

# **Régressão e Continentalidade**

**Luciano Araújo Dourado Filho, Alexandre Ribeiro Carneiro,  
Lokisley e Marcos Vinícius Cruz Lima**

<sup>1</sup>Uefs - Universidade Estadual de Feira de Santana  
Av. Transnordestina, s/n, Novo Horizonte  
Feira de Santana – BA – Brasil – 44036-900

lucianoadfilho@gmail.com

marcosviniciuscl@outlook.com

## **1. Introdução**

Em engenharia, é utilizado o ajuste de curvas em análise de tendência, e teste de hipótese. Seguindo esse ponto, iremos analisar o clima no estado da Bahia que é influenciado por um conjunto de fatores regionais e continentais, tais fatores lidam com diferentes características, essas sendo temperatura, umidade, chuvas e ventos ao decorrer de um certo tempo.

## **2. Objetivos**

Este trabalho tem por objetivo geral realizar um estudo sobre a interpolações e régressões com polinômios de primeiro a quinto grau, explicando a correlação da variação térmica para o conteúdo de continentalidade abordado ao longo deste trabalho.

## **3. Metodologia Experimental**

Como demonstrado na Figura 3, o conjunto de dados proposto para o estudo dispõe de 6300 registros com entradas para as informações de temperatura média, altitude, latitude, longitude, mês e ano para alguns municípios baianos.

Em primeira instância, constatou-se que haviam informações para apenas 25 municípios (Figura 3) de um total de 417 dos quais a Bahia é constituída. Levando em consideração que para analisar os efeitos da continentalidade sobre o estado da Bahia seriam necessários dados para todos os municípios a fim de garantir maior precisão, propôs-se utilizar o conjunto de dados obtido para obtenção de um modelo para realizar a inferência de temperatura para os municípios em que os dados encontram-se faltantes.

	CIDADE	TEMPERATURA MÉDIA	ALT(M)	LAT (º DEC.)	LONG (ºdec)	MÊS	ANO
0	Alagoinhas	26.7	128.8	12.1	38.4	1	1999
1	Alagoinhas	26.7	128.8	12.1	38.4	2	1999
2	Alagoinhas	27.0	128.8	12.1	38.4	3	1999
3	Alagoinhas	25.6	128.8	12.1	38.4	4	1999
4	Alagoinhas	24.2	128.8	12.1	38.4	5	1999
...	...	...	...	...	...	...	...
6295	Vitória da Conquista	18.1	879.5	14.9	40.8	8	2019
6296	Vitória da Conquista	20.2	879.5	14.9	40.8	9	2019
6297	Vitória da Conquista	21.7	879.5	14.9	40.8	10	2019
6298	Vitória da Conquista	23.0	879.5	14.9	40.8	11	2019
6299	Vitória da Conquista	23.2	879.5	14.9	40.8	12	2019

**Figura 1. Conjunto de dados**

```
['Alagoinhas',
 'Barra',
 'Barreiras',
 'Bom Jesus da Lapa',
 'Caetité',
 'Caravelas',
 'Carinhanha',
 'Cipó',
 'Correntina',
 'Cruz das Almas',
 'Feira de Santana',
 'Guaratinga',
 'Irecê',
 'Itaberaba',
 'Itiruçu',
 'Jacobina',
 'Lençóis',
 'Monte Santo',
 'Morro do Chapéu',
 'Paulo Afonso',
 'Remanso',
 'Senhor do Bonfim',
 'Serrinha',
 'Santa Rita de Cássia',
 'Vitória da Conquista']
```

**Figura 2. Municípios**

Para isso, como o foco do trabalho consistiu em avaliar modelos de regressão, nossa proposta consistiu em avaliar 3 abordagens que constituiram:

1. Regressão Linear
2. Regressão Polinomial
3. Regressão Multivariada

Para as abordagens de regressão Linear e Polinomial, a proposta consistiu em relacionar os atributos dos conjuntos de dados com a temperatura dois a dois e.g., Temperatura x Altitude, Temperatura x Latitude, Temperatura vs Longitude ... Isso possibilitou a obtenção de 5 curvas (lineares e não lineares) que possibilitam a inferência de um valor de temperatura para uma determinada variável de entrada.

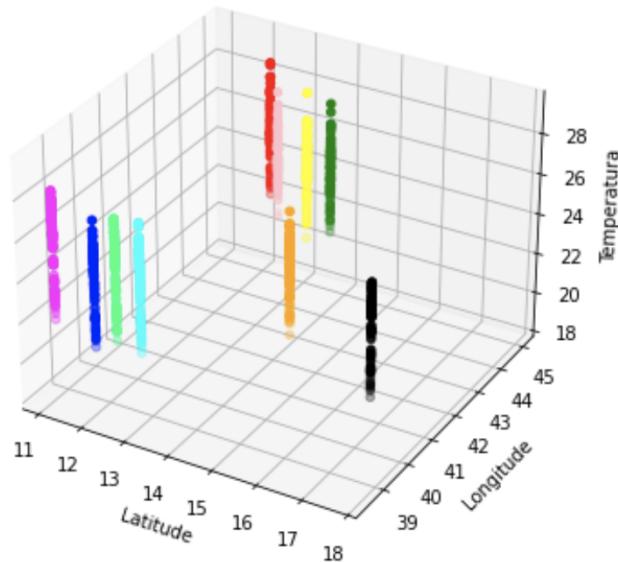
A heurística empregada para relacionar de forma unificada os cinco modelos (obtidos à partir das cinco variáveis de entrada) a fim de possibilitar a obtenção de uma só saída de temperatura, foi a de computar a média da temperatura das saídas de cada uma das cinco retas. Essa heurística foi utilizada para as retas de regressão linear e também para as curvas de regressão polinomial.

