



LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

MESTRADO

MÉTODOS QUANTITATIVOS PARA A DECISÃO ECONÓMICA E EMPRESARIAL

TRABALHO FINAL DE MESTRADO

TRABALHO DE PROJETO

MODELAÇÃO E PREVISÃO DA OCUPAÇÃO EM UNIDADES
HOTELEIRAS. O CASO DO PESTANA HOTEL GROUP

JOANA RITA ALVES SÍTIMA

MAIO-2018



LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

MESTRADO EM
MÉTODOS QUANTITATIVOS PARA A DECISÃO
ECONÓMICA E EMPRESARIAL

TRABALHO FINAL DE MESTRADO
TRABALHO DE PROJETO

MODELAÇÃO E PREVISÃO DA OCUPAÇÃO EM UNIDADES
HOTELEIRAS. O CASO DO PESTANA HOTEL GROUP

JOANA RITA ALVES SÍTIMA

ORIENTAÇÃO:

PROFESSORA DRA. ESMERALDA ARRANHADO RAMALHO

MAIO-2018

Agradecimentos

Em primeiro lugar, gostaria de agradecer à equipa de *Business Intelligence* do *Pestana Hotel Group*, em especial à Catarina Figueiredo e ao Edgar Silva por toda a ajuda, cooperação e incentivo que me deram no decorrer deste projeto.

Em segundo lugar, à minha orientadora, a Professora Doutora Esmeralda Ramalho, por toda a ajuda, paciência e dedicação que mostrou em orientar-me ao longo destes meses.

Por fim, gostaria de agradecer à minha família, ao Diogo e aos meus amigos pelo ânimo e encorajamento que me deram durante o desenvolvimento deste trabalho.

A todos o meu obrigada.

Resumo

Atualmente, o turismo revela-se um setor com alto relevo na economia nacional, e estando este em expansão, torna-se essencial perceber de que forma continuarão as atividades turísticas a desenvolver-se no futuro.

No sector particular da hotelaria, a previsão da ocupação é um tema fulcral para uma gestão eficaz. De facto, no âmbito de uma unidade hoteleira, saber com alta precisão quantos quartos vão estar ocupados, para um determinado dia, constitui um forte contributo para a gestão da receita.

O objetivo central deste projeto é obter as previsões relativas à ocupação das diversas unidades hoteleiras do *Pestana Hotel Group*. Este tema foi abordado como um problema de previsão com séries temporais, assentes em duas metodologias: modelação através de decomposição clássica e modelos da classe SARIMA.

O processo de modelação foi feito com auxílio dos *softwares Microsoft Excel 2016 e Eviews*. Em primeiro lugar, começou-se pela decomposição clássica, onde se assume que captando a tendência e a sazonalidade da série, é possível conjugá-las de forma a obter previsões. Sendo esta uma abordagem genérica, que pode ser aplicada indistintamente a diferentes unidades, foi utilizada para previsão no âmbito de cada uma das unidades da amostra. Ao invés da metodologia anterior, os modelos da classe SARIMA requerem uma aplicação diferenciada para cada uma das unidades hoteleiras, pelo que foram apenas utilizados num caso de estudo de um hotel relevante para o Grupo Pestana.

Ao comparar os resultados de previsão obtidos pelas duas metodologias para a unidade hoteleira em estudo, verifica-se que os modelos da classe SARIMA geram menores erros de previsão. De facto, o melhor modelo SARIMA exibiu um EPAM de 8% enquanto a decomposição produziu um EPAM de 22%.

Contudo, convém ter presente que a hotelaria é um mercado extremamente volátil, facilmente afetado por fatores extrínsecos e/ou intrínsecos, que fortemente influenciam as previsões. Exige-se assim cuidado na interpretação e uso das previsões produzidas.

Palavras-chave: previsão, séries temporais, modelação, hotelaria, ocupação

Abstract

Nowadays, touristic activities are a significant part of the Portuguese economy. Tourism is an expanding sector, for which it is crucial to predict growth and its implications for the future.

In the particular area of hotel management, occupancy's forecasting is essential for an efficient management of each unit. Thus, knowing precisely the number of occupied rooms for a given day is a major contribution for a good revenue management.

This project aims at obtaining the occupancy's forecast for a number of hotel units of the Pestana Hotel Group. This issue was approached as a forecasting problem with time series, based on two distinct methodologies: classical decomposition modelling and SARIMA models.

The modelling process was implemented with the support of the *Microsoft Excel 2016* and the *Eviews* softwares. At first, we began with classical decomposition, where it is assumed that, by capturing the time series's trend and seasonality, it is possible to combine them in order to obtain the desired forecasts. Because this approach is generic, in the sense that can be applied indistinctively to all the units in the sample, it is used to predict occupancy in the framework of each of the hotels in analysis. In contrast, the SARIMA models involve a different approach for each hotel. Therefore, their application is restricted to a case study concerning an hotel which is particularly relevant for Pestana Hotel Group.

When comparing the forecast results obtained by both methodologies, in the framework of the case study, SARIMA models produce more precise forecasts, with a MAPE (mean absolute percentage error) of 8% vs. a 22% MAPE obtained by the classical decomposition models.

Notwithstanding, is very important to highlight the extremely volatile behavior of the hospitality business, that is easily affected by extrinsic and/ or intrinsic factors, which have strong impact over forecasts. This has to be accounted for in both the interpretation and the use of the produced predictions.

Keywords: forecast, time series, modeling, hospitality, occupancy

Índice

Agradecimentos	i
Resumo	ii
Abstract.....	iii
Índice de tabelas	v
Índice de figuras	v
Glossário.....	vi
1. Introdução.....	1
2. Contextualização do Estudo	2
2.1. A indústria	2
2.2. A empresa	3
2.3. O projeto	5
3. Metodologias	6
3.1. Modelos de decomposição clássica de séries temporais.....	6
3.2. Modelos da classe SARIMA	9
4. Caracterização dos dados e das variáveis	12
4.1. Descrição do tratamento dos dados	12
4.2. Análise das principais características amostrais	15
4.2.1. Globalidade das unidades abrangidas	15
4.2.2. Unidade “caso de estudo”	17
5. Resultados empíricos.....	18
5.1. Modelação e previsão através de modelos clássicos de séries temporais.....	18
5.2. Modelação através de modelos da classe SARIMA	22
6. Conclusão e investigações futuras	26
7. Bibliografia.....	28
8. Anexos.....	29

Índice de tabelas

Tabela 1- Índices sazonais obtidos pelas duas metodologias	19
Tabela 2- Modelos da classe SARIMA selecionados	23
Tabela 3- Valores obtidos para os critérios de seleção	24
Tabela 4- Erros de previsão com base na amostra de treino	24

Índice de figuras

Figura 1- Cronograma da série de “caso de estudo” e respetiva linha de tendência.....	17
Figura 2- Previsão de OCC para a unidade “caso de estudo”	20
Figura 3- Análise dos erros por unidade	21
Figura 4- Previsão de OCC para a unidade “caso de estudo” pelo modelo SARIMA.....	25

Glossário

ADF: *Augmented Dickey-Fuller*

AIC: *Akaike information criterion*

ARR: Preço médio do quarto

BIC: *Bayesian information criterion ou Schwarz criterion*

EAM: Erro absoluto médio

EPAM: Erro percentual absoluto médio

EQM: Erro quadrático médio

FAC: Função autocorrelação

FACP: Função autocorrelação parcial

FoodRev: Receita obtida por restaurantes/bares das unidades hoteleiras

INE: Instituto Nacional de Estatística

OCC: Ocupação de uma unidade hoteleira

PIB: Produto Interno Bruto

REQM: Raiz do erro quadrático médio

RN: Noites dormidas

RoomRev: Receita obtida por venda de quartos

SARIMA: *Seasonal Autoregressive Integrated Moving Average*

TotalRev: Receita total

VAB: Valor acrescentado bruto

1. Introdução

Em Portugal, o turismo tem-se revelado como um dos setores mais dinâmicos da Economia. De facto, só no ano de 2016, o setor do turismo produziu cerca de 11,5 mil milhões de euros, o que se traduz em aproximadamente 7% do valor acrescentado bruto (VAB) produzido por todo o sistema económico português (Relvas, 2017). Dentro do setor turismo, a hotelaria assume uma grande visibilidade, o que torna relevante a realização de estudos neste âmbito, essencialmente vocacionados para a área da gestão hoteleira, onde o foco do negócio é a receita.

Um dos principais instrumentos de apoio à gestão da receita no ramo da hotelaria é a obtenção de valores de previsão, que funcionam como indicadores e alertas para períodos futuros. Para além disto, ter uma capacidade de antevisão do negócio permite a gestão plena de muitas outras variáveis, sobretudo o preço médio do quarto (ARR).

Neste contexto, o presente projeto tem como principal objetivo a obtenção de valores de previsão referentes à ocupação das diversas unidades hoteleiras do *Pestana Hotel Group*. Pretende-se com este estudo analisar os dados referentes a 66 unidades hoteleiras da cadeia, com o sentido de obter previsões para cada uma delas, para o ano de 2018, em termos de ocupação mensal e diária. Desta forma, a variável em estudo é a ocupação, sendo esta definida percentualmente como o quociente entre o número de quartos vendidos e o número de quartos disponíveis, numa determinada data. Porém, como a data de abertura de cada unidade é, naturalmente, diferente, o período de observações disponível para cada hotel é diferente. No entanto, o horizonte temporal definido para análise corresponde aos anos de 2010 a 2017, para o qual a empresa contém os dados informatizados.

As metodologias subjacentes ao projeto assentam em duas categorias principais: modelos clássicos de decomposição de séries temporais e modelos da classe SARIMA. Para todas as unidades envolvidas no projeto são obtidas previsões através da decomposição clássica, podendo a empresa consultá-las de forma automática em relatório. Por outro lado, para aplicar a modelação SARIMA, foi escolhida uma unidade hoteleira, de forte importância para o grupo (representou 4,4% das RN do grupo em 2017), que servirá de “caso de estudo, e para a qual se poderão comparar as previsões obtidas pelas duas classes de modelos. Esta modelação é feita apenas para um hotel do

grupo, uma vez que seria incomportável modelar todas as unidades hoteleiras através de modelos SARIMA, pois estes requerem uma modelação individual e intrínseca para cada série. Para a implementação das duas vertentes metodológicas, os dados são trabalhados com auxílio dos *softwares Excel e Eviews*. No âmbito dos modelos clássicos de decomposição, por solicitação da empresa, produziu-se uma aplicação para *Excel*. Esta apresenta a previsão de acordo com modelo que minimiza o erro de previsão, mediante a seleção da unidade hoteleira de interesse. Para isto, a aplicação recolhe automaticamente os dados mais recentes e atualiza os modelos e as respetivas previsões.

Note-se que poderia ainda ter sido considerada outra metodologia preditiva, baseada em regressão múltipla. Assim, recorrendo a variáveis *dummy* sazonais, a variáveis de tendência, e a regressores, seria possível obter previsões alternativas. Contudo, esta abordagem foi inviabilizada por não estarem disponíveis dados que pudessem ser utilizados como variáveis explicativas para a variável de interesse, tais como: o estado do clima nos diferentes destinos, dados de voos e fluxos de movimentos em aeroportos, índice de preços da concorrência, datas de eventos festivos, etc.

Em suma, pode aferir-se que este projeto será um forte contributo para o grupo, uma vez que os vários departamentos interessados poderão consultar de forma automática e rápida os valores de previsão mensais e diários respetivos para cada hotel. Isto permitirá auxiliar na aplicação de estratégias de *pricing* e de negócio, que conduzam à maximização da receita.

2. Contextualização do Estudo

Nesta secção é feita uma contextualização do estudo em termos do setor turismo, seguindo-se uma apresentação da empresa e conclui-se com a descrição detalhada do projeto.

2.1. A indústria

Quando se observa os valores registados em 2016 para o PIB nacional, percebe-se a relevância que o turismo está a assumir na Economia. Para além do forte contributo

do setor do turismo no VAB e no PIB, anteriormente mencionado para 2016, também se sabe que os turistas estão a despende mais: em 2016, os consumos realizados pelos turistas superaram os 23 mil milhões de euros, o que corresponde a cerca de 12% do PIB do país (Relvas, 2017). Dados os factos, Portugal apresenta-se como o segundo país da Europa onde o setor do turismo revela maior impacto na economia, só sendo ultrapassado por Malta.

Face a estes resultados, é importante conhecer na sua plenitude o negócio da hotelaria, uma vez que este detém um forte impacto no PIB nacional. Na gestão de unidades hoteleiras ou de uma cadeia de hotéis, a atenção está concentrada maioritariamente na receita gerada com a venda de quartos - *Room Revenue (RoomRev)* em que $Room\ Revenue = RN * ARR$, isto significa que a receita obtida pela venda dos quartos é resultado da multiplicação do número de noites (RN) pelo preço médio do quarto (ARR).¹

Ter uma visão na íntegra da empresa, dos mercados, e da indústria, irá permitir prestar serviços de extrema qualidade, atendendo às necessidades e gostos dos diversos tipos de hóspedes. Com uma gestão eficiente e equilibrada é também possível agir rapidamente às mudanças impostas pelo mercado ou pela concorrência. Assim, a previsão de determinadas variáveis é por isso um tema central a desenvolver no mundo empresarial. Atualmente, as empresas estão a investir fortemente em técnicas inovadoras, que permitem prever a evolução do negócio e compreender as causas que têm um impacto crítico no mesmo (Weatherford & Kimes, 2003). Certamente que, com previsões rigorosas e precisas é possível agir de forma mais célere às mudanças impostas pelas necessidades do mercado, sendo esta a força motriz deste projeto.

