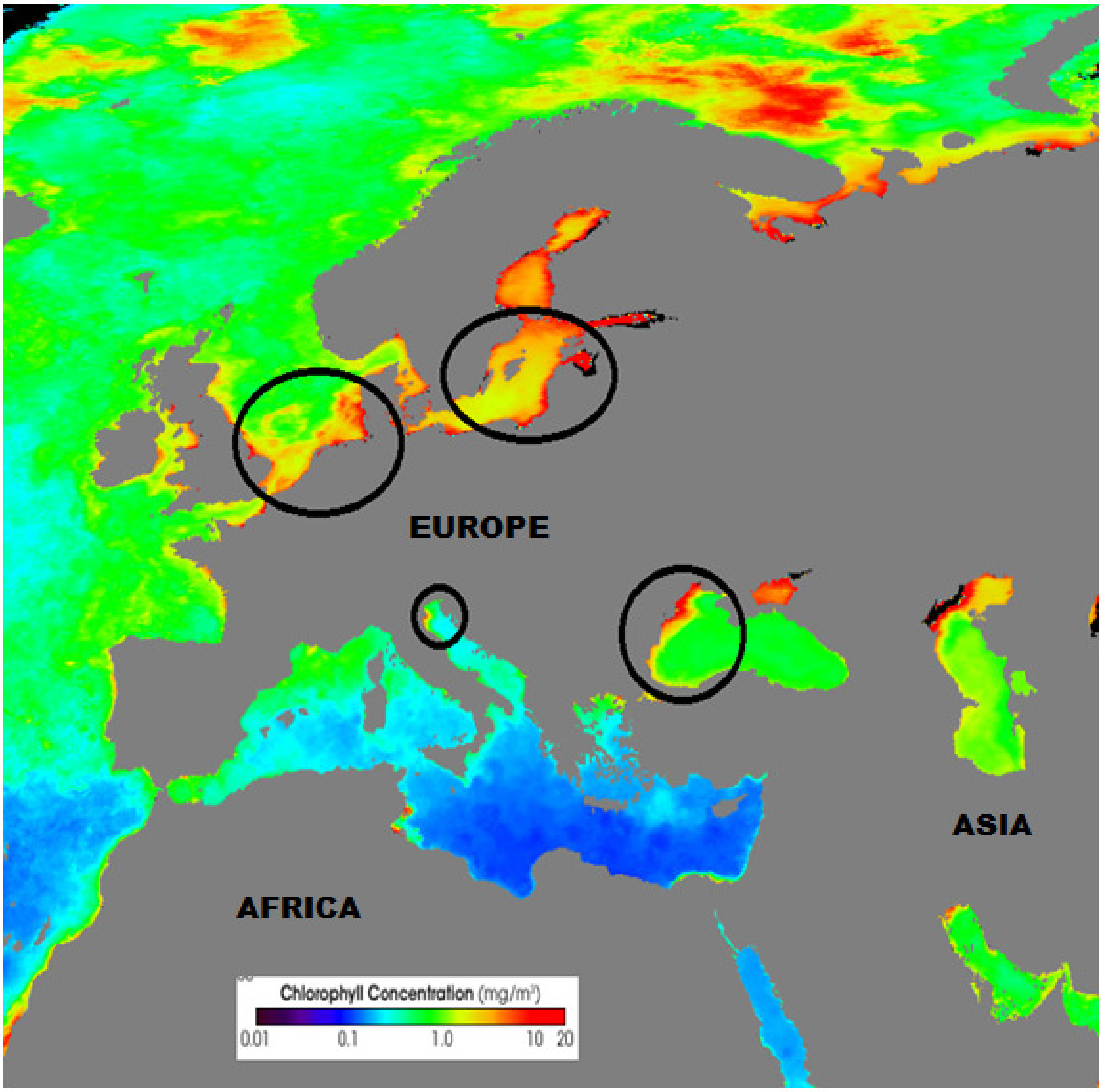

Figure 7.3 Seawifs composite image of chlorophyll *a* for the year 2006 (21. March-20. June). The four areas are evidenced. From upper left proceeding clockwise: coastal North Sea, Baltic Proper, North-western Black Sea shelf and northern Adriatic (http://earthobservatory.nasa.gov/IOTD/view.php?id=6735)

After short introduction on eutrophication, nutrient enrichment, MSFD and GES from the results of scenarios evaluation shown in Figure 7.2, as it was expected the optimal and proper watershed management leads to an improvement of the trophic state of the marine ecosystem. However, this improvement is not so strong but it is enough to change value of TRIX.

Scenario analysis gives us an insight on how various nutrient loads can assess to marine ecosystem state. Although TRIX value of 4.0 present boundary between **High** and **Good State** from Figure 7.2 can be seen how little it takes to change this TRIX value. Changing from secondary wastewater treatment (WWT) level which presents “Current State” to primary WWT level TRIX will amount 4.13, and if there is no treatment of wastewater TRIX will amount 4.2. In Sixth Scenario (see description in Table 7.1) TRIX is raised to 4.31, while Fifth Scenario gives the lowest TRIX (3.72). Reducing nutrient loads with adopting tertiary WWT level TRIX will fall from 4.0 to 3.83 (Third Scenario). For further decreasing of TRIX it is necessary to work in optimization of nutrients in agriculture, such is rational use of fertilizers, use of fertilizer with lower share of nutrients, etc.