No caso do modelo por regressão multivariada, não foi necessário empregar nenhuma heurística pois o método já emprega uma heurística própria de equacionamento das curvas provenientes das relações das variáveis de entrada x temperatura que consiste na atribuição e determinar coeficientes de contribuição para cada uma das curvas individualmente.

### 3.1. Análise Preliminar

A fim de melhor investigar a natureza do fenômeno que desejamos estudar, assim como a viabilidade de realização da proposta, realizou-se uma inspeção visual de alguns dos atributos do conjunto de dados para alguns subconjuntos de amostras. Na Figura 3.1, por exemplo, são apresentados os dados de temperatura para alguns municípios escolhidos arbitrariamente para o período entre 1999 e 2005. O gráfico permite visualizar a relação entre Temperatura, Latitude e Longitude, e com isso investigar analiticamente se por exemplo seria possível inferir a temperatura média de um município com base nos dados de Latitude, Longitude.

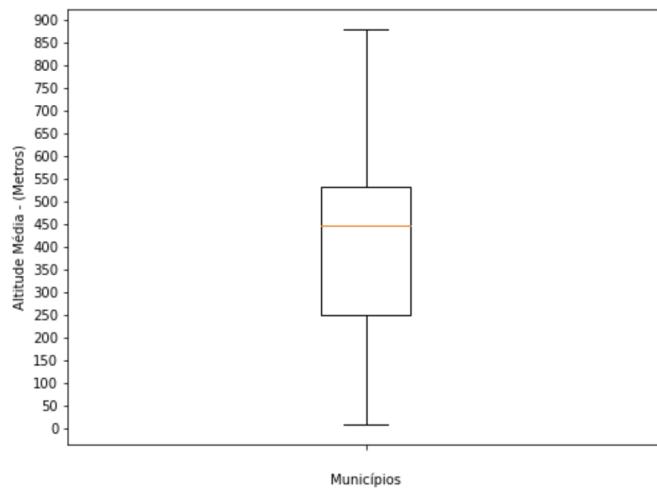
Temperatura vs LAT-LONG Em Algumas cidades entre os anos de 1999 e 2005



**Figura 3. Temperatura vs Latitude vs Longitude**

Com isso é possível observar-se que, espacialmente, municípios mais próximos apresentaram variações de temperatura semelhantes, o que leva a crer que seja viável inferir a temperatura de um município para qual não se tem dados, e que maior será a precisão se houverem dados correspondentes no conjunto de treinamento, em que a proximidade com o município do qual deseja-se inferir a temperatura seja a mínima possível.

Portanto, em contrapartida quanto melhor, em termos de precisão, o modelo for capaz de capturar a relação entre disposição espacial (latitude x longitude) e variação de temperatura, mais precisa consequentemente será a temperatura inferida para o município com dado ausente.



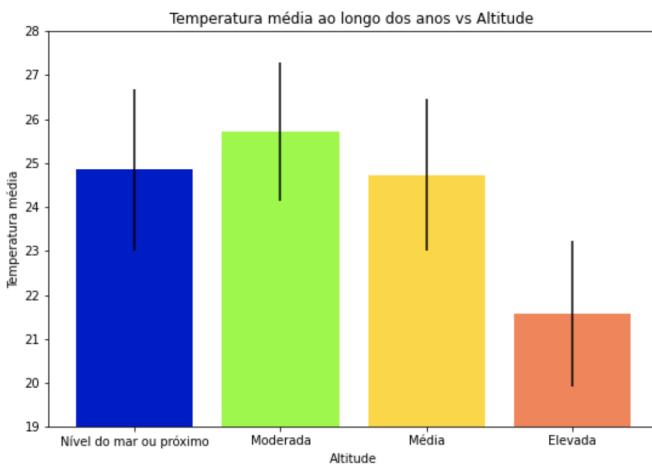
**Figura 4. Boxplot Altitude**

Em relação à Altitude, pudemos plotar um gráfico que permitisse analisar a variação de Altitude por município presente no conjunto. Como demonstrado na Figura 4, os quartis representados nos gráficos ilustram em termos percentuais, as frações dos municípios que se encontram em determinados intervalos de altitude. Por exemplo, o primeiro quartil é formado por municípios que apresentam altitude média entre 0 e aproximadamente 250 metros, e correspondem a 25% dos dados do conjunto. O segundo quartil representa outra fração de 25% dos dados, que apresentam altitude média entre aproximadamente 250 e 450 metros. A linha laranja representa a mediana e divide o segundo e terceiro quartil.

