

*Research Paper***Modelação e Previsão do fluxo de Turismo em Portugal: Perspetivas para uma gestão estratégica****Modeling and forecasting tourism flow in Portugal: Perspectives for a strategic management***Submitted in 10, October 2022**Accepted in 16, November 2022**Evaluated by a double-blind review system*

FILIPE R. RAMOS^{1*}
LUÍSA P. LACERDA²
JOSÉ D. CURTO³

RESUMO

Objetivo: Com aumento do Turismo em Portugal, bem como das empresas afetas ao mesmo, impõe-se a necessidade de análise e previsão do fluxo de turistas para que a gestão dos negócios seja dotada de uma estratégia competitiva. Face às alterações na ‘recente’ dinâmica dos dados afetos ao turismo, neste artigo discutem-se os contributos e limitações do recurso às metodologias clássicas de previsão, quando aplicadas a este setor, nomeadamente ao número de dormidas em estabelecimentos de alojamento turístico em Portugal.

Metodologia: O estudo incide na modelação e previsão de séries temporais afetas ao número de dormidas mensais, em estabelecimentos de alojamento turístico em Portugal, entre janeiro de 2002 e março de 2022. Decorrente de algumas sugestões constantes na literatura científica, recorreram-se às metodologias de Alisamento Exponencial (*Exponential Smoothing* – ETS). Em termos computacionais, foi utilizado o ambiente computacional Jupyter, com linguagem de programação *Python* (versão 3.7.3).

Resultados: Os resultados foram apresentados e discutidos mediante a análise de duas séries temporais: (1) Número total de dormidas em estabelecimentos de alojamento turístico em Portugal – série Total; (2) Número de dormidas de residentes em Portugal em estabelecimentos de alojamento turístico em Portugal – série Residentes. Globalmente, da análise das séries temporais, verificou-se o crescimento do Turismo em Portugal desde 2002, com uma quebra visível em 2020, decorrente da situação pandémica. Relativamente às metodologias ETS utilizadas na modelação e previsão, embora estas tenham correspondido de forma positiva na previsão da série Total (embora com algum erro), o mesmo não se verificou na série Residentes. Nesta série, por ter uma dinâmica recente completamente atípica (os dados históricos não replicam o comportamento atual) verifica-se que os modelos ETS, potencialmente mais adequados (com componente de tendência e sazonalidade), não convergem, em geral. Contudo,

^{1*} Corresponding author. FCUL e CEAUL, Universidade de Lisboa, Portugal, frramos@fc.ul.pt

² ISCTE, Instituto Universitários de Lisboa, Portugal. mlppb@iscte-iul.pt

³ ISCTE e BRU, Instituto Universitários de Lisboa, Portugal, dias.curto@iscte-iul.pt

importante será referir que foram as dormidas dos residentes que, no período pandémico, ditaram a dinâmica presente na série Total.

Limitações: A literatura aponta para um bom desempenho das metodologias ETS em séries temporais com características presentes nas séries em estudo (com presença de um ciclo de tendência e clara sazonalidade), facto que motivou a sua escolha. Contudo, a dificuldade destas metodologias em lidar com quebras abruptas de estrutura, no histórico dos dados, ficou patente neste estudo. Não obstante de previsões mais ou menos ajustadas, o destaque vai para a não convergência de alguns modelos que, em certa medida, se poderiam ajustar melhor ao histórico de dados. Neste sentido, impõe-se a procura de metodologias de previsão alternativas, onde as metodologias de *Machine Learning*, nomeadamente *Deep Learning (Deep Neural Networks)* têm sido apontadas na literatura científica como bastante promissoras. Esse será o caminho seguinte de investigação.

Originalidade: Face à importância que o Turismo tem no tecido económico e social em Portugal, dado que se trata de um setor bastante volátil e em constante mudança, impõe-se a definição de uma estratégia de atuação futura para perceber como, internamente, o sector pode definir políticas por forma a evitar situações de dependência externa. Para tal, além de uma análise atual ao histórico de dados, decorrente de um período atípico de pandemia, impõem-se avaliar de forma crítica a capacidade preditiva de modelos econométricos (clássicos), passíveis de serem utilizados pela indústria afeta ao turismo. Tal contribui não só para um melhor entendimento do fenómeno em estudo, como constitui uma ferramenta de apoio à tomada de decisão.

Palavras-Chave: Turismo, Gestão Estratégica, Séries Temporais, Modelos de Alisamento Exponencial, Previsão, Erro de Previsão.

ABSTRACT

Purpose: The increase of Tourism in Portugal, as well as the companies related to it, it is necessary to analyze and forecast the flow of tourists so that the management of the business is endowed with a competitive strategy. Given the changes in the 'recent' dynamics of tourism data, this article discusses the contributions and limitations of using classical forecasting methodologies, when applied to this sector, namely to the number of overnight stays in tourist accommodation establishments in Portugal.

Methodology: The study focuses on the modeling and forecasting of time series related to the number of monthly overnight stays, in tourist accommodation establishments in Portugal, between January 2002 and March 2022. As a result of some suggestions contained in the scientific literature, it was resorted the Exponential Smoothing (ETS) methodologies. In computational terms, we used the Jupyter computational environment, with the Python programming language (version 3.7.3).

Findings: The results were presented and discussed through the analysis of two time series: (1) Total number of overnight stays in tourist accommodation establishments in Portugal – Total series; (2) Number of overnight stays spent by residents in Portugal in tourist accommodation establishments in Portugal – Residents series. Overall, from the analysis of the time series, there was a growth of Tourism in Portugal since 2002, with a

visible drop in 2020, due to the pandemic situation. Regarding the ETS methodologies used in the modeling and forecasting, although they corresponded positively in the forecast of the Total series (with some error), the same did not happen in the Residents series. In this series, due to the recent dynamics that are completely atypical, it appears that the ETS methodologies, potentially more adequate, do not converge, in general. However, it is important to mention that it was the overnight stays of residents that, in the pandemic period, dictated the dynamics present in the Total series.

Research limitations: The literature points to a good performance of ETS methodologies in time series with characteristics present in the series under study (with the presence of a trend cycle and clear seasonality), a fact that motivated its choice. However, the difficulty of these methodologies in dealing with abrupt breaks in the data history was evident in this study. Despite how adjusted the forecasts are, the highlight is the non-convergence of some models that could be better adjusted to the historical data. In this sense, it is necessary to search for alternative forecasting methodologies, where Machine Learning methodologies, namely Deep Learning (Deep Neural Networks) have been pointed out in the scientific literature as quite promising. This will be the next step of the investigation.

Originality: Given the importance that Tourism has both in the economic and social dimension of Portugal, and being a very volatile and constantly changing sector, it is imperative to define a strategy for future action to understand how, internally, the sector can define policies to avoid situations of external dependence. In addition to a current analysis of the data history, resulting from an atypical period of pandemic, we need to critically evaluate the predictive capacity of (classical) econometric models, which can be used by the industry related to tourism. This not only contributes to a better understanding of the phenomenon under study, but also constitutes a tool to support decision-making.

Keywords: Tourism, Time Series, Exponential Smoothing Models, Forecasting, Business Management, Strategic Management.

1. Introdução

O setor do Turismo é um motivo para se criar, manter e desenvolver empresas, sendo uma atividade económica fundamental para a geração de emprego e de riqueza em Portugal. Por se tratar de uma atividade em constante crescimento, as empresas devem conhecer a sua dinâmica e definir uma gestão estratégica que sustente o negócio. A gestão estratégica consiste em continuamente planear, monitorizar, analisar e avaliar todas as necessidades que uma organização precisa para atingir as suas metas e objetivos.

Como é do conhecimento geral, resultante da situação pandémica a nível mundial – COVID-19 – os valores para onde apontavam as previsões do Turismo (nomeadamente a nível nacional) sofreram uma quebra impensável. De acordo com os dados da Organização Mundial do Turismo (OMT), comparativamente a 2020 (o pior ano desde que há registo), embora o turismo mundial tenha recuperado 4% em 2021, os valores permaneceram 72% abaixo dos níveis de 2019 (ano pré-pandémico). Em linha com outros destinos mundiais, apesar de, em 2021, o turismo em Portugal ter registado uma

recuperação face a 2020 (mais 37,4 milhões de dormidas em alojamentos turístico, correspondendo a um aumento de 45,2%), as dormidas ainda se encontram 53% abaixo das dormidas registadas nos alojamentos turístico em 2019.

Surge então a necessidade de uma análise cuidada à tendência recente dos valores afetos ao Turismo, bem como identificar modelos preditivos capazes de auxiliar os gestores na tomada de decisões. A robustez e precisão da previsão será um fator preponderante para que a informação obtida contribua, efetivamente, para a definição de uma estratégia que permita a sustentabilidade e a prosperidade dos negócios afetos ao Turismo.

