

UNIVERSIDADE ABERTA



**Temperatura do Oceano Atlântico em Mudança:
Modelação Espaço-Temporal em Regiões de Portugal e Cabo Verde**

Arciolindo de Conceição Pinheiro

**Mestrado em Estatística, Matemática e Computação
Ramo Estatística Computacional**

2026

UNIVERSIDADE ABERTA



**Temperatura do Oceano Atlântico em Mudança:
Modelação Espaço-Temporal em Regiões de Portugal e Cabo Verde**

Arciolindo de Conceição Pinheiro

**Mestrado em Estatística, Matemática e Computação
Ramo Estatística Computacional**

**Dissertação de mestrado orientada pela
Prof^a. Doutora Elisabete Carolino
e coorientada pela
Prof^a. Doutora Maria do Rosário Ramos**

Abril, 2026

Declaração relativa à utilização por terceiros

Eu, Arciolindo de Conceição Pinheiro, autor do trabalho académico intitulado “Temperatura do Oceano Atlântico em Mudança: Modelação Espaço-Temporal em Regiões de Portugal e Cabo Verde”, apresentado à Universidade Aberta para obtenção do grau de Mestre em Estatística, Matemática e Computação, Ramo Estatística Computacional, orientado pela Prof.^a Doutora Elisabete Carolino e coorientado pela Prof.^a Doutora Maria do Rosário Ramos,

Declaro que:

1. Autorizo a disponibilização deste trabalho no Repositório Aberto da Universidade Aberta (integrante da rede RCAAP), em regime de acesso aberto, para fins de preservação, divulgação científica e consulta pública.
2. Autorizo a utilização deste trabalho por terceiros, nos termos da licença *Creative Commons*.



<https://creativecommons.org/licenses/by-nc-nd/4.0/>



DECLARAÇÃO DE INTEGRIDADE

STATEMENT OF INTEGRITY

Declaro ter atuado com integridade na elaboração da presente dissertação/tese. Confirmo que em todo o trabalho conducente à sua elaboração não recorri à prática de plágio ou a qualquer outra forma de falsificação de resultados.

Mais declaro que tomei conhecimento integral do Regulamento Disciplinar da Universidade Aberta, publicado no Diário da República, 2.ª série, n.º 215, de 6 de novembro de 2013.

I hereby declare having conducted my thesis with integrity. I confirm that I have not used plagiarism or any form of falsification of results in the process of the thesis elaboration.

I further declare that I have fully acknowledged Disciplinary Regulations of the Universidade Aberta (regulation published in the official journal Diário da República, 2.ª série, N.º 215, de 6 de novembro de 2013).

Universidade Aberta, 06 de abril de 2026

Nome completo/Full name: Arciolindo de Conceição Pinheiro

Assinatura/Signature:

manuscrita ou digital / handwritten or digital

Agradecimentos

Em primeiro lugar, agradeço aos meus pais, António João Pinheiro e Albertina de Conceição Neves, pelo exemplo de dedicação, pelo incentivo incondicional aos estudos e pelo apoio constante ao longo de toda a minha formação académica.

À minha esposa Maria José Soares Mendes pela paciência, compreensão e carinho em todos os momentos, especialmente nos mais exigentes deste percurso. O seu apoio foi essencial para a conclusão deste trabalho.

Ao meu amigo José Félix Moniz Delgado, pelo incentivo contínuo, motivação constante e pelas palavras de encorajamento nos momentos mais desafiantes. A sua amizade e apoio tiveram um papel importante ao longo deste percurso académico.

À minha orientadora, Professora Doutora Elisabete Carolino, e à minha coorientadora, Professora Doutora Maria do Rosário Ramos, pelo rigor científico, disponibilidade, orientação segura e contributos fundamentais em todas as fases deste trabalho. Agradeço profundamente a ambas pela partilha de conhecimento, pelos conselhos sempre pertinentes e pelo apoio constante, que foram decisivos para a concretização desta dissertação.

Um agradecimento à Dra. Dora Carinhas, do Instituto Hidrográfico (Portugal), pela sugestão de analisar o fenómeno a diferentes profundidades e pela colaboração e disponibilidade na obtenção dos dados oceanográficos.

Por fim, à Universidade Aberta, pela oportunidade de realizar este mestrado e pelo enquadramento académico que tornou possível a concretização desta dissertação.

Resumo

Esta dissertação analisa a evolução da temperatura média mensal do mar no Atlântico entre janeiro de 1968 e junho 2023 em quatro regiões: Nazaré, Faro e São Miguel (Portugal) e Praia (Cabo Verde), nas profundidades de 1 m e 5 m, utilizando dados *European Marine Observation and Data Network (EMODnet)*. Observou-se um padrão sazonal anual bem definido em todas as localidades, bem como um aquecimento estatisticamente significativo, com aumentos entre 0,0196°C e 0,0234°C por ano, equivalentes a 0,20-0,23°C por década. Praia apresenta temperaturas consistentemente mais elevadas e maior estabilidade térmica.

Foram aplicadas várias metodologias de séries temporais, incluindo decomposição clássica, *Seasonal-Trend Decomposition*, Holt-Winters e modelos SARIMA. Os modelos SARIMA apresentaram melhor desempenho, captando de forma robusta a tendência e a sazonalidade, com erros de previsão reduzidos (MAE entre 0,59°C e 0,97°C; MAPE entre 2,94% e 4,97%). As previsões para julho de 2023 a junho de 2024 reproduzem adequadamente o ciclo sazonal e alinham-se com os valores observados.

Para integrar a dimensão espacial, ajustou-se um modelo STARMA às séries a um metro de profundidade, baseado numa matriz de pesos construída a partir de distâncias geodésicas. O modelo final revelou dependência temporal dominante e influência espacial moderada, apresentando bom desempenho e resíduos próximos de ruído branco. A comparação entre os modelos SARIMA e STARMA revelou desempenhos preditivos globalmente semelhantes, com ligeira vantagem do modelo STARMA na maioria das localidades analisadas (Nazaré, Faro e Praia). As previsões STARMA mantiveram a hierarquia térmica entre regiões e reforçaram a coerência dos resultados univariados.

Globalmente, o estudo confirma o aquecimento progressivo das águas do Atlântico e demonstra que modelos SARIMA e STARMA são ferramentas eficazes para compreender e prever a variabilidade térmica regional, fornecendo suporte científico relevante para ações de monitorização oceânica e adaptação às alterações climáticas.

Palavras-chave: Temperatura do Mar; Modelos SARIMA; Modelação Espaço-Temporal (STARMA); Oceano Atlântico; Alterações Climáticas; Previsão.

Abstract

This dissertation analyzes the evolution of the monthly mean sea temperature in the Atlantic Ocean between January 1968 and June 2023 in four regions: Nazaré, Faro and São Miguel (Portugal), and Praia (Cape Verde), at depths of 1 m and 5 m, using European Marine Observation and Data Network (EMODnet) data. A well-defined annual seasonal pattern was observed in all locations, as well as a statistically significant warming trend, with increases ranging from 0.0196°C to 0.0234°C per year, equivalent to 0.20-0.23°C per decade. Praia exhibits consistently higher temperatures and greater thermal stability.

Several time series methodologies were applied, including classical decomposition, Seasonal-Trend Decomposition, Holt-Winters and SARIMA models. SARIMA models showed the best performance, robustly capturing both the trend and seasonality, with low forecasting errors (MAE between 0.59°C and 0.97°C; MAPE between 2.94% and 4.97%). Forecasts for the period from July 2023 to June 2024 adequately reproduced the seasonal cycle and aligned well with the observed values.

To incorporate the spatial dimension, a STARMA model was fitted to the time series at one-meter depth, based on a spatial weight matrix constructed from geodesic distances. The final model revealed dominant temporal dependence and moderate spatial influence, showing good performance and residuals close to white noise. A comparison between the SARIMA and STARMA models revealed broadly similar predictive performance, with a slight advantage of the STARMA model in most of the locations analysed (Nazaré, Faro, and Praia). STARMA forecasts preserved the thermal hierarchy among regions and reinforced the structural consistency of the univariate results.

Overall, the study confirms the progressive warming of Atlantic waters and demonstrates that SARIMA and STARMA models are effective tools for understanding and predicting regional thermal variability, providing relevant scientific support for ocean monitoring and climate adaptation strategies.

Keywords: Sea Temperature; SARIMA Models; Spatio-Temporal Modelling (STARMA); Atlantic Ocean; Climate Change; Forecasting.

Índice

1- Capítulo I- Introdução	1
1.1- Enquadramento e objetivos da dissertação	1
2- Capítulo II – Séries Temporais.....	5
2.1- Conceito de série temporal	5
2.2- Objetivos da análise de séries temporais.....	5
2.3- Componentes de uma série temporal	6
2.4- Modelos de decomposição.....	6
2.5- Fundamentos estocásticos de séries temporais	7
2.5.1- Processo estocástico	7
2.5.2- Estacionaridade (estrita e fraca)	8
2.5.3- Funções de dependência temporal.....	8
2.5.4- Exemplos clássicos de processos estocásticos	10
2.5.5- Modelos para Séries Estacionárias.....	11
2.5.6- Modelos para Séries Não Estacionárias.....	18
2.6- Métodos determinísticos (alisamento exponencial).....	21
2.6.1- Alisamento exponencial simples	21
2.6.2- Alisamento exponencial de Holt	22
2.6.3- Alisamento exponencial de Holt-Winters	22
2.7- Diagnóstico e validação.....	24
2.7.1- Teste de raiz unitária	24
2.7.2- Análise dos resíduos.....	29
2.8- Critérios para a seleção de modelos de séries temporais.....	31
2.9- Previsão em séries temporais	32
2.10- Medidas de avaliação	33
3- Capítulo III – Análise e Modelação da Temperatura do Mar	36
3.1- Introdução.....	36
3.2- Descrição dos dados	36
3.3- Análise exploratória dos dados	37
3.4- Avaliação de estacionaridade	43
3.5- Modelação das séries por região e profundidade	46
3.5.1- Modelos de decomposição clássica	46
3.5.2- Decomposição da tendência e sazonalidade utilizando Loess (STL).....	49
3.5.3- Modelos de alisamento exponencial de Holt-Winters.....	52

3.5.4-	Modelos ARIMA/SARIMA.....	55
3.6-	Previsões utilizando os modelos SARIMA selecionados.....	63
3.7-	Avaliação da qualidade das previsões	65
4-	Capítulo IV – Fundamentos Teóricos dos Modelos Espaço–Temporais	69
4.1-	Limitações dos modelos univariados e motivação para modelos espaço– temporais.....	69
4.2-	Operador de defasagem espacial.....	69
4.3-	Definição formal do modelo STARMA	71
4.4-	Condições de estacionaridade, causalidade e invertibilidade	72
4.4.1-	Estacionaridade de segunda ordem no domínio espaço–temporal	72
4.4.2-	Condições algébricas de causalidade e invertibilidade.....	73
4.4.3-	Extensão dos modelos STARMA a processos não estacionários	74
4.5-	Identificação, estimação e diagnóstico em modelos STARMA	74
4.5.1-	Identificação de modelos espaço-temporais	74
4.5.2-	Estimação dos parâmetros dos modelos STARMA	76
4.5.3-	Diagnóstico dos modelos STARMA	81
4.5.4-	Seleção do modelo final	83
4.6-	Previsão com modelos STARMA.....	83
5-	Capítulo V – Modelação Espaço–Temporal da Temperatura do Mar com Modelos STARMA.....	85
5.1-	Introdução à aplicação espaço–temporal.....	85
5.2-	Construção da matriz de pesos espaciais	85
5.3-	Identificação da estrutura do modelo STARMA	86
5.4-	Modelos candidatos e critérios de seleção	89
5.5-	Diagnóstico, validação e seleção do modelo final	92
5.6-	Equação do modelo final selecionado.....	97
5.7-	Previsões utilizando o modelo STARMA selecionado	97
5.8-	Avaliação da qualidade das previsões	99
6-	Capítulo VI – Considerações finais	102
6.1-	Conclusão	102
6.2-	Trabalhos futuros.....	103
	Bibliografia.....	104
	ANEXOS.....	107

Anexo I- Modelos SARIMA candidatos por localização e profundidade: BIC, AICc e valor-p do teste Ljung-Box	107
Anexo II- Previsões intervalares da temperatura da água (80% de confiança), modelos SARIMA.....	109
Anexo III- Previsões intervalares da temperatura da água (95% de confiança), modelos SARIMA.....	110
Anexo IV - Tabela de valores críticos do teste Dickey-Fuller	111
Anexo V - Tabela de valores críticos do teste KPSS de estacionariedade.....	112

Índice de tabelas

Tabela 3.1: Estatísticas descritivas das temperaturas mensais registadas em Nazaré, Faro, São Miguel e Praia.	38
Tabela 3.2: Valores-p do teste de Ljung-Box aplicado às séries anuais de temperatura do mar (1968-2023).	40
Tabela 3.3: Aumento médio da temperatura do mar (1968-2023) estimado pelo método de Sen, com valores-p do teste de Mann-Kendall modificado.	41
Tabela 3.4: Estatísticas dos testes ADF aplicados às séries de temperatura.	44
Tabela 3.5: Estatísticas dos testes de PP aplicados às séries de temperatura.	44
Tabela 3.6: Estatísticas dos testes de KPSS aplicados às séries de temperatura.	45
Tabela 3.7: Valores-p do Teste de Ljung-Box para os Resíduos da Decomposição Clássica.	48
Tabela 3.8: Valores-p do Teste de Ljung-Box para os Resíduos da Decomposição STL. ..	51
Tabela 3.9: Valores-p do Teste de Ljung-Box para os Resíduos dos modelos Holt-Winters.	54
Tabela 3.10: Modelos SARIMA automáticos ajustados às séries.	55
Tabela 3.11: Valor-p do teste Ljung-Box aplicado aos resíduos dos modelos SARIMA automáticos.	56
Tabela 3.12: Resultados dos testes ADF e KPSS para as séries diferenciadas sazonalmente.	57
Tabela 3.13: Modelos SARIMA selecionados por localização e profundidade com base na aleatoriedade dos resíduos e critérios de informação.	58
Tabela 3.14: Previsões pontuais mensais de temperatura (jul/2023–jun/2024) por localização e profundidade.	63
Tabela 3.15: Valores observados e estimados da temperatura da água do mar em diferentes profundidades (Nazaré e Faro), 2023-2024.	65
Tabela 3.16: Valores observados e estimados da temperatura da água do mar em diferentes profundidades (S. Miguel e Praia), 2023-2024.	66
Tabela 3.17: Métricas de avaliação da qualidade preditiva dos modelos SARIMA.	66
Tabela 4.1: Propriedades teóricas das STACF e STPACF dos modelos STARMA.	76
Tabela 5.1: Distância (em quilómetros) entre as regiões em análise.	86
Tabela 5.2: Identificação dos modelos STARMA candidatos.	89

Tabela 5.3: BIC e variância dos modelos candidatos.	89
Tabela 5.4: Estimativas dos parâmetros do modelo M1.	90
Tabela 5.5: Estimativas dos parâmetros do modelo M2.	90
Tabela 5.6: Estimativas dos parâmetros do modelo M3.	90
Tabela 5.7: Estimativas dos parâmetros do modelo M4.	90
Tabela 5.8: Estimativas dos parâmetros do modelo M5.	90
Tabela 5.9: Estimativas dos parâmetros do modelo M6.	91
Tabela 5.10: Estimativas dos parâmetros do modelo M7.	91
Tabela 5.11: Estimativas dos parâmetros do modelo M8.	91
Tabela 5.12: Estimativas dos parâmetros do modelo M9.	91
Tabela 5.13: Previsões pontuais da temperatura mensal da superfície do mar (°C), julho 2023 – junho 2024.	98
Tabela 5.14: Valores observados e previstos da temperatura da água do mar à profundidade de 1 m, para as localizações de Nazaré, Faro, São Miguel e Praia, no período 2023-2024.	99
Tabela 5.15: Métricas de avaliação da qualidade preditiva do modelo STARMA.	100

Índice de Figuras

Figura 3.1: Representação gráfica das temperaturas mensais registadas em Nazaré, Faro, São Miguel e Praia.	37
Figura 3.2: Boxplots das Séries de Temperaturas Mensais por Região e Profundidade.	42
Figura 3.3: Representações gráficas da ACF e PACF.	43
Figura 3.4: Decomposição Clássica do tipo aditivo das séries temporais das temperaturas médias mensais em cada local e profundidade.	47
Figura 3.5: ACF e PACF da Componente Residual da Decomposição Clássica.	48
Figura 3.6: Decomposição STL para as séries temporais em cada local e profundidade.	50
Figura 3.7: ACF e PACF da Componente residual da Decomposição STL.	51
Figura 3.8: Séries de temperaturas observadas (linhas pretas) e respetivos ajustes pelo modelo de Holt-Winters (linhas coloridas) em cada local e profundidade.	53
Figura 3.9: ACF e PACF dos resíduos dos modelos Holt-Winters ajustados.	54
Figura 3.10: ACF e PACF dos resíduos dos modelos ARIMA automático ajustados.	56
Figura 3.11: Previsões das temperaturas médias mensais do mar (°C) para julho de 2023 a junho de 2024, por localidade e profundidade, incluindo previsões pontuais e intervalos de confiança a 80% e 95%.	64
Figura 3.12: Análise de concordância de Bland-Altman para as previsões Nazaré e Faro (jul. 2023 – jun. 2024).	67
Figura 3.13: Análise de concordância de Bland-Altman para as previsões São Miguel e Praia (jul. 2023 – jun. 2024).	68
Figura 5.1: funções de autocorrelação espaço-temporal (STACF).	87
Figura 5.2: funções de autocorrelação parcial espaço-temporal (STPACF).	87
Figura 5.3: Função de autocorrelação espaço-temporal dos resíduos do Modelo M1: STARMA(1,0).	92
Figura 5.4: Função de autocorrelação espaço-temporal dos resíduos do Modelo M2: STARMA(2,0).	93
Figura 5.5: Função de autocorrelação espaço-temporal dos resíduos do Modelo M3: STARMA(3,0).	93
Figura 5.6: Função de autocorrelação espaço-temporal dos resíduos do Modelo M4: STARMA(1,1).	94

Figura 5.7: Função de autocorrelação espaço-temporal dos resíduos do Modelo M5: STARMA(2 ₁ ,1).	94
Figura 5.8: Função de autocorrelação espaço-temporal dos resíduos do Modelo M6: STARMA(1 ₁ ,1 ₁).	95
Figura 5.9: Função de autocorrelação espaço-temporal dos resíduos do Modelo M7: STARMA(2 ₁ ,1 ₁).	95
Figura 5.10: Função de autocorrelação espaço-temporal dos resíduos do Modelo M8: STARMA(3 ₁ ,1 ₁).	96
Figura 5.11: Função de autocorrelação espaço-temporal dos resíduos do Modelo M9: STARMA(3 ₁ ,2 ₁).	96
Figura 5.12: Análise de concordância de Bland-Altman para as previsões modelo STARMA (jul. 2023 - jun. 2024).	101

Lista de abreviaturas, siglas e acrónimos

AIC	Critério de Informação de Akaike
AICc	Critério de Informação de Akaike Corrigido
ACF	Função de Autocorrelação
ADF	Teste de Dickey–Fuller Aumentado
AR	Modelo Autoregressivo
ARIMA	Modelo Autoregressivo Integrado de Médias Móveis
BIC	Critério de Informação Bayesiano
EMODnet	European Marine Observation and Data Network
KPSS	Teste de Kwiatkowski–Phillips–Schmidt–Shin
MA	Média Móvel
MAE	Erro Absoluto Médio
MAPE	Erro Percentual Absoluto Médio
MK	Teste de Mann-Kendall
MODIS	Moderate Resolution Imaging Spectroradiometer
MQO	Mínimos Quadrados Ordinários
NOAA	National Oceanic and Atmospheric Administration
PACF	Função de Autocorrelação Parcial
PP	Teste de Phillips-Perron
RMSE	Raiz do Erro Quadrático Médio
SARIMA	Modelo Autoregressivo Integrado de Médias Móveis Sazonal
SSE	Soma dos Quadrados dos Erros
STACF	Função de Autocorrelação Espaço-Temporal
STL	Seasonal and Trend decomposition using Loess
STPACF	Função de Autocorrelação Parcial Espaço-Temporal
STAR	Modelo Autoregressivo Espaço-Temporal
STARMA	Modelo Autoregressivo de Médias Móveis Espaço-Temporal
TSM	Temperatura da Superfície do Mar

1- Capítulo I- Introdução

1.1- Enquadramento e objetivos da dissertação

As alterações climáticas, impulsionadas pelas emissões humanas de gases com efeito de estufa, estão a transformar os oceanos em reservatórios térmicos em expansão. De acordo com o Sexto Relatório de Avaliação do IPCC (2021), mais de 90 % do excesso de energia retido no sistema climático desde meados do século XX foi absorvido pelas águas oceânicas, um facto com consequências diretas nos ecossistemas marinhos e nas comunidades que deles dependem.

Em Portugal, o Instituto Hidrográfico registou em 2023 anomalias térmicas ao largo de Faro, com temperaturas significativamente superiores à média dos últimos vinte anos (UNRIC, 2023). Longe de constituir um episódio localizado, esse aquecimento inscreve-se num padrão mais amplo: em Praia, Cabo Verde, observam-se tendências térmicas ascendentes igualmente persistentes.

A modelação estatística destas variáveis ambientais tem sido amplamente explorada na literatura científica, com particular destaque para as abordagens baseadas em séries temporais. Os primeiros estudos sobre a evolução das temperaturas do mar recorreram essencialmente a abordagens univariadas, sobretudo a modelos ARIMA e SARIMA. Estas metodologias mostraram-se eficazes na captura de padrões sazonais e tendências de longo prazo, comprovando a sua robustez estatística em investigações que abrangem tanto escalas regionais como globais, e desde a superfície até diversas profundidades da coluna de água.

Entre essas investigações, destacam-se, por exemplo, Harshada et al. (2022), que aplicaram o SARIMA à previsão da temperatura da superfície do Mar Arábico com dados recolhidos entre julho de 2002 e julho de 2020, identificando o modelo $SARIMA(1,0,1)(1,1,2)[12]$ como o mais adequado, com elevada precisão ($RMSE = 0,351\text{ }^{\circ}\text{C}$; $MAPE = 0,905\%$). Seguindo uma lógica semelhante, mas numa perspetiva global, Bai (2024) modelou a temperatura média mensal na coluna de água (0-700 m) utilizando dados da NOAA (2000-2023); o modelo $SARIMA(3,1,2)(2,0,0)[12]$ permitiu prever um ligeiro aumento de $0,01\text{ }^{\circ}\text{C}$ para 2024. Ainda em 2024, Bilgili et al. (2024) expandiram a análise temporal (1975-2023) para a temperatura da superfície do mar (TSM) global. A seleção do modelo $SARIMA(1,1,1)(1,1,1)[12]$, baseada no AIC, resultou em erros de previsão reduzidos ($MAPE = 0,1377\%$; $MAE = 0,0250\text{ }^{\circ}\text{C}$) e indicou um aumento de $1,52\%$ na temperatura

anual global entre 2023 e 2050. Mais recentemente, Khan et al. (2025) focaram-se na região norte do Oceano Índico, utilizando dados de satélite MODIS recolhidos entre agosto de 2002 e abril de 2020. A aplicação do modelo $SARIMA(1,0,0)(2,1,1)$ [12] obteve um MAPE de 0,592 % e RMSE de 0,22 °C.

Em conjunto, estes estudos evidenciam a robustez dos modelos SARIMA na modelação de séries de temperatura oceânica, justificando a sua adoção como uma das abordagens de referência nesta dissertação.

Contudo, apesar da eficácia dos modelos univariados na captura de padrões como tendência e sazonalidade, estes não são capazes de representar a dependência espaço-temporal que pode existir entre diferentes regiões. Para colmatar essa limitação, foi desenvolvida uma nova classe de modelos, denominada modelos autoregressivos de médias móveis espaço-temporais (STARMA), introduzida por Pfeifer e Deutsch (1980a). Esta abordagem, que modela simultaneamente as dimensões temporal e espacial, foi mais tarde aprimorada para aplicação em dados com componente sazonal (Pfeifer e Deutsch, 1981). Os próprios autores ilustraram o procedimento iterativo de três etapas (identificação, estimação e diagnóstico) em estudos pioneiros: inicialmente com dados sobre registos de detenções por agressão em 14 distritos de Boston (1969-1974) e, posteriormente, reconhecendo a sazonalidade em muitas séries, através de dados de caudal do rio Mississippi.

Desde então, diversos estudos têm explorado esta metodologia. Nesse âmbito, destacam-se, por exemplo, Subba Rao e Antunes (2004), que compararam a modelação espaço-temporal com abordagens univariadas, aplicando modelos STARMA a registos de temperatura em nove locais do Reino Unido. Apesar dos modelos ARMA analisados individualmente apresentarem desempenhos preditivos semelhantes, o STARMA mostrou-se superior por integrar todas as localizações num único modelo, evidenciando maior eficiência paramétrica. Num contexto distinto, Hu et al. (2011) aplicaram o modelo STARMA para detetar fissuras em estruturas de placas, explorando a correlação espacial entre vibrações em diferentes pontos, o que permitiu identificar eficazmente a localização e dimensão das fissuras sem conhecimento prévio da estrutura não danificada. Mais recentemente, Alam et al. (2024) recorreram a abordagens univariadas (ARIMA) e multivariadas (STARMA) para modelar a precipitação em 119 localizações de Bengala Ocidental (Índia), utilizando 50 anos de dados (1970-2019). Enquanto os modelos ARIMA apresentaram desempenho satisfatório, o

STARMA revelou-se significativamente superior, produzindo erros preditivos substancialmente inferiores (MAE, MAPE e RMSE).

No entanto, a aplicação dos modelos STARMA na modelação da temperatura oceânica permanece praticamente ausente na literatura, lacuna que esta dissertação procura ajudar a colmatar. Face ao exposto, torna-se essencial ir além da análise isolada de cada série. Compreender como as temperaturas em diferentes pontos do Atlântico estão estatisticamente associadas entre si é fundamental para produzir previsões robustas, capazes de orientar políticas de adaptação em regiões particularmente vulneráveis.

Objetivos do Estudo

Esta dissertação tem como **objetivo geral** desenvolver abordagens estatísticas para a modelação e previsão da temperatura média mensal do mar em quatro regiões do Atlântico: Nazaré, Faro e São Miguel (Portugal) e Praia (Cabo Verde), integrando tanto a dimensão temporal como a espacial da variabilidade térmica, de modo a gerar evidências científicas que possam informar políticas de adaptação aos impactos das alterações climáticas e apoiar a conservação e uso responsável dos recursos oceânicos.

Para concretizar este objetivo, definem-se os seguintes **objetivos específicos**:

- Caracterizar a estrutura temporal das séries de temperatura do mar registadas em Nazaré, Faro, São Miguel (Portugal) e Praia (Cabo Verde), nas profundidades de 1 e 5 metros;
- Selecionar, com base em critérios de diagnóstico (análise residual, ACF/PACF, testes estatísticos), os modelos mais adequados para cada série;
- Estender a análise à dimensão espacial mediante a aplicação de modelos STARMA, de modo a capturar as dependências térmicas entre as quatro regiões à profundidade de 1 metro;
- Comparar o desempenho preditivo dos modelos SARIMA univariados com o do modelo STARMA a 1 metro de profundidade; e
- Produzir previsões pontuais e intervalares para o horizonte julho de 2023 a junho de 2024 e avaliar a sua qualidade.

Para alcançar os objetivos propostos, esta dissertação organiza-se em seis capítulos, incluindo o presente capítulo introdutório.

O **Capítulo II** introduz os fundamentos teóricos da análise de séries temporais univariadas, abordando conceitos como estacionaridade, decomposição, funções de autocorrelação (ACF e PACF), modelos AR, MA, ARIMA e SARIMA, e critérios de seleção e validação. Seguidamente, o **Capítulo III** aplica esses fundamentos às séries de temperatura do mar nas quatro localidades e duas profundidades, conduzindo à seleção de modelos SARIMA, validados pela ausência de autocorrelação residual e pelos critérios de informação de Akaike corrigido (AICc) e Bayesiano (BIC), e à realização e avaliação da qualidade das previsões para o horizonte de julho de 2023 a junho de 2024. Toda a análise foi implementada no ambiente estatístico R.

Por sua vez, o **Capítulo IV** desenvolve a teoria dos modelos espaço-temporais (STARMA), abordando a construção de matrizes de ponderação espacial com base em distâncias geodésicas, a identificação da ordem do modelo com base nas funções de autocorrelação espaço-temporal (STACF) e autocorrelação parcial espaço-temporal (STPACF), a estimação por máxima verossimilhança, o diagnóstico residual e a teoria da previsão. Posteriormente, o **Capítulo V** aplica estes modelos às séries de temperatura do mar registradas nas quatro localidades à profundidade de 1 metro, conduzindo à seleção do modelo espaço-temporal mais adequado, com base no Critério de Informação Bayesiano (BIC), na significância estatística dos parâmetros e na adequação residual. Inclui ainda a comparação entre as métricas de erro dos modelos SARIMA e do modelo STARMA, avaliando o valor acrescentado da integração da dimensão espacial. Todas as análises foram implementadas em R.

Finalmente, o **Capítulo VI** apresenta as considerações finais da dissertação, recapitulando os principais resultados obtidos, avaliando se os objetivos traçados foram plenamente alcançados, refletindo sobre as limitações inerentes ao estudo e sugerindo caminhos concretos para investigação futura.

2- Capítulo II – Séries Temporais

2.1- Conceito de série temporal

Segundo Brockwell e Davis (2016), uma série temporal é um conjunto de observações x_t , registadas em momentos específicos t , em que t representa o tempo (ou o índice temporal) associado a cada observação. Quando as observações são realizadas em intervalos de tempo fixos (por exemplo, diariamente, mensalmente), a série é denominada discreta. Por outro lado, se as observações são feitas de forma contínua ao longo do tempo, a série é classificada como contínua.

2.2- Objetivos da análise de séries temporais

Embora a literatura identifique diversos objetivos na análise de séries temporais, entre os quais se contam os descritivos, preditivos, explicativos e de controlo (Chatfield, 2019), o presente estudo centra-se essencialmente em dois:

- (i) **Descrição:** a análise de uma série temporal inicia-se habitualmente pela representação gráfica dos valores observados ao longo do tempo, o que permite identificar padrões estruturais, como tendências de longo prazo, componentes sazonais e eventuais anomalias (por exemplo, *outliers* ou observações em falta). Esta etapa descritiva é fundamental para compreender o comportamento das temperaturas ao longo do tempo.
- (ii) **Previsão:** a análise de séries temporais permite prever valores futuros, constituindo uma base para a tomada de decisões antecipadas. No presente estudo, esta abordagem é aplicada à previsão das temperaturas médias mensais do mar nas diversas localidades e profundidades em análise, nomeadamente à superfície e a 5 metros de profundidade em Praia (Cabo Verde), Nazaré, São Miguel e Faro (Portugal).

2.3- Componentes de uma série temporal

As fontes de variabilidade de uma série temporal podem ser decompostas em quatro componentes (Morettin, 1987): tendência, sazonalidade, componente cíclica e componente residual.

Tendência: consiste na variação (crescente ou decrescente) de longo prazo no nível médio de uma série, podendo ser linear ou não linear.

Sazonalidade: reflete um padrão repetido em intervalos regulares, no presente estudo, com período anual (12 meses). Segundo Ehlers (2009), a sazonalidade pode ser **aditiva**, quando a amplitude das flutuações sazonais se mantém aproximadamente constante ao longo do tempo, ou **multiplicativa**, quando essa amplitude varia proporcionalmente ao nível geral da série.

Componente cíclica: refere-se a variações de longa duração que não possuem periodicidade fixa nem duração constante, diferenciando-se da sazonalidade precisamente pela ausência de regularidade temporal (Hyndman & Athanasopoulos, 2021).

Componente residual

A componente residual ou irregular representa o conjunto de variações aleatórias não explicadas pelas outras componentes da série temporal e deve, após um bom ajustamento do modelo, comportar-se como um ruído branco (Chatfield, 2019).

2.4- Modelos de decomposição

Após a identificação das componentes de uma série temporal, realiza-se a sua decomposição, que consiste em separá-las para analisar individualmente o seu comportamento. Esse processo permite uma melhor compreensão da estrutura da série, facilitando a análise e a previsão da série temporal.

Segundo Hyndman & Athanasopoulos (2021), na decomposição clássica de séries temporais, as componentes de tendência e ciclo não são separadas. Em vez disso, são representadas conjuntamente por uma única componente designada tendência-ciclo (T_t). Assim, a decomposição clássica considera apenas três componentes: tendência-ciclo (T_t), sazonalidade (S_t) e resíduos (R_t).

Tradicionalmente, assume-se que essas componentes podem ser combinadas de forma **aditiva** ou **multiplicativa**, conforme a natureza da série. Os modelos correspondentes são:

1. modelo aditivo:

$$X_t = T_t + S_t + R_t \quad (1)$$

2. modelo multiplicativo:

$$X_t = T_t \times S_t \times R_t \quad (2)$$

Em ambos os casos, X_t representa a série temporal observada.

A decisão entre utilizar um **modelo aditivo** ou **multiplicativo** depende da análise da componente sazonal. Quando a amplitude das variações sazonais não depende da tendência, isto é, não muda com o nível da série temporal, a decomposição aditiva é mais adequada. Por outro lado, quando a amplitude sazonal depende do nível da série temporal, isto é, aumenta ou diminui proporcionalmente ao nível da série, utiliza-se a decomposição multiplicativa (Hyndman e Athanasopoulos, 2021).

Contudo, é fundamental distinguir este método clássico (ou simples) de métodos de decomposição mais robustos, como o STL (*Seasonal and Trend decomposition using Loess*). A decomposição clássica baseia-se no cálculo de médias móveis e assume que a componente sazonal permanece constante ao longo do tempo, o que a torna sensível à presença de valores atípicos (outliers), capazes de distorcer as estimativas da tendência e da sazonalidade (Hyndman & Athanasopoulos, 2021).

Por sua vez, o método STL baseia-se na suavização local (*Loess*) e apresenta-se como uma alternativa robusta e flexível. Destaca-se, sobretudo, por permitir que a componente sazonal varie ao longo do tempo e por recorrer a um procedimento iterativo que torna a decomposição resistente à presença de outliers (Hyndman & Athanasopoulos, 2021).

2.5- Fundamentos estocásticos de séries temporais

2.5.1-Processo estocástico

Um processo estocástico é uma família de variáveis aleatórias que variam ao longo do tempo. Pode ser representado por $X = \{X(t): t \in T\}$, onde $X(t)$ é uma variável aleatória e T é o conjunto de índices ordenados. A classificação do processo depende da natureza do conjunto

de índices T . Caso T seja discreto, o processo é classificado como discreto; se for contínuo, é considerado um processo em tempo contínuo (Ross, 1996).

2.5.2- Estacionaridade (estrita e fraca)

Segundo Shumway e Stoffer (2025), os processos estocásticos podem ser classificados como estritamente estacionários ou fracamente estacionários. Um processo estocástico $\{X(t), t \in T\}$ é denominado **estritamente estacionário** se a sua distribuição conjunta $(X(t_1), X(t_2), \dots, X(t_n))$ for igual à distribuição conjunta de $(X(t_1 + h), X(t_2 + h), \dots, X(t_n + h))$, para qualquer $n \in \mathbb{N}$, quaisquer instantes $t_1, t_2, \dots, t_n \in T$ e qualquer deslocamento $h \in \mathbb{R}$ (ou $h \in \mathbb{Z}$ dependendo da natureza do processo). Por outras palavras, a sua distribuição conjunta permanece constante, independentemente do instante do tempo em que o processo é observado.

Contudo, é impraticável verificar se um processo estocástico é estritamente estacionário, na medida em que envolve o conhecimento de todas as distribuições conjuntas do processo. Na maioria dos casos, dispomos apenas de uma única realização do processo estocástico, ou seja, de uma única série temporal, o que impossibilita a observação de todas as distribuições possíveis. Por isso, a forma mais simples e prática de lidar com esta questão tem sido recorrer ao conceito de estacionaridade fraca ou estacionaridade de segunda ordem, que exige apenas as seguintes condições:

- $E[X_t^2] < +\infty$, ou seja, o processo tem variância finita;
- O valor médio $\mu_t = E(X_t) = \mu$, constante ao longo do tempo;
- A autocovariância γ_{t_1, t_2} depende apenas do desfasamento $h = t_2 - t_1$, e não dos instantes específicos t_1 e t_2 .

