


APPLICATION PAPER  

Simulating the air quality impact of prescribed fires using graph neural network-based PM_{2.5} forecasts

Kyleen Liao¹ , Jatan Buch² , Kara D. Lamb² and Pierre Gentine²

¹Department of Computer Science, Stanford University, Stanford, CA, USA

²Department of Earth and Environmental Engineering, Columbia University, New York City, NY, USA

Corresponding author: Kyleen Liao; Email: kyleenliao@stanford.edu

Received: 18 May 2024; **Revised:** 01 December 2024; **Accepted:** 06 January 2025

Keywords: air quality; forecasting; graph neural network; wildfires

Abstract

The increasing size and severity of wildfires across the western United States have generated dangerous levels of PM_{2.5} concentrations in recent years. In a changing climate, expanding the use of prescribed fires is widely considered to be the most robust fire mitigation strategy. However, reliably forecasting the potential air quality impact from prescribed fires, which is critical in planning the prescribed fires' location and time, at hourly to daily time scales remains a challenging problem. In this paper, we introduce a spatio-temporal graph neural network (GNN)-based forecasting model for hourly PM_{2.5} predictions across California. Utilizing a two-step approach, we use our forecasting model to predict the net and ambient PM_{2.5} concentrations, which are used to estimate wildfire contributions. Integrating the GNN-based PM_{2.5} forecasting model with simulations of historically prescribed fires, we propose a novel framework to forecast their air quality impact. This framework determines that March is the optimal month for implementing prescribed fires in California and quantifies the potential air quality trade-offs involved in conducting more prescribed fires outside the peak of the fire season.

Impact Statement

PM_{2.5} pollution poses significant health risks and is responsible for millions of deaths per year. Our work forecasts the PM_{2.5} concentration at sparse locations and estimates the fire-specific PM_{2.5} contribution to assess their impact on air quality. Furthermore, prescribed fires, while preventing wildfires, also generate PM_{2.5}, raising concerns about the air quality trade-offs. To the best of our knowledge, our work is the first to apply machine learning to predict the PM_{2.5} concentration from simulated prescribed fires. We use our forecasting model to conduct novel experiments that can help the fire service better understand and minimize the pollution exposure from prescribed fires.

1. Introduction

Across many parts of the western United States (WUS), wildfire size, severity, and fire season length have increased due to climate change (Williams et al., 2019). Wildfires across the WUS have led to the largest daily mean PM_{2.5} (particulate matter <2.5 microns) concentrations observed by ground-based sensors in



This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

recent years (Burke et al., 2021) and exposure to PM_{2.5} is responsible for 4.2 million premature deaths worldwide per year (World Health Organization, 2022). Within California, additional PM_{2.5} emissions from extreme wildfires over the past 8 years have reversed nearly two decades of decline in ambient PM_{2.5} concentrations (Burke et al., 2023).

Due to the numerous severe health consequences from PM_{2.5} pollution exposure, performing accurate and temporally fine-grained PM_{2.5} predictions has become increasingly significant. A recent study by Aguilera et al. (2021) found the PM_{2.5} emitted from wildfires to be more toxic than the PM_{2.5} emitted from ambient sources. Accurate PM_{2.5} predictions are also important in the context of prescribed fires, or controlled burns, which have been widely accepted as an effective land management tool and could have the potential to reduce the resulting smoke from future wildfires (Kelp et al., 2023). Since air quality is a major public concern surrounding prescribed fires (McCaffrey, 2006), land managers conducting these burns require access to robust, near real-time predictions of downwind air pollution, to determine suitable locations and burn windows.

However, most data-driven PM_{2.5} forecasting algorithms do not distinguish between ambient PM_{2.5} concentration and the additional concentration due to fire emissions. Previous work studying the effect of prescribed fires on pollution used chemical transport models (CTMs) like the Community Multiscale Air Quality (CMAQ) and Goddard Earth Observing System Atmospheric Chemistry (GEOS-Chem) models to calculate the PM_{2.5} impact of prescribed fires at different locations (Kelp et al., 2023). Jin et al. (2023) and Schollaert et al. (2024) used CTMs to study prescribed fires in the Western United States. Although CTMs can model the complex chemical processes in PM_{2.5} transport, their computational expense is a drawback in generating accurate predictions as well as exploring a large parameter space for simulating prescribed burns (Askariyeh et al., 2020; Byun and Schere, 2006; Zaini et al., 2022; Rybarczyk and Zalakeviciute, 2018).

Recent analyses have combined satellite-based observations with meteorological data to derive daily wildfire-specific PM_{2.5} for all ZIP codes in California (Aguilera et al., 2023) and in 10 km grid cells across the contiguous United States (Childs et al., 2022). Both these studies relied on smoke plume boundaries identified by expert input to identify smoke exposure per day and grid cell, and used the PM_{2.5} measurements on non-smoke days to estimate the PM_{2.5} contribution from background sources. Specifically, Childs et al. (2022) (Childs et al., 2022) defined wildfire-specific PM_{2.5} as anomalies relative to monthly mean PM_{2.5} concentration in each grid cell and adopted gradient-boosted trees to predict the anomalous PM_{2.5} as a function of fire-related predictors. On the other hand, Aguilera et al. (2023) trained an ensemble of machine learning models to first estimate the total PM_{2.5} concentration for each ZIP code, then obtained the wildfire-specific PM_{2.5} by subtracting the background PM_{2.5} imputed from model predictions of PM_{2.5} for non-smoke days. Meanwhile, Qiu et al. (2024) found that CTMs overestimated the PM_{2.5} concentrations for extreme smoke events in 2020 while the wildfire-specific PM_{2.5} predictions from the Childs et al. (2022) machine learning model were in much better agreement with surface PM_{2.5} measurements.

Our research builds upon the GNN-GRU machine learning model named PM_{2.5}-GNN from Wang et al. (2020), which was used to forecast non-wildfire-influenced PM_{2.5} pollution in China. In contrast, our work focuses on predicting fire-influenced PM_{2.5} at different sensor locations across California. The spatio-temporal modeling capabilities of the PM_{2.5}-GNN coupled with domain knowledge make the model especially valuable for PM_{2.5} prediction with spatially sparse monitor observations, enabling the PM_{2.5}-GNN to outperform baseline machine learning architectures such as random forest (RF), long short-term memory (LSTM), and multilayer perceptron (MLP) models. Our PM_{2.5}-GNN model predicts the PM_{2.5} pollution at an hourly resolution in California and considers two applications: (1) quantifying fire-specific PM_{2.5} concentration and (2) forecasting the pollution levels emitted from simulated prescribed fire events. Specifically, we incorporate satellite-derived data on fire intensity within the PM_{2.5}-GNN model to forecast PM_{2.5} concentration from ambient sources, observed fires, and simulated controlled burns.

Our GNN-based forecasting framework can help policymakers better isolate the PM_{2.5} concentration emitted from wildfires. While studies like Childs et al. (2022) and Qiu et al. (2024) rely on analyst-derived information, such as plume boundaries, to identify non-smoke days for calculating median PM_{2.5} values

per grid cell and month, our method operates without the need for expert input. This makes it suitable for rapid deployment as a fast and efficient emergency management tool. Additionally, our forecasts can aid forest managers in minimizing the $PM_{2.5}$ exposure of vulnerable populations during controlled burns and facilitate community discussions of potential locations and burn windows for prescribed fires. Thus, while several studies have used machine learning to forecast air quality (Wang et al., 2020; Li et al., 2023) as well as to estimate wildfire-specific $PM_{2.5}$ (Aguilera et al., 2023; Childs et al., 2022), this is the first paper, to the best of our knowledge, that utilizes machine learning to predict the $PM_{2.5}$ concentration at sensor locations from simulated prescribed fires. Although we focus on $PM_{2.5}$ predictions at sparse sensor locations, our method could be extended to predictions over both ZIP codes and a regular grid.

