

E-STGCN: extreme spatio-temporal graph convolutional networks for air quality forecasting

Madhurima Panja^{1,†} , Tanujit Chakraborty^{1,2,†} , Anubhab Biswas³
and Soudeep Deb⁴ 

¹SAFIR, Sorbonne University Abu Dhabi, Al Reem Island, Abu Dhabi 38044, UAE

²Sorbonne Center for Artificial Intelligence, Sorbonne University, Paris 75005, France

³SUPSI, University of Applied Sciences and Arts of Southern Switzerland, Manno, Switzerland

⁴Decision Sciences Area, Indian Institute of Management Bangalore, Bangalore, India

Address for correspondence: Tanujit Chakraborty. Email: tanujit.chakraborty@sorbonne.ae

Abstract

Modelling and forecasting air quality is crucial for effective air pollution management and protecting public health. Air quality data, characterized by nonlinearity, nonstationarity, and spatio-temporal correlations, often include extreme pollutant levels in severely polluted cities (e.g. Delhi, the capital of India). This is ignored by various geometric deep learning models, such as spatio-temporal graph convolutional networks (STGCNs), which are otherwise effective for spatio-temporal forecasting. This study develops an extreme value theory (EVT) guided modified STGCN model (E-STGCN) for air pollution data to incorporate extreme behaviour across pollutant concentrations. E-STGCN combines graph convolutional networks for spatial modelling and EVT-guided long short-term memory units for temporal sequence learning. Along with spatial and temporal components, it incorporates a generalized Pareto distribution to capture the extreme behaviour of different air pollutants and embed this information into the learning process. The proposal is then applied to analyse air pollution data of 37 monitoring stations across Delhi, India. The forecasting performance for different test horizons is compared to benchmark forecasters (both temporal and spatio-temporal). It is found that E-STGCN has consistent performance across all seasons. The robustness of our results has also been evaluated empirically. Moreover, combined with conformal prediction, E-STGCN can produce probabilistic prediction intervals.

Keywords: air quality, graph convolutional networks, extreme value modelling, spatio-temporal forecasting

1 Introduction

Rapid industrialization and urbanization have spurred global economic growth but also intensified environmental issues, with air pollution ranking among the most pressing concerns (Brunekreef & Holgate, 2002; Shaddick et al., 2020). According to the World Health Organization (WHO), approximately seven million premature deaths per year are linked to air pollution,¹ underscoring the need for urgent air quality management. Major air pollutants including particulate matters (PM), nitrogen dioxide (NO₂), ozone (O₃), sulphur dioxide (SO₂), and carbon monoxide (CO), pose significant risks to cardiovascular and respiratory health (Lelieveld et al., 2015; Olaniyan et al., 2020). Recognizing this growing threat, the United Nations included air quality improvement

[†] M.P. and T.C. contributed equally.

¹ <https://www.who.int/health-topics/air-pollution>.

within the Sustainable Development Goals,² and many countries have adopted National Ambient Air Quality Standards (NAAQS) to regulate pollution levels. In India, the Central Pollution Control Board (CPCB) mandates that hourly average concentrations of PM with a diameter of 2.5 μm or less ($\text{PM}_{2.5}$) and PM with a diameter of 10 μm or less (PM_{10}) pollutants should not exceed 60 and 100 $\mu\text{g}/\text{m}^3$, respectively.³ However, observations from 37 monitoring stations across Delhi from 2019 to 2023⁴ reveal average $\text{PM}_{2.5}$ and PM_{10} levels surpassing 100 and 200 $\mu\text{g}/\text{m}^3$, far above regulatory limits. These concentrations intensify further during winter due to low temperatures and Delhi's landlocked geography, which restricts pollutant dispersion, contributing to heightened risks of respiratory, cardiovascular, and neurological diseases (Salvi et al., 2018). Pandey et al. (2021) further highlights that air pollution adversely affects India's economic growth as well. Given these challenges, our study aims to develop a spatio-temporal forecasting model to improve air quality prediction in urban environments. Such models are essential for informing public behaviour and helping authorities implement timely interventions to mitigate health risks.

Air quality forecasting methods can be broadly classified into two categories: physical models and data-driven approaches. Physical models, such as community multiscale air quality (Byun & Schere, 2006) and the nested air quality prediction model system (Wang et al., 2014), simulate pollutant emissions, transport, and dispersion using atmospheric principles. However, they often require extensive domain expertise, region-specific calibration, and heavy computation, making them less suitable for real-time monitoring; simpler models like Gaussian plume and Operational street canyon formulations also struggle with accuracy due to limited parameterization (Byun & Schere, 2006; Vardoulakis et al., 2003). In contrast, data-driven methods that leverage historical information to capture pollution trends have shown some promise (Lei et al., 2019). Thus, traditional statistical models such as ARIMA (autoregressive integrated moving average) and dynamic factor models are widely used (Kumar & Jain, 2010), but they cannot fully capture the nonlinear structure of air quality processes. Deep learning frameworks have addressed these limitations, with long short-term memory (LSTM) (Li et al., 2017), hybrid convolutional neural networks (CNNs)–LSTM architectures (Du et al., 2019), recurrent networks (Ong et al., 2016), transformers (Vaswani, 2017), and temporal convolutional networks (TCNs) (Samal et al., 2021), yielding improved predictive performance. However, most of these models primarily focus on temporal patterns alone and overlook the spatial interactions between monitoring locations. Since neighbouring locations influence pollutant levels at any given station, a spatio-temporal approach is essential to model air pollution dispersion (Zhou et al., 2024) accurately. Graph-based methods, particularly graph neural networks (GNNs) and graph convolutional networks (GCNs), have emerged as powerful tools for modelling spatio-temporal dependencies (Scarselli et al., 2008). In the current context, Gao and Li (2021) leveraged GNNs with LSTMs to capture spatio-temporal transmission of $\text{PM}_{2.5}$. GCNs are also effective for air quality modelling because they propagate information across connected locations, enabling localized feature aggregation from neighbouring stations (Yu et al., 2018). For a comprehensive discussion in this context, refer to Atluri et al. (2018) and Jin et al. (2024).

The application of GCN-based approaches for modelling spatial dynamics in air pollution networks remains limited due to scalability and data sparsity issues. Moreover, existing forecasting architectures often struggle to accurately predict peaks in a time series, which is critical for air pollution forecasting to anticipate exceedances beyond regulatory thresholds. Due to the catastrophic nature of extreme pollution levels, accurately modelling values above NAAQS thresholds is crucial for effective early warning systems. Extreme value theory (EVT) offers a principled statistical framework for analysing rare events and has been widely applied in hydrology, climate science, and air quality to characterize the probability distribution of extreme pollutant concentrations (Coles et al., 2001; Ray et al., 2023). In air quality studies, EVT methods such as block maxima (BM) and peaks over threshold (POT) have been employed to model pollutant extremes using generalized extreme value (GEV) and generalized Pareto (GP) distributions, respectively (Reiss et al., 1997). These techniques help in estimating the likelihood of extreme occurrences, allowing for the detection of potential rare events. EVT-based studies have been instrumental in forecasting pollution exceedances, thus supporting effective intervention strategies (Kan & Chen, 2004; Sfetsos

² <https://sdgs.un.org/goals>.

³ https://cpcb.nic.in/upload/NAAQS_2019.pdf.

⁴ Delhi remains the world's most polluted capital city in 2023; <https://www.statista.com>.

et al., 2006). The reader is further referred to [Martins et al. \(2017\)](#) for a comprehensive review of EVT tools in air pollution problems.

Interestingly, despite its proven utility, EVT has not been combined with spatio-temporal forecasting methods to build early warning systems for environmental preparedness. Our study aims to bridge this gap by introducing a novel EVT-guided modified spatio-temporal graph convolutional network (E-STGCN) model to handle the nonlinear, nonstationary, and extreme behaviour for major air pollutants in Delhi, specifically $PM_{2.5}$, PM_{10} , and NO_2 . We examine the extreme behaviour of these pollutants across 37 sensor locations using the POT method, modelled by the GP distribution. Integrating these insights into a modified version of spatio-temporal GCN (modified STGCN) enhances its ability to forecast air quality time series data while also capturing the extreme observations. This enables the E-STGCN model to capture extremes in comparison with standard spatio-temporal GCN (STGCN), which is designed for inference on graph structures with temporal dependencies in traffic flow applications ([Yu et al., 2018](#)). Standard STGCN ([Yu et al., 2018](#)) leverages graph convolutions to model spatial dependencies and 1-D convolutional layers with gating mechanisms to capture short-term temporal patterns. On the contrary, the proposed E-STGCN framework builds upon the foundational spatial convolution blocks of STGCN; however, it differs from the temporal forecasting perspective and loss function construction.

In the E-STGCN architecture, we retain the graph convolutional structure of STGCN to encode the spatial dependencies among monitoring stations effectively. However, we introduce two key innovations in the framework: (1) employing LSTM units in place of temporal convolution layers to better capture long-memory temporal dynamics (evident in air pollutant time series), we name this architecture as modified STGCN and (2) introducing a new hybrid loss function based on data loss and negative log-likelihood of the GP distribution (we call it POT loss). The use of the POT loss function enhances the modelling ability of rare but impactful extreme pollution events. The EVT component in E-STGCN enhances the model's sensitivity to tail behaviour via introducing a task-specific inductive bias that prioritizes accuracy in the extreme regime, which is typically under-represented in conventional data loss formulations. There is a conceptual connection between the proposed E-STGCN and that of Physics-informed neural networks (PINNs) ([Raissi et al., 2019](#)), which combines (noisy) data with physical models and implements them through deep neural networks. However, in our method, instead of a physical model, we use an EVT-based GP distribution-fitted model as the building block for E-STGCN. Moreover, to evaluate the effect of the POT loss, we compare it with the modified version of STGCN and the standard STGCN model, as proposed by [Yu et al. \(2018\)](#). Modified STGCN shares the same architecture as E-STGCN, including the graph convolutional and LSTM components, but is trained solely using the standard data loss, without incorporating the EVT module. As we shall demonstrate in our empirical study, the proposed E-STGCN, which enhances the original STGCN by incorporating both LSTM-based temporal modelling and an EVT-based loss function, achieves superior performance while capturing extreme air pollutant concentrations. In contrast, modified STGCN, which retains the architectural enhancements of E-STGCN but excludes the EVT component, shows improved baseline performance over STGCN in capturing long memory but lacks the same precision as E-STGCN in modelling extreme events. Furthermore, our proposed framework is scalable and capable of generating multistep forecasts for both low and high-frequency spatio-temporal datasets. Unlike traditional GNN models, typically optimized for hourly predictions and shorter horizons (e.g. 12–72 h), E-STGCN handles daily air quality data, providing reliable long-term forecasts at 30-, 60-, and 90-day horizons. Additionally, we apply conformal prediction methods to quantify forecast uncertainties, offering critical probabilistic insights for policy planning.

The remainder of this paper is organized as follows. Section 2 provides a brief description of EVT. In Section 3, we introduce the proposed E-STGCN architecture. Section 4 outlines the experimental setup and reports the air quality forecasting results. In Section 5, we discuss the implications of our approach to air quality forecasting. Finally, Section 6 concludes the paper and suggests future research directions.

2 Preliminaries on EVT

EVT focuses on analysing the stochastic behaviour of rare or extreme events within a given stochastic process. The goal of extreme value analysis is to quantify unusually large or small events

and estimate the probability of these extreme occurrences, which differ significantly from the more common observations in the data. EVT deals with the asymptotic distribution of extreme order statistics, especially in the context of large datasets. This theory has been implemented in diverse domains, including earth sciences (Katz et al., 2002), economics and finance (Marimoutou et al., 2009), public health (Thomas et al., 2016), and engineering (Castillo, 2012), among others. As mentioned above, statistical methods for modelling extreme events primarily rely on two approaches, i.e. BM and POT. Below, we briefly summarize the EVT methods that were utilized in this study.

2.1 BM approach

The BM method analyses extreme events in a time series dataset (Gumbel, 1958). Given a sequence of time-dependent observations, this method divides the dataset into equal-sized nonoverlapping blocks and considers the maximum value from each block as the extreme value of the time series. The probability distribution of the extremes is modelled using the generalized extreme value (GEV) distribution. To mathematically explain this, let X_1, X_2, \dots, X_n be independent and identically distributed (iid) random variables with continuous distribution function $F(\cdot)$. Then, as $n \rightarrow \infty$, the distribution of $M_n = \max_{1 \leq i \leq n} X_i$ converges to $G(x)$, called the GEV distribution, defined by (following Fisher & Tippett, 1928)

$$G(x) = \begin{cases} \exp \left\{ - \left(1 + \zeta_G \left(\frac{x - \mu_G}{\sigma_G} \right) \right)^{-1/\zeta_G} \right\} & \text{if } \zeta_G \neq 0, \\ \exp \left\{ - \exp \left(- \left(\frac{x - \mu_G}{\sigma_G} \right) \right) \right\} & \text{if } \zeta_G = 0. \end{cases}$$

In the above distribution, $\zeta_G \in \mathbb{R}$ is the extreme value index and it controls the shape of the distribution, $\mu_G \in \mathbb{R}$ is the location parameter, and $\sigma_G > 0$ is the scale parameter. Depending on the tail behaviour of the distribution, which is influenced by ζ_G , the GEV family can be classified into three extreme value distributions: Gumbel ($\zeta_G = 0$), Fréchet ($\zeta_G > 0$), and Weibull ($\zeta_G < 0$). While the Gumbel type distributions are suitable for modelling the extremes of the exponentially decaying-tailed distribution, Fréchet and Weibull families are the reference classes for the extremes of heavy-tailed and finite-tailed distributions, respectively (Rocco, 2014).

Although the BM method has been widely used for extreme value analysis, it has several drawbacks. The partitioning of the dataset in this approach leads to significant information loss, as only the maximum value from each block is retained, potentially missing multiple extreme observations within a block. Also, usually, multiple extreme observations happen within a short time interval, which cannot be captured by the BM method. POT tries to overcome the disadvantages of the BM approach.

2.2 POT approach

The POT approach is a key technique in EVT that identifies observations exceeding a preselected threshold, known as extreme values (Balkema & De Haan, 1974). By concentrating only on observations above the threshold, the POT approach offers an efficient and accurate mechanism for modelling tail behaviour compared to conventional methods that assess the entire distribution. Given a time series dataset $\{e_1, e_2, \dots, e_l\}$ and a threshold τ^* (any observations that exceed the threshold are called extreme events), the POT approach selects extreme events when $e_i > \tau^*$. The distribution of exceedances over the large threshold τ^* asymptotically follows a GP distribution. To explain it mathematically, let Z_1, Z_2, \dots, Z_n be a series of iid random variables with a marginal distribution $Q(\cdot)$. Pickands III (1975) approximated the exceedance distribution for sufficiently large threshold values using a GP distribution, defined by

$$\mathcal{H}(z) = \begin{cases} 1 - \left(1 + \frac{\zeta z}{\sigma} \right)^{-1/\zeta} & \text{if } \zeta \neq 0, \\ 1 - \exp \left(- \frac{z}{\sigma} \right) & \text{if } \zeta = 0. \end{cases} \quad (1)$$

Here, $\zeta \in \mathbb{R}$ is the shape parameter and $\sigma > 0$ is the scale parameter of the GP distribution. The shape parameter ζ plays a key role in determining the qualitative behaviour of the GP distribution and influences its domain of attraction. When $\zeta = 0$, $\mathcal{H}(z)$ belongs to the Gumbel distribution family, where the probability of extreme observations decreases exponentially, as indicated by its light tails. For $\zeta > 0$, $\mathcal{H}(z)$ follows a Fréchet distribution characterized by heavy tails, suggesting more frequent extreme observations. Conversely, when $\zeta < 0$, $\mathcal{H}(z)$ corresponds to a Weibull distribution with short tails, implying a lower probability of extreme observations. The POT approach offers a robust technique for effectively modelling extreme observations with minimum data loss. It is particularly suited for capturing the clustering effect, which is a prominent phenomenon in extreme events. The advantages of this method for modelling extreme air pollution levels are demonstrated in [AL-Dhurafi et al. \(2018\)](#), where the POT approach has been applied to investigate air pollution index exceedances in urban areas of Peninsular Malaysia.

2.3 Methods for threshold selection in POT approach

The choice of threshold plays a key role in identifying the extreme observations in the dataset, thus significantly impacting the effectiveness of the POT approach. If a low threshold is selected, usual observations can be treated as extreme and violate asymptotic assumptions. On the contrary, a high threshold value can overlook potential extreme observations by treating too few data points as extreme. The threshold selection can be done objectively through a bias-variance trade-off or determined subjectively, with input from domain experts. Among various statistical procedures, the mean excess plot (MEP) is a popular approach for determining the threshold in the POT method ([Benktander & Segerdahl, 1960](#)). The mean excess function of the random variable Z with distribution function $Q_Z(z)$ and right endpoint z_R is given by

$$ME(\tau^*) := E(Z - \tau^* \mid Z > \tau^*) = \int_{\tau^*}^{z_R} \left(\frac{1 - Q_Z(s)}{1 - Q_Z(\tau^*)} \right) ds,$$

provided $E(Z) < \infty$ ([Embrechts et al., 2013](#)). Thus, if we model the statistical properties of exceedance for any arbitrarily chosen random variable Z among Z_1, Z_2, \dots, Z_n , with GP (σ, ζ) distribution [as in Eq. (1)], then the expected value of Z will be finite if and only if $\zeta < 1$ and the mean excess function can be computed as

$$ME(\tau^*) = \frac{\sigma}{1 - \zeta} + \frac{\zeta}{1 - \zeta} \tau^*,$$

where $0 \leq \tau^* < \infty$ if $0 \leq \zeta < 1$ and $0 \leq \tau^* \leq -\frac{\sigma}{\zeta}$ if $\zeta < 0$. A natural estimate of the mean excess function, $\widehat{ME}(\tau^*)$, is defined by

$$\widehat{ME}(\tau^*) = \frac{\sum_{i=1}^n (z_i - \tau^*) I_{[z_i > \tau^*]}}{\sum_{i=1}^n I_{[z_i > \tau^*]}}; \quad \tau^* \geq 0,$$

where I denotes the indicator function. This method considers the set of all points $\{(\tau^*, \widehat{ME}(\tau^*)) : \tau^* < z_{(n)}\}$, where $z_{(n)}$ is the highest order statistic from the sample. In principle, the MEP will appear linear if the exceedance observations are fitted with a GP distribution, which has a finite mean ([Das et al., 2025](#)).

The MEP approach serves as an effective and objective tool for threshold selection in the absence of specific regulatory thresholds. For example, in a study on extreme influenza cases in Zhejiang, China, [J. Chen et al. \(2015\)](#) employed the MEP method to determine an optimal threshold for influenza incidence, owing to the lack of a standardized cut-off. This illustrates how MEP can successfully identify a threshold that balances the representation of both regular and extreme events. Similar applications can be found in fields such as hydrology ([Durocher et al., 2019](#)), finance ([Chukwudum et al., 2020](#)), and environmental science ([Ghosh & Resnick, 2010](#)), where predefined regulatory thresholds are absent. In contrast, our study examines air pollution concentrations where public health and policy alignment are critical. The NAAQS, specified by the

CPCB, offers scientifically established thresholds for air pollutant levels aimed at protecting public health. Although the MEP-derived threshold may be useful in the absence of regulatory cut-offs, it often corresponds to pollutant levels substantially higher than those defined by regulatory standards, reducing its practical utility. If the MEP threshold deviates significantly from the NAAQS, it may overlook numerous events of public-health concern. Therefore, we adopt the NAAQS threshold to ensure that our findings align with public-health policy and aid in designing intervention strategies.

3 Proposed methodology

This section introduces the proposed E-STGCN along with the modified STGCN method for spatio-temporal forecasting of air pollution concentration levels in the presence of extreme observations. Specifically, Section 3.1 outlines the mathematical formulation of the spatio-temporal air pollution forecasting problem, while Section 3.2 provides an overview of the E-STGCN architecture, with detailed descriptions of the components within each module of the proposed framework.

3.1 Problem formulation

In this study, we address the air quality prediction problem as a spatio-temporal forecasting task, where the key challenge is to model temporal patterns from historical data while simultaneously capturing the spatial relationships between multiple air quality monitoring stations. Let there be N monitoring stations and suppose $\mathbf{X}_t = [X_t^1, X_t^2, \dots, X_t^N] \in \mathbb{R}^N$ denote the vector of air pollutant concentrations across all stations at time t , where $t = [1, T]$. Our objective is to generate q -step-ahead ($q \geq 1$) forecasts $\{\hat{\mathbf{X}}_{T+1}, \hat{\mathbf{X}}_{T+2}, \dots, \hat{\mathbf{X}}_{T+q}\}$ based on T past observations $\mathcal{X}_{1:T} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T]^T \in \mathbb{R}^{T \times N}$ which represents the air pollutant concentration across T historical observations and N monitoring stations. To achieve this, we develop a forecasting model that integrates EVT with GCN and LSTM networks for modelling the spatio-temporal correlations among the N monitoring stations, accounting for the presence of extreme observations in the dataset. Our goal is to learn a forecasting function $F_{E-STGCN}$ that maps spatio-temporal inputs into a sequence of future predictions. To understand the objective formally, let $G = \{V, E\}$ denote an undirected graph, where $V \in \mathbb{R}^N$ represents the set of nodes, corresponding to the monitoring stations and the set of edges $E \in \mathbb{R}^{N \times N}$ indicates the spatial correlations between the stations. Mathematically, the air quality forecasting problem can then be expressed as

$$F_{E-STGCN} : (\mathcal{X}_{1:T}, A) \rightarrow \{\hat{\mathbf{X}}_{T+1}, \hat{\mathbf{X}}_{T+2}, \dots, \hat{\mathbf{X}}_{T+q}\},$$

where the adjacency matrix $A \in \mathbb{R}^{N \times N}$ encode the spatial connections and $\hat{\mathbf{X}}_{T+i} = (\hat{X}_{T+i}^1, \hat{X}_{T+i}^2, \dots, \hat{X}_{T+i}^N) \in \mathbb{R}^N$ represents the i -step-ahead ($i = [1, q]$) forecast of air pollution concentrations for the N monitoring stations, computed based on the T historical observations.

