Revista Ibero-Americana de Ciências Ambientais
Ibero-American Journal of Environmental Sciences

Out a Nov 2020 - v.11 - n.6

ISSN: **2179-6858**

This article is also available online at:
www.sustenere.co

Substantive audit testing of sewer systems using Brazilian open database: stat methods for compliance screening

The monitoring of data and the management of information are important tools for the decision-making and guidance of investments, which is a crucial procedure for the effective capital opening of the sanitation sector. The research aims at analyzing the correlation of urban population growth with the variables of importance for evaluating the overall performance of sanitary sewage systems. The study was developed based on raw data and indicators, in the long term period (1999 to 2017), with the information provided by the service providers themselves to the SNIS (national system of information on sanitation). The methodology consists of an application of a substantive test, following the guidelines of ISO 19.011 :2018, NBC TI 01 and, NBC PI 01: 2012, where the structure is presented in the format of the analytical review. For the validation of the test, statistical routines were used. Scores were elaborated to determine the probable inconsistencies found during the statistical routine, where the results, denominated evidence by the audit, were divided into conformities and non-conformities, being indicated 19 conformities and seven non-conformities. It is concluded that the methodology used can help evaluate public domain data and generate subsidies for sector managers and investors.

Keywords: Conformity; Sewage; Statistics.

Teste de auditoria substantiva de sistemas de esgoto usando banco de dados aberto brasileiro: métodos estatísticos para triagem de conformidade

O monitoramento de dados e a gestão de informações são ferramentas importantes para a tomada de decisão e direcionamento de investimentos, procedimento fundamental para a efetiva abertura de capital do setor de saneamento. A pesquisa tem como objetivo analisar a correlação do crescimento da população urbana com as variáveis de importância para a avaliação do desempenho geral dos sistemas de esgotamento sanitário. O estudo foi desenvolvido com base em dados brutos e indicadores, no período de longo prazo (1999 a 2017), com as informações prestadas pelos próprios prestadores de serviço ao SNIS (Sistema Nacional de Informações sobre Saneamento). A metodologia consiste na aplicação de um teste substantivo, seguindo as diretrizes da ISO 19.011: 2018, NBC TI 01 e, NBC PI 01: 2012, onde a estrutura é apresentada no formato de revisão analítica. Para a validação do teste, foram utilizadas rotinas estatísticas. Foram elaborados escores para apurar as prováveis inconsistências encontradas durante a rotina estatística, onde os resultados, denominados evidências pela auditoria, foram divididos em conformidades e não conformidades, sendo apontadas 19 conformidades e sete não conformidades. Conclui-se que a metodologia utilizada pode auxiliar na avaliação de dados de domínio público e gerar subsídios para gestores e investidores do setor.

Palavras-chave: Conformidade; Esgoto; Estatisticas.

Topic: **Sistemas de Gestão Ambiental**

Reviewed anonymously in the process of blind peer.

Received: **06/10/2020**
Approved: **28/11/2020**

Bianca Harumi Yamaguti Garcia
Universidade Federal de Rondonópolis, Brasil
http://lattes.cnpq.br/8559038420337922
http://orcid.org/0000-0002-9509-6767
bianca.yg@hotmail.com

Ricardo Alves de Olinda
Universidade Estadual da Paraíba, Brasil
http://lattes.cnpq.br/7767223263366578
http://orcid.org/0000-0002-0509-8428
ricardo.estat@yahoo.com.br

Domingos Sávio Barbosa
Universidade Federal de Rondonópolis, Brasil
http://lattes.cnpq.br/6896725721269796
http://orcid.org/0000-0001-6793-0956
domingosbar@gmail.com

Camila Leonardo Mioto
Universidade Federal de Rondonópolis, Brasil
http://lattes.cnpq.br/2846355335199799
http://orcid.org/0000-0002-6951-9527
ea.mioto@gmail.com

