# iSpray: Reducing Urban Air Pollution with Intelligent Water Spraying

YUN CHENG, ETH Zurich, Switzerland
ZIMU ZHOU, Singapore Management University, Singapore
LOTHAR THIELE, ETH Zurich, Switzerland

Despite regulations and policies to improve city-level air quality in the long run, there lack precise control measures to protect critical urban spots from heavy air pollution. In this work, we propose iSpray, the first-of-its-kind data analytics engine for fine-grained $PM_{2.5}$ and $PM_{10}$ control at key urban areas via cost-effective water spraying. iSpray combines domain knowledge with machine learning to profile and model how water spraying affects $PM_{2.5}$ and $PM_{10}$ concentrations in time and space. It also utilizes predictions of pollution propagation paths to schedule a minimal number of sprayers to keep the pollution concentrations at key spots under control. In-field evaluations show that compared with scheduling based on real-time pollution concentrations, iSpray reduces the total sprayer switch-on time by 32%, equivalent to $1,782\ m^3$ water and $18,262\ kWh$ electricity in our deployment, while decreasing the days of poor air quality at key spots by up to 16%.

4

CCS Concepts: • **Human-centered computing** → *Ubiquitous and mobile computing systems and tools*; • **Hardware** → *Sensor applications and deployments.*

Additional Key Words and Phrases: Air Pollution; Water Spraying

**ACM Reference Format:**
Yun Cheng, Zimu Zhou, and Lothar Thiele. 2022. iSpray: Reducing Urban Air Pollution with Intelligent Water Spraying. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 1, Article 4 (March 2022), 29 pages. https://doi.org/10.1145/3517227

## 1 INTRODUCTION

Urban air pollution threatens the health of residents. Epidemiological studies find a positive correlation between exposure to high concentrations of small particulate matter (with diameter less than 10 and 2.5 micrometer, *i.e.*, $PM_{10}$ and $PM_{2.5}$) and cardiovascular or respiratory diseases [3, 26]. Yet 91% of the world's population lives in areas where air quality levels exceed WHO limits [39]. In response, many cities have deployed large-scale sensor networks [9, 16, 24] to monitor urban air pollution [9, 40], generate fine-grained air quality maps [5, 16, 45], and forecast heavily polluted areas [10, 46] for citizens to adjust their travel plans accordingly.

In addition to *passive monitoring* of urban air pollution, *active control* strategies are also crucial. Governments and authorities have launched various policies and regulations to reduce emissions from factories, transport, and household to improve the *overall* (*e.g.*, city-level, annual average) air quality [4, 29]. However, there lacks measures for *fine-grained* (*e.g.*, specific districts, hourly average) air pollution control. Such measures are complementary to the city-level policies and regulations and aim to offer precise protection to critical points of interest (POIs) such as residential areas, schools, hospitals, etc. within the city.

Corresponding Author: Zimu Zhou.
Authors' addresses: Yun Cheng, ETH Zurich, Computer Engineering and Networks Laboratory, Zurich, Switzerland, chengyu@ethz.ch; Zimu Zhou, Singapore Management University, School of Computing and Information Systems, Singapore, Singapore, zimuzhou@smu.edu.sg; Lothar Thiele, ETH Zurich, Computer Engineering and Networks Laboratory, Zurich, Switzerland, thiele@ethz.ch.


2474-9567/2022/3-ART4 $15.00
https://doi.org/10.1145/3517227

In this paper, we explore water spraying for precise $PM_{2.5}$ and $PM_{10}$ control at key urban POIs. Water spraying proves effective for dust control at construction and mining sites [20, 21, 36] and has recently been applied for PM reduction in urban areas [42]. The principle is to atomize water into micro droplets to fall in combination with ambient dusts [30]. The fog produced by commodity sprayers can spread 10 to 100 meters and our field studies show water spraying reduces $PM_{2.5}$ and $PM_{10}$ concentrations by 20% to 30% (up to 13 $\mu g/m^3$ for $PM_{2.5}$ and 19 $\mu g/m^3$ for $PM_{10}$), in various weather conditions (see Sec. 4.1), which is considered significant improvements in air pollution control [15]. Note that a reduction of $10\mu g/m^3$ in $PM_{2.5}$ and $PM_{10}$ concentrations is valuable for the health of residents, especially on human respiratory system [41]. Clinical research indicated that the average life span was extended by 0.35 years for every $10\mu g/m^3$ decrease of $PM_{2.5}$ [11], whereas the mortality of cardiopulmonary diseases and lung cancer increased by 6% and 8%, respectively, for every $10\mu g/m^3$ increase of $PM_{2.5}$ [38]. It is also shown that for each increase of $PM_{10}$ by $10\mu g/m^3$, the overall morbidity increased by 0.38% [19], and the mortality related to respiratory diseases increased by 0.58% [1]. Therefore, water spraying holds potential for effective urban $PM_{2.5}$ and $PM_{10}$ control at fine spatiotemporal granularity, and provides valuable benefits for human health, especially when the pollution reduction is over 10 $\mu g/m^3$.

Designing an urban water spraying system, however, faces multiple technical challenges. *(i)* There lacks quantitative models on how water spraying reduces $PM_{2.5}$ and $PM_{10}$ concentrations in the urban outdoor space. Existing models are primarily derived for indoor environments with controlled ventilation [12, 42]. They are unfit for profiling pollution reduction outdoors due to the complex aerodynamics and meteorological factors in the open urban space. It is difficult to decide which sprayers to switch on without a quantitative pollution reduction model. *(ii)* The water spraying system should be cost-effective, *i.e.*, a minimal number of sprayers are switched on to keep the $PM_{2.5}$ and $PM_{10}$ concentrations at the key POIs within the desired range. For example, a single sprayer in our deployment consumes 0.6 $m^3$ water and 5 $kWh$ electricity per hour, which adds up to 792 $m^3$ water and 6600 $kWh$ electricity a day if all the sprayers are operating non-stop. We empirically show that a strategically selected sprayer subset would suffice to ensure the air quality level at given POIs (see Sec. 6.2.3).

To this end, we propose iSpray, a data analytics engine for fine-grained air pollution control at key urban POIs via cost-effective water spraying. We exploit both domain knowledge and data-driven approaches to characterize and model the spraying-induced pollution reduction in time and space. The hybrid approach enables accurate pollution reduction modeling even with limited spraying data for training. We further propose a sprayer scheduling scheme based on the predictions of pollution propagation paths. By prioritizing water spraying along the pollution propagation paths, we avoid unnecessary spraying that only marginally suppresses the pollution at the targeting POIs. The main contributions of this paper are summarized as follows.

- To the best of our knowledge, we are the first to characterize the effect of commodity water sprayers on $PM_{2.5}$ and $PM_{10}$ reduction in outdoor urban areas. Field studies show that the spraying-induced pollutant reduction at the sprayer's location is non-linearly weather-dependent, which can be modeled via a neural network, and the model generalizes across sprayer locations.
- We design an explainable model to integrate water spraying into urban air quality map generation. We exploit domain knowledge to isolate the impact of spraying on the pollutant's spatial distribution for easy sprayer scheduling and accurate map generation with limited spraying data. Evaluations show that our approach outperforms pure data-driven map generation by 7.9 to 9.3 in mean absolute error (MAE).
- We propose a propagation-aware sprayer scheduling algorithm for cost-effective air pollution control at key urban spots. Compared with the baseline strategy that switches on sprayers according to the current pollutant concentration, our scheduling scheme reduces the total sprayer switch-on time by 32%, or equivalently $1,782$ $m^3$ water and $18,262$ $kWh$ electricity for our deployment, while decreasing the days of poor $PM_{2.5}$ and $PM_{10}$ air quality at key POIs by 13% and 16%.

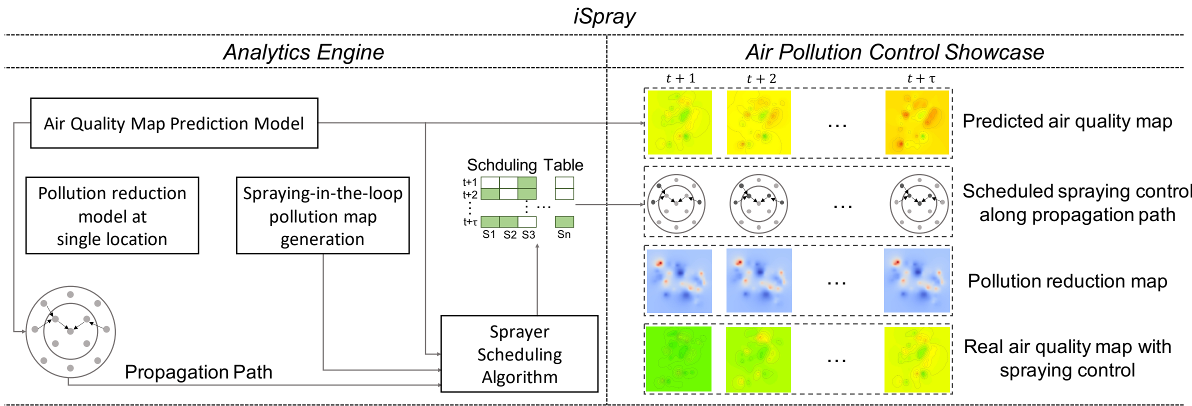



**Fig. 1.** iSpray overview: functional modules of iSpray (left); workflow of iSpray (right).

In the rest of this paper, we provide an overview of iSpray in Sec. 2, explain the deployment and data collection in Sec. 3, and elaborate on each module in Sec. 4, Sec. 5 and Sec. 6. We present the overall evaluations of iSpray in Sec. 7, review related work in Sec. 8 and conclude in Sec. 9.

## 2 ISPRAY OVERVIEW

iSpray is a data analytics engine for urban air pollution control with commodity sprayer hardware. It offers *(i)* pollution reduction modeling at single sprayer locations, *(ii)* pollution map generation, and *(iii)* cost-effective sprayer scheduling. Fig. 1 illustrates the functional modules in iSpray. Table 1 summarizes the major notations that will be used throughout this paper.

The *pollution reduction modeling module* (see Sec. 4) characterizes and quantifies the impact of water spraying on $PM_{2.5}$ and $PM_{10}$ concentrations at the locations where the sprayers are installed. It is the foundation to integrate the impact of water spraying into air quality map generation (*i.e.*, spatial distribution of pollutant concentration). Existing pollution reduction models for water spraying either halt at simulations [42] or are designed for indoor scenarios with controlled ventilation [20, 21]. They are unfit for modeling pollution reduction outdoors because they fail to account for the complex aerodynamics and meteorological factors in the open urban space. iSpray takes a data-driven approach to model how water spraying reduces outdoor air pollution under various environmental conditions. Through in-field studies, iSpray learns a neural network that quantifies the reduction in $PM_{2.5}$ or $PM_{10}$ concentration at single sprayer locations given specific spraying time, meteorological conditions, and other environmental factors.

The *pollution map generation module* (see Sec. 5) models how water spraying affects the *spatial* distribution of $PM_{2.5}$ and $PM_{10}$ concentrations. Due to limited spraying data for effective training, we model the spatial pollution reduction with both domain knowledge and data-driven approaches. Instead of feeding all data into a machine learning model as previous studies [5, 8, 16], we exploit a Gaussian plume model [2, 43] to simulate pollution reduction in space by regarding the sprayer as a sink that absorbs pollution. We also propose a parameter learning strategy to estimate the inaccessible parameters in the Gaussian plume model from historical data. Evaluations show our hybrid modeling method outperforms pure data-driven schemes in modeling spraying-induced pollution reduction maps (see Sec. 5.4.2).

The *sprayer scheduling module* (see Sec. 6) aims to keep the air pollution at crucial POIs under predefined thresholds by switching on a minimal set of sprayers. Our measurements show that the spraying-induced pollution reduction is non-uniform across space (*e.g.*, due to wind direction) and is non-linear to multiple environmental

**Table 1.** Summary of major notations.

| Notation | Explanation |
|---|---|
| $C(g)$ | ground truth pollution concentration in grid $\boldsymbol{g}$ |
| $\hat{C}(g)$ | estimated pollution concentration in grid $\boldsymbol{g}$ |
| $C_d(g)$ | pollution concentration in grid $\boldsymbol{g}$ due to dispersion |
| $R(g)$ | overall pollution reduction in grid $g$ due to water spraying |
| $d_{cross}(.)$ | crosswind distance between two grids |
| $d_{down}(.)$ | downwind distance between two grids |
| $g$ | grid in 2-dimensional space |
| $i$ | index for POI |
| $k$ | index for sprayer |
| $L_i$ | location of POI |
| $\boldsymbol{m}$ | mean function of Gaussian process |
| $\boldsymbol{v}$ | covariance function of Gaussian process |
| $K$ | total number of sprayers in the region of interest |
| $o$ | operating status of a sprayer, which can be *on* or *off* |
| $r(g\|s_k)$ | pollutant reduction in grid $g$ with sprayer $s_k$ switched on |
| $s_k =< g_k, o_k >$ | sprayer $s_k$ in grid $g_k$ with operating status $o_k$ |
| $t$ | discrete time index |
| $\Delta t$ | time duration, set to 1 to 6 hours for map generation and sprayer scheduling |
| $\phi(.)$ | learned function for Gaussian plume dispersion parameter $\sigma$ |
| $Q_{air}$ | pollution emission rate |
| $Q_{s_k}(\Delta t)$ | accumulative pollution reduction in grid of sprayer $s_k$ |
| $Q_{s_k}$ | abbreviation for $Q_{s_k}(\Delta t)$ when $\Delta t$ is set to 1 hour |
| $\sigma$ | Gaussian plume dispersion parameter, which is a function of downwind distance and other features |
| $\bar{w}$ | average horizontal wind speed |

factors (*e.g.*, weather). Therefore, the amount of pollution reduction at a given POI varies if a different sprayer is switched on. iSpray proposes a propagation-path-based heuristic to rank the importance of sprayers to the pollutant reduction at each POI, so as to turn on a minimal number of sprayers without exceeding the targeting pollution threshold at crucial POIs.

## 3 HARDWARE DEPLOYMENT AND DATA COLLECTION

iSpray is designed as a software solution that works with commodity sprayer hardware. This section presents the sprayer hardware deployment and data collection in this study.

