www.czasopisma.pan.pl PAN www.journals.pan.pl

Archives of Environmental Protection
Vol. 44 no. 2 pp. 3–13

PL ISSN 2083-4772
DOI 10.24425/119688

DE GRUYTER OPEN

Dependence of water quality assessment on water sampling frequency – an example of Greater Poland rivers

Małgorzata Loga[1]*, Michał Jeliński[1], Niina Kotamäki[2]

[1]Warsaw University of Technology, Poland
Faculty of Building Services, Hydro and Environmental Engineering
[2]Finnish Environment Institute (SYKE), Finnland

*Corresponding author's e-mail: malgorzata.loga@is.pw.edu.pl

Keywords: water quality, sampling frequency, river monitoring.

Abstract: Under the EU Water Framework Directive, the status of a surface water body (*wb*) is determined from a water sampling program. The reliability of this determination partly depends on the frequency of such sampling and is the subject of this paper. Available water quality data were acquired from the national environmental monitoring system in Greater Poland (Wielkopolska). At any given monitoring station, an assumption of normally distributed data was checked for all water quality indicators (*wqi*) and relevant statistical parameters identified. For particular sampling frequencies, a Monte Carlo simulation model was used to generate synthetic wqi data series from which *wb* status could be estimated. Assuming that 24 annual measurements of all *wqi* is the maximum economically affordable sampling frequency and taking this frequency as a reference, it has been shown that in about 22% of cases, a water quality class assessed from sampling at the standard frequency of 12 times per year is inaccurate. In less than 50% of cases was the reference assessment better than classifications based on lower frequencies. Nevertheless, in 33% of cases the indicator class was correctly assessed from measurements taken 4 times annually. The correlation between water quality class assessment and sampling frequency is not simple nor can it be arbitrarily assumed. When choosing the sampling frequency as a basis for reliable water quality assessment it is necessary to take into account characteristics of natural and man-made pressures acting on rivers in their catchments.

Introduction

Implementation of the Water Framework Directive (WFD) has entailed considerable changes to the monitoring systems of the EU member states (EC 2000), (EC 2013). An obligation to assess the ecological and chemical status of all surface waters and groundwater has led to extensive monitoring of a large number of water quality indicators in all significant water bodies of the EU countries. Assessing the state of surface waters in the EU countries is at present realized by three types of WFD monitoring: i) surveillance monitoring for assessing the overall water body status, ii) operational monitoring required for water bodies that are at risk of failing to meet the WFD objectives, and iii) investigative monitoring – for finding the reasons for failing to meet 'good status'.

The assessment of the ecological status of a surface water body relies mostly on biological quality elements, although hydro-morphological and physico-chemical quality elements play a supportive role. Technically, the status class of an individual water body is assessed through the use of the mean values of water quality indicators measured in the water body within each of the six-year consecutive periods of the WFD monitoring cycle. The mean value of any water quality indicator (*wqi*) within a water body, being the estimate of the true mean of the indicator, is inevitably associated with some degree of uncertainty. This uncertainty can lead to erroneous classification of a given water body and, consequently, can cause irrelevant or wrong decisions concerning the water body.

Many researchers have reported that the uncertainty in the estimate of the mean value is strongly related to the frequency of the water sampling (Brauer et al. 2009, Reynolds et al. 2016, Loga 2016). Usually, they also ascertain that it is economic factors that constrain the collection of sufficient monitoring data. However, new emerging monitoring methods and in situ techniques make water monitoring more efficient and less constrained economically. An overview of current technological advances in real time monitoring using various in-situ techniques can be found in the literature (Blaen et al. 2016.) More water quality elements can now be measured in real time by using ion-selective electrodes, uv--vis and fluorescence spectroscopy, or colorimetry. However, periodic calibration (even once a week), regular cleaning (due to fouling) or replacement of reagents remain indispensable for obtaining reliable readings from the sensor instruments. Also when a monitoring location is far from a power source, reliability of electrical supply is very important as the development of water quality sensors powered by solar panels (Amruta et al. 2016) is not yet widespread in monitoring practice.

So far, monitoring using in-situ, on-line water quality measurements with high temporal resolution is only being developed in research projects. Such monitoring schemes are run mostly on a catchment scale and focus on improving an understanding of complex nutrient cycles rather than for routine monitoring for status assessment. High and low frequency monitoring based on sensor measurements undertaken during dry weather and/or high intensity rainfall made it possible to track the concentrations of various substances as they are flushed from the catchment and subsequently diluted in the river (Ivanovsky et al. 2016).

Also, easy and relatively cheap access to satellite or air--borne remote sensing images has increased our ability to analyze water quality related features having large spatial extent such as algal blooms in large lakes or marine oil spills. However, the imagery still has inadequate spatial resolution as far as monitoring of water quality in rivers is concerned. The most serious constraint for successful application of this new technology in water quality monitoring is that most natural water constituents relevant for water quality monitoring do not have contrasting properties (e.g. color) and are therefore difficult to detect with satellite sensors. Despite some obvious advantages of using satellite imagery, other factors such as cloudiness and atmospheric disturbances considerably reduce the attractiveness of satellite remote sensing (Kiefer et al. 2015).

Apart from these new emerging technologies, the traditional water samples from existing monitoring programs are still the basis of the classification of *wb* status. Designing robust monitoring programs and, in particular, setting annual frequencies for water sampling is a challenging issue. An extensive review of water quality monitoring strategies can be found in Behmel et al. (2016). The paper contains several "use cases" understood as a "sequence of actions to achieve a goal" whereby sampling frequencies and recurrence are discussed as the possible scientific requirements for a satisfactory monitoring design. There are also other requirements, challenges and constraints present, all playing an important role in planning, optimizing and managing a water quality monitoring program; monitoring objectives, technical means, financial and human resources will all influence each other. Moreover, Behmel et al. (2016) emphasize, "No holistic solution exists to cover all steps of water quality programs", so as to encourage regional water managers to work out methods for *constrained* optimization of monitoring networks.

An interesting study of status assessment by Skeffington et al. (2015), comparing the results of weekly and monthly sampling frequencies, not surprisingly, concludes that weekly data present a narrower 95% confidence interval in comparison to monthly data. A similar but smaller narrowing of the 95% confidence interval occurred when sampling time was restricted to working days only (from Monday to Friday) and for limited hours when compared to 7 days per week and unconstrained sampling hours. Restriction of sampling time to a 3 h window introduced in this study slightly improved the precision of estimates in comparison to the full working hours (from 9 h to 18 h) strategy, depending on the water quality parameter being measured.

Although many researchers and practitioners have been attempting to determine optimal sampling frequencies and to minimize inherent uncertainty, reviews of the literature suggest that there are no universally valid, nor sufficiently well-detailed guidelines for the design of water quality monitoring networks (Kovacs et al. 2012, Skarbovik et al. 2012, Canadian Council of Ministers of the Environment 2015). In essence, the task has proven to be challenging and difficult to resolve in one go. The need for a more coherent theory using quantitative measures of this uncertainty is acknowledged widely in the WFD-related literature (e.g. Hering et al. 2010, Borja et al. 2013, Reyjol et al. 2014). The quantification issue relates to errors in water sampling procedures, errors in chemical and biological analyses, estimates of uncertainty in *wqi* and their impact on the classification of river status are addressed in many research reports and papers devoted to monitoring of lakes and rivers (Clarke 2013, Carstensen and Lindegarth, 2016, Kotamäki et al. 2015).

