Brandenburg University of Technology Cottbus
Faculty of Environmental Sciences and
Process Engineering

Investigating Ecological Indicators of Freshwater Ecosystems Using Signal Analysis Methods

A Doctoral Dissertation, Approved by the Faculty of Environmental Sciences and Process Engineering of the Brandenburg University of Technology Cottbus and Submitted in Fulfilment of the Requirements for the Award of the Degree of Ph.D

Submitted by

Jean Duclos Alegue Feugo, M.Sc.

From Bamenda, Cameroon

Supervision

Univ. - Prof. Dr. rer. nat. habil. Albrecht Gnauck, BTU Cottbus
Priv. - Doz. Dr. - Ing. habil. Thomas Rauschenbach, TU Ilmenau

Date of disputation: 31th of March 2008

Brandenburg University of Technology, Cottbus
Faculty of Environmental Science and
Process Engineering

Investigating Ecological Indicators of Freshwater Ecosystems Using Signal Analysis Methods

A dissertation submitted in partial fulfilment of the requirements for the Degree of Ph.D. in Environmental and Resource Management under the Department of Ecosystems and Environmental Informatics.

by

Jean Duclos Alegue Feugo, M.Sc.

Approved by supervisors

Univ. - Prof. Dr. rer. nat. habil. Albrecht Gnauck, BTU Cottbus

Priv. - Doz. Dr. - Ing. habil. Thomas Rauschenbach, TU Ilmenau

Declaration

I, Jean Duclos Alegue Feugo, declare that this dissertation is solely the result of my effort and has never been submitted in another institution as a requirement for any certificate. The dissertation was an independent investigation under the supervision of Prof. Dr. habil. Albrecht Gnauck (main supervisor) and Priv. -Doz. Dr. habil. Thomas Rauschenbach (second supervisor). All the sources used for the writing of this dissertation have been adequately referenced.

Jean Duclos Alegue Feugo (M.Sc.)……

Dept. of Ecosystem and Environmental Informatics, BTU

Dedication

I dedicate this work to my Mum and late Dad who always taught me that excellence is granted to a man only as the reward of hard work.

Acknowledgements

Without the comfort, wisdom, strength and joy from the Lord Jesus, this dissertation would not have been possible. The encouragement and financial support from my entire family enabled me to take this challenge to the end. My professor, Albrecht Gnauck, spent quality time in training me and pushing me to work harder and harder so as to develop the necessary skills required for a PhD. I also express my gratitude to my second supervisor, PD Dr. Rauschenbach who was very busy, but took quality time to advice me on how to improve my work. My colleagues Shaffi Noor, Shouke wei, Ralf Heinrich, Bernhard Luther and Hartmut Nemitz all contributed in creating a very nice working environment.

Abstract

Human life depends on water of good quality. Freshwater ecosystems are being degraded as a result of anthropogenic activities. Managing freshwater resources require a good understanding of the dynamics of the processes affecting water quality and the interrelationship existing between them. Signal analysis methods are used to extract information from water quality time series. The information obtained from such analysis is required for developing dynamic models as an aid in investigating the outcome of different management scenarios. They are equally useful in determining the appropriate sampling frequency required for monitoring ecological indicators. Modern methods developed in other field such as mathematical statistics need to be investigated in the fields of water quality management so as to enhance the knowledge of the functioning and the structure of water bodies.

Classical methods consisting of time domain and frequency domain methods in combination with a modern method, wavelet analysis, were used to extract information from water quality indicators from the River Havel in the State of Brandenburg in Germany. The indicators were dissolved oxygen, chlorophyll-a and water temperature which are respectively chemical, biological and physical indicators. The time domain methods revealed the behavior of the signal across time as well as the relationship between them. The frequency domain methods proved quite inadequate because the signals from water quality are non-stationary with changing variance across time. The wavelet methods were quite good in unraveling the behavior of these signals at different time-scales.

This analysis revealed that the high frequency changes have no significant effect on the long term dynamics of water quality signals. Given that only the low frequency components influence the long term behavior of these signals, it was found that it makes more sense to sample most of the signals at a weekly or two weekly intervals so as to avoid noisy or redundant information. In

addition, the time-scale decomposition allows for noisy or redundant information that blurs the long term dynamics thereby negatively affecting the quality of models to be detected and kicked out. Moreover, it was also found that the use of a time delay in a dynamic model should be based on the delay from the time-scale that influence the long term dynamics the most in the freshwater body by the help of the wavelet cross-correlation rather than the classical cross-correlation. Finally, the same indicator from different water bodies of the same watershed portrayed different time and frequency dependent behavior implying that each freshwater body needs to be uniquely investigated. Applying the same management techniques to different water bodies without prior investigation will not be sound.

Freshwater ecosystems are more and more threatened by phenomena such as global warming, pharmaceuticals in water bodies, invasive species requiring more investigation on the applicability of tools developed in other fields of science in the field of water resource management. Enhanced knowledge will improve the existing knowledge on the structure and functioning of freshwater bodies, the quality of models developed as management decision aid and the quality of the data used for decision making.

Table of Contents

Abbreviations

ACF	Auto Correlation Function
AIC	Akaike Information Criterion
AR	Auto Regressive
ARIMA	Auto Regressive Integrated Moving Average
ARMA	Auto Regressive Moving Average
BJ	Box-Jenkins
CWT	Continuous Wavelet Transform
DDT	Dichlorodiphenyltrichloroethane
DNA	Deoxyribonucleic Acid
DWT	Discrete Wavelet Transform
LAWA	Länderarbeitsgemeinschaft Wasser
MA	Moving Average
MODWT	Maximal Overlap Discrete Wavelet Transform
OECD	Organisation for Economic Cooperation and Development
PACF	Partial Auto Correlation Function
PCB	Polychlorinated Biphenyl
SARIMA	Seasonal Auto Regressive Integrated Moving Average
USEPA	United States Environmental Protection Agency

List of Symbols

$\ddot{A}$	Temperature dependent parameter
A_{t-i}	White noise
Cov_{xy}	Covariance of x and y
D_f	Daubechies wavelet family
D_j	Wavelet details
d	Particle diameter
$d_{j,b}$	Coefficient of mother wavelet at all scales from 1 to j
D_t	Oxygen defficit
DO_{sat}	Oxygen saturation coefficient
$F(f_k)$	Discrete Fourier transform
$F_n(t)$	Fourier polynomial
$F(\omega)$	Fourier transform
$F_{x,y}(\omega)^2$	Squared coherence of x and y
g	Gravity
$G_{x,y}(\omega)$	Gain of the cross-spectrum of x and y
$I(f_i)$	Intensity of periodogram
$I_{x,y}(\omega)$	Cross-periodogram of x and y
KN	Half saturation constant of nitrogen
KP	Half saturation constant of phosphorus
N	Sample size
p	Order of autoregressive process
q	Order of moving average process
r_j	Autocorrelation function
R_{DO}^{aer}	Contribution to re-aeration to the conversion rate of dissolved

$S_{j,b}$	Coefficient of father wavelet for maximum scale 2^J
	oxygen
t_c	Maximum dissolved oxygen
u	Wavelet location parameter
W_o	Central frequency of Morlet wavelet
$W_{j,t}$	Scale wavelet coefficients computed from X_t
$W(u,s)$	Wavelet transformed signal
$W(n)$	Window function
V_s	Settling velocity
$\vec{V}_w$	Kinematic viscosity of ambient water
$\bar{x}$	Mean value of x
X_t	Time series
$\bar{y}$	Mean value of y
$\overline{Y_g}$	Geometric mean of original data
z	Transfer constant

Greek symbols

λ	Power of transformation
τ	Tau - time lag
Δ	Window
ω	Angular velocity
μ	Mean of process
δ	Model constant
$\varphi_{x,y}(\omega)$	Phase of the cross-spectrum of x and y
Φ	Model parameters
$\phi_{j,b}$	Father wavelet
$\psi(t)$	Mother wavelet

$\psi^m(t)$	Morlet wavelet
$\psi^g(t)$	Gaussian wavelet
$\Psi_{u,s}(t)$	Wavelet basis function
$\gamma_{x,\tau}(S_j)$	Wavelet covariance
ρ_w	Density of ambient water
ρ_s	Particle density
$\rho_{xy}(\tau)$	Cross correlation of x and y
$\rho_x(S_j)$	Wavelet correlation
$\rho_{x,\tau}(S_j)$	Wavelet cross-correlation
$\sigma_x^2(S_j)$	Wavelet variance

List of Figures

List of Tables

1 Introduction

Though a common substance, life on earth cannot exist without water. The quality of water affects the survival, health and growth of plants and animals in and around aquatic ecosystems. Throughout the world, the quality of freshwater bodies has been dramatically altered by human actions. These alterations occur as a result of the over extraction of water resources and the use of freshwater ecosystems as waste disposal sites. Political concern for the quality of freshwater ecosystems has arisen in recent years as a result of pollution problems which are increasing as a result of population growth, industrial development resulting in domestic, industrial waste and agricultural run-off being discharged directly or indirectly into rivers, lakes and reservoirs (Roberts, 1993; Ongley, 1999).

The main pollutants of freshwater bodies are petroleum products, synthetic agricultural chemicals, heavy metals, hazardous wastes, organic matter, sediments and infectious organisms. The main sources of pollution of water bodies are municipal, industrial and agricultural wastes. They usually contain many synthetic chemical which cannot be broken down by natural processes and are referred to as persistent as well as non-persistent or degradable substances. The persistent wastes consist of pesticides, herbicides, fungicides and insecticides like dieldrin, DDT, metals like lead, mercury and cadmium (Cunningham and Saigo, 2003; Dunne and Leopold, 1988). They degrade very slowly or cannot be broken down at all, remaining in freshwater bodies for hundreds of years. The damage is usually irreversible or reparable only after decades or centuries. The non-persistent pollutants come from domestic sewage, fertilisers and some industrial wastes.

The ecological state of a freshwater resource at anytime is the result of the complex interaction between many different physical, chemical and biological

components or processes. Sound and effective management of a freshwater body cannot be done without a good understanding of the behavior of the components and the interrelationship between them that affect the ecological state of the water body. The physical, chemical and biological components are linked such that any action targeting one component will directly or indirectly affect another (Chiras, 1998; Boyd, 2000; Gerry and Michael, 2002). Given that it is at times not possible and very expensive to measure the different processes influencing water quality, decision makers need indicators that capture critical dynamics or serve as a clue of changes in ecological systems. Hence, in order to assess the current status of a freshwater body, appropriate water quality indicators have been developed and used. When measured across time, they indicate whether the state of the freshwater body is improving, deteriorating or remaining unchanged with regards to pre-determined standards. Effective indicators are able to alert or act as early warning system of a given problem before it gets too bad and they also assist in identifying the corrective measures that need to be taken.

The purpose of analyzing ecological signals is usually to understand the underlying structure of the natural processes that generate the signals and to build models that capture as many aspects as possible of the processes (Pahl-Wostl, 1995; Grigg, 1996; Klapper, 2003). Data points, recorded over time usually have an internal structure such as autocorrelation, trend or seasonal variation that should be accounted for. Analyzing the observations of a process, extracting important features regarding its morphological and statistical structure, provides information helpful for building models (Brillinger, 1981; Kumar and Foufoula, 1998). Statistical and mathematically rigorous tools of signal analysis that can reveal important features of a process and structure not apparent from direct observation are a key component of process understanding. Understanding the time varying and frequency behavior, the relationship that exists between indicators is a prerequisite for establishing sound models useful for policy analysis and sustainable management of freshwater bodies.

Water quality indicators are recorded as a function of time and have been usually investigated by means of appropriate time domain techniques (e.g correlation functions, test functions) (Straškraba and Gnauck, 1985). Some few papers make use of the frequency domain methods in investigating water quality indicators (Nesmeřak and Straškraba, 1985). Given that many natural systems have frequency dependent variability, an understanding of this frequency dependence gives more information concerning the underlying physical mechanism that produced the signal. The Fourier transformation is used to project a signal from the time domain into the frequency domain so as to extract frequency dependent information and hidden periodic components contained in the signal. The Fourier analysis (periodogram, spectral and cross spectral analysis) offers the researcher the possibility to examine the periods and amplitudes of fluctuations in a time series and between time series as a function of frequency (Alegue and Gnauck, 2006b). The time domain methods reveal the time at which events occur but give no information concerning the frequency at which they occur. The frequency domain methods simply give information on the frequency content of the signal, but no information on when the different frequencies occur. To obtain frequency and time dependent information, both methods have to be used complementarily.

In addition to classical methods of signal analysis, new development in the field of mathematical statistics has led to wavelet analysis which reveals not only when the signal occurs but the frequency content (Strang and Nguyen, 1997). Wavelet analysis as a modern signal analysis tool, has been used to investigate the time scale behavior of signals in fields such as geophysics, medicine, electronics (Lau and Weng, 1995; Mallat, 1998), and also needs to be used for investigating indicators of freshwater ecosystems.

The recorded observations of many ecological indicators of freshwater ecosystems are an amalgam of components or processes operating at different time scales (corresponding to different ranges of frequencies active at specific time intervals in the Fourier domain) which cannot be revealed by classical signal analysis methods (Donoho and Johnstone, 1995). Answering scientific

questions about the processes represented by the observed data are often inherently linked to understanding the behavior within and between processes not just at a given time scale, but at different time scales. Effective signal analysis provides starting information for modeling and enables a general description of causal relationships (Straškraba and Gnauck, 1985). The information extracted may also be helpful in estimating the weight of input and disturbance relating to the output variables of the model concerned as well as serve as a useful tool when appropriate initial values for estimate procedures are to be chosen. The frequency domain methods are not suitable for time series which are non stationary and do not offer the possibility to identify in time each of the active frequency components present in the time series (Rebecca, 1998; Granger and Engle, 1983; Bhagavan, 1985). The time domain methods on the other hand offer no possibility to examine the frequency content of a time series. When things happen is known, but not the frequency at which they happen. During modeling of freshwater quality, planning and management of freshwater ecosystems, the variations occurring at different time scales are not investigated and taken into consideration because of the limited capabilities of classical methods.

Today, the analysis and management of freshwater ecosystems cannot be restricted exclusively to a particular field of study, but must be interdisciplinary. Useful theories are being developed in other scientific fields. In order to acquire a deeper insight of these systems, it will make sense to examine modern interesting methods in order to investigate how they can be adapted and experimented or put into practice in the field of the analysis and management of freshwater ecosystems. Such an interdisciplinary approach will lead to more informed monitoring, modeling and management of these ecosystems. Wavelets analysis, a modern tool that has proven quite useful for time scale based signal analysis, offering a solution to the time scale analysis problem because it provides an effective approach for extracting both the information on the time localization and the frequency content of the time series (Mallat, 1998; Daubechies, 1992; Chui, 1992). It has the ability to decompose time series into

several sub-series which may be associated with particular time scales (Bruce and Gao, 1996; Keinert, 2004). As a result, the interpretation of features in water quality time series may be facilitated by first applying an appropriate wavelet transform and subsequently interpreting each individual sub-series. Modifications can be done on the individual sub-series before recombining them in order to improve on the quality of the signal.

The ability of wavelet analysis to capture variability in time and scale has been performed primarily on univariate processes to unravel the microstructure of signals. Investigating how such a tool, in combination with the classical methods, can be used to extract useful information from freshwater indicators is of paramount importance. The application of wavelets to bivariate analysis of processes can be quite helpful in revealing relationships between signals at different time scale (Whitcher, 1998). Hence, as a complement to the classical methods, wavelet analysis can enable the researcher to understand the dominant scale of variability in indicators, the optimal sampling frequency, and the interrelationship between the indicators especially at different time scale in a particular freshwater body. It is quite important to practically use this modern method in combination to the classical methods for analyzing freshwater signals. Usually, the question of how complex scientific theories, methods or useful scientific results can be used by water quality managers is not answered. There is a need for better models, expert judgement for management of freshwater ecosystems to easily comply with established water quality standards. It will also be interesting to examine how the results obtained from classical and modern signal analysis methods can be used in a practical way for enhancing the monitoring, modelling and sustainable management of freshwater ecosystems.

The main objective of this thesis is to investigate how wavelet analysis and classical methods can be used to extract useful information concerning the underlying structure and relationships between water quality signals with the aim of enhancing the monitoring, modelling and management of these ecosystems. In addition, the methods will be used to compare the underlying

structure of the same indicators from different freshwater bodies. The emphasis is on conveying the ideas behind the methods as opposed to presenting a comprehensive mathematical development which is available in many texts. The approach will concentrate on exactly what wavelets and classical methods can tell us concerning water quality time series and it's applicability to water quality management.

The pollution and indicators of freshwater ecosystems are introduced in chapter two. A description of the type of pollution, pollutants and their main sources, relevant equations as well as their effect on human and freshwater ecosystems are presented. In addition, the development and use of freshwater quality indicators as well as some essential freshwater quality indicators are given. The standards, which are the main administrative tools available for evaluating freshwater quality by the help of the indicators, are described. Finally, the methods of pre-treatment of time series of these indicators before any analysis so as to avoid misleading results are examined.

Chapter three begins by presenting the classical signal analysis methods which consist of the time and frequency domain approaches with references given for full mathematical overview of the methods. The classical methods are provided before the wavelet analysis, a modern signal analysis approach is then introduced. Different wavelet families such as the Daubechies, Morlets and Gaussian families as well as different wavelet transforms are presented. The wavelet multiresolution analysis, wavelet variance and covariance, wavelet correlation and cross-correlation are introduced with some examples.

Chapter four contains applications of how the methodology introduced in chapter three performs on real time series of freshwater quality indicators. A chemical, biological and a physical water quality indicator from the Havel River in the State of Brandenburg in Germany are investigated by means of both the classical and wavelet analysis methods. The wavelet approach gives new insight and responds to some pertinent questions that classical methods were not able to answer. In addition, a comparative investigation of dissolved oxygen signal from the River Elbe, Havel and Oder by means of univariate signal

analysis methods is done. Despite its ability to act as a lens, it is found that wavelet analysis cannot replace the classical methods which also have their specificity. Rather, the modern methods should be used to complement the classical ones.

A summary of the findings of this thesis is presented in chapter 5 while the conclusions drawn from the thesis are given in chapter 6 along with open questions and future directions.

2 Pollution and Indicators of Freshwater Ecosystems

Escalating concern about the quality of freshwater ecosystems as a result of pollution caused by anthropogenic activities has motivated efforts to monitor and assess their ecological status and trends (Gore and Petts, 1989; Botkin and Keller, 2005). Monitoring physical, chemical and biological water quality is the key means of assessing freshwater ecosystem health by providing a direct measure of concentration of substances known or believed to affect humans and the ecological status of a given water body. Measuring and interpreting all factors and variables interplaying in a given ecological issue causing concern is not possible, hence indicators are used. Ecological indicators include physical, chemical and biological measurements and indices that attempt to characterize or summarize critical and often complex freshwater ecosystems components (US EPA, 2000; Dale and Bayeler, 2001). They act as signs or signals that transmit a complex message from numerous sources in a simple and useful manner. They can be easily used in assessing the condition and changes occurring in a freshwater resource across time. Ecological indicators are also able to act as an early warning signal of ecological changes and assist in diagnosing the cause of an ecological problem. Depending on the desired use of a water body, water quality standards are established. Water quality standards give concentrations of substances which are indicators of the suitability of the water body for a particular use. Effluent standards are the main administrative tool used for the control of point source pollution while the ambient standard enable the determination of the suitability of the water body for a particular use (Koren, 1991).

In this section, the pollution of freshwater ecosystems, which is a real threat to human, animal and plant health, with water quality which is becoming the limiting factor to sustainable development in some parts of the world, will be

examined. The notion of indicator as well as some water quality indicators used for investigating freshwater ecosystems is examined. In addition, water quality standards used for the control of freshwater quality are presented. Given that most water quality data from monitoring programs contain errors or violate certain assumptions required for analysis, some approaches of data pre-treatment are given.

2.1 Pollution and pollutants of freshwater ecosystem

Pollution is a term applied to any environmental state or manifestation which is harmful or unpleasant to life, resulting from failure to achieve or maintain control over the chemical, physical or biological consequences or side effects of human scientific, industrial and social habits (Collocott and Dobson, 1974). From another perspective, water pollution refers to any physical, biological or chemical change in water quality that adversely affects living organisms and makes water unsuitable for desired uses as a result of anthropogenic activities (Uhlmann, 1991; Cunningham et al., 2003). Many human activities from water supply and sanitation to transport, mining and the chemical industry have the potential to pollute water bodies. Freshwater pollution comes from point and non point sources. Sewage treatment plants, factories, power plants, underground coal mines, oil wells are usually considered as point sources because they come from specific locations such as drain pipes, ditches or sewer outflows. It is relatively easy to monitor and regulate such sources by means of effluent standards. On the contrary, non point sources such as run-off from fields and animal feedlots, lawns and gardens, golf courses, construction sites, street and parking lots, logging areas and roads, are scattered, having no specific location where they discharge into a particular water body. Rainfall flushing polluting substances such as gasoline, oil, and lead from the above mentioned sites is the main route into freshwater bodies. A more subtle diffuse or non point pollution is the atmospheric deposition of contaminants carried by air currents and precipitated into watersheds or directly into water bodies. Though the types and sources of freshwater pollutants are usually interrelated,

it is quite convenient to separate them into major categories in order to study them.

2.1.1 Oxygen demanding wastes

When certain organic material such as sewage, food processing wastes, paper pulp are discharged into freshwater bodies, oxygen is used up by decomposers to break down these substances. High concentrations of organic matter released into rivers by untreated industrial and municipal wastes causes a sharp decrease in dissolved oxygen (which at times result in anoxia) and a consequent suffocation of aquatic organisms like fishes and a release in ammonia and nitrite downstream of the effluent input. Along the length of a river, the most obvious effect of high concentrations of organic matter is the oxygen sag curve which can be observed from a few kilometers to 100 km downstream of the input (Pepper et al., 1996; Chapman, 1996). Industrial activities like the pulp paper, palm oil extraction and sugar beet processing may produce waste waters with BOD and COD values exceeding 1,000 mg/l capable of totally depleting dissolved oxygen present in a freshwater body. The impact of oxygen demanding wastes in a freshwater body will depend on the amount of waste discharged and on the size and flow of water through the freshwater body. The following general equation represents the decomposition.

Organic matter + bacteria + O_2 —> CO_2 + H_2O

Most dissolved oxygen models are based on the Streeter-Phelps equation (1925) which is given by

$$\frac{dD}{dt} = KL \text{ or } \frac{dD}{dt} = -K_2 D$$

where D is the dissolved oxygen deficit, L is the concentration of organic matter, K is the coefficient of de-oxygenation, and K_2 the coefficient of re-aeration .

The mass transfer of oxygen can be expressed as

$$R_{DO}^{aer} = K_2.(DO_{sat} - DO)$$

where R_{DO}^{aer} is the contribution of re-aeration to the conversion rate of dissolved oxygen, DO_{sat} is the oxygen saturation coefficient (Elmore and Hayes, 1960) given by

$$DO_{sat} = 14.652 - 0.41022*T + 0.007991*T^2 - 0.000077774*T^3$$

where T is the temperature (°C).

The oxygen sag models as a result of oxygen demanding wastes are given by

$$t_c = \frac{1}{(K_2 - K)} \ln\left[\left(\frac{K_2}{K}\right)\left(1 - D_0\left\{\frac{K_2 - K}{KL_0}\right\}\right)\right]$$

where t_c is the maximum dissolved oxygen, DO is the initial dissolved oxygen at the point of waste discharge. Determining the oxygen deficit (D_t) at time t is given by

$$D_t = \left(\frac{KL_0}{K_2 - K}\right)\left(e^{-K_t} - e^{-K_{2t}}\right) + D_0 e^{-K_{2t}}$$

where L_0 is the concentration of organic matter at time zero.

2.1.2 Nutrients

Nutrients such as nitrogen, phosphorus, carbon, sulphur, silica and potassium stimulate plant growth. While nitrogen is primarily found in nitrates and ammonia, phosphorus is found in phosphates. These two nutrients are the main limiting factors for the populations of algae and other plants. When depleted, Silica can limit the growth of diatoms. The main sources of these nutrients are from excess fertilizers from croplands. Laundry detergents are equally an important anthropogenic source of nutrients. Nitrates and nitric acid can also enter freshwater bodies through the atmosphere. When the levels of phosphorus and nitrogen drastically increase in a freshwater body, the growth of plants goes wild practically choking lakes and rivers with thick mats of algae

or dense growth of aquatic plants (Laws, 2000; Boon, 1992). This condition impairs fishing, navigation, swimming, recreational boating. The death, decay and decomposition of these plants cause a fall in the level of dissolved oxygen concentration killing aquatic organisms. The equation from Tchobanoglous and Schroeder (1985) suggest that

$CO_2 + H_2O$ + Nutrients (in the presence of light) —> CH_2O [new algal cells] + O_2

Assuming that the stoichoimetry of the algal biomass is constant, Monod-type equations for growth limitation factor is given by

$$f(N) = \frac{NH_4 + NO_3}{KN + (NH_4 + NO_3)}$$

and

$$f(P) = \frac{DIS - P}{KP + (DIS - P)}$$

where *KN* and *KP* are the half saturation constants for nitrogen (*N*) and phosphorus (*P*).

2.1.3 Suspended solids

Suspended solids are suspended mineral and organic particles which do not settle immediately creating turbidity plumes making the water body murky. They come from agricultural soil erosion, forestry or construction, urban runoff, industrial effluents and excess phytoplankton growth (US EPA, 1997). When they settle, they create deposits that may suffocate benthic organisms. The source of suspended solids is usually as a result of human activities in the watershed like soil erosion from agriculture, forestry or construction, urban run-off, industrial effluents which results in sedimentation. Sediments reduce water depth favoring infestation by aquatic macrophytes, and when highly organic

result in anaerobic conditions in shallow areas of the water body. The sedimentation of suspended particles is given by Stokes equation,

$$V_s = \frac{1}{18}\left[\frac{(\rho_s - \rho_f)gd^2}{V_w}\right]$$

where V_s is the settling velocity, ρ_s is the density of the particle, ρ_w is the density of ambient water, d is the particle diameter, V_w is the kinematic viscousity of ambient water and g is the gravity.

2.1.4 Toxic organic and inorganic pollutants

Thousands of natural and synthetic organic chemicals are used in the chemical industry in making pesticides, plastics, pharmaceutical, pigments and other products used in everyday life. Most synthetic chemicals like pesticides, herbicide and PCBs, manufactured artificially are problematic in water bodies of developed and developing countries (Ripl and Wolter, 2003). They enter water bodies as point source from sewers and effluents from industrial and municipal sources as well as leaching of solid and liquid waste dumps or agricultural run-off. Many of them are extremely toxic even at small concentrations, being able to cause birth defects, genetic disorders and cancer. Most of them are persistent, degrading very slowly, some bioaccumulate and are biomagnified in the food chain (Landner, 1989; Dunne and Leopold, 1988). The most important sources of toxic organic chemicals are from improper disposal of industrial and household wastes, run-off from farm fields, golf courses, forest, and roadsides.

Inorganic water pollutants encompass metals, acids and salts. Acids, salts, nitrates and chlorine are normally not toxic at low concentration, but affect biological communities at high concentrations. Acidification of running waters is a result of direct inputs of acidic waste waters from mining or specific industries or indirect inputs through acidic atmospheric deposition mainly as nitric and sulphuric acids resulting mainly from motor exhaust and fossil fuel consumption. Rainfall in densely populated areas and highly industrialized zones are a major cause of acidification (Jørgensen, 1997). Acidification

causes the solubilization of certain metals particularly Al^{3+} when the pH falls below 4.5. This increase in concentration of the metal can be toxic to fish causing the water body to be unsuitable for certain uses. Metals such as mercury, lead, cadmium and nickel are highly toxic and can cause death even at low concentrations. Due to their highly persistent nature, these metals accumulate in food chains having a cumulative effect in humans.

2.1.5 Microbial contaminants

Pathogenic organisms are the most serious water pollutants in terms of human health all round the world. The main source of these pathogens is from improperly treated human wastes. In addition, animal wastes from feedlots or fields near waterways and food processing industries with improper waste treatment facilities are a source of pathogens into freshwater bodies. Contrary to the dramatic situation in less developed countries, effective sewage treatment plants in developed countries have reduced or eliminated most of the worst sources of pathogens in freshwater bodies. The problem in developing countries is exacerbated in fast growing cities and by the high population growth rate that far exceeds the development rate of waste water collection and treatment facilities (Maybeck et al., 1989). In rivers flowing through major cities like New Delhi and Djakarta, faecal coliform counts can be in numbers of 10^6 per 100ml which far exceeds safety levels prescribed by the WHO (Chapman, 1996). These pathogens, which mainly consist of bacteria, viruses and protozoan, cause water borne diseases such as cholera, typhoid, bacteria and amoebic dysentery, enteritis, polio, infectious hepatitis and schistosomiasis. In addition, filariasis, yellow fever and malaria which are rampant in developing countries, are transmitted by insects which have aquatic larvae.

2.1.6 Thermal pollution

Changes in water temperature can be caused by urbanization, agriculture, forestry, impoundments and the discharge of industrial effluents. Downstream of coal and nuclear electrical power generating plants, where heated water is discharged into the water body on a continual basis are most affected by this

type of pollution because temperature affects the basic physical and chemical processes necessary for life. Changes in the natural temperature pattern in a freshwater body are able to eliminate species that are adapted to the natural cycle of temperatures of the water body. All biological reactions strongly depend on environmental conditions with temperature being the most influential. This influence is most often expressed by means of the general function:

$$\ddot{A}_{(T)} = \ddot{A}_{(20°C)} e^{c(T-20)}$$

where T is the temperature (°C); Ä is the temperature dependent parameter; *c* is the temperature constant.

2.1.7 Some health impacts

For a healthy living, clean water is an absolute necessity. Because of the effects of certain anthropogenic activities on freshwater bodies, millions of humans are deprived of an adequate supply of fresh and clean drinking water. As a result of the contamination of water bodies with heavy metals, persistent organic pollutants, faecal material and nutrients, serious health problems have resulted with 80% of diseases in developing countries being water related (UNEP, 2002). Chemicals causing health disorders may be naturally present in water bodies or may be introduced by human activities. Pesticides contain organophosphates and carbonates which damage the nervous system. The presence of chlorides results in a damage of the reproductive and endocrine system. Most pesticides contain carcinogenic substances well above safety levels which may result in cancer. Lead has the ability to accumulate in the human body and affect the central nervous system with pregnant women and little children being the population most at risk (Borney, 1994). Fluorides, when in excess results in yellow teeth, damage of the spinal cord and cause crippling disease. High concentrations of nitrates in drinking water cause the blue body syndrome, a condition whereby a very restricted amount of oxygen reaches the brain resulting in death (US EPA, 1992). Nitrate is also associated with cancer of the digestive tract. The presence of arsenic in drinking water has the

potential to cause skin cancer, vascular diseases as well as liver and nervous system damages (Viessman and Hammer, 1998). The presence of petrochemicals like benzene in drinking water causes cancer even at very low concentrations. Heavy metals cause serious damages to the kidney and nervous system as well as vascular diseases. And the list of human health problems resulting from unsustainable use of freshwater bodies is extremely long and needs to be dealt with more seriously. Figure 2.1 provides a three compartment model consisting of blood, liver and bones with the respective toxicological equations blood, liver and bone based on Jørgensen (1994).

The mass balance on blood is given by

$$\frac{dp_1}{dt} = R - Z_{3,1}p_1 - Z_{2,1}p_1 - Z_{0,1}p_1 + Z_{1,2}p_2 + Z_{1,2}p_3$$

The mass balance on liver is given by

$$\frac{dp_2}{dt} = Z_{2,1}p_1 - Z_{1,2}p_2 - Z_r\left(\frac{p_2}{Z_m + p_2}\right)$$

The mass balance on bones is given by

$$\frac{dp_3}{dt} = Z_{3,1}p_1 - Z_{1,3}p_3$$

with *Z* representing the different transfer constants, *R* the constant intake rate and P_1, P_2, P_3 representing the pollutant concentration in blood, liver and bone respectively as indicated by figure 2.1.

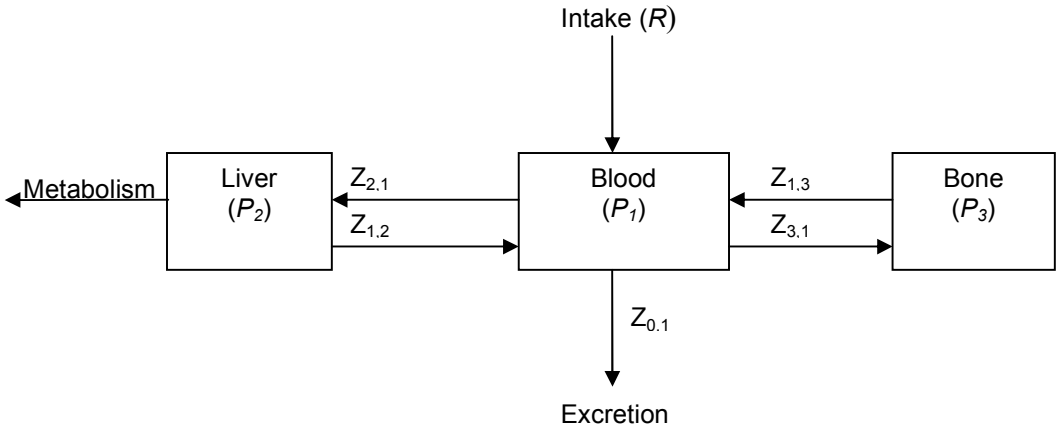

Figure 2.1: Interaction between compartments of a toxicological model (adapted from Nirmalakhandan, 2002).

With emerging issues such as global warming, invasive species, pharmaceuticals and personal care products in water bodies, freshwater ecosystems are further threatened. Water quality is linked to all of the eight millennium development goals, and directly affects the livelihood and survival of humans, there is need to efficiently and effectively protect and manage these fragile resources.

2.2 Freshwater quality indicators

Water quality change can be associated with changes in water quality parameters such as dissolved oxygen levels, nutrient concentrations, pH and temperature. To assess and monitor changes in the state of freshwater ecosystems, data of water quality parameters, collected through monitoring programs can be analyzed, interpreted and ecosystem health information characterized. A freshwater quality indicator is a parameter or value that reflects the condition of an ecological component, usually with a significance that extends beyond the measurement or value itself (Shear et al., 2005). When used alone or in combination with others, indicators are able to provide the means to assess progress towards desired goals and objectives. It enables the water resource manager to determine whether his objectives are being met,

are closer to being met or whether the conditions are deteriorating. Hence, indicators of freshwater ecosystems can be used to assess the condition of these ecosystems, to provide an early warning signal of changes in the freshwater body, or to diagnose the cause of a problem in the freshwater body. Using indicators is practical because not everything can be recorded in the environment. It requires a measurement which describes the observed state of the ecosystem component and a reference value which reflects the desired state of the ecosystem component. A comparism between the measured and the desired state is necessary in assessing the magnitude of change objectively, the gravity of the change and the success of amelioration efforts. The value of an ecological indicator rest on the premise that better understanding of what is happening in the freshwater ecosystems leads to better and more effective policies for encouraging desirable changes, discouraging undesirable changes, and maintaining variability within tolerable limits (Belnap, 1998; Stork et al., 1997).

Ecological indicators of freshwater ecosystems can be used for diagnosis, warning and for monitoring change. They are usually good in describing the present conditions and facilitating prediction of the future state by the help of appropriate models. Good ecological indicators quantify information so that its significance is clear, simplify information about complex phenomena to improve communication, and is a cost-effective and accurate alternative to monitoring individual processes, species, et cetera (Landres, 1992). In developing ecological indicators, it is necessary to ensure that they are complete enough to capture the dynamics of key processes without being so complex that their indication is fuzzy.

2.2.1 Types of water quality indicators

The variables recorded by monitoring programs may be of different origins and can be denoted as stressors to the ecosystem (e.g. nutrients), state variable of the ecosystem (e.g. biomass) or driving force of the ecosystem (solar radiation). These give rise to stressor indicators, ecological state indicators and

driving force indicators (Shear et al., 2005). Some variables can both be stressors and state variables depending on the specific situation.

Table 2.1: Essential freshwater quality indicators and their indication.

Indicator	Indication	Type of Indicator	Category
Water temperature	Energy	Physical	Driving force
Solar radiation	Light		
Secchi depth	Transparency		Ecological state
Turbidity			
Suspended solids			
Nitrogen, phosphorus	Nutrients	Chemical	Stressor
BOD	Pollution		
Conductivity			
Dissolved oxygen	Productivity, Respiration		Ecological state
pH	Acidity, CO_2, Alkalinity,		Stressor
Chlorophyll-a	Algal biomass	Biological	Ecological state
Faecal coliform	Faecal material		Stressor

There are a number of classification schemes or models for indicators like the pressure-state-response model (OECD, 2001). The most used ecological indicators are the ecological state indicators and the pressure or stressor indicators that have and effect on the ecological state indicators. In table 2.1, the essential freshwater quality indicators are given. The measurements of the indicators obtained depend on both natural and anthropogenic driving forces.

Fluctuations from the mean value over time may be as a result of parallel acting internal and external driving forces (figure 2.2). Seasonal and periodic processes influence the variable causing most of them to exhibit a cycling behavior over time (Alegue and Gnauck, 2006).

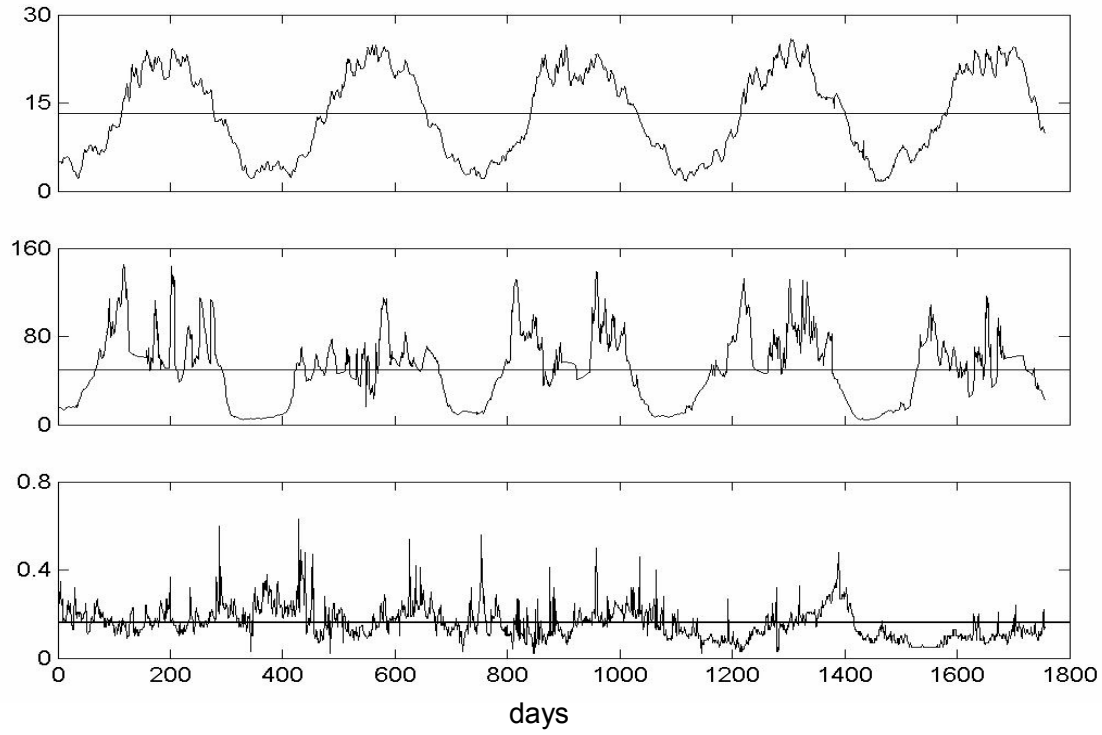

Figure 2.2: Time variations and fluctuations of indicators (water temperature (°C)(top), chlorophyll-a (µg/l)(middle), phosphorus (mg/l)(bottom)).

An appropriate set of ecological state indicator enables the water quality manager to describe and diagnose any current problems while stressor indicators enables to know some of the causes and predict the future state of the system. As an example, the ecological state indicator, dissolved oxygen captures the balance between primary production and respiration. When the ambient concentration becomes much higher than the saturation value at a given temperature and pressure, it indicates a dominance of photosynthetic activities, high nutrient concentration, algal blooms and or the excessive growth

of macrophytes. On the other hand, when the ambient concentration becomes much lower than the saturation value at a given temperature and pressure, it indicates a dominance of respiration as a result of organic enrichment from wastewater or upstream increase in primary production as a result of nutrient enrichment. This information is quite useful because when dissolved oxygen concentration falls below certain threshold values, fishes as well as invertebrates that process organic waste decline (Barthelmes, 1981).

2.2.2 Use of indicators

Developing and using freshwater quality ecological indicators is challenging with regards to determining which of the numerous measures of the ecological systems effectively characterize the system while being quite simple so as to be monitored and modeled effectively and efficiently. They should be able to quantify the magnitude of stress, the degree of exposure to stress, or the degree of ecological exposure (Hansaker and Carpenter, 1990; Suter, 1993) and provide a simple and efficient means to examine the composition, structure and function of the freshwater ecosystem (Karr, 1997a; 1997b). A baseline or reference state is necessary to compare status and trends measured by an ecological indicator. Hence, the range of values the indicator can take and what the values mean as well as the temporal and spatial scale over which the indicator is likely to change should be clear. The use of ecological indicators of freshwater ecosystems rest on the assumption that, the presence or absence of fluctuations in indicators reflects changes occurring at different levels in the system (Dale and Beyeler, 2001).

Using freshwater quality indicators as an aid in managing freshwater ecosystems faces the following challenges. Firstly, monitoring programs usually depend on a small number of indicators and do not consider the full complexity of freshwater ecosystems. Secondly, the choice of indicators is confounded in management and monitoring programs which have fuzzy long term goals and objectives. When the goals and objectives are fuzzy, inappropriate variables for management decision making may be used. Early setting of clear goals and

objectives help monitoring to focus on pertinent current and future issues, with the selection of ecological indicators that measure the characteristics of the freshwater body closely related to management concerns. Finally, management and monitoring programs often lack scientific rigor as a result of the failure to use a defined protocol for the identification of relevant ecological indicators. Establishing and using standard procedures for the selection of ecological indicators will enable repeatability, avoid bias as well as impose some discipline during the selection process. This will thereby ensure that the indicator selected will encompass management concerns (Slocombe, 1998; Belnap, 1998). Hence, freshwater quality indicators are required to capture the complexities of the ecosystem while remaining simple enough to be easily and routinely monitored.