When using TRIX We have to refer on research of Giovanardi and Vollenweider (2004) which illustrate some of the arising interpretative problems. One of such problems is how many samples are needed to obtain a reliable estimate of the difference between two contiguous TRIX means. In this research a monthly sampling frequency was used (12 samples per year). With this it is reached discrimination level between two measurement stations equal to 0.76 TRIX units, which is not indeed favourable. But, also it must be said that data for the top 10 m of the water column were averaged (sampling at 0.3, 5 and 10 m) which increase discrimination level to acceptable limit of 0.5.

Through above scenarios evaluation which is described in Section 7.1 of this Chapter following WWT level according to marine ecosystem state is determined (see Table 7.2). Proposal for WWT level given in Table 7.2 concerns only on effect of nutrient loads, and not on effect of organic (carbon) load to marine ecosystems. Because of this certain WWT (Secondary) is needed for removal of organic load.

The scenarios evaluation indicates that for the present state of marine ecosystem **Secondary** WWT is more than enough and **there is no need to invest in Tertiary** WWT (with respect to nutrient removal only).

Table 7.2 Proposal for proper WWT level

TRIX	State	WWT level
0-4	High	Secondary (Preliminary)
4-5	Good	Secondary (Primary)
5-6	Moderate	Secondary
6-8	Poor	Tertiary
8-10	Bad	Tertiary

Sufficient WWT level for **High State** of marine ecosystem is **Preliminary,** and for **Good State Primary** WWT level can be used. Taking into consideration that in this research effect of organic (carbon) load to marine ecosystem was not considered, it is supposed that **Secondary**

WWT level which has been adopted for all the watersheds in base scenario (describes the "Current State" of ecosystem, see Chapters 4 and 6) will be adequate for removal of this kind of load. **Secondary** WWT level which removes 30 % of P and 40 % of N is sufficient for **Moderate State.** For **Poor** and **Bad State** of marine ecosystem **Tertiary** WWT level which removes more than 90% of nutrients is recommended.

Considering the Urban Waste Water Treatment Directive (91/271/EEC) and research done in this Ph.D Thesis most parts of NA can be treated as less sensitive areas. **Preliminary** WWT will be suitable for open coastal areas (regards only to nutrient loads, not organic load). Exception can be put on sensitive areas with high nutrient load that are eutrophic or which in the near future may become eutrophic if protective actions are not taken (Po River delta, Venice lagoon and freshwater bodies inside the watershed).

Proposal for WWT level given in Table 2 is obtained from the model described in Chapter 6 which is not dynamic. Questions that arise from this issue are few, but the main is: **What will happen if Secondary level of treatment is kept all the time and how long the state of marine ecosystem will remain Moderate?** The answer is that **dynamic model has to be developed, or the series of current situations can be stated to imitate dynamic of the system.** In the case if nutrient loads from watershed (conditions in watershed) remain the same the state of marine ecosystem will **not change**.

Chapter 8.

Conclusions and further work

Summary of conclusions are given at the end of individual chapter's describing specific problem. Here are given only summaries of those conclusions as well as some specific remarks.

Influence of different pressures from surrounding watersheds on functioning of large scale marine ecosystem was explored in this Ph.D Thesis. Understanding the linkage between water quality in marine ecosystems and river watersheds is important in order to better assess marine ecosystems processes and to evaluate different management options in watersheds aimed at improving the marine ecosystem state.

The area of interest was placed on the northern Adriatic (NA, see Chapter 2) which is the most productive part of Adriatic Sea and as it was mentioned in previous chapters has many problems due to nutrient enrichments which could cause eutrophication, leading to mucilage and toxic algal blooms damaging tourism, to anoxia near the sea bottom causing the death of benthic fish and invertebrates.

To manage these problems several models have been done mainly using machine learning (ML) tools (see Chapter 3). These models give insight and help us to understanding of functioning of NA marine ecosystem.

"Standard" modelling as it is known is a key part of the ecosystem approach when laboratory experiments cannot be made or when historical data are limited. When modelling nutrient loads (e.g. Pressures) from watershed ML could not be used because of limited set of data, so in this case ArcView Generalized Watershed Loading Function (see Chapter 3) application was used for modelling. Data obtained from model (see Chapter 4) have then been used for further processing with ML tools where measured and modelled/simulated data were coupled to get model for defining the State of NA marine ecosystem. From model results (regarding specific loads in t/km^2) reduction efforts should be redirected mainly there where specific loads are high and not predominantly to Po River watershed like it is typically suggested.

Data analysis through which in this Ph.D Thesis was described the State of marine ecosystem is a key part of the ecosystem approach, both to isolate the effects of single pressures on single ecosystem compartments (e.g. laboratory data) or to understand more about emergent ecosystem dynamics (e.g. historical data). While most commonly, data analyses of other authors (see Chapter 5) were performed with only classical and just recently with advanced statistical approaches such as principal component analysis (PCA), in this Ph.D Thesis ML tools were used for this purpose. Although "standard techniques" provide very useful insights in the data, they are sometimes limited in terms of interpretability due to their black-box nature. On the other hand, a branch of ML methods and tools were proven to produce descriptive, e.g. transparent-box models, which generally allow much easier interpretation (Kompare, 1995, Kompare *et al.*, 2001, Atanasova *et al.*, 2008, Džeroski, 2009, Volf *et al.*, 2011). The advantage of these ML tools (Weka, Cubist) to build understandable and interpretable models which provide important insights on ecosystem functioning and his state are shown in this Ph.D Thesis (see Chapter 5).