An interesting analysis of various sources of uncertainty, based on a comparison of inter-laboratory data and internal quality control data, can be found in Guigues et al. (2016). This study showed that scale of contribution to sampling uncertainty was dependent on the properties of the water quality parameters.

The application of non-parametric tests as a tool for supporting sampling frequency optimization on an example of 13 rivers and 44 sampling stations located in the Campania region was reported by Naddeo et al. (2013). Vilmin et al. (2016) presented a case study of optimal sampling frequency for fast varying physico-chemical components in the Seine River. Optimality criterion used in this case study was an accurate estimation of six WFD-defined *wqi* in a large river influenced by significant agglomeration. The research confirmed the general opinion that water flow and pollution dynamics in rivers are dependent on the characteristics of their catchments, in terms of geological structure and land usage, efficiency of operating sewage treatment plants (Facchi et al. 2007), distribution of point and non-point sources of pollution and on meteorological conditions (Neal et al. 2012).

With regard to the above conclusion, and based on reports from several case studies, it is apparent that the task of designing monitoring programs, including both the location of monitoring points and the frequency of water sampling, have to be dealt with locally and adaptively, taking into account changes in the natural environment (e.g. increasing occurrence of hydrological extremes, climate warming and pressures from human occupation).

As the operational and maintenance costs of the water monitoring system directly depend on the number of measured water quality elements and their sampling frequencies, it is economically vital to find a compromise between the total costs of a monitoring system and the expected investments in the water sector whenever the assessed water body (*wb*) status is "not good" or "at risk" and obligatory remediation measures need to be implemented. This is of particular concern when the *wb* status classification "not good" is false, as the result of high uncertainty in the *wb* assessment (which in many cases results from low sampling frequency). Clearly, if a higher water sampling frequency had been used, a false assessment would have been less likely to happen and consequent investments in river remediation would not be necessary.

According to EEA (2016) there are 20 monitoring programs running for rivers in Europe. In most of these programs, samples are taken on an annual cycle with a sampling frequency ranging from 4 to 26 samples per year. This more than six-fold ratio in sampling frequency very likely originates from countries' economic constraints but has its consequences in the corresponding uncertainty in the status assessment of their river waters. As the economies of the EU countries are slowly converging (as are their ecological budgets), it may be imagined that in the long run the water sampling frequencies in river monitoring systems could be optimized in order to achieve similar credibility in water status assessment for rivers across the EU.

The key theoretical problem considered in this study is a relationship between water sampling frequency and the uncertainty of the water body status assessed from the measurements. We present in this paper the first steps towards rationalizing the water monitoring efforts in Poland and towards formulating guidelines for improving the country's surface water monitoring system when operating under the WFD. The analysis is based on the monitoring data of the State Environmental Monitoring System in the Greater Poland region.

Materials and methods

Monitoring data

This study utilizes the river water quality monitoring data collected within the Polish State Environmental Monitoring System in Greater Poland (Wielkopolska region) Fig. 1. This part of Greater Poland is located in the central-western Poland. It is subdivided into two macro regions – south Wielkopolska lowland and Wielkopolska lake district. There are 521 surface water bodies within the province among which 383 are rivers and 138 are lakes. The prevailing abiotic water body types are large lowland rivers, rivers with an organic substrate and lowland sandy clay and gravel rivers. The region is mostly lowland landscape with arable lands (74%), forest (26%) – mainly coniferous stands and many lakes (Urząd Marszałkowski 2017). A majority of the area (88%) is drained by the river Warta and its tributaries, all in the Odra river basin. There is also intensive cattle breeding in the region. As a consequence of pressures from agriculture and livestock, the prevailing problems of water quality are high concentrations of nitrogen and phosphorus compounds causing eutrophication of water ecosystems.

The Polish regulations concerning water status assessment (Dz.U. 2014, item. 1482) state that "when the class of any physico-chemical indicator in a water body is lower than the class resulting from an assessment based on biological elements, the overall ecological status of the *wb* needs to be lowered by one class", and "when any of the biological quality elements in a *wb* does not meet requirements of good status, there is no need to assess the physico-chemical status of that water body". Henceforth, physico-chemical quality elements are classified into three classes: *high*, *good* and *below good.*

Although WFD-oriented water monitoring programs were launched in Poland in 2006, location of measuring points, types of monitoring network and methods of water quality assessment in the country have been modified a number of times since then (GIOS 2017). There are over 4500 river water bodies in the country from which about 1800 are monitored within surveillance and operational networks (GIOS 2013). The status assessment of unmonitored water bodies is estimated on the basis of the status of the monitored ones using similarity in catchment landscapes, geological type and the type and intensity of pressures. Although the final report on water body status in Poland, required under the WFD at the end of 2015, has not yet been published, it can be partially estimated (Dz.U.2016, poz.1911, poz.1967, 2016): 76% of the assessed river water bodies and 67% of lake water bodies failed to reach good status (Soszka et al. 2016).