The remainder of the paper is organized as follows: the fire and meteorological data used in our analyses are detailed in Section 2. In Section 3.1, we describe the $PM_{2.5}$ -GNN model and validate its $PM_{2.5}$ predictions with baseline machine learning models. The model setup for estimating wildfire-specific $PM_{2.5}$ is outlined in Section 3.2. Sections 4.1 and 4.2 present, respectively, idealized experiments using a framework based on the $PM_{2.5}$ -GNN model to determine the optimal time to implement prescribed fires and to quantify the reduction in air quality impact due to mitigation of larger wildfires. We discuss the paper's conclusions and potential directions for future work in Section 5.

2. Dataset

Our dataset consists of $PM_{2.5}$, meteorological, and fire data at an hourly resolution over 5 years (2017–21). The $PM_{2.5}$ concentration data, at a total of 112 sparse air quality sensor locations in California shown in Figure 9, is collected from both the California Air Resources Board (CARB) as well as the Environmental Protection Agency (EPA) (California Air Resources Board, n.d.; US Environmental Protection Agency, n.d.). The MissForrest algorithm (Stekhoven and Bühlmann, 2011) was used to impute the missing $PM_{2.5}$ observations from offline sensors. The data for the seven meteorological predictors, which include u and v horizontal components of wind, total precipitation, and air temperature, are retrieved from the ERA5 Reanalysis database (Hersbach et al., 2020). We provide the full list of predictors in Table 1. Though the meteorological predictors may capture the diurnal $PM_{2.5}$ cycles and seasonal patterns, the Julian date and hour of the day are also included as predictors to provide the model with additional context.

The fire radiative power (FRP) provides information about the fire intensity. The FRP at each fire location is taken from the Visible Infrared Imaging Radiometer Suite (VIIRS) (Schroeder and Giglio, 2017) instrument on board the Suomi satellites. To assess the impact of nearby fires at the location of a

Table 1. Dataset predictors and $PM_{2.5}$ -GNN node features

Predictor name	Unit	Source
Planetary boundary layer height (PBLH)	m	ERA5 reanalysis
u-component of wind	m/s	ERA5 reanalysis
v-component of wind	m/s	ERA5 reanalysis
m Temperature	K	ERA5 reanalysis
Dewpoint temperature	K	ERA5 reanalysis
Surface pressure	Pa	ERA5 reanalysis
Total precipitation	m	ERA5 reanalysis
WIDW FRP within 25 km, 50 km, 100 km, 500 km	MW	VIIRS
Number of fires within 500 km	1	VIIRS
Julian date	1	N/A
Time of day	1	N/A

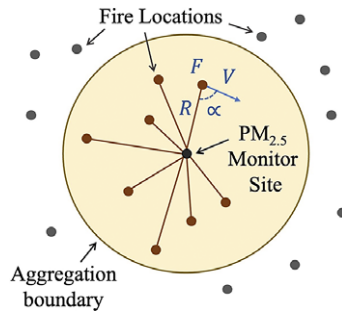


Figure 1. FRP aggregated around a given radius for each $PM_{2.5}$ monitor location using wind and distance information.

$PM_{2.5}$ monitor, we aggregate the FRP values of all active fires within radii of 25, 50, 100, and 500 km. To emphasize the fires that would likely have a more substantial downwind effect on $PM_{2.5}$ concentration, we use inverse distance weighting (IDW) and wind-based weighting in the FRP aggregation. For each $PM_{2.5}$ monitor location, aggregations are performed on radii of 25, 50, 100, and 500 km to derive the wind and inverse-distance weighted (WIDW) FRP using the process described in Figure 1 and

$$F_{WIDW} = \sum_{i=1}^n \frac{F_i |V_i| \cos(|\alpha_i|)}{4\pi R_i^2} \tag{1}$$

where n is the number of fire locations within a certain radius of the $PM_{2.5}$ monitor site, F is the FRP value at the fire location, $|V|$ is the magnitude of the wind speed at the fire location, α is the relative angle between the wind direction and the direction from the fire to the $PM_{2.5}$ monitor, and R is the distance between the fire site and $PM_{2.5}$ monitor. The number of fires within 500 km of a $PM_{2.5}$ site is also included in the dataset. The prescribed fire latitude, longitude, and duration data retrieved from the California Department of Forestry and Fire Protection (Cal Fire) (Cal Fire, n.d.) is not represented as a variable in the training dataset but instead used in Experiments 1 and 2 when simulating prescribed fires.

3. $PM_{2.5}$ forecasts

3.1. $PM_{2.5}$ -GNN model

3.1.1. Methods

We trained the spatio-temporal $PM_{2.5}$ -GNN model from Wang et al. (2020), to predict $PM_{2.5}$ concentration at an hourly temporal resolution utilizing spatially sparse observations, as illustrated in Figure 2. In Figure 2, v_i, v_j, F^{v_i} , and $F^{e_{ij}}$, respectively, refer to nodes i and j , the feature vector for node i , and the feature vector for edge e_{ij} . The $PM_{2.5}$ -GNN model consists of two components: a graph neural network (GNN) to learn the $PM_{2.5}$ spatial propagation between monitor sites and a Gated Recurrent Unit (GRU) to capture

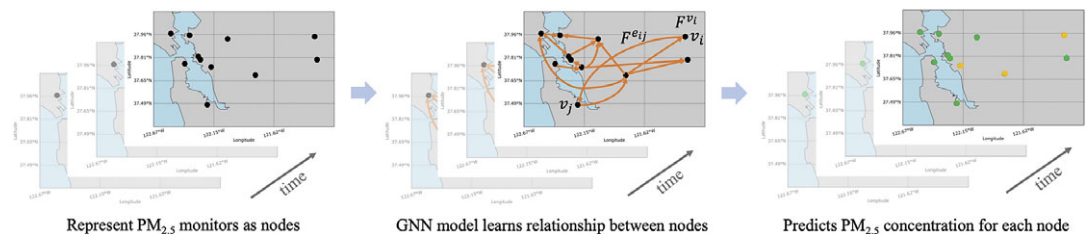


Figure 2. Graph neural network (GNN) used in our $PM_{2.5}$ forecasting model considers $PM_{2.5}$ monitors as nodes in the graph and produces node-level predictions.

Table 2. Training/validation/testing split

Training	Validation	Testing
1/1/2017–12/31/2018	1/1/2020–12/31/2020	1/1/2021–12/31/2021

the temporal diffusion process. The GNN model has a directed graph, where nodes represent the locations of PM_{2.5} monitors and edges model the PM_{2.5} movement and interactions between monitors, making it well-suited to leverage non-homogeneous data. Thus, the node features include all the meteorological and fire-related predictors, as shown in Table 1. The edge attributes consist of the wind speed of the source node, the distance between the source and sink, the wind direction of the source node, the direction from source to sink, and an advection coefficient that quantifies the degree of wind impact on the sink node by the source node. In the graph model of PM_{2.5} transport, the source node represents the site where the PM_{2.5} was initially detected at a specific time point and the sink node represents a location downwind of the initial site. Domain knowledge is incorporated in the graph representation through the choice of node and edge attributes, as well as through the choice of the graph connectivity. For instance, the graph explicitly includes wind direction information and considers geographical elevation differences. The GNN only models the transport of PM_{2.5} between two sites if they are within 300 km of one another and if the elevation difference between the sites is < 1200 m. The altitude threshold was established on the assumption that differences in elevation >1200 m between sites would hinder PM_{2.5} transport.

