Kardos Máté Krisztián

Increasing the efficiency of a surface water quality monitoring system using statistical models

Ph.D. thesis booklet

Advisor: Adrienne Clement, Ph.D., Associate Professor

Budapest University of Technology and Economics

Faculty of Civil Engineering,

Vásárhelyi Doctoral School of Civil Engineering and Earth Sciences

Budapest, the 30th April 2020.

The current situation of water quality monitoring and status evaluation

The operation of water quality monitoring systems is a complex task, ranging from the designation of sites, measurement dates / frequencies and components to be measured, through the laboratory tests to the verification, evaluation, storage, and publication of results along with the allocation of the necessary human and material resources. Monitoring systems are also complex in terms of function, as it is usually the task of the one and same system to assess the state of the water network for a period and to detect long-term changes (trends), but also the short-term outliers (accidental pollution). Besides these, monitoring systems provide the information needed to understand the processes taking place in the river basins. The condition for the creation of water quality data is the sometimes costly sampling and laboratory tests: running the surface water quality monitoring system costs HUF 2–10 billion per year.

While the history of water quality measurements is nearly a century old, there have been significant developments in four related areas in the recent decades: (i) availability of water quality influencing geographical information data have increased (e.g., remote sensing-based topographic and land cover maps); (ii) computer systems for managing data (GIS software) have developed; (iii) statistical methods capable of unveiling relationships between data have developed and become easy-to-apply due to user-friendly programming techniques; and (iv) besides the human needs, ecosystems are gaining raising attention, which fact is mapped in specific laws (see e.g. European Union's Water Framework Directive). All this makes it possible and necessary to examine and reevaluate our monitoring systems – with an optimization point of view.

Throughout the work, I focused on small watercourses in Hungary (watershed < 1000 km²), as this is the type of water for which the amount of water quality data has multiplied in the last decades, and that makes it possible to establish statistical relationships. My primary goal was to develop the physico-chemical status evaluation system, besides which – like a "byproduct" – I investigated the efficient design of monitoring systems. I conducted my studies with the application of the surface water quality database, the related, already existing geographical information data applying state of the art data mining and statistical modeling techniques.

Summary of the Thesis

As a first step, I collected the most up-to-date and accurate data available on water quality, the water network, and the river basins. These include water network spatial data sets, water flow data, topographic and land use maps, and point source data. I thoroughly evaluated the water quality databases; and - where necessary – I filtered out or corrected erroneous data. I organized the collected data into a common database, which meant, on the one hand, assigning the location, and, on the other hand, accessing them from the same platform. This work in described in chapters 2.1 ("Databases used") and 2.2. ("Methods used").

Chapter 3 of the dissertation is about the relationships between river basin characteristics and water quality. Using two types of statistical models (binary logistic regression and linear discriminant analysis), I examine whether the achievement of the target status ("achieved" / "not achieved") and the rating of the water body ("high" / "good" / "moderate") / "poor" / "bad") can be predicted as a function of the terrain, land use and point source load characteristics of the river basin. For the study, I separate hilly and lowland watercourses. I found that models perform better on hilly watersheds than on lowland ones; that the achievement of the good status can be predicted with an accuracy above 90% and the status class (on the five-class scale) can be predicted with an accuracy of ~65% for highland and ~50% for lowland rivers. For both high- and lowland types, the ratio of forests and naturally covered areas and the amount of total phosphorus load of the point emissions relative to the long term mean water flow of the receiving watercourse have the most substantial influence. With the help of the mentioned statistical models, (i) I give an estimate of the status of the water bodies, (ii) give a plan for allocating the monitoring resources to dubious sites and (iii) estimate the water courses' wastewater load capacity.

The basic idea of chapter 4 ("Examination of the extent of similarity between sampling sites") is that water quality data can not only be interpreted assigned to one measurement site, but it has sense to group similar sites and to interpret associated measurement results together. With this method, groups supplied with much / scarce information can be distinguished, enabling to increase the efficiency of the monitoring system further. I characterize the measurement sites on the one hand with the water quality measurement results and, on the other hand, with the already mentioned characteristics of the associated catchment as background variables. The grouping is performed based on the water quality measurement data by the hierarchical cluster analysis method; subsequently, the reproducibility of the formed groups based on background variables is examined by linear discriminant analysis. I find that there is a higher degree of similarity between unpolluted watercourses, so relatively less monitoring is required to assess their status. Heavily polluted water bodies are characterized by a sufficient, relatively large number of measurements. However, moderately polluted waters are underrepresented in the current monitoring programs.

