

<https://doi.org/10.1590/2318-0331.241920180077>

Spectral analysis in determining water quality sampling intervals

A análise espectral na determinação de intervalos de amostragem da qualidade da água

Régis Leandro Lopes da Silva¹ , André Luiz Lopes da Silveira²  and Geraldo Lopes da Silveira³ 

¹Universidade Federal de Santa Maria, Cachoeira do Sul, RS, Brasil

²Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

³Universidade Federal de Alfenas, Poços de Caldas, MG, Brasil

E-mails: regis.silva@ufsm.br (RLLS), andre@iph.ufrgs.br (ALLS), geraldo.silveira@unifal-mg.edu.br (GLS)

Received: May 16, 2018 - Revised: July 09, 2019 - Accepted: September 02, 2019

ABSTRACT

To make water quality series more representative, real-time monitoring techniques are developed. However, these techniques have obstacles in their use, such as high costs and difficulties in equipment installation, maintenance, and calibration. One alternative is near-real time water quality monitoring (NRTWQM), with sampling done less frequently than daily. The study objective was to evaluate, through spectral analysis, the water quality sampling frequency representativity for different catchments. For this purpose, a historical series of real time water quality monitoring stations were used in Brazil, Canada, and the USA. These series were submitted to spectral analysis to identify the denser frequencies and their representativeness across the series. To obtain the sampling intervals, the Nyquist-Shannon theorem was applied. Weekly intervals accounted for 65% of cumulative frequencies for the three verified parameters, and the sampling intervals obtained by means of the characteristic frequencies were shown to be executable in the NRTWQM models for up to the 90% of cumulative frequency. For cumulative frequency above 90%, the intervals approach the daily values.

Keywords: Sampling frequency; Monitoring; Near real time.

RESUMO

Para gerar séries históricas de qualidade da água mais representativas desenvolvem-se técnicas de monitoramento em tempo real. Contudo, estas técnicas possuem obstáculos quanto ao uso: alto custo, dificuldades de instalação, manutenção e calibração dos equipamentos. Uma alternativa seria o chamado monitoramento da qualidade da água em tempo quase real (MQATQR), que envolve amostragens com frequência menor que a diária. O objetivo deste trabalho foi avaliar, por meio da análise espectral, a representatividade de frequências de amostragem da qualidade da água para diferentes tipos de bacias. Para isso foram utilizadas séries históricas de estações de monitoramento da qualidade da água em tempo real no Brasil, Canadá e EUA. Estas séries foram submetidas à análise espectral utilizando o janelamento de Hann, para identificação das frequências mais densas e da representatividade delas dentro das séries, com o auxílio de curvas de frequência acumuladas. Para obtenção dos intervalos de amostragem foi aplicado o teorema de Nyquist-Shannon. Intervalos semanais representaram 65% de frequências acumuladas para os três parâmetros analisados, e os intervalos de amostragem obtidos se mostraram executáveis nos moldes do MQATQR até a frequência acumulada de 90%. Para as frequências acumuladas maiores que 90% os intervalos se aproximam do diário.

Palavras-chave: Frequência de amostragem; Monitoramento; Tempo quase-real.



INTRODUCTION

Demands for water have led man to create and refine techniques for monitoring water quantity and quality in a manner that the data acquired guide help management, engineering, and environmental decisions. For quantitative uses, that is, those that will derive, extract, divert water from a course or change its regime, modern availability of data has not been enough despite being more abundant. Rainfall and fluviometric monitoring networks with daily temporal frequencies already operate in many stations in Brazil, with representative historical series for many catchments (ANA, 2017). However, when it comes to water quality data, time availability at monitoring sites is very low and often does not provide a solid basis for water management support.

In Brazil, this monitoring is generally done in a traditional manner by collecting samples and analyzing them in laboratories. Although this form allows a larger number of analyzed parameters, it has low temporal density (HANISCH; FREIRE-NORDI, 2015).

Real-time water quality monitoring (RTWQM) (DONG et al., 2016; ZAMYADI et al., 2016; MEYER et al., 2019) using automatic monitoring stations generates a well-established database much larger in terms of temporal representativeness, despite little use in Brazil. There are few stations with this type of operation in the country due to high costs, long-term maintenance, and various operational difficulties.

One alternative is a compromise between the two previously mentioned forms of monitoring, which is near-real-time water quality monitoring (NRTWQM). This form of monitoring is still not widespread and consists of the on-site measurement of water quality parameters by an operator, e.g. a research institution or environmental control body. This operator makes the measurements by means of mobile sensors with frequencies above the conventional and enough for the purposes of the proposed monitoring program and lower than the daily frequency. Disclosure of the data collected solely depends on the time it takes the operator to reach a location with possible transmission, with the possibility of being immediate. Thus, there may be greater spatial coverage with only one measuring equipment (LINKLATER; ORMECI, 2013; SILVA, 2018). This monitoring may result in lower costs compared to real-time monitoring and a potential option for developing a more robust water quality database. Developing countries with few or no resources for environmental control and large areas that require numerous stations for reasonable spatial coverage demand alternative forms of monitoring.

In this context, the question arises: what kind of monitoring is enough in terms of sampling frequency as a database for different aspects of water resource management? To answer this, we assume that a RTWQM data series is nothing but the recording of a signal over time. In addition, the recording of this signal holds several characteristics that are of fundamental importance to understand the phenomenon that generated such signal, including duration, amplitude, and frequency (OPPENHEIM; WILLSKY; YOUNG, 1983). However, recording this signal is not a simple task, because to do so, it is necessary to collect several samples of it over time to estimate its characteristics or reproduce it if necessary (OPPENHEIM; SCHAFFER; BUCK, 1999). Furthermore, how

do we know if we are sampling correctly? Are we missing any important features of this signal? If the signal is relatively simple, such as a constant amplitude sine wave, defining the sampling rate is not difficult, although the more complex the signal, the more complex to define the sampling rate.

For the historical series of water quality data, several factors may influence the frequencies present in the signal, such as rainfall seasonality, intraday variations in atmospheric pressure, variations in metabolism of aquatic organisms, and even cyclic effluent discharges (CARVALHO et al., 2015; DAMASCENO et al., 2015; GIRARDI et al., 2016). Nevertheless, it is possible to verify which frequencies have the highest density in the signal, that is, which are present and which predominate in the historical series, showing which periodicities are relevant.

Spectral analysis is a powerful tool to analyze historical series, and although not many studies have used it for water quality data series, it is already commonly used for environmental data series analysis (NEEFF et al., 2005; MENEZES; SOUZA; PESSANHA, 2014).

Thus, the objective of this study was to evaluate, through spectral analysis, the representativeness of water quality sampling frequencies for different types of catchments and determine possible sampling for the NRTWQM model. The hypothesis is that there are sampling intervals that provide greater representativeness of the water quality series than the conventional monitoring applied today, and which are technically feasible, considering the difficulties of implementing real-time monitoring. Greater representativeness of the series gives the water resource managers a better database for more technically based decisions or better monitoring the evolution of river water quality conditions.

