Introdução à Assimilação de Dados (MET 563-3)

Motivação - Equação de Análise Empírica

Dr. Carlos Frederico Bastarz
Dr. Dirceu Luis Herdies

Programa de Pós-Graduação em Meteorologia (PGMET) do INPE

22 de Setembro de 2025

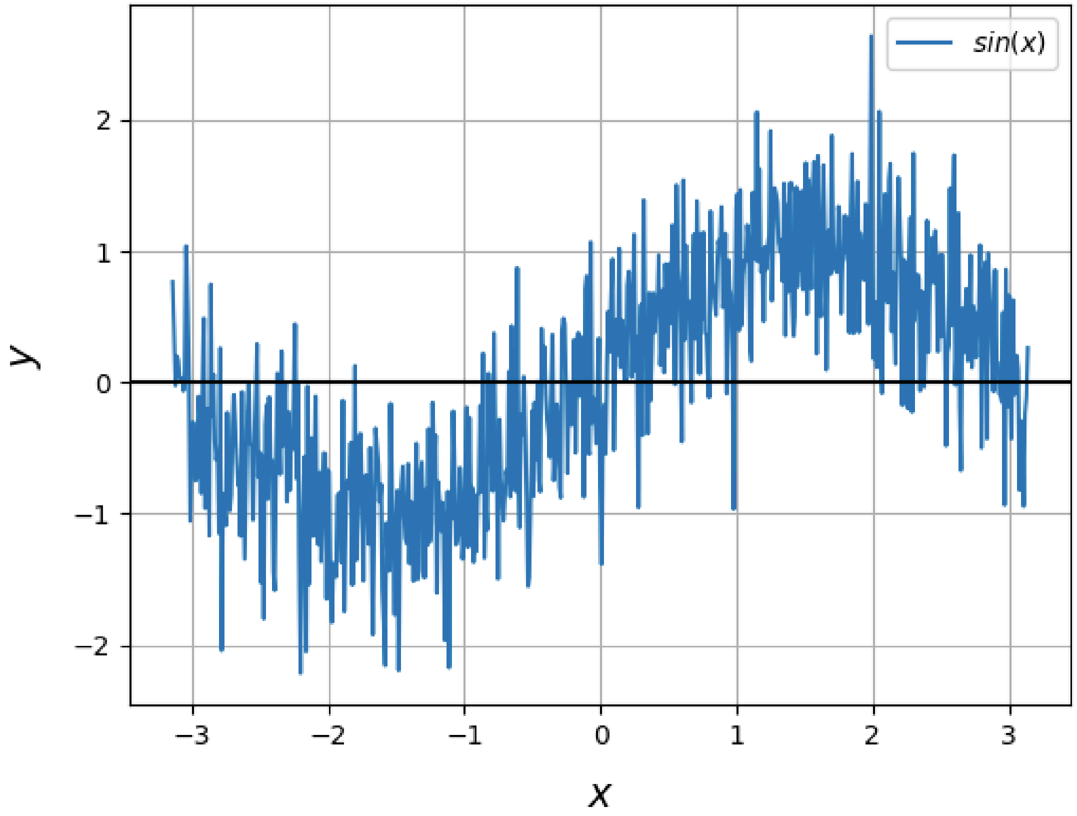
Equação de Análise Empírica

Motivação

- Esta é uma primeira aproximação para o problema de análise univariada
- Isto significa, na prática, que estamos tratando de um conjunto (ou uma série) de observações de uma mesma quantidade (e.g., temperatura)
- Considere um modelo matemático simples:

$$f(\mathbf{x}) = \sin(\mathbf{x}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2), \quad -\pi \leq x \leq \pi$$

- A função seno com a adição de um ruído normalmente distribuído



Equação de Análise Empírica

Motivação

- Suponha que possamos utilizar este modelo para ajustar uma curva produzida a partir de uma distribuição normal randômica utilizando uma **equação de análise empírica**:

$$\mathbf{x_a} = \alpha \mathbf{y_o} + (1 - \alpha)\mathbf{x_b}$$

- Onde:
  - $\mathbf{x_a}$: é o vetor análise
  - $\mathbf{x_b}$: é o vetor background
  - $\mathbf{y_o}$: é o vetor observação
  - α: é um peso escalar dado à observação e ao background

Equação de Análise Empírica

O modelo

- Já sabemos que o nosso modelo é a função seno. Então, vamos definir um domínio para a nossa função. Seja $\mathbf{x_0}$ um vetor com 629 elementos de 1 a 629:

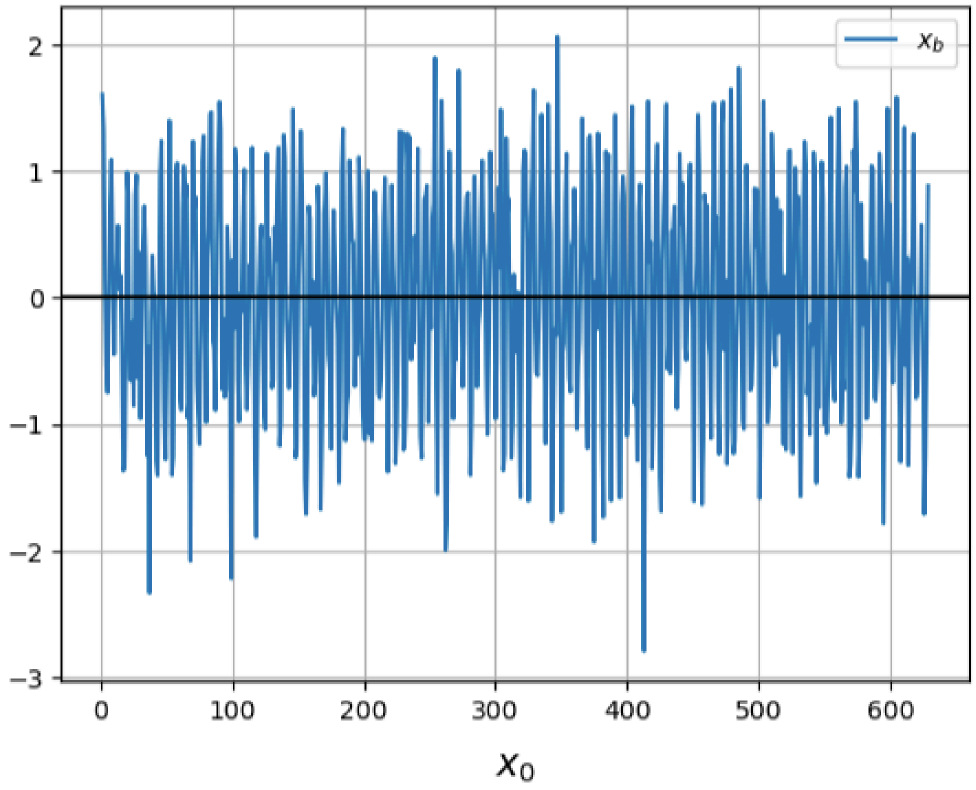
```
x0 = np.arange(1,630,1)
```

- Como nosso modelo é a função seno, vamos aplicar a função aos elementos do nosso domínio e vamos nomear de $\mathbf{x_b}$ o vetor com a imagem da nossa função, ou melhor, $\mathbf{x_b}$ é o nosso background:

```
xb = np.sin(x0) + ruido
```

Como é $\mathbf{x_b}$?

```
xb =
array([ 1.60624883e+00,  1.32688359e+00,  1.36485484e-01, -5.28018657e-01, ...
       -1.70776972e+00, -1.23210461e+00, -4.33982552e-01, 8.86349535e-01])
```



Equação de Análise Empírica

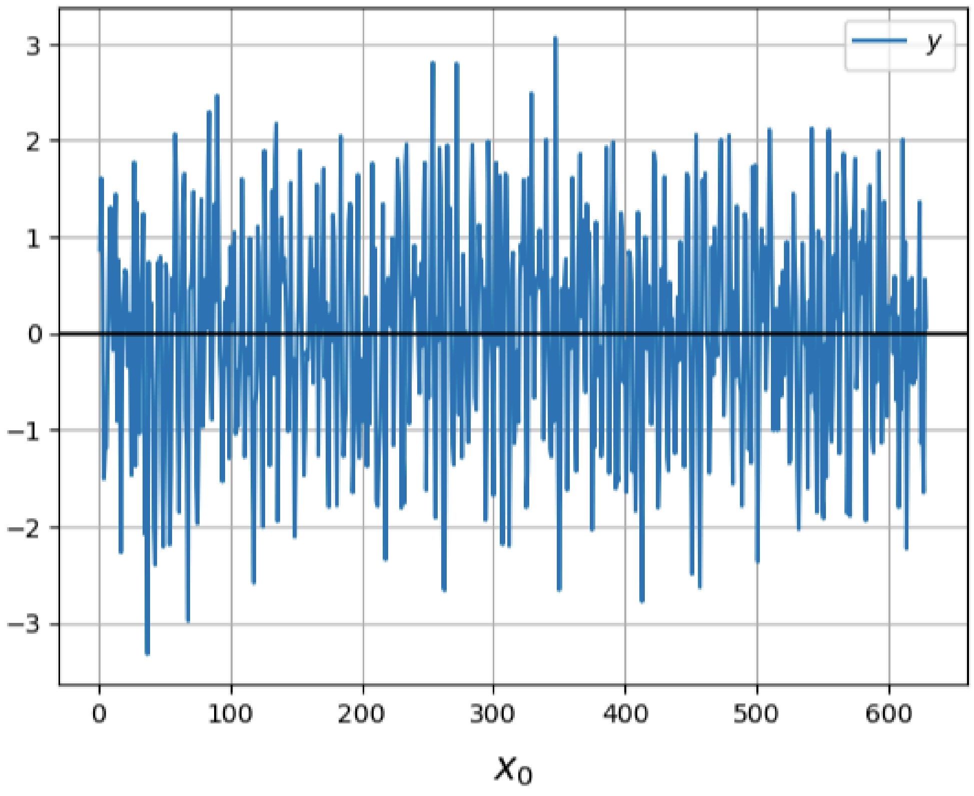
As observações

- O vetor observação $\mathbf{y_o}$, pode ser definido de forma semelhante ao vetor background $\mathbf{x_b}$:

```
mu_true = 0
sigma_true = 1
s = np.random.normal(mu_true, sigma_true, 629)
y = xb + np.sin(s)
```

Como é $\mathbf{y_o}$?

