CHANGE IN AIR POLLUTANTS AND SPATIAL HAZARD ANALYSIS USING HYBRID MODELLING IN THE SAKARYA BASIN (TURKIYE)

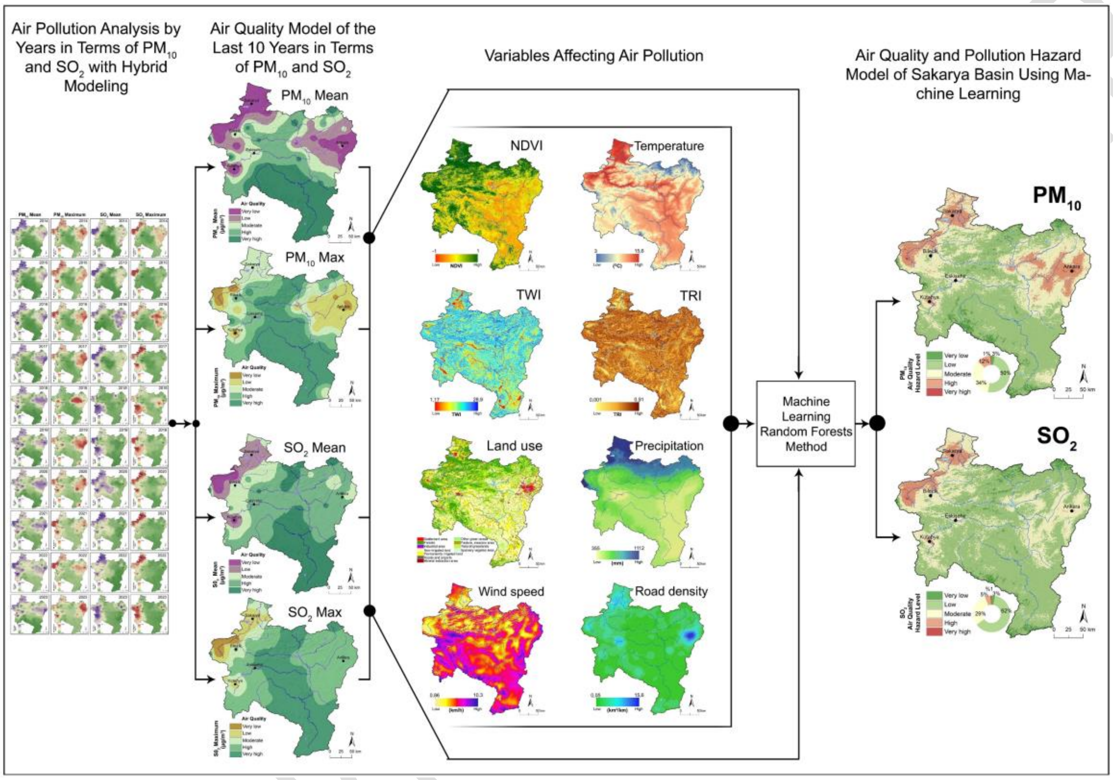
S. Murat UZUN

Marmara University, Faculty of Humanities and Social Sciences, Department of Geography, İstanbul, Türkiye. murat.uzun@marmara.edu.tr https://orcid.org/0000-0003-2191-3936

Nuriye GARİPAĞAOĞLU

Marmara University, Faculty of Humanities and Social Sciences, Department of Geography, İstanbul, Türkiye. nuriyeg@marmara.edu.tr https://orcid.org/0000-0003-4967-8536

Graphical Abstract



Abstract

The temporal and spatial distribution of air pollution can be analysed with high accuracy using various methods. This study examines the temporal and spatial variation of particulate matter (PM_{10}) and sulphur dioxide (SO_2) pollutants in the Sakarya Basin (Turkiye) using a hybrid modelling approach and presents a spatial hazard analysis. In the study, ground-based measurement data obtained from 16 monitoring stations for the period 2014–2023 were integrated with Sentinel-5 Tropomi satellite data, and calibrationand interpolation methods were applied to produce annual distributions of average and maximum values. Subsequently, the cumulative sum of data from the last ten years was used to reveal the spatial distribution of air quality. The air quality hazard analysis of the basin was modelled using the Random Forest Machine Learning method, employing -ten-year average and maximum result distributions together with topographic wetness index (TWI), land use, precipitation, temperature, road density, wind speed, topographic roughness index (TRI), and -normalized difference vegetation index (NDVI) analysis variables. The study show that PM_{10} concentrations generally decreased over the last decade, although temporal fluctuations were observed. The highest PM_{10} values were recorded in 2017 and 2021 and were mainly concentrated around Ankara, Kütahya,

Sakarya, and İnegöl. Short-term improvements were observed during the years 2020–2021 when COVID-19 restrictions were in effect. It was determined that SO_2 concentrations entered a significant downward trend after 2015. According to machine learning-based hazard analysis, the areas with the highest risk in terms of PM_{10} and SO_2 are the city centres of Ankara, Kütahya, Sakarya and İnegöl. The results indicate that air pollution in the Sakarya Basin is critical, particularly around industrial centres and major transport routes, while air quality levels are relatively lower in the northern and southern parts of the basin.

Keywords: Air quality, PM_{10}, SO_2, Hybrid modelling, Machine Learning, Random Forests Air pollution hazard analysis.

1. INTRODUCTION

Air quality is influenced by human activities, built structures, and the processes they generate, as well as by climatic conditions and topographical elements, and it is subject to change due to various air pollutants. When the concentrations of chemical elements in the atmosphereexceed certain threshold values, air quality deteriorates and reaches critical levels in terms of the environmental conditions, human health, ecosystem balance, and quality of life (Butler et al., 2008). In particular, increasing demands associated with population growth lead to intensification in many human activities, greater use of chemicals from multiple sources, increased fossil fuel consumption, and air pollution (Garipağaoğlu, 2015). Industrialisation, migration to large cities, unplanned urbanisation, and mining activities, significantly alter the concentration of chemical substances in the atmosphere and result in intense air pollution (Ceylan & Bulkan, 2018; Yener & Demirarslan, 2024). A notable example is the Great Smog of London in 1952, during which approximately 4,000 people lost their lives over a four-day period (Ferreira et al., 2013). Similarly, an increase in air pollution–related mortality was observed in Chinese cities such as Beijing and Shanghai between 2000 and 2015 (Mao & Zhao, 2014; Pyae & Kallawicha; 2024; Zhai et al., 2024). Beyond these examples, the rapid increase in air pollution sources has led to declining in air quality and growing environmental problems across many countries, basins, and cities worldwide (Demirarslan & Akıncı, 2018; García Nieto et al., 2018; Coşkun et al., 2023).

Particulate matter (PM_{10}) and sulphur dioxide (SO_2), which can cause significant problems when their concentrations exceed certain levels, are among the primary air pollutants prioritised for monitoring by international organisations such as the World Health Organisation (WHO) and the European Environment Agency (EEA) (WHO, 2021). Particulate matter (PM) consists of a mixture of solid and liquid particles suspended in the air, varying in size, shape, source and chemical composition (Pope & Dockery, 2006). Particle size plays a critical role in human health impacts and can trigger diseases such as emphysema, diabetes, and hypertension, potentially leading to mortality (Chen et al., 2018; Al Suwaidi et al., 2024). Changes in PM concentrations are governed by key factors including fossil fuel emission volumes, time of day, temperature, sunshine duration, humidity, precipitation, wind speed, and wind direction (Chen et al., 2013). Sulphur dioxide (SO_2), is a colourless, pungent-smelling reactive gas that reaches high concentrations, primarily during fossil fuel combustion, metal smelting, and other industrial processes (Azmi et al., 2025). During winter months, SO_2 concentrations tend to increase due to domestic heating andthe types of materials used to meet energy demands, which can result in significant environmental problems. Under such conditions, elevated SO_2 levels can cause increases in respiratory and other diseases, particularly among human populations (Kotan & Erener, 2023).

Basin-based analyses are increasingly recognised as essential for assessing air quality at both global and national scales, including in Türkiye (Şişman, 2019; Garipağaoğlu, 2024). Basins represent primary natural units in which geographical, meteorological, and topographical characteristics vary considerably and play a key role in the distribution and accumulation of pollutants (Uzun & Garipağaoğlu, 2022). At the same time, the spacial concentration of anthropogenic activities in certain areas and the interaction of these pollutant sources with natural environmental conditions, , which affect air quality in different areas, also stems from the holistic systematic structure of basins. The examination of different air pollutant parameters at the basin scale is of great importance for identifying regional pollution sources and developing effective local air quality management strategies. Basin-based approaches offer significant advantages in analysing air quality by enabling the assessment of local and regional pollution sources within natural boundaries (Zhou et al., 2019).

Air pollution exhibits a non-linear dynamic nature, making its prediction highly challenging. Furthermore, the observation and analysis process required to closely understand the distribution of air pollutants in the atmosphere involve significant costs (Liu et al., 2026). Therefore, various methods are used to determine the distribution and quantity of air pollutants (Vorapracha et al., 2015; Di et al., 2019; Kumari & Toshniwal, 2020). The most reliable results are obtained from stations that perform direct measurements. However, as there no stations measuring air pollutant concentrations in every area or city worldwide, alternative techniques are required. One such approach involves the use of satellite data to detect air quality and pollutants. In particular, the Sentinel 5 Tropomi satellite provides spatial distribution data for certain pollutants and aerosols in the atmosphere at specific spatial scales (Liv et al., 2017c; Coşkun et al., 2022). In addition, air pollutant analyses based on mobile measurements or short-term monitoring campaigns conducted over specific periods are also used. In this context, deterministic, statistical, heuristic, empirical, and stochastic methods are applied to simulate air pollutant concentrations in areas where measurements or sufficient data are not available (Liv et al., 2017a; Huang & Kuo, 2018). The validity and reliability of these methods depend on data availability, quality, and spatial scale (Prajul vd., 2025b). However, recent developments in artificial intelligence technologies, such as deep learning and machine learning, enable reliable results to be produced through the integration of different datasets (Bai et al., 2018; Liv d., 2017b; Delavar et al., 2019, Choubin et al., 2020; Yılmaz & Karagözoğlu, 2022; Mampitiya et al., 2024; Yang et al., 2024; Aydın & Kılar, 2025; Prajul & Subramanian, 2025). Accordingly, in recent years, hybrid models that integrate different methods or data sources have increasingly been used in the spatial distribution analysis of air pollutants (Chen et al., 2015; Irmak et al., 2024; Prajul vd., 2025a).

Air pollution occurring in any region or city can spread to other areas under the influence of natural conditions or produce global-scale effects through the release of different chemical substances into the atmosphere. These effects may reach hazardous levels for humans, other living organisms, and ecosystems. Consequently, air pollution is not only a phenomenon of present conditions but also has the potential to affect broader areas over the long term through its interaction with global climate change. For this reason, the extent of air pollution experienced in countries or cities with different levels of development has been continuously examined worldwide, and various measures have been implemented to address this issue (Garipağaoğlu, 2020). Among these measures, monitoring air pollution in both temporal and spatial dimensions is of primary importance (Agustine et al., 2017). Accurately determining the extent of air pollution using spatially appropriate models is crucial for the implementation of effective mitigation strategies (Toros et al., 2018). In recent years, air pollution problems have intensified in large cities within countries experiencing rapid industrial development and economic growth, together with migration and urbanisation processes (Vural & Şahinalp, 2023; Uyar, 2024; Akyürek, 2025; Aydın & Raja, 2025). Although air pollution has been examined from global and regional perspectives in numerous studies (Chen et al., 2015; Chelhaoui et al., 2024; Aydın & Kılar, 2025) the modelling approaches employed and the spatial limitations of the study areas constitute important constraints.Within the scope, this study utilises hybrid modelling approach to analyse changes in air pollutants and to conduct a spatial hazard analysis in the Sakarya Basin, which covers a large area in Turkiye and exhibits geographical diversity and air quality problems.

