

Krigagem dos resíduos - Aplicação para estimativa de sólidos inorgânicos em suspensão através de sensoriamento remoto

Daniel Andrade Maciel

Análise Espacial – SER 301-3 Instituto Nacional de Pesquisas Espaciais

Resumo

De fundamental importância para os ciclos biogeoquímicos existentes em um corpo d'água, a carga de sólidos em suspensão afeta diretamente os meios físico, químico e biológico dos rios, lagos e oceanos. Devido a esta importância, seu monitoramento e quantificação é essencial para um eficiente gerenciamento dos recursos hídricos. Este monitoramento é feito, principalmente, através de amostragens em campo e técnicas de laboratório, o que são custosas e demandam tempo. O sensoriamento remoto surge, assim, para estimar a concentração de sólidos em suspensão através de modelos empíricos que relacionam a interação da radiação eletromagnética com os componentes opticamente ativos presentes no curso d'água. Estes modelos porém, devido aos fatores ambientais, apresentam diversos erros. Assim, técnicas de geoestatística como a krigeagem podem ser aplicadas visando a diminuição destes erros, caso estes possuam alguma dependência espacial. Foram usadas técnicas chamadas de Regression Kriging e Kriging With External Drift, que foram comparadas a um modelo empírico simples. Foram observados melhores resultados para a aplicação do modelo de krigagem, com valores de R² (0,55 para validação), RMSE (134 mg L⁻¹) e MAPE (27,7 %) maiores para a krigagem aplicada.

Estado da Arte

Importância dos sólidos em suspensão

A carga de sólidos em suspensão afeta de diferentes maneiras os ambientes aquáticos, sendo controlados pela concentração, tempo de exposição, composição química e tamanho de partículas dos sólidos em suspensão. O aumento de sua concentração provoca alterações físicas, químicas e, consequentemente biológicas no corpo d'água (BILLOTA e BRAZIR, 2008). Estas alterações prejudicam todo o ecossistema dependente do curso d'água, causando prejuízos a todos os utilizadores deste sistema.

Como alterações físicas, incluem-se a diminuição da profundidade da zona eufótica, a criação de gradientes de circulação, mudanças de temperatura e assoreamento de rios, canais e reservatórios (KIRK, 1985; TUNDISI e TUNDISI, 2008; BILLOTA e BRAZIR, 2008). Além destes problemas, associados ao aumento da concentração de sólidos em suspensão, destacam-se também a redução da navegabilidade, a diminuição da vida útil dos reservatórios e o aumento nos custos do tratamento da água para consumos diversos, principalmente o humano (VON SPERLING, 1996; BILLOTA e

BRAZIR, 2008). Segundo Billota e Brazir (2008), alterações químicas podem ser atribuídas ao lançamento de contaminantes adsorvidos pelos sólidos, como pesticidas, metais pesados, provenientes de zonas de agricultura e mineração, por exemplo (DAWNSON e MACKLIN, 1998; KRONVANG et al., 2003; NASRABADI et al., 2016) e nutrientes (WU et al., 2015) no curso hídrico. Este aumento na quantidade de nutrientes, principalmente nitrogênio e fósforo está diretamente ligado à eutrofização do curso d'água, ligada a agricultura extensiva e também ao esgoto (VON SPERLING, 1996).

Em relação as implicações biológicas do aumento da concentração de sólidos em suspensão, destaca-se seu impacto sobre diferentes populações de organismos planctônicos da base da cadeia alimentar, pequenos invertebrados até ao comprometimento da produção pesqueira (CORDONE e KELLEY, 1961; BILLOTA e BRAZIR, 2008). A diminuição da quantidade e da qualidade da energia radiante incidente no corpo d'água devido ao aumento da turbidez é um fator limitante na produtividade primária, já que estes seres autotróficos aquáticos irão competir não apenas entre si pela radiação incidente, mas também com outros componentes absorvedores presentes no meio (KIRK, 1985; TUNDISI e TUNDISI, 2008; KIRK, 2011). Para os perífitons, este aumento na carga de sólidos em suspensão também acarreta na redução do desenvolvimento de filamentos e, em velocidades altas de escoamento, o fluxo pode ser abrasivo e danificar as suas estruturas fotossintéticas (STEINMAN e McINTIRE, 1990; BILOTTA e BRAZIER, 2008; CHAPMAN, HAYWARD e FAITHFUL, 2017).

Com a diminuição da principal fonte de alimento de pequenos invertebrados aquáticos (e.g. zooplâncton, invertebrados bentônicos) e peixes, essas populações também decrescem em ambientes turvos (CORDONE e KELLEY, 1961; RAVEN, EVERT e EICHHORN, 2007; CHAPMAN, HAYWARD e FAITHFUL, 2017). Além disso, outros efeitos são relatados a estas populações com o aumento da carga de sólidos em suspensão. A sedimentação destes sólidos afeta diretamente a comunidade de invertebrados bentônicos, causando sufocação física e redução da permeabilidade dos sedimentos (CHAPMAN, HAYWARD e FAITHFUL, 2017). Já para o zooplâncton, causa toxicidade e ingestão de partículas sem valor nutricional, ao mesmo tempo em que pode proporcionar proteção e abrigo contra predadores, como peixes (DONOHUE e MOLINOS, 2009). Segundo Kirk (1992), a diminuição na taxa de crescimento de populações de zooplancton em ambientes turvos, pode ser atribuída também à diminuição da sobrevivência e da fecundidade com o aumento das cargas de sedimentos, sendo a

população mais jovem a mais afetada (KIRK e GILBERT, 1990). De acordo com Billota e Brazier (2008), as comunidades de peixes podem ser afetadas de diferentes maneiras. A sedimentação de partículas pode bloquear os poros na estrutura de cascalho do fundo do curso d'água, com grandes efeitos nos ovos e larvas dos peixes (CORDONE e KELLEY, 1961; HARROD e THEURER, 2002; SUEDEL, WILKENS e KENNEDY, 2016). Além disto, a presença de sólidos em suspensão pode também afetar peixes adultos, entupindo e sendo abrasiva a suas estruturas branquiais e também estressando e afetando seu sistema imunológico, deixando os peixes mais vulneráveis a doenças, além de também prejudicar a migração (CORDONE e KELLEY, 1961; HARROD e THEURER, 2008).