Key to the success of a monitoring program is the selection of effective indicators. As a result of the great variety of freshwater quality issues, the complexity of environmental data, and the necessity for management decisions, many types of indicators have been developed for many different purposes. Due to their great diversity, the selection of successful ecological indicators for with regards to an issue needs to be done by the help of appropriate criteria. Some essential criteria are as follows:

Firstly, it must be conceptually relevant. Useful freshwater quality indicators are based on clear conceptual models of the structure and functioning of the freshwater body. The chosen ecological indicator should provide answers to relevant assessment questions which are developed as a result of a particular environmental issue of concern. Using the wrong indicators provides accurate information that is totally useless for management decision making on the issue. In addition, it must be easy to measure and implement. The sampling techniques for the ecological indicator should be feasible, straightforward and relatively inexpensive and appropriate for use in the monitoring program. The indicators should be easy to understand, simple to apply and provide information to managers and policy makers that is relevant, scientifically sound, easily documented and cost-effective (Stork et al., 1997; Lorenz et al., 1999).

Next, it must be sensitive to stresses on the system. The water quality indicator should display a high degree of sensitivity to a particular and, perhaps, subtle stress, thereby serving as an early warning of imminent water quality problem. Finally, it must be easily interpretable and useful. The water quality indicator should convey information on the ecological condition in a manner that makes sense to freshwater resource managers with no fuzziness. Its usefulness will depend on its ability to report or predict changes that can be averted by management actions. Using a set of indicators representing the structure, function and composition of a freshwater resource is quite necessary.

2.2.3 Essential freshwater quality indicators

2.2.3.1 Dissolved oxygen and biochemical oxygen demand

Nearly all aquatic forms of life require oxygen to survive resulting in the frequent measurement of dissolved oxygen (DO) and biochemical oxygen demand (BOD) by monitoring programs. The concentration of DO indicates how well the water body is aerated as well as the organisms capable of surviving in it. Variations of DO concentration is as a result of season, time of the day, temperature and salinity. The BOD serves as a measure of the amount of oxygen consumed in the water by chemical and biological processes. Being one of the most important ingredients of freshwater ecosystems, the concentration of oxygen in the water is one of the most appropriate indicators of the freshwater ecosystem health. A decline in DO concentration in a water body forces most animals to quickly move to areas with higher levels of oxygen or perish. A water body with little or no oxygen is not able to support healthy levels of animal or plant life making the oxygen concentration an important indicator for assessing water quality.

Dissolved oxygen

Oxygen enters freshwater ecosystems through photosynthesis by aquatic plants and from the atmosphere through diffusion. The action of wind on water surfaces puts more water in contact with the atmosphere further boosting the

amount of oxygen in the water. It is essential to all forms of aquatic life including the organisms responsible for the self-purification processes in natural water bodies. It varies as a function of temperature, salinity, turbulence, atmospheric pressure, the photosynthetic activity of algae and macrophytes and the respiration of plants and animals.

Concentrations of dissolved oxygen in unpolluted waters are generally close to less than 10mg/l (Novotny, 2003). In addition to being used for respiration, oxygen is needed to aid in decomposition of organic matter which is an integral part of freshwater ecosystem function. The dissolved oxygen of a water body can be depleted as a result of the decomposition of large quantities of organic matter by bacteria, making it not habitable for many species. This can be enhanced as a result of an overload of nutrients from wastewater treatment plants or runoff from urban storm water or runoff from urban storm water or other land uses. An increase in nutrients results in an overgrowth of phytoplankton. When the phytoplankton dies and gets to the bottom, its decomposition uses oxygen in the deep waters. On the other hand, when oxygen concentration is low, nutrients bound to bottom sediments are released into the water column, enhancing more phytoplankton growth and subsequent oxygen depletion.

Dissolved oxygen concentration in a water body can vary greatly with depth during certain times of the year as a result of vertical stratification. Stratification blocks the transfer of oxygen between the upper and lower layers and may result in hypoxia and even anoxia in the lower layers. During overturn which usually occurs during changing seasons, oxygen rich surface water mixes with the oxygen poor deep water. When the level of dissolved oxygen fall below a certain threshold value, mobile animals move to areas with high dissolved oxygen and immobile species perish. When dissolved oxygen levels are above 5 mg/L, most plants and animals can grow and reproduce freely (Chiras, 1998). They have become stressed at concentrations between 3-5 mg/l and at concentrations below 3 mg/l, a condition known as hypoxia, mobile species move elsewhere and most immobile species die (Boyd, 2000). At

concentrations of less than 0.5 mg/l known as anoxia, all species requiring oxygen for survival die.

Dissolved oxygen saturation is a measure of the highest dissolved oxygen concentration possible under the environmental limits of atmospheric pressure, temperature and salinity. As temperature and salinity increases, the amount of oxygen capable of being held by water decreases significantly. The percent saturation is the amount of oxygen present in the water relative to the potential dissolved oxygen saturation of the water at a given temperature and pressure. This can be calculated as follows:

Percentage saturation = (measured DO/ DO saturation) x 100

Biochemical oxygen demand

BOD is a measure of the amount of biochemical degradable organic matter present in a sample of water. It measures the amount of oxygen that aerobic microorganisms consume or require to decompose or oxidize an organic matter to a stable inorganic form. The rate of oxygen consumption in this reaction is determined by the water temperature, the type of organic and inorganic material present and the presence of certain types of microorganisms. The amount of oxygen present in a water body is directly affected by the BOD. The consequences of high BOD are similar to that of low dissolved oxygen concentration. Hence, the greater the BOD in a water body, the more rapidly oxygen is depleted causing less oxygen to be available to aquatic animals. The result is that many aquatic animals become stressed, suffocate and even die. Unpolluted waters have a BOD of less than 5 mg/l, raw sewage has a value of between 150 to 300 mg/l and wastewater treatment plant effluent have a value of between 8 to 150 mg/l and industrial wastes may have a value of up to 25,000 mg/l (Novotny, 2003).

2.2.3.2 Nitrogen and phosphorus

The nutrients nitrogen and phosphorus are key water quality indicators. Depending on the chemical form, these nutrients significantly affect in a direct or indirect way the growth of plants, concentration of oxygen, clarity of water, rates of sedimentation and other freshwater ecosystem processes. The primary role of nitrogen is for DNA and protein synthesis and for photosynthesis. The presence of phosphorus is critical for metabolic processes involving the transfer of energy. Their concentration varies according to the surrounding land use, geology and season.

As chemical substances critical for the maintenance, growth, reproduction, protection from diseases and survival of plants, a number of nutrients namely, nitrogen, phosphorus, carbon, oxygen, silica, magnesium, potassium, iron, zinc, calcium and copper, are required. Of these nutrients, phosphorus and nitrogen are the two most important nutrients essential for aquatic plant growth; their amounts in freshwater ecosystems have increased significantly as a result of anthropogenic activities.

Though essential for the survival and growth of aquatic plants, excess nitrogen and phosphorus set off a set of processes that seasonally deplete DO in freshwater ecosystems. An overabundance of these nutrients leads to eutrophication, a condition whereby high nutrient concentrations stimulate excessive algal bloom. This causes the water to be cloudy preventing light from reaching submerged aquatic vegetation which needs light for photosynthesis, causing them to die. A resultant fall in oxygen concentration as oxygen consuming bacteria decompose dead phytoplankton causes hypoxia or even anoxia.

Certain species of bacteria and blue-green algae can fix nitrogen gas, converting it into an inorganic nitrogen form, making it available to plants. In the freshwater body, nitrogen exist in a variety of chemical forms namely, ammonium, nitrate, and nitrite and particulate and dissolved forms. By nitrification, some bacteria transform ammonia into nitrite and then nitrate in an oxygen consuming process. In oxygen deficient conditions, nitrification is

inhibited and ammonia or nitrite forms of nitrogen accumulate. Under anoxic conditions, bacteria converts nitrate to nitrite and to nitrogen gas in a process called denitrification. Nitrate and urea as well as ammonia are very soluble. All these forms promote phytoplankton, algae and bacterial blooms (Phinney, 1999). Several forms of phosphorus also exist in freshwater bodies, namely, organic phosphate, orthophosphate (inorganic, dissolved phosphorus), total phosphorus (dissolved and particulate), and polyphosphate (from detergents). Under oxygen rich conditions, phosphate forms chemical complexes with minerals like manganese, aluminum, and iron which fall to the sediments. Under anoxic conditions, phosphate bound to the sediments is released into the water and can initiate another round of phytoplankton blooms.

2.2.3.3 pH

As a measure of the hydrogen ion concentration of water, pH is used as an indication of acidity with high pH values in standing waters during the warmer months being associated with high phytoplankton densities. The consumption of carbon dioxide by phytoplankton during the day in the process of photosynthesis causes a rise in pH and subsequent fall at night with the release of carbon dioxide due to respiration. In productive standing waters, carbon dioxide levels fall to very low levels causing the pH to rise to around 9 standard units and above.

The pH of a water body in critical to the survival of almost all aquatic plants and animals. It acts as a measure of the degree of acidity or alkalinity of a solution by determining the amount of hydrogen ion (H^+) present in the solution, expressed as a negative logarithm on a scale of 0 to 14 standard units. A pH of 7 (pure water) is basic, below seven is acidic and above 7 is basic or alkaline. The optimum for most organisms is between 6.5 and 8.5. pH values are based on a logarithmic scale implying that for each 1.0 change of pH, the alkalinity or acidity changes by a factor of 10. Hence, a pH of 5.0 is ten times more acidic than a pH of 6.0 and 100 times more acidic than a pH of 7.0.

A freshwater body with pH values outside the range of 5.0 to 10.0 is considered as an indication of industrial pollution or some natural catastrophe. Most aquatic organisms have difficulties surviving when pH drops below 5.0 or above 9.0. pH changes also affect the solubility of some metals such a iron and copper. Low pH cause the re-suspension of toxic metals from the sediments into the water column, negatively affecting aquatic species. pH levels fluctuate between day and night and between seasons. Carbon dioxide is removed from the water body by photosynthetic aquatic plants during the day, significantly increasing the pH.

2.2.3.4 Temperature

The temperature of freshwater body significantly affects biological and chemical processes in a water body. When the water temperature increases, the capacity of water to hold dissolved oxygen reduces and increases with a decrease in water temperature at a given pressure. The rate of plant photosynthesis, metabolic rates of aquatic organisms and, the sensitivity of organisms to toxic wastes, parasites and disease is also influenced by the water temperature (USEPA, 1997). Many aquatic organisms carry out certain important activities like reproduction and migration as a function of specific water temperatures. Most organisms need an optimum temperature to function at maximum and generally require the slow change in water temperature as seasons change to acclimate. Hence, any rapid shifts in water temperature will negatively affect aquatic plants and animals, with shifts of more than 1-2 °C being able to cause thermal stress and shock (Campbell and Wildberger, 1992). Thermal stratification of standing water governs certain critical processes occurring in the water body. The temperature of a water body varies with depth, season, and inflowing water temperature and wind effect. By collecting the water temperature at different depth, thermal layers for the water body are determine, which are necessary for interpreting other ecological indicators.

2.2.3.5 Total coliforms and fecal coliforms

The main pathogenic micro-organisms are bacteria, viruses and protozoan associated with fecal waste causing disease such as cholera, typhoid, giardiasis and hepatitis when contaminated water is ingested in one way or another. Directly testing for each pathogen is quite expensive and impracticable, hence, indicator species among which are total coliforms, fecal coliforms, E. coli and enterococci are used. Total coliforms constitute a group of very closely related genera of bacteria all sharing a useful diagnostic feature. A very fecal specific indicator is the subgroup of total coliform called fecal coliform. The coliforms bacteria inhabits the colon or lower intestine of poikilothermic organisms (wildlife, humans, pets and farm animals) and usually constitute as much as 50 % of fecal waste. Though not usually pathogenic in themselves, their presence is a good indication of sewage contamination, usually accompanied by disease causing organisms or pathogens.

2.3 Evaluation of water quality by standards

Both the ambient water quality of a water body and the release of pollutants to the water body (effluents) which influence the ambient water quality need to be controlled. There exist different water quality standards for different freshwater uses. For example, drinking, fishing and bathing water quality standard. They generally comprise two elements, namely, the ambient standard (maximum allowable concentration) and the effluent standards (maximum permissible discharge).

2.3.1 Ambient water quality standards

The ambient water quality standard provides the concentration of a given substance above which the water resource is unsuitable for a particular use. Above this concentration, human and ecosystem health are affected (Viessman and Hammer, 1998). The ambient water quality standard serves as a basis or reference point for assessing the current status of water quality and its suitability for various uses.

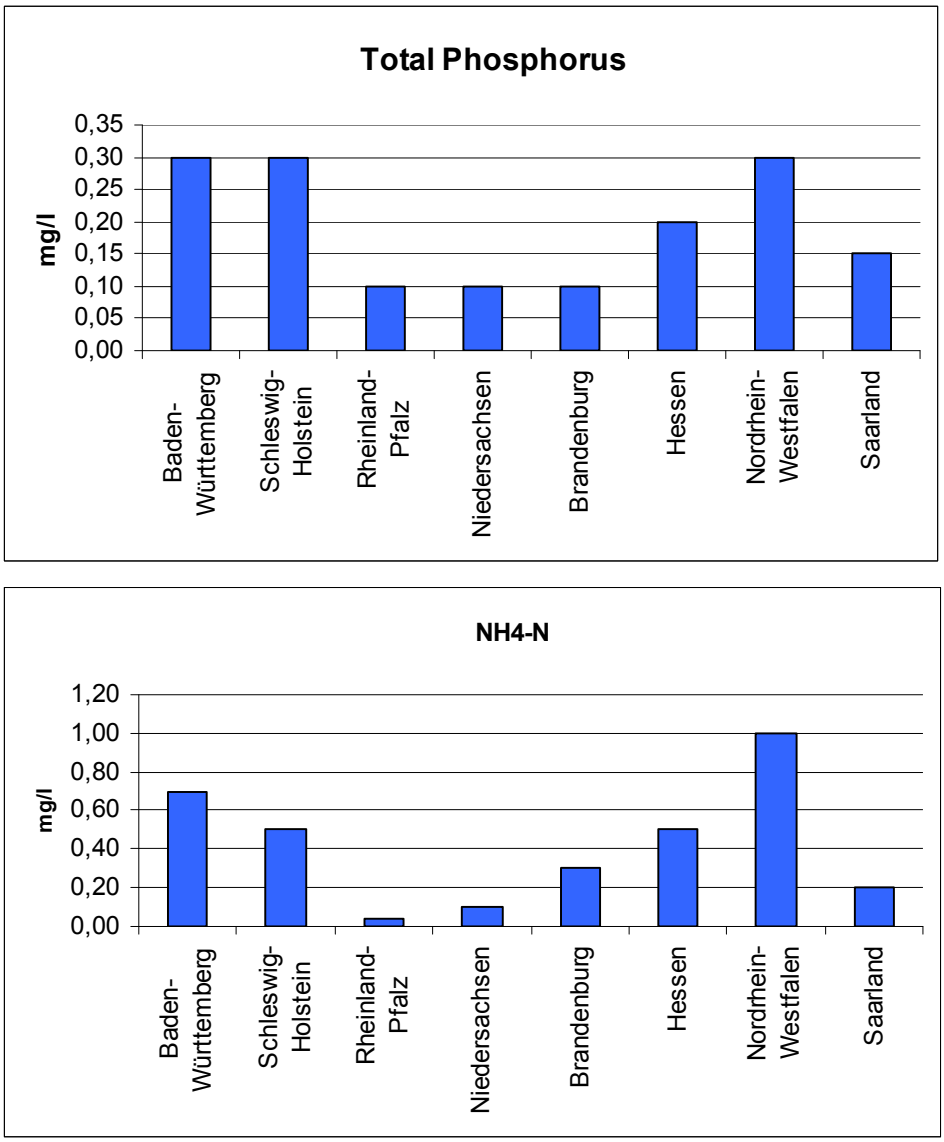

Figure 2.3: LAWA class II water quality standard for some states in Germany.

The European Union for example has different standards depending on the use assigned to a freshwater ecosystem (for example fishing, recreation, drinking water abstraction). The rationale for setting water quality standards is to ensure that the concentration of certain substances in freshwater ecosystems should have no direct or indirect harmful effect on human and ecosystem health. The standard assumes that concentrations below or above a certain threshold posses no threat or risk to human and ecosystem health under the designated

use of the freshwater ecosystem (Koren, 1991). Hence, the standard is established based on the designated use for the water resource.

In setting a realistic water quality standard, the climatic, geologic and land use characteristics of the watershed must be taken into consideration. In Germany for example, each state establishes the standard for substances based on the characteristics of the different watersheds of the state. Figure 2.3 illustrates the existence of great differences between the standards of the same substances from one state to another. Given that the watersheds of the different states possess different characteristics, it is not possible to have one standard for all water bodies in the country. In addition, economic motivations may also influence the establishment of the standards. A compromise usually has to be met between the economic growth of a region and environmental constraints.

2.3.2 Effluent water quality standards

Effluent limitations refers to the maximum amount of a pollutant that a polluter may discharge into a water body taking into consideration the effluent from all other source and is generally derived by the help of mathematical models (Viessman and Hammer, 1998). The appropriate effluent standard helps in the attainment of the ambient water quality standard for a given water body. The effluent standard may allow some discharge or no discharge of pollutants based on the type of pollutant (Koren, 1991). Generally, the discharge of certain pollutants like high level radioactive wastes, biological warfare materials and any radiological wastes into freshwater bodies are strictly prohibited. Such pollutants are able to cause death, disease, behavioural abnormality, cancer, genetic mutations, physiological malfunctions, etc, to humans and other organisms (Koren, 1991). In setting effluent standards, an adequate margin of safety must exist. Industries generally have to meet effluent limits that require or reflect the current best practical technology. Due to the present increasing challenges facing freshwater ecosystems as a result of anthropogenic pollution, zero discharge of pollutants should be the standard when possible (Barthelmes, 1981).

A number of freshwater quality indicators portray a cycling time structure (figure 2.4) as a result of parallel acting seasonal and periodic internal and external driving forces. Establishing an appropriate water quality standard without taking this cycling behaviour into consideration will render the standard inappropriate as a result of the water body complying and not complying to the standard in the lower and upper regions of the cycle respectively. In order to assess the compliance of a freshwater ecosystem to the established standard for a designated water use, monitored indicator data is compared to the standard. It is quite interesting to closely examine cycling water quality indicators and standards. Some water quality indicators, sampled at a daily interval, from the Havel River in the state of Brandenburg in Germany were compared against the LAWA (1998) water quality standards for the State of Brandenburg. The red lines in the figure represent the established ambient water quality standards for the State.

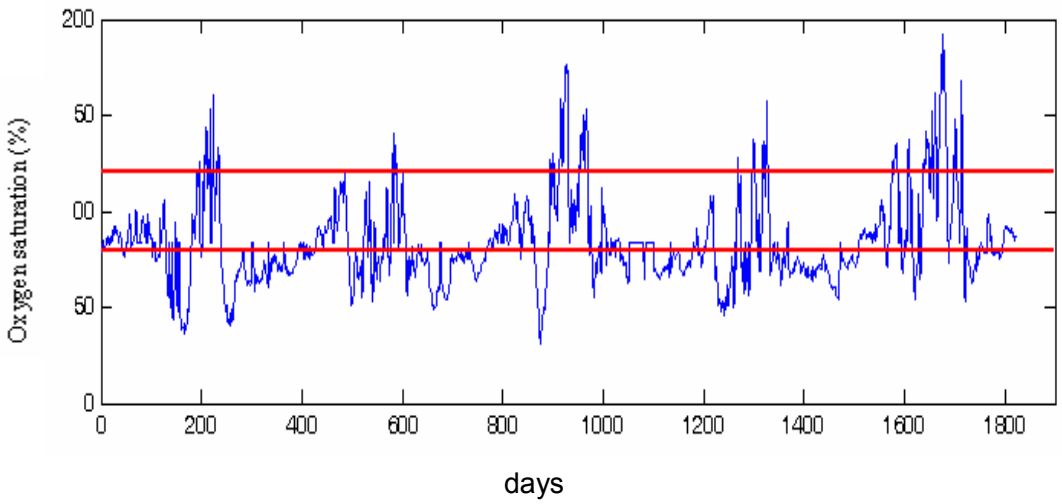

Figure 2.4: Water quality standard for oxygen saturation with an upper limit of 120 % and a lower limit of 80 % for the state of Brandenburg, Germany.

It can be observed from figure 2.4 that the lower limit for oxygen saturation is more or less the mean value (83.3) for this indicator of respiration and productivity. This water body is eutrophic with high oxygen production especially during the warm periods. There seam to be two cycles present which

are as a result of the two dominant algal species namely diatoms and cyanobacteria present in Spring and late Summer. Oxygen saturation is charcterized by high fluctuations which result in an abnormal situation of compliance and non compliance within an extremely short period of time, putting in question the appropriateness of the standard for this indicator.

Figure 2.5 presents the cycling behaviour exhibited by chlorophyll-a, an indicator of biomass or biological productivity which also causes some difficulties in assessing the compliance of this freshwater ecosystem to the required standard.

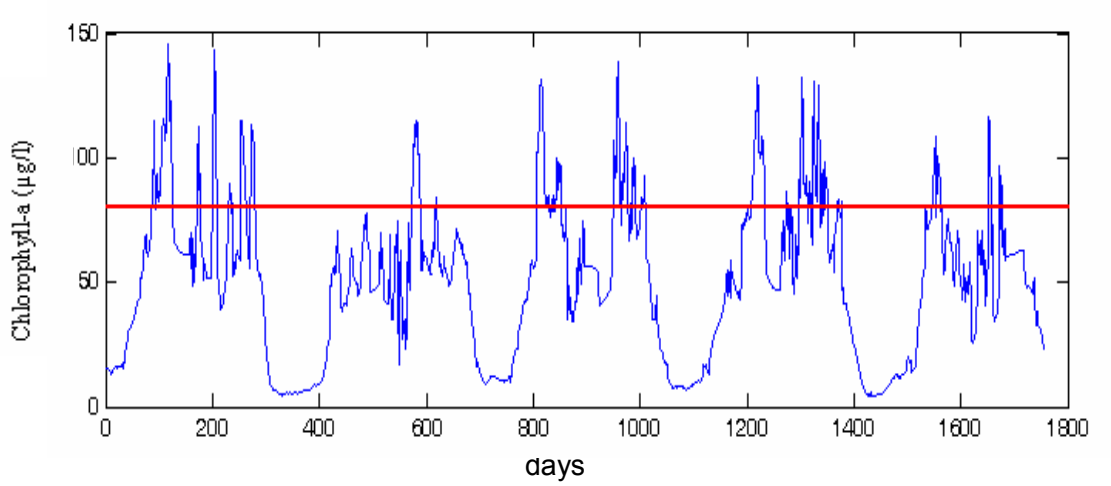

Figure 2.5: Water quality standard for chlorophyll-a with an upper limit of 80 µg/l.

There exist a longer cycle as a result of seasonal variations and a shorter cycle because of the presence of diatoms and cyanobacteria, the two dominant algal species appearing at different periods of the year. The result is that, the combined effects of these cyclic behaviors causes the water body to comply and not comply to the established standard within a relatively short period of time. It may be necessary for such an indicator to have different standards for different relevant seasons taking into consideration the cycling behaviour.

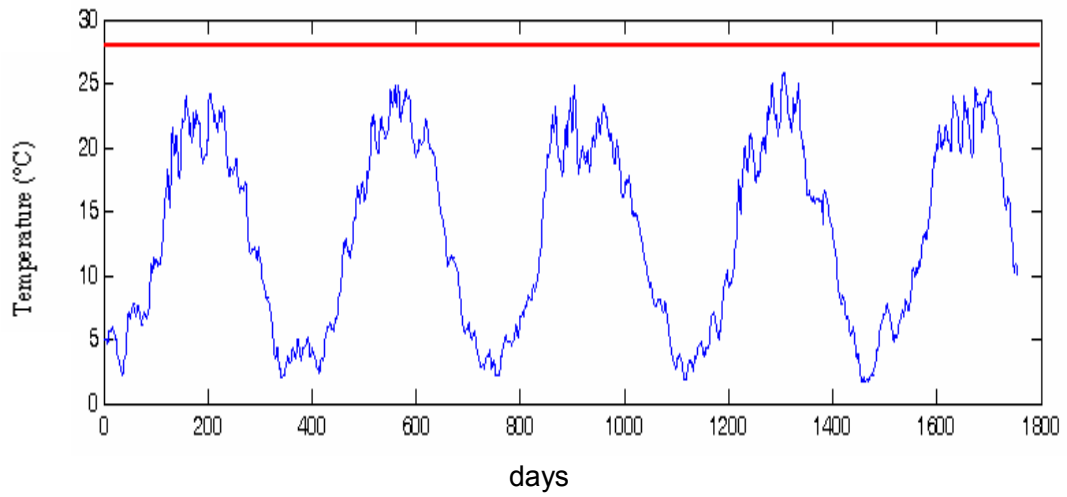

Figure 2.6: Water temperature with an upper limit of 28°C.

In Figure 2.6 the water temperature and its standard are displayed. The water body complies with the standard. The rationale for establishing such a high standard for water temperature is difficult to understand given that an increase in temperatures during Winter which comply to this standard is able to destroy the structure of the biota present. Temperatures of water bodies can usually be altered by a consistent input of warm water (cooling water) from thermal and power plants, which can even raise the water temperature by 8 – 11°C (Thanh and Tam, 1993). A more realistic standard for the different seasons of the year taking into consideration the biota present needs to be considered.

Figure 2.7 clearly indicates that the pH, an indicator of acidity and carbondioxide content of the water body, complies with the lower limit of the standard. The two cycles observed are driven by the two algal biomass present at different periods of the year which strongly influence the carbondioxide content of the water body. There however exist situations where the upper limit of the standard complies and do not comply within a very short period of time.

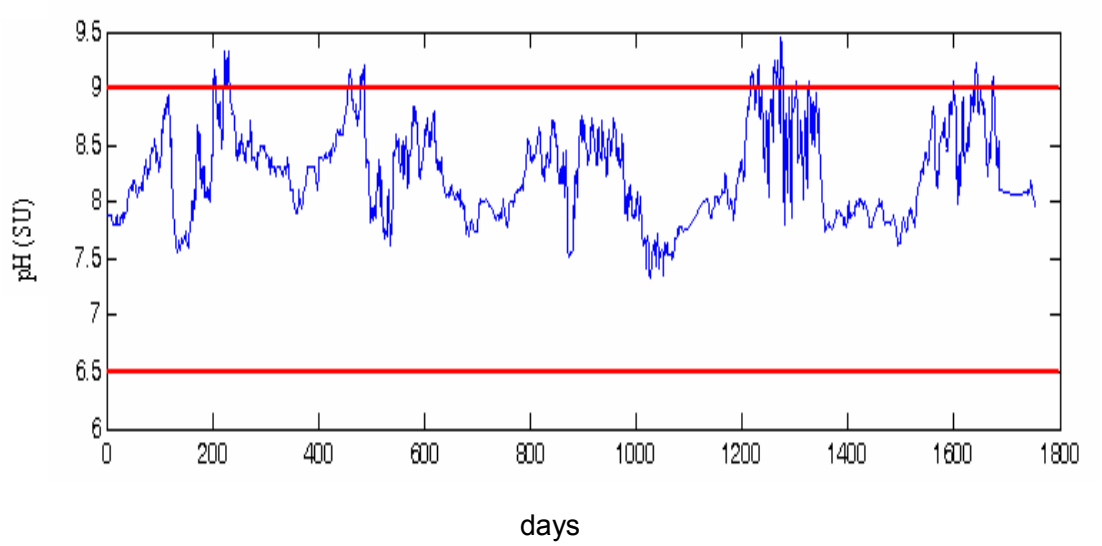

Figure 2.7: Water quality standard for pH with an upper limit of 9 standard units and a lower limit of 6.5 standard units (SU).

2.4 Pre-treatment of water quality time series

Freshwater quality data collected by monitoring programs is extremely important for assessing the condition of a freshwater body and developing an understanding of the interrelationship between the components of the freshwater body. Unfortunately, data from monitoring programs contain a lot of inconsistencies that affect the results of any analysis. Errors in water quality time series lead to some general problems in water quality research and simulation. They are as a result of missing values, impossible values, inconsistent values and unlikely values. They cause not only difficulties in process identification and parameter estimation but also misinterpretations of spatial and temporal variations of water quality indicators. Mostly, time series represent samples of data at discrete time events based on various sampling intervals. For modelling and simulation of freshwater ecosystem processes time series must be mapped on a regular time grid. This procedure is known as re-sampling of time series and consists on data interpolation or, in the case of disturbed signals, on data estimation. Some well-known linear and nonlinear interpolation methods exist while data estimation can be done by static and dynamic approximation procedures. In addition, most approaches for analysing

statistical data are linear methods which assume constant variance, a normal distribution and independence of the data set. Most often, water quality time series violate these assumptions. The signals are often not normally distributed and are either serially or spatially correlated or require non-linear models (Hunter, 1977, 1980; Hunter, 1982; Berthouex et al., 1981). For a water quality signal to be analyzed, the most common data distortions have to be rectified with acceptable methods like outlier removal, rendering the data equidistant, smoothing or filtering the data and applying transformations to the data set.

2.4.1 Preliminary data examination

Before any analysis is carried out on water quality time series data, it is essential to do some preliminary checks. This checks help reassure the analyst the data at hand is meaningful, helps detect missing data and errors in data entry, helps detect some patterns and ensure the assumptions of the analysis are met, helps detect departures from the assumptions and the possible transformation required for the data, helps detect unusual values termed outliers, etc. These methods were originally developed by John Tukey (1977) and were extended by Haoglin et al. (1983).

2.4.1.1 Run sequence plot

The run sequence plot is reveals the behavior of the signal across time. This is the first step in analyzing any signal. A look at the signal gives an idea about the general tendency of the signal over time like the presence of a trend and any periodic or seasonal behavior. Figure 2.8 is an example of a run sequence plot of conductivity sampled at daily intervals from 1998 to 2002 from the River Havel.

At a glance, the run sequence plot of a signal is able to reveal the existence of seasonality, an upward or downward trend, exponentially changing variance and other tendencies which need to be further verified by more rigorous methods. Some of these tendencies cause the data set to violate the assumptions on which many statistical or time series analysis techniques

depend. A technique like the Box-Jenkins time series modeling approach requires the data to be stationary.

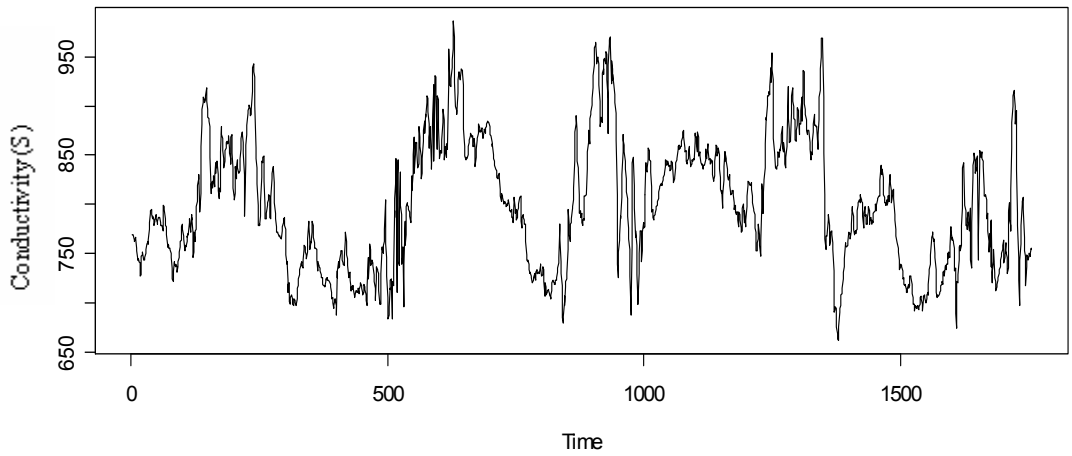

Figure 2.8: Run sequence plot of daily values of conductivity.

2.4.1.2 Box plot

The box plot is an effective method for displaying a five number data summary. Also referred to as the box and whisker plot, it summarizes the median, upper and lower quartiles, minimum and maximum values (Berthouex and Brown, 1994). In other words, it gives the median, variability of the data around the median, skew of the data, range of the data and the size of the data set. The box plot may be represented horizontally or vertically as shown in figure 2.9. The box contains the middle 50% of the data with the upper edge of the box showing the 75^{th} percentile of the data and the lower edge indicating the 25^{th} percentile. The interquartile range is the range of the middle two quartiles. The median value is represented by the line in the box. When the median line is not equidistant from the edges, it is an indication that the data is skewed. The whiskers or end of the vertical lines shows the minimum and maximum values. In the presence of outliers in the data set, they may extend to a maximum of 1.5 times the interquartile range. Any point outside the edges of the whiskers are considered or suspected to be outliers. The main advantage of the box plot

is its ability to display the location and spread of the data set at a glance. It gives the information on the symmetry or skewness, shows outliers and enables the comparism of different data sets when laid side by side.

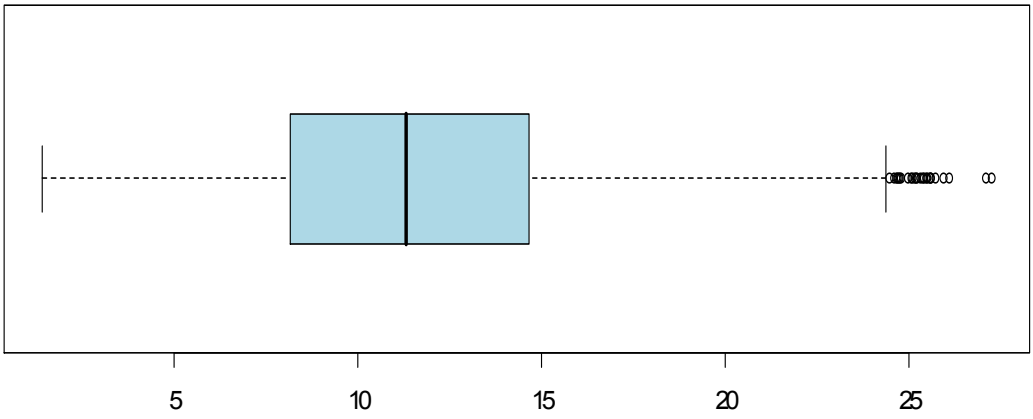

Figure 2.9: Box plot of turbidity.

2.4.2 Outlier removal and data transformation

An outlier is considered as an observation that lies at an abnormal distance from other values in a random sample from a population. Often, some values stand out from the general trend and may be as a result of gross errors in sampling measurement, or may be due to mistakes in recording the data. The analyst must be careful not to think in a too simplistic manner and make considerable errors by simply discarding such values without closer and critical examination. As an example, some early observers of stratospheric ozone concentration failed to detect the hole in the ozone layer because their computers had been programmed to screen incoming data for outliers. Therefore, the values that defined the hole in the ozone layer were discarded (Berthouex and Brown, 1994). Graphical techniques such as the box-plot, histogram and scatter plots can enable the analyst to easily distinguish between normal observations and abnormal ones. Outliers and temporal change in the value of an indicators like in the case of a storm (the value may

be associated to an outlier) can cause distortion in the location and variance of the data, bias estimates and result in poor models. The outliers should be taken out prior to analysis or modelling or dealt with by means of an appropriate transformation like the power, logarithmic or square root transformation. In many situations, values which are above 2 standard deviations are considered as outliers and taken out. This is not possible when dealing with cyclic processes where much care has to taken when extracting outliers.

Transformations are used to eliminate distortions that may violate statistical assumptions on which the statistical analysis method is based. The technical reason for data transformation are to make the spread equal in different data sets (uniform variance) make the distribution of residuals normal and to make the effects of treatments additive (Box et al., 1978). Variance stabilisation transformations such as the square root and logarithm are used to obtain constant variance. The effect of square root and log transformations is to make the larger values less important relative to the smaller ones. For the log transformation given by

$$x = \log(y)$$

the range of the transformed variables is relatively smaller than for the square root given by

$$x = \sqrt{y}$$

The Box-Cox power transformation can enable the analyst to transform a data so as to simultaneously satisfy the conditions of normality and constant variance. This transformation is applicable for almost any kind of statistical model and any kind of transformation and is given by

$$y_i^{\lambda} = \frac{y_i^{\lambda} - 1}{\lambda \bar{y}_g^{\lambda-1}}$$

where $\bar{y}_g$ is the geometric mean of the original data, λ is the power of the transformation. When λ is 0, we obtain the log transformation, λ= -1 gives the

reciprocal transformation, $\lambda = ½$ gives a square root transformation and $\lambda \neq 1$ is no transformation (Box et al., 1978).

For linearizing a signal, transformations using reciprocals, ratios, or logarithms are used. The use of transformations should be done with care because they can turn a good situation into a bad one like the case where the signal is linearized but the variance becomes unequal. There should be a good statistical justification for making transformations and then analysing the transformed data. They may be used for variance stabilisation, make the distribution of the errors normal or effects additive.

2.4.3 Equidistancy of data

Time series methods require equidistant time series data. Unfortunately, many water quality time series are not equidistant as a result of missing data. The missing data can be replaced by interpolation methods such as linear, spline, nearest neighbour, and cubic spline methods. However, long gaps in a signal cannot be replaced by interpolation methods, but require approximation techniques such as the autoregressive, moving average, autoregressive integrated moving average and seasonal autoregressive integrated moving average method. The idea behind approximation techniques is to fit an appropriate model to the data and use the model to estimate the missing data.

2.4.3.1 Interpolation methods

Missing data within a given signal creates a complicated problem (Little and Rubin, 1983; 1987). The process of estimating the outcome in-between sampled data points is called interpolation. There are a number of interpolation methods which are available for reconstructing freshwater quality time series among which are the nearest neighbour interpolation, linear interpolation, cubic hermite interpolation, spline interpolation and Fourier interpolation. Investigations of different interpolation method for selected measuring point of the Rivers Spree, Dahme, Elbe, Havel, Oder and Teltow Channel by Gnauck and Luther (2004) enabled the linear interpolation to be chosen as the most

appropriate approach for replacing missing data for most indicators in these freshwater bodies. The interpolation algorithms used for the different interpolation techniques are presented in table 2.2.

Table 2.2: Algorithms of some interpolation methods (Gnauck and Luther, 2004).

Method	Equation	Characteristic of β
Linear	$\beta(t) = \begin{cases} x_k & t<(t_k+t_{k+1})/2 \\ x_{k+1} & t\geq(t_k+t_{k+1})/2 \end{cases}$	$\beta \notin C^{(0)}[t_0,t_n]$ (β discontinuous)
Nearest neighbour	$\beta(t) = \frac{x_{k+t} - x_k}{t_{k+1} - t_k}(t + t_k) + x_k$	$\beta \in C^{(0)}[t_0,t_n]$ (β continuous)
Cubic spline polynomial	$\beta(t)\Big\|_{[t_k,t_{k+1}]} = a_k t^3 + b_k t^2 + c_k t + h_k$	$\beta \in C^{(1)}[t_0,t_n]$ (β continuous differentiable)
Cubic Hermite polynomial	$\beta(t)\Big\|_{[t_k,t_{k+1}]} = e_k t^3 + f_k t^2 + g_k t + h_k$	$\beta \in C^{(2)}[t_0,t_n]$ (β continuous, two times continuous differentiable)

2.4.3.2 Approximation methods

The main idea behind approximation methods is to fit a model to a time series data and use the model to predict the missing values. As with any modelling problem, selecting the most suitable model is quite necessary because the better the model, the better the prediction used to replace the missing values. There are a number of parametric and nonparametric approaches for modelling time series data with the most common parametric approach for univariate time series modelling being the Box-Jenkins approach (Jones, 1980; Grifoni and Passerini, 2004). An example of using the method is presented. The Fourier polynomial can also be used for approximating stationary and periodic time

series. The main nonparametric approaches are artificial neural network and fuzzy logic with the main approximation methods shown in figure 2.10.

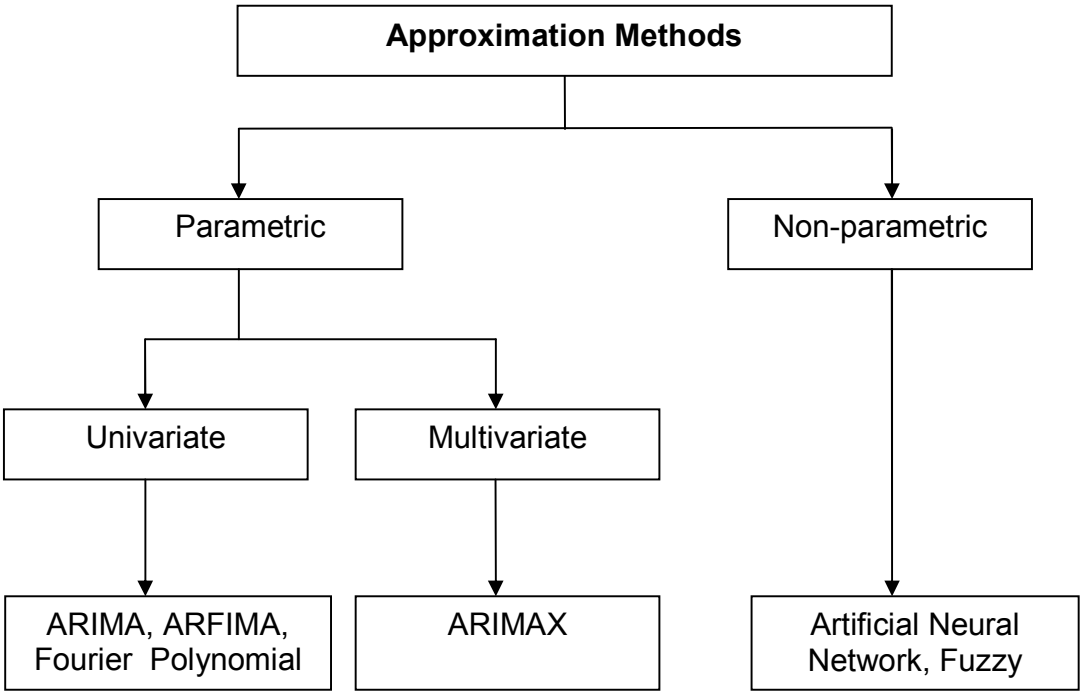

Figure 2.10: Approximation methods (adapted from Grifoni and Passerini, 2004).

The Box-Jenkins method

This approach combines the moving average model (MA) and the autoregressive model (AR) which gives rise to the autoregressive moving average model (ARMA). The time series must be stationary (having a mean and variance that remains constant across time).