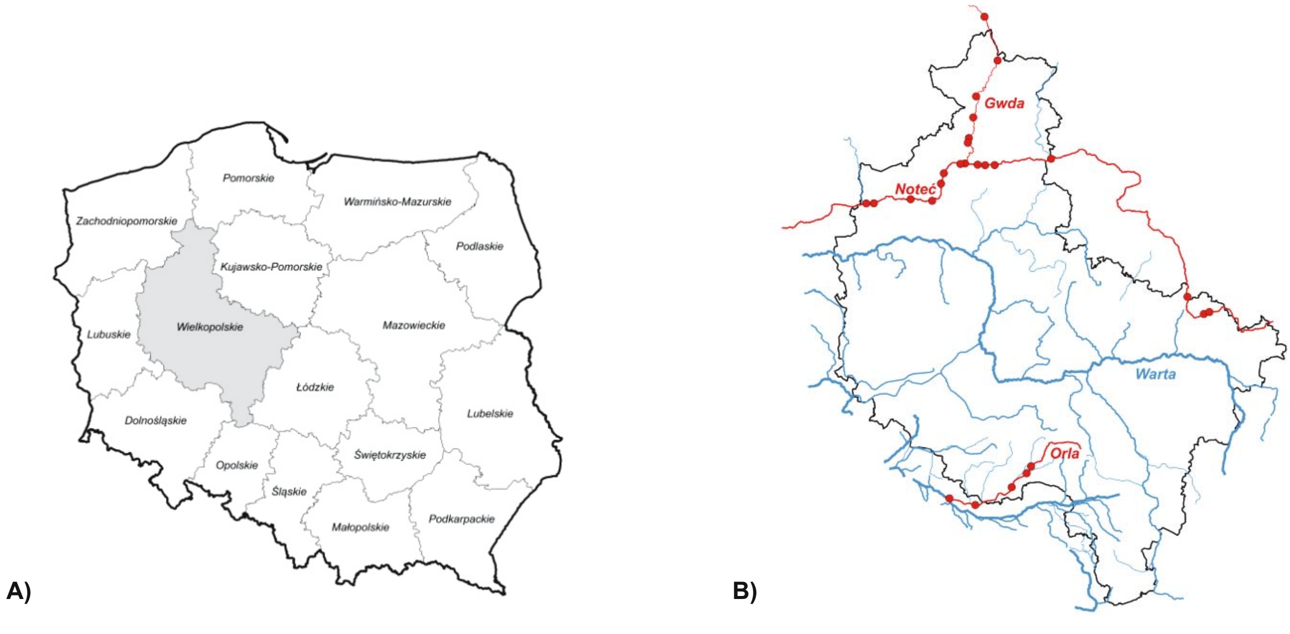


Fig. 1A) Location of Greater Poland (Wielkopolska) region on the map of Poland. **B)** River network in Wielkopolska Region. Three rivers – the Noteć, the Gwda and the Orla, marked in red, are referred to in the text

Changes observed in populations of biological elements, which are considered the most important for *wb* status assessment, represent the integrated reaction of river ecosystems to various pressures. As the changes are rather slow, it is not envisaged to undertake biological measurements with a frequency higher than once every two or three years. In that case, it will be necessary to measure biological indicators for several dozen years to infer any meaningful statistical distributions from the time series of biological indicators. Noting that biological quality elements were introduced in the Polish monitoring system routine in 2006, and given that up to now only two WFD water management cycles have been completed (1st 2003–2009, 2nd 2010–2015), it has been realized that biological indicators are rarely available more than twice for any water body. Therefore, in this work, the scope of analysis of how water quality assessment depends on sampling frequency is restricted to physico-chemical elements only and drawing on values of these indicators measured by the river monitoring system of Greater Poland.

Water quality data used in this study were collected in rivers of the region during the period 2006–2014 by the Voivodeship Inspectorate of Environmental Protection of Greater Poland as a part of its statutory activities within the State Environmental Monitoring system.

As the majority of monitoring stations in Poland serve several monitoring purposes within each water management cycle, data for different lists of water quality elements are being produced annually. Consequently, values for most of the water quality parameter are available almost every year. At present, long series of measurement data for physico-chemical and chemical elements, i.e. temperature, pH, indicators of oxygen conditions, salinity indicators, concentrations of nutrients, hardness, suspended solids, etc. – in total, time series for 37 indicators, are available for analysis and for estimation of the statistical distributions of these elements. All these water quality elements have been measured in water samples taken up to 12 times per year in rivers of Greater Poland.

Also, concentrations of so-called *specific pollutants*, such as arsenic, chromium, fluorides, and phenols have been measured within this regional water monitoring system. However, due to the scarcity of measurements made within each year (4 times) it was decided not to take these *specific pollutants* into consideration in further analysis.

From over 11 thousand available *data sets* representing measurements of physico-chemical water quality indicators at monitoring points in Greater Poland for the years 2006–2014, data from 16 rivers, consisting of 57 water bodies, have been ultimately chosen for further analysis. Hereafter, one *data set* is understood to be a set of all measurement outcomes of one physico-chemical element, within one year, at one monitoring point. For a water quality data set to be adopted for analysis, a minimum series length of at least 8 values per annum was the criterion.

Method adopted

It is assumed in this study that the indicators (water quality parameter concentrations) themselves are normally distributed random variables and hence the corresponding sample (series) means are also random variables normally distributed around the corresponding (true) mean values (D'Agostino 1986, Loga and Nawalany 2009). As the measure of an indicator uncertainty the mean standard deviation of the distribution of the sample mean is adopted. The variability of the sample mean is smaller than the standard deviation of the indicator itself by the factor of square root of *n,* where *n* is the length of the indicator series.

Using these assumptions it was possible to calculate one of the WFD required measures of uncertainty – *the confidence of the status class* for particular indicator. It has been estimated in literature as a half of the width of confidence interval of the indicator mean for the chosen level of significance. In this paper the issue of estimating confidence of the status class is not addressed – it can be found elsewhere (Loga 2012, 2016) .

The analysis carried out in this work started by testing the normality assumption for each physico-chemical indicator. Statistical parameters of normal distributions describing random variability of physico-chemical elements in water bodies have been estimated from the available measurement data. The main aspects of the tests used in this first phase are presented below.

Monitoring data for each of physico-chemical indicator measured within the same year in different water bodies of a given river were checked to determine whether they are normally distributed, using the Shapiro-Wilk test and then checked for equality of variance and mean value using the T.test (t-Student test) and F.test (Snedecor test). When the outcome of the normality test was negative, the empirical data were log-transformed and tested once again for normality. The data were then pooled together to secure a sufficient number of values in order to estimate the indicator probability distribution, provided conditions of variance and mean equality in the corresponding sets of data were fulfilled.

Similarly, the series of data originating from the same water body, which have been collected over several years, before pooling into one data set, were checked for their variance and mean equality using a T.test and F.test and for normality of distribution using Shapiro-Wilk test. In case the outcome of the normality test was negative, the data were transformed by a logarithmic function and tested once again for normality.

In the Polish monitoring system the number of water samples (and thus, measurements of the indicators) taken in water bodies during one year is only slightly greater than twelve, which still is too few for estimating parameters of a statistical distribution. Therefore, attempts have been made to pool data originating from different water bodies of the same river or from the same water body, but from consecutive years, into one data set. Whenever the equality of the means and the variances of two individual sets of a water quality element were confirmed, the sets have been pooled together. This consolidation allowed for calculating overall mean value and variance for a statistical distribution corresponding to this element. Pooled sets of measured data having at least 24 measured values were considered as sufficient for estimating parameters of statistical distributions.

In the second phase of the analysis, a number of simple Monte-Carlo (M-C) models have been employed to simulate a process of random sampling of *wqi* using selected annual frequencies. In each M-C model, a generator of pseudo--random numbers served to imitate the Gaussian behaviour of "measurement outcomes" of an indicator normally distributed around its *mean value* and with its variability equal to the *standard deviation*, both estimated from the indicator pooled data.

Specifically, for all indicators the data sets with 4, 6, 12 and 24 measurement values per annum have been M-C generated to simulate the assumed annual frequency of surface water sampling: once in every quarter, once in every two months, once a month and twice a month. For each indicator (and given water body/river) a number of M-C runs were executed. It turned out that carrying out 500 runs for each frequency of sampling was sufficient to obtain a stable estimate of the standard deviation. For each M-C--generated statistical sample of any indicator (a model of the series of the indicator's measured values), an estimate of the sample mean value was used as the key parameter for assigning a water quality class for the given indicator. In some cases, the sample means corresponding to generated data sets of different sizes (4, 6, 12 or 24) led to assigning different classes of water quality for the same given indicator. The procedure of M-C sampling with frequencies of 4, 6, 12 and 24, was repeated 500 times for each indicator. This resulted in a set of series of sufficient size to allow statistical inferences to be made with respect to the relationship between the frequency of sampling and the uncertainty of class status assessment, on the basis of "simulated measurements".