Estas alterações físico-químico-biológicas, provocam, também, um grande prejuízo econômico. Os prejuízos relacionados ao aumento da carga de sólidos podem ser encontrados em diversas áreas. Koteen, Alexander e Loomis (2002) mostram que o aumento da carga de sólidos em suspensão e da salinidade em áreas de agricultura, ao longo do Rio Colorado, nos Estados Unidos provocou um aumento nos custos anuais, passando de cerca de U\$\$ 10 para uma concentração de 800 mg L-1 para cerca de U\$\$ 175, para cada acre-pé de água utilizado. Já Tundisi e Tundisi (2010) apontam que, em áreas protegidas e com mananciais de boa qualidade, o custo do tratamento da água é de cerca de R\$ 2,00 a R\$ 3,00 por 1000m³. Já em mananciais desmatados, este custo pode chegar a R\$250,00 ou R\$300,00 por 1000m³, aumento este, devido a necessidade de maior uso de floculantes, coagulantes e desinfetantes para atender aos padrões de potabilidade. Ainda, Heberling et al. (2015) mostram que uma redução de 1% no valor da turbidez da água bruta leva a uma diminuição nos custos de tratamento de U\$\$ 1123,00 ao ano. Pimentel et al. (1995) estimaram o custo anual causado pela erosão hídrica nos Estados Unidos, em corpos d'água, chegando a marca de 5,063 bilhões de dólares. Estes custos se relacionam à prejuízos em atividades recreacionais (2,44 bilhões), instalações de armazenamento de água (841,8 milhões), navegação (683 milhões) e outros usos do curso d'água (1,098 bilhões).

Assim, vê-se a necessidade do monitoramento da concentração dos sólidos inorgânicos em suspensão em corpos d'água como medida de controle, fiscalização e avaliação de todo o comportamento do uso do solo na bacia hidrográfica. Além disto, este monitoramento serve como base para planejamentos governamentais de recuperação de áreas degradadas, avaliação de processos de assoreamento e transporte de sedimentos e

poluentes, determinação da quantidade de coagulantes em estações de tratamento de água, psicultura, além de outros usos.

A determinação da carga de sólidos inorgânicos em suspensão pode ser obtida através de algumas técnicas de laboratório. No estado de São Paulo, a Norma Técnica Interna SABESP NTS 013 (SABESP, 1999) segue a metodologia proposta por APHA (1999), em que uma alíquota da amostra de água é filtrada através de um filtro com diâmetro de poros entre 0,7 à 2 µm. Esta passagem separa os sólidos suspensos dos sólidos dissolvidos. Após isto, o filtro é pesado com o resíduo retido é submetido de 105 °C por uma hora para ser determinado a carga de sólidos em suspensão. Para a separação da fração inorgânica (chamada de fixa na metodologia), submete-se o resíduo a uma temperatura de 550°C até peso constante. Esta metodologia, porém, é custosa e demanda muito tempo. Assim, são necessárias técnicas de avaliação mais eficientes da concentração de sedimentos inorgânicos em suspensão. Entre estas técnicas, a avaliação da interação da energia eletromagnética com o corpo d'água e seu comportamento com a variação dos componentes opticamente ativos vem sendo de grande auxílio para a caracterização da concentração de sedimentos em suspensão. Como a refletância da água aumenta com a concentração de sólidos em suspensão, principalmente dos sólidos inorgânicos, devido ao seu alto valor de índice de refração em relação à água (KIRK, 2011), é possível através de dados das propriedades ópticas aparentes e inerentes do corpo d'água, a criação de modelos empíricos e semi-analíticos para relacionar estes dados com a concentração de sólidos em suspensão (MUNDAY JR e ALFOLDI, 1979; NOVO, HANSOM e CURRAN, 1989; NOVO, STEFEN e BRAGA, 1991; GOODIN et al., 1993; BOSS et al., 2009; NECHAD et al., 2010; BRANIGAN, 2013; LOBO et al., 2015; KNAEPS et al., 2015; ROTTA el al., 2016; CHU et al., 2018).

Modelos empíricos e semi-empíricos funcionam a partir da relação estatística entre a variável dependente e a variável independente, que pode ser representada por uma ou mais variáveis. A análise de regressão, para os modelos empíricos é o método mais utilizado (MONTANHER, 2013). A análise de regressão contempla uma família de funções chamadas de Modelos Lineares Generalizados (GLM), em que assumem uma relação linear entre as variáveis dependentes e independentes. Já para relações nãolineares, modelos aditivos generalizados (GAM) são utilizados (HENGL, 2009). A regressão linear simples, com n-variáveis, pode ser representada através da Equação 1 (HENGL, 2007)

$$\hat{z}(s_0) = \sum_{k=0}^p \hat{\beta}_k * q_k(s_0) + e(s_0)$$
(1)

Onde $q_k(s_0)$ são os valores das variáveis auxiliares, $\hat{\beta}_k$ são os coeficientes estimados da regressão, que podem ser estimados por ajustes como o método dos mínimos quadrados ou por técnicas iterativas, como redes neurais e simulação Monte Carlo. p é o número de variáveis auxiliares utilizadas e $e(s_0)$ é o resíduo da regressão. Assim, os métodos empíricos e semi-empíricos irão utilizar os coeficientes obtidos na regressão para estimar a concentração de TSS em locais não amostrados.

