

Explorando a Relação entre Variáveis Ambientais e a Produção de
Clorofila-a em Cianobactérias: Uma Abordagem com Modelos
Lineares Generalizados

Thiago Tavares Lopes

05 dezembro 2024

Sumário

1	Introdução	2
1.1	Cianobactérias	2
1.2	Modelos Lineares Generalizados	3
2	Seleção do modelo	4
2.1	Análise de Resíduos	6

Resumo

XXXXX

1 Introdução

Foi proposto um modelo linear generalizado para avaliar a produção de clorofila a em cianobactérias em diferentes condições climáticas. O *dataset* utilizado foi obtido do catálogo de dados público do governo dos Estados Unidos, disponível em *Data.Gov*. Esse *dataset* possui informações detalhadas sobre diferentes espécies de cianobactérias, sendo estas informações: Informações genéticas, condições climáticas do local de coleta das amostras e localização geográfica da coleta. Para a construção do modelo, foram consideradas exclusivamente as informações de clorofila a quantificada e condições ambientais (físicas e químicas) do local da coleta das amostras. O trabalho aqui desenvolvido foi fundamentado no artigo disponível no seguinte link.

Sobre os dados utilizados, temos as seguintes variáveis e suas respectivas descrições (tabela 1):

Tabela 1: Variáveis utilizadas

Variável	Descrição
chlorophyll_a	Quantidade de Clorofila a ($\mu\text{g/L}$)
total_nitrogen	Quantidade Total de nitrogênio ($\mu\text{g/L}$)
total_phosphorus	Quantidade Total de Fósforo ($\mu\text{g/L}$)
temp_water_celsius	Temperatura da água ($^{\circ}\text{C}$)
dissolved_oxygen	Oxigênio dissolvido (mg/L)
pH_water	pH da água
carbon_dioxide_water	Dióxido de Carbono (mg/L)
total_nitrogen_water	Quantidade de Nitrato, Nitrito, Amônia e Nitrogênio Orgânico (mg/L)
nitrite_water	Quantidade de Nitrito (mg/L)
nitrate_water	Quantidade de Nitrato (mg/L)
phosphorus_water	Quantidade de Fósforo (mg/L)
sulfate_water	Quantidade de Sulfato (mg/L)
total_nitrogen_water	Quantidade de Nitrato, Nitrito, Amônia e Nitrogênio Orgânico (mg/L) - água filtrada
ammonia(NH3 + NH4+)_water	Quantidade de NH3+ e NH4+ (mg/L) como NH4

1.1 Cianobactérias

Devido a atividade fotossintetizante das cianobactérias estima-se que as primeiras tiveram origem entre 2,6 a 3,5 bilhões de anos atrás (LAU e colab., 2015) e são uma das principais responsáveis pela atmosfera oxigenada como conhecemos hoje, participando do “Grande Evento de Oxigenação” (HUISMAN e colab., 2018; PLANAVSKY e colab., 2014; RASMUSSEN e colab., 2008). O estromatólitos são uma evidência da atividade de microrganismos que ocorreu a , aproximadamente 3,700 milhões de anos atrás (NUTMAN e colab., 2016). As cianobactérias são classificadas como microrganismo procariontes autotróficos com sistemas adaptativos particulares, como a capacidade de fixar nitrogênio do ar atmosférico devido a presença da enzima nitrogenase localizada nos heterócitos (PETERS e colab., 2015). São capazes de realizar a fotossíntese na presença ou ausência de oxigênio. Existem espécies que se desenvolvem na ausência de luz ou em condições anaeróbicas utilizando sulfetos como doadores de elétrons para a fotossíntese, além disso são bactérias gram – negativas e dispõem da estrutura chamada bainha mucilaginosa e tricoma, podem ou não apresentar o acinetos e o heterócitos que são estruturas especializadas na sobrevivência da espécie em ambientes não favoráveis (ABED e colab., 2009; COHEN e colab., 1986; HUISMAN e colab., 2018; LAU e colab., 2015; STAL e MOEZELAAR, 1997). A figura 1, apresenta algumas das cianobactérias encontradas na região amazônica. As mesmas podem apresentar estrutura filamentosas, colonial, etc.