2.2. A empresa

O *Pestana Hotel Group*, conta com mais de quatro décadas de existência, e é atualmente uma das mais prestigiadas cadeias hoteleiras portuguesas. O grupo contém aproximadamente 90 hotéis dispersos mundialmente, campos de golfe, entre outros

¹ Note-se, no entanto que a maioria das unidades hoteleiras não limita o seu negócio apenas na venda de quartos como gerador de receita. Durante as estadias são disponibilizados aos hóspedes outros serviços (restaurantes, bares, spa, etc.) que contribuem para a totalidade da receita produzida por cada hotel- *Total Revenue (Total Rev)*, onde $Total\ Rev = Room\ Rev + Food\ Rev + Other_{services}$

serviços. Na indústria, são várias áreas de negócio nas quais o grupo se encontra presente: hotelaria, golfe, casinos, imobiliário turístico, distribuição turística, *timesharing*, indústria e serviços (Pestana.com, 2018). Contudo, é na vertente da hotelaria, que o *Pestana Hotel Group* detém um forte posicionamento no mercado. Após 40 anos de existência, o grupo continua a desenvolver-se e a investir de forma sustentável. Atualmente dentro do setor da hotelaria, o grupo conta com quatro marcas de excelência:

- *Pestana hotels and resorts*: é constituída por cerca de 50 hotéis de requinte, entre quatro e cinco estrelas, localizados em 3 continentes: América, Europa, África.
- *Pousadas de Portugal*: estão localizadas em propriedades históricas e icónicas do país. As pousadas permitem que os hóspedes usufruam de estadias em edifícios genuinamente portugueses desde castelos, palácios, fortes, mosteiros ou conventos.
- *Pestana Collection Hotels*: é constituída por seis unidades luxuosas localizadas em Lisboa, Cascais, Porto, e a mais recente em Amesterdão. Todas as propriedades estão localizadas em luxuosas edificações e monumentos históricos, situados em locais privilegiados.
- *Pestana CR7 Lifestyle Hotels*: é a marca mais recente do grupo e conjuga a vasta experiência hoteleira do grupo com o estilo de vida do futebolista Cristiano Ronaldo. Os hotéis desta marca estão localizados no Funchal e em Lisboa, apresentando um estilo moderno alusivo à vida de CR7.

Em cooperação com o *Pestana Hotel Group* o presente trabalho tem como principal objetivo comparar diferentes modelos de previsão para séries temporais, de forma a conseguir, com um erro reduzido, prever a ocupação das unidades. Formalmente, a ocupação em hotelaria é medida através da variável OCC, que se expressa pelo rácio entre o número de quartos vendidos e o número total de quartos disponíveis para uma determinada data:

$$OCC = \frac{\#quartos_{ocupados_{mês}}}{\#quartos_{disponíveis_{mês}} * 100 \quad (1)$$

Para efeitos de previsão, esta variável é preferível à variável “número de quartos ocupados” dado que, por ser um quociente, considera variações no número de quartos disponíveis. Desta forma, garante-se que, as unidades que expandiram o seu número de quartos, não ficam prejudicadas com registos históricos de ocupação inferiores.

2.3. O projeto

Este projeto pretende implementar modelos de teor estatístico para ocupação de unidades hoteleiras de forma automática. Isto permitirá que os *stakeholders* consigam de forma rápida consultar as previsões de ocupação detalhadamente para cada hotel.

Sabe-se que previsões precisas são cruciais para uma gestão e controlo da receita eficiente. (Lee, 1990) provou que um aumento de 10% na precisão da previsão no setor aéreo aumentou a receita em 0,5-3,0% em voos com elevada procura. Dado este facto, obter previsões detalhadas e precisas é um forte contributo para a maioria dos sistemas de gestão de receita, e sem as quais as advertências de preço e/ou de disponibilidade podem conduzir a conclusões imprecisas (McCartney, 2000).

Os dados tipicamente utilizados para previsão da ocupação de um hotel podem ser divididos em duas dimensões: a data em que a reserva é feita e a data em que efetivamente o hóspede usufrui do quarto. Estes dados são utilizados para dois tipos de modelos. Os primeiros têm por base dados históricos, que consideram apenas o número final de quartos (ou as chegadas) numa determinada data de estadia. Por outro lado, os modelos de reserva consideram o incremento de reservas ao longo do tempo, para uma noite de estadia específica (modelos de *pick-up*). Neste trabalho utilizam-se dados históricos.

Este tipo de dados pode ser utilizado para obter previsões através dos métodos clássicos de decomposição de séries temporais, tais como o alisamento exponencial, o método de médias móveis, ou a regressão linear (Weatherford & Kimes, 2003). Por outro lado (Sa, 1987) dedicou-se a estudar os modelos da classe SARIMA para obter previsões que auxiliassem na gestão do preço para um determinado voo. Contudo, os resultados obtidos não foram auspiciosos, pelo que o autor alterou o seu estudo para o uso de regressão múltipla. (Lee, 1990) também se dedicou ao estudo de modelos SARIMA para

desenvolver previsões com base em dados históricos de reservas. Assim, de acordo com os estudos produzidos na área e apresentados anteriormente, comparar os resultados de previsão de acordo com diferentes classes metodológicas, constitui também um dos objetivos deste projeto.

Por fim, torna-se importante frisar que, das duas metodologias em estudo, os modelos clássicos de séries temporais são de cariz genérico, utilizando procedimentos comuns a todas as séries. Desta forma, foi possível aplicá-los a todas as unidades hoteleiras abrangidas no estudo. Em resultado disto, foi criado um relatório em *Excel* que permite selecionar uma unidade e posteriormente apresentar, de forma automática, as previsões de OCC para o ano de 2018.

Para a aplicação da metodologia SARIMA foi selecionada apenas uma unidade que serve de “caso de estudo” ao longo do projeto, dado que cada unidade, potencialmente requereria um modelo SARIMA diferente.

Em geral, a partir dos valores de ocupação mensal fornecidos pelo grupo Pestana, produzem-se previsões mensais e diárias, através de uma aplicação *Excel*, a partir das quais, o grupo Pestana terá um conhecimento antecipado da ocupação das suas unidades, sendo este um apoio nas áreas de gestão de *pricing*, *revenue* e atribuição de *budgets*.

3. Metodologias

3.1. Modelos de decomposição clássica de séries temporais

Os métodos clássicos de séries temporais têm por base a decomposição em forças componentes das quais a tendência, a sazonalidade e outros movimentos oscilatórios são as principais, e às quais se junta a componente residual.

A tendência pode ser entendida como a inércia da série temporal, isto é, o caminho que a série segue no longo prazo. Esta componente abrange os movimentos que se manifestam de forma suave e consistente durante períodos longos (Murteira, Muller, & Turkman, 2000). A componente sazonal traça as variações em relação à tendência, que geralmente ocorrem durante um ano, revelando comportamentos repetidos ciclicamente. Estes movimentos repetidos, nem sempre seguem um padrão e, portanto, podem ter causas sociais ou naturais inerentes. Os movimentos oscilatórios estão

associados às fases intercaladas de expansão e recessão que afetam a economia. Contudo, apesar de poderem ser frequentes, estes movimentos não exibem qualquer periodicidade previamente definida, o que faz com que, para ciclos longos seja difícil a disjunção destes com a tendência. Por fim, a componente residual engloba fatores aleatórios que não são possíveis de modelar e prever, tais como catástrofes naturais, atentados terroristas, surtos ou epidemias, etc. (Murteira, Muller, & Turkman, 2000).

Os modelos clássicos combinam as forças descritas anteriormente de forma aditiva ou de forma multiplicativa. Considere-se T_t a tendência, S_t a sazonalidade, C_t os movimentos oscilatórios e E_t a parte residual. O modelo *standard* multiplicativo é:

$$OCC_t = T_t * S_t * C_t * E_t \quad (2)$$

No presente estudo, o modelo acima, revela-se potencialmente adequado, pois regra geral é apropriado para situações em que a amplitude da componente sazonal aumenta ou diminui de acordo com a tendência da série.

Seguidamente descreve-se de forma breve a estimação das componentes de tendência e de sazonalidade, por duas vias.

Obtenção da tendência por regressão linear

Numa primeira abordagem, a tendência pode ser entendida através do comportamento de uma reta, na qual o sinal do declive fornece informação sobre a orientação positiva ou negativa da série, o que pode ser indicador de períodos de crescimento ou de recessão. Posto isto, um modelo a considerar é (2), em que a tendência é obtida por: $OCC_t = \alpha + \beta t + u_t$ onde os parâmetros α e β são estimados pelo método dos mínimos quadrados (Murteira, Muller, & Turkman, 2000).

Obtenção da tendência por médias móveis

Outro método frequentemente utilizado para a obtenção da tendência é o das médias móveis. Neste caso, as médias móveis permitem suavizar os dados referentes à ocupação para formar um indicador de tendência. Como o próprio nome indica, uma

média móvel é uma média que se movimenta, no sentido em que os dados mais antigos vão sendo desconsiderados à medida que os mais recentes vão estando disponíveis. A ordem de médias móveis, $MM(OCC)$, a utilizar depende da frequência dos dados, sendo neste caso aconselhada a ordem 12, por se tratarem de observações mensais. Assim, uma das vertentes a considerar para o modelo (2) assume que a tendência é descrita por $T_t = MM(OCC)_{12}$.

Obtenção da componente sazonal

Uma das formas de obter esta componente remete para o cálculo dos índices sazonais. Estes índices indicam a diferença percentual entre a média dos valores em cada período e a média de todos os períodos (Neves, Cruz, & Correia, 2014). Ao serem trabalhados dados mensais, é de esperar a obtenção de doze índices. Contudo, a obtenção dos índices sazonais varia consoante a forma como é obtida a tendência. Quando esta é captada através de regressão linear, então têm-se que: $OCC_t = T_t * S_t * E_t$ onde, $\hat{T}_t = \hat{\alpha} + \hat{\beta}t$, e, portanto $\frac{OCC_t}{\hat{T}_t} = S_t * E_t$. Assim, quando a tendência é obtida recorrendo a regressão linear, os respetivos fatores sazonais são obtidos dividindo a série original pela tendência estimada, calculando-se em seguida a média de cada mês ao longo da série (Murteira & Black, 1993):

$$I_{jan} = \frac{1}{7} (OCC_{jan2011} + \dots + OCC_{jan2017})$$

...

$$I_{dez} = \frac{1}{7} (OCC_{dez2011} + \dots + OCC_{dez2017})$$

Por outro lado, quando a tendência é captada por médias móveis, a obtenção dos índices sazonais para cada mês, pode ser obtida através dos seguintes passos (Caiado, 2012):

1. Alisamento da série de OCC através do cálculo das médias móveis de ordem 12 (métrica já calculada para obtenção da tendência - secção 3.1.2.);

2. Cálculo do rácio entre a série original e a série alisada;
3. Determinação dos índices sazonais para cada um dos 12 meses: $(I_{jan}, I_{fev}, \dots, I_{dez})$.
4. Obtenção dos 12 fatores sazonais (S_t) pelo quociente entre cada um dos índices e a sua média geométrica.

$$S_{jan} = \frac{I_{jan}}{\sqrt[12]{I_{jan} * I_{fev} * \dots * I_{dez}}}$$

...

$$S_{dez} = \frac{I_{dez}}{\sqrt[12]{I_{jan} * I_{fev} * \dots * I_{dez}}}$$

Após captadas as forças constituintes da série, nomeadamente a tendência e a sazonalidade, é possível obter previsões, conjugando-as através do modelo multiplicativo (2).

3.2. Modelos da classe SARIMA

Na abordagem de modelos de classe SARIMA a metodologia a considerar neste projeto é a Box-Jenkins (Box et al. (1994)). Esta é aplicada iterativamente, segundo 3 etapas fulcrais: identificação do modelo, estimação dos parâmetros e avaliação de diagnóstico.

Etapa 1: identificação do modelo

Nesta fase pretende-se encontrar os valores de (p, d, q) e $(P, D, Q)_{12}$, onde p representa a ordem dos termos autorregressivos, d o número de diferenciações, q a ordem dos termos das médias móveis, P a ordem dos termos autorregressivos sazonais (múltiplos de 12), D o número de diferenciações sazonais e Q a ordem dos termos das médias móveis sazonais (múltiplos de 12). A escolha destes valores é feita com base na análise gráfica das funções autocorrelação e autocorrelação parcial. Contudo, para iniciar

esta fase é necessário confirmar que se está a trabalhar com uma série estacionária. Este princípio é avaliado recorrendo ao teste aumentado de Dickey-Fuller (ADF), que permite detetar a presença de raízes unitárias na série. Note-se que a presença de pelo menos uma raiz unitária é um sinal indicativo de que a série não é estacionária. Posto isto, se o resultado do teste revelar que existe evidência estatística de que a série tem pelo menos uma raiz unitária, a metodologia Box-Jenkins recomenda a aplicação das primeiras diferenças para obter uma série estacionária, o que significa que o parâmetro d se passa a definir como 1. Por outro lado, a ordem de diferenciação referente à sazonalidade foi encontrada por tentativas, baseadas na observação das funções de autocorrelação e autocorrelação parcial.

Uma vez obtida a estacionaridade, é necessário observar as funções autocorrelação e autocorrelação parcial de forma a conseguir encontrar os valores (p, q) e (P, Q) que servirão para modelar a série.

Etapa 2: estimação do modelo

Após a escolha dos modelos passíveis de descrever o comportamento da série, está-se em condições de proceder à respetiva estimação dos parâmetros do modelo. Esta fase é feita com o auxílio do *software Eviews*, que apura uma estimativa dos parâmetros do modelo, recorrendo ao método de máxima verosimilhança - *ARIMA maximum Likelihood*.

Etapa 3: diagnóstico

Por fim, após a estimação dos modelos candidatos, é necessário efetuar uma avaliação que meça a qualidade das estimativas obtidas e do seu ajustamento. Esta avaliação é feita com base na análise dos resíduos do modelo, que segundo as hipóteses subjacentes, estes deverão assumir um comportamento próximo de um ruído branco.

Box e *Ljung* sugerem o teste de *Portmanteu* para verificar a correlação dos resíduos. A hipótese nula subjacente ao teste estabelece que: $H_0 = \rho_1 = \rho_2 = \dots = \rho_k = 0$ em que a estatística de teste é dada por $Q = n * (n + 2) * \sum_{j=1}^k \frac{\hat{\rho}_j^2}{n-j}$, que sob

H_0 segue uma distribuição assintótica de qui-quadrado com $k - m$ graus de liberdade, sendo m o número de parâmetros estimados e j o intervalo entre as observações. Posto isto, deverão ser analisados cuidadosamente todos os *lags* residuais, de forma a garantir que os *p-values* referentes ao teste de *Portmanteau* são suficientemente altos, de modo a que a hipótese nula não seja rejeitada. Contudo, se efetivamente o modelo não for adequado, será necessário retornar à etapa inicial e repetir iterativamente todo o procedimento, de forma a obter modelos mais razoáveis. Se por outro lado, o modelo apresentar resultados satisfatórios na etapa de diagnóstico, então procede-se à fase de previsão.

