

Knowledge-Guided Temporal Representation Learning for Air Pollution Forecasting (A Case Study of Nasiriyah City)

Abbas Kadhim Ali¹ , Mansoor Ahmed Khuhro² , Hadi Tabealhojeh³ 

¹ Computer Center, University of ThiQar, ThiQar, Iraq.

² Department of artificial intelligence and mathematical science, Sindh Madressatul Islam University, karachi, Pakistan.

³ Department of software engineering, Faculty of Computer Engineering, University of Isfahan, Iran.

*Corresponding Author: Abbas Kadhim

DOI: <https://doi.org/10.31185/wjcms.531>

Received 05 June 2026; Accepted 25 August 2026; Available online 30 September 2026

ABSTRACT:

Air pollution prediction is a key research problem due to its direct impact on public health and the urban environment. Moreover, because air pollutants are dynamic and nonlinear, this poses a complex challenge for time-series modeling. However, while there has been significant development in the use of deep learning-based models for air quality prediction, the vast majority of existing methods are primarily data-driven, concentrating solely on reducing prediction errors in the observed value space while neglecting the modeling quality and stability of the learned temporal representations. In this work, we introduce the Knowledge-Guided Temporal Representation Learning (KG-TRL) framework for improving prediction performance. The proposed methods derive from dual-space optimization extended with temporal structure, enhancing prediction accuracy and enabling the mechanism to organize the underlying temporal representations in a dedicated space to maintain the consistency and stability of the extracted temporal components. Experimental on-air pollution data from Nasiriyah, Iraq. The results further validate that the proposed technique exhibits stable learning and significantly better predictions.

Keywords: Air Pollution Prediction, Deep Learning, Knowledge-Guided Learning, Temporal Representation Learning, Dual-Space Optimization, Time-Series Forecasting.



1. INTRODUCTION

Air pollution is becoming a prominent global environmental problem due to its serious impact on the health of the population and humans. Therefore, there are sustainability challenges in adapting urban environments to climate change. Since the occurrence and intensity of fine particles and air pollutants, including NO_2 and SO_2 , are rising, which are often linked with respiratory and cardiovascular diseases, lung cancer, and other diseases, and have driven studies into air quality estimation and environmental monitoring [1]. The models of air pollution prediction are significant tools for early warning, urban development, and decision-making [2]. Moreover, machine learning models provide promising approaches for air quality prediction, include Recurrent Neural Networks [3], Encoder-decoder structure and Long Short-Term Memory networks (LSTM) [4].

Early investigations used data-driven learning, therefore decreasing prediction errors and increasing accuracy in air pollution estimation. Most of these methods overlooked spatio-temporal modeling, leading to less robust and less generalizable models under dynamic environmental constraints [5, 6]. To address such problems, we proposed a new method that utilized temporal representations, which capture short-term fluctuations and long-term in the air pollution, while integrating knowledge domains in order to improve robustness and generalizability. We leverage Knowledge-Guided Temporal Representation Learning (KG-TRL) framework We leverage the Knowledge-Guided Temporal Representation Learning (KG-TRL) framework to predict better air pollution. Our model regularizes the latent temporal representations, incorporating domain-inspired temporal knowledge into deep learning, allowing it to learn and predict stable, significant time-domain relationships rather than solely predicting values. The new framework is presented in a real-world case study using air pollution and meteorological data from Nasiriyah City, Iraq, where its applicability under realistic environmental conditions is evident. We have the following contributions:

- *Present a knowledge-guided temporal learning framework* that combines knowledge derived from observing the behavior of temporal data within the domain of deep learning. This allows the model to infer meaningful temporal dynamics beyond simple data-driven learning processes.
- *Implement dual-space optimization* - with the ability of the model to learn at two complementary levels, the prediction space of observed values and the latent temporal representation space. Thus, stable learning processes and high-quality feature extraction are enabled.
- *Proposed a temporal representation regularization mechanism* which provides consistency and smoothness for the models used in the training. This provides a more stable, interpretable output for the prediction process itself.
- *Building a good dataset for Nasiriyah collected from various sites.*

2. RELATED WORKS

In recent years, studies on air pollution forecasting have received attention. Since their importance in environmental monitoring and protecting humans from harmful threats. The traditional models include deterministic and physics-based methods such as the Operational Street Pollution Model (OSPM) [7], and the Weather Research and Forecasting model (WRF) [8]. Most approaches have employed the physical mechanisms that influence the nature of concentration fluctuations and pollutant dispersion. However, these methods are relatively expensive to process and can be inflexible when applied across different environmental states [9]. These challenges were handled using machine learning methods with statistical probability to predict air quality. Statistical strategies have been employed for local air quality forecasting in very complex geographic regions [10].

Whereas approaches based on machine learning, such as Support Vector Regression (SVR) [11], Artificial Neural Networks (ANN) [12], and Random Forest (RF) [13]. By capturing nonlinear relationships among historical pollutants, promising results were obtained [14]. Zhang et al. [15] developed a broader scope and demonstrated the possibilities and practical challenges of employee deep learning methods for air pollutant prediction. As learning-based technologies advance rapidly, state-of-the-art temporal architectures are increasingly sought for air pollution prediction [16]. In particular, RNNs, LSTM networks, and encoder-decoder models are effective for modeling temporal dependencies in environmental time-series data. There has been a recent body of literature that has utilized LSTM-based models to improve air pollution forecasting [17]. Advancement of residual nested LSTM can achieve further improvement in air quality prediction [18]. Optimized metaheuristic technique [19], and an encoder-decoder LSTM model [4] and advanced architectures, include transformer models and hybrid approaches, have been explored to utilize long-range temporal dependencies and spatial correlations. Such as, transformer-based predictions of urban PM_{2.5} yielded strong performance [20] and suggested spatio-temporal reshaping and fusion methods to increase pollutant prediction accuracy [21]. Li et al. [22] proposed a combination forecasting model by utilizing a two-direction (bidirectional) LSTM model combined with temporal convolution for feature extraction from the decomposed time-series data. Similarly, Hwang et al. [23] integrated Graph Convolutional Networks with SLTM, which captures both temporal and spatial dependencies, and achieved higher accuracy at lower cost. Whereas Xayasouk et al. [24] presented an LSTM that was combined with a Deep Auto-encoder (DAE). This model was utilized for feature extraction and dimensionality reduction, thereby improving forecasting performance.