ML was also used for linking two different domains: (1) the Pressures (nutrients) from surrounding watershed and (2) the State of NA marine ecosystem (see Chapter 6). For this purpose was used ML tool called Multi Target Stepwise Model Tree Induction. Results obtained by this model were then used to calculate the trophic index (TRIX).

Final task of this Ph.D Thesis was to propose optimal watershed management and to assess the proper wastewater treatment level according to marine ecosystem state which was determined through TRIX (Vollenweider *et al.*, 1998). For this task scenarios evaluation have been done (see Chapter 7).

Further research will be focused to: (1) bring better insight into nutrient dynamics, (2) more measurements of data in surrounding watersheds in order to create more extensive database which can then be processed with ML tools and (3) developing of a nutrient dynamic model. Also, for better understanding of functioning of the marine ecosystem and for more reliable predictions of future scenarios it would be of great help to link nutrient dynamic model from this research with hydrodynamic model.

8.1 Original contributions

Original contributions of this Ph.D Thesis are:

(1) Descriptive model for the dynamics of phytoplankton concentration, explains the dynamics of phytoplankton concentration in the NA for period 1972 to 2007. The model successfully identifies some of the triggers of changes in the phytoplankton dynamics by confirming the hypothesis made in previous researches.

(2) Predictive model for phytoplankton concentration 14 days in advance, gives accurate predictions of phytoplankton concentration for 14 days in advance correctly predicting the peak values of the phytoplankton concentration. As such, it can be efficiently used for water management purposes, e.g. as a phytoplankton concentration prediction supplement to watershed models that simulate nutrients loadings and concentrations in the aquatic environment as a consequence of human and natural activities in the watershed (e.g. land use, untreated wastewater, and so on).

(3) Model for TIN/PO_4 ratio, the model strongly confirmed the assumption that the mucilage events are connected with the changes of this ratio in the system, e.g. the model says that mucilage events coincide with the significant change of the TIN/PO_4 ratio.

(4) NA watershed nutrient loads, calculated nutrient loads (Pressures) from watershed for NA in period 1999 to 2007. Also, here are given contribution of nutrient loads by each sub-watershed and major sources of nutrients.

(5) Model linking the Pressures from watershed with the State of the marine ecosystem, model combining nutrient loads from watershed and marine data calculates TRIX through which the State of marine ecosystem is defined.

(6) Controlling and managing the activities in the watersheds. One of these activities is to determinate the proper wastewater treatment level according to marine ecosystem state.

(7) Defining proper wastewater treatment level according to marine ecosystem state.

(8) Positive and negative impacts of none treated, partially treated and treated wastewater to the marine ecosystem.

(9) Use of ML (new techniques) in solving the problems above.

8.2 The application of research results

Results of this Ph.D Thesis research can/will be used in: (1) assessing the proper wastewater treatment level and agricultural fertilization intensity in watersheds that lead to coastal waters, (2) modelling natural processes, e.g. marine ecosystems, (3) modelling the interaction between marine ecosystems and processes in watershed, (4) testing and reconstruction of existing models, (5) understanding, both negative and positive impacts of WWTP to marine ecosystems and (6) sustainable management of watersheds areas.

References

Adeka, J.E., Strobl, R.O., Becht, R., 2008, An Environmental system analysis of lake Elementaita, Kenya with reference to water quality, Proceedings of Taal 2007, The 12th world lake conference, 1365-1372.

Aha, D.W., Kibler, D., Albert, M.K., 1991, Instance-based learning algorithms, *Machine Learning*, 6, 1, 37-66.

Appice, A. and Džeroski, S., 2007, Inducing Multi-Target Model Trees in a Stepwise Fashion, Proceedings of the fifteenth Italian symposium on advanced database systems, editors: Michelangelo Ceci, Donato Malerba, Letizia Tanca. Torre Canne di Fasano (BR), Italy, 17-20 June, 2007.

Arhounditsis, G., Giourga, C., Loumou, A., Koulouri, M., 2002, Quantitative Assessment of agricultural runoff and soil erosion using mathematical modelling: Application in the Mediterranean Region, *Environmental Management*, 30, 434-453.

Arnold, J., Binger, R., Burrows, D., DeBarry, P.A., Engman, E.T., Evans, B.M., Evans, T.A., Garbrecht, J., Garcia, L., Johnson, L.E., Jorgeson, J.D., Krysanova, V., Leavesley, G., Maidment, D.R., Nelson, E.J., Ogden, F.L., Olivera, F., Quimpo, R.G., Seybert, T.A., Sloan, W.T., Theurer, F., 2000, GIS Modules and Distributed Models of the Watershed, American Society of Engineers, Reston, VA, 120 pp.

Artegiani, A., Bregant, D., Paschini, E., Pinardi, N., Raicich, F., Russo, A., 1997, The Adriatic Sea general circulation, Part II: baroclinic circulation structure, *Journal of Physical Oceanography*, 27, 1515-1532.

Artioli, Y., Bendoricchio, G., Palmeri, L., 2005, Defining and modelling the coastal zone affected by the Po river (Italy), *Ecological Modelling*, 184, 55-68.

