DOI: 10.2429/proc.2016.10(1)047 2016;10(2)

Elena NEVEROVA-DZIOPAK[1] and Zbigniew KOWALEWSKI[1]

MODELLING OF CHLOROPHYLL-A CONTENT IN RUNNING WATERS

MODELOWANIE ZAWARTOŚCI CHLOROFILU-A W WODACH PŁYNĄCYCH

Abstract: Chlorophyll-a is one of the key parameters for assessment of trophic status of surface waters. However, Polish standard environmental monitoring procedures assume a low frequency of chlorophyll measurements in running waters, which does not provide the possibility of permanent control of eutrophication process and taking the appropriate preventive and protective measures sufficiently in advance. The article is focusing on constructing of predicting model of chlorophyll-a content based on data obtained within monitoring realized by Regional Inspectorates for Environmental Protection. Multivariate linear regression (MLR) model for chlorophyll-a content prediction was formulated on the base of chosen parameters like: pH, oxygen saturation, different forms of nitrogen and phosphorus. Formulation of the model was followed by a test of the applicability of each of the individual components of the regression equation. The main purpose was to develop an algorithm allowing for quick adaptation of model to local conditions in the rivers in order to make a reliable prediction of chlorophyll content.

Keywords: eutrophication, chlorophyll-a, running waters, prediction models

Eutrophication

Anthropogenic eutrophication is a priority global environmental problem. The negative consequences associated with this process are the disturbance of ecological balance, deterioration of water quality, loss of aesthetic properties and decline of biospheric role of aquatic ecosystems [1]. For effective prevention of eutrophication development, its control, prediction and management the properly designed monitoring procedures based on reliable methods of water trophic status assessment are essential. Classical methods for assessing of trophic status are based on the measuring of nitrogen and phosphorus content in surface waters, which are considered to be the major cause of eutrophication, as well as chlorophyll-a content which represents the level of algal biomass production and informs about ecosystem reaction on water enrichment in biogenic matter and the intensity of the process. These parameters are used most frequently in monitoring of eutrophication [2].

There are various methods for measuring the chlorophyll-a content: weight methods (Polish Norm), optical methods (*i.e.* Algaltorch©) or spatial analysis from planes or satellites. Unfortunately, measurements of chlorophyll conducting by Polish monitoring services (Regional Inspectorates for Environmental Protection - RIEP) in the frames of surface water state monitoring program are performed rarely and in small number of measuring points. The situation worsened after 2008 when the part of the measurement point was liquidated and the number of measured standard parameters was reduced. The reason was the high labor- and cost-consuming measurements of chlorophyll. This is an

[1] Department of Environmental Management and Protection, AGH University of Science and Technology, al. A. Mickiewicza 30, 30-059 Kraków, Poland, phone +48 12 617 47 04, email: kowalew@agh.edu.pl, elenad@agh.edu.pl

Contribution was presented during ECOpole'15 Conference, Jarnoltowek, 14-16.10.2015

argument for developing a more pragmatic and low-cost methods for monitoring based on the results of forecasting models. Most of the forecasting statistical models are developed for the lentic systems; in the case of streams and rivers they are less common but equally necessary. The complexity of chlorophyll modelling in running waters is conditioned not only by the specifics of these ecosystems, main limiting factors and water movement, but also by the annual variations of flow intensity caused by the floods [1]. The primary objective of this study is to develop an algorithm for creation of "high-speed" model of chlorophyll-a content, based on data obtained within routine, monthly monitoring measurements carried out by RIEP. Such model is expected to predict the chlorophyll content and fill a gap in the research.

Existing solutions for modeling of chlorophyll-a

In the modern practice of modeling of chlorophyll-a content, there are the different approaches used. The first one is the determination of chlorophyll content based on images obtained from aerial or satellite photography. The second approach is the method in which the selection of the components of the model and its formulation are implemented using artificial neural networks or similar algorithms. The third approach are methods based on statistical modeling using mainly the regression analysis.

