

A Deep Learning Approach for Spatiotemporal Estimation of PM_{2.5} Concentration in Urban Areas: A Case Study of Greater Taipei Area

Vincent Wiso^{1*} Hsuan Ren² Tang-Huang Lin³

Abstract

Air pollution in urban areas remains a threat, with particulate matter being one of the key indicators of pollution. Numerous studies show that particulate matter with a diameter of 2.5 μm or smaller (PM_{2.5}) affects public health, hence the need for more accurate models. Traditional numerical methods fail to capture pollutants' spatiotemporal patterns; however, neural networks can learn from environmental data inputs to obtain correlations. In this study, a Temporal Convolutional Network–Bidirectional Long Short-Term Memory (TCN–Bi-LSTM) hybrid model is proposed. TCN–Bi-LSTM is for PM_{2.5} monthly estimation in 2023 using Taiwan's Ministry of Environment air quality historical data from July 2015 to December 2022 – including weather parameters and land use data – to produce 1km PM_{2.5} maps via k nearest neighbors – inverse distance weighting (KNN – IDW). The results compare the proposed model against classical models like Convolution Neural Network and Long Short-Term Memory. The proposed model performs better compared to the others with a metric evaluation of $R^2 = 0.7960$ and root mean square error (RMSE) = 2.1606 $\mu\text{g}/\text{m}^3$ in the GroupKFold cross-validation. This study aims to contribute to air quality monitoring research by reducing reliance on aerosol optical depth (AOD) derived PM_{2.5} estimates, which are usually affected by cloud cover gaps during satellite AOD retrieval. This study, therefore, addresses the missing data issue by producing continuous PM_{2.5} concentration exposure maps. Future work aims to incorporate independent surface PM_{2.5} observations to check whether new findings and results will be obtained.

Keywords: Particulate Matter (PM), Spatiotemporal, Temporal Convolutional Network, Long Short-Term Memory

1. Introduction

Outdoor air pollution, caused by various pollutants, remains a threat in urban areas. The World Health Organization (WHO) estimated that, in 2019, the majority of the world's population lived in areas that did not meet air quality guidelines, and this was projected to cause approximately 7 million premature deaths annually. This mortality is attributed to exposure to fine particulate matter (PM), which leads to multiple diseases like respiratory and lung cancer (WHO, 2024; WHO, 2025). Particulate matter refers to a blend of solid and liquid particles, of different sizes, that are inhalable and found in the air (US EPA, 2025; WHO, 2025). In Taiwan, studies have shown positive associations between long-term PM_{2.5} exposure and cardiovascular and respiratory morbidity, including hospital visits and outpatient visits (Chien *et al.*, 2018; Wang *et al.*, 2021; Lo *et al.*, 2022). In line with this, it is crucial to estimate the PM_{2.5} concentration accurately spatially and temporally. High-resolution PM_{2.5}

concentration estimates complement the monitoring stations' observations, aid in exposure assessment of cohort studies, and policy evaluation of air pollution.

Currently, properly calibrated air quality monitoring systems offer coarse resolution (spatial and temporal) surface-level measurements (Song *et al.*, 2019). The uneven distribution and sometimes few air quality monitoring stations make it difficult to accurately estimate PM_{2.5} concentration at a high spatial resolution in huge geographical areas.

Modeling for surface-level PM_{2.5} concentration includes a myriad of techniques like land use regression, numerical model simulations, geostatistical interpolation, and machine learning, with each having its pros and cons. Numerical models simulate the transformation, transport, and deposition of fine particulate matter (Li & Jiang, 2023). However, this method requires a high amount of computing resources, making it difficult to obtain fine spatial and temporal resolution PM_{2.5} concentrations (Yu *et al.*, 2022). Geostatistical interpolation methods utilize spatial autocorrelation among observation data to produce

¹ Master Student, Centre for Space & Remote Sensing Research, National Central University

² Professor, Centre for Space & Remote Sensing Research, National Central University

³ Distinguished Professor, Centre for Space & Remote Sensing Research, National Central University

* Corresponding Author, Tel: +886 981 038441, E-mail: vincletus2050@gmail.com

Received Date: Jun. 12, 2026

Revised Date: Jun. 29, 2026

Accepted Date: Aug. 19, 2026

continuous PM_{2.5} concentration maps. However, this method assumes spatial stationarity and, in some cases, may have over-smoothing or underestimation, especially in sparsely covered areas (Lee *et al.*, 2012; Requia *et al.*, 2019). Land use regression (LUR) estimates PM_{2.5} concentrations by modeling the relationship between land-use variables, such as cover types and distance to emission sources, and pollution levels. Low temporal resolution and low capacity to capture nonlinear relationships among predictors compromise their use (Zhang *et al.*, 2021b; Zhang *et al.*, 2018). Machine learning enables capturing of nonlinear relationships between surface-level PM_{2.5} concentrations and variables like land use characteristics, weather parameters, and satellite aerosol optical depth (AOD), offering improved accuracy (Chen *et al.*, 2018). However, machine learning models may have limited capacity to learn deeper relations, for example, global spatial information and pollutant concentrations' temporal change (Qin *et al.*, 2019).

The majority of surface-level PM_{2.5} concentration methods (apart from machine learning) adopt a linear approach and ignore spatiotemporal relationships. This makes it difficult to estimate PM_{2.5} concentration, especially in areas with intricate atmospheric sources and diverse sources of pollution. On the other hand, deep learning utilizes neural networks to learn the (nonlinear) spatiotemporal relationships between PM_{2.5} concentration and influencing variables like land use characteristics, meteorological parameters, and satellite AOD (Li *et al.*, 2025; Wu *et al.*, 2024).

In this study, we aim to use a hybrid TCN – Bi-LSTM model to estimate monthly PM_{2.5} concentration in the Greater Taipei Area. The final product will be a 1km monthly PM_{2.5} concentration map. The study aims to contribute to the bulk of air quality monitoring by mitigating the missing data issue, mainly caused by cloud cover on satellite observations during aerosol optical depth retrieval.

The report is organized into five sections, with part one being the introduction. Part two explores related work in this line of research, while part three focuses on the study area, dataset, and methodology. Part four discusses the results, while Part five concludes the report.

2. Related Work

The related work reviews approximately 30 articles from the last 30 years, as existing research. This section discusses the importance of PM_{2.5} estimation, challenges associated with PM_{2.5} estimation, data inputs employed, deep learning (and other) modeling techniques employed, results verification, and interpretation, from the perspective of existing research.

This section also describes the gaps in knowledge from existing research, which will influence our research study's contribution to PM_{2.5} estimation studies.

The applications of PM_{2.5} estimation from the previous works were threefold. Some studies focused on modeling PM_{2.5} estimation for health outcomes and on-hospital visits (Chien *et al.*, 2018; Wang *et al.*, 2021), epidemiological assessments for public health research (Jung *et al.*, 2018; Zhang *et al.*, 2018; Song *et al.*, 2019), and other uses, such as modeling biomass burning transportation (Lin *et al.*, 2014) and forecasting accuracy of PM_{2.5} estimation (Ong *et al.*, 2016).

The majority of previous studies argue that mapping continuous PM_{2.5} exposure surfaces enables the realization of local spatial heterogeneity, leading to the reconstruction of long-term spatiotemporal patterns for regional comparisons (Chen *et al.*, 2018; Yang *et al.*, 2020; Cheng & Hsu, 2019). Mapping continuous exposure surfaces has been a strong motivation factor, with some studies developing AOD – PM_{2.5} models to estimate pollution in areas with limited air quality monitoring stations, various land use contexts, and complex terrain (Jung *et al.*, 2018; Wang *et al.*, 2025).

This is because output from air quality observation stations is point data, which deteriorates with an increase in distance (from the observation stations) when interpolated (Lee *et al.*, 2012). Furthermore, the observation stations in most areas are sparse and unevenly distributed due to various factors such as operational constraints and suitable site accessibility affecting their installation (Hacıoğlu *et al.*, 2016; Gupta *et al.*, 2018). Empirical evidence of continuous exposure mapping from some studies showed the realization of fine-resolution map products (Wei *et al.*, 2023; Li *et al.*, 2025; Wang *et al.*, 2025). This improvement leads to variations in spatial and temporal scale choices, with the specifications to be employed depending on the objectives to be achieved and tradeoffs to be made.

The previous works concur that multisource data is crucial, with the majority utilizing satellite AOD, chemical transport model (CTM) scaled AOD (reanalysis input data), weather parameters like precipitation, and land use context (Li *et al.*, 2017 & 2022; Zhang *et al.*, 2018; Berrocal *et al.*, 2020). This is because various factors influence PM_{2.5} concentrations; for instance, wind speed affects the accumulation, dispersion, and transportation of PM_{2.5} (Chang, 2019). The issue of satellite AOD missing data during retrieval was addressed mainly by imputing the AOD to obtain a gapless product (Wei *et al.*, 2023; Wang *et al.*, 2025). The former study (Wei *et al.*, 2023) further supplemented the AOD product with a CTM-scaled AOD to stabilize observations (trading off fine resolution for coarse resolution).

The literature review identified three major modelling approaches for PM_{2.5} estimation. Some

adopted a geostatistical approach with a spatial regression aspect where PM_{2.5} as a spatiotemporal input is encoded in a myriad of ways, such as spatial weights or distance, for instance, stratified land use regression and kriging interpolation (Zhang *et al.*, 2021b) and IDW (Wu *et al.*, 2024). Nonlinear function modelling was another approach, where (hybrid) neural network regressors were employed, for instance, geographic regression neural network (Li *et al.*, 2017), and a forest-based machine learning method (Requia *et al.*, 2019). Finally, deep sequence approaches, which model spatiotemporal dependence in their architecture, for instance, a hybrid CNN – LSTM model was used to estimate PM_{2.5} concentration (Qin *et al.*, 2019), a Fast Fourier Transform (FFT) LSTM was employed to forecast PM_{2.5} (Yu *et al.*, 2022). Furthermore, deep sequence approaches consider the observation station sparsity issue; for instance, one approach modeled PM_{2.5} estimation using a correlation-based station selection and spatial attention to prevent some stations from dominating (Lin *et al.*, 2022).

Different authors handled model generalization and evaluation in different ways. Some opted for a cross-validation design, where samples are held out for testing (Chen *et al.*, 2018; Wang *et al.*, 2025), while others opted for the train, validation, and test splits design (Qin *et al.*, 2019; Lin *et al.*, 2022; Hu *et al.*, 2024). Model spatial generalization was explored in two ways, where one was based on the global-to-local framework, with evaluation focusing on localized method behavior under site-based splits using a geographically and temporally weighted nearest neighbor logic (Li *et al.*, 2022). The other exploration way was spatial generalization at the provincial level. However, such regionalization introduced boundary issues that made spatial generalization difficult (Chen *et al.*, 2018). On this note, with no empirical evidence, Qin *et al.* (2019) argued that single-city generalization is preferable to multi-city generalization; hence, the latter should be explored before warranting generalization claims. The temporal generalization perspective focused on the model performance with temporal regime changes. For instance, Zhang *et al.* (2021a) concentrated on the short-term and long-term behavior of PM_{2.5} concentration, Yu *et al.* (2022), and Li & Jiang (2023) centered on forecasting, prioritizing predictive ability across time rather than the (unmonitored) station areas. Jung *et al.* (2018) addressed the seasonality and regime dependence of PM_{2.5} concentrations across seasons and locations, and Lin *et al.* (2014) focused on the Southeast Asian springtime biomass-burning window. There was an observation that the temporal aggregation changes one's claim; for instance, daily estimation does not translate to monthly estimation reliability and vice versa. This is because coarse temporal aggregation can give the impression of improved results while masking local errors (Chen *et al.*, 2018; Wei *et al.*, 2023).

Three patterns emerged in the interpretation of PM_{2.5} findings. First, interpretability by design, whereby model output is readable, for instance, in spatial statistical relationships, through which interpretability is realized in spatial guidelines, such as distance- conditioned blending that explains the reliability of a kriging surface (Lee *et al.*, 2012). Second, is attribution after fitting, whereby models rank variables influencing predictions. For instance, Chen *et al.* (2018) employed ensemble learners to check variables and interpret their influence, and Wei *et al.* (2023) used xAI screening to interpret large-scale aerosol-influencing variables. Third, a pattern to process narrative, whereby authors interpret the mapped output into plausible drivers by diagnostic structure rather than formal causal claims. For instance, Lin *et al.* (2014) framed springtime PM_{2.5} elevation levels around biomass-burning influence and utilized emission comparisons to support attribution. Wang *et al.* (2021) and Wang *et al.* (2025) employed regime stratification as an interpretation method to observe behavior across pollution categories. Moreover, Basith *et al.* (2022) used mechanistic plausibility from biomedical synthesis to support particular exposure patterns in subsequent health contexts without claiming the exposure model itself to be causal.

The literature review of previous works showed two gaps in knowledge. First, the generalization under the station held-out validation design was not consistent because most authors relied on random splits, with few paying attention to the station's sparse areas. Second, model output uncertainty is not consistently showcased as a map product, with few authors producing an uncertainty surface grounded in held-out error behavior.

We endeavor to address these gaps in knowledge in different ways. Our study will adopt GroupKFold by station, which is station -holdout generalization to enable out-of-fold (OOF) samples to represent unseen locations. Moreover, the distance-stratified OOF error will be translated to an uncertainty surface grounded in evaluation, for instance, root mean square (RMSE) by distance.

3. Methodology

3.1 Study Area

The Greater Taipei Area is home to about 7 million people, according to the 2019 census. It is made up of Taipei, Keelung, and New Taipei cities, boasting an area of 2457km². This makes it one of the most densely populated areas in the country, at approximately 3000 people per square kilometer (New Taipei City Government, 2017).