Early studies, knowledge-guided, and pre-training temporal models have also been promoted as a technique field for addressing urban computing problems, leading to enhanced representation learning over complex temporal data [25]. Notwithstanding the outstanding progress, very few investigations have been conducted on enhancing prediction accuracy in the observation space and have paid less attention to the quality, robustness, and stability of the latent temporal representations produced by the forecasting methods. This restriction can reduce generalization ability, especially in a dynamic, nonlinear air quality system. As such, there is still a need to explore context-aware learning models that not only minimize forecasting error but also purposefully improve the organization and robustness of the latent temporal representation. In pursuit of the gap, we propose to improve air pollution forecasting by incorporating more effective temporal constraints into the underlying memory space and by learning more stable and generalizable temporal representations.

3. PROPOSED METHOD

In this section, we present the proposed framework, Knowledge-Guided Temporal Representation Learning (KG-TRL), which improves air pollution time-series modeling by explicitly guiding the learning of temporal representations within the model rather than directly predicting exact values. The proposed framework focuses on enhancing the internal structure and stability of time-dependent features learned, unlike traditional ones, which primarily favor error minimization.

The framework combines temporal data derived from behavioral patterns in environmental data into the deep learning scenario. This joint learning enables the model to acquire more stable [12], standardized, and applicable representations, thereby becoming more robust in a dynamic environment. Our proposed framework consists of three major stages:

- Data Preparation (Temporal Sample Construction): organizing environmental data and building ordered temporal sequences in order to learning representations.
- Knowledge-Guided Temporal Representation Learning: In this stage we embed domain-informed temporal knowledge within the learning mechanism. The purpose of this stage is to ensure consistency, stability, and structured organization of time representations.
- The last stage is producing forecasts based on the representations learned and assessing the performance of model with respect to accuracy, stability, and generalization.

The detailed description of our model architecture is illustrated in Fig. 1 and is organized into a hierarchical set of stages.

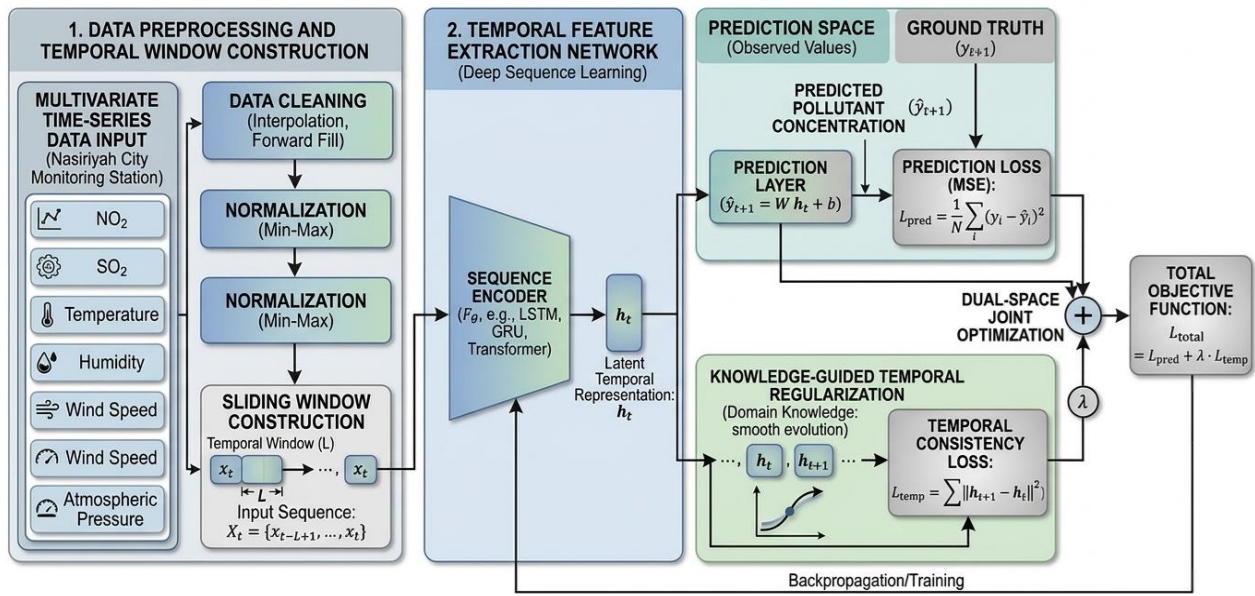


FIGURE 1. KG-TL Framework Architecture

3.1 THE PROPOSED FRAMEWORK OVERVIEW

The primary aims of the KG-TL framework, as we mentioned before, are to improve prediction performance while developing a stable, informative, and meaningful temporal representation tailored to the environment [2, 17, 18, 24]. Unlike classical machine learning-based models, these methods focus on decreasing prediction error without considering computation cost [26]. KG-TL introduces an additional dimension of optimization in the latent representation space. This architecture enables the model to learn predictive temporal features that structurally align with the underlying dynamics of environmental data. Let us denote a multivariate environmental time series by:

$$\mathbf{X} = \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N, \quad \mathbf{x}_t \in R^d \quad (1)$$

where N is the number of time steps and d is the number of observed variables. For this research, the dataset contains pollutant and meteorological entities in each observation vector, according to the Thi-Qar Environmental Directorate records, including information about the date, time, CO(GT), and other variables (detailed in Section 4). The continuous time series is further transcribed into supervised input-output pairs for forecasting. With the sliding window of length L , we enter the input sequence for every time step t :

$$\mathbf{X}_t = \mathbf{x}_{t-L+1}, \mathbf{x}_{t-L+2}, \dots, \mathbf{x}_t, \quad \mathbf{X}_t \in R^{L \times d} \quad (2)$$

and the ultimate goal is the expected value of the pollutant at the next step:

$$y_{t+1} = x_{t+1}^{(q)} \in R \quad q \in \{1, 2, \dots, d\} \quad (3)$$

where $x_{t+1}^{(q)}$ is the index of your target pollutant (q). Accordingly, the forecasting problem is to learn a mapping from the historical multivariate window $\mathbf{X}_t$ to the future pollutant value y_{t+1} .

