AQX: Explaining Air Quality Forecast for Verifying Domain Knowledge using Feature Importance Visualization

RESHIKA PALANIYAPPAN VELUMANI, The Hong Kong University of Science and Technology, Hong Kong
MENG XIA, Korea Advanced Institute of Science and Technology, South Korea
JUN HUN, Univeristy of Norte Dame, USA
CHAOLI WANG, Univeristy of Norte Dame, USA
ALEXIS LAU, The Hong Kong University of Science and Technology, Hong Kong
HUAMIN QU, The Hong Kong University of Science and Technology, Hong Kong

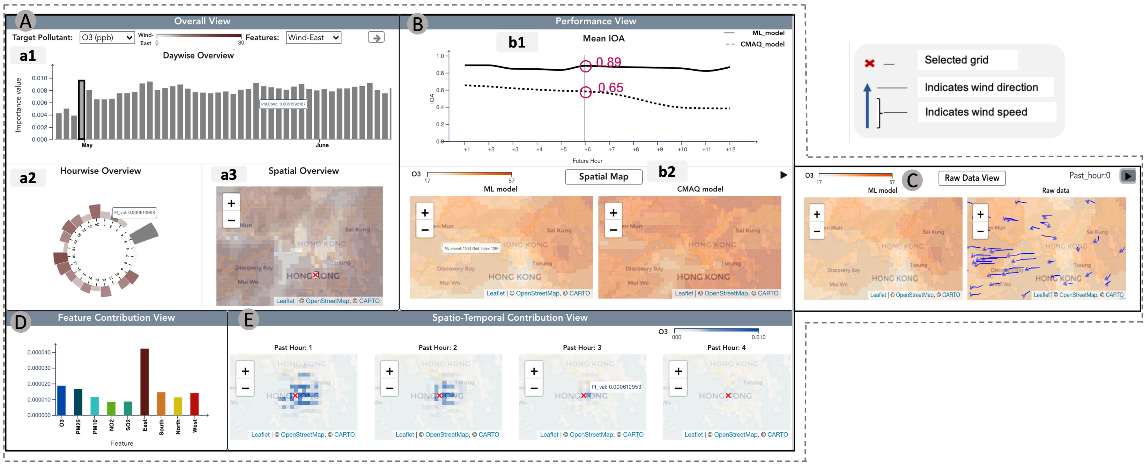

Fig. 1. AQX contains multiple coordinated views to support exploring, analyzing, and verifying the ML model's learning with domain knowledge using feature contribution information along with performance and raw data information. To understand the global pattern of feature contribution, the Overview (A) displays the feature contribution aggregated and presented as Daywise Overview (a1), Hourwise Overview (a2), and Locationwise Overview (a3). It also aids in narrowing down to the instance of interest. The *Performance View* (B) displays and compares the ML and CMAQ models' forecast accuracy on monitoring stations in Mean IOA(Index of Agreement) view (b1) and the spatial patterns captured by the model for the entire Hong Kong region in Spatial Map View (b2) for the target timestamp and pollutant to understand what the model can and cannot learn. The *Raw data View* (C) shows the wind trajectories for the input time period using animation which aids in understanding how wind carries pollutants from one place to another. The Feature-Temporal Importance view (D) shows the overall contribution of input features for the instance of interest and this helps in knowing the highly contributing features for the forecast. The Spatio-Temporal View (E) shows the contribution of grid locations from different timestamps which helps to understand the contribution of features from different spatial locations for the input time period.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2022 Association for Computing Machinery.
Manuscript submitted to ACM

Air pollution forecast has become critical because of its direct impact on human health and its increased production caused by rapid industrialization. Machine learning (ML) solutions are being drastically explored in this domain because they can potentially produce highly accurate results with access to historical data. However, experts in the environmental area are skeptical about adopting ML solutions in real-world applications and policy making due to their black-box nature. In contrast, despite having low accuracy sometimes, the existing traditional simulation model (e.g., CMAQ) are widely used and follows well-defined and transparent equations. Therefore, presenting the knowledge learned by the ML model can make it transparent as well as comprehensible. In addition, validating the ML model's learning with the existing domain knowledge might aid in addressing their skepticism, building appropriate trust, and better utilizing ML models. In collaboration with three experts with an average of five years of research experience in the air pollution domain, we identified that feature (meteorological feature like wind) contribution, towards the final forecast as the major information to be verified with domain knowledge. In addition, the accuracy of ML models compared with traditional simulation models and raw wind trajectories are essential for domain experts to validate the feature contribution. Based on the identified information, we designed and developed AQX, a visual analytics system to help experts validate and verify the ML model's learning with their domain knowledge. The system includes multiple coordinated views to present the contributions of input features at different levels of aggregation in both temporal and spatial dimensions. It also provides a performance comparison of ML and traditional models in terms of accuracy and spatial map, along with the animation of raw wind trajectories for the input period. We further demonstrated two case studies and conducted expert interviews with two domain experts to show the effectiveness and usefulness of AQX.

CCS Concepts: • **Applied computing → Environmental sciences**; • **Human-centered computing → Visual analytics**.

Additional Key Words and Phrases: Machine Learning, Explainable AI, Air pollution, Visual Analytics, Spatio-Temporal Data, Validation

ACM Reference Format:
Reshika Palaniyappan Velumani, Meng Xia, Jun Hun, Chaoli Wang, Alexis Lau, and Huamin Qu. 2022. AQX: Explaining Air Quality Forecast for Verifying Domain Knowledge using Feature Importance Visualization. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 24 pages. https://doi.org/10.1145/3490099.3511150

1 INTRODUCTION

Due to rapid urbanization, environmental pollution, specifically air pollution, has become more serious. It directly impacts human health and causes severe health complications like chronic respiratory diseases, heart diseases, and lung cancer [24]. To tackle this issue, modeling, forecasting, and monitoring air quality has become a hot spot among the scientific community [27]. In particular, with the increased availability of historical data, machine learning (ML) has gained considerable attention in this critical domain. It can potentially model and forecast complex data like air quality data accurately, which is dynamic, volatile, and highly variable in space and time [9, 37, 53]. Various ML techniques [14, 30] have been proposed by machine learning researchers as a solution that can forecast air quality like traditional statistical methods like CMAQ (Community Multi-scale Air Quality modeling) [5]. Moreover, National Science Foundation (NSF) has funded 100 million USD to establish AI institutes that can accelerate the research in AI for ES(Environmental Science) [1].

However, experts in the environmental science domain are skeptical about adopting ML solutions in real-world applications and policy-making due to their black-box nature. In contrast, the traditional model like CMAQ is currently widely used in many real-world applications and policy decisions on air quality management [2] as it works transparently based on clearly defined physical and chemical equations. However, these traditional models are not good at modeling sudden changes or non-linear behavior, which often results in less accurate forecast results [10]. Therefore, presenting the knowledge learned by the ML model might make it transparent and comprehensible for the domain experts.

[1] https://www.ai2es.org/
[2] https://www.epa.gov/cmaq/cmaq-models-0

Furthermore, validating the ML model's learning with the existing domain knowledge might aid in addressing their skepticism, building appropriate trust, and better utilizing ML models. Concurrently, numerous techniques have been proposed in the XAI (Explainable AI) field to uncover the black box, which has proven to be successful in illuminating the workings of machine learning models [1, 6]. However, little work has systematically investigated the prospect of using XAI for validating domain knowledge and what domain knowledge needs to be validated to gain the appropriate trust towards the ML models [16, 45].

Domain experts are end-users of ML models or XAI tools with more domain knowledge than common public but little-to-no technical background. Moreover, predominantly experts in the air pollution domain are end-users of ML models rather than developers themselves. Unlike common users, whose interest lies in the model results and performance, domain experts are more interested to understand what the ML model can learn from the data. In addition, they might be interested in verifying whether it is consistent with their knowledge [3]. Therefore, existing XAI tools explaining the workings of hidden layers of the ML model are neither easy for them to comprehend nor can be used to corroborate the domain knowledge.

To build a tool that facilitates experts in verifying their domain knowledge, it is essential to identify and distill critical domain knowledge that needs to be verified by domain experts to establish an appropriate trust in ML models. Therefore, we conducted a formative study following the design study methodology proposed in [43], with three domain experts with an average of five years of experience in conducting research pertaining to air pollution and air quality. From the formative study, we identified that feature (meteorological feature like wind) contribution towards the final forecast as the major information to be verified with domain knowledge. In addition, the performance of ML model compared with traditional simulation model and visualizing the raw wind trajectories are essential for domain experts to validate the feature contribution information.

Based on the findings, we derived seven design requirements to guide the overall design and development of AQX: A visual analytics system for verifying domain knowledge using feature importance visualization. In particular, AQX uses multiple coordinated views to present the contributions of input features at the different levels of aggregation in both temporal and spatial dimensions. In addition, it shows the performance information of both the ML model and the traditional CMAQ model for comparison. The system also visualises and presents the raw wind trajectories using animation to facilitate the validation process. The system was then evaluated by two case studies and an expert interview with two domain experts to demonstrate its effectiveness and usefulness. To summarize, we list our contributions as follows:

- A list of design requirements for an XAI tool that verifies domain knowledge in air pollution area.
- A visual analytics system for domain experts to explore, understand, and verify their knowledge by showing the contributions of input features at different levels of aggregation, the model performance, and the raw data.
- A comprehensive evaluation to demonstrate the usefulness and effectiveness of the system by two case studies and an expert interview.

2 RELATED WORK

The related work of this paper includes ML Models for Air Pollutant Forecasting, ML Interpretations, Visual Analytics for XAI and Visual Analytics for Spatio-Temporal (ST) data.

