

Desempenho Operacional Ferroviário

Abordagem Frequentista e Bayesiana em
Perspectiva

Karina Barros de Oliveira
Laís Tebas Santana



Sumário



3 Introdução

16 Considerações Finais

5 Objetivos

17 Referências

6 Revisão (PI IV)

8 Procedimentos Metodológicos

11 Resultados



Relevância do setor ferroviário

- ▶ Papel vital no transporte de grandes cargas.
- ▶ Oferece custos mais baixos, segurança e sustentabilidade em comparação com o transporte rodoviário.
- ▶ A eficiência operacional é crucial para atender à crescente demanda da indústria e economia.
- ▶ SIADE (Sistema de Acompanhamento do Desempenho Operacional das Concessionárias) - Ferramenta para monitorar, coletar e analisar dados da operação ferroviária.



Abordagem Bayesiana

Caracterizada pela incorporação de informações prévias (prior), juntamente com a evidência obtida dos dados (likelihood) para formar uma probabilidade posterior (posterior).

Diferenças entre Estatística Bayesiana e Frequentista

Bayesiana	Frequentista
Usa informações prévias e as atualiza com novos dados.	Não usa informações prévias; baseia-se exclusivamente nos dados observados.
Grau “Pessoal” de incerteza	Repetição
Intervalos de credibilidade, que são intervalos onde o parâmetro está com certa probabilidade.	Intervalos de confiança, que têm uma interpretação baseada em longas sequências de amostras.

Objetivos



Objetivo Geral
Comparar a eficácia das abordagem frequentista e bayesiana na modelagem preditiva do desempenho operacional ferroviário.



Objetivos Específicos

- Revisar e atualizar a abordagem frequentista;
- Implementar a abordagem bayesiana;



Objetivos Específicos

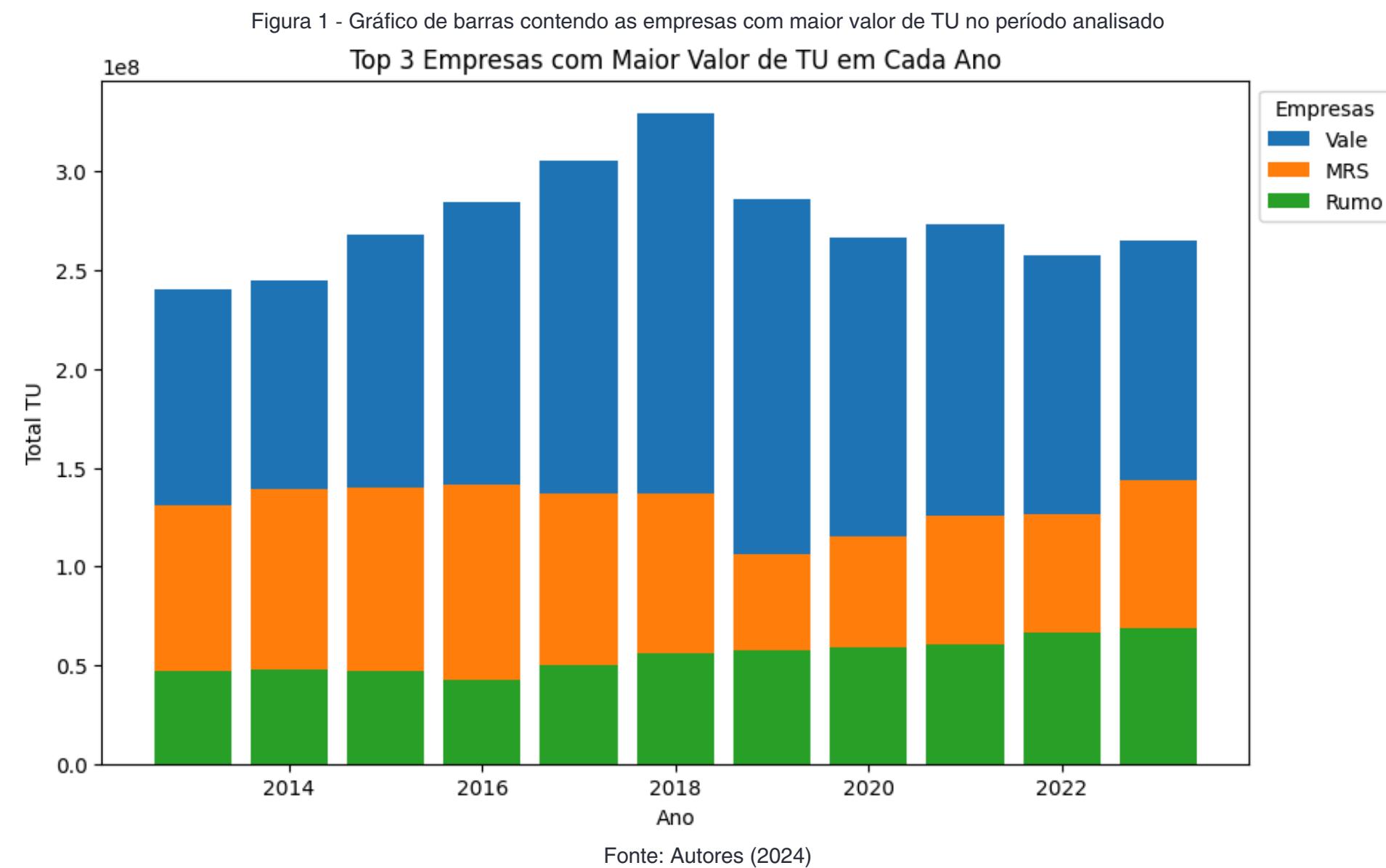
- Comparar os resultados preditivos;
- Avaliar benefícios e limitações;
- Propor recomendações.

Abordagem Frequentista

Revisão de análise exploratória utilizada na abordagem frequentista.

Quadro 1 - Resumo estatístico			
	TU	Ano	Mes
count	83.600,00	83.600,00	83.600,00
mean	66.400,00	2.020,00	6,55
std	466.000,00	3,19	3,44
min	1,00	2.010,00	1,00
25%	1.140,00	2.020,00	4,00
50%	5.270,00	2.020,00	7,00
75%	21.600,00	2.020,00	10,00
max	13.900.000,00	2.020,00	12,00

Fonte: Autores (2024)



Abordagem Frequentista

Previsão de movimentação de cargas em TU para as empresas MRS, Vale, Rumo, VLI, EFPO, FTC, FTL, com auto-ARIMA.

