REVISTA ÁGUAS SUBTERRÂNEAS

Artigos

Comparison of methods to assess the accuracy of the incorporation of censored chemical data in descriptive statistical analysis of contaminated groundwater

Comparação de métodos para avaliar a acurácia da incorporação de dados químicos censurados na análise estatística descritiva de águas subterrâneas contaminadas

Vinicius Rodrigues dos Santos[1]; Luis de Almeida Prado Bacellar[1]; Cícero Antônio Antunes Catapreta[1]

[1] Universidade Federal de Ouro Preto (UFOP), Ouro Preto, MG, Brasil.

✉ vinicius.santos@aluno.ufop.edu.br, bacellar@ufop.edu.br, catapret@pbh.gov.br

Keywords:

Censored data;
Hydrochemistry;
Detection limit;
Descriptive statistics.

Abstract

Chemical analyses of groundwater often present data sets with censored values, i.e., below the detection limit (LOD). When the proportion of censored values is significant, descriptive (mean, median and standard deviation) or exploratory geochemical analysis may be impaired. Ignoring such data or replacing them with some predetermined value is not always the recommended alternative. Thus, the objective of this research is to investigate the applicability of four methods in estimating censored chemical data from an area with contaminated groundwater. Three statistical methods were used: parametric (Maximum Likelihood Estimation, MLE), non-parametric (Kaplan-Meier, KM) and robust (Order Regression Methods, ROS), in addition to the traditional method of direct replacement of censored data, using LOD/2. The MLE, assuming a Gaussian distribution of the data (MLE-no), yielded allowable substitution factors, close to 0.5, similarly to the traditional substitution method (LOD/2). Validation with complete datasets with the same estimation methods and considering three artificial LOD attested to the good results of MLE-no and ROS with 25% and 50% of censored data, respectively, as well as LOD/2. The first two methods are preferable to LOD/2 as they are statistically based. It is recommended in future studies that such estimation methods be combined with other geostatistical treatments to improve the spatial analysis of hydrochemical datasets.

Palavras-chave:

Dados censurados;
Hidroquímica;
Limite de detecção;
Estatística descritiva.

Revisão por pares.
Recebido em: 31/12/2021.
Aprovado em: 19/12/2022.

Resumo

Análises químicas da água subterrânea frequentemente apresentam conjuntos de dados com valores censurados, ou seja, abaixo do limite de detecção (LD). Quando a proporção de valores censurados é significativa, a análise geoquímica descritiva (média, mediana e desvio padrão) ou a exploratória podem ser prejudicadas e ignorar tais dados ou substituí-los por algum valor pré-determinado nem sempre é a alternativa recomendável. Assim, neste trabalho objetiva-se investigar a aplicabilidade de quatro métodos na estimativa de dados químicos censurados de uma área com águas subterrâneas contaminadas. Foram empregados três métodos estatísticos: paramétrico (Estimativa de Máxima Verossimilhança, MLE), não paramétrico (Kaplan-Meier, KM) e robusto (Métodos de Regressão na Ordem, ROS), além do método tradicional de substituição direta de dados censurados, utilizando o LD/2. O MLE, admitindo uma distribuição gaussiana dos dados (MLE-no), rendeu fatores de substituição admissíveis, próximos a 0,5, à semelhança do método tradicional de substituição (LD/2). A validação com conjuntos de dados completos com os mesmos métodos de estimativa e considerando três LD artificiais atestou os bons resultados de MLE-no e ROS com 25% e 50% de dados censurados, respectivamente, bem como de LD/2. Acredita-se que os dois primeiros métodos sejam preferíveis ao LD/2, por serem estatisticamente baseados. Recomenda-se em estudos futuros que tais métodos de estimativa sejam combinados com outros tratamentos geoestatísticos para melhorar a análise espacial de conjuntos de dados hidroquímicos.

DOI: http://doi.org/10.14295/ras.v37i1.30104

1. INTRODUCTION

Templ *et al.* (2008) define censored data as values of a measurement that are below or above the limit of detection (LOD) of an equipment or analytical method. Groundwater contamination studies usually have to deal with chemical data censored to the left, that is, values below the LOD. The proportion of samples containing censored data in a data set can affect simple calculations, such as descriptive statistics (mean, median and standard deviation) and even exploratory data analysis. According to Helsel (2006), the most unfavorable procedure of treating censored data is excluding them from the analysis, as bias is introduced in the descriptive statistics, *e.g.* in the mean and median values. Some authors (HORNUNG; REED, 1990;

TEMPL *et al*., 2008) recommend replacing the censored data with the direct substitution (DS) method, usually with LOD/2, but only when there are a significant number of uncensored values (at least 75%), which usually is impractical in many contaminated sites. The analysis of groundwater contamination depends on complete information, which justifies the acquisition of appropriate values to substitute for censored data.

Three types of statistical data analysis for the estimation of censored data sets have been applied as alternative methods to DS: linear regression, parametric and non-parametric methods. Usually, these methods are employed in controlled chemical data sets created by computer simulation, provided that the censoring value is known (LIU *et al*, 1997; CLARKE, 1998; HEWETT; GANSER, 2007; ANTWEILER; TAYLOR, 2008; ANTWEILER, 2015). On the other hand, when treating environmental analytical datasets containing non-simulated censored values, as in the case of present study, it is difficult to assess the accuracy of each method.

Many authors have used a variety of methods in order to incorporate censored data in the descriptive statistical analysis (ANTWEILER; TAYLOR, 2008; ANTWEILER, 2015; BACCARELLI *et al*., 2005; LEE; HELSEL, 2005; FIÉVET; VEDOVA, 2010; CARRANZA, 2011). Some authors recommend Regression in Order (ROS) (LEVITAN *et al*., 2014; SINGH; NONCERINO, 2002), whereas others Maximum Likelihood Estimation (MLE) (GILLIOM; HELSEL, 1986; GIBBONS, 2001. LEITH *et al.* (2010) stated that DS method are equivalent to more complex ones, such as Kaplan-Meier (KM) and MLE, when the proportion of censored data is low.

Helsel (2005), in turn, recommends that: (i) KM should be chosen when the proportion of censored data is less than 50% of the data; (ii) MLE, ROS or DS should be employed when censored data represent 50 to 80% of small sample (n <50), and MLE or DS when the sample is large (n> 50); (iii) When the censored data proportion is higher than 80%, no method could be employed.

That is, if on one hand there is the recommendation to not ignore censored data (HELSEL, 2006), on the other hand there is no consensus on which method should be more appropriate to incorporate them. Therefore, the main objective of this study is to analyze four (parametric, non-parametric, robust, and DS) methods of incorporating censored data in the geochemical database of a real contamination site. The database results from the quarterly monitoring of tens of groundwater monitoring wells of a solid waste landfill (CTRS-BR040) area in Belo Horizonte (MG) (19º54'58"S, 44º00'54"W), SE Brazil. This is one of the largest waste landfills in Brazil, which operated until 2005, receiving a variety of domestic, industrial, public and healthcare solid waste (BARELLA *et al*., 2013). The groundwater in the landfill area of influence has shown evidence of chemical contamination since 2005 (BARELLA *et al*., 2013). The solid waste was initially unselectively disposed, which makes it difficult to envisage the source of the three contamination plumes identified in the area (BACELLAR; OLIVEIRA FILHO, 2009).