Nota: Ao longo deste trabalho, o termo *estacionaridade* será utilizado para designar a estacionaridade de segunda ordem (também designada por estacionaridade fraca).

2.5.3-Funções de dependência temporal

A **função de autocovariância**, para um processo estacionário, é definida como (Shumway e Stoffer, 2025):

$$\gamma_k = Cov[X_t, X_{t+k}] = E[(X_t - \mu)(X_{t+k} - \mu)], \quad (3)$$

onde k representa o desfasamento temporal entre observações.

Segundo Ehlers (2009), a forma clássica de estimar a função de autocovariância de um processo fracamente estacionário é dada pela seguinte expressão:

$$\hat{\gamma}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}), \quad (4)$$

onde $\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t$ representa a média amostral da série temporal.

A **função de autocorrelação (ACF)**, para um processo estacionário, é definida por (Shumway e Stoffer, 2025):

$$\rho_k = \text{corr}(X_t, X_{t+k}) = \frac{\text{Cov}(X_t, X_{t+k})}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t+k})}} = \frac{\text{Cov}(X_t, X_{t+k})}{\text{Var}(X_t)} = \frac{\gamma_k}{\gamma_0} \quad (5)$$

De forma clássica, a função de autocorrelação é estimada utilizando o seguinte estimador (Ehlers, 2009):

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{n-k} (X_t - \bar{X})^2} = \frac{\hat{\gamma}_k}{\gamma_0} \quad (6)$$

A função de autocorrelação goza das seguintes propriedades (Ehlers, 2009):

1. A função de autocorrelação é simétrica, ou seja, $\rho_k = \rho_{-k}$. Isto significa que a correlação entre X_t e X_{t+k} é igual à correlação entre X_t e X_{t-k} .
2. A função assume valores no intervalo $[-1, 1]$, ou seja, $-1 \leq \rho_k \leq 1$.

De acordo com Ehlers (2009), de forma análoga à função de autocorrelação, a função de autocorrelação parcial (PACF) representa a correlação condicional entre observações separadas por k períodos, condicionada às observações nos desfasamentos intermédios. Em outras palavras, a PACF no desfasamento k , denotada por ϕ_{kk} , mede a correlação direta entre X_t e X_{t+k} após remover o efeito das dependências lineares das observações intermédias X_{t+i} , $i = 1, \dots, k-1$. A PACF no desfasamento k , ϕ_{kk} , é dada por:

$$\rho_k = \sum_{i=1}^k \phi_{ki} \rho_{k-i} = \phi_{k1} \rho_{k-1} + \phi_{k2} \rho_{k-2} + \dots + \phi_{kk} \rho_0, \quad k > 1. \quad (7)$$

Como $\rho_0 = \frac{\gamma_0}{\gamma_0} = 1$, tem-se a seguinte expressão simplificada:

$$\rho_k = \phi_{k1} \rho_{k-1} + \phi_{k2} \rho_{k-2} + \dots + \phi_{kk}, \quad k > 1.$$

Onde:

- ρ_k é a autocorrelação no desfasamento;
- ϕ_{ki} são os coeficientes da PACF; e
- $\rho_{k-1}, \rho_{k-2}, \dots, \rho_0$ são as autocorrelações nos desfasamentos anteriores.

Para visualizar os coeficientes da ACF ou da PACF ao longo dos desfasamentos, utiliza-se o correlograma que é um gráfico que exibe os primeiros k coeficientes estimados em função do respetivo desfasamento. A sua análise revela padrões estruturais da série, tais como tendência, sazonalidade ou comportamento aleatório (Ehlers, 2009).

2.5.4-Exemplos clássicos de processos estocásticos

Sequência Aleatória

Segundo Ehlers (2009), sequência aleatória, também conhecida como **processo puramente aleatório** ou **ruído branco**, refere-se a uma série temporal constituída por variáveis aleatórias ε_t , independentes e que seguem a mesma distribuição (i.i.d.). Estas variáveis possuem uma média constante, $E(\varepsilon_t) = \mu$, e uma variância constante, $Var(\varepsilon_t) = \sigma_\varepsilon^2$.

Passeio aleatório

Segundo Ehlers (2009), chama-se **passeio aleatório** a um processo $\{X_t\}$, definido recursivamente por $X_t = X_{t-1} + \varepsilon_t$, onde $\{\varepsilon_t\}$ representa um processo discreto puramente aleatório, ou seja, uma sequência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com média μ e variância σ_ε^2 constantes.

Por meio de substituições sucessivas, é possível reescrever $X_t = X_{t-1} + \varepsilon_t$ como:

$$X_t = X_{t-2} + \varepsilon_{t-1} + \varepsilon_t = X_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t = \dots = X_0 + \sum_{j=1}^t \varepsilon_j.$$

Assumindo que o processo se inicia em $X_0 = 0$, tem-se:

$$X_t = \sum_{j=1}^t \varepsilon_j.$$

A esperança matemática do processo é dada por:

$$E(X_t) = \sum_{j=1}^t E(\varepsilon_j) = \sum_{j=1}^t \mu = \mu t.$$

Uma vez que a média varia com t , conclui-se que este processo é não estacionário.

2.5.5- Modelos para Séries Estacionárias

Processo de Média Móvel de ordem q

Segundo Ehlers (2009), uma série temporal X_t admite uma representação em **médias móveis de ordem q** , denotada por **MA(q)**, se for definida como:

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}. \quad (8)$$

Onde:

- ε_t representa um processo discreto puramente aleatório, com média $\mu_\varepsilon = 0$ e variância σ_ε^2 constante.
- θ_i , com $i = 1, \dots, q$, são os parâmetros do modelo, pertencentes ao conjunto dos números reais ($\mathbb{R}$), com $\theta_q \neq 0$. Assume-se que a série temporal possui média nula.

A esperança matemática (média) do processo X_t é dada por:

$$E(X_t) = E(\varepsilon_t) + \sum_{j=1}^q \theta_j E(\varepsilon_{t-j}) = 0.$$

A variância é dada por:

$$Var(X_t) = Var(\varepsilon_t) + \sum_{j=1}^q \theta_j^2 Var(\varepsilon_{t-j}) = (1 + \theta_1^2 + \dots + \theta_q^2) \sigma_\varepsilon^2.$$

A função de autocovariância é definida por:

$$\gamma_k = Cov(X_t, X_{t+k}) = \begin{cases} 0, & k > q \\ \sigma_\varepsilon^2 \sum_{j=0}^{q-k} \theta_j \theta_{j+k}, & k = 0, \dots, q \\ \gamma_{-k}, & k < 0 \end{cases}$$

com $\theta_0 = 1$.

E a função de autocorrelação (ACF):

$$\rho_k = \begin{cases} 1, & k = 0 \\ \frac{\sum_{j=0}^{q-k} \theta_j \theta_{j+k}}{\sum_{j=0}^q \theta_j^2}, & k = 1, \dots, q. \\ 0, & k > q \\ \rho_{-k}, & k < 0 \end{cases}$$

Esta série temporal pode ser considerada **fracamente estacionária**, também designada por **estacionária de 2ª ordem**, uma vez que a média e a variância se mantêm constantes ao longo do tempo e a autocovariância depende apenas da diferença $|t_2 - t_1|$, sendo, por conseguinte, independente do instante t , para qualquer combinação dos coeficientes $\theta_1, \dots, \theta_q$. A série será considerada **estritamente estacionária**, caso os termos ε_t sigam uma distribuição normal (Ehlers, 2009).

Considerando o caso específico em que $q = 1$ e a constante do modelo é nula, obtém-se o seguinte processo de média móvel de primeira ordem $MA(1)$ (Ehlers, 2009):

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}.$$

A sua variância será dada por:

$$Var(X_t) = (1 + \theta_1^2) \sigma_\varepsilon^2.$$

A sua função de autocovariância será dada por:

$$\gamma_k = Cov(X_t, X_{t+k}) = \begin{cases} 0 & |k| > 1 \\ (1 + \theta_1^2) \sigma_\varepsilon^2 & k = 0 \\ \theta_1 \sigma_\varepsilon^2 & |k| = 1 \end{cases}$$

E a sua função de autocorrelação (ACF) será dada por:

$$\rho_k = \begin{cases} 1 & k = 0 \\ \frac{\theta_1}{1 + \theta_1^2} & |k| = 1. \\ 0 & |k| > 0 \end{cases}$$

O processo sob análise é estacionário, independentemente do valor de θ_1 . Contudo, como diferentes valores de θ_1 podem resultar na mesma função de autocorrelação, é necessário recorrer ao conceito de invertibilidade para garantir a unicidade do modelo correspondente a cada função de autocorrelação (Ehlers, 2009).

A invertibilidade garante que os resíduos possam ser expressos de forma única como combinações lineares infinitas dos valores passados de X_t . Neste âmbito, o modelo:

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

pode ser reescrito de forma a isolar ε_t :

$$\varepsilon_t = X_t - \theta_1 X_{t-1} + \theta_1^2 X_{t-2} - \theta_1^3 X_{t-3} + \dots$$

Caso $|\theta_1| \geq 1$, a condição de invertibilidade não se verificaria, pois observações mais antigas teriam um peso maior sobre os valores ε_t , o que não faz sentido num modelo de séries temporais. Portanto, a condição $|\theta_1| < 1$ torna-se imperativa para garantir a invertibilidade, assegurando que observações mais antigas tenham menor influência sobre a variação X_t (Ehlers, 2009).

A condição de invertibilidade para um processo de **média móvel de ordem q , $MA(q)$** , é estabelecida utilizando o operador de atraso, denotado por B , descrito como:

$$B^j X_t = X_{t-j}, \text{ para todo } j.$$

A equação pode ser escrita como:

$$X_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) \varepsilon_t = \Theta(B) \varepsilon_t,$$

onde $\Theta(B)$ representa um polinómio de ordem q em B .

O processo é invertível se todas as raízes da equação característica $\Theta(z) = 0$ se situarem fora do círculo unitário (Ehlers, 2009).

Processo autorregressivo de ordem p

Segundo Ehlers (2009), uma série temporal X_t admite uma representação como **processo autorregressivo de ordem p , ou $AR(p)$** , se satisfaz a equação:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t, \quad (9)$$

onde ε_t é um processo puramente aleatório com média zero e variância constante σ_ε^2 , e os coeficientes ϕ_i , para $i = 1, \dots, p$, são constantes reais.

Considerando o caso mais simples, em que $p = 1$, obtém-se o seguinte processo autorregressivo de primeira ordem ($AR(1)$), conforme descrito por:

$$X_t = \phi_1 X_{t-1} + \varepsilon_t,$$

O modelo pode ser reescrito de forma equivalente como:

$$X_t = \phi_1^{r+1} X_{t-r-1} + \sum_{j=0}^r \phi_1^j \varepsilon_{t-j},$$

onde r representa o número de substituições recursivas (ou iterações) realizadas na expansão do processo $AR(1)$, com $r \in \mathbb{N}$.

Supondo que X_t representa um processo estacionário, com média nula e variância constante, podemos inferir que:

$$E \left[\left(X_t - \sum_{j=0}^r \phi_1^j \varepsilon_{t-j} \right)^2 \right] = \phi_1^{2r+2} E(X_{t-r-1}^2) = \phi_1^{2r+2} \sigma^2.$$

Caso $|\phi_1| < 1$ e $r \rightarrow \infty$, então ϕ_1^{2r+2} tende para zero, permitindo que o processo seja reescrito como:

$$X_t = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} + \dots$$

Esse processo é análogo a um processo de médias móveis de ordem infinita, $M(\infty)$, no qual o efeito dos valores passados diminui com o tempo, o que é esperado devido à natureza decrescente da função exponencial para $|\phi_1| < 1$.

Recorrendo ao operador de atraso B , definido por $B^j X_t = X_{t-j}$, a equação anterior pode ser escrita da seguinte forma:

$$X_t = \sum_{j=0}^{\infty} \phi_1^j B^j \varepsilon_t = \frac{1}{1 - \phi_1 B} \varepsilon_t.$$

Assumindo que $E(\varepsilon_t) = 0$, o processo X_t terá como propriedades:

Média:

$$E(X_t) = 0$$

Variância:

$$Var(X_t) = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2}$$

Função de autocovariância:

$$\gamma_k = \phi_1^k \text{Var}(X_t) = \phi_1^k \gamma_0 = \phi_1^k \frac{\sigma_\varepsilon^2}{1 - \phi_1^2}, k \geq 0$$

Função de autocorrelação:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1^k$$

Desta forma, conclui-se que o processo $AR(1)$, para $|\phi| < 1$, é estacionário, pois a média e a variância são constantes ao longo do tempo e a função de autocovariância γ_k depende apenas do desfasamento (lag) k , sendo, portanto, invariável no tempo.

Conforme Ehlers (2009), um processo autorregressivo de ordem p pode ser representado como um processo de médias móveis de ordem infinita, da seguinte forma:

$$X_t = \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots = (\psi_0 + \psi_1 B + \psi_2 B^2 + \dots) \varepsilon_t = \psi(B) \varepsilon_t.$$

Deste modo, o modelo será estacionário se a soma dos quadrados dos coeficientes ψ_j for finita, ou seja, $\sum_j \psi_j^2 < \infty$.

Recorrendo ao operador de atraso B , a equação:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t,$$

pode ser reescrita da seguinte forma:

$$X_t = \phi_1 B X_t + \dots + \phi_p B^p X_t + \varepsilon_t.$$

Após fatorizar, obtém-se,

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) X_t = \varepsilon_t$$

Ou, de forma matricial:

$$\phi(B) X_t = \varepsilon_t$$

$$\Leftrightarrow X_t = \phi(B)^{-1} \varepsilon_t$$

$$\Leftrightarrow X_t = \psi(B) \varepsilon_t$$

com $\psi(B) = \phi(B)^{-1}$.

O processo autorregressivo X_t é considerado estacionário se, e somente se, todas as raízes da equação $\phi(z) = 0$ estiverem localizadas fora do círculo unitário, ou seja, os valores absolutos das raízes devem ser maiores que 1.

Para um processo autorregressivo estacionário de ordem p , $AR(p)$, com valor médio $E(X_t) = 0$ e variância σ^2 , as funções de autocovariância e autocorrelação seguem as equações de Yule-Walker.

A função de autocovariância γ_k é dada por:

$$\gamma_k = E(X_t X_{t+k}) = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p} = \sum_{i=1}^p \phi_i \gamma_{k-i}, k \geq 1.$$

E a função de autocorrelação ρ_k (ACF), é dada por:

$$\rho_k = \sum_{i=1}^p \phi_i \rho_{k-i}, k \geq 1.$$

Modelos ARMA(p, q) e suas propriedades

Segundo Ehlers (2009), o modelo ARMA combina as representações autorregressiva e de médias móveis. Assim, uma série temporal X_t admite uma representação como processo autorregressivo e de médias móveis de ordens p e q , denotado por $ARMA(p, q)$, se satisfaz a equação:

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \quad (10)$$

onde ε_t é um processo puramente aleatório com média zero e variância constante σ_ε^2 , e os coeficientes $\phi_i (i = 1, \dots, p)$ e θ_j para $j = 1, \dots, q$ são constantes reais.

Utilizando o operador de atraso, o modelo pode ser reescrito como:

$$\Phi_p(B)X_t = \Theta_q(B)\varepsilon_t,$$

onde Φ_p e Θ_q são polinômios de ordem p e q , respetivamente.

Para que o processo seja estacionário, é necessário que as raízes de $\Phi_p(z) = 0$ estejam fora do círculo unitário. A invertibilidade do processo é garantida se as raízes de $\Theta_q(z) = 0$

estiverem fora do círculo unitário. Estas são as mesmas condições que garantem a estacionaridade e a invertibilidade nos processos autorregressivo e de médias móveis.

Função de autocovariância

De acordo com Box et al. (2015), para um processo $ARMA(p, q)$ estacionário, a função de autocovariância $\gamma_k = Cov(X_t, X_{t+k})$ depende do desfasamento k .

Para $0 \leq k \leq q$:

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} + \sigma_\varepsilon^2 \sum_{j=0}^{q-k} \theta_j \theta_{j+k},$$

onde $\theta_0 = 1$ por convenção e $\sigma_\varepsilon^2 = Var(\varepsilon_t)$.

Para $k > q$:

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p},$$

Para $k > q$, os termos de médias móveis não contribuem para a autocovariância porque os erros $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ não estão correlacionados com as observações X_{t-k} .

Função de autocorrelação

A função de autocorrelação é definida por:

$$\rho_k = \frac{\gamma_k}{\gamma_0}.$$

Ou seja, o valor de ρ_k depende exclusivamente do desfasamento k .

Para $k = 0$:

$$\rho_0 = \frac{\gamma_0}{\gamma_0} = 1.$$

Para $1 \leq k \leq q$:

Os valores $\rho_1, \rho_2, \dots, \rho_q$ dependem simultaneamente dos parâmetros autorregressivos ϕ_i e dos parâmetros de médias móveis θ_i . São obtidos por meio da resolução do sistema de equações lineares formado pelas equações de autocovariância para os desfasamentos $k = 0, 1, \dots, q$.

Após o cálculo das autocovariâncias $\gamma_0, \gamma_1, \dots, \gamma_q$, determina-se:

$$\rho_k = \frac{\gamma_k}{\gamma_0}.$$

Para $k > q$:

$$\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p}.$$

Nota: Para $k > q$, o comportamento de γ_k e de ρ_k é igual ao de um processo autorregressivo puro $AR(p)$.

2.5.6-Modelos para Séries Não Estacionárias

Necessidade de diferenciação

De acordo com Ehlers (2009), quando se trata de séries temporais não estacionárias, os modelos anteriormente citados como o Autorregressivo ($AR(p)$), de Médias Móveis ($MA(q)$), e o Autorregressivo de Médias Móveis ($ARMA(p, q)$), são inapropriados para a sua modelação, uma vez que estes modelos partem do princípio de que as séries são estacionárias, ou seja, que características como a média e a variância se mantêm invariantes no tempo. Assim, para lidar com séries temporais não estacionárias, pode-se recorrer a modelos como o Autorregressivo e de Médias Móveis Integrado de ordem p , d e q , ou seja, $ARIMA(p, d, q)$. Estes modelos combinam três componentes: o **Autorregressivo (AR)**, a **Integração (I)** e as **Médias Móveis (MA)**.

Antes de proceder à diferenciação, é fundamental verificar a estabilidade da variância da série. Na presença de heterocedasticidade, recorre-se a transformações matemáticas, como a transformação logarítmica ou a transformação de Box-Cox (Box & Cox, 1964). Esta etapa prévia assegura a homocedasticidade da variância ao longo do tempo, cumprindo um pressuposto essencial para a correta modelação.

Modelo $ARIMA(p, d, q)$

Os modelos $ARIMA$ transformam uma série temporal não estacionária X_t , numa série estacionária W_t através da aplicação de uma ou mais diferenciações simples. A série transformada é definida como:

$$W_t = \nabla^d X_t = (1 - B)^d X_t,$$

onde:

- d é o número de diferenças necessárias para transformar a série não estacionária X_t numa série estacionária W_t ;
- $\nabla = 1 - B$ é o operador de primeira diferença;
- $\nabla^d = (1 - B)^d$ é o operador de diferenciação de ordem d .

Sendo o modelo autorregressivo de médias móveis integrado ($ARIMA(p, d, q)$) dado pela equação:

$$W_t = \phi_1 W_{t-1} + \dots + \phi_p W_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}. \quad (11)$$

Ou, de forma matricial:

$$\Phi(B)(1 - B)^d X_t = \Theta(B)\varepsilon_t,$$

onde:

- ε_t é um processo puramente aleatório com média zero e variância constante σ_ε^2 ;
- $\Phi(B)$ é o polinómio autorregressivo de ordem p ;
- $\Theta(B)$ é o polinómio de médias móveis de ordem q ;
- Os parâmetros ϕ_i ($i = 1, \dots, p$) e θ_j ($j = 1, \dots, q$) são constantes reais.

Para que o processo seja estacionário e invertível, as raízes de todos os polinómios ($\Phi(z)$ e $\Theta(z)$) devem situar-se fora do círculo unitário.

Uma das limitações deste modelo é o facto de não ser adequado para lidar com séries que apresentam um padrão sazonal.

Modelos $SARIMA(p, d, q)(P, D, Q)$

De acordo com Ehlers (2009), o processo autorregressivo e de médias móveis integrado (ARIMA), acima descrito, é adequado apenas para modelar séries temporais não estacionárias que não apresentam um padrão de sazonalidade claro, podendo ser útil para séries temporais com tendência, a qual é removida por meio de diferenciações. Assim, no caso de séries temporais com uma componente sazonal, recorre-se aos **processos autorregressivos e de médias móveis integrado sazonal (SARIMA)** para a sua modelação. Para garantir que a série se torne estacionária nestes modelos, é necessário efetuar uma ou

mais diferenciações múltiplas associadas ao padrão sazonal s . Combinando o operador de diferenciação regular de ordem d (∇^d) com o operador de diferenciação sazonal de ordem D e período s (∇_s^D), obtemos o operador $\nabla^d \nabla_s^D$, que remove tanto a tendência como a sazonalidade da série temporal.

O modelo ARIMA sazonal é uma extensão do modelo $ARIMA(p, d, q)$, desenvolvido para integrar a componente sazonal, resultando no modelo SARIMA multiplicativo de ordem $(p, d, q) \times (P, D, Q)_s$, representado por:

$$\phi(B)\Phi(B^s)W_t = \theta(B)\Theta(B^s)\varepsilon_t, \quad (12)$$

onde:

- $W_t = \nabla^d \nabla_s^D X_t$
- $\nabla^d = (1 - B)^d$
- $\nabla_s^D = (1 - B^s)^D$
- p : ordem da componente autorregressiva (AR)
- d : grau de diferenciação
- q : ordem da componente de médias móveis (MA)
- P : ordem da componente autorregressiva sazonal
- D : grau de diferenciação sazonal
- Q : ordem da componente de médias móveis sazonal
- s : período sazonal
- $\varepsilon_t \sim WN(0, \sigma^2)$
- $\phi(B)$: polinómio autorregressivo regular
- $\Phi(B^s)$: polinómio autorregressivo sazonal
- $\theta(B)$: polinómio de média móvel regular
- $\Theta(B^s)$: polinómio de média móvel sazonal

Para que o processo seja estacionário, as raízes dos polinómios autorregressivos $\phi(z) = 0$ e $\Phi(z^s) = 0$ devem situar-se fora do círculo unitário. Para que o processo seja invertível, as raízes dos polinómios de médias móveis $\theta(z) = 0$ e $\Theta(z^s) = 0$ devem igualmente situar-se fora do círculo unitário.

As ordens p, q, P e Q são determinadas a partir da análise das funções de autocorrelação (ACF) e autocorrelação parcial (PACF). As ordens d e D dificilmente assumem valores

superiores a 1, pois, geralmente, uma única diferenciação é suficiente para tornar a série estacionária.

2.6- Métodos determinísticos (alisamento exponencial)

Os métodos de alisamento exponencial surgiram no final dos anos 1950 com base nos trabalhos realizados por Brown (1959), Holt (1957) e Winters (1960), tendo contribuído para o desenvolvimento de técnicas de previsão mais eficazes. Nesta abordagem, as previsões são calculadas através de médias ponderadas das observações passadas, em que os pesos atribuídos diminuem exponencialmente com o tempo, o que significa que as observações mais recentes têm maior influência nas previsões geradas (Hyndman e Athanasopoulos, 2021).

2.6.1-Alisamento exponencial simples

Segundo Hyndman e Athanasopoulos (2021), quando se trata de uma série temporal estacionária, isto é, que não possui tendência nem sazonalidade, o método mais adequado para antecipar o seu comportamento é o **alisamento exponencial simples**. Considerando uma série temporal X_t estacionária, a previsão de um valor futuro $\hat{X}_{t+1}$ é obtida através de uma média ponderada entre o valor atual da série e a previsão anterior, com pesos definidos por um parâmetro de suavização $\alpha \in [0,1]$.

O modelo é representado pela seguinte equação:

$$\hat{X}_{t+1} = \alpha X_t + (1 - \alpha)\hat{X}_t, \quad (13)$$

onde:

- $\hat{X}_{t+1}$ é a previsão no próximo período ($t + 1$),
- X_t é o valor observado no instante t ,
- $\hat{X}_t$ é a previsão efetuada no período t ,
- α é o parâmetro de suavização que determina a importância atribuída ao valor mais recente da série.

Alternativamente, o modelo pode ser representado na **forma por componentes**, introduzindo o conceito de nível da série no instante t , representado por L_t . Neste caso, o modelo é dado pelas seguintes equações:

$$\begin{cases} \hat{X}_{t+1} = L_t \\ L_t = \alpha X_t + (1 - \alpha)L_{t-1} \end{cases}'$$

onde L_t representa a estimativa suavizada do nível no instante t , sendo atualizada constantemente com base no valor observado X_t e o nível estimado no período L_{t-1} . Quando se pretende prever h passos à frente ($h > 0$), o valor estimado é igual ao último nível observado:

$$\hat{X}_{t+h} = L_t.$$

Uma das limitações do método é não ser adequado para realizar previsões em séries com tendência ou sazonalidade.

2.6.2-Alisamento exponencial de Holt

De acordo com Hyndman e Athanasopoulos (2021), o método de alisamento exponencial de Holt constitui uma generalização do alisamento exponencial simples para a realização de previsões em séries que apresentam tendência linear, seja positiva ou negativa. O método é definido pelas três equações:

$$\begin{aligned} \text{Previsão:} \quad & \hat{X}_{t+h} = L_t + b_t h \quad \text{para } h > 0 \\ \text{Nível:} \quad & L_t = \alpha X_t + (1 - \alpha)(L_{t-1} + b_{t-1}) \\ \text{Tendência:} \quad & b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}, \end{aligned} \tag{14}$$

onde:

- L_t representa uma estimativa do nível da série no instante t ,
- b_t representa uma estimativa da tendência (inclinação) da série no instante t ,
- $\alpha \in [0,1]$ é o parâmetro de alisamento para o nível,
- $\beta \in [0,1]$ é o parâmetro de alisamento para a tendência.

Uma das limitações do método é não ser adequado para realizar previsões em séries que apresentam um padrão sazonal.

2.6.3-Alisamento exponencial de Holt-Winters

Segundo Hyndman e Athanasopoulos (2021) e Makridakis, Wheelwright e Hyndman (1997) os modelos de alisamento exponencial simples e de Holt, acima descritos, não são adequados

para realizar previsões em séries temporais que possuem as componentes de tendência e sazonalidade em simultâneo. Nesses casos, os autores recomendam a utilização do modelo de **Holt-Winters**, que é uma extensão do alisamento exponencial simples, desenvolvida para captar diretamente a sazonalidade da série.

O modelo é composto pelos parâmetros de alisamento α , β e γ , estimados pelo método dos mínimos quadrados. Correspondem ao nível médio da série (L_t), à sua tendência (b_t) e ao padrão sazonal (s_t), respetivamente.

O método de Holt-Winters apresenta duas variantes. A escolha entre elas depende da natureza da sazonalidade observada na série temporal: aditiva ou multiplicativa.

Modelo aditivo de Holt-Winters

A variante aditiva aplica-se quando a amplitude das variações sazonais não muda com o nível da série temporal, sendo caracterizado pelas seguintes equações (Makridakis, Wheelwright e Hyndman, 1997):

$$\begin{aligned}
 \text{Nível:} & \quad L_t = \alpha(X_t - s_{t-s}) + (1 - \alpha)(L_{t-1} + b_{t-1}), \\
 \text{Tendência:} & \quad b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}, \\
 \text{Sazonalidade:} & \quad S_t = \gamma(X_t - L_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-s}, \\
 \text{Previsão:} & \quad \hat{X}_{t+h} = L_t + b_t h + s_{t-s+h},
 \end{aligned} \tag{15}$$

onde:

- S_t : **Componente sazonal** no instante t .
- $\gamma \in [0,1]$: Parâmetro de suavização da sazonalidade.
- s : Período da sazonalidade, ou seja, o número de períodos que compõem um ciclo sazonal completo (por exemplo, $s = 12$ para dados mensais com sazonalidade anual).

Modelo multiplicativo de Holt-Winters

A variante multiplicativa aplica-se quando a amplitude sazonal depende do nível da série temporal. É caracterizada pelas seguintes equações (Makridakis, Wheelwright e Hyndman, 1997):

$$\begin{aligned}
\text{Nível:} \quad L_t &= \alpha \frac{X_t}{s_{t-s}} + (1 - \alpha)(L_{t-1} + b_{t-1}), \\
\text{Tendência:} \quad b_t &= \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}, \\
\text{Sazonalidade:} \quad S_t &= \gamma \frac{X_t}{L_t} + (1 - \gamma)s_{t-s}, \\
\text{Previsão:} \quad \hat{X}_{t+h} &= (L_t + b_t h)s_{t-s+h}.
\end{aligned} \tag{16}$$

Os parâmetros de alisamento α , β e γ pertencem ao intervalo $[0,1]$.

2.7- Diagnóstico e validação

2.7.1-Teste de raiz unitária

A verificação da estacionariedade constitui uma etapa fundamental na análise de séries temporais, permitindo avaliar a adequação da série à aplicação de determinados métodos de modelação. A presença de uma raiz unitária indica que a série não é estacionária, devendo, por isso, ser submetida a diferenciação (Hyndman & Athanasopoulos, 2021). Para avaliar a presença de raízes unitárias, recorre-se frequentemente a três procedimentos complementares:

- Testes de **Dickey-Fuller Aumentado** e **Phillips-Perron**: Partem da hipótese nula de não estacionariedade (presença de raiz unitária).
- Teste de **Kwiatkowski-Phillips-Schmidt-Shin**: Parte da hipótese nula de estacionariedade.

Esta diferença na formulação das hipóteses implica que a regra de decisão e a interpretação dos resultados dependem do teste utilizado.

Dada uma série temporal X_t , a presença de uma raiz unitária num processo AR(p) significa que o polinómio autorregressivo

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

satisfaz a condição $\phi(1) = 0$. Isto acontece quando

$$1 - \phi_1 - \phi_2 - \dots - \phi_p = 0$$

ou, de forma equivalente,

$$\sum_{i=1}^p \phi_i = 1.$$

Teste Dickey-Fuller Aumentado (ADF)

Proposto por Dickey e Fuller (1979), o teste parte do pressuposto de que a série possui raiz unitária (não é estacionária) até que se prove o contrário.

Fundamentação Matemática

Definindo $\delta = \sum_{i=1}^p \phi_i - 1$, a hipótese nula de raiz unitária passa a ser expressa de forma equivalente por $H_0: \delta = 0$.

Para testar esta hipótese, transforma-se o modelo original em termos da primeira diferença, $\Delta X_t = X_t - X_{t-1}$, obtendo-se a versão aumentada do teste:

$$\Delta X_t = \delta X_{t-1} + \sum_{i=1}^k \gamma_i \Delta X_{t-i} + \varepsilon_t,$$

em que os termos ΔX_{t-i} (com $i = 1, \dots, k$) visam corrigir possíveis autocorrelações nos resíduos, daí a designação “aumentado”. Consoante as características da série, o modelo pode ainda incluir:

- uma constante a_0 , para capturar uma média não nula:

$$\Delta X_t = a_0 + \delta X_{t-1} + \sum_{i=1}^k \gamma_i \Delta X_{t-i} + \varepsilon_t$$

- ou uma constante e uma tendência determinística $a_1 t$, quando a série exibe tendência linear:

$$\Delta X_t = a_0 + a_1 t + \delta X_{t-1} + \sum_{i=1}^k \gamma_i \Delta X_{t-i} + \varepsilon_t$$

A escolha entre estas versões do modelo depende das propriedades estocásticas e determinísticas da série em análise, e influencia diretamente os valores críticos utilizados no teste (Enders, 2014).

Hipóteses

$$\begin{cases} H_0: \delta = 0, & \text{a série não é estacionária (possui raiz unitária)} \\ H_1: \delta < 0, & \text{a série é estacionária} \end{cases}$$

Estatística de teste

A estatística de teste corresponde ao valor- t de regressão associado ao coeficiente δ , definido por:

$$t_{\delta} = \frac{\hat{\delta}}{SE(\hat{\delta})}$$

onde $\hat{\delta}$ é o estimador de mínimos quadrados e $SE(\hat{\delta})$ é o erro padrão da estimativa.

Distribuições Assintóticas:

Segundo Hamilton (1994), considerando $W(r)$, com $r \in [0,1]$, um processo de Wiener padrão, definido como um processo estocástico de tempo contínuo que satisfaz as seguintes propriedades:

1. $W(0) = 0$;
2. Para quaisquer $0 \leq s < t \leq 1$, o incremento $W(t) - W(s)$ segue uma distribuição normal com média zero e variância $t - s$, sendo os incrementos independentes, isto é, $W(t) - W(s) \sim N(0, t - s)$;
3. Continuidade: $W(s)$ é contínuo em s .

O mesmo autor demonstra que a distribuição limite da estatística t_{δ} depende da especificação do modelo:

1. Modelo sem constante nem tendência

$$t_{\delta} \Rightarrow \frac{\frac{1}{2}(W(1)^2 - 1)}{\left(\int_0^1 W(r)^2 dr\right)^{\frac{1}{2}}}$$

2. Modelo com constante

$$t_{\delta} \Rightarrow \frac{\frac{1}{2}(W(1)^2 - 1) - W(1) \int_0^1 W(r) dr}{\left(\int_0^1 W(r)^2 dr - \left(\int_0^1 W(r) dr\right)^2\right)^{\frac{1}{2}}}$$

3. Modelo com constante e tendência linear:

$$t_{\delta} \Rightarrow \frac{\frac{1}{2}(\tilde{W}(1)^2 - 1)}{\left(\int_0^1 \tilde{W}(r)^2 dr\right)^{\frac{1}{2}}},$$

onde $\tilde{W}(r) = W(r) + (6r - 4) \int_0^1 W(s)ds - (12r - 6) \int_0^1 sW(s)ds$.

Regra de decisão: Rejeita-se H_0 ao nível de significância α , concluindo que a série é estacionária, se: $t_{\delta} < \text{valor crítico}(\alpha)$. Caso contrário, não se rejeita H_0 , e recomenda-se a diferenciação da série.

Os valores críticos do teste de Dickey-Fuller correspondem aos percentis da distribuição limite não padrão derivada por Dickey e Fuller (1979), obtidos por simulação de Monte Carlo e tabelados em Dickey (1976) (Anexo IV).

Teste Phillips-Perron (PP)

O teste de Phillips-Perron (PP), proposto por Phillips e Perron (1988), constitui uma alternativa ao teste ADF para detecção de raízes unitárias. Ambos os testes partilham a mesma hipótese nula e as mesmas distribuições assintóticas, diferenciando-se fundamentalmente na forma como tratam a autocorrelação e a heterocedasticidade dos resíduos.

Hipóteses

As hipóteses são idênticas às do teste ADF:

$$\begin{cases} H_0: \delta = 0, & \text{a série não é estacionária (possui raiz unitária)} \\ H_1: \delta < 0, & \text{a série é estacionária} \end{cases}$$

Estatística de Teste

Ao contrário do teste ADF, que adiciona termos de diferenças desfasadas à equação de regressão para corrigir a autocorrelação (abordagem paramétrica), o teste PP utiliza a equação de Dickey-Fuller original (sem *lags* aumentados) e aplica uma correção não-paramétrica à estatística t_{δ} .

A estatística de teste do PP, denotada por $Z_{t_{\delta}}$, é calculada ajustando o valor- t do estimador de mínimos quadrados ordinários (MQO) para ter em conta a estrutura de dependência serial e heterocedasticidade dos resíduos, recorrendo à estimativa de variância de longo prazo de

Newey-West. Desta forma, o teste PP é geralmente mais robusto na presença de heterocedasticidade nos resíduos do que o teste ADF.

