Łukasz MAKOWSKI[1]

Warsaw University of Technology (1)

Distributed measurement system with applied Bayesian method to estimate cyanobacteria levels in Dobczyckie lake

Abstract. *Distributed measurement systems are more common in protection of strategic natural resources such as clean water. Bayesian methods can be applied to algorithms of such systems to improve quality of immediate situation assessment and hazard notification. In this paper a practical example of such approach is discussed.*

Streszczenie. *Rozproszone systemy pomiarowe zdobywają popularność w ochronie strategicznych zasobów naturalnych takich jak woda pitna. Metody bayesowskie mogą zostać użyte w algorytmach takich systemów, w celu poprawy jakości natychmiastowej oceny sytuacji i powiadamiania o zagrożeniach. W tym artykule omówiony jest praktyczny przykład realizacji takiego podejścia.* **(Rozproszony system pomiarowy wykorzystujący metody bayesowskie do estymacji liczebności sinic w jeziorze Dobczyckim)**

Keywords: distributed measurement systems, signal prediction algorithms, Bayesian methods
Słowa kluczowe: rozproszone systemy pomiarowe, algorytmy predykcji sygnału, metody bayesowskie

Introduction

Goal of presented research is to provide an appropriate model for computation of estimated map chart, and probability distribution of cyanobacteria levels in Dobczyckie lake. In biochemical processes these unicellular organisms create toxins as a waste product, which might be hazardous to ecosystem. First step to solve this problem could be a measurement system working directly on water reservoir [1].

Some limitations of current measurement system

Cyanobacteria biology is researched for a long time and it is known that their quantity is related to level of total phosphorus (TP) and total nitrogen (TN) dissolved in water [2]. Measurements of these compound groups are easier *in situ* than counting cyanobacteria and may lead to earlier estimation of their number and faster call for alert of a colony bloom.

Data fusion with Levenberg-Marquardt Algorithm (LMA) was successfully used to estimate level of cyanobacteria in Dobczyckie lake [3]. However, explanation of difference between computed and real values, which was considerably high at some points, was considered to be insufficient. Furthermore simple curve fitting does not yield prediction outside sampled data window and is heavily parametrised.

Therefore there is a need for more probabilistic approach to provide better information about confidence level over calculated estimations.

Bayesian approach

Cyanobacteria level that is considered to be unknown variable and other, measurable water parameters are continuous functions which may be observed at times $t = 1 \ldots T$.

Modelling starts from simple autoregression of order 1, which has the form

$$(1) \qquad y_t = \mu + \rho_1 y_{t-1} + u_t \qquad t = 2, \ldots T$$

Predicted, discrete cyanobacteria level is estimated as y_t and y_{t-1} is one-step lagged observation. Autocorrelation between successive values is given by ρ. Inevitable signal error is denoted as u_t. As the environment limits values of y_t, signal is centred and the level of outcome μ is equal to zero, which simplifies equation.

Having operator B that denotes earlier samples of signal y_t, general equation for order p autoregression is

$$(2) \qquad y_t - \rho_1 B^1 y_t - \rho_2 B^2 y_t \ldots - \rho_p B^p y_t = u_t$$

which can be further simplified to

$$(3) \qquad \rho(B) y_t = u_t$$

However, values that represent actual water state must be taken under consideration. Except of TP and TN also dissolved oxygen (DO), acidity (pH) and turbidity (FTU) have high correlation with cyanobacteria quantity level. Therefore problem changes to search for a multivariate model.

Lets denote variables having impact on cyanobacteria as a vector $Y_t = (y_{1t}, y_{2t}, \ldots y_{Kt})'$, and provide coefficients matrices Φ of size $K \times K$, and signal errors $U_t \sim N_K(0, V)$ where V is a covariance matrix. Then we can rewrite above equations with multivariate autoregression

$$(4) \qquad Y_t - \Phi_{p1} Y_{t-1} - \Phi_{p2} Y_{t-2} \ldots - \Phi_{pp} Y_{t-p} = U_t$$

Applying presented approach to normal distribution we can calculate conditional likelihood between cyanobacteria level and measurable water parameters with equation

(5)

$$f(Y_2|Y_1, \rho, \frac{1}{\sigma^2}) \propto \frac{1}{\sigma^{(n-p)}} exp \left(-\frac{1}{2\sigma^2} \sum_{t=p+1}^{n} |\rho(B) Y_t|^2 \right)$$

Conclusion

To gain better accuracy of proposed model we need to calculate equations with higher polynomial levels. Therefore there is a need for more frequent and consecutive gathering of water data samples, which can be achieved only with monitoring station continually working on reservoir, preferably performing measurements and calculations in real-time.

This work has been supported by the European Union in the framework of European Social Fund through the Warsaw University of Technology Development Programme.

BIBLIOGRAPHY

[1] Dziadak B., Kalicki A., Staroszczyk Z., Makowski Ł., Michalski A., Framework architecture of a large scale distributed measurement system for environmental protection. Institute of Sustainable Technologies NRIP, ch. 5, pp. 235 - 263, 2010
[2] Smith V.H. Havens K., East T., James T., N:P ratios, light limitation, and cyanobacterial dominance in a subtropical lake impacted by non-point source nutrient pollution. Environmental Pollution, 122, pp. 379 - 390, 2003
[3] Bogdan Dziadak, Andrzej Kalicki, Łukasz Makowski, Andrzej Michalski, Zbigniew Staroszczyk Wykorzystanie fuzji danych do estymacji liczebności sinic w jeziorze Dobczyckim. Przegląd Elektrotechniczny, 2011 (in proofreading)

Authors: Ph.D. Łukasz Makowski, Institute of Theory of Electrical Engineering Measurement and Information Systems, Faculty of Electrical Engineering, Warsaw University of Technology, ul. Koszykowa 75, 00-662 Warszawa, Poland, email: lukasz.makowski@ee.pw.edu.pl