

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

Thiago Tavares Lopes

03 dezembro 2024

## Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Cianobactérias	1
<b>2</b>	<b>Seleção do modelo</b>	<b>3</b>
2.1	Análise de Resíduos	4

## 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.

### 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 acineto 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 filamentosa, colonial, etc.

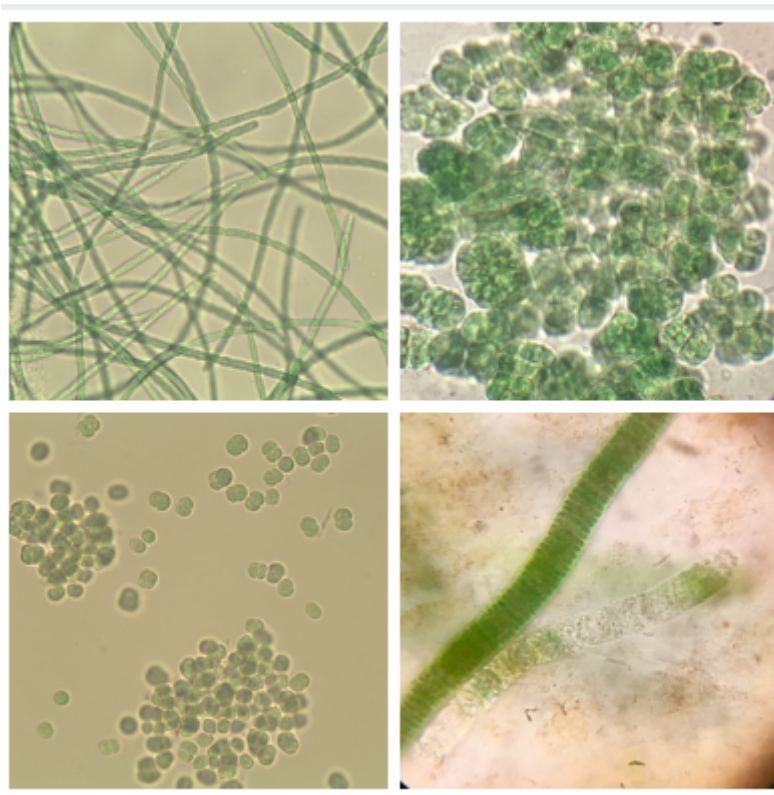


Figura 1: Cianobactérias-Fonte: Coleção Amazônica de Cianobactérias e Microalgas

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.

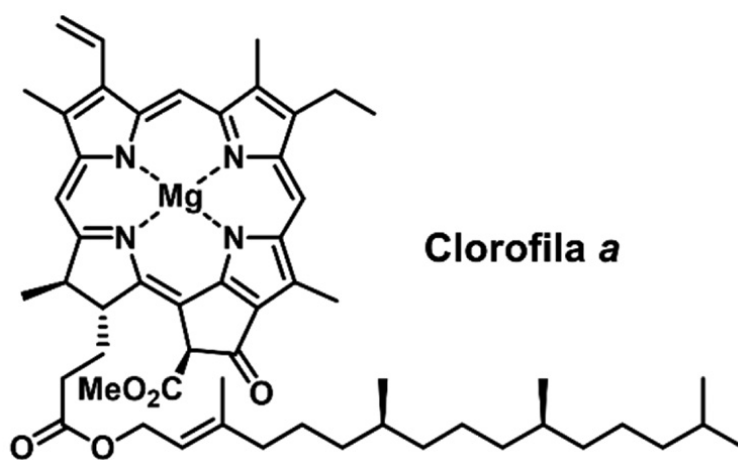


Figura 2: Clorofila a

## 2 Seleção do modelo

Nesta seção são apresentados os resultados da elaboração dos modelos para explicar a variabilidade de produção de clorofila a. Neste modelo, a variável resposta \$ Y (chlorophyll\\_a) \$ segue uma distribuição Gama. A distribuição Gama é usada para modelar variáveis contínuas positivas e assimétricas, como concentrações químicas. Seja  $Y \sim \text{Gamma}(\mu, \phi)$  em que:  $\mu$  é a média e  $\phi$  é o parâmetro de dispersão.

A tabela 1, 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.

