MODELLING OF DYNAMIC WATER QUALITY CHANGES BY MEANS OF TIME SERIES METHODS

Jean Duclos Alegue[1], Albrecht Gnauck[1], Bernhard Luther[1]

[1]Brandenburg University of Technology at Cottbus,
Dept. of Ecosystems and Environmental Informatics,
Konrad-Wachsmann-Allee 1, D – 03046 Cottbus,
Phone +49-355-692713, Fax +49-355-692743,

jeduclo@yahoo.com *(Jean Duclos Alegue)*

Abstract

Water quality indicators are used to capture physical, chemical and biological changes occurring in freshwater bodies. These dynamic changes result from activities of internal and external driving forces which may be natural or anthropogenic. Modeling of water quality dynamics is of importance in understanding the underlying network structure of water quality changes and to forecast their general tendency within reasonable limits. For modeling and simulation of water quality processes it is necessary that all data sets are based on regularly time grids. The dynamics captured reveals changing amplitudes and variances across time. This paper examines the applicability of three approaches (the autoregressive moving average method, the Fourier polynomial, and digital filter algorithm methods) for modeling water quality changes by time series methods. The different types of indicators namely physical (water temperature), chemical (dissolved oxygen) and biological (chlorophyll-a) are taken from rivers of different hydraulic structures in Germany, namely the River Havel, the Elbe River and the Oder River. The autoregressive moving average method gives acceptable results, but is not helpful for forecasting. The Fourier polynomial is useful for approximating physical indicators and gives unacceptable results for chemical and biological indicators while the filter algorithms give acceptable approximations for all indicators and possibilities for forecasting.

Keywords: ARMA, Fourier polynomial, Filter algorithm, Water quality time series.

Presenting Author's biography

Jean Duclos Alegue. The presenting author is a Cameroon national. After an M.Sc. in Environmental and Resource Management at BTU Cottbus, Germany (2002-2004), he commenced a PhD at the Department of Ecosystems and Environmental Informatics at the same university where he teaches Modeling of Ecosystem and Data Analysis with Matlab. He is expected to present his dissertation before winter 2007. The author is very familiar and comfortable with ecosystem modeling, environmental data analysis, GIS / remote sensing and Java programming language.

ISBN 978-3-901608-32-2

1 Introduction

Freshwater quality data collected by monitoring programs is extremely important for assessing the ecological condition of a freshwater body and for developing an understanding of the interrelationship between the components of the freshwater body. The data collected usually reveals the existence of continuous change over time resulting from internal and external natural and anthropogenic driving forces [1]. It is necessary to model these time series so as to develop an understanding of the structure and the mechanism producing the signals. This is also helpful for predicting the simulating changes within a freshwater body. Unfortunately, data from monitoring programs contain a lot of inconsistencies that affect the results of any analysis [2]. They are as a result of missing values, impossible values, inconsistent values and unlikely values. Errors in water quality time series lead to some general problems in water quality research and simulation. They cause not only difficulties in process identification and parameter estimation but also misinterpretations of spatial and temporal variations of water quality indicators [3]. Mostly, time series represent samples of data at discrete time events based on various sampling intervals. For modeling 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 analyzing 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 [4]. 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.

This paper investigates the use of three time series methods for approximating water quality time series so as to identify the most appropriate approach for modelling these signals from the investigated watershed. The Fourier polynomial, the autoregressive moving average and the digital filter algorithm methods for time series approximation will be used for the investigation. The data consist of water quality indicators like conductivity, chlorophyll-a and dissolved oxygen taken from rivers of different hydraulic structures in Germany, namely the River Havel, the Elbe River and the Oder River. Outliers were removed and missing data was replaced by means of linear interpolation method from the data sampled at daily interval.

