

# Estimativa de concentração de sedimentos em suspensão por imagens de sensoriamento remoto

Prof. Dr. Édipo Henrique Cremon  
edipo.cremon@ifg.edu.br



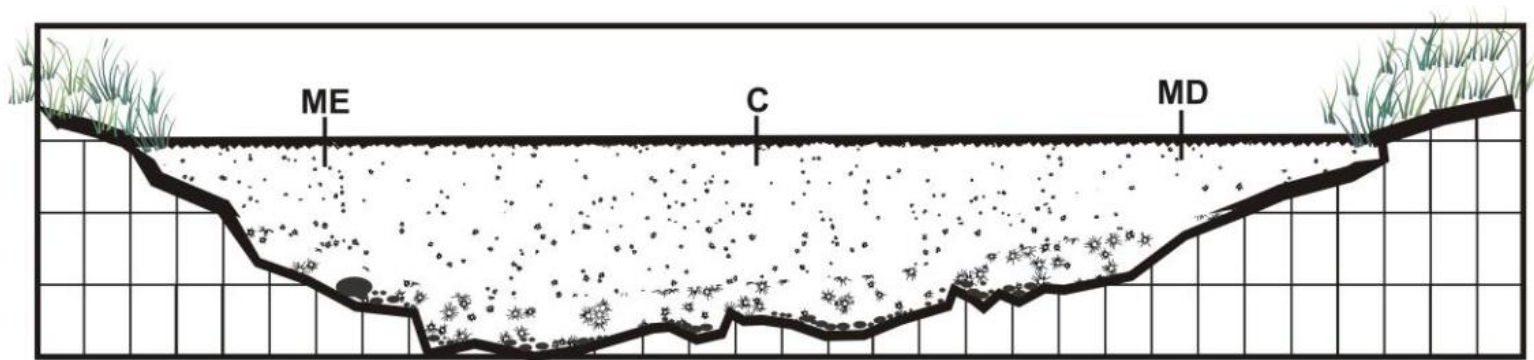
## Como medir a concentração de sedimentos em suspensão??

O sedimento suspenso é medido através da coleta de recipientes de água e seu envio a um laboratório para determinar a concentração. Como a quantidade de sedimentos que um rio pode transportar muda com o tempo, os hidrólogos fazem medições e amostras à medida que o fluxo do rio aumenta e diminui durante o ciclo hidrológico ou evento meteorológico.



<https://doi.org/10.4025/bolgeogr.v28i1.8472>

## Como medir a concentração de sedimentos em suspensão??



- As coletas são realizadas em seções transversais do canal;
- E pode ser realizada em diferentes profundidades ou percorrendo toda a coluna d'água da base para o topo, enquanto a garrafa é preenchida.
- Uma integralização ou média é calculada para obter a concentração de sedimentos para um determinado trecho do canal.

## Como medir a concentração de sedimentos em suspensão??

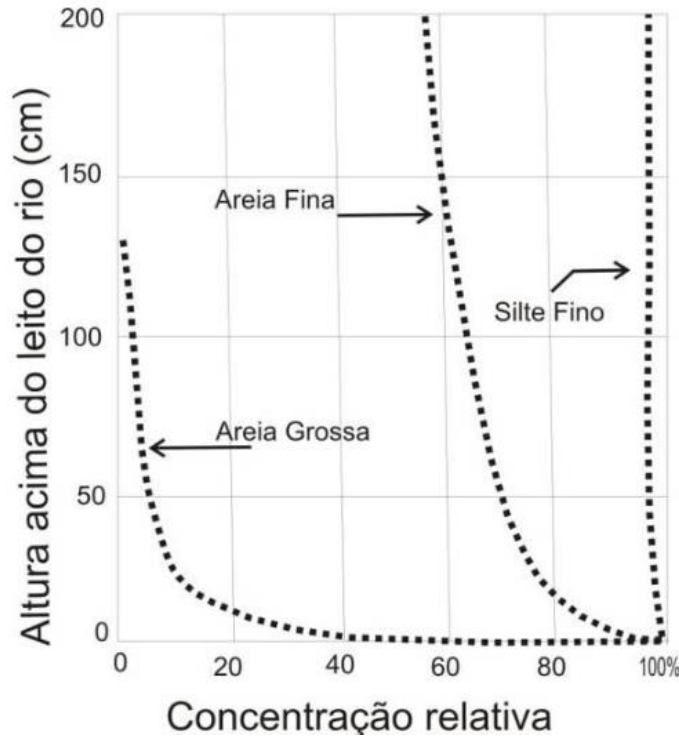


Figura - Distribuição vertical das partículas em suspensão. Mod. (SUGUIO; BIGARELLA, 1979).

- A amostragem percorrendo toda a coluna d'água da base para o topo, enquanto a garrafa é preenchida é utilizada muitas vezes pelas características da granulometria dos sedimentos na coluna d'água.

## Como medir a concentração de sedimentos em suspensão??

Uma vez que sabemos quanta água está fluindo e a quantidade de sedimentos na coluna d'água em diferentes condições de fluxo, podemos calcular a quantidade de sedimentos em toneladas que passam pelo local de medição durante um dia, durante eventos meteorológicos extremos ou durante o ano inteiro.

## Problemas

- A obtenção da concentração de sedimentos em suspensão através dos métodos convencionais demandam grandes custos, como mão de obra qualificada, reagentes, materiais, tempo e de logística.
- Além das condições instáveis, como rios com grande vazão ou ventos fortes, é difícil acessar o local e, assim, obter amostras de água nesses locais.
- Em outras palavras, é difícil uma análise temporal e regional (espacial) confiável porque requer um grande número de pontos amostrados no tempo e espaço, o que geralmente não é possível devido ao alto custo.

# Demanda

- Há demanda de metodologias para medir ou estimar a concentração de sedimentos em suspensão com eficiência e baixo custo.

## Sensoriamento Remoto



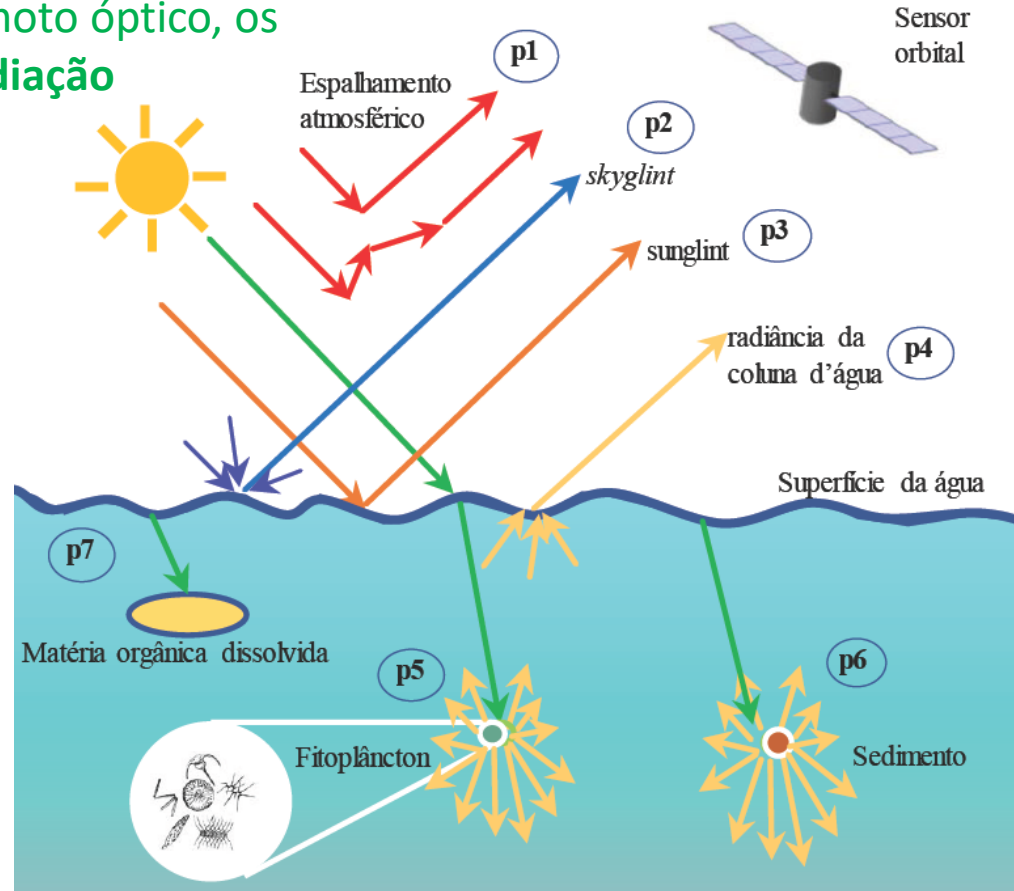
Michigan Tech Research Institute



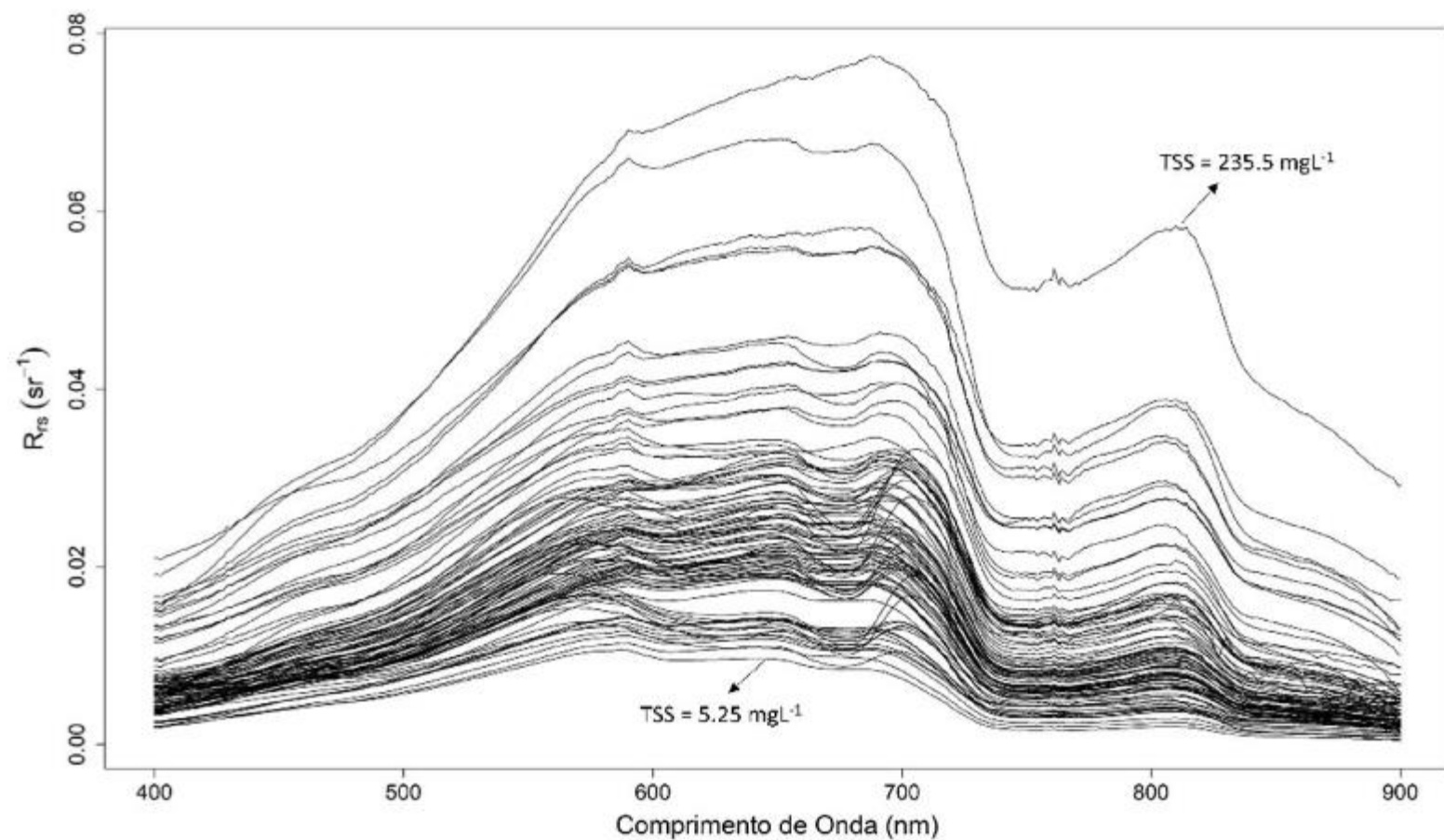
# O que um sensor mede?