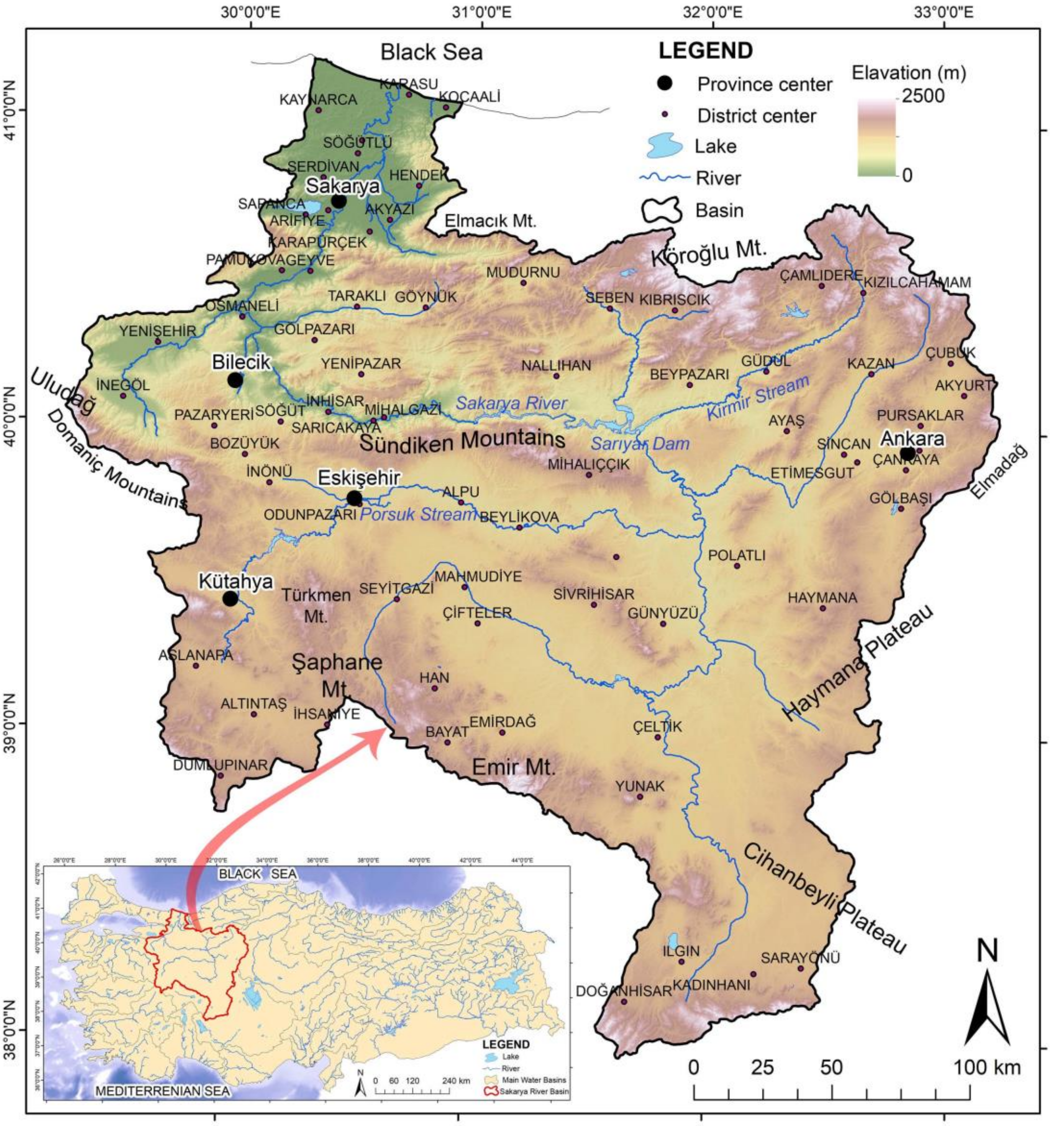


Figure 1: *Location of the study area*

The Sakarya Basin is one of Turkiye's 25 main water basins. The Sakarya River originates in the southern parts of the basin and from various upland areas in the west, flowing northwards and discharging into the Black Sea (Figure 1). The basin covers an area of 63,268.4 km², with an mean elevation of 980 m and an elevation amplitude of 2,325 m. The mean annual precipitation in the basin is 549.7 mm. Türkiye's capital city, Ankara, as well as densely populated cities such as Sakarya and Eskişehir are located within the basin boundaries. In addition, the basin contains important transportation networks, including the D-100 and O-4 motorways and railway lines, as well as industrial facilities distributed across differentareas. Located in northwestern Türkiye, the Sakarya Basin has a highly complex structure shaped by both natural and socio-economic dynamics (Figure 1). The coexistence of agricultural activities, settlements, transportation infrastructure, and intensive industrial areas results in air quality conditions influenced by a wide range of regionally variable pollutant sources.

The aim of this study is to model and analyse the spatial and temporal distribution of PM_{10} and SO_2 concentrations in the Sakarya Basin over the last ten years (2014-2023) using Geographic Information Systems (GIS), Remote Sensing (RS) and Artificial Intelligence (AI) technologies. Furthermore, this study aims to reveal the spatial distribution of air quality and produce air quality hazard distribution data for the basin using the Random Forests (RF)Machine Learning (ML) method in order to identify contributing processes. The hybrid modelling of recent air pollutant distributions, basin-based determination of risk sensitivity, and identification of possible mitigation measures constitute additional focal points of the study.

2. DATA AND METHOD

This research consists of two main stages. In the first stage, the temporal and spatial analysis of PM_{10} and SO_2 values, examined as air pollutants in the basin for the period 2014–2023, was carried out using a hybrid modelling approach. In the second stage, a hazard analysis was performed using the Random Forest (RF) method within Machine Learning (ML) framework, based on the ten-year mean and maximum datasets derived from the hybrid model results, together with eight different variables affecting air pollution.

The air pollutant data used in the study were obtained from the air quality database of the Turkish Ministry of Environment, Urbanisation and Climate Change and cover 16 monitoring stations *(1-Sakarya-Adapazarı, 2-Bilecik, 3-Bozüyük, 4-Eskişehir-Tepebaşı, 5-Eskişehir-Odunpazarı, 6-Bolu-Abant, 7-Konya-Sarayönü, 8-İnegöl, 9-Kütahya-Kentpark, 10-Polatlı-Ankara, 11-Bahçelievler-Ankara, 12-Demetevler-Ankara, 13-Sincan-Ankara, 14-Keçiören-Ankara, 15-Kayaş-Ankara, 16-Siteler-Ankara)* The dataset consists of daily mean, maximum, and minimum PM_{10} and SO_2 values for each year between 2014 and 2023 In addition, air quality data, including aerosol and SO_2 information for the period 2017–2023, were obtained from the Sentinel 5 Tropomi satellite via the Copernicus prgramme (Figure 2). A total of ten different variables were identified for the basin-scale air pollution hazard model. Two of the variables consist ofair pollutant model outputs. For the remaining variables, Normalized Difference Vegetation Index (NDVI) and land use data were obtained from the Sentinel 2 MSI multispectral satellite image dated 09.07.2025. Temperature and precipitation data were obtained from the Turkish General Directorate of Meteorology based on long-term average records from 28 stations. Road network data were derived from OpenStreetMap, For Topographic Roughness Index (TRI) and Topographic Wetness Index (TWI) analyses, a 10 m resolution Digital Elevation Model (DEM) obtained from the General Directorate of Mapping was used, while wind speed data in TIFF format were obtained from the Global Wind Atlas.

2.1. Spatial Distribution of PM_{10} and SO_2 Using Hybrid Modelling

The study employed a multi-step procedural framework to analyse the temporal and spatial changesin air pollutants (Figure 2). First, the air pollutants to be analysed within the basin were identified. In this context, particulate matter (PM_{10}) and sulphur dioxide (SO_2) were included in the scope of the study, as increases in their atmospheric concentrations have the potential to cause both short-term and long-term health problems (Azmi et al., 2024). PM_{10} refers to particles with diameters smaller than 10 micrometres, while SO_2 is a gas mainly produced by the combustion of fossil fuels (Al Suwaidi et al., 2024). Both pollutants are associated with respiratory diseases and cardiovascular disorders and can affect all living organisms within the ecosystem.

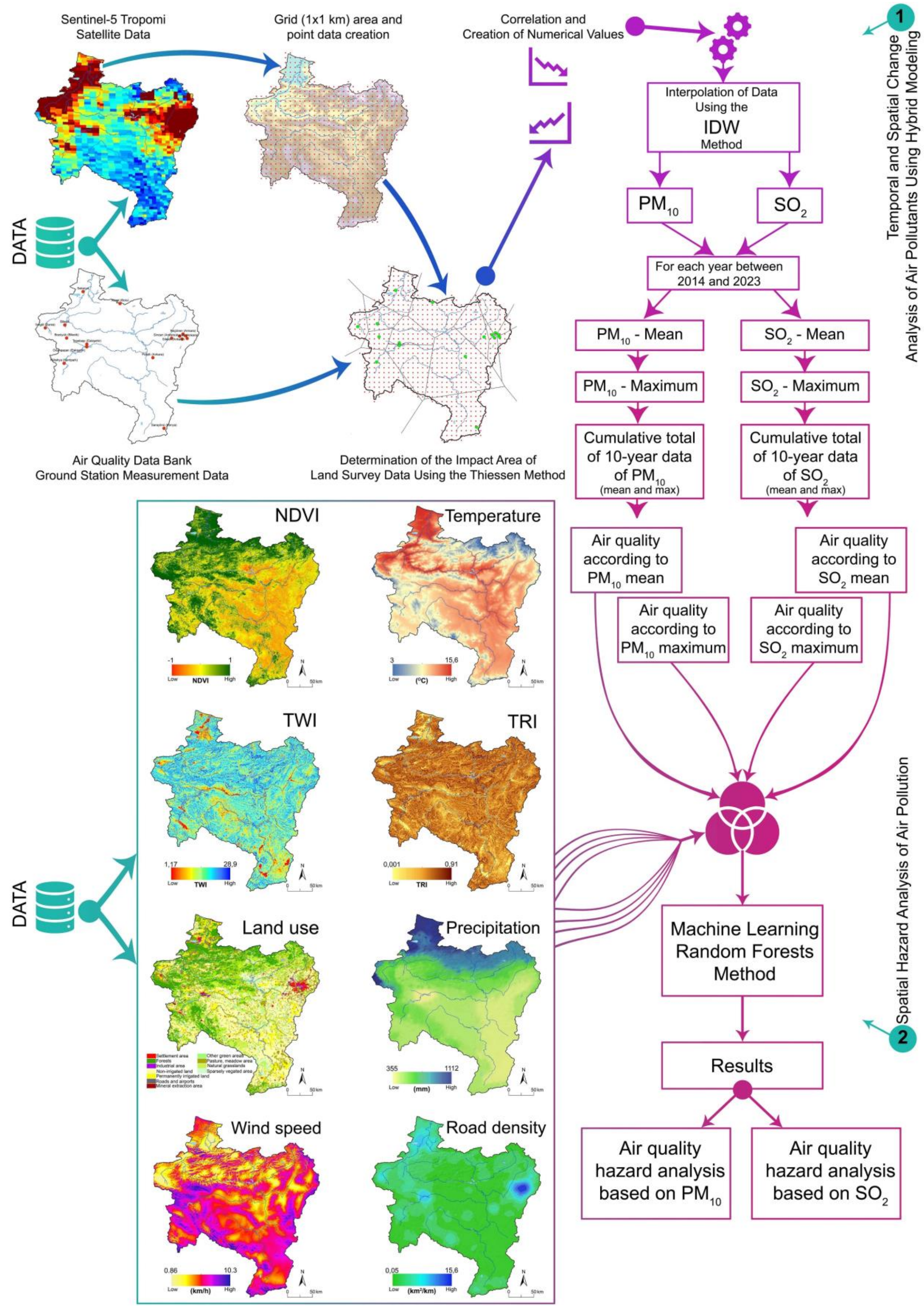


Figure 2: *The workflow diagram and hybrid model of the study: 1, the hybrid modelling of air pollutants between 2014 and 2023 is presented, while. 2, a hazard sensitivity model using Machine Learning with various data is demonstrated.*

The study aimed to produce annual average and maximum distribution maps of PM_{10} and SO_2 by correlating ground-based station measurement data obtained from the Turkish air quality database with satellite data. Within this scope, the required coding was first performed in Google Earth Engine (GEE) without downloading data for the Sakarya Basin, using the Sentinel-5 Tropomi satellite imagery, and spatial distributions of aerosol and SO_2 data were generated. The Sentinel-5 Tropomi (Tropospheric Monitoring Instrument) satellite data have been widely used in studies on air quality and pollutant distribution, ultraviolet radiation, climate monitoring and prediction (Coşkun et al., 2022). The satellite is funded by the European Space Agency (ESA) and provides free access through Copernicus programme. The aerosol and SO_2 datasets used in this study have a spatial resolution of 1132.2 metres and are provided in mol/m^2 units.

