

ARIMA-Based Forecasting of Air Pollutant Concentrations in North Cyprus: A Multi-Station, Multi-Pollutant Time Series Study

Saba Waseem

Muhammad Ali

Zakria Qadir

zakria.q@sistc.edu.au

Sydney International School of Technology and Commerce

Research Article

Keywords: North Cyprus; ARIMA; time series forecasting; air quality; Mediterranean region

Posted Date: October 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11504281/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

ARIMA-Based Forecasting of Air Pollutant Concentrations in North Cyprus: A Multi-Station, Multi-Pollutant Time Series Study

Saba Waseem^a, Muhammad Ali^a, Zakria Qadir^{b,*}

^a Sustainable Environment and Energy Systems Program, Middle East Technical University Northern Cyprus Campus, Kalkanlı, Güzelyurt, TRNC

^b Department of Information Technology, Sydney International School of Technology and Commerce, Sydney, Australia

* Corresponding author. E-mail address: zakria.q@sistc.edu.au

Abstract

Air pollution monitoring and forecasting in the Turkish Republic of Northern Cyprus (TRNC) remains critically understudied. This study presents the first comprehensive multi-pollutant, multi-station time series forecasting study for North Cyprus, employing Autoregressive Integrated Moving Average (ARIMA) models to predict concentrations of SO₂, NO₂, PM₁₀, and PM_{2.5} across four monitoring stations — Kyrenia, Nicosia, Famagusta, and Teknecik — over the period 2016–2019, with forecasts generated for 2020. Monthly average pollutant concentrations were derived from hourly (SO₂, NO₂) and daily (PM₁₀, PM_{2.5}) observations collected using calibrated instrumentation. Three model types — Simple Seasonal, Winters’ Additive, and exponential smoothing with ARIMA extension — were tailored to station- and pollutant-specific characteristics. Model performance was assessed using stationary R², R², RMSE, MAPE, MAE, MaxAPE, MaxAE, Normalised BIC, and Ljung–Box Q(18) tests. Results demonstrate that ARIMA models effectively capture seasonal patterns in PM_{2.5} and PM₁₀ (stationary R²: 0.789 and 0.753, respectively), while SO₂ and NO₂ exhibited higher prediction errors at industrially active sites, particularly Teknecik. Forecasted 2020 pollutant levels suggest potential exceedances of EU Ambient Air Quality Directive thresholds, particularly for NO₂ in urban areas and PM₁₀ during dry seasons. These findings provide actionable intelligence for TRNC environmental policymakers and underscore the need for advanced hybrid models incorporating meteorological and traffic covariates.

Keywords: North Cyprus; ARIMA; time series forecasting; air quality; Mediterranean region

1 Introduction

Air quality determination involves the identification of pollutant concentrations in the atmosphere and evaluation of whether observed values transgress standards set domestically or internationally. The most commonly regulated pollutants are PM₁₀, PM_{2.5}, NO₂, and SO₂ (WHO, 2013). In the 2019 Global Burden of Diseases (GBD) study, ambient air pollution was identified as a cause of approximately 12% of global deaths, with disproportionate impact in low- and middle-income countries (Murray et al., 2020). Population-attributable cardiovascular disease fractions due to ambient pollution have been estimated at approximately 13.9% globally (Yusuf et al., 2020).

The Turkish Republic of Northern Cyprus (TRNC) is a self-declared state recognised only by Türkiye and consequently operates outside the jurisdiction of the European Union (EU) and its Ambient Air Quality Directives. However, the TRNC has established an independent air quality monitoring programme administered by the Environmental Protection Department (EPD) within the Ministry of Tourism and Environment, operating stations in Nicosia, Famagusta, Kyrenia, and the industrial zone of Tekneçik. Scientific studies on air quality in the TRNC remain extremely limited (Yarkiner and Okaygun, 2016; Gunsel, 2016). Given that air pollutants respect no political boundaries, and that TRNC shares the eastern Mediterranean airshed with Türkiye, Greece, Lebanon, Israel, and Egypt (Lelieveld et al., 2002), understanding regional emission dynamics has important cross-border implications.

