



**Universidade de Brasília
Departamento de Estatística**

**Um estudo sobre modelagem preditiva para a demanda de eletricidade no
Sudeste e Centro-Oeste**

Marco Aurélio Izidro Machado

Trabalho de Conclusão de Curso, apresentado para o Departamento de Estatística da Universidade de Brasília, como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2021**

Marco Aurélio Izidro Machado

**Um estudo sobre modelagem preditiva para a demanda de eletricidade no
Sudeste e Centro-Oeste**

Orientador: Prof. Dr. José Augusto Fiorucci

Trabalho de Conclusão de Curso, apresentado para o Departamento de Estatística da Universidade de Brasília, como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2021**

Agradecimentos

Agradeço aqueles que contribuíram direta e indiretamente na construção desse trabalho, em especial:

À Universidade de Brasília, por ter me proporcionado um ensino gratuito e de excelência.

Ao Departamento de Estatística e seu corpo docente, por terem garantido com todo o zelo a graduação de sublimidade.

Ao meu orientador, Prof. Dr. José Augusto Fiorucci, pelo tempo e contribuição à minha formação.

À minha família, por entenderem as minhas ausências, pela paciência, pelo carinho e por sempre torcerem pelo meu sucesso, mesmo que muitas vezes nem eu o visualize.

À minha namorada, pela inspiração de dedicação e energia para concluir esse projeto.

”O que eu faço, é uma gota no meio de um oceano. Mas sem ela, o oceano será menor.”

Madre Teresa de Calcutá

Resumo

A previsão da demanda por eletricidade é necessária para a programação em tempo real da geração de eletricidade. Tal processo deve ser eficiente para minimizar tanto o desperdício, bem como o risco de faltar eletricidade para a população. Esse tipo de série temporal apresenta múltiplas sazonalidade, geralmente com ciclos diário, semanal e anual. Este estudo tem como objetivo avaliar a capacidade preditiva dos principais modelos disponíveis na literatura, para tal, é considerada a série histórica da demanda de eletricidade na região Sudeste/Centro-Oeste do Brasil. A performance preditiva dos modelos é avaliada por meio de um processo conhecido como janela deslizante. Os resultados revelam que o modelo SARIMA de sazonalidades intra-diária e intra-semanal possui resultados competitivos frente a abordagens mais sofisticadas, como o TBATS.

Palavras-chave: Demanda de Energia, Previsão, Múltiplas Sazonalidades, Modelos SARIMA, Modelos de Alisamento, Modelos TBATS.

Sumário

Lista de siglas	7
1 Introdução	8
2 Material e Métodos	11
2.1 Dados da demanda de energia no Sudeste/Centro-Oeste	11
2.2 <i>Seasonal autoregressive integrated moving average</i> (SARIMA)	14
2.2.1 SARIMA com apenas sazonalidade semanal.	14
2.2.2 SARIMA com duas sazonalidades, intra-diária e semanal.	15
2.3 Regressão por termos de Fourier com erros ARIMA.	15
2.4 Alisamento exponencial Holt-Winters (HWT).	16
2.4.1 Alisamento exponencial HWT com uma sazonalidade semanal.	16
2.4.2 Alisamento exponencial HWT com duas sazonalidades, intra-diária e semanal.	17
2.5 TBATS	17
2.6 Validação Cruzada para séries temporais	21
3 Comparação Empírica	23
4 Conclusão	31
Referências	32

Lista de siglas

AIC	- Critério de informação de Akaike
ARIMA	- <i>Autoregressive integrated moving average</i>
ARMA	- <i>Autoregressive moving average</i>
HWT	- Holt-Winters
IC	- <i>Intraday cycle</i>
MAPE	- <i>Mean absolute percentage error</i>
MSTL	- Múltiplas sazonalidades, Tendência e Loess
MWh/h	- Megawatt-hora por hora
ONS	- Operador nacional do sistema elétrico
SARIMA	- <i>Seasonal autoregressive integrated moving average</i>
TBATS	- <i>Exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components</i>

1 Introdução

Na atualidade, é fácil assumir a presença da eletricidade disponível em qualquer lugar e a qualquer momento. Porém, como muitos outros serviços que são tomados como garantidos, existem complexos estudos e modelos estatísticos por trás, garantindo estabilidade e evitando o desperdício de recursos naturais e monetários. Segundo Lisi e Shah (2020), modelos acurados para a previsão de demanda e preços são cruciais no bom planejamento e operação dos sistemas da rede elétrica. Uma vez que eletricidade não pode ser armazenada em grandes quantidades e precisa ser entregue imediatamente ao consumidor final. Dessa forma, estimar incorretamente a demanda de energia implica em perdas financeiras para todos participantes do mercado. À vista disso, superestimar a demanda por energia leva à compra excessiva e desperdício de recursos, e subestimar faz necessário rebalancear o sistema a qualquer custo.

Também apontado por Lisi e Shah (2020), diferentes participantes do mercado podem ter interesse em diferentes horizontes de previsão. Previsões de algumas horas ou alguns dias, são chamadas de previsões a curto prazo e são usados para garantir a confiabilidade, eficiência nas operações da rede elétrica, gerenciamento de carga, programação e despacho de unidades geradoras. Previsões de médio prazo usam o horizonte de algumas semanas ou meses, normalmente, aplicadas para planejar manutenções, reabastecimentos ou cálculo de lucros. Por fim, as previsões de longo prazo com horizonte de meses ou anos, comumente se interessam nos máximos e mínimos para realizar planejamento e cálculos de risco.

Apesar das séries de demanda de eletricidade serem amplamente estudadas na literatura pela relevância monetária, essas são exemplo de múltiplos períodos sazonais, ao serem observadas múltiplas vezes todos os dias. Os padrões sazonais intra-diários são perceptíveis, juntamente com o comportamento intra-semanal e anual. De fato, Taylor (2010) em um estudo empírico, mostra que a inclusão da sazonalidade anual, juntamente com a sazonalidade intra-semanal e intra-diária podem melhorar as previsões de curto prazo para modelos de alisamento exponencial, *Seasonal autoregressive integrated moving average* (SARIMA) e de redes neurais aplicados às séries de demanda de eletricidade da França e Grã Bretanha dos anos de 2001 a 2007.

Diversos estudos usam de modelos univariados, como o modelo SARIMA, para prever a demanda de eletricidade a curto prazo (HAGAN, BEHR, 1987; MBAMALU, EL-HAWARY, 1993; TAYLOR, McSHARRY, 2007), outros estudos usam variáveis explicativas, como temperatura, tentando justificar os padrões sazonais identificados e realizar previsões (DARBELLAY, SLAMA, 2000; HOR et al., 2005; CANCELO, 2008). Em um estudo empírico, Taylor (2008) compara o resultado de ambas abordagens, e mostra que

a combinação das duas pode resultar em previsões ainda melhores.

Contudo, diversas variáveis explicativas podem ser usadas na previsão de séries temporais, como por exemplo o estudo de Hyde e Hodnett (1997) que considera a temperatura, horas de insolação, velocidade do vento e umidade do ar. Além dessas variáveis, Vu, Muttaqi e Agalgaonkar (2015) consideram também a população, Produto Interno Bruto (PIB), preço, evaporação e dias chuvosos na construção de um modelo de regressão multivariado. Deina (2020) propõe uma metodologia para auxiliar no processo de seleção das variáveis explicativas para incluir no modelo de previsão de energia elétrica.

Existem diferentes maneiras nas quais feriados e outros dias anômalos são tratados na literatura. Smith (2000) realiza previsões de curto prazo à demanda de eletricidade do estado de *New South Wales* na Austrália, utilizando regressão Bayesiana semi-paramétrica, e trata feriados como se tivessem o mesmo comportamento de domingos, por apresentarem comportamento semelhante na série.

Para a série de demanda de eletricidade Irlandesa, Hyde e Hodnett (1997) usa regressão linear com abordagem baseada em regras que diferenciam cada tipo de dia, calculadas extraíndo o componente estimado de influência do clima e então fazendo o valor médio das horas de cada tipo de dia, sendo dias da semana, feriados ou dias santos. Com a suposição de que a magnitude e o formato de cada tipo de dia seria diferente.

Cancelo (2008) realiza previsões com horizonte de 24 horas para a *Red Eléctrica de España*, o operador do sistema espanhol. Por meio de um modelo *Autoregressive integrated moving average* (ARIMA) com variáveis *dummy* distinguindo em dias normais, dias de semana, fins de semana e dias especiais.

Taylor e McSharry (2007) realiza um estudo empírico para comparar diversos modelos na previsão da demanda de eletricidade europeia, incluindo ARIMA, *Autoregressive moving average* (AR) periódico, Holt-Winters (HWT) de duas sazonalidades e um modelo proposto baseado em análise de componentes principais, o qual não apresenta resultado satisfatórios. Nesse estudo, os feriados são alisados ao substituir os valores de tais datas pela média das respectivas horas da semana anterior e posterior combinadas. Justificando que o operador do sistema elétrico usaria experiência própria em tais dias para sobrescrever as previsões nos dias de feriado. Hippert (2005) realiza alisamento semelhante ao estudar a hipótese de sobre-ajuste em redes neurais com dados do Rio de Janeiro, substitui valores de feriados, pelos valores do equivalente dia da semana anterior.

Srinivasan, Chang e Liew (1995) estudam o uso de redes neurais difusas na previsão da demanda de eletricidade, com ênfase em feriados e fins de semana. Com o uso de variáveis difusas para a temperatura, intensidade de chuvas e proximidade a feriados. Mostram que redes neurais podem obter boa performance para dia anômalos sem necessidade de alterar a série.

Em contraste, Arora e Taylor (2013) introduzem a abordagem baseada em regras aos modelos com tripla sazonalidade de alisamento exponencial HWT, SARIMA, redes neurais artificiais e alisamento exponencial por decomposição em valores singulares. Em análise empírica, Arora e Taylor (2013) concluem que a combinação dos modelos SARIMA e HWT, ambos adaptados para considerar dias anômalos, resultam nos melhores valores *Mean absolute percentage error* (MAPE) para a série de demanda de eletricidade da Grã Bretanha avaliada nos anos de 2001 a 2009.

Além dos modelos comumente usados na literatura para previsões univariadas, temos como objetivo comparar o modelo TBATS descrito por De Livera, Hyndman e Snyder (2011). TBATS é até então pouco explorado na literatura por ser relativamente novo, porém apresenta diversas vantagens. Brozyna et al. (2018) testa a capacidade do modelo TBATS, apontando a facilidade de gerar previsões detalhadas a longo prazo, em vista de não apresentar restrições quanto a sazonalidade. Diferente dos modelos ARIMA ou HWT, que suportam apenas períodos sazonais inteiros, com dificuldades para sazonalidades de longo prazo.

