

Previsão da qualidade do ar utilizando modelos ARDL

Fernando Batista

Escola Superior de Tecnologia e Gestão do Politécnico de Leiria

CDRSP - Centre for Rapid and Sustainable Product Development, Polytechnic of Leiria

fernando.batista@ipleiria.pt

Jorge Siopa

Escola Superior de Tecnologia e Gestão do Politécnico de Leiria

jorge.siopa@ipleiria.pt

Resumo

A qualidade do ar nas zonas urbanas afeta a saúde e o bem-estar da população. Sabendo da importância deste problema, é necessário, numa primeira fase, monitorizar todos os parâmetros atmosféricos que podem afetar a qualidade do ar e, numa segunda fase, criar modelos preditivos. Com estes modelos preditivos é possível analisar a importância dos parâmetros atmosféricos que influenciam a concentração dos gases que degradam a qualidade do ar. Utilizando um pacote *Python* para uma análise *AutoRegressive Distributed Lag* (ARDL) de séries temporais, foi possível caracterizar esta importância.

Na tentativa de avaliar o bem-estar dos utilizadores [1] do Campus 2 da Escola Superior de Tecnologia e Gestão do Politécnico de Leiria, foram recolhidos dados a partir da monitorização da qualidade do ar obtidos pela Unidade Móvel de Monitorização da Qualidade do Ar local. A partir destes dados, foi realizado um estudo que visa caracterizar a importância de determinados parâmetros meteorológicos, utilizando para esse fim uma ferramenta de análise preditiva de séries temporais, ARDL (*Statsmodels Python*). Este método é muito utilizado em economia para realizar previsões. As séries temporais do modelo são divididas em variáveis endógenas e exógenas. Os parâmetros que caracterizam as condições atmosféricas são sempre assumidos como variáveis exógenas. As concentrações dos gases na atmosfera que influenciam a qualidade do ar são considerados variáveis endógenas e/ou exógenas.

Foram obtidas as séries temporais para a concentração dos seguintes gases: ozono (O₃), monóxido de carbono (CO), hidrocarbonetos (NO, NO_x e NO₂) e partículas PM10 e PM2,5. Para os parâmetros meteorológicos temos: a precipitação (mmH₂O), a radiação global (Rglobal), a temperatura (Temp), a humidade (Humid), a pressão atmosférica (Press), a velocidade do vento (Vwind) e a sua direção (Dwind). Para tentar caracterizar o tráfego automóvel, estimou-se uma série temporal de horas de ponta (Rhour). Esta série temporal composta por “0” e “1” foi sugerida no trabalho realizado nesta área [2], que utilizou o método dos mínimos quadrados para gerar o modelo preditivo. Para desenhar de forma mais robusta conclusões com séries temporais, precisamos de muitos dados de muitos dias. Apesar de neste estudo apenas termos dados de 17 dias, foi possível chegar a algumas conclusões.

Foram realizados diversos testes estatísticos às séries temporais para avaliar a sua estacionariedade, tais como ADF e KPSS. A relação entre os dois tipos de variáveis também foi avaliada [3].

O método ARDL utiliza dados prévios das variáveis e consegue captar o efeito sazonal da série. A equação (1) mostra o processo iterativo caracterizado por desfasamentos.

$$Y_t = \delta + \sum_{i=1}^{P-1} \gamma_i S_{[(\text{mod}(t,P)+1)=i]} + \sum_{p=1}^A \phi_p Y_{t-p} + \sum_{k=1}^M \sum_{j=0}^{Q_k} \beta_{k,j} X_{k,t-j} + \epsilon_t \quad (1)$$

Onde δ é uma constante, $\gamma_i S_{[(\text{mod}(t,P)+1)=i]}$ capta as mudanças sazonais, P é o período da sazonalidade, A é o lag de defasamento da variável endógena, M é o número de variáveis exógenas (X_k), Q_k é o número de lag atrasados de X_k e ϵ_t é o ruído branco.

Métricas estatísticas como o "erro quadrático ajustado", $Adj.R^2$, a "raiz do erro quadrático médio", $RMSE$ e métodos de estacionariedade foram utilizados para chegar a conclusões, Tabela 1.

Tabela 1 – Teste de 6 dias – Métricas estatísticas e estacionariedade dos residuais preditivos

	TEST residuais preditivos - estacionariedade											
	TRAIN - Adj.R2			TEST - RMSE			ARDL1		ARDL2		ARDL3	
	ARDL1	ARDL2	ARDL3	ARDL1	ARDL2	ARDL3	ADF(c)	KPSS(c)	ADF(c)	KPSS(c)	ADF(c)	KPSS(c)
O ₃	0.978	0.991	0.992	9.87	8.58	8.24	Sim	Não	Sim	Sim	Sim	Não
CO	0.877	0.905	0.907	0.19	0.16	0.14	Sim	Não	Sim	Sim	Sim	Não
NO	0.746	0.815	0.815	25.41	18.77	18.77	Sim	Não	Sim	Sim	Sim	Sim
NO _x	0.769	0.923	0.923	32.98	17.00	17.00	Sim	Não	Sim	Sim	Sim	Sim
NO ₂	0.895	0.975	0.975	8.00	7.74	5.80	Sim	Não	Sim	Sim	Sim	Não
PM10	0.990	0.990	0.990	13.91	10.32	10.32	Não	Sim	Sim	Sim	Sim	Sim
PM2,5	0.993	0.994	0.994	10.82	6.06	6.06	Sim	Sim	Sim	Sim	Sim	Sim

Foram realizados três modelos de ARDL, o primeiro com parâmetros meteorológicos como variáveis exógenas, o segundo com a concentração dos gases poluentes como variáveis exógenas e o terceiro foi uma combinação das concentrações dos gases e dos parâmetros meteorológicos identificadas no primeiro modelo. Os três modelos tiveram um bom desempenho nos dados de treino, mas, quando o RMSE e a estacionariedade dos resíduos preditivos foram avaliados, o segundo modelo, que tem apenas variáveis de concentração dos gases poluentes, teve o melhor ajuste e foi o único a ter resíduos preditivos estacionários, Tabela 1.

A partir dos dados recolhidos podemos inferir que existe uma ligação mais forte entre as concentrações dos gases poluentes na atmosfera do que destas com os parâmetros meteorológicos, para previsões de curto prazo.

Referências

- [1] M. Pacione. (2003). *Urban environmental quality and human wellbeing – A social geographical perspective*. Landscape and Urban Planning, 65(1-2):19–30.
- [2] J. Siopa, B. Gonçalves, L. Aires, and M. Gaspar. *Monitoring and prediction of the air quality towards sustainable work environments*. In: JOCLAD 2023, Viana do Castelo, abril 2023. Associação Portuguesa de Classificação e Análise de Dados.
- [3] Eberlein, E. (2001). Application of Generalized Hyperbolic Lévy Motions to Finance. In: Barndorff-Nielsen, O.E., Resnick, S.I., Mikosch, T. (eds) Lévy Processes. Birkhäuser, Boston, MA.