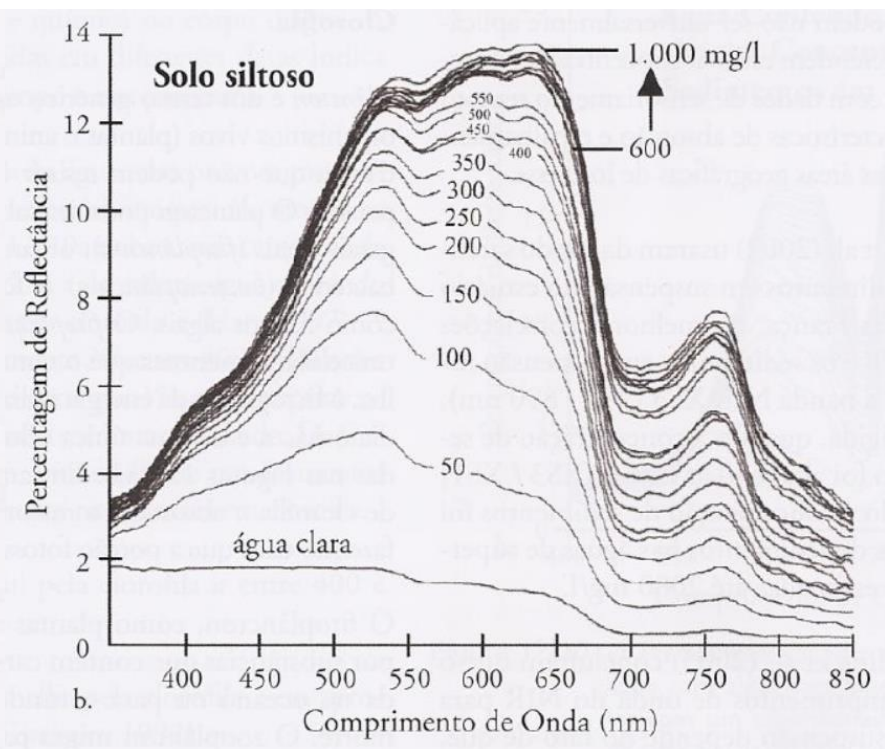
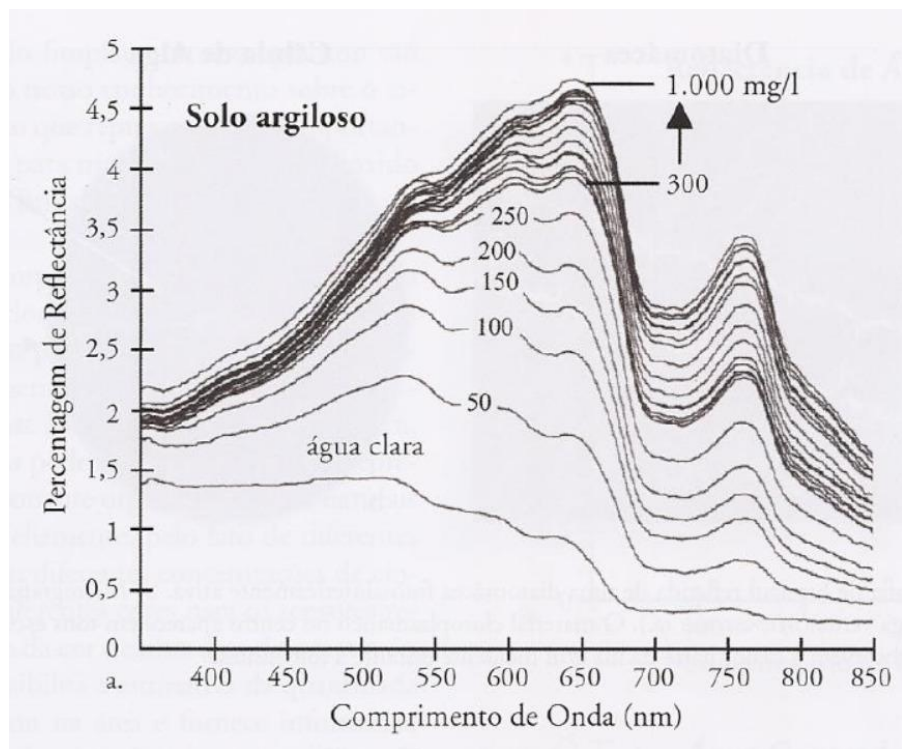
No sensoriamento remoto óptico, os sensores **medem a radiação refletida** pelos alvos.

Trajetória da luz entre fonte (sol), alvo (corpo d'água) e sensor orbital. (1) processo de espalhamento atmosférico, (2) processo de reflexão especular da luz difusa do céu, (3) processo de reflexão especular da luz solar direta, (4) radiação emergente da coluna d'água, (5) processo de espalhamento e absorção da luz por fitoplâncton, (6) processo de espalhamento da luz por sedimentos, (7) processo de absorção da luz por matéria orgânica dissolvida. Trajetória da luz entre fonte (sol), alvo (corpo d'água) e sensor orbital. (1) processo de espalhamento atmosférico, (2) processo de reflexão especular da luz difusa do céu, (3) processo de reflexão especular da luz solar direta, (4) radiação emergente da coluna d'água, (5) processo de espalhamento e absorção da luz por fitoplâncton, (6) processo de espalhamento da luz por sedimentos, (7) processo de absorção da luz por matéria orgânica dissolvida.









Medidas in situ de reflectância com vários níveis de CSS variando de 0 a 1000 mg/L.  
Fonte: Lodhi et al. (1997)

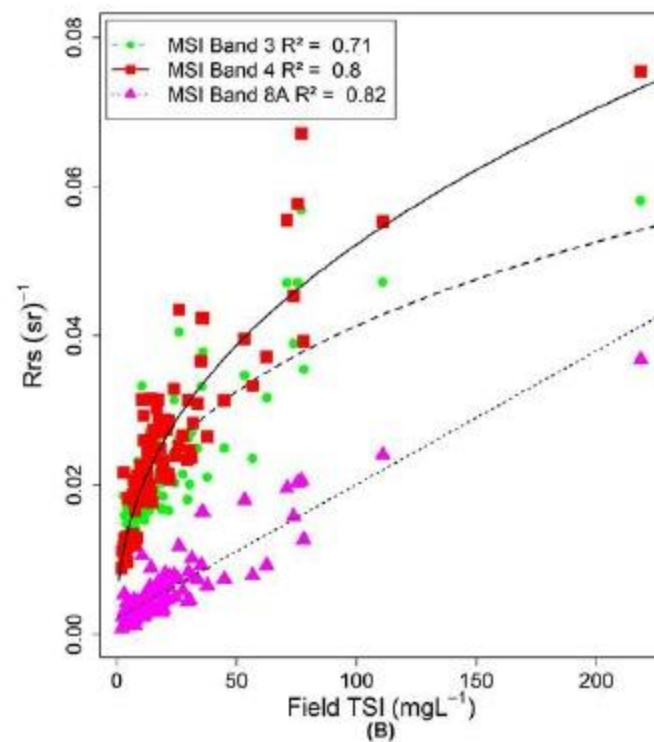
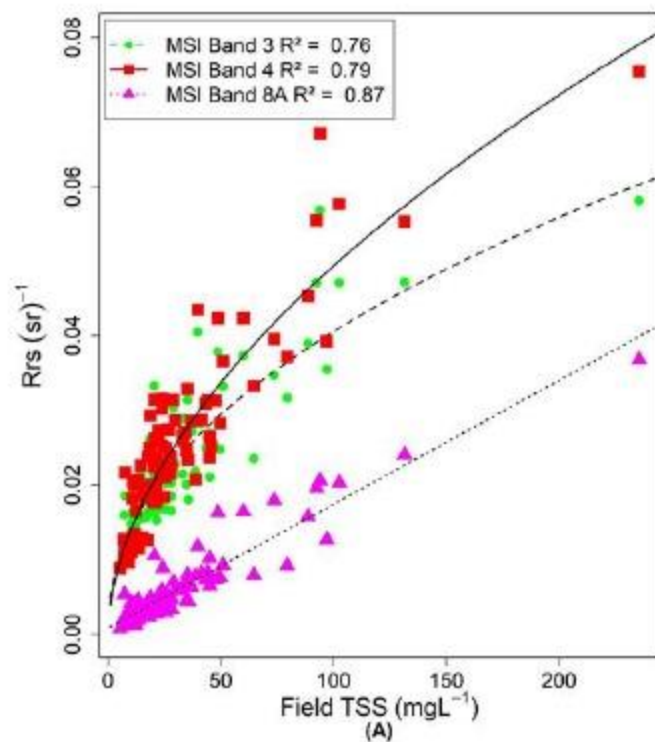
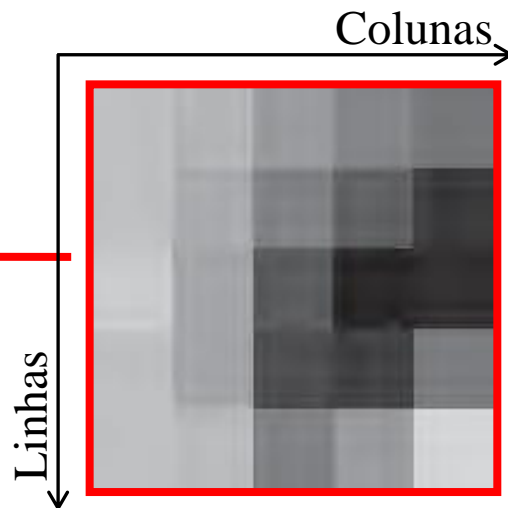


Figura 5.16. Exemplo de saturação das bandas 3 (550 nm) e 4 (665 nm) dos Sentinel-2, em relação à banda 8A (860 nm), com dados obtidos no Lago Curuai, estado do Pará (adaptado de (Maciel, et al., 2019) (in prep.).

# O que significa o valor do pixel na imagem de SR?



181	168	134	107	87
181	154	101	60	20
195	148	53	6	13
181	148	40	26	94
181	175	101	134	208

- Usualmente, os valores dos pixels nas imagens de SR representam os chamados números digitais (ND). Possuem relação com a radiância refletida, entretanto não possuem significado físico para relacionar os alvos com suas propriedades físico-químicas.

# O que significa o valor do pixel na imagem de SR?

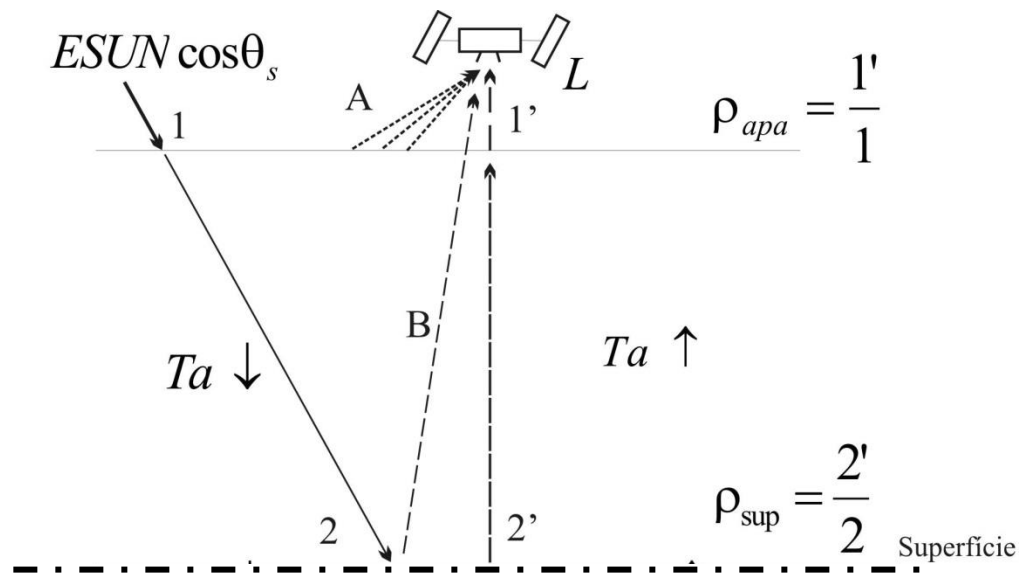
- A radiância ( $L_\lambda$ ) registrada pelos sensores é convertida em número digital ( $ND_\lambda$ ) para permitir a manipulação e armazenamento em ambiente computacional;
- As imagens são disponibilizadas pelos distribuidores geralmente nesta unidade.
- O  $ND_\lambda$  não possui significado físico, sendo apenas uma representação computacional necessária para a codificação da grandeza física, não devendo, portanto, ser diretamente correlacionado com as variáveis biofísicas ou geofísicas de interesse (PONZONI; SHIMABUKURO, 2007).