As shown in Table 2, for the PM_{2.5}-GNN model, 2 years are used for training and 1 year each for validation and testing. In total, we use ~1.96 million samples for training and 0.98 million samples each for validation and testing. The year 2019 is excluded during training, validation, and testing because the 2019 fire season was an outlier and was less damaging than the other years. Validating and testing the model on data from the years 2020 and 2021, respectively, would help us gain a better understanding of the model's performance during intense fires, as both 2020 and 2021 had severe wildfire seasons. Our model produces PM_{2.5} forecasts for a prediction window of 48 hours into the future based on a model initialized with PM_{2.5} observations in a historical window of 240 hours.

3.1.2. Results

For the evaluation of the PM_{2.5}-GNN's performance, we analyzed heat maps of the observed versus predicted PM_{2.5} concentrations for various future time points (1-hour, 12-hour, and 48-hour forecasts), as shown in Figure 3. The color scale represents the density of points within a bin, with bin boundaries selected on a logarithmic scale. Perfect predictions would align along the 1:1 line, while deviations from this line in specific directions indicate prediction errors. Similar heat map graphs of observed versus predicted PM_{2.5}

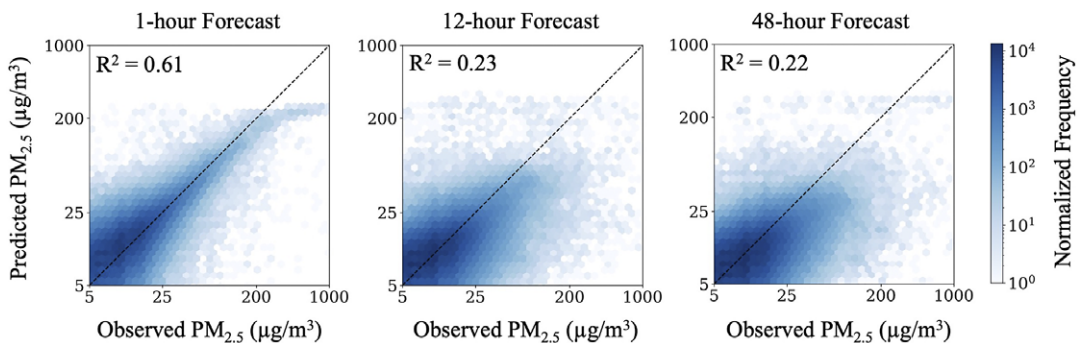


Figure 3. Heat maps illustrating the observed PM_{2.5} concentrations versus the PM_{2.5}-GNN model predictions for 1-, 12-, and 48-hour forecasts, with log-transformed axes and color scale. Also indicated is the identity line (dotted black) and R² values of the best-fit linear model.

were presented in Considine et al. (2023) (Considine et al., 2023). For our 1-hour forecasts, there is a high density of points along the identity ($y = x$) line, indicating the model’s ability to predict $PM_{2.5}$ levels close to the observed values. However, the heat maps also reveal that the model begins to underpredict more severely for observed $PM_{2.5}$ concentrations exceeding $\approx 200 \mu\text{g}/\text{m}^3$. The $PM_{2.5}$ -GNN model’s R^2 value is lower than those reported in Considine et al. (2023) and Qiu et al. (2024), which both studied $PM_{2.5}$ predictions using machine learning, because our predictions are at an hourly resolution, whereas theirs are at a daily resolution. As the $PM_{2.5}$ -GNN model predicts further into the future, the R^2 value and the density of the points along the identity line for elevated concentrations decrease. The decreasing accuracy of predictions for time steps further into the future is a challenge inherent to prediction tasks.

To further evaluate the $PM_{2.5}$ -GNN model, its performance was compared to three baseline models, the random forest (RF), long short-term memory (LSTM), and multilayer perceptron (MLP) models. The $PM_{2.5}$ -GNN model had the lowest mean absolute error (MAE) and root mean squared error (RMSE) values, as shown in Table 3. As a reference for the error, the US Air Quality Index (AQI) very unhealthy and hazardous $PM_{2.5}$ levels are $\geq 150.5 \mu\text{g}/\text{m}^3$. Heat maps of the observed versus predicted $PM_{2.5}$ concentrations for all baseline models are given in Appendix A.

Additionally, the time series results of the $PM_{2.5}$ -GNN, RF, LSTM, and MLP were graphed to analyze the results. Figure 4 displays predictions 1 hour into the future from the testing results of two example sites. These sites were selected due to observed $PM_{2.5}$ levels reaching the US AQI’s very unhealthy ($\geq 150.5 \mu\text{g}/\text{m}^3$) levels and demonstrate typical performance of the models for elevated $PM_{2.5}$. The graphs showed that the $PM_{2.5}$ -GNN was better able to predict elevated concentrations in comparison to the RF, MLP, and LSTM. Although Table 3 shows that the $PM_{2.5}$ -GNN model’s MAE and RMSE metrics surpass

Table 3. Results of the $PM_{2.5}$ -GNN, RF, LSTM, and MLP models

	$PM_{2.5}$ -GNN	RF	LSTM	MLP
MAE	5.23	5.29	5.73	6.24
RMSE	6.72	6.78	7.32	7.83

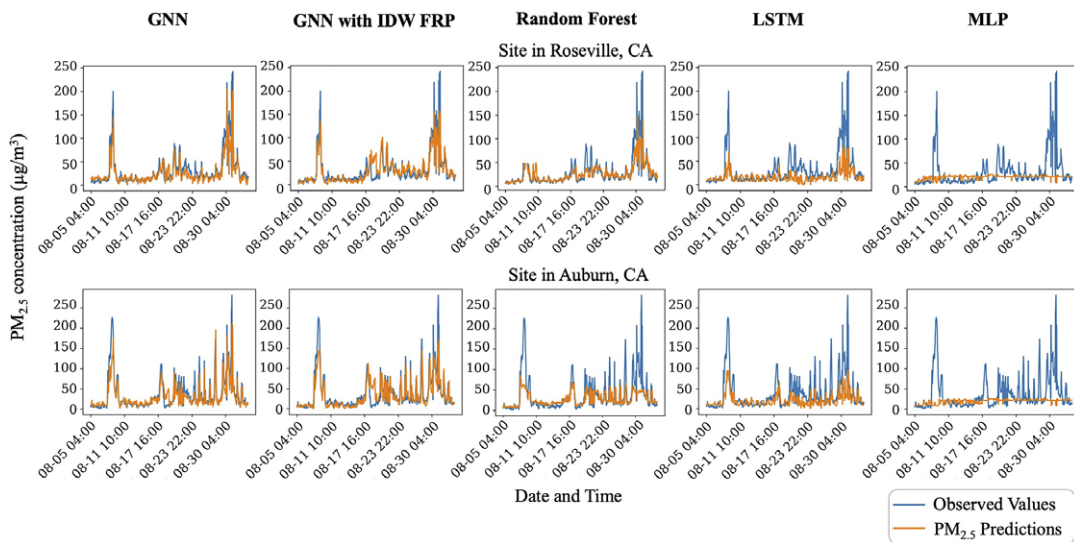


Figure 4. $PM_{2.5}$ predictions 1 hour into the future from a temporal subset of testing results for example sites. The $PM_{2.5}$ -GNN (Column 1), Random Forest, LSTM, and MLP all use the WIDW FRP predictor, while the $PM_{2.5}$ -GNN with IDW FRP (Column 2) uses the IDW FRP predictor.

