



UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GEOQUÍMICA: PETRÓLEO E MEIO
AMBIENTE – POSPETRO

GILSON BARBOSA DOURADO

**AJUSTE DO MODELO ADITIVO GENERALIZADO E
IDENTIFICAÇÃO DE PADRÃO ESPACIAL EM DADOS
GEOQUÍMICOS DO MUNICÍPIO DE MADRE DE DEUS**

Salvador
2019

GILSON BARBOSA DOURADO

**AJUSTE DO MODELO ADITIVO GENERALIZADO E
IDENTIFICAÇÃO DE PADRÃO ESPACIAL EM DADOS
GEOQUÍMICOS DO MUNICÍPIO DE MADRE DE DEUS**

Tese apresentada ao Programa de Pós-Graduação em Geoquímica: Petróleo e Meio Ambiente (Pospetro), do Instituto de Geociências da Universidade Federal da Bahia, como requisito parcial para obtenção do título de Doutor em Geoquímica do Petróleo e Ambiental

Orientadora: Profa. Dra. Gisele Mara Hadlich
Coorientadora: Profa. Dra. Denise Nunes Viola

Salvador
2019

Ficha catalográfica elaborada pelo Sistema Universitário de Bibliotecas (SIBI/UFBA),
com os dados fornecidos pelo(a) autor(a).

Dourado, Gilson Barbosa
AJUSTE DO MODELO ADITIVO GENERALIZADO E
IDENTIFICAÇÃO DE PADRÃO ESPACIAL EM DADOS GEOQUÍMICOS
DO MUNICÍPIO DE MADRE DE DEUS / Gilson Barbosa
Dourado. -- Salvador, 2019.
104 f.

Orientador: Gisele Mara Hadlich.
Coorientador: Denise Nunes Viola.
Tese (Doutorado - Programa de Pós-Graduação em
Geoquímica: Petróleo e Meio Ambiente (Pospetro)) --
Universidade Federal da Bahia, Universidade Federal
da Bahia, Instituto de Geociências, 2019.

1. Geoquímica Ambiental. 2. Análise Fatorial. 3.
Padrão espacial. 4. Geoestatística. 5. Mapeamento
Científico. I. Hadlich, Gisele Mara. II. Viola, Denise
Nunes. III. Título.

**AJUSTE DO MODELO ADITIVO GENERALIZADO E
IDENTIFICAÇÃO DE PADRÃO ESPACIAL DE DADOS
GEOQUÍMICOS DO MUNICÍPIO DE MADRE DE DEUS, BAHIA**

por

Gilson Barbosa Dourado

(Bacharel em Estatística – Universidade Federal da Bahia – UFBA - 2001, Salvador - BA)

Orientadora: Profa. Dra. Gisele Mara Hadlich

Coorientadora: Profa. Dra. Denise Nunes Viola

TESE DE DOUTORADO

**Submetida em satisfação parcial dos requisitos do grau de
DOUTOR EM GEOQUÍMICA DO PETRÓLEO E AMBIENTAL**

À Câmara de Ensino de Pós-Graduação e Pesquisa da Universidade Federal da Bahia

APROVAÇÃO


BANCA EXAMINADORA



Profa. Dra. Gisele Mara Hadlich (IGEO/UFBA)



Prof. Dr. Sérgio Augusto de Moraes Nascimento (IGEO/UFBA)



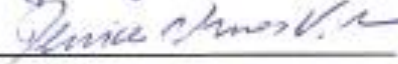
Profa. Dra. Giovana Oliveira Silva (IME/UFBA)



Prof. Dr. Antônio Fernando de Souza Queiroz (IGEO/UFBA)



Prof. Dr. Joll José Celino (IGEO/UFBA)



Profa. Dra. Denise Nunes Viola (IME/UFBA)

Data da Defesa Pública: 27/06/2019

SALVADOR-BAHIA

JUNHO/2019

À Guiomar Dourado (*in memoriam*) e a José Dourado, mãe e pai queridos, por ter me ensinado a aprender e dar valor à vida.

À Silvana Dourado, esposa amada, pela grande paciência e abnegação.

AGRADECIMENTOS

À Gisele Mara Hadlich, orientadora querida, sempre muito paciente e tão atenciosa, receptiva e, acima de tudo, uma grande orientadora.

À Denise Nunes Viola, coorientadora e querida amiga, pela paciência em ajudar a alinhar a estatística com a geoquímica, e pela generosidade.

A Antônio Fernando Queiroz, pela imensa ajuda e colaboração nas correções e adequação do tema à geoquímica, pela generosidade, pela disponibilidade e amizade.

Ao Programa de Pós-Graduação em Geoquímica: Petróleo e Meio Ambiente (Pospetro), da UFBA, pelo apoio, a infraestrutura, a qualidade e a simpatia dos seus professores, pesquisadores e funcionários.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

A Marcio Silva, amigo e colega de curso, na paciência em ensinar geoquímica, pela grande amizade e ajuda nos momentos de descontrole emocional.

À Milena Marlim, pela amizade, pela motivação em buscar concluir o doutorado, além das grandes resenhas.

À Camila Paim, pela amizade, ajuda e colaboração na compreensão da geoquímica, além das grandes resenhas nos corredores.

A todos os colegas de curso que muito contribuíram e cooperaram para a minha formação em geoquímica. Tive o grande privilégio de ter colegas empenhados e comprometidos que sem eles, não conseguiria finalizar o curso.

Muito obrigada por possibilitarem essa experiência enriquecedora e gratificante, da maior importância para meu crescimento como ser humano e profissional.

Por último, mas não menos importante, a Deus (Jeová), pois sem Ele seria impossível concluir o curso e a quem também devo a minha vida.

RESUMO

Esta pesquisa teve por objetivo analisar e modelar o comportamento de parâmetros de geoquímica inorgânica em solos e sedimentos na área litorânea da Baía de Todos os Santos (BTS), Bahia, especificamente no município de Madre de Deus. Visou, também, estudar a amostragem espacial em Geoquímica. Os métodos geoestatísticos unem aspectos espaciais com aspectos aleatórios de dados. Esses métodos baseiam-se na teoria das variáveis regionalizadas a partir da qual é possível estudar a estrutura espacial possibilitando calcular o erro cometido na avaliação, levando em conta a posição real das amostras. Inicialmente, para maior conhecimento do material bibliográfico sobre o tema, foi realizado um estudo sobre o mapeamento científico relacionado com pesquisas publicadas em periódicos indexados no Scopus, pertinentes à amostragem espacial para observação da evolução das formas de abordagem. As análises realizadas foram através do software R com o pacote/ferramenta bibliometrix. Constatou-se um aumento no número de artigos publicados ao longo dos anos avaliados (2007 a 2017); os Estados Unidos da América e a China são os países com maior número de publicações sobre o tema. As palavras-chave mais frequentes nos resumos dos artigos e nos títulos foram mapeadas, mostrando o aumento da utilização, na área de geociências, de ferramentas estatísticas cada vez mais sofisticadas nas soluções de problemas. Em seguida foi construído um modelo de regressão linear utilizando o GAMLSS para compreender a influência dos parâmetros físicos e químicos no índice de partição reduzida (I_R) das frações geoquímicas de apicuns em Madre de Deus. O I_R descreve quantitativamente a distribuição do metal nas frações em um intervalo entre 0 a 1, e foi aplicado em 14 amostras de sedimentos. Obteve-se as concentrações dos metais Cu, Cr e Ni em cada fração geoquímica (trocável, ligada a carbonatos, ligada à matéria orgânica, ligada a óxidos-hidróxidos de Fe e Mn, e na forma residual) e calculou-se o I_R para cada metal analisado. Para modelar o comportamento de parâmetros físico e químicos (pH, Eh, salinidade, granulometria, matéria orgânica - MO, capacidade de troca catiônica (CTC), nitrogênio total (N) e fósforo assimilável (P)), foram construídos modelos considerando o I_R como variável de interesse e os referidos parâmetros como o conjunto de variáveis explicativas. As variáveis pH, Eh e areia influenciaram a distribuição dos metais. A influência da salinidade foi constatada somente para o Cu e Ni, e não para o Cr. Os modelos GAMLSS produziram resultados coerentes para o apicum que apresenta elevada dispersão nos dados, demonstrando que modelos de regressão mais sofisticados podem ser utilizados na modelagem de dados ambientais. Por último, realizou-se um estudo para identificar a existência de padrão espacial dos parâmetros geoquímicos inorgânicos e de metais em solos e sedimentos em Madre de Deus por meio de técnicas geoestatísticas e da análise multivariada (Análise Fatorial Exploratória). Inicialmente foi feita análise fatorial robusta que gerou quatro fatores que representaram, juntos, 81,6% da variação total dos dados. Os dados não apresentaram o pressuposto de distribuição normal e não foi possível fazer o variograma para verificar a existência de padrão espacial, optando-se pelo teste de aleatorização de Mantel. De acordo com esse teste, metais, parâmetros físicos e químicos apresentam padrão espacial não aleatório, indicando que essas quantidades/ concentrações são autocorrelacionadas espacialmente.

Palavras-chave: Geoquímica Ambiental, Análise Fatorial, Padrão espacial, Geoestatística, Mapeamento Científico

ABSTRACT

The objective of this research was to analyze and model the behavior of inorganic geochemistry parameters in soils and sediments in the Bay of All Saints (BTS), Bahia, specifically in the municipality of Madre de Deus. It also aimed to deal with spatial sampling in Geochemistry. Geostatistical methods link spatial aspects with random aspects of data. These methods are based on the theory of regionalized variables from which it is possible to study the spatial structure, making it possible to calculate the error made in the evaluation, taking into account the actual position of the samples. Initially, for a better knowledge of the bibliographic material on the subject, a study was carried out on the scientific mapping related to research published in periodicals indexed in Scopus, pertinent to the spatial sampling to observe the evolution of the forms of approach. The analyzes were performed through software R with the bibliometrix package / tool. There was an increase in the number of articles published over the evaluated years (2007 to 2017); the United States of America and China are the countries with the largest number of publications on the subject. The most frequent keywords in the article summaries and in the titles were mapped, showing the increasing use in the area of geosciences of increasingly sophisticated statistical tools in problem solutions. A linear regression model was then constructed using the GAMLSS to understand the influence of physical and chemical parameters on the reduced partition index (IR) of apicuns geochemical fractions in Madre de Deus. IR quantitatively describes the distribution of the metal in the fractions in a range of 0 to 1, and was applied to 14 sediment samples. Concentrations of Cu, Cr and Ni metals were obtained in each geochemical fraction (exchangeable, bound to carbonates, bound to organic matter, bound to Fe and Mn oxides-hydroxides, and in residual form) and IR was calculated for each metal analyzed. To model the behavior of physical and chemical parameters (pH, Eh, salinity, granulometry, organic matter - OM, cation exchange capacity (CTC), total nitrogen (N) and assimilable phosphorus (P)), as variable of interest and said parameters as the set of explanatory variables. The pH, Eh and sand variables influenced the distribution of metals. The influence of salinity was verified only for Cu and Ni, and not for Cr. The GAMLSS models have produced consistent results for apicum that show high dispersion in the data, demonstrating that more sophisticated regression models can be used in the modeling of environmental data. Finally, a study was carried out to identify the existence of a spatial pattern of inorganic and metal geochemical parameters in soils and sediments in Madre de Deus using geostatistical techniques and multivariate analysis (Exploratory Factor Analysis). Initially, a robust factor analysis was performed, which generated four factors that together represented 81.6% of the total data variation. The data did not present normal distribution and it was not possible to make the variogram to verify the existence of spatial pattern, choosing the Mantel randomization test. According to this test, metals, physical and chemical parameters have a spatial pattern, indicating that these quantities / concentrations are spatially autocorrelated.

Keyword: Environmental Geochemistry, Factor Analysis, Spatial Pattern, Geostatistics, Scientific Mapping

SUMÁRIO

1	INTRODUÇÃO	10
2	HIPÓTESES	14
3	OBJETIVOS	15
4	MAPEAMENTO CIENTÍFICO SOBRE AMOSTRAGEM ESPACIAL EM GEOCIÊNCIAS	16
	RESUMO	16
4.1	INTRODUÇÃO	16
4.2	AMOSTAGEM ESPACIAL	18
4.2.1	Amostragem com base no desenho (<i>based-design</i>)	21
4.2.2	Amostragem como base no modelo (<i>basead-model</i>)	24
4.3	MATERIAIS E MÉTODOS	25
4.4	RESULTADOS E DISCUSSÃO	26
4.5	CONCLUSÃO	35
5	IDENTIFICAÇÃO DE PADRÃO ESPACIAL EM DADOS GEOQUÍMICOS INORGÂNICOS – ESTUDO DE CASO: MADRE DE DEUS, BAHIA, BRASIL	36
	RESUMO	36
5.1	INTRODUÇÃO	36
5.2	MATERIAIS E MÉTODOS	38
5.2.1	Dados geoquímicos (parâmetros físicos e químicos)	40
5.2.2	Análise exploratória de dados e Análise multivariada	40
5.2.4	Análise de padrão espacial de dados	41
5.3	RESULTADOS E DISCUSSÕES	42
5.4	CONCLUSÃO	52
6	MODELO ADITIVO GENERALIZADO PARA POSIÇÃO, ESCALA E FORMA APLICADO A AMBIENTE HIPERSALINO ASSOCIADO A MANGUEZA: ESTUDO DE CASO	56

RESUMO	56
6.1 INTRODUÇÃO	57
6.2 MATERIAIS E MÉTODOS	59
6.2.1 Procedimentos Analítico	60
6.2.2 Índice de partição reduzida dos metais: I_R	61
6.2.3 Análise Estatística	62
6.3 RESULTADOS E DISCUSSÃO	63
6.3.1 Modelo estatístico	64
6.3.2 Modelo ajustado	68
6.4 CONCLUSÃO	70
7 CONCLUSÃO GERAL	72
REFERÊNCIAS	73
APÊNDICE A	81
APÊNDICE B	93
APÊNDICE C	97

1 INTRODUÇÃO

Há algumas definições para a geoquímica ambiental, no geral, ela tem como objetivos estudar, analisar e entender as relações entre os elementos químicos que interagem entre a litosfera, hidrosfera, atmosfera, biosfera; tem, como campo de análise, o espaço antroposférico, ou seja, o espaço no qual é mais perceptível a influência humana nos recursos naturais, conseqüentemente interferindo nos espaços socioeconômicos. Para Siegel (2002) é a disciplina que estuda a química da terra sólida, seus componentes aquosos e gasosos e formas de vida para avaliar os impactos da contaminação por metais pesados nos ecossistemas do nosso planeta. Trata-se de condições fisiológicas, químicas e biológicas em um ambiente como temperatura, estado da matéria, acidez (pH), potencial de redução-oxidação (redox), atividade bacteriana e demanda biológica de oxigênio (DBO). Esses fatores e outros influenciam a mobilização, dispersão, deposição, distribuição e concentração de metais/metalóides potencialmente tóxicos que podem prejudicar a saúde dos organismos em um ecossistema.

Devido ao seu objeto de estudo, estudos geoquímicos ambientais, gera grande quantidade de dados analíticos que podem ser utilizados para avaliar a qualidade e/ou contaminação do ambiente, do solo e da água. A interpretação de dados analíticos em um estudo geoquímico deve ter duas grandes preocupações: 1) confiabilidade dos dados - com grandes quantidades de dados gerados é comum enfrentar problemas com a qualidade desses resultados; conclusões diferenciadas podem ser obtidas em função da coleta de amostras, diferenciadas no tempo, ou de procedimentos analíticos distintos; 2) georreferenciamento dos dados - é fundamental que os dados sejam georreferenciados para que seja identificado o padrão espacial dos parâmetros analisados, tornando possível fazer uma modelagem espacial sobre a área de estudo. Compreender a distribuição espacial oriunda de fenômenos ocorridos no espaço geográfico constitui um grande desafio para elucidação de questões centrais em diversas áreas do conhecimento. O objetivo da análise espacial é mensurar propriedades e correlação entre os vizinhos, levando em conta a localização espacial do fenômeno em estudo de forma explícita. Nesse contexto surgiu, na década de 50, a Estatística Espacial (DRUCK, 2004).

A Estatística Espacial é uma ferramenta que possibilita estudar o comportamento espacial de variáveis e permite estimar seu valor em áreas geograficamente amostradas com base mais informações de seus vizinhos mais próximos. Tal ferramenta é utilizada para

entendimento e modelagem da variabilidade espacial de atributos. A distribuição espacial de dados oriundos de fenômenos ocorridos no espaço constitui um grande desafio para a compreensão de questões centrais em estudos ambientais, e analisar a dependência espacial do fenômeno de estudo vem se tornando cada vez mais comum, o que é possibilitado pelos Sistemas de Informações Georreferenciadas (SIGs) (CÂMARA et al., 2004). Os SIGs têm contribuído para o desenvolvimento de estudos em monitoramento ambiental, mas para fazer uma análise espacial devem-se ter, inicialmente, dados georreferenciados e confiáveis. Nesse sentido, Lago et al. (2014) apresentaram um Sistema de Banco de Dados Geoquímicos Georreferenciados (SBDGG/NEA/POSPETRO/UFBA) com objetivo de sistematizar, armazenar e disponibilizar dados geoquímicos gerados a partir de resultados relacionados à Baía de Todos os Santos - Ba (BTS) e seu entorno. Isto inclui dados de geoquímica inorgânica, em diferentes matrizes ambientais: solo, sedimentos, água e vegetação.

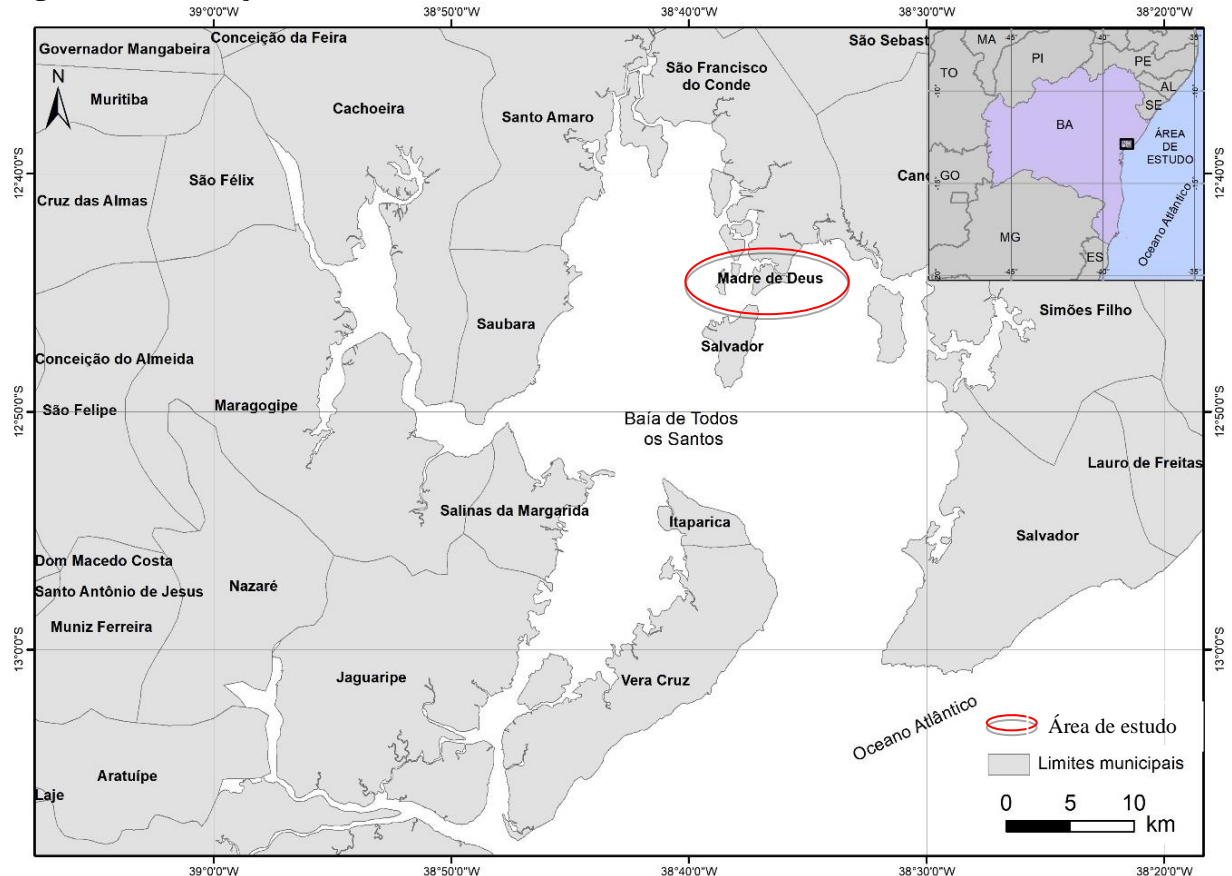
Assim, esta pesquisa, propõe a integração de técnicas estatísticas de Análise Espacial de Dados e Geoestatística para analisar a existência de padrão espacial e a geração de mapas dos parâmetros físico-químicos, contribuindo para uma melhor compreensão da dinâmica da distribuição desses parâmetros especialmente na região norte da BTS, tomando como espaço o Município de Madre de Deus.

A BTS está centrada na latitude 12°50' S e longitude 38°38' W, apresentando uma área de 1.233km²; isso a coloca na posição de segunda maior baía do Brasil. Ao longo dos anos, a BTS sofreu importantes ações antrópicas que alteraram suas características originais. Destaca-se que em setembro de 1950 entrou em operação a refinaria Landulpho Alves de Mataripe (RLAM), no Município de São Francisco do Conde. Em 1967 foi implantado o Centro Industrial de Aratu e em junho de 1978 entrou em operação e primeiro complexo petroquímico planejado do país: o Polo Petroquímico de Camaçari. Esses complexos estão localizados na parte mais interna da BTS, parte norte. Nesta baía, existem 10 terminais portuários de grande porte, um canal de vias navegáveis, além de canais internos profundos. Sua riqueza natural, com expressiva extensão de estuários e manguezais e sua forte relação com a história do Brasil, faz da BTS um polo turístico, além de polo econômico da região (HATJE; ANDRADE, 2009).

Devido às intervenções antrópicas de porte e com o crescimento das cidades em seu entorno, a BTS tem sido objeto de estudos e de monitoramento ambiental permanente. A relevância socioeconômica e ambiental desta baía no Estado da Bahia e a existência de diversas pesquisas que geraram dados de geoquímica ambiental, destacando-se os trabalhos realizados pelos pesquisadores do Núcleo de Estudos Ambientais do Instituto de Geociências

da UFBA (NEA/IGEO/UFBA), foram motivadores da escolha desta área para o presente estudo. A área de estudo específica desta pesquisa abrange o município de Madre de Deus na zona costeira norte da BTS (Figura 1).

Figura 1 - Localização de Madre de Deus na Baía de Todos os Santos (BTS), Estado da Bahia



Fonte: Elaboração do autor (2019).

Fundamentação teórica sobre estatística espacial, amostragem e modelagem são apresentados no Apêndice A, e algumas complementações sobre Materiais e Métodos usados constam no Apêndice B.

Os resultados estão sendo apresentados sob a forma de três artigos científicos, que contemplam:

- uma narrativa sobre amostragem espacial e um levantamento bibliométrico para analisar a evolução da abordagem dos termos relacionados com o tema, de forma a compreender melhor esse campo de pesquisa;
- uma verificação da existência de padrão espacial dos parâmetros geoquímicos inorgânicos do sedimento por meio da associação de técnicas geoestatísticas e de análise multivariada, visando contribuir com a melhoria no entendimento dos padrões espaciais existentes.

- uma avaliação do modelo de regressão, apropriada para dados correlacionados, utilizando o GAMLSS para compreender a influência dos parâmetros físicos e químicos no índice de partição reduzida (I_R), nas frações geoquímicas de um ambiente costeiro hipersalino denominado de apicum, na região de Madre de Deus, na BTS.

2 HIPÓTESES

Grande parte dos processos geoquímicos que ocorre na superfície terrestre está associada à presença de elementos químicos característicos do processo gerador e espacialmente relacionados à sua área de ocorrência. Em pesquisas geoquímicas, portanto, é fundamental que sejam traçados planos de amostragem espacial e que os dados analíticos obtidos sejam adequadamente tratados para retratar com confiabilidade o ambiente e permitir interpretações não equivocadas.

Na Baía de Todos os Santos, sedimentos e solos no Município de Madre de Deus apresentam parâmetros geoquímicos com padrão espacial não aleatório. Portanto, modelos de análise multivariada e geoestatísticos devem ser utilizados para explicar os comportamentos de parâmetros físicos e químicos analisados (pH, Eh, salinidade, granulometria, matéria orgânica - MO, capacidade de troca catiônica - CTC, nitrogênio total - N e fósforo assimilável - P) e de metais associados (Cu, Zn, Mn, Fe e Cr).

Com base nisso, as hipóteses são:

- os modelos estatísticos semiparamétrico explicam satisfatoriamente a variabilidade que ocorre entre os parâmetros físicos e químicos oriundos do ambiente apicuns (ambientes hipersalinos associados a manguezais) em Madre de Deus;
- a autocorrelação espacial entre os ambientes (mangue, apicum, encosta e urbano) influencia as concentrações de metais, dos parâmetros físicos e químicos nos sedimentos e solo no Município de Madre de Deus.

3 OBJETIVOS

O objetivo geral da pesquisa é analisar e modelar, através de técnicas estatísticas de Análise Espacial e Modelagem Geoestatística, o comportamento de parâmetros de geoquímica inorgânica - parâmetros físicos e químicos (pH, granulometria, salinidade, matéria orgânica, fósforo, nitrogênio e metais) em solos e/ou sedimentos no Município de Madre de Deus, localizado na região NE da Baía de Todos os Santos (BTS).

São objetivos específicos:

- discutir alguns aspectos sobre amostragem espacial, sua evolução e mapear o uso de técnicas de amostragem espacial nas Geociências e em estudos geoquímicos ambientais;
- analisar e avaliar o padrão espacial não aleatório dos dados geoquímicos de sedimentos ou solos presentes na área de estudo;
- gerar mapas que evidenciem a relação entre diferentes parâmetros na distribuição espacial das variáveis físicas e químicas de solos e sedimentos;
- utilizar método geoestatístico multivariado para estimar os parâmetros físicos e químicos da matriz ambiental deste estudo.

4 MAPEAMENTO CIENTÍFICO SOBRE AMOSTRAGEM ESPACIAL EM GEOCIÊNCIAS

RESUMO

Este estudo é uma revisão narrativa da literatura sobre trabalhos publicados relacionados à amostragem espacial para observação da evolução das formas de abordagem. Foi realizado levantamento bibliográfico e foram utilizados artigos da base de dados Scopus utilizando os descritores “sampling spatial”, e a partir do tratamento desses resultados, realizou-se uma análise bibliométrica utilizando-se o software R através do pacote/ferramenta bibliometrix. Foram selecionados os estudos considerando os critérios de inclusão a partir dos descritores, o aspecto cronológico, a adequação e pertinência dos estudos e o impacto destes artigos na comunidade científica. Como produto, observou-se que ocorreu um aumento no número de artigos publicados ao longo de anos e redução nas citações dos artigos. Os Estados Unidos da América e a China são os países com maior número de publicação que abrangem o tema. No total, 331 artigos foram selecionados da base de dados da área de Geociências ou correlatas. Desses, 1.363 autores publicaram trabalhos relacionados ao tema da amostragem espacial, como autor principal ou coautor, sendo o periódico com maior número de publicações o Plos One. As palavras-chave mais frequentes nos resumos dos artigos e nos títulos foram mapeadas evidenciando o aumento da utilização, na área de geociências, de ferramentas estatísticas cada vez mais sofisticadas nas soluções de problemas de interesse.

Palavras chaves: amostragem espacial, mapeamento científico, bibliometria, geociências

4.1 INTRODUÇÃO

Em estudos geoquímicos é comum o uso de diferentes matrizes (rochas, solos, sedimentos, água, atmosfera, biota) para o desenvolvimento de pesquisas e/ou monitoramento, pois essas matrizes colaboram para uma melhor explicação do fenômeno em investigação. Entretanto, quando o pesquisador trabalha com diferentes matrizes deve ter bastante cautela, pois diferentes fontes de informações ampliam a variabilidade que não é controlada no estudo. Paralelamente, como não é possível multiplicar o número de amostras e/ou matrizes analisadas, haja visto o aumento de custo operacionais das pesquisas, essa variabilidade incorporada ao estudo acaba sendo desprezada na análise. Para não correr esse erro, é fundamental realizar um adequado planejamento amostral que acarrete a redução da variabilidade dos dados.

Um procedimento inicialmente comum em estudos geoquímicos, destacando-se os ambientais, para se obter conhecimento de algumas características intrínsecas ao espaço, é observar uma ou mais amostras de seus elementos, e a partir do resultado amostral, obter

estimativas para a característica de interesse (amostra piloto). O problema maior deste procedimento é garantir que a amostra seja representativa, ou seja, saber se os resultados obtidos na análise desse espécime são confiáveis para avaliar as características da população e se são representativos da heterogeneidade do ambiente avaliado.

Há diversas razões para realizar estudos geoquímicos por amostragem. Os motivos mais comuns se devem: a) à busca de redução de custos para estudar uma população, à redução do tempo e de mão-de-obra para a realização da coleta dos dados; b) à necessidade de confiabilidade e qualidade dos dados, considerando uma perda de precisão que seja insignificante em comparação com a enumeração da população (COCHRAN, 1977). A ressalva para esse procedimento está nos casos em que a população é pequena, quando a característica é de fácil mensuração ou quando há necessidade de alta precisão.

Nas Geociências, técnicas de amostragens tradicionais não são adequadas, visto que a variabilidade da população, na maioria das vezes, depende de sua localização. Para lidar com esse problema, propõem-se utilizar a amostragem espacial em que as escolhas das amostras são realizadas com mais de uma dimensão, por exemplo, tempo ou espaço (localização). Acompanhando essa linha de pensamento, Matheron (1971) propôs o uso da teoria da variável regionalizada, que foi bem construída e amplamente aplicada nas Ciências da Terra. Outra abordagem é utilizar a autocorrelação espacial para melhorar a eficiência da amostragem em termos da estimativa do erro da variância, no tocante ao desenho amostral e ao tamanho da amostra (STEIN; ETTEMA, 2003). Outros autores, como J.F. Wang (2009; 2010), propuseram ainda o uso de estratificação espacial heterogênea para a obtenção de uma amostragem espacial com melhor eficiência na inferência.

A quantidade de informações disponíveis aumentou a dificuldade de garantir uma amostra representativa e relevante a respeito de um tema. Ao estudar um tema, é frequente se deparar com métodos e resultados, às vezes, contraditórios. Segundo Moreira (2008), no atual contexto da informação, estar informado é, paradoxalmente, o grande problema. Assim, em uma era da informação que produz um número cada vez maior de publicações, identificar a pesquisa relevante e suas interconexões tornou-se uma tarefa cada vez mais desafiadora e valiosa.

Um caminho coerente para tentar esclarecer controvérsias é apoiar-se em mapeamento bibliográfico do tema em estudo. Partindo desse pressuposto, um bom delineamento da pesquisa é iniciado com uma revisão sistemática da literatura, isso porque a revisão sistemática é um tipo de investigação focada em questão bem definida, que visa identificar, selecionar, avaliar e sintetizar as evidências relevantes disponíveis.