Distribuições Assintóticas:

Sob a hipótese nula, a estatística Z_{t_δ} converge para as mesmas distribuições funcionais de um processo de Wiener apresentadas para o teste ADF. Consequentemente, as expressões matemáticas para os modelos sem constante, com constante e com constante e tendência linear são idênticas às descritas na secção anterior.

Regra de Decisão: A regra de decisão é análoga à do teste ADF: rejeita-se H_0 ao nível de significância α se $Z_{t_\delta} < \text{valor crítico}(\alpha)$ concluindo-se pela estacionariedade da série. Os valores críticos utilizados são os mesmos tabelados por Dickey (1976).

Teste Kwiatkowski–Phillips–Schmidt–Shin (KPSS)

Este teste, desenvolvido por Kwiatkowski et al. (1992), parte do pressuposto de que a série é estacionária (em torno de um nível ou de uma tendência determinística) até que se prove o contrário. Diferentemente do ADF e do PP, a hipótese nula aqui assume a estacionariedade.

Hipóteses do teste:

$$\begin{cases} H_0: \sigma_\varepsilon^2 = 0, & \text{a série é estacionária} \\ H_1: \sigma_\varepsilon^2 > 0, & \text{a série não é estacionária} \end{cases}$$

A estatística do teste é um multiplicador de Lagrange (LM) dado por:

$$LM = \frac{1}{T^2} \sum_{t=1}^T \frac{S_t^2}{\hat{\sigma}^2},$$

onde $S_t = \sum_{i=1}^t e_i$ é a soma acumulada dos resíduos e $\hat{\sigma}^2$ é um estimador consistente da variância de longo prazo.

Distribuições Assintóticas

Considerando $W(r)$, com $r \in [0,1]$, um processo de Wiener padrão, e definindo a Ponte Browniana como $V(r) = W(r) - rW(1)$, a distribuição limite da estatística LM depende da especificação do modelo (Kwiatkowski et al., 1992):

1. Modelo com constante (Estacionariedade de Nível):

$$LM \Rightarrow \int_0^1 V(r)^2 dr$$

2. Modelo com constante e tendência linear (Estacionariedade de Tendência):

$$LM \Rightarrow \int_0^1 V^*(r)^2 dr$$

onde $V^*(r) = W(r) + (6r - 4) \int_0^1 W(s)ds - (12r - 6) \int_0^1 sW(s)ds$.

Sob H_0 , esta estatística segue uma distribuição assintótica não padrão, especificamente um funcional do movimento browniano ponte (*Brownian bridge*) (Kwiatkowski et al., 1992).

Regra de decisão: Rejeita-se H_0 ao nível de significância α se $LM > \text{valor crítico}(\alpha)$, onde os valores críticos são obtidos da distribuição assintótica não-padrão do funcional de *Brownian Bridge*, conforme tabelado em Kwiatkowski et al. (1992). Os valores críticos para diferentes níveis de significância estão apresentados no Anexo V.

Por serem distribuições de difícil tratamento analítico, na prática, compara-se o valor da estatística de teste (calculado computacionalmente) com os valores críticos tabelados específicos para cada procedimento. Para os testes ADF e PP, utilizam-se os valores críticos de Dickey (1976); para o teste KPSS, recorrem-se aos valores críticos propostos por Kwiatkowski et al. (1992). Em todos os casos, a escolha dos valores críticos adequados deve ser feita consoante a especificação do modelo (sem constante, com constante, ou com constante e tendência).

2.7.2-Análise dos resíduos

No contexto da modelação de séries temporais, assume-se que os resíduos do modelo sejam não autocorrelacionados. Este pressuposto é fundamental para garantir a validade estatística e a fiabilidade preditiva dos modelos ajustados. Segundo Chatfield (2003), a presença de autocorrelação nos resíduos indica que o modelo não capturou integralmente a dinâmica subjacente à série, podendo conduzir a previsões enviesadas ou ineficientes e à subestimação da incerteza associada às estimativas, o que compromete a amplitude e a credibilidade dos intervalos de previsão.

A verificação desse pressuposto pode ser realizada através de testes estatísticos específicos. Entre os mais utilizados encontram-se os testes de *portmanteau*, originalmente propostos por

Box e Jenkins em 1970 e amplamente descritos por Ehlers (2009). As versões mais conhecidas são o teste de **Box–Pierce** e o de **Ljung–Box**. Esses testes avaliam, de forma conjunta, se as m primeiras autocorrelações amostrais dos resíduos diferem significativamente de zero, permitindo identificar dependências temporais remanescentes após o ajustamento do modelo. Formalmente, testam-se as hipóteses:

$$H_0: \rho_1 = \rho_2 = \rho_3 = \dots = \rho_m = 0$$

$$H_1: \rho_k \neq 0, \text{ para pelo menos um } k \in \{1, \dots, m\}.$$

em que m representa o número de autocorrelações consideradas no teste.

Segundo Hyndman e Athanasopoulos (2021), os resíduos devem também apresentar média nula, uma vez que resíduos com média diferente de zero indicam a existência de viés nas previsões. Essa condição pode ser avaliada por meio de um teste t aplicado à média dos resíduos.

No teste de **Box–Pierce**, a estatística de teste é dada por (Ehlers, 2009):

$$Q = n \sum_{k=1}^m \hat{\rho}_k^2. \quad (17)$$

Em amostras pequenas, o teste de Box–Pierce torna-se menos fiável, pois a distribuição da estatística Q deixa de aproximar-se adequadamente da distribuição qui-quadrado teórica. Nestes casos, recomenda-se a aplicação do teste de Ljung-Box (Ehlers, 2009), cuja estatística de teste corrigida é:

$$Q^* = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k}. \quad (18)$$

Esta correção melhora o desempenho do teste em amostras pequenas.

Sob H_0 , tanto Q quanto Q^* seguem assintoticamente uma distribuição χ_{m-p-q}^2 , onde p e q são, respetivamente, os números de parâmetros autorregressivos (AR) e de médias móveis (MA) estimados no modelo.

Assim, ao nível de significância de 5%, rejeita-se a hipótese nula sempre que $Q^* > \chi_{\alpha, m-p-q}^2$. Nesse caso, conclui-se que há evidência estatística de autocorrelação residual,

ou seja, os resíduos não se comportam como ruído branco, pelo que o modelo ajustado não é adequado, uma vez que não capturou toda a estrutura dependente da série temporal.

2.8- Critérios para a seleção de modelos de séries temporais

Na análise de séries temporais, nenhum modelo consegue capturar perfeitamente todos os aspetos da realidade subjacente, o que torna indispensável considerar e testar diferentes abordagens. Para uma mesma série temporal, podem ser ajustados diversos modelos, como os anteriormente apresentados, bem como o mesmo modelo com diferentes parametrizações. Por isso, torna-se necessário definir critérios para avaliar a qualidade dos modelos ajustados, de forma a permitir a escolha daquele que apresenta o melhor ajustamento aos dados observados.

De acordo com Ehlers (2009), perante a adequação de vários modelos ao mesmo conjunto de dados observados, pode-se recorrer ao critério mais amplamente utilizado que é o ***Critério de Informação de Akaike (AIC)***, para auxiliar na seleção do modelo mais apropriado. Assim, quanto menor for o valor do **AIC**, mais adequado será considerado o modelo.

Segundo Hyndman e Athanasopoulos (2021), no contexto de modelos de séries temporais com erros normalmente distribuídos, o **AIC** é dado por:

$$AIC = n \log \left(\frac{SSE}{n} \right) + 2(k + 2), \quad (19)$$

onde n é o número de observações, k é o número de parâmetros do modelo e o SSE é a soma dos quadrados dos erros.

No entanto, os mesmos autores apontam que uma das limitações do **AIC**, para amostras de dimensões reduzidas, é a sua tendência para favorecer modelos com um número excessivo de parâmetros. Para ultrapassar essa limitação, foram desenvolvidos outros critérios, como o ***Critério de Informação de Akaike corrigido (AICc)*** e o ***Critério de Informação Bayesiano (BIC)***, variantes do **AIC**, que penalizam a inclusão de parâmetros adicionais no modelo.

De acordo com Hyndman e Athanasopoulos (2021), o **AICc**, no contexto de modelos de séries temporais com erros normalmente distribuídos, é dado por:

$$AICc = AIC + \frac{2(k+2)(k+3)}{n-k-3}, \quad (20)$$

E o **BIC** é dado por:

$$BIC = n \log \left(\frac{SSE}{n} \right) + (k+2) \log(n), \quad (21)$$

onde n é o número de observações, k é o número de parâmetros do modelo e o SSE é a soma dos quadrados dos erros.

Assim como no caso do critério **AIC**, os modelos com menores valores de **AICc** e **BIC** são preferidos.

Adicionalmente, a seleção do modelo deve reger-se pelo Princípio da Parcimónia, o qual preconiza que, entre modelos com desempenho preditivo e capacidade de ajustamento semelhantes, deve ser preferido aquele com o menor número de parâmetros.

2.9-Previsão em séries temporais

Um dos objetivos da modelação por séries temporais, como descrito na Secção 2.1.2, é efetuar previsões de observações futuras com base nas observações passadas. Independentemente do modelo final selecionado, a previsão pontual para um horizonte $h > 0$ consiste na estimação do valor da série no instante futuro $t + h$, utilizando todas as observações disponíveis até ao instante t e as estimativas dos parâmetros do modelo.

Assim, dada uma série temporal $\{X_t\}$, a previsão de X_{t+h} é dada pelo valor esperado condicional da observação futura, dados os valores observados até ao instante t . Sob a hipótese de que os erros do modelo são independentes e normalmente distribuídos, esta previsão é dada por (Hyndman & Athanasopoulos, 2021):

$$\hat{X}_{t+h|t} = E(X_{t+h} | X_t, X_{t-1}, X_{t-2}, \dots). \quad (22)$$

Contudo, a previsão pontual não fornece informações sobre a incerteza associada à previsão futura. Para colmatar esta limitação, recorre-se à construção de intervalos de previsão, que indicam um intervalo plausível no qual se espera que a observação futura venha a ocorrer, com um determinado nível de confiança (por exemplo, 80 % ou 95 %).

O intervalo de previsão a $(1 - \alpha) \times 100\%$ para o horizonte h , sob a hipótese de que os resíduos do modelo são independentes e normalmente distribuídos, é dado por (Hyndman & Athanasopoulos, 2021):

$$\hat{X}_{t+h|t} \pm z_{1-\alpha/2} \hat{\sigma}_h, \quad (23)$$

onde:

- $\hat{X}_{t+h|t}$ é a previsão pontual;
- $z_{1-\alpha/2}$ é o quantil da distribuição normal padrão correspondente ao nível de confiança pré-definido;
- $\hat{\sigma}_h$ é o desvio padrão de erro de previsão.

Importa salientar que a fórmula utilizada para os intervalos de previsão assume a normalidade dos resíduos do modelo. Na prática, se esta hipótese falhar, os intervalos podem não ter exatamente a precisão do nível de confiança declarado. Contudo, nesta dissertação, como estamos a trabalhar com amostras de grande dimensão (várias décadas de observações mensais), a aproximação normal mantém-se adequada, apoiada pelas propriedades assintóticas dos estimadores.

No presente trabalho, as previsões são realizadas após a seleção do modelo mais adequado, isto é, daquele cujos resíduos não apresentem evidência de autocorrelação ($p > 0,05$ no teste de Ljung-Box) e, dentre os modelos que satisfazem essa condição, aquele que minimize os critérios *AICc* e *BIC*, em conformidade com as recomendações de diagnóstico e seleção de modelos de Hyndman & Athanasopoulos (2021).

2.10-Medidas de avaliação

Segundo Hyndman e Athanasopoulos (2021), existem várias medidas para avaliar o desempenho de modelos de séries temporais. Estas medidas podem ser aplicadas tanto na avaliação da qualidade do ajustamento do modelo aos dados observados como na avaliação da precisão das previsões. No primeiro caso, os erros (resíduos) são obtidos pela comparação entre os valores observados e os valores ajustados pelo modelo na amostra de treino. No segundo, a avaliação é realizada comparando os valores previstos com os valores efetivamente observados na amostra de teste, quando esses valores se tornam disponíveis. Desta forma, é possível quantificar a verdadeira capacidade preditiva do modelo.

Nesta dissertação, a avaliação do desempenho dos modelos incidirá exclusivamente sobre a precisão das previsões (*out-of-sample*), através da comparação entre os valores previstos e os valores reais observados no período de teste, compreendido entre julho de 2023 e junho de 2024. Para esse efeito, serão utilizadas duas categorias de medidas de erro:

Medidas dependentes da escala

Estas métricas avaliam a qualidade das previsões na mesma unidade em que os dados da série estão expressos, pelo que o seu valor depende diretamente da magnitude e da escala da própria série temporal. Por esta razão, não devem ser utilizadas para comparar séries com escalas diferentes.

As duas medidas dependentes da escala mais utilizadas são o ***Erro Absoluto Médio (MAE)*** e a ***Raiz do Erro Quadrático Médio (RMSE)***.

O ***Erro Absoluto Médio*** é dado por:

$$MAE = \frac{1}{n} \sum_{t=1}^n |X_t - \hat{X}_t| = \frac{1}{n} \sum_{t=1}^n |e_t|. \quad (24)$$

A ***Raiz do Erro Quadrático Médio*** é dada por:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (X_t - \hat{X}_t)^2} = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2}. \quad (25)$$

Dada a sua simplicidade de cálculo e facilidade de interpretação, o ***MAE*** é uma das métricas mais utilizadas para comparar diferentes modelos aplicados a uma mesma série temporal.

Medidas baseadas em erros percentuais

O erro percentual é calculado por $p_t = \frac{X_t - \hat{X}_t}{X_t} \times 100$. Uma das principais vantagens desta medida é a sua independência em relação à escala dos dados observados, o que permite comparar o desempenho das previsões em diferentes séries temporais. Por esta razão, o erro percentual é uma das métricas mais frequentemente utilizadas. No entanto, esta medida apresenta algumas limitações, nomeadamente o facto de poderem assumir valores infinitos ou indefinidos se $X_t = 0$, para qualquer t , além de apresentarem valores extremos quando algum X_t está próximo de zero.

A medida mais utilizada é o ***Erro Percentual Absoluto Médio (MAPE)***, definido como a média dos valores absolutos dos erros percentuais:

$$MAPE = \frac{1}{n} \sum_{t=1}^n |p_t|. \quad (26)$$

Uma outra limitação das medidas baseadas em erros percentuais é que elas pressupõem a existência de um zero significativo, ou seja, que um valor de zero representa a ausência total daquilo que está a ser medido. Assim sendo, não são adequadas para medir a precisão de previsões de temperatura, porque a temperatura possui um ponto zero arbitrário.

3- Capítulo III – Análise e Modelação da Temperatura do Mar

3.1- Introdução

O presente capítulo analisa as séries temporais de temperatura média mensal do mar, observadas entre janeiro de 1968 e junho de 2024 nas localidades de Nazaré, Faro e São Miguel (Portugal) e Praia (Cabo Verde), nas profundidades de 1 m (superfície) e 5 m. Organiza-se da seguinte forma: após uma breve descrição dos dados, realiza-se uma análise exploratória, tanto gráfica como descritiva; em seguida, avalia-se a estacionaridade das séries por meio de testes estatísticos formais (ADF, PP e KPSS), comparam-se distintas abordagens de modelação, nomeadamente decomposição clássica, decomposição STL, alisamento exponencial de Holt-Winters e modelos SARIMA, e procede-se à seleção do modelo final com base na ausência de autocorrelação residual (teste de Ljung–Box) e nos critérios de informação AICc e BIC. Por fim, apresentam-se as previsões para o período entre julho de 2023 e junho de 2024, acompanhadas da avaliação da sua qualidade.

3.2- Descrição dos dados

Os dados analisados neste capítulo correspondem às temperaturas médias mensais da água do mar (em °C), registadas por boias oceanográficas localizadas em Nazaré, Faro e São Miguel (Portugal) e Praia (Cabo Verde), e obtidas a partir da plataforma *EMODnet* (*European Marine Observation and Data Network*), disponível publicamente através do *EMODnet Geoviewer* (<https://emodnet.ec.europa.eu/geoviewer/>). Para cada uma destas localidades, dispõem-se de séries temporais nas profundidades de 1 metro e 5 metros, perfazendo um total de oito séries. Pretendia-se incluir profundidades adicionais, à luz de outros estudos; contudo, não existiam dados disponíveis para além dos níveis mencionados. Embora houvesse medições em outras profundidades, a sua distribuição era irregular entre as diferentes localizações.

O registo temporal disponível estende-se de janeiro de 1960 a junho de 2024. Contudo, os primeiros oito anos (1960–1967) apresentam valores mensais repetidos de forma exata, o que sugere resultarem de imputação, por exemplo por ausência de medições reais. Por esta razão, a análise foi circunscrita ao período de janeiro de 1968 a junho de 2024, em que a variabilidade temporal exhibe consistência com padrões climatológicos reais, conforme verificado na análise exploratória.

3.3- Análise exploratória dos dados

A representação gráfica dos dados ao longo do tempo constitui uma etapa essencial na análise de séries temporais, permitindo uma compreensão preliminar da sua estrutura. Esta visualização inicial facilita a identificação de padrões relevantes, como tendências ascendentes ou descendentes, ciclos sazonais, alterações estruturais e valores atípicos. O reconhecimento destes comportamentos é crucial para compreender a dinâmica da série e orientar adequadamente as etapas subsequentes da análise e da modelação (Ehlers, 2009).

Neste contexto, procedeu-se à análise gráfica das séries em estudo.

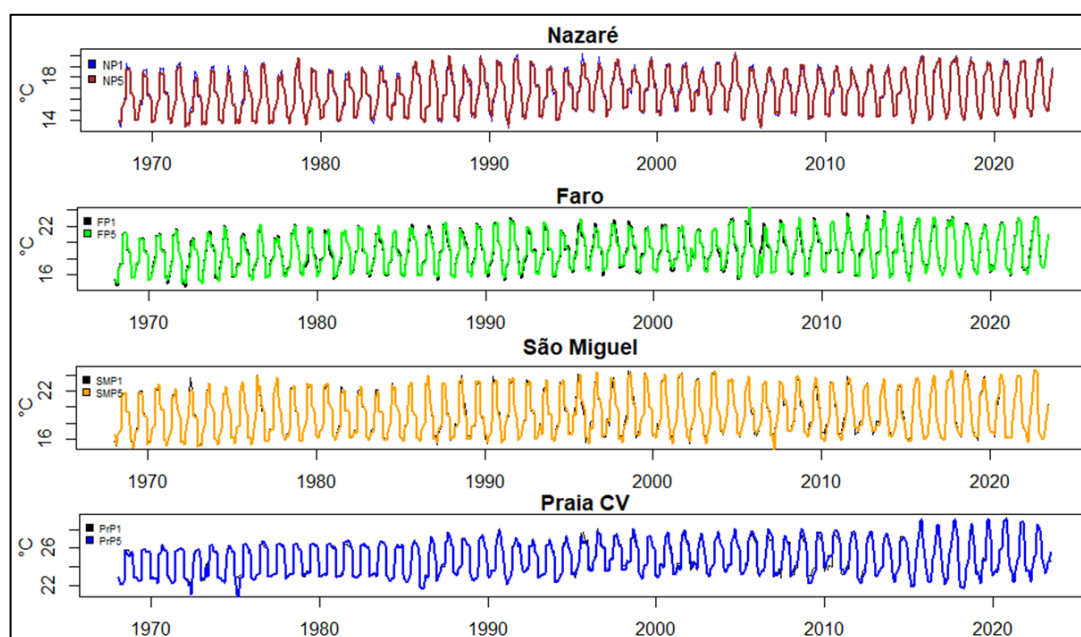


Figura 3.1: Representação gráfica das temperaturas mensais registadas em Nazaré, Faro, São Miguel e Praia.

Legenda das séries: NP1/NP5 – Nazaré, 1 m/5 m; FP1/FP5 – Faro, 1 m/5 m; SMP1/SMP5 – São Miguel, 1 m/5 m; PP1/PP5 – Praia, 1 m/5 m.

Fonte: Elaboração própria com dados disponibilizados no site: <https://emodnet.ec.europa.eu/geoviewer/>

A análise das séries temporais ilustradas na [Figura 3.1](#) mostra que há um padrão sazonal evidente em todas as localizações e profundidades, indicando que nenhuma das séries pode ser considerada estacionária. Observa-se também uma tendência de aumento progressivo na temperatura entre 1968 e meados de 2023, tanto a um (1) metro como a cinco (5) metros de profundidade, sendo mais acentuado à superfície, possivelmente devido à sua maior exposição à radiação solar direta, e por isso mais quente.

No quadro que se segue são apresentadas as estatísticas descritivas da temperatura média mensal por localidade e profundidade.

Tabela 3.1: Estatísticas descritivas das temperaturas mensais registadas em Nazaré, Faro, São Miguel e Praia.

Localidade	Profundidade	Média (°C)	Mediana (°C)	Mínimo (°C)	Máximo (°C)	Desvio Padrão (°C)	Coefficiente de Variação (%)
Nazaré	1 m	16,69	16,64	13,35	20,31	1,79	10,73
	5 m	16,65	16,62	13,44	20,07	1,75	10,51
Faro	1 m	18,91	18,85	14,57	23,92	2,21	11,69
	5 m	18,90	18,86	14,92	24,92	2,20	11,64
São Miguel	1 m	19,20	18,80	15,18	24,67	2,66	13,85
	5 m	19,15	18,80	14,79	24,64	2,68	13,96
Praia	1 m	24,92	24,74	21,12	29,31	1,84	7,38
	5 m	24,89	24,78	20,41	29,18	1,87	7,51

Fonte: Elaboração própria com dados disponibilizados no site: <https://emodnet.ec.europa.eu/geoviewer/>

Da análise da *Tabela 3.1* e dos gráficos da *Figura 3.1*, conclui-se ainda que, em termos médios, a diferença de temperatura entre as profundidades é pequena, como era expectável. As temperaturas mais elevadas foram registadas na Praia (Cabo Verde), enquanto as mais baixas ocorreram em Nazaré (Portugal), independentemente da profundidade observada.

De forma geral, observa-se que a mediana se aproxima bastante da média em todas as localidades, exceto em São Miguel, o que sugere uma distribuição aproximadamente simétrica das temperaturas nestas localidades. Em contraste, em São Miguel, a diferença mais acentuada entre estes dois valores pode indicar uma assimetria à direita, possivelmente causada por temperaturas atipicamente elevadas que influenciam a média.

Por fim, os coeficientes de variação indicam que São Miguel apresenta a maior instabilidade térmica relativa ao longo do tempo, seguida de Faro, uma vez que regista os valores mais elevados deste indicador. Em contraste, na Praia, apesar de apresentar as temperaturas médias mais altas, os coeficientes de variação são os mais baixos, evidenciando uma maior estabilidade térmica relativa, em ambas as profundidades.

Análise da taxa de variação das temperaturas por localidade e profundidade

Para avaliar a evolução da temperatura no período 1968-2023, utilizou-se o estimador do declive de Sen (Sen, 1968), aplicado às médias anuais, um método robusto que calcula a tendência a partir da mediana dos declives entre pares de observações. Esta abordagem é

especialmente útil porque não é influenciada por valores extremos (*outliers*), presentes nas séries em estudo.

Formalmente, considerando n observações ordenadas temporalmente (t_i, Y_i) , $i = 1, \dots, n$, onde t_i representa o ano da observação e Y_i a temperatura média anual, calcula-se para cada par (i, j) com $1 \leq i < j \leq n$ o declive:

$$D_{ij} = \frac{Y_j - Y_i}{t_j - t_i}.$$

O estimador do declive $\hat{\beta}$ é então definido como a mediana dos

$$N = \frac{n(n-1)}{2}$$

valores de D_{ij} . Conforme demonstrado por Sen (1968), $\hat{\beta}$ é um estimador não enviesado.

Teste de significância da tendência

Entre os métodos não paramétricos disponíveis, o teste de Mann-Kendall (Mann, 1945) é um dos mais utilizados para avaliar a presença de tendências em séries temporais, destacando-se pela sua robustez face a violações de normalidade. O teste avalia as hipóteses:

$$\begin{cases} H_0: \text{Não existe tendência na série (os dados são aleatoriamente ordenados)} \\ H_1: \text{Existe tendência na série (crescente ou decrescente)} \end{cases}.$$

Os resultados do teste são válidos se as observações forem independentes, pois a presença de autocorrelação serial tende a subestimar a variância real da estatística de teste S , dada por

$$Var(S) = \frac{n(n-1)(2n+5)}{18},$$

inflacionando a probabilidade de erro do Tipo I (rejeição indevida de H_0).

Caso os dados apresentem autocorrelação serial, recomenda-se recorrer ao teste de Mann-Kendall modificado proposto por Hamed e Rao (1998).

Este procedimento corrige a variância da estatística de teste S , multiplicando $Var(S)$ por

$$\frac{n}{n_S^*},$$

onde n_S^* corresponde ao número de observações corrigido pela autocorrelação da série. Esta correção assegura que os resultados do teste não sejam enviesados pela dependência temporal presente nas séries.

O fator de ajuste é estimado por:

$$\frac{n}{n_S^*} = 1 + \frac{2}{n(n-1)(n-2)} \times \sum_{i=1}^{n-1} (n-i)(n-i-1)(n-i-2)\rho_S(i),$$

onde n é o número real de observações, e $\rho_S(i)$ é a função de autocorrelação dos postos das observações.

Neste sentido, iniciou-se por verificar o pressuposto de independência mediante a aplicação do teste de Ljung-Box às séries anuais de temperatura. Os resultados encontram-se na tabela seguinte:

Tabela 3.2: Valores-p do teste de Ljung-Box aplicado às séries anuais de temperatura do mar (1968-2023).

Localidade	Profundidade	Valor- p do teste de Ljung-Box
Nazaré	1 m	< 0,0001
	5 m	< 0,0001
Faro	1 m	< 0,0001
	5 m	< 0,0001
S. Miguel	1 m	< 0,0001
	5 m	< 0,0001
Praia	1 m	< 0,0001
	5 m	< 0,0001

Fonte: Elaboração própria.

Ao nível de significância de 5%, rejeita-se a hipótese nula de não autocorrelação em todas as séries analisadas ($valor - p < 0,0001$), evidenciando a presença de dependência temporal estatisticamente significativa entre observações consecutivas.

Assim, a significância estatística do declive de Sen foi avaliada mediante o teste de Mann-Kendall modificado. Os resultados obtidos encontram-se apresentados na tabela seguinte.

*Tabela 3.3: Aumento médio da temperatura do mar (1968-2023) estimado pelo método de Sen, com valores-
p do teste de Mann-Kendall modificado.*

Localidade	Profundidade	Aumento Médio Anual (°C/ano)	Aumento Médio por Década (°C/década)	Valor- <i>p</i> (MK modificado)
Nazaré	1 m	0,01956	0,19555	< 0,0001
	5 m	0,02090	0,20904	< 0,0001
Faro	1 m	0,02336	0,23355	< 0,0001
	5 m	0,02342	0,23419	< 0,0001
S. Miguel	1 m	0,02035	0,20351	< 0,0001
	5 m	0,02038	0,20383	< 0,0001
Praia	1 m	0,02015	0,20149	< 0,0003
	5 m	0,02169	0,21698	< 0,0001

Fonte: Elaboração própria.

Os resultados apresentados na [Tabela 3.3](#), evidenciam tendências de aquecimento estatisticamente significativas (*valor – p* ≤ 0,0003) em todas as séries temporais analisadas.

Os aumentos médios anuais oscilaram entre 0,01956 °C/ano (Nazaré, 1m) e 0,02342 °C/ano (Faro, 5m), correspondendo a um aquecimento médio de aproximadamente 0,20-0,23 °C por década. Estas estimativas são de magnitude comparável às reportadas na literatura recente (ex.: Shaltout, 2019; Saes et al., 2022; European Space Agency, 2025).

Para confirmar o padrão sazonal identificado na análise gráfica das séries temporais, iniciou-se pela observação dos *boxplots* representados na [Figura 3.2](#).

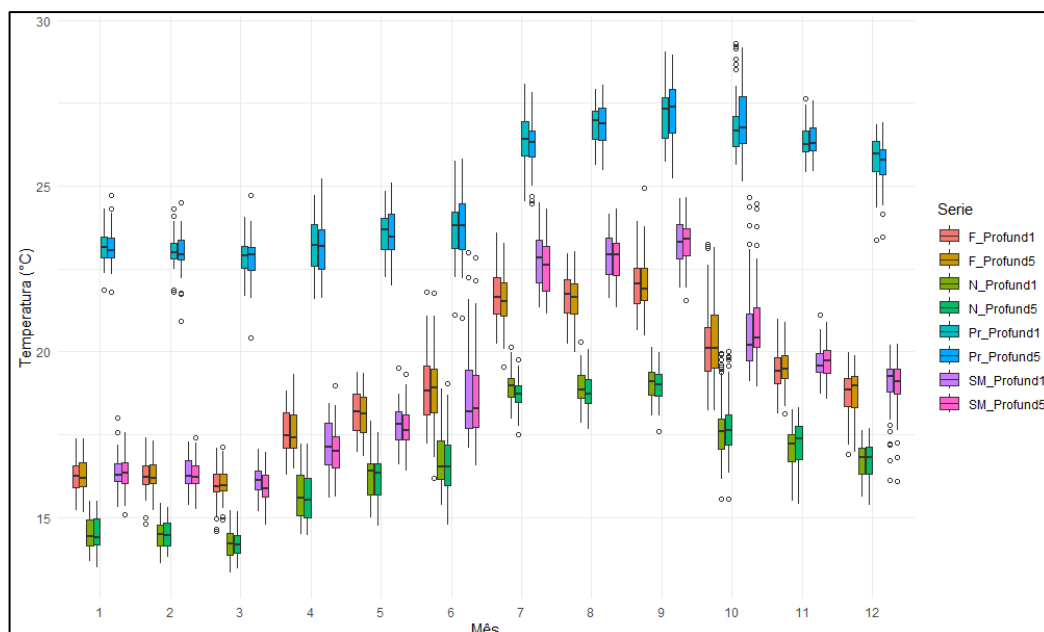


Figura 3.2: Boxplots das Séries de Temperaturas Mensais por Região e Profundidade.

Legenda das séries: N_Profund1 e N_Profund5 (Nazaré, 1 m e 5 m); F_Profund1 e F_Profund5 (Faro, 1 m e 5 m); SM_Profund1 e SM_Profund5 (São Miguel, 1 m e 5 m); Pr_Profund1 e Pr_Profund5 (Praia, 1 m e 5 m).

Fonte: Elaboração própria com dados disponibilizados no site: <https://emodnet.ec.europa.eu/geoviewer/>

A análise dos *boxplots* confirma que todas as séries exibem um padrão sazonal bem definido, com temperaturas mais baixas nos meses de inverno e temperaturas mais elevadas nos meses de verão, conforme esperado.

De modo a aprofundar as análises anteriormente realizadas, procedeu-se à representação gráfica e a análise visual das funções de autocorrelação (ACF) e autocorrelação parcial (PACF) das séries temporais de temperatura média mensal relativas às quatro regiões e profundidades em estudo (*Figura 3.3*). Verifica-se que as referidas séries não se comportam como ruído branco, ou seja, não apresentam um comportamento puramente aleatório, uma vez que os coeficientes de autocorrelação não são estatisticamente nulos (situam-se fora do intervalo de confiança de $\pm 0,076$, para o nível de significância de 5%), revelando, assim, uma dependência temporal significativa entre as observações em todos os desfasamentos k .

A presença de autocorrelações significativas na ACF, com valores máximos em desfasagens múltiplas de 12 meses (12, 24, 36, 48) e valores mínimos em desfasagens múltiplas de 6 meses (6, 18, 30, 42), valida estatisticamente a sazonalidade anual identificada na análise gráfica das *Figuras 3.1* e *3.2*.

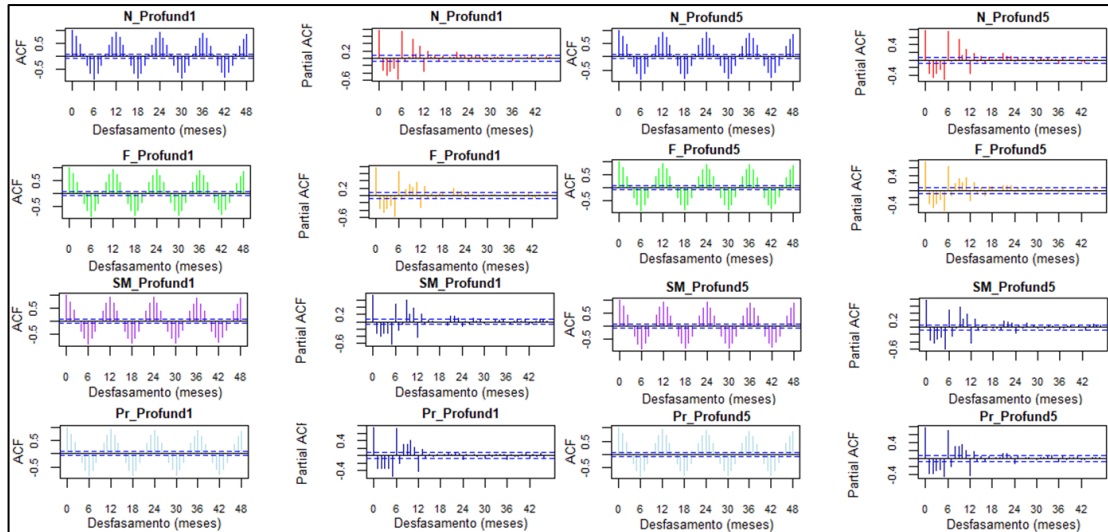


Figura 3.3: Representações gráficas da ACF e PACF.

Legenda das séries: *N_Profund1* e *N_Profund5* (Nazaré, 1 m e 5 m); *F_Profund1* e *F_Profund5* (Faro, 1 m e 5 m); *SM_Profund1* e *SM_Profund5* (São Miguel, 1 m e 5 m); *Pr_Profund1* e *Pr_Profund5* (Praia, 1 m e 5 m).

Fonte: Elaboração própria.

3.4- Avaliação de estacionaridade

Dickey-Fuller aumentado (ADF)

Para confirmar a possível não estacionaridade sugerida pela análise gráfica ([Figura 3.1](#)) e pelas funções de autocorrelação ([Figura 3.3](#)), foi aplicado o teste de **ADF** às séries temporais. Este teste avalia a hipótese nula de que a série possui uma raiz unitária, ou seja, que não é estacionária.

Considerando que as séries apresentam uma tendência ao longo do tempo, o teste foi realizado com o parâmetro *type = "trend"*. O número de desfasamentos foi selecionado automaticamente pelo R, com base no Critério de Informação de Akaike (*selectlags = "AIC"*), o que permite identificar o número ótimo de desfasamentos a incluir no modelo (Hyndman e Athanasopoulos, 2021). O seguinte comando foi utilizado como referência: `ur.df(x, type = "trend", selectlags = "AIC")`, em que *x* representa a série de temperatura média mensal.

Resultado dos testes:

Tabela 3.4: Estatísticas dos testes ADF aplicados às séries de temperatura.

Séries Temperatura Mensais	Valor Observado da Estatística de Teste ADF	Valor Crítico
Nazaré (1 m)	-12,6475	-3,41
Nazaré (5 m)	-12,7158	-3,41
Faro (1 m)	-12,9179	-3,41
Faro (5 m)	-13,0631	-3,41
São Miguel (1 m)	-13,0352	-3,41
São Miguel (5 m)	-13,1223	-3,41
Praia (1 m)	-12,9745	-3,41
Praia (5 m)	-13,3375	-3,41

Fonte: Elaboração própria.

Conforme se observa na [Tabela 3.4](#), os valores da estatística de teste são inferiores aos valores críticos ao nível de significância de 5%, pelo que se rejeita a hipótese nula de presença de raiz unitária. Assim, conclui-se que todas as séries analisadas podem ser consideradas estacionárias em torno da média.

Teste de Phillips-Perron (PP)

Com o objetivo de avaliar a robustez dos resultados obtidos através do teste de **ADF**, foram também realizados os testes de **PP**, os quais complementam a análise da estacionaridade.