### 3.1 Sprayer Hardware and Deployment

Water spraying is widely used for dust control in the construction and mining industries [20, 21, 36] and has also been applied for ambient particulate matter reduction in urban areas [42]. The principle of water spraying for dust suppression is to atomize water into droplets of size comparable to fine particulate matters *e.g.*, $1\mu m$ to $8\mu m$. These droplets can stay suspended in the air for a long time and will then fall in combination with the ambient dusts and particulate matters [30].

A commodity sprayer exploits an electric motor to press water through high-pressure resistant pipes and atomizing nozzles to produce micro droplets. A single nozzle can produce fog lengths of 3 to 5 meters, which can spread 10 to 30 meters in windless conditions and 100 meters in windy conditions. A typical sprayer consists of an atomization system, a water tank, a multi-nozzle sprinkler and other control modules. The atomization system and water tank are normally installed on the ground (see Fig. 2-a and Fig. 2-b) while the sprinkler is


( a ) ( b ) ( c )

**Fig. 2.** Ground hardware components of a commodity water sprayer: (a) exterior of the atomization system and the water tank; (b) internal design of the atomization system. (c) Example sprayer deployment at critical urban POIs: School, Hospital, Road and Factory. Note that only the multi-nozzle sprinkler of the sprayers are shown.

usually installed high above the ground *e.g.*, at the edge of rooftops, for better dust suppression performance (see Fig. 2-c). iSpray is designed as part of the control module to intelligently switch on and off the sprayer.

Since we aim at cost-effective air pollution control at critical urban POIs, we deploy sprayers at various pollution-sensitive POIs such as schools and hospitals. We also deploy sprayers at locations of representative $PM_{2.5}$ and $PM_{10}$ sources such as factories and roadsides to profile the impact of water spraying on air pollution reduction. 55 sprayers were installed at diverse urban POIs covering an area of 18 km $\times$ 24 km in a metropolis in China. A portable air quality sensing box is also installed in the close vicinity of each sprayer to collect real-time $PM_{2.5}$ and $PM_{10}$ concentrations as well as weather measurements including air temperature, relative humidity, air pressure, wind speed and wind direction. All the data are transmitted via NB-IoT to a central server.

We partition our sprayer deployment into three groups: *Research Area*, *Target Area* and *Control Area* (see Fig. 3-c). The principles of area selection are as follows.

- The *Research Area* covers sprayers with co-located air quality sensing boxes. That is, each site within the *Research Area* consists of a sprayer and an air quality sensing box as shown in Fig. 3-b. There are 55 such pairs of sprayers and air quality sensing boxes in the *Research Area*. This area is used for modeling and testing single-location pollution reduction (Sec. 4) as well as pollution reduction maps (Sec. 5).
- The *Target Area* is a sub-area of the *Research Area* where we would like to control the $PM_{2.5}$ and $PM_{10}$ levels. It covers critical POIs as those shown in Fig. 2-(c). We randomly pick two Target Areas that contain diverse POIs. *Target Area 1* contains 7 pairs of sprayers and sensing boxes and *Target Area 2* contains 8. We mainly use *Target Area 1* to test the sprayer scheduling performance of iSpray in the evaluations (Sec. 7) and the slightly smaller *Target Area 2* to assess the generalization of iSpray (Sec. 7.4).
- The *Control Areas* are used as control groups against the *Target Areas* to evaluate the effectiveness of water spraying. Each site in the *Control Areas* only has an air quality sensing box without a sprayer. We select *Control Areas* with the following criteria. First, the $PM_{2.5}$ and $PM_{10}$ distributions of the *Control Area* should be similar to those in the *Target Areas* when the sprayers are closed. The similarity is measured by the Kullback-Leibler (KL) divergence as in [6]. Second, the *Control Areas* are located at different orientations relative to the *Target Areas*. In total, three *Control Areas* are chosen, with 5,4 and 5 air quality boxes, respectively. We use the average $PM_{2.5}$ and $PM_{10}$ concentrations of the three *Control Areas* as the control group for the *Target Areas*.

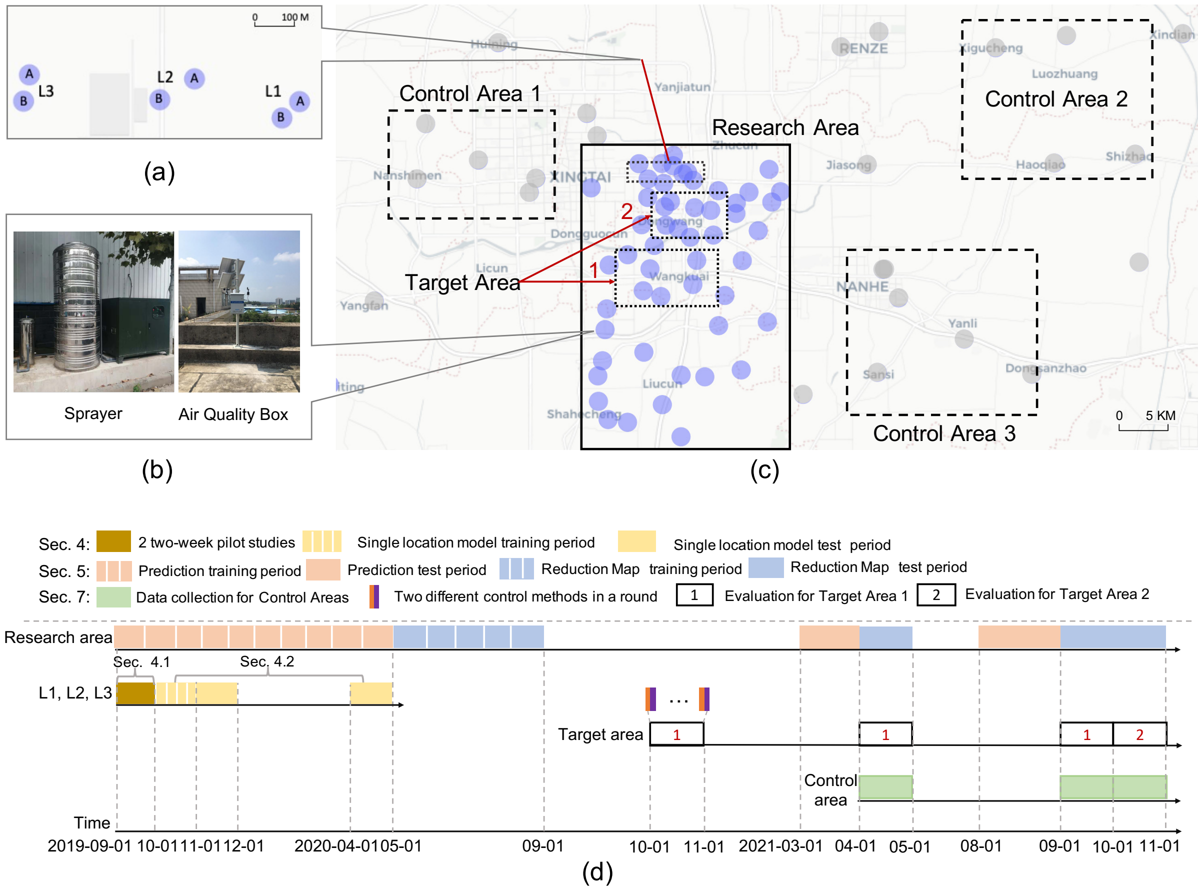



**Fig. 3.** Summary of data collection campaigns. (a) Sprayers deployed at three locations. $L_1$ is factory, $L_2$ is roadside and $L_3$ is residential area. At each location, there are two closely deployed sprayers (also with air quality sensing boxes) $A$ and $B$. We use these sprayers to analyze and model spraying-induced pollution reduction at single locations. (c) Overall deployment. Each site within the Research Area consists of a sprayer and a co-located air quality sensing box as shown in (b). Each site in the Control Area only has an air quality sensing box. The Target Areas are sub-areas of the Research Area for pollution control. (d) Time split of data campaigns.

## 3.2 Data Collection

We collected measurements from the 55 sprayers and their co-located air quality sensing boxes in *Research Area*, including data from 7 air quality sensing boxes from *Target Area 1*, and 8 from *Target Area 2*. Meanwhile, data from the 14 (5 + 4 + 5) air quality sensing boxes data in the three *Control Areas* are also collected. The dataset contains the following data collected spanning from September 1st, 2019 to November 1st, 2021.

- **Air quality and local weather data:** We sample real-time air quality and weather-related readings from the air quality sensing boxes at every minute. The air quality readings include $PM_{2.5}$, and $PM_{10}$. The local weather information includes Air Temperature ($AT$), Relative Humidity ($RH$), Air Pressure ($AP$), Wind Speed ($WS$), and Wind Direction ($WD$) measured at the location of the sensing box. Prior research [5, 9, 27, 45] showed that these factors affect the $PM_{2.5}$ and $PM_{10}$ concentration. The values of all the weather variables are normalized to the range of $[0, 1]$.

- **Sprayer data:** We record sprayer data including sprayer operating status, which is either *on* or *off*, as well as the usage of water and electricity. The sampling rate is also every minute.
- **Forecast weather data:** In addition to the local weather data sampled at each air quality sensing box, we also collect public weather *forecast* data [1] for the entire region of interest. These weather records contain $1km \times 1km$ grid-level air temperature, relative humidity, air pressure, wind speed, and wind direction for every hour. These data will be used in the *air quality map prediction* module (see Sec. 5.1).

Fig. 3-(d) summarizes our data collection campaigns and their usage.

- **Data collection for single-location pollution reduction**. We use data correspond to the three locations $L_1$, $L_2$, and $L_3$ in Fig. 3-(a) for characterizing and modeling air pollution at single locations (see Sec. 4). The selection of these three locations is deferred to Sec. 4.1.1. Its data collection period is from September 1st, 2019 to April 30th, 2020. Specifically, two two-week pilot studies, from September 1st, 2019 to September 15th, 2019, and from September 16th, 2019 to September 30th, 2019, respectively, are adopted to analyze the spraying-induced air pollution reduction at single locations (see Sec. 4.1). Afterwards, we use the data collected from October 2019, as well as from November 2019 (Autumn dataset) and April 2020 (Spring dataset), to train and test our single-location air pollution reduction model (see Sec. 4.2.2).
- **Data collection for pollution reduction map generation**. We use data collected from the *Research Area* for training and testing air pollution map generation (Sec. 5). Specifically, we use the data from September 1st, 2019 to April 30th, 2020 for training the pollution map prediction without spraying (Sec. 5.1) and data in March and August 2021 for testing (Sec. 5.4.1). Similarly, we use the data from May 1st, 2020 to August 31st, 2020 for training the pollution reduction map (Sec. 5.2), and data in April, September, and October 2021 for testing (Sec. 5.4.2).
- **Data collection for sprayer scheduling**. Note that the scheduling algorithm of iSpray does not involve training other than the above models for pollution reduction (see Sec. 6). Therefore, we only need datasets for testing. Specifically, we use the data collected *(i)* in October 2020 from *Target Area 1* to compare different scheduling strategies (see Sec. 7.2); *(ii)* in April 2021 and September 2021 from *Target Area 1* to test the performance of iSpray scheduling (see Sec. 7.3); and *(iii)* in October 2021 from *Target Area 2* to test the generalization of iSpray scheduling (see Sec. 7.4). Meanwhile, we collect the data from the *Control Areas* for the corresponding months as the control group, *i.e.*, without any water spraying.

Note that both $PM_{2.5}$ and $PM_{10}$ are particulate matters and the only difference lies in size of the particle. Also $PM_{2.5}$ is more critical to the human health [41]. Therefore, in the rest of this paper, we will mainly use $PM_{2.5}$ to illustrate our technical details, but provide the evaluations for $PM_{10}$ mainly in Sec. 7.

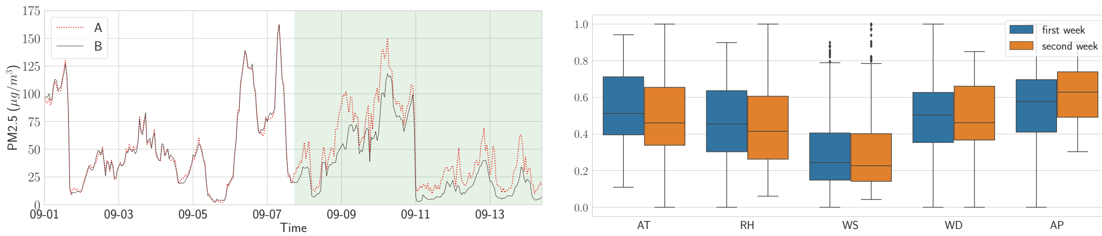
## 4 CHARACTERIZING WATER SPRAYING ON SINGLE-SPOT AIR POLLUTION REDUCTION

In this section, we conduct preliminary studies to answer the following two questions: *(i)* Does water spraying reduce air pollution at single POIs? *(ii)* Can we model the amount of air pollution reduction at a single POI as a function of sprayer time and other environmental factors? We answer these questions with data collected at the three locations in Fig. 3-a.

### 4.1 Water Spraying Suppresses Air Pollution at Single Locations

We first investigate whether water spraying notably decreases air pollution concentrations in the outdoor open air via two field studies.

[1] https://www.ecmwf.int/en/forecasts/datasets visited 2021-11-01




**(a)** **(b)**

**Fig. 4.** Impact of water spraying on hourly averaged $PM_{2.5}$ concentration measured at two closely deployed sprayers ($A$ and $B$) at $L_1$. (a) $PM_{2.5}$ concentrations measured at $A$ and $B$ over time. Both $A$ and $B$ were switched off in the first week and $B$ was switched on in the second week (portions with green background). (b) Distributions of local weather data in the first and the second week.