2. THEORY

In the DS Method, arbitrary values for certain variables can substitute the censored values, either by LOD or a fraction of it, such as LOD/2 or LOD/√2 (HEWETT; GANSER, 2007). This method is widely used, although a scientific basis that justifies such application is still lacking. Some researchers encourage the DS method only when censored data represent less than 50% (HORNUNG; REED, 1990) or 25% (REIMANN, 2008; TEMPL; REIMANN, 2008) of a population.

KM is a non-parametric estimator that does not require the previous knowledge of the data distribution shape (YOUNG *et al*., 1999) and that is based on the Bayes' multiplication rule (KAPLAN; MEIER, 1958). It was initially developed to estimate the survival curve of right-censored data in Medical Sciences and was later adapted by Helsel (2011) to left-censored data sets. According to Young *et al.* (1999), the survival analysis is originally used to analyze data representing the time for an event of interest. When analyzing censored data, it is considered that the probability of a censored value be estimated above its censoring limit S(X) equals the probability of the censored value X to have an estimated value less than its limit, named A, and the probability of the censoring value X be equal to its limit, named B. This is the survival function:

$$S(X) = P(A \cap B) = P(A)P(A|B) \quad (1)$$

The KM method allows to estimate the proportion of concentrations below each observed level, classifying the sample values and yielding the probability of the censored value to be below the limit i. Equation (2) shows the probability of a parameter X to be less than the threshold x_i.

$$S(X < x_i) = \prod_{j=1}^{k} \frac{b_j - d_j}{b_j} \quad (2)$$

where b_j and d_j are, respectively, the number of uncensored and censored data above limit i (HELSEL, 2011).

Thus, the estimate of the mean of a data set containing censored values ($\hat{\mu}$) is obtained by:

$$\hat{\mu} = \sum_{i=1}^{m} x_i [S(X < x_i) - S(X < x_{i-1})] \quad (3)$$

The MLE method (FISHER, 1925) is based on the supposition of the knowledge of the probability density function (pdf) for the phenomena under investigation. It consists of the iterative optimization of the likelihood (L) function for each limit i:

$$L(x,,(\hat{\mu},\hat{\sigma})|i) = k^{-1}\sum_{j=1}^{k}\log\left(pdf\left(x_j,(\hat{\mu},\hat{\sigma})|i\right)\right) \tag{4}$$

where:

k = number of data above a certain limit

x = uncensored values above a certain limit

$\hat{\sigma}$ = estimated standard deviation

As the MLE overestimates mean values and underestimates variance, Cohen (1961) proposed the correction of the results with a linear function factor (λ). The mean corresponding to the uncensored values (ῃ)

$$\mu c = \hat{\mu} - \lambda(\hat{\mu} - x_c) \tag{5}$$

where x_c is the value of the censored variable to be estimated. Small data sets may not yield reliable results with MLE estimators, once it is difficult to identify the previous data set distribution.

The ROS method estimates censored data by least squares regression for logarithms of data *versus* their normal scores. The method assumes that the uncensored data fit normal or lognormal distributions. After adjusting a regression equation with the uncensored observations in a probability graph, the values for the individual censored observations are predicted from the regression model based on their normal scores (HELSEL, 2011; SINGH; NOCERINO, 2002). According to Helsel (2011), the probability to exceed the reported limit is:

$$pe_j = pe_{j+1} + \frac{A_j}{A_j + B_j}\left[1 - pe_{j+1}\right] \tag{6}$$

where A_j is the number of observations detected between the threshold of *j* and (*j* + 1) and B_j is the number of observations of censored and uncensored data below limit *j*.

3. MATERIALS AND METHODS

The adopted geochemical dataset derives from quarterly groundwater monitoring campaigns that were caried out from 2010 to 2015 in the influence area of the CTRS-BR040 landfill. The campaign code that was used is “Cn_year”, where the subscript “n” refers to the trimester of sampling (n=1, 2, 3 or 4) in a certain year. Due to management problems, only data from 20 campaigns of the years 2014 and 2015 are available. Following Helsel (2005) recommendations, only data sets of physical-chemical and chemical parameters with less than 80% censored values were used, resulting in 26 parameters.

Probability graphs were used to assess the parameter distributions and possible outliers and to evaluate if the distribution is normal (THODE, 2002).

Two scenarios were considered for censored data analysis and summary statistics calculations. The Scenario 1 is made up of chemical analyses of each monitoring campaign, which contain a maximum of 80% censored data. In the Scenario 2, only monitoring campaign samples with no censored data were used. In this scenario, the censoring limits were artificially adjusted, choosing a value within the dataset that generates new sets of data with approximately 25%, 50% and 75% of censored data. This was done in order to assess the best methods and validate the analysis of Scenario 1.

The 20 campaigns were assessed independently in both scenarios, once along the years the groundwater samples were analyzed in different laboratories, with different LOD values. The mean for each parameter were calculated applying the three methods (ROS, KM and MLE), assuming for the MLE method two types of data distribution (normal and lognormal). All calculations were made using the NADA package (LEE, 2010) of software R.

These methods, however, did not attribute a substitution value for the censored data. As proposed by Sanford *et al.* (1993), a substitution factor (r_x), which is a LOD multiplying fraction, can be used as a substitution value to estimate the mean:

$$r_x = (\mu c)/LOD_j \tag{7}$$

where μ_c is the estimated mean for the censored data, and LOD_j is the limit of detection for a certain variable j.

As the censored data of Scenario 1 are hidden, that is, the real value of a result below LOD is never known, Equation 7 can be used to assess the performance of the statistical methods (KM, ROS and MLE). It is worth mentioning that r_x should be between 0 and 1, as it is a multiplying fraction. Therefore, if r_x from a certain variable j fall outside the 0<rx<1 interval, its respective mean estimated by any of the methods will not be valid.

The substitution factor r_x can be obtained from the weighted mean of the censored and uncensored (SANFORD *et al.*,1993):

$$n\hat{\mu} = n_u\mu_u + n_c\mu_c \tag{8}$$

which can be re-written as:

$$\mu c = (n\hat{\mu} - n_u\mu_u)/n_c \tag{9}$$

where:

$\hat{\mu}$ = estimated mean for the whole dataset (censored and uncensored);

n = total number of samples;

n_c = number of censored data;

μ_u = mean for the uncensored data;

n_u = number of uncensored data.