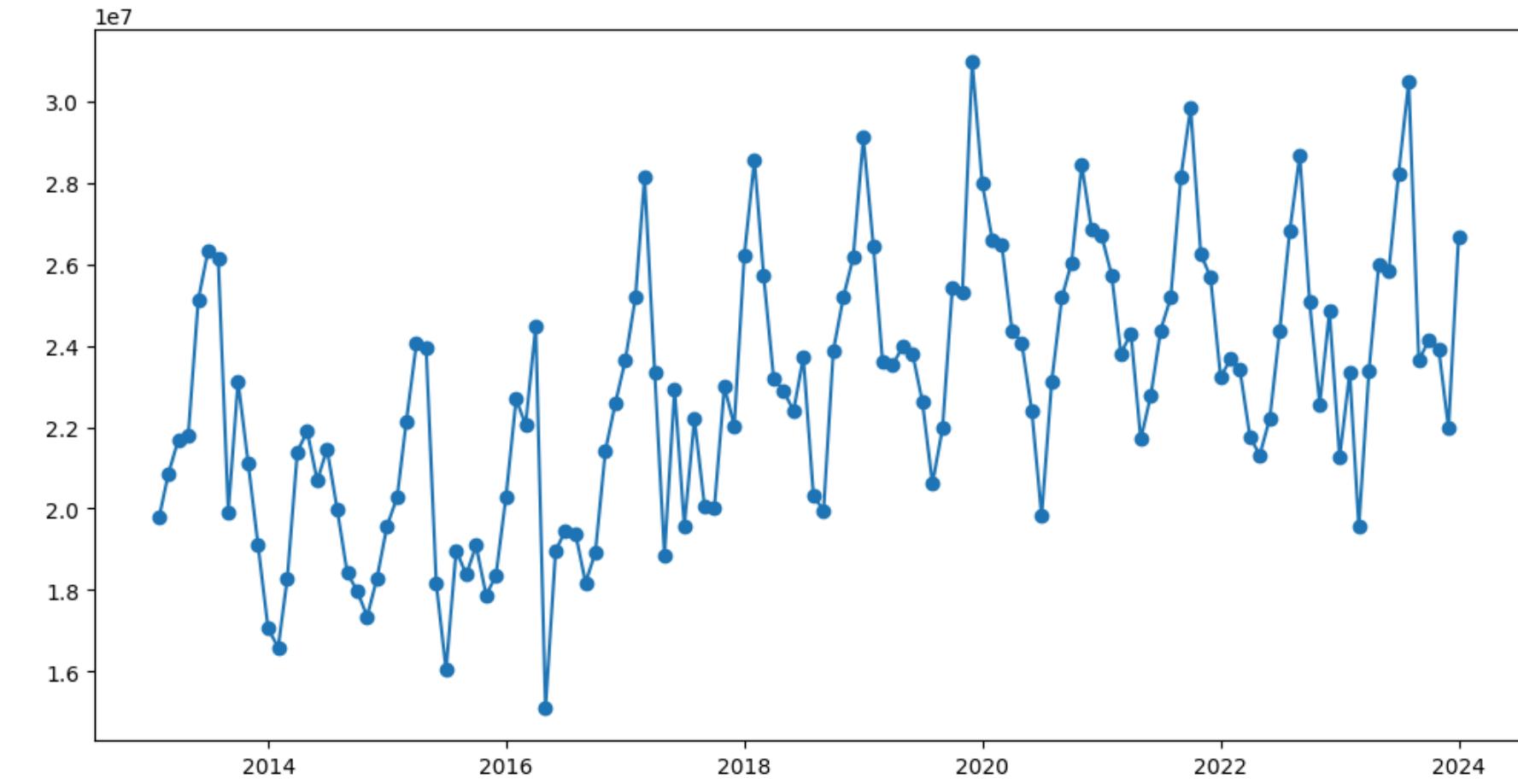
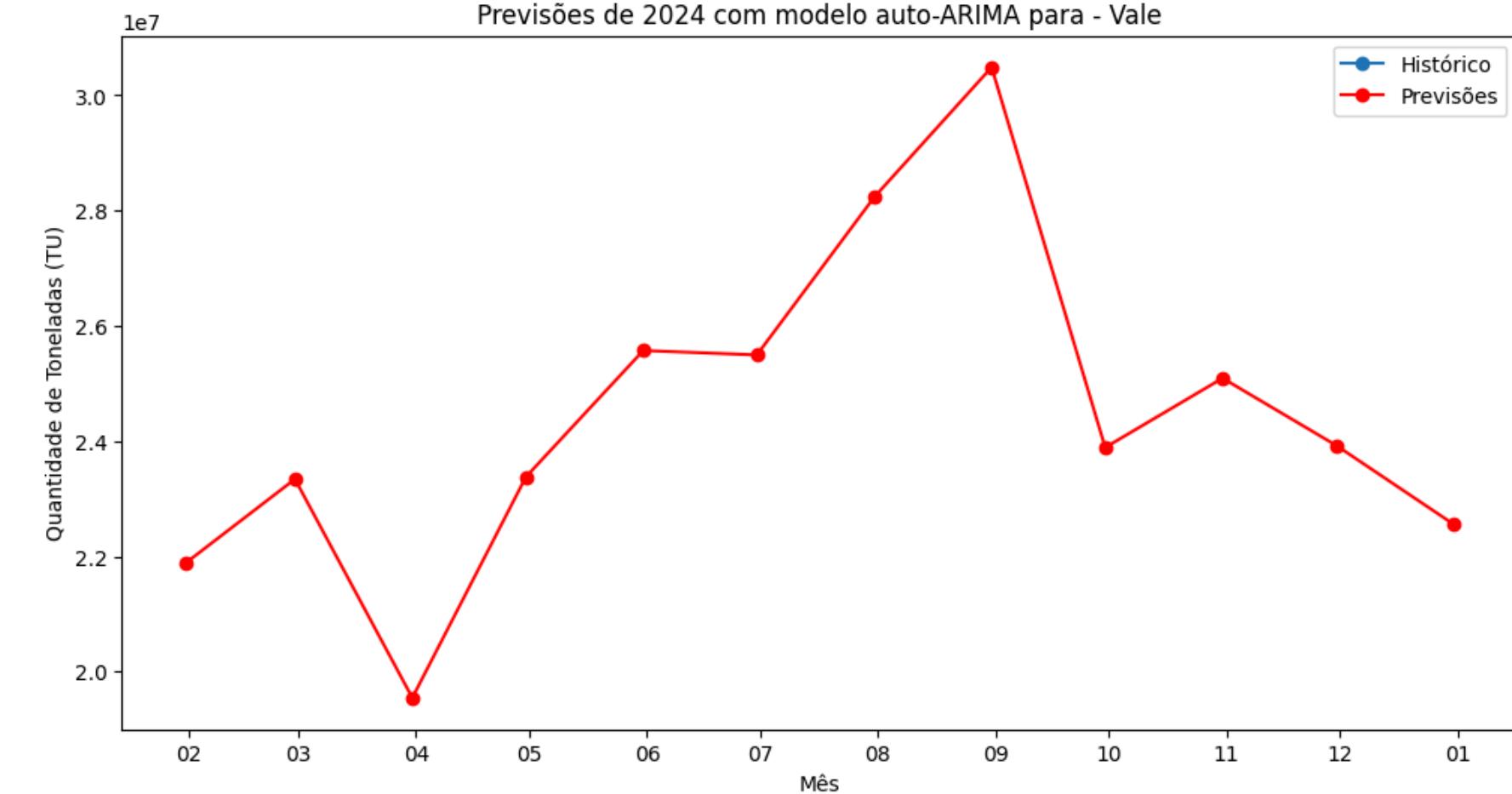
Exemplo de resultado do auto-ARIMA:

Best model: ARIMA(3,1,2)(0,0,0)[0]

- Componente Autorregressivo (AR): Ordem 3
- Componente de Diferenciação (I): Ordem 1
- Componente de Média Móvel (MA): Ordem 2
- Componente de Sazonalidade (ar, ma, sa): (0, 0, 0)

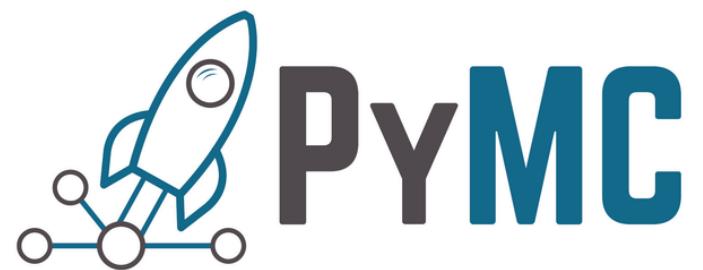
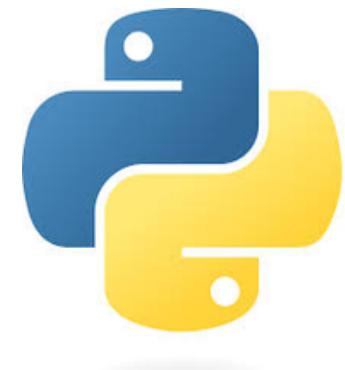
Figura 2 - Previsão de TU para a empresa Vale com auto-Arima

Previsões de 2024 com modelo auto-ARIMA para - Vale



Implementação Bayesiana

Ferramentas utilizadas



Procedimentos
Metodológicos

Análise preliminar

Tratamento e normalização de dados.

Quadro 2 - Dataset inicial

Mes_Ano	Mercadoria_ANTT	Estacao_Origem	UF_Origem	Estacao_Destino	UF_Destino	TU	Ano	Mes	Empresas
jan/13	Ferro Gusa	Açailândia	MA	Ponta da Madeira Pêra do Píer	MA	109295.0	2013	1	Vale
jan/13	Ferro Gusa	Marabá	PA	Ponta da Madeira Pêra do Píer	MA	53155.0	2013	1	Vale
jan/13	Ferro Gusa	Rosário	MA	Ponta da Madeira Pêra do Píer	MA	10185.0	2013	1	Vale
jan/13	Ferro Gusa	Santa Inês	MA	Ponta da Madeira Pêra do Píer	MA	9443.0	2013	1	Vale
jan/13	Gasolina	Itaqui Base Combustível	MA	Açailândia	MA	10410.0	2013	1	Vale
...
dez/23	Soja	Maringa	PR	D Pedro II	PR	32208.0	2023	12	Rumo
dez/23	Soja	Rolandia	PR	D Pedro II	PR	16889.0	2023	12	Rumo
dez/23	Soja	Tupancireta	RS	Rio Grande	RS	3161.0	2023	12	Rumo
dez/23	Ureia	D Pedro II	PR	Maringa	PR	3089.0	2023	12	Rumo
dez/23	Ureia	km 5	PR	Maringa	PR	4868.0	2023	12	Rumo

Fonte: Autores (2024)

Quadro 3 - Dataset normalizado

days	TU_normalized
0.000000	0.121070
0.007777	0.000000
0.014802	0.187020
0.022579	0.171180
0.030105	0.279772
...	...
0.969393	0.756958
0.977170	0.823028
0.984696	0.757704
0.992474	0.671020
1.000.000	0.819880

Fonte: Autores (2024)

Análise preliminar

Divisão dos dados.