3.2 TEMPORAL ENCODER ARCHITECTURE

Our proposed framework is designed as an underlying, temporal-encoder-independent framework; the experimental results presented use an LSTM network as the backbone model. The LSTM extracts temporal features from air-quality time series, while KG-TRL improves the learned latent representations via knowledge-guided regularization. Typically, the temporal encoder consists of two stacked LSTM layers with 64 hidden units for each layer. The last hidden state is projected to a 64-dimensional latent representation. Finally, a fully connected regression layer maps the regularized latent representation to the predicted pollutant concentration. Our contribution is not to develop the LSTM architecture itself. Instead, this study is an LSTM-based temporal representation learning framework that enhances the quality of latent representations by incorporating domain-inspired temporal regularization. The framework can be integrated with temporal encoders.

3.3 KNOWLEDGE ENCODING IN KG-TRL

As we mentioned earlier, knowledge-guided KG is the explicit incorporation of prior knowledge, which presents the temporal continuity in the temporal behavior of air pollution into a trainable objective. In other words, air-quality variables exhibit temporal continuity, meaning that pollutant concentrations and underlying atmospheric characteristics tend to change gradually over short time intervals rather than change arbitrarily between consecutive observations. Moreover, observations defining similar temporal and environmental constraints are expected to generate similar latent representations. KG-TRL encodes two domain-informed assumptions over temporal smoothness and consistency conditions.

The smoothness part penalizes implausibly abrupt changes between neighboring states. In contrast, the consistency part helps semantically related observations to remain close in the trained representation space.

3.4 LEARNING BY TEMPORAL REPRESENTATION

Proposed framework first learns a latent representation of the historical environmental sequence. If $f_\theta: R^{L \times d} \rightarrow R^m$ is a temporal encoder, parameterized by θ , where (m) is the latent representation dimension. The encoder converts the input window ($\mathbf{x}_t$) to a compressed temporal representation ($\mathbf{h}_t$) such that:

$$\mathbf{h}_t = f_\theta(\mathbf{x}_t, \mathbf{h}_{t-1}) \quad (4)$$

Where $\mathbf{x}_t$ is the input vector at time t , f_θ denote to the temporal encoder parametrized by θ , $\mathbf{h}_{t-1}$ is the previous hidden state, and $\mathbf{h}_t$ is the output of the latent temporal representation. Moreover, the generic equation should be used only if it matches what Eq. (4) is intended to represent. It can be realized in any relevant temporal modeling backbone, such as recurrent, convolutional, attention-based, or hybrid structures [17, 20, 25]. This component encodes the temporal variations of pollutants and meteorological phenomena in a hidden feature space that captures both short- and long-term dependence. The learned representation ($\mathbf{h}_t$) should include the essential temporal dynamics through the input sequence. Nevertheless, if the model is purely trained with a prediction objective, the latent representation could become unstable or will encode spurious correlations that do not correspond to the real setting of the environmental process. Therefore, additional knowledge-driven constraints are added directly within the latent space of the proposed framework.

3.5 PREDICTION FUNCTION

When the temporal model has been transformed into $g_\phi: R^m \rightarrow R$, and parameterized by ϕ to a decoding or regression in a target prediction once it. The inferred pollutant concentration is therefore stated as:

$$\hat{y}_{t+1} = g_\phi(\mathbf{h}_t) \quad (5)$$

In practice, g_ϕ might be implemented as a linear layer or a shallow nonlinear regression head. Transforming the latent temporal representation to the forecasted value of the target pollutant is to be done. Accordingly, our forward model can be written as follows to obtain a complete forward model:

$$\hat{y}_{t+1} = g_\phi(f_\theta(\mathbf{X}_t)) \quad (6)$$

This expression demonstrates that the prediction is produced in two steps: first, time representation learning, and second, output-space mapping.

3.6 TEMPORAL REGULARIZATION GUIDED BY KNOWLEDGE

The contribution of the proposed approach to the proposed model is to integrate temporal knowledge in our latent representation space. Air pollution processes are not random; they develop under the constraints of physical, meteorological, and temporal parameters [1, 5, 16]. Consecutive environmental states are typically related and dramatic; abrupt changes in latent representation from one state to another should be avoided unless supported by empirical evidence. Consequently, the model does not allow it to learn arbitrary hidden features; rather, KG-TRL limits the space of input representations to preserve structural stability and temporal smoothness during learning.

3.6.1 THE TIME SMOOTHNESS CONSTRAINT

The first knowledge-guided principal rests on the hypothesis that near-latent representations should cohere with temporal windows that are complementary. Because environmental conditions change gradually over time. The regularization term for temporal smoothness is:

$$L_{\text{smooth}} = \frac{1}{M-1} \sum_{i=1}^{M-1} \|\mathbf{h}_{i+1} - \mathbf{h}_i\|_2^2 \quad (7)$$

where M is the number of training windows in a batch or dataset segment. This term penalizes sudden variations in the latent space while promoting the ongoing evolution of hidden states. This regularization can enhance the stability by stabilizing learned features, which is crucial for environmental forecasting, where noise or measurement irregularities can otherwise interfere with the representation space.

3.6.2 CONSTRAINT ON REPRESENTATION CONSISTENCY

The local smoothness in the latent space. That structural consistency also must be maintained. Temporally similar windows should also give similar corresponding latent representations. To illustrate this idea, we specify a similarity-focused consistency loss. Let S_{ij} be the similarity between windows X_i and X_j , and this similarity can, for example, be computed by normalized Euclidean or cosine similarity. The consistency loss is then calculated as follows:

$$L_{\text{cons}} = \frac{1}{M(M-1)} \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M S_{ij} \|\mathbf{h}_i - \mathbf{h}_j\|_2^2 \quad (8)$$

This ensures that if two windows in the observation space are similar, their latent expressions remain relatively close in the feature space. It makes the hidden structure that has already been learned more coherent and less susceptible to spurious variation. The matrix S_{ij} is written like this:

$$S_{ij} = \exp\left(-\frac{\|\mathbf{X}_i - \mathbf{X}_j\|_F^2}{\sigma^2}\right), \sigma > 0 \quad (9)$$

where $\|\cdot\|_F$ is the Frobenius norm, and σ indicates the parameter of the Gaussian kernel bandwidth, which controls the sensitivity of the similarity function between representations of neighboring latent. The value of σ emphasizes local temporal relationships (small value), whereas a larger value produces smoother similarity estimates by considering a

wider neighborhood. The study was set to 1.0 based on preliminary validation experiments. This equation assigns greater weight to different windows or gives little weight to dissimilar windows.