[3] NSF has established AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES) to develop XAI methods aligned with perspectives and priorities of environmental science domain. https://www.ai2es.org/research/foundational-research-in-trustworthy-ai-ml/

2.1 ML Models for Air Pollutant Forecasting

Air pollutant dataset falls under a specific type of Spatio-Temporal (ST) data which can be either represented as tensors or as spatial maps. Various machine learning models have been proposed to handle these two representatives of air pollution datasets and make the forecast. Since air pollutant data is often represented as spatial maps we will focus on related works which use a sequence of spatial maps as input to the ML model and produce a sequential spatial map as output. Here, the length of the sequences represents the input and output time period. Spatial maps can be essentially considered as image-like matrices, and thus Convolutional Neural Networks (CNN) has been used for the forecasting task [25, 28, 59, 60]. Due to the temporal attribute associated with these ST data, various model architectures consisting of Recurrent Neural Networks (RNN) [7, 8] have also been proposed for the forecasting task. However, performing forecasts with a sequence of spatial maps involves modeling temporal and spatial correlations, which requires a combined function of both CNN and RNN. One such approach that combines the convolutional structure of CNN and Long Short-Term Memory (LSTM) units is the convolutional LSTM network (ConvLSTM) layer [46] which was initially proposed for precipitation forecasting. The work used a sequence-to-sequence model whose input and output were both sequential spatial maps, which are ST data. Many variants of ConvLSTM like [54–56] have achieved impressive results on modeling and forecasting ST data. Few works focused on modeling air pollution data using ConvLSTM [46]. They used spatial maps generated from air quality and meteorological data collected from monitoring station as the input to the ConvLSTM model and forecast the future air quality for the study region. In this work, we adopt the model architecture from [2], it used a sequence-to-sequence model architecture with ConvLSTM as the building block and further utilized the results of a simulation model to make forecasts for future hours. We modified the architecture according to our needs and further enhanced the model by feeding it with fine-grained interpolated data.

2.2 ML Interpretation

There are two general categories under which the XAI works fall. One is intrinsic explainability, and the other one is post-hoc explainability. Simple models like linear/logistic regression [40], decision tree [15], k-nearest neighbors, etc., are transparent models which are self-explainable while complex models like neural networks require post-hoc explainability [1, 6]. In this literature review, we mainly focus on post-hoc techniques as our paper tries to explain neural network learning, which is a complex black-box model. Post-hoc explainability uses various methods like text explanations, visual explanations, local explanations, explanations by example, explanations by simplification, and feature relevance explanations techniques to alleviate the interpretability of the complex models [6]. We will further discuss feature relevance explanations in this section and visual explanations in subsection 2.3, as we adopt the aforementioned method to present the relational link between input-output.

Feature relevance explanations methods usually assign the input features, an importance score to show the individual impact of each feature on the final prediction, which help users understand the relationship between features and predictions. Some recent works which were based on sensitivity analysis methods, like Partial Dependence Plot (PDP) [18], SHAP[33], are commonly adopted for illustrating how a change in feature value affects the prediction result. However, one major limitation of sensitivity analysis methods like PDP, SHAP, etc. is that they are computationally expensive. It becomes infeasible when the dataset has an exceptionally large number of features like ST data [36] with multiple variables having spatial and temporal dimensions, each of which will be considered as an input feature. A feasible approach that can be adopted in our scenario is the gradient-based method [29]. Gradient-based methods provide feature relevance by calculating the first derivative of the output with respect to the input [29]. In this paper, we produce our explanation based on gradient-based methods as they are computationally inexpensive, provide more certain and reliable results, are well supported by most of the ML frameworks, and can be implemented with ease [12, 35, 38, 42, 44].

2.3 Visual Analytics for XAI

Representing the ML model's inner workings using visual analytics (VA) is the most inherent way to explain for non-ML-expert people like domain experts. VA interfaces utilise novel interactions to enable users to interact, which can help them in exploring, understanding, diagnosing the model and the underlying data as well [26]. There has been a recent surge in XAI works which makes use of visual analytics for explaining complex ML model's behavior [11, 19, 20, 34, 41, 47, 49–52, 57]. These VA interfaces were designed based on requirements for a particular set of end-users who can be ML experts, domain experts, or common public and evaluated using case studies and qualitative feedback. This literature review mainly focuses on visual analytics for XAI work on ST data, especially air pollution datasets. One closely related work [45] tried to explain RNNs in high-dimensional time-series forecasts from two aspects: model mechanism and feature importance to the domain experts. Another work [48] visualized the influence of input (space-time and data features) for each prediction using correlation charts. [22] explained the input(Temperature, wind, and humidity)- output (pollutants concentration) relationship using SHAP values on data collected from the monitoring stations (sparse ST data). Despite having the forecasting model itself as its main contribution, some works like ADAIN [14] and GeoMAN [30] interpreted the ML model in terms of local spatial dependency by visualizing the weights learned by the attention layers using heat maps and scatter plots. However, these XAI works in air quality focused on explaining the ML models based on what the domain experts do not know about the model. But not many works investigate the prospect explaining ML model based what the experts already know (i.e.) domain expert's domain knowledge, to increase transparency and trust. As a result, in this paper, we design a visual analytics solution that explains ML model to help domain experts corroborate their domain knowledge.

2.4 Visual Analytics for ST data

One of the distinct attributes of the ST dataset is the data volume, and it might not be easy to process and visualize these kinds of extensive data without the cost of time. Some works, such as imMens [32], Datavore [17], DICE [23], speeded up queries by pre-aggregating the data. In addition, they also utilized GPUs to achieve faster query results. However, data reduction like aggregation can be made effectively only with prior knowledge of the domain field. Otherwise, some interesting outliers and patterns might remain hidden from the users. A more recent work TPFlow [31] used a dataset subdivision algorithm to identify subsets with similar trends/patterns along multiple dimensions for further observation and comparison. Some works which deal with air pollution data in specific [16, 39, 61] used clustering based on similarity to aggregate and handle the massiveness of the dataset. Previous studies [4] suggested that ST data especially air pollution concentration, differs based on geographical locations and temporal cycles. Thereby spatial and temporal dimensions of the data should be considered while performing aggregation or data reduction to reveal interesting patterns. As such, we present the data at different levels of aggregation in both spatial and temporal dimensions to perform in-depth analysis.

Current works visualizing multidimensional ST data either uses multi-coordinated visualization or multivariate visualization, to summarise and display information from temporal and spatial dimensions and different input features like pollutants, meteorological data. There are works [3, 13] that leveraged the combination of both methods to mitigate the problem of visual encoding exhaustion. However, visual analytics for ST data require extra considerations for the dynamic nature of the data and their features, especially in air pollution domain where meteorological features like wind which is highly dynamic and moves over space and time dimensions. Considering this aspect of the ST data features is crucial as it helps domain experts verify their domain knowledge. As such, we propose a novel VA system that visualize raw wind trajectories and presents it in the form of animation for the validating the existing knowledge with that of the ML model's learning.

3 INFORMING THE DESIGN

3.1 Formative Study

The formative study helped us to collect information about the key domain knowledge of different features especially wind and its contribution to air pollutant's concentration forecast. We designed the formative study as shown below.

Participants and Procedure: We associated with 3 domain experts (2 Male, 1 Female) from the Environmental department for conducting the formative study. E1 is a professor in environmental science department with a research experience of more than 20 years. He predominantly studies numerical modeling of the atmosphere, regional and urban air pollution. E2 is a researcher with 5 years of experience in research which focuses on applying statistical methods in deterministic models and data fusion for air pollution forecast. E3 is a phd student conducting research in utilizing various ML models for air quality forecast. All the domain researchers have interest in utilizing ML models for their research and as well as in real world applications and therefore would like to understand their behavior. The study included a semi-structured interview asking the domain experts a series of questions for about 90 minutes. We started the Q&A session with questions about the evaluation methods in the air quality domain, as this can help us evaluate the ML model. We advanced with questions about input features to understand input-output relationship. Based on the answers given we asked follow-up questions to have a better understanding. The questions interchanged with the domain experts are listed in Tab. 1 and the answers provided by them are listed in appendix subsection A.1. Finally we drafted the initial design requirements based on the Q&A session. We iteratively collaborated with domain experts for a period of five months (November 2020 - March 2021) by holding biweekly meetings. Inputs from the experts helped us to ensure that the results produced by the ML model and feature contribution information generated by the gradient based method are acceptable as well as the developed visual analytical system meets their requirements.

3.2 Design Requirements

We identified three primary information the domain experts needed for verifying their domain knowledge from the formative study. The most important information required is **(I1) Feature Contribution:** Contributions of input features (air quality and meteorological features) presented at different levels of aggregation can reveal global and instance level patterns, which can be further verified with domain knowledge. Other essential information that's required is **(I2) Performance of ML and Simulation models:** Performance of the ML model compared with that of simulation model can provide a holistic understanding of what the ML model can and cannot learn. The last critical information required is **(I3) Raw Wind data:** Raw Wind data is used to validate the input feature contribution (I1). Since wind helps in carrying pollutants from one place to another, visual presentation of the raw wind trajectories for the input time period can aid in verifying the contribution of spatial dimension to the final forecast. Based on the three information described above; we further summarised the following seven design requirements. Design requirements R1 to R6 are related to the air quality domain, and the requirement R7 is related to UI (user interface) design. The

Table 1. Questions from the formative study.

Questions	Information
Q1: How to evaluate machine learning model? • **Q1(a):** How to evaluate spatial consistency and coherence? **Q2:** Why CMAQ is widely accepted?	**I2:** Performance of Ml and simulation models.
Q3: What are the most important features in air quality forecast? • **Q3(a):** How does wind affect air pollutant concentration? • **Q3(b):** Effect of wind on different pollutants? **Q4:** Does air pollutants affect concentration of another air pollutant?	**I1:** Feature Contribution
Q5: Other useful information about wind?? • **Q5(a):** What are the influential features other than wind?.	**I3:** Raw Wind data

relationship among the identified information from formative study, derived design requirements,and proposed design elements for the system are shown in Figure 2.

