SOFIA UNIVERSITY
ST. KLIMENT OHRIDSKI

APPLICATION OF BIG DATA FOR IMPROVEMENT OF THE ASSESSMENT OF THE SELF-PURIFICATION POTENTIAL IN WATER AND SEDIMENTS OF CASCADE FROM SHPP SREDEN ISKAR

INTERNATIONAL SCIENTIFIC CONFERENCE "KLIMENT'S DAYS" 2020

I. Yotinov[1]*, S. Tsolova[3], Y. Todorova[1], E. Daskalova[1], S. Lincheva[2], Y. Topalova[1]

1 – Department of General and Applied Hydrobiology, Faculty of Biology, Sofia University "St. Kliment Ohridski", Sofia; 2 – Sofia waste treatment plant; 3 – Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski"
**Corresponding author: yotinov@yahoo.com*

One of the professional environmental challenges of today is the assessment, control and biomanagement of river ecosystems used for renewable energy production. A similar ecosystem is the Iskar River in the area of the Cascade of 5 functioning HPPs "Sreden Iskar". Various problems are intertwined in this technological ecosystem, which require an interdisciplinary solution for control and management of the risk associated with the entry of pollutants into water and sediments in this area. Big data development offers realistic opportunities for more efficient operation, assessment and management of such complex technological ecosystems. Big data refers to all the data we have and we transform it into knowledge that we can use directly for management. A very important element is the right framework for the use of data, analysis and solutions, which is the basis of optimized efficiency.

The aim of this study is to use the accumulation of large databases to improve the assessment of the self-purification potential in sediments and waters of the Sreden Iskar cascade, with microbiological and hydrochemical parameters over a long period of time. From this point of view, the assessment of this potential and the forecasts for its impact from the operation of small hydropower plants is essential for preserving the ecological integrity of the region.

The use of big data in water management and control is a relatively new concept. Improvements in computerized analysis allow the large amount of data that is generated to be transformed so that problems can be identified earlier. Ultimately, big data applied in an ecological direction could increase the productivity and sustainability of technological aquatic ecosystems.

Materials and Methods

Fig.1 Study area and location of sampling site

To analyze the river Iskar the following for 2012, 2013, 2014, 2015 (High water – June; Low water – September):

- hydrochemical parameters - chemical oxygen demand (COD), Dissolved Suspended Solids (DSS), total suspended solids (TSS), Unsoluble Suspended Solids(USS), NH_4^+, NO_2^-, NO_3^-, PO_4^{3-}.
- microbiological parameters - quantity of aerobic heterotrophic bacteria (AeH) cultivated on meat peptone agar (Scharlau, Brit. Phar.) for 1 day at 28^{0}C and quantity of endo-bacteria (Endo) cultivated on Endo medium (Scharlau, Brit. Phar.) for 1 day at 37^{0}C.

St. Prokopanik

St. Lakatnik

St. Gabrovnitsa

Data from microbiological and hydrochemical parameters were processed with IBM Cognos Analytics.

Causality models

Granger himself defines the simple causal two-variable model as follows (Granger, 1969: 431):

$$X_t = \sum_{j=1}^{m} a_j X_{t-j} + \sum_{j=1}^{m} b_j Y_{t-j} + \varepsilon_t$$

$$Y_t = \sum_{j=1}^{m} c_j X_{t-j} + \sum_{j=1}^{m} d_j Y_{t-j} + \eta_t$$

The definition of causality given by Grander "implies that Y_t is causing X_t provided some b_j is not zero. Similarly X_t is causing Y_t if some c_j is not zero. If both of these events occur, there is said to be a feedback relationship between X_t and Y_t." (Granger, 1969: 431)

Granger sets the following requirements (Granger, 1969: 431):

1. X_t and Y_t must be stationary time series with zero means.
2. ε_t and η_t must be two uncorrelated white-noise series.
3. m must be finite and shorter than the given time series.

