UNIVERSIDADE ESTADUAL DE MARINGÁ
CENTRO DE CIÊNCIAS EXATAS
CURSO DE ESTATÍSTICA

AVALIAÇÃO DE MÉTODOS NÃO PARAMÉTRICOS PARA PREDIÇÃO EM MODELOS ADITIVOS

Marco Aurelio Valles Leal

Maringá
2022

MARCO AURELIO VALLES LEAL

AVALIAÇÃO DE MÉTODOS NÃO PARAMÉTRICOS PARA PREDIÇÃO EM MODELOS ADITIVOS

Trabalho de conclusão de curso apresentado como requisito parcial para a obtenção do título de bacharel em Estatística pela Universidade Estadual de Maringá.

Orientador: Profº Drº George Lucas Moraes Pezzot a

Maringá
2022

AVALIAÇÃO DE MÉTODOS NÃO PARAMÉTRICOS PARA PREDIÇÃO EM MODELOS ADITIVOS

MARCO AURELIO VALLES LEAL

Trabalho de conclusão de curso apresentado como requisito parcial para a obtenção do título de bacharel em Estatística pela Universidade Estadual de Maringá.

Aprovado em: _____/_____/_____.

BANCA EXAMINADORA

_____

Orientador
Profº Drº George Lucas Moraes Pezzot
Universidade Estadual de Maringá

_____

Membro da banca
Profº Drº Brian Alvarez Ribeiro de Melo
Universidade Estadual de Maringá

_____

Membro da banca
Profº Drº Willian Luís de Oliveira
Universidade Estadual de Maringá

RESUMO

É comum, nas mais diversas áreas, investigar e modelar a relação entre variáveis. O modelo mais simples é denominado modelo de regressão linear simples e assume que a média da variável resposta é modelada como uma função linear das variáveis explicativas, supondo erros aleatórios com média zero, variância constante e não correlacionados. Entretanto, nem sempre a relação existente é perfeitamente linear. Neste contexto, é possível flexibilizar o modelo de regressão linear modelando a dependência da variável resposta com cada uma das variáveis explicativas em um contexto não paramétrico. Esta nova classe de modelos é dita modelos aditivos e mantêm a característica dos modelos de regressão lineares de serem aditivos nos efeitos preditivos. Portanto, este projeto visa apresentar os modelos aditivos, além de técnicas de suavização utilizadas para ajustar modelos no contexto não paramétrico. Por fim, a metodologia é aplicada em dados artificiais (simulados) e em dados reais, dando enfoque à qualidade das predições.

Palavras-chave : Regressão. Modelo aditivo. Suavizadores.

1 Resultados e Discussão

1.1 Estudo de simulação

Nesta seção, serão utilizadas simulações de dados para gerar situações nas quais possam ser aplicadas as técnicas estudadas, analisando suas respectivas performances. Para os resultados obtidos, quatro técnicas de suavização serão empregadas, sendo elas: o suavizador de *kernel*, *Loess*, *splines* de regressão de grau 1 e grau 3. Realizar-se-ão ajustes para o primeiro cenário, considerando distintos parâmetros de suavização para avaliar, visualmente, os comportamentos das curvas em diagramas de dispersão. Em seguida, adotando o método de *data splitting*, *leave one out cross-validation*, encontrar-se-á um parâmetro de suavização que forneça a ocorrência do menor erro quadrático médio possível, desta forma, evitando um super-ajuste do modelo. Ademais, ajustes serão executados considerando tais parâmetros e, em seguida, calcular-se-á os erros quadráticos médios para cada técnica suavizadora, entre os valores observados e estimativas do modelo. Portanto, será considerado o método mais aderente aquele que apresentar o menor erro quadrático possível.

Posteriormente, este procedimento será repetido para cada cenário em mil amostras, contabilizando a quantidade de vezes em que cada técnica apresenta o menor erro quadrático médio. Por exemplo, para o primeiro cenário, será gerado mil amostras aleatórias de tamanho n. Para cada amostra será empregado o procedimento acima, salvando seus repectivos erros quadráticos médio. Ao final da simulação, será contabilizado se a ocorrência do erro quadrático médio em cada técnica foi mínima e, por fim, comparar e verificar qual técnica obtém o melhor resultado em uma simulação de mil amostras. Serão avaliados as duas visões, primeiramente analisando o comportamento para EQM's obtidos do processo de *LOOCV*, para escolha do melhor parâmetro de suavização concluindo qual melhor técnicas de suavização obtém um melhor desempenho de predição. A segunda comparar dentros dos mesmos cenários os EQM's, considerando os dados observados e preditos e concluir qual dos suavizadores são mais aderente aos dados.

Vale ressaltar que serão empregados dois comportamentos, uma proveniente de uma função senoidal e outra de uma função Gamma: Cenário 1 e Cenário 2. Ainda, serão gerados nove sub-cenários, valendo-se da combinação de três tamanhos amostrais (150, 250 e 350), em três valores de desvio padrão distintos.

1.1.1 Cenário 1

Para este cenário, será considerado X uma sequência de 0 a 50 e Y, definido pela função

$$y = 10 + 5sen\pi\frac{x}{24} + \varepsilon,$$

onde ε é um termo aleatório, normalmente, distribuído com média zero e variância constante.

Os tamanhos amostrais utilizados serão iguais a 150, 250 e 350 e valores de desvio padrão 0.5, 1 e 2. Na Figura 1, tem-se o comportamento dos dados para cada desvio padrão, considerando 350 observações com a curva : $10 + 5sen\pi\frac{x}{24}$.

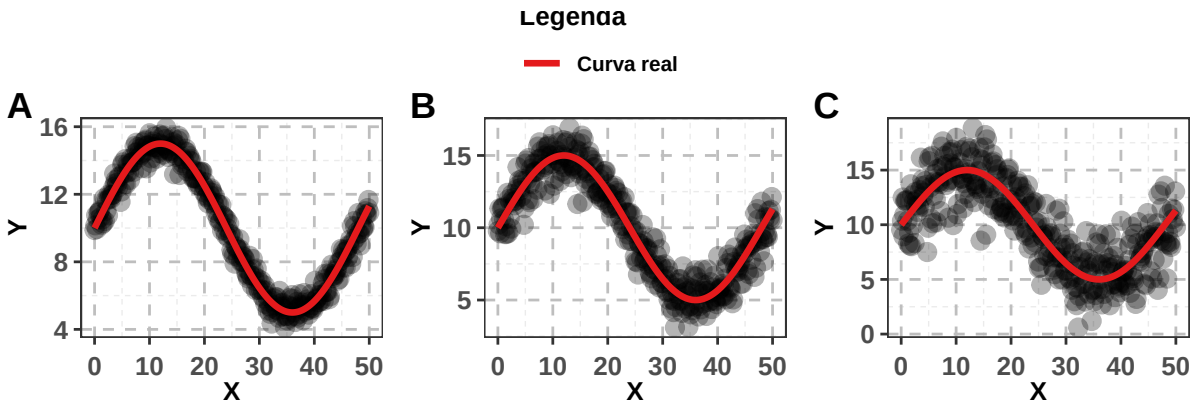

Figura 1: Gráfico de dispersão dos dados simulados e curva real, para o Cenário 1. (A) DP = 0.5, (B) DP = 1 e (C) DP = 2

Levando em conta a configuração do gráfico C (Figura 1), curvas distintas para cada método de suavização, foram ajustadas, adotando parâmetros de suavização arbitrários e são apresentados na Figura 2.