In this study, the quantitative values obtained from ground-based monitoring stations were organised and processed. At this stage, the Sentinel satellite data with a spatial resolution of 1132.2-were converted into point data using ArcGIS 10.5 software. The resulting point data were clustered within the zones of influence of the ground station measurements using the Thiessen polygon method. Satellite data and ground-based measurement data were then correlated to generate the values within these clusters. Relevant previous studies were reviewed, and calibration procedures were applied accordingly (Qi et al., 2019; Pyae & Kallawicha, 2024). Linear regression analysis was used to establish the relationship between ground station measurements and satellite-derived values. Regression analysis is a functional method that explains the mathematical relationship and correlation between two or more variables.

The calibrated numerical values were assigned to point features, and spatial distribution maps were subsequently generated using the inverse distance weighted (IDW) interpolation method. In hybrid air pollutant modelling, the spatial and temporal scales of emissions are of critical importance. At global or regional scales, spatial resolutions of 10 and 12 km are generally sufficient to capture pollutant distributions. However, for urban or basin-based modelling, data sources with grid spacing of less than 1-4 km are required (Ferreira et al., 2013). For this reason, both ground-based measurements and satellite data were used in the study, and PM_{10} and SO_2 distribution maps with an 1 km across the basin were produced. Inverse distance weighted (IDW) interpolation is method used to estimate pixel values at points where data are unavailable based on the quantitative analysis of values at known sample points (Loyd, 2007). The fundamental principle of the IDW interpolation technique is that estimated values decrease as the distance from known points increases, according to predefined mathematical relationships (Watson & Philip, 1985). Accordingly, the cell values are calculated by considering multiple surrounding points and their distances from the location being estimated (Jumaah et al., 2019). The estimated values are a function of both distance and the magnitude of nearby points, with the influence of each point decreasing as distance increases. IDW interpolation is applied using the following formula.

$$Z\,(X_0) = \frac{\sum_{1=1}^{n} Z(X_i).\, d_{i0}^{-r}}{\sum_{1=1}^{n} d_{i0}^{-r}}$$

In the formula, the predicted value at location X_0 is a function of the neighbouring measurements n (z(X0i) and i=1,2,..,n); r represents the upper bound defining the range of influence of each observation, and d denotes the distance between the measurement location Xi and the prediction location X0 (Watson & Philip, 1985). As the exponent increases, the weight assigned to observations located farther from the estimated location decreases. An increase in the exponent results in estimates that more closely resemble the nearest measurement locations.

The study produced annual distribution maps of the mean and maximum values PM_{10} and SO_2 values and presented graphs illustrating their temporal changes. Subsequently, the mean and maximum values for the period 2014-2023 were cumulatively aggregated. The resulting raster data data were classified into five categories using natural breaks method, and air quality data for PM_{10} and SO_2 values in the Sakarya Basin were generated (Coşkun et al., 2022).

2.2. Air Pollution Spatial Hazard Model Variables

Various variables are used in hazard sensitivity analyses related to air pollution and air quality (Tella et al. 2021). Variable selection differs across studies due to factors such as the characteristics of the research area and access to data sources. In this study, ten variables were used in the air pollution/air quality hazard sensitivity model. These include the cumulative average and maximum totals of PM_{10} and SO_2 representing air quality, as well as land use, Normalised Difference Vegetation Index (NDVI), Topographic Wetness Index (TWI), temperature, precipitation, Topographic Roughness Index (TRI), road density, and wind speed data for the basin. The selected variables correspond to those showing the highest correlation with air quality in previous studies (Eeftens et al., 2012; Rani et al., 2018; Choubin et al., 2020; Bounakhla et al., 2023; Azmi et al., 2024).

Land use data was produced using Sentinel 2 MSI multispectral satellite imagery using supervised classification and the nearest neighbour method. Land use is considered one of the key parameters influencing air pollution distribution, particularly as it reflects urbanisation patterns, population density, and industrial areas (Tella et al., 2021). The NDVI was calculated using the 5th and 8th bands of the Sentinel satellite imagery according to the formula [NDVI= (PNIR-PRED) / (PNIR+PRED)] (Myneni et al., 1995). Vegetation density derived from NDVI explains important spatial patterns in the relationship between biological processes and air quality.

Temperature and precipitation data for the basin were generated using the lapse rate adjustments, Thiessen polygons, and IDW interpolation methods based on long-term (1990-2023) average values from 28 stations located within the basin. As climatic factors, temperature and precipitation influence air pressure, condensation processes, and radiation levels, and are therefore directly related to air quality. Road density data was created using line density tool in ArcGIS 10.5 software based on linear features obtainedfrom OpenStreetMap. Road density highlights areas with elevated air pollution levels, particularly those associated with fossil fuel consumption. Wind speed data were obtained in TIFF format from the Global Wind Atlas and used directly. The TWI and TRI variables were derived from the basin's 10 m resolution DEM data using the following formulas.

$$TWI = \frac{([\![FA]\!] + 1)}{([\![E]\!] + 1).Log}$$

$$TRI = Grid(H_{mean} - H_{min}) \,/\, Grid(H_{max} - H_{min})$$

In the TWI formula, FA represents the slope value, and e represents the cumulative flow (Parker, 1982). In the TRI formula, H_{mean} denotes the mean elevation within the grid cell, while *Hmin* and *Hmax* represent the minimum and maximum elevations, respectively (Riley et al., 1999). Wind speed, TWI, and TRI are closely related to the air circulation, circulation, and gas deposition processes within the basin, thereby affecting air quality.

2.3. Spatial Hazard Model for Air Pollution Using Machine Learning Random Forests Method

The relationship between air pollution distribution and the parameters affecting it is complex, therefore, appropriate methods are required to to explore these relationships (Bai et al., 2018). In particular, recent advances in GIS and AI technologies provide effective approaches for revealing the relationship between variables affecting air pollution (Choubin et al., 2020). In machine learning (ML) modelling, where air quality hazard models are simulated using predictive variables, the selection of an appropriate number of predictive variables is essential (Carslaw & Ropkins, 2012). In this context, in order to identify the most relevant and effective variables, machine learning pre-processing includes a step known as feature selection, during which unnecessary variables are eliminated and the most influential variables are determined (Jia et al., 2016; Ceylan & Bulkan, 2018; Sayed et al., 2019). Simulated annealing methods and stochastic search algorithms can be used to determine the degree of relationship and interaction between variables (Kirkpatrick et al., 1983; Rere et al., 2015; Pinedo, 2016; Joharestani et al., 2019; Choubin et al., 2020).

The Random Forests (RF) method within Machine Learning is a combinatorial algorithm that is widely used in air pollution prediction studies (Choubin et al., 2020). This method is capable of handling multidimensional classification, regression, and correlation problems with high accuracy (Breiman 2001). In this type of algorithm, multiple classifications are employed for decision-making and for splitting each node. Each node represents a homogeneous region that contains observational data classified and divided by the algorithm. In the RF method, each decision tree corresponds to a subset of randomly selected data. Each decision tree then performs decision-making and and classifications based on the variables within its corresponding data subset. For the classification of new data, the final RF output is determined based on the aggregated prediction results of all decision trees rather than a single algorithm (Choubin et al., 2020). In this study, the Random Forest method within Machine Learning was selected to reveal air pollution hazard levels for both PM_{10} and SO_2 (Zimmerman et al., 2018). The Caret package (Kuhn, 2015) was used in R software together with the 10-fold cross-validation (CV) method to select the most suitable predictor variables for PM_{10} and SO_2 hazard prediction from the available set of ten variables. The k-fold cross-validation methodology (k = 10) was applied to calibrate the models using the Caret R package used for hazard modelling for both pollutants (Kuhn, 2008, Kuhn, 2015). The optimal value of the model variables was generated using tuning function of the Caret R package, employing the random-search tuneLength function, which produces the maximum number of tuning parameter combinations (Kuhn, 2015; Choubin et al., 2020). The operational methodology of RF consists of creating datasets (model predictions in this study), using these subsets to create decision trees, and combining the inferences. The decision tree classification process is applied using the following formula.

$$C = argmax\left(\sum_{K=1}^{I} Ck\right)$$

Here, $\mathrm{C_k}$ denotes the class prediction of the k-th decision tree, I denotes the total number of trees, and the final class prediction C is determined as the class with the maximum summed votes across all trees (Wu et al., 2024). In this study, the Random Forest method assigns a randomly varying weight value based on misclassification in the observational data to evaluate the importance of each predictor used in classification. In this study, 50 ntree values were used for RF decision trees. The study also used 10 as the mtry parameters, which is the number of randomly drawn variables that are candidates for insertion at each node (Tella et al., 2021). According to the Random Forest machine learning method, the hybrid model shows the highest impact, that is, importance level, for the air hazard model based on air quality data (Figure 3). The highest impact ranking among the dependent variables is land use, road density, TWI, NDVI, precipitation, TRI, temperature, and wind speed. At the end of the hazard model, the maximum number of terminal nodes in the decision tree (maxnodes) was left at the default setting, and PM_{10} and SO_2 air pollution hazard maps were generated.

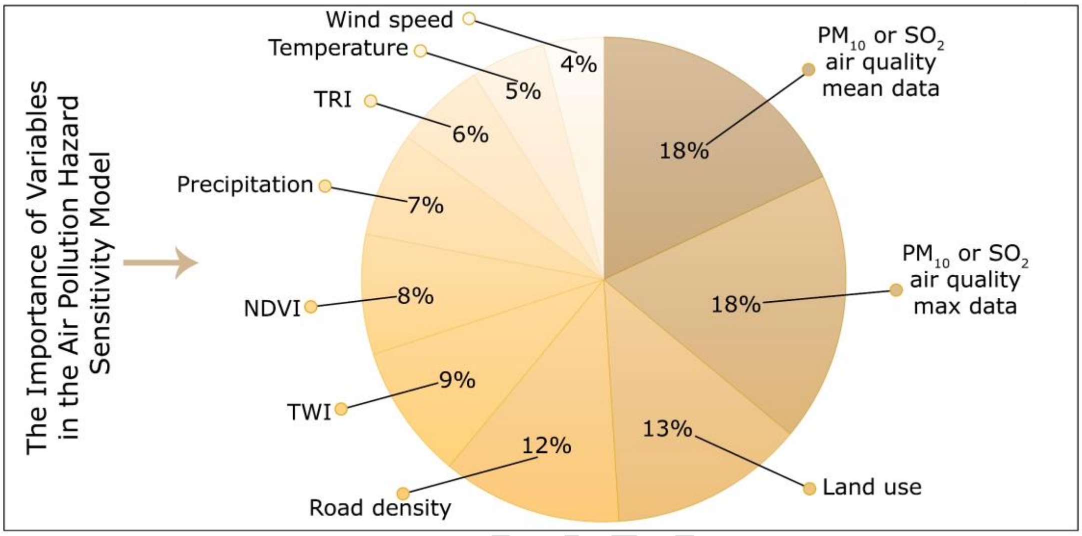


Figure 3: *Importance (impact) ratios of model variables*

2.4. Methods Used for Model Accuracy Assessments

The Root Mean Square Error (RMSE) formula was used to evaluate the performance of the hybrid model method proposed and used in this study. RMSE is calculated using the following formula.

$$MSE = \left(\frac{1}{n}\right) * \sum(y - y\_pred)^2 \qquad RMSE = \sqrt{MSE}$$

In the formula, n is the number of data points, y represents the actual values, and y_pred represents the predicted values (Chapi et al., 2017). In the formula, the mean square error (MSE) is first calculated, and then the RMSE is calculated as the square root of this value. The study used ground measurement station data as actual point data and PM_{10} and SO_2 hybrid model data as predicted data.