DOI: 10.6008/CBPC2179-6858.2020.006.0057

Referencing this:
GARCIA, B. H. Y.; OLINDA, R. A.; BARBOSA, D. S.; MIOTO, C. L.. Substantive audit testing of sewer systems using Brazilian open database: stat methods for compliance screening. **Revista Ibero Americana de Ciências Ambientais**, v.11, n.6, p.716-724, 2020. DOI: http://doi.org/10.6008/CBPC2179-6858.2020.006.0057

INTRODUCTION

The population growth, urbanization, and modifications in the sanitation systems are processes potentially important for the assessment of the environmental quality of the river basins. The interest to study the methods of evaluating and audit the efficiency of the sanitation services is growing, particularly to energy efficiency, reduce water losses and security (XIN et al., 2014; LIVSHITZ et al., 2017; FRAIA et al., 2018; BYLKA et al., 2019).

In Brazil, the main milestone for the development of sanitation is regulated by Federal Law 11.445 (BRASIL, 2007). In these, the Brazilian Government was created the National System of Information on Sanitation (SNIS), when the open data of structural and management sewer systems of cities are registered.

According to Pupin et al. (2015), Brazilian cities may be suffering financial penalties from the Federal Government if they do not report the sewer data. Therefore, only 64% of the national cities report to SNIS in 2013. However, a decrease in government political activity about the sanitation of Brazil was observed in the last five years (2015 to 2020).

Since it is a voluntary report, although, obligatory for the receiving of financial resources in the area, the SNIS needs tools to raise the compliance of the information sent by the service provider. The methodology to audit and certificate the data sent to the SNIS has been instituted, to gather techniques which aggregates sufficient evidence or proofs for the determination of its statement, nominated Acertar Project (INTERAGUAS, 2017).

The main parameters for the application of the audit procedures of the Acertar Project method, it is necessary to apply a substantive test. One possibility of this can be a screening hypothesis if the total urban population (response variable) influences the behavior of the other raw data sewage structure indicators (covariate variable). The hypothesis can be analyzed by the audit requirements from ISO 19.011 (2018) and Brazilian accounting standards NBC TI 01 and NBC PI 01 (2012). The research aims at analyzing the correlation of urban population growth with the variables of importance for evaluating the overall performance of sanitary sewage systems.

MATERIALS AND METHODS

Study location

The Vermelho River Basin (VRB) has around 1.508,2 Km2 of a catchment area, located on in the south of Mato Grosso State (population estimated 3.484.466), and is an important agricultural commodities producer (IBGE, 2020). Vegetation typically of Savanna and tropical humid climate. The Vermelho River drains on the São Lourenço River Basin (SLRB), and according to Silva et al. (2020), has located northeastern of the Pantanal Wetland, the largest floodplain of the world. On the São Lourenço Water Basin, has located the largest urban centers of Mato Grosso State, with significative anthropic effect under the Pantanal floodplain. Thus, the monitoring of water quality is necessary, especially in the final disposal effluent sites of sewage. Tha SLRB concentrates around 31% and the VRB around 8,9% of the total population of Mato Grosso State.

Figure 1 presents a map of the location of VRB highlighting the perimeters of the cities which contain information in the database of the SNIS, as well as its most important tributaries and areas of special environmental and humanitarian vulnerability (Jarudore, Tadarimana, and Tereza Cristina Indigenous and Community Conserved Area) and the Conservation Units of Dom Osório Stoffel and the João Basso State Ecological Parks.

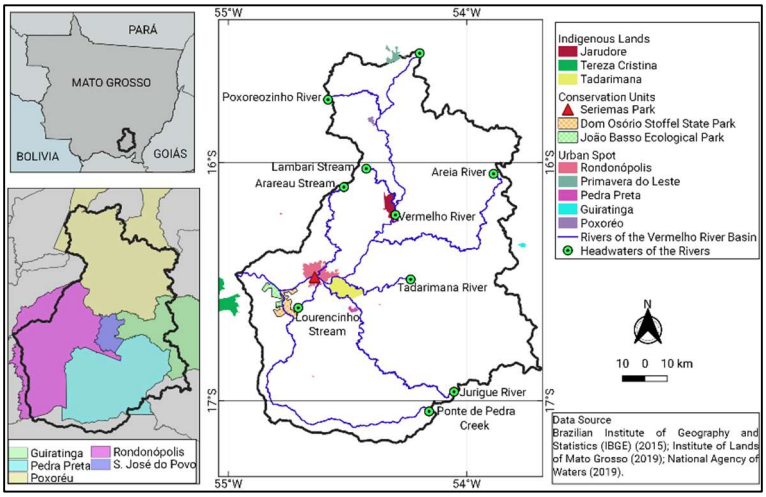

Figure 1: The study area of Vermelho River Watershed (VRW) highlighting the with the cites, tributaries, and vulnerable areas.