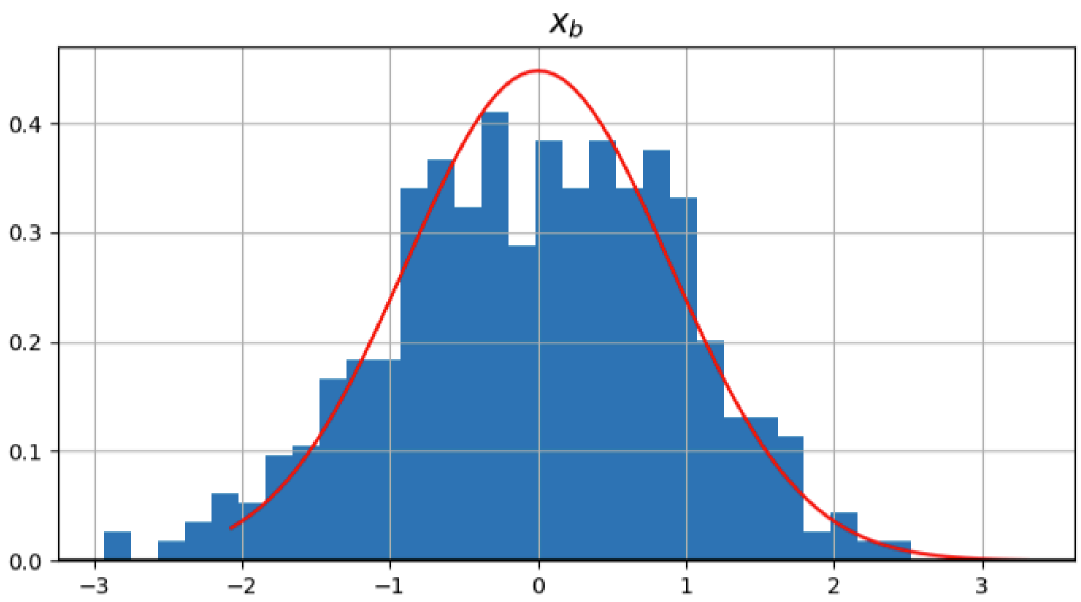
```
y =
array([ 8.63703635e-01,  1.61015360e+00,  5.87457172e-01, -1.50481750e+00, ...
       -9.50091116e-01, -1.64829028e+00,  5.60348293e-01, 5.84897816e-02])
```

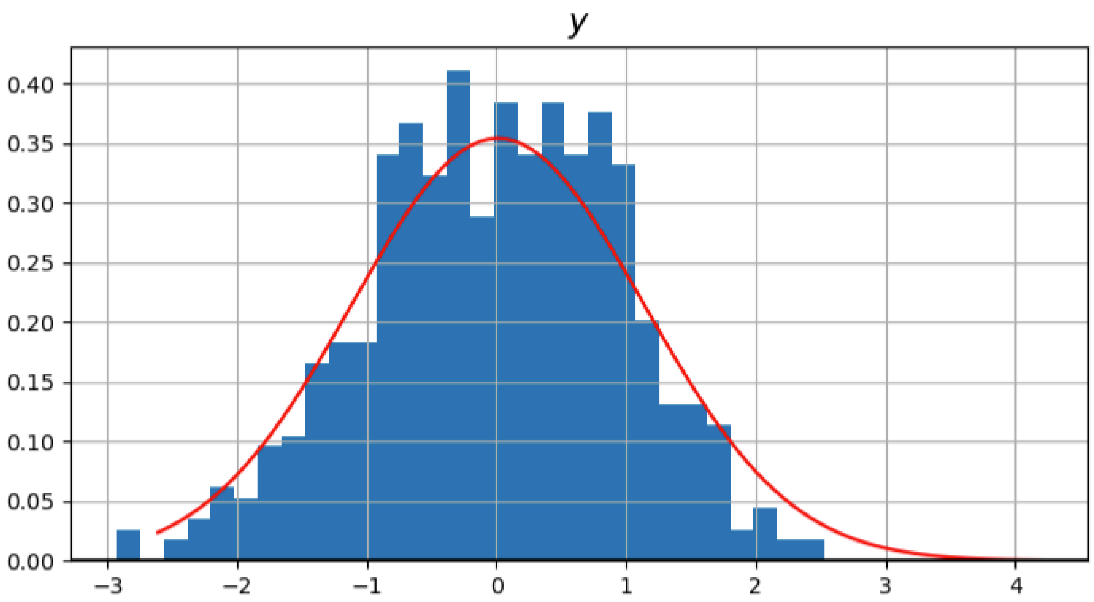


Equação de Análise Empírica

Distribuição Normal

- Observe que ambos, $\mathbf{x_b}$ e $\mathbf{y}$, possuem distribuição normal, isto é, ambos são representados por valores aleatórios distribuídos sobre uma curva normal com $\mu_{xb} = 0,0019$ e $\sigma_{xb} = 0,8909$ e $\mu_y = -0.011$ e $\sigma_y = 0.8563$:





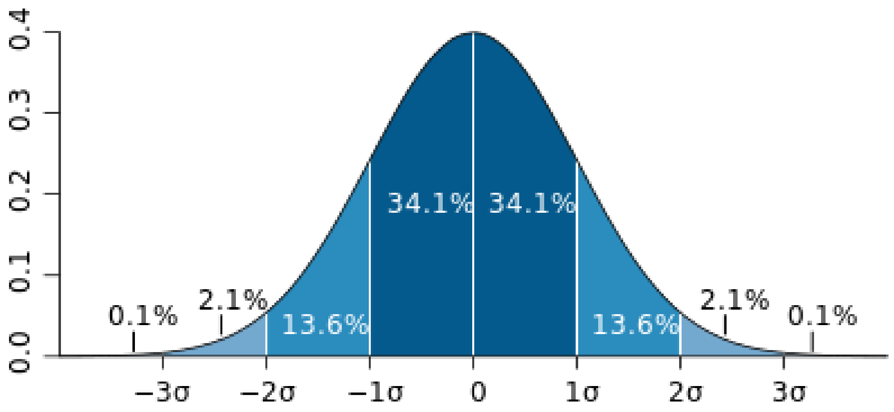
Equação de Análise Empírica

Distribuição Normal

- Estamos mantendo as distribuições de $\mathbf{x_b}$ e $\mathbf{y_o}$ próximas à distribuição normal, porque esta distribuição possui as seguintes propriedades:

$$f(\psi) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\psi-\mu)^2}{2\sigma^2}}$$

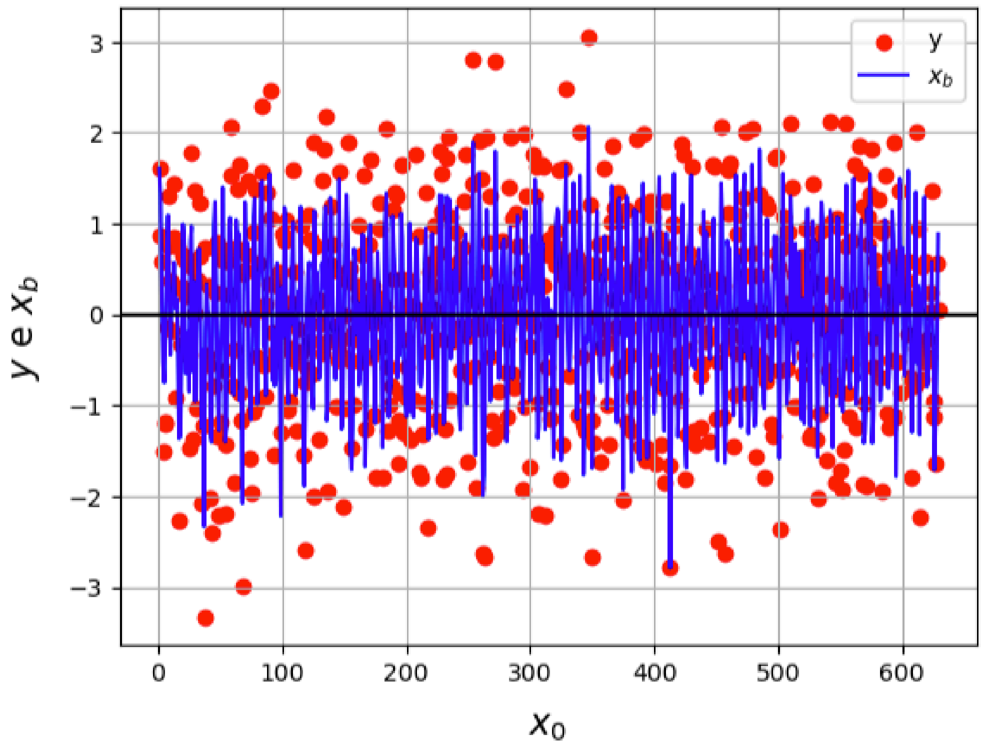
- ~68% dos valores encontram-se a uma distância da média inferior a um desvio-padrão
- ~95% dos valores encontram-se a uma distância da média inferior a duas vezes o desvio-padrão
- ~99,7% dos valores encontram-se a uma distância da média inferior a três vezes o desvio-padrão



Equação de Análise Empírica

Séries de $\mathbf{x_b}$ e $\mathbf{y_o}$

- Com $\mathbf{x_b}$ e $\mathbf{y_o}$ definidos, podemos plotar os seus elementos:



Equação de Análise Empírica

Equação de Análise Empírica

- Olhando para nossa equação de análise empírica, percebemos que os elementos $\mathbf{x_b}$ e $\mathbf{y_o}$ já estão definidos
- Ainda precisamos determinar o parâmetro α, que é o peso a ser atribuído às observações
- $1-\alpha$ é um outro peso que será atribuído ao background - **por que?**
- Uma vez determinado o valor de α, determinaremos $1-\alpha$ e, consequentemente, o valor de $\mathbf{x_a}$, o vetor análise (representado da mesma forma que $\mathbf{x_b}$ e $\mathbf{y_o}$):

$$\mathbf{x_a} = \alpha\mathbf{y_o} + (1-\alpha)\mathbf{x_b}$$

Equação de Análise Empírica

Determinação de α

- Antes de determinarmos o parâmetro α, precisamos saber o que ele é e como pode ser definido
- α é um parâmetro que relaciona as medidas das variâncias das parcelas:

$$\alpha = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$$

- Onde:
  - σ_b^2 e σ_o^2 são as variâncias do background e das observações
- Para calcular α, precisamos calcular as variâncias dos vetores $\mathbf{x_b}$ e $\mathbf{y_o}$

Equação de Análise Empírica

Erros $E(\mathbf{x_b})$ e $E(\mathbf{y_o})$

- A variância é uma medida de dispersão
- Ela pode ser calculada com base no erro da distribuição dos valores
- Vamos fazer as seguintes considerações:
  1. Não há relação entre os elementos dos dois vetores $\mathbf{x_b}$ e $\mathbf{y_o}$
  2. Os erros dos elementos dos vetores $\mathbf{x_b}$ e $\mathbf{y_o}$ são radômicos, ou seja, não há relação entre os erros dos elementos do vetor background e entre os elementos do vetor observação

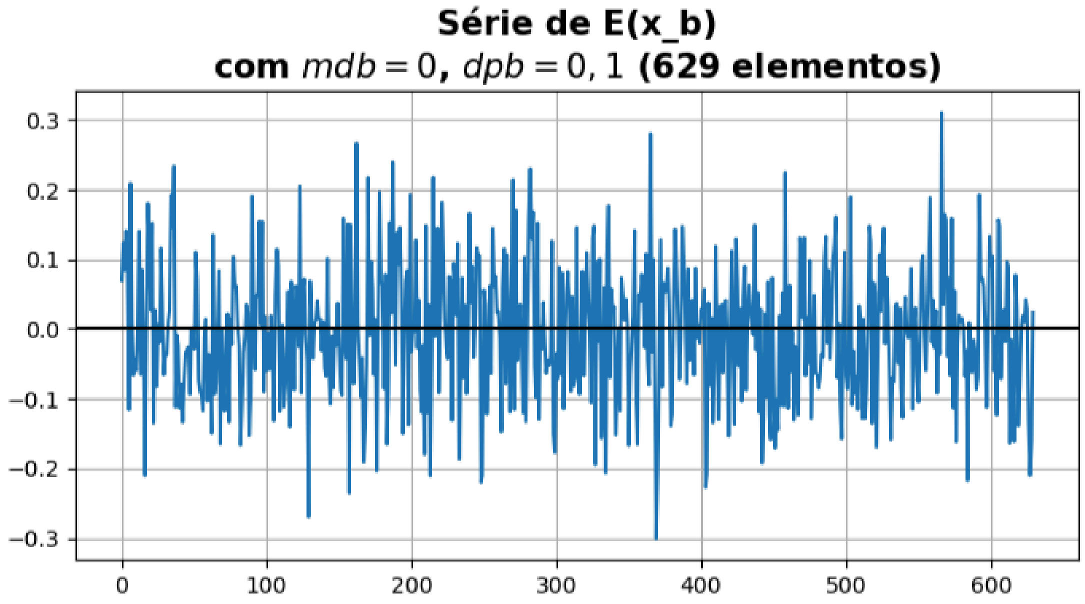
```
errb = mdb + dpb * np.random.randn(len(x0))
erro = mdo + dpo * np.random.randn(len(x0))
```

