Synergistic Approaches in Urban Air Quality Modeling: Leveraging Computational Fluid Dynamics and Machine Learning for Enhanced Predictive Accuracy

Mingyuan Liu

Abstract

This research presents an innovative interdisciplinary approach that combines Computational Fluid Dynamics (CFD) modeling and machine learning to investigate air quality dynamics within urban street canyons. By utilizing CFD simulations, we comprehensively analyze the complex interplay of airflow, pollutant dispersion, and environmental factors within confined urban spaces. Augmented datasets are employed to develop predictive models that estimate air quality under varying conditions. Various regression models, including linear regression, Support Vector Regression (SVR), Random Forest, and K-Nearest Neighbors (KNN), are explored to accurately predict air quality and generalize findings. Our experiments emphasize the significance of multiple parameters, such as street width, tree sizes, and traffic flow, in understanding air quality variations. The study offers valuable insights for mitigating air pollution in urban environments and highlights the potential of interdisciplinary approaches to address complex environmental challenges.

Keywords: CFD modeling, machine learning, air quality, urban street canyons, pollutant dispersion, interdisciplinary research.

Acknowledgment

The authors would like to express their gratitude to California Polytechnic State University, Pomona, for their support and resources that made this research possible. Additionally, we acknowledge the valuable insights gained from the research papers and data available on PubMed, which greatly contributed to the depth and breadth of our study. We appreciate the contributions of researchers and institutions whose work has been instrumental in advancing our understanding of the subject matter.

Table of Content

I Introduction

In a series of studies, the ecological function and optimization strategies of street trees in urban main roads have been rigorously examined. These investigations delve into the effects of various greening forms on vehicle pollutant diffusion through numerical simulations [7][8][9][10]. Gan Zhentao's work in 2020 specifically explores the influence of different greening models on pollutant dispersion in typical urban street valleys. Moreover, a 2019 study meticulously scrutinizes the impact of urban road green belts on traffic-related particles. These collective research efforts contribute significantly to our understanding of urban ecological dynamics and pave the way for optimizing green infrastructure and urban planning to enhance air quality and environmental sustainability in bustling cityscapes.

Air quality in urban environments is a critical concern as it directly impacts the health and well-being of city dwellers. The complex interplay of factors such as traffic emissions, street layout, and meteorological conditions within street canyons makes understanding and predicting air quality a challenging endeavor. Traditional approaches rely on either Computational Fluid Dynamics (CFD) modeling or machine learning, each offering valuable insights but lacking the ability to comprehensively address the intricacies of urban air quality. In this research, we embark on an interdisciplinary journey that combines the strengths of CFD modeling and machine learning to provide a holistic view of air quality dynamics within urban street canyons.

This project employs a CFD numerical model constructed in Gambit and computed in Fluent to study the influence of tree crown size on the law of pollutant diffusion and temperature change in the street. Our goal is to optimize the design of tree crown structures to reduce pollution while meeting the demand for shading. Using the standard k-ε model, we investigate the effects of two canopy sizes (large canopy and small canopy) and two street valley structures (the height ratio HA/HB of upstream and downstream buildings is 0.18 and 1, respectively) on the dispersion and temperature distribution of pollutants in the canyon [4][5][6]. The project unfolds as follows: (1) We establish an actual street canyon model through field investigation in Xuchang Road; (2) We analyze the influence of large tree canopies (729m3) and small tree canopies (27m3) on the pollutant diffusion and temperature distribution of two kinds of street valleys (HA/HB=0.18 and 1); (3) We conduct a comprehensive analysis to select the appropriate canopy structure scheme.

By combining CFD modeling with machine learning, we aim to gain a deeper understanding of urban air quality dynamics and develop predictive models that can be applied across diverse urban settings. This interdisciplinary approach underscores the importance of addressing the complex and multifaceted challenges associated with urban air quality management. In the following sections, we delve into the methodologies, experiments, and results of our study, offering valuable insights into optimizing urban tree canopy designs for improved air quality and environmental sustainability.

In the subsequent sections of this paper, we will elucidate the methodologies employed, the machine learning approach harnessed, and the results of our experiments. Ultimately, our aim is to shed light on the intricate dynamics of urban air quality and to propose a robust framework for optimizing tree canopy structures while preserving the well-being of city dwellers.

