Previsão da qualidade do ar em Delhi

Maria Pedroso
Dep. de Ciência de Computadores
FCUP
Porto, Portugal
up201805037@up.pt

Nuno Fernandes
Dep. de Ciência de Computadores
FCUP
Porto, Portugal
up202109069@up.pt

Ricardo Andrade
Dep. de Ciência de Computadores
FCUP
Porto, Portugal
up201805015@up.pt

***Abstract*—A qualidade do ar em Delhi coloca sérios riscos para a população. O índice de qualidade do ar (AQI) tem sido utilizado por diversos investigadores para analisar e prever os valores da qualidade do ar. Desta forma, o objetivo do presente estudo foi analisar os dados de observações diárias do AQI em Delhi a partir do ano de 2015 e determinar o algoritmo com maior capacidade de previsão. Testámos um modelo Arima, um modelo exponencial e uma rede neuronal (LSTM). O modelo que apresentou melhor capacidade preditiva num horizonte temporal de um ano foi o LSTM. Este estudo suporta investigações recentes acerca da adequabilidade das LSTM na previsão de séries temporais. Os resultados obtidos servem como ponto de referência a futuros estudos, assim como poderá auxiliar os decisores políticos no planeamento de intervenções com vista ao combate da poluição do ar.**

I. Introdução

Os níveis preocupantes de poluição do ar têm aumentado a consciencialização da população para a qualidade do ar, particularmente em zonas urbanas [1]. A poluição do ar afeta severamente a saúde humana, a fauna, flora e coloca em causa a preservação de todos os ecossistemas [2]. Como tal, os níveis futuros de qualidade do ar tem sido um tópico fulcral na agenda dos principais decisores políticos em todo o mundo.

Na Índia, particularmente na cidade de Delhi, a poluição do ar coloca sérios riscos para a saúde humana [3]. O Inverno em Delhi é severo, enevoado, e conhecido pelos seus níveis elevados de poluição do ar [4]. Assim, a previsão da qualidade do ar tem sido alvo de interesse por diversos investigadores [5], [6].

A qualidade do ar, medida através do índice de qualidade do ar (AQI), é um indicador importante para a população perceber facilmente se os níveis de poluição do ar acarretam perigos para a sua saúde [6]. E ainda, serve como suporte aos decisores políticos na tomada de decisão de estratégias com vista à mitigação deste perigo [6].

Os principais objetivos do presente estudo são compreender as diferenças na qualidade do ar em Delhi ao longo dos anos e determinar qual o melhor algoritmo de previsão dos valores diários do AQI.

A presente investigação tem implicações teóricas, já que irá comparar a adequabilidade de diferentes modelos aos dados da qualidade do ar em Delhi, e também implicações práticas uma vez que permitirá aos cidadãos ter uma estimativa futura da qualidade do ar, assim como poderá auxiliar os decisores políticos no planeamento de estratégias de combate à poluição.

II. Método

A. Características do dataset

A base de dados com os valores diários da qualidade do ar na Índia [7] do período de 2015 a 2020 já foi usado em vários estudos anteriores [13], [14]. O dataset é composto por 16 variáveis: cidade da recolha da medição, data da medição, concentração de poluentes por dia presentes no ar, e, por fim, duas variáveis correspondentes à classificação da qualidade do ar. Na presente investigação apenas iremos utilizar os dados referentes à cidade de Delhi e ao indicador AQI.

O AQI é definido como Air quality index e usa uma escala de 0 a 500 para classificar a qualidade do ar. O limiar para que o ar seja considerado prejudicial é 201 (Apêndice A). Este índice é calculado através dos valores designados como sub-índices dos poluentes, que por sua vez, são calculados usando a média do valor da concentração em 24 horas (8 horas para CO e O3) e o health breakpoint concentration range. Porém, é necessário haver pelo menos três poluentes com medição e um deles têm de ser ou PM2.5 ou PM10. Por fim, o AQI corresponde ao pior sub-índice obtido.

B. Pré-processamento e análise exploratória

Numa fase inicial efetuámos interpolação de dez registos de observações. De seguida, procedemos à divisão da série em duas componentes, a componente de treino (82%) e teste (18%). A componente de treino, constituída pelas observações compreendidas entre Janeiro de 2015 e Junho de 2019, irá ser utilizada para ajustar os modelos, enquanto que a componente de teste, constituída pelas observações de Julho de 2019 a Julho de 2020, permitirá avaliar a qualidade das previsões.

Posteriormente, avaliámos as características da série temporal (Fig.1). Verificamos um padrão decrescente do AQI ao longo dos anos, por isso podemos concluir que existe tendência negativa, logo uma melhoria na qualidade do ar (Apêndice L). Relativamente, à sazonalidade observam-se ciclos constantes ao longo do tempo, concluindo que existe sazonalidade, o que também pode ser comprovado através do *seasonal plot* (Apêndice B). Para além disso, através dos *seasonal plot* podemos concluir que os valores do AQI são mais elevados para os meses de Janeiro e Novembro.