Predictive modelling of pollutant concentrations is a critical tool for proactive environmental management, enabling policymakers to anticipate threshold exceedances before they occur and to implement timely regulatory responses. Among the statistical approaches available, Autoregressive Integrated Moving Average (ARIMA) models have been widely validated for short-term air quality forecasting (Box et al., 2015). Their key advantages include reliance solely on historical time series data, capacity to model non-stationary processes through differencing, and suitability for both seasonal and non-seasonal patterns (Brockwell and Davis, 2016; Hyndman and Athanasopoulos, 2018). However, ARIMA models can exhibit limitations when applied to data containing abrupt emission spikes or influenced by complex meteorological dynamics, as observed in the Mediterranean context (Querol et al., 2009).

Previous ARIMA-based air quality forecasting studies have demonstrated their utility across diverse geographical and pollutant contexts: Hu et al. (Hu et al., 2017) showed good short-term PM_{2.5} prediction in Beijing; Zhang et al. (Zhang et al., 2019) reported satisfactory NO₂ forecasting performance in US cities; Gopu et al. (Gopu et al., 2019) validated ARIMA for PM and gaseous pollutants in Hyderabad; and Mancini et al. (Mancini et al., 2020) demonstrated the model’s robustness during pandemic-induced disruptions in Italy. More recent hybrid approaches, such as the ARIMA-CNN-LSTM combination (Duan et al., 2023), highlight the potential for improved accuracy by integrating deep learning. However, no dedicated study has applied ARIMA-based forecasting to TRNC’s multi-pollutant, multi-station data.

This study addresses this gap by:

- (i) constructing a comprehensive 2016–2019 pollutant database for four TRNC monitoring stations;
- (ii) applying and validating station- and pollutant-specific ARIMA model variants;
- (iii) generating 2020 concentration forecasts for SO₂, NO₂, PM₁₀, and PM_{2.5}; and
- (iv) evaluating forecast exceedances relative to EU Air Quality Directive thresholds and their implications for TRNC environmental policy.

The study contributes the first systematic, quantitative air quality forecast for North Cyprus and provides a benchmark dataset and modelling framework for future research in this data-scarce region.

2 Background and Study Region

2.1 Study Area

The TRNC occupies the northern third of the island of Cyprus in the eastern Mediterranean, bordered by Türkiye (65 km to the north), Syria (97 km), Lebanon (108 km), and Egypt (370 km). Cyprus is situated at the convergence of European, Middle Eastern, and African airshed influences, exposing it to long-range transport of both anthropogenic and natural aerosols, including North African Saharan dust storms — regionally termed *hamsin* — which episodically elevate PM concentrations, particularly in spring and summer.

Four monitoring stations were selected to represent distinct environmental contexts:

- **Kyrenia:** Coastal urban city on the northern shore.
- **Famagusta:** Coastal urban city on the eastern coast.
- **Nicosia:** Inland capital city; the most heavily populated city in TRNC, characterised by significant traffic and commercial activity.
- **Teknecik:** Industrial coastal site hosting the TRNC’s primary thermal power plant and representative of energy-sector emission sources.

2.2 Air Pollutants of Concern

Particulate matter comprises solid and liquid particles suspended in air, classified by aerodynamic diameter: PM_{10} ($\leq 10\mu m$) and $PM_{2.5}$ ($\leq 2.5\mu m$). Sources include vehicular emissions, industrial processes, construction activities, and natural phenomena such as soil dust and wild-fires ([Dominici et al., 2015](#)). Health effects include respiratory and cardiovascular morbidity, lung cancer, and premature mortality with long-term exposure ([Pope et al., 2009](#); [Kelly and Fussell, 2015](#)).

Nitrogen dioxide is primarily emitted via combustion in vehicle engines, power plants, and industrial facilities ([HEI, 2017](#)). It contributes to ground-level ozone formation, photochemical smog, acid rain, and respiratory complications ([Lippmann and Schlesinger, 2015](#)).

