



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

Análise do modelo de regressão GWR em águas túrbidas para dados Landsat 5 - TM

Rogério Flores Júnior¹

¹Instituto Nacional de Pesquisas Espaciais

INPE

21/12/2017

Resumo

Os meios aquáticos são compostos por uma mistura de materiais suspensos e dissolvidos. Estes materiais interagem com a energia incidente de forma a absorver ou espalhar, controlando assim a disponibilidade desta dentro do corpo d'água. Portanto, os meios são espectralmente dependentes dos materiais que os compõem. É de grande interesse a utilização de imagens orbitais para determinação destes materiais, a fim de estimar os parâmetros biofísicos da água. A determinação destes parâmetros é dada através de algoritmos que propõem a derivação das correlações, a partir de modelos, das propriedades inerentes do meio com dados radiométrico dos sensores orbitais. Estes algoritmos utilizam-se de modelos de regressão, sendo o ajuste linear a forma mais comum. A aplicação de modelos de regressão espacial em variáveis contínuas podem apresentar estimativas mais realistas que os modelos lineares. A regressão geográfica ponderada (GWR) foi aplicada neste trabalho em comparação com o método linear simples. Os resultados mostraram que a GWR foi capaz de estimativas melhores, devido a adição da componente espacial. A análise dos coeficientes *betas* locais gerados pela GWR mostrou uma boa concordância com Dinâmica espacial do lago na época de seca.

Palavras-chave: Análise espacial, Modelos de regressão, GWR, lago Grande Curuai, óptica hidrológica.

1 Introdução

Os sistemas aquáticos de água doce são considerados os mais vulneráveis a atividades humanas (VÖRÖSMARTY et al., 2000). Com o avanço dos plantios e urbanização em torno de barragens, o descarte de resíduos e a lixiviação têm aumentado os problemas da deposição de sedimentos e eutrofização deste corpos d'água. A utilização de imagens de satélite para determinação de parâmetros de regiões aquáticas vêm crescendo, principalmente com o desenvolvimento de sensores com uma melhor capacidade de detecção em alvos com baixas respostas como água (PALMER; KUTSER; HUNTER, 2015). Segundo Gordon (1989) podemos determinar a presença dos componentes opticamente ativos (COA) através das medidas de reflectância adquiridas através de imagens de sensores orbitais. A utilização de correções atmosféricas é essencial para a determinação de parâmetros dos componentes presentes nos corpos d'água, visto que a qualidade da reflectância de superfície das imagens é diretamente dependente do sensor, método de correção e as características da atmosfera-superfície (OKIN; GU, 2015).

A derivação de parâmetros de qualidade água a partir de sensoriamento remoto é realizada através de modelos empíricos e semi-empíricos, que propõem uma relação estatística entre a variável dependente e um conjunto de variáveis dependentes. As técnicas de geoestatística, são frequentemente utilizadas para estimação de parâmetros onde os dados de campo não são suficientemente disponíveis para o completo entendimento da variável. O método mais utilizado para aplicação em modelos empíricos é o de regressão (MONTANHER, 2013). Este tipo de análise assume uma relação linear entre as variáveis dependentes e independentes, sendo descrita pelo grupo dos Modelos Lineares Generalizados (GLM). A regressão linear simples (RLS) leva em consideração apenas uma variável independente.

A utilização de modelos de regressão espacial, como a regressão ponderada geográfica (GWR), pode ser útil para a avaliação de correlação de duas ou mais variáveis que possuem suas variações correlacionadas com o espaço em que estão inseridas. Dessa forma, pode-se abordar questões de auto-correlação espacial que são encontradas em um conjunto de dados, além de poder investigar a distribuição espacial dos parâmetros analisados e o significado dos parâmetros de previsão, como os coeficientes beta. Segundo Nakaya (2014) a modelagem por GWR apresenta melhor desempenho em relação aos modelos tradicionais de regressão e permite uma leitura diferenciada de seus coeficientes, contribuindo para agregação de novas interpretações do fenômeno analisado.

Portanto, este trabalho teve como objetivo comparar as estimativas de parâmetros de qualidade da água através da utilização da regressão geograficamente ponderada (GWR) em contraste com a regressão linear simples em dados Landsat 5 TM.

2 Materiais e métodos

2.1 Área de estudo

A planície do Lago Grande de Curuai, localizada a 900 km da foz do rio Amazonas (Figura 1), tem uma área de aproximadamente 3500 Km² e uma flutuação de até 7 metros no nível da água (BARBOSA, 2005). Esta planície é formada por cerca de 20 lagos que estão interligados por canais de comunicação que se mantêm durante todo o ciclo hidrológico e é caracterizada como planície fluvial inundável sujeita ao regime natural de águas do rio Amazonas, pois, durante o período de nível máximo das águas do rio Amazonas, as mesmas extravasam para a várzea aumentando o volume do Lago Curuai (BARBOSA, 2005).



Figura 1: Planície do lago Grande Curuai.

A planície recebe, em diferentes proporções, águas brancas, claras e pretas. As águas pretas originam-se de florestas que fazem fronteira com a margem sul do lago, as águas claras de pequenos rios da margem sul, e as águas brancas, que possuem a maior proporção, vem do rio Amazonas e apresentam uma grande quantidade de sólidos em suspensão (BARBOSA, 2005).

2.2 Dados

Os dados orbitais utilizados neste trabalho foram obtidos através da plataforma Earth explorer da USGS. Foi selecionada uma cena Landsat 5 do sensor TM para a data de 24 de novembro de 2003, devido a sua proximidade temporal com os dados em situ e sua baixa cobertura de nuvens. Para utilização dos dados radiométricos foi utilizado do método de correção atmosférica ACOLITE. O ACOLITE é um método de correção alternativo, que se baseia nas características da imagem para remoção dos efeitos da atmosfera sobre os dados. Os dados radiométricos de cada banda utilizada foram extraídos para a mesma localização dos pontos de coleta in situ utilizando as ferramentas de extração por pontos do Software Arcgis 10.2.

Os dados de campo foram coletados em uma campanha de campo durante o mês de novembro de 2003, para medir as concentrações de componentes opticamente ativos nas águas da planície de inundação do Curuai por Barbosa (2005). Foram coletadas amostras de água para análise laboratorial de clorofila-a e sedimentos totais em suspensão de aproximadamente 70 pontos. As amostras foram coletadas integrando a coluna de água da superfície até a profundidade de Secchi. O TSS foi determinado com base na metodologia de *Wetzel e Likens*, e as análises de clorofila-a foram baseadas em *Nush* (BARBOSA, 2005). Foi feita uma remoção de pontos que estavam sob a cobertura de nuvens na área de estudo, e também pontos em que sua localização não correspondiam a massa de água presente nos dados orbitais. A localização dos pontos de coleta dos dados está apresentado na Figura 2.