According to current Polish regulation (Dz.U. 2014, item. 1482), the classification of a water body with reference to a given physico-chemical indicator, is based on the mean value of the indicator measured within one year. However, in order to obtain a legally binding assessment, it is necessary to secure at least 4 values of measurement data per year. This justifies the use of the Monte Carlo generator, which can simulate as many "measurements" per annum as may be needed for the statistical inferences. There is, however, one particular problem with the M-C simulation, that of judging the 'correctness' of the classification of the indicator when the sample mean is estimated from 4, 6, 12 or 24 "measured" values. Because it is not possible to decide objectively upon the "true class" of water body for any *wqi*, it was decided to adopt a reference "base line" to allow analysis and comparison of classifications determined from the different sampling frequencies.

To resolve this problem, the classification based on 24 water samples per annum was assumed as a reference. This frequency (24 per annum) is the recommended annual sampling frequency for water stations on rivers and streams under the GEMS program (WMO 2013). This sampling frequency has also been recommended as potentially the maximum frequency which would be economically feasible to adopt by the Voivodship Inspectorates of Environmental Protection and their laboratories.

Therefore a water body classification based on the indicator's average value, derived from 24 measurements, has been assumed to represent its "true class" and to be the reference for classifications obtained when measuring with frequencies lower than 24 samples per annum.

Results

From the 7914 data sets which remained after screening, 39% of water quality indicators (*wqi*) were fitted by a normal distribution, over 13% were converted into normal distributions via a logarithmic transformation, and 8% of water quality parameters gave a negative result when tested for normality, despite logarithmic transformation. For 40% of water quality data sets, the time series were too short (less than 12 values) to allow estimation of their distribution functions. A summary of the tests for normality is presented in Table 1.

By using monitoring data from the rivers of Greater Poland, and classifying their waters into one of three classes (1 – *high*, 2 – *good* and 3 – *below good*), a percentage of particular *wqi* classes together with their dependence on sampling frequency has been estimated – see Table 2 for an example of the Noteć river. Differences are generally small between the "true" *wqi* class (i.e. class assessed from measurements made with the reference frequency 24 per annum) and classes determined from indicators sampled with smaller frequencies. In the case of the Noteć river, less than 3% of the water bodies were 'wrongly' classified when using indicator sampling frequencies of 4 times per annum compared to the reference frequency of 24 per annum.

Similarly, samples generated by the Monte-Carlo model for the Mała Wełna water body for two sampling frequencies, 4 and 24, show little difference in the determined class distribution (Fig. 2).

When comparing number of water quality elements determining a particular class for the cases of 4, 6, 12 and 24 measurements per year, the differences are typically smaller than 10%. The greatest difference between results from 4 measurements per year and 24 measurements per year was observed for Mała Wełna in 2011 (Fig. 2). However, even in this case, the difference in percentage of water quality indicators determining 2^{nd} class (good status) is only slightly greater than 13%.

An extreme case of the lack of difference in the assessed class arising from sampling frequency is presented in Table 3. It can be observed that for lower part of the Gwda river (0.3 km from the river mouth) there are no differences in classification for any of the sampling schemes. Generally, the occurrence of physico-chemical classes being 'wrongly' assessed based on a less frequent sampling scheme happens in less than 20% of cases, with the exception of one location (the Gwda 48.3 km). As the differences in the *wqi*-based classification corresponding to frequencies 4, 6 or 12 and 24 samples per year are very small in the case of the Gwda river, the water sampling frequency in that river can be reduced from the present 12 times to even 6 times per year without losing precision in the classification and assessment of the river status.

However, the assessment of water body status does not respond to a decrease in sampling frequency in the same way for all analyzed rivers.

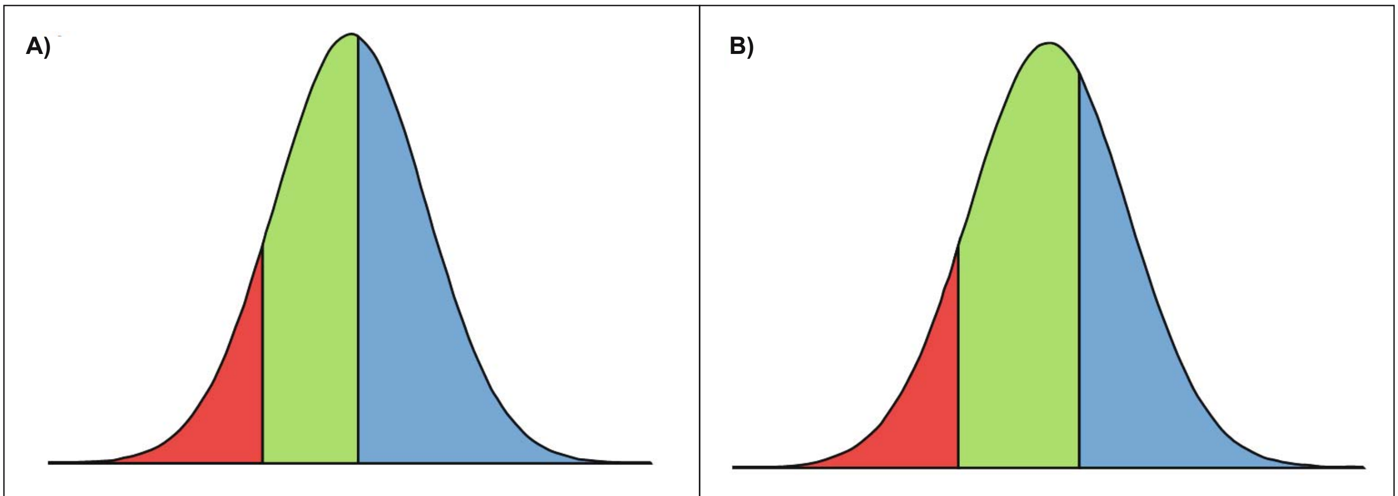
The example of the Orla river (Table 4) shows that in two locations, at 22.1 km and 52.6 km, even standard sampling

Table 1. Fraction of water quality indicators (*wqi*) passing the test for normality (based on measurement data from the monitoring stations in Greater Poland for the years 2006–2014)

	Number of monitoring stations	*wqi* passing the test for normality	*wqi* passing the test for normality after logarithmic transformation	*wqi* not passing the test for normality	*wqi* not tested for normality as the result of insufficient number of measurements	Total
All rivers in Greater Poland	127	**3116**	**1063**	**606**	**3133**	**7914**
		39%	13%	8%	40%	100%

Table 2. Percentage of water quality indices classified in one of three classes depending on sampling frequency – example of Noteć river