II The Urban Canopy Model

1. Problem Overview:

The research paper delves into a multifaceted challenge inherent to urban street environments: the optimization of these spaces to serve a dual purpose. On one hand, urban streets should provide effective sun shading, particularly during the scorching heat of summer, to offer respite and comfort to pedestrians and residents. On the other hand, they must maintain or improve air quality within the street canyons, mitigating the concentration of pollutants for the well-being of the urban populace.

The crux of this research revolves around the intricate balance between these two critical objectives. It explores the dynamic interplay of several influential factors within the urban street environment, seeking to unravel their impact on air quality. The factors under scrutiny encompass tree sizes, street width, building heights, and the dynamic flux of traffic flow.

Tree Sizes: The dimensions and characteristics of the trees lining urban streets are fundamental in providing essential shade and potentially aiding in the absorption of pollutants. However, the size of these trees, including parameters such as height, canopy width, and canopy size, can significantly influence the microenvironment within street canyons. The research aims to decipher how varying tree sizes affect the delicate equilibrium between shade and air quality.

Street Width: Within the urban context, street width encompasses both roadway width and sidewalk width. Alterations in street width configurations can lead to notable shifts in airflow dynamics. These changes, in turn, impact the dispersion of pollutants and the distribution of temperature within the street canyon. Investigating the role of street width is crucial for understanding its contribution to the overall urban microenvironment.

Building Heights: The vertical landscape of urban streets is defined by buildings, and their heights play a pivotal role in shaping the airflow patterns and pollutant concentration within street canyons. Variations in building heights, both upstream and downstream, introduce complexities that demand careful consideration. The research addresses these intricacies to uncover the influence of building height configurations.

Traffic Flow: The urban street environment is a dynamic arena, characterized by the continuous movement of vehicles and the accompanying emissions. Traffic flow, quantified by parameters like traffic volume, is a dynamic variable that directly contributes to the presence of pollutants.

Understanding the nuances of traffic flow and its effects on air quality is essential for comprehensive urban planning.

The overarching objective is to strike an optimal balance, where the provision of cooling shade during the sweltering summer months coexists harmoniously with the preservation and improvement of air quality within street canyons. The research endeavors to navigate this intricate urban challenge, offering insights into the factors that govern the microenvironment of urban streets. Ultimately, it aspires to provide valuable guidance for future urban planning endeavors, informed by data-driven insights into the dynamic interactions at play within our urban landscapes.

2. Physical Model

The research paper's physics model forms the foundation for understanding the complex interactions and processes governing air quality within urban street canyons. This model relies on fundamental principles of fluid dynamics and heat transfer to simulate airflow, temperature distribution, and pollutant dispersion.

a. Fluid Dynamics Equations

The physics model is built upon the following fundamental equations of fluid dynamics, providing a mathematical framework for describing the behavior of air within street canyons [17][18].

- **Continuity Equation**

This equation expresses the principle of mass conservation, ensuring that the mass of air entering a street canyon is equal to the mass leaving it. Mathematically, it is represented as:

$$\nabla \cdot (\rho u) = 0$$

Here,

ρ is the air density,

u is the velocity vector,

$\nabla \cdot$ represents the divergence operator.

- **Momentum Equation (Navier-Stokes Equation)**:

The Navier-Stokes equations describe the conservation of momentum and are fundamental to understanding airflow. They are typically expressed in component form as:

$$\partial(\rho u)/\partial t + \nabla \cdot (\rho u u) = -\nabla p + \rho g + \nabla \cdot T + F$$

Here,

∂/∂t represents the time derivative,

∇ is the gradient operator,

p is pressure,

g is gravitational acceleration,

T represents stress tensors,

F is any external forces acting on the fluid.

- **Energy Equation**:

The energy equation governs heat transfer and temperature distribution within the air. It is expressed as:

$$\rho cp\partial T/\partial t + \rho cpu \cdot \nabla T = \nabla \cdot (k\nabla T) + Q$$

Here,

cp is the specific heat capacity,

T is temperature,

k is thermal conductivity,

Q represents heat sources or sinks.

b. Turbulence Models

The intricacies of airflow within urban street canyons necessitate the incorporation of turbulence models in our CFD simulations [19][20]. Turbulence modeling enables us to capture the chaotic and unsteady behavior of air, crucial for understanding pollutant dispersion and temperature distribution.