A partir dessas informações foi possível agrupar os municípios de acordo com categorias correspondentes aos níveis de altitude a fim de analisarmos seus efeitos em conformidade com a variação de temperatura. Para isso, a partir dos intervalos apresentados na Figura 4, definiu-se as categorias:

- Altitude nível do mar ou próximo: entre 0 e 255 metros.
- Altitude Moderada: entre 255 e 477.8 metros.
- Altitude Média: entre 477.8 e 551.7 metros.
- Altitude Elevada: acima de 551.7 metros.

A partir das categorias definidas, pode-se agrupar os municípios em conformidade com os níveis de altitude correspondentes e calcular a média e desvio padrão da temperatura ao longo dos anos, a fim de analisarmos seu comportamento, como demonstrado na Figura 5.



**Figura 5. Altitude média por categoria vs Temperatura média ao longo dos anos.**

Em primeira instância observou-se uma alta dispersão das medidas, a qual se atribui a característica de se tratar de uma média de temperatura dos 12 meses ao decorrer de 21 anos, o que se justifica pela temperatura variar com a sazonalidade, além dos efeitos diversos da interação humana com o meio ambiente, como por exemplo emissão de carbono, que frequentemente é relacionada com a elevação das médias de temperatura ao longo dos últimos anos pelo globo.

Além disso, observa-se que o comportamento esperado seria que a temperatura média decrescesse com a elevação de altitude, o que não é verificado para os municípios de altitude Moderada, em relação aos de Nível do mar ou próximo. Acredita-se que esse fator se deu devido ao critério de categorização dos dados de altitude em conjunto à falta de representatividade do conjunto, que dispõe de uma cobertura de apenas 6% dos municípios da Bahia. No entanto, observou-se que de modo geral a temperatura apresentou uma relação inversa com a elevação de altitude e portanto a altitude também pode ser relevante para estimar a temperatura.

### 3.2. Preprocessamento e Avaliação Experimental

Para garantir melhor rigor à avaliação separamos o conjunto de dados em treino e teste através de um protocolo de amostragem aleatória estratificada (70% treino, 30% teste) por município, ou seja, de forma a garantir que houvesse representação para os mesmos municípios em ambos os conjuntos de treino e de teste.

Para construção dos modelos de regressão linear e polinomial nenhum preprocessamento além dos ajustes de formatação de representação de casas decimais por ponto ao invés de vírgula foi realizado. Já para construir o modelo baseado em regressão multivariada, considerando que as variáveis apresentam-se em escalas e com unidades de medidas diferentes, os atributos numéricos foram normalizados utilizando a normalização Min-Max. Já os atributos categóricos ordinais (Mês e ano), foram codificados para eliminar a ordinalidade. Para isso foram criados vetores de tamanhos correspondentes ao tamanho do intervalo dos dados (mês: 12, ano: 21) preenchidos com o valor nominal 1 na posição correspondente ao mês ou ano e demais posições preenchidas com zeros, como ilustrado na Figura 6.

	CIDADE	TEMPERATURA MÉDIA	ALT(M)	LAT (° DEC.)	LONG (°dec)	mes_n1	mes_n2	mes_n3	mes_n4	mes_n5	...	ano_n12	ano_n13	ano_n14	ano_n15	...
0	Alagoinhas	25.9	0.123242	0.325301	0.029412	0	0	0	0	0	...	0	0	0	0	0
1	Alagoinhas	21.7	0.123242	0.325301	0.029412	0	0	0	0	0	...	0	0	0	0	0
2	Alagoinhas	21.8	0.123242	0.325301	0.029412	0	0	0	0	0	...	0	0	0	0	0
3	Alagoinhas	24.7	0.123242	0.325301	0.029412	0	0	0	0	1	...	0	0	0	0	0
4	Alagoinhas	24.4	0.123242	0.325301	0.029412	0	0	0	0	1	...	0	0	1	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4420	Vitória da Conquista	21.3	0.877260	0.662651	0.382353	0	0	0	0	0	...	0	0	0	0	0
4421	Vitória da Conquista	21.9	0.877260	0.662651	0.382353	0	0	0	0	0	...	0	0	0	0	1
4422	Vitória da Conquista	18.3	0.877260	0.662651	0.382353	0	0	0	0	0	...	0	0	0	0	0
4423	Vitória da Conquista	21.9	0.877260	0.662651	0.382353	0	0	1	0	0	...	0	0	0	0	0
4424	Vitória da Conquista	21.8	0.877260	0.662651	0.382353	0	1	0	0	0	...	0	0	0	0	0

4425 rows × 38 columns

**Figura 6. Dataset preprocessado**

Para avaliar o desempenho dos modelos computamos, além da média do coeficiente de determinação  $R^2$  para as curvas de regressão, a medida Erro Quadrático Médio (Mean Squared Error - MSE), medida de avaliação para modelos de regressão que consiste na raiz quadrada da média do erro quadrático, onde o erro é definido como a diferença entre o valor inferido e o valor real, como demonstrado na equação apresentada na Figura 7. Adicionalmente para o modelo de regressão polinomial também avaliamos o erro absoluto médio MAE (Figura 7).