New Taipei area has 120km of coastline

replenished by huge rivers like Xindian and Keelung, with terrain features like hills and mountains. The city experiences a warm and humid climate with high rainfall caused by typhoons and monsoons. New Taipei is the most populous municipality with 29 districts led by a mayor and an elected city council as the legislature. Major economic activities are tech and business firms and tourism, for instance, Yingge ceramics (New Taipei City Government, 2017).

Keelung, a natural deep-water harbor, is situated on the northeast coast of Taiwan. Keelung faces the Pacific Ocean, surrounded by offshore islets like Heping and Keelung Islets, and hills and mountains. The climate of the area is maritime with a wet winter influenced by the northeasterly monsoon. The city has 7 districts with multiple departments, like finance and public works. Economically, Keelung, being a port city, logistics and transport depends on the harbor (Keelung City Government, 2022).

Taipei City is located at the northeast tip of Taipei Basin, bordered by New Taipei on all sides. The terrain has the basin floor, overthrust ridges, and the Datun volcanic group, with Keelung, Tamsui, and Xindian rivers shaping the urban lowlands. The city experiences a subtropical monsoon climate with hot summers and mild winters characterized by heavy rainfall and seasonal typhoons. The city is the country's capital and functions as the economic, political, cultural, and financial center. As a municipality, it is divided into 12 districts led by a mayor and an elected council as the legislature (Taipei City Government, 2025).

The Greater Taipei Area has 20 observation stations under the custodianship of the Ministry of Environment (MoENV), with the majority located in the northern part and none in the South, as shown in Figure 1. Therefore, Taoyuan City and Yilan County were included within the study domain to incorporate observations from 8 MoENV stations for spatiotemporal PM_{2.5} mapping, as shown in Figure 2. All model input data were masked to the study area shown in Figure 2. The resulting PM_{2.5} maps were subsequently clipped to the Greater Taipei Area, and only results within this final area were reported, excluding Taoyuan City and Yilan County.

3.2 Data Collection

The choice of dataset was influenced by a review of previous works, as discussed above. The dataset contains static and dynamic variables, with the former remaining stable (changing over long periods of time) while the latter changing over short periods of time. The former includes population density, land use, slope, aspect, building height, building footprint, and road network, while the latter consists of satellite AOD, wind speed, temperature, precipitation, air pressure, relative humidity, and surface-level PM_{2.5}, with its data

collection (monthly) from July 2015 to December 2023. However, only data from July 2015 to December 2022 were used for model development, with the 2023 data input channels being used exclusively to generate the station-assisted maps described in Section 3.8 and not used in the model training.

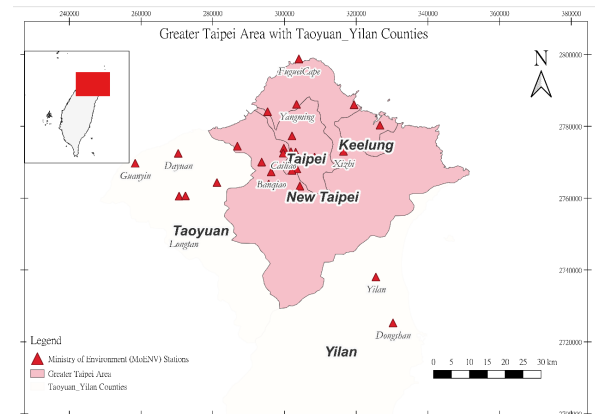


Figure 1 The study Area showing Ministry of Environment (formerly Environmental Protection Agency) stations

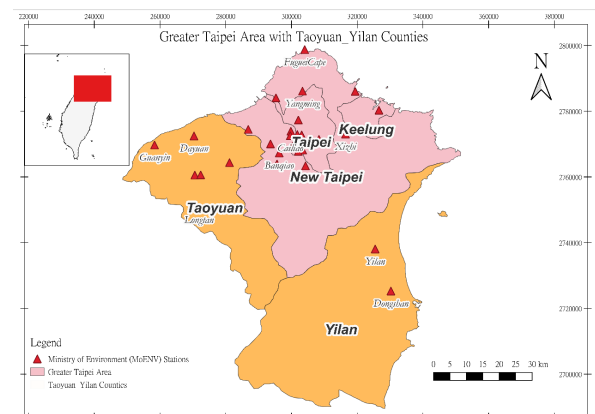


Figure 2 The study Area (including Taoyuan & Yilan) showing Ministry of Environment (formerly Environmental Protection Agency) stations

Satellite AOD is a valuable source of satellite-observed pollutants, which is a proxy for PM_{2.5} to aid the model in learning spatiotemporal patterns. The Himawari product was chosen due to its rich temporal resolution (10 minutes), and a tradeoff in spatial resolution (5km) was made. The downloaded (daily) product was then aggregated into a monthly product and masked and clipped to the study area (including Taoyuan & Yilan areas).

The AOT daily valid count was obtained from aggregating satellite AOD to a monthly product to represent the days contributing to each pixel's monthly AOD, with a higher count indicating that the monthly AOD product is reliable. The AOD daily valid count, ranging from 0 to 31 days, was retained in its original scale rather than standardized to maintain the direct

interpretation (mentioned in beginning of this paragraph).

Meteorological parameters (wind speed, temperature, precipitation, air pressure, and relative humidity) generally influence the formation, loading, transportation, and dispersion of pollutant particles in the atmosphere.

Population density is a vital predictor of PM_{2.5} concentration since it is an indicator of human activities generating emissions, for instance, industrial and commercial activity.

Different land use contexts exhibit distinct emission patterns, hence influencing PM_{2.5} concentration. For instance, agricultural lands are associated with biomass burning and dust.

Slope and aspect influence PM_{2.5} concentration, especially in areas with complex topography. For instance, slope affects how air and aerosol particles traverse the terrain, while aspect influences moisture conditions and solar radiation exposure, depending on the slope direction.

Building height and footprint are urban grain parameters that showcase the built environment conditions that affect emission concentration, airflow obstruction, and pollutant dispersion, hence influencing PM_{2.5} concentration.

Road networks (roads and railways) are associated with suspended dust and exhaust fumes from vehicles, hence influencing PM_{2.5} concentration.

Surface-level PM_{2.5} observations from ground stations served as the dependent target variable for supervised model development, whereas the other 26 variables served as independent predictor inputs. The 2023 PM_{2.5} observations were not used to train the deep-learning model but only for station-level tuning and anchoring of the final station-assisted maps, as described in Section 3.8.

Positional embeddings and temporal encodings were developed by the author for different purposes. The former was to provide the spatial location of each grid cell, enabling the proposed model to learn geographic patterns and location-dependent PM_{2.5} variations. The latter was to represent cyclic temporal patterns, such as seasonality, while preserving continuity between periods, for instance, December and January being treated as temporally close.

3.3 Data Preprocessing

Resampling techniques were used to harmonize the multisource data to a 1km EPSG3826 grid (and monthly temporal resolution). Bilinear interpolation was used to resample continuous variables (satellite AOD, meteorological parameters, slope, aspect, building height, and footprint) to produce smooth transitions. AOT valid daily count was resampled using

nearest neighbor to avoid creating artificial validity values.

The static and dynamic data channels were examined to check their distributions, including mean, minimum, maximum values, standard deviation, and percentiles. Dynamic variables were also checked to ensure that no monthly observations were missing. Based on this analysis, a scaling profile was developed to normalize the data inputs and prevent variables with large magnitudes from dominating model training. Satellite AOD, air pressure, air temperature, relative humidity, and slope were standardized using z-scores. Building height, population density, precipitation, road-rail density, and wind speed were transformed using log1p followed by z-score normalization. AOD daily valid count, aspect sine and cosine, building footprint, and land-use were retained in their original form since they were already standardized; for instance, aspect values ranged from -1 to 1, while that of building footprint fractional coverage ranged from 0 to 1. The scaling parameters were derived exclusively from 72 months (from July 2015 to June 2021) and subsequently applied, without refitting, to the other monthly time steps (till December 2022) and 2023 datasets, as will be discussed in the following subsections.

The data channels were stacked into a 147 x 113 x 90 x 26 THWC tensor representing grid height, width, 90 monthly time steps, and 26 channels. A channel-specific monthly mask identified valid and missing entries, retaining months with at least 24 channels of valid data while the masked values were left in the tensor with zero effective weight. The relaxed masking threshold was performed to prevent otherwise usable months from being discarded and limit the propagation of zero-weight and missing samples. The resulting tensor was then ready for model input.

3.4 Flowchart

The selection of the CNN-Bi-LSTM autoencoder, TCN, and Bi-LSTM was informed by a review of related work (Li *et al.*, 2022; Wei *et al.*, 2023; Hu *et al.*, 2024; Li *et al.*, 2025; Wang *et al.*, 2025). The specific model configurations, integration strategy, calibration workflow, and station-assisted mapping procedure were developed and adapted for the present study.

A CNN – Bi-LSTM autoencoder was trained on the first 72 months and validated on the subsequent 18 months using a masked Huber reconstruction loss. The autoencoder checkpoint with the lowest masked validation loss was selected, achieving a reconstruction performance of $R^2 = 0.9687$, RMSE = 0.1589, mean absolute error (MAE) = 0.0993, and mean bias error (MBE) = 0.0382. During training, the decoder reconstructed the initial 26 channels, and at the same time, the encoder learned a 64-dimensional

representation for each month pixel observation, with the selected encoder (with the lowest validation loss) applied to all the 90-time steps to produce a 90 x 64 latent sequence for each pixel for subsequent PM_{2.5} modelling. The pretrained encoder was then fine-tuned (for 300 epochs) using (90 months) surface-level PM_{2.5} observations (at station pixels) to enable learning features more directly related to PM_{2.5} variability, and its (encoder) weights were frozen and used as a fixed feature extractor. Each (monthly) time step was weighted according to predictor availability (see Table 1), PM_{2.5} availability, and monthly observation coverage raised to 0.58, with invalid and missing months receiving zero weight to prevent some months from dominating. The CNN – Bi-LSTM autoencoder specifications and hyperparameters are reported in the appendix.

Training samples for the TCN – Bi-LSTM (proposed) model were generated from the MoENV stations' core grid cells and nearby jittered cells. Temporal encodings (6) and 2 spatial offsets (dx and dy) were appended to each 64-dimensional latent vector. The 6 temporal encodings were: $\sin(2\pi m/12)$, $\cos(2\pi m/12)$, $\sin(4\pi m/12)$, $\cos(4\pi m/12)$, $\sin(\pi \cdot y_norm)$, and $\cos(\pi \cdot y_norm)$, where m is the month-of-year and y_norm is the normalized year position across the study period. The temporal cues represented the annual cycle, second seasonal harmonic, and normalized interannual position, respectively, using sine–cosine pairs, while dx and dy described each sampled pixel's displacement from its source station. The variables (6 temporal encodings and 2 spatial offsets) were not inadvertently entered twice. The original positional and temporal encodings (see Table 1) were included among the 26 channels processed by the autoencoder and compressed into the 64-dimensional latent representation. The six temporal cues were then explicitly appended to the latent features to preserve

clear seasonal and long-term temporal information for the prediction heads. Likewise, dx and dy represented relative offsets of the jittered samples from their monitoring stations and were therefore distinct from the absolute x and y coordinate embeddings. This produced a final model input sequence of 90×72.

The TCN and Bi-LSTM regression heads used the derived latent-sequence samples and were trained using station-grouped GroupKFold cross-validation. Entire monitoring stations were withheld in each fold, and both models were optimized using a weighted masked Huber loss. The resulting station-held-out out-of-fold predictions from the two heads were linearly blended using weights estimated exclusively from the OOF results. Calibration was learned from the blended station-held-out OOF predictions after aggregating jittered samples to the station–month level. Candidate station and cluster-based bias or affine corrections were retained only when they improved RMSE without worsening absolute bias, followed by a small global bias adjustment to remove residual drift.

For 2023, rolling 90-month input windows were processed through the fine-tuned PM-aware encoder, after which the blended TCN–Bi-LSTM heads generated raw station and grid-level PM_{2.5} estimates. Station predictions were tuned using the best-performing calibration function against 2023 observations to define station anchors. Monthly maps were then produced by interpolating the anchor-based correction deltas with KNN–IDW using up to three stations within 120 km, adding these corrections to the raw pixel estimates, and fixing station-core pixels to the tuned anchors. The resulting 1km GeoTIFF maps were finally clipped to the Greater Taipei Area. The following sections of the methodology will further explore the proposed model training, calibration, and station-assisted mapping.