Analisando o ACF confirma-se que a série apresenta uma tendência, traduzida pelo seu lento decaimento para 0 (Apêndice D).

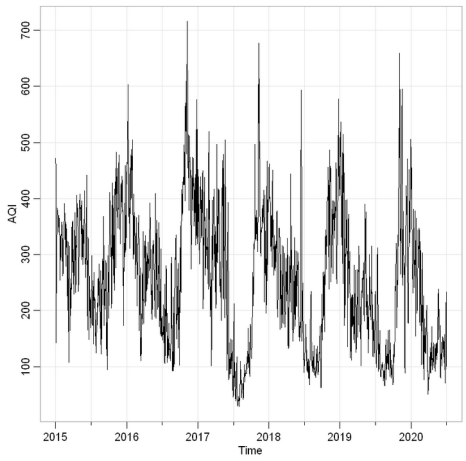

Fig. 1: Série temporal do AQI de Janeiro 2015 a Julho de 2020

C. Modelos ARIMA e SARIMA

Para a escolha dos parâmetros destes modelos foram introduzidas duas funções do R a partir das quais se conclui que, para tornar a série temporal estacionária é necessária a aplicação de uma diferença simples bem como de uma diferença sazonal (Apêndice C). No entanto, o erro inerente a estas funções faz com que uma verificação destes resultados através de outros métodos seja essencial. Portanto, será introduzida uma análise ao ACF e PACF bem como a implementação de testes de raíz unitária, nomeadamemte o teste ADF (Augmented Dickey–Fuller) e o teste KPSS (Kwiatkowski-Phillips-Schmidt-Shin) cujas hipóteses nulas são, respetivamente, H_0 : "A série possui uma raíz unitária" e H_0 : "A série é estacionária". Alertamos para a importância da implementação conjunta destes dois testes, de forma a obterem-se resultados mais robustos. [8]

Analisando a série temporal, ACF e PACF conclui-se que a série apresenta tendência bem como sazonalidade (Apêncide D), logo não é estacionária. Este dado é corroborado com os testes estatísticos (Apêndice E) e evidencia a necessidade da implementação de uma diferença simples. Repare-se que usualmente a implementação da diferença sazonal deve ser precedida à diferença simples no entanto, devido à natureza diária dos dados e à sazonalidade anual, a implementação da diferença sazonal implica a perda de 365 observações. Deste modo, foi optado primeiro pela aplicação da diferença simples com o intuito de obter a estacionariedade e contornar esta perda de informação.

Aplicada a diferença simples, procedeu-se novamente à análise descrita anteriormente, onde se verificou através do ACF e PACF que a série é aproximadamente estacionária (Apêndice F), mais uma vez corroborado com os testes estatísticos (Apêndice G). Assim, dá-se uma contradição de resultados.

Portanto, tendo em conta a dicotomia das conclusões obtidas e assumindo que a perda das 365 observações não acarreta consequências negativas para as previsões, a modelação será realizada com a aplicação de uma diferença simples, consistindo na escolha dos parâmetros p_1, q_1 no modelo ARIMA$(p_1, 1, q_1)$ e com a aplicação de uma diferença sazonal e uma simples, consistindo na escolha dos parâmetros p_2, q_2 no modelo SARIMA$(p_2, 1, q_2)(0, 1, 0)_{365}$.

1) Modelos ARIMA: De forma a sistematizar a escolha da ordem dos polinómios auto-regressivo e média-móvel, iniciou-se com um valor reduzido de q e prossegui-se com a testagem de valores gradualmente mais pequenos de p, partindo de um valor elevado. Seguidamente, fixou-se o valor p num valor adequado para os dados tendo em conta os resultados obtidos anteriormente e prossegui-se com a testagem de valores gradualmente mais pequenos de q, partindo de um valor elevado. Por último, foi verificado o modelo indicado pelo R como sendo o mais adequado para os dados, através do comando *auto.arima()* desconsiderando a componente sazonal. Repare-se que durante este procedimento, sempre que existissem coeficientes não significativos, procedeu-se à sua anulação com o intuito de se obterem modelos mais ajustados para os dados. Assim, de entre os vários modelos testados será apresentada uma amostra ilustrativa das várias fases descritas acima: ARIMA$(1, 1, 8)$ com e sem coeficientes removidos, ARIMA$(3, 1, 6)$, ARIMA$(1, 1, 2)$ e, por fim, ARIMA$(3, 1, 2)$. Constatou-se que todos estes modelos apresentam resíduos não correlacionados, evidenciado no teste de *Ljung-Box* e aparentam descrever bem os dados, pelo que a escolha do mais adequado deve residir nos valores de AIC obtidos sendo que quanto menor, melhor (Apêndice H). Portanto, conclui-se que o modelo mais indicado, é o modelo ARIMA(8,1,1) com os coeficientes não significativos anulados (Apêndice I).