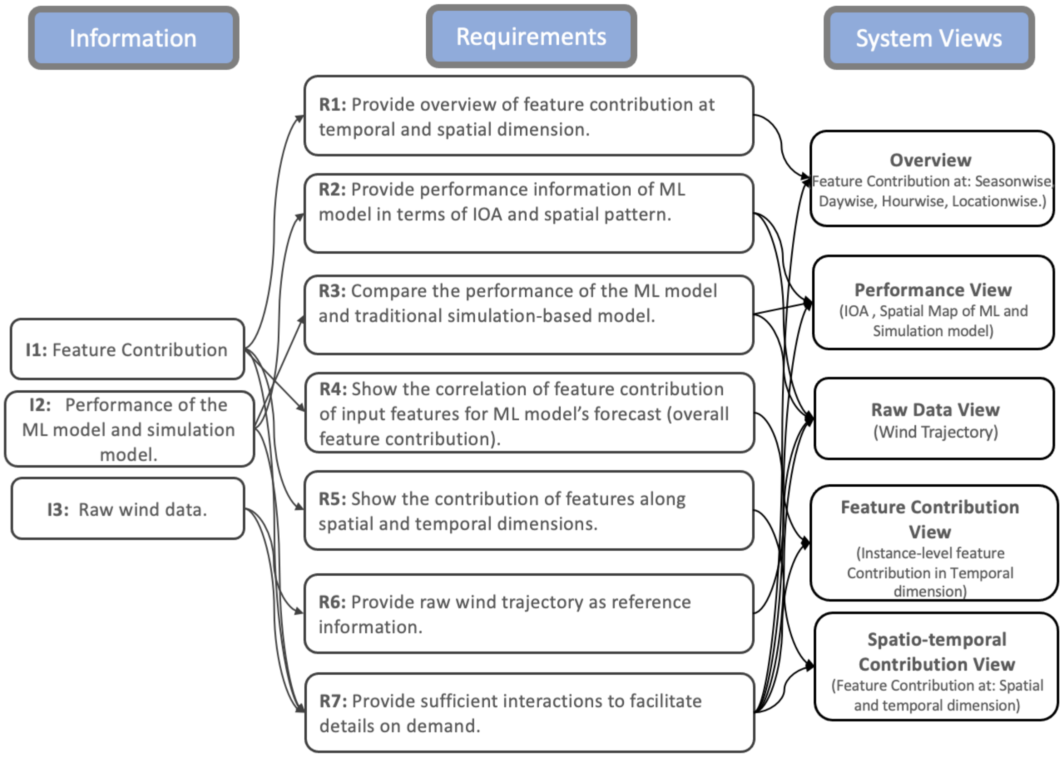
Information
Requirements
System Views
R1: Provide overview of feature contribution at temporal and spatial dimension.
R2: Provide performance information of ML model in terms of IOA and spatial pattern.
R3: Compare the performance of the ML model and traditional simulation-based model.
R4: Show the correlation of feature contribution of input features for ML model's forecast (overall feature contribution).
R5: Show the contribution of features along spatial and temporal dimensions.
R6: Provide raw wind trajectory as reference information.
R7: Provide sufficient interactions to facilitate details on demand.
I1: Feature Contribution
I2: Performance of the ML model and simulation model.
I3: Raw wind data.
Overview
Feature Contribution at: Seasonwise, Daywise, Hourwise, Locationwise.)
Performance View
(IOA , Spatial Map of ML and Simulation model)
Raw Data View
(Wind Trajectory)
Feature Contribution View
(Instance-level feature Contribution in Temporal dimension)
Spatio-temporal Contribution View
(Feature Contribution at: Spatial and temporal dimension)

Fig. 2. The relationship among identified information from formative study, derived design requirements,and proposed design elements for the system.

R1: Provide an overview of feature contribution at temporal and spatial dimensions: All the experts (**E1, E2, E3**) mentioned that the input feature's contribution, especially wind and its direction on air pollutants differs for different seasons. For example, during summer, the contribution of easterly wind to air quality is high, whereas, in winter, northerly winds have a high contribution. Moreover, wind's contribution to air quality differs for different geographical locations. For instance, the wind might have a higher contribution to air quality over places near the sea than over places in the city center. Therefore, an overview of the input feature's contribution across different seasons and its comparison is needed to verify high-level domain knowledge in the temporal dimension. In addition, the overview of the contribution of features in various geographical locations is also required to verify domain knowledge in spatial dimensions.

R2: Provide the performance of ML model in terms of IOA and spatial pattern: All the experts (**E1, E2, E3**) agreed that showing the performance of the forecast model can give an instant insight about what the model can and cannot learn from the data. **E1** mentioned that *"A ML model's performance on air quality data is usually assessed based on two factors, one is accuracy which is measured in terms of Index of Agreement (IOA) calculated on monitoring stations, and the other one is spatial pattern and consistency which can be evaluated by spatial map"*. Therefore, the system needs to present the performance information of the ML model in terms of IOA and spatial map.

R3: Compare the performance of ML model, and traditional simulation-based model: E1 mentioned that CMAQ is a commonly used simulation model in the air pollution domain and should be considered as a baseline for evaluating the ML model's performance, especially to assess the spatial consistency and coherence of the ML model's forecast. **E3** added *"CMAQ is a widely accepted traditional simulation model and is currently being used for policy making regarding air pollution. It cannot be ignored completely, at least CMAQ should be considered as baseline for comparing and evaluating the ML model's performance."* Therefore, providing the details of the simulation model's performance in terms of IOA and spatial map can facilitate the comparison of performances of the ML model and CMAQ model.

R4: Show the correlation of feature contribution of input features for ML model's forecast (Overall Feature Contribution): E2 and **E3** stated that certain air pollutant features are highly correlated and can influence the concentration of each other. For example, PM pollutants and O3 are correlated. They can contribute to each other's concentration either positively or negatively, but PM and small-scale pollutants like SO2 and NO2 usually do not show any correlation. **E2** mentioned *"NO2, SO2 and PM pollutants have no correlation because NO2 and SO2 is highly reactive and its presence in the air is for a short period of time, thus they have less contribution for PM pollutant's concentration."* Thus showcasing the input feature contribution for air pollutants forecast can aid in verifying the knowledge about the correlation between features.

R5: Show the contribution of features along spatial and temporal dimensions: Experts (**E1, E2, E3**) stated that each input feature varies across spatial and temporal dimensions, particularly wind, which fluctuates highly in both dimensions. So, it is vital to show the contribution of input features from spatial and temporal dimensions for the final forecast. This information can be used to verify the common understanding regarding the contribution of these dimensions in forecasts. **E3** mentioned that *"Usually features from the nearest input time periods and spatial locations have a strong influence over the final forecast for a particular instance."*

R6: Provide raw wind trajectory as reference information: E2 highlighted the importance of presenting raw data, especially raw wind trajectories. Wind helps in carrying pollutants from one place to another and can be responsible for sudden changes in air quality, so it is essential to show the raw wind trajectory for the input time period, and this can function as additional information to support the ML model's learning and validate the feature contribution information in the spatial dimension.

R7: Provide sufficient interactions to facilitate details on demand: Experts **E1, E2, and E3** expressed their interest in performing case-by-case analysis apart from analyzing global or high-level seasonal and geographical patterns. In particular, the experts wanted to understand the role played by spatial and temporal dimensions of the input features in the final forecast of an instance. **E2** also added that he is very much interested in enriching his knowledge by exploring some exciting instances. So the system needs to support interactions like selection, filter, and tool-tips.

4 FORECAST MODELING

In this section, we describe the application dataset (air quality dataset), the model architecture we use for the forecast, and the method we used for data interpolation.

4.1 Application Domain and Dataset

In this paper, we focus on explaining ML model behavior on air pollution forecast. We mainly focus on five pollutants that drastically affect human health, namely PM10, PM25, O3, SO2, and NO2. The study area for this paper is Hong Kong. There are 16 air quality and 28 meteorological monitoring stations installed across Hong Kong to collect air quality and meteorological data respectively on an hourly basis. We use both air quality and meteorological data collected over one year from 1st January 2018 to 31st December 2018 for training and evaluating the ML model. Each input feature has both spatial and temporal dimensions (i.e., its values change over space and time). The features of the data are listed in Tab. 2.

Table 2. Features taken as input: air pollutant and meteorology.

Category	#Monitoring Stations	Feature Type	Units
Air Pollutant Data	16	PM2.5	µg/m3
		PM10	µg/m3
		O3	ppb
		SO2	ppb
		NO2	ppb
Meteorological Data	28	Wind speed	Meter/Second
		Wind direction	Degree
		Temperature	Celsius

4.2 Data Processing

The domain experts aim to forecast air quality for the entire Hong Kong region with the data collected from the monitoring stations. Therefore, we divided the study area into 64 × 41 grids with a resolution of 1km. We then interpolated the data for locations without monitoring stations using a well-established technique called the Gradient Vector Flow (GVF) [58]. GVF, which is a two-stage diffusion approach, is leveraged to obtain the velocity and values of air pollutants at all locations from the sparsely sampled data. In the first stage, it estimates the velocities and air pollutants at these pixels, which are closest to the ground-truth grid cells (i.e.) grid cells with monitoring stations, through a weighted linear interpolation. Then, the interpolated values are iteratively diffused to the whole region by minimizing the Laplacian. Similarly, data were interpolated for all timestamps from January 2018 to 31st December 2018 and for all the input features (PM10, PM25, O3, SO2, NO2, Wind).

For generating and visualizing the trajectories of wind from the raw data, we used Euler's method [21], a numerical method for tracing particle's flow in a two-dimensional vector field. It is a two-step process; first, the user needs to select a constant named "Step-size", and second, based on the selected constant, the algorithm calculates the next position of the particle (wind) along the direction of the vector using bilinear interpolation method. These two steps are repeated until the wind position reaches the extent of the study area. These extracted points can be further used to visualize wind trajectories for the given time period.