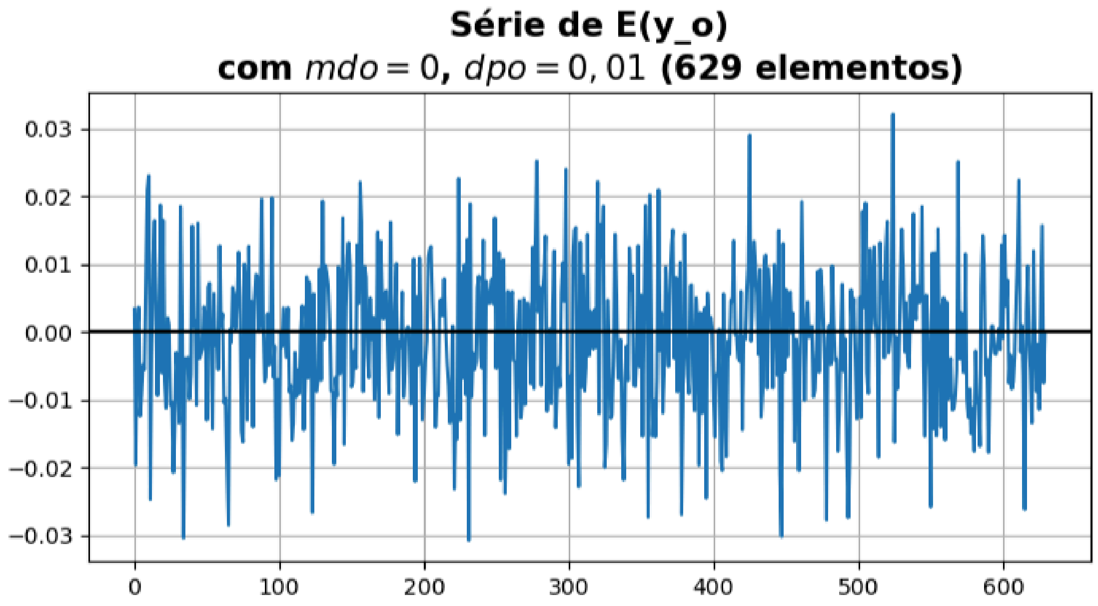
- Onde:
  - mdo: média da observação (μ_o)
  - mdb: média do background (μ_b)
  - dpo: desvio-padrão da observação (σ_o)
  - dpb: desvio-padrão do background (σ_b)

Equação de Análise Empírica

Testando alguns valores de σ_b e σ_o

- Exemplo da série dos erros de background $E(\mathbf{x_b})$ e observação $E(\mathbf{y_o})$:

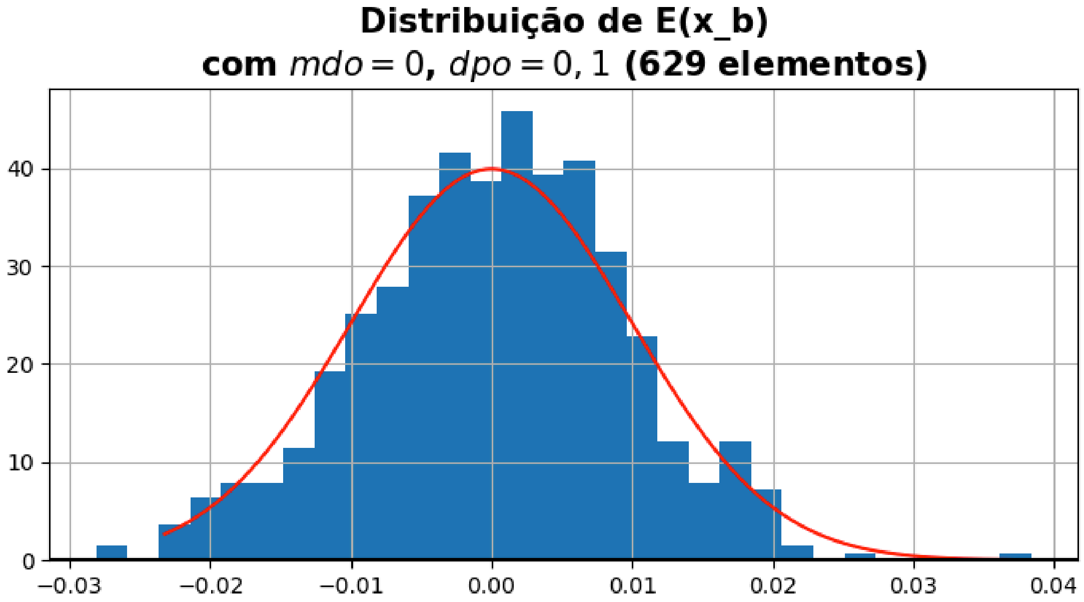


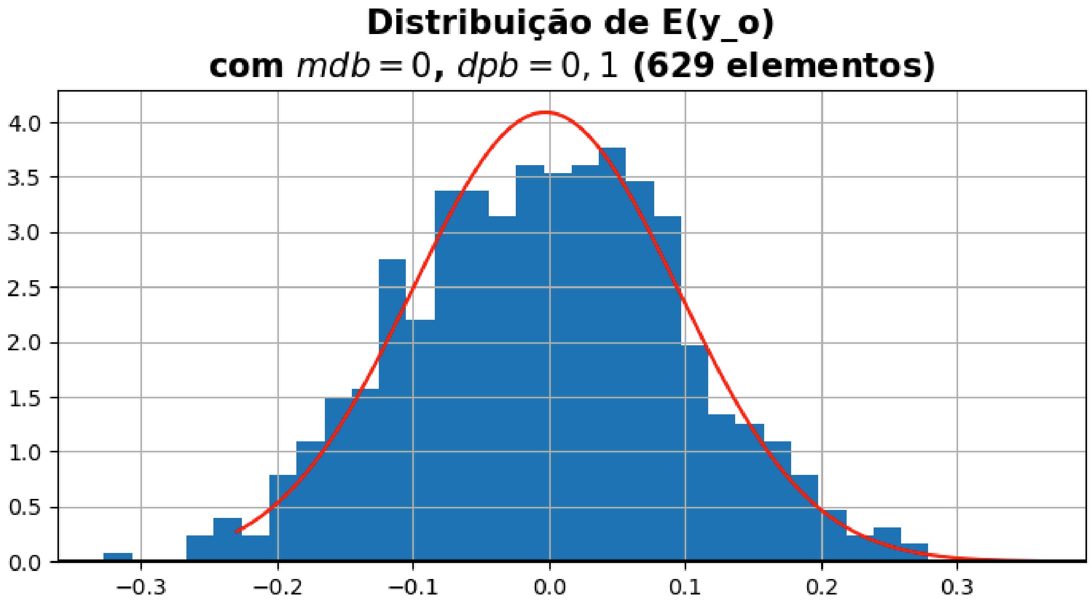


Equação de Análise Empírica

Testando alguns valores de σ_b e σ_o

- Exemplo da distribuição dos erros de background $E(\mathbf{x_b})$ e observação $E(\mathbf{y_o})$:





Equação de Análise Empírica

Variância dos erros σ_b^2 e σ_o^2

- Dado que α depende dos valores das variâncias das distribuições dos erros de background e observação, calculamos σ_b^2 e σ_o^2:

```
sigmab2 = np.var(errb)
sigmao2 = np.var(erro)
```

- Partindo-se dos valores das distribuições de $E(\mathbf{x_b})$ e $E(\mathbf{y_o})$, obtemos as seguintes variâncias:

```
sigmab2 = 0.0095226361060977
sigmao2 = 0.00011333207595536619
```

A variância dos erros de observação é muito menor do que a variância dos erros de background

Equação de Análise Empírica

Cálculo de α

- Com os valores de σ_b^2 e σ_o^2, o valor de α é calculado:

```
alpha = sigmab2 / (sigmab2 + sigmao2)
alpha = 0.9882386415340758
```

- Da equação de análise empírica, observamos que 99% do peso é dado para as observações, enquanto que 1% $(1-\alpha)$ de peso é dado para o background

$$\mathbf{x_a} = \alpha\mathbf{y_o} + (1-\alpha)\mathbf{x_b}$$

Equação de Análise Empírica

Cálculo de x_a

- Observando novamente a equação da análise, notamos que todos os parâmetros estão determinados:

$$\mathbf{x_a} = \alpha \mathbf{y_o} + (1 - \alpha)\mathbf{x_b}$$

- α: é um valor único ($\alpha \approx 0,99$)
- $\mathbf{y_o}$: é um vetor com valores "observados" de apenas uma grandeza (e.g., temperatura)
- $\mathbf{x_b}$: é um vetor com valores produzidos (calculados) por um modelo matemático (neste caso, a função seno adicionada de um ruído de distribuição próxima à Normal)