Ao fazer o levantamento bibliográfico de um tema, o pesquisador se depara com uma massa de dados que necessita de boas técnicas que possibilitem uma análise confiável. Dessa necessidade surgiu a bibliometria. Nascimento (2016), defini bibliometria como: “conjunto de técnicas quantitativas de avaliação que permite medir a produção e a qualidade de pesquisadores, grupos ou instituições de pesquisa, com base no número de publicações e citações bibliográficas”.

O mapeamento científico, ou mapeamento bibliométrico, é um importante tópico de pesquisa no campo da bibliometria. Ele tenta encontrar representações de conexões intelectuais dentro do sistema dinamicamente mutável do conhecimento científico. Em outras palavras, o mapeamento científico visa mostrar os aspectos estruturais e dinâmicos da pesquisa científica (MORRIS; VAN DERVEER MARTENS, 2008).

O mapeamento científico ou mapeamento bibliométrico é uma representação espacial de como as disciplinas, campos, especialidades e documentos individuais ou autores são relacionados uns com os outros. É focado em monitorar um campo científico e delimitar áreas de pesquisa para determinar sua estrutura cognitiva e sua evolução. O fluxo de trabalho geral em uma análise de mapeamento científico tem diferentes etapas: recuperação de dados, pré-processamento, extração de rede, normalização, mapeamento, análise e visualização. No final desse processo, o analista tem que interpretar e obter algumas conclusões a partir dos resultados (COBO et al., 2011).

O objetivo deste estudo é analisar a evolução da abordagem dos termos relacionados a métodos de amostragem espacial, através de um levantamento bibliométrico, de forma a compreender melhor esse campo de pesquisa. Para tanto, questões sobre amostragem espacial são inicialmente abordadas.

Os dados bibliométricos foram levantados utilizando-se o software R (versão 3.2.1, R DEVELOPMENT CORE TEAM, 2015), partindo-se da base de dados da Scopus. Os resultados obtidos foram analisados e interpretados para melhor se compreender este campo de estudo, conhecendo-se os principais periódicos, as principais palavras-chave, os autores mais citados e os mais produtivos dentre outros.

4.2 AMOSTAGEM ESPACIAL

A diferença básica do método de amostragem espacial em relação à amostragem clássica está no fato de que uma amostra é selecionada com base na localização geográfica e/ou nas

suas características associadas. Diferente da teoria de amostragem clássica, uma série de métodos de amostragem espacial tem sido desenvolvidos e testados. Esses métodos têm sido explorados intensivamente para provar fenômenos naturais usando-se um padrão geométrico regular (AMBLARD-GROSS et al., 2004; DESSARD; BARHEN, 2005; STEVENS; OLSEN, 2004). Segundo Cressie (1990), a principal ideia por trás da amostragem espacial é determinar um tamanho ideal da amostra, selecionando-se os locais de amostragem $\{s_1, s_2, \dots, s_n\}$ a partir do qual os dados $Z = (Z(s_1), Z(s_2), \dots, Z(s_n))$ possam ser utilizados para estimar $g(Z(.))$, em que g refere-se a alguma área/extensão geográfica.

Considerando-se a estratégia tríade de amostragem $(\mathfrak{R}, \mathfrak{T}, \psi)$, em que \mathfrak{R} é o campo aleatório, \mathfrak{T} é o desenho amostral e ψ é um estimador, o símbolo \mathfrak{R} representa as características de um campo aleatório, como o erro aleatório, tendências, autocorrelação ou estratificação, que são determinantes da eficiência do cálculo da amostra. O desenho amostral \mathfrak{T} refere-se ao procedimento de determinação de seleção da amostra em termos de densidade e localização, enquanto o estimador ψ define o procedimento utilizado para calcular uma estimativa de um parâmetro de população a partir dos dados amostrais. É inútil discutir os méritos de um desenho amostral sem considerar o estimador a ser utilizado e as características do campo aleatório, ou vice-versa. É a combinação dos três (campo aleatório, desenho amostral e estimador) que determina a precisão das estimativas (WANG et. al., 2012).

O processo de amostragem espacial pode ser dividida em seis etapas (COCHRAN, 1977):

1. o esclarecimento dos objetivos de amostragem para estimar a média ou total de uma população, para prever estimacao de valores em locais sem amostragem, para o mapa de uma área, ou para identificar a posição de um alvo;
2. definição da população a ser amostrada e unidade amostral. No contexto espacial, a população da qual são extraídas as unidades amostrais e da população a ser estimada pode ser espacialmente inconsistente; a primeira é chamada de população de amostragem e a segunda, população a estimar;
3. escolha do método de amostragem e cálculo do tamanho da amostra sob restrições orçamentárias e de precisão específicas. A decisão correta pode ser feita utilizando-se a tríade do desenho amostral $(\mathfrak{R}, \mathfrak{T}, \psi)$;
4. desenho de um plano de amostragem que descreva onde, quando e como unidades de amostragem devem ser tomadas;
5. desenho da amostra e realização de medições nas unidades amostrais obtidos. A recolha de medidas é comumente chamada de amostra;

6. a análise da amostra à luz do conhecimento prévio para se obter as características do campo estocástico, tais como aleatoriedade, tendência, autocorrelação ou estratificação. Essas características e os objetivos de amostragem fornecem a base para a escolha da melhor estatística para fazer inferências para os parâmetros populacionais.

Duas teorias são comumente distintas para a amostragem e a inferência: a amostragem com base no **desenho** e amostragem com base no **modelo**. Elas diferem em seu elemento de aleatoriedade na atribuição de uma estrutura estocástica para inferências (SÄRNDAL, 1978; DE DEGRUIJER; TER BRAAK, 1990). Em um contexto espacial, o "aqui" (local) e o "agora" (tempo) característicos de um campo aleatório são obtidos uma vez que toda a área é exaustivamente pesquisada. A incerteza da estimativa da amostra decorre da aleatoriedade dos locais de amostragem, que pode ser reduzida através do aumento da dimensão da amostra até que a área esteja totalmente coberta. Esse tipo de abordagem é uma amostragem com base em desenho/planejamento, que pode ser implementada por meio de amostragem tradicional, por exemplo, amostragem aleatória simples, amostragem estratificada e etc (COCHRAN, 1977). Em alternativa, se um valor já foi uma realização de um processo estocástico subjacente (uma superpopulação), pode não ser obtido mesmo se a área inteira é exaustivamente pesquisada. Essa abordagem é chamada de amostragem com base no modelo (SÄRNDAL, 1978) e os métodos envolvidos estão principalmente relacionados com estatísticas espaciais e geoestatística (MATHERON, 1963), em que o campo aleatório observado é considerado como uma única realização de um processo espacial. Alguns pressupostos têm que ser feitos e o modelo do processo tem que ser bem definido para estimar a superpopulação, porque apenas uma observação transversal está disponível, o que muitas vezes não é suficiente para suportar a inferência de população não imensurável.

O processo de amostragem de matrizes ambientais (rocha, solo, água, sedimento ou outra) pode envolver duas fontes de aleatoriedade: os locais de amostragem podem ou não ser selecionados aleatoriamente, dadas as localizações, e os valores a serem medidos podem ou não ser considerados como aleatórios (BRUS; DE GRUIJTER, 1997). O Quadro 1 apresenta as possíveis combinações de estratégias de amostragem para os correspondentes local e valores versus fixo e aleatório. Cada uma dessas quatro combinações representa uma abordagem diferente para a amostragem.

Quadro 1 - Estratégias amostrais definidas com base em duas fontes de aleatorização

Locais /regiões de amostragem	Valores em determinados locais/regiões	
	Fixo	Aleatório
Fixo	Estratégia totalmente determinística	Estratégia com base no modelo: p.e, Krigagem probabilística
Aleatório	Estratégia com base no desenho/planejamento: Teoria clássica de amostragem	Estratégia totalmente aleatória

Fonte: Brus e Gruijter (1997).

Quando o objetivo da amostragem espacial é prever/estimar a variável de interesse com o menor número de locais de monitoramento, dois métodos são geralmente sugeridos para alcançar esse objetivo: o **espaço de enchimento** (apud NYCHKA; SALTZMAN, 1998) e **desenho Bayesiano ótimo** utilizando simulação de cadeias de Markov não homogêneas (MÜLLER, 1998). Grande parte do foco em amostragem espacial nos últimos anos tem sido a estrutura geométrica para a geração de uma grade aleatória e a otimização em relação ao tamanho da amostra, a maximização da variância e autocorrelação espacial (DESSARD; BARHEN, 2005; GETIS et al., 2000; SALEHI, 2004; ZHU; STEIN, 2006; ZHU; ZHANG, 2006). As grades regulares mais comuns são a grade triangular equilátera, a grade retangular / quadrada e a grade hexagonal (CRESSIE, 1990).

O plano de amostragem ideal busca capturar a variabilidade máxima pelo mínimo de unidades amostrais possíveis, o que envolve uma adequada distância/espacamento entre unidades amostrais para que autocorrelação espacial possa ser minimizada ou eliminada. Em vista das dificuldades do método, a autocorrelação pode ser descartada por duas razões. Em primeiro lugar, evitar autocorrelação espacial reduz a redundância. Em segundo lugar, a estrutura de correlação espacial pode ainda ter implicações para escolhas em relação ao desenho espacial de amostragem, a estimativa e previsão, e o método de observação (SALEHI, 2004). Uma abordagem com base em modelo de utilização de tais padrões de correlação tem sido particularmente influente na pesquisa das ciências da Terra, em relação aos recursos minerais e fósseis de combustível. Tais comportamentos são discutidos por Cressie (1990), quando analisa diferentes métodos de estrutura de correlação de modelagem em dados espaciais.

4.2.1 Amostragem com base no desenho (*based-design*)

Segundo Brus e Gruijter (1997), na abordagem/estratégia com base no desenho/planejamento, a aleatoriedade é introduzida na fase da amostragem, ou seja, os locais

de amostragem são selecionados por um processo predominantemente aleatório. As regras do experimento aleatório (processo de seleção) são descritas com precisão pelo desenho amostral, daí o termo abordagem com base no desenho. Nessa abordagem, o desenho amostral não faz referência ao conjunto real de locais de amostragem, mas com o procedimento utilizado para selecionar a amostra. Um resultado possível deste experimento aleatório é composto por um conjunto de locais de amostragem, da amostra e do número de objetos a serem coletados locais. Assim, na terminologia da amostragem clássica, muitas amostras diferentes podem ser selecionadas a partir de uma determinada área utilizando o mesmo desenho amostral aleatório.

A teoria da abordagem com base no desenho é muitas vezes referida como teoria clássica de amostragem, que não deve ser confundida com a estatística clássica. A teoria da amostragem clássica baseia-se em probabilidades a partir da aleatorização e desenvolvendo-se separadamente na corrente principal da estatística clássica. Essa última baseia-se principalmente em probabilidades inferidas a partir de modelos estocásticos, por exemplo, com os pressupostos de forma independente, de maneira idêntica e observações normalmente distribuídas.

Além de erros de medição, o processo de amostragem é a única fonte de aleatoriedade considerada na abordagem com base no desenho. Isto implica que o valor desconhecido em qualquer determinado local e hora é considerado como fixo, não aleatório.

A teoria da amostragem, desde os anos 40, busca por estratégias de recolhimento de espécimes que reduzam o custo do processo sem perda da robustez estatística (NEYMAN, 1934, 1938; COCHRAN, 1946; HORTVITZ; THOMPSON, 1952). Isso deu origem a diversos métodos nos quais a seleção destes materiais pode ter probabilidades iguais ou não. Amostragem Aleatória Simples (AAS) e amostragem sistemática (AS) servem como ponto de partida para a compreensão do recolhimento de amostras. Em AAS ou AS, as unidades de espécimes são desenhadas de forma independente uma da outra com a mesma probabilidade. É, teoricamente, uma abordagem relativamente simples, mas raramente é utilizada na amostragem devido à sua baixa eficiência. Outros métodos de amostragem rentáveis incluem a estratificada, aquela por conglomerados, a por múltiplos estágios, em duas fases, e ou a sequencial.

Por exemplo, na AAS um número pré-determinado de locais de amostragem é selecionado aleatoriamente a partir da área, com probabilidades iguais de seleção e de forma independente uma da outra. Isso é feito tomando-se as coordenadas geográficas de cada local a ser amostrado a partir de um gerador de números aleatórios ou de uma tabela de números

aleatórios. Com a Amostragem Estratificada (AE), a área é primeiramente dividida em certo número de sub-regiões chamadas estratos, e, em seguida, a AAS é aplicada a cada um dos estratos separadamente. Os tamanhos de amostra nos estratos podem ser escolhidos de tal forma que as probabilidades dos locais a serem amostrados diferem entre os estratos, podendo ainda ser escolhido proporcional a seu tamanho.

Em se tratando de amostragem espacial, há dois fatores muito importantes a serem consideradas, a autocorrelação e a heterogeneidade, visto que são fatores característicos dos fenômenos naturais. A autocorrelação espacial viola a suposição de que a população é independente e identicamente distribuída (i.i.d), mas, segundo o princípio de Tobler em geografia, ou primeira lei da geografia, observações próximas umas das outras são mais susceptíveis de ser semelhantes às observações mais distantes (TOBLER, 1970). O não reconhecimento da autocorrelação espacial pode implicar na eficácia com que a amostragem é efetuada. A variância do erro da estimativa pode, por exemplo, ser estimada de forma incorreta no que diz respeito às escolhas de projeto de amostragem e tamanho da amostra (RIPLEY, 1981; HAINING, 1988). Para solucionar o problema da autocorrelação espacial nos métodos de amostragem, por exemplo, AAS, AS, AE, alguns autores propõem correções nos estimadores levando-se em consideração a autocorrelação espacial (RIPLEY, 1981; DUNN; HARRISON, 1993).

Quenouille (1949) e Das (1950) foram os primeiros trabalhos a avaliar estratégias de amostragem espacial. Esses estudos teóricos foram realizados com base em campos aleatórios que produziram resultados dos processos espacialmente homogêneos. Ripley (1981) mostrou que AS supera outros delineamentos de amostragem quando não existe conhecimento prévio está disponível sobre campos aleatórios, exceto se houver periodicidade no atributo.

Dunn e Harrison (1993), em seu trabalho realizado com base em amostras de uso de solo, mostraram que, embora a AS tenha sido a mais eficaz dos três métodos de amostragem, os ganhos em eficiência, em relação à AE, foram altamente variáveis. Os seus resultados sugerem que a presença de não-estacionariedade e anisotropia em dados espaciais podem ter um efeito grave sobre a eficiência da AS.

Outra característica dos fenômenos naturais de suma importância em diferentes áreas de análise de dados em ciências ambientais é a heterogeneidade. A heterogeneidade dos atributos de um campo aleatório compreende dois elementos de variação de segunda ordem: a variância global (variância da população) e a estrutura espacial (autocorrelação espacial) (WANG et al., 2012). Ambos os elementos desta variação precisam ser reconhecidos na concepção e avaliação de projetos de amostra, incluindo o plano de amostragem e de escolha

do estimador (COCHRAN, 1977). Por exemplo, na amostragem estratificada espacial, a área heterogênea é dividida em várias subáreas que são cada uma mais homogênea, o que reduz a variância total. Após a divisão, o número de pontos de amostragem é atribuído a cada subárea de acordo com a área e/ou a proporção e, em seguida, AAS é aplicada a cada subárea.

4.2.2 Amostragem como base no modelo (*basead-model*)

Na amostragem com base no modelo, os valores do local/região de interesse não são fixados e um conjunto de valores observados em toda a região representa uma realização de um modelo estocástico da variação. A aleatoriedade nesta abordagem é introduzida através de um conjunto de modelos estocásticos. Os locais de amostragem são fixos e a inferência baseia-se assumindo a validade dos modelos estocásticos. Neste método, para calcular um estimador ou um preditor, os pesos dos dados de amostra são determinados por covariâncias entre as observações, que são dadas pelo modelo como uma função das coordenadas dos pontos de amostragem. Diferentemente da amostragem com base no desenho, a amostragem com base no modelo dispõe de uma função objetivo predefinido que minimiza a variância e, portanto, leva a um único esquema de coleta de materiais ideal para este objetivo específico. Wang et al. (2012), em um artigo de revisão sobre métodos de amostragem espacial, discutem três funções objetivos específicos que considera relevante em um contexto espacial: minimização da variância do erro de estimativa, atingir a cobertura espacial igual para polígonos irregulares e alcançar uma cobertura igual no espaço recurso.

Embora tenha havido grande progresso no desenvolvimento de teorias espaciais de amostragem, há poucos programas de computador abertos para amostragem espacial (SPOCK, 2012), porque o conhecimento prévio, autocorrelação espacial e heterogeneidade espacial não são facilmente implementáveis em programas. Assim, o programa de desenvolvimento para esse efeito é visto como solução de maneira a promover a utilização destas técnicas sofisticadas.

Do ponto de vista computacional, a diferença principal entre as duas abordagens é a forma como os dados da amostra são ponderados para se obter estimativas ou previsões. Na abordagem com base no desenho, os pesos são derivados a partir do delineamento. Na abordagem com base no modelo, os pesos são derivados do modelo escolhido e a configuração real de localizações da amostra. O mesmo aplica-se ao cálculo das estimativas de erro.

Quando os estatísticos da pesquisa começaram a desenvolver estratégias baseadas em modelos, discutiram a teoria e méritos de suas propostas em relação a estratégias baseadas em projetos já existentes. Assim, a literatura estatística geral sobre amostragem aponta, desde o início, uma consciência clara das duas abordagens e seu relacionamento. Paralelamente, as Ciências da Terra tinham desenvolvido uma teoria estatística especial para a estimativa espacial, conhecida como "teoria das variáveis regionalizadas" (MATHERON, 1971). Inicialmente, o desenvolvimento desta teoria e a metodologia da krigagem ocorreram fora dos canais regulares de comunicações estatísticas. Foi em grande parte um desenvolvimento separado, que criou a sua própria publicação e conferências em circuitos e terminologias, coletivamente referidos como "geoestatística". Apesar deste desenvolvimento separado, teoria das variáveis regionalizadas tem uma ligação mais estreita com a estatística clássica do que a teoria da amostragem clássica tem, no sentido de que ambas são baseadas em modelos estocásticos similares. Isso se reflete, por exemplo, pelo fato de que várias formas de krigagem podem ser reformuladas como casos especiais de regressão generalizada.

Durante este desenvolvimento separado, a geoestatística não discutiu sua metodologia em relação à teoria da amostragem clássica. Uma exceção é Borgman e Quimby (1988). Mais recentemente os métodos de krigagem foram reconhecidos e discutidos como casos especiais do que tinha sido definidas antes, como estratégias de amostragem com base no modelo na literatura sobre teoria da amostragem (DE GRUIJER; TER BRAAK, 1990). Provavelmente devido a esse isolamento, equívocos surgiram na literatura geoestatística sobre questões relacionadas com a amostragem aleatória.

4.3 MATERIAIS E MÉTODOS

No contexto deste estudo, o levantamento preliminar, que resultou no referencial teórico, e no mapeamento bibliométrico foi realizado em uma das mais populares e importantes bases de dados que indexa periódicos científicos, a Scopus. Optou-se pelo levantamento apenas dos artigos de periódicos.

A busca sistemática na base de dados do Scopus foi realizada para artigos científicos que abordassem amostragem espacial ("Spatial Sampling"). Optou-se por usar o termo em inglês nos títulos dos artigos, resumos ou palavras-chave, para o período compreendido entre 2007 e 2017. Foram localizados 1070 (mil e setenta) artigos que tratavam sobre o tema, em diferentes áreas de conhecimento. Foi então aplicado um critério/filtro para selecionar artigos

científicos das subáreas: Agricultura e Ciências Biológicas, e Ciências Ambientais. A base de dados Scopus forneceu, através de download, o banco de dados pesquisado na extensão CSV. O procedimento metodológico empregado no levantamento bibliográfico está representado na Figura 2.

Os resultados do levantamento dos dados sobre amostragem espacial foram analisados para o ano de publicação, assim como a revista de publicação, a frequência de termos e palavras-chave utilizadas pelos autores, além dos tópicos de pesquisa tratados e o país de origem do autor.

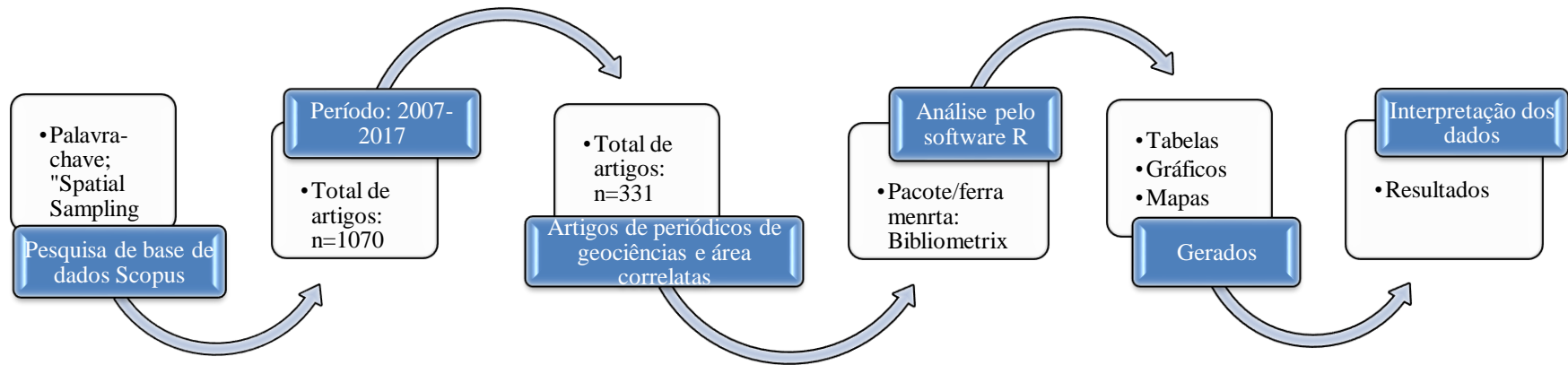
O software estatístico R de código aberto foi utilizado para análise bibliométrica e textual. Com uma grande comunidade de usuários, o software R tem muitas ferramentas disponíveis para realizar uma poderosa análise bibliométrica e de texto, utilizando comandos e pacotes (WELBERS et al., 2017). O pacote do R utilizado para análise bibliométrica e para o mapeamento bibliométrico foi o *bibliometrix* (ARIA, CUCCURULLO, 2017).

4.4 RESULTADOS E DISCUSSÃO

O total de artigos correspondentes à consulta (*sampling spatial*) na base de dados Scopus entre os anos de 2007 a 2017 foi de 1070 artigos científicos em periódicos revisados por pares. Como o objetivo é mapear esta temática na Geociência e em áreas correlatas, foi aplicado um filtro que resultou em 331 artigos de 180 periódicos distintos, produzidos por 1363 autores, com uma média de citação de 16,32.

Assim, obteve-se desse levantamento um índice de autor de 4,1 por artigo obtido pela razão do número de autores por artigo. O índice de 4,55 coautores por artigo foi calculado dividindo o número de aparições do autor pelo número de artigos expressando, portanto, a de coautores por artigo. Neste caso, o índice levou em conta as aparências do autor, enquanto para os “autores por artigo” um autor, mesmo que tenha publicado mais de um artigo, é contado apenas uma vez. Por esse motivo, o Índice de coautores por Artigo \geq Índice de autores por artigos. O Índice de Colaboração (IC) de 4,28 foi calculado como Total de Autores de Artigos Multi-Criados dividido pelo Total de Artigos Multi-Autorizados, ou seja, o índice de colaboração é um índice de coautores por artigo calculado utilizando o conjunto de artigos de várias autorias. Estes resultados mostram a grande cooperação existente entre os autores sobre a temática (Quadro 2).

Figura 2 - Fluxograma esquemático da metodologia aplicada para o levantamento bibliométrico sobre amostragem espacial na área de Geociências e afins, na base de dados Scopus



Fonte: Elaboração do autor.

Quadro 2 - Informações gerais sobre as publicações

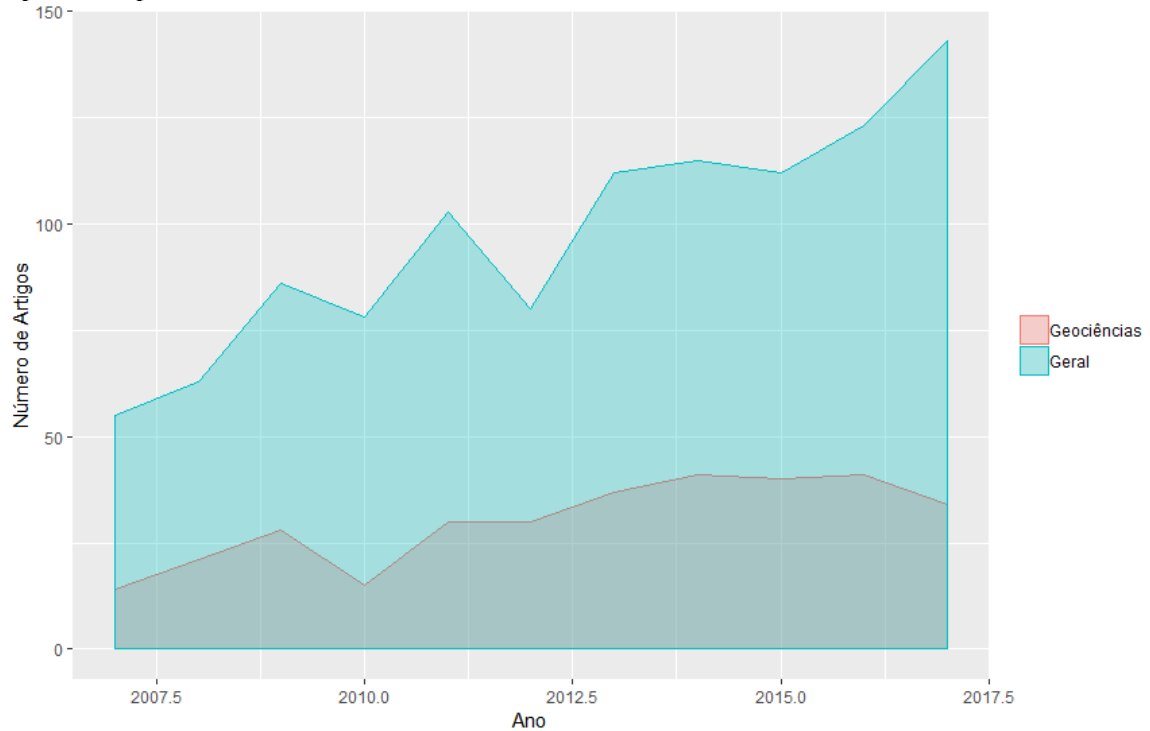
Total de artigos	331
Artigos com um único autor	15
Artigos com mais de um autor	1348
Fontes (periódicos revisados por pares)	180
Palavras-chave Plus (ID)	3370
Palavras-chave do autor (DE)	1187
Período	2007 - 2017
Média de citações por artigos	16,32
Autores	1363
Aparições do autor	1506
Autores de artigos de autoria única	15
Autores de artigos com vários autores	1348
Artigos por Autor	0,243
Autores por artigo	4,12
Co-autores por artigos	4,55
Índice de Colaboração	4,28

Fonte: Elaboração do autor.

A evolução temporal de todos os artigos publicados no ano de 2007 a 2017 sobre a temática amostragem espacial é mostrada na Figura 3. As primeiras pesquisas sobre o tema tiveram seu início em 1990 com Cressie, desde então, a quantidade de artigos que cita e estuda a temática tem aumento, o que também pode ser visto pelo ponto de corte dado a esta pesquisa, 2007 a 2017. Esta evolução positiva no número de artigos ao longo dos anos reflete um interesse crescente neste campo de pesquisa. Essa tendência crescente também é observada nas Geociências e áreas correlatas, apesar da pequena redução no último ano (Figura 3).

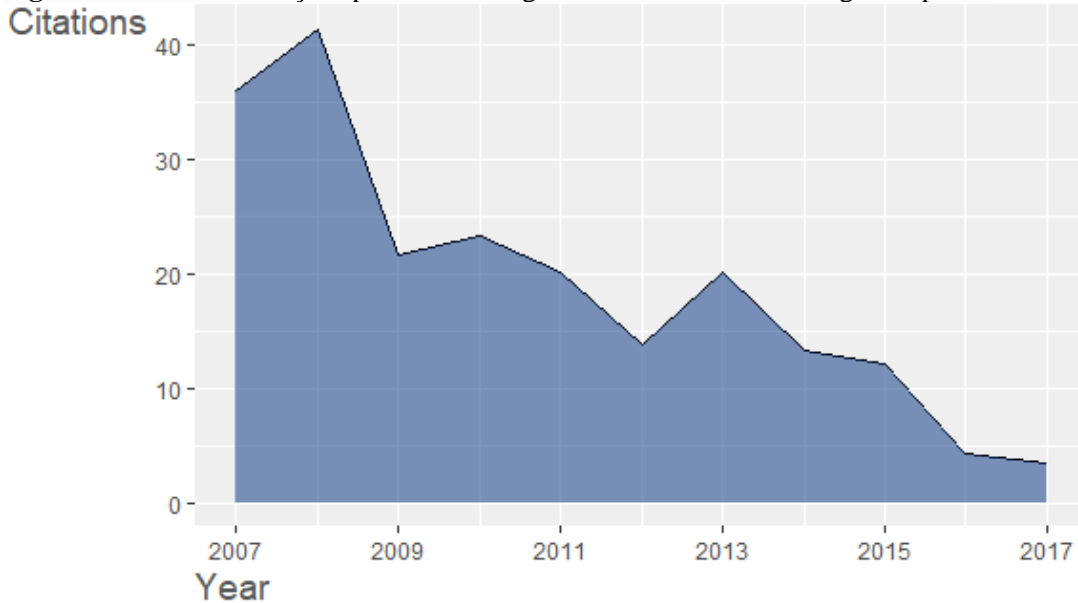
Algo importante a se observar é a tendência que citações dos artigos ao longo do tempo. Nota-se que ao longo do tempo a média de citações de artigos por ano se comportou de forma decrescente (Figura 4). A ocorrência desta tendência indica a presença de novos autores surgiram e com isso artigos mais antigos passaram a ser menos citados.

Figura 3 - Evolução da publicação por ano da base de dados Scopus para busca por amostragem espacial no período de 2007 – 2017



Fonte: Elaboração do autor.