Processo de seleção

Decorridas as três etapas mencionadas anteriormente, pode ser obtido mais do que um modelo que cumpra todos os pressupostos, e que, portanto, seja adequado para explicar o comportamento da série. Assim, é necessário recorrer a critérios de seleção que auxiliem na escolha do “melhor” modelo. No presente projeto, são considerados os critérios de seleção mais usuais, AIC e BIC, em que o modelo preferível é o que apresenta um valor mais reduzido para estes critérios.

Previsão

A previsão para a variável OCC representa o objetivo fulcral deste projeto. Note-se que a previsão da ocupação para os momentos futuros m , é descrita por:

$$\widehat{OCC}_T(m) = P_T(m) = E(OCC_{T+m} | OCC_T, OCC_{T-1}, OCC_{T-2}, \dots)$$

Em que $P_T(m)$ representa o preditor de OCC_{T+m} , T o ponto de origem de previsão e m caracteriza o horizonte temporal de previsão. É importante salientar que, este último está relacionado com a amostra: a previsão estática (feita a um passo com os verdadeiros valores do período anterior) ou a previsão dinâmica que é aplicada a h passos à frente, usando os valores previstos (Caiado, 2012).

Neste projeto apenas será desenvolvida a previsão dinâmica (h passos à frente).

Segundo esta metodologia preditiva, os valores de previsão para o momento futuro $T + h$, são obtidos considerando apenas as observações até ao instante T . Posteriormente, recorre-se às observações até ao instante $T + 1$ para se obter a previsão para o momento $T + h + 1$, e assim sucessivamente (Caiado, 2012). No final das consecutivas iterações, é necessária uma avaliação de diagnóstico, que deverá ser sustentada pelo cálculo dos erros de previsão. Estes erros são calculados de acordo com os valores preditivos obtidos para a amostra de teste e com os valores reais da amostra.

Por fim, quando filtrados os modelos com menor erro de previsão (erro percentual absoluto médio - EPAM), obtêm-se as previsões da variável de interesse para o horizonte temporal pretendido.

4. Caracterização dos dados e das variáveis

4.1. Descrição do tratamento dos dados

De acordo com o referido na secção 1, os dados foram obtidos em parceria com o *Pestana Hotel Group* e dizem respeito à ocupação, dada pela equação (1), para as diferentes unidades hoteleiras. Os dados contêm informação referente a 66 hotéis, especificados no Anexo I, com os respetivos valores mensais de OCC. Note-se que as unidades hoteleiras não abriram todas no mesmo ano, o que conseqüentemente implica que cada amostra tenha uma dimensão distinta. Os registos mais antigos dizem respeito ao ano de 2010 e prolongam-se até ao ano 2017.

Com base no que é explicitado na literatura, as previsões em hotelaria podem ser feitas em termos de noites dormidas (RN) em vez de ocupação percentual. Porém, a conselho da empresa, foi utilizada a variável OCC, pois esta acrescenta informação que a variável RN não consegue alcançar. Isto é, OCC consegue captar o aumento/diminuição do número de quartos disponíveis, por se tratar de um quociente. Contudo, dado o interesse em obter a previsão para RN, obteve-se esta variável através de $RN = OCC * \#quartos\ disponíveis$.

Por outro lado, é de extrema importância referir que a ocupação de uma unidade hoteleira é muito díspar, consoante o segmento a partir do qual provêm as reservas. Assim, as RN totais de uma unidade dependem do número de quartos reservados procedentes de cada segmento, o que se traduz em $RN_{totais} = RN_{business} +$

$$RN_{complimentary} + RN_{contracted} + RN_{corporate} + RN_{direct}.$$

Todavia, em hotelaria o segmento de grupos não é homogéneo e pode conter grupos de vários segmentos de mercado (como grupos religiosos, associações nacionais, eventos sociais locais, reuniões empresariais, excursões, etc.) de dimensões variadas. Os grupos normalmente reservam um bloco de quartos, em que a dimensão do grupo pode ser distinta. Devido a esta imprevisibilidade que os segmentos de grupos representam para a ocupação de um hotel, quer pela dimensão do grupo, quer pela duração da estadia, não é totalmente espectável que volte a ocorrer o mesmo volume de ocupação. Os grupos representam por isto uma exceção na previsão. Desta forma, a previsão da presença de grupos depende, em grande parte, dos perfis de reserva de grupo, das vendas, do tamanho do grupo, da duração da estadia, do tipo de unidade (negócio ou lazer), e não só de dados históricos, que para este segmento são pouco representativos da realidade (Kimes, 1999).

Dadas estas circunstâncias, os dados utilizados foram filtrados de forma a não incluir informação de RN e OCC proveniente de reservas de grupos e, portanto, apenas foram incluídos os segmentos *contracted*, *corporate* e *direct*.

Outro aspeto a ter em consideração é o facto de os dados terem frequência mensal. Dado que para um setor tão volátil quanto o da hotelaria se torna crucial conseguir uma previsão diária, de forma a sustentar as oscilações de preço do quarto, foi necessário obter resultados em termos diários. Assim, obteve-se um peso percentual que representa o valor de cada RN diária no mês a que corresponde, de forma a ser possível atribuir esse mesmo peso aos dias a prever.

Para que mais facilmente se perceba a extrapolação imposta, considere-se o seguinte exemplo genérico: pretende-se obter as previsões para o dia 10 de março de 2018, onde se assume que os pesos são sempre calculados tendo por base o ano anterior, neste caso 2017. Assim, se no dia 10 de março de 2017 para uma determinada unidade houve 9 RN e no total do mês de março de 2017 se registaram 265 RN, então assume-se que o décimo dia de março tem um peso de 3,4% ($\frac{9}{265}$) do total das RN do mês de março de 2017. Posto isto, as previsões do dia 10 de março de 2018 representam 3,4% das RN previstas para todo o mês de março de 2018.

A esta necessidade é fundamental acrescentar outra exigência que remete para a calendarização de dias festivos. Note-se que os fins de semana, feriados, férias escolares, épocas festivas e a época balnear são fatores fundamentais para a OCC de uma unidade,

pois esta tende a aumentar consideravelmente nestas situações. Assim, ao fazer a extrapolação citada anteriormente é preciso considerar que se o dia 10 de março de 2017 aconteceu a uma sexta-feira, por conseguinte, o dia 10 de março de 2018 ocorrerá num sábado, o que fará com que seja necessário calcular o peso percentual do dia 10 de março de 2018 com base no dia 11 de março de 2017. A este paralelismo dá-se o nome de comparação com o mesmo dia da semana do ano anterior.

Por outro lado, esta regra contém algumas exceções, como é o caso dos feriados e das épocas festivas que ocorrem sempre na mesma data (exceto Páscoa). Para estes casos, o peso percentual foi calculado com base no mesmo dia numérico do ano passado, pois apenas assim se consegue perceber o impacto que determinado feriado teve para a ocupação da unidade.

O último aspeto a ter em atenção assenta na existência de unidades (especialmente no Algarve) que encerram o serviço em meses de época baixa. Para unidades que se sabe à partida que todos os anos fecham em determinados meses, é preciso preparar computacionalmente os modelos para este facto, quer prevenindo a obtenção de valores previstos diferentes de zero, quer excluindo estas observações de cálculos que interfiram diretamente em métricas criadas para a obtenção de previsões (como é o caso de médias móveis, índices sazonais, etc.). Por outro lado, se uma unidade apresentar um zero e não se tratar de uma unidade sazonal, então assume-se que a mesma esteve sob um processo de obras.

Em suma, os aspetos que foram considerados para a modelação e tratamento dos dados podem sumarizar-se nos seguintes:

- Exclusão de unidades com dois ou menos anos de histórico;
- Exclusão do segmento de grupos;
- Necessidade de prever em OCC e posteriormente converter em RN;
- Extrapolação de valores de previsão mensais para diários, tratando de forma diferenciada os feriados e épocas festivas;

- Unidades que encerram o serviço em meses de inverno ou que estiveram em processos de obras foram tratadas com especial atenção.

4.2. Análise das principais características amostrais

4.2.1. Globalidade das unidades abrangidas²

De acordo com as características amostrais das diversas unidades abrangidas no projeto, percebe-se que as unidades F, J e L, são as que apresentam médias mais elevadas (>80%) em termos de ocupação. Ao invés destas, as unidades P, Q e AB, são as que apresentam menores valores em média de OCC, sendo esta entre 38% e 32%.

Por outro lado, analisando os valores máximos de OCC, observa-se que as unidades AG, F e T são as unidades que registam os maiores valores máximos de ocupação, oscilando entre os 100% e os 98%.

No que diz respeito ao desvio padrão, as unidades C, R, Q são as que apresentam menores valores. Isto permite concluir que estas apresentam poucas oscilações em termos de OCC e, portanto, pouca dispersão na amostra, com valores de ocupação próximos da média.

De acordo com os cronogramas de cada série (Anexo I) torna-se evidente que existem unidades que apresentam uma componente sazonal mais vincada que outras, o no setor hoteleiro é um comportamento padrão. Verifica-se que as unidades que apresentam altos níveis sazonais (atingindo picos elevados nos meses de julho a setembro) são sobretudo as unidades *resort* dedicadas à vertente de lazer, localizadas essencialmente nas zonas do Algarve e da Madeira. Note-se que algumas das unidades nestas zonas encerram o serviço em épocas de baixa procura (por exemplo unidade B), o que se traduz numa sazonalidade mais acentuada face a hotéis que mantêm as atividades sucessivamente.

Observando o comportamento das séries referentes às unidades da Europa (F, L e J) verifica-se que estas apresentam uma sazonalidade menos evidente, com ciclos pouco delineados, característicos de unidades da vertente de negócio.

Atente-se também nas unidades do Brasil (R, G, AO, AE) que apresentam maiores

² Nesta secção, as unidades são referenciadas por letras de forma a manter a confidencialidade dos dados

valores de ocupação nos meses de janeiro e fevereiro, que correspondem à estação de verão local. Contudo, à exceção de G, um hotel de carácter de lazer, os restantes hotéis do Brasil não apresentam ciclos bem definidos. Observam-se, portanto, séries bastante recortadas, onde parece não existir uma tendência crescente, com valores de OCC próximos da média, comportamento típico de unidades que não se encontram em períodos de expansão.

No que diz respeito às unidades hoteleiras em África, destacam-se à primeira vista o comportamento de AB e de AJ que parecem aparentar uma quebra de estrutura. A primeira trata-se de uma série com valores iniciais de OCC relativamente baixos, para a qual a partir do ano de 2013 se observa uma mudança estrutural, passando a desenvolver um comportamento cíclico mais acentuado. No entanto, também AJ sofre uma mudança de comportamento, passando a apresentar ciclos menos longos, com um pico isolado na mesma altura de cada ano. Note-se que estas mudanças bruscas no comportamento das séries resultam, em muitos casos, de: mudanças na gestão, promoções, remodelação da unidade/edifício, entre outros fatores.

Por fim, analisando o comportamento das pousadas (unidade AK; a as unidades AR a BO) verifica-se que estas em geral apresentam um comportamento sazonal, com picos de ocupação nos meses de verão e quedas nos meses de inverno. Porém, na grande maioria, as pousadas apresentam valores mínimos de ocupação, para meses fora da época balnear, mais elevados que unidades das outras marcas. Este comportamento é explicado pelo facto de as pousadas serem vistas como um destino de fim de semana, mini férias e escapadas, o que faz com que, apesar dos valores máximos serem registados no verão, existam outros fluxos de hóspedes durante todo o ano.

Notam-se, portanto, dois tipos de comportamentos de séries referentes a pousadas. Um tipo apresentam ciclos longos com valores de OCC elevados em dois e três meses consecutivos, como é o caso das Pousadas BN, AX, AR e AK, etc. Por outro lado, as outras apresentam um pico extremo isolado, caindo depois abruptamente para valores mais reduzidos, como é o caso das pousadas AU, BD, BH e BC.

4.2.2. Unidade “caso de estudo”

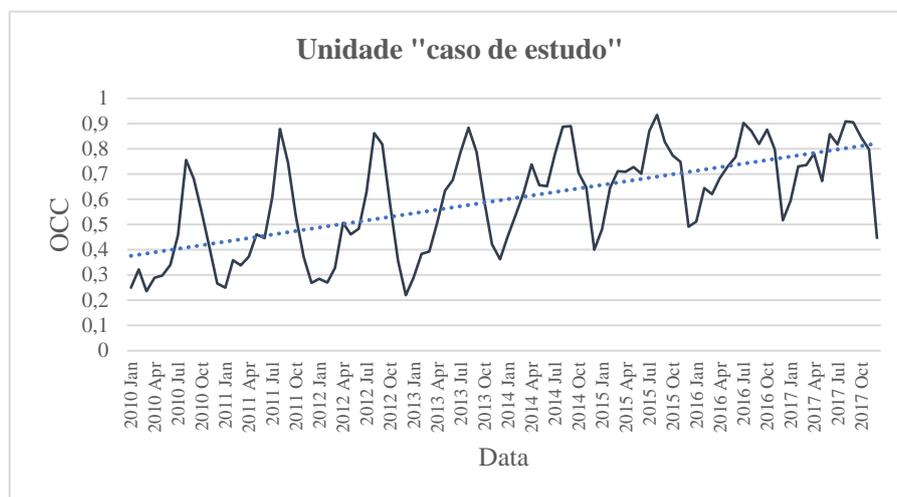


Figura 1- Cronograma da série de “caso de estudo” e respetiva linha de tendência (azul picotado)

A unidade escolhida como “caso de estudo” é uma unidade situada na Madeira, apresentado valores consideravelmente elevados de OCC durante todo o ano. De acordo com a análise do cronograma da série, verifica-se a existência de uma ligeira tendência crescente em termos de ocupação (Figura 1). Por outro lado, é possível perceber que o máximo de ocupação é atingido maioritariamente no mês de agosto, com exceção dos anos de 2014 e 2016, em que o ponto máximo ocorre no mês de setembro e de outubro, respetivamente.

Pode considerar-se ainda que, esta unidade hoteleira parece estar num período de expansão, no que diz respeito à procura. Tal conclusão é possível de ser verificada quando se analisam os valores mínimos de ocupação, que se mostram crescentes. Assim, por exemplo, nos anos de 2010 e 2011 a ocupação mínima para os meses de inverno rondava os 24%, o que contrasta com os anos de 2015, 2016 e 2017, em que a ocupação mínima se aproxima de 50%.

Comparando o período de 2010 a 2013 com o de 2014 a 2017 percebe-se que a estrutura da série apresenta ligeiras discrepâncias. No primeiro, é notório um comportamento mais extremo entre meses de verão e de inverno, o que faz com que a amplitude seja maior, traduzindo-se em picos de ocupação extremamente acentuados. Por outro lado, no segundo período mencionado, com o atenuar da variabilidade, percebe-se

claramente o impacto que as viagens *city break* e “fora de estação” têm na ocupação.