Table 1 Dataset used for the proposed model input, with their original resolutions

Variables	No. of variables	Resolution (spatial/temporal)	Source
Satellite AOD: Himawari L3 Monthly(500nm)	1	5km/Daily	Himawari 8
AOT daily valid count	1	NA/NA	
Precipitation, Relative Humidity, Temperature, Atmospheric pressure,	4	2.5km/Monthly	CWA
Wind speed	1	2km/ Daily	TCCIP
Population Density	1	1km/NA	WorldPop Hub
Land Use	6	10m/NA	ESA WorldCover 2021
Slope/Aspect (cos & sin)	3	30m/NA	ALOS DEM
Building height	1	100m/NA	GHSL of EU-JRC
Building footprint	1	10m/NA	GHSL of EU-JRC
Road network	1	NA/NA (rasterized)	NLSC
Surface PM _{2.5}	1	NA/Hourly	MoENV
Positional embeddings (x & y)	2	NA/NA	
Temporal encodings: cos/sin (Month & Year)	4	NA/NA	

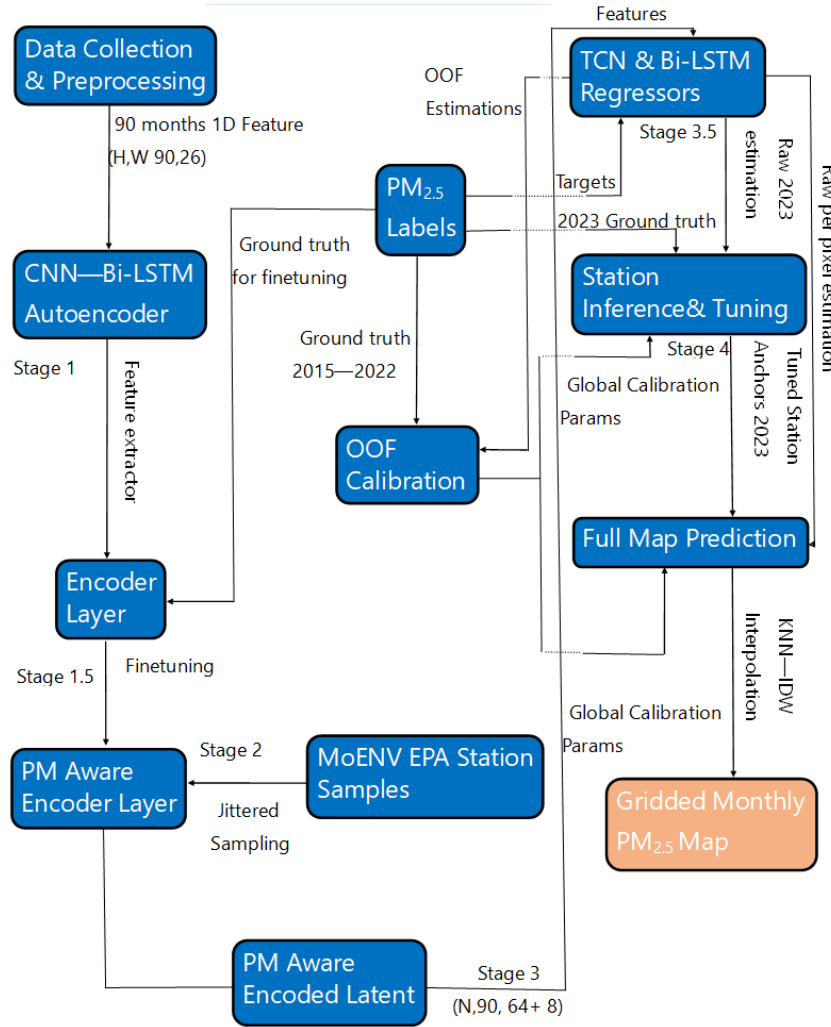


Figure 3 Overall workflow for PM_{2.5} estimation and 2023 station-assisted spatial mapping

3.5 Model Architecture

The proposed model comprises parallel dual heads – TCN and Bi-LSTM, as shown in Figure 4. The architecture was selected based on the complementary strengths of the two models and their demonstrated effectiveness in related studies (as discussed earlier). The TCN head captures local and multi-scale temporal patterns using causal convolutions, whereas the Bi-LSTM captures longer-range dependencies in both temporal directions.

The TCN head is a causal, dilated 1D convolution block in residual stacks. Two cycles of four dilated residual blocks with dilations, $d \in \{1, 2, 4, 8\}$ were utilized. Each of these blocks contains two Conv1D layers (kernel size, $k = 3$ and 64 filters) with ReLU, Dropout, Layer Normalization, and a 1×1 skip projection. The receptive field, R , is represented by equation (1).

$$R = 1 + (k - 1) \sum_{l=1}^L d_l \quad \text{..... (1)}$$

where the sum includes all the Conv1D layers in the

stack. L stands for the total number of all layers, while d is the dilation rate.

$$\begin{aligned} \sum_{l=1}^L d_l &= 2(1 + 2 + 4 + 8) \times 2 = 60 \\ \Rightarrow R &= 1 + (3 - 1) \cdot 60 = 121 \end{aligned}$$

This enables the head to see 121-time steps in one forward pass with no future leakage and, therefore, can cover our historical period, 90 months, with ease.

The Bi-LSTM head is made up of two layers with 64 and 32 units of direction, respectively. For a monthly sequence, $X = \{x_t\}_{t=1}^T$, with $T = 90$, $x_t \in R^{72}$, per each time step, the PM_{2.5} concentration is mapped by stacking the two bidirectional layers. For the l^{th} Bi-LSTM layer,

$$h_{t,\text{fwd}}^{(1)} = \text{LSTM}_{\text{fwd}}^{(1)}(z_t^{(l-1)}, h_{t-1,\text{fwd}}^{(1)}, c_{t-1,\text{fwd}}^{(1)}) \quad \text{..... (2)}$$

$$h_{t,\text{bwd}}^{(1)} = \text{LSTM}_{\text{bwd}}^{(1)}(z_t^{(l-1)}, h_{t+1,\text{bwd}}^{(1)}, c_{t+1,\text{bwd}}^{(1)}) \quad \text{..... (3)}$$

(2) and (3) concatenated to obtain (4);

$$h_t^{(1)} = [h_{t,\text{fwd}}^{(1)}; h_{t,\text{bwd}}^{(1)}] \dots\dots\dots (4)$$

where $z_t^{(1)} = x_t$, $z_t^{(2)} = h_t^{(1)}$. After the second Bi-LSTM, layer normalization and dropout are applied, obtaining the normalized state $\hat{h}_t = \text{Dropout}(\text{Layer Norm}(h_t^{(2)}))$

$$\hat{y}_t = W_o \hat{h}_t + b_o, \quad \hat{y}_t \in R \dots\dots\dots (5)$$

where W_o is the readout weight of the Dense (1) layer while b_o is bias.

3.6 Model Training

Each 90×72 latent-sequence sample was screened before training. A sample-month was considered valid when at least 24 data channels and a $\text{PM}_{2.5}$ observation were available; invalid jitter-station associations and zero-weight samples were removed. Its effective weight was represented as

$$w_{i,t} = 1_{\text{valid}_{i,t}} \left(\frac{\text{hour}_{\text{coverage}_{i,t}}}{100} \right)^{0.58} s_i \dots\dots\dots (6)$$

where $1_{\text{valid}_{i,t}}$ is a time mask while $\left(\frac{\text{hour}_{\text{coverage}_{i,t}}}{100} \right)^{0.58}$ refers to coverage weight and s_i is spatial weight. The samples and time steps (month, in this case) were represented by i and t respectively. Jitter sampling utilized a ring-based spatial weight that decreased with distance from the source station ($r \leq 2$ and a guarded $r=3$ for Yangming station) to ensure well-observed months and pixels nearer the station contributed more to the effective weight. The sampling radius r represents ± 2 pixels (approximately ± 2 km on the $1 \text{ R}^2 \text{ km}$ grid employed).

The dual heads were optimized by a weighted Masked Huber Loss ($\delta = 0.6$) across valid samples and time steps

$$L = \frac{\sum_{i,t} w_{i,t} \rho_{\delta}(\hat{y}_{i,t} - y_{i,t})}{\sum_{i,t} w_{i,t} + \epsilon} \dots\dots\dots (7)$$

$$\rho_{\delta}(e) = \begin{cases} 1/2e^2, & |e| \leq \delta \\ \delta(|e| - 1/2\delta), & |e| > \delta \end{cases} \dots\dots\dots (8)$$

where $\rho_{\delta}(e)$ is the Huber penalty applied to an individual prediction residual, with a cutoff δ , ϵ is a small numerical constant to prevent division by zero, $\hat{y}_{i,t}$ is the model estimation while $y_{i,t}$ are the observations.

TCN & Bi-LSTM heads were trained using *Adam optimizer* with a learning rate of $1e-3$. *L2 regularization* was $2e-4$, batch size 64, and a maximum of 200 epochs for OOF training (and up to 300 epochs for final heads refitting on all samples). *Early Stopping* was employed with (patience =8, restore best) and *ReduceLROnPlateau* ($x0.5$, patience =3, min lr= $1e-5$).

Computations ran on float32 with fixed seeds for reproducibility, and during cross-validation, RMSE, MAE, MBE, and R^2 were logged on station-held-out valid months.

Model training and pre-calibration evaluation used 10-fold station-grouped cross-validation across 27 stations. 28 stations were initially considered (20 in the Greater Taipei Area plus 8 in Taoyuan and Yilan), however, Cape Fuguei was excluded, leaving 27 stations for the GroupKFold cross-validation. This exclusion was due to the sample-weighting issue identified at the beginning of model training, which made Cape Fuguei unsuitable for inclusion in the weighted training and evaluation framework. All samples associated with a station, including its jittered neighboring pixels, were assigned to the same fold, so that 2–3 complete stations were withheld in each iteration. For every fold, new TCN and Bi-LSTM heads were trained on the remaining stations using the weighted masked Huber loss, with Adam optimization, a batch size of 64, and a maximum of 200 epochs subject to early stopping and learning-rate reduction. The trained heads then generated predictions only for the withheld stations, and these fold-wise outputs were combined to form complete station-held-out OOF predictions for all samples, as shown

$$\begin{aligned} \hat{y}_t^{\text{ens}} &= w_{\text{TCNs},m} \hat{y}_t^{\text{TCN}} + (1 - w_{\text{TCNs},m}) \hat{y}_t^{\text{BiLSTM}}, \\ 0 &\leq w_{\text{TCNs},m} \leq 1 \dots\dots\dots (9) \end{aligned}$$

where $\hat{y}_t^{\text{TCN}}$ is the TCN OOF prediction, $\hat{y}_t^{\text{BiLSTM}}$ is the Bi-LSTM OOF prediction, i denotes a valid OOF sample month prediction, $w_{\text{TCNs},m}$ represents the TCN ensemble weight assigned for station, s and calendar month, m and $1 - w_{\text{TCNs},m}$ is the Bi-LSTM ensemble weight.

Station-calendar-month-specific TCN weights, $w_{\text{TCN},s,m}$, were estimated from the corresponding historical OOF prediction-observation pairs using constrained least squares. A global TCN weight, $w_{\text{TCN},\text{global}}$, was additionally estimated from all valid OOF pairs and used as a fallback when insufficient information was available for a station-month-specific weight.

$$w_{\text{TCN},\text{global}}^* = \text{clip}_{[0,1]} \left[\frac{\sum_i (y_i^{\text{TCN}} - y_i^{\text{BiLSTM}})(y_i - y_i^{\text{BiLSTM}})}{\sum_i (y_i^{\text{TCN}} - y_i^{\text{BiLSTM}})^2 + \epsilon} \right] \dots\dots\dots (10)$$

Where $w_{\text{TCN},\text{global}}^*$ represents the optimal global TCN ensemble weight derived from the historical OOF predictions, i indexes all valid historical OOF prediction-observation pairs, $\text{clip}_{[0,1]}$ constrains the weight to the interval $[0,1]$ and ϵ prevents division by zero.

The resulting raw ensemble predictions were evaluated at both the sample-month level and, after weighted aggregation of jittered samples, the station-

month level. These results represent the predictive performance of the TCN–Bi-LSTM ensemble before calibration and specifically assess spatial generalization to unseen station locations rather than strict future time forecasting.

3.7 Calibration

The raw OOF ensemble predictions were first aggregated from the jittered sample-month level to the station-month level using the corresponding coverage and spatial weights. Stations were grouped into three clusters using six standardized descriptors: the observed PM_{2.5} mean, standard deviation, seasonal amplitude, winter–summer difference, and two principal components derived from the station-associated latent features. Cluster-wise monthly affine parameters were subsequently estimated using ridge regression:

$$\widehat{y}_{s,t}^{\text{cal}} = a_{k,m} \widehat{y}_{s,t}^{\text{raw}} + b_{k,m} \dots \dots \dots (11)$$

Where s is the station, t *monthly step*, k denotes the station cluster and m is the calendar month corresponding to t . Residual winsorization, slope clipping, and shrinkage toward a unit slope were used to limit the influence of extreme residuals and prevent excessive correction.

For each station, five calibration alternatives were evaluated: identity, constant-bias, monthly-bias, shrink-affine, and cluster-wise monthly affine correction. Identity retained the raw prediction unchanged. A non-identity station-specific calibrator was considered only when at least 12 valid OOF months were available, except for the cluster-based correction, which could also support stations with fewer observations. The correction was retained only when it reduced RMSE by at least 8% and did not increase the absolute MBE. Among eligible alternatives, the calibrator with the lowest RMSE was selected, with absolute MBE used as the tie-breaker. The rationale for the choice of calibration functions is discussed in the appendix.

The selected station-level corrections were applied to the OOF predictions, after which a small global bias anchor was calculated from the remaining mean residual

$$B_{\text{global}} = -\frac{1}{N} \sum_{i=1}^N (\widehat{y}_i^{\text{cal},0} - y_i) \dots \dots \dots (12)$$

where B_{global} represents the global bias, N is the number of valid OOF station-month prediction-observation pairs, $\widehat{y}_i^{\text{cal},0}$ represents the calibrated PM_{2.5} value before global bias and y_i as surface-level PM_{2.5} value (observed). The final calibrated estimate was obtained as follows

$$\widehat{y}_i^{\text{cal}} = \widehat{y}_i^{\text{cal},0} + B_{\text{global}} \dots \dots \dots (13)$$

where $\widehat{y}_i^{\text{cal},0}$ represents the calibrated PM_{2.5} value before global bias, and $\widehat{y}_i^{\text{cal}}$ represents the final calibrated PM_{2.5} value after global bias. The ensemble weight, station assignments, cluster-specific monthly parameters, selected station calibrators, and global bias were then saved and frozen.

After calibration had been derived from the OOF predictions, new final TCN and Bi-LSTM heads were trained from scratch using all available samples for up to 300 epochs, with early stopping and learning-rate reduction based on training loss. Their predictions were blended using the frozen OOF-derived station-month-specific TCN weights, with the global TCN weight used as a fallback, after which the frozen calibration rules were applied without further fitting to produce final calibrated estimates. Thus, calibration was **not incorporated into the neural-network training loss**; it was applied as a separate post-processing stage to the final blended predictions. Therefore, these results are calibration-adjusted OOF performance (and not a full independent validation outcome) since calibration functions were fitted and selected using the same OOF prediction-observation pairs used for evaluation.