**Table 2.** Results of t-tests for weather conditions.

**(a)** p-values for weather conditions between the first and second week.

| | AT | AP | RH | WD | WS |
|---|---|---|---|---|---|
| $L_1$ | 0.32 | 0.44 | 0.66 | 0.69 | 0.79 |
| $L_2$ | 0.33 | 0.42 | 0.65 | 0.71 | 0.77 |
| $L_3$ | 0.30 | 0.42 | 0.66 | 0.72 | 0.76 |

**(b)** p-values for weather conditions between sprayer A and B.

| | AT | AP | RH | WD | WS |
|---|---|---|---|---|---|
| $L_1$ | 0.90 | 0.90 | 0.89 | 0.92 | 0.93 |
| $L_2$ | 0.89 | 0.91 | 0.90 | 0.92 | 0.93 |
| $L_3$ | 0.92 | 0.90 | 0.89 | 0.93 | 0.94 |

*4.1.1 Pollution Reduction over Time.* We randomly choose six sprayers (with co-located air quality sensing boxes) from the *Research Area* in Fig. 3-(c) for a two-week field study (from September 1st, 2019 to September 15th, 2019). Specifically, sprayers from three locations are selected, where there are two closely deployed ($<$ 100 meters) sprayers at these three locations (see Fig. 3-(a)). The sprayers labeled as $A$ at each location are used as the control group. That is, they are kept switch off during the entire two weeks. The sprayers labeled as $B$ at each location are switched off in the first week and switched on in the second week. We use the $PM_{2.5}$ and $PM_{10}$ concentrations as well as the local weather data (*i.e.*, air temperature $AT$, relative humidity $RH$, air pressure $AP$, wind speed $WS$, wind direction $WD$) for this study.

Fig. 4-(a) plots the $PM_{2.5}$ concentrations measured at sprayer $A$ and $B$ at location $L_1$ (factory) in these two weeks. We average the minute-resolution $PM_{2.5}$ values into hourly resolution to highlight the general trend over two weeks. In the first week, where both sprayers were switched off, the mean absolute difference in the $PM_{2.5}$ readings of sprayer $A$ and $B$ is within $0.5\mu g/m^3$. In contrast, this difference in $PM_{2.5}$ concentration increases to $13.0\mu g/m^3$ for the second week, where sprayer $A$ remained off while sprayer $B$ was switched on (portions with green background in Fig. 4-(a)). Similar results are observed for sprayer $A$ and $B$ at location $L_2$ and $L_3$. Specifically, for location $L_2$, the mean absolute difference between sprayer $A$ and $B$ is $0.6\mu g/m^3$ in the first week, and $10.7\mu g/m^3$ in the second week. For location $L_3$, the mean absolute difference between sprayer $A$ and $B$ is $0.8\mu g/m^3$ in the first week, and $9.5\mu g/m^3$ in the second week. The significant change in the $PM_{2.5}$ measurements at the two closely deployed sprayers indicates that water spraying notably affects the air pollution.

The difference in $PM_{2.5}$ concentration might be caused by notable changes in environmental conditions in the first and the second week. For example, factors such as *wind* are known to affect the spatiotemporal distribution

**Table 3.** Difference in hourly averaged $PM_{2.5}/PM_{10}$ concentration of sprayer $B$ compared with sprayer $A$ in the second week. $(A \rightarrow B)/reduction$ means the mean value changes from A to B, and the reduction percentage.

| | $L_1$ (Factory) | $L_2$ (Roadside) | $L_3$ (Residential) |
|---|---|---|---|
| $PM_{2.5}$ | $(49.5 \rightarrow 36.5)/-26.2\%$ | $(45.0 \rightarrow 34.3)/-23.6\%$ | $(41.2 \rightarrow 31.8)/-22.9\%$ |
| $PM_{10}$ | $(65.8 \rightarrow 46.7)/-29.1\%$ | $(62.1 \rightarrow 46.3)/-25.4\%$ | $(55.0 \rightarrow 42.3)/-23.1\%$ |

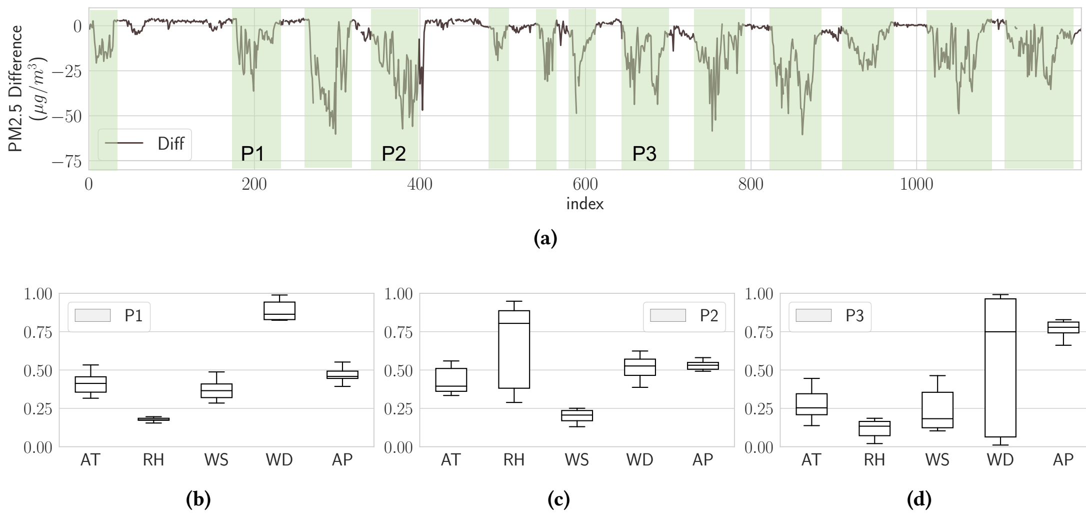



**Fig. 5.** Illustration of weather-dependent pollution reduction: (a) difference of $PM_{2.5}$ between $B$ and $A$ at location $L1$, where $P1$ to $P3$ are three random periods of the same duration when $B$ is switched on; (b)-(d): local weather data during $P1$-$P3$.

of $PM_{2.5}$ concentrations [5, 45]. Fig. 4-(b) plots the distributions of the weather data (*i.e.*, air temperature $AT$, relative humidity $RH$, wind speed $WS$, wind direction $WD$ and air pressure $AP$ in the first and the second week. It is observed that the weather conditions are similar for the first and the second week. It implies that the change in $PM_{2.5}$ concentrations at the two sprayers is mainly due to change in sprayer status *i.e.*, $B$ was switched on in the second week. As a more quantitative measure, we use t-test to assess the difference in weather conditions across both weeks for $L_1, L_2$ and $L_3$. For each weather variable measured at each location, its measurements in the first week and the second week are used as the two independent inputs for the t-test. The hypothesis is that two independent samples have identical average (expected) values and a p-value larger than 0.05 is explained as a positive signal to support the hypothesis. Table 2a shows the p-values for all meteorological variables, which range from 0.30 to 0.79. Therefore, the measurements of weather conditions from these two weeks can be considered drawn from the same distribution, *i.e.*, similar to each other.

Table 3 shows the difference in pollution concentrations measured at sprayer $A$ and $B$ at the three locations in the second week. As is shown, water spraying decreases $PM_{2.5}$ and $PM_{10}$ concentrations by over 20% at representative urban POIs, which is considered remarkable improvements in air pollution control [15].

*4.1.2 Pollution Reduction at Finer Time Granularity.* In this field study, the setups follow those in Sec. 4.1.1, except that instead of keeping the sprayers $B$ at locations $L_1$ to $L_3$ switched on continuously in the second week, we regularly switched these sprayers on and off for a random duration from 15 minutes to 24 hours. The study was

**Table 4.** List of input features for single-location pollution reduction.

| Category | Features |
|---|---|
| Weather | air temperature, air pressure, relative humidity, wind speed, wind direction |
| Pollution Level | We defined 6 discrete PM2.5 levels [9] and use one-hot encoding to represent them |
| POI | We selected 10 common POIs from [17] and use one-hot-encoding to represent them |
| Time Unit | Time unit after the opening of spraying system, one unit means one quarter-hour |
| Date | Hour-of-Day, Day-of-Week, Month-of-year, isHoliday |

conducted from September 16th, 2019 to September 30th, 2019. The local weather conditions can be considered as similar between two nearby locations (*i.e.*, $A$ and $B$ at each location) during these two weeks. As a quantitative measure, we conduct a t-test for all weather variables between $A$ and $B$. The p-values are between 0.89 to 0.94, which are larger than 0.05 (see Table 2b), indicating the weather data at $A$ and $B$ are similar. Thus, the difference of pollution concentrations between sprayer $A$ and $B$ at these locations is primarily due to spraying.

Fig. 5-(a) plots the difference of the $PM_{2.5}$ concentrations (averaged for every 15 minutes) between $B$ and $A$ for these 14 days. The zones colored in green are periods with sprayer $B$ switched on. We make the following observations. *(i)* The $PM_{2.5}$ difference in the uncolored zones is almost zero, meaning the $PM_{2.5}$ concentrations at $A$ and $B$ are almost the same. This is expected because $A$ and $B$ experience similar weather conditions and there is no air pollution reduction by water spraying during these periods. *(ii)* The $PM_{2.5}$ difference in the green zones ranges from $5.5\mu g/m^3$ to $62.0\mu g/m^3$. The air pollution reduction owes to water spraying. However, the amount of reduction varies over time. To understand the reasons for such variations, we investigate the air pollution reduction from three random periods of the same duration (10 hours), $P_1$ to $P_3$ in Fig. 5-(a). Our hypothesis is that the spraying-induced pollution reduction is weather-dependent. Fig. 5-(b) to Fig. 5-(d) show the local normalized weather data during the three periods $P_1$ to $P_3$. The average $PM_{2.5}$ reduction in these three periods are $8.8\mu g/m^3$, $22.9\mu g/m^3$, and $15.4\mu g/m^3$, which notably differ. The local weather data during these three periods also vary. For example, the wind direction of $P1$, $P2$ and $P3$ differs from each other (with mean values of 0.81, 0.49 and 0.62). This indicates pollution propagates to location $L_1$ from different locations during $P1$, $P2$ and $P3$, which might partially explain the difference in spraying-induced pollution reduction. In fact, the heavy precipitation and strong wind in $P2$ facilitates pollution dispersion and increases the pollution reduction rate. The analysis implies that the varied $PM_{2.5}$ reduction in the same time duration at the same location attributes to the difference in *local weather conditions*, as will be shown next.

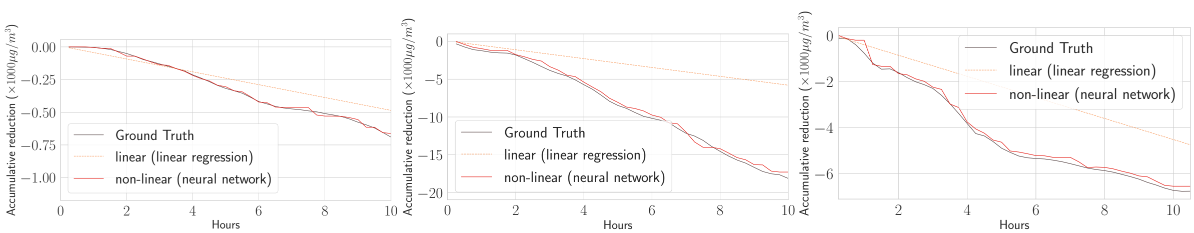
## 4.2 Modeling Air Pollution Reduction at Single Locations

From the field studies in Sec. 4.1, water spraying reduces air pollution at single locations but the reduction varies and is likely *weather-dependent*. In this subsection, we aim to quantify the *accumulative pollution reduction* over time as a function of *weather conditions*. We prefer modeling *accumulative* to *instant* pollution reduction since the accumulative reduction model facilitates decisions on whether to switch off a sprayer after a given period. That is, given a time slot $\Delta t$ at sprayer $s_k$ and all the needed features, the air pollution reduction model will predict the accumulative pollution reduction $Q_{s_k}(\Delta t)$.

*4.2.1 Neural Network Based Pollution Reduction Model.* To model the accumulative pollution reduction as a function of weather conditions, we explore both linear (multi-variant linear regression) and non-linear (neural network) models. Specifically, we feed all the forecast weather data as input features. Additionally, we also include *(i)* pollution levels features, such as $PM_{2.5}$ levels; *(ii)* POI features such as *road, park, factory*; *(iii)* time unit features and *(iv)* date features such as *hour of day*. Previous studies [6, 16] show that these features also benefit air pollution related modeling. Table 4 summarizes all the input features.

**Table 5.** Accuracy of single-location pollution reduction models ($\mu g/m^3$).

| #hours | Model | Autumn | | | Spring | | |
|---|---|---|---|---|---|---|---|
| | | $RM \rightarrow L_1$ | $RM \rightarrow L_2$ | $RM \rightarrow L_3$ | $RM \rightarrow L_1$ | $RM \rightarrow L_2$ | $RM \rightarrow L_3$ |
| 2 | linear (linear regression) | 20.2 | 25.6 | 19.5 | 25.9 | 24.3 | 29.0 |
| | non-linear (neural network) | 3.2 | 3.4 | 3.9 | 4.6 | 4.1 | 3.5 |
| 4 | linear (linear regression) | 28.3 | 32.1 | 26.7 | 34.9 | 33.1 | 38.0 |
| | non-linear (neural network) | 5.9 | 7.1 | 4.3 | 6.1 | 6.5 | 4.8 |
| 6 | linear (linear regression) | 33.2 | 35.6 | 38.1 | 40.3 | 48.8 | 45.6 |
| | non-linear (neural network) | 6.3 | 7.8 | 5.2 | 6.1 | 8.0 | 9.1 |




**(a)** **(b)** **(c)**

**Fig. 6.** Testing single-location pollution reduction models on (a) $P1$ (b) $P2$ and (c) $P3$ in Fig. 5-(a).

*4.2.2 Comparisons of Single-Location Pollution Reduction Models.* We empirically explore whether the non-linear or the linear model is suited for single-location pollution reduction.

**Setups.** We collect data from $L_1$, $L_2$, and $L_3$ during October 2019 for training, and November 2019 and April 2020 for testing. During these periods, the sprayers were set to be switched on when the $PM_{2.5}$ concentration exceeded $35\mu g/m^3$, the excellent air quality level defined in Sec. 7.1; and switched off when the $PM_{2.5}$ concentration dropped below the threshold. We define the data from November 2019 as the Autumn dataset and the one from April 2020 as the Spring dataset. The architecture and hyperparameters of the neural network are automatically optimized using grid-search in Sweeps[2], the final structure used for MLP model is $24(input\ layer) \times 35(first\ hidden\ layer) \times 10(second\ hidden\ layer) \times 1(output\ layer)$ with a dropout rate of 0.2.