Desta forma, verifica-se, que os picos de OCC de verão e de inverno se encontram mais diluídos (ciclos mais longos) no decorrer do ano. Este fenómeno é resultado de fatores, que têm vindo a contribuir para a mudança de comportamento dos clientes. Assim, nota-se que os viajantes começam a abandonar a atitude característica de realizar apenas uma viagem extensa na época de verão, realizando deslocações mais curtas fora de época alta. Esta mudança de comportamento é fruto de diversas modificações, das quais se destacam a diminuição das tarifas de deslocação, o crescente número de ligações áreas, expansão de companhias áreas *low cost*, etc.

5. Resultados empíricos

Para a estimação dos modelos das secções 5.1 e 5.2, foi considerada uma “amostra de treino” com dados referentes ao período de janeiro de 2010 a dezembro de 2016. Para efeitos de avaliação da capacidade preditiva, foi considerada a “amostra de teste” com dados de OCC compreendidos entre janeiro e dezembro de 2017. Após avaliada a capacidade preditiva de cada modelo, realizaram-se previsões mensais de OCC, para o ano 2018 (previsão a 12 passos à frente).

De seguida, apresentam-se os modelos de decomposição clássica para todas as unidades abrangidas no estudo e os modelos SARIMA de cariz univariado aplicados para uma unidade hoteleira do *Pestana Hotel Group*.

5.1. Modelação e previsão através de modelos clássicos de séries temporais

Esta secção descreve os modelos aplicados para todas as unidades abrangidas pelo estudo. O processo foi totalmente automatizado, computando todas as restrições mencionadas na secção 4.1. Este mecanismo permitiu obter um relatório *Excel* automático com as previsões de OCC para o ano de 2018, de acordo com os dois modelos definidos através da decomposição clássica (o *output* deste relatório encontra-se no anexo II).

Modelação:

Modelo 1: de acordo com a equação (2) este modelo consiste na obtenção da componente tendência através do ajuste de uma reta de regressão, multiplicada pelos índices sazonais apropriados (anexo III).

Modelo 2: este modelo, também descrito pela equação (2), considera que a tendência é captada através de médias móveis, multiplicadas pelos índices sazonais mensais adequados (anexo III).

Tabela 1- Índices sazonais obtidos pelas duas metodologias

Mês	Unidade “caso de estudo”		Unidade Cidadina- Exemplo	
	Índices sazonais obtidos para o modelo 1	Índices sazonais obtidos para o modelo 2	Índices sazonais obtidos para o modelo 1	Índices sazonais obtidos para o modelo 2
Jan	0,68	0,66	0,82	0,80
Fev	0,84	0,82	0,91	0,89
Mar	0,88	0,86	1,03	1,01
Abr	1,01	1,00	1,06	1,03
Mai	1,02	1,01	0,96	0,94
Jun	1,07	1,06	0,97	0,94
Jul	1,27	1,26	1,09	1,06
Ago	1,48	1,49	1,17	1,16
Set	1,36	1,39	1,02	1,02
Out	1,13	1,17	1,07	1,09
Nov	0,80	0,99	0,87	1,03
Dez	0,54	0,66	0,85	1,08

Na tabela 1 encontram-se os índices sazonais referentes à unidade “caso de estudo” e a uma unidade de cariz menos sazonal, obtidos para os dois modelos. Verifica-se que, pelas duas metodologias (médias móveis e regressão) os valores dos índices não apresentam acentuadas discrepâncias. Para a unidade “caso de estudo”, os índices sazonais estão abaixo da referência de 1 nos meses de janeiro a março e de novembro a dezembro, que correspondem a meses de ocupação diminuída. Por outro lado, há uma clara evidência que para os meses de julho, agosto e setembro existe uma acentuada ocupação, fruto de fluxos de época balnear.

No que diz respeito à outra unidade, pode concluir-se que esta apresenta valores relativamente elevados de ocupação durante todo o ano. Assim, constata-se que a amplitude entre os índices de cada mês é muito menor do que a registada para a unidade que serve de “caso de estudo”. Este facto, leva a crer que se trata de um hotel citadino, vocacionado para uma vertente de negócio apresentando ciclos sazonais pouco definidos, com valores de ocupação consideravelmente elevados durante todo o ano.

Previsão:

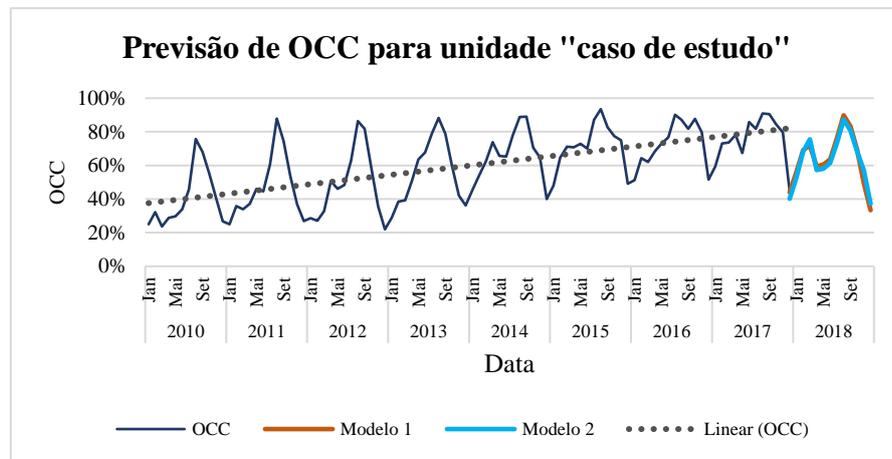


Figura 2- Previsão de OCC para a unidade “caso de estudo”

De acordo com os resultados de previsão obtidos pelo método de decomposição clássica (Figura 2) para a unidade “caso de estudo”. Observa-se que apesar de existir uma tendência crescente ao longo da série, os dois modelos não estendem essa tendência na previsão. Este facto pode ser explicado devido à mudança estrutural dos ciclos nos últimos dois anos. Repare-se que em termos de registos históricos, a série só contém dois ciclos (referentes aos anos de 2016 e 2017) mais díspares dos restantes. Dado este facto, os modelos preditivos consideraram fortemente o formato dos ciclos mais antigos, visto serem os mais perseverantes na série. Assim, observando-se os resultados da previsão verifica-se um comportamento mais semelhante com o decorrido nos anos de 2014 e 2015, do que nos últimos dois anos.

Por fim, conclui-se que através desta metodologia, em 2018 são esperados dois picos de ocupação mais elevada, um primeiro na altura dos meses de março e abril e um

segundo nos meses de julho e agosto. Nos meses de janeiro e fevereiro está prevista uma ocupação na ordem dos 40%, tal como para os meses finais do ano.

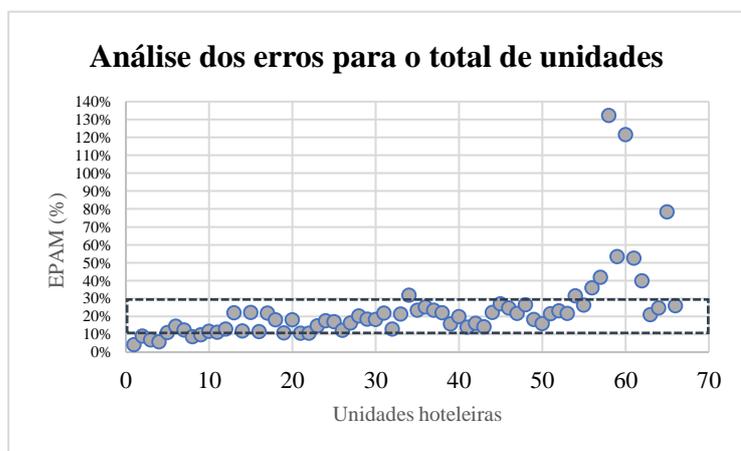


Figura 3- Análise dos erros por unidade

Para todas as unidades abrangidas no projeto foram obtidas previsões por esta metodologia e, para todas elas foi medido o EPAM (também foram consideradas as outras medidas de erro, contudo não se apresentam, por revelarem conclusões iguais às do EPAM), recorrendo à amostra de teste. De acordo com análise destes erros (Figura 3) verifica-se que, a maioria dos hotéis conduzem erros preditivos entre os 10% e os 30%.

Contudo, existem três unidades que apresentam erros de previsão extremamente elevados, sendo estas H, AJ e V. Tal facto justifica-se pela quebra de estrutura destas séries, o que dificulta as previsões.

No que diz respeito às unidades que apresentam valores de EPAM superiores a 25%, foi feita uma análise que é incorporada no *output* realizado para a empresa. De acordo com esta análise foi possível concluir que, na grande maioria, os erros de previsão elevados ocorrem pelo crescimento/diminuição do segmento de grupos, existência de eventos ligados a casamentos, presença do segmento *extra* ou ainda pela transferência do segmento pelo qual provêm as reservas.

Os resultados obtidos por esta metodologia, permitem concluir que 46 unidades produzem erros de previsão menores quando é utilizado o modelo 1, e as restantes 20 quando é aplicado o modelo 2.

Pode concluir-se ainda que, em termos de EPAM este é próximo dos 24% para os dois modelos, sendo que para a unidade que serve de “caso de estudo” se obteve um EPAM de 22%. Note-se por fim, que dada a estrutura simples desta metodologia, e a

forma universal como todas as séries são tratadas, espera-se obter erros de previsão mais reduzidos através da metodologia SARIMA, onde cada série é modelada individualmente.

5.2. Modelação através de modelos da classe SARIMA

O ponto de partida na metodologia *Box-Jenkins*, apresentada para a criação de modelos da classe SARIMA, consiste na verificação da estacionaridade da série, através do teste ADF. De acordo com o anexo IV, verifica-se que a série não é estacionária, uma vez que o *p-value* obtido no teste leva à não rejeição da hipótese nula de que a série tem pelo menos uma raiz unitária, isto é, a série não é estacionária.

Para além deste facto, ao analisar as funções autocorrelação (FAC) e autocorrelação parcial (FACP), verifica-se que na FAC existe um comportamento em forma de onda sinusoidal, típico de séries não estacionárias. Posto isto, é necessário proceder à primeira diferenciação da série, afim de alcançar a estacionaridade da mesma. Após esta diferenciação verifica-se, através do teste ADF que a série já é estacionária, (anexo IV), cumprindo assim o pressuposto. Desta forma, obtém-se uma série integrada de ordem 1, ficando definido o parâmetro *d* do modelo SARIMA.

Porém, de acordo com a análise inicial da série e atendendo aos correlogramas respetivos (anexo IV), percebe-se que a série apresenta um alto cariz sazonal. Assim, verifica-se que, na FAC e FACP da série diferenciada, o *lag* 12, apresenta um valor extremamente elevado de correlação, sobretudo na FAC. Para corroborar esta conjectura, confirma-se também que, tanto os primeiros *lags* (que indicam uma estrutura não sazonal da série), como os *lags* sazonais, são lentos a atingir o valor zero, o que torna necessário a aplicação de uma diferenciação sazonal ($\Delta OCC_t = OCC_t - OCC_{t-12}$). Após a aplicação desta diferenciação, está-se em condições de começar a primeira etapa para a modelação da série temporal.

Por análise gráfica da FAC e da FACP da série com uma diferenciação apropriada, nota-se que o *lag* 1 da FACP decai exponencialmente a zero, tal como o *lag* sazonal 12, o que sugere um modelo $SARIMA(1,1,0)(1,1,0)_{12}$. Por outro lado, também na FAC o *lag* 1 decresce rapidamente a zero, o que leva a considerar um termo MA (1), para além dos já considerados no modelo anterior, o que se traduz em: $SARIMA(1,1,1)(1,1,0)_{12}$.

Em alternativa aos dois modelos apresentados, surge ainda um terceiro onde se inclui um termo AR (2), considerando-se que o decaimento exponencial a zero na FACP

é muito mais acelerado a partir do *lag* 2, o que se expressa no seguinte modelo: $SARIMA(2,1,1)(1,1,0)_{12}$.

Por fim, o último modelo incluído no estudo é $SARIMA(0,1,1)(1,1,0)_{12}$, que considera um termo MA (1), fruto do corte acentuado a partir do *lag* 1 na FAC, e um termo sazonal- SAR (12) bastante evidente na FACP.

Tabela 2- Modelos da classe SARIMA selecionados

	Equação
Modelo 1	$SARIMA(1,1,0)(1,1,0)_{12}$
Modelo 2	$SARIMA(1,1,1)(1,1,0)_{12}$
Modelo 3	$SARIMA(2,1,1)(1,1,0)_{12}$
Modelo 4	$SARIMA(0,1,1)(1,1,0)_{12}$

Concretizada a primeira etapa, que conduziu à identificação de quatro modelos passíveis de descrever a série temporal (Tabela 2), está-se em condições de avançar para a etapa de estimação. Nesta fase, recorre-se ao auxílio do *software Eviews*, que apura uma estimativa dos parâmetros do modelo, através do método de máxima verosimilhança. As estimativas dos parâmetros para cada modelo encontram-se no anexo V.

De acordo com os *outputs* obtidos pelo *software*, em todos os modelos existe pelo menos um parâmetro que não se mostra estatisticamente significativo em termos individuais. Tal facto verifica-se quando se observa os *p-values* do teste estatístico *t* de significância individual dos parâmetros. Porém prossegue-se para a etapa de diagnóstico, com parâmetros que não se revelam estatisticamente significativos individualmente, uma vez que de acordo com o teste estatístico F de significância global, se comprova que os parâmetros em conjunto se revelam estatisticamente significativos. Assim, no que diz respeito ao teste de avaliação conjunta, para os quatro modelos em análise, foram obtidos *p-values* próximos de zero para o teste F global, pelo que, há evidência estatística de que os parâmetros são estatisticamente significativos em conjunto.

Na fase de avaliação de diagnóstico, analisando os resultados do teste de *Portmanteu*, verifica-se que para os quatro modelos em estudo, todos os *lags* apresentam *p-values* superiores ao nível de significância de 5% (anexo VI). De acordo com estes resultados, é possível afirmar com uma confiança de 95%, que os erros dos modelos não são correlacionados e, portanto, evidenciam um comportamento semelhante ao de um

ruído branco.