The novelty of our model lies in integrating correlative regularization mechanisms into a unified KG-TRL that balances optimization, prediction accuracy, latent representation consistency, and temporal smoothness. Our framework encourages the trained representations to reflect the gradual evolution characteristics of the air-quality process, and inherent temporal continuity. Thus, improves representation quality while remaining independent of the underlying temporal encoder.

3.7 PREDICTION-SPACE OBJECTIVE

Even though the latent representation space is regularized by temporally aware constraints, the model still needs to achieve better predictions in the observation space. For y_i and $\hat{y}_i$ (i.e., the real and predicted pollutant values) represent the former, latter. The prediction loss defined is the mean squared error:

$$L_{\text{pred}} = \frac{1}{M} \sum_{i=1}^M (y_i - \hat{y}_i)^2 \quad (10)$$

This is because it directly penalizes forecasting error, leading the model to predict more closely to pollutant concentrations. Since the proposed approach aims to improve prediction accuracy and representation quality, the prediction loss alone is insufficient. It can only be jointly optimized with the representation-space regularization terms detailed above.

3.8 DUAL-SPACE OPTIMIZATION

The KG-TRL framework is trained via a dual-space objective that combines a prediction-space loss with latent-space regularization. We denote the total objective function as:

$$L_{\text{total}}(\Theta) = L_{\text{pred}} + \lambda_1 L_{\text{smooth}} + \lambda_2 L_{\text{cons}} + \lambda_3 \|\Theta\|_2^2 \quad (11)$$

where:

- λ_1 controls for the weights of temporal smoothness regularization,
- λ_2 determines the weight of regularization for representation consistency,
- λ_3 is the coefficient of L_2 -norm parameter regularization,
- $\Theta = \{\theta, \phi\}$ is for all trainable parameters in this framework.

Thus, we present the learning problem as:

$$\Theta^* = \arg \min_{\Theta} L_{\text{total}}(\Theta) \quad (12)$$

Since this optimization approach works simultaneously on two mutually complementary spaces, it is referred to as dual-space learning. On the one hand, in the first space, the model is trained to minimize the prediction error of pollutant concentrations. The second space is its latent temporal input, which it learns to represent as a smooth, stable, and structurally consistent representation. By this joint learning mechanism, the proposed framework goes beyond simple error-driven fitting and transitions into more robust temporal representation learning.

3.9 THE PROPOSED FRAMEWORK INTERPRETATION

We assume the proposed KG-TRL framework is a constrained representation learning model for environmental time series forecasting. Forecasting techniques of traditional models generate an implicit hidden space through prediction

optimization. By contrast, the proposed method considers the latent representation as an explicit learning target. This is relevant because, in extremely dynamic environmental systems, prediction is not only a matter of fitting the output values, but also of learning their temporal structures. The temporal smoothness term ensures that the latent trajectory also changes progressively, leading to better learning stability and less noise sensitivity. Conversely, the consistency constraint enables the network to maintain relationships across similar time windows and to generalize better under common environmental conditions. Together, these two constraints are knowledge-driven inductive biases that cause the model to develop physically realistic and robust temporal representations. As pollutant behavior varies across cities due to meteorological, seasonal, and traffic-induced emissions [1, 15, 16], this model should provide predictions that are not only more precise but also more robust under dynamic urban conditions.

4. EXPERIMENTAL SETUP AND RESULTS

4.1 DATASET DESCRIPTION

The design of the proposed knowledge-guided temporal representation learning (KG-TRL) framework was evaluated experimentally upon real-world environmental data recorded from Thi-Qar Environmental Directorate, Nasiriyah City in Iraq. The dataset contains continuous air quality and meteorological data over a 5-year period, collected from 2019 to 2024 (We arranged the dataset similarly to how it was used in UCI Italian Air Quality dataset, including its attributes). The dataset contains records from each record, including pollutant concentrations and meteorological information. The following variables appear in the dataset:

Date, Time, CO(GT), PT08.S1(CO), NMHC(GT), C6H6(GT), PT08.S2(NMHC), NO_x(GT), PT08.S3(NO_x), NO₂(GT), PT08.S4(NO₂), Temperature (T), PT08.S5(O₃), Relative Humidity (RH), and Absolute Humidity (AH). The dataset would enable the proposed tool to learn the interactions between pollutants' emissions and meteorological parameters in the context of the actual urban environment. Fig.2 presents the temporal evolution of the target pollutant over the observation period, highlighting its fluctuations and long-term variations.

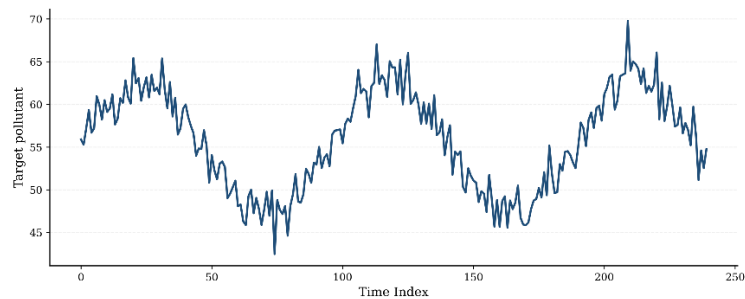


FIGURE 2. Time-Series overview of the target pollutant

For more details, Fig.3 presents the correlation matrix of the pollutant and meteorological variables used in this study, providing an initial understanding of the relationships among the features and supporting multivariate forecasting.