3.2 E-STGCN model overview

The overall architecture of the E-STGCN framework, depicted in Figure 1, consists of three primary modules: the spatial module, the temporal module, and the EVT module. The spatial module maps the input data onto attributed spatio-temporal graphs and learns the underlying spatial correlations. These learned graph structures and historical air pollutant concentrations are processed through the spatial blocks comprising GCNs and fully connected neural networks, which capture dynamic temporal information and spatial influences. The output from the spatial module is then fed into the temporal module, where the future trajectories of air pollutant concentrations are predicted using recurrent LSTM layers and a fully connected dense layer (modified to capture long-memory dynamics and differs from standard STGCN). The EVT module, another key component of the E-STGCN architecture, is specifically designed to model rare but impactful extreme pollution events. It fits a GP distribution to the historical air pollutant concentrations that exceed permissible thresholds. This fitted distribution is used to augment the conventional data-driven loss with POT loss (discussed in Section 3.2.3). The hybrid loss function ensures that predictions exceeding the regulatory

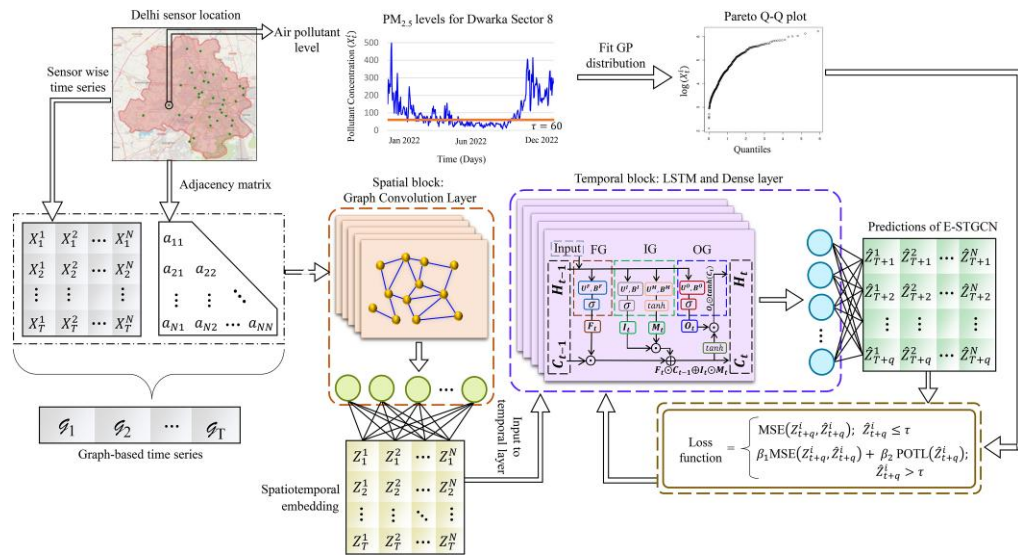


Figure 1. Extreme spatio-temporal graph convolutional networks (E-STGCNs). Daily air pollution concentration levels from different regions of Delhi, along with the corresponding adjacency matrix, are processed through a graph convolutional network (GCN) and a dense layer to generate spatio-temporal embeddings. To account for extreme values, each sensor’s time series data is modelled using a generalized Pareto (GP) distribution. The GCN-embedded data is then passed through an long short-term memory (LSTM) layer, followed by a dense layer, to produce accurate forecasts. The network is trained using a modified loss function that combines the conventional mean squared error (MSE) loss with a peaks-over-threshold loss (POTL) when predictions exceed a predefined threshold (τ).

thresholds are not only penalized for their squared error but are also encouraged to conform to the statistical structure of the learned tail distribution. The primary objective of E-STGCN is to capture extremes, not merely to be robust to them. Specifically, the model aims to accurately forecast both the occurrence and magnitude of rare pollution events, a crucial drawback of prior spatio-temporal deep learning models. The EVT module directly addresses this gap by embedding domain-specific tail behaviour into the training process. By integrating EVT-based statistical knowledge with the spatio-temporal representations learned from the spatial and temporal modules, E-STGCN can effectively model the dynamics of air pollutant concentrations, particularly in scenarios involving threshold exceedances. Our framework is conceptually aligned with the idea of PINN (Karniadakis et al., 2021; Raissi et al., 2019), where domain knowledge is embedded into neural networks via loss constraints (not architectural changes). PINN uses physical laws to inform model behaviour in regions where data is scarce or uncertain by taking the total loss as $Loss_{total} = Loss_{data} + \lambda Loss_{physics}$, where $Loss_{data}$ is the data-centric loss and $Loss_{physics}$ is the residual from the governing ordinary differential equations or partial differential equations. In E-STGCN, we use EVT to regularize the model where the data is sparse in the tail (certainly many extreme values). In this way, the GP distribution penalty serves as a probabilistic prior grounded in statistical theory, analogous to physical laws in the mechanistic model, thus allowing E-STGCN to accurately forecast extreme observations in pollution concentration levels.

3.2.1 Spatial module

In the spatial domain, the air pollutant concentrations at different sensor locations influence each other with varying intensities, and most interactions are dynamic. To capture the spatial correlations among the monitoring stations, we employ graph convolution operations. Typically, GCNs allow convolution operations on arbitrary graph structures, enabling the learning of node-order invariant representations. In the E-STGCN framework, we model the historical air pollutant concentrations using GCN by considering the geographical locations of the monitoring stations as nodes, which form the basis of spatial dependencies. Thus, the undirected graph $G = \{V, E\}$, with V nodes and E connecting edges, can be represented using an adjacency matrix A for efficient computer processing. Specifically, the adjacency matrix A is static and is constructed based on the

weighted Haversine distance (d_{ij}) between the geographical locations of the i th station (with latitude ϕ^i , longitude λ^i) and the j th station as

$$d_{ij} = 2R \sin^{-1} \left[\sqrt{\sin^2 \left(\frac{\Delta\phi}{2} \right) + \cos(\lambda^i) \cos(\lambda^j) \sin^2 \left(\frac{\Delta\lambda}{2} \right)} \right], \quad (2)$$

where $\Delta\phi = \phi^i - \phi^j$, $\Delta\lambda = \lambda^i - \lambda^j$, and R represents the earth's radius. The weighted adjacency matrix, indicating the similarity between the corresponding nodes, are computed using a Gaussian kernel as

$$a_{ij} = \exp\left(-\frac{d_{ij}^2}{\tilde{\sigma}^2}\right), \quad \text{when } i \neq j \quad \text{and} \quad \exp\left(-\frac{d_{ij}^2}{\tilde{\sigma}^2}\right) \geq \epsilon, \quad (3)$$

where $\tilde{\sigma}^2$ and ϵ are the parameters that control the distribution and sparsity of the adjacency matrix A . Specifically, if the distance between the nodes exceeds $\sqrt{-\tilde{\sigma}^2 \ln \epsilon}$, no edges are considered between the nodes.

With a slight abuse of terminology, let us represent the air pollutant concentrations monitored at N stations over T timestamps as spatio-temporal graphs $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_T\}$, where each graph $\mathcal{G}_t = \{\mathbf{X}_t, A\}$ consists of $\mathbf{X}_t \in \mathbb{R}^N$, representing the pollutant attributes at time t , and $A \in \mathbb{R}^{N \times N}$, providing the structural information for the N stations. To map the non-Euclidean spatio-temporal graphs to spatio-temporal node embeddings, we perform localized convolutions of the node neighbourhood using GCN layers. GCNs generalize the concept of CNN filters to graph-structured data by applying polynomial filters over neighbouring nodes. These filters can be approximated using Chebyshev polynomials of order d as

$$\mathcal{P}_w(L) = \sum_{u=0}^d w_u \mathcal{C}_u(\tilde{L}),$$

where \mathcal{C}_u denotes the u th Chebyshev polynomial and $\tilde{L} = \frac{2L}{\zeta_{\max}} - I_N$ is the normalized graph Laplacian. Here, $L = D - A$ is the Laplacian, D is the diagonal degree matrix with $D_{ii} = \sum_j A_{ij}$, and ζ_{\max} is the largest eigenvalue of L . The graph convolution on input data \mathbf{X}_t is then given by

$$\mathbf{X}'_t = \mathcal{P}_w(L)\mathbf{X}_t. \quad (4)$$

As shown by Kipf and Welling (2016), a first-order approximation ($d = 1$) is often sufficient and computationally efficient, leading to a simplified graph convolution. Hence, Eq. (4) can be simplified as

$$\mathbf{X}'_t = w_0 \mathbf{X}_t + w_1 \left(\frac{2L}{\zeta_{\max}} - I_N \right) \mathbf{X}_t, \quad (5)$$

where w_0 and w_1 are filter weights shared across all N nodes. By applying a stack of K different polynomial filter layers, which corresponds to a sequence of K graph convolution layers with first-order approximations, the spatio-temporal node embeddings for $\mathcal{G}_t = \{\mathbf{X}_t, A\}$; $t = 1, 2, \dots, T$, can be computed as

$$\begin{aligned} h_t^{i,(0)} &= X_t^i, \\ h_t^{i,(k)} &= f_t^{(k)} \left(W_t^{(k)} \frac{\sum_{j \in \mathcal{N}(i)} h_t^{j,(k-1)}}{|\mathcal{N}(i)|} + B_t^{(k)} h_t^{i,(k-1)} \right); \quad k = 1, 2, \dots, K, \\ Z_t^i &= \text{Dense} \left(h_t^{i,(K)} \right), \end{aligned}$$

where in the k th iteration, the function $f_t^{(k)}$ and filter weights $\{W_t^{(k)}, B_t^{(k)}\}$ are shared to update the initial embedding using 1-hop localized convolutions, repeated K times based on a neural message passing mechanism (Gilmer et al., 2017). Thus, $h_t^{i(k)}$ is the embedding of node i at timestamp t during iteration k , computed by taking the mean of its neighbouring nodes and its self-embedding from the previous iteration at time t . The final spatio-temporal representation, Z_t^i , from the spatial block is computed by modelling the GCN output from the K th layer using a fully connected dense layer. Consequently, the spatio-temporal embedding $\mathbf{Z}_t = [Z_t^1, Z_t^2, \dots, Z_t^N] \in \mathbb{R}^N$ generated from \mathcal{G}_t updates \mathbf{X}_t with the encoded information from $(K - 1)$ -order neighbourhood of the central node through K successive filtering operations. It is of the essence here to point out that the first-order approximation of the polynomial filter is highly effective and scalable for large-scale graph structures (Yu et al., 2018).

3.2.2 Temporal module

The temporal module of the E-STGCN framework is designed to model the spatio-temporal embeddings, $[\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_T]^T$, learned in the spatial block. Due to the complex sequential dependencies within $\{\mathbf{Z}_t\}$, we employ an LSTM network, a robust variant of Recurrent Neural Networks (RNNs), that efficiently overcomes optimization challenges of conventional RNN architectures (Hochreiter & Schmidhuber, 1997). This temporal module differs from the architecture of the standard STGCN, where 1-D convolutional layers with gating mechanisms are used. The LSTM network in E-STGCN introduces specialized memory cells that replace standard hidden units, improving stability, speed, and accuracy. These cells maintain self-connected recurrent edges with fixed weights, enabling stable gradient flow across long sequences. The cell state stores long-term information, while hidden states manage short-term context. A gating mechanism, comprising forget, input, and output gates, controls updates to both states, allowing the network to effectively capture and retain long and short-term dependencies in sequential data. At each timestamp t , for each of the N nodes, the LSTM receives p lagged values $\mathbf{z}_t^i = \{Z_{t-p+1}^i, Z_{t-p+2}^i, \dots, Z_t^i\} \in \mathbb{R}^p$ along with the previous hidden state vector H_{t-1}^i as input. It then generates the q -steps-ahead projections of Z_t^i along with the new memory M_t^i , updating both the hidden state H_t^i and the cell state C_t^i . The forget gate determines which parts of the previous cell state C_{t-1}^i should be retained by computing the forget gate activation as

$$F_t^i = \phi_1(U_{ZF}^i \mathbf{z}_t^i + U_{HF}^i H_{t-1}^i + B_F^i),$$

where $U_{ZF}^i \in \mathbb{R}^{m \times p}$, $U_{HF}^i \in \mathbb{R}^{m \times m}$, and $B_F^i \in \mathbb{R}^m$ are learnable parameters, m is the number of hidden layers, and ϕ_1 is a sigmoid function that constrains values of $F_t^i \in [0, 1]$. A value of F_t^i close to 1 retains the corresponding part of C_{t-1}^i ; while a value close to 0 discards it. The input gate controls how much new information from the current input \mathbf{z}_t^i should influence the cell state by computing the activation (I_t^i) and new memory vector (M_t^i) as

$$I_t^i = \phi_1(U_{ZI}^i \mathbf{z}_t^i + U_{HI}^i H_{t-1}^i + B_I^i), \quad M_t^i = \phi_2(U_{ZM}^i \mathbf{z}_t^i + U_{HM}^i H_{t-1}^i + B_M^i),$$

where ϕ_2 is tangent hyperbolic activation function, $U_{ZI}^i, U_{ZM}^i, U_{HI}^i, U_{HM}^i, B_I^i$, and B_M^i are learnable parameters. Using F_t^i, I_t^i , and M_t^i current cell state is updated as

$$C_t^i = F_t^i \odot C_{t-1}^i \oplus I_t^i \odot M_t^i,$$

where \odot denotes the element-wise multiplication. Finally, the current hidden state is calculated in the output gate based on the activation vector (O_t^i) of the output gate as

$$H_t^i = O_t^i \odot \phi_2(C_t^i), \quad \text{where } O_t^i = \phi_1(U_{ZO}^i \mathbf{z}_t^i + U_{HO}^i H_{t-1}^i + B_O^i),$$

with U_{ZO}^i , U_{HO}^i , and B_O^i being the output gate parameters. To compute H_t^i and C_t^i , the initial values are set to $H_0^i = C_0^i = 0$. Consequently, the q -steps-ahead forecast of the air pollutant concentrations for the i th node is obtained using a fully connected dense layer as

$$\{\widehat{Z}_{t+1}^i, \widehat{Z}_{t+2}^i, \dots, \widehat{Z}_{t+q}^i\} = \text{Dense}(H_t^i).$$

The final output generated by the temporal module effectively captures the sequential patterns of the air pollutant series. However, the model struggles to forecast sudden peaks, which are particularly common in Delhi's air pollutant concentrations during winter months. To address this, we design the EVT module within the E-STGCN architecture, enabling the framework to accurately forecast spatio-temporal dependencies in situations of threshold exceedances.

3.2.3 EVT module

In the field of air pollution control, Roberts (1979) emphasized that rare events often hold more significance than regular observations. Therefore, prior knowledge of these rare occurrences is crucial for accurate modelling and forecasting of air pollutant concentrations. The spatial and temporal modules of the E-STGCN architecture leverage historical pollutant data from various proximal monitoring stations to predict future trends. However, their inability to differentiate between common and rare events limits their effectiveness in modelling extreme occurrences. To address this issue, the EVT module, a key component of the E-STGCN framework, utilizes EVT to identify the underlying patterns of air pollutant concentrations associated with rare observations.

In the EVT module, we employ the POT approach to analyse the extreme observations and integrate them into the spatio-temporal forecasts of the previous modules. In the POT method (as discussed in Section 2.2), we examine the behaviour of exceedances by fitting a GP distribution to the pollutant concentrations that exceed the NAAQS threshold (τ). Following Eq. (1), the conditional GP distribution for the i th monitoring station at time t can be mathematically formulated as

$$P[X_t^i - \tau \leq x_t^i \mid X_t^i \geq \tau] = \begin{cases} 1 - \left(1 + \frac{\zeta^i x_t^i}{\sigma^i}\right)^{-1/\zeta^i} & \text{if } \zeta^i \neq 0, \\ 1 - \exp\left(-\frac{x_t^i}{\sigma^i}\right) & \text{if } \zeta^i = 0, \end{cases}$$

where $\zeta^i \in \mathbb{R}$ is the shape parameter and $\sigma^i > 0$ is the scale parameter for the GP distribution fitted to the pollutant concentrations of the i th monitoring station. The shape parameter ζ^i is particularly important as it influences the tail behaviour of the GP distribution. To estimate the shape and scale parameters of the GP distribution for the i th station, we consider the \tilde{k}^i observations $\{X_1^i, X_2^i, \dots, X_{\tilde{k}^i}^i\}$ exceeding the threshold τ and compute the log-likelihood function based on the threshold exceedance as follows:

$$l(\sigma^i, \zeta^i) = -\tilde{k}^i \log(\sigma^i) - \left(1 + \frac{1}{\zeta^i}\right) \sum_{m=1}^{\tilde{k}^i} \log\left(1 + \frac{\zeta^i x_m^i}{\sigma^i}\right),$$

provided $(1 + \frac{\zeta^i x_m^i}{\sigma^i}) > 0$ for $m = 1, 2, \dots, \tilde{k}^i$; otherwise $l(\sigma^i, \zeta^i) = -\infty$ (Coles et al., 2001; Grimshaw, 1993). Since the log-likelihood of the GP distribution lacks a closed-form analytical solution, we employ the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, a quasi-Newton optimization method, to estimate the parameters σ^i and ζ^i (Fletcher, 2000). The BFGS algorithm iteratively updates parameter estimates using gradient information and approximates the inverse Hessian to maximize the log-likelihood, while ensuring stability and convergence by enforcing constraints such as the positivity of σ^i and validity of the likelihood domain. The resulting estimates

$(\hat{\sigma}^i, \hat{\xi}^i)$ of the scale (σ^i) and shape (ξ^i) parameters obtained from the air pollutant concentration levels of the i th monitoring station enables the design of the POT loss function which serves as a prior information in the E-STGCN framework.

The POT loss function is computed as the negative log-likelihood of the fitted GP distribution. Now using a single predicted value of air pollution concentration level (\hat{Z}_t^i) for monitoring station i at time t that exceeds the threshold τ , we compute the negative log-likelihood (NLL) of the fitted GP distribution (POT loss) as

$$\text{POTL}(\hat{Z}_t^i) = \text{NLL} = \log(\hat{\sigma}^i) + \left(1 + \frac{1}{\hat{\xi}^i}\right) \log\left(1 + \frac{\hat{\xi}^i \hat{Z}_t^i}{\hat{\sigma}^i}\right). \tag{6}$$

We then incorporate the negative log-likelihood function [Eq. (6)] while designing the loss function of the temporal module. This approach enhances the modelling of threshold exceedances by incorporating knowledge from EVT as prior information to the model.

3.2.4 Optimization

The objective function of the E-STGCN framework is formulated as a combination of the data loss, computed by the mean squared error (MSE), and the POT-based loss function [Eq. (6)] depending on whether predicted values exceed a specified threshold τ . This allows the E-STGCN to capture the dynamics of air pollution, not just the average behaviour, but especially its rare and high-impact extremes (ones that violate regulatory thresholds by NAAQS). The hybrid loss function, combining MSE for general predictive accuracy with negative log-likelihood penalty from a GP distribution for threshold exceedance, in the form of POT loss, can be expressed as

$$\text{Loss}(Z_{t+q}^i, \hat{Z}_{t+q}^i) = \begin{cases} \text{MSE}(Z_{t+q}^i, \hat{Z}_{t+q}^i), & \hat{Z}_{t+q}^i \leq \tau, \\ \beta_1 \text{MSE}(Z_{t+q}^i, \hat{Z}_{t+q}^i) + \beta_2 \text{POTL}(\hat{Z}_{t+q}^i), & \hat{Z}_{t+q}^i > \tau, \end{cases} \tag{7}$$

where β_1 and β_2 are the hyperparameters that regulate the contributions of data loss and the POT-based loss, respectively. Since the loss function is differentiable almost everywhere, we utilize the backpropagation method to train the corresponding weights. Thus, for predictions below the threshold τ , the model is optimized solely based on MSE loss that draws the predictions toward the conditional mean. This is appropriate for the bulk of the data. However, for predictions above the threshold τ , the loss function becomes: $\beta_1 \text{MSE}(Z_{t+q}^i, \hat{Z}_{t+q}^i) + \beta_2 \text{POTL}(\hat{Z}_{t+q}^i)$, where the model is encouraged not just to minimize MSE but to conform to the statistical structure of extreme values learned from the data. Note that the GP distribution penalty term does not force the prediction to match the mean of the GP distribution, but it acts like a regularization term, encouraging the prediction to fall within the plausible tail shape defined by EVT. This acts as a distribution-aware regularization, but not a replacement for the predictive loss. Using only the GP distribution likelihood for exceedance would ignore the distance between observed and predicted values; therefore, it can significantly harm the point prediction accuracy (which is crucial in air pollution forecasting for policy intervention). Hence, the combined loss provides a trade-off where the model learns to be accurate (stay close to the observed values) while regularizing the tail behaviour via alignment with the empirical tail shape of the pollutant distribution. Additionally, to ensure an effective balance between data-based learning and the EVT knowledge, the coefficients β_1 and β_2 in the modified loss function are selected through a cross-validation approach by minimizing the root mean square error on the hold-out validation set (as indicated in Figure 2). These hyperparameters govern the trade-off between accurately forecasting average concentration levels and effectively capturing the extreme behaviour of the dataset. As a result, the E-STGCN model minimizes a modified loss function that captures the underlying spatio-temporal dynamics while also integrating distributional characteristics through the POT-based component to better model extreme concentration levels. A detailed visualization showcasing the working principle of the E-STGCN framework is provided in Figure 1. Although designed for air pollution, the proposed

hybrid loss function is model-agnostic and can be integrated into other spatio-temporal architectures for extreme forecasting tasks (e.g. finance, epidemics, etc.).

Remark 1 E-STGCN extends the foundational principles of the STGCN framework (Yu et al., 2018) by adopting a spatio-temporal learning paradigm that combines graph-based spatial modelling with temporal sequence processing. Both architectures utilize graph convolutional layers to capture spatial dependencies among monitoring stations using a predefined adjacency structure and are designed to forecast multivariate signals over a graph topology. Despite this shared foundation, E-STGCN introduces several important modifications, particularly in its temporal modelling strategy and training objective (loss function), that distinguish it from the original STGCN. The key differences are outlined below:

1. *Temporal Module*: STGCN uses 1-D convolutional layers coupled with gated linear units, which are effective at capturing local temporal patterns within a fixed window. However, this design is limited in its ability to model long-term memory dependencies, which are frequently observed in air quality time series due to seasonal variation, meteorological shifts, and policy interventions. To address these limitations, E-STGCN replaces the temporal convolutional block with LSTM units (sequence-to-sequence architecture), which are specifically designed to capture (both short- and long memory) dynamic temporal dependencies and preserve contextual information over extended horizons. This modification significantly improves classical STGCN's ability to model persistent pollution episodes and non-stationary temporal dynamics.
2. *Tail Modelling and Novel Loss Function*: E-STGCN incorporates a POT approach to explicitly separate moderate and extreme events, based on NAAQS regulatory thresholds. This allows the model to learn distinct dynamics for each regime and better capture extreme pollution events, which often exhibit different statistical behaviour from regular observations. E-STGCN introduces a hybrid loss function that combines the traditional MSE loss with a negative log-likelihood penalty derived from the GP distribution (POT loss), forming an EVT-based loss component. This additional term is activated only when the predicted values exceed a specified threshold τ , guiding the model to produce forecasts that are not only accurate on average but also statistically consistent with the empirical tail behaviour of the data. In contrast, standard STGCN (Yu et al., 2018) lacks any mechanism to explicitly handle extreme events and treats all prediction errors uniformly, regardless of their magnitude or severity. This EVT-guided learning mechanism of E-STGCN enables it to generate risk-aware forecasts, a critical gap in STGCN's design.