# Conversão de ND para valores físicos

$$ND_{\lambda} \rightarrow L_{\lambda} \rightarrow \rho_{apa(\lambda)} \rightarrow \rho_{sup(\lambda)}$$



# Conversão de ND para valores físicos



$ESUN_{(\lambda)}$  = irradiância solar média no topo da atmosfera ( $W/m^2/\mu m$ );

$\theta_s$  = ângulo zenital solar (graus);

$Ta \downarrow_{(\lambda)}$  = transmitância atmosférica para a irradiância incidente no topo da atmosfera;

$Ta \uparrow_{(\lambda)}$  = transmitância atmosférica para a irradiância ascendente da superfície;

$\rho_{apa(\lambda)}$  = reflectância aparente ou ToA (*Top of Atmosphere*), definida pela razão entre a estimativa da irradiância ascendente a partir da radiância captada pelo sensor e a irradiância solar incidente no topo da atmosfera;

$\rho_{sup(\lambda)}$  = reflectância de superfície, definida pela razão entre a irradiância ascendente da superfície e a irradiância incidente nela;

$A_{(\lambda)}$  = fluxo radiante refletido pela atmosfera e captado pelo sensor;

$B_{(\lambda)}$  = fluxo radiante refletido pela superfície e captado pelo sensor.

# Conversão de ND para valores físicos

- Na prática, as bases de dados já tem disponibilizado as imagens em diferentes níveis de processamento.
- É possível adquirir imagens com valores de pixel em reflectância de superfície (com correção atmosférica);
- Ou mesmo em reflectância ToA (sem correção atmosférica).



# Correção atmosférica

“Os métodos de correção atmosférica são constantemente avaliados e não existe consenso sobre o melhor modelo, visto que a qualidade dos produtos variam amplamente em função das condições atmosféricas, tipo de alvo, disponibilidade de dados auxiliares e familiaridade com o modelo” (Martins et al., 2017).

## Tome nota

É necessário que as imagens estejam em reflectância ToA ou de superfície (pode-se avaliar diferentes métodos de correção atmosférica) para relacionar as propriedades físico-químicas da superfície terrestre com o valor de pixel de imagens de Sensoriamento Remoto.

**Dica: se você não é *hardcore* do SR, busque imagens já corrigidas.**

# Relacionamento entre Variáveis

Por exemplo:

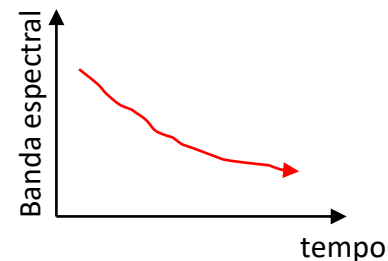
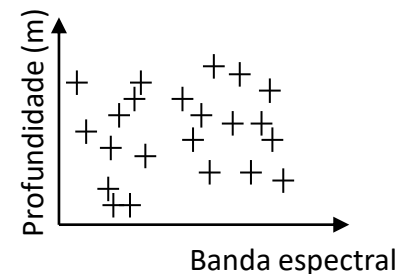
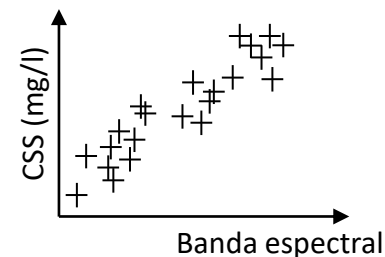
Considere que uma determinada banda espectral (Banda) apresenta valores de reflectância baixos para concentração de sedimentos em suspensão e apresente valores altos de reflectância para maiores concentrações de sedimentos em suspensão.

Por outro lado, esta mesma banda não tem qualquer relação com a profundidade do canal.

Se observarmos uma diminuição do valor desta banda ao longo do tempo, o que podemos concluir quanto a dinâmica da concentração de sedimentos suspensos deste lugar?

**Quanto à CSS, espera-se que tenha havido uma diminuição**

**Quanto à profundidade, nada podemos afirmar**



## O que almejamos?

- Elaborar modelos para estimar propriedades físico-químicas dos alvos por sensoriamento remoto!