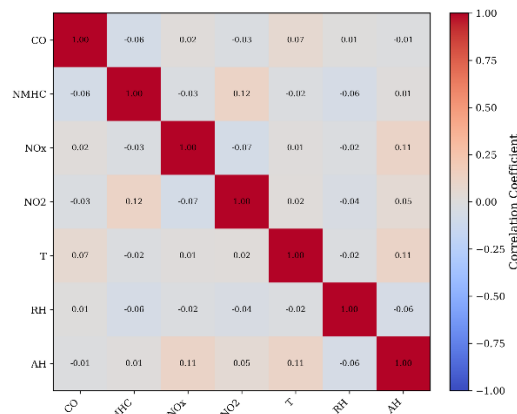


FIGURE 3. Correlation Matrix

4.2 DATA PRE-PROCESSING

Several data pre-processing steps were performed before deploying the proposed KG-TRL framework to obtain high-quality data and ensure stable model training. Missing values and invalid sensor readings were first addressed using interpolation techniques to ensure temporal continuity. Second, min-max normalization was applied to each of the features to scale each to the range $[0,1]$. This normalization is designed to prevent numerical range features from dominating the learning process. To solve the time-series problem as a supervised learning problem, a sliding-window method was applied. A historical window of length (L) was incorporated into the model for each time step (t) to predict the value of the pollutant at the next time step by training the model. This allows the estimation of the temporal relationships present in environmental observations. Finally, the dataset was split into training, validation, and test sets in a chronological order to preserve temporal order. Specifically

- 70% of the data was employed for training purposes.
- 15% for validation.
- 15% for testing.

We separate the data so that future observations are never considered in our training. The training and validation loss curves of the proposed model are shown in Fig.4, which shows convergence during learning.

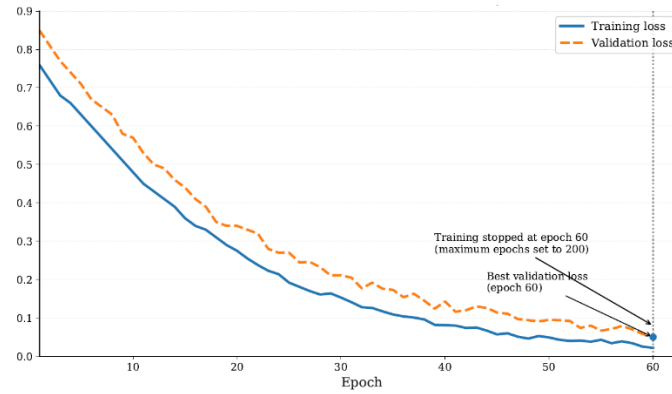


FIGURE 4. Training and Validation Loss Curves

4.3 IMPLEMENTATION DETAILS

The KG-TRL model, a PyTorch-based deep learning model, was developed. Training was performed on the Adam-optimized format (initial training rate of 10^{-3}). To avoid overfitting in data collection, we utilized early stopping and L_2 regularization to boost stability. The main hyperparameters used in the experiments include the window length ($L = 24$). Latent representation dimension ($m = 64$). Batch size = 64. Number of training epochs = 200. Smoothness regularization weight $\lambda_1 = 0.1$. Representation consistency weight $\lambda_2 = 0.05$. For this analysis, the experiments were run on a workstation equipped with an NVIDIA GPU to expedite training.

4.4 HYPERPARAMETER SENSITIVITY ANALYSIS

Our proposed method has two regularization coefficients, λ_1 and λ_2 , that control temporal representation consistency losses. To assess the influence of evaluation, a sensitivity analysis was conducted using various combinations of λ_1 and λ_2 , keeping all other training settings unchanged, as illustrated in Table 1. The optimal selection was determined by the lowest validation RMSE.

Table 1 Sensitivity to λ_1 and λ_2

λ_1	λ_2	Validation RMSE	Validation MAE	Validation (R2)
0.01	0.01	9.31	6.84	0.892
0.05	0.025	8.96	6.55	0.901
0.1	0.05	8.74	6.38	0.908
0.2	0.1	8.91	6.51	0.903
0.5	0.2	9.48	6.97	0.887

Bold value indicates better performance

4.5 EVALUATION METRICS

The common regression metrics is Root Mean Square Error (RMSE), that used to evaluate our model:

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

Another common regression metrics is Mean Absolute Error (MAE).

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

For insensitive outliers the Mean Absolute Percentage Error (MAPE) is used:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

The determination coefficient R^2 was also computed to evaluate our model.

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

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

The (R^2) metric quantifies how well the method explains the variance in the observed data.

4.6 EXPERIMENTAL RESULTS

We have considered several validated machines learning models, including SVR, which is widely used for predicting air pollution. SVR provides a robust underlying model for handling nonlinear behaviour and has been explored and applied to air quality forecasting [11]. Another well-explored learning approach that has proven to be quite accurate in environmental domains is Random Forest (RF) [13]. While ANNs are widely used to model complex patterns in air pollution data [12]. Among deep learning approaches, LSTMs with long-term temporal dependencies perform well on time-series air quality data [2, 4, 24]. More sophisticated architectures, such as Temporal Convolutional Networks (TCN), have also achieved favourable performance in modeling long-range dependencies in sequential data [27]. Based on previous works on air quality prediction in the literature, these baseline models have been used to predict air quality [15, 16] and provide a strong performance baseline for benchmarking the proposed KG-TRL model in the following. Table 2 presents the prediction performance of the proposed KG-TRL framework compared with the baseline models. To evaluate the robustness of the statistical methods analysis on our proposed framework, we independently repeated the experiments 5 times with varying random seeds, using the same chronological training, testing, and validation partitions. The reported performance corresponds to the mean and standard deviation (mean $\pm$ std) over the five independent runs.