In Chapter 5 of my dissertation, I examine the extent to which the transition from weekly to fortnightly sampling to less frequent (monthly, quarterly) sampling reduces the reliability of the classification for various statistical indicators and water quality variables using the artificial random sampling method (Monte Carlo analysis). I find that both the median and the mode can be applied with a higher hit rate than the average for the very low sample numbers I examined and with the existing thresholds. The reliability of the classification is generally higher for variables such as pH, electric conductivity, dissolved oxygen, and total organic carbon; biochemical oxygen demand, ammonium, and phosphorus forms, on the other hand, weaken the reliability. Based on the above, I propose a revision of the status assessment system.

Chapter 6 of the dissertation contains a summary discussion of the results of the studies. Chapter 7 summarizes the results in a thesis-like manner.

The studies presented in the dissertation; their input data, goals, and results, along with the relationships among them are presented in Fig. 1.

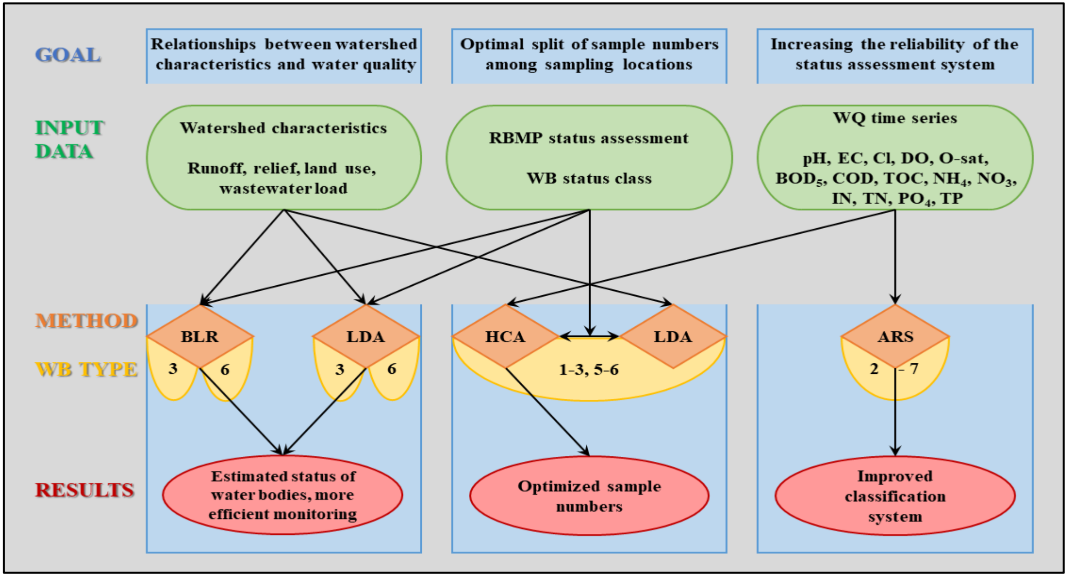

Fig. 1. *Studies presented in the Thesis. RBMP: River Basin Management Plan; WB: Water body; WQ: Water Quality; VARIABLES: BOD_5: Biochemical Oxygen Demand; COD: Chemical Oxygen Demand; EC: Electric conductivity; DO: Dissolved oxygen; IN: Inorganic Nitrogen; TN: Total Nitrogen; TOC: Total Organic Carbon; TP: Total Phosphorus; METHODS: ARS: Artificial Random Sampling; BLR: Binary Logistic Regression; HCA: Hierarchical Cluster Analysis, LDA: Linear Discriminant Analysis. WB TYPES: 1-4 hilly; 5-7 flat; 1-3 & 5-6 small; 4 & 7 large watershed.*

New scientific results

Thesis 1 *Status estimation of water bodies*

Based on the examination of Hungarian hilly and lowland river basins, I develop methods based on linear discriminant analysis and binary logistic regression to assess the physicochemical status of small watercourses. Using the methods, taking into account the topographic, land use and point load characteristics of the river basin, as well as the results of water quality measurements in the region, the status assessment can be clarified: the need for intervention on water bodies is> 90%, the classification of water bodies on a five-point scale can be given with 50-60% accuracy.

Related papers: [3, 6-8]

Thesis 2 *Specification of maximum wastewater load values by watershed land use*

The following formulas can determine the sewage-origin phosphorus load capacity of watercourses with a catchment area of less than 1000 km².

On hilly areas:

$$\frac{P}{Q} \leq \max\left(0\,,-0{,}1+\frac{E}{5}\right)$$

On flatlands:

$$\frac{P}{Q} \leq \max(0\,,-0{,}2+\frac{2E}{3})$$

Where

P is the phosphorus load capacity (g/s) of the water course per long term mean discharge (m^3 / s).

Q is the long term mean discharge of the water course at the outflow point (m^3/s),

E is the share of forests and naturally vegetated areas in the basin (km^2 / km^2).