Inteligência Artificial

Machine Learning

Deep Learning

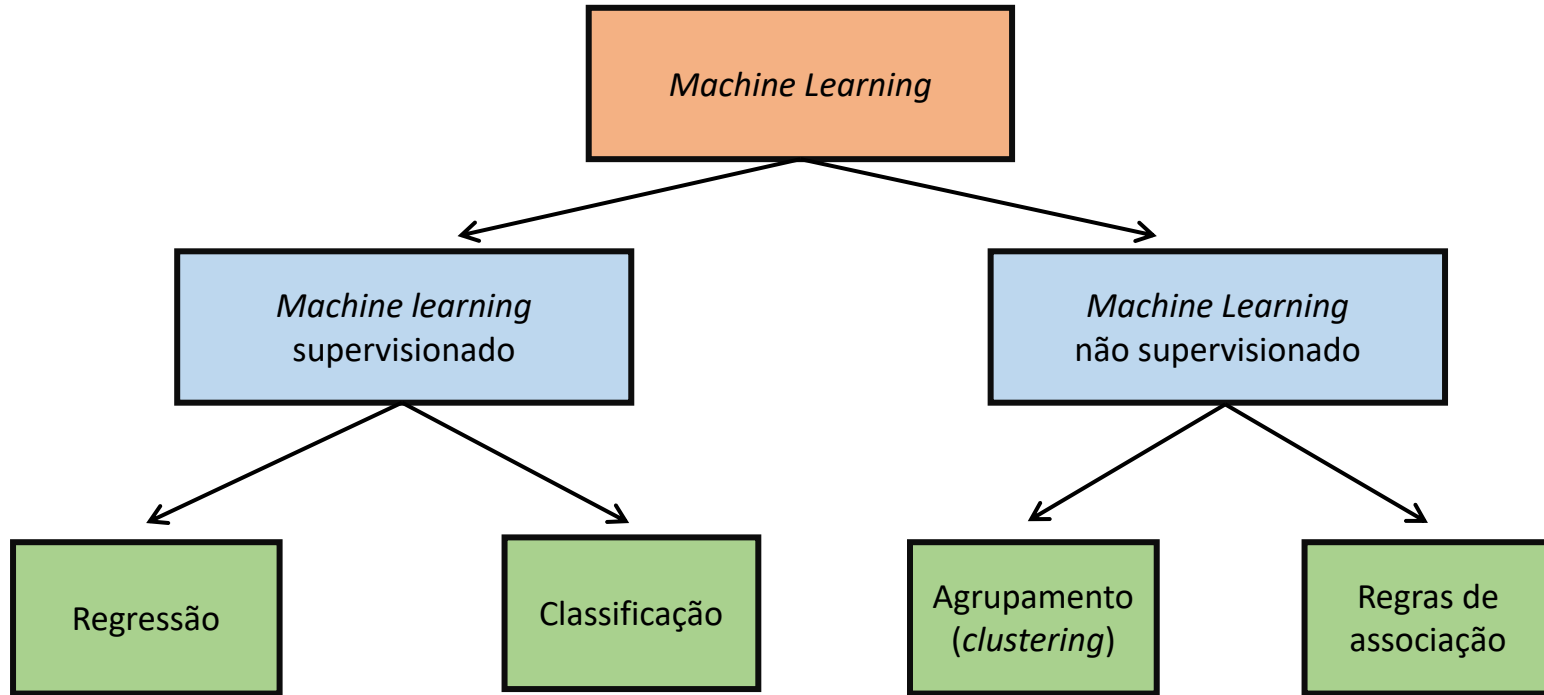
Ciência de  
Dados

Estatística

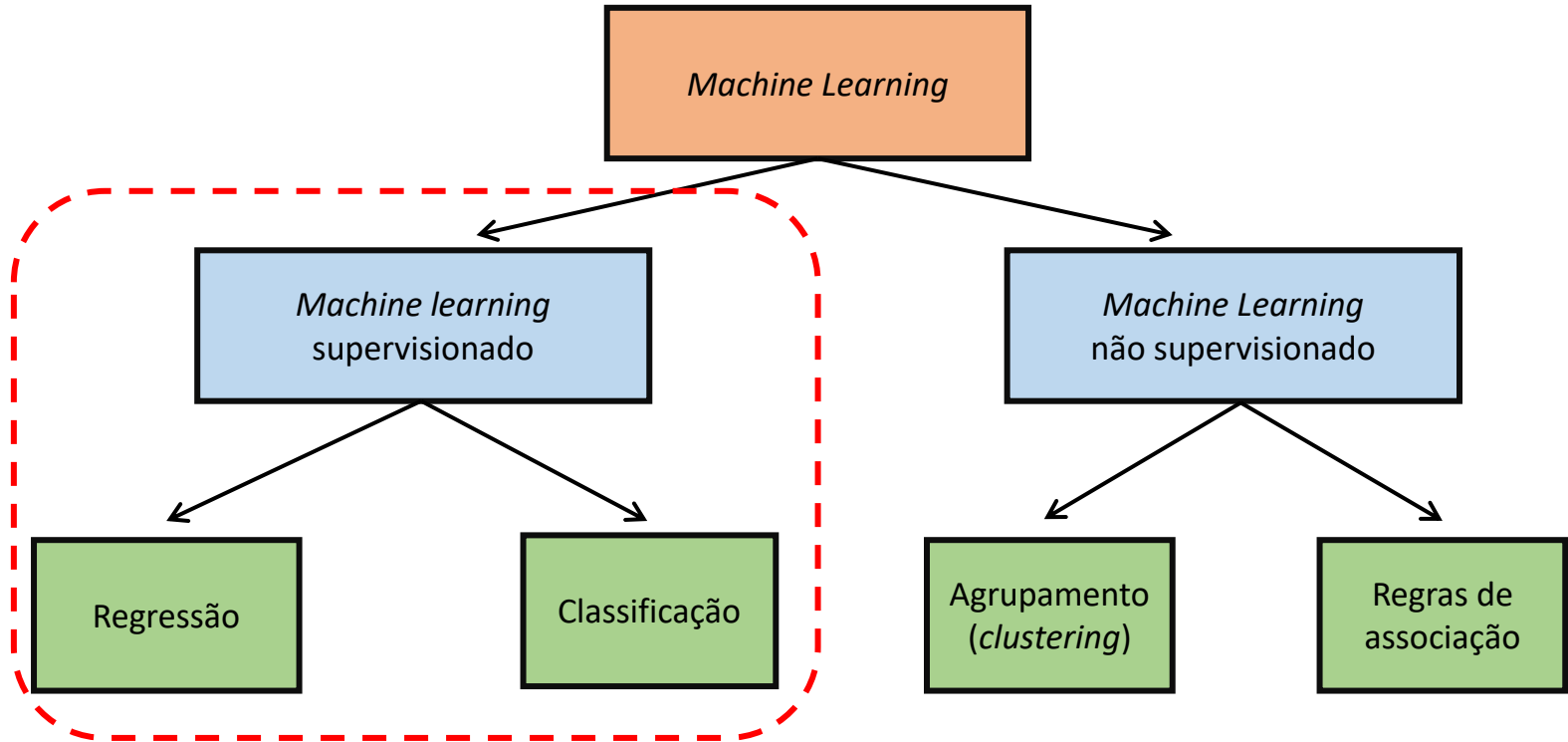
Mineração de  
Dados



# Machine Learning

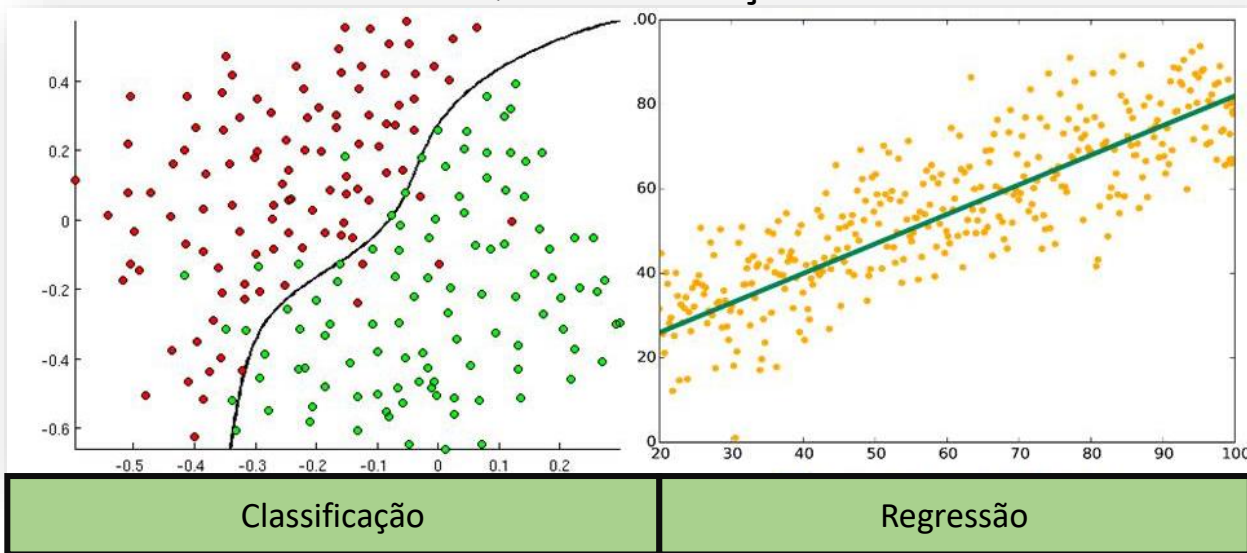


# Machine Learning



# Machine Learning

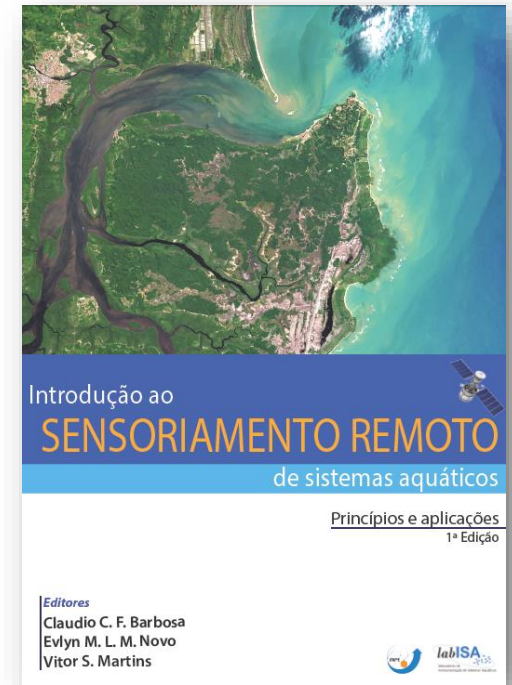
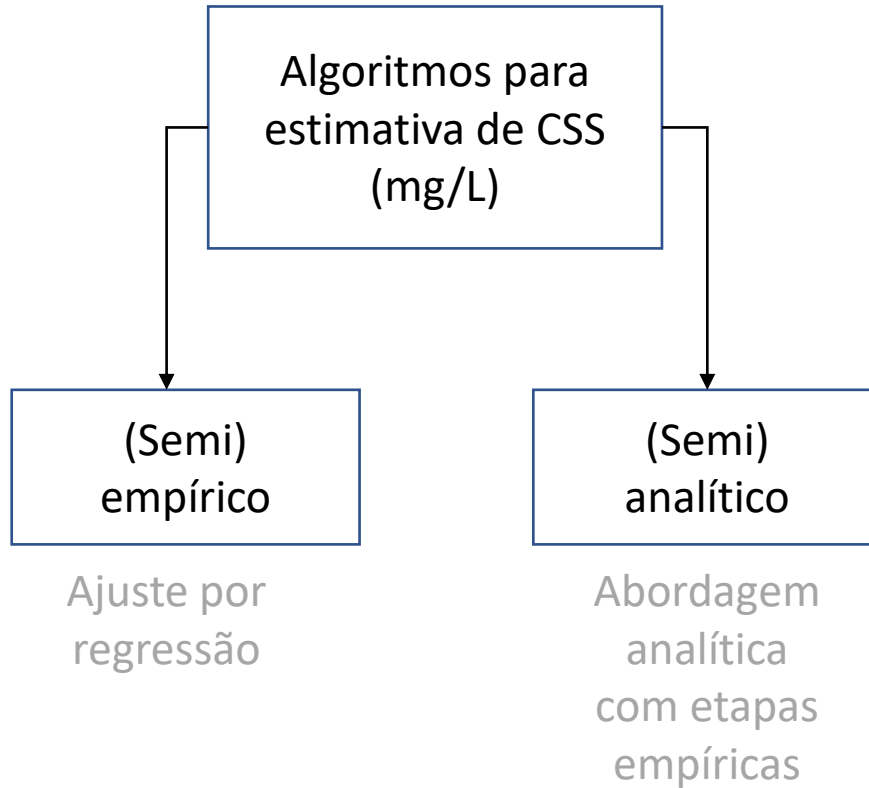
Qual a diferença entre?



O algoritmo busca prever dados classificando-as em diferentes categorias

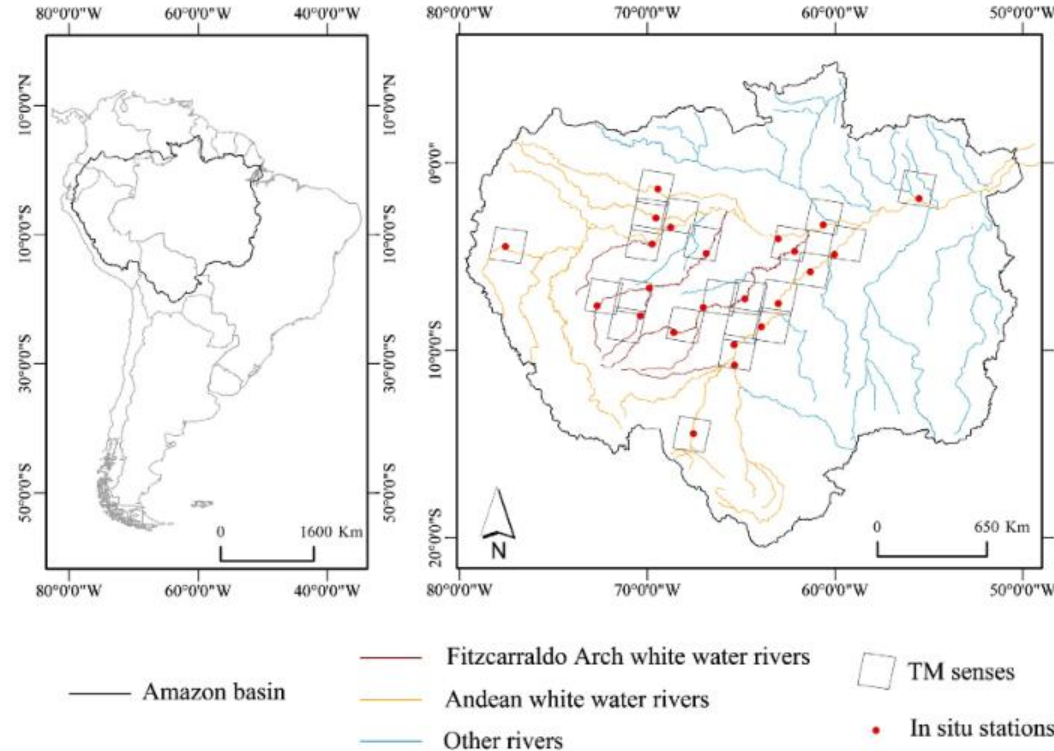
O algoritmo tenta prever um valor contínuo para uma entrada de dados

# Sedimentos suspensos





# Sedimentos suspensos



## Empirical models for estimating the suspended sediment concentration in Amazonian white water rivers using Landsat 5/TM

Otávio C. Montanher<sup>a,b,\*</sup>, Evelyn M.L.M. Novo<sup>a</sup>, Cláudio C.F. Barbosa<sup>a</sup>, Camilo D. Rennó<sup>a</sup>, Thiago S.F. Silva<sup>c</sup>

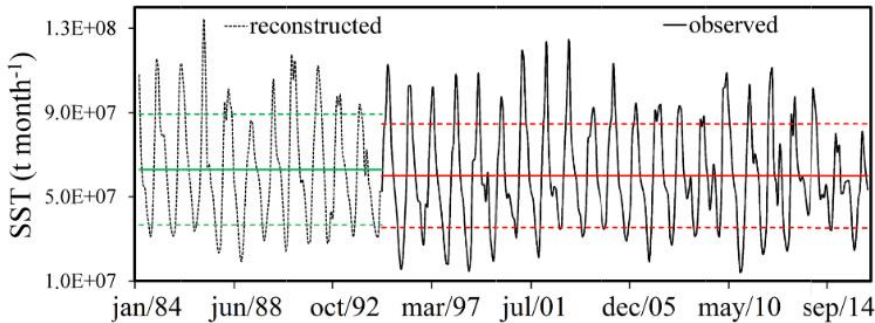
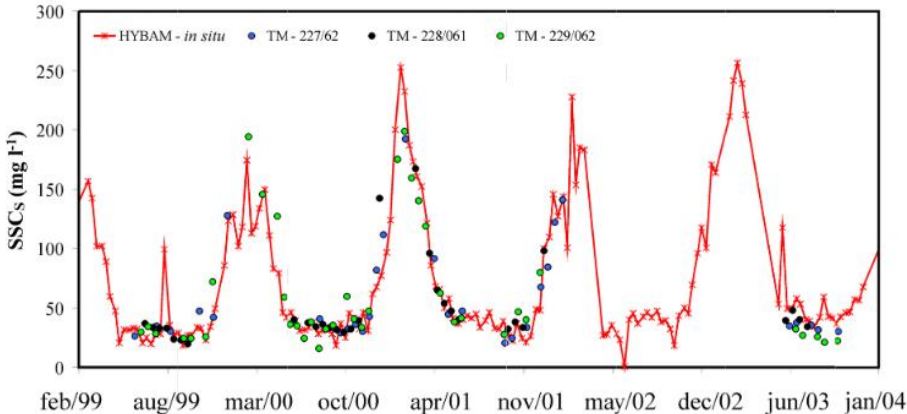
<sup>a</sup> Divisão de Sensoriamento Remoto, Instituto Nacional de Pesquisas Espaciais, 12201-970, São José dos Campos, SP, Brazil

<sup>b</sup> Departamento de Tecnologia, Universidade Estadual de Maringá, 87506-370, Umuarama, PR, Brazil

<sup>c</sup> Departamento de Geografia, Instituto de Geociências e Ciências Exatas (IGCE) – Universidade Estadual Paulista (UNESP), Rio Claro, SP, Brazil



# Sedimentos suspensos

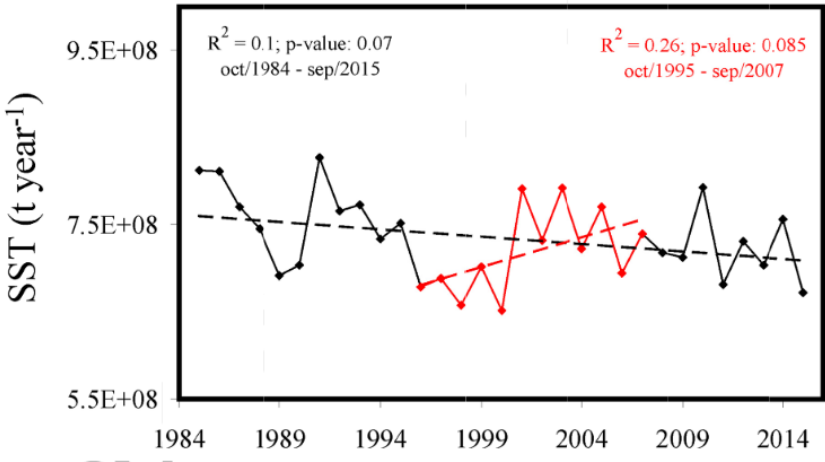


## Temporal trend of the suspended sediment transport of the Amazon River (1984–2016)

Otávio Cristiano Montanher, Evlyn Márcia Leão de Moraes Novo & Edvard Elias de Souza Filho

To cite this article: Otávio Cristiano Montanher, Evlyn Márcia Leão de Moraes Novo & Edvard Elias de Souza Filho (2018): Temporal trend of the suspended sediment transport of the Amazon River (1984–2016), Hydrological Sciences Journal, DOI: [10.1080/02626667.2018.1546387](https://doi.org/10.1080/02626667.2018.1546387)

To link to this article: <https://doi.org/10.1080/02626667.2018.1546387>



# Sedimentos suspensos



Contents lists available at ScienceDirect

Catena

journal homepage: [www.elsevier.com/locate/catena](http://www.elsevier.com/locate/catena)



Reversal of the sediment load increase in the Amazon basin influenced by divergent trends of sediment transport from the Solimões and Madeira Rivers