High water

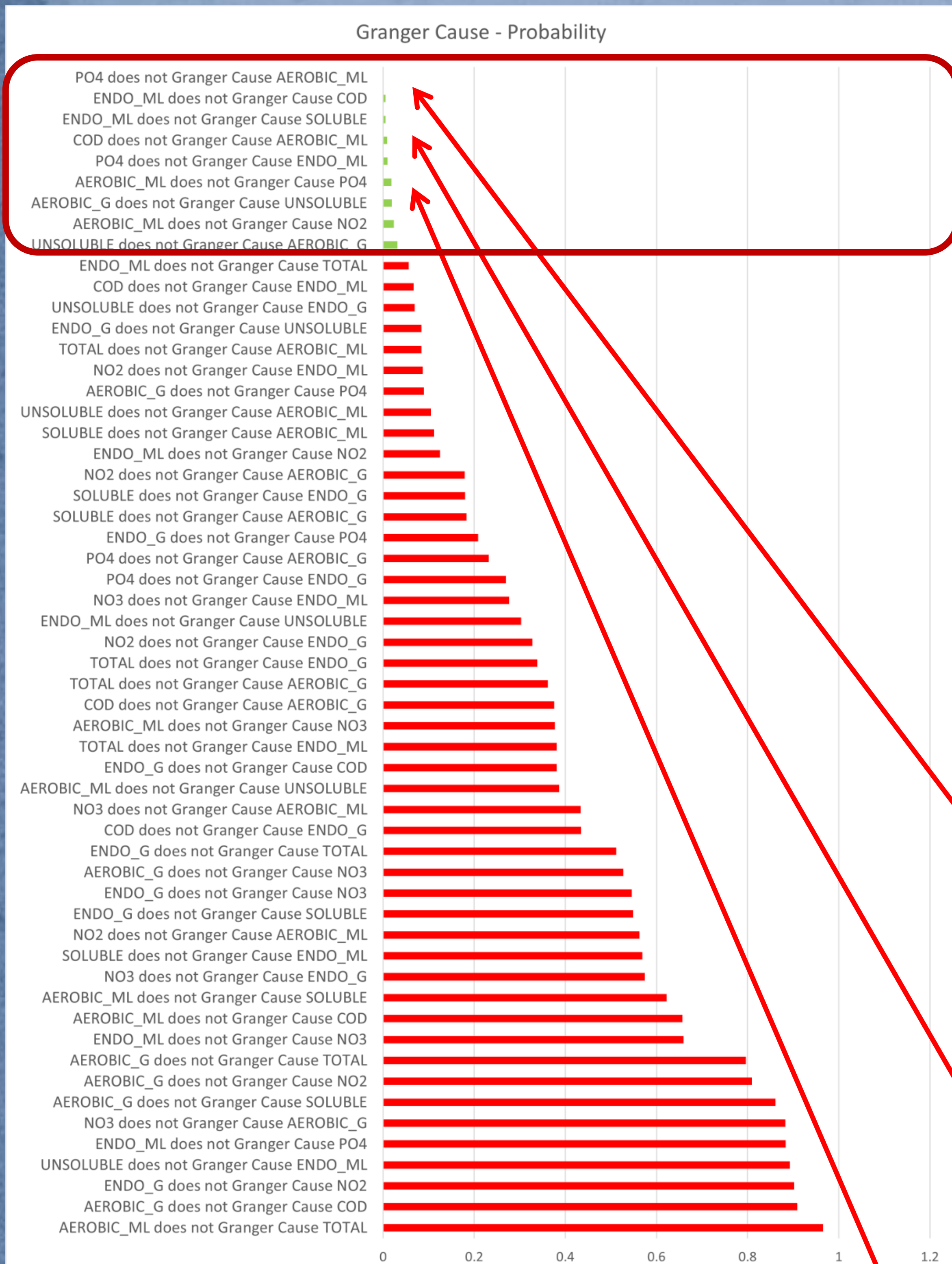

Fig. 2 Granger causality test for high water in river Iskar

Table 1. Panel unit-root test statistics

Variables	Deterministic	Level form	
		Statistics[a]	Probability
total suspended solids, mg/l	Individual intercept	-2.92 (0)	0.002
dissolved suspended soils, mg/l	Individual intercept	-7.44 (0)	0.000
unsoluble suspended solids, mg/l	Individual intercept	-3.16 (0)	0.001
NH4, mg/l	Individual intercept	2.58 (0)	0.995
NO2, mg/l	Individual intercept	-2.85 (0)	0.002
NO3, mg/l	Individual intercept	-2.43 (0)	0.008
PO4, mg/l	Individual intercept	-2.57 (0)	0.005
COD, mgO2/l	Individual intercept	-69.32 (0)	0.000
Aerobic heterotrophic bacteria, CFU/g	Individual intercept	-4.89 (0)	0.000
Endo bacteria, CFU/g	Individual intercept	-3.72 (0)	0.000
Aerobic heterotrophic bacteria, CFU/ml	Individual intercept	-4.11 (0)	0.000
Endo bacteria, CFU/ml	Individual intercept	-2.61 (0)	0.005

Only the time series of NH4 is non-stationary. All other time series are stationary.