O erro associado $e(s_0)$ é dependente de diversos fatores, como interferências atmosféricas, reflexão especular, resoluções radiométricas dos sensores, heterogeneidade de substâncias presentes no corpo d'água, além de diferenças entre o horário de medição em campo e a passagem do satélite (para dados orbitais) (MONTAHNHER, 2013). Além disto, segundo Hengl (2009), devido à complexidade das variáveis ambientais, esta relação é tão complexa que não pode ser modelada precisamente. Isto acontece porque não se conhece precisamente: a) todas as variáveis auxiliares necessárias para o modelo; b) as equações requeridas para modelar as entradas e saídas; c) a significância da componente aleatória no sistema (HENGL, 2009). Desta maneira, se o resíduo de um modelo empírico gerado apresenta uma variação de valores no espaço é possível modelala, através da interpolação por diferentes técnicas, principalmente as conhecidas como krigeagem.

A técnica de Krigagem foi inicialmente criada por um engenheiro de minas para melhorar as estimativas de minério em jazidas (Krige, 1951) e, após cerca de 10 anos, Matheron (1962) deriva as formulas e estabelece o campo da geoestatística linear. A versão padrão da krigeagem é chamada de Krigagem Ordinária (Ordinary Kriging, OK). Dados geoestatísticos são dados que, em teoria, poderiam ser avaliados em qualquer local, porém tipicamente vem apenas de uma pequena amostragem, como por exemplo, dados de amostragem de qualidade de água (BIVAND et al., 2008). Em uma aproximação puramente geoestatística, as predições em um determinado ponto e posição ($\hat{z}(s_0)$) são comumente feitas calculando pesos médios sobre determinadas observações (WEBSTER e OLIVER, 2001), sendo que

$$\hat{z}(s_0) = \sum_{i=0}^n \lambda_i * z(s_i) \tag{2}$$

Em que: $\hat{x}(s_0)$ é o valor predito em um local (s₀) não visitado. Os pesos λ_i são escolhidos para que o a variância do erro de predição seja minimizado, sendo estes pesos dependentes da estrutura de autocorrelação espacial dos dados. Este é o método conhecido como Krigagem Ordinária. Alternativamente a krigagem, existem os métodos empíricos de regressão (Equação 1). Considerando estas duas técnicas, um método híbrido de interpolação é conhecido como Regression-Kriging, que inicialmente usa um modelo empírico para estimar a parte empírica da Equação 3 e utiliza uma krigeagem simples com média zero para interpolar os valores dos resíduos obtidos na parte empírica (HENGL, 2007). Desta maneira, se o resíduo de um modelo empírico gerado apresenta uma dependência espacial, é possível modela-la, através da interpolação pela Regression Kriging, por exemplo. Sendo o valor de uma variável em determinado ponto a soma das componentes determinísticas (e.g. modelo empírico) e da componente estocástica (resíduo) (Equação 2), podemos calcular estatisticamente a parte determinística da equação e estimar a parte estocástica através da interpolação, por exemplo, através da krigagem (HENGL, 2007; 2009).

$$\hat{z}(s_{0}) = \sum_{k=0}^{p} \hat{\beta}_{k} * q_{k}(s_{0}) + \hat{e}(s_{0})$$

$$= \hat{m}(s_{0}) + \hat{e}(s_{0})$$

$$= \sum_{k=0}^{p} \hat{\beta}_{k} * q_{k}(s_{0}) + \sum_{i=1}^{n} \lambda_{i} e(s_{i})$$
(3)

Em que, $\hat{e}(s_0)$ são os resíduos interpolados, λ_i são os pesos da krigeagem, obtidos através da estrutura de dependência espacial dos resíduos e $e(s_i)$ são as localizações destes resíduos e $\hat{m}(s_0)$ é o modelo empírico ajustado. Se os coeficientes da regressão forem estimados pelo método dos quadrados mínimos generalizados (GLS), temos que os betas serão estimados pela Equação 4.

$$\hat{\beta}_{GLS} = (q^T * C^{-1} * q)^{-1} * q^t * C^{-1} * z$$
(4)

Sendo que, $\hat{\beta}_{GLS}$ é o vetor para os coeficientes estimados da regressão, **C** é a matriz de covariância dos resíduos, **q** é a matriz dos preditores em uma localização amostrada e **z** é o vetor dos valores medidos da variável amostrada. Uma vez que os coeficientes

empíricos são estimados, os resíduos são interpolados por krigagem e adicionados ao valor de \hat{z} (s_0). Se considerarmos uma denotação matricial, temos que:

$$\hat{z}(s_0) = q_0^T * \hat{\beta}_{GLS} + \lambda_0^T * (z - q \,\hat{\beta}_{GLS})$$
(5)

Assim, $\hat{z}(s_0)$ é o valor predito pela RK em uma posição s_0 , q_0 é o vetor de predição para $p + 1 e \lambda_0$ é o vetor de **n** pesos da krigageam, que é utilizado para a interpolação dos resíduos, dado na Equação 5 por $*(z - q \hat{\beta}_{GLS})$. Segundo Hengl (2007), os erros de predição do modelo podem ser dados pela Equação 6.

$$s_{RK}^{2}(s_{0}) = (C_{0} + C_{1}) - C_{0}^{T} * C^{-1} * c_{0} + (q_{0} - q^{T} * C^{-1} * c_{0})^{T} * (q^{T} * C^{-1} * q)^{-1} + (q_{0} - q^{T} * C^{-1} * c_{0})$$
(6)

Em que, $C_0 + C_1$ é a variação de alcance e c_0 é o vetor de covariâncias dos resíduos em locais não visitados. Este processo de estimativa é iterativo. Inicialmente, o modelo empírico é estimado por OLS e a função de covariância dos resíduos é utilizada para estimar os coeficientes da GLS. Estes são usados para recalcular os resíduos e uma nova função de covariância é calculada, até a estabilização.