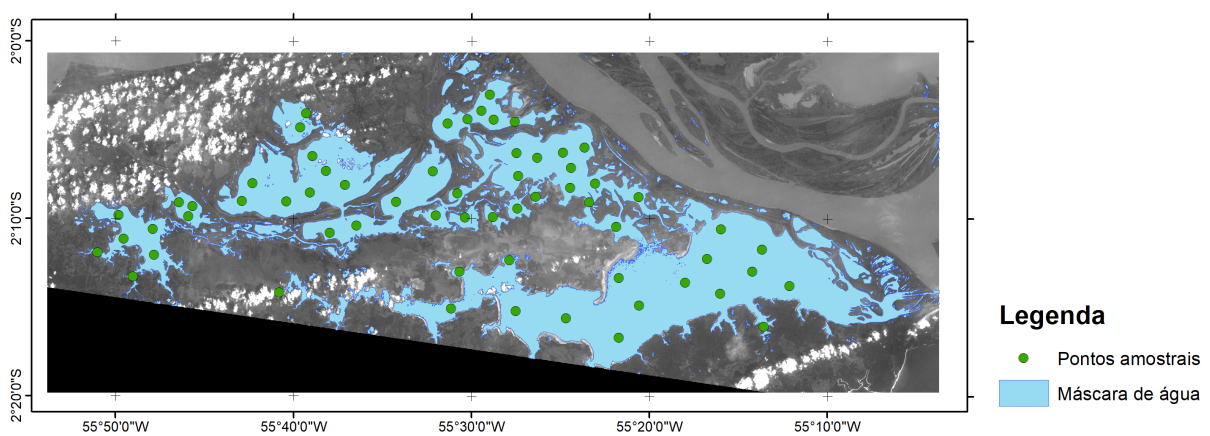


Figura 2: Mapa de localização dos pontos de coleta.

2.3 Análise Exploratória

Os parâmetros disponíveis na série de dados utilizadas neste trabalho são Total de Sólidos Orgânicos em Suspensão; Total de Sólidos Inorgânicos em Suspensão; Total de Sólidos em Suspensão; Carbono Inorgânico Dissolvido e Carbono Orgânico Dissolvido. A Tabela 1 apresenta a estatística descritiva para os parâmetros disponíveis nos dados de campo, com seus valores mínimos, máximos e sua média. Todas as análises exploratórias foram realizadas utilizando as funções estatísticas do Software Rstudio.

Tabela 1: Estatística básica dos parâmetros disponíveis.

	Clorofila ($\mu\text{g/L}$)	TSI (mg/L)	TSO (mg/L)	TSS (mg/L)	DIC (ppm)	DOC (ppm)
Mínimo	0,7966	9,0000	3,743	12,7430	0,736	1,037
Média	32,5175	458,6544	66,6683	525,3226	7,456356	6,043125
Máximo	87,8607	1007,06	130,69	1137,7500	18,53	11,38

Os valores extremamente altos dos sólidos inorgânicos totais na Tabela 1 são dado pelo baixo nível da água neste período do ano, que em conjunto com o vento promove a ressuspensão destes materiais inorgânicos. Por tanto, esta é uma variável de grande interesse para este estudo, e juntamente com a clorofila, que possui grande importância na determinação da produtividade primária e dos níveis de eutrofização do corpo d'água, foram selecionadas para serem aplicadas aos modelos de regressão.

2.3.1 Boxplot

As variáveis foram inicialmente analisadas por um diagrama de dispersão para a escolha das que apresentavam maior correlação com os parâmetros disponíveis. Para verificação de dados espúrios foram gerados gráficos de boxplot para os parâmetros selecionados. O diagrama de caixa (BoxPlot) é um método gráfico tipicamente representado por quartis e inter quartis que auxilia na definição dos limites superior e inferior da série de dados analisados. Valores extremos nos dados podem ser identificados através do gráfico de boxplot, podendo ser classificados como valores espúrios, e removidos das análises posteriores (GUIMARÃES, 2008).

2.3.2 Teste de normalidade

O primeiro método para verificação do formato da distribuição de uma variável contínua é a construção do histograma. O histograma é um gráfico de barras justapostas em que no eixo horizontal está a variável de interesse dividida em classes e no eixo vertical a frequência da classe correspondente. O método gráfico do histograma têm a desvantagem de ser subjetivo, pois depende de interpretação visual. Para um resultado mais objetivo, pode-se usar testes paramétricos de aderência à distribuição Normal. Portanto para este trabalho foi utilizado o teste estatístico de Shapiro-Wilk para se testar a normalidade dos parâmetros.

2.4 Análise Espacial

2.4.1 Regressão Linear Simples (RLS)

A regressão abrange uma ampla gama de métodos para modelar a relação entre uma variável dependente e um conjunto de uma ou mais variáveis independentes. A variável dependente pode ser conhecida como a variável y , resposta ou de regressão. As variáveis independentes são algumas vezes conhecidas como variáveis x , preditoras ou regressora (CHARLTON; FOTHERINGHAM; BRUNSDON, 2009). O objetivo da análise de regressão linear global é quantificar a relação linear entre a variável dependente e uma ou mais variáveis independentes, expressa pela Equação 1

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

Onde Y_i é a i -ésima observação da variável dependente, é X_i a i -ésima observação da variável dependente, os parâmetros β_0 e β_1 são, respectivamente, o intercepto e a inclinação da reta, e a constante ε representa os erros, ou resíduos da estimativa do modelo (NETER et al., 1996). No modelo de regressão linear simples assume-se que as observações não são correlacionadas, e que os resíduos do modelo também são independentes e não correlacionados com a variável dependente, além de apresentar variância constante e distribuição normal com média zero.

2.4.2 Teste de Moran Unidimensional

Após a aplicação da regressão linear realiza-se o teste I de Moran nos resíduos desta regressão para verificação da dependência espacial. A estatística I de Moran é simplesmente o valor do coeficiente de inclinação de uma reta de regressão para o diagrama apresentado. Considerando uma reta de regressão no diagrama de Moran, quanto maior a aderência dessa reta, maior a evidência de correlação espacial. Quanto mais próximo de +1, mais forte a correlação espacial positiva e quanto mais próximo de -1, mais forte a correlação espacial negativa (MARQUES et al., 2010). Foi utilizada neste trabalho a métrica dos K-vizinhos mais próximos para a aplicação do diagrama de Moran unidimensional, sendo a mesma forma em que o kernel da GWR é calculado. Para a criação dos diagramas de Moran foi utilizado o Software GeoDa, selecionando-se a ponderação de vizinhos mais próximos no valor de três.