Uma vez cumpridos os pressupostos impostos na fase de diagnóstico, prossegue-se para a fase de seleção dos modelos. Nesta fase final, recorre-se aos critérios de seleção AIC e BIC para comparar os diferentes modelos.

Tabela 3 - Valores obtidos para os critérios de seleção

Modelos	AIC	BIC
Modelo 1	-2.588937	-2.462456
Modelo 2	-2.624289	-2.466187
Modelo 3	-2.627634	-2.501153
Modelo 4	-2.597088	-2.407365

De acordo com a Tabela 3 verifica-se que o modelo 3 é o que apresenta um menor valor para os critérios, pelo que em termos de ajustamento, este será o modelo preferível. Contudo, mantendo o objetivo transversal a todo este projeto, foram obtidas as previsões para a amostra de treino (janeiro a dezembro de 2017) com os quatro modelos, e posteriormente avaliada a capacidade preditiva de cada um.

Tabela 4- Erros de previsão com base na amostra de treino

	REQM	EQM	EAM	EPAM
Modelo 1	6,05%	0,59%	6,05%	8,88%
Modelo 2	5,91%	0,60%	5,91%	8,69%
Modelo 3	5,89%	0,59%	5,89%	8,65%
Modelo 4	5,91%	0,60%	5,91%	8,70%

De acordo com a tabela 4 pode concluir-se que, o modelo 3 para além de ser o preferível em termos de ajustamento, é também o que apresenta os menores valores para todas as tipologias de erros. Posto isto, está-se em condições de prosseguir com este modelo para a obtenção de previsões da variável OCC para os doze meses do ano de 2018 (Figura 4).

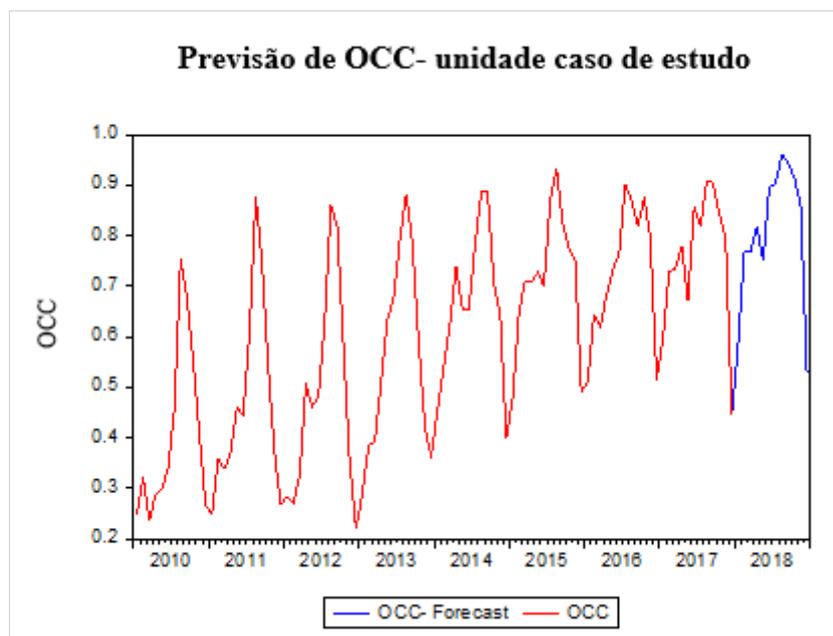


Figura 4- Previsão de OCC para a unidade “caso de estudo” pelo modelo SARIMA

Tal como já foi referido anteriormente, esta unidade hoteleira até 2017, parecia caminhar num período expansão, conjectura que foi corroborada pelos resultados preditivos. De acordo com as previsões obtidas pelo modelo da classe SARIMA, no ano de 2018 a tendência crescente da série prolongar-se-á (Figura 4), aumentando assim os seus níveis de OCC para os meses de época alta. Assim, espera-se um ligeiro aumento de ocupação no mês de agosto face aos registos anteriores para o mesmo mês.

Por outro lado, espera-se que no início do ano de 2018 ocorra uma ligeira queda na ocupação, o que se traduz num aumento dos valores mínimos registados para o período de época baixa. Desta forma, estima-se que nos primeiros meses de 2018, a OCC passará a rondar os 40% ao invés de valores na ordem dos 50%, registados nos últimos dois anos.

Os resultados de previsão apontam para um comportamento, em geral, semelhante ao ano de 2017. Contudo, analisando detalhadamente as previsões, espera-se que nos meses de junho e julho de 2018 não ocorra a mesma queda de OCC que se registou em 2017. Por fim, pode concluir-se que os ciclos sazonais desta unidade tendem a tornar-se cada vez mais uniformes e longos. Para isto, basta reparar que o modelo prevê que o comportamento do ano 2018, seja mais igualável ao dos últimos dois anos, deixando de existir ciclos curtos, com picos bruscos isolados, tal como se registou nos primeiros anos.

6. Conclusão e investigações futuras

Na ótica dos modelos de decomposição clássica de séries temporais, acredita-se que o comportamento das forças constituintes seja próximo do que irá ocorrer futuramente, o que serviu de conjectura na obtenção de previsões para o ano de 2018. Entre os dois modelos desta classe, o modelo 1 (que recorre à regressão) mostra-se preferível na obtenção de resultados de previsão para 48 hotéis, e o modelo 2 para 16 hotéis.

De acordo com a modelação SARIMA, foram analisados quatro modelos para descrever a ocupação de uma unidade “caso de estudo”. O terceiro modelo foi o que apresentou um EPAM inferior aos restantes (8%) e, portanto, foi através deste que se obtiveram os valores de previsão referentes à ocupação para 2018.

Comparando os erros de previsão obtidos pelas duas metodologias para a mesma unidade hoteleira, conclui-se que os modelos da classe SARIMA são preferíveis, apresentando um EPAM de 8% face aos 22% obtidos pelo método de decomposição.

Os resultados de previsão obtidos pelas duas metodologias, para a unidade que serviu de “caso de estudo” ao longo de todo o projeto, revelam evidentes discrepâncias. De acordo com a metodologia de decomposição clássica, a tendência crescente da série não é acompanhada para valores futuros. Assim, em 2018 esperam-se dois picos isolados de ocupação elevada, um nos meses de março e abril e outro nos meses de julho e agosto. Nos meses de janeiro, fevereiro, novembro e dezembro é prevista uma ocupação perto dos 40%.

Através da metodologia SARIMA, os resultados obtidos mostram-se bastante diferentes. Os valores de previsão para 2018, são mais semelhantes aos valores de ocupação registados em 2017. Contudo, é esperado que nos meses de junho e julho de 2018 não ocorra a mesma queda de OCC que se registou em 2017 no mesmo período. Segundo esta metodologia, os resultados de previsão acompanham a tendência crescente da série, refletindo-se em ciclos sazonais cada vez mais invariáveis e longos. Face a esta mudança do comportamento dos ciclos sazonais, deixa de ser esperado um mês isolado com elevado valor de ocupação, passando a esperar-se 3 a 4 meses de valores consecutivamente elevados.

As mudanças, que já se começam a notar nos ciclos sazonais, e que se preveem para tempos futuros, são explicadas pela mudança da sazonalidade do turismo. Esta

mudança de comportamento representa um desafio às políticas de desenvolvimento, nomeadamente no ramo de gestão e planeamento de destinos. Desta forma, já é visível nos dados que a concentração densa da população turística apenas num período do ano começa a ser amenizada, através de ofertas turísticas diversificadas. Este facto, levará ao afluxo de pessoas durante todo o ano, o que dará espaço para viagens *city break* e *travel off season*, que ao não serem registadas de forma tão frequente em dados precedentes, poderão dificultar as previsões

Com a elaboração deste projeto e conseqüente criação de um *output* automático, que inclui as previsões para as diferentes unidades, as partes interessadas do grupo Pestana podem escolher qualquer uma das unidades abrangidas e obter em tempo real as previsões de ocupação para o ano de 2018 e anos futuros. Este facto auxiliará os departamentos responsáveis pela gestão da receita e de estratégia de *pricing*. Por outro lado, os valores de previsão obtidos tornam-se para o Grupo Pestana uma ferramenta de apoio no processo de criação do *budget* para cada unidade.

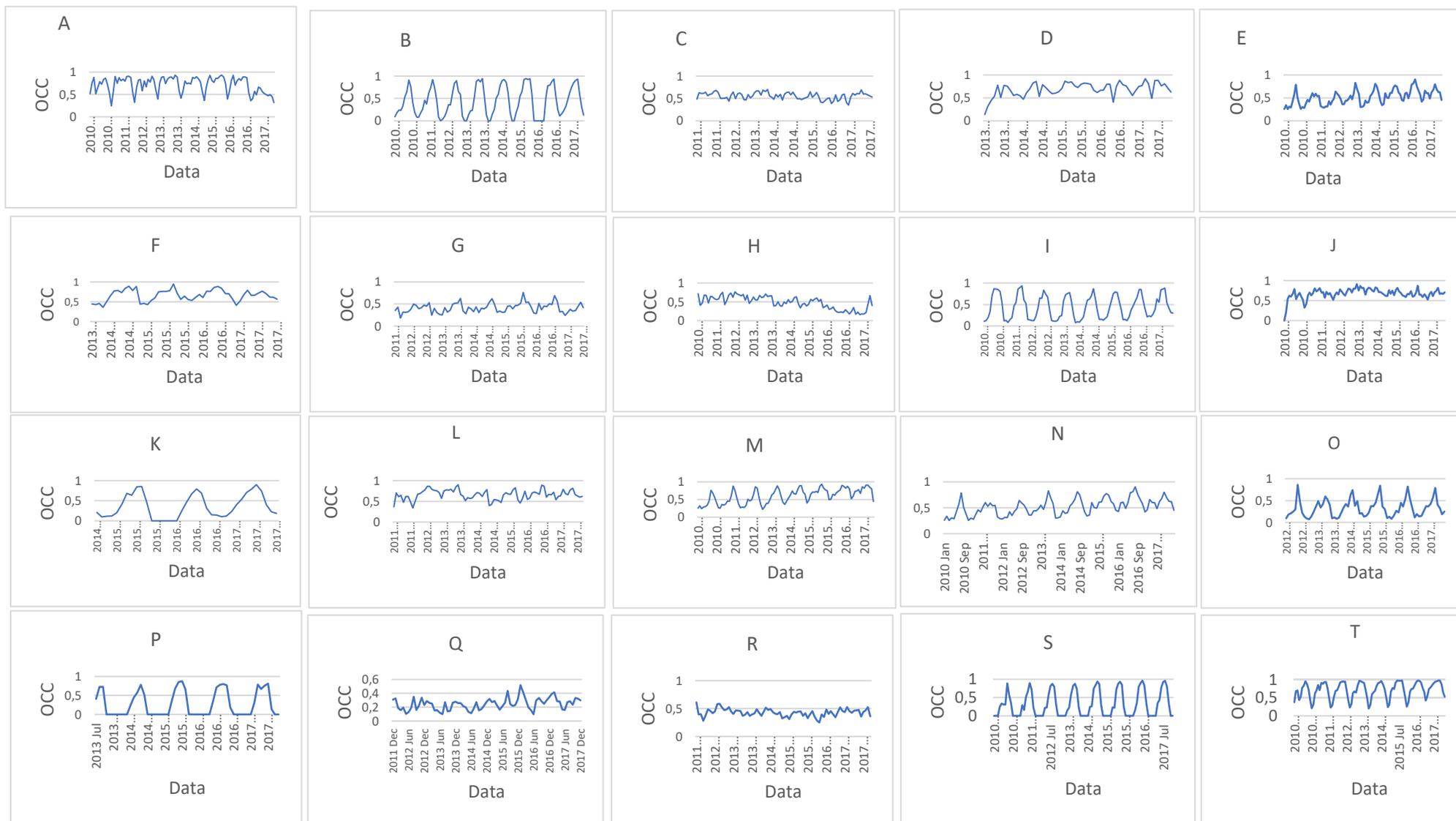
Para investigações futuras, sugere-se o uso de modelos de regressão baseados em variáveis explicativas como: temperaturas climatéricas dos destinos, procura das cidades, dados de voos e de movimentos em aeroportos, índice de preços da concorrência e datas de eventos festivos. Ao abarcar estas novas variáveis, será possível obter previsões que potencialmente apresentam menor erro e, simultaneamente, perceber de que forma estas variáveis influenciam a ocupação das unidades hoteleiras.