Na literatura sobre geoestatística, diversos nomes são dados a técnicas em que os processos são muito similares ou ainda são os mesmos, em essência (HENGL, 2007). Basicamente, todos seguem os pressupostos de Matheron (1969), que introduziu a Krigagem Universal (UK). Esta, é baseada em uma modelagem de tendência a partir valores das coordenadas (i.e. posição das amostras). Quando o *trend* é feito a partir de variáveis auxiliares externas, ao invés das coordenadas, dá-se o nome de Kriging With External Drift (KED) (WACHERNAGEL, 1998; HENGL, 2007; BIVAND et al., 2008). Na KED e na UK são feitas com a krigeagem, com a diferença que a matriz de covariância dos resíduos é extendida para as variáveis auxiliares incluídas na KED. Entretanto, se nós podemos estimar o modelo de tendência empírico utilizado na KED separadamente dos resíduos, e depois soma-los, chegando na Regression-Kriging, vista na Equação 5. Goovaerts (1997) nomeia esta família de interpoladores baseados na relação entre as variáveis auxiliares e o espaço de *Kriging with a trend model* e denomina a RK de *Simple Kriging with varying local mean*. Hengl (2007) prova, matematicamente, que a KED é

igual a RK, dados os mesmos valores de entrada. Este tipo de técnica é conhecido, em inglês como *Regression-Kriging* (RK). Diversos trabalhos apresentam melhorias em resultados utilizando a RK ao invés de outros métodos de krigagem (, porém, para a relação entre dados de sensoriamento remoto e a predição de componentes opticamente ativos no corpo d'água, não foram encontrados relatos de utilização desta técnica na literatura. Webster e Oliver (2007) apontam como desvantagem da RK a estimativa dos coeficientes do modelo empírico pelo método OLS que, apesar de ser imparcial, não irá estimar a mínima variância a menos que a amostragem seja aleatória. Outra desvantagem é que as estimativas de semivariância obtidas dos resíduos são tendenciosas, porque dependem de uma relação não-linear entre os parâmetros estimados, que são, pela propagação de erros, estimados erroneamente. Assim, o variograma pode ser estimado erroneamente, aumentando o viés com o aumento dos valores dos *lags* (i.e. distância em que os pares do semivariograma são calculados).

Hengl (2007) apresenta a utilização da RK para o mapeamento da temperatura de superfície na Croácia, com cerca de 95% da variação explicada. Já Riso-Decelis et al. (2017), utilizaram a técnica de RK em um rio no México, utilizando como variáveis auxiliares pontos de descarga de efluentes, obtendo um RMSE de 20 NTU para turbidez, que variaram em um range de 0 - 80 NTU, muito altos quando se comparados a modelos utilizando sensores remotos. Dogliotti et al. (2015), para um *range* alto de turbidez (1,8 – 988) encontraram valores de RMSE de 20,5 mg L⁻¹. Claramente, vale lembrar que as variáveis auxiliares e também o ambiente utilizado é diferente. Porém, ainda não se encontram na literatura estudos utilizando a RK, com variáveis auxiliares de sensoriamento remoto para predizer os resíduos derivados de modelos empíricos em ambientes aquáticos.

Assim, o objetivo deste trabalho foi avaliar a Regression Kriging e a Kriging With External Drift na predição dos valores de sólidos inorgânicos em suspensão no Lago Curuai – PA e compara-las com os resultados obtidos por uma regressão empírica tradicional.

Metodologia

Área de Estudo

A área de estudo deste trabalho é denominada de planície do Lago Grande de Curuai (Figura 1) e está localizada a 90 km a noroeste da cidade de Santarém – PA. Este local é considerado representativo da planície de inundação do médio Amazonas e é alvo de diversas pesquisas (ARRAUT et al., 2005; BARBOSA, 2005; NOVO et al., 2006; CARVALHO et al., 2015). A planície do Lago Grande do Curuai possui uma área de aproximadamente 3500 km², constituída por mais de 20 lagos que são ligados por canais que se mantém durante todo o ciclo hidrológico, possui uma variação até 7 metros do nível d'água durante os períodos de cheia / seca.

Tabela 1 Localização da área de estudo com os pontos amostrais utilizados por Barbosa (2005) e cedidos para este trabalho.



Dados de concentração de Sólidos Inorgânicos em suspensão

Os dados de concentração de sólidos inorgânicos em suspensão foram obtidos através de Barbosa (2005). Foram utilizados 73 pontos com amostras de TSI. As amostragens foram realizadas em um período de 10 dias, com início em 22/11/2003 e término em 02/12/2013. Na Tabela 1 estão apresentados os valores máximos, mínimos e os valores médios da concentração dos componentes opticamente ativos presentes na amostragem.

Tabela 2 Valores máximos, mínimos e médios para os componentes opticamente ativos presentes na amostragem da planície do Lago Grande do Curuai. TSO = Total de Sólidos Orgânicos em Suspensão; TSI = Total de Sólidos Inorgânicos em Suspensão; TSS = Total de Sólidos em Suspensão; DIC = Carbono Inorgânico Dissolvido; DOC = Carbono Orgânico Dissolvido.

TSO	TSI	TSS	CLOROFILA	DIC	DOC
-----	-----	-----	-----------	-----	-----

	(mg L ⁻¹)	(mg L ⁻¹)	(mg L ⁻¹)	(µG L ⁻¹)	(ppm)	(ppm)
MÉDIA	66,66826	458,6544	525,3226	32,51748	7,456356	6,043125
MÍNIMO	3,743	9	12,743	0,796631	0,736	1,037
MÁXIMO	130,69	1007,06	1137,75	87,8607	18,53	11,38

É possível analisar através da Tabela 1 os altos valores de sólidos inorgânicos em suspensão nesta data utilizada (Média de 458 mg L⁻¹). Estes valores são muito mais altos dos que os encontrados em outras épocas na mesma região (Setembro/2003 TSI = 49 mg L⁻¹; Fevereiro/2004 TSI = 94,6 mg L⁻¹; Junho/2004 TSI = 8,8 mg L⁻¹) (BARBOSA, 2005). Segundo Novo et al. (2006), isto acontece principalmente devido ao baixo nível das águas nesta época do ano, o que favorece a influência dos ventos na ressuspensão dos sedimentos inorgânicos que estão depositados no fundo do corpo d'água.