4 Experimental evaluation

In this study, we assess the efficiency of the proposed E-STGCN framework by comparing its forecasting performance with several temporal and spatio-temporal forecasters. We use daily data on Delhi's air pollutant concentration levels from 1 January 2019 to 31 December 2022, to train the models and generate forecasts for different months of 2023. To demonstrate the generalizability of our proposal, we evaluate its forecasting performance across three forecast horizons, namely short-term, medium-term, and long-term, spanning over 30 days, 60 days, and 90 days, respectively, using a rolling window approach. For the short-term horizon, forecasts are computed for each of the 12 months of 2023, separately. The forecast window covers two consecutive months in the medium-term horizon, resulting in six cases. There are four forecast windows for the long-term horizon, each covering three successive months. Figure 2 visually represents the training, validation, and test periods used in the forecasting tasks. The following subsections present a brief description of the air pollutant datasets and their global characteristics (Section 4.1), extreme

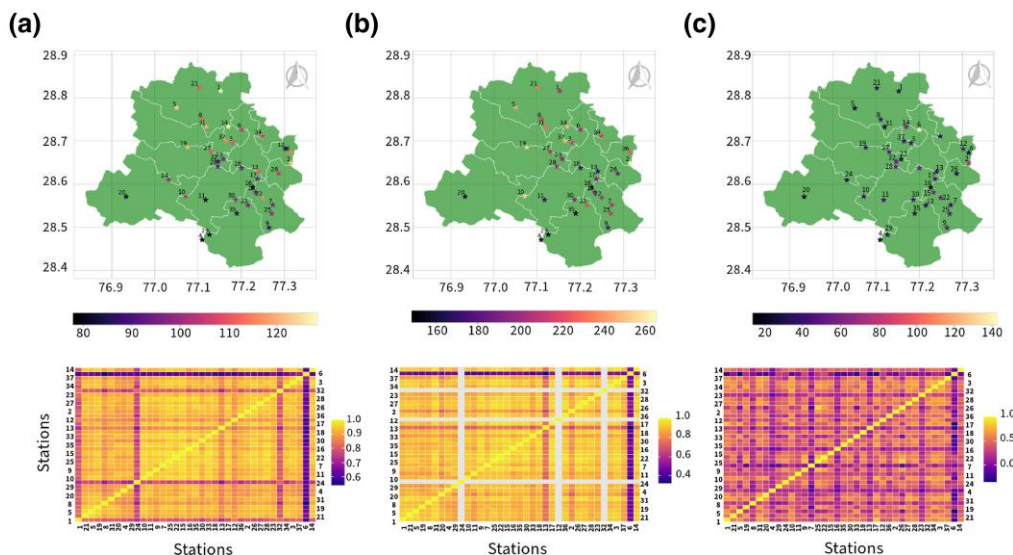


Figure 3. Upper panel: Spatial distribution of the monitoring stations in Delhi and average pollution level of (a) $PM_{2.5}$, (b) PM_{10} , and (c) NO_2 . Lower panel: Pairwise correlation between pollutant levels from each station for (a) $PM_{2.5}$, (b) PM_{10} , and (c) NO_2 . Stations are reordered such that geographically proximate locations appear adjacent to each other, emphasizing correlations driven by spatial proximity. For PM_{10} , stations 12, 24, and 32 did not record data and are represented as white cells in the heat map to denote missing observations.

patches farther from the diagonal correspond to station pairs that are geographically more distant and, hence, less correlated. Notably, $PM_{2.5}$ and PM_{10} concentrations show consistently strong positive correlations across most station pairs, while NO_2 exhibits moderate to strong correlations, primarily among nearby stations. Moreover, some geographically distant station pairs (e.g. stations 9 and 18; stations 21 and 23) also display high correlations, suggesting the influence of nonlocal emission sources. These spatial correlation structures reveal the complex interplay of local and regional factors influencing air quality in Delhi and provide critical insights for modelling both spatial and temporal dependencies in the forecasting process.

4.2 Extreme value modelling of air quality data

In this section, we employ the BM method and the POT approach to detect and model the extreme observations in the air pollutant concentrations. Figure 4 presents the results of extreme value analysis using the BM method for daily pollution concentrations of $PM_{2.5}$, PM_{10} , and NO_2 , measured at the Alipur monitoring station from 2019 to 2022. The other stations display similar behaviour as well. In our analysis, we consider a block size of 30 days, representing the maximum value in each block with a red circle and the remaining observations with green circles. From the plots, it can be observed that in several blocks, the maximum values are not necessarily extreme. Conversely, in other blocks, multiple extreme values, apart from the maximum, are abandoned by this method. To address these limitations, we employ the POT approach in our study. For determining the optimal threshold in the POT method, we utilize the MEP approach and demonstrate the results for pollution concentrations of $PM_{2.5}$, PM_{10} , and NO_2 monitored at the same station in Figure 5. The plot highlights the mean excess value for various thresholds (τ^*) with a 95% confidence interval. From the MEP, we can observe that the mean excess value becomes linear beyond the green straight line, indicating that the corresponding value represents the threshold. Specifically, the MEP-based thresholds are 583 for $PM_{2.5}$, 658 for PM_{10} , and 116 for NO_2 datasets. However, using these thresholds results in only 0.14% extreme values for $PM_{2.5}$, PM_{10} , and NO_2 dataset, which is insufficient for effective POT analysis. Nevertheless, in scenarios where extremely high observations are scarce, the MEP technique for threshold selection within the POT framework can be useful. For example, in case of NO_2 concentration levels from selected low-pollution monitoring stations in Delhi, specifically Station 4 (Aya Nagar), Station 16 (Lodhi

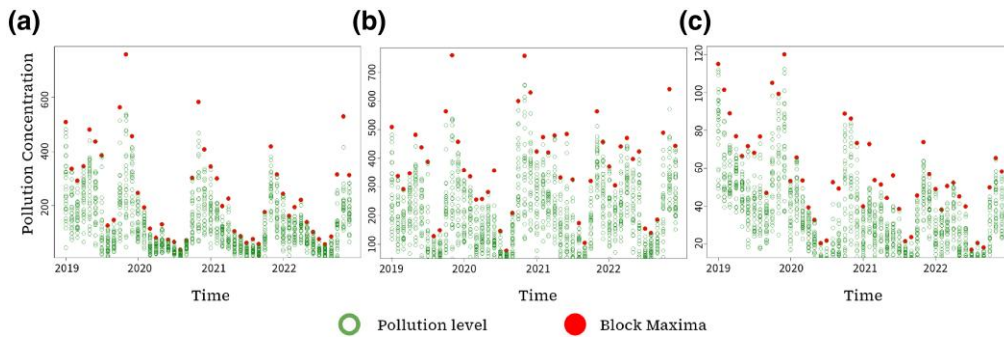


Figure 4. Block maxima plot for extreme value analysis of (a) $\text{PM}_{2.5}$, (b) PM_{10} , and (c) NO_2 pollutant concentration in Alipur, Delhi monitoring station with each month representing a block. Green points indicate the pollution levels, and red circles are the maximum values identified for each block.

Road IMD), and Station 20 (Najafgarh) of Delhi, where the frequency of extreme NO_2 levels ranges between 0.07% and 0.21% (as reported in [online supplementary material, Table A.3 of Appendix](#)), the MEP approach can be utilized. [Figure 6](#) displays the mean excess value computed for these datasets across varying thresholds (τ^*), along with 95% confidence intervals. The estimated MEP-based thresholds are $72 \mu\text{g}/\text{m}^3$ for Station 4, $74 \mu\text{g}/\text{m}^3$ for Station 16, and $80.5 \mu\text{g}/\text{m}^3$ for Station 20, all of which closely align with the NAAQS regulatory threshold of $80 \mu\text{g}/\text{m}^3$ for daily NO_2 exposure. These results demonstrate the practical validity of the MEP approach in relatively clean environments, where exceedance events are rare but still important to model. They further highlight the effectiveness of MEP in identifying reliable thresholds for tail modelling, even in low-exceedance regimes where capturing extreme behaviour remains critical and a regulatory threshold is absent (e.g. epidemic datasets).

In this study, we opt for a subjective method of threshold selection, utilizing the NAAQS established by the CPCB for industrial, residential, rural, and other areas. Domain experts determine these thresholds to protect public health, vegetation, and the environment. Following the NAAQS recommendation, we set the threshold values as $60 \mu\text{g}/\text{m}^3$ for $\text{PM}_{2.5}$, $100 \mu\text{g}/\text{m}^3$ for PM_{10} , and $80 \mu\text{g}/\text{m}^3$ for NO_2 pollutants and examine the exceedance of pollution concentration levels over these thresholds. From [online supplementary material, Tables A.1, A.2, A.3 of Appendix A.1](#), the average exceedance levels are 61% for $\text{PM}_{2.5}$, 77% for PM_{10} , and 10% for NO_2 . Additionally, to verify the iid assumption of the POT approach for these exceedance datasets, we perform the Durbin–Watson (DW) test ([Durbin & Watson, 1971](#)), which detects autocorrelation at lag 1 in the residuals from the regression analysis. The DW test p -values (refer to the above-mentioned [online supplementary material, tables in Appendix A.1](#)) indicate that for most exceedance time series, lag 1 residuals are uncorrelated. However, for certain stations with limited observations above the threshold, the DW test statistic could not be computed. We also demonstrate the fitting of the GP distribution for different air pollutant concentrations with the selected thresholds in [Figure 7](#). Due to the absence of enough extreme observations in the NO_2 dataset, the GP distribution does not provide a good fit as opposed to the $\text{PM}_{2.5}$ and PM_{10} datasets.

4.3 Forecasting performance evaluation measures

In our experimental evaluation, we utilize four deterministic and three probabilistic forecast performance indicators. The deterministic measures, including mean absolute error (MAE), mean absolute scaled error (MASE), root mean squared error (RMSE), and symmetric mean absolute percent error (SMAPE), quantify the average performance of different forecasters under various error formulations ([Hyndman, 2018](#)). Additionally, to evaluate how well E-STGCN captures extreme pollution behaviour, we include three probabilistic metrics, such as continuous ranked probability score (CRPS), threshold weighted CRPS (twCRPS), and Pinball Loss (quantile loss). These metrics provide insight into the model's ability to forecast not only central tendencies but also distributional tail characteristics, aligning with the goals of strictly proper scoring rules as emphasized by [Gneiting and Raftery \(2007\)](#) and [Gneiting and Ranjan \(2011\)](#). Specifically, the

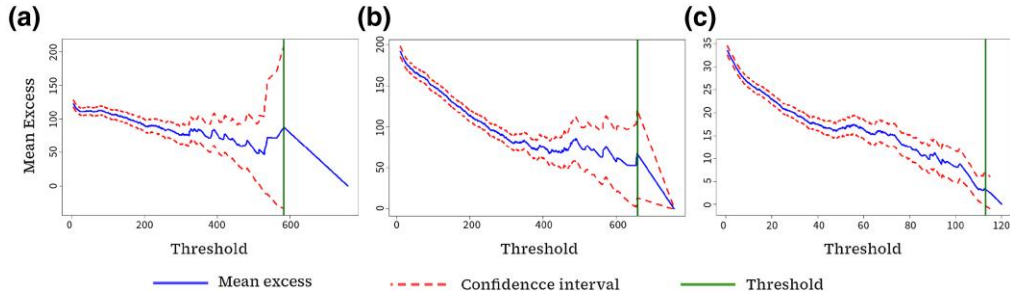


Figure 5. Mean excess plot (MEP) for (a) $\text{PM}_{2.5}$, (b) PM_{10} , and (c) NO_2 pollutant concentration in Alipur, Delhi monitoring station. The blue solid line indicates the mean excess level, the red dotted lines represent the 95% confidence interval, and the green solid line is the threshold obtained from the MEP.

quantile-based Pinball Loss evaluates performance at specific quantiles, the CRPS measures overall predictive performance across the entire distribution, and the twCRPS emphasizes model performance above a preselected threshold. The mathematical formulations of these metrics are as follows:

$$\begin{aligned} \text{MAE} &= \frac{1}{q} \sum_{t=1}^q |X_t^i - \widehat{X}_t^i|, & \text{MASE} &= \frac{\sum_{t=T+1}^{T+q} |\widehat{X}_t^i - X_t^i|}{\frac{q}{T-1} \sum_{t=2}^T |X_t^i - X_{t-1}^i|}, \\ \text{RMSE} &= \sqrt{\frac{1}{q} \sum_{t=1}^q (X_t^i - \widehat{X}_t^i)^2}, & \text{SMAPE} &= \frac{1}{q} \sum_{t=1}^q \frac{2|\widehat{X}_t^i - X_t^i|}{|\widehat{X}_t^i| + |X_t^i|} \times 100\%, \\ \text{Pinball Loss}(\rho^*) &= \frac{1}{q} \sum_{t=1}^q \max(\rho^*(X_t^i - \widehat{X}_t^i), (1 - \rho^*)(X_t^i - \widehat{X}_t^i)), \\ \text{CRPS} &= \frac{1}{q} \sum_{t=1}^q \left[\int_{-\infty}^{\infty} (F_t^i(\tilde{u}) - \mathbb{1}(\tilde{u} \geq X_t^i))^2 d\tilde{u} \right], & \text{and} \\ \text{twCRPS} &= \frac{1}{q} \sum_{t=1}^q \left[\int_{-\infty}^{\infty} (F_t^i(\tilde{u}) - \mathbb{1}(\tilde{u} \geq X_t^i))^2 w(\tilde{u}) d\tilde{u} \right], \end{aligned}$$

where q denotes the forecast horizon, ρ^* is the quantile, \widehat{X}_t^i is the forecast of the actual value X_t^i , F_t^i is the predicted probability distribution function (PDF) for the i th station at time t , T is the size of the training sample, and $w(\tilde{u}) = \mathbb{1}(\tilde{u} \geq \tau)$ is the weight function with τ representing the NAAQS threshold. In our analysis, we consider two quantiles at $\rho^* = \{0.8, 0.95\}$ and report the values of the Pinball Loss based on 80% and 95% quantiles (Gneiting et al., 2023). By definition, the minimum value of these performance measures suggests the ‘best-fitted’ model.

4.4 Benchmark forecasting models

To establish the efficacy of the proposed algorithm, we rigorously evaluate the model against state-of-the-art approaches. A list of these competing methods and their modelling capabilities is summarized in Table 1. Among time-dependent models, we consider the ubiquitous ARIMA approach (Box et al., 1970) as well as several deep learning techniques, including LSTM (Hochreiter & Schmidhuber, 1997), TCN (Y. Chen et al., 2020), DeepAR (Salinas et al., 2020), Transformers (Wu et al., 2020), and NBeats (Oreshkin et al., 2019). For the spatio-temporal models, we evaluate the performance of space–time autoregressive moving average (STARMA) (Pfeifer & Deutrich, 1980), generalized space–time autoregressive (GSTAR) (Cliff & Ord, 1975), fast Gaussian process (GpGp) (Guinness, 2018), STGCN (Yu et al., 2018), spatio-temporal neural network (STNN) (Saha et al., 2020), modified STGCN (introduced here), and DeepKriging (Nag et al., 2023). In

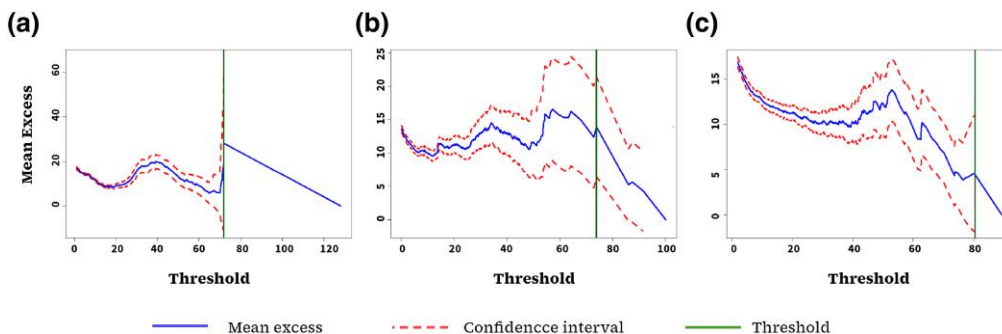


Figure 6. Mean excess plot (MEP) for NO₂ pollutant concentration monitored in (a) Station 4 (Aya Nagar), (b) Station 16 (Lodhi Road IMD), and (c) Station 20 (Najafgarh) of Delhi. The blue solid line indicates the mean excess level, the red dotted lines represent the 95% confidence interval, and the green solid line is the threshold obtained from the mean excess plot (MEP).

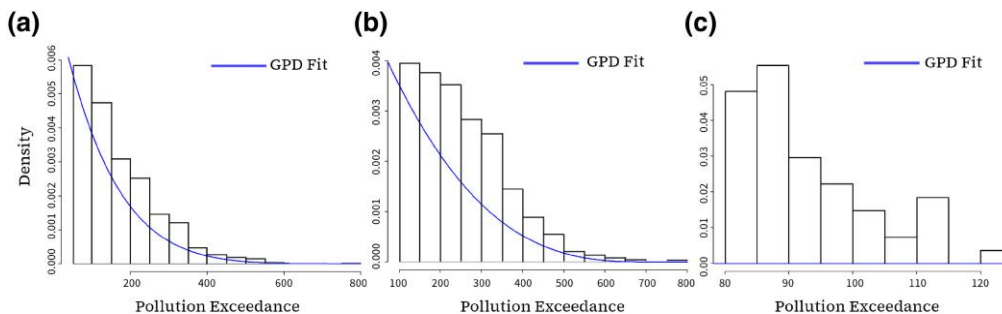


Figure 7. (a)–(c) Probability density plots of PM_{2.5}, PM₁₀, and NO₂ pollutant concentration extremes in Alipur, Delhi monitoring station, respectively. All histograms are fitted with the probability density (blue) of the generalized Pareto distribution.

the interest of space, more details of these models are provided in [online supplementary material, Appendix A.2](#).

4.5 Experimental setup and forecasting accuracy

In this section, we discuss the implementation of the proposed E-STGCN approach for forecasting air pollutant concentrations in Delhi. To train the sequential workflow of our model, we first utilize the ‘fgpd’ function from the *evmix* package in R. This function computes the maximum likelihood estimates for the scale parameter ($\sigma^i > 0$) and the shape parameter ($\zeta^i \in \mathbb{R}$) of the GP distribution, based on the training dataset for the i th station whenever an exceedance over the NAAQS threshold occurs. These estimated parameters provide prior information regarding extreme values in the training data. Subsequently, we implemented the E-STGCN model in Python to generate the spatio-temporal forecasts for the proposed approach. For modelling the spatial dependencies in the dataset, we compute the adjacency matrix (A) based on the weighted Haversine distance, as in Eq. (3). This matrix identifies the neighbours for each sensor, organizing their locations into a graphical structure by identifying relevant nodes and edges. Next, we employ CNNs and a dense layer from the *TensorFlow* library to encode the training data’s structural and feature-based information. To model the temporal dependencies, the output of the spatial module is passed through an LSTM layer and a dense layer. The weights of the temporal layer are optimized using a custom loss function, which combines the MSE loss with a POT-based loss [as in Eq. (7)]. This modified loss function leverages prior information about the NAAQS exceedances to enhance the accuracy of air pollution forecasts. Once the E-STGCN model and other benchmark forecasters are implemented, we generate out-of-sample forecasts using a rolling window approach for different forecast horizons. Below, we summarize the performance of our proposal and baseline models from temporal and spatio-temporal paradigms based on several key performance

Table 1. Comparison of forecasting frameworks

Models	Spatio-temporal	Nonlinear	Nonstationarity	Probabilistic forecasting	Scalability	Extreme value handling
ARIMA (Box et al., 1970)	X	X	✓	✓	X	X
LSTM (Hochreiter & Schmidhuber, 1997)	X	✓	✓	X	✓	X
TCN (Y. Chen et al., 2020)	X	✓	✓	X	✓	X
DeepAR (Salinas et al., 2020)	X	✓	✓	X	✓	X
Transformers (Wu et al., 2020)	X	✓	✓	X	✓	X
NBeats (Oreshkin et al., 2019)	X	✓	✓	X	✓	X
STARMA (Pfeifer & Deutrch, 1980)	✓	X	X	✓	X	X
GSTAR (Cliff & Ord, 1975)	✓	X	✓	X	X	X
GpGp (Guinness, 2018)	✓	X	X	✓	✓	X
STGCN (Yu et al., 2018)	✓	✓	✓	X	✓	X
STNN (Saha et al., 2020)	✓	✓	✓	X	X	X
DeepKriging (Nag et al., 2023)	✓	✓	✓	✓	X	X
Modified STGCN	✓	✓	✓	✓	✓	X
Proposed E-STGCN	✓	✓	✓	✓	✓	✓

Note. The columns indicate whether each model can address spatio-temporal correlations, nonlinearity, and nonstationarity in time series data. Additional columns assess whether the method at its core can produce probabilistic forecasts, scalability (size of the data does not pose a problem), and handle extreme observations.

indicators. A brief description of the benchmark temporal and spatio-temporal baseline models used in the experimental analysis, along with their implementation details, is outlined in [online supplementary material, Appendix A.2](#).

