

Previsão da qualidade do ar utilizando modelos ARDL

Fernando Batista*, **Jorge Siopa****

*ESTG, CDRSP, Instituto Politécnico de Leiria

** ESTG, Instituto Politécnico de Leiria



Séries Temporais

Gases poluentes

Parâmetros meteorológicos

Dados

Frequência: 15 min

TRAIN: 11 dias

TEST: 6 dias

2009							Fevereiro									Março
Sex	Sab	Dom	Seg	Ter	Qua	Qui	Sex	Sab	Dom	Seg	Ter	Qua	Qui	Sex	Sab	Dom
13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	1
TRAIN											TEST					

Python package: Numpy, Pandas, Sklearn, Statsmodels and Matplotlib

Descrição dos dados recolhidos

Dados estatísticos das séries

	Gases poluentes							Rhour	Parâmetros meteorológicos						
	O3	CO	NO	NOX	NO2	PM10	PM25		mmH2O	Press	Rglobal	Temp	Humid	Vwind	Dwind
count	1632	1632	1632	1632	1632	1632	1632	1632	1632	1632	1632	1632	1632	1632	1632
mean	18.23	0.33	14.52	32.42	18.01	47.28	35.43	0.11	0.00	1015.3	185.52	12.18	71.89	0.94	0.93
std	16.67	0.26	30.63	39.07	12.99	21.60	16.33	0.31	0.03	4.23	269.25	6.34	23.31	0.82	0.07
min	0.65	0.00	0.00	0.23	0.54	9.32	6.23	0.00	0.00	1007.0	0.45	1.96	19.71	0.08	0.20
50%	9.66	0.29	4.55	21.64	14.70	46.06	36.07	0.00	0.00	1015.2	0.81	11.18	80.75	0.62	0.94
max	51.32	1.48	384.35	433.52	79.13	94.21	75.69	1.00	0.62	1025.7	807.47	25.17	99.10	5.60	1.00

Rhour - Hora de ponta (0 ou 1)

O3 - Azoto

CO - Monóxido de Carbono

NO - Monóxido de Nitrogênio

NOX - Todos os óxidos de Nitrogênio

NO2 - Dióxido de Nitrogênio

PM10 - Partículas $\leq 10 \mu\text{m}$

PM25 - Partículas $\leq 2,5 \mu\text{m}$

mmH2O - Precipitação

Press - Pressão atmosférica

Rglobal - Radiação global

Temp - Temperatura

Humid - Humidade

Vwind - Velocidade do vento

Dwind - Direcção do vento

ARDL - AutoRegressive Distributed Lag

Dados estatísticos das séries														
Gases poluentes							Parâmetros meteorológicos							
O3	CO	NO	NOX	NO2	PM10	PM25	Rhour	mmH2O	Press	Rglobal	Temp	Humid	Vwind	Dwind
3 modelos de ARDL														
ARDL1	Variáveis Endógenas						Variáveis Exógenas							
ARDL2	Variáveis Endógenas & Variáveis Exógenas													
ARDL3	Variáveis Endógenas & Variáveis Exógenas						Variáveis Exógenas							

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

Variável Endógena

Constante

Sazonalidade
 $P = 4 * 24 = 96 \text{ lag}$

Variável Endógena com lag p

Variáveis Exógenas com lag j

Ruído branco

Numero máximo de lag: $A \leq 4$ e $Q_k \leq 4$

Estacionaridade das séries temporais

Teste de Estacionaridade - Determinar se os *valores médios* e a *variância* de uma série permanecem constantes ou com uma tendência ao longo do tempo.

ADF - Augmented Dickey Fuller test

H0 (nula): Tem uma raiz unitária

H1 (alternativa): Série estacionária (p-value < 0.05)

Não foi feita qualquer diferenciação para tornar as séries estacionárias

KPSS - Kwiatkowski-Phillips-Schmidt-Shin test.

H0 (nula): Série estacionária (p-value > 0.05)

H1 (alternativa): Tem uma raiz unitária

Teste de Estacionaridade															
	O3	CO	NO	NOX	NO2	PM10	PM25	HPonta	mmH20	Press	Rglobal	Temp	Humid	Vwind	Dwind
ADF(c)	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim
KPSS(c)	Sim	Sim	Não	Não	Sim	Sim	Não	Sim	Não	Não	Sim	Sim	Sim	Sim	Sim
ADF(ct)	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim
KPSS(ct)	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Sim	Sim	Sim	Sim	Sim