Figura 2 - Algoritmo para divisão dos dados

```
# Dividir os dados em treinamento e teste
prop_train = 0.8
n_train = round(prop_train * len(df_data))

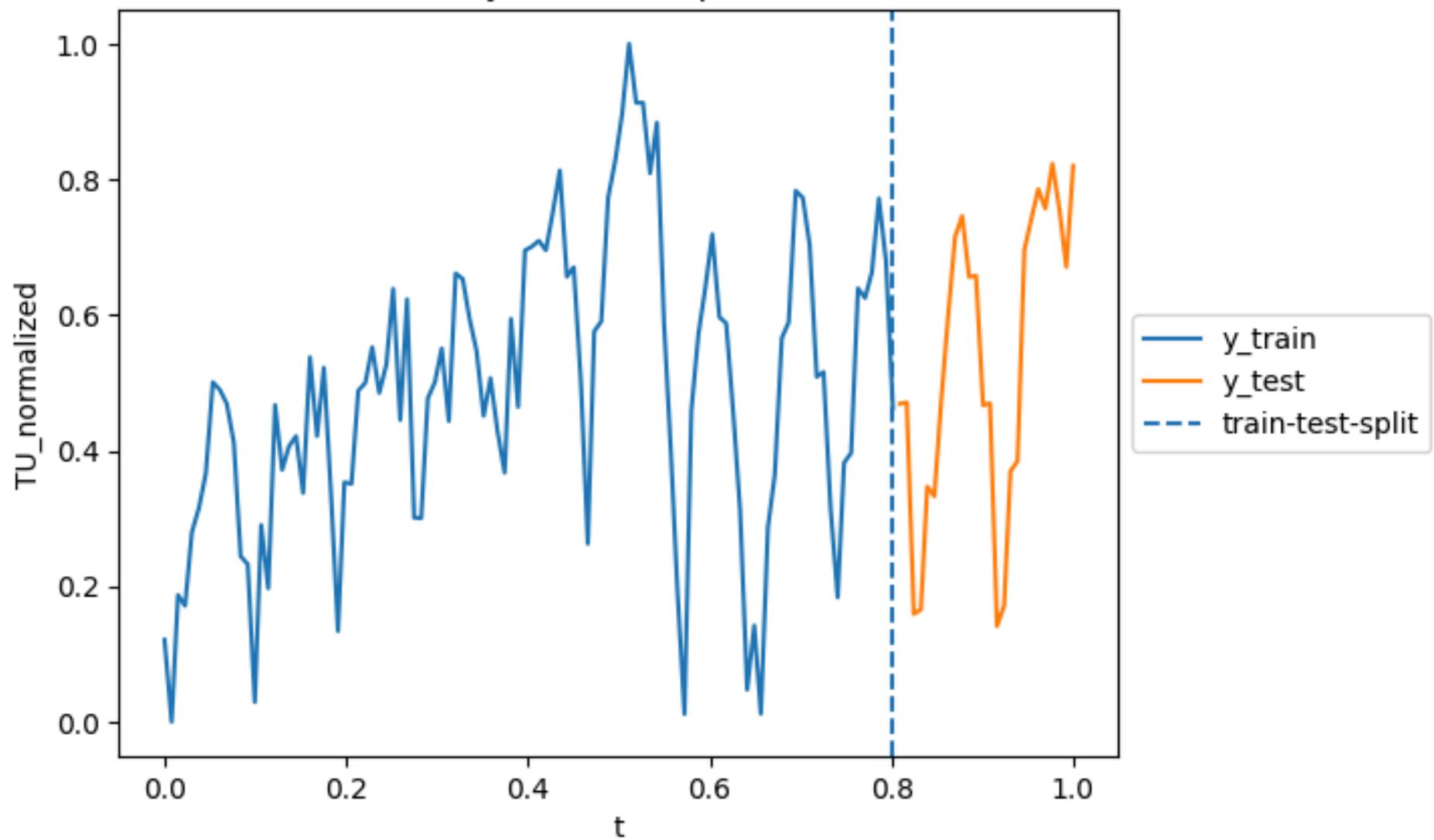
x_train = df_data['days'][:n_train].values
y_train = df_data['TU_normalized'][:n_train].values

x_test = df_data['days'][n_train:].values
y_test = df_data['TU_normalized'][n_train:].values

# Plot
fig, ax = plt.subplots()
sns.lineplot(x=x_train.flatten(), y=y_train.flatten(), label='y_train', ax=ax)
sns.lineplot(x=x_test.flatten(), y=y_test.flatten(), label='y_test', ax=ax)
ax.axvline(x=x_train.flatten()[-1], linestyle='--', label='train-test-split')
ax.legend(loc='center left', bbox_to_anchor=(1, 0.5))
ax.set(title='y train-test split', xlabel='t', ylabel='TU_normalized')
plt.show()
```

Fonte: Autores (2024)

Figura 3 - Gráfico para os dados de treinamento e de teste
y train-test split



Fonte: Autores (2024)

Definição do modelo

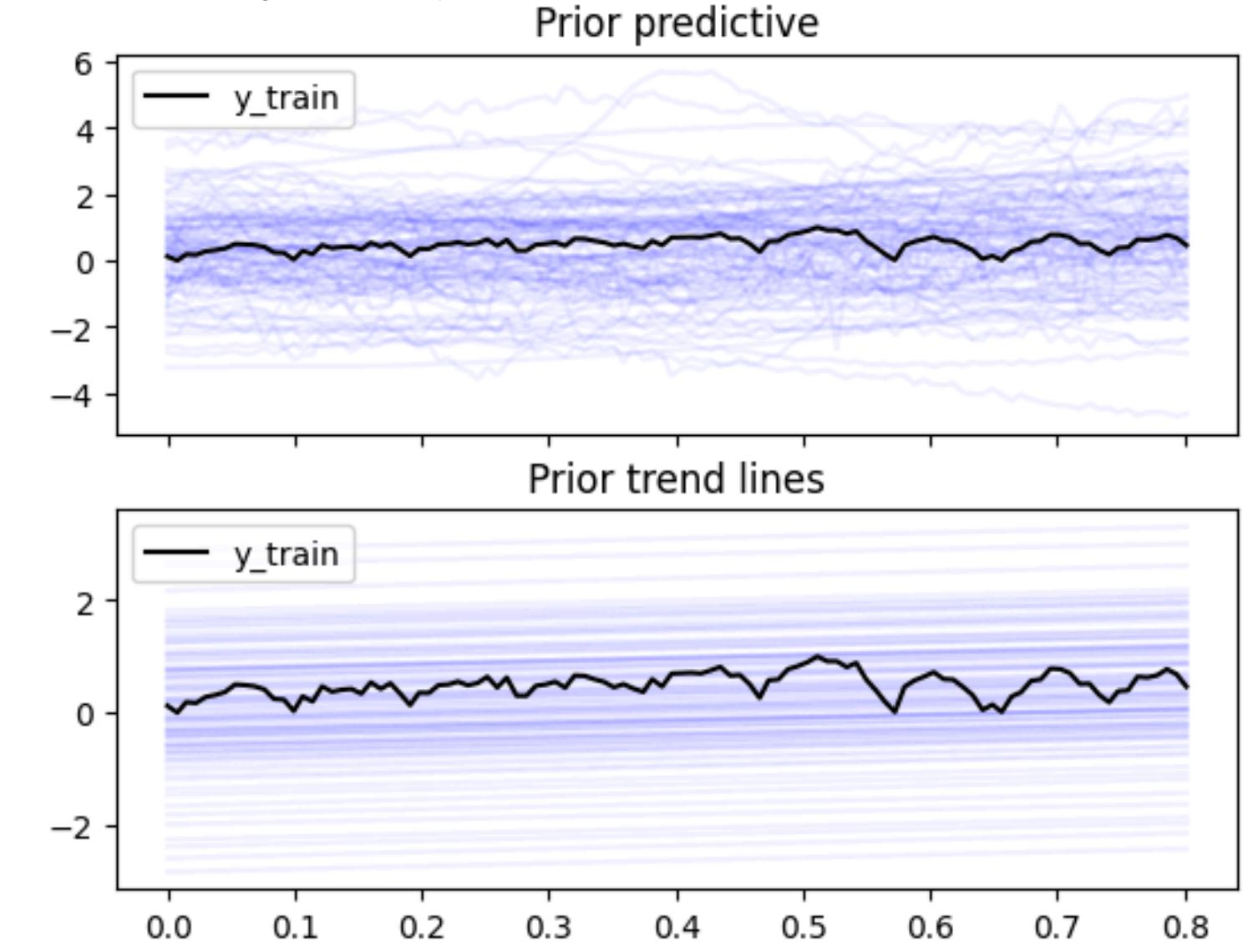
Análise Preditiva a Priori.