Tables 2, 3, and 4 present the performance of the proposed model and the baseline architectures in generating short-term forecasts for PM_{2.5}, PM₁₀, and NO₂ levels, respectively, in terms of selected performance indicators. The results for the remaining evaluation metrics are summarized in [online supplementary material, Appendix A.3](#) to maintain the readability of the tables. As indicated in Table 2, the proposed E-STGCN model achieves state-of-the-art performance for several months of 2023. In particular, during the onset (November) and end (February) of winter, E-STGCN generates the most accurate forecasts and efficiently captures the overall variability, including extreme observations, for the PM_{2.5} concentration levels. While NBeats and ARIMA outperformed the proposed framework in December and January, their ability to capture threshold exceedances remains limited than that of the E-STGCN model, as highlighted by the twCRPS metric. Additionally, the E-STGCN model improved ARIMA's forecast by 22.4% (based on the MAE metric) in March, and its performance remained competitive with NBeats. In April, our proposal shows a similar trend by efficiently capturing the extreme values in the dataset, while its average performance remains comparable to ARIMA and NBeats. During the summer months, from May to September, the proposed framework outperformed the benchmark models across most of the deterministic and probabilistic performance indicators, except in July, when LSTM produced more accurate average forecasts, and E-STGCN provided the most precise predictions for extreme values. In October, GpGp and Transformers recorded the lowest average forecast error; however, E-STGCN recorded the least values for twCRPS and Pinball Loss (95%) metrics. For short-term forecasting of PM₁₀, the E-STGCN model consistently performed best for the first three months of

2023, as measured by most point-based and distribution-based performance metrics. During April and May, the PM_{10} concentration levels rarely exceeded the NAAQS threshold, leading to similar performance between the E-STGCN and modified STGCN models, as the framework was trained primarily on the data loss. The average performance of the STNN model is better during June; however, the E-STGCN framework provides a better performance in capturing the tail distribution of the data. In July and December, although the NBeats framework outperforms all the forecasting approaches, our proposed model regains its forecasting superiority from August to October, as indicated by the average and distributional error metrics. In November, the spatio-temporal GpGp model showed competitive performance with our method. For 30-day ahead forecasts of NO_2 concentration levels, the E-STGCN and modified STGCN models showed similar performance. Since the average exceedance of NO_2 levels over the NAAQS threshold was around 10%, the use of a POT-based loss function was limited. Nevertheless, both architectures outperformed the standard STGCN, underscoring the effectiveness of incorporating an LSTM-based temporal module in capturing long-memory dynamics of air pollutant datasets. As shown in Table 4, the E-STGCN and modified STGCN models provided the lowest forecast errors in several months, including January, February, April, May, and August–November. For March and December, the STARMA, ARIMA, and NBeats frameworks outperform the competitive forecasting models, whereas in July, the ESTCGN framework accurately predicts the tail behaviour of the dataset.

The 60-day and 90-day ahead forecasting results, as presented in Tables 5, 6, and 7, demonstrate how the proposed E-STGCN architecture improves upon the baseline models for longer forecast horizons. For both $PM_{2.5}$ and PM_{10} pollutants, our model delivers the most accurate forecasts and quantifies the extreme observations during the first two 60-day windows, improving forecast accuracy by 9.73% over the best-performing baseline model. In the subsequent two forecast periods (May–June and July–August), the GSTAR model performs best for $PM_{2.5}$ levels, while for PM_{10} , the proposed model provides similar average performance as the GpGp and NBeats frameworks and outperforms others in capturing the extreme patterns in the dataset. During the September–October period, the E-STGCN framework achieves the lowest forecast error for both pollutants. For the final medium-term forecast period of 2023, the ARIMA and the E-STGCN model surpass the performance of all other approaches in terms of point-based and distribution-based measures, respectively; however, their performance lagged behind E-STGCN in modelling the extreme pollution levels during this period. For the medium-term and long-term forecasting of NO_2 concentration levels, we observe similar patterns to those seen in the short-term forecasts. The proposed E-STGCN and modified STGCN models generate similar results and outperform the baseline models in most periods, except for July–August (in the medium-term) and the last two long-term forecast windows, where NBeats, DeepAR, and ARIMA perform better. For other forecast windows, the STARMA and GSTAR models offer competitive performance compared to the best-performing frameworks. For the long-term forecasting task of $PM_{2.5}$, the E-STGCN architecture performs well during the first and last quarters of 2023, when the data predominantly exhibits more extreme behaviour. During the summer months, the NBeats and DeepAR frameworks provide more accurate forecasts of $PM_{2.5}$ concentration levels. In case of long-term forecasting of PM_{10} , the proposed framework outperforms the competing models in capturing overall concentration trends during the first two and the last quarters of 2023. However, during the third forecast window (July–September), DeepAR, ARIMA, and NBeats achieve superior performance for forecasting PM_{10} levels.

The experimental results reported in our study align with the *No Free Lunch* theorem, which suggests that any forecasting model performing best on a particular dataset is likely to perform poorly on others (Wolpert & Macready, 1997). Overall, the E-STGCN framework consistently achieved superior forecast performance across most tasks. Among the temporal models, ARIMA and NBeats performed well, while from the spatio-temporal paradigm, most of the baseline architectures, namely STARMA, GSTAR, GpGp, STNN, and STGCN, demonstrated competitive performance. The DeepKriging framework, however, performed poorly in most forecasting tasks due to scalability issues, which hindered its ability to handle medium-sized spatio-temporal datasets. Additionally, the performance of models like LSTM, TCN, DeepAR, and Transformers lagged behind the E-STGCN framework due to their inability to capture the spatial dependencies associated with pollutant transmission. We also observed that the proposed E-STGCN consistently

Table 2. Forecasting performance of the proposed extreme value theory guided modified spatio-temporal graph convolutional networks (E-STGCN) model in comparison to the temporal-only and spatio-temporal forecasting techniques for 30-day ahead forecast horizon of PM_{2.5} pollutant (best results are given in bold)

Forecast Period	Metric	Temporal-only model										Spatio-temporal model							Proposed E-STGCN
		ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified STGCN	DeepKrigging					
JAN	MAE	54.88	160.54	108.88	164.82	93.72	61.72	95.87	63.48	86.09	66.51	65.90	69.11	179.16	56.07				
	MASE	0.96	2.85	1.94	2.93	1.63	1.10	1.67	1.10	1.49	1.16	1.18	1.20	3.21	0.98				
	RMSE	71.18	174.93	137.98	178.87	114.86	78.46	112.01	82.27	107.61	84.94	84.55	90.21	192.41	75.34				
FEB	SMAPE	31.20	152.10	79.80	160.70	60.40	36.50	69.40	39.50	53.30	39.01	37.50	41.50	187.10	31.60				
	Pinball Loss	35.68	128.43	79.47	131.86	74.32	41.03	76.03	47.16	66.68	49.42	38.33	52.14	143.33	39.43				
	CRPS	78.95	113.95	117.19	113.95	113.43	89.85	110.15	94.29	111.56	96.64	93.40	102.28	114.01	83.20				
MAR	MAE	36.26	91.75	74.64	96.10	37.45	38.97	73.25	70.54	37.82	36.04	143.11	36.78	142.16	30.00				
	MASE	1.23	3.11	2.57	3.27	1.27	1.35	2.47	2.40	1.31	1.23	5.27	1.26	4.95	1.01				
	RMSE	45.68	101.86	96.43	105.83	50.23	49.71	88.40	85.35	49.05	47.79	168.10	46.61	224.10	38.81				
APR	SMAPE	32.00	127.70	74.70	139.70	33.10	34.00	89.10	83.80	33.50	31.87	76.60	32.50	167.80	26.70				
	Pinball Loss	17.44	73.40	42.27	76.88	24.60	20.09	58.44	56.29	21.67	22.58	29.29	17.88	105.34	15.86				
	CRPS	50.74	71.20	75.02	71.21	54.57	54.71	72.02	71.81	53.89	52.50	115.34	51.70	92.99	46.22				
MAY	MAE	33.32	56.96	61.43	60.59	25.88	25.90	36.81	31.18	31.99	30.40	74.22	37.43	76.95	25.83				
	MASE	1.70	2.78	3.03	2.98	1.35	1.24	1.79	1.53	1.75	1.60	3.93	1.94	3.83	1.27				
	RMSE	38.68	61.09	80.25	64.47	30.17	31.38	42.52	35.75	36.87	35.75	88.74	42.73	80.11	31.84				
JUN	SMAPE	38.60	108.50	76.50	120.90	31.90	30.00	59.10	44.70	38.20	36.24	63.10	42.00	193.20	35.30				
	Pinball Loss	7.21	45.57	31.55	48.47	7.76	7.90	27.62	20.59	8.40	8.75	15.63	8.68	61.56	7.72				
	CRPS	39.74	47.91	57.34	47.91	34.81	35.24	45.98	42.02	38.08	38.21	68.44	42.45	47.93	39.80				
JUL	MAE	22.49	47.35	47.05	51.54	33.61	22.44	50.17	37.83	36.48	38.27	46.06	25.31	68.90	25.37				
	MASE	1.22	2.52	2.57	2.75	1.94	1.20	2.68	2.02	2.16	2.18	2.52	1.38	3.75	1.35				
	RMSE	27.92	53.81	59.37	57.56	38.34	28.55	57.09	45.31	41.67	43.10	53.42	30.26	73.88	33.21				
AUG	SMAPE	32.20	93.50	82.00	108.50	45.10	33.00	107.10	68.90	48.40	48.92	71.60	35.90	195.20	38.20				

(continued)

Table 2. Continued

Forecast Period	Metric	Temporal-only model							Spatio-temporal model							Proposed E-STGCN
		ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified STGCN	DeepKrigging		
MAY	Pinball Loss	10.43	37.86	30.94	41.23	10.21	14.01	40.11	29.17	10.56	10.37	28.91	10.19	55.04	10.08	
	CRPS	37.21	43.49	47.57	43.48	37.75	37.15	43.67	43.50	39.32	40.42	46.46	37.42	43.63	37.12	
	MAE	35.10	45.77	47.65	49.80	39.24	30.31	42.19	35.32	40.23	42.47	42.10	34.01	85.44	28.95	
	MASE	1.43	1.83	1.93	2.00	1.63	1.22	1.69	1.42	1.72	1.75	1.69	1.39	3.52	1.16	
	RMSE	42.79	56.97	62.95	60.38	45.89	37.60	52.70	45.30	47.76	49.82	53.44	40.22	100.97	37.37	
JUN	SMAPE	48.40	85.80	83.90	100.40	53.70	44.20	80.10	59.30	53.70	55.87	76.60	48.70	141.70	43.80	
	Pinball Loss	11.73	36.55	32.31	39.83	11.32	14.87	33.20	26.50	11.32	12.32	33.06	12.17	59.68	11.30	
	CRPS	41.50	44.36	47.31	44.35	42.15	38.98	44.11	43.11	43.95	45.37	44.32	39.76	52.59	38.94	
	MAE	23.59	22.77	31.88	26.79	53.82	33.72	27.43	24.78	54.13	27.94	29.15	22.14	43.22	22.03	
	MASE	2.20	1.99	2.89	2.39	5.22	3.08	2.47	2.24	5.34	2.66	2.60	2.03	3.97	2.04	
JUL	RMSE	27.23	27.32	40.01	30.91	55.83	40.94	30.66	28.29	57.68	32.84	33.82	27.25	46.10	26.17	
	SMAPE	45.10	58.40	76.80	75.10	78.80	54.00	85.50	73.60	78.00	52.91	65.70	45.30	174.60	42.80	
	Pinball Loss	6.85	18.11	18.35	21.39	11.10	8.16	21.94	19.80	11.37	9.29	17.32	9.77	34.57	6.82	
	CRPS	26.06	27.51	28.62	28.16	39.23	33.94	27.80	27.49	41.02	31.62	30.42	26.97	28.19	25.17	
	MAE	23.00	12.98	22.36	16.23	64.44	17.71	20.32	16.44	60.83	32.21	17.16	18.85	92.72	15.69	
AUG	MASE	3.20	1.63	2.84	2.00	8.76	2.47	2.47	1.99	8.39	4.60	2.09	2.53	12.38	2.12	
	RMSE	26.16	15.84	28.49	19.05	65.33	21.69	23.09	19.42	63.79	34.43	20.20	23.40	107.91	18.87	
	SMAPE	53.20	41.20	79.40	55.30	99.60	43.00	81.20	60.20	94.60	69.07	56.70	46.10	133.90	41.10	
	Pinball Loss	5.07	9.72	14.59	12.71	12.90	4.55	16.09	12.50	12.18	10.02	12.07	4.67	47.89	4.41	
	CRPS	21.87	18.25	22.84	20.49	40.58	20.01	20.95	20.20	40.44	25.73	20.90	21.56	52.53	18.23	
AUG	MAE	20.00	16.06	25.41	19.67	60.41	18.38	20.52	25.65	58.39	51.29	14.87	14.51	35.37	11.85	
	MASE	3.25	2.30	3.87	2.87	10.15	2.98	3.06	3.84	9.78	8.63	2.30	2.38	5.40	1.84	
	RMSE	22.20	19.01	30.73	22.42	61.48	22.64	24.71	28.93	59.55	52.64	18.41	17.73	37.24	15.00	

(continued)

Table 2. Continued

Forecast Period	Metric	Temporal-only model										Spatio-temporal model							Proposed E-STGCN
		ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified STGCN	DeepKrigging					
SEP	SMAPE	44.00	45.90	88.60	61.10	90.70	38.40	67.40	94.20	89.20	82.54	36.60	33.90	160.80	31.00				
	Pinball Loss	4.66	12.54	18.20	15.63	12.17	5.59	15.65	20.45	11.75	11.40	6.58	5.90	28.29	5.99				
	CRPS	20.38	21.45	24.83	23.31	39.88	21.37	24.02	23.97	38.91	35.56	20.01	18.73	23.71	17.38				
OCT	MAE	22.48	20.76	30.91	24.06	53.97	19.66	22.97	22.30	52.60	56.07	28.39	17.33	35.54	15.06				
	MASE	2.46	2.15	3.26	2.50	6.01	2.12	2.41	2.34	5.84	6.25	2.99	1.88	3.74	1.64				
	RMSE	26.80	25.72	37.13	29.06	57.04	24.10	30.48	29.16	56.34	59.07	33.63	21.73	40.25	18.90				
NOV	SMAPE	49.90	57.10	94.50	70.30	84.10	44.80	67.70	63.90	82.50	85.41	91.50	42.30	139.20	38.60				
	Pinball Loss	5.11	15.62	21.00	18.66	10.95	7.06	16.80	15.88	10.66	12.39	21.24	5.96	28.39	4.91				
	CRPS	24.43	26.35	28.67	26.68	39.72	23.40	27.57	27.29	40.13	40.62	26.92	21.46	26.73	19.65				
DEC	MAE	35.69	75.37	76.73	79.84	35.38	41.20	67.20	65.33	33.87	36.12	84.67	36.51	68.69	36.37				
	MASE	1.93	4.19	4.33	4.45	1.98	2.25	3.72	3.63	1.89	1.98	4.77	2.04	3.89	1.99				
	RMSE	47.38	85.49	90.79	89.45	42.93	52.15	83.63	81.76	42.76	46.58	94.30	43.08	76.29	44.76				
NOV	SMAPE	35.90	111.30	116.80	124.00	35.70	44.90	90.90	86.90	34.00	36.89	145.70	36.70	62.10	36.70				
	Pinball Loss	26.47	60.30	60.76	63.87	18.36	29.21	53.67	52.13	18.59	24.77	67.64	20.87	26.30	20.83				
	CRPS	56.05	63.26	64.88	63.27	47.22	56.02	64.94	64.86	49.04	52.69	63.24	50.68	63.59	49.06				
DEC	MAE	102.46	216.26	177.96	219.77	147.62	119.58	141.29	143.39	136.72	142.82	218.78	95.48	239.19	77.08				
	MASE	1.77	3.74	3.09	3.80	2.52	2.06	2.42	2.46	2.32	2.44	3.77	1.65	4.15	1.33				
	RMSE	119.19	228.49	197.04	231.84	164.44	137.38	157.71	159.42	153.82	159.83	230.81	111.04	250.58	90.35				
NOV	SMAPE	49.30	159.30	116.60	164.40	81.10	61.80	81.10	81.80	72.50	77.45	168.90	45.40	198.10	34.60				
	Pinball Loss	79.49	173.01	142.28	175.82	117.81	93.07	112.22	114.07	108.65	114.85	172.29	69.16	191.35	51.83				
	CRPS	142.51	148.22	149.34	148.23	147.73	145.81	147.36	147.19	146.60	147.55	145.67	133.40	148.31	110.10				
DEC	MAE	74.03	176.33	134.41	181.01	106.47	68.77	97.15	106.11	96.54	81.80	159.13	104.60	194.85	116.77				
	MASE	1.97	4.76	3.63	4.89	2.85	1.84	2.59	2.86	2.55	2.20	4.37	2.82	5.26	3.15				

(continued)

Table 2. Continued

Forecast Period	Metric	Temporal-only model						Spatio-temporal model						Proposed	
		ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified STGCN	DeepKrigging	E-STGCN
	RMSE	64.04	186.95	163.73	191.37	122.82	89.40	123.92	133.80	116.90	102.78	170.62	121.14	205.76	132.24
	SMAPE	39.80	151.90	81.40	159.90	66.10	35.90	58.20	68.30	56.80	46.32	139.00	64.90	183.90	76.60
	Pinball Loss	50.95	141.06	88.75	144.81	85.07	46.63	71.73	84.21	74.21	65.71	115.13	83.60	155.88	93.39
	CRPS	106.68	123.46	133.07	123.46	122.15	105.74	124.31	125.80	120.71	124.10	116.47	122.10	123.88	122.61

Note. In the table pinball loss at 80% quantile is reported.

Table 3. Forecasting performance of the proposed extreme value theory guided modified spatio-temporal graph convolutional networks (E-STGCN) model in comparison to the temporal-only and spatio-temporal forecasting techniques for 30-day ahead forecast horizon of PM₁₀ pollutant (best results are given in bold)

Forecast Period	Metric	Temporal-only model										Spatio-temporal model						Proposed E-STGCN
		ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified STGCN	DeepKrigging				
JAN	MAE	99.31	276.07	262.61	275.41	165.16	87.24	121.40	82.99	86.54	94.94	115.31	89.89	233.97	82.67			
	MASE	1.23	3.46	3.31	3.45	2.05	1.10	1.49	1.02	1.09	1.18	1.47	1.12	2.94	1.03			
	RMSE	126.85	294.14	307.81	293.53	192.87	112.07	147.81	110.33	109.86	123.03	145.59	119.16	273.87	108.41			
	SMAPE	35.90	175.50	124.70	174.60	70.50	30.80	47.10	29.70	30.30	34.22	37.00	32.00	120.60	29.10			
	Pinball Loss	73.38	220.85	197.76	220.33	131.76	49.21	95.66	58.43	44.33	66.80	38.69	67.15	174.11	55.12			
FEB	CRPS	148.08	181.85	191.48	181.85	181.60	126.42	166.91	133.98	125.78	140.99	149.49	142.56	193.51	123.22			
	MAE	59.32	218.02	212.75	216.10	102.77	68.89	157.65	153.76	106.83	57.98	77.69	70.35	201.06	56.49			
	MASE	1.25	4.72	4.58	4.68	2.14	1.50	3.38	3.32	2.22	1.23	1.75	1.48	4.41	1.21			
	RMSE	77.99	228.33	247.69	226.50	122.29	83.55	175.50	171.40	125.72	74.10	101.26	88.41	223.28	71.36			
	SMAPE	25.40	172.10	116.00	169.00	49.80	29.50	96.70	93.30	52.60	24.96	32.20	32.20	140.50	24.30			
MAR	Pinball Loss	42.62	174.42	151.88	172.88	81.67	34.85	126.07	122.95	85.05	38.28	36.88	51.30	158.45	31.21			
	CRPS	96.23	143.51	157.80	143.52	135.78	101.07	144.43	144.32	137.02	92.88	109.52	111.50	149.42	90.31			
	MAE	51.41	159.77	168.58	154.64	52.33	52.00	54.18	51.87	60.80	49.95	252.19	56.17	167.16	42.54			
	MASE	1.29	3.93	4.20	3.80	1.23	1.31	1.30	1.26	1.61	1.26	6.54	1.39	4.12	1.05			
	RMSE	62.16	166.91	203.62	162.02	63.41	64.42	64.10	61.11	71.00	60.05	309.41	70.32	174.16	51.96			
APR	SMAPE	29.60	169.10	115.00	158.10	31.30	29.50	33.40	31.60	34.40	29.01	78.20	36.20	185.20	25.30			
	Pinball Loss	14.42	127.82	113.45	123.71	37.08	14.23	34.13	30.48	18.40	16.54	52.64	40.34	133.73	16.90			
	CRPS	73.61	105.90	117.70	105.90	82.38	75.46	80.74	77.35	78.00	72.26	212.69	89.06	105.94	66.12			
	MAE	59.82	181.01	171.35	179.36	71.49	67.72	151.14	106.86	61.34	55.68	62.35	53.04	229.48	53.04			
	MASE	1.37	4.18	3.89	4.14	1.61	1.54	3.46	2.45	1.42	1.31	1.45	1.24	5.38	1.24			
APR	RMSE	75.86	191.41	191.06	189.86	88.30	85.42	165.07	124.18	74.55	66.76	77.04	62.40	311.57	62.40			
	SMAPE	30.90	164.50	130.50	161.40	37.90	36.10	117.10	68.50	31.90	29.03	32.70	27.70	143.40	27.70			

(continued)

Table 3. Continued

Forecast Period	Metric	Temporal-only model							Spatio-temporal model							Proposed E-STGCN
		ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified STGCN	DeepKriging		
MAY	Pinball Loss	43.18	144.81	132.12	143.49	53.27	51.59	120.90	83.76	38.60	27.75	39.20	25.20	156.05	25.20	
	CRPS	96.36	122.26	126.72	122.26	108.53	108.55	122.69	122.69	92.03	82.19	93.58	78.66	145.23	78.66	
	MAE	71.01	176.32	167.71	170.78	79.39	71.75	117.99	88.82	78.77	78.42	94.51	77.46	188.01	77.46	
	MASE	1.12	2.81	2.67	2.72	1.24	1.15	1.86	1.41	1.23	1.25	1.53	1.24	3.01	1.24	
	RMSE	91.48	199.93	205.51	195.07	107.15	92.81	147.03	118.57	105.91	97.83	117.57	96.88	211.09	96.88	
JUN	SMAPE	38.00	167.30	119.70	155.40	42.90	38.60	77.90	49.30	42.50	41.91	47.90	41.40	194.50	41.40	
	Pinball Loss	35.98	141.06	122.77	136.63	53.76	33.77	91.62	61.73	52.29	36.63	38.56	31.28	150.38	31.28	
	CRPS	90.80	122.60	127.58	122.60	108.91	90.80	122.96	119.12	107.33	96.06	111.76	96.87	122.88	96.87	
	MAE	67.66	117.99	123.54	114.96	47.59	43.32	87.20	74.67	69.23	72.83	43.44	53.60	130.25	51.65	
	MASE	2.00	3.36	3.53	3.27	1.42	1.27	2.46	2.11	2.09	2.17	1.25	1.58	3.78	1.51	
JUL	RMSE	81.88	129.68	148.78	126.93	58.79	56.31	100.34	89.19	82.67	82.26	55.60	65.15	154.60	65.44	
	SMAPE	45.10	150.10	124.00	142.20	36.00	33.40	90.00	70.90	47.00	48.73	33.00	39.00	137.00	37.90	
	Pinball Loss	20.16	94.39	90.49	91.96	20.32	23.77	69.76	59.68	23.20	19.06	23.70	23.90	98.71	16.17	
	CRPS	80.27	84.39	87.31	84.39	63.00	62.06	83.63	83.14	81.28	77.44	61.75	67.01	91.18	69.55	
	MAE	101.08	60.00	72.71	57.24	72.83	30.07	38.87	31.80	96.52	125.26	60.67	48.40	76.07	48.40	
AUG	MASE	5.72	3.21	3.77	3.05	4.37	1.65	2.01	1.66	5.83	7.09	3.52	2.74	4.11	2.74	
	RMSE	108.53	65.60	88.01	63.07	77.47	36.78	46.32	38.07	103.92	128.13	75.32	55.25	80.74	55.25	
	SMAPE	81.10	120.10	117.10	110.20	69.50	36.90	62.10	46.70	79.90	93.13	56.40	51.70	186.90	51.70	
	Pinball Loss	20.77	48.00	51.51	45.79	14.95	10.36	30.43	22.67	20.41	26.11	16.01	10.67	60.86	10.67	
	CRPS	74.62	47.71	52.75	47.70	58.41	38.46	46.71	43.11	72.39	83.19	61.67	48.14	47.75	48.14	
RMSE	MAE	60.84	108.73	112.74	105.64	48.29	47.00	71.27	84.91	60.04	42.14	87.94	47.44	94.30	39.70	
	MASE	2.86	4.88	5.02	4.73	2.23	2.03	3.16	3.80	2.88	1.90	3.98	2.19	4.25	1.80	
	RMSE	69.72	116.57	124.58	113.71	56.02	59.81	87.31	97.19	69.23	51.87	97.95	55.36	114.38	47.97	