The validation of each method in Scenario 2 was performed calculating the bias, which is the difference between the real (μ_r) and the estimated mean ($\hat{\mu}$). In Scenario 2 the mean values of each dataset (censored and uncensored) are known previously to the insertion of the artificial censoring limits. Thus, it is possible to compare the results of the four methods in this scenario (MLE, KM, ROS and DS). The results of Scenario 2 aid the interpretation of the results of Scenario 1, helping to select the most adequate method substitution for censored values.

4. RESULTS

Table 1 presents the total number of 27 chemical and physical-chemical parameters (variables) from the 20 groundwater sampling campaigns, the number corresponding to the censored and uncensored data (Scenario 1) and the maximum and minimum concentrations for each variable. The minimum concentration value refers to the least LOD value obtained for each Variable.

Table 1 – Statistical summaries of the groundwater parameters in the 20 quarterly monitoring campaigns in scenario 1

Parameters	Total number of data	Censored data number	Number of uncensored data number	Censored data (%)	Minimum concentration (mg/L)	Maximum concentration (mg/L)
Aluminium(s)	216	97	119	44,91	0,005	9,5
Aluminium(t)	641	255	386	39,78	0,001	14
Barium(s)	280	102	178	36,43	0,002	7,2
Barium(t)	73	27	46	36,99	0,0025	7,2
Lead(s)	72	38	34	52,78	0,003	0,15
Lead(t)	220	135	85	61,36	0,003	3,1
Cloride	109	36	73	33,03	0,001	357
Cobalt(t)	72	47	25	65,28	0,002	0,089
Copper(t)	36	17	19	47,22	0,003	0,088
BOD	535	303	232	56,64	0,24	245
COD	504	325	179	64,48	0,01	393
Iron(s)	570	257	313	45,09	0,001	118
Iron(t)	356	53	303	14,89	0,002	211
Fluoride	281	92	189	32,74	0,003	19
Mercury(s)	36	26	10	72,22	0,00002	0,0036
Mercury(t)	36	25	11	69,44	0,017	0,26
Manganese(s)	568	211	357	37,15	0,00002	95
Manganese(t)	603	144	459	23,88	0,002	12
Nickel(t)	37	16	21	43,24	0,00002	0,087
Nitrate	504	89	415	17,66	0,005	112
Nitrite	243	171	72	70,37	0,008	0,98
Oils and Grease	217	89	128	41,01	0,01	44
Sulfate	320	48	272	15	0,008	288
Hydrogen Sulfide	37	7	30	18,92	0,002	0,011
Surfactantes	37	15	22	40,54	0,02	0,12
Zinc(t)	282	97	185	34,4	0,005	50

s – soluble; t – total

An estimated mean ($\hat{\mu}$) for all data from all 20 monitoring campaigns in Scenario 1 was calculated. To optimize space, only the results of 6 campaigns is presented here (Table 2). The mean values ($\hat{\mu}$) estimated with the KM, ROS and MLE present less variability, when compared to those obtained with MLE-log. However, as it will be discussed later, even with relatively close mean values, this little variability can yield large differences in substitution factor.

The substitution factors or replacement factor (r_x) for 10 groundwater monitoring campaigns in Scenario 1 (Table 3) were calculated with Equations 9 and 7. The MLE-no was the sole method that resulted in an admissible r_x (between 0 and 1) for 100% of the cases. On the other hand, circa 30% of the r_x calculated from the mean with ROS are inconsistent.

There is a notable difference between r_x values obtained by MLE-no and ROS. In fact, r_x obtained by MLE-no showed little variance, with values close to 0.5. In turn, r_x obtained by ROS varied from 0.01 to 0.90, within consistent results. The results with ROS showed larger variance because it considers the proportion and magnitude of the uncensored data. That is, depending on the amount of uncensored data in a sample and whether these values are of magnitude much higher than LOD, the tendency is a super estimation of the mean, resulting in r_x greater than 1.

The r_x calculated with KM and MLE-log were consistent for only 30% and 11%, respectively. Thus, the MLE-no and ROS methods resulted in most coherent mean estimates for Scenario 1, when validated by the substitution factor.

Table 2 – Groundwater parameters for 6 sampling campaigns and their respective mean values (μ̂) with KM, ROS and MLE with normal (MLE-no) and lognormal (MLE-log) distribution for scenario 1.

Campaing	Parameter	KM mean	ROS mean	MLE-log mean	MLE-no mean
C1_2010	Aluminum(s)	0,3640	0,3566	3,2724	0,3654
	Aluminum(t)	1,0925	1,0859	1,5516	1,1012
	BOD	1,5550	1,5578	2,0479	1,5929
	COD	13,6111	8,5903	21,5525	10,2781
	Sol. Iron	0,7778	0,7757	15,6197	0,7767
	Manganese(s)	3,3769	3,3764	11,6286	3,3767
	Manganese(t)	1,0900	1,0899	2,1683	1,0899
	Nitrates	7,4762	7,4743	1670,5836	7,4792
C3_2011	Aluminum(t)	0,8411	0,8359	4,7557	0,8359
	BOD	4,3235	3,1506	12,9036	3,2320
	COD	20,8108	13,3123	12,6510	15,7396
	Iron(t)	4,5634	4,5632	10,6810	4,5633
	Manganese(s)	0,8594	0,8592	2,3089	0,8593
	Nitrates	10,4673	10,4456	56,5715	10,4422
	Sulfides	4,4865	4,4860	9,2988	4,4859
C1_2012	Aluminum(t)	0,1198	0,1063	1,6160	0,1077
	Barium(s)	0,1779	0,1764	283,2785	0,1769
	Barium(t)	0,2051	0,2037	8037,4385	0,2026
	Sol. Lead(s)	0,0047	0,0041	0,0042	0,0040
	Lead(t)	0,0112	0,0111	0,0108	0,0110
	COD	21,4324	16,5954	14,6282	17,7351
	Iron(s)	0,1682	0,1652	2,0968	0,1666
	Iron(t)	2,2028	2,2007	2301,7277	2,2015
	Manganese(t)	0,6210	0,6128	51244,7000	0,6141
	Nitrates	9,5935	9,5882	367,5338	9,5845
	Sulfides	3,5944	3,5886	22,8132	3,5884