Source data

Data were extracted from the database of the SNIS, between the years of 1999 and 2017, from five cities of the VRB (Guiratinga, Pedra Preta, Poxoréu, Primavera do Leste and Rondonópolis cities). The data was downloaded from the SNIS website with a total of (SNIS, 2019). Were analyzed a total of 22 data and sewage indicators selected from the original database. In the results, we represented the cities names by codes, due to ethical and legal criteria.

Statistical Analysis

The research follows a methodology where the verification of the parametricity of the data for the discovery of the type of correlation to be conducted is the first step, with: i) tests of normality, with the presence of descriptive statistics, in the descriptive analysis; ii) test of normality, with the presence of inferential statistics, with the application of the Anderson-Darling Test; soon after, it was used iii) Kendall's test of correlation, for reasons eventually explained ahead; iv) time series, in graph format, for the confirmation of the obtained data through the correlation and at last, v) adoption of scores, relating the results found in the Kendall correlation. For the test cited previously, the software Rstudio was used as a supporting tool in the statistical analysis.

Substantive test

After the execution of the statistical analysis, it is carried out a scoring through scores based on the Level of Compliance which the information provides the research. This system of scores is adapted from the

Acertar Project (ACERTAR, 2017), which proposes a Matrix of Information Certification based on the observance tests and substantive tests. Since the research is centralized in the application of substantive tests, the scores are shown in a linear format, according to Table 1.

The application of a substantive test, following the guidelines of ISO 19.011:2018, NBC TI 01, and NBC PI 01: 2012, where the structure is presented in the format of the analytical review.

Table 1: Score for audit of information based on its Level of Compliance.

	Scores		
	1	**2**	**3**
Decision/Verdict	**The information presents a low level of compliance or the compliance was not assessed**	**The information presents medium levels of compliance**	**The information presents high levels of compliance**
Criterion	**p-value >0,05**	**p-value 0,01<p<0,05**	**p-value < 0,01**

The Levels of Compliance are defined after the explanation of the results through the correlation analysis, proposed by the hypothesis of the research.

RESULTS AND DISCUSSION

One important parameter for a decision to conduct the statistical analysis is the information available for each city in the SNIS database. According to the hypothesis of this research, three cities previously mentioned do not present their data periodically. Thus, because of the scope of the data over the eighteen years, the study of Vermelho River Basin was based on two cities, denominated P and R, which of 83% of all the population of the VRB.

For the identification of the requirements applicable to the substantive test, the variables were collected in the platforms available by the SNIS, and organized for the treatment of the data, according to Table 2. Therefore, only eight raw data and eight indicators are suitable to use in the research, due to the same factors cited previously. The abbreviation of the variables, the original name, and the position (response or covariate variable) are described in Table 2.

Tests of normality and Anderson-Darling test

For the establishing of correlation between variables, the verification of the normality of the data to choose which type of correlation to use is necessary. Thus, initially, it has been conducted a verification through an inferential test of data normality.

The Anderson-Darling (AD) test was used, in a way that if the (p-value < 0,05), the null hypothesis that the variable follows a normal distribution is rejected. Table 2 presents the values obtained through the test in all the variables. Since all the variables had a significant p-value (p-value < 0,05), then, none of these follow normality, certified by an inferential test.

Table 2: Response and the covariate variables in Anderson-Darling Normality Test. N= number. Abr.= Abbreviation. U= Unidimensional.