Figura 4 - Algoritmo para definição do modelo

```
with pm.Model() as model:  
  
    l_ratquad = pm.HalfNormal('l_ratquad', sigma=1)  
    alpha_ratquad = pm.HalfNormal('alpha_ratquad', sigma=1)  
    sigma_ratquad = pm.HalfNormal('sigma_ratquad', sigma=1)  
  
    l_matern32 = pm.HalfNormal('l_matern32', sigma=1)  
    sigma_matern32 = pm.HalfNormal('sigma_matern32', sigma=1)  
  
    cov_ratquad = sigma_ratquad * pm.gp.cov.RatQuad(1, l_ratquad, alpha_ratquad)  
    cov_matern32 = sigma_matern32 * pm.gp.cov.Matern32(1, l_matern32)  
  
    cov = cov_ratquad + cov_matern32  
  
    a = pm.Normal("a",0,1)  
    b = pm.Normal("b",0.5,0.05)  
    trend = pm.Deterministic("trend",a + b*x_train)  
    M = pm.gp.mean.Linear([b],a)  
  
    gp = pm.gp.Marginal(mean_func=M, cov_func=cov)  
    sigma = pm.HalfNormal('sigma', sigma=0.1)  
    #Antes sigma=1; Atribuiu-se sigma mínimo - incerteza adicionada aos dados observados  
    y_obs = gp.marginal_likelihood('y', x=x_train.reshape(-1,1), y=y_train, noise=sigma)  
  
with model:  
    prior = pm.sample_prior_predictive()
```

Fonte: Autores (2024)

Figura 5 - Gráfico para amostra da Priori Preditiva e Linhas de Tendência da Priori



Fonte: Autores (2024)

Resultados

Definição do modelo

Análise Preditiva a Priori.

```
Figura 6 - Algoritmo para distribuição a priori do modelo  
with model:  
    trace_approx = pm.fit() # Fit da distribuição aproximada usando inferência variational  
    trace = trace_approx.sample()  
  
# Plot do progresso do ELBO (Evidence Lower Bound)  
plt.plot(trace_approx.hist)  
plt.title('ELBO Progress')  
plt.xlabel('Iterations')  
plt.ylabel('ELBO value')  
plt.show()
```

Fonte: Autores (2024)

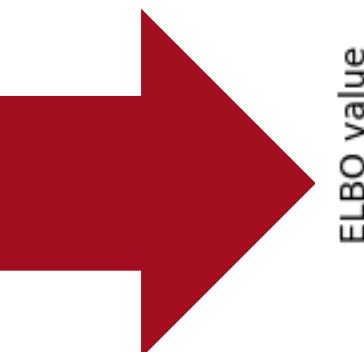
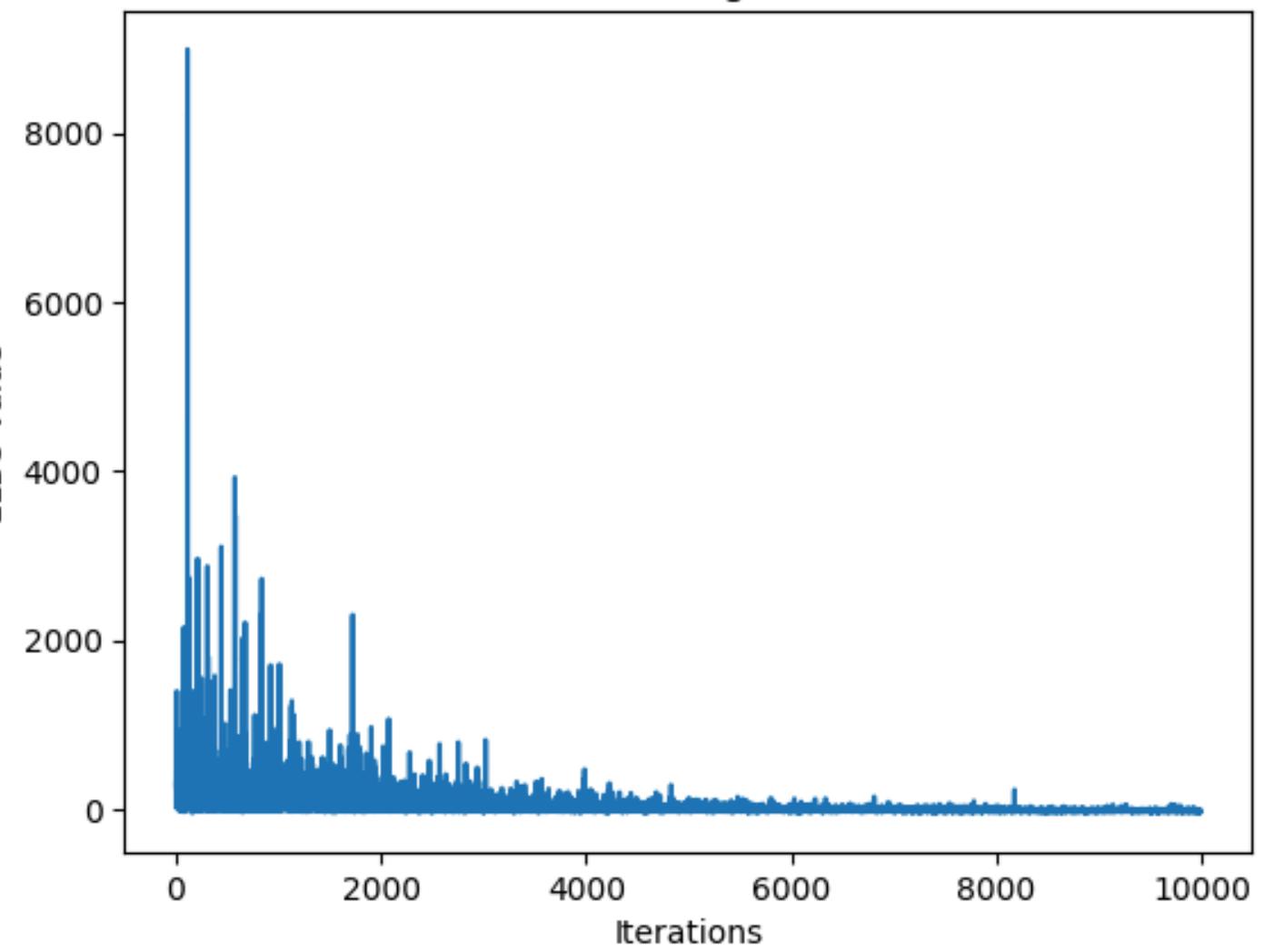


Figura 7 - Gráfico para ELBO
ELBO Progress



Fonte: Autores (2024)

Resultados

Validação do modelo

Análise Preditiva a Posteriori.