An autoregressive model is a linear regression of the current value of the series against one or more prior values of the series.

$$X_t = \partial + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots + \Phi_p X_{t-p} + A_t$$

$$\partial = \left(1 - \sum_{i=1}^{p} \phi_i\right)\mu$$

Where X_t is the time series A_t is the white noise, μ is the mean of the process and the value of p is the order of the autoregressive process.
A moving average model is a linear regression of the current value of the series against the white noise or random shock of one or more prior value of the series.

$$X_t = \mu + A_t - \varnothing_1 A_{t-1} - \varnothing_2 A_{t-2} + \ldots + - \varnothing_q A_{t-q}$$

where X_t is the time series, μ is the mean and A_{t-i} are white noise, $\varnothing_1, \ldots, \varnothing_q$ are parameters of the model. The value of q is the order of the moving average model.
Box and Jenkins made popular an approach which combines the MA and the AR models by developing a methodology for identifying and estimating the model.

$$X_t = \partial + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots + \Phi_p X_{t-p} + A_t - \varnothing_1 A_{t-1} - \varnothing_2 A_{t-2} + \ldots + - \varnothing_q A_{t-q}$$

where the terms of the equation have the same meaning like in the AR and MA models. P is the AR order and q is the MA order which together give rise to ARMA (p, q). The Box-Jenkins approach assumes that the time series is stationary. Box-Jenkins recommends the differencing of non-stationary time series in order to achieve stationarity. This differencing give rise to the ARIMA model with the “I” meaning integrated (Box et al., 1994). The Box-Jenkins approach can be extended to include a seasonal MA and a seasonal AR component.

Three important steps are necessary for constructing the Box-Jenkins model, namely model identification, model estimation and the model diagnosis. The first step which can be quite subjective depending on the chosen method is the determination of the integers p, d and q. These represent the autoregressive order, the differencing order and the moving average order of the ARIMA model respectively. This may entail selecting the structure and order of model from the appearance of the plotted autocorrelation plot and the partial autocorrelation plot making it quite subjective. This process may involve much trial and error. Automated techniques like the final prediction error and the Akaike information

criterion (AIC) automate the model identification process rendering it more efficient and objective (Akaike, 1969; 1974). Model estimation entails estimating the coefficients of the model that has been identified.

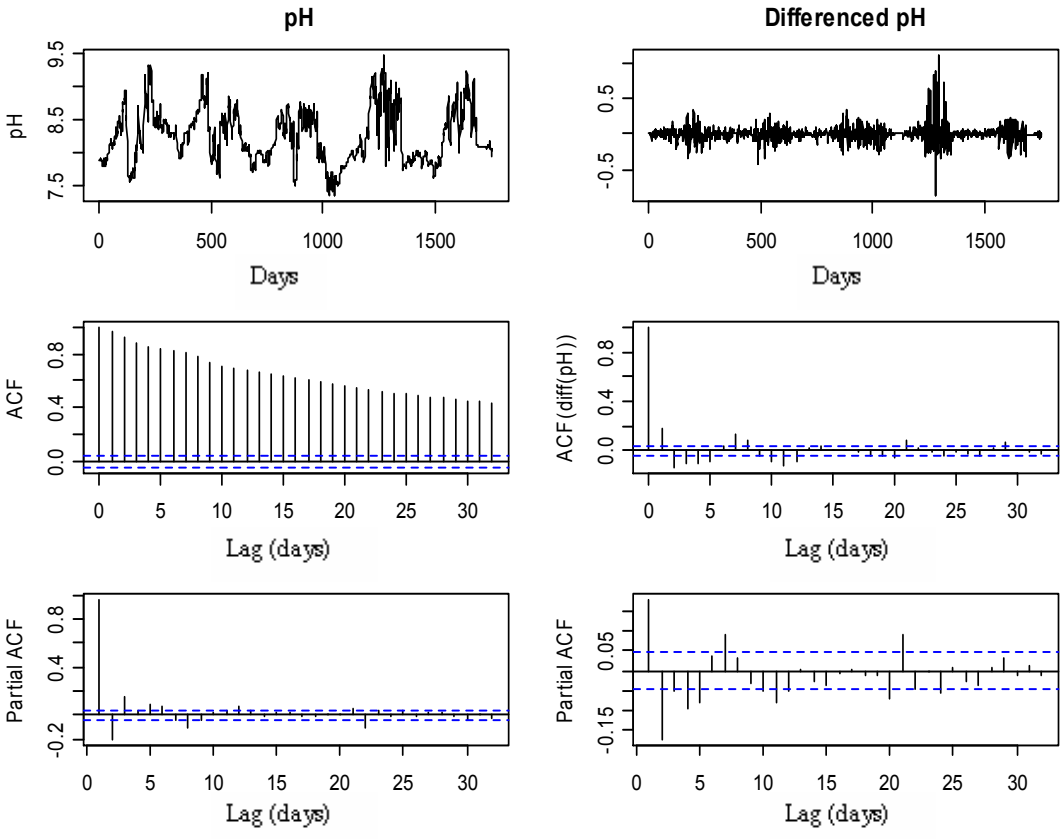

Figure 2.11: Original and differenced pH signal, ACF and PACF.

The parameters p, d, and q are given as input and an appropriate software tool like SAS or SPSS which performs the iterative calculations needed by means of the maximum likelihood or the non-linear least square estimation method. The ability to diagnose how well a model fits the experimental data is a vital part of the analysis. If the BJ model is a good fit for the data, the residuals should be random, with fixed distribution, location and scale. A plot of the residuals and autocorrelation of the residuals can be used. As an example, pH time series sampled at daily interval was used to investigate this approach. The AIC in

table 2.3, the differenced (to render it stationary) ACF and PACF in figure 2.11 respectively were used to select the parameters of the model.

Table 2.3: AIC values of the different ARIMA models.

ARIMA model	AIC value
(1,1,0)	- 3133,5
(2,1,0)	- 3184,359
(2,1,1)	**- 3198,566**
(2,1,2)	- 3197,878
(0,1,1)	- 3155,534
(0,1,2)	- 3174,238

The ARIMA (2, 1, 1) is the most appropriate from among the set of investigated models because it has the lowest AIC and is confirmed from the graphical model diagnostic techniques. The coefficients of the model ARIMA (2, 1, 1) with AIC = - 3198.57 are displayed in table 2.4 with figures 2.12 and 2.13, exhibiting the approximation and the model diagnostic plots respectively.

Table 2.4: AIC values of the different ARIMA models.

Model	Coefficient	Standard error
ar1	0.7276	0.0814
ar2	-0.2604	0.0235
ma1	-0.5371	0.0832
Sigma square	0.009419	

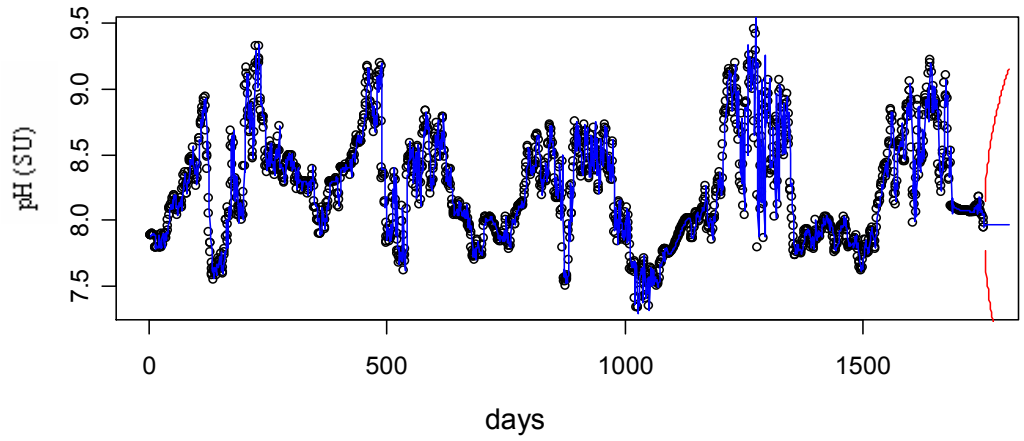

Figure 2.12: pH signal and its ARIMA(2, 1, 1) approximation with a forecast of 50 periods ahead.

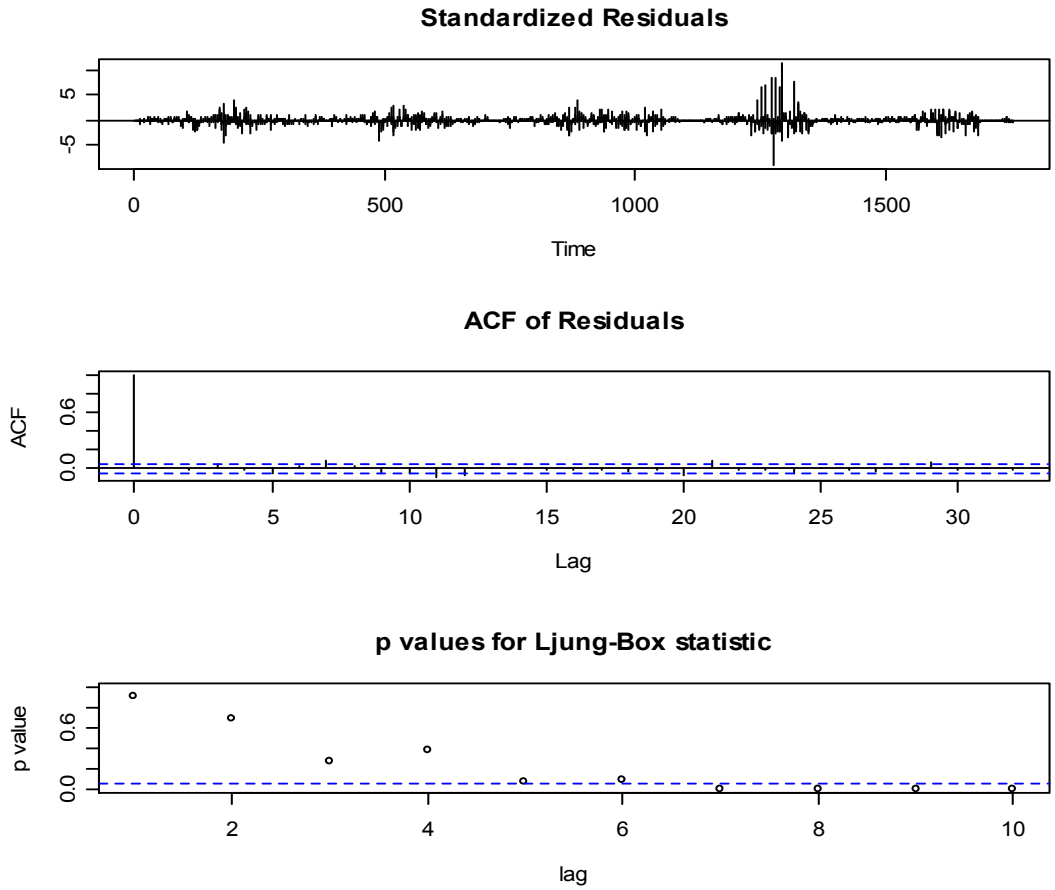

Figure 2.13: Model diagnostics plot indicates that the residuals are random, the ACF of the residual is zero and the p value for the Ljung-Box statistics falls to zero indicating the model ARIMA (2, 1, 1) fits that data.

Pre-treatment of water quality signals is extremely important because the quality of the results of any data based analysis depends on the quality of the data used. It is dangerous when the quality of the data series bad, because management decisions will be taken based on wrong result. Rather than enhancing the ability of water quality managing to make sound decisions, it may rather enhance poor decision making. When monitored data is pre-treated so as to improve on the quality, robust methods are required to extract the information contained in the ecological signals. The following chapter will present signal analysis methods, most of which will be used for investigating chosen water quality signals.

3 Methods of Signal Analysis

Analyzing freshwater quality signals so as to develop a better understanding of the changes occurring in these ecosystems is essential for sustainably managing them. Signal analysis methods are a set of techniques used for extracting information contained in signals which may be contaminated with noise or other signals or not (Smith, 1997). Two classical approaches exist for the analysis of time series data namely the time and frequency domain methods. In the time domain approach, values of the signal, recorded as a function of time are analyzed by means of techniques such as the autocorrelation, correlation, cross-correlation functions, etc. Unfortunately, time domain methods give no information concerning the frequency at which the changes in the signal occur. Given that many natural systems have frequency dependent variability, an understanding of this frequency dependence gives more information concerning the underlying physical mechanism that produced the signal. The Fourier transformation is used to project a signal from the time domain into the frequency domain so as to reveal periodic components present in the signal, the active frequency band in the signal and their intensity or relative importance (Bloomfield, 1976; Percival, 1995). The Fourier analysis therefore provides an effective tool for examining the variability of a signal as a function of frequency by means of the periodogram and cross-spectral analysis, but gives no information concerning the time at which the different frequencies occur. Wavelet analysis is a modern mathematical tool used for the analysis of a signal both in the frequency and time domain. Wavelet analysis methods act as a lens for investigating the characteristics of a signal on a scale by scale basis as well as the interrelationship between signals at different time scales (Mallat, 1989; McCoy and Walden, 1996; Hess and Wickerhauser, 1996). Such investigations bring about new insight and a deeper understanding of many complex processes and cause-effect relationships.

This chapter first presents the main time domain approaches used for analyzing univariate and bivariate time series with special regards to freshwater quality indicators. Techniques such as the autocorrelation, correlation, covariance, cross-correlation, etc. are presented. In addition, the frequency domain approaches such as the periodogram analysis, spectral analysis and cross-spectral analysis - coherency and phase - are examined. Finally, wavelet methods such as the multiresolution analysis, wavelet variance and covariance, correlation and cross-correlation useful for investigating water quality signals are examined. In addition to the above mentioned methods, there are other signal analysis methods such as the Hilbert transforms, the Wigner-Ville transforms, the short time Fourier transform and others as presented in table 3.1 which will not be on display.

Table 3.1: Some main signal analysis methods and formulae.

Method / domain	**Formula**
Time domain Autocorrelation	$r_j = \frac{\sum_{t=1}^{N-j}(Y_t - \bar{Y})(Y_{t+j} - \bar{Y})}{\sum_{i=1}^{N}(Y_t - \bar{Y})^2}$
Covariance	$Cov_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{N-1}$
Cross-correlation	$\rho_{xy}(\tau) = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_{i+\tau} - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_{i+\tau} - \bar{y})^2}}$
Frequency domain Periodogram	$I(f_i) = \frac{2}{N}\left\{\left[\sum_{k=1}^{N} y(k)\cos(2\pi f_i k)\right]^2 + \left[\sum_{k=1}^{N} y(k)\sin(2\pi f_i k)\right]^2\right\}$

Spectral analysis	$\overline{S_y}(f_j) = \sum_{l=-L}^{L} b_l \hat{S}(f_{j+l})$
Coherency	$F_{xy}(\omega)^2 = \frac{\left\|f_{xy}(\omega)\right\|^2}{f_x(\omega) f_y(\omega)}$
Phase	$\varphi_{x,y}(\omega) = \arctan\left[\frac{\mathrm{Im}\, f_{xy}(\omega)}{\mathrm{Re}\, f_{xy}(\omega)}\right]$
Gain	$G_{xy}(\omega) = \frac{\left\|f_{xy}(\omega)\right\|}{f_x(\omega)}$
Fourier polynomial	$F_n(t) = a_0 + \sum_{k=1}^{k=n} (a_k \cos(kt) + b_k \sin(kt))$
Discrete cosine transform	$X_k = \frac{1}{2}\left(x_o + (-1)^k xN - 1\right) + \sum_{n=1}^{N-2} x_n \cos\left[\frac{\pi}{N-1} nk\right]$ K=0, … N-1
Hilbert transform	$\overleftrightarrow{h}(t) = \int_{-\infty}^{\infty} s(\tau) h(t-\tau) d\tau$ where $h(t) = \frac{1}{\pi t}$
Time-frequency Short time Fourier transform	$F(\omega, \Delta) = \int w(t-\Delta) f(t) e^{-2\pi w t} dt$
Wigner- Ville transform	$W_{s(\tau,f)} = \int_{-\infty}^{\infty} x(t+\frac{\tau}{2}) x^*(t-\frac{\tau}{2}) e^{-i2\pi f t} dt$
Time-scale Wavelet analysis	$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right)$

3.1 Time domain methods

The detection and description of the underlying patterns and structure of water quality signals need to be done in a rigorous manner so as to develop decent models which explain these patterns. Some techniques are available for investigating the behavior of a signal in the time domain. The autocorrelation function, the covariance, correlation, cross-covariance and cross-correlation are the main tools available for bivariate signal analysis in the time domain (Hipel and McLeod, 1994).

3.1.1 Autocorrelation

The autocorrelation of a signal refers to the correlation of the signal with its past and future values. In other words, it is a method for characterizing the correlation within a time series over time. It is given by

$$r_j = \frac{\sum_{t=1}^{N-j}(Y_t - \bar{Y})(Y_{t+j} - \bar{Y})}{\sum_{i=1}^{N}(Y_t - \bar{Y})^2}$$

where the measurements are Y_1, $Y_{2,}$ . . ., Y_N and k is the time lag.

The data series must be equidistant. The autocorrelation function measures the correlation between two values of the same variable at times X_i and X_{i+k}. The autocorrelation function of a signal is used to detect non-randomness in the data and for the identification of an appropriate model for a non-random data (AR, MA, ARIMA, SARIMA models) (Box et al., 1994; Vanables and Ripley, 1994). For the detection of non-randomness of the data, the first lag (lag 1) is of interest. For the identification of an appropriate model for a time series, the autocorrelation for many lags are plotted. A correlogram is a plot of the autocorrelation function against the lag *k*. when the data is random, the correlogram shows values which rapidly decay to zero. In the situation where the data contains cyclic and stochastic components, the correlogram will exhibit a cyclic movement with the period of the cycle being given by the distance between two successive points where the curve cuts the x-axis. Correlograms

are very practical for the determination of the dependence between successive observations of a time series. If the correlogram indicates the existence of correlation between successive terms x(t) and x(t + *k*), the signal is assumed dependent or said to exhibit long memory (Beran, 1994). A strong autocorrelation or a time series that exhibits long memory behavior can complicate the identification of any significant covariance or correlation between signals. Figure 3.1 represents the autocorrelation of conductivity.

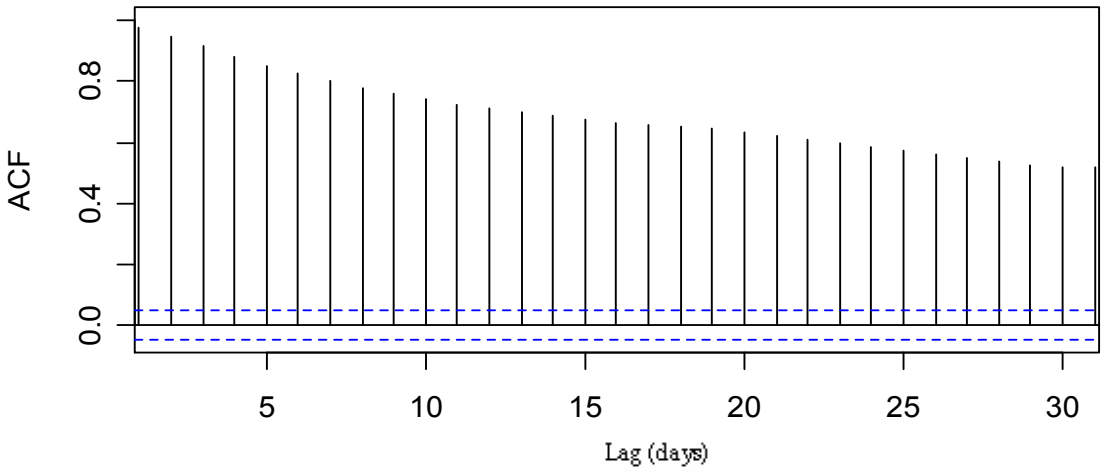

Figure 3.1: Correlogram of conductivity.

The behavior of the autocorrelation function can also be used for the identification of Box-Jenkins models. When exponential, decaying to zero or shows alternating positive and negative, decaying to zero, it indicates an autoregressive model. When it contains one or more spikes, it indicates a moving average model and when close to zero, the data are essentially random. Periodic high value indicates a seasonal autoregressive term (Haan, 2002). When there is no decay to zero, it means the data is not stationary.

3.1.2 Covariance and correlation

The most common statistical techniques for measuring the strength of the relationship between two signals are the covariance and correlation. Measuring

the strength of a linear relationship between two continuous random variables implies determining how much the two variables covary or vary together. If one variable increases (or decreases) as the other increases (or decreases), then, the two variables covary. If on the other hand one variable does not change as the other variable increases (or decreases), then the variables do not covary. The covariance measures how much two variables covary in a sample of observations. This is given by

$$Cov_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{N-1}$$

The covariance ranges from -∞ to +∞. One limitation of the covariance as a measure of the strength of a linear relationship is that its absolute magnitude is dependent on the units of the two variables. For example, the magnitude will be larger if the measurements of a study are done expressed in grams rather than in kilograms. Figure 3.2 reveals the covariance of conductivity with water temperature exhibiting a relatively high value as a result of the high magnitude of the values of the original conductivity time series units. The cloud reveals areas where the data series are present with the darkest portions indicating the strongest clusters of the two signals. The most significant clusters exist during winter and summer.

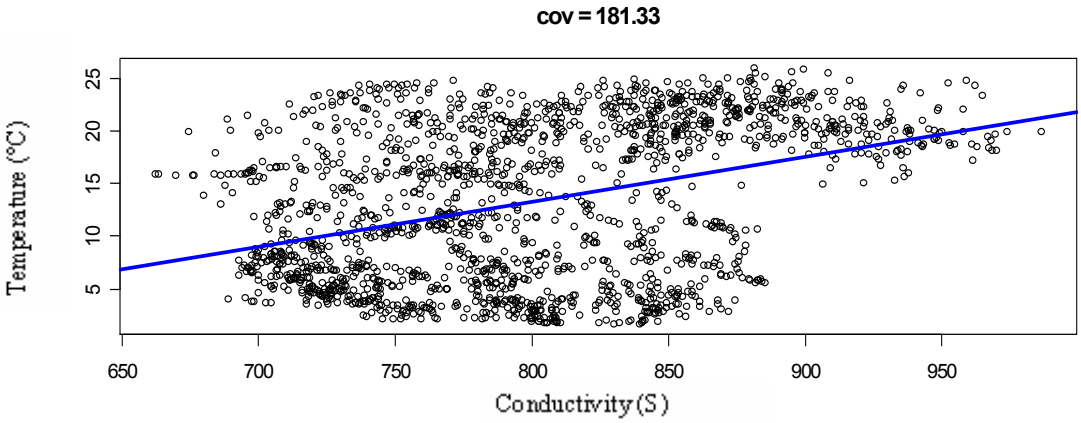

Figure 3.2: Covariance of conductivity and water temperature.

The covariance can be standardized by dividing the standard deviations of the two variables so that our measure of the strength of the linear relationship lies between -1 and +1. This is known as the Pearson product-moment correlation and measures the strength of linear relationships. The correlation coefficient measures the strength of the linear relationship between two signals. Correlation analysis serves as a tool for describing the degree to which one signal is linearly related to another and is also used as a measure of the degree of association between two variables. It is given by

$$\rho_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

When the relationship is curvilinear, the true strength is not measured by the correlation coefficient. A high value of the correlation coefficient between two signals simply indicates that departures from the mean value at time *t* in the two signals are of the same sign. A value of positive one implies the two variables have a perfect and increasing relationship while a value of negative one implies that the two signals have a perfect but decreasing relationship.

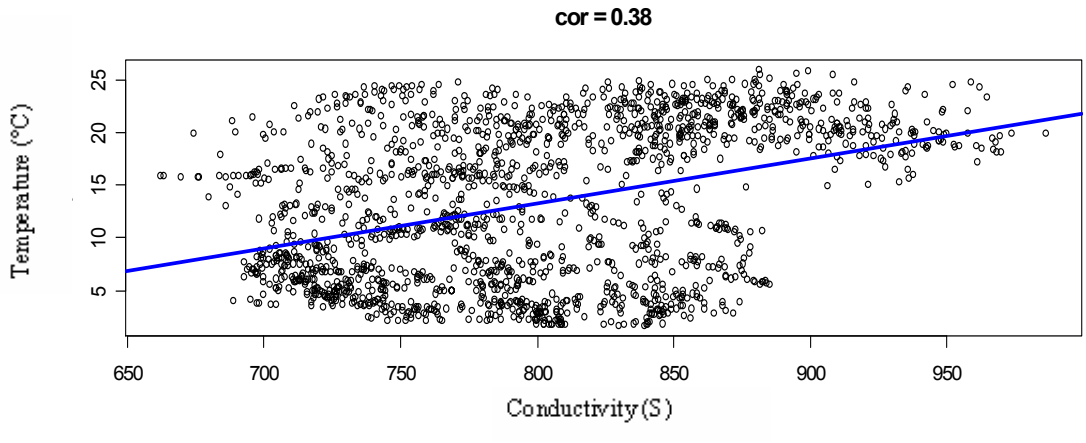

Figure 3.3: Correlation of conductivity and water temperature.

Figure 3.3 shows the correlation of conductivity and water temperature, a normalized version of the covariance and the value, when compared to that of the covariance, it seams quite low. This normalization enables correlations between indicators to be compared with others since they vary between 1 and -1 contrary to their covariances.

3.1.3 Cross-covariance and cross-correlation

The cross-correlation function between two signals can be a useful tool in determining the amount of linear dependence between two signals at different time lags. Often, one would like to measure the predictability of another signal Y_t from the signal X_t, leading to the notion of cross-covariance. The scaled version of the cross-covariance is called the cross correlation. Hence, the lagged-correlation is the correlation between two signals at different time lags. This is useful because at times, departures at a given time in a signal are related to departures at a different time in another signal. The cross-correlation function which is the correlation as a function of lag is used to investigate the lagged relationships which are characteristic of many physical systems. Given two signals X_t and Y_t, the sample cross-covariance function is given by

$$C_{uy}(\tau) = \frac{1}{N}\sum_{t=1}^{N-\tau}(u_t - \bar{u})(y_{t+\tau} - \bar{y})$$

for k = 0, 1, …, (N - 1)

where N is the series length, ū and ŷ are the sample means, and k is the lag. The sample cross-covariance and cross-correlation scaled by the variances of the two signals is given by

$$r_{xy}(\tau) = \frac{c_{xy}(\tau)}{\sqrt{c_{xx}(o)c_{yy}(o)}}$$

where $C_{xx}(O)$, $C_{yy}(O)$ are the sample variances of X_t and Y_t.

Contrary to the autocovariance function and the autocorrelation function which are symmetric, the cross-covariance function and the cross-correlation function are asymmetric (value of lag *k* not equal to the value at lag -k). Hence, the first part represents Y_t lagging X_t and the second part applies to X_t lagging Y_t. Figures 3.4 and 3.5 give the cross-covariance and cross-correlation of conductivity with water temperature respectively.

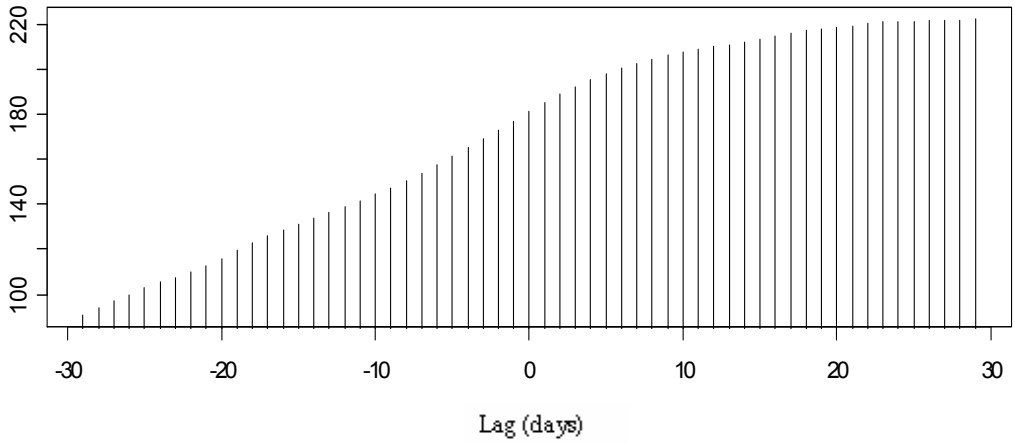

Figure 3.4: Cross-covariance of conductivity and water temperature.

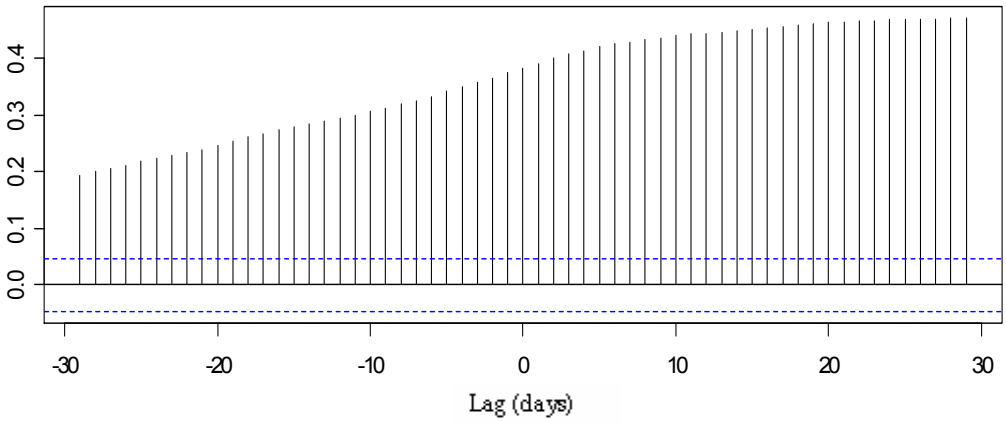

Figure 3.5: Cross-correlation of conductivity and water temperature.

These time domain techniques which have been presented are quite useful in extracting the time varying behavior of the signal, but give no information on the

frequency content or the active frequencies in the signal as well as the behavior of the signal at different time scales.

3.2 Frequency domain methods

In 1807, the French mathematical physicist, Jean Baptist Joseph Fourier developed a technique of expressing a function by superposing sine and cosine sinusoidal functions (Granger and Engle, 1983; Smith, 1997). The Fourier analysis makes use of sine and cosine functions used as basis functions and are most suitable for periodic and stationary signals. The frequency domain analysis exposes the frequency components active in a signal as well as the relationship between signals at certain frequency bands. The univariate spectral analysis examines a signal in the frequency domain, exploring cyclical patterns by a representation of the signal in terms of a linear combination of sinusoids of different frequencies and amplitudes. The periodogram and spectral analysis are the main tools for univariate frequency domain analysis. The bivariate spectral analysis enables the determination of correlations between signals at different frequencies. The cross spectral analysis is the main tool for studying the co-variation between signals as a function of frequency and consists of the squared coherency and the phase (Koopmans, 1974; 1983; Rebecca, 1998).

The Fourier polynomial provides a means of approximating periodic functions by sums of sine and cosine functions, shifted and scaled. This polynomial is an expression of the form

$F_n(t) = a_0 + a_1\cos(t) + b_1\sin(t) + \ldots + a_n\cos(nt) + b_n\sin(nt)$

which may be rewritten as

$$F_n(t) = a_0 + \sum_{k=1}^{k=n} (a_k \cos(kt) + b_k \sin(kt))$$

with

$$a_k = \int_0^{2\pi} f(t)\cos(kt)dt$$

for k = 0, 1, 2, …

$$b_k = \int_0^{2\pi} f(t)\sin(kt)dt$$

for k = 0, 1, 2, …

The integer k gives the frequency of the sine and cosine function; hence large values of k correspond to very wriggly graphs. The numbers a_k and b_k represent the amplitudes. The constants a_0, a_i and b_i, *i =1, . . ., n* are called the coefficients of the $F_n(t)$ and the Fourier polynomial is 2π periodic. By making use of a trigonometric identity called the Euler identity which is given by

$$e^{ix} = \cos(t) + i\sin(t)$$

the Fourier transform is derived and is given by

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-iwx}dt$$

This transform projects a signal from the time domain to the frequency domain and serves as the basis for frequency domain analysis.

3.2.1 Periodogram analysis

One of the simplest and most useful procedures which make use of the Fourier approach is the periodogram, introduced by Schuster in the nineteenth century. The periodogram is the frequency domain counterpart of the autocorrelation in the time domain, which provides information on the behavior of a time series over time with special regards to the memory of the process or how it is at one instance of time dependent on or related to the process at some prior time. The periodogram attempts to quantify the variability in the time series in terms of repeating patterns having fixed frequencies or periods. It graphs the spectral density by representing the intensity as a function of frequency (Parzan, 1983).

In other words, it gives the spectrum of a stationary signal which is a distribution of the variance of the signal as a function of frequency or partitions its variance among all possible frequencies so that the predominant frequencies can be identified. The frequency components that account for the largest share of the variance are revealed with each peak representing the part of the variance of the signal that is due to a cycle of a different period or length (Jenkins and Watts, 1968; Koopmans, 1995). The spectrum thus derived contains no new information beyond that which is in the autocovariance, its time domain counterpart. The two techniques simply present the information concerning the variance of the signal in different and complementary ways. The periodogram is given by

$$I(f_i) = \frac{2}{N}\left\{\left[\sum_{k=1}^{N} y(k)\cos(2\pi f_i k)\right]^2 + \left[\sum_{k=1}^{N} y(k)\sin(2\pi f_i k)\right]^2\right\}$$

for i = 1, 2, …, q

where q = (N -1)/2 for odd N and q = N/2 for even N.

The periodogram thus derived is a plot of I(f_i) against f_i, where f_i = i/N is the i^{th} harmonic of the fundamental frequency 1/N up to the Nyquist frequency of 0.5 cycles per sampling. As a result, the periodogram maps out the spectral content of the signal, indicating how its relative power varies over the range of frequencies between f_i = 0 and 0.5. Since I(f_i) is obtained by multiplying y(k)by sine and cosine functions of the harmonic frequency, it will take on relatively large values when this frequency coincides with a periodicity of this frequency occurring in y(k). Significant periodicity in the signal with period T = 1/f_i samples will induce a sharp peak in the periodogram at f_i cycles/sample. Figure 3.6 presents the raw periodogram of conductivity. The area under the curve represents the variance of the signal. The raw periodogram at times may not be an appropriate estimator of the spectrum due to leakages. This occurs when variations in one frequency leaks into periodogram terms at frequencies different from the true frequency of the variation. To reduce the leakage

problem the periodogram is smooth by a windowing technique in a controlled way.

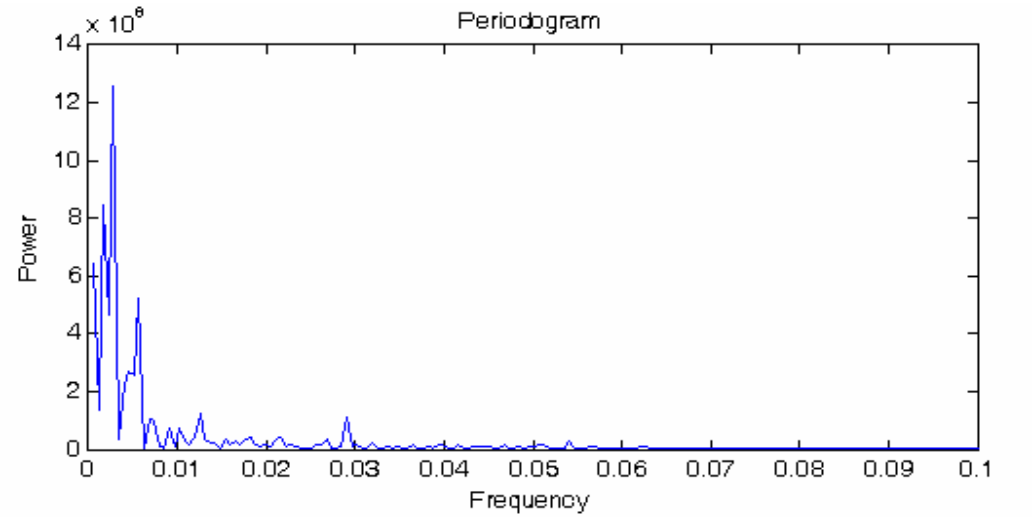

Figure 3.6: Periodogram of conductivity.

3.2.2 Spectral analysis

A power spectrum is a slightly modified version of the periodogram done by smoothing the periodogram. In this technique, the intensities are averaged together across neighboring frequencies to provide a smoother and more reliable estimate of the distribution of variance continuously across the entire range of frequencies from 1/N to ½. Though smoothing reduces leakage, excessive smoothing makes it difficult to distinguish between the contributions of neighboring frequencies or to make judgments about the period or cycle length of major periodic components (Martin and Thomson, 1982; Brillinger, 1981; Grenander, 1981).

There are various smoothing or windowing techniques which differ from each other mainly in two ways. Firstly, the width of the window, which is the number of neighboring frequencies that are included in the weighted average, can vary.

Secondly, the weights used for this weighted forms can be of different forms (Percival, and Walden, 1993). Before smoothing, the graph of the periodogram is usually relatively jagged, while the power spectrum is somewhat flatter. Most windows such as the Hamming or Hanning window give larger weights to the frequencies near the centre of the window than to frequencies near the edge of the window. Smoothing is not necessary when there is no evidence of leakage like in figure 3.6. Table 3.2 provides the algorithms of some main window functions with P denoting the width of the windows. *n* represents an integer, with values that range from zero to N-1. Hence, it gives a time shifted form of the windows. Figure 3.7 and 3.8 both represent examples of a Hamming and Bartlett window of different widths.

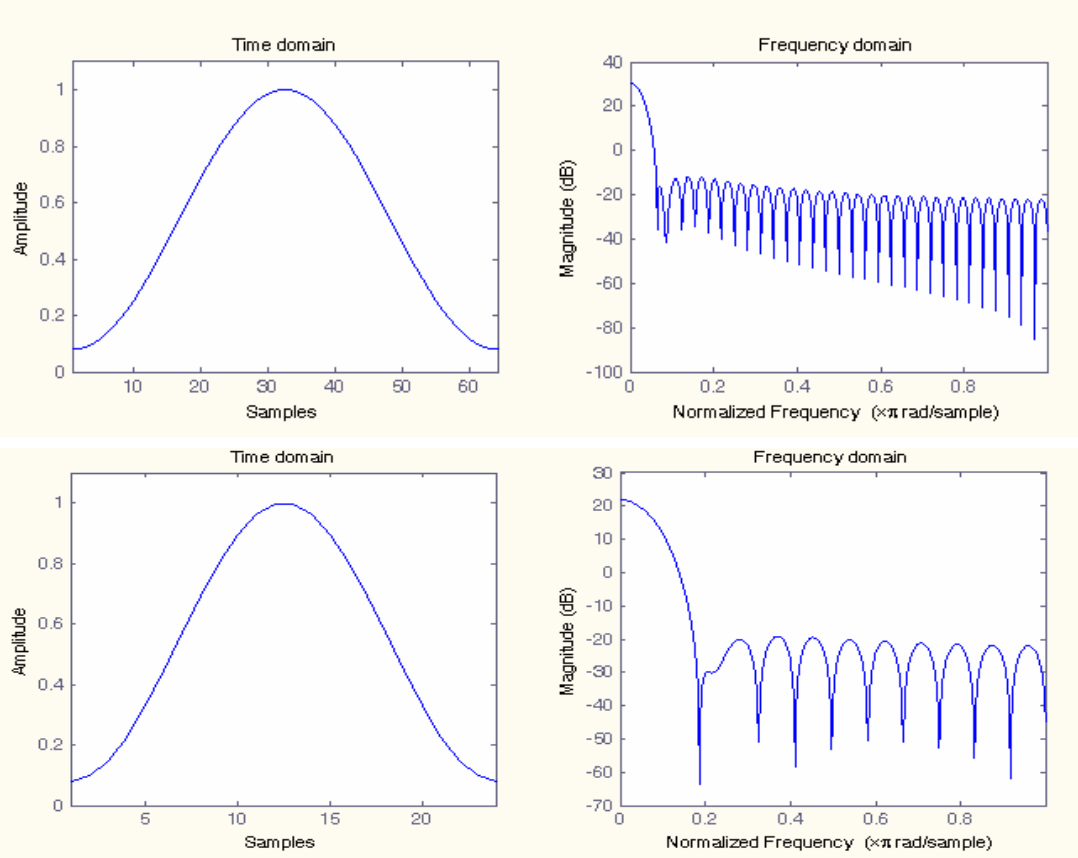

Figure 3.7: Hamming window of width 64 (top) and of width 24 (bottom).

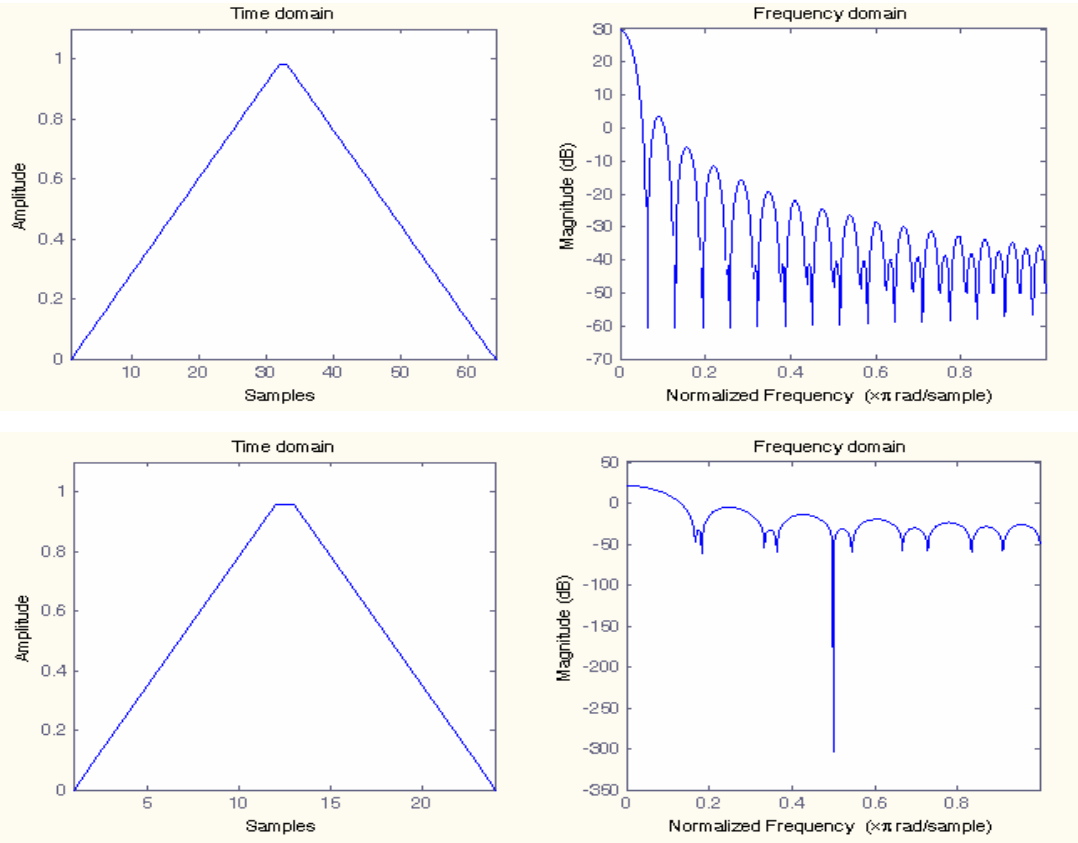

Figure 3.8: Bartlett window of width 64 (top) and of width 24 (bottom).