Tong Li, Shuai Wang\*, Yanxu Liu, Bojie Fu, Dexin Gao

State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China



Hydrological Sciences Journal

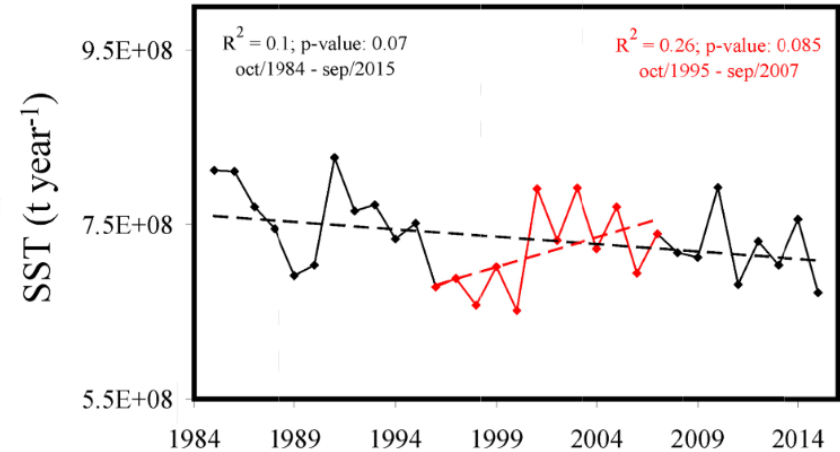
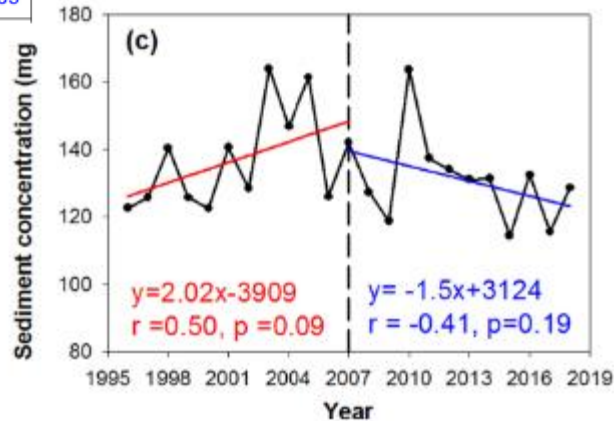
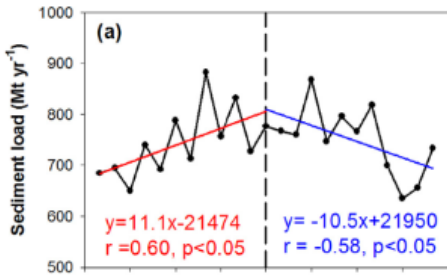
ISSN: 0262-6667 (Print) 2150-3435 (Online) Journal homepage: <http://www.tandfonline.com/loi/thsj20>

## Temporal trend of the suspended sediment transport of the Amazon River (1984–2016)

Otávio Cristiano Montanher, Evlyn Márcia Leão de Moraes Novo & Edvard Elias de Souza Filho

To cite this article: Otávio Cristiano Montanher, Evlyn Márcia Leão de Moraes Novo & Edvard Elias de Souza Filho (2018): Temporal trend of the suspended sediment transport of the Amazon River (1984–2016), Hydrological Sciences Journal, DOI: 10.1080/02626667.2018.1546387

To link to this article: <https://doi.org/10.1080/02626667.2018.1546387>



<https://doi.org/10.1080/02626667.2018.1546387>

# Sedimentos suspensos



## Estimating the suspended sediment concentration from TM/Landsat-5 images for the Araguaia River – Brazil

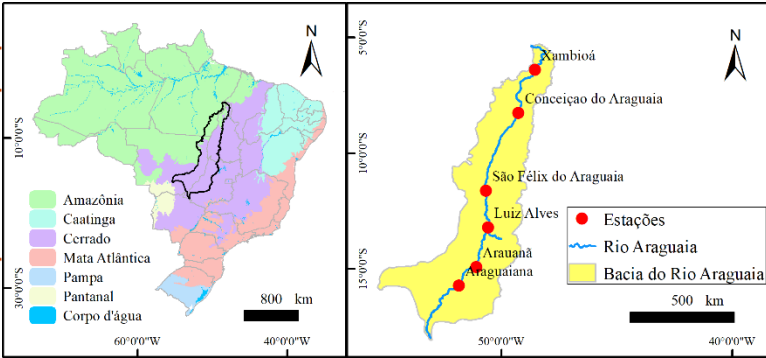
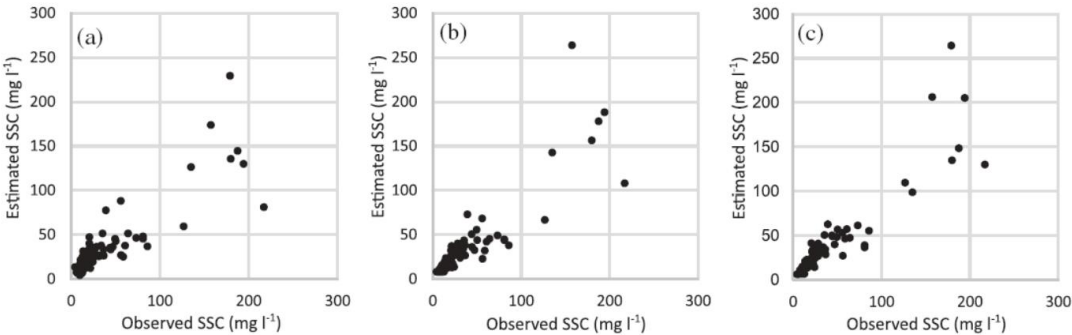
É. H. Cremon<sup>a</sup>, Ana Maria S. da Silva<sup>a</sup> and Otávio C. Montanher<sup>b</sup>

<sup>a</sup>Federal Institute of Goiás - Department of Geomatics, Goiânia, GO, Brazil; <sup>b</sup>State University of Maringá - Department of Geography/GEMA, Maringá, PR, Brazil

**Table 2.** Performance analysis of the models generated by regression analysis.

Approach		$R^2$ *	RMSE	NRMSE
Surface reflectance	Simple regression	0.59	39.20	15.47
	Multiple regression	0.60	39.09	15.43
ToA reflectance	Simple regression	0.74	29.48	12.59
	Multiple regression	0.82	26.78	13.86
	Multiple regression with categorical variable	0.87	21.46	10.09

\* For multiple regression, adjusted  $R^2$  was used.





# Sedimentos suspensos

Com dados MODIS



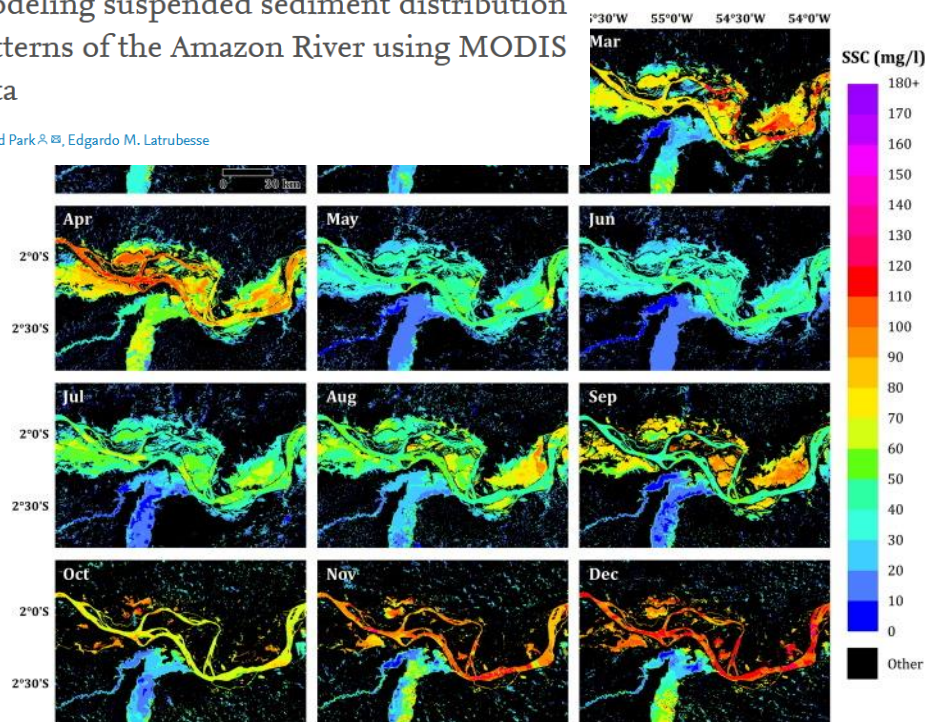
Remote Sensing of Environment

Volume 147, 5 May 2014, Pages 232–242



## Modeling suspended sediment distribution patterns of the Amazon River using MODIS data

Edward Park <sup>a,✉</sup>, Edgardo M. Latrubesse



Contents lists available at ScienceDirect

Comptes Rendus Geoscience

www.sciencedirect.com



Hydrology, Environment

## Purus River suspended sediment variability and contributions to the Amazon River from satellite data (2000–2015)

Andre Luis Martinelli Real dos Santos <sup>a,b,\*</sup>, Jean Michel Martinez <sup>c</sup>, Naziano Pantoja Filizola Jr. <sup>b,d</sup>, Elisa Armijos <sup>b,e</sup>, Luna Gripp Simões Alves <sup>a</sup>



International Journal of Applied Earth Observation and Geoinformation

Volume 21, April 2013, Pages 341–355



## Discharge and suspended sediment flux estimated along the mainstream of the Amazon and the Madeira Rivers (from *in situ* and MODIS Satellite Data)

S. Mangiarotti <sup>a,✉</sup>, J.-M. Martinez <sup>b,c</sup>, M.-P. Bonnet <sup>b,c</sup>, D.C. Buarque <sup>d</sup>, N. Filizola <sup>a</sup>, P. Mazzega <sup>b,c</sup>



Journal of South American Earth Sciences

Volume 44, July 2013, Pages 45–54



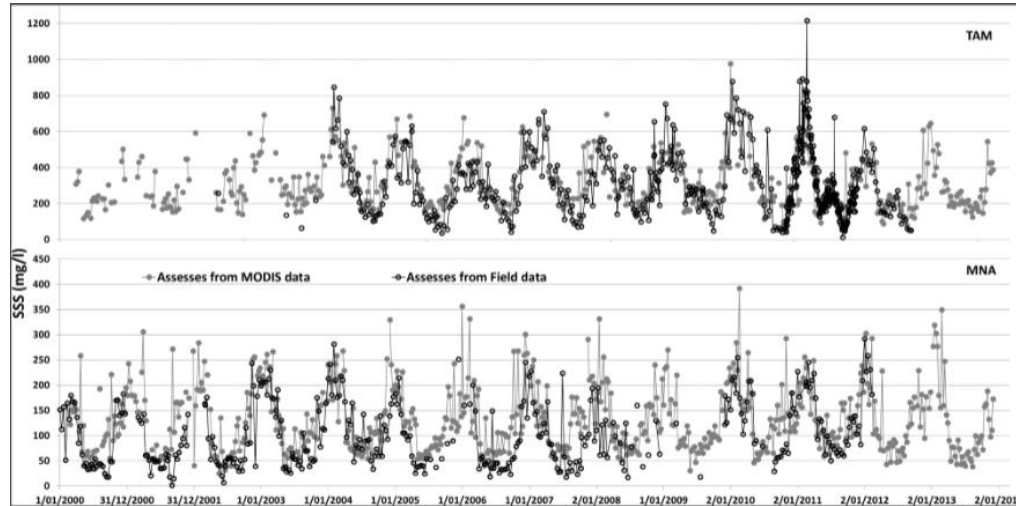
## A study of sediment transport in the Madeira River, Brazil, using MODIS remote-sensing images

Raúl Espinoza Villar <sup>a,✉</sup>, Jean-Michel Martinez <sup>a,b,c,\*</sup>, Marie Le Texier <sup>c,d</sup>, Jean-Loup Guyot <sup>b,c</sup>, Pascal Fraizy <sup>b,c</sup>, Paulo Roberto Meneses <sup>a</sup>, Eurides de Oliveira <sup>a</sup>

<https://doi.org/10.1080/02626667.2018.1546387>

# Sedimentos suspensos

## Com dados MODIS



Espinoza-Villar et al. (2018)



Hydrology, Environment

Purus River suspended sediment variability and contributions to the Amazon River from satellite data (2000–2015)

Andre Luis Martinelli Real dos Santos<sup>a,b,\*</sup>, Jean Michel Martinez<sup>c</sup>, Naziano Pantoja Filizola Jr.<sup>b,d</sup>, Elisa Armijos<sup>b,e</sup>, Luna Gripp Simões Alves<sup>a</sup>



International Journal of Applied Earth Observation and Geoinformation

Volume 21, April 2013, Pages 341–355



Discharge and suspended sediment flux estimated along the mainstream of the Amazon and the Madeira Rivers (from *in situ* and MODIS Satellite Data)

S. Mangiarotti<sup>a,\*,</sup>, J.-M. Martinez<sup>b,\*,</sup>, M.-P. Bonnet<sup>b,\*,</sup>, D.C. Buarque<sup>d</sup>, N. Filizola<sup>a</sup>, P. Mazzega<sup>b,\*,</sup>