4.3 Model Architecture

This section describes the components, input and output of the ML model we built, while the proposed visual analytics solution can support any differentiable ML model (i.e.) models with functions for which derivatives can be computed. The ML model we built follows an encoder-decoder structure with ConvLSTM as the building block, inspired from the model architecture of [2], to forecast the pollutant's concentration. The model architecture shown in Figure 3 consists of two parts: encoder and decoder. Both encoder and decoder have the same architecture, which comprises two ConvLSTM blocks with 64 and 32 feature maps, respectively, and 3×3 convolutional kernel, each followed by a batch normalization layer. The input for the ML model is past 12 hour interpolated data (spatial map) and the output is forecasted future 12 hour data (spatial map). We use the data collected from the monitoring station, issued by the Hong Kong Observatory which are publicly available, as ground truth to evaluate the model. Each data point given as an input for the encoder is a 4D tensor with the dimension of $\left(N_{in}, N_x, N_y, N_{pol+mete}\right)$, where N_{in} is the number of timestamps of input historical air quality data (in our case, $N_{in} = 12$), (N_x, N_y) is the grid location, N_{pol} is the number of pollutants, and N_{mete} is the number of meteorological features. The encoder outputs a state which encodes the historical data in the dimension of $(N_x, N_y, N_{h_{enc}})$, where $N_{h_{enc}}$ is the number of feature maps generated in the last ConvLSTM in the encoder. Then the decoder takes this as input and outputs a 4D tensor of dimension $(N_{out}, N_x, N_y, N_{h_{dec}})$, where N_{out} is the number of future timestamps we want to forecast (in our case, it is 12 hours in future), and $N_{h_{dec}}$ is the number feature maps produced in the last ConvLSTM in the decoder. The decoder is then followed by a 2D convolutional layer with 1×1 kernel size and ReLU activation function and outputs 4D tensor of dimension $(N_{out}, N_x, N_y, N_{p_{out}})$, where $N_{p_{out}}$ is the number of pollutant to forecast (in our case $N_{p_{out}} = 1$). The model is optimized using mean squared error and Adam optimizer with a learning rate of 0.001. The model was evaluated on observed data from March, June, September and December of the year 2018 and trained on the remaining eight-month data for 100 epochs with a batch size of 16.

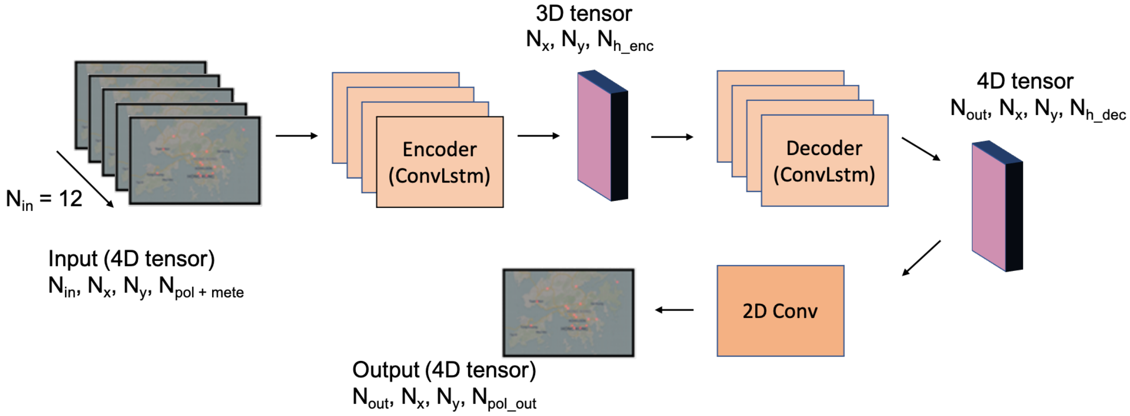
3D tensor
N_x, N_y, N_{h_enc}
4D tensor
N_{out}, N_x, N_y, N_{h_dec}
N_{in} = 12
Encoder (ConvLstm)
Decoder (ConvLstm)
Input (4D tensor)
N_{in}, N_x, N_y, $N_{pol + mete}$
2D Conv
Output (4D tensor)
N_{out}, N_x, N_y, N_{pol_out}

Fig. 3. ConvLSTM model architecture.

4.4 Extracting feature importance

Our model interpretation method is mainly based on first-derivative saliency [29]. We calculate the saliency score for individual units and aggregate across spatial locations or the specific timestamp to derive the spatial feature contribution or temporal feature contribution. Below we discuss the implementation details of the model interpretation method.

4.5 First-Derivative Saliency

Generally, the input is denoted as I, and the output of an ML model M is denoted as $M(I)$. According to the first-order Taylor expansion, we can approximate the model's output with a linear function of the input

$$M(I) \approx w(I)^T I + b. \tag{1}$$

where w and b are the weights and bias, respectively. Since we are using the first-order Taylor expansion, the value of $w(I)$ is the first-order derivative with respect to the model's output

$$w(I) = \left. \frac{\partial (M)}{\partial I} \right|_I . \tag{2}$$

Such derivatives can measure how sensitive the input unit is to the final forecast results [29]. We can use the derivative's absolute value to indicate the importance of this input unit to the final forecast, which is the saliency score $S(I)$

$$S(I) = |w(I)|. \tag{3}$$

4.6 Spatial and Temporal Feature Importance

Based on the definition of first-derivative saliency, we further define the spatial and temporal feature contribution as follows.

In our case, let's denote the input data as $I = \{I_1, I_2, , ..., I_{t_{in}}\}$ where t_{in} denotes the input time period. For a given timestamp, the forecast is a 2D tensor with a dimension of $W * H$, where W and H denote the width and height of the 2D tensor, respectively.

Considering $M_{(t_{out},x,y)}(I_{t_{in}})$ as the output of the model, (i.e., forecast instance, at a timestamp t_{out} on a location (x, y) in the 2D tensor), we can calculate the saliency score $S_t^{(x,y)}(I_{t_{in}})$ using first-derivative saliency which is also considered as the spatial feature contribution

$$S_t^{(x,y)}(I) = \left| \frac{\partial \left(M_{(t_{out},x,y)} \right)}{\partial I_{t_{in}}} \right|_I . \tag{4}$$

For temporal feature contribution calculation at a specific timestamp t, we simply consider the sum of the saliency score for all (x, y) locations at this timestamp as the temporal feature importance $S_t(I)$

$$S_t(I) = \sum_{1 \le x \le W, 1 \le y \le H} S_t^{(x,y)}(I). \tag{5}$$

5 AQX

In this section, we introduce AQX, a visual analytic system that explains air quality forecast for verifying domain knowledge.

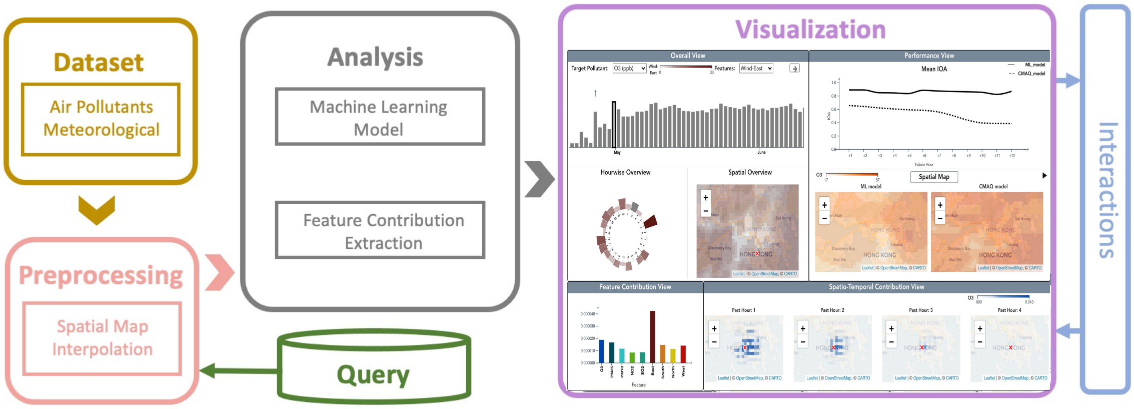

Fig. 4. AQX includes the preprocessing, analysis, visualization, and interaction modules.

5.1 System Overview

AQX comprises of four modules, namely: (1) Preprocessing; (2) Analysis; (3) Visualization; and (4) Interaction as shown in Figure 4. The preprocessing module interpolates the air pollutant and meteorological data for the entire Hong Kong region from the data collected at the monitoring station. The analysis module incorporates two main parts: the ML model part (ConvLSTM architecture) for forecasting the air pollutant, and the feature contribution extraction part calculated using the first derivative gradient method. The visualization module visualizes the feature contribution at various levels of aggregation and as well as includes visualization of performance information of the models and raw wind trajectories. In addition, the interaction module supports exploratory analysis with rich interactions.

The visualization module consists of five main views: (1) *Overview* displays the overall feature contribution at Temporal (Days and Hours) and Spatial dimensions; (2) *Performance View* shows the performance of ML model and traditional simulation model in terms of IOA (index of Agreement) values calculated at monitoring stations and Spatial Maps showing the spatial pattern and consistency of the forecast; (3) *Raw data View* presents an animation of wind trajectories, (i.e., wind movement along with speed and direction, for the input time period) (4) *Feature Contribution View* shows the instance level overall input feature's contribution in temporal dimension; and finally (5) *Spatio-Temporal Contribution View* displays the instance level input feature's contribution at both spatial and temporal dimensions.

5.2 Overview

Overview (Figure 1(A)) displays the overall feature contribution at different temporal and spatial resolutions. The feature contribution information is aggregated and presented as Daywise (Figure 1(a1)), Hourwise (Figure 1(a2)), and Locationwise (Figure 1(a3)) feature contributions. The *Overview* has filters to select the target pollutant and feature to analyse which are the inputs to the visualization system. The system uses different colors to encode the input features (NO2, PM25, O3, SO2, PM10, Wind-North, Wind-South, Wind-East, Wind-West) and use gradient of the respective colors to denote the feature contribution value. For example, a darker color gradient represents a higher feature contribution value and a lighter color gradient denotes a lower feature contribution value.

Daywise Overview (Figure 1(a1)) part uses bar chart to present the contribution of the selected feature for every day in a year. The view aids the domain experts in understanding the contribution of different features for the forecast of target air pollutant on different days over the year. Here the x-axis denotes days of the year, and the y-axis denotes feature contribution value. The users can interact with the bars by clicking on them to select a day of interest to explore further. A tooltip appears with the date and feature contribution information as the user hovers over the bars.

Hourwise Overview (Figure 1(a2)) This view helps the domain experts in understanding the contribution of the selected feature for the forecast of target pollutant throughout the day (for 24 hours), using a circular barplot, after the user selects a date of interest from the *Daywise Overview* chart. This chart provides interaction to select a particular hour of interest by clicking on the bar. A tooltip appears on hovering over these bars, showing the feature contribution information. The bar's color denotes the selected feature, and the gradient of the color represents the feature contribution value. For example, a darker gradient represents a higher feature contribution value, while a lighter gradient denotes a lower value. Barchart was considered as an alternate design for this view. However, utilizing a bar chart to present the feature contribution for 24 hour will be difficult due to space constraints. Moreover, a circular barplot resembles a clock. Hence, we chose the circular bar plot for this view.