(continued)

Table 3. Continued

Forecast Period	Metric	Temporal-only model							Spatio-temporal model							Proposed E-STGCN
		ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified STGCN	DeepKrigging		
SEP	SMAPE	42.60	146.70	141.80	138.40	37.20	39.20	71.80	94.80	43.10	32.29	104.10	36.00	93.00	31.00	
	Pinball Loss	14.02	86.98	88.22	84.51	16.72	35.18	56.80	67.91	15.76	18.73	69.87	14.09	59.35	14.25	
	CRPS	67.48	77.86	79.92	77.86	60.84	68.49	79.02	78.59	67.32	58.65	77.33	60.26	85.83	54.41	
	MAE	82.97	92.66	104.00	91.22	52.34	68.16	48.30	47.63	66.26	89.70	79.02	44.58	119.67	35.90	
	MASE	4.60	5.68	6.07	5.59	3.08	3.89	2.95	2.90	3.91	5.15	4.93	2.54	7.04	2.05	
OCT	RMSE	94.06	102.86	122.03	101.53	63.62	81.38	60.46	58.86	78.06	101.05	90.40	55.68	166.10	44.51	
	SMAPE	60.70	130.70	122.30	126.70	46.50	53.40	48.50	46.60	53.70	63.63	96.90	41.50	98.90	35.70	
	Pinball Loss	16.86	74.08	74.96	72.92	13.99	14.91	26.92	24.25	14.97	19.36	59.08	14.96	63.87	13.98	
	CRPS	78.86	70.71	76.44	70.71	62.09	69.93	65.91	64.31	69.96	84.24	71.51	59.82	84.12	50.39	
	MAE	56.62	214.64	202.97	206.32	81.46	94.69	120.44	118.95	67.79	60.10	178.89	62.18	218.93	56.42	
NOV	MASE	1.54	6.00	5.57	5.75	2.17	2.61	3.27	3.29	1.80	1.64	5.03	1.70	6.12	1.55	
	RMSE	76.41	225.67	223.43	217.78	101.38	114.30	148.51	147.03	88.08	79.08	192.39	80.64	231.48	74.86	
	SMAPE	25.00	176.90	140.50	162.80	37.90	47.60	66.20	65.60	30.20	26.74	134.70	27.80	181.60	25.00	
	Pinball Loss	38.14	171.71	160.34	165.06	62.31	74.59	95.89	94.70	47.51	40.09	142.75	39.99	175.13	33.52	
	CRPS	93.92	139.86	142.87	139.86	119.26	133.38	141.47	141.89	105.70	95.93	138.79	95.26	140.35	90.50	
DEC	MAE	160.63	370.52	342.23	365.48	230.28	114.14	164.50	163.94	101.36	139.78	248.39	105.00	378.61	105.00	
	MASE	2.00	4.64	4.27	4.57	2.86	1.43	2.04	2.04	1.28	1.74	3.08	1.32	4.77	1.32	
	RMSE	181.25	387.20	370.87	382.37	255.63	137.63	183.15	182.67	123.56	161.36	273.40	125.82	406.14	125.82	
	SMAPE	49.20	182.20	140.80	176.90	78.10	31.90	51.80	51.80	27.60	40.51	88.70	29.40	169.70	29.40	
	Pinball Loss	124.79	296.41	268.32	292.38	183.96	73.64	125.90	125.97	49.25	105.57	198.49	71.86	302.31	71.86	
DEC	CRPS	222.28	237.70	241.89	237.70	237.70	178.26	216.77	219.46	159.67	206.01	237.94	165.00	243.63	165.00	
	MAE	111.36	320.05	303.66	311.54	174.14	82.85	131.11	148.68	112.51	134.13	100.10	117.10	326.28	134.80	
	MASE	2.02	5.87	5.58	5.71	3.12	1.54	2.36	2.71	2.02	2.43	1.95	2.14	6.03	2.50	

(continued)

Table 3. Continued

Forecast Period	Metric	Temporal-only model						Spatio-temporal model						Proposed	
		ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified STGCN	DeepKrigging	E-STGCN
	RMSE	139.68	329.61	342.88	321.36	191.44	100.84	163.30	183.15	139.38	155.58	118.07	141.74	347.20	157.71
	SMAPE	36.60	184.50	129.80	174.50	66.20	24.70	45.00	55.20	36.10	46.92	28.90	39.50	168.30	48.20
	Pinball Loss	83.53	256.04	228.05	249.23	139.30	39.40	95.89	117.67	80.35	108.21	32.82	93.29	259.98	107.55
	CRPS	174.92	202.04	219.05	202.04	200.90	128.38	189.84	201.10	171.97	190.48	140.95	177.56	209.45	189.76

Note. In the table Pinball Loss at 80% quantile is reported.

Table 4. Forecasting performance of the proposed extreme value theory guided modified spatio-temporal graph convolutional networks (E-STGCN) model in comparison to the temporal-only and spatio-temporal forecasting techniques for 30-day ahead forecast horizon of NO₂ pollutant (best results are given in bold)

Forecast Period	Metric	Temporal-only model										Spatio-temporal model						Proposed	
		ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified STGCN	DeepKrigging	E-STGCN				
JAN	MAE	13.76	23.48	38.72	26.73	16.49	15.71	17.87	15.76	16.37	17.68	28.46	12.28	40.02	12.28				
	MASE	3.33	8.21	15.68	9.85	11.36	4.58	4.99	2.48	4.48	10.59	8.15	4.74	13.43	4.74				
	RMSE	17.01	26.50	41.37	29.63	19.45	19.34	21.80	19.37	19.92	20.64	36.38	15.01	48.68	15.01				
	SMAPE	33.70	61.60	160.40	76.20	39.50	41.50	44.20	38.70	39.70	40.26	62.50	28.30	109.50	28.30				
	Pinball Loss	7.80	18.36	30.95	21.20	8.07	10.73	13.71	11.03	9.43	9.49	13.02	6.50	23.98	6.50				
FEB	CRPS	19.37	24.59	26.04	25.26	20.56	21.88	23.64	22.00	21.54	21.13	29.21	18.30	32.98	18.30				
	MAE	14.21	25.80	39.27	27.09	16.76	13.64	20.57	18.42	16.95	17.47	33.94	11.92	33.84	11.92				
	MASE	2.32	4.85	8.16	5.18	4.47	2.65	3.61	2.71	3.58	6.28	7.55	2.50	6.94	2.50				
	RMSE	17.06	28.35	41.70	29.61	19.36	16.64	23.51	21.46	20.09	20.08	47.18	14.56	36.77	14.56				
	SMAPE	34.60	70.80	158.20	76.70	40.20	37.00	56.20	47.40	41.40	39.56	67.40	28.50	123.20	28.50				
MAR	Pinball Loss	8.33	20.27	31.39	21.39	8.20	6.95	15.70	13.93	10.21	9.69	14.18	7.05	26.70	7.05				
	CRPS	19.39	25.05	26.07	25.30	20.27	19.15	24.85	23.57	22.18	20.67	32.54	17.85	27.05	17.85				
	MAE	12.39	17.26	31.02	18.48	16.78	12.68	8.64	9.94	14.91	14.81	21.44	8.96	30.13	8.96				
	MASE	4.65	6.43	12.07	6.82	9.83	4.20	3.16	2.63	5.53	8.44	8.56	3.53	11.56	3.53				
	RMSE	14.36	19.22	32.85	20.42	18.54	15.22	10.60	11.92	17.07	16.80	27.80	11.23	33.18	11.23				
APR	SMAPE	38.90	53.90	157.50	59.50	48.30	39.00	27.20	31.40	46.50	43.66	58.50	27.40	144.90	27.40				
	Pinball Loss	3.97	13.15	24.79	14.33	5.30	3.84	4.81	5.22	5.26	6.20	9.37	4.46	23.50	4.46				
	CRPS	15.57	18.43	21.04	18.92	17.63	15.99	13.64	14.24	17.23	16.92	22.17	13.92	21.17	13.92				
	MAE	15.70	22.31	36.47	23.05	19.74	16.72	18.32	18.19	17.41	19.48	25.79	13.63	37.75	13.63				
	MASE	3.57	3.15	6.41	3.24	5.02	2.78	2.56	2.94	4.37	4.62	6.08	2.28	6.95	2.28				
APR	RMSE	19.05	25.77	39.24	26.49	22.29	20.62	22.58	22.24	21.18	22.00	33.53	16.22	40.30	16.22				
	SMAPE	40.50	60.90	168.20	64.30	49.60	47.70	49.20	50.00	45.20	48.22	64.20	34.40	196.80	34.40				

(continued)

Table 4. Continued

Forecast Period	Metric	Temporal-only model							Spatio-temporal model							Proposed E-STGCN
		ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified STGCN	DeepKrigging		
MAY	Pinball Loss	8.17	17.26	29.17	17.95	8.13	12.16	14.33	13.75	9.77	9.42	13.23	7.24	30.20	7.24	
	CRPS	19.90	22.70	24.22	22.83	21.53	22.09	22.49	22.50	21.36	21.03	26.38	18.45	23.74	18.45	
	MAE	14.82	19.30	32.75	19.71	20.06	18.59	14.77	13.94	16.85	19.98	27.00	11.33	23.67	11.33	
	MASE	3.37	2.59	5.80	2.59	6.46	6.02	2.18	2.62	4.64	7.01	7.54	2.20	8.26	2.20	
	RMSE	18.05	22.61	35.85	23.03	22.99	22.09	18.47	17.38	20.32	22.70	35.38	14.35	28.05	14.35	
JUN	SMAPE	41.70	56.70	155.60	58.60	52.60	55.40	41.30	40.40	47.50	50.63	66.00	32.40	65.40	32.40	
	Pinball Loss	6.41	14.75	26.16	15.20	7.26	8.62	9.99	8.76	7.63	8.46	11.37	6.29	10.81	6.29	
	CRPS	18.55	20.24	22.93	20.43	21.12	21.10	18.83	18.30	19.78	20.86	26.20	16.18	23.40	16.18	
	MAE	13.79	12.89	23.74	13.38	24.21	11.55	7.43	10.03	13.84	19.91	25.04	8.05	26.23	8.05	
	MASE	5.61	3.72	9.24	3.77	10.91	5.10	1.84	4.18	10.88	9.26	14.84	2.71	10.78	2.71	
JUL	RMSE	15.80	14.59	25.68	15.09	25.40	13.58	9.28	11.79	16.15	21.21	33.49	9.88	27.36	9.88	
	SMAPE	49.80	48.50	144.50	50.70	70.10	45.60	29.30	39.80	51.30	60.29	70.60	31.30	197.50	31.30	
	Pinball Loss	4.03	8.73	18.81	9.48	5.76	4.49	5.39	7.27	4.47	6.27	8.70	3.82	20.98	3.82	
	CRPS	14.75	14.52	17.40	14.66	19.25	13.74	11.54	13.29	14.83	17.15	22.43	11.60	16.25	11.60	
	MAE	15.21	10.11	19.58	10.19	25.66	8.11	6.74	7.92	13.51	12.63	23.31	7.56	14.86	7.56	
AUG	MASE	6.35	3.40	6.25	3.28	10.64	2.77	2.15	2.50	6.09	5.20	9.64	2.68	5.08	2.68	
	RMSE	16.93	11.74	21.19	11.88	26.97	10.41	8.48	9.63	15.33	14.46	31.45	9.95	16.94	9.95	
	SMAPE	61.60	46.90	153.80	47.50	82.10	47.50	35.00	40.10	59.30	55.29	74.90	37.20	80.00	37.20	
	Pinball Loss	3.68	5.83	15.60	6.30	5.80	5.24	4.50	5.29	3.87	5.12	7.51	3.26	9.30	3.26	
	CRPS	13.84	11.39	13.65	11.53	18.77	11.15	9.57	10.69	13.27	12.78	20.03	10.39	13.44	10.39	
AUG	MAE	17.45	9.59	17.26	9.62	25.21	8.56	7.89	8.40	14.07	19.32	19.88	6.95	12.12	6.95	
	MASE	7.14	4.18	6.10	4.22	10.73	3.11	2.44	2.76	7.98	8.73	12.51	2.13	5.74	2.13	
	RMSE	19.36	11.41	19.24	11.45	27.10	10.80	9.92	10.40	16.21	20.87	28.54	8.66	13.93	8.66	

(continued)

Table 4. Continued

Forecast Period	Metric	Temporal-only model							Spatio-temporal model							Proposed E-STGCN
		ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified STGCN	DeepKrigging		
SEP	SMAPE	71.50	52.90	154.50	53.10	86.80	47.50	40.80	46.60	64.80	77.15	75.40	38.40	74.90	38.40	38.40
	Pinball Loss	4.21	5.21	13.78	5.19	5.72	4.84	4.79	5.84	3.98	6.22	6.96	3.25	7.72	3.25	3.25
	CRPS	14.78	10.81	12.04	10.80	18.80	10.31	9.91	10.59	13.13	15.14	17.39	9.04	11.65	9.04	9.04
	MAE	13.27	10.27	20.46	10.32	20.50	9.90	8.50	10.48	12.52	19.13	15.40	6.46	22.24	6.46	6.46
	MASE	4.74	3.50	5.65	3.49	7.77	3.24	2.30	2.62	6.35	7.36	6.97	1.87	6.37	1.87	1.87
OCT	RMSE	14.91	11.93	22.27	11.98	21.82	11.92	10.62	12.61	14.62	20.57	19.91	8.32	23.50	8.32	8.32
	SMAPE	53.50	48.50	148.30	48.80	69.10	49.60	41.00	52.20	53.30	65.75	66.50	32.10	195.10	32.10	32.10
	Pinball Loss	3.35	6.40	16.29	6.56	4.72	5.92	6.04	7.93	4.20	5.64	7.80	2.87	17.78	2.87	2.87
	CRPS	13.11	12.42	15.94	12.47	16.65	12.64	11.63	12.87	13.47	16.16	15.71	9.69	13.97	9.69	9.69
	MAE	13.92	19.79	34.59	20.17	18.54	16.77	16.12	16.93	15.79	18.41	24.02	12.09	32.98	12.09	12.09
NOV	MASE	2.55	3.05	6.12	3.09	3.81	2.77	2.48	2.61	3.46	3.64	4.22	1.91	6.01	1.91	1.91
	RMSE	19.20	24.99	38.91	25.41	23.56	22.57	22.37	23.21	21.30	23.34	30.25	17.30	37.23	17.30	17.30
	SMAPE	35.40	55.90	165.90	57.50	46.20	48.50	44.40	47.60	41.90	45.57	79.90	30.30	136.30	30.30	30.30
	Pinball Loss	7.32	15.22	27.65	15.61	8.05	12.37	12.44	12.73	8.85	9.45	17.04	7.23	26.09	7.23	7.23
	CRPS	18.13	21.68	23.23	21.77	20.34	21.46	21.09	21.47	19.78	20.32	23.39	17.00	22.84	17.00	17.00
DEC	MAE	19.56	31.39	42.55	31.40	25.12	22.51	18.01	20.86	20.54	26.53	34.00	16.62	46.43	16.62	16.62
	MASE	3.11	4.04	5.61	4.03	4.83	3.39	3.02	3.42	4.35	5.75	5.18	2.19	5.95	2.19	2.19
	RMSE	24.01	35.90	47.07	35.92	29.31	27.62	22.48	25.60	24.75	30.55	42.28	21.10	49.98	21.10	21.10
	SMAPE	43.80	77.40	140.20	77.40	51.40	54.70	42.20	46.30	47.60	51.18	84.10	36.20	190.20	36.20	36.20
	Pinball Loss	11.74	24.45	33.62	24.45	14.70	14.09	11.40	13.20	10.93	14.60	23.12	10.31	37.08	10.31	10.31
DEC	CRPS	24.64	29.61	31.59	29.59	27.63	28.09	23.75	26.04	25.12	28.31	31.36	22.93	29.92	22.93	22.93
	MAE	17.57	30.04	41.09	30.01	23.40	20.50	18.74	24.55	18.18	22.21	31.54	21.44	47.41	21.44	21.44
	MASE	2.62	4.24	6.53	4.24	3.30	3.91	2.56	3.42	2.99	3.41	5.07	3.30	8.38	3.30	3.30

(continued)

Table 5. Forecasting performance of the proposed extreme value theory guided modified spatio-temporal graph convolutional networks (E-STGCN) model in comparison to the temporal-only and spatio-temporal forecasting techniques for 60-day ahead forecast horizon of PM_{2.5} and PM₁₀ pollutants (best results are given in bold)

Pollutant	Forecast	Metric	Temporal-only model										Spatio-temporal model					Proposed
			ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified STGCN	DeepKrigging	E-STGCN		
PM _{2.5}	JAN-FEB	MAE	73.48	132.20	116.87	131.90	66.52	59.28	80.56	100.35	58.53	62.00	76.26	56.73	84.47	55.00		
		MASE	1.72	3.11	2.73	3.10	1.54	1.39	1.87	2.36	1.36	1.44	1.82	1.32	1.99	1.28		
		RMSE	86.23	148.60	135.33	148.33	89.53	77.12	96.46	113.61	78.81	84.09	94.98	76.44	100.93	73.00		
	Pinball Loss	SMAPE	46.50	152.00	126.90	151.30	47.40	45.20	81.50	57.00	39.90	43.58	48.80	39.60	77.20	37.90		
		CRPS	83.42	95.40	94.98	95.40	50.69	45.07	62.52	23.37	39.18	46.61	34.45	38.63	58.54	37.44		
		MAE	35.03	59.96	49.14	55.51	29.62	35.06	50.95	50.86	37.44	34.42	143.20	27.68	68.10	22.35		
MAR-APR	Pinball Loss	MASE	1.82	3.02	2.46	2.78	1.58	1.84	2.56	2.55	2.07	1.81	7.71	1.39	3.47	1.13		
		RMSE	40.90	65.06	58.64	60.99	34.89	41.08	57.72	58.23	42.69	39.94	154.31	35.06	77.92	27.69		
		SMAPE	43.70	131.80	93.70	114.10	39.20	42.80	111.50	112.70	46.50	43.28	98.80	41.70	144.30	31.20		
	CRPS	Pinball Loss	8.10	47.97	35.48	44.41	8.60	8.89	40.47	39.91	9.54	9.65	28.74	18.35	48.39	8.08		
		MAE	40.07	45.49	46.80	45.50	36.68	40.11	45.67	45.84	40.96	39.48	98.36	41.11	49.50	33.74		
		MASE	37.54	40.38	42.47	38.31	46.70	22.53	43.36	22.15	49.65	50.80	43.12	30.41	37.31	43.53		
MAY-JUN	Pinball Loss	MASE	2.14	2.23	2.35	2.11	2.71	1.24	2.41	1.23	2.95	2.92	2.33	1.70	2.03	2.44		
		RMSE	42.70	49.77	51.29	48.09	51.41	32.55	50.27	30.71	54.99	55.57	52.49	42.70	44.79	56.56		
		SMAPE	57.70	97.10	117.20	88.10	66.60	39.10	127.60	38.70	68.60	69.57	70.40	58.00	99.80	94.50		
	CRPS	Pinball Loss	11.64	32.29	33.46	30.63	11.17	15.27	34.65	13.32	11.69	19.84	21.46	19.52	29.10	19.19		
		MAE	30.37	21.33	26.49	18.44	60.75	24.10	29.51	12.67	62.11	62.03	20.80	29.41	36.68	28.05		
		MASE	4.07	2.68	3.38	2.30	8.17	3.25	3.77	1.60	8.35	8.33	2.60	3.89	4.71	3.70		
JUL-AUG	Pinball Loss	RMSE	33.44	24.35	30.50	21.69	61.90	28.32	32.59	16.06	64.19	63.43	24.31	35.63	38.78	34.11		
		SMAPE	60.90	74.00	109.00	59.70	93.40	51.00	132.80	38.10	93.60	93.83	61.90	56.70	197.50	55.40		

(continued)