3.8 Station-Assisted Mapping

The twelve-monthly data channel layers for 2023 were first stacked into a four-dimensional tensor of shape (12, H, W, C). Its channels were reordered to match the canonical channel sequence used during model development. Each channel was transformed using the scaling rule fitted from the 72-month (July 2015 – June 2021) training period used in the CNN–Bi-LSTM autoencoder, including *raw*, *z-score*, or *loglp-z-score* scaling. The corresponding means and standard deviations (in the 72 months) were reused unchanged, and invalid values were masked and replaced with zero only after scaling. Consequently, no 2023 observations were used to estimate preprocessing parameters, avoiding scaling leakage.

The scaled 2023 tensor was then combined progressively with the historical 90-month tensor to construct a rolling 90-month sequence for each 2023 month. For instance, January used the latest 89 historical months, including January 2023, whereas December used the latest 78 historical months, including all 12 months of 2023. For each station and jittered sampling location, the 90-month rolling window was passed through the frozen encoder to obtain a 90×64 latent sequence. Six temporal cues and the corresponding jitter offsets (discussed in 3.4 *Flowchart*) were then appended, producing a 90×72 input required by the final TCN and Bi-LSTM heads.

The final heads, which had been refitted on all available historical samples, generated one prediction for the final timestep of each rolling window. Their

outputs were combined to obtain the 2023 pre-calibration model estimates:

$$\widehat{y_{s,m}^{\text{estimate}}} = w_{s,m}^{\text{TCN}} \widehat{y_{s,m}^{\text{TCN}}} + (1 - w_{s,m}^{\text{TCN}}) \widehat{y_{s,m}^{\text{BiLSTM}}} \dots\dots\dots (14)$$

The $w_{\text{TCN},s,m}$, derived from historical OOF predictions was applied to each station-associated sample. The frozen $w_{\text{TCN},\text{global}}$, was used as the fallback when corresponding station-month weights were unavailable. No ensemble weights were estimated using the 2023 observations. The resulting jitter-level estimates were finally averaged to produce one pre-calibrated monthly estimate for each station.

This procedure should not be interpreted as future forecasting because the estimation data for the target month were included in the corresponding 90-month sequence. However, the 2023 surface-level PM_{2.5} observations were not introduced to the 2023 encoder or the dual heads during this stage, but in the subsequent 2023 calibration.

The pre-calibrated monthly estimates were compared with the 2023 surface-level PM_{2.5} observations, and several calibration functions were tested as tuning alternatives for each station. The function producing the highest R² score was chosen, and its adjusted values became the tuned 2023 station estimates. Accordingly, the tuned station estimates were model-derived values adjusted using the 2023 observations; they were not the observations themselves. These calibration functions included exponential moving (EMA) bias, affine EMA, EMA affine blend, change-point bias, change-point affine, and hybrid correction, with the rationale and their corresponding equations described in the appendix. Each station correction was clipped using the station-specific adjustment limit K_s associated with its selected 2023 calibration function; the tested limits did not exceed $5\mu\text{g}/\text{m}^3$.

For each month of 2023, the corresponding 90-month rolling predictor window was applied across every valid grid cell. The frozen encoder, Bi-LSTM, and TCN heads were used to generate an initial PM_{2.5} estimate at each pixel, with its horizontal and vertical offset features set to zero because grid pixels have no jitter displacement. The two regression outputs were combined using the global OOF-derived TCN weight. At each grid cell, the TCN and Bi-LSTM outputs were combined using the frozen global OOF-derived TCN weight, $w_{\text{TCN},\text{global}}$, because station-month-specific weights were not applicable to non-station grid cells. Station-specific corrections were then calculated from the difference between the tuned station estimates and their corresponding pre-calibration model estimates. Each station correction limit was constrained using the correction limit selected earlier.

The station corrections were spatially propagated across the study grid using (KNN – IDW), with each grid cell, up to the three nearest stations within a 120-km search radius, being selected. Their clipped

corrections were weighted using inverse squared distance, and the interpolated correction field was calculated as

$$\Delta_{g,m}^{\text{IDW}} = \left(\sum_{s \in \mathcal{N}_{g,m}} w_{g,s} \Delta_{s,m}^{\text{clipped}} \right) \div \left(\sum_{s \in \mathcal{N}_{g,m}} w_{g,s} \right) \dots (15)$$

where g denotes the target grid cell, m denotes the 2023 month, s is the station anchor, $\mathcal{N}_{g,m}$ represents the three nearest stations with valid corrections in the search radius, and $w_{g,s}$ represents the inverse-squared distance assigned to stations s the relative to the grid cell g . The final station-assisted grid estimate was then obtained by adding the interpolated correction to the pre-calibrated model estimate

$$\widehat{y_{g,m}^{\text{final}}} = \widehat{y_{g,m}^{\text{initial}}} + \Delta_{g,m}^{\text{IDW}} \dots\dots\dots (16)$$

where $\widehat{y_{g,m}^{\text{final}}}$ denoted the final station-assisted grid estimate, $\widehat{y_{g,m}^{\text{initial}}}$ denotes the PM_{2.5} estimate generated by the dual heads at that grid cell, and $\Delta_{g,m}^{\text{IDW}}$ represents the KNN – IDW interpolated station correction.

Where no valid station correction was available within the search radius, the pre-calibrated model estimate was retained unchanged. A representative anchor pixel was assigned to each station using the available core or audited jitter-sampling locations. After spatial correction, the value at each anchor pixel was replaced with the corresponding tuned station estimate. This hard anchoring ensured exact agreement between the mapped and tuned station values at the station locations.

Finally, pixels outside the valid study-domain mask were assigned the NoData value. The corrected monthly grids were exported as georeferenced GeoTIFF files using the original grid dimensions, coordinate reference system, spatial transform, and resolution. A final quality-control check compared the GeoTIFF values at the anchor pixels with the tuned station estimates to confirm that the hard-anchoring procedure was correctly applied (with the difference in values being $1\text{e-}8$, which is negligible).

3.9 Model Evaluation

The model was evaluated using R² score, root mean square (RMSE), mean absolute error (MAE), and mean bias error (MBE). R² score significance is to showcase the model's spatiotemporal variance and is given by

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \widehat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \dots\dots\dots (17)$$

where y_i are observations, $\widehat{y}_i$ are predictions, $\bar{y}$ is the sample mean, and n is the number of evaluated pairs. RMSE monitors the error magnitude from the model estimations and is given by

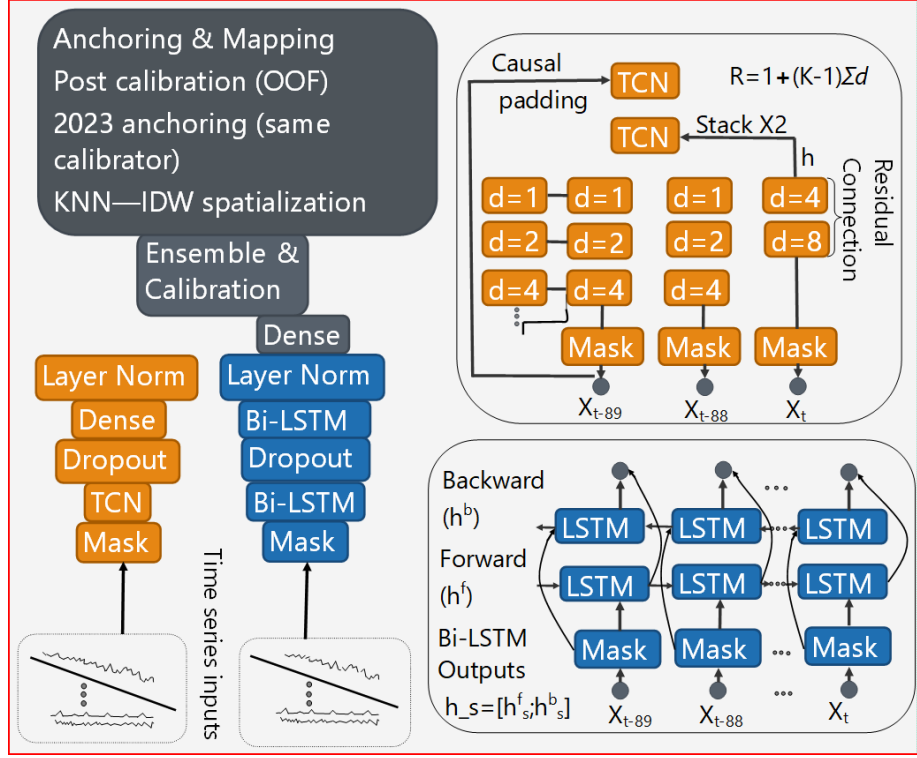


Figure 4 Architecture of the proposed model, calibration, and mapping framework, modified from Wang *et al.* (2025)

$$MSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \dots \dots \dots (18)$$

appendix.

MAE showcases the model average absolute error and is formulated as shown

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \dots \dots \dots (19)$$

Model bias is reported by MBE using this formula with positive values signifying overestimation, and the reverse is true.

$$MBE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \dots \dots \dots (20)$$

The model validation was from OOF estimations. Here, in each fold, all samples from the held-out station group form the validation set while the rest form the training set. This achieves spatial OOF estimation without station leakage.

The proposed model was benchmarked against classical models – a convolutional neural network (CNN) and long short-term memory (LSTM). The selection of the classical models was informed by previous works (Qin *et al.*,2019; Zhang *et al.*,2021a; Lin *et al.*,2022; Yu *et al.*,2022; Li & Jiang, 2023; Wu *et al.*,2024). The specific model configurations were developed and adapted for the present study. The classical models were developed under the same conditions as the proposed model, to be fair, and then all were evaluated using the four metrics discussed above. The hyperparameters and the specifications of the classical baseline models are reported in the

4. Results

4.1 Comparison with Classical Models

The proposed model outperformed the CNN and LSTM baseline models both before and after calibration. The proposed model, after calibration, led with $R^2 = 0.7960$ and $RMSE = 2.1606 \mu g/m^3$. The calibrated CNN and LSTM models achieved R^2 values of 0.6581 and 0.6226, and $RMSE$ values of 2.7971 $\mu g/m^3$, and 2.9387 $\mu g/m^3$, respectively, as shown in Tables 2 and 3.

Table 2 Models' GroupKFold performance before calibration

Model	R^2	RMSE ($\mu g/m^3$)	MAE ($\mu g/m^3$)	MBE ($\mu g/m^3$)
CNN	0.5978	3.0339	2.3205	-0.6738
LSTM	0.5835	3.0874	2.3434	-0.5586
TCN & Bi-LSTM (proposed)	0.7475	2.4035	1.8144	-0.0227

Table 3 Models' GroupKFold performance after calibration

Model	R^2	RMSE ($\mu g/m^3$)	MAE ($\mu g/m^3$)	MBE ($\mu g/m^3$)
CNN	0.6581	2.7971	2.1741	0.0134
LSTM	0.6226	2.9387	2.2696	-0.0102
TCN & Bi-LSTM (proposed)	0.7960	2.1606	1.6401	0.1428

Calibration improved the agreement between the OOF estimates and station observations by reducing systematic bias and overall error. The proposed model's R^2 score increased from 0.7475 to 0.7960, while RMSE decreased from $2.4035 \mu\text{g}/\text{m}^3$ to $2.1606 \mu\text{g}/\text{m}^3$, with comparable improvements observed for the CNN and LSTM models. The shift of MBE toward zero indicates that the systematic underestimation present in the pre-calibrated estimates was substantially reduced, while RMSE and MAE describe the remaining overall error magnitude.

Figures 5–7 visualize the calibrated results reported in Table 3 (rounded off to 3 decimal places), with the solid blue line denoting the 1:1 reference, the dashed orange line showing the fitted regression, and the color indicating sample density.

The proposed model showed the closest agreement with the 1:1 line, achieving the highest R^2 score and lowest RMSE after calibration, followed by the CNN and LSTM models. The fitted regression lines indicate that all models retained some prediction compression at higher $\text{PM}_{2.5}$ concentrations, although this effect was least pronounced for the proposed model.