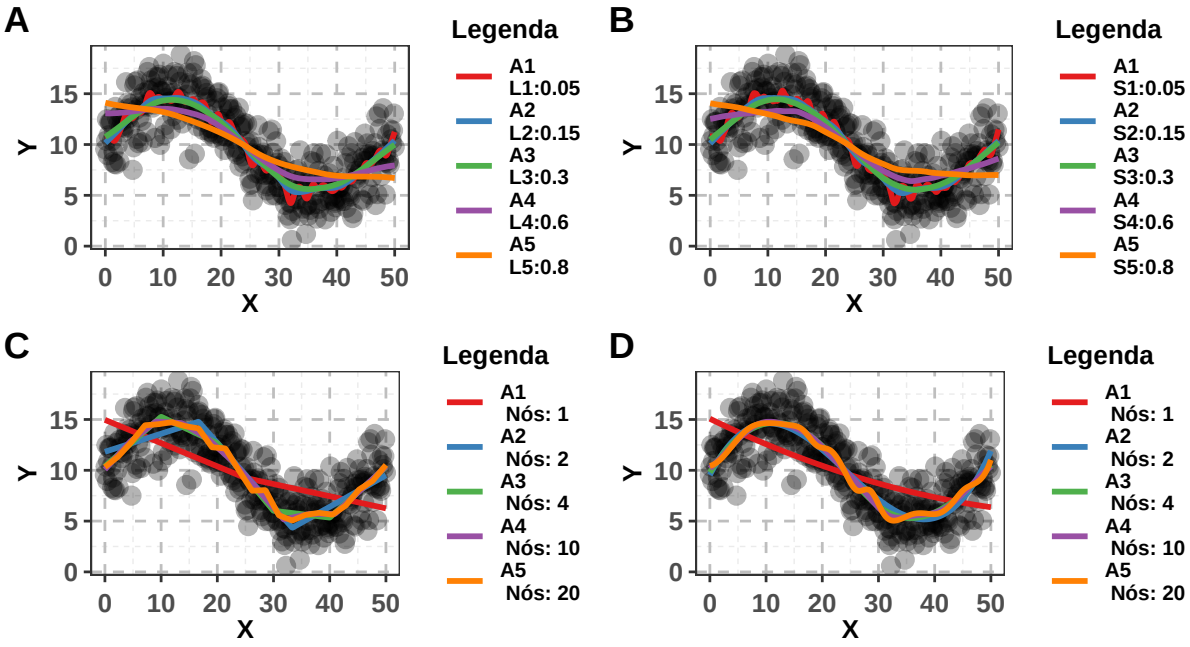

Figura 2: Comparação entre diferentes ajustes (Cenário 1), com parâmetros de suavização distintos, considerando os suavizadores. (A) Kernel, (B) Loess, (C) Splines de Regressão Grau 1 e (D) Splines de Regressão Grau 3.

Ao avaliar os ajustes dos gráficos A e B, verifica-se que, conforme o parâmetro de suavização aumenta, estes tendem a ser muito suaves. Ou seja, na medida em que o parâmetro se torna suficientemente grande, a curva tenderá a ser uma reta. Porém, quando este parâmetro

tende a ser muito pequeno, o ajuste realizado interpolará os dados. Para os gráficos C e D, na proporção em que o parâmetro de suavização aumenta, a curva ajustada apresenta rugosidade em sua forma. Conforme este parâmetro se torna pequeno, a curva tenderá a ser uma reta.

Os suavizadores em diagramas de dispersão remetem a uma ideia visual de como o ajuste está se comportando em relação aos dados. Na Figura 3, são apresentados os gráficos contendo os resultados do erro quadrático médio mínimo obtidos, realizando o *leave one out cross-validation*. Neste gráfico, verifica-se o comportamento dos erros quadrático médio em relação a seus repectivos parâmetros de suavização. Ainda, nota-se para qual parâmetro de suavização haverá o melhor ajuste (evitando super-ajuste), levando em consideração menor erro quadrático possível dentre todos os candidatos. Em outras palavras, ao realizar ajustes controlando o valor do parâmetro de suavização, obter-se-á um ajuste no qual o Erro Quadrático Médio (EQM) será o menor de todos, logo, este será o melhor candidato para representar os dados. Observa-se, então, os parâmetros que, supostamente, induzirão um melhor ajuste para cada suavizador.

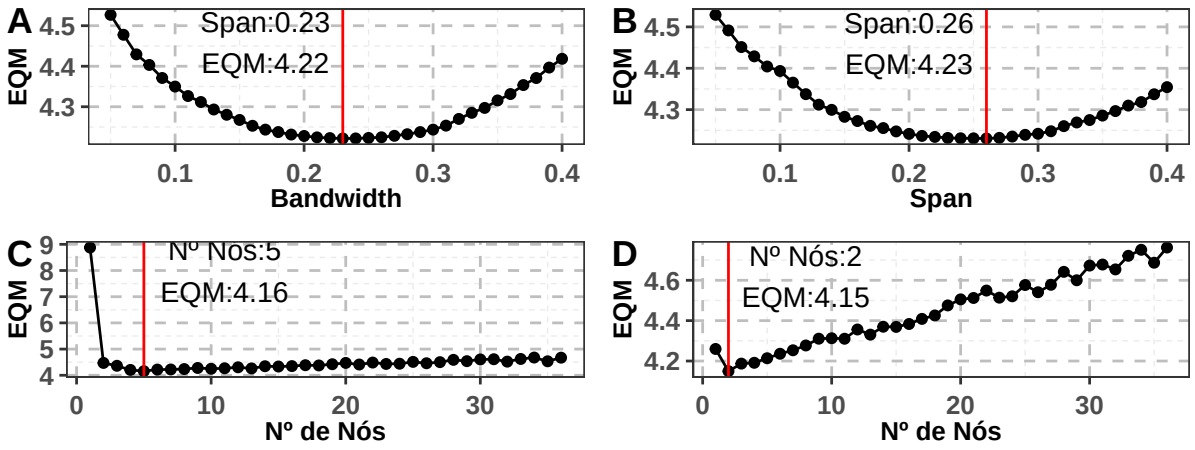

Figura 3: Erro quadrático médio versus parâmetro de suavização (Cenário 1) pós aplicação do Leave One Out Cross-Validation. (A) Kernel, (B) Loess, (C) Splines de Regressão Grau 1 e (D) Splines de Regressão Grau 3.