2) Modelos SARIMA: De forma a sistematizar a escolha da ordem dos polinómios auto-regressivo e média-móvel, foi seguido o mesmo paradigma aquele usado nos modelos ARIMA. Inclusive, foi novamente aplicado o comando *auto.arima()* considerando a componente sazonal, o qual indicou o modelo SARIMA$(3, 0, 0)(0, 1, 0)_{365}$. Este modelo considera apenas a aplicação de uma diferença sazonal, contrastando com as conclusões obtidas anteriormente. Este dado ilustra a necessidade de uma exploração dos dados cuidadosa, visto que a partir de diferentes métodos podem surgir conclusões distintas que devem ser abordadas e testadas de forma a obterem-se os melhores resultados. De entre os vários modelos testados, salientamos a seguinte amostra ilustrativa das várias fases da escolha dos parâmetros: SARIMA$(1, 1, 5)(0, 1, 0)_{365}$, SARIMA$(4, 1, 4)(0, 1, 0)_{365}$, SARIMA$(2, 1, 2)(0, 1, 0)_{365}$, SARIMA$(1, 1, 2)(0, 1, 0)_{365}$ e SARIMA$(3, 0, 0)(0, 1, 0)_{365}$. Constatou-se que todos estes modelos apresentam resíduos não correlacionados, evidenciado no teste de *Ljung-Box* e aparentam descrever bem os dados, pelo que a escolha do mais adequado deve residir nos valores de AIC obtidos que quanto menor, melhor (Apêndice J). Portanto, conclui-se que o modelo mais indicado, é o modelo SARIMA$(1, 1, 2)(0, 1, 0)_{365}$ (Apêndice K).

Comparando os dois modelos ARIMA e SARIMA cujo desempenho foi superior conclui-se, através da dos valores do AIC (Apêndice H e J), que o modelo SARIMA

apresenta um valor significativamente inferior pelo que é mais adequado para estes dados. Deste modo, o modelo SARIMA$(1,1,2)(0,1,0)_{365}$ será utilizado para efeitos de previsão, numa fase posterior do trabalho.

D. Modelos ES

Nos modelos ES (Exponential Smoothing), podemos usar a série temporal com tendência e sazonalidade, não sendo necessário aplicar diferenciação ou qualquer outro método para as remover.

Posto isto, poderemos escolher o modelo através das características da série ou através das medidas como AIC, BIC e AICc. Sendo assim, decidimos abordar as duas questões.

Para perceber as características, inicialmente é necessário aplicar uma decomposição a série temporal que consiste num método para representar a série temporal em três componentes: tendência, sazonalidade e os resíduos. A tendência e a sazonalidade podem ser multiplicativas, aditivas ou não ser consideradas, enquanto que os resíduos tem de ser considerados, tendo apenas as duas primeiras classificações, respectivamente.

Para a nossa série obtivemos que a a tendência era aditiva, já que decrescia com um comportamento aproximadamente linear, a sazonalidade era aditiva, uma vez que a variação ao longo do tempo era constante, e, por fim, os resíduos eram multiplicativos já que a variação não era constante ao longo do tempo (Apêndice L). Posto isto, obtivemos a combinação MAA (Resíduos, Tendência, Sazonalidade).

Em relação a aproximação da escolha do modelo através das medidas, decidimos tentar todas as outras combinações possíveis para aferir as melhores. Para isso, usamos as diferentes 9 combinações possíveis entre tendência e a sazonalidade, aplicando primeiro com os resíduos aditivos e, depois com os multiplicativos (Apêndice M e Apêndice N). Para escolher usamos as medidas AIC, BIC e AICc, mas como, por vezes, os melhores modelos tinham previsões que estavam longe da realidade ou com intervalos de confiança com amplitudes bastante elevadas, decidimos que também iríamos usar a qualidade das previsões como motivo da escolha (Apêndice O).

Dentro de cada uma das tabelas e através das previsões, obtivemos que os melhores modelos seriam: MAA e MNA (Apêndice R, Apêndice N e Apêndice P). Entre estes dois, tendo em conta que o segundo tem medidas inferiores e previsões mais realistas, respetivamente, decidimos que iria ser o escolhido.

E. Modelos LSTM

Long Short-Term Memory (LSTM) [9] é um tipo de uma rede neuronal recorrente (RNN) com a capacidade de armazenar valores passados com o objetivo de prever valores futuros. Uma rede neuronal é constituída por três camadas: 1) uma camada de entrada, 2) uma camada oculta, e 3) uma camada de saída. O número de nódulos na camada de entrada é determinada pelo número de atributos presentes no dataset. Estes nódulos comunicam com a camada de saída através de ligações (sinapses) com os nódulos ocultos na camada escondida. Estes nódulos ocultos são os responsáveis por filtrar e atribuir um peso à importância da informação proveniente da camada de entrada. A particularidade das LSTM, e o motivo da sua adequabilidade na previsão de séries temporais, é que ao contrário das RNN comuns, as LSTM têm componentes adicionais capazes de memorizar longas sequências de dados passados [11].

A lstm proposta foi treinada nos dados brutos, sem qualquer preprocessamento, usando 50 épocas e minimizando o erro quadrado médio através do método de otimização Adam, que é um algoritmo de otimização adaptativo para gradiente descendente estocástico.