7. Bibliografia

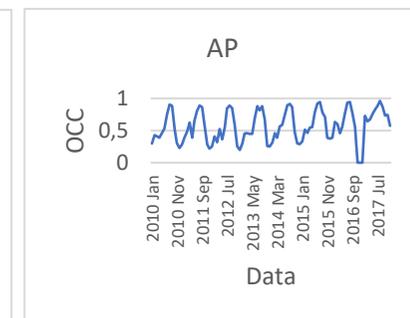
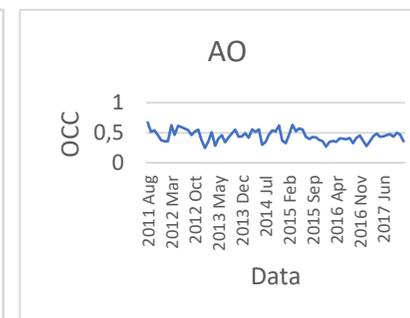
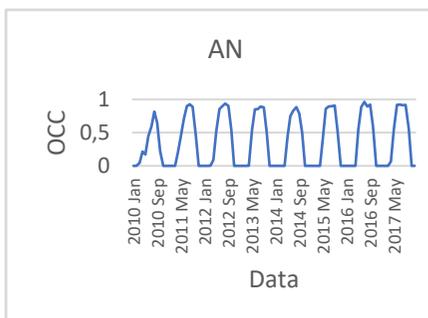
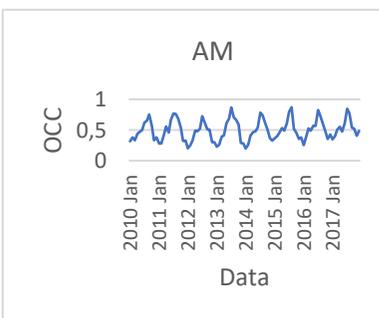
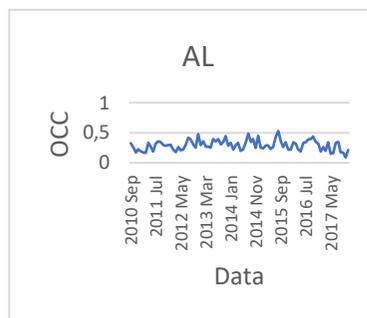
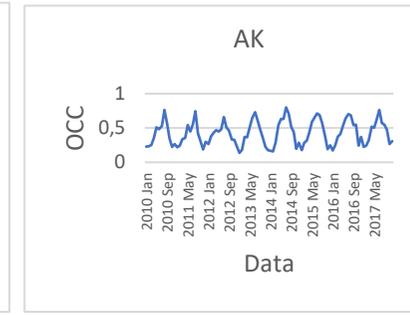
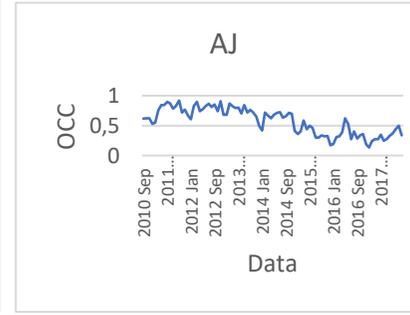
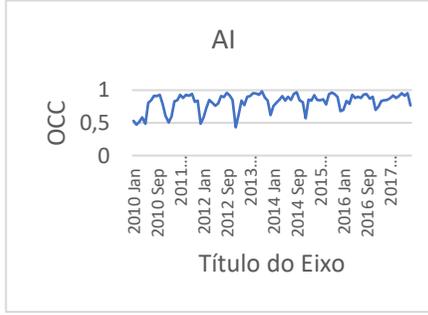
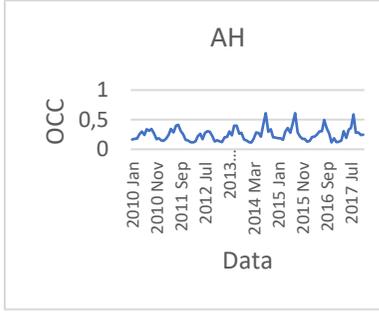
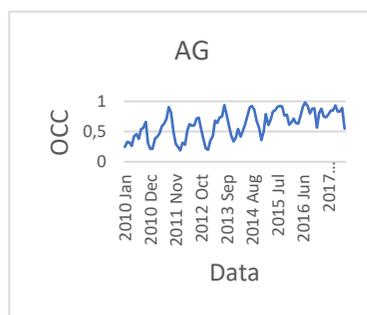
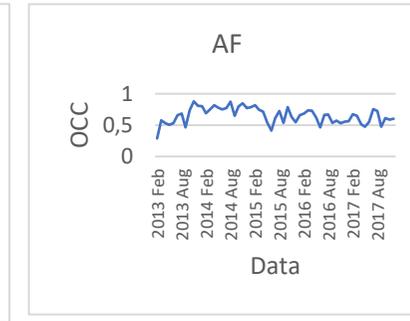
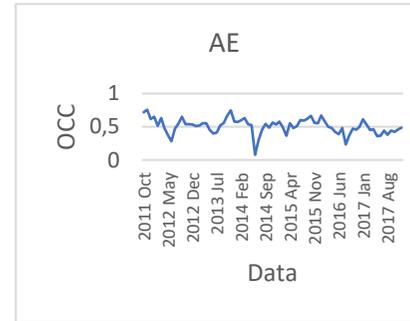
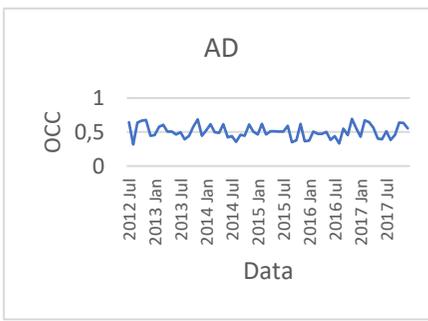
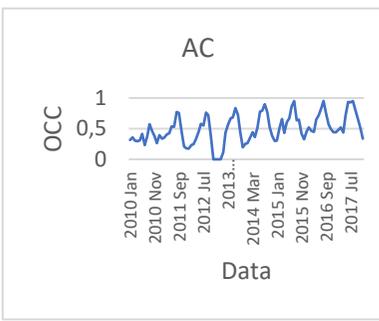
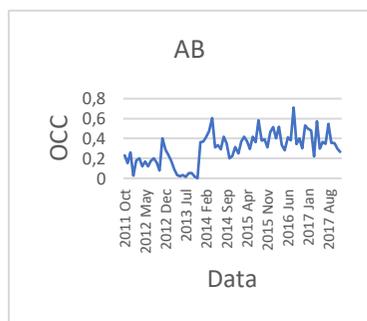
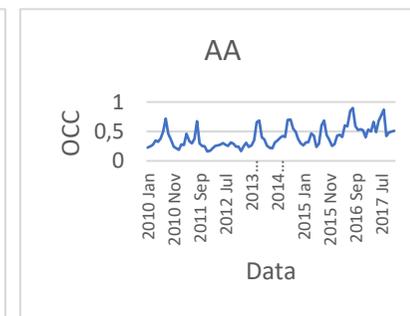
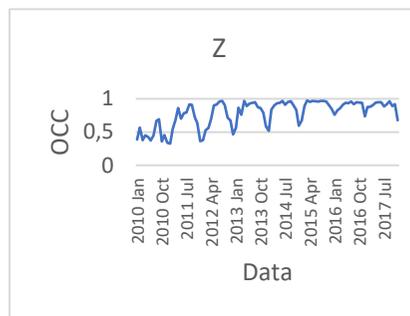
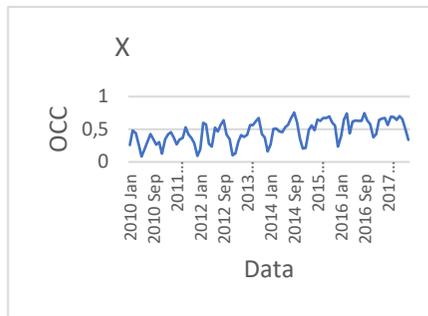
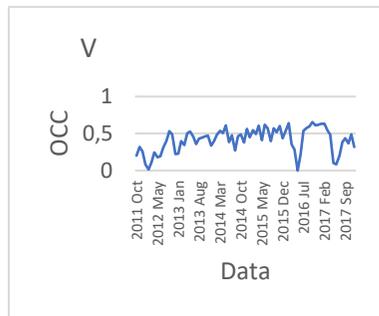
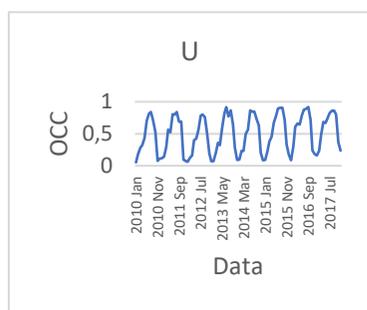
- Bayer, D. M., Castro, N. M., & Bayer, F. M. (2012). Modelagem e Previsão de Vazões Médias Mensais do Rio Potiribu Utilizando Modelos de Séries Temporais. *Instituto de Pesquisas Hidráulicas — UFRGS*, vol. 17., p. 11.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (1994). *Time series analysis: Forecasting and Control*, 3ª Ed., San Francisco: Prentice-Hall.
- Caiado, J. (2012). *Métodos de Previsão em Gestão com aplicações em Excel*, 1ª Ed., Lisboa: Edições Sílabo.
- Kimes, S. E. (1999). Group Forecasting Accuracy in Hotels. *Journal of the Operational Research Society*, Cornell University School of Hotel Administration, 50(11), p. 1104-1110.
- Lee, A. O. (1990). Airline Reservations Forecasting: Probabilistic and Statistical Models of the Booking Process. *Flight Transportation Laboratory Report R90-5*. Massachusetts Institute of Technology.
- Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Winkler, R., Newton, J., Parzen, E. (1982). The accuracy of extrapolation (time series) methods: results of a forecasting competition. *Journal of Forecasting*, vol. 1., p. 111-153.
- McCartney, S. (2000). In Airlines Find a Bag of High-Tech Tricks to Keep Income Aloft. *Wall Street Journal*, January, issue 20
- Murteira, B. J., & Black, G. H. (1993). *Estatística descritiva*. Portugal: McGraw-Hill.
- Murteira, B. J., Muller, D. A., & Turkman, K. F. (2000). *Análise de Sucessões Cronológicas*. Portugal: McGraw-Hill.
- Neves, H., Cruz, A. R., & Correia, A. (2014). A sazonalidade da procura turística na ilha de Porto Santo. *Revista Portuguesa de estudos regionais*, vol. 17., p. 25-43.
- Pestana.com*. (25 de Janeiro de 2018). Obtido de Pestana Hotel Group: <https://www.pestana.com/pt>
- Relvas, R. B. (2017). Turismo responde por 7% da economia nacional. *ECO- Economia online*, 1.
- Sa, J. (1987). Reservations forecasting in airline yield management. *Flight Transportation Lab Report R87-1*. Massachusetts Institute of Technology, p. 116-117.
- Weatherford, L. R., & Kimes, S. E. (2003). A Comparison of Forecasting Methods for Hotel. *International Journal of Forecasting*, 19(3), p.401-415.

8. Anexos

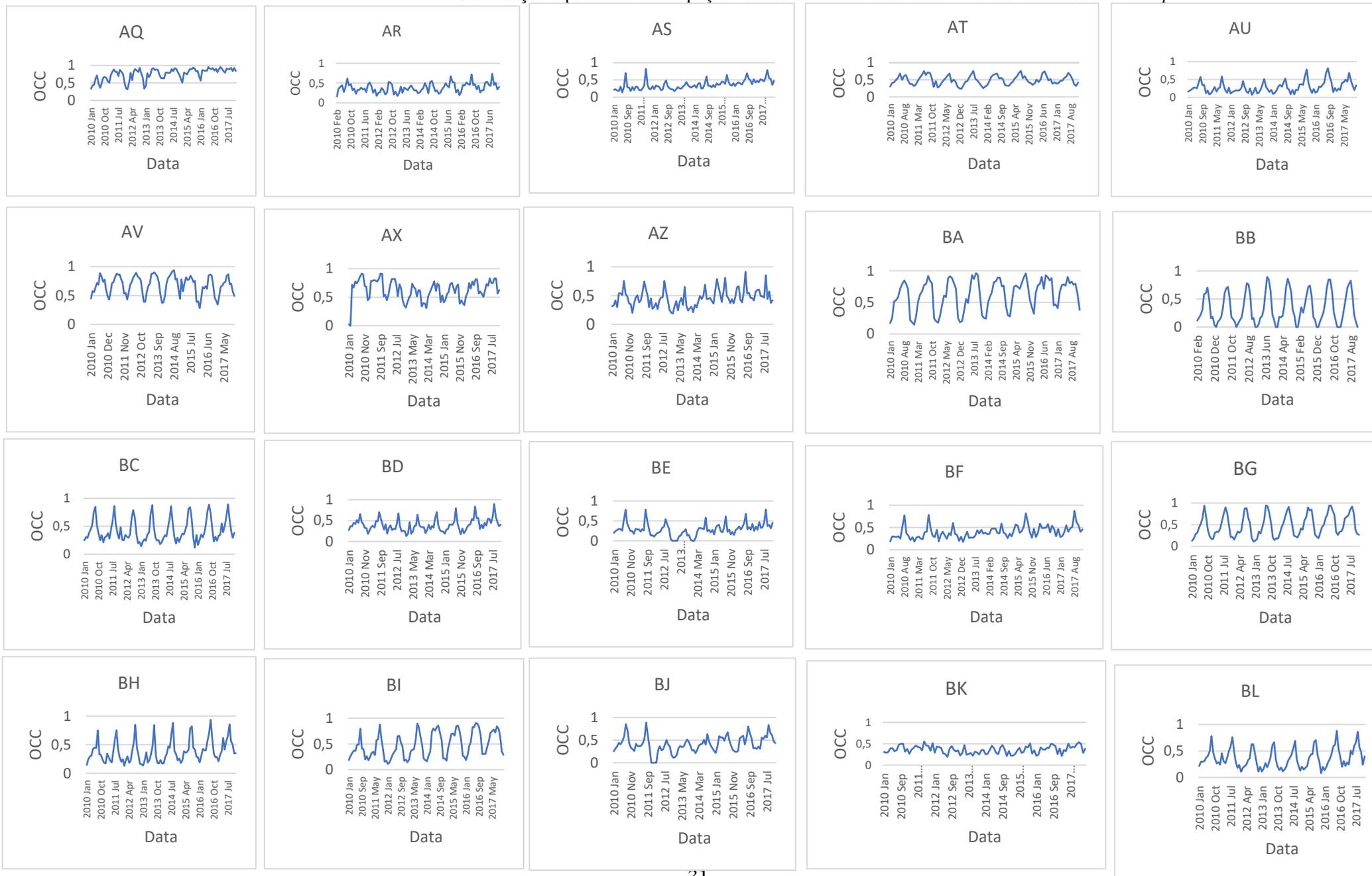
Anexo I- Cronogramas da totalidade de unidades hoteleiras