Table 3.2: Some main window functions and their formulae

Window	**Formula**
Gauss	$w(n) = e^{-\frac{1}{2}\left(\frac{n-(N-1)/2}{\sigma(N-1)/2}\right)^2}$ $\sigma \le 0.5$
Hamming	$w(n) = e^{-\frac{1}{2}\left(\frac{n-(N-1)/2}{\sigma(N-1)/2}\right)^2}$

Hanning	$w(n) = 0.5\left(1 - \cos\left(\frac{2\pi n}{N-1}\right)\right)$
Bartlett	$w(n) = \frac{2}{N-1}\left(\frac{N-1}{2} - \left\lvert n - \frac{N-1}{2}\right\rvert\right)$
Triangular	$w(n) = \frac{2}{N}\left(\frac{N}{2} - \left\lvert n - \frac{N-1}{2}\right\rvert\right)$
Kaiser	$w(n) = \frac{I_o\left(\pi\alpha\sqrt{1-\left(\frac{2n}{N-1}-1\right)^2}\right)}{I_o(\pi\alpha)}$
Bartlett-Hanning	$w(n) = a_o - a_1\left\lvert\frac{n}{N-1} - \frac{1}{2}\right\rvert - a_2\cos\left(\frac{2\pi n}{N-1}\right)$ $a_o = 0.62$; $a_1 = 0.48$; $a_2 = 0.8$

The standard approach for computing the smooth periodogram is as follows. Given the discrete Fourier transform

$$F(f_k) = \frac{1}{n}\sum_{t=1}^{n} w_t y_t e^{i2\pi f_k t}, k = 0,...,n-1,$$

computed from the signal with values $y_1,\ldots,y_n$, where (w_t) is a data window or taper used for the mitigation of leakages due to end effects (Bloomfield, 1976; Brillinger, 1981). The periodogram, given by

$$\hat{I}_y(f_k) = \left\lvert F(f_k)\right\rvert^2$$

is computed, and finally the smoothed periodogram estimate

$$\overline{I_y}(f_k) = \sum_{l=-L}^{L} b_l \hat{I}(f_{k+l})$$

is computed, with the frequency domain smoothing weights $b_l = b_{n,l}$ chosen so as to achieve a good compromise between bias and variability.

3.2.3 Cross spectral analysis

Using the univariate spectral analysis enables one to detect periodic movements or patterns within a signal while the bivariate spectral analysis enables to describe pairs of signals in the frequency domain (Parzan, 1983). This is achieved by decomposing their covariance in to different frequency components. The goal in bivariate spectral analysis is the detection and measurement of the strength of the relationship between two signals as well as the direction of the relationship. The cross spectral analysis may be regarded as the frequency domain equivalent of the time domain correlation analysis and enables the determination of the correlations between signals at different frequencies. In other words, cross spectral analysis enables the estimation of the co-variation between signals a as function of frequency (Priestly, 1981). It is composed of three major components namely: the squared coherency spectrum, the phase spectrum and the gain.

The squared coherency, derived from the cross spectrum, measures the degree to which two signals can be represented as a linear function of the other in the frequency domain. It is the squared correlation of cyclical components of the two signals (Platt and Denman, 1975) and serves as a measure of the variance. This can also be regarded as the squared correlation of cyclical components of two signals for a given time period and is used as a measure for the explained variance.

As given by Rebecca (1998), the discrete Fourier transform F_x of a signal X can be calculated for each frequency ω by

$$F_x(\omega) = \frac{1}{N} \sum_{t=0}^{n-1} X_t e^{-i\omega t}$$

with the Fourier transform for signal Y being F_y at frequency ω. Bloomfield (1976) shows that the cross-periodogram or cross-spectrum $I_{x,y}$ between signal X and Y can be derived from

$$I_{x,y}(\omega) = \frac{N}{2\pi} F_x(\omega) F_y(\omega)$$

$I_{x,y}$ consist of complex numbers which are not so interesting in themselves but can be used for estimating the coherence and the phase between the signals X and Y at each frequency. The squared coherence between signals X and Y at frequency ω is computed from the cross spectra and the individual spectra of X and Y (Rebecca, 1998). Given $g_{x,y}$ as the cross-spectrum for signals X and Y, $g_{x,x}$ as the spectrum for signal X, $g_{y,y}$ as the spectrum for signal Y. The squared coherence between signal X and Y at frequency ω is given by

$$F_{xy}(\omega)^2 = \frac{\left| f_{xy}(\omega) \right|^2}{f_x(\omega).f_y(\omega)}$$

and is analogous to the formula for the Pearson r^2 which is calculated by taking the covariance of X and Y divided by the Variance of X and Y. The coherence lies between zero and one and is interpreted similarly like the r^2. At each frequency, it is the estimated proportion of variance that is shared between the two signals within that particular frequency band.

The phase of the cross-spectrum, denoted by $\varphi_{x,y}(\omega)$ is derived from the imaginary (Im) and the real (Re) parts of the cross-spectrum as follows

$$\varphi_{x,y}(\omega) = \arctan\left[\frac{\mathrm{Im}\, f_{xy}(\omega)}{\mathrm{Re}\, f_{xy}(\omega)} \right]$$

which measures the phase difference between the frequency components of the two series.

The gain is given by

$$G_{xy}(\omega) = \frac{\left| f_{xy}(\omega) \right|}{f_x(\omega)}$$

which indicates the extend to which the spectrum $F_x(\omega)$ has been modified to approximate the corresponding frequency component of $F_y(\omega)$. The analysis of these three quantities together with the spectrum of each signal with the amplitude of their cross spectrum give one an overall view of the frequency interaction of the two signals.

The coherence indicates the percentage of shared variance between the two signals at a given frequency band while the phase indicates the timing of the peaks in the Y signal relative to peaks in the X signal at a particular frequency band (the phase is however usually given in terms of the fractions of the cycle)(Rebecca, 1998). The coherence from one or several frequency bands in the cross-spectrum summarizes how strongly or weakly related the two signals are across the entire set of Nyquist frequencies. It is important to note that the squared coherency spectrum has to be interpreted taking into consideration the individual spectra of the two signals (only frequency bands having a relatively high variance are important) while the phase spectrum has to be interpreted taking into consideration the coherence spectrum. The coherence lies between 0 and +1 and is an estimation of the percentage of covariance of the signals within a certain frequency band. If there is however a negligible amount of variance in a given frequency band, a large coherence in this band is not important. The univariate spectrum therefore serves as a screening device for deciding which frequency band contains enough variance in one or both signals to be interesting, and then proceed to examine the coherence only for those frequency bands.

Similarly, unless there exist an acceptably high coherence between the two signals within a particular frequency band, it will be totally senseless to search for the phase relationship between the two signals at that frequency band. It will only be reasonable to investigate the phase at frequency bands where the coherence is reasonably high (in addition, sampling errors of the phase estimate is inversely related to the squared coherence meaning that when the coherence gets small, the error in estimating the phase increases and vice-

versa). Hence, any interpretation of the phase spectrum should exclusively be limited to the frequency bands where the coherence is reasonably high.

The outcome of the squared coherence spectrum can either be a uniformly low value across all frequency bands suggesting that the two signals are not related within any frequency band or at any time lag or it can be fairly high across all frequency bands of is selectively high only within one or few narrow frequency bands. To conclude that there is an existence of synchronized cycles within the two signals, the following observations must have been made. A high percentage of the power of the two signals should be contained within the same frequency band or closely neighboring bands (to allow for estimation errors) and a high coherence between the two signals should occur within this same frequency band. Then, the phase spectrum can be used to examine the phase relationship at this same frequency band.

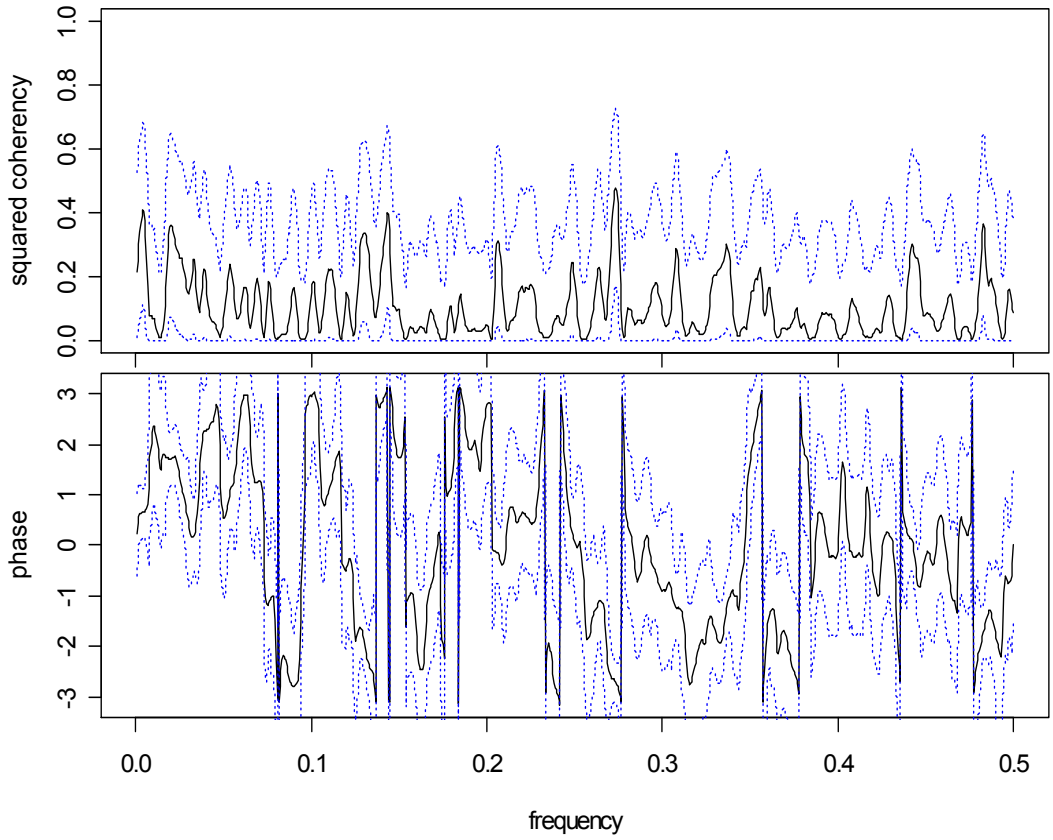

Figure 3.9: Squared coherency (top) and phase spectra between conductivity and water temperature.

Figure 3.9 reveals the squared coherency and phase spectra of water temperature and conductivity. The frequency domain analysis reveals the frequency content of the time series, but no information is given concerning the time at which these frequencies occur. In addition, the frequency domain methods are valid only for stationary time series, which is contrary to the behavior exhibited by water quality time series. Hence, these time series usually need to be transformed to render them stationary or are simply assumed stationary within certain intervals prior to frequency domain analysis. The frequent change in period and amplitude of periodic patterns observed in water quality indicators as a result of time varying internal and external driving forces cannot be captured or investigated by these methods. In an attempt to remedy this situation and use a time-frequency representation of a time series in the frequency domain was given by the windowed Fourier integral transform, which attempts to approximate a signal to a local Fourier expansion of *f*. In order to attain time localization, *f* is multiplied with a time window *w(t)* and the usual Fourier integral transform is performed (Golubev et al., 2000). This implies the Fourier analysis is applied only to a section of the entire signal and the signal is assumed stationary within the length of the window. The window is then moved along the time axis, thereby covering the complete time domain of the signal. Hence, the windowed signal $f_w(t)$ depends on both the position of the window μ and the time t such that

$$f_w(t,\Delta) = w(t-\Delta)f(t)$$

Applying the Fourier transform to this windowed signal gives the windowed Fourier transform as a function of the frequency ω and the position of the window Δ. The windowed Fourier transform is therefore given by

$$F(\omega,\Delta) = \int w(t-\Delta)f(t)e^{-2\pi wt}dt$$

The major setback of the windowed Fourier transform is that the same time window is utilized over the whole frequency domain. Hence, the time series exhibiting high frequency changes, the window may be too large and too small

for very low frequency changes. The wavelet transform, by the translation and dilation of a single function called the wavelet, provides decomposition into components from different scales whose degree of localization is connected to the size of the scale window as will be shown in the following section.

3.3 Wavelet analysis

The main idea behind wavelet analysis is to imitate the windowed Fourier analysis, but using basis functions (wavelets) that are better suited to capture local behavior of non-stationary signals (Shumway, 2005).

3.3.1 Wavelets

A wavelet is a small wave and a wave is a real valued function that is defined over the entire real axis and oscillates back and forth about zero with the amplitude of the oscillations remaining relatively constant everywhere (Percival et al, 2004) like the sine wave. A wavelet ψ(t) is a function of time that obeys the following wavelet admissibility condition (Debnath, 2002; Gençay et al., 2002).

1. $\int_{-\infty}^{\infty} \psi(t)dt = 0$
2. $\int_{-\infty}^{\infty} \phi\ (t)dt = 1$

Wavelets come in families generated by the father wavelet denoted by Φ and a mother wavelet denoted by ψ. Father wavelets, used to represent the long scale smooth or low frequency component of a signal integrates to one while the mother wavelet, used to capture the detailed and high frequency components or deviations from the smooth components, integrates to zero (Wickerhauser, 1994; Whitcher et al., 2000). The father wavelet gives rise to the scaling coefficients while the mother wavelet gives rise to the differencing coefficients. Hence, the father wavelet acts as a low pass filter while the mother wavelet acts as a high pass filter. The father wavelet is given by

$$\phi_{J,b} = 2^{-\frac{J}{2}} \phi\left(\frac{t - 2^J b}{2^J}\right)$$

with

$$\int \phi(t)dt = 1$$

and the mother wavelet given by

$$\psi_{J,b} = 2^{-\frac{j}{2}} \psi\left(\frac{t - 2^j b}{2^j}\right)$$

with

$$\int \psi(t)dt = 0$$

for j = 1, …, J

The shift or translation parameter is $2^j b$, and the scale parameter is 2^j. Wavelet analysis serves as a mathematical tool acting as a lens for inspecting the time varying structure of signals and relationships between signals. It decomposes signals on a scale by scale basis by projecting them onto a basis function and can be expressed by

$$\Psi_{u,s}(t) = \frac{1}{\sqrt{s}} \Psi\left(\frac{t-u}{s}\right)$$

where *S* is the scale parameter, u is the location parameter. Changing *S* produces dilating effects (*S* > 1) or contracting effects (*S* < 1) of the function ψ(t). Changing *U* analysis the signal f(t) around different points of *U*. One of the earliest wavelet functions, the Haar wavelet are useful for demonstrating properties of wavelets, but they do not have good time frequency localisation properties (Shumway, 2005). A classic example of a continuous wavelet function is the Morlet wavelet (figure 3.10) given by

$$\psi^m(t) = \frac{1}{\sqrt{2\pi}} e^{-i\omega_0 t} e^{-\frac{t^2}{2}}$$

where i = √-1 is the imaginary part. ω_o is the central frequency of the wavelet.

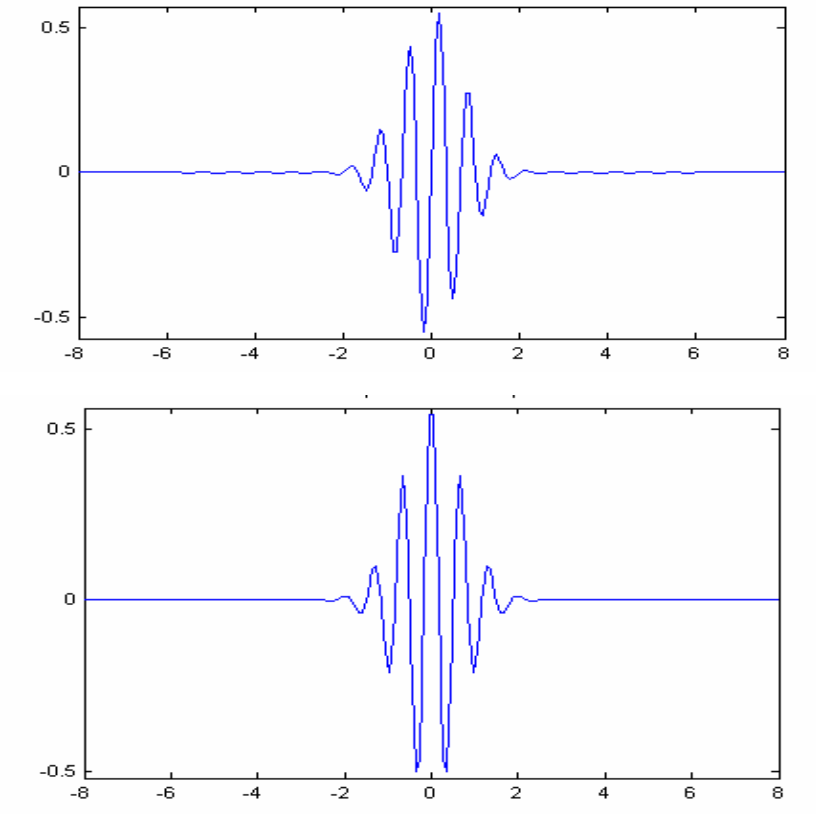

Figure 3.10: Father (top) and mother (bottom) wavelet of the Morlet family.

The wavelet functions are spread out for larger values of *j* (or scale parameter 2^j) and tall and narrow for small values of the scale.

The Gaussian family (figure 3.11) which relates to the first derivative of the Gaussian probability density function is given by

$$\psi^g(t) = \frac{\sqrt{2t}}{\sigma^{3/2}\pi^{1/4}} e^{-t^2/2\sigma^2}$$

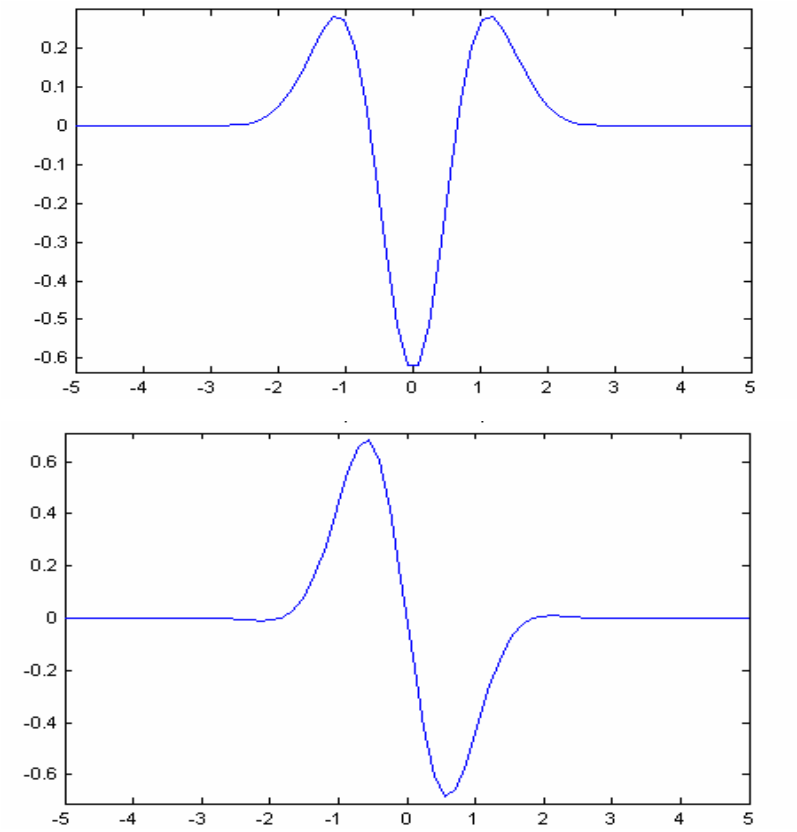

Figure 3.11: Father (top) and mother (bottom) wavelet of the Gaussian family.

Ingrid Daubechies invented the compactly supported orthonormal wavelets consisting of nine members called the Daubechies wavelet family and can be represented by

$$D(f) = 2\cos^L(\pi f)\sum_{i=0}^{L/2-1}\binom{L/2-1+l}{l}\sin^{2l}(\pi f)$$

In the Daubechies approach, the father wavelet Φ and the mother wavelet ψ are produced with compact support such that the moments of Φ and ψ of order from 1 to *n* vanish (Daubechies, 1988; 1990). This is a necessary property to guarantee good approximation properties of the corresponding wavelet expansions. Some members of the Daubechies family are shown in figures 3.12, 3.13, 3.14 and 3.15.

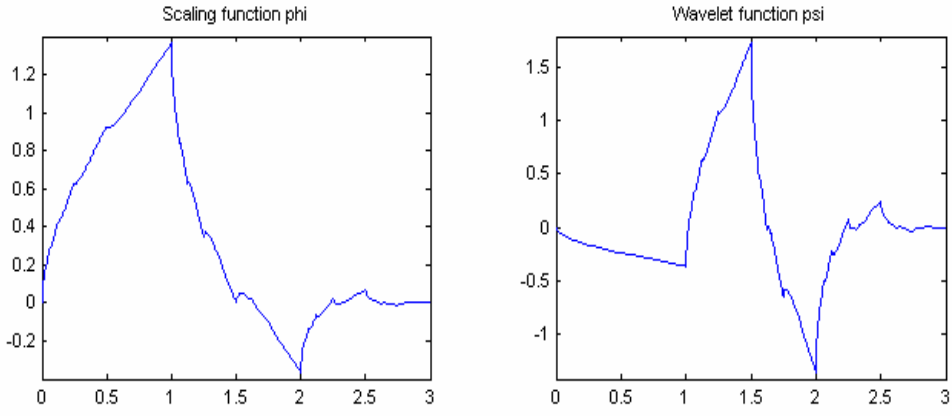

Figure 3.12: Father (left) and mother (right) wavelet of the Daubechies 2 family (db2).

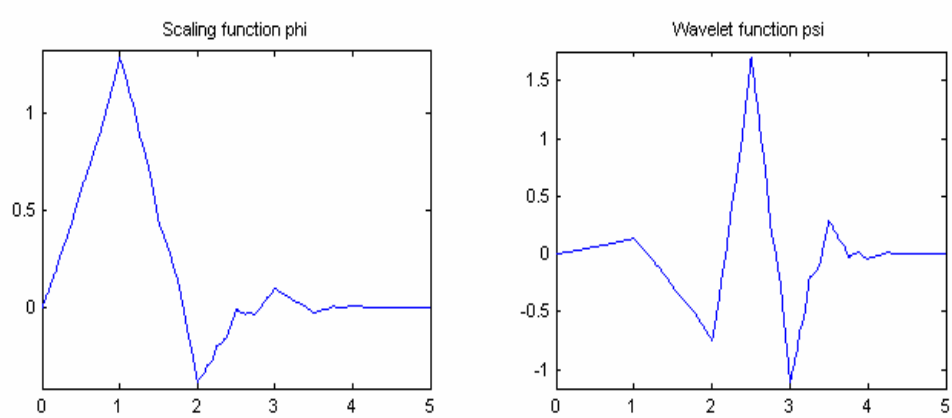

Figure 3.13: Father (left) and mother (right) wavelet of the Daubechies 3 family (db3).

The number after the name of the wavelet refers to the width and smoothness of the wavelet; for example, the daublet 8 (db8) is wider and smoother than the

daublet 4 (db4) (Shumway, 2005). The daublets are among the first type of continuous orthogonal wavelets with compact support while a family like the symmlets are constructed to be closer to symmetry than the daublets (Shumway, 2000).

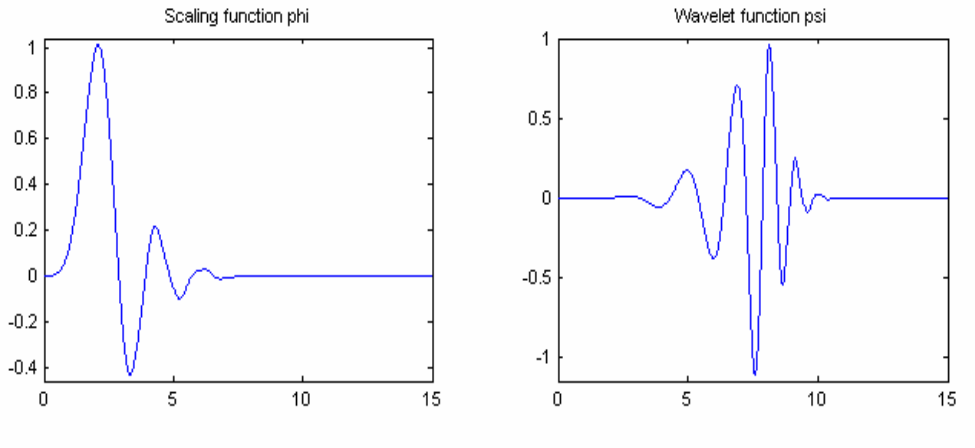

Figure 3.14: Father (left) and mother (right) wavelet of the Daubechies 4 family (db4).

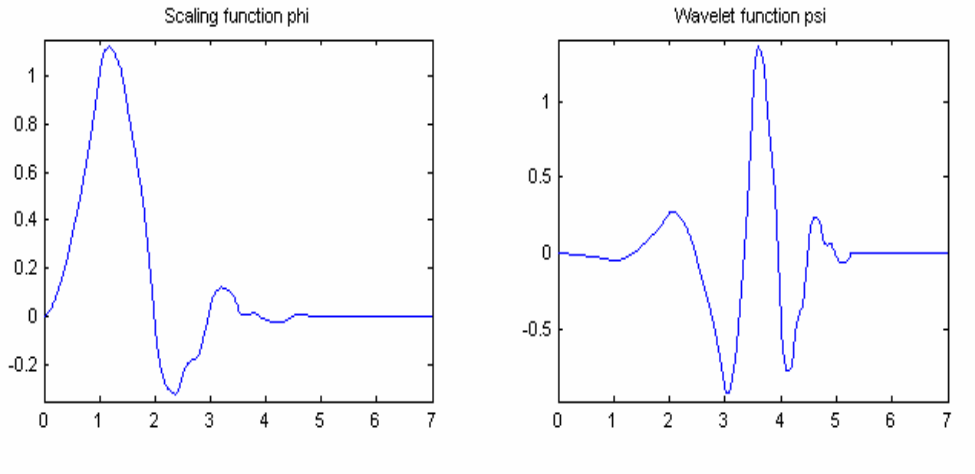

Figure 3.15: Father (left) and mother (right) wavelet of the Daubechies 8 family (db8).

A scaled and translated version of the wavelet function is given by

$$\psi_{j,b}(t) = 2^{-j/2}\psi\left(\frac{t-2^j b}{2^j}\right)$$

for $j = 1, \ldots, J$, where the shift or translation parameter is given by $2^j b$, with the scale parameter given by 2^j, the wavelets are spread out and shorter for larger values of j (or scale parameter2^j) and tall and narrow for smaller values of the scale. The larger values of the scale represent slower, smoother or coarser movements of the signal while the smaller values of the scale represent the faster, choppier or finer high frequency movement of the signal.

3.3.2 Wavelet transforms

There are three essential wavelet transforms, namely the continuous wavelet transform, the discrete wavelet transform and the maximal overlap discrete wavelet transform.

3.3.2.1 The continuous wavelet transform

This is a function of two variables W(u,s) obtained by projecting a signal X(t) on to a particular wavelet Ψ and is given by

$$W(u,s) = \int_{-\infty}^{\infty} X(t)\Psi_{u,s}(t)dt,$$

where

$$\Psi_{u,s}(t) = \frac{1}{\sqrt{s}}\Psi\left(\frac{t-u}{s}\right)$$

gives a translated and dilated version of the original wavelet function. The coefficients that are obtained are a function of the location and scale parameters. Applying shifted and scaled versions of a wavelet function decomposes the signal into simpler components (Torrence and Compo, 1998). The continuous wavelet transform is applied to a function X(t) defined over the entire real axis. On the other hand, in practical applications, we have only a

finite number of N of sampled values, rendering the CWT inadequate, hence, the need for a discrete version.

3.3.2.2 Discrete wavelet transform

Implementing the wavelet transform on sampled data requires the discretization of the scale and location parameters. Kumar and Foufoula (1997) demonstrate that in discretizing the two parameters (s,u), one can choose $S = S^m_o$, where m is an integer and S_o is a fixed dilation step greater than one. Given $\sigma_s = S\sigma_1$, one can choose $t = nt_o S^o_m$, where $t_o > 0$ is dependent on ψ(t) with *n* being an integer. By defining

$$\psi_{m,n}(t) = \frac{1}{\sqrt{s_o^m}}\psi(\frac{t - nt_o\lambda_o^m}{s_o^m}) = s_o^{\frac{-m}{2}}\psi(s_o^{-m}t - nt_o)$$

is valid and the discrete wavelet transform is given by

$$wf(m,n) = s_o^{\frac{-m}{2}}\int f(t)\psi(s_o^{-m}t - nt_o)dt$$

when the DWT is applied to a time series or vector of observations X, it gives N wavelet coefficients

$$w = Wx$$

the coefficients can be organized into J+1 vectors $w = [w_1, \ldots, w_j, v_J]^T$, with w_j being the length and $N/2^J$ vector of scaling coefficients associated with averages on a scale of length 2^J (Whitcher, 1998).

3.3.2.3 Maximal overlap discrete wavelet transform (MODWT)

Despite the fact that the DWT is a very useful operation, it has some limitations. An alternative transform, the MODWT, is a modified version of the discrete wavelet transform. Given a signals x, the MODWT coefficients are given by

$$\hat{W} = \hat{W}x$$

With $\hat{W}$ being a (J+1) N x N matrix that defines the MODWT with the coefficients being organized into J + 1 vectors.

$$\hat{w} = [\hat{w}1, \hat{w}2, \ldots, \hat{w}_J, \acute{U}_J]^T$$

with $\hat{w}_J$ being the length $N/2^J$ vector of scaling coefficients associated with averages on a scale of length $2^J = 2S_j$ similar to the DWT.

Four important properties available for distinguishing the DWT from the MODWT are provided by Percival and Mofjeld (1997).

1. The MODWT is able to handle any given sample size N contrary to the DWT which restricts the sample size to a multiple of 2^J.
2. The smooths and details produced by the multiresolution analysis of a MODWT are associated with zero phase filters implying a proper aligning of the original signal with features in a multiresolution analysis.
3. Contrary to the DWT, the MODWT wavelet and scaling coefficients as well as MRA are shift invariant to a circular shifting of the original signal. Shifting the signal by a given amount simply shifts the MODWT wavelet and scaling coefficient by the same amount.
4. The MODWT estimator of variance is statistically more efficient than the DWT estimator.

3.3.3 Multiresolution analysis

The idea behind wavelet multiresolution analysis is to study a signal or process represented at different resolutions and to develop a mechanism for moving from one resolution to another. Given that the wavelet transform breaks a signal by projecting it onto scaled and shifted versions of the basis function results in representing the signal at multiple scales (Percival and Guttorp, 1994; Percival and Walden, 2000). The effect of the shifting and scaling process is what makes this representation possible and is referred to as multiresolution analysis. The wavelet transform is usually applied in the form of a filter bank, comprising two filters. The scaling filter known as the father wavelet is a low

pass filter while the wavelet filter known as the mother wavelet is a high pass filter.

Given a signal X(t) of length N = 2^j, the filtering procedure can be performed a maximum of j time, giving rise to j different wavelet scales. The wavelet coefficients or detail coefficients are produced by the wavelet filter while the scaling filter gives rise to the smooth version of the signal used at the next scale. Given the respective father and mother wavelets,

$$\phi_{J,b} = 2^{\frac{-J}{2}} \phi\left(\frac{t - 2^J b}{2^J}\right)$$

$$\int \phi(t)dt = 1$$

and

$$\psi_{j,b} = 2^{\frac{-j}{2}} \psi\left(\frac{t - 2^j b}{2^j}\right)$$

for j=1, . . ., J

$$\int \psi(t)dt = 0$$

where $\Phi_{J,b}$ is the father wavelet and $\Psi_{j,b}$ is the mother wavelet with the scale parameter "s" being restricted to the dyadic scale 2^j. When a signal is projected onto a given basis function, we obtain

$$S_{J,b} = \int f(t)\phi_{J,b}$$

and

$$d_{j,b} = \int f(t)\psi_{j,b}$$

with $S_{J,k}$ being the coefficients for the father wavelet at a maximum scale of 2^J (the smooth coefficients) and $d_{j,k}$ being the detail coefficients from the mother wavelet at all scales from 1 to J, to the maximal scale. Based on these coefficients, the function *f(t)* can be represented by

$$f(t) = \sum_b S_{J,b}\phi_{J,b}(t) + \sum_b d_{j,b}\psi_{J,b}(t) + ... + \sum_b d_{1,b}\psi_{1,b}(t)$$

and can be equally represented by

$$f(t) = S_J + D_j + D_{j-1} + \ldots + D_j + \ldots D_1$$

where

$$S_J = \sum_b S_{J,b}\phi_{J,b}(t)$$

$$D_j = \sum_b d_{j,b}\psi_{j,b}(t)$$

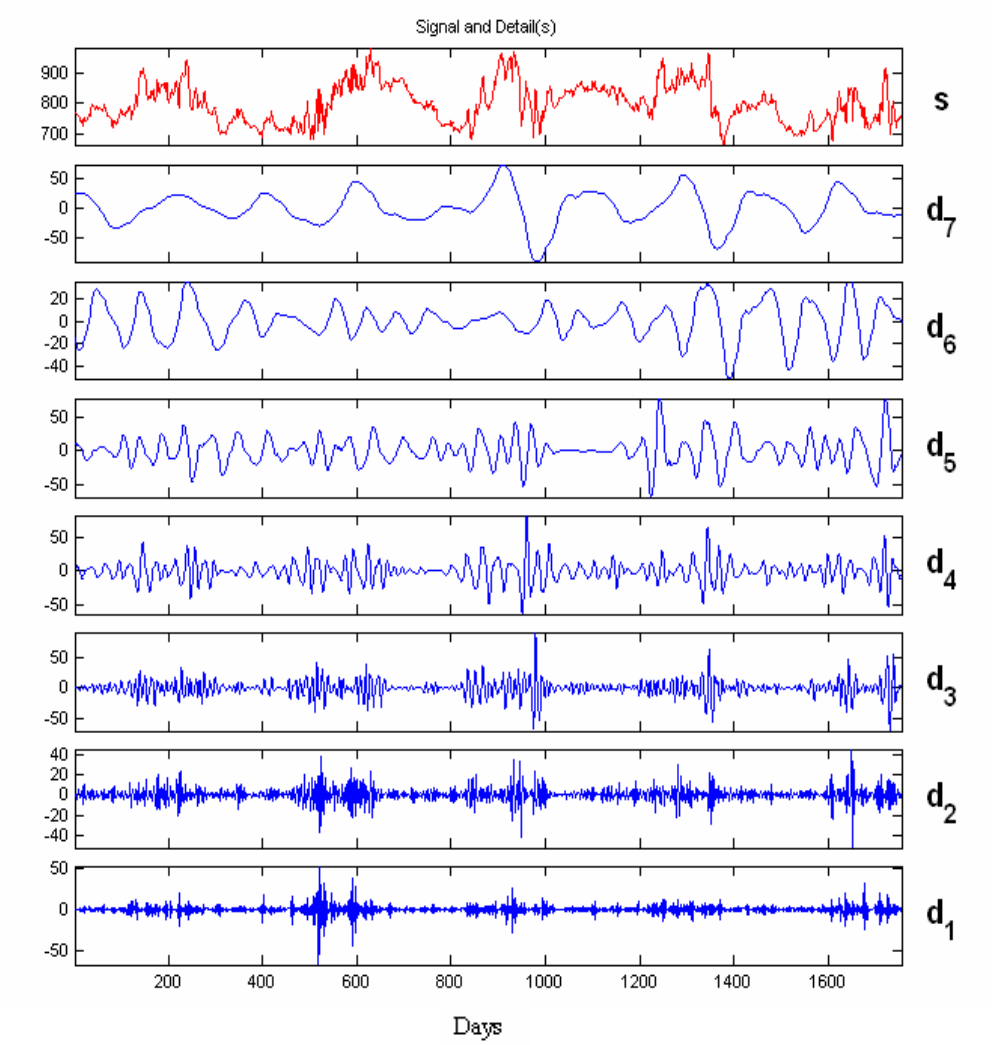

Figure 3.16: Details of conductivity at level 7 with db4.

Figures 3.16 and 3.17 respectively present the details and the approximations from the multiresolution analysis of conductivity at level 7 using the Daubechies 4 mother wavelet. The details unravel the changes occurring in the signal at different time scale. This is quite important in elucidating the time scale at which the changes that influence the long term dynamics of the signal occur. As can be observed, the low scales or high frequency components have a relatively low intensity compared to the low frequency or higher scale changes. This simply reveals that the changes influencing the general tendency of the signal occurs at the higher time scales.

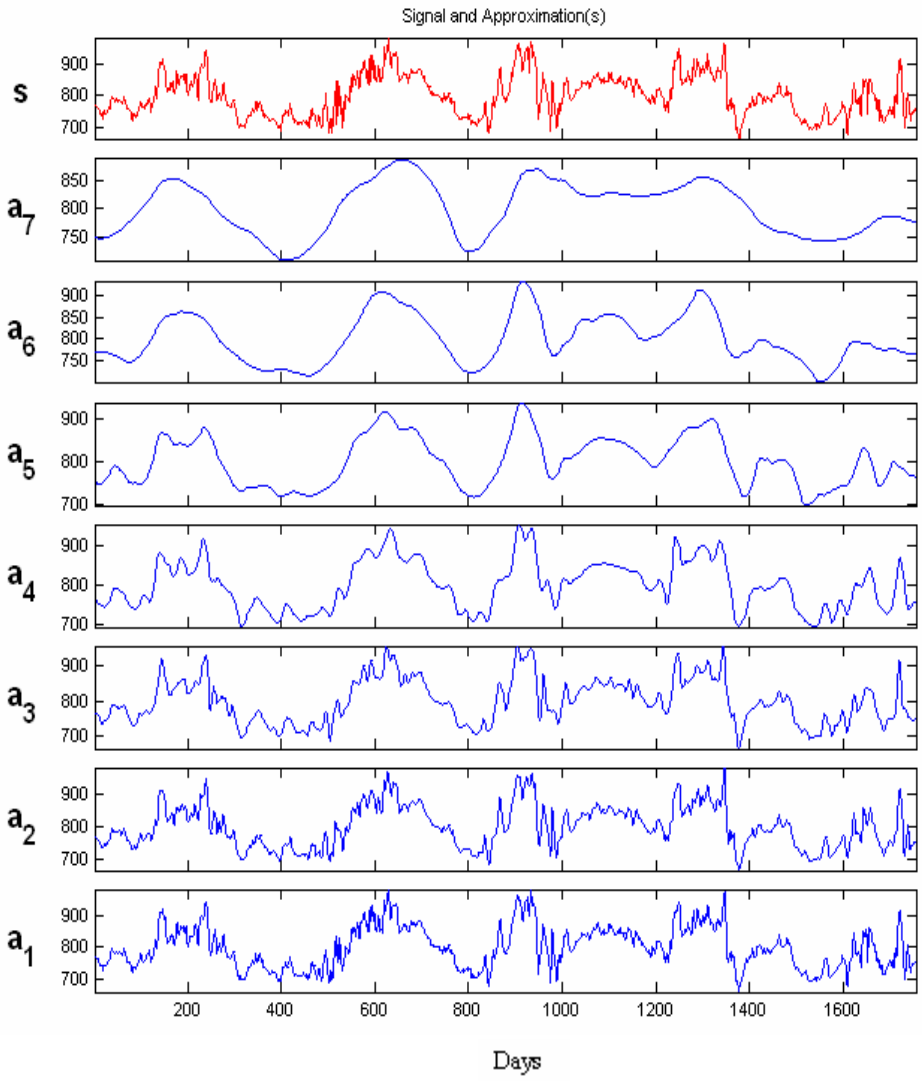

Figure 3.17: Approximations of conductivity at level 7 with db4.

The approximations in figure 3.17 shows the progressively filtered version of the time series from one scale to the next, acting like a lens. This clearer and less blur version, as a result of the progressive removal of the active fluctuations from one scale to another enables the long term or the low frequency behavior of the time series to be revealed. Features such as trends can be seen as the resolution of the lens is increased as one move from one scale to the next.

3.3.4 Wavelet variance and covariance

As earlier shown by the MRA, the wavelet transform is able to decompose a signal on a scale by scale basis. The variance of a signal can equally be decomposed using this technique. Gençay et al. (2002) showed that the time varying variance for a signal X_t is the variance of the scale S_j wavelet coefficient $W_{j,t}$ using

$$\sigma_{x,t}^2(S_j) = \frac{1}{2S_j}\mathrm{var}(w_{j,t})$$

Assuming that the wavelet variance is dependent of t, the time independent wavelet variance can be summarized using the time dependent wavelet variance given by

$$\sigma_x^2(S_j) = \frac{1}{2S_j}\mathrm{var}(w_{j,t})$$

As presented by Percival (1995), the wavelet variance decomposes the variance of X_t on a scale by scale basis using

$$\sum_{j=1}^{\infty}\sigma_x^2(S_j) = \mathrm{var}(X_t)$$

Gençay et al. (2002) show that the wavelet covariance is the covariance between the scale S_j wavelet coefficients from a bivariate signal. Given a bivariate signal $X_t = (X_{1,t}, X_{2,t})$ of a stochastic process, let $W_{j,t} = (w_{1,j,t}, w_{2,j,t})$ be the scale S_j wavelet coefficients computed from X_t. the wavelet covariance of the bivariate signal $X_{1,t}$, $X_{2,t}$ for the scale S_j is given by

$$\gamma_x(S_j) = \frac{1}{2S_j}\mathrm{cov}(w_{1,j,t}, w_{2,j,t})$$

Similar to the wavelet variance of a univariate signal, the wavelet covariance decomposes the covariance between two signals on a scale by scale basis (Whitcher, 1998) by

$$\sum_{j=1}^{\infty}\gamma_x(S_j) = \mathrm{cov}(x_{1,t}, x_{x,t})$$

If a time lag τ is introduced between $w_{1,j,t}$ and $w_{2,j,t}$, the wavelet cross-covariance can be obtained by

$$\gamma_{x,\tau}(S_j) = \frac{1}{2S_j}\mathrm{cov}(w_{1,j,t}, w_{2,j,t+\tau})$$

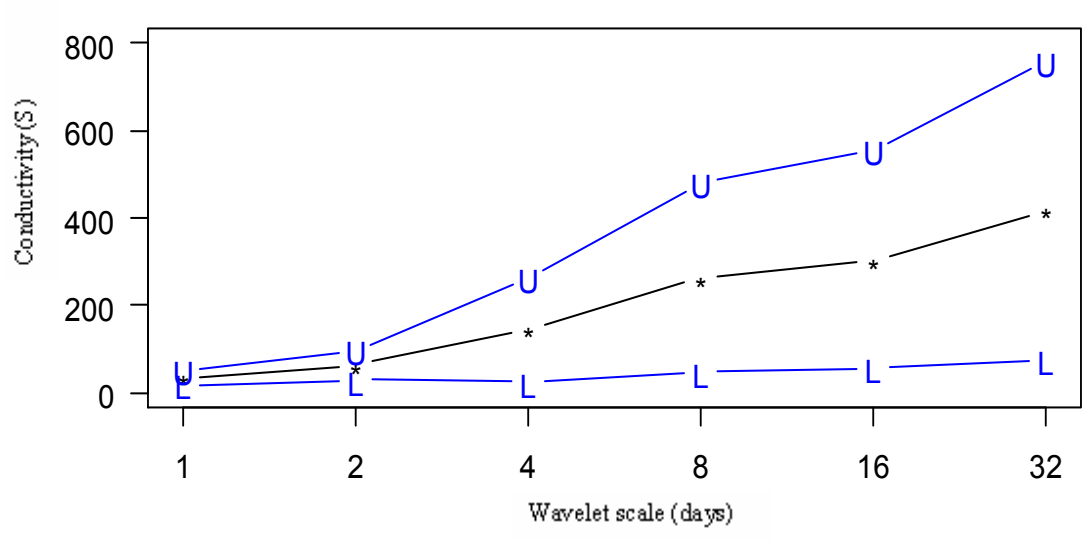

Figure 3.18: Wavelet variance of conductivity (S) with db4 wavelet.