The accuracy analysis of the air pollution hazard sensitivity model developed using the Random Forest machine learning method in this study was subjected to the Receiver Operating Characteristic (ROC) method. The ROC curve and analysis is one of the methods that measures the efficiency of sensitivity analyses produced at different scales (Tehrany et al., 2013). The analysis is generally calculated by comparing the sensitivity data with the inventory of the event data as training and test data. The ROC method is calculated using the following formula.

$$FPR = \frac{FP}{FP+TN} \qquad TPR = \frac{TP}{TP+TN}$$

In the formula, FP denotes the number of false positive cases, TN denotes the number of true negative cases, TP denotes the number of true positive cases, and FN denotes the number of false negative cases (Tehrany et al., 2013). The ROC curve distributes false positive and true positive values vertically and horizontally on the x and y axes. The value remaining on the obtained ROC curve (AUC) approaching 1 indicates high model accuracy, while approaching 0 indicates low accuracy. In this study, 258 random point data were used in the accuracy analyses of the models. In the ROC analysis, 70% of these data points were randomly selected as training data (181 data points) and 30% as test data (77 data points). The obtained ROC results were evaluated for the accuracy of the hazard sensitivity model.

3. FINDINGS

3.1. Temporal and Spatial Variation of PM_{10} and SO_2 in the Sakarya Basin

The highest mean PM_{10} values in the Sakarya Basin between 2014 and 2023 were recorded in 2017 and 2021 (Figure 4). In terms of the distribution of PM_{10} mean values, the highest air pollution was found to be concentrated around Ankara, Sakarya, Kütahya, and İnegöl (Figures 5 and 6). The lowest values were observed in the southern and northern parts of the basin and towards the central part of the basin from these areas. According to the hybrid model, the mean PM_{10} values in the basin ranged between 26.63 and 38.1 during the 10-year period examined. According to the trend analysis of the mean values, the R2 value is 0.874, indicating a downward trend. When examining the temporal distribution of the basin's PM_{10} mean data, it can be seen that in 2014, 2015, and 2016, the highest values were in the Sakarya, Kütahya, and İnegöl areas, in 2017 in eastern Ankara, in 2018 in Kütahya, in 2019 in Eskişehir and Kütahya, and between 2020 and 2023 in large industrial cities (Ankara, Eskişehir, Kütahya, Sakarya).

According to the hybrid model results, when the Sakarya Basin was examined in terms of PM_{10} maximum values, the highest value was determined to be in 2018 (Figure 4). The R2 value of the PM_{10} maximum data, which shows a downward trend from 2014 to 2023, was determined to be 0.425. It is noteworthy that values were high in 2020 and 2021, when the COVID-19 pandemic was prevalent, and then began to decline. An analysis of the last 10 years of data shows that PM_{10} maximum values are particularly concentrated in Ankara, Sakarya, İnegöl and Kütahya (Figures 5 and 6). However, maximum PM_{10} values have varied in certain years. The highest values detected between 2014 and 2017 around the city centres of Ankara, Bilecik, Kütahya, İnegöl, and Sakarya were recorded in Nallıhan, Kütahya, and Bozüyük in 2018. In 2019, 2020, and 2021, the highest values were detected in Ankara, Polatlı, Eskişehir, Kütahya, İnegöl, and Bozüyük. In 2022 and 2023, Ankara has the highest values due to its population density. In recent years, high values have also been observed in İnegöl, Kütahya, Bozüyük, and Eskişehir outside this area.

When examining the hybrid model results for SO_2 mean values, it was determined that the highest value was measured in 2021. It was found that the trend in SO_2 mean values remained almost constant between 2014 and 2023, with an R2 value of 0.022 (Figure 4). It was determined that the SO_2 mean data, which showed temporal fluctuations, mostly did not exceed the threshold value for air pollution according to international standards. SO_2 values with temporal fluctuations also showed spatial variability (Figures 5 and 6). The areas with the highest mean SO_2 values in 2014, 2015, and 2016 were İnegöl, Sakarya, Bilecik, Ankara, and Polatlı, while the highest values in 2017 and 2019 were measured in Kütahya, İnegöl, and Sakarya, and in 2018 in Kütahya. Between 2020 and 2023, high measurements were detected in the western part of the basin, particularly in Kütahya, Eskişehir, İnegöl, and Sakarya.

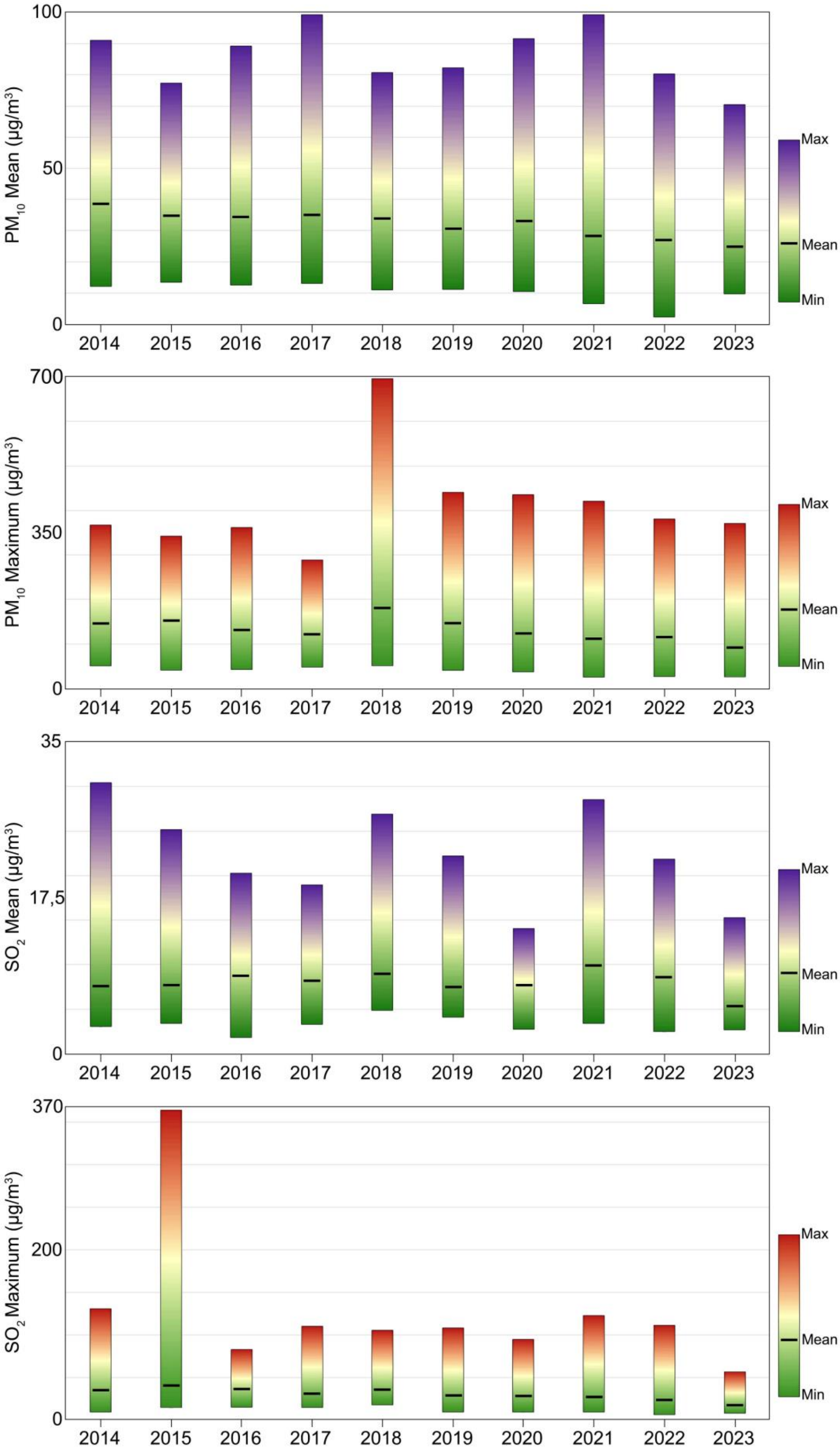


Figure 4: *Temporal variation of mean and maximum PM_{10} and SO_2 values in the Sakarya Basin*

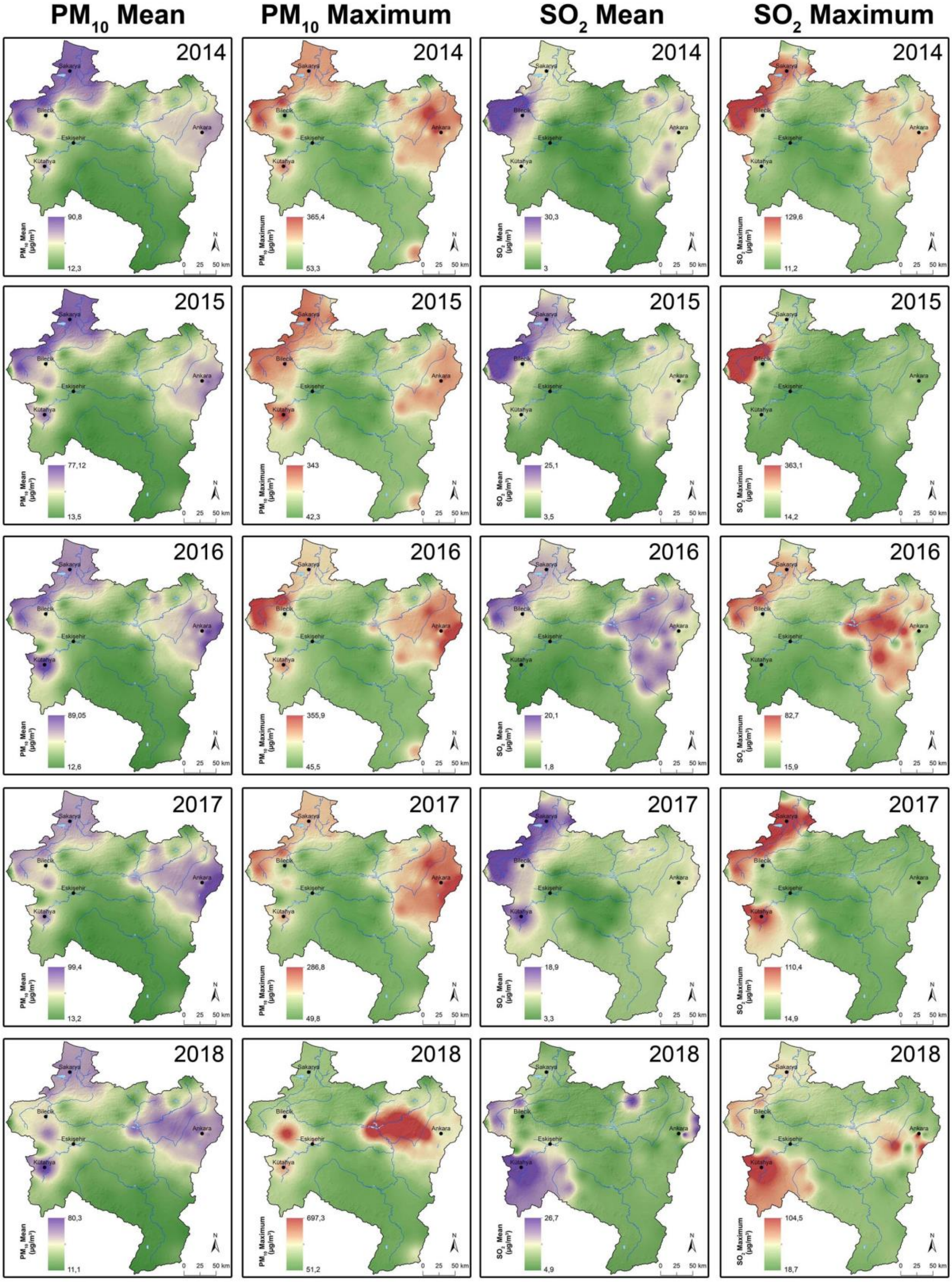


Figure 5: *Spatial variation of PM10 and SO2 mean and maximum values (2014–2018)*