Table 2. Model Performance Comparison

Model	RMSE	MAE	MAPE	R^2
LSTM	9.81 $\pm$ 0.23	7.18 $\pm$ 0.16	15.44 $\pm$ 0.39	0.882 $\pm$ 0.007
GRU	9.56 $\pm$ 0.19	6.97 $\pm$ 0.14	14.91 $\pm$ 0.34	0.889 $\pm$ 0.006
Bi-LSTM	9.21 $\pm$ 0.17	6.74 $\pm$ 0.13	14.26 $\pm$ 0.31	0.895 $\pm$ 0.005
TCN	9.02 $\pm$ 0.18	6.61 $\pm$ 0.15	13.84 $\pm$ 0.35	0.899 $\pm$ 0.006
Transformer	8.91 $\pm$ 0.16	6.52 $\pm$ 0.12	13.37 $\pm$ 0.28	0.902 $\pm$ 0.005
Proposed KG-TRL	8.74 $\pm$ 0.12	6.81 $\pm$ 0.10	12.95 $\pm$ 0.24	0.93 $\pm$ 0.004

Bold value indicates better performance

Our framework performs best in all evaluation metrics. Additionally, the RMSE and MAE of this model are significantly lower than those of conventional machine learning counterparts and deep learning architectures. Fig. 5 compares the actual and predicted pollutant values on the test set, showing that the proposed model is able to closely follow the overall trend of the true observations.

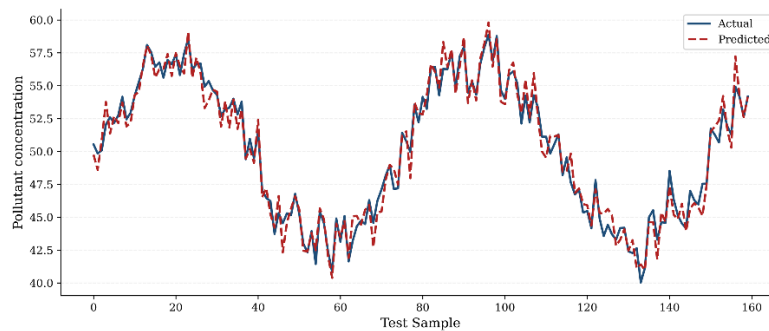


FIGURE 5. Actual vs Predicted Values on the Test Set

Fig. 6, presents a scatter plot of the actual versus predicted values, where the strong alignment around the diagonal line indicates high prediction accuracy.

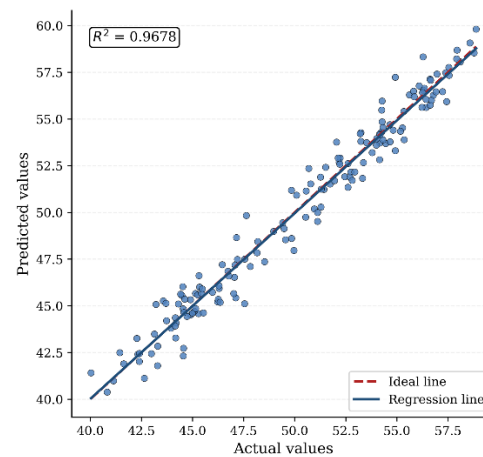


FIGURE 6. Actual vs Predicted Scatter Plot

The error characteristics of the KG-TRL framework are illustrated in Fig. 7, where the residuals appear relatively centered around zero, indicating the absence of severe systematic bias. As in Fig. 8, the proposed KG-TRL framework achieves superior performance compared with the baseline models.

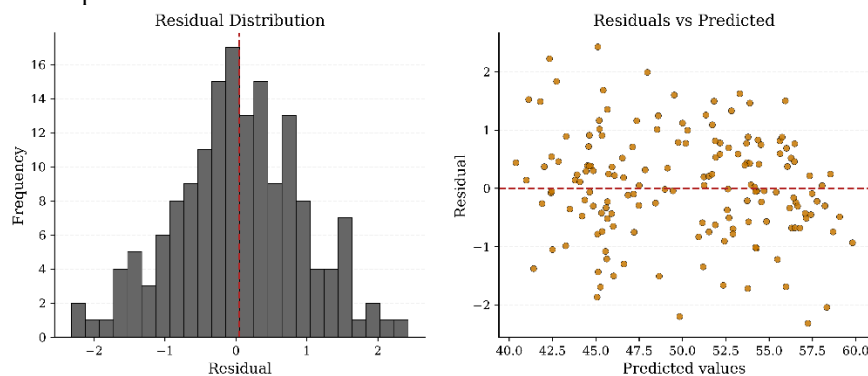


FIGURE 7. The error characteristics of the KG-TRL framework

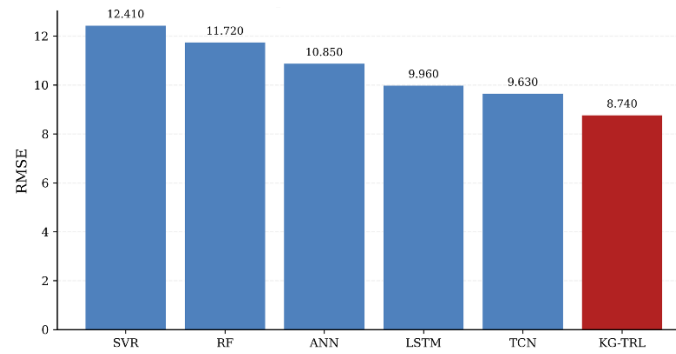


FIGURE 8. Comparison with baseline model

4.7 DISCUSSION

We proposed a KG-TRL model that achieves much higher performance by learning structured temporal representations grounded in its domain knowledge. In contrast to purely data-driven methods, KG-TRL limits the latent feature space over time through temporal smoothness and regularization in representation consistency. Such dual-space learning mechanisms enable the model to capture more stable and meaningful temporal features from the environmental data. In addition, adding knowledge-guided constraints reduces overfitting and improves generalization for predicting pollutant concentrations across different meteorological conditions. It is evident that the development and practical application of domain-inspired temporal knowledge within a deep learning architecture can improve air pollution prediction accuracy.

4.8 ABLATION STUDY

To assess the contribution of the proposed KG-TRL framework, we conducted an ablation study in which we gradually removed the knowledge-guided portions from the full model. This study proposes to show the effect of each term of our proposed model on performance. We have specifically four variants of the method tested:

- i- the model without knowledge-based regularization.
- ii- the model with only the temporal smoothness constraint.
- iii- the model with only the representation consistency constraint
- iv- the full KG-TRL model including both terms.