**Results.** Table 5 shows the accuracy of single-location pollution reduction models (RM) on predicting next 2 to 6 hours reduction values using the Autumn and Spring test sets. RM is trained using the data from $L_1$ on Autumn dataset and used to test the performance from $L_1$ to $L_3$. Neural network works best for all test sets with MAE errors ranging from 3.2 to 9.1, much less than the results from linear model (MAEs from 19.5 to 48.8). Also, we can find that neural network generalize well in a different season (Spring) and locations ($L_2$ and $L_3$). Those results reveal the necessity of using neural network in modeling the single-location pollution reduction.

As a case study, we also test the above models on data collected from $P1$, $P2$, and $P3$ in Fig. 5-(a). As shown in Fig. 6, the linear model fails to capture the complex relationship between the input features and the accumulative $PM_{2.5}$ reduction (MAE of 59.2), whereas the estimations of the neural network are highly accurate (MAE of 5.5).

[2]https://docs.wandb.ai/guides/sweeps

## 5 SPATIAL MODELING OF WATER SPRAYING ON AIR POLLUTION

In addition to the pollution reduction at single locations, we also need to model how water spraying affects the *spatial* distribution of pollutant concentration so as to schedule the sprayers for effective pollution control at key urban POIs. Specifically, suppose a set of sprayers are switched on at time $t$ and operate for $\Delta t$, we aim to generate an *air pollution reduction map* to depict spraying-induced pollution reduction *in space* at time $t + \Delta t$. In this section, we first present a scheme for air quality map prediction without water spraying (Sec. 5.1), based on which we propose an accurate air pollution reduction map generation (Sec. 5.2) and its parameter learning method (Sec. 5.3). Finally, we present the evaluations for air pollution reduction map generation (Sec. 5.4).

### 5.1 Air Quality Map Prediction Without Water Spraying

Although a new air quality map prediction model (without water spraying) is not our focus, highly accurate predictions are important because they will be used for parameter learning of the pollution reduction model (see Sec. 5.3) and pollution propagation path generation (see Sec. 6.2.1). In response, we adapt a state-of-the-art air quality prediction model [22] which is built upon convolutional long-short-term-memory (convLSTM) modules [33]. Specifically, we add two modifications to improve the prediction accuracy. *(i)* We feed the model with air quality readings from a dense deployment rather than a sparse one to improve the sensor data quality. *(ii)* We design a new weather encoder module to better incorporate the weather influence on air quality changes. Fig. 7 shows our air quality prediction model called *Air-convLSTM*. Assume that air quality map data and weather data are both on grid-level with shape of $(M \times N)$, the historical length and prediction steps are equal as $\tau$, our model consists of the following submodules:

- Air Encoder: it takes the historical air quality map data as input with shape of $(\tau \times M \times N)$ and produces the hidden encoding state $H_{air}$ with shape of $(M \times N \times |H_{air}|)$, where $|H_{air}|$ denotes its hidden dimension.
- Weather Encoder: it inputs the gird-level weather data with shape of $(\tau \times M \times N \times N_{wea})$, where $N_{wea}$ is the weather data dimensions. We use the hidden state of each convLSTM cell as the output of this encoder with the shape of $(\tau \times M \times N \times |H_{wea}|)$, where $|H_{wea}|$ is the hidden dimension of the weather encoder.
- Air Decoder: it takes the air encoder results as input and produces an output with the shape of $(\tau \times M \times N \times |H_{air\_dec}|)$, where $|H_{air\_dec}|$ is its hidden state dimension.
- Weather Fusion: For each decoder step $i, i \in (1 \ldots \tau)$, concatenate the hidden state of air decoder and weather encoder and prepare the input to a fully-connected network (*FCN*). The input dimension is $(M \times N \times |H_{air+wea}|)$, where $|H_{air+wea}| = |H_{air\_dec}| + |H_{wea}|$. The FCN will incorporate the weather influence on air quality changes and produce the adapted values with shape of $(M \times N)$ at each decoding step. The overall dimension of prediction map is $(\tau \times M \times N)$.

### 5.2 Building Air Pollution Reduction Map with Domain Knowledge

Following the conventions in the air pollution map generation literature [8, 16], we discretize the entire 2-dimensional region of interest into grids $\{g\}$. Consider $K$ sprayers deployed in the entire region where a set of sprayers are switched on at time $t$ and will be operating for duration $\Delta t$, our aim is to estimate the reduction $R(g)$ in pollutant concentration for every grid at time $t + \Delta t$. We consider a grid size of $1km \times 1km$ and a time resolution of 1 hour because *(i)* $1km \times 1km$ is widely used in related research [5, 45, 46]; *(ii)* 1 hour is a common time resolution to evaluate the air quality. We use $Q_{s_k}$ to represent the accumulative pollution reduction at a single location over $\Delta t$ hours afterwards.

One may integrate the sprayer data with emission source and weather data to directly learn an air quality map prediction model. We *separately* consider pollution absorption and dispersion due to *limited water spraying data* for effective training. The limited water spraying data also motivate us to model air pollution reduction

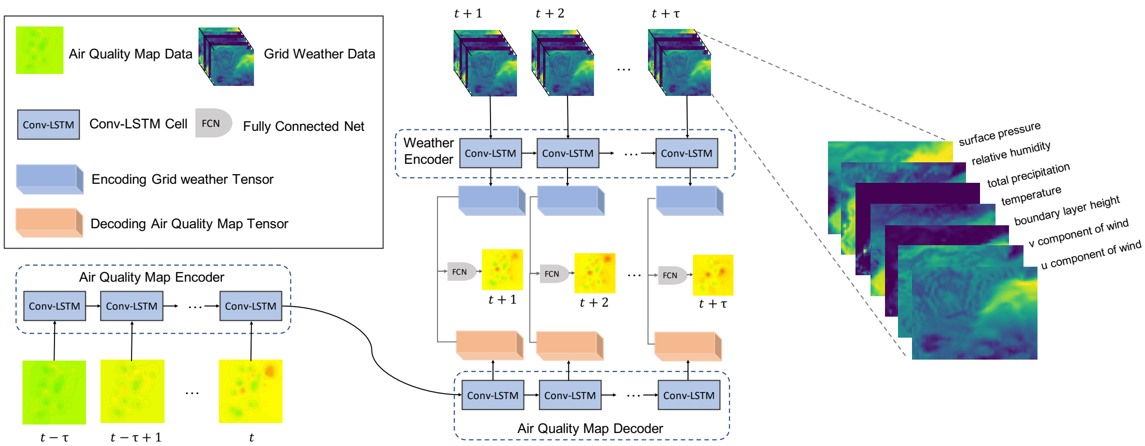



**Fig. 7.** Air quality map prediction model (*Air-convLSTM*) used in iSpray.

maps with *domain knowledge*. We empirically compare the accuracy of our approach with jointly learning of both pollution absorption and dispersion in Sec. 5.4.2.

*5.2.1 Spatial Pollution Reduction of a Single Sprayer.* We first model the pollution reduction $r(g|s_k)$ in grid $g$ due to sprayer $s_k$. The model is inspired by the classical Gaussian plume model to assess the impacts of emission sources on urban air pollution [2, 43]. Specifically, the Gaussian plume model describes the pollution dispersion $c(g|e)$ in grid $g$ (in 2-dimension) due to an emission source $e$ as a Gaussian distribution in vertical directions.

$$c(g|e) = \frac{Q_{air}}{2\pi\sigma\bar{w}} \exp\left(-\frac{1}{2}\left(\frac{d_{cross}(g,e)}{\sigma}\right)^2\right) \tag{1}$$

where $Q_{air}$ is the pollution emission rate, $d_{cross}(g,e)$ is the crosswind distance [3] between $g$ and the grid of $e$. $\bar{w}$ is the average horizontal wind speed, and $\sigma$ is the Gaussian plume dispersion parameter, which is a function of the downwind distance $d_{down}(g,e)$ between $g$ and the grid of $e$ (see footnotes for definition).

Since pollution *absorption*, *i.e.*, pollution reduction due to water spraying in our case, can be considered as the inverse process of *dispersion*, we hypothesize the pollution reduction $r(g|s_k)$ in grid $g$ due to sprayer $s_k$ behaves similar as in Eq. (1). Due to the difficulty to obtain parameters such as $\delta$, we modify the original Gaussian plume model to characterize pollution reduction $r(g|s_k)$ in grid $g$ due to sprayer $s_k$ as follows:

$$r(g|s_k) = \frac{Q_{s_k}}{2\pi\phi(d_{down}(g,g_k), f_{eta})\bar{w}_{s_k}} \exp\left(-\frac{1}{2}\left(\frac{d_{cross}(g,g_k)}{\phi(d_{down}(g,g_k), f_{eta})}\right)^2\right) \tag{2}$$

where $Q_{s_k}$ is the pollution reduction over time period $\Delta t$ in the grid where sprayer $s_k$ is installed, which is modeled as Sec. 4.2, $d_{down}(g,g_k)$ and $d_{cross}(g,g_k)$ are the downwind and crosswind distances between $g$ and the grid $g_k$ of sprayer $s_k$, respectively. $\bar{w}_{s_k}$ is the average horizontal wind speed. $\phi(.)$ is a learnable function to determine $\delta$. The input to $\phi(.)$ is downwind distance $d_{cross}(g,g_k)$ and extra features $f_{eta}$ as described in Table 4. We use a multi-layer perceptron (MLP) to implement the function of $\phi(.)$. The detailed parameter learning procedure is deferred to Sec. 5.3.

[3] Let's make a 2-D coordinate axis with the wind direction as $x$ and the orthogonal one as $y$, which centers at $e$. For grid $g$, the distance between the vertical mapping of $g$ to $x$ axis and $e$ is called the downwind distance, while the distance between vertical mapping of $g$ to $y$ axis and $e$ is called the crosswind distance.

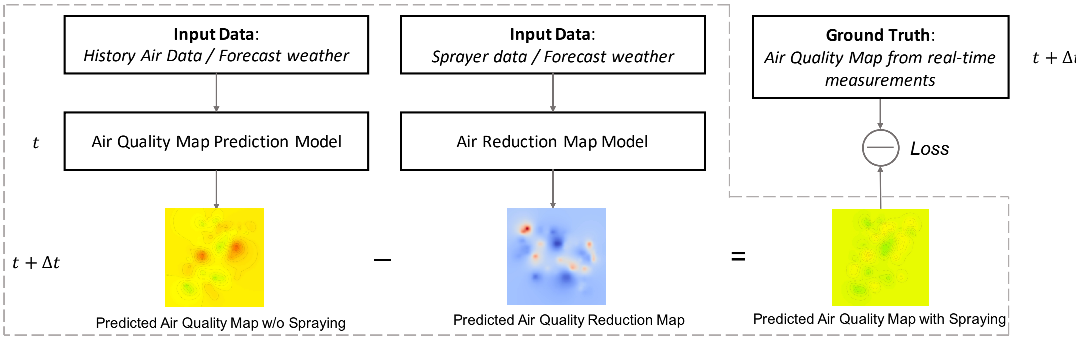



**Fig. 8.** Parameter learning in air pollution reduction map $R(g)$. The air quality map without water spraying $C_d(g)$ is generated via the air quality prediction model in Sec. 5.1. The ground truth air quality map $C(g)$ is generated by interpolating the air quality sensor measurements.

*5.2.2 Spatial Pollution Reduction of Multiple Sprayers.* Consider $K$ sprayers deployed in the entire region of interest. Then the total pollution reduction $R(g)$ in grid $g$ over time period $\Delta t$ is given by:

$$R(g) = \sum_{k=1}^{K} I(s_k)c(g|s_k) \tag{3}$$

where $I(s_k)$ is an indicator function of the operating status $o_k$ of sprayer $s_k$, *i.e.*,

$$I(s_k) = \begin{cases} 1 & \text{if } o_k \text{ is } \textit{on} \\ 0 & \text{if } o_k \text{ is } \textit{off} \end{cases} \tag{4}$$

## 5.3 Parameter Learning for Air Pollution Reduction Maps

Although our air pollution reduction map modeling is built upon domain knowledge, some parameters are still difficult to access, which are captured by the MLP parameters in $\phi(.)$ of Eq. (2). This subsection explains how to learn these MLP parameters (see Fig. 8). Our idea is to first generate the air quality map due to pollution dispersion $C_d(g)$ for $t + \Delta t$ via the air quality prediction model in Sec. 5.1. Then we calculate the air quality reduction map $R(g)$ at $t + \Delta t$ following Eq. (3). The final air quality map at $t + \Delta t$ is calculated as:

$$\hat{C}(g) = C_d(g) - R(g) \tag{5}$$

This map can be compared with the ground truth air quality map $C(g)$ by interpolating the air quality sensor measurements at $t + \Delta t$. The difference between these two maps enables us to update the parameters in $\phi(.)$.

As we will show in Sec. 5.4.1, the air quality map prediction model without spraying influence is accurate for small $\Delta t$. It means the main air quality estimation error comes from $R(g)$, *i.e.*, the MLP parameters we would like to learn. Given the sensor measurements at $t + \Delta t$, we can generate the ground truth air quality map $C(g)$ by Gaussian processes [31]:

$$C(g) \sim \mathcal{GP}(\boldsymbol{m}, \boldsymbol{v}) \tag{6}$$

where $\boldsymbol{m}$ and $\boldsymbol{v}$ are the mean and covariance function, respectively. Gaussian process based interpolation proves highly accurate with a dense pollutant sensor deployment [7, 9], which is the case in our scenario. We define the loss function as the difference between ground truth $C(g)$ and predicted one $\hat{C}(g)$. Using this procedure and loss function, we can successfully learn the parameters in $\phi(.)$ by optimizing and decreasing the loss.

**Table 6.** Accuracy of air quality map prediction (without water spraying) measured by MAE.

| Model | Spring (Mar. 2021) | | Summer (Aug. 2021) | |
|---|---|---|---|---|
| | $PM_{2.5}$ ($\mu g/m^3$) | $PM_{10}$ ($\mu g/m^3$) | $PM_{2.5}$ ($\mu g/m^3$) | $PM_{10}$ ($\mu g/m^3$) |
| Naive | 4.5 | 5.6 | 3.8 | 4.9 |
| ConvLSTM | 3.4 | 4.7 | 3.2 | 4.1 |
| w-ConvLSTM | 1.6 | 2.8 | 1.5 | 2.6 |
| Air-ConvLSTM | **1.2** | **1.5** | **0.9** | **1.4** |

## 5.4 Evaluations on Air Pollution Reduction Map Generation

As mentioned, accurate air pollution reduction map generation is crucial for effective sprayer scheduling. Next, we assess the accuracy of air pollution map generation ignoring and considering water spraying in sequel.