those of the other three models by $<1 \mu\text{g}/\text{m}^3$, comparing Figure 3 with Figure 10 from the Appendix demonstrates that the $\text{PM}_{2.5}$ -GNN significantly outperforms the other methods in predicting elevated $\text{PM}_{2.5}$ levels. At longer time horizons, although the performance of all models declines for the 12- and 48-hour forecasts, the $\text{PM}_{2.5}$ -GNN model, with its recurrent neural network component, achieves the highest R^2 values due to both its spatial and temporal inductive biases. The LSTM ranks second, whereas the RF and MLP, lacking temporal inductive bias, exhibit a more significant decrease in prediction skills for forecasts further into the future. The small differences in MAE and RMSE can be attributed to the distribution of the observed $\text{PM}_{2.5}$ data from the testing set with a 99th percentile of only $50.70 \mu\text{g}/\text{m}^3$. However, from the graphs of all three models, it was evident that there was a tendency for the model to output a value close to the observed value at the previous time point as its prediction. This issue is also prevalent in other studies in this area and the broader machine-learning field.

The $\text{PM}_{2.5}$ -GNN’s performance was also compared to the performance of a $\text{PM}_{2.5}$ -GNN trained on a dataset with only inverse-distance weighted (IDW) FRP, not wind and inverse-distance weighted (WIDW). The $\text{PM}_{2.5}$ -GNN with WIDW FRP has slightly higher MAE and RMSE, as seen in Table 4. However, the graphs of the results showed that the $\text{PM}_{2.5}$ -GNN with WIDW FRP was better able to predict elevated $\text{PM}_{2.5}$ values. This is significant, as current prediction models under-predict fire-influenced $\text{PM}_{2.5}$ concentration (Reid et al., 2021). The reason for the slightly higher MAE and RMSE for the $\text{PM}_{2.5}$ -GNN with WIDW FRP seems to be that the model tends to overpredict low concentration values.

3.2. Fire-specific $\text{PM}_{2.5}$ forecasts

3.2.1. Methods

For the task of distinguishing the pollution emitted from wildfires, a two-step process is used. A $\text{PM}_{2.5}$ -GNN model is first trained to predict the total $\text{PM}_{2.5}$ concentration and a second $\text{PM}_{2.5}$ -GNN is trained to predict the $\text{PM}_{2.5}$ emitted from ambient sources. The predictions from the ambient-focused $\text{PM}_{2.5}$ -GNN are subtracted from the forecasts from the first $\text{PM}_{2.5}$ -GNN to produce an estimate of the fire-specific $\text{PM}_{2.5}$. This is similar to the methodology of Aguilera et al. (2023), which also subtracts the predicted non-smoke $\text{PM}_{2.5}$ concentrations from the net $\text{PM}_{2.5}$ predictions to estimate the fire-specific $\text{PM}_{2.5}$. Our process is outlined in Figure 5.

The detailed methodology for the first $\text{PM}_{2.5}$ -GNN model, which is trained to predict the total $\text{PM}_{2.5}$ concentration, is outlined in Section 3.1.1. This $\text{PM}_{2.5}$ -GNN model is trained on all meteorological and fire-related predictors. The second $\text{PM}_{2.5}$ -GNN model, on the other hand, focuses only on predicting the ambient $\text{PM}_{2.5}$ and is thus trained only on the meteorological data. For this model, fire predictors are excluded during training to prevent the model from learning the effect of fires on $\text{PM}_{2.5}$ concentration. All

Table 4. Results of the $\text{PM}_{2.5}$ -GNN with WIDW FRP and IDW FRP

	$\text{PM}_{2.5}$ -GNN with WIDW FRP	$\text{PM}_{2.5}$ -GNN with IDW FRP
MAE	5.23	5.11
RMSE	6.72	6.62

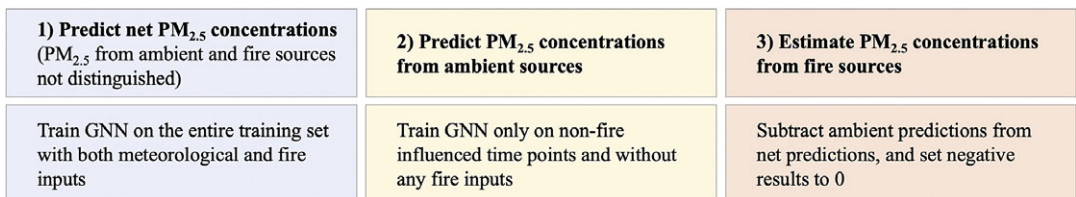


Figure 5. Conceptual diagram of the methodology for distinguishing fire-specific and ambient $\text{PM}_{2.5}$ concentrations.

the data during fire events are also excluded because including time points during fires would allow the model to learn the influence of fires from meteorological predictors like temperature. Therefore, for the second PM_{2.5}-GNN model to only predict ambient pollution, the model should only be trained on meteorological data and non-fire-influenced time points. However, selecting time points without fire events is challenging, as PM_{2.5} particles emitted by a fire can persist in the air for weeks and secondary aerosol formation resulting from fires can also impact PM_{2.5} concentrations for several days following a fire (World Health Organization, 2006). Our analysis revealed that relatively high FRP values continued to affect the PM_{2.5} concentration for over a week. Thus, during training, all time points with WIDW FRP 500 km values >0.15 within a 12-day window (240 hours before and 48 hours after the time point) are excluded, resulting in the model being trained on 21% of the data from the original training set. However, the training set remains balanced. When analyzing the percentage distribution of the training set by month, the mean percentage is 8.33% and the standard deviation is 0.91%. The second PM_{2.5}-GNN is then validated and tested at all time points to obtain ambient PM_{2.5} predictions, which is necessary to quantify the fire-specific PM_{2.5}.

3.2.2. Results

The PM_{2.5}-GNN model trained to predict only ambient PM_{2.5} had an MAE of 5.66 $\mu\text{g}/\text{m}^3$ and RMSE of 6.85 $\mu\text{g}/\text{m}^3$ for its predictions at time points without fire influence (i.e., times where no WIDW FRP within 500 km values exceeds 0.15 within a 12-day window—240 hours before and 48 hours after the time point). Analysis of the ambient-focused model indicates a tendency to underpredict at time points classified as non-fire-influenced. This underprediction is likely due to limitations in the criteria used to define fire influence, as many elevated, unhealthy PM_{2.5} observations are still present during periods categorized as not fire-influenced. Figure 6 visually distinguishes the ambient and fire-specific PM_{2.5} forecasts. Since there is no ground truth value to validate the attribution of PM_{2.5} concentration produced by ambient and wildfire sources (Aguilera et al., 2023), there is no metric to describe the accuracy of our fire-specific predictions. However, as shown in Figure 6, the result aligns with expectations, as there are significant levels of PM_{2.5} attributed to fires during elevated PM_{2.5} concentrations.