Assim, este estudo tem como objetivo analisar o fluxo de turismo (histórico e atual) e discutir o modo como as metodologias clássicas podem potenciar a forma de prever o fluxo de Turismo em Portugal. Para o efeito, o estudo assenta na modelação e previsão de séries temporais afetas ao número de dormidas mensais, em estabelecimentos de alojamento turístico em Portugal (continental e ilhas), entre janeiro de 2002 e março de 2022. Decorrente de algumas sugestões constantes na literatura científica, na modelação e previsão das séries temporais serão utilizadas as metodologias de Alisamento Exponencial (*Exponential Smoothing – ETS*). A análise crítica à qualidade preditiva dos modelos será feita com base em métricas de mediação do erro, nomeadamente a métrica *Mean Absolute Percentage Error (MAPE)*.

Como resultado dessa análise, pretende-se que as decisões tomadas, na gestão dos negócios deste setor, estejam assentes em dados concretos, garantindo o sucesso das organizações e das pessoas envolvidas.

Em termos de estrutura, além da presente Introdução (secção 1), este trabalho está organizado em mais 4 secções, sucintamente descritas em seguida.

Numa primeira (secção 2), é feita uma revisão da literatura, balizando esta investigação.

As secções 3 e 4 respeitam à parte prática, sendo que na secção 3 se expõem os procedimentos metodológicos seguidos na análise de dados e onde são apresentadas as séries temporais em estudo. A secção 4 é dedicada na íntegra à apresentação de resultados, começando pela análise exploratória das séries temporais, seguindo-se a parte afeta à modelação e previsão.

Finalmente, na secção 5, são apresentadas e sintetizadas as principais conclusões deste estudo.

2. Revisão da Literatura

2.1. Turismo em Portugal

Definir o conceito de Turismo gera alguma controvérsia, uma vez que, enquanto conceito e atividade, este tem evoluído ao longo do tempo, em especial nas últimas décadas (Caldas et al., 2020). Bucho (2010) afirma que, independentemente da definição que se escolha e seja qual for o prisma pelo qual se analise este fenómeno, é indiscutível a componente principal que o caracteriza: o Turismo presume a deslocação das pessoas, a viagem.

Em Portugal, a atividade turística emprega aproximadamente meio milhão de pessoas (Ramos & Costa, 2017). Portugal passou a ter um rácio superior a 1 (um turista por cada habitante) desde de 2014, indicador que não tem parado de subir (Botelho, 2019).

Segundo López (2021), Lisboa foi por larga margem o destino de viagem mais popular de Portugal para turistas internacionais em 2019 e 2020, e como resultado, teve o maior número de aberturas de hotéis e albergues no último ano. A mesma autora demonstra ainda que a esmagadora maioria das empresas afetas ao Turismo são microempresas (90,2%). A capital portuguesa não é caso único. A importância que o Turismo teve/tem na zona norte de Portugal, nomeadamente na cidade do Porto, é algo também notado, tendo este fenómeno merecido particular reflexão em Ramos et al. (2019) pela forma como este contribuiu para o potenciar de uma economia num período de recessão.

Devido ao seu crescimento, o Turismo de Portugal tem recebido vários prémios,⁴ distinções e incentivos para se desenvolver e tornar o país mais próspero. Estes reconhecimentos e conseqüente crescimento exigem uma Gestão Estratégica dos negócios associados ao Turismo.

2.2. Paradigmas do Turismo

O Turismo, nesta fase da história, apresenta um crescimento muito acentuado, nomeadamente em algumas zonas de Portugal como o Algarve, a zona de Lisboa e a Madeira (Ramos & Costa, 2017). Invernos amenos, temperaturas quentes de verão e uma oferta diversificada de destinos costeiros e de montanha fizeram de Portugal um destino de férias ideal na Europa — tanto que em 2021 Portugal liderou o ranking dos melhores destinos europeus para visitar de carro, pelo seu clima, segurança, acessibilidade e pontos de referência, entre outras coisas (López, 2021).

Contudo, temos vindo a assistir a um novo paradigma, onde não só os estrangeiros querem desfrutar destes destinos para férias, como os próprios locais têm aproveitado para explorar e conhecer mais a fundo o país onde vivem. Esta situação destacou-se nos últimos dois anos, tendo sido motivo de várias notícias nos meios de comunicação social. Porém, se nos últimos anos já vínhamos a assistir a campanhas com o slogan “Vá para fora cá dentro” (Turismo de Portugal, 2013), em redes e meios de comunicação social, fatores como a COVID-19 foram determinantes para a dinâmica dos dados recentes.

Certo é que, ao ter mais portugueses a explorar e conhecer novas zonas de Portugal, desde o litoral ao interior, estes acabam por desempenhar o papel de turistas no seu próprio país, desfrutando de programas culturais, gastronómicos e lúdicos, o que se revela benéfico para a Economia interna.

2.3. Gestão Estratégica e recuperação de um período de pandemia

Ter e fazer uso de uma “Gestão Estratégica” é tido pelos profissionais como um motor para o desenvolvimento de qualquer atividade. O Turismo não é exceção. Se a atividade turística for dotada de uma linha (de gestão) estratégica, mais facilmente se promove a eficiência organizacional em qualquer entidade envolvida no Turismo (Kirovska, 2011).

Com o aumento da concorrência e da globalização, a dinâmica e o desenvolvimento do Turismo são cada vez mais intensificados e complexos, impondo-se cada estado, cada sector e cada agente dotar-se de informação clara objetiva que fundamente as linhas de atuação. É neste prisma que o papel da Gestão Estratégica se torna proeminente, conduzindo ao assegurar de operações comerciais eficientes e eficazes nos mais diversos

⁴ A lista de prémios recebidos (por ano e categorias) pode ser consultada no site do Turismo de Portugal http://www.turismodeportugal.pt/pt/quem_somos/Organizacao/Premios_Distincoes/Paginas/default.aspx

domínios afetos ao Turismo (Kirovska, 2011). São exemplo a gestão de recursos físicos e humanos, a gestão de sistemas de informação, a manutenção e desenvolvimento de infraestruturas, ou mesmo a expansão e desenvolvimento do negócio turístico conducentes a projetos de Turismo novos e inovadores.

Deste modo, será plausível assumir que, decorrente de uma Gestão Estratégica sólida e devidamente estruturada e fundamentada, se pode continuar a encarar o Turismo como um pilar estratégico para a (recuperação da) economia, com contribuições cada vez mais significativas para o produto interno bruto local (Zhang et al., 2021).

2.4. Previsão do Turismo

Hoje a necessidade de adaptação e aprendizagem é largamente justificada pela globalização, incerteza e mudança (Machado et al., 2020), onde a previsão pode constituir uma excelente ferramenta no apoio à tomada de decisão. É por isto que a previsão da procura de Turismo se torna algo fundamental para a definição de uma estratégia de planeamento, uma vez que permite a expansão dos serviços turísticos, tanto no investimento em recursos humanos como em estruturas físicas, para poder atender às necessidades das pessoas (Casanova, 2012) e aos desafios atuais e futuros. Mais, parte dos serviços/recursos turísticos têm uma natureza perecível, o que reforça a necessidade prever, o melhor possível, a procura do Turismo (Gunter & Önder, 2015).

De acordo com López (2021), antes da pandemia do COVID-19, o setor do Turismo representava quase um quinto da economia portuguesa, mas a crise global fez com que a contribuição do Turismo para o PIB caísse mais de metade em 2020.

Deste modo, encontrar modelos preditivos capazes de lidar com quebras (inesperadas) constitui um desafio para as empresas e investigadores, uma vez que, alinhados com a literatura científica, fenómenos como este podem constituir um verdadeiro obstáculo à previsão, tal como destacado em Ramos (2021). Mais, a natureza sazonal da procura turística é também uma característica a ter em conta antes da identificação e seleção do modelo de previsão ideal (Goh & Law, 2002).

Com base nestas premissas, os desafios para a previsão do fluxo do turismo não é um tema estanque na literatura. Mais, face a conjuntura atual, onde muitas organizações contam com a previsão para preparar os seus anos de recuperação de crise (Zhang et al., 2021), impõe-se perceber a dinâmica atual do fluxo de turismo e de que forma a previsão pode levar a uma perspetiva estratégica do desenvolvimento do negócio, com benefícios indiscutíveis para a estrutura social e económica do país.

3. Metodologia

Nesta seção, além de ser feita uma referência aos procedimentos metodológicos adotados na investigação (metodologia usada na modelação e previsão das séries temporais e respetiva implementação computacional), são apresentadas as séries que servirão de base à parte empírica do nosso estudo.

3.1. Séries Temporais: Modelação e Previsão

A articulação de técnicas/ferramentas estatísticas, matemáticas e computacionais, no processo de análise, modelação e previsão de séries temporais, manifesta-se um claro

suporte de apoio à tomada de decisão (Ramos, 2021). O estudo de séries temporais tem-se mostrado com grande interesse prático não só no campo da investigação, como no campo da ação, com vasta aplicabilidade, permitindo descobrir, com alguma margem de erro, os valores futuros a partir de um histórico de valores (Tealab, 2018).