Figura 4 - Média de citações por ano de artigos relacionados à amostragem espacial

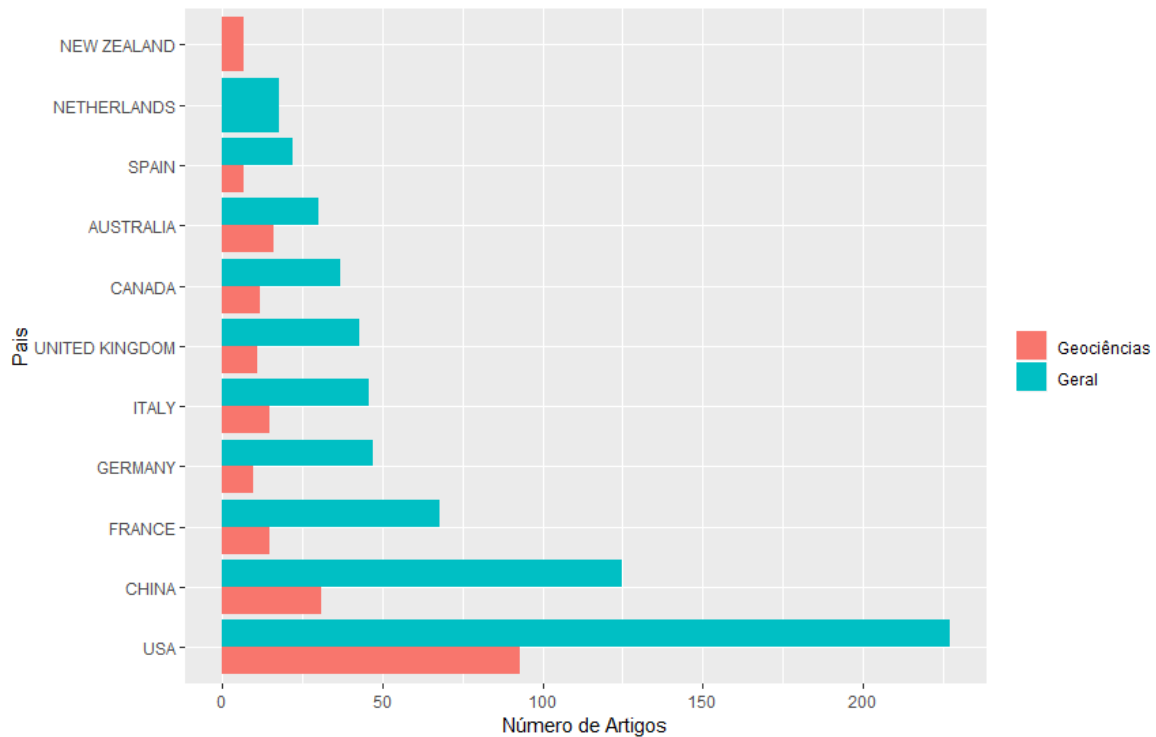


Fonte: Elaboração do autor.

Os dez países mais produziram artigos sobre o tema pesquisa está representado na Figura 5. A maior parte das publicações nesse campo de estudo são originadas dos centros de pesquisas dos Estados Unidos da América (227 artigos, 27,19%), seguidos pela China

(125 artigos, 14,97%), França (68 artigos, 8,14%) e Alemanha (47 artigos, 5,63%). Esses quatro países totalizam juntos a maioria das publicações. Essa mesma tendência ocorre nas Geociências e áreas correlatas: EUA (93 artigos, 34,06%), China (31, 11,3%), Austrália (16 artigos, 5,86%) e França (15 artigos, 5,49%) (Figura 5). Essa produtividade ficará melhor ilustrada refletida quando analisar a formação das redes de colaboração.

Figura 5 - Distribuição do número de artigos sobre amostragem espacial com base no país de origem do autor principal, geral na base Scopus e especificamente em Geociências e correlatas



Fonte: Elaboração do autor.

O Quadro 3 apresenta informações sobre os dez periódicos com maior número de artigos com a lista dos 10 principais periódicos com publicações sobre o tema “amostragem espacial” é mostrada no Quadro 3. Os 10 principais periódicos incluíram aproximadamente 26% do total de publicações ($n = 331$, 26,28%). As revistas científicas “Plos One” ($n = 19$, 5,74%) e “Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural” ($n = 13$, 3,93%) foram os dois principais periódicos, contribuindo com mais de 10 documentos cada. Isso mostra que, com o tema de “amostragem espacial”, há possibilidades de publicação em diversos periódicos na área de Geociências ou áreas correlatas.

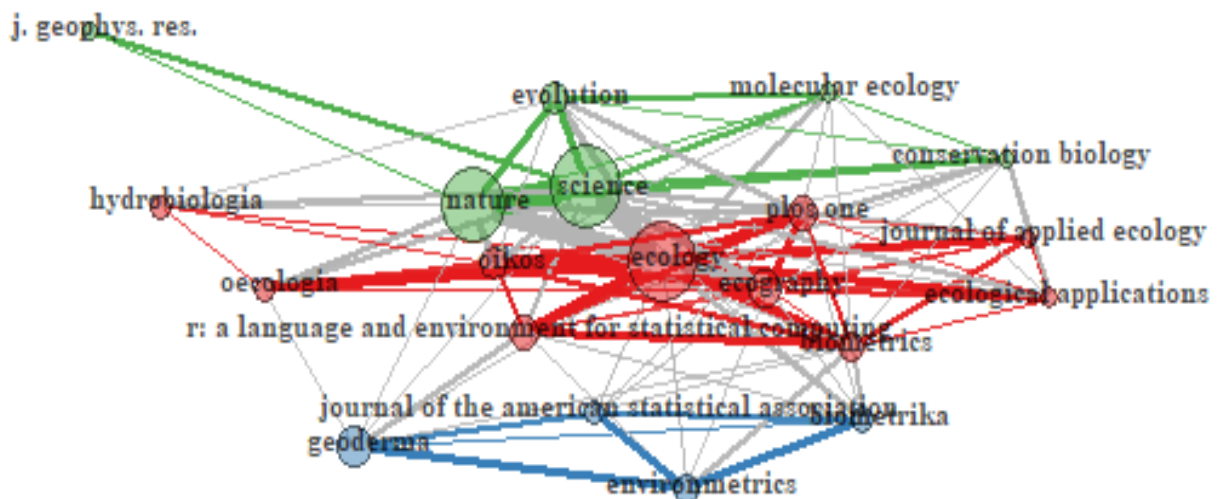
Quadro 3 – Periódicos por número de artigos sobre “amostragem espacial” da base de dados Scopus no período de 2007 e 2017

Periódicos	Número de artigos	%
PLOS ONE	19	5,74
NONGYE GONGCHENG XUEBAO/TRANSACTIONS OF THE CHINESE SOCIETY OF AGRICULTURAL ENGINEERING	13	3,93
ENVIRONMETRICS	9	2,72
REMOTE SENSING OF ENVIRONMENT	9	2,72
AGRICULTURAL AND FOREST METEOROLOGY	7	2,11
ENVIRONMENTAL AND ECOLOGICAL STATISTICS	7	2,11
GEODERMA	7	2,11
ENVIRONMENTAL MONITORING AND ASSESSMENT	6	1,81
BIOLOGICAL CONSERVATION	5	1,51
ECOLOGY AND EVOLUTION	5	1,51

Fonte: Elaboração do autor.

A Figura 6 mostra um mapa da a rede de cocitação entre os 20 periódicos sobre amostragem espacial. Mapas de cocitação entre periódicos ilustram a impacto do periódico. O tamanho dos nós representando os periódicos autores é proporcional à contribuição absoluta de cada periódico (Quadro 3). Os laços relacionais (linhas) que unem os nós representam a cocitação entre os periódicos específicos. As larguras dos laços relacionais que ligam os periódicos representam o número de citações entre eles. Nota-se a formação de três grandes grupos (identificadas pelas cores verde, vermelha e azul) mostrando uma alta conexão entre os principais periódicos neste campo.

Figura 6 – Mapa da rede de cocitação entre os períodos da base de dados Scopus referente a área de geociência e correlatas sobre amostragem espacial no período de 207 a 2017



Fonte: Elaboração do autor.

Os 10 autores com mais artigos publicados sobre “amostragem espacial” e maior índice H são mostrados no Quadro 4.

Quadro 4 - Os 10 principais autores sobre a temática “amostragem espacial” e o índice h, considerando a base Scopus para artigos relacionados a geociências e áreas correlatas no período 2007 a 2017

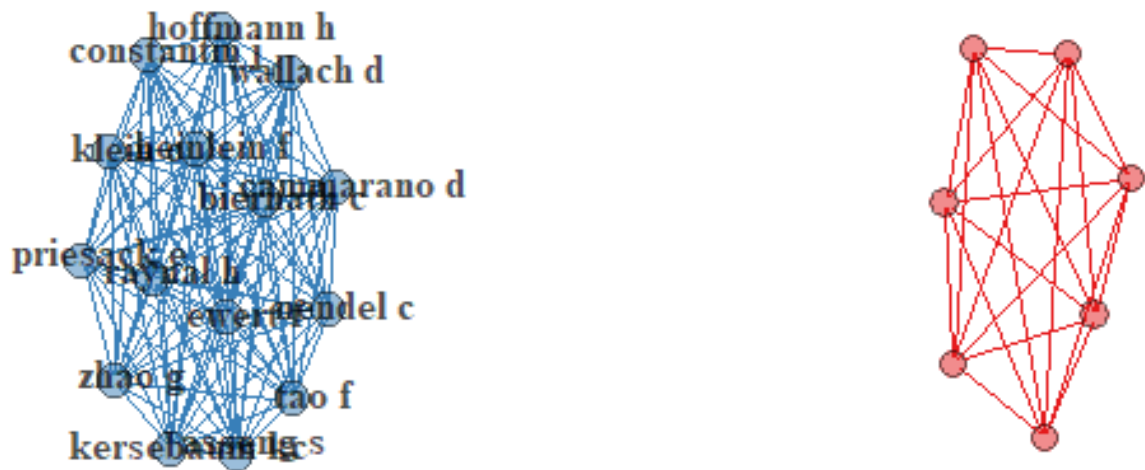
Autor	Número de artigos	%	Índice h
YANG J	6	1,81	4
LIU J	4	1,21	3
WANG D	4	1,21	3
WANG JF	4	1,21	4
ZHANG H	4	1,21	4
ALBRECHT C	3	0,91	3
BARET F	3	0,91	3
CESCATTI A	3	0,91	3
CHEN J	3	0,91	3
GRAFSTRM A	3	0,91	3

Fonte: Elaboração do autor.

O índice H é uma métrica de nível de autor que tenta medir o impacto da produtividade e da citação das publicações de um cientista ou acadêmico. Esses 10 autores, ao longo dos anos de 2007 a 2017, tiveram impacto na sua produtividade e em citações com índices ente 3 e 4. A maioria dos autores são chineses.

De acordo com van Eck e Waltman (2014), através da análise de redes de coautoria é possível identificar como pesquisadores, instituições de pesquisa ou países se relacionam de acordo com a quantidade de estudos que realizam e publicam conjuntamente. Visando a identificar as características de colaboração entre os autores mais relevantes, a Figura 7 apresenta as principais redes de coautoria mapeadas da base de dados Scopus nas áreas de geociências e correlatas no período entre 2007 e 2017. O mapa mostra que os autores mais produtivos sobre a temática trabalham em uma única rede, com isso, esclarece a quantidades de artigos publicados em uma década sobre a temática entre os autores.

Figura 7 - Mapa da rede de coautoria entre os períodos da base de dados Scopus referente a área de geociência e correlatas sobre amostragem espacial no período de 2007 a 2017



Fonte: Elaboração do autor.

Analisando a ocorrência das palavras-chave encontradas nos títulos e no resumo dos autores, os termos “spatial sampling” (amostragem espacial), “remote sensing” (sensoriamento remoto), “sampling” (amostragem), “geostatistics” (geoestatísticas), “uncertainty”, (incerteza), “kriging” (krigagem), “monitoring” (monitoramento), “GIS” (Sistema de Informações Georreferenciadas), “spatial variability” (variação espacial) e “soil moisture” (umidade do solo) são os que apresentaram alta ocorrência (Quadro 5).

Quadro 5 - Palavras-chaves com maior frequência nos títulos, resumo do autor e no manuscrito da base Scopus para artigos relacionados a geociências e áreas correlatas no período 2007 a 2017

Palavras-chave (autor)	Número de Artigos	Palavras-chave (manuscrito)	Número de Artigos
SPATIAL SAMPLING	42	SAMPLING	110
REMOTE SENSING	13	SPATIAL ANALYSIS	60
SAMPLING	11	ARTICLE	50
GEOSTATISTICS	9	SPATIAL SAMPLING	44
UNCERTAINTY	8	UNITED STATES	42
KRIGING	7	REMOTE SENSING	39
MONITORING	7	ENVIRONMENTAL MONITORING	38
GIS	6	ANIMALS	34
SPATIAL VARIABILITY	6	UNCERTAINTY ANALYSIS	31
SOIL MOISTURE	5	ANIMAL	28

Fonte: Elaborada pelo autor.

4.5 CONCLUSÃO

Este artigo apresentou um panorama abrangente das publicações realizadas sobre “amostragem espacial”. As publicações compreendem o período entre 2007 e 2017, com um total de 331 artigos científicos publicados em 180 periódicos indexados na base de dados Scopus, considerando a área de Geociências e correlatas. Destaca-se que as Ciências Agrárias possuem várias publicações sobre o assunto, referindo-se às amostragens de solos para fins de cultivo agrícola, silvicultura ou pecuária.

Os países com maior número de autores que publicam sobre o assunto são os Estados Unidos da América (227 artigos, 27,19%) e a China (125 artigos, 14,97%). Muitos autores ou coautores trabalham em rede, evidenciando o potencial de aumento no número de publicações através dessas interações. Entre os periódicos que mais publicam trabalhos nesse campo de pesquisa, a revista Plos One encontra-se em destaque, com um total de 19 trabalhos, e a revista Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering apresenta 13 publicações.

Em mapeamento, as palavras-chave que mais aparecem nos resumos e títulos dos trabalhos relacionados a amostragem espacial são “remote sensing” (sensoriamento remoto), “sampling” (amostragem), “geostatistics” (geoestatísticas), “uncertainty”, (incerteza), “kriging” (krigagem), “monitoring” (monitoramento), “GIS” (Sistema de Informação Georreferenciada), “spatial variability” (variação espacial) e “soil moisture” (umidade do solo). Esses termos facilitam a busca de artigos na base de dados da Scopus ou outras bases, representando os temas mais significativos para estudos sobre amostragens espaciais em Geociências ou afins.

5 IDENTIFICAÇÃO DE PADRÃO ESPACIAL EM DADOS GEOQUÍMICOS INORGÂNICOS – ESTUDO DE CASO: MADRE DE DEUS, BAHIA, BRASIL

RESUMO

Este estudo tem por objetivo identificar a existência de padrão espacial de parâmetros geoquímicos inorgânicos e de metais em solos ou sedimentos superficiais encontrados no município de Madre de Deus, Bahia, através de teste de aleatorização e da análise multivariada (Análise Fatorial), buscando entender o comportamento desses parâmetros. Foram utilizados dados analíticos de 96 pontos amostrais, georreferenciados, distribuídos em manguezais, apicuns, áreas urbanizadas e encostas não urbanizadas do Município de Madre de Deus. Através da análise fatorial robusta gerou quatro fatores que juntos explicam 81,6% da variação total dos dados. Haja vista os dados não apresentarem distribuição normal, optou-se, então, pelo teste de aleatorização de Mantel, para identificar a existência de padrão espacial. De acordo com o teste, metais, pH e Eh e nutrientes e M.O. apresentam padrão espacial, indicando que estas concentrações são autocorrelacionadas espacialmente.

Palavras-chave: Geoquímica inorgânica, Análise fatorial Robusta, Geostatística, Padrão Espacial, Interpolação Espacial.

5.1 INTRODUÇÃO

Avaliar as variações espaciais das características/propriedades de solos/sedimentos tem recebido grande atenção nos estudos ambientais e na área de Geociências devido à necessidade de produção de mapas ou superfícies contínuas de respostas confiáveis e de compreensão da distribuição espacial oriunda de fenômenos ocorridos no espaço geográfico, que possam colaborar com uma melhor interpretação do fenômeno em estudo (CAFARELLI et al., 2015; MANZIONE et al., 2005). Em geral, estudos sobre distribuição ou variação espacial das características/propriedades do solo são tipicamente baseados em um conjunto finito de localizações georreferenciadas obtidas através de técnicas de amostragem direta de campo e análises laboratoriais. No entanto, os custos de amostragem e de trabalhos analíticos são onerosos, o que dificulta obter espécimes suficientes para caracterizar com precisão a variabilidade espacial do ambiente (CAFARELLI et al., 2015). Alguns autores têm demonstrado que os atributos de sedimento/solo, muitas vezes, não revelam uma variação puramente aleatória ao longo de um terreno, apresentando padrão espacial não aleatório (GOMES et al., 2007).

No entanto, construir uma superfície de resposta para estimar os valores dessas propriedades/características de solos/sedimentos em locais não amostrados requer

conhecimento sobre o fenômeno em estudo e sobre as características geográficas da área. Como no geral, em estudos geoquímicos, esses dados de campos são amostras pontuais georreferenciadas. Os valores de um atributo em pontos não amostrados precisam ser estimados para gerar dados no espaço contínuo. Li e Andrew (2011 e 2014), em uma revisão abrangente sobre a temática, afirmam que nos casos observados por eles os métodos de interpolação espacial fornecem uma ferramenta eficaz para estimar os valores de uma variável ambiental em locais não amostrados, utilizando dados de observações pontuais.

Além do conhecimento dos métodos de interpolação existentes, estudos das variações das características/propriedades de solos/sedimentos requerem empregar técnicas estatísticas adequadas, pois as técnicas clássicas de análise estatística são construídas com o pressuposto de que as amostras são aleatórias, independentes e identicamente distribuídas (i.i.d.), o que não é válido para dados com estrutura de padrão espacial e/ou temporal. Quando existe evidência de padrão espacial e/ou temporal, é esperado que as observações mais próximas sejam mais parecidas que as distantes, logo, existem correlação entre elas (VIOLA, 2007).

Dessa forma, a estatística espacial tem sido utilizada como importante ferramenta de análise dos dados correlacionados, a fim de modelar e estudar a estrutura de padrão espacial dos atributos de solos/sedimentos, comumente avaliada por meio do ajuste do semivariograma experimental e teórico (MANZIONE; ZIMBACK, 2011).

Quando o interesse está em identificar o padrão espacial oriundo do fenômeno em estudo, diversas técnicas de estatística espacial são utilizadas. A existência de padrão espacial pode ser verificada por indicadores que confirmam a veracidade desse padrão espacial não aleatório a partir de uma comparação dos valores de uma amostra e de seus vizinhos. Por exemplo: quando o interesse está na obtenção de superfícies contínuas para dados pontuais, obtidas através da geoestatística, o padrão espacial pode ser constatado por intermédio do semivariograma com o envelope simulado ou teste de aleatorização de Mantel (RIBEIRO JÚNIOR et al., 2009); enquanto que para dados de área, utiliza-se o índice de Moran (MORAN, 1950).

A questão é como verificar a existência de padrão espacial em dados ambientais, visto que estudos ambientais envolvem grande quantidade de variáveis. Alguns autores sugerem o uso de técnicas multivariadas em concomitância com análise espacial. Essa integração auxilia no tratamento de grande volume de variáveis. Nesse ínterim, a associação da geoestatística com a análise multivariada (análise fatorial) busca entender a interação entre as variáveis, estabelecendo, para esse fim, um conjunto de fatores, calculados pela combinação linear dos dados originais (FIGUEIREDO FILHO; SILVA JÚNIOR, 2010).

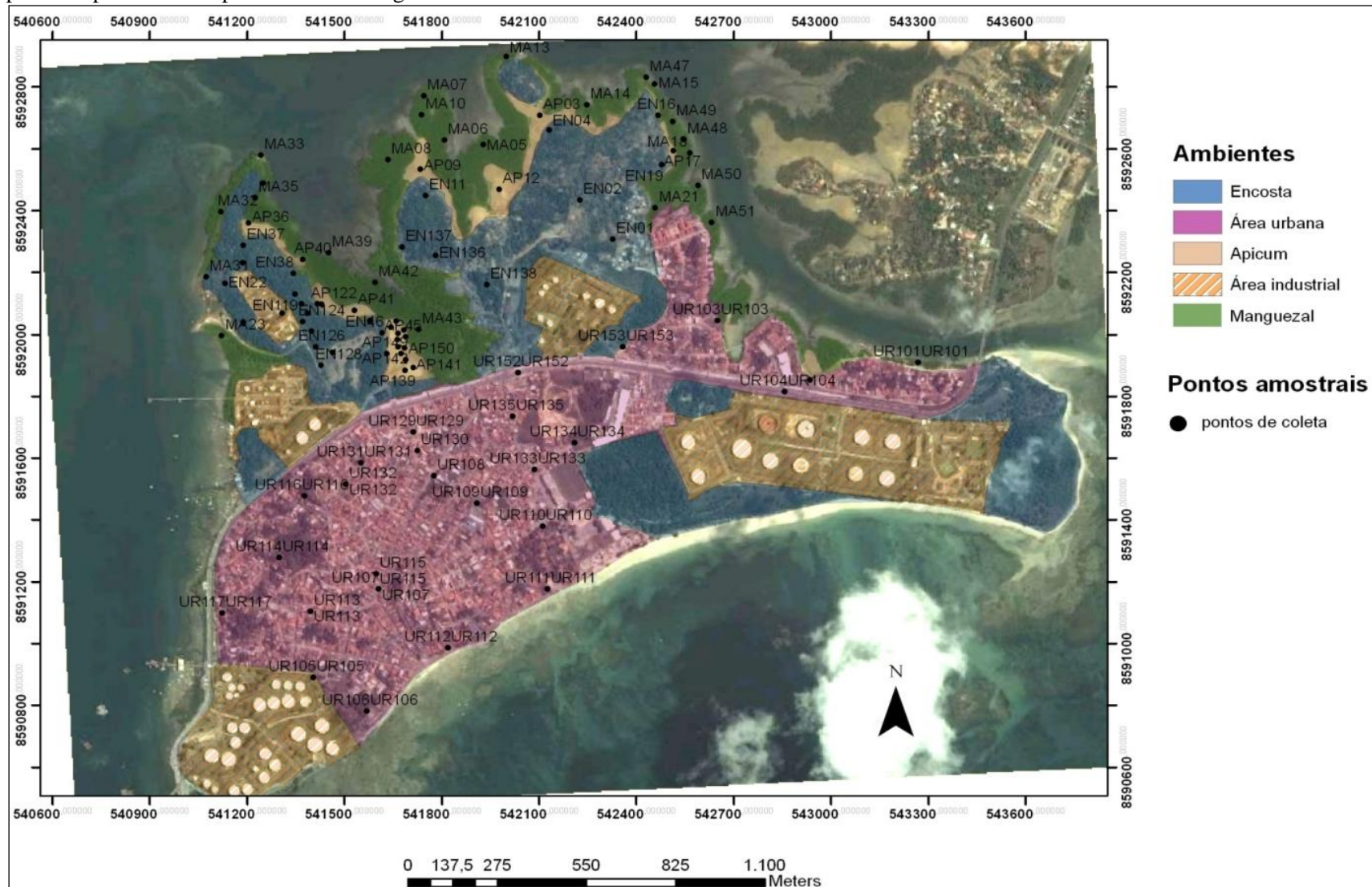
Silva (2012) estudou geoquimicamente o fracionamento de metais traços em diferentes tipos de ambientes (manguezais, apicuns, encostas não urbanizadas e áreas urbanizadas), na região NE da Baía de Todos os Santos - Bahia (Município de Madre de Deus). Entretanto esse autor encontrou dificuldades para caracterizar diferenças significativas em relação a diversos parâmetros físico-químicos e concentração de metais obtidos nos ambientes estudados.

No intuito de auxiliar na compreensão de problemas geoquímicos verificados com a manipulação de muitos resultados analíticos, a exemplo daquele verificado por Silva (2012), foi empregado o método de identificação de padrão espacial empregado por Viola (2007). Assim, este trabalho relacionando resultados de parâmetros físicos e químicos (pH; Eh, capacidade de troca catiônica (CTC); salinidade; granulometria, nas frações areia, silte e argila; matéria orgânica (MO); fósforo assimilável (P); nitrogênio (N); cobre (Cu); zinco (Zn); manganês (Mn); ferro (Fe); cromo (Cr)), e padrão espacial, objetivou contribuir com soluções que permitissem ampliar as certezas das interpretações geoquímicas quando são avaliadas as relações entre dados analíticos e ambientes de obtenção desses parâmetros.

5.2 MATERIAIS E MÉTODOS

Os dados geoquímicos utilizados para a identificação do padrão espacial foram obtidos na região NE da Baía de Todos os Santos-Bahia, nas cercanias do Município de Madre de Deus (SILVA, 2012) (Figura 9). Para a coleta das amostras, a região foi dividida em quatro ambientes distintos, designados: manguezal, apicum, encosta (zonas não urbanizadas) e urbano. Foram coletadas, no total, 96 amostras de sedimentos superficiais (entre 0 a 5 cm de profundidade), assim distribuídas: manguezal, n = 24; apicum, n = 19; encosta, n = 27; urbano, n = 26. Esse material coletado foi conduzido e analisado nas unidades laboratoriais “Laboratório de Estudos de Manguezal-(LEM)” e “Laboratório de Estudos do Petróleo-LEPETRO”, do Núcleo de Estudos Ambientais (NEA) do Instituto de Geociências (IGEO) da Universidade Federal da Bahia (UFBA).

Figura 9 – Imagem aérea do município de Madre de Deus, sobre a qual se pode diferenciar os ambientes manguezal, apicum, encostas e áreas urbanas. Os pontos representam os pontos de amostragem de solos ou sedimentos



Fonte: Assumpção (2014).

5.2.1 Dados geoquímicos (parâmetros físicos e químicos)

Os parâmetros selecionados foram escolhidos em função de sua importância geoquímica para os ambientes estudados (manguezais, apicuns, encostas não urbanizadas e áreas urbanizadas). Os dados utilizados foram oriundos da pesquisa de Silva (2012), coletados no Município de Madre de Deus em 2011. São eles: pH; Eh; capacidade de troca catiônica (CTC); salinidade; granulometria, nas frações areia, silte e argila; matéria orgânica (MO); fósforo assimilável (P); nitrogênio (N); cobre (Cu); zinco (Zn); manganês (Mn); ferro (Fe); cromo (Cr).

5.2.2 Análise exploratória de dados e Análise multivariada

Para a análise exploratória, empregando aos dados geoquímicos, foram realizadas as seguintes etapas: a) medidas descritivas (média, mediana desvio padrão, assimetria, curtose, coeficiente de variação); b) testes não-paramétrico para diferenças de grupos de Kuskal-Wallis, devido os dados não terem distribuição normal; c) verificação da existência de valores discrepantes: *outliers* (valores extremos) e *leverage* (pontos de alavancagem); d) testes de normalidade multivariada (MVN). Todos os teste foram avaliados ao nível de significância de 5% ($\alpha = 0,05$).

Um processo alternativo de mensuração é identificar variáveis que apresentam a mesma estrutura subjacente. Tecnicamente, isso pode ser implementado através de técnica multivariada de agrupamento das variáveis, por exemplo, Análise de Componentes Principais (ACP) e Análise Fatorial (AF). Uma das técnicas estatísticas multivariadas de grande aplicabilidade em Geociências é a ACP, a qual se constitui na mais popular forma de extração de fatores, sendo essa fermenta encontrada em quase todos programas estatísticos. No entanto, a comparação entre a AF e a ACP mostra que essas duas técnicas multivariadas diferem em relação ao método de extração. Enquanto a AF tem o objetivo de representar com precisão as correlações fora da diagonal entre as variáveis como dimensões latentes subjacentes, a ACP explica apenas a variação na matriz de escores brutos para menor número possível de componentes (FIGUEIREDO FILHO; SILVA JÚNIOR, 2010).

A principal função das diferentes técnicas de análise fatorial é reduzir uma grande quantidade de variáveis observadas a um número reduzido de fatores. Esses fatores representam as dimensões latentes (construtos) que resumem ou explicam o conjunto de

variáveis observadas, desse modo, obtém-se dimensões latentes que descrevem os dados em menos conceitos do que as variáveis individuais originais (HAIR et al, 2006, p. 91).

A Análise Fatorial (AF) é obtida com base na estimativa da matriz de correlação. Uma vez que a matriz esteja disponível, as cargas e os fatores, a serem considerados na análise, são determinados. Eles são geralmente estimados para a matriz de correlação de amostras, que é a matriz de covariância da amostra empírica padronizada pelas variâncias empíricas. Essa abordagem clássica é adequada se os dados forem provenientes de uma população com distribuição normal multivariada. No entanto, se esse pressuposto não for atendido, a matriz de correlação estimada pode ser severamente enviesada (TREIBLMAIER; FILZMOSER, 2010).

Várias abordagens alternativas foram sugeridas para estimar a matriz de correlação confiável na violação do pressuposto de normalidade multivariada. Uma maneira proeminente é utilizar o estimador da Determinante de Covariância Mínima (DCM), que procura um subconjunto de observações com determinante mínimo da matriz de covariância empírica, que é então multiplicada por um fator de consistência. A matriz de correlação robusta é obtida pela padronização das covariâncias robustas com a raiz quadrada dos elementos diagonais. O estimador DCM é atraente por sua alta robustez, especialmente porque há um algoritmo rápido disponível para uso em vários pacotes de software estatístico amplamente utilizados, como, por exemplo, o software R (TREIBLMAIER; FILZMOSER, 2010).

Neste trabalho utilizaremos a Análise Fatorial Robusta (ZHANG; RONG; LI, 2017) para realizar os agrupamentos dos parâmetros físicos e químicos. A análise foi realizada utilizando o Software R por meio do pacote Facov proposto por Zhang (2013).

5.2.4 Análise de padrão espacial de dados

Na análise para identificação da existência de padrão espacial aleatório dos parâmetros físicos e químicos obtidos na região NE da Baía de Todos os Santos-Bahia, nas cercanias do Município de Madre de Deus utilizou-se o teste de aleatorização de Mantel para detecção de padrão espacial (VIOLA, 2007).

Os testes aleatorizados, no geral, apresentam algumas vantagens em relação aos testes estatísticos clássicos, pois são computacionalmente fáceis de serem calculados, são baseados em estatísticas não padronizadas e não necessitam de informações prévias a respeito da população da qual a amostra foi retirada (MANLY, 2006; VIOLA 2007). Esses testes são mais justificados quando as amostras analisadas são aleatórias ou o próprio delineamento do

experimento sugere um teste de aleatorização (VIOLA, 2007). No contexto de dados espaciais, em geral, deseja-se testar a hipótese nula de padrão espacial aleatório versus a alternativa de padrão espacial não aleatório. Um teste para tal hipótese foi utilizado estudar, a partir da configuração dos pontos observados, se a dispersão dos parâmetros físicos e químicos obtidos na região NE da Baía de Todos os Santos-Bahia, nas cercanias do Município de Madre de Deus segue um padrão aleatório.

Desse modo, com os fatores obtidos pela AF e no intuito de utilizar a geoestatística para representação espacial, realizaram-se algumas verificações primordiais: análise descritiva espacial para verificar tendências; análise de padrão espacial, teste de aleatorização de Mantel; e, com base nos resultados do teste de aleatorização de Mantel, aplicar interpolação que melhor se adequa aos resultados obtidos para a construção do mapa. Todas as análises foram realizadas utilizando o programa R versão 3.2.1 (R DEVELOPMENT CORE TEAM, 2015), para a interpolação utilizou-se o pacote de dados geoR (RIBEIRO JÚNIOR; DIGGLE, 2001) (Apêndice C).

5.3 RESULTADOS E DISCUSSÕES

Nas Tabelas 1 a 3 são apresentados os resultados das medidas descritivas dos parâmetros físicos e químicos obtidos na região NE da Baía de Todos os Santos-Bahia, nas cercanias do Município de Madre de Deus, estratificada pelo ambiente em análise (apicum - ap; manguezal - ma; encosta - en.; e urbano - ur).