Location of monitoring points along the Noteć river		4 measurements per year, % of *wqi* indicating a given class			12 measurements per year % of *wqi* indicating a given class			24 measurements per year % of *wqi* indicating a given class		
[km]	year	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
49.9	2006	88.8	10.8	0.4	92.7	7.3	0.0	91.2	8.8	0
	2007	90.0	9.2	0.8	88.3	10.8	0.8	86.7	13.3	0
	2013	85.0	14.2	0.8	85.0	15	0	84.2	15.8	0
55.4	2007	82.3	16.2	1.5	80.0	20	0	80	20	0
	2008	87.5	10.8	1.7	85.0	15	0	83.3	16.7	0
	2010	83.8	15.4	0.8	85.4	14.6	0	85.4	14.6	0
87.0	2006	88.6	10.9	0.5	87.7	11.4	0.9	89.1	10.9	0
94.1	2007	85.4	13.8	0.8	85.8	14.2	0	84.2	15.8	0
	2010	84.3	13.6	2.1	85.0	13.6	1.4	85.7	12.9	1.4
100.0	2006	88.4	11.1	0.5	86.3	13.7	0	85.3	14.7	0
106.7	2013	88.9	10.4	0.7	89.3	10.4	0.4	90	10	0
117.0	2006	88.5	11.5	0.0	86.5	13.5	0	86.5	13.5	0
120.3	2006	66.7	28.1	5.2	59.5	35.2	5.2	59	34.8	6.2
	2007	68.0	26.0	6.0	70.4	24.8	4.8	68.8	26	5.2
	2008	63.6	31.4	5.0	62.9	32.9	4.3	62.1	30.7	7.1
131.2	2006	52.3	40.9	6.8	57.3	33.6	9.1	56.8	34.5	8.6
135.0	2013	70.4	25.0	4.6	71.2	24.2	4.6	71.5	24.6	3.8
164.0	2006	54.3	39.5	6.2	56.2	37.1	6.7	55.2	36.7	8.1
	2007	62.6	30.0	7.4	61.9	29.3	8.9	63.7	26.7	9.6
	2008	55.3	39.3	5.3	54.7	42.7	2.7	56.7	38.0	5.3
	2010	51.3	34.7	14	57.3	29.3	13.3	54.7	32.0	13.3
320.6	2007	67.0	26.5	6.5	67.5	27.0	5.5	66.5	28.0	5.5
	2008	80.0	20.0	0.0	79.3	20.7	0.0	78.7	21.3	0.0
339.1	2007	81.1	13.3	5.6	78.9	17.4	3.7	79.6	16.7	3.7
	2008	71.2	18.8	10.0	72.4	17.1	10.6	72.4	16.5	11.2
	2013	73.8	23.8	2.3	72.3	26.2	1.5	70.0	29.2	0.8
341.8	2007	70.0	19.1	10.9	70.9	16.4	12.7	72.7	15.5	11.8



45.8% high (blue), 41.7% good (green), 12.5% below good (red) | 35% high (blue), 55% good (green), 10% below good (red)

Fig. 2. Percentage of physico-chemical water quality elements classified into three classes as a result of sampling with A) frequency 4 times per year (left picture) and B) 24 times per year (right picture) for water body Mała Wełna

frequency (i.e. once per month) can lead to a false class assessment for more than 50% of *wqi*. Discrepancies in the assessed status class assessments for water bodies in this river can occur in almost 70% of cases.

A possible explanation of the observed differences between the results presented in Table 3 and Table 4 is that all water bodies in the Gwda river were assessed as being in *good ecological status* for all of those years and were characterized by a small variance in *wqi* measurement values, whereas water bodies in the Orla river, in the majority of cases were assessed as *below good* status, with accompanying relatively large variance in indicator values.

Table 3. Percentage of incorrect physico-chemical indicator based classes resulting from sampling 4, 6, 12 times per year with reference to classification based on 24 samples per year for river Gwda in years 2016 and 2014

River	Location along the River [km]	Year	% of incorrect classification based on measurements with frequencies 4, 6, 12 and 24 measurements per year		
			4–24	6–24	12–24
Gwda	0.3	2006	9.1	9.1	9.1
		2007	16.7	16.7	16.7
		2008	6.7	13.3	13.3
		2012	8.3	8.3	8.3
		2013	0.0	0.0	0.0
		2014	0.0	0.0	0.0
	16.0	2006	5.6	5.6	5.6
	34.0	2007	11.5	15.4	3.8
		2010	7.7	15.4	7.7
	37.9	2006	11.8	5.9	11.8
	48.3	2006	41.2	35.3	17.6
		2008	21.4	14.3	7.1
		2012	0.0	14.3	7.1
	78.0	2006	15.8	15.8	10.5
		2007	19.0	14.3	9.5
		2008	13.3	13.3	13.3
		2012	16.7	0.0	0.0
	112.0	2006	15.0	5.0	5.0
		mean	**12.2**	**11.2**	**8.1**

Table 4. Percentage of incorrect physico-chemical indicator based classes resulting from sampling 4, 6, 12 times per year with reference to classification based on 24 samples per year for river Orla in years 2006–2013

River	Location along the river [km]	Year	% of incorrect classifications based on measurements with frequencies 4, 6, 12 and 24 measurements per year		
			4–24	6–24	12–24
Orla	1.1	2007	30.0	10.0	10.0
		2010	35.7	35.7	28.6
	22.1	2007	35.7	50.0	35.7
		2008	46.2	69.2	53.8
		2009	66.7	41.7	25.0
		2010	50.0	21.4	14.3
	39.4	2006	26.1	21.7	26.1
		2007	44.4	38.9	27.8
		2008	42.9	42.9	42.9
		2009	46.2	69.2	46.2
		2010	46.2	30.8	23.1
	49.0	2013	50.0	16.7	41.7
	52.6	2006	41.7	29.2	20.8
		2007	42.9	35.7	35.7
		2008	61.5	46.2	61.5
		2009	69.2	46.2	38.5
		2010	58.3	25.0	25.0
		mean	**46.7**	**37.1**	**32.7**

An example of differences between the Orla and the Gwda rivers in their water quality characteristics are median and interquartile ranges of ammonia nitrogen (N NH4) observed at monitoring stations on both rivers within the period 2006–2013. They are presented in Fig. 3 in the form of box plots.

It can be observed that in order to compare ranges of concentration of *wqi* in the two rivers, the vertical scale for the Orla river must be made ten times greater than for the Gwda river. The corresponding means and standard deviations for BOD5, ammonia nitrogen (N-NH4), phosphates (PO4), total phosphorus (Ptot) and total nitrogen (Ntot), showing consequently for the Orla higher values of means and variance than for the Gwda, are presented in Table 5 for two selected monitoring stations on both rivers. Also the results of F.test for variance in the two rivers show that the ratios of variances of corresponding indicators for the Gwda and the Orla are considerably lower than one. Additional test for equality of variances *wqi* in the two rivers has resulted in rejecting this hypothesis, which confirms that the difference in the random fluctuations in concentration of the of rivers' indicators is statistically significant.

The overall discrepancy in classifications based on mean values of water quality parameters resulting from sampling with frequency 4, 6, 12 and 24 for all analyzed water bodies in Greater Poland are presented in Table 6.