The blue lines indicate the upper and lower limits of the wavelet variance of conductivity. Figure 3.18 confirms that the intensity of the changes occurring in the signal are negligible at scale 1 and 2 and starts increasing significantly as the time scale increases. This indicates that no new information is obtained by sampling at a time scale lower than 4 because the changes taking place below

4 are near zero. The long term behaviour is determined by the changes occurring at the higher time scales.

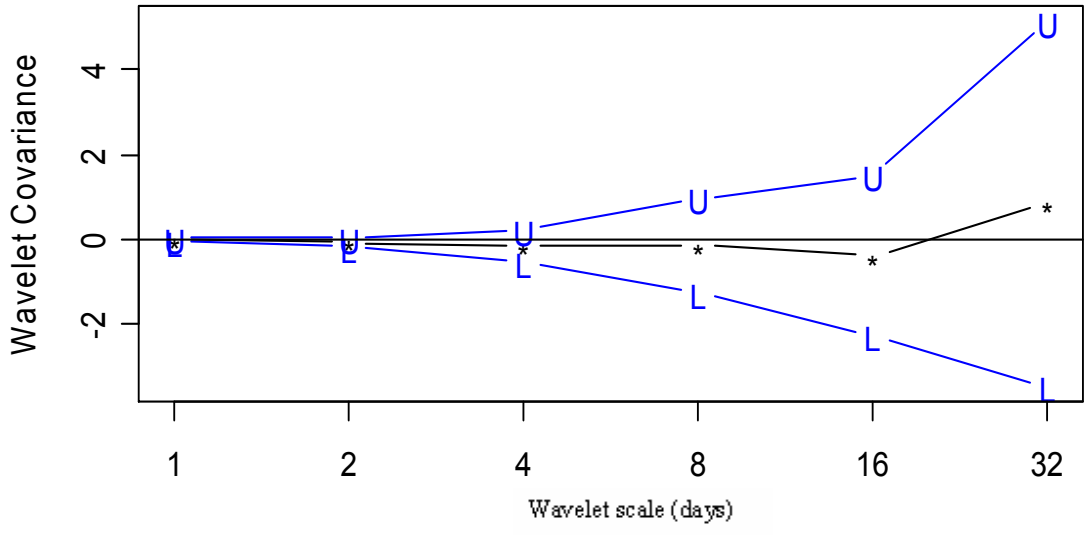

Figure 3.19: Wavelet covariance of conductivity (S) and water temperature (°C) with db4 wavelet.

Figure 3.19 gives the covariance between conductivity and water temperature, elucidating the interrelationship between indicators at different time scales. At different scales, there seam to exist no significant covariance between the signals though the limits increase as the time scale increases. This is quite important in detecting the behavior of signals at the dominant scales of variability in a freshwater body.

3.3.5 Wavelet correlation and cross-correlation

The wavelet correlation makes use of the wavelet variance and covariance to decompose the correlation between two signals on a scale by scale basis. Normalizing the wavelet covariance by the variability inherent in the observed wavelet coefficients gives the wavelet correlation (Gençay et al., 2002). This is given by

$$\rho_x(s_j) = \frac{\gamma_x(s_j)}{\sigma_1(s_j)\sigma_2(s_j)}$$

where σ_1^2 (S_j) is the wavelet variance for $X_{1,t}$ and σ_2^2 (S_j) is the wavelet variance for $X_{2,t}$ associated with the scale S_j. The value of wavelet correlation coefficient varies between 1 and -1. Figure 3.20 gives the wavelet correlation, a normalized version of the wavelet covariance. It presents more or less the same results like the like the covariance by indicating the absence of a significant relationship between conductivity and water temperature.

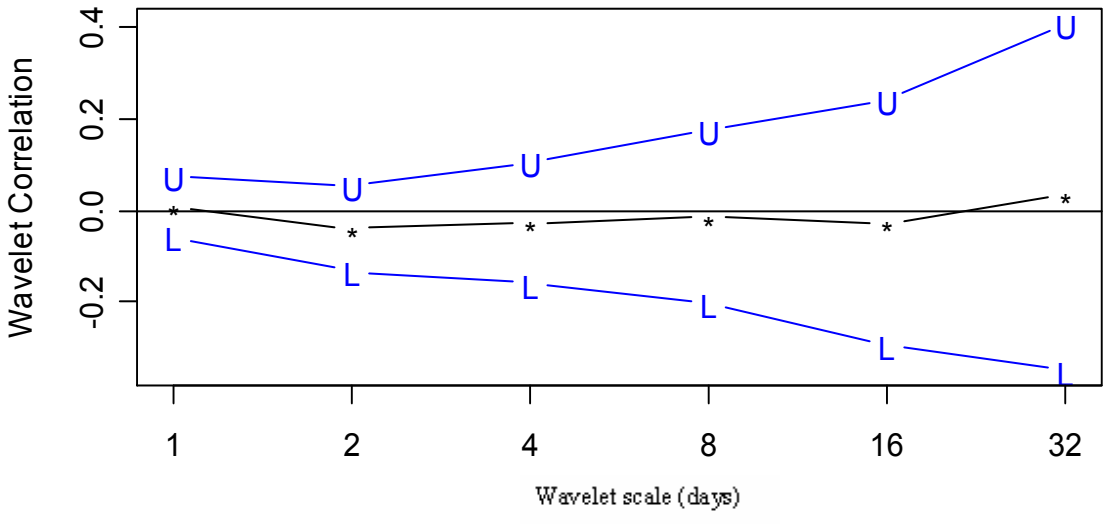

Figure 3.20: Wavelet correlation of conductivity (S) and water temperature (°C) with db4 wavelet.

The wavelet cross correlation provides insight into the lead-lag relationship between two signals on a scale by scale basis. Introducing a time lag of τ so as to investigate the lead lag relationship between the two signals gives the wavelet cross correlation which is given by

$$\rho_{x,\tau}(s_j) = \frac{\gamma_{x,\tau}(s_j)}{\sigma_1(s_j)\sigma_2(s_j)}$$

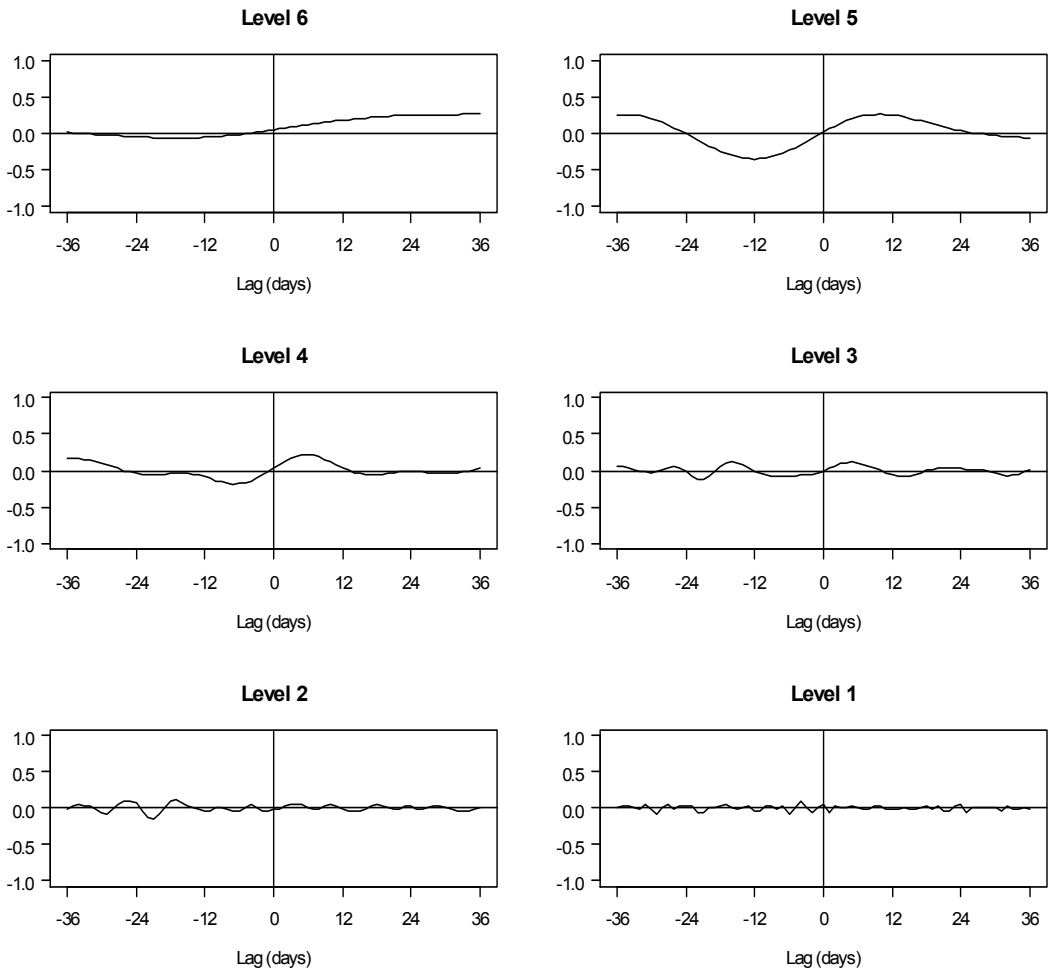

Figure 3.21: Wavelet cross-correlation of conductivity and water temperature with db4 wavelet.

The lead-lag relationship from one scale to the next as unveiled by figure 3.21 is quite interesting for investigating the scale at which the greatest correlation occurs and the delay. If this scale has a significant or the dominant influence on the long term tendency of the time series, the delay will significantly affect the relationship between the indicators.

The ability of wavelet analysis to decompose a time series into components at different time scale makes it an effective tool for water quality signal analysis. Despite its merits, it is not able to replace the classical signal analysis methods.

A practical combination of these methods for the analysis of freshwater quality indicators will significantly give a very deep insight to the time varying structure of the investigated indicators as well as their interrelationship.

4 Application to Water Quality Indicators

Managing freshwater ecosystems should be done on the basis of information extracted from freshwater quality indicators. The indicators need to be analyzed by means of appropriate methods so as to extract as much information as possible. Some of the analysis may reveal inconsistencies in the data structure such as missing values which have to be dealt with, non conformity to the assumptions of the statistical technique one desires to apply on the data as well as reveal the most appropriate technique or model for a particular signal. In the first part of this chapter, the methods covered in chapter three will be applied to a chemical, biological and a physical water quality indicator of the River Havel in Germany. This enables one not only to investigate the information that can be derived from a signal by using the methods, but also to compare the structural differences that exist between these different types of water quality indicators. In the second part of the chapter, a water quality indicator, dissolved oxygen from three rivers namely River Elbe, Havel and Oder will be investigated using the same methods. The intention is to detect similarities and differences that exist in an indicator from different freshwater bodies in the same watershed. The emphasis is on the useful information that can be extracted so as to enhance the management of freshwater ecosystems which are of paramount importance.

4.1 Study area

The River system under study is that of the Havel, which belongs to the greatest tributaries on the right hand site of the River Elbe. The length of the river is about 325 km. It is characterized by a very small elevation difference between source (63 m above sea level) and mouth into the River Elbe (22 m above sea level) (Gnauck and Luther, 2003). For low-flow situations a slope of the water level of 2 cm/km was observed. The watershed is characterized by

shallow lakes, wetlands and marshy country, as well as by high evaporation rates. Only 25% of precipitation contributes to flow. Hydraulic works and banked-up water levels influence water flow and the intensity and kinetics of nutrient dynamics along the course of the river. The active sediment layer is given by 2 cm to 15 cm. Up to 1990 the water quality of the River Havel was characterized by high eutrophication rates (Klose, 1995). After 1990 the nutrient concentrations of effluents of sewage water treatment plants are diminished according to German environmental laws. Furthermore, a reduced usage of fertilizers as well as changes in the land use of agricultural areas has diminished the amount of nutrients from diffuse sources. Figure 4.1 is the map of the investigation area. The monitoring stations are denoted by the notations Tek, SPK, HK, HV, Nu for the canals and Rivers present followed by different numbers for the respective sampling points along them.

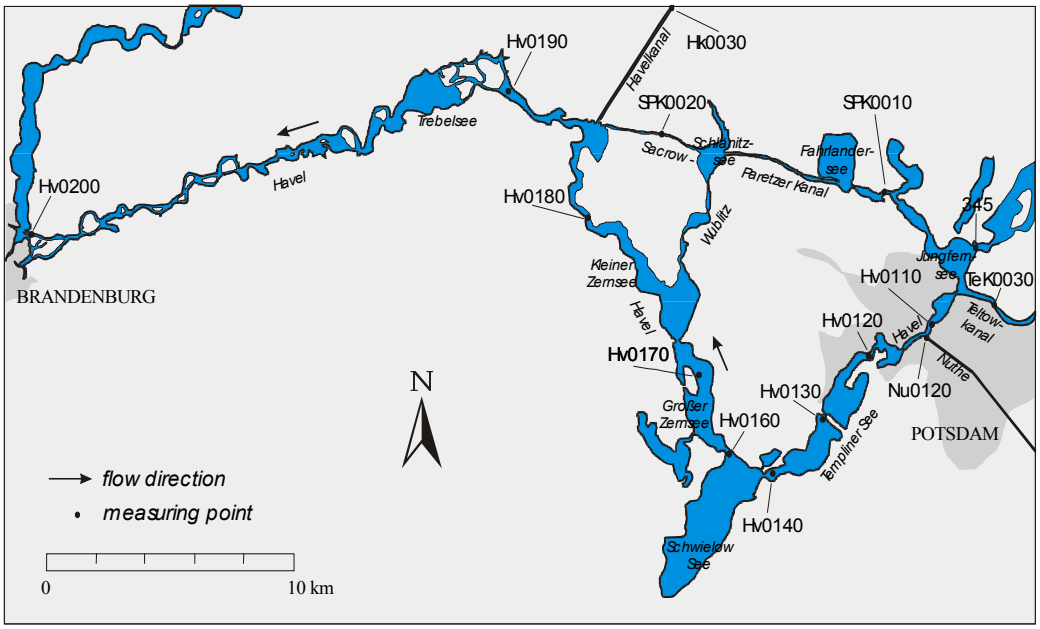

Figure 4.1: Map of the Lower Havel River, Germany (from, Gnauck and Luther, 2003).

4.2 Water quality signals and their analysis

This section of the chapter uses dissolved oxygen, chlorophyll-a and water temperature time series of the River Havel and the second section makes use of dissolved oxygen from Rivers Elbe, Havel and Oder for the investigations. The reason is because these indicators give a first impression of the quality, state of the freshwater body and an idea of the types of organisms it may sustain and covers the different types of water quality indicators. The data was collected by means of automatic sensors with a sampling frequency of ten minutes and spans from 1998 till 2002 for the river Havel and from 1997 to 2000 for river Elbe and Oder. Missing data were replaced by means of linear interpolation and outliers were removed. From previous analysis on the data by Alegue and Gnauck (2006a), it was found necessary to resample the data to a daily interval before investigation.

The plot of dissolved oxygen in figure 4.2 reveals the existence of yearly cycles with high irregular fluctuations throughout the year. The highest concentrations occur during the warm periods of the year as a result of the high algal biomass present producing oxygen.

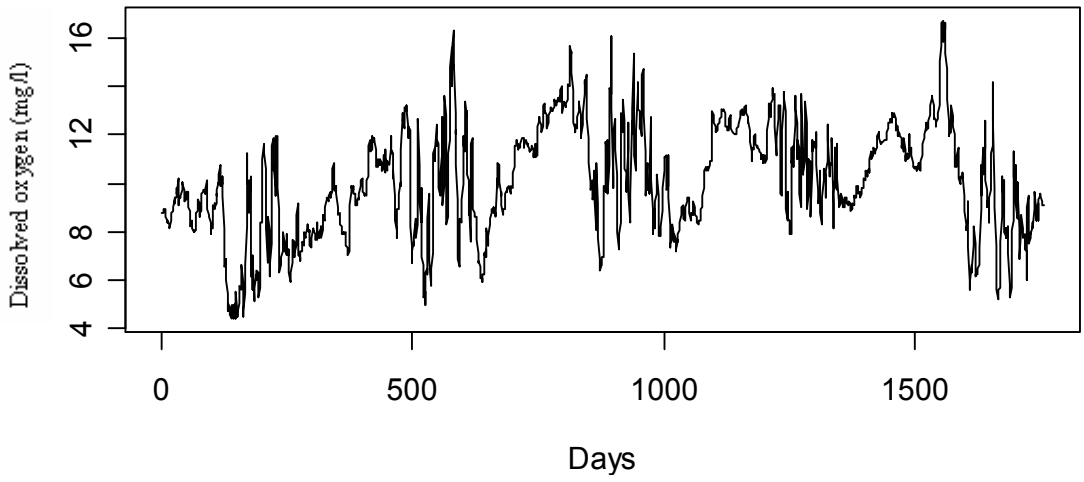

Figure 4.2: Run sequence plot of dissolved oxygen from River Havel.

The chlorophyll-a plot (cf. figure 2.2 (middle)) portrays a major yearly cycle and two shorter significant cycles within the yearly cycle. The yearly cycle is as a result of the yearly seasonal variations while the two shorter cycles are as a result of the life cycle of the two dominant algal species present. These are namely diatoms and cyanobacteria which appear during spring and summer respectively. The water temperature signal (cf. figure 2.2 (top)) clearly reveal a distinct yearly cycle as a result of the seasonal driving forces namely solar radiation and air temperature.

Autocorrelation

The correlogram of dissolved oxygen signal in figure 4.3 (top) reveals the existence of a strong autocorrelation between the dissolved oxygen values. This simple means the values are not random with the existence of dependence between the present dissolved oxygen values with prior values. This is referred to as the exhibition of long memory where the present values still has a memory of the prior values. The reason for this strong autocorrelation structure is because the concentration of dissolved oxygen does not rapidly change within a short time interval. The correlogram does not unveil any periodic behavior probably as a result of the strong autocorrelation between the values. Such a persistent autocorrelation structure may affect modeling results (Straškraba and Gnauck, 1985), so it has to be dealt with before modeling begins.

The correlogram of chlorophyll-a in figure 4.3 (middle) also presents the existence of a strong autocorrelation or persistence present in the data structure. This also indicates the absence of randomness in the time series data and the existence of dependence between present values and prior values of chlorophyll-a concentration. The decay is very slow as the time lag increases; hence, long memory is exhibited by the signal whereby present values have a memory of past prior value. This slow decay is because the algal biomass changes slowly within a short period of time. The driving forces

responsible for the algal biomass do not also changes rapidly within short time intervals. However, the autocorrelation of chlorophyll-a is stronger than that exhibited by the dissolved oxygen signal. This implies that the sampling frequency of dissolved oxygen requires a higher sampling frequency than that of chlorophyll-a.

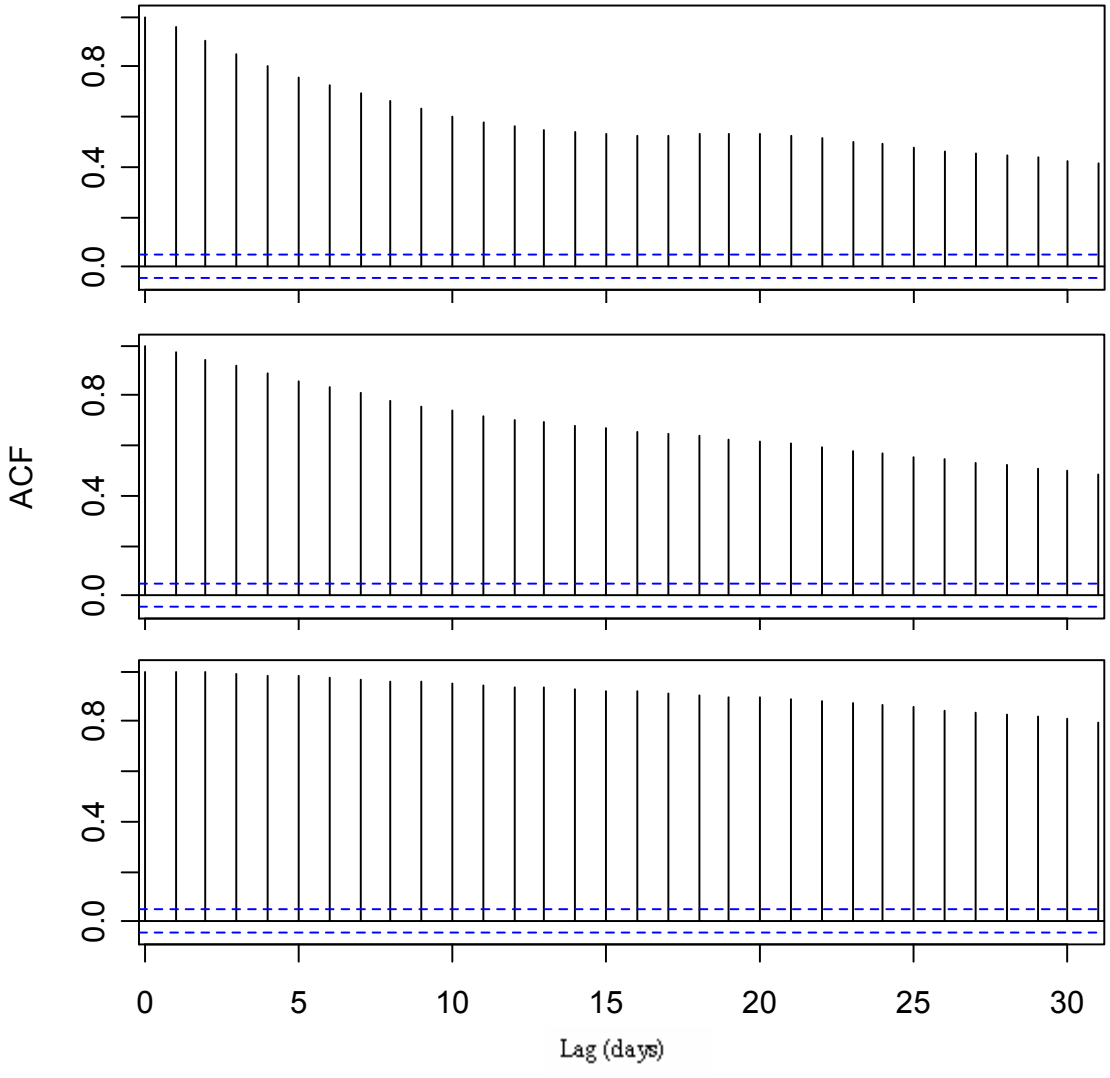

Figure 4.3: Correlogram of dissolved oxygen (top), chlorophyll-a (middle) and water temperature (bottom) of River Havel.

The correlogram of water temperature in figure 4.3 (bottom) unveils the existence of an extremely strong autocorrelation or persistence between the values of water temperature. This very strong persistence means the values are not independent, but exhibit a long memory with the present values strongly influence by the prior values like in the case of dissolved oxygen and chlorophyll-a signals. The correlogram is not able to indicate the presence of the cyclic behavior present in the signal due to the very strong persistence between subsequent values and very slow decay as the time lag increases. This indicates that the temperature of the water body does not rapidly change within short time intervals. This stability is important for aquatic organisms which may be killed by rapid temperature fluctuations.

All the three indicators portray a strong autocorrelation or the existence of persistence between present and prior values with the water temperature having the strongest autocorrelation. This implies the present values of the physical indicators are more strongly influenced by the prior value of the biological and the chemical indicators. One of the implications is that the sampling frequency of the physical indicator should be less than that of the biological indicator which should also be less than that of the chemical indicator. The reason behind this persistence is because the underlying driving forces (e.g solar radiation) determining the long term dynamics of these indicators equally change slowly across time.

Covariance and correlation

Figure 4.4 (top) indicates a negative covariance exists between dissolved oxygen and water temperature with a value of - 4.98. This means there is a decrease in the concentration of dissolved oxygen as the water temperature increases and an increase in concentration as the water temperature decrease in this freshwater body. The covariance value is not very informative because it depends on the magnitude of the units of the indicators under investigation. The correlation is the normalized version of the covariance and is very informative because it varies between -1 and +1. The correlation in this case is

-0.31 reveals the negative correlation is not so strong. The reason for the negative correlation is that at a given pressure, the amount of oxygen a water body can hold decreases as the temperature increases. Hence, a water body is more rapidly saturated of dissolved oxygen during high temperature with the dissolved oxygen escaping through simple diffusion into the atmosphere. The blue line is a fitted ordinary least square regression line. The strong cluster observed between 0 to 10 °C and 15 to 25 °C is as a result of the proliferation of diatoms and cyanobacteria biomass releasing oxygen at these temperatures respectively.

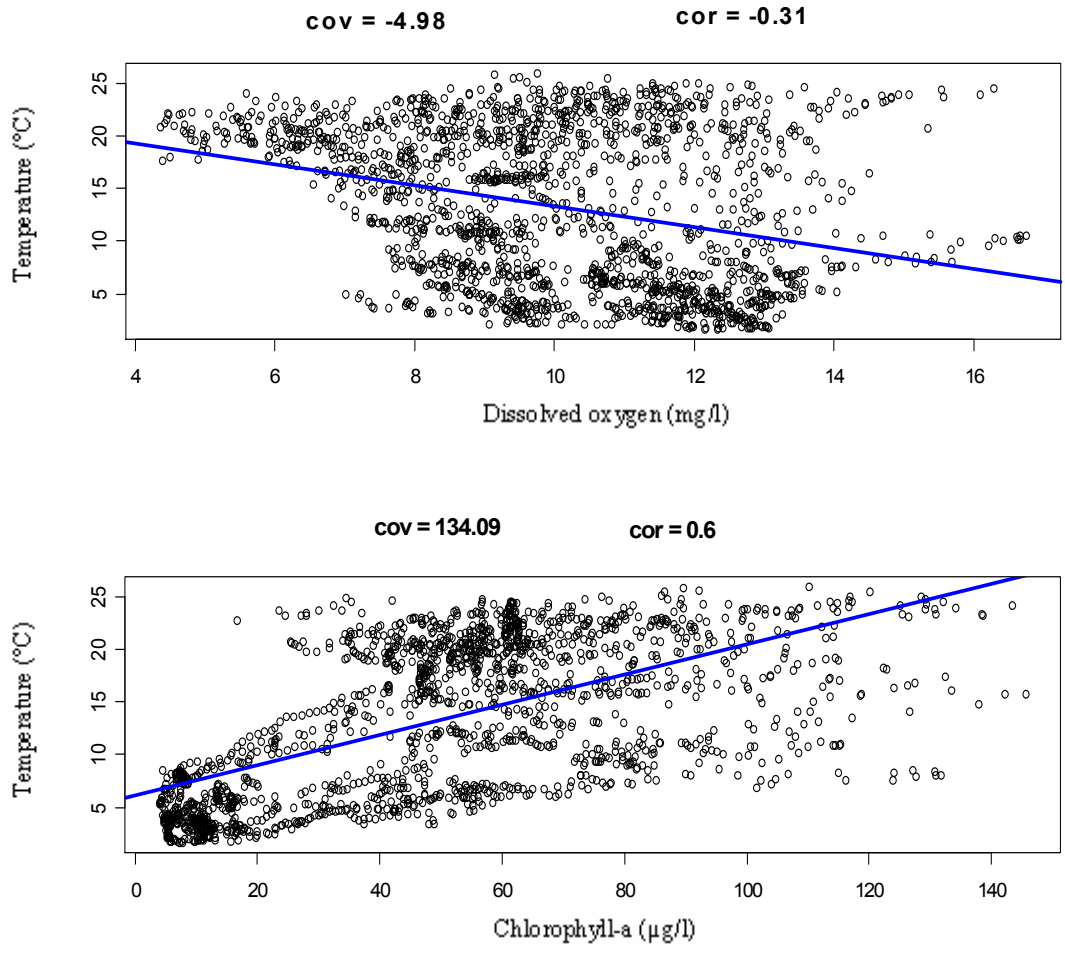

Figure 4.4: Covariance and correlation of dissolve oxygen (top) and chlorophyll-a (bottom) with water temperature.

Figure 4.4 (bottom) reveals the existence of a positive covariance between chlorophyll-a and water temperature indicators having a value of +134.09. This indicates that the concentration of biomass in the water body increases as water temperature increases. This is confirmed by the strongest cluster occurring between 0 and 10 °C and 15 to 25 °C as a result of the proliferation of diatoms and cyanobacteria biomass at these respective temperatures. Hence, the highest concentrations of biomass will be found in the water body during the warmer months of the year due to the availability of light and a suitable temperature necessary for biomass production by photosynthesis. This is also because the amount of nutrients content available for biomass production is significantly greater during the warmer periods of the year. The high value of the covariance is not quite informative because it depends on the magnitude of the units of the indicators. A normalized version of the covariance known as the correlation with values varying between -1 to +1 was used and the value of 0.6 was obtained indicating a relatively weak correlation between biomass and the water temperature. The blue line is simply the regression line between chlorophyll-a and water temperature fitted by ordinary least square method.

The co-variation of the chemical and biological signal with water temperature is in different direction. The physical signal significantly influences the chemical and biological indicators in different directions but is not influenced by them.

Cross-correlation

The lead-lag relationship between dissolved oxygen and water temperature as revealed by figure 4.5 (top) indicates that there exists a time lag of about 12 days for the highest correlation to be observed with the correlation being negative. The correlation of -0.31 is not the highest value that exist. At a lag of 12 days, the correlation between the two indicators is -0.4. This can also be explained by the fairly strong autocorrelation structure of dissolved oxygen which does not rapidly change over a short time interval. Hence, it requires 12 days for an increase in temperature to cause the highest decrease in dissolved

oxygen concentration, given that the correlation is negative. This has to be taken into consideration in modeling, planning and management effort which takes into consideration water temperature and dissolved oxygen.

The lead-lag relationship between chlorophyll-a and water temperature presented in figure 4.5 (bottom) indicates that there exist a positive cross-correlation with a maximum value of +0.6 at lag zero with a decrease in correlation as the time lag increases. Hence, the effect of an increase in water temperature decreases as the time lag increases. From another perspective, it indicates the strongest variation in the same direction of chlorophyll-a concentration as the temperature of water increase, occurs at lag zero and decreases as the time lag changes.

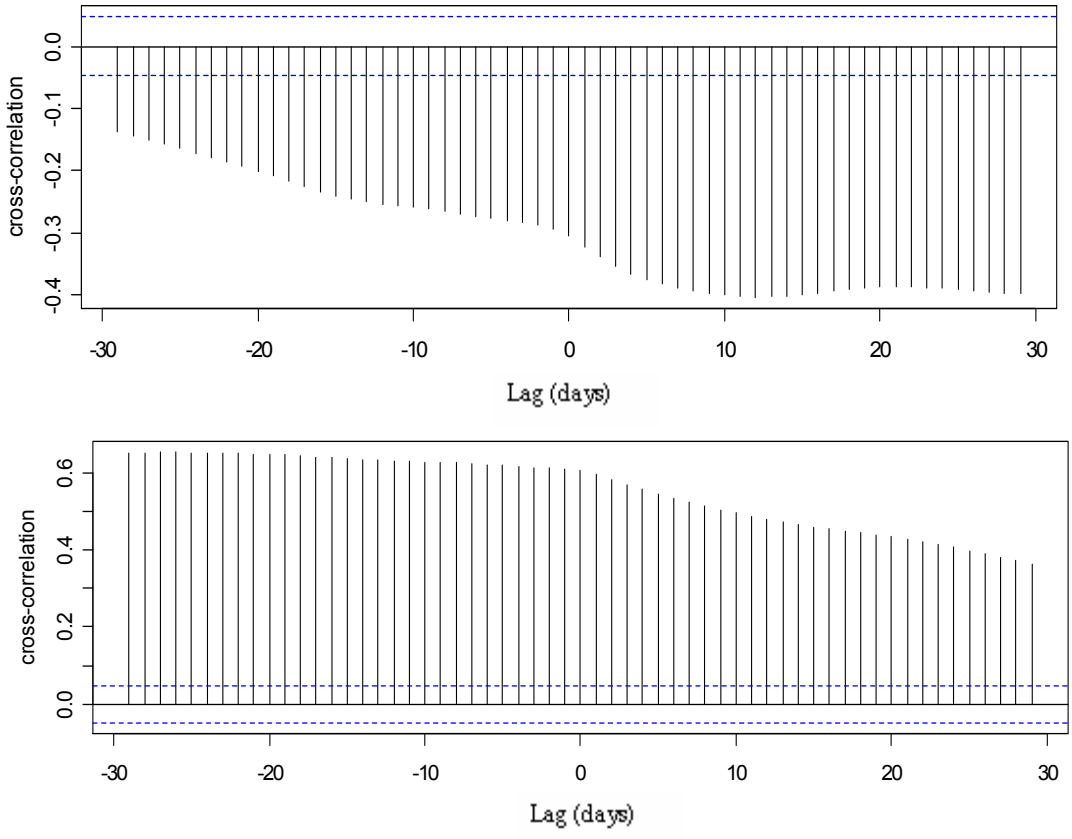

Figure 4.5: Cross-correlation of dissolved oxygen (top) and chlorophyll-a (bottom) with water temperature.

Periodogram

The periodogram of dissolved oxygen in figure 4.6 (top) indicates that the strongest variations in the signal occur in the low frequency region. Hence, the long term tendency of the dissolved oxygen curve is exclusively influenced by the low frequency variations in the time series. This dominance of low frequency components confirm the findings that changes in the signal occur slowly and not rapidly within a short time interval. The periodogram also reveals two major peaks with one of them having a higher intensity and a lower frequency that the other. The lower frequency indicates a longer period and the higher intensity indicates and stronger influence on the long term tendency. The first peak representing the longer period may be the influence of seasonal variations in water temperature. The lower peak indicates a shorter cycles which occurs between the seasonal cycles as a result of the two species of algal biomass present in the water body. The other parts of the periodogram reveal more or less variations which have no significant effect on the long term tendency of the dissolved oxygen time series.

Figure 4.6 (middle) gives the periodogram of chlorophyll-a and reveals that the variations in the time series have their highest intensity in the low frequency bands. This means the long term dynamics of this signal is exclusively governed by the low frequency variations. The periodogram also reveals the presence of one major and one smaller peak indicating cycles or periodic components with different periods. The peak with the highest intensity has the lowest frequency, hence the longer period of about one year revealing the seasonal cycle present in the chlorophyll-a time series. The longer peak is influenced by seasonal changes in water temperature. The smaller peak has a shorter period and is caused by the cycle introduced as a result of the life cycle of diatoms and cyanobacteria, which are the dominant group of the phytoplankton biomass present in the freshwater body. Actually, of the two dominant groups present, the diatoms appear in spring while the cyanobacteria are present in summer. Some insignificant high frequency changes are present, but are not able to influence the general tendency of the chlorophyll-a signal.

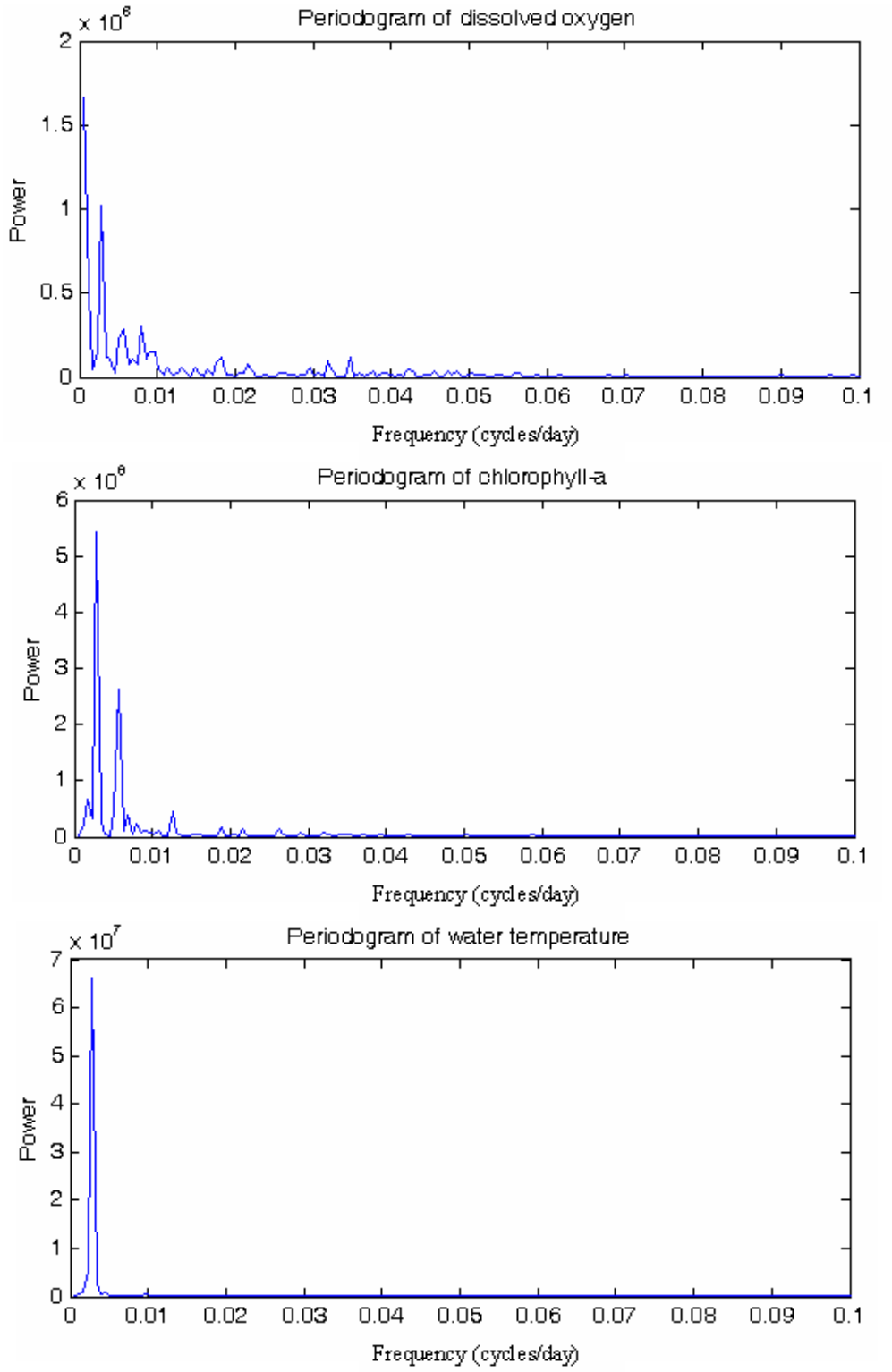

Figure 4.6: Periodogram of dissolved oxygen (top), chlorophyll-a (middle) and water temperature (bottom).

The periodogram of water temperature (figure 4.6 (bottom)) clearly indicates that the variations influencing the long term behavior of the water temperature signal are essentially low frequency variations. There are very little random fluctuations with little or no high frequency behavior having a noticeable intensity. It also reveals the presence of a clear periodic behavior indicated by the sharp peak of high intensity present. The periodic or cyclic behavior is as a result of seasonal variations of the external driving forces that influence water temperature, namely solar radiation and air temperature. These seasonal cycles influence the behavior of the other freshwater water quality indicators because most freshwater processes and reactions are temperature dependent. Any increase or decrease in the intensity of the periodogram or any shift in the frequency bands will automatically have an influence on most water quality indicators in this fresh water body.

Squared coherency and phase spectra

For the squared coherency spectrum of dissolved oxygen and water temperature given in figure 4.7 (top) to be interpreted, the periodogram of the two signals must be examined to determine the frequency bands of interest. In this case, the frequency bands of interest derived from the periodograms of dissolved oxygen and water temperature are from zero to 0.01. The higher frequency bands show negligible intensity in their variation, making it totally senseless to examine the squared coherency of the high frequency bands. The squared coherence, like an r^2, is an estimate of the percentage of variance in dissolved oxygen, within the frequency bands of interest, which is predictable from the variance in water temperature within the same frequency band. Within the band of interest, a maximum coherence of 0.44 is observed, implying that 44% of the variance in the dissolved oxygen signal is predictable from the water temperature signal. The higher frequency bands are of no interest because the variations present in these bands have no significant effect on the long term dynamics of the dissolved oxygen signal; hence, these frequencies are of no particular interest for squared coherency analysis.