*5.4.1 Effectiveness of Air Quality Map Prediction without Spraying.* We first evaluate the air quality map prediction without considering water spraying.

**Setups.** We compare our *Air-convLSTM* (see Sec. 5.1) with three baselines:

- *Naïve:* use the current timestamp value as the predictions for future hours.
- *ConvLSTM:* use historical air quality map as input and ConvLSTM [33] to predict future maps [22],
- *w-ConvLSTM:* concatenate weather maps with air quality maps, and use ConvLSTM [33] for prediction.

We use the air quality sensor box data of the *Research Area* from September 1st, 2019 to April 30th, 2020 for training of each algorithm. Then we use the data in March, 2021 as the *Spring* test dataset and those from August, 2021 as the *Summer* test dataset. We assess the air quality map prediction accuracy for the next 6 hours since the prediction for the next 6 hours suffice for our scheduling algorithm (see Sec. 6).

**Results.** Table 6 shows the MAEs for air quality map prediction. *Air-ConvLSTM* acquires the best overall prediction accuracy in both test sets. Compared with *w-ConvLSTM*, our method decreases the prediction error of $PM_{2.5}$ and $PM_{10}$ by 25.0% and 46.4% in Spring test period, and 40.0% and 46.2% in Summer test period. More importantly, *Air-ConvLSTM* also successfully predicts the air quality changing patterns. *i.e.*, forecasting the air quality map sudden changing time slots and pollution evolving patterns. This is the key prerequisite for the success of the air pollution propagation path algorithm in Sec. 6.2.1. Fig. 9 shows an example of predicting the pollution sudden change patterns using all methods. We can find that *ConvLSTM fails to predict accurately without using the weather data. w-ConvLSTM* partially solves the problem and improves the performance by including weather features. However, simply concatenating weather data fails to learn the air quality changing patterns, and leads to constant good air quality predictions in all future hours as shown in Fig. 9. Instead, *Air-ConvLSTM* concatenates the weather encoder information with the air quality map predictions, thus predicting the changing patterns from bad to good with precise time slots. This greatly helps for the pollution propagation path detection and thus the overall success of iSpray.

We also evaluate the impact of prediction steps (in hours) and training data length on the accuracy of *Air-ConvLSTM*. As shown in Fig. 10-(a), increasing the prediction steps from 1 to 8, the MAE of *Air-ConvLSTM* increases from 0.6 to 1.8. We choose to predict the next 6 hours in iSpray because the prediction MAE 1.4 is relatively low and 6-hour predictions are also suitable for our scheduling algorithm (see Sec. 6). When increasing the training data length from 2 to 8 months, the prediction MAE of *Air-ConvLSTM* decreases from 2.3 to 1.4 (see Fig. 10-(b)). This is expected because more historical data improves prediction accuracy.

*5.4.2 Effectiveness of Spraying-induced Air Pollution Reduction Map.* Next we verify the effectiveness of our air pollution reduction map for iSpray. The main aim is to show the necessity to separate pollution reduction map generation as in Sec. 5.2.

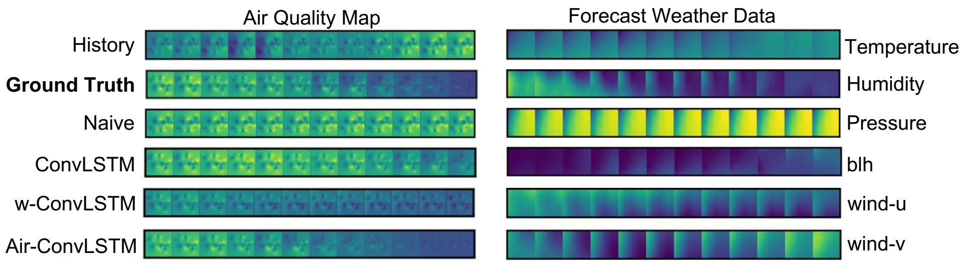



**Fig. 9.** Air quality map prediction case with weather data.

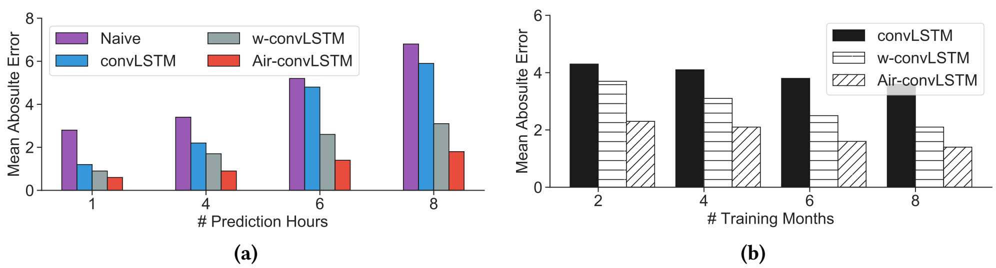



**Fig. 10.** Parameter study of air quality map prediction. (a) Impact of prediction steps; (b) Impact of training data amount.

**Setups.** We compare our method for pollution reduction map generation with the following three baselines:

- *Naïve:* use the current air quality map as the prediction of next step with spraying influence.
- *Land Use Regression:* use Land use regression model [16] and spraying information as input and predict the air quality map at next step.
- *Prediction-based:* concatenate spraying data to the input of FCN module in *Air-ConvLSTM* for prediction.

We use data from the *Research Area* during May 1st, 2020 to August 31st, 2020 for training, and data collected in April 2021 (Spring dataset), and September 1st, 2021 to October 31st, 2021 (Autumn dataset) for testing. The sprayer scheduling strategies for the training and testing periods are as follows. During the training period, we follow the same sprayer scheduling scheme as described in Sec. 4.2.2, *i.e.*, opening the sprayer once the local air quality in above the good air quality threshold. During the testing periods, the sprayer devices are operated by following the schedule timetable produced by iSpray Sec. 6.2.2.

Given the current air quality sensing box measurements and the sprayer status information for the next hour, the task is to predict the air quality map in the next hour. We use MAE to quantify the prediction accuracy of each algorithm.

**Results.** Table 7 shows the overall results. Land use regression model and Prediction-based model fail to generate accurate air quality map with the spraying influence. By decomposing the problem into air quality map prediction without spraying influence and spraying-based air quality modeling, iSpray successfully learns the unknown parameters in the model and make accurate air quality map with a small amount of data. iSpray achieves MAEs of 1.9 and 2.6 for Spring and Autumn datasets, a significant improvement over all the baselines.

Fig. 11 shows the example maps generated by different methods. iSpray benefits from the air pollution reduction model and generates the most accurate map with spraying influence, which is an essential component for our scheduling algorithm.

**Table 7.** Air pollution reduction map accuracy comparison for $PM_{2.5}$ ($\mu g/m^3$) measured by MAE.

| Dataset | Naïve | Land use regression | Prediction-based | Reduction Map |
|---|---|---|---|---|
| Spring | 16.4 | 14.3 | 9.8 | **1.9** |
| Autumn | 19.2 | 15.6 | 11.9 | **2.6** |

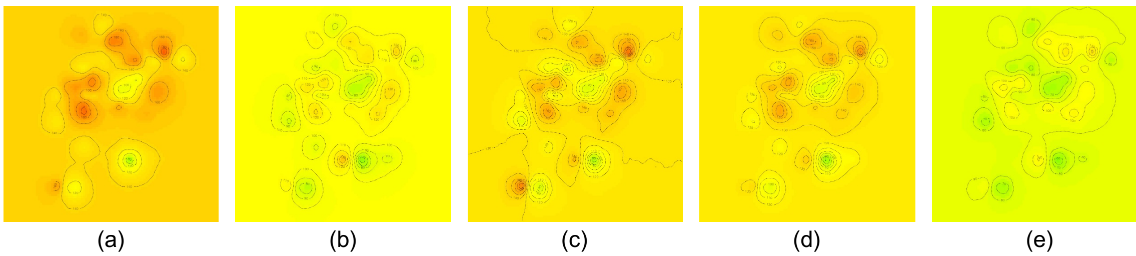



**Fig. 11.** Air pollution reduction map accuracy comparison case study: (a) air quality map at $t$, which is also the output of Naïve; (b) ground truth air quality map at $t + 1$; (c) Land use regression baseline; (d) Prediction-based baseline; (e) iSpray.

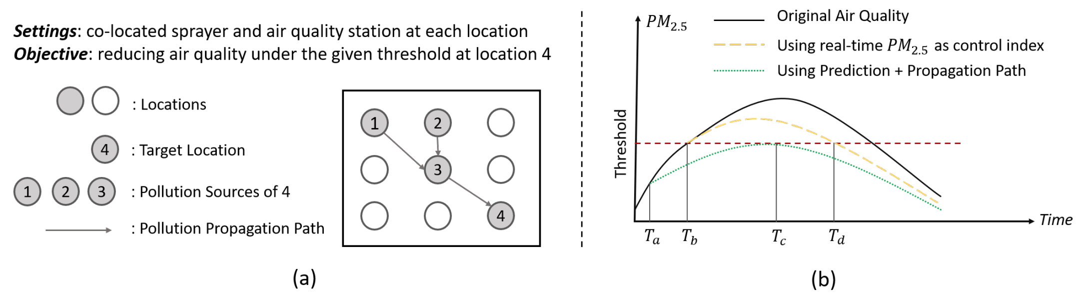



**Fig. 12.** Example of cost-effective scheduling intuition: (a) an example setting; (b) scheduling based on real-time $PM_{2.5}$ values fails, while switching on the sprayer (and those along the pollution propagation path) in advance may keep the $PM_{2.5}$ concentration within the threshold with less total switch-on time.

## 6 COST-EFFECTIVE SPRAYER SCHEDULING

The air pollution reduction map (Sec. 5) enables us to quantify the impact of switching on each sprayer on the spatial distribution of pollutant concentrations. We now present our cost-effective sprayer scheduling scheme to the control the air pollution at key POIs with minimal number of operating sprayers.

### 6.1 Feasibility of Cost-Effective Sprayer Scheduling

Given certain POIs in the region of interest, we aim to make a spraying schedule for the next $\tau$ hours (*i.e.*, whether each sprayer should be switched on or off in each hour) such that *(i)* the pollution concentrations in the grids where the POIs reside are within a given threshold and *(ii)* the total switch-on hours are minimized. Minimizing the total switch-on hours is necessary because a single sprayer consumes 120 $kWh$ electricity and 14.4 $m^3$ water if operating non-stop in a day. We explain the intuitions for cost-effective scheduling via an example below.

Fig. 12-(a) shows a simplified setting of our problem, where the space is partitioned into 9 grids and there is a sprayer and a co-located air quality sensing box in each grid. Our goal is to decide the scheduling timetable for

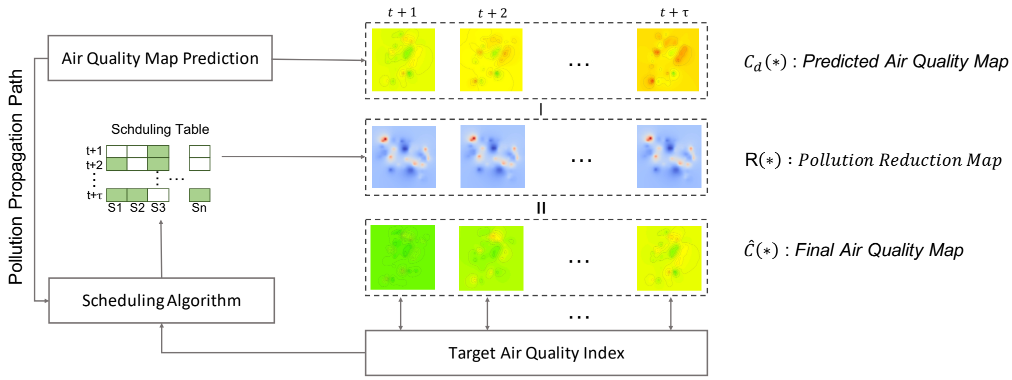



**Fig. 13.** Overview of our cost-effective sprayer scheduling scheme.

all the 9 sprayers such that the $PM_{2.5}$ concentration in the target grid *i.e.*, grid 4 in Fig. 12-(a) is under a given threshold, which is shown by the red dotted line in Fig. 12-(b). One scheduling strategy is to decide whether to switch on a sprayer according to the real-time $PM_{2.5}$ measurements at the co-located air quality sensing box. Suppose the real-time $PM_{2.5}$ concentration in grid 4 exceeds the threshold at time $T_b$. This strategy will then switch on the sprayer in grid 4 at time $T_b$ till $T_d$, when the real-time $PM_{2.5}$ concentration falls below the threshold, as shown by the yellow dotted line in Fig. 12-(b). Since the pollution reduction is not instant, this method will fail to keep the $PM_{2.5}$ concentration within the threshold during $T_b$ to $T_d$. Our solution is to switch on the sprayer in grid 4 *in advance* as well as the a set of sprayers *along the pollution propagation path* towards grid 4. That is, we switch on the sprayer in grid 4 at time $T_a$, when the $PM_{2.5}$ concentration is still within the threshold. This way, the peak $PM_{2.5}$ concentration at $T_c$ will be under the threshold, as shown by the green dotted line in Fig. 12-(b). Note that the switch-on time of the sprayer in grid 4 can be short if certain sprayers along the propagation path (*i.e.*, grid 1, 2 and 3, where the arrow denotes the direction of pollution propagation) have been switched on before the pollution propagates to grid 4. The example implies the following:

- Switching on sprayers based on real-time pollution concentration fails to keep the pollution at target POIs under control due to delays in spraying-induced pollution reduction. Therefore, it is important to predict the future pollution concentration and switch on the sprayers in advance.
- Switching on sprayers along the pollution propagation path holds promise to keep the pollution at target POIs under control and reduce the total switch-on time by suppressing the pollution near the source.

## 6.2 Scheduling Method

Inspired by the motivation example in Sec. 6.1, Fig. 13 illustrates our cost-effective sprayer scheduling scheme. It first predicts the air quality maps for the next $\tau$ hours without water spraying (Sec. 5.1) and then generate the pollution propagation path towards the target grids (Sec. 6.2.1). Then we take the predictions, propagation path and the given threshold to make a scheduling timetable to guarantee the air quality in the target grids with small total sprayer switch-on time (Sec. 6.2.2). We give the implementation details below.