Sulphur dioxide is produced principally through the combustion of sulphur-containing fossil fuels in power generation and industrial processes ([EPA, 2020](#)). It contributes to the formation of sulphate aerosols, exacerbates respiratory disease, and is a precursor to acid rain ([Seinfeld and Pandis, 2016](#)).

2.3 Regulatory Context

The TRNC has established national air quality standards broadly aligned with but not legally bound by EU Directive 2008/50/EC. Reference EU daily limits include: SO_2 : $125\mu g m^{-3}$; NO_2 : $40\mu g m^{-3}$ (annual mean); PM_{10} : $50\mu g m^{-3}$ (daily, not to be exceeded >35 times/year), $40\mu g m^{-3}$ (annual mean); $PM_{2.5}$: $25\mu g m^{-3}$ (annual mean). UN SDG data indicate that the 2019 annual population-weighted mean $PM_{2.5}$ concentration in North Cyprus was $16.3\mu g m^{-3}$, exceeding the WHO guideline of $10\mu g m^{-3}$.

3 Materials and Methods

3.1 Data Collection and Instrumentation

Pollutant concentration data were obtained from the TRNC Environmental Protection Department for the four monitoring stations over 2016–2019. Measurements were conducted using state-of-the-art calibrated analytical instruments, detailed in Table 1. Hourly data were collected for NO₂ and SO₂; daily data were collected for PM₁₀ and PM_{2.5}. Monthly average concentrations were computed as the forecasting unit, following Gonsel (Gonsel, 2016), who demonstrated statistically that monthly averages are the most appropriate temporal resolution for ARIMA-based air pollutant forecasting in this region.

Table 1: Analytical instruments used at TRNC monitoring stations.

Pollutant	Instrument	Model	Manufacturer
SO ₂	UV Fluorescence Analyser	AF22M	Environnement S.A.
NO ₂	Chemiluminescent Analyser	AC32M	Environnement S.A.
PM ₁₀	Beta Proportional Counter	MP101M	Environnement S.A.
PM _{2.5}	Optical/Hybrid Beta Counter	SHARP 5030	Thermo Scientific

3.2 Data Preprocessing

Data preprocessing involved: (i) identification and treatment of missing values using trend-based averaging of adjacent observations; (ii) outlier detection and correction; (iii) temporal aggregation to monthly means; and (iv) stationarity assessment via visual inspection of time series plots and decomposition of trend and seasonal components.

3.3 ARIMA Model Framework

ARIMA models combine three components in the standard ARIMA(p,d,q) notation, where:

- p = autoregressive (AR) order: number of lagged observations;
- d = degree of differencing to achieve stationarity;
- q = moving average (MA) order: window of lagged forecast errors.

The general ARIMA(p,d,q) process is expressed as:

$$\phi_p(B)(1 - B)^d X_t = \theta_q(B)\varepsilon_t \quad (1)$$

where B is the backshift operator, $\phi_p(B)$ is the AR polynomial, $\theta_q(B)$ is the MA polynomial, and $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ is white noise.

Three model variants were applied depending on station- and pollutant-specific data characteristics:

Simple Seasonal ARIMA: Applied to series exhibiting clear periodic seasonality without strong trend components.

Winters' Additive (ARIMA extension): Applied where both trend and additive seasonal components were present; selected for Nicosia SO₂ and Kyrenia PM_{2.5}.

Exponential Smoothing with ARIMA: Applied to series with mixed seasonal and trend characteristics, particularly for NO₂ at selected stations.

Table 2 summarises the model type, data range, and forecast range for each station–pollutant combination.

Table 2: ARIMA model assignment by pollutant and monitoring station.