MATERIAL AND METHODS

Monitoring stations

In order to apply the spectral analysis proposed in this study, water quality monitoring stations with significant temporal density were required to evaluate the most common frequencies and sampling intervals that characterize them. Seventy-six stations that had long historical series of RTWQM in rivers with automatic monitoring platforms were used, being four of them in Brazil and operated by CETESB. Five are in Canada and operated by the Water Resources Management Division (WRMD) of the government body Newfoundland and Labrador and Nova Scotia Environment, and 67 in the United States that are operated by the USGS (United States Geological Survey).

The characteristics of water quality monitoring stations and their respective historical series are summarized in Table 1, where N is the number of points, HSMT is the historical series mean time, and the areas refer to the drainage areas of the monitoring stations. The location of the stations is summarized in Figure 1. The monitoring performed at these points and that compose the historical series is RTWQM, with a temporal resolution of 1 hour. The parameters considered for the analysis were dissolved oxygen (DO), pH, and electrical conductivity.

Spectral analysis

A historical series is nothing more than a record of a signal composed of different frequencies over time. Determining the importance of each of these frequencies enables the use of tools such as spectral analysis. For this, by using the Fourier series in its exponential form, a signal $f(t)$ can be expressed as follows:

$$f(t) \approx \sum_{k=-\infty}^{\infty} c_k e^{j\omega_0 kt}, k = \pm 1, \pm 2, \pm 3, \dots \quad (1)$$

Table 1. Characteristics of historical series and drainage areas of selected water quality monitoring stations.

Country	N	HSMT	Minimum Area	Maximum Area
Brazil	4	9.2 years	240 km ²	16,100 km ²
Canada	5	8.4 years	64 km ²	7,900 km ²
USA	67	6 km ²	2,916,000 km ²	

N = number of points; HSMT = historical series mean time.

Being t time; j a constant for the signal; ω_0 the fundamental angular frequency; and c_k the Fourier exponential series coefficients, or spectral coefficients, which are given by Equation 2, where T is the fundamental period:

$$c_k = \frac{1}{T} \int_T f(t) e^{-j\omega_0 kt} dt, k = \pm 1, \pm 2, \pm 3, \dots \quad (2)$$

Representation of the spectral coefficients of a signal as a function of frequency results in a graph of frequency densities, which is also called PSD (Power Spectral Density). This graph is evaluated by spectral analysis.

The exponential form of the Fourier series is used for its application in continuous signal analysis. In the case of high temporal resolution data series such as RTWQM, it can be considered a continuous signal.

To eliminate the representation of frequencies generated by possible discontinuities that this signal may present and highlight

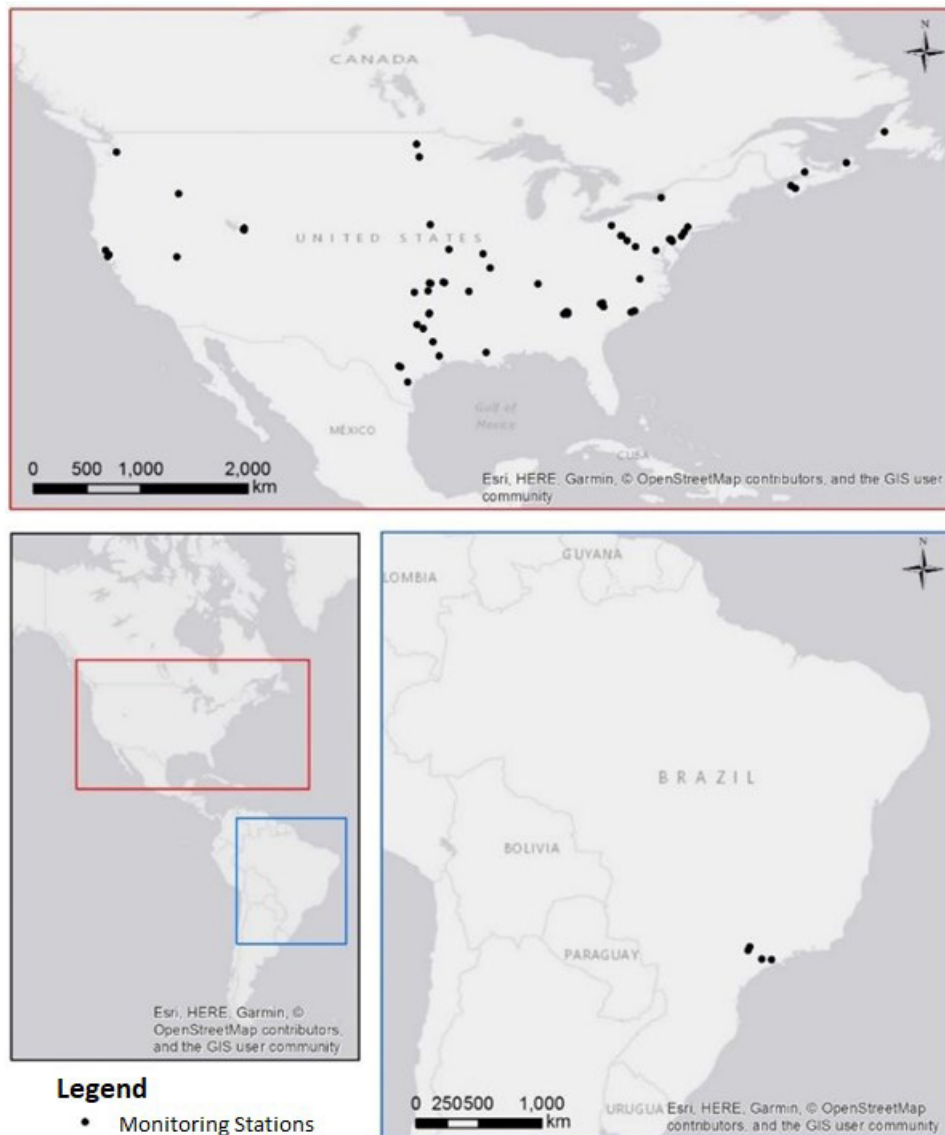


Figure 1. Location of selected water quality monitoring stations.

the frequencies of interest, windowing techniques are used, which smooth out the graph and make interpretation easier.

For this study, the Hann window (HARRIS, 1978) was used, which is best suited for general use or when the nature of the signal to be analyzed is unknown. The application of the Hann window for a window of size N is defined by Equation 3.

$$w(n) = 0.5 - 0.5 \cos\left(\frac{2\delta n}{N}\right), n = 0, 1, 2, \dots, N - 1 \quad (3)$$

The Eviews® 9 statistical software was used to perform the spectral analysis, which generated all the frequency spectra for the 76 monitoring stations used. Three parameters were considered in each station: conductivity, DO, and pH. After generating the spectra, the next stage of the study covered the accumulation of the frequency densities obtained in the spectral analysis.