III. Resultados

A qualidade do ar em Delhi tem vindo a melhorar ao longo dos anos (Apêndice L). Contudo, os valores relativamente elevados do AQI ainda colocam problemas severos à população. Através do modelo de LSTM conseguimos prever o registo diário do AQI a 365 dias com um RMSE = 46 (Figura 2). Adicionalmente, o modelo exponencial revelou um RMSE = 205 (Apêndice P) e o modelo SARIMA um RMSE = 93 (Apêndice Q).

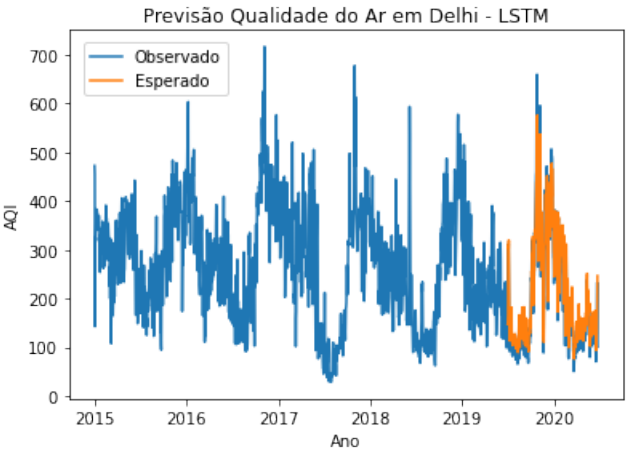

Fig. 2: Previsões 365 dias modelo LSTM

IV. Discussão

O presente estudo teve como objetivos avaliar as principais caraterísticas da qualidade do ar em Delhi ao longo dos anos e testar qual o melhor tipo de modelo para prever a qualidade do ar no próximo ano. Assim, utilizámos três tipos de modelos diferentes para prever o valor do AQI. Os nossos resultados sugerem que o modelo com melhor capacidade de previsão para os valores do AQI é a lstm.

A. Implicações Teóricas

Apesar do AQI ainda se manter elevado, a tendência negativa encontrada poderá ser explicada pelas intervenções nos últimos anos com vista à melhoria da qualidade do ar em Delhi [10].

O efeito de sazonalidade encontrado nos dados atuais corrobora resultados de estudos anteriores, de que o AQI em Delhi aumenta durante os meses de Inverno [4], [10].

O resultado das previsões do nosso modelo LSTM corrobora resultados de estudos anteriores relativamente ao bom desempenho dos modelos LSTM na previsão de séries temporais [11]. Apesar das vantagens dos modelos de aprendizagem automática, uma vez que não carecem de tanto investimento por parte do investigador na parte do pré-processamento e escolha dos parâmetros, os modelos clássicos estatísticos apresentados continuam relevantes dado serem passíveis de interpretação humana [11], [12]. Além disso, os modelos de aprendizagem automática geralmente necessitam de um elevado volume de dados, caso contrário, poderemos incorrer em problemas de sobre-ajustamento [9].

Por último, comparámos o resultados das previsões do nosso melhor modelo com resultados de estudos anteriores, e verificámos que: o nosso modelo (RMSE = 46) obteve melhores resultados do que o modelo ARIMA com previsões para o ano de 2006 [13]; e um valor próximo ao ARIMA de outros autores para o ano de 2020 [14]. Contudo gostaríamos de realçar que estes estudos utilizaram valores mensais do AQI, ao contrário das nossas previsões que foram referentes a observações diárias. E ainda, que o ano de 2020, devido à melhoria da qualidade do ar devido ao confinamento pela COVID-19 poderá ser considerado um outlier, e assim, dificultar a previsão [14], [15].

B. Implicações Práticas

O modelo proposto poderá possibilitar o cidadão comum dos níveis previstos da qualidade do ar para um determinado período, e ainda auxiliar os decisores de políticas públicas no planeamento de intervenções eficazes com vista ao combate à poluição do ar. Por exemplo, poderão adotar medidas atempadas que previna níveis severos de poluição, ou até mesmo, colocar restrições à circulação caso a situação assim o exija.

C. Limitações/Estudos Futuros

O facto dos dados de teste serem referentes ao período do confinamento pelo COVID-19, e este ser considerado um ano atípico, poderá ser considerado como uma limitação ao nosso estudo uma vez que os nossos modelos estatísticos tiveram dificuldades em prever o declive acentuado na melhoria da qualidade do ar a partir de março de 2020.

Estudos futuros poderão utilizar os valores dos nossos modelos como referência para encontrar modelos com melhor capacidade de previsão. Por exemplo, poderão relacionar outras séries temporais, através de correlação cruzada ou adicionar estas séries como preditores adicionais ao modelo LSTM. Algumas variáveis apontadas como das maiores causas de poluição e que que poderão ser tidas em consideração são: a produção de energia eléctrica, a produção industrial, o tráfego automóvel, a produção agrícola, e a temperatura [16].