Com o auxilio da Tabela 1, concluir-se-á que o melhor método de suavização para predição, considerando o Cenário 1 com apenas uma amostra, é o *splines* de regressão cúbico por obter o menor erro quadrático possível. Porém, observando e analisando os resultados obtidos apenas em uma amostra, pode levar à decisões equivocadas.

Tabela 1: Erro Quadrático Médio para os suavizadores Loess, Kernel e Spline Cúbico

Suavizador	Parâm. Suavizador	EQM
Kernel	0.23	4.2216
Loess	0.26	4.2303
Splines Grau 1	5.00	4.1564
Splines Grau 3	2.00	4.1497

Na Figura 4, são demonstrados os ajustes, levando em consideração os melhores parâmetros obtidos por meio do processo de validação cruzada.

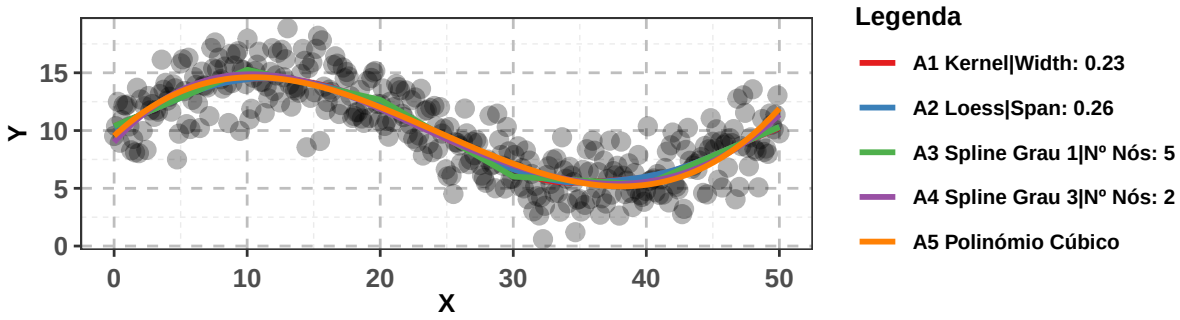

Figura 4: Comparação entre os ajustes, considerando parâmetros de suavização obtidos por meio da validação cruzada.

Na Tabela 1, é mostrado os EQM's provenientes dos ajustes apresentados na Figura 4. Concluímos que a téncnica suavizadora que melhor se ajusta aos dados simulados (Cenário 1) é o *Loess*.

Tabela 2: Erro Quadrático Médio para os suavizadores Loess, Kernel e Spline Cúbico

Suavizador	Parâm. Suavizador	EQM
Kernel	0.23	0.2652
Loess	0.26	0.2634
Splines Grau 1	5	0.2674
Splines Grau 3	2	0.2723
Polinômio Cúbico		0.2751

Neste momento, será reanalisado este procedimento em um processo de geração de amostras aleatórias. Serão considerados os cenários apresentados no começo deste capítulo: três tamanhos amostrais, para três variabilidades, totalizando nove cenários distintos. Para cada cenário serão geradas mil amostras aleatórias, aplicando o procedimento *leave one out cross-validation* para cada método de suavização. Em seguida, contabilizar-se-á a quantidade de vezes em que cada técnica obteve erro quadrático mínimo. Ainda, será avaliado qual suavizador apresenta o melhor ajuste para representar os dados simulados, por meio do cálculo dos EQM's, levando em conta o valores observados e ajustados de cada técnica. Por fim, avaliar-se-á qual técnica obteve o melhor desempenho de predição e ajuste.

Na Tabela 3, estão esquematizados os resultados obtidos, por meio do *leave one out cross-validation* para cada cenário, considerando mil amostras simuladas para os suavizadores *Kernel*, *Loess*, *Splines* de regressão grau 1 e *Splines* de regressão grau 3.

Tabela 3: Percentual do Erro quadrático mínimo, obtidos por meio de aplicação do Procedimento 1, para seleção do melhor parâmetro de suavização, considerando cada suavizador em 1000 amostras.

Sub-Cenário	Tamanho	Desvio Padrão	Kernel	Loess	Sp. Reg. 1	Sp. Reg. 3
1	150	0.5	0.9%	5.3%	19.7%	74.1%
2	150	1.0	1.4%	8.4%	31.1%	59.1%
3	150	2.0	1.4%	10.8%	48%	39.8%
4	250	0.5	0.7%	3.8%	16.9%	78.6%
5	250	1.0	1.1%	7%	23.5%	68.4%
6	250	2.0	1.7%	13.1%	35.8%	49.4%
7	350	0.5	0.8%	3.7%	15.8%	79.7%
8	350	1.0	1.3%	7%	20.5%	71.2%
9	350	2.0	1.8%	11.7%	33.5%	53%

Destaca-se o suavizador *splines* cúbico por obter o melhor desempenho em quase todos os cenários, exceto o terceiro (Sub-Cenário 3), no qual o *splines* de grau 1 obteve erro quadrático médio mínimo em cerca de 48% das amostras simuladas. Ainda quando fixamos o valor do tamanho amostral, constata-se que, conforme a variabilidade dos dados aumenta, há indicios de que os resultados obtidos para a técnica *splines* cúbico estejam se dispersando para as demais técnicas. Observa-se ainda que conforme o tamanho amostral aumenta, há indícios de que os percentuais estejam convergindo e estabilizando, envidenciando que os *splines* cúbico tendem a obter um desempenho melhor em relação aos demais métodos. Além disso, ressalta-se que os percentuais para o suavizador de *kernel* foram os piores, obtendo EQM's mínimo, de até no máximo 1,8% das amostras em relação aos cenários simulados.

Na Figura 5, são apresentados os comportamentos dos erros quadráticos médios obtidos do processo de simulação das amostras. De forma sucinta, verifica-se uma tendência decrescente conforme o tamanho amostral aumenta, assim como a sua aplitude tende a ficar menor. Ademais, visualmente, não é observada diferença significativa entre as técnicas *Kernel* e *Loess*. Ainda, percebe-se que os *splines* (grau 1 e grau 3) apresentam um comportamento mediano relativamente inferior quando comparado aos demais.

Portanto, levando em consideração a Tabela 3 e a Figura 5 pode-se concluir que, para o Cenário 1, o suavizador *splines* de regressão cúbico tende apresentar um melhor desempenho de predição em relação às demais técnicas.