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$$

**Figura 7. Mean Squared Error & Mean Absolute Error.**

#### 4. Análise de Resultados

A Figura 8 demonstra os resultados obtidos para cada uma das curvas correspondentes às ilustradas na Figura 10 no conjunto de treinamento para o modelo de Regressão Linear, em que a saída de temperatura é obtida à partir da temperatura média das 5 retas. Esse modelo obteve um valor médio de coeficiente de determinação  $R^2 = 0.0623$ , o que significa que grande parte dos pontos não foram adequados à reta, o que é justificável, levando-se em consideração a alta dispersão dos dados que é observável nos gráficos da Figura 10. Em contrapartida, no conjunto de teste, o modelo teve desempenho de MSE=22.4034.

	name	slope	intercept	r	p	std_err
0	temperatura x altitude	-0.004802156145257135	26.48989277730599	-0.4778938763595215	2.4490551516958633e-251	0.00013272335776116055
1	temperatura x latitude	-0.3035086151310716	28.123055916306583	-0.23959276294536205	8.28569632966416e-59	0.01849274982439446
2	temperatura x longitude	0.1215040238585787	19.360151563442646	0.10036957266073883	2.2064036055868554e-11	0.018110542936992668
3	temperatura x mes	-0.06231987194202965	24.756615898828162	-0.08685116746258519	7.178146730084067e-09	0.010748509487236563
4	temperatura x ano	0.03698365082128744	-49.950808661495856	0.09057645970128686	1.5806480087357979e-09	0.0061143053552236756

**Figura 8. Resultados de avaliação (conjunto de treino) das curvas obtidas para o modelo.**

Observa-se que à essa dispersão se atribui o fato do conjunto de dados apresentar diversas leituras para as mesmas variáveis, ou seja, para mesmos valores de altitude, latitude, longitude, temperatura, etc... tem-se as medições para cada um dos 12 meses dos 21 anos apresentados no conjunto. Tendo em vista isso, propusemos reduzir a dispersão por do treinamento do modelo com uma redução de dados referentes aos anos. Assim avaliamos o desempenho do modelo treinado com dados apenas dos últimos 10, 5 e 2 anos e comparamos com o modelo treinado sobre o conjunto inteiro. Adicionalmente avaliamos também o desempenho dos modelos treinados com e sem os atributos: mês e ano.

A Figura 9, mostra os resultados para os modelos propostos. Como se pode observar, a redução na granularidade dos dados impactou positivamente no desempenho do modelo, de forma a reduzir o erro quadrático médio até o ponto em que os dados já começavam a se tornar escassos demais (corte2). Em contrapartida os resultados mostraram que a remoção do atributo ano impactou negativamente o desempenho, com aumento significativo do erro quadrático médio no conjunto de testes, e mais significativamente ainda se removidos os atributos Mês e Ano conjuntamente.