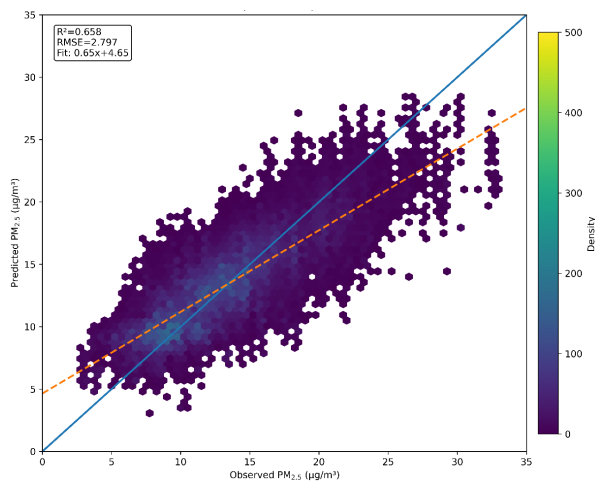


Figure 5 Observed versus calibrated OOF $\text{PM}_{2.5}$ estimates from the CNN model

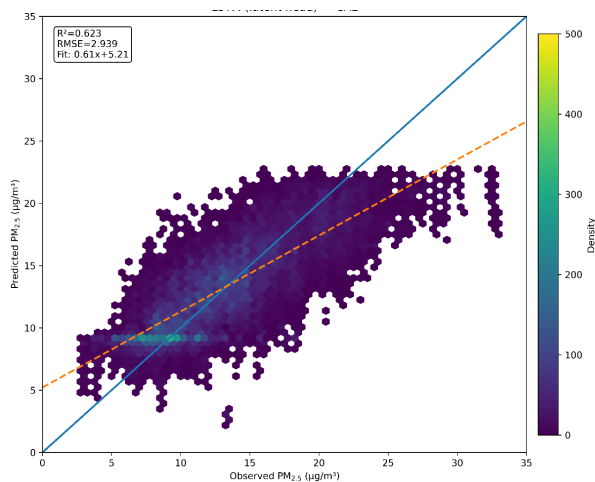


Figure 6 Observed versus calibrated OOF $\text{PM}_{2.5}$ estimates from the LSTM model

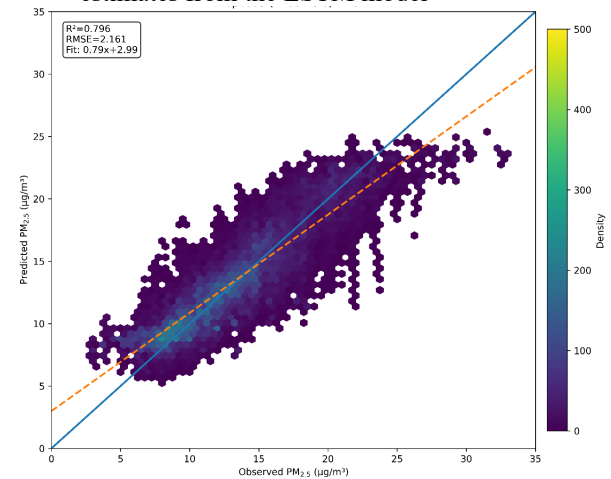


Figure 7 Observed versus calibrated OOF $\text{PM}_{2.5}$ estimates from the proposed model

4.2 $\text{PM}_{2.5}$ Concentration Maps

The calibrated monthly $\text{PM}_{2.5}$ maps, presented in Figure 8 using a common concentration scale, allow direct spatial and temporal comparison throughout 2023. $\text{PM}_{2.5}$ concentrations were generally higher from January to May, with peaks in March and April, declined from June to September, and increased slightly from October to December, as shown in Figure 9. This pattern is plausibly consistent with Taiwan's seasonal cycle, where winter and early spring commonly favor pollutant accumulation, whereas summer rainfall and stronger atmospheric dispersion reduce $\text{PM}_{2.5}$ concentrations.

The maps also reveal substantial spatial variability across the study area. Banqiao and Cailiao represent densely urbanized environments, Keelung provides a coastal reference, Xizhi represents the eastern urban corridor, and Yangming represents a higher-elevation environment. Although the five stations exhibited similar seasonal patterns, Yangming consistently showed lower concentrations plausibly due to its higher elevation.

The accompanying spatial uncertainty magnitude map in Figure 10 can be interpreted as a station-assisted support index derived from calibrated station-held-out OOF errors. Weighted RMSE and MAE values were first computed for distance bands and then assigned to each grid pixel according to its distance from the nearest station anchor. Lower values indicate stronger station-assisted support and lower expected error, whereas higher values indicate weaker support and greater expected error farther from monitoring stations. The layer is expressed in $\mu\text{g}/\text{m}^3$ and should therefore be interpreted as an empirical error-support indicator, rather than a formal probabilistic uncertainty estimate.

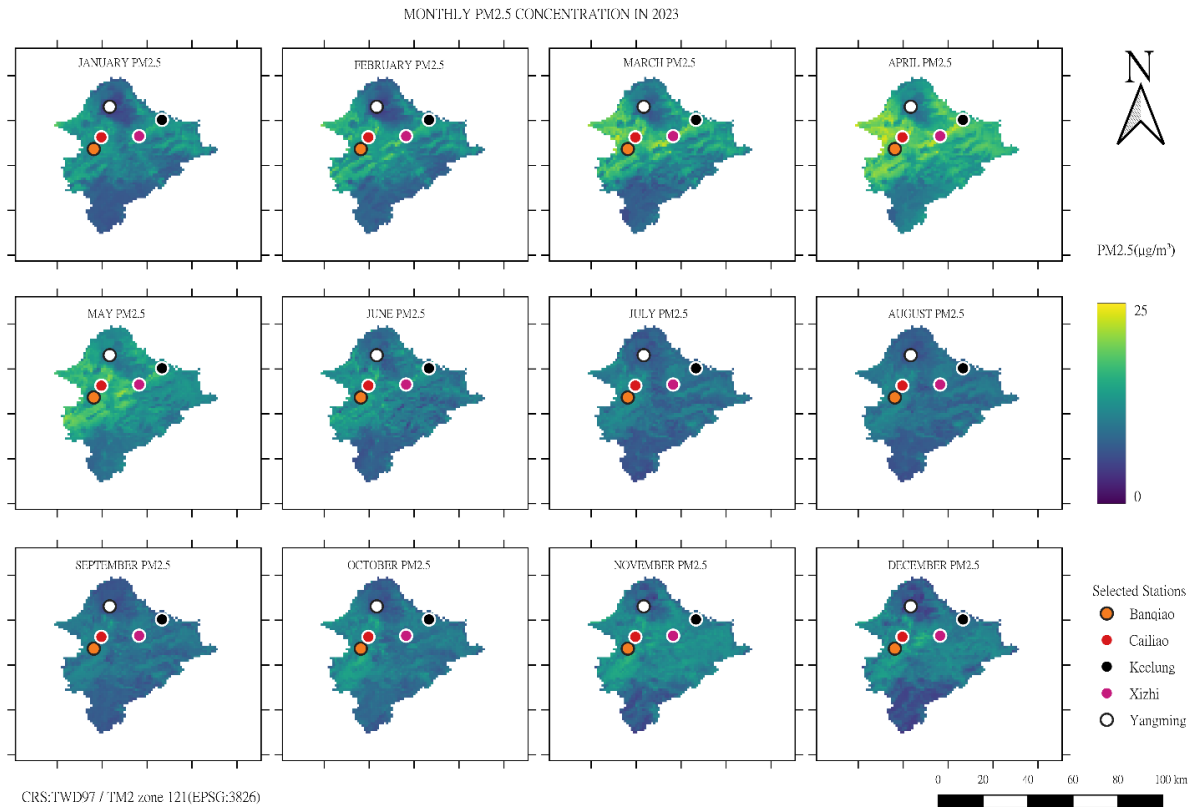


Figure 8 Calibrated monthly station-assisted PM_{2.5} concentration maps for the Greater Taipei Area in 2023

Estimated Monthly PM_{2.5} at Selected Stations (2023)

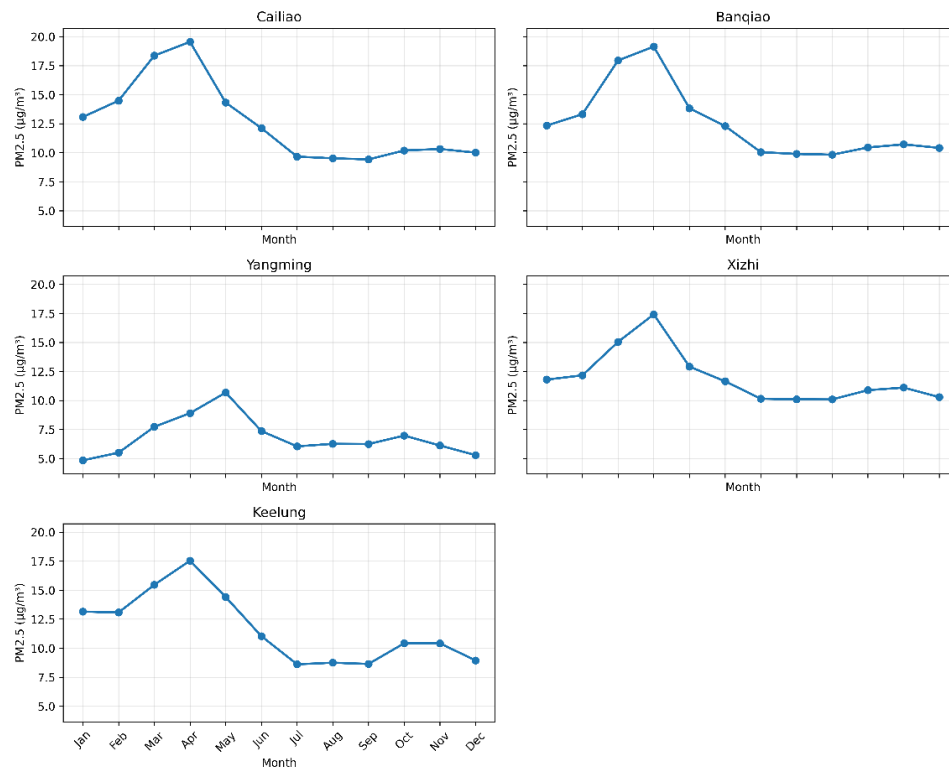


Figure 9 Monthly calibrated PM_{2.5} estimates at the five selected monitoring stations in 2023

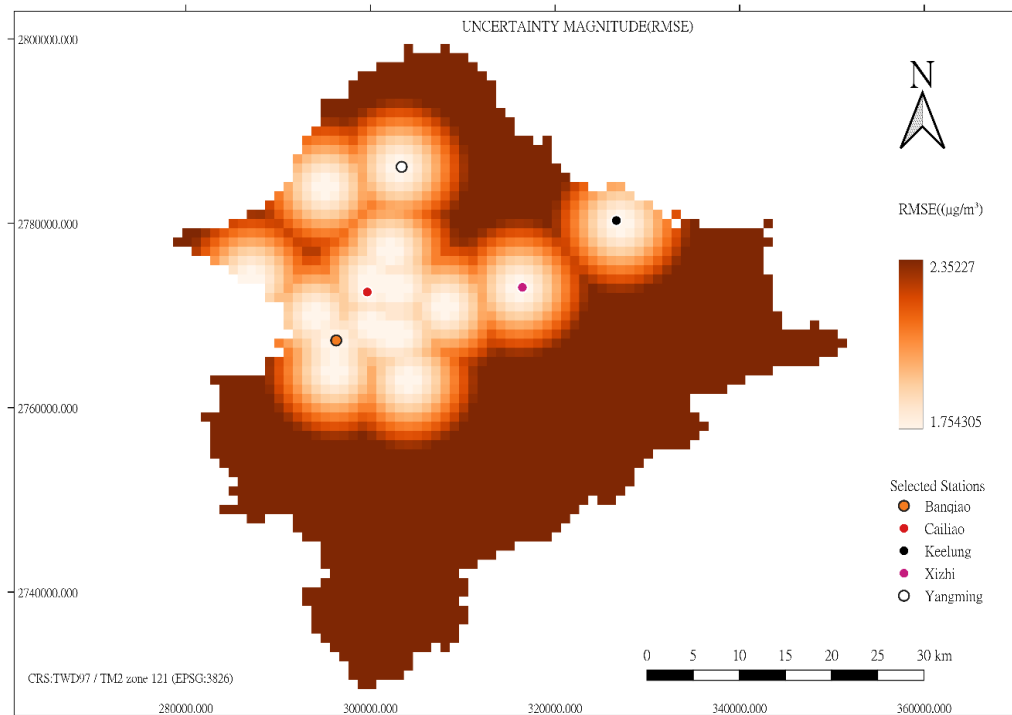


Figure 10 Empirical station-assisted support map for the 2023 PM_{2.5} product, derived from calibrated OOF RMSE and distance to the nearest station anchor

4.3 Seasonal PM_{2.5} Concentration Maps

The 2023 calibrated PM_{2.5} estimates were aggregated by season to assess whether the monthly spatiotemporal patterns persisted. The seasonal definitions commonly adopted in previous studies were employed, with winter defined as December–February, spring as March–May, summer as June–August, and autumn as September–November. As shown in Table 4, concentrations were highest in spring, followed by winter, autumn, and summer. The similar mean and median values indicate that these seasonal patterns were not driven by a small number of extreme pixels, while the higher standard deviations in spring and winter reflect greater spatial variability. Lower summer concentrations are consistent with rainfall and wet scavenging, whereas autumn indicates the gradual return of pollutant accumulation before winter.

Table 4 Pixel-level summary for the calibrated 2023 seasonal PM_{2.5} exposure maps

Season	Mean ($\mu\text{g}/\text{m}^3$)	Std deviation ($\mu\text{g}/\text{m}^3$)	Median ($\mu\text{g}/\text{m}^3$)	Valid pixels
Winter	10.489	2.651	10.840	2478
Spring	13.542	3.272	13.464	2478
Summer	9.520	1.711	9.532	2478
Autumn	10.000	1.797	10.201	2478

Although winter conditions generally favor PM_{2.5} accumulation, the elevated spring concentrations shown in Figure 12 may also reflect peculiar transport processes in northern Taiwan. The region is influenced

by long-range aerosol transport, especially biomass burning plumes from Southeast Asia, with previous studies showing Indochina-sourced aerosols reaching Taiwan from February to April. Moreover, wind speed and transboundary pollution during the winter-to-spring transitional period drive this seasonal aerosol accumulation (Lin *et al.*, 2014; Jung *et al.*, 2018; Cheng *et al.*, 2019; Chang, 2019).

Table 5 shows that calibration produced the largest positive adjustment in winter, indicating systematic underestimation in the pre-calibrated estimates. Spring required only a minor overall reduction but showed the greatest spatial variability in calibration-induced changes.

Table 5 Pixel-level summary of calibration-induced changes in the 2023 seasonal PM_{2.5} exposure maps (post-calibration minus pre-calibration)

Season	Mean Δ ($\mu\text{g}/\text{m}^3$)	Std deviation Δ ($\mu\text{g}/\text{m}^3$)	Median Δ ($\mu\text{g}/\text{m}^3$)	Valid pixels
Winter	1.665	0.593	1.820	2478
Spring	-0.160	1.088	-0.038	2478
Summer	-0.663	0.651	-0.519	2478
Autumn	-0.283	0.799	-0.456	2478

Figures 11–14 show higher seasonal PM_{2.5} concentrations across the western, central, and eastern study area, with lower levels in the north and south. Elevated western and central concentrations likely reflect urban emissions and pollutant accumulation, while eastern concentrations may be associated with transport along the urban corridor and weaker dispersion.