4. Prescribed fire simulations

A major contribution of this work is the novel framework integrating simulations of prescribed fires with GNN-based predictions of the resulting PM_{2.5} concentrations. The prescribed fires are simulated by transposing historical prescribed fires to target times. The Cal Fire (Cal Fire, n.d.) latitude, longitude, and duration data for the prescribed burns are matched with the VIIRS FRP data. The transposed prescribed fire FRP information is combined with the observed meteorological data at the target times and input into the PM_{2.5}-GNN model, which produces the PM_{2.5} predictions.

Using this framework, we perform two model experiments. Experiment 1 demonstrates how the PM_{2.5}-GNN forecasting model can determine the optimal time to implement prescribed fires and focuses

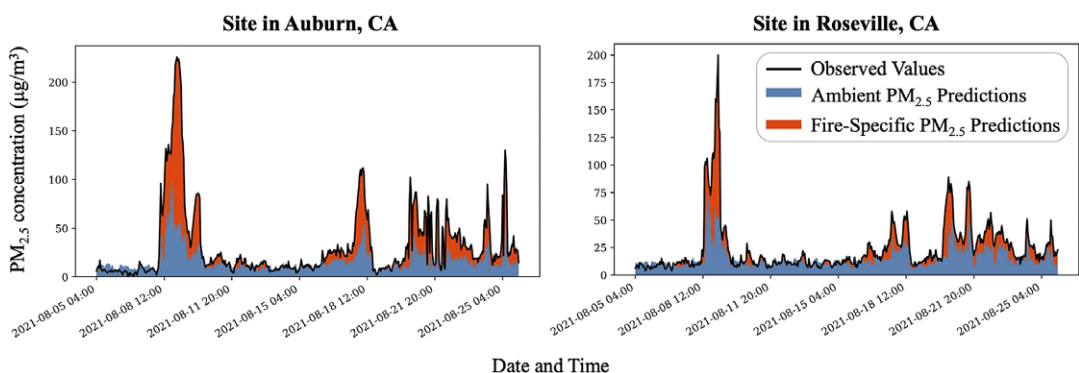


Figure 6. Ambient and fire-specific PM_{2.5} predictions 1 hour into the future from a temporal subset of testing results for example sites.

on the short-term pollution effect of prescribed fires. Experiment 2, on the other hand, focuses on quantifying the pollution impact of prescribed fires across months. In the rest of the section, we discuss each experiment in more detail:

4.1. Experiment 1: minimizing prescribed fire PM_{2.5} impact

4.1.1. Methods

To determine the optimal time to implement prescribed fires, we consider the immediate effect of prescribed fires. In effect, this experiment simulates historically prescribed fires and predicts the resulting PM_{2.5} pollution under different meteorological and fire conditions. That is, we transpose the FRP values from actual prescribed fire events to target time points and add them to the observed FRP values at those points. As these FRP values are combined, they are weighted using inverse distance and wind (both direction and magnitude), as outlined in Section 2.

In this experiment, we transpose a window of time containing prescribed fires (1/3/21–1/15/21) to target times throughout the year 2021 at 24-hour time steps to simulate the air quality impacts of controlled burns, as shown in Figure 7. This window contains 10 prescribed fires with burned areas above 100 acres.

4.1.2. Results

As shown in Table 5, the month of August was the least optimal time to implement prescribed fires since it resulted in the most significant PM_{2.5} concentration. As August is during the peak wildfire season, implementing prescribed fires would only exacerbate the already hazardous air quality. August’s mean PM_{2.5} was 29.61% greater than the average mean of other months and August’s maximum was 44.27% greater than the average maximum of other months. On the other hand, March, which had the lowest mean value, was found to be the most optimal month to implement prescribed fires. The mean and maximum values were calculated by averaging the mean and maximum PM_{2.5} predictions of the locations whose PM_{2.5} observations were ≥ 50 µg/m³ during the window 1/3/21–1/15/21. As the PM_{2.5} observations at

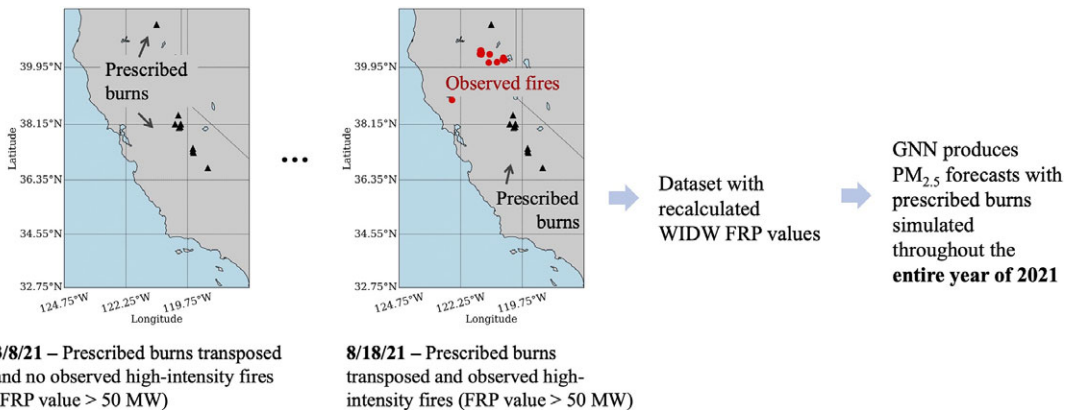


Figure 7. Schematic illustration of the methodology used in Experiment 1 to generate PM_{2.5} predictions based on simulated prescribed burns and observed fire events throughout 2021.

Table 5. Comparing the results of PM_{2.5} predictions based on simulated prescribed fires in Experiment 1 (see text for more details) for each month of 2021

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Mean (µg/m ³)	15.62	15.64	14.69	15.18	14.76	15.92	18.44	21.73	18.55	16.95	19.94	18.73
Max (µg/m ³)	36.49	39.06	39.10	38.76	40.85	42.04	47.25	60.13	43.44	45.61	40.54	45.32
Std Dev (µg/m ³)	7.53	7.13	6.93	6.08	5.62	5.58	7.90	19.30	7.42	7.64	8.57	8.98

those locations were elevated during 1/3/21–1/15/21, they were likely influenced by the fire events transposed across the year 2021.

4.2. Experiment 2: quantifying prescribed fire $PM_{2.5}$ trade-off

4.2.1. Methods

This idealized experiment aims to quantify the pollution trade-off of implementing larger prescribed fires to mitigate emissions from the Caldor Fire, one of the largest Californian wildfires in 2021. Specifically, we use the $PM_{2.5}$ -GNN model to simulate the counterfactual scenario of performing scaled-up controlled burns in 2021 at the location of the Caldor Fire as shown in Figure 8. We employ two simulation techniques: one corresponding to the immediate air quality impact of more intense prescribed fires and the other to the longer-term effect of prescribed burning, related to mitigating the emissions from a larger wildfire such as the Caldor Fire.

For the first case, we simulated the effect of three historical prescribed fires, which were all within 20 km of the 2021 Caldor Fire, were active from 3/21 to 5/31 in 2018, 2019, and 2020, respectively, and burned around 6300 acres each. Since the Caldor Fire burned around 221,835 acres (Cal Fire, n.d.), we assume that preventing a fire of that scale would require a larger controlled burn. Thus, when creating the fire-related input predictors, the FRP values from the prescribed fires are artificially increased by a factor of 100 and transposed together to 2021, thereby simulating large prescribed fires from 3/21/21 to 5/31/21. As described in Section 4.1, the prescribed fires are transposed by combining the FRP values of the prescribed fires with the observed FRP values from other fires at the target time point, followed by aggregating those values using inverse distance weighting and wind information.