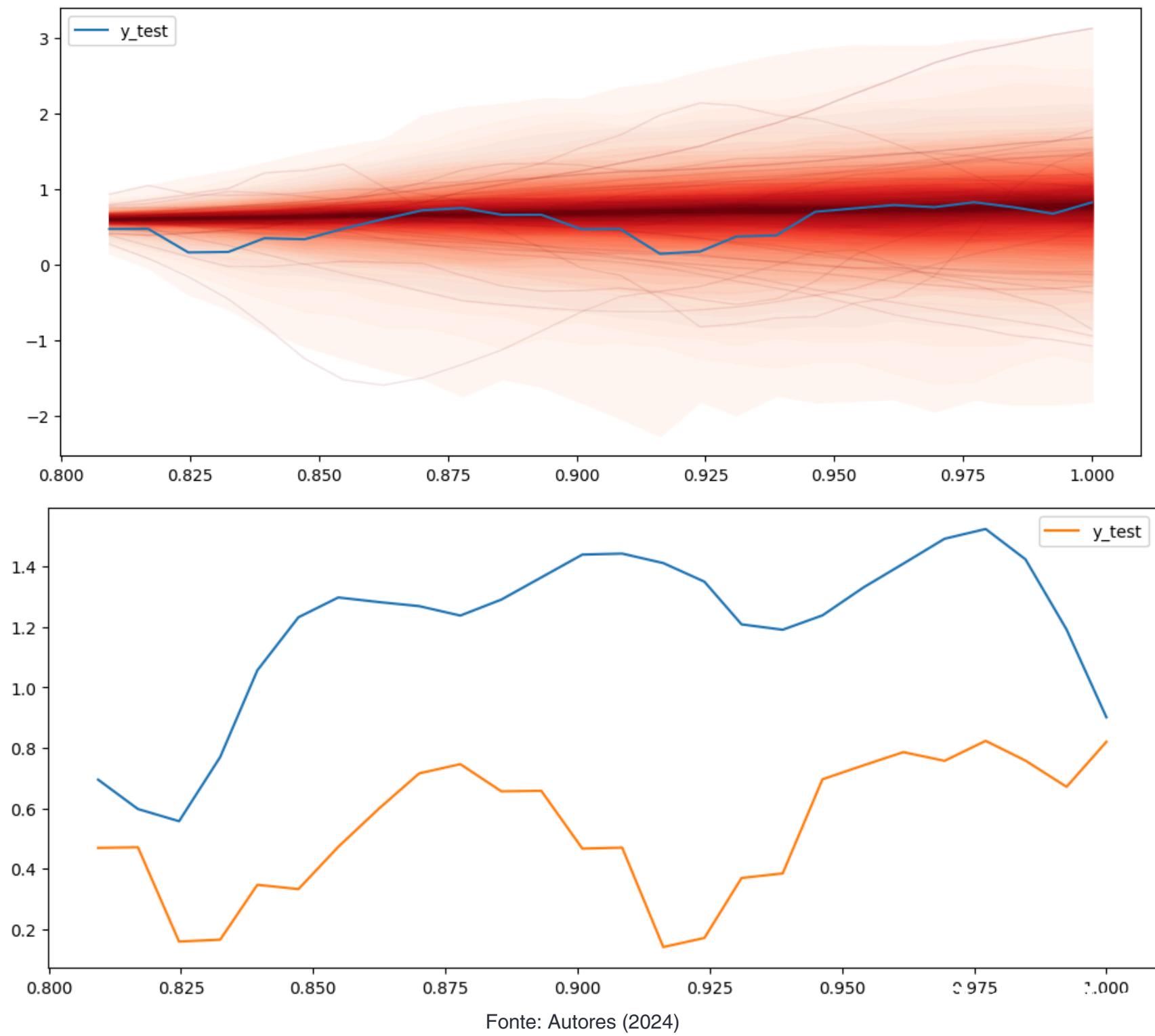
Figura 8 - Algoritmo para distribuição posteriori do modelo

```
with model:  
    x_test_c = gp.conditional('x_test', x_test.reshape(-1,1))  
    pred_m = pm.sample_posterior_predictive(trace, var_names=["x_test"])  
pred_samples = az.extract(pred_m, group="posterior_predictive", var_names=["x_test"])  
pred_samples  
  
from pymc(gp.util import plot_gp_dist  
fig = plt.figure(figsize=(12,5)); ax = fig.gca()  
plot_gp_dist(ax, pred_samples.T, x_test);  
sns.lineplot(x=x_test.flatten(), y=y_test.flatten(), label='y_test', ax=ax)  
y_pred = pred_m.posterior_predictive["x_test"].values.mean(axis=1)[0]  
print(f"MAE: {abs(y_pred - y_test).sum()/len(y_test)}")  
  
MAE: 0.1690958007929488
```

Fonte: Autores (2024)

Resultados

Figura 9 - Gráficos para Amostragem da predição posteriori



Fonte: Autores (2024)