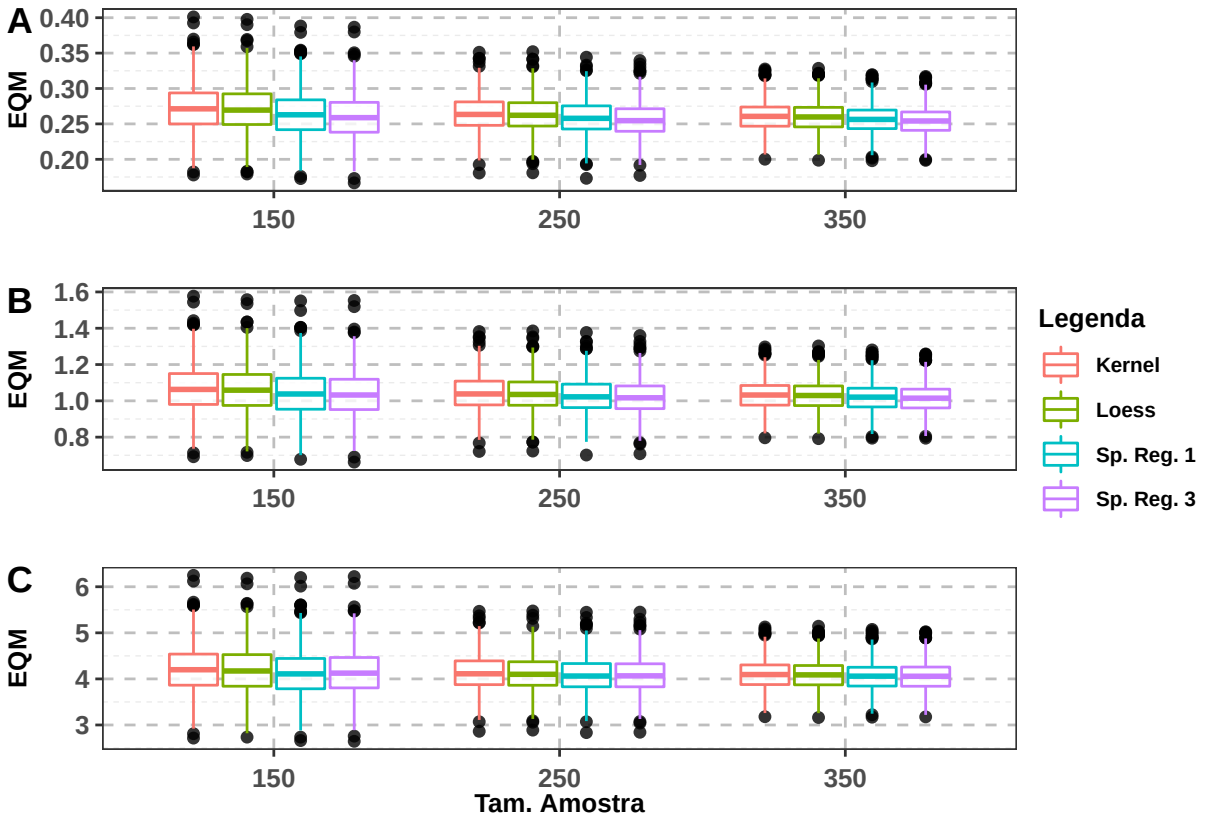

Figura 5: Comparação do erro quadrático para as 1000 amostras para cada suavizador. (A) DP = 0.5, (B) DP = 1 e (C) DP = 2

Na Tabela 4 é apresentado os percentuais, resultantes da contabilização do menor erro quadrático médio completo levando em conta os melhores ajustes obtidos por meio da seleção do melhor parâmetro suavizador.

Tabela 4: Percentual do erro quadrático mínimo completo, levando em conta o ajuste com o melhor parâmetro de suavizador, para todas a técnicas em relação à 1000 amostras.

Sub-Cenário	Tamanho	Desvio Padrão	Kernel	Loess	Sp. Reg. 1	Sp. Reg. 3	Pol. Cúbico
1	150	0.5	65.3%	15.4%	11.7%	7.6%	0%
2	150	1.0	61.3%	11.8%	14.6%	12.3%	0%
3	150	2.0	63.2%	11.2%	11.3%	14.3%	0%
4	250	0.5	67.2%	17.1%	11%	4.7%	0%
5	250	1.0	66.1%	14%	12.7%	7.2%	0%
6	250	2.0	56.2%	12.5%	13.9%	17.4%	0%
7	350	0.5	79.9%	8.8%	8.7%	2.6%	0%
8	350	1.0	63.3%	16.8%	13.1%	6.8%	0%
9	350	2.0	56.2%	12.4%	13.8%	17.6%	0%

Observa-se que o suavizador com *kernel* gaussiano obteve o melhor desempenho de ajuste em todos os cenários propostos, sendo considerado a melhor técnica em no mínimo 56,2% (Sub-Cenário 6 e 9) das amostras simuladas. Em seguida o suavizador *Loess* apresenta um melhor desempenho em quatro sub-cenários, em relação as técnicas restantes, seguido do *splines* de regressão linear e por fim o *splines* o cúbico. A Figura 6 apresenta o comportamento para os EQM_c. Quando comparado as técnicas *kernel* e *Loess*, em relação ao seus comportamento mediano, visualmente não aparentam ser significativamente diferentes, porém em alguns dos cenários, a mediana dos $EQM_c's$ para suavizador *kernel* apresenta ser relativamente inferior. Ao comparar os $EQM_c's$ medianos das técnicas, splines de regressão linear e cúbico estes aparesentam ser relativamente maiores em relação ao *kernel* e *loess*, e não apresentando diferença significativa entre si. Ressalta-se o ajuste paramétrico, regressão cúbico, apresentando o apresentando o pior comportamento para os $EQM_c's$, por apresentar um comportamento mediano significativo quando comparada com as demais técnicas.