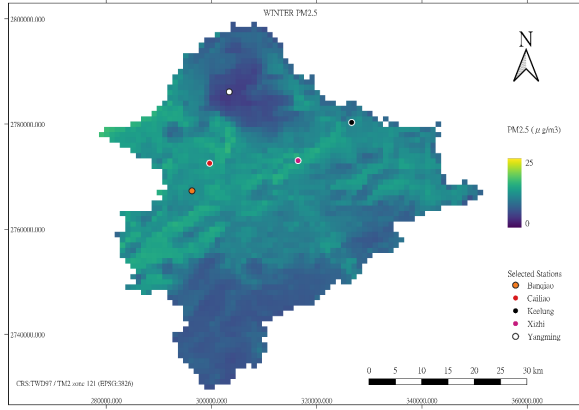


Figure 11 Calibrated winter PM_{2.5} concentration map for 2023

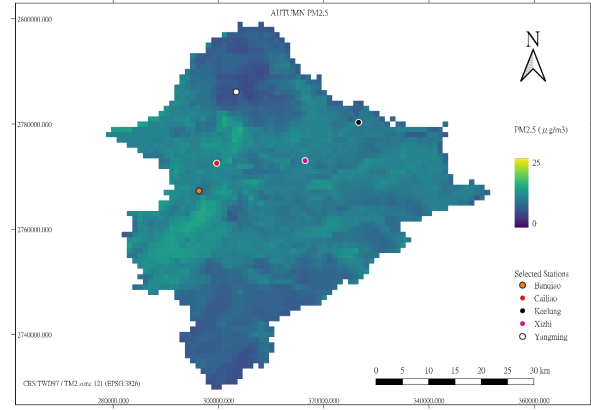


Figure 14 Calibrated autumn PM_{2.5} concentration map for 2023

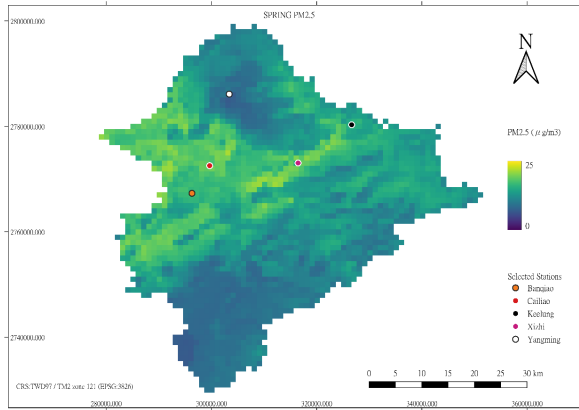


Figure 12 Calibrated spring PM_{2.5} concentration map for 2023

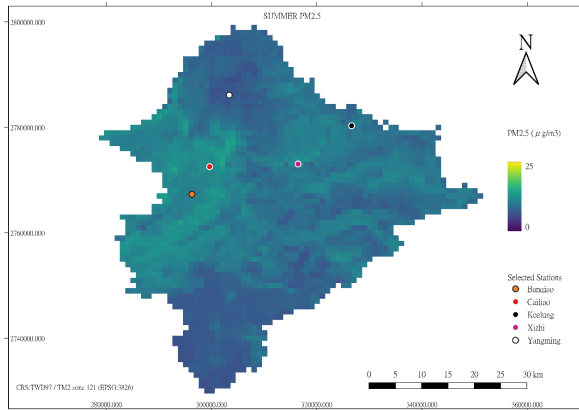


Figure 13 Calibrated summer PM_{2.5} concentration map for 2023

Tables 6 and 7 summarize the proposed model's season-wise pooled OOF performance before and after calibration.

Table 6 Seasonal pooled OOF performance of the proposed model before calibration

Season	R ²	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	MBE ($\mu\text{g}/\text{m}^3$)
Winter	0.7026	2.5557	1.9987	0.2150
Spring	0.6920	2.8779	2.1393	-0.6990
Summer	0.5250	1.8858	1.4335	0.1456
Autumn	0.6196	2.2331	1.7205	0.1924

Seasonal performance varied by metric, with winter achieving the highest R^2 , indicating the strongest variance capture, whereas summer produced the lowest RMSE and MAE, partly reflecting its lower PM_{2.5} concentration range. Spring and autumn showed intermediate performance, as shown in Table 6. Despite spring having higher PM_{2.5} concentrations, the proposed model captured winter's concentration variance more accurately, reflecting spring's spatially uneven concentration structure, which is also consistent with its elevated MBE and the patterns discussed in Tables 4 and 5. The concentration variance captured in summer and autumn was low, evidenced by the R^2 score. However, the overall reduced concentrations yielded better RMSE and MAE, indicating lower absolute errors despite weaker variance capture.

Table 7 Seasonal pooled OOF performance of the proposed model after calibration

Season	R ²	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	MBE ($\mu\text{g}/\text{m}^3$)
Winter	0.7760	2.2181	1.7433	0.3002
Spring	0.7504	2.5903	1.9538	-0.1746
Summer	0.5736	1.7868	1.3446	0.1846
Autumn	0.6916	2.0107	1.5495	0.2339

In Table 7, the seasonal performance ranking remained unchanged after calibration. Seasonal performance varied by metric, with winter achieving the highest R^2 , whereas summer produced the lowest RMSE and MAE, and the arguments in Table 6 still hold. A slight positive residual bias persisted in winter, summer, and autumn as the calibration prioritized correcting systematic bias through R^2 , RMSE, and MAE rather than centering MBE around zero, consistent with the Table 3 discussion. Overall, Tables 6 and 7 confirm that the spatiotemporal patterns observed in the monthly PM_{2.5} concentration maps persist across the aggregated seasonal map products.

5. Conclusion

This study developed a hybrid TCN–Bi-LSTM framework to estimate monthly PM_{2.5} concentrations

and produce 1km exposure maps for the Greater Taipei Area. The main objective was achieved by integrating multi-source environmental data with station observations to generate continuous monthly PM_{2.5} surfaces for 2023.

Based on pooled station-grouped OOF evaluation, the proposed model outperformed the CNN and LSTM baselines before and after calibration. After calibration, the proposed model achieved $R^2 = 0.7960$ and $RMSE = 2.1606 \mu\text{g}/\text{m}^3$, compared with $R^2 = 0.6581$ and $RMSE = 2.7971 \mu\text{g}/\text{m}^3$ for CNN and $R^2 = 0.6226$ and $RMSE = 2.9387 \mu\text{g}/\text{m}^3$ for LSTM. Calibration reduced systematic bias and improved overall agreement between the estimates and observations. These results should, however, be interpreted as calibration-adjusted OOF performance rather than fully independent station-level validation because the PM-aware encoder refinement preceded fold separation.

The 2023 maps showed coherent monthly and seasonal patterns. Concentrations were highest in spring, followed by winter, autumn, and summer. Higher PM_{2.5} levels were generally observed across the western and central urbanized areas, whereas lower levels occurred in elevated environments such as Yangming. Spring exhibited the greatest spatial variability and remained comparatively difficult to model, while winter showed the strongest seasonal variance capture. The consistency between the monthly and seasonal products indicates that the framework captured the broader temporal and spatial structure of PM_{2.5} across the study area.

The final maps combined a deep-learning background field with spatialized station-level corrections. The deep-learning model represented the wider spatiotemporal concentration structure, while KNN-IDW propagated bounded corrections from available stations and preserved the tuned values at station pixels. Therefore, the 2023 products should be interpreted as station-assisted exposure maps rather than purely independent model extrapolations. The accompanying station-assisted support layer further identifies where the mapped surface is better supported by observations and where interpretation should be more cautious.

The principal contribution of this study is a practical framework that combines station-held-out evaluation, OOF calibration, deep-learning estimation, and observation-assisted mapping under limited station coverage. The resulting exposure maps can complement existing monitoring networks and provide continuous spatial information where satellite AOD retrievals are incomplete. Nevertheless, sparse station distribution, monthly temporal resolution, limited representation of fine-scale urban heterogeneity, and dependence on 2023 station tuning constrain the generalizability of the final products. Accordingly, the maps are most appropriate as decision-support and exposure-assessment products used alongside, rather than in place of, ground-based observations.

ABBREVIATIONS

The following abbreviations are used in this article:

AOD Aerosol Optical Depth
Bi-LSTM Bidirectional Long Short-Term Memory
CNN Convolutional Neural Network
CTM Chemical Transport Model
CWA Central Weather Administration
EPA Environmental Protection Agency
EC-JRC GHSL European Commission, Joint Research Centre, Global Human Settlement Layer
FFT Fast Fourier Transform
IDW Inverse Distance Weighting
KNN k Nearest Neighbor
LSTM Long Short-Term Memory
LUR Land Use Regression
MAE Mean Absolute Error
MBE Mean Bias Error
MoENV Ministry of Environment
NLSC National Land Surveying & Mapping Centre
OOF Out-of-fold
PM Particulate Matter
TCCIP Taiwan Climate Change Projection Information & Adaptation Knowledge Platform
TCN Temporal Convolutional Network
RMSE Root Mean Square Error
WHO World Health Organization

APPENDIX A

The specifications of the models used in this research study are listed below

CNN – Bi-LSTM Autoencoder

- a) Conv1D (64,128 k=3)
- b) SE channel = 128
- c) Bi-LSTM (64 to 128, 32 to 64 'concat')
- d) Dense (128,64, residual)
- e) TimeDistributed Dense (26)
- f) Temporal SoftMax
- g) Epochs =120
- h) Batch size = 64
- i) Seed= 42
- j) Early Stopping, patience=10
- k) ReduceLROnPlateau, patience=5
- l) Loss – Masked Huber ($\delta=0.6$)
- m) Optimizer – Adam
- n) Activation – ReLu, tanh, GeLu
- o) Learning rate=0.001
- p) Weight decay = 0.0001
- q) Layer Norm = 0.000001
- r) Dropout= 0.08
- s) MHA head/dim= 4/16

APPENDIX B

The calibration functions were selected to address different forms of systematic prediction error while avoiding unnecessary post-processing. The *identity function* retained the raw estimate whenever no correction was warranted and therefore served as the baseline option. *Cluster-wise monthly affine* calibration applied month-specific slopes and intercepts estimated from groups of stations with similar temporal and latent characteristics, allowing stations with limited OOF support to benefit from a shared correction structure. *Constant bias* correction introduced a single station-specific offset to remove persistent overestimation or underestimation. *Shrunk affine* calibration adjusted both prediction scale and offset, while slope clipping and shrinkage toward unity were used to reduce instability and prevent excessive correction. *Monthly bias* calibration applied calendar-month-specific offsets to account for seasonally varying bias without directly modifying the prediction amplitude.

The 2023 station-assisted mapping used a separate set of observation-assisted station-tuning functions from those applied to the historical OOF predictions. The historical calibration was developed for the 2015–2022 station-held-out OOF setting and, therefore, used comparatively restrained functions to limit overfitting. In contrast, the 2023 procedure retrospectively adjusted monthly station estimates within the target year before spatial mapping, allowing station-specific short-term bias changes to be represented.

The candidate functions included EMA bias, which recursively updated an additive correction from recent residuals; affine EMA, which updated both slope and intercept to address changing level and scale bias; changepoint bias, which identified a shift in the within-year residual pattern and applied different additive corrections across regimes; and changepoint affine, which allowed both slope and intercept to differ between the detected regimes. A hybrid alternative selected between EMA and changepoint-bias corrections according to recent performance, while the blend combined EMA-bias and affine-EMA outputs using a rolling performance-based weight. For each station, the function producing the highest 2023 station-level (R^2) was retained, and its adjusted outputs were used as the tuned station estimates for the subsequent delta-based KNN-IDW mapping.

CNN Model

- a) Conv1D (64,32 k=5)
- b) Conv1D (16 k=3)
- c) Dense (1)
- d) Epochs =300
- e) Batch size = 32
- f) Seed= 42
- g) Early Stopping, patience=12
- h) ReduceLROnPlateau, patience=6
- i) Loss – Masked Huber (δ .=0.6)
- j) Optimizer – Adam
- k) Activation – ReLu, tanh
- l) Learning rate=0.0003
- m) Coverage power=0.58
- n) Dropout = 0.15
- o) L2 regularizer=0.00001

LSTM Model

- a) Conv1D (64,32 k=5)
- b) LSTM (64 ‘true’)
- c) TimeDistributedDense(1)
- d) Epochs =300
- e) Batch size = 32
- f) Seed=42
- g) Early Stopping, patience=12
- h) ReduceLROnPlateau, patience=6
- i) Loss – Masked Huber (δ .=0.6)
- j) Optimizer – Adam
- k) Activation – tanh
- l) Learning rate=0.0003
- m) Dropout=0.15
- n) Coverage power =0.58
- o) L2 regularizer=0.00001

TCN – Bi-LSTM Regressor

- a) Conv1D (1,2,4,8 k=3) X 2
- b) Bi-LSTM (64,32)
- c) Dense (1)
- d) Epochs =200
- e) Batch size = 64
- f) Seed=42
- g) EarlyStopping, patience=8
- h) ReduceLROnPlateau, patience=8
- i) Loss – Masked Huber (δ .=0.6)
- j) Optimizer – Adam
- k) Activation – ReLu, tanh
- l) Learning rate=0.001
- m) Dropout=0.12
- n) Coverage power =0.58
- o) L2 regularizer=0.0002

Appendix X. Historical OOF Calibration Functions

(2015–2022 station-held-out out-of-fold calibration)

Notation: $y_{s,t}$ is the observed $\text{PM}_{2.5}$ concentration for station s at time t , $\hat{y}_{s,t}^{\text{pre}}$ is the corresponding pre-calibration OOF ensemble estimate, and $m(t)$ denotes the calendar month associated with time t .