Accumulated frequency density curves

After obtaining the frequency spectra for each monitoring station, the next step was to produce accumulated density curves, which are also called Cumulative Spectral Power (CSP). These curves are nothing more than the normalized integral of the frequency density spectrum (Figure 2).

To obtain this cumulative frequency density curve, just apply Equation 4:

$$\tilde{I}(\omega) = \int_0^{\omega} \tilde{S}(\omega') d\omega' \quad (4)$$

Being $\tilde{S}(\omega)$ the spectral density; $\tilde{I}(\omega)$ a positive and non-decreasing function, and:

$$\tilde{I}(0) = 0 \quad (5)$$

$$\tilde{I}(\infty) = 1 \quad (6)$$

After accumulation, the values were replaced by relative values (percentages of the total accumulated in each curve) to normalize all curves. The accumulated spectral density curves were generated for the historical series of the conductivity, DO, and pH parameters for the 76 monitoring stations analyzed. Here, these curves will be called cumulative frequency curves. In each of these curves, the frequency values corresponding to the accumulated spectral densities were extracted from 5 to 95% and in increments of 5% (5, 10, 15, ..., 95%) (Figure 3).

These characteristic values were used in the next methodological steps, which aim to relate the characteristic frequencies to the different monitoring stations used, and then determine the sampling intervals for each accumulated density.

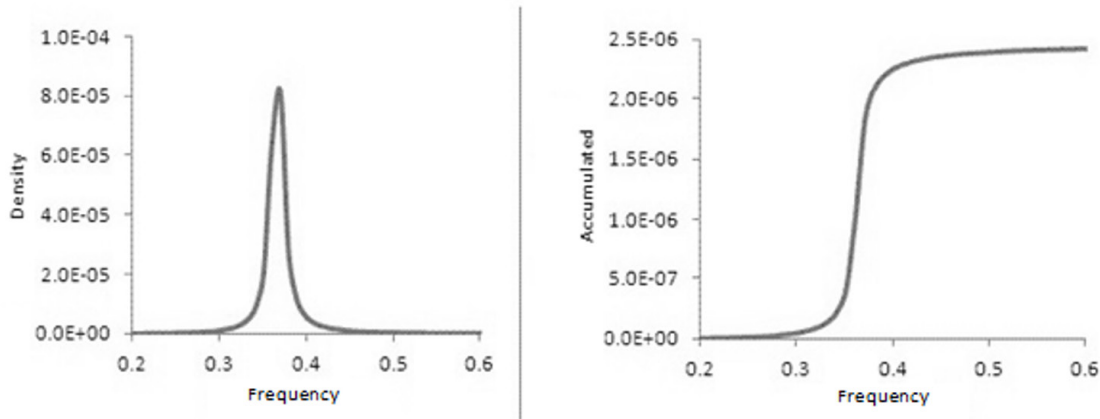


Figure 2. Frequency density and accumulated density curve.

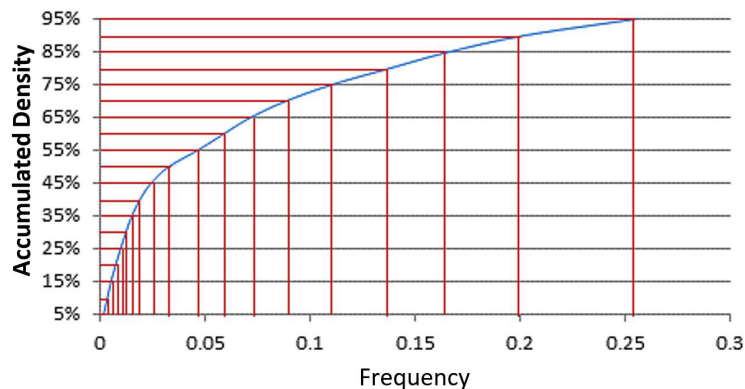


Figure 3. Characteristic frequencies extracted from the accumulated density curve.

Separation of monitoring stations into orders

Monitoring stations located in catchments with different formats, areas, and subject to different hydrological regimes were used in the present study. To verify the existence of any association between these catchment attributes and the characteristic frequencies extracted from the accumulated frequency curves, the monitoring stations were separated into five orders according to the catchment size:

- Small catchment: up to 100 km² in area;
- Medium-Small Catchment: between 100 and 1,000 km²;
- Medium Catchment: between 1,000 and 10,000 km²;
- Medium-Large Catchment: between 10,000 and 100,000 km²;
- Large Catchment: with an area larger than 100,000 km².

Of the four Brazilian catchments used in the study, two are medium-small, one medium-sized, and one medium-large. Then, the frequency averages obtained from the members of each order were compared to verify the existence of any significant difference between averages. For these comparisons, two multiple comparison tests were used: ANOVA and Welch (1951), which is the most indicated when the variances of the subgroups are heterogeneous. If comparison of the averages of two or more orders show no significant differences, the average of all stations that make up the orders that were compared is used.

Determination of sampling intervals

The main concern when sampling a signal is to do so with a temporal density that allows the estimation of the characteristics of this sample and that these characteristics are representative of the signal. In addition, it may be necessary to reproduce this signal, and if sampling is insufficient, such reproduction may be severely impaired by effects such as aliasing. For this effect not to occur, sampling must follow the Nyquist-Shannon Theorem (SHANNON, 1949), which is also known as the Sampling Theorem, which is given by Equation 7:

$$f_{max} \leq \frac{f_s}{2} \tag{7}$$

where f_{max} is the maximum frequency of the sampled signal and f_s is the sampling frequency. For this study, the maximum frequencies are the frequencies assigned to each catchment order at their level of representativeness of the frequency spectra (5, 10, 15, ..., 95% of the cumulative frequency curve).

Upon determining the sampling frequency, in order to have the sampling interval T, Equation 8 is applied:

$$T = \frac{1}{f_s} \tag{8}$$

Thus, the sampling intervals that correspond to the certain accumulated frequencies are obtained, which translate into representative ranges of the monitoring performed in the NRTWQM molds.

RESULTS AND DISCUSSION

Spectral analysis

Strong influence of seasonality was observed in all historical series in all spectral analyses, since the highest frequency densities were observed around 0.00274, which represents the approximate period of 365 days.

Graphs appear to have high densities close to zero, although this is an effect of the smoothing promoted by the Hann windowing applied to the graph. In fact, the density values for frequencies lower than the annual are very close to zero, but viewing them is only possible when smoothing is not applied.

The frequency density graphs of some of the 76 monitoring stations for the DO parameter are shown in Figure 4. They are organized by the size of the catchment, the top one being a small catchment, the middle one a medium catchment, and the bottom a large one.