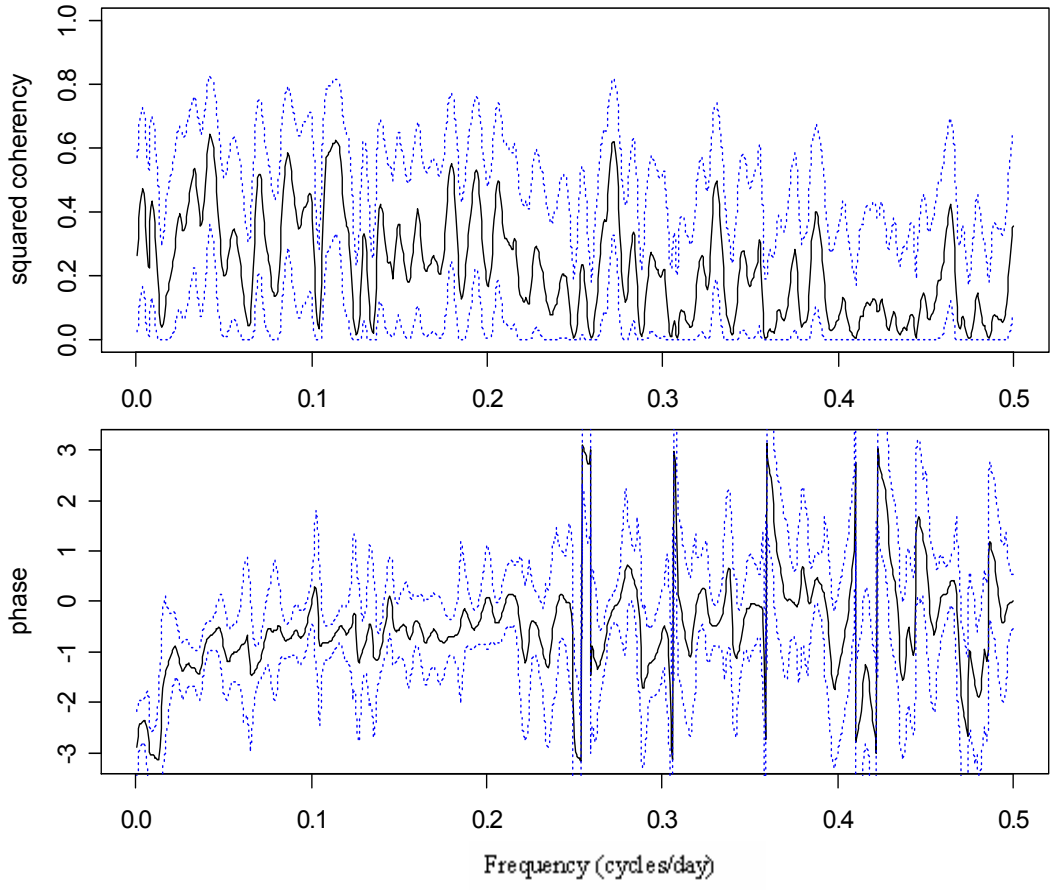

Figure 4.7: Squared coherency (top) and phase (bottom) spectra of dissolved oxygen with water temperature.

The phase spectrum of dissolved oxygen and water temperature portrayed in figure 4.7 (bottom) equally needs to be interpreted taking into consideration the frequency bands of the squared coherency which are of interest. These are the bands ranging from zero to 0.01 where the squared coherency is high. The phase within these bands is negative with the lowest value being -3. Hence, the highest coherence between dissolved oxygen and water temperature exist when a phase difference of -3 occur. From this analysis, the squared coherency and the phase spectra enables one to conclude that there is the existence of synchronized cycles between the dissolved oxygen and water temperature signals as indicated by figure 4.7.

The squared coherency spectrum of chlorophyll-a and water temperature shown in figure 4.8 (top) needs to be interpreted, using the periodograms of the two signals so as to determine the frequency bands of interest. In this case, the frequency bands of interest derived from the periodograms (cf. figure 4.6) are from zero to 0.01. The higher frequency bands show negligible intensity in their variation, making it totally senseless to examine the squared coherency of the high frequency bands.

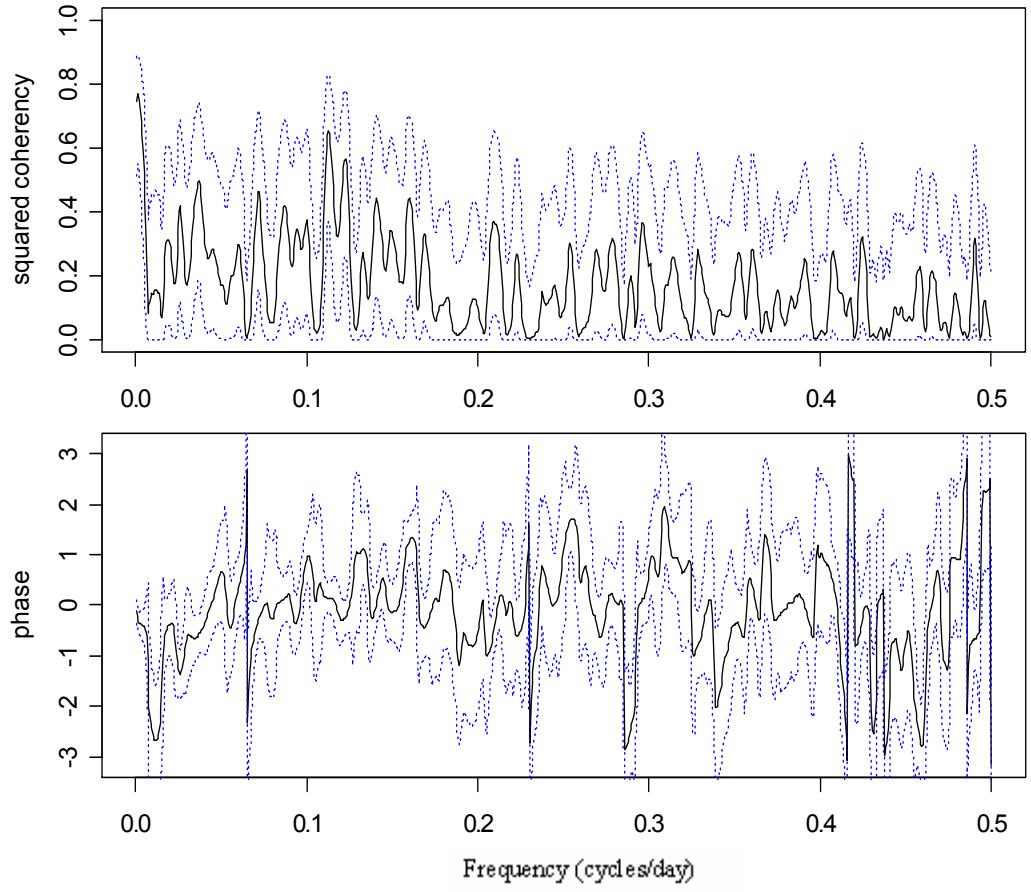

Figure 4.8: Squared coherency (top) and phase (bottom) spectra between chlorophyll-a with water temperature.

The squared coherency is high within the bands zero to 0.01, which are bands of interest, with a maximum coherence being 0.75. This implies that within the

frequency bands of interest, 75 % of the observed variance of the chlorophyll-a signal is predictable from the water temperature signal. The higher frequency bands are of no interest because the variations present in these bands have no significant effect on the long term dynamics of the chlorophyll-a signal; hence, these frequencies are of no particular interest for squared coherency analysis.

The phase spectrum of figure 4.8 (bottom) equally needs to be interpreted taking into consideration the frequency bands of the squared coherency which are of interest. These are the bands ranging from zero to 0.01 where the squared coherency is high. The phase within these bands is zero. Hence, the highest coherence between chlorophyll-a and water temperature exist when a phase difference of zero. From this analysis, the squared coherency and the phase spectra enables one to conclude that there is the existence of synchronized cycles between the chlorophyll-a and water temperature signals as indicated by figure 4.8.

Fourier polynomial

The Fourier polynomial makes use of the Eulerian identity, which does not take into consideration changes in frequency of the sine and cosine basis function across time. Approximating natural signals with this polynomial poses certain problems due to the basic assumptions. The approximation requires the frequencies and amplitudes of the constituent sinusoids to be stationary (constant mean and variance across time). This is not the case of fresh water quality signals which exhibit changing frequency, amplitudes and phase shift across time requiring methods such as wavelets.

The Fourier polynomial is not quite successful in attempting to approximate the dissolved oxygen signal by making use of a combination of sine and cosine sinusoidal functions. Several small fluctuations are present and affect the result of the Fourier approximation presented in figure 4.9. Eight terms are used, but the coefficient of determination is very low and may be as a result of the multiplicity of internal and external driving forces operating at different frequencies and time scales. Increasing the number of terms give no

significantly better result. The Fourier polynomial of seventh order is given by the following equation:

$$f(t) = 10.23 - 0.13\cos(0.006t) - 0.75\sin(0.006t) - 0.41\cos(0.012t) - 0.52\sin(0.012t) - 0.26\cos(0.018t) + 1.2\sin(0.018t) + 0.2\cos(.024t) - 0.27\sin(0.24t) - 0.38\cos(.03t) - 0.07\sin(0.03t) - 0.58\cos(0.036t) + 0.53\sin(0.036t) - 0.16\cos(0.042t) + 0.39(0.042t).$$

The goodness of fit delivers $r^2 = 0.29$ and adjusted $r^2 = 0.28$. This means, that approximately 30 % of signal variations are described by a Fourier polynomial with fixed periodicity.

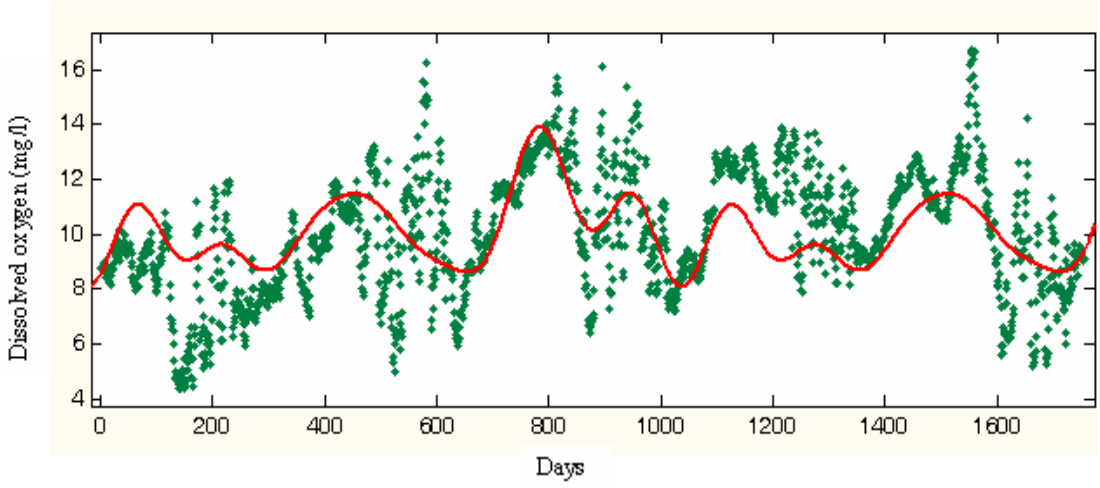

Figure 4.9: Seventh order Fourier polynomial of dissolved oxygen.

The Fourier approximation, shown in figure 4.10 requires several sinusoidal terms to obtain a decent result. This is because several internal and external driving forces are responsible for the behaviour portrayed by the chlorophyll-a signal. Increasing the terms does not yield a significantly better result. The best approximation was got by a Fourier polynomial of fourth order with $r^2 = 0.68$ and adjusted $r^2 = 0.67$ with the following equation:

$$f(t) = 48.76 + 4.7\cos(0.0087t)+0.47\sin(0.0087t) - 29.64\cos(0.02t) - 1.4\sin(0.02t) - 2.48\cos(0.03t) + 1.3\sin(0.03t) - 20.35\cos(0.04t) + \sin(0.04t)$$

As can be seen, the Fourier approximation describes the time behaviour of the signal more or less correctly during the transient reaches from winter to summer and vice-versa. High amplitudes in spring due to an algal bloom of diatoms and in summer caused by an algal bloom of cyanobacteria as well as low amplitudes in winter are not correctly approximated.

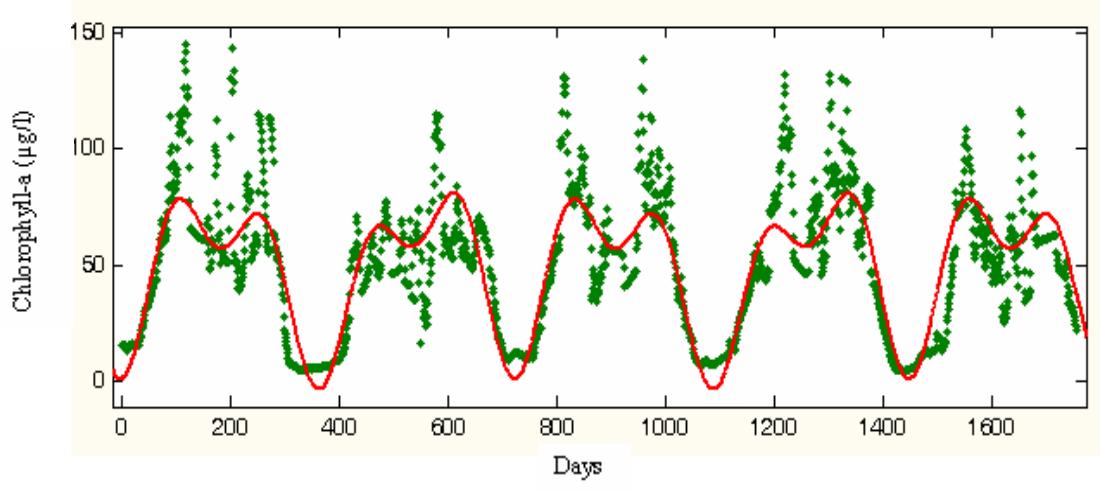

Figure 4.10: Fourth order Fourier polynomial of chlorophyll-a.

While signal levels of amplitudes and phases vary from year to year the periodic rhythm of the approximating polynomial remains constant. This fact leads to differences in the time courses between the real signal and its approximation, and, therefore, to a reduced predictive power of the polynomial (Gnauck, 2006).

Figure 4.11 reveals a decent result for the Fourier approximation for water temperature. Being a physical water quality indicator, its main driving force is solar radiation which also has a distinct cyclic behaviour. The Fourier approximation of third order is given by

$$f(t) = 13 + 0.39\cos(0.009t) + 0.018\sin(0.009t) - 9.71\cos(0.018t) - 2.47\sin(0.018t) - 0.4018\cos(0.027t) + 0.23\sin(0.027t)$$

where $r^2 = 0.95$ and adjusted $r^2 = 0.95$. Approximating physical indicators by means of this method often gives a high coefficient of determination with the chemical indicators giving the lowest coefficient of determination for this freshwater ecosystem (Gnauck, 2006).

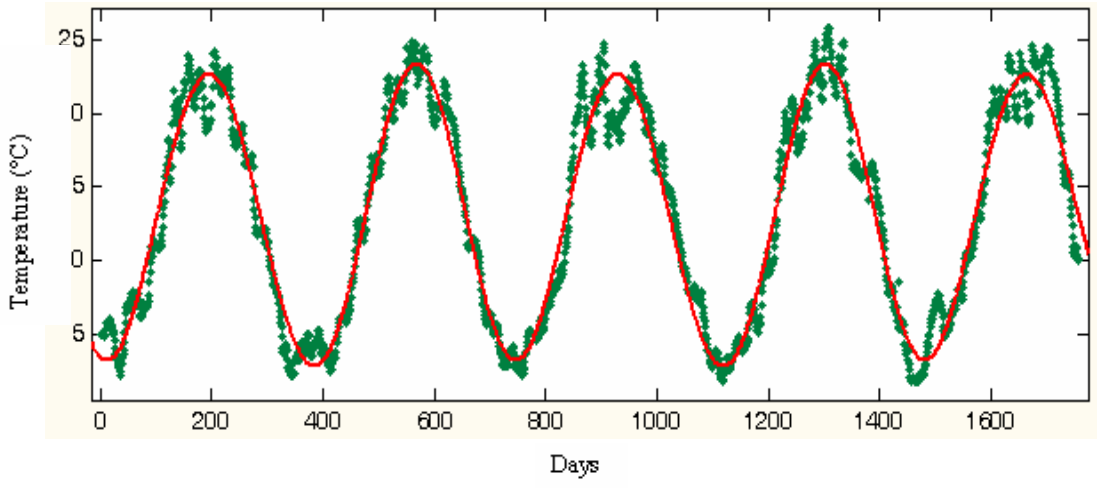

Figure 4.11: Third order Fourier polynomial of water temperature.

Wavelet multiresolution analysis

The multiresolution analysis provided in figure 4.12 using Daubechies 4 wavelet family at level 7 shows an additive decomposition of the dissolved oxygen signal. The approximations reveal a progressively finer version of the signal from level 1 to level 7 or the low frequency variations present in the dissolved oxygen time series. It acts as a lens unveiling a finer and finer resolution from one level to the next, revealing a lesser and lesser blur version of the signal as we move to the higher levels with a progressive elimination of high level fluctuations. No long term trend is present in the signal, though there exist a fall in the concentration of dissolved oxygen during the year 1998

probably due to some pollution event, with a subsequent recovery to more or less normal value during the subsequent years. This can be clearly seen at level 7 which is a version of the signal without the high frequency fluctuations of scale 1, 2, 4, 8, 16, 32 and 64 (approximately two months.).

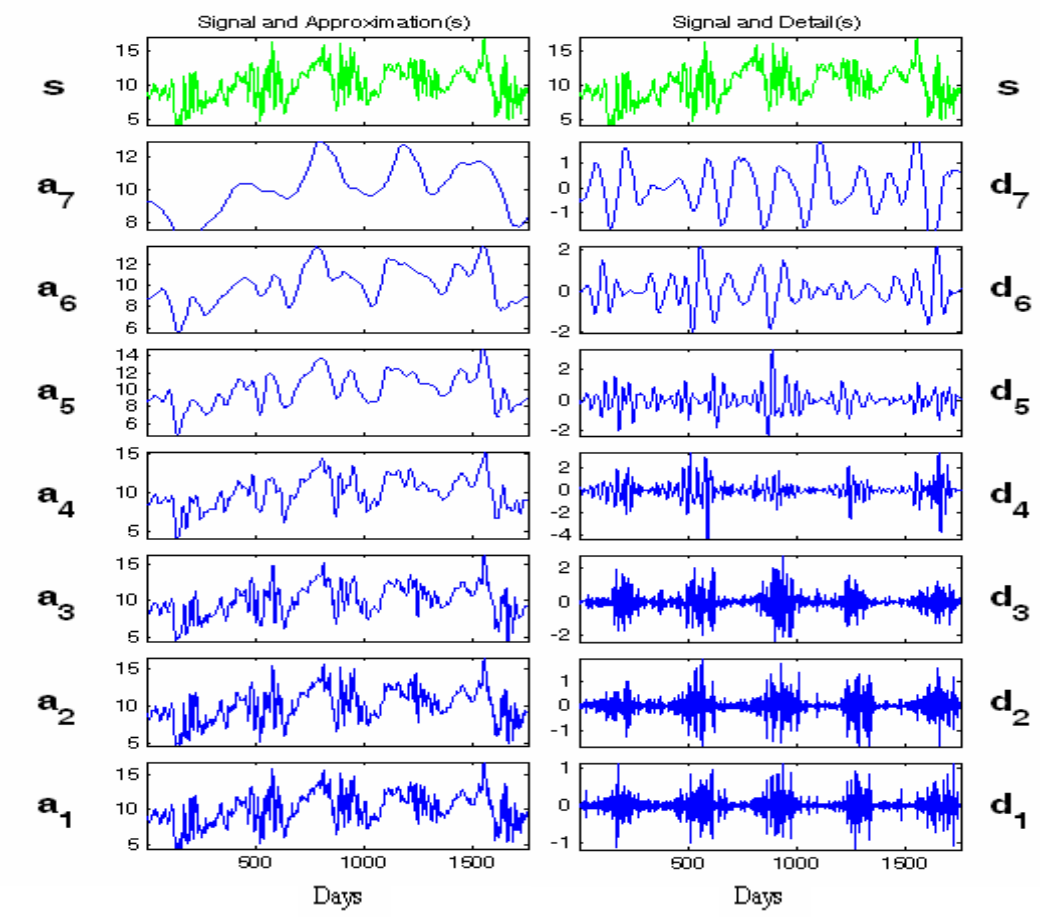

Figure 4.12: Multiresolution analysis of dissolved oxygen with db4 at level 7.

The details reveal the high frequency variations or changes present in the dissolved oxygen time series by providing an additive decomposition of the high frequency fluctuations on a scale by scale basis. It reveals the daily changes (level one), two day variations (level 2), 4 days variations (level 3), 8 days variations (level 4), 16 days variations (level 5), 32 days variations (level 6) and 64 days variations (level 7). This progressive decomposition reveals the existence of differences in the magnitude of the fluctuations from one scale to

another. It effectively shows that the lower scales have changes of relatively lower intensity compared to the higher scales of variation. The implication is that, the long term tendency observed in the dissolved oxygen time series is significantly influenced only by the fairly strong changes occurring at the higher scales and not the lower scales. At the lower scales, the fluctuations are greater during the warmer months than during the colder months. At the higher scales such as scale 64, the fluctuations are high throughout the year. It is quite interesting to examine the variance at different scales.

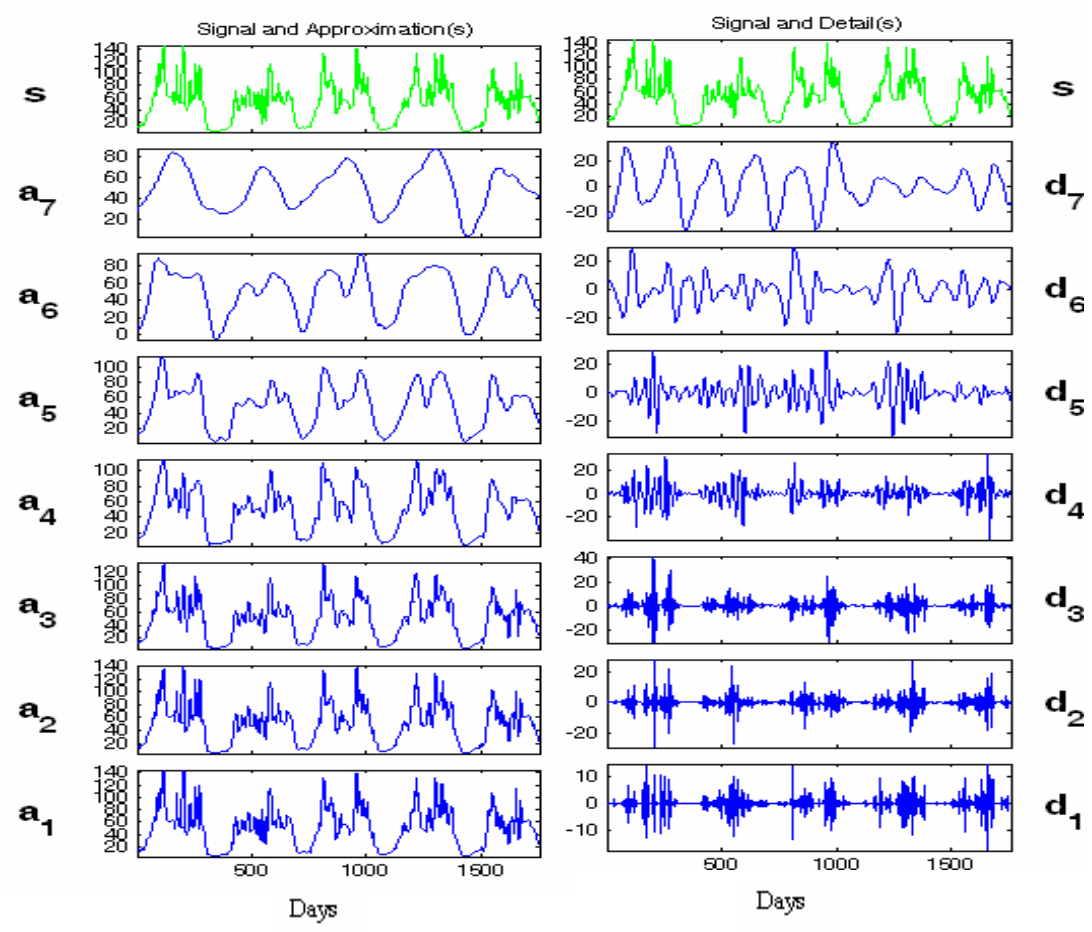

Figure 4.13: Multiresolution analysis of chlorophyll-a with db4 at level 7.

The multiresolution analysis presented in figure 4.13 using Daubechies 4 wavelet family at level 7 unveils an additive decomposition of the chlorophyll-a signal. The approximations reveal a progressively finer version of the signal

from level 1 to level 7 or the low frequency variations present in the chlorophyll-a time series. It acts as a lens with a finer and finer resolution from one level to the other revealing a less blur version of the signal as we move to the higher levels with a progressive decrease of high level fluctuations. No long term trend is present in the signal as can be clearly seen from level 4 upwards. At level 5, the shorter cycle resulting from the life cycle of two major phytoplankton groups can clearly be seen as well as the longer cycle resulting from seasonal variation of the indicator.

The details unravel the high frequency variations present in the chlorophyll-a time series or provide an additive decomposition of the high frequency variation. This progressive decomposition reveals the differences in fluctuations from one scale to another. It effectively shows that the lower scales are less important compared to the higher scales of variation. The implication is that, the long term behavior observed in the chlorophyll-a time series is mostly influenced by the fluctuations occurring at the higher scales and not the lower scales. At the lower scales, the fluctuations are more important during the warmer months with no change occurring during the colder months. At the higher scales such as scale 32, the fluctuations are high throughout the year with the fluctuations at level 7 clearly portray the two cycles present in the signal. It is quite interesting to examine the variance at different scales so as to quantify the change in variation from one scale to the next.

The multiresolution analysis of water temperature which is given in figure 4.14 using the Daubechies 4 wavelet family at level 7 also portrays the additive decomposition of the signal at different time scales. From the approximations starting from scale 1 till scale 7, it can be observed that there is no upward or downward trend present in the water temperature time series. Acting as a lens with a progressive increase of the resolution, it offers a lesser and lesser blur version of the signal as the fluctuations are removed in an orderly manner. Clear cycles are observed in the signal as a result of seasonal changes in the intensity of solar radiation, which is the main driving force of this indicator.

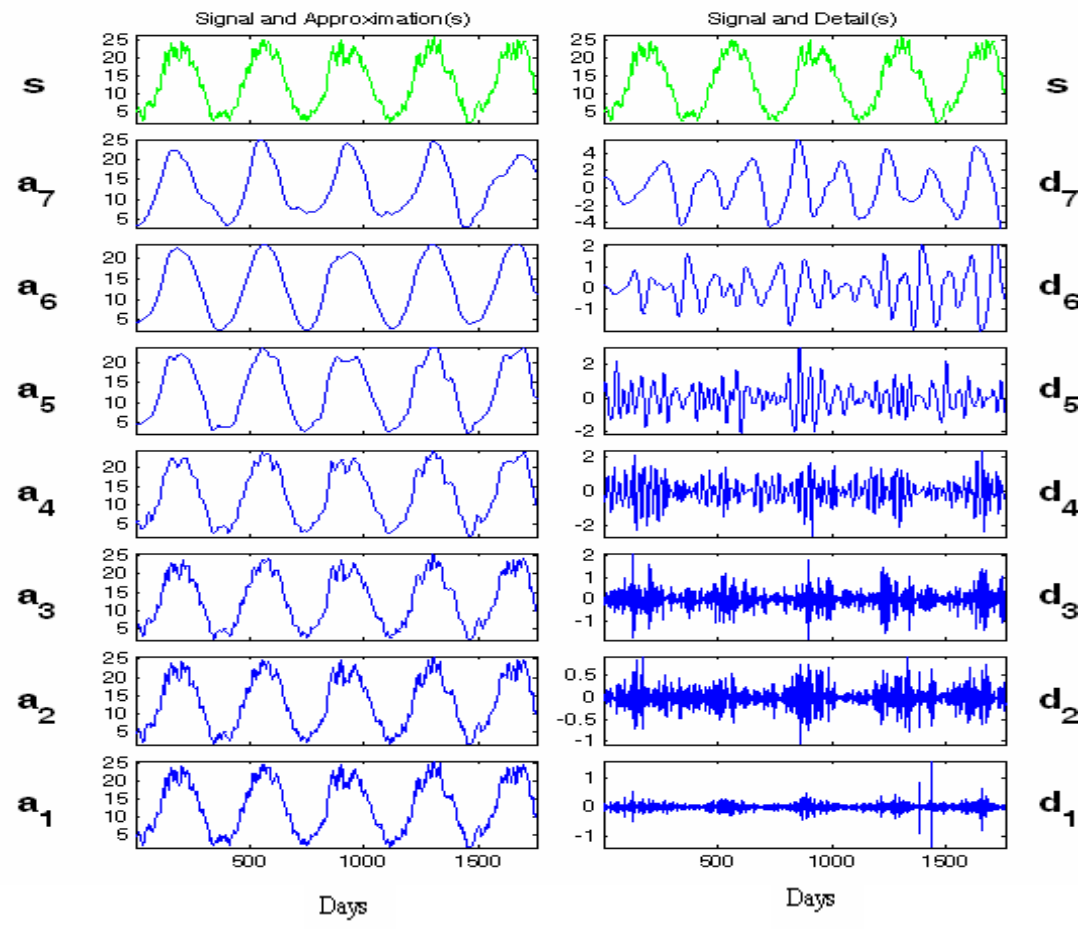

Figure 4.14: Multiresolution analysis of water temperature with db4.

The details give an insight into the high frequency variations and their intensity acting in the signal of water temperature on a time scale basis. The daily fluctuations given by d1 are quite small, indicating the presence of very little changes in water temperature on a daily interval. These changes seem more or less uniform throughout the years with very little changes from one season to the next. As the time scale increases, the intensity of the fluctuations increase with the strongest fluctuations being observed from scale 32 and above or from level 7 which is represented by d7. This means the long term behavior of the water temperature time series in determined by the low frequency components or variations occurring at a monthly interval. This results in the fact that abrupt changes in the environment cannot significantly change the long term behavior

of the water body, therefore acting as a buffer to rapid weather changes. This creates stable conditions and an acceptable environmental for freshwater organisms.

Wavelet variance

The wavelet variance exhibited in figure 4.15 portrays the intensity of the changes from one scale to the other of the dissolved oxygen time series. This graphical representation of the wavelet variance enables the researcher to answer questions concerning the dominant scale of variation in the time series, the homogeneity of variations from one scale to the next, the importance of the variations at one scale compared to the variations occurring at another scale. It also enables one to assess the information gained when the sampling frequency is increased.

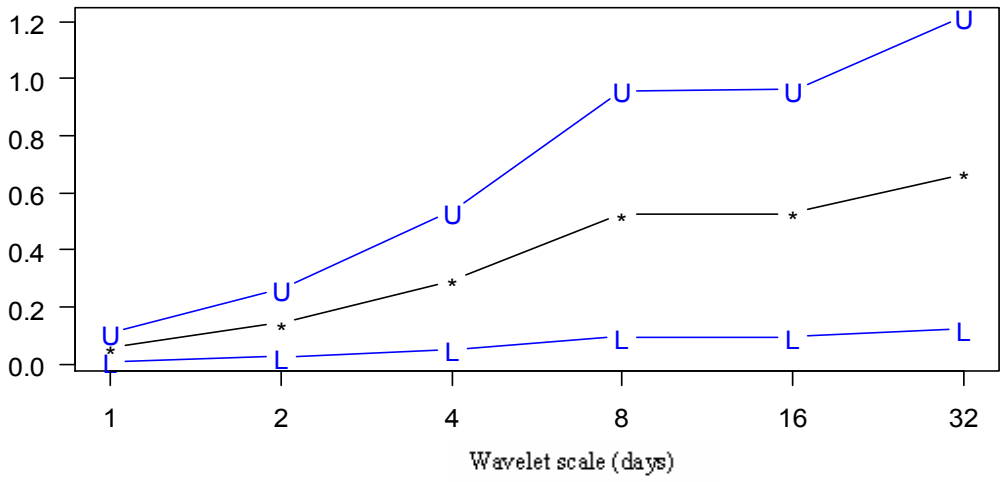

Figure 4.15: Wavelet variance of dissolved oxygen with db4.

Figure 4.15 reveals that the variation in the time series increases progressively till scale 8 where a local maximum can be observed. There is no significant increase in the variations that occur after this scale. This has important implications regarding the monitoring of dissolved oxygen time series in this fresh water ecosystem. Monitoring at scales lower than 8 is good, but give

redundant information because the time series will contain information on fluctuations that have no significant influence on the long term behavior of the time series. Monitoring at scales above 8 will cause a loss of important information that has an influence on the long term behavior of the signal contained in scale 8. Hence, the optimal scale of monitoring is 8 for this indicator and this freshwater ecosystem.

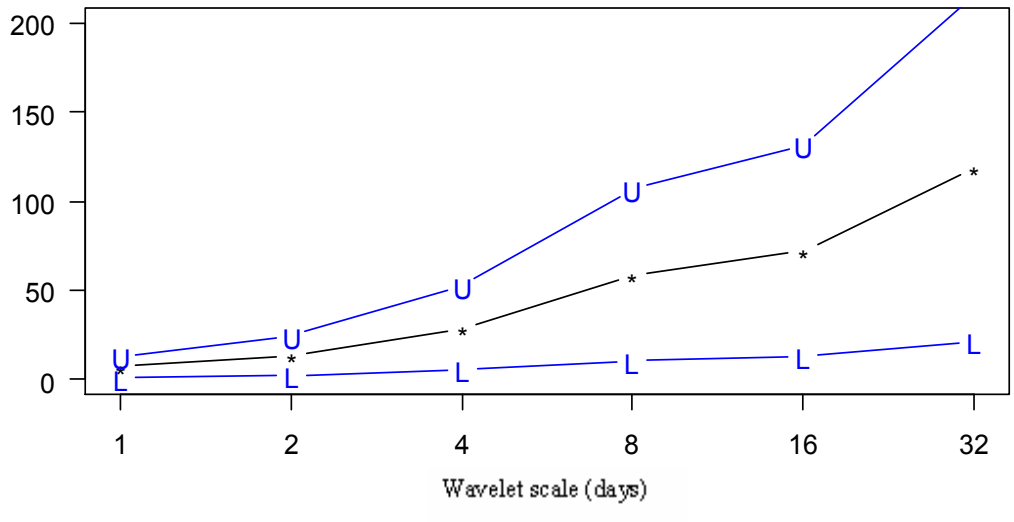

Figure 4.16: Wavelet variance of chlorophyll-a with db4.

The wavelet variance shown in figure 4.16 reveals the intensity of variation from one scale to the other of the chlorophyll-a time series. It is observed that the higher scales are the dominant scales of variation with the absence of any significant changes occurring at the lower scales. Hence, the variations are not homogenous across the different time scale. The variation in the time series increases consistently and linearly till scale 8 where it seems to stabilize till scale 16, then it increases significantly from scale 32. To effectively interpret this plot of wavelet variance, the approximations from the wavelet multiresolution analysis must also be examined. The approximations at level 4 (scale 8) and 5 (scale 16) clearly shows the presence of the shorter cycles and the longer cycle. From level 6 (scale 32), only the longer cycle is clearly present. The high variance at scale 8 and 16 capture the variations causing the

shorter cycles while the longer cycle provoke the variance at scale 32 which is stronger producing the longer cycle.

This has important implications regarding the monitoring of chlorophyll-a time series in this fresh water ecosystem. Monitoring at scales lower than 8 is good, but give redundant information because the time series will contain information on fluctuations that have no significant influence on the long term tendency of the time series. Monitoring at scales 8 or 16 will enable the variations causing the shorter and the longer cycle to be captured while monitoring at scale 32 will capture the variations causing the seasonal cycle, but not the shorter ones. This will result in a loss of important information on the shorter cycles that has an influence on the long term behavior of the signal contained in scale 8 and 16. Hence, the optimal scale of monitoring is 8 for this indicator and this freshwater ecosystem, so as to capture all the variations significantly influencing the general tendency of the signal while avoiding insignificant fluctuations.

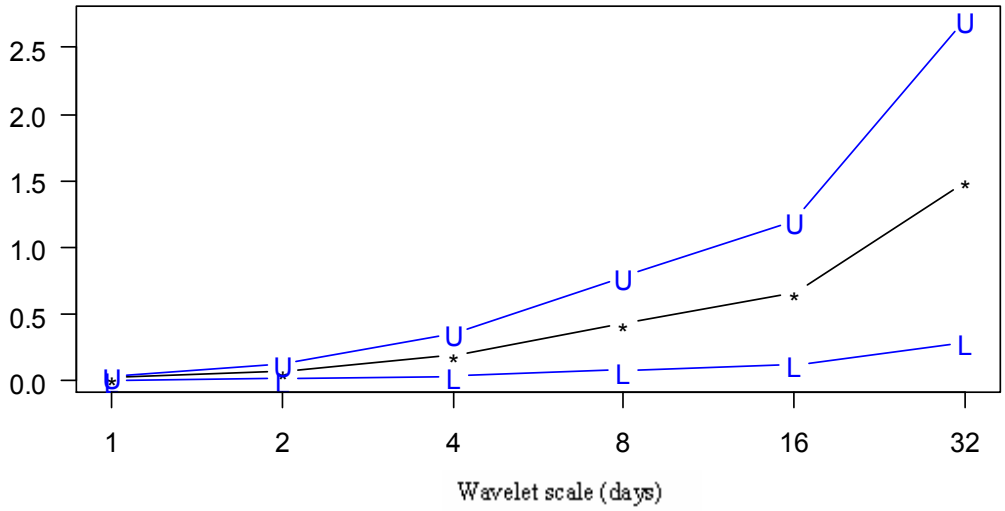

Figure 4.17: Wavelet variance of water temperature with db4.

The wavelet variance of figure 4.17 gives a clearer insight into the intensity of the variations occurring at different time scale in the water temperature signal. It reveals that there is a progressive increase in the intensity of the changes as

one move from one scale to the next, though the intensity of the variations is relatively low. From scale 16 to 32, there is a sharp increase in the intensity, implying that low frequency changes or changes occurring from scale 32, are those that are able to significantly influence the long term behavior of the water temperature time series. This is very important for aquatic organisms and the stability of chemical processes in the water body, with the variance structure enabling the system not to suffer from rapid changes in temperature as a result of short term fluctuations in weather conditions. If such were the case, it will result in a chaotic disruption of the freshwater ecosystem given that many organisms depend on the regularity of the changes in water temperature to perform certain activities at certain periods in their life cycle. Monitoring water temperature of this ecosystem at a monthly interval will be effective in capturing the variation that influences the long term dynamics of this signal. Much information is not gained by significantly increasing the sampling frequency.

Wavelet covariance

The wavelet covariance between dissolved oxygen and water temperature in figure 4.18 unravels the co-variation or the interrelationship at different time scale between these two indicators. This is quite interesting because it indicates a covariance of close to zero at scale 1, representing the daily variations present in the signal as a result of the absence of any significant change present at this time scale. There is a linear increase in the positive co-variation or relationship up till scale 8 after which the co-variation starts dropping and increases negatively till it reaches -0.4 at scale 32. From literature and other analysis, the correlation between dissolved oxygen and water temperature should portray a negative correlation. But the wavelet decomposition of the covariance at different time scales shows that the covariance from scale 1 to scale 16 is positive, and the value are different from one scale to the other due to the difference in the intensity of co-variation of the different processes operating at these different time scale. Scale 32 shows a negative covariance which is what could have been expected. This observed behavior in the lower scales is as a result of the high production of oxygen by

the photosynthetic activity of the high phytoplankton biomass present in this freshwater body that turns to overshadow the effect of water temperature at lower scales.

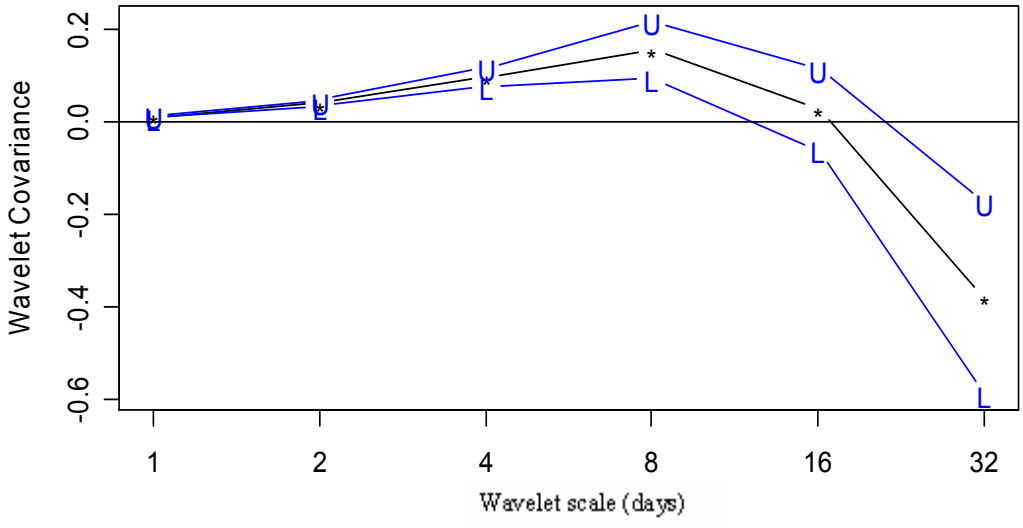

Figure 4.18: Wavelet covariance of dissolved oxygen with water temperature using db4.

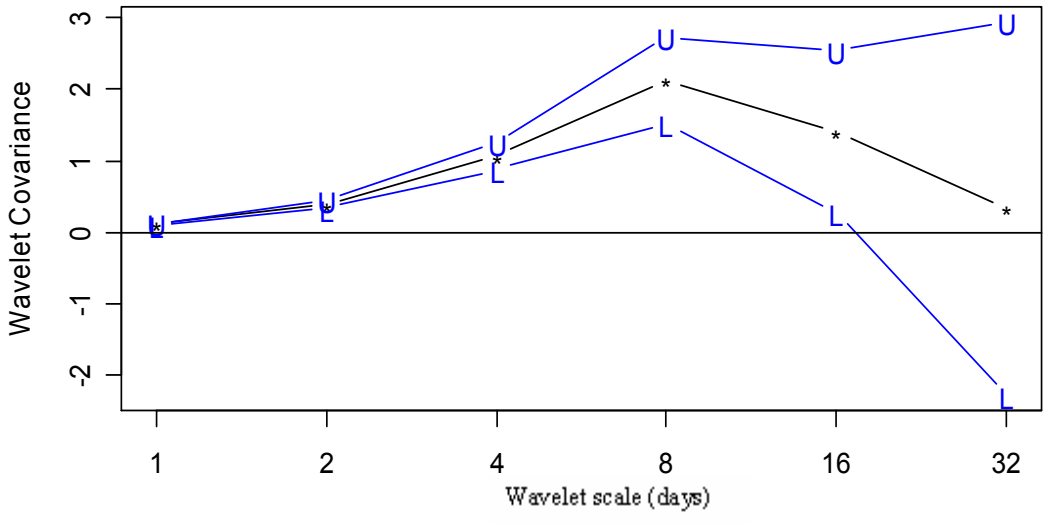

Figure 4.19: Wavelet covariance of chlorophyll-a with water temperature using db4.