Type	Variable Name	Unit	Abr.	Position	A	p-value

Raw data	Total urban population supplied with sanitary sewage	inh.	Tot. Urb. Pop.	Response	2,0762	2,1E-05
	Quantity of active sewage connections	con.	Con.		2,294	6E-06
	Quantity of active sewage economies	econ,	Econ.		1,8821	6,2E-05
	Extension of the sewage network	km	Ext.		0,94772	0,0145
	Volume of the collected sewage	1000 m^3/year	Vol. Col.		0,7471	0,04625
	Volume of the treated sewage	1000 m^3/year	Vol. Treat.		0,8085	0,03233
	Volume of the invoiced sewage	1000 m^3/year	Vol. Fat.		0	0
	Total consumption of electric energy in the sewage systems	1000 kWh/year	Cons. Ener.		4,9047	0
Indicator	Sewage collection index	%	Ind. Col.		1,0436	0,00822
	Sewage treatment index	%	Ind. Treat.		2,2257	8,7E-06
	Extension of the network of sewage through connection	m/con	Ind. Ext.	Covariate	0,78945	0,03641
	Index of urban service of sewage referred to the cities supplied with water	%	Ind. Water		2,3093	5,5E-06
	Index of treated sewage referred to the consumed water	%	Ind. Sew. Water Cons.		1,4487	0,00078
	Index of urban service of sewage referred to the cities supplied with sewage	%	Ind. Sew. Sew.		2,3093	5,5E-06
	Index of total service of sewage referred to the cities supplied with water	%	Ind. Sew. Water		0	0
	Index of electric energy consumption in systems of sanitary sewage	kWh/m^3	Ind. Cons. Ener.		5,3054	0

Correlation Test: Kendall Correlation

After the inferential test, was conducted the correlation between the independent variables and the response variable. The test evaluates the correlation between the ordinal variables as Kendall correlation, are presented in Table 2, and none of them follow normality. This test, when p-value < 0,01 it means that there is a dependence between the covariate variables (raw data and indicators) and the response variable (urban population) Thus, Table 3 contains the Kendall correlation test for the two cities in the study, denominated P and R.

Table 3: Kendall correlation test for the variables analyzed in the cities P and R in study. Bold values highlight p-values > 0,01.

Covariate Variable	City	τ (tau)	p-value
Con.	P	0,79	1,138E-03
	R	0,849	2,13E-03
Econ.	P	0,84	1,035E-03
	R	0,54	2,856E-03
Ext.	P	0,51	**0,01262**
	R	0,88	3,57E-06
Vol. Col.	P	0,74	4,822E-04

Vol. Treat.	R	0,79	5,45E-04
	P	0,69	1,172E-03
Vol. Inv.	R	0,66	8,22E-05
	P	0,67	9,405E-04
Ind. Col.	R	0,86	1.41E-06
	P	0,66	1,825E-03
Ind.Treat.	R	0,25	**0,1767**
	P	0,17	**0,4911**
Ind. Ext.	R	0,26	**0,1481**
	P	-0,43	**0,03618**
Ind. Water	R	-0,68	**0,0519**
	P	0,84	3,555E-05
Ind. Sew. Water Cons.	R	0,70	8.98E-05
	P	0,58	5,95E-03
Ind. Sew. Sew.	R	0,25	**0,1767**
	P	0,84	3,555E-05
Ind. Sew. Water	R	0,70	8.98E-05
	P	0,72	4,266E-04
	R	0,65	1,277E-04

As presented in the Kendall analysis, there are covariate variables that do not correlate with the response variable (Tot. Urb. Pop.), having a p-value < 0,01, where these are: Extension of the sewage network (city P); Sewage collection index (city R); Sewage treatment index (cities P and R); Extension of the network of sewage through connection (cities P and R) and Index of treated sewage referred to the consumed water (city R). The covariate variables cited above (with p-value <0,01) represent possible non-conformities of the sewage system of two cities of the VRB.

Time series

To search for more evidence, besides the Kendall correlation (Table 3), plots to the variables over time were created. The type of plot selected was a time series, which demonstrates the evolution of the covariate variables in a certain periodicity (represented by the horizontal axis 'x' of the graph), with the addition of a double-axis in the vertical, known as 'double y', as observed in Figure 2 (a-g).