Aerial and satellite methods

Among the wide use of spatial applications there is MERIS satellite, designed to control the state of the oceans. It has also become an important tool in estimating the quantities of chlorophyll and biomass in freshwaters [3]. On the basis of satellite NIR measurements (near infrared) the direct correlation models were formulated [4] as well as the models consisting of several parameters with the construction similar to the construction of the aggregate indicators of eutrophication [5, 6]. Previously, the direct measurements were conducted with the help of satellites like: Adeos and Orbview-2 [7].

Artificial neural networks (*ANN*)

The use of artificial neural networks (ANN) is quite common in the modelling of chlorophyll-a content and other biological components in aquatic environment. In modelling process there are used the different neural networks and different methods of recognition, so there is no one universal algorithm for chlorophyll content modelling. It is popular to use the ANN in comprehensive models for predicting chlorophyll changes using the range of parameters including hydrological, hydro-morphology, meteorological data and physical and chemical parameters of the examined area [8]. Next to the neural network it is possible to use heuristic evolutionary algorithm (HEA). This algorithm could create a model that is changing in time simultaneously to environmental data changing [9]. There are also the models based on fuzzy logic co-working with neural networks and evolutionary algorithms [10].

Statistical modelling

Among the statistical models the linear regression should be distinguished. The basis of this method is the least squares method. In addition to the selection of the parameters the correlation analysis can be used (Pearson or Spearman rank) or variance inflation factor (VIF). For determining of coefficients for regression equations there is frequently used the so-called stepwise regression: backward, forward or less often bidirectional. With a limited number of suggested parameters it can be used the Akaike information criterion (AIC) [11, 12]. The formulation of statistical models of chlorophyll content can be based on Principal component analysis (PCA) and Principal component regression (PCR), in particular for the choice of model parameters [13]. There are also statistical methods based on the values of biotic indicators and their coefficient of variations allowing for the formulation of chlorophyll prediction model [14]. An interesting issue is the use of phytoplankton growth model BLOOM II to estimate the chlorophyll content [15]. According to specialist literature the ANN and statistical methods, primarily Multivariate linear regression (MLR) will finally give similar results [16]. High compliance of prediction, both for the test data values as well as for real values of parameter, gives a real possibility to use both of these methods in practice of routine monitoring.

Assessment of model adjust to real values

The basic estimation test for the convergence of prediction values with factual values is correlation analysis. The most common indicator is the Pearson coefficient, which can range from –1 to 1. The best fit is observed when its value is a positive integer and equal 1, while the minimum accepted level is usually determined individually [14]. Other parameters describing the degree of prediction reliability are: Theil Index (TI), Mean Absolute Error (MAE), Standard Uncertainty, Root Mean Square Error (RMSE), Mean Relative Error (MRE) [13, 17, 18], Occasionally Receiving Operating Characteristic (ROC) curve (AUC) [10] or even scatterplots [19].

Research area

The object of research is the river Ner, located in north-western Poland, in the province of Lodz. Together with its tributaries, such as Dobrzynka, Pisia, Beldowka and Nida the river Ner is monitored by the Regional Inspectorate for Environmental Protection (RIEP) located in Lodz city, within the framework of state monitoring. Ner is a river of abiotic type 20 - lowland gravel river. The rivers of this type are characterized by varying flow intensity, with mainly gravel and sand bottom substrate. It was necessary to select the appropriate measurement point where the measurement of chlorophyll was conducted regularly. The research was based on data obtained from monitoring provided by the Inspectorate in the period of years 1992-2014 in chosen measurement point located at 111.7 km of the Ner river, where the content of chlorophyll was measured in the period of 1993-2006. During overall multi-year monitoring the RIEP measured the values of more than 60 parameters, but there are only about 25 parameters were measured regularly every year. The main of them are: water and air temperatures, color of water, solid substances, pH value and dissolved oxygen, BOD and COD measured by two methods, total nitrogen and total phosphorus, conductivity, hardness, concentrations of chlorides, calcium and magnesium and others. The set of 25 values of regularly measured parameters was used to construct a predictive model of chlorophyll-a content.