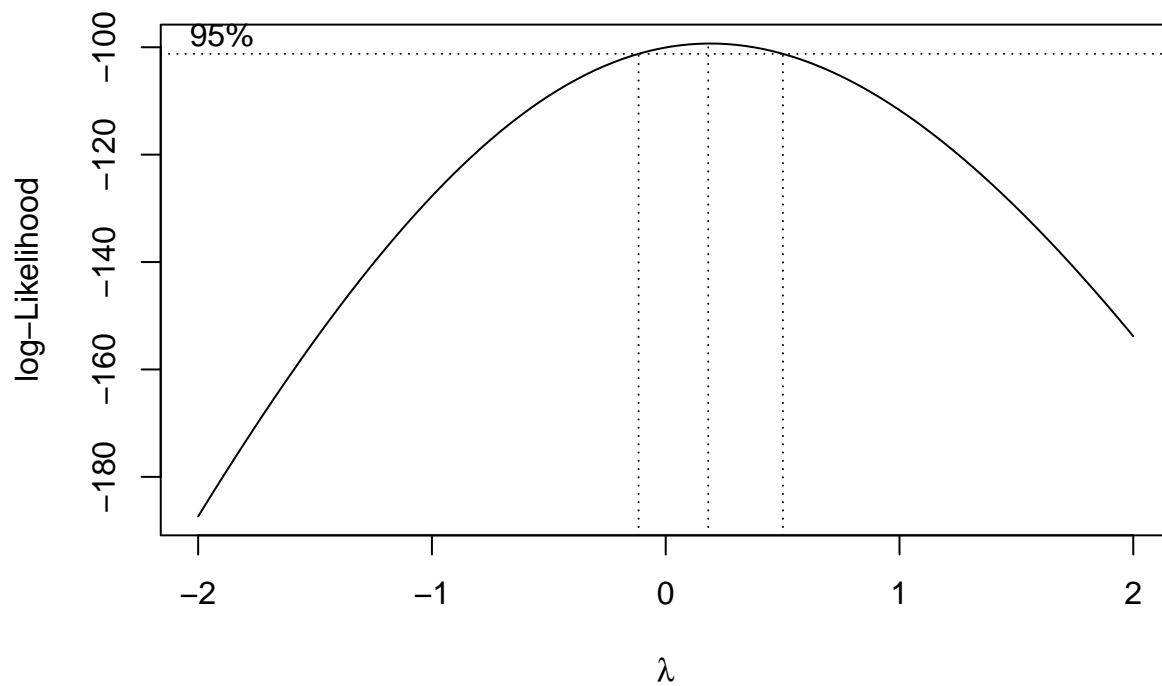
Tabela 1: Resultados para o modelo 1

	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	

<sup>a</sup> Nota:\*\*\* p<0.001; \*\* p<0.01; \* p<0.05; . p<0.1

Foi elaborado um novo modelo com  $AIC = 317$ ,  $Residual\ Deviance = 28,3$ . O mesmo é apresentado na tabela 2, observa-se que as variáveis significativas para o modelo são, *total\_nitrogen*, *total\_phosphorus*, *dissolved\_oxygen*, *pH\_water*, *carbon\_dioxide\_water*.

*#Para verificar se a função de ligação log é apropriada.*  
**boxcox(fit2)**



*#Certifique-se de que todas as variáveis explicativas têm uma relação linear na escala da função de ligação*

*#crPlots(fit2)*

*#bptest(fit2)*

Tabela 2: 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	*

<sup>a</sup> Nota:\*\*\* p<0.001; \*\* p<0.01; \* p<0.05; . p<0.1

## 2.1 Análise de Resíduos

*# RESIDUAL ANALYSIS*

*# coeficiente de determinacao generalizado (Nagelkerke, 1991)*

`RsqGLM(fit2,plot=F)`

## NOTE: Tjur R-squared applies only to binomial GLMs

## \$CoxSnell

## [1] 0,41826

##

```
## $Nagelkerke
## [1] 0,41957
##
## $McFadden
## [1] 0,093991
##
## $Tjur
## [1] NA
##
## $sqPearson
## [1] 0,38694
## DIAGNOSTIC ANALYSIS

# td = resid(fit2,type="deviance")
#
# # RESIDUOS X INDICES
#
# abline(plot(td), lty=c(2,2,3), h=c(-2,2,0))
#
#
# abline(plot(fitted(fit2),td) , lty=c(2,2,3), h=c(-2,2,0))

#fitted(fit2) - aqui os valores já estão na escala original da variável resposta
#valores estimados para o preditor linear
#fit2$linear.predictors

## Gamma model
```

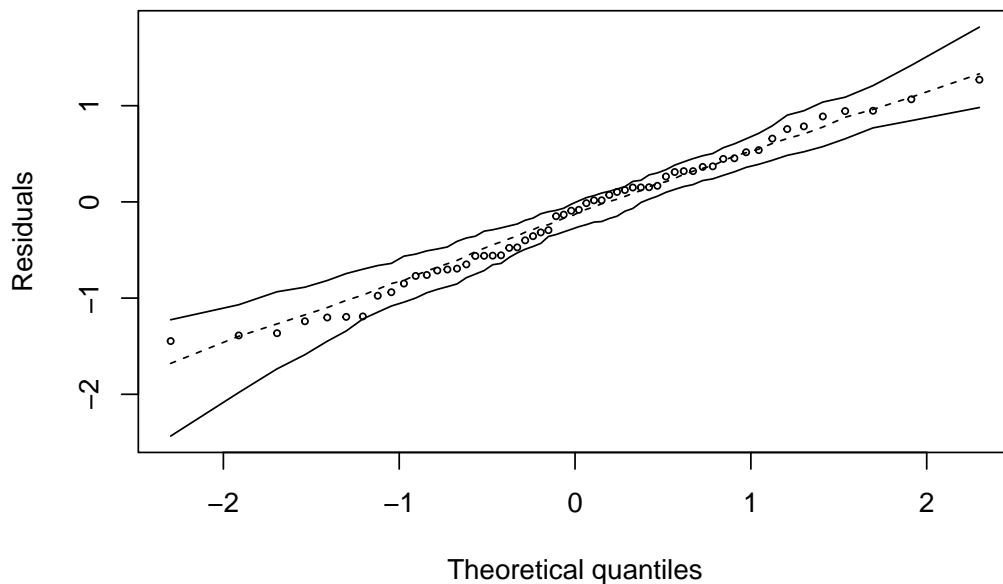


Figura 3: Envelope Simulado

```
## Total points: 58
```

## Points out of envelope: 0 ( 0 %)

