Annals of Agri-Bio Research **27** (1) : 35-41, 2022

**Impact Score : 0.28 (Scopus)**

# GIS and Time Series Modelling Approach to Predict Tropospheric Nitrogen Dioxide

CHADETRIK ROUT, GAURAV SHUKLA, VIKAS BENIWAL[1], SURENDRA PAL SINGH[2] AND RAHUL GROVER*

*Department of Civil Engineering, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala-133 207 (Haryana), India*
**(e-mail : rahulgrover@mmumullana.org; Mobile : 96710 20303)*

(Received : January 4, 2022; Accepted : March, 10, 2022)

**ABSTRACT**

Time series is a time-oriented or chronological sequence of observations on a variable of interest. Auto-Regressive Integrated Moving Average (ARIMA) model approach was used in this study for time series analysis of $NO_2$ concentration in Punjab region, India. Kriging Spatial Interpolation method was also used. This study integrated the satellite observed data with statistical methods. The predicted $NO_2$ concentration was used for spatial distribution and estimation of $NO_2$. OMI satellite data for tropospheric $NO_2$ from the year 2012 to 2019 were used to make a forecast of $NO_2$ concentrations for the year 2020. The $R^2$ value showed good agreement between the observed and predicted concentrations of $NO_2$ in both the approaches.

**Key words :** Tropospheric $NO_2$, ARIMA, Kriging interpolation, time series modelling, spatial distribution

## INTRODUCTION

The Earth's atmosphere is a mixture of various gases, which consists of ~78.08% $N_2$, 20.95% $O_2$, 0.93% Ar, 0.04% $CO_2$ and the rest other trace gases. The layer of greatest interest from the point of pollution is the troposphere in which most living things exist. One of the air pollutants in the troposphere is nitrogen dioxide ($NO_2$), which participates in a catalytic chain reaction producing ozone. This gas plays a key role in tropospheric chemistry with important implications for air quality and climate change. National Ambient Air Quality Standards (NAAQS) by Central Pollution Control Board (CPCB) of India has set its permissible limit as 40 μg/$m^3$ in ambient air. The CPCB also executes a nationwide programme of ambient air quality monitoring known as National Air Quality Monitoring Programme (NAMP). The ground monitoring network of NAMP has 342 operating stations covering major cities/towns, but still India's capacity to monitor and assess the problem of air pollution remains abysmally weak. $NO_2$ is one of the pollutants measured under NAMP. Ground stations provide information at the sampled locations but leave gaps in our understanding of air quality in non-monitored areas.

Satellite observations of tropospheric $NO_2$ are useful to address some of these issues because of their good spatial coverage and long-term measurements over the entire Indian domain (Singh *et al.*, 2018).The strong absorption lines of the $NO_2$ molecule in the visible wavelength range of the spectrum facilitate the use of optical absorption spectroscopy for measuring atmospheric $NO_2$ abundance (Sur *et al.*, 2017). Nadir viewing instruments have been deployed on satellite platforms since the mid-1990s. This has resulted in the global monitoring of $NO_2$ concentrations under consistent measurement conditions. The Ozone Monitoring Instrument (OMI) measures atmospheric nitrogen dioxide for various time zones around the world (Goldberg *et al.*, 2017). The OMI on board NASA's Earth Observing satellite–Aura is a wide-field imaging grating spectrometer, its horizontal resolution is 13 × 24 km at the nadir point, but gets considerably larger with higher viewing angles. The satellite

[1]Department of Biotechnology, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala-133 207 (Haryana), India.
[2]Surveying Engineering Department, Wollega University, Nekemte City, Post box-395 (Oromia), Ethiopia.

performs spectral measurements in the range of 270-500 nm at a spectral resolution of 0.63 nm. This instrument provides a global coverage of spectral measurements.

In this study, a time series model was formulated for forecasting and geo-statistical technique was used for interpolating the spatial concentration distribution of tropospheric $NO_2$ concentration in the region of the Indian state Punjab. Auto-Regressive Integrated Moving Average (ARIMA) model approach was used in this study for time series analysis and forecasting of $NO_2$. Comparison of predicted values of $NO_2$ with respect to the observed values was analyzed using R-squared statistics.