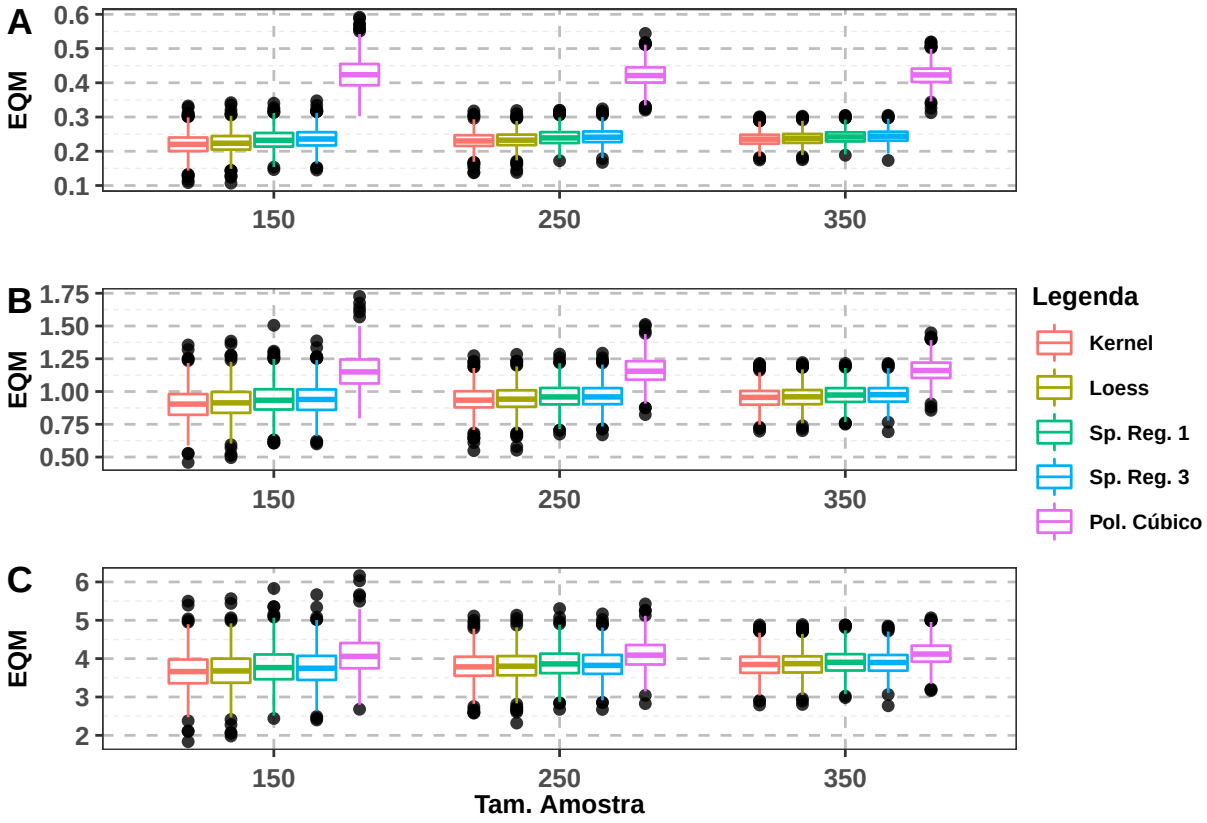

Figura 6: Comparação do erro quadrático para as 1000 amostras para cada suavizador. (A) DP = 0.5, (B) DP = 1 e (C) DP = 2

1.1.2 Cenário 2

Para este cenário, os valores de x serão uma sequência de 0.1 à 2. Ainda, temos que $y = f(x) + \varepsilon$, com $f(x) \sim Gamma(6, 10)$ e $\varepsilon \sim N(0, \sigma^2)$. Serão considerados tamanhos amostrais iguais a $150, 250$ e 350 e valores de desvio padrão $0.05, 0.1$ e 0.15. Na Figura 7, tem-se o comportamento dos dados para cada desvio padrão, considerando 350 observações.

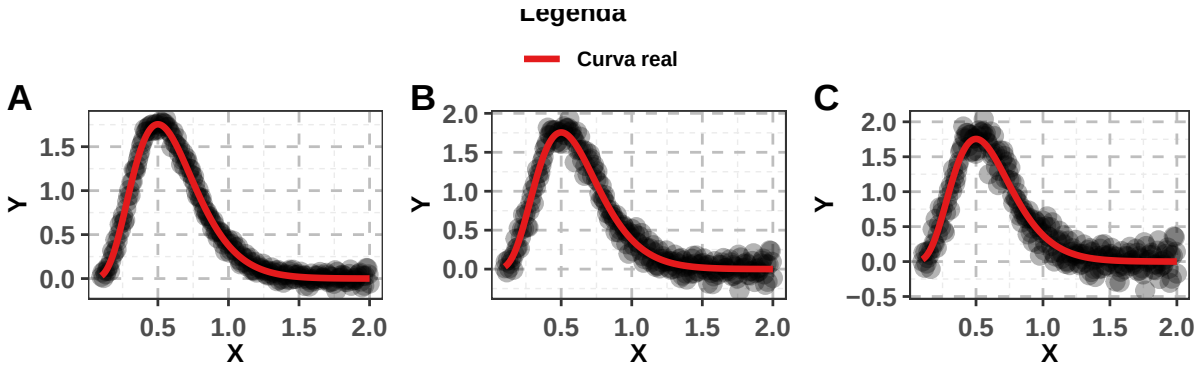

Figura 7: Gráfico de dispersão dos dados gerados para o estudo de simulação.

Quando realizamos a análise aplicando o procedimento de simulação de amostras, que se encontra sumarizado na Tabela 5, observa-se que o suavizador *splines* de regressão cúbico apresentou o melhor desempenho em todos cenários, obtendo o EQM_{LOOCV} mínimo, de aproximadamente 70% à 92% das amostras entre os cenários simulados. Ainda, ressalta-se que, conforme o tamanho amostral aumenta, o percentual apresenta uma têndencia de aumento e indícios de estabilização. Quando fixado um tamanho amostral, o desempenho entre as amostras simuladas com variabilidade distinta apresenta uma tendência decrescente.

Tabela 5: Percentual do Erro quadrático mínimo, obtidos por meio de aplicação do Procedimento 1, para seleção do melhor parâmetro de suavização, considerando cada suavizador em 1000 amostras.

Sub-Cenário	Tamanho	Desvio Padrão	Kernel	Loess	Sp. Reg. 1	Sp. Reg. 3
1	150	0.05	0.6%	2.6%	7.4%	89.4%
2	150	0.10	0.7%	6.3%	13.5%	79.5%
3	150	0.15	0.7%	9.1%	20.6%	69.6%
4	250	0.05	0.4%	1.7%	8.2%	89.7%
5	250	0.10	0.7%	3.4%	10.3%	85.6%
6	250	0.15	0.8%	6.5%	12.7%	80%
7	350	0.05	0.2%	2%	5.9%	91.9%
8	350	0.10	0.6%	2.5%	8.9%	88%
9	350	0.15	1%	4.7%	11.4%	82.9%

De forma análoga ao Cenário 1, o comportamento para os $EQM'_{LOOCV}s$ (vide Figura 8), apresenta uma tendência decrescente conforme o tamanho amostral aumenta, e, percebe-se que a amplitude tende a ficar menor. Ainda, percebe-se que não há diferença signficativa entre os suavizadores *kernel* e *loess*. Os *splines* apresentam ter um comportamento mediano inferior, com destaque ao *splines* cúbico que, visualmente, aparenta ser menor quando comparados com as demais técnicas. Sendo assim, baseado na Tabela 5 e na Figura 8, os dados evidenciam que o suavizador *splines* cúbico possui o melhor desempenho de predição em relação aos outros suavizadores.