Related papers: [1, 3]

Thesis 3 *Improving the allocation of locations based on similarities*

Using the method of hierarchical cluster analysis, I examined the degree of similarity between small watercourses based on their water quality measurement data. I found that there was a higher similarity between unloaded watercourses than between those loaded with pollutants. It follows that the diversity of water quality – and thus, the importance of monitoring, the necessity to densify measurements in space and time – is increasing in the presence of anthropogenic impacts.

Related papers: [2, 7, 10]

Thesis 4 *Improving the efficiency of the Hungarian status assessment-oriented monitoring system*

The Hungarian water quality monitoring system can be optimized with the integration of the national topographic, land use and point sources databases, historical water quality data sets, and modern statistical methods.

The method can be used to prioritize type 3 and 6 (small hilly and lowland) Hungarian river water bodies based on the extent to which on-site measurements are needed to assess their status: ca. 50% of hilly water bodies and ca. 25% of lowland water bodies -can be classified with high safety without on-site measurements (Fig. 2). In order to increase the reliability of the status evaluation, it is recommended to modify the site allocation, the time-frequency, and the inclusion of water quality variables.

Related papers: [3, 4–5, 7, 9]

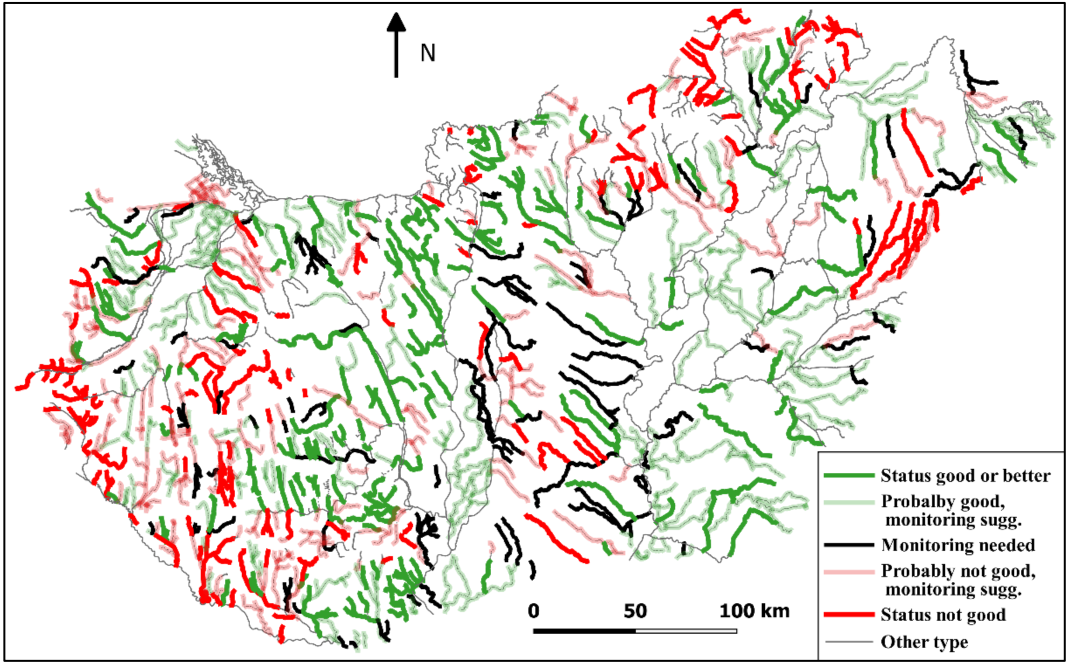

Fig. 2. *Model-estimated status of type 3 and 6 small water bodies, and the need for on-site sampling.*

Theses-related papers

Peer-reviewed journal articles

[1] Clement Adrienne, Jolánkai Zsolt & Kardos Máté Krisztián: Results of the river basin management planning related to municipal water management: the role of municipal wastewater treatment in the development of surface water quality and the planned measures. (In Hungarian: A vízgyűjtő-gazdálkodási tervezés települési vízgazdálkodással kapcsolatos eredményei: a kommunális szennyvíztisztítás szerepe a felszíni vízminőség alakulásában és a tervezett intézkedések) *MASZESZ Hírcsatorna*. 2015. (5) 6–16.

[2] Kardos Máté Krisztián & Clement Adrienne: Similarities among small water courses based on multiparameter physico-chemical measurements. *Central European Geology*. Accepted for publication.

[3] Kardos Máté Krisztián & Clement Adrienne: Prediciting Small water courses' Physico-chemical status from watershed characteristics with two multivariate statistical methods. *Open Geosciences*. 2020. (12) 71–84. DOI:10.1515/geo-2020-0006.

