Multi-Pollutant Ground-level Air Pollution Prediction through Deep MeteoGCN-ConvLSTM

Pratyush Muthukumar*, Shaurya Pathak*, Kabir Nagrecha*, Hirad Hosseini*, Dawn Comer**, Navid Amini*
Jeanne Holm**, and Mohammad Pourhomayoun *
* Department of Computer Science, California State University Los Angeles, Los Angeles CA,
[pmuthuk2, knagrec2, namini, mpourho]@calstatela.edu,
** City of Los Angeles, Los Angeles CA, [dawn.comer, jeanne.holm]@lacity.org

CSCI-RTAI '22 Full Research Paper

***Abstract*—Air pollution is the fourth-largest threat to human health. The harmful effects of air pollutants have costed the global economy nearly $3 trillion. It is imperative that a solution for mitigating the harmful effects of the most pervasive ground-level air pollutants – Carbon Monoxide (CO), Nitric Oxide (NO), Nitrogen Dioxide (NO_2), Ozone (O_3), and particulate matter 2.5 ($PM2.5$) – is implemented, especially in urban areas. Recent advances in deep learning such as the Convolutional Long Short Term Memory (ConvLSTM) architecture are capable of learning complex spatiotemporal patterns with multisource data. We propose a novel sequential encoder-decoder ConvLSTM architecture capable of predicting hourly CO, NO, NO_2, O_3, and $PM2.5$ spatially continuously over Los Angeles. Our model utilizes multisource satellite imagery collected from the ESA Tropospheric Monitoring Instrument (TROPOMI), remote-sensing data collected by the NASA Moderate Resolution Imaging Spectroradiometer (MODIS) instrument onboard the NASA Terra+Aqua satellites, and site-monitoring sensor observations of atmospheric and ground-level air pollution, meteorological data, and wildfire data. Our results show that our MeteoGCN-ConvLSTM model is competitive with state-of-the-art approaches across all predicted air pollutants. Moreover, our results show the versatility of our model when provided with solely atmospheric satellite imagery and remote-sensing data as input.**

***Index Terms*—spatiotemporal prediction, air pollution prediction, graph convolutional network, convolutional long short term memory, ground-level air pollution data, site monitoring stations**

I. Introduction

The harmful effects of air pollution on mankind are very well documented. According to Ritchie and Roser (2017), 11.65% of deaths or at least 1 in 10 people die from air pollution related causes. Air pollution can generally be categorized into traffic-related air pollution, ozone, noxious gases, and particulate matter. The most detrimental air pollutants to human health include particulate matters under 2.5 micrometers in diameter ($PM2.5$), carbon monoxide (CO), nitric oxides (NO), nitrogen dioxide (NO_2), and Ozone (O_3).

In order to tackle the issue of air pollution, we must understand the sources, similarities, and effects amongst various types of air pollution. This paper delves into our understanding of air pollutants and our ability to predict them through deep learning. In the last decade, deep learning has revolutionized various fields and has proven to be particularly valuable for air pollution prediction (Bellinger et al., 2017). When developing these deep learning models, we must consider both spatial and temporal features due to the spatiotemporal nature of air pollution. Concentrations of pollutants are are strongly correlated to pollutants in neighboring areas (spatial correlation) and earlier or later concentrations (temporal correlation). Many methods applying deep learning for air pollution prediction focus on either capturing spatial or temporal patterns in data, but considerably more complexity and development is requisite for a model capable of capturing both (Abrahamsen et al., 2018; Grover et al., 2015; Narejo and Pasero, 2017; Weyn et al., 2020).

To learn spatiotemporal patterns in multisource remote-sensing data, ground-level observations, and satellite imagery, we apply a two-stage model called MeteoGCN-ConvLSTM. The first stage applies a Graph Convolutional Network (GCN) for learning high-level patterns within complex ground-level meteorological features. GCNs are a powerful deep learning architecture which applies convolution to graph structures, an abstraction well suited for meteorological and air pollution data collected at sparse site monitoring stations. he goal of the Graph Convolutional Network is to learn the feature embeddings and patterns of nodes and edges in a graph. The GCN learns the features of an input graph $G(V, E)$ typically expressed with an adjacency matrix A as well as a feature vector x_i for every node i in the graph expressed in a matrix of size $V \times D$ where V is the number of vertices in the graph and D is the number of input features for each vertex. The output of the GCN is an $V \times F$ matrix where F is the number of output features for each vertex. We can then construct a deep neural network with an initial layer embedding of $h_v^0 = x_i$ to perform convolution neighborhoods of nodes, similar to a Convolutional Neural Network (CNN). Then, the k-th layer of the neural network's embedding on vertices h_v^k is

$$h_v^k = \sigma\left(W_k \sum_{u \in N(v) \cup v} \frac{h_v^{k-1}}{\sqrt{|N(u)||N(v)|}}\right), \forall k > 0,$$

where σ is some non-linear activation function, h_v^{k-1} is the previous layer embedding of v, W_k is a transformation matrix for self and neighbor embeddings, and $\sum_{u \in N(v)} \frac{h_u^{k-1}}{|N(v)|}$ is the average of a neighbor's previous layer embeddings. The

979-8-3503-2028-2/22/$31.00 ©2022 IEEE
DOI 10.1109/CSCI58124.2022.00009

neural network can be trained efficiently through sparse batch operations on a layer wise propagation rule

$$H^{(k+1)} = \sigma(D^{-\frac{1}{2}}\tilde{A}D^{-\frac{1}{2}}H^{(k)}W_k),$$

where I is the identity matrix, $\tilde{A} = A + I$, and D is the diagonal node degree matrix defined as $D_{ii} = \sum_j A_{i,j}$ (Kipf and Welling, 2016). In this way, the GCN can train a neural network to output a graph with feature vectors for each node in the graph.