2 Autoregressive Moving Average Method

Combining an autoregressive model (AR) with a moving average model (MA) an ARMA (autoregressive moving average) model is obtained which is given by:

$$X_t = \partial + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots + \Phi_p X_{t-p} + A_t - \o_1 A_{t-1} - \o_2 A_{t-2} + \ldots + - \o_q A_{t-q} \qquad (1)$$

where X_t is the time series, μ is the mean and A_t is the white noise, $\Phi_1,\ldots,\Phi_p$ and $\o_1, \ldots,\o_q$ are parameters of the model [5]. The p denotes the order of AR and q is the MA order which together give rise to ARMA (p, q). The first step consist in rendering the time series stationary, determining the order (p and q) of the most appropriate model by means of the lowest AIC, the autocorrelation and autocorrelation plots. Next, the parameters of the model are estimated and finally, a diagnostic test is carried out to investigate the goodness of the fit. The ARIMA technique proved effective in capturing water quality changes across time, provided the signal where rendered stationary [6]. However, no reasonable forecast could be obtained from ARIMA models as can be seen from figs. 1, 2 and 3.

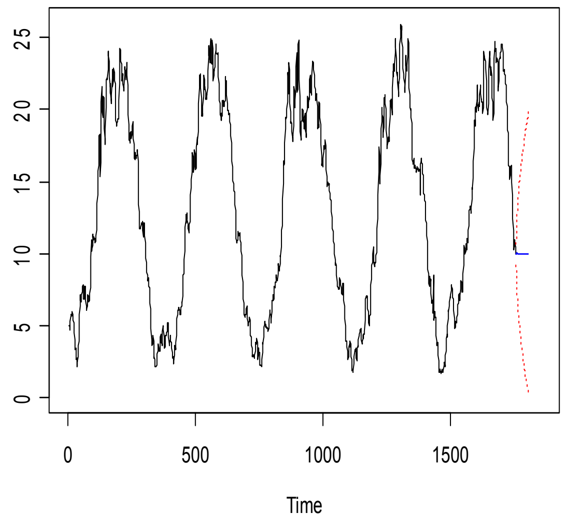

Fig 1: ARIMA(3,1,3) model of water temperature with a forecast of 50 periods

For water temperature time series, an ARIMA(3,1,3) model was found to be the most appropriate as a result of its low AIC(1441.61) and is given by the following equation :

$$X_t = 0.36 + 1.29_{(0.16)}X_{t-1} - 0.99_{(0.26)}X_{t-2} + 0.21_{(0.13)}X_{t-3} + 0.75_{(0.16)}A_{t-1} - 0.55_{(0.21)}A_{t-2} - 0.14_{(0.05)}A_{t-3}$$

The standard errors of the estimated coefficients are given in the subscripted values in brackets. Given a lower and upper 95 % confidence interval as shown in figure 2, a forecast of 50 days ahead is extremely poor. Hence, the model is suitable for approximating but not forecasting the time series.

An ARIMA(1,1,2) model was found to be the most appropriate for dissolved oxygen as a result of the lowest AIC(3060.75) and is given by the following equation:

$$X_t = 0.58+ 01.9_{(0.02)}X_{t-1} + 0.61_{(0.03)}A_{t-1} + 0.37_{(0.02)}A_{t-2}$$

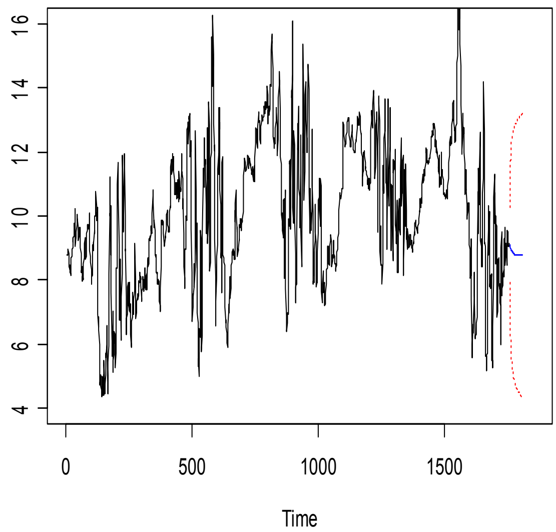

Fig 2: ARIMA(1,1,2) model of dissolved oxygen with a forecast of 50 periods

Similar to the water temperature model, a forecast of 50 periods ahead for the dissolved oxygen ARIMA model given unacceptable values as if revealed in figure (5).