Detailed inventory reveals that the various emission sources of $NO_2$ in India belong to industrial, transportation and biomass burning sectors (Saikia *et al.*, 2019). The usage of diesel and petrol, dense population, heavy biomass fuel, heavy traffic usage was the major source of $NO_2$ (Ileperuma, 2020). Historically, air quality monitoring was done through ground-based measurements. Modified Jacobs and Hochheiser Method was used for the measurement of nitrogen dioxide in ground-based monitoring. Satellite-based observations were more suitable for monitoring and analyzing environmental parameters due to their good spatial and temporal coverage (El-Khoury *et al.*, 2021).

Various studies showed good capability of satellite observation for $NO_2$ concentration. Observations of tropospheric $NO_2$ measurements from space using satellite observations began in 1995. Global Ozone Monitoring Experiment (GOME) instrument on the Second European Remote Sensing Satellite (ERS-2) launched in April, 1995 allowed the retrieval of Vertical Column Density (VCD) of $NO_2$ on a global scale. After GOME, similar instruments such as SCanning Imaging Absorption SpectroMeter for Atmospheric CartograpHY(SCIAMACHY), Ozone Monitoring Instrument (OMI) and GOME2 TROPOspheric Monitoring Instrument (TROPOMI) were launched. The tropospheric $NO_2$ VCD maps derived from these instruments were used to study many scientific applications, pollution emissions and pollutant distribution. The OMI column measurements were useful proxy for ground-level variability in $NO_2$ concentrations (Li and Wu, 2021). The mixed effect model (MEM) was developed to predict tropospheric $NO_2$ in China region (Chi *et al.*, 2021). This study used seven years OMI tropospheric $NO_2$ data with other topographical and climate parameters to develop MEM. Results showed the effectiveness of the time series based developed model and OMI sensor for $NO_2$ measurements. Machine learning methods, such as Random Forest, Support Vector machine, Gradient boosting, and linear regression were also found suitable for the prediction of $NO_2$ using OMI retrieved tropospheric $NO_2$ data. Zhang *et al.* (2021) used random forest method to develop a prediction model using OMI $NO_2$ data. Other satellite data, such as TROPOspheric Monitoring Instrument (TROPOMI), data of Sentinel-5 Precursor (S5P) were also a good source for monitoring various atmospheric constituents (such as $NO_2$, $O_3$, $SO_2$, CO, $CH_4$, etc.). These data also provided high-resolution spatial coverage for monitoring, assessment and prediction approaches. Kang *et al.* (2021) compared several machine learning methods (Random Forest, Support Vector Machine, gradient boosting, regression etc.) to assess the ability of the model to predict $NO_2$ and $O_3$ using TROPOMI data. Comparative study with the SCIAMACHY and OMI data indicated that OMI tropospheric $NO_2$ columns revealed a clearer picture because of the higher resolution.

Besides machine learning methods, time series analysis was also used by various studies for assessing and understanding the environmental parameters and to develop a prediction model (Akdi *et al.*, 2021). Studies by Akdi *et al.* (2021) used ARIMA and regression model to analyze the 20 years trend of particulate matter ($PM_{2.5}$) in Paris to develop a prediction model which showed that both models were suitable for forecasting. High-resolution spatial distribution of $NO_2$, NO and $O_3$ was observed in Lebanon region (El-Khoury *et al.*, 2021). This study used a land use regression model for analysis and prediction. Zhang *et al.* (2021) also used a land use regression model with dispersion modelling approach to predict $NO_2$ concentration at high spatial resolution for New York city. This study used Research LINE source (R-LINE) model predicting $NO_2$ with other climate and topographical parameters to develop a new regression model. This hybrid approach suggested a good ability to predict $NO_2$

concentration. Multilinear Regression Model (MLR), Autoregressive moving average (ARMA) and ARIMA methods were widely applied to model time series. Based on the literature in our study, OMI sensor was used to retrieve $NO_2$ measurements and a time series model was formulated using ARIMA approach to predict $NO_2$.