V. Conclusão

O presente estudo corrobora resultados de estudos anteriores que sugerem uma melhoria da qualidade do ar na cidade de Delhi ao longo dos anos. E ainda, gostaríamos de realçar os níveis frequentemente elevados de poluição do ar registados nos meses de Inverno. O modelo que demonstrou maior capacidade na previsão do AQI foi uma LST, sendo que o desempenho deste modelo aproxima-se, e até mesmo supera resultados obtidos em estudos anteriores. Assim, o nosso modelo poderá ser utilizado como referência em estudos futuros e ter implicações práticas relativamente ao auxílio no planeamento de estratégias de combate à poluição do ar.

References

[1] Kurt, Atakan AND Oktay, Ayse. (2010). *Forecasting air pollutant indicator levels with geographic models 3days in advance using neural networks*. Expert Systems with Applications. 37. 7986-7992. 10.1016/j.eswa.2010.05.093.

[2] Von Schneidemesser Erika, Driscoll Charles, Rieder Harald E. and Schiferl Luke D. (2020) *How will air quality effects on human health, crops and ecosystems change in the future?* Phil. Trans. R. Soc. A. http://doi.org/10.1098/rsta.2019.0330

[3] Sharma, Arun Kumar, Baliyan, Palak and Kumar, Prashant. *Air pollution and public health: the challenges for Delhi, India* Reviews on Environmental Health, vol. 33, no. 1, 2018, pp. 77-86. https://doi.org/10.1515/reveh-2017-0032

[4] Guttikunda, S.K., Gurjar, B.R. *Role of meteorology in seasonality of air pollution in megacity Delhi, India*. Environ Monit Assess 184, 3199–3211 (2012). https://doi.org/10.1007/s10661-011-2182-8

[5] Gourav, Rekhi J.K., Nagrath P., Jain R. (2020) *Forecasting Air Quality of Delhi Using ARIMA Model. In: Jain V., Chaudhary G., Taplamacioglu M., Agarwal M. (eds) Advances in Data Sciences, Security and Applications*. Lecture Notes in Electrical Engineering, vol 612. Springer, Singapore. https://doi.org/10.1007/978-981-15-0372-6_25~

[6] Anikender Kumar, P. Goyal *Forecasting of daily air quality index in Delhi, Science of The Total Environment*, Volume 409, Issue 24, 2011, Pages 5517-5523, ISSN 0048-9697. https://doi.org/10.1016/j.scitotenv.2011.08.069.

[7] https://data.gov.in/(Accessed: 01 dezembro 2021)

[8] https://www.statsmodels.org/dev/examples/notebooks/generated/stationarity_detrending_adf_kpss.html (Accessed: 22 Janeiro 2022)

[9] J. Patterson, *Deep Learning: A Practitioner's Approach*, OReilly Media, 2017.

[10] Mohan, M., Kandya, A. *An Analysis of the Annual and Seasonal Trends of Air Quality Index of Delhi*. Environ Monit Assess 131, 267–277 (2007). https://doi.org/10.1007/s10661-006-9474-4

[11] S. Siami-Namini, N. Tavakoli and A. Siami Namin, *A Comparison of ARIMA and LSTM in Forecasting Time Series*, 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018, pp. 1394-1401, doi: 10.1109/ICMLA.2018.00227.

[12] Anita Yadav, C. K. Jha, Aditi Sharan. 2020. *Optimizing LSTM for time series prediction in Indian stock market*, 167, 2091-2100 Procedia Computer Science. https://doi.org/10.1016/j.procs.2020.03.257

[13] Kumar A, Goyal P. *Forecasting of daily air quality index in Delhi*. Sci Total Environ. 2011 Nov 15;409(24):5517-23. doi: 10.1016/j.scitotenv.2011.08.069.

[14] R. Mangayarkarasi, C. Vanmathi, Mohammad Zubair Khan, Abdulfattah Noorwali, Rachit Jain, Priyansh Agarwal. (2021). *COVID19: Forecasting Air Quality Index and Particulate Matter (PM2.5)*,3,3363-3380. Computers, Materials & Continua. https://doi.org/10.32604/cmc.2021.014991

[15] Shouraseni Sen Roy, Robert C. Balling. *Impact of the COVID-19 lockdown on air quality in the Delhi Metropolitan Region*, Applied Geography, Volume 128, 2021, 102418. https://doi.org/10.1016/j.apgeog.2021.102418

[16] Conserve Energy Future. *Causes, Effects and Impressive Solutions to Air Pollution*. https://www.conserve-energy-future.com/causes-effects-solutions-of-air-pollution.php (Accessed: 25 Janeiro 2022)

APPENDIX

Appendix A

AQI	Remark	Color Code	Possible Health Impacts
0-50	Good		Minimal impact
51-100	Satisfactory		Minor breathing discomfort to sensitive people
101-200	Moderate		Breathing discomfort to the people with lungs, asthma and heart diseases
201-300	Poor		Breathing discomfort to most people on prolonged exposure
301-400	Very Poor		Respiratory illness on prolonged exposure
401-500	Severe		Affects healthy people and seriously impacts those with existing diseases

Fig. 3: Escala de gravidade dos valores de AQI

Appendix B

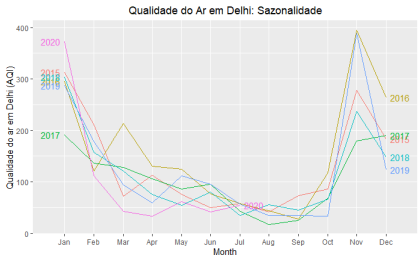

Fig. 4: Gráfico da sazonalidade

Appendix C

```r
```{r}
ndiffs(train) # operador de diferenças
nsdiffs(train) # componente sazonal
```
```

```
[1] 1
The time series frequency has been rounded to support seasonal differencing.[1] 1
```

Fig. 5: Funções do R.