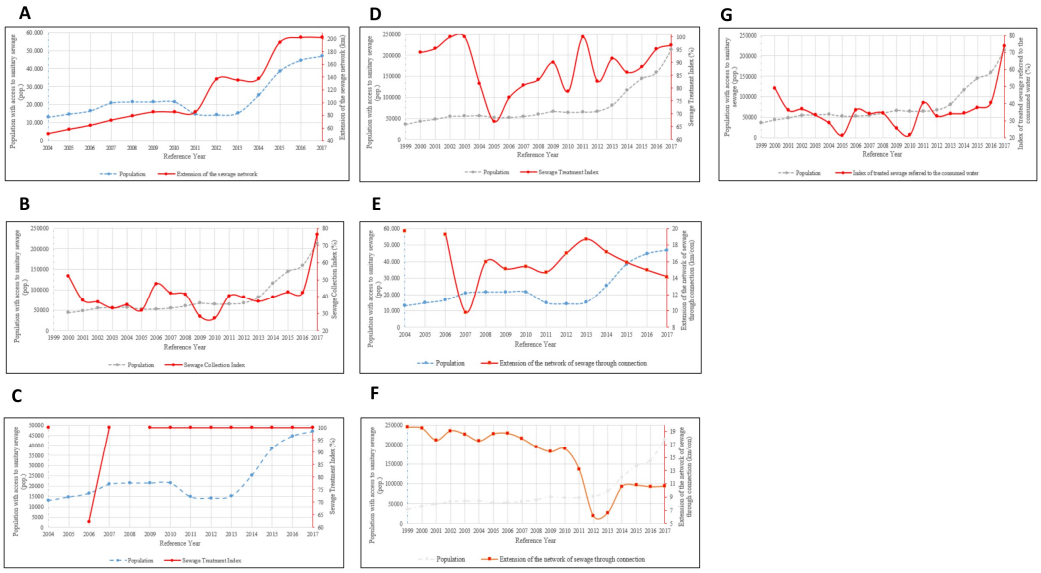

Figure 2 (a-g): Time Series in double 'y' which show the covariate variables that present non-conformities in the Kendall Correlation: (a) Ext. – City P; (b) Ind. Col. – City R; (c) Ind. Treat. – City P; (d) Ind. Treat. – City R; (e) Ind. Ext. – City P; (f) Ind. Ext. – City R; (g) Ind. Sew. Water Cons. – City R.

The behavior of the double 'y' plots through time series in Figure 2, explains the initial idea set by the Analysis of Kendall, where the covariate variables do not correlate with the response variable and can be associated with a non-conformities, for example, as show the graphs b) and c). In the first, the R city, the waveform of a variable 'extension of sewage system' over the years, can indicate an error in input data, supposedly causing the inconsistency; and the second (b), probably, calculations mistakes cause gaps and chaotic shape os the index. Presumably, it is do not expected the length of a city sewage system or the collection sewage index to reduce along the time when the population city grows continually. According to Costa et al. (2014) that the Brazilian average of Ind. Col. in 2010 was around 54,8%, and Maia et al. (2017) report that in 2014 the national average of Ind. Treat. Was of 48,6%.

Scores application: compliance level

After the application of correlation and the time series, the scores of compliance were determined, according to the criterion established in the methodology of this study, as shown in Table 4.

Table 4: Scores based on the Kendall correlation test oh the compliance level of the raw data and indicators used in this study. Highlights of compliance scores: High (Gren), Medium (Yellow), and Low (Red).

Covariate Variable	Score City P	Score City R
Con.	3	3
Econ.	3	3
Ext.	2	3
Vol. Col.	3	3
Vol. Treat.	3	3
Vol. Inv.	3	3
Ind. Col.	3	1
Ind. Treat.	1	1
Ind. Ext.	2	1
Ind. Water	3	3
Ind. Sew. Water Cons.	3	1
Ind. Sew. Sew.	3	3
Ind. Sew. Water	3	3

The scores of Compliance Level can denote a conformity and non-conformity status of a dataset or indicator. The supposed non-conformities, in what concerns the variable "network extension", there are possible inconsistencies found in the Medium and Low Level of Compliance. In city P, the raw data Ext., and the indicator Ind. Ext., both on Medium level; in R, the indicator Ext., of Low level.