Comparamos empiricamente os modelos univariados SARIMA, regressão por termos de Fourier, alisamento exponencial de HWT e TBATS na estimação da série de demanda por eletricidade no Sudeste e Centro-Oeste. Identifica-se que o modelo SARIMA de sazonalidades intra-diária e intra-semanal obtém as melhores previsões de curto prazo. Evidencia-se que os ajuste automáticos dos modelos falharam em identificar a sazonalidade anual presente, em vista da literatura considerada, conclui-se que os resultados obtidos servem como passo inicial para uma análise mais completa, nos quais variáveis explicativas podem auxiliar na modelagem das sazonalidades complexas presentes.

2 Material e Métodos

2.1 Dados da demanda de energia no Sudeste/Centro-Oeste

Dados da demanda de eletricidade nacional são disponíveis publicamente no site do Operador Nacional do Sistema elétrico (ONS). Foram trabalhados os dados de demanda de eletricidade do Sudeste/Centro-Oeste em megawatts-hora por hora (MWh/h), avaliados 24 vezes ao dia, dentre os anos de completos de 2010 a 2020. Os feriados presentes na série foram substituídos pelos dados do dia equivalente da semana anterior, em vista que esses *outliers* viriam a influenciar a estimativa dos parâmetros. Como consequência, as previsões resultantes não são aplicáveis apenas na previsão de dias normais. Na prática, dias anômalos, como feriados, não podem ser ignorados, Arora e Taylor (2013) propõem uma alternativa que pode ser usada, em estudos futuros, para complementar os resultados aqui apresentados.

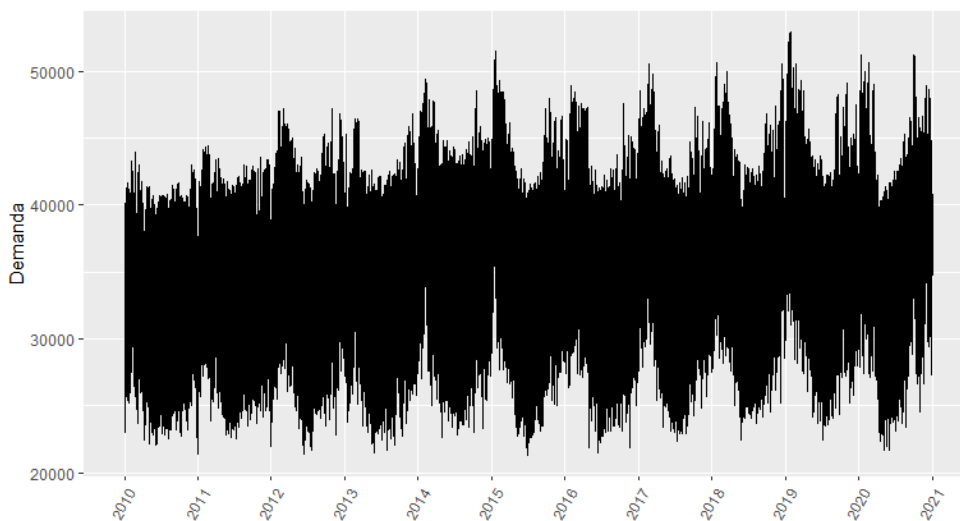


Figura 1: Demanda de energia em MWh/h . No eixo Y há a demanda de energia elétrica e no eixo X há os anos da análise.

Os dados trabalhados são ilustrados na Figura 1, com as datas especiais já substituídas. Fica aparente a existência de sazonalidade anual. Em vista da oscilação observada, que apresenta picos em torno da virada de cada ano e vales no período central de cada ano.

Nas Figuras 2 e 3, são ilustrados as séries de demanda máxima e mínima para cada mês com o passar dos anos. Tanto a demanda máxima como a mínima apresentam comportamento semelhante, com maior variabilidade para os meses próximos à virada do ano e menor variabilidade para os meses próximos ao meio do ano.

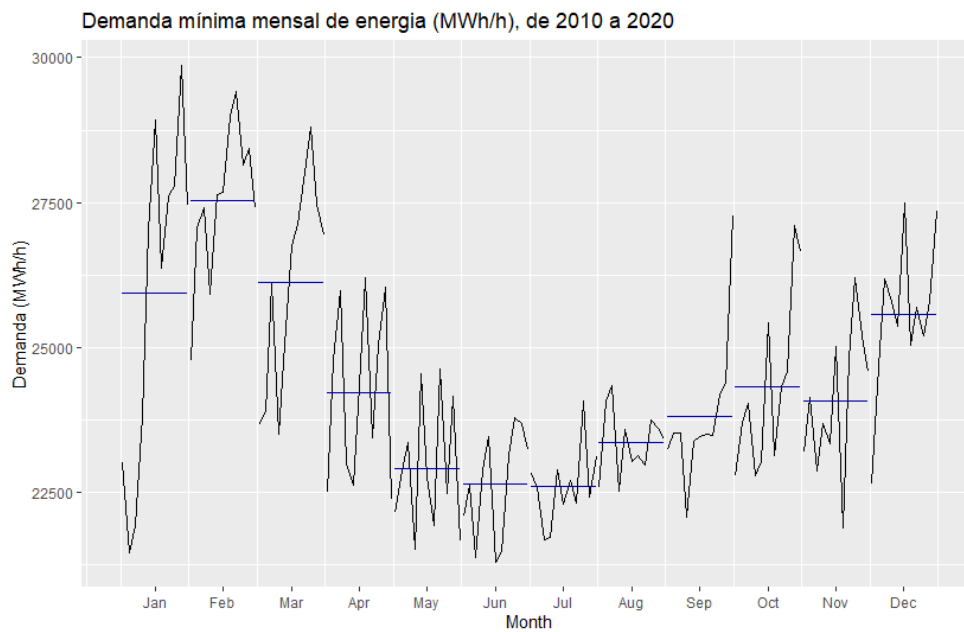


Figura 2: As linhas pretas representam a demanda mínima por eletricidade na região Sudeste/Centro-Oeste de cada mês entre os anos de 2010 a 2020 e as linhas azuis representam a média dos respectivos meses, os quais são indicados no eixo x .

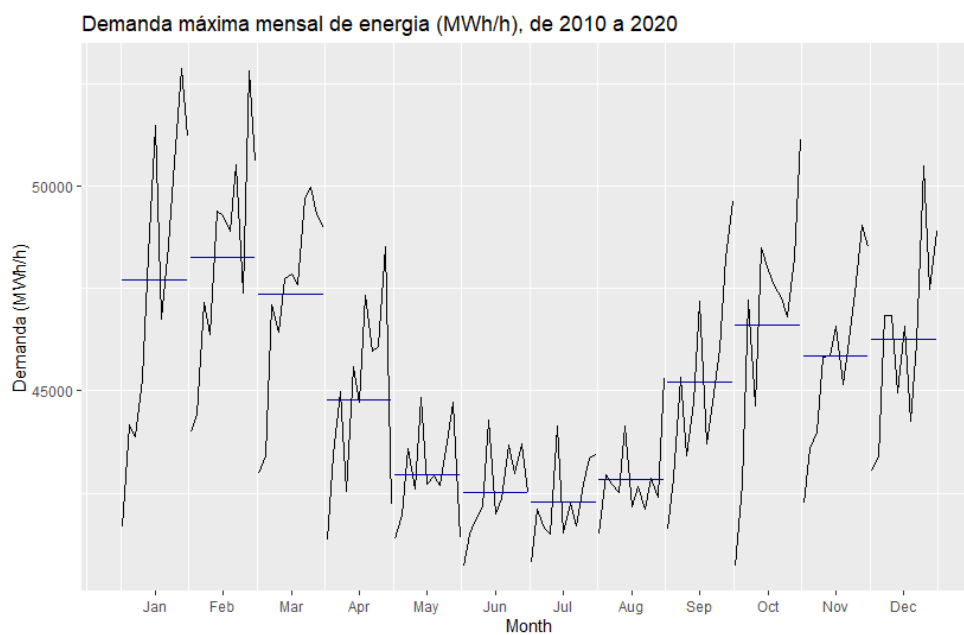


Figura 3: As linhas pretas representam a demanda máxima por eletricidade na região Sudeste/Centro-Oeste de cada mês entre os anos de 2010 a 2020 e as linhas azuis representam a média dos respectivos meses, os quais são indicados no eixo x .

Na Figura 4, observa-se a demanda de eletricidade das três primeiras semanas de outubro de 2018, como exemplo do comportamento intra-semanal presente nos dados. Para os dias úteis, observam-se vales em torno das 3 horas e picos em torno das 15 horas, em contraste, os fins de semana, que apresentam demanda reduzida e comportamento distinto.

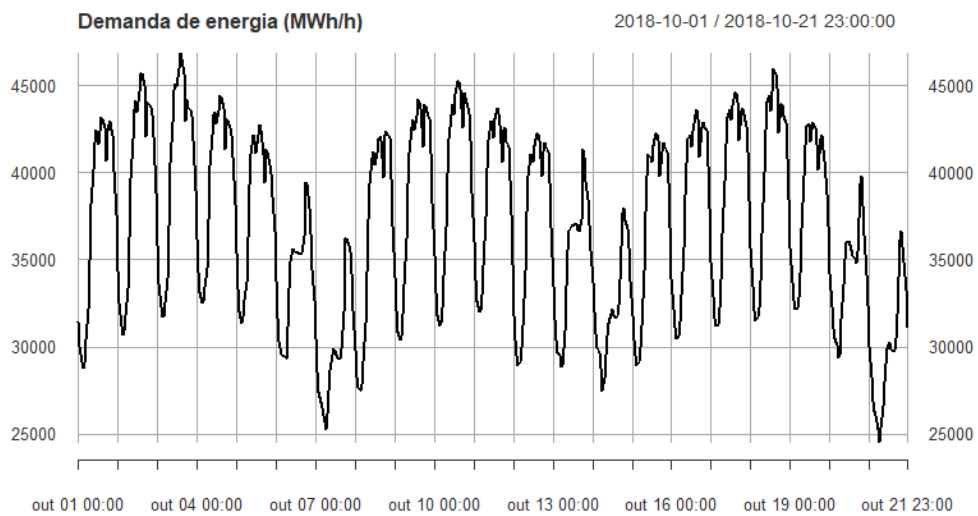


Figura 4: Intervalo da demanda de energia em *MWh/h*, referente as três primeiras semanas do mês de outubro de 2018