A. Identity Calibration

The identity function retained the pre-calibration estimate without adjustment:

$$\hat{y}_{s,t}^{(I)} = \hat{y}_{s,t}^{\text{pre}}$$

B. Constant-Bias Calibration

A station-specific mean residual was calculated as:

$$b_s = \frac{1}{N_s} \sum_{t \in \mathcal{T}_s} (y_{s,t} - \hat{y}_{s,t}^{\text{pre}})$$

$$\hat{y}_{s,t}^{(B)} = \hat{y}_{s,t}^{\text{pre}} + b_s$$

where N_s is the number of valid OOF observations available for station s .

C. Monthly-Bias Calibration

A separate additive correction was estimated for each station and calendar month:

$$b_{s,m} = \frac{1}{N_{s,m}} \sum_{t \in \mathcal{T}_{s,m}} (y_{s,t} - \hat{y}_{s,t}^{\text{pre}})$$

$$\hat{y}_{s,t}^{(MB)} = \hat{y}_{s,t}^{\text{pre}} + b_{s,m(t)}$$

where $N_{s,m}$ is the number of valid OOF observations for station s in month m .

D. Shrink-Affine Calibration

A station-specific slope and intercept were estimated using a regularized affine model:

$$(a_s^*, b_s^*) = \arg \min_{a_s, b_s} \left[\sum_{t \in \mathcal{T}_s} (y_{s,t} - a_s \hat{y}_{s,t}^{\text{pre}} - b_s)^2 + \lambda_a (a_s - 1)^2 + \lambda_b b_s^2 \right]$$

$$\hat{y}_{s,t}^{(SA)} = a_s^* \hat{y}_{s,t}^{\text{pre}} + b_s^*$$

The regularization terms shrink the slope toward 1 and the intercept toward 0; the fitted slope was constrained to the interval 0.70–1.40.

Appendix X. Historical OOF Calibration Functions (continued)

E. Cluster-Wise Monthly Affine Calibration

Stations were grouped into clusters, and a monthly affine correction was estimated using pooled OOF observations within each cluster:

$$(a_{k,m}^*, b_{k,m}^*) = \arg \min_{a,b} \sum_{(s,t) \in D_{k,m}} (y_{s,t} - a \hat{y}_{s,t}^{\text{pre}} - b)^2 + \lambda_a (a - 1)^2 + \lambda_b b^2$$

$$\hat{y}_{s,t}^{(CA)} = a_{k(s),m(t)}^* \hat{y}_{s,t}^{\text{pre}} + b_{k(s),m(t)}^*$$

where $k(s)$ denotes the cluster assigned to station s .

F. Calibrator Eligibility and Selection

A non-identity calibrator was retained only when it reduced station-level RMSE by at least 8% relative to the identity function and did not increase the absolute mean bias error:

$$\text{RMSE}_s(c) \leq 0.92 \text{RMSE}_s(I)$$

$$|\text{MBE}_s(c)| \leq |\text{MBE}_s(I)|$$

$$c_s^* = \arg \min_{c \in \mathcal{C}_s^{\text{eligible}}} \text{RMSE}_s(c)$$

The candidate set was $\mathcal{C} = \{I, B, MB, SA, CA\}$.

G. Global Bias Anchoring

After applying the selected calibration functions, the remaining mean residual was calculated as:

$$B_{\text{global}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i^{\text{cal},0})$$

$$\hat{y}_i^{\text{cal}} = \hat{y}_i^{\text{cal},0} + B_{\text{global}}$$

where $\hat{y}_i^{\text{cal},0}$ is the calibrated estimate before global bias anchoring.

This appendix summarizes the historical OOF calibration used during model development. It is distinct from the separate 2023 observation-assisted station-tuning functions applied prior to station-assisted mapping.

Appendix X. Candidate Functions for 2023 Station Tuning

The equations below summarize the six candidate functions used to retrospectively adjust the 2023 pre-calibration station estimates before delta-based spatial mapping. For station s and month m , $y_{s,m}$ is the observed surface-level $\text{PM}_{2.5}$ concentration and $\hat{y}_{s,m}^{\text{pre}}$ is the corresponding pre-calibration model estimate.

Common notation. Define $\text{clip}(x; L, U) = \min\{\max(x, L), U\}$ and $C_{K_s}(x) = \text{clip}(x; -K_s, K_s)$, where K_s is the station-specific maximum adjustment selected from the tested clipping grid. Residual updates were bounded to $\pm 6 \mu\text{g m}^{-3}$.

A. Exponential moving-average (EMA) bias

The most recent residual updates a single additive bias term:

$$\bar{\epsilon}_{s,m-1} = \text{clip}\left(y_{s,m-1} - \hat{y}_{s,m-1}^{\text{pre}}; -6, 6\right), \quad (1)$$

$$b_{s,m}^{(B)} = (1 - \alpha_b) b_{s,m-1}^{(B)} + \alpha_b \bar{\epsilon}_{s,m-1}, \quad (2)$$

$$\hat{y}_{s,m}^{(B)} = \hat{y}_{s,m}^{\text{pre}} + C_{K_s}\left(b_{s,m}^{(B)}\right). \quad (3)$$

The bias state is initialized at $b_{s,0}^{(B)} = 0$.

B. Affine EMA

Affine EMA updates both the scale and additive bias from the previous month:

$$\rho_{s,m-1} = \text{clip}\left(\frac{y_{s,m-1}}{\max(\hat{y}_{s,m-1}^{\text{pre}}, 10^{-3})}; 0.70, 1.40\right), \quad (4)$$

$$\bar{\epsilon}_{s,m-1}^{(A)} = \text{clip}\left(y_{s,m-1} - a_{s,m-1}^{(A)} \hat{y}_{s,m-1}^{\text{pre}}; -6, 6\right), \quad (5)$$

$$a_{s,m}^{(A)} = (1 - \alpha_a) a_{s,m-1}^{(A)} + \alpha_a \rho_{s,m-1}, \quad (6)$$

$$b_{s,m}^{(A)} = (1 - \alpha_b) b_{s,m-1}^{(A)} + \alpha_b \bar{\epsilon}_{s,m-1}^{(A)}, \quad (7)$$

$$\hat{y}_{s,m}^{(A)} = \hat{y}_{s,m}^{\text{pre}} + C_{K_s}\left(a_{s,m}^{(A)} \hat{y}_{s,m}^{\text{pre}} + b_{s,m}^{(A)} - \hat{y}_{s,m}^{\text{pre}}\right). \quad (8)$$

The states are initialized at $a_{s,0}^{(A)} = 1$ and $b_{s,0}^{(A)} = 0$.

Interpretation: EMA bias represents gradually evolving additive error, whereas affine EMA also permits the prediction scale to change.

Appendix X – 2023 station-tuning functions | 1

Appendix X. Candidate Functions for 2023 Station Tuning (continued)

C. Changepoint bias and changepoint affine

Let $\epsilon_{s,m} = y_{s,m} - \hat{y}_{s,m}^{\text{pre}}$. A candidate changepoint separates the residual record into two regimes:

$$\tau_s^* = \arg \min_{\tau} \sum_{m \leq \tau} (\epsilon_{s,m} - \bar{\epsilon}_{s,1})^2 + \sum_{m > \tau} (\epsilon_{s,m} - \bar{\epsilon}_{s,2})^2, \quad (9)$$

where $\bar{\epsilon}_{s,1}$ and $\bar{\epsilon}_{s,2}$ are the mean residuals before and after τ_s^* .

The changepoint-bias estimate is

$$\hat{y}_{s,m}^{(\text{CB})} = \hat{y}_{s,m}^{\text{pre}} + C_{K_s}\left[\bar{\epsilon}_{s,1} \mathbb{I}(m \leq \tau_s^*) + \bar{\epsilon}_{s,2} \mathbb{I}(m > \tau_s^*)\right]. \quad (10)$$

For changepoint affine, separate ridge-regularized slopes and intercepts are fitted to the two regimes:

$$\hat{y}_{s,m}^{(\text{CPA})} = \begin{cases} \hat{y}_{s,m}^{\text{pre}} + C_{K_s}(a_{s,1} \hat{y}_{s,m}^{\text{pre}} + b_{s,1} - \hat{y}_{s,m}^{\text{pre}}), & m \leq \tau_s^*, \\ \hat{y}_{s,m}^{\text{pre}} + C_{K_s}(a_{s,2} \hat{y}_{s,m}^{\text{pre}} + b_{s,2} - \hat{y}_{s,m}^{\text{pre}}), & m > \tau_s^*. \end{cases} \quad (11)$$

D. Hybrid correction

The hybrid uses changepoint bias only when its recent rolling mean-squared error is at least 8% lower than that of EMA bias:

$$I_{s,m} = \mathbb{I}\left[\text{MSE}_{s,m}^{(\text{CB})} < 0.92 \text{MSE}_{s,m}^{(B)}\right], \quad (12)$$

$$\hat{y}_{s,m}^{(H)} = I_{s,m} \hat{y}_{s,m}^{(\text{CB})} + (1 - I_{s,m}) \hat{y}_{s,m}^{(B)}. \quad (13)$$

E. EMA-affine blend

The blend combines the EMA-bias and affine-EMA estimates using a rolling, performance-based convex weight:

$$\hat{y}_{s,m}^{(\text{BL})} = (1 - \beta_{s,m}) \hat{y}_{s,m}^{(B)} + \beta_{s,m} \hat{y}_{s,m}^{(A)}, \quad (14)$$

$$\beta_{s,m} = \text{clip}\left(\frac{\sum_{k \in \mathcal{H}_{s,m}} d_{s,k} (y_{s,k} - \hat{y}_{s,k}^{(B)})}{\sum_{k \in \mathcal{H}_{s,m}} d_{s,k}^2}; 0, 1\right), \quad (15)$$

$$d_{s,k} = \hat{y}_{s,k}^{(A)} - \hat{y}_{s,k}^{(B)}, \quad (16)$$

$$\beta_{s,m} = \text{clip}\left(0.7 \beta_{s,m}^* + 0.3 \beta_{s,m}^{\text{prior}}; 0, 1\right). \quad (17)$$

Here, $\mathcal{H}_{s,m}$ contains the available recent months used to judge performance.

F. Station-specific selection

For each station, the candidate with the highest 2023 station-level coefficient of determination was retained:

$$k_s^* = \arg \max_{k \in \mathcal{C}} R_s^2(k), \quad (18)$$

$$\hat{y}_{s,m}^{\text{tuned}} = \hat{y}_{s,m}^{(k_s^*)}, \quad (19)$$

$$\mathcal{C} = \{B, A, \text{CPB}, \text{CPA}, H, \text{BL}\}. \quad (20)$$

RMSE, MAE, and MBE were calculated as supporting diagnostics but were not used to select the winning function.

Use note: the tuned values are observation-assisted model estimates, not the station observations themselves. They were subsequently used to derive the station correction deltas for KNN-IDW mapping.

Appendix X – 2023 station-tuning functions | 2

APPENDIX C

Daily AOD was obtained from the JAXA Himawari Monitor Level 3 Daily Aerosol Property product (1DARP) at 500 nm. The *AOT_L3_Merged_Mean* field was used (derived from cloud-screened AOT retrievals), representing the daily mean of the Level 3 merged AOD product. Because this Level 3 product had already undergone JAXA retrieval, cloud-screening, and quality-control processing, no additional QA bit filtering was applied in this study. Missing or invalid values provided in the product were excluded, while no additional numerical AOD threshold was imposed. Monthly AOD for each pixel was calculated as the arithmetic mean of all available valid daily *AOT_L3_Merged_Mean* observations within the month. The number of valid contributing days was retained separately as the AOD daily-valid-count variable.

Land cover information was obtained from the ESA WorldCover 10 m 2021 v200 product. This is a global land-cover classification for the reference year 2021 derived from Sentinel-1 and Sentinel-2 observations. The WorldCover product contains 11 original land cover classes and was produced using a gradient-boosting decision-tree classifier (CatBoost), followed by expert-rule-based refinement. The product has a reported global overall accuracy of $76.7 \pm 0.5\%$, while the reported overall accuracy for Asia is $82.1 \pm 1.0\%$. In this study, the 10 m categorical map was converted to 1km fractional land cover layers. Five classes were retained explicitly: Tree cover (class 10), Grassland (30), Cropland (40), Built-up (50), and Permanent water bodies (80). The remaining WorldCover classes—Shrubland, Bare/sparse vegetation, Snow and ice, Herbaceous wetland, Mangroves, and Moss and lichen—were combined as ‘Other.’ Fractional coverage within each 1km pixel was calculated by averaging the corresponding binary 10 m class masks.

Population density was derived from the **WorldPop 2020 Taiwan population density raster** and adjusted using **2020 Ministry of the Interior (MOI) population totals** for Taipei, New Taipei, Keelung, Taoyuan, and Yilan to improve consistency with official local statistics. The resulting 2020 population density surface was treated as a static spatial covariate throughout 2015–2023. This assumption was adopted because population density patterns were expected to vary more gradually than the monthly environmental variables; however, it does not imply

that population remained unchanged over the study period, and interannual demographic changes were therefore not explicitly represented, which is an acknowledged challenge.

Building footprint and building height were derived from the GHSL 2018 built surface and average building height products. These layers were resampled to the 1km study grid and treated as static urban form covariates throughout 2015–2023. This assumption reflects the relatively slow evolution of urban morphology compared with monthly environmental variables, while acknowledging that construction and redevelopment during the study period were not explicitly represented.

Road network density was calculated by combining road, conventional railway, and high-speed railway vector networks (as of 2021) and summing their line lengths within each 1km grid cell, expressed as km². The resulting transport network density was treated as a static spatial covariate during 2015–2023 because major transport infrastructure changes relatively slowly compared with the monthly environmental variables. Nevertheless, network additions or modifications during the study period were not explicitly represented.