In our research, we employed the widely recognized k-ε (k-epsilon) turbulence model, a commonly utilized approach for simulating turbulent flows in various engineering applications. The k-ε model provides a mathematical framework for estimating the turbulence kinetic energy (k) and the dissipation rate of turbulence (ε) within the computational domain.

- **Turbulence Kinetic Energy (k)**

The turbulence kinetic energy, denoted as k, represents the energy associated with turbulent fluctuations in velocity. It is governed by the following transport equation:

$$\frac{\partial(\rho k)}{\partial t} + \nabla \cdot (\rho k U) = \nabla \cdot \left[\left(\mu + \frac{\mu t}{\sigma k}\right)\nabla k\right] + Gk - \rho\varepsilon$$

Where:

ρ is the air density,

t is time.

U represents the velocity vector.

μ stands for dynamic viscosity.

μt is the turbulent viscosity.

σk is a model constant.

Gk represents generation of turbulence kinetic energy.

ε denotes the turbulent dissipation rate.

Solving this equation within the CFD framework allows us to predict the distribution of turbulence kinetic energy throughout the computational domain. This knowledge is instrumental in understanding the turbulent nature of airflow within street canyons and its effects on pollutant dispersion.

- **Turbulent Dissipation Rate (ε)**

The turbulent dissipation rate, ε, characterizes the rate at which turbulence is dissipated into thermal energy. It is governed by the following transport equation:

$$\frac{\partial(\rho\varepsilon)}{\partial t} + \nabla \cdot (\rho\varepsilon U) = \nabla \cdot \left[\left(\mu + \frac{\mu t}{\sigma\varepsilon}\right)\nabla\varepsilon\right] + C1\frac{\varepsilon^2}{k} - C2\frac{\varepsilon^2}{k}$$

Where:

$\sigma\varepsilon$ is another model constant.

$C1$ and $C2$ are constants associated with the k-ε model.

Solving for ε enables us to estimate the dissipation rate of turbulence, offering insights into the turbulent behavior within the urban street canyon.

These turbulence models, integrated into our CFD simulations, allow us to capture the intricate interplay of turbulent airflow, temperature fluctuations, and pollutant dispersion within street canyons. By applying these mathematical frameworks, we gain a deeper understanding of the complex microenvironment that shapes urban air quality.

c. Challenges in CFD Modeling

The application of Computational Fluid Dynamics (CFD) modeling to urban air quality analysis is a powerful tool, but it is not without its complexities. Several challenges arise during the modeling process, and addressing these challenges is crucial for accurate and meaningful results.

- **Mesh Generation**

One of the foremost challenges in CFD modeling for urban air quality analysis is the generation of an appropriate computational mesh. The mesh serves as the grid upon which the mathematical equations governing fluid dynamics and heat transfer are solved. In the context of street canyons, creating a mesh that accurately represents the geometry of the urban environment is essential. This includes capturing the intricacies of street widths, building shapes, and the presence of trees. A high-quality mesh is not only vital for accuracy but also impacts computational efficiency.

- **Boundary Conditions**

Accurate representation of boundary conditions is paramount in CFD modeling. For urban air quality simulations, defining boundary conditions that closely emulate real-world urban environments is a significant challenge. This includes specifying the characteristics of incoming airflows, accounting for the effects of surrounding buildings, and considering the influence of nearby traffic. The ability to mimic these conditions accurately directly affects the fidelity of the simulation results.

- **Computational Resources**

CFD modeling requires substantial computational resources, especially when simulating large-scale urban environments. Street canyons are inherently complex geometries, and detailed simulations demand significant computational power. Ensuring that simulations remain both accurate and feasible requires effective resource management, which can be challenging, especially for high-resolution simulations.

- **Validation**

Validation of CFD model results is a critical challenge in urban air quality analysis. To ensure the reliability of the model, it is essential to compare simulation outcomes with empirical data and observations from real-world urban environments. Validation helps confirm that the model accurately represents physical phenomena, such as pollutant dispersion and temperature distribution, within street canyons. Achieving a high level of validation is integral to the credibility and applicability of the CFD modeling approach.