Artioli, Y., Friedrich, J., Gilbert, A.J., Quatters-Gollop, A., Mee, L.D., Vermaat, J.E., Wulff, F., Humborg, C., Palmeri, L., Pollehne, F., 2008, Nutrient budgets for European seas: a measure of the effectiveness of nutrient reduction policies, *Marine Pollution Bulletin*, 56, 1609-1617.

Atanasova, N., Gal, G., Kompare, B., 2008, Modelling Dinoflagellate Dynamics in Lake Kinneret, *Mitteilungen Internationale Vereiningung fuer Theoretische und Angewandte Limnologie*, 31, 100-104.

Autorita di bacino del fiume Po, 2008, Studio per L˝approfondimento delle variazioni dei carici di azoto e fosforo transitati nella stazione di Pontelagoscuro, per L˝analisi di processi rilevanti ai fini della compresione della composizione e della transformazione dei carichi, Responsabile del progeto: prof. Viaroli, P.

Azam, F., Fonda Umani, S., Funari E., 1999, Significance of bacteria in the mucilage phenomenon in the northern Adriatic Sea, *Annale dell Istitutio superior di sanita,* 35, 411-419.

Barausse, A., Duci, A., Carlotta Mazzoldi, C., Artioli, Y., Luca Palmeri, L., 2009, Trophic network model of the Northern Adriatic Sea: Analysis of an exploited and eutrophic ecosystem, *Estuarine, Coastal and Shelf Science*, 83, 577-590.

Barnes, R.S.K. and Hughes, R.N., 1999, An Introduction to Marine Ecology, Blackwell Publishing.

Bernardi Aubry, F., Berton, A., Bastianini, M., Socal, G., Acri, F., 2004, Phytoplankton succession in a coastal area of the NW Adriatic, over a 10-year sampling period (1990-1999), *Continental Shelf Research*, 24, 97-115.

Blockeel, H., Raedt, L.D., Ramon, J., 1998, Top-down induction of clustering trees, In J. Shavlik, editor, Proceedings of the 15th International Conference on Machine Learning, 55-63.

Caddy, J.F., 2000, Marine catchment basin effects versus impacts of fisheries on semi-enclosed seas, *ICES Journal of Marine Science*, 57, 628-640.

Cataletto, B., Feoli, E., Fonda Umani, S., Sun. C.Y., 1995, 11 years of time-series analysis on the net-zooplankton community in the Gulf of Trieste, *ICES Journal of Marine Sciences*, 52, 669-678.

Ceci, M., Appice, A., Malerba D., 2003, Comparing simplification methods for model trees with regression and splitting nodes. In N.Zong, Z. Ras, S. Tsumoto, and E. Suzuki, editors, Foundations of Intelligent Systems, 14th International Symposium, volume 2871 of LNAI, pages 49-56. Springer-Verlag.

Cestnik, B. and Bratko, I., 1991, On estimating probabilities in tree pruning. In Proceedings of the 5th European Working Session on Learning, 151-163.

Cloern, J.E., 2001, Our evolving conceptual model of the coastal eutrophication problem, *Marine Ecology Progress Series*, 210, 223-253.

Conversi, A., Peluso, T., Fonda Umani. S., 2009, The Gulf of Trieste: A changing ecosystem, *Journal of Geophysical Research*, 114, C03S90, doi: 10.1029/2008JC004763.

Cozzi, S. and Giani, M. 2011, River water and nutrient discharges in the Northern Adriatic Sea: Current importance and long term changes, *Continental Shelf Research*, doi:10.1016/j.csr.2011.08.010.

Cozzi, S., Ivančić, I., Catalano, G., Djakovac, T., Degobbis, D., 2004, Dynamics of the oceanographic properties during mucilage appearance in the northern Adriatic Sea: analysis of the 1977 event in comparison to earlier events, *Journal of Marine Systems*, 50, 223-241.

Cucchiari, E., Guerrini, F., Penna, A., Totti, C., Pistocchi, R., 2008, Effect of salinity, emperature, organic and inorganic nutrients on growth of cultured Fibrocapsa japonica (Raphidophyceae) from the northern Adriatic Sea, *Harmful Algae*, 7, 405-414.

De Lazzari, A., Berto, D., Cassin, D., Boldrin, A., Giani, M., 2008, Influence of winds and oceanographic conditions on the mucilage aggregation in the Northern Adriatic Sea in 2003-2006, *Marine Ecology*, 29, 469-482.

de Wit, M. and Bendoricchio. G., 2001, Nutrient fluxes in the Po basin, *Science of the Total Environment*, 273, 147-161.

Degobbis, D., 1988, Biogeokemijski ciklus hranjivih soli u sjevernom Jadranu, Doktorski rad, Sveučilište u Zagrebu, Institut Ruđer Bošković, Rovinj.

Degobbis, D., Fonda-Umani, S., Franco, P., Malej, A., Precali, R., Smodlaka, N., 1995, Changes in the northern Adriatic ecosystem and appearance of hypertrophic gelatinous aggregates, *Science of the Total Environment,* 165, 43-58.

Degobbis, D. and Gilmartin, M., 1990, Nitrogen phosphorus and silicon budgets for the northern Adriatic Sea, *Oceanologica Acta*, 13, 31-45.