Esses resultados corroboraram para a hipótese de que a temperatura é significativamente afetada pela sazonalidade e que consequentemente a informação relativa ao mês é imprescindível para obter um modelo representativo. O mesmo vale para o atributo Ano, cujos resultados refletiram a capacidade do modelo de capturar as consequências da ampla variedade de fenômenos que causam as variações de temperatura ao longo dos anos.

```

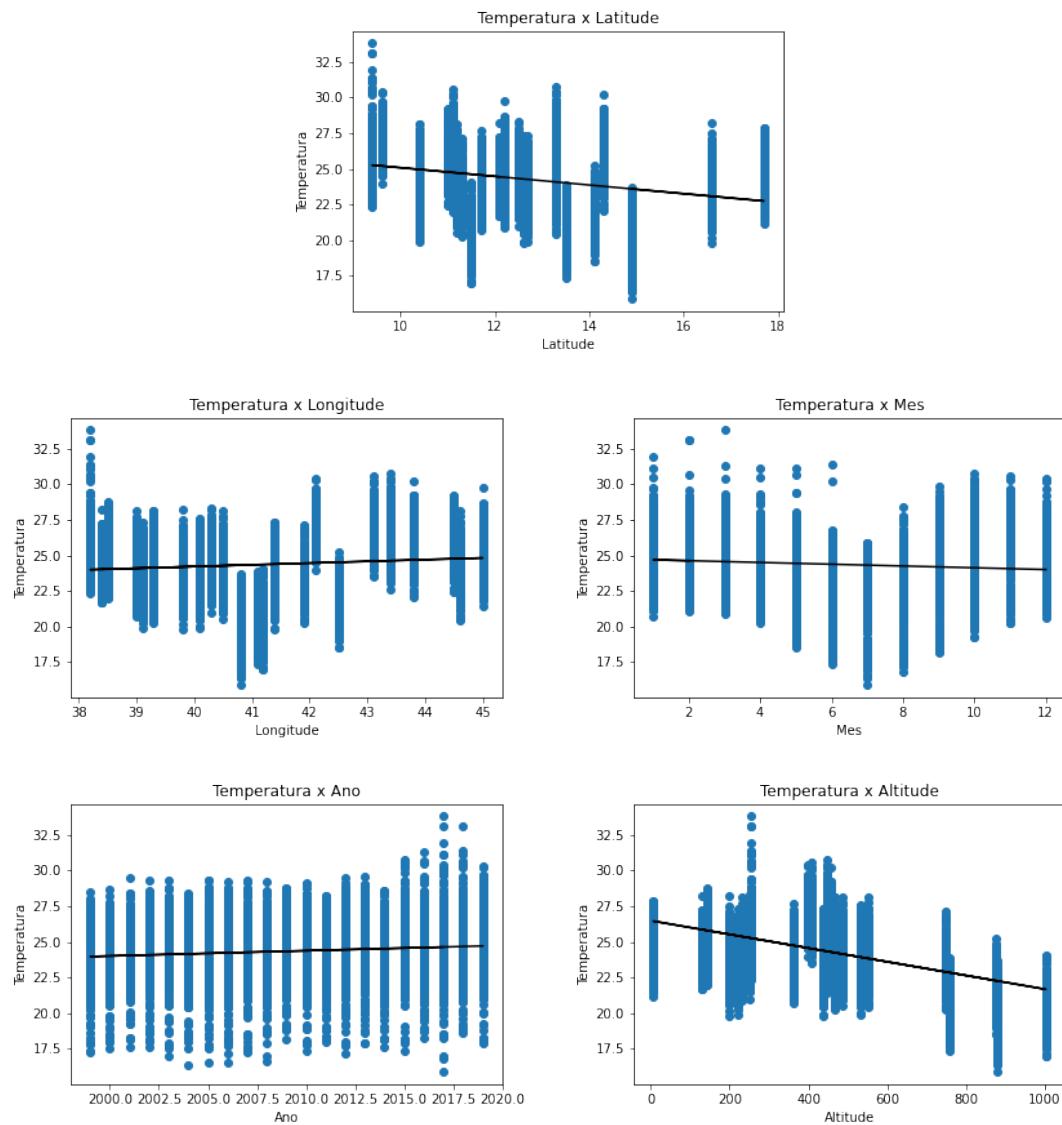
Modelo: linear_regression_model_baseline
Mean Squared Error: 22.403457955405784 (Com feature ano)
Mean Squared Error: 29.680547476275855 (Sem feature ano)
Mean Squared Error: 43.08268568038433 (Sem feature mes e ano)
Modelo: linear_regression_model_corte10
Mean Squared Error: 21.463723182805623 (Com feature ano)
Mean Squared Error: 28.07238348498347 (Sem feature ano)
Mean Squared Error: 41.20175841552263 (Sem feature mes e ano)
Modelo: linear_regression_model_corte5
Mean Squared Error: 19.99844147026798 (Com feature ano)
Mean Squared Error: 26.675350853172688 (Sem feature ano)
Mean Squared Error: 39.54925492319826 (Sem feature mes e ano)
Modelo: linear_regression_model_corte2
Mean Squared Error: 24.153000073609974 (Com feature ano)
Mean Squared Error: 26.340505609501363 (Sem feature ano)
Mean Squared Error: 39.159501169682216 (Sem feature mes e ano)

```

**Figura 9. Resultados de avaliação (conjunto de teste) dos modelos de regressão linear com quantidades reduzidas de dados.**

#### 4.1. Regressão polinomial

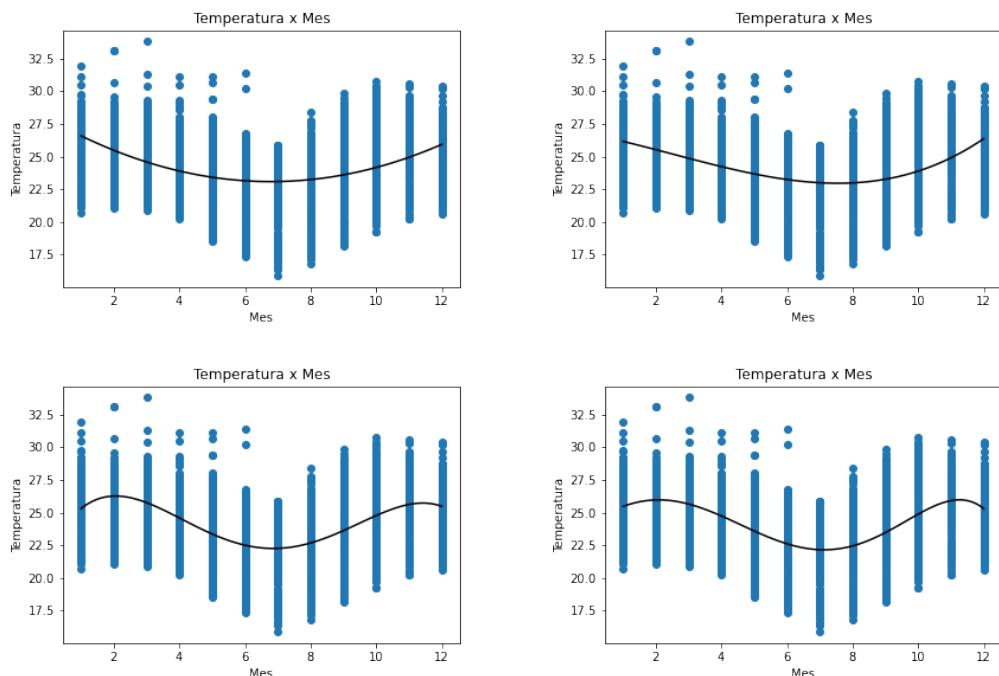
Como observado na Figura10, apesar da dispersão notavelmente evidente, os dados demonstraram comportamento não linear, fator que também contribuiu para o baixo desem-