The regression prediction model of chlorophyll-a content

Algorithm of statistical analysis

Statistical analysis was carried out in stages using Statistica software version 10.0 developed by StatSoft. The first step of the analysis was to verify the compliance of distribution of the parameters derived from the annual monitoring with the normal distribution. Confirmation of the normal character of their distribution allowed for further statistical analysis. Globally for multi-year data it was created the Pearson correlation coefficient matrix to pre-eliminate the parameters depending on each other. It assumes two methods of its creation and verification. The first method is the separation of data groups forming the model and the test group, the second method presupposes that all data will be used for construction of the model, and the whole will be verified by comparing the data for individual years. For formulation of the model the multiple stepwise linear regression (MLR) was used. Then the validity of the model elements was tested by the AIC criterion. The parameters of the model were selected using the linear regression provided the correct final model. For the assessment of conformity with factual values the Pearson's correlation coefficient and root main squared error (RMSE) were used. The elaborated algorithm applied for statistical analysis and further formulation of the model is shown in the Figure 1.

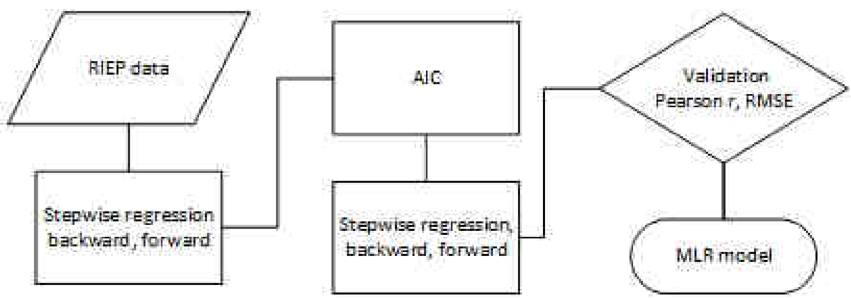

Fig. 1. Algorithm of model creation

Model calculation

The database was formed by the results of measurements of selected 25 parameters in the measurement point located at the 111.7 km of the river Ner, carried out during the monitoring period 1993-2006. In total, there were conducted 136 series of measurements and 346 values of different parameters were determined. The values of 21 parameters were selected as the model input data: chlorophyll-a, air and water temperature, water color, total suspended solids, pH, oxygen saturation, BOD_5, COD_{Mn}, $ChZT_{Cr}$, ammonia, nitrates, nitrites, total nitrogen, phosphates, total phosphorus, total hardness, sulphates, chlorides, calcium, magnesium. The data was characterized by a distribution similar to normal distribution. The highest coefficient of Pearson (r = 0.744) characterized the interdependence of chlorophyll-a and COD_{Mn} values. Then stepwise regression analysis was applied, both backward and forward. The obtained results of statistic test F (Fisher test) allowed the selection of the following parameters for model formulation: suspended solids, BOD_5, COD_{Mn}, nitrites, nitrates, total hardness, dissolved substances, sulfates, calcium and magnesium. The applied AIC analysis made it possible to rank the selected parameters in the following order of priority: sulphates, COD_{Mn}, nitrites, total suspended solids, magnesium, total hardness, dissolved substances, calcium, BOD_5. Then, the construction of appropriate model based on MLR method of five-stage backward analysis were conducted. The last stage of the analysis allowed for final selection of following parameters for model formulation: solid substances (SS), BOD_5, COD_{Mn}, nitrates ($N\text{-}NO_2$) and dissolved substances (DS) Finally the statistical model for prediction of chlorophyll content (Chl^p-a) takes the form of following equation (1):

$$Chl^p\text{-a} = -27.4835 + 0.1678\ SS + 3.2788\ BOD + 6.9097\ COD + (-33.6520)\ N\text{-}NO_2 + (-0.0695)\ DS \quad (1)$$

The graph in Figure 2 shows the values of chlorophyll-a concentrations obtained by modelling during the subsequent stages of the analysis against the factual values of chlorophyll- a concentration. Table 1 shows the values of selected parameters of the model in each of five steps of statistical analysis.