2.4.3 Regressão geograficamente Ponderada (GWR)

A regressão geograficamente ponderada (Geographically weighted regression–GWR) ajusta um modelo de regressão linear a cada ponto observado, ponderando as demais em funções da distância a esse ponto. O resultado deste modelo é um conjunto de parâmetros para cada ponto, onde observações próximas para o ponto de ajuste da regressão local possuem pesos mais elevados do que observações mais afastadas. Este procedimento é realizado para cada

ponto em relação a todas as observações (DRUCK, 2004). Este tipo de regressão é expressa matematicamente pela Equação 2

$$y_1(\mathbf{u}) = \beta_{0i}(\mathbf{u}) + \beta_{1i}(\mathbf{u}).x_{1i} + \beta_{2i}(\mathbf{u}).x_{2i} + \dots + \beta_{mi}(\mathbf{u}).x_{mi} \quad (2)$$

A matriz de peso no GWR representa a importância diferente de cada observação individual no conjunto de dados usado para estimar os parâmetros no local. Em geral, quanto mais se aproxima uma observação, maior o peso. Assim, cada estimativa de ponto tem uma matriz de peso única. Na essência, existem dois regimes de ponderação que podem ser usados: kernel fixo e kernel adaptativo. Para o kernel fixo, a distância é constante, mas o número de vizinhos mais próximos varia. Para o kernel adaptativo, a distância varia, mas o número de vizinhos permanece constante (CHARLTON; FOTHERINGHAM; BRUNSDON, 2009).

$$w_i(\mathbf{u}) = e^{-0,5[\frac{d_i(\mathbf{u})}{h}]^2} \quad (3)$$

Para a ponderação das observações foi utilizada a função gaussiana com largura de banda adaptativa (Equação 3), onde $w_i(\mathbf{u})$ é peso definido para as observações na localização \mathbf{u} , $d_i(\mathbf{u})$ é a distância euclidiana de \mathbf{u} em relação aos dados, h é definido com "Bandwidth" ou largura de banda. O algoritmo utiliza-se da menor distância que abrange os vizinhos mais próximos através da seleção da largura de banda de forma dinâmica, com base no valor do Akaike Information Criterion –AICc (CHARLTON; FOTHERINGHAM; BRUNSDON, 2009).

A espacialização dos valores estimados pela regressão GWR foi realizada utilizando o método de interpolação pelo inverso da distância ponderada (inverse distance weighted - IDW) com peso 2, ou seja, o inverso do quadrado da distância. Este método foi aplicado sobre os coeficientes da regressão de acordo com Chu, Kong e Chang (2018). Todos os passos realizados para as regressões simples (RLS) e GWR foram desenvolvidos no ambiente do software estatístico *Rstudio* (TEAM, 2017) com o auxílio do pacote "*spgwr*". Os mapas de regressão foram gerados a partir do software ArcGis 10.2 com os valores preditos para cada modelo utilizado.

Para a validação dos dados estimados pelos dois tipos de regressão em relação aos dados de campo foram inicialmente removidos 15 dos pontos amostrais, para um conjunto de validação, a fim de se evitar a influência destes pontos sobre a regressão. Os pontos restantes configuraram o conjunto para aplicação dos métodos. Por fim, os resultados dos modelos foram extrapolados para a posição dos pontos removidos e comparados com seus valores de campo. O processo foi repetido por 40 iterações, contabilizando um número de 600 pontos de validação para cada modelo. Foram extraídos destes resultados os coeficientes de determinação (R^2), o erro quadrático médio (RMSE) e a média percentual absoluta do erro (MAPE), estes últimos que são expressos respectivamente pelas Equações 4 e 5, onde x_i é o valor predito, \hat{x}_i é o valor de referência e n é

o número de observações.

$$RMSE = \sqrt{\sum \frac{(X_i - \hat{x}_i)^2}{n}} \quad (4)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{X_i - \hat{x}_i}{X_i} \right| * 100 \quad (5)$$

O fluxograma apresentado na Figura 3 demonstra de forma objetiva os passos realizados neste trabalho, desde a obtenção dos dados brutos, sua preparação, as análises exploratórias e espaciais, a aplicação dos métodos de regressão, até a criação dos mapas de regressão.

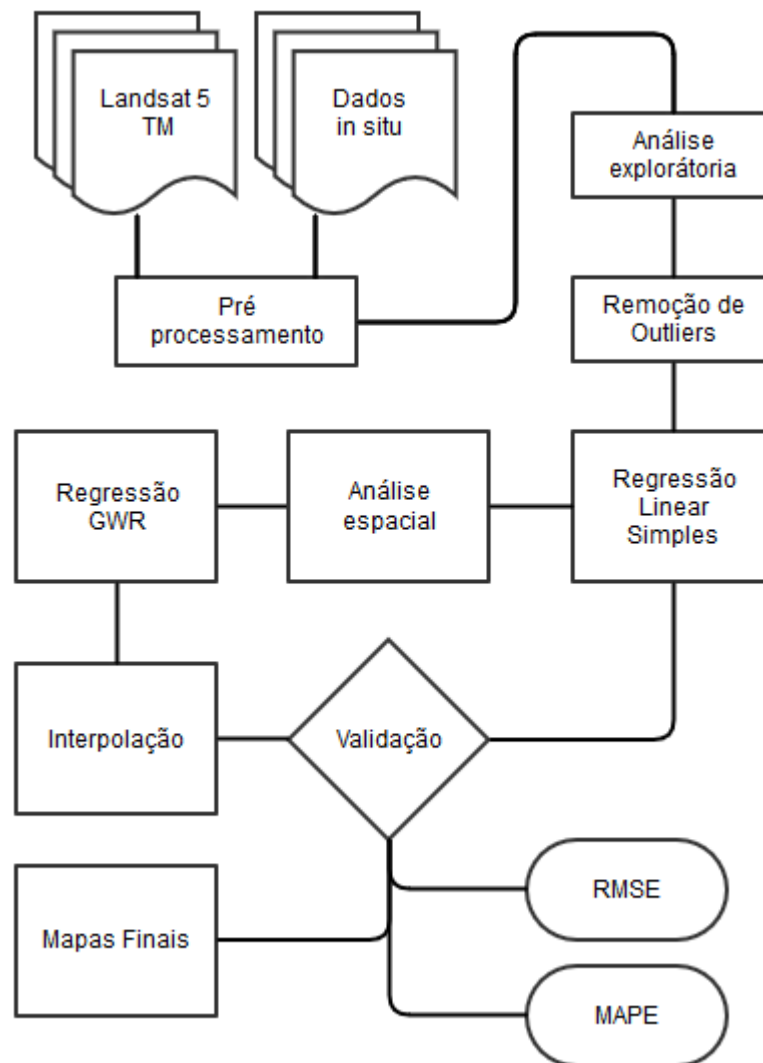


Figura 3: Fluxograma da metodologia utilizada.

3 Resultados e Discussão

3.1 Análise exploratória

A correlação entre os dados de campo e as bandas do sensor utilizado foi analisada através dos gráficos de dispersão apresentados na Figura 4. Foram selecionados os parâmetros clorofila, sólidos totais em suspensão e sólidos inorgânicos em suspensão, os quais foram comparados com as bandas do visível (bandas 1,2 e 3) e infravermelho próximo (banda 4).