The second stage of our model applies the Convolutional Long Short-Term Memory (ConvLSTM) architecture to sets of time-indexed sequences of gridded data. The ConvLSTM model is a variant of the traditional Long Short-Term Memory (LSTM) model capable of performing convolution within the cells of the LSTM to allow for multidimensional video-like inputs and outputs. This is done by replacing the Hadamard products used to define the key equations for the FC-LSTM with the convolution operation. The key equations for the ConvLSTM are

$$\begin{aligned}
i_t &= \sigma(W_i x_t + W_i h_{t-1} + W_i * c_{t-1} + b_i) \\
f_t &= \sigma(W_f x_t + W_f h_{t-1} + W_f * c_{t-1} + b_f) \\
c_t &= f_t * c_{t-1} + i_t * \tanh(W_x x_t + W_h h_{t-1} + b_c) \\
o_t &= \sigma(W_x x_t + W_h x_{h-1} + W_c * c_t + b_o) \\
h_t &= o_t * \tanh(c_t),
\end{aligned}$$

where $*$ denotes the convolution operation (Shi et al., 2015).

Prior works on applying the ConvLSTM structure for air pollution prediction utilized air pollutant and meteorological information for prediction of a single air pollutant, typically PM2.5 or NO_2 (Muthukumar et al., 2020a,b). Further work has looked into applying wildfire and smoke data as input features for prediction (Muthukumar et al., 2021a,b, 2022). However, current research in this field has not shown an effective architecture for general-purpose multi pollutant prediction. Moreover, many prior models apply solely ground-level air pollution information or solely atmospheric data, but seldom do we see a model utilizing both (Cocom et al., 2020; Muthukumar et al., 2020c; Nagrecha et al., 2020). It is also not clear how effective satellite imagery and remote-sensing data is compared to ground-level observations for pollution forecasting with deep learning.

II. Methods

In this section, we describe our methodology for developing a two-stage prediction model for forecasting spatiotemporal carbon monoxide, nitric oxide, nitrogen dioxide, ozone, and PM2.5. One key consideration when developing our architecture is the variety of formats in our input data sources. The goal of our predictive algorithm is to perform spatiotemporal forecasting on time-indexed series of images. For satellite imagery and remote-sensing data, the format of input samples fits this constraint. However, for sparse ground-based site monitoring stations, we must apply additional preprocessing to generate time-series input images for the ConvLSTM architecture.

We make the following contributions to the field of deep learning for spatiotemporal air pollution prediction: (1) we develop a versatile two-stage MeteoGCN-ConvLSTM deep learning architecture effective for multi pollutant prediction, (2) we utilize a unique combination of remote-sensing, satellite imagery, and ground-level pollution, wildfire and meteorological data, and (3) we investigate the significance of ground-level site monitoring data in spatiotemporal air pollution forecasting.

A. Data

We collect input data from six major sources: ground-level pollution data, remote-sensing atmospheric pollution data, atmospheric pollution satellite imagery, ground-level meteorological data, remote-sensing meteorological data, and remote-sensing wildfire data. For all data, we define a bounding box geographical region within the bounds of $-118.75°W$ to $-117.5°W$ and $33.5° - 34.5°$. For all data sources, we collect hourly historical data from January 1 2019 to September 1 2022, corresponding to 1340 days worth of data or 32160 hourly timesteps.

a) Ground-Level Pollution Data: The goal of our model is to predict hourly ground-level air pollution over Los Angeles in the future. We utilize ground-level pollution data collected from site monitoring stations as both input to our deep learning model and as ground truth information for model evaluation purposes. In our architecture, we use historical ground-level pollution data as an input feature to predict spatially continuous air pollution and evaluate against held out samples of ground level pollution data in the future. For all experiments, we collect hourly ground-level pollution data of Carbon Monoxide (CO), Nitric Oxide (NO), Nitrogen Dioxide (NO_2), Ozone (O_3), and $PM2.5$. Ground-level site monitoring stations for air pollution measurement typically are either low-cost community maintained or highly-regulated government maintained. To ensure we introduce the least amount of bias during prediction, we utilize solely government maintained site monitoring stations for ground-level pollution data, as we can ensure a low data collection error under uncertainty.

We seek to predict spatiotemporal air pollution spatially continously over Los Angeles in the form of forecasted images in the future. As a result, we must transform these sparse ground-based site monitoring station measurements into an image bounded by the geographical bounds we define for Los Angeles. To achieve this, we apply a simple spatioteporal interpolation of Inverse Distance Weighting (IDW) assuming a linear relationship between the similarity of sensor readings and physical distance between sensors. Figure 1 displays an example ground-truth frame for the PM2.5 measurements collected from the 7 EPA AirNow site monitoring stations as well as the result of applying IDW interpolation to these measurements. Both are overlayed on the terrain map of the geographical bounds we define for Los Angeles.

Within the geographic bounding area we define for Los Angeles, there are two providers of government-maintained site monitoring data. First, we utilize the California Air Resources

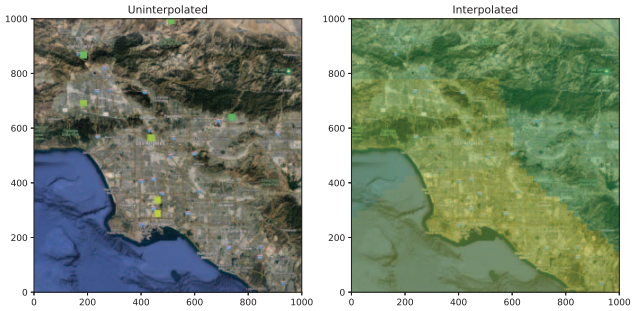


Fig. 1: January 1 2019 Hour 0 PM2.5 Ground-truth Data: Uninterpolated EPA AirNow Sites vs IDW-Interpolated EPA AirNow Sites