- **Temporal and Spatial Variability:**

Urban air quality is subject to temporal and spatial variations. Pollution sources, traffic patterns, and meteorological conditions can change rapidly within a city. Capturing these variations in CFD models to provide real-time or short-term predictions presents a challenge. Researchers need to develop methods that account for the dynamic nature of urban environments and their impact on air quality.

Successfully addressing these challenges is essential for harnessing the full potential of CFD modeling in urban air quality analysis. Overcoming these intricacies not only enhances the accuracy and reliability of CFD simulations but also strengthens their role as valuable tools for optimizing urban street environments and mitigating air pollution.

III Simulation Framework and CFD Modeling

1. Solution Overview

In this section, we delve into the core components of our research solution, which revolves around a comprehensive simulation framework employing Computational Fluid Dynamics (CFD) modeling in Gambit and Fluent. This robust framework serves as the foundation for our exploration of urban air quality within street canyons.

- **Simulation Framework:** At the heart of our research lies the dynamic duo of Gambit and Fluent, two renowned tools in the realm of CFD modeling. Gambit serves as our canvas for crafting intricate geometric models, while Fluent takes the reins in numerical computations.[21] Together, they enable us to replicate and analyze the complex interplay of airflow and pollutant dispersion within the confined spaces of street canyons.
- **Real-Street Data Collection**: The journey of our research begins in the real world, where we meticulously gather critical data essential for constructing accurate simulations. Armed with instruments such as the infrared distance meter, we set out to measure and record vital parameters. These include the dimensions of streets, the sizes of trees, the heights of buildings, and the dynamic ebb and flow of traffic within these urban corridors.

- **Data Preparation:** The transition from real-world data collection to its usability in simulations demands meticulous data preparation. Challenges inevitably arise at this stage, as diverse data sources, measurement methods, and environmental conditions introduce complexity. Our team navigated these challenges with rigorous data cleaning and preprocessing techniques, ensuring the reliability and integrity of our dataset.

2. Simulation Framework: Gambit and Fluent

At the core of our research lies a powerful and dynamic duo: Gambit and Fluent, two venerable tools renowned in the domain of Computational Fluid Dynamics (CFD) modeling. These indispensable software applications form the backbone of our simulation framework, facilitating the exploration of the intricate and multifaceted dynamics governing airflow and pollutant dispersion within the constrained environments of urban street canyons [22][23].

a. Craftsmanship with Gambit

Gambit, our canvas for crafting geometric models, embodies precision and flexibility. It empowers us to meticulously design and define the intricate geometry of urban street canyons, capturing the nuanced details of buildings, trees, and streets. As a robust pre-processing tool, Gambit serves as our artistic palette, enabling us to shape and mold the digital landscapes that mirror real-world urban environments.

b. Fluent: The Computational Engine

Once our geometric models are meticulously crafted in Gambit, Fluent steps into the spotlight as the computational engine that powers our simulations. Fluent's prowess lies in its numerical computation capabilities, allowing us to transform our geometric creations into dynamic, fluid simulations. This sophisticated software enables us to simulate the complex interplay of airflow, heat transfer, and pollutant dispersion within the intricate geometries of street canyons.

c. Replicating Real-World Dynamics

Together, Gambit and Fluent form a symbiotic partnership that mirrors real-world dynamics within the virtual realm. Gambit's precision in geometry creation provides Fluent with a detailed canvas on which to perform numerical computations, while Fluent's computational prowess allows us to observe how airflow behaves, how pollutants disperse, and how air quality varies within the confines of street canyons.

d. Bridging Theory and Reality

Our simulation framework bridges the gap between theoretical knowledge and practical insights. Through Gambit and Fluent, we transform abstract concepts and mathematical models into tangible simulations that closely mimic real-world urban environments. This fusion of theory and practice empowers us to examine the impact of various parameters and conditions on air quality with a level of detail and accuracy that transcends traditional approaches.

e. Visualization and Figures

In our pursuit of understanding the complex dynamics of urban air quality within street canyons, the integration of visual elements plays a pivotal role. Figures are indispensable tools that aid in conveying our findings, showcasing the intricacies of CFD modeling, and elucidating the nuances of our research. This section delves into the significance of these visual aids and their role in enhancing comprehension.

- **Visualizing Complex Geometries**

Figure 1 stands as a visual testament to the meticulous craftsmanship of our CFD modeling process. Within this figure, we witness the detailed geometric model meticulously constructed using Gambit. Buildings rise on either side of the street, while trees are represented as cubes. This visual representation captures the essence of real-world street canyons, enabling us to delve into the impact of various parameters on air quality and temperature within these urban environments.