Panel cointegration test

The time series are too short for the panel cointegration test. However only the time series of NH4 is non-stationary therefore we will exclude this time series. Since all other time series are stationary, we will use Granger causality test directly.

The analyzes show clear relationships between the amount of aerobic heterotrophs and the content of organic matter (COD), which means that during the study period we have induction of the heterotrophic group of bacteria from the high content of organic matter in the water and sediments. From these dependencies it is concluded that the self-purification processes in the Iskar River are intensive. Aerobic heterotrophs confirm their importance as an indicator group of bacteria and are an important part of the indicative apparatus assessing the potential of the aquatic ecosystem to self-purify. From the relationship between endo-bacteria and COD, it is clear that these bacteria have a positive effect on the mineralization of organic pollutants.

From the obtained data in fig. 2, a double relationship is established between aerobic heterotrophs and the concentration of phosphate ions in the waters of the Iskar River. This relationship is fully expected given that high nutrient loads correspond to the increased number of heterotrophic bacteria. These relationships can be used as a very good indicator of intensive self-purification processes in the waters of the Iskar River, with aerobic heterotrophs playing a very important role in them. The relationship between aerobic heterotrophic bacteria and unsoluble suspended solids confirms that bacteria (in this case aerobic heterotrophs) can agglomerate on undissolved particles, which can also use these particles as a nutrient substrate.

An important relationship is established between Endo bacteria and dissolved suspended solids in the waters of the Iskar River. There is a relationship between the increased content of endo bacteria caused by contaminants, which make up a significant part of the dissolved suspended solids. The established long-term relationships between endo bacteria and dissolved suspended solids prove that these bacteria play a significant role in the mineralization of organic matter. This is evidenced by the long-term relationship of the same bacteria with the presence of phosphate ions, which confirms the assumption that these bacteria have a supporting role in the mineralization of organic matter.