Imagens de satélite

Para este trabalho, foi selecionada uma cena do satélite Landsat 5, sensor TM, órbitaponto 228/061, datada de 24 de novembro de 2003. As imagens foram obtidas gratuitamente através do portal Earth Explorer (USGS – <u>https://earthexplorer.usgs.gov/</u>). Foi realizada a correção atmosférica através do modelo ACOLITE (VANHELLEMONT E RUDDICK, 2014) para a obtenção dos dados em refletância de superfície para a área de estudo. Os valores dos pixels da cena, para cada banda (1 a 4) e para cada ponto amostral de TSI, foram extraídos através da ferramenta *extract*, disponível no pacote Raster, do *software* estatístico R (R CORE TEAM, 2017). Foram excluídos *outliers* presentes nos dados, bem como pontos de nuvem que existiam na imagem. Assim, restaram 61 pontos para a análise final.

Relação Refletância x TSI – Modelo Empírico

Para a obtenção do modelo empírico, a relação entre os valores de refletância e as concentrações de sedimentos inorgânicos em suspensão foram avaliadas. Inicialmente, equações lineares foram testadas e observada a relação entre as variáveis. Foi também aplicada uma transformação logarítimica na variável TSI, já que esta apresenta este tipo de comportamento quando concentrações muito altas de sólidos inorgânicos em suspensão estão presentes (KNAPES et al., 2015).

Análise Dos Resíduos

Após a geração dos modelos com a aplicação do logarítimo na base 10 para o TSI, os resíduos destes modelos foram testados. Avaliou-se a distribuição espacial dos resíduos, para observar se existiam padrões espaciais e também foi aplicado um teste de Moran para os resíduos, com hipótese h0 de zero autocorrelação espacial e também foi gerado um semivariograma para avaliar a estrutura de dependência espacial destes resíduos

Regression Kriging e Kriging With External Drift

Foi realizada uma análise de relação entre a KED e a RK, pois, segundo Hengl (2007; 2009) estes métodos são matematicamente equivalentes¹. O que é mostrado por Hengl (2007) é confirmado através da Figura 2, que apresenta um gráfico de dispersão entre os valores preditos pela RK e pela KED dentro da área de água do lago Curuai. Para esta validação, foram utilizados todos os 61 pontos de amostragem. A realização da RK e da KED foi feita em ambiente R, utilizando o pacote gstat (PEBESMA, 2004; GRÂLER, PEBESMA e HEUVELINK, 2016), um dos principais pacotes disponíveis em ambiente R para análise espacial e krigagem (BIVAND, 2008). A metodologia proposta por Hengl (2015) foi utilizada para o processamento da RK e a metodologia proposta por ROSSITER (2017) foi utilizada para o processamento da KED. Os resultados demonstram alta concordância entre os métodos, corroborando com o postulado por Hengl (2007). Um R² de 0,999 foi obtido para a relação entre as duas metodologias. Nesta análise, percebeu-se uma demanda computacional muito alta para a RK, com um tempo de processamento de cerca de 20 minutos. Diferentemente, a KED apresentou tempo de processamento bem mais rápido (~2 minutos), assim, já que estas apresentaram esta concordância, nos próximos processamentos de validação apenas a KED foi utilizada, devido a menor demanda computacional e tempo para o processamento.

¹ Em Hengl (2007), Apêndice 1 há a demonstração matemática provando a igualdade entre os métodos Regression Kriging e Kriging With External Drift.

Figura 1 Gráfico de dispersão com densidade de pontos. Count = Número de pontos encontrados para determinados valores de x e y. TSI - RK é concentação de sedimentos inorgânicos estimados pela Regression Kriging e TSI – KED é a concentração de sedimentos inorgâni



Validação dos Modelos

Como o objetivo deste trabalho é avaliar se o método de krigagem que considera a dependência espacial do resíduo apresenta melhores resultados para a variável predita de TSI, foi realizada uma simulação de Monte Carlo, com 200 iterações. Em cada iteração, dez pontos aleatórios dentro dos 61 pontos amostrais são retirados. Os 51 pontos restantes são utilizados para gerar a krigeagem e também para gerar um modelo linear. Foi utilizada como variável dependente em todas as análises o logarítimo na base 10 dos sólidos inorgânicos em suspensão. Esta transformação é realizada em diversos trabalhos (KRIVORUCHKO, 2001; SILA et al., 2017; HU, 2017, XU, 2018) e foi realizada neste devido ao aumento da correlação entre as variáveis e também do próprio comportamento das mesmas. Para a avaliação dos modelos, os valores de coeficiente de determinação (R²), Erro Médio Quadrático (RMSE) e o Erro Médio Absoluto em Percentagem (MAPE - Equação 7)

$$MAPE = 100 * \frac{1}{n} * \sum_{i=1}^{n} \frac{|x_i - y_i|}{x_i}$$
(7)

Em que, n é o número de amostras, x_i é a concentração de TSI medida em campo, para a amostra **i**, e y_i é o valor da amostra estimado pelo modelo. Os códigos para o cálculo do MAPE e RMSE foram implementados em R.

Mapas Finais

Após a validação, foram gerados dois mapas da Planície de Inundação do Lago Grande do Curuai, para avaliar os resultados. Os arquivos no formato .tiff foram gerados em ambiente R e posteriormente exportados para o *software* ArcGis 10.2, devido a limitações do R em relação a composição de mapas.