Para tal, foi utilizado o seguinte comando do R, que implementa o teste PP para detecção de tendência: `ur.pp(x, type = "Z-tau", model = "trend", lags = "long")`.

Resultados dos testes:

Tabela 3.5: Estatísticas dos testes de PP aplicados às séries de temperatura.

Séries Temperatura Mensais	Valor Observado da Estatística de Teste	Valor Crítico
Nazaré (1 m)	-6,9662	-3,4187
Nazaré (5 m)	-6,8539	-3,4187
Faro (1 m)	-6,8496	-3,4187
Faro (5 m)	-6,7695	-3,4187
São Miguel (1 m)	-7,2088	-3,4187
São Miguel (5 m)	-6,9108	-3,4187
Praia (1 m)	-7,2215	-3,4187
Praia (5 m)	-6,8511	-3,4187

Fonte: Elaboração própria.

À semelhança dos resultados obtidos com o teste de **ADF**, os testes de **PP**, apresentados na [Tabela 3.5](#), confirmam as mesmas decisões ao mesmo nível de significância.

Teste de Kwiatkowski–Phillips–Schmidt–Shin (KPSS)

Por fim, procedeu-se à aplicação do teste KPSS de Kwiatkowski et al, (1992) às séries em estudo. Atendendo à presença de uma tendência determinística, recorreu-se ao *argumento type = "tau"*, adequado para testar estacionaridade em torno de uma tendência. Para lidar com possível autocorrelação de longo alcance nos resíduos, utilizou-se ainda a opção *lags = "long"*, que implementa uma correção mais conservadora e reduz a probabilidade de rejeição indevida da hipótese nula de estacionaridade (Hyndman e Athanasopoulos, 2021).

A execução do teste foi realizada através do comando *ur.kpss(x, type = "tau", lags = "long")*, sendo os respetivos resultados apresentados na [Tabela 3.6](#).

Tabela 3.6: Estatísticas dos testes de KPSS aplicados às séries de temperatura.

Série de Temperatura média Mensal	Valor Observado da Estatística de Teste	Valor Crítico
Nazaré (1 m)	0,1382	0,146
Nazaré (5 m)	0,1883	0,146
Faro (1 m)	0,1126	0,146
Faro (5 m)	0,1326	0,146
São Miguel (1 m)	0,0812	0,146
São Miguel (5 m)	0,0826	0,146
Praia (1 m)	0,2021	0,146
Praia (5 m)	0,183	0,146

Fonte: Elaboração própria.

Os resultados indicam que, para a maioria das séries, Nazaré (1 m), Faro (ambas as profundidades) e São Miguel (ambas as profundidades), os valores observados da estatística de teste são inferiores ao valor crítico, pelo que não se rejeita a hipótese nula de estacionaridade. Contudo, nas séries de Praia (ambas as profundidades) e de Nazaré a 5 m, os valores observados da estatística de teste são superiores ao valor crítico, levando à rejeição da hipótese nula e sugerindo não estacionaridade.

Em síntese, embora os testes ADF e PP sugiram estacionaridade em torno de uma tendência determinística, os resultados do teste KPSS, que rejeitam a hipótese nula de estacionaridade para as séries da Praia e de Nazaré (5 m), indicam a presença de não estacionaridade estocástica. Perante esta divergência, e tendo em conta a evidência gráfica de tendência e

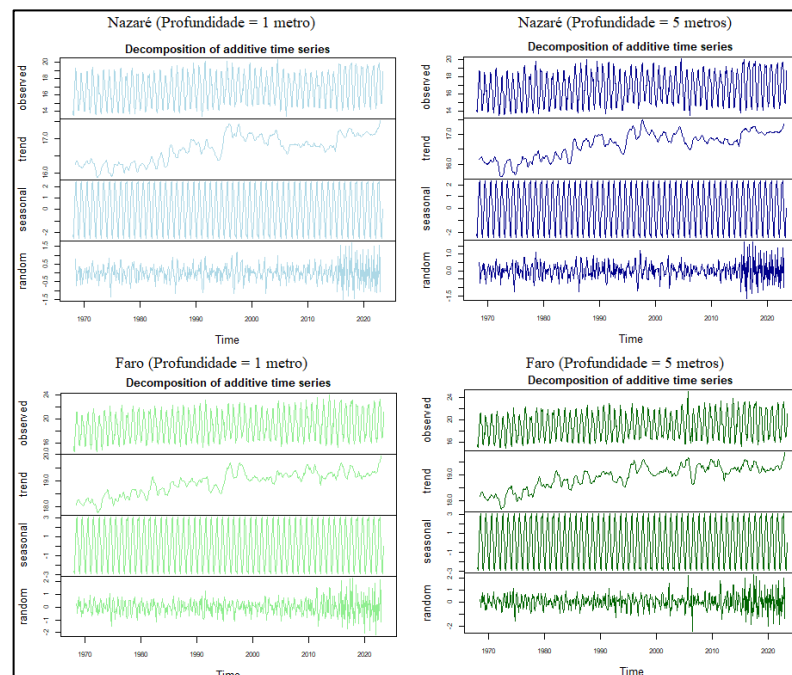
sazonalidade, adotou-se uma abordagem conservadora. Para fins de modelação, todas as séries serão tratadas como não estacionárias.

3.5- Modelação das séries por região e profundidade

Com o objetivo de identificar o modelo que melhor descreve o comportamento das séries de temperatura média mensal nas diferentes regiões e profundidades, foram ajustadas e comparadas quatro abordagens: (i) decomposição clássica, (ii) decomposição STL (*Seasonal and Trend decomposition using Loess*), (iii) modelo de alisamento exponencial de Holt-Winters e (iv) modelos ARIMA/SARIMA. Esta sequência permitiu avaliar progressivamente a capacidade de cada abordagem em representar a tendência, a sazonalidade e a componente residual das séries, com base em critérios como o AICc, a análise gráfica e estatística dos resíduos (ACF/PACF e teste de Ljung-Box).

3.5.1-Modelos de decomposição clássica

Foram ajustados modelos de decomposição clássica do tipo aditivo a todas as séries em estudos, pois ao analisar a [Figura 3.1](#), verifica-se que a variação da amplitude sazonal das temperaturas registadas em ambas as profundidades parece manter-se aproximadamente constante ao longo do tempo. A implementação foi realizada através do comando: `decompose(x)`.



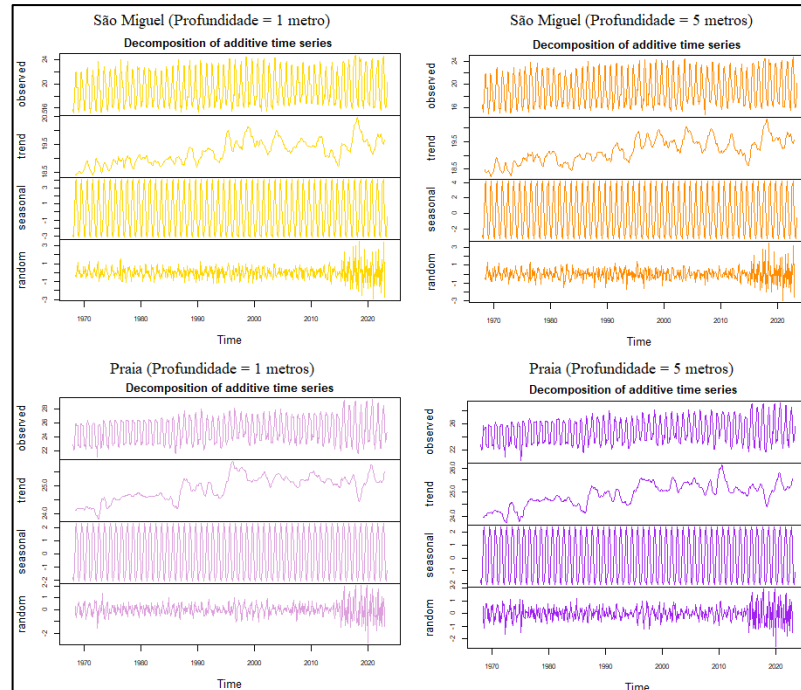


Figura 3.4: Decomposição Clássica do tipo aditivo das séries temporais das temperaturas médias mensais em cada local e profundidade.

Fonte: Elaboração própria.

A decomposição das séries temporais, apresentada na [Figura 3.4](#), permitiu identificar três principais componentes: a **tendência**, a **sazonalidade** e a **componente residual**. A componente de tendência extraída confirmou o padrão de aumento progressivo descrito na Secção 3.3.

A componente dos resíduos mostra que, embora inicialmente estes apresentem um comportamento aparentemente aleatório em torno de zero, nos últimos anos observa-se um aumento significativo da sua variabilidade.

Para confirmar a análise anterior, recorreu-se à representação gráfica das funções ACF e PACF ([Figura 3.5](#)), aplicadas aos resíduos dos modelos de decomposição clássica. Observam-se autocorrelações significativas em vários desfasamentos, nomeadamente nos desfasamentos 1, 6 e 12, o que indica que os resíduos não são completamente aleatórios.

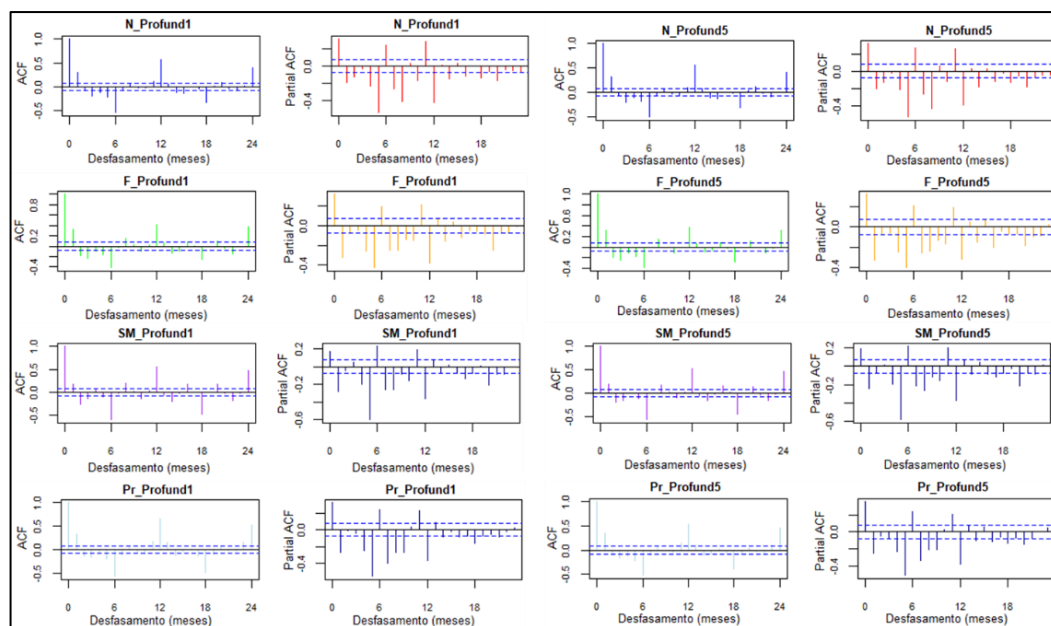


Figura 3.5: ACF e PACF da Componente Residual da Decomposição Clássica.

Legenda das séries: N_Profund1 e N_Profund5 (Nazaré, 1 m e 5 m); F_Profund1 e F_Profund5 (Faro, 1 m e 5 m); SM_Profund1 e SM_Profund5 (São Miguel, 1 m e 5 m); Pr_Profund1 e Pr_Profund5 (Praia, 1 m e 5 m).

Fonte: Elaboração própria.

Teste de Ljung-Box Aplicado aos Resíduos da Decomposição Clássica

Aplicou-se o teste de Ljung-Box aos resíduos da decomposição clássica, considerando um desfasamento $k = 24$ (correspondente a dois anos completos de observações).

Tabela 3.7: Valores-p do Teste de Ljung-Box para os Resíduos da Decomposição Clássica.

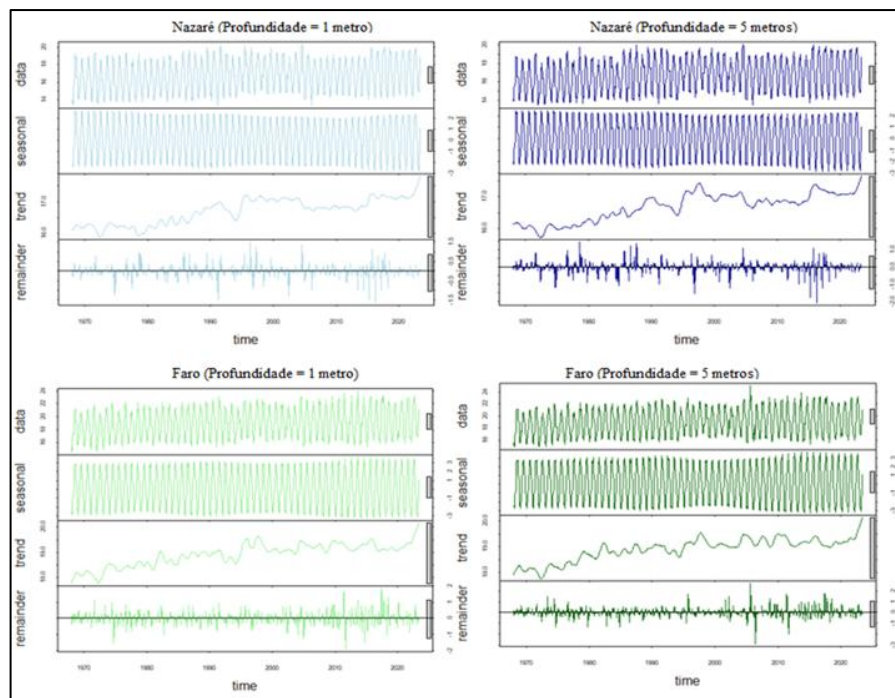
Série de Temperatura média Mensal	Valor-p
Nazaré (1 m)	< 0,001
Nazaré (5 m)	< 0,001
Faro (1 m)	< 0,001
Faro (5 m)	< 0,001
São Miguel (1 m)	< 0,001
São Miguel (5 m)	< 0,001
Praia (1 m)	< 0,001
Praia (5 m)	< 0,001

Fonte: Elaboração própria.

Em todos os casos, os valores- p são próximos de zero ($p < 0,001$), levando à rejeição da hipótese nula de ausência de autocorrelação. Conclui-se, assim, que os resíduos apresentam dependência temporal significativa, indicando que o modelo não capturou integralmente a estrutura dinâmica das séries.

3.5.2-Decomposição da tendência e sazonalidade utilizando Loess (STL)

A decomposição STL foi aplicada às séries temporais através do comando `stl(x, s, window = 13, robust = TRUE)`. A parametrização `s.window = 13`, segundo Hyndman e Athanasopoulos (2021), permite que a componente sazonal varie ao longo do tempo, ajustando-se a oscilações tanto na forma como na amplitude do padrão sazonal, o que torna o modelo mais flexível face a alterações estruturais na série. Considerando que as séries em estudo dizem respeito a temperaturas do mar, variáveis sujeitas a fenómenos climáticos extremos e a eventuais erros de medição, considerou-se apropriado ativar a opção robusta (`robust = TRUE`). Esta decisão é corroborada empiricamente pela presença recorrente de valores atípicos ao longo de praticamente todos os meses e séries analisadas, conforme evidenciado na [Figura 3.2](#). De acordo com os mesmos autores, esta configuração reduz a influência de observações atípicas nas estimativas das componentes de tendência e sazonalidade. As observações extremas passam a ser absorvidas pela componente residual, o que contribui para preservar a estrutura da decomposição e facilita a sua identificação e análise separada.



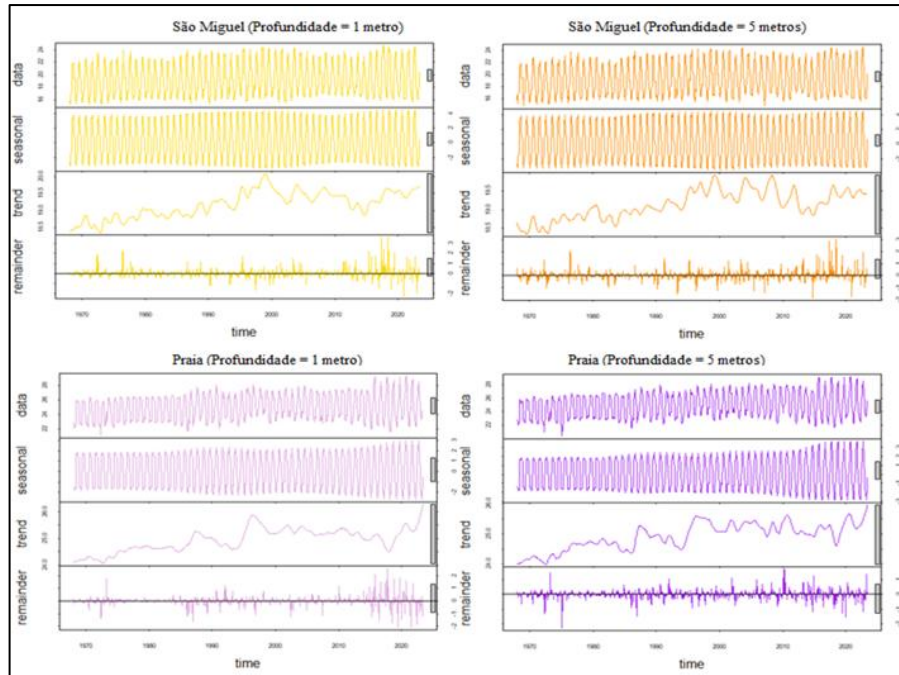


Figura 3.6: Decomposição STL para a séries temporais em cada local e profundidade.

Fonte: Elaboração própria.

A decomposição STL apresentada na [Figura 3.6](#) resultou numa separação mais clara das três componentes da série: a **tendência**, a **sazonalidade** e a **componente residual**. Tal como na decomposição clássica, a componente de tendência corroborou os resultados da análise exploratória (Secção 3.3).

Constata-se, ainda que de forma menos acentuada do que na decomposição clássica, um aumento da variabilidade dos resíduos da decomposição STL, particularmente nas temperaturas registadas em Faro a um metro de profundidade, bem como nas séries de São Miguel e Praia em ambas as profundidades. Assim, apesar da aparente melhoria, a decomposição STL ainda não foi totalmente eficaz na captação das flutuações mais recentes nestas regiões.

Os resíduos da decomposição STL apresentaram, de forma geral, menor autocorrelação e menor variabilidade ao longo do tempo do que os obtidos com a decomposição clássica, aproximando-se mais de um comportamento típico de ruído branco. Conforme ilustrado na [Figura 3.7](#), os gráficos ACF e PACF dos resíduos da STL exibem picos menos acentuados e menor persistência ao longo dos desfasamentos, o que indica que as componentes de tendência e sazonalidade conseguiram captar melhor a estrutura dos dados.

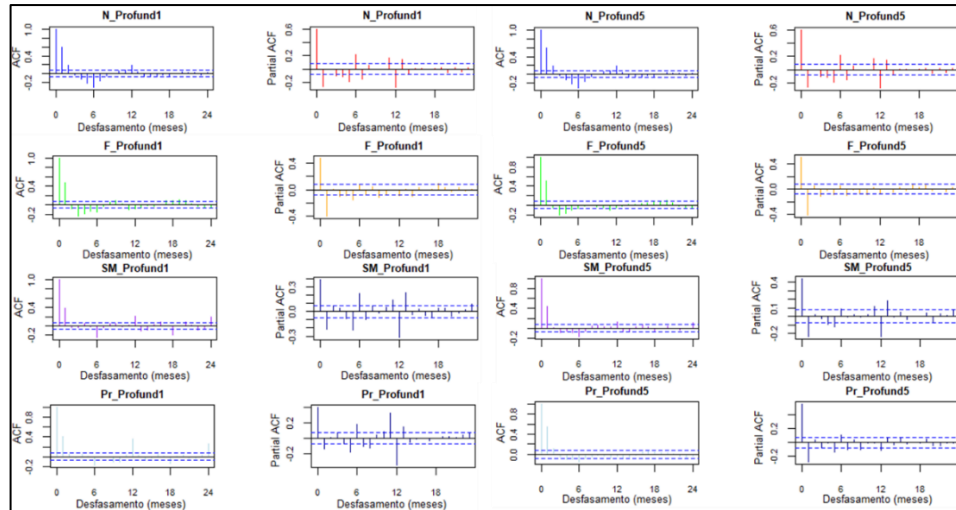


Figura 3.7: ACF e PACF da Componente residual da Decomposição STL.

Legenda das séries: N_Profund1 e N_Profund5 (Nazaré, 1 m e 5 m); F_Profund1 e F_Profund5 (Faro, 1 m e 5 m); SM_Profund1 e SM_Profund5 (São Miguel, 1 m e 5 m); Pr_Profund1 e Pr_Profund5 (Praia, 1 m e 5 m).

Fonte: Elaboração própria.

Teste de Ljung-Box Aplicado aos Resíduos da Decomposição STL

Para uma avaliação mais objetiva, procedeu-se à aplicação do teste de Ljung-Box ($k = 24$) aos resíduos das diferentes séries temporais.

Tabela 3.8: Valores-p do Teste de Ljung-Box para os Resíduos da Decomposição STL.

Série de Temperatura média Mensal	Valor-p
Nazaré (1 m)	< 0,001
Nazaré (5 m)	< 0,001
Faro (1 m)	< 0,001
Faro (5 m)	< 0,001
São Miguel (1 m)	< 0,001
São Miguel (5 m)	< 0,001
Praia (1 m)	< 0,001
Praia (5 m)	< 0,001

Fonte: Elaboração própria.

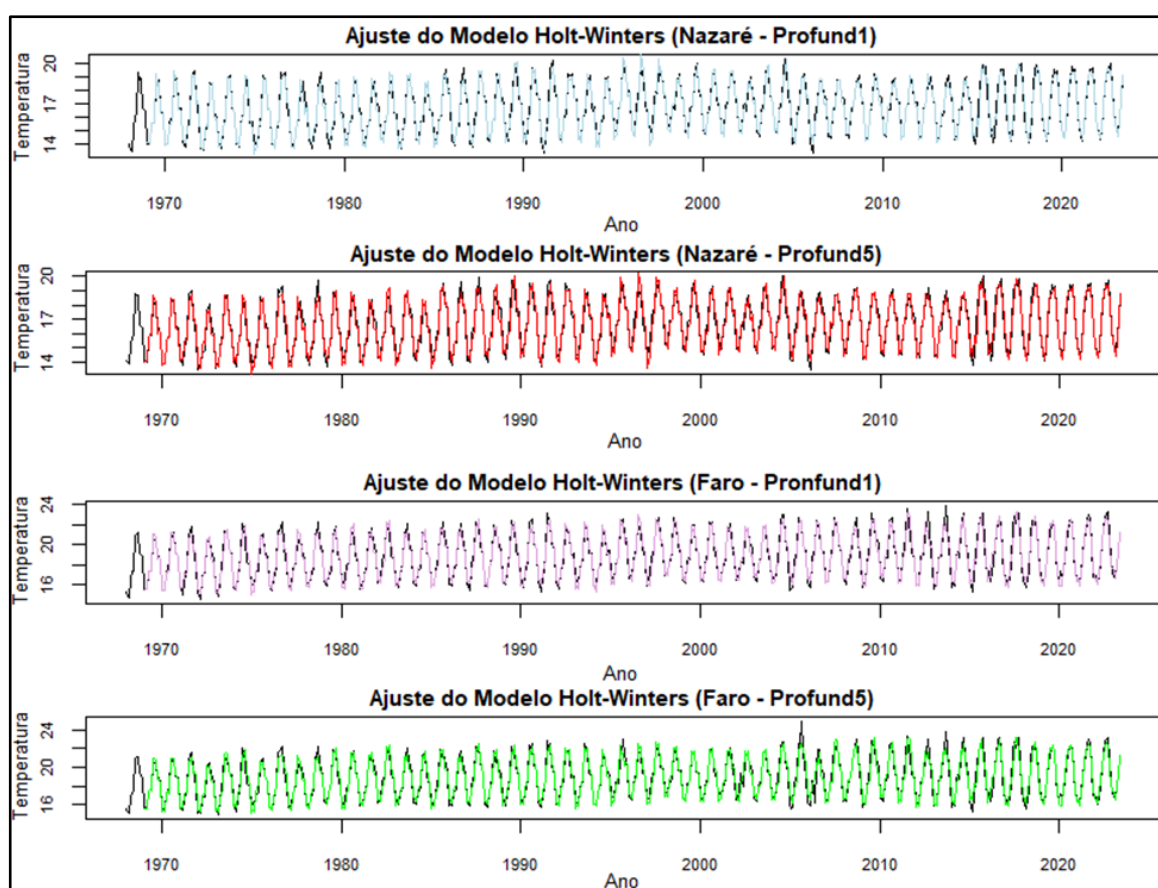
Os resíduos da decomposição *STL* são autocorrelacionados ($p < 0,001$), o que confirma que este método também não consegue captar totalmente a estrutura temporal presente nas séries.

Para aprofundar o estudo e a perspetiva, são explorados modelos adicionais, como o modelo de suavização exponencial de **Holt-Winters** e o **ARIMA** com componentes sazonais (**SARIMA**), com o objetivo de captar de forma mais completa os padrões subjacentes nas séries temporais.

3.5.3-Modelos de alisamento exponencial de Holt-Winters

Foram ajustados modelos de Holt-Winters às séries de temperatura mensal registadas em todas as localizações e em ambas as profundidades. Tal como justificado na Secção 3.5.1, optou-se pelo modelo aditivo.

Na [Figura 3.8](#) apresentam-se os ajustamentos obtidos, em que se comparam os valores observados com as trajetórias estimadas pelos respetivos modelos.



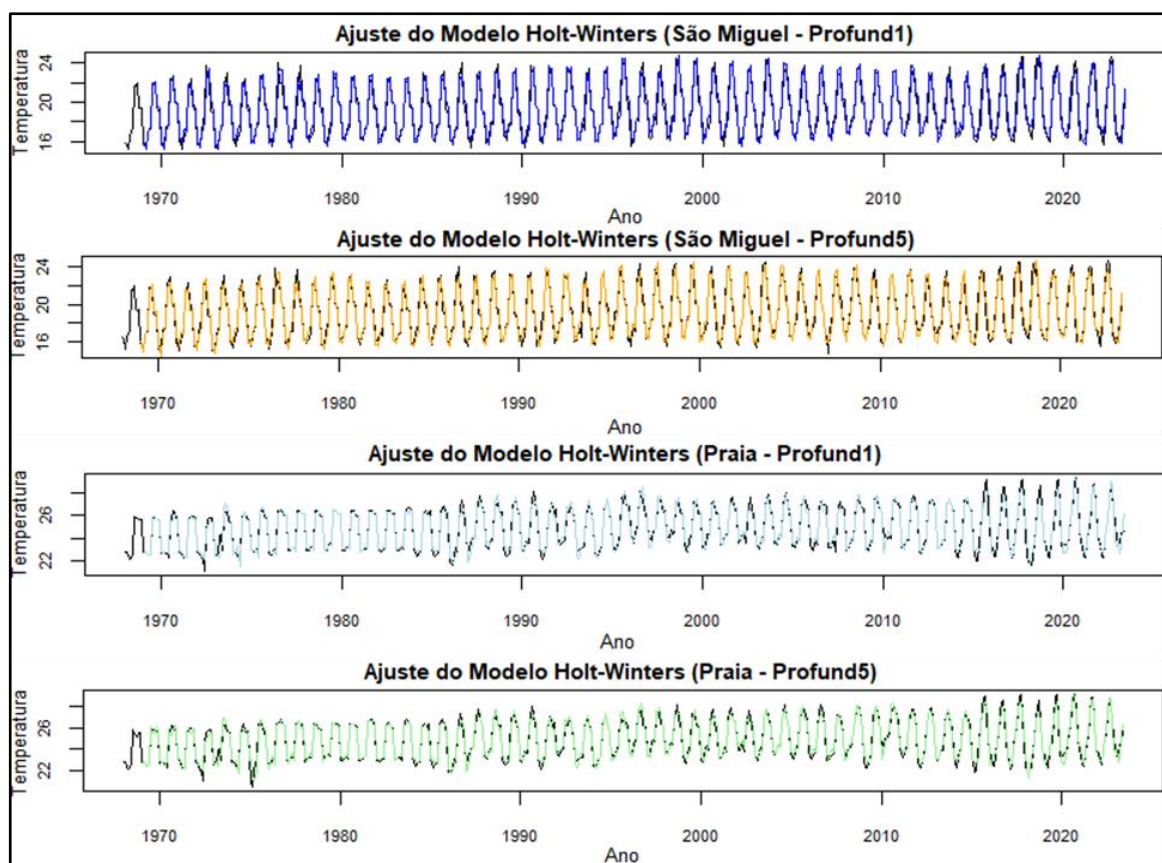


Figura 3.8: Séries de temperaturas observadas (linhas pretas) e respectivos ajustes pelo modelo de Holt-Winters (linhas coloridas) em cada local e profundidade.

Fonte: Elaboração própria.

A análise visual dos gráficos apresentados na [Figura 3.8](#) revela um ajustamento satisfatório dos modelos de Holt-Winters às séries temporais de temperatura média mensal, em ambas as profundidades. O modelo parece captar de forma adequada tanto a tendência de longo prazo quanto a componente sazonal, evidenciando-se um bom alinhamento entre os valores observados e os ajustados ao longo do tempo.

Avaliação quantitativa da qualidade dos modelos

Foram construídos os gráficos da ACF e da PACF dos resíduos, apresentados na [Figura 3.9](#), que permitem avaliar a presença de autocorrelação e identificar eventuais padrões não captados pelos modelos.

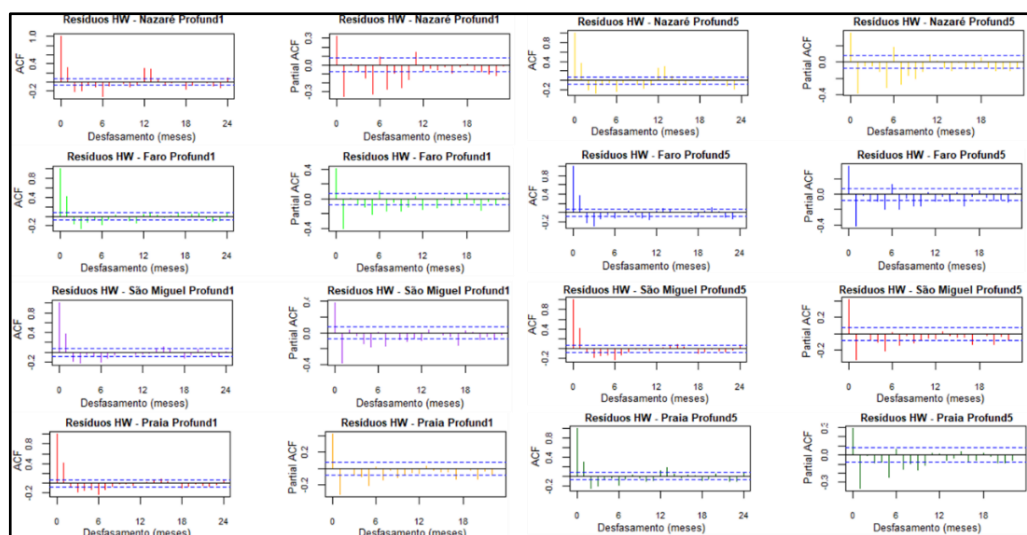


Figura 3.9: ACF e PACF dos resíduos dos modelos Holt-Winters ajustados.

Fonte: Elaboração própria.

A análise da [Figura 3.9](#) revela uma forte autocorrelação nos primeiros desfasamentos (em especial no *lag* 1), comum a todos os modelos Holt-Winters ajustados, independentemente da localização e da profundidade considerada. Este padrão consistente sugere que os modelos, apesar de capturarem parte da tendência e da sazonalidade das séries, não conseguem captar integralmente toda a estrutura dos dados.

Teste de Ljung-Box Aplicado aos Resíduos dos modelos Holt-Winters

Os resultados do teste aplicado aos resíduos, considerando um desfasamento $k = 24$, encontram-se na tabela que se segue.

Tabela 3.9: Valores-*p* do Teste de Ljung-Box para os Resíduos dos modelos Holt-Winters.

Série de Temperatura média Mensal	Valor- <i>p</i>
Nazaré (1 m)	$< 2,2 \times 10^{-16}$
Nazaré (5 m)	$< 2,2 \times 10^{-16}$
Faro (1 m)	$< 2,2 \times 10^{-16}$
Faro (5 m)	$< 2,2 \times 10^{-16}$
São Miguel (1 m)	$< 2,2 \times 10^{-16}$
São Miguel (5 m)	$< 2,2 \times 10^{-16}$
Praia (1 m)	$< 2,2 \times 10^{-16}$
Praia (5 m)	$< 2,2 \times 10^{-16}$

Fonte: Elaboração própria.

Para todas as séries, os valores-*p* são extremamente próximos de zero, levando à rejeição da hipótese nula de ausência de autocorrelação residual, confirmando as conclusões obtidas a partir da análise visual das funções ACF e PACF ([Figura 3.9](#)).

3.5.4-Modelos ARIMA/SARIMA

Tendo em conta os resultados da análise de estacionaridade realizada na secção 3.4, optou-se por iniciar a modelação com a função ***auto.arima()*** do pacote *forecast* em R. Esta abordagem permite identificar automaticamente a estrutura do modelo *ARIMA/SARIMA* mais adequada com base em critérios de informação como o AIC e o BIC, ajustando simultaneamente os parâmetros de tendência, de sazonalidade e de ruído (Hyndman & Khandakar, 2008).

A função *auto.arima()* avalia automaticamente a necessidade de diferenciação, identificando tanto o tipo (regular ou sazonal) como o número de diferenciações requeridas para garantir a estacionaridade da série.

Ajustou-se o modelo SARIMA automático a todas as séries temporais analisadas, tendo identificados os seguintes modelos, conforme a tabela seguinte:

Tabela 3.10: Modelos SARIMA automáticos ajustados às séries.

Séries Temperatura Média Mensal Por Localidade e Profundidade	Modelo Auto ARIMA Ajustado
Nazaré (1 m)	SARIMA(3,0,3)(1,1,2)[12]
Nazaré (5 m)	SARIMA(4,0,5)(0,1,2)[12]
Faro (1 m)	SARIMA(5,0,3)(0,1,1)[12]
Faro (5 m)	SARIMA(1,0,1)(0,1,1)[12]
São Miguel (1 m)	SARIMA(0,0,2)(0,1,1)[12]
São Miguel (5 m)	SARIMA(1,0,1)(0,1,1)[12]
Praia (1 m)	SARIMA(5,0,3)(0,1,1)[12]
Praia (5 m)	SARIMA(2,0,2)(0,1,1)[12]

Fonte: Elaboração própria.

Todos os modelos automáticos identificados apresentam $D = 1$ e $d = 0$, sugerindo a necessidade de diferenciação sazonal e ausência de tendência estocástica. Note-se que isto não invalida a tendência determinística de aumento da temperatura observada na análise exploratória; esta é capturada através de uma constante, não exigindo diferenciação ($d = 0$) (Hyndman & Athanasopoulos, 2021). Este resultado reforça a constatação da presença de um padrão sazonal em todas as séries, tal como identificado nas secções anteriores. Além disso, a necessidade de diferenciação sazonal indica a existência de não estacionaridade sazonal.

Análise da qualidade dos modelos ajustados

Para a análise dos modelos identificados automaticamente, recorreu-se aos gráficos da ACF e da PACF dos resíduos da [Figura 3.10](#).