Parte dos modelos econométricos afetos ao estudo de séries temporais baseiam-se apenas no histórico de dados (valores passados) para efetuar previsões. O comportamento de uma série temporal é marcado por algumas componentes (tendência, ciclicidade, sazonalidade e aleatoriedade)⁵, onde a existência de quebras de estrutura poderá ser um claro obstáculo ao processo de modelação e previsão (Pesaran et al., 2006).

Quando uma série apresenta uma dinâmica simples, com clara evidência de tendência e/ou sazonalidade, as metodologias clássicas (muito em particular as metodologias de alisamento exponencial) mostram-se suficientes por apresentarem previsões relativamente ajustadas, com um baixo custo computacional (Ramos, 2021). Simões et al. (2019) destacam, também, estas metodologias pela facilidade de utilização e entendimento, referindo-se à forte aplicabilidade no estudo de séries temporais. Em geral, estas metodologias devolvem resultados satisfatórios, além de permitir uma boa compreensão dos resultados.

3.2. Modelos de Alisamento Exponencial

A metodologia de alisamento exponencial, ETS (*Exponential Smoothing*), embora tenha sido proposta no final da década de 1950, é um clássico na modelação de séries temporais, não só por apresentar previsões bastante fiáveis, como pelo facto de serem de rápida implementação computacional e compreensão simples, fatores que se revertem como uma clara vantagem para o mundo empresarial (Hyndman & Athanasopoulos, 2018).

Sucintamente, os modelos ETS baseiam-se em médias ponderadas de observações anteriores, podendo os pesos decrescer, ou não, exponencialmente, à medida que as observações tendem para o passado. Ou seja, ao invés de os pesos serem considerados uniformes, há possibilidade de uma ‘ponderação’, podendo as observações recentes terem mais peso que as observações antigas. Em modelos mais simples de ETS apenas se observa que quanto mais recentes forem as observações, maior será o peso associado (ou vice-versa). Já modelos mais avançados podem integrar outras componentes, como tendência e/ou sazonalidade. Importante será ainda referir que, comparativamente a outros modelos (clássicos), em termos de implementação prática, estes modelos não requerem etapas prévias relativas a pré-processamento dos dados (como são exemplo os modelos autorregressivos que requerem a estacionaridade das séries).

Segundo a bibliografia (Hyndman & Athanasopoulos, 2018), podemos considerar três tipos de modelos ETS, *Single Exponential Smoothing* (SETS), *Double Exponential Smoothing* (DETS) e *Triple Exponential Smoothing* (TETS). Sucintamente, se o método mais simples assume um comportamento estacionário, o segundo lida explicitamente com existência de uma tendência, enquanto o terceiro adiciona à tendência a ocorrência de

⁵ Tendência reflete uma evolução, a longo prazo, no sentido da monotonia da série (crescente/decrescente), podendo apresentar uma dinâmica linear ou não linear, sendo facilmente identificada através de uma análise gráfica; Ciclicidade é traduzida por um padrão de flutuações de médio prazo, podendo ou não ser periódicas, as quais podem afetar diretamente a tendência global da série; Sazonalidade reflete-se em variações cíclicas regulares que ocorrem em períodos constantes e menores comparativamente à ciclicidade; Aleatoriedade está associada a flutuações imprevisíveis, não modeladas pelas componentes anteriores (Ramos, 2021).

dinâmicas de sazonalidade. Assim, no seguimento do exposto, em termos de implementação computacional, consideraram-se 9 modelos ETS, como se descreve na Tabela 1, em que cada modelo é rotulado por um par de letras, (T,S), relativas à componente de Tendência (T) e à componente de Sazonalidade (S).

Tabela 1. Classificação dos modelos ETS

Componente de tendência (T)	Componente de sazonalidade (S)		
	Nenhuma (N)	Modelo Aditivo (A)	Modelo Multiplicativo (M)
Nenhuma (N)	(N, N)	(N, A)	(N, M)
Modelo Aditivo (A)	(A, N)	(A, A)	(A, M)
Modelo Aditivo amortecido (A _d)	(A _d , N)	(A _d , A)	(A _d , M)

Fonte: Adaptado de Hyndman & Athanasopoulos (2018, sec. 7.4)

De acordo com o exposto, podemos identificar três grandes grupos de modelos:

- *Single Exponential Smoothing*: (N, N)
- *Double Exponential Smoothing*: (A, N) e (A_d, N)
- *Triple Exponential Smoothing*: (N, A), (A, A), (A_d, A), (N, M), (A, M) e (A_d, M)

Com efeito, na implementação das metodologias ETS, é útil reconhecer a presença de comportamentos de tendência (ou cliço de tendência), de sazonalidade e do "ruído" presente nas observações, avaliando-se a 'força de tendência' e a 'força de sazonalidade' (Wang et al., 2006). Assim, denotando por CT_t a componente do ciclo de tendência, por S_t a componente sazonal e por R_t a componente residual (aleatória), foi calculada a força de tendência (\mathcal{F}_T) e a força de sazonalidade (\mathcal{F}_S) de cada uma das séries temporais estudadas. Em dados que evidenciam uma forte tendência, a variância da soma das componentes CT_t com R_t , $Var(CT_t + R_t)$, apresenta um valor considerável e, consequentemente, o valor do quociente entre as duas variâncias $Var(R_t)$ e $Var(CT_t + R_t)$, $Var(R_t)/Var(CT_t + R_t)$, deverá ser relativamente pequeno. Definimos, assim, a 'força de tendência' pela Equação (1)

$$\mathcal{F}_T = \max\left(0, 1 - \frac{Var(R_t)}{Var(CT_t + R_t)}\right) \tag{1}$$

onde uma série com valor de \mathcal{F}_T próximo de 1 evidencia uma forte tendência, enquanto numa série sem tendência, o valor será 0.

De forma análoga podemos definir 'força de sazonalidade' pela Equação (2)

$$\mathcal{F}_S = \max\left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t + R_t)}\right) \quad (2)$$

onde uma série com valor de \mathcal{F}_S próximo de 0 exibe pouca ou nenhuma sazonalidade, enquanto numa série com forte sazonalidade, o valor será próximo de 1.

Deste modo, além da determinação das medidas supramencionadas, para cada uma das séries, procurando evitar uma escolha *ad hoc* do melhor modelo para efetuar cenários de previsão, a qual poderia ser feita mediante análise gráfica das principais componentes da série (tendência e sazonalidade) e do modo como elas entram no método de suavização (por exemplo, de forma aditiva, amortecida ou multiplicativa), optou-se por considerar uma forma mais robusta analisando-se, para cada série, todos os modelos em paralelo com a informação obtida a partir dos critérios AIC (Akaike, 1974) e BIC (Schwarz, 1978), onde a ‘convergência’ dos parâmetros de cada modelo, para efeitos de previsão pode ser avaliada, escolhendo-se o(s) melhor(es) modelo(s).

3.3. Dados e implementação computacional

Os dados que permitiram fazer o presente estudo foram recolhidos do site do INE (2022) e referem-se ao número de dormidas em estabelecimentos de alojamento turístico em Portugal (continental e ilhas), entre janeiro de 2002 e abril de 2022. Com o intuito de dar resposta aos objetivos traçados na investigação, considerou-se útil considerar duas séries temporais:

- A. Número total de dormidas em estabelecimentos de alojamento turístico em Portugal – série Total;
- B. Número de dormidas de residentes em Portugal em estabelecimentos de alojamento turístico em Portugal – série Residentes.

Dada que a informação estava dispersa, foi necessário desenhar um plano sistemático de recolha, tratamento e organização de dados numa folha de cálculo *Microsoft Excel*.