Chlorophyll-a time series is best modelled by an ARIMA(3,1,3) which yields the lowest AIC(11460) and is given by the following equation:

$$X_t = 6.32 - 0.11_{(0.12)}X_{t-1} - 0.86_{(0.07)}X_{t-2} + 0.19_{(0.11)}X_{t-3} - 0.29_{(0.12)}A_{t-1} + 0.90_{(0.02)}A_{t-2} + 0.39_{(0.11)}A_{t-3}$$

The ARIMA model also yields an unsatisfactory forecast for the chlorophyll-a time series. This simply implies that the method is inappropriate for predicting the time series of water quality indicators, though they provide excellent approximations.

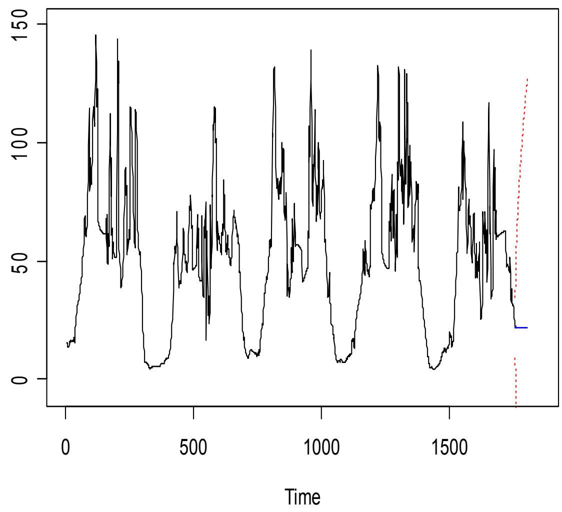

Fig 3: ARIMA(3,1,3) model of chlorophyll-a with a forecast of 50 periods

3 Fourier Polynomial Method

The Fourier polynomial provides a means of approximating periodic functions by sums of sine and cosine functions, shifted and scaled [7]. This polynomial is of the form

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

with

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

and

$$b_k = \int_0^{2\pi} f(t)\sin(kt)dt \qquad (3)$$

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, a_0, a_i and b_i, i =1, . . ., n are the Fourier coefficients of the $F_n(t)$, and the Fourier polynomial is 2π periodic.

Fig. 4 shows a Fourier polynomial of 7th order for dissolved oxygen, one of the most important chemical water quality indicators with the following equation:

$$\begin{aligned} f(t) = {} & 10.23 - 0.13\cos(0.006t) - 0.75\sin(0.006t) \\ & - 0.41\cos(0.012t) - 0.52\sin(0.012t) \\ & - 0.26\cos(0.018t) + 1.2\sin(0.018t) \\ & + 0.2\cos(0.024t) - 0.27\sin(0.24t) \\ & - 0.38\cos(0.03t) - 0.07\sin(0.03t) \\ & - 0.58\cos(0.036t) + 0.53\sin(0.036t) \\ & - 0.16\cos(0.042t) + 0.39(0.042t) \end{aligned}$$