Resultados

Análise de Regressão e resíduos

Inicialmente, foi avaliada a relação entre o TSI e as bandas espectrais do sensor TM, Landsat 5 (Figura 2). A partir da análise da Figura 2, vê se que as faixas do visível (banda 1, 2 e 3) do sensor TM não apresentaram altos valores de correlação. Isto é esperado, visto os altos valores de concentração de sólidos em suspensão saturam os valores de refletância da água nestas faixas espectrais (ARST, 2002; KNAEPS et al., 2015; GIARDINO et al., 2017). A melhor correlação obtida foi para a Banda 4 (760 – 900 nm). Esta relação também corrobora com a literatura, já que altas concentrações de sedimentos em suspensão elevam a refletância da água nas faixas do infravermelho próximo, faixa em que o único componente opticamente ativo influenciador são os sólidos em suspensão (KNAEPS et al., 2015). Figura 2 Análise de correlação entre as faixas espectrais do sensor TM e o material inorgânico em suspensão e do logarítimo na base 10 do material inorgânico em suspensão. MISusp = Material Inorgânico em Suspensão, LOG10_TSI = Logarítimo na base 10 do TSI; b1, b2, b3, b4 são as bandas do sensor TM, Landsat 5. Cores mais próximas do preto indicam alta correlação positiva.



Assim, aplicou-se a regressão linear nas imagens tendo como variável dependente o logartímo do TSI e independente a banda 4 do sensor TM. A regressão linear utilizada é apresentada na Equação 8 (n = 61, $R^2 = 0.456$).

$$Log10(TSI) = 1,185 + 10,960 * b4$$
(8)

Segundo Hengl (2007), a aplicação da krigagem nos resíduos é efetiva apenas se o R² do modelo empírico apresenta valores de R² próximos de 0,5. Isto se dá porque, caso o valor de R² for muito alto, os resíduos tendem a zero, e se o valor de R² for muito baixo, não faz sentido a utilização do modelo empírico e a krigagem universal é mais indicada.

Como o valor de R² está na faixa indicada por Hengl (2007), a Equação 8 foi aplicada os resíduos foram analisados (Figura 3). É possível identificar padrões na distribuição dos resíduos da regressão linear no Lago Grande de Curuai. O teste do índice

de Moran para autocorrelação espacial dos resíduos, com hipótese h_0 de zero correlação espacial foi aplicado através da função Moran.I do pacote APE (PARADIS, CLAUDE e STRIMMER, 2004) do *software* estatístico R, com p-valor de 2,3 * 10⁻¹³, não havendo razões para discordar que à 5% de significância exista a correlação espacial.

Figura 3 Mapa dos resíduos para a planície do Lago Grande do Curuai. As cores representam os valores dos resíduos da regressão linear aplicada (LogTSI x B4) no Lago Grande do Curuai



Com a autocorrelação espacial atestada através do teste do Índice de Moran, prosseguiu-se para a simulação de Monte Carlo para avaliar o comportamento da regressão linear sem considerar a dependência espacial dos resíduos e da KED. Na Tabela 2 os resultados da validação são apresentados. Tabela 3 Resultado da validação pelo método de Monte Carlo para a KED e para a Regressão Simples. n = 2000.

TÉCNICA	D 2	RMSE	MAPE	
IECNICA	R-	$(mg L^{\cdot 1})$	(%)	
KED	0,57	134,8	27,7	
REGRESSÃO	0,25	184,4	33,35	

Os resultados da Tabela 2 mostram que houve uma melhora em todos os métodos de avaliação dos modelos para os resultados dos dados preditos pela KED comparados com a Regressão simples. Os resultados obtidos de MAPE foram um pouco mais altos do que os obtidos por Knaeps et al. (2015) (MAPE = 25%), que avaliaram a relação entre a faixa em 1020 nm e a concentração de material em suspensão total, que variaram em um *range* de $0 - 1000 \text{ mg L}^{-1}$. Vale lembrar que a faixa em 1020 é mais indicada para a relação dos sólidos em suspensão em altas concentrações, já que não é saturada por este aumento, conforme evidenciado no estudo de Knaeps et al. (2015). Na Figura 4 é possível avaliar a maior aproximação da reta 1:1 para os valores de TSI estimados pela KED ao se comparar com a regressão linear simples.

Figura 4 TSI Estimado pela KED e pela Regressão (REG) comparado com os dados obtidos em campo. Linha em preto representa a reta 1:1



Como última etapa, o mapa da KED e da Regressão Simples foi gerado (Figuras 5 e 6). Ao aplicar a KED, percebe-se que principalmente na parte noroeste houve uma maior aproximação dos valores reais existentes pela KED, que subestimou os valores neste ponto. Também, na parte sudeste, houve uma superestimação pela regressão simples e, ao considerar os resíduos, os valores nesta área ficam mais próximos dos valores reais.

Figura 5 Aplicação do modelo empírico (Equação 8) para a imagem Landsat 5 TM. Os pontos coloridos representam os valores de concentração nos pontos amostrais.



Figura 6 Aplicação da KED para a imagem Landsat 5 TM. Os pontos coloridos representam os valores de concentração nos pontos amostrais.



Conclusões

Ao considerar a dependência espacial dos resíduos, obteve-se melhores resultados, com os dados avaliados, ao se comparar com uma regressão linear simples. Claramente, deve-se testar outras épocas, já que o regime hidrológico na região é extremamente dependente da época do ano avaliada. Ainda assim, estes resultados evidenciam que, ao se considerar a interpolação dos resíduos por krigeagem, uma melhoria nos resultados da predição do TSI por sensoriamento remoto foi evidenciada. Mais estudos devem ser feitos avaliando outras épocas no Lago Grande do Curuai (e.g. épocas de cheia) e outras regiões (e.g. zonas costeiras, oceano). Ainda, pode-se recomendar o uso de séries temporais ligadas a modelagem espacial (e.g. *Geographic and Temporal Weighted Regression* (e.g. CHU et al. (2018)), *Time-Regression-Kriging)* para avaliar a dependência espaço-tempo da variação dos constituintes opticamente ativos em águas interiores, costais e oceânicas.

Referências

ARST, H. **Optical properties and remote sensing of multicomponental water bodies.** Springer Science & Business Media, 2003.