Verifica-se na Tabela 1, quando são exibidos os dados dos parâmetros referentes ao pH, Salinidade, Eh, P, N, M.O., CTC, que as medidas de tendência central (média e mediana) são semelhantes em aproximadamente 50% dos parâmetros, indicando assimetria desses em metade dos ambientes analisados. Essa indicação pode ser observada quando analisado o Coeficiente de Variação (C.V.) que é uma medida relativa que avalia a dispersão dos dados. Warrick e Nielson (1980) propuseram a seguinte classificação para o C.V.: baixa variação ($C.V.\% < 12\%$), variação moderada ($12\% \leq C.V.\% \leq 60\%$) e alta variação ($C.V.\% > 60\%$). De acordo com essa classificação, 48% dos parâmetros, em relação aos ambientes, analisados apresentam alta variabilidade, enquanto apenas 8,3% dos parâmetros, em relação aos ambientes, apresentam baixa variabilidade. Contudo, poucos parâmetros apresentam coeficiente de assimetria próximo a zero (33,3% dos parâmetros), ou seja, os dados não são simétricos indicando possível violação da suposição de normalidade.

Para verificar se os parâmetros em estudo 8 difere em relação ao ambiente em análise utilizou-se do teste não-paramétrico de Kruskal-Wallis. Os resultados dos testes revelam que os parâmetros analisados diferem em relação ao ambiente (p -valor < 0,05) (Tabela 1). Essas variações são esperadas devido às peculiaridades de cada ambiente, sinalizando que esses parâmetros podem se alterar em relação ao espaço.

Tabela 1 - Estatísticas descritivas dos parâmetros físicos e químicos determinados em solos ou sedimentos coletados na região NE da Baía de Todos os Santos – cercanias do município de Madre de Deus, Bahia

	Ambiente ¹	n	min - max	média	mediana	desvio padrão	C.V%*	Assimetria	Curtose	p -valor**
pH	ma	24	5,86 – 8,38	7,01	7,00	0,74	10,56	0,47	-0,73	0,0000
	ap	19	6,24 - 8,08	7,40	7,49	0,50	6,76	-0,71	-0,41	
	en	27	4,44 - 8,24	5,33	5,05	0,89	16,70	1,76	2,59	
	ur	26	5,43 - 8,11	7,04	7,30	0,85	12,07	-0,65	-0,89	
Salinidade	ma	24	30,00 – 70,00	49,25	50,00	11,67	23,70	0,14	-1,22	0,0000
	ap	19	64,00 – 100,00	93,21	100,00	10,80	11,59	-1,28	0,44	
	en	27	0,00 – 7,00	3,00	3,00	2,11	70,33	0,02	-1,38	
	ur	26	0,00 - 5,00	2,23	2,00	2,01	90,13	0,12	-1,67	
Eh	ma	24	-59 - 66	14,67	13,00	35,12	239,40	-0,54	-0,48	0,0000
	ap	19	-82 - 48	-25,58	-40,00	40,83	159,62	0,43	-1,29	
	en	27	-79 - 159	92,85	106,00	66,44	71,56	-1,14	0,40	
	ur	26	-83 - 71	-19,23	-36,00	48,78	253,67	0,58	-0,99	
P	ma	24	12,50 – 337,50	95,52	63,89	88,69	92,85	1,55	1,41	0,0000
	ap	19	0,02 – 54,01	61,69	51,47	43,66	70,77	0,57	-0,81	
	en	27	0,00 - 213,93	1,73	3,71	56,73	178,79	2,01	3,1	
	ur	26	2,48 - 1715,69	315,81	184,07	398,38	126,15	2,1	4,19	
N	ma	24	0,04 – 0,64	0,18	0,12	0,15	83,33	1,58	1,82	0,0000
	ap	19	0,04 – 0,21	0,07	0,08	0,04	57,14	2,27	5,89	
	en	27	0,04 - 0,36	0,15	0,14	0,08	53,33	0,72	0,13	
	ur	26	0,08 - 0,30	0,13	0,12	0,06	46,15	1,44	1,49	
MO	ma	24	1,36 – 11,43	4,23	3,89	2,51	59,34	0,91	0,41	0,0000
	ap	19	0,31 – 2,38	1,11	1,10	0,48	43,24	0,59	0,46	
	en	27	0,40 - 5,34	1,85	1,82	1,05	56,76	1,19	2,15	
	ur	26	0,25 – 4,40	2,24	2,50	1,17	52,23	-0,07	-1,11	
CTC	ma	24	11,56 - 42,77	20,34	18,06	7,52	36,97	1,23	1,09	0,0000
	ap	19	12,08 – 36,81	24,65	25,75	7,05	28,60	-0,06	-1,04	
	en	27	4,50 - 32,87	15,54	13,11	8,40	54,05	0,67	-0,84	
	ur	26	5,29 – 14,04	9,16	9,27	2,68	29,26	0,37	-0,98	

¹ ma - manguezal; ap - apicum; en - encostas não urbanizadas; ur - áreas urbanizadas

* C.V.% - coeficiente de variação em percentual

** p -valor do Teste de Kuskal-Wallis

Fonte: Elaboração do autor (2019).

Tabela 2 - Estatísticas descritivas das frações granulométricas dos solos ou sedimentos coletados na região NE da Baía de Todos os Santos – cercanias do município de Madre de Deus, Bahia

	Ambiente ¹	n	min - max	média	mediana	desvio padrão	C.V%*	Assimetria	Curtose	p-valor**
Areia	ma	24	20,59 – 93,19	78,54	82,72	15,86	20,19	-2,2	4,94	0,0000
	ap	19	73,72 - 97,28	90,61	93,66	6,93	7,65	-1,31	0,23	
	en	27	3,50 – 94,63	52,38	61,44	29,33	55,99	-0,53	-1,2	
	ur	26	24,60 – 92,11	74,96	77,16	15,73	20,98	-1,77	2,97	
Silte	ma	24	6,80 – 76,38	20,75	16,66	15,18	73,16	2,23	5,03	0,0000
	ap	19	2,72 – 25,38	9,20	6,33	6,70	72,83	1,31	0,23	
	en	27	5,36 – 92,99	46,53	40,10	28,41	61,06	0,51	-1,22	
	ur	26	7,78 – 73,67	24,46	22,30	15,30	62,55	1,79	3,04	
Argila	ma	24	0,02 - 3,03	0,60	0,39	0,69	115,00	2,06	4,19	0,0000
	ap	19	0,00 – 0,91	0,20	0,07	0,28	140,00	1,28	0,2	
	en	27	0,02 - 3,74	1,49	1,47	1,00	67,11	0,46	-0,53	
	ur	26	0,10 – 1,73	0,59	0,46	0,49	83,05	0,94	-0,17	

¹ ma - manguezal; ap - apicum; en - encostas não urbanizadas; ur - áreas urbanizadas

* C.V.% - coeficiente de variação em percentual

** p-valor do Teste de Kuskal-Wallis

Fonte: Elaboração do autor (2019).

Quanto a granulometria (areia, Silte e Argila) das amostras obtidos na região NE da Baía de Todos os Santos-Bahia, nas cercanias do Município de Madre de Deus (Tabela 2), observa-se que as médias e medianas das variáveis analisadas diferem em relação aos ambientes, que é comprovado pelo teste não-paramétrico de Kurskal_Wallis ($p\text{-valor} < 0,05$). No que tange a variabilidade dos dados (C.V.%), Silte e Argila apresentam alta variabilidade (C.V% >60%), nos ambientes analisados, já para Areia, apresenta variabilidade moderada entre os ambientes: mangue, encosta e urbano ($12\% \leq \text{C.V.}\% \leq 60\%$).

Para os metais analisados, Cu, Zn, Mn, Fe e Cr, as médias e medianas diferem em relação aos ambientes, além de apresentarem assimetria positiva, na maioria dos parâmetros analisados (coeficiente de assimetria > 0) (Tabela 3). Também se nota que os metais analisados diferem as suas quantidades quando comparado entre os ambientes analisados (teste de não-paramétrico de Kustal-Wallis, $p\text{-valor} < 0,05$).

Após as análises descritivas e dos testes de comparação entre os ambientes analisados, o passo seguinte, antes de aplicar o teste de aleatorização de Mantel, é verificar se os parâmetros físicos e químicos analisados apresentam distribuição normal. Para isso realizou-se o teste de normalidade univariado e multivariado. Foi realizado o teste de normalidade de univariado e multivariado de Shapiro-Wilks, com nível significância de 5%, em todas as variáveis em estudo (Tabela 4), quando foi verificado que esses parâmetros não atendem aos requisitos de normalidade.

Tabela 3 - Estatísticas descritivas dos metais Cu, Zn, Mn, Fe e Cr nos solos ou sedimentos coletados na região NE da Baía de Todos os Santos – cercanias do município de Madre de Deus, Bahia

	Ambiente	n	min - max	média	mediana	desvio padrão	C.V%*	Assimetria	Curtose	p-valor**
Cu	ma	24	1,49 – 39,23	10,85	8,20	8,46	77,97	1,93	3,39	0,0012
	ap	19	0,18 – 30,46	5,64	3,12	7,08	125,53	2,33	5,19	
	en	27	0,07 - 39,79	14,89	11,72	12,60	84,62	0,61	-1,11	
	ur	26	1,92 - 19,83	7,78	7,01	3,87	49,74	1,01	1,36	
Zn	ma	24	4,54 – 54,07	17,34	14,00	10,90	62,86	1,74	3	0,0006
	ap	19	4,26 – 27,78	12,20	11,61	6,24	51,15	1,14	0,59	
	en	27	4,19 - 157,13	28,74	26,37	30,19	105,05	2,8	9,28	
	ur	26	10,75 - 44,21	23,38	20,57	9,59	41,02	0,9	-0,28	
Mn	ma	24	6,08 – 82,27	25,87	16,92	19,63	75,88	1,3	0,72	0,0000
	ap	19	3,15 – 821,51	75,19	17,06	191,63	254,86	3,16	9,2	
	en	27	16,80 – 844,52	202,81	64,17	230,81	113,81	1,17	0,23	
	ur	26	10,78 – 299,47	69,85	48,29	71,43	102,26	1,99	3,14	
Fe	ma	24	1825,48 – 19081,00	6133,65	5285,43	3766,17	61,40	1,92	3,73	0,0000
	ap	19	1317,33 – 37446,61	4614,52	2247,83	8155,25	176,73	3,42	10,84	
	en	27	727,61 – 25728,10	5261,29	2479,01	5806,43	110,36	1,72	3,14	
	ur	26	663,27 - 5245,61	2074,22	1680,86	1235,48	59,56	1,13	0,22	
Cr	ma	24	1,25 – 22,94	6,16	4,70	5,41	87,82	1,19	1,32	0,0027
	ap	19	0,05 – 16,36	2,29	0,87	4,13	180,35	2,33	4,69	
	en	27	0,05 - 15,51	5,04	3,92	4,09	81,15	0,91	0,05	
	ur	26	0,05 – 29,16	4,46	3,19	6,15	137,89	2,55	7,33	

¹ ma - manguezal; ap - apicum; en - encostas não urbanizadas; ur - áreas urbanizadas

* C.V.% - coeficiente de variação em percentual

** p-valor do Teste de Kuskal-Wallis

Fonte: Elaboração do autor (2019).

Alguns pressupostos são importantes serem garantidos para não ter influência negativa nas análises. Segundo Figueiredo Filho e Silva Junior (2010) a não normalidade dos dados e a assimetria prejudica as estimativas da análise fatorial (AF) bem com as estimativas das técnicas geoestatística (YAMAMOTO; LANDIN, 2013). Como metodologia antes do emprego da análise fatorial, Zygmunt e Smith (2014) sugerem verificar previamente, como pressuposto, a normalidade multivariada, que forma a base para a estatística correlacional sobre a qual AF e vários procedimentos usados na análise de Máxima Verossimilhança (LM) se apoiam. Como os parâmetros físicos e químicos obtidos na região NE da Baía de Todos os Santos-Bahia tem métricas diferentes optou-se por fazer uma padronização dos dados, desse modo à influência das diferentes métricas utilizada para mensurar as variáveis em estudo.

Tabela 4 - Teste de normalidade univariado e multivariado (MVN) de Shapiro-Wilks para dados de solos e sedimentos coletados na região NE da Baía de Todos os Santos – cercanias do Município de Madre de Deus, Bahia

Variáveis	Estatística de teste	<i>p</i> -valor
pH	0,9348	0,0001
Salinidade	0,7832	<0,001
Eh	0,9446	0,0005
P	0,5129	<0,001
N	0,7729	<0,001
MO	0,8267	<0,001
Cu	0,7955	<0,001
Zn	0,6422	<0,001
Mn	0,5569	<0,001
Fe	0,6242	<0,001
Cr	0,7968	<0,001
CTC	0,9350	0,0001
Areia	0,7953	<0,001
Silte	0,7953	<0,001
Argila	0,8197	<0,001
MVN	0,41246	0,0000

Fonte: Elaboração do autor (2019).

Um total de 15 parâmetros físicos e químicos forma o conjunto de variáveis analisadas. Para melhor compreender como estas variáveis se correlacionam, calculou-se o coeficiente de correlação de Spearman (Tabela 5), visto que, os parâmetros analisados não apresentam distribuição normal (Tabela 4). Observou-se, no geral, que a maioria das variáveis apresentou coeficientes de correlação abaixo de 0,5, indicando que as variáveis são pouco correlacionáveis, com exceção de pH x Eh, Salinidade x CTC, Areia X Silte, Areia x Argila, Silte x Argila, Cu x Zn e Zn x Mn. Analisando esse tratamento estatístico é possível visualizar quais grupos de variáveis serão agrupadas através da Análise Fatorial

Tabela 5 - Coeficiente de correlação de Spearman e teste de correlação para os parâmetros analisados em solos e sedimentos superficiais coletados na região NE da Baía de Todos os Santos – cercanias do Município de Madre de Deus, Bahia

	pH	Salinidade	Eh	P	N	MO	Cu	Zn	Mn	Fe	Cr	CTC	Areia	Silte
pH	1													
Salinidade	0,378	1												
Eh	-0,918	-0,318	1											
P	0,419	-0,014	-0,424	1										
N	-0,293	-0,320	0,206	0,219	1									
MO	-0,037	-0,119	0,042	0,249	0,460	1								
Cu	-0,210	-0,181	0,171	0,016	0,388	0,313	1							
Zn	-0,091	-0,340	-0,038	0,163	0,415	0,282	0,747	1						
Mn	-0,246	-0,440	0,147	0,027	0,273	0,076	0,575	0,638	1					
Fe	-0,005	0,255	0,174	0,007	0,020	0,201	0,361	0,004	-0,009	1				
Cr	-0,114	-0,094	0,137	0,057	0,332	0,285	0,720	0,543	0,383	0,475	1			
CTC	0,157	0,669	-0,135	-0,118	0,074	0,026	0,129	-0,026	-0,101	0,393	0,140	1		
Areia	0,567	0,5	-0,484	0,158	-0,389	-0,050	-0,522	-0,495	-0,540	-0,065	-0,460	0,156	1	
Silte	0,569	-0,511	0,482	-0,156	0,387	0,044	0,517	0,498	0,540	0,057	0,458	-0,166	-0,999	1
Argila	-0,538	-0,420	0,471	-0,175	0,366	0,011	0,580	0,453	0,509	0,163	0,471	-0,103	-0,928	0,924

Fonte: Elaboração do autor (2019).

Após a interpretação das correlações, foi iniciada a análise fatorial (AF). Para tanto, foi necessário calcular medidas de adequação da amostra, neste trabalho consideraremos os índice de Kaiser-Meyer-Olkin (KMO) e teste de esfericidade de Bartlett, para verificação da possibilidade da adequação dos parâmetros físicos e químicos ao emprego da AF (Tabela 6). Como regra para interpretação dos índices de KMO, valores menores que 0,5 são considerados inaceitáveis, valores entre 0,5 e 0,7 são considerados medíocres; valores entre 0,7 e 0,8 são considerados bons; valores maiores que 0,8 e 0,9 são considerados ótimos e excelentes, respectivamente (HUTCHESON; SOFRONIOU, 1999). Observou-se que os parâmetros físicos e químicos apresentam medidas de adequação KMO para os parâmetros superior a 0,50 e a medida de adequação KMO geral de 0,74, indicando que os parâmetros físicos e químicos podem ser agrupados através da Análise Fatorial (Tabela 6). Desse modo, os resultados obtidos com os parâmetros físicos e químicos indicam que se tem bons resultados como emprego da análise fatorial ao conjunto de dados. O teste de esfericidade de Bartlett ($p\text{-valor} < 0,05$) indicou que as correlações entre os itens são suficientes para a realização da AF.

Tabela 6 - Medida de adequação da amostra de Kaiser-Meyer-Olkin (KMO), para os parâmetros analisados em solos e sedimentos superficiais coletados na região NE da Baía de Todos os Santos – cercanias do município de Madre de Deus, Bahia

Variável	pH	Salinidade	Eh	P	N	MO	Cu	Zn	Mn	Fe
KMO	0,71	0,72	0,63	0,56	0,75	0,68	0,76	0,85	0,72	0,6

Variável	Cr	CTC	Areia	Silte	Argila
KMO	0,8	0,51	0,76	0,76	0,88

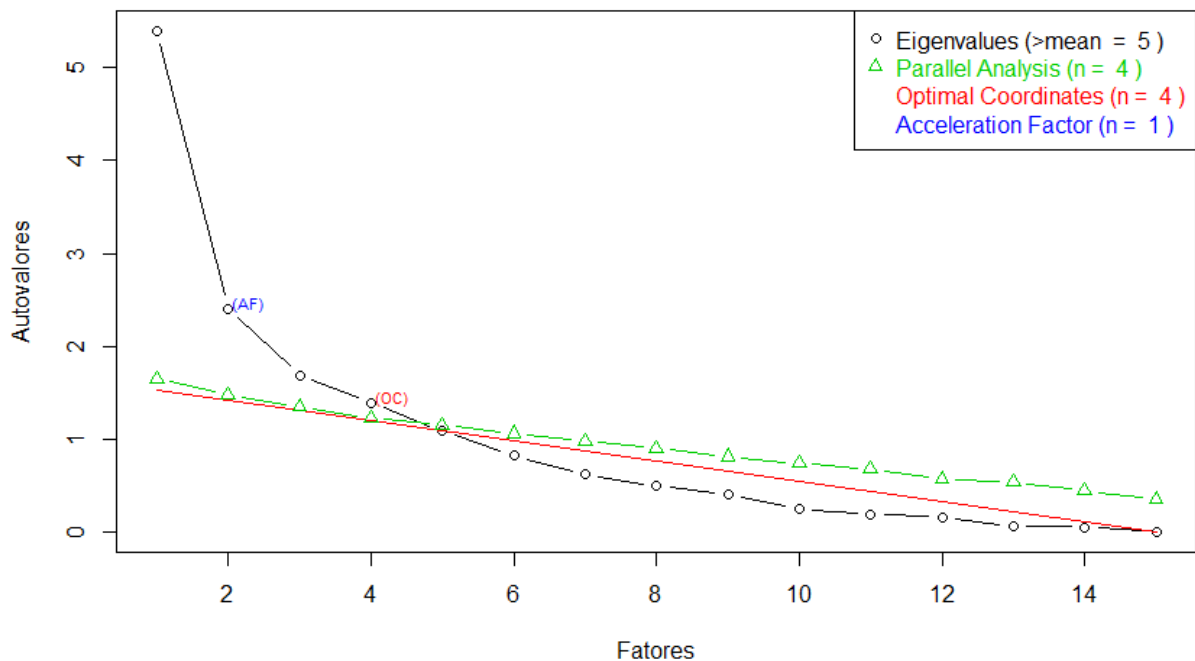
Teste de esfericidade de Bartlett, $p\text{-valor} = 0,000$

Fonte: Elaboração do autor (2019).

Verificada a adequação do emprego da AF, o passo seguinte foi determinar o número de fatores ou agrupamentos. Para tal utilizou-se o gráfico de *Screenplot*, que é um critério que permite a determinação do número de fatores (Figura 10). O resultado obtido permitiu que fossem considerados 4 fatores, sugeridos pelos testes da Análise Paralelo e Coordenadas Ótimo. O teste de coordenadas ótimas para determinar o número de fatores através da localização do *screen* medindo os gradientes associados a autovalores e suas coordenadas antecedentes. Já o teste da Análise do Paralelo leva em conta a proporção de variância

resultante de erro de amostragem, sendo considerado um dos testes mais fortemente recomendado para determinar o número de fatores (GARRIDO; ABAD; PONSODA, 2012).

Figura 10 - Gráfico dos testes para determinar o número de fatores para os parâmetros analisados em solos e sedimentos superficiais coletados na região NE da Baía de Todos os Santos – cercanias do município de Madre de Deus, Bahia

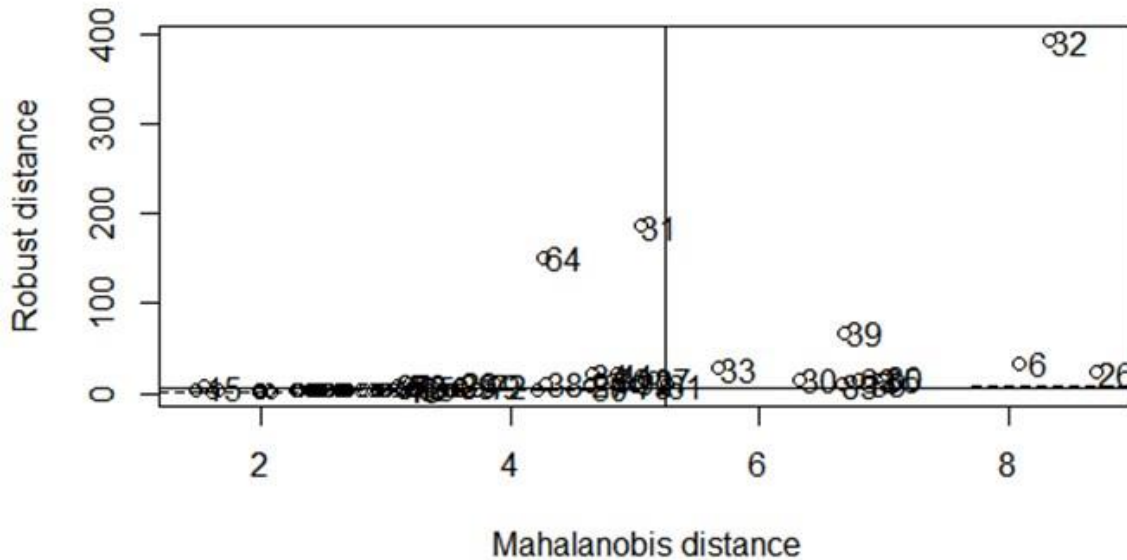


Fonte: Elaboração do autor (2019).

Com a finalidade de saber se os parâmetros físicos e químicos analisados apresentam valores discrepantes (outliers) foi utilizada a distância de *Mahalanobis* para identificar os pontos amostrais com valores discrepantes entre os dados (Figura 11). Quatro pontos amostrais, 31, 32, 39 e 64, são identificados como pontos discrepantes. Segundo Zygmunt e Smith (2014), um único *outlier* (valor discrepantes) pode distorcer as estimativas, tais como, coeficientes de correlação, medidas de congruência, a exemplo do de Cronbach, e estimativas dos fatores oriundos da análise fatorial. *Outliers* podem eventualmente levar à especificação de modelos incorretos.

Como o conjunto de dados geoquímicos analisados apresentou valores discrepantes, optou-se por se utilizar uma AF robusta, proposta por Pison et al. (2003). Esse método estatístico foi atualizado por Zang (2013). Deve ser destacado que os tratamentos estatísticos determinados por esses autores se mostram robustos na presença *outliers*, por isso têm sido amplamente empregados para o reconhecimento desse efeito.

Figura 11 - Gráfico da Distância de Mahalanobis para os pontos amostrais coletados na região NE da Baía de Todos os Santos – cercanias do município de Madre de Deus, Bahia



Fonte: Elaboração do autor (2019).

Após a constatação da necessidade do emprego dessa AF robusta, a análise procedeu utilizando o número de fatores igual a quatro, como foi constatado através da Figura 11, e método de extrações dos fatores o das componentes principais. Com base nesse procedimento os 4 fatores explicam 81,70% da variabilidade do conjunto dos dados geoquímicos tratados. Também é importante sublinhar que essa AF robusta permitiu que as 15 variáveis fossem agrupadas em quatro (Tabela 7).

A partir da Tabela 7 observou-se que: i) o Fator 1 apresenta como cargas fatoriais mais altas as variáveis salinidade, CTC, areia, silte e argila; logo, foi denominado Grupo Geoquímico “granulometria-salinidade”; ii) o Fator 2 apresentou cargas fatoriais positivas altas para Cu, Fe, Cr; sendo denominado Grupo Geoquímico “metais”; iii) o Fator 3 possui altas cargas positivas para pH e Eh, e foi nomeado Grupo Geoquímico “pH-Eh”; iv) o Fator 4 apresenta apenas uma carga fatorial alta para P, N, MO e Zn, portanto, recebeu o nome de Grupo Geoquímico “nutrientes-MO”.

Após análise fatorial foi realizado o teste de aleatorização de Mantel nos fatores obtidos na AF (Tabela 8). O teste de aleatorização de Mantel indicou que os fatores “metais”, “pH-Eh” e “nutrientes-MO” apresentam padrão espacial, ou seja, o comportamento da distribuição dos fatores não é aleatório.

Tabela 7 - Análise Fatorial, método de extração pelas componentes principais, dos parâmetros físicos e químicos coletados na região NE da Baía de Todos os Santos – cercanias do município de Madre de Deus, Bahia

Variáveis	Fator 1	Fator 2	Fator 3	Fator 4	Comunalidade
pH	0,367	0,035	-0,834	-0,134	0,858
Salinidade	0,823	0,128	-0,071	-0,293	0,797
Eh	-0,245	0,095	0,920	-0,001	0,911
P	-0,127	-0,059	-0,375	0,733	0,614
N	-0,110	0,293	0,283	0,835	0,869
MO	0,075	0,347	0,433	0,733	0,849
Cu	-0,143	0,819	-0,138	0,403	0,865
Zn	-0,471	0,325	-0,240	0,590	0,790
Mn	-0,490	0,116	-0,547	0,163	0,588
Fe	0,173	0,925	0,095	0,069	0,919
Cr	-0,132	0,901	0,038	0,104	0,836
CTC	0,834	0,154	-0,079	0,061	0,728
Areia	0,924	-0,066	-0,166	-0,131	0,906
Silte	-0,922	0,057	0,170	0,136	0,903
Argila	-0,856	0,292	0,037	-0,043	0,845
Cargas dos fatores	4,567	2,811	2,405	2,468	
% da Variância explicada	30,40%	18,70%	16,00%	16,50%	
% Acumulado da variância explicada	30,40%	49,20%	65,20%	81,70%	

Fonte: Elaboração do autor (2019).

Tabela 8 – Teste de aleatorização de Mantel para os fatores

	Correlação de Pearson	<i>p</i> -valor	Correlação de Spearman	<i>p</i> -valor
Fator1 – Granulometria – salinidade	0,09364	0,17830	0,05359	0,30420
Fator2 – Metais	0,23034	0,01170	0,18151	0,03970
Fator3 – pH-Eh	0,17899	0,04110	0,25597	0,00620
Fator4 – Nutrientes-MO	0,25760	0,00590	0,26542	0,00370

Fonte: Elaboração do autor (2019).

Com os resultados dos tratamentos estatísticos realizados sobre os parâmetros geoquímicos trabalhados, foram gerados os mapas de superfícies contínuas para os fatores analisados. Com base nos resultados do teste de aleatorização de Mantel (Tabela 8), os parâmetros granulométricos (fator 1) não apresentam padrão espacial, utilizou-se a interpolação espacial do interpolar Inverso da Distância Ponderada (IDP) para construção do mapa de superfície (Figura 12). Segundo Andriotti (2005) e Landim (2000), o IDP é um modelo determinístico indicado para áreas com alta densidade de pontos (amostras). Esse método baseia-se na Lei de Tobler, ou seja, estima as medidas desconhecidas através da

média ponderada das medidas conhecidas, dando maior peso a pontos mais próximos. Devido à simplicidade conceitual e facilidade de programação, é um dos métodos de interpolação mais difundidos (LONGLEY et al., 2013).

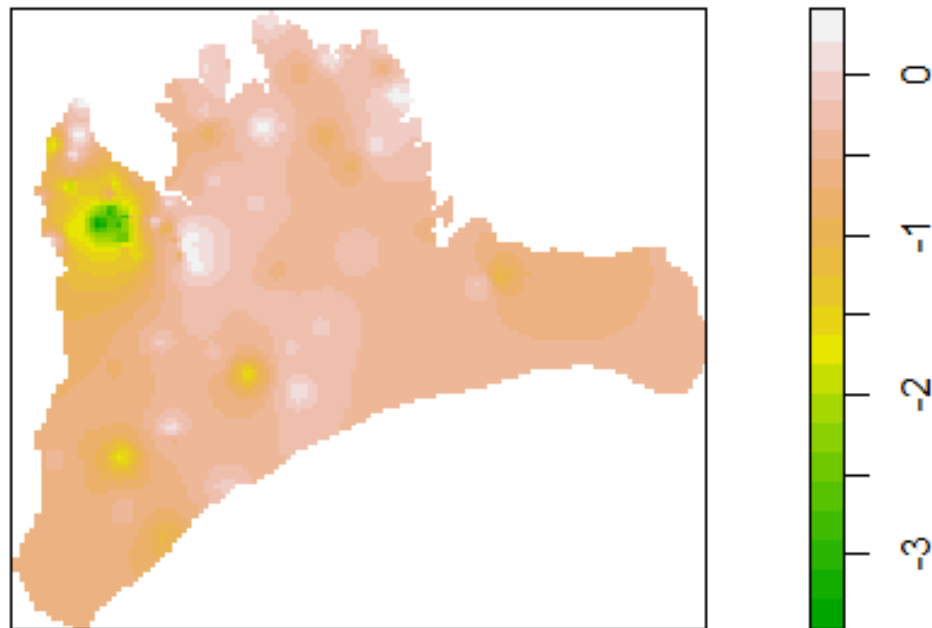
O modelo gerado pelo método IDP apresentou contornos concêntricos, em formato de vértices, ao redor dos pontos amostrados, principalmente daquelas que apresentam maiores concentrações-quantidades referentes aos parâmetros analisados (Figura 12). As feições concêntricas representam a ponderação da distância no entorno das amostras. De acordo com Landim (2000), o IDP é um algoritmo que destaca as variações de pequena amplitude (irregularidades locais) em regiões com dados dispersamente distribuídos.

Para os demais fatores utilizou-se a krigagem ordinária, visto que, esses fatores apresentam padrão espacial (Tabela 8). Krigagem ordinária (JOURNEL; HUIJBREGTS, 1978) é aceita por muitos cientistas do solo como uma técnica apropriada para estimar os valores das propriedades do solo em locais não amostrados. Quanto a metais (Figura 13), observa-se que nas proximidades dos manguezais as concentrações de metais são menores e nas áreas de encostas as concentrações são maiores. A interpolação por krigagem ordinária reflete bem ao que se observa nos dados pontuais (Tabela 3). Já para o fator que reflete as relações entre os parâmetros pH-Eh (Figura 14), as maiores concentrações encontram-se no entorno da costa. Quanto ao fator que traz informações sobre os nutrientes-MO encontrados nos solos e sedimentos (Figura 15), a maior concentração ocorre em local de proximidade dos ambientes de manguezal e urbano, sugerindo a contaminação antrópica por resíduos orgânicos.