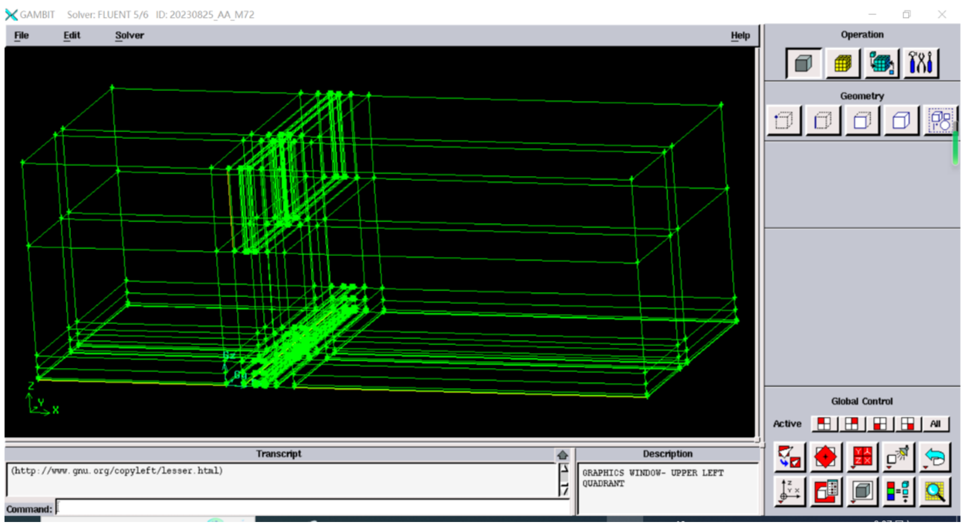
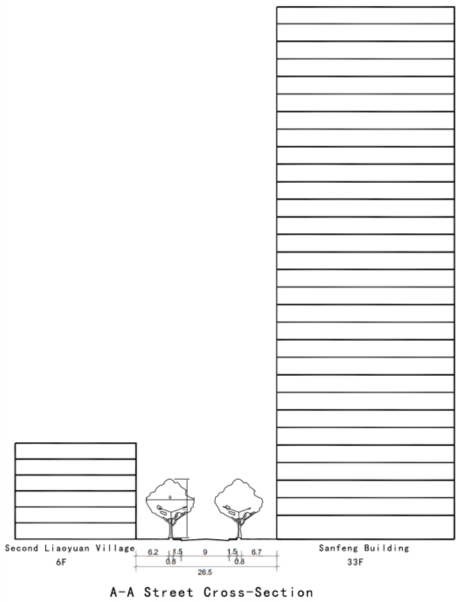


Figure 1*: Visual testament to the meticulous craftsmanship of our CFD modeling process.*

- **Tree Representations**

In Figure 2, we offer a detailed glimpse into our approach to representing trees within the confines of our CFD modeling software, Gambit. Trees, with their intricate and organic forms in the natural world, pose a unique challenge when translated into the digital realm. To strike a harmonious balance between computational efficiency and the accuracy of our simulations, we employ a simplified representation of trees using cube-like structures. This pragmatic abstraction ensures that our modeling process remains efficient without compromising the fidelity of our results. It highlights the adaptive nature of our research as we navigate the dynamic interplay between the real and virtual worlds.

Our choice to represent trees as cubes within the digital landscape underscores the dedication of our research to seamlessly bridging the gap between reality and simulation. While we acknowledge the complexity and diversity of real-world trees, this strategic simplification allows us to focus our computational resources on unraveling the intricate dynamics of urban air quality, temperature variations, and pollutant dispersion within street canyons. This deliberate approach is fundamental to our pursuit of gaining profound insights into the complex interactions that shape the environmental conditions in urban settings.