**Figura 10. Resultados para o modelo de regressão linear.**

penho do modelo de regressão linear. Motivado por isso, avaliamos o desempenho dos modelos de Regressão Polinomial, com polinômios de grau 2,3,4 e 5. Além disso, apesar de constatar-se que a redução da granularidade dos dados impacta positivamente até um certo limite, realizamos a mesma avaliação, porém considerando apenas os resultados do modelo para o conjunto de testes com corte de 5 anos, em relação ao desempenho sobre o conjunto completo.

A Figura 11 demonstra as curvas resultantes dos modelos de regressão polinomial para o atributo Mês para os polinômios de graus 2,3,4 e 5 respectivamente (da esquerda para direita e de cima para baixo). Como podemos observar, de fato o modelo de regressão polinomial obteve retas que se adequaram melhor aos dados de forma geral.



**Figura 11. Adequação das curvas aos dados (Temperatura x Mês) para crescentes graus de polinômios.**

Em relação ao desempenho versus granularidade, as Figuras 12 e 13 demonstram a diferença significativa de desempenho entre os modelos treinados com dados dos 21 anos em relação ao modelo treinado com apenas os dados dos últimos 5 anos. Além disso ressalta-se o desempenho em função do aumento do grau do polinômio do modelo de regressão, que muda mais significativamente do grau 2 para o 3, e tende a reduzir o erro quadrático médio ainda mais conforme o aumento do grau. Como os resultados demonstram, para modelos de característica polinomial, a diferença de performance obtida ao reduzir a granularidade/amplitude dos dados foi muito mais expressiva em relação ao modelo de regressão linear.

```

Modelo Polinomial corte 20
Mean Squared Error: 21.64010725457557 (Polynomial Model, degree: 2 )
Mean Squared Error: 21.487006749289467 (Polynomial Model, degree: 3 )
Mean Squared Error: 21.19977989693724 (Polynomial Model, degree: 4 )
Mean Squared Error: 21.16104436580146 (Polynomial Model, degree: 5 )

```

**Figura 12.** .

```

Modelo Polinomial corte 5 (com feature ano)
Mean Squared Error: 12.248525644578875 (Polynomial Model, degree: 2 )
Mean Squared Error: 4.6011494399493715 (Polynomial Model, degree: 3 )
Mean Squared Error: 4.277855967996905 (Polynomial Model, degree: 4 )
Mean Squared Error: 4.21774820792532 (Polynomial Model, degree: 5 )

```

**Figura 13.** .

## 5. Regressão Multivariada

Como os resultados das análises puderam reforçar, os modelos lineares e polinomiais tenderam a obter melhor desempenho quando se reduziu a dispersão dos dados. Como a Figura 10 ilustra, a dispersão apresentada nos dados, que é justificada pelo fato de se apresentarem medições diferentes de temperatura para anos e meses diferentes apesar de mantidas as condições de latitude, longitude, altitude, etc... Em contrapartida, é possível observar um padrão de organização na dispersão, que faz com que haja uma tendência de crescimento da temperatura a despeito da preservação das condições mencionadas anteriormente.

Como também mencionado, essa "dispersão organizada" é prevista, haja vista que a temperatura média do globo tem sido crescente ao decorrer dos anos, no entanto, o fato dos modelos de regressão apresentarem melhor performance quando atenua-se a dispersão corrobora para a hipótese de que a regressão, polinomial ou linear equacionada por meio da média dos resultados das retas tenha capacidade suficiente de representação. Como as informações de temperatura relativas ao mês e ano dos municípios são significativamente relevantes para obtenção de modelos com melhor capacidade preditiva, resta testar modelos que possam, hipoteticamente tirar melhor proveito dessas relações, ou seja, que tenham maior capacidade de representação do fenômeno estudado.

Nesse sentido, como a heurística empregada para realizar a previsão de temperatura com base nas variáveis independentes é a média entre as saídas correspondentes das variáveis dependentes das equações, despreza-se as informações que as relações entre os dados num espaço n-dimensional permitiria capturar para tornar o modelo mais representativo. Por conta disso, propôs-se a utilização de um algoritmo de regressão multivariada, que analogamente, permite definir adequar os dados de um fenômeno multidimensional, num hiperplano, portanto supostamente tendo maior capacidade de representação. Para verificar essa hipótese avaliamos o desempenho dos modelo de regressão multivariada treinado sobre o conjunto de dados inteiro em relação ao desempenho do modelo treinado sobre o conjunto com apenas os dados dos últimos 5 anos.