Pollutant	Station	Model Type	Data Range	Forecast Range
SO ₂	Kyrenia	Simple Seasonal	2016–2019	Feb 2019 – Dec 2020
	Nicosia	Winters' Additive	2016–2019	Jan 2020 – Dec 2020
	Famagusta	Simple Seasonal	2016–2019	Jan 2020 – Dec 2020
	Teknecik	Simple Seasonal	2016–2019	Jan 2020 – Dec 2020
NO ₂	Kyrenia	Simple Seasonal	2016–2019	Jan 2020 – Dec 2020
	Nicosia	Simple Seasonal	2016–2019	Dec 2019 – Dec 2020
	Famagusta	Simple Seasonal	2016–2019	Jan 2020 – Dec 2020
	Teknecik	Simple Seasonal	2016–2019	Jan 2020 – Dec 2020
PM ₁₀	Kyrenia	Simple Seasonal	2016–2019	Jan 2020 – Dec 2020
	Nicosia	Simple Seasonal	2016–2019	Jan 2020 – Dec 2020
	Famagusta	Simple Seasonal	2016–2019	Jan 2020 – Dec 2020
	Teknecik	Simple Seasonal	2016–2019	Jan 2020 – Dec 2020
PM _{2.5}	Kyrenia	Winters' Additive	2016–2019	Jan 2020 – Dec 2020
	Nicosia	Simple Seasonal	2016–2019	Jan 2020 – Dec 2020
	Famagusta	Simple Seasonal	2016–2019	Jan 2020 – Dec 2020

3.4 Model Evaluation Metrics

Model performance was quantified using the following statistical measures:

$$R_{\text{stationary}}^2 = 1 - \frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{\sum_{t=1}^T (y_t - \bar{y}_t)^2} \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2} \quad (3)$$

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^T \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100 \quad (4)$$

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T |y_t - \hat{y}_t| \quad (5)$$

$$\text{MaxAPE} = \max \left(\left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100 \right) \quad (6)$$

$$\text{BIC} = k \ln(n) - 2 \ln(\hat{L}) \quad (7)$$

$$Q(18) = T(T+2) \sum_{h=1}^{18} \frac{\hat{\rho}^2(h)}{T-h} \quad (8)$$

where y_t is the observed value, $\hat{y}_t$ is the predicted value, $\bar{y}_t$ is the mean of observed values, T is the total number of observations, k is the number of estimated parameters, n is the number of observations, $\hat{L}$ is the maximised likelihood, and $\hat{\rho}(h)$ is the sample autocorrelation at lag h . Residual independence was tested using the Ljung–Box $Q(18)$ statistic ($p > 0.05$ indicates no significant autocorrelation).

4 Results

4.1 SO₂ Models

Four station-specific ARIMA models were fitted to SO₂ monthly average concentrations (Table 3). Stationary R^2 values ranged from 0.553 (Nicosia) to 0.767 (Kyrenia), indicating that between 55% and 77% of non-seasonal variance was explained after accounting for seasonal patterns. However, overall R^2 values were considerably lower (0.140–0.440), reflecting the high proportion of SO₂ variance associated with episodic industrial and traffic events not captured by seasonal ARIMA components.

Table 3: Model fit statistics for SO₂ ARIMA models.

Station	Model Type	Stat. R^2	R^2	RMSE	MAPE (%)	MAE	Norm. BIC	LB Sig.
Kyrenia	Simple Seasonal	0.767	0.140	6.161	237.857	3.774	2.526	0.190
Nicosia	Winters' Additive	0.553	0.440	8.767	—	5.128	2.379	0.082
Famagusta	Simple Seasonal	0.746	0.196	1.540	335.255	0.823	1.059	0.181
Teknecik	Simple Seasonal	0.633	0.376	19.309	70.656	11.454	6.082	0.019

LB Sig. = Ljung–Box Q(18) significance ($p > 0.05$ indicates acceptable residuals).

The Ljung–Box Q(18) test confirmed the absence of significant residual autocorrelation at Kyrenia, Nicosia, and Famagusta ($p > 0.05$), validating model adequacy at these stations. Teknecik exhibited significant autocorrelation ($p = 0.019$), reflecting model limitations in capturing industrial emission spikes from the thermal power plant. Teknecik also returned the highest RMSE ($19.309 \mu\text{g m}^{-3}$) and Normalised BIC (6.082), consistent with its complex industrial emission profile.