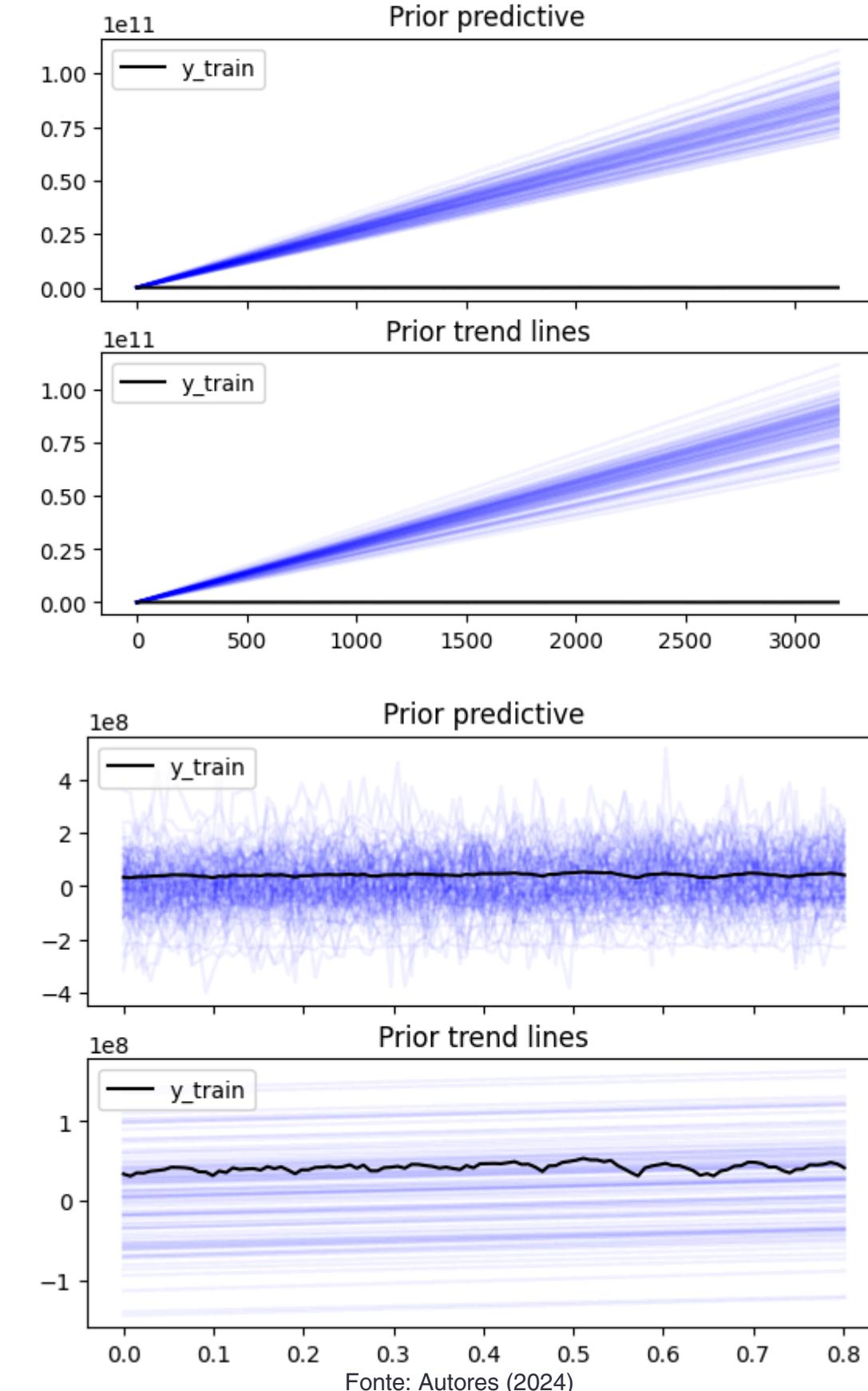
Comparações

Quadro 4 - Quadro comparativo de Modificações e Mean Absolute Error (Erro médio absoluto)

Modificações	MAE
Sigma 1	25.103.754.517.493.100
Data Normalizada; Sigma = 1	5.872.042.117.858.100
Dados Normalizados; Sigma = 1	0.18565568338640817
Dados Normalizados; Sigma = 0.1	0.1758655998420332
Dados Normalizados; Sigma = 0.5	0.18798419117226547

Fonte: Autores (2024)

Figura 10 - Gráfico para amostra da Priori Preditiva das Modificações 1 e 2



Fonte: Autores (2024)

Modelo Prophet

Com ferramenta de modelagem de séries temporais automatizada.