It can be concluded from the Tab.6 that in average, in about 22% of cases, classes of *wqi* assessed from the standard monitoring program (that from sampling with frequency 12 times per year) are not correct. In 33% of cases, an indicator class could be correctly assessed based on sampling frequency three times smaller than once per month and in only about 50% of cases, classifications based on the reference frequency of water sampling are better than the ones obtained with smaller sampling frequency.

Discussion

When seeking new, more reliable water quality assessment rules for future water monitoring programs, it seems necessary to accept the fact that the dependence of water quality assessment on water sampling frequency is not simply monotonic nor it can be arbitrarily assumed. The analyses made in this work suggest that the type of monitoring program – surveillance, operational,

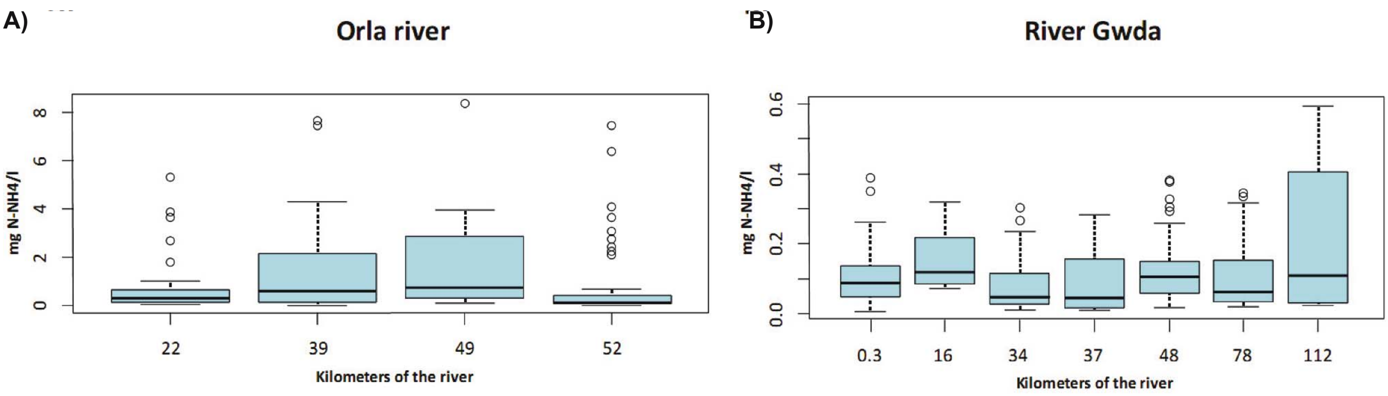


Fig. 3. Box plots for concentration of ammonia nitrogen in different water bodies of A) the Orla river B) the Gwda river during period 2006–2013

Table 5. Comparison of means – i) and standard deviations – ii) for water bodies of Orla and Gwda rivers

i)	Orla 39				Gwda 34			
	2006	2007	2008	2009	2007	2008	2009	2012
BOD5	1.917	3.747	4.578	2.600	2.483	2.08	1.96	2.192
N-NH4	1.097	1.055	1.294	1.509	0.077	0.045	0.094	0.091
PO4	2.084	1.144	1.714	1.230	0.160	0.202	0.182	0.161
Ptot	0.781	0.594	0.748	0.552	0.119	0.113	0.113	0.106
Ntot	11.456	14.346	9.660	9.953	1.953	1.806	1.328	1.965
ii)	**Orla 39**				**Gwda 34**			
	2006	2007	2008	2009	2007	2008	2009	2012
BOD5	0.878	3.317	3.390	1.956	1.411	0.789	0.884	1.013
N-NH4	2.189	1.262	1.272	2.278	0.672	0.032	0.098	0.077
PO4	0.874	0.768	1.219	0.651	0.081	0.112	0.078	0.055
Ptot	0.331	0.406	0.493	0.269	0.034	0.034	0.033	0.023
Ntot	8.311	11.074	6.384	7.067	1.004	1.155	0.353	0.940

Table 6. Summary table – a comparison of w.b. classification results obtained for water sampling frequencies 4, 6, 12 and 24 samples per year

	4–24 measurements per year	6–24 measurements per year	12–24 measurements per year
The average percentage of the amount of indicators for which the sampling frequency less than 24 per year resulted in different class than the "true class"	33.5%	28.2%	22.1%
The percentage of cases when class assessment based on 4,6 and 12 samples per year respectively was better than for the other two frequencies.	3.4%	14.6%	50.7%

investigative or for protected areas, cannot predetermine the choice of water sampling frequency as it is practiced in Poland now. The example of two Greater Poland rivers – the Gwda and the Orla, indicate that status classifications made on the basis of measurements with different water sampling frequency (Tables 3 and 4), lead to completely opposite conclusions as to whether the present water sampling frequency should be kept unchanged, reduced or increased. It seems therefore important to formulate principles for designing sampling frequencies which are adequate to keep the uncertainty of the status class assessment acceptably low. This is indeed a fundamental issue for urgent research. It is also evident from the analysis that the standard deviations of water quality indicators which are classified "*below good*" are greater than for elements which result in classifications of "*good*" and "*high*" status. Similar dependence has been documented by Wierzchołowska (2012). For this reason, when comparing the variances of the indicators in the rivers Orla and Gwda, it was evident that standard deviation values for the Orla river were much higher than for the Gwda river.

The *wqi* variance is an essential parameter on which a choice of water sampling frequency should be based. To obtain further insight into the magnitudes and differences in the variances, the data from the former WFD assessment cycles should be utilized more effectively. This idea is confirmed indirectly in the review paper of Strobl and Robillard (2008), who state conclusively that "the design of a monitoring network needs to be periodically re-assessed and accordingly modified to account for changing environmental conditions". This is needed because of the natural evolution of aquatic systems and due to growing pressures from human activities in the catchment. In particular, when facing the effects of climate warming including its influence on droughts and flood frequencies, monitoring programs should be systematically revised especially in terms of the frequency of water sampling.

The differences in the assessed water body classes determined from the response to different sampling frequencies, and exemplified by the two extreme cases of the rivers Orla and Gwda, are also in agreement with the observations reported by Skeffington et al. (2015) who analyzed 95% confidence intervals for weekly and monthly sampling. As for the polluted water of the Orla river, the mean values in many samples were close to the class boundary and in such a case, consistent assignment to a single class under repetitive sampling was unlikely; classification from several indicators characterizing water bodies in this river was distributed across two or three classes. For the clean Gwda river, results of status classifications were practically invariant with the frequency of sampling.

The uncertainty originating from the complexity of spatial and temporal processes in the aquatic environment as well as the confusions and errors accompanying the classification protocols should be taken into account when assessing water body status. To study the influence of various factors on the classification uncertainty, additional information such as flow rates, vegetation phases, changes in land use, and air temperature should be collected and stored in the water monitoring data bases. Also identifiers for the measurement team (different team members can be characterized by different variabilities in measurement outcomes) responsible for taking measurements and sampling should be stored together with the measurements.