*6.2.1 Deriving Pollution Propagation Path.* We identify pollution propagation paths by adapting the method in [23]. The key observation is that if the uptrend interval (pollution propagation) of grid $a$ is ahead of $b$, then $a$ is considered a causal parent node of $b$. Their method builds causal graphs and finds the *top-k* patterns from all generated graphs using historical data. These patterns are the *statistically* frequent pollution propagation behaviors. For our case, however, we should identify the predicted pollution propagation patterns *at the current*

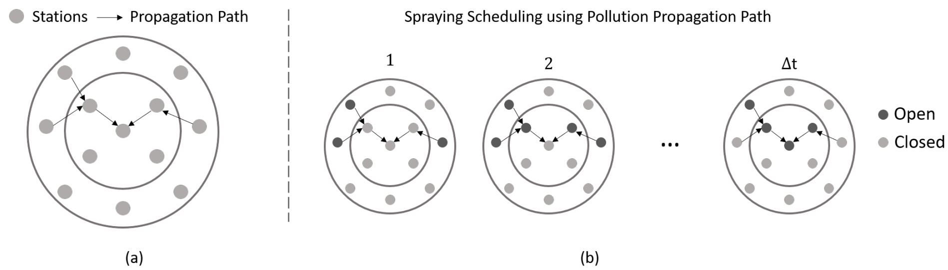



**Fig. 14.** (a) Air pollution propagation path to one target location. (b) one example of spraying scheduling along the air pollution propagation path.

---

**Algorithm 1:** Deriving propagation paths

**Input:** Predicted air quality readings at all locations for next $\Delta t$ hours, pollution influence circles list $PC$, target locations $L$, distance threshold $d$

**Output:** Air pollution propagation paths $Path$ to target locations $L$

**for** *each target location* $L_i$ *in* $L$ **do**
- $Path_i = [L_i]$ ; `// initialize the propagation path for location` $L_i$
- Conduct pollution influence circles ($C_j, j \in (1 \ldots l)$, where smaller index represents small diameter) centered in target location $L_i$ using the given values in $PC$ (black circles in Fig. 14);
- **for** $j \in (1 \ldots l)$ **do**
  - Find all stations $S_c$ located in $C_j$ apart from stations in $Path_i$;
  - **for** *for each station* $x$ *in* $S_c$ *and each* $y$ *in* $Path_i$ **do**
    - **if** *(the distance between* $(x, y) \leq d$*) and (max value of* $x \geq$ *max value of* $y$*) and the uptrend interval of* $x$ *is ahead of* $y$ **then**
      - add station $x$ to $Path_i$

Concatenate all propagation path $Path_i$ and get the overall propagation path $Path$

**return** $Path$

---

*timestamp*. Given the spatiotemporal air quality data, we aim to estimate the pollution propagation paths. Algorithm 1 illustrates the entire process.

- Firstly, we characterize the causal parent nodes of each grid as [23]. However, we add two more constraints: *(i)* the maximum values of parent nodes should be larger than the ones of child node; and *(ii)* the distance between them should be smaller than some threshold. These two constraints facilitate finding the pollution propagation path for a single timestamp.
- Secondly, we introduce the *pollution influence circles* to iteratively find the causal parent nodes of target grid from inner circles to outside ones (see the circles centered at the target grid in Fig. 14-(a)).
- Finally, we apply the above two steps for each target grid, and then we can derive the propagation path, as the arrows shown in Fig. 14-(a).

*6.2.2 Putting it Together.* Algorithm 2 illustrates our sprayer scheduling algorithm, which generates a timetable for all the sprayers in the next $\tau$ hours. The algorithm works as follows.

- For each target location, we first compute the highest prediction pollution concentration in next $\tau$ hours. If the peak concentration exceeds the threshold, the propagation path estimation module is called to identify the propagation path to this target location, *e.g.*, the arrows in Fig. 14-(a).

**Algorithm 2:** Scheduling algorithm

**Input:** Predicted air quality map in next $\tau$ hours; target locations $L$; threshold value $V_{thres}$; air quality map prediction model and air pollution reduction map model

**Output:** Schedule timetable $T_{skd}$ for all sprayer systems in next $\tau$ hours

1 For all target locations $L$, find the highest predicted air quality readings $V_{L_i}$ at location $L_i$ in next $\tau$ hours, assume the timestamp is $\Delta t$

2 **if** $V_{L_i} > V_{thres}$ **then**

3     Conduct the pollution propagation path for target location $L_i$ using Algorithm 1

4     **for** $t' \in (1 \ldots \Delta t)$ **do**

5         Set sprayer to open if the local predicted $PM_{2.5}$ values are greater than $V_{thres}$

6         Conduct air pollution reduction map for $t'$ using the method in Sec. 5

7         Update air quality map predictions after the current timestamp `// update predictions after each scheduling step to incorporate the influence of spraying`

8     Get the scheduling timetable $T^i_{skd}$ for location $L_i$

9 Go to step **1** until convergence or all sprayer systems have been scheduled.

10 Concatenate all scheduling timetable $T^i_{skd}$ and get the overall scheduling timetable $T_{skd}$

11 **return** $T_{skd}$

- Given the pollution propagation path in previous step, we greedily decide the sprayer status. We start with the first time slot and set the sprayers along the propagation path as open if the concentrations in the grids of these sprayers are above the threshold. After each time slot, we update the future air quality predictions to incorporate the influence of spraying. We continue the process till the predicted pollution concentrations are below the threshold or all sprayers are switched on. One example scheduling timetable is shown in Fig. 14-(b).
- We apply the same pipeline to all target locations until their predicted peak concentrations are below the threshold or all sprayers are used.

*6.2.3 Visualization of iSpray Scheduling.* We use one real water spraying control case to illustrate the effectiveness of the scheduling algorithm in iSpray, *i.e.*, Algorithm 2. Assume the current timestamp is $t$ and we try to decide the scheduling timetable for next $\tau$ hours (in our case, $\tau = 6$) to suppress the air pollution in the target area. Following the scheduling algorithm in Algorithm 2, iSpray works as follows to produce the scheduling timetable for all spraying systems.

- iSpray first predicts the air quality maps for the next 6 hours as shown in first row of Fig. 15. The highest prediction values for target area is in $t + 4$, so $\Delta t = 4$.
- Using the propagation path found using Algorithm 1, iSpray schedules the sprayers from the sources to target area step by step, and generates the new air quality map with spraying influence. iSpray only schedules those sprayers along the propagation path instead of all where the predicted concentrations are above the threshold.
- After deciding the sprayers to switch on in the next 4 hours, iSpray generates new air quality maps with spraying influence (third row in Fig. 15). The air quality readings in the target area are now below the threshold, so iSpray terminates.

We can see that iSpray only schedules the necessary sprayers along the propagation path which affect the *Target Area*. Therefore our sprayer scheduling is cost-effective. The difference between the predicted air quality map and the ground truth one (generated by Gaussian process interpolation, see Sec. 5.3) is small (see Sec. 5.4.1 and Sec. 5.4.2 for quantitative results), which also validating the effectiveness of iSpray.

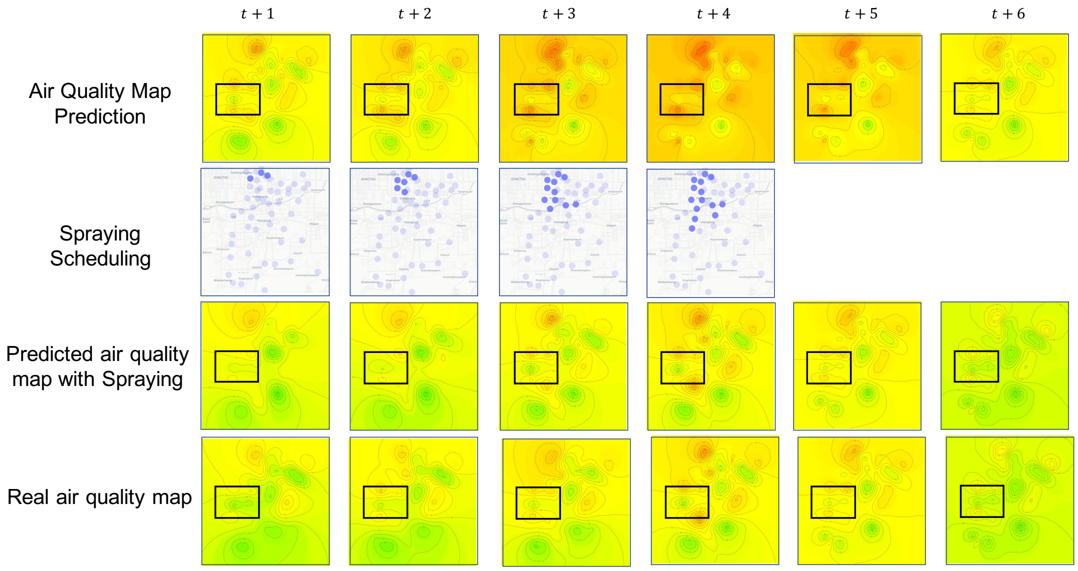



**Fig. 15.** One real water spraying control use case using iSpray. Black box represents the *Target Area*.

## 7 EVALUATION OF ISPRAY SCHEDULING

This section evaluates the scheduling of iSpray and discuss its limitations and extensions.

### 7.1 Overall Experiment Setups

Since it is impossible to test sprayer scheduling schemes simultaneously at the same location, we test sprayer scheduling in the *Target Areas* and use the *Control Areas* as the control group without spraying to derive quantitative performance metrics. Note that the scheduling algorithm involves training expect those for air quality prediction and pollution reduction map generation. In the following evaluations, these models are trained using the datasets in Sec. 5. We only explain the detailed setups for the testing datasets in each experiment below.

We use water usage ($m^3$) and electricity usage ($kWh$) to compare the cost-effectiveness of different sprayer scheduling methods. We use the mean of real-time and 24-hour average value of $PM_{2.5}$ and $PM_{10}$ ($\mu g/m^3$), as well as the excellent quality rate of $PM_{2.5}$ and $PM_{10}$ ($PM_{2.5} \leq 35\mu g/m^3$ according to China Strandard [4], $PM_{10} \leq 40\mu g/m^3$ as adopted in the paper) to assess the air quality, which are also used as the threshold in our model. One sprayer consumes 5 $kW$ electricity 0.6 $m^3$ water per hour. We consider a scheduling resolution of an hour, as in Sec. 6.1. For all the evaluations below, we use the average performance of the three *Control Areas* shown in Fig. 3-(c) to mitigate the impact of relative orientations to the *Target Areas*.

### 7.2 Performance of Different Scheduling Algorithms

We mainly compare iSpray with the baseline method that controls sprayers based on the real-time pollution concentrations measured at the co-located sensing box, which is denoted as *Real-Time-Values* afterwards.

**Setups.** We test these two sprayer scheduling schemes in October 2020 in *Target Area 1*. For fair comparison, we choose the first 30 days and split them into 15 pairs. In each pair of two days, we randomly choose iSpray or

[4]China National Standard: https://healthandsafetyinshanghai.com/china-air-quality/

**Table 8.** Performance comparison between different scheduling methods.

| Method | Water ($m^3$) | Electricity ($kWh$) | Real-time $PM_{2.5}$/ $PM_{10}$ ($\mu g/m^3$) | 24-hour average $PM_{2.5}$/ $PM_{10}$ ($\mu g/m^3$) | Excellent Quality Rate $PM_{2.5}$/ $PM_{10}$ |
|---|---|---|---|---|---|
| *Real-time-values (15 days)* | 5,108 | 42,571 | 40 / 46 | 38/44 | 72% / 75% |
| *iSpray (15 days)* | 3,326 | 24,309 | 34 / 36 | 33/36 | 85% / 91% |
| *iSpray over Real-time-values* | -34.8% | -42.3% | -15.0% / -21.7% | -13.2% / -18.2% | +13% / +16% |

the baseline for scheduling in the first day and the other for the second day. Therefore we have 15 test rounds in total, as shown in Fig. 3-(d). We report the average performance of these 15 test rounds.

**Results.** Table 8 summarizes the performance of iSpray and the *Real-time-values* baseline. If all the sprayers are operating non-stop for 15 days, the water and electricity usage are $11,880\,m^3$ ($0.6*24*55*15$) and $99,000\,kWh$ ($5*24*55*15$), respectively. Both scheduling schemes notably reduce the usage of water and electricity, where our iSpray requires only $3,326\,m^3$ water and $24,309\,kWh$ electricity, which reduces the water and electricity usage by 34.8% and 42.3% compared with the *Real-time-values* baseline. Meanwhile, the mean values of real-time $PM_{2.5}$ and $PM_{10}$ decrease by $6\mu g/m^3$ and $10\mu g/m^3$, which accounts for 15.0% and 21.7%. The mean values of 24-hour average $PM_{2.5}$ and $PM_{10}$ also decrease by $5\mu g/m^3$ and $8\mu g/m^3$, which accounts for 13.2% and 18.2%. The excellent quality rate of $PM_{2.5}$ and $PM_{10}$ increase by 13% and 16%.

## 7.3 Performance With and Without iSpray Scheduling

This experiment quantifies the air pollution reduction due to iSpray. Since it is difficult to directly measure the air quality with and without water spraying at the same location and time, we adopt the distribution similarity concept [6] for indirect comparison. Specifically, it is observed that the air quality distributions of different regions within a city are similar in the same time period [6]. This allows us to assess the impact of water spraying for the same time period by comparing with the *Control Areas*.

**Setups.** We select March to April 2021 for testing in Spring, and August to September 2021 for testing in Autumn. Specifically, we switch off all the sprayers in *Target Area 1* in March 2021 and August 2021, and schedule the sprayers by iSpray in *Target Area 1* in April 2021 and September 2021. We use the air quality data during the same months from the three *Control Areas* as the control group (by averaging across the three *Control Areas*.