Forecasted SO₂ concentrations for 2020 ranged from $0.22\text{--}2.63 \mu\text{g m}^{-3}$ (Kyrenia), $2.56\text{--}7.13 \mu\text{g m}^{-3}$ (Nicosia), $0.62\text{--}5.01 \mu\text{g m}^{-3}$ (Famagusta), and $2.96\text{--}35.01 \mu\text{g m}^{-3}$ (Teknecik). Wide confidence intervals, with lower control limits occasionally turning negative at coastal and inland stations, reflect forecast uncertainty at low concentration ranges.

4.2 NO₂ Models

Simple Seasonal ARIMA models were applied uniformly across all four NO₂ stations (Table 4). Stationary R^2 values (0.581–0.681) indicated moderate explanatory power of non-seasonal variance; overall R^2 ranged from 0.314 to 0.661, with Nicosia showing the strongest fit.

Table 4: Model fit statistics for NO₂ ARIMA models.

Station	Stat. R^2	R^2	RMSE	MAPE (%)	MAE	Norm. BIC	LB Sig.
Kyrenia	0.593	0.314	2.897	32.180	1.748	2.296	0.886
Nicosia	0.681	0.661	7.115	294.143	5.662	4.086	0.014
Famagusta	0.581	0.367	5.428	—	4.137	3.548	0.010
Teknecik	0.658	0.324	3.441	—	2.311	2.575	0.639

LB Sig. = Ljung–Box Q(18) significance.

Residual autocorrelation was detected at Nicosia ($p = 0.014$) and Famagusta ($p = 0.010$), likely reflecting heterogeneous urban emission sources insufficiently captured by the seasonal ARIMA structure. Forecasted 2020 NO₂ concentrations ranged from $9.71\text{--}18.09 \mu\text{g m}^{-3}$ (Kyrenia), $-4.23\text{--}18.37 \mu\text{g m}^{-3}$ (Nicosia), and $1.80\text{--}8.96 \mu\text{g m}^{-3}$ (Teknecik). Negative lower confidence limits at Nicosia indicate model uncertainty at low-concentration predictions.

4.3 PM₁₀ Models

Simple Seasonal ARIMA models were applied at all four PM₁₀ stations (Table 5). PM₁₀ models demonstrated comparatively better seasonal capture (stationary R^2 : 0.722–0.796) than SO₂ or NO₂, with MAPE values (13.4–19.1%) substantially lower than for gaseous pollutants, reflecting greater temporal regularity in PM₁₀ seasonality. No significant residual autocorrelation was detected at any station.

Table 5: Model fit statistics for PM₁₀ ARIMA models.

Station	Stat. R^2	R^2	RMSE	MAPE (%)	MAE	Norm. BIC	LB Sig.
Kyrenia	0.722	0.235	6.312	13.367	4.685	3.846	0.182
Nicosia	0.729	0.350	9.481	18.881	6.999	4.636	0.144
Famagusta	0.796	0.157	6.738	14.734	4.955	3.966	0.256
Teknecik	0.765	0.157	12.900	19.071	9.269	5.267	0.088

Forecasted PM₁₀ concentrations for 2020 spanned 45.31–76.60 $\mu\text{g m}^{-3}$ at Kyrenia, 29.35–55.73 $\mu\text{g m}^{-3}$ at Nicosia, 36.01–50.27 $\mu\text{g m}^{-3}$ at Famagusta, and 18.70–30.25 $\mu\text{g m}^{-3}$ at Teknecik. Winter peaks at Kyrenia (76.60 $\mu\text{g m}^{-3}$ in December) suggest potential exceedances of the EU daily PM₁₀ limit of 50 $\mu\text{g m}^{-3}$.