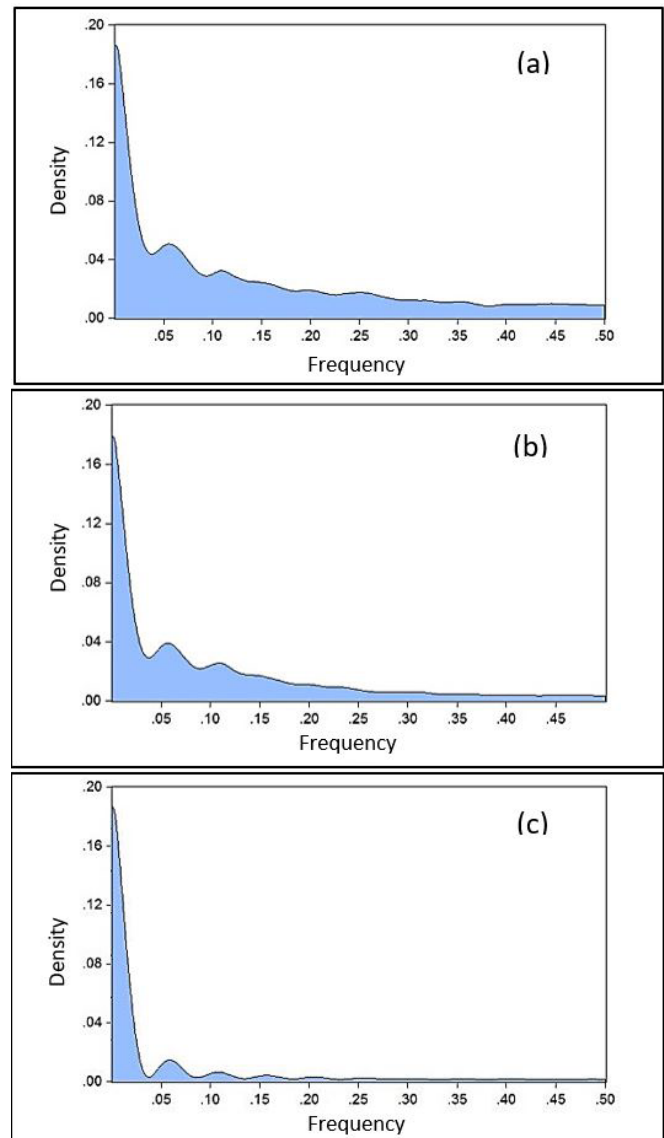


Figure 4. Frequency density spectra of historical DO series for a small (a), medium (b), and large (c) catchment.

Still concerning the DO parameter, regardless of the catchment size, a density peak around 0.065 was observed, which represents a period of approximately 15 days.

In smaller catchments, high frequencies have relatively higher densities than the ones for the high frequencies in larger catchments, that is, a higher presence of high frequencies in small catchments than in large catchments. This behavior was repeated in the spectral analyses of the historical series of the conductivity parameter.

For the pH parameter, the reduction in the higher frequency densities with the increase of the catchment was observed until the order of medium-large catchments. At stations of large catchments, densities were higher for higher frequencies, which represents increased presence of frequencies in the historical pH series for large catchments.

Feng, Kirchner and Neal (2004) also worked with spectral analyses of historical water quality series. The authors used Cl^- and Na^+ concentrations in historical flow series to determine the travel time of river catchment in Wales. They compared the frequency densities of two databases, one daily with three years and one weekly with 17 years of data, and found strong seasonal influence, with well-marked peaks in frequency corresponding to the annual period. Additionally, the authors also found differences between spectral densities when comparing analyses of the two parameters collected at the same station.

This difference between spectral behaviors between different quality parameters for the same monitoring station demonstrates that, even with strong influence of hydrological processes such as seasonality, other phenomena are independent of the hydrological behavior of the catchment and regulate the physical and chemical characteristics of waters. This only reinforces that decisions such as wastewater licensing are not based solely on qualitative aspects of the effluent and dilution capacity of the receiving body, but consider the evolution of the water quality of that course over time, which justifies the need to intensify water quality monitoring strategies.

In this study, although the spectral analyses showed the most common frequencies in the series, they were only the input to find the representation of each sampling interval. This representativeness was obtained by creating the accumulated frequency curves.

Accumulated frequency density curves

Cumulative frequency density curves are the integration of the frequency density function and allow us to assess how representative the sampling of any given signal is.

Assuming the sampling theorem, if we perform enough sampling to reproduce a given frequency ω present in a signal, this sampling will be enough to reproduce all frequencies below ω in this signal, although not the highest ones. Since the accumulated frequency curve accumulates frequencies from the lowest to highest, when we establish a point on this curve equivalent to a ω frequency, we are representing the signal as the cumulative densities of all frequencies up to ω .

In this context, cumulative frequency curves were generated from all 76 historical series of monitoring stations used here for

the three selected parameters. Some accumulated frequency curves for the conductivity parameter are shown in Figure 5.

They are organized by the size of the catchment, the top one being a small catchment, the middle one a medium catchment, and the bottom a large catchment.

As the catchment increases, so does the concavity of the curve in relation to the frequency axis, which shows a gain in representativeness of low to medium frequencies that increases the catchment. In most curves, the change in slope is always gradual and smooth, although only for the DO parameter and around the cumulative frequency of 50% is there usually any sudden change in the slope of the curve.

Based on the analyzed series, this means that it is easier to promote a gain in the increase of representativeness with an increase of the samples of up to the 50% representativity range. Above such range, even if sampling with higher frequencies, the incremental gain of representativity will be lower.

The accumulated pH frequency curves behaved similarly to the conductivity curves, but in some cases of small catchments, the curve approached more than one straight line than conductivity

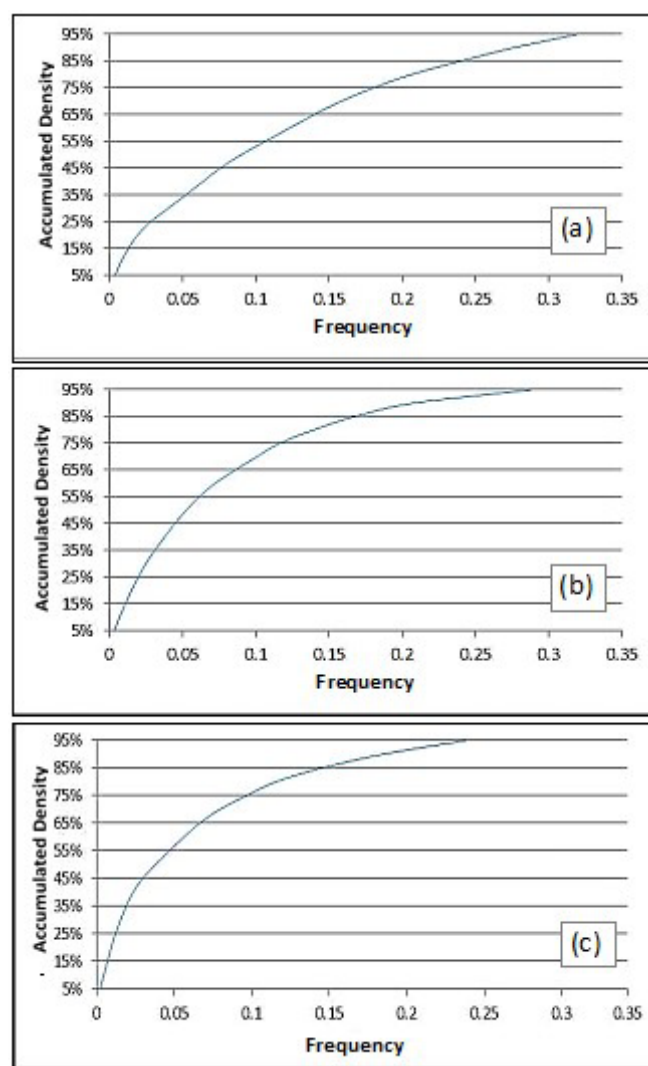


Figure 5. Cumulative frequency density curves from historical conductivity series for small (a), medium (b), and large (c) catchments.