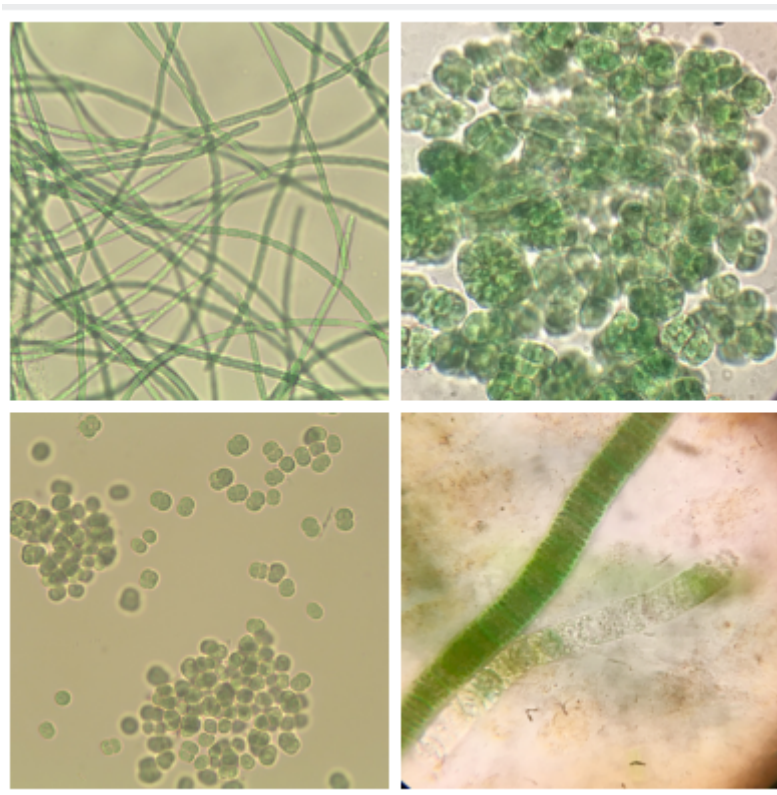


Figura 1: Cianobactérias-Fonte: Coleção Amazônica de Cianobactérias e Microalgas

A clorofila a 2 é um pigmento verde ou azul que capta a luz natural ou sintética e é essencial para a fotossíntese, é encontrado em todos os grupos de vegetais e outros organismos autótrofos, utilizada como indicadora da biomassa em ambientes aquáticos.

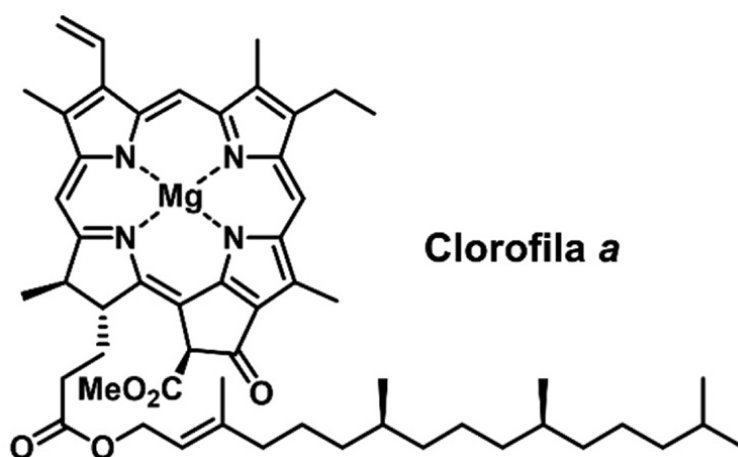


Figura 2: Clorofila a

1.2 Modelos Lineares Generalizados

Um modelo linear generalizado (MLG), é uma extensão dos modelos lineares generalizados em que a variável dependente não segue, obrigatoriamente, uma distribuição normal. Podemos definir a variável resposta como o componente aleatório do modelo, pertencente a mais variadas distribuições como a normal, binomial, exponencial, etc. O componente sistemático, variável independente, do modelo se encaixa de forma linear

ao mesmo. Por último, pode-se definir a função de ligação, que a estrutura responsável por fazer a ligação entre o componente aleatório e o componente sistemático. Em sua definição, temos que $g(\cdot)$ é uma função que transforma a média da variável resposta $E(y) = \mu$ de modo a estabelecer uma relação linear com os preditores da seguinte forma $g(\mu) = X\beta$.

2 Seleção do modelo

Nesta seção são apresentados os resultados da elaboração dos modelos para explicar a variabilidade de produção de clorofila a. Devido a natureza positiva assimétrica da variável resposta $Y(chlorophyll_a)$, foi utilizado a distribuição Gamma com função de ligação logarítmica. Seja $Y \sim Gamma(\mu, \phi)$ em que: μ é a média e ϕ é o parâmetro de dispersão.

A tabela 2, apresenta os resultados do primeiro modelo construído. Nota-se, que apenas as variáveis *total_nitrogen*, *pH_water* e *total_nitrogen_water...8*, foram significativas para o modelo. Por conseguinte, foi utilizado a técnica **stepwise** para escolher um novo modelo baseado no menor AIC.