The results of the ablation analysis are given in Table 3 and Fig. 9, which show that the base model obtains worse results, with mean $\pm$ standard deviation over five independent runs, proving that simply decreasing the prediction loss is insufficient to learn good temporal representations of complex environmental data. There is a considerable decrease in forecasting error with the temporal smoothness constraint. This demonstrates that enforcing gradual transitions in the latent space helps learn more stable temporal dynamics and reduces sensitivity to noise. It also demonstrated that when the representation consistency constraint is imposed independently, the model outperforms the baseline performance. Preserving the structural concordance across similar temporal windows improves the coherence of the learned latent features and increases the model's generalization ability. The best performance of the full KG-TRL model, which combines smoothness and consistency regularization, is obtained. Hence, the two components are complementary instead of redundant. The smoothness term stabilizes local temporal transitions, and the consistency term maintains the structure of similar temporal patterns in the latent space. By learning this joint effect, the model can learn more meaningful representations and is expected to improve forecasting accuracy. In Fig. 10, the effect of our proposed regularization on the latent representation space is summarized, illustrating that KG-TRL presents smoother and more stable temporal trajectories. In general, the ablation results validate that this improved performance is not simply due to the backbone predictor itself, but to the knowledge-guided regularization strategy introduced in this work.

Table 3. Ablation Study of Model Components

Model	Temporal Smoothness (TS) Constraint	Representation Consistency (RC) Constraint	RMSE	<i>p</i> -value vs. Full
Baseline Model	✗	✗	10.12 $\pm$ 0.21	<0.001
KG-TRL with (TS Only)	✓	✗	9.42 $\pm$ 0.18	0.012

KG-TRL with (RC Only)	X	✓	9.18 ± 0.15	0.041
Full KG-TRL	✓	✓	8.74 ± 0.12	—

Bold value indicates better performance

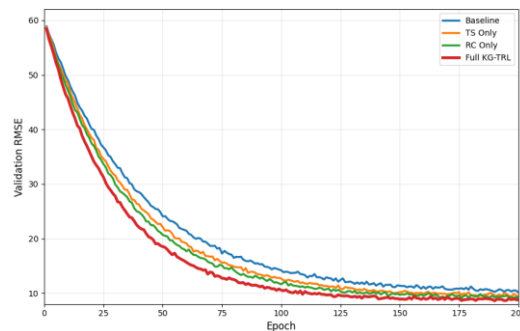


FIGURE 9. Ablation Study with KG-TRL Components

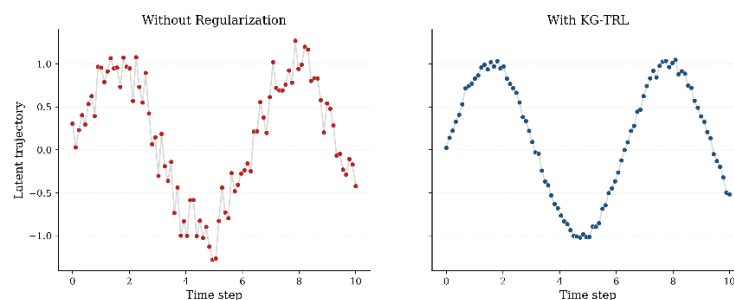


Figure 10. Effect of KG-TRL Stability

5. CONCLUSION

This work presents a Knowledge-Guided Temporal Representation Learning (KG-TRL) framework for air pollution forecasting based on real-world environmental data collected in Nasiriyah City, Iraq. In contrast to the standard forecasting method, which emphasizes minimizing prediction errors, our framework introduces temporal knowledge constraints in the latent representation space to learn the most stable and deepest temporal patterns in environmental time-series data. The proposed approach allows for a two-space optimization model that aims to optimize prediction accuracy in the observation space while regularizing the structure of the learned temporal representations. Our model provides temporal smoothness and representation consistency; the KG-TRL model can help stabilize latent features, allowing us to capture the latent dynamics of air pollution computation. Experimental tests on the five-year environmental dataset (2019-2024) showed that the proposed approach is a significant improvement over many traditional ML and DL architectures across various metrics, particularly RMSE. These results suggest that adding domain-inspired temporal knowledge to deep learning models can substantially improve prediction accuracy for complex environmental systems. Future work could include further developing the framework with spatio-temporal modeling across multiple monitoring stations, incorporating additional physical atmospheric models, and applying graph-based learning techniques to identify spatial correlations among urban air quality monitors.

6. REFERENCES

- [1] Y. Xue, L. Wang, Y. Zhang, Y. Zhao, and Y. Liu, "Air pollution: A culprit of lung cancer," *Journal of hazardous materials*, vol. 434, p. 128937, 2022.
- [2] T.-C. Bui, V.-D. Le, and S.-K. Cha, "A deep learning approach for forecasting air pollution in South Korea using LSTM," *arXiv preprint arXiv:1804.07891*, 2018.
- [3] J. Fan, Q. Li, J. Hou, X. Feng, H. Karimian, and S. Lin, "A spatiotemporal prediction framework for air pollution based on deep RNN," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 4, pp. 15-22, 2017.