As funções *ndiffs* e *nsdiffs* correspondem, respetivamente, ao número de diferenças simples e ao número de diferenaças sazonais que se devem aplicar a uma série temporal de forma a torná-la estacionária.

Appendix D

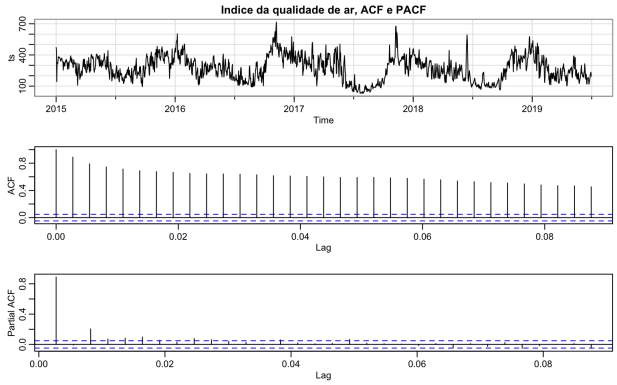

Fig. 6: Série temporal, ACF e PACF.

Analisando o ACF, conclui-se que a série apresenta tendência, traduzida pelo seu lento decaimento para 0, que pode passar despercebida exclusivamente através da observação do primeiro gráfico. Logo, o índice da qualidade do ar não é estacionário e o ACF não tem significado. Inclusive, é possível inferir que a série apresenta sazonalidade, patente na série e ACF e explorada anteriormente. Por outro lado, a série apresenta fases de rápido crescimento seguidas de um lento decrescimento que remetem para a sua não linearidade e consequente inadequabilidade de determinados modelos de previsão.

Appendix E

```
p-value smaller than printed p-value
        Augmented Dickey-Fuller Test

data:  ts
Dickey-Fuller = -4.0386, Lag order = 11, p-value = 0.01
alternative hypothesis: stationary


        KPSS Test for Trend Stationarity

data:  ts
KPSS Trend = 0.18221, Truncation lag parameter = 8, p-value = 0.02267
```

Fig. 7: Testes ADF e KPSS para a série temporal.

Relativamente ao teste ADF o valor-p $= 0.01 < 0.05$, de onde se rejeita a hipótese nula e, relativamente ao teste KPSS o valor-p $\approx 0.05 \leq 0.05$, de onde se rejeita a hipótese nula. Portanto, a rejeição da hipótese nula em ambos testes permite inferir que a série temporal não é estacionária. [4]

Appendix F

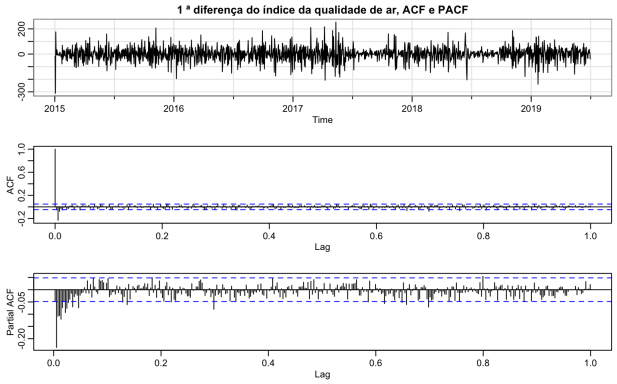

Fig. 8: Série temporal com uma diferença simples, ACF e PACF.

Após a aplicação da 1ª diferença e através da observação do primeiro gráfico infere-se que a série não apresenta tendência nem sazonalidade e possui uma média de zero. Este dado é corroborado com o gráfico do ACF.

Appendix G

Relativamente ao teste ADF o valor-p$= 0.01 < 0.05$, de onde se rejeita a hipótese nula e, relativamente ao teste KPSS o valor-p$= 0.1 \leq 0.05$, de onde se infere que não existe evidência estatística para a rejeição da hipótese nula. Portanto, pode-se concluir que a série temporal é estacionária. [4]

```
p-value smaller than printed p-value
        Augmented Dickey-Fuller Test

data:  ts_1
Dickey-Fuller = -16.021, Lag order = 10, p-value = 0.01
alternative hypothesis: stationary

p-value greater than printed p-value
        KPSS Test for Trend Stationarity

data:  ts_1
KPSS Trend = 0.0040125, Truncation lag parameter = 7, p-value = 0.1
```

Fig. 9: Testes ADF e KPSS para a série temporal com uma diferença simples.