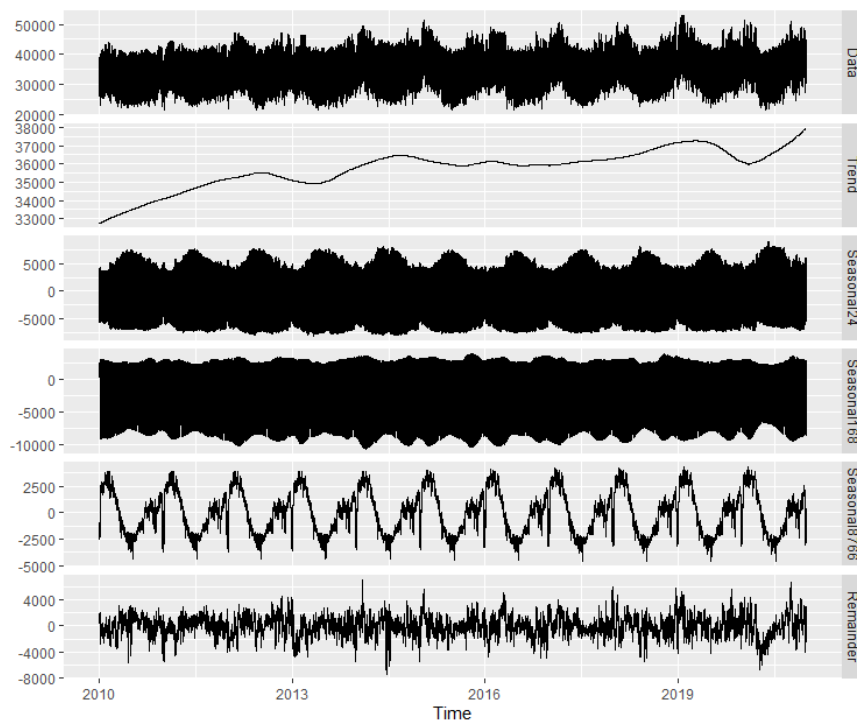


Figura 5: Decomposição MSTL para Sudeste/Centro-Oeste

Além disso, uma vez que se trata de dados avaliados de hora em hora, existe a possibilidade da presença de sazonalidades intra-diária, intra-semanal e anual. Com a decomposição de Múltiplas Sazonalidades, Tendência e Loess (MSTL), ilustrada na Figura 5, foram identificados os comportamentos sazonais: *Seasonal24*, *Seasonal168* e *Seasonal8766*. Os quais são, respectivamente, os componentes sazonais: diária de 24 observações, semanal de 168 observações e anual de 8766 observações identificadas por este método. A tendência presente é crescente, porém o crescimento de 5000 MWh/h no

período de 10 anos é pequeno, evidenciando que previsões de curto prazo podem ter boa performance sem considerar a tendência global. Vale notar, que parte da sazonalidade anual aparenta ter sido captada nas sazonalidades intra-diária e intra-semanal, porém esta decomposição ainda serve como ponto de partida para interpretação do comportamento dos dados, em que os modelos que serão ajustados podem tratar melhor das sazonalidades em questão.

2.2 *Seasonal autoregressive integrated moving average* (SARIMA)

Essa seção apresenta os modelos da família SARIMA com diferentes tipos de modelagem para a componente sazonal.

2.2.1 SARIMA com apenas sazonalidade semanal.

Como argumentado por Taylor (2010), um pesquisador poderia optar por um modelo SARIMA de apenas sazonalidade semanal, se a modelagem for boa o suficiente. O modelo SARIMA com uma componente sazonal é descrito na equação,

$$\phi_p(L)\Omega_{P_2}(L^{S_2})(y_t - a - bt) = \theta_q(L)\Psi_{Q_2}(L^{S_2})\varepsilon_t \quad (2.2.1)$$

em que y_t é a observação no momento t ; a é uma constante; b é o coeficiente linear da tendenciada; ε_t é ruído branco (erro); L é um operador de atraso; ϕ_p , Ω_{P_2} , θ_q e Ψ_{Q_2} são polinômios de ordens p , P_2 , q e Q_2 .

A ordem dos polinômios são ajustadas conforme necessário para obter um melhor ajuste, descritos na forma $SARIMA(p,d,q)(P,D,Q)[m]$ onde d e D indicam respectivamente as ordens dos atrasos L e L^{S_2} e m indica a frequência da respectiva sazonalidade S .

Com a função *auto.arima* do pacote *forecast* no *software* R, foi ajustado um primeiro modelo $SARIMA(5,0,3)[1](0,1,0)[168]$, visando minimizar o critério de informação de Akaike (AIC). Modelos semelhantes a esse, não avaliados pelo método automático, foram testados manualmente com a função *msarima* do pacote *smooth*. O modelo final escolhido foi $SARIMA(5,0,3)[1](0,1,1)[168]$, com AIC 0,72% menor.

2.2.2 SARIMA com duas sazonalidades, intra-diária e semanal.

Modelos SARIMA simples, podem ser estendidos para comportarem múltiplas sazonalidades como indicado por Box et al. (1994, p. 333) apud Taylor (2010). O modelo SARIMA de duas sazonalidades é descrito na equação,

$$\phi_p(L)\Phi_{P_1}(L^{S_1})\Omega_{P_2}(L^{S_2})(y_t - a - bt) = \theta_q(L)\Theta_{Q_1}(L^{S_1})\Psi_{Q_2}(L^{S_2})\varepsilon_t \quad (2.2.2)$$

em que os termos adicionados a equação (2.2.1) para comportar sazonalidade múltipla foram: Φ_{P_1} e Θ_{Q_1} ; polinômios de ordens P_1 e Q_1 habilitando a modelagem intra-diária.

Com a função *msarima* do pacote *smooth* no *software* R, foi selecionado manualmente o modelo SARIMA(5,0,3)[1](1,0,1)[24](0,1,1)[168] em consequência as autocorrelações dos resíduos observados no modelo anterior.

2.3 Regressão por termos de Fourier com erros ARIMA.

Análogo ao modelo SARIMA de duas sazonalidades, pode-se substituir em (2.2.2) S_1 por S_3 . Onde S_3 indica o atraso (*lag*) anual, obtendo um modelo SARIMA de sazonalidade intra-semanal e anual.

Porém, segundo Hyndman (2010), sazonalidades de longo período apresentam performance computacional ruim e não fazem muito sentido. Isso por implicar na comparação da demanda avaliada contra a demanda de exatamente um ano antes, no qual o padrão sazonal pode apresentar comportamento errático.

Nos casos de longos períodos sazonais, Hyndman (2010) indica utilizar séries de Fourier para modelar as sazonalidades desejadas e ajustar um modelo ARIMA para contabilizar a autocorrelação dos resíduos. A abordagem consiste em utilizar termos de Fourier, os quais, são escritos como $y_t = a + \sum_{k=1}^K [\alpha_k \sin(2\pi kt/m) + \beta_k \cos(2\pi kt/m)] + N_t$ onde a é uma constante, a frequência m da sazonalidade não precisa ser um valor inteiro, N_t representa o processo ARIMA e K é a ordem escolhida, onde valores maiores de K permitem modelar séries mais complexas. Assim, a adição de sazonalidades para o ajuste de um modelo, pode ser feita incluindo um termo de Fourier com a respectiva sazonalidade m e ordem K .

Ao ajustar modelos de uma, duas e três sazonalidades com a função *auto.arima*, definindo $K = 7$ para todas sazonalidades, os modelos obtidos são descritos na Tabela 1.

Tabela 1: Componente ARIMA dos modelos de regressão por termos de Fourier

Sazonalidades	Erros ARIMA
Semanal	ARIMA(3,1,4)
Diária e Semanal	ARIMA(3,1,2)
Semanal e Anual	ARIMA(5,1,4)
Tripla	ARIMA(3,1,2)

2.4 Alisamento exponencial Holt-Winters (HWT).

Métodos de previsão por alisamento exponencial consistem em ponderar as observações anteriores com os pesos decaindo exponencialmente, para que as observações recentes tenham influência maior na previsão do modelo. O alisamento exponencial de HWT se diferencia pela presença do termo de tendência. Uma limitação desse método é a maneira limitada como trata sazonalidades. Hyndman (2010) comenta esse problema, indicando que uma sazonalidade de período m utiliza de $m - 1$ parâmetros, dificultando a estimação para m grande. Em decorrência dessa limitação, foram estimados modelos com apenas sazonalidade intra-diária ou semanal.

2.4.1 Alisamento exponencial HWT com uma sazonalidade semanal.

O modelo de alisamento exponencial HWT com apenas uma componente sazonal, semelhante ao modelo SARIMA de uma sazonalidade, poderia, teoricamente, apresentar resultados bons o suficiente como mencionado por Taylor (2010). Tal formulação é apresentada nas equações de (2.4.1) à (2.4.3)

$$\hat{y}_t(k) = l_t + W_{t-s_2+k} + \phi^k(y_t - (l_{t-1} + W_{t-s_2})) \quad (2.4.1)$$

$$l_t = \lambda(y_t - W_{t-s_2}) + (1 - \lambda)l_{t-1} \quad (2.4.2)$$

$$W_t = \omega(y_t - l_{t-1}) + (1 - \omega)W_{t-s_2} \quad (2.4.3)$$

em que, l_t é chamado de *level* e é usado para fazer a previsão pontual em t ; W_t é o termo de crescimento (tendência) obtido a partir de um alisamento exponencial dos níveis passados; λ e ω são parâmetros de alisamento; k se refere ao número de passos a frente da origem de previsão t ; assim $\hat{y}_t(k)$ é a previsão k passos a frente do momento t ; ϕ é o coeficiente do termo que ajusta o erro por auto-correlação de primeira ordem. Tal modelo foi ajustado com a função *HoltWinters* do pacote *stats*.

2.4.2 Alisamento exponencial HWT com duas sazonalidades, intra-diária e semanal.

O método de HWT foi adaptado por Taylor (2003) para conter mais de um termo de sazonalidade. Assim, sendo uma extensão do método demonstrado anteriormente, o qual é representado pelas equações de (2.4.4) à (2.4.7),

$$\hat{y}_t(k) = l_t + d_{t-s_1+k} + W_{t-s_2+k} + \phi^k(y_t - (l_{t-1} + d_{t-s_1} + W_{t-s_2})) \quad (2.4.4)$$

$$l_t = \lambda(y_t - d_{t-s_1} - W_{t-s_2}) + (1 - \lambda)l_{t-1} \quad (2.4.5)$$

$$d_t = \delta(y_t - l_{t-1} - W_{t-s_2}) + (1 - \delta)d_{t-s_1} \quad (2.4.6)$$

$$W_t = \omega(y_t - l_{t-1} - d_{t-s_1}) + (1 - \omega)W_{t-s_2} \quad (2.4.7)$$

em que o termo d_t é o índice, que representa a sazonalidade intra-diária; W_t é responsável pelo índice da sazonalidade semanal; δ é mais um parâmetro de alisamento. Tal modelo foi ajustado com a função *dshw* do pacote *forecast*.

Semelhante ao caso dos modelos SARIMA, sazonalidades de período longo demandam de muitos parâmetros e, conseqüentemente, tempo computacional. Assim, foi omitido o uso de sazonalidades anuais nos modelos de Alisamento exponencial HWT.