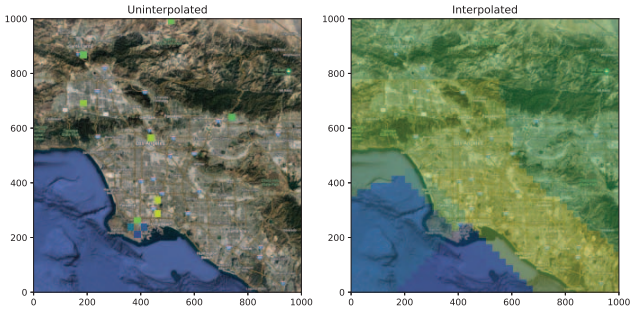


Fig. 2: January 1 2019 Hour 0 PM2.5 Ground-truth Data: Uninterpolated EPA AirNow + POLA Sites vs IDW-Interpolated EPA AirNow + POLA Sites

Board AQMIS2 tool to collect ground-level measurements from 7 site monitoring stations managed by the EPA AirNow program (White et al., 2004). For all air pollutants, there are seven quality assured, validated sites providing hourly measurements at the following locations: Lancaster, Santa Clarita, Reseda, Glendora, Los Angeles – North Main Street, Long Beach, and South Long Beach. Note that for Ozone and Carbon Monoxide, the South Long Beach site was unavailable and instead replaced with an equivalent site in Azusa.

We also utilize government-maintained data of hourly $PM2.5, CO, NO, NO_2$, and O_3 from the Port of Los Angeles site monitoring stations (POLA). The Port of Los Angeles provides ground-level air pollution measurements at four locations within the Los Angeles Harbor: Wilmington Community Station, San Pedro Community Station, Coastal Boundary Station, and the Source-Dominated Station.

In total, we collect hourly samples of $PM2.5, CO, NO, NO_2$ and O_3 from 11 government-maintained site monitoring stations within Los Angeles from both the EPA AirNow and Port of Los Angeles. Figure 2 displays a sample ground-truth frame for PM2.5 measurements collected from the 11 EPA AirNow and Port of Los Angeles sites as well as the result of applying IDW interpolation.

b) Remote-Sensing Atmospheric Pollution Data: When constructing our model, we seek to include an unique combination of varied data sources for air pollution prediction. Particularly, we focused on providing our architecture with input data describing both ground-level and atmospheric air pollution patterns. Data on atmospheric aerosols are typically collected as remote-sensing data through precise instruments onboard low earth orbiting satellites.

We collect remote-sensing satellite imagery from the NASA Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm (Lyapustin et al., 2018). The MAIAC algorithm is a data preprocessing algorithm applied on measurements collected by the NASA Moderate Resolution Imaging Spectroradiometer (MODIS) instrument onboard the NASA Terra and Aqua satellites that converts raw measurements to analytics-ready samples by atmospheric aerosol and air pollutant data from MODIS images, normalizing pixel values, interpolating daily data for hourly use, and removing cloud cover masks (Justice et al., 2002). The Terra and Aqua satellites provide remote-sensing measurements over 36 spectral bands using the MODIS instrument.

We collect data from one of these spectral bands measuring blue-band Aerosol Optical Depth at a central wavelength of 0.47 μm. AOD or Aerosol Optical Depth is a measure of the direct amount of sunlight being blocked by atmospheric aerosols and air pollutants. AOD is perhaps the most comprehensive measure of ambient air pollution and years of research has shown a strong correlation between AOD readings and pollutant concentrations in both atmospheric and ground-level settings (Li et al., 2015). To integrate AOD data with our model, we select the data within our geographic gounds and apply a downsampling layer to reduce the resoltion to a 40 pixel by 40 pixel image, which is the standard resolution we utilize for all data source inputs to our model.

c) Atmospheric Pollution Satellite Imagery: In addition to remote-sensing data, we include satellite imagery of various air pollutants as input to our model. Research in atmospheric sciences shows that there exists positive correlations between pairs of differing air pollutants (Elsom, 1978). We collect satellite imagery of various air pollutants from the TROPOspheric Monitoring Instrument (TROPOMI) onboard the ESA Sentinel-5P satellite (Veefkind et al., 2012). TROPOMI is a spectrometer capable of sensing ultraviolet (UV), visible (VIS), near (NIR) and short-wavelength infrared (SWIR) light and provides high-resolution global hourly data of atmospheric ozone, methane, formaldehyde, aerosol, carbon monoxide, nitrogen dioxide, and sulfur dioxide. For our models, we utilize hourly imagery of methane (CH_4), nitrogen dioxide (NO_2), and carbon monoxide (CO) due to its product's provided spatial resolution and correlation to the five target air pollutants. Similarly to the remote-sensing atmospheric data, we crop the imagery to our geographic bounds and apply downsampling to generate hourly 40 pixel by 40 pixel samples. We choose this standard resolution across all input features because 1 pixel corresponds to roughly a 1 kilometer by 1 kilometer square area of real-world geography within our

defined bounds.

d) Ground-level Meteorological Data: We collect ground-level meteorological data from the Iowa State University MesoNet database (Herzmann et al., 2004). The Environmental Mesonet database collects and records hourly Meteorological Aerodrome (METAR) Reports from Automated Surface Observing Systems (ASOS) located near various airports and municipal airstrips within the continental United States. The ASOS data are primarily used by airlines and air traffic controllers to monitor meteorological features near and around airport runways. The METAR data provides comprehensive hourly reports of 17 ground-level meteorological features including wind speed, wind direction, relative humidity, dew point, precipitation, Air Quality Index (AQI), air pressure, and air temperature. Within our geographic boundaries, there are 24 ASOS sensors providing hourly validated meteorological feature readings. Due to the heterogeneity between stationary (e.g. AQI, temperature, humidity) and non-stationary features (e.g. wind speed, wind direction), we found that applying an IDW interpolation to generate images out of sparse data did not work well for meteorological features. Instead, we developed a novel Graph Convolutional Network (GCN) architecture to perform spatiotemporal kriging on the meteorological data that we represent as weighted directed graphs for each hour.