Journal of South American Earth Sciences

Volume 44, July 2013, Pages 45–54



A study of sediment transport in the Madeira River, Brazil, using MODIS remote-sensing images

Raúl Espinoza Villar<sup>a,\*,</sup>, Jean-Michel Martinez<sup>a,\*,</sup>, Marie Le Texier<sup>c,\*,</sup>, Jean-Loup Guyot<sup>b,\*,</sup>, Pascal Fraizy<sup>b,\*,</sup>, Paulo Roberto Meneses<sup>a</sup>, Eurides de Oliveira<sup>a</sup>



# QUALIDADE DE ÁGUA



## FILTROS

### Corpo Hídrico

72

### Quantidade de Estações

#### Município

#### UF

- ☐ RIACHO FUNDO
- ☐ RIO ACAUÃ
- ☐ RIO APODÍ-MOSSORÓ
- ☐ RIO ARACOIABA
- ☐ RIO ARAGUAIA
- ☐ RIO BANABUIU
- ☐ RIO CANINDÉ
- ☐ RIO CHORO
- ☐ RIO ITARARÉ
- ☐ RIO JAGUARIBE
- ☐ RIO MADEIRA
- ☐ RIO PACOTI
- ☐ RIO PARAGUAI
- ☐ RIO PARAÍBA
- ☐ RIO PARANAPANEMA
- ☐ RIO PARAPEBA
- ☐ RIO PIANÓ

#### Estação

- ☒ Açude Aracoiaba Virtual
- ☒ Açude Armando Ribeiro Gonçalves Virtual
- ☒ Açude Banabuiu Virtual
- ☒ Açude Boqueirão das Parelhas Virtual
- ☒ Açude Castanhão Virtual
- ☒ Açude Coremas Virtual
- ☒ Açude Engenheiros Ávidos Virtual
- ☒ Açude Eptácio Pessoa Virtual
- ☒ Açude Fogareiro Virtual
- ☒ Açude Gargalheiras Virtual
- ☒ Açude Mãe D'água Virtual
- ☒ Açude Orós Virtual
- ☒ Açude Pacoti Virtual

- ☐ ACARI
- ☐ AÇU
- ☐ ALMEIRIM
- ☐ ALVORADA DO SUL
- ☐ APODI
- ☐ ARACOIABA
- ☐ BANABUIU
- ☐ BOQUEIRÃO
- ☐ BURITIZERO
- ☐ CÁCERES
- ☐ CARINHANHA
- ☐ CHAVANTES
- ☐ CHORO
- ☐ COARI
- ☐ CONCEIÇÃO DO ARAGUAIA
- ☐ COREMAS
- ☐ CORUMBÁ
- ☐ EUCLIDES DA CUNHA PAULISTA
- ☐ FARTURA
- ☐ FELIXLÂNDIA
- ☐ FLORÍNIA
- ☐ FONTE BOA
- ☐ HUMAITÁ

- ☐ AM
- ☐ BA
- ☐ CE
- ☐ MG
- ☐ MS
- ☐ MT
- ☐ PA
- ☐ PB
- ☐ PR
- ☐ PU
- ☐ RN
- ☐ RO
- ☐ SP
- ☐ TO

#### Parâmetro

- ☐ Clorofila
- ☐ Sedimentos
- ☐ Turbidez

**UNIDADES**  
**Clorofila:** µg/L  
**Sedimentos:**  
mg/L **Turbidez:**  
NTU



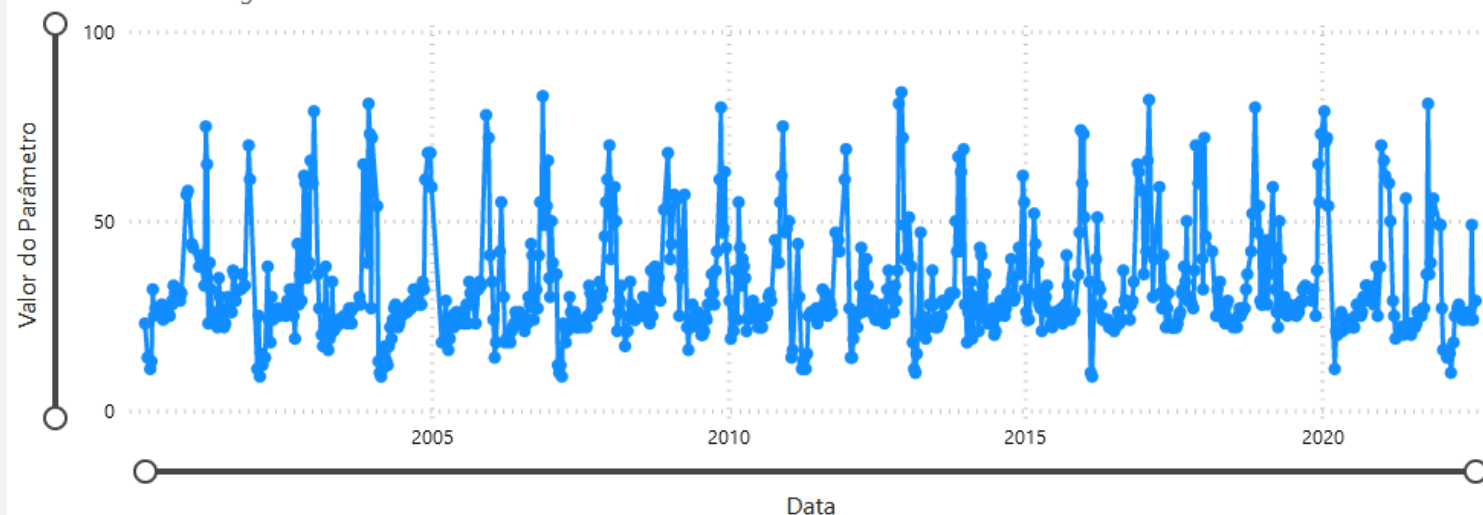
# SÉRIES TEMPORAIS ESTAÇÕES DE QUALIDADE

1

Quant. Estações

UNIDADES: Clorofila:  $\mu\text{g/L}$  Sedimentos:  $\text{mg/L}$  Turbidez: NTU

São Félix do Araguaia Virtual



Código	Data	Valor
1137S05039W0	05/08/2022	29
1137S05039W0	20/07/2022	24
1137S05039W0	12/07/2022	23
1137S05039W0	12/07/2022	26
1137S05039W0	04/07/2022	24
1137S05039W0	26/06/2022	26
1137S05039W0	18/06/2022	25
1137S05039W0	10/06/2022	25
1137S05039W0	02/06/2022	24
1137S05039W0	25/05/2022	26
1137S05039W0	17/05/2022	24
1137S05039W0	02/05/2022	25
1137S05039W0	23/04/2022	28
1137S05039W0	15/04/2022	27
1137S05039W0	07/04/2022	25
1137S05039W0	30/03/2022	25
1137S05039W0	22/03/2022	18

DATA

24/02/2000 29/08/2022

CORPO HÍDRICO

☐ RIO ARAGUAIA

MUNICÍPIO

☐ SÃO FÉLIX DO ARAGUAIA

CÓDIGO

☐ 1137S05039W0

NOME

- ☐ Porto velho virtual
- ☐ Santo Antônio do Içá Montante Virtual
- ☒ São Félix do Araguaia Virtual
- ☐ São Francisco Virtual
- ☐ Tabatinga Virtual
- ☐ Tamshiyacu Virtual
- ☐ Tafá Lucante Virtual

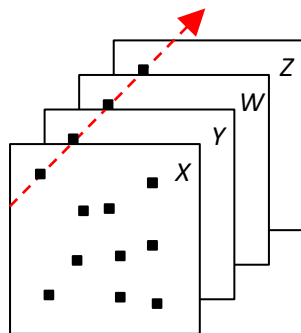
PARÂMETRO

☐ Sedimentos



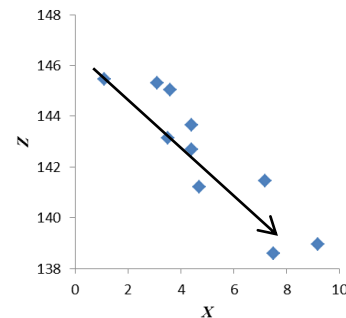
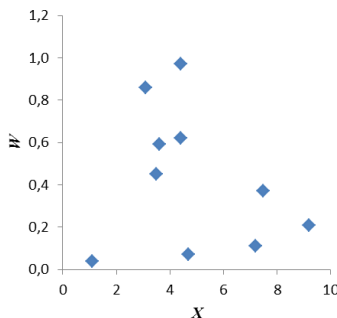
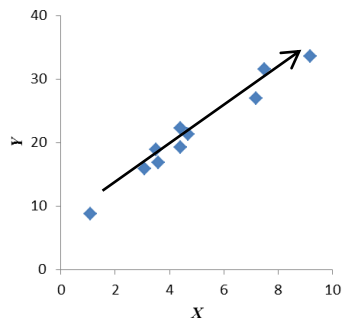
# Relacionamento entre Variáveis

Em imagens ou mapas, o relacionamento aparece pela posição geográfica

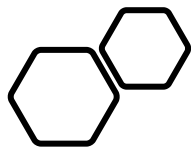


Amostra	X	Y	W	Z
1	3,5	18,9	0,45	143,2
2	7,5	31,5	0,37	138,6
3	4,4	22,2	0,62	142,7
4	1,1	8,7	0,04	145,5
5	4,4	19,2	0,97	143,7
6	4,7	21,3	0,07	141,2
7	7,2	27,0	0,11	141,5
8	3,6	16,8	0,59	145,1
9	9,2	33,6	0,21	139,0
10	3,1	15,9	0,86	145,3

← mesma posição geográfica



Muitos estudos buscam entender as relações de dependência entre variáveis de modo a construir modelos que permitam prever o comportamento de uma variável conhecendo-se os valores de outra ou outras variáveis



# Tipo de Regressão

## Tipos de Regressão

Regressão Linear

Regressão Polinomial

Regressão SVM

Regressão por Árvore de Decisão

Regressão por ensemble (ex.: *Random Forest*)

Regressão *Ridge*

Regressão Lasso

Regressão Logística



## Regressão Linear Simples

- Apenas duas variáveis são usadas, uma dependente ( $Y$ ) e uma independente ( $X$ )

# Terminologia

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- $X$  é a variável **independente**, também chamada de **preditor**
- $Y$  é a variável **dependente**, também chamada de **variável resposta**
- $\beta_0$  e  $\beta_1$  são os **parâmetros** ou **coeficientes** da regressão
- $\varepsilon$  é o **termo de erro**

# Exemplos de Regressão Linear Simples

$$Y = \beta_0 + \beta_1 X$$

$$Produtividade_{ton/ha} = -3,719 + 10,08.NDVI$$

$$\ln(CSS)_{mg/l} = -0,75 + 4,38 \cdot \frac{B3}{B2}$$

$$\ln(Biomassa_{total}) = -1,96722 + 1,99320 \cdot \ln(DAP)$$

$$CSS = -159.3 + 254.1 (B3/B1)$$



# Regressão Linear Simples

- Os modelos de regressão são formados a partir de combinações lineares de variáveis, através de parâmetros
- $Y = \beta_0 + \beta_1 X$  é uma combinação linear de um parâmetro linear  $\beta_0$  e um parâmetro linear  $\beta_1$  que multiplica  $X$
- $Y = \beta_0 + \beta_1 X + \beta_2 X^2$  também é uma combinação linear, de
  - um parâmetro linear  $\beta_0$ , um parâmetro linear  $\beta_1$  que multiplica  $X$ , e um parâmetro linear  $\beta_2$  que multiplica  $X^2$

# Regressão Linear Simples

$Y = \beta_0 + \beta_1/\beta_2 X$  não é uma combinação linear

$Y = \beta_0 + e^{\beta_1 X}$  não é uma combinação linear

A linearidade se refere aos parâmetros, e não às variáveis

# Características de um modelo de regressão

Os modelos de regressão expressam essencialmente:

- Uma tendência de  $Y$  em variar sistematicamente com o preditor  $X$
- Uma dispersão de pontos ao redor da curva que descreve uma relação estatística

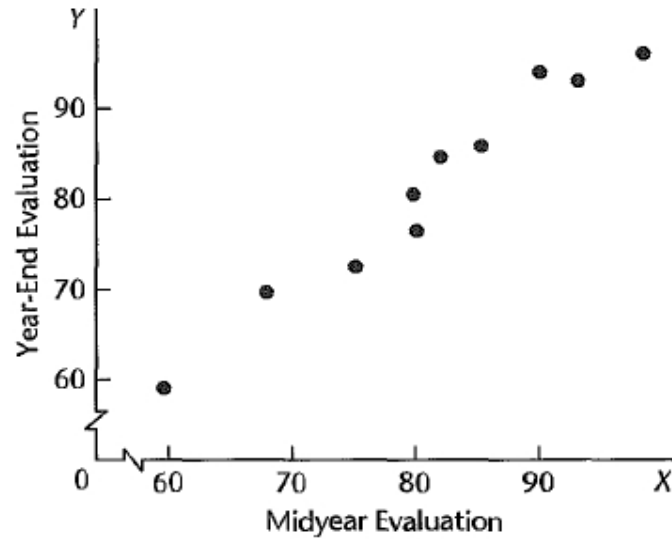
Essas características são expressas através das pressuposições:

- Existe uma distribuição de probabilidade de  $Y$  para cada nível (valor) de  $X$
- A média destas distribuições varia sistematicamente com  $X$

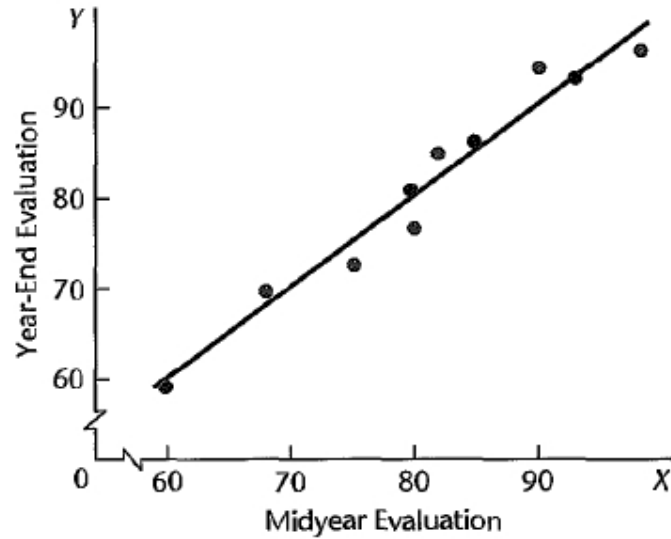


# Exemplo

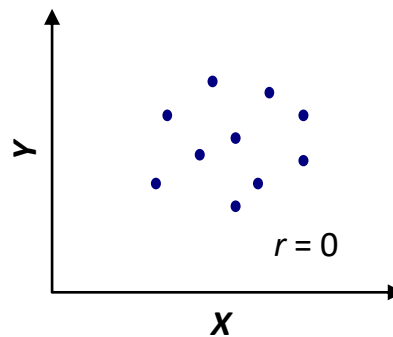
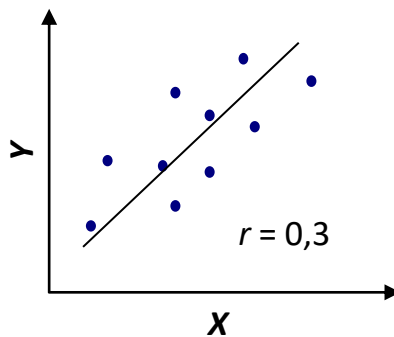
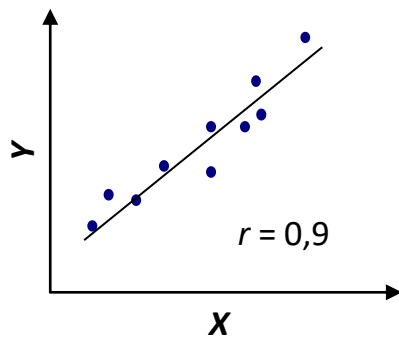
Scatter Plot



Scatter Plot and Line of Statistical Relationship

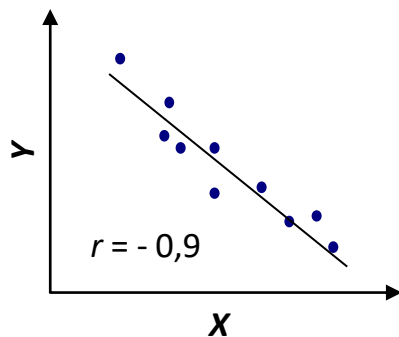


# Coeficiente de Correlação (r)



Coeficiente de Correlação (de Pearson)

mede o grau de relação **linear** entre X e Y



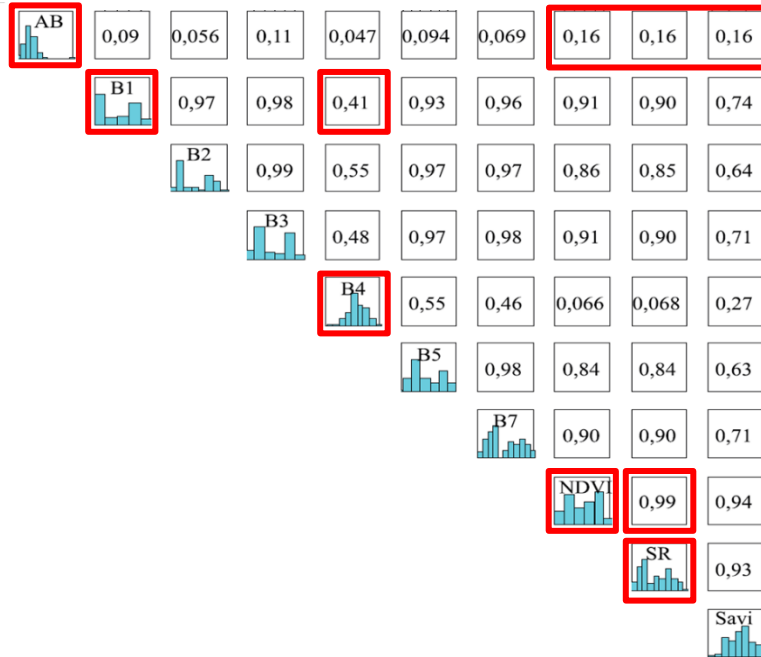
$$r = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

$$-1 \leq r \leq 1$$

$$r = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{X})^2 \sum_{i=1}^n (y_i - \bar{Y})^2}} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\left[ n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \right] \left[ n \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2 \right]}}$$

# Coeficiente de Correlação

Correlações (em módulo) entre características dendrométricas da Caatinga brasileira e dados TM Landsat 5 (Almeida et al., 2014\*)



AB - área basal

B1 a B7 – bandas do TM/Landsat

$NDVI = (B4 - B3) / (B4 + B3)$

$SR = B4 / B3$

$Savi = 0,5(B4 - B3) / (B4 + B3 + 0,5)$

É fundamental analisar o **gráfico de dispersão** para verificar se a relação é linear e os pontos estão bem distribuídos!

\*Fonte: <http://www.scielo.br/pdf/pab/v49n4/0100-204X-pab-49-04-0306.pdf>

# Coeficiente de Correlação

Correlações (em módulo) entre características dendrométricas da Caatinga brasileira e dados TM Landsat 5 (Almeida et al., 2014\*)



AB - área basal

B1 a B7 – bandas do TM/Landsat

$NDVI = (B4 - B3) / (B4 + B3)$

$SR = B4 / B3$

$Savi = 0,5(B4 - B3) / (B4 + B3 + 0,5)$

É fundamental analisar o

**gráfico de dispersão** para

verificar se a relação é

linear e os pontos estão

bem distribuídos!

O coeficiente de correlação

nem sempre representa

bem a relação entre

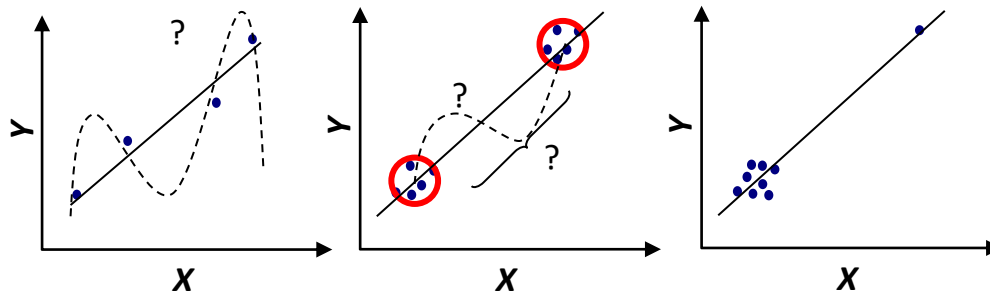
variáveis!

\*Fonte: <http://www.scielo.br/pdf/pab/v49n4/0100-204X-pab-49-04-0306.pdf>

# Coeficiente de Correlação

## Interpretações errôneas do coeficiente de correlação

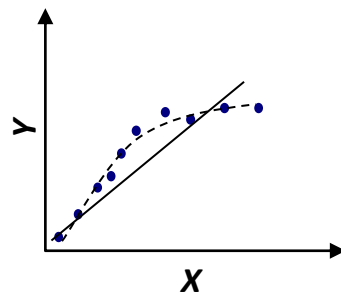
- Um alto coeficiente de correlação nem sempre indica que a equação de regressão estimada está bem ajustada aos dados.



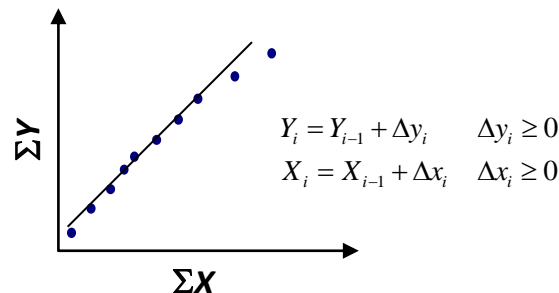
Poucos pontos

Grupos de pontos

Ponto isolado



Relação quase linear

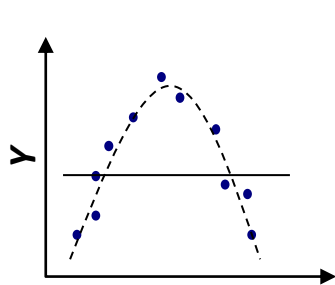


Variáveis cumulativas

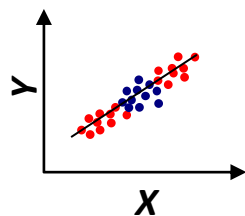
# Coeficiente de Correlação

## Interpretações errôneas do coeficiente de correlação

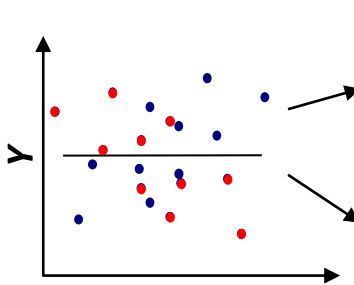
- Um coeficiente de correlação próximo de zero nem sempre indica que  $X$  e  $Y$  não são relacionadas.



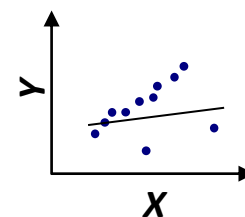
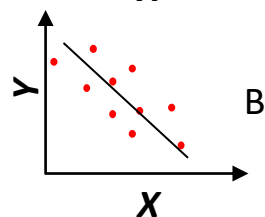
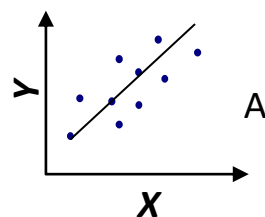
Relação não linear



Amostragem não representativa



Mistura de grupos  
com relações  
diferentes



Presença de outliers

# Coeficiente de Determinação ( $r^2$ )

A partir da formulação abaixo temos o Coeficiente de Determinação ( $r^2$ ):

$$r^2 = \frac{SQR}{SQT} = \frac{SQR}{SQR + SQE}$$

SQT = Soma dos Quadrados Totais

SQR = Soma dos Quadrados da Regressão

SQE = Soma dos Quadrados dos Erros



# Coeficiente de Determinação ( $r^2$ )

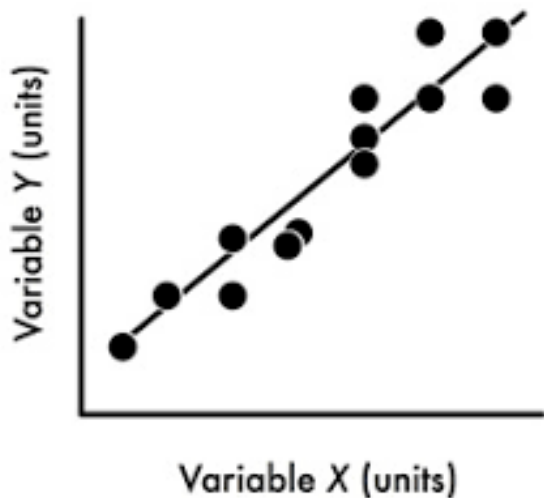
Como podemos interpretar o valor de  $r^2$ ?