2.5 TBATS

Proposto por De Livera, Hyndman e Snyder (2011), se trata de uma modelagem para séries temporais com: termos **T**rigonométricos para sazonalidade, transformação de **B**ox-cox para heterogeneidade, erros *Autoregressive moving average* (**A**RMMA), **T**endência via alisamento exponencial e **S**azonalidade múltipla; podendo trabalhar com sazonalidade de períodos não inteiros. O modelo TBATS é representado pelo conjunto das equações (2.5.1) à (2.5.10),

$$y_t = \text{série original no tempo } t \quad (2.5.1)$$

$$y_t^{(w)} = \begin{cases} (y_t^w - 1)/w & \text{se } w \neq 0 \\ \log y_t & \text{se } w = 0 \end{cases} \quad (2.5.2)$$

$$y_t^{(w)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^M s_{t-m_i}^{(i)} + e_t \quad (2.5.3)$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha e_t \quad (2.5.4)$$

$$b_t = (1 - \phi)b + \phi b_{t-1} + \beta e_t \quad (2.5.5)$$

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)} \quad (2.5.6)$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} e_t \quad (2.5.7)$$

$$s_{j,t}^{*(i)} = -s_{j,t-1}^{(i)} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} e_t \quad (2.5.8)$$

$$\lambda_j^{(i)} = 2\pi j/m_i \quad (2.5.9)$$

$$e_t = \sum_{i=1}^p \phi_i e_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (2.5.10)$$

em que $y_t^{(w)}$ em (2.5.2) é o resultado da transformação de Box-Cox com parâmetro ω . As equações (2.5.3)-(2.5.5) se referem a um modelo HWT com tendência b_t do tipo *damped* (amortecida) e parâmetro de alisamento ϕ ; M é o número de componentes sazonais, onde m_1, m_2, \dots, m_M são os tamanhos dos ciclos sazonais.

As equações (2.5.6)-(2.5.9) apresentam a modelagem da sazonalidade por termos de Fourier, onde, k_i denota o número de funções harmônicas requeridas para cada componente. Por fim, na equação (2.5.10) a autocorrelação da componente de erro e_t é ajustada por um modelo ARMA(p,q).

O modelo TBATS pode ser ajustado com a função *tbats* do pacote *forecast*. Os parâmetros dos ajustes realizados são apresentados na Tabela 2.

Tabela 2: Parâmetros dos modelos TBATS calculados

Sazonalidade	ω	p	q	ϕ	m_1	k_1	m_2	k_2	m_3	k_3
Semanal	0	5	4	-			168	7		
Diária e Semanal	0,016	0	0	0.8	24	7	168	6		
Semanal e Anual	0,053	4	2	1			168	14	8766	3
Tripla	0,046	1	0	-	24	7	168	5	8766	6

Nas Figuras de 6 a 9 observam-se as decomposições dos modelos TBATS ajustados. Com os dados observados reescalados pelo parâmetro ω de cada modelo, a componente de *level*, a componente de tendência, identificada apenas nos modelos de duas

sazonalidades, e as respectivas componentes sazonais especificadas. No eixo x é representado o tempo, em função do número de anos desde o início da série estudada.

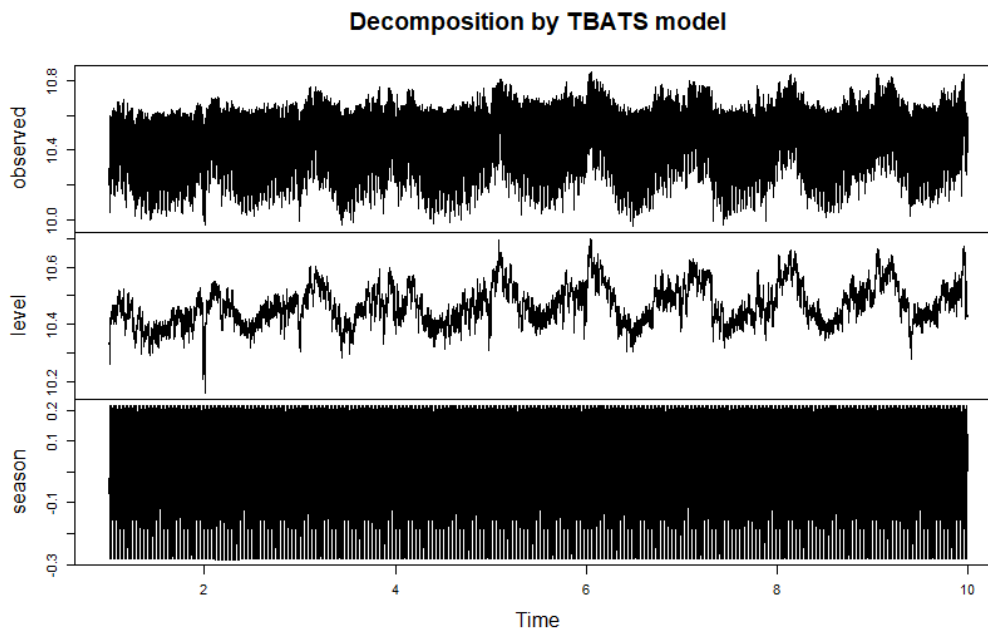


Figura 6: Decomposição realizada pela modelo TBATS de sazonalidade intra-semanal, apresentando o logaritmo da série de demanda de eletricidade em cima, seguido pela componente de *level* e a componente de sazonalidade intra-diária. No eixo x são representados os anos desde o início da série.

Na Figura 6, referente ao modelo de sazonalidade semanal, observamos o logaritmo dos valores originais, em vista do $\omega = 0$ no modelo, seguidas pela componente de *level* que aparenta captar parte da sazonalidade anual e a componente de sazonalidade intra-semanal.

A Figura 7, referente ao modelo de sazonalidade intra-diária e intra-semanal, ilustra os dados observados reescalados por $\omega = 0,016$, seguido pelas componentes de *level*, tendência, sazonalidade intra-diária e sazonalidade intra-semanal. Tanto a componente de *level*, como a componente de sazonalidade diária apresenta ter captado parte da sazonalidade anual presente na série, a componente de *level* oscila com os mesmos picos e vales identificados no período anual e a componente de sazonalidade intra-diária apresenta maior variabilidade no período central do ano.

Na decomposição ilustrada pela Figura 8, observamos os dados reescalados por $\omega = 0,053$, as componentes de *level*, tendência e sazonalidades intra-semanal e anual. A componente de *level* apresenta comportamento oposto a sazonalidade anual identificada. A componente de sazonalidade semanal apresenta ter captado parte da sazonalidade anual, com maior oscilação nos períodos do meio do ano.

Para a decomposição do modelo de sazonalidade tripla na Figura 9, são apresentados os dados de treinamento reescalados por $\omega = 0,046$, seguido da componente de

level e as respectivas sazonalidades intra-diária, intra-semanal e anual. Visualmente, a componente de sazonalidade anual não aparenta ter captado um padrão bem estabelecido.

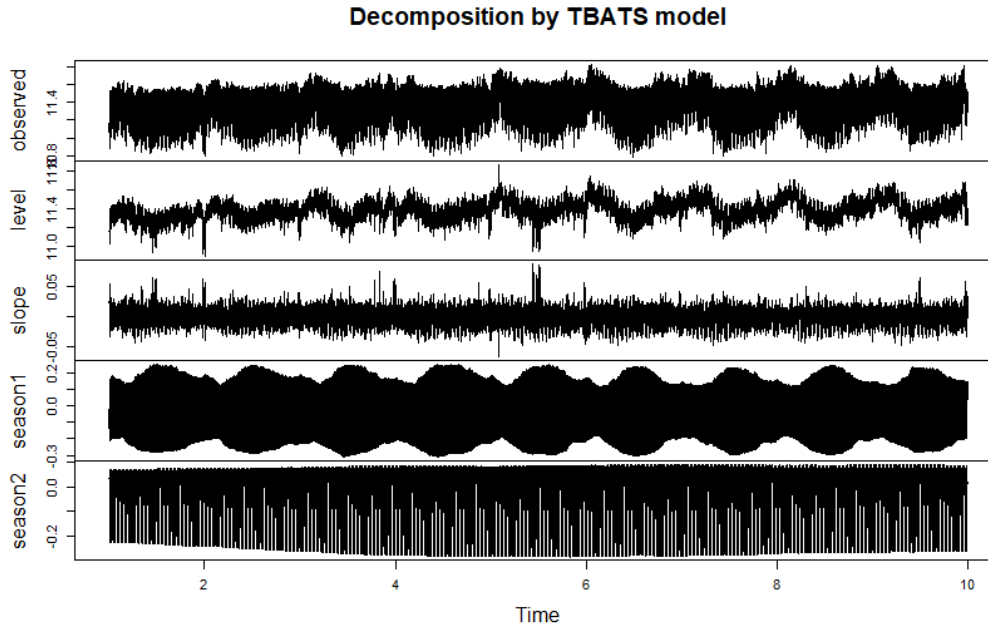


Figura 7: Decomposição realizada pela modelo TBATS de sazonalidade intra-diária e intra-semanal, apresentando a série de demanda de eletricidade reescalada por $\omega = 0,016$, seguido pelas componente de *level*, tendência, sazonalidade intra-diária e sazonalidade intra-semanal. No eixo x são representados os anos desde o início da série.

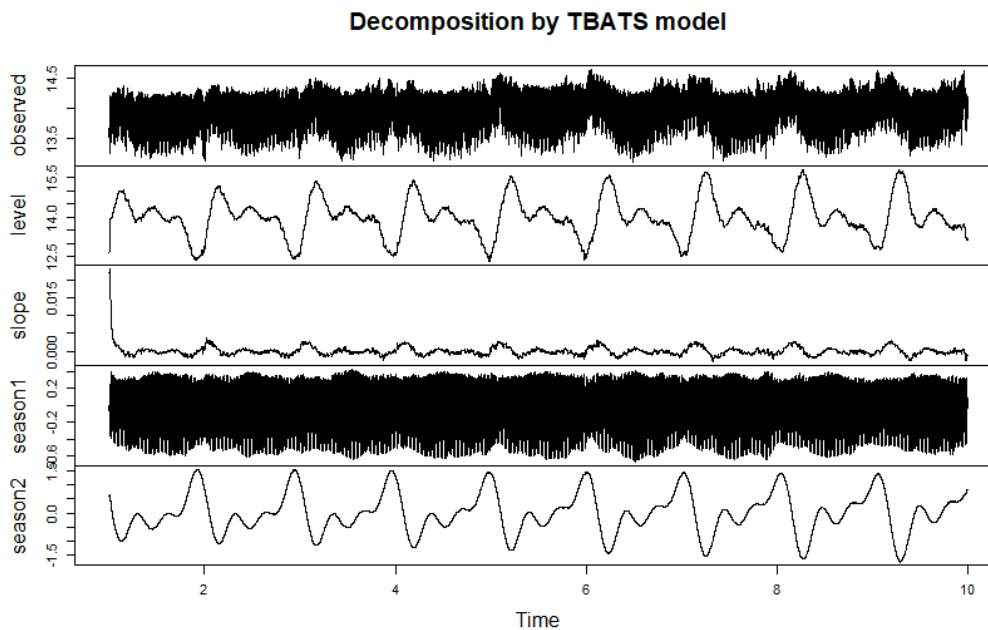


Figura 8: Decomposição realizada pela modelo TBATS de sazonalidade intra-semanal e anual, apresentando a série de demanda de eletricidade reescalada por $\omega = 0,053$, seguido pelas componente de *level*, tendência, sazonalidade intra-semanal e sazonalidade anual. No eixo x são representados os anos desde o início da série.