e) Remote-Sensing Meteorological Data: In addition to ground-level meteorological data, we utilize remote-sensing data of gridded climate features from the GridMET dataset. GridMET applies an assimilation algorithm on top of temporally rich data from the North American Land Data Assimilation System Phase 2 remote-sensing dataset (Mitchell et al., 2004) and spatially rich data from the Parameter-elevation Regressions on Independent Slopes Model (Cosgrove et al., 2003). GridMET provides hourly high-resolution gridded remote-sensing data of various meteorological features, of which we utilize precipitation, wind speed, wind direction, and maximum relative humidity for our predictive model. For preprocessing, we crop raw data to within our geographic bounds and downsample each frame. A visualization of GridMET data for a particular sample for these meteorological features for both the contiguous US and LA county is shown in Figures 3, 4, 5, 6.

f) Remote-sensing Wildfire Data: The geographic location of our study area poses unique challenges in terms of air pollution prediction. Wildfires in Los Angeles county have been increasingly common, and research in the field shows that the smoke/heat generated from a wildfire while it is burning is the largest contributor to ambient air pollution (Liu et al., 2016). We collect remote-sensing wildfire data from two sources: NASA MODIS data and NASA MERRA-2 data (Gelaro et al., 2017; Savtchenko et al., 2004). From the NASA MODIS instrument, we collect remote-sensing data on Fire Radiative Power (FRP). Fire Radiative Power (FRP) is a measure of the radiant heat output from a fire. The main contributors to increased levels of FRP include smoke from wildfires and emissions from the burning of carbon-

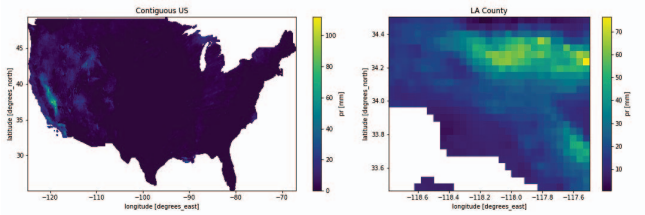


Fig. 3: Remote-Sensing Data of Precipitation (mm) for 1-1-2019 Hour 0: Contigouous US vs LA County

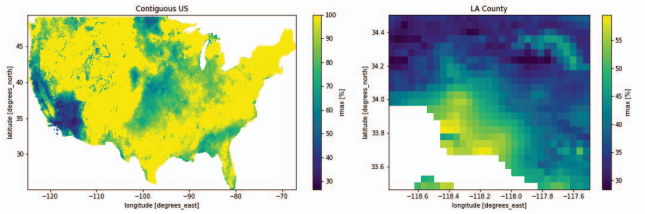


Fig. 4: Remote-Sensing Data of Relative Humidity (%) for 1-1-2019 Hour 0: Contigouous US vs LA County

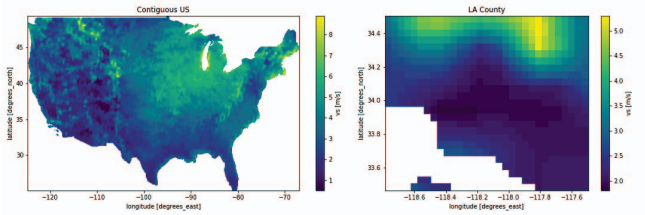


Fig. 5: Remote-Sensing Data of Wind Speed (m/s) for 1-1-2019 Hour 0: Contigouous US vs LA County

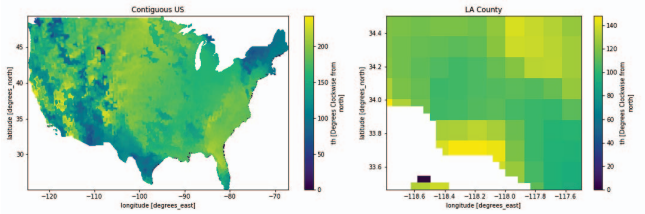


Fig. 6: Remote-Sensing Data of Wind Direction (° from North) for 1-1-2019 Hour 0: Contigouous US vs LA County

based fuel, such as carbon monoxide CO and carbon dioxide CO_2 emissions. There is a strong positive correlation between wildfires and FRP readings as well as a weaker positive correlation between carbon emissions (CO_2, CO) and FRP readings. We also use wildfire and heat data from the NASA MERRA-2 data source. The Modern Era Retrospective analysis for Research and Applications, version 2 (MERRA-2) is a global atmospheric reanalysis produced by the NASA Global Modeling and Assimilation Office (GMAO). The data format of the MERRA-2 features we use in our predictive model are time-indexed series of gridded remote-sensing data. We use MERRA-2 imagery of three wildfire/heat features: Planetary Boundary Layer (PBL) height, surface air temperature, and surface exchange coefficient for heat. For all wildfire remote-sensing data, we similarly crop and downsample each hourly

sample to a 40 px by 40 px frame.

B. Model Architecture and Implementation

Our model utilizes a two-stage architecture to achieve the following: (1) interpolate sparse ground-level meteorological features to denser time-indexed series of gridded data and (2) apply a spatiotemporal predictive model on various time-indexed series of gridded data features. Of the data sources described earlier, all but the ground-level meteorological features are in the format of hourly sequential gridded data. We refer to the first stage of our architecture as the MeteoGCN and the second stage as the ConvLSTM. We apply the Graph Convolutional Network (GCN) architecture on the sparse ground level meterological data. The GCN architecture is primarily effective in producing high-quality interpolation, allowing us to convert sparse graph inputs to denser, more connected graphs. We adapt the architecture proposed by Wu et al. (2021) for spatiotemporal kriging with GCNs. First, we represent the raw meteorological features for a particular hour as a weighted directed graph. We construct this weighted directed graph by defining the set of vertices V for each graph to be the meteorological monitoring station. The vertex attributes are the stationary features recorded by a particular station, while the edge attributes are the non-stationary features. Here, we refer to stationary features as those which can be defined as bound to the physical location of the monitoring site, such as temperature, relative humidity, precipitation. Non-stationary features include wind speed, wind direction, and wind gust. The meteorological graph creation process is described in Algorithm 1.