5.4 CONCLUSÃO

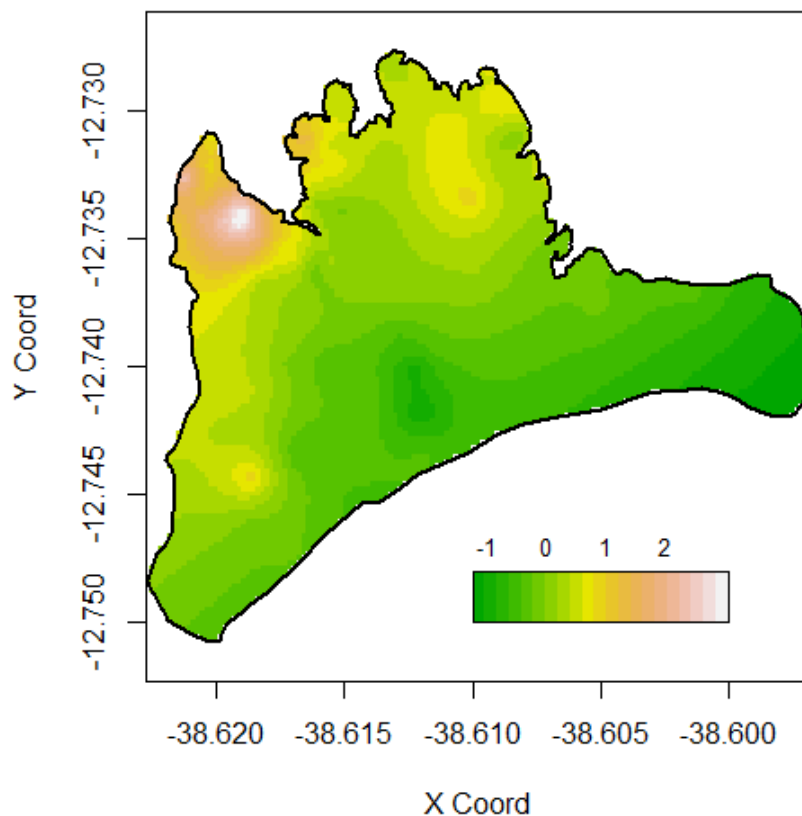
Os parâmetros referentes ao pH, Salinidade, Eh, P, N, M.O., CTC, apresentam médias e medianas diferentes para os ambientes analisados, indicando assimetria e alta variabilidade entre os ambientes analisados, justificando o resultado encontrado no teste não-paramétrico de Kruskal-Wallis que revelam que os parâmetros analisados diferem em relação ao ambiente (p -valor < 0,05) (Tabela 1). Essas variações são esperadas devido às peculiaridades de cada ambiente, sinalizando que esses parâmetros podem se alterar em relação ao espaço.

Figura 12 – Mapa da distribuição espacial para o fator compostos pelos parâmetros granulométricos interpolados por IDP para na região NE da Baía de Todos os Santos – cercanias do município de Madre de Deus, Bahia



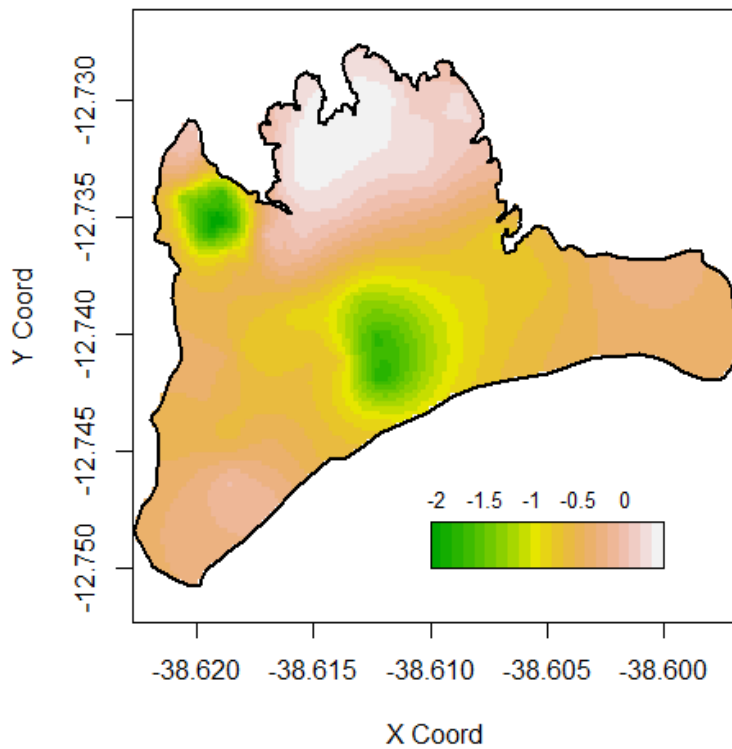
Fonte: Elaboração do autor (2019).

Figura 13 – Mapa da distribuição espacial para o fator compostos pelos metais interpolados por Krigagem para na região NE da Baía de Todos os Santos – cercanias do município de Madre de Deus, Bahia



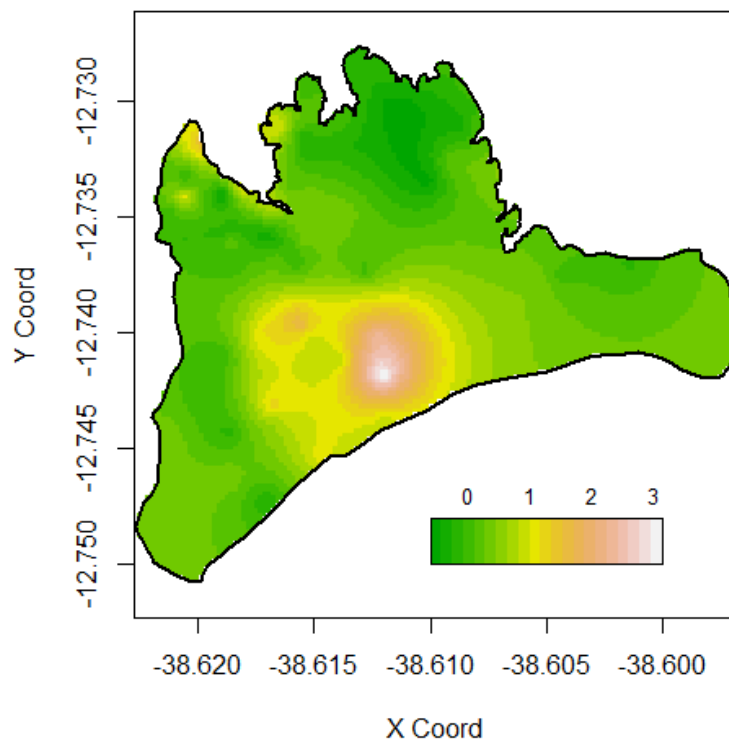
Fonte: Elaboração do autor (2019).

Figura 14 – Mapa da distribuição espacial para o fator composto por pH e Eh interpolado por krigagem para na região NE da Baía de Todos os Santos – cercanias do município de Madre de Deus, Bahia



Fonte: Elaboração do autor (2019).

Figura 15 – Mapa da distribuição espacial para o fator composto por nutrientes e M.O. interpolado por krigagem para na região NE da Baía de Todos os Santos – cercanias do município de Madre de Deus, Bahia



Fonte: Elaboração do autor (2019).

Para os parâmetros granulométricos observou-se que as medias e medianas das variáveis analisadas difere em relação aos ambientes, que é comprovado pelo teste não-paramétrico de Kurskal_Wallis ($p\text{-valor} < 0,05$). No que tange a variabilidade dos dados (C.V.%), Silte e Argila apresentam alta variabilidade (C.V.% $>60\%$), nos ambientes analisados. Já para Areia, apresenta variabilidade moderada entre os ambientes mangue, encosta e urbano ($12\% \leq C.V.\% \leq 60$).

Quanto aos metais analisados, as medias e medianas diferem em relação aos ambientes, indicando assimetrias entre os dados que pode ser observados com os valores encontrados nos coeficiente de assimetria, valores > 0 . Também se notou, que os metais analisados deferem as suas quantidades quando comparado entre os ambientes analisados (teste de não-paramétrico de Kustal-Wallis, $p\text{-valor} < 0,05$).

O presente trabalho avaliou, através de ferramentas de análise multivariada e análise espacial, a existência de padrão espacial nos parâmetros físicos e químicos oriundos de amostras de solos e sedimentos superficiais coletadas em diferentes ambientes no Município de Madre de Deus, Bahia.

A partir destas técnicas, conclui-se que os dados podem ser agrupados em quatro fatores com alta explicação da variabilidade encontrada nos dados originais: o que mais diferencia os ambientes é a granulometria e a salinidade, e, como consequência da salinidade, e o CTC. Essa é, de fato, uma característica que retrata bem a realidade. Nas encostas e áreas urbanas, os solos possuem salinidade menor, enquanto a quantidade de sais presentes em manguezais, e, sobretudo em apicuns, é bem mais elevada. Nesse caso, a CTC deve ser considerada com cuidado, haja vista que o apicum, por ser um ambiente hipersalino, exacerba o valor de CTC (que considera as quantidades de sais em seu cálculo). Quanto à granulometria, o apicum é um ambiente totalmente diferenciado, tipicamente arenoso. Já manguezais e encostas apresentam maiores quantidades de materiais mais finos, como silte e argila.

Verificou-se também a existência de pontos de alavancagem assim mostrando que a técnica robusta de análise fatorial é a mais indicada para obter resultados mais robustos.

Haja vista os dados não apresentarem distribuição normal, optou-se, então, pelo teste de aleatorização de Mantel. De acordo com o teste, metais, pH e Eh e nutrientes e M.O. apresentam padrão espacial, indicando que estas concentrações não estão distribuídas aleatoriamente no espaço.

6 MODELO ADITIVO GENERALIZADO PARA POSIÇÃO, ESCALA E FORMA APLICADO A AMBIENTE HIPERSALINO ASSOCIADO A MANGUEZA: ESTUDO DE CASO

RESUMO

A variação de dados demonstrada em estudos ambientais gera uma preocupação em torno de suas interpretações. O modelo aditivo generalizado para posição, escala e forma (GAMLSS - Generalized Additive Models for Location Scale and Shape) pode ser utilizado na modelagem de dados ambientais, pois permite ser aplicado a dados que não apresentam distribuição normal, não necessitando de abordagem paramétrica. O objetivo do trabalho foi construir um modelo de regressão utilizando o GAMLSS para compreender a influência dos parâmetros físicos e químicos no índice de partição reduzida (IR) em frações geoquímicas de um ambiente costeiro hipersalino denominado apicum. O IR descreve quantitativamente a distribuição de cada metal estudado nas frações dentro de um intervalo entre 0 a 1. O fracionamento químico foi aplicado em 14 amostras de apicuns distribuídos no Município de Madre de Deus, Bahia. A partir de análises das amostras, obteve-se as concentrações dos metais Cu, Cr e Ni em cada fração geoquímica (troçável; relacionada a carbonatos; à matéria orgânica; a óxidos e hidróxidos de Fe e Mn e fração residual, adaptado de Tessier et al., 1979). A partir destes valores de concentração calculou-se o IR para cada metal analisado. Foram analisados também parâmetros físico-químicos (pH, Eh, salinidade, granulometria, matéria orgânica - MO, capacidade de troca catiônica - CTC, nitrogênio total - N e fósforo assimilável - P). A média dos valores de IR encontrados foram 0,66 para o Cu; 0,45 para o Cr e para o Ni foi 0,98, indicando a interação existente entre os metais e determinados parâmetros físicos e químicos. As variáveis silte e areia, P e CTC apresentaram alta variabilidade (coeficiente de variação 77,63%; 68,64%; 78,48%, respectivamente). Para a construção do modelo considerou-se o IR como variável de interesse e os parâmetros físicos e químicos como o conjunto de variáveis explicativas. Primeiramente, gerou-se um modelo básico sem a utilização das variáveis explicativas e obteve-se um p-valor estatisticamente significativo. Em seguida estabeleceu-se a função de ligação com o menor valor de AIC (Critério de Informação de Akaike generalizado) e SBC (Critério Bayesiano de Schwarz). Para IR do metal Cu, a função de ligação foi Cauchit; para o Cr, foi a Log; para o Ni, a Probit. No modelo final esperava-se que a salinidade tivesse uma influência na distribuição de todos os metais nas frações geoquímicas, mas essa influência foi constatada somente para o Cu e Ni, e não para o Cr. As variáveis pH, Eh e areia influenciam a distribuição dos metais em estudo. O GAMLSS produziu resultados coerentes para um apicum com elevada dispersão nos dados, demonstrando que modelos de regressão mais sofisticados podem ser utilizados na modelagem de dados ambientais.

Palavras chaves: GAMLSS; apicum; fracionamento químico; Índice de partição reduzida

6.1 INTRODUÇÃO

A heterogeneidade demonstrada em dados ambientais gera uma grande preocupação em estudos geoquímicos. Na tentativa de solucionar este problema, a literatura sugere diversas metodologias, que vão desde o processo de coleta das informações até o uso de ferramentas estatísticas de análise. Com relação às ferramentas estatísticas mais utilizadas para análise de dados com essas características, a Análise de Variância (ANOVA) é a mais utilizada identificar e caracterizar a variabilidade existente nos dados. Porém, para utilizar a ANOVA alguns pressupostos devem ser observados, tais como: normalidade dos dados e homogeneidade da variância; que no geral, em estudos ambientais, não são verificados. Como possibilidade a esse Outra possibilidade é usar o modelo de regressão linear que tem a função de investigar a relação entre variáveis aleatórias de interesse, variável resposta ou dependente, ($Y_i, i = 1, 2, \dots, n$) com um conjunto de variáveis explicativas ou exploratórias ($X_i, i = 1, 2, \dots, n$). A relação entre a variável resposta com a variáveis explicativas é expressa na equação 1:

$$Y_i = \beta_0 + \beta_1 X_i + e_i \quad (1)$$

em que: β_0 – constante de regressão, representa o intercepto da reta com o eixo do Y ; β_1 – coeficiente de regressão, representa a variação de Y em função da variação de uma unidade da variável X ; e e_i – é o erro que está associado a distância entre o valor observado Y_i e o correspondente ponto na curva, do modelo proposto, para o mesmo nível i de X .

Quando a variável resposta (Y_i) não apresentam distribuição normal e/ou homogeneidade, o modelo clássico de regressão linear não é o mais adequado na modelagem dos dados. Apesar de, no geral, não ser possível ter total conhecimento da forma funcional verdadeira do modelo ou as suposições probabilísticas das variáveis envolvidas, é possível, a priori, ter conhecimento das hipóteses da distribuição real ou forma das funções de algumas variáveis. Assim, a regressão semiparamétrico torna-se uma opção prática, consistente e robusta para este tipo de análise ao permitir a modelagem de formas funcionais complexas que contemplam simultaneamente componentes paramétricas e não-paramétricas. Por este motivo, pesquisadores têm dedicado especial atenção ao desenvolvimento de técnicas estatísticas de modelagem mais flexíveis e menos restritivas como forma de minimizar possíveis fontes de erros de especificação do modelo e aumentar a acurácia das estimativas de quantidades de interesse (FLORÊNCIO, 2010).

Nesta linha, a classe de modelos estatísticos de regressão semiparamétrica GAMLLS (modelos aditivos generalizados para posição, escala e forma) (RIGBY; STASINOPOULOS,

2005) tem uma ampla aceitação entre a comunidade científica por conta de promover bons ajustes a um número variado de observações. Outro ponto importante está associado com a verificação da contribuição individual de cada variável explicativa com a variabilidade da variável resposta, devido à flexibilização na modelagem dos dados. O modelo a ser proposto neste estudo envolve a avaliação e/ou compreensão da interação dos metais com as frações geoquímicas no sedimento.

A dinâmica entre as frações geoquímicas oferece indício das suas biodisponibilidades, que por sua vez permitem avaliar o risco de suas presenças nos ecossistema terrestre e aquático, como também é útil para determinar o grau de associação entre os metais e os sedimentos e solos e a possível extensão de remobilização destes elementos no ambiente (Santos et. al, 2017). Os metais contidos nos solos irão se distribuir nas frações geoquímicas em decorrência das constantes modificações dos parâmetros físicos e químicos do ambiente. Desse modo, faz-se necessário compreender o grau de associação dos metais nas diferentes frações geoquímicas do sedimento/solo (FEVAS et al., 2011; WANG et al., 2012; MANN, 2010; PASSOS et al., 2010). Para tanto, utiliza-se métodos de fracionamento químico que consistam em avaliar o comportamento dos metais nas frações geoquímicas. Esse método baseia-se numa série de reações químicas consecutivas (troca iônica, oxirredução e/ou dissolução) com a utilização de reagentes apropriados em condições específicas para cada fração (GLEYZES; TELLIER, 2002; RAO et al., 2008).

Reconhecendo que o fracionamento permite determinar a concentração de metais em diferentes frações geoquímicas de solo, assume-se que essas frações apresentam diferentes disponibilidades desses metais para o ambiente (PASSOS et al., 2010). Assim, é possível verificar a interação deles com as frações geoquímicas através do índice de partição reduzida (I_R). Esse índice é indicado para descrever quantitativamente a distribuição do metal nessas frações dentro de uma escala no intervalo de 0 a 1. Na literatura o I_R foi utilizado, por exemplo, nos trabalhos de Han e Banin (1997a; 1999b), Han e Banin (2012), Han et al. (2003), Gusiatin e Klimuk (2012), Miretzky et al. (2011) e Nowack et al. (2010) para delinear os metais em amostras de solos.

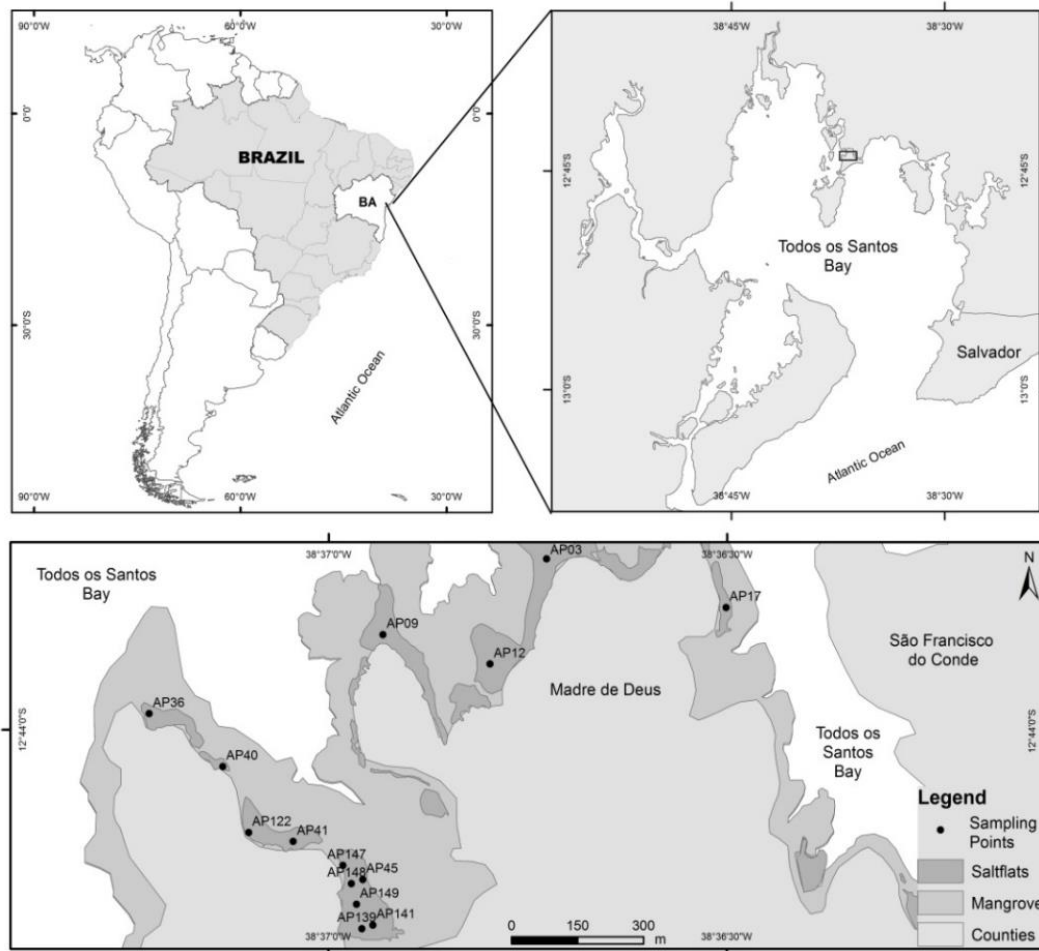
Neste contexto, o objetivo do trabalho foi construir um modelo de regressão utilizando o GAMLSS para compreender a influência dos parâmetros físicos e químicos no índice de partição reduzida (I_R), nas frações geoquímicas de um ambiente costeiro hipersalino associado a manguezais, denominado apicum.

6.2 MATERIAIS E MÉTODOS

O ambiente de apicum é uma área plana hipersalina localizada na interface manguezal-encosta ou no interior do manguezal, geralmente com ausência de vegetação (UCHA et al., 2008; HADLICH et al., 2010). Em pesquisas realizadas nesse ambiente constatou-se uma variação nos parâmetros físicos e químicos, o que pode ser atribuído à elevada salinidade e à sedimentação de partículas oriundas de ambientes adjacentes (manguezais e encostas).

Em um ambiente de apicum, no Município de Madre de Deus, localizado na região norte da Baía de Todos os Santos, Bahia, foram coletadas 14 amostras de sedimentos de superfície (0-5 cm) em novembro de 2010 (Figura 16). As amostras foram analisadas nos laboratórios do Núcleo de Estudos Ambientais (NEA) do Instituto de Geociências da Universidade Federal da Bahia. Todos os reagentes utilizados foram de alto grau analítico.

Figura 16 - Localização dos pontos de coleta de amostra de sedimento



Elaboração: Daniele Vasconcelos.

6.2.1 Procedimentos Analítico

Em laboratório, os parâmetros pH, Eh, nitrogênio total (N), capacidade de troca catiônica (CTC) foram determinados segundo Embrapa (1997); a salinidade foi determinada com o auxílio de um refratômetro, com valor máximo de 100. A análise granulométrica foi realizada em um analisador de partículas com difração a Laser, com pré-tratamento das amostras de sedimentos com H_2O_2 e $[(\text{NaPO}_3)_6]$ $0,1 \text{ mol L}^{-1}$; após leitura no equipamento, os dados foram tratados no programa GRADSTAT. A matéria orgânica (MO) foi determinada pelo método de Walkey-Black (1947). O fósforo assimilável (P) foi determinado através dos métodos de Grasshoff et al. (1999) e Aspilla (1976).

Foi aplicado o método de fracionamento proposto por Tessier et al. (1979) com adaptação em relação à extração dos metais na última etapa do procedimento (Quadro 6). O método compreende cinco frações: trocável (F1), ligada a carbonato (F2), ligada a óxido-hidróxido Fe-Mn (F3), ligada a matéria orgânica (F4) e residual (F5). Essa última etapa foi realizada em um forno de micro-ondas Provecto Analítica DG T 100 plus, seguindo a metodologia EPA 3052. O fracionamento químico foi realizado em tubos de polipropileno de 50 mL empregando 2,5 g de amostra.

Quadro 6 - Protocolo adaptado do fracionamento químico proposto por Tessier et al. (TESSIER; CAMPBELL; BISSON, 1979)

Fração	Reagente / concentração	Volume (mL)	pH	Condição experimental
F1	MgCl_2 1 mol L^{-1}	20	7	1 h, temperatura ambiente agitação contínua
F2	NaOAc 1 mol L^{-1} / HOAc	20	5	5 h, temperatura ambiente agitação contínua
F3	$\text{NH}_2\text{OH.HCl}$ $0,04 \text{ mol L}^{-1}$ em solução 25% (v/v) CH_3COOH	40	-	6 h, $96 \text{ }^\circ\text{C}$, agitação periódica
F4	HNO_3 $0,02 \text{ mol L}^{-1}$ / H_2O_2 30% / $\text{CH}_3\text{COONH}_4$ em solução 20% (v/v) HNO_3	32,5	2	5 h, 852°C , agitação periódica / 30 min, temperatura ambiente, agitação contínua
F5	HNO_3 65%, HF 40%, HCl 37%	-	-	US EPA 3052 (1996)

A quantificação dos metais (Cu, Cr e Ni) nas frações geoquímicas foi determinada por espectrometria de absorção atômica com chama acetileno/ar (FAAS), equipamento marca Analytik Jena, modelo contrAA 300, com lâmpada de Xenon short-arc como fonte de radiação operando no modo hot spot. A Tabela 9 mostra os limites de detecção (LD) e de quantificação (LQ) dos metais em questão. O LD e LQ foram calculados a partir do desvio

padrão do branco analítico e do coeficiente angular da curva de calibração para cada elemento (IUPAC, 1975). A condições experimentais do forno de micro-ondas para a digestão total da amostra encontra-se no Quadro 7.

Quadro 7 - Programas de microondas para extração parcial e total de metais em solos e sedimentos de Madre de Deus, Bahia

Passos	1 ^a	2 ^a	3 ^a	4 ^a	5 ^a
Digestão parcial					
Tempo (min)	5	1	4	3	0
Potencia (watts)	400	790	320	000	000

Fonte: Silva (2012).

Tabela 9 - Condições de operação do EAA utilizadas para a determinação dos metais em solos e sedimentos de Madre de Deus, Bahia

Elemento	λ (nm) *	LD (mg.kg ⁻¹)	LQ (mg.kg ⁻¹)
Cu	324,754	0,07	0,24
Cr	359,349	0,09	0,30
Ni	232,003	0,47	1,55

*Comprimento de onda do elemento

Fonte: Silva (2012).

6.2.2 Índice de partição reduzida dos metais: I_R

A interação dos metais Cu, Cr e Ni com as frações geoquímicas foram verificadas através do I_R . Esse índice de partição é definido do seguinte modo:

$$I_R = \frac{1}{k^2} \sum_{i=1}^k F_i i^2 \quad (2)$$

em que: $k = 5$ (número máximo de etapas do fracionamento); F_i é a razão entre a concentração do metal na respectiva fração geoquímica e o conteúdo total; i é número de etapas do fracionamento.

O menor valor de I_R (mínimo igual a 0,03) representa uma distribuição padrão onde muitos metais residem nas frações potencialmente disponíveis; por outro lado, um valor elevado de I_R (máximo igual a 1) demonstra uma situação onde a maior proporção do metal está relacionada a fração residual. Valor intermediário (ponto médio entre o mínimo e o máximo, 0,52) evidencia que os metais estão distribuídos entre todas as frações da amostra (HAN; BANIN, 1997a; 1999b; HAN et al., 2003).

6.2.3 Análise Estatística

Os modelos utilizados neste trabalho são baseados nas classes de modelos estatísticos de regressão semiparamétrico, denominados de modelos aditivos generalizados para posição, escala e forma (GAMLSS) propostos por Rigby e Stasinopoulos (2005). No modelo GAMLSS a premissa de que a variável resposta pertence à família exponencial é relaxada e substituída por uma família de distribuição mais geral. Além disto, a parte sistemática do modelo é ampliada para permitir a modelagem não apenas da média (ou posição), mas de todos os parâmetros da distribuição condicional da variável resposta, sejam através de funções paramétricas ou não-paramétricas das variáveis explanatórias e/ou termos de efeitos aleatórios.

A classe de modelos de regressão GAMLSS permite a construção de modelos de regressão para os parâmetros de muitas distribuições, tais como Binomial, Exponencial, Gama, Weibull, Beta, para citar algumas possibilidades. Nesta abordagem é possível modelar separadamente todos os parâmetros da distribuição como função de variáveis explicativas.

Matematicamente na estrutura da classe de modelos de regressão GAMLSS os p parâmetros $\theta = (\theta_1, \theta_2, \dots, \theta_p)^T$ de uma função de probabilidade $f(y/\theta)$ são modelados utilizando termos aditivos. Presume-se que para $i = 1, \dots, n$ as observações y_i são independentes e condicionais a θ , com função densidade de probabilidade $f(y_i/\theta_i)$, onde $\theta_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{ip})^T$ é um vetor de p parâmetros relacionados às variáveis explanatórias e aos efeitos aleatórios. Em 2005, Rigby e Stasinopoulos (2005) definem a classe de modelos de regressão GAMLSS da seguinte forma:

Seja $y = (y_1, y_2, \dots, y_n)^T$ um vetor de variáveis respostas. Considere uma função de ligação monótona $g_k(\cdot)$, $k = 1, 2, \dots, p$ relacionando o k -ésimo parâmetro θ_k as variáveis explanatórias e aos efeitos aleatórios por meio de um modelo aditivo (Z_{jk}, γ_{jk}) dado:

$$g_k(\theta_k) = \eta_k = X_k \beta_k + \sum_{j=1}^{J_k} Z_{jk} \gamma_{jk}$$

em que θ_k e η_k são vetores $n \times 1$, por exemplo $\theta_k = (\theta_{1k}, \theta_{2k}, \dots, \theta_{nk})^T$, $\beta_k = (\beta_{1k}, \beta_{2k}, \dots, \beta_{nk})^T$ é um vetor de parâmetros de tamanho J'_k , X_k e Z_k são matrizes de planejamento (covariáveis) fixas, conhecidas e de ordem $(n \times J'_k)$ e $(n \times q_{jk})$, respectivamente. Já γ_{jk} é uma variável aleatória q_{jk} -dimensional. O uso da fórmula acima se justifica pela

facilidade no uso dos algoritmos de reajuste (*backfitting*) e na incorporação ao modelo diferentes tipos de termos aditivos e/ou termos de efeitos aleatórios.

6.3 RESULTADOS E DISCUSSÃO

A Tabela 10 apresenta informações das medidas descritivas utilizadas (valores máximos e mínimos, média, mediana, desvio padrão e coeficiente de variação) para parâmetros físicos e químicos e dos indicadores de partição reduzida (I_R) dos metais Cu, Cr e Ni. Estes resultados são oriundos das 14 amostras georreferenciadas coletas nos apicuns da cidade de Madre de Deus. Os indicadores de partição reduzida apresentam valores de médias bem próximas aos valores de variância. Quanto a dispersão (Coeficiente de variação %), verifica-se que as medias não apresentam alta variabilidade, pois tem coeficiente de variação inferior a 50%. Já o comportamento dos parâmetros físicos e químicos, mesmo se tratando do mesmo ambiente em estudo (apicum), para alguns parâmetros físicos e químicos apresentaram alta variabilidade (Silte – C.V.% = 77,63%, Argila – C.V.% = 137,57%, P – C.V.% = 68,64%, e CTC – C.V.% = 78,48%), esses parâmetros apresentam uma grande amplitude nos dados (Mín – Max) (Tabela 10).