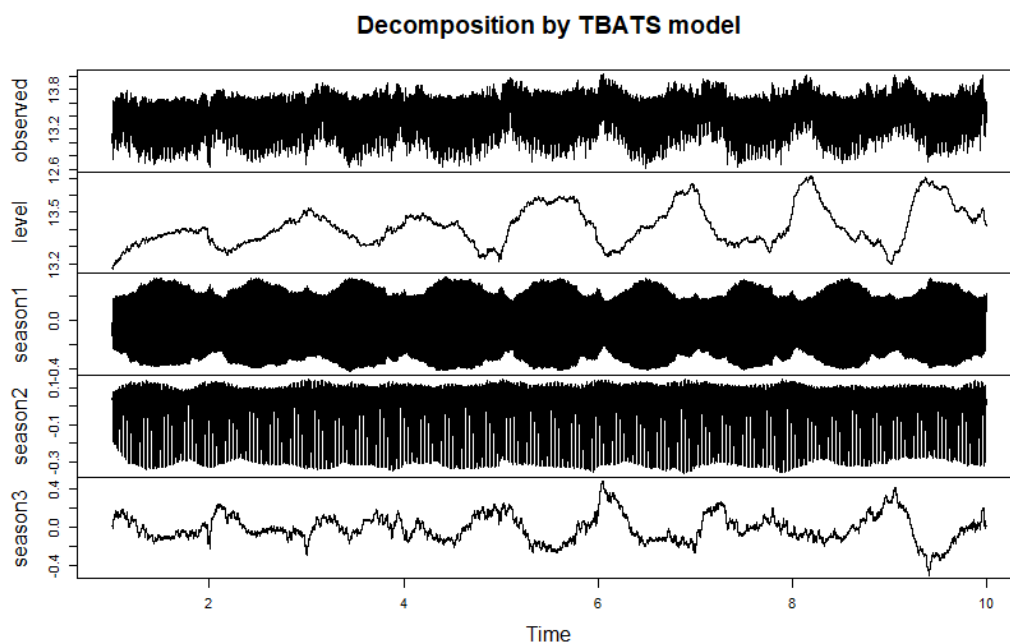


Figura 9: Decomposição realizada pela modelo TBATS de sazonalidade tripla, apresentando a série de demanda de eletricidade reescalada por $\omega = 0,046$, seguido pelas componente de *level*, e sazonalidades intra-diária, intra-semanal e anual. No eixo x são representados os anos desde o início da série.

2.6 Validação Cruzada para séries temporais

Os modelos estimados foram treinados com dados da demanda de eletricidade dos anos completos de 2010 a 2018, reservando os anos completos de 2019 e 2020 para testes.

Tais testes foram realizados via validação cruzada, semelhante aos estudos de Taylor (2010) e Arora e Taylor (2013), nos quais os n dados de treinamento são inicialmente usados para prever um horizonte $h = 168$, equivalente a uma semana de observações e avaliados em relação ao MAPE. Em seguida, $n + 1$ observações são consideradas para novamente prever 168 pontos e calcular o MAPE de cada um. Esse processo é repetido até que a ultima semana disponível no banco de dados tenha sido estimada, totalizando 17377 repetições. Consequentemente, as observações do banco de testes não pertencentes a primeira ou última semanas foram estimadas 168 vezes, uma para cada horizonte de previsão.

Isolar parte do banco de dados como banco de testes não é necessário quando se usa validação cruzada. Bergmeir (2012) argumenta que o uso validação cruzada resulta em modelos mais robustos, em parte por permitir usar o banco completo na estimação dos modelos. Bergmeir (2018) demonstra que validação cruzada pode ser aplicada a todos modelos de séries temporais, nos quais os preditores são os valores defasados, como todos modelos aqui analisados.

Contudo, os ajustes obtidos não sucederam em eliminar a autocorrelação dos resíduos, violando a suposição necessária para calcular intervalos de confiança, limitando a avaliação para apenas previsões pontuais. Assim, foram usados apenas valores desconhecidos para os testes, mostrando que as previsões ainda podem ser aproveitadas no planejamento da distribuição de energia, apesar da possibilidade de existência de ajustes melhores.

3 Comparação Empírica

Com o objetivo de identificar o modelo com as melhores previsões da demanda de energia elétrica no Sudeste e Centro-Oeste brasileiro, os modelos foram ajustado com funções automatizadas no *software* R, por conta do grande número de parâmetros tal processo seria inviável manualmente. Utilizado em seguida o método de validação cruzada para séries temporais na geração de previsões pontuais e calculando o respectivo MAPE para cada horizonte de previsão, o qual é definido pela seguinte equação

$$MAPE = \frac{100}{N} \sum_{t=1}^N \left| \frac{A_t - F_t}{A_t} \right|.$$

O calculo do MAPE, é dado pela média dos erros relativos absolutos de previsão. Onde N se refere ao número de vezes que cada observação foi estimada, A_t é o valor de demanda verdadeiro e F_t é o respectivo valor previsto. O valor é então multiplicado por 100 para que o resultado seja dado em porcentagem. Onde valores próximos de zero são preferíveis.

MAPE, apesar de comum na literatura, apresenta limitações. Essa é uma estatística que não pode ser usada para dados que passem pelo valor zero, pro conta da divisão por A_t . Consequentemente, para valores positivos, previsões muito baixas não podem exceder 100%, porém, previsões altas podem. Tofallis (2015) mostra que a seleção de modelos pela estatística MAPE, leva sistematicamente a seleção de previsões com valores mais baixos. Os dados aqui trabalhados devem ser pouco afetados por tais problemas, em vista que são estritamente positivos. Tofallis (2015) propõe o uso do logaritmo da razão de precisão como alternativa superior ao MAPE. Contudo, os resultados em MAPE são apresentados com o intuito de comparar a qualidade dos resultados em relação aos outros estudos disponíveis na literatura.

É apresentada a comparação para as diferentes configurações sazonais de cada método de modelam de séries temporais testados. Em seguida, são comparados os modelos de melhor performance.

Nas Figuras 10 e 11, pode-se comparar o resultado da validação cruzada para os modelos de regressão por séries harmônicas de Fourier com erros ARIMA, observa-se que as previsões para horizontes curtos de previsão tem performance pior e a adição de sazonalidades resultou em melhorias perceptíveis para esse método. Porém os resultados obtidos são superados pelos demais modelos avaliados.

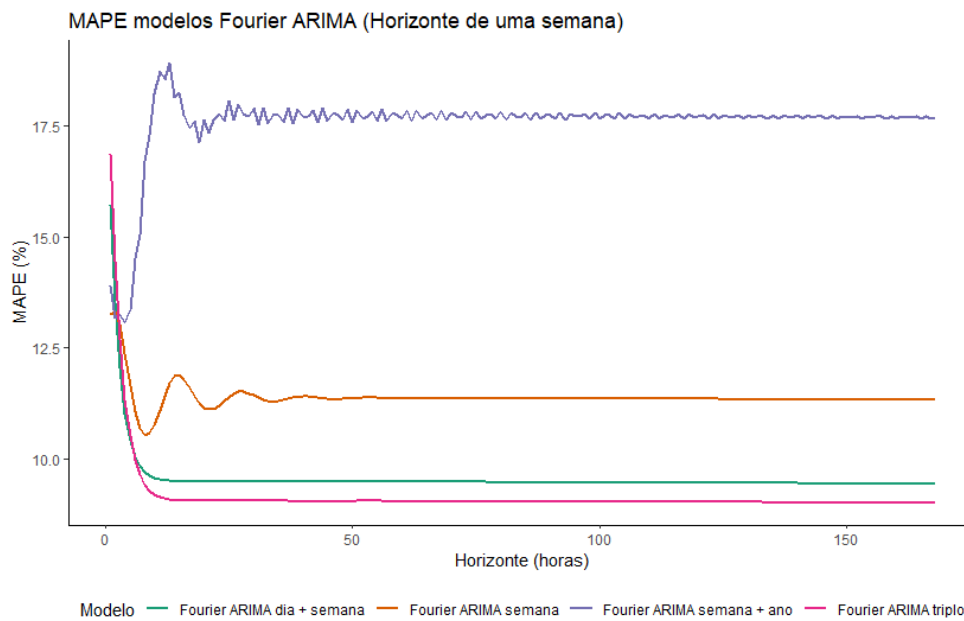


Figura 10: MAPE referente aos modelos de regressão por termos de Fourier com erros ARIMA, com horizonte de uma semana de previsões.

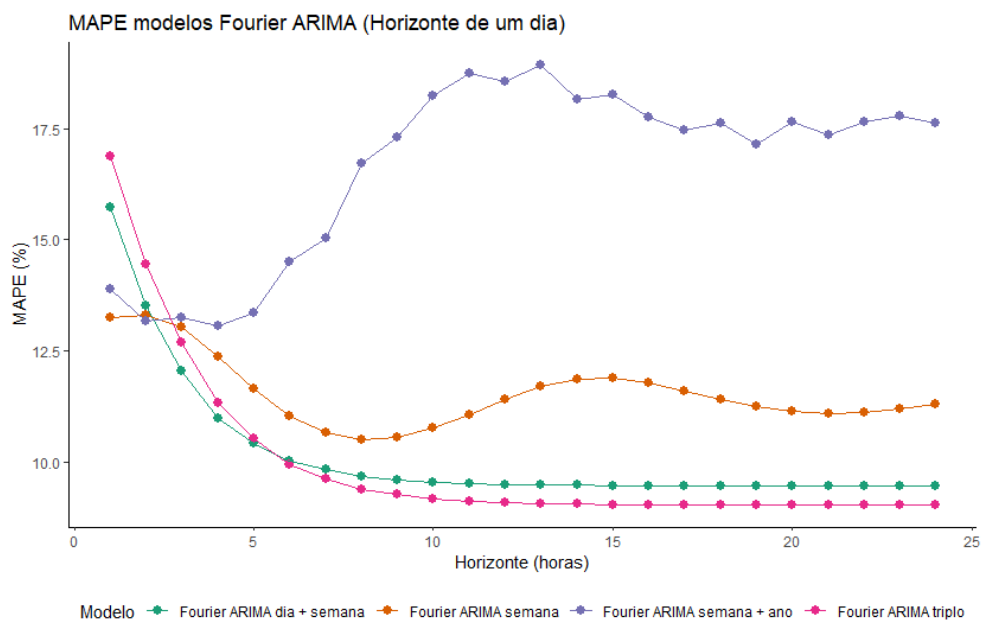


Figura 11: MAPE referente aos modelos de regressão por termos de Fourier com erros ARIMA, com horizonte de um dia de previsões

O MAPE dos modelos SARIMA com horizonte de uma semana, ficou próximo de 5% como ilustrado na Figura 12 e abaixo de 3% para horizonte de 24 horas, como ilustrado na Figura 13. Com MAPE mínimo de 0,5467% para no modelo de SARIMA de sazonalidade dupla.