The GCN is then trained by hiding a set of nodes and vertices along with their corresponding attribute vectors and instructing the GCN to interpolate for these missing structures using ground truth data from the neighboring set of nodes and edges. At the end of training, GCN is capable of interpolating a sparse meteorological graph into a dense graph.

An intermediate step between the MeteoGCN and ConvLSTM is to decode the dense meteorological graph into image-like gridded data. We apply an unsupervised learning graph representation learning algorithm included within the StellarGraph Python library to perform this embedding extraction. Additionally, we ensure that the outputted embedding image is geographically bounded within our defined bounds, as to maintain homogeneity with other input features.

Finally, we apply the ConvLSTM architecture on the preprocessed input features and output of the MeteoGCN. The ConvLSTM architecture learns on data in the format of a 5D tensor with the following dimensions: (samples, frames, rows, columns, channels). Each input data source can be represented as a channel with rows $= 40$ and columns $= 40$. Each sample contains 24 frames corresponding to 24 hours. In total, our historical dataset contains 1340 days worth of data corresponding to 1340 samples with 24 frames of 40 by 40 image bundles with 10 channels. The 10 channels is derived as 1 channel from the AQMIS ground-level pollutant grid, 1 channel from the output of the MeteoGCN, 1 channel from MODIS AOD, 3 channels from TROPOMI imagery, 1 channel from MODIS FRP, and 3 channels from MERRA-2 features. A breakdown of the input dataset is displayed in Figure 7.

Algorithm 1 Meteorological Graph Construction

Input: Meteorological site features $f_i \in F$, where each f_i contains site coordinates x_i, y_i and a set of site-specific stationary $s_i \in S$ and non-stationary $n_i \in N$ feature values. Boundary latitude values lat_{max}, lat_{min}. Boundary longitude values long_{max}, long_{min}.

```
Initialize 40 × 40 array grid A.
Initialize weighted directed graph G = (V, E)
for f_i ∈ F do
    grid_x, grid_y = ⌊ x_i·40 / (long_max − long_min) ⌋, ⌊ y_i·40 / (lat_max − lat_min) ⌋
    A[grid_x][grid_y] = vector of site-specific stationary values s_i
    Set A[grid_x][grid_y] as vertex of G
end for
for f_i ∈ F do
    for n_i ∈ N do
        Let start_x, start_y be the starting coordinates of a weighted directed edge in G
        start_x, start_y = grid_x, grid_y
        Recover end_x, end_y from site-specific non-stationary value n_i.
        Create weighted directed edge in G starting from vertex located at (start_x, start_y) and ending at vertex located at (end_x, end_y) with weight of |n_i|.
    end for
end for
```

Output: Geographically bound graph feature matrix grid A, Weighted Directed Graph G

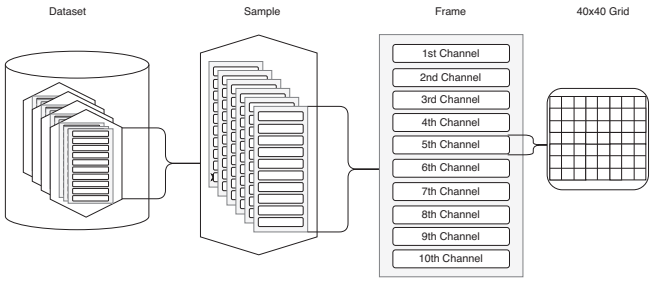


Fig. 7: ConvLSTM Input 5-Dimensional Tensor Breakdown

We use an 80-10-10 training-validation-testing split corresponding to 1072 samples for training, 134 samples for validation and 134 samples for testing. We generate labels for our model by adding a 1-hour time-lag to our input frames in each sample, which implies our model does hour-by-hour prediction for 24 hours in the future given the last 24 hours of data. For example, an input frame in our model would include data for frames (hours) 1 through 24, while the label would include data from frames 25 through 48. We can then compare our model's predicted results for frames 25-48 and evaluate

against the label's frames 25-48 to evaluate performance. Since the architecture of an LSTM dictates that the forecast horizon depends on the number of samples passed as input, we can extend our prediction for further than 24 hours given additional historical data (i.e. using 48 hours in the past to predict 48 hours in the future, hour-by-hour).

We also study the impact of ground-level features on air pollution prediction, so we additionally perform an ablation study where we train a ConvLSTM model with solely satellite imagery and remote-sensing data. The 5D input tensor for this satellite only model contains 8 channels, as we do not utilize ground-level meteorological or air pollution features as input. The label for this model remains as the IDW-interpolated ground-level air pollution data collected from the 11 site monitoring stations.

III. Results

Our model predicts spatiotemporal PM2.5 in terms micrograms per cubic meter ($\frac{\mu g}{m^3}$) and NO_2, NO, CO, O_3 in terms of parts per million (ppm) spatially continously over Los Angeles county hourly. Our model utilizes 24 hours of data in the past to make hour-by-hour predictions for these five air pollutants 24 hours in the future. To evaluate the accuracy of our model, we can consider two scenarios. Firstly, if we want to perform spatially continuous predictions over Los Angeles county in the future, then our predicted samples will be in the form of time-indexed series of images. The ground truth labels would then be the IDW-interpolated ground-level air pollution data from the 7 EPA AirNow site monitoring stations. To evaluate this model, we could utilize a per-pixel Root Mean Square Error (RMSE) or similar metrics and average these values to get a composite error value. However, computing per-pixel error is not entirely representative of the underlying shared structures between the predicted image and ground truth. Additionally, apart from the 7 pixels corresponding within a 1 km by 1km physical distance of the ground truth site monitoring stations, the remaining pixels of the 40 px by 40 px ground truth image is generated via IDW-interpolation. Thus, the alternative scenario is to predict air pollution in the future as time-indexed series of images spatially continuously but evaluate model performance against solely the 7 ground-based site monitoring stations by extracting the 7 pixels within the predicted image that lie within a 1 km^2 area of the 7 ground truth monitoring sites.