A Figura 14 apresenta os resultados para os modelos propostos. Como demonstrado, o modelo treinado sobre o conjunto inteiro apresentou performance substancial-

mente superior ao modelo treinado apenas com dados dos últimos 5 anos. Acredita-se que por se tratar de tentar traçar um plano que contenha a máxima quantidade de dados possível, torna-se mais difícil em termos de precisão traçar o plano na medida em que se diminui a quantidade dos dados. Isso se justifica quando comparamos os resultados de  $R^2$  obtidos pelos modelos, em que o modelo treinado com apenas dados dos últimos 5 anos (cut 5) apresenta  $R^2$  negativo, enquanto o modelo treinado sobre o conjunto inteiro,  $R^2 = 0.75$ .

```
R2 score: 0.755234342905827
Mean Squared Error: 1.4403472634057204 (Multivariate Linear Regression model, cut 21)
Mean Absolute Error: 0.9365769075520833 (Multivariate Linear Regression model, cut 21)

R2 score: -3.3717435475011775e+27
Mean Squared Error: 1.9841352129235584e+28 (Multivariate Linear Regression model, cut 5)
Mean Absolute Error: 113266933057630.03 (Multivariate Linear Regression model, cut 5)
```

**Figura 14. Resultados para os modelos de Regressão Multivariada.**

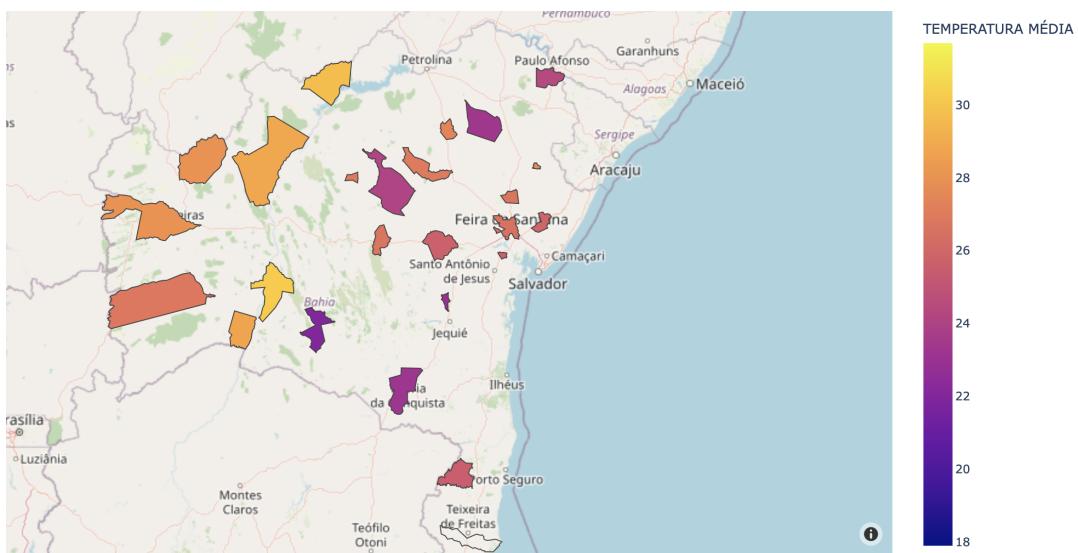
### 5.1. Análise dos efeitos da continentalidade

A continentalidade é um fenômeno natural que apresenta efeitos sobre a amplitude térmica de um clima terrestre. Esse efeito é principalmente decorrente da ação de troca de calor do sol com as superfícies terrestres (sólidas ou líquidas) fazendo com que se altere a temperatura atmosférica. Para o caso de meios sólidos, os efeitos se dão devido à rapidez com que o solo se aquece e irradia o calor de sua superfície para a atmosfera. Em contrapartida, no período noturno devido baixa capacidade de absorção da temperatura durante o dia, quanto chega a noite ocorre também a rapidez do resfriamento de sua superfície e isso causa a queda da temperatura atmosférica.

Assim, o coeficiente de dilatação térmica da superfície irradiada pelo sol determinará a velocidade com que a troca de calor será realizada com o ar atmosférico. Portanto, no caso de regiões mais próximas à superfícies líquidas como mares, lagos, rios, oceanos, o ar atmosférico poderá ser aquecido lentamente, por conta do calor específico da água. Neste caso a água atuará como mecanismo regulador de temperatura, fazendo com que em regiões litorâneas, por exemplo, a temperatura se mantenha praticamente constante, pois durante o dia a água absorve o calor do sol e, à noite, a irradiação lenta do calor absorvido pela massa de água faz com que o ar em torno se aqueça, mantendo a temperatura.

Considerando a baixa cobertura do conjunto de dados em relação aos municípios (Figura 15), utilizamos o modelo de regressão multivariada para inferir a temperatura dos municípios para o conjunto de dados não cobria. Para isso obtivemos um conjunto de dados que dispõe das informações referentes à latitude, longitude e altitude média de cada município do estado da Bahia, realizamos os pre-processamentos de forma a garantir compatibilidade estatística com os dados de treinamento (normalização, codificação, etc...). À partir disso pode-se inferir a temperatura mensal para cada município para um determinado ano entre 1999 e 2019. Diante disso, pudemos inferir as temperaturas médias mensais para cada município no ano de 2019, calculamos a média de temperatura anual para cada município e plotamos os resultados no mapa da Bahia.