Figura 11 - Gráficos com decomposição da série

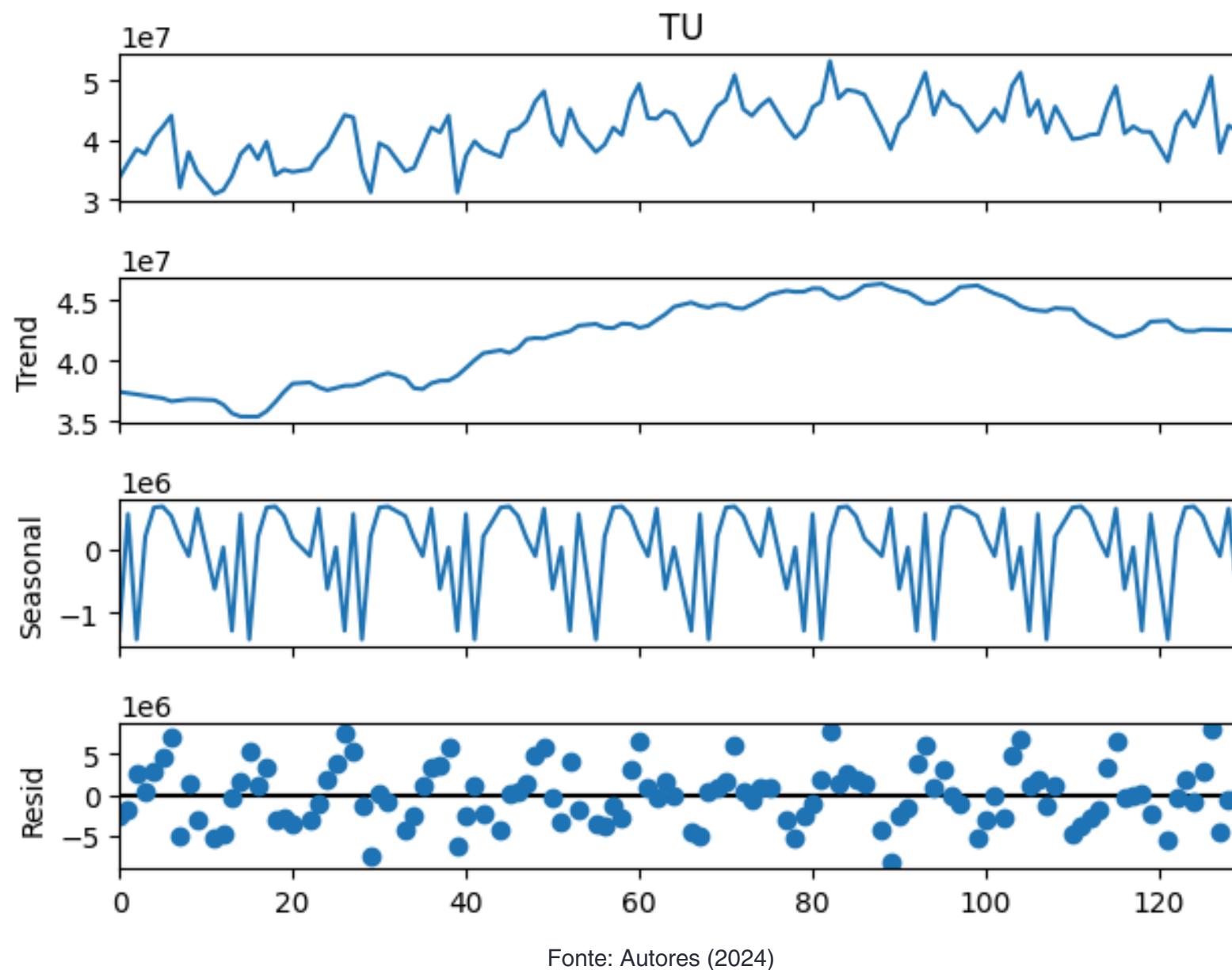
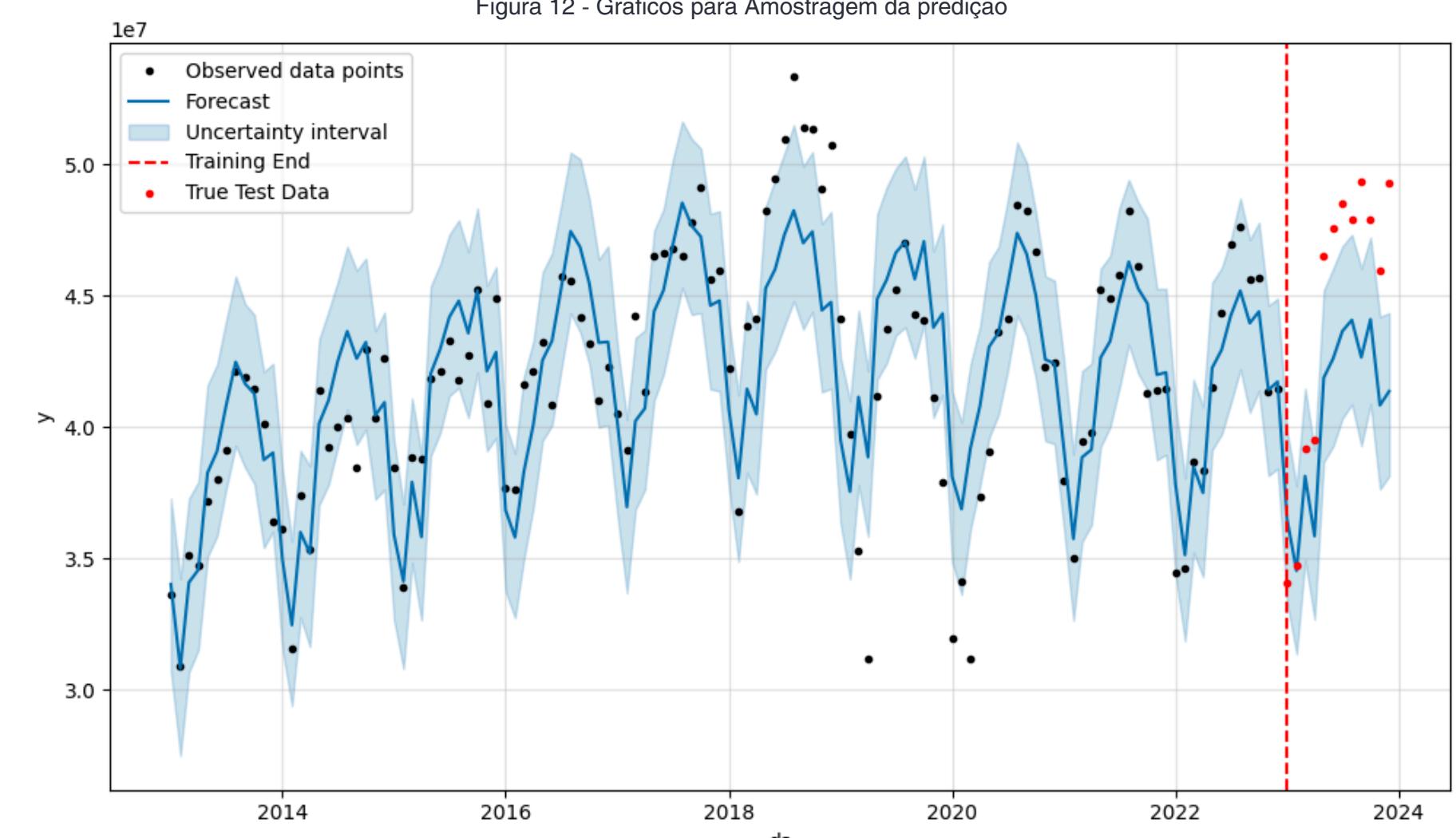


Figura 12 - Gráficos para Amostragem da predição



Fonte: Autores (2024)

Total previsto para o ano de 2023:
486.227.218,81 (TU)
Valor real = 530.596.848 TU

Considerações Finais

- Para profícua abordagem Bayesiana necessita-se de mais configuração e conhecimento de inferência estatística, incluindo definição de prioris e amostragem.
- Processos Gaussianos para a previsão do desempenho operacional permite capturar a incerteza nas previsões e modelar relações não lineares nos dados.
- Propõe-se maior exploração do uso de Kernels Gaussianos, além da expansão do conjunto de dados e a inclusão de variáveis adicionais.



Referências

- BRASIL. Agência Nacional de Transportes Terrestres. **Portal de Dados Abertos**: Sistema de Acompanhamento do Desempenho Operacional das Concessionárias - SIADE. Brasília, 31 ago. 2023a. Disponível em: <https://dados.antt.gov.br/dataset/sistema-de-acompanhamento-do-desempenho-operacional-das-concessionarias-siade>. Acesso em: 03 maio 2024.
- SGROTT, P. R. **Uma análise do cenário do transporte ferroviário de cargas**. 2019. Monografia (Graduação em Engenharia Ferroviária e Metroviária) – Centro Tecnológico de Joinville da Universidade Federal de Santa Catarina, Joinville, 2019. Disponível em: <https://repositorio.ufsc.br/handle/123456789/197312>. Acesso em: 03 maio 2024.
- DIECKMANN, J. Getting Started Predicting Time Series Data with Facebook Prophet. **Medium**, 2024. Disponível em: <https://towardsdatascience.com/getting-started-predicting-time-series-data-with-facebook-prophet-c74ad3040525>. Acesso em: 4 jun. 2024