Appendix H

ARIMA(8,1,1)	ARIMA(8,1,1)mod	ARIMA(3,1,6)	ARIMA(1,1,2)	ARIMA(3,2,1)
17582	17576	17578	17585	17581

Fig. 10: AIC dos modelos ARIMA.

Appendix I

```
Series: train
ARIMA(8,1,1)

Coefficients:
         ar1      ar2     ar3  ar4      ar5  ar6  ar7      ar8      ma1
      0.6803  -0.2062  0.0480    0  -0.0463    0    0  -0.0479  -0.8285
s.e.  0.0446   0.0298  0.0301    0   0.0258    0    0   0.0244   0.0378

sigma^2 estimated as 2594:  log likelihood=-8780.89
AIC=17575.78   AICc=17575.84   BIC=17613.6
```

Fig. 11: Modelo ARIMA(8,1,1) com coeficientes não significativo anulados.

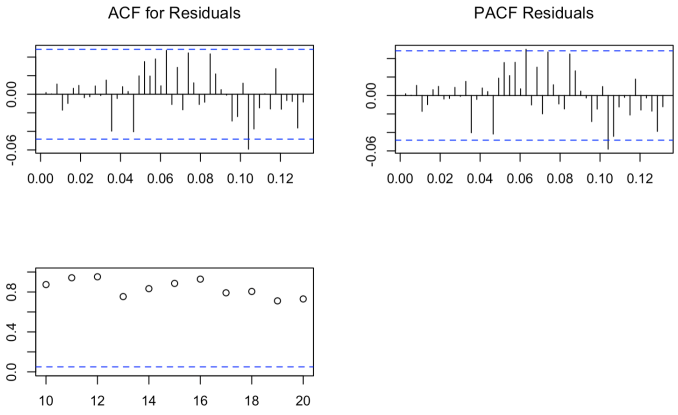

Fig. 12: Teste de Ljung-Box para o modelo ARIMA(8,1,1) com coeficientes não significativo anulados.

Appendix J

SARIMA(1,1,5)(0,1,0)	SARIMA(4,1,4)(0,1,0)	SARIMA(2,1,2)(0,1,0)	SARIMA(1,1,2)(0,1,0)	SARIMA(3,0,0)(0,1,0)
14525	14527	14522	14520	14535

Fig. 13: AIC dos modelos SARIMA.

Tabela com os valores do AIC para os vários modelos.
Appendix K

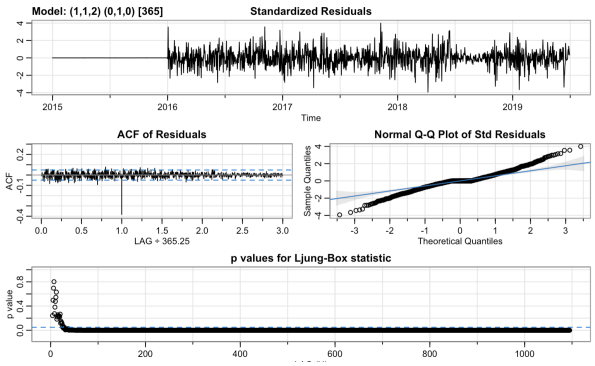

Fig. 14: AIC dos modelos SARIMA.

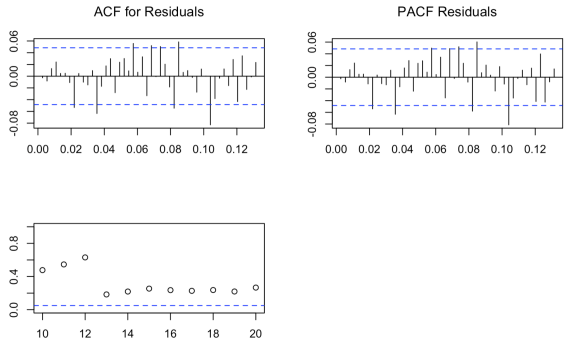

Fig. 15: AIC dos modelos SARIMA.

Appendix L

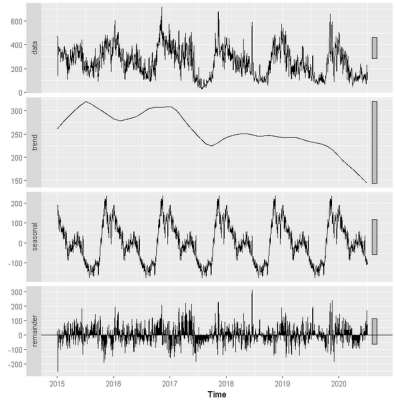

Fig. 16: Decomposição da série temporal

Appendix M

ETS	AIC	BIC	AICc
ANN	21640.92	21657.74	21640.93
ANA	21986.18	24048.96	22151.78
ANM	22048.15	24116.54	22214.75
AAN	21642.89	21665.32	21642.91
AAA	22350.86	24424.85	22518.47
AAM	21951.22	24025.21	22118.82
AAdN	21640.13	21668.16	21640.16
AAdA	22352.58	24432.18	22521.20
AAdM	21920.00	23988.39	22086.60