Degobbis, D., Malej, A., Fonda Umani, S., 1999, The mucilage phenomenon in the northern Adriatic Sea-A critical review of the present scientific hypotheses, *Annali dell'Istituto Superiore di Sanità*, 35, 373-383.

Degobbis, D., Precali, R., Ferrari, C.R., Djakovac, T., Rinaldi, A., Ivančić, I., et al., 2005, Changes in nutrient concentrations and ratios during mucilage events in the period 1999-2002, *Science of the Total Environment*, 353, 103-114.

Degobbis, D., Precali, R., Ivančić, I., Smodlaka, N., Fuks, D. and Kveder. S., 2000, Long-term changes in the northern Adriatic ecosystem related to anthropogenic eutrophication, *International Journal of Environment and Pollution*, 13, 495-533.

Deliman, P.N., R.H. Glick, and C.E. Ruiz, 1999, Review of Watershed Water Quality Models. U.S. Army Corps of Engineers, Tech. Rep. W-99-1, 26 pp.

Deserti, M., Cacciamani, C., Chiggiato, J., Rinaldi, A., Ferrari, C.R., 2005, Relationship between northern Adriatic Sea mucilage events and climate variability, *Science of the Total Environment*, 353, 82-88.

Draper, N. R. and Smith, H., 1982, Applied regression analysis. John Wiley & Sons.

Dzeroski, S., 2009, Machine learning applications in habitat suitability modeling, in Artificial Intelligence Methods in the Environmental Sciences, S.E. Haupt, A. Pasini, and C. Marzban, Editors, Springer: Berlin. p. 397-412.

Džeroski, S., Kobler, A., Gjorgjioski, V., Panov, P., 2006, Using decision trees to predict forest stand height and canopy cover from landsat and lidar data. *In Enviromental Info*.

EEA, 2001a, Eutrophication in Europe’s coastal waters, EEA report 7, EEA.

EEA, 2001b, Marine and coastal environment-Annual topic update 2000, Topic report 11, EEA, Copenhagen (DK).

EEA, 2003a, Europe's water: an indicator-based assessment, Topic report 1, EEA, Copenhagen (DK).

EEA, 2003b, Hazardous substances in the European marine environment-trends in metals and persistent organic pollutants, Topic report 2, EEA, Copenhagen (DK).

EEA, 2006, The changing faces of Europe's coastal areas. EEA report 6, EEA, Copenhagen (DK).

Evans, B.M. and Corradini, K.J., 2007, AVGWLF, Version 7.2, A guide to creating software-compatible data sets, Pennsylvania State Institutes of Energy and the Environment, The Pennsylvania State University, University Park, PA 16802.

Evans, B.M., Lehning, D.W., Corradini, K.J., 2008, AVGWLF, Version 7.1, Users Guide, Pennsylvania State Institutes of Energy and the Environment, The Pennsylvania State University, University Park, PA 16802.

Evans, B.M., Lehning, D.W., Corradini, K.J., Petersen, G.W., Nizeyimana, E., Hamlett, J.M., Robillard, P.D., Day, R.L, 2002, A Comprehensive GIS-Based Modeling approach for Predicting Nutrient Loads in Watersheds, *Journal of Spatial Hydrology*, 2, 1-18.

Fonda Umani, S., Beran, A., Parlato, S., Virgilio, D., Zollet, T., De Olazabal, A., Lazzarini, B., Cabrini. M., 2004, Noctiluca scintillans Macartney in the Northern Adriatic Sea: Long-term dynamics, relationships with temperature and eutrophication, and role in the food web, *Journal of Plankton Research*, 26, 545-561.

Gabrielsen, P. and Bosch, P., 2003, Environmental Indicators: Typology and Use in Reporting, EEA internal working paper, 20 pp.

Giovanardi, F. and Vollenweider, R.A., 2004, Trophic conditions of marine coastal waters: experience in applying the Trophic Index TRIX to two areas of the Adriatic and Tyrrhenian seas, *Journal of Limnology*, 63, 199-218.

Tchobanoglous, G., Franklin L., Burton, H., Stensel, D., 2003, Wastewater Engineering, Treatment and Reuse, Fourth Edition, Metcalf and Eddy, Inc., McGraw Hill, New York.

Giani, M., Savelli, F., Boldrin, A., 2003, Temporal variability of particulate organic carbon, nitrogen and phosphorus in the Northern Adriatic Sea, *Hydrobiologia*, 494, 319-325.

Glen, G., 2010, The Impact of Climate Change on European Lakes, Springer, 1st Edition, 507 p.

Grilli, F., Marini, M., Degobbis, D., Ferrari, C. R., Fornasiero, P., Russo, A., Gismondi, M., Djakovac, T., Precali, R., Simonetti, R., 2005, Circulation and horizontal fluxes in the northern Adriatic Sea in the period June 1999-July 2002. Part II: Nutrients transport, *Science of the Total Environment*, 353, 115-125.

Haith, D.A. and Shoemaker, L.L., 1987, Generalized watershed loading functions for stream flow nutrients, *Water Resources Bulletin*, 23, 471-478.

Haith, D.A. and Tubbs, L.J., 1981, Watershed loading functions for Nonpoint sources. *ASCE Journal of Environmental Engineering*, 107, 121-137.

Harding Jr., L.W., Degobbis, D., Precali. R., 1999, Production and fate of phytoplankton: Annual cycles and interannual variability. Ecosystems at the land-sea margin: Drainage basin to coastal sea, *Coastal and Estuarine Studies,* 55, 131-172.