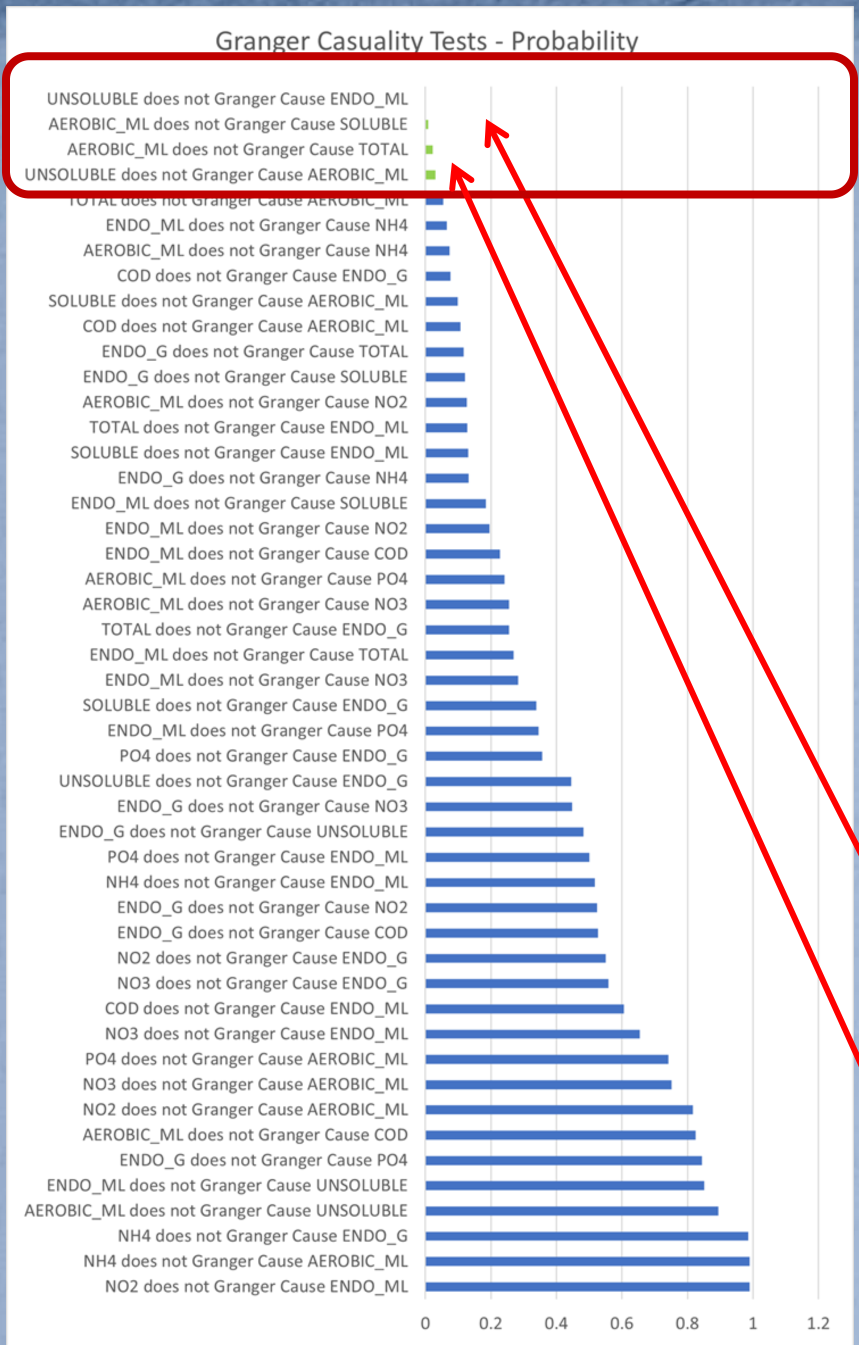

Fig. 2 Granger causality test for low water in river Iskar

Low water

Table 2. Panel unit-root test statistics

Variables	Deterministic	Level form	
		Statistics[a]	Probability
total suspended solids, mg/l	Individual intercept	-2.58 (0)	0.005
dissolved suspended solids, mg/l	Individual intercept	-2.26 (0)	0.012
unsoluble suspended solids, mg/l	Individual intercept	-3.34 (0)	0.000
NH4, mg/l	Individual intercept	-28.45 (0)	0.000
NO2, mg/l	Individual intercept	-6.30 (0)	0.000
NO3, mg/l	Individual intercept	-3.98 (0)	0.000
PO4, mg/l	Individual intercept	-8.75 (0)	0.000
COD, mgO2/l	Individual intercept	-13.81 (0)	0.000
Aerobic heterotrophic bacteria, CFU/g	Individual intercept	-1.20 (0)	0.116
Endo bacteria, CFU/g	Individual intercept	-5.30 (0)	0.000
Aerobic heterotrophic bacteria, CFU/ml	Individual intercept	-31.92 (0)	0.000
Endo bacteria, CFU/ml	Individual intercept	-4.66 (0)	0.000

Only the time series of Aerobic heterotrophic bacteria, CFU/g is non-stationary. All other time series are stationary.

Panel cointegration test

The time series are too short for the panel cointegration test. However only the time series of aerobic heterotrophic bacteria, CFU/g is non-stationary therefore we will exclude this time series. Since all other time series are stationary, we will use Granger causality test directly.

- The analysis establishes an interesting relationship between aerobic heterotrophs and the content of dissolved suspended solids in the waters of the Iskar River. This dependence is caused by pollutants, which make up a significant proportion of dissolved suspended solids. The established relationships between endo-bacteria and solutes in low water prove once again that endo-bacteria also play a significant role in the mineralization of organic matter.
- An interesting relationship is that between aerobic heterotrophs and unsoluble suspended solids, which is also found during high water. By providing a substrate for the attachment and feeding of bacteria, high amounts of unsoluble suspended solids, which may be organic or inorganic in nature, may be associated with high concentrations of bacteria.

Conclusions

The analysis of large data for the Iskar River in the area of the “Middle Iskar” Cascade revealed the dependences with the highest degree of binding between AeH-phosphate ions, AeH-COD and Endo-COD during the high water season, while during the low water season such are Endo-USS, AeH-SSD. The established relationships can be developed with an even larger volume of data and be a key element in the construction of an indicative apparatus for improve assessing the self-purification potential in water and sediments. The application of big data in the management of water technological ecosystems would help to identify environmental problems earlier.

Acknowledgment: *This document was supported by the Grant № BG05M2OP001-1.002-0019: "Clean Technologies for Sustainable Environment - Waters, Waste, Energy for a Circular Economy", financed by the Science and Education for Smart Growth Operational Program (2014-2020) and co-financed by the EU through the ESIF by means financing of labor. This work has been carried out in the framework of the National Science Program "Environmental Protection and Reduction of Risks of Adverse Events and Natural Disasters", approved by the Resolution of the Council of Ministers № 577/17.08.2018 and supported by the Ministry of Education and Science (MES) of Bulgaria (Agreement № Д01-322/18.12.2019).*