

Air quality prediction for work environments using ARDL models

F. C. Batista^{1,2}, Jorge Siopa¹,

¹ School of Technology and Management, Polytechnic of Leiria

² CDRSP - Centre for Rapid and Sustainable Product Development, Polytechnic of Leiria
fernando.batista@ipleiria.pt jorge.siopa@ipleiria.pt

Air quality in urban areas affects the health and well-being of the population. Knowing the importance of this problem, it is necessary in the first phase to monitor all the atmospheric parameters that can affect air quality and in the second phase, to create predictive models. With these predictive models it is possible to analyze the importance of the atmospheric parameters that influence the gases that degrade air quality. Using a Python package for an Autoregressive Distributed Lag (ARDL) analysis of time series, it was possible to characterize this importance.

Keywords: sustainable campus network, air quality prediction, atmosphere parameters, time series analysis, ARDL model

In an attempt to assess well-being [1] at Campus 2 of the School of Technology and Management of the Polytechnic of Leiria, air quality monitoring data obtained by the on-site Mobile Air Quality Monitoring Unit was collected. A study was carried out using this data to characterise the importance of certain atmospheric parameters, using a predictive time series analysis tool, ARDL (Statsmodels Python). This method is widely used in economics to estimate forecasts. The time series in the model are divided into endogenous and exogenous variables. The parameters that characterize the atmospheric conditions are always assumed to be exogenous variables. The gases in the atmosphere that influence air quality are assumed to be endogenous variables or exogenous variables.

The time series obtained for the gases were ozone (O_3), carbon monoxide (CO), hydrocarbons (NO , NO_X and NO_2) and particles PM_{10} and $PM_{2.5}$. For the atmospheric conditions we have precipitation (mmH_2O), global radiation (R_{global}), temperature ($Temp$), humidity ($Humid$), atmospheric pressure ($Press$), wind speed (V_{wind}) and direction (D_{wind}). To try to characterize car traffic, a time series of rush hour was estimated (R_{hour}). This time series consisting of "0" and "1" was suggested in the work done in this field, which used OLS to generate the predictive model [2]. In order to draw more robust conclusions with time series, we need a lot of data from many days. Although in this study we only have data from 17 days, it was possible to reach some conclusions.

Various statistical tests were carried out on the time series to assess their stationarity, such as ADF and KPSS. The relationship between the two types of variables was also assessed [3].

The ARDL method uses previous data on the variables and manages to capture the seasonal effect of the series. Equation (1) shows of the iterative process characterized by lags.

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

where δ is constant, $\gamma_i S_{[(\text{mod}(t,s)+1)=i]}$ capture seasonal shifts, P is the period of the seasonality, A is the lag length of the endogenous variable, M is the number of exogenous variables X_k , Q_k is included the lag length of X_k and ϵ_t is a white noise shock.

Statistical metrics such as the $Adj.R^2$, $RMSE$ and stationarity methods.

Table 1: Test - 6 days - Statistical metric and stationarity predictive residuals

	RMSE		ARDL1		ARDL2		ARDL3		
	ARDL1	ARDL2	ARDL3	ADF	KPSS	ADF	KPSS	ADF	KPSS
O_3	9.87	8.58	8.24	Yes	No	Yes	Yes	Yes	No
CO	0.19	0.16	0.14	No	No	Yes	Yes	Yes	No
NO	25.41	18.77	18.77	Yes	No	Yes	Yes	Yes	Yes
NOX	32.98	17.00	17.00	Yes	No	Yes	Yes	Yes	Yes
NO_2	8.00	7.74	5.80	Yes	No	Yes	Yes	Yes	No
$MP10$	13.91	10.32	10.32	No	Yes	Yes	Yes	Yes	Yes
$MP2.5$	10.82	6.06	6.06	Yes	Yes	Yes	Yes	Yes	Yes

Three models of ARDL were carried out, the first with atmospheric conditions as exogenous variables, the second with atmospheric gases as exogenous variables and the third was a combination of atmospheric gases and the atmospheric conditions identified in the first model. The three models performed well on the training data, but when the RMSE and the stationarity of the predictive residuals were evaluated, model 2, which only has gas variables, had the best fit and was the only one to have stationary predictive residuals, Table 1. From the collected data we can infer that there is a stronger link between pollutant gases in the atmosphere than with meteorological data, for short time forecasting.

References

- [1] M. Pacione. Urban environmental quality and human wellbeing-a social geographical perspective. *Landscape and Urban Planning*, 65(1):19–30, 2003.
- [2] J. Siopa, B. Goncalves, L. Aires, and M. Gaspar. Monitoring and prediction of the air quality towards sustainable work environments. In *JOCLAD 2023*, Viana do Castelo, April 2023. Associação Portuguesa de Classificação e Análise de Dados.
- [3] Asif Raihan, Dewan Ahmed Muhtasim, Md Nafiul Alam Khan, Monirul Islam Pavel, and Omar Faruk. Nexus between carbon emissions, economic growth, renewable energy use, and technological innovation towards achieving environmental sustainability in Bangladesh. *Cleaner Energy Systems*, 3:100032, 2022.