Herndl, G.J., 1992, Marine snow in the northern Adriatic Sea: possible causes and consequences for a shallow ecosystem, *Marine Microbial Food Webs*, 6, 149-172.

Hopkins, T.S., Artegiani, A., Kinder, C., Pariante. R., 1999, A discussion of the northern Adriatic circulation and flushing as determined from the ELNA hydrography. The Adriatic Sea, Ecosystem Research Report 32, ed. T.S. Hopkins, *et al.*, 85-106.

Ignatiades, L., 2008, Scaling the trophic status of the Aegean Sea, eastern Mediterranean, *Journal of Sea Research*, 54, 51-57.

Jennings, E., Allott, N., Pierson, D.C., Schneiderman, E.M., Lenihan, D., Samuelson, P., Taylor, D., 2009, Impacts of climate change on phosphorus loading from a grassland catchment: Implications for future management, *Water Research*, 43, 4316-4326.

Justic, D., Legovic, T., Rottini-Sandrini, L. 1987. Trends in oxygen content 1911-1984 and occurrence of benthic mortality in the northern Adriatic Sea, *Estuarine Coastal and Shelf Science*, 25, 435-445.

Kamburska, L. and Fonda Umani. S., 2006, Long-term copepod dynamics in the Gulf of Trieste (Northern Adriatic Sea): Recent changes and trends, *Marine Ecology Progress Series*, 31, 195-203.

Karalic, A., 1992, Linear regression in regression tree leaves, In Proceedings of International School for Synthesis of Expert Knowledge, 151-163, Slovenia.

Kennish M.J. and Paerl, H.W., 2010, Coastal Lagoons-Critical Habitats of Environmental Change, CRC Press, 483-511.

Kompare, B., 1995, The Use of Artificial Intelligence in Ecological Modelling, Ph.D Thesis, FGG, Ljubljana, Royal Danish School of Pharmacy, Copenhagen, Ljubljana.

Kompare, B., Todorovski, L., Ďzerovski, S., 2001, Modelling and prediction of phytoplankton growth with equation discovery: case study-LakeGlumsø, Denmark, *Mitteilungen Internationale Vereiningung fuer Theoretische und Angewandte Limnologie*, 27, 3626-3631.

Malerba, D., Esposito, F., Ceci, M., Appice, A., 2004, Top down induction of model trees with regression and splitting nodes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26, 612-625.

Mehta, M., Agrawal, R., Rissanen, J., 1996, SLIQ: A fast scalable classifier for data mining. In Proceedings of the 5th International Conference on Extending Database Technology, pages 18-32.

Moore, I., Turner, A.K., Wilson, J.P., Jenson, S.K., Band, L.E., 1991, GIS and Land-Surface-Subsurface Process Modeling. In: Proc. 1st International Conference, Workshop on Integrated Geographic Information Systems and Environmental Modeling, Boulder, CO.

Mozetič, P., Fonda Umani, S., Cataletto, B., Malej, A., 1998, Seasonal and inter-annual plankton variability in the Gulf of Trieste (northern Adriatic), *ICES Journal of Marine Sciences*, 55, 711-722.

Mozetič, P., Solidoro, C., Cossarini, G., Socal, G., Precali, R., Francé, J., Bianchi, F., De Vittor, C., Smodlaka, N., Fonda Umani, S., 2009, Recent Trends Towards Oligotrophication of the Northern Adriatic: Evidence from Chlorophyll *a* Time Series, *Estuaries and Coasts*, 33, 362-375.

Myklestad, SM., 1995, Release of extracellular products by phytoplankton with special emphasis on polysaccharides, *Science of the Total Environment*, 165, 155-164.

Navarro, E., Caputo, L., Marce, R., Carol, J.,Benejam, L., Garcıa-Berthou, E., Armengol, J., 2009, Ecological classification of a set of Mediterranean reservoirs applying the EU Water Framework Directive: A reasonable compromise between science and management, *Lake and Reservoir Management*, 25, 364-376.

Niblett, T. and Bratko, I., 1986, Learning decision rules in noisy domains. In M. Bramer, editor, Research and Development in Expert Systems, volume 3, pages 25-34. Cambridge University Press.

Nixon, S.W. and Fulweiler, R.W., 2009, Nutrient pollution, eutrophication and the degradation of coastal marine ecosystems, Global Loss of Coastal Habitats, Rates, Causes and Consequenses, ed. Duarte, C.M., Fundacion BBVA.

Ogrinc N. and Faganeli J., 2006, Phosphorus regeneration and burial in near-shore marine sediments (the Gulf of Trieste, northern Adriatic Sea), *Estuarine, Coastal and Shelf Science*, 67, 579-588.

Organization for Economic Co-Operation and Development (OECD), 1982, Eutrophication of Waters: Monitoring, Assessment and Control, Paris.

Pagnotta, R., Caggiati, G., Piazza, D., Ferrari, F., 1995, Il controllo della qualita` delle acque superficiali del bacino padano, *Inquinamento*, 4, 8-14.

Palmeri, L., Bendoricchio, G., Artioli, Y., 2005, Modelling nutrient missions from river systems and loads to the coastal zone: Po River case study, Italy, *Ecological Modelling*, 184, 37-53.