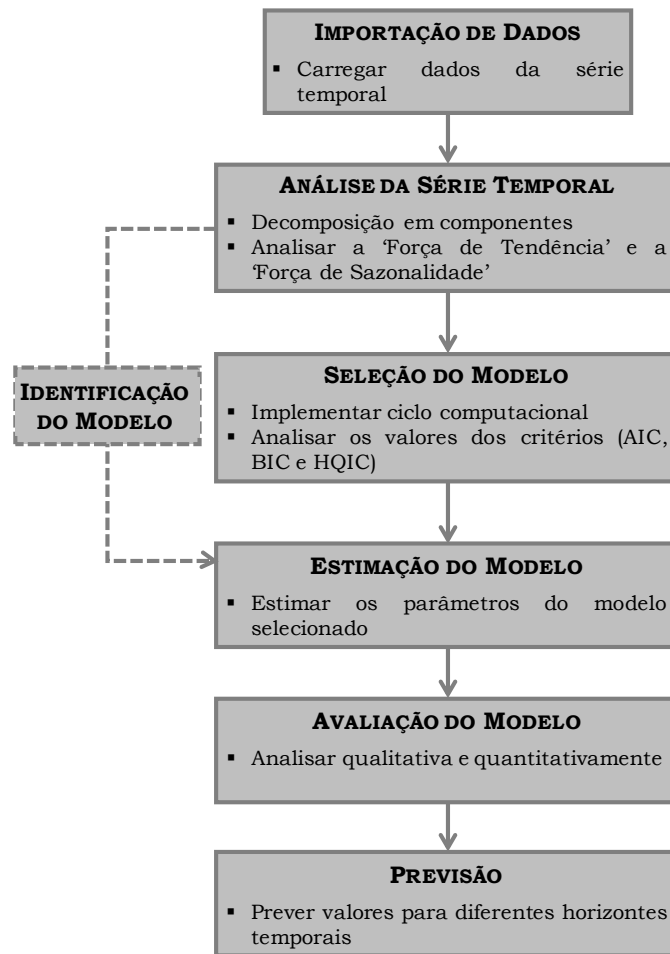
Para o estudo das séries temporais, utilizou-se o ambiente computacional *Jupyter Notebook*, com a linguagem de programação *Python* (em específico a versão 3.7.3). O código utilizado foi o desenvolvido em Ramos (2021), estando *open source* em Lopes e Ramos (2020). Metodologicamente, na Figura 1 estão apresentadas as etapas relativas à implementação computacional dos modelos ETS.

Conforme esquematizado, para desenvolver o estudo empírico, a primeira etapa passa pela análise e caracterização das séries temporais em estudo. Essa análise envolve:

1. o recurso a representações gráficas;
2. a análise de existência de quebras de estrutura;
3. a determinação de medidas descritivas;
4. a implementação de alguns testes estatísticos (normalidade, estacionariedade e independência).

Embora todos os detalhes relativos à implementação computacional, bibliotecas, funções e testes (de hipótese) possam ser consultados com detalhe em Ramos (2021), destacamos um aspeto fundamental – quebras de estrutura.

Figura 1. Metodologia de implementação computacional dos modelos ETS



Fonte: Ramos (2021)

Face a importância que a existência quebras de estrutura/mudanças de regime tem no processo de modelação e previsão de series temporais, considerou-se útil analisar a presença deste(s) fenómeno(s). Assim, para avaliar a existência de quebras de estrutura/mudanças de regime na série temporal, recorreremos ao algoritmo CUSUM (Brown et al., 1975), cuja implementação em *Python*, função `detect_cusum`, foi desenvolvida por Duarte e Watanabe (2018).

3.4. Forecasting e Avaliação do Erro (Métricas)

Para se obter uma previsão, ou seja, ao ajustar uma série de forma a projetá-la para além do período da amostra, surge o conceito de ‘erro de previsão’ (Ramos, 2021).

Assim, uma série temporal y no instante t (num intervalo de tempo T), $\{y_t\}$ com $t \in T$, sendo y_{t+h} o valor desconhecido no período futuro $t+h$ e \hat{y}_{t+h} a sua previsão obtida com base na informação disponível até ao momento t , o erro de previsão corresponde à diferença entre esses dois valores (ver equação (3)).

$$e_{t+h} = y_{t+h} - \hat{y}_{t+h} \quad (3)$$

Sendo o propósito avaliar o desempenho de um modelo de previsão, torna-se essencial analisar os valores de indicadores capazes de traduzir, em termos numéricos, a medida de erro. De entre as métricas de erro comumente referidas na bibliografia, destacamos o *Mean Absolute Percentage Error* (MAPE), como a métrica usada na avaliação dos modelos propostos neste trabalho. O MAPE é uma ‘grandeza adimensional’ que avalia a dimensão do erro em termos percentuais, sendo que, em termos matemáticos, é definido pela equação (4), onde s denota o número de previsões a efetuar (janela de previsão).

$$MAPE = \frac{\sum_{i=1}^s \left| \frac{y_{t+i} - \hat{y}_{t+i}}{y_{t+i}} \right|}{s} \times 100 \tag{4}$$

4. Resultados

Traçada uma perspetiva teórica, bem como os procedimentos metodológicos seguidos e feita uma apresentação sucinta das séries temporais em estudo, na presente secção, após a sua análise exploratória, passar-se-á ao processo de modelação e previsão.

4.1. Análise exploratória das séries temporais

As séries temporais consideradas respeitam aos dados mensais relativos a dormidas em estabelecimentos de alojamento turístico em Portugal entre janeiro de 2002 e abril de 2022, sendo que a série “Residentes” corresponde ao número de dormidas de residentes em Portugal e a série “Total” corresponde ao número total de dormidas ($\times 10^7$), dado pela soma das dormidas de residentes em Portugal com as de turistas residentes no estrangeiro. Os dados históricos relativos às séries apresentam-se nas Figuras 2 e 3.

Figura 2. Série Total ($\times 10^7$): Representação gráfica (por anos)

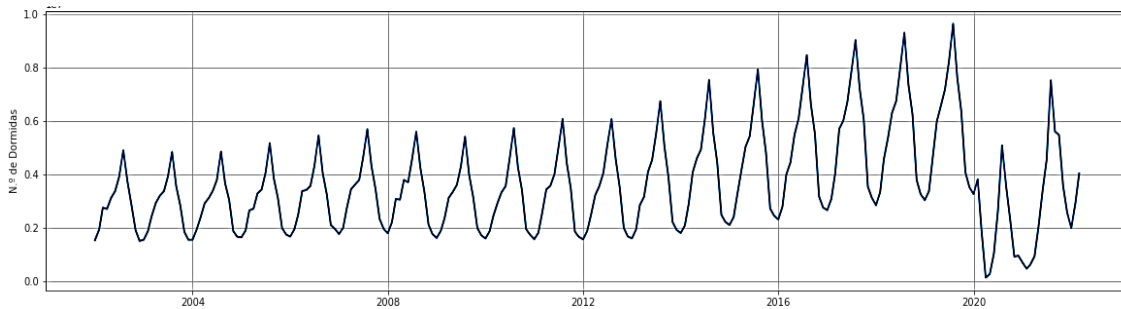
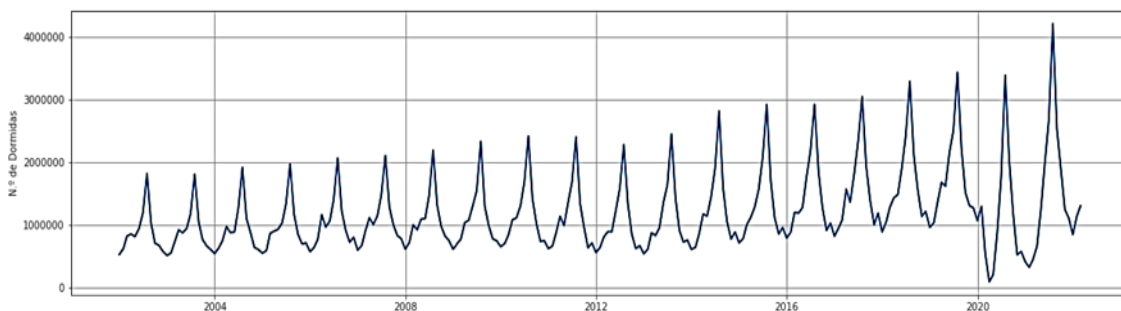


Figura 3. Série Residentes: Representação gráfica (por anos)



Das representações gráficas depreende-se em ambas as series uma tendência linear crescente, com evidências claras de comportamento sazonal (ao longo do ano) e um aumento da variabilidade ao longo dos anos, destacando-se uma quebra evidente em 2020, originando uma mudança de regime em 2020 e período subsequente.

Através da representação de *box-plots* anuais agrupados (Figuras 4 e 5), verifica-se um aumento gradual da amplitude amostral e interquartil (nomeadamente desde 2011 até 2019) com a existência de observações *outliers* na série Residentes (mês de férias de Verão). Em 2020 devido à pandemia, com o fecho de fronteiras por todo o mundo, as implicações no fluxo de Turismo foram evidentes.

Assim, pode-se verificar na Figura 4 e na Figura 5 uma quebra abrupta em 2020 e a existência de indícios de uma recuperação em 2021 e nos primeiros três meses de 2022 (de reforçar que em 2022 só foram considerados os primeiros três meses do ano).