In the latter case, we simulate the effect of controlled burns later in the year by excluding all FRP values within 25 km of the Caldor Fire between 8/14/21 and 10/21/21, assuming that a prescribed fire implemented earlier in the year (or even during the previous one to two fire seasons) could effectively mitigate a large fire in the same location a few months later. Since we use $PM_{2.5}$ observations from the previous 10 days to initialize each 48-hour window of $PM_{2.5}$ predictions, using the $PM_{2.5}$ data from 2021 would implicitly include the Caldor Fire’s influence. To avoid this bias in our counterfactual scenario, we used as input the average of $PM_{2.5}$ observations on the same date and hour from 8/14 to 10/2 in 2017, 2018, 2019, and 2020. Thus, the inputs will be fire-influenced but do not include the impact of the Caldor Fire. Limitations of our methodology are discussed in Section 4.2.3.

The $PM_{2.5}$ predictions from this experiment’s counterfactual scenario are compared to baseline predictions derived using observed meteorological and fire inputs from 2021 without any prescribed fires around the Caldor Fire locations.

4.2.2. Results

The results support that although prescribed fires slightly increase $PM_{2.5}$ in the short term, they reduce future $PM_{2.5}$ resulting from wildfires. As shown in Table 6, the simulated prescribed burns led the mean of

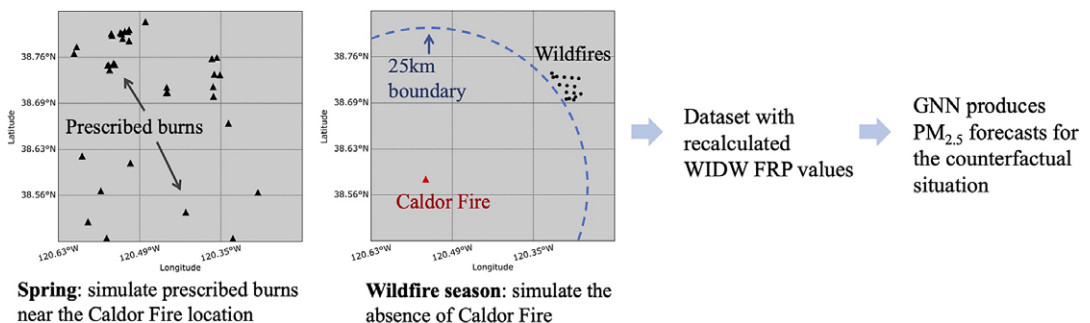


Figure 8. Schematic illustration of the methodology used in Experiment 2 for simulating prescribed burns during spring and the absence of the Caldor Fire during the wildfire season.

Table 6. Comparing the predicted $PM_{2.5}$ effect of simulated prescribed burns in Experiment 2 (see text for more details) with baseline $PM_{2.5}$ predictions

	3/21/21–5/31/21		8/14/21–10/21/21	
	Simulated prescribed burn	Without prescribed burn (Baseline)	Removed Caldor fire	With Caldor fire (baseline)
Mean ($\mu\text{g}/\text{m}^3$)	6.83	6.52	12.49	16.24
Max ($\mu\text{g}/\text{m}^3$)	55.78	54.12	81.00	177.32

the $PM_{2.5}$ predictions to be increased by an average of $0.31 \mu\text{g}/\text{m}^3$ and the maximum $PM_{2.5}$ prediction to be increased by 3.07%. The mean and maximum values were calculated by averaging the mean and maximum $PM_{2.5}$ predictions of the 13 $PM_{2.5}$ monitor locations within 100 km of the Caldor Fire. Table 6 also quantifies that the maximum of the predictions with the Caldor Fire’s influence removed was 54.32% lower than the maximum of the baseline predictions. Thus, the magnitude of the immediate $PM_{2.5}$ increase from the prescribed fire was significantly lower than the magnitude of the $PM_{2.5}$ decrease experienced during the fire season. Furthermore, excluding the influence of the Caldor Fire reduced the number of days with an unhealthy daily average $PM_{2.5}$ concentration from a mean of 3.54 days to 0.80 days. The reduction in $PM_{2.5}$ pollution after excluding the Caldor Fire influence is illustrated in Figure 9, where the $PM_{2.5}$ monitoring sites are color-coded depending on the $PM_{2.5}$ pollution’s US AQI level.

4.2.3. Discussion

A limitation of our prescribed fire simulation methodology is the absence of ground truth $PM_{2.5}$ data for counterfactual or simulated prescribed fire scenarios. This creates challenges in both defining $PM_{2.5}$ inputs for the model and evaluating the accuracy of the simulation. Although no direct solution currently exists, our results can be compared to and validated against CTMs.

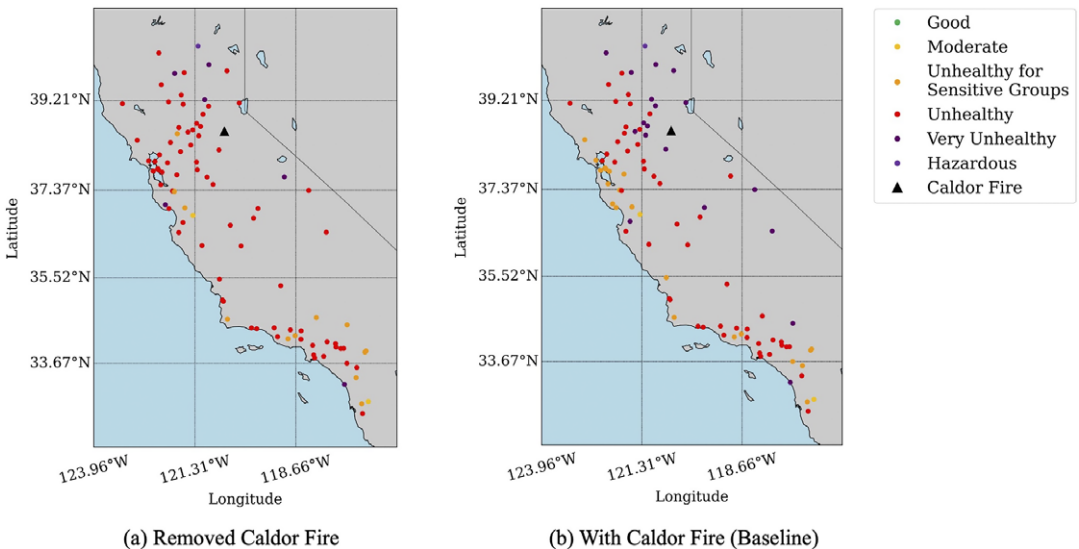


Figure 9. Maximum $PM_{2.5}$ predictions per site from 8/14/21–12/31/21 under conditions (a) with prescribed burns at the Caldor Fire location during the spring and without Caldor Fire during the wildfire season and (b) without prescribed burns at the Caldor Fire location and with the Caldor Fire during the wildfire season.