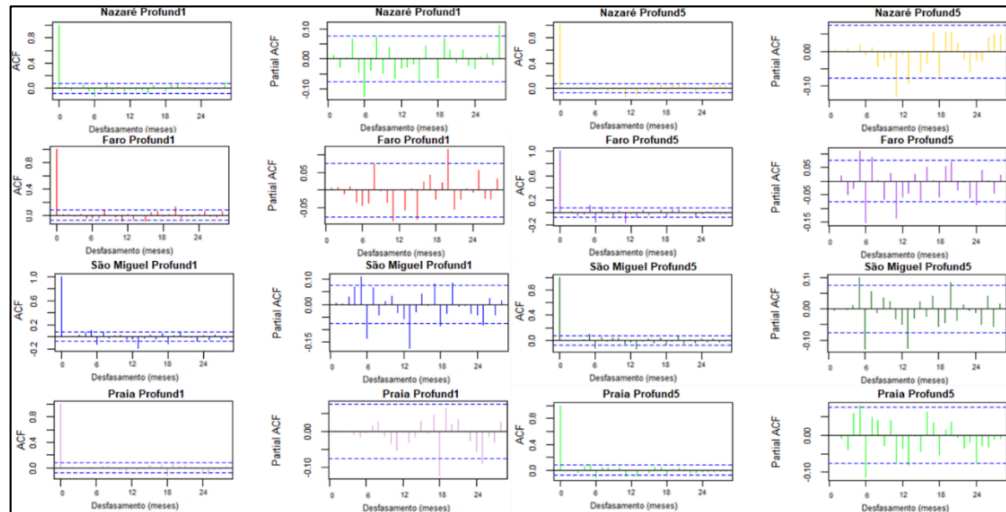


Figura 3.10: ACF e PACF dos resíduos dos modelos ARIMA automático ajustados.

Fonte: Elaboração própria.

A análise dos resíduos dos modelos SARIMA automáticos, apresentada na [Figura 3.10](#), revela uma melhoria substancial em relação aos modelos Holt-Winters. Ao contrário dos resíduos dos modelos de suavização exponencial, que exibiam forte autocorrelação no desfasamento 1, os resíduos dos modelos SARIMA apresentam-se predominantemente dentro dos limites de confiança da ACF e PACF, indicando que a estrutura temporal foi adequadamente capturada. Embora se observem alguns picos isolados fora dos limites, estes não seguem um padrão sistemático e são de magnitude reduzida, sugerindo que os modelos ajustados são estatisticamente adequados.

Teste de Ljung-Box Aplicado aos Resíduos dos Modelos SARIMA Automáticos

Tabela 3.11: Valor-p do teste Ljung-Box aplicado aos resíduos dos modelos SARIMA automáticos.

Séries Temperaturas Médias Mensais Por Localidade e Profundidade	Valor-p Teste Ljung-Box Aplicados Resíduos Modelos ARIMA Ajustados
Nazaré (1 m)	0,001191
Nazaré (5 m)	0,0006429
Faro (1 m)	0,002095
Faro (5 m)	3,682e-07
São Miguel (1 m)	6,258e-08
São Miguel (5 m)	0,0004659
Praia (1 m)	0,09395
Praia (5 m)	0,005219

Fonte: Elaboração própria.

Os resultados da aplicação do teste indicam que os resíduos são autocorrelacionados ($\text{valor} - p < 0,05$) em sete das oito séries, evidenciando que os modelos automáticos não captaram integralmente a estrutura de dependência temporal. A única exceção foi a série da Praia a 1 metro de profundidade ($\text{valor} - p = 0,0940 > 0,05$), cujos resíduos podem ser considerados aproximadamente ruído branco.

Tendo em conta esta evidência, proceder-se-á ao refinamento manual dos modelos cujos resíduos apresentam autocorrelação. Esse refinamento será orientado pela análise das funções de autocorrelação (ACF) e autocorrelação parcial (PACF) dos resíduos, com o objetivo de especificar modelos que incorporem termos adicionais (sazonais ou não sazonais) capazes de eliminar a autocorrelação remanescente.

Análise de estacionaridade após diferenciação sazonal de primeira ordem sugerido pelos modelos SARIMA automáticos

A tabela seguinte apresenta os resultados dos testes de estacionaridade ADF e KPSS.

Tabela 3.12: Resultados dos testes ADF e KPSS para as séries diferenciadas sazonalmente.

Série diferenciada	Profundidade	ADF (valor-p)	KPSS (valor-p)
Nazaré	1 m	< 0,01	> 0,10
	5 m	< 0,01	> 0,10
Faro	1 m	< 0,01	> 0,10
	5 m	< 0,01	> 0,10
São Miguel	1 m	< 0,01	> 0,10
	5 m	< 0,01	> 0,10
Praia	1 m	< 0,01	> 0,10
	5 m	< 0,01	> 0,10

Fonte: Elaboração própria.

Os resultados apresentados na [Tabela 3.12](#) confirmam a estacionaridade das séries após diferenciação sazonal de ordem 1, evidenciando convergência entre os testes de raiz unitária aplicados:

- Teste ADF: Os valores- p inferiores a 0,01 permitem rejeitar a hipótese nula de presença de raiz unitária (H_0), ao nível de significância de 5%, indicando que as séries diferenciadas são estacionárias.
- Teste KPSS: Os valores- p superiores a 0,10 implicam a não rejeição da hipótese nula de estacionariedade (H_0), corroborando a conclusão anterior.

Ajuste manual de modelos às séries

O ajuste manual foi realizado para todas as séries, exceto para a Praia (1 m), cujo modelo automático satisfaz os pressupostos estatísticos. O conjunto completo de modelos testados manualmente encontra-se detalhado no Anexo I.

Modelos selecionados para cada série

Os modelos finais foram selecionados com base na hierarquia metodológica proposta por Hyndman & Athanasopoulos (2021): prioridade à aleatoriedade dos resíduos (*valor* – $p > 0,05$ no teste de Ljung-Box com $k = 24$ defasagens), seguida da minimização de AICc e BIC. Na série de temperatura mensal registada em São Miguel, a um metro de profundidade, em que nenhum dos modelos ajustados produziu resíduos totalmente ruído branco, foi escolhido aquele com menor grau de autocorrelação residual. A tabela seguinte sintetiza os modelos que se revelaram mais adequados para a modelação das séries em estudo.

Tabela 3.13: Modelos SARIMA selecionados por localização e profundidade com base na aleatoriedade dos resíduos e critérios de informação.

Localidade	Profundidade	Modelo Selecionado	BIC	AICc	p-valor (Ljung-Box)
Nazaré	1 m	SARIMA(6,0,3)(1,1,2)[12]	266.37	208.66	0.1189
	5 m	SARIMA(6,0,4)(0,1,2)[12]	290.05	232.34	0.1655
Faro	1 m	SARIMA(6,0,6)(0,1,1) [12]	788.8	726.69	0.3124
	5 m	SARIMA(2,0,6)(3,1,0)[12]	863.17	809.86	0.08367
São Miguel	1 m	SARIMA(5,0,5)(0,1,1)[12]	904.95	851.64	0.0288
São Miguel	5 m	SARIMA(6,0,6)(0,1,1)[12]	925.73	863.62	0.08473
Praia	1 m	SARIMA(5,0,3)(0,1,1)[12]	512.89	468.4	0.09395
	5m	SARIMA(4,0,4)(0,1,1)[12]	614.73	570.24	0.1373

Fonte: Elaboração própria.

Equações dos modelos SARIMA estimados

A seguir apresentam-se as equações dos modelos selecionados, estimadas para cada série temporal com base na estimação por máxima verosimilhança. Em todos os casos, o período sazonal é $s = 12$.

Nazaré – Um metro de profundidade

Modelo selecionado: $SARIMA(6,0,3)(1,1,2)[12]$.

A equação geral do modelo é dada por:

$$\begin{aligned} (1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4 - \phi_5 B^5 - \phi_6 B^6)(1 - \Phi_1 B^{12})(1 - B^{12})X_t = \\ \underbrace{\hspace{10em}}_{AR \text{ não sazonal}} \underbrace{\hspace{10em}}_{AR \text{ sazonal}} \underbrace{\hspace{10em}}_{Diferença sazonal} \\ = \underbrace{(1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3)}_{MA \text{ não sazonal}} \underbrace{(1 + \theta_1 B^{12} + \theta_2 B^{24})}_{MA \text{ sazonal}} \varepsilon_t \end{aligned}$$

Substituindo os coeficientes pelas respectivas estimativas, obtêm-se:

$$\hat{\phi}_1 = -0,3265, \hat{\phi}_2 = 0,2020, \hat{\phi}_3 = 0,1833,$$

$$\hat{\phi}_4 = -0,0916, \hat{\phi}_5 = 0,1488, \hat{\phi}_6 = -0,2822,$$

$$\hat{\theta}_1 = 1,3503, \hat{\theta}_2 = 0,6464, \hat{\theta}_3 = -0,0019,$$

$$\hat{\Phi}_1 = -0,5460, \hat{\Theta}_1 = -0,0024 \text{ e } \hat{\Theta}_2 = -0,3629.$$

A equação estimada do modelo SARIMA resulta então em:

$$\begin{aligned} (1 + 0,3265B - 0,2020B^2 - 0,1833B^3 + 0,0916B^4 - 0,1488B^5 \\ + 0,2822B^6)(1 + 0,5460B^{12})(1 - B^{12})X_t \\ = (1 + 1,3503B + 0,6464B^2 - 0,0019B^3)(1 - 0,0024B^{12} \\ - 0,3629B^{24}) \varepsilon_t \end{aligned} \quad (27)$$

com $\varepsilon_t \sim \text{ruído branco}(0, \hat{\sigma}^2 = 0,07814)$.

Nazaré – Cinco metros de profundidade

Modelo selecionado: $SARIMA(6,0,4)(0,1,2)[12]$.

A equação geral do modelo selecionado é dada por:

$$\begin{aligned} (1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4 - \phi_5 B^5 - \phi_6 B^6)(1 - B^{12})X_t = \\ = (1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3 + \theta_4 B^4)(1 + \theta_1 B^{12} + \theta_2 B^{24}) \varepsilon_t \end{aligned}$$

Após a estimação dos respectivos coeficientes, obteve-se a seguinte equação do modelo:

$$\begin{aligned}
& (1 + 0,5438B + 0,2597B^2 + 0,0590B^3 - 0,0907B^4 - 0,3061B^5 \\
& \quad + 0,3123B^6)(1 - B^{12})X_t \\
& = (1 + 1,5642B + 1,3898B^2 + 0,9057B^3 + 0,4053B^4)(1 - \\
& \quad 0,6577B^{12} - 0,0209B^{24})\varepsilon_t
\end{aligned} \tag{28}$$

com $\varepsilon_t \sim \text{ruído branco}$ ($0, \hat{\sigma}^2 = 0,08085$).

Faro – Um metro de profundidade

Modelo selecionado: *SARIMA*(6,0,6)(0,1,1) [12].

A equação geral do modelo selecionado é dada por:

$$\begin{aligned}
& (1 - \phi_1B - \phi_2B^2 - \phi_3B^3 - \phi_4B^4 - \phi_5B^5 - \phi_6B^6)(1 - B^{12})X_t = \\
& = (1 + \theta_1B + \theta_2B^2 + \theta_3B^3 + \theta_4B^4 + \theta_5B^5 + \theta_6B^6)(1 + \theta_1B^{12})\varepsilon_t
\end{aligned}$$

Substituindo as estimativas dos coeficientes na equação geral, obtém-se:

$$\begin{aligned}
& (1 + 2,3182B + 2,6712B^2 + 2,3206B^3 + 1,3581B^4 + 0,0747B^5 \\
& \quad - 0,2892B^6)(1 - B^{12})X_t \\
& = (1 + 3,1745B + 4,9424B^2 + 5,3054B^3 + 4,1872B^4 + \\
& \quad 2,1140B^5 + 0,4819B^6)(1 - 0,7923B^{12})\varepsilon_t
\end{aligned} \tag{29}$$

com $\varepsilon_t \sim \text{ruído branco}$ ($0, \hat{\sigma}^2 = 0,1713$).

Faro – Cinco metros de profundidade

Modelo selecionado: *SARIMA*(2,0,6)(3,1,0)[12].

A equação geral do modelo selecionado é dada por:

$$\begin{aligned}
& (1 - \phi_1B - \phi_2B^2)(1 - \phi_1B^{12} - \phi_2B^{24} - \phi_3B^{36})(1 - B^{12})X_t = \\
& = (1 + \theta_1B + \theta_2B^2 + \theta_3B^3 + \theta_4B^4 + \theta_5B^5 + \theta_6B^6)\varepsilon_t
\end{aligned}$$

Fazendo a substituição dos coeficientes pelas suas estimativas, tem-se:

$$\begin{aligned}
& (1 + 1,4731B + 0,6772B^2)(1 + 0,7114B^{12} + 0,4545B^{24} \\
& \quad + 0,2213B^{36})(1 - B^{12})X_t \\
& = (1 + 2,3578B + 2,2455B^2 + 0,9905B^3 + 0,1213B^4 + \\
& \quad 0,0020B^5 + 0,0586B^6)\varepsilon_t
\end{aligned} \tag{30}$$

com $\varepsilon_t \sim \text{ruído branco}(0, \sigma^2 = 0,1955)$.

São Miguel – Um metro de profundidade

Modelo selecionado: *SARIMA*(5,0,5)(0,1,1)[12].

A equação geral do modelo selecionado é dada por:

$$\begin{aligned}
& (1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4 - \phi_5 B^5)(1 - B^{12})X_t = \\
& = (1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3 + \theta_4 B^4 + \theta_5 B^5)(1 + \theta_1 B^{12}) \varepsilon_t
\end{aligned}$$

Efetuada a substituição dos coeficientes pelas respectivas estimativas, vem:

$$\begin{aligned}
& (1 + 0,3213B + 0,2016B^2 + 0,5152B^3 + 0,5498B^4 \\
& \quad - 0,234B^5)(1 - B^{12})X_t \\
& = (1 + 1,1103B + 0,7317B^2 + 0,8733B^3 + 1,1963B^4 \\
& \quad + 0,5399B^5)(1 - 0,5130B^{12})\varepsilon_t
\end{aligned} \tag{31}$$

com $\varepsilon_t \sim \text{ruído branco}(0, \hat{\sigma}^2 = 0,2088)$.

São Miguel – Cinco metros de profundidade

Modelo selecionado: *SARIMA*(6,0,6)(0,1,1)[12].

A equação geral do modelo selecionado é dada por:

$$\begin{aligned}
& (1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4 - \phi_5 B^5 - \phi_6 B^6)(1 - B^{12})X_t = \\
& = (1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3 + \theta_4 B^4 + \theta_5 B^5 + \theta_6 B^6)(1 + \theta_1 B^{12}) \varepsilon_t
\end{aligned}$$

Uma vez substituídos os coeficientes pelas estimativas respectivas, obtém:

$$\begin{aligned}
& (1 + 0,0455B + 0,1089B^2 + 0,473B^3 + 0,4603B^4 - 0,429B^5 \\
& \quad + 0,0497B^6)(1 - B^{12})X_t
\end{aligned}$$

$$= (1 + 0,8297B + 0,4449B^2 + 0,7245B^3 + 0,978B^4 + 0,2015B^5 - 0,1356B^6)(1 - 0,5785B^{12})\varepsilon_t \quad (32)$$

com $\varepsilon_t \sim \text{ruído branco}(0, \hat{\sigma}^2 = 0,2105)$.

Praia – Um metro de profundidade

Modelo selecionado: *SARIMA*(5,0,3)(0,1,1)[12].

A equação geral do modelo selecionado é dada por:

$$\begin{aligned} (1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4 - \phi_5 B^5)(1 - B^{12})X_t = \\ = (1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3)(1 + \theta_4 B^{12})\varepsilon_t \end{aligned}$$

Após a substituição dos coeficientes pelas suas estimativas, tem-se:

$$\begin{aligned} (1 + 0,6566B + 0,289B^2 - 0,1162B^3 - 0,2416B^4 \\ - 0,1592B^5)(1 - B^{12})X_t \\ = (1 + 1,5532B + 1,2759B^2 + 0,5788B^3)(1 - 0,4811B^{12})\varepsilon_t \end{aligned} \quad (33)$$

com $\varepsilon_t \sim \text{ruído branco}(0, \hat{\sigma}^2 = 0,1169)$.

Praia – Cinco metros de profundidade

Modelo selecionado: *SARIMA*(4,0,4)(0,1,1)[12].

A sua equação geral do modelo é:

$$\begin{aligned} (1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4)(1 - B^{12})X_t = \\ = (1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3 + \theta_4 B^4)(1 + \theta_5 B^{12})\varepsilon_t \end{aligned}$$

Procedendo à substituição dos seus coeficientes pelas respectivas estimativas, vem:

$$\begin{aligned} (1 + 0,5035B + 0,1977B^2 - 0,3816B^3 - 0,2051B^4)(1 \\ - B^{12})X_t \\ = (1 + 1,4589B + 1,1717B^2 + 0,3383B^3 - \\ 0,1513B^4)(1 - 0,5435B^{12})\varepsilon_t \end{aligned} \quad (34)$$

com $\varepsilon_t \sim \text{ruído branco}(0, \hat{\sigma}^2 = 0,1363)$.

3.6- Previsões utilizando os modelos SARIMA selecionados

Utilizando os modelos SARIMA validados na Secção 3.6, procedeu-se ao cálculo de previsões para o horizonte de julho de 2023 a junho de 2024. Este período corresponde a doze meses consecutivos, permitindo capturar um ciclo sazonal completo.

As previsões foram obtidas recorrendo à função *forecast()* do pacote *forecast* em R, com os seguintes parâmetros: Horizonte: $h = 12$ meses; Níveis de confiança: 80% e 95%, em conformidade com as boas práticas de previsão (Hyndman & Athanasopoulos, 2021).

Apresentam-se na [Tabela 3.14](#) as previsões pontuais mensais para as localidades e respetivas profundidades em estudo.

Tabela 3.14: Previsões pontuais mensais de temperatura (jul/2023–jun/2024) por localização e profundidade.

Mês/Ano	Nazaré 1m	Nazaré 5m	Faro 1m	Faro 5m	S. Miguel 1m	S. Miguel 5m	Praia 1m	Praia 5m
Jul. 2023	19,19	18,98	21,49	21,35	22,19	21,82	24,67	25,65
Ago. 2023	19,86	19,72	22,37	22,42	23,92	23,67	27,69	27,96
Sep. 2023	19,53	19,33	22,61	22,63	24,02	23,97	28,73	28,90
Out. 2023	19,61	19,53	22,50	22,49	23,44	23,36	28,31	28,55
Nov. 2023	17,97	17,98	19,81	19,88	19,43	19,54	26,80	27,19
Dez. 2023	16,22	16,26	17,56	17,42	17,32	17,38	24,90	24,64
Jan. 2024	15,27	15,29	16,89	16,92	17,17	17,04	23,91	23,88
Fev. 2024	14,79	14,75	16,69	16,78	16,45	16,40	23,28	23,44
Mar. 2024	14,58	14,51	16,43	16,48	16,03	16,07	23,18	23,21
Abr. 2024	15,45	15,28	17,24	17,33	16,14	16,15	23,46	23,34
Mai 2024	16,89	16,69	18,74	18,66	18,17	18,09	24,17	24,14
Jun. 2024	18,54	18,33	20,68	20,55	20,76	20,70	24,77	25,31

Fonte: Elaboração própria com base em modelos SARIMA (R, pacote forecast).

Nota: As previsões intervalares por localização e profundidade encontram-se nos Anexos II e III.

A [Figura 3.11](#) ilustra graficamente as séries de temperaturas médias mensais, as previsões pontuais (em azul) e os respetivos intervalos de confiança a 80% e 95% (em tons sombreados de azul), gerados pelos modelos SARIMA selecionados para cada localidade e profundidade.

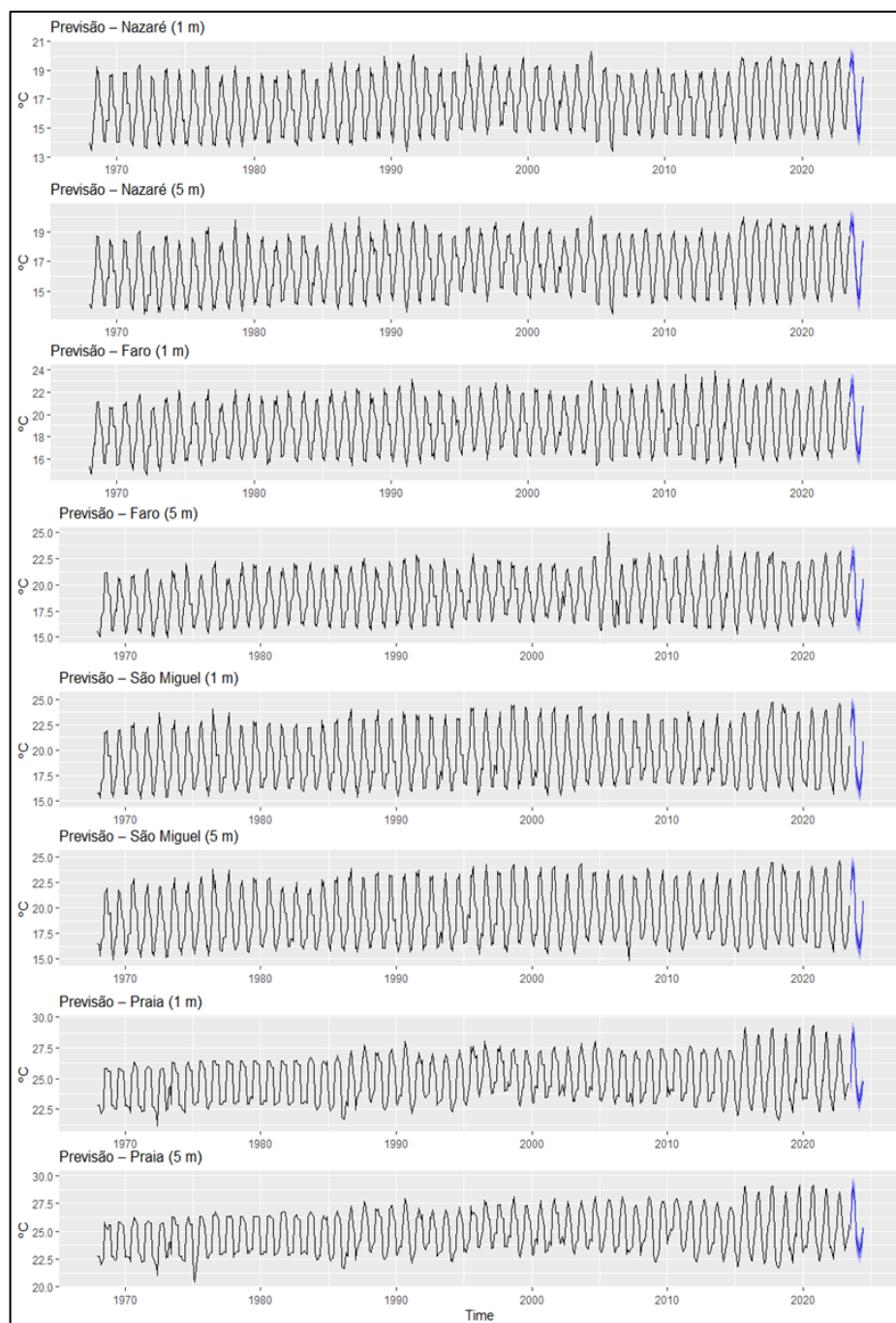


Figura 3.11: Previsões das temperaturas médias mensais do mar (°C) para julho de 2023 a junho de 2024, por localidade e profundidade, incluindo previsões pontuais e intervalos de confiança a 80% e 95%.

Fonte: Fonte: Elaboração própria.

As previsões calculadas mantêm a estrutura sazonal anual observada nas séries históricas. Quanto à validade estatística, reforça-se que a maioria dos modelos satisfaz os critérios de diagnóstico (Secção 3.6). A única exceção é o modelo de São Miguel (1m), cujas previsões devem ser interpretadas com cautela, pois os seus resíduos são autocorrelacionados.

À data da presente redação, todo o horizonte previsto (julho de 2023 a junho de 2024) já era observável, permitindo uma validação *ex post* da capacidade preditiva dos modelos SARIMA propostos.

3.7- Avaliação da qualidade das previsões

Com o objetivo de avaliar o desempenho preditivo dos modelos SARIMA selecionados, procedeu-se à comparação entre as previsões pontuais e os valores observados. Os modelos foram ajustados com base nos dados históricos compreendidos entre janeiro de 1968 e junho de 2023. A partir desses modelos ajustados, foram geradas previsões para o período de julho de 2023 a junho de 2024.

Tabela 3.15: Valores observados e estimados da temperatura da água do mar em diferentes profundidades (Nazaré e Faro), 2023-2024.

Mês	Nazaré 1 m		Nazaré 5 m		Faro 1 m		Faro 5 m	
	Valor Observado	Previsto	Valor Observado	Previsto	Valor Observado	Previsto	Valor Observado	Previsto
Jul. 2023	19,99	19,19	19,85	18,98	21,93	21,49	21,86	21,35
Ago. 2023	20,91	19,86	20,85	19,72	21,94	22,37	21,80	22,42
Set. 2023	20,98	19,53	20,94	19,33	22,15	22,61	22,10	22,63
Out. 2023	20,26	19,61	20,22	19,53	22,28	22,50	22,28	22,49
Nov. 2023	18,60	17,97	18,59	17,98	20,63	19,81	20,63	19,88
Dez. 2023	17,03	16,22	17,00	16,26	18,87	17,56	18,84	17,42
Jan. 2024	15,63	15,27	15,64	15,29	17,77	16,89	17,79	16,92
Fev. 2024	15,16	14,79	15,15	14,75	17,42	16,69	17,42	16,78
Mar. 2024	14,98	14,58	14,95	14,51	17,22	16,43	17,20	16,48
Abr. 2024	15,67	15,45	15,63	15,28	17,64	17,24	17,69	17,33
Maio 2024	16,87	16,89	16,82	16,69	18,91	18,74	18,87	18,66
Jun. 2024	18,15	18,54	18,04	18,33	20,29	20,68	20,18	20,55

Fonte: Elaboração própria.

Tabela 3.16: Valores observados e estimados da temperatura da água do mar em diferentes profundidades (S. Miguel e Praia), 2023-2024.

Mês	S. Miguel 1 m		S. Miguel 5 m		Praia 1 m		Praia 5 m	
	Valor Observado	Previsto	Valor Observado	Previsto	Valor Observado	Previsto	Valor Observado	Previsto
Jul. 2023	22,83	22,19	22,76	21,82	27,47	24,67	27,45	25,65
Ago. 2023	24,51	23,92	24,42	23,67	28,69	27,69	28,66	27,96
Set. 2023	23,90	24,02	23,79	23,97	28,93	28,73	28,88	28,90
Out. 2023	22,29	23,44	22,26	23,36	28,98	28,31	28,94	28,55
Nov. 2023	20,06	19,43	20,19	19,54	27,96	26,80	27,95	27,19
Dez. 2023	18,90	17,32	19,09	17,38	26,27	24,90	26,31	24,64
Jan. 2024	18,23	17,17	18,30	17,04	24,54	23,91	24,57	23,88
Fev. 2024	17,62	16,45	17,59	16,40	24,08	23,28	24,12	23,44
Mar. 2024	17,01	16,03	17,08	16,07	23,92	23,18	23,94	23,21
Abr. 2024	17,19	16,14	17,26	16,15	24,17	23,46	24,12	23,34
Mai. 2024	18,96	18,17	18,98	18,09	24,81	24,17	24,85	24,14
Jun. 2024	21,55	20,76	21,53	20,70	25,08	24,77	25,05	25,31

Fonte: Elaboração própria.

Para quantificar de forma objetiva o erro das previsões, calcularam-se as métricas de desempenho: Erro Absoluto Médio (MAE), Raiz do Erro Quadrático Médio (RMSE) e Erro Percentual Absoluto Médio (MAPE), cujos resultados são apresentados na [Tabela 3.17](#).

Tabela 3.17: Métricas de avaliação da qualidade preditiva dos modelos SARIMA.

Localidade	Profundidade	MAE (° C)	RMSE (° C)	MAPE (%)
Nazaré	1 m	0,5958	0,7040	3,20
	5 m	0,6342	0,7484	3,41
Faro	1 m	0,5867	0,6645	3,06
	5 m	0,6008	0,6797	3,12
S. Miguel	1 m	0,8792	0,9483	4,54
	5 m	0,9683	1,0317	4,97
Praia	1 m	0,9192	1,1235	3,48
	5 m	0,7658	0,9083	2,94

Fonte: Elaboração própria.

Com base na análise das [Tabelas 3.15 a 3.17](#), observa-se que as previsões estão, de forma geral, próximas dos valores observados. Os valores mais reduzidos de MAE e RMSE,

observados em Nazaré e Faro, sugerem que os modelos SARIMA capturam com maior precisão a dinâmica térmica nestas regiões, relativamente a São Miguel e Praia.

Análise de Concordância de Bland-Altman

Para além das métricas de erro apresentadas na [Tabela 3.17](#), recorreu-se ao método de Bland-Altman para avaliar graficamente a concordância entre valores observados e previstos (Bland & Altman, 1986). Esta abordagem permite identificar visualmente se o modelo apresenta tendência para subestimar ou sobrestimar a temperatura, e se o erro se mantém estável ao longo do intervalo de valores analisado.

As [Figuras 3.12](#) e [3.13](#) ilustram esta análise para as quatro localidades e profundidades em estudo.

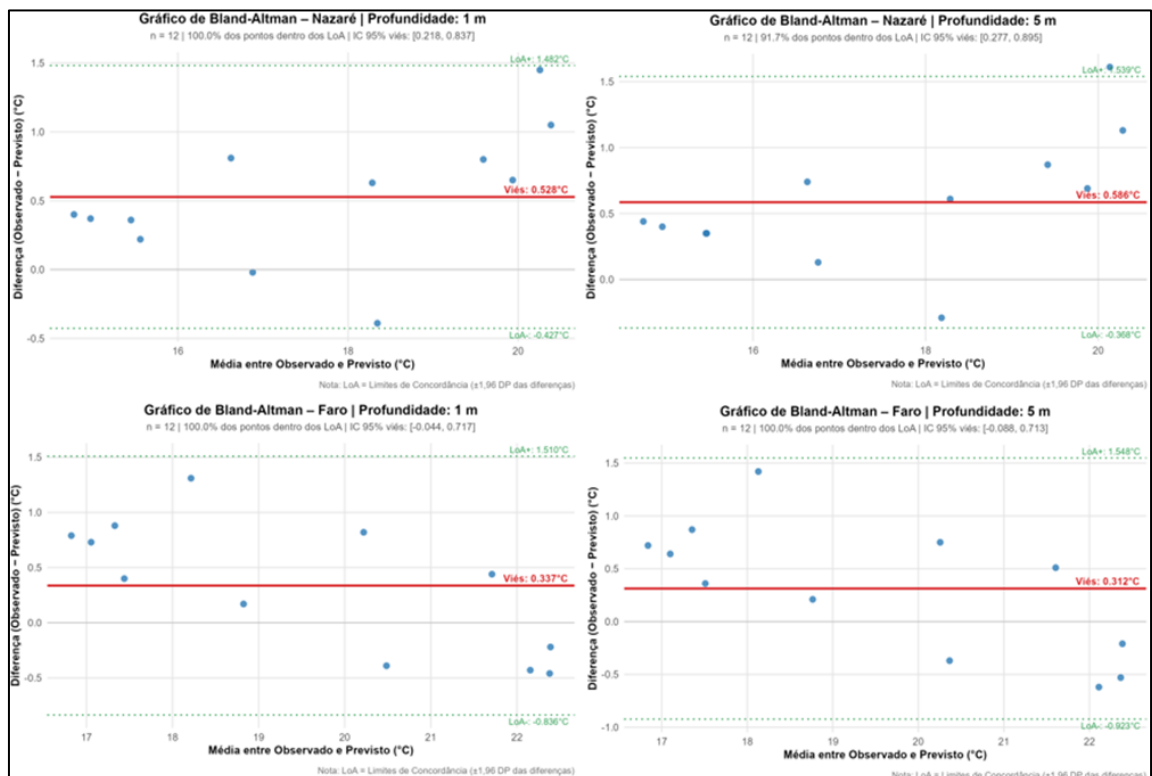


Figura 3.12: Análise de concordância de Bland-Altman para as previsões Nazaré e Faro (jul. 2023 – jun. 2024).

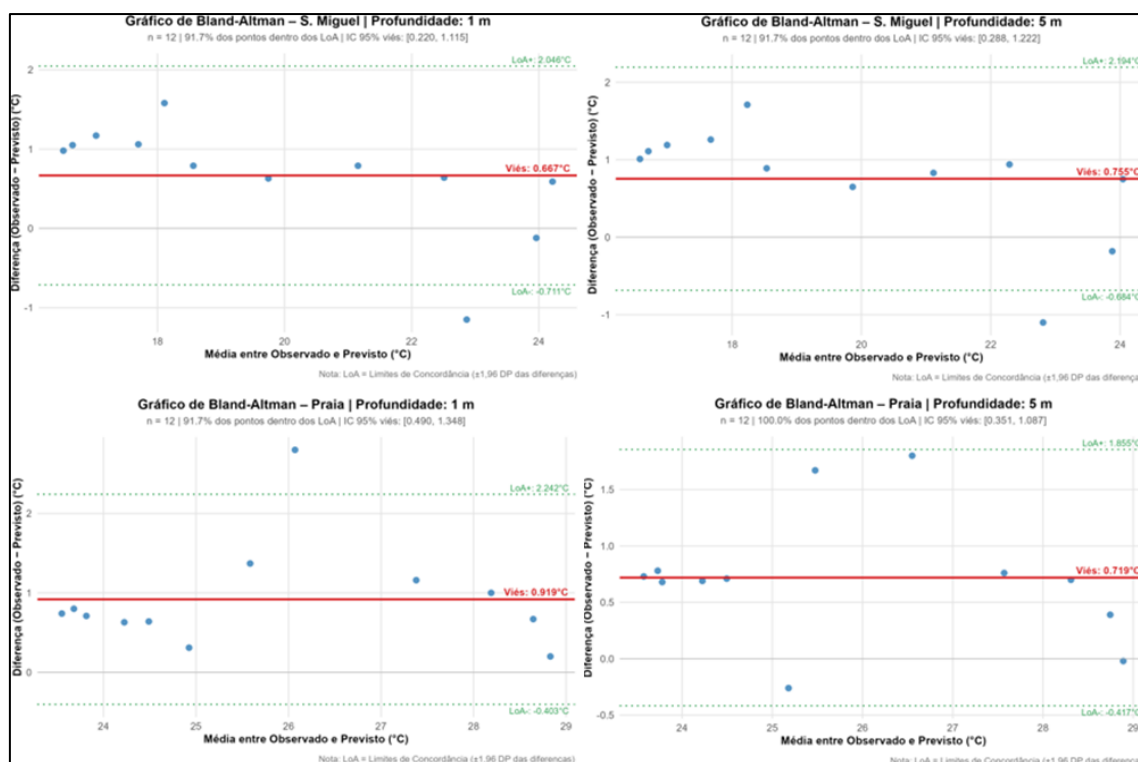


Figura 3.13: Análise de concordância de Bland-Altman para as previsões São Miguel e Praia (jul. 2023 – jun. 2024).

A análise de concordância de Bland-Altman ([Figuras 3.12 e 3.13](#)) revela que o modelo SARIMA apresenta viés positivo estatisticamente significativo em três das quatro localidades em estudo (Nazaré, São Miguel e Praia), em ambas as profundidades, indicando uma tendência para subestimar a temperatura da água do mar. Em Faro, o viés foi menor (0,31-0,34°C) e não estatisticamente significativo em ambas as profundidades (IC 95%: [-0,044; 0,717] a 1 m e [-0,088; 0,713] a 5 m, ambos contendo zero), sugerindo melhor desempenho preditivo nesta localidade.

Em todas as localidades e profundidades, a grande maioria das observações (pelo menos 91,7%) situou-se dentro dos limites de concordância. Estes resultados corroboram as métricas de erro da [Tabela 3.17](#) e sugerem que, apesar da tendência consistente para subestimar a temperatura na maioria das localidades, o modelo SARIMA é adequado para realizar previsões.