ARRAUT, E.M. et al. Estudo do comportamento espectral da clorofila e dos sólidos em suspensão nas águas do lago grande de Curuai (Pará), na época da seca, através de técnicas de espectroscopia de campo. **Simpósio Brasileiro de Sensoriamento Remoto**, v. 12, p. 2447-2456, 2005.

APHA - AMERICAN PUBLIC HEALTH ASSOCIATION et al. Standard methods for the examination of water and wastewater. American Public Health Association., 1999

BARBOSA, C. C. F. Sensoriamento Remoto da dinâmica da circulaçnao da água do sistema planície de Curuai/ Rio Amazonas. n. September, p. 255, 2005.

GRÄLER, B, PEBESMA, E. AND HEUVELINK, H,. Spatio-Temporal Interpolation using gstat. **The R Journal** 8(1), 204-218, 2016.

BILOTTA, G.S., BRAZIER, R.E.. Understanding the influence of suspended solids on water quality and aquatic biota. **Water Research**, [s.l.], v. 42, n. 12, p.2849-2861, jun. 2008. Elsevier BV. <u>http://dx.doi.org/10.1016/j.watres.2008.03.018</u>.

BOSS, E. et al. Comparison of inherent optical properties as a surrogate for particulate matter concentration in coastal waters. Limnology and Oceanography: Methods, v. 7, n. 11, p. 803-810, 2009.

BRANIGAN, J. Development of a Field Test for Total Suspended Solids Analysis. 2013.

CARVALHO, L. A. A. et al. Implications of scatter corrections for absorption measurements on optical closure of Amazon floodplain lakes using the Spectral Absorption and Attenuation Meter (AC-S-WETLabs). **Remote Sensing of Environment,** v. 157, p. 123-137, 2015.

CHAPMAN, P. M., HAYWARD, A., FAITHFUL, J. Total Suspended Solids Effects on Freshwater Lake Biota Other than Fish. **Bulletin Of Environmental Contamination And Toxicology**, [s.l.], v. 99, n. 4, p.423-427, 17 ago. 2017. Springer Nature. http://dx.doi.org/10.1007/s00128-017-2154-y. CHU, H.-J.; KONG, S.-J.; CHANG, C.-H. Spatio-temporal water quality mapping from satellite images using geographically and temporally weighted regression. **International Journal of Applied Earth Observation and Geoinformation**, v. 65,, p. 1–11, 2018.

CORDONE, A. J., KELLEY, D. W. The influences of inorganic sediment on the aquatic life of streams. California Department of Fish and Game, 1961.

DAWSON, E.J., MACKLIN, M.G. Speciation of Heavy Metals in Floodplain and Flood Sediments: a Reconnaissance Survey of the Aire Valley, West Yorkshire, Great Britain. **Environmental Geochemistry And Health,** [s.l.], v. 20, n. 2, p.67-76, 1998. Springer Nature. <u>http://dx.doi.org/10.1023/a:1006541724394</u>.

DONOHUE, I; GARCIA MOLINOS, J. Impacts of increased sediment loads on the ecology of lakes. **Biological Reviews**, v. 4, n. 4, p. 517-531, 2009.

DOGLIOTTI, A. I. et al. A single algorithm to retrieve turbidity from remotely-sensed data in all coastal and estuarine waters. **Remote Sensing of Environment**, v. 156, p. 157–168, 2015.

GIARDINO, C. et al. Bio-optical Modeling of Total Suspended Solids. In: **Bio-optical Modeling and Remote Sensing of Inland Waters.** 2017. p. 129-156.

GOODIN, D. G. et al. Analysis of suspended solids in water using remotely sensed high resolution derivative spectra. **Photogrammetric Engineering & Remote Sensing**, v. 59, n. 4, p. 505-510, 1993.

GOOVAERTS, P.. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York, p. 496, 1977.

HARROD, T. R.; THEURER, F. D. 7 Sediment. Agriculture, hydrology and water quality, p. 155, 2002.

HEBERLING, M. T. et al. Comparing drinking water treatment costs to source water protection costs using time series analysis. **Water Resources Research**, v. 51, n. 11, p. 8741-8756, 2015.

HENGL, T. A practical guide to geostatistical mapping. 2009

HENGL, T; HEUVELINK, G. B. M; ROSSITER, D. G. About regression-kriging: from equations to case studies. **Computers & geosciences**, v. 33, n. 10, p. 1301-1315, 2007.

HENGL, T et al. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. **PloS one**, v. 10, n. 6, p. e0125814, 2015.

HU, Dangui et al. Spatiotemporal regression Kriging to predict precipitation using timeseries MODIS data. **Cluster Computing**, v. 20, n. 1, p. 347-357, 2017.

KIRK, J. T. O. Effects of suspensoids (turbidity) on penetration of solar radiation in aquatic ecosystems. **In**: Perspectives in Southern Hemisphere Limnology. Springer Netherlands, 1985. p. 195-208.

_____. Light and photosynthesis in aquatic ecosystems. Cambridge university press, 2011.

KIRK, K. L. Effects of suspended clay on Daphnia body growth and fitness. **Freshwater Biology**, v. 28, n. 1, p. 103-109, 1992.

KNAEPS, E. et al. A SWIR based algorithm to retrieve total suspended matter in extremely turbid waters. **Remote Sensing of Environment**, v. 168, p. 66-79, 2015.

KOTEEN, J; ALEXANDER, S. J.; LOOMIS, J. B. Evaluating benefits and costs of changes in water quality. 2002.

KRIGE, D G. A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Southern African Institute of Mining and Metallurgy, v. 52, n. 6, p. 119-139, 1951.

KRIVORUCHKO, K. Using linear and non-linear kriging interpolators to produce probability maps. Disponível em: http://www. esri. com/software/arcgis/arcgisxtensions/geostatistical/research_papers. html, 2001.

KRONVANG, B. et al. Pesticides and heavy metals in Danish streambed sediment. **Hydrobiologia**, [s.l.], v. 494, n. 1-3, p.93-101, mar. 2003. Springer Nature. http://dx.doi.org/10.1023/a:1025441610434.