Parsons, T.R., Maita, Y., Lalli, C.M., 1985, A Manual of Chemical and Biological Methods for Seawater Analysis, Pergammon Press, Oxford, p. 173.

Pečar, O., Precali, R., Giani, M., Blažina, M., Smodlaka, N., 2004, Northern Adriatic mesocosm experiment Rovinj 2003: Primary production and total organic carbon, *Periodicum Biologorum*, 106, 23-29.

Pettine, M., Casentini, B., Fazi, S., Giovanardi, F., Pagnotta, R., 2007, A revisitation of TRIX for trophic status assessment in the light of the European Water Framework Directive: Application to Italian coastal waters, *Marine Pollution Bulletin*, 54, 1413-1426.

Pojed, I. and Kveder, S., 1977, Investigation of nutrient limitation of phytoplankton production in the northern Adriatic by enrichment experiments, *Thallasso Jugoslavia*, 13, 13-24.

Precali, R., Giani, M., Marini, M., Grilli, F., Pečar, O., Paschini, E., 2005, Mucilage events in the northern Adriatic in the period 1999-2002, *Science of the Total Environment*, 353, 10-23.

Provini, A., Crosa, G., Marchetti, R., 1992, Nutrient export from the Po and Adige river basins over the last 20 years. In: Vollenweider RA, Marchetti R, Viviani R, editors. Int conf marine coastal eutrophication, 21-24 March 1990, Bologna, Italy, *Science of the Total Enviroment*, 291-313.

Pugnetti, A., Armeni, M., Camatti, E., Crevatin, E., Dell'Anno, A., Del Negro, P., Milandri, A., Socal, G., Umani, S. F., Danovaro, R., 2005, Imbalance between phytoplankton production and bacterial carbon demand in relation to mucilage formation in the Northern Adriatic Sea, *Science of the Total Environment,* 353, 162-177.

Quinlan, J. R., 1986, Induction of Decision Trees, *Machine Learning*, 1, 1, 81-106.

Quinlan, J.R., 1992, Learning with continuous classes, Proc. AI'92, 5th Australian Joint Conference on Artificial Intelligence, Adams & Sterling (eds.), World Scientific, pp 343-348, Singapore.

Quinlan, J.R., 1993, Combining instance-based and model-based learning, In proceedings ML'93, ed. Morgan Kaufmann.

Raicich, F., 1994, Note on flow rates of the Adriatic rivers. Technical Report. CNR Istituto Talassografico Sperimentale, Trieste. RF 02/94, 8 pp.

Raicich, F., 1996, On the fresh water balance of the Adriatic Sea, *Journal of Marine Systems*, 9, 305-319.

Raymont, J.E.G., 1980, Plankton and Productivity in the Oceans. Vol. 1. Phytoplankton. Pergamon Press, Oxford, 489.

Redfield A.C., 1934, On the proportions of organic derivations in seawater and their relation to the composition of plankton. James Johnstone memorial volume. (ed. R.J. Daniel). University Press of Liverpool, 177-192.

Redfield, A.C., Ketchum, B.H., Richards, F.A., 1963, The influence of organisms on the composition of seawater. In: The Sea, edited by M.N. Hill, New York, *Interscience Publishing*, 2, 27-77.

Revelante, N. and Gilmartin, M., 1983, The phytoplankton of the Adriatic Sea: community structure and characteristics, *Thallasso Jugoslavia*, 19, 303-318.

Russo, A., Maccaferri, S., Djakovac, T., Precali, R., Degobbis, D., Deserti, M., Paschini, E., Lyons, D. M., 2005, Meteorological and oceanographic conditions in the northern Adriatic Sea during the period June 1999-July 2002: Influence on the mucilage phenomenon, *Science of the Total Environment*, 353, 24-38.

Salas, F., Teixeira, H., Marcos, C., Marques, J.C., Perez-Ruzafa, A., 2008, Applicability of the trophic index TRIX in two transitional ecosystems: the Mar Menor lagoon (Spain) and the Mondego estuary (Portugal), *International Council for the Exploration of the Sea*, Published by Oxford Journals, 1442-1448.

Santojanni, A., Arneri, E., Bernardini, V., Cingolani, N., Di Marco, M., Russo, A., 2006, Effects of environmental variables on recruitment of anchovy in the Adriatic Sea. *Climate Research*, 31, 181-193.

Shun Dar Lin, 2007, Water and Wastewater Calculation Manual, Second edition, McGraw Hill, New York.

Singh, V.P. and Frevert, D.K., 2006, Watershed models, Taylor and Francis Group.

Smith, G.H.S., Best, J.L., Bristow, C.S., Petts, G.E., 2006, Braided rivers: process, deposits, ecology and management, International Association of Sedimentologists.

Solidoro, C., Bandelj, V., Barbieri, P., Cossarini, G., Fonda Umani. S., 2007, Understanding dynamic of biogeochemical properties in the northern Adriatic Sea by using self-organizing maps and k-means clustering. *Journal of Geophysical Research*, 112, C07S90, doi: 10.1029/2006JC003553.

Solidoro, C., Bastianini, M., Bandelj, V., Codermatz, R., Cossarini, G., Melaku Canu, D., Ravagnan, E., Salon, S., Trevisani. S., 2009, Current state, scales of variability and decadal trends of biogeochemical properties in the Northern Adriatic Sea, *Journal of Geophysical Research*, 114, C07S91, doi: 10.1029/2008JC004838.