Locationwise Overview (Figure 1(a3)) presents the contribution of the selected features in different spatial locations for the selected date and hour using a heatmap overlaid on top of the Hong Kong map. Based on the suggestions from domain experts, we selected grid cells having monitoring stations as representative locations to calculate feature contribution as they have the ground truth data to evaluate the ML model. The color of the heatmap denotes the selected feature, and the gradient of the color denotes the value of feature contribution. The users can select a particular location of interest by clicking on any grid cells on the heatmap.

5.3 Performance View

Performance View (Figure 1(B)) aims to show the performance of the ML model for the selected pollutant and timestamp and further facilitate comparing it with the traditional simulation (baseline) model. The model's performances are presented in terms of IOA (Index of Agreement) values, which measures the accuracy of forecast on the monitoring stations, and Spatial map, which showcases the model's ability to capture spatial patterns. The **mean IOA** line chart (Figure 1(b1)) presents the IOA values of ML and simulation models, averaged over all sixteen monitoring stations, for the forecast time period. The x-axis represents the forecast period, and the y-axis represents the mean IOA values. The lines are encoded as solid and dashed to represent the ML and simulation models. A tooltip appears on hovering over the lines, displaying the ML and simulation model's mean IOA value for the corresponding future hour. The **spatial map** (Figure 1(b2)) shows the pollutant concentration forecasted by ML and simulation model using two heat maps overlaid on top of Hong Kong geographical map, respectively. The heatmap helps to understand the spatial pattern of the forecasted pollutant. The color gradient denotes the concentration of the pollutant. For example, a darker gradient represents a higher concentration, and a lighter gradient represents a lower concentration. Tooltip appears on hovering over the heatmaps to display information like the grid number and pollutant concentration of the corresponding grid. This view also incorporates a legend on top, which helps the users understand the visual encodings.

5.4 Raw data View

Raw data View (Figure 1(C)) visualizes the animation of wind trajectories for the input time period. This view can be seen when the user clicks on the **Spatial Map** button present in the *Performance View*. This view has two heatmaps overlaid on top of the Hong Kong map, placed side-by-side. The left one shows the ML model's forecast for the selected pollutant and timestamp. The right one shows the movement of wind trajectories overlaid on the top of heatmap of the target pollutant for the input period from input hour 12 to input hour 1 using animation. The wind trajectories are represented using blue lines with arrow marks at the end, which indicate wind directions, and the lengths of the lines denote the wind speed. We used the pathline tracing method [21], which is a type of flow visual, to extract the trajectories from the wind vector data as it helps in making invisible flow patterns of wind visible. We adopted animation over static visualization since wind is a highly dynamic feature and moves in space and time simultaneously. Visualizing wind trajectories overlaid on top of pollutant concentration heatmap using animation can help in better understanding of how wind moves pollutants from one place to another.

5.5 Feature Contribution View

Feature Contribution View (Figure 1(D)) visualizes the overall contribution values of the input features aggregated across all spatial locations for the input time period for the selected instance using a bar chart. After domain experts select an instance of interest in the *Overview*, they can use the *Feature Contribution View* to understand what are the highly contributing features for this instance's forecast. The color and x-axis of the bar chart represent different input features, as seen in the *Overview*. The y-axis denotes the *Feature contribution value*.

5.6 Spatio-Temporal Contribution View

Spatio-Temporal Contribution View (Figure 1(E)) aims to provide the contribution of features in spatial and temporal dimension. This view helps to understand the contribution of features from different spatial locations of the input time period towards the forecast for the selected instance. This view uses 12 heatmaps overlaid on top of the Hong Kong maps showing the feature importance with a color gradient for the past 12 input hours. The color of the grid cells represents the selected feature, and the gradient represents the feature contribution of the spatial location. For example, a darker gradient denotes a higher feature contribution, while a lighter gradient denotes a lower contribution. The heatmaps in the Spatio-Temporal View zoom to the target grid location, highlighted via a red cross, to reveal fine-grained spatial contribution information. On hovering over the grid cells, a tooltip appears showing information about grid index and feature contribution value of the grid. Since usually, grid locations from near past hours have higher contribution values than the later past hours, we show the past hour 1, 2, 3, 4, and 5 in the main view, and further scrolling down the Spatio-Temporal View reveals the heatmaps of later timestamps from past hour 6 to past hour 12.

6 EVALUATION

In this section, we demonstrate two case studies and expert interview with two domain experts to show the effectiveness and usefulness of AQX in verifying domain knowledge.

6.1 Case Studies

This section describes the case studies observed by two domain experts (E1, E2). Both E1 and E2 participated in the formative study and were involved in the iterative design process, and hence they are familiar with the system. E1 used the system to verify whether the ML model's learning (feature contribution information) is consistent with their domain knowledge. And E2 used the system to analyze and understand the behavior of the ML model during extreme weather conditions and enrich his knowledge.

Case1: Verifying the domain knowledge Firstly, we, along with E1, summarised and categorized the following key domain knowledge that needs to be verified based on the information collected during the formative study.

D1: Wind.

- **D1(a):** Easterly winds have high contribution on air quality during the summer months.
- **D1(b):** Wind has high contribution on air quality in places near sea or open area and less contribution in places near city center.

D2: Pollutants.

- **D2(a):** Among PM pollutants (PM10, PM2.5) and O3, strong correlations can be observed (i.e.) the PM pollutants and O3 can contribute to each other's concentration.

D3: Spatial and Temporal dimension.

- **D3(a):** The concentration of pollutants at a particular time and place exhibits spatial and temporal dependencies (i.e.) features from the nearest previous timestamps and the nearest locations have a higher contribution towards the concentration of pollutants at the current location and timestamp.

E1 started the verification process with the *Overview* (Figure 1(A)); he selected O3 as the target pollutant using the drop-down menu (Target Pollutant) in the *Overview* as shown in Figure 1(A). E1 chose O3 since it is a highly toxic pollutant when present at ground level and is often analyzed with much importance[4]. E1 then analyzed the contribution of each features one by one in the *Daywise Overview* Figure 1(a1) of Overview part Figure 1(A) **(R1)**. He noticed that Wind-East has a relatively higher feature contribution during May which is a summer month in Hong Kong. This observation verifies the domain knowledge about wind during summer **(D1(a))**. While further analyzing, E1 noticed from the *Daywise Overview* that during a particular day in May (2018-05-11), the feature contribution of Wind-East was high as shown in Figure 1(a1). He selected the day of interest from the Bar chart in *Daywise Overview* to explore its feature contribution in *Hourwise Overview* (**(R7)**) Guided by the visual cue (color and height of the circular bar plot), E1 further narrowed down to a particular hour (05:00:00) which had a darker color gradient and higher bar height, as highlighted in Figure 1(a2). He later found that the wind speed at 2018-05-11 05:00:00 HKT was 40-50 km/hr, which is a strong wind, and the Hong Kong observatory issued Typhoon signal-3 warning for the selected timestamp (2018-05-11 05:00:00 HKT). From the *Locationwise Overview* as seen in Figure 1(a3), E1 observed that the color gradient of Wind-East feature has a relatively darker shade on open areas indicating higher contribution than in the city center **(R1)**. This verifies the second domain knowledge about wind as in **D1(b)**. Before analyzing the instance-level information, E1 went to the *Performance View* (Figure 1(B)) to check and compare the performance of the ML model and CMAQ model (**R2**, **R3**) for the selected timestamp (2018-05-11 05:00:00 HKT). From the *Mean IOA* view as shown in Figure 1(b1), E1 saw that the IOA values calculated at the monitoring stations for the results generated by the ML model are higher than that of the CMAQ model. E1 understood that the ML model performs better in forecasting at the monitoring stations than CMAQ model. Having known the performance of the models on monitoring stations, E1 analyzed the capability of the models in capturing the spatial pattern in the *Spatial Map* view as seen in Figure 1(b2). He understood that the CMAQ model and ML model have different results in terms of spatial patterns as shown in Figure 1(b2). This is because of the difference in input given to these models. CMAQ (traditional simulation model) takes input like geographical features, elevation data, traffic data, etc. on the other hand ML model takes only the data from the monitoring stations as the input. Further, based on his domain knowledge about O3, E1 mentioned that the CMAQ model's spatial map is more acceptable than the ML model's as O3 is a small scale pollutant and its concentration has abrupt spatial changes, which can be observed from the changes in color intensities of the spatial map as seen in 1(b2) CMAQ spatial map. In conclusion, from the *Performance View*, E1 was able to get a high-level understanding of what the ML model can and cannot learn from the ST (i.e.) air quality data. Upon selecting a location near the sea in the Hong Kong map from

[4] https://www.epa.gov/ground-level-ozone-pollution

Locationwise Overview 1(a3), E1 analyzed the instance-level contribution information. From the *Feature Contribution View* E1 observed that Wind-East has a higher contribution for the selected instance (Timestamp and grid location). He also noticed that apart from O3 pollutant, PM pollutants (PM10, PM25) had relatively high contribution for O3 forecast as seen in Figure 1(D) **(R4)**. This observation verifies the domain knowledge about the correlation that exists among pollutants **(D2(a))**. E1 moved to the *Spatio-Temporal Contribution View* (Figure 1(E)) to observe how features from spatial and temporal dimensions contribute to the final forecast for the selected instance (**R5**). From the Figure 1(E), he observed that features from near input past hours like Past Hours 1, 2, 3 have high feature contribution value as noted from the higher color gradient of neighboring grid cells surrounding the target grid cell indicated by the red cross sign. This observation validates the domain knowledge about the spatial and temporal attributes of the features (**D3(a)**). Finally, E1 used *Raw Data View* (Figure 1(C)) to check the raw wind trajectories for the input time period (2018-05-11 05:00:00 HKT) (**R6**). After viewing the animation, E1 noted that the wind was blowing from the east direction for the input time period. He also further noticed from *Spatio-Temporal Contribution View* as seen in 1(D) that the contribution of features from grid cells located in east direction has darker color gradient indicating high feature contribution which is aligned with the wind trajectory. E1 stated that the above observation made the feature contribution information shown in other views are more reasonable and acceptable. Through this case study E1 used *AQX* to verify some key domain knowledge.