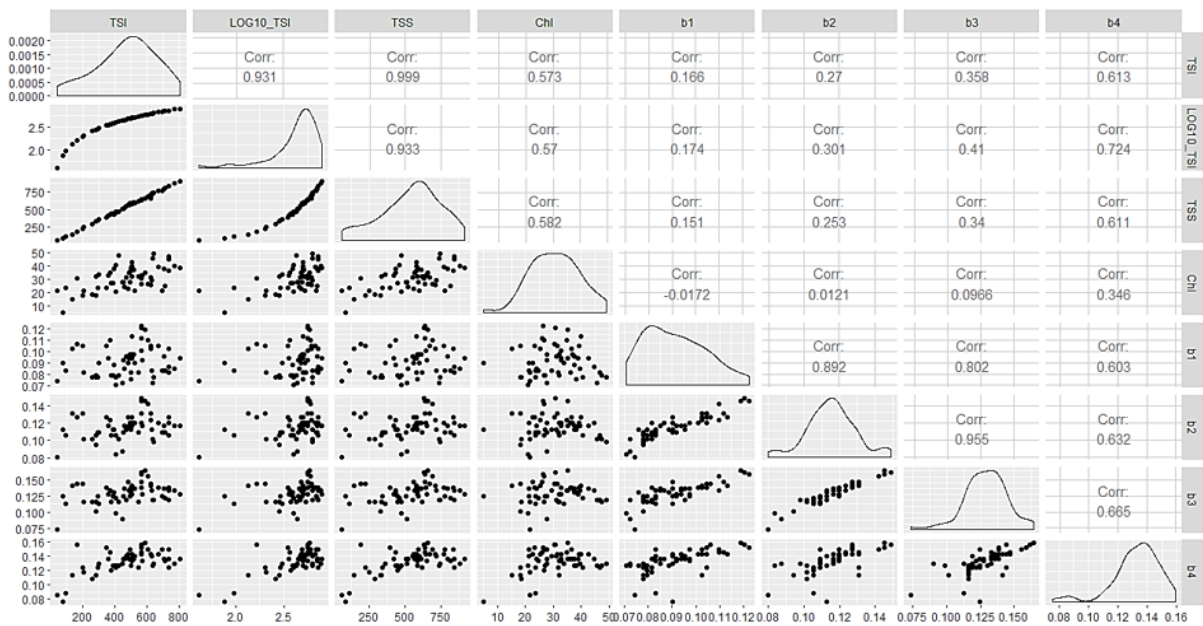


Figura 4: Gráficos de dispersão das variáveis.

As bandas do sensor TM que apresentaram melhores correlações foram as bandas 3 e 4. Estas bandas foram escolhidas para a aplicação do modelo de regressão linear e posteriormente o modelo geográfico GWR. As bandas apresentaram uma melhor correlação com os dados de TSI transformados para escala logarítmica de base 10 (\log_{10}), o que corrobora com Knaeps et al. (2015) e portanto os dados foram trabalhados desta forma nas regressões. Os resultados foram convertidos de volta aos valores normais após a regressão.

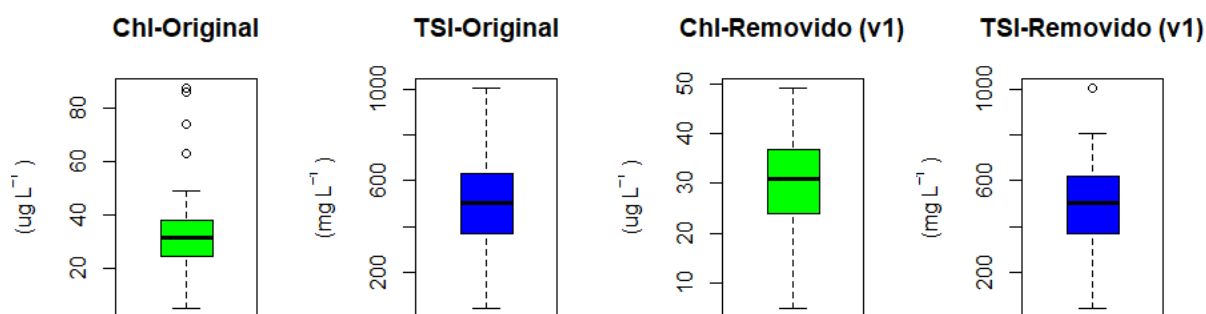
3.1.1 Identificação de dados espúrios

Para análise de "outliers" foram gerados os gráficos de boxplot para os dados originais de clorofila e os sólidos inorgânicos em suspensão. Foram constatados a presença de 4 pontos acima do 3º quartil, sendo "outliers" para a os dados de clorofila, apresentados na Figura 5a.

Os pontos que continham valores espúrios foram removidos da análise e novos gráficos de boxplot foram confeccionados. Com a redução da variação dos dados de TSI após a remoção inicial, foi detectado um novo ponto "outlier", como mostra a Figura 5b. Novamente o valor ex-

(a) Gráfico Boxplot para os da originais.

(b) Gráfico Boxplot após a primeira remoção.



tremo foi retirado da série de dados e uma nova análise boxplot foi realizada, não se constatando a presença de valores espúrios, como mostra a Figura 6.