Campaing	Parameter	KM mean	ROS mean	MLE-log mean	MLE-no mean
C1_2012	Surfactants	0,0347	0,0316	0,0322	0,0311
C3_2013	Aluminum(t)	0,1400	0,1330	0,4209	0,1321
	BOD	4,0833	2,5068	14,7318	2,4714
	COD	17,0556	10,7058	10,9223	12,7923
	Iron(t)	8,1780	8,1770	31,7987	8,1771
	Fluorides	0,0473	0,0461	0,0480	0,0462
	Manganese(t)	0,5273	0,5288	232,9152	0,5225
	Nitrates	0,0381	0,0354	0,0539	0,0361
	Oils	1,4639	1,1741	1,1726	1,1509
	Zinc(t)	0,0411	0,0381	0,0337	0,0387
C3_2014	Aluminum(t)	0,1471	0,1425	1,6750	0,1408
	Color	53,9706	52,1862	153,4956	52,7257
	BOD	8,6559	7,0210	26,7929	6,8031
	COD	25,7059	20,2868	18,0120	21,6147
	Iron(s)	0,2330	0,2118	7,8219E+05	0,2109
	Iron(t)	7,0411	7,0405	30,9818	7,0405
	Manganese(s)	0,7197	0,7222	1,3491E+06	0,7138
	Manganese(t)	0,9575	0,9553	872,7964	0,9528
	Nitrites	0,0922	0,0831	2,1276	0,0845
C1_2015	Aluminum(t)	0,1263	0,1197	2,0917	0,1195
	Barium(s)	0,2476	0,2020	2,0887E+05	0,1968
	Iron(t)	4,0378	4,0355	324,4090	4,0351
	Manganese(t)	0,6149	0,6116	32,3452	0,6101
	Nitrites	0,0097	0,0046	0,0056	0,0061
	Sulfides	3,3201	3,3194	12,8598	3,3187

s – soluble; t – total; units: mg/L

Table 3 – Groundwater parameters and their respective replacement factor (r_x) with means calculates by KM, ROS, MLE-no and MLE-log.

Campaing	Parameter	KM r_x	ROS r_x	MLE-log r_x	MLE-no r_x
C1_2010	Aluminum(s)	0,42	0,01	161,50	0,50
	Aluminum(t)	0,32	0,13	14,00	0,58
	BOD	0,28	0,29	2,66	0,46
	COD	1,04	0,26	2,29	0,52
	Iron(s)	1,00	0,08	6,29E+03	0,50
	Manganese(s)	0,67	0,34	5,40E+03	0,50
	Manganese(t)	1,00	0,26	7,76E+03	0,50
	Nitrates	0,30	0,18	1,11E+05	0,50
C3_2011	Aluminum(t)	1,86	0,50	1,04E+03	0,50
	BOD	2,30	0,37	16,41	0,50
	COD	1,20	0,17	0,08	0,51
	Iron(t)	1,40	0,15	4,53E+04	0,50
	Manganese(s)	1,00	0,15	5,36E+03	0,50
	Nitrates	16,00	2,61	2,84E+04	0,50
	Sulfides	1,00	0,58	4452,38	0,50
C1_2012	Aluminum(t)	2,10	0,32	199,82	0,50
	Barium(s)	1,00	0,29	1,40E+05	0,50
	Barium(t)	1,67	1,01	3,81E+06	0,50
	Sol. Lead	1,00	0,62	0,66	0,57
	Lead(t)	1,00	0,73	-0,05	0,52
	COD	1,10	0,32	0,01	0,51
	Sol. Iron	1,00	0,06	595,63	0,50
	Iron(t)	1,00	0,19	8,96E+05	0,50
	Manganese(t)	2,40	0,14	1,40E+07	0,50
	Nitrates	4,20	2,01	1,47E+05	0,50
	Sulfides	1,60	0,52	3557,08	0,50
C1_2012	Surfactants	1,00	0,61	0,69	0,56
C3_2013	Aluminum(t)	2,00	0,67	55,24	0,50
	BOD	4,80	0,60	33,20	0,50
	COD	1,01	0,10	0,13	0,40
	Iron(t)	4,00	0,25	8,50E+04	0,50
	Fluorides	1,25	0,45	1,70	0,51
	Manganese(t)	2,40	3,02	9,30E+04	0,50
	Nitrates	1,00	0,34	4,74	0,51
	Oils	1,10	0,60	0,60	0,56
	Zinc(t)	1,00	0,39	-0,49	0,51
C3_2014	Aluminum(t)	3,00	1,19	614,16	0,50
	Color	1,00	0,29	40,81	0,50
	BOD	6,80	1,24	68,47	0,50
	COD	1,20	0,28	-0,11	0,50
	Iron(s)	15,00	1,14	5,11E+08	0,50
	Iron(t)	6,00	0,50	2,04E+05	0,50
	Manganese(s)	5,50	7,65	1,15E+09	0,50
	Manganese(t)	6,50	3,63	1,11E+06	0,50
	Nitrites	1,87	0,27	362,32	0,50
C1_2015	Aluminum(t)	26,89	0,54	671,03	0,50
	Barium(s)	36,50	4,16	1,48E+08	0,50
	Iron(t)	5,00	1,19	5,45E+05	0,50
	Manganese(t)	8,00	2,89	4,90E+04	0,50
	Nitrites	1,12	0,31	0,47	0,55
	Sulfides	3,62	2,12	2,03E+04	0,50

(s) soluble; (t) total

Three parameters (iron, manganese and barium) were selected to compose the database in Scenario 2 among the 10 monitoring campaigns, because 100% of their original data were uncensored, that is above LOD. Therefore, artificial censoring limits of circa 25, 50 and 75%, were simulated. Table 4 presents, the artificial censoring limits, the proportions of censored data, and the maximum and minimum concentrations of the original samples.

In this scenario, besides the three methods focused on Scenario 1, the DS method (LOD/2) was also incorporated with the aim of better comparing the real mean of the data sets before the input of the artificial censoring limits.

The estimated $\hat{\mu}$, considering three proportions of censored data (25, 50 and 75%) in Scenario 2 (Table 5), indicate an increase in the estimated mean values as the proportion of censored data increases.

Table 4 – Maximum, minimum and percentage of censored data of groundwater chemical parameters in scenario 2

Censored Data Range	Campaing	Parameter	Total Number of Elements	Censorship Limit/ Minimum (mg/L)	Maximum (mg/L)	Percentage of Censored Data
25%	C4_2011	Barium(s)	36	0,10	2,20	27,78
		Barium(t)	36	0,10	4,10	22,22
		Chlorides	36	10	385	25,00
	C3_2012	Barium(t)	36	0,10	2	27,78
		Chlorides	36	15	395	25,00
		Total Iron	36	0,10	39	27,78
	C2_2013	Barium(t)	37	0,11	4,70	21,62
		Chlorides	37	11	370	24,32
		Sulfates	37	1,10	50	24,32
	C1_2014	Barium(t)	34	0,10	2,40	23,53
		Chlorides	34	20	330	26,47
		Sulfates	34	6	87	23,53
	C2_2014	Barium(t)	34	0,11	4,30	26,47
		Chlorides	34	10	306	23,53
	C1_2015	Barium(t)	34	0,10	2,60	26,47
		Chlorides	34	20	297	23,53
50%	C4_2011	Barium(s)	36	0,20	2,20	52,78
		Barium(t)	36	0,20	4,10	50,00
		Chlorides	36	31	385	47,22
	C3_2012	Barium(t)	36	0,20	2	52,78
		Chlorides	36	35	395	52,78
		Iron(t)	36	0,50	39	52,78
	C2_2013	Barium(t)	37	0,20	4,70	54,05
		Chlorides	37	30	370	48,65
		Sulfates	37	3	50	48,65
	C1_2014	Barium(t)	34	0,20	2,40	52,94
		Chlorides	34	35	330	50,00
		Sulfates	34	8,50	87	50,00
	C2_2014	Barium(t)	34	0,20	4,30	50,00
		Chlorides	34	30	306	47,06
	C1_2015	Barium(t)	34	0,20	2,60	52,94
		Chlorides	34	35	297	52,94
75%	C4_2011	Barium(s)	36	0,50	2,10	72,22
		Barium(t)	36	0,50	4,10	66,67
		Chlorides	36	70	385	77,78
	C3_2012	Barium(t)	36	0,80	2	77,78
		Chlorides	36	75	395	77,78
		Iron(t)	36	2,00	39	69,44
	C2_2013	Barium(t)	37	0,80	4,70	75,68
		Chlorides	37	90	370	78,38
		Sulfates	37	7	50	75,68
	C1_2014	Barium(t)	34	0,70	2,40	73,53
		Chlorides	34	70	330	70,59
		Sulfates	34	22,00	87	70,59
	C2_2014	Barium(t)	34	0,80	4,30	73,53
		Chlorides	34	70	306	70,59
	C1_2015	Barium(t)	34	0,40	2,60	79,41
		Chlorides	34	75	297	76,47