$R^2$  nos dá a proporção da  
variância total de Y  
explicada pelo modelo de  
regressão

# Análise de Regressão

“Método estatístico que utiliza a relação entre duas ou mais variáveis para que uma variável possa ser estimada (ou predita) a partir da outra ou das outras”



A existência de uma relação estatística entre a variável **dependente Y** e a variável **independente X** não implica que Y realmente dependa de X, ou que exista uma relação de causa-efeito entre X e Y.

Neter, J. et al. **Applied Linear Statistical Models**. McGraw Hill, 1996.

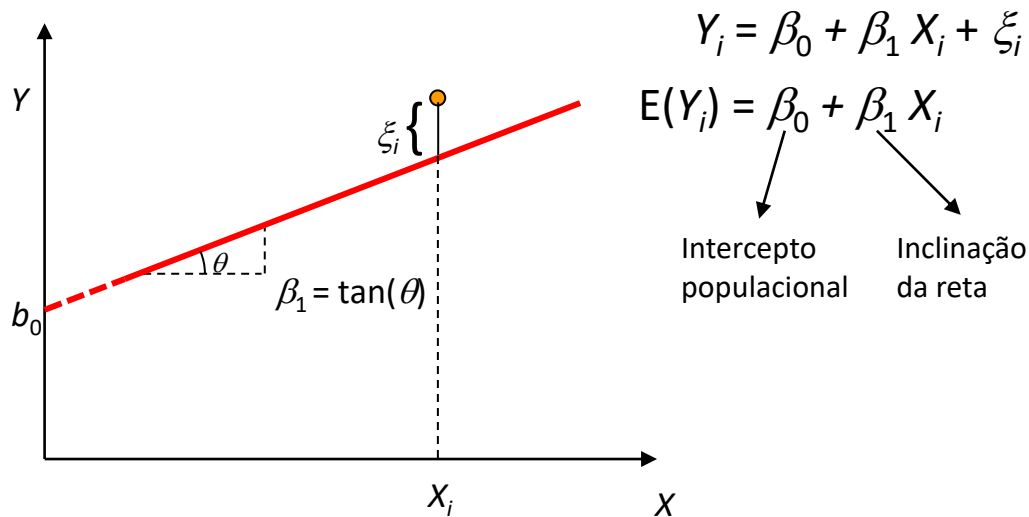
# Análise de Regressão

Para que serve uma análise de regressão?

- Encontrar as variáveis mais relevantes que se relacionam com a variável dependente ( $Y$ )
- Encontrar a função que descreve como uma ou mais variáveis se relacionam com a variável dependente ( $Y$ ) e estimar os parâmetros que definem esta função (equação ajustada)
- Usar a equação ajustada para prever valores da variável dependente ( $Y$ )

## Regressão Linear Simples

# Regressão Linear Simples



$\beta_0$  representa o valor de  $E(Y_i)$  quando  $X_i = 0$

$\beta_1$  é o coeficiente angular da reta e representa o aumento em  $E(Y_i)$  quando  $X_i$  é incrementado em uma unidade

# Regressão Linear Simples

Como ajustar a reta na regressão?

Em geral não se conhece os valores de  $\beta_0$  e  $\beta_1$

Eles podem ser estimados através de dados obtidos por amostras

O método comumente utilizado na estimação dos parâmetros é o **método dos mínimos quadrados**, o qual considera os desvios quadráticos dos  $Y_i$  em relação a seu valor esperado:

$$\xi_i = Y_i - E(Y_i) \qquad \xi_i = Y_i - (\beta_0 + \beta_1 X_i)$$

Em particular, o método dos mínimos quadrados requer que consideremos a soma de  $n$  desvios quadrados, denotado por  $Q$ :

$$Q = \sum_{i=1}^n \xi_i^2 = \sum_{i=1}^n [Y_i - \beta_0 - \beta_1 X_i]^2$$

De acordo com o método dos mínimos quadrados, os estimadores de  $\beta_0$  e  $\beta_1$  são aqueles, denotados por  $b_0$  e  $b_1$ , que tornam mínimo o valor de  $Q$ . Isso é feito derivando-se  $Q$  em relação a  $\beta_0$  e  $\beta_1$  e igualando-se as expressões encontradas a zero.

# Modelos Linearizáveis

**Modelo Padrão:**  $Y_i = \beta_0 + \beta_1 X_i + \xi_i$

exponencial

$$Y_i = \beta_0 e^{\beta_1 X_i} \xi_i$$

$$\ln Y_i = \ln \beta_0 + \beta_1 X_i + \ln \xi_i$$

$$Y'_i = \beta'_0 + \beta_1 X_i + \xi'_i$$

$$\xi'_i \sim N(0, \sigma^2)$$

potencial

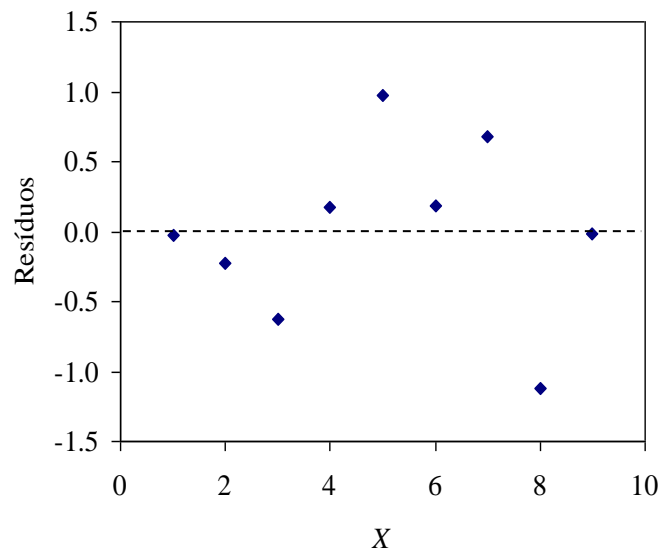
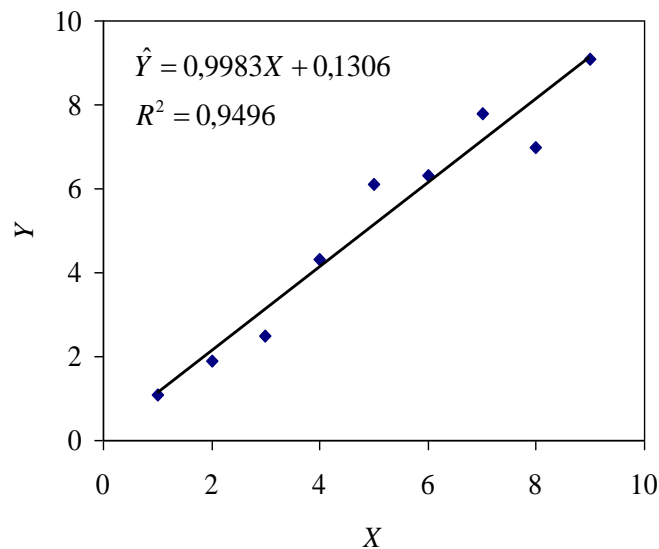
$$Y_i = \beta_0 X_i^{\beta_1} \xi_i$$

$$\ln Y_i = \ln \beta_0 + \beta_1 \ln X_i + \ln \xi_i$$

$$Y'_i = \beta'_0 + \beta_1 X'_i + \xi'_i$$

$$Y'_i = \beta_0 + \beta_1 X'_i + \xi_i \quad \left\{ \begin{array}{l} \text{logaritmo} \\ \text{potência} \\ \text{inverso} \end{array} \right.$$

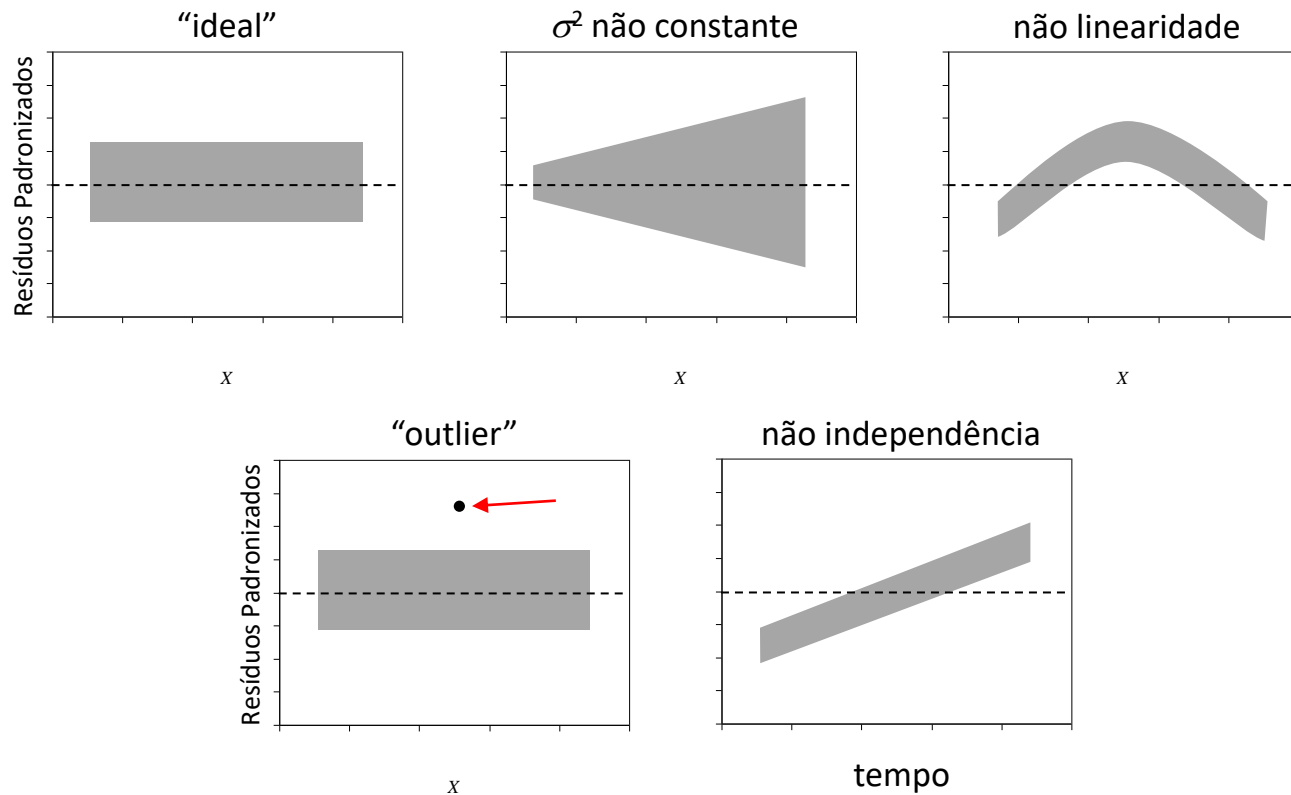
# Análise de Resíduos



$$\text{Resíduo} = e_i = Y_i - \hat{Y}_i$$



# Análise de Resíduos



# Regressão Linear Múltipla

## Modelo Geral

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_{p-1} X_{p-1,i} + \xi_i$$

$\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1}$  são parâmetros do modelo ( $p$  parâmetros no total)

$X_{1,i}, X_{2,i}, \dots, X_{p-1,i}$  são valores fixos conhecidos

$\xi_i$  são erros independentes  $\xi_i \sim N(0, \sigma^2)$

$i = 1, 2, \dots, n$

Fazendo  $X_{0,i} = 1$ , podemos reescrever o modelo como

$$\begin{aligned} Y_i &= \beta_0 X_{0,i} + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_{p-1} X_{p-1,i} + \xi_i \\ &= \sum_{k=0}^{p-1} \beta_k X_{k,i} + \xi_i \end{aligned}$$

# Pressupostos da regressão linear

1. a relação entre a variável dependente e as variáveis independentes deve ser linear;
2. as variáveis foram medidas adequadamente, ou seja, assume-se que não há erro sistemático de mensuração;
3. a expectativa da média do termo de erro (resíduos) é igual a zero;
4. homocedasticidade, ou seja, a variância do termo de erro é constante para os diferentes valores da variável independente;
5. ausência de autocorrelação, ou seja, os termos de erros são independentes entre si;
6. a variável independente não deve ser correlacionada com o termo de erro;
7. nenhuma variável teoricamente relevante para explicar Y foi deixada de fora do modelo e nenhuma variável irrelevante para explicar Y foi incluída no modelo;
8. as variáveis independentes não apresentam alta correlação, o chamado pressuposto da não multicolinearidade;
9. assume-se que o termo de erro tem uma distribuição normal; e
10. há uma adequada proporção entre o número de casos e o número de parâmetros estimados.

# Estimativa de concentração de sedimentos em suspensão por imagens de sensoriamento remoto

Prof. Dr. Édipo Henrique Cremon  
edipo.cremon@ifg.edu.br