A Figura 16 permite visualizar os resultados de média anual de temperatura obtidos para o ano de 2019 à partir do modelo de regressão multivariada. Como os dados apre-



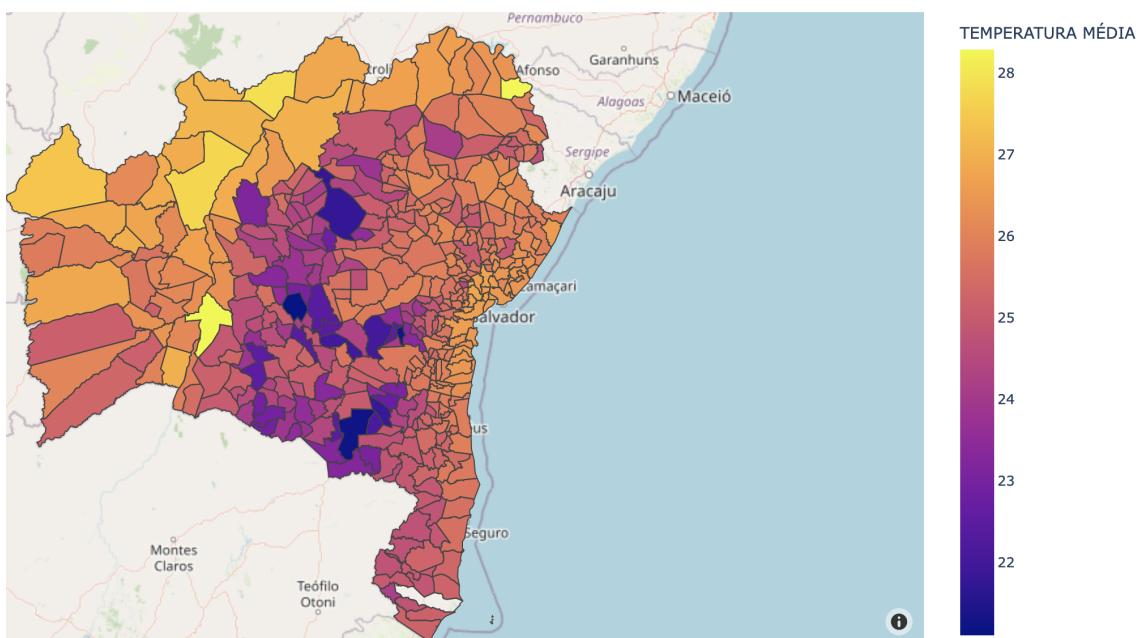
**Figura 15. Disposição das médias de temperatura do ano de 2019 para os municípios disponíveis no conjunto de dados original.**

sentam uma granularidade de média mensal, não há como estimar as amplitudes térmicas diárias e suas variações ao longo dos meses. Portanto para analisarmos a relação com os efeitos de Continentalidade vs Maritimidade pudemos verificar que, latitudinalmente, nos pontos extremos do estado há uma diferença significativa de média anual de temperatura.

Observa-se que na medida em que se aumenta a continentalidade, ou seja, distância em relação ao litoral, as temperaturas médias são mais elevadas. No entanto, como as temperaturas são influenciadas por outros fatores climáticos como altitude e longitude, ou seja, amplificam-se tanto os efeitos de maritimidade como de continentalidade conforme aproxima-se do equador, pois aumenta-se a irradiação solar e consequentemente a troca de calor se intensifica. Em contrapartida na medida em que se eleva a altitude têm-se temperaturas mais baixas por conta da menor pressão atmosférica.

Diante disso, é possível correlacionar também a temperatura média anual inferida para os municípios, com o nível de elevação territorial ilustrada na Figura 17, que demonstra um modelo digital de elevação da Bahia<sup>1</sup>. Isso possibilitou constatar que o modelo de regressão multivariada permitiu gerar uma correspondência entre os efeitos esperados da atuação dos fatores climáticos sobre as médias de temperatura obtidas para os municípios.

<sup>1</sup>Disponível em <http://www.bahia-turismo.com/mapas/relevo.htm>



**Figura 16. Disposição das médias de temperatura estimadas para o ano de 2019 para todos municípios do estado da Bahia.**



**Figura 17. Modelo de elevação digital do estado da Bahia.**

## 6. Conclusão

As análises realizadas por meio dos resultados encontrados possibilitaram evidenciar a possibilidade de estimar a temperatura média mensal para um determinado município, com base em sua localização geográfica e altitude. Além disso os resultados possibilitaram verificar a hipótese da possibilidade de inferência de temperatura para um município

com base nos dados de outro em que se haja certa proximidade geográfica. Diante disso ressalta-se a capacidade dos modelos obtidos em representar com satisfatória precisão, os efeitos das variáveis que influenciam sobre as condições climáticas, no que dizem respeito à temperatura.

Como trabalhos futuros tem-se em vista avaliar a precisão dos dados gerados para os municípios "faltantes" por meio de algum conjunto de dados com as leituras de temperaturas médias mensais das estações desses municípios. Além disso julgamos que seria interessante também avaliar a capacidade do modelo de estimar os efeitos das mudanças climáticas em relação à temperatura ao longo dos anos, que advém das causas da interação dos seres humanos com o meio, como mencionado, através da emissão de co<sub>2</sub>, etc...