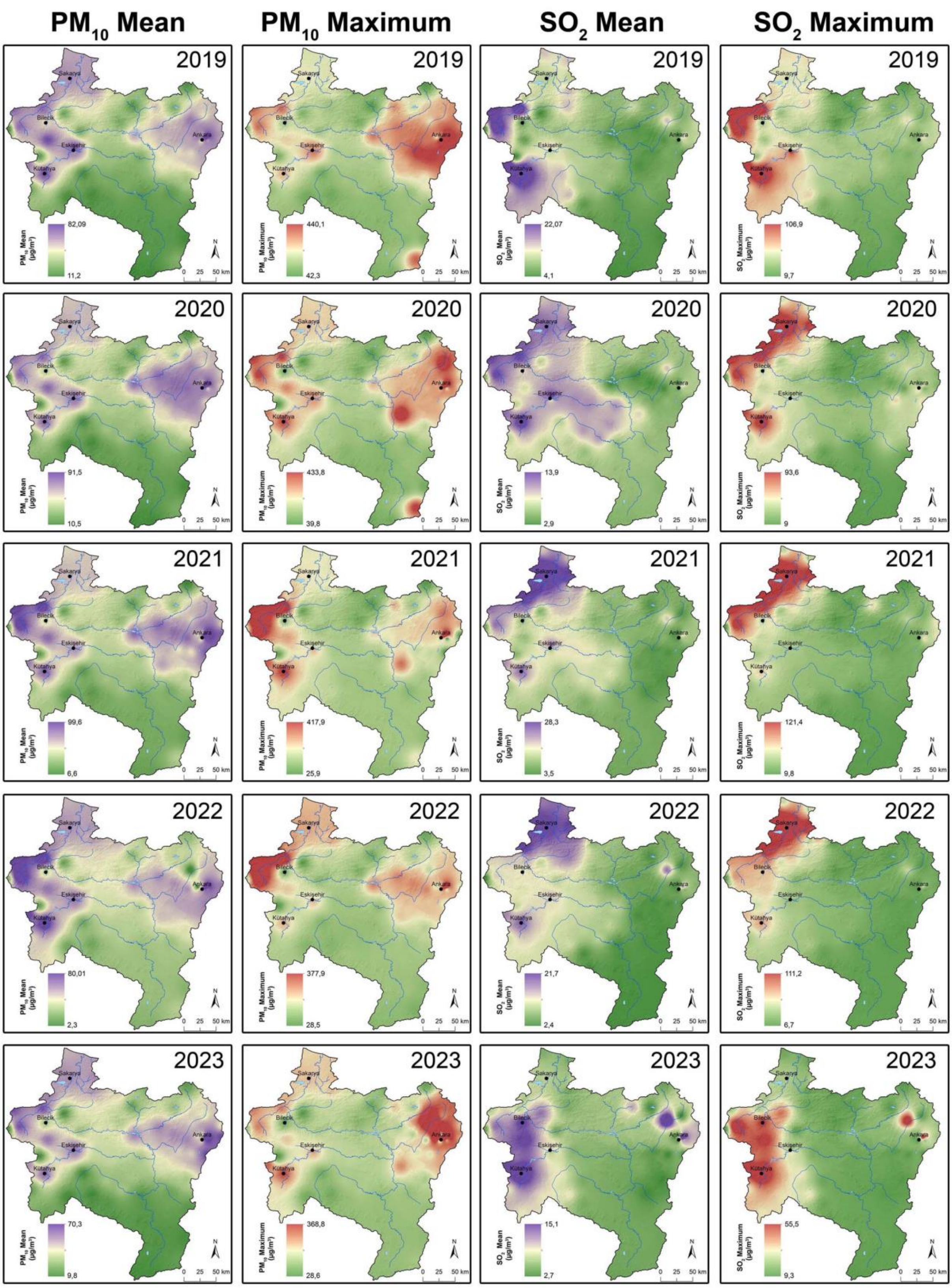


Figure 6: *Spatial variation of PM_{10} and SO_2 mean and maximum values (2019–2023)*

The year with the highest SO_2 maximum values was 2015 (Figure 4). In that year, a significantly higher SO_2 concentration was detected in İnegöl compared to other years. From that year onwards, it was observed that SO_2 maximum values showed a decreasing trend, with an R^2 value of 0.879. The temporal and spatial distribution of SO_2 maximum values showed spatial differences. In 2014 and

2016, around Ankara, Polatlı, Sakarya and İnegöl; in 2015, around İnegöl; in 2017, to the west of the Kütahya, İnegöl, Sakarya line; in 2018 in the south of Ankara and Kütahya, in 2019 and 2020 in Kütahya, İnegöl and Sakarya, in 2021 and 2022 in Sakarya and its southern parts, and in 2023 in the vicinity of Kütahya and Bilecik (Figures 5 and 6).

According to the European Union Air Quality Standard, the limit values are 40 µg/m³ for PM_{10} annually and µg/m³ for SO_2 annually. The limit values are the same for the winter period, when maximum values are mostly observed. In this regard, when PM_{10} air pollutants were analysed, it was determined that the mean values did not exceed the limit values for the basin average, but that pollution was well above this value in some local areas. In particular, it was determined that PM_{10} air pollutants were consistently above the limit values in the Sakarya, Ankara, Kütahya, and İnegöl areas. When PM_{10} maximum data were examined in terms of EU air quality limit values, it was found that the air average has consistently exceeded the limit value over the last 10 years. In this regard, it was determined that PM_{10} air pollution occurs during the winter season, particularly in the southern and northern parts, where pollution may be lower than that of the basin as a whole. SO_2 mean values are below EU air quality standards. Furthermore, in terms of maximum SO_2 mean values, the limit value was not exceeded only in 2017, 2020, and 2023. In terms of maximum SO_2 values, it was determined that EU air quality standards were not exceeded only in 2023 in terms of basin average values, while they were exceeded in all years in terms of basin maximum values.

3.2. Air Quality in the Basin in Terms of PM_{10} and SO_2 Based on Hybrid Model Data for the Period 2014-2023

PM_{10} and SO_2 values, whose spatial distribution was determined separately for each year over a 10-year period using the hybrid model, were cumulatively aggregated. The data obtained reveal the air quality in terms of the average and maximum values of PM_{10} and SO_2 based on the data from the Sakarya Basin over the last 10 years (Figures 7 and 8).

PM_{10} average values in the air quality results data show a spatial distribution in the basin at the following rates: lowest 16%, low 16%, moderate 17%, high 21%, and very high 30%. The areas with the highest air pollution and consequently the lowest air quality are concentrated in the city of Ankara and its surroundings, Kütahya, Bozüyük, İnegöl, Yenişehir, Sakarya and its surroundings (Figure 7A). It has been determined that air quality is very high in the mountainous and densely vegetated areas in the north of the basin and in the central and southern parts of the basin. Population distribution, urbanisation, and industrial centres within the basin have been particularly influential in creating this situation.

When examining the results data for PM_{10} maximum values in the Sakarya Basin in terms of air quality, the basin covers an area with the lowest percentage of 3%, low 15%, moderate19%, high 21% and very high 42%. The areas with the lowest air quality in terms of PM_{10} maximum values are the eastern part of Ankara city, İnegöl, Yenişehir, Osmaneli, and Kütahya city centres and their surroundings (Figure 7B). Population density and industry have been particularly influential factors in this situation. The areas with the highest air quality in terms of particulate matter are located in the area following the central line of the basin from north to south and expanding towards the south.

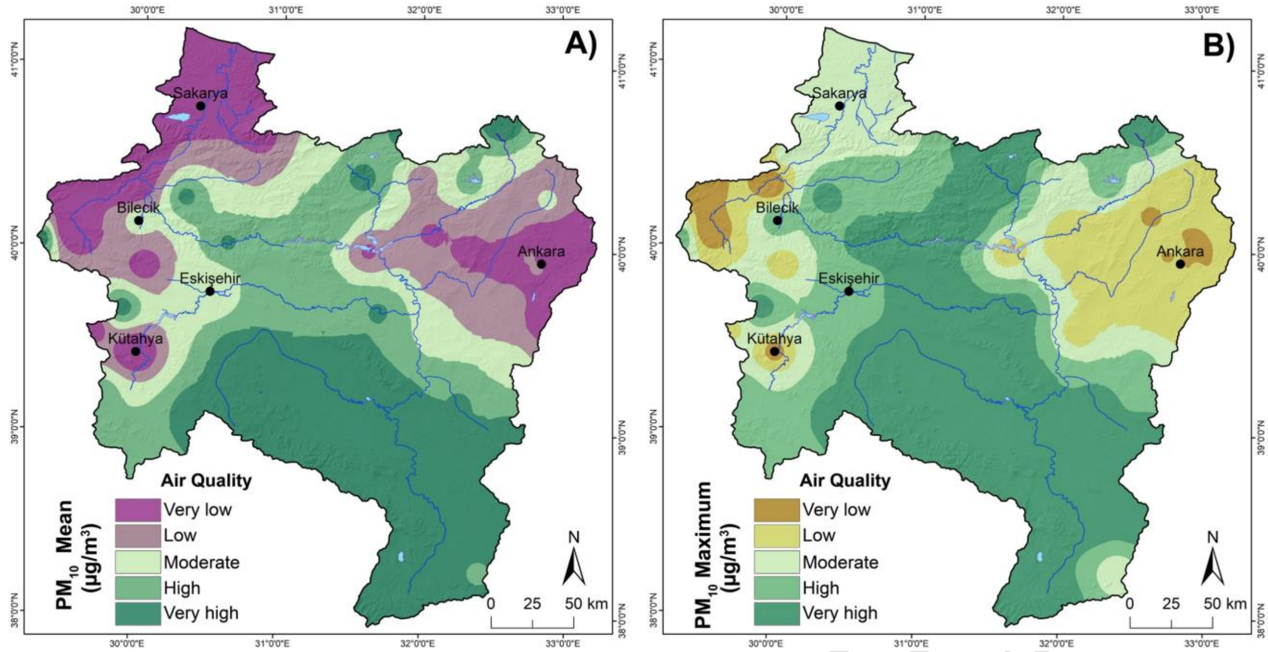


***Figure 7:** Air quality distribution based on the cumulative total of the basin's 10-year PM_{10} values. A) mean, B) maximum*

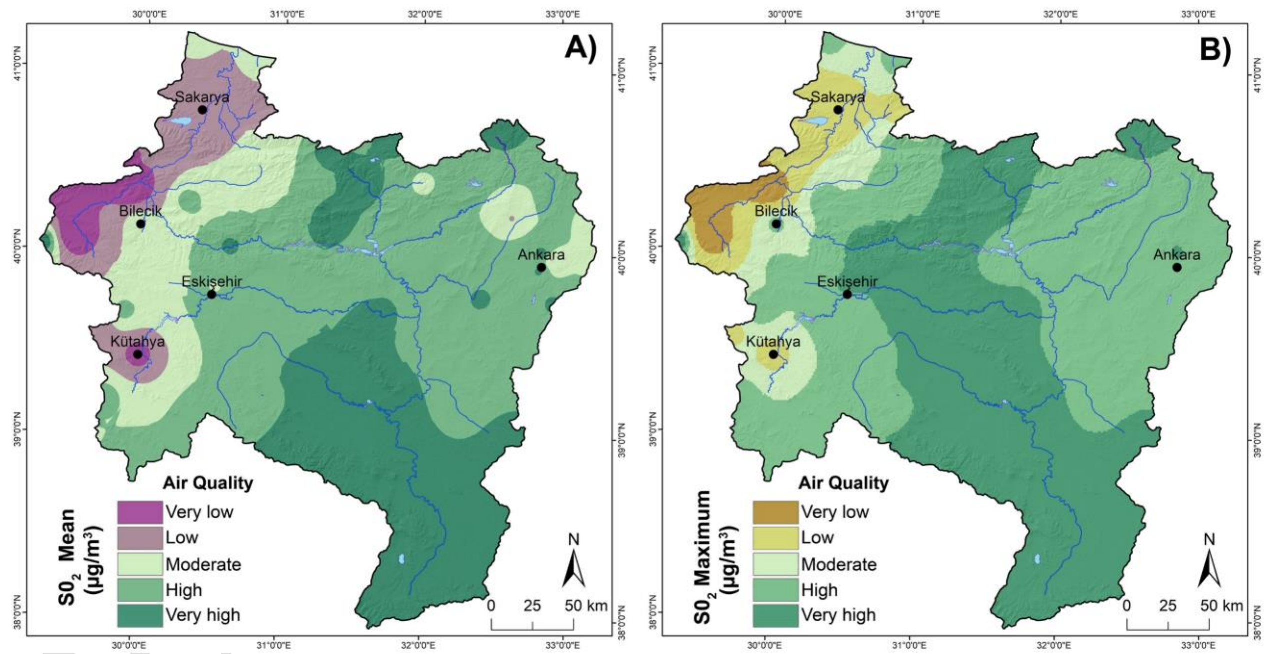


***Figure 8:** Air quality distribution based on the cumulative total of the basin's 10-year SO_2 values. A) mean, B) maximum*

When examining the air quality results data for SO_2 average values, the spatial distribution within the basin shows the following percentages: lowest 3%, low 9%, moderate 16%, high 44%, and very high 28%. The 10-year cumulative total of SO_2 average values indicates that air quality is high in two-thirds of the basin. According to the analysis data, the areas with the lowest air quality are around Kütahya, İnegöl, Yenişehir, Osmaneli, and Pamukova (Figure 8A). In particular, industrial facilities around Kütahya and İnegöl, as well as the geomorphologically-induced heat inversion effect in Yenişehir, İnegöl, and Osmaneli, have caused high SO_2 levels in these areas. The areas with the highest air quality are concentrated in the north and south of the basin.