We evaluate the performance of our model using the Root Mean Square Error (RMSE) metric. For possible comparisons against baselines and future research, we also compute the Normalized Root Mean Square Error (NRMSE), which normalizes RMSE scores by the mean of the ground truth samples. Table I displays the RMSE and NRMSE error values calculated over all 24-hour predictions in the testing set (134 days from April 20 2022 to September 1 2022) at the 7 EPA AirNow ground-truth site locations for all air pollutant targets. The best performing site location per pollutant is italicized in each target pollutant section and the best performing site location across all target pollutants is bolded.

TABLE I: RMSE and NRMSE Error Values over Testing Set at each AirNow Sensor Location for all Target Air Pollutants

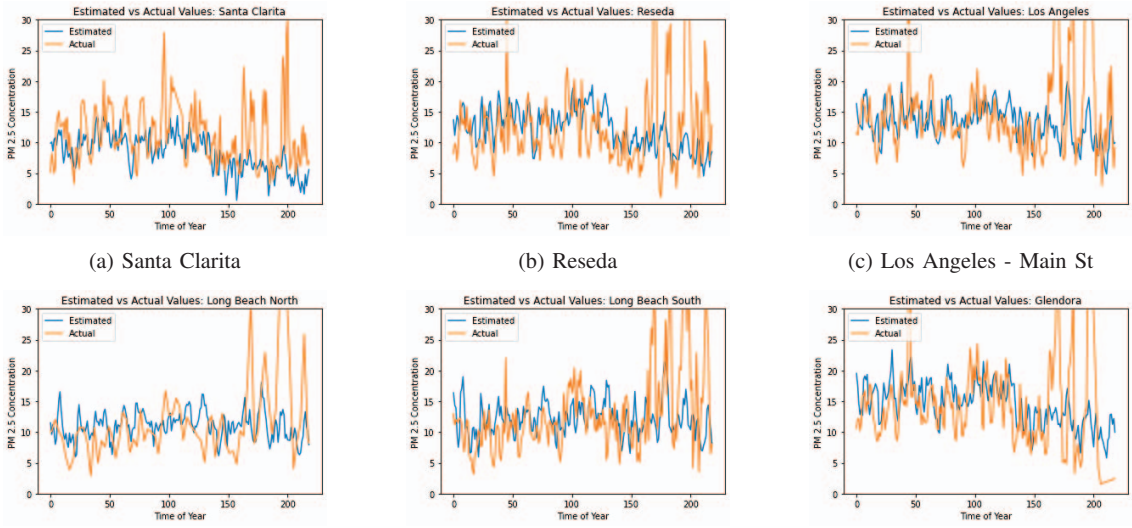
Target	Sensor Location	Testing Set Metrics RMSE	NRMSE
PM2.5	Lancaster	1.0341	0.0523
	Glendora	1.1494	0.0557
	Santa Clarita	*1.0691*	*0.0503*
	Reseda	1.2834	0.0579
	LA—Main St	1.5725	0.0621
	Long Beach	1.4451	0.0598
	South Long Beach	1.4528	0.0610
Nitrogen Dioxide	*Lancaster*	*0.0015*	*0.0535*
	Glendora	0.0017	0.0593
	Santa Clarita	0.0016	0.0562
	Reseda	0.0016	0.0566
	LA—Main St	0.0023	0.0661
	Long Beach	0.0019	0.0602
	South Long Beach	0.0018	0.0593
Carbon Monoxide	***Lancaster***	**0.5461**	**0.0456**
	Glendora	0.5582	0.0475
	Santa Clarita	0.5847	0.0495
	Reseda	0.5696	0.0483
	LA—Main St	0.6961	0.0590
	Long Beach	0.6353	0.0568
	Azusa	0.5681	0.0481
Nitric Oxide	*Lancaster*	*0.0009*	*0.0518*
	Glendora	0.0012	0.0535
	Santa Clarita	0.0014	0.0581
	Reseda	0.0013	0.0575
	LA—Main St	0.0023	0.0689
	Long Beach	0.0021	0.0641
	South Long Beach	0.0019	0.0623
Ozone	Lancaster	0.0045	0.0596
	Glendora	0.0041	0.0569
	Santa Clarita	*0.0039*	*0.0517*
	Reseda	0.0048	0.0605
	LA—Main St	0.0053	0.0649
	Long Beach	0.0055	0.0628
	Asuza	0.0041	0.0591

We also evaluate the error of a model trained with eight input channels not including ground-level meteorological or air pollution data to understand the impact of satellite imagery for PM2.5 prediction. The RMSE scores calculated on 7 ground-truth EPA AirNow sensors for 24-hour prediction averaged over the testing set for PM2.5 prediction for a satelite data only and full MeteoGCN-ConvLSTM model is described in Table II.

TABLE II: RMSE Error Values over Testing Set at each AirNow Sensor Location for PM2.5 Prediction: Satellite Data Only MeteoGCN-ConvLSTM vs Full MeteoGCN-ConvLSTM

Target	Sensor Location	Testing Set RMSE Satellite Only	Full
PM2.5	Lancaster	1.8276	1.0341
	Glendora	3.1834	1.1494
	Santa Clarita	1.8496	1.0691
	Reseda	2.4263	1.2834
	LA—Main St	3.1145	1.5725
	Long Beach	1.9805	1.4451
	South Long Beach	2.3314	1.4528

Additionally, in Figure 8 we provide visualizations of the



(a) Santa Clarita (b) Reseda (c) Los Angeles - Main St

(d) Long Beach (e) South Long Beach (f) Glendora

Fig. 8: PM2.5 Predicted vs Actual Plots of Ground-truth Site Locations for Satellite Data Only MeteoGCN-ConvLSTM

remote-sensing data and satellite imagery only MeteoGCN-ConvLSTM model for PM2.5 prediction from January 1 2022 to September 1 2022 comparing predicted PM2.5 against ground truth values.