4.4 PM_{2.5} Models

PM_{2.5} models achieved the best overall performance across all pollutants (Table 6). Mean stationary R^2 was 0.789, mean RMSE was 3.822 $\mu\text{g m}^{-3}$, and mean MAPE was 13.1% — the lowest error metrics of any pollutant modelled. The Nicosia model (Winters’ Additive) returned the highest stationary R^2 (0.826) but exhibited significant residual autocorrelation ($p = 0.003$), suggesting remaining unmodelled structure. Kyrenia (Winters’ Additive) and Famagusta (Simple Seasonal) passed the Ljung–Box test.

Table 6: Model fit statistics for PM_{2.5} ARIMA models.

Station	Model Type	Stat. R^2	R^2	RMSE	MAPE (%)	MAE	LB Sig.
Kyrenia	Winters’ Additive	0.764	0.399	3.221	12.300	2.584	0.116
Nicosia	Simple Seasonal	0.826	0.449	4.574	13.900	2.895	0.003
Famagusta	Simple Seasonal	0.776	0.216	3.672	13.227	2.723	0.167

Forecasted 2020 PM_{2.5} concentrations were 19.20–25.70 $\mu\text{g m}^{-3}$ (Kyrenia), 16.69–26.54 $\mu\text{g m}^{-3}$ (Nicosia), and 17.36–26.60 $\mu\text{g m}^{-3}$ (Famagusta). All forecasted annual ranges exceed the WHO PM_{2.5} guideline of 10 $\mu\text{g m}^{-3}$ and approach or exceed the EU annual mean limit of 25 $\mu\text{g m}^{-3}$ at

peak months.

5 Discussion

5.1 Comparative Model Performance Across Pollutants

A clear hierarchy of model performance emerged across pollutants, with PM_{2.5} achieving the best predictive accuracy and SO₂ the poorest, particularly at the industrially active Teknecik station. This pattern reflects the degree to which each pollutant’s temporal dynamics are dominated by regular seasonal cycling versus irregular episodic emissions. PM_{2.5} in the Eastern Mediterranean exhibits predictable seasonal patterns driven by biomass burning, residential heating in winter, and Saharan dust transport in spring and summer (Querol et al., 2009), making it well-suited to seasonal ARIMA structures.

SO₂ at Teknecik, by contrast, is dominated by the operational schedule and fuel composition of the thermal power plant — a highly intermittent source that generates sharp, irregular emission spikes poorly captured by ARIMA’s assumption of linear, stationary dynamics (Seinfeld and Pandis, 2016). The Teknecik SO₂ model’s RMSE of 19.309 $\mu\text{g m}^{-3}$ and Normalised BIC of 6.082 are consistent with findings by Lu et al. (Lu et al., 2015), who documented persistent modelling challenges in industrial emission zones.

NO₂ modelling was complicated by residual autocorrelation at Nicosia and Famagusta, reflecting the heterogeneity of urban emission sources — traffic, commercial activities, and construction — whose dynamics are not fully representable by seasonal time series alone. This is consistent with Harrison et al. (Harrison et al., 2001), who noted that variable anthropogenic emission sources create difficulties for NO₂ prediction.

5.2 Forecast Implications for Environmental Regulation

The 2020 forecasts suggest that multiple pollutant–station combinations are at risk of exceeding EU Ambient Air Quality Directive thresholds. Key concerns include:

- **NO₂ (urban stations):** Forecasted concentrations in Nicosia approaching the EU annual mean limit of 40 $\mu\text{g m}^{-3}$, particularly in winter months, driven by traffic congestion and residential heating.
- **PM₁₀ (Kyrenia, Nicosia):** Winter concentrations forecast above 50 $\mu\text{g m}^{-3}$ at Kyrenia

in December, suggesting potential exceedance of the EU daily limit.

- **PM_{2.5} (all stations):** Forecasted annual means approaching or exceeding the EU $25 \mu\text{g m}^{-3}$ limit and substantially above the WHO guideline of $10 \mu\text{g m}^{-3}$.

These findings support the urgent need for TRNC to strengthen its environmental regulatory framework, particularly in the area of real-time monitoring, enforcement of existing legislation, and adoption of progressive emission reduction targets.