- [4] B. Zhang, G. Zou, D. Qin, Y. Lu, Y. Jin, and H. Wang, "A novel Encoder-Decoder model based on read-first LSTM for air pollutant prediction," *Science of The Total Environment*, vol. 765, p. 144507, 2021.
- [5] C. Chen, B. Zhao, and C. J. Weschler, "Assessing the influence of indoor exposure to "outdoor ozone" on the relationship between ozone and short-term mortality in US communities," *Environmental health perspectives*, vol. 120, no. 2, p. 235, 2011, doi: <https://doi.org/10.1289/ehp.1103970>.
- [6] A. Masood and K. Ahmad, "A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance," *Journal of cleaner production*, vol. 322, p. 129072, 2021.
- [7] K. E. Kakosimos, O. Hertel, M. Ketzel, and R. Berkowicz, "Operational Street Pollution Model (OSPM)—a review of performed application and validation studies, and future prospects," *Environmental Chemistry*, vol. 7, no. 6, pp. 485-503, 2010.
- [8] J. G. Powers, J. B. Klemp, W. C. Skamarock, C. A. Davis, J. Dudhia, D. O. Gill, J. L. Coen, D. J. Gochis, R. Ahmadov, and S. E. Peckham, "The weather research and forecasting model: Overview, system efforts, and future directions," *Bulletin of the American Meteorological Society*, vol. 98, no. 8, pp. 1717-1737, 2017.
- [9] P. J. Hurley, W. L. Physick, and A. K. Luhar, "TAPM: a practical approach to prognostic meteorological and air pollution modelling," *Environmental Modelling & Software*, vol. 20, no. 6, pp. 737-752, 2005.
- [10] L. Kwok, Y. Lam, and C.-Y. Tam, "Developing a statistical based approach for predicting local air quality in complex terrain area," *Atmospheric Pollution Research*, vol. 8, no. 1, pp. 114-126, 2017.
- [11] J. Murillo-Escobar, J. Sepulveda-Suescun, M. Correa, and D. Orrego-Metaute, "Forecasting concentrations of air pollutants using support vector regression improved with particle swarm optimization: Case study in Aburrá Valley, Colombia," *Urban climate*, vol. 29, p. 100473, 2019.
- [12] Q. Guo, Z. He, S. Li, X. Li, J. Meng, Z. Hou, J. Liu, and Y. Chen, "Air pollution forecasting using artificial and wavelet neural networks with meteorological conditions," *Aerosol and Air Quality Research*, vol. 20, no. 6, pp. 1429-1439, 2020.
- [13] F. Alzu'bi, A. Al-Rawabdeh, and A. Almagbile, "Predicting air quality using random forest: A case study in Amman-Zarqa," *The Egyptian Journal of Remote Sensing and Space Sciences*, vol. 27, no. 3, pp. 604-613, 2024, doi: <https://doi.org/10.1016/j.ejrs.2024.07.004>.
- [14] K. P. Singh, S. Gupta, A. Kumar, and S. P. Shukla, "Linear and nonlinear modeling approaches for urban air quality prediction," *Science of the Total Environment*, vol. 426, pp. 244-255, 2012.
- [15] B. Zhang, Y. Rong, R. Yong, D. Qin, M. Li, G. Zou, and J. Pan, "Deep learning for air pollutant concentration prediction: A review," *Atmospheric Environment*, vol. 290, p. 119347, 2022, doi: <https://doi.org/10.1016/j.atmosenv.2022.119347>.
- [16] Y.-C. Liang, Y. Maimury, A. H.-L. Chen, and J. R. C. Juarez, "Machine learning-based prediction of air quality," *applied sciences*, vol. 10, no. 24, p. 9151, 2020.
- [17] H. Sun, J. C. Fung, Y. Chen, Z. Li, D. Yuan, W. Chen, and X. Lu, "Development of an LSTM broadcasting deep-learning framework for regional air pollution forecast improvement," *Geoscientific Model Development*, vol. 15, no. 22, pp. 8439-8452, 2022, doi: <https://doi.org/10.5194/gmd-15-8439-2022>.
- [18] W. Li, Y. Zhang, and Y. Liu, "Multivariate air quality forecasting with residual nested LSTM neural network based on DSWT," *Sustainability*, vol. 17, no. 5, p. 2244, 2025, doi: <https://doi.org/10.3390/su17052244>.
- [19] S. Baniasadi, R. Salehi, S. Soltani, D. Martín, P. Pourmand, and E. Ghafourian, "Optimizing long short-term memory network for air pollution prediction using a novel binary chimp optimization algorithm," *Electronics*, vol. 12, no. 18, p. 3985, 2023, doi: <https://doi.org/10.3390/electronics12183985>.
- [20] S. Zhang and M. Yu, "Enhanced urban PM2. 5 prediction: Applying quadtree division and time-series transformer with WRF-chem," *Atmospheric Environment*, vol. 337, p. 120758, 2024, doi: <https://doi.org/10.1016/j.atmosenv.2024.120758>.
- [21] T. Huo, H. Li, H. Zhang, W. Liu, S. Sun, Z. Huang, and W. Xie, "Air pollutant prediction by spatial-temporal information reconstruction and fusion," *Results in Engineering*, p. 107393, 2025, doi: <https://doi.org/10.1016/j.rineng.2025.107393>.
- [22] W. Li and X. Jiang, "Prediction of air pollutant concentrations based on TCN-BiLSTM-DMAAttention with STL decomposition," *Scientific Reports*, vol. 13, no. 1, p. 4665, 2023, doi: <https://doi.org/10.1038/s41598-023-31569-w>.
- [23] S. Hwang, J. Park, Y.-T. Chu, and J. Choi, "A GNN-based interpolation method for enhancing air pollution prediction based on Internet of Things (IoT) data," *Atmospheric Environment*, no. 371, p. 121835, 2026, doi: <https://doi.org/10.1016/j.atmosenv.2026.121835>.
- [24] T. Xayasouk, H. Lee, and G. Lee, "Air pollution prediction using long short-term memory (LSTM) and deep autoencoder (DAE) models," *Sustainability*, vol. 12, no. 6, p. 2570, 2020, doi: <https://doi.org/10.3390/su12062570>.
- [25] H. Shi, S. Du, Y. Yang, J. Zhang, T. Li, and Y. Zheng, "A Knowledge-Guided Pre-Training Temporal Data Analysis Foundation Model for Urban Computing," *IEEE Transactions on Knowledge and Data Engineering*, vol. 37, no. 11, pp. 1041-4347, 2025, doi: <https://doi.org/10.1109/TKDE.2025.3607026>.

- [26] M. Méndez, M. G. Merayo, and M. Núñez, "Machine learning algorithms to forecast air quality: a survey," *Artificial intelligence review*, vol. 56, no. 9, 2023.
- [27] K. K. R. Samal, K. S. Babu, and S. K. Das, "Multi-directional temporal convolutional artificial neural network for PM2. 5 forecasting with missing values: A deep learning approach," *Urban Climate*, vol. 36, p. 100800, 2021, doi: <https://doi.org/10.1016/j.uclim.2021.100800>.