s – soluble; t - total

Table 5 – Estimated mean for ROS, KM, MLE and LOD/2 for scenario 2

Censored Data Range	Campaing	Parameter	KM	ROS	MLE-no	MLE-log	LOD/2
25%	C4_2011	Barium(s)	0,4758	0,4619	0,5067	0,4637	0,4546
		Barium(t)	0,5728	0,5574	0,5934	0,5607	0,5606
		Chlorides	70,2500	69,1107	78,9302	69,0146	68,8611
	C3_2012	Barium(t)	0,4525	0,4355	0,4980	0,4390	0,4386
		Chlorides	79,3333	76,9190	83,2387	76,9916	76,9583
		Iron(t)	4,8550	4,8341	17,7315	4,8411	4,8411
	C2_2013	Barium(t)	0,5584	0,5423	0,5586	0,5466	0,5465
		Chlorides	74,1892	73,0418	86,7736	72,8718	72,8514
		Sulfates	7,0892	6,9156	7,5603	6,9325	6,9311
	C1_2014	Barium(t)	0,4100	0,3957	0,4158	0,3985	0,3982
		Chlorides	74,9706	71,8399	74,9112	72,1340	72,0588
		Sulfates	20,5235	19,5420	20,2162	19,4481	19,4176
	C2_2014	Barium(t)	0,5421	0,5227	0,5421	0,5277	0,5275
		Chlorides	66,5882	65,4598	74,6305	65,1941	65,1765
	C1_2015	Barium(t)	0,4071	0,3881	0,3893	0,3914	0,3912
		Chlorides	70,8529	68,1006	68,8713	68,5693	68,5000
50%	C4_2011	Barium(s)	0,5217	0,4683	0,5513	0,4709	0,4689
		Barium(t)	0,6200	0,5653	0,6658	0,5613	0,5628
		Chlorides	78,6667	68,5313	73,4444	71,1100	70,8750
	C3_2012	Barium(t)	0,5167	0,4602	0,5284	0,4504	0,4481
		Chlorides	91,4722	76,7436	85,9536	77,8202	77,4861
		Iron(t)	5,0422	4,9057	18,9901	4,9110	4,9103
	C2_2013	Barium(t)	0,6238	0,5520	0,6546	0,5438	0,5454
		Chlorides	82,5135	73,8684	83,5842	74,0262	73,7568
		Sulfates	7,9865	6,9309	7,4209	7,1297	7,1108
	C1_2014	Barium(t)	0,4550	0,4003	0,4418	0,4043	0,4021
		Chlorides	81,0588	70,8873	76,8686	72,1900	71,8088
		Sulfates	20,4706	19,5944	20,0914	19,4293	19,1471
	C2_2014	Barium(t)	0,5874	0,5221	0,5819	0,5334	0,5324
		Chlorides	74,6471	64,5538	69,9560	66,4521	66,1765
	C1_2015	Barium(t)	0,4594	0,3795	0,4040	0,3975	0,3959
		Chlorides	76,8529	65,5526	72,0854	67,9938	67,5882
75%	C4_2011	Barium(s)	0,7133	0,5638	0,5329	0,5482	0,5328
		Barium(t)	0,7844	0,5851	0,6164	0,6269	0,6247
		Chlorides	108,4722	73,1977	77,4498	78,1872	76,5833
	C3_2012	Barium(t)	1,0208	0,6019	0,6062	0,6420	0,6042
		Chlorides	113,8056	86,0520	84,9050	84,1357	82,3056
		Iron(t)	8,1944	5,9769	12,0082	5,2924	5,2778
	C2_2013	Barium(t)	1,0038	0,5254	0,6296	0,7048	0,6859
		Chlorides	123,7297	91,3123	78,7973	86,8185	83,7568
		Sulfates	13,4595	7,3407	7,5644	7,9146	7,7838
	C1_2014	Barium(t)	0,8521	0,5008	0,5682	0,5767	0,5432
		Chlorides	111,2941	73,7574	76,7564	79,2564	77,4118
		Sulfates	31,5882	21,6649	21,0745	21,8643	21,0000
	C2_2014	Barium(t)	1,0062	0,5440	0,6553	0,7173	0,6974
		Chlorides	99,6765	67,0329	72,7854	76,2184	74,2647
	C1_2015	Barium(t)	0,7803	0,3879	0,4117	0,4308	0,4229
		Chlorides	106,9706	71,3165	71,3400	76,7116	70,6471

s – soluble; t – total; unit: mg/L

The first validation of the mean values in Scenario 2, except DS method, was implemented by checking substitution factor (r_x), determined with Equations 9 and 7. The acceptable values were those with $0<r_x<1$ (Table 6).

The substitution factor (r_x) was greater than 1 for the majority of the variables with KM. These results apparently contradict Helsel (2005), who stated that the non-parametric models (KM) would be more adequate when the dataset contains less than 50% censored data.

The MLE-no and ROS methods presented 100% of consistent results. As expected, the MLE-log method was inconsistent, mainly with lower proportions of censored data. As in Scenario 1, the substitution factors obtained using mean values calculated by MLE-no were close to 0.5. Differently from Scenario 1, there is less variability in the r_x results obtained by ROS, which is related to less variability between the uncensored data and the artificial censoring limits.

A second validation of the results and the comparison of the accuracy between the methods was performed by calculating the bias of the mean values (Table 7). A bias close to 0 indicates better accuracy and aids the validation of the results obtained in Scenario 1.