Table 5. Continued

Pollutant	Forecast	Metric	Temporal-only model						Spatio-temporal model						Proposed	
			ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified DeepKrigging		E-STGCN
PM ₁₀	SEP-OCT	Pinball Loss	6.55	16.98	20.22	14.57	12.21	5.82	23.60	9.45	12.50	13.47	13.26	7.14	29.24	6.91
		CRPS	26.97	22.71	23.22	22.46	39.78	24.77	22.95	18.45	41.13	40.81	23.37	28.61	22.72	22.80
	MAE	36.12	54.88	63.93	52.05	44.53	40.25	58.62	34.86	44.33	34.29	53.65	33.99	57.96	33.60	
	MASE	2.63	3.95	4.64	3.74	3.32	2.91	4.25	2.54	3.34	2.52	3.90	2.50	4.15	2.46	
	RMSE	51.27	69.38	78.49	66.93	50.97	55.78	76.23	49.13	51.34	41.26	68.24	41.05	71.90	40.34	
	SMAPE	55.00	107.90	148.20	97.40	59.80	64.00	120.70	52.40	59.40	49.63	108.80	49.60	123.00	48.90	
NOV-DEC	Pinning Loss	CRPS	26.24	43.73	50.83	41.36	14.96	30.90	46.10	24.12	14.66	15.37	42.29	13.65	46.12	13.52
		MAE	46.88	47.82	48.53	47.83	45.98	47.67	49.68	45.92	46.10	39.24	47.71	39.02	47.83	38.67
	MASE	90.77	204.62	195.80	200.48	127.97	132.34	159.92	157.59	120.80	107.25	198.17	95.34	224.42	91.09	
MAR-APR	Pinning Loss	MAE	1.91	4.30	4.11	4.21	2.67	2.78	3.35	3.31	2.50	2.24	4.18	2.00	4.73	1.91
		RMSE	111.64	217.05	210.92	213.17	146.81	149.75	175.57	173.31	139.69	127.11	211.16	115.40	244.57	112.98
	SMAPE	46.50	169.40	155.70	162.30	74.50	80.80	117.80	113.50	68.20	57.87	163.30	49.60	189.00	46.30	
	CRPS	71.38	163.70	156.29	160.39	102.18	105.13	127.54	125.72	96.23	85.93	158.54	74.93	177.00	69.81	
	MAE	131.98	136.38	137.02	136.39	135.65	136.02	136.20	136.15	134.78	134.44	136.42	133.45	139.13	131.94	
	MASE	94.47	252.19	245.06	245.01	135.51	95.19	120.61	116.70	101.61	82.81	102.18	80.35	125.87	72.53	
JAN-FEB	Pinning Loss	MAE	1.53	3.99	3.87	3.88	2.11	1.53	1.88	1.83	1.58	1.30	1.63	1.27	1.97	1.15
		RMSE	113.55	269.06	280.32	262.36	163.79	116.76	147.89	143.09	130.47	110.99	127.15	106.36	152.58	94.10
	SMAPE	35.10	183.00	148.20	171.80	61.20	35.40	61.00	56.50	40.70	32.34	40.00	31.40	58.10	28.20	
	CRPS	30.47	201.75	189.48	196.01	107.88	34.42	91.92	90.83	74.52	60.41	56.99	55.78	89.08	45.19	
	MAE	124.50	163.35	164.21	163.35	162.62	124.74	152.08	153.03	149.54	131.16	125.64	149.42	131.16	109.46	
	MASE	61.48	168.41	174.61	164.51	62.63	61.36	103.94	95.72	57.14	54.37	165.95	51.94	181.87	49.59	
NOV-DEC	Pinning Loss	MAE	1.43	3.87	4.02	3.78	1.39	1.42	2.35	2.18	1.32	1.27	4.06	1.20	4.19	1.14
		RMSE	73.69	178.80	203.57	175.13	78.19	75.55	125.97	117.40	70.86	65.85	212.95	63.17	191.54	60.07

(continued)

Table 5. Continued

Pollutant	Forecast	Metric	Temporal-only model						Spatio-temporal model						Proposed	
			ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified DeepKrigging		E-STGCN
MAY-JUN	Pinball Loss	SMAPE	33.40	167.10	135.50	159.10	35.80	33.80	76.30	67.70	32.20	30.45	60.90	29.20	198.90	28.10
		RMSE	19.10	134.73	127.74	131.61	45.14	28.45	79.35	72.03	32.15	24.67	36.13	21.88	145.49	23.67
		CRPS	83.08	113.16	121.38	113.16	95.51	87.98	115.24	114.65	85.66	77.98	155.32	75.52	113.16	73.97
MAY-JUN	MAE	MAE	71.99	145.59	145.40	143.09	83.16	67.01	129.13	111.88	67.53	75.81	87.48	76.14	77.28	66.47
		MASE	1.48	2.95	2.92	2.90	1.65	1.35	2.61	2.27	1.40	1.57	1.83	1.55	1.53	1.36
		RMSE	89.34	167.01	173.28	164.85	113.32	95.91	150.33	134.00	87.32	90.26	107.85	100.07	107.10	89.67
JUL-AUG	Pinball Loss	SMAPE	43.60	155.20	133.70	149.30	56.00	42.90	128.80	100.80	41.70	45.48	48.30	47.80	50.50	41.10
		CRPS	30.71	116.47	111.33	114.47	64.84	50.75	103.25	89.01	33.56	32.42	28.02	44.45	58.81	37.86
		MAE	86.48	104.87	109.70	104.87	104.04	98.39	103.80	103.36	87.12	86.72	97.49	101.79	101.82	86.46
JUL-AUG	MAE	MAE	58.10	87.62	95.24	82.17	59.70	42.82	76.47	77.41	70.26	104.24	58.97	49.79	72.79	49.79
		MASE	2.90	4.18	4.53	3.90	3.05	2.04	3.62	3.69	3.65	5.22	2.79	2.46	3.43	2.46
		RMSE	65.40	97.74	109.72	92.87	67.62	54.78	91.86	93.15	78.60	111.69	72.70	59.03	84.92	59.03
SEP-OCT	Pinball Loss	SMAPE	50.40	142.70	142.60	125.20	52.50	40.70	110.10	112.70	58.20	73.42	64.00	43.70	105.10	43.70
		CRPS	13.75	70.09	72.58	65.73	15.87	21.15	60.88	61.64	17.58	22.79	39.40	12.37	57.79	12.37
		MAE	59.61	64.63	68.49	64.63	61.12	57.72	66.10	66.26	66.90	82.76	66.03	56.00	64.96	56.00
SEP-OCT	MAE	MAE	64.76	153.86	160.36	148.98	67.84	77.56	119.55	118.63	70.16	74.00	132.91	69.95	164.33	63.70
		MASE	2.38	5.74	5.96	5.55	2.50	2.88	4.44	4.42	2.62	2.74	4.95	2.59	6.14	2.36
		RMSE	81.32	174.76	184.58	170.42	87.21	97.58	151.29	150.95	88.90	89.91	155.66	84.98	185.21	81.31
NOV-DEC	Pinball Loss	SMAPE	40.50	156.20	149.10	145.10	42.70	47.90	93.70	92.00	43.80	44.80	119.00	43.20	175.80	40.00
		CRPS	27.79	123.08	124.20	119.15	38.93	42.61	89.79	88.02	35.99	29.03	105.59	28.50	130.88	26.66
		MAE	85.45	109.81	113.53	109.81	92.85	101.60	115.41	116.21	92.69	91.66	109.09	88.29	110.79	87.29
NOV-DEC	MAE	MAE	114.94	342.22	328.89	338.54	202.89	156.01	204.62	188.63	186.24	164.65	175.93	162.09	350.87	162.09
		MASE	1.71	5.07	4.85	5.01	2.97	2.30	3.00	2.79	2.71	2.43	2.58	2.40	5.20	2.40
		RMSE	1.71	5.07	4.85	5.01	2.97	2.30	3.00	2.79	2.71	2.43	2.58	2.40	5.20	2.40

(continued)

Table 5. Continued

Pollutant	Forecast	Metric	Temporal-only model						Spatio-temporal model						Proposed	
			ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified STGCN		DeepKrigging
		RMSE	139.63	356.98	351.94	353.46	226.80	177.83	226.54	210.06	208.70	191.68	208.93	187.66	366.66	187.66
		SMAPE	35.00	179.80	157.70	175.80	72.60	52.30	81.90	72.00	64.90	54.60	61.20	53.80	187.60	53.80
		Pinball Loss	87.31	273.77	260.48	270.84	162.18	121.19	160.86	148.22	134.06	146.36	127.14	128.73	280.63	128.73
		CRPS	180.03	220.48	222.80	220.48	220.48	209.88	216.08	219.38	215.31	219.38	214.57	219.31	220.86	219.31

Note. In the table Pinball Loss at 80% quantile is reported.

Table 6. Forecasting performance of the proposed extreme value theory guided modified spatio-temporal graph convolutional networks (E-STGCN) model in comparison to the temporal-only and spatio-temporal forecasting techniques for 60-day ahead forecast horizon of NO₂ pollutant. (best results are given in bold)

Pollutant	Forecast	Metric	Temporal-only model										Spatio-temporal model										Proposed				
			ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified STGCN	DeepKrigging	E-STGCN											
NO ₂	JAN-FEB	MAE	15.04	29.56	39.64	26.76	17.21	18.62	18.45	18.37	18.52	17.63	35.17	13.50	42.27	13.50	42.27	13.50	42.27	13.50	42.27	13.50	42.27	13.50	42.27	13.50	
		MASE	2.85	6.53	9.23	5.79	5.78	4.21	3.85	2.60	3.95	8.79	7.72	4.69	8.48	4.69	8.48	4.69	8.48	4.69	8.48	4.69	8.48	4.69	8.48	4.69	
	RMSE	18.17	32.57	42.33	29.93	20.18	22.76	22.59	22.23	21.72	20.62	47.45	16.28	52.00	16.28	52.00	16.28	52.00	16.28	52.00	16.28	52.00	16.28	52.00	16.28	52.00	
	SMAPE	37.30	91.50	172.00	76.70	41.50	51.50	46.50	47.90	45.90	40.49	67.30	31.90	108.70	31.90	108.70	31.90	108.70	31.90	108.70	31.90	108.70	31.90	108.70	31.90	108.70	
	Pinball Loss	7.58	23.55	31.70	21.17	7.52	13.67	13.11	12.73	8.65	9.62	12.74	5.86	22.36	5.86	22.36	5.86	22.36	5.86	22.36	5.86	22.36	5.86	22.36	5.86	22.36	
MAR-APR	MAE	CRPS	19.90	25.64	25.99	25.36	20.66	24.23	23.65	23.62	22.33	21.06	33.93	18.55	34.44	18.55	34.44	18.55	34.44	18.55	34.44	18.55	34.44	18.55	34.44	18.55	34.44
		MAE	15.60	21.86	33.86	20.66	17.82	14.20	14.39	15.75	17.59	17.79	26.32	12.04	27.12	12.04	27.12	12.04	27.12	12.04	27.12	12.04	27.12	12.04	27.12	12.04	
	MASE	4.38	3.44	6.65	3.25	5.23	3.61	2.61	3.03	5.67	4.97	6.98	2.74	5.56	2.74	5.56	2.74	5.56	2.74	5.56	2.74	5.56	2.74	5.56	2.74	5.56	
	RMSE	18.92	25.14	36.57	24.02	20.56	18.02	18.53	20.26	21.19	20.54	34.52	15.32	31.41	15.32	31.41	15.32	31.41	15.32	31.41	15.32	31.41	15.32	31.41	15.32	31.41	
	SMAPE	43.60	68.30	170.60	62.10	48.20	40.80	40.50	47.80	50.30	46.41	63.10	33.80	103.60	33.80	103.60	33.80	103.60	33.80	103.60	33.80	103.60	33.80	103.60	33.80	103.60	
MAY-JUN	Pinball Loss	CRPS	6.15	17.16	27.08	16.06	6.77	5.68	9.55	11.27	6.89	7.84	10.49	6.02	20.22	6.02	20.22	6.02	20.22	6.02	20.22	6.02	20.22	6.02	20.22	6.02	20.22
		MAE	16.25	18.31	28.86	16.63	19.45	21.93	15.58	14.83	17.27	19.61	22.31	12.43	29.55	12.43	29.55	12.43	29.55	12.43	29.55	12.43	29.55	12.43	29.55	12.43	
	MASE	5.05	3.09	6.64	2.92	7.01	7.33	2.72	3.33	6.70	7.34	6.69	3.66	6.44	3.66	6.44	3.66	6.44	3.66	6.44	3.66	6.44	3.66	6.44	3.66	6.44	
	RMSE	19.17	21.37	31.48	19.76	21.69	25.46	18.72	17.95	20.48	21.95	29.14	15.34	32.01	15.34	32.01	15.34	32.01	15.34	32.01	15.34	32.01	15.34	32.01	15.34	32.01	
	SMAPE	50.80	64.00	163.30	55.10	56.20	62.10	54.40	53.40	55.00	54.94	65.30	40.80	143.80	40.80	143.80	40.80	143.80	40.80	143.80	40.80	143.80	40.80	143.80	40.80	143.80	
JUL-AUG	Pinball Loss	CRPS	5.86	14.17	23.06	12.43	6.24	6.96	11.81	11.12	7.09	7.38	10.18	5.01	23.17	5.01	23.17	5.01	23.17	5.01	23.17	5.01	23.17	5.01	23.17	5.01	23.17
		MAE	17.86	18.36	20.67	17.86	18.84	21.39	17.71	17.27	18.60	18.93	22.49	15.58	19.05	15.58	19.05	15.58	19.05	15.58	19.05	15.58	19.05	15.58	19.05	15.58	
	MASE	18.64	9.98	18.39	9.90	25.17	8.69	9.28	10.07	16.33	18.84	24.76	12.05	18.77	12.05	18.77	12.05	18.77	12.05	18.77	12.05	18.77	12.05	18.77	12.05	18.77	
	RMSE	7.28	3.48	5.83	3.40	9.63	2.77	2.80	2.96	7.62	7.07	10.54	4.11	5.96	4.11	5.96	4.11	5.96	4.11	5.96	4.11	5.96	4.11	5.96	4.11	5.96	
	SMAPE	20.75	12.36	20.89	12.27	26.64	11.41	11.63	12.44	19.01	21.10	35.55	14.79	21.13	14.79	21.13	14.79	21.13	14.79	21.13	14.79	21.13	14.79	21.13	14.79	21.13	
MAY-JUN	MAE	72.20	50.70	154.10	50.30	84.30	49.00	52.00	57.70	67.60	71.21	78.50	53.30	162.50	53.30	162.50	53.30	162.50	53.30	162.50	53.30	162.50	53.30	162.50	53.30	162.50	
	RMSE	18.92	25.14	36.57	24.02	20.56	18.02	18.53	20.26	21.19	20.54	34.52	15.32	31.41	15.32	31.41	15.32	31.41	15.32	31.41	15.32	31.41	15.32	31.41	15.32	31.41	

(continued)

Table 6. Continued

Pollutant	Forecast	Metric	Temporal-only model						Spatio-temporal model						Proposed	
			ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified STGCN		DeepKrigging
SEP-OCT	Pinball Loss		4.40	5.87	14.65	5.83	5.64	5.67	6.84	7.55	4.43	5.81	7.54	3.57	15.02	3.57
		CRPS	15.84	11.55	13.35	11.46	18.42	11.20	11.39	11.83	14.97	16.05	20.92	12.84	12.92	12.84
		MAE	15.03	15.67	27.75	15.31	19.61	14.92	17.39	19.72	15.21	19.74	18.80	12.42	19.78	12.42
		MASE	3.82	3.19	5.89	3.11	5.13	3.18	3.42	3.91	4.43	5.12	4.66	2.72	4.81	2.72
		RMSE	19.63	20.84	32.21	20.47	23.89	20.35	23.63	25.64	19.91	24.07	25.21	16.95	24.92	16.95
NOV-DEC	Pinball Loss	SMAPE	48.70	55.20	162.60	53.50	57.90	56.50	68.10	84.80	51.10	57.14	68.10	39.90	86.90	39.90
		CRPS	16.60	17.78	19.26	17.63	19.03	18.32	18.90	19.16	17.24	19.12	19.97	15.36	19.15	15.36
		MAE	20.08	30.52	43.84	30.71	23.70	21.87	21.51	24.54	20.97	24.19	41.25	19.76	36.24	19.76
		MASE	2.68	3.81	5.99	3.83	3.37	2.93	2.82	3.23	3.10	3.49	6.36	2.72	5.28	2.72
		RMSE	24.78	35.49	48.37	35.66	28.20	27.05	26.81	29.82	25.37	28.71	62.27	24.71	40.88	24.71
	Pinball Loss	SMAPE	43.80	75.00	156.00	75.80	48.50	51.40	49.80	57.90	47.00	48.60	77.60	43.30	81.60	43.30
		CRPS	12.25	23.90	34.87	24.07	13.95	14.48	15.24	17.40	11.55	15.23	20.21	13.02	25.77	13.02
		MAE	25.36	29.25	30.95	29.29	27.16	27.68	26.99	28.25	25.52	27.38	36.46	25.61	28.06	25.61
		MASE	2.68	3.81	5.99	3.83	3.37	2.93	2.82	3.23	3.10	3.49	6.36	2.72	5.28	2.72
		RMSE	24.78	35.49	48.37	35.66	28.20	27.05	26.81	29.82	25.37	28.71	62.27	24.71	40.88	24.71

Note. In the table Pinball Loss at 80% quantile is reported.

Table 7. Forecasting performance of the proposed extreme value theory guided modified spatio-temporal graph convolutional networks (E-STGCN) model in comparison to the temporal-only and spatio-temporal forecasting techniques for 90-day ahead forecast horizon of different pollutants (best results are given in bold)

Pollutant	Forecast	Metric	Temporal-only model										Spatio-temporal model										Proposed
			ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified	DeepKrigging	E-STGCN	STGCN	STGCN	STNN	Modified	DeepKrigging	E-STGCN	
PM _{2.5}	JAN-FEB-MAR	MAE	87.20	108.92	101.25	108.36	52.24	50.23	77.69	80.22	50.66	51.90	91.36	47.42	65.84	46.00							
		MASE	2.47	3.10	2.87	3.09	1.48	1.43	2.20	2.28	1.45	1.47	2.66	1.35	1.94	1.31							
		RMSE	97.95	127.48	118.65	127.00	75.09	66.30	90.16	92.99	68.84	73.24	110.19	64.49	76.98	62.59							
		SMAPE	59.70	143.60	137.90	142.00	41.70	47.90	110.70	110.30	40.20	41.39	67.80	37.90	50.20	36.60							
		Pinball Loss	27.03	87.13	80.52	86.69	36.48	38.46	60.86	63.74	29.13	34.91	36.50	28.28	27.89	26.83							
APR-MAY-JUN	APR-MAY-JUN	CRPS	83.53	82.09	81.26	82.09	77.78	68.19	73.77	75.95	69.47	74.53	90.73	65.27	72.05	62.27							
		MAE	44.46	47.61	59.85	43.16	40.92	25.30	52.92	47.06	47.07	45.87	79.39	40.23	58.46	36.04							
		MASE	2.55	2.63	3.36	2.38	2.39	1.42	2.95	2.62	2.81	2.64	4.72	2.28	3.32	2.05							
		RMSE	50.55	55.62	68.24	51.84	45.93	30.93	59.96	54.56	52.37	51.03	85.98	47.66	70.62	42.94							
		SMAPE	59.90	116.20	171.10	96.50	58.00	40.80	157.20	133.60	63.50	61.84	83.50	58.20	131.70	52.10							
JUL-AUG-SEP	JUL-AUG-SEP	Pinball Loss	10.97	38.09	47.62	34.52	10.34	10.78	42.32	37.29	11.54	11.83	16.38	15.72	42.83	11.32							
		CRPS	43.12	39.28	40.16	39.29	39.77	37.64	39.09	39.06	43.32	42.44	61.47	43.01	43.52	39.95							
		MAE	37.00	20.14	28.17	19.75	61.16	41.07	31.89	29.92	61.83	63.09	20.28	49.34	43.87	45.66							
		MASE	4.64	2.36	3.35	2.31	7.69	5.17	3.80	3.56	7.76	7.92	2.41	6.18	5.32	5.73							
		RMSE	40.83	24.13	32.83	23.78	62.78	47.51	35.94	34.54	64.02	64.92	24.56	60.88	53.88	57.29							
JUL-AUG-SEP	JUL-AUG-SEP	SMAPE	68.30	64.30	112.40	62.50	93.20	69.80	146.50	132.70	93.10	94.02	61.30	74.20	135.20	71.50							
		Pinball Loss	7.71	15.84	21.37	15.51	12.32	8.72	25.46	23.72	12.43	13.71	13.79	11.19	29.61	10.29							
		CRPS	30.49	23.50	24.27	23.49	40.74	34.56	23.93	24.10	41.48	42.08	23.69	41.79	29.04	29.56							

(continued)

Table 7. Continued

Pollutant	Forecast	Metric	Temporal-only model					Spatio-temporal model					Proposed			
			ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN		STNN	Modified DeepKrigging	E-STGCN
PM ₁₀	OCT– NOV–DEC	MAE	99.14	158.91	176.00	157.95	98.80	136.28	163.98	163.83	96.85	96.12	162.02	97.34	177.69	95.10
		MASE	2.66	4.29	4.76	4.26	2.65	3.68	4.43	4.43	2.58	2.91	4.37	2.62	4.81	2.56
	RMSE	126.15	180.61	196.24	179.76	125.13	161.44	190.17	189.36	123.31	128.35	183.52	125.07	197.34	122.44	
	SMAPE	64.30	151.40	187.30	149.20	63.90	110.10	157.20	157.30	61.80	65.22	162.00	64.30	199.50	61.00	
	Pinball Loss	78.49	127.12	140.79	126.36	76.57	108.97	131.16	131.07	74.97	75.82	129.53	77.11	142.15	74.14	
	CRPS	114.80	115.44	115.53	115.44	114.58	115.60	117.16	116.93	114.31	114.41	115.45	114.91	115.45	114.10	
	MAE	111.07	219.16	219.77	215.14	109.07	110.28	128.22	118.87	84.76	73.02	135.20	68.88	207.63	68.81	
	MASE	2.04	3.94	3.94	3.86	1.92	2.02	2.28	2.12	1.50	1.31	2.51	1.24	3.73	1.24	
	RMSE	128.20	237.79	251.17	234.10	139.64	129.64	149.21	139.84	113.21	98.71	176.45	92.72	235.06	92.55	
	SMAPE	44.00	174.70	161.10	167.80	52.20	43.80	86.60	73.20	37.20	31.84	48.00	29.90	151.50	29.90	
Pinball Loss	29.99	175.32	171.44	172.12	85.71	32.37	99.53	93.41	59.62	46.66	38.52	41.73	161.88	41.64		
CRPS	125.56	146.24	143.68	146.24	144.56	126.52	136.51	136.65	127.41	110.51	148.12	104.01	150.70	103.99		
APR–MAY– JUN	MAE	69.36	162.31	159.67	153.56	65.89	69.75	150.98	123.69	68.75	72.40	134.97	71.32	143.21	71.32	
MASE	1.49	3.45	3.37	3.29	1.39	1.47	3.20	2.62	1.48	1.56	2.97	1.54	3.08	1.54		
RMSE	85.71	180.11	181.27	174.06	88.02	95.17	169.22	144.79	86.74	86.75	161.53	85.70	163.18	85.70		
SMAPE	40.00	169.50	158.80	154.00	38.70	42.40	152.90	112.20	42.50	41.43	64.80	40.80	94.30	40.80		
Pinball Loss	33.59	129.85	125.98	124.44	43.33	52.72	120.78	98.34	38.58	27.46	48.90	26.74	98.02	26.74		
CRPS	89.98	111.26	112.71	111.26	96.81	104.82	110.78	110.55	92.48	89.62	132.31	88.48	109.68	88.48		
JUL–AUG– SEP	MAE	67.13	89.83	98.76	84.03	57.10	44.69	84.31	86.80	70.51	101.56	52.56	78.42	71.77	59.51	
MASE	3.48	4.48	4.92	4.18	3.03	2.21	4.20	4.35	3.82	5.27	2.67	4.04	3.54	3.04		