**Results.** We first show that the distribution similarity proposed in [6] holds for our deployment. Specifically, we plot the air quality distributions of *Target Area 1* and the *Control Areas* in March 2021 and August 2021, when all sprayers were switched off. As shown in Fig. 16-(a),(e) and Fig. 16-(c),(g), the two distributions of the target and the control groups are similar. Therefore, the differences in air quality distributions during the same time period are mainly due to water spraying. This is shown in Fig. 16-(b),(f) and Fig. 16-(d),(h), where the only difference is that the sprayers in *Target Area 1* were switched on by iSpray. We can observe notable air pollution reduction.

More quantitatively, we use a 5-number-summary (*min,1st quartile, median, 3rd quartile and max*) to compare the distribution difference between the target area and the control area in April 2021 and September 2021. Table 9 summarizes the differences. According to Table 9, in April 2021, the $PM_{2.5}$ median and 3rd quartile in *Target Area 1* are 27 and 33, which are reduced by 21.2% and 29.8% compared with those in the *Control Areas*. The $PM_{10}$ median and 3rd quartile of *Target Area 1* are 46 and 65, a reduction of 39.5% and 41.2% compared with those in the *Control Areas*. Similarly, in September 2021, the $PM_{2.5}$ median and 3rd quartile in *Target Area 1* are 25 and 39, which are reduced by 24.2% and 36.1% compared with those in the *Control Areas*. The $PM_{10}$ median and 3rd quartile of *Target Area 1* are 58 and 81, yielding a reduction of 23.7% and 37.2% compared with those in the *Control Areas*.

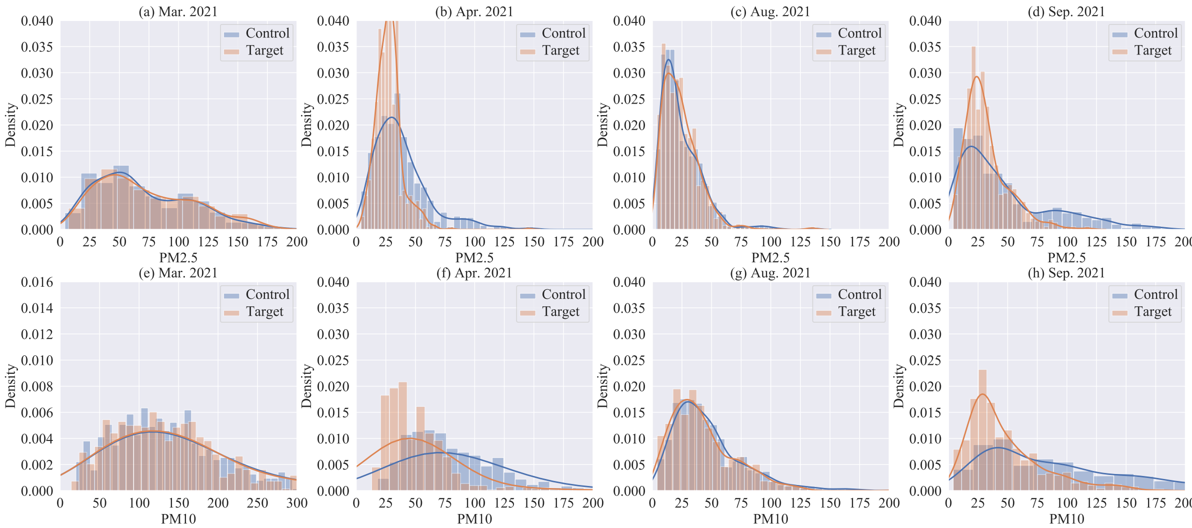


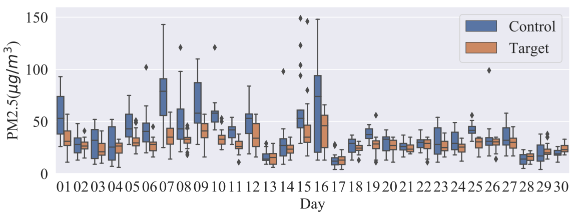

**Fig. 16.** Visualization of air quality ($PM_{2.5}/PM_{10}$) distributions of the *Control Areas* and *Target Area 1*.

**Table 9.** Comparison between the $PM_{2.5}/PM_{10}$ distributions of *Control Areas* and *Target Area 1*.

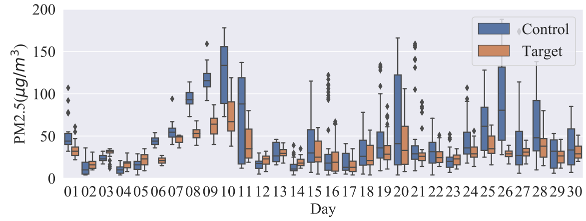
| Time | Region | iSpray Control | min | 1st quartile | median | 3rd quartile | max |
|---|---|---|---|---|---|---|---|
| March 2021 | *Control Areas* | - | 4/20 | 39/86 | 60/135 | 100/193 | 194/371 |
| | *Target Area 1* | OFF | 7/18 | 41/87 | 63/133 | 103/187 | 186/366 |
| April 2021 | *Control Areas* | - | 4/18 | 22/52 | 33/76 | 47/111 | 267/455 |
| | *Target Area 1* | ON | 4/13 | 20/33 | 27/46 | 33/65 | 146/379 |
| Aug. 2021 | *Control Areas* | - | 3/4 | 12/26 | 19/40 | 33/59 | 138/201 |
| | *Target Area 1* | OFF | 4/5 | 14/22 | 21/36 | 32/56 | 137/215 |
| Sep. 2021 | *Control Areas* | - | 4/5 | 17/41 | 33/76 | 61/129 | 188/326 |
| | *Target Area 1* | ON | 4/4 | 20/31 | 25/58 | 39/81 | 119/265 |

To clearly illustrate the pollution reduction on a daily basis, we plot the daily $PM_{2.5}$ box-plots of *Target Area 1* and the *Control Areas* for April and September, 2021 in Fig. 17-(a),(b). We have the following observations. *(i)* The $PM_{2.5}$ distributions of *Target Area 1* and the *Control Areas* are similar if the $PM_{2.5}$ concentrations are low. This is because iSpray will switch off the sprayers when the pollution level is low. *(ii)* The median and max values are significantly reduced during high $PM_{2.5}$ period, indicating that iSpray switches on the sprayers to suppress pollution during these times. The observations also hold for $PM_{10}$, as shown in Fig. 17-(c),(d).

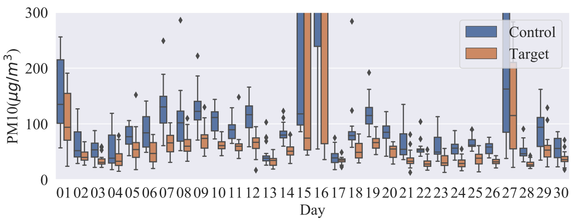
To further analyze the performance of iSpray in reducing the air pollution, especially its effectiveness in dropping the pollution from a polluted level to an excellent one, the total number of days above the excellent air quality level is calculated by comparing the 24-hour average value. The results in Table 10 show that the total number of pollutant days in April are 14 and 28 for $PM_{2.5}$ and $PM_{10}$ in Control Area, and they are reduced to 3 and 17 days by applying iSpray in Target Area 1, which accounts for a reduction of 79% and 39%, respectively. Similarly, a reduction of 53% and 41% can also be found for $PM_{2.5}$ and $PM_{10}$ in the September dataset.



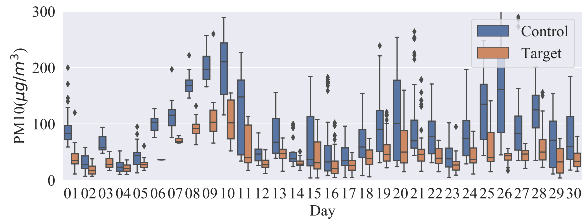

**(a)** $PM_{2.5}$, April




**(b)** $PM_{2.5}$, September




**(c)** $PM_{10}$, April




**(d)** $PM_{10}$, September

**Fig. 17.** Daily pollution control results in Target Area 1 compared with Control Areas in 2021.

**Table 10.** Total days above excellent air quality level in Target Area 1 compared with Control Areas in 2021.

| Pollution Type | April | | September | |
|---|---|---|---|---|
| | Control Area | Target Area | Control Area | Target Area |
| PM2.5 ($> 35\mu g/m^3$) | 14 | 3 | 19 | 9 |
| PM10 ($> 40\mu g/m^3$) | 28 | 17 | 29 | 17 |

## 7.4 Performance in Different Target Areas

This experiment demonstrates the generality of iSpray.

**Setups.** We select a different target area as shown in Fig. 3-(c), denoted as *Target Area 2* to test our iSpray scheduling algorithm. The test took place in October 2021. As with Sec. 7.3, we use the air quality data of the *Control Areas* from the same period as the control group.

**Results.** Fig. 18-(a) plots the pollution concentration distribution in *Target Area 2* and the *Control Areas*. We observe notable reduction of high pollution concentrations in the *Target Area 2*. Quantitatively, the $PM_{2.5}$ median and 3rd quartile in *Target Area 2* are 30 and 45, which are reduced by 23.1% and 35.7% compared with those in the *Control Areas*. The $PM_{10}$ median and 3rd quartile of *Target Area 2* are 51 and 75, resulting in a reduction of 32.6% and 37.5% compared with those in the *Control Areas*. Fig. 18-(b) further illustrates the daily air pollution distributions. We observe the same patterns for *Target Area 1*. The days above the excellent air quality level for Control Area are 16 and 26 for $PM_{2.5}$ and $PM_{10}$, and iSpray reduces them to 8 and 16 days for Target Area 2, yielding a reduction of 50% and 31%, respectively.

In summary, the extent of pollution reduction by iSpray is similar in *Target Area 2* as in *Target Area 1*, validating the generality of our method.

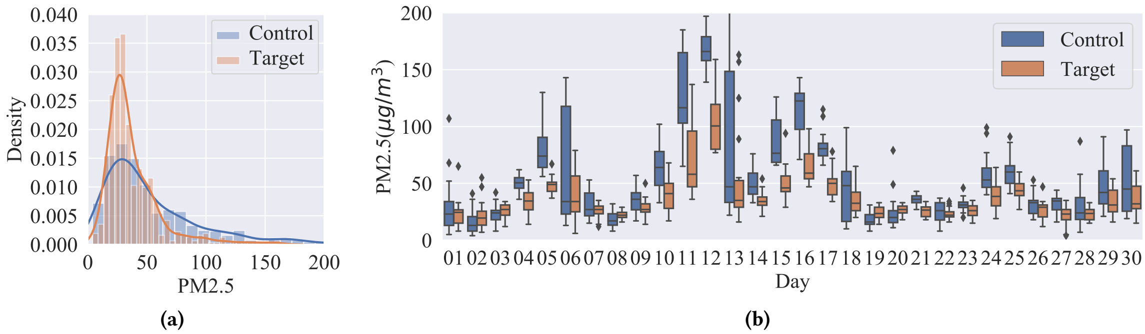



**Fig. 18.** (a) $PM_{2.5}$ and (c) $PM_{10}$ distributions of the *Control Areas* and *Target Area 2*; Daily (b) $PM_{2.5}$ and (d) $PM_{10}$ control results comparison between *Target Area 2* and the *Control Areas*.

## 7.5 Discussions

We briefly discuss the hyperparameter selection in iSpray and the potential extensions to mobile deployments.

*7.5.1 Hyperparameters in iSpray.* We set the threshold values ($V_{thres}$ in Algorithm 2) of $PM_{2.5}$ and $PM_{10}$ to $35\mu g/m^3$ and $40\mu g/m^3$, respectively, which are also the excellent air quality threshold in China, where our system is deployed. In practice, reducing the threshold values $V_{thres}$ to near zero tends to keep all the sprayer switched on in Algorithm 2, leading to non-stop water spraying. Increasing the $V_{thres}$ to higher values will decrease the water sprayer usage time. For our evaluation, we aim to control the pollution level according to the local standards. The spray schedule timetable of the proposed approach depends on the selected threshold, and iSpray aims to clean the local air quality to a level below the threshold. When it is impossible to achieve the local standard, all available sprayers will be open. In this case, the mobile sprayer solutions or hybrid one may help to further suppress the pollution level, as described in Sec. 7.5.2.

*7.5.2 Extensions to Mobile Sprayer Deployments.* In addition to the static deployments like iSpray, there is also extensive research interest in exploiting mobile sensor deployments for air pollution monitoring [13, 14, 16, 18, 27, 40]. We can also extend our scheduling algorithms to mobile settings *e.g.*, with water sprayers mounted on trucks as follows. *(i)* Derive the pollution propagation paths in the next few hours as Algorithm 1. *(ii)* Revise the schedule algorithm in Algorithm 2 with two more considerations: limited water storage and the travel time of the mobile sprayer to specific locations. One solution is to decide locations where sprayers are needed in each time slot using Algorithm 2, and then adapt existing route planning algorithms for spatial crowdsourcing [34, 37, 44] to dispatch mobile sprayers to these locations at the targeting time slots while satisfying the water storage constraints.

A further extension is a hybrid mobile and static water spraying system where mobile sprayers act as backups when pollution control with static water spraying fails (line 9 in Algorithm 2). In this case, mobile sprayers can be scheduled to further suppress the pollution level.

## 8 RELATED WORK

Our work is related to the following two categories of research.

### 8.1 Ubiquitous Urban Air Pollution Sensing and Inference

The availability of portable sensors and urban data has enabled ubiquitous urban air pollution monitoring and inference services. Installed at hot spots [7, 9, 32], vehicles [16, 18, 27, 40] or carried by citizens [28, 35], the low-cost gas and dust sensors provide real-time and fine-grained measurements to analyze urban airborne pollutant concentrations. With measurements collected from a large-scale deployment, accurate air quality map can be generated via spatial interpolation such as Gaussian process [7, 9, 18]. Access to air quality related urban data such as meteorological conditions, traffic flows, emission sources has enabled accurate air quality map generation with sparse sensor deployments by designing dedicated inference models such as spatiotemporal co-training [45], weather-aware auto-encoder [27], etc. Integrating sensor measurements with urban data also facilitates analytics beyond map generation. Examples include simultaneous air quality estimation and prediction [5, 10], pollution propagation pattern discovery [23], and sensor calibration function transfer [6].