ISBN 978-3-901608-32-2

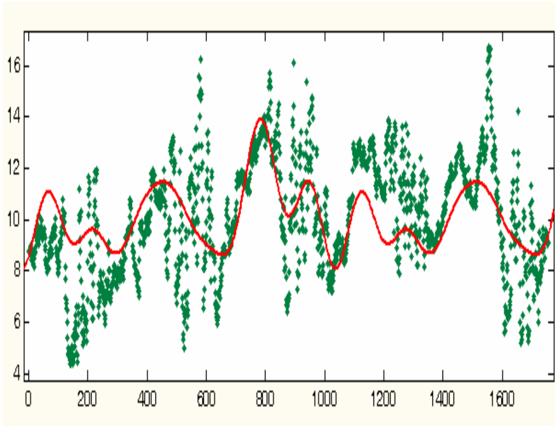

Fig 4: Seventh order Fourier polynomial of dissolved oxygen

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

In the case of a biological indicator chlorophyll-a given in fig. 5, 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:

$$
\begin{aligned}
f(t) = {} & 48.76 + 4.7\cos(0.0087t) + 0.47\sin(0.0087t) \\
& - 29.64\cos(0.02t) - 1.4\sin(0.02t) \\
& - 2.48\cos(0.03t) + 1.3\sin(0.03t) \\
& - 20.35\cos(0.04t) + \sin(0.04t)
\end{aligned}
$$

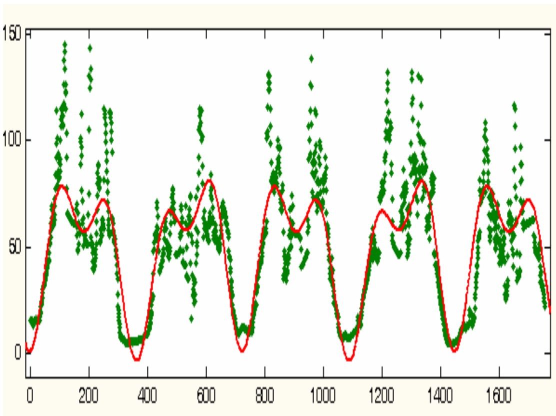

Fig 5: Fourth order Fourier polynomial of phytoplankton biomass

As can be seen, the Fourier approximation describes the time dependent changes of phytoplankton biomass more or less correctly during the transient reaches spring and in fall. 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.

Fig. 6 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 polynomial of third order is given by

$$
\begin{aligned}
f(t) = {} & 13 + 0.39\cos(0.009t) + 0.018\sin(0.009t) \\
& - 9.71\cos(0.018t) - 2.47\sin(0.018t) \\
& - 0.4018\cos(0.027t) + 0.23\sin(0.027t)
\end{aligned}
$$

where $R^2 = 0.95$ and adjusted $R^2 = 0.95$. The investigation revealed that the Fourier approximation is not appropriate for approximating the dynamics of chemical and biological changes in a freshwater body. It however proved quite effective for approximating variations of a physical water quality indicator such as water temperature. This is because the physical changes do not significantly vary their amplitudes and phases across time compared to the biological and chemical changes.

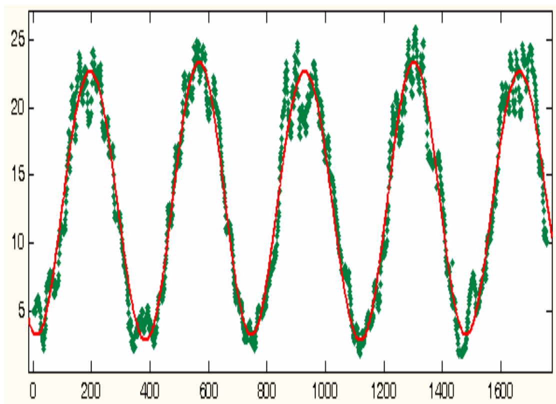

Fig 6: Third order Fourier polynomial of water temperature

4 Digital Signal Filter Algorithm Methods

In opposite of the first two methods, digital signal filter models deliver consistent equidistant data estimates based on major signal frequencies. An ideal low pass with the gain characteristic $|H(\omega)|^2 = 1/(1+F(\omega^2))$ work as distortionless system. In table 1 some well-known low pass filters are given. Mainly filters of order 1 to 3 are useful for modelling water quality changes (table 2) where the 95% significance level was used as decision criteria for acceptance.