Figura 4. Série Total ($\times 10^7$): Representação gráfica dos *box-plots* anuais

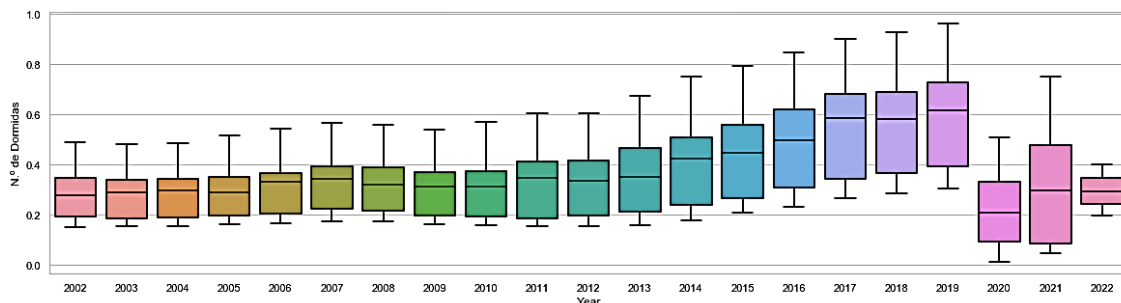
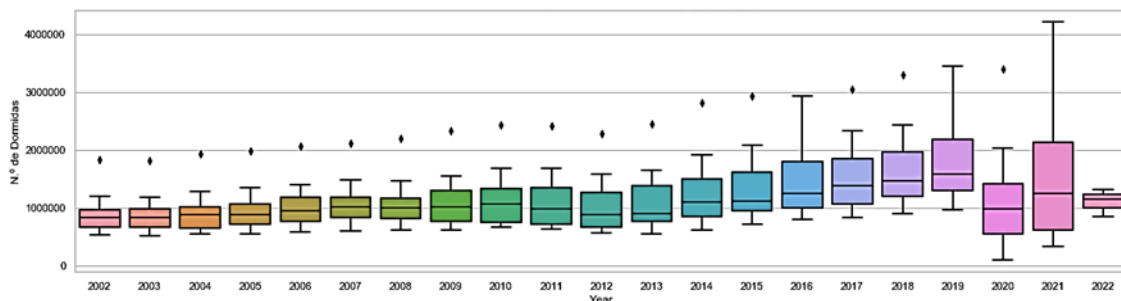


Figura 5. Série Residentes: Representação gráfica dos *box-plots* anuais



Pela representação gráfica dos *box-plots* mensais (Figura 6 e Figura 7) é evidente o comportamento sazonal das séries: o mês com maior número de dormidas é agosto, sendo os mais baixos os meses de janeiro e dezembro. De referir, ainda, a existência de observações *outliers* nos cinco primeiros meses, parte delas consequente dos confinamentos e das restrições de circulação resultantes da pandemia (anos de 2020 e 2021).

Figura 6. Série Total($\times 10^7$): Representação gráfica dos *box-plots* mensais

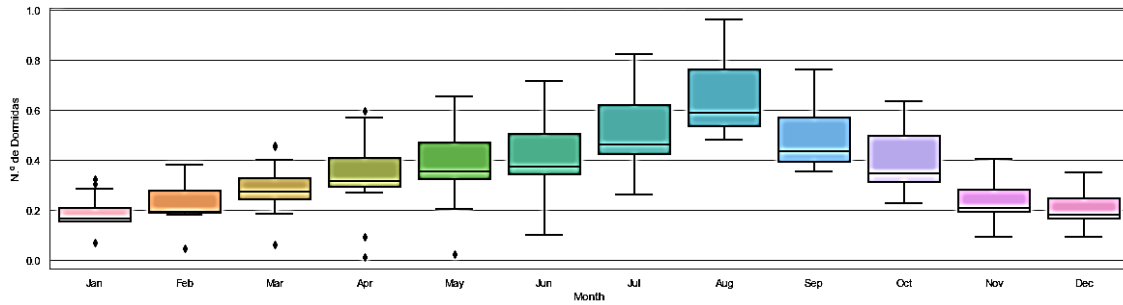
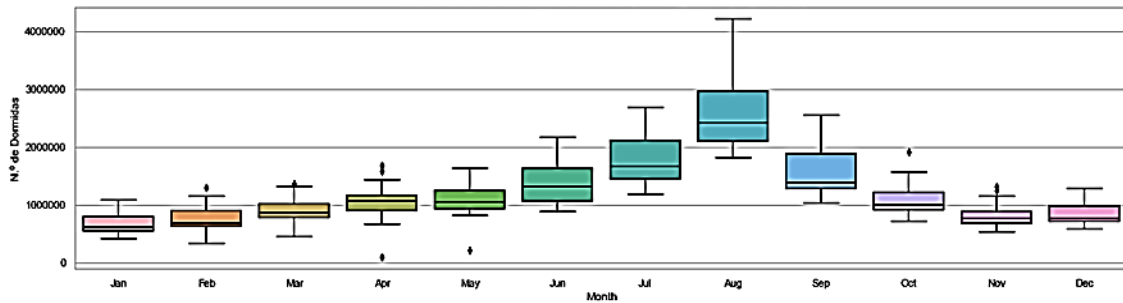


Figura 7. Série Residentes: Representação gráfica dos *box-plots* mensais



As principais estatísticas descritivas das séries temporais são apresentadas na Tabela 2, onde se pode verificar que, em ambas as séries, o valor máximo corresponde às dormidas em agosto de 2019 e o valor mínimo reflete as dormidas em abril de 2020.

Tabela 2. Estatísticas descritivas

	<i>Count</i>	<i>Mean</i>	<i>Std</i>	<i>Min</i>	<i>Q1</i>	<i>Q2</i>	<i>Q3</i>	<i>Max</i>	<i>Kurtosis</i>	<i>Skewness</i>
Série Total	243	3644629	1817023	133212	2208905	3365240	4583488	9633427	0.4438	0.8370
Série Residentes	243	1201920	625718	96673	77319.5	1035100	1400719	4214577	3.2374	1.6136

As duas séries têm uma média superior à mediana, impacto esperado no valor da assimetria (*skewness*) que é positivo – distribuição assimétrica à direita. Relativamente à curtose (*kurtosis*) verifica-se em ambos os casos uma distribuição leptocúrtica do número de dormidas. Estas informações reforçam algumas características observáveis na curva de densidade resultante da distribuição do número de dormidas (Figura 8 e Figura 9).

Figura 8. Série Total ($\times 10^7$): Histograma (com curva de densidade)

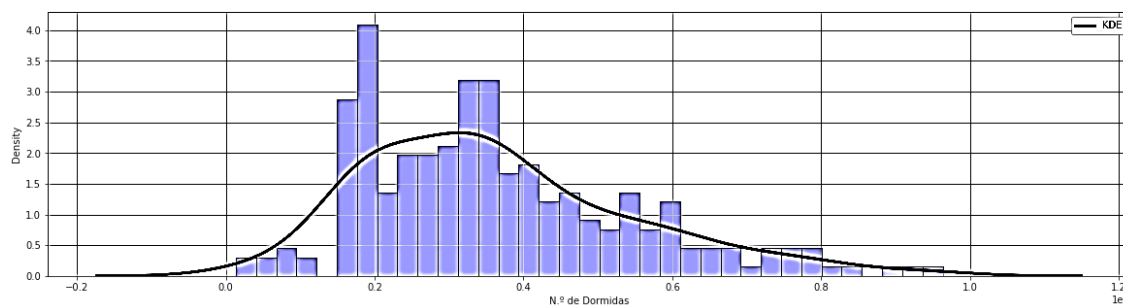
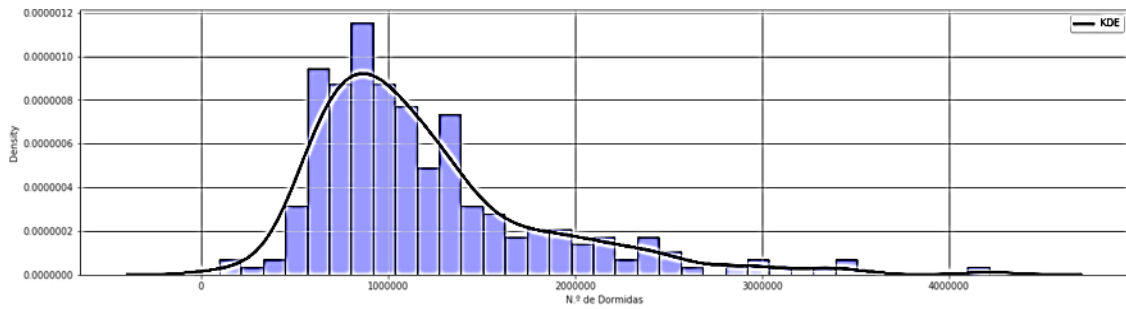


Figura 9. Série Residentes: Histograma (com curva de densidade)



Na Tabela 3 apresentam-se os resultados dos testes de hipóteses referentes à normalidade, estacionariedade e independência (sob hipótese nula de que os dados são independentes e identicamente distribuídos) para cada uma das séries em estudo.