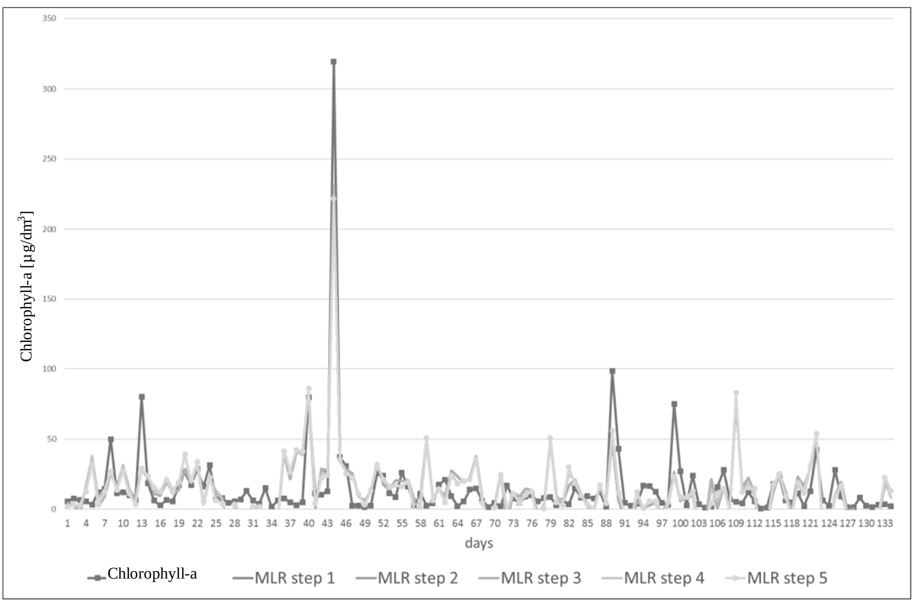

Fig. 2. Chlorophyll values - real and predicted

Using backward linear regression algorithm, with selecting the key parameters of the model using AIC criterion and using the MLR analysis it was possible to formulate the final variant of the model and to achieve the satisfactory results. Table 2 presents the values of

Pearson correlation coefficient and root main squared error RMSE. In each step they achieved a high degree of correlation between the factual values of chlorophyll-a content and the values obtained on the base of formulated model. Due to the fact that all the parameters used for the final MLR analysis are the parameters regularly monitored by RIEP, there is the possibility to formulate the predictive models based on the maximum amount of parameters. The research also justified the possibility of application of the elaborated algorithm for selection of the parameters of predictive model. The permanent application of mentioned algorithm would help to avoid the problem of remodelling and ensure the satisfactory level of model adjustment to real aquatic conditions.

Table 1

Parameters of linear regression of each step statistical analysis

MLR stages	**Step 1**	**Step 2**	**Step 3**	**Step 4**	**Step 5**
Absolute term	–46.146	–45.996	–46.061	–37.916	–27.483
Suspension	0.199	0.199	0.198	0.178	0.167
BOD	3.278	3.274	3.240	3.071	3.278
COD_{Mn}	7.271	7.269	7.284	7.127	6.909
Nitrate	–34.760	–34.840	–35.054	–30.914	–33.652
Dissolved substances	–0.092	–0.092	–0.093	–0.098	–0.069
Hardness	0.119	0.121	0.132	0.071	out
Sulphates	–0.161	–0.162	–0.159	out	out
Calcium	0.036	0.034	out	out	out
Magnesium	0.052	out	out	out	out