Tab 1: Digital filter functions

Filter method	Filter equation
Butterworth	$\lvert H(\omega)\rvert^2 = 1/(1 + \omega^{2n})$
Chebychev 1	$\lvert H(\omega)\rvert^2 = 1/(1 + \varepsilon^2 c_n^2(\omega))$ ε - ripple factor, $c_n(\omega)$ - Chebychev-polynomial of order *n*
Chebychev 2	$\lvert H(\omega)\rvert^2 = 1/(1 + \varepsilon^{*2} c_n^{2n}(\omega))$, $\varepsilon^* = 2\varepsilon/(1-\varepsilon)$
Cauer	$\lvert H(\omega)\rvert^2 = 1/(1 + \varepsilon^2 F_n^*(\omega^2))$, $F_n^*(\omega^2)$ - characteristic function

Tab 2: Applications of digital low pass filters on water quality changes

Indicator	1. order filter	2. order filter	3. order filter
DOC	Cauer	no filter acceptable	no filter acceptable
conductivity	Butterworth, Cauer Chebychev1	Butterworth	Butterworth, Cauer, Chebychev1
NH4-N, NO2-N, NO2-N	Cauer	no filter acceptable	no filter acceptable
DO	Butterworth, Cauer Chebychev1	Butterworth	Butterworth, Chebychev1, Cauer
o-PO4-P	Cauer	no filter acceptable	no filter acceptable
pH	Butterworth, Cauer Chebychev1	Butterworth, Chebychev2	Butterworth, Chebychev1, Chebychev2,Cauer
water flow	Butterworth, Chebychev1, Cauer	Butterworth, Chebychev1	Butterworth, Chebychev1, Cauer
water temperature	Butterworth, Chebychev1, Cauer	Butterworth, Chebychev1, Chebychev2, Cauer	Butterworth, Chebychev1, Chebychev2, Cauer

Digital signal filter models will be distinguished by differences in the pass band and by the ripple effect. Higher order filters cause strong ripple effects in the pass band as well as in stop band. Examples of water quality modelling by means of digital signal filter algorithms for running waters are presented in figs. 7, 8, and 9. Mainly filters of order 1 to 3 are useful for modelling water quality changes (tab. 2) where the 95% significance level was used as decision criteria for acceptance. Limit frequencies are selected by means of power spectra.

As can be seen in fig. 7, the filter algorithm gives an acceptable result for a chemical indicator concerning periodicity and phase. The amplitudes of the model output are lower than in reality.