Correlações da séries temporais

Correlação de Pearson

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Correlação de Pearson													
	O3	CO	NO	NOX	NO2	PM10	PM25	Rhour	Rglobal	Temp	Humid	Vwind	Dwind
O3	1.000	-0.722	-0.350	-0.435	-0.481	-0.688	-0.717	-0.048	0.662	0.857	-0.787	0.617	0.004
CO		1.000	0.408	0.530	0.629	0.542	0.529	0.068	-0.558	-0.612	0.621	-0.506	-0.034
NO			1.000	0.958	0.521	0.113	0.131	0.269	-0.088	-0.140	0.109	-0.188	-0.003
NOX				1.000	0.743	0.114	0.119	0.367	-0.215	-0.187	0.164	-0.264	-0.024
NO2					1.000	0.073	0.046	0.467	-0.440	-0.232	0.237	-0.349	-0.063
PM10						1.000	0.945	-0.197	-0.333	-0.651	0.605	-0.385	0.065
PM25							1.000	-0.187	-0.352	-0.691	0.641	-0.403	0.066

NOX \approx NO + NO2

Teste de causalidade de Granger

Teste de causalidade

Determinar se uma série temporal X pode prever outra série temporal Y .

Teste de causalidade de Granger

H_0 (nula): X não causa Y

H_1 (alternativa): X causa Y (p-value < 0.05)

Teste de causalidade de Granger - p-value < 0.05													
	O3	CO	NO	NOX	NO2	PM10	PM25	Rhour	Rglobal	Temp	Humid	Vwind	Dwind
O3	1	0	0.001	0.014	0	0	0	0.262	0	0	0	0	0
CO	0	1	0.003	0.509	0	0	0	0	0	0	0	0	0.005
NO	0.001	0	1	0.382	0.365	0.702	0.690	0.069	0.001	0	0.001	0	0.087
NOX	0.001	0	0.006	1	0.007	0.284	0.300	0.001	0	0	0	0	0.028
NO2	0	0	0.003	0.003	1	0	0	0	0	0	0	0	0.001
PM10	0	0	0	0	0	1	0	0.003	0	0	0	0	0.158
PM25	0	0	0	0	0	0	1	0	0	0	0	0	0.039

Critério de informação: BIC

Numero máximo de *lag*: $A \leq 4$ e $Q_k \leq 4$

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

ARDL1	Lags		Q1	Q2	Q3	Q4	Q5	Q6
	P	A	Rglobal	Temp	Humid	Vwind	Dwind	Rhour
O3	96	1				3		
CO	96	3			1	0		0
NO	96	1				0		
NOX	96	1				0		
NO2	96	1		2		0		0
PM10	96	2						
PM25	96	4						

Modelos ARDL 2 e 3

ARDL2	Lags		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
	P	A	O3	CO	NO	NOX	NO2	PM10	PM25	Rhour
O3	96	1	---		3	1				
CO	96	3		---	2		1		1	
NO	96	2	1	2	---					
NOX	96	3	1	2		---	3			
NO2	96	3	1	2	3		---			
PM10	96	2		1				---		
PM25	96	3		0			2		---	

ARDL3	Lags		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13
	P	A	O3	CO	NO	NOX	NO2	PM10	PM25	Rhour	Rglobal	Temp	Humid	Vwind	Dwind
O3	96	1	---		3	1								2	
CO	96	3		---	1		2		1				0		
NO	96	2	1	2	---										
NOX	96	3	1	2		---	3								
NO2	96	3	1	2	3		---		0	0		0			
PM10	96	2		1				---							
PM25	96	3		0			2		---						

Teste de Durbin-Watson

Avaliar a presença de auto-correlação nos resíduos do modelo de regressão.

$$Adj.R^2 = 1 - (1 - R^2) \frac{n - 1}{n - k - 1}$$

Teste de Durbin-Watson			
	ARDL1	ARDL2	ARDL3
O3	2.07	2.04	2.08
CO	2.01	2.02	2.03
NO	2.01	2.00	2.00
NOX	2.00	2.06	2.06
NO2	2.05	2.08	2.06
PM10	2.00	2.01	2.01
PM25	1.97	1.97	1.97