The wavelet covariance between chlorophyll-a with water temperature in figure 4.19 shows the decomposition of the co-variation or the interrelationship at different time scales between these two indicators. At scale 1 which reflects the daily variations, there exists little co-variation between the two signals. The there is a linear increase in the positive co-variation or relationship up till scale 8 and the co-variation starts dropping till scale 32, but does not become negative. The signals have the highest covariance at scale 8, which is the scale at which the variations significantly influencing the shorter cycles in the chlorophyll-a time series occur. This implies, the variations of the two dominant groups that compose the biomass strongly vary in the same direction as variations in water temperature compared to the variation of other component operating at different time scales. The variations at scale 32 representing the fluctuations that have the most influence on the seasonal cycle are not do not strongly change in the same direction as the temperature of water changes. Once again, the scale of 8 seem to be a determining scale of activities in this fresh water body making it quite necessary to incorporate this information into planning, monitoring, modeling and management of the freshwater body.

Wavelet cross-correlation

Figure 4.20 presents the lead-lag relationship between dissolved oxygen and water temperature. This behavior by means of wavelet cross-correlation has been decomposed into different time scales. Hence, a cross correlation at different time scale is revealed. The highest cross-correlation seems to occur at a scale of 8 or level 4 having a value of -0.5 with a time lag of 8 days. It is clearly observed that the cross- correlation differ from one scale to the next with a net increase in its value from one scale to the next with the lower scales having the lowest cross-correlation despite increasing the time lag.

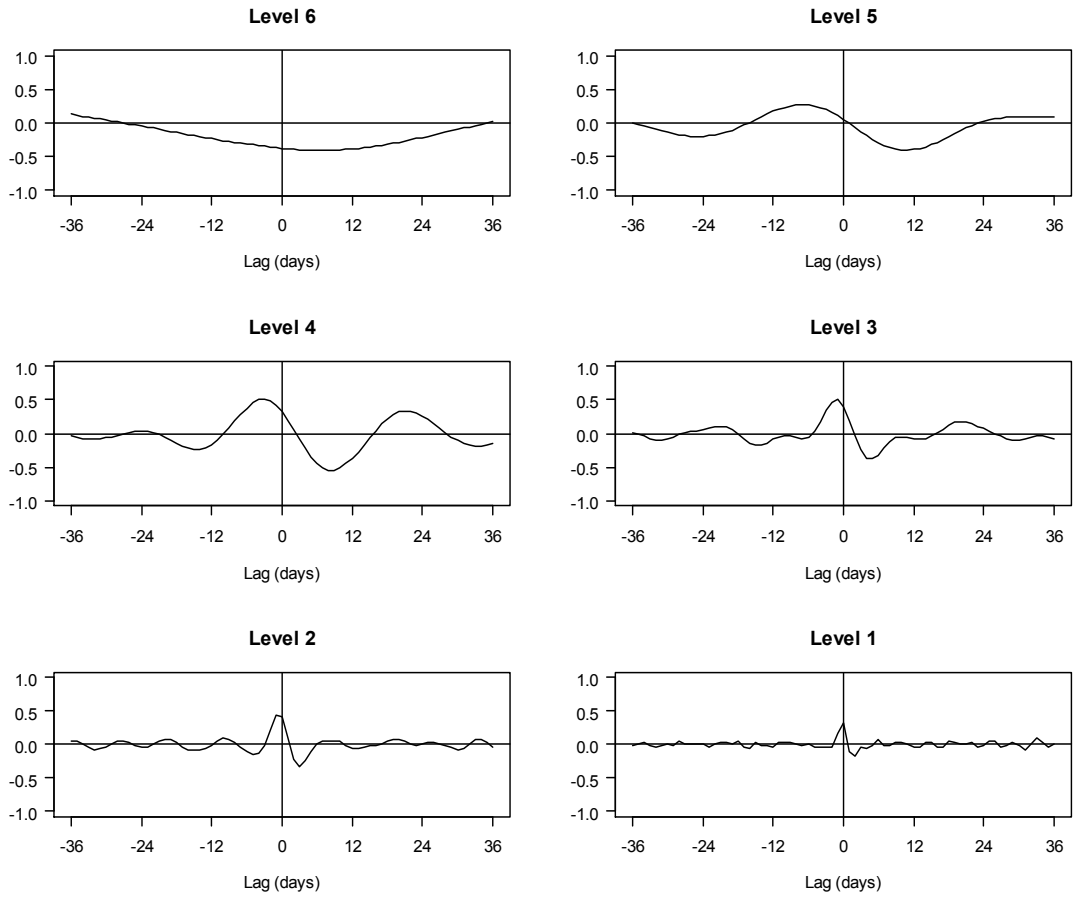

Figure 4.20: Cross-correlation between dissolved oxygen with water temperature using db4.

Given that the dominant scale of variability of this indicator is scale 8 and it also has its greatest cross-correlation at this scale, it will be very reasonable to use the results from the wavelet cross-correlation at scale 8 for modeling, planning and management that takes into consideration the effects of water temperature on dissolved oxygen.

The time scale decomposition of the lead-lag relationship of chlorophyll-a and water temperature is presented in figure 4.21. The highest cross-correlation occurs at level 4 or scale 8 with a correlation of -0.5 at a time lag of 10 days. As is clearly portrayed by the time scale decomposition of the cross-correlations, there is a net difference in the behavior of the cross-correlations from one scale to the other.

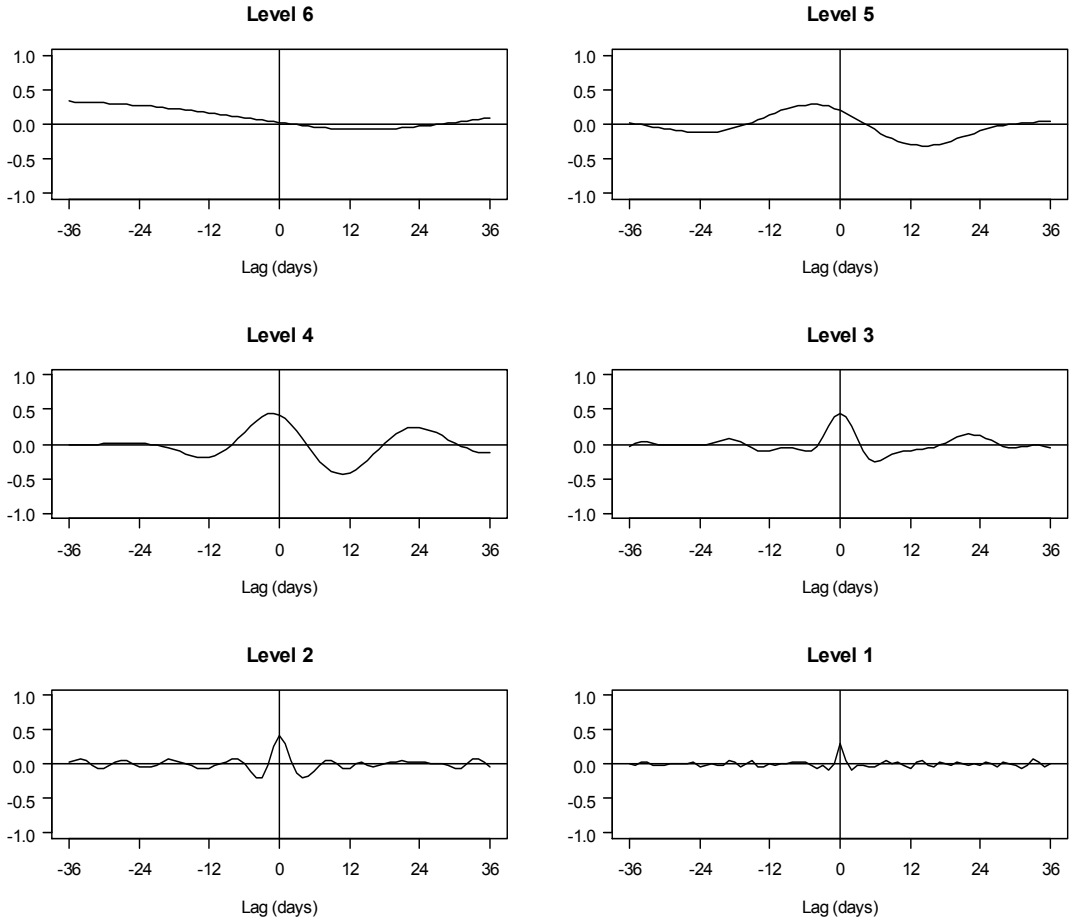

Figure 4.21: Wavelet cross-correlation between chlorophyll-a with water temperature using db4.

Since it is observed that scale 8 seems to be the time scale which captures the most important variability influencing the long term dynamics of the time series, the cross-correlation value at this scale is the highest implying the scale is the determining one. Any planning, modeling or management strategies have to take into consideration variations occurring at weekly intervals.

4.3 Comparing dissolved oxygen signal from Rivers Elbe, Havel and Oder

In the preceding part, a physical, chemical and biological ecological indicator from a River, the Havel was investigated. In this section, an ecological indicator, namely dissolved oxygen from three Rivers of a watershed will be investigated and compared. The data is from the Rivers Elbe, Havel and Oder, sampled at 10 minutes interval and latter aggregated to a daily interval was used as portrayed in figure 4.22. The main idea is to investigate whether the same indicator from different water bodies of a given watershed have the same time varying structural properties. Hence, the following univariate signal analysis methods will be used; the autocorrelation function, Box-Jenkins approximation, periodogram, Fourier polynomial, wavelet multiresolution analysis, and the wavelet variance.

The dissolved oxygen signal from the river Elbe exhibits a clear strong regular seasonal cycle from one year to the other with little high frequency fluctuations within the yearly cycles. The signal from the river Havel also contains a seasonal cycle that is punctuated with strong irregular high frequency fluctuations. The signal from the river Oder equally demonstrates a yearly seasonal cycle with the highest high frequency fluctuations within the yearly cycle. From this preliminary examination, it is obvious that this same indicator, recorded from different freshwater bodies of the same watershed display significantly different time varying behaviour which should be confirmed by further analysis.

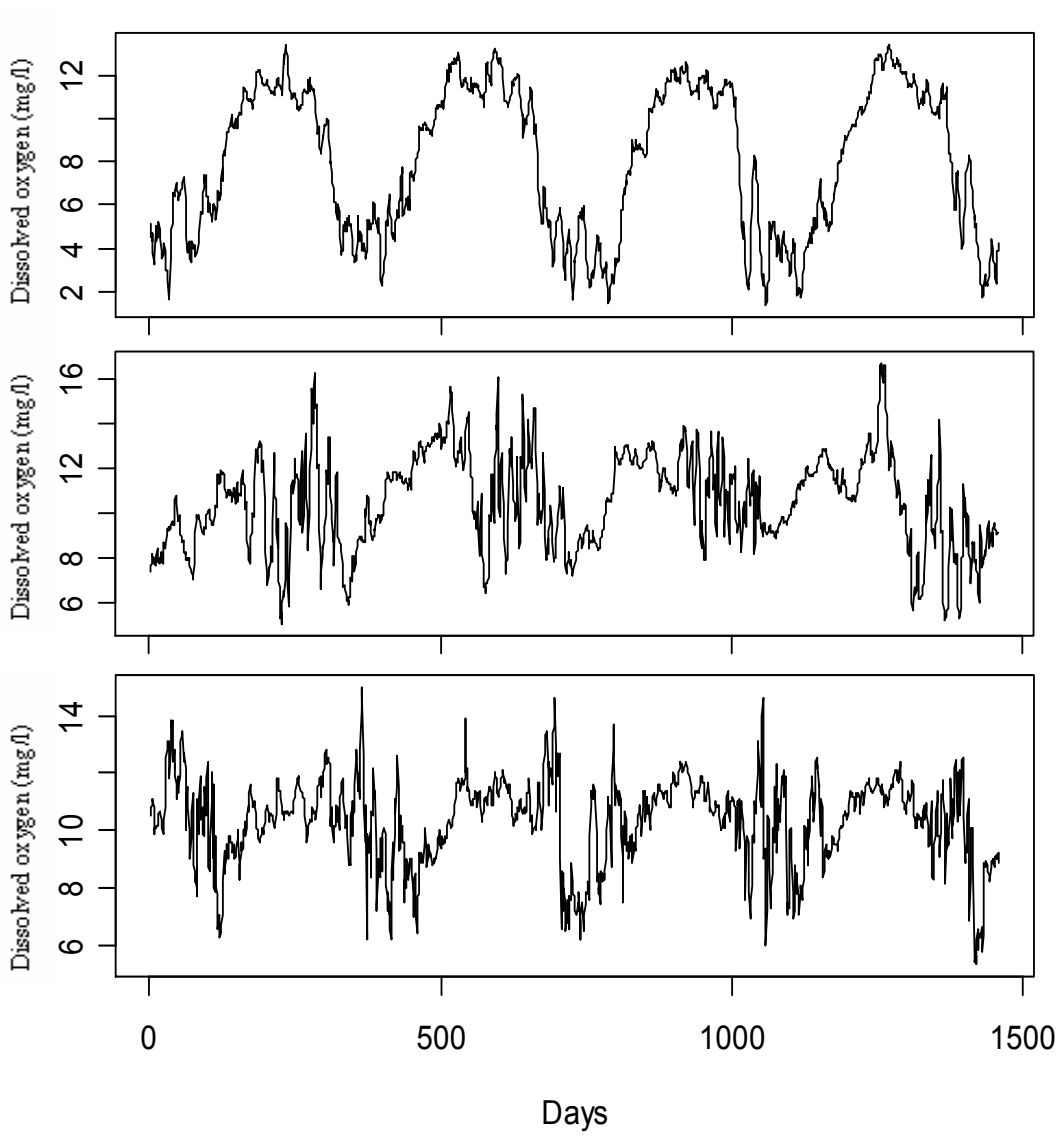

Figure 4.22: Run sequence plots of dissolved oxygen of River Elbe (top), River Havel (middle) and River Oder (bottom).

Autocorrelation

An examination of the autocorrelation structure of the dissolved oxygen signal (figure 4.23) confirms the net structural difference. The correlogram of the dissolved oxygen signal from the Elbe River reveals a very strong autocorrelation between prior and present values. This simply means that the present values are very strongly influenced by the prior values indicating the

absence of independence or the existence of long memory between the time series values. Such a strong autocorrelation may indicate that the sampling frequency for such a signal should not be high.

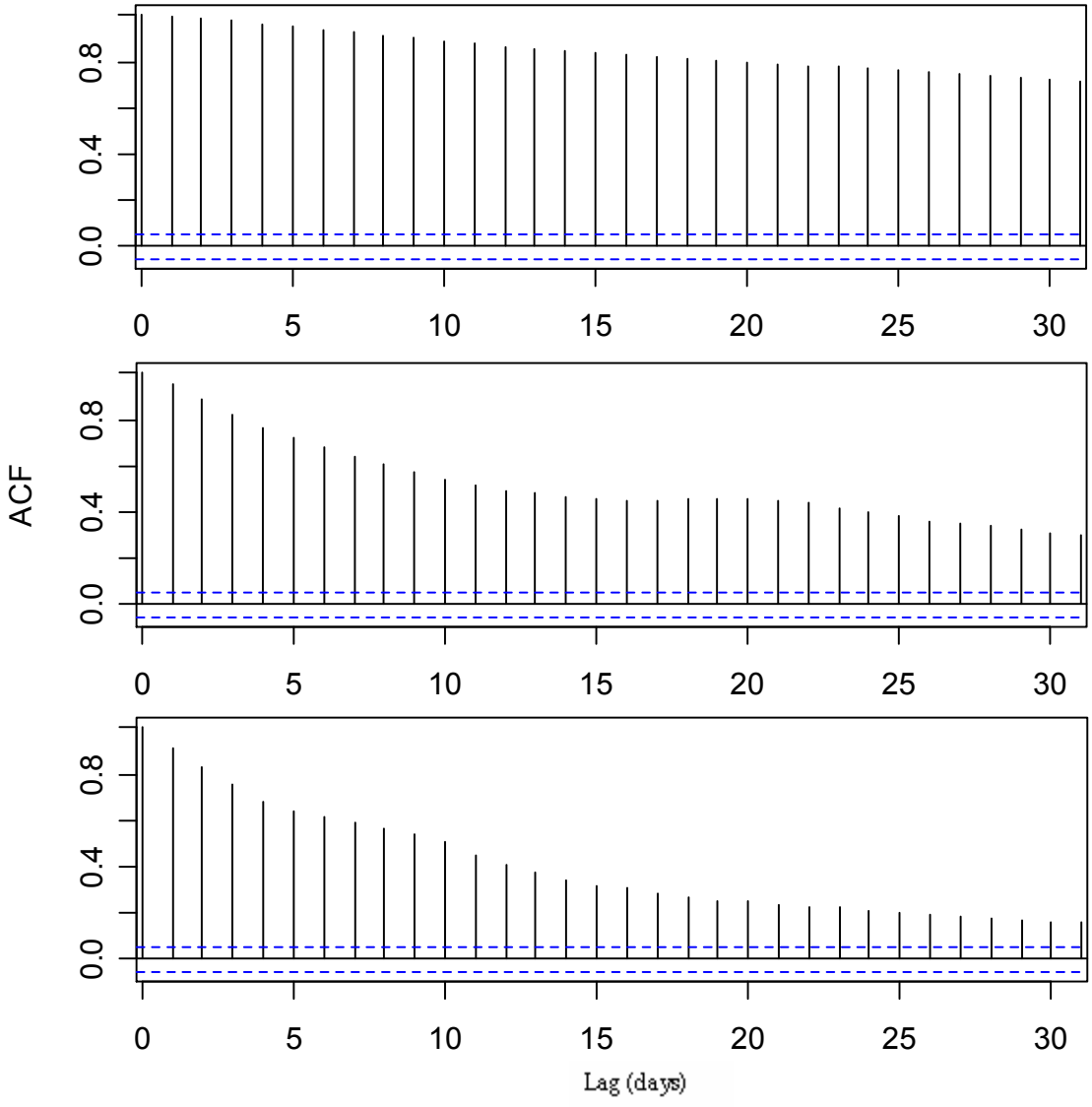

Figure 4.23: Correlogram of dissolved oxygen of Rivers Elbe (top), Havel (middle) and the Oder (bottom).

The correlogram of the signal from the Havel River portrays a different structure than that of the Elbe. It exhibits a less strong autocorrelation structure with the

presence of a slow decay of the autocorrelation as the time lag increases. Hence, the present values of the signal are not very strongly influenced by the prior values. This may indicate that such a signal requires a higher sampling frequency compared to that from the Elbe.

The correlogram of the dissolved oxygen signal from the Oder River further confirms the structural differences by portraying a fairly weak autocorrelation structure. Here, one can observe a more rapid decay of the autocorrelation as compared to that of the other signals. Hence, the influence of prior values on the present values of the signal is fairly weak and may be an indication of the need of a higher sampling frequency.

Box-Jenkins approximation

The difference in autocorrelation structure also implicates that different time series models are required to capture the long term dynamics of the signals. An investigation of the most appropriate Box-Jenkins approximation approach (ARMA/ARIMA) was done. The most appropriate approximation is selected by examining the standardized residuals, the autocorrelation function of the residuals, the p values of the Ljung-Box statistics and Akaike information criterion (the approximation with the smallest value is usually the most appropriate). A forecast of a period of 50 was done.

The following results were obtained for the approximation from the Elbe River. An ARIMA (2,1,0) model was found to be the most appropriate with the smallest AIC being 984.5 with the coefficients for the model parameters shown in table 4.1.

Table 4.1: Coefficients of ARIMA (2, 1, 0)

Parameter	Coefficient	Standard error
ar1	0.1877	0.0259
ar2	0.1559	0.0259

The estimated sigma square is 0.1145, the log likelihood is -489.25 and the AIC is 984.5 with figure 4.24 showing the signal, the approximation and the

forecast. The red lines indicate the 95 % confidence interval with the blue line in-between representing the forecast for 50 periods.

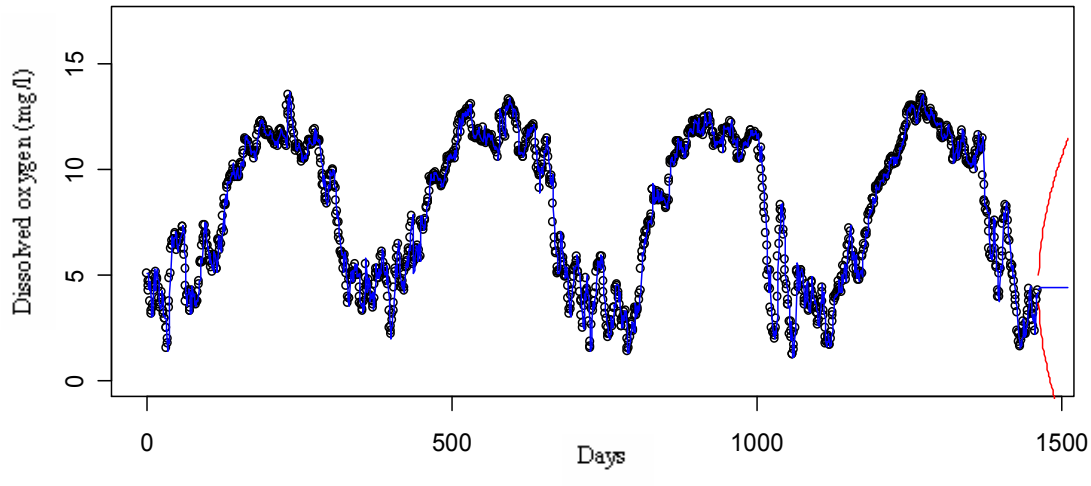

Figure 4.24: ARIMA (2, 1, 0) approximation of dissolved oxygen signal from River Elbe with forecast of 50 periods.

Using the Box-Jenkins method for the Havel River yielded the following. An ARIMA (4, 1, 3) model was found to be the most appropriate with the smallest AIC being 2582.72 with the coefficients for the model parameters shown in table 4.25.

Table 4.2: Coefficients of ARIMA (4, 1, 3)

Parameter	Coefficient	Standard error
ar1	-0.1982	0.1249
ar2	0.6532	0.1357
ar3	0.3587	0.1049
ar4	-0.1401	0.0551
ma1	0.5003	0.1215
ma2	-0.7035	0.1094
ma3	-0.6712	0.0824

The estimated sigma square is 0.3399, the log likelihood is -1283.36 and the AIC is 2582.72 with figure 4.25 showing the signal, the approximation and the forecast.

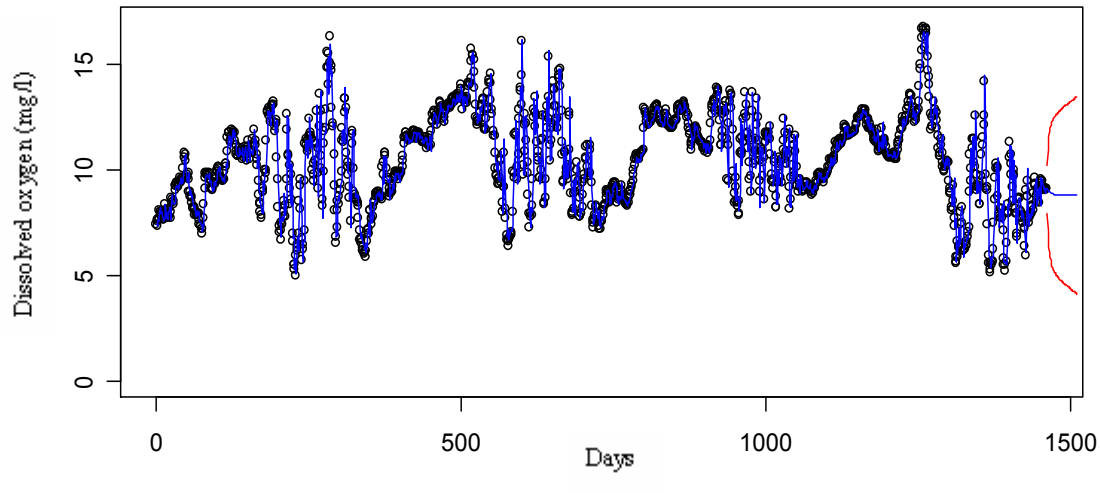

Figure 4.25: ARIMA (4, 1, 3) approximation of dissolved oxygen signal from River Havel with forecast of 50 periods.

Table 4.3: Coefficients of ARIMA (4, 1, 3)

Parameter	Coefficient	Standard error
ar1	0.3801	0.0370
ar2	-0.4857	0.0339
ar3	0.8638	0.0325
ar4	-0.0610	0.0299
ma1	-0.4535	0.0268
ma2	0.4307	0.0309
ma3	-0.9163	0.0371

Using the Box-Jenkins method for the Oder River yielded the following. An ARIMA (4, 1, 3) model was found to be the most appropriate with the smallest AIC being 2735.16 with the coefficients for the model parameters shown in table 4.3. The estimated sigma square is 0.3773, the log likelihood is -1359.58 and the AIC is 2735.16 with figure 4.26 showing the signal, the approximation and the forecast.

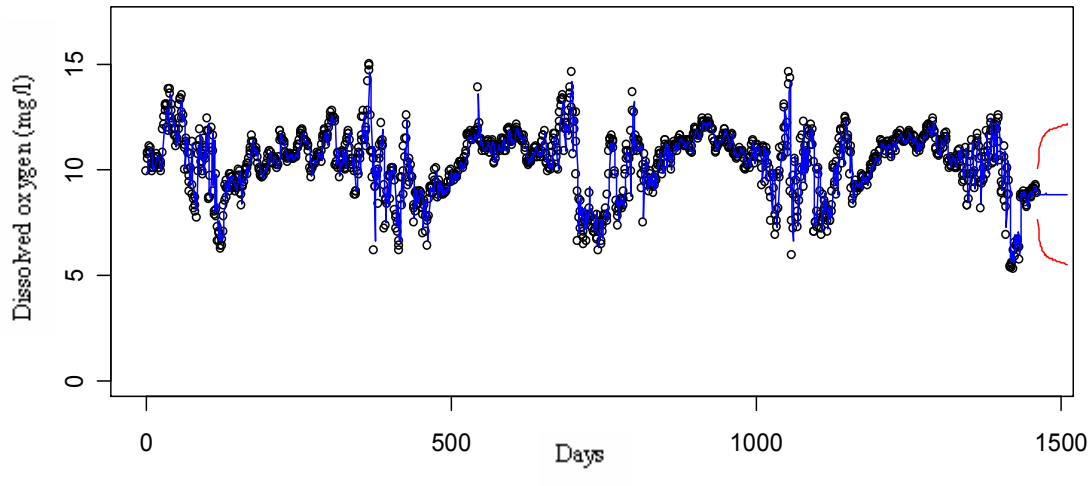

Figure 4.26: ARIMA (4, 1, 3) approximation of dissolved oxygen signal from River Oder with forecast of 50 periods.

Periodogram

Further investigating the structures of the signals using periodogram analysis reveals the following results. The periodogram of the signal from the Elbe River (figure 4.27 (top)) portrays a clear distinct peak with the strongest intensity. Negligible high fluctuations are present indicating that the strong seasonal cycle is exclusively influenced by a strong low frequency component. The signal from the Havel River (figure 4.27 (middle)) reveals a strong low frequency component represented by a peak of lower intensity than that from the Elbe, but contains some random fluctuations. The signal from the Oder River (figure 4.27 (bottom)) exhibits a low frequency peak with the lowest intensity of the three signal and a significant number of random fluctuations. Hence, the periodogram analysis enables the analyst to estimate the difference in intensity of the yearly cycle that exists between the signals.

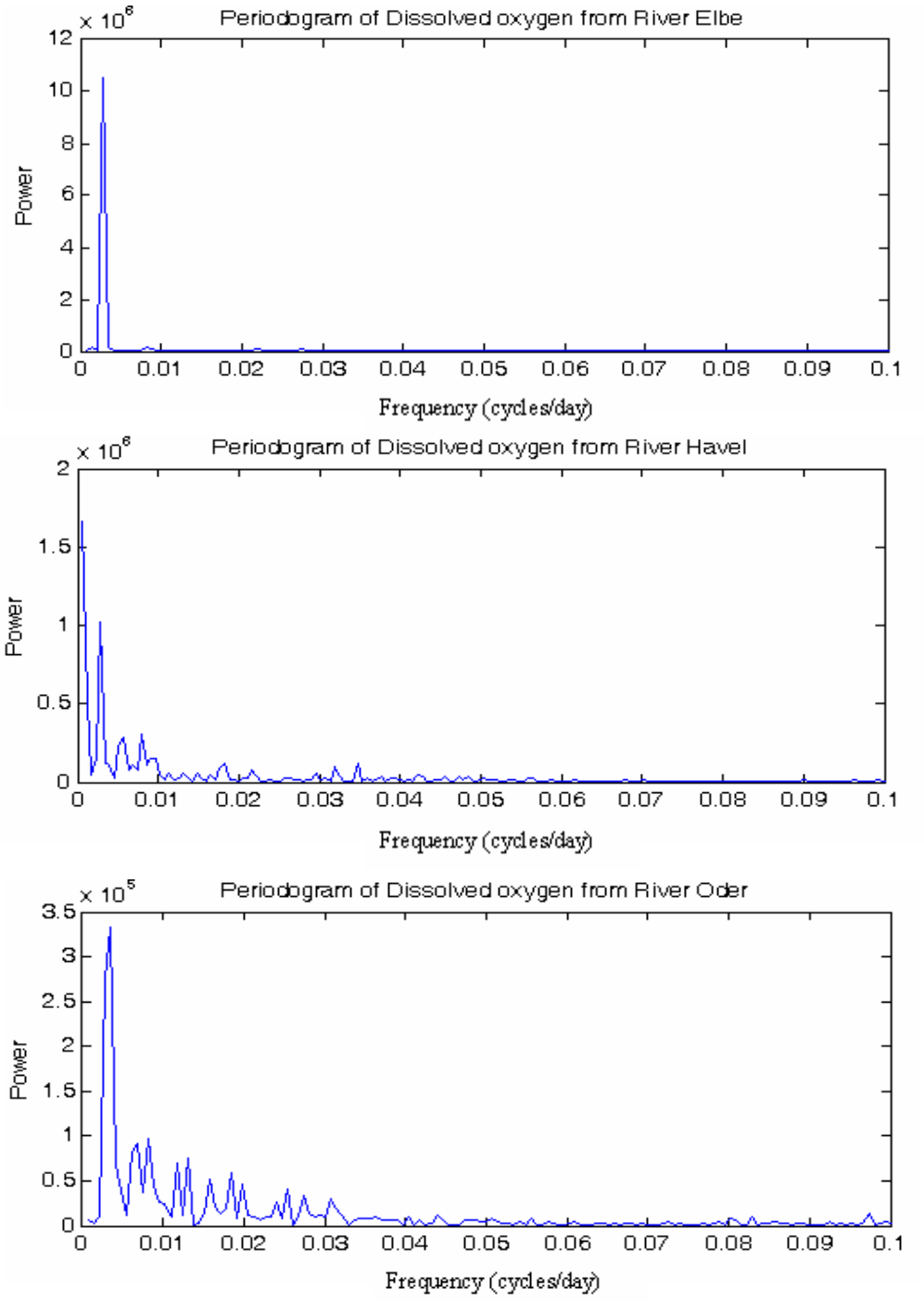

Figure 4.27: Periodogram of dissolved oxygen of the river Elbe (top), Havel (middle) and the Oder (bottom).

Fourier polynomial

In an attempt to approximate the signal using a combination of sine and cosine sinusoidal functions provided by the Fourier polynomial technique, the following results were obtained.

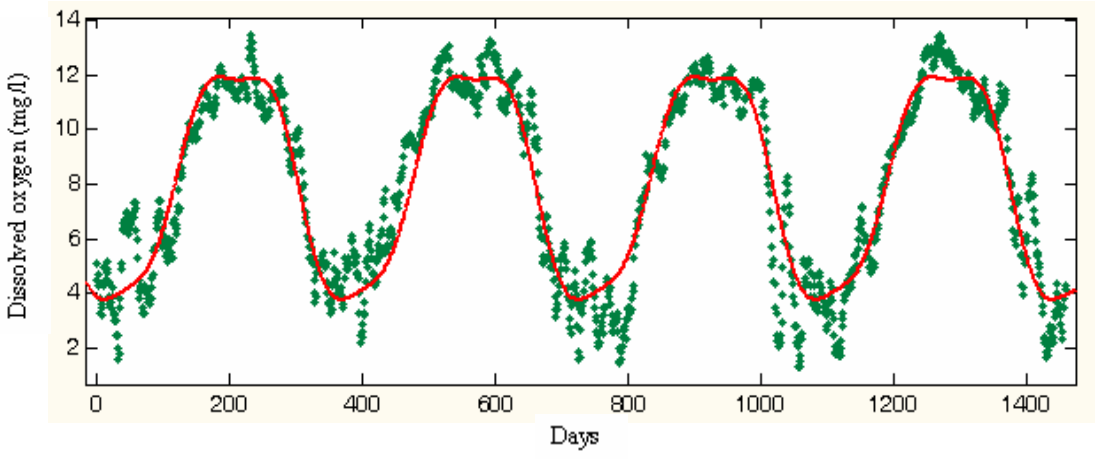

Figure 4.28: Third order Fourier polynomial of dissolved oxygen from the River Elbe.

As a result of the distinct cyclic behavior exhibited by the signal from the Elbe River, the Fourier polynomial of third order as shown in figure 4.28 gives a fairly decent approximation of the signal. The approximation provides a coefficient of determination r^2 = 0.88 and adjusted r^2 = 0.88 implying approximately 80 % of the variations are captured by the polynomial and the equation it is given by:

$$f(t) = 8.12 - 3.7\cos(0.018t) - 2.4\sin(0.018t) - 0.27\cos(0.036t) - 0.18\sin(0.036t) - 0.24\cos(0.054t) + 0.5\sin(0.054t).$$

As a result of several fluctuations present in the signal from the Havel River, the result of the Fourier polynomial is not able to give a decent approximation as represented in figure 4.29. The best approximation is obtained with a polynomial of third order with further increase in the number of polynomial terms yielding no better results. The coefficient of determination r^2 = 0.30 and

adjusted r^2 = 0.30 implying approximately 30% of the variations are captured by the polynomial and the equation it is given by:

$$f(t) = 10.65 - 0.16\cos(0.01t) - 0.38\sin(0.01t) - 043\cos(0.02t) - 0.23\sin(0.02t) - 1.58\cos(0.03t) + 0.12\sin(0.03t).$$

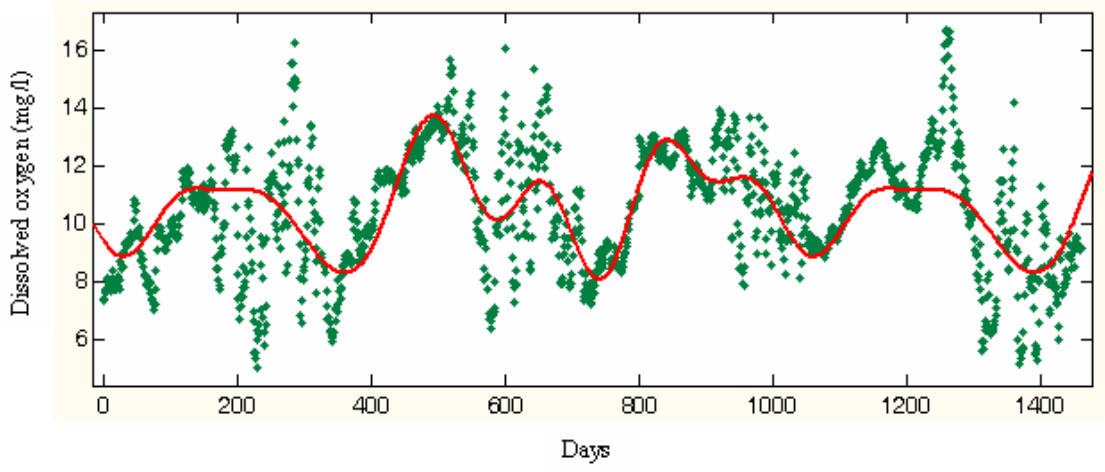

Figure 4.29: Third order Fourier polynomial of dissolved oxygen from the River Havel.

Similar to the signal from the Havel River, the signal from the River Oder is plagued with several irregular high frequency fluctuations that significantly decrease the ability of the Fourier polynomial to approximate the signal as shown in figure 4.30. As a result, a polynomial of second order gave the best approximation with any further increase in the number of terms yielding no better result. The coefficient of determination r^2 = 0.30 and adjusted r^2 = 0.29 implying approximately 30% of the variations are captured by the polynomial and the equation it is given by:

$$f(t) = 10.39 - 0.06\cos(0.01t) - 0.03\sin(0.01t) - 0.95\cos(0.02t) - 0.67\sin(0.02t).$$

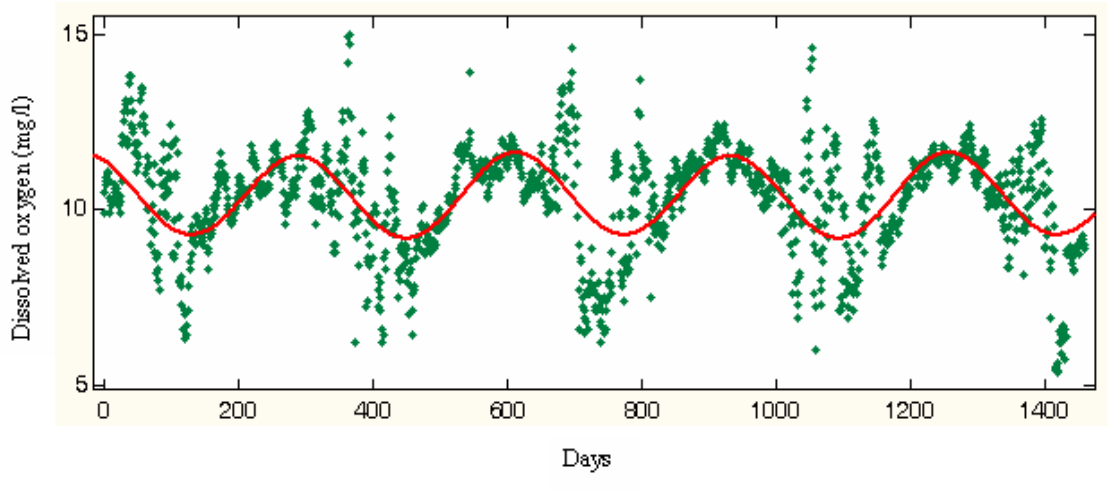

Figure 4.30: Second order Fourier polynomial of dissolved oxygen from the River Oder.

Wavelet multiresolution analysis

The approximations provided in figures 4.31, 4.32 and 4.33 give a clearer and clearer version of the signal giving the clearest version of the signal at level 6. The approximation of the signal from River Elbe clearly unveils a clear cycling behaviour with a distinct sinusoidal signal void of any high frequency changes that blur the underlying long term signal is revealed at level 6. The behaviours of the other signals from River Havel and Oder exhibit cycles with much high frequency fluctuations. The details on the same figures show the changes present in the signals at different time scales. The intensity of the daily changes of the signal from River Elbe are of fairly low intensity and do not significantly change from one season to the next. The intensity of the fluctuations present in the signals from the Rivers Havel and Oder are clearly distinct from one season to the next with the intensity of the signal equally increasing as the time scale increases. The strongest intensity of the fluctuations in all the is observed in the higher time scales.

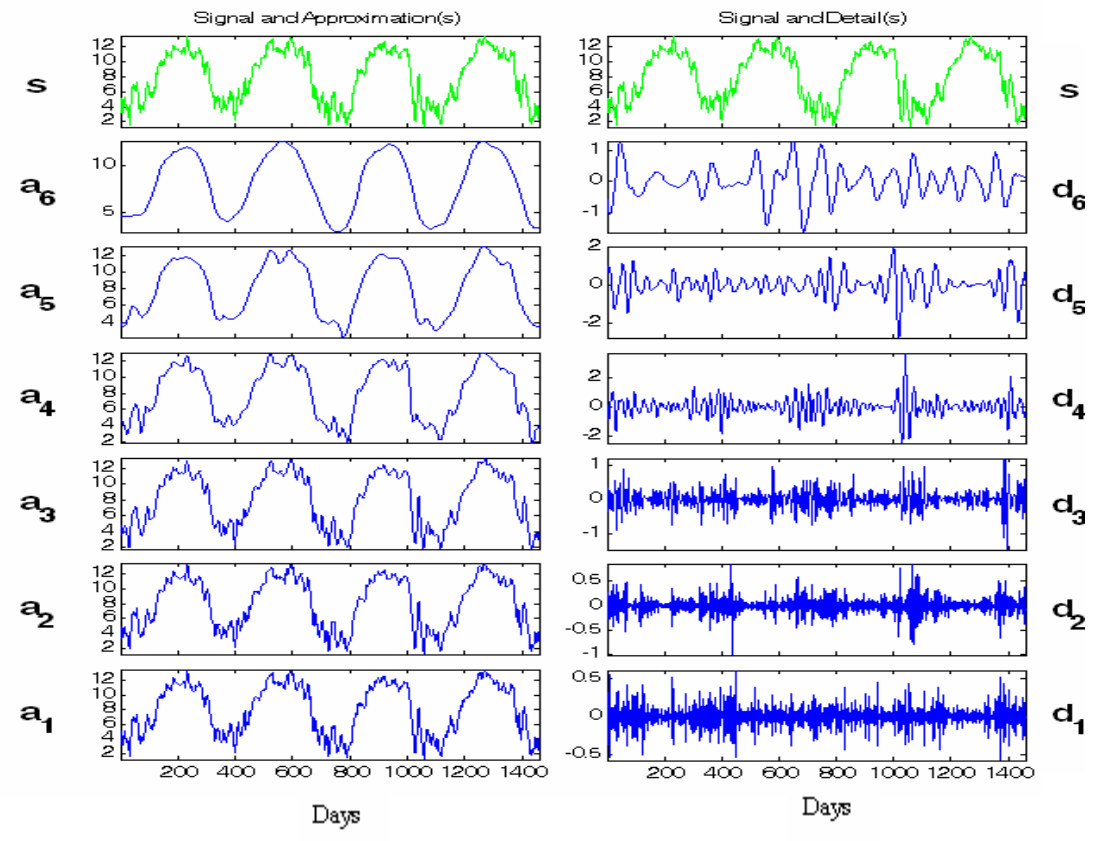

Figure 4.31: Multiresolution analysis of dissolved oxygen of the River Elbe with db4.