Fig. 17: Medidas obtidas para os resíduos aditivos

Appendix N

ETS1	AIC1	BIC1	AICc1
MNN	21542.84	21559.65	21542.85
MNA	22502.86	24571.25	22669.46
MNM	21849.30	23917.69	22015.90
MAN	21461.60	21484.02	21461.62
MAM	21878.96	23947.35	22045.57
MAdN	21460.75	21483.17	21460.77
MAdA	22675.31	24749.30	22842.91
MAdM	21890.89	23976.10	22060.52

Fig. 18: Medidas obtidas para os resíduos multiplicativos

Appendix O

Um exemplo dos modelos com as melhores medidas apresentarem previsões longe da realidade ou com intervalos de confiança maus, é um dos melhores modelos para os resíduos multiplicativos é a combinação MAN. Mas se verificarmos o gráfico das previsões com o intervalo verificamos que o intervalo abrange valores incoerentes com a realidade.

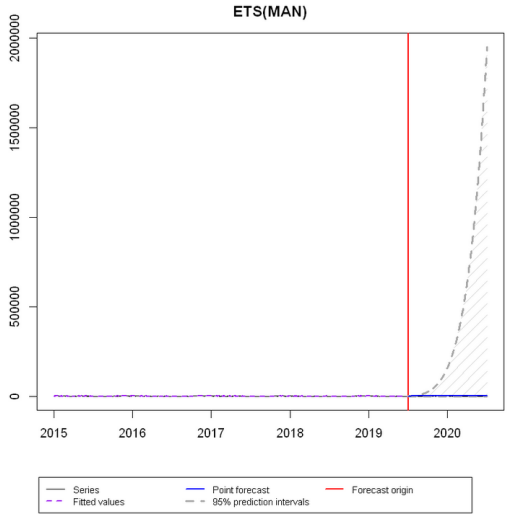

Fig. 19: Previsões obtidas para o modelo ES com a combinação MAN

Appendix P

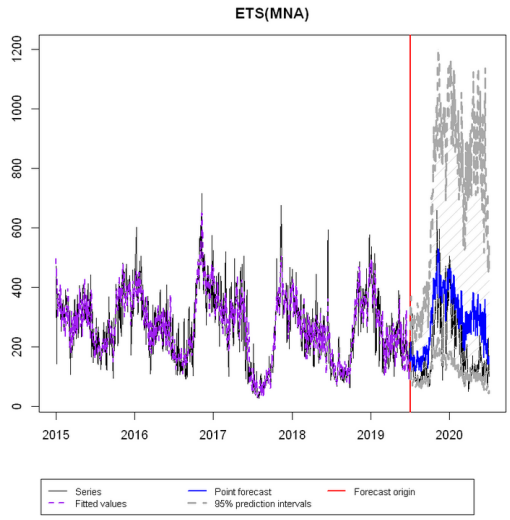

Fig. 20: Previsões obtidas para o modelo ES com a combinação MNA

Appendix Q

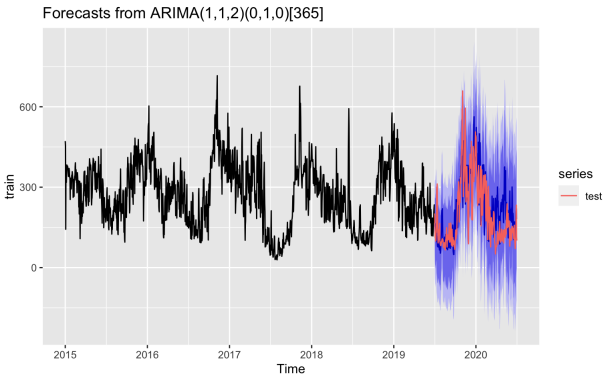

Fig. 21: Previsões obtidas para o modelo SARIMA.

Appendix R

ETS	AIC	AICc	BIC
MAA	22670.89	22839.5	24750.49

Fig. 22: Medidas obtidas para o modelo ES, MAA

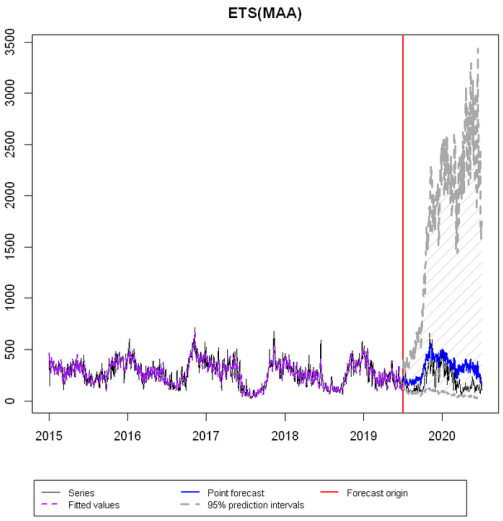

Fig. 23: Previsões para o modelo ES, MAA