**METHODOLOGY**

ARIMA model approach was used in this study for time series analysis of $NO_2$. This approach consisted of three phases :

- Identification, in which the characteristics and statistics of a time series were examined and attempts were made to relate them to those of specific models.
- Estimation, in which the parameters of the tentatively identified model(s) were estimated using the data at hand; and
- Diagnostic checking, in which the estimated model(s), and residuals of the fitted model(s) were examined to see if the model(s) make sense and were consonant with our assumptions.

The time series ARIMA model is referred as Box-Jenkins model. In time series analysis stationary stochastic process is explained by its mean, variance and autocorrelation function. In $Y_t$, where t = 1 to t, is a time series. The use of Box-Jenkins backshift operator (B), the simplified autoregressive moving average model ARMA (*p,q*) is expressed as Equation 1.

$$\phi_p (B)\, Y_t = \theta_q (B)\, a_t \qquad ...(1)$$

Where, $\emptyset_p(B) = (1 - \sum_{i=1}^{p} \emptyset_i B^i)$ and $\theta_q(B) = (1 - \sum_{j=1}^{q} \theta_j B^j)$

Where, $\emptyset_p$ and $\theta_q$($p$ = 1 to $p$ and $q$ = 1 to $q$) were the autoregressive and moving average parameters, and $a_t$ was the residuals.

If a time series was non-stationary stochastic process, then it was represented by differences and forms the ARIMA (*p, d, q*) model and expressed as Equation 2.

$$\phi_p (B)\, (1 - B)^d Y_t = \phi_p (B)\, a_t \qquad ...(2)$$

Where, d was the order of differencing.

Spatial interpolation was the procedure of estimating the value of a desired parameter ($NO_2$ in this study) at various non-monitored locations within the area covered by existing observations. The spatial interpolation method used in this study was Kriging (Ikechukwu *et al.*, 2017). Kriging utilized the variogram model to create best linear unbiased values at each location. The semi-variogram was used to demonstrate the spatial correlation structure in the data.

In Kriging, one must model the spatial autocorrelation used a semi-variogram instead of assuming a direct linear relationship with the separation distance. The semi variogram captured the spatial dependence between sample locations by plotting semi-variance against the separation distance. The general expression to estimate the semi-variance $\gamma(h)$ defined over the observed data at a distance '$h$' given as Equation 3.

$$\gamma(h) = \frac{1}{2n(h)}\left(\sum_{i=1}^{n(h)} [M_i - M_j]^2\right) \qquad ...(3)$$

Where, $M_i$ and $M_j$ were the measured concentration of tropospheric $NO_2$ at two locations lagged by the distance '$h$' and n($h$) is the number of pairs of sample locations with distance '$h$'.

As the distance between sample locations increased, the semi-variance was expected to increase because near samples were more similar than distant samples. However, the experimental semi-variogram values did not approach to "0" at the origin. This was due to residuals and was known as Nugget. The semi-variogram values were sill and range was estimated by a keen observation of the model. $NO_2$ data were retrieved for Punjab region from *http://disc.gsfc.nasa.gov* from year 2012 to 2020. The data available were from NASA Aura satellite's Ozone Monitoring Instrument. Initially, trend analysis was conducted for temporal variations of $NO_2$ from the year 2012 to 2019. A time series model was formulated using ARIMA approach with the help of STATISTICA software. Using the model as above, forecasting was done for the year 2020. R squared statistics were used to compare observed and forecast values of $NO_2$. Using the forecast value for a certain year, geo-statistical techniques were applied to obtain the spatial distributions.

*(i) Geo-statistical analysis :* An experimental semi-variogram was generated. Diagnostics were performed and an appropriate semi-variogram model and interpolation method was selected. Spatial distribution of $NO_2$ was obtained using Kriging interpolation method with the help of ILWIS software. Comparison of predicted values of $NO_2$ with respect to the observed values was done with the help of R-squared statistics.

Generally Kriging worked on a selected hypothetical semi-variogram models used for calculating semi variance esteems at the places where tropospheric $NO_2$ was assessed. A few hypothetical semi-variogram models were conceivable, such as linear, spherical, exponential, Gaussian, etc. The most appropriate semi-variogram model might be found depending on the RMS error between the semi-variance values acquired from the experimentally observed semi-variance and the theoretical model predicted semi-variance values.