Tabela 10 - Medidas descritivas dos parâmetros físicos e químicos determinados em de solos ou sedimentos coletados na região NE da Baía de Todos os Santos – cercanias do município de em Madre de Deus, Bahia

Variáveis	Mín-Máx	Média	Mediana	Desvio padrão	Coef. Variação (%)
I_R -Cu	0,416-0,835	0,661	0,669	0,123	18,66
I_R -Cr	0,405-0,569	0,449	0,44	0,038	8,41
I_R -Ni	0,959-0,995	0,984	0,987	0,011	1,08
pH	6,24-8,08	7,359	7,51	0,556	7,56
Eh	-82,00-48,00	-18,929	-16,5	45,554	-
Salinidade	77-100	95,857	100	7,912	8,25
Areia	73,72-97,28	89,853	93,675	7,94	8,84
Silte	2,72-25,38	9,916	6,28	7,698	77,63
Argila	0,00-0,91	0,21	0,075	0,289	137,57
N-total	0,040-0,080	0,057	0,04	0,021	35,95
MO	0,308-2,378	1,104	1,06	0,491	44,49
P	12,50-154,01	58,471	48,135	40,137	68,64
CTC	1,475 – 154,753	54,291	36,225	42,61	78,48

Fonte: Elaboração do autor.

Em vista dados não apresentarem distribuição normal utilizou-se o coeficiente de correlação de Spearman (ρ) para identificar a correlações entre os índices de partição reduzida dos metais avaliados e os parâmetros físicos e químicos (Tabela 11). Observa-se que o I_R para o metal Cu apresentou correlação com as variáveis Areia ($\rho = -0,61$) e Silte ($\rho = 0,61$). Já o índice de partição reduzida de Cr apresentou correlação com Areia ($\rho = -0,57$), Silte ($\rho = 0,57$), P ($\rho = 0,78$) e CTC ($\rho = 0,76$). Enquanto o índice de partição reduzida de Ni não apresentou correlação significativa estatisticamente com nenhum dos parâmetros físicos e químicos (nível de significância de 5%, $\alpha = 0,05$). Por conta da pouca informação entre os I_R dos metais em estudo com os parâmetros físicos e químicos. Para descrever e verificar as relações entre o índice de partição reduzidas dos metais e os parâmetros físicos e químicos, que não são vistos através das correlações lineares, utilizou-se os modelos de regressão semiparamétrica GAMLSS.

6.3.1 Modelo estatístico

As relações existentes entre os I_R dos metais estudados e as demais variáveis foram investigadas por análise de regressão múltipla. Visto que o I_R é uma variável contínua que assume valores no intervalo $[0, 1]$, utilizou-se a distribuição beta para ajustar a variável resposta. A distribuição beta é indicada para modelagem de variáveis resposta que englobam taxas ou proporções num intervalo unitário $[0,1]$, ou seja, $0 < y < 1$. Para os procedimentos computacionais de estimação utilizou-se o pacote GAMLSS desenvolvido por Stasinopoulos e Rigby (2005), escrito para o software livre R (versão 3.2.1, R DEVELOPMENT CORE TEAM, 2015).

Para selecionar o modelo final foi seguido algumas etapas. Na primeira etapa gerou-se um modelo básico com apenas os indicadores de partição reduzidas de Cu, Cr e Ni sem a interferência de variáveis explicativas (parâmetros físicos e químicos). Para tanto foi aplicado o modelo GAMLSS para modelar os valores de I_{rCu} , I_{rCr} e I_{rNi} . Utilizou-se esses modelos sem variáveis explicativas para avaliar o ganho no ajusto dos indicadores de partição reduzida com o acréscimo dos parâmetros físicos e químicos. Os resultados estão apresentados na Tabela 12.

Tabela 11 - Coeficiente de correlação de Spearman dos parâmetros físicos e químicos determinados em de solos ou sedimentos coletados na região NE da Baía de Todos os Santos – cercanias do município de em Madre de Deus, Bahia

Variável	pH	Salinidade	Eh	Areia	Silte	Argila	N	MO	P	CTC	I _R -Cu	I _R -Cr	I _R -Ni
pH	1,00												
Salin.	-0,02	1,00											
Eh	-0,93	-0,11	1,00										
Areia	0,21	0,50	-0,22	1,00									
Silte	-0,21	-0,40	0,23	-1,00	1,00								
Argila	-0,20	-0,61	0,17	-0,91	0,90	1,00							
N	-0,69	0,28	-0,90	0,24	-0,24	-0,17	1,00						
MO	-0,45	0,17	0,38	0,01	-0,01	0,01	-0,23	1,00					
P	0,31	0,04	-0,33	-0,19	0,19	0,11	0,28	-0,11	1,00				
CTC	0,39	-0,03	-0,43	-0,30	0,31	0,20	0,37	-0,17	0,89	1,00			
I_R-Cu	-0,14	-0,49	0,26	-0,61	0,61	0,45	-0,36	0,12	0,18	0,35	1,00		
I_R-Cr	0,27	-0,25	-0,30	-0,57	0,57	0,48	0,24	-0,20	0,78	0,76	0,42	1,00	
I_R-Ni	0,49	-0,01	-0,49	0,42	-0,42	-0,33	0,41	-0,35	-0,28	0,04	-0,05	-0,34	1,00

**p*-valor < 0,05

Fonte: Elaboração do autor.

A seleção e a construção dos modelos foram realizadas em etapa. Na primeira etapa gerou-se um modelo básico sem utilização de variáveis explicativas, obtendo-se estimativas diretas para os parâmetros da distribuição aplicada à modelagem dos valores de I_rCu , I_rCr e I_rNi . Esses modelos serão considerados como referências na comparação com outros modelos. As estimativas são apresentadas na Tabela 12.

Tabela 12 - Modelo GAMLSS para as respostas I_rCu , I_rCr e I_rNi sem variáveis os parâmetros físicos e químicos determinados em de solos ou sedimentos coletados na região NE da Baía de Todos os Santos – cercanias do município de em Madre de Deus, Bahia

	Parâmetros	Estimativas	Erro Padrão	Estat. t	p-valor
I_rCu	μ^*	0,6657	0,1382	4,816	0,00422
	σ^{**}	-11,2440	0,2289	-4,912	0,00036
I_rCr	μ^*	-0,2035	0,0391	-5,201	0,0002
	σ^{**}	-2,5441	0,2022	-12,580	0,0000
I_rNi	μ^*	4,1354	0,1681	24,600	0,0000
	σ^{**}	-2,4659	0,2112	-11,680	0,0000

* μ é a média (escala)

** σ é o desvio-padrão (forma)

Fonte: Elaboração do autor.

Antes de verificar quais variáveis farão parte do modelo final, é necessário verificar qual função de ligação apresenta o melhor ajuste para o modelo. Nessa etapa, foi rodado o modelo GAMLSS com variável resposta os indicadores de partição reduzida e com variáveis explicativas todos os parâmetros físicos e químicos em estudo explicativas como objetivo de verificar qual função de ligação proporciona o melhor ganho pela associação existente entre as variáveis explicativas. Pode-se definir função de ligação com a função que associa a média da resposta aos preditores lineares no modelo. Os resultados estão apresentados na Tabela 13. A seleção da função de ligação através do Critério de Informação de Akaike (AIC) e do Critério de Informação Bayesiano (SBC) (BURNHAM; ANDERSON, 2002). Através dos critérios AIC e BIC as funções de ligação para o índice de partição reduzida serão: *Cauchit* para o Cu, *Log* para Cr e *Probit* para Ni (Tabela 13).

Após escolha da função de ligação far-se-á a seleção das variáveis que comporá o modelo final. Nessa etapa, primeiro fara a seleção das variáveis explicativas (parâmetros físicos e químicos) para modelar o parâmetro de posição (μ) aplicando técnicas de *stepwise* pare seleção de covariáveis para modelar μ . Essa técnica retira sucessivamente as variáveis menos significativas (o nível de significância de $\alpha = 10\%$), ou seja, variáveis com *p-valor* > 0,1 de modo a se chegar a um modelo com menor valor da estatística GAIC (Critério de

Informação de Akaike Generalizado), com penalidade 2, entre todos os modelos possíveis. Após seleção das covariáveis para modelar μ , repete-se a mesma metodologia para seleção das covariáveis para modelar σ (escala). Na Tabela 6 apresenta os resultados dos critérios utilizados para a seleção do modelo final oriundo da metodologia de seleção empregada. Com base nos resultados apresentado (Tabela 14), verifica-se que a introdução de variáveis para modelar os parâmetros de locação e de forma contribui com um melhor ganho no ajuste do modelo; isto acontece para todos os indicadores estudados.

Tabela 13 - Escolha da função de ligação com base no AIC e SBC

	Função de ligação	AIC	SBC
IrCu	Log	-18,8268	-11,7971
	Logit	-19,6782	-12,6486
	Cauchit	-20,9985	-13,9688
	Probit	-19,5526	-12,523
	Cloglog	-19,3258	-12,2962
IrCr	Log	-59,3446	-52,315
	Logit	-57,3851	-50,0342
	Cauchit	-57,0638	-50,3555
	Probit	-57,0174	-49,9877
	Cloglog	-58,0501	-51,0205
IrNi	log*	-104,45	-97,42
	Logit	-107,774	-100,745
	Cauchit	-105,917	-98,8868
	Probit	-107,847	-100,818
	cloglog*	-107,837	-100,807

Fonte: Elaboração do autor.

Tabela 14 - Estatísticas para seleção dos modelos avaliados

	Parâmetros	A	B	C
Ir-Cu	GAIC	-16,9498	-19,5006	-21,5460
	SBC	-15,6717	-15,0272	-16,4335
	Pseudo-R ²	-	0.5920	0.6944
Ir-Cr	GAIC	-49,1760	-67,5980	-71,1305
	SBC	-47,8978	-63,7636	-66,6571
	Pseudo-R ²	-	0.8485	0.8980
Ir-Ni	GAIC	-89,7974	-110,3204	-114,3932
	SBC	-88,5193	-104,5689	107,3635
	Pseudo-R ²	-	0.9151	0.9475

A – Básico; B – Para média; C – Para média e variância.

Fonte: Elaboração do autor.

Com base nos resultados apresentado, verifica-se que a introdução de variáveis para modelar os parâmetros de locação e de forma contribui com um melhor ganho no ajuste do modelo; isto acontece para todos os indicadores estudados.

6.3.2 Modelo ajustado

A heterogeneidade das frações geoquímicas e a falta de uma teoria que determine de forma funcional a relação entre o índice de partição reduzida dos metais e parâmetros físicos e químicos dificultam a aplicação de ferramentas estatísticas que resultem em modelos simultaneamente parcimoniosos, abrangentes e fidedignos as relações geoquímicas existentes no ambiente. Desse modo, a estrutura de regressão utilizada para descrever tais relações, I_R x parâmetros físicos e químicos, sejam flexíveis a ponto de acomodar as peculiaridades do ambiente apicum.

O modelo final ajustado para todos os I_R está representado na Tabela 15. Verificou-se que, dos parâmetros físicos e químicos analisados, poucas variáveis descrevem a influência do Cu na interação das frações geoquímicas (I_R), indicando que pH, Eh, Salinidade, Areia e CTC contribuem para explicar o parâmetro de posição (μ); o pH teve a maior contribuição quando comparado com as outras variáveis. Nos apicuns, a presença de sais favorece a elevação do pH através do efeito tampão derivado das águas marinhas que atingem o ambiente, o que explica a variação de pH entre levemente ácido (6,24) e básico (8,08). Neste modelo o parâmetro de escala (σ) é ajustada pela M.O devido ao Cu ter a tendência em formar complexos estáveis com a matéria orgânica (GOMES et al., 1997; JARADAT et al., 2006; MAHANTA; BAHTTACHARYYA, 2011), com os grupos funcionais R-COOH (carboxila) e R-OH (fenólico) presentes nesta fração. Para o Cr, as variáveis pH, Eh, Areia e P contribuíram para explicar a interação deste metal nas frações geoquímicas para modelar o parâmetro de posição. Já o parâmetro de escala foi ajustada pela CTC. A elevada salinidade dos apicuns contribui para elevar o valor de CTC. Ao contrário da correlação de Spearman apresentadas na Tabela 11, que não indicou correlação entre as variáveis para o I_R do Ni o modelo GAMLSS apontou uma relação deste índice com as variáveis pH, Eh, salinidade, areia, MO, P e CTC, que contribuíram para modelar o parâmetro de posição. O pH também apresentou a maior contribuição quando comparado com as outras variáveis. O parâmetro de escala no modelo I_R do Ni foi ajustada pelas variáveis Eh e CTC.

Tabela 15 - Modelos GAMLSS selecionados para as respostas IrCu, IrCr e IrNi selecionado para os coeficientes μ e σ

I _R	Coeficiente μ				Coeficiente σ			
	Variáveis	Estimativa	Erro Padrão	p-Valor	Variáveis	Estimativa	Erro Padrão	p-Valor
Cu	Intercepto	-4,6162	3,3738	0,2203	Intercepto	1,0809	0,4046	0,0370
	pH	0,9643	0,4828	0,0928	MO	-2,6163	0,1719	0,0000
	Eh	0,0157	0,0065	0,0511				
	Salinidade	-0,0046	0,0152	0,7712				
	Areia	-0,0191	0,0175	0,3177				
	CTC	0,0126	0,0029	0,0052				
Cr	Intercepto	-0,2862	0,2819	0,3439	Intercepto	-4,3848	0,3179	0,0000
	pH	-0,0369	0,0372	0,3545	CTC	0,0120	0,0047	0,0387
	Eh	-0,0007	0,0005	0,1953				
	Areia	-0,0036	0,0009	0,0040				
	P	0,0011	0,0002	0,0019				
Ni	Intercepto	0,3676	0,3949	0,4207	Intercepto	-3,9723	0,4989	0,0041
	pH	0,1078	0,0522	0,1307	Eh	-0,0238	0,0037	0,0076
	Eh	-0,0008	0,0009	0,4135	CTC	-0,0111	0,0073	0,2269
	Salinidade	-0,0034	0,0013	0,0855				
	Areia	0,0186	0,0019	0,0023				
	MO	-0,1486	0,0494	0,0574				
	P	-0,0084	0,0005	0,0004				
	CTC	0,0067	0,0005	0,0009				

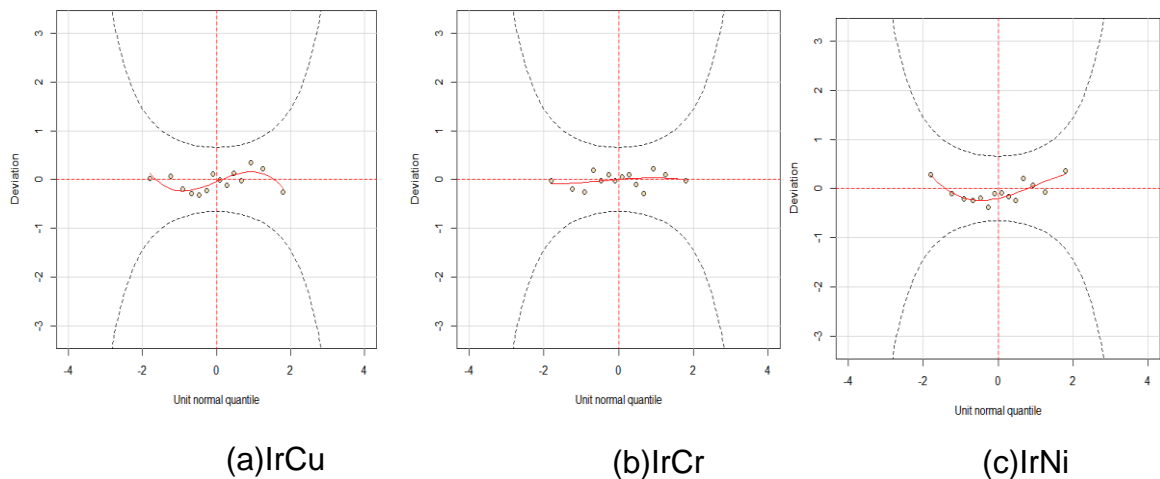
*Erro Padrão

Fonte: Elaboração do autor.

Vale ressaltar que nos modelos apresentados para descrever as relações entre os índices de partição reduzidas para metais e os parâmetros físicos e químicos as covariáveis pH, Eh e Areia estiveram presentes, o que caracteriza uma maior influência destas variáveis com relação à interação dos metais Cu, Cr e Ni nas frações geoquímicas.

Com objetivo de verificar se o modelo final ajusta adequadamente os dados, será usado os resíduos do modelo para verificação. Nesse trabalho será utilizado o gráfico de *worm plot* para realizar a análise de resíduos (Figura 17). Este gráfico foi introduzido por Buuren e Fredriks (2011) e consiste em uma ferramenta de diagnóstico para análise dos resíduos em diferentes regiões (intervalos) da variável explanatória. Como podem ser observados, os pontos dos modelos avaliados estão situados no interior da região de “aceitação” (entre as curvas elípticas), ou seja, os modelos fornecem um bom ajuste.

Figura 17 - Gráfico de ajuste *worm-plot* dos modelos para as repostas IrCu, IrCr e IrNi para os resíduos



Fonte: Elaboração do autor.

6.4 CONCLUSÃO

O apicum demonstrou ser um ambiente bastante complexo, pois os dados dos parâmetros físicos e químicos apresentaram uma elevada variação. Como os resultados analíticos dos sedimentos não seguem uma distribuição normal e, portanto, não podem ser analisados por métodos paramétricos. Por conta deste fato, a correlação de Spearman não apresentou resultados coerentes que pudessem explicar a influência dos parâmetros físicos e

químicos no I_R. Com isso foi necessária a utilização de ferramentas estatísticas mais sofisticadas, selecionando-se o GAMLSS.

Os modelos GAMLSS podem ser aplicados para vários tipos de distribuições de probabilidade da variável resposta, não ficando limitado à distribuição normal; portanto, apresentam uma flexibilidade quanto à modelagem de dados. Esse tipo de modelo produziu resultados coerentes para um ambiente costeiro hipersalino com elevada dispersão nos dados, demonstrando que modelos de regressão mais sofisticados melhor se ajustam à modelagem de dados ambientais devido à grande quantidade de variabilidade não controlada.

O uso da classe de modelos GAMLSS não deve ser confundido com preciosismo da análise de regressão, mas como um método eficiente de modelagem oriundo de técnicas avançadas da pesquisa científica que contribui para o aumento da acurácia da pesquisa. A importância desta classe de modelos advém de o fato desta metodologia constituir uma abordagem unificada de muitos procedimentos estatísticos.

7 CONCLUSÃO GERAL

Para modelar o indicador de fracionamento de metais em um ambiente hipersalino, os modelos GAMLSS podem ser aplicados por apresentar flexibilidade quanto à modelagem de dados. Esse tipo de modelo produziu resultados coerentes para um ambiente costeiro hipersalino com elevada dispersão nos dados demonstrando que modelos de regressão mais sofisticados melhores se ajustam à modelagem de dados ambientais, devido à grande quantidade de variabilidade não controlada.

Através técnicas multivariadas (análise fatorial robusta) e análise espacial foi possível comprovar a existência de padrão espacial nos parâmetros físicos e químicos oriundos de amostras de sedimentos/solo retiradas na cidade de Madre de Deus, Bahia. Haja vista, os dados não apresentarem distribuição normal, utilizou-se o teste de aleatorização de Mantel para identificar o padrão espacial entre os parâmetros físicos e químicos analisados. Concluiu-se que os parâmetros granulométricos apresentam padrão espacial aleatório, deste modo, o mapa de distribuição foi construído utilizando a interpolação espacial (IDP). Para os fatores referentes ao pH-Eh, metais e nutrientes-MO o teste de aleatorização de Mantel identificou que esses fatores apresentam padrão espacial e para a construção dos mapas de distribuição foi utilizado a interpolação por Krigagem.

REFERÊNCIAS

- ARIA, M.; CUCCURULLO, C. Bibliometrix: An R-tool for comprehensive science mapping analysis. **Journal of Informetrics**, v. 11, n. 4, p. 959-975, 2017.
- AMBLARD-GROSS, G.; MAUL, A.; FERARD, J. F.; CARROT, F.; AYARAULT, S. Spatial variability of sampling: Grid size impact on atmospheric metals and trace elements deposition mapping with mosses. **Journal of Atmospheric Chemistry**, v. 49, n. 1–3, p. 39–52, 2004.
- ASSUMPCÃO, H. C. P. **Modelagem ambiental e krigagem de contaminantes inorgânicos no município de Madre de Deus, Bahia**. 88 f., 2014. Dissertação (Mestrado em Geoquímica: Petróleo e Meio Ambiente) – Instituto de Geociências, Universidade Federal da Bahia. Salvador, 2014.
- ASSUNÇÃO, R. M. **Estatística espacial com aplicações em epidemiologia, economia e sociologia**. São Carlos: UFSCar, 2001. 136 p.
- ASPILA, K. I.; AGEMIAN, H.; CHAU, A. S. Y. A semi-automated method for the determination of inorganic, organic and total phosphate in sediments. **Analyst**, v. 101, p. 187-197, 1976.
- BRUS, D.J.; DE GRUIJTER, J.J. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). **Geoderma**, v. 80, p. 1–59, 1997.
- BUUREN, S.; FREDRIKS, M. Worm plot: a simple diagnostic device for modelling growth reference curves. **Statistic in Medicine**, v. 20, n. 8, p. 1259-1277, 2001.
- BURROUGH, P.A. Spatial aspects of ecological data. In: JONGMAN, R.H.; TER BRAAK, C.J.F.; TONGEREN, O.F.R. (Eds). **Data analysis in community and landscape ecology**. Wageningen: Pudoc, 1987. p. 213-251.
- CAFARELLI, BARBARA; CASTRIGNANÒ, ANNAMARIA; DE BENEDETTO, DANIELA; PALUMBRO, ANGELO DOMENICO; BUTAFUOCO, GABRIELE BUTTAFUOCO. A linear mixed effect (LME) model for soil water content estimation based on geophysical sensing: a comparison of an LME model and kriging with external drift. **Environmental Earth Sciences**, v. 73, n. 5, p. 1951–1960, 2015
- CÂMARA, G. et al. Análise espacial de áreas. In: _____. **Análise espacial de dados geográficos**. São José dos Campos: INPE, 2004. cap. 5. Disponível em: <http://www.dpi.inpe.br/gilberto/livro/analise/cap5-areas.pdf>. Acesso em: 24 out. 2016.
- CAO, LI; CHENG, QIUMING. Quantification of anisotropic scale invariance of geochemical anomalies associated with Sn-Cu mineralization in Gejiu, Yunan Province, China. **Journal of Geochemical Exploration**, v 122, p 47-54, 2012.
- CARVALHO, C. N. Geoquímica ambiental: conceitos, métodos e aplicações. **Geochimica Brasiliensis**, v.3, n.1, p.17-22, 1989.

CHENG, QIUMING. Singularity theory and methods for mapping geochemical anomalies caused by buried sources and for predicting undiscovered mineral deposits in covered áreas. **Journal of Geochemical Exploration**, v 122, p 55-70, 2012.

CRESSIE, N. A. C. **Statistics for Spatial Data**. New York: John Wiley & Sons, 1993.

COBO, M. J.; LÓPEZ-HERRERA, A. G.; HERRERA-VIDEIRA, E.; HERRERA, F. Science mapping software tools: Review, analysis, and cooperative study among tools. **Journal of the American Society for Information Science and Technology**, v. 62, n. 7, p. 1382-1402, 2011.

CONCEIÇÃO, G. M. S.; SALDIVA, P. H. N.; SINGER, J. M. Modelos MLG e MAG para análise da associação entre poluição atmosférica e marcadores de morbi-mortalidade: uma introdução baseada em dados da cidade de São Paulo, **Revista Brasileira de Epidemiologia**, v. 4, n. 3, p. 206-219, 2001.

COCHRAN, W.G. Relative accuracy of systematic and stratified random samples for a certain class of populations. **Annals of Mathematical Statistics**, v. 17, p. 164-177, 1946.

COCHRAN, W. G. **Sampling techniques**. 3. ed. New York: John Wiley, 1977. 428 p.

DRUCK, S.; CARVALHO, M.S.; CÂMARA, G.; MONTEIRO, A.V.M. (Eds). **Análise espacial de dados geográficos**. Brasília: EMBRAPA, 2004. Não existe, p. ex., a figura 13...

DAS, A. C. Two-dimensional systematic sampling and the associated stratified and random sampling. **Sankhyā**, v. 10, p. 95–108, 1950.

DE GRUIJTER, J.J.; TER BRAAK, C.J.F. Model-free estimation from spatial samples: a reappraisal of classical sampling theory. **Mathematical Geology**, v. 22, p. 407–415, 1990.

DESSARD, H.; BAR-HEN, A. Experimental design for spatial sampling applied to the study of tropical forest regeneration. **Canadian Journal of Forest Research-Revue Canadienne Recherche Forestiere**, v. 35, n. 5, p. 1149–1155, 2005.

DUNN, R.; HARRISON, A.R. Two-dimensional systematic sampling of land use. **Applied Statistics**, v. 42, p. 585–601, 1993.

EMBRAPA – EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Manual de métodos de análise de solo**. 2. ed. rev. atual. Rio de Janeiro: Embrapa - Centro Nacional de Pesquisa de Solo, 1997.

FAVAS, P. J. C.; PRATAS, J.; GOMES, M. E. P.; CALA, V. Selective chemical extraction of heavy metals in tailings and soils contaminated by mining activity: Environmental implications. **Journal of Geochemical Exploration**, v. 111, p. 160–171, 2011.

FIGUEIREDO FILHO, D. B.; SILVA JÚNIOR, J. A. Visão além do alcance: uma introdução à análise fatorial. **Opinião Pública**, Campinas, v. 16, n. 1, p. 160-185, 2010.

FLORÊNCIO, L. A. **Engenharia de avaliação com base em modelos GAMLSS**. 2010. Dissertação (Mestrado em Estatística) - Universidade Federal de Pernambuco.

GARRIDO, L. E.; ABAD, F. J.; PONSODA, V. A new Look at Horn's parallel analysis with ordinal variables. **Psychological Methods**, v. 18, n. 4, p. 454-474, 2013.

GETIS, A.; DRUMMY, P.; GARTIN, J.; GORR, W.; HARRIES, K.; ROGERSON, P.; STOE, D.; WRIGHT, R. Geographic information science and crime analysis. **URISA Journal**, v. 12, n. 2, p. 7-14, 2000.

GLEYZES, C.; TELLIER, S., M. Astruc Fractionation studies of trace elements in contaminated soils and sediments: a review of sequential extraction procedures. **Trens in analytical chemistry**, v. 21, n. 6-7, p. 451-467, 2002.

GOMES, N. M.; FARIA, M. A.; SILVA, A. M.; MELLO, C. R.; VIOLA, M. R. Variabilidade espacial de atributos físicos do solo associados ao uso e ocupação da paisagem. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 11, n. 4, p. 427-435, 2007.

GOMES, P. C.; FONTES, M. P. F.; COSTA, L. M.; MENDONÇA, E. S. Extração fracionada de metais pesados em latossolo vermelho-amarelo. **Revista Brasileira de Ciência do Solo**, v. 21, p. 543-551, 1997.

GRASSHOFF, K.; KREMLING, K.; EHRHARDT, M. **Methods of Seawater Analyses**. 3. ed, Willey-VCH, Weinhein, 1999. 599 p.

GUERRA, P.A.G. **Geoestatística operacional**. Brasília: Ministérios das Minas e Energia, 1988.

GUSIATIN, Z. M.; KLIMIUK, E. Metal (Cu, Cd and Zn) removal and stabilization during multiple soil washing by saponin. **Chemosphere**, v. 86, p. 383-391, 2012.

HADLICH, G. M.; CELINO, J. J.; UCHA, J. M. Diferenciação físico-química entre apicuns, manguezais e encostas na Baía de Todos os Santos, Nordeste do Brasil. **Geociências**, v. 29, n.4, p. 633-641, 2010.

HAN, F. X.; BANIN, A. 1999. Long-term transformation and redistribution of potentially toxic heavy metals in arid-zone soils: II. Incubation at the field capacity moisture content. **Water, Air, and Soil Pollution**, v. 114, p. 221-250, 1990.

HAN, F. X.; BANIN, A. Long-term transformations and redistribution of potentially toxic heavy metals in arid-zone soils incubated: I. **Under saturated conditions, Water, Air, and Soil Pollution**, v. 95, p. 399-423, 1997.

HAN, F. X.; BANIN, A.; KINGERY, W. L.; TRIPLETT, G. B.; ZHOU, L. X.; ZHENG, S. J.; DING, W. X. New approach to studies of heavy metal redistribution in soil. **Advances in Environmental Research**, v. 8, p. 113-120, 2003.

HATJE, V.; ANDRADE, J.B. (Org.). **Baía de Todos os Santos: aspectos oceanográficos**. Salvador: EDUFBA, 2009. 304 p.

HORVITZ, D.G.; THOMPSON, D.J. A generalization of sampling without replacement from a finite universe. **Journal of the American Statistical Association**, v. 47, p. 663-685, 1952.

IUPAC. International Union of Pure and Applied Chemistry. Nomenclature, symbols, units and their usage in spectrochemical analysis-III. **Spectrochimica Acta**, v. 338, p. 247-269, 1975

JARADAT, Q. M.; MASSADEH, A. M.; ZAILTOUN, M. A.; MAITAH, B. M. Fractionation and sequential extraction of heavy metals in the soil of scrapyards of discarded vehicles. **Environmental Monitoring and Assessment**, v. 112, p. 197–210, 2006.

KRIGE, D. G. A statistical approach to some basic mine evaluation problems on the Witwatersrand. **Journal South African Institute Mining Metallurgy**, v. 52, p. 119-139, 1951.

LAGO, J.B. **Banco de Dados Geoquímicos Georreferenciados Aplicados a Região do Litoral Sul e da Baía de Todos os Santos – Bahia**. 2013. Dissertação (Mestrado em Geoquímica) – Programa de Pós-Graduação em Geoquímica: Petróleo e Meio) – Instituto de Geociências, Universidade Federal da Bahia, 2013.

LANDIM, P.M.B. **Introdução aos métodos de estimação espacial para confecção de mapas**. Rio Claro: UNESP, 2000. 20 p.

LESSA, G.C.; BITTENCOURT, A.C.S.P.; BRICHTA, A.; DOMINGUEZ, J.M.L. A Reevaluation of the Late Quaternary Sedimentation in Todos os Santos Bay (BA), Brazil. **Anais da Academia Brasileira de Ciências**, v. 72, n. 4, p. 573-590, 2000.

LI, JIN e HEAP, ANDREW D. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. **Ecological Informatics**, v. 6, p. 228-241, 2011.

LI, JIN e HEAP, ANDREW D. Spatial interpolation methods applied in the environmental sciences: A review. **Environmental Modelling & Software**, v 53, p 173-189, 2014.

LONGLEY, P.A.; GOODCHILD, M.F.; MAGUIRE, D.J.; RHIND, D.W. **Sistemas e ciência da informação geográfica**. Porto Alegre: Bookman, 2013. 540p.

MAHANTA, M. J.; BHATTACHARYYA, K. G. Total concentrations, fractionation and mobility of heavy metals in soils of urban area of Guwahati, India. **Environmental Monitoring and Assessment**, v. 173, p. 221-240, 2011.

MANLY, B. F. J. **Randomization, Bootstrap and Monte Carlo Methods in Biology**. Florida: Chapman & Hall, 2006.