Tabela 2: Resultados para o modelo 1

	Estimate	Std. Error	t value	Pr(> t)	Significance
(Intercept)	-7,17864	5,53834	-1,29617	0,20167	
total_nitrogen	0,00109	0,00054	2,02436	0,04903	*
total_phosphorus	0,00698	0,00441	1,58323	0,12053	
temp_water_celsius	-0,02700	0,05012	-0,53868	0,59282	
dissolved_oxygen	-0,25390	0,18544	-1,36920	0,17788	
pH_water	1,40239	0,61068	2,29644	0,02647	*
carbon_dioxide_water	0,24729	0,13331	1,85499	0,07031	.
total_nitrogen_water...8	0,28761	0,79233	0,36300	0,71834	
nitrite_water	-0,07889	5,14524	-0,01533	0,98784	
nitrate_water	0,18158	0,21271	0,85367	0,39791	
phosphorus_water	-1,14418	1,13485	-1,00823	0,31886	
sulfate_water	-0,00460	0,00567	-0,81218	0,42106	
total_nitrogen_water...13	-0,46093	0,84540	-0,54521	0,58836	
'ammonia(NH3 + NH4+)_water'	-1,31316	1,93832	-0,67747	0,50165	

^a Nota:*** p<0.001; ** p<0.01; * p<0.05; . p<0.1

Foi ajustado um novo modelo com $AIC = 317$ e Residual Deviance (28,252) que foi menor que a Deviance Nula (46,207). Os resultados são apresentados na tabela 3, observa-se que as variáveis significativas para o modelo são, *total_nitrogen*, *total_phosphorus*, *dissolved_oxygen*, *pH_water*, *carbon_dioxide_water*. Ademais, teste de aderência baseado na razão entre a deviance residual (28,252) e o valor crítico de qui-quadrado $qchisq(0.95, 52)$ (69,832 para 52 GL) sugere que o modelo se ajusta adequadamente aos dados.

Por conseguinte, A clorofila-a é positivamente influenciada por nitrogênio total, dióxido de carbono e pH da água, sendo essas variáveis significativas para o modelo. Por último, Fósforo total e oxigênio dissolvido têm tendências que podem ser relevantes em análises futuras, com mais dados ou em níveis de significância menos conservadores.

Tabela 3: Resultados para o modelo 2

	Estimate	Std. Error	t value	Pr(> t)	Significance
(Intercept)	-4,23976	2,71284	-1,5629	0,12415	
total_nitrogen	0,00089	0,00039	2,2833	0,02653	*
total_phosphorus	0,00693	0,00384	1,8039	0,07704	.
dissolved_oxygen	-0,13273	0,07892	-1,6817	0,09862	.
pH_water	0,83629	0,31895	2,6220	0,01144	*
carbon_dioxide_water	0,07590	0,03738	2,0305	0,04744	*

^a Nota:*** p<0.001; ** p<0.01; * p<0.05; . p<0.1

Por conseguinte, as expressões 1 e 2, representam a estrutura do modelo e o modelo final com os β estimados, respectivamente

$$\eta = \beta_0 + \beta_1 \cdot \text{total_nitrogen} + \beta_2 \cdot \text{total_phosphorus} + \beta_3 \cdot \text{dissolved_oxygen} + \beta_4 \cdot \text{pH_water} + \beta_5 \cdot \text{carbon_dioxide_water} \quad (1)$$

$$\eta = -4.239760 + 0.000889 \cdot \text{total_nitrogen} + 0.006927 \cdot \text{total_phosphorus} - 0.132727 \cdot \text{dissolved_oxygen} + 0.836286 \cdot \text{pH_water} + 0.075904 \cdot \text{carbon_dioxide_water} \quad (2)$$

Por último, a tabela 4 apresenta os coeficientes estimados para o modelo log-linear aplicado a μ , juntamente com a interpretação de cada estimativa. Esses coeficientes indicam como cada variável preditora está associada à resposta em termos do logaritmo natural da média da variável resposta.

Tabela 4: Resultados da regressão log-linear com coeficientes estimados e suas interpretações

Variável	Coef. (Estimativa)	Interpretação
(Intercept)	-4,239760	Valor médio de $\log(\mu)$ quando todos os preditores são iguais a zero.
total_nitrogen	0,000889	Para cada aumento unitário em total_nitrogen, μ cresce em $e^{0,000889}$.
total_phosphorus	0,006927	Para cada aumento unitário em total_phosphorus, μ cresce em $e^{0,006927}$.
dissolved_oxygen	-0,132727	Para cada aumento unitário em dissolved_oxygen, μ diminui em $e^{-0,132727}$.
pH_water	0,836286	Para cada aumento unitário em pH_water, μ cresce em $e^{0,836286}$.
carbon_dioxide_water	0,075904	Para cada aumento unitário em carbon_dioxide_water, μ cresce em $e^{0,075904}$.