In this study, spherical, exponential and Gaussian semi-variogram models were tried over the analytically developed semi-variogram and the spherical model was found to give the least RMS error for the $NO_2$ data under study. Thus, ordinary Kriging spatial interpolation was implemented using the following steps :

Out of the total 85 selected coordinate points of Punjab region, 20 points were reserved for data validation and the rest 65 points were used as input data for performing spatial interpolation using Kriging method. The value of tropospheric $NO_2$ was estimated using time series modelling at these 65 points were fed as input to ILWIS to interpolate $NO_2$ concentration at 20 validation points.

*(ii) Conducting semi-variogram analysis :* This investigation normally comprised first producing an experimental variogram by estimating the Nugget, Sill and Range from the spatial correlations obtained. Then, a parametric model that satisfactorily caught the design of the observational variogram was fitted. This was done by obtaining semi-variograms for Gaussian, spherical and exponential models and the model with the least RMS error was selected as the best fit (Fig. 1).

*(iii) Spatial prediction using ordinary Kriging :* The evaluated parameters of the variogram were fed as input for the Kriging spatial forecast from

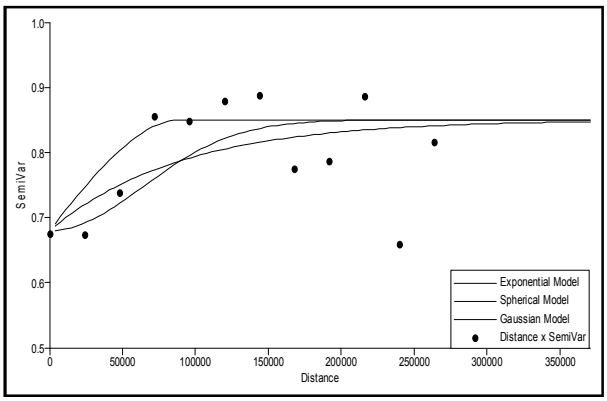


Fig. 1. Experimental semi-variogram plot.

which a spatially interpolated surface was created by ILWIS.

*(iv) Cross validation :* The resultant interpolated values were cross-validated and the relational coefficient ($R^2$ value) was analyzed between the interpolated value and the satellite observed data (Equation 4).

$$R^2 = 1 - \frac{\text{Unexplained Variation}}{\text{Total Variation}} \quad ...(4)$$

## RESULTS AND DISCUSSION

The methodology described above was applied to forecast the monthly tropospheric $NO_2$ concentrations for the year 2020 at 85 locations. The forecasted values of $NO_2$ concentration were obtained for all locations. Sixty-five values were taken to interpolate $NO_2$, and the remaining 20 locations were used for validation purpose.

First of all, time series of satellite observed monthly data from 2012 to 2019 for tropospheric $NO_2$ were plotted for all 85 locations. The trend observed in the time series was removed by differencing and logarithmic transformations. The order of differencing (*d*), the auto-regressive component (*p*) and the moving average component (*q*) were determined i. e. p = 0, d = 1 and q = 1. ARIMA (0, 1, 1) model was found to fit the Punjab region data. Thus, ARIMA (0, 1, 1) model was used to forecast the values of tropospheric $NO_2$ at all 85 locations for the year 2020. The forecast values and the satellite observed values of $NO_2$ concentration have been shown in Fig. 2.

For the year 2020, the co-efficient of determination ($R^2$) was computed to compare the average forecast values of $NO_2$ with the

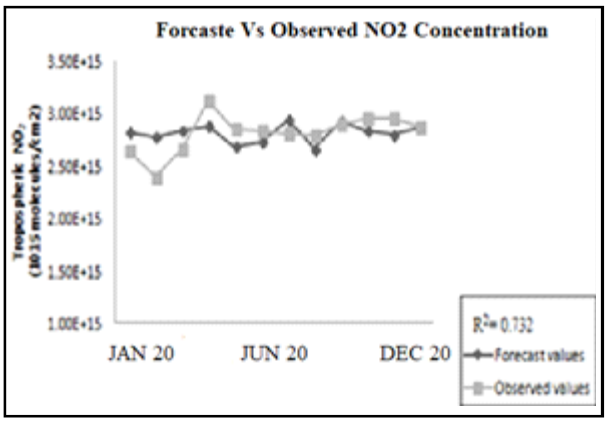


Fig. 2. The satellite observed $NO_2$ concentration and forecast $NO_2$ (molecules/$cm^2$) concentration by time series analysis.