Tabela 3. Testes de normalidade, estacionariedade e independência

	Testes de Normalidade			Testes de Raiz Unitária/ Estacionariedade		Teste de Independência
	<i>Kurtosis</i>	<i>Skewness</i>	<i>Jarque-Bera</i>	ADF	KPSS	BDS (Dim. 2 – Dim. 6)
Série Total						
<i>statistic</i>	1.4453	4.8491	30.3689	-2.8783	0.5982	18.8298 – 22.1105
<i>p-value</i>	0.1484	0.0000*	0.0000*	0.0479	-----	0.0000*
Série Residentes						
<i>statistic</i>	4.8882	7.7890	211.56449	-1.3561	1.0849	5.0778 – 9.7241
<i>p-value</i>	0.0000*	0.0000*	0.0000*	0.6031	-----	0.0000*

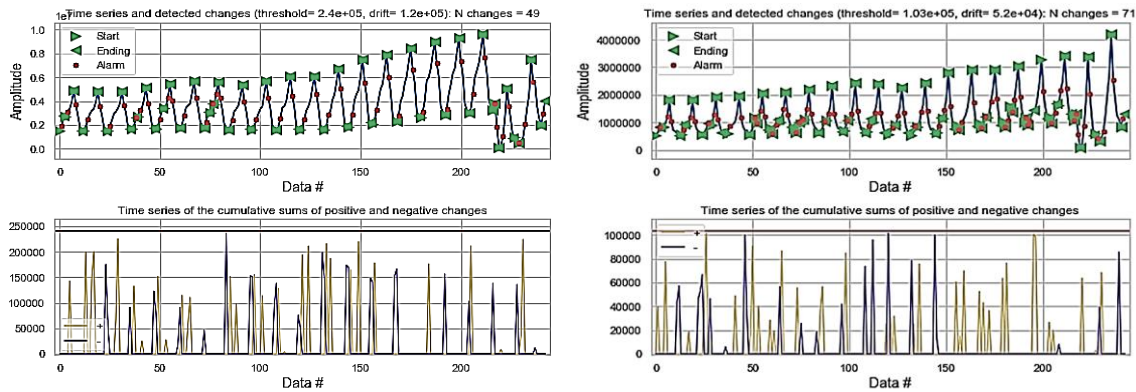
*Rejeita-se H0 para os níveis de significância de 1%, 5% e 10%

Alinhado com a literatura, face à importância que a presença de quebras de estrutura pode ter no processo de modelação e previsão, foi analisada a presença de quebras de estrutura com a aplicação do algoritmo CUSUM, onde identificamos quebras de estrutura constantes que resultam, na sua maioria, das alterações de comportamento sazonal, conforme representado na Figura 10. Porém, no seguimento do já exposto, de salientar, uma vez mais, a mudança de regime observada em cada uma das séries (embora com impactos distintos) na sequência da pandemia em 2020.

Figura 10. Mudanças/quebras estruturais: (A) Série Total; (B) Série Residentes

(A)

(B)



4.2. Modelação e Previsão: Alisamento Exponencial (ETS)

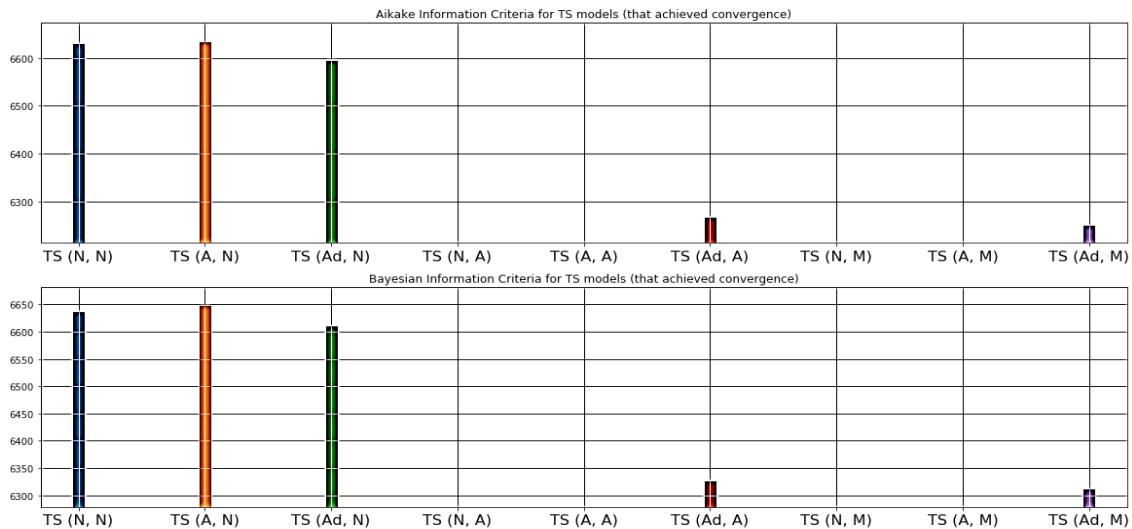
Para a modelação e previsão usando metodologia ETS, considerou-se útil o cálculo da “força de tendência” (\mathcal{F}_T) e “força de sazonalidade” (\mathcal{F}_S). Conforme descrito na secção 3.2., os valores observados evidenciam que ambas as séries apresentam uma forte tendência (valores de \mathcal{F}_T iguais a 1 numa vertente multiplicativa) e uma forte sazonalidade (valores de \mathcal{F}_S próximos de 1), factos que confirmam o referido anteriormente: da análise da representação gráfica de cada uma das séries temporais (Figura 2 e Figura 3) depreendem-se evidencias tendência e sazonalidade.

Tabela 4. Valores da “força de tendência” e da “força de sazonalidade”

	Série Total		Série Residentes	
	\mathcal{F}_T	\mathcal{F}_S	\mathcal{F}_T	\mathcal{F}_S
Modelo Aditivo	0.7232	0.8536	0.1882	0.8247
Modelo Multiplicativo	1.0000	0.8141	1.0000	0.8706

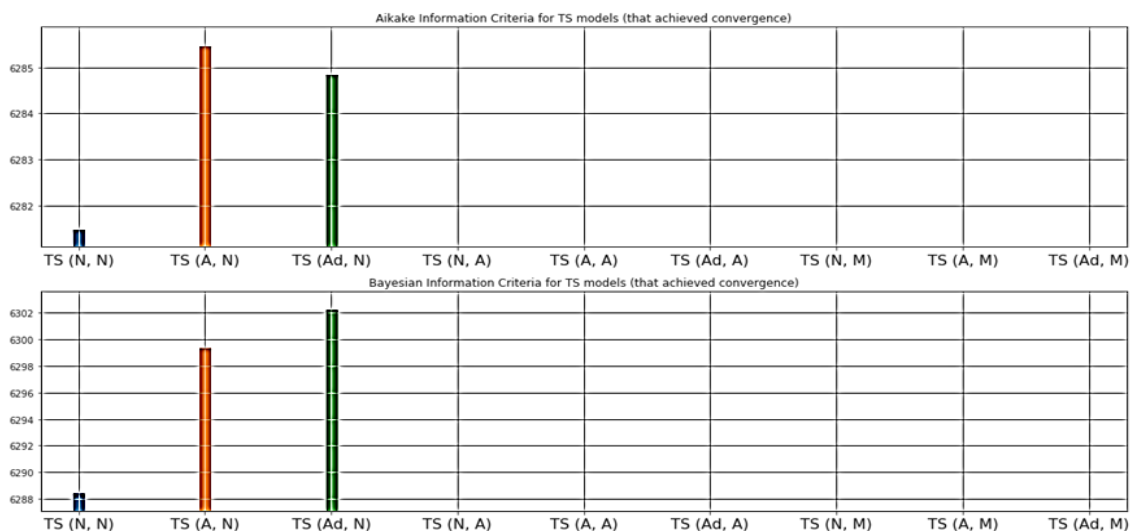
Seguindo os procedimentos de implementação computacional descritos na secção 3.3., para identificar os modelos ETS mais apropriados (de entre os modelos que apresentam uma convergência dos parâmetros, para efeitos de previsão), de acordo com os critérios de seleção de AIC e BIC, conforme Figura 11, os mais adequados para a série Total são: TS (Ad, M) e TS (Ad, A) (graficamente, corresponde às barras com menor valor – quanto menor o valor, mais adequado é o modelo (ver Akaike (1974) e Schwarz (1978))).