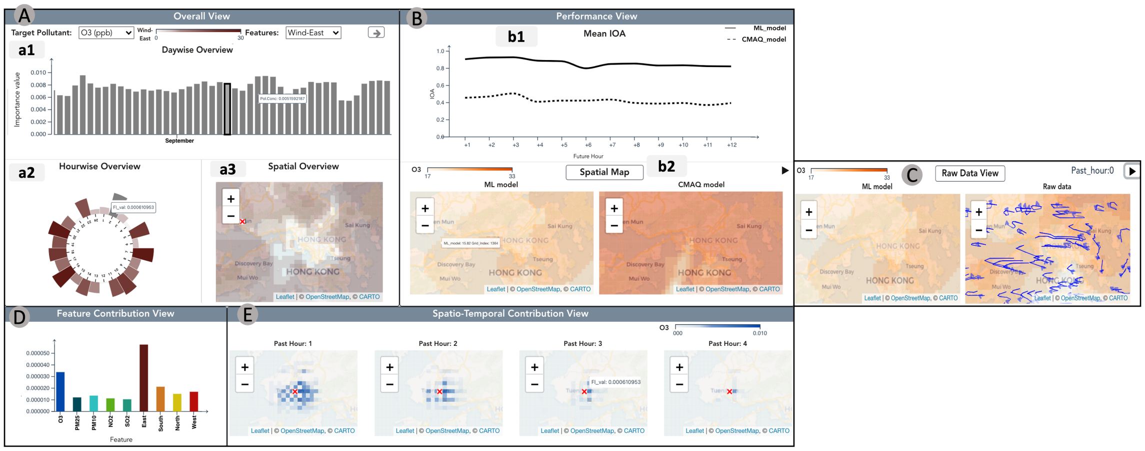

Fig. 5. ML model's learning during extreme weather conditions. Daywise (a1), Hourwise (a2), and Locationwise (a3) feature contributions of the selected pollutant. *Performance View* showing Mean_IOA (b1) and Spatial Map (b2) of ML and simulation model. *Raw data View* (c) showing wind trajectories. *Feature Contribution View* (d) showing the overall feature importance of the selected instance, and *Spatio-Temporal Contribution View* (e) showing the grid level contribution.

Case 2: Exploring the ML behavior during extreme weather conditions E2 was very interested in analyzing and understanding how the ML model learns from the data and makes forecasts during extreme weather conditions since the simulation model's forecasts are not-so-good during this period. In 2018, Hong Kong encountered a super typhoon from 15th September 2018 to 17th September 2018 and E2 wanted to analyze timestamp from this particular time period. E2 selected O3 as the target pollutant from the drop-down menu in the *Overview* (Figure 5(A)) to analyse and understand further.

E2 started analyzing the contribution of features for O3 forecast one-by-one using the filter in the *Daywise Overview* (Figure 5(a1)) **(R1)** for September month for the target pollutant. He found that Wind-East had a higher contribution than other features especially wind from other directions during September. Particularly on 12-16 September 2018, during which hurricane struck Hong Kong. Since the HK government issued a warning signal 10 (Highest warning signal) on 16th September 2018, E2 selected the corresponding bar to explore further **(R7)**. In the *Hourwise Overview* chart based on darker shade and the height of the bar as shown in Figure 5(a2), E2 picked second hour of the day as a point of interest as Wind-East has high contribution value at this particular hour. During this particular hour of the day, the wind speed was 144km/hr. E2 then wanted to analyze the performance of the ML and CMAQ model **(R2, R3)** during this extreme weather condition by checking the *IOA* chart as shown in Figure 5(b2). He noted that IOA values on monitoring stations for CMAQ was low and ML model was high. E2 stated that he expected the CMAQ to have low performance but was surprised to note the performance of the ML model. E2 checked the *Spatial map* (Figure 5(b2)) , to analyze the performance of the models in terms of capturing spatial patterns. From observing Figure 5(b2), E2 mentioned that CMAQ model's forecast has acceptable spatial patterns than the ML model as CMAQ's spatial map has rapid changes in color gradients in the grid cells which correlates with behaviour of O3 pollutant. After examining the *Performance view* (Figure 5(B)), E2 selected a location in the center of the city from the *Locationwise Overview* (Figure 5(a3)) to understand its instance-level feature contribution. From the *Feature Contribution* view Figure 5(d) , E2 noticed that the wind from the east has high importance. Further observing the spatiotemporal importance of the input features in Figure 5(E), E2 noticed that grid locations from past hours 1 and 2 has relatively high importance than the later past hours **(R5)**. The above observation indicates that sudden changes in wind speed and directions has happened during these nearby timestamps. So, E2 moved to the *Raw data View* (Figure 5) to check the wind trajectory animation and validate the contribution information shown **(R6)**. E2 observed that wind indeed moved from the east to the west for the input time period during which its direction changes rapidly, as seen in Figure 5(C). Through this case, E2 understood how the ML model behaves during extreme weather and the information captured by it. E2 concluded that the ML model was able to capture the sudden change of the pollutant in extreme weather conditions better than the CMAQ model. He also said that while in the spatial dimension, CMAQ maintained a better spatial consistency.

6.2 Expert Interview

For the expert interview, we invited two (E4, E5) domain experts, who were not involved in the formative study and case studies, to evaluate the system based on its usability and effectiveness. E4 is a researcher who predominantly works on modeling regional or local air quality and is interested in understanding the ML model's behavior on air pollutant datasets, and E5 is also a researcher interested in XAI for environmental science.

Procedure. The interview was a semi-structured one conducted with experts separately, each of which lasted for 50 minutes. We first introduced the objective of the research, the data we used, and our visualization system (AQX). After this, we presented the case study found by E1 and E2. Followed by this, we invited the experts to explore and analyze the functionalities of the system. Finally, we collected their feedback regarding the visual designs, interactions, and the overall usability of the system. We summed up our observations and the experts' feedback as follows.

System Usefulness Both the experts mentioned that AQX is a useful system for verifying domain knowledge. The experts commented that the *Overview* helps them see the overall pattern and narrow the analysis to the point of interest. They further stated that separating the feature contribution and aggregating it in different levels of temporal and spatial resolution is intuitive. Furthermore, it aids in verifying domain knowledge that can be observed in temporal and spatial dimensions separately. However, they also mentioned that the system must include interactions to facilitate the comparison of different features' contributions in *Daywise Overview*. E4 also mentioned that *"It is difficult to remember so much information when comparing with the contribution information in other month."* expressing their difficulty in remembering when comparing contribution information of all months in *Daywise Overview*. The experts commented that the *Performance View* gives them a high-level understanding of what the model can and cannot learn. This view helped them quickly understand that even though the ML model's forecast accuracy is better than the simulation model on the monitoring stations, and it fails to capture the pollutant's spatial patterns that the simulation model can achieve. Both the experts stated that the *Feature Contribution View* is intuitive and easy to understand. E3 stated that *"The bar chart showing the importance of various features is easy to understand and is indeed a useful information to visualize."* For the *Spatio-Temporal Contribution* view, E4 stated that it is interesting to know how the spatial importance changes over the input time period and how the grid locations in the direction of wind have high importance value as shown in the case study. E5 added that this view helps her understand how the ML model uses the spatial attribute of the input feature. In terms of *Raw data View*, both experts commented that visualizing wind trajectory to verify the other view's information is intuitive. E4 stated that *"Air pollution forecast is a complex mechanism. It depends on multiple features like physical, chemical and geographical. However, the wind is regarded as an important feature though we might not have statistical data to support because of the changes in climatic condition. So visualizing wind trajectory to support the feature importance information is cool and intuitive."*

Visual Designs and Interactions. We also collected feedback regarding various visuals used in the system. Both experts felt that system navigation is easy to understand and follow. They also mentioned that the visualizations were elegant and self-explanatory. E5 said, *"the Feature Contribution view is very useful and easy to understand which feature is more important for the forecast."* The experts complimented the interactions supported by the system. E4 appreciated the zoomed display of the target grid location in the spatial importance view upon selecting the target grid location in the spatial map view. He further added that *"The zoom option helps me to gain a clear picture of spatial contribution information."*. However, E5 suggested that the system should incorporate some visual cues to indicate the available interactions. She stated that *"For the Spatial-Temporal importance view shows the spatial importance for past 12-hours, the system should indicate that the view is scroll-able to reveal the spatial importance at later time stamps."*

7 DISCUSSION

In this section, we discussed the social impact of our research, the limitations of the designs, and the possible future work for the study.

7.1 Social impact of the research

In this paper, we propose a visual analytics system to aid domain experts in verifying the ML model's learning with their domain knowledge. Air pollution is critical domain. And forecasting, analysing and monitoring air quality is essential for policy making to maintain a healthy environment. If ML models can produce highly accurate forecasts, then verifying and validating domain knowledge can aid in establishing appropriate trust in ML solutions or approaches. Moreover, this can further increase the possibility of adopting ML solutions to air pollution and other domains instead of adopting it without understanding or completely ignoring it because of its black-box nature.

7.2 Limitations and future work

Scalability Some of the views like the *Feature Contribution* view might suffer from scalability issues when the number of input features increases. For example, in the *Feature Contribution View*, it might not be easy to visually differentiate features using color when the number of features increases to more than nine.

Visual Cognitive Load The system currently supports the exploration of the contribution of features one at a time. However, the domain experts in the expert interview expressed that having interactions to facilitate the comparison of the contribution of multiple features in the *Overview* might be helpful. This design requirement can be added to R1 to enhance its usability in future work. The experts also mentioned that they have to retain a lot of information in the memory to compare the contribution of a feature across different months in the *Daywise Overview*. This increases the cognitive load in the users.

Limited number of subjects The limited number of domain experts is due to the availability of experts in air quality field within our university. However, the experts in the formative study and final evaluation were deeply involved throughout the process. The feedback and design requirements derived from the formative study can hold valid despite the number of subjects involved in the study, while including more subjects might help in fine-tuning the design requirements.