satellite observed data i.e. $R^2 = 0.732$, indicating good fit with the model. Forecast for all 85 locations was carried out for each month from January 2020 to December 2020.

The 20 validation locations which were used for comparison of the satellite observed $NO_2$ concentration and the predicted $NO_2$ concentrations were marked as triangles (Fig. 3). The solid dots represented the locations used for spatial interpolation.

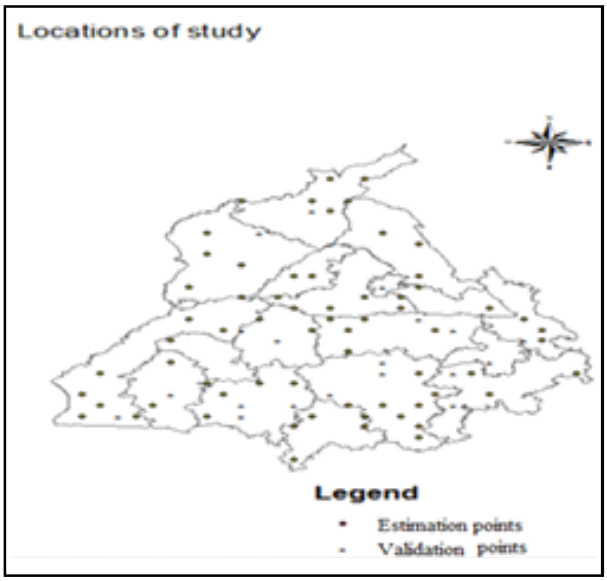


Fig. 3. The 85 locations of Punjab region used in this study.

The $R^2$ values lied in the range of 0.695 and 0.803 indicating that ARIMA (0, 1, 1) model satisfactorily predicted the monthly $NO_2$ concentrations over the Punjab region (Table 1). The solid dots in Fig. 3 show the 65 locations that were used in spatial interpolation and triangles showing the 20 validation locations.

**Table 1.** $R^2$ statistics obtained on comparing the ARIMA (0, 1, 1) forecast $NO_2$ concentration with the satellite observed $NO_2$ concentration for 2020

| Months | $R^2$ |
|---|---|
| January | 0.746 |
| February | 0.780 |
| March | 0.722 |
| April | 0.803 |
| May | 0.754 |
| June | 0.695 |
| July | 0.726 |
| August | 0.714 |
| September | 0.772 |
| October | 0.726 |
| November | 0.758 |
| December | 0.734 |

Experimental semi-variance ($\gamma_{(h)}$) values for each month were computed using equation 1 with the help of ILWIS software (Meena *et al.*, 2019). From these values, Nugget, Sill and Range were estimated for each month. The semi-variance values were computed for Spherical, Gaussian and Exponential models using ILWIS. Root mean square error (RMSE) was computed for these three models for January, February and April 2020. For these three months, all three models were plotted over the semi-variance values obtained from the experimental semi-variogram on the same graph (Fig. 4).

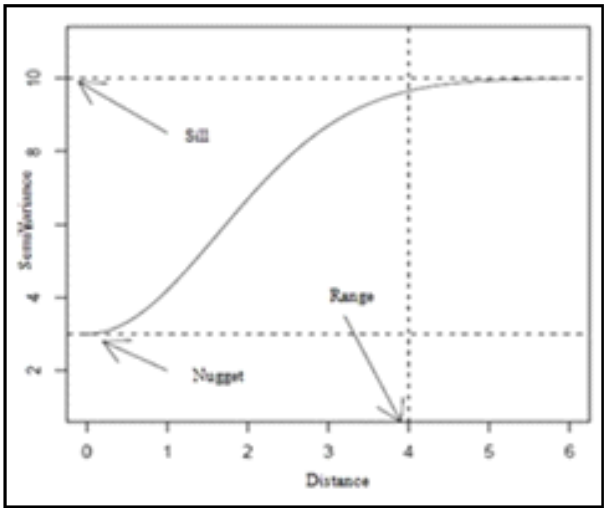


Fig. 4. Variogram model.