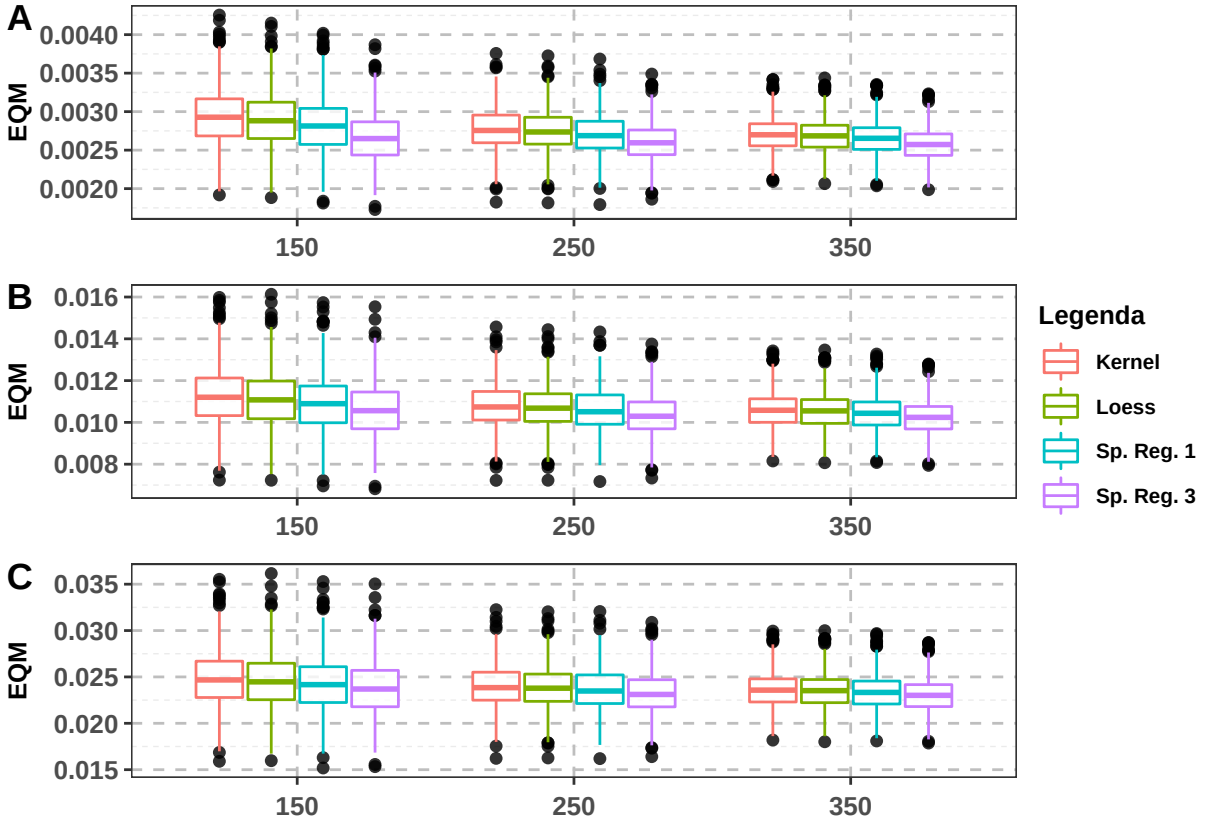

Figura 8: Comparação do erro quadrático para as 1000 amostras para cada suavizador por (A) DP = 0.05, (B) DP = 0.1 e (C) DP = 0.15

Em relação ao desempenho do melhor ajuste, pode ser observado na Tabela **??**. Novamente, destaca-se o suavizador *kernel*, por obter o melhor desempenho em todos os sub-cenários, de no mímino 49,1% das amosntras simuladas. Em seguida, comparado as técnicas restantes destaca-se o método de splines de regressão linear por obter melhor desempenho em quatro sub-cenários, seguidos do *loess* e por último o *splines* de regressão cúbico.

Tabela 6: Percentual do erro quadrático mínimo completo, levando em conta o ajuste com o melhor parâmetro de suavizador, para todas a técnicas em relação à 1000 amostras.

Sub-Cenário	Tamanho	Desvio Padrão	Kernel	Loess	Sp. Reg. 1	Sp. Reg. 3	Pol. Cúbico
1	150	0.05	90.1%	6%	3.8%	0.1%	0%
2	150	0.10	71.1%	9.3%	16.3%	3.3%	0%
3	150	0.15	49.1%	13.5%	25.6%	11.8%	0%
4	250	0.05	91.9%	4.2%	3.5%	0.4%	0%
5	250	0.10	79.9%	7.6%	10.7%	1.8%	0%
6	250	0.15	61.9%	10.6%	20.3%	7.2%	0%
7	350	0.05	91.9%	3.7%	3.7%	0.7%	0%
8	350	0.10	82%	7.6%	8.7%	1.7%	0%
9	350	0.15	65.4%	10.3%	20.3%	4%	0%

Na Figura 9, observa-se que o suavizador *kernel* apresenta um comportaemnto mediano relativamente inferior aos demais suavizadores. Seguidos do *loess*, *splines* de regressão linear e por último o *splines* cúbico. Portanto, levando em conta a Tabela 6 e Figura 9, há indícios de que o suavizador com *kernel* gaussiano apresenta o melhor desempenho para ajustar os dados.