When analysing the 10-year cumulative total of maximum SO_2 values, it was determined that the lowest air quality value was 2%, low 6%, moderate 8%, high 43% and very high 41% in the basin area. The findings reveal that air quality in 80% of the basin is high in terms of SO_2. This situation has been caused, in particular, by the use of natural gas-based domestic heating in the urban centres of the Sakarya Basin over the last 10-15 years, as has been the case throughout Turkiye. The areas with the lowest air quality are the urban centres of Kütahya, İnegöl, Yenişehir, Osmaneli and Pamukova (Figure 8B).

3.3. Air Pollution and Quality Spatial Hazard Analysis

The study presents average and maximum air quality data for PM_{10} and SO_2 based on the cumulative total of 10 years of data produced using the hybrid model. These data were supplemented with NDVI, land use, TWI, TRI, temperature, precipitation, road density, and wind speed data affecting air pollution and quality, and a hazard sensitivity analysis was performed for both pollutants using the Random Forest (RF) machine learning method. The analysis results were categorised into five levels using the natural breaks method and produced as separate maps for PM_{10} and SO_2. According to the findings, in the PM_{10} air quality hazard analysis, 3% of the total area of the basin has very low hazard, 50% has low hazard, 34% has medium hazard, 12% has high hazard, and 1% has very high hazard. According to the PM_{10} air pollutant model, the highest risk was found in the city centres of Ankara, İnegöl, Kütahya, and Sakarya (Figure 9A). Areas covering a wider area with high levels of risk were found to be the immediate vicinity of the cities mentioned, between Yenişehir, Osmaneli, and Pamukova, north and south of Ankara, and along the D-100 and O-4 roads in Sakarya. The use of lignite at the Ankara Çayırhan thermal power plant has increased the risk level in terms of PM_{10} in this region (around the Sarıyar Dam).

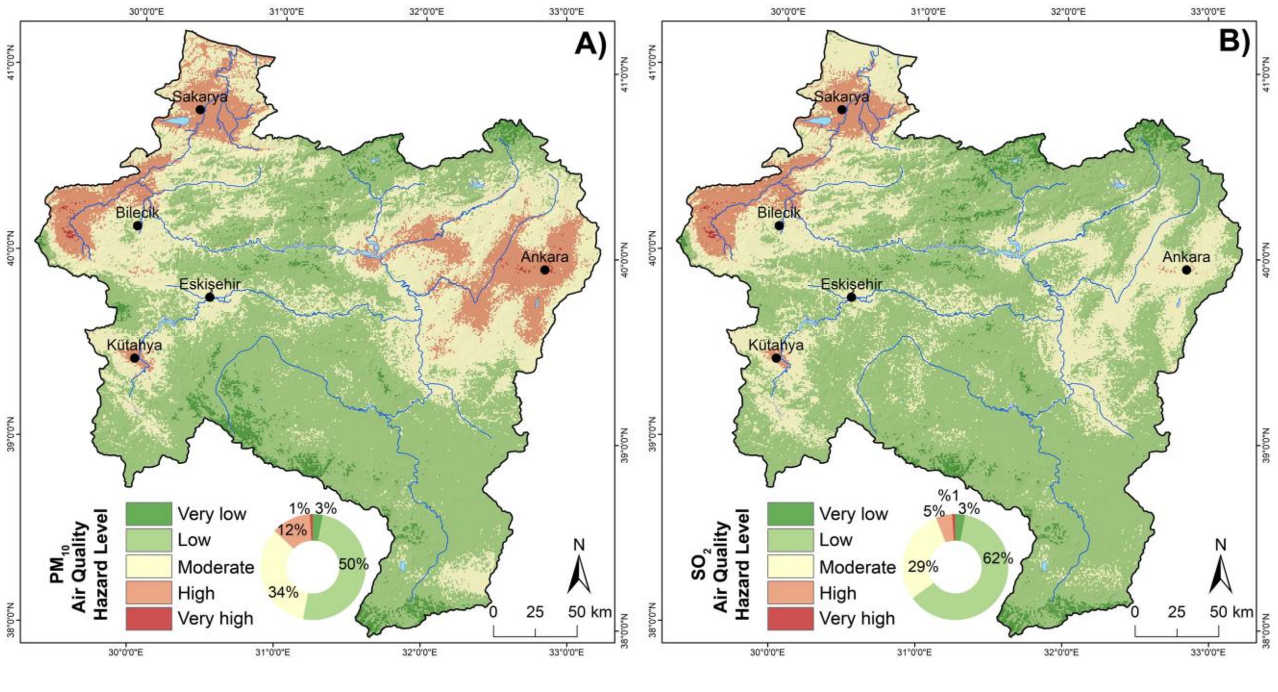


Figure 9: A) PM_{10} air quality hazard analysis result B) SO_2 air quality hazard analysis result

In the SO_2 air quality hazard analysis, 3% of the basin's total area is classified as very low hazard, 56% as low, 29% as moderate, 5% as high, and 1% as very high hazard. In terms of SO_2, the hazard level in 90% of the basin is fairly medium or below. This situation has been influenced by the widespread use of natural gas in city centres in recent years, the use of electric and hybrid cars, and exhaust emission controls. The concentration of industrial facilities that increase SO_2 concentrations in certain areas has also increased the risk of air pollution in specific areas. According to the SO_2 hazard model findings, the highest level of risk is in the city centres of İnegöl, Sakarya, and Kütahya (Figure 9B). At the same time, the immediate surroundings of these cities and the city centre of Ankara are at high risk. The presence of different types of industrial establishments in these areas and the lignite-fired thermal power plants in Kütahya province have increased the risk level in terms of SO_2. Air pollution risk levels were modelled as quite low in the northern, southern and central parts of the basin in terms of both SO_2 and PM_{10}.

4. DISCUSSION

One of the key contributions and findings of this study is the combined use of satellite data (Sentinel-5 Tropomi) and ground station measurements through hybrid modelling. This approach has increased spatial resolution and produced reliable distribution maps at a scale of 1 km. The literature shows that hybrid models are increasingly being used in air pollution analyses (Prajul et al. 2025a). For example, Chen et al. (2015) noted that hybrid models provide more accurate predictions, especially during periods of high PM concentration, while Chelhaoui et al. (2024) achieved similar high accuracy by combining WRF-CHIMERE and machine learning (Chen et al., 2015; Chelhaouiet al., 2024). Furthermore, the similarity between the accuracies in hybrid model studies and this study indicates that they can be used in air quality and pollution modelling (Delavar et al., 2019; Irmak et al., 2024). The application of hybrid modelling in the Sakarya Basin has reliably revealed the spatial distribution despite the limited number of stations in the region. This finding is consistent with the results of Coşkun et al. (2022), which showed that the calibration of satellite data with station measurements in the Kocaeli example achieved high success in air quality prediction (Coşkun et al., 2022).

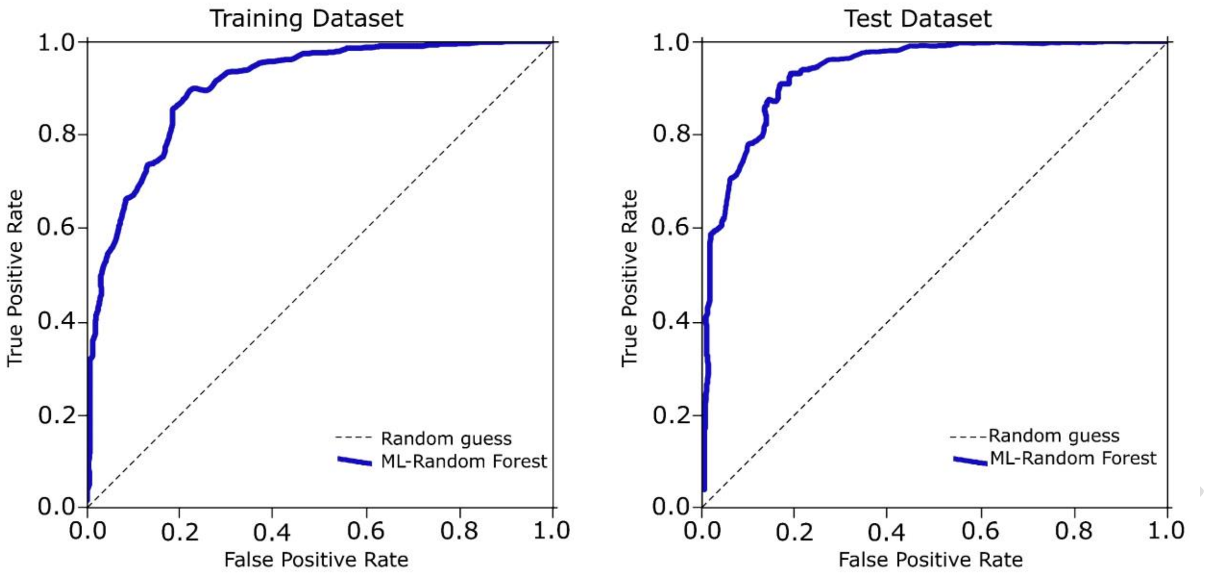
According to the research results presented in this study using hybrid modelling, PM_{10} concentrations fluctuated during the 2014–2023 period but generally showed a downward trend. High values were recorded particularly in 2017 and 2021, while short-term improvements in air quality were observed during the 2020–2021 period due to the impact of COVID-19 restrictions. This situation is similar to studies conducted in different cities in Turkiye. For example, Kotan & Erener (2023) observed temporary decreases in PM_{10} values in Kocaeli during the COVID-19 period, while Yener & Demirarslan (2024) also noted similar trends across Turkiye (Kotan & Erener, 2023; Yener & Demirarslan, 2024). However, the persistence of high PM_{10} levels around Ankara, Kütahya, İnegöl, and Sakarya creates the evidence that industrial activities, dense transport networks, and population pressure have created a lasting impact. This finding supports studies emphasising that industrialisation and urbanisation are decisive factors in air pollution in the Marmara and Central Anatolia regions (Demirarslan & Akıncı, 2018; Vural & Şahinalp, 2023; Akyürek, 2025; Aydın & Raja, 2025). Similar results have also been obtained in studies conducted on an international scale. Chen et al. (2018) examined $PM_{2.5}$ and PM_{10} concentrations in China and demonstrated that industry and transport are particularly decisive factors in large urban centres (Chen et al., 2018). Similarly, Bounakhla et al. (2023) emphasised that, along with meteorological factors, heavy traffic and industry in urban areas increase PM pollution in Morocco (Bounakhla et al., 2023). Similar trends are observed in many densely populated areas around the world where industry is concentrated in specific areas (Al Suwaidi et al., 2024). The fact that the results for the Sakarya Basin are consistent with these findings indicates that the region faces a problem that parallels global trends while also possessing its own unique spatial characteristics.

The findings of SO_2 values have shown a clear downward trend over time. Particularly in 2015, very high values were observed around İnegöl, with a downward trend becoming dominant in subsequent years. This situation can be explained by the widespread use of natural gas in many cities in Turkiye, the development of filtration technologies in thermal power plants, the monitoring of EU and WHO standards and practices, and the transformation in energy policies (increased investment in renewable energy). Indeed, Garipağaoğlu's (2020, 2024) studies emphasise a long-term downward trend in SO_2 values across Turkiye (Garipağaoğlu, 2020, 2024). However, the fact that high SO_2 levels are still observed around Kütahya, İnegöl, and Sakarya indicates that local risks persist in areas with a high concentration of industrial facilities. Similar results have been created in studies conducted in the Marmara Region (Kotan & Erener, 2023; Ceylan & Bulkan, 2018). Internationally, it is known that SO_2 is a persistent problem, particularly in regions where coal is heavily used for energy production (Mao & Zhao, 2014; Zhai et al., 2024). In this context, the findings from the Sakarya Basin are consistent with global trends but create the impression that risks persist in certain industrial centres at the regional level.