Ordinary Kriging method of interpolation was used to obtain the spatial distribution map of Punjab region for each month of 2020. The highest estimates of $NO_2$ concentration over Punjab region in the year 2020 were observed for October (Fig. 5). Fig. 4 shows the spatial distribution map for October 2020, the areas

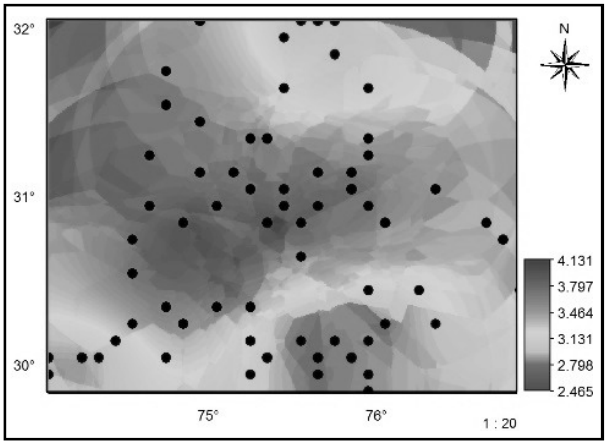


Fig. 5. Spatial distribution of $NO_2$ concentration (1015 molecules/$cm^2$) for October 2020.

having the highest $NO_2$ concentration ranged from 3.79 x 1015 to 4.13 x 1015 molecules/ $cm^2$ were depicted in magenta were among the areas having $NO_2$ concentration in this range. Ferozpur, Mukstar and Faridkot were/are agricultural areas. More water applications in the agricultural fields activated microbial activity contributing more soil emissions which was the main reason for higher levels of $NO_2$. It was further increased due to the use of fertilizers in agricultural fields. Majority of farmers practised crop residue burning to save time. Substantial amount of $NO_2$ was released into the atmosphere from large-scale crop residue burning in wheat-rice crop rotation. Ludhiana and Jalandhar were the industrial regions of Punjab. High values of $NO_2$ in these areas might be attributed to the usage of diesel and petrol, dense population, heavy biomass fuel and heavy traffic usage.

For each month i. e. from January to December 2020 in the validation locations, comparison between the satellite observed $NO_2$ concentration and the Kriging interpolated $NO_2$ concentration was done. The co-efficient of determination $R^2$ was used to show the goodness of fit between the interpolated values of tropospheric $NO_2$ from Kriging and the satellite observed data. The $R^2$ results ranged between 0.631 to 0.760 indicating that the estimated values of $NO_2$ concentration at the validation locations were adequately fitted (Table 2).

**Table 2.** $R^2$ values obtained on comparing the Kriging estimated value with the satellite observed $NO_2$ value for 2020

| Months | $R^2$ |
|---|---|
| January | 0.712 |
| February | 0.754 |
| March | 0.684 |
| April | 0.760 |
| May | 0.726 |
| June | 0.631 |
| July | 0.675 |
| August | 0.669 |
| September | 0.714 |
| October | 0.685 |
| November | 0.706 |
| December | 0.713 |

## CONCLUSION

The estimation methods of time series modelling and spatial interpolation were quite robust and easy to apply, and the integration of their outputs was straight forward. Time series forecasts of $NO_2$ concentration values were compared with satellite observed data; the $R^2$ 0.695 and 0.803. Likewise, in spatial interpolation $R^2$ values ranged in between 0.631 to 0.76. Major hotspots identified in this study lied in Ludhiana, Mukstar, Jalandhar and Ferozpur areas.