Table 6 – Replacement factors (r_x) calculated from the means obtained by the ROS, KM, MLE-log and MLE-no in scenario 2

Censored Data Range	Campaing	Parameter	KM r_x	ROS r_x	MLE-log r_x	MLE-no r_x
25%	C4_2011	Barium(s)	0,91	0,45	1,92	0,51
		Barium(t)	1,00	0,37	1,85	0,51
		Chloride	1,00	0,54	4,47	0,51
	C3_2012	Barium(t)	1,00	0,39	2,64	0,51
		Chloride	1,13	0,49	2,17	0,51
		Iron(t)	1,00	0,25	464,56	0,50
	C2_2013	Barium(t)	1,00	0,32	1,01	0,51
		Chloride	1,00	0,57	5,70	0,51
		Sulfates	1,09	0,44	2,85	0,51
	C1_2014	Barium(t)	1,00	0,39	1,25	0,51
		Chloride	1,05	0,46	1,04	0,51
		Sulfates	1,28	0,59	1,07	0,52
	C2_2014	Barium(t)	1,00	0,34	1,00	0,51
		Chloride	1,10	0,62	4,52	0,51
	C1_2015	Barium(t)	-6,83	0,39	0,43	0,51
		Chloride	1,00	0,42	0,58	0,51
50%	C4_2011	Barium(s)	1,00	0,49	1,28	0,52
		Barium(t)	1,10	0,55	1,56	0,51
		Chloride	1,03	0,34	0,68	0,52
	C3_2012	Barium(t)	1,15	0,62	1,26	0,52
		Chloride	1,26	0,46	0,96	0,52
		Iron(t)	1,00	0,48	53,85	0,50
	C2_2013	Barium(t)	1,25	0,59	1,54	0,51
		Chloride	1,10	0,51	1,17	0,52
		Sulfates	1,10	0,38	0,71	0,51
	C1_2014	Barium(t)	1,00	0,48	0,88	0,52
		Chloride	1,03	0,45	0,79	0,52
		Sulfates	0,71	0,56	0,65	0,53
	C2_2014	Barium(t)	1,05	0,40	1,00	0,51
		Chloride	1,10	0,39	0,77	0,52
	C1_2015	Barium(t)	1,10	0,35	0,58	0,52
		Chloride	1,00	0,39	0,74	0,52
75%	C4_2011	Barium(s)	1,00	0,59	0,50	0,54
		Barium(t)	1,00	0,40	0,50	0,53
		Chloride	1,09	0,44	0,52	0,53
	C3_2012	Barium(t)	1,19	0,51	0,52	0,58
		Chloride	1,04	0,56	0,54	0,53
		Iron(t)	2,60	1,00	5,35	0,51
	C2_2013	Barium(t)	1,03	0,23	0,41	0,53
		Chloride	1,07	0,61	0,43	0,54
		Sulfates	1,57	0,42	0,46	0,52
	C1_2014	Barium(t)	1,10	0,42	0,55	0,56
		Chloride	1,19	0,43	0,49	0,54
		Sulfates	1,18	0,54	0,50	0,56
	C2_2014	Barium(t)	1,03	0,24	0,43	0,53
		Chloride	1,01	0,35	0,47	0,54
	C1_2015	Barium(t)	1,62	0,39	0,46	0,52
		Chloride	1,07	0,45	0,45	0,54

s – soluble; t - total

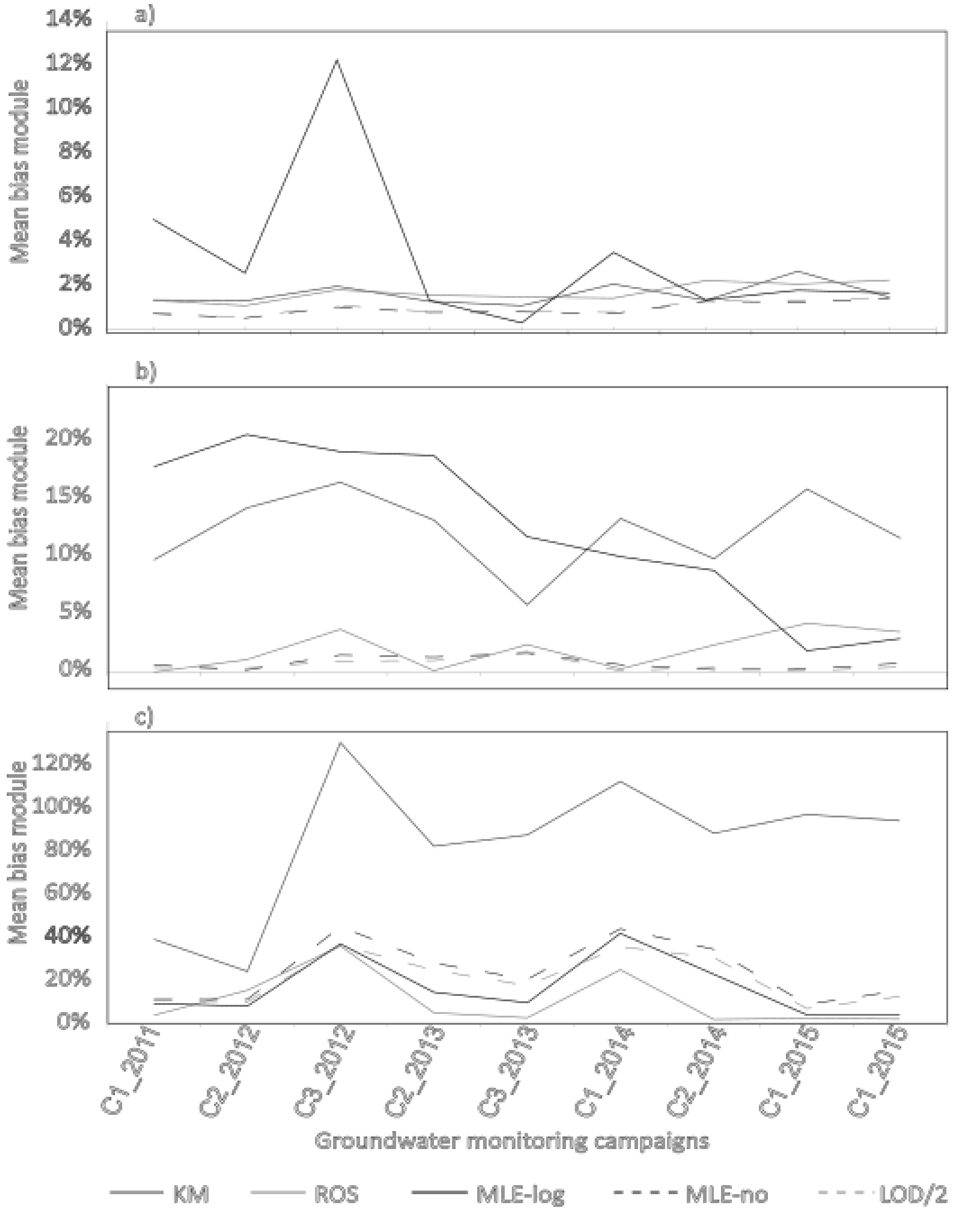
Table 7 – Bias of the original mean calculated from the means obtained by the ROS, KM, MLE-log and MLE-no in scenario 2.