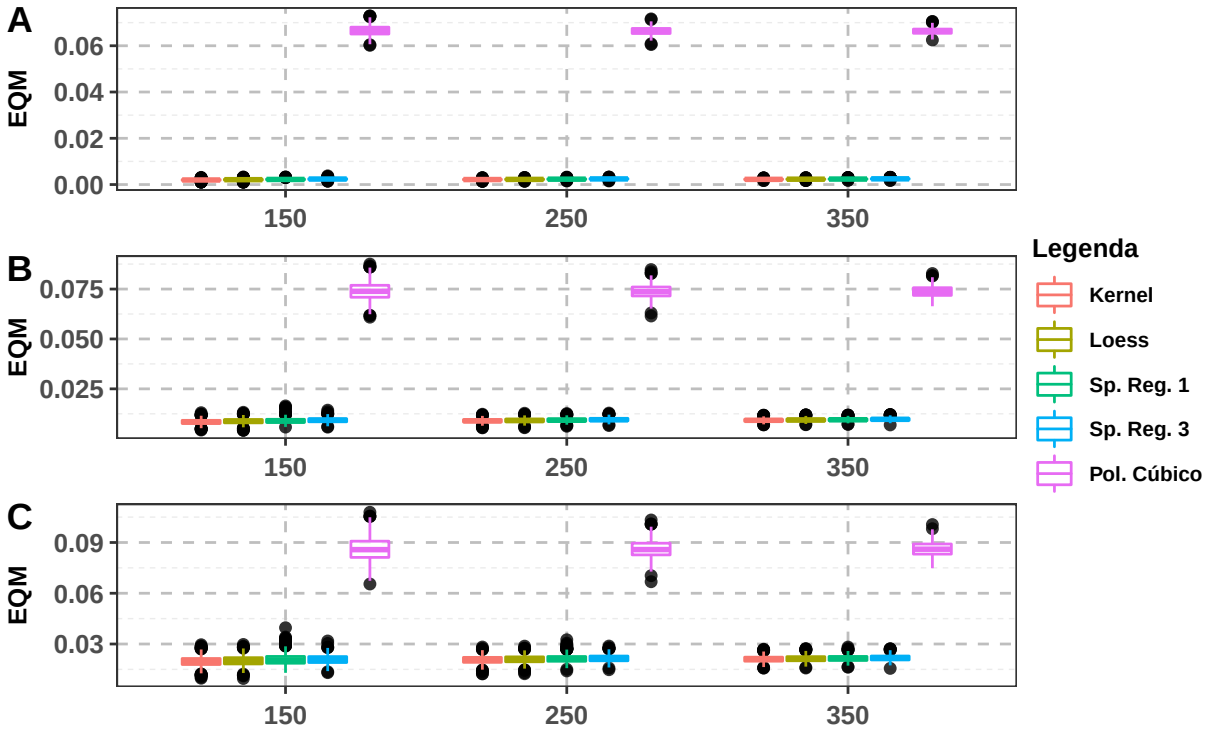

Figura 9: Comparação do erro quadrático para as 1000 amostras para cada suavizador por (A) DP = 0.05, (B) DP = 0.1 e (C) DP = 0.15

1.1.3 Considerações finais do estudo de simulação

Pode-se comparar as técnicas de suavização apresentadas neste trabalho em diversos cenários mostrando quais delas apresentam o melhor poder preditivo e quais delas ajustam melhor os dados. Segundo IZBICKI (2020), nem sempre o modelo que se adequa melhor aos dados, necessariamente irá obter o melhor pode preditivo. Verificou-se dentro das simulações que o suavizador *splines* de regressão cúbico apresentou o melhor desempenho de predição, avaliando-se por meio da métrica EQM_{LOOCV}, em ambos os cenários propostos (Cenário 1 e Cenário 2). Em contrapartida, ao avaliar os cenários por meio da métrica EQM_c, pode-se verificar que o suavizador com *kernel* gaussiano obteve o melhor desempenho.

2 Aplicação

Para esta aplicação serão empregadas as técnicas de suavização em dados reais. Os dados foram retirados do site NIST Standard Reference Database 140 [1] . É um estudo referente a expansão térmica de cobre. A variável resposta é o coeficiente de expansão térmica e a variável preditora é a temperatura em graus kelvin. Neste trabalho sera abordado um modelo com apenas uma covariável neste caso, sendo o modelo aditivo da seguinte forma

$$y = \alpha + f(X) + \epsilon,$$

onde os erros ϵ são independentes, com $E(\epsilon) = 0$ e $var(\epsilon) = \sigma^2$. A $f(X)$ é uma função univariada arbitrária, que será suavizada pelos métodos vistos até o momento.

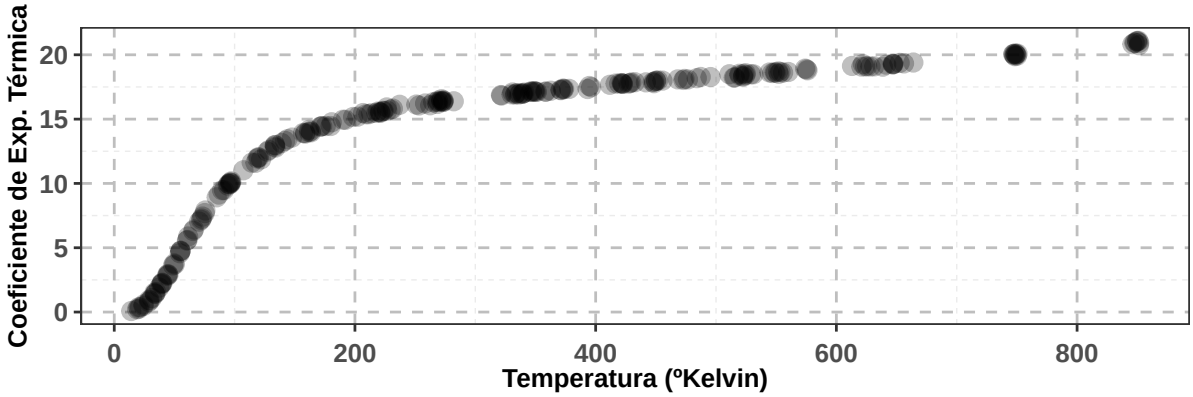

Figura 10: Gráfico de dispersão referente ao estudo de espansão térmica do cobre versus a temperatura em graus Kelvin.

Na Tabela 7, é apresentado o processo de seleção do melhor parâmetro, com seus repectivos parâmetros selcionados por meio da métrica EQM_{LOOCV}.

Tabela 7: Erro Quadrático Médio para os suavizadores Loess, Kernel e Spline Cúbico

Suavizador	Parâm. Suavizador	EQM
Kernel	0.1	0.0091
Loess	0.1	0.0078
Splines Grau 1	20.0	0.0078
Splines Grau 3	14.0	0.0065

Ressalta-se que o *splines* de regressão cúbico apresenta o menor EQM_{LOOCV}. Na Figura 11, são demonstrados os ajustes, levando em consideração os melhores parâmetros obtidos por meio do processo de validação cruzada. Visualmente, observa-se que as técnicas suavizadoras apresentam um bom ajuste, para captar a tendência de expansão térmica em relação a

[1] https://www.itl.nist.gov/div898/strd/index.html

temperatura. Como pode-se observar o ajuste paramétrico não consegue descrever o comportamento de forma adequada para estes dados.