Scope of the paper The scope of the study is to develop a visual analytics XAI tool for experts to verify their knowledge. However, the study can be further extended to understand whether domain knowledge verification alleviates the trust in the ML model and whether it persuades the domain experts to use it in their application domain through a large-scale qualitative study. Furthermore, the design requirements and VA system can also be extended to develop a tool for debugging and improving the performance of the ML model, where distilling domain-specific knowledge to the ML model can improve its performance.

Generalizability We discuss the generalizability of the design requirements and the system based on its applicability to other stakeholders as well as other domains. **Other Stakeholders:** Our study considers domain experts with little-to-no technical background as the target users. Political leaders from the environmental bureau with domain knowledge can use the system to understand the ML model's behavior and decide whether to use ML models for policy-making. Domain experts with ML knowledge (Model developers) can utilize the design requirements and the visualization system with modifications to understand the ML model's ability to learn key domain knowledge and further steer and improve the model. Medical professional can also utilize the visual system to understand the behavior of ML models and better use these models to anticipate and manage the health risks related to poor air quality. **Other Scenarios:** *AQX* can be generalized to domains other than air quality domain. Precipitation-nowcasting forecasts future rainfall over a study area with data collected from monitoring stations at regular intervals. This domain uses meteorological data like wind and satellite images to perform the forecast. In particular, wind plays a vital role in precipitation nowcasting. Therefore, with some minor changes in the view, *AQX* can be adopted for precipitation nowcasting. Furthermore, a few design requirements and views of *AQX* can be used in the aerodynamics domain, where prediction of pressure field around the aircraft helps to isolate and localize the source of air acoustics. When an aircraft lands and takes off from a runway, its interaction with the wind changes the pressure around the aircraft and produces noise. The domain experts have critical knowledge about the changes in pressure field depending on wind speed and its direction, which can be analyzed and verified using *Raw data* view, and *Spatio-temporal Contribution* view. In addition, *AQX* is model agnostic and can support feature contribution explanation for any differentiable ML model.

8 CONCLUSION

In this paper, we formulated the need to explain the ML model's learning to domain experts and verify it with their knowledge. We conducted a formative study and identified that feature contributions towards the final air pollution forecast, along with the prediction accuracy and raw data information, are essential for the domain experts to verify their knowledge. We introduced AQX, a visual analytics system designed to help experts validate and verify the ML model's learning with their domain knowledge. We presented two case studies and expert interviews to demonstrate the effectiveness and usefulness of the proposed system. The feedback from the experts states that AQX has helped verify and validate their knowledge. As a future work, we want to conduct more longitude studies on improving the tool to build domain experts' appropriate trust in ML models and the awareness of the risk.

ACKNOWLEDGMENTS

This research is supported in part by the HSBC 150th Anniversary Charity Programme through the PRAISE-HK project under grant HKBF17RG0. We would like to thank Yuzhe Jiang, Zhihua Jin, Furui Cheng, Yumiao Zhang, Meilan Wang, Li Lu for their valuable contribution for the paper.

REFERENCES

[1] Amina Adadi and Mohammed Berrada. 2018. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access* 6 (2018), 52138–52160.

[2] Antoine Alléon, Grégoire Jauvion, Boris Quennehen, and David Lissmyr. 2020. PlumeNet: Large-scale air quality forecasting using a convolutional LSTM network. *arXiv preprint arXiv:2006.09204* (2020).

[3] Gennady Andrienko, Natalia Andrienko, Wei Chen, Ross Maciejewski, and Ye Zhao. 2017. Visual analytics of mobility and transportation: State of the art and further research directions. *IEEE Transactions on Intelligent Transportation Systems* 18, 8 (2017), 2232–2249.

[4] Gennady Andrienko, Natalia Andrienko, Urska Demsar, Doris Dransch, Jason Dykes, Sara Irina Fabrikant, Mikael Jern, Menno-Jan Kraak, Heidrun Schumann, and Christian Tominski. 2010. Space, time and visual analytics. *International journal of geographical information science* 24, 10 (2010), 1577–1600.

[5] K Wyat Appel, Alice B Gilliland, Golam Sarwar, and Robert C Gilliam. 2007. Evaluation of the Community Multiscale Air Quality (CMAQ) model version 4.5: sensitivities impacting model performance: part I—ozone. *Atmospheric Environment* 41, 40 (2007), 9603–9615.

[6] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58 (2020), 82–115.

[7] V Athira, P Geetha, Rab Vinayakumar, and KP Soman. 2018. Deepairnet: Applying recurrent networks for air quality prediction. *Procedia computer science* 132 (2018), 1394–1403.

[8] Sagar V Belavadi, Sreenidhi Rajagopal, R Ranjani, and Rajasekar Mohan. 2020. Air quality forecasting using LSTM RNN and wireless sensor networks. *Procedia Computer Science* 170 (2020), 241–248.

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

[10] Daewon Byun and Kenneth L Schere. 2006. Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. (2006).

[11] Carrie J Cai, Jonas Jongejan, and Jess Holbrook. 2019. The effects of example-based explanations in a machine learning interface. In *Proceedings of the 24th international conference on intelligent user interfaces*. 258–262.

[12] Aditya Chattopadhay, Anirban Sarkar, Prantik Howlader, and Vineeth N Balasubramanian. 2018. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 839–847.

[13] Wei Chen, Fangzhou Guo, and Fei-Yue Wang. 2015. A survey of traffic data visualization. *IEEE Transactions on Intelligent Transportation Systems* 16, 6 (2015), 2970–2984.

[14] Weiyu Cheng, Yanyan Shen, Yanmin Zhu, and Linpeng Huang. 2018. A neural attention model for urban air quality inference: Learning the weights of monitoring stations. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

[15] Mark W Craven and Jude W Shavlik. 1996. Extracting tree-structured representations of trained networks. *Advances in neural information processing systems* (1996), 24–30.

[16] Zikun Deng, Di Weng, Jiahui Chen, Ren Liu, Zhibin Wang, Jie Bao, Yu Zheng, and Yingcai Wu. 2019. Airvis: Visual analytics of air pollution propagation. *IEEE transactions on visualization and computer graphics* 26, 1 (2019), 800–810.

[17] Mohamed Ben Ellefi, Zohra Bellahsene, and Konstantin Todorov. 2015. Datavore: a vocabulary recommender tool assisting Linked Data modeling. In *ISWC: International Semantic Web Conference*.

[18] Jerome H Friedman. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* (2001), 1189–1232.

[19] Oscar Gomez, Steffen Holter, Jun Yuan, and Enrico Bertini. 2020. ViCE: visual counterfactual explanations for machine learning models. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 531–535.

[20] Md Naimul Hoque and Klaus Mueller. 2021. Outcome-explorer: A causality guided interactive visual interface for interpretable algorithmic decision making. *arXiv preprint arXiv:2101.00633* (2021).

[21] Kenneth I Joy. 2007. Numerical methods for particle tracing in vector fields. *On-Line Visualization Notes* (2007), 1–7.

[22] Ilias Kalamaras, Ioannis Xygonakis, Konstantinos Glykos, Sigmund Akselsen, Arne Munch-Ellingsen, Hai Thanh Nguyen, Andreas Jacobsen Lepperod, Kerstin Bach, Konstantinos Votis, and Dimitrios Tzovaras. 2019. Visual analytics for exploring air quality data in an AI-enhanced IoT environment. In *Proceedings of the 11th International Conference on Management of Digital EcoSystems*. 103–110.

[23] Niranjan Kamat, Prasanth Jayachandran, Karthik Tunga, and Arnab Nandi. 2014. Distributed and interactive cube exploration. In *2014 IEEE 30th International Conference on Data Engineering*. IEEE, 472–483.

[24] Marilena Kampa and Elias Castanas. 2008. Human health effects of air pollution. *Environmental pollution* 151, 2 (2008), 362–367.

[25] Jintao Ke, Hai Yang, Hongyu Zheng, Xiqun Chen, Yitian Jia, Pinghua Gong, and Jieping Ye. 2018. Hexagon-based convolutional neural network for supply-demand forecasting of ride-sourcing services. *IEEE Transactions on Intelligent Transportation Systems* 20, 11 (2018), 4160–4173.

[26] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. 2008. Visual analytics: Definition, process, and challenges. In *Information visualization*. Springer, 154–175.

[27] Lester B Lave and E Seskin. 1973. Air pollution and human health. *Readings in Biology and Man* 169 (1973), 294.

[28] Doyup Lee, Suehun Jung, Yeongjae Cheon, Dongil Kim, and Seungil You. 2018. Forecasting taxi demands with fully convolutional networks and temporal guided embedding. In *NIPS 2018 Spatiotemporal Workshop*.

[29] Jiwei Li, Xinlei Chen, Eduard Hovy, and Dan Jurafsky. 2015. Visualizing and understanding neural models in nlp. *arXiv preprint arXiv:1506.01066* (2015).

[30] Yuxuan Liang, Songyu Ke, Junbo Zhang, Xiuwen Yi, and Yu Zheng. 2018. Geoman: Multi-level attention networks for geo-sensory time series prediction.. In *IJCAI*. 3428–3434.

[31] Dongyu Liu, Panpan Xu, and Liu Ren. 2018. TPFlow: Progressive partition and multidimensional pattern extraction for large-scale spatio-temporal data analysis. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 1–11.

[32] Zhicheng Liu, Biye Jiang, and Jeffrey Heer. 2013. imMens: Real-time visual querying of big data. In *Computer Graphics Forum*, Vol. 32. Wiley Online Library, 421–430.

[33] Scott Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874* (2017).

[34] Yao Ming, Huamin Qu, and Enrico Bertini. 2018. Rulematrix: Visualizing and understanding classifiers with rules. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 342–352.

[35] Takayuki Miura, Satoshi Hasegawa, and Toshiki Shibahara. 2021. MEGEX: Data-Free Model Extraction Attack against Gradient-Based Explainable AI. *arXiv preprint arXiv:2107.08909* (2021).

[36] Christoph Molnar. 2019. *Interpretable Machine Learning*. https://christophm.github.io/interpretable-ml-book/.