Censored Data Range	Campaing	Parameter	Mean Bias KM (%)	Mean Bias ROS (%)	Mean Bias MLE-log (%)	Mean Bias MLE-no (%)	Mean Bias LOD/2 (%)
25%	C4_2011	Barium(s)	1,35	-1,61	7,92	-1,23	-3,17
		Barium(t)	1,37	-1,36	5,03	-0,76	-0,79
		Chloride	3,00	1,33	15,72	1,19	0,96
	C3_2012	Barium(t)	1,99	-1,84	12,25	-1,06	-1,14
		Chloride	3,58	0,43	8,68	0,52	0,48
		Iron(t)	0,28	-0,15	266,25	-0,01	-0,01
	C2_2013	Barium(t)	1,32	-1,60	1,37	-0,81	-0,84
		Chloride	2,43	0,84	19,80	0,61	0,58
		Sulfates	1,91	-0,59	8,68	-0,34	-0,37
	C1_2014	Barium(t)	2,10	-1,46	3,54	-0,76	-0,83
		Chloride	5,23	0,84	5,15	1,25	1,15
		Sulfates	4,39	-0,60	2,83	-1,08	-1,23
	C2_2014	Barium(t)	1,36	-2,25	1,37	-1,32	-1,36
		Chloride	3,21	1,46	15,68	1,05	1,03
	C1_2015	Barium(t)	2,68	-2,09	-1,81	-1,27	-1,33
		Chloride	5,63	1,53	2,68	2,23	2,12
50%	C4_2011	Barium(s)	11,12	-0,25	17,43	0,31	-0,12
		Barium(t)	9,73	0,04	17,83	-0,67	-0,40
		Chloride	15,34	0,48	7,68	4,26	3,91
	C3_2012	Barium(t)	16,45	3,73	19,10	1,53	0,99
		Chloride	19,43	0,20	12,22	1,60	1,17
		Iron(t)	4,15	1,33	292,24	1,44	1,42
	C2_2013	Barium(t)	13,19	0,16	18,79	-1,32	-1,03
		Chloride	13,92	1,98	15,40	2,20	1,83
		Sulfates	14,81	-0,37	6,68	2,49	2,22
	C1_2014	Barium(t)	13,31	-0,32	10,02	0,68	0,12
		Chloride	13,78	-0,50	7,90	1,33	0,79
		Sulfates	4,12	-0,33	2,19	-1,17	-2,61
	C2_2014	Barium(t)	9,83	-2,37	8,82	-0,26	-0,45
		Chloride	15,70	0,06	8,43	3,00	2,58
	C1_2015	Barium(t)	15,88	-4,28	1,90	0,27	-0,14
		Chloride	14,57	-2,27	7,47	1,37	0,76
75%	C4_2011	Barium(s)	51,94	20,09	13,51	16,77	13,48
		Barium(t)	38,83	3,55	9,09	10,95	10,56
		Chloride	59,04	7,32	13,55	14,63	12,28
	C3_2012	Barium(t)	130,09	35,66	36,63	44,69	36,18
		Chloride	48,59	12,35	10,85	9,85	7,46
		Iron(t)	69,26	23,46	148,03	9,32	9,01
	C2_2013	Barium(t)	82,14	-4,66	14,23	27,88	24,47
		Chloride	70,82	26,07	8,79	19,86	15,63
		Sulfates	93,48	5,52	8,74	13,77	11,89
	C1_2014	Barium(t)	112,19	24,71	41,50	43,61	35,28
		Chloride	56,22	3,53	7,74	11,25	8,66
		Sulfates	60,67	10,20	7,19	11,21	6,82
	C2_2014	Barium(t)	88,15	1,74	22,54	34,12	30,40
		Chloride	54,50	3,90	12,82	18,14	15,11
	C1_2015	Barium(t)	96,82	-2,16	3,86	8,66	6,68
		Chloride	59,47	6,32	6,36	14,36	5,32

s – soluble; t - total

This analysis reveals that the application of the four methods resulted in low bias values when the percentage of censored data is 25%, especially with the last three methods, especially with MLE-no and LOD/2, with bias values close to zero (Figure 1). The ROS, LOD/2 and MLE-no methods in general yielded better results when the percentage of censored data was 50%, especially the first two. The mean values estimated using these methods tend to increase with increasing censored data percentages. Although Helsel (2005) recommends a maximum of 80% of censored data in a dataset, it was possible to observe that datasets with more than 70% of censored data in scenario 2 showed a relatively high average bias. Therefore, it is possible in the future, through more careful analyses, to revise this limit.

Figure 1 – Graphs of the bias modules for mean values of barium concentration in groundwater for scenario 2. (a) 25% censored data; (b) 50% censored data; (c) 75% censored data



Therefore, the results for Scenario 2 highlight a better performance of ROS and MLE-no, as attested by other authors (GILLIOM; HELSEL, 1986; KROLL; STEDINGER, 1996; GIBBONS *et al*., 2001; CROGHAN; EGEGHY, 2003). The results obtained with these methods were similar to those obtained by DS with LOD/2, but the use of statistically-based methods of fast and simple execution using basic software is always preferable.

It is worth mentioning that the results obtained here contradict Helsel (2005), who recommends the application of the KM method to data sets containing up to 50% censored data.

5. CONCLUSIONS

Censored data are commonly related to groundwater geochemical sampling and to find solutions to substitute them is fundamental when analyzing the behavior of aquifer contamination. To assess alternatives of substitution for censored data, two scenarios were designed to validate the generated estimates. Scenario 1 consisted of variables containing data censored by the laboratory analytical methods, with a proportion of censored data less than 80%. Scenario 2 consisted of a data set containing no censored data, in which censored values were introduced artificially.

The ROS, MLE-no and LOD/2 methods yielded the best results for mean estimates for geochemical database from the landfill area in both scenarios, in particular MLE-no and LOD/2. The MLE-no and ROS methods yielded the best quantification limits respectively for database containing 25% and 50% censored data. The reason for the good performance of the ROS method in estimating mean values for data containing percentages up to 50% censured data is that it is a robust method, in which the magnitude and the distribution of uncensored data have great influence in the fraction determined by the substitution factor.

These methods are a good alternative for the traditional direct substitution of hydrogeochemical parameters below the detection limit, as they are statistically based and easy to apply with open-source software.

REFERENCES

ANTWEILER, R.C., TAYLOR, H.E. Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets: I. Summary statistics. *Environmental Science & Technology*, v. 42, p. 3732–3738, 2008. https://doi.org/10.1021/es071301c

ANTWEILER, R.C., 2015. Evaluation of Statistical Treatments of Left-Censored Environmental Data Using Coincident Uncensored Data Sets. II. Group Comparisons. *Environmental Science & Technology*, v. 49, p. 13439-13446, 2015 https://doi.org/10.1021/acs.est.5b02385

BACELLAR L.A.P., OLIVEIRA FILHO W.L. *Caracterização definitiva da pluma de contaminação das águas subterrâneas da área do aterro sanitário da CTRS-BR040*. Unplubished report, Belo Horizonte, 2009.