IV. Conclusion

In this paper, we apply a complex two-stage deep learning architecture for spatiotemporal ground-level air pollution prediction. We developed the novel MeteoGCN-ConvLSTM architecture for hourly ground-level prediction of various air pollutants including nitric oxide (NO), nitrogen dioxide (NO_2), carbon monoxide (CO), ozone (O_3), and PM2.5 spatially continuously over Los Angeles county. Our model applies ground-level pollution data, remote-sensing atmospheric pollution data, atmospheric pollution satellite imagery, ground-level meteorological data, remote-sensing meteorological data, and remote-sensing wildfire data for spatiotemporal air pollution prediction. We evaluated our predictions through RMSE and NRMSE error metrics against ground-truth EPA AirNow site monitoring stations.

Our results show that the MeteoGCN-ConvLSTM model performs the best in predicting carbon monoxide spatially continuously in Los Angeles county. The average NRMSE score across all site locations for carbon monoxide prediction showed a 10.12% decrease in error compared to the second best performing pollution target of PM2.5 and a 20.73% decrease in error compared to the worst performing pollutant target of nitric oxide. We theorize that carbon monoxide is the best predicted target for our model due to MODIS Fire Radiative Power (FRP) remote-sensing input data. Since FRP is strongly correlated to wildfire emissions and weakly correlated to carbon emissions such as carbon monoxide and carbon dioxide, the model has more historical information to utilize in prediction for carbon monoxide over other pollutant targets.

Further, we can see that across all pollutant targets, the RMSE and NRMSE scores show that the MeteoGCN-ConvLSTM model performs the best at either the Lancaster or Santa Clarita sites. Geographically, this could be explained due to the site monitoring stations' proximity to the Angeles National Forest, which MERRA-2 and MODIS FRP provide particularly valuable information about during wildfire season, as this location is a hotspot for potential wildfires. Also, these sites are the furthest from the metropolitan hub of downtown Los Angeles, so they may experience less variability in sensor readings compared to other sites.

Finally, we can see that ground-level meteorological and pollutant data is crucial for air pollution prediction, as the full MeteoGCN-ConvLSTM outperforms the satellite imagery and remote-sensing data only model with a 36.35% decrease in average RMSE testing set error over all site locations for PM2.5 prediction. From Figure 8, we can visually identify significantly more variability in predicted PM2.5 during the later months of 2022 (i.e. August through September 2022), coinciding with the start of wildfire season in Los Angeles.

V. Future Work

In the future, we hope to include low-cost community maintained air pollution site readings as input to our model. We also hope to understand and account for the data measurement error for ground-based site readings. Additionally, we can increase the spatial resolution of our input grid data and thus our predictions for finer-grained air pollution forecasting. Finally, we hope to extend our predictive results for understanding the health effects and impacts of forecasted air pollutants on residents in the area, such as risk scores for developing asthma, emphysema, and cardiovascular illnesses.

References

Erik Boye Abrahamsen, Ole Magnus Brastein, and Bernt Lie. Machine Learning in Python for Weather Forecast based on Freely Available Weather Data. *Exergy Analysis for Combined Heat and Power (CHP) Plants*, 2018.

Colin Bellinger, Mohomed Shazan Mohomed Jabbar, Osmar Zaïane, and Alvaro Osornio-Vargas. A systematic review of data mining and machine learning for air pollution epidemiology. *BMC public health*, 17(1):1–19, 2017.

Emmanuel Cocom, Pratyush Muthukumar, Jeanne Holm, Dawn Comer, Anthony Lyons, Irene Burga, Christa A Hasenkopf, Chisato Calvert, and Mohammad Pourhomayoun. Particulate Matter Forecasting in Los Angeles County with Sparse Ground-based Sensor Data Analytics. In *AGU Fall Meeting Abstracts*, volume 2020, pages A061–0004, 2020.

Brian A Cosgrove, Dag Lohmann, Kenneth E Mitchell, Paul R Houser, Eric F Wood, John C Schaake, Alan Robock, Curtis Marshall, Justin Sheffield, Qingyun Duan, et al. Real-time and retrospective forcing in the north american land data assimilation system (NLDAS) project. *Journal of Geophysical Research: Atmospheres*, 108(D22), 2003.

DM Elsom. Spatial correlation analysis of air pollution data in an urban area. *Atmospheric Environment (1967)*, 12(5): 1103–1107, 1978.

Ronald Gelaro, Will McCarty, Max J Suárez, Ricardo Todling, Andrea Molod, Lawrence Takacs, Cynthia A Randles, Anton Darmenov, Michael G Bosilovich, Rolf Reichle, et al. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of climate*, 30(14):5419–5454, 2017.

Aditya Grover, Ashish Kapoor, and Eric Horvitz. A Deep Hybrid Model for Weather Forecasting. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 379–386, 2015.

Daryl Herzmann, Ray Arritt, and Dennis Todey. Iowa environmental mesonet. *Available at mesonet. agron. iastate. edu/request/coop/fe. phtml (verified 27 Sept. 2005). Iowa State Univ., Dep. of Agron., Ames, IA*, 2004.

CO Justice, JRG Townshend, EF Vermote, E Masuoka, RE Wolfe, Nazmi Saleous, DP Roy, and JT Morisette. An overview of MODIS land data processing and product status. *Remote sensing of Environment*, 83(1-2):3–15, 2002.

Thomas N Kipf and Max Welling. Semi-supervised Classification with Graph Convolutional Networks. *arXiv preprint arXiv:1609.02907*, 2016.