The $R^2$ values obtained in this study were adequate for the estimation of tropospheric $NO_2$ concentration. The procedures followed in this study should be capable of assisting future predictions as well as relying on them for estimating the $NO_2$ concentrations at sites where ground monitoring was not done. In particular, such an approach allowed to carry out a good estimate of $NO_2$, the analyzed pollutant, and to reconstruct the eventual missing data in the time and space domain. Thus, the models used in this study could be used in future studies for the estimation of tropospheric $NO_2$ concentration over any region.

## REFERENCES

Akdi, Y., Gölveren, E., Ünlü, K. D. and Yücel, M. E. (2021). Modelling and forecasting of monthly $PM_{2.5}$ emission of Paris by periodogram-based time series methodology. *Environ. Monit. Assess.* **193** : 1-15.

Chi, Y., Fan, M., Zhao, C., Sun, L., Yang, Y., Yang, X. and Tao, J. (2021). Ground-level $NO_2$ concentration estimation based on OMI tropospheric $NO_2$ and its spatiotemporal characteristics in typical regions of China. *Atmos. Res.* **264** : 2411-2502.

El-Khoury, C., Alameddine, I., Zalzal, J., El-Fadel,

M. and Hatzopoulou, M. (2021). Assessing the intra-urban variability of nitrogen oxides and ozone across a highly heterogeneous urban area. *Environ. Monit. Assess.* **193** : 1-22.

Goldberg, D. L., Lamsal, L. N., Loughner, C. P., Swartz, W. H., Lu, Z. and Streets, D. G. (2017). A high-resolution and observationally constrained OMI $NO_2$ satellite retrieval. *Atmos. Chem. Phys.* **17** : 11403-11421.

Ikechukwu, M. N., Ebinne, E., Idorenyin, U. and Raphael, N. I. (2017). Accuracy assessment and comparative analysis of IDW, spline and kriging in spatial interpolation of landform (Topography) : An experimental study. *J. Geogr. Inf. Syst.* **9** : 354-371.

Ileperuma, O. A. (2020). Review of air pollution studies in Sri Lanka. *Ceylon J. Sci.* **49** : 225-238.

Kang, Y., Choi, H., Im, J., Park, S., Shin, M., Song, C. K. and Kim, S. (2021). Estimation of surface-level $NO_2$ and $O_3$ concentrations using TROPOMI data and machine learning over East Asia. *Environ. Pollut.* **288** : 117711. *doi : 10.1016/j.envpol.2021.117711.*

Li, L. and Wu, J. (2021). Spatiotemporal estimation of satellite-borne and ground-level $NO_2$ using full residual deep networks. *Remote Sens. Environ.* **254** : 112257. *doi : 10.1016/ j.rse.2020.112257.*

Meena, H. M., Machiwal, D., Santra, P., Moharana, P. C. and Singh, D. V. (2019). Trends and homogeneity of monthly, seasonal and annual rainfall over arid region of Rajasthan, India. *Theor. Appl. Climatol.* **136** : 795-811.

Saikia, A., Pathak, B., Singh, P., Bhuyan, P. K. and Adhikary, B. (2019). Multi-model evaluation of meteorological drivers, air pollutants and quantification of emission sources over the upper Brahmaputra basin. *Atmosphere* **10** : 703. *https:// doi.org/10.3390/atmos10110703.*

Singh, R. P., Kumar, S. and Singh, A. K. (2018). Elevated black carbon concentrations and atmospheric pollution around Singrauli coal-fired thermal power plants (India) using ground and satellite data. *Int. J. Environ. Res. Public Health* **15** : 2472.

Sur, R., Peng, W. Y., Strand, C., Spearrin, R. M., Jeffries, J. B., Hanson, R. K., Bekal, A., Halder, P., Poonacha, S. P., Vartak, S. and Sridharan, A. K. (2017). Mid-infrared laser absorption spectroscopy of $NO_2$ at elevated temperatures. J. *Quant. Spectrosc. Radiat. Transf.* **187** : 364-374.

Zhang, X., Just, A. C., Hsu, H. H. L., Kloog, I., Woody, M., Mi, Z., Rush, J. Georgopoulos, P., Wright R. O. and Stroustrup, A. (2021). A hybrid approach to predict daily $NO_2$ concentrations at city block scale. *Sci. Total Environ.* **761** : 143279.