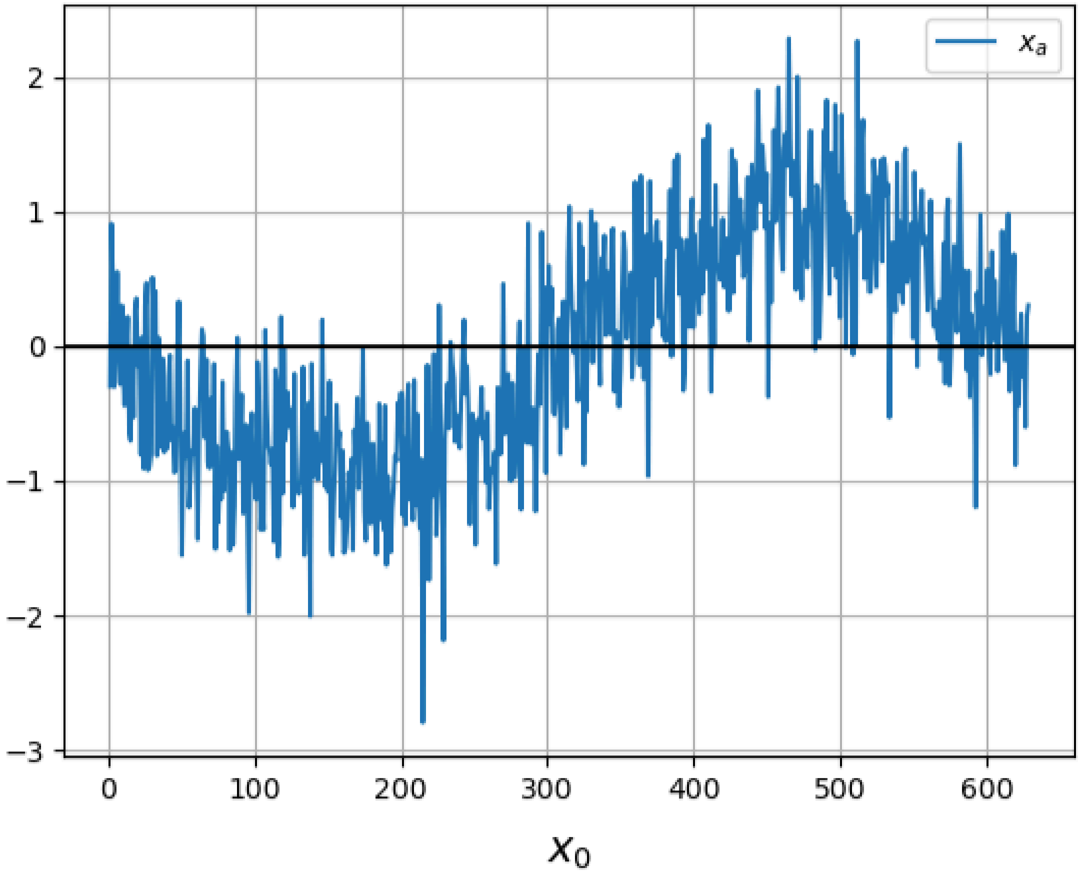
Equação de Análise Empírica

Cálculo de x_a

$$\mathbf{x_a} = \alpha \mathbf{y_o} + (1 - \alpha)\mathbf{x_b}$$

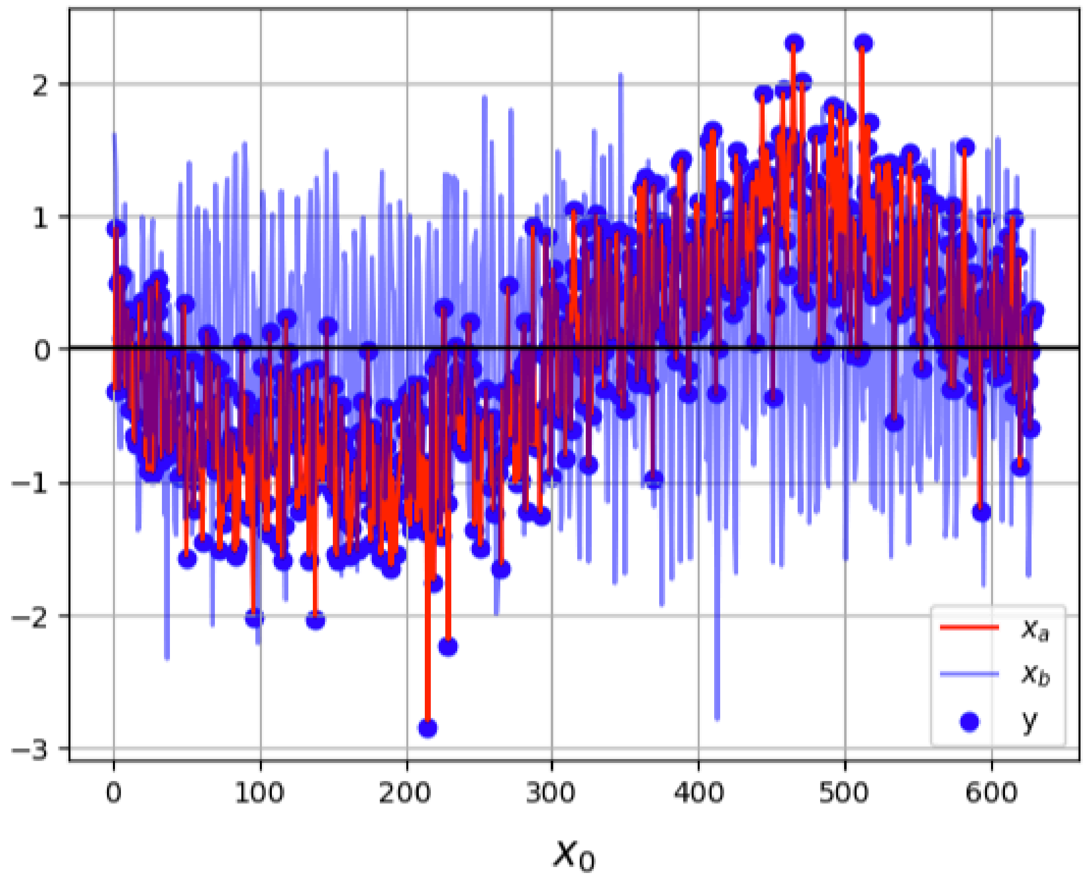
```
xa = alpha * y + (1 - alpha) * xb

xa =
array([-2.97708719e-01,  9.10349665e-01,  4.99888690e-01, -3.03402470e-01, ...
       -2.54933099e-02, -6.02816581e-01,  2.19151299e-01, 3.04442057e-01])
```



Equação de Análise Empírica

Plotando todos os resultados juntos



- Observe que a análise (curva vermelha) representa o ajuste do background (curva azul) às observações (pontos azuis)
- Quanto mais precisa a observação, melhor o ajuste

🎲 Notebook com [Atividade Prática 1]

🤔 Dúvidas

🔗 https://cfbastarz.github.io/met563-3/
🐙 https://github.com/cfbastarz/MET563-3
📧 carlos.bastarz@inpe.br