and DO, reducing their concavity in relation to the frequency axis. This demonstrates a more homogeneous frequency density distribution in the spectral analysis for catchments of this order. This fact can be attributed to lower pH variability for different water quality conditions.

Separation of monitoring stations into orders

Since the drainage area is an easily accessible parameter, the monitoring stations used in orders were divided. The average values of frequencies linked to the accumulated density levels used were assigned to each order (5, 10, 15, ..., 95%). Afterwards, these averages were compared to establish the frequencies to be later used in determining the ideal sampling intervals for the NRTWQM strategy. The average frequencies of the stations belonging to each order are organized in Figure 6, for each of the accumulated frequencies extracted from the accumulated density curve, for the parameter conductivity (top), DO (center), and pH (lower). It is possible to notice the proximity of the mean values for some orders, which often overlap the curves.

For conductivity, until the cumulative frequency of 50%, the medium-small to large catchments are very close, and the order of small catchments is well detached, showing the well-differentiated behavior of frequencies for catchments up to 100 km². After 50% of accumulated frequency, the medium-large and large catchments begin to create distance from each other, and from these two the curves of the medium-small and medium catchments also begin to separate, although still joined together until the accumulated frequency of 95%. This union between the curves of the medium-small and medium orders shows that, in terms of the accumulated frequency for the conductivity parameter, the behavior does not differ between catchments from 100 km² to 10,000 km².

For the DO parameter, it is possible to observe that the frequencies of small catchments are always higher and with a certain distance from the frequencies of the medium-small to medium-large catchments, which have very close values. The frequencies of large catchments are always much lower than the others. The small catchment order curves remained close until the cumulative frequency of 35%, then taking off and resuming proximity at the larger cumulative frequencies. For the large catchment curve, detachment was quite marked after the cumulative frequency of 45%.

For pH, as in the other parameters, the order of small catchments maintained a certain isolation from the accumulated frequency of 30%, which once again corroborates the need for special attention to small catchments. This need for a special approach to small catchments is justified by the fact that, for all parameters, the mean accumulated density curves of the small catchments were distanced from the other curves. This behavior can be attributed to the greater hydrological variability of the small catchments as well as lower capacity to absorb sludge in these catchments.

The curves of the orders from medium-small to medium-large remained close, just not as overlapping as in the other two parameters. On the other hand, the curve of the large catchments presented an unexpected behavior and different from

the other two parameters, presenting frequencies even higher than the small catchment for accumulated frequency above 85%. However, this behavior observed for the pH parameter of large catchments should be evaluated sparingly, since any statement about this order has a great load of uncertainty due to the small number of stations under analysis compared to other orders.

The fact that some orders presented averages very close to others motivated the comparison between these averages through the ANOVA and Welch (1951) tests in order to verify if these orders actually have different averages or just sample variance and thus should receive a single value that is the average of all stations belonging to these comparing orders. Thus, after the mean comparison tests, new frequency values were assigned to each order for each accumulated frequency. The frequency

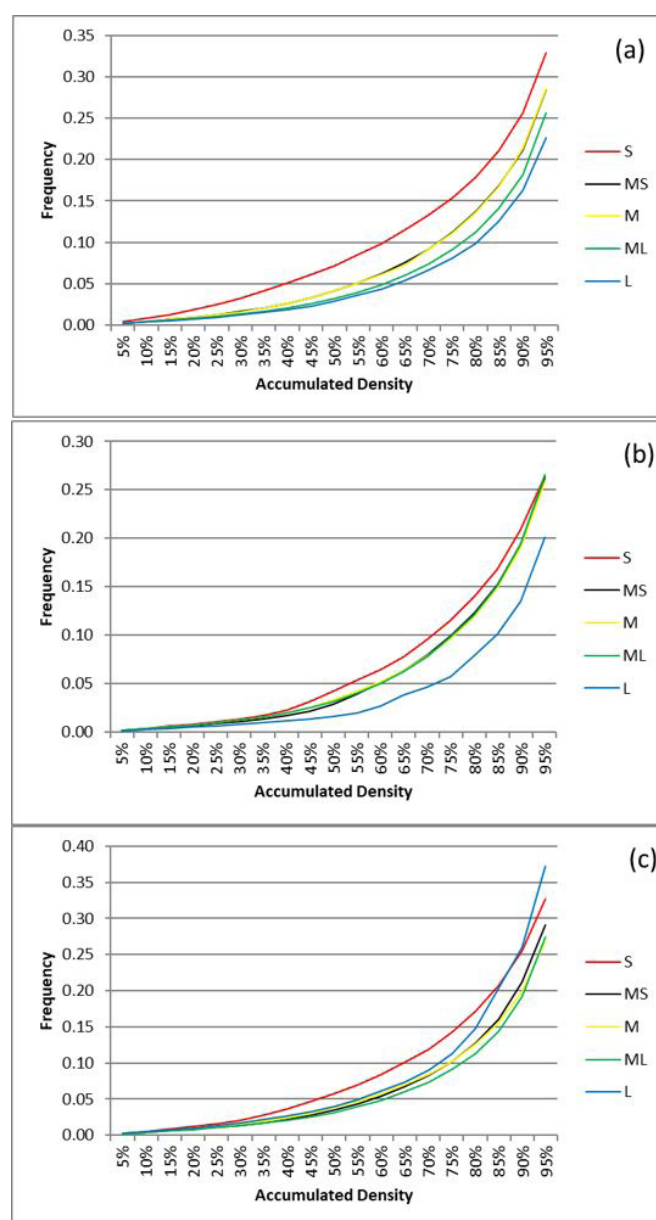


Figure 6. Accumulated frequency curves of each order for conductivity (a), DO (b), and pH (c). (S = small; MS = medium-small; M = medium; ML = medium-large; L = large).

values assigned to the orders after the mean comparisons for the conductivity (top), DO (center), and pH (bottom) parameter are reported in Figure 7.

In some cases, one order had its curve merged entirely with the other. This behavior is due to the indifference between the averages of some orders up to certain accumulated frequencies. In other cases, the values assigned to the orders were the same until a certain accumulated frequency, differing later.