Evidencia-se que o modelo SARIMA de sazonalidade dupla supera a performance do modelo de sazonalidade única para todos horizontes testados, apesar de ambos apre-

sentarem resultados bons. Indicando que o modelo SARIMA de tripla sazonalidade deve ser considerado, apesar da significativa demanda computacional para ser ajustado.

A diferença média entre os dois modelos foi apenas de 0,19%, mas no contexto da demanda de eletricidade do Sudeste e Centro-Oeste, tal diferença implica dezenas de MWh/h de energia.

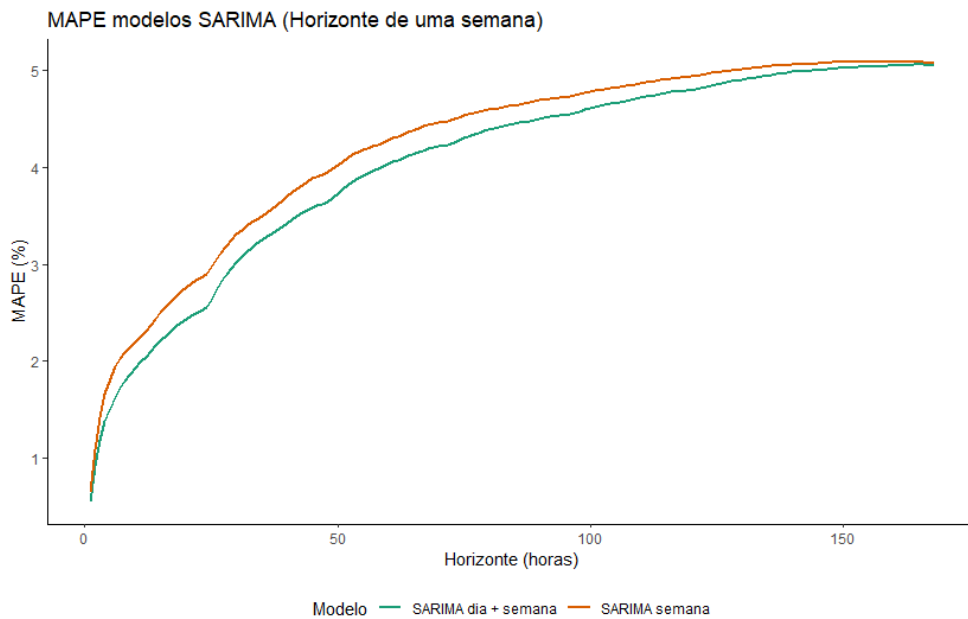


Figura 12: MAPE referente aos modelos SARIMA, com horizonte de uma semana de previsões

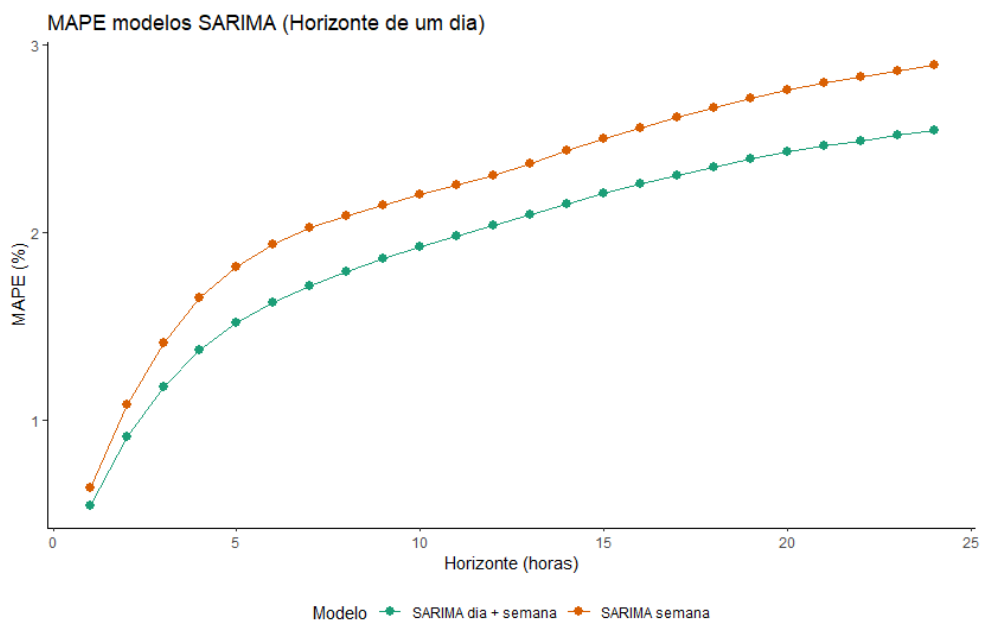


Figura 13: MAPE referente aos modelos SARIMA, com horizonte de um dia de previsões

Em relação aos ajustes por TBATS é apresentado, nas Figuras 14 e 15, o melhor modelo para previsões com horizonte de uma semana ou um dia, foi o modelo de sazonalidades intra-diária e intra-semanal. Dentre estes, o modelo de sazonalidades intra-semanal e anual apresenta os melhores resultados para os horizontes de 3 a 14 horas. No entanto, há uma piora significativa para horizontes mais longos. Já o modelo de sazonalidade tripla, juntamente com o modelo intra-diário intra-semanal apresentam MAPE de aproximadamente 1,33% para a primeira hora de previsão.

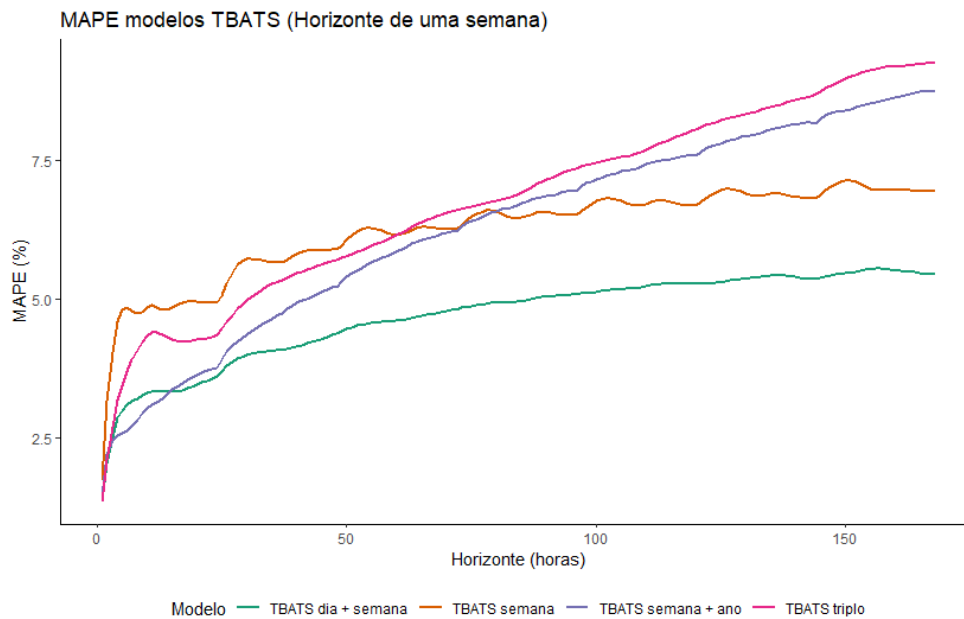


Figura 14: MAPE referente aos modelos TBATS, com horizonte de uma semana de previsões

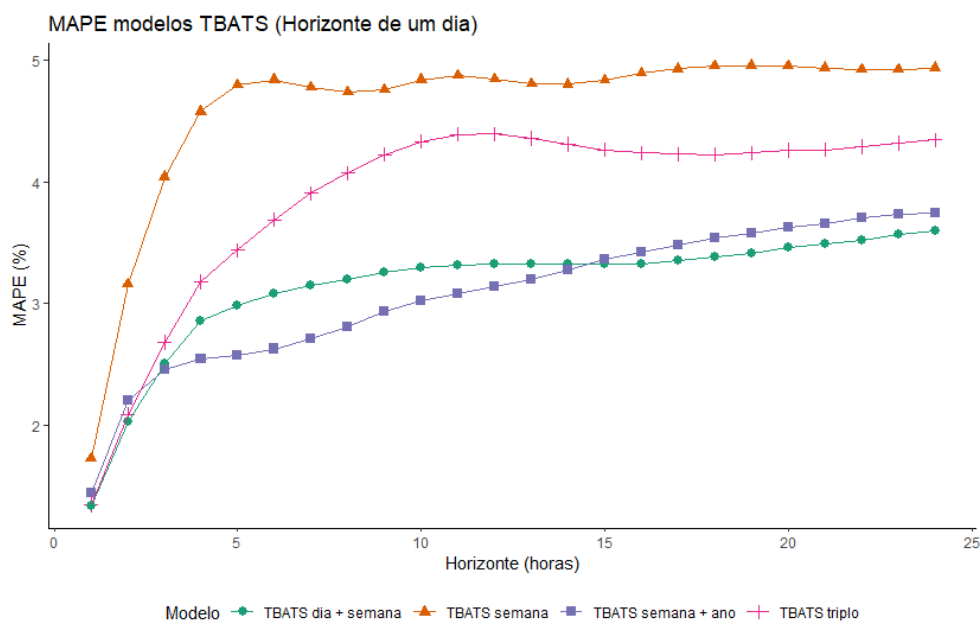


Figura 15: MAPE referente aos modelos TBATS, com horizonte de um dia de previsões

Em vista das Figuras 16 e 17, os modelos de Alisamento Exponencial de HWT aparenta ser menos preciso para a previsão dos momentos que seguem cada período sazonal, em vista dos saltos após cada período de 24 horas. De fato, controlando para a sazonalidade intra-diária este efeito é reduzido. Contudo, o modelo de Alisamento Exponencial de HWT com duas sazonalidades, apresentou previsões de MAPE 2,64% para o horizonte de 24 horas.