5. Conclusion and future work

This work generated hourly PM_{2.5} predictions for California using a PM_{2.5}-GNN model and demonstrated that due to its spatio-temporal modeling capabilities, the PM_{2.5}-GNN outperformed other machine learning models like the Random Forest, LSTM, and MLP. The temporally fine-grained, hourly PM_{2.5} predictions can help people better plan their outdoor activities to stay healthy. Additionally, our work focused on exploring two novel applications of the PM_{2.5}-GNN model: (1) estimating the fire-specific contribution in PM_{2.5} forecasts and (2) predicting the PM_{2.5} for simulated fire events. Our work demonstrates the versatility of the PM_{2.5}-GNN: the model can be applied to real-time prediction tasks, the GNN-based simulations from Experiment 1 can assess optimal windows for future prescribed fires, and the GNN-based simulations from Experiment 2 can be used for retrospective analysis.

Following previous machine learning-based studies (Aguilera et al., 2023; Childs et al., 2022), we apply GNNs to the task of distinguishing the PM_{2.5} pollutant concentration emitted from fires versus ambient sources. This is significant as machine learning has higher computational efficiency than chemical transport models (CTMs) and the PM_{2.5}-GNN model has been shown to outperform other machine learning models for pollution forecasting. Our hourly PM_{2.5}-GNN predictions may also be a promising method to improve the skill of 2-day air quality forecasts. For the 48-hour prediction window that we considered here, the PM_{2.5}-GNN model was shown to outperform other machine learning-based air quality forecasts due to its spatio-temporal inductive bias. Although we focused here on predictions at 112 sparse air quality sensor locations where there are historical PM_{2.5} measurements available, the PM_{2.5}-GNN model has the ability to make predictions at sites that were not included in the original training dataset because of the spatiotemporal inductive bias of the PM_{2.5}-GNN model.

To the best of our knowledge, this is the first research paper to apply machine learning for simulating the PM_{2.5} impact of prescribed fires, which is significant given the limitations of CTMs. A major contribution of this work is the prescribed fire simulation framework, which integrates prescribed fire simulations with GNN-based PM_{2.5} predictions. Future work will focus on improving the air quality simulations by incorporating fire risk models (Buch et al., 2023; Langlois et al., 2024) and physics-based smoke plume dynamics (Mallia and Kochanski, 2023) in the PM_{2.5}-GNN model. Another promising direction is developing an auxiliary GNN model that incorporates the impact of individual fires and obviates the need to aggregate FRP values in the vicinity of each PM_{2.5} monitor, thereby making it easier to remove or add specific fire influences. Our framework provides land managers and the fire service with a useful tool to minimize the PM_{2.5} exposure of vulnerable populations, while also informing local communities of potential air quality impacts and beneficial trade-offs when implementing controlled burns.

Open peer review. To view the open peer review materials for this article, please visit <http://doi.org/10.1017/eds.2025.4>.

Author contribution. Conceptualization: Kyleen Liao; Jatan Buch; Kara Lamb; Pierre Gentine. Methodology: Kyleen Liao; Jatan Buch; Kara Lamb; Pierre Gentine. Data curation: Kyleen Liao. Data visualization: Kyleen Liao. Writing original draft: Kyleen Liao. All authors approved the final submitted draft.

Competing interest. None.

Data availability statement. Data and code for reproducing all the results can be found on GitHub: https://github.com/kyleenliao/PM2.5_Forecasting_GNN/tree/main

Funding statement. We acknowledge funding from NSF through the Learning the Earth with Artificial Intelligence and Physics (LEAP) Science and Technology Center (STC) (Award #2019625). Jatan Buch, Kara Lamb, and Pierre Gentine were also supported in part by the Zegar Family Foundation.

Ethical standard. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

References

- Aguilera R, Corringham T, Gershunov A et al. (2021) Wildfire smoke impacts respiratory health more than fine particles from other sources: observational evidence from Southern California. *Nature Communications* 12, 1493 <https://doi.org/10.1038/s41467-021-21708-0>
- Aguilera R, Luo N, Basu R et al. (2023) A novel ensemble-based statistical approach to estimate daily wildfire-specific PM_{2.5} in California (2006–2020). *Environment International* 171, 107719. <https://doi.org/10.1016/j.envint.2022.107719>

- Askariyeh MH, Khreis H and Vallamsundar S** (2020) Air pollution monitoring and modeling. In Khreis H (ed.), *Traffic-Related Air Pollution*. Elsevier, pp. 111–135. <https://doi.org/10.1016/B978-0-12-818122-5.00005-3>
- Buch J, Williams AP, Juang CS et al.** (2023) SMLFire1.0: a stochastic machine learning (SML) model for wildfire activity in the western United States. *Geoscientific Model Development* 16, 12. <https://doi.org/10.5194/gmd-16-3407-2023>
- Burke M, Childs ML, de la Cuesta B et al.** (2023) The contribution of wildfire to PM_{2.5} trends in the USA. *Nature* 622, 761–766. <https://doi.org/10.1038/s41586-023-06522-6>
- Burke M, Driscoll A, Heft-Neal S et al.** (2021) The changing risk and burden of wildfire in the United States. *Proceedings of the National Academy of Sciences* 118(2), e2011048118. <https://doi.org/10.1073/pnas.2011048118>
- Byun D and Schere KL** (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, 51–77. <https://doi.org/10.1115/1.2128636>
- California Air Resources Board.** Air Quality and Meteorological Information System [internet database]. Available at <https://www.arb.ca.gov/aqmis2/aqmis2.php>
- Childs ML, Li J, Wen J et al.** (2022) Daily local-level estimates of ambient wildfire smoke PM_{2.5} for the contiguous US. *Environmental Science & Technology* 56, 19. <https://doi.org/10.1021/acs.est.2c02934>
- Considine EM, Hao J, de Souza P et al.** (2023) Evaluation of model-based PM_{2.5} estimates for exposure assessment during wildfire smoke episodes in the western U.S. *Environmental Science and Technology* 57(5), 2031–2041. <https://doi.org/10.1021/acs.est.2c06288>
- Cal Fire.** (n.d.) Fire Perimeters [Internet Database]. Available at <https://www.fire.ca.gov/what-we-do/fire-resource-assessment-program/fire-perimeters>
- Cal Fire.** (n.d.) Caldor Fire. Cal FIRE. Available at <https://www.fire.ca.gov/incidents/2021/8/14/caldor-fire/>
- Hersbach H, Bell B, Berrisford P, et al.** (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* 146, 1999–2049. <https://doi.org/10.1002/qj.3803>
- Jin L, Permar W, Selimovic V et al.** (2023) Constraining emissions of volatile organic compounds from western US wildfires with WE-CAN and FIREX-AQ Airborne Observations. *Atmospheric Chemistry and Physics* 23(10), 5969–5991. <https://doi.org/10.5194/acp-23-5969-2023>
- Kelp MM, Carroll MC, Liu T et al.** (2023) Prescribed burns as a tool to mitigate future wildfire smoke exposure: lessons for states and Rural Environmental Justice Communities. *Earth's Future* 11(6), e2022EF003468. <https://doi.org/10.1029/2022EF003468>
- Langlois GP, Buch J, and Darbon J** (2024) Efficient first-order algorithms for large-scale, non-smooth maximum entropy models with application to wildfire science. *Entropy* 26(8), 691. <https://doi.org/10.3390/e26080691>
- Li L, Wang J, Franklin M et al.** (2023) Improving air quality assessment using physics-inspired deep graph learning. *npj Climate and Atmospheric Science* 6, 152. <https://doi.org/10.1038/s41612-023-00475-3>
- Mallia DV and Kochanski AK** (2023) A review of modeling approaches used to simulate smoke transport and dispersion. In *Landscapes Fire, Smoke, and Health*, American Geophysical Union (AGU), chapter 8. <https://doi.org/10.1002/9781119757030.ch8>
- McCaffrey SM** (2006) Prescribed fire: What influences public approval. In Dickinson MB (ed.), *Fire in Eastern Oak Forests: Delivering Science to Land Managers, Proceedings of a Conference*, 15–17 November 2005. Columbus, OH: Gen. Tech. Rep. NRS-P-1. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station, pp. 192–198.
- Qiu M, Kelp M, Heft-Neal S et al.** (2024) Evaluating estimation methods for wildfire smoke and their implications for assessing health effects. Available at <https://eartharxiv.org/repository/view/7250/>
- Reid CE, Considine EM, Maestas MM et al.** (2021) Daily PM_{2.5} concentration estimates by county, ZIP code, and census tract in 11 Western States 2008–2018. *Science Data* 8, 112. <https://doi.org/10.1038/s41597-021-00891-1>
- Rybarczyk Y and Zalakeviciute R** (2018) Machine learning approaches for outdoor air quality modelling: a systematic review. *Applied Sciences* 8(12), 2570. <https://doi.org/10.3390/app8122570>
- Schollaert CL, Marlier ME, Marshall JD et al.** (2024) Exposure to smoke from wildfire, prescribed, and agricultural burns among at-risk populations across Washington, Oregon, and California. *GeoHealth* 8(4). <https://doi.org/10.1029/2023gh000961>
- Schroeder W and Giglio L** (2017) *VIIRS/NPP Thermal Anomalies/Fire 6-Min L2 Swath 750m V001* [Data set]. NASA EOSDIS Land Processes Distributed Active Archive Center. <https://doi.org/10.5067/VIIIRS/VNP14.001>
- Stekhoven DJ and Bühlmann P** (2011) Missforest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 28(1), 112–118. <https://doi.org/10.1093/bioinformatics/btr597>
- US Environmental Protection Agency.** (n.d.) Air Quality System Data Mart [internet database]. Available at <https://www.epa.gov/outdoor-air-quality-data>
- Wang S, Li Y, Zhang J et al.** (2020) PM_{2.5}-GNN: a domain knowledge enhanced graph neural network for PM_{2.5} forecasting. In *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*. <https://doi.org/10.1145/3397536.3422208>
- Williams AP, Abatzoglou JT, Gershunov A et al.** (2019) Observed impacts of anthropogenic climate change on wildfire in California. *Earth's Future* 7(8), 892–910. <https://doi.org/10.1029/2019ef001210>
- World Health Organization** (2006) Regional Office for Europe & Joint WHO/Convention Task Force on the Health Aspects of Air Pollution. *Health Risks of Particulate Matter from Long-Range Transboundary Air Pollution*. Copenhagen: WHO Regional Office for Europe. Available at <https://apps.who.int/iris/handle/10665/107691>