(continued)

Table 7. Continued

Pollutant	Forecast	Metric	Temporal-only model					Spatio-temporal model					Proposed			
			ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN		STNN	Modified DeepKrigging	E-STGCN
NO ₂	RMSE		76.53	100.04	111.97	94.82	66.28	56.63	98.63	103.53	80.25	109.53	65.99	100.54	84.83	76.20
			54.90	144.20	156.50	125.70	50.80	45.60	130.20	130.80	57.80	71.70	57.30	54.10	104.50	53.30
	Pinball Loss		14.92	71.86	76.77	67.20	15.10	29.33	67.22	68.49	16.98	22.03	34.79	18.24	55.97	20.38
			65.28	65.90	68.55	65.90	60.97	62.90	67.20	68.62	68.29	82.83	63.37	82.50	66.10	65.91
	MAE		181.14	296.82	300.74	292.10	162.13	246.24	259.03	263.90	155.14	137.84	268.84	145.11	284.10	135.20
			3.24	5.32	5.38	5.23	2.86	4.41	4.63	4.73	2.72	2.81	4.83	2.60	5.09	2.41
	RMSE		212.44	317.30	325.67	312.89	194.85	274.96	293.93	297.78	188.18	171.56	291.76	179.30	309.89	169.99
			74.10	177.60	175.10	171.40	61.40	122.30	135.20	139.80	57.40	50.71	151.00	54.50	164.10	49.30
	Pinball Loss		144.49	237.46	240.11	233.68	128.70	196.98	207.09	211.09	122.13	109.33	215.04	115.22	227.10	106.91
			195.50	195.37	197.42	195.37	195.27	196.71	198.95	198.78	193.24	192.65	195.40	193.15	196.75	191.84
MAE		15.26	25.05	37.10	24.14	17.52	18.14	19.04	18.62	18.74	17.32	30.57	13.74	22.91	13.74	
		3.59	6.34	9.93	6.11	6.33	4.74	4.85	3.07	4.64	9.41	7.52	5.73	7.03	5.73	
RMSE		18.22	28.35	39.76	27.45	20.35	21.85	22.75	22.10	21.95	20.21	40.92	16.56	26.15	16.56	
		40.40	76.40	174.30	71.60	44.80	55.00	54.30	55.00	49.50	42.54	69.60	35.00	72.90	35.00	
Pinball Loss		6.85	19.83	29.67	19.02	6.82	13.46	14.05	13.55	8.20	8.72	13.15	5.61	16.27	5.61	
		19.26	23.72	24.25	23.54	20.07	22.57	22.60	22.35	21.84	20.12	29.88	18.17	22.94	18.17	
MAE		17.56	20.04	31.76	18.86	20.23	14.51	19.61	19.57	19.08	19.89	23.83	15.58	22.12	15.58	
		4.91	3.17	6.53	3.05	5.85	2.98	3.24	3.73	6.69	5.55	6.49	3.72	4.02	3.72	
RMSE		20.81	24.36	35.15	23.23	23.11	18.73	23.74	23.30	22.80	22.92	31.31	18.60	26.23	18.60	
		50.20	64.60	173.50	58.50	55.30	47.80	69.50	69.10	56.20	53.12	65.80	44.70	81.10	44.70	

(continued)

Table 7. Continued

Pollutant	Forecast	Metric	Temporal-only model					Spatio-temporal model					Proposed			
			ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN		STNN	Modified DeepKrigging	E-STGCN
		Pinball Loss	6.79	15.54	25.40	14.38	6.90	10.45	15.49	14.88	8.20	8.07	11.70	5.84	16.94	5.84
		CRPS	19.72	20.41	21.48	20.18	20.92	19.18	20.09	20.37	20.57	20.66	24.05	18.08	20.63	18.08
	JUL–AUG–SEP	MAE	19.29	10.10	19.57	10.00	25.88	10.09	10.84	11.91	18.41	23.07	19.59	15.59	23.39	15.59
		MASE	7.18	3.38	6.06	3.38	9.38	3.14	3.16	3.45	8.80	8.43	8.51	5.42	8.10	5.42
		RMSE	21.28	12.74	22.09	14.01	27.37	13.01	13.71	14.90	20.72	24.74	26.72	18.40	32.27	18.40
		SMAPE	72.40	50.50	167.30	50.00	83.80	58.90	61.20	71.60	72.30	77.38	72.20	60.50	124.50	60.50
		Pinball Loss	4.46	6.25	15.64	8.16	5.72	7.10	8.05	9.02	4.69	6.27	6.92	4.00	15.25	4.00
		CRPS	16.19	12.02	13.59	12.52	18.97	12.28	12.45	12.88	15.83	17.70	18.31	14.94	17.01	14.94
	OCT–NOV–DEC	MAE	19.43	28.26	42.40	27.22	22.33	25.21	28.42	32.14	20.41	22.73	35.98	21.08	43.38	21.08
		MASE	2.79	3.86	6.44	3.69	3.54	3.56	4.02	4.63	3.09	3.53	5.42	2.97	6.63	2.97
		RMSE	24.92	34.10	47.19	33.11	27.58	31.41	35.32	38.73	25.60	28.09	41.82	26.91	48.01	26.91
		SMAPE	43.50	74.60	179.40	70.00	48.20	68.00	80.00	101.20	47.10	48.30	120.90	48.00	196.40	48.00
		Pinball Loss	11.42	22.21	33.92	21.28	11.40	19.48	22.35	25.48	12.26	13.21	27.13	14.21	34.70	14.21
		CRPS	24.20	27.32	27.92	27.21	25.51	27.22	27.94	28.14	24.57	25.92	28.50	25.58	27.92	25.58

Note. In the table Pinball Loss at 80% quantile is reported.

outperformed or performed similarly to the modified STGCN model. This improvement arises from the training mechanism in E-STGCN, which incorporates the POT-based loss function. By integrating information about threshold exceedances, this hybrid loss enables the framework to better capture spikes in pollutant concentrations over NAAQS, particularly during the onset and end of winter months when extreme pollution episodes are most frequent. Since such extreme events often prompt regulatory actions and public-health interventions, underestimating peak concentrations can delay emergency responses and hinder timely policy implementation during high-risk periods. Therefore, a model's ability to accurately forecast pollution exceedances beyond regulatory thresholds is critical for assessing its operational relevance. In this regard, the POT-based loss in E-STGCN plays a key role by explicitly enhancing its ability to capture rare but impactful extremes, which conventional temporal and spatio-temporal architectures often struggle to model. Consequently, the empirical evaluation highlights that the E-STGCN framework offers reliable and operationally meaningful predictions of air pollutant concentrations across different horizons.

Additionally, we evaluated the computational feasibility of the E-STGCN model and baseline forecasters by recording the average training and inference time for each framework. Table 8 reports the total time (in seconds) required to train the models and generate 90-day ahead (January–March) forecasts for PM_{2.5} concentrations across all monitoring stations of Delhi using a single core of an Intel i7 12th generation processor. Among the statistical approaches, ARIMA and STARMA require the least training time due to their low parameter complexity. In contrast, temporal-only deep learning models, particularly the Transformer and NBeats architectures, incur higher training durations due to their attention mechanisms and deep neural architectures, respectively. For spatio-temporal models, STGCN and modified STGCN require moderate training time, while the computational intensity of the GpGp framework increases its training duration. The proposed E-STGCN requires comparable training time to that of standard STGCN, despite the inclusion of an LSTM network in the temporal module and EVT-based loss function. While training times vary across models, inference remains fast across different frameworks except for GpGp and DeepKriging. This runtime analysis underscores that while E-STGCN introduces some additional computational demands, it remains competitive in terms of both training and inference time, and thus is well-suited for near real-time forecasting applications.

4.6 Statistical tests for model robustness

To validate the robustness of our experimental results, we employ multiple comparison with the best (MCB) test (Koning et al., 2005) and the Diebold–Mariano (DM) test (Diebold & Mariano, 2002). The MCB test aims to identify the ‘best’ forecasting model among all \mathcal{F} architectures based on their performance across \mathcal{D} datasets. For a specific evaluation metric, this nonparametric procedure ranks all models based on their performance across different forecasting tasks and computes the mean rank. The model with the lowest mean rank is considered the ‘best’ forecasting architecture. Next, the critical distance (CD) for each of the \mathcal{F} models is computed as $\delta_\theta \sqrt{\mathcal{F}(\mathcal{F} + 1)/6\mathcal{D}}$, where δ_θ represents the critical value of the Tukey distribution at significance level θ . The CD of the ‘best’ performing model serves as the reference value against which all other models are compared. We apply the MCB test and visualize the results based on the deterministic RMSE and distribution-based CRPS metric for PM_{2.5}, PM₁₀, and NO₂ in Figure 8. From the MCB plots (based on RMSE), we observe that the proposed E-STGCN architecture achieves the ‘best’ performance, with a minimum rank of 3.27 for PM_{2.5} and 2.34 for PM₁₀ datasets. The MCB test results based on the probabilistic CRPS metric showcase that the E-STGCN framework outperforms other models in capturing the overall patterns in the dataset, hence achieving the least rank of 2.77 for PM_{2.5} and 2.70 for PM₁₀ datasets. For NO₂ forecasting, the performance of E-STGCN and modified STGCN is similar, and they jointly obtain the lowest rank of 2.41 (in terms of RMSE) and 2.59 (in terms of CRPS). Among the competing models, the modified STGCN, NBeats, and ARIMA frameworks consistently showcase better performance, having competitive ranks with E-STGCN. Spatio-temporal models such as GSTAR, STGCN, STARMA, and GpGp outperform the majority of the time-dependent frameworks by effectively capturing space–time correlations. Moreover, the CD values for most of the baseline models lie above the reference value (shaded region), indicating that their performance is significantly worse than the ‘best-fitted’ E-STGCN model. Additional MCB test results based on the other deterministic and probabilistic evaluation metrics, provided in [online supplementary](#)

Table 8. Training and inference time (in seconds) for long-term forecasting of PM_{2.5} concentration levels across Delhi's monitoring stations

Time	Temporal-only model						Spatio-temporal model						Proposed E-STGCN		
	ARIMA	LSTM	TCN	DeepAR	Transformers	NBeats	STARMA	GSTAR	GpGp	STGCN	STNN	Modified STGCN	DeepKrigging		
Training	0.33	29.85	28.82	29.37	181.75	544.54	0.16	2.82	285.95	170.01	0.18	155.02	116.95	176.83	
Inference	0.01	0.13	0.10	0.12	0.21	0.32	0.05	0.34	13.59	0.01	0.02	0.57	11.93	0.65	

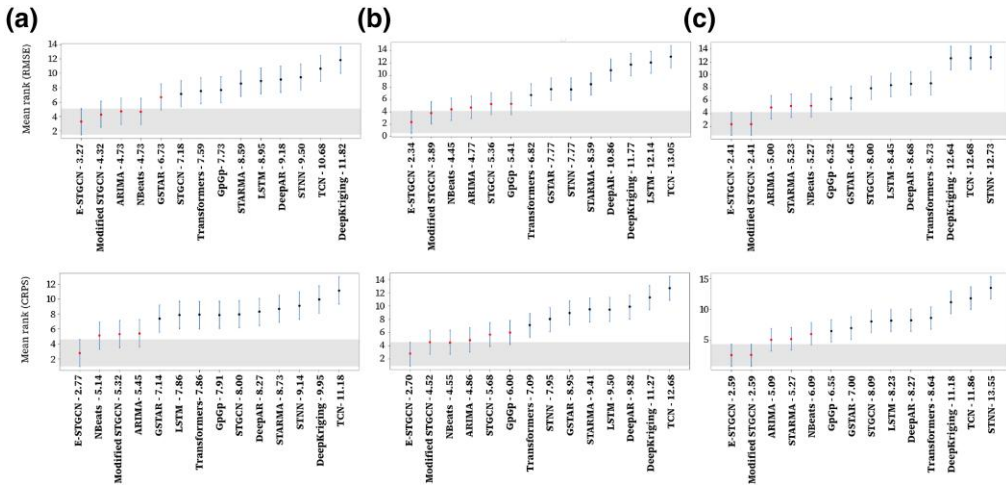


Figure 8. The multiple comparison with the best (MCB) test results for (a) $PM_{2.5}$, (b) PM_{10} , and (c) NO_2 pollutant concentration levels based on root mean squared error (RMSE; upper panel) and continuous ranked probability score (CRPS; lower panel) metric. In the figure, for example, ‘E-STGCN-3.27’ means that the average rank of the proposed E-STGCN algorithm based on the RMSE metric is 3.27 for the $PM_{2.5}$ dataset; the same explanation applies to other algorithms and datasets. The shaded region depicts the reference value of the test.

material, Figures A.1 and A.2 of Appendix A.4, show similar patterns. Overall, the MCB test results, evaluated using both average and distributional forecast metrics, highlight that the proposed E-STGCN approach consistently produces accurate forecasts and effectively captures extreme concentration levels across various air pollutants.

Next, we employ the DM test to assess whether the forecasting performance of the proposed E-STGCN framework significantly differs from that of the baseline models. Specifically, for any baseline architecture \mathcal{A} and the proposed E-STGCN model, we compute the multivariate loss differential series for a given station as

$$\Lambda_{t,\mathcal{A}}^i = \left| X_t^i - \widehat{X}_{t,\mathcal{A}}^i \right| - \left| X_t^i - \widehat{X}_{t,E-STGCN}^i \right|,$$

where X_t^i represents the ground truth data for station i at time t with $\widehat{X}_{t,E-STGCN}^i$ and $\widehat{X}_{t,\mathcal{A}}^i$ being the corresponding forecasts generated by the E-STGCN and model \mathcal{A} , respectively. This statistical testing procedure checks whether the expected loss differential is zero using the DM statistic as

$$DM \text{ statistic for station } i = \sqrt{q} \frac{\mu_{\Lambda}^i}{\alpha_{\Lambda}^i},$$

where q is the forecast horizon, μ_{Λ}^i and α_{Λ}^i are, respectively, the sample mean and standard deviation of the loss differential series $\Lambda_{t,\mathcal{A}}^i$. Using this statistic, we test the null hypothesis $H_0: \mathbb{E}(\Lambda_{t,\mathcal{A}}^i) \leq 0$ against the alternative $H_1: \mathbb{E}(\Lambda_{t,\mathcal{A}}^i) > 0$, where $\mathbb{E}(\cdot)$ denotes expectation. If the p -value of the test is less than the significance level, we reject H_0 and conclude that the forecasting performance of the E-STGCN framework is superior to that of \mathcal{A} architecture. In our analysis, we conduct the DM test to assess the statistical significance of the performance differences between E-STGCN and the second and third-best-performing baselines, modified STGCN and NBeats. Figure 9 present the test results for forecasting $PM_{2.5}$ and PM_{10} concentrations during the October–November–December period. This plot evaluates the station-wise forecasting performance of E-STGCN with the benchmarks, where the X-axis represents station indices and the Y-axis indicates DM test statistics. A positive DM test statistic value indicates the superiority of the E-STGCN over the baselines, while a negative value suggests that the baselines perform better.

As highlighted in the plot, the E-STGCN method performs similarly to or better than the baselines across most stations, except for PM_{2.5} forecasting of CRRM Mathura Road (station no. 7), where modified STGCN achieves superior results. Moreover, the significant p -values at 1%, 5%, 10%, and 20% levels are marked using orange, green, blue, and violet-coloured stars, respectively. As evident from Figure 9, E-STGCN significantly outperforms NBeats for PM_{2.5} forecasting in 19 out of 37 stations at a 1% significance level. Compared to modified STGCN, E-STGCN demonstrates significantly different performance for multiple stations at varied levels. For the PM₁₀ forecasting, we observe that E-STGCN achieves significantly better results than both the modified STGCN and NBeats for several monitoring stations at 1% and 5% levels of significance. The overall findings of the DM test are consistent with the MAE metrics reported in the experimental evaluations. Hence, this test underscores the statistical significance of our findings. For NO₂ concentration levels, the forecasts from the E-STGCN and the modified STGCN models are very similar due to the absence of many significant extreme observations, resulting in $\Lambda_{t,STGCN}^i \approx 0$, rendering the DM statistic undefined in this case.

4.7 Uncertainty quantification

In addition to producing the point forecasts of the air pollutant concentrations through the E-STGCN approach, we quantify the uncertainty inherent with these forecasts using the conformal prediction technique (Vovk et al., 2005). This distribution-free approach generates the probabilistic intervals around the point estimates based on a conformal score (γ_t). The computation of γ_t at time t involves modelling p -lagged values of the target series \mathbf{X}_t using both E-STGCN and an uncertainty model \mathcal{U} as follows:

$$\gamma_t = \frac{|\mathbf{X}_t - \text{E-STGCN}(\mathbf{X}_{t-p})|}{\mathcal{U}(\mathbf{X}_{t-p})}.$$

Subsequently, using the sequential nature of \mathbf{X}_t and γ_t , we derive the conformal quantile by applying a weighted aggregation technique with a fixed window $\{v_t = 1(\chi \geq t - v), \chi < t\}$ of size v as

$$\kappa_t = \inf \left\{ \omega : \frac{1}{\min(v, \chi - 1)} \sum_{\chi=1}^{t-1} \gamma_\chi v_\chi \geq 1 - \rho \right\},$$

where ρ is the significance level. Then, the computation of the conformal prediction interval using the conformal quantiles κ_t can be expressed as

$$[\text{E-STGCN}(\mathbf{X}_{t-p}) \pm \kappa_t \mathcal{U}(\mathbf{X}_{t-p})].$$

In Figure 10, we present the point and interval estimate of air pollutant concentrations generated by the E-STGCN model along with the results of the two best-performing models, modified STGCN and NBeats, as identified by the MCB plots in Section 4.6. The conformal prediction intervals demonstrated in Figure 10 are calculated with $\rho = 0.20$ for three selected forecasting scenarios. The plot highlights the generalizability of the proposed E-STGCN model in providing valuable insights about air pollutant concentration levels, mainly modelling their threshold exceedance values. These findings are pivotal for environmentalists in designing awareness and mitigation strategies.

5 Policy implications

Rapid urbanization and industrialization have significantly impacted air quality in many developing and underdeveloped countries. In its 2021 Global Air Quality Guidelines (AQGs), WHO recommended critical air pollutants such as PM, NO₂, SO₂, O₃, and CO based on their effects on mortality and human health. Among these, PM and NO₂ have gained particular attention from air quality researchers due to their direct links to increased mortality, as evidenced in epidemiological studies (Olaniyan et al., 2020). NO₂ is a highly reactive gas primarily emitted from

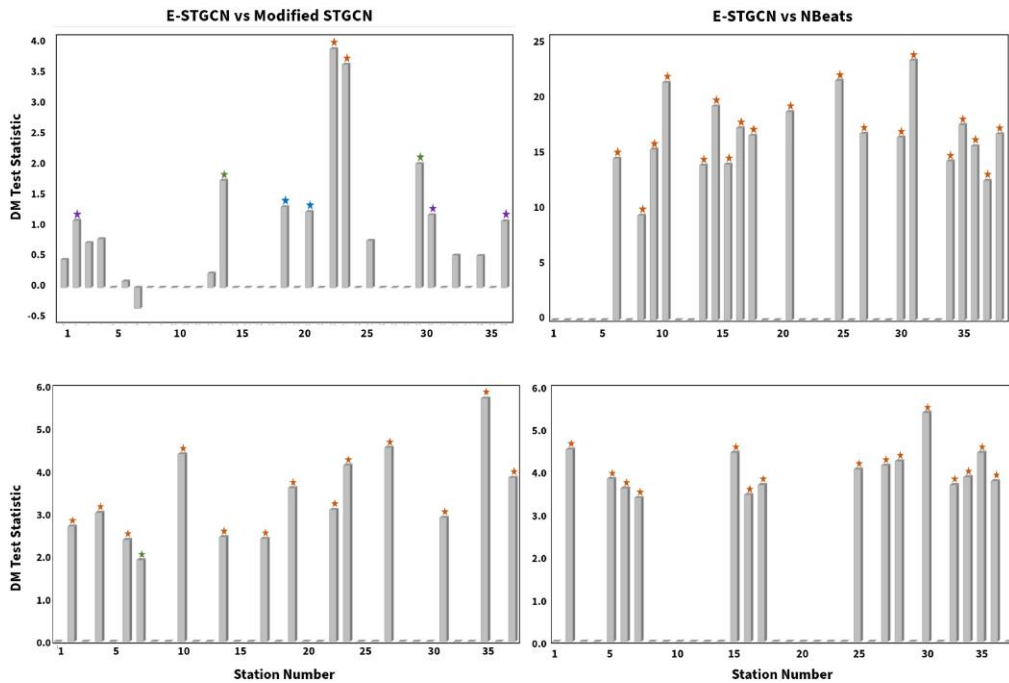


Figure 9. Diebold–Mariano (DM) test results comparing (left) extreme value theory guided modified spatio-temporal graph convolutional networks (E-STGCN) and modified spatio-temporal graph convolutional networks (STGCN), and (right) E-STGCN and NBeats for forecasting PM_{2.5} (upper panel) and PM₁₀ (lower panel) pollutant concentrations over the 90-day OCT–NOV–DEC forecast window. The Y-axis represents DM test statistic values based on the mean absolute error (MAE) metric, while the X-axis indicates the monitoring station indices. Stars denote significant p -values, with colours representing 1% (orange), 5% (green), 10% (blue), and 20% (violet) significance levels, respectively.

automobile exhaust, power plants, and industrial machinery. In urban areas, NO₂ levels are mainly driven by the transportation sector. For instance, the urban regions of North America and Europe often report higher NO₂ despite low levels of PM_{2.5} and PM₁₀ (Ji et al., 2022). It was found that acute exposure to NO₂ can aggravate respiratory diseases, such as asthma and other pulmonary symptoms, although no causal relationship between NO₂ exposure and health mortality was established (Faustini et al., 2014). On the other hand, PM comprises a mix of acids (such as nitrates and sulphates), organic chemicals, metals, soil, dust particles, and allergens. These particles originate from various sources like fossil fuel combustion, industrial emissions, construction activities, wildfires, stubble burning, and household cooking. In Delhi, PM levels frequently exceed the NAAQS, even when NO₂ concentrations remain relatively low (Abirami & Chitra, 2021). It is particularly critical as PM is identified as a causal factor for cardiovascular and respiratory mortality and remains a serious concern for India’s capital. Given that the population of Delhi and the National Capital region (NCR) of India is particularly vulnerable to PM exposure, which can cause health emergencies, this study aims to forecast PM levels in Delhi by considering extreme behaviours of air pollutants. The proposed E-STGCN framework offers a technological solution for real-time monitoring and forecasting of hazardous air pollutants in Delhi. This approach is particularly valuable when extreme observations and nonlinear patterns characterize the observed spatio-temporal data. The proposed methodology has the potential to advance future research endeavours on enhancing air quality forecasting models and to promote environmental sustainability. Although E-STGCN has been developed specifically for air pollution data in this study, it can also be extended to other applied fields, including epidemiology, seismology, and transportation research, where similar patterns of extreme events and complex dependencies are frequently observed.