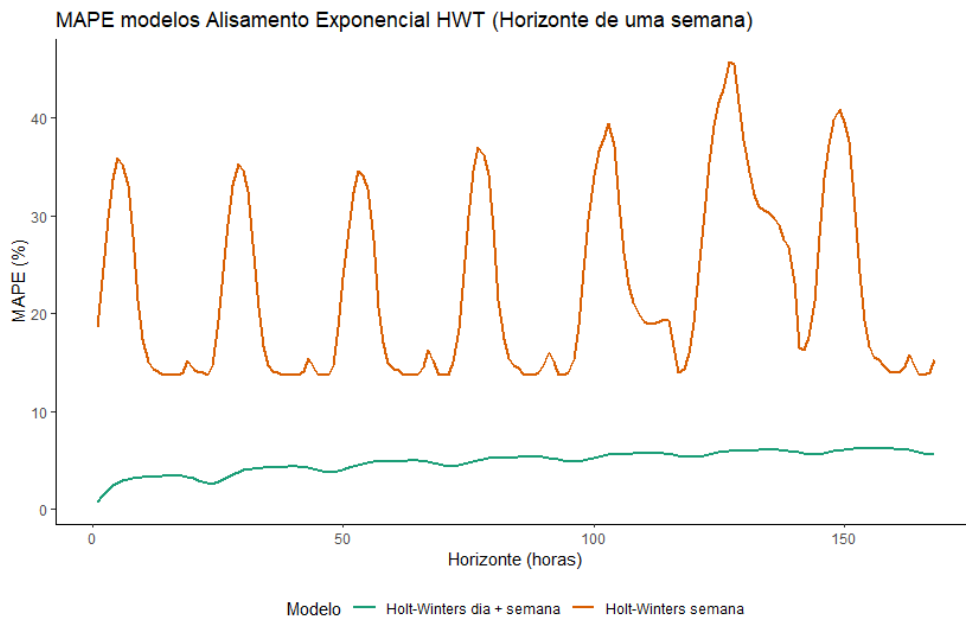


Figura 16: MAPE referente aos modelos de Alisamento Exponencial HWT, com horizonte de uma semana de previsões

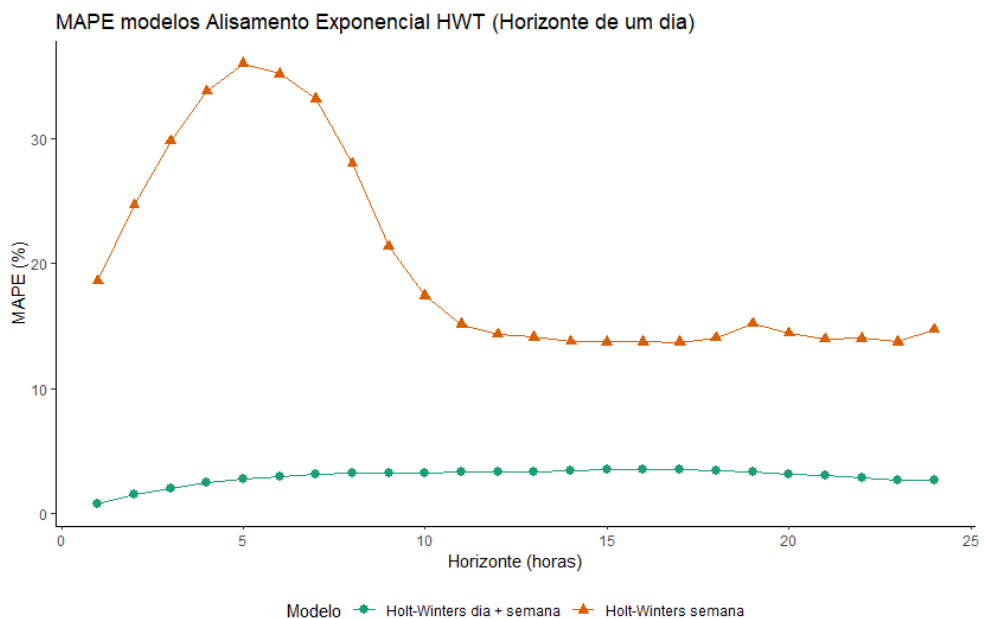


Figura 17: MAPE referente aos modelos de Alisamento Exponencial HWT, com horizonte de um dia de previsões

A Figura 18 compara os valores MAPE médios para as primeiras 24 horas de previsão dentre todos os modelos. Observa-se que os modelos por termos de Fourier e o modelo de alisamento exponencial HWT com sazonalidade semanal, obtiveram MAPE médio superior a 5% nas primeiras horas. Comparando os demais modelos para as primeiras 6 horas de previsão na Figura 19, tem-se que ambos modelos SARIMA apresentam os melhores resultados médios em previsões de curto prazo.