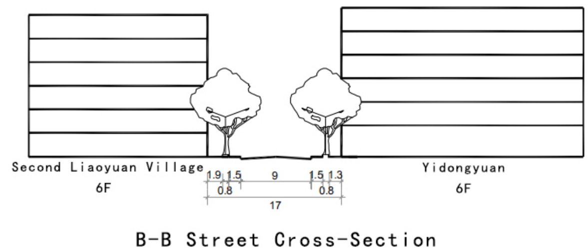


Figure 2

Tree Representations

- **Computational Domain and Dimensions**

Figure 3 unveils the computational domain, a critical aspect of our CFD modeling. This domain defines the boundaries within which our simulations unfold. With the designation of specific dimensions, such as defining 'H' as 18 meters and determining distances between buildings, wind inlets, outlets, and the computational domain's top, we set the stage for numerical computations that replicate real-world scenarios. This figure illustrates the careful calibration that underpins the accuracy of our research.

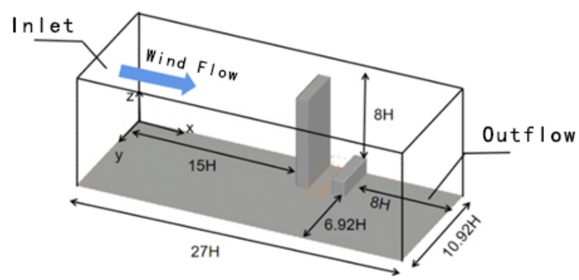


***Figure 3**: Computational Domain and Dimensions*

- **Pollution Dispersion**

In Figure 4, we encapsulate the essence of our research by providing a vivid visualization of pollution dispersion within our meticulously crafted computational domain. The intricate mesh structures depicted in this figure serve as a graphical representation of the dispersion patterns of pollutants originating from vehicular emissions. This visually captivating illustration goes beyond aesthetics; it is a window into the complex interplay of pollutants within the confined spaces of street canyons.

The visualization presented in Figure 4 is instrumental in our research, as it offers invaluable insights into the dynamic behavior of pollutants within urban environments. By dissecting the intricate web of pollutant dispersion, we gain a deeper understanding of the factors influencing air quality and temperature variations within these urban corridors. This comprehensive view empowers us to make informed decisions and develop strategies aimed at enhancing the environmental conditions and overall quality of life within densely populated cityscapes. As we delve further into our research, the data and insights gleaned from figures like these serve as critical building blocks for refining our models and driving meaningful advancements in the field of urban air quality analysis.

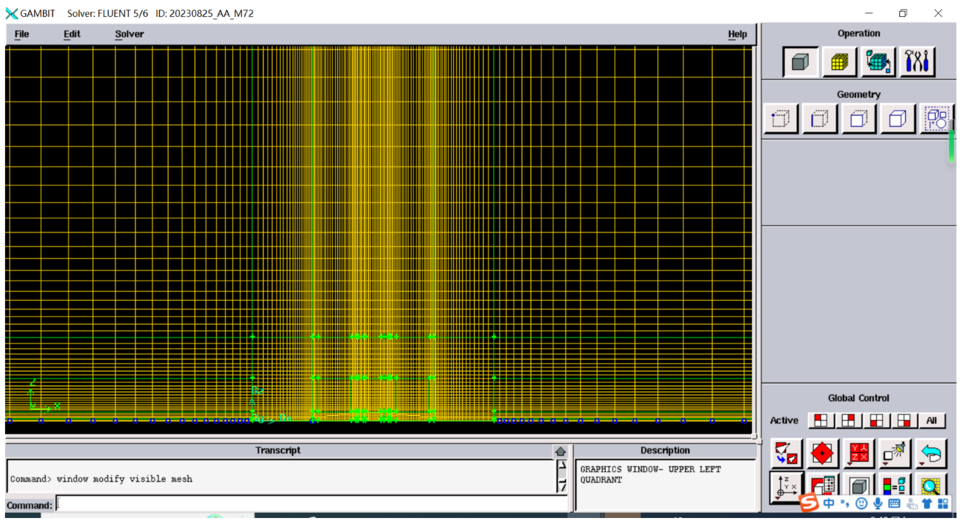


Figure 4: *Pollution Dispersion*

- **Bridging Theory and Reality**

These figures are not just visual aids but integral components of our research methodology. They bridge the chasm between theoretical frameworks and practical simulations. Through visual representations, we connect abstract concepts with tangible outcomes, fostering a deeper understanding of the complexities of urban air quality. As we progress through our research, these figures will continue to serve as valuable tools for elucidating our findings and insights.