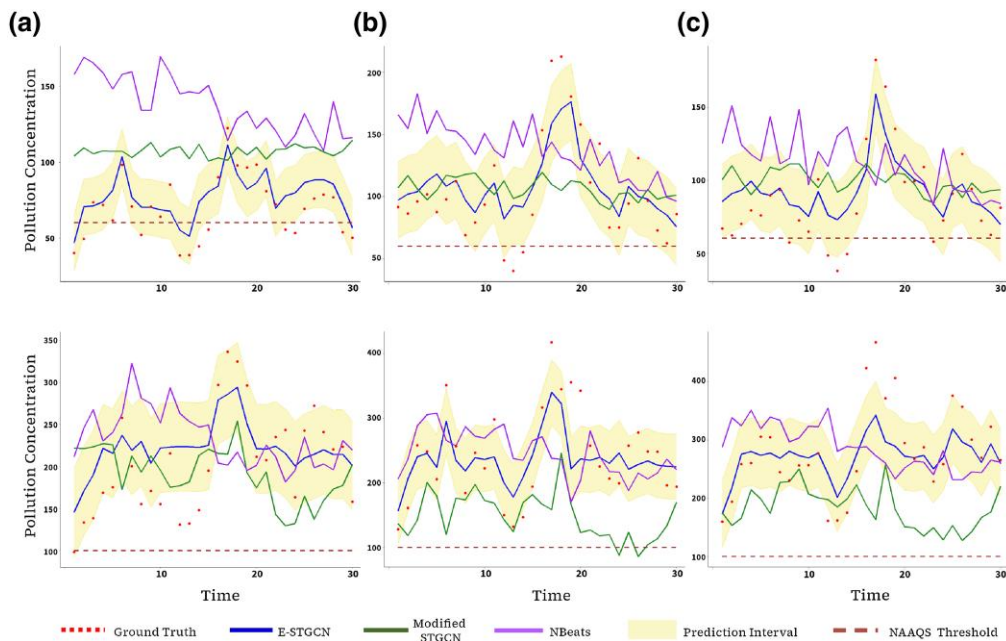


Figure 10. Upper panel presents ground truth (red dots) $PM_{2.5}$ pollutant concentrations monitored at (a) DTU, (b) Dr. Karni Singh Shooting Range, and (c) IGI Airport stations during February 2023 window and corresponding point forecasts of E-STGCN (blue line), modified STGCN (green line), and NBeats (violet line) framework. The conformal prediction interval (yellow-shaded) of the E-STGCN model quantifies the associated uncertainty. The lower panel highlights similar information about PM_{10} concentrations monitored at the corresponding stations.

6 Conclusion

Accurate air quality forecasting remains a challenging problem due to complex spatio-temporal dependencies in pollutant concentration levels. Pollutants such as $PM_{2.5}$, PM_{10} , and NO_2 often exhibit extreme behaviours while also displaying nonlinear and nonstationary properties. Among these hazardous pollutants, $PM_{2.5}$ and PM_{10} concentrations are consistently high in some of the world’s majorly polluted cities, leading to serious health hazards and restricting economic growth. In particular, air pollution levels can intensify with seasonal variations. For instance, in Delhi, the concentration of $PM_{2.5}$ and PM_{10} increases rapidly during winter due to low wind speed, stubble burning, firecracker emissions, and other contributing sources. To address these challenges, public awareness through early warning systems is of paramount importance.

In this study, we propose the E-STGCN model, which aims to provide actionable insights by generating real-time forecasts of air pollutant concentrations. Our approach bridges the gap between existing EVT-based models, which focus on predicting extreme behaviour, and data-driven forecasting methods that predict future trajectories without accounting for the tail behaviour of the extreme observations. By integrating EVT knowledge with modified spatio-temporal GCNs, our proposed framework effectively performs spatio-temporal forecasting while tackling extreme observations. Experimental results, conducted on real-world air pollutant data (daily frequencies) of $PM_{2.5}$, PM_{10} , and NO_2 from 37 monitoring stations in Delhi demonstrate that the E-STGCN approach is well suited for predicting the future dynamics of nonlinear and nonstationary datasets with spatio-temporal dependence and extreme events. Additionally, the model generates appropriate probabilistic bands along with point forecasts, enabling environmental advocates to monitor air pollution trends and design effective control strategies. Further, the forecastability and statistical significance tests conducted in this study verify the effectiveness and robustness of the proposed architecture for air pollution forecasting over various time horizons.

An interesting avenue for future research is to identify various climatic, transportation, and industrial indices that have a causal impact on rising air pollution levels. Future studies could explore how incorporating these causal covariates might enhance the accuracy of the E-STGCN

approach. Additionally, assessing model performance on BM (e.g. monthly or seasonal peaks) could offer deeper insights into the model's ability to forecast extreme events. Such an evaluation would complement the EVT-guided training objective by directly linking statistical tail modelling to empirical performance during high-impact pollution events. Another potential direction would be to extend the air pollution modelling capabilities of E-STGCN on a global scale, analysing its impact on pollution-related mortality and morbidity.

Conflicts of interest: The authors declare no competing interest.

Funding

No funding was received to assist with the preparation of this manuscript.

Data availability

Data and codes are available in the GitHub repository: https://github.com/mad-stat/E_STGCN.

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series B*.

References

- Abirami S., & Chitra P. (2021). Regional air quality forecasting using spatiotemporal deep learning. *Journal of Cleaner Production*, 283, Article 125341. <https://doi.org/10.1016/j.jclepro.2020.125341>
- AL-Dhurafi N. A., Masseran N., Zamzuri Z. H., & Razali A. M. (2018). Modeling unhealthy air pollution index using a peaks-over-threshold method. *Environmental Engineering Science*, 35(2), 101–110. <https://doi.org/10.1089/ees.2017.0077>
- Atluri G., Karpatne A., & Kumar V. (2018). Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys (CSUR)*, 51(4), 1–41. <https://doi.org/10.1145/3161602>
- Balkema A. A., & De Haan L. (1974). Residual life time at great age. *Annals of Probability*, 2(5), 792–804. <https://doi.org/10.1214/aop/1176996548>
- Benktander G., & Segerdahl C.-O. (1960). On the analytical representation of claim distributions with special reference to excess of loss reinsurance. In *Transactions of the international congress of actuaries* (pp. 1–10).
- Box G. E., Jenkins G. M., Reinsel G. C., & Ljung G. M. (1970). *Time series analysis: Forecasting and control*. John Wiley & Sons.
- Brunekreef B., & Holgate S. T. (2002). Air pollution and health. *The Lancet*, 360(9341), 1233–1242. [https://doi.org/10.1016/S0140-6736\(02\)11274-8](https://doi.org/10.1016/S0140-6736(02)11274-8)
- Byun D., & Schere K. L. (2006). Review of the governing equations, computational algorithms, and other components of the models-3 community multiscale air quality (CMAQ) modeling system. *Applied Mechanics Reviews*, 59(2), 51–77. <https://doi.org/10.1115/1.2128636>
- Castillo E. (2012). *Extreme value theory in engineering*. Elsevier.
- Chen J., Lei X., Zhang L., & Peng B. (2015). Using extreme value theory approaches to forecast the probability of outbreak of highly pathogenic influenza in Zhejiang, China. *PLoS One*, 10(2), Article e0118521. <https://doi.org/10.1371/journal.pone.0118521>
- Chen Y., Kang Y., Chen Y., & Wang Z. (2020). Probabilistic forecasting with temporal convolutional neural network. *Neurocomputing*, 399(1), 491–501. <https://doi.org/10.1016/j.neucom.2020.03.011>
- Chukwudum Q. C., Mwita P., & Mung'atu J. K. (2020). Optimal threshold determination based on the mean excess plot. *Communications in Statistics: Theory and Methods*, 49(24), 5948–5963. <https://doi.org/10.1080/03610926.2019.1624772>
- Cliff A., & Ord J. (1975). Model building and the analysis of spatial pattern in human geography. *Journal of the Royal Statistical Society: Series B (Methodological)*, 37(3), 297–328. <https://doi.org/10.1111/j.2517-6161.1975.tb01548.x>
- Coles S., Bawa J., Trenner L., & Dorazio P. (2001). *An introduction to statistical modeling of extreme values (Vol. 208)*. Springer.
- Das D., Athulya R., Chakraborty T., Ray A., Hens C., Dana S. K., Ghosh D., & Murukesh N. (2025). Pattern change of precipitation extremes in Svalbard. *Scientific Reports*, 15(1), 8754. <https://doi.org/10.1038/s41598-025-92339-4>
- Diebold F. X., & Mariano R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 20(1), 134–144. <https://doi.org/10.1198/073500102753410444>

- Du S., Li T., Yang Y., & Horng S.-J. (2019). Deep air quality forecasting using hybrid deep learning framework. *IEEE Transactions on Knowledge and Data Engineering*, 33(6), 2412–2424. <https://doi.org/10.1109/TKDE.2019.2954510>
- Durbin J., & Watson G. S. (1971). Testing for serial correlation in least squares regression. III. *Biometrika*, 58(1), 1–19. https://doi.org/10.1007/978-1-4612-4380-9_20
- Durocher M., Burn D. H., & Ashkar F. (2019). Comparison of estimation methods for a nonstationary index-flood model in flood frequency analysis using peaks over threshold. *Water Resources Research*, 55(11), 9398–9416. <https://doi.org/10.1029/2019WR025305>
- Embrechts P., Klüppelberg C., & Mikosch T. (2013). *Modelling extremal events: For insurance and finance* (Vol. 33). Springer Science & Business Media.
- Faustini A., Rapp R., & Forastiere F. (2014). Nitrogen dioxide and mortality: Review and meta-analysis of long-term studies. *European Respiratory Journal*, 44(3), 744–753. <https://doi.org/10.1183/09031936.00114713>
- Fisher R. A., & Tippett L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society* (Vol. 24, pp. 180–190). Cambridge University Press.
- Fletcher R. (2000). *Practical methods of optimization*. John Wiley & Sons.
- Gao X., & Li W. (2021). A graph-based LSTM model for PM_{2.5} forecasting. *Atmospheric Pollution Research*, 12(9), Article 101150. <https://doi.org/10.1016/j.apr.2021.101150>
- Ghosh S., & Resnick S. (2010). A discussion on mean excess plots. *Stochastic Processes and Their Applications*, 120(8), 1492–1517. <https://doi.org/10.1016/j.spa.2010.04.002>
- Gilmer J., Schoenholz S. S., Riley P. F., Vinyals O., & Dahl G. E. (2017). Neural message passing for quantum chemistry. In *International Conference on Machine Learning* (pp. 1263–1272). PMLR.
- Gneiting T., & Raftery A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378. <https://doi.org/10.1198/016214506000001437>
- Gneiting T., & Ranjan R. (2011). Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, 29(3), 411–422. <https://doi.org/10.1198/jbes.2010.08110>
- Gneiting T., Wolfrum D., Resin J., Kraus K., Bracher J., Dimitriadis T., Hagenmeyer V., Jordan A. I., Lerch S., Phipps K., & Schienle M. (2023). Model diagnostics and forecast evaluation for quantiles. *Annual Review of Statistics and Its Application*, 10(1), 597–621. <https://doi.org/10.1146/statistics.2023.10.issue-1>
- Grimshaw S. D. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics: A Journal of Statistics for the Physical, Chemical, and Engineering Sciences*, 35(2), 185–191. <https://doi.org/10.1080/00401706.1993.10485040>
- Guinness J. (2018). Permutation and grouping methods for sharpening Gaussian process approximations. *Technometrics: A Journal of Statistics for the Physical, Chemical, and Engineering Sciences*, 60(4), 415–429. <https://doi.org/10.1080/00401706.2018.1437476>
- Gumbel E. J. (1958). *Statistics of extremes*. Columbia University Press.
- Hochreiter S., & Schmidhuber J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hyndman R. (2018). *Forecasting: Principles and practice*. OTexts.
- Ji J. S., Liu L., Zhang J., Kan H., Zhao B., Burkart K. G., & Zeng Y. (2022). NO₂ and PM_{2.5} air pollution co-exposure and temperature effect modification on pre-mature mortality in advanced age: a longitudinal cohort study in China. *Environmental Health*, 21(1), 97. <https://doi.org/10.1186/s12940-022-00901-8>
- Jin M., Koh H. Y., Wen Q., Zambon D., Alippi C., Webb G. I., King I., & Pan S. (2024). A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46, 10466–10485. <https://doi.org/10.1109/TPAMI.2024.3443141>
- Kan H.-D., & Chen B.-H. (2004). Statistical distributions of ambient air pollutants in Shanghai, China. *Biomedical and Environmental Sciences*, 17(3), 366–372.
- Karniadakis G. E., Kevrekidis I. G., Lu L., Perdikaris P., Wang S., & Yang L. (2021). Physics-informed machine learning. *Nature Reviews: Physics*, 3(6), 422–440. <https://doi.org/10.1038/s42254-021-00314-5>
- Katz R. W., Parlange M. B., & Naveau P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, 25(8–12), 1287–1304. [https://doi.org/10.1016/S0309-1708\(02\)00056-8](https://doi.org/10.1016/S0309-1708(02)00056-8)
- Kipf T. N., & Welling M. (2016). ‘Semi-supervised classification with graph convolutional networks’, arXiv, arXiv:1609.02907, preprint: not peer reviewed. <https://doi.org/10.48550/arXiv.1609.02907>
- Koning A. J., Franses P. H., Hibon M., & Stekler H. O. (2005). The M3 competition: Statistical tests of the results. *International Journal of Forecasting*, 21(3), 397–409. <https://doi.org/10.1016/j.ijforecast.2004.10.003>
- Kumar U., & Jain V. (2010). ARIMA forecasting of ambient air pollutants (O₃, NO, NO₂ and CO). *Stochastic Environmental Research and Risk Assessment*, 24(5), 751–760. <https://doi.org/10.1007/s00477-009-0361-8>
- Lei M. T., Monjardino J., Mendes L., Gonçalves D., & Ferreira F. (2019). Macao air quality forecast using statistical methods. *Air Quality, Atmosphere & Health*, 12(9), 1049–1057. <https://doi.org/10.1007/s11869-019-00721-9>

- Lelieveld J., Evans J. S., Fnais M., Giannadaki D., & Pozzer A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, 525(7569), 367–371. <https://doi.org/10.1038/nature15371>
- Li X., Peng L., Yao X., Cui S., Hu Y., You C., & Chi T. (2017). Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental Pollution*, 231, 997–1004. <https://doi.org/10.1016/j.envpol.2017.08.114>
- Marimoutou V., Raggad B., & Trabelsi A. (2009). Extreme value theory and value at risk: Application to oil market. *Energy Economics*, 31(4), 519–530. <https://doi.org/10.1016/j.eneco.2009.02.005>
- Martins L. D., Wikuats C. F. H., Capucim M. N., de Almeida D. S., da Costa S. C., Albuquerque T., Carvalho V. S. B., de Freitas E. D., de Fátima Andrade M., & Martins J. A. (2017). Extreme value analysis of air pollution data and their comparison between two large urban regions of South America. *Weather and Climate Extremes*, 18(6), 44–54. <https://doi.org/10.1016/j.wace.2017.10.004>
- Nag P., Sun Y., & Reich B. J. (2023). Spatio-temporal deepkriging for interpolation and probabilistic forecasting. *Spatial Statistics*, 57(4), Article 100773. <https://doi.org/10.1016/j.spasta.2023.100773>
- Olaniyan T., Jeebhay M., Rööslä M., Naidoo R. N., Künzli N., de Hoogh K., Saucy A., Badpa M., Baatjies R., Parker B., Leaner J., & Dalvie M. A. (2020). The association between ambient NO₂ and PM_{2.5} with the respiratory health of school children residing in informal settlements: A prospective cohort study. *Environmental Research*, 186, Article 109606. <https://doi.org/10.1016/j.envres.2020.109606>
- Ong B. T., Sugiura K., & Zettsu K. (2016). Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting PM_{2.5}. *Neural Computing & Applications*, 27(6), 1553–1566. <https://doi.org/10.1007/s00521-015-1955-3>
- Oreshkin B. N., Carpo D., Chapados N., & Bengio Y. (2019). ‘N-beats: Neural basis expansion analysis for interpretable time series forecasting’, arXiv, arXiv:1905.10437, preprint: not peer reviewed. <https://doi.org/10.48550/arXiv.1905.10437>
- Pandey A., Brauer M., Cropper M. L., Balakrishnan K., Mathur P., Dey S., Turkoglu B., Kumar G A., Khare M., Beig G., Gupta T., Krishnakutty R. P., Causey K., Cohen A. J., Bhargava S., Aggarwal A. N., Agrawal A., Awasthi S., Bennett F., ... Dandona L. (2021). Health and economic impact of air pollution in the states of India: The Global Burden of Disease Study 2019. *The Lancet: Planetary Health*, 5(1), e25–e38. [https://doi.org/10.1016/S2542-5196\(20\)30298-9](https://doi.org/10.1016/S2542-5196(20)30298-9)
- Pfeifer P. E., & Deutch S. J. (1980). A three-stage iterative procedure for space-time modeling. *Technometrics: A Journal of Statistics for the Physical, Chemical, and Engineering Sciences*, 22(1), 35–47. <https://doi.org/10.2307/1268381>
- Pickands III J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3, 119–131.
- Raissi M., Perdikaris P., & Karniadakis G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378, 686–707. <https://doi.org/10.1016/j.jcp.2018.10.045>
- Ray A., Chakraborty T., Radhakrishnan A., Hens C., Dana S. K., Ghosh D., & Murukesh N. (2023). ‘Pattern change of precipitation extremes in bear island’, arXiv, arXiv:2312.04502, preprint: not peer reviewed. <https://doi.org/10.48550/arXiv.2312.04502>
- Reiss R.-D., Thomas M., & Reiss R. (1997). *Statistical analysis of extreme values (Vol. 2)*. Springer.
- Roberts E. (1979). Review of statistics of extreme values with applications to air quality data: Part II. applications. *Journal of the Air Pollution Control Association*, 29(7), 733–740. <https://doi.org/10.1080/00022470.1979.10470856>
- Rocco M. (2014). Extreme value theory in finance: A survey. *Journal of Economic Surveys*, 28(1), 82–108. <https://doi.org/10.1111/joes.2014.28.issue-1>
- Saha A., Singh K., Ray M., & Rathod S. (2020). A hybrid spatio-temporal modelling: An application to space-time rainfall forecasting. *Theoretical and Applied Climatology*, 142(3–4), 1271–1282. <https://doi.org/10.1007/s00704-020-03374-2>
- Salinas D., Flunkert V., Gasthaus J., & Januschowski T. (2020). Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3), 1181–1191. <https://doi.org/10.1016/j.ijforecast.2019.07.001>
- Salvi S., Kumar G A., Dhaliwal R S, Paulson K., Agrawal A., Koul P. A., Mahesh P A, Nair S., Singh V., Aggarwal A. N., Christopher D J, Guleria R., Mohan B V M., Tripathi S. K., Ghoshal A. G., Kumar R V., Mehrotra R., Shukla D K, Dutta E., ... Dandona L. (2018). The burden of chronic respiratory diseases and their heterogeneity across the states of India: The Global Burden of Disease Study 1990–2016. *The Lancet: Global Health*, 6(12), e1363–e1374. [https://doi.org/10.1016/S2214-109X\(18\)30409-1](https://doi.org/10.1016/S2214-109X(18)30409-1)
- Samal K. K. R., Panda A. K., Babu K. S., & Das S. K. (2021). Multi-output TCN autoencoder for long-term pollution forecasting for multiple sites. *Urban Climate*, 39, Article 100943. <https://doi.org/10.1016/j.uclim.2021.100943>
- Scarselli F., Gori M., Tsoi A. C., Hagenbuchner M., & Monfardini G. (2008). The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1), 61–80. <https://doi.org/10.1109/TNN.2008.2005605>

- Sfetsos A., Zoras S., Bartzis J. G., & Triantafyllou A. G. (2006). Extreme value modeling of daily PM10 concentrations in an industrial area. *Fresenius Environmental Bulletin*, 15(8), 841–845.
- Shaddick G., Thomas M. L., Mudu P., Ruggeri G., & Gumy S. (2020). Half the world's population are exposed to increasing air pollution. *NPJ Climate and Atmospheric Science*, 3(1), 1–5. <https://doi.org/10.1038/s41612-020-0124-2>
- Thomas M., Lemaitre M., Wilson M. L., Viboud C., Yordanov Y., Wackernagel H., & Carrat F. (2016). Applications of extreme value theory in public health. *PLoS One*, 11(7), Article e0159312. <https://doi.org/10.1371/journal.pone.0159312>
- Vardoulakis S., Fisher B. E., Pericleous K., & Gonzalez-Flesca N. (2003). Modelling air quality in street canyons: A review. *Atmospheric Environment*, 37(2), 155–182. [https://doi.org/10.1016/S1352-2310\(02\)00857-9](https://doi.org/10.1016/S1352-2310(02)00857-9)
- Vaswani A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 31, 6000–6010. <https://doi.org/10.5555/3295222.3295349>
- Vovk V., Gammelman A., & Shafer G. (2005). *Algorithmic learning in a random world* (Vol. 29). Springer.
- Wang Z., Li J., Wang Z., Yang W., Tang X., Ge B., Yan P., Zhu L., Chen X., Chen H., Wand W., Li J., Liu B., Wang X., Wand W., Zhao Y., Lu N., & Su D. (2014). Modeling study of regional severe hazes over mid-eastern China in January 2013 and its implications on pollution prevention and control. *Science China: Earth Sciences*, 57(1), 3–13. <https://doi.org/10.1007/s11430-013-4793-0>
- Wolpert D. H., & Macready W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1), 67–82. <https://doi.org/10.1109/4235.585893>
- Wu N., Green B., Ben X., & O'Banion S. (2020). 'Deep transformer models for time series forecasting: The influenza prevalence case', arXiv, arXiv:2001.08317, preprint: not peer reviewed. <https://doi.org/10.48550/arXiv.2001.08317>
- Yu B., Yin H., & Zhu Z. (2018, July). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-2018* (pp. 3634–3640). ACM, International Joint Conferences on Artificial Intelligence Organization. <https://doi.org/10.24963/ijcai.2018/505>
- Zhou X., Wang J., Wang J., & Guan Q. (2024). Predicting air quality using a multi-scale spatiotemporal graph attention network. *Information Sciences*, 680(7955), Article 121072. <https://doi.org/10.1016/j.ins.2024.121072>