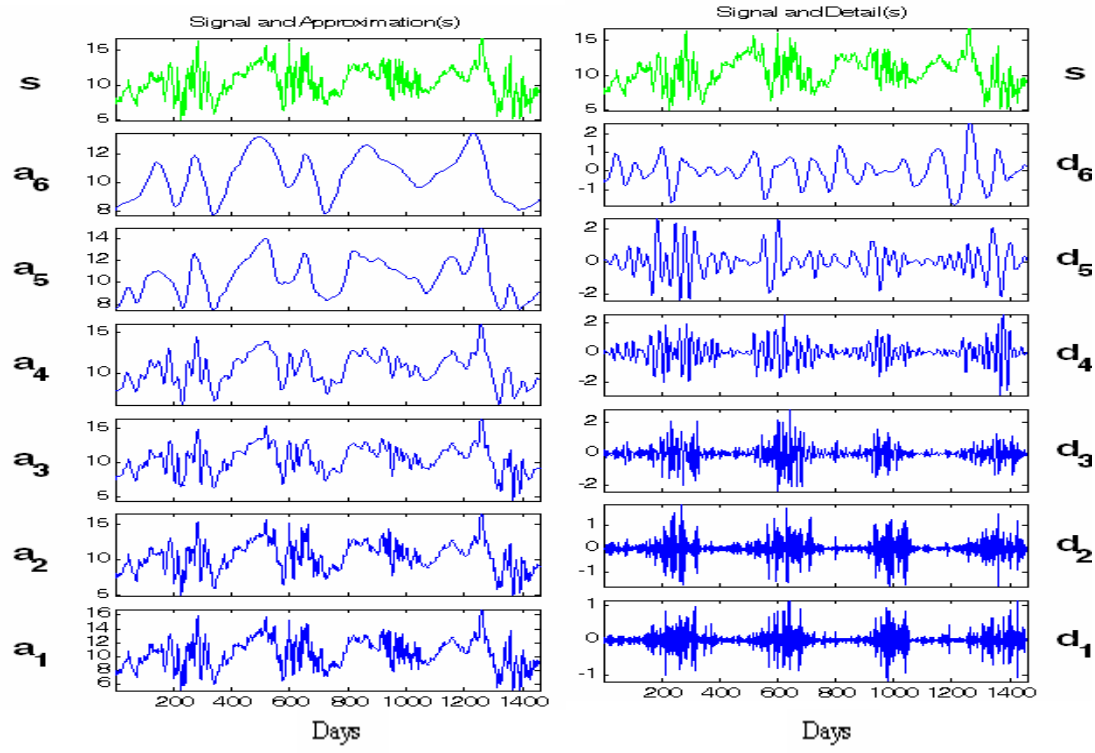

Figure 4.32: Multiresolution analysis of dissolved oxygen of the River Havel with db4.

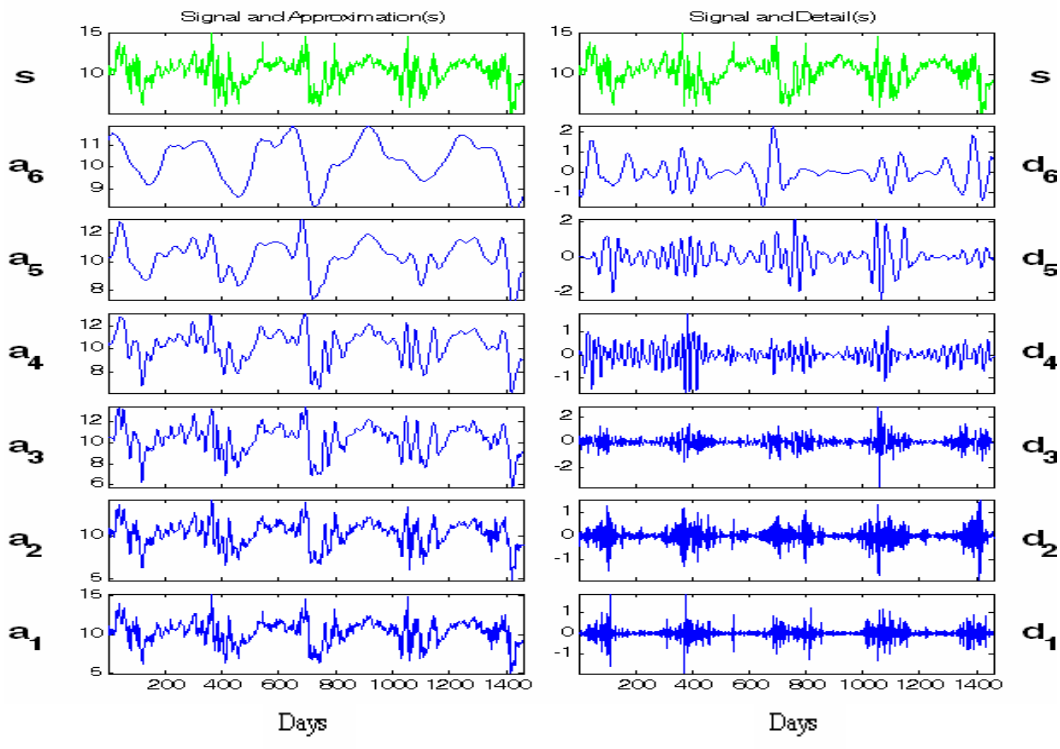

Figure 4.33: Multiresolution analysis of dissolved oxygen of the River Oder with db4.

Wavelet variance

The wavelet variance of the signal from the Elbe River reveals that the changes with the highest intensity occur at scales 8, 16 and 32. It is evident from figure 4.34 that there is no significant change from scale 1, 2 and 4.

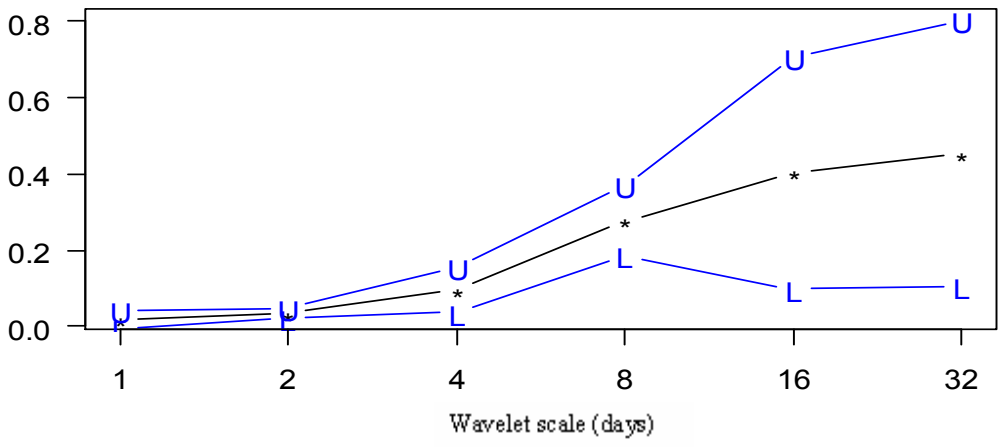

Figure 4.34: Wavelet variance of dissolved oxygen from the River Elbe using db4.

This has some implications on choosing an optimal sampling strategy for the signal. When the sampling frequency is increased from scale 8 to scale 4 or 2 or 1, little or no new significant information is obtained. This will not be cost effective and may result in recording redundant high frequency information which turns to blur the underlying long term dynamics of the signal.

The wavelet variance of the dissolved oxygen signal from the Havel River shown in figure 4.15 exhibits a linear increase in the intensity of the changes from scale 1 up to scale 8, where a local maximum is reached. Increasing the time scale doest not significantly alter the intensity of changes occurring in the signal. A close examination of the magnitude of the changes occurring are scales lower than 8 reveals changes of low intensity that do not significantly affect the general tendency of the signal as confirmed by the multiresolution analysis. Hence, sampling the signal at a weekly interval will be cost effective.

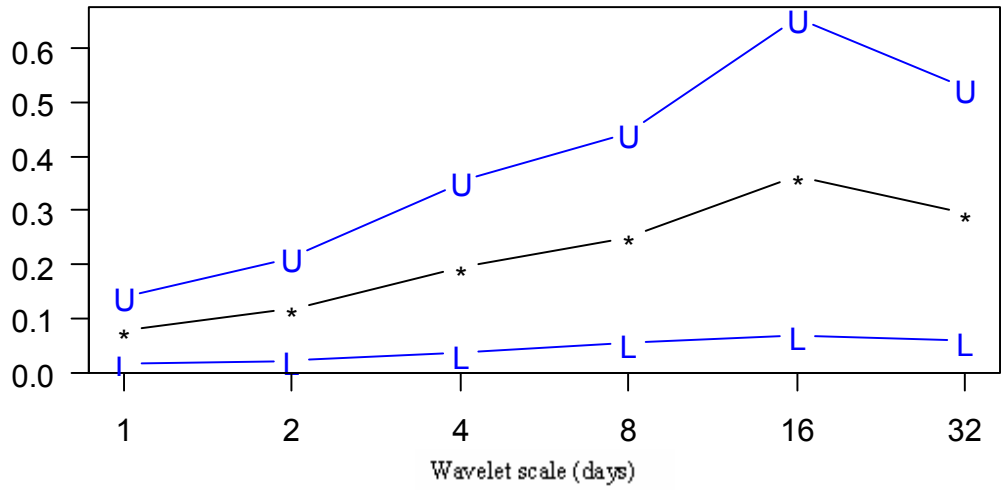

Figure 4.35: Wavelet variance of dissolved oxygen from the River Oder using db4.

The wavelet variance of the dissolved oxygen signal from the River Oder unveiled in figure 4.35 equally reveals a linear increase in the magnitude of the changes occurring from scale 1 till scale 16, where a maximum is reached. Increasing the time scale gives no further increase in the intensity of the changes occurring in the signal. Similar to the signal from the Havel River, the

lower scales or high frequency changes are of relatively low intensity, implying that the information gained from sampling at a fairly high frequency is negligible. The high frequency information generally contains information that does not influence the general tendency of the signal, but rather blur it and negatively affect modeling results.

The analysis reveals that the signals of dissolved oxygen, from different freshwater bodies in the same watershed exhibits different time varying structure. The macro and the micro structure of the signals show clear differences, thou they have some similarities. Their autocorrelation structures are significantly different with their long term behavior being governed by low frequency components as revealed by the periodogram and wavelet methods. Due to the different microstructures, the signals require different time series models to successfully approximate them.

5 Summary

Freshwater ecosystems are deteriorating as a result of anthropogenic activities and needs to be effectively managed. The major pollutants of freshwater ecosystems are presented in the second chapter of this thesis. It was found that the main pollutants are oxygen demanding wastes, nutrients like nitrogen and phosphorous, suspended solids, toxic organic and inorganic pollutants like pesticides and herbicides, microbial contaminants and thermal pollution. The sources of these pollutants are point and non-point domestic, municipal, industrial and agricultural wastes. Some these wastes substances not only negatively affect the biota of the freshwater body, but contain carcinogenic substances and infectious agents that endanger human health.

Freshwater indicators are used in order to assess the ecological state, provide an early warning signal of changes in the water body, or to diagnose the cause of a problem (type of pollutant) in the freshwater ecosystem. For an indicator to be chosen and used, it has to conceptually relevant to the issue being investigated, easy to measure and implement using the available resources. In addition, it should be easy to understand and apply, sensitive to stresses on the system, easily interpretable and provide useful and timely information for decision making. Not only do they help in describing the present ecological conditions, but also facilitate the prediction of the future ecological state by the help of appropriate models. The models also enable the investigation of the behaviour of the system under different management scenarios.

In order to evaluate the ecological state of a water body, water quality standards which provide the concentration of substances above which the water resource is unsuitable for a particular use is compared to monitored data. This is the main administrative tool for evaluating the suitability of a water body for different uses have to be established taking into consideration the climatic,

geologic, land use characteristics of the watershed as well as the time varying behaviour of the indicator. The investigation of cycling water quality indicators revealed that a standard is not appropriate as a result of alternating periods of compliance and non compliance within a very short interval of time as observed in the State of Brandenburg, Germany. A more realistic upper and lower limit that takes into consideration the microstructure of the indicator is proposed.

For the water quality signals to be analysed so as to extract useful data from them, they need to be pre-treated so as to improve the data quality which is usually messy in structure. This is usually as a result of missing data, inconsistent data, impossible data and data that do not conform to the requirements of the analysis technique. Extracting information from an unreliable data series will mislead the decision makers. Techniques such as the run sequence plot, histogram and box-plot permits an initial examination of the data series. When required, the data series may be transformed by appropriate transformations, missing data can be replaced by appropriate interpolation or approximation method. The approximation methods are appropriate when the time interval over which data points are missing is fairly long.

In order to develop useful freshwater quality models, appropriate signal analysis methods are required to extract the starting information concerning the behaviour and relationship between the relevant indicators. There exist classical as well as modern bivariate and univariate signal analysis methods. The time domain methods (correlation functions and test functions) enables information to be extracted only from the time vary behaviour of the signal. The frequency domain methods (periodogram, spectral and cross-spectral analysis) enable cyclic components as well as the active frequency bands and their intensity to be uncovered. These two classical methods are not able to reveal the time at which the different frequencies are active. It was found that the water quality signals of natural processes are an amalgam of fluctuations occurring at different time-scale with different intensities. A modern method, wavelet analysis is able to unravel this changes giving deeper insight to the long and short term dynamics of water quality indicators.

Applying the univariate and bivariate classical and modern signal analysis methods to a physical, chemical and biological water quality signal (water temperature, dissolved oxygen and chlorophyll-a respectively) give the following results. The results for dissolved oxygen analysis are as follows: The signal of the chemical indicator, dissolved oxygen, exhibits long memory structure, indicating that the present values of the signal are influenced by the prior values, making the time series to be non-independent. The time series portrays two periodic or cyclic behaviors with the longer cycle being caused by seasonal variations and a shorter cycle embedded in the longer cycle. The short cycle occurs probably as a result of the life cycle of the strong phytoplankton biomass present in the water body producing oxygen during photosynthesis. This phenomenon is able to significantly influence the long term tendency or behavior of dissolved oxygen in this freshwater body. Examining the different components acting at different time scale in the dissolved oxygen signal also reveal that the intensity of the fluctuations occurring during the warmer periods is significantly stronger than those occurring during the colder periods. The strength or intensity of the changes increases as we move from lower to higher time scale. The daily changes are quite insignificant, increasing as the time scale increases till scale 8, where a first local maximum is reached. From here, the fluctuations are not significantly different as we increase the time scale. The fluctuations at the scale of 8 and above are those that significantly influence the long term tendency displayed by the dissolved oxygen signal while the variations occurring at the lower scale are not able to change the long term behavior of the signal. This implies it will be cost effective for monitoring programs to sample at a scale of 8 or at a weekly interval if the objective of the monitoring program is to investigate changes in the long term behavior of the indicator.

The bivariate signal analysis between dissolved oxygen and water temperature using the time domain, frequency domain and wavelet analysis methods reveal that there exists a weak negative correlation of -0.31 implying a decrease in the concentration of dissolved oxygen as water temperature increases. This result

is obtained when the signal, which is an amalgam of processes occurring at different time scales, is analyzed. Further investigation by decomposing the signal on a time scale basis shows that there exist weak positive correlations which increases from the lower scales till the scale of 8, then drops at scale 16 and finally becomes significantly negative at scale 32. Scale 8, which has strong fluctuations capable of influencing the long term dynamics of the indicator, shows a positive correlation contrary to what may be expected, may be due to the strong photosynthetic activity of the strong algal biomass producing much oxygen capable of overshadowing the effects of water temperature on the dissolved oxygen concentration at this scale. The situation returns to what may be expected from theory at scale 32. There is evidence of the existence of synchronized cycles between the dissolved oxygen and water temperature signal within the frequency bands from zero to 0.01 with a phase difference of -3. The cross-correlation of the non-decomposed dissolved oxygen signal portrays a strong correlation of about -0.4 at a time lag of 12 while the decomposed version-the wavelet cross-correlation- shows that the strongest most important cross-correlation occurs at scale 8 or level 4 with a correlation of -0.5 at a time lag of 8 days. The wavelet analysis helps us to understand that the weekly changes strongly influence the long term dynamics of the dissolved oxygen concentration in this water body and should be taken into consideration in modeling, planning and management of this freshwater body.

The signal of the biological indicator, chlorophyll-a exhibits long memory or persistence in its autocorrelation structure, implying the existence of dependence between present values and prior values. The most important variations in the time series are determined by the low frequency fluctuations with two important periodic event caused by seasonal change (longer period) and the life cycle of the two main constituent groups present in the phytoplankton biomass (shorter period). Examining the variations at different time scales reveal that the scales lower than 8 are not able to significantly influence the long term tendency of the signal with scale 8 being the most

important scale. This is because it is able to capture the variations that significantly influence the life cycle of the constituent species in the biomass as well as those affecting the seasonal change while leaving out insignificant variations. For monitoring to be cost effective, it should not be done at scales lower than 8 or weekly interval. To study exclusively the variations influencing the seasonal cycles while excluding the shorter cycles, scale 32 is the appropriate scale.

Bivariate signal analysis between chlorophyll-a and water temperature reveal the existence of a relatively strong correlation between the two signals with a value of 0.6 implying that there is an increase of biomass as water temperature increases and vice-versa. This result is obtained when the signal, which is amalgam of processes operating at different time scales, is analyzed. When decomposed, it is found that the correlation is quite different from one scale to the other with the highest value occurring at scale 8. The presence of synchronized cycles between the dissolved oxygen and water temperature signals within the frequency bands from zero to 0.01 is revealed. The cross-correlation of the non-decomposed signal indicates a decrease of correlation as the time lag is increase while the decomposed version indicates the highest cross-correlation value is obtained at scale 8 at a time lag of 10 with a strange value -0.5. The time scale of 8 seem to be the determining scale with regards to significant changes affecting the long term dynamics of the biomass present in this freshwater body and must be taken into consideration or incorporated in modeling, planning and management of the freshwater body. A more detailed analysis of the variation and interrelationships at scale 8 is required to give more insight to changes occurring in this freshwater body.

It is observed that the signal of the physical indicator, water temperature, has a dependent time structure with a very strong autocorrelation and the exhibition of long memory behavior. The signal contain a single periodic component of high intensity which indicates in a clear cyclic behavior as a result of seasonal changes from the warm to the colder periods of the year. This periodic behavior is influenced mainly be the low intensity variation present in the time series.

The high frequency variations have very minimal influence of the long term characteristics of the time series. An examination of the time scale variations confirm that the high frequency variations or variations occurring at the lower time scale have no significant influence on the signal's long term characteristic. The fluctuations occurring from scale 32 which represent the monthly variation and above significantly affect the long term behavior of the time series. This has a very important implication for aquatic life and processes in this freshwater body in the sense that rapid changes of climatic conditions have no significant effect on the freshwater body. Only the low frequency or monthly variations and above are able to affect the system significantly. The bivariate analysis of water temperature is provided with the other indicators.

It is observed that all the investigated signal have a strong correlation structure with water temperature exhibiting the strongest correlation. The implication for monitoring is that the signal does not need to be sampled at a high frequency. It is also observed that the changes that significantly influence the long term dynamics of the signal is caused by the low frequency components of the signal. It was found that the chemical and biological indicators can be cost effectively sampled at a weekly interval while the physical indicator could be sampled at a monthly interval for this freshwater ecosystem. It was also found that the highest cross-correlation for the signals occur at a scale of 8. A deeper insight of the microstructure and dynamics of the dissolved oxygen, chlorophyll-a and water temperature signal is obtained by combining the classical and wavelet signal analysis methods. No method can replace the other, they are simple complimentary with some giving a deeper insight of certain aspects than the others.

An investigation of the dissolved oxygen signal from different water bodies in the same watershed shows that the time varying behavior of the indicators are quite different, despite some similarities. The implication of this dissimilarity in the same watershed is that each freshwater body offers unique modeling and management challenges. The Fourier and Box-Jenkins models required to approximate the signals are different. A given management approach can not

simply be transferred to another water body even in the same watershed as a result of differences in the behavior of the processes. The signals obtained from each water body have to be analyzed so as to extract the information necessary for effectively modeling and managing the water body of interest.

6 Conclusion

With the advent of new threats such as global warming, the need for an effective conservation and management of water resources in the world is becoming more and more urgent as a result of various reasons. Rapidly growing human population with a resultant increase in resource consumption, increasing discharge of waste into water bodies are the main reasons. Water quality indicators are used in order to assess the state of freshwater bodies, detect trends and the long term dynamics of the quality of a freshwater body. In order to efficiently model water quality changes and to manage water bodies by the help of indicators so as to comply with established standards, the information contained in the signals has to be extracted by means of classical and modern signal analysis methods. By combining these methods in investigating a physical, a biological and a chemical water quality indicator new insight into the structural and behavioural dynamics are obtained.

The classical signal analysis methods consisting of the time and frequency domain methods enable an analyst to extract important characteristics of water quality signals which serve as starting information for developing water quality models. The time domain methods only give information on the time vary structure of the signal as a function of time, but no information on the frequency behaviour of the signal at that particular time. The information obtained is very useful, for example the autocorrelation and partial autocorrelation are important for identifying the order of ARIMA models. The frequency dependency of water quality signals needs to be extracted by means of appropriate frequency domain methods. These methods however, exclusively reveal only the frequency content of a signal, but no information on the time at which the different frequency bands present are active in the signal.

Modern methods such as wavelets are able not only to reveal the frequency content of the signal, but also the time at which the changes are occurring. Actually, wavelet methods enable the water quality analyst to extract the changes taking place in the signal at different time scale, given that the behaviour of water quality signals is an amalgam of changes occurring at different time scale. This is quite important in determining the most cost effective sampling frequency for the indicator as well as they detect the scale at which the changes that influence the long term dynamics of the signal occur. The information on the relationship between signals at different time scales can also be extracted by means of wavelet bivariate analysis methods like the cross-correlation. The main implication of using the wavelet methods is that it enables the water quality analyst and modeller to extract from a signal only the important changes that significantly affect the long term dynamics of the indicator and leave out the high frequency changes which turn to blur the underlying important information and negatively affect the results of models. In addition, an appropriate sampling frequency can be determined for the signal. Finally, the bivariate methods like the wavelet cross-correlation enables the analyst to introduce the most effective time delay when using dynamic water quality models so as to improve the quality of the results.

The same indicators from different freshwater bodies pose unique modelling challenges as a result of significant differences in the processes that generated the signals. As a result, each water body requires effective signal analysis of the water quality indicators sampled in order to model and make informed management decision. Though the modern methods offer a lot of advantages, they however are not able to replace the classical methods. The two techniques should be used in order to gain all the necessary insight and information for effectively managing freshwater bodies. Given the challenges facing the world with respect to water quality management, decision making cannot be restricted to one discipline or field of studies. Modern methods developed in other fields have to be adapted and used in investigating

freshwater quality signals so as to enhance the modelling and management of freshwater ecosystems.

The methods however are limited only to indicators that can recorded at an acceptable sampling frequency as opposed the used of indicator species in identifying the ecological status and changes in a freshwater body. These indicator species reflect the real status of the freshwater body, but are quite unable to give information on sporadic changes of the concentration of certain pollutants. A combination of parameter based indicators and indicators species is of great important for sound management of freshwater ecosystems.

7 References

Akaike, H. (1969). Fitting Autoregressive Models for Prediction. Ann. Inst. Stat. Math., 21, 243-247.

Akaike, H. (1974). A New Look at Statistical Model Identification. IEEE trans: Automat. Contribution., AC-19, 716-723.

Alegue, J. D. and Gnauck, A. (2006a). Time Series Modeling of Water Quality Indicators. In: I. Troch, and F. Breitenecker (eds.): Proc. MATHMOD, ARGESIM-Verlag, Vienna.

Alegue, J. D. and Gnauck, A. (2006b). Using Fourier Analysis in Investigating Water Quality Indicators. In: Gnauck, A. (Hrsg.): Modellierung und Simulation von Ökosystemen. Workshop Kölpinsee 2005. Shaker Verlag, Aachen, pp. 42-54.

Barthelmes, D. (1981). Hydrobiologische Grundlagen der Binnenfischerei. Fischer, Jena.

Belnap, J. (1998). Environmrntal Auditing: Chossing Indicators of Natural Resource Condition: A Case Study in Arches National Park, UT, USA. Environ. Manage. 22, 635-642.

Berthouex, P. M. and Brown, L. C. (1994). Statistics for Environmental Engineers. Lewis Publishers, London.

Berthouex, P. M., Hunter, W. G. and Pallesen, L. (1981). Waste Water Treatment: A Review of Statistical Applications. Econometrics 81-selected Papers, pp. 77-99.

Bhagavan, C. (1985). On Non-stationary Time Series. In: Hanna, E., Rao, M. and Krishnaiah, P (ed.): Handbook of statistics, vol. 3. Elsevier science publishers B.V. (1985) 311-320.

Bloomfield, P. (1976). Fourier Analysis of Time Series: An introduction. Wiley, New York.

Boon, P. J. (1992). Essential Elements in the Case for River Conservation. In: Boon PJ, Calow P, Petts GE (eds) River Conservation and Management, pp. 11-34. John Wiley and Sons, Chichester.

Borney, H. (1994). Pesticides in Ground and Surface Water. Vol. 9 of Chemistry of Plant Protection. Essays on the Fate of Pesticides in Surface Water and Ground Water Including Methods to Minimize Water Pollution from Pesticides. Springer-Verlag, New York.

Botkin, B. D. and Keller, E. A. (2005). Environmental Science: Earth as a Living Planet. John Wiley and Sons, Inc. pp 436.

Box, G. E. P., Hunter, W. G. and Hunter, J. S. (1978). Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building. Wiley Interscience, New York.

Box, G. E. P., Jenkins, G. M. and Reinsel. G. C. (1994). Time Series Analysis, Forecasting and Control, 3rd ed. Prentice Hall, NJ.

Boyd, E. (2000). Water Quality: An Introduction. Kluwer Academic Publishers, Boston.

Brillinger, D. R. (1981). Time Series: Data Analysis and Theory. Holden-Day Series in Time Series Analysis. Holden-Day, San Francisco.

Bruce, A. and Gao, H. Y. (1996). Applied Wavelet Analysis with S-PLUS. Springer-Verlag, New York.

Campbell, G., and S. Wildberger. (1992). The Monitor's Handbook. LaMotte Company, Chestertown, MD. 71 pp.

Chapman, D. (1996). Water Quality Assessment: A Guide to the Use of Biota, Sediments and Water in Environmental Monitoring. Chapman & Hall, New York.

Chiras, D. (1998). Environmental Science: A Systems Approach to Sustainable Development. Wadsworth Publishing Company, Boston. pp 407-425.

Chui, C. K. (1992). An Introduction to Wavelets. Academic Press, New York.

Cunningham, M. A. and Saigo, B. W. (2003). Environmental Science: A global Concern. McGraw Hill, Montreal. pp 447-474.

Dale, H. V. and Beyeler, C. S. (2001). Challenges in the Development and Use of Ecological Indicators. Ecological Indicators, 1:3-10.

Daubechies, I. (1988). Orthogonal Bases of Compactly Supported Wavelets. Comm. Pure Appl. Math. 41, 909-996.

Daubechies, I. (1990). The Wavelet Transform, Time-Frequency Localization and Signal Analysis. IEEE Trans. Info. Theory, 36(5):961-1005.

Daubechies, I. (1992). Ten Lectures on Wavelets. Volume 61 of CBMS-NSF Regional conference series in applied mathematics. Society for industrial and applied mathematics, Philadelphia.

Debnath, L. (2002). Wavelet Transforms and their Applications. Birkhäuser, Berlin.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to Unknown Smoothness via Wavelet Shrinkage. Biometrika, 81, 425-455.

Dunne, T. and Leopold, L. B. (1988). Water and Environmental Planning. In: Freeman W. H. (ed.). A great summary and detailed examination of water resources and problems. San Francisco.

Elmore, H. L. and Hayes, T. W. (1960). Solubility of Atmospheric Oxygen in Water. J. of Env. Eng. Div. ACSE 86 (SA4), 41-53.

Foufoula-Georgiou, E. and Kumar, P. (1994). Wavelets in Geophysics. Vol. 4 of wavelet analysis and its applications. Academic press, Inc., San Diego.

Gencay, R., Selcuk, F. and Whitcher, B. (2002). An Introduction to Wavelet and other Filtering Methods in Finance and Economics. Academic Press, New York.

Gerry, P. Q. and Michael, J. K. (2002). Experimental Design and Data Analysis for Biologists. Cambridge University Press.

Gnauck, A. (2006). Functoren und Signale-zur Signaleanalyse Ökologischer prozesse. In: Gnauck, A. (Hrsg.): Modellierung und Simulation von Ökosystemen. Workshop Kölpinsee 2004. Shaker Verlag, Aachen, pp. 192-224.

Gnauck, A. and Luther, B. (2004). Simulation and Parameter Optimisation of a Shallow Lake Eutrophication Model. In: Gnauck, A. (Hrsg.): Modellierung und Simulation von Ökosystemen. Workshop Kölpinsee 2004. Shaker Verlag, Aachen, pp. 13-29.

Golubev, Y., Härdle, W., Hlavka, Z., Klinke, S., Neumann, M. H., and Stefan S. (2000). Wavelets. In: Härdle, W., Klinke, S. and Müller, M. (eds). X-plore-Learning Guide. Springer-verlag, Heidelberg.

Gore, J. A. and Petts, G. E. (1989). Alternatives in Regulated River Management. CRC Press, Boca Raton, Florida.

Granger, C. and Engle, R. (1983). Applications of Spectral Analysis in Econometrics. In: Brillinger, D. and Krishnaiah, P (ed.): Handbook of Statistics, vol. 3. Elsevier science publishers B.V. (1983) 93-109.

Grenander, U. (1981). Abstract Inference. Wiley, New York.

Grifoni, R. C. and Passerini, G. (2004). Statistical Modelling of the Remediation of Environmental Data Time Series. In: Latini, G. and Passerini, G. (eds.). Handling missing data : Applications to Environmental Analysis. Witpress, Boston. pp 51-89.

Grigg, N. (ed.)(1996). Water Resource Management: Principles, Regulations and Cases. McGraw Hill, Montreal.

Haan, C. T. (2002). Statistical Methods in Hydrology. Second Edition. Iowa State University Press, Ames, Iowa, USA: 128 -160.

Hansaker, C. J. Carpenter, D. L. (1990). Environmental Monitoring and Assessment Program. Ecological Indicators. EPA 600/3-90/060. US EPA. Office Research Development Research, Research Triangle Park, NC, USA.

Hess-Nielsen, N. and Wickerhauser, M. V. (1996). Wavelet and Time-Frequency Analysis. Proceedings of the IEEE, 84, No. 4, 523-540.

Hipel, K. W. and McLeod, A. I (1994). Time Series Modelling of Water Resources and Environmental Systems. Elsevier, Amsterdam.

Hoaglin, J. D., Mosteller, F. and Tukey, J. W. (1983). Understanding Robust and Exploratory Data Analysis. Wiley, New York.

Hunter, J. S. (1977). Incorporating Uncertainty into Environmental Regulations. National Academy of Science, Washington, D.C.

Hunter, J. S. (1980). The National Measurement System. Science, 210, 869-874.

Hunter, W. G. (1982). Environmental Statistics. In: Kotz and Johnson (eds.). Encyclopedia of Statistical Sciences. Vol. 2. John Wiley, New York.

Jenkins, G. M. and Watts, D. G. (1968). Spectral Analysis and its Applications. Holden-Day, San Francisco.

Jones, R. H. (1980). Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations. Technotrics 22(3), 389-395

Jørgensen, S. E. (1997). Acidification of Scandinavian Lakes. In: Jorgensen, S. E. and Matsui, S. (eds). Guidelines of Lake Management. Vol.8. The World Lakes in Crisis. ILEL and UNEP.

Jørgensen, S. E. (1994). Fundamentals of Ecological Modelling. Elsevier, London.

Karr, J. R. (1997a). Rivers as Sentinels: Using the Biology of Rivers to Guide Landscape Management. In: Bilby, R. E., and Naiman, R. J., (eds). The Ecology and Management of Streams and Rivers in the Pacific Northwest Coastal Ecoregion. Springer-Verlag, New York.

Karr, J. R. (1997b). The Future is Now: Biological Monitoring to Ensure Healthy Waters. In: Streamkeepers: Aquatic Insects as Biomonitors. Xerces Society, Portland. pp. 31-36.

Keinert, F. (2004). Wavelets and Multi-wavelets. Chapman and Hall/CRC, London.

Klapper, H. (2003). The Assault, Management and Reversal of Eutrophication. In: O'Sullivan, P. E. and Reynolds, C. S. (eds). The Lake Handbook. Vol. 2: Lake Restoration and Rehabilitation. Blackwell, Malden.

Klose, H. (1995). Die Eutrophirung der Havel und ihr Bestimmender Einfluß auf Ökosystem und Nutzungen. In: Landesumweltamt Brandenburg (Hrsg.): Die Havel. Studien und Tagungsberichte, Bd. 8, Potsdam, pp. 16-32.

Koopmans, L. H. (1974). The Spectral Analysis of Time Series. Academic Press, New York.

Koopmans, L. H. (1983): A Spectral Analysis Primer. In: D. R. Brillinger and P. R. Krishnaiah (eds.): Handbook of Statistics, vol 3. Elsevier Science Publishers, Amsterdam. pp. 169-183.

Koopmans, L. H. (1995). The Spectral Analysis of Time Series. San Diego, CA. Academic Press.

Koren, H. (1991). Handbook of Environmental Health and Safety: Principles and Practices. Lewis Publishers, Florida.

Kumar, P. and Foufoula, E. (1997). Wavelet Analysis of Geophysical Applications. Review of Geophysics. 35(4), 385-412.

Landner, L. (1989). Chemicals in the Aquatic Environrment: Advanced Hazard Assessment. Springer-Verlag, Berlin.

Landres, P. B. (1992). Ecological Indicators: Panacea or Liability? In: McKenzie, D.H., Hyatt, D.E. and McDonald, V.J., eds, Ecological Indicators, Vol. 2. New York, Elsevier Applied Science, 1295–318.

Lau, K. M. and Weng, H. Y. (1995). Climate Signal Detection using Wavelet Transform: How to Make a Time Series Sing. Bulletin American Meteorological Society, 76, 2391-2402.

LAWA (1998). Beurteilung der Wasserbeschaffenheit von Fließgewässern in der Bundesrepublik Deutschland – Chemische Gewässergüteklassifikation. Kulturbuchverlag, Berlin.

Laws, A.E. (2000). Aquatic Pollution: An Introductory Text. John Wiley and Sons, New York.

Little, R. J. A. and Rubin, D. B. (1983). Missing Data in Large Data Sets. In: Wright, T. (ed.): Statistical Methods and the Improvement of Data Quality. Academic Press, London, pp. 73-82.

Little, R. J. A. and Rubin, D. B. (1987). Statistical Analysis of Missing Data. Wiley, Chichester.

Lorenz, C. M., Gilbert, A. J., Calino, W. P. (1999). Indicators for Transboundary River Management. Environmental management. 22:483-493.

Mallat, S. (1989). A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. IEEE Trans. Patt. Recog. and Mach. Intell., 11(7):674-693.

Mallat, S. (1998): A Wavelet Tour of Signal Processing . Academic Press, New York.

Maybeck, M., Chapman, D. and Helmer, R. (eds.)(1989). Global Freshwater Quality: A First Assessment . Blackwell Rference, Oxford.

McCoy, E. J. and Walden, A. T. (1996). Wavelet Analysis and Synthesis of Stationary Long-Memory Processes. J. Comp. Graph. Stats., 5, 26-56.

Nesmeřak and Straškraba, M. (1985). Spectral Analysis of Automatically Recorded Data from Slapy Reservoir, Czechoslovakia. Int. Reveue ges. Hydrobiol. 70(1), 27-46.

Nirmalakhandan, N. (2002). Modelling Tools for Environmental Engineers and Scientists. CRC Press, New York.

Novotny, V. (2003). Water Quality: Diffuse Pollution and Watershed Management. John Wiley and Sons, New Jersey.

OECD. (2001). Environmental Indicators for Agriculture. Vol. 3. Methods and Results. OECD, Paris.

Ongley, D. E. (1999). Water Quality, an Emerging Global Crisis. In: Walling, D. E. and Bruce, W. (eds). Water Quality, Processes and Policy. Wiley and Sons, New York.

Pahl-Wostl, C. (1995). The Dynamic Nature of Ecosystems. Wiley, Chichester.

Parzan, E. (1983). Autoregressive Spectral Estimation. In: Brillinger, D. R. and Krishnaiah, P. R. (eds). Time Series in the Frequency Domain, Handbook of Statistics, vol. 3. Amsterdam: North Holland. pp. 211-243.

Pepper, I. L., Gerba, C. P., Brusseau, M. L. (1996). Pollution Science. Academic Press, London.

Percival, D. B. (1995). On Estimation of Wavelet Variance. Biometrika 82(3), 619-631.

Percival, D. B. and Guttorp, P. (1994). Long-memory Processes, the Allan Variance and Wavelets. Journal od the American statistical association 92 (439), 868-880

Percival, D. B. and Walden, A. T. (1993). Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques. Cambridge Unversity Press.

Percival, D. B. and Walden, A. T. (2000). Wavelet Methods for Time Series Analysis. Cambridge University Press.

Percival, D. B., Wang, M. and Overland, J. E. (2004). An Introduction to Wavelet Analysis with Applications to Vegetation Monitoring, Community Ecology, 5(1), pp. 19-30.

Phinney, J. (1999). “Nurients and Toxics.” In: Meeting Notes—U.S. Environmental Protection Agency (USEPA)/Center for Marine Conservation (CMC) workshop: Volunteer Estuary Monitoring: Wave of the Future. Astoria, OR: May 19-21, 1999.

Platt, T. and Denman, K.L. (1975). A General Equation for the Mesoscale Distribution of Phytoplankton in the Sea. Mémoires de la Société Royale des Sciences de Liège, 6th Series, 7, 31-42.

Priestley, M. B. (1981). Spectral Analysis and Time Series. Academic Press, London.

Rebecca, W. (1998). Spectral Analysis of Time-series Data. Guilford Press, New York.

Ripl, W. and Wolter, K. D. (2003). The Assault on the Quality and Value of Lakes: In: O’Sullivan, P. E. and Reynolds, C. S. (eds). The Lake Handbook. Vol. 2: Lake Restoration and Rehabilitation. Blackwell, Malden. pp 25-64

Roberts, C. N. (1993). The Changing Global Environment. Blackwell Publishers, Oxford.

Shear, H., Bertram, P., Forst, C. and Horvatin, P. (2005). Development and Application of Ecosystem Health Indicators in the North American Great Lakes Basin. In: Jørgensen, E. S., Costanza, R. and Xu, F. (eds.): Handbook of Ecological Indicators for Assessment of Ecosystem Health. Taylor and Francis, New York.

Shumway, R. H. (2005). Time Series Analysis and its Applications. Springer, New York.

Shumway, R. H. (2000). Time Series Analysis and its Applications. Springer-Verlag, Heidelberg.

Smith, W. S. (1997). The Scientist and Engineer's Guide to Digital Signal Processing. California Technical Publishing, San Diego, pp. 141-167.

Stork, N. E., Boyle, T. J. B., Dale, V. H., Eeley, H., Finegam, B., Lawes, M., Manokaran, N., Prabhu, R., Sorberon, J. (1997). Criteria and Indicators for Assessing the Sustainability of Forest Management: Conservation of biodiversity. Centre for international forestry research. Working paper No. 17, Bogar, Indonesia.

Strang, G. and Nguyen, T. (1997). Wavelets and Filter Banks. Wellesley-Cambridge Press, Wellesley MA.

Straškraba, M. (2003). Reservoirs and other Artificial Waterbodies. In: O'Sullivan, P. E. and Reynolds, C. S. (eds). The Lake Handbook. Vol. 2: Lake restoration and rehabilitation. Blackwell, Malden. Pp. 300-320.

Straškraba, M. and Gnauck, A. (1985). Freshwater Ecosystems: Modelling and Simulation. Elsevier, Amsterdam.

Streeter, W. H. and Phelps, E. B. (1925). A Study of the Pollution of Natural Purification of the Ohio River. Public Health Bull. 146, US Public Health service, Washington DC.

Suter, G. W. (1993). A Critique of Ecosystem Health Concepts and Indexes. Environ. Toxicol. Chem. 12, 1533–1539.

Tchobanoglous, G. and Schroeder, E.D.(1985). Water Impurities and Public Health. Water Quality. Addison-Wesley Publishing Company, Massachusetts. pp. 192-200.

Thanh, N. C. and Tam, D. M. (1993). Surface Water Quality Management. In: Thanh, N. C. and Biswas, A. K. (eds): Environmentally Sound Water Management. Oxford University Press, Bombay.

Torrence, C. and Compo, G. P. (1998). A Practical Guide to Wavelet Analysis. Bulletin American Meteorological Society, 79, 61-78.

Tukey, J. W. (1977). Exploratory Data Analysis. Addison- Wesley, Reading.

Uhlmann, D. (1991). Anthropogenic Pertubation of Ecological Systems. A Need of Transfer from Principles to Applications. In: Ravera, O.(ed.): Terrestrial and Aquatic Ecosytem Pertubation and Recouvery. Ellis Horwood, New York, pp. 47-61.

UNEP. (2002). Global Environment Outlook – 3. http://grida.no/geo/geo3/english/pdf.htm- (Last accessed, May 14, 2007)

US EPA (1992). The Quality of Our Nations Water. Washington Doc:EPA. Good survey of problems and solutions.

US EPA. (1997). Volunteer Stream Monitoring: A Method Manual. EPA 841-B-97003.http://www.swrcb.ca.gov/funding/docs/grantinfo/10monitoring_questions.pdf (last accessed May 9, 2007).

US EPA. (2000). Evaluation Guidelines for Ecological Indicators. In: Jackson, L. E., Kurtz, J. C, Fisher, W. S. (eds). EPA/620/R-99/005. US EPA Office of research and Development, Research triangle Park, NC, USA.

Venables, W. N. and Ripley, B. D. (1994). Modern Applied Statistics with S-PLUS. Springer Verlag, New York.

Viessman, W. Jr. and Hammer, J. M. (1998). Water Supply and Pollution Control. Addison Wesley Longman, Menlo Park, California.

Whitcher, B. (1998). Assessing Non-stationary Time Series using Wavelets. Ph.D. Thesis, University of Washington.

Whitcher, B., Guttorp, P. and Percival, D. B. (2000). Wavelet Analysis of Covariance with Application to Atmospheric Time Series. Journal of Geophysical Research, 105, No. D11, 14, 941-14, 962.

Wickerhauser, M. V. (1994). Adapted Wavelet Analysis from Theory to Software. A K Peters.

Young, P. C. and Young, T. (1992): Environmentric Methods of Non-stationary Time-Series Analysis: Univariate Methods. In: Hewitt, C. N. (ed.) Methods of Environmental Data Analysis. Elsevier Applied Science, London. pp 37-77.