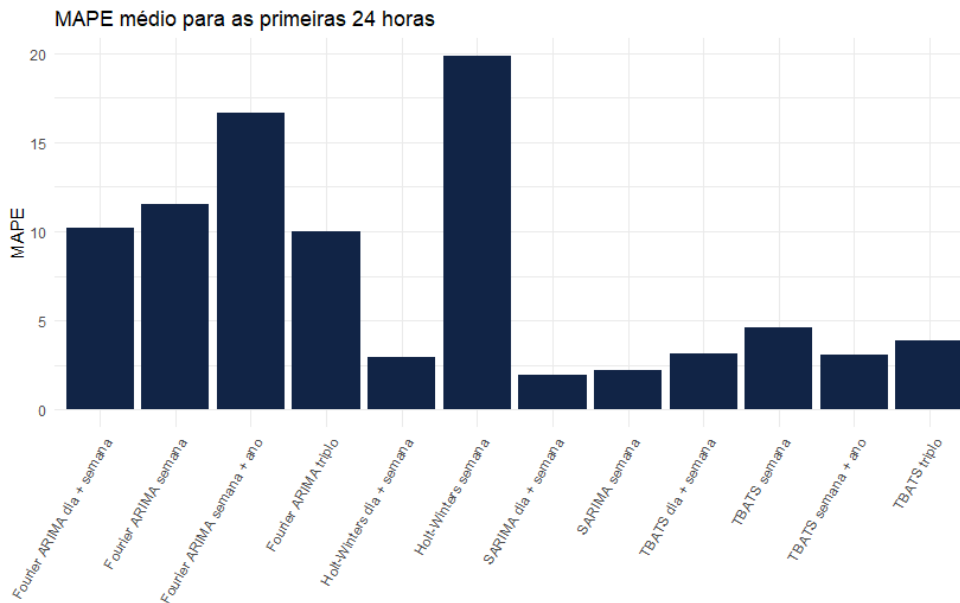


Figura 18: MAPE médio para as primeiras 24 horas

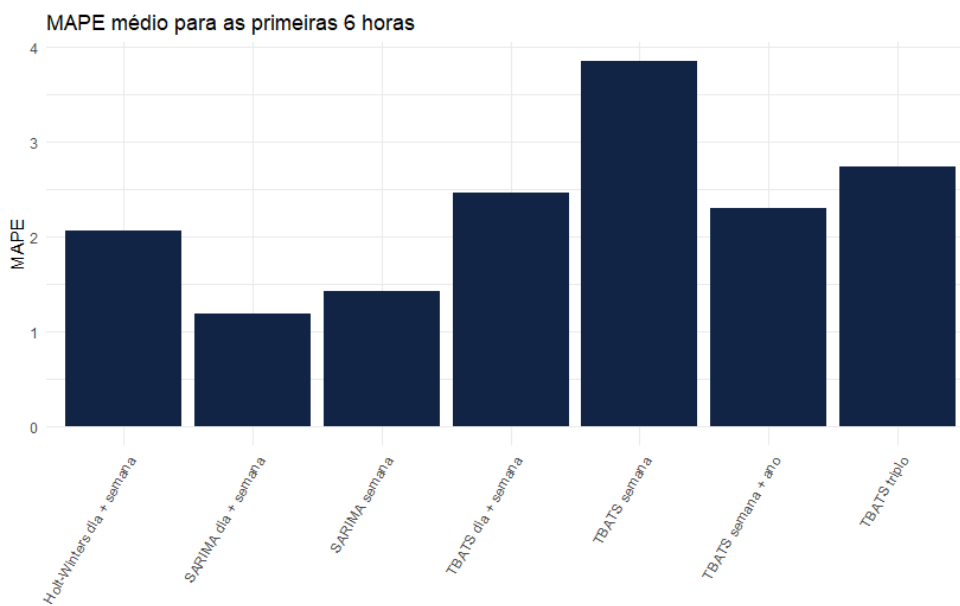


Figura 19: MAPE médio para as primeiras 6 horas

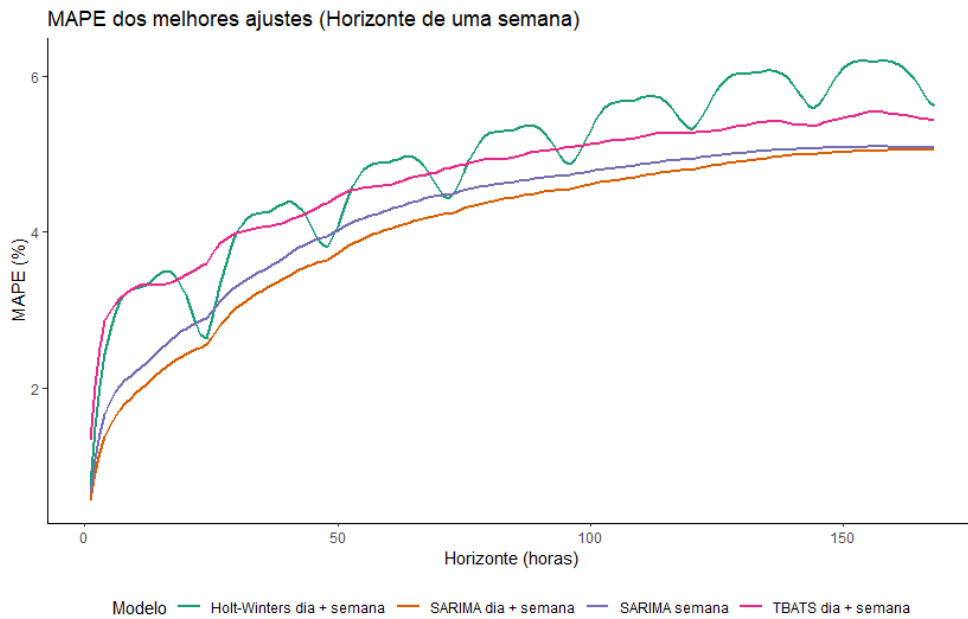


Figura 20: MAPE dos melhores ajustes para uma semana de previsões

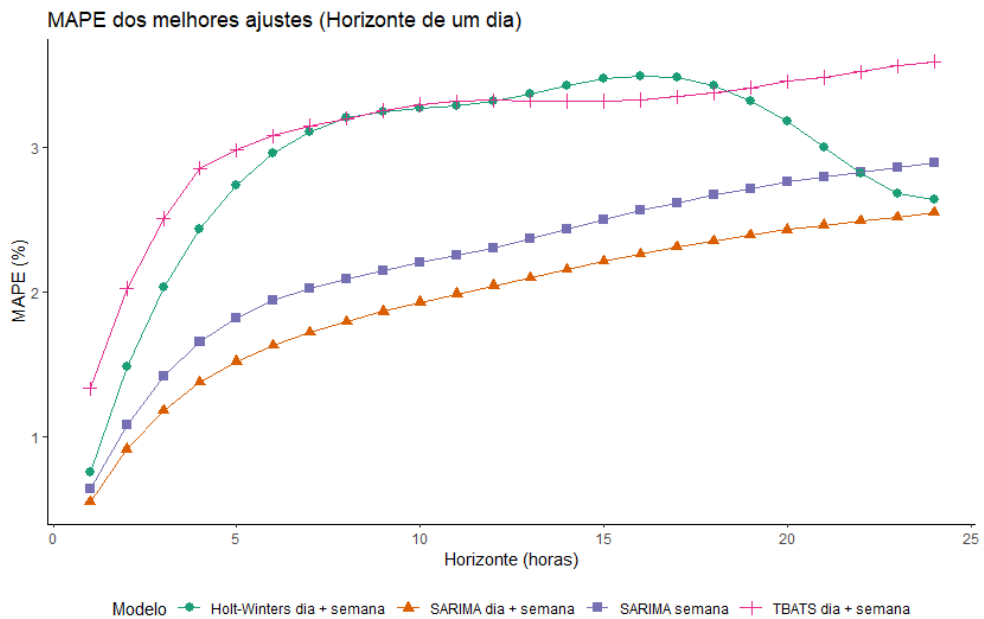


Figura 21: MAPE dos melhores ajustes para um dia de previsões

Até então, observamos que a adição de componentes sazonais apresentam resultados benéficos nos modelos apresentados. Dentre modelos com os melhores resultados observados são comparados nas Figuras 20 e 21. Verifica-se que o modelo SARIMA de sazonalidade diária e semanal, $SARIMA(5,0,3)[1](1,0,1)[24](0,1,1)[168]$, apresenta MAPE menor para todos os horizontes de previsão.

Além disso, ao estudar o efeito da sazonalidade anual na demanda de eletricidade da França e Grã Bretanha, Taylor (2010), realiza validação cruzada nos dados de um ano

a frente dos dados usados para modelagem. Obtendo MAPE em torno de 0,5% para a primeira hora de previsão e MAPE entre 1,5% e 2% para as previsões com horizonte de 24 horas dos melhores modelos no estudo. Semelhante ao experimento de Taylor (2010), os modelos com mais sazonalidades obtiveram de fato os melhores resultados.

No estudo de Arora e Taylor (2013), que trata de dias anômalos, nos quais foi utilizada a mesma metodologia de teste que Taylor (2010), os resultados são semelhantes para dias normais, com MAPE próximo de 0,5% para a primeira hora de previsão e próximo de 2% para o horizonte de 24 horas nos melhores modelos obtidos.

4 Conclusão

Neste estudo foram utilizados dados da série de demanda de eletricidade no Sudeste e Centro-Oeste, disponibilizada publicamente pela ONS. Por se tratar de uma série formada por observações horárias, foram identificados múltiplos comportamentos sazonais referentes a períodos diários, semanais e anuais. Tais padrões de comportamento foram então incorporados aos modelos estatísticos disponíveis na literatura, esses foram mensurados quanto a capacidade preditiva utilizando o procedimento conhecido como janela deslizante.

A revisão de literatura identificou que a inclusão da sazonalidade anual é preferível nos modelos ajustados, e variáveis explicativas podem auxiliar nesse processo. No entanto ao comparar empiricamente os modelos univariados: ARIMA com termos de Fourier, SARIMA, Alisamento exponencial HWT e TBATS para diferentes arranjos de sazonalidade, identificou-se que o modelo SARIMA de sazonalidades intra-diária e intra-semanal apresenta os melhores resultados para previsões de curto prazo (até uma semana), é esperado que o modelo SARIMA de sazonalidade tripla apresente ganhos, porém esse se mostrou inviável computacionalmente.

Analogamente ao estudo apresentado em Taylor (2010), os modelos obtidos apresentaram ganhos ao considerar mais períodos sazonais. No entanto, ao observar o ajuste do modelo TBATS aparenta ser possível obter um melhor ajuste para a componente sazonal anual que o alçado pelo método automático utilizado nesse estudo, tal aspecto deve ser investigado em estudos futuros. Além do mais, sugere-se que estudos futuros procurem incorporar variáveis explicativas, principalmente variáveis que estejam relacionadas a eventos comemorativos, feriados e pontos facultativos que não possuem calendário fixo.

Referências

ARORA, S.; TAYLOR, J. W. Short-term forecasting of anomalous load using rule-based triple seasonal methods. *IEEE Transactions on Power Systems*, v. 28, n. 3, p. 3235–3242, 2013.

BERGMEIR, C.; BENÍTEZ, J. M. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, v. 191, p. 192–213, 2012. ISSN 0020-0255. Data Mining for Software Trustworthiness. Disponível em: [⟨https://www.sciencedirect.com/science/article/pii/S0020025511006773⟩](https://www.sciencedirect.com/science/article/pii/S0020025511006773).

BERGMEIR, C.; HYNDMAN, R. J.; KOO, B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, v. 120, p. 70–83, 2018. ISSN 0167-9473. Disponível em: [⟨https://www.sciencedirect.com/science/article/pii/S0167947317302384⟩](https://www.sciencedirect.com/science/article/pii/S0167947317302384).

BROZYNA, J. et al. MULTI-SEASONALITY IN THE TBATS MODEL USING DEMAND FOR ELECTRIC ENERGY AS A CASE STUDY. *ECONOMIC COMPUTATION AND ECONOMIC CYBERNETICS STUDIES AND RESEARCH*, 52, n. 1, p. 229–246, 2018. ISSN 0424-267X.

CANCELO, J. R.; ESPASA, A.; GRAFE, R. Forecasting the electricity load from one day to one week ahead for the Spanish system operator. *INTERNATIONAL JOURNAL OF FORECASTING*, 24, n. 4, SI, p. 588–602, OCT-DEC 2008. ISSN 0169-2070.

DARBELLAY, G.; SLAMA, M. Forecasting the short-term demand for electricity - Do neural networks stand a better chance? *INTERNATIONAL JOURNAL OF FORECASTING*, 16, n. 1, p. 71–83, JAN-MAR 2000. ISSN 0169-2070.

GOULD, P. G. et al. Forecasting time series with multiple seasonal patterns. *European Journal of Operational Research*, v. 191, n. 1, p. 207–222, 2008. ISSN 0377-2217. Disponível em: [⟨https://www.sciencedirect.com/science/article/pii/S0377221707008740⟩](https://www.sciencedirect.com/science/article/pii/S0377221707008740).

HAGAN, M.; BEHR, S. THE TIME-SERIES APPROACH TO SHORT-TERM LOAD FORECASTING. *IEEE TRANSACTIONS ON POWER SYSTEMS*, 2, n. 3, p. 785–791, AUG 1987. ISSN 0885-8950.

HIPPERT, H.; BUNN, D.; SOUZA, R. Large neural networks for electricity load forecasting: Are they overfitted? *International Journal of Forecasting*, v. 21, n. 3, p. 425–434, 2005. ISSN 0169-2070. Disponível em: [⟨https://www.sciencedirect.com/science/article/pii/S016920700400130X⟩](https://www.sciencedirect.com/science/article/pii/S016920700400130X).

HOR, C.; WATSON, S.; MAJITHIA, S. Analyzing the impact of weather variables on monthly electricity demand. *IEEE TRANSACTIONS ON POWER SYSTEMS*, 20, n. 4, p. 2078–2085, NOV 2005. ISSN 0885-8950.

HYDE, O.; HODNETT, P. An adaptable automated procedure for short-term electricity load forecasting. *IEEE TRANSACTIONS ON POWER SYSTEMS*, 12, n. 1, p. 84–93, FEB 1997. ISSN 0885-8950. 1996 IEEE / PES Winter Meeting, BALTIMORE, MD, JAN 21-25, 1996.

- HYNDMAN, R. J. Forecasting with long seasonal periods. 2010. Disponível em: <https://robjhyndman.com/hyndsight/longseasonality/>.
- HYNDMAN, R. J.; KHANDAKAR, Y. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, v. 26, n. 3, p. 1–22, 2008. Disponível em: <https://www.jstatsoft.org/article/view/v027i03>.
- LISI, F.; SHAH, I. Forecasting next-day electricity demand and prices based on functional models. *Energy Systems*, Springer, v. 11, n. 4, p. 947–979, 2020.
- LIVERA, A. M. D.; HYNDMAN, R. J.; SNYDER, R. D. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American statistical association*, Taylor & Francis, v. 106, n. 496, p. 1513–1527, 2011.
- MBAMALU, G.; ELHAWARY, M. LOAD FORECASTING VIA SUBOPTIMAL SEASONAL AUTOREGRESSIVE MODELS AND ITERATIVELY REWEIGHTED LEAST-SQUARES ESTIMATION. *IEEE TRANSACTIONS ON POWER SYSTEMS*, 8, n. 1, p. 343–348, FEB 1993. ISSN 0885-8950.
- OPERADOR NACIONAL DO SISTEMA ELÉTRICO – ONS. *CURVA DE CARGA HORÁRIA*. 2021. Disponível em: www.ons.org.br. Acesso em: 26 fevereiro 2021.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2020. Disponível em: <https://www.R-project.org/>.
- SMITH, M. Modeling and short-term forecasting of New South Wales electricity system load. *JOURNAL OF BUSINESS & ECONOMIC STATISTICS*, 18, n. 4, p. 465–478, OCT 2000. ISSN 0735-0015.
- SRINIVASAN, D.; CHANG, C.; LIEW, A. DEMAND FORECASTING USING FUZZY NEURAL COMPUTATION, WITH SPECIAL EMPHASIS ON WEEKEND AND PUBLIC HOLIDAY FORECASTING. *IEEE TRANSACTIONS ON POWER SYSTEMS*, 10, n. 4, p. 1897–1903, NOV 1995. ISSN 0885-8950. 1995 IEEE/PES Winter Meeting, NEW YORK, NY, JAN 29-FEB 02, 1995.
- TAYLOR, J. W. Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, Taylor & Francis, v. 54, n. 8, p. 799–805, 2003.
- TAYLOR, J. W. An evaluation of methods for very short-term load forecasting using minute-by-minute British data. *INTERNATIONAL JOURNAL OF FORECASTING*, 24, n. 4, SI, p. 645–658, OCT-DEC 2008. ISSN 0169-2070.
- TAYLOR, J. W. Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, v. 204, n. 1, p. 139–152, 2010.
- TAYLOR, J. W.; MCSHARRY, P. E. Short-term load forecasting methods: An evaluation based on European data. *IEEE TRANSACTIONS ON POWER SYSTEMS*, 22, n. 4, p. 2213–2219, NOV 2007. ISSN 0885-8950.
- TOFALLIS, C. A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, Springer, v. 66, n. 8, p. 1352–1362, 2015.

VU, D. H.; MUTTAQI, K. M.; AGALGAONKAR, A. P. A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. *APPLIED ENERGY*, 140, p. 385–394, FEB 15 2015. ISSN 0306-2619.