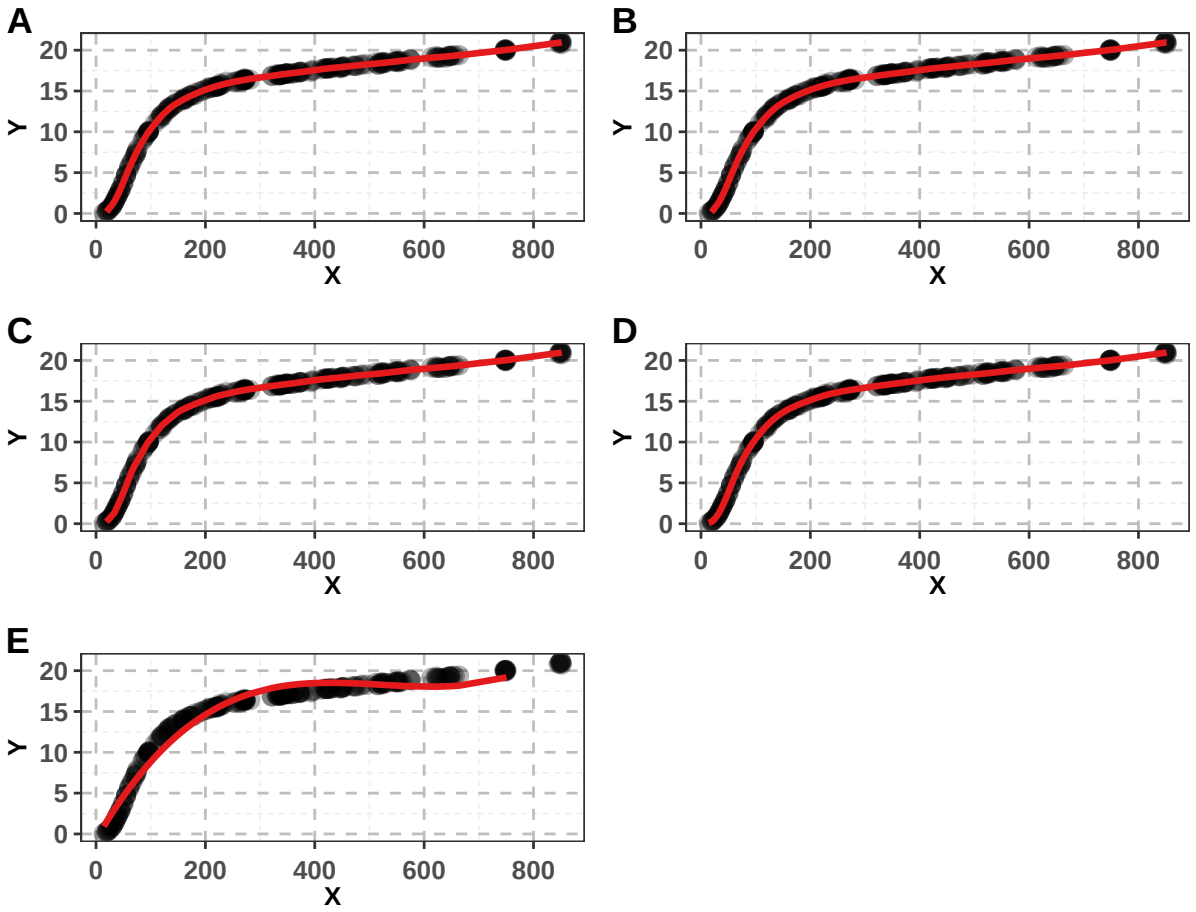

Figura 11: Comparação entre os ajustes, considerando parâmetros de suavização obtidos por meio da validação cruzada.

Ao comparar os $EQM_c's$, na Tabela 8, ressalta-se que a regressão polinomial obteve o pior resultado para esta métrica. Para as técnicas *kernel*, *loess* e *splines* de regressão linear apresentaram valore bem próximos entre si. Destaca-se o *splines* de regressão por obter o menor erro quadrático completo.

Tabela 8: Erro Quadrático Médio para os suavizadores Loess, Kernel e Spline Cúbico

Suavizador	Parâm. Suavizador	EQM
Kernel	0.1	0.000200
Loess	0.1	0.000211
Splines Grau 1	20	0.000211
Splines Grau 3	14	0.000171
Polinômio Cúbico		0.031052

Portanto, levando em consideração os resultado obtido, suavizador em *splines* de re-

gressão cúbico, apresenta o melhor desempenho tanto em relação ao poder de predição quanto melhor técnica para ajustar, o comportamento de expansão térmica em relação a temperatura.

3 Conclusão

Para investigar e modelar relações entre variáveis o modelo de regressão linear pode ser utilizado, porém quando essa relação não possui forma linear, uma alternativa é o uso de ferramentas que não impõem suposições paramétricas. Nesse contexto, existem técnicas de suavização que podem ser utilizadas, inclusive na estimação das funções do componente sistemático dos modelos aditivos.

As técnicas de suavização *kernel*, *loess* e splines de regressão, em particular os de grau um e grau três foram aprsentados, que são utilizadas para estimar as funções presentes em modelos aditivos. Fora introduzido um método para obtenção no melhor parâmetro suaviador, a fim de evitar sub ou super-ajuste, realizado por meio de método de validação cruzada. Duas métrias foram abordadas uma para verificar a qualiadade de predição dos ajustes (EQM_{LOOCV}) e a outra para verificar a qualiade do ajuste (EQM_c).

Para validar a metodologia estudada, foram realizadas análises em dados simulados e dados reais. Em dados simulados em diferentes cenários, foram observados os resultados em relação ao comportamento de técnicas de suavização, comparando os modelos obtidos, por meio das métricas introduzidas anteriormente. Ainda, verificou-se que o ajuste mais adequado para descrever o comportamento dos dados não obtém necessariamente o melhor poder preditivo.

Por fim, os métodos discutidos fora aplicados em dados reais, onde mais uma vez os suavizadores foram avaliados e verificou-se qual apresentou o melhor poder preditivo e qual representa de forma mais adequada os dados.

4 Referências

BUJA, A., HASTIE, T. & TIBSHIRANI, R. (1989). **Linear smoothers and additive models**. The Annals of Statistics, 17, 453-510.

CLEVELAND, W. S. (1979). **Robust locally weighted regression and smoothing scatterplots**. Journal of the American Statistical Association, 74, 829-836.

DELICADO, P., 2008 **Curso de Modelos no Paramétricos** p. 200.

EUBANK, R. L(1999) **Nonparametric Regression and Spline Smoothing**. Marcel Dekker, 2o edição. Citado na pág. 1, 2, 29

FAHRMEIR, L. & TUTZ, G. (2001) **Multivariate Statistical Modelling Based on Generalized Linear Models**. Springer, 2o edição. Citado na pág. 15

GREEN, P. J. & YANDELL, B. S. (1985) **Semi-parametric generalized linear models**. Lecture Notes in Statistics, 32:4455. Citado na pág. 15

GREEN P. J. & SILVERMAN B. W. (1994). **Nonparametric regression and generalized linear models: a roughness penalty approach**. Chapman & Hall, London.

HASTIE, T. J. & TIBSHIRANI, R. J. (1990). **Generalized additive models**, volume 43. Chapman and Hall, Ltd., London. ISBN 0-412-34390-8.

MONTGOMERY, D. C. & PECK, E. A. & VINING, G. G. **Introduction to Linear Regression Analysis**. 5th Edition. John Wiley & Sons, 2012.

IZBICK, R. & SANTOS, T. M. **Aprendizado de máquina: uma abordagem estatística**. ISBN 978-65-00-02410-4.

TEAM, R. CORE. R: **A language and environment for statistical computing**. (2013).