World Health Organization. (2022) *Ambient (Outdoor) Air Pollutidon*. Available at [https://www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health)

Zaini N, Ean LW, Ahmed AN et al. (2022) PM_{2.5} forecasting for an urban area based on deep learning and decomposition method. *Scientific Reports* 12(1). <https://doi.org/10.1038/s41598-022-21769-1>

A. Appendix. Comparing the heat maps of the baseline models in Figure 10 with those of the PM_{2.5}-GNN in Figure 3 illustrates that the PM_{2.5}-GNN performs best as indicated by the high R² values and density of points along the identity (y = x) line.

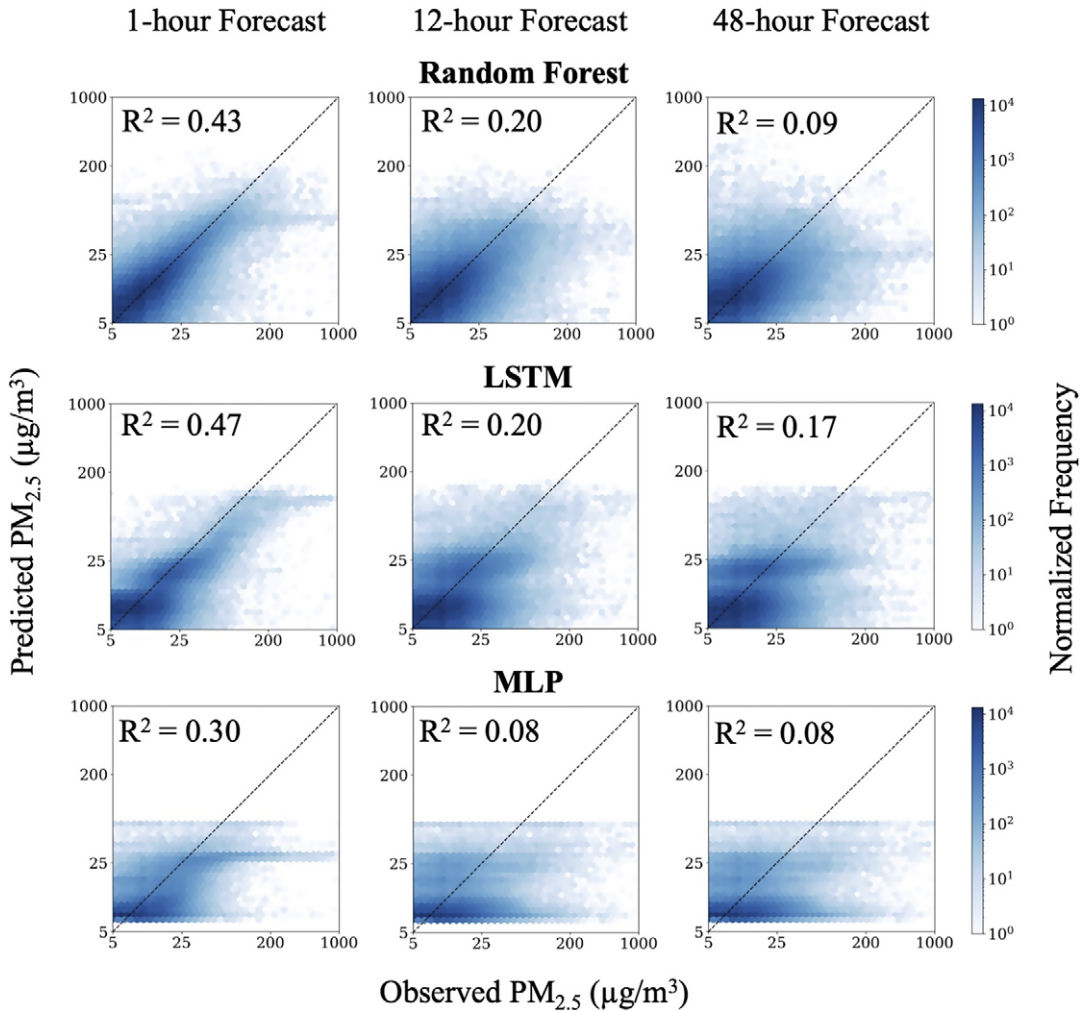


Figure 10. Heat maps illustrating the observed versus predicted PM_{2.5} concentrations for 1-hour; 12-hour; and 48-hour forecasts, with log-transformed axes and color scale, for the Random Forest, LSTM, and MLP models. Also indicated are the identity line (dotted black) and R² values of the best-fit linear model.