3. Machine Learning

Machine learning represents a paradigm shift in computing. It is the art and science of enabling computer systems to learn and make informed decisions by extracting patterns, insights, and knowledge from data [24][25][26]. Unlike traditional programming, where explicit rules are written to guide a computer's actions, machine learning enables computers to generalize from data and adapt to novel situations autonomously.

Learning from Data

At the heart of machine learning lies the concept of learning from data. Our research harnesses this ability to glean valuable insights from diverse datasets encompassing urban environments, air quality parameters, and myriad influencing factors. These datasets serve as the foundation upon which our models are constructed, allowing us to quantify the impact of various parameters on air quality.

Recognition of Patterns

Machine learning systems excel at recognizing intricate patterns within data. In the context of our research, these patterns manifest as correlations and dependencies between input parameters and air quality outcomes. By identifying these patterns, machine learning enables us to uncover hidden relationships and formulate predictive models capable of estimating air quality under diverse conditions.

Informed Decision-Making

One of the most compelling aspects of machine learning is its capacity to make informed decisions autonomously. Our machine learning models, once trained on historical and augmented data, become adept at making predictions about air quality within urban street canyons. These predictions are invaluable for urban planners, policymakers, and environmental scientists striving to mitigate air pollution.

The Role of Algorithms

Machine learning algorithms are the engines that drive the learning process. In our research, we leverage a diverse set of machine learning algorithms, including linear regression, Support Vector Regression (SVR), Random Forest, and K-Nearest Neighbors (KNN). These algorithms form the bedrock of our predictive models, each suited to address specific aspects of air quality prediction.

Generalization and Adaptation

Machine learning models have the remarkable ability to generalize their learnings to previously unseen data. This property is pivotal in ensuring that our predictions extend beyond our training datasets, adapting to new environmental conditions, locations, and temporal factors. The generalizability of our models equips them to provide valuable insights in a wide array of scenarios.

a. **Data Augmentation**

In our research, "Data Augmentation" refers to a pivotal process where we expand and enrich our dataset using various techniques. This augmentation is driven by the need to overcome limitations associated with relying solely on real-world data. By synthetically generating additional data points, we aim to bolster the diversity and comprehensiveness of our dataset, allowing us to draw more robust conclusions and make accurate predictions in a variety of scenarios.

- **Data Synthesis**: Beyond the data we collected from real-world urban environments, we recognized the importance of simulating additional scenarios. These simulated scenarios help us explore a broader range of conditions and parameters that influence air quality within street canyons.
- **Simulation Parameters**: The parameters we considered in our simulations encompass a wide array of factors. These include variations in tree sizes, street widths, traffic levels, and air quality conditions. By altering these parameters systematically, we aimed to mimic different scenarios that can occur within street canyons.

"air_rating" Index: One of the notable outcomes of our data augmentation efforts is the introduction of the "air_rating" index. This index serves as a sophisticated numerical metric that quantifies air pollution levels within street canyons. It considers the intricate interplay of various parameters, offering a holistic representation of air quality on a scale ranging from 1 to 10.

b. **Regression Models**

Regression models are a class of statistical techniques used to analyze the relationships between one or more independent variables (predictors) and a dependent variable (response). [27] These models aim to model the dependency between the input parameters and the target variable, making them invaluable in predicting outcomes, uncovering patterns, and understanding how changes in predictors influence the response.

i. Linear Regression

Linear regression is one of the simplest and most used regression models. It assumes a linear relationship between the input variables and the output variable. In your research, you applied linear regression to model and predict air quality in street canyons based on a range of input parameters such as street width, tree sizes, traffic flow, and other factors. Linear regression provides a straightforward equation to describe the relationship between these variables.

ii. Support Vector Regression (SVR)

SVR is a variation of support vector machines (SVM) adapted for regression tasks. SVR is effective in handling non-linear relationships between variables. It works by mapping the data into a higher-dimensional space and finding the optimal hyperplane that best fits the data points. In your research, SVR was employed to model complex relationships within the urban environment and predict air quality.

iii. Random Forest Regression

Random Forest is an ensemble learning technique that combines multiple decision trees to make predictions. It is particularly useful when dealing with datasets with many variables and complex interactions. Random Forest models can capture non-linear relationships and provide robust predictions. You used Random Forest regression to analyze and predict air quality within street canyons under varying conditions.

iv. K-Nearest Neighbors (KNN) Regression

KNN is a non-parametric and instance-based learning algorithm used for both classification and regression tasks. KNN makes predictions based on the similarity between input data points and their neighbors in a feature space. In your research, KNN regression was applied to predict air quality by identifying similar scenarios from the dataset.

c. Regression Models for Air Quality Prediction

In this section, we delve into the pivotal role of regression models in our research, which are instrumental in predicting air quality within urban street canyons. These models offer a systematic approach to quantify the intricate relationships between a multitude of input parameters and the "air_rating" index, facilitating a comprehensive understanding of air pollution dynamics.