About the raw data Ext., city P, as observed in Figure 6-a, the extension of the network in the first years of data gathering did not grow with the populational development, which justifies a possible non-conformity. However, the responsibility of the administrators of sewage systems is not exclusive. According to Suita (2018), in the city of Caxias do Sul – RS, the collect sewage network was installed in front of residences but, in some cases, does not connect to the municipal network.

The indicators Ind. Ext., in cities P and R, represents the average distance between the connections of sewage. According to MINUTA (2014), that a low value in the indicator demands bigger investments in the increase of the coverage of the networks of sanitary sewage

Through the time series represented by Figures 2-a and 2-g, it is possible to see the instability of the

indicators concerning the urban population, which is indicated as the cause of the supposed non-conformity by the statistical routine. As for the volume of water consumed by the population, according to what is stated by the National Agency of Waters – ANA (2018), of 500 thousand liters of water removed from nature, 400 thousand liters are reverted into sewage in the cities, where less than 40% of the national average are collected and treated. Therefore, in the past years, both cities walk towards the universalization of the coverage and treatment of sewage, seeing that the coefficient of return is close to 80%.

Lastly, the possible inconsistencies of Low Level of Compliance concerning two indicators, one in each city, about the variable "treated volume": Ind. Treat. In the cities P. and R. The reason of the treated volume by the collected volume brings time series with a stagnation of the indicator, in city P (Figure 6-c) and variability over the years in city R (Figure 6-d), justifying the inconsistency in the statistical routine. According to the Diagnosis of Water and Sewage Services – 2017 (2019), of all the sewage collected in Brazil, only 73,7% goes through treatment. This indicates that the values treated by the analyzed cities are superior to the national average, and this emphasizes the results, improved micro measurement of the values set would increase the level of the administrators of both cities.

The use of statistics for analyzes of the conformities and compliance of dataset of sewage systems, there are one of the methods to identify them, therefore making the management and the access to information public domain, applying on the third-party audits. The Project Acertar establishes some guidelines for the localization of inconsistencies in the governmental platforms, aiming at the quality in the management of the information in the area of sanitation in Brazil.

CONCLUSIONS

About the method, the results were considered satisfactory, because of the using of open and public domain dataset and in the application of a substantive test through statistical routine, followed by the distribution in scores, contributed for the gathering of supposed conformities, non-conformities and evidence in the systems of sanitation of the cities of the VRB.

The supposed non-conformities were identified possible inconsistencies with the hypothesis of correlation which guides the research, in the Kendall correlation as well as in the Time Series with graphs "double y" type. These data were emphasized through the scores of Level of Compliance medium and low and should be verified and accompanied in the process of post-audit, not covered by the research.

In general, the study can be replicated due to the facility in the handling of the information gathering by the SNIS platform and by the tools of audit, which involve aspects of the observance tests, in the internal scope, and the substantive tests, that can be external.

Finally, we strongly recommend the caution of the data sent to the SNIS, and encourage other studies to develop stats methods using sewage and water quality open data.

REFERENCES

ANA. Agência Nacional de Águas. **Conjuntura dos recursos hídricos no Brasil 2018:** Informe anual. Brasília: ANA, 2018.

BYLKA, J.; TOMASZ, M.. A Review of Energy Assessment Methodology for Water Supply Systems. **Energies**, v.12, p.23, 2019. **DOI:** https://doi.org/10.3390/en12234599

BRASIL. **Portaria n. 719 de 12 de dezembro de 2018.** Institui metodologia para auditoria e certificação de informações do Sistema Nacional de Informações sobre Saneamento (SNIS). Edição 239. Seção 1. Página 85. Ministério das Cidades. Brasília: DOU, 2018.