Sournia, A., 1973, La production primaire planctonique en Méditerranée. *Bulletine Ėtude Communicacione Mediterane,* 5, 128.

Spillman, C.M., Imberger, J., Hamilton, D.P., Hipsey, M.R., Romero J.R., 2007, Modelling the effects of Po River discharge, internal nutrient cycling and hydrodynamics on biogeochemistry of the Northern Adriatic Sea, *Journal of Marine Systems*, 68, 167-200.

Stachowitsch, M., Fanuko, N., Richter, M., 1990, Mucus aggregates in the northern Adriatic sea: an overview of types and occurrences, *Marine Ecology*, 11, 327-350.

Stanfill, C. and Waltz, D., 1986, Toward memory-based reasoning, Communications of the *ACM* 29, 12, 1213-1228.

Strickland, J.D.H. and Parsons, T.R., 1972, A practical handbook of seawater analysis, *Bulletin of Fish Research*, Bd. Canada, Ottawa, 167, 310.

Supić, N., Đakovac, T., Krajcar, V., Kuzmić, M., Precali, R., 2006, Effects of excessive Po River discharges in the northern Adriatic, *Fresenius Environmental Bulletin,* 15, 3, 193-199.

Supić, N., Orlić, M., Degobbis, D., 2000, The Istrian countercurrent and its year to year variability, *Estuarine, Coastal and Shelf Science*, 51, 385-397.

Tedesco, L., Socal, G., Bianchi, F., Acri, F., Veneri, D., Vichi, M., 2007, NW Adriatic Sea biogeochemical variability in the last 20 years (1986-2005), *Biogeosciences*, 4, 673-687.

Thornthwaite, C.W., 1948, An approach toward a rational classification of climate, *Geographical Review*, 38, 55-94.

Thornton, D., Santillo, D., Thake, B., 1999, Prediction of Sporadic Mucilaginous Algal Blooms in the Northern Adriatic Sea, *Marine Pollution Bulletin,* 38, 10, 891-898.

U.S. Environmental Protection Agency (EPA), 1999, Protocols for Developing Nutrient TMDLs, EPA 841-B-99-007, Office of Water (4503 F), Washington, D.C.

United Nations Educational, Scientific and Cultural Organization (UNESCO), 1986, Progress on Oceanographic Tables and Standards 1983–1986: Work and recommendations of the Unesco/SCOR/ICES/IAPSO Joint Panel., UNESCO Technical Papers in Marine Science, No. 50, p. 59.

United Nations Environment Programme (UNEP), 1995, Assesment of the state of eutrophication in Mediterranean sea, First draft, Mediterranean action plan, Athens, 3-8 April 1995.

Utermöhl, H., 1958, Zur Verfollkommnung der quantitative Phytoplankton-Methodik, *Mitteilungen Internationale Vereiningung fuer Theoretische und Angewandte Limnologie*, 17, 47-71.

Volf, G., Atanasova, N., Kompare, B., Precali, R., Ožanić, N., 2011, Descriptive and prediction models of phytoplankton in the northern Adriatic, *Ecological Modelling*, 222, 2502-2511.

Vollenweider, R.A. and Kerekes, J.J., 1982, Eutrophication of Waters, Monitoring, Assessment and Control, O.E.C.D., Paris, 154 pp.

Vollenweider, R.A. and Rinaldi, A., Eds., 1995, Marine Mucillages, with special reference to mucilage events in the northern Adriatic Sea, the Tyrrhenian Sea and the North Sea, *Science of the Total Environment*, 165, 1-230.

Vollenweider, R.A., Giovanardi, F., Montanari, G., Rinaldi, A., 1998, Characterization of the trophic conditions of marine coastal waters, with special reference to the NW Adriatic Sea: proposal for a trophic scale, turbidity and generalized water quality index, *Environmetrics*, 9: 329-357.

Vollenweider, R.A., Rinaldi, A., Montanari, G., 1992, Eutrophication, structure and dynamics of a marine coastal system: results of ten-year monitoring the Emilia-Romagna coast (Northwest Adriatic Sea). In: R.A. Vollenweider, R. Marchetti, R. Viviani (Eds.), Marine Coastal Eutrophication, Elsevier, Amsterdam, 63-106.

Wang, Y. and Witten, I.H., 1997, Induction of model trees for predicting continuous classes, Proceedings of the European Conference on Machine Learning, University of Economics, Faculty of Informatics and Statistics, Prague.

Wilson, J.P., 1996, GIS Based Land Surface/Subsurface Modeling: New Potential for New Models? In: Proc. 3rd International Conference/Workshop on Integrated Geographic Information Systems and Environmental Modeling, Santa Fe, NM.

Witten, I.H. and Frank, E., 2000, Data mining-Pratical machine learning Tools and Techniques with java implementations, Academic Press, USA.

Zavatarelli, M., Baretta, J.W., Baretta-Bekker, J.G., Pinardi, N., 2000, The dynamics of the Adriatic Sea ecosystem, An idealized model study, *Deep-Sea Research*, 47, 937-970.

Zavatarelli, M., Raicich, F., Bregant, D., Russo, A., Artegiani. A., 1998, Climatological biogeochemical charatcteristics of the Adriatic Sea, *Journal of Marine Systems*, 18, 227-263.