MANN, A. W. Strong versus weak digestions: ligand-based soil extraction geochemistry. **Geochemistry: Exploration**. **Environment Analysis**, v. 10, p. 17-26, 2010.

MANTHERON, G. **The theory of regionalized variables and its applications**. Paris: Les Cahiers du Centre de Morphologie Mathématique de Fontainebleu, 1971.

MANZIONE, R. L.; ZIMBACK, C. R. L. Análise espacial multivariada aplicada na avaliação da fertilidade do solo. **Engenharia na Agricultura**, Viçosa, v. 19, n. 3, p. 227-235, 2011.

MANZIONE, R.L., CAMARA, G. MONTEIRO, A.M.V., ZIMBACK, C.R.L., DRUCK, S. Exploring the multivariate spatial structure of soil acidity data. In: RENARD, P. et al. (Ed.) **Geostatistics for Environmental Applications**. Berlin: Springer, 2005.

MORAN, P. A. P. Notes on continuous stochastic phenomena. **Biometrika**, v. 37, p. 17-23, 1950.

MOREIRA, W. Revisão de literatura e desenvolvimento científico: conceitos e estratégias para confecção. **Janus**, v. 1, n.1, 2008.

MIRETZKY, P.; AVENDAÑO, M. R.; UMÑOZ, C.; CARRILLO-CHAVEZ, A. Use of partition and redistribution indexes for heavy metal soil distribution after contamination with a multi-element solution. **Journal of Soils Sediments**, v. 11, p. 619–627, 2011.

MORRIS, S.; VAN DERVEER MARTENS, B. Mapping research specialties. **Annual Review of Information Science and Technology**, v. 42, p. 213–295, 2008.

MÜLLER, W. G. **Collecting spatial data: Optimum design of experiments for random fields**. New York: Physica-Verlag, 1998.

NASCIMENTO, ANDRÉA GONÇALVES. **Altimetria para bibliotecários: guia prático de métricas alternativas para avaliação da produção científica**. Rio de Janeiro: edição do autor, 2016.

NEYMAN, J. On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. **Journal of the Royal Statistical Society**, v. 97, n. 4, p. 558-625, 1934.

NEYMAN, J. Contributions to the theory of sampling human populations. **Journal of the American Statistical Association**, v. 33, p. 101-116, 1938.

NYCHKA, D.; SALTZMAN, N. Design of air-quality monitoring networks. In: NYCHKA, D.; PIEGORSCH, W.; COX, L. (Eds.). **Case studies in environmental statistics**. New York: Springer-Verlag, 1998.

NOWACK, B.; SCHULIN, R.; LUSTER, J. Metal fractionation in a contaminated soil after reforestation: Temporal changes versus spatial variability. **Environmental Pollution**, v. 158, p. 3272-3278, 2010.

PASSOS, E. A.; ALVES, J. C.; SANTOS, I. S.; ALVES, J. P. H.; GARCIA, C. A. B.; COSTA, A. C. S. Assessment of trace metals contamination in estuarine sediments using a sequential extraction technique e principal component analysis. **Microchemical Journal**, v. 96, p. 50–57, 2010.

PISON, G.; ROUSSEEUW, P. J.; FILZMOSER, P; CROUX, C. “Robust Factor Analysis.” **Journal of Multivariate Analysis**, v. 84, p. 145-172, 2003.

QUENOUILLE, M.H. Problems in plane sampling. **Annals of Mathematical Statistics**, v. 20, p. 355–375, 1949.

- RAO, C. R. M.; SAHUQUILLO, A.; SANCHEZ, J. F. L. A Review of the Different Methods Applied in Environmental Geochemistry For Single and Sequential Extraction of Trace Elements in Soils and Related Materials. **Water, Air & Soil Pollution**, v. 189, p. 291-333, 2008.
- RIBEIRO JUNIOR, P. J.; VIOLA, D. N.; DEMETRIO, C. G. B.; MANLY, B., F.; FERNANDES, O. A. Spatial detection patterns of onion thrips (*Thrips tabaci*) on onion fields. **Scientia Agricola**, v. 66, p. 90-99, 2009.
- RIBEIRO JÚNIOR, P. J.; DIGGLE, P. J. geoR: a package for geostatistical analysis. **R-NEWS**, Pelotas, v. 1, n. 2, p. 15-18, 2001.
- RIGBY, R. A., STASINOPOULOS, D. M. Generalized Additive Models for Location, Scale and Shape, (with discussion). **Applied Statistic**, v. 54, p. 507–554, 2005.
- RIPLEY, B. **Spatial Statistics**. New York: Wiley, 1981.
- RIGBY, R. A.; STASINOPOULOS, D. Generalized additive models for location, scale and shape. **Journal of the Royal Statistical Society**, v. 54, p. 507-554, 2005.
- SALEHI, M. Optimal sampling design under a spatial correlation model. **Journal of Statistical Planning and Inference**, v. 118, p. 9–18, 2004.
- SÄRNDAL, C. E. Design-based and model-based inference in survey sampling. **Scandinavian Journal of Statistics**, v. 5, p. 27–52, 1978.
- SIEGEL, Frederic R. **Environmental geochemistry of potentially toxic metals**. 1.ed. New York: Editora Springer, 2002.
- SILVA, M. J. **Fracionamento de metais traço (Cr, Cu, Ni e Zn) em solos e sedimentos do município de Madre de Deus, Bahia**. 90 f., 2012. Dissertação (Mestrado em Geoquímica: Petróleo e Meio Ambiente) – Instituto de Geociências, Universidade Federal da Bahia. Salvador, 2012.
- SIMOS, N.; COSTANTINO, C. J. Soil Spatial Variability Effect on Soil Structure Interaction Studies: Enveloping Uncertainties in Structural Response. In: UJNR WORKSHOP ON SOIL-STRUCTURE INTERACTION, 3. **Proceedings**, California, USA.
- SPÖCK, G. Spatial sampling design based on spectral approximations to the random field. **Environmental Modelling & Software**, v. 33, p. 48-60, 2012.
- STEIN, A., ETTEMA, C. An overview of spatial sampling procedures and experimental design of spatial studies for ecosystem comparisons. **Agriculture, Ecosystems and Environment**, v. 94, p. 31–47, 2003.
- STEVENS, D. L., Jr.; OLSEN, A. R. Spatially balanced sampling of natural resources. **Journal of the American Statistical Association**, v. 99, n. 465, p. 262–278, 2004.
- TESSIER, A.; CAMPBELL, P. G. C.; BISSON, M. 1979. Sequential extraction procedure for the speciation of particulate trace metals. **Analytical Chemistry**, v. 51, n. 7, p. 844–851, 1979.

- TOBLER, W. R. A Computer Movie Simulating Urban Growth in the Detroit Region. **Economic Geography**, v. 46, p. 234-240, 1970.
- UCHA, J. M.; HADLICH, G. M.; CELINO, J. J. Apicum: transição entre solos de encostas e de manguezais. **Revista E. T. C.**, p. 58-63, 2008.
- VAN ECK, N. J.; WALTMAN, L. Visualizing bibliometric networks. In Y. DING, R. ROUSSEAU; D. WOLFRAM (Eds.). **Measuring scholarly impact: methods and practice.**, p. 285– 320, London: Springer, 2014.
- VIOLA, D. N. **Detecção e modelagem de padrão especial em dados binários de contagem.** 2007. 118p. Tese (Doutorado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura Luiz de Queiroz/USP, Piracicaba, 2007.
- WALKLEY-BLACK, A. A critical examination of a rapid method for determining organic carbon soils: Effect of variations in digestion conditions and of inorganic soil constituents. **Soil Science**, v. 63, p 251-263, 1947.
- WANG, J.F.; CHRISTAKOS, G.; HU, M.G., Modeling spatial means of surfaces with stratified nonhomogeneity. **IEEE. Transactions on Geoscience and Remote Sensing**, v. 47, 4167–4174, 2009.
- WANG, J.F.; HAINING, R.P.; CAO, Z.D. Sample surveying to estimate the mean of a heterogeneous surface: reducing the error variance through zoning. **International Journal of Geographical Information Science**, v. 24, 523–543, 2010.
- WANG, WENLEI; ZHAO, JIE; CHENG, QIUMING. Fault trace-oriented singularity mapping technique to characterize anisotropic geochemical signatures in Gejiu mineral district, China. **Journal of Geochemical Exploration**, v. 134, p 27-37, 2013.
- WANG, J., STEIN, A., GAO, B. GE, Y. A review of spatial sampling. **Spatial Statistics**, v. 2, p. 1-14, 2012.
- WANG, X.; LI, Y. Distribution and fractionation of heavy metals in long-term and short-term contaminated sediments. **Environmental Engineering Science**, v. 29, n. 7, p. 617-622, 2012.
- WELBERS, K.; VAN ATTEVELDT, W.; BENOIT, K. Text Analysis in R. **Communication Methods and Measures**, v. 11, n. 4, p. 245-265, 2017.
- WOLGEMUTH, K. M.; BURNETT, W. C.; MOURA, P. L. Oceanography and suspended material in Todos os Santos Bay. **Revista Brasileira de Geociências**, v. 11, n. 3, p. 172-178, 1981.
- YAMAMOTO, J. K.; LANDIM, P. M. B. **Geoestatística: conceitos e aplicações.** São Paulo:Ofcina de Textos, 2013. p. 215.
- ZHANG, YY. **Robustfa: An Object Oriented Solution for Robust Factor Analysis.** 2013, R package version 1.0-5, URL <http://CRAN.R-project.org/package=robustfa>.

ZHANG, YING-YING; RONG, TENG-ZHONG; LI, MAN-MAN. A Recommendation for Classical and Robust Factor Analysis. **Journal of Mathematics and Computer Science**, v 21, n 2, p. 1-15, 2017.

ZHU, Z. Y.; STEIN, M. L. Spatial sampling design for prediction with estimated parameters. **Journal of Agricultural Biological and Environmental Statistics**, v. 11, n. 1, p. 24–44, 2006.

ZHU, Z. Y.; ZHANG, H. Spatial sampling design under the infill asymptotic framework. **Environmetrics**, v. 17, n. 4, p. 323–337, 2006.

ZUO, RENGUANG; CARRANZA; E. J. M.; WANGA, Jian. Spatial analysis and visualization of exploration geochemical data. **Earth-Science Reviews**, v. 158, p. 9-18, 2016.

ZUO, RENGUAG; WANG, JIAN. Fractal/multifractal modeling of geochemical data: A review. **Journal of Geochemical Exploration**. v 164, p 33-41, 2016.

ZYGMONT, C., SMITH, M. R. Robust factor analysis in the presence of normality violations, missing data, and outliers: Empirical questions and possible solutions. **The Quantitative Methods for Psychology**, v. 10, n. 1, p. 40-55. DOI: 10.20982

APÊNDICE A

FUNDAMENTOS TEÓRICOS

A abordagem clássica para estudo e tratamento de dados ambientais, sociais ou econômicos, utiliza ferramentas estatísticas bem consolidadas, como teste de hipóteses, análise de variância, análise de regressão, dentre outras. Nessa abordagem, o espaço geográfico ou a localização do dado não é considerada. Com o avanço dos Sistemas de Informações Geográficas (SIGs), “onde” ocorrem os fenômenos, torna-se muito importante e em alguns casos é essencial à compreensão do problema, principalmente em se tratando de estudos ambientais. Ferramentas estatísticas foram desenvolvidas e a localização do fenômeno em estudo tornou-se fundamental para compreensão do problema. Essas ferramentas foram disponibilizadas a várias áreas do conhecimento, seja pela inclusão de funções de estatística espacial nos SIGs, seja pelo desenvolvimento de interfaces eficientes para combinar a capacidade de visualização dos SIGs com o poder analítico dos pacotes de estatística.

Nesse capítulo são apresentados alguns conceitos básicos sobre a evolução da estatística espacial e de modelos utilizados.

A) ESTATÍSTICA ESPACIAL

Muitos autores apontavam para a importância de considerar a componente espacial em seus estudos. Um dos primeiros e importante estudo foi desenvolvido por Simos e Costantino (2004) em pesquisa sobre a variação do teor de água no solo. Os autores afirmaram que, ao fazer um estudo sobre o solo ou suas propriedades, deve-se sempre levar em consideração a variabilidade que ocorre no espaço e/ou no tempo. Portanto, fazer uma amostragem em um número finito de locais daria uma descrição incompleta da variação natural, pois a variabilidade existente no solo é o produto de muitos fatores que operam em diferentes escalas espaciais e temporais. Vale ressaltar que as propriedades do solo exibem variações mesmo dentro de perfis homogêneos ou de camadas subsuperficiais distintas.

O estudo de variações ambientais, sejam elas geoquímicas, biológicas ou outras, requer o uso de ferramentas estatísticas adequadas. Os métodos clássicos de análise estatística são construídos com o pressuposto de que as amostras são aleatórias, independentes e

identicamente distribuídas (i.i.d.), o que não é válido para dados com estrutura de dependência espacial e/ou temporal. Para dados que apresentam evidência de dependência espacial é esperado que as observações vizinhas sejam mais parecidas do que as distantes, ou seja, correlacionadas espacialmente (RIBEIRO JÚNIOR et al., 2007).

Em situações em que o padrão espacial está possivelmente presente, deve-se utilizar ferramentas de análise específicas para identificar e modelar o padrão existente. A estatística espacial é uma área da estatística que estuda metodologias para coleta, descrição, visualização e análise dos dados associados a posições geográficas que possam ser modelados como processos estocásticos (ASSUNÇÃO, 2001). Um processo estocástico em um espaço de probabilidade (Ω, F, P) é uma coleção de variáveis aleatórias indexadas por um conjunto não vazio G , isto é, $Z(x) = \{z(x_i); \forall x_i \in G\}$. Neste trabalho, o conjunto G será o espaço real p -dimensional, ou seja, $G \in R^p$. Cada um dos valores $z(x_i)$ pode ser visto como uma realização particular de uma variável aleatória $Z(x_i)$.

Os dados espaciais podem ser classificados em três tipos: a) processos pontuais – a variável de interesse é a localização do evento; b) superfície contínua (geoestatística) – são fenômenos caracterizados pela continuidade espacial da variável aleatória; c) dados de área – são caracterizados por eventos agregados por regiões poligonais e nesse caso não se conhece a localização exata do evento, apenas o número de ocorrências por região. Há métodos de análise específicos para cada tipo de dados. Assim, a análise espacial é composta por um conjunto de procedimentos encadeados cuja finalidade é a escolha de um modelo inferencial que considere o padrão espacial presente no fenômeno estudado. Geralmente o procedimento de análise espacial é precedido de uma análise exploratória com objetivo de identificar padrões de dependência espacial no fenômeno em estudo (CRESSIE, 1993).

Alguns conceitos são importantes para a análise de dados espaciais. O primeiro deles é a **dependência espacial**. A noção de dependência espacial parte da primeira lei da geografia de Waldo R. Tobler (1970) - “Tudo está relacionado a tudo mais, mas as coisas próximas estão mais relacionadas do que as coisas distantes”. Cressie (1993) afirma que a dependência espacial está presente em todas as direções e fica mais fraca à medida que aumenta a distância na localização dos dados. Pode-se afirmar, assim, que a maior parte das ocorrências, sejam naturais ou sociais, apresentam uma relação que depende da distância.

A dependência espacial é um importante conceito para compreensão dos fenômenos espaciais e pode ser avaliada por meio da **autocorrelação espacial**. Entende-se por autocorrelação espacial a correlação espacial entre realizações de uma mesma variável

aleatória medida em locais distintos do espaço. A autocorrelação espacial pode ser mensurada por diversos indicadores que verificam a existência de padrão espacial não aleatório a partir de uma comparação dos valores de uma amostra e de seus vizinhos. No geral, o padrão espacial é verificado através de: a) semivariograma como envelope simulado; b) índice de Moran (MORAN, 1950); ou c) teste de aleatorização de Mantel (RIBEIRO JÚNIOR et al., 2017). Valores significativos de índices de autocorrelação espacial são evidência de dependência espacial e indicam que o pressuposto de amostras independentes é inválido e que modelos inferenciais para esses casos devem considerar o espaço explicitamente em suas formulações.

Outra forma de verificar se um conjunto de dados tem padrão espacial e/ou temporal não aleatório é utilizando os automodelos nos quais a informação da vizinhança é considerada como covariável do modelo e, se essa for significativa, indica que existe padrão espacial.

a) Estacionariedade e isotropia

Um conceito importante na análise espacial, que pode ser aplicado a estudos geoambientais, é o da estacionariedade. Um processo estocástico é estacionário se todos os momentos são constantes e independentes da origem no espaço (GUIMARÃES, 2004). Assim, se um processo é dito estacionário de ordem K , temos:

$$E[Z(x)] = m_1(x) = m \text{ é constante para } \forall x \in G;$$

$$E[Z^2(x)] = m_2(x) \text{ é constante para } \forall x \in G;$$

$$\vdots$$

$$E[Z^k(x)] = m_k(x) \text{ é constante para } \forall x \in G.$$

Portanto, se um processo estocástico é estacionário de ordem K , então o processo será estacionário para todas as ordens inferiores K . No geral, os métodos estatísticos espaciais precisam apenas de estacionariedade de primeira e segunda ordem. Nessa situação, Guimarães (2004) preconiza que:

$E[Z(x)] = m_1(x)$ é constante para $\forall x \in G$: a média do processo estocástico é constante e independente da origem no espaço;

$E[Z^2(x)] = m_2(x)$ é constante para $\forall x \in G$: implica que a variância do processo estocástico é constante, independente da origem no espaço e que para cada par $\{Z(x), Z(x+h)\}$ a função de covariância existe e depende apenas da distância h .

Um processo é isotrópico se, além de estacionário, a covariância depende somente da distância entre os pontos e não da direção entre eles.

Quando o fenômeno em estudo, como no caso de variáveis de geoquímica inorgânica, revela diferentes padrões de dependência espacial, ou seja, apresenta uma variabilidade que não é a mesma em todas as direções, o fenômeno em estudo é chamado de anisotrópico. A anisotropia é uma característica muito frequente nos elementos da natureza, isto é, a variabilidade ou distribuição espacial de tais elementos ocorre mais intensamente numa direção e menos intensamente em outra direção. Inúmeros estudos demonstram a anisotropia em estudos geoquímicos (CHENG, 2012; CAO, CHENG, 2012; WANG, ZHAO, CHENG, 2013; ZUO, WANG, 2016; ZUO, CARRANZA, WANG, 2016).

b) Semivariograma

Existem fenômenos nitidamente não estacionários que tornam inadequados o uso de processos estocásticos estacionários. Para solucionar esse problema é habitual refletir sobre considerar uma nova classe de funções que, considerando a hipótese intrínseca, chamam-se funções aleatórias intrínsecas. Essa hipótese impõe a estacionariedade dos dois primeiros momentos dos acréscimos espaciais.

Com base nesse pressuposto pode-se definir o semivariograma. Primeiro, se considerar os acréscimos a distância h dados por:

$$\varepsilon(h) = Z(x + h) - Z(x), \forall x \in G. \quad (\text{A.1})$$

tem-se que as duas condições de estacionariedade são expressas por:

$$E\{\varepsilon(h)\} = 0, \quad (\text{A.2})$$

$$E\{[\varepsilon(h)]^2\} = 2\gamma(h). \quad (\text{A.3})$$

A partir do segundo momento dos acréscimos, define-se o variograma da variável Z como

$$\gamma(h) = \frac{1}{2}E\{[Z(x + h) - Z(x)]^2\} \quad (\text{A.4})$$

Desenvolvendo a função (A.4), tem-se

$$\begin{aligned} \gamma(h) &= \frac{1}{2}E\{[Z(x + h)]^2\} + \frac{1}{2}E\{[Z(x)]^2\} - E\{[Z(x + h)Z(x)]\} \\ 2\gamma(h) &= E\{[Z(x)]^2\} - E\{[Z(x + h)Z(x)]\}. \end{aligned} \quad (\text{A.5})$$

Somando-se e subtraindo-se m^2 ao 2º membro, obtém-se uma expressão simples que relaciona as duas funções:

$$\gamma(h) = \frac{c(0) - c(h)}{2}, \quad (\text{A.6})$$

em que $\gamma(h)$ representa o semivariograma, que é a metade do variograma. A equação (A.6) indica que, sob a hipótese de estacionariedade de 2ª ordem, a covariância e o semivariância são formas alternativas de caracterizar a autocorrelação dos pares $Z(x)$ e $Z(x + h)$ separados pelo vetor h .

O semivariograma é uma ferramenta gráfica útil para medir o grau de dependência espacial entre os elementos da amostra de uma variável, possibilitando medir a magnitude e a forma de tal dependência espacial.

A partir dos dados coletados, o semivariograma experimental é constituído considerando-se todos os pares de amostra $z(x)$ e $z(x + h)$, separadas pelo vetor distância h , a partir da equação:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i + h) - z(x_i)]^2, \quad (\text{A.7})$$

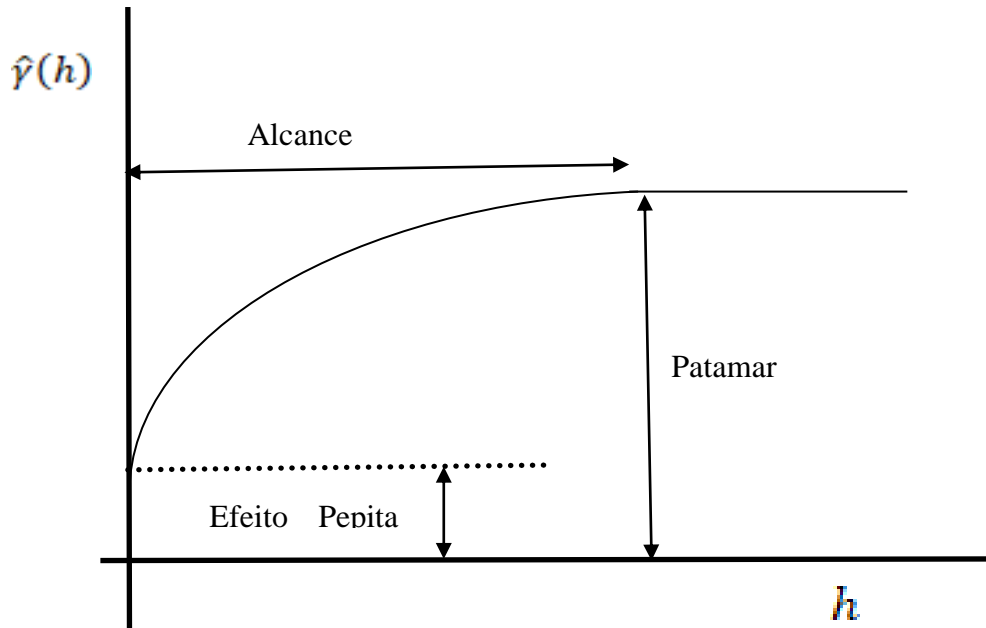
em que, $\hat{\gamma}(h)$ é a semivariância estimada e $N(h)$ é o número de pares de valores medidos com distância h , $z(x)$ e $z(x + h)$, separados pela distância h .

Há situações em que a variabilidade local não é constante e se modifica ao longo da área de estudo (heteroscedasticidade). Na prática pode-se testar a hipótese adicional de que o fenômeno é isotrópico; nesse caso, a determinação experimental do semivariograma depende apenas das distâncias entre as amostras e não da direção relativa entre elas.

Os parâmetros do semivariograma podem ser observados na Figura A.1:

- alcance (a): distância em que as amostras apresentam padrão espacial, para distâncias superiores a h o padrão é aleatório;
- patamar (C): valor do semivariograma correspondente ao seu alcance (a). Considera-se que a partir desse ponto a função se estabiliza e não existe mais dependência espacial entre as amostras;
- efeito pepita (C_0): este parâmetro refere-se à descontinuidade na origem do semivariograma. Idealmente, $\gamma(0) = 0$. Esta descontinuidade na origem do semivariograma, correspondendo à diferença entre as amostras de maior proximidade e gerada por micro regionalização, erros de amostragens ou erros de medidas.
- contribuição (C_1): diferença entre o patamar (C) e o efeito pepita (C_0).

Figura A.1 - Parâmetros obtidos em um semivariograma



B) GEOESTATÍSTICA

Os métodos geoestatísticos tiveram início em 1951 quando Krige, em seu trabalho sobre dados de concentração de ouro (Au), concluiu que somente a informação dada pela variância seria insuficiente para explicar o fenômeno em estudo, uma vez que os dados não eram independentes e apresentavam um padrão. Nesse trabalho foi sugerido considerar as distâncias entre as observações. A partir das observações de Krige (1951), Matheron (1971) desenvolveu a teoria das variáveis regionalizadas.

A teoria da variável regionalizada serviu de base para as suposições das ferramentas em geoestatística. Essa teoria pode ser entendida como uma função que varia, no espaço, com certa aparência de continuidade, ou seja, é uma variável cujos valores são correlacionados (de algum modo) como a posição espacial que ocupam.

Segundo Burrough (1987), a teoria das variáveis regionalizadas pressupõe que a variação pode ser expressa pela soma de três componentes: uma componente estrutural, associada a um vetor médio constante ou a uma tendência constante; uma componente aleatória, espacialmente correlacionada; e um ruído aleatório ou erro residual.

Na estimativa de variáveis regionalizadas, os métodos não estatísticos utilizados na geologia levam em conta apenas o aspecto espacial das amostras com base em apreciações pessoais ou simplesmente de acordo com a disposição do espaçamento das amostras obtidos

(GUERRA, 1988). Esses métodos supõem uma longa experiência no fenômeno de interesse; assim, a aplicação dessas técnicas tem como base a “intuição” do pesquisador deixando de considerar características importantes do fenômeno, tais como a variabilidade espacial.

Os métodos/técnicas geoestatísticas conseguem englobar o aspecto espacial com o aspecto aleatório. Esses métodos baseiam-se na teoria das variáveis regionalizadas a partir da qual é possível estudar a estrutura espacial possibilitando calcular o erro cometido na avaliação, levando-se em conta a posição real das amostras. Na definição de Matheron (1962; citado por JOURNAL; HUIJBREGTS, 1978), geoestatística é a aplicação do formalismo de funções aleatórias ao reconhecimento e predição de fenômenos naturais.

Há muitas ferramentas **geoestatísticas univariadas**, sendo o método da **krigagem** o mais utilizado. Krigagem é um processo estocástico de obtenção do melhor preditor linear não viciado de uma variável desconhecida. A krigagem pressupõe que o fenômeno estudado atende a hipótese de estacionariedade (MATHERON, 1971).

A grande diferença entre a krigagem e outros métodos de interpolação é a maneira como os pesos são atribuídos às diferentes amostras. Por exemplo, na interpolação por média simples, os pesos são todos iguais a $\frac{1}{n}$, em que n é o número de amostras; já na interpolação com base no inverso do quadrado das distâncias, os pesos são definidos como o inverso do quadrado da distância que separa o valor interpolado dos valores observados. Na krigagem o procedimento é semelhante ao da interpolação por média móvel ponderada, porém as informações são extraídas a partir do semivariograma para encontrar pesos ótimos a serem associados às amostras para a estimativa do valor em um determinado ponto (CÂMARA et al., 2004).

Outra abordagem de ferramentas geoestatísticas é a **multivariada**. As ferramentas geoestatísticas multivariadas são preferíveis quando as relações entre as variáveis são dependentes de escala e podem ser associadas a diferentes fontes de variabilidade. Essa abordagem denomina-se Análise de Cokrigagem Fatorial. Essa é uma combinação da tradicional Análise de Componentes Principais que descreve a correlação de um conjunto de dados multivariado, e a geoestatística que leva em consideração a natureza regionalizada das variáveis. Uma importante vantagem da Análise de Cokrigagem Fatorial é a possibilidade de extrair um número restrito de fatores dependentes da escala regionalizada (CÂMARA et al., 2004).

C) AMOSTRAGEM ESPACIAL

Em estudos ambientais há uma grande preocupação com o monitoramento do meio ambiente, motivo pelo qual se faz uso de diversos instrumentos diagnósticos como, por exemplo, a geoquímica¹. Em estudos geoquímicos, onde são utilizadas diferentes matrizes (rochas, solos, sedimentos, água, atmosfera, biota), é observada uma grande variabilidade nos dados que não é apenas explicada pela matriz utilizada. Assim, é fundamental possuir um bom planejamento de coleta de amostras que acarrete a redução dessa variação de resultados.

Um procedimento inicialmente comum em estudos ambientais, para se obter conhecimento de algumas características de um ambiente, é observar uma ou mais amostras de seus elementos, e a partir do resultado amostral, obter estimativas para a característica de interesse. O problema maior desse procedimento é garantir que a amostra seja representativa, ou seja, saber se os resultados obtidos pela amostra são confiáveis para avaliar as características da população e se são representativos da heterogeneidade do ambiente avaliado.

Há diversas razões para se fazer uso por amostragem. Os motivos mais comuns se devem: a) à busca de redução de custos para se estudar uma população, à redução do tempo e de mão-de-obra para a realização da coleta dos dados; b) à necessidade de confiabilidade e qualidade dos dados, considerando uma perda de precisão que seja insignificante em comparação com a enumeração da população (COCHRAN, 1977). A ressalva para esse procedimento está nos casos em que a população é pequena, quando a característica é de fácil mensuração ou quando há necessidade de alta precisão.

Em estatística, amostragem é a utilização de processos ou técnicas para obtenção de dados que possibilitem inferir sobre características de um conjunto (universo ou população) por meio do exame de uma parte desse conjunto (amostra). A finalidade da teoria da amostragem é utilizar a técnica mais eficiente para representar uma população com informações obtidas em uma amostra. Essa teoria tenta desenvolver métodos de seleção de amostras que forneçam estimativas, ao menor custo e com maior precisão possível ao objetivo de estudo (COCHRAN, 1977).

¹Segundo Goldschmidt(1923), a geoquímica se ocupa de: determinar a abundância relativa e absoluta de elementos da Terra; estudar a distribuição e migração de elementos individuais com o objetivo de descobrir os princípios que controlam a distribuição e migração dos elementos.

A pesquisa sobre amostragem tem uma longa história. Há evidências de que os chineses e os egípcios foram os primeiros a fazer aplicações intuitivas dos princípios de amostragem, mas a primeira tentativa conhecida para fazer declarações sobre uma população utilizando informações apenas de uma parte dela foi do comerciante inglês John Graunt (1620-1674). Seu famoso tratado descreve um método para estimar a população de Londres com base em informação parcial. Desde então, a teoria da amostragem tem se desenvolvido separadamente da teoria clássica da estatística e evoluiu para um extenso corpo teórico, de métodos e operações amplamente utilizados em diversas áreas (COCHRAN, 1946; COCHRAN, 1977).

Em algumas áreas, como nas ciências geográficas, técnicas de amostragens unidimensional/tradicional não são adequadas, visto que a variabilidade da população depende de sua localização (WANG, 2009). Para lidar com esse problema da amostragem bidimensional (amostragem espacial), foram propostas algumas soluções. Matheron (1971) propôs o uso da teoria da variável regionalizada, que foi bem construída e amplamente aplicada em Geociências. Uma outra abordagem é utilizar a autocorrelação espacial para melhorar a eficiência da amostragem em termos da estimativa do erro da variância em relação ao desenho amostral e o tamanho da amostra (STEIN; ETTEMA, 2003). Wang (2009; 2010) propôs o uso de estratificação espacial heterogênea para obter uma amostragem espacial com melhor eficiência na inferência.