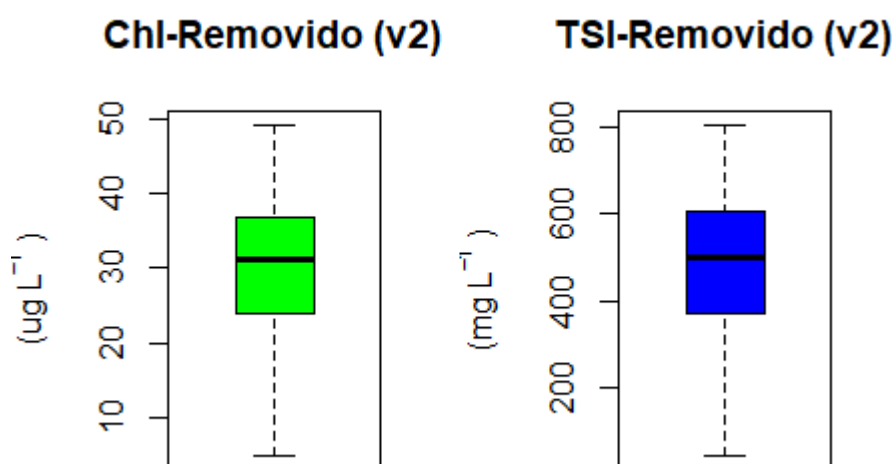


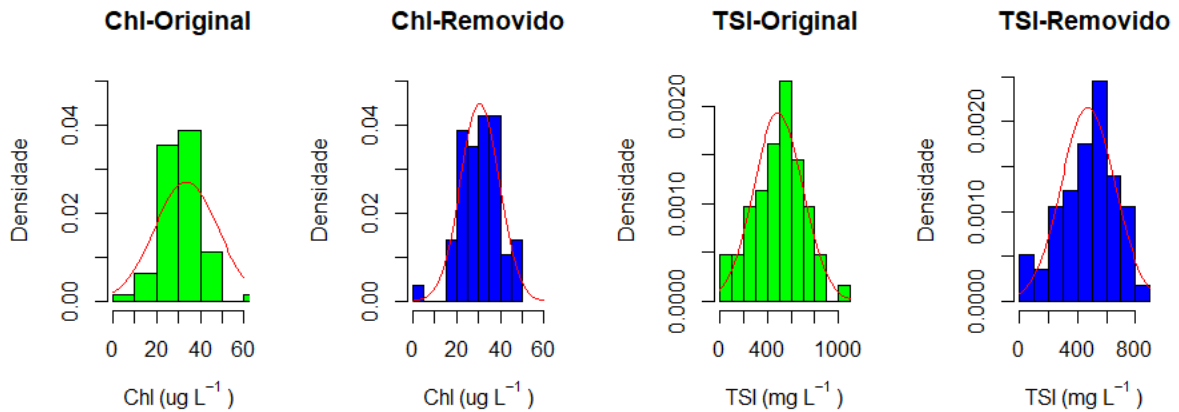
Figura 6: Gráfico Boxplot após a segunda remoção.

3.1.2 Teste de normalidade dos dados

Para analisar a normalidade dos dados utilizados foram confeccionados os histogramas dos dados antes e após a remoção dos "outliers" conjuntamente com a aplicação do teste de Shappiro-Wilk para normalidade. A Figura 7a apresenta a comparação da distribuição de densidade para a clorofila, em verde os dados originais e azul após a remoção. Os valores p obtidos através do teste de Shapiro-Wilk para normalidade apresentaram para os dados originais da clorofila um valor de $7,662e^{-07}$, e o valor de 0,8138 após a remoção dos dados espúrios, atestando assim a normalidade da série. Para o TSI o valor p original obtido foi de 0,803, e de 0,2302 após a remoção, confirmando em ambos os casos a normalidade da série.

(a) Histograma dos dados originais.

(b) Histograma dos dados selecionados.

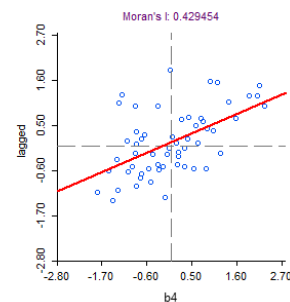
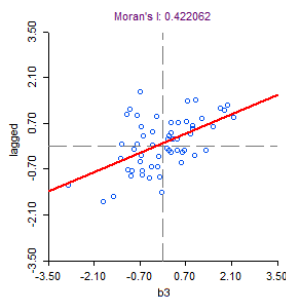


3.2 Diagrama de Moran unidimensional

Os resultados obtidos para o diagrama de Moran utilizando os resíduos das RLS para clorofila em função das bandas (Figura 8a e 8b) mostraram um valor de autocorrelação de 0,422063 para a banda 3 e de 0,429454 para a banda 4. Para o TSI os valores de autocorrelação foram mais baixos, de 0,212962 para a banda 3 e de 0,249212 para a banda 4.

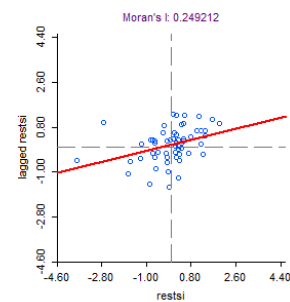
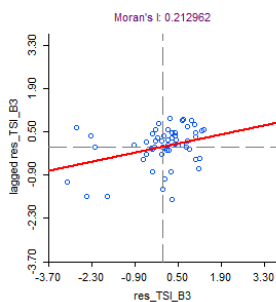
(a) Diagrama de Moran para os Resíduos ChlxB3

(b) Diagrama de Moran para os Resíduos ChlxB4



(c) Diagrama de Moran para os Resíduos TSIXB3

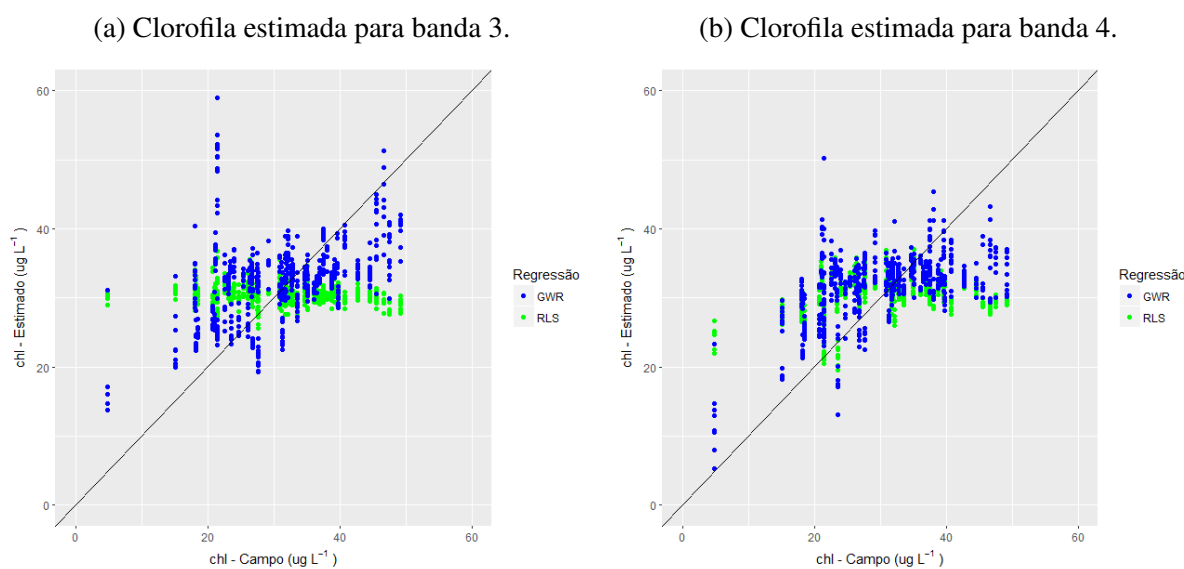
(d) Diagrama de Moran para os Resíduos TSIXB4



O valores obtidos para os diagramas de Moran apresentaram valores próximos ao zero o que indica baixa correlação espacial das variáveis, porém esta correlação existe e possibilita a aplicação da regressão GWR para estes dados.