LOBO, F. L.; COSTA, M. P. F; NOVO, E. M. L. M. Time-series analysis of Landsat-MSS/TM/OLI images over Amazonian waters impacted by gold mining activities. **Remote Sensing of Environment**, v. 157, p. 170-184, 2015. MATHERON, G. Le krigeage universel. 1969.

MONTANHER, O. C. Modelos empíricos para estimativa da concentração de sedimentos em suspensão em rios amazônicos de águas brancas a partir de imagens Landsat 5. 2013. 115 p. 2013. Dissertação (Mestrado em Sensoriamento Remoto). Instituto Nacional de Pesquisas Espaciais, São José dos Campos.

MUNDAY JR, John C.; ALFÖLDI, Thomas T. Landsat test of diffuse reflectance models for aquatic suspended solids measurement. **Remote sensing of environment**, v. 8, n. 2, p. 169-183, 1979.

NASRABADI, T. et al. Using total suspended solids (TSS) and turbidity as proxies for evaluation of metal transport in river water. **Applied Geochemistry**, [s.l.], v. 68, p.1-9, maio 2016. Elsevier BV. <u>http://dx.doi.org/10.1016/j.apgeochem.2016.03.003</u>.

NECHAD, B.; RUDDICK, K. G.; PARK, Y. Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. **Remote Sensing of Environment**, v. 114, n. 4, p. 854-866, 2010

NOVO, E. M. M.; HANSOM, J. D.; CURRAN, P. J. The effect of sediment type on the relationship between reflectance and suspended sediment concentration. **Remote Sensing**, v. 10, n. 7, p. 1283-1289, 1989.

NOVO, E. M. L. M; STEFFEN, C. A.; BRAGA, C. Z. F. Results of a laboratory experiment relating spectral reflectance to total suspended solids. **Remote Sensing of Environment,** v. 36, n. 1, p. 67-72, 1991.

NOVO, E. M. L. M. et al. Seasonal changes in chlorophyll distributions in Amazon floodplain lakes derived from MODIS images. **Limnology**, v. 7, n. 3, p. 153-161, 2006.

PARADIS, E., CLAUDE. J., STRIMMER, K. 2004. APE: analyses of phylogenetics and evolution in R language. **Bioinformatics** 20: 289-290.

PEBESMA, E. J. Multivariable geostatistics in S: the gstat package. Computers &Geosciences, 30: 683-691, 2004.

PIMENTEL, D et al. Environmental and economic costs of soil erosion and conservation benefits. **Science-AAAS-Weekly Paper Edition**, v. 267, n. 5201, p. 1117-1122, 1995.

RAVEN, P. H.; EVERT, R. F.; EICHHORN, S. E. **Biologia vegetal**. In: **Biologia vegetal**. Guanabara, 2007.

RIZO-DECELIS, L. D.; PARDO-IGÚZQUIZA, E.; ANDREO, B. Spatial prediction of water quality variables along a main river channel, in presence of pollution hotspots. **Science of the Total Environment**, v. 605–606, p. 276–290, 2017.

ROSSITER, D. G. An introduction to geostatistics with R/gstat. Cornell University. Retrieved May, v. 18, p. 2016, 2016.

SABESP. Norma Técnica Interna SABESP NTS 013. 1999. Disponível em: http://www2.sabesp.com.br/normas/nts/nts013.pdf>. Acesso em: 20 dez. 2017.

SILA, A; POKHARIYAL, G; SHEPHERD, K. Evaluating regression-kriging for midinfrared spectroscopy prediction of soil properties in western Kenya. Geoderma Regional, v. 10, p. 39-47, 2017.

R CORE TEAM (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

ROTTA, L. H. S et al. Analyzing the status of submerged aquatic vegetation using novel optical parameters. **International Journal of Remote Sensing**, v. 37, n. 16, p. 3786-3810, 2016.

STEINMAN, A. D., MCINTIRE, C. D. Recovery of lotic periphyton communities after disturbance. **Environmental Management**, v. 14, n. 5, p. 589-604, 1990.

TUNDISI, J. G.; TUNDISI, T. M. Impactos potenciais das alterações do Código Florestal nos recursos hídricos. **Biota Neotropica**, v. 10, n. 4, 2010.

VANHELLEMONT, Q; RUDDICK, K. Turbid wakes associated with offshore wind turbines observed with Landsat 8. **Remote Sensing of Environment**, v. 145, p. 105-115, 2014.

Von SPERLING, Marcos. **Introdução à qualidade das águas e ao tratamento de esgotos:** Princípios do Tratamento biológico de águas resíduarias. Belo Horizonte: Departamento de Engenharia Sanitária e Ambiental, 1996. 243 p.

WACKERNAGEL, H. **Multivariate geostatistics: an introduction with applications**, 2nd Edition. Springer-Verlag, p. 381, 2003.

WEBSTER, R.; OLIVER, M. Geostatistics for Environmental Scientists. 2007. Statistics in Practice, Wiley, Chichester.

WU, J et al. Nitrogen and phosphorus associating with different size suspended solids in roof and road runoff in Beijing, China. **Environmental Science And Pollution Research**, [s.l.], v. 22, n. 20, p.15788-15795, 4 jun. 2015. Springer Nature. http://dx.doi.org/10.1007/s11356-015-4743-9.

XU, Y; SMITH, S,E; AMR, S. G.; SUHAS P. A.; WANI, V. D. Nair, Estimating soil total nitrogen in smallholder farm settings using remote sensing spectral indices and regression kriging, **CATENA**, Volume 163, Pages 111-122, ISSN 0341-8162, 2018. https://doi.org/10.1016/j.catena.2017.12.011.

APÊNDICE 1

O Scritp utilizado neste trabalho está disponível no link abaixo, hospedado no OneDrive.

https://1drv.ms/u/s!AsplDRqiCbwigYklusQ1S5jpqP-7Ug