Studies published in volumes of studies

[4] Clement Adrienne, Kardos Máté Krisztián & Szilágyi Ferenc: Classification of surface waters according to the physico-chemical characteristics supporting ecology - lessons learnt from the status assessment for the planning of action programs. (In Hungarian: Felszíni vizek minősítése az ökológiát támogató fiziko-kémiai

jellemzők szerint - az állapotértékelés tanulságai az intézkedési programok tervezése szempontjából). In: *Lajos Szlávik, Tamás Gampel & Edit Szigeti (ed.): Papers of the 33rd National Assembly organized by the Hungarian Hydrological Society. Szombathely, July 1-3, 2015*. pp. 1-11.

[5] Kardos Máté Krisztián: Physico-chemical classification of surface water bodies based on sparse samples according to the Water Framework Directive. (In Hungarian: Víz Keretirányelv szerinti fiziko-kémiai minősítés alacsony mintaszám esetén.). In: *Jakab Gusztáv & Csengeri Erzsébet (ed.): XXI. century water management at the intersection of sciences: II. International Conference on Water Science. 22nd March, 2019, Szarvas, Hungary ISBN: 978-963-269-808-3*. pp. 98–110.

[6] Kardos Máté Krisztián & Clement Adrienne: Indication of the impact of land use on water quality with pattern recognition algorithms. (In Hungarian: A területhasználat vízminőségre gyakorolt hatásának indikációja mintázatfelismerő algoritmusokkal). In: *Fazekas István & Lázár István (ed.): Functioning and image of landscapes MTA DTB Földtudományi Szakbizottság, Debrecen, 2019. ISBN: 978-963-7064-39-5*. pp. 197-202.

Abstracts (not to be considered as scientific papers)

[7] Kardos Máté Krisztián & Clement Adrienne: Spatiotemporal optimization of monitoring networks with respect to water body classification. In: *Kronvang, Brian; Fraters, Dico &*

Kovar, Karel (editors): International Interdisciplinary Conference on Land Use and Water Quality, Agriculture and the Environment, Aarhus, Denmark, 2019. 06. 02-05. p. 148.

[8] Kardos Máté Krisztián & Clement Adrienne: Indication of the impact of land use on water quality with pattern recognition algorithms. (In Hungarian: A területhasználat vízminőségre gyakorolt hatásának indikációja mintázatfelismerő algoritmusokkal.) In: *Fazekas István & Lázár István (ed..): Abstracts of the VIII. Hungarian Conference of Landscape Ecology, 29-31 August 2019, Kisvárda, Hungary. ISBN: 978-963-508-915-4.* p. 53.

Technical Reports (not to be considered as scientific papers)

[9] Krisztián Máté Kardos: Study required to comply with the Nitrates Directive - General physico-chemical assessment - summary. Technical Report. *BME VKKT*, May 2018. pp 1-11.

[10] Kardos Máté Krisztián: Development of an algorithm for grouping river network monitoring sites.Technical Report. *BME VKKT*, March 2019. pp 1-28.

Selected references

Behmel S., Damour M., Ludwig R., & Rodriguez M. J. 2016. Water quality monitoring strategies — A review and future perspectives. *Science of the Total Environment* 571:1312–29.

Clement A. & Szilágyi F. 2015. Physico-chemical status assessment system of surface water bodies. Background document no 6-2. to the National River Basin Management Plan In Hungarian: Felszíni víztestek fizikai kémiai állapotértékelési rendszere – Az OVGT 6-2. számú háttéranyaga. Országos Vízügyi Főigazgatóság, Budapest.

European Commission 2003. Common Implementation Strategy for the Water Framework Directive – Guidance Document No. 7: Monitoring under the Water Framework Directive. European Commission, Bruxelles.

Giri S. & Qiu Z. 2016. Understanding the relationship of land uses and water quality in twenty first century: A review. *Journal of Environmental Management* 173:41–48.

Hastie T., Tibshirani R. & Friedman J. 2009. The Elements of Statistical Learning – Data Mining, Inference, and Prediction. *Springer Series in Statistics*. Springer, Berlin.

Hatvani I. G. 2014. Application of state-of-the-art geomathematical methods in water protection - on the example of the data series of the Kis-Balaton Water Protection System. Doktori (PhD) értekezés. Eötvös Loránd Tudományegyetem Természettudományi Kar, Budapest.

Somlyódy L. (ed.) 2011. Water management in Hungary – State of the art and future strategic tasks. (In Hungarian: Magyarország vízgazdálkodása: helyzetkép és stratégiai feladatok.) Hungarian Academy of Sciences, Budapest.

Somlyódy L. 2018. Surface water quality – modeling and control. (In Hungarian: Felszíni vizek minősége – Modellezés és szabályozás). Typotex, Budapest.

Strobl R. O. & Robillard P. D. 2008. Network design for water quality monitoring of surface freshwaters: A review. *Journal of Environmental Management* 87(4):639–48.