3.2.1 Validação das Estimativas da RLS x GRW

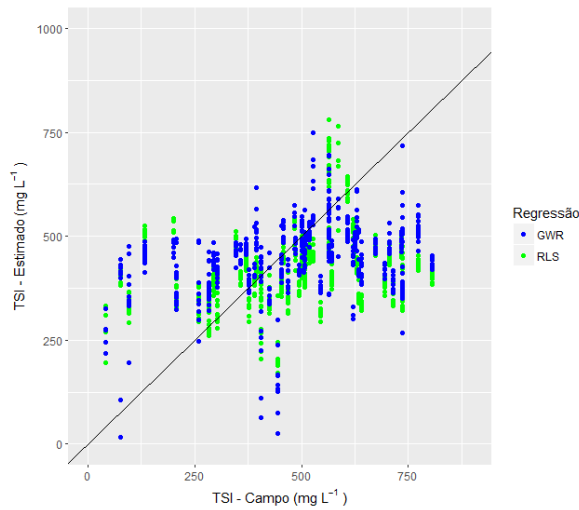
Os dados de campo dos pontos removidos para o conjunto de validação foram comparados com os valores estimados para cada banda em relação aos tipos de regressão. Os pontos em verde representam as 600 observações estimadas pela RLS e os pontos em azul às estimadas pela GWR. A linha central representa a razão 1:1, ou seja, valores preditos sem erros em relação ao campo.



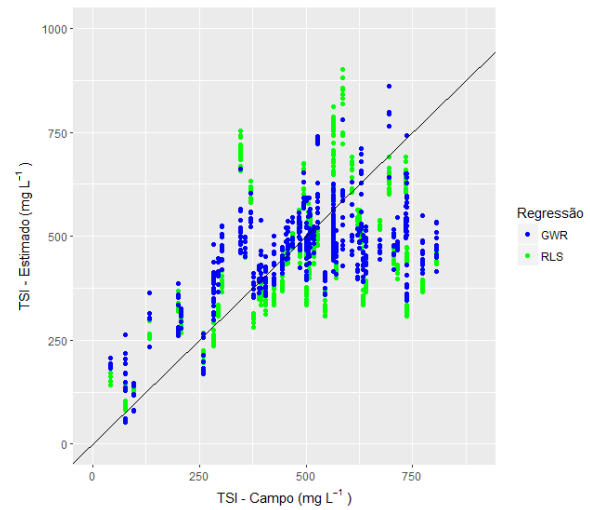
Para as estimativas de clorofila através da banda 3 os dados de validação mostraram (Figura 9a) que a RLS (verde) não foi capaz de reproduzir os dados de campo, superestimando as baixas concentrações e subestimando as altas. Já o GWR (azul) melhorou as estimativas nos extremos mas apresentou uma variância grande nos resultados. Para a banda 4 (Figura 9b) os 2 modelos superestimaram de forma parecida as baixas concentrações, porém a GWR apresentou estimativas melhores para altas concentrações, subestimando menos que a RLS.

As estimativas para o TSI se mostraram bem análogas entre a RLS e a WGR. Para a banda 3 (Figura 10a) se evidencia uma melhora sutil nos valores com concentrações baixas e altas, com pouca diferença entre os modelos na parte central. Na banda 4 (Figura 10b) WGR obteve uma variância menor dos valores estimados do que a RLS, porém ambas as regressões subestimaram as altas concentrações.

(a) Sólidos inorgânicos estimados para banda 3.



(b) Sólidos inorgânicos estimados para banda 4.



A tabela 2 contém os resultados dos coeficientes de determinação, RMSE e MAPE para os pontos utilizados para a validação, discutidos na secção anterior. Para a regressão da clorofila a melhora nos valores de R^2 é expressiva, passando de 0,02 para 0,23 na banda 3 e de 0,07 para 0,28 na banda 4. Isto pode ser explicado pelo valor de auto correlação do diagrama de Moran. Já para o TSI o R^2 não apresenta melhoras muito significativas, de 0,07 para 0,1 na banda 3 e de 0,26 para 0,32 na banda 4. Os Parâmetros apresentaram reduções em seus valores de RMSE, de aproximadamente 15 mg/l para os TSI, porém o seu erro médio absoluto percentual (MAPE) não apresentou grandes variações <2%. Para a clorofila o RMSE reduziu 1 $\mu\text{g/l}$ com uma redução no MAPE de aproximadamente 6%.

Tabela 2: R^2 , RMSE, e MAPE para chl e TSI

	TSI (B3)	TSI (B4)	CHL (B3)	CHL (B4)
R^2 RLS	0,07	0,26	0,02	0,07
R^2 GWR	0,1	0,32	0,23	0,28
RMSE RLS (mg/l $\mu\text{g/l}$)	190,32	171,64	8,92	8,48
RMSE GWR (mg/l $\mu\text{g/l}$)	178,24	154,37	7,75	7,58
MAPE RLS (%)	47,64	28,71	29,17	29,17
MAPE GWR (%)	45,98	28,91	23,49	24,24

3.2.2 Análise dos coeficientes da GWR

Para esta etapa selecionou-se apenas os modelos de regressão dos sólidos totais inorgânicos estimados pela banda 4, devido ao seu melhor desempenho nas análises anteriores, e a presença majoritária deste componente em relação aos outros (Tabela 1). A Figura 11 apresenta os coeficientes derivados da regressão GWR dos sólidos inorgânicos em função da banda 4 para cada ponto de coleta de dados. Estes coeficientes estão dispostos com cores mais escuras para intervalos de valores menores e mais claras para intervalos de valores maiores.

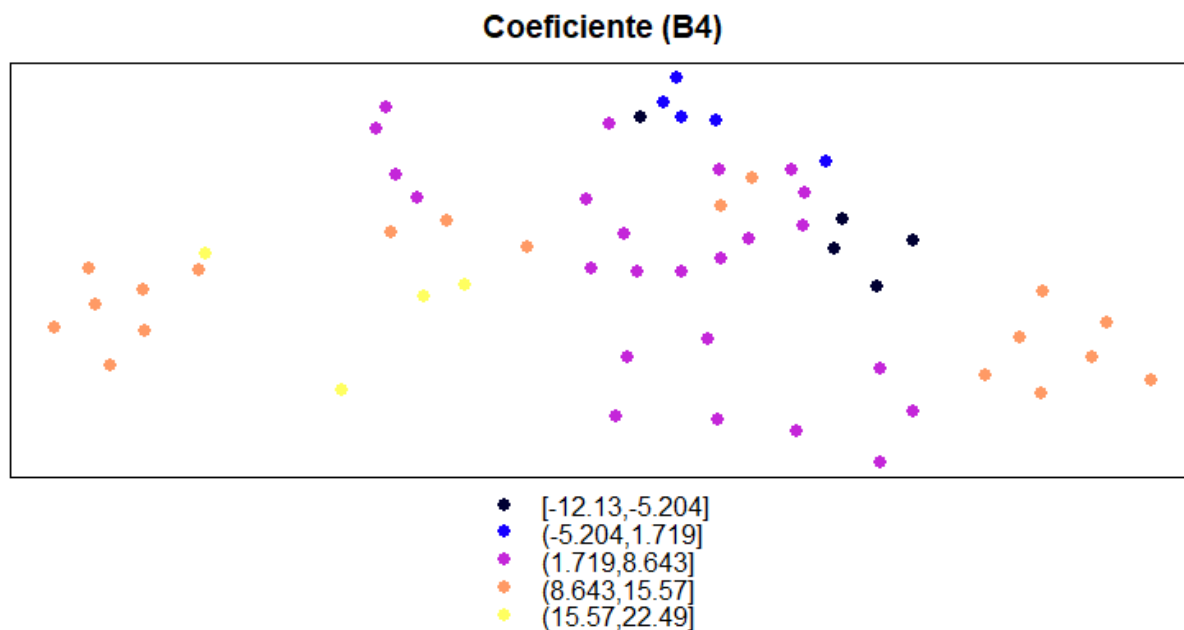


Figura 11: Coeficientes da GWR para a banda 4.