APPENDIX D

A total of 28 monitoring stations were initially considered. Jitter sampling was performed on the 1km grid using offsets of ± 2 km in both spatial directions, yielding a (5 x 5) neighborhood (25 candidate samples) for each standard station. Yangming used a ± 3 km guarded neighborhood ((7 x 7), 49 candidates). Thus, a theoretical maximum of 724 candidates’ locations was possible. Nearest-station ownership was enforced, duplicate station pixel locations were removed and samples with no valid samples were discarded, to remain with 590 samples. Each candidate sample retained its parent-station identity together with its horizontal and vertical offsets dx , dy and ring distance r from the station core. Candidate samples were subsequently screened using the temporal validity mask (≥ 24 valid channels), the nearest-station consistency guard, and the effective weight criterion, as discussed previously in chapter 3. Samples failing these checks were removed before model training to remain with 517 valid samples from 27 stations, with station-specific counts provided in the table below. There was no observation of nearest station violation from the samples.

Station	Initial Samples	Final Samples	Discarded Samples
Banqiao	16	16	0
Cailiao	12	12	0
Cape Fuguei	25	0	25
Datong	5	5	0
Dayuan	25	24	1
Dongshan	25	25	0
Guanyin	25	11	14

Guting	15	15	0
Keelung	25	14	11
Linkou	25	25	0
Longtan	25	25	0
Pingzhen	20	20	0
Sanchong	11	11	0
Shilin	19	19	0
Songshan	25	25	0
Tamsui	25	22	3
Taoyuan	25	25	0
Tucheng	20	20	0
Wanhua	12	12	0
Wanli	25	6	19
Xindian	25	25	0
Xinzhuang	22	22	0
Xizhi	25	25	0
Yangming	49	49	0
Yilan	25	25	0
Yonghe	9	9	0
Zhongli	15	15	0
Zhongshan	15	15	0

References

- Basith, S., Manavalan, B., Shin, T.H., Park, C.B., Lee, W.S., Kim, J., and Lee, G., 2022. The impact of fine particulate matter 2.5 on the cardiovascular system: a review of the invisible killer, *Nanomaterials*, 12(15): 2656, DOI:10.3390/nano12152656.
- Berrocal, V.J., Guan, Y., Muyskens, A., Wang, H., Reich, B.J., Mulholland, J.A., and Chang, H.H., 2020. A comparison of statistical and machine learning methods for creating national daily maps of ambient PM_{2.5} concentration, *Atmospheric Environment*, 222: 117130, DOI:10.1016/j.atmosenv.2019.117130.
- Chang, I.-C., 2019. Identifying leading nodes of PM_{2.5} monitoring network in Taiwan with big data-oriented social network analysis, *Aerosol and Air Quality Research*, 9(12): 2844–2864, DOI:10.4209/aaqr.2019.11.0554.
- Chen, G., Li, S., Knibbs, L.D., Hamm, N.A.S., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M.J., and Guo, Y., 2018. A machine learning method to estimate PM_{2.5} concentrations across China with remote sensing, meteorological, and land use information, *Science of The Total Environment*, 636: 52–60, DOI:10.1016/j.scitotenv.2018.04.251.
- Cheng, F.Y., and Hsu, C.H., 2019. Long-term variations in PM_{2.5} concentrations under changing meteorological conditions in Taiwan, *Scientific Reports*, 9(1): 6635, DOI:10.1038/s41598-019-43104-x.
- Chien, L.C., Chen, Y.A., and Yu, H.L., 2018. Lagged influence of fine particulate matter and geographic disparities on clinic visits for children's asthma in Taiwan, *International Journal of Environmental Research and Public Health*, 15(4): 829, DOI:10.3390/ijerph15040829.
- Gupta, S., Pebesma, E., Mateu, J., and Degbelo, A., 2018. Air quality monitoring network design optimisation for robust land use regression models, *Sustainability*, 10(5): 1442, DOI:10.3390/su10051442.
- Hacıoğlu, H.İ., Arı, A., Özkan, A., Elbir, T., Tuncel, G., Yay, O.D., and Gaga, E.O., 2016. A new approach for site selection of air quality monitoring stations: Multi-criteria decision-making, *Aerosol and Air Quality Research*, 16(6): 1390–1402, DOI:10.4209/aaqr.2014.11.0273.
- Hu, J., Jia, Y., Jia, Z.H., He, C.B., Shi, F., and Huang, X.H., 2024. Prediction of PM_{2.5} concentration based on deep learning for high-dimensional time series, *Applied Sciences*, 14(19): 8745, DOI:10.3390/app14198745.
- Jung, C.-R., Hwang, B.-F., and Chen, W.-T., 2018. Incorporating long-term satellite-based aerosol optical depth, localized land use data, and meteorological variables to estimate ground-level PM_{2.5} concentrations in Taiwan from 2005 to 2015, *Environmental Pollution*, 237: 1000–1010, DOI:10.1016/j.envpol.2017.11.016.
- Keelung City Government, 2022. Keelung City, Available at: <https://www.klccg.gov.tw/en/klcg2>, Assessed April 6, 2026.
- Lee, S.J., Serre, M.L., van Donkelaar, A., Martin, R.V.,

- Burnett, R.T., and Jerrett, M., 2012. Comparison of geostatistical interpolation and remote sensing techniques for estimating long-term exposure to ambient PM_{2.5} concentrations across the continental United States, *Environmental Health Perspectives*, 120(12): 1727–1732, DOI:10.1289/ehp.1205006.
- Li, J., Chen, J., You, R., and He, Q., 2025. PM_{2.5} concentration prediction: Ultrahigh spatiotemporal resolution achieved by combining machine learning and low-cost sensors, *Sensors*, 25(17): 5527, DOI:10.3390/s25175527.
- Li, T., Shen, H., Yuan, Q., and Zhang, L., 2022. A locally weighted neural network constrained by global training for remote sensing estimation of PM_{2.5}, *IEEE Transactions on Geoscience and Remote Sensing*, 60: 4102513, DOI:10.1109/TGRS.2021.3074569.
- Li, T., Shen, H., Zeng, C., Yuan, Q., and Zhang, L., 2017. Point-surface fusion of station measurements and satellite observations for mapping PM_{2.5} distribution in China: Methods and assessment, *Atmospheric Environment*, 152: 477–489, DOI:10.1016/j.atmosenv.2017.01.004.
- Li, W., and Jiang, X., 2023. Prediction of air pollutant concentrations based on TCN-BiLSTM-DMAAttention with STL decomposition, *Scientific Reports*, 13(1): 4665, DOI:10.1038/s41598-023-31569-w.
- Lin, C.-Y., Zhao, C., Liu, X., Lin, N.-H., and Chen, W.-N., 2014. Modelling of long-range transport of Southeast Asia biomass-burning aerosols to Taiwan and their radiative forcings over East Asia, *Tellus B: Chemical and Physical Meteorology*, 66(1): 23733, DOI:10.3402/tellusb.v66.23733.
- Lin, S., Zhao, J., Li, J., Liu, X., Zhang, Y., Wang, S., Mei, Q., Chen, Z., and Gao, Y., 2022. A spatial-temporal causal convolution network framework for accurate and fine-grained PM_{2.5} concentration prediction, *Entropy*, 24(8): 1125, DOI:10.3390/e24081125.
- Lo, W.-C., Ho, C.-C., Tseng, E., Hwang, J.-S., Chan, C.-C., and Lin, H.-H., 2022. Long-term exposure to ambient fine particulate matter (PM_{2.5}) and associations with cardiopulmonary diseases and lung cancer in Taiwan: A nationwide longitudinal cohort study, *International Journal of Epidemiology*, 51(4): 1230–1242, DOI:10.1093/ije/dyac082.
- New Taipei City Government, 2017. New Taipei City, Available at: <https://foreigner.ntpc.gov.tw/home.jsp?id=d127e0ce0f4f407b&utm>, Assessed April 6, 2026.
- Ong, B.T., Sugiura, K., and Zettsu, K., 2016. Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting PM_{2.5}, *Neural Computing and Applications*, 27(6): 1553–1566, DOI:10.1007/s00521-015-1955-3.
- Qin, D., Yu, J., Zou, G., Yong, R., Zhao, Q., and Zhang, B., 2019. A novel combined prediction scheme based on CNN and LSTM for urban PM_{2.5} Concentration, *IEEE Access*, 7: 20050–20059, DOI:10.1109/ACCESS.2019.2897028.
- Requia, W.J., Coull, B.A., and Koutrakis, P., 2019. Evaluation of predictive capabilities of ordinary geostatistical interpolation, hybrid interpolation, and machine learning methods for estimating PM_{2.5} constituents over space, *Environmental Research*, 175: 421–433, DOI:10.1016/j.envres.2019.05.025.
- Song, Y., Huang, B., He, Q., Chen, B., Wei, J., and Mahmood, R., 2019. Dynamic assessment of PM_{2.5} exposure and health risk using remote sensing and geo-spatial big data, *Environmental Pollution*, 253: 288–296, DOI:10.1016/j.envpol.2019.06.057.
- Taipei City Government, 2025. Taipei City, Available at: <https://english.gov.taipei/>, Assessed April 6, 2026.
- US EPA, 2025. Particulate Matter (PM) Pollution, Available at: <https://www.epa.gov/pm-pollution/particulate-matter-pm-basics>, Assessed April 6, 2026.
- Wang, F., Chen, T., Chang, Q., Kao, Y.-W., Li, J., Chen, M., Li, Y., and Shia, B.-C., 2021. Respiratory diseases are positively associated with PM_{2.5} concentrations in different areas of Taiwan, *PLoS One*, 16(4): e0249694, DOI:10.1371/journal.pone.0249694.
- Wang, Z., Crooks, J.L., Regan, E.A., and Karimzadeh, M., 2025. High-resolution estimation of daily PM_{2.5} levels in the contiguous US using Bi-LSTM with attention, *Remote Sensing*, 17(1): 126, DOI:10.3390/rs17010126.
- Wei, J., Li, Z., Lyapustin, A., Wang, J., Dubovik, O., Schwartz, J., Sun, L., Li, C., Liu, S., and Zhu, T., 2023. First close insight into global daily gapless 1 km PM_{2.5} pollution, variability, and health impact, *Nature Communications*, 14(1): 8349, DOI:10.1038/s41467-023-43862-3.
- WHO, 2024. Ambient(outdoor) air pollution, Available at: <https://www.who.int/news-room/factsheets/detail/ambient-%28outdoor%29-air-quality-and-health>, Assessed April 6, 2026.
- WHO, 2025. Air quality, energy and health, Available at:

- [climate-change-and-health/air-quality-and-health/health-impacts/types-of-pollutants](#), Assessed April 6, 2026.
- Wu, Y., Xu, Z., Xu, L., and Wei, J., 2024. An improved deep learning approach considering spatiotemporal heterogeneity for PM_{2.5} prediction: A case study of Xinjiang, China, *Atmosphere*, 15(4):460, DOI:10.3390/atmos15040460.
- Yang, K., Shi, Y., Luo, Y., Liu, R., Sun, W., and Sun, M., 2020. Assessing spatiotemporal air environment degradation and improvement represented by PM_{2.5} in China using a two-phase hybrid model, *Sustainable Cities and Society*, 59: 102180, DOI:10.1016/j.scs.2020.102180.
- Yu, T., Wang, Y., Huang, J., Liu, X., Li, J., and Zhan, W., 2022. Study on the regional prediction model of PM_{2.5} concentrations based on multi-source observations, *Atmospheric Pollution Research*, 13(4): 101363, DOI:10.1016/j.apr.2022.101363.
- Zhang, J., Peng, Y., Ren, B., and Li, T., 2021a. PM_{2.5} concentration prediction based on CNN-BiLSTM and attention mechanism, *Algorithms*, 14(7): 208, DOI:10.3390/a14070208.
- Zhang, Y., Cheng, H., Huang, D., and Fu, C., 2021b. High temporal resolution land use regression models with POI characteristics of the PM_{2.5} distribution in Beijing, China, *International Journal of Environmental Research and Public Health*, 18(11):6143, DOI:10.3390/ijerph18116143.
- Zhang, Z., Wang, J., Hart, J.E., Laden, F., Zhao, C., Li, T., Zheng, P., Li, D., Ye, Z., and Chen, K., 2018. National scale spatiotemporal land-use regression model for PM_{2.5}, PM₁₀, and NO₂ concentration in China, *Atmospheric Environment*, 192: 48–54, DOI:10.1016/j.atmosenv.2018.08.046.

基於遙測之都會區 PM_{2.5} 濃度估測之深度學習方法： 以大臺北地區為例

魏文森^{1*} 任玄² 林唐煌³

摘要

都市空氣污染仍為重要環境威脅，細懸浮微粒(PM_{2.5})對人體健康具有不利影響，因此提升其估算精度十分重要。本研究提出時間卷積網路與雙向長短期記憶網路(TCN - Bi-LSTM)混合模型，使用臺灣環境部 2015 年 7 月至 2022 年 12 月歷史空氣品質、氣象及土地利用資料，估算 2023 年逐月 PM_{2.5}，並以 K 近鄰 - 反距離加權法(KNN - IDW)產製 1 km 濃度圖。GroupKFold 交叉驗證顯示，本模型優於 CNN 與 LSTM， $R^2 = 0.7960$ ， $RMSE = 2.1606 \mu g/m^3$ 。本研究整合多源環境資料，以降低對衛星 AOD 直接推估 PM_{2.5} 的依賴，並減少雲層造成的 AOD 缺值影響，建立連續 PM_{2.5} 暴露圖。未來將納入獨立地表 PM_{2.5} 觀測資料，以進一步驗證模型表現。

關鍵詞：懸浮微粒(PM)、時空變化、時間卷積網路(TCN)、長短期記憶(LSTM)、雙向長短期記憶(Bi-LSTM)

¹ 國立中央大學 太空及遙測研究中心 碩士生

² 國立中央大學 太空及遙測研究中心 教授

³ 國立中央大學 太空及遙測研究中心 特聘教授

* 通訊作者，電話：0981-038-441, E-mail：vincletus2050@gmail.com

收到日期：民國 115 年 06 月 12 日

修改日期：民國 115 年 06 月 29 日

接受日期：民國 115 年 08 月 19 日