Figura 11. Série Total: Critérios de Informação para modelos ETS



Por outro lado, para a série Residentes (Figura 12), selecionamos o modelo TS (N, N), de entre os possíveis modelos (os que apresentam uma convergência dos parâmetros), que não conta com tendência nem sazonalidade da série.⁶

Figura 12. Série Residentes: Critérios de Informação para modelos ETS



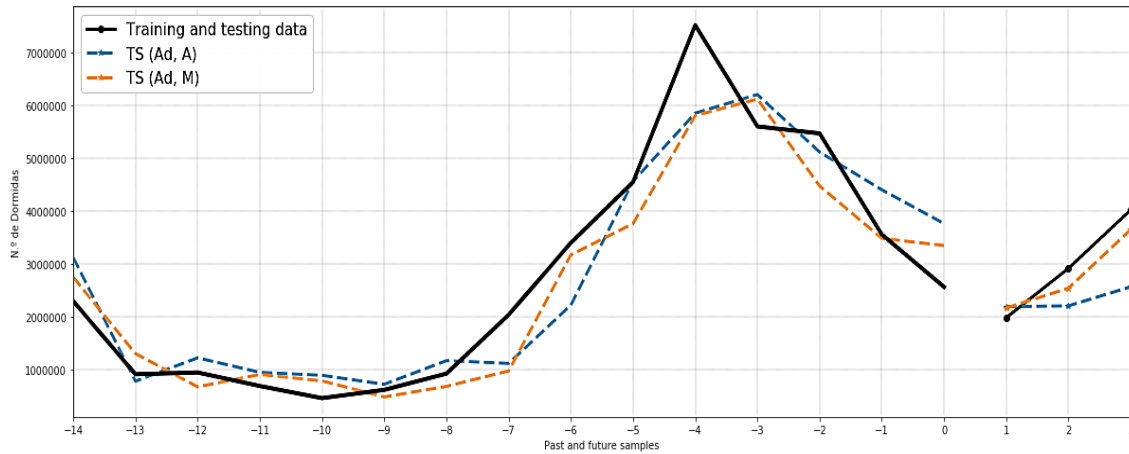
Sendo um dos objetivos avaliar o ‘ajustamento’ global dos modelos ETS, considerámos útil analisar a qualidade preditiva dos modelos identificados pelos critérios descritos como sendo (de entre aqueles cujos parâmetros convergem) os potencialmente mais adequados. Para o efeito, foram obtidas as estimativas de previsão *out-of-sample* produzidas, por cada modelo para o primeiro trimestre de 2022 (1º mês – janeiro; 2º mês – fevereiro e 3º mês – março).

Na Figura 13 e Figura 14 encontram-se as respetivas representações gráficas (para cada série), onde, além da comparação dos valores preditos (*forecasting*) com os que efetivamente foram observados, considerámos útil representar (com uma janela temporal das últimas 15 observações) os valores ajustados pelo modelo (*fitting*), donde se pode aferir a adequabilidade de cada modelo aos dados reais (previsões *in-sample*).

⁶ Características que contradizem o comportamento da série, comprometendo-se a qualidade preditiva.

Relativamente à série Total, verificamos na Figura 13 que existe um ajuste bastante satisfatório entre a linha de previsão de ambos os modelos considerados e a flutuação global dos dados reais (linha preta), dado que a linha de previsão acompanha, em geral, a tendência da série.

Figura 13. Série Total: Modelos ETS (fitting e forecasting)

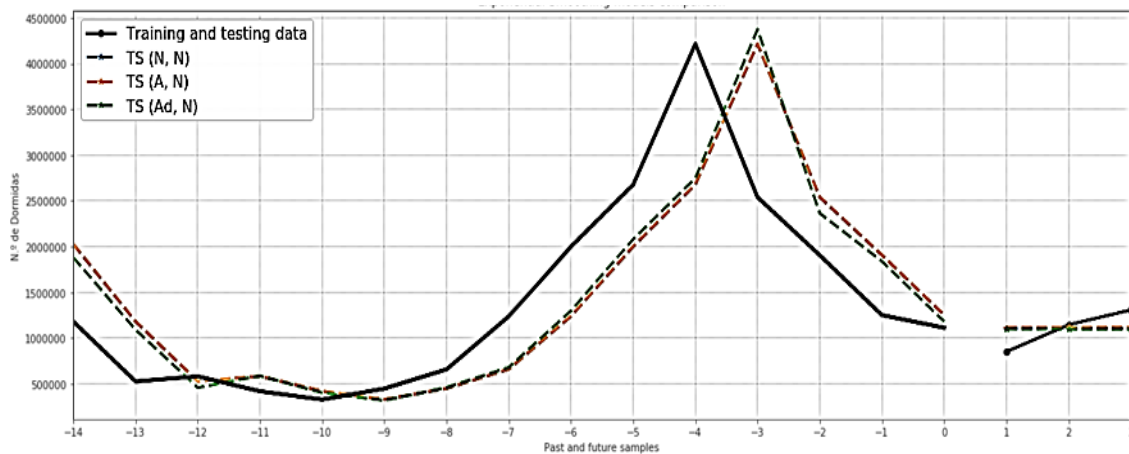


Globalmente, pode verificar-se que a linha do modelo TS (Ad, M), a linha cor de laranja, está mais próxima dos valores reais.

Na série Residentes, dado que foi utilizado um modelo sem tendência e/ou sazonalidade, verifica-se na Figura 14 que a previsão é praticamente um valor constante.

Comparando as previsões nos vários modelos identificados (e possíveis), selecionou-se, para cada série, um modelo considerado globalmente mais adequado (e aquele cujos valores do MAPE eram menores): para a série Total selecionou-se o modelo TS (Ad, M) e para a série Residentes o modelo TS (N, N).

Figura 14. Série Residentes: Modelos ETS (fitting e forecasting)



Na Figura 15 e Figura 16 são apresentados os gráficos relativos aos resíduos correspondentes ao modelo selecionado para cada caso.

Figura 15. Série Total: Resíduos de modelos ETS

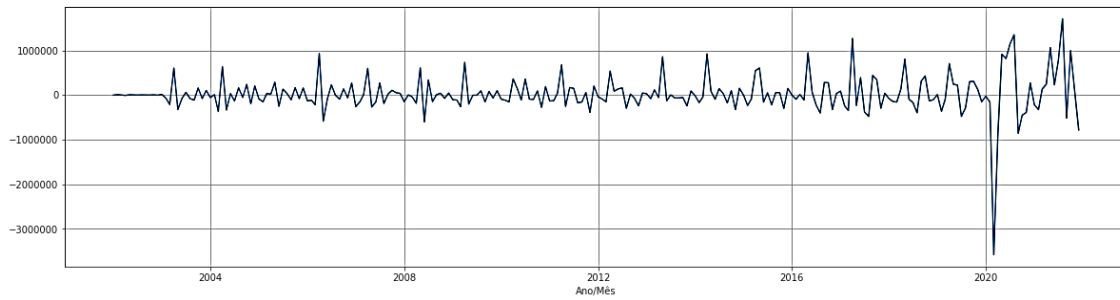
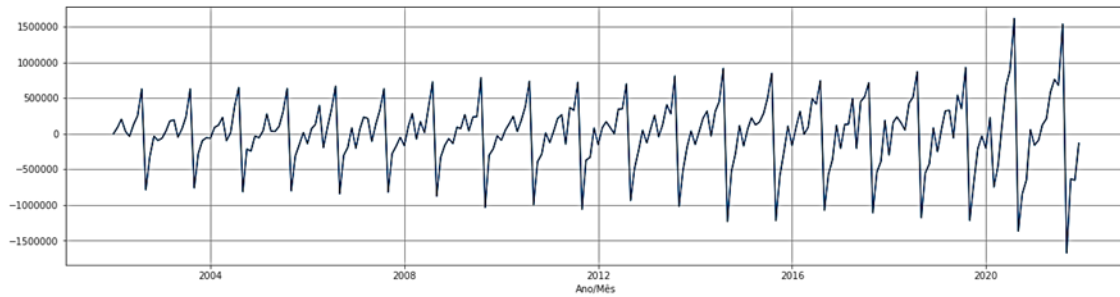


Figura 16. Série Residentes: Resíduos de modelos ETS



Na Figura 15 (Série Total) identificamos, globalmente, um ruído branco, existindo um pico que foge ao padrão no ano 2020. Contudo, na Figura 16 (série Residentes), verifica-se um evidente comportamento sazonal nos resíduos, uma vez que o modelo selecionado não capta esta dinâmica (evidenciando limitações das metodologias ETS).

Para uma avaliação mais robusta dos modelos selecionados, apresentamos na Tabela 5 os valores reais registados nos meses utilizados para avaliação dos modelos, bem como os valores previstos, para cada caso, calculando-se os valores correspondentes ao MAPE para as previsões (*out-of-sample*).

Analisando os valores apresentados, se, para efeitos de previsão da série Total, o modelo ETS selecionado, TS (Ad, M), corresponde de forma positiva, o mesmo não se verificou para a série Residentes. Esta série, por apresentar uma dinâmica recente e completamente atípica, verifica-se que os modelos ETS potencialmente mais adequados (modelos com tendência e sazonalidade), não apresentam, em geral, uma convergência dos respetivos parâmetros para efeitos de previsão. Os únicos modelos que apresentam convergência são modelos sem componente sazonal, o que não é concordante com o histórico de dados, comprometendo, assim, a qualidade preditiva.