When formulating guidelines for choosing a water sampling frequency as a basis for water quality assessment, it is important also to determine the relationship between different ranges of *wqi* variance and particular types of pressures acting on the river waters, characteristic flow rates in a river and other hydrogeomorphological parameters. It seems that for catchments with prevailing land use in form of extensively used meadows and pastures and for forests, the sampling frequency could be lower than for agricultural watersheds or urban catchments (Shupe 2017). Particularly under storm water conditions (Paule-Mercado et al. 2017), a sampling frequency should be adopted accordingly.

The uncertainty related to a water body status assessment should always be compared to some reference (or acceptable) level of uncertainty, preferably the very same as the one decided on and used by water managers in their risk analysis protocol. In general, the uncertainty related to water body status assessment is a risk factor that should be taken into account in water management, especially when making a tradeoff between the costs of a too intensive water monitoring program, on the one hand, and the cost of investment in the (possibly) unnecessary corrective or remedial measures arising from water sampling of a given frequency, on the other hand. Essentially, the necessary frequency of water sampling for a given water body can be derived from a relationship between water sampling frequency and the uncertainty embedded in the water body status assessment. In case studies reported in this article such a relationship has been assessed from the monitoring data. Although important, the frequency of water sampling is only one parameter in the design of water monitoring schemes. Other monitoring parameters and aspects like timing, location of representative points within a water body, lists of quality indicators to be measured, instrumental as well as analytical methods used, statistical procedures applied to data interpretation, all seem to be no less important and require more systematic and integrated evaluation.

Despite the lack of sufficient biological data and measurements of specific pollutants, and analyzing only physico-chemical indicators, the issue of water body status classification in rivers, discussed in this study, has revealed concepts more widely applicable in the future developments of river water monitoring programs.

Conclusions

In conclusion, the greater the variance of the indicator and the smaller the sampling frequency, the more likely the assessed water body status will differ from the reference status assessed. Hence, for the rivers with large variances in measurement data, a greater frequency of measurement should be considered. Although this conclusion seems generally true and unsurprising, this analysis and experimental (monitoring) investigation has revealed its complexity with respect to the assessment of water body status. Indeed, the study has paved the way for new research on river water monitoring as a basis for water resources protection and, in general, for water management.

The effect of water quality indicator variance on the assessed status class can be applied to other quality indicators as well, not only to the limited number of physic-chemical indicators as presented in this paper. In particular, when designing future monitoring schemes, it is important to consider the value of more frequent measurements of specific pollutants and the corresponding link between their variability and the water sampling frequency.

In a simple case, when there are no other objectives for monitoring programs other than status assessment, the frequency of sampling can be reduced in two cases. In the first case, if the water body status reported to date is "*high*" and there are no pressures in the catchment so that the water quality monitoring system is implemented mainly as a regulatory or precautionary measure, and the sampling frequency can be safely reduced. In the second case, if the water body status is *"below good"* and there is little chance of water quality improvement, it seems better to refrain from frequent measurements and invest in remedial actions. An exception to this reasoning is for the safety of water resources used for drinking water supply. The ultimate goal is to choose an appropriate frequency of water sampling from the spectrum of options ranging between these two extremes, and to justify the costs of water sampling corresponding to the chosen frequency.

This study is setting a base for extending the investigations and analysis to the regional scale, and to include relationships between pressures within a river catchment, land use and the magnitude of variance of the *wqi* in the river systems. It can be envisaged that the *change in the present paradigm of river water monitoring* may be made feasible by explicitly considering the involvement of water managers and other stake-holders in water management planning and by tailoring new monitoring programs to appropriate environmental goals and economic constraints.

As there cannot be universal guidelines for determining the optimal frequency of water sampling, the task has to be approached locally and catchment-wise. By analyzing hard monitoring data and revealing relationships which are not necessarily obvious, this paper can be considered as a starting point for proposing guidelines for shaping new water monitoring programs, at least in Poland. The guidelines should respect the environmental criteria for water resources protection by making use of analyses made and ideas developed in this study. In particular the guidelines should take into account the links between the variance in water quality indicators and the catchment's processes and pressures. They should also include the relationship between the uncertainty in the assessment of water body status and the frequency of water sampling from which it is determined. Reference must also be made to water protection practices and practitioners in order to keep the monitoring budget on a cost-effective level.

References

Amruta, M.K. & Satish, M.T. (2013). Solar powered water quality monitoring system using wireless sensor network, In: Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013 International Multi-Conference on IEEE (pp. 281–285).

Behmel, S., Damour, M., Ludwig, R. & Rodriguez, M. J. (2016). Water quality monitoring strategies – A review and future perspectives. *Science of the Total Environment,* Tom 571, pp. 1312–1329.

Elliott, M., Andersen, J.H., Cardoso, A.C., Carstensen, J., Ferreira, J.G., Heiskanen, A.-S., Marques, J.C., Neto, J.M., Teixeira, H., Uusitalo, L., Uyarra, M.C. & Zampoukas, N. (2013). Good Environmental Status of marine ecosystems: What is it and how do we know when we have attained it?, *Marine Pollution Bulletine*, pp. 16–27.

Canadian Council of Ministers of the Environment (2015). *Guidance manual for optimising water quality monitoring program design,* brak miejsca: Canadian Council of Ministers of the Environment.

Carstensen, J. & Lindegarth, M. (2016). Confidence in ecological indicators: A framework for quantifying uncertainty components from monitoring data, *Ecological Indicators*, Issue 67.

CIS (2003). *Guidance document No 7. Monitoring under Water Framework Directive,* Luxembourg: Office for Official Publications of the European Communities.

Clarke, R. (2012). Estimating confidence of European WFD ecological status class and WISER Bioassessment Uncertainty Guidance Software (WISERBUGS), *Hydrobiologia*, Issue 704.

D'Agostino, R.B. (1986). Tests for the Normal Distribution, In: R.B. D'Agostino & M. A. Stephens, redaktorzy Goodness-of--Fit Techniques*,* New York: Marcel Dekker.

Dz.U. 2016, poz. 1911 (2016). *Rozporządzenie Rady Ministrów w sprawie Planu gospodarowania wodami na obszarze dorzecza Wisły.*

Dz.U. 2009, nr 122, poz. 1018 (2009). *Rozporządzenie Ministra Środowiska w sprawie klasyfikacji stanu ekologicznego, potencjału ekologicznego i stanu chemicznego jednolitych części wód powierzchniowych.*

Dz.U. 2014, poz. 1482 (2014). *Rozporządzenie Ministra Środowiska z dnia 22 października 2014 r. w sprawie sposobu klasyfikacji stanu jednolitych części wód powierzchniowych oraz środowiskowych norn jakości dla substancji priorytetowych.*

Dz.U. 2016, poz. 1967 (2016). *Rozporządzenie Rady Ministrów w sprawie Planu gospodarowania wodami na obszarze dorzecza Odry.*

Dz.U. 2016, poz. 1911, poz. 1967 (2016). *Rozporządzenie Rady Ministrów w sprawie Planu gospodarowania wodami na obszarze* *dorzecza Wisły. Rozporządzenie Rady Ministrów w sprawie Planu gospodarowania wodami na obszarze dorzecza Odry.*

Dz.U. 2014, poz. 1 (2014). *Rozporządzenie Ministra Środowiska z dnia 22 października 2014 r. w sprawie sposobu klasyfikacji stanu jednolitych wód powierzchniowych oraz środowiskowych norm jakości dla substancji priorytetowych.*

EC (2000). *Directive 2000/60/EC of the European Parliament and of the Council of 23 October establishing a framework for communinty action in the field of water policy,* Official J. of European Union.