[37] Sayali Nemade. 2019. A Survey on Different Machine Learning Techniques for Air Quality Forecasting for Urban Air Pollution. *International Journal for Research in Applied Science and Engineering Technology* 7 (04 2019), 2185–2194. https://doi.org/10.22214/ijraset.2019.4395

[38] Quoc Phong Nguyen, Kar Wai Lim, Dinil Mon Divakaran, Kian Hsiang Low, and Mun Choon Chan. 2019. GEE: A gradient-based explainable variational autoencoder for network anomaly detection. In *2019 IEEE Conference on Communications and Network Security (CNS)*. IEEE, 91–99.

[39] Huamin Qu, Wing-Yi Chan, Anbang Xu, Kai-Lun Chung, Kai-Hon Lau, and Ping Guo. 2007. Visual analysis of the air pollution problem in Hong Kong. *IEEE Transactions on visualization and Computer Graphics* 13, 6 (2007), 1408–1415.

[40] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1144.

[41] Dominik Sacha, Matthias Kraus, Daniel A Keim, and Min Chen. 2018. Vis4ml: An ontology for visual analytics assisted machine learning. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 385–395.

[42] Sam Sattarzadeh, Mahesh Sudhakar, Konstantinos N Plataniotis, Jongseong Jang, Yeonjeong Jeong, and Hyunwoo Kim. 2021. Integrated Grad-Cam: Sensitivity-Aware Visual Explanation of Deep Convolutional Networks Via Integrated Gradient-Based Scoring. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1775–1779.

[43] Michael Sedlmair, Miriah Meyer, and Tamara Munzner. 2012. Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2431–2440. https://doi.org/10.1109/TVCG.2012.213

[44] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*. 618–626.

[45] Qiaomu Shen, Yanhong Wu, Yuzhe Jiang, Wei Zeng, KH Alexis, Anna Vianova, and Huamin Qu. 2020. Visual interpretation of recurrent neural network on multi-dimensional time-series forecast. In *2020 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, 61–70.

[46] 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. *arXiv preprint arXiv:1506.04214* (2015).

[47] Akshat Shrivastava and Jeffrey Heer. 2020. ISEQL: Interactive sequence learning. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 43–54.

[48] Hyesook Son, Seokyeon Kim, Hanbyul Yeon, Miyeon Lee, Yejin Kim, and Yun Jang. [n. d.]. Visual Deep Learning Models Analysis for Air Pollution Predictions. ([n. d.]).

[49] Thilo Spinner, Udo Schlegel, Hanna Schäfer, and Mennatallah El-Assady. 2019. explAIner: A visual analytics framework for interactive and explainable machine learning. *IEEE transactions on visualization and computer graphics* 26, 1 (2019), 1064–1074.

[50] Hendrik Strobelt, Sebastian Gehrmann, Hanspeter Pfister, and Alexander M Rush. 2017. Lstmvis: A tool for visual analysis of hidden state dynamics in recurrent neural networks. *IEEE transactions on visualization and computer graphics* 24, 1 (2017), 667–676.

[51] Junpeng Wang, Liang Gou, Han-Wei Shen, and Hao Yang. 2018. Dqnviz: A visual analytics approach to understand deep q-networks. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 288–298.

[52] Junpeng Wang, Liang Gou, Hao Yang, and Han-Wei Shen. 2018. Ganviz: A visual analytics approach to understand the adversarial game. *IEEE transactions on visualization and computer graphics* 24, 6 (2018), 1905–1917.

[53] Senzhang Wang, Jiannong Cao, and Philip Yu. 2020. Deep learning for spatio-temporal data mining: A survey. *IEEE Transactions on Knowledge and Data Engineering* (2020).

[54] Yunbo Wang, Zhifeng Gao, Mingsheng Long, Jianmin Wang, and S Yu Philip. 2018. Predrnn++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. In *International Conference on Machine Learning*. PMLR, 5123–5132.

[55] Yunbo Wang, Mingsheng Long, Jianmin Wang, Zhifeng Gao, and Philip S Yu. 2017. Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 879–888.

[56] Yunbo Wang, Jianjin Zhang, Hongyu Zhu, Mingsheng Long, Jianmin Wang, and Philip S Yu. 2019. Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9154–9162.

[57] Daniel Karl I Weidele, Justin D Weisz, Erick Oduor, Michael Muller, Josh Andres, Alexander Gray, and Dakuo Wang. 2020. AutoAIViz: opening the blackbox of automated artificial intelligence with conditional parallel coordinates. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 308–312.

[58] Chenyang Xu and Jerry L Prince. 1997. Gradient Vector Flow: A New External Force for Snakes. In *Proceedings of IEEE International Conference on Computer Vision*. 66–71.

[59] Junbo Zhang, Yu Zheng, Dekang Qi, Ruiyuan Li, and Xiuwen Yi. 2016. DNN-based prediction model for spatio-temporal data. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 1–4.

[60] Pengpeng Zhao, Anjing Luo, Yanchi Liu, Fuzhen Zhuang, Jiajie Xu, Zhixu Li, Victor S Sheng, and Xiaofang Zhou. 2020. Where to go next: A spatio-temporal gated network for next poi recommendation. *IEEE Transactions on Knowledge and Data Engineering* (2020).

[61] Zhiguang Zhou, Zhifei Ye, Yanan Liu, Fang Liu, Yubo Tao, and Weihua Su. 2017. Visual analytics for spatial clusters of air-quality data. *IEEE computer graphics and applications* 37, 5 (2017), 98–105.

A FORMATIVE STUDY:

A.1 Question and Answers

Q1: How to evaluate machine learning model?

- Usually for ST forecasts, especially air quality forecast accuracy is evaluated based on the performance of the model at the monitoring stations, which is measured in terms of IOA (Index of Agreement). And the spatial pattern of the model's forecast, which can be evaluated from the spatial consistency and coherence of the spatial map.
- In terms of IOA, a model with an error of 15 percent or accuracy of above 80 percent is acceptable.

Q1(a): How to evaluate spatial consistency and coherence?

- Spatial coherence/consistency can only be evaluated using domain knowledge as it is location and feature dependent. E.g., Temperature is a large-scale feature, and it has the same value for a large geographical extent. In contrast, air pollutants are small-scale features, and they can have different values in two different streets located in the same area.

- There are no quantitative metrics to evaluate Spatial coherence and consistency. It can be evaluated with the help of domain experts by visualizing the spatial map of the model's forecast. They can help to verify if it is acceptable or not.
- But if there is a need for a baseline for comparison, the spatial map of the CMAQ model's forecast can be used. If the spatial map of the ML model's forecast is similar to the spatial map CMAQ model's forecast, then it can be accepted.

Q2:Why CMAQ is widely accepted?

- Usually, a model is widely accepted depending on the scenario in which the model is being used. There are two traditional ways to forecast meteorological features like temperature. One way: Give the average of measured recordings from previous years at a particular timestamp as the forecast for the same timestamp next year. Second, give the previous hour or previous minute recorded measurement as the forecast for future timestamp (persistent forecast). These two methods can be used alternatively depending on the scenario and the feature we want to forecast. However, CMAQ has better performance than these two methods for all features, which is why it is widely accepted.

Q3: What are the most important features in air quality forecast?

- Air pollutant's concentration depends on multiple factors like emission source, emission duration, meteorological condition, location of the monitoring stations, season, time period of the day, geographical boundaries, etc.
- All the factors/ features are equally important. If we have to point out the most important or influential feature, it should be wind.
- Wind is a critical and highly regarded feature. Even though we do not have statistically significant data to prove this claim given the variations in seasons and climatic conditions, it is a commonly accepted fact that wind is important in scattering the pollutants and thereby affects the forecasts.

Q3(a): How does wind affect air pollutant concentration?

- Wind's influence over air pollutants varies between different seasons and geographical locations.
- In the summer months of May to September, the air quality is affected by wind from the East as it is the prevailing wind direction during the summer season.
- In the winter months of November to February, the air quality is largely affected by wind from the North as it is the prevailing wind direction during the winter season. These are some high-level seasonal patterns that can be observed.
- In terms of high-level spatial pattern, the wind has a higher impact on air quality in open areas (i.e.) places near the sea than the city center.

Q3(b):Does air pollutants affect concentration of another air pollutant?

- PM10 and PM2.5 are positively correlated with each other. PM pollutants and O3 are sometimes negatively correlated and sometimes positively correlated. So, PM and O3 pollutants can influence each other's concentration in a given location and time.
- But PM pollutants are less similar to NO2 and SO2. So, correlation of any kind cannot be observed amongst these pollutants.
- It is because PM10 and PM2.5 are large scales, long term features, and NO2, SO2 are local or small-scale features because of their highly reactive nature.

- NO2, SO2 and O3 doesn't have any correlation. This is because NO2 and SO2 are highly reactive and stay in the air for very short time period, thus they have less contribution for concentration of pollutants like O3, PM pollutants

Q4: Effect of wind on different pollutants?

- Wind has a similar effect on all the five major air pollutants (PM10, PM25, O3, SO2, NO2).
- All the five pollutants are microscopic, and PM10 is the largest among them, but they can still be scattered by the wind for long distances (a few kilometers). The only difference is that SO2 and NO2 are highly reactive, so they exist for a shorter time.

Q5: Any other information about wind?

- As wind flows from one location to another, it might bring pollutants along with it, and this might increase pollutant's concentration in one location and decrease in another location.
- Knowing the wind movement can help understand the source of pollutants or why there is a sudden change in the air quality at a given point of time or location.
- And also, wind moves both spatially and temporally. Usually, air quality in a place will be highly influenced by the wind from the nearest previous timestamps and flowing from the nearest spatial location. As wind flowing from farther timestamps and farthest location lose its speed as it travels and might not have a heavy influence.

Q9: What are the influential features other than wind?

- Temperature is another important meteorological feature. The influence of temperature over air quality can be considered for analysis depending on the data quality (data should be reliable, with no missing or erroneous values).
- As temperature increases, the air moves faster, and thereby the pollutants can get scattered easily.
- Temperature and PM pollutants are positively correlated features.
- But since the temperature is large-scale (remains same for entire study area) and long term (remains same for longer time period) feature and is not as dynamic and fluctuating as wind. It will be more insightful to understand how wind influences air pollutants.