Table 2

The values of statistical characteristics at each step of MLR

MLR step	**Number of parameters**	**Mean deviation**	**Standard deviation SD**	**Pearson *r***	**RMSE value**
1	9	14.539	25.20	0.79	18.43
2	8	14.539	25.20	0.79	18.43
3	7	14.534	25.20	0.79	18.43
4	6	14.392	25.01	0.79	18.52
5	5	14.380	24.93	0.79	18.60

Summary and conclusions

The increasing pollution of rivers and the continuing threat of eutrophication problem forces to expand, optimize and improve water monitoring programs for eutrophication. It is therefore the prior necessity for regular control of the value of such parameters as various forms of nitrogen and phosphorus and chlorophyll-a content. Under the conditions where the key parameters of eutrophication are not subjected to routine monitoring it is the need of other ways of obtaining the necessary information. In the absence of information on the monthly values of chlorophyll-a concentrations it is possible to obtain the necessary values on the base of mathematical modelling. In order to create a model that would be readily and widely used by monitoring authorities (RIES) the model must be easy to use and based on regularly measured parameters. The elaborated algorithm of statistical analysis presented in the article allows for the selection of key model input parameters and formulation of a simple mathematical model which is characterized by high degree of reliability of the obtained results. Application of the developed algorithm of statistical analysis which provides the base for formulation of prognostic mathematical models allows the prediction of chlorophyll-a content in order to forecast the dynamics of trophic status changes and to take the appropriate decisions on the integrated management in river basins.

References

[1] Biggs B. Eutrophication of streams and rivers: dissolved nutrient-chlorophyll relationships for benthic algae. J North Am Benthol Soc. 2000;19:17-31.

[2] Neverova-Dziopak E. Podstawy zarządzania procesem eutrofizacji antropogenicznej (Fundumentals of anthropogenic eutrophication management) Kraków: Wydawnictwa AGH; 2010.

[3] Bezy JL, Delwart S, Rast M. A new generation of ocean-colour sensor onboard Envisat. Esa Bull Sp Agency. 2000;103:48-56.

[4] Yacobi Z, Wesley J, Kaganovsky S, Sulimani B, Leavitt B, Gitelson AA. NIR-red reflectance-based algorithms for chlorophyll-a estimation in mesotrophic inland and coastal waters: Lake Kinneret case study. Water Res. 2011;45(7):2428-2436. DOI: 10.1016/j.watres.2011.02.002.

[5] Mishra S, Mishra DR. Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. Remote Sens Environ. 2012;117:394-406. DOI: 10.1016/j.rse.2011.10.016.

[6] Reboreda R, Cordeiro GF, Nolasco R. Modeling the seasonal and interannual variability (2001-2010) of chlorophyll-a in the Iberian margin. J Sea Res. 2014;93:133-149. DOI: 10.1016/j.seares.2014.04.003.

[7] Saitoh S, Iida T, Sasaoka K. A description of temporal and spatial variability in the Bering Sea spring phytoplankton blooms (1997-1999) using satelite multi-sensor remote sensing. Prog Oceanogr. 2002;55(1-2):131-146. DOI: 10.1016/S0079-6611(02)00074-5.

[8] Jeong K-S, Kim D-K, Joo G-J. River phytoplankton prediction model by Artificial Neural Network: Model performance and selection of input variables to predict time-series phytoplankton proliferations in a regulated river system. Ecol Inform. 2006;1(3):235-245. DOI: 10.1016/j.ecoinf.2006.04.001.

[9] Zhang X, Recknagel F, Chen Q, Cao H, Li R. Spatially-explicit modelling and forecasting of cyanobacteria growth in Lake Taihu by evolutionary computation. Ecol Modell. 2015;306:216-225. DOI: 10.1016/j.ecolmodel.2014.05.013.

[10] Malek S, Syed A, Sharifah M, Singh S, Milow P, Salleh A. Assessment of predictive models for chlorophyll-a concentration of a tropical lake. BMC Bioinformatics. 2011;12 Suppl 1(Suppl 13):S12. DOI: 10.1186/1471-2105-12-S13-S12.