2.1 Análise de Resíduos

Neste capítulo, é apresentada a análise de resíduos do segundo modelo desenvolvido. O R^2 generalizado foi calculado utilizando a métrica de Nagelkerke, cujo valor obtido foi 0,41957. Esse resultado indica que o modelo explica aproximadamente 41,96% da variância observada nos dados, sugerindo um ajuste moderado. Este desempenho é coerente com a natureza do conjunto de dados, composto por informações reais coletadas em ambiente natural, que geralmente apresentam alta variabilidade intrínseca.

A figura 3, apresenta o envelope simulado com $\alpha = 5\%$.

Gamma model

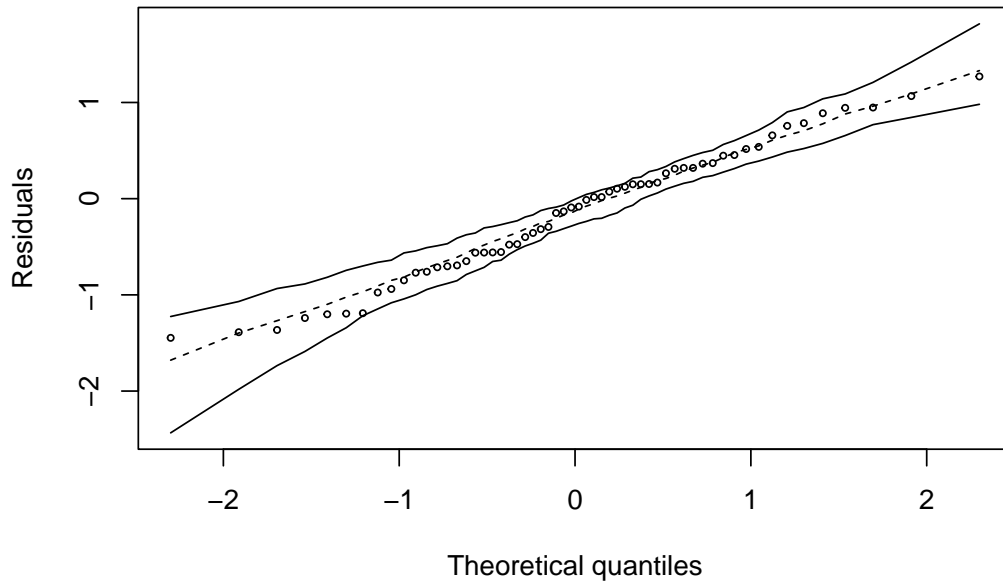


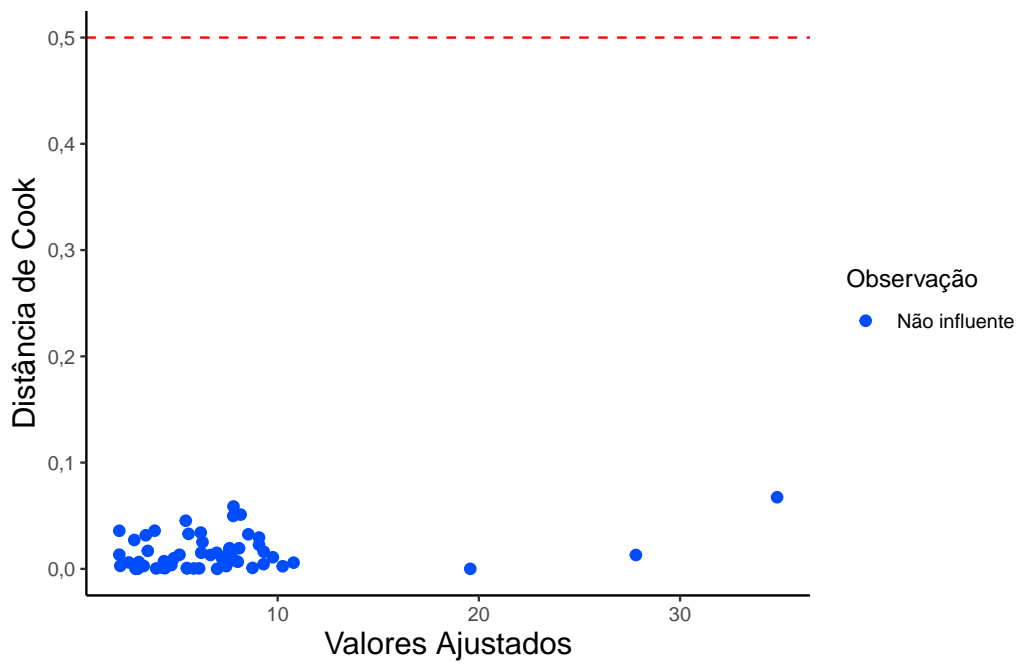
Figura 3: Envelope Simulado

Total points: 58

Points out of envelope: 0 (0 %)

A figura ??, apresenta a distância de cook.

Gráfico de Distância de Cook



A figura ?? apresenta os gráficos para os resíduos.