BARELLA, C.F., BACELLAR, L.A.P., NALINI, H.A. Influence of the natural oxidation of the leachate organic fraction from a landfill on groundwater quality, Belo Horizonte: Minas Gerais, south-eastern Brazil. *Environ Earth Sci, v.* 70, p. 2283–2292, 2013. https://doi.org/10.1007/s12665-013-2284-4

BACCARELLI, A. *et al*. Handling of dioxin measurement data in the presence of non-detectable values: overview of available methods and their application in the Seveso chloracne study. *Chemosphere*, v. 60, n. 7, p. 898-906, 2005. https://doi.org/10.1016/j.chemosphere.2005.01.055

CARRANZA, E.J.M. Analysis and mapping of geochemical anomalies using logratio-transformed stream sediment data with censored values. *J. Geochem. Explor*. v. 110, n. 2, p. 167-185, 2011. https://doi.org/10.1016/j.gexplo.2011.05.007

CLARKE, J. U. Evaluation of Censored Data Methods To Allow Statistical Comparisons among Very Small Samples with Below Detection Limit Observations. *Environmental Science & Technology* v. 32, p. 177-183, 1998. https://doi.org/10.1021/es970521v

COHEN, A.C. Tables for maximum likelihood estimates: singly truncated and singly censored samples. *Technometrics,* v. 3, p. 535–541, 1961. https://doi.org/10.1080/00401706.1961.10489973

CROGHAN, C. W., EGEGHY, P. P. *Methods of dealing with values below the limit of detection using SAS. Southeastern SAS User Group September*, p. 22-24, 2003.

FIÉVET, B., VEDOVA, D.C. Dealing with non-detect values in time series measurements of radionuclide concentration in the marine environment. *Journal Environmental Radioactive* v. 101, n. 1, p. 1-7, 2010. https://doi.org/10.1016/j.jenvrad.2009.07.007

FISHER, R.A. Theory of statistical estimation. *In*: MATHEMATICAL PROCEEDINGS OF THE CAMBRIDGE PHILOSOPHICAL SOCIETY. Cambridge University Press, p. 700–725, 1925. https://doi.org/10.1017/S0305004100009580

GIBBONS, R.D. *Statistical methods for detection and quantification of environmental contamination*. United States: John Wiley & Sons, 2001.

GILLIOM, R.J., HELSEL, D.R. Estimation of distributional parameters for censored trace level water quality data: 1. Estimation techniques. *Water Resources Research,* v. 22, p. 135–146, 1986. https://doi.org/10.1029/WR022i002p00135

HELSEL, D.R. Insider censoring: distortion of data with nondetects. *Human and Ecological Risk Assessment*, v. 11, p. 1127–1137, 2005. https://doi.org/10.1080/10807030500278586

HELSEL, D.R. Fabricating data: how substituting values for nondetects can ruin results, and what can be done about it. *Chemosphere*, v. 65, p. 2434–2439. https://doi.org/10.1016/j.chemosphere.2006.04.051

HELSEL, D.R. *Statistics for censored environmental data using Minitab and R*. Second Edition. United States: John Wiley & Sons, 2011. https://doi.org/10.1002/9781118162729

HEWETT, P., GANSER, G.H. A comparison of several methods for analyzing censored data. *Annals of Occupational Hygiene* v. 51, p. 611-632, 2007.

HORNUNG, R.W., REED, L.D. Estimation of average concentration in the presence of nondetectable values. *Applied occupational and environmental hygiene*, v. 5, p. 46–51, 1990. https://doi.org/10.1080/1047322X.1990.10389587

KAPLAN, E.L., MEIER, P. Nonparametric estimation from incomplete observations. J*ournal of the American Statistical association, v.* 53, p. 457–481, 1958. https://doi.org/10.1080/01621459.1958.10501452

KROLL, C.N., STEDINGER, J.R. Estimation of moments and quantiles using censored data. *Water Resources Research, v.* 32, p. 1005–1012, 1996. https://doi.org/10.1029/95WR03294

LEE, L. *NADA*: Nondetecs and Data Analysis for Environmental Data. R Package, 2010.

LEE, L., HELSEL, D. Baseline models of trace elements in major aquifers of the United States. *Applied Geochemistry*, v. 20, n. 8, p. 1560-1570, 2005. https://doi.org/10.1016/j.apgeochem.2005.03.008

LEITH, K.F. *et. al.* A comparison of techniques for assessing central tendency in left-censored data using PCB and pp'DDE contaminant concentrations from Michigan's Bald Eagle Biosentinel Program. *Chemosphere*, v. 80, p. 7-12, 2010. https://doi.org/10.1016/j.chemosphere.2010.03.056

LEVITAN, D.M., SCHREIBER, M.E., SEAL, R.R., BODNAR, R.J., AYLOR J.G. Developing protocols for geochemicalbaeline studies: An exemple from the Cole Hill uranium deposit, Virginia, USA. *Applied Cheochemistry, v.* 43, p. 88-100, 2014. https://doi.org/10.1016/j.apgeochem.2014.02.007

LIU, S., LU, J.C., KOPLIN, D. W., MEEKER, W. Q. Analysis of environmental data with censored observations. *Environmental Science & Technology*, v. 31, p. 3358-3362, 1997. https://doi.org/10.1021/es960695x

REIMANN, C. *Statistical data analysis explained*: applied environmental statistics with R. United States: John Wiley & Sons, 2008. https://doi.org/10.1002/9780470987605

SANFORD, R.F., PIERSON, C.T., CROVELLI, R.A.. An objective replacement method for censored geochemical data. *Mathematical Geology*, v. 25, p. 59–80, 1993. https://doi.org/10.1007/BF00890676

SINGH, A., NOCERINO, J. Robust estimation of mean and variance using environmental data sets with below detection limit observations. *Chemometrics and Intelligent Laboratory Systems,* v. 60, p. 69-86, 2002. https://doi.org/10.1016/S0169-7439(01)00186-1

TEMPL, M., FILZMOSER, P., REIMANN, C. Cluster analysis applied to regional geochemical data: problems and possibilities. *Applied Geochemistry*, v. 23, p. 2198–2213, 2008. https://doi.org/10.1016/j.apgeochem.2008.03.004

THODE, H.C. *Testing for normality*. First Edition. United States: CRC press, 2002. https://doi.org/10.1201/9780203910894

YOUNG, K.D., MENEGAZZI, J.J., LEWIS, R.J. Statistical method: IX. Survival analysis. *Academic emergency medicine*, v. 6, p. 244–249, 1999. https://doi.org/10.1111/j.1553-2712.1999.tb00165.x