Our work is also built upon the fusion of sensor and urban data. However, it differs from existing literature on ubiquitous air quality inference in two aspects. *(i)* Prior urban air pollution map generation proposals [7, 9, 18, 27, 45] mainly model the *dispersion* of airborne pollutants from emission sources. We characterize and model the *absorption* of pollutants due to water spraying, which integrates pollution control measures into accurate air quality map generation. *(ii)* Previous air quality analytics services only offer passive monitoring of pollution to raise awareness [9, 16, 23, 27, 45]. We propose a cost-effective spraying scheduling strategy to keep air pollution at critical POIs under control.

### 8.2 Water Spraying Systems for Dust Control

Water spraying is widely adopted for dust control in factories and mines [20, 21, 36] and its usage has recently been extended for particulate matter control in urban areas [12, 25, 42]. Del Corno *et al.* [12] carry out an experiment of removing aerosols with the help of high-pressure water spray nozzle as they generate water droplets that are smaller in size compared to those from regular, low-pressure nozzles. The experiment was conducted in a transparent glass chamber of size $0.5m \times 1m \times 1.5m$, equipped with a high-pressure spray nozzle system. Yu *et al.* [42] proposes a geoengineering scheme to reduce air pollution in the cities of China with water spray technology. The indoor experiment results show that the $PM_{2.5}$ concentration can be reduced significantly, the extent of which depends on the scavenging coefficients. However, the authors did not evaluate the spraying system in outdoor environment. Liu *et al.* [25] propose to use a sprinkling system along the roadside to mitigate $PM_{2.5}$ and $PM_{10}$ concentrations. However, it is only a conceptual system without quantitative analysis and results.

From the above reviews of related works, we can find that: *(i)* The current works on water spraying systems for dust control still focus on single location (*e.g.*, pollution sources) or indoor evaluations. How to characterize the pollution reduction in outdoor environment for multiple sprayer devices in still an unsolved research problem. *(ii)* Regarding the air pollution sensing research, current research are mainly about improving sensing data quality, generating air quality maps, doing spatial inference or temporal predictions, etc. However, how to improve the air quality with existing pollution control systems (*e.g.*, spraying system) is still missing. For the first time, we propose a framework to fill the gap between air quality sensing and air pollution control with spraying system.

## 9 CONCLUSION

In this work, we propose iSpray, a data analytics engine for $PM_{2.5}$ and $PM_{10}$ control at critical POIs by cost-effective water spraying. Its design systematically combines domain knowledge from environmental sciences and machine learning techniques. iSpray offers learnable pollution reduction modeling at single locations, accurate air pollution reduction map generation, and propagation-path-aware sprayer scheduling. Evaluations with in-field sprayer deployments show that iSpray reduces the total sprayer switch-on time by 32%, while decreasing the days of high $PM_{2.5}$ and $PM_{10}$ concentrations at key POIs by 12% and 16%. We envision our work as one of the first endeavors for precise urban air pollution control with ubiquitous data and commodity hardware.

## ACKNOWLEDGEMENTS

We thank the anonymous reviewers and associate editors for the valuable comments. This research was supported by the Lee Kong Chian Fellowship awarded to Zimu Zhou by Singapore Management University.

## REFERENCES

[1] Antonis Analitis, Klea Katsouyanni, Konstantina Dimakopoulou, Evangelia Samoli, Aristidis K Nikoloulopoulos, Yannis Petasakis, Giota Touloumi, Joel Schwartz, Hugh Ross Anderson, Koldo Cambra, et al. 2006. Short-term effects of ambient particles on cardiovascular and respiratory mortality. *Epidemiology* 17, 2 (2006), 230–233.

[2] N Kh Arystanbekova. 2004. Application of Gaussian plume models for air pollution simulation at instantaneous emissions. *Mathematics and Computers in Simulation* 67, 4-5 (2004), 451–458.

[3] B Brunekreef and B Forsberg. 2005. Epidemiological evidence of effects of coarse airborne particles on health. *European Respiratory Journal* 26, 2 (2005), 309–318.

[4] Siyi Cai, Yangjun Wang, Bin Zhao, Shuxiao Wang, Xing Chang, and Jiming Hao. 2017. The impact of the "air pollution prevention and control action plan" on PM2. 5 concentrations in Jing-Jin-Ji region during 2012–2020. *Science of the Total Environment* 580 (2017), 197–209.

[5] Ling Chen, Yifang Ding, Dandan Lyu, Xiaoze Liu, and Hanyu Long. 2019. Deep Multi-Task Learning based Urban Air Quality Index Modelling. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 2:1–2:17.

[6] Yun Cheng, Xiaoxi He, Zimu Zhou, and Lothar Thiele. 2019. ICT: In-Field Calibration Transfer for Air Quality Sensor Deployments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 6:1–6:19.

[7] Yun Cheng, Xiaoxi He, Zimu Zhou, and Lothar Thiele. 2020. MapTransfer: Urban Air Quality Map Generation for Downscaled Sensor Deployments. In *Proceedings of the IEEE/ACM International Conference on Internet-of-Things Design and Implementation*. IEEE Press, Piscataway, NJ, USA, 14–26.

[8] Yun Cheng, Xiucheng Li, Zhijun Li, Shouxu Jiang, and Xiaofan Jiang. 2014. Fine-grained air quality monitoring based on gaussian process regression. In *International Conference on Neural Information Processing*. Springer, Berlin, Germany, 126–134.

[9] Yun Cheng, Xiucheng Li, Zhijun Li, Shouxu Jiang, Yilong Li, Ji Jia, and Xiaofan Jiang. 2014. AirCloud: a Cloud-based Air-Quality Monitoring System for Everyone. In *Proceedings of the ACM Conference on Embedded Networked Sensor Systems*. ACM, New York, NY, USA, 251–265.

[10] Yun Cheng, Olga Saukh, and Lothar Thiele. 2021. TIP-Air: Tracking Pollution Transfer for Accurate Air Quality Prediction. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*. ACM, New York, NY, USA, 589–599.

[11] Andrew W Correia, C Arden Pope III, Douglas W Dockery, Yun Wang, Majid Ezzati, and Francesca Dominici. 2013. The effect of air pollution control on life expectancy in the United States: an analysis of 545 US counties for the period 2000 to 2007. *Epidemiology (Cambridge, Mass.)* 24, 1 (2013), 23.

[12] Ada del Corno, Sonia Morandi, Flavio Parozzi, Lucio Araneo, and Francesco Casella. 2017. Experiments on aerosol removal by high-pressure water spray. *Nuclear Engineering and Design* 311 (2017), 28–34.

[13] Rong Ding, Zhaoxing Yang, Yifei Wei, Haiming Jin, and Xinbing Wang. 2021. Multi-agent reinforcement learning for urban crowd sensing with for-hire vehicles. In *Proceedings of the IEEE Annual International Conference on Computer Communications*. IEEE Press, Piscataway, NJ, USA, 1–10.

[14] Yi Gao, Wei Dong, Kai Guo, Xue Liu, Yuan Chen, Xiaojin Liu, Jiajun Bu, and Chun Chen. 2016. Mosaic: A low-cost mobile sensing system for urban air quality monitoring. In *Proceedings of the IEEE Annual International Conference on Computer Communications*. IEEE Press, Piscataway, NJ, USA, 1–9.

[15] Michael Greenstone and Patrick Schwarz. 2018. Is China winning its war on pollution?

[16] David Hasenfratz, Olga Saukh, Christoph Walser, Christoph Hueglin, Martin Fierz, Tabita Arn, Jan Beutel, and Lothar Thiele. 2015. Deriving High-Resolution Urban Air Pollution Maps using Mobile Sensor Nodes. *Pervasive and Mobile Computing* 16 (2015), 268–285.

[17] Baidu Inc. 2021. Baidu POI. http://lbsyun.baidu.com/index.php?title=androidsdk/guide/search/poi.

[18] Arnaud Jutzeler, Jason Jingshi Li, and Boi Faltings. 2014. A Region-based Model for Estimating Urban Air Pollution. In *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI, Palo Alto, CA, USA, 424–430.

[19] Haidong Kan and Bingheng Chen. 1989. Analysis of exposure-response relationships of air particulate matter and adverse health outcomes in China. *Journal of Environment and Health* 6 (1989).

[20] Fred N Kissell. 2003. *Handbook for Dust Control in Mining*. NIOSH, Cincinnati, OH, USA.

[21] Jundika Candra Kurnia, Agus Pulung Sasmito, and Arun Sadashiv Mujumdar. 2014. Dust Dispersion and Management in Underground Mining Faces. *International Journal of Mining Science and Technology* 24, 1 (2014), 39–44.

[22] Van-Duc Le, Tien-Cuong Bui, and Sang-Kyun Cha. 2020. Spatiotemporal deep learning model for citywide air pollution interpolation and prediction. In *Proceedings of the IEEE International Conference on Big Data and Smart Computing*. IEEE Press, Piscataway, NJ, USA, 55–62.

[23] Xiucheng Li, Yun Cheng, Gao Cong, and Lisi Chen. 2017. Discovering Pollution Sources and Propagation Patterns in Urban Area. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, USA, 1863–1872.

[24] Liang Liu, Wu Liu, Yu Zheng, Huadong Ma, and Cheng Zhang. 2018. Third-eye: a mobilephone-enabled crowdsensing system for air quality monitoring. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 20:1–20:26.

[25] Shiyong Liu, Konstantinos Triantis, and Lan Zhang. 2014. The Design of an Urban Roadside Automatic Sprinkling System: Mitigation of PM2. 5–10 in Ambient Air in Megacities. *Chinese Journal of Engineering* 2014 (2014), 1–12.

[26] Feng Lu, Dongqun Xu, Yibin Cheng, Shaoxia Dong, Chao Guo, Xue Jiang, and Xiaoying Zheng. 2015. Systematic review and meta-analysis of the adverse health effects of ambient PM2. 5 and PM10 pollution in the Chinese population. *Environmental Research* 136 (2015), 196–204.

[27] Rui Ma, Ning Liu, Xiangxiang Xu, Yue Wang, Hae Young Noh, Pei Zhang, and Lin Zhang. 2020. Fine-Grained Air Pollution Inference with Mobile Sensing Systems: A Weather-Related Deep Autoencoder Model. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020), 52:1–52:21.

[28] Balz Maag, Zimu Zhou, and Lothar Thiele. 2018. W-Air: Enabling Personal Air Pollution Monitoring on Wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 24:1–24:25.

[29] Yang Ou, J Jason West, Steven J Smith, Christopher G Nolte, and Daniel H Loughlin. 2020. Air pollution control strategies directly limiting national health damages in the US. *Nature Communications* 11, 1 (2020), 1–11.

[30] Dariusz Prostański. 2013. Use of air-and-water spraying systems for improving dust control in mines. *Journal of Sustainable Mining* 12, 2 (2013), 29–34.

[31] Carl Edward Rasmussen. 2004. Gaussian Processes in Machine Learning. *Advanced Lectures on Machine Learning: ML Summer Schools* 3176 (2004), 63.

[32] Federica Rollo, Bharath Sudharsan, Laura Po, and John G Breslin. 2021. Air Quality Sensor Network Data Acquisition, Cleaning, Visualization, and Analytics: A Real-world IoT Use Case. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*. ACM, New York, NY, USA, 67–68.

[33] Xingjian SHI, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun WOO. 2015. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (Eds.). Curran Associates, Inc., Red Hook, NY, USA, 802–810.

[34] Qian Tao, Yongxin Tong, Shuyuan Li, Yuxiang Zeng, Zimu Zhou, and Ke Xu. 2021. A Differentially Private Task Planning Framework for Spatial Crowdsourcing. In *Proceedings of the IEEE International Conference on Mobile Data Management*. IEEE Press, Piscataway, NJ, USA, 9–18.

[35] Rundong Tian, Christine Dierk, Christopher Myers, and Eric Paulos. 2016. Mypart: Personal, Portable, Accurate, Airborne Particle Counting. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1338–1348.

[36] Evelyn Tjoe Nij, Simone Hilhorst, Ton Spee, Judith Spierings, Friso Steffens, Mieke Lumens, and Dick Heederik. 2003. Dust Control Measures in the Construction Industry. *Annals of Occupational Hygiene* 47, 3 (2003), 211–218.

[37] Yongxin Tong, Yuxiang Zeng, Zimu Zhou, Lei Chen, Jieping Ye, and Ke Xu. 2018. A unified approach to route planning for shared mobility. *Proceedings of the VLDB Endowment* 11, 11 (2018), 1633–1646.

[38] Michelle C Turner, Daniel Krewski, C Arden Pope III, Yue Chen, Susan M Gapstur, and Michael J Thun. 2011. Long-term ambient fine particulate matter air pollution and lung cancer in a large cohort of never-smokers. *American journal of respiratory and critical care medicine* 184, 12 (2011), 1374–1381.

[39] World Health Organization. 2018. Ambient Air Pollution - a Major Threat to Health and Climate. http://www.who.int/airpollution/ambient/en/.

[40] Di Wu, Tao Xiao, Xuewen Liao, Jie Luo, Chao Wu, Shigeng Zhang, Yong Li, and Yike Guo. 2020. When Sharing Economy Meets IoT: Towards Fine-grained Urban Air Quality Monitoring through Mobile Crowdsensing on Bike-share System. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020), 61:1–61:26.

[41] Yu-Fei Xing, Yue-Hua Xu, Min-Hua Shi, and Yi-Xin Lian. 2016. The impact of PM2. 5 on the human respiratory system. *Journal of thoracic disease* 8, 1 (2016), E69.

[42] Shaocai Yu. 2014. Water Spray Geoengineering to Clean Air Pollution for Mitigating Haze in China's Cities. *Environmental Chemistry Letters* 12, 1 (2014), 109–116.

[43] Paolo Zannetti. 1990. Gaussian models. In *Air Pollution Modeling*. Springer, Berlin, Germany, 141–183.

[44] Yuxiang Zeng, Yongxin Tong, and Lei Chen. 2019. Last-mile delivery made practical: An efficient route planning framework with theoretical guarantees. *Proceedings of the VLDB Endowment* 13, 3 (2019), 320–333.

[45] Yu Zheng, Furui Liu, and Hsun-Ping Hsieh. 2013. U-air: When Urban Air Quality Inference Meets Big Data. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, USA, 1436–1444.

[46] Yu Zheng, Xiuwen Yi, Ming Li, Ruiyuan Li, Zhangqing Shan, Eric Chang, and Tianrui Li. 2015. Forecasting fine-grained air quality based on big data. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, USA, 2267–2276.