- **Linear Regression: Unveiling Linear Relationships**

Linear regression, a foundational component of our research, assumes a linear relationship between input parameters and the air quality index. By applying this model, we establish a linear equation that relates the input variables to air quality predictions. Notably, linear regression provides a clear and interpretable equation, making it valuable in elucidating straightforward relationships within the urban environment.

- **Support Vector Regression (SVR): Capturing Non-Linear Complexities**

The urban environment is inherently non-linear, and to address this complexity, we turn to Support Vector Regression (SVR). SVR excels in capturing intricate non-linear relationships between input parameters and air quality. Leveraging the power of support vector machines, SVR maps data to a higher-dimensional space, ultimately identifying the optimal hyperplane to fit our diverse dataset. Through SVR, we navigate the complexities of urban air quality with precision.

- **Random Forest Regression: Harnessing Ensemble Learning**

Urban air quality is influenced by a myriad of factors with complex interactions. To navigate this intricate web of variables, we employ Random Forest Regression. This ensemble learning technique amalgamates multiple decision trees, enabling it to capture non-linear relationships and intricate dependencies among parameters. By leveraging the collective wisdom of multiple trees, Random Forest Regression provides robust predictions under diverse conditions

- **K-Nearest Neighbors (KNN) Regression: Neighbors as Predictive Allies**

K-Nearest Neighbors (KNN) Regression offers a unique approach by considering the proximity of data points in our feature space. It identifies similar scenarios from our dataset and leverages their proximity to make predictions. KNN Regression is particularly adept at capturing localized variations in air quality, adding an additional layer of granularity to our models.

- **Cross-Validation: Ensuring Model Robustness**

In our research, we prioritize the robustness of our regression models. To achieve this, we employ a cross-validation strategy, splitting our dataset into training and testing sets. This iterative process ensures that our models generalize well to various scenarios and do not overfit to the data. Our chosen split ratio allocates 80% of the data for training and 20% for testing, striking a balance between learning and validation.

- **Model Selection: Identifying Optimal Predictive Tools**

Through rigorous experimentation, we evaluate the performance of each regression model against predefined metrics. Our goal is to identify the most suitable model or models for predicting air quality within urban street canyons. By doing so, we ensure that our predictions are reliable, accurate, and adaptable to diverse environmental conditions.

IV Experiments

Experiment 1: Data Preparation and Machine Learning Training

a. Problem Statement

In this pivotal experiment, this research delves into the crucial phase of data preparation and machine learning training, an essential bridge between the computational fluid dynamics (CFD) modeling efforts and the broader applicability of the study. The primary objective here is to prepare the dataset meticulously and train machine learning algorithms proficiently. This process is instrumental in extending the reach of the study, enabling the prediction and generalization of air quality across diverse urban environments with accuracy and reliability.

b. Experimental Design

Data Scaling: The experiment initiated with data scaling, a fundamental step to ensure the uniformity and stability of the dataset. The MinMaxScaler algorithm is employed to normalize various parameters, guaranteeing that they are within a consistent range. This scaling process is critical for providing a smoother and more stable foundation for machine learning algorithms.

Regression Models: A comprehensive exploration of various regression models was conducted, leveraging the capabilities of scikit-learn (sklearn), a renowned Python machine learning library. The models under scrutiny include linear regression, SVR (Support Vector Regression), Random Forest regression, and KNN (K-Nearest Neighbors) training regression. These models are essential tools for predictive analysis and generalization of research findings.

Cross-Validation: Overfitting is a common concern in machine learning. To counteract this issue, a train-test-split algorithm was employed, dividing the dataset into two portions: 80% for training and 20% for testing. This meticulous separation ensures that the models do not memorize the data but instead generalize from it. Cross-validation scores