Therefore, by considering the statistical analysis of comparison between means of different orders, the accumulated frequency density curves obtained after the adjustment suggested by ANOVA and Welch (1951) tests should be used. However, when applying these results to establish a monitoring strategy, the decision maker should be aware that, depending on the cumulative

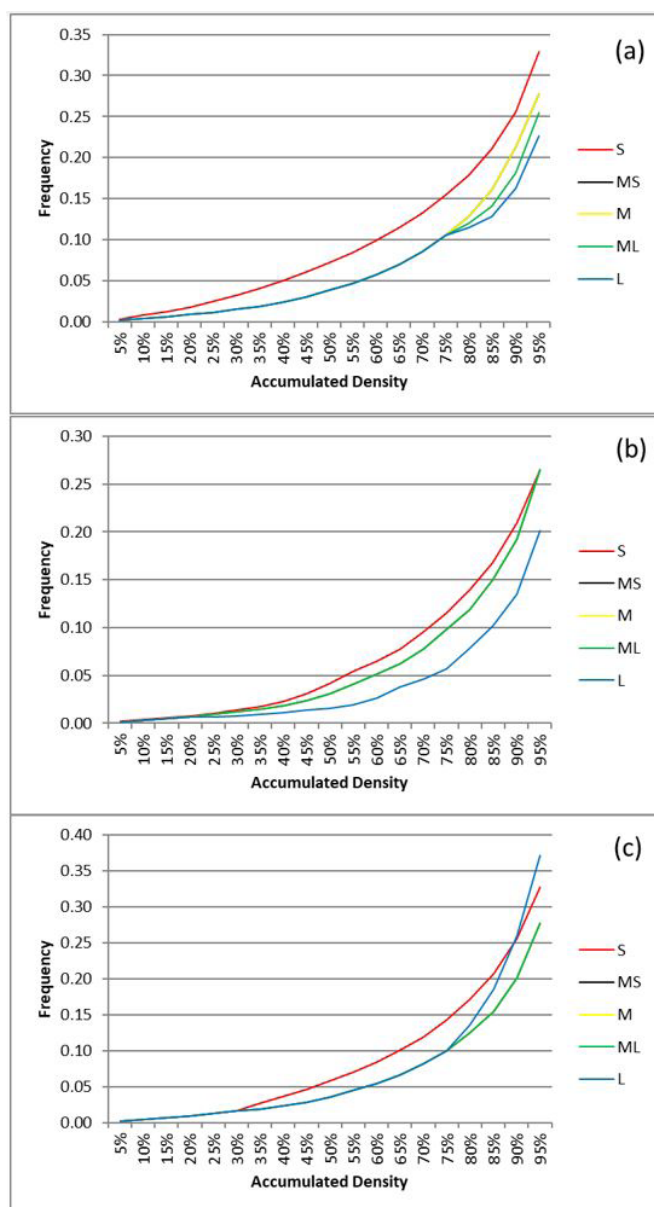


Figure 7. Accumulated frequency curves of each order for conductivity (a), DO (b), and pH (c). (S = small; MS = medium-small; M = medium; ML = medium-large; L = large).

frequency that he or she determines is ideal for the program, some aspects will be well addressed and others not.

For better characterization of the series and monitoring long and short-term trends, the NRTWQM strategy applies perfectly, as very high-accumulated frequencies are not required for the most frequent frequencies in the series to be considered, or with higher densities in the spectral analysis. This makes it possible to use the tool with a control instrument in the management of grants, for example, to monitor the evolution of water quality according to certain releases or withdrawals of flows, or even for monitoring conditions of an environmental license, in the case of potentially polluting enterprises.

Nevertheless, when determining an accumulated frequency, it is necessary to stop monitoring all frequencies higher than the frequency related to the accumulated frequency. In addition, higher frequencies, such as those near 100% accumulated frequency, represent sudden and rapid variations in water quality, which are often linked to accidental events or intentional spills of large unauthorized volumes. This means that the NRTWQM strategy cannot be applied as a Water Quality Alert System, otherwise it will be ineffective. This non-application of NRTWQM extends to monitoring programs aimed at tracking intraday variations. However, a great potential of the strategy for the improvement of water quality series can be observed in order to better subsidize framing, granting or licensing studies.

Determination of sampling intervals

With the frequencies for the different catchment types established, which sampling interval is required to ensure a representation of the associated frequency remained to be determined.

The accumulated frequencies and respective sampling intervals for the conductivity parameter organized by catchment order is shown in Table 2. By comparing the accumulated frequencies with the conventional monitoring used today in Brazil in which three to four samplings are performed per year, this method may guarantee an accumulated frequency between 10 and 15% for the conductivity parameter.

It is also possible to see that, for a cumulative frequency of 95%, the NRTWQM is not a good strategy, because in order to guarantee this cumulative frequency, daily sampling is necessary, in which case the use of real-time monitoring is more interesting. For small catchments of up to 100 km², the cumulative frequency of 90% is no longer interesting from an operational point of view.

For the conductivity parameter, weekly sampling intervals ensure 65% of cumulative frequency for medium-small to large catchments, and around 50% for small. If two samples are taken per week, the accumulated frequencies are around 80%, except for small catchments that are between 65 and 70% accumulated frequency.

It is also important to mention that the sampling interval values were rounded down after calculating from frequencies, as it does not make sense to use day fractions in an NRTWQM strategy. Moreover, they were rounded down in order not to violate the Sampling Theorem because of rounding.

Table 2. Accumulated frequencies and their respective sampling intervals for conductivity.

Accumulated Frequency	Sampling Intervals by Order (Days) - conductivity				
	S	MS	M	ML	L
5%	135	252	252	252	252
10%	62	120	120	120	120
15%	39	77	77	77	77
20%	27	55	55	55	55
25%	20	42	42	42	42
30%	15	32	32	32	32
35%	12	25	25	25	25
40%	9	20	20	20	20
45%	8	16	16	16	16
50%	6	12	12	12	12
55%	5	10	10	10	10
60%	5	8	8	8	8
65%	4	7	7	7	7
70%	3	5	5	5	5
75%	3	4	4	4	4
80%	2	3	3	4	4
85%	2	3	3	3	3
90%	1	2	2	2	3
95%	1	1	1	1	2

S = small; MS = medium-small; M = medium; ML = medium-large; L = large.

For the DO parameter (Table 3), weekly sampling intervals guarantee 60% cumulative frequency for small catchments, between 65 and 70% for medium-small to medium-large orders and between 75 and 80% for large catchments. If two samples are taken per week, the cumulative frequencies between 75 and 80% for small catchments are reached, between 80 and 85% for medium-small to medium-large orders and up to 90% for large.

In a similar study about sampling frequencies and using the DO parameter for a catchment with an area of 10,128 km², Silva (2012) found a representativity of 85% for a 4-day sampling interval. To evaluate this representativeness, the author used the correlation coefficient between the original monitoring series and sampling subsets with frequencies lower than the daily one, unlike the present study that used the accumulated frequencies as an indication of representativeness of the sampling frequencies.