In the study, the average and maximum data obtained from the cumulative sum of PM_{10} and SO_2 data were tested using RMSE. The predicted data in RMSE were generated by the hybrid model data in the study, while the observed data were generated by ground measurement station data. According to the evaluation results, the average RMSE value for PM_{10} was calculated as 0.173, the maximum for PM_{10} as 0.191, the average RMSE value for SO_2 as 0.167, and the maximum for SO_2 as 0.198. The fact that the evaluation results are close to zero indicates that the model data has a high level of accuracy and reliability.

In the machine learning (Random Forest)-based hazard analysis, the city centres of Ankara, Kütahya, İnegöl, and Sakarya emerged as high-risk areas in terms of PM_{10}. Although the risk for SO_2 is more limited, it is similarly concentrated in areas with high industrial and transport activity. This result is similar to the spatial hazard analyses conducted by Choubin et al. (2020) for PM_{10} in Barcelona (Choubin et al., 2020). Furthermore, the analysis found that the topographic wetness index (TWI), land use, and road density variables were decisive. This situation creates the evidence that topography and meteorological factors play a critical role in pollutant distribution (Rani et al., 2018; Bounakhla et al., 2023). In particular, the prominence of the road density variable also shows that transport-related emissions are a key determinant of pollution distribution across the basin.

In the study, the hazard sensitivity analysis data modelled using Random Forest was evaluated for accuracy using ROC. According to the ROC data, the model's AUC value was calculated as 0.93 (Figure 10). The high accuracy rate of the model is similar to previous artificial intelligence-based air pollution distribution hazard analyses (Tella et al., 2021). The high value of the ROC analysis result ensures that this model can be used in basin-based air pollution hazard sensitivity analyses in different areas, particularly in Turkiye.



Figurel 10: *ROC analysis of the spatial hazard model for air pollution in the workplace*

5. CONCLUSION

The study has been one of the few studies conducted at the basin scale in Turkiye, revealing the spatio-temporal variation of PM_{10} and SO_2 specifically in the Sakarya Basin. The combined use of hybrid modelling and machine learning has enabled both the mitigation of data deficiencies and the reliable determination of hazard levels. However, there are some limitations. Firstly, the limited number of stations has complicated model calibration. Additionally, the resolution of satellite data may prevent a more detailed examination of urban sub-scale differences. Furthermore, in the Mediterranean Basinand transition zones, not only specific climatic, geographical, and anthropogenic emissions, but also the long-range transport of desert dust cause relatively high levels of PM and other air pollutants in certain areas. In this regard, fundamental factors such as emission volume, time of day, temperature, sunshine, humidity, precipitation, wind speed and direction govern changes in the concentration of pollutants such as PM_{10} and SO_2. Therefore, simulating air pollution is a highly complex and challenging process for both hybrid models and artificial intelligence-based models. The hybrid model used in this study, which yielded relatively good results (according to RMSE and ROC evaluation results), can be further improved in terms of accuracy and reliability in future studies by using a denser station network, high-resolution satellite data, and different machine learning algorithms (XGBoost, CNN-LSTM, and similar methods).

The hybrid model used in the study revealed the average and maximum distribution of PM_{10} and SO_2 values between 2014 and 2023. The model results show that both pollutant values are decreasing over time, but the highest values are concentrated in certain centres. It was concluded that air quality is at its lowest level in areas with dense populations and industrial activities, such as Sakarya, Ankara, Kütahya, and İnegöl. According to the air quality hazard analysis modelled using the Machine Learning-Random Forests method based on 10 different variables, it was determined that problems could arise in Sakarya, İnegöl, and Kütahya in terms of PM_{10} and SO_2 pollutants. Therefore, domestic, fossil fuel, and industrial air pollutants in metropolitan centres must be strictly controlled.

The findings of this study demonstrate the importance of considering the basin-scale air quality management. Implementing emission reduction policies is critical, particularly in the areas of Kütahya, İnegöl, Sakarya, and Ankara, where industrial concentrations are concentrated within the study area. Controlling transportation-related emissions, increasing green spaces, and strengthening continuous monitoring systems at industrial facilities will contribute to air quality management. In conclusion, this study demonstrates that the combined use of hybrid modelling and machine learning offers a powerful approach when revealing the spatial and temporal dimensions of air pollution in the Sakarya Basin. The model and hazard analysis results yielded reliable data consistent with both national and international literature. However, another dimension of the results suggests that specific measures must be taken at the regional scale in the Sakarya Basin.

5.1. The Framework of Limitations in the Research

The study presents a spatial hazard model for air pollution using a hybrid model for the annual distribution analysis of air pollutants and the Machine Learning Random Forests method based on 10 different variables at the basin level, which has certain limitations. The fundamental limitation is that the stations providing the most reliable measurement results in the hybrid modelling of the annual distribution of air pollutants are distributed in specific areas within the basin. In this regard, Sentinel 5 Tropomi satellite imagery was used in the study, and basin-based modelling was implemented. The IDW interpolation method was preferred in this model. Kriging and natural neighbourhood methods can also be used in subsequent studies based on this model. However, since each interpolation method uses different algorithms in terms of distribution, the result maps and, consequently, the distribution of air pollutants will vary.

In this study, analyses were conducted on PM_{10} and SO_2 pollutants, which had the least data gaps at local monitoring stations during the last annual period. In this regard, the data gaps for other air pollutants such as PM2.5, CO2, O3, and NO2 constitute another limitation of the study. Another limitation in this part of the study is that the analyses were not conducted seasonally. This approach was chosen because modelling based on average and maximum values for PM_{10} and SO_2 would have required a large amount of data, and analysing four different seasons would have made it difficult to detail this data. However, as winter air pollutants particularly contain maximum values, winter data was included in the study to a certain extent. The study did not include the effect of desert dust coming to Turkiye from the southeast during certain periods. This data, which is effective during the summer months and increases PM_{10} levels, is not available for each year within the 10-year period from the relevant public institutions. Therefore, it could not be used in the study.

A Random Forest method was used to model 10 different variables in the basin-based air quality-air pollution hazard sensitivity analysis. A significant limitation of the model is the lack of airborne access to building data. However, the sheer size of the building data for Ankara alone demonstrates the difficulty and limitations of the model during processing. For this reason, land use data capable of showing population density, urbanisation, and building data at certain scales was preferred in the spatial hazard model. At the same time, the mining and industrial areas within the land use data also made important contributions to the model.

References

Agustine, I., Yulinawati, H., Suswantoro, E., Gunawan, D., (2017). Application of open air model (R package) to analyze air pollution data. Indonesian J. Urban and Environ. Technol. 1 (1), 94–109.

Akyürek, Ö. (2025). Analysis of Temporal Changes of Air Pollutants Concentrations in Kocaeli, Turkey. Journal of Natural Hazards and Environment, 11(2), 406-417. https://doi.org/10.21324/dacd.1599855

Al Suwaidi, A., Ali, T., Atabay, S. et al. (2024). Particulate matter (pm10) monitoring in the United Arab Emirates using a satellite remote sensing based model. Discov Environ 2, 50 (2024). https://doi.org/10.1007/s44274-024-00068-8

Aydın, O. & Kılar, H. (2025). Development and Future Perspectives of Air Pollution Research Using Artificial Intelligence-Based Methods: A Bibliometric Review. Journal of Natural Hazards and Environment, 11(2), 471-487. https://doi.org/10.21324/dacd.1628030

Aydın, O. & Raja, N. B. (2025). Ankara'da Partikül Madde (PM10) ve Kükürt Dioksit (SO2) Konsantrasyonlarının İklim Parametreleri İle Mekânsal-Zamansal Analizi. Doğal Afetler Ve Çevre Dergisi, 11(1), 249-267. https://doi.org/10.21324/dacd.1533641

Azmi, W. N. F., Pillai, T. R., Latif, M. T., Shaharudin, R., & Koshy, S. (2024). Development of land use regression model to estimate particulate matter (PM10) and nitrogen dioxide (NO2) concentrations in Peninsular Malaysia. Atmospheric Environment: X, 21, 100244. https://doi.org/10.1016/j.aeaoa.2024.100244

Bai, L., Wang, J., Ma, X., & Lu, H. (2018). Air Pollution Forecasts: An Overview. International Journal of Environmental Research and Public Health, 15(4), 780. https://doi.org/10.3390/ijerph15040780

Bounakhla, Y., Benchrif, A., Costabile, F., Tahri, M., El Gourch, B., El Hassan, E.K., Zahry, F., Bounakhla, M., (2023). Overview of PM10, PM2.5 and BC and their dependent relationships with meteorological variables in an urban area in northwestern Morocco. Atmosphere 14 (1), 162.

Breiman, L. (2001). Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Butler, T.M., Lawrence, M.G., Gurjar, B.R., van Aardenne, J., Schultz, M., Lelieveld, J., (2008). The representation of emissions from megacities in global emission inventories. Atmospheric Environment 42, 703e719. http://dx.doi.org/10.1016/ j.atmosenv.2007.09.060.

Chapi, K., Singh, V. P., Shirzadi, A., Shahabi, H., Bui, D. T., Pham, B. T., & Khosravi, K. (2017). A novel hybrid artificial intelligence approach for flood susceptibility assessment. Environmental Modelling & Software, 95, 229-245. https://doi.org/10.1016/j.envsoft.2017.06.012

Carslaw, D. C. & Ropkins, K. (2012). Openair – An R package for air quality data analysis. Environmental Modelling and Software, 27–28, 52–61. https://doi.org/10.1016/j.envsoft.2011.09.008

Ceylan, Z. & Bulkan, S., (2018). Forecasting PM10 levels using ANN and MLR: a case study for Sakarya City. Glob. Nest J. 20, 281–290.

Chen, Y.Y., Shi, R.H., Shu, S.J., Gao, W., (2013). Ensemble and enhanced PM10 contamination forecast model based on stepwise regression and wavelet analysis. Atmos. Environ. 74, 346–359.

Chen, D., Xu, T., Li, Y., Zhou, Y., Lang, J., Liu, X. and Shi, H. (2015). A Hybrid Approach to Forecast Air Quality during High-PM Concentration Pollution Period. Aerosol Air Qual. Res. 15: 1325-1337. https://doi.org/10.4209/aaqr.2014.10.0253

Chen, G., Li, S., Knibbs, L. D., Hamm, N. A. S., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M. J., & Guo, Y. (2018). A machine learning method to estimate PM2.5 concentrations across China with remote sensing, meteorological and land use information. Science of the Total Environment, 636, 52–60. https://doi.org/10.1016/j.scitotenv.2018.04.251

Chelhaoui, Y., El Ass, K., Lachatre, M. et al. (2024). A new optimized hybrid approach combining machine learning with WRF-CHIMERE model for PM10 concentration prediction. Model. Earth Syst. Environ. 10, 5687–5701 (2024). https://doi.org/10.1007/s40808-024-02086-0

Choubin, B., Abdolshahnejad, M., Moradi, E., Querol, X., Mosavi, A., Shamshirband, S., & Ghamisi, P. (2020). Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. Science of The Total Environment, 701, 134474. https://doi.org/10.1016/j.scitotenv.2019.134474

Çoşkun, M., Şahiner, H., Abujayyab, S. K., & Canbulat, O. (2022). Evaluation of air quality based on satellite data and ground station measurements in Kocaeli provinceTürk Coğrafya Dergisi, (81), 53-68. https://doi.org/10.17211/tcd.1108123

Coşkun, M., Sahiner, H., Canbulat, O., Ozturk, A. (2023). Iğdır merkez ilçesinde hava kalitesi sorunu ve yerleşim uygunluk analizi. Cografya Dergisi. Cografya Dergisi, 47, 45-59. https://doi.org/10.26650/JGEOG2023-1233378

Delavar, M. R., Gholami, A., Shiran, G. R., Rashidi, Y., Nakhaeizadeh, G. R., Fedra, K., & Hatefi Afshar, S. (2019). A Novel Method for Improving Air Pollution Prediction Based on Machine Learning Approaches: A Case Study Applied to the Capital City of Tehran. ISPRS International Journal of Geo-Information, 8(2), 99. https://doi.org/10.3390/ijgi8020099

Demirarslan, K. O. & Akıncı, H. (2018). Assessment of Winter Season Air Quality in the Marmara Region Using GIS and Air Quality Data. Journal of Natural Hazards and Environment, 4(1), 11-27. https://doi.org/10.21324/dacd.344564

Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M. B., Choirat, C., Koutrakis, P., Lyapustin, A., Wang, Y., Mickley, L. J., & Schwartz, J. (2019). An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. Environment International, 130, Article 104909. https://doi.org/10.1016/j.envint.2019.104909

Eeftens, M., Beelen, R., de Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., Declercq, C., et al., (2012). Development of Land Use Regression models for PM2.5, PM2.5 absorbance, PM10 and PMcoarse in 20 European study areas; results of the ESCAPE project. Environ. Sci. Technol. 46 (20), 11195–11205. https://doi.org/10.1021/es301948k.