BRASIL. **Lei n. 11.445 de 5 de janeiro de 2007.** Estabelece diretrizes nacionais para o saneamento básico; [...] dá outras providências. Brasília: DOU, 2007.

CFC. Conselho Federal de Contabilidade. **Normas brasileiras de contabilidade: auditoria interna:** NBC TI 01 e NBC PI 01. Brasília: CFC, 2012.

COSTA, S. A. B.; CÔRTES, L. S.; MOY, H. L.; FREITAS JÚNIOR, M. M.. Indicadores em saneamento: análise da prestação dos serviços de água e de esgoto no Brasil. In: SIBESA, ASSOCIAÇÃO BRASILEIRA DE ENGENHARIA SANITÁRIA E AMBIENTAL, 12. **Anais.** Natal: UFRN, 2014.

FRAIA, S.; MASSAROTTI, N.; VANOLI, L.. A Novel Energy Assessment of Urban Wastewater Treatment Plants. **Energy Conversion & Management**, Londres, v.163 p.304-13, 2018 **DOI:** https://doi.org/10.1016/j.enconman.2018.02.058

INTERAGUAS. Programa de Desenvolvimento do Setor Água. **Projeto Acertar: Manual de Melhores Práticas de Gestão da Informação sobre Saneamento.** Brasília: MDC, 2017.

ISSO. International Organization for Standardization. I**SO n. 19011 de 20 de dezembro de 2018.** Guidelines for auditing management systems. Rio de Janeiro: ABNT, 2018

LIVSHITZ, I.; LONTSIKH, P.; ELISEEV, S.. The optimization method of the integrated management system security audit. In: CONFERENCE OF OPEN INNOVATIONS ASSOCIATION, 20. **Anais.** Russia: IEEE, 2017. p.24853.

MAIA, C. V. A.; BATISTA, L. J. R.; LIMA, J. G. A.. Análise da prestação municipal de serviços de abastecimento de água e esgotamento sanitário em municípios do nordeste brasileiro. **Revista Espacios**, v.38, n.8, p.25,2017.

MDR. Ministério do Desenvolvimento Regional. Secretaria Nacional de Saneamento. **Diagnóstico dos Serviços de Água e Esgoto, 2017.** Brasília: MDR, 2019.

MINUTA. **Relatório de avaliação de desempenho**. Agência Reguladora. São Paulo: ARESPCJ. 2014.

PUPIN, P. L. F.; BRUMATTI, L. M.; BORGES, A. C. G.. Análise dos dados sobre resíduos sólidos nas bases da PNSB e do SNIS. **Revista Nacional de Gerenciamento de Cidades,** v.3, n.21, 2015. **DOI:** http://dx.doi.org/10.17271/2318847232120151151

SNIS. Sistema Nacional de Informações sobre Saneamento. **Série Histórica.** Brasília: SNS, 2019.

SILVA, M. O.; OLINDA, R.; MIOTO, C. L.; BARBOSA, D. S.. Análise plurianual da qualidade das águas de bacia tributária do Pantanal brasileiro**. Revista Ibero-Americana de Ciências Ambientais**, v.11, n.2, p.172-181, 2020. **DOI:** https://doi.org/10.6008/CBPC2179-6858.2020.002.0019

SOUZA, A. V. V.; LOVERDE-OLIVEIRA, S. M.. Análise da qualidade da água do Rio Vermelho em Mato Grosso: no período de cheia no ano de 2014. **Biodiversidade**, v.13, n.2, p.115-126, 2014.

SUITA, M. C. A.. **Análise dos fatores determinantes para a ocorrência de ociosidade das redes de esgoto sanitário no município de Caxias do Sul/RS.** Dissertação (Mestrado em Engenharia e Ciências Ambientais) - Universidade de Caxias do Sul, Porto Alegre, 2018.

XIN, K.; TAO, T.; YINGJUN, L.; XIAOLAN, X.; FEI, L.. Apparent Losses Analysis in District Metered Areas of Water Distribution Systems. **Water Resources Management,** v.28, n.3, p.683-696. 2014. **DOI:** http://doi.org/10.1007/s11269-013-0508-8