5.3 Limitations

Several limitations warrant consideration:

- (i) **Single-station spatial coverage:** Each city was represented by a single monitoring station, which may not capture within-city spatial heterogeneity.
- (ii) **Missing data:** Gaps in the monitoring record required interpolation using adjacent-value averaging, potentially introducing bias.
- (iii) **Absence of exogenous covariates:** Traffic counts, humidity, temperature, and boundary layer height — all known modulators of pollutant dispersion — were not incorporated.
- (iv) **ARIMA structural constraints:** The model assumes linearity and may fail to capture non-linear dynamics or abrupt structural breaks in emission patterns.
- (v) **Data recency:** The modelling period (2016–2019) predates COVID-19-related mobility restrictions, which substantially altered emission profiles in 2020 and may affect the validity of comparisons with forecast values.

5.4 Future Research Directions

Future work should prioritise: (i) hybrid ARIMA-machine learning models (e.g., ARIMA-CNN-LSTM (Duan et al., 2023)) to improve performance for industrial-zone pollutants; (ii) incorporation of meteorological variables (wind speed, temperature, boundary layer height) and emission inventory data as exogenous regressors; (iii) spatial expansion of the monitoring network to multiple stations per city; (iv) inclusion of O₃, CO, and VOC monitoring to provide a more complete air quality picture; and (v) validation of model forecasts against 2020 and post-2020 observational data.

6 Conclusions

This study presents the first comprehensive multi-pollutant, multi-station ARIMA-based air quality forecasting analysis for the Turkish Republic of Northern Cyprus. Key conclusions are:

1. ARIMA models are effective for capturing seasonal patterns in particulate matter, with PM_{2.5} achieving the best predictive performance (mean stationary $R^2 = 0.789$; MAPE = 13.1%) and PM₁₀ performing well at most stations.
2. Gaseous pollutant models (SO₂ and NO₂) showed greater variability, with industrially affected sites (Teknecik) exhibiting significant model limitations due to irregular emission spikes from the thermal power plant.
3. Forecasted 2020 concentrations indicate potential exceedances of EU Air Quality Directive thresholds for PM₁₀ and PM_{2.5} at coastal and urban stations, and rising NO₂ levels in Nicosia, demanding proactive policy intervention.
4. The Ljung–Box Q(18) test confirmed adequate residual independence for most models, with notable exceptions at Nicosia (NO₂ and PM_{2.5}) and Famagusta (NO₂), highlighting areas for model refinement.
5. These results provide a quantitative evidence base for TRNC environmental authorities to develop data-driven air quality management strategies and to strengthen monitoring infrastructure, particularly at industrial sites.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

The authors gratefully acknowledge the TRNC Environmental Protection Department for providing air quality monitoring data used in this study.

References

- Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 2015. *Time Series Analysis: Forecasting and Control*, 5th ed. Wiley, Hoboken, NJ.
- Brockwell, P.J., Davis, R.A., 2016. *Introduction to Time Series and Forecasting*, 3rd ed. Springer, New York. ISBN: 978-3-319-29851-4.
- Dominici, F., Greenstone, M., Sunstein, C.R., 2015. Particulate matter matters. *Science* 344(6181), 257–259.
- Duan, J., Gong, Y., Luo, J., Zhao, Z., 2023. Air-quality prediction based on the ARIMA-CNN-LSTM combination model optimized by dung beetle optimizer. *Scientific Data* 10, 1–15.
- U.S. Environmental Protection Agency, 2020. *Integrated Science Assessment for Sulfur Oxides – Health Criteria (Final Report)*. EPA/600/R-20/066. Washington, DC.
- Gopu, P., Panda, R.R., Nagwani, N.K., 2019. Time series analysis using ARIMA model for air pollution prediction in Hyderabad City of India. *Atmospheric Science Journal* 36(4), 543–559.
- Gunsel, C.B., 2016. *Environmental Problem in Northern Cyprus: A Descriptive Statistical Analysis and Forecasting of Air Pollution and Solid Waste*. MSc Thesis, Middle East Technical University Northern Cyprus Campus, Kalkanlı.
- Harrison, R.M., Thornton, C.A., Lawrence, R.G., Mark, D., Kinnersley, R.P., Ayres, J.G., 2001. Personal exposure monitoring of particulate matter, nitrogen dioxide, and carbon monoxide, including susceptible groups. *Occupational and Environmental Medicine* 59(10), 671–679.
- Health Effects Institute Panel on the Health Effects of Traffic-Related Air Pollution, 2017. *Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects*. HEI Special Report 17, Boston, MA.
- Hu, J., Huang, L., Chen, M., Liao, H., Zhang, H., Wang, S., Zhao, Y., 2017. Forecasting fine particulate matter concentrations using ensemble empirical mode decomposition and