Pode-se notar que existem agrupamentos nos valores dos coeficientes em diversas áreas do lago, sendo relativamente concomitantes com as barreiras físicas desenvolvidas pelo baixo nível da água no período de estudo.

O coeficiente global gerado pela regressão linear simples para as mesmas condições da Figura 11 foi de 11,21. Apenas os valores na cor amarela para os coeficientes da GWR estão acima do valor global, as demais classes tem um coeficiente próximo ao da RLS (rosa) ou menor (roxo, azul e preto). O isolamento de 8 pontos a sudeste do lago (rosa) pode ter influenciado valores altos, bem como para o único ponto isolado em amarelo ao sul.

3.2.3 Mapas de regressão

Devido as altíssimas concentrações de TSI e a baixa relação das bandas escolhidas com a clorofila, foram gerados apenas mapas interpolados para o parâmetro TSI. Os mapas a seguir apresentam os resultados da espacialização para o TSI, estimados a partir dos coeficientes obtidos para as regressões, e aplicados sobre a grade da imagem recortada para máscara de água do lago Grande Curuai. A Figura 12 apresenta as estimativas para os dados de TSI para a banda 4 a partir da regressão linear simples. Os valores do parâmetro medidos em campo estão representados pelos pontos coloridos, os quais seguem a mesma temática de cores, para a comparação com as estimativas do modelo.

Nota-se uma super estimação dos valores na região central-sul do lago (em verde) para a RLS, tal fato é suavizado pela GWR (Figura 13) devido ao agrupamento de pontos com coeficientes menores (1,7 a 8,6) do que o coeficiente global da RLS (11,21).

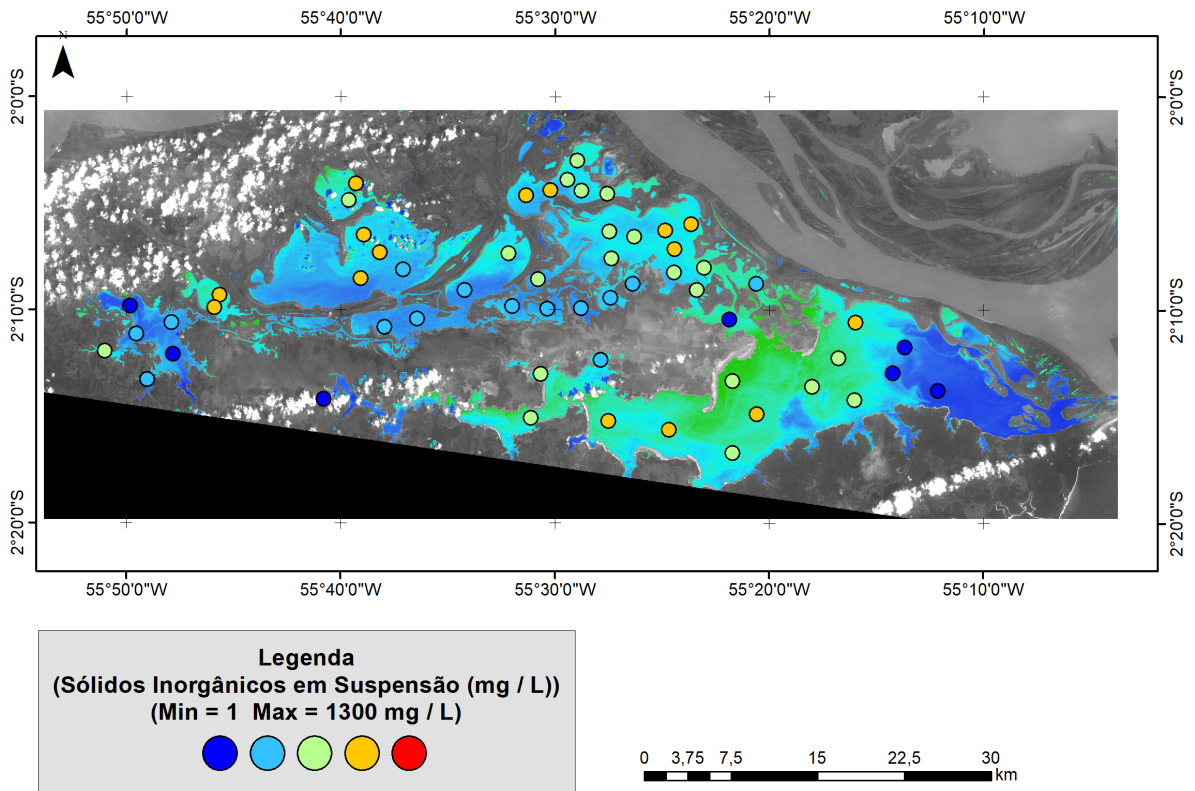


Figura 12: Mapa de regressão estimados pela RLS.