The sampling interval values for the pH parameter are shown in Table 4. Using weekly samplings has a cumulative frequency of 55% for small catchments and 65% for medium-small to large orders. If two samples are taken each week, the cumulative frequencies are between 70 and 75% for small catchments, between 80 and 85% for medium-small to medium-large orders, and between 75 and 80% for large catchments.

By synthesizing the research results, when opting for the use of near-real-time water quality monitoring strategies, the sampling intervals described in Table 5 were identified.

The monitoring stations of the 4 Brazilian catchments used in the study are installed in the Cotia River (240 km²), Piracicaba River (8,900km²), Alto Tietê (636 km²), and Medio Tietê (16,100 km²). As an example, to implement NRTWQM in these catchments, with a Dissolved Oxygen focus and 80% frequency representation, monitoring should have a sampling interval of 4 days as all of them are in the size range of 100 to 100,000 km². Other

Table 3. Cumulative frequencies and respective sampling intervals for DO.

Accumulated Frequency	Sampling Intervals by Order (Days) - DO				
	S	MS	M	ML	L
5%	276	317	317	317	518
10%	134	152	152	152	152
15%	86	99	99	99	99
20%	62	71	71	71	71
25%	47	53	53	53	78
30%	36	42	42	42	63
35%	28	33	33	33	52
40%	22	26	26	26	44
45%	16	21	21	21	37
50%	11	16	16	16	31
55%	9	12	12	12	25
60%	7	9	9	9	18
65%	6	8	8	8	13
70%	5	6	6	6	10
75%	4	5	5	5	8
80%	3	4	4	4	6
85%	2	3	3	3	4
90%	2	2	2	2	3
95%	1	1	1	1	2

S = small; MS = medium-small; M = medium; ML = medium-large; L = large.

Table 4. Accumulated frequencies and respective sampling intervals for pH.

Accumulated Frequency	Sampling intervals by order (days) - pH				
	S	MS	M	ML	L
5%	247	247	247	247	247
10%	117	117	117	117	117
15%	75	75	75	75	75
20%	53	53	53	53	53
25%	40	40	40	40	40
30%	31	31	31	31	31
35%	18	27	27	27	27
40%	13	21	21	21	21
45%	10	17	17	17	17
50%	8	14	14	14	14
55%	7	11	11	11	11
60%	5	9	9	9	9
65%	4	7	7	7	7
70%	4	6	6	6	6
75%	3	5	5	5	5
80%	2	4	4	4	3
85%	2	3	3	3	2
90%	1	2	2	2	1
95%	1	1	1	1	1

S = small; MS = medium-small; M = medium; ML = medium-large; L = large.

important catchments in Brazil, such as Rio dos Sinos (3,690 km²), Paraíba do Sul river (56,500 km²), and Rio Doce (83,400 km²), with weekly sampling, guarantee 65% representativeness for the three parameters analyzed in this study. When considering application to large Brazilian catchments such as the Xingu River (531,250 km²), the São Francisco River (641,000 km²), and the Amazon River

Table 5. Recommended sampling intervals for the NRTWQM strategy.

Parameter	Accumulated Frequency	Catchment Size		
		<100 km ²	100<>100,000 km ²	>100,000 km ²
Electric Conductivity	35%	12	25	25
	50%	6	12	12
	65%	4	7	7
	80%	2	3	4
Dissolved Oxygen	35%	28	33	52
	50%	11	16	31
	65%	6	8	13
	80%	3	4	6
pH	35%	18	27	27
	50%	8	14	14
	65%	4	7	7
	80%	2	4	3

(7,050,000 km²), 50% representativeness of the dissolved oxygen parameter can be obtained as a monthly sampling example.

Unlike the approach of this study, which used spectral analysis as input to verify the representativeness of a sampling frequency in a series, most studies that evaluate the effectiveness of sampling intervals use the comparison between the maximum errors in relation to the mean and the standard deviation of each type of sampling. According to recommendations by CCME (2015), this is one of the statistical ways of sampling frequency optimization.

Sanders et al. (1983) used this approach and concluded that the smaller the area of the catchment to be monitored, the higher the sampling frequency required, and in addition to the use of the area as a subsidy for choosing the frequency. The authors also defined two recommendation thresholds for sampling intervals based on hydrological regime, more specifically using an index resulting from the division between the mean maximum flow rate and the mean minimum flow rate.

Biswas and Lawrence (2013) recommend that it is possible to reduce sampling frequencies for monitoring stations where concentration is more important than load. They also cite the use of different sampling frequencies for tributaries and main channels, recommending the use of lower frequencies for tributaries and higher ones for the main channels. This final hypothesis goes in the opposite direction of the results obtained here, since smaller sample interval values were obtained for the same catchment for the small catchment than for the large ones. Although the authors may be referring to complementarity in monitoring between points within the same catchment, the reduction in the sampling frequency of tributaries may cause significant loss of representativeness of the monitoring program.

CONCLUSIONS

The following items can be concluded from this study:

- Spectral analysis of historical water quality data series can be used as input to create cumulative frequency density curves, which from a sampling point of view, translates into cumulative frequency curves. Using cumulative frequency

curves as an indicator of representativeness made it possible to determine the ability of a sampling frequency to store the characteristics of the original series of water quality data;

- There is a relationship between the sampling frequencies required for the different accumulated frequencies and physical attributes of the catchments. The relationship with the area is a good alternative because of the ease of obtaining this data, allowing the prescription of different sampling intervals for different types of catchments when no water quality data are available to guide the establishment of a sampling frequency;
- It was possible to obtain the sampling intervals from the sampling frequencies for the different accumulated densities. Sampling intervals were generally executable in the NRTWQM molds up to 90% frequency. For accumulated frequencies higher than 90%, the intervals approach the daily values, being more advisable to use real time strategies;
- With respect to catchment sizes, for NRTWQM to be used in catchments of less than 100 km², smaller cumulative frequencies should be chosen to ensure a sampling interval that is executable within the strategy proposal.

As recommendations for use of the strategy:

- It is valid to improve the technique, as it is useful for the optimization of water resource management, but as it is technically complex, it still requires great statistical knowledge, which is often not the case for the water resources manager;
- When choosing the strategy, the decision maker should be aware that frequencies higher than the chosen cumulative frequency will not be represented and may have important environmental meanings. An example of this is the use of the strategy in small catchments, where if the catchment concentration time is less than the maximum sampling frequency, no important natural runoff processes that influence water quality will be detected;
- The objectives of the monitoring program in which the NRTWQM strategy will be used must be well defined,

and the sampling interval should ensure an accumulated frequency that meets these objectives;

- When defining the cumulative frequency that meets the objectives of the monitoring program, the sampling interval chosen should be that of the quality parameter that is most important for that monitoring program, as it is not feasible to perform a sampling interval for each parameter;
- Ranges should also be approximated to operational realities, for example, instead of using an 8-day interval, use the more convenient 7-day (weekly) interval.