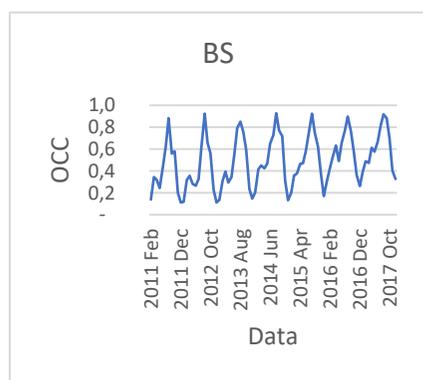
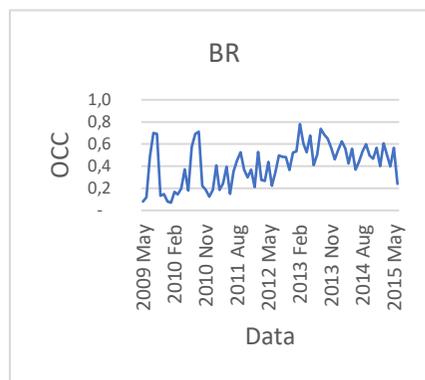
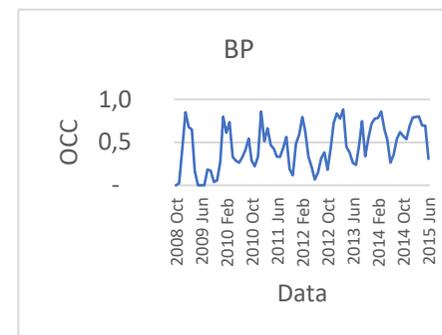
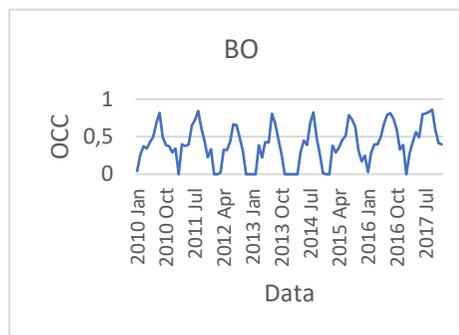
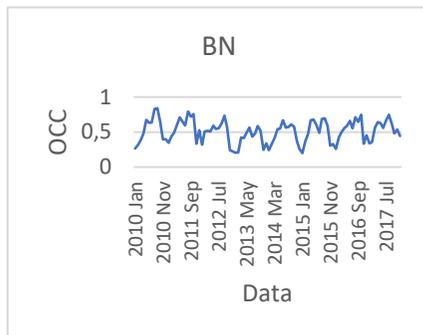
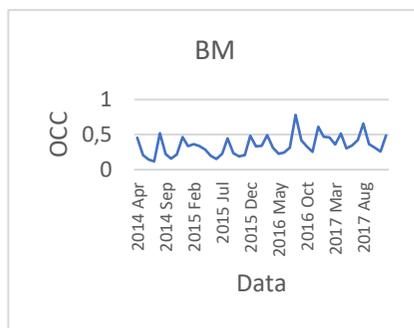
Modelação e previsão da ocupação em unidades hoteleiras. O caso do *Pestana Hotel Group*



Modelação e previsão da ocupação em unidades hoteleiras. O caso do *Pestana Hotel Group*



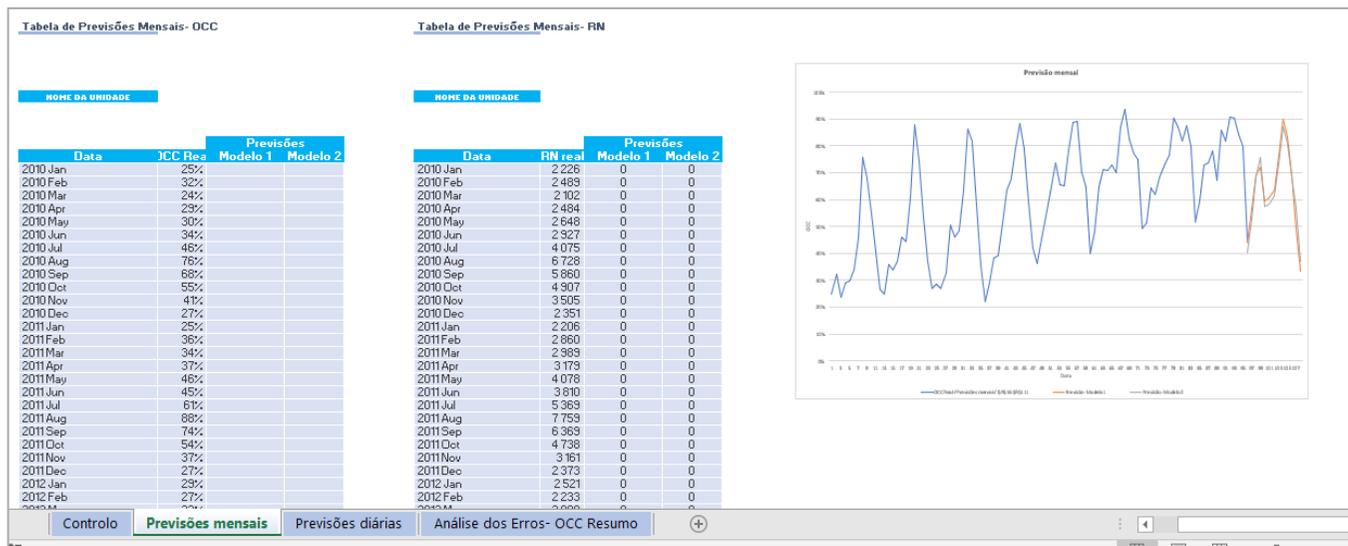
Modelação e previsão da ocupação em unidades hoteleiras. O caso do *Pestana Hotel Group*



Anexo II- *Output* do relatório para a empresa, com previsões pelo método de decomposição clássica- *Excel*

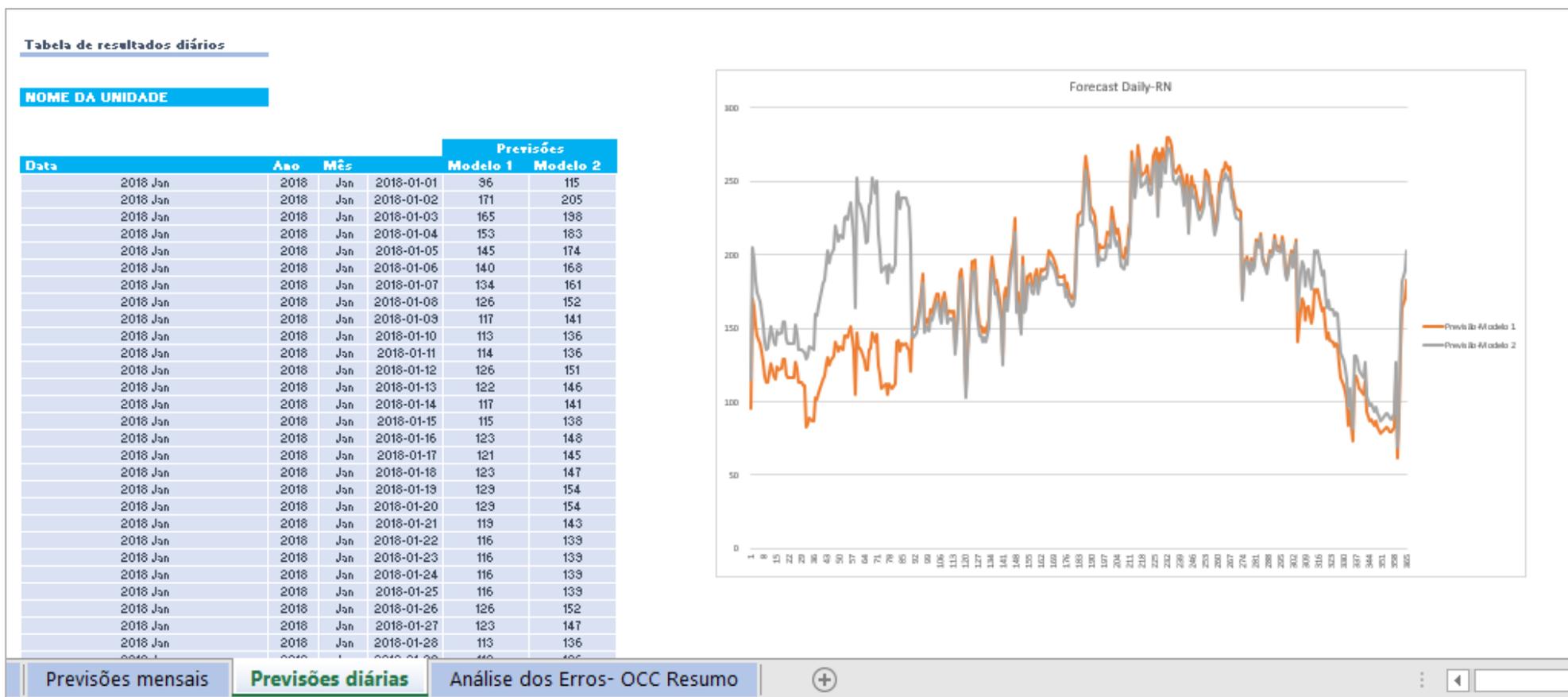


Sheet 1: “Controlo” - Possibilidade de escolher a unidade pretendida



Sheet 2: “Previsões mensais” - valores de previsão para a unidade escolhida em OCC e respetiva conversão em RN

Modelação e previsão da ocupação em unidades hoteleiras. O caso do *Pestana Hotel Group*



Sheet 3: “Previsões diárias” - valores de previsão para a unidade escolhida em RN

Modelação e previsão da ocupação em unidades hoteleiras. O caso do *Pestana Hotel Group*

Unidade	Erro	Modelo	Comentários
	4%	Modelo 1	
	9%	Modelo 1	
	7%	Modelo 1	
	6%	Modelo 1	
	11%	Modelo 1	
	15%	Modelo 1	
	12%	Modelo 1	
	9%	Modelo 2	
	10%	Modelo 1	
	12%	Modelo 1	
	11%	Modelo 1	
	13%	Modelo 1	
	22%	Modelo 1	
	12%	Modelo 1	
	22%	Modelo 1	
	12%	Modelo 1	
	22%	Modelo 1	
	18%	Modelo 1	
	11%	Modelo 1	
	18%	Modelo 1	
	11%	Modelo 2	
	11%	Modelo 2	
	15%	Modelo 1	
	18%	Modelo 1	
	17%	Modelo 1	
	12%	Modelo 1	
	16%	Modelo 2	
	20%	Modelo 1	
	18%	Modelo 2	
	18%	Modelo 2	
	22%	Modelo 1	
	13%	Modelo 2	
	21%	Modelo 1	

Sheet 4: “Análise dos Erros- OCC Resumo” – Parte I: Para cada unidade são apresentados os erros mínimos e o modelo respetivo

Anexo III - Detalhes dos modelos de decomposição clássica- *Excel*

Modelo 1:

Year	Month	OCC	St	Trend 1	Modelo 1
2010	Jan	26%	0,53	0,51	=SE(X16="S";0;SE.ERRO(AX16*BB16;""))
	Fev	32%	0,73	0,51	0,37
	Mar	39%	0,95	0,51	0,48
	Abr	48%	1,11	0,51	0,56
	Mai	67%	1,20	0,51	0,61
	Jun	63%	1,21	0,51	0,61
	Jul	63%	1,08	0,51	0,55
	Ago	83%	1,29	0,51	0,65
	Set	84%	1,30	0,51	0,66
	Out	63%	1,19	0,51	0,60
	Nov	40%	0,66	0,51	0,33
	Dez	40%	0,61	0,51	0,31
2011	Jan	35%	0,53	0,51	0,27
	Fev	44%	0,73	0,51	0,37
	Mar	49%	0,95	0,51	0,48
	Abr	60%	1,11	0,51	0,56
	Mai	71%	1,20	0,51	0,61
	Jun	64%	1,21	0,51	0,61
	Jul	59%	1,08	0,51	0,55
	Ago	79%	1,29	0,51	0,65

Cálculo dos coeficientes da recta	
b=	0,0001
a=	0,5

a)

b)

Legenda: a) obtenção do modelo 1: multiplicação da tendência, obtida através de regressão linear, pelos índices sazonais; b) valores estimados para os coeficientes da reta

Modelo 2:

Year	Month	OCC	Mt=Tt*Ct	Rt=Yt/Mt
2010	Jan	26%		
	Fev	32%		
	Mar	39%		
	Abr	48%		
	Mai	67%		
	Jun	63%		
	Jul	63%	= $(\text{SOMA}(\text{AD16:AD28})+\text{SOMA}(\text{AD17:AD27}))/24$	
	Ago	83%	55%	150%
	Set	84%	56%	149%
	Out	63%	57%	111%
	Nov	40%	58%	69%
	Dez	40%	58%	68%
2011	Jan	35%	58%	60%
	Fev	44%	57%	76%
	Mar	49%	57%	86%
	Abr	60%	57%	105%
	Mai	71%	57%	124%
	Jun	64%	57%	112%

Year	Month	OCC	Mt=Tt*Ct	Rt=Yt/Mt
2010	Jan	26%		
	Fev	32%		
	Mar	39%		
	Abr	48%		
	Mai	67%		
	Jun	63%		
	Jul	63%	54%	= $\text{AD22}/\text{AG22}$
	Ago	83%	55%	150%
	Set	84%	56%	149%
	Out	63%	57%	111%
	Nov	40%	58%	69%
	Dez	40%	58%	68%
2011	Jan	35%	58%	60%
	Fev	44%	57%	76%
	Mar	49%	57%	86%
	Abr	60%	57%	105%
	Mai	71%	57%	124%
	Jun	64%	57%	112%

a)

b)

Legenda: a) Obtenção da tendência por alisamento da série, através do cálculo de médias móveis de ordem 12 b) cálculo do rácio entre a série original e alisada.

Mês	Factores sazonais
Jan	0,53
Fev	0,70
Mar	0,86
Abr	1,06
Mai	1,18
Jun	1,17
Jul	1,21
Ago	1,26
Set	1,35
Out	1,17
Nov	0,90
Dez	0,73

Índices sazonais
0,72
0,88
1,09
1,21
1,20
1,24
1,29
1,38
1,20
0,92
0,75

= $\text{CP14}/\text{MÉDIA.GEOMÉTRICA}(\text{SCPS14:SCPS25})$

Legenda: obtenção dos fatores sazonais e dos respetivos índices sazonais por média geométrica.

Data	Month	opç OC	OCC	Modelo 2	Mt= Tt*Ct	Rt=Yt/Mt	St MM
2010 Jan	Jan	25%	25%				0,66
2010 Feb	Fev	32%	32%				0,82
2010 Mar	Mar	24%	24%				0,86
2010 Apr	Abr	29%	29%				1,00
2010 May	Mai	30%	30%				1,01
2010 Jun	Jun	34%	34%				1,06
2010 Jul	Jul	46%	46%	=SE.ERRO(AQ22*AO22;"")			1,26
2010 Aug	Ago	76%	76%	60%	41%	186%	1,49
2010 Sep	Set	68%	68%	57%	41%	165%	1,39
2010 Oct	Out	55%	55%	49%	42%	131%	1,17
2010 Nov	Nov	41%	41%	42%	43%	95%	0,99
2010 Dec	Dez	27%	27%	29%	44%	60%	0,66
2011 Jan	Jan	25%	25%	30%	45%	55%	0,66
2011 Feb	Fev	36%	36%	38%	46%	77%	0,82
2011 Mar	Mar	34%	34%	41%	47%	72%	0,86
2011 Apr	Abr	37%	37%	47%	47%	79%	1,00
2011 May	Mai	46%	46%	48%	47%	98%	1,01
2011 Jun	Jun	45%	45%	50%	47%	95%	1,06
2011 Jul	Jul	61%	61%	59%	47%	129%	1,26
2011 Aug	Ago	88%	88%	70%	47%	187%	1,49
2011 Sep	Set	74%	74%	64%	46%	160%	1,39
2011 Oct	Out	54%	54%	55%	47%	114%	1,17
2011 Nov	Nov	37%	37%	47%	48%	78%	0,99
2011 Dec	Dez	27%	27%	31%	48%	56%	0,66

Legenda: obtenção do modelo 2: multiplicação da tendência, captada através de médias móveis, pelos índices sazonais.