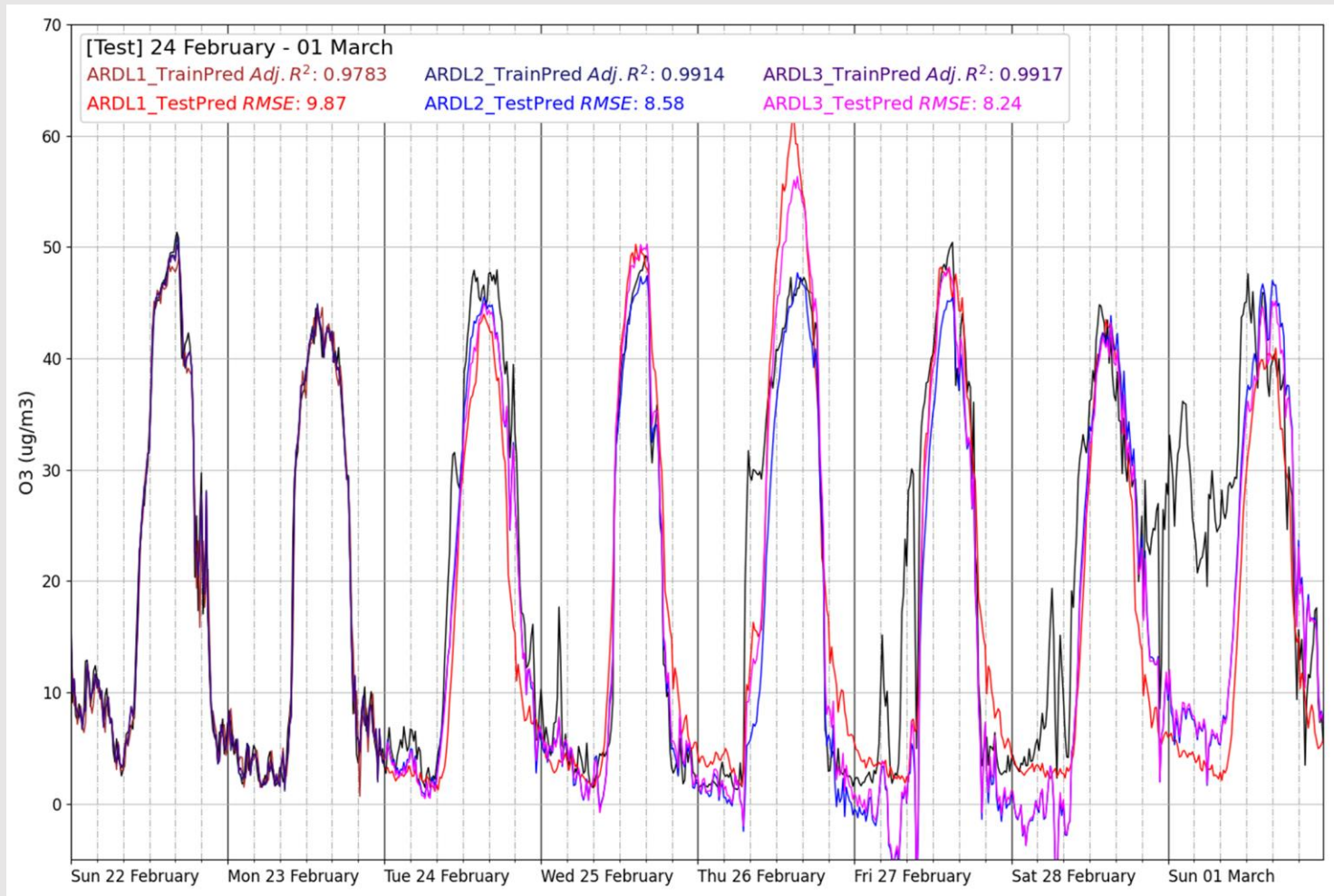
	TRAIN - Adj.R2			TEST - RMSE		
	ARDL1	ARDL2	ARDL3	ARDL1	ARDL2	ARDL3
O3	0.9783	0.9914	0.9917	9.87	8.58	8.24
CO	0.8774	0.9054	0.9065	0.19	0.16	0.14
NO	0.7461	0.8150	0.8150	25.41	18.77	18.77
NOX	0.7686	0.9230	0.9230	32.98	17.00	17.00
NO2	0.8949	0.9745	0.9749	8.00	7.74	5.80
PM10	0.9897	0.9902	0.9902	13.91	10.32	10.32
PM25	0.9931	0.9935	0.9935	10.82	6.06	6.06

$$d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}$$

$$0 < d < 4$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\text{Predicted}_i - \text{Actual}_i)^2}{N}}$$

O3 - TRAIN + TEST com ARDL1, ARDL2 e ARDL3



Resíduo do TEST - Teste de Estacionaridade						
	ARDL1		ARDL2		ARDL3	
	ADF(c)	KPSS(c)	ADF(c)	KPSS(c)	ADF(c)	KPSS(c)
O3	Sim	Não	Sim	Sim	Sim	Não
CO	Sim	Não	Sim	Sim	Sim	Não
NO	Sim	Não	Sim	Sim	Sim	Sim
NOX	Sim	Não	Sim	Sim	Sim	Sim
NO2	Sim	Não	Sim	Sim	Sim	Não
PM10	Não	Sim	Sim	Sim	Sim	Sim
PM25	Sim	Sim	Sim	Sim	Sim	Sim

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



Obrigado pela atenção!

fernandobatista.net