[11] Heiskary S, Bouchard W. Development of eutrophication criteria for Minnesota streams and rivers using multiple lines of evidence. Freshw Sci. 2015;34(2):574-592. DOI: 10.1086/680662.

[12] Rocha RRA, Thomaz SM, Carvalho P, Gomes LC. Modeling chlorophyll-a and dissolved oxygen concentration in tropical floodplain lakes (Paraná River , Brazil). Brazilian J Biol. 2009;69(2):491-500. DOI: 10.1590/S1519-69842009000300005.

[13] Cho Kyung H, Kang J-H, Ki Seo J, Park Y, Cha Sung M, Kim Joon H. Determination of the optimal parameters in regression models for the prediction of chlorophyll-a: a case study of the Yeongsan Reservoir, Korea. Sci Total Environ. 2009;407(8):2536-2545. DOI: 10.1016/j.scitotenv.2009.01.017.

[14] Håkanson L, Malmaeus JM, Bodemer U, Gerhardt V. Coefficients of variation for chlorophyll, green algae, diatoms, cryptophytes and blue-greens in rivers as a basis for predictive modelling and aquatic management. Ecol Modell. 2003;169(1):179-196. DOI: 10.1016/S0304-3800(03)00269-2.

[15] Niu L, Van Gelder P, Guan Y, Zhang C, Vrijling JK. Probabilistic analysis of phytoplankton biomass at the Frisian Inlet (NL). Estuar Coast Shelf Sci. 2015;155:29-37. DOI: 10.1016/j.ecss.2014.12.049.

[16] Wu N, Huang J, Schmalz B, Fohrer N. Modeling daily chlorophyll a dynamics in a German lowland river using artificial neural networks and multiple linear regression approaches. Limnology. 2013;15(1):47-56. DOI: 10.1007/s10201-013-0412-1.

[17] Zhao X, Zhang H, Tao X. Predicting the short-time-scale variability of chlorophyll a in the Elbe River using a Lagrangian-based multi-criterion analog model. Ecol Modell. 2013;250:279-286. DOI: 10.1016/j.ecolmodel.2012.11.018.

[18] Mulia Iyan E, Tay H, Roopsekhar K, Tkalich P. Hybrid ANN-GA model for predicting turbidity and chlorophyll-a concentrations. J Hydro-Environ Res. 2013;7(4):279-299. DOI: 10.1016/j.jher.2013.04.003.

[19] Qiuhua I, Lihai S, Tingjing G. Use of principal component scores in multiple linear regression models for simulation of chlorophyll-a and phytoplankton abundance at a karst deep reservoir, southwest of China. Acta Ecol Sin. 2014;34(1):72-78. DOI: 10.1016/j.chnaes.2013.11.009.

MODELOWANIE ZAWARTOŚCI CHLOROFILU-A W WODACH PŁYNĄCYCH

AGH Akademia Górniczo-Hutnicza w Krakowie, Katedra Kształtowania i Ochrony Środowiska

Abstrakt: Chlorofil-a jest jednym z kluczowych parametrów służących do oceny stanu troficznego wód. W Polsce w ramach standardowego monitoringu rzek jest jednak badany rzadko. Artykuł skupia się na skonstruowaniu modelu predykcji zawartości chlorofilu-a w oparciu o dane pochodzące z rutynowego monitoringu realizowanego przez Wojewódzkie Inspektoraty Ochrony Środowiska. W tym celu na podstawie parametrów jakości wód, takich jak pH, nasycenie tlenem oraz różne formy azotu i fosforu, został sformułowany model regresyjny, a następnie przeprowadzono test zasadności zastosowania w nim poszczególnych składników równania regresji. Ostatnim etapem było opracowanie algorytmu pozwalającego na szybkie dostosowywanie modelu do lokalnych warunków w rzekach w celu dokonania wiarygodnej prognozy zawartości chlorofilu.

Słowa kluczowe: eutrofizacja, chlorofil-a, wody płynące, modele predykcyjne