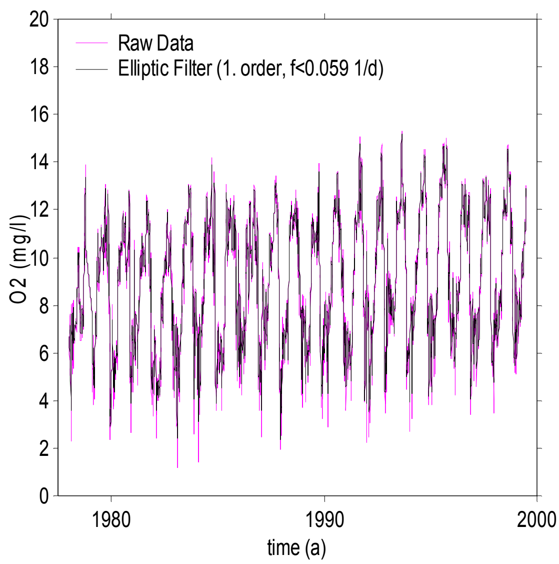

Fig 7: Digital signal filter model of DO

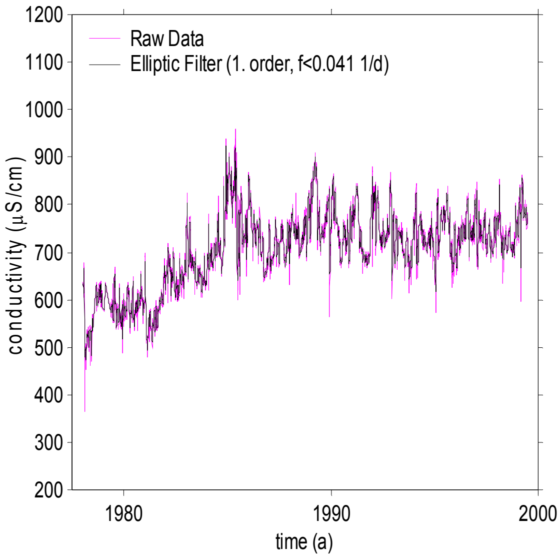

Fig 8: Digital signal filter model of conductivity indicating industrial pollution

In opposite of that, dynamic changes of water flow can be modelled quite well by different filter algorithms (fig 9, see tab. 2). In the case of biologically induced water quality changes high pass filters have to be used.

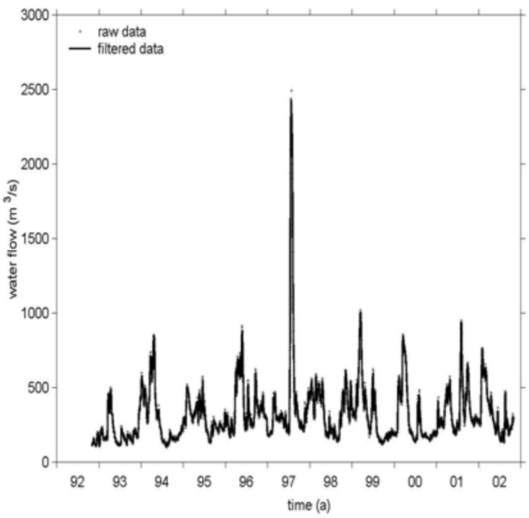

Fig 9: Elliptic filter model for water flow,

f < 0.067 1/d

5 Conclusions

The investigations reveal that the Fourier polynomial is inappropriate for approximating biological and chemical water quality time series. Physical water quality signals such as water temperature yield decent results. This is because the Fourier approach requires stationary time series because it uses sinusoidal sine and cosine basis functions. The changing amplitudes and periods present especially in chemical and biological water quality time series render this method ineffective in capturing such changes in the signals. In addition, the autoregressive moving average method is quite able to provide decent approximations for water quality indicator only when they have been rendered stationary either by differencing or removing any trend present. Unfortunately, any attempt to forecast or predict future values by means of this method gives unacceptable results. Finally, the digital filter algorithm methods give acceptable results and prove quite useful for approximating and capturing changes in the water quality time series with great possibilities for forecasting the approximated signals. It becomes necessary to also investigate the use of non-linear time series modelling methods by genetic algorithms for freshwater quality data.

5 References

[1] Straškraba, M., Gnauck, A. (1985): Freshwater Ecosystems – Modelling and Simulation. Fischer, Stuttgart/Jena.

[2] Bhangu, I. and P. H. Whitfield (1997): Seasonal and Long-Term Variations in Water Quality of the Skeena River at USK, *British Columbia. Water Research* ***31****(9), 2187-2194.*

[3] 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

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

[5] Vecchia, A. V. (1985): Periodic autoregressive-moving averages (PARMA) modeling with applications to water resources. *Water Resource. Bull.* ***21****(4), 721-730.*

[6] Young, P. C (1999): Nonstationary time series analysis and forecasting. *Progress in Environm. Sci.* ***1****(1*), *3-48.*

[7] Franses, P. H. (1999): Periodicity and Structural Breaks in Environmetric Time Series. In: Mahendrarajah, S., A. J. Jakeman and M. McAleer (eds.) Modelling Change in Integrated Economic and Environmental Systems. Wiley, New York.

ISBN 978-3-901608-32-2