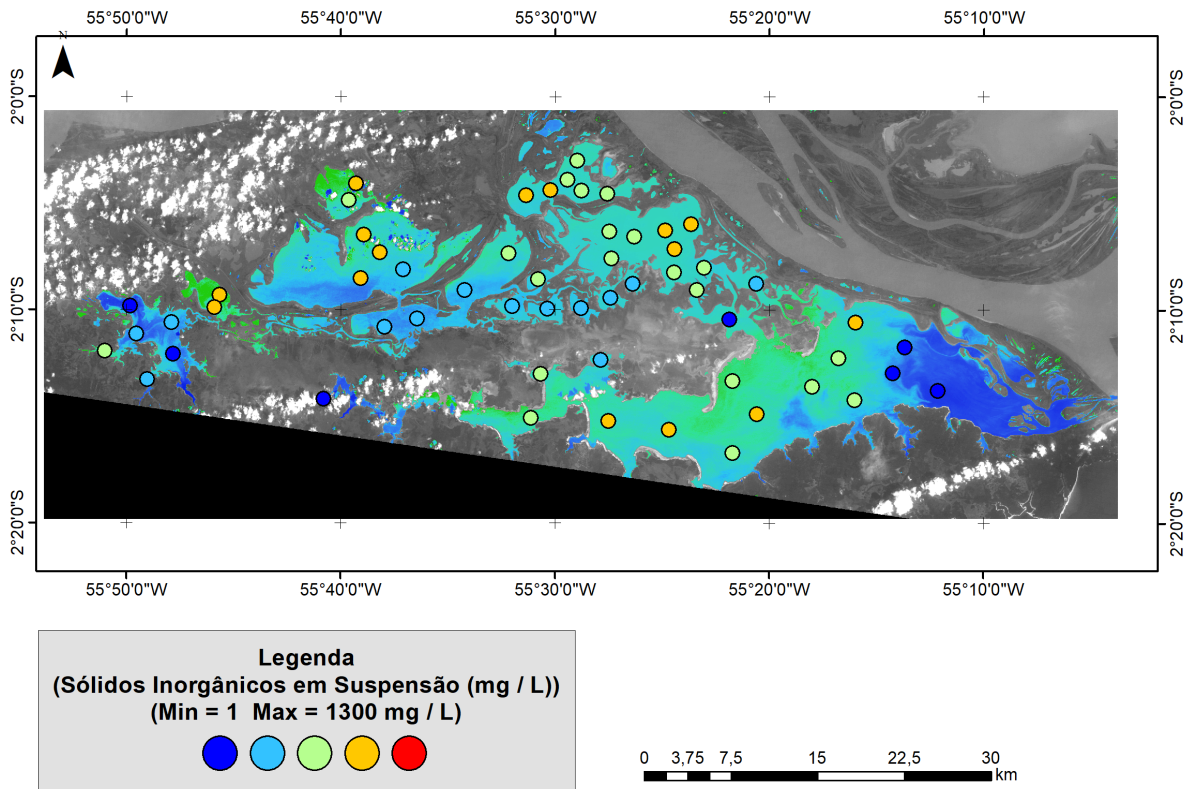


Figura 13: Mapa de regressão estimados pela GWR.

No geral a GWR apresentou uma tendência de estimativas mais altas do que a RLS para todo o lago para os dados de TSI em relação a banda 4. Devido a sua banda adaptativa utilizando vizinhos mais próximos pode-se notar a influência de pontos próximos a pontos com valores menores se comparados aos dados de campo, podendo ser bem notado nas margens da região oeste e norte. Esta influência também pode ser vista nos pontos em azul claro na região central do lago, os quais foram influenciados tanto pelo norte como pelo sul.

A existência de uma grande área seca na região central do lago não foi contabilizada pelo modelo GWR para a determinação do coeficientes baseados nos pontos vizinhos, o qual pode acarretar grandes incertezas às estimativas do parâmetros analisados.

4 Conclusão

A partir deste trabalho foi possível analisar a aplicação de uma técnica de regressão que se utiliza de ponderações geográficas em dados de qualidade da água. Os modelos de regressão puderam ser comparados entre si e em relação as estimativas dos dados de campo.

Pode-se notar o desenvolvimento de aglomerados nos coeficientes da regressão GWR, os quais são concomitantes com a dinâmica do lago no período de estudo, época da seca, devido as divisões naturais.

Foi possível destacar os seguintes fatores como resultados deste trabalho:

- Baixo RMSE geral, devido a relação simples de bandas;
- Método GWR obteve resultados melhores r^2 e RMSE;
- Coeficientes locais próximos devido ao baixo nível da água;
- Baixíssima correlação da clorofila devido a águas altamente túrbidas;
- Os valores de autocorrelação podem explicar as diferentes magnitudes na recuperação da GWR entre a clorofila e os sólidos inorgânicos.

Para o futuro desenvolvimento deste trabalho, busca-se a aplicação de algoritmos empíricos ou semi-analíticos que expressem uma melhor correlação com os parâmetros estudados, bem como, a exploração de técnicas mais robustas para a interpolação dos coeficientes para toda a área de estudo. E por fim inserir como uma nova condicionante a relação temporal dos dados, com a aplicação da "Geographical and Temporal Weighted Regression"(GTWR).

Agradecimentos

O autor agradece os orientadores Cláudio Clemente Barbosa, Lino Augusto Sander de Carvalho, os colegas de trabalho do LabISA e a CNPq pela bolsa de mestrado.

Referências Bibliográficas

- BARBOSA, C. *Sensoriamento Remoto da dinâmica da circulação da água do sistema planície de Curuai/Rio Amazonas. São José dos Campos. 281 p.* Tese (Doutorado) — Tese (Doutorado em Sensoriamento Remoto)—INPE, 2005.
- CHARLTON, M.; FOTHERINGHAM, S.; BRUNSDON, C. Geographically weighted regression. *White paper. National Centre for Geocomputation. National University of Ireland Maynooth*, 2009.
- 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*, Elsevier, v. 65, p. 1–11, 2018.
- DRUCK, S. *Análise espacial de dados geográficos*. [S.l.]: Empraba Cerrados, 2004.
- GORDON, H. R. Dependence of the diffuse reflectance of natural waters on the sun angle. *Limnology and oceanography*, Wiley Online Library, v. 34, n. 8, p. 1484–1489, 1989.
- GUIMARÃES, P. R. B. *Métodos quantitativos estatísticos*. 1ª Edição. Curitiba: IESDE Brasil SA, 2008.
- KNAEPS, E. et al. A swir based algorithm to retrieve total suspended matter in extremely turbid waters. *Remote Sensing of Environment*, Elsevier, v. 168, p. 66–79, 2015.
- MARQUES, A. d. S. et al. Análise exploratória de dados de área para índices de furto na mesorregião de presidente prudente-sp. *Trabalho apresentado no III Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação-PE*, p. 01–08, 2010.
- 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.
- NAKAYA, T. Gwr4 user manual. *WWW Document. Available online: http://www.st-andrews.ac.uk/geoinformatics/wp-content/uploads/GWR4manual_201311.pdf (accessed on 4 November 2013)*, 2014.
- NETER, J. et al. *Applied linear statistical models*. [S.l.]: Irwin Chicago, 1996. v. 4.
- OKIN, G. S.; GU, J. The impact of atmospheric conditions and instrument noise on atmospheric correction and spectral mixture analysis of multispectral imagery. *Remote Sensing of Environment*, Elsevier, v. 164, p. 130–141, 2015.
- PALMER, S. C.; KUTSER, T.; HUNTER, P. D. *Remote sensing of inland waters: Challenges, progress and future directions*. [S.l.]: Elsevier, 2015.
- TEAM, R. C. R. *A language and environment for statistical computing [Internet]. Vienna, Austria; 2014*. 2017.
- VÖRÖSMARTY, C. J. et al. Global water resources: vulnerability from climate change and population growth. *science*, American Association for the Advancement of Science, v. 289, n. 5477, p. 284–288, 2000.