EC (2013). *Directive 2013/39/EC of the European Parliament and the council amending Directives 2000/60EC and 2008/106EC as regards priority substances in the field of water policy.* Official Journal of the European Union.

Facchi, A., Gandolfi, C. & Whelan, M.J. (2007). A comparison of river quality sampling methodologies under highly variable load conditions, *Chemosphere*, pp. 746–756.

GIOS (2013). *Ocena stanu jednolitych części powierzchniowych wód płynących (w tym zbiorników zaporowych) w 2013 roku, z uwzględnieniem monitoringu w latach 2011 i 2012.* (http://www.gios.gov.pl/images/dokumenty/pms/monitoring_wod/Ocena_stanu_wod_powierzchniowych_plynacych_w_2013_roku.pdf (09.05.2018)).

GIOS (2017). (http://www.gios.gov.pl/pl/stansrodowiska/monitoring-wod (26.04.2018)).

Guigues, N., Desenfant, M., Lalere, B., Vaslin-Reimann, S., Eyl, D., Mansuit, P. & Hance, E. (2016). Estimating sampling and analysis uncertainties to assess the fitness for purpose of water quality monitoring network, *Accreditation and Quality Assurance*, pp. 101–112.

Kovacs, J., Korponai, J., Kovacs, I. & Hatvani, I.G. (2012). Introducing sampling frequency estimation using variograms in water research with the example of nutrient loads in the Kis-Balaton Water Protection System (Hungary), *Ecological Indicator*, 42, pp. 237–243.

Loga, M. (2012). *Hierarchical approach to water body status misclassification,* Wessex, WIT Press.

Loga, M. (2016). An estimation of uncertainty measures, confidence and precision, of surface water status assessment, *Ochrona Środowiska*, 38, pp. 15–23.

Loga, M. & Nawalany, M. (2009). *The mew methods for assessing the state of uniform bodies of water.* (http://www.wios.tarnow.pl/pl0302/rezultaty/oceny_stanu_wod_powierzchniowych/raport_c_1.pdf (03.2017).

Naddeo, V., Scannapieco, D., Zarra, T. & Belgiorno, V. (2013). River water quality assessment: Implementation of non-parametric tests for sampling frequency optimization, *Land Use Policy*, pp. 197–205.

Neal, C., Reynolds, B., Rowland, P., Norris, D., Kirchner, J.W., Neal, M., Sleep, D., Lawlor, A., Woods, C., Thacker, S., Guyatt, H., Vincent, C., Hockenhull, K., Wickham, H., Harman, S. & Armstrong, L. (2012). High-frequency water quality time series in precipitatio nad straem flow: From fragmentary signals to scientific challenge, *Science of the Total Environment*, pp. 3–12.

R Core, T. (2012). *A language and environment for statistical computing. R Foundation for Statistical Computing,* Vienna, Austria.

Skarbovik, E., Stalnacke, P. & Bonsnes, T.E. (2012). Impact of sampling frequency on mean concentrations and estimated loads of suspended sediments ina Norwegian river: Implications for water management, *Science of the Total Environment*, 433, pp. 462–471.

Soszka, H., Kolada, A., Pasztaleniec, A., Ochocka, A., Kutyła, S. & Bielczyńska, A. (2016). *Opracowanie zbiorczej oceny stanu jezior. Załacznik 4.8. do opracowania „Przetworzenie i zwryfikowanie danych PMŚ w zakresie monitoringu jezior z lat 2013–2015 wraz z opracowanie oceny stanu oraz nadzorem merytorycznym”,* GIOŚ.

Strobl, R.O. & Robillard, P.D. (2008). Network design for water quality monitoring of surface freshwaters: A review, *Journal of Environmental Management*, pp. 639–648.

Urząd Marszałkowski (2017). (https://www.umww.pl/informacja-o-regionie_informacje-oregionierolnictwo ((20.03.2017).

Vilmin, L., Flipo, N., Escoffier, N. & Groleau, A. (2016). Estimation of water quality of large urbanized river as defined by European WFD: what is the optimal sampling frequency., *Environmental Science and Pollution Research*, pp. 1–17.

Wierzchołowska, A. (2012). *Zastosowanie hierarchicznej metody oceny prawdopodobieństwa błędnej klasyfikacji stanu wody dla rzek wybranych dwóch województw,* Politechnika Warszawska.

WMO (2013). Planning of Water Quality Monitoring Systems, WMO-No.1113, World Meteorological Organization, Geneva, Switzerland 2013.

Wpływ częstości pobierania próbek na wynik oceny stanu części wód – na przykładzie rzek w woj. wielkopolskim

Streszczenie: Zgodnie z Ramową Dyrektywą Wodną ocena stanu jednolitych części wód jest wyznaczana na podstawie pomiarów monitoringowych. Celem artykułu jest analiza wiarygodności tej oceny w zależności od częstotliwości pobierania próbek. Do analizy wykorzystano dane dotyczące rzek w województwie wielkopolskim, uzyskane z systemu Państwowego Monitoringu Środowiska. Przeprowadzono testy zgodności rozkładów wartości wskaźników jakości wody z rozkładem normalnym, a następnie wyznaczono ich podstawowe charakterystyki statystyczne. Do generowania, na podstawie wyznaczonych rozkładów, syntetycznych serii pomiarowych wskaźników o różnej liczności próbek, zastosowano modele Monte-Carlo. Przyjmując jako częstotliwość referencyjną 24 pomiary w ciągu roku (największą częstotliwością monitoringu, ekonomicznie i organizacyjnie możliwą do realizacji) wykazano, że w 22% przypadków klasyfikacja wskaźników uzyskana na podstawie 12 pomiarów w ciągu roku jest obarczona błędem. W nieco mniej niż 50% przypadków wyniki oceny uzyskane przy zastosowaniu referencyjnej częstotliwości nie mogły być zastąpione pomiarami o mniejszej częstotliwości. W 33% przypadków klasa wskaźnika była prawidłowo wyznaczona na podstawie zaledwie 4 pomiarów w ciągu roku. Związek pomiędzy oceną klasy wskaźnika jakości wody a częstotliwością pobierania próbek jest skomplikowany i nie może być przyjmowany arbitralnie. Przy wyborze częstotliwości próbkowania w monitoringu rzek konieczne jest wzięcie pod uwagę naturalnych i antropogenicznych czynników i presji z obszaru zlewni wpływających na wody rzeczne.