Ferreira, J., Guevara, M., Baldasano, J.M., Tchepel, O., Schaap, M., Miranda, A.I., Borrego, C. (2013). A comparative analysis of two highly spatially resolved European atmospheric emission inventories. Atmos. Environ. 75, 43–57. https://doi.org/10.1016/j.atmosenv.2013.03.052

García Nieto, P.J., S´anchez Lasheras, F., García-Gonzalo, E., de Cos Juez, F., (2018). Estimation of PM10 concentration from air quality data in the vicinity of a majör steelworks site in the metropolitan area of Avil´es (Northern Spain) using machine learning techniques. Stoch. Environ. Res. Risk Assess. 32, 3287–3298.

Garipağaoğlu, N. (2015). Türkiye ortam sorunları coğrafyası. (Genişletilmiş ve güncellenmiş 2. Baskı), Yeditepe.

Garipağaoğlu, N. (2020). Türkiye'nin hava kalitesinin maksimumlar ve limit değerler açısından incelenmesi. Türkiye'de Coğrafya araştırmaları içinde (s.186-222). Gazi Kitabevi.

Garipağaoğlu, N. (2024). Investigation of Meriç-Ergene Basin Air Quality in Terms of Averages, Maximums, and Limit Values. Eastern Geographical Review, 29(52), 1-18. https://doi.org/10.17295/ataunidcd.1492625

Huang, C. & Kuo, P. (2018). A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities. Sensors, 18, Article 2220. https://doi.org/10.3390/s18072220

Irmak, E., Yeşilbudak, M., & Taşdemir, O. (2024). Enhanced PV Power Prediction Considering PM10 Parameter by Hybrid JAYA-ANN Model. Electric Power Components and Systems, 52(11), 1998–2007. https://doi.org/10.1080/15325008.2024.2322668

Jia, H., Ding, S., Du, M., Xue, Y., (2016). Approximate normalized cuts without Eigendecomposition. Inf. Sci. 374, 135–150. https://doi.org/10.1016/j.ins.2016.09.032

Joharestani, M. Z., Cao, C., Ni, X., Bashir, B., & Talebiesfandarani, S. (2019). PM2.5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data. Atmosphere, 10, Article 373. http://doi.org/10.3390/atmos10070373

Jumaah, H. J., Ameen, M. H., Kalantar, B., Rizeei, H. M., & Jumaah, S. J. (2019). Air quality index prediction using IDW geostatistical technique and OLS-based GIS technique in Kuala Lumpur, Malaysia. Geomatics, Natural Hazards and Risk, 10(1), 2185-2199.

Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., (1983). Optimization by simulated annealing. Science 220 (4598), 671–680. http://doi.org/10.1126/science.220.4598.671

Kotan, B., & Erener, A. (2023). Seasonal analysis and mapping of air pollution (PM10 and SO2) during Covid-19 lockdown in Kocaeli (Türkiye). International Journal of Engineering and Geosciences, 8(2), 173-187. https://doi.org/10.26833/ijeg.1111699

Kuhn, M., (2008). Building predictive models in R using the caret package. J. Stat. Softw. 28, 1–26.

Kuhn, M., (2015). Caret: classification and regression training. Astrophysics Source Code Library.

Kumari, P., & Toshniwal, D. (2020). Impact of Lockdown on Air Quality Over Major Cities Across the Globe During COVID-19 Pandemic. Urban Climate, 34.

Li, Y., Wang, Y., Rui, X., Li, Y., Li, Y., Wang, H., Zuo, J., & Tong, Y. (2017a). Sources of atmospheric pollution: a bibliometric analysis. Scientometrics, 112, 1015–1045. https://doi.org/10.1007/s11192-017-2421-z

Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C., & Chi, T. (2017b). Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation. Environmental Pollution, 231, 997–1004. https://doi.org/10.1016/j.envpol.2017.08.114

Li, T., Shen, H., Yuan, Q., Zhang, X. X., & Zhang, L. (2017c). Estimating ground-level PM2.5 by fusing satellite and station observations: a geo-intelligent deep learning approach. Geophysical Research Letters, 44(23), 985–993. https://doi.org/10.1002/2017GL075710

Liu Y, Ma S, Zhang X and Chen J (2026). New energy development, land system transformation, and grain productivity quality: evidence from China. Front. Sustain. Food Syst. 10:1765501. doi: 10.3389/fsufs.2026.1765501

Loyd C.D., (2007), Local Modelsfor Spatial Analysis, CRC Press, 21-22p.

Mampitiya, L., Rathnayake, N., Hoshino, Y., & Rathnayake, U. (2024). Forecasting PM10 levels in Sri Lanka: A comparative analysis of machine learning models PM10. Journal of Hazardous Materials Advances, 13, 100395. https://doi.org/10.1016/j.hazadv.2023.100395

Ma, D., & Zhao, T. (2014). Talking about the current situation of air pollution in China and its governance recommendations. J. Hebei Inst. Archit. Civ. Eng, 2, 53-54.

Myneni, R. B., Hall, F. G., Sellers, P. J., & Marshak, A. L. (1995). The interpretation of spectral vegetation indexes. IEEE Transactions on Geoscience and Remote Sensing, 33(2), 481–486.

Parker, A. J. (1982). The topographic relative moisture index: an approach to soil moisture assessment in mountain terrain. Physical Geography, 3(2), 160-168.

Pope, C. A., & Dockery, D. W. (2006). Health effects of fine particulate air pollution: lines that connect. Journal of the Air & Waste Management Association, 56(6), 709-742.

Prajul, M. & Subramanian P. (2025). Optimizing Air Pollution Prediction in Urban Environments using a Hybrid RNN-PBO Model with IoT Data. 2025 International Conference on Visual Analytics and Data Visualization (ICVADV), 480-487.

Prajul, M., Subramanian, P., & Surendran, R. (2025a). Air pollution monitoring system using stacked attentional vectormap convolutional bidirectional network with bobcat optimization and IoT-cloud. Global NEST Journal, 27(3), 1-12. https://doi.org/10.30955/gnj.06937

Prajul, M., Subramanian P, Surendran, R. (2025b). Deep Feedforward Air Pollution Classification Empowered by Sophia-G Optimization, 2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Tirupur, India, 2025, pp. 1696-1702, doi: 10.1109/ICIMIA67127.2025.11200570.

Pyae, T. S., & Kallawicha, K. (2024). First temporal distribution model of ambient air pollutants (PM2.5, PM10, and O3) in Yangon City, Myanmar during 2019–2021. Environmental Pollution, 347, 123718. https://doi.org/10.1016/j.envpol.2024.123718

Rani, N.L.A., Azid, A., Khalit, S.I., Juahir, H., Samsudin, M.S., (2018). Air pollution index trend analysis in Malaysia, 2010-15. Pol. J. Environ. Stud. 27 (2), 801–808. https://doi.org/10.15244/pjoes/75964.

Rere, L.R., Fanany, M.I., Arymurthy, A.M., (2015). Simulated annealing algorithm for deep learning. Procedia Comput. Sci. 72, 137–144. https://doi.org/10.1016/j.procs.2015.12.114

Riley S. J, De Gloria S. D, & Elliot R. (1999). A terrain ruggedness index that quantifies topographic heterogeneity. Intermountain Journal of Sciences, 5, 23-27.

Qi, Y., Li, Q., Karimian, H., & Liu, D. (2019). A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory. Science of the Total Environment, 664, 1–10. https://doi.org/10.1016/j.scitotenv.2019.01.333

Sayed, G.I., Hassanien, A.E., Azar, A.T., (2019). Feature selection via a novel chaotic crow search algorithm. Neural Comput. Appl. 31 (1), 171–188. https://doi.org/10.1007/s00521-017-2988-6

Şişman, E. (2019). Trend Assessments for Selected Air Quality Monitoring Stations in Turkey. Journal of Natural Hazards and Environment, 5(1), 134-152. https://doi.org/10.21324/dacd.444503

Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2014). Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. Journal of Hydrology, 512, 332–343. https://doi.org/10.1016/j.jhydrol.2014.03.008

Tella, A., Balogun, A.-L., Adebisi, N., & Abdullah, S. (2021). Spatial assessment of PM10 hotspots using Random Forest, K-Nearest Neighbour and Naïve Bayes. Atmospheric Pollution Research, 12(10), 101202. https://doi.org/10.1016/j.apr.2021.101202

Toros, H., Bağış, S., & Gemici, Z. (2018). Ankara'da Hava Kirliliği Mekansal Dağılımının Modellenmesi. Ulusal Çevre Bilimleri Araştırma Dergisi, 1(1), 20-53.

Uyar, N. (2024). Monitoring and Analysis of Air Quality in Zonguldak Province by Remote Sensing. Turkish Journal of Remote Sensing, 6 (1), 57-67.

Uzun, M., & Garipağaoğlu, N. (2022). Integrated and sustainable watershed management model of Göksu River Basin (Sakarya River Basin) with spatial autocorrelation and cluster analysis approach. Turkish Geographical Review, (81), 23-38. https://doi.org/10.17211/tcd.1173420

Watson, D.F. & Philip, G.M. (1985) A Refinement of Inverse Distance Weighted Interpolation. Geoprocessing, 2, 315-327.

World Health Organization (WHO). (2021). Air quality guidelines: global update 2021. Geneva: WHO.

Wu, Y. Zhang, Z. Qi, X. Hu, W., Si, S. (2024) Prediction of flood sensitivity based on Logistic Regression, eXtreme Gradient Boosting, and Random Forest modeling methods. Water Sci Technol,89 (10): 2605–2624. https://doi.org/10.2166/wst.2024.146

Vorapracha, P., Phonprasert, P., Khanaruksombat, S., & Pijarn, N. (2015). A Comparison of Spatial Interpolation Methods for Predicting Concentrations of Particle Pollution (PM10). International Journal of Chemical, Environmental & Biological Sciences, 3(4), 302–306.

Vural, E., & Şahinalp, M.S. (2023). Investigation of particulate matter pollution in Şanlıurfa city under the influence of topographic and climatic factor. Türk Coğrafya Dergisi, (84), 53-66. https://doi.org/10.17211/tcd.1342144

Yang, Y., Cermak, J., Chen, X., Chen, Y., & Hou, X. (2024). High-Resolution PM_{10} Estimation Using Satellite Data and Model-Agnostic Meta-Learning. *Remote Sensing*, *16*(13), 2498. https://doi.org/10.3390/rs16132498

Yener, İ., & Demirarslan, K. O. (2024). Did Turkey experience reductions in air pollution during the COVID-19 lockdown and partial lockdown? Doğal Afetler ve Çevre Dergisi, 10(1), 179–191. https://doi.org/10.21324/dacd.1339741

Yılmaz, Z., & Karagözoğlu, M. B. (2022). Evaluatıion Of Air Pollution (Pm10 And So2) By Anova Method–The Case Of Mardin (Turkey) Province Kirklareli University Journal of Engineering and Science, 8(2), 343-356. https://doi.org/10.34186/klujes.1201251

Zhai, H., Huang, L., Emery, C., Zhang, X., Wang, Y., Yarwood, G., Fu, J. S., & Li, L. (2024). Recommendations on benchmarks for photochemical air quality model applications in China—NO2, SO2, CO and PM10. Atmospheric Environment, 319, 120290. https://doi.org/10.1016/j.atmosenv.2023.120290

Zhou, Y., Levy, J. I., Hammitt, J. K., & Evans, J. S. (2019). Regional public health benefits of reduced ambient fine particulate matter through energy scenario analysis. Environmental Health Perspectives, 117(1), 103-109.

Zimmerman, N., Presto, A. A., Kumar, S. P. N., Gu, J., Hauryliuk, A., Robinson, E. S., Robinson, A. L., & Subramanian, R. (2018). A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. Atmospheric Measurement Techniques, 11, 291–313. https://doi.org/10.5194/amt-11-291-2018