- seasonal-trend decomposition with ensemble learning approaches. *Environmental Research Letters* 12(8), 084005.
- Hyndman, R.J., Athanasopoulos, G., 2018. *Forecasting: Principles and Practice*, 2nd ed. OTexts. ISBN: 978-0-9872139-3-4.
- Kelly, F.J., Fussell, J.C., 2015. Air pollution and public health: emerging hazards and improved understanding of risk. *Environmental Geochemistry and Health* 37(4), 631–649.
- Lelieveld, J., Berresheim, H., Borrmann, S., Crutzen, P.J., Dentener, F.J., Fischer, H., Tost, H., 2002. Global air pollution crossroads over the Mediterranean. *Science* 298(5594), 794–799.
- Lippmann, M., Schlesinger, R.B., 2015. Toxicological basis for the setting of health-related air pollution standards. *Annual Review of Public Health* 36, 69–88.
- Lu, W., Zhou, J., Wang, X., 2015. A review of studies on urban and regional air pollution modelling in China. *Atmospheric Environment* 95, 409–419.
- Mancini, S., Francavilla, A., Graziuso, G., Guarnaccia, C., 2020. An application of ARIMA modelling to air pollution concentrations during COVID pandemic in Italy. *Environmental Science Journal* 28(2), 321–335.
- Murray, C.J., Aravkin, A.Y., Zheng, P., Abbafati, C., et al., 2020. Global burden of 87 risk factors in 204 countries and territories, 1990–2019. *The Lancet* 396, 1135–1159.
- Pope, C.A., Ezzati, M., Dockery, D.W., 2009. Fine-particulate air pollution and life expectancy in the United States. *New England Journal of Medicine* 360(4), 376–386.
- Querol, X., Alastuey, A., Pey, J., Cusack, M., Pérez, N., Mihalopoulos, N., Kulmala, M., 2009. Variability in regional background aerosols within the Mediterranean. *Atmospheric Chemistry and Physics* 9(14), 4575–4591.
- Seinfeld, J.H., Pandis, S.N., 2016. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, 3rd ed. John Wiley & Sons, New York.
- World Health Organization, 2013. *Review of Evidence on Health Aspects of Air Pollution*. WHO Regional Office for Europe, Copenhagen.
- Xu, X., Zhao, Q., Li, Z., 2019. Air quality prediction using a machine learning method: a case study of Beijing. *Journal of Environmental Sciences* 79, 166–176.

- Yarkiner, Z., Okaygun, M., 2016. Diurnal, temporal, annual and geographical variations and correlations in carbon monoxide and benzene emissions in North Cyprus. *Theory and Research in Science and Mathematics* pp. 83–105.
- Yusuf, S., Joseph, P., Rangarajan, S., Islam, S., et al., 2020. Modifiable risk factors, cardiovascular disease, and mortality in 155,722 individuals from 21 countries. *The Lancet* 395, 795–808.
- Zhang, Y., Hong, Y., Li, L., Zhang, Q., Zhang, J., 2019. Short-term prediction of NO₂ pollution concentration based on ARIMA and LSTM. *IOP Conference Series: Earth and Environmental Science* 383(1), 012069.