Tabela 5. Valores reais e valores previstos pelos modelos ETS (e valor do MAPE)

	Série Total			Série Residentes		
	Real	Previsão ETS	MAPE	Real	Previsão ETS	MAPE
Janeiro	1988869	2173621	9.29%	851424	1113740	30.81%
Fevereiro	2920471	2641112	11.05%	1145829	998626.5	16.8%
Março	4023327	3713308	10.44%	1307984	1145101	16.15%

5. Conclusões

Face à importância que o Turismo tem no tecido económico e social português, considerou-se pertinente a investigação desenvolvida, sobretudo na conjuntura atual, motivada pela COVID-19. Decorrente do período pandémico, onde se registaram quebras acentuadas no Turismo, a dimensão empresarial afeta ao setor sofreu consequências cujo impacto a nível económico e social foi inegável. Com efeito, para definição de uma estratégia de atuação futura e perceber como nos podemos reinventar, impõe-se uma análise aos dados históricos do fluxo de turismo, apontando previsões.

Alinhado com a Revisão da Literatura, os dados analisados demonstraram o quão Portugal estava dependente do fluxo de turismo externo, ficando os negócios associados ao Turismo reféns dos turistas estrangeiros. Com a pandemia, com todos os países a fecharem as fronteiras, o setor ficou obrigados a ‘sobreviver’ e a estar dependente de movimentações internas. Precisamente, uma análise cuidada ao histórico mais recente (últimos dois anos) permite-nos perceber que foram os próprios portugueses que seguraram alguma atividade turística no seu próprio país. Deste modo, tornou-se evidente a necessidade de uma gestão estratégica capaz de fomentar o Turismo interno, criando mecanismos de ‘não dependência’ externa. Ou seja, em termos estratégicos, sendo importante identificar alternativas, a fim de evitar falência e declínio dos negócios, há que realçar o interesse em não esquecer e fomentar políticas como o “Vá para fora cá dentro”.

No seguimento de um outro objetivo desta investigação, ficou claro, pela Revisão da Literatura, que a previsão constitui uma efetiva ferramenta de apoio à tomada de decisão. Como tal, interessa identificar metodologias de previsão robustas e capazes de dar resposta às oscilações bruscas ocorrentes nos dados históricos. Reconhecendo o aspeto redutor de se analisar o fluxo do Turismo com base no histórico do “número de dormidas em estabelecimentos de alojamento turístico em Portugal”, impunha-se a definição de uma estratégia de análise que, em certa medida, tornasse o fenómeno quantificável e que se identificasse com o fenómeno em estudo. Não obstante dessa limitação, certo é que as séries em estudo muito descrevem algumas das características de outras séries temporais afetas ao Turismo, com um padrão claro de tendência e/ou sazonalidade nos dados históricos. Sendo essas dinâmicas facilmente captáveis pelas metodologias clássicas, seria de esperar que os modelos ETS não só se ajustassem aos dados históricos, como produzissem boas previsões. Contudo, motivado por quebra de estrutura/mudança de regime decorrente da pandemia, o nosso estudo traz a tona a dificuldade das metodologias clássicas em lidar com situações que saem fora dos padrões. Embora este facto não seja novidade, a forma como ele surge neste estudo é merecedor de particular atenção.

Apesar da quebra verificada no fluxo de turismo total, como no recuperar existe uma certa semelhança com os dados históricos (com claro padrão de sazonalidade), os modelos ETS produziram previsões razoáveis, embora com alguma margem de erro. Porém, quando se centra particular atenção no fluxo de turismo interno (residentes), não só existe uma quebra, como existe uma clara mudança no fluxo de turismo, onde os dados históricos não parecem encaixar de modo algum na dinâmica atual – os portugueses começaram a fazer turismo interno em massa. Este fenómeno evidenciou limitações nos modelos ETS em produzir previsões precisas (agravadas pela não convergência, para efeitos de previsão, dos parâmetros de alguns modelos). Com efeito, partindo desta limitação, com

intuito de dar cumprimento ao objetivo de encontrar metodologias de previsão adequadas, surge a motivação na procura de metodologias de previsão alternativas.

Assim, numa perspetiva de desenvolvimentos futuros, procurar-se-á investir nessa direção, analisando a possibilidade e a existência de vantagens na implementação de metodologias de *Machine Learning*, nomeadamente de *Deep Learning*. Tal escolha é motivada e está alinhada com a literatura científica recente, onde estas metodologias têm sido apontadas como bastante promissoras face à ocorrência de fenómenos atípicos no conjunto de dados históricos (Ramos, 2021).

Em suma, apesar de se identificarem limitações, reconhecemos que o trabalho já desenvolvido se reverte de uma mais-valia, não só pelo estudo empírico desenvolvido, como, partindo das conclusões referidas, se identificam perspetivas e desenvolvimentos futuros, alinhados com a literatura científica atual, avaliando as potencialidades de metodologias de *Deep Learning*.

Referências Bibliográficas

- Akaike, H. (1974). A new look at the statistical model identification. In *Selected Papers of Hirotugu Akaike* (pp. 215–222). New York: Springer.
- Botelho, N. (2019). *Mais turistas que habitantes? Portugal faz parte de clube restrito*. Retrieved from <https://visao.sapo.pt/atualidade/economia/2019-02-08-Mais-turistas-que-habitantes--Portugal-faz-parte-de-clube-restrito/>
- Brown, R. L., Durbin, J., & Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, 37(2), 149-163.
- Bucho, D. (2010). *Património, Animação e Turismo*. Portalegre: Instituto Politécnico de Portalegre.
- Caldas, I., Machado, H., Sousa, B., & Vereiro, L. (2020). Um olhar exploratório sobre o turismo voluntário. *European Journal of Applied Business Management, IWTHM Special Issue*, 15-30.
- Casanova, S. (2012). Previsão da demanda turística da cidade de Foz do Iguaçu: Uma aplicação com os modelos ARIMA. *Revista Turismo, Visão e Ação*, 14(3), 366–385.
- Duarte, M., & Watanabe, R. N. (2018). *Notes on Scientific Computing for Biomechanics and Motor Control*. GitHub. Retrieved from: <https://github.com/BMClab/BMC>
- Goh, C., & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Management*, 23(5), 499–510.
- Gunter, U., & Önder, I. (2015). Forecasting international city tourism demand for Paris: Accuracy of uni- and multivariate models employing monthly data. *Tourism Management*, 46, 123–135. <https://doi.org/10.1016/j.tourman.2014.06.017>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice (2nd ed.)*. Melbourne, Austrália: OTexts. Retrieved from <https://otexts.com/fpp2/>
- INE (2022). *Estatísticas do Turismo*. Retrieved from: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACO

ESpagenumber=1&PUBLICACOEScolecao=107668&PUBLICACOESTipo=ea&selTab=tab0

- Kirovska, Z. (2011). Strategic management within the tourism and the world globalization. *Journal of Economics*, 2(1), 69–76.
- Lopes, D. R., & Ramos, F. R. (2020). Univariate time series forecast. Retrieved from <https://github.com/DidierRLopes/UnivariateTimeSeriesForecast>
- López, A. M. (2021). *Travel and Tourism in Portugal - statistics & facts*. Retrieved from <https://www.statista.com/topics/6863/travel-and-Tourism-in-portugal/>
- Machado, A., Sousa, B., Dieguez, T., Ribeiro, M. & Cunha, F. (2020). Empreendedorismo social, inovação e benchmarking no instagram: Combate aos efeitos negativos da COVID-19 numa visão luso-brasileira. *European Journal of Applied Business Management*, 6(2), 59-82.
- Pesaran, M. H., Pettenuzzo, D., & Timmermann, A. (2006). Forecasting time series subject to multiple structural breaks. *Review of Economic Studies*, 73(4), 1057–1084.
- Ramos, D. M., & Costa, C. M. (2017). Turismo: Tendências de evolução. *PRACS: Revista Eletrônica de Humanidades Do Curso de Ciências Sociais Da UNIFAP*, 10(1), 21–33.
- Ramos, F. R. (2021). *Data Science na Modelação e Previsão de Séries Económico-financeiras: das Metodologias Clássicas ao Deep Learning*. Instituto Universitário de Lisboa - ISCTE Business School, Lisboa, Portugal.
- Ramos, F. R., Fidalgo, A. & Ramires, A. (2019). Tourism as a growing force in a receding economy: The Oporto City case. *7th Annual Winter Global Business Conference*. Tignes.
- Simões, P., Costa, J., Provenza, M., Xavier, V., & Goulart, J. (2019). Modelos de previsão para temperatura e salinidade no fenómeno de ressurgência: Análise dos dados da boia 19°00's34°00'w no período entre 2005 e 2014. *XIX Simpósio de Pesquisa Operacional e Logística Da Marinha*, 1–23.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464.
- Tealab, A. (2018). Time series forecasting using artificial neural networks methodologies: A systematic review. *Future Computing and Informatics Journal*, 3(2), 334–340.
- Turismo de Portugal (2013). *Vá para fora cá dentro*. Retrieved from <https://www.facebook.com/Vaparaforacadentro/>
- Zhang, H., Song, H., Wen, L., & Liu, C. (2021). Forecasting tourism recovery amid COVID-19. *Annals of Tourism Research*, 87, 1-10. <https://doi.org/10.1016/j.annals.2021.103149>