Anexo IV- Detalhes dos modelos SARIMA- *Views*

Null Hypothesis: OCC has a unit root
 Exogenous: Constant
 Lag Length: 11 (Automatic - based on SIC, maxlag=11)

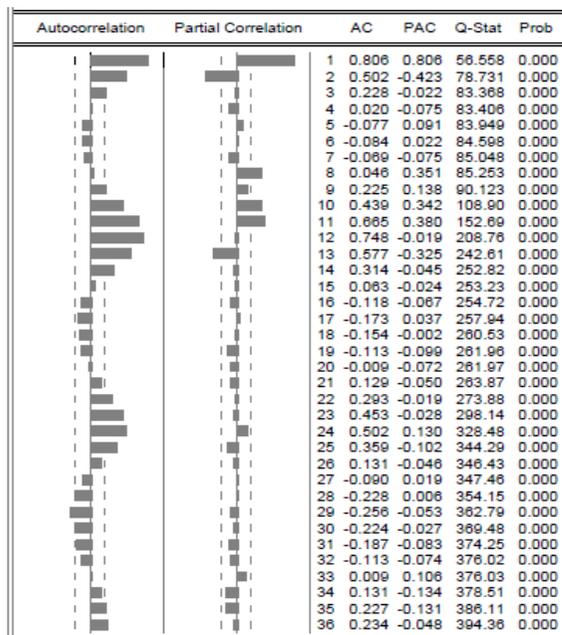
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-0.842390	0.8006
Test critical values:		
1% level	-3.524233	
5% level	-2.902358	
10% level	-2.588587	

Null Hypothesis: D(OCC) has a unit root
 Exogenous: Constant
 Lag Length: 10 (Automatic - based on SIC, maxlag=11)

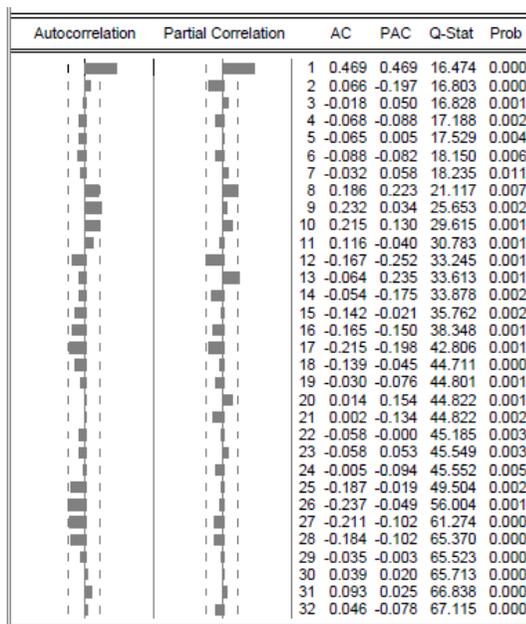
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-9.759659	0.0000
Test critical values:		
1% level	-3.524233	
5% level	-2.902358	
10% level	-2.588587	

Legenda: Teste de raízes unitárias da série original

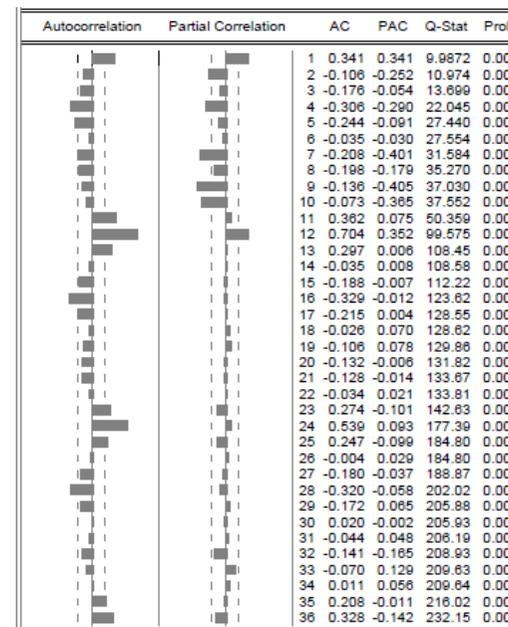
Legenda: Teste de raízes unitárias após 1ª diferenciação



Legenda: Correlograma da série original



Correlograma da série com 1 diferenciação simples



Correlograma da série com 1 diferenciação sazonal

Anexo V- Estimativas dos parâmetros dos modelos SARIMA- *Eviews*

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.054423	0.012377	4.397080	0.0000
AR(1)	0.528279	0.101630	5.198043	0.0000
SAR(12)	-0.318657	0.124607	-2.557308	0.0128
SIGMASQ	0.003848	0.000748	5.145592	0.0000
R-squared	0.306758	Mean dependent var	0.053956	
Adjusted R-squared	0.276174	S.D. dependent var	0.075022	
S.E. of regression	0.063827	Akaike info criterion	-2.588937	
Sum squared resid	0.277025	Schwarz criterion	-2.462456	
Log likelihood	97.20174	Hannan-Quinn criter.	-2.538585	
F-statistic	10.02994	Durbin-Watson stat	1.734909	
Prob(F-statistic)	0.000015			

Legenda: output do modelo 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.055221	0.010279	5.372026	0.0000
AR(1)	0.314574	0.485231	0.648297	0.5190
AR(2)	-0.049609	0.309897	-0.160081	0.8733
SAR(12)	-0.356931	0.133046	-2.682759	0.0092
MA(1)	0.368925	0.451688	0.816769	0.4170
SIGMASQ	0.003585	0.000671	5.344635	0.0000
R-squared	0.354050	Mean dependent var	0.053956	
Adjusted R-squared	0.305114	S.D. dependent var	0.075022	
S.E. of regression	0.062538	Akaike info criterion	-2.597088	
Sum squared resid	0.258127	Schwarz criterion	-2.407365	
Log likelihood	99.49515	Hannan-Quinn criter.	-2.521559	
F-statistic	7.235019	Durbin-Watson stat	1.977550	
Prob(F-statistic)	0.000019			

Legenda: output do modelo 3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.055180	0.010530	5.240177	0.0000
AR(1)	0.249244	0.192130	1.297268	0.1990
SAR(12)	-0.354815	0.128417	-2.762999	0.0074
MA(1)	0.430628	0.187798	2.293041	0.0250
SIGMASQ	0.003588	0.000671	5.347938	0.0000
R-squared	0.353457	Mean dependent var	0.053956	
Adjusted R-squared	0.314858	S.D. dependent var	0.075022	
S.E. of regression	0.062098	Akaike info criterion	-2.624289	
Sum squared resid	0.258364	Schwarz criterion	-2.466187	
Log likelihood	99.47440	Hannan-Quinn criter.	-2.561348	
F-statistic	9.157029	Durbin-Watson stat	1.971573	
Prob(F-statistic)	0.000006			

Legenda: output do modelo 2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.055339	0.009023	6.133443	0.0000
AR(12)	-0.321717	0.134953	-2.383914	0.0199
MA(1)	0.580667	0.108022	5.375445	0.0000
SIGMASQ	0.003696	0.000674	5.485622	0.0000
R-squared	0.334103	Mean dependent var	0.053956	
Adjusted R-squared	0.304725	S.D. dependent var	0.075022	
S.E. of regression	0.062556	Akaike info criterion	-2.627634	
Sum squared resid	0.266098	Schwarz criterion	-2.501153	
Log likelihood	98.59483	Hannan-Quinn criter.	-2.577281	
F-statistic	11.37264	Durbin-Watson stat	1.792291	
Prob(F-statistic)	0.000004			

Legenda: output do modelo 4

Anexo VI- Correlograma dos resíduos referentes aos modelos SARIMA- *Eviews*

Modelo 1:

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.128	0.128	1.2272	
		2 -0.176	-0.196	3.5913	
		3 -0.029	0.024	3.6573	0.056
		4 -0.018	-0.055	3.6837	0.159
		5 -0.088	-0.083	4.3018	0.231
		6 -0.151	-0.147	6.1410	0.189
		7 -0.147	-0.151	7.9096	0.161
		8 0.198	0.197	11.188	0.083
		9 0.136	0.019	12.759	0.078
		10 0.071	0.119	13.191	0.105
		11 0.178	0.178	15.951	0.068
		12 -0.007	-0.057	15.956	0.101
		13 0.013	0.114	15.971	0.142
		14 -0.043	-0.037	16.138	0.185
		15 -0.131	-0.005	17.746	0.167
		16 -0.103	-0.076	18.747	0.175
		17 -0.172	-0.201	21.605	0.119
		18 -0.110	-0.103	22.808	0.119
		19 0.030	-0.160	22.896	0.153
		20 0.120	0.067	24.360	0.144
		21 0.071	-0.060	24.894	0.164
		22 -0.006	-0.080	24.897	0.205
		23 -0.029	-0.026	24.990	0.248
		24 0.025	-0.040	25.060	0.294
		25 -0.137	-0.057	27.184	0.248
		26 -0.147	-0.016	29.673	0.196
		27 -0.134	-0.076	31.795	0.164
		28 -0.106	-0.097	33.158	0.158
		29 -0.002	-0.065	33.159	0.192
		30 0.042	-0.012	33.384	0.222
		31 0.125	0.076	35.424	0.191
		32 0.031	-0.086	35.553	0.223

Modelo 2:

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.011	0.011	0.0084	
		2 -0.000	-0.001	0.0084	
		3 -0.006	-0.006	0.0113	
		4 -0.008	-0.008	0.0168	0.897
		5 -0.106	-0.106	0.9185	0.632
		6 -0.085	-0.084	1.4998	0.682
		7 -0.160	-0.163	3.6036	0.462
		8 0.211	0.217	7.2929	0.200
		9 0.110	0.112	8.3201	0.216
		10 0.054	0.046	8.5733	0.285
		11 0.200	0.197	12.053	0.149
		12 -0.002	-0.047	12.053	0.210
		13 0.021	0.043	12.094	0.279
		14 -0.058	-0.037	12.404	0.334
		15 -0.145	-0.067	14.360	0.278
		16 -0.099	-0.074	15.297	0.289
		17 -0.177	-0.224	18.343	0.192
		18 -0.096	-0.087	19.260	0.202
		19 -0.003	-0.145	19.261	0.255
		20 0.098	0.049	20.243	0.262
		21 0.044	0.011	20.443	0.308
		22 0.011	-0.089	20.456	0.368
		23 -0.055	-0.050	20.780	0.410
		24 0.011	-0.031	20.793	0.472
		25 -0.163	-0.063	23.815	0.357
		26 -0.128	-0.024	25.727	0.314
		27 -0.144	-0.060	28.166	0.253
		28 -0.106	-0.090	29.539	0.242
		29 -0.006	-0.048	29.544	0.287
		30 0.010	-0.023	29.556	0.334
		31 0.107	0.078	31.056	0.315
		32 0.033	-0.052	31.205	0.356

Modelo 3

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.100	0.100	0.7480	
		2 0.107	0.098	1.6189	
		3 -0.041	-0.062	1.7511	0.186
		4 -0.003	-0.004	1.7518	0.416
		5 -0.136	-0.128	3.2289	0.358
		6 -0.082	-0.061	3.7714	0.438
		7 -0.139	-0.104	5.3579	0.374
		8 0.196	0.231	8.5433	0.201
		9 0.140	0.133	10.208	0.177
		10 0.104	0.017	11.133	0.194
		11 0.209	0.189	14.946	0.092
		12 -0.005	-0.092	14.948	0.134
		13 0.025	0.026	15.004	0.182
		14 -0.079	-0.031	15.574	0.212
		15 -0.164	-0.105	18.091	0.154
		16 -0.143	-0.077	20.034	0.129
		17 -0.207	-0.238	24.193	0.062
		18 -0.108	-0.065	25.340	0.064
		19 -0.019	-0.117	25.377	0.087
		20 0.091	0.060	26.225	0.095
		21 0.037	-0.015	26.367	0.120
		22 0.022	-0.104	26.419	0.152
		23 -0.070	-0.034	26.954	0.172
		24 -0.008	-0.021	26.961	0.213
		25 -0.196	-0.066	31.309	0.115
		26 -0.158	-0.026	34.190	0.081
		27 -0.178	-0.063	37.947	0.047
		28 -0.132	-0.097	40.045	0.039
		29 -0.021	-0.032	40.100	0.050
		30 0.013	-0.009	40.120	0.065
		31 0.104	0.066	41.523	0.062
		32 0.037	-0.067	41.705	0.076

Modelo 4:

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.008	0.008	0.0043	
		2 0.004	0.004	0.0057	
		3 0.012	0.012	0.0164	
		4 -0.013	-0.013	0.0299	
		5 -0.101	-0.101	0.8490	0.357
		6 -0.081	-0.081	1.3792	0.502
		7 -0.161	-0.162	3.4920	0.322
		8 0.213	0.223	7.2810	0.122
		9 0.108	0.115	8.2683	0.142
		10 0.052	0.048	8.5019	0.204
		11 0.204	0.191	12.122	0.097
		12 -0.002	-0.048	12.122	0.146
		13 0.021	0.038	12.160	0.204
		14 -0.059	-0.045	12.475	0.255
		15 -0.148	-0.069	14.517	0.206
		16 -0.098	-0.078	15.436	0.218
		17 -0.177	-0.223	18.484	0.140
		18 -0.059	-0.087	19.442	0.149
		19 -0.005	-0.144	19.445	0.194
		20 0.093	0.050	20.323	0.206
		21 0.043	-0.008	20.512	0.249
		22 0.008	-0.088	20.520	0.304
		23 -0.056	-0.052	20.861	0.345
		24 0.008	-0.032	20.868	0.405
		25 -0.165	-0.062	23.960	0.295
		26 -0.131	-0.025	25.938	0.254
		27 -0.144	-0.059	28.380	0.202
		28 -0.109	-0.088	29.806	0.191
		29 -0.007	-0.047	29.813	0.231
		30 0.005	-0.024	29.816	0.275
		31 0.106	0.080	31.272	0.260
		32 0.034	-0.049	31.430	0.298