Jing Li, Barbara E Carlson, and Andrew A Lacis. How well do satellite AOD observations represent the spatial and temporal variability of PM2.5 concentration for the United States? *Atmospheric Environment*, 102:260–273, 2015.

Jia Coco Liu, Loretta J Mickley, Melissa P Sulprizio, Francesca Dominici, Xu Yue, Keita Ebisu, Georgiana Brooke Anderson, Rafi FA Khan, Mercedes A Bravo, and Michelle L Bell. Particulate air pollution from wildfires in the western us under climate change. *Climatic change*, 138(3):655–666, 2016.

Alexei Lyapustin, Yujie Wang, Sergey Korkin, and Dong Huang. MODIS collection 6 MAIAC algorithm. *Atmospheric Measurement Techniques*, 11(10):5741–5765, 2018.

Kenneth E Mitchell, Dag Lohmann, Paul R Houser, Eric F Wood, John C Schaake, Alan Robock, Brian A Cosgrove, Justin Sheffield, Qingyun Duan, Lifeng Luo, et al. The multi-institution north american land data assimilation system (NLDAS): Utilizing multiple gcip products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research: Atmospheres*, 109(D7), 2004.

Pratyush Muthukumar, Emmanuel Cocom, Jeanne Holm, Dawn Comer, Anthony Lyons, Irene Burga, Christa Hasenkopf, and Mohammad Pourhomayoun. Real-time Spatiotemporal NO2 Air Pollution Prediction with Deep Convolutional LSTM through Satellite Image Analytics. In *AGU Fall Meeting Abstracts*, volume 2020, 2020a.

Pratyush Muthukumar, Emmanuel Cocom, Jeanne Holm, Dawn Comer, Anthony Lyons, Irene Burga, Christa Hasenkopf, and Mohammad Pourhomayoun. Real-time Spatiotemporal Air Pollution Prediction with Deep Convolutional LSTM through Satellite Image Analysis. In *16th International Conference on Data Science (ICDATA '20)*, pages 317–328. Springer Nature, 2020b.

Pratyush Muthukumar, Emmanuel Cocom, Kabir Nagrecha, Jeanne Holm, Dawn Comer, Anthony Lyons, Irene Burga, Chisato Fukuda Calvert, and Mohammad Pourhomayoun. Satellite Image Atmospheric Air Pollution Prediction through Meteorological Graph Convolutional Network with Deep Convolutional LSTM. In *7th Annual Conference on Computational Science and Computational Intelligence (CSCI-ISAI '20)*. IEEE CPS, 2020c.

Pratyush Muthukumar, Emmanuel Cocom, Kabir Nagrecha, Dawn Comer, Irene Burga, Jeremy Taub, Chisato Calvert, Jeanne Holm, and Mohammad Pourhomayoun. Predicting PM2.5 Atmospheric Air Pollution using Deep Learning with Meteorological Data and Ground-based Observations and Remote-sensing Satellite Big Data. *Air Quality, Atmosphere & Health*, pages 1–14, 2021a.

Pratyush Muthukumar, Kabir Nagrecha, Emmanuel Cocom, Dawn Comer, Irene Burga, Jeremy Taub, Chisato Calvert, Jeanne Holm, and Mohammad Pourhomayoun. Predicting PM2.5 Air Pollution using Deep Learning with Multisource Satellite and Ground-based Observations and Meteorological and Wildfire Big Data. In *AGU Fall Meeting Abstracts*, volume 2021, 2021b.

Pratyush Muthukumar, Kabir Nagrecha, Dawn Comer, Chisato Fukuda Calvert, Navid Amini, Jeanne Holm, and Mohammad Pourhomayoun. PM2.5 Air Pollution Prediction through Deep Learning Using Multisource Meteorological, Wildfire, and Heat Data. *Atmosphere*, 13(5):822, 2022.

Kabir Nagrecha, Pratyush Muthukumar, Emmanuel Cocom, Jeanne Holm, Dawn Comer, Irene Burga, and Mohammad Pourhomayoun. Sensor-Based Air Pollution Prediction

using Deep CNN-LSTM. In *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 694–696. IEEE, 2020.

Sanam Narejo and Eros Pasero. Meteonowcasting using Deep Learning Architecture. *(IJACSA) International Journal of Advanced Computer Science and Applications*, 8(8), 2017.

POLA. URL https://www.portoflosangeles.org/environment/air-quality/air-quality-monitoring. (Accessed on 15 September 2022).

Hannah Ritchie and Max Roser. Air Pollution. *Our World in Data*, 2017. https://ourworldindata.org/air-pollution.

A Savtchenko, D Ouzounov, S Ahmad, J Acker, G Leptoukh, J Koziana, and D Nickless. Terra and aqua MODIS products available from NASA GES DAAC. *Advances in Space Research*, 34(4):710–714, 2004.

Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional LSTM Network: A machine learning approach for precipitation nowcasting. *arXiv preprint arXiv:1506.04214*, 2015.

JP Veefkind, I Aben, K McMullan, H Förster, J De Vries, G Otter, J Claas, HJ Eskes, JF De Haan, Q Kleipool, et al. TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote sensing of environment*, 120:70–83, 2012.

Jonathan A Weyn, Dale R Durran, and Rich Caruana. Improving Data-Driven Global Weather Prediction using Deep Convolutional Neural Networks on a Cubed Sphere. *Journal of Advances in Modeling Earth Systems*, 12(9): e2020MS002109, 2020.

John E White, Richard A Wayland, Timothy S Dye, and Alan C Chan. AIRNow air quality notification and forecasting system. In *Beijing International Environment Forum, Beijing, China, September*, pages 14–15, 2004.

Yuankai Wu, Dingyi Zhuang, Aurelie Labbe, and Lijun Sun. Inductive graph neural networks for spatiotemporal kriging. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4478–4485, 2021.