D) MODELAGEM DOS PARÂMETROS FÍSICOS E QUÍMICOS

O mapeamento de parâmetros físicos e químicos é usualmente feito a partir de dados agregados por área. Embora os mapas resultantes sejam facilmente interpretáveis, com a agregação perde-se a estrutura espacial detalhada dos dados e, conseqüentemente, as variações espaciais em pequena escala não são detectáveis.

Outro problema na análise de dados de parâmetros físicos e químicos é que são raros os modelos usualmente empregados para a agregação estatística de dados dessa natureza, e geralmente carecem de interpretação espacial adequada, já que são baseados em suposições pouco realistas de independência. Portanto, foi preferível derivar um modelo estatístico para esses dados a partir da suposição da dependência espacial. Através desse processo foi possível estimar uma superfície para os parâmetros físicos e químicos. A abordagem semiparamétrica

de Modelos Aditivos Generalizados (GAM) e sua generalização (Modelos Aditivos Generalizados para posição, escala e forma – GAMLSS) (FLORÊNCIO, 2010), permitiram, para este estudo, a inclusão de covariáveis que caracterizaram a região para investigar a variação espacial dos parâmetros físicos e químicos trabalhados, tornando possível estimar uma superfície para cada parâmetro analisado.

E) MODELOS ADITIVOS GENERALIZADOS (GAM)

O GAM é uma extensão do modelo linear generalizado, em que o termo relacionado às covariáveis $\sum_i x_{ij}\beta_j$ é substituído por $\sum_j g_j(X_j)$, em que X_j são as covariáveis e $g_j(X_j)$ denota uma função semi paramétrica estimada por meio de funções de suavização (HASTIE; TIBSHIRANI, 1990). Com essa mudança, não é necessário assumir uma relação linear entre a variável resposta e as covariáveis, não sendo nem necessário conhecer a forma dessa relação, mas sim estimá-la a partir de um conjunto de dados. A função estimada $g_j(X_j)$, também é chamada de função de suavização ou curva suavizada.

A função de suavização permite descrever a forma da variável resposta e também revelar uma possível ausência de linearidade nas relações estudadas, uma vez que não apresenta a estrutura rígida de uma função paramétrica. Portanto, os GAM permitem especificar a relação entre a variável resposta e as covariáveis de forma flexível, evitando modelos complexos.

O GAM é considerado modelo semiparamétrico. Porém, como o modelo linear generalizado é paramétrico no que diz respeito à distribuição de probabilidade da variável resposta, é possível modelar alguns preditores de forma não-paramétrica por meio de termos lineares e polinômios (FLORÊNCIO, 2010).

F) MODELOS ADITIVOS GENERALIZADOS PARA POSIÇÃO, ESCALA E FORMA (GAMLSS)

Os modelos utilizados neste trabalho foram fundamentados nas classes de modelos estatísticos de regressão semiparamétrico, denominados de modelos aditivos generalizados

para posição, escala e forma (GAMLSS) propostos por Rigby e Stasinopoulos (2005). No modelo GAMLSS a premissa de que a variável resposta pertence à família exponencial é relaxada e substituída por uma família de distribuição mais geral. Além disso, a parte sistemática do modelo é ampliada para permitir a modelagem não apenas da média (ou posição), mas de todos os parâmetros da distribuição condicional da variável resposta, seja através de funções paramétricas ou não-paramétricas das variáveis explanatórias e/ou termos de efeitos aleatórios.

A classe de modelos de regressão GAMLSS permite a construção de modelos de regressão para os parâmetros de muitas distribuições, tais como Binomial, Exponencial, Gama, Weibull, Beta, dentre outras distribuições. Nessa abordagem é possível modelar separadamente todos os parâmetros da distribuição como função de variáveis explicativas.

Na estrutura da classe de modelos de regressão GAMLSS os p parâmetros $\theta = (\theta_1, \theta_2, \dots, \theta_p)^T$ de uma função de probabilidade $f(y/\theta)$ são modelados utilizando-se termos aditivos. Presume-se que para $i = 1, 2, \dots, n$ as observações y_i são independentes e condicionais a θ_i , com função densidade de probabilidade $f(y_i/\theta_i)$, em que $\theta_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{ip})^T$ é um vetor de p parâmetros relacionados às variáveis explanatórias e aos efeitos aleatórios. Em 2005, Rigby e Stasinopoulos definiram a classe de modelos de regressão GAMLSS da seguinte forma:

Seja $y^T = (y_1, y_2, \dots, y_n)$ um vetor de variáveis respostas, considera-se uma função de ligação monótona $g_k(\cdot)$, $k = 1, 2, \dots, p$ relacionando o k -ésimo parâmetro θ_k às variáveis explanatórias e aos efeitos aleatórios por meio de um modelo aditivo (Z_{jk}, γ_{jk}) dado:

$$g_k(\theta_k) = \eta_k = X_k \beta_k + \sum_{j=1}^{J_k} Z_{jk} \gamma_{jk} \quad (\text{A.8})$$

em que θ_k e η_k são vetores com dimensão n , por exemplo, $\theta_k = (\theta_{1k}, \theta_{2k}, \dots, \theta_{nk})^T$, $\beta_k = (\beta_{1k}, \beta_{2k}, \dots, \beta_{nk})^T$ é um vetor de parâmetros de tamanho J_k , X_k e Z_{jk} são matrizes de planejamento (covariáveis) fixas, conhecidas e de ordem $(n \times J_k)$ e $(n \times q_{jk})$, respectivamente. Já γ_{jk} é uma variável aleatória q_{jk} -dimensional. O uso da equação 8 justifica-se, para a análise de dados em geoquímica ambiental e mais particularmente para os dados da região da BTS sob análise neste trabalho, pela facilidade no uso dos algoritmos de reajuste (*backfitting*) e na

incorporação ao modelo de diferentes tipos de termos aditivos e/ou termos de efeitos aleatórios.

APÊNDICE B

MATERIAIS E MÉTODOS

Seguem algumas informações complementares sobre materiais e métodos adotados para atingir os objetivos traçados.

A) BANCO DE DADOS – SBDGG

Nesta pesquisa foram utilizados dados geoquímicos disponibilizados pelo SBDGG construído para a Baía de Todos os Santos por Lago et al. (2014).

Com o desenvolvimento dos estudos geoquímicos e aprimoramento tecnológico, são gerados muitos dados que devem ser tratados para avaliar quali-quantitativamente diversos ambientes pesquisados. Porém, inúmeros projetos enfrentam problemas com a quantidade de dados geoquímicos obtidos, advindos da coleta de amostras diferenciadas no tempo, com procedimentos analíticos distintos e arquivos de dados nem sempre recuperáveis ou confiáveis. O armazenamento dos dados é um grande desafio: há necessidade de organização de dados geoquímicos em arquivos sistematizados e confiáveis, que possam ser disponibilizados para diversos trabalhos. Assim, a proposta de serem conservados em sistemas tecnológicos do tipo computacional, como “Bancos de Dados” (BD), configura-se como uma importante decisão técnica para profissionais da geoquímica.

Nesse contexto, foi criado o Sistema Banco de Dados Geoquímicos Georreferenciados (SBDGG) vinculado ao NEA/IGEO/UFBA e ao Programa de Pós-Graduação em Geoquímica: Petróleo e Meio Ambiente (Pospetro) do IGEO. Esse BD foi construído visando possibilitar a entrada, organização e armazenamento de dados geoquímicos produzidos no âmbito das pesquisas realizadas junto ao NEA e ao Pospetro, em diferentes áreas do Estado da Bahia (LAGO et al., 2014). Os dados inseridos no SBDGG referem-se a análises (físico-químicas e de compostos inorgânicos e orgânicos) de sedimentos, água, vegetação (folhas, raízes, caules) e biota (moluscos, peixes), correspondentes a materiais coletados em locais georreferenciados.

Para a realização deste trabalho foram utilizados dados de geoquímica inorgânica de amostras de solos e/ou sedimentos do SBDGG entre os anos de 2011 a 2016, especificamente de Madre de Deus. Os dados originais são de Silva (2012).

B) PARÂMETROS FÍSICO-QUÍMICOS E METAIS

Para este estudo foram utilizados dados dos parâmetros físicos e químicos oriundos do projeto de pesquisa financiado pela Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) de número 4493/2009, intitulado “Disponibilidade de metais traço na Baía de Todos os Santos”. Essa pesquisa longitudinal propiciou um estudo que se traduziu em uma investigação mais aprofundada no município de Madre de Deus. Nesse projeto foram coletadas amostras de sedimentos em manguezais e em apicuns, e de solos em áreas urbanizadas ou não no município. Segundo Silva (2012), as amostras coletadas em campo foram liofilizadas, depois passaram por um pré-tratamento que consistiu na maceração, peneiramento com malha de 2mm e retirada de objetos estranhos na amostra. Para retirada das porções a serem analisadas, foi realizado o quarteamento dessas amostras. Com as amostras prontas para análise, realizou-se o fracionamento (segundo o método Tessier, 1979, adaptado, no qual se obteve as seguintes frações químicas: metais sob forma trocável; metais associados a carbonatos, à matéria orgânica; associados a óxidos e hidróxidos de Fe e de Mn; metais presentes na fração residual), extração parcial e digestão total para determinação dos metais: Pb, Cu, Zn, Cr, Ni, Fe, Mn. As amostras também foram analisadas para os seguintes parâmetros: potencial hidrogeniônico (pH); potencial redox (Eh); nitrogênio total (N); capacidade de troca catiônica (CTC); salinidade; granulometria; matéria orgânica (MO); fósforo assimilável (P). Os resultados analíticos obtidos por Silva (2012) foram inseridos no SBDGG e que foram utilizados na presente investigação científica.

C) ANÁLISE ESTATÍSTICA

Para a análise dos dados geoquímicos foram utilizadas algumas técnicas estatísticas relacionadas com os objetivos deste trabalho.

O primeiro objetivo desta pesquisa foi a análise e avaliação do padrão espacial dos parâmetros físicos e químicos trabalhados. No geral, a dependência espacial é verificada através do semivariograma com envelope simulado, índice de Moran, quando o conjunto de dados apresenta distribuição normal, ou teste de aleatorização de Mantel (RIBEIRO JÚNIOR et al., 2009).

Pesquisas ambientais por amostragem podem, a depender do problema avaliado, ser bastante onerosas, principalmente quando se contabiliza o custo laboratorial. Na geoquímica pouco é empregada a amostragem em que o fator espaço seja tão importante quanto o custo. Pretende-se avaliar, nesta primeira análise, a viabilidade do emprego de técnicas de amostragem espacial em estudos geoquímicos, no intuito de se obter a menor quantidade possível de amostras para redução de custos de coleta de campo e analíticos.

Para a geração de mapas temáticos da distribuição espacial dos parâmetros físicos e químicos, foram utilizadas técnicas de geoestatística multivariada com o objetivo de descrever o fenômeno estudado com base em informações da matriz ambiental em estudo.

Todas as análises estatísticas foram realizadas pelo software livre R (versão 3.2.1, R DEVELOPMENT CORE TEAM, 2015). Esse software possui uma linguagem e um ambiente de desenvolvimento integrado para cálculos estatísticos e gráficos. O uso desse software justifica-se por ser altamente expansível com uso de pacotes, bibliotecas para funções específicas ou áreas de estudo típicas, além de ser largamente utilizado para análise de dados. Ele é utilizado em integração com o programa QGis, por exemplo, que corresponde a um sistema de informações geográficas (SIG). Sendo necessário um georreferenciamento para tratar os dados espaciais e gerar mapas com modelos de distribuição de diferentes parâmetros e suas correlações em sedimentos e solos da região de Madre de Deus, na BTS.

D) APRESENTAÇÃO DOS RESULTADOS

Os resultados estão sendo apresentados sob a forma de artigos científicos, que contemplam:

- uma narrativa sobre amostragem espacial e um levantamento bibliométrico para analisar a evolução da abordagem dos termos relacionados com o tema, de forma a compreender melhor esse campo de pesquisa;
- uma avaliação do modelo de regressão utilizando o GAMLSS para compreender a influência dos parâmetros físicos e químicos no índice de partição reduzida (I_R), nas frações geoquímicas de um ambiente costeiro hipersalino denominado de apicum, na região de Madre de Deus, na BTS;

- uma verificação da existência de padrão espacial dos parâmetros geoquímicos inorgânicos do sedimento por meio da associação de técnicas geoestatísticas e de análise multivariada, visando contribuir com a melhoria no entendimento dos padrões espaciais existentes.

APÊNDICE C

COMANDO UTILIZADOS PARA ANÁLISE DO ARTIGO: IDENTIFICAÇÃO DE PADRÃO ESPACIAL EM DADOS GEOQUÍMICOS INORGÂNICOS – ESTUDO DE CASO: MADRE DE DEUS, BAHIA, BRASIL

```

#=====
# Análise Descritivas e análise fatorial dos dados de 2008 de Madre de Deus
#=====
require("MVA")
require("ellipse")
require(MVN)
require(CCA)
library(psych) # biblioteca para análise descritiva por grupo
dadosMFA = DadosMD[, 4:23] # dados para analise MFA
rownames(dadosMFA)<- DadosMD$X__1 # nomear as linhas do conjunto de dados
by(dadosMFA[, 2:17], dadosMFA$Ambiente, describe)# medidas descritivas para o conjunto de dados
by(dadosMFA[, 18:20], dadosMFA$Ambiente, describe)# medidas descritivas para o conjunto de dados
library(ggpubr) # pacote para plotar varios graficos em uma janela
#=====
# Teste de Kruskal
#=====
kruskal.test(pH ~ as.factor(Ambiente), data = dadosMFA)
kruskal.test(Salinidade ~ as.factor(Ambiente), data = dadosMFA)
kruskal.test(Eh ~ as.factor(Ambiente), data = dadosMFA)
kruskal.test(N ~ as.factor(Ambiente), data = dadosMFA)
kruskal.test(MO ~ as.factor(Ambiente), data = dadosMFA)
kruskal.test(P ~ as.factor(Ambiente), data = dadosMFA)
kruskal.test(Ca ~ as.factor(Ambiente), data = dadosMFA)
kruskal.test(Mg ~ as.factor(Ambiente), data = dadosMFA)
kruskal.test(K ~ as.factor(Ambiente), data = dadosMFA)
kruskal.test(Na ~ as.factor(Ambiente), data = dadosMFA)
kruskal.test(CTC ~ as.factor(Ambiente), data = dadosMFA)
# Granulometria
kruskal.test(Areia ~ as.factor(Ambiente), data = dadosMFA)
kruskal.test(Silte ~ as.factor(Ambiente), data = dadosMFA)
kruskal.test(Argila ~ as.factor(Ambiente), data = dadosMFA)
# Metais
kruskal.test(Cu ~ as.factor(Ambiente), data = dadosMFA)
kruskal.test(Zn ~ as.factor(Ambiente), data = dadosMFA)
kruskal.test(Mn ~ as.factor(Ambiente), data = dadosMFA)
kruskal.test(Fe ~ as.factor(Ambiente), data = dadosMFA)
kruskal.test(Cr ~ as.factor(Ambiente), data = dadosMFA)
#=====
# Correlacao dos dados
#=====
library(ggplot2)
library(GGally)

```

```

ggpairs(dadosMFA,columns=2:8, upper = list(continuous = wrap('cor', method = "spearman", size = 3,
hjust=0.5)), mapping = aes(color = Ambiente, alpha = .5)) + theme(text =
element_text(size=10),axis.text.x=element_text(angle=45,hjust=1,size=10))
ggpairs(dadosMFA,columns=9:11, upper = list(continuous = wrap('cor', method = "spearman", size = 3,
hjust=0.5)), mapping = aes(color = Ambiente, alpha = .5)) + theme(text =
element_text(size=10),axis.text.x=element_text(angle=45,hjust=1,size=10))
ggpairs(dadosMFA,columns=12:16, upper = list(continuous = wrap('cor', method = "spearman", size = 3,
hjust=0.5)), mapping = aes(color = Ambiente, alpha = .5)) + theme(text =
element_text(size=10),axis.text.x=element_text(angle=45,hjust=1,size=10))
ggpairs(dadosMFA,columns=2:16, upper = list(continuous = wrap('cor', method = "spearman", size = 3,
hjust=0.5)), mapping = aes(color = Ambiente, alpha = .5)) +
theme(text = element_text(size=9),axis.text.x=element_text(angle=45,hjust=1,size=9))
ggplot(dadosMFA, aes(x=CTC, y=Salinidade, color = Ambiente)) +
geom_point() + geom_density_2d()
#=====
# teste para as correlações
#=====
cor(dadosMFA[,2:20], method = "spearman")
tc <- corr.test(dadosMFA[,2:20], method = "spearman")
print(tc, digits = 4)
#=====
# Teste de normalidade multivariada
#=====
library(mvnormtest)
C <- t(as.matrix(dadosMFA[2:20])) # para usar o teste vamos transpor a matriz
mshapiro.test(C) # Teste de normalidade multivariada de Shapiro
#=====
# Teste de adequação da amostra - KMO e Bartlett
#=====
KMO(dadosMFA[, -c(1, 13:16)])# teste KMO para todas as variáveis dos dados (KMO=0,73)
cortest.bartlett(dadosMFA[, -c(1, 13:16)])# teste de esfericidade de Barlet para todas
#=====
# Determinação do numero de fatores
#=====
library(nFactors)
ev <- eigen(cor(scale(dadosMFA[, -c(1, 13:16)]))) # autovalores
ap <- parallel(subject=nrow(dadosMFA[, -c(1, 13:16)]),var=ncol(dadosMFA[, -c(1, 13:16)]),
rep=100,cent=.05)
nS <- nScree(x=ev$values, aparallel=ap$eigen$qevpea)
nS #
plotnScree(nS, ylab="Autovalores", xlab="Fatores", main = "")#gráfico para determinar o número de fatores
#=====
# Análise Fatorial Classica
#=====
library(GPArotation)
# primeiro rodou a análise fatorial classica e depois verificar qual apresenta melhor ajuste
# - Minimum Residual Factor Analysis
print( fa(scale(dadosMFA[, -c(1, 13:16)]), center = TRUE), 5, fm = "minres"))
# - Principal Axis Factor Analysis
print( fa(scale(dadosMFA[, -c(1, 13:16)]), center = TRUE), 5, fm = "pa"))
# - Weighted Least Squares Factor Analysis

```

```

print( fa(scale(dadosMFA[, -c(1, 13:16)]), center = TRUE), 4, fm = "wls"))
# - generalized least squares factor analysis
print( fa(scale(dadosMFA[, -c(1, 13)], center = TRUE), 5, fm = "gls"))
# - maximum likelihood factor analysis
print( fa(scale(dadosMFA[, -c(1, 13:16)]), center = TRUE), 4, fm = "ml"))
# - OLS
print( fa(scale(dadosMFA[, -c(1, 13:16)]), center = TRUE), 5, fm = "ols"))
# Componentes Principais
print(principal(scale(dadosMFA[, -c(1, 13:16)]), center = TRUE), nfactors = 4, covar = FALSE))
#=====
# Análise fatorial Robusta
#=====
library(robustfa)
summary(af_clas_pca = FaClassic(x = scale(dadosMFA[, -c(1, 13:16)]), center = TRUE), factors = 4, method =
"pca", scoresMethod = "none") # imprimir a análise de componentes principais
summary(FaClassic(x = dadosMFA[, -c(1, 13:16)]), factors = 4, method = "pca", scoresMethod = "none", cor =
TRUE)) # imprimir os resultados da análise fatorial clássica
af_rob_cor = FaCov(scale(dadosMFA[, -c(1, 13:16)]), center = TRUE), factors = 4, cor = TRUE, method =
"pca", scoresMethod = "regression", cov.control = CovControlSde())
summary(af_rob_cor) # imprimir a análise fatorial robusta
#=====
covMcd=CovRobust(x=scale(dadosMFA[, -c(1, 13:16)]), center = TRUE))
myplotDD(x = covMcd)
plot(x=covMcd, which="dd", main="")
#=====
# Armazenamento dos escores
#=====
escores = getScores(af_rob_cor) # escores da AFR
colnames(escores) <- c("Fator1", "Fator2", "Fator3", "Fator4")
dadosMD_fator = cbind(DadosMD[, 2:3], dadosMFA[, ], escores)
#=====
write.table(dadosMD_fator, "MD_Geo_Fatores.txt", sep = "\t")

```

TESTE DE ALEATORIZAÇÃO DE MANTEL

```

#=====
# Teste de aleatorização de Mantel
#=====
library(ggplot2)
# Matriz com as distancias entre os pontos de coleta das amostras: A (coordenadas x e y)
L = dadosMD_fator[, 1:2] #coordenadas UTM em metros
n = nrow(L) # dimensão da matriz A a ser criada com as distancias euclidianas
# cria a matriz a nula
A=matrix(c(0),n,n) # matriz para receber as distancias
# Calculo das distancias euclidianas
for (i in 1:n) {
  for (j in 1:n) {
    if (i==j) {
      A[i,j]=0 }
    else {
      A[i,j]=sqrt((L[i,1]-L[j,1])^2+(L[i,2]-L[j,2])^2) }
  }
}

```

```

}}
# Matriz B: matriz das distâncias entre as repostas - fatores
Z = as.matrix(dadosMD_fator[, 19]) # fatores
m=nrow(Z) # dimensão/observações da matriz Z
# Como em A, antes de encontrar a matriz B
# propriamente dita, cria-se uma matriz nula B
B=matrix(c(0),m,m) # recebe as distancias entre os fatores caculados
for (i in 1:m) {
  for (j in 1:m) {
    if (i==j) {
      B[i,j]=0 }
    else {
      B[i,j]=sqrt((Z[i,1]-Z[j,1])^2) }
    } }
#=====
# estatística de teste é dada pelo coeficiente de correlação de Spearman entre os elementos
# correspondentes de A e B
#=====
A1=matrix(c(A),n,1)
B1=matrix(c(B),m,1)
cor(A1,B1,method='spearman')
# Teste de Mantel
proc=function(q){
  # Lendo os dados
  aux=NULL
  # Teste de aleatorização
  for (i in 1:q){
    A1.al=sample(A1)
    correlA1=cor(A1.al,B1, method="spearman")
    if (correlA1>=cor(A1,B1, method="spearman"))
      aux[i]=1
    else aux[i]=0
  }
  pvalor=mean(aux)
  pvalor
}
proc(10000)
dados = as.data.frame(cbind(A1, B1))
graf_disp <- ggplot(dados, aes(x=B1, y=A1)) +
  geom_point() + geom_smooth(method="lm", se = F) +
  labs(y = "Distância Localização", x = "Distância resposta")
plot(graf_disp)
#=====
# Teste de aleatorização de Mantel por ambiente
#=====
# Matriz com as distancias entre os pontos de coleta das amostras: A (coodenadas x e y)
data_ma <- subset(dadosMD_fator, Ambiente == "ma", select=c(X, Y, Fator1, Fator2, Fator3, Fator4))
data_ap <- subset(dadosMD_fator, Ambiente == "ap", select=c(X, Y, Fator1, Fator2, Fator3, Fator4))
data_en <- subset(dadosMD_fator, Ambiente == "en", select=c(X, Y, Fator1, Fator2, Fator3, Fator4))
data_ur <- subset(dadosMD_fator, Ambiente == "ur", select=c(X, Y, Fator1, Fator2, Fator3, Fator4))

```

```

L = data_ur[, 1:2] #coordenadas UTM em metros
n = nrow(L) # dimensão da matriz A a ser criada com as distancias euclidianas
# cria a matriz a nula
A=matrix(c(0),n,n) # matriz para receber as distancias
# Calculo das distancias euclidianas
for (i in 1:n) {
  for (j in 1:n) {
    if (i==j) {
      A[i,j]=0 }
    else {
      A[i,j]=sqrt((L[i,1]-L[j,1])^2+(L[i,2]-L[j,2])^2) }
    } }
# Matriz B: matriz das distâncias entre as repostas - fatores
Z = as.matrix(data_ur[, 6]) # fatores
m=nrow(Z) # dimensão/observações da matriz Z
# Como em A, antes de encontrar a matriz B
# propriamente dita, cria-se uma matriz nula B
B=matrix(c(0),m,m) # recebe as distancias entre os fatores calculados
for (i in 1:m) {
  for (j in 1:m) {
    if (i==j) {
      B[i,j]=0 }
    else {
      B[i,j]=sqrt((Z[i,1]-Z[j,1])^2) }
    } }
#=====
# estatística teste é dada pelo coeficiente de correlação de Pearson entre os elementos
# correspondentes de A e B
#=====
A1=matrix(c(A),n,1)
B1=matrix(c(B),m,1)
cor(A1,B1,method='spearman')
# Teste de Mantel
proc=function(q){
  # Lendo os dados
  aux=NULL
  # Teste de aleatorização
  for (i in 1:q){
    A1.al=sample(A1)
    correlA1=cor(A1.al,B1, method="spearman")
    if (correlA1>=cor(A1,B1, method="spearman"))
      aux[i]=1
    else aux[i]=0
  }
  pvalor=mean(aux)
  pvalor
}
proc(10000)

```

```

#=====
# Converter coordenadas UTM em lat/log
#=====
library(PBSmapping) # carregando o pacote PBS mapping
# coordenadas em UTM importadas contendo duas colunas, X e Y
# converter as coordenadas
coord = dadosMD_fator[,1:2]
# A função solicita que seja feitas duas coisas com os dados:
# especificar para os dados a zona e a projeção, neste caso 22 e UTM.
attr(coord, "zone")<- 24
attr(coord, "projection")<-"UTM"
# Convertendo de UTM para lat/long
# use "km=F" caso as coordenadas UTM estejam em metros e "southern=T" sendo no hemisfério sul.
dataconv<-convUL(coord, km=F, southern=T)
head(dataconv)
plot(dataconv)
#=====
# Criando a variável Amb - categorização da variável Ambiente
dadosMD_fator$Amb[dadosMD_fator$Ambiente=="ma"] <- 1
dadosMD_fator$Amb[dadosMD_fator$Ambiente=="ap"] <- 2
dadosMD_fator$Amb[dadosMD_fator$Ambiente=="en"] <- 3
dadosMD_fator$Amb[dadosMD_fator$Ambiente=="ur"] <- 4
#=====
# Juntar as coordenadas convertidas aos dados
dadosMD_espacial = cbind(dataconv, dadosMD_fator[, 23:27])
#=====
# carregar a biblioteca geoR
library(geoR)
# colocar os fatores no formato "geodata" do geoR
fator1 = as.geodata(dadosMD_espacial, coords.col = 1:2,data.col=3, covar.col = 7)
fator2 = as.geodata(dadosMD_espacial, coords.col = 1:2,data.col=4, covar.col = 7)
fator3 = as.geodata(dadosMD_espacial, coords.col = 1:2,data.col=5, covar.col = 7)
fator4 = as.geodata(dadosMD_espacial, coords.col = 1:2,data.col=6, covar.col = 7)
# lendo dados tipo shapefile
library(maptools)
borda_md = readShapePoly(file.choose())
plot(borda_md)
# Armazenando as bordas do município em um objeto
bordas = borda_md@polygons[[1]]@Polygons[[1]]@coords
#inserindo as bordas nos dados
fator1$borders = bordas
fator2$borders = bordas
fator3$borders = bordas
fator4$borders = bordas

#=====
# Construção dos Mapas
#=====
library(spatstat)
library(inline)
library(rgdal)

```

```

library(maptools)
library(ncf)
library(raster)
dis <- pointDistance(dadosMD_espacial[, 1:2], lonlat=FALSE)
max(dis)

borda_idw <- as.owin(borda_md)
fator1_idw      =      ppp(dadosMD_espacial$X,dadosMD_espacial$Y,      window=borda_idw,
marks=dadosMD_espacial$Fator1)

# Estimaco e impresso do mapa utilizando o IDW para o fator 1

idw_res1 <- idw(fator1_idw, power = 2) #only takes a minute
plot(idw_res1, main = "", col = terrain.colors(21))
image(idw_res1, col = terrain.colors(21), x.leg = c(-38.610, -38.600), y.leg = c(-12.750, -12.748))
#=====
# Como os fatores 2 a 4 apresentam padro espacial para realizar a composio dos mapas
# ser utilizada a Krigagem.
#=====
# visualizar os dados para o fator 2
fator2.ml <- likfit(fator2, trend = "1st", ini=c(1.0, 0.002), nug = 0.1)
fator2.gr <- pred_grid(fator2$borders, by = 0.00018)
KC <- krige.control(obj.m = fator2.ml, trend.d = "1st", trend.l = "1st")
OC <- output.control(simulations = TRUE, n.pred = 1000, quantile = c(0.1, 0.25, 0.5, 0.75,
0.9), threshold = 350)
fator2.kc <- krige.conv(fator2, loc = fator2.gr, krige = KC, output = OC)
names(fator2.kc)
image(fator2.kc, col = terrain.colors(21), x.leg = c(-38.610, -38.600), y.leg = c(-12.750, -12.748))
#=====
fator3.ml = likfit(fator3, ini=c(1.0, 0.002), nug = 0.4)
summary(fator3.ml)
fator3.gr <- pred_grid(fator3$borders, by = 0.00018)
KC1 <- krige.control(obj.m = fator3.ml)
OC1 <- output.control(simulations = TRUE, n.pred = 1000, quantile = c(0.1, 0.25, 0.5, 0.75,
0.9), threshold = 350)
fator3.kc <- krige.conv(fator3, loc = fator3.gr, krige = KC1, output = OC1)
names(fator3.kc)
image(fator3.kc, col = terrain.colors(21), x.leg = c(-38.610, -38.600), y.leg = c(-12.750, -12.748))
#=====
fator4.ml <- likfit(fator4, ini=c(1.0, 0.002), nug = 0.4)
summary(fator4.ml)
fator4.gr <- pred_grid(fator4$borders, by = 0.00018)
KC2 <- krige.control(obj.m = fator4.ml)
OC2 <- output.control(simulations = TRUE, n.pred = 1000, quantile = c(0.1, 0.25, 0.5, 0.75,
0.9), threshold = 350)
fator4.kc <- krige.conv(fator4, loc = fator4.gr, krige = KC2, output = OC2)
names(fator4.kc)
image(fator4.kc, col = terrain.colors(21), x.leg = c(-38.610, -38.600), y.leg = c(-12.750, -12.748))

```


Ficha catalográfica elaborada pelo Sistema Universitário de Bibliotecas (SIBI/UFBA),
com os dados fornecidos pelo(a) autor(a).

Dourado, Gilson
AJUSTE DO MODELO ADITIVO GENERALIZADO E
IDENTIFICAÇÃO DE PADRÃO ESPACIAL EM DADOS GEOQUÍMICOS
DO MUNICÍPIO DE MADRE DE DEUS / Gilson Dourado. --
Salvador, 2019.
104 f.

Orientadora: Gisele Hadlich.
Coorientadora: Denise Viola.
Tese (Doutorado - Programa de Pós-Graduação em
Geoquímica: Petróleo e Meio Ambiente (Pospetro)) --
Universidade Federal da Bahia, Instituto de
Geociências, 2019.

1. Geoquímica Ambiental. 2. Análise Fatorial. 3.
Padrão espacial. 4. Geoestatística. 5. Mapeamento
Científico. I. Hadlich, Gisele. II. Viola, Denise.
III. Título.