REFERENCES

- ANA – AGÊNCIA NACIONAL DAS ÁGUAS. *Relatório de conjuntura dos recursos hídricos no Brasil*. Brasília: ANA, 2017.
- BISWAS, T.; LAWRENCE, B. *Revision of the River Murray Water Quality Monitoring Program*. Queensland: Murray-Darling Basin Authority, 2013.
- CARVALHO, K. Q.; LIMA, S. B.; PASSIG, F. H.; GUSMÃO, L. K.; SOUZA, D. C.; KREUTZ, C.; BELINI, A. D.; ARANTES, E. J. Influence of urban area on the water quality of the Campo river basin, Paraná state, Brazil. *Brazilian Journal of Biology = Revista Brasileira de Biologia*, v. 75, n. 4, p. 96-106, 2015. Supplement 2. <http://dx.doi.org/10.1590/1519-6984.00413suppl>. PMID:26628235.
- CCME – CANADIAN COUNCIL OF MINISTERS OF THE ENVIRONMENT. *Guidance manual for optimizing Water Quality Monitoring Program design*. Quebec: Canadian Council of Ministers of the Environment, 2015.
- DAMASCENO, M. C. S.; RIBEIRO, H. M. C.; TAKIYAMA, L. R.; PAULA, M. T. Avaliação sazonal da qualidade das águas superficiais do rio Amazonas na orla da cidade de Macapá, Amapá, Brasil. *Revista Ambiente & Água*, v. 10, n. 3, p. 598-613, 2015. <http://dx.doi.org/10.4136/ambi-agua.1606>.
- DONG, Z.; LI, F.; BEHESHTI, B.; MICKELSON, A.; PANERO, M.; ANID, N. Autonomous real-time water quality sensing as an alternative to conventional monitoring to improve the detection of food, energy and water indicators. *Journal of Environmental Studies and Sciences*, v. 6, n. 1, p. 200-207, 2016. <http://dx.doi.org/10.1007/s13412-016-0383-8>.
- FENG, X.; KIRCHNER, J. W.; NEAL, C. Spectral analysis of chemical time series from long-term catchment monitoring studies: hydrochemical insights and data requirements. In: WIEDER R. K.; NOVÁK M.; VILE, M. A. (Org.). *Biogeochemical investigations of terrestrial, freshwater, and wetland ecosystems across the globe*. Dordrecht: Springer, 2004. http://dx.doi.org/10.1007/978-94-007-0952-2_16.
- GIRARDI, R.; PINHEIRO, A.; TORRES, E.; KAUFMANN, V.; GARBOSSA, L. H. P. Evolution of physicochemical species concentration in stream based on heavy rainfall event data obtained for high-frequency monitoring. *Brazilian Journal of Water Resources*, v. 21, n. 4, p. 653-665, 2016. <http://dx.doi.org/10.1590/2318-0331.011616055>.
- HANISCH, W.; FREIRE-NORDI, C. S. Monitoramento remoto em tempo real de mananciais visando as florações de cianobactérias. In: POMPÊO, M.; MOSCHINI-CARLOS, V.; NISHIMURA, P. Y.; SILVA, S. C.; DOVAL, J. C. L. (Org.). *Ecologia de reservatórios e interfaces*. São Paulo: Instituto de Biociências, Universidade de São Paulo, 2015.
- HARRIS, J. On the use of Windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, v. 66, n. 1, p. 51-83, 1978. <http://dx.doi.org/10.1109/PROC.1978.10837>.
- LINKLATER, N.; ORMECI, B. Real-time and near real-time monitoring options for water quality. In: AHUJA, S. (Org.). *Monitoring water quality pollution assessment, analysis, and remediation*. Oxford: Elsevier, 2013. p. 189-225.
- MENEZES, M. L.; SOUZA, R. C.; PESSANHA, J. F. M. Combining singular spectrum analysis and PAR(p) structures to model wind speed time series. *Journal of Systems Science*, v. 27, n. 1, p. 29-46, 2014. <http://dx.doi.org/10.1007/s11424-014-3301-8>.
- MEYER, A. M.; KLEIN, C.; FUNFROCKEN, E.; KAUTENBURGER, R.; BECK, H. P. Real-time monitoring of water quality to identify pollution pathways in small and middle scale rivers. *The Science of the Total Environment*, v. 651, n. Pt 2, p. 2323-2333, 2019. <http://dx.doi.org/10.1016/j.scitotenv.2018.10.069>. PMID:30332665.
- NEEFF, T.; DUTRA, L. V.; SANTOS, J. R.; FREITAS, C. C.; ARAUJO, L. S. Power spectrum analysis of SAR data for spatial forest characterization in Amazônia. *International Journal of Remote Sensing*, v. 26, n. 13, p. 2851-2864, 2005. <http://dx.doi.org/10.1080/01431160500104301>.
- OPPENHEIM, A. V.; SCHAFER, R. W.; BUCK, J. R. *Discrete-time signal processing*. Upper Saddle River: Prentice Hall, 1999.
- OPPENHEIM, A. V.; WILLSKY, A. S.; YOUNG, I. T. *Signals and systems*. Englewood Cliffs: Prentice-Hall, 1983.
- SANDERS, T. G.; WARD, R. C.; LOFTIS, J. C.; STEELE, T. D.; ADRIAN, D. D.; YEVJEVICH, V. *Design of networks for monitoring water quality*. Littleton: Water Resources Publications, 1983.
- SHANNON, C. E. Communication in the presence of noise. *Proceedings of the Institute of Radio Engineers*, v. 37, n. 1, p. 10-21, 1949.
- SILVA, R. L. L. *Estratégia de monitoramento em apoio ao licenciamento ambiental em trecho de vazão reduzida*. 2012. 150 f. Dissertação (Mestrado em Engenharia Civil) - Universidade Federal de Santa Maria, Santa Maria, 2012.
- SILVA, R. L. L. *Monitoramento da qualidade da água em tempo quase-real: uma alternativa para a gestão de recursos hídricos*. 2018. 192 f.

Tese (Doutorado em Recursos Hídricos e Saneamento Ambiental)
- Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

WELCH, B. L. On the comparison of several mean values: an alternative approach. *Biometrika*, v. 38, n. 3-4, p. 330, 1951. <http://dx.doi.org/10.2307/2332579>.

ZAMYADI, A.; CHOO, F.; NEWCOMBE, G.; STUETZ, R.; HENDERSON, R. K. A review of monitoring technologies for real-time management of cyanobacteria: recent advances and future direction. *Trends in Analytical Chemistry*, v. 85, p. 83-96, 2016. <http://dx.doi.org/10.1016/j.trac.2016.06.023>.

Authors contributions

Régis Leandro Lopes da Silva: Conception, literature review, development, analysis and discussion of the results, and writing of the paper.

André Luiz Lopes da Silveira: Analysis and discussion of the results and writing of the paper.

Geraldo Lopes da Silveira: Conception, analysis, and discussion of results.