4- Capítulo IV – Fundamentos Teóricos dos Modelos Espaço–Temporais

4.1- Limitações dos modelos univariados e motivação para modelos espaço–temporais

No capítulo anterior, a análise concentrou-se exclusivamente na dimensão temporal das temperaturas médias mensais observadas nas regiões de Nazaré, Faro e São Miguel (Portugal) e Praia (Cabo Verde), em ambas as profundidades, recorrendo a abordagens clássicas de séries temporais univariadas, como os modelos de decomposição, o método STL, técnicas de alisamento exponencial e os modelos ARIMA e SARIMA (Box et al., 2015; Hyndman & Athanasopoulos, 2021). Embora estas metodologias permitam modelar adequadamente a evolução temporal de cada série individual, ignoram possíveis relações espaciais entre diferentes locais (Pfeifer & Deutsch, 1980a).

Com o intuito de ultrapassar esta limitação, surgiram os modelos espaço-temporais, que permitem capturar simultaneamente a dinâmica temporal dos dados e as dependências espaciais entre diferentes regiões. Entre as várias propostas, destacam-se os modelos STARMA (*Space-Time Autoregressive Moving Average*), formalizados por Pfeifer e Deutsch (1980a), que se revelam particularmente adequados quando existe autocorrelação espacial estatisticamente significativa (Pfeifer & Deutsch, 1980a).

Os modelos STARMA resultam da combinação de três componentes fundamentais: a componente autorregressiva, a componente de médias móveis e a dimensão espacial definida sobre um conjunto de N locais fixos. A dependência entre os diferentes locais é integrada no modelo através de matrizes de ponderação, que representam a influência que cada local exerce sobre os seus vizinhos. Estas matrizes são definidas a priori pelo investigador e devem refletir, tanto quanto possível, as características reais do fenómeno que se pretende modelar (Pfeifer & Deutsch, 1980a).

4.2- Operador de desfasagem espacial

Seja $L^{(l)}$ o operador de desfasagem espacial de ordem l , tal que:

$$L^{(0)}X_t = X_t,$$

e, para $l > 0$,

$$(\mathbf{L}^{(l)} \mathbf{X}_t)_i = \sum_{j=1}^N w_{ij}^{(l)} \mathbf{X}_{tj}, \quad (35)$$

para cada $i = 1, \dots, N$, onde $w_{ij}^{(l)}$ representa o peso atribuído ao local j na influência exercida sobre o local i , considerando apenas os vizinhos de ordem l . Os pesos $w_{ij}^{(l)}$ são não negativos e satisfazem a condição de normalização:

$$\sum_{j=1}^N w_{ij}^{(l)} = 1, \quad (36)$$

$$\forall i = 1, \dots, N.$$

Na forma matricial, os pesos $w_{ij}^{(l)}$ são organizados numa matriz quadrada $\mathbf{W}^{(l)}$ de ordem N , onde a soma dos elementos de cada linha é igual a 1. Assim, temos:

$$\mathbf{L}^{(0)} \mathbf{X}_t = \mathbf{W}^{(0)} \mathbf{X}_t = \mathbf{I}_N \mathbf{X}_t,$$

e, para $l > 0$,

$$\mathbf{L}^{(l)} \mathbf{X}_t = \mathbf{W}^{(l)} \mathbf{X}_t.$$

A construção da matriz de ponderação espacial $\mathbf{W}$ não obedece a um procedimento único. Existem diversas abordagens possíveis, e a escolha mais adequada depende tanto da natureza do fenómeno em análise como da configuração geográfica dos locais observados.

Quando os locais se distribuem de forma regular, como numa malha, é comum definir os vizinhos com base numa hierarquia discreta: os vizinhos de primeira ordem correspondem aos pontos mais próximos; os de segunda ordem são aqueles ligeiramente mais distantes, mais afastados que os de primeira ordem, mas ainda mais próximos do que os de terceira ordem, e assim sucessivamente para ordens superiores.

Os elementos da matriz de pesos, neste caso, são definidos como:

$$w_{ij}^{(l)} = \begin{cases} \frac{1}{n_i^{(l)}}, & \text{se } j \text{ é vizinho de } i \text{ na ordem } l \\ 0, & \text{caso contrário} \end{cases}, \quad (37)$$

onde $n_i^{(l)}$ é o número de vizinhos do local i na l -ésima ordem de vizinhança.

Na prática, a grande maioria dos conjuntos de dados espaciais costuma estar distribuída de forma irregular, como é o caso das boias oceânicas analisadas neste trabalho (Nazaré, Faro, São Miguel e Praia), o que inviabiliza a aplicação direta dessa abordagem hierárquica baseada em malhas regulares.

Nestes contextos, é comum recorrer a definições de vizinhança que incorporam explicitamente a distância geográfica entre os locais. Uma das abordagens mais utilizadas consiste na ponderação por distância inversa, segundo a qual o peso atribuído ao par (i, j) é inversamente proporcional à distância entre os locais i e j . Formalmente, define-se:

$$w_{ij} = \begin{cases} \frac{1}{d_{ij}}, & \text{se } i \neq j \\ 0, & i = j \end{cases},$$

onde d_{ij} corresponde à distância geodésica (em quilómetros) entre as boias i e j .

4.3- Definição formal do modelo STARMA

Seja $\mathbf{X}_t \in \mathbb{R}^N$ o vetor de observações no instante t nas N localizações consideradas. Um modelo STARMA é dado por (Pfeifer & Deutsch, 1980a):

$$\mathbf{X}_t = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{k\ell} \mathbf{W}^{(l)} \mathbf{X}_{t-k} - \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{k\ell} \mathbf{W}^{(l)} \boldsymbol{\varepsilon}_{t-k} + \boldsymbol{\varepsilon}_t, \quad (38)$$

onde:

- p é a ordem autorregressiva;
- q é a ordem de médias móveis;
- λ_k é a ordem espacial associada ao k -ésimo termo autorregressivo;
- m_k é a ordem espacial associada ao k -ésimo termo de médias móveis;
- $\phi_{k\ell}$ é o parâmetro autorregressivo correspondente ao desfasamento temporal k e à ordem espacial l ;
- $\theta_{k\ell}$ é o parâmetro de médias móveis correspondente ao desfasamento temporal k e à ordem espacial l ;
- $\mathbf{W}^{(l)}$ é a matriz de pesos da ordem espacial l ;

- $\boldsymbol{\varepsilon}_t$ representa o vetor de erros aleatórios, com distribuição normal multivariada, com as seguintes propriedades:

$$E[\boldsymbol{\varepsilon}_t] = \mathbf{0}$$

e

$$E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_{t+s}^\top] = \begin{cases} \sigma^2 \mathbf{I}_N, & s = 0 \\ \mathbf{0}, & s \neq 0 \end{cases}.$$

Segundo Pfeifer & Deutsch (1980a), dentro da família de modelos STARMA destacam-se dois casos particulares. Quando $q = 0$, o modelo reduz-se apenas a termos autoregressivos, sendo designado por **modelo autoregressivo espaço-temporal (STAR)**, dado pela seguinte equação:

$$\mathbf{X}_t = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} \mathbf{W}^{(l)} \mathbf{X}_{t-k} + \boldsymbol{\varepsilon}_t.$$

Já quando o modelo não possui termos autoregressivos, ou seja, $p = 0$, é denominado **modelo de médias móveis espaço-temporal (STMA)**, expresso por:

$$\mathbf{X}_t = - \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{kl} \mathbf{W}^{(l)} \boldsymbol{\varepsilon}_{t-k} + \boldsymbol{\varepsilon}_t.$$

4.4- Condições de estacionaridade, causalidade e invertibilidade

4.4.1-Estacionaridade de segunda ordem no domínio espaço-temporal

De acordo com Pfeifer & Deutsch (1980a) os modelos STARMA pressupõem que o processo estocástico seja estacionário de segunda ordem no domínio espaço-temporal. Isto significa que:

- O valor médio é constante em todos os locais e instantes:

$$E[X_{i,t}] = \mu, \quad \forall i = 1, \dots, N, \forall t \in \mathbb{Z};$$

- A covariância entre $X_{i,t}$ e $X_{j,t+s}$ depende apenas da ordem de vizinhança espacial l e k associadas aos locais i e j e do desfasamento temporal s , isto é:

$$\text{Cov}(X_{i,t}, X_{j,t+s}) = \gamma_{lk}(s),$$

onde γ_{lk} representa a função de covariância espaço-temporal entre vizinhos de l -ésima e k -ésima ordem, e a estrutura de covariância não varia com o tempo.

4.4.2-Condições algébricas de causalidade e invertibilidade

De acordo com Pfeifer & Deutsch (1980a), para que um modelo STARMA espaço-temporal descreva um processo estacionário, deve satisfazer duas condições estruturais:

- **Causalidade:** as raízes da equação

$$\det \left(x^p \mathbf{I}_N - \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} \mathbf{W}^{(l)} x^{p-k} \right) = 0, \quad (39)$$

devem situar-se no interior do círculo unitário ($|x| < 1$).

- **Invertibilidade:** as raízes da equação

$$\det \left(x^q \mathbf{I}_N - \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{kl} \mathbf{W}^{(l)} x^{q-k} \right) = 0, \quad (40)$$

devem situar-se no interior do círculo unitário ($|x| < 1$).

Estas condições garantem que o processo $\{X_t\}$ possa ser representado como uma combinação linear ponderada de erros passados em que os pesos decrescem assintoticamente para zero.

Como os STAR não têm componente de médias móveis, são sempre invertíveis, já os STMA, ao não incluírem termos autoregressivos, são estacionários por natureza.

Nota: Segue-se aqui a convenção de Pfeifer & Deutsch (1980a), em que a variável x representa a inversa das raízes do polinómio autoregressivo tradicional. Assim, a condição $|x| < 1$ equivale à condição padrão $|r| > 1$, onde r denota as raízes do polinómio autoregressivo escrito na forma tradicional.

4.4.3- Extensão dos modelos STARMA a processos não estacionários

Em contextos práticos de modelação espaço-temporal, como o analisado neste estudo, que envolve séries mensais de temperatura em diferentes localidades e profundidades, os processos raramente são estacionários, isto é, é muito comum os dados apresentarem não apenas a tendência, mas também uma componente sazonal. Para fazer face a estas fontes de não estacionaridade, adota-se uma abordagem análoga à utilizada nos modelos SARIMA no Capítulo III: aplica-se diferenciação às séries antes do ajuste do modelo STARMA. Contudo, tal como verificado na análise univariada, a diferenciação sazonal de ordem 1 (com período $S=12$) foi suficiente para estabilizar a média e remover a componente sazonalidade. Assim, não houve necessidade de aplicar diferenciação regular ($d = 0$).

Assim, neste estudo, ajusta-se um modelo STARMA à série sazonalmente diferenciada, definida por

$$Y_t = (1 - B^{12})X_t,$$

onde $X_t = (X_{1,t}, X_{2,t}, X_{3,t}, X_{4,t})'$ representa o vetor de temperaturas nas quatro localizações em estudo (Nazaré, Faro, São Miguel e Praia) e B é o operador de desfasamento temporal.

4.5- Identificação, estimação e diagnóstico em modelos STARMA

4.5.1-Identificação de modelos espaço-temporais

Segundo Pfeifer & Deutsch (1980a), na identificação de modelos espaço-temporais que melhor descrevem a estrutura de dependência dos dados observados, recorre-se à combinação de informações visuais extraídas do gráfico da função de autocorrelação espaço-temporal, complementadas pela análise da função de autocorrelação parcial espaço-temporal. Essa função de autocorrelação é obtida a partir do cálculo da função de autocovariância entre todos os pares de locais em estudo, resultando num total de N^2 covariâncias.

Recorrendo ao operador de desfasamento espacial $L^{(l)}$, definido anteriormente tal que $L^{(l)}X_t = W^{(l)}X_t$, define-se a covariância entre os vizinhos ponderados de ordens l e k de um mesmo local, separados por um desfasamento temporal s , da seguinte forma:

$$\gamma_{lk}(s) = E \left\{ \sum_{i=1}^N \frac{L^{(l)}(X_t)_i L^{(k)}(X_{t+s})_i}{N} \right\}.$$

Em termos matriciais, esta expressão pode ser reescrita como:

$$\gamma_{lk}(s) = \text{tr} \left\{ \frac{\mathbf{W}^{(k)'} \mathbf{W}^{(l)} \boldsymbol{\Gamma}(s)}{N} \right\},$$

onde $\boldsymbol{\Gamma}(s) = E(\mathbf{X}_t \mathbf{X}_{t+s}')^*$ é a matriz de covariâncias cruzadas.

O estimador amostral centrado de $\boldsymbol{\Gamma}(s)$ é dado por:

$$\hat{\boldsymbol{\Gamma}}(s) = \frac{1}{T-s} \sum_{t=1}^{T-s} (\mathbf{X}_t - \bar{\mathbf{X}})(\mathbf{X}_{t+s} - \bar{\mathbf{X}})',$$

onde $\bar{\mathbf{X}} = \frac{1}{T} \sum_{t=1}^T \mathbf{X}_t$.

A partir de $\hat{\boldsymbol{\Gamma}}(s)$, obtém-se os valores estimados da autocovariância:

$$\hat{\gamma}_{lk}(s) = \text{tr} \left\{ \frac{\mathbf{W}^{(k)'} \mathbf{W}^{(l)} \hat{\boldsymbol{\Gamma}}(s)}{N} \right\}.$$

Uma vez definida a função de autocovariância, a função de autocorrelação espaço-temporal entre vizinhos de ordem l e k , desfasados s unidades no tempo, é dada por:

$$\rho_{lk}(s) = \frac{\gamma_{lk}(s)}{\sqrt{\gamma_{ll}(0)\gamma_{kk}(0)}}.$$

Os valores de ρ_{lk} pertencem ao intervalo $[-1,1]$, o que permite comparações entre pares de ordens de vizinhança. Segundo Pfeifer & Deutsch (1980a), esta definição é preferível, porque o seu estimador amostral apresenta variabilidade aproximadamente constante para todos os pares (l, k) . As suas estimativas são dadas por:

$$\hat{\rho}_{lk}(s) = \frac{\hat{\gamma}_{lk}(s)}{\sqrt{\hat{\gamma}_{ll}(0)\hat{\gamma}_{kk}(0)}}.$$

Apesar da utilidade da função de autocorrelação espaço-temporal, a identificação do modelo STARMA adequado permanece um desafio. Para contornar esta dificuldade, recorre-se à **função de autocorrelação parcial espaço-temporal (STPACF)**, introduzida por Martin e Oeppen (1975), que se baseia na estrutura do modelo STAR (*Space-Time Autoregressive*).

Para obter uma equação análoga às de Yule-Walker no domínio espaço-temporal, multiplica-se ambos os membros da equação por $(\mathbf{W}^{(h)} \mathbf{X}_{t-s})'$ e toma-se o valor esperado.

$$E \left[(\mathbf{W}^{(h)} \mathbf{X}_{t-s})' \mathbf{X}_t \right] = \sum_{j=1}^p \sum_{l=0}^{\lambda_j} \phi_{jl} E \left[(\mathbf{W}^{(h)} \mathbf{X}_{t-s})' \mathbf{W}^{(l)} \mathbf{X}_{t-j} \right],$$

Dividindo ambos os membros por N , obtém-se a equação de Yule–Walker espaço-temporal

$$\gamma_{h0}(s) = \sum_{j=1}^p \sum_{l=0}^{\lambda_j} \phi_{jl} \gamma_{hl}(s-j), \quad s > 0.$$

A STPACF de ordem espacial λ é definida como o último coeficiente $\phi'_{k\lambda}$ obtido ao resolver o sistema de equações de Yule–Walker espaço-temporais para $l = 0, \dots, \lambda$ em cada defasamento temporal $k = 1, 2, \dots$

Note-se que o apóstrofo em ϕ' distingue os coeficientes da STPACF dos parâmetros do modelo STAR (ϕ_{jl}). Enquanto ϕ_{jl} descrevem a estrutura do processo gerador, $\phi'_{k\lambda}$ são quantidades derivadas usadas exclusivamente na fase de identificação.

Ainda, conforme Pfeifer & Deutsch (1980a), a tabela que se segue, apresenta as seguintes diretrizes que orientam a identificação das ordens de um modelo STARMA, a partir da análise dos gráficos das funções de autocorrelação espaço-temporal (STACF) e autocorrelação parcial espaço-temporal (STPACF).

Tabela 4.1: Propriedades teóricas das STACF e STPACF dos modelos STARMA.

Modelo	STACF	STPACF
$STAR(p_\lambda)$	Decai gradualmente em tempo e espaço	Corta após o defasamento p e ordem espacial λ
$STMA(q_m)$	Corta após defasamento q e ordem espacial m	Decai gradualmente em tempo e espaço
$STARMA(p_\lambda, q_m)$	Decai gradualmente em tempo e espaço	Decai gradualmente em tempo e espaço

Fonte: Adaptado de Pfeifer & Deutsch (1980a, pp. 39–41).

4.5.2-Estimação dos parâmetros dos modelos STARMA

Conforme Pfeifer & Deutsch (1980a), uma vez escolhido o modelo STARMA mais apropriado, procede-se à estimação dos parâmetros Φ e Θ pelo método da máxima verossimilhança, sob o pressuposto de que os resíduos seguem uma distribuição normal multivariada.

Assumindo que os erros são independentes no tempo e no espaço, com média nula e variância constante, tem-se:

$$E[\boldsymbol{\varepsilon}] = \mathbf{0}, \text{ e } Var(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_{NT},$$

onde N denota o número de localizações espaciais, T o número de instantes temporais, e

$\mathbf{I}_{NT}$ a matriz identidade de ordem NT .

A função de verossimilhança associada aos erros é dada por:

$$\begin{aligned} L(\boldsymbol{\varepsilon}|\boldsymbol{\Phi}, \boldsymbol{\Theta}, \sigma^2) &= (2\pi)^{-\frac{NT}{2}} |\sigma^2 \mathbf{I}_{NT}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}\right) \\ &= (2\pi)^{-\frac{NT}{2}} (\sigma^2)^{-\frac{NT}{2}} \exp\left(-\frac{S(\boldsymbol{\Phi}, \boldsymbol{\Theta})}{2\sigma^2}\right) \\ &= (2\pi\sigma^2)^{-\frac{NT}{2}} \exp\left(-\frac{S(\boldsymbol{\Phi}, \boldsymbol{\Theta})}{2\sigma^2}\right), \end{aligned} \quad (41)$$

onde

$$S(\boldsymbol{\Phi}, \boldsymbol{\Theta}) = \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon} = \sum_{i=1}^N \sum_{t=1}^T \varepsilon_i(t)^2$$

representa a soma dos quadrados dos resíduos.

Sabendo que os erros são aleatórios e não observáveis, enquanto a série X_t é observável, os resíduos podem ser estimados recursivamente a partir da expressão do modelo STARMA reescrita em função dos próprios resíduos. Deste modo, obtém-se:

$$\boldsymbol{\varepsilon}_t = \mathbf{X}_t - \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} \mathbf{W}^{(l)} \mathbf{X}_{t-k} + \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{kl} \mathbf{W}^{(l)} \boldsymbol{\varepsilon}_{t-k}$$

para $t = 1, 2, \dots, T$.

Os valores de $\mathbf{X}_t$ e $\boldsymbol{\varepsilon}_t$, para instantes anteriores ao início da amostra ($t \leq 0$) são desconhecidos, o que impede o cálculo direto dos resíduos nos primeiros períodos. Esse procedimento de inicialização é complexo e indispensável, pois sem ele não é possível obter

estimativas dos resíduos e, consequentemente, dos parâmetros Φ e Θ . Para ultrapassar esta limitação, assume-se que X_t e ε_t são nulos para todo $t \leq 0$.

Com esta simplificação, a função de verossimilhança passa a ser condicional aos primeiros valores observados, permitindo a estimação dos parâmetros Φ e Θ mesmo na ausência de observações prévias, isto é, dada por:

$$L(\Phi, \Theta, \sigma^2 | X) = (2\pi\sigma^2)^{-\frac{NT}{2}} \exp\left(-\frac{S_*(\Phi, \Theta)}{2\sigma^2}\right),$$

onde

$$S_*(\Phi, \Theta) = \hat{\varepsilon}'\hat{\varepsilon}$$

representa a soma condicional dos quadrados dos resíduos, calculada com $X_t = 0$ e $\varepsilon_t = 0$ para todo $t \leq 0$.

O estimador de máxima verossimilhança condicional de σ^2 é dado por

$$\hat{\sigma}_{ML}^2 = \frac{S_*(\hat{\Phi}, \hat{\Theta})}{NT},$$

sendo Φ , Θ os valores que minimizam a função $S_*(\Phi, \Theta)$ e NT o número total de observações espaço-temporais utilizadas na estimação.

Para séries longas, o procedimento condicional produz estimativas muito próximas das obtidas pelo método exato de máxima verossimilhança. Contudo, quando se trata de séries com poucas observações temporais, este método revela-se inadequado, pois as suas estimativas são enviesadas devido à substituição dos valores iniciais por zero.

Tendo em conta que, no caso do modelo STAR (isto é, sem componente de médias móveis), os estimadores de máxima verossimilhança condicional coincidem com os estimadores de mínimos quadrados, a estimação dos seus parâmetros pode ser feita com base na formulação da regressão linear, embora as hipóteses clássicas dessa teoria (ex.: regressores fixos e não correlacionados com os erros) não sejam satisfeitas (Mann & Wald, 1943).

No presente trabalho, as séries analisadas correspondem a registos médios mensais de temperatura ao longo de várias décadas, o que assegura um número suficientemente elevado de observações temporais, pelo que as estimativas obtidas pelo método condicional são, na

prática, aproximadamente iguais às que seriam obtidas com o método exato de máxima verossimilhança.

Reescrevendo o modelo STAR em forma de regressão linear, obtém-se:

$$Y = X\Phi + \varepsilon,$$

onde Φ é o vetor que contém todos os parâmetros ϕ_{jl} . Assim, os estimadores de mínimos quadrados condicionais são dados por:

$$\hat{\Phi} = (X'X)^{-1}X'Y.$$

A estimativa da variância dos resíduos é dada por:

$$\hat{\sigma}^2 = \frac{S_R(\hat{\Phi})}{NT - K},$$

onde $S_R(\hat{\Phi}) = (Y - X\hat{\Phi})'(Y - X\hat{\Phi})$ é a soma condicional dos quadrados dos resíduos.

Com base nos resultados assintóticos de Mann e Wald (1943), a estatística

$$\frac{(NT - K)\hat{\sigma}^2}{\sigma^2}$$

é aproximadamente qui-quadrado com $NT - K$ graus de liberdade, ou seja,

$$\frac{(NT - K)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{NT-K}.$$

Esta propriedade permite a construção de intervalos de confiança assintóticos para σ^2 cuja precisão aumenta com o tamanho da amostra.

Com base na teoria da regressão linear, cujos resultados são assintoticamente válidos para modelos STAR (Mann e Wald, 1943), é possível estabelecer intervalos de confiança para os parâmetros do modelo STAR utilizando a estatística

$$\frac{(\Phi - \hat{\Phi})'X'X(\Phi - \hat{\Phi})/K}{\frac{S_R(\hat{\Phi})}{NT - K}} \sim F(K, NT - K),$$

onde $F(K, NT - K)$ representa a distribuição de Fisher-Snedecor com K e $NT - K$ graus de liberdade e K o número de parâmetros estimados.

Importa salientar que todos os resultados de inferência estatística apresentados assentam em propriedades assintóticas, cuja validade se justifica no limite em que a dimensão da amostra tende para infinito. No caso dos modelos STAR, apesar de violarem as hipóteses da regressão linear clássica, nomeadamente o facto de os regressores serem aleatórios e estarem correlacionados com os erros, o trabalho de Mann e Wald (1943) assegura que, em amostras grandes, os estimadores por mínimos quadrados são consistentes e assintoticamente normais. Para os modelos STMA e STARMA, embora a estimação seja realizada por otimização não linear, a robustez das inferências assintóticas também depende do número de observações espaço-temporais. Quanto maior for a série, mais precisas e estáveis se tornam as inferências.

No presente estudo, tais aproximações são razoáveis, uma vez que as séries de temperatura média mensal abrangem várias décadas, o que confere robustez suficiente às inferências assintóticas utilizadas na estimação e no diagnóstico dos modelos considerados.

Região de confiança para os parâmetros

Dado que os modelos STMA e STARMA são não lineares, não é possível construir regiões de confiança exatas para parâmetros com base na teoria clássica de regressão linear. Entretanto, seguindo a abordagem proposta por Pfeifer & Deutsch (1980a), pode-se construir regiões de confiança aproximadas, por meio da expansão quadrática da função soma dos quadrados dos resíduos, $S(\Phi, \Theta)$, em torno do vetor de estimativas de mínimos quadrados da seguinte forma:

$$S(\Phi, \Theta) = S(\delta) \approx S(\hat{\delta}) + (\delta - \hat{\delta})' Q (\delta - \hat{\delta}),$$

onde $\delta' = (\Phi', \Theta')$ é o vetor de todos os parâmetros do modelo e Q é uma aproximação da matriz Hessiana de $S(\delta)$, avaliada em $\hat{\delta}$, isto é,

$$Q = \frac{1}{2} \left[\frac{\partial^2 S(\delta)}{\partial \delta_i \partial \delta_j} \right]_{\hat{\delta}}$$

para $i = 1, 2, \dots, K$, $j = 1, 2, \dots, K$, sendo K a dimensão de δ , ou seja, o número de parâmetros. Sob a hipótese de que o modelo está adequadamente ajustado, ou seja, os resíduos se comportam como ruído branco, o termo de segunda ordem envolvendo as derivadas dos erros pode ser negligenciado (Pfeifer & Deutsch, 1980a), resultando na

aproximação $Q \approx X'X$, onde X é a matriz jacobiana das derivadas dos resíduos em ordem aos parâmetros, avaliadas em $\hat{\delta}$.

A região de confiança aproximada a $100(1 - \alpha)\%$ é então definida pelo conjunto de valores δ que satisfazem

$$S(\delta) \leq S(\hat{\delta}) + \frac{K}{TN - K} S(\hat{\delta}) F_{K, TN-K; \alpha},$$

com K o número de parâmetros, T o número de observações temporais, N o número de séries espaciais e $F_{K, TN-K; \alpha}$ o quantil de ordem α da distribuição F de Snedecor com K e $TN - K$ graus de liberdade.

4.5.3-Diagnóstico dos modelos STARMA

Com o modelo candidato identificado e os seus parâmetros já estimados, inicia-se a fase de validação diagnóstica, cujo objetivo é avaliar se o modelo representa adequadamente a estrutura espaço-temporal dos dados observados. Esta validação contempla a análise dos resíduos e a significância estatística dos parâmetros estimados.

Análise dos resíduos

A análise dos resíduos tem por objetivo verificar se estes se comportam como ruído branco gaussiano, ou seja, os resíduos devem apresentar:

- Média nula;
- Variância constante;
- Covariâncias nulas entre observações em diferentes locais ou momentos; e
- Distribuição normal.

Esta análise pode ser realizada através do cálculo das funções de autocorrelação e autocorrelação parcial espaço-temporais amostrais dos resíduos e compará-las com a sua variância. Conforme demonstrado por Pfeifer & Deutsch (1980a), sob a hipótese de ruído branco puro, a variância do estimador da autocorrelação espaço-temporal $\hat{\rho}_{l_0}(s)$ satisfaz

$$Var(\hat{\rho}_{l_0}(s)) \approx \frac{1}{N(T - s)}.$$

Adicionalmente, a literatura propõe testes formais para a validação dos pressupostos dos resíduos na dimensão espacial. Pfeifer e Deutsch (1980c) apresentam, entre outros, o teste de independência, que avalia a ausência de covariâncias espaciais entre os resíduos, e o teste de esfericidade, que permite verificar simultaneamente a independência espacial e a homogeneidade da variância residual.

Contudo, no presente trabalho, optou-se por seguir a prática habitual na literatura, baseando-se na inspeção visual da função de autocorrelação espaço-temporal amostral (STACF) dos resíduos. Assim, consideram-se os resíduos compatíveis com a hipótese de ruído branco sempre que a STACF não apresente picos significativos, isto é, quando os coeficientes de autocorrelação se mantêm próximos de zero para a maioria das defasagens temporais e espaciais. Complementarmente, a análise gráfica da série de resíduos permite identificar eventuais padrões sistemáticos que possam indiciar violações dos pressupostos de média nula, variância constante ou normalidade.

Significância estatística dos parâmetros estimados

A avaliação da significância estatística dos parâmetros estimados pode ser realizada utilizando duas abordagens: testando todos os coeficientes em conjunto ou avaliando cada coeficiente individualmente.

A primeira abordagem baseia-se na região de confiança anteriormente apresentada e consiste em submeter todos os coeficientes do modelo a um teste conjunto. Neste contexto, testam-se as seguintes hipóteses:

$$H_0: \Phi = \Theta = \mathbf{0} \quad \text{vs} \quad H_1: (\Phi, \Theta) \neq (\mathbf{0}, \mathbf{0}),$$

isto é, a hipótese nula assume que todos os coeficientes do modelo, quer espaciais, quer temporais, são em conjunto iguais a zero, enquanto a hipótese alternativa admite que pelo menos um deles seja diferente de zero.

A segunda abordagem, mais útil em termos práticos, consiste em testar a significância de cada coeficiente individualmente, mantendo os restantes parâmetros livres. Ou seja, são testadas as seguintes hipóteses:

$$H_0: \delta_k = 0 \quad \text{vs} \quad H_1: \delta_k \neq 0,$$

onde δ_k representa o parâmetro genérico (por exemplo ϕ_{kl} ou θ_{kl}).

Seja $\hat{\delta}$ a estimativa de mínimos quadrados do vetor completo de coeficientes a serem testados e $\hat{\delta}^*$ a estimativa obtida quando δ_k é fixado a zero. Então, sob H_0 , a estatística é dada por:

$$F = \frac{(TN - K)[S_*(\hat{\delta}^*) - S_*(\hat{\delta})]}{S_*(\hat{\delta})} \sim F_{1, TN-K}.$$

Caso um coeficiente não seja estatisticamente significativo, deve ser removido do modelo, e a versão resultante passa a constituir um novo candidato. Este procedimento repete-se iterativamente até se obter um modelo cujos coeficientes são todos estatisticamente significativos e cujos resíduos se comportam como ruído branco.

4.5.4-Seleção do modelo final

Nos casos em que mais do que um modelo satisfaz todas as condições descritas na secção anterior, diagnóstico dos modelos STARMA, recorre-se a critérios de informação para a seleção do modelo final. Optou-se pelo ***Critério de Informação Bayesiano (BIC)*** como critério de comparação entre modelos candidatos, pois, segundo Hyndman e Athanasopoulos (2021), o *BIC* penaliza mais fortemente a inclusão de parâmetros do que o *AICc*. Assim, no contexto espaço-temporal, com T períodos e N localizações, considera-se $n = TN$, sendo o critério definido por:

$$BIC = -2\log(L) + k\log(TN). \quad (42)$$

4.6- Previsão com modelos STARMA

Após a seleção do modelo STARMA que melhor descreve a estrutura espaço-temporal, com base nos critérios *AICc* e *BIC*, condicionada à ausência de autocorrelação residual, procede-se à etapa de previsão.

As previsões *h-passos* à frente, denotadas por $\hat{X}_{t+h|t}$, são dadas pela esperança condicional de X_{t+h} dada a informação disponível até ao instante t , isto é, com base nas observações passadas $\{X_t, X_{t-1}, X_{t-2}, \dots\}$. Assumindo que $\{\varepsilon_t\}$ constitui um ruído branco espaço-temporal, tem-se, $E(\varepsilon_{t+h}|X_t, X_{t-1}, X_{t-2}, \dots) = 0$ para $h \geq 1$.

Para $h = 1$, todos os termos envolvidos são conhecidos e a previsão é dada por:

$$\hat{\mathbf{X}}_{t+1|t} = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{k\ell} \mathbf{W}^{(l)} \mathbf{X}_{t+1-k} - \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{k\ell} \mathbf{W}^{(l)} \boldsymbol{\varepsilon}_{t+1-k}. \quad (43)$$

Para $h > 1$, a previsão é obtida de forma recursiva. Os erros futuros $\boldsymbol{\varepsilon}_{t+h-k}$ (com $h - k \geq 1$) são substituídos por $\mathbf{0}$ e os valores futuros $\mathbf{X}_{t+h-k}$ não disponíveis são substituídos pelas previsões correspondentes já obtidas, $\hat{\mathbf{X}}_{t+h-k|t}$.

Adicionalmente, é possível construir intervalos de previsão para as previsões, de modo a quantificar a incerteza associada às mesmas. Assumindo que os resíduos seguem uma distribuição normal espaço-temporal, o intervalo de previsão a $(1 - \alpha)\%$ para a previsão h -passos à frente é dado por:

$$\hat{\mathbf{X}}_{t+h|t} \pm z_{1-\alpha/2} \hat{\sigma}_h \quad (44)$$

onde:

- $\hat{\mathbf{X}}_{t+h|t}$ é a previsão pontual;
- $z_{1-\alpha/2}$ é o quantil da distribuição normal padrão correspondente ao nível de confiança pré-definido;
- $\hat{\sigma}_h$ é o desvio padrão do erro de previsão.

Este procedimento não é apresentado de forma explícita por Pfeifer & Deutsch (1980a), entretanto é consequência direta da estrutura recursiva do modelo STARMA e da hipótese de ruído branco. Na prática, trata-se de uma extensão da abordagem de previsão dos modelos ARMA univariados ao contexto multivariado e espaço-temporal.

5- Capítulo V – Modelação Espaço–Temporal da Temperatura do Mar com Modelos STARMA

5.1- Introdução à aplicação espaço–temporal

Após a modelação da dependência temporal das séries históricas de temperaturas médias mensais, pretende-se agora explorar de forma conjunta a dependência no tempo e no espaço, considerando, numa primeira fase, apenas as temperaturas registadas a um metro de profundidade em todas as regiões analisadas. Esta integração tem como objetivo não apenas compreender as dinâmicas temporais das temperaturas em análise, mas também apreender a sua variabilidade espacial.

Essa abordagem requer, numa primeira etapa, a definição das matrizes de pesos $W^{(0)}$, $W^{(1)}$, $W^{(2)}$, ..., $W^{(r)}$ em que r representa a ordem espacial máxima considerada. A matriz $W^{(0)}$ representa a identidade, ou seja, a própria localização, enquanto as matrizes $W^{(1)}$, $W^{(2)}$, ..., $W^{(r)}$ descrevem relações de vizinhança sucessivas entre as diferentes localizações (Pfeifer & Deutsch, 1980a).

Tendo estabelecido os fundamentos teóricos dos modelos STARMA no Capítulo IV, este capítulo dedica-se à sua aplicação empírica. A análise organiza-se da seguinte forma: (i) construção da matriz de pesos espaciais com base nas distâncias geodésicas; (ii) identificação dos modelos candidatos com base nas funções STACF e STPACF; (v) seleção do modelo final com base no critério BIC, na significância dos parâmetros e na análise residual; (v) previsões geradas pelo modelo STARMA e (vi) avaliação da qualidade das previsões.

5.2- Construção da matriz de pesos espaciais

Para a escolha do valor de r , adotou-se a abordagem clássica de Pfeifer e Deutsch (1980b) para modelos STARMA. Ao focarem em modelos de primeira ordem, os autores limitam as defasagens espaciais às ordens 0 e 1, correspondentes às matrizes de pesos $W^{(0)}$ e $W^{(1)}$, uma restrição que visa evitar a sobreparametrização e simplificar a estimação do modelo.

A matriz $W^{(0)}$ é a matriz identidade de ordem N , expressa por:

$$W^{(0)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

A matriz de pesos $W^{(1)}$ foi definida a partir das distâncias, em quilómetros, entre as diferentes regiões em estudo, apresentadas na [Tabela 5.1](#).

Tabela 5.1: Distância (em quilómetros) entre as regiões em análise.

Localização	Nazaré	Faro	São Miguel	Praia
Nazaré	0	361,8	1450,7	3081,3
Faro	361,8	0	1559,6	2839,6
São Miguel	1450,7	1559,6	0	2527,9
Praia	3081,3	2839,6	2527,9	0

Fonte: Elaboração própria.

A matriz $W^{(1)}$, de primeira ordem, representa as relações de vizinhança imediata entre as regiões. Quanto menor for a distância entre duas regiões, maior será o peso atribuído à interação entre elas. Esta matriz foi construída a partir do inverso das distâncias, em quilómetros, sendo normalizada por linha de forma a garantir a comparabilidade entre os valores atribuídos, isto é, a soma dos pesos em cada linha é igual a 1.

$$W^{(1)} = \begin{bmatrix} 0 & 0,73 & 0,18 & 0,09 \\ 0,74 & 0 & 0,17 & 0,09 \\ 0,4 & 0,37 & 0 & 0,23 \\ 0,3 & 0,33 & 0,37 & 0 \end{bmatrix}$$

5.3- Identificação da estrutura do modelo STARMA

Para a identificação dos parâmetros do modelo *STARMA*, recorreu-se às funções de autocorrelação espaço-temporal (*STACF*) e de autocorrelação espaço-temporal parcial (*STPACF*), após se ter assegurado a estacionaridade das séries mediante diferenciação sazonal, conforme discutido na Secção 4.4.3.

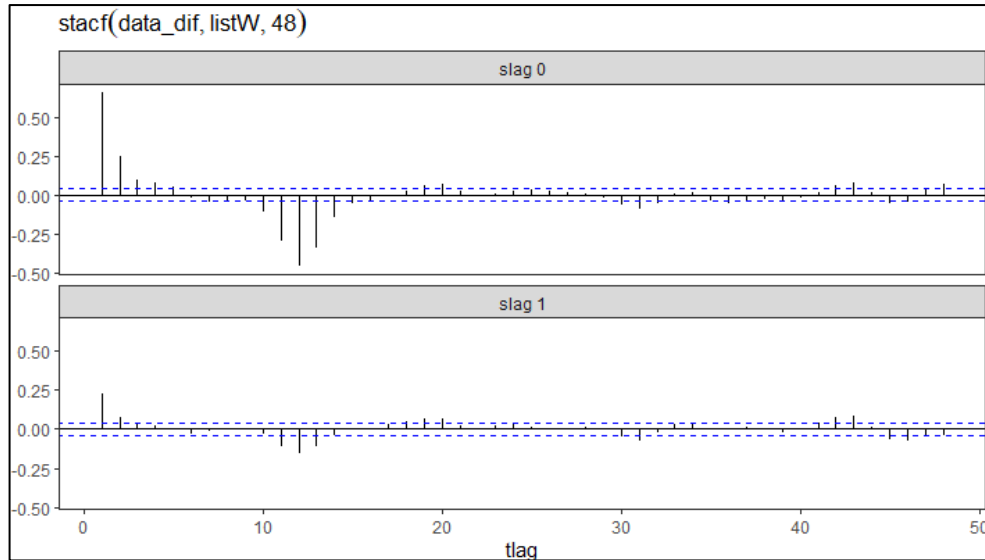


Figura 5.1: funções de autocorrelação espaço-temporal (STACF).

Fonte: Elaboração própria.

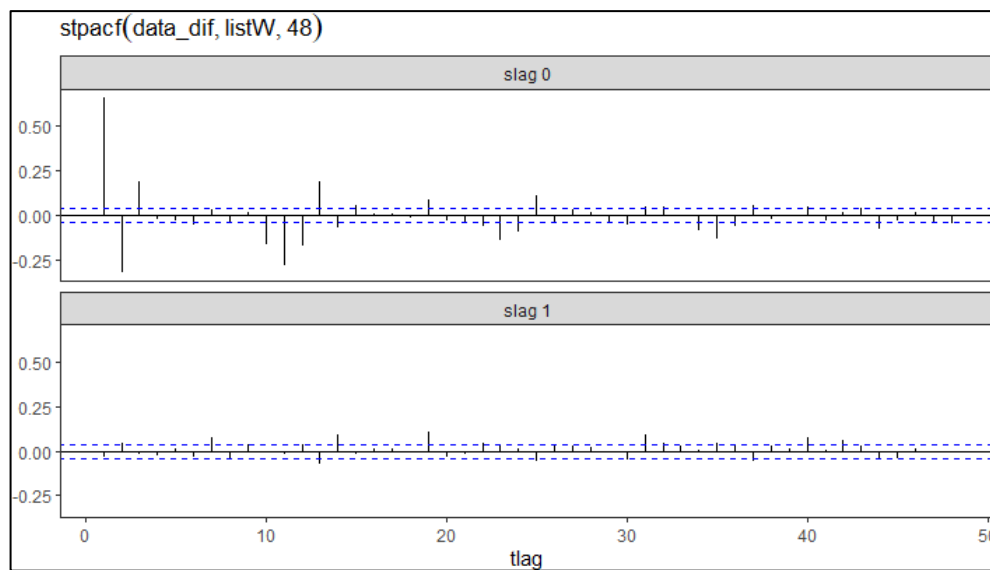


Figura 5.2: funções de autocorrelação parcial espaço-temporal (STPACF).

Fonte: Elaboração própria.

A análise do gráfico da função de autocorrelação espaço-temporal (STACF), sugere um decaimento relativamente rápido tanto na dimensão temporal como espacial, indicando que não há evidências fortes sobre a existência de componentes de médias móveis. No que se refere à análise da função de autocorrelação parcial espaço-temporal (STPACF), constata-se que o pico mais expressivo ocorre no desfasamento 1, sugerindo uma estrutura autoregressiva predominante de primeira ordem ($AR(1)$). No entanto, há picos com alguma

relevância estatística em desfasamentos superiores, ainda que menores, o que levou à consideração de modelos de ordem 2 e 3, com o intuito de captar melhor a estrutura de dependência residual.

A análise conjunta das funções STACF e STPACF indica, adicionalmente, a presença de alguma dependência espacial, embora de intensidade relativamente fraca. Tal conclusão decorre do facto de surgirem apenas algumas autocorrelações estatisticamente significativas nos desfasamentos espaciais (*slag 1*) e com baixa magnitude. Estes picos sugerem que existe algum grau de associação entre as regiões vizinhas, mas não suficiente para caraterizar uma estrutura espacial dominante no processo.

Por outro lado, a análise das referidas funções indica ainda que não há evidência estatística de componentes sazonais remanescentes relevantes. Embora se observem autocorrelações significativas em torno do desfasamento correspondente ao período sazonal ($lag = 12$), estas não se repetem nos seus múltiplos (24, 36, ...), nem exibem um padrão sistemático que sustente a introdução de termos AR ou MA sazonais no modelo.

Assim, o modelo ajustado às séries sazonalmente diferenciadas será designado por:

$$STARMA(p_{\lambda}, q_m)$$

Onde:

- p_{λ} : ordem da componente autoregressiva (AR) temporal, com dependência espacial até a ordem λ ;
- q_m : ordem da componente de médias móveis (MA) temporal, com dependência espacial até a ordem m .

Dado que as funções STACF e STPACF não apresentaram padrões claros, como cortes abruptos ou decaimentos regulares, que permitem identificar diretamente a ordem dos componentes autoregressivos ou de médias móveis, serão testados vários modelos candidatos. A seleção do modelo mais adequado basear-se-á no critério BIC e na análise da estrutura de autocorrelação dos resíduos, tanto no domínio temporal como espacial.

5.4- Modelos candidatos e critérios de seleção

A tabela que se segue apresenta os modelos candidatos identificados:

Tabela 5.2: Identificação dos modelos STARMA candidatos.

Modelos Candidatos	Descrição
M1: STARMA(1 ₁ , 0)	AR(1) temporal + AR(1) espacial
M2: STARMA(2 ₁ , 0)	AR(2) temporal + AR(1) espacial
M3: STARMA(3 ₁ , 0)	AR(3) temporal + AR(1) espacial
M4: STARMA(1 ₁ , 1)	AR(1) temporal + AR(1) espacial + MA(1) temporal
M5: STARMA(2 ₁ , 1)	AR(2) temporal + AR(1) espacial + MA(1) temporal
M6: STARMA(1 ₁ , 1 ₁)	AR(1) temporal + AR(1) espacial + MA(1) temporal + MA(1) espacial
M7: STARMA(2 ₁ , 1 ₁)	AR(2) temporal + AR(1) espacial + MA(1) temporal + MA(1) espacial
M8: STARMA(3 ₁ , 1 ₁)	AR(3) temporal + AR(1) espacial + MA(1) temporal + MA(1) espacial
M9: STARMA(3 ₁ , 2 ₁)	AR(3) temporal + AR(1) espacial + MA(2) temporal + MA(1) espacial

Fonte: Elaboração própria.

Em seguida, procedeu-se à análise detalhada de cada um dos modelos candidatos. Para cada modelo, avaliaram-se os valores do Critério de Informação Bayesiano ([Tabela 5.3](#)) e a significância estatística dos parâmetros estimados ([Tabelas 5.4 a 5.12](#)). Adicionalmente, inspecionaram-se as funções de autocorrelação espaço-temporal dos resíduos, com o objetivo de verificar se a estrutura de dependência foi adequadamente capturada pelo modelo.

Tabela 5.3: BIC e variância dos modelos candidatos.

Modelo	BIC	$\hat{\sigma}^2$
M1	3492,838	0,2243283
M2	3214,644	0,2013556
M3	3132,519	0,1960210
M4	3144,298	0,1973198
M5	3148,084	0,1972866
M6	3155,491	0,1985161
M7	4273,823	0,5741259
M8	3151,789	0,2060957
M9	3159,341	0,2051762

Fonte: Elaboração própria.

Tabela 5.4: Estimativas dos parâmetros do modelo M1.

Parâmetro	Estimativa	Erro Padrão	Valor-t	Valor-p
ϕ_{10}	0,667844	0,016056	41,5950	<2e-16
ϕ_{11}	-0,037168	0,021055	-1,7653	0,07763

Fonte: Elaboração própria.

Tabela 5.5: Estimativas dos parâmetros do modelo M2.

Parâmetro	Estimativa	Erro Padrão	Valor-t	Valor-p
ϕ_{10}	0,879363	0,019550	44,9798	<2e-16
ϕ_{11}	-0,038072	0,019950	-1,9084	0,05645
ϕ_{20}	-0,322157	0,018687	-17,2400	<2e-16

Fonte: Elaboração própria.

Tabela 5.6: Estimativas dos parâmetros do modelo M3.

Parâmetro	Estimativa	Erro Padrão	Valor-t	Valor-p
ϕ_{10}	0,937530	0,020263	46,2681	<2e-16
ϕ_{11}	-0,037758	0,019698	-1,9169	0,05537
ϕ_{20}	-0,481536	0,024983	-19,2748	<2e-16
ϕ_{30}	0,184834	0,019518	9,4700	<2e-16

Fonte: Elaboração própria.

Tabela 5.7: Estimativas dos parâmetros do modelo M4.

Parâmetro	Estimativa	Erro Padrão	Valor-t	Valor-p
ϕ_{10}	0,468601	0,018376	25,5004	<2e-16
ϕ_{11}	-0,029671	0,019751	-1,5022	0,1332
θ_{10}	0,468758	0,024780	18,9171	<2e-16

Fonte: Elaboração própria.

Tabela 5.8: Estimativas dos parâmetros do modelo M5.

Parâmetro	Estimativa	Erro Padrão	Valor-t	Valor-p
ϕ_{10}	0,568719	0,041032	13,8604	< 2,2e-16
ϕ_{11}	-0,034323	0,019752	-1,7377	0,082382
ϕ_{20}	-0,090580	0,032706	-2,7696	0,005653
θ_{10}	0,377611	0,043982	8,5856	< 2,2e-16

Fonte: Elaboração própria.

Tabela 5.9: Estimativas dos parâmetros do modelo M6.

Parâmetro	Estimativa	Erro Padrão	Valor-t	Valor-p
ϕ_{10}	0,4615400	0,0190432	24,2364	< 2e-16
ϕ_{11}	-0,0033153	0,0247400	-0,1340	0,89341
θ_{10}	0,4862507	0,0275665	17,6392	< 2e-16
θ_{11}	-0,0665277	0,0354523	-1,8765	0,06069

Fonte: Elaboração própria.

Tabela 5.10: Estimativas dos parâmetros do modelo M7.

Parâmetro	Estimativa	Erro Padrão	Valor-t	Valor-p
ϕ_{10}	0,5702263	0,0704214	8,0973	8,525e-16
ϕ_{11}	-0,0060995	0,0419412	-0,1454	0,88438
ϕ_{20}	-0,0984505	0,0559220	-1,7605	0,07844
θ_{10}	0,3886961	0,0775293	5,0135	5,702e-07
θ_{11}	-0,0699759	0,0604046	-1,1585	0,24678

Fonte: Elaboração própria.

Tabela 5.11: Estimativas dos parâmetros do modelo M8.

Parâmetro	Estimativa	Erro Padrão	Valor-t	Valor-p
ϕ_{10}	0,9339558	0,0904959	10,3204	< 2,2e-16
ϕ_{11}	-0,0081519	0,0251948	-0,3236	0,74630
ϕ_{20}	-0,4869855	0,0921517	-5,2846	1,364e-07
ϕ_{30}	0,1869964	0,0421862	4,4326	9,693e-06
θ_{10}	0,0176075	0,0941692	0,1870	0,85169
θ_{11}	-0,0717262	0,0364697	-1,9667	0,04932

Fonte: Elaboração própria.

Tabela 5.12: Estimativas dos parâmetros do modelo M9.

Parâmetro	Estimativa	Erro Padrão	Valor-t	Valor-p
ϕ_{10}	0,9497319	0,1419272	6,6917	2,690e-11
ϕ_{11}	-0,0061387	0,0251643	-0,2439	0,80729
ϕ_{20}	-0,4926703	0,1034595	-4,7620	2,022e-06
ϕ_{30}	0,1857873	0,0408751	4,5452	5,737e-06
θ_{10}	0,0028356	0,1454599	0,0195	0,98445
θ_{11}	-0,0771323	0,0366692	-2,1035	0,03552
θ_{20}	-0,0128113	0,0725684	-0,1765	0,85988

Fonte: Elaboração própria.

Ao analisar as [Tabelas 5.3 a 5.12](#), constata-se que o modelo $STARMA(3_1, 0)$ (M3) é o mais adequado, pois apresenta o menor BIC ($BIC = 3132,52$) e a menor variância residual ($\hat{\sigma}^2 = 0,1960$). A diferença de BIC face ao segundo melhor modelo com componente espacial (M4) é superior a 11 unidades. Embora a estimativa do parâmetro ϕ_{11} apresente um valor- p ligeiramente superior a 5% ($p = 0,055$) e o modelo $AR(3)$ puramente temporal (equivalente ao M3 com $\phi_{11} = 0$) tenha um BIC marginalmente inferior (3128,03), a inclusão de ϕ_{11} é justificada por três razões complementares:

(i) evidência empírica de dependência espacial residual, ainda que fraca; (ii) preservação da capacidade explicativa da estrutura espaço-temporal, que se perderia completamente no modelo puramente temporal; e (iii) expectativa teórica a priori de dependência entre localidades geograficamente próximas. A adequação final do modelo será validada pela análise das funções de autocorrelação espaço-temporal dos resíduos.

5.5- Diagnóstico, validação e seleção do modelo final

Para a análise dos resíduos, recorreu-se à representação gráfica da função de autocorrelação espaço-temporal (STACF) dos resíduos, que permite avaliar estatisticamente a presença (ou ausência) de correlação serial residual, garantindo assim a adequação final do modelo selecionado.

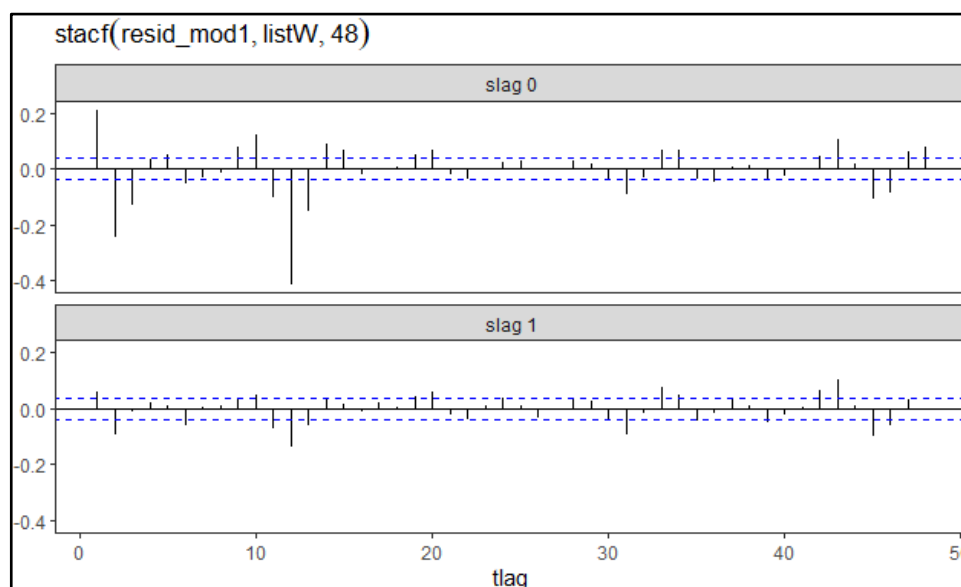


Figura 5.3: Função de autocorrelação espaço-temporal dos resíduos do Modelo M1: $STARMA(1_1,0)$.

Fonte: Elaboração própria.

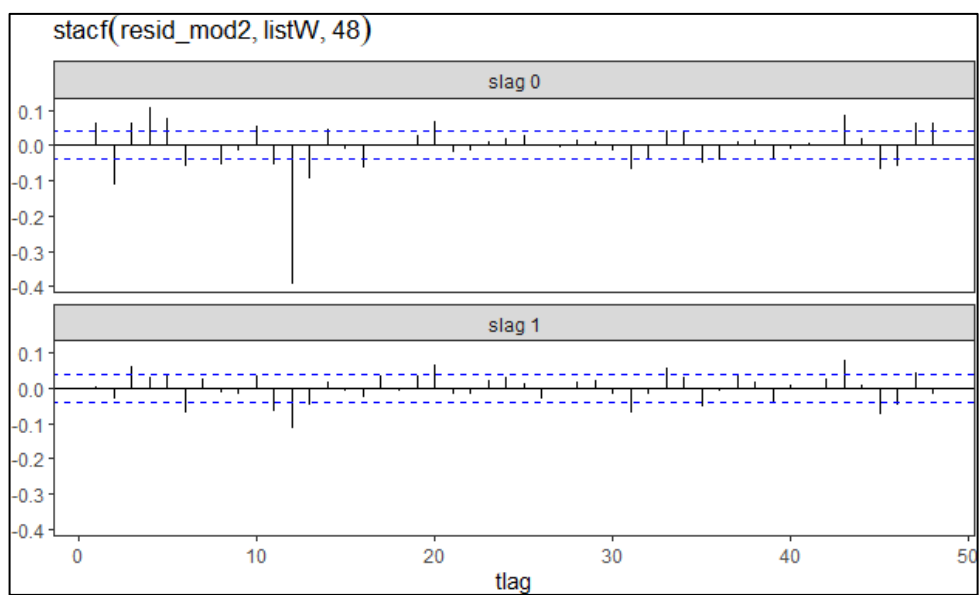


Figura 5.4: Função de autocorrelação espaço-temporal dos resíduos do Modelo M2: STARMA(2,0).

Fonte: Elaboração própria.

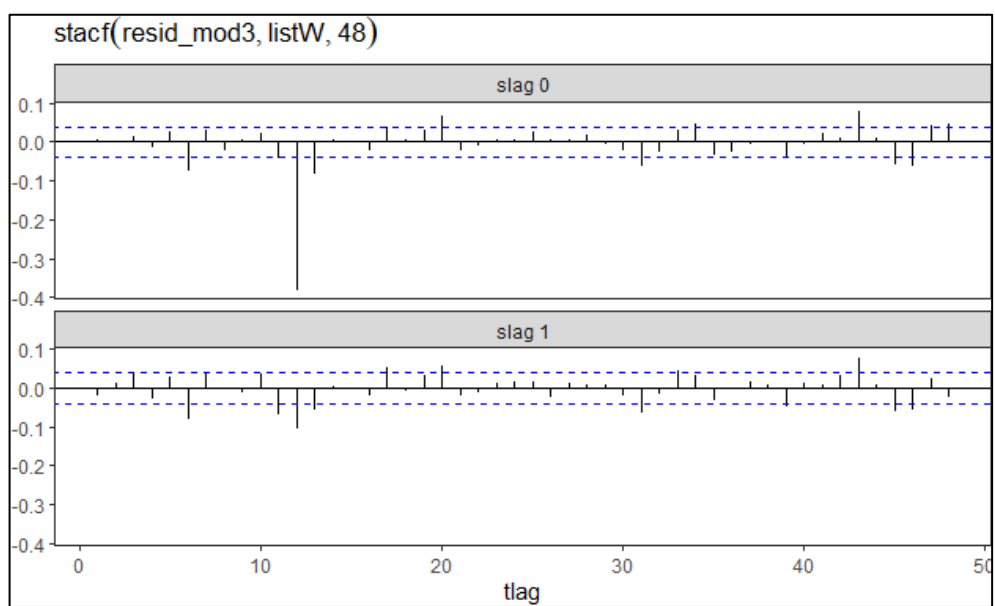


Figura 5.5: Função de autocorrelação espaço-temporal dos resíduos do Modelo M3: STARMA(3,0).

Fonte: Elaboração própria.

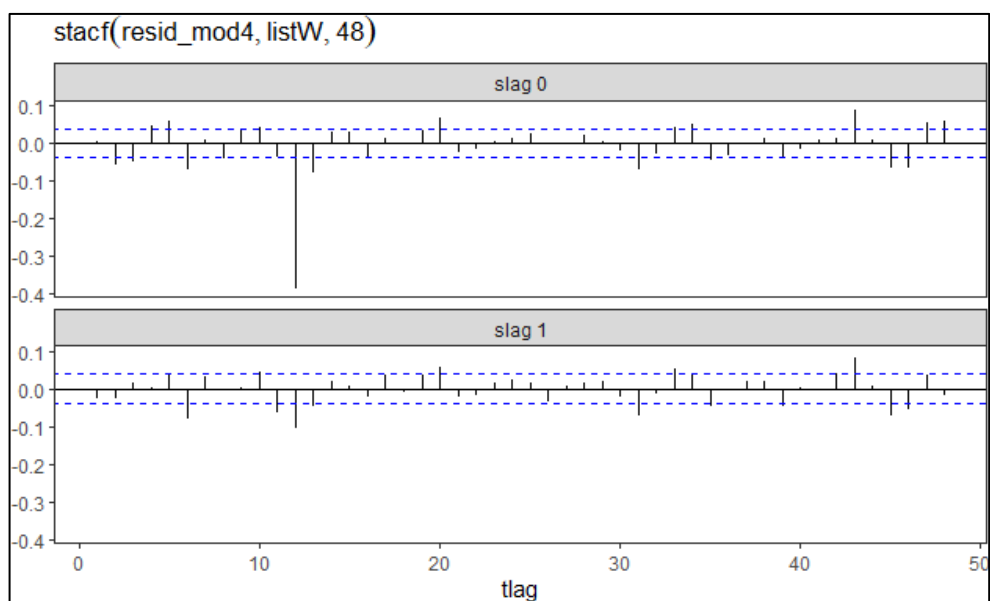


Figura 5.6: Função de autocorrelação espaço-temporal dos resíduos do Modelo M4: STARMA(1,1).

Fonte: Elaboração própria.

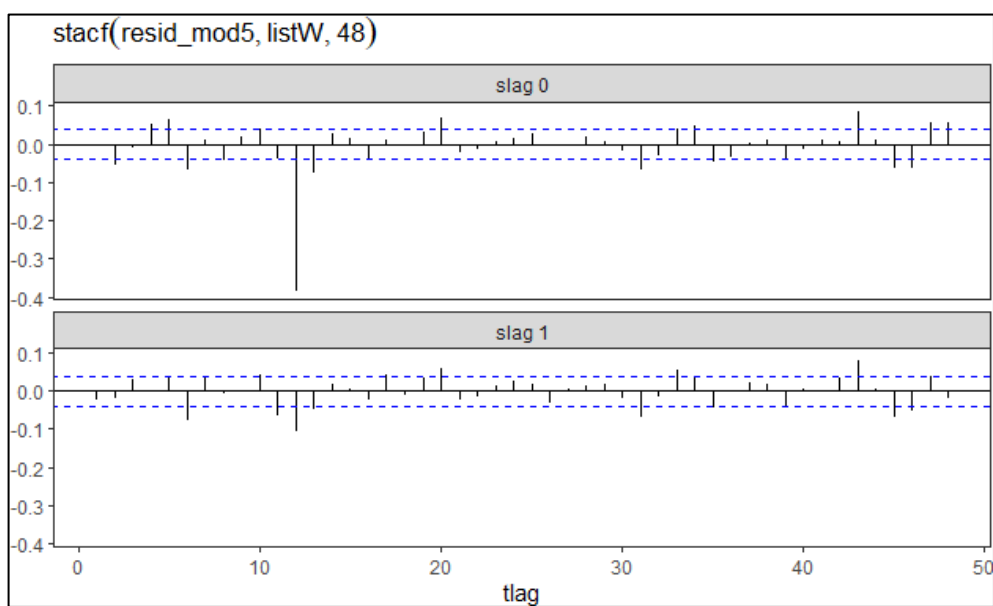


Figura 5.7: Função de autocorrelação espaço-temporal dos resíduos do Modelo M5: STARMA(2,1).

Fonte: Elaboração própria.

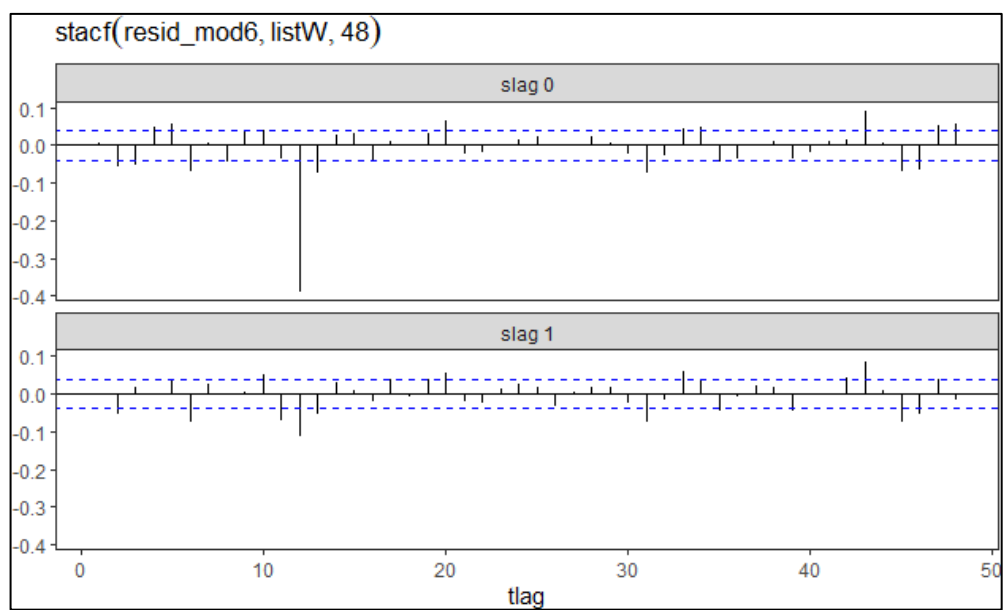


Figura 5.8: Função de autocorrelação espaço-temporal dos resíduos do Modelo M6: STARMA(1,1₁).

Fonte: Elaboração própria.

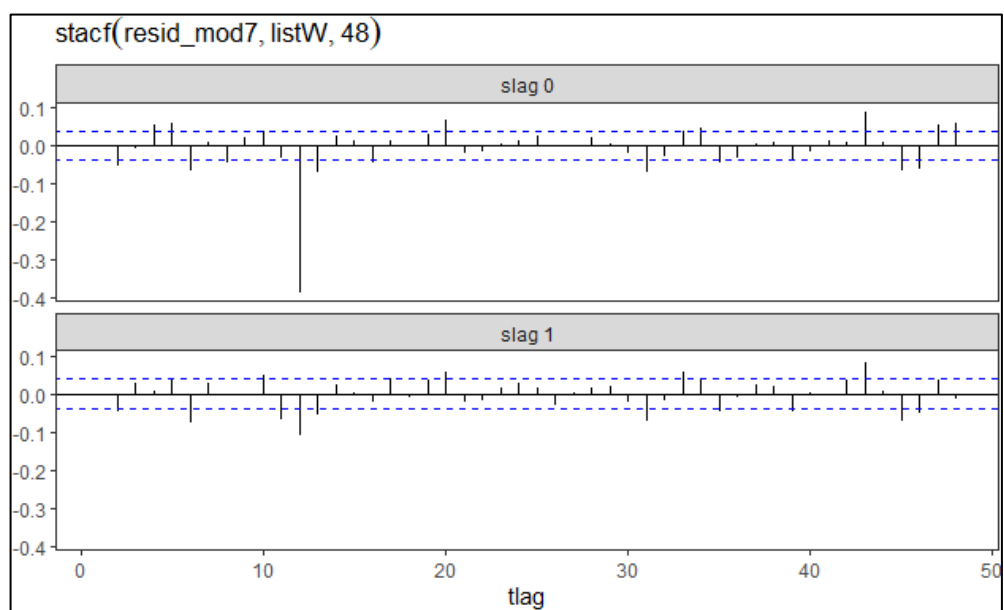


Figura 5.9: Função de autocorrelação espaço-temporal dos resíduos do Modelo M7: STARMA(2,1₁).

Fonte: Elaboração própria.

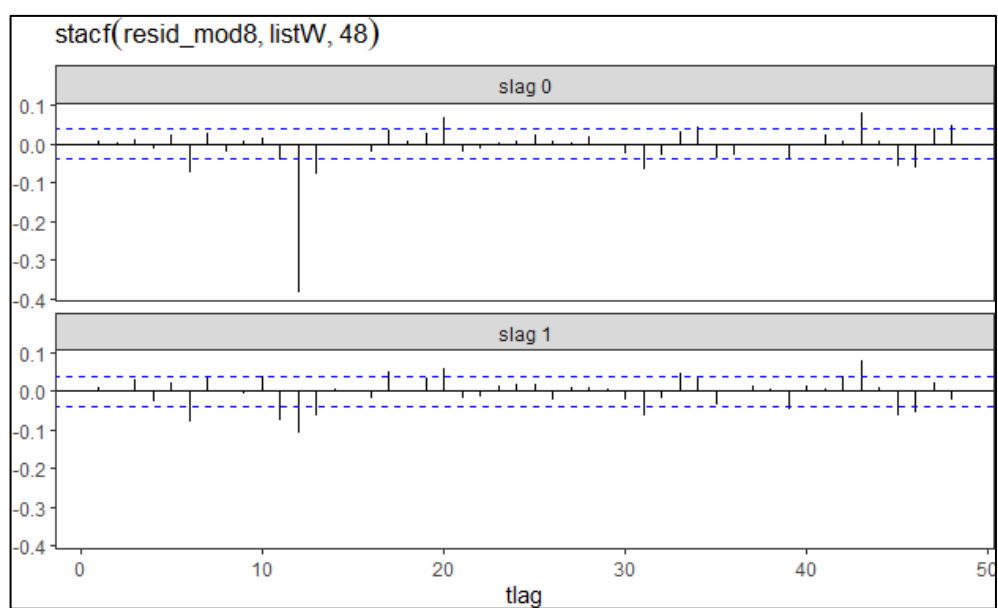


Figura 5.10: Função de autocorrelação espaço-temporal dos resíduos do Modelo M8: STARMA(3,1₁).

Fonte: Elaboração própria.

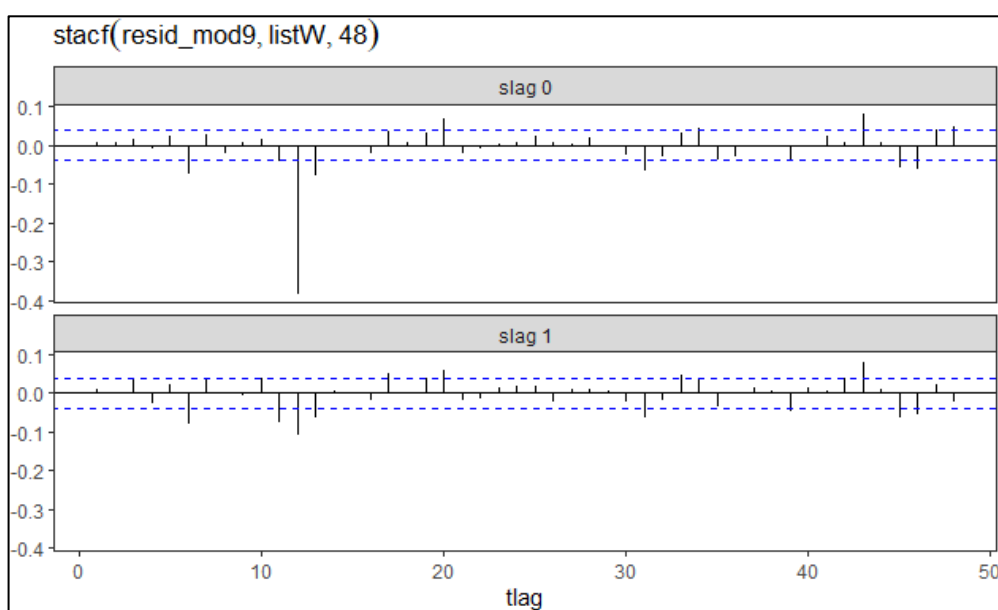


Figura 5.11: Função de autocorrelação espaço-temporal dos resíduos do Modelo M9: STARMA(3,2₁).

Fonte: Elaboração própria.

Da análise das [Figuras 5.3 a 5.11](#), constata-se que a grande maioria das autocorrelações dos resíduos do modelo M3 ([Figura 5.5](#)) se encontra dentro dos limites de confiança, tanto na dimensão temporal (*slag* 0) como na espacial (*slag* 1). Este comportamento é consistente com ruído branco espaço-temporal, indicando que o modelo conseguiu captar adequadamente a estrutura de dependência presente na série.

Além disso, quando comparado com modelos mais simples (por exemplo, M1) ou excessivamente complexos (como M7, com BIC= 4273) não se observa uma melhoria visual na função STACF dos resíduos que justificasse a sua escolha face ao modelo M3. Assim, confirma-se que o modelo $STARMA(3_1, 0)$ apresenta o melhor compromisso entre ajuste e parcimónia, sendo o modelo final selecionado.

5.6- Equação do modelo final selecionado

O modelo $STARMA(3_1, 0)$ foi ajustado à série sazonalmente diferenciada:

$$Y_t = (1 - B^{12})X_t = X_t - X_{t-12},$$

A equação do modelo ajustado é:

$$Y_t = \phi_{10}Y_{t-1} + \phi_{11}W^{(1)}Y_{t-1} + \phi_{20}Y_{t-2} + \phi_{30}Y_{t-3} + \varepsilon_t.$$

Com base nas estimativas dos parâmetros ([Tabela 5.6](#)), obtém-se:

$$\begin{aligned} \hat{Y}_t = & 0,93753Y_{t-1} - 0,037758W^{(1)}Y_{t-1} - 0,481536Y_{t-2} \\ & + 0,184834Y_{t-3}. \end{aligned} \quad (45)$$

5.7- Previsões utilizando o modelo STARMA selecionado

Após a seleção do modelo STARMA que melhor descreve a estrutura espaço-temporal das quatro séries temporais de temperatura mensal da superfície do mar, relativas às localizações de Nazaré, Faro e São Miguel (Portugal), e Praia (Cabo Verde), procedeu-se à geração de previsões pontuais para o horizonte de 12 meses, compreendendo o período de julho de 2023 a junho de 2024.

As previsões foram obtidas recursivamente, seguindo-se a reversão da diferenciação sazonal de ordem 12 ($X_t = X_{t-12} + Y_t$), de modo a restituir os valores na escala original das

temperaturas mensais (em °C). A tabela que se segue apresenta os valores previstos para cada localização no período de julho de 2023 a junho de 2024, com base no modelo.

Tabela 5.13: Previsões pontuais da temperatura mensal da superfície do mar (°C), julho 2023 – junho 2024.

Mês	Nazaré	Faro	São Miguel	Praia
Julho 2023	19,17	21,85	22,07	24,81
Agosto 2023	19,82	22,81	24,07	27,50
Setembro 2023	19,76	23,14	24,67	28,62
Outubro 2023	19,99	23,35	24,38	28,04
Novembro 2023	18,28	21,06	19,15	26,65
Dezembro 2023	16,51	18,62	16,70	24,98
Janeiro 2024	15,51	17,32	16,56	24,28
Fevereiro 2024	14,99	16,87	16,00	23,31
Março 2024	14,94	17,04	16,17	23,74
Abril 2024	15,93	17,87	16,99	24,36
Mai 2024	17,29	19,37	18,36	24,57
Junho 2024	18,84	21,08	20,29	24,57

Fonte: Elaboração própria.

Verifica-se, com base na análise exploratória das séries realizada no Capítulo III, que as previsões geradas pelo modelo $STARMA(3_1, 0)$, apresentadas na [Tabela 5.13](#), são coerentes, plausíveis e alinhadas com a dinâmica observada ao longo das últimas décadas. As previsões reproduzem fielmente a componente sazonal das séries, evidenciando máximos típicos nos meses de setembro/outubro e mínimos em fevereiro/março, comportamento característico da evolução histórica das temperaturas da superfície do mar nas quatro localizações.

Mantém-se igualmente a hierarquia térmica espacial verificada ao longo dos últimos anos: Praia (Cabo Verde) continua a apresentar as temperaturas mensais mais elevadas, seguindo-se São Miguel e Faro, cujos valores permanecem relativamente próximos, enquanto Nazaré mantém consistentemente os valores mais baixos. Esta ordenação térmica, preservada nas previsões, confirma a capacidade do modelo em capturar a estrutura espaço-temporal das séries.

Em síntese, o modelo $STARMA(3_1, 0)$ revelou-se adequado para capturar a dinâmica espaço-temporal conjunta das temperaturas à superfície do mar nas quatro localidades

analisadas, com previsões coerentes com os padrões históricos de sazonalidade e hierarquia térmica.

5.8- Avaliação da qualidade das previsões

Com o objetivo de analisar a capacidade preditiva do modelo STARMA selecionado, apresenta-se na [Tabela 5.14](#), a comparação entre os valores previstos e os valores observados. O modelo foi ajustado com base na série histórica das temperaturas registradas a um metro de profundidade nas quatro localidades em estudo, compreendida entre janeiro de 1968 e junho de 2023, tendo sido posteriormente utilizado para gerar previsões referentes ao período de julho de 2023 a junho de 2024.

Tabela 5.14: Valores observados e previstos da temperatura da água do mar à profundidade de 1 m, para as localizações de Nazaré, Faro, São Miguel e Praia, no período 2023-2024.

Mês	Nazaré 1 m		Faro 1 m		S. Miguel 1 m		Praia 1m	
	Valor Observado	Previsto	Valor Observado	Previsto	Valor Observado	Previsto	Valor Observado	Previsto
Jul. 2023	19,99	19,17	21,93	21,85	22,83	22,07	27,47	24,81
Ago. 2023	20,91	19,82	21,94	22,81	24,51	24,07	28,69	27,50
Set. 2023	20,98	19,76	22,15	23,14	23,90	24,67	28,93	28,62
Out. 2023	20,26	19,99	22,28	23,35	22,29	24,38	28,98	28,04
Nov. 2023	18,60	18,28	20,63	21,06	20,06	19,15	27,96	26,65
Dez. 2023	17,03	16,51	18,87	18,62	18,90	16,70	26,27	24,98
Jan. 2024	15,63	15,51	17,77	17,32	18,23	16,56	24,54	24,28
Fev. 2024	15,16	14,99	17,42	16,87	17,62	16,00	24,08	23,31
Mar. 2024	14,98	14,94	17,22	17,04	17,01	16,17	23,92	23,74
Abr. 2024	15,67	15,93	17,64	17,87	17,19	16,99	24,17	24,36
Mai 2024	16,87	17,29	18,91	19,37	18,96	18,36	24,81	24,57
Jun. 2024	18,15	18,84	20,29	21,08	21,55	20,29	25,08	24,57

Fonte: Elaboração própria.

De modo a avaliar a precisão das previsões, foram calculadas três métricas de erro: Erro Absoluto Médio (MAE), Raiz do Erro Quadrático Médio (RMSE) e Erro Percentual Absoluto Médio (MAPE), conforme se detalha na [Tabela 5.15](#).

Tabela 5.15: Métricas de avaliação da qualidade preditiva do modelo STARMA.

Localização	MAE (°C)	RMSE (°C)	MAPE (%)
Nazaré 1 m	0,4950	0,6164	2,61
Faro 1 m	0,5292	0,6166	2,61
S. Miguel 1 m	1,1133	1,2750	5,61
Praia 1 m	0,8208	1,0762	3,04

Fonte: Elaboração própria.

Da análise das [Tabelas 5.14 e 5.15](#), constata-se que as previsões pontuais apresentam desvios reduzidos em relação aos valores observados em todas as localizações. As métricas de erro sugerem maior precisão preditiva do modelo STARMA em Nazaré e Faro em comparação com São Miguel e Praia.

Comparação com os modelos SARIMA univariados

A comparação entre as métricas de erro dos modelos SARIMA ajustados em Nazaré, Faro, São Miguel e Praia, a um metro de profundidade ([Tabela 3.17](#)) e as do modelo STARMA ([Tabela 5.15](#)), revela que ambas as abordagens apresentam desempenhos preditivos muito semelhantes e erros reduzidos. O modelo espaço-temporal STARMA mostrou-se ligeiramente mais preciso em Nazaré, Faro e Praia, apresentando valores inferiores de MAE, RMSE e MAPE. Em contrapartida, os modelos univariados SARIMA evidenciaram um desempenho superior em São Miguel, onde registaram menores erros de previsão.

Análise de Concordância de Bland-Altman para a Previsões

Para complementar a avaliação da qualidade preditiva do modelo STARMA, procedeu-se à análise de concordância de Bland-Altman (Bland & Altman, 1986), seguindo a mesma metodologia aplicada anteriormente aos modelos SARIMA.

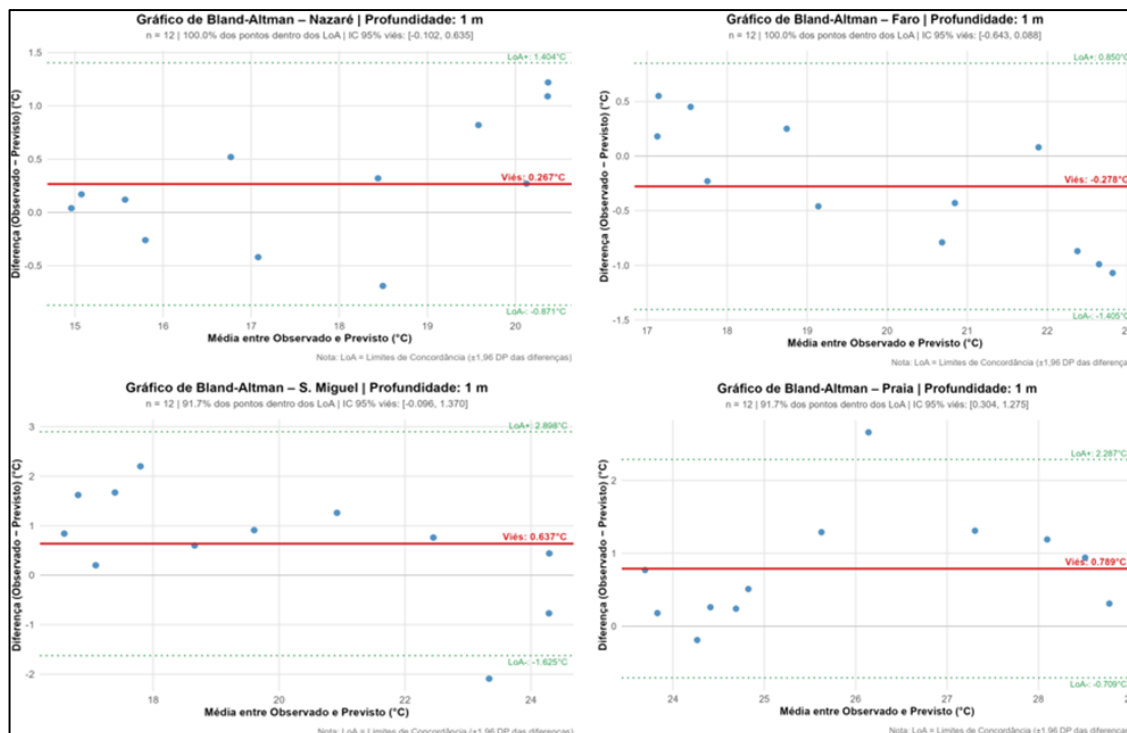


Figura 5.12: Análise de concordância de Bland-Altman para as previsões modelo STARMA (jul. 2023 - jun. 2024).

Fonte: Elaboração própria.

Da análise dos gráficos de Bland-Altman da [Figura 5.12](#), observa-se que Nazaré e Faro são as localidades onde o viés não é estatisticamente significativo, com intervalos de confiança a 95% que incluem o valor zero, sugerindo um melhor desempenho preditivo nestas localidades. Em contrapartida, Praia e São Miguel apresentam viés positivo estatisticamente significativo, indicando uma tendência consistente para subestimar a temperatura da água do mar, particularmente acentuada na Praia, consistente com a análise das medidas de erro.

Em todas as localidades analisadas, pelo menos 91,7% das observações encontraram-se dentro dos limites de concordância.

6- Capítulo VI – Considerações finais

6.1- Conclusão

A presente dissertação cumpriu o seu objetivo central de desenvolver e comparar abordagens estatísticas para a modelação e previsão da temperatura média mensal do mar no Atlântico, integrando as dimensões temporal e espacial. A análise abrangeu quatro regiões geograficamente distintas: Nazaré, Faro e São Miguel (Portugal) e Praia (Cabo Verde), entre janeiro de 1968 e junho de 2023 nas profundidades de 1 m e 5 m. Os resultados obtidos fornecem evidências robustas sobre a evolução térmica destas águas e validam a aplicação de modelos estocásticos avançados neste contexto oceanográfico.

No âmbito da modelação univariada, todas as séries analisadas revelaram tendências de aquecimento estatisticamente significativas, com aumentos médios entre $0,0196^{\circ}\text{C}$ e $0,0234^{\circ}\text{C}$ por ano, o que corresponde a aproximadamente $0,20\text{-}0,23^{\circ}\text{C}$ por década. Estes valores são consistentes com as projeções reportadas na literatura científica recente. Entre as metodologias testadas, os modelos SARIMA demonstraram superioridade face às abordagens de decomposição clássica, STL e Holt-Winters, apresentando erros de previsão reduzidos (MAE entre $0,59^{\circ}\text{C}$ e $0,97^{\circ}\text{C}$; RMSE entre $0,66^{\circ}\text{C}$ e $1,2^{\circ}\text{C}$; MAPE entre 2,94% e 4,97%). A validação *ex post* das previsões para o horizonte julho de 2023 a junho de 2024, já observável à data da redação, reforçou a credibilidade e a capacidade preditiva destes modelos.

Relativamente à dimensão espacial, a aplicação do modelo $STARMA(3_1, 0)$ às séries de superfície (1 m) permitiu integrar a dependência entre as quatro regiões, baseada numa matriz de pesos construída a partir de distâncias geodésicas. A comparação direta com os modelos SARIMA univariados a um metro de profundidade revelou que, apesar da fraca dependência espacial identificada ($\hat{\phi}_{11} = -0,038$), o modelo STARMA preserva a qualidade preditiva. Embora a influência espacial se tenha revelado moderada face à dependência temporal dominante, o modelo preservou a hierarquia térmica regional (Praia > São Miguel $\approx$ Faro > Nazaré) e produziu previsões coerentes com os padrões históricos.

Globalmente, o estudo confirma o aquecimento progressivo das águas do Atlântico e demonstra que a combinação de modelos SARIMA e STARMA constitui uma ferramenta estatística robusta para compreender e prever a variabilidade térmica regional. Os resultados obtidos fornecem evidência científica para sistemas de monitorização oceânica e para o

planeamento de políticas de adaptação climática em setores vulneráveis, como as pescas e o turismo costeiro. Reconhecem-se, contudo, limitações relacionadas com a restrição da modelação espacial à profundidade de 1 m e com a utilização de matrizes de pesos baseadas apenas em distâncias geodésicas. Estas condicionantes enquadram o alcance das conclusões apresentadas e motivam as linhas de investigação futuras detalhadas na secção seguinte.

6.2- Trabalhos futuros

Nesta secção são apresentadas sugestões de linhas de investigação a serem exploradas em trabalhos futuros, tais como:

- **Modelação espaço-temporal com dimensão vertical:** desenvolver modelos STARMA espaço-temporais que integrem simultaneamente múltiplas profundidades, tratando a profundidade como uma terceira dimensão (para além do espaço e do tempo). Esta extensão permitiria captar não apenas as dependências horizontais entre localidades, mas também as interações verticais, proporcionando uma caracterização mais completa da estrutura térmica oceânica.
- **Análise de sensibilidade à matriz de ponderação espacial:** refazer a modelação espaço-temporal considerando matrizes de pesos alternativas (por exemplo, matriz de pesos baseada na correlação entre valores observados e distância, etc.). Esta análise permitiria avaliar a robustez do modelo selecionado face a diferentes especificações da estrutura de dependência espacial.
- **Reestimação dos modelos com dados mais atualizados:** a atualização das séries com a inclusão de observações de 2024 e 2025 permitirá reestimar os modelos SARIMA e STARMA para avaliar se a estrutura identificada (ordens, parâmetros e matriz de vizinhança) permanece válida ou se exige ajustes.
- **Análise de quebras estruturais e modelação por regimes:** propõe-se investigar a existência de *breakpoints* nas séries, de modo a identificar subperíodos com tendências ou variabilidades distintas. Esta análise permitirá avaliar se os modelos SARIMA e STARMA devem ser ajustados por regimes temporais, melhorando a representação da dinâmica estudada.

Bibliografia

- Alam, N. M., Mitra, S., Pandey, S. K., Jana, C., Ray, M., Ghosh, S., Paul Mazumdar, S., Shankar, S. V., Saha, R., & Kar, G. (2024). Enhanced spatio-temporal modeling for rainfall forecasting: A high-resolution grid analysis. *Water*, 16(13), 1891. <https://doi.org/10.3390/w16131891>
- Bai, Z. (2024). Prediction of ocean temperature based on ARIMA model. In *Proceedings of the 2nd International Conference on Mathematical Physics and Computational Simulation* (pp. 114–121). <https://doi.org/10.54254/2753-8818/36/20240530>
- Bilgili, M., Pinar, E., & Durhasan, T. (2024). Global monthly sea surface temperature forecasting using the SARIMA, LSTM, and GRU models. *Earth Science Informatics*, 18, Article 10. <https://doi.org/10.1007/s12145-024-01585-z>
- Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 327(8476), 307–310. [https://doi.org/10.1016/S0140-6736\(86\)90837-8](https://doi.org/10.1016/S0140-6736(86)90837-8)
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 211–252. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.
- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting* (3rd ed.). Springer.
- Chatfield, C. (2003). *The analysis of time series: An introduction* (6th ed.). Chapman & Hall/CRC.
- Chatfield, C., & Xing, H. (2019). *The analysis of time series: An introduction with R* (7th ed.). Chapman and Hall/CRC.
- Dickey, D. A. (1976). *Estimation and hypothesis testing in nonstationary time series* (Doctoral dissertation, Iowa State University). Iowa State University Digital Repository. <https://lib.dr.iastate.edu/rtd/6267>
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366), 427–431. <https://doi.org/10.2307/2286348>
- Ehlers, R. S. (2009). *Análise de séries temporais*. Blucher.
- Enders, W. (2014). *Applied econometric time series* (4th ed.). Wiley.
- European Space Agency (2025). *Sea surface warming faster than expected*. https://www.esa.int/Applications/Observing_the_Earth/FutureEO/Space_for_our_climate/Sea_surface_warming_faster_than_expected.
- Hamed, K. H., & Rao, A. R. (1998). A modified Mann–Kendall trend test for autocorrelated data. *Journal of Hydrology*, 204(1–4), 182–196. [https://doi.org/10.1016/S0022-1694\(97\)00125-X](https://doi.org/10.1016/S0022-1694(97)00125-X)
- Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.
- Harshada, S., Satyanarayana, S., & Jayappa, K. S. (2022). Forecast of sea surface temperature and chlorophyll-a using ARIMA, south-eastern Arabian Sea, Karnataka, India. In A. K. Dikshit et al. (Eds.), *Innovative trends in hydrological and environmental systems* (Lecture Notes in Civil Engineering, Vol. 234, pp. 189–203). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-0304-5_15
- Hu, X. T., Qin, Z. Y., & Chu, F. L. (2011). Damage detection in plate structures based on space-time autoregressive moving average processes. *Journal of Physics: Conference Series*, 305, 012119. <https://doi.org/10.1088/1742-6596/305/1/012119>

- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). <https://otexts.com/fpp3/>
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3), 1–22. <https://doi.org/10.18637/jss.v027.i03>
- IPCC. (2021). *Climate change 2021: The physical science basis*. Cambridge University Press. <https://www.ipcc.ch/report/ar6/wg1/>
- Khan, I. A., Lahori, A. H., Vahidi, H., Arsalan, M. H., Shaukat, M. H., Ilyas, S., & Zia, I. (2025). Sea surface temperature dynamics and seasonal forecasts based on ARIMA modelling for the Indian Ocean using MODIS satellite data (2002–2020). *Ecological Questions*, 36(3), 137–150. <https://doi.org/10.12775/EQ.2025.024>
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1–3), 159–178. [https://doi.org/10.1016/0304-4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y)
- Makridakis, S. G., Wheelwright, S. C., & Hyndman, R. J. (1997). *Forecasting: Methods and applications* (3rd ed.). Wiley.
- Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica*, 13(3), 245–259. <https://doi.org/10.2307/1907187>
- Mann, H., & Wald, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica*, 11, 173–200.
- Martin, R. L., & Oeppen J. E. (1975). The identification of regional forecasting models using space-time correlation functions. *Transactions of the Institute of British Geographers*, 66, 95–118.
- Morettin, P. A., & Toloi, C. M. C. (1987). *Análise de séries temporais: Modelos lineares univariados*. Blucher.
- Pfeifer, P. E., & Deutsch, S. J. (1980a). A three-stage iterative procedure for space-time modeling. *Technometrics*, 22, 35–47.
- Pfeifer, P. E., & Deutsch, S. J. (1980b). Identification and interpretation of first-order space-time ARMA models. *Technometrics*, 22, 397–408.
- Pfeifer, P. E., & Deutsch, S. J. (1980c). Independence and sphericity tests for the residuals of space-time ARMA models. *Communications in Statistics - Simulation and Computation*, 9(5), 533–549.
- Pfeifer, P. E., & Deutsch, S. J. (1981). Seasonal space-time ARIMA modelling. *Geographical Analysis*, 13, 117–133.
- Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2), 335–346. <https://doi.org/10.1093/biomet/75.2.335>
- Ross, S. M. (1996). *Stochastic processes*. Wiley.
- Saes, M. J. M., Gjelstrup, C. V. B., Visser, A. W., & Stedmon, C. A. (2022). Separating annual, interannual, and regional change in sea surface temperature in the Northeastern Atlantic and Nordic Seas. *Journal of Geophysical Research: Oceans*, 127, e2022JC018630. <https://doi.org/10.1029/2022JC018630>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63(324), 1379–1389. <https://doi.org/10.1080/01621459.1968.10480934>
- Shaltout, M. (2019). Recent sea surface temperature trends and future scenarios for the Red Sea. *Oceanologia*, 61(4), 484–504. <https://doi.org/10.1016/j.oceano.2019.05.002>

- Shumway, R. H., & Stoffer, D. S. (2025). *Time series analysis and its applications: With R examples* (5th ed.). Springer.
- Subba Rao, T., & Antunes, A. M. C. (2004). Spatio-temporal modelling of temperature time series: A comparative study. In D. R. Brillinger et al. (Eds.), *Time series analysis and applications to geophysical systems* (pp. 123–147). Springer-Verlag.
- UNRIC. (2023). Qual o impacto das ondas de calor marinhas? *United Nations Regional Information Centre for Western Europe*. <https://unric.org/pt/qual-o-impacto-das-ondas-de-calor-marinhas/>

ANEXOS

Anexo I- Modelos SARIMA candidatos por localização e profundidade: BIC, AICc e valor-p do teste Ljung-Box

<i>Séries</i>	<i>Modelo</i>	<i>Ordem ARIMA</i>	<i>BIC</i>	<i>AICc</i>	<i>Valor-p</i>
Nazaré (1 metro)	Manual	SARIMA(3,0,3)(1,1,3)[12]	277,25	228,35	0,0008943
		SARIMA(3,0,2)(1,1,3)[12]	312,59	268,1	3,352e-09
		SARIMA(2,0,3)(1,1,3)[12]	297,45	252,96	6,547e-09
		SARIMA(3,0,2)(1,1,2)[12]	306,15	266,08	6,592e-09
		SARIMA(2,0,3)(1,1,2)[12]	291	250,93	1,403e-08
		SARIMA(6,0,3)(1,1,2)[12]	266,37	208,66	0,1189
	Automático	SARIMA(3,0,3)(1,1,2)[12]	275,26	226,35	0,001191
Nazaré (5 metros)	Manual	SARIMA(4,0,5)(0,1,3)[12]	307,3	249,59	0,0004401
		SARIMA(4,0,4)(0,1,3)[12]	306,61	253,3	0,0001231
		SARIMA(3,0,5)(0,1,3)[12]	302,24	248,92	0,0005982
		SARIMA(5,0,4)(0,1,2)[12]	302,24	248,93	0,002828
		SARIMA(6,0,4)(0,1,2)[12]	290,05	232,34	0,1655
		SARIMA(6,0,5)(0,1,2)[12]	295,95	233,85	0,1739
	Automático	SARIMA(4,0,5)(0,1,2)[12]	305,88	248,17	0,0006429
Faro (1 metro)	Manual	SARIMA(5,0,3)(0,1,2)[12]	793,03	744,13	0,001801
		SARIMA(5,0,2)(0,1,2)[12]	788,49	744	0,001273
		SARIMA(4,0,3)(0,1,2)[12]	795,39	750,9	5,546e-05
		SARIMA(6,0,3)(0,1,1)[12]	790,89	741,99	0,005652
		SARIMA(6,0,2)(0,1,1)[12]	784,92	740,44	0,006529
		SARIMA(6,0,3)(0,1,2)[12]	797,31	744	0,003102
		SARIMA(6,0,4)(0,1,1)[12]	792,15	738,83	0,009941
		SARIMA(6,0,5)(0,1,1) [12]	793,38	735,66	0,0102
		SARIMA(6,0,5)(1,1,1) [12]	811,83	749,72	0,0007978
		SARIMA(6,0,6)(0,1,1) [12]	788,8	726,69	0,3124
	Automático	SARIMA(5,0,3)(0,1,1)[12]	788,03	739,12	0,002095
Faro (5 metros)	Manual	SARIMA(1,0,1)(0,1,2)[12]	859	836,68	3,253e-07
		SARIMA(1,0,2)(0,1,2)[12]	865,4	838,64	2,318e-07
		SARIMA(2,0,1)(0,1,2)[12]	870,76	839,55	2,609e-07
		SARIMA(3,0,3)(0,1,1)[12]	835,05	799,41	0,003246
		SARIMA(4,0,3)(0,1,1)[12]	838,55	798,48	0,00809
		SARIMA(6,0,2)(0,1,1)[12]	847,3	802,81	0,009263
		SARIMA(6,0,2)(3,1,0)[12]	868,64	815,33	0,03857
		SARIMA(6,0,3)(3,1,0)[12]	876,75	819,04	0,01995
		SARIMA(2,0,6)(0,1,3)[12]	851,64	798,33	0,03734
		SARIMA(2,0,6)(3,1,0)[12]	863,17	809,86	0,08367
	Automático	SARIMA(1,0,1)(0,1,1)[12]	855,89	833,57	3,682e-07
São Miguel (1 metro)		SARIMA(0,0,2)(0,1,2)[12]	910,79	888,47	1,52e-07
		SARIMA(0,0,3)(0,1,2)[12]	916,72	889,95	1,583e-07
		SARIMA(1,0,2)(0,1,2)[12]	916,83	890,07	1,308e-07
		SARIMA(4,0,2)(0,1,1)[12]	916,21	880,57	7,348e-08
		SARIMA(5,0,2)(0,1,1)[12]	904,64	864,57	3,557e-06

	Manual	SARIMA(6,0,2)(0,1,1)[12]	898,68	854,19	0,001919
		SARIMA(6,0,3)(0,1,1)[12]	903,89	854,99	0,001401
		SARIMA(6,0,2)(0,1,2)[12]	905,14	856,24	0,001194
		SARIMA(5,0,5)(0,1,1)[12]	904,95	851,64	0,0288
		SARIMA(5,0,5)(0,1,2)[12]	910,68	852,96	0,02972
		SARIMA(5,0,5)(0,1,3)[12]	917,14	855,04	0,01876
		SARIMA(5,0,4)(0,1,1)[12]	900,35	851,45	0,001881
		SARIMA(6,0,6)(0,1,1)[12]	916,26	854,15	0,0206
		SARIMA(6,0,6)(0,1,2)[12]	921,96	855,47	0,01947
	Automático	SARIMA(0,0,2)(0,1,1)[12]	905,47	887,59	6,258e-08
São Miguel (5 metros)	Manual	SARIMA(1,0,1)(0,1,2)[12]	904,21	881,89	0,0004135
		SARIMA(1,0,2)(0,1,2)[12]	910,21	883,44	0,0003268
		SARIMA(2,0,1)(0,1,2)[12]	909,03	882,26	0,001843
		SARIMA(3,0,1)(0,1,2)[12]	911,82	880,61	0,003298
		SARIMA(4,0,1)(0,1,2)[12]	917,22	881,58	0,001849
		SARIMA(6,0,1)(0,1,1)[12]	912,18	872,12	0,02144
		SARIMA(6,0,1)(0,1,2)[12]	918,51	874,02	0,01825
		SARIMA(6,0,2)(0,1,1)[12]	910,64	866,15	0,04347
		SARIMA(6,0,2)(0,1,2)[12]	916,93	868,03	0,03517
		SARIMA(6,0,3)(0,1,1)[12]	916,72	867,82	0,03076
		SARIMA(6,0,4)(0,1,1)[12]	916,6	863,29	0,04506
		SARIMA(6,0,4)(0,1,2)[12]	922,76	865,04	0,02509
		SARIMA(6,0,5)(0,1,1)[12]	920,33	862,61	0,03804
		SARIMA(6,0,5)(0,1,2)[12]	926,48	864,37	0,01978
		SARIMA(6,0,6)(0,1,1)[12]	925,73	863,62	0,08473
	Automático	SARIMA(1,0,1)(0,1,1)[12]	898,05	880,18	0,0004659
Praia (5 metros)	Manual	SARIMA(2,0,2)(0,1,2)[12]	615,99	584,78	0,003398
		SARIMA(2,0,3)(0,1,2)[12]	622,46	586,81	0,002152
		SARIMA(3,0,2)(0,1,2)[12]	622,47	586,83	0,002148
		SARIMA(4,0,3)(0,1,1)[12]	617,44	577,37	0,01411
		SARIMA(3,0,3)(0,1,1)[12]	622	586,36	0,002255
		SARIMA(3,0,3)(0,1,2)[12]	628,48	588,41	0,001384
		SARIMA(4,0,3)(0,1,2)[12]	623,82	579,33	0,01019
		SARIMA(4,0,4)(0,1,1)[12]	614,73	570,24	0,1373
		SARIMA(4,0,4)(0,1,2)[12]	620,99	572,09	0,09992
		SARIMA(3,0,4)(0,1,1)[12]	610,62	570,55	0,1068
		SARIMA(3,0,4)(0,1,2)[12]	617,08	572,59	0,07773
		SARIMA(2,0,4)(0,1,1)[12]	616,24	580,6	0,009331
	Automático	SARIMA(2,0,2)(0,1,1)[12]	609,51	582,74	0,005219

**Anexo II- Previsões intervalares da temperatura da água (80% de confiança),
modelos SARIMA**

Local / Prof,	Intervalo	Jul 2023	Ago 2023	Set 2023	Out 2023	Nov 2023	Dez 2023	Jan 2024	Fev 2024	Mar 2024	Abr 2024	Mai 2024	Jun 2024
Nazaré 1 m	LI	18,83	19,35	18,98	19,05	17,42	15,66	14,72	14,23	14,02	14,89	16,33	17,98
	LS	19,55	20,37	20,07	20,16	18,52	16,77	15,83	15,35	15,14	16,02	17,45	19,10
Nazaré 5 m	LI	18,61	19,20	18,77	18,96	17,41	15,68	14,71	14,17	13,93	14,70	16,11	17,74
	LS	19,34	20,24	19,89	20,10	18,55	16,84	15,87	15,33	15,09	15,86	17,27	18,91
Faro 1 m	LI	20,96	21,67	21,89	21,78	19,10	16,84	16,17	15,97	15,72	16,52	18,02	19,96
	LS	22,02	23,06	23,32	23,21	20,53	18,27	17,61	17,41	17,15	17,96	19,46	21,41
Faro 5 m	LI	20,78	21,66	21,86	21,72	19,11	16,64	16,15	16,01	15,70	16,56	17,89	19,77
	LS	21,91	23,17	23,40	23,26	20,65	18,19	17,70	17,56	17,25	18,11	19,44	21,32
São Miguel 1 m	LI	21,60	23,17	23,26	22,68	18,66	16,55	16,40	15,67	15,25	15,36	17,39	19,98
	LS	22,77	24,67	24,78	24,21	20,20	18,10	17,95	17,22	16,81	16,92	18,95	21,55
São Miguel 5 m	LI	21,23	22,92	23,20	22,59	18,76	16,60	16,26	15,62	15,29	15,37	17,30	19,92
	LS	22,41	24,42	24,74	24,14	20,32	18,16	17,82	17,18	16,85	16,94	18,87	21,49
Praia 1 m	LI	24,23	27,10	28,12	27,69	26,18	24,26	23,27	22,65	22,54	22,83	23,53	24,14
	LS	25,11	28,28	29,35	28,93	27,42	25,53	24,54	23,92	23,81	24,10	24,81	25,41
Praia 5 m	LI	25,17	27,31	28,20	27,84	26,48	23,91	23,16	22,71	22,49	22,62	23,42	24,59
	LS	26,12	28,62	29,59	29,25	27,90	25,36	24,61	24,16	23,94	24,06	24,86	26,03

Nota: LI e LS correspondem aos limites inferior e superior dos intervalos de confiança, Valores em °C.

**Anexo III- Previsões intervalares da temperatura da água (95% de confiança),
modelos SARIMA**

Local / Prof,	Intervalo	Jul 2023	Ago 2023	Set 2023	Out 2023	Nov 2023	Dez 2023	Jan 2024	Fev 2024	Mar 2024	Abr 2024	Mai 2024	Jun 2024
Nazaré 1 m	LI	18,64	19,07	18,69	18,76	17,13	15,37	14,42	13,93	13,72	14,60	16,03	17,68
	LS	19,74	20,64	20,36	20,45	18,81	17,07	16,13	15,64	15,44	16,31	17,75	19,40
Nazaré 5 m	LI	18,42	18,93	18,47	18,66	17,11	15,38	14,41	13,86	13,62	14,39	15,80	17,44
	LS	19,53	20,52	20,19	20,40	18,86	17,14	16,18	15,64	15,40	16,17	17,58	19,22
Faro 1 m	LI	20,68	21,30	21,51	21,40	18,72	16,46	15,79	15,59	15,34	16,14	17,64	19,58
	LS	22,31	23,43	23,70	23,59	20,91	18,65	17,99	17,79	17,53	18,34	19,84	21,79
Faro 5 m	LI	20,48	21,26	21,45	21,31	18,70	16,23	15,74	15,60	15,29	16,15	17,48	19,36
	LS	22,21	23,57	23,81	23,67	21,06	18,60	18,11	17,97	17,66	18,52	19,85	21,73
São Miguel 1 m	LI	21,29	22,78	22,85	22,27	18,25	16,14	15,99	15,26	14,84	14,95	16,97	19,57
	LS	23,08	25,06	25,19	24,62	20,61	18,51	18,36	17,64	17,22	17,33	19,37	21,96
São Miguel 5 m	LI	20,92	22,52	22,79	22,18	18,35	16,19	15,85	15,21	14,87	14,96	16,89	19,51
	LS	22,72	24,81	25,14	24,55	20,73	18,57	18,24	17,60	17,26	17,35	19,28	21,90
Praia 1 m	LI	24,00	26,79	27,80	27,36	25,85	23,93	22,94	22,31	22,20	22,49	23,20	23,80
	LS	25,34	28,59	29,67	29,25	27,75	25,87	24,88	24,25	24,15	24,44	25,14	25,74
Praia 5 m	LI	24,92	26,96	27,83	27,46	26,10	23,53	22,78	22,33	22,11	22,23	23,03	24,20
	LS	26,37	28,97	29,96	29,63	28,28	25,74	24,99	24,54	24,32	24,45	25,25	26,42

Anexo IV - Tabela de valores críticos do teste Dickey-Fuller

τ								
n	.01	.025	.05	.10	.90	.95	.975	.99
25	-2.66	-2.26	-1.95	-1.60	.92	1.33	1.70	2.16
50	-2.62	-2.25	-1.95	-1.61	.91	1.31	1.66	2.07
100	-2.60	-2.24	-1.95	-1.61	.90	1.29	1.64	2.03
250	-2.58	-2.24	-1.95	-1.62	.89	1.28	1.63	2.00
500	-2.58	-2.24	-1.95	-1.62	.89	1.28	1.62	2.00
750	-2.58	-2.24	-1.95	-1.62	.89	1.28	1.62	2.00
∞	-2.58	-2.23	-1.95	-1.62	.89	1.28	1.62	1.99

τ_{μ}								
n	.01	.025	.05	.10	.90	.95	.975	.99
25	-3.75	-3.33	-2.99	-2.64	-.37	.00	.34	.71
50	-3.59	-3.23	-2.93	-2.41	-.41	-.04	.28	.66
100	-3.50	-3.17	-2.90	-2.59	-.42	-.06	.26	.63
250	-3.46	-3.14	-2.88	-2.58	-.44	-.07	.24	.62
500	-3.43	-3.13	-2.87	-2.57	-.44	-.07	.23	.61
750	-3.43	-3.13	-2.87	-2.57	-.44	-.08	.23	.61
∞	-3.42	-3.12	-2.86	-2.57	-.44	-.08	.23	.60

τ_{τ}								
n	.01	.025	.05	.10	.90	.95	.975	.99
25	-4.38	-3.95	-3.60	-3.24	-1.14	-.81	-.50	-.15
50	-4.15	-3.80	-3.50	-3.18	-1.19	-.87	-.58	-.24
100	-4.04	-3.73	-3.45	-3.15	-1.22	-.92	-.64	-.27
250	-3.98	-3.69	-3.42	-3.13	-1.23	-.92	-.64	-.27
500	-3.95	-3.67	-3.41	-3.13	-1.24	-.93	-.65	-.32
750	-3.95	-3.67	-3.41	-3.13	-1.24	-.93	-.65	-.32
∞	-3.95	-3.67	-3.41	-3.13	-1.25	-.94	-.66	-.32

Fonte: Dickey (1976, pp. 53)

Valores críticos assintóticos do teste Dickey-Fuller para as três especificações:

- τ - Modelo sem constante
- τ_{μ} – Modelo com constante
- τ_{τ} – Modelo com constante e tendência

Anexo V - Tabela de valores críticos do teste KPSS de estacionariedade

η_{μ}				
Critical level	0.10	0.05	0.025	0.01
Critical value	0.347	0.463	0.574	0.739
η_{τ}				
Critical level	0.10	0.05	0.025	0.01
Critical value	0.119	0.146	0.176	0.216

Fonte: Kwiatkowski et al. (1992, pp. 166)

Valores críticos assintóticos do teste KPSS para as especificações:

- η_{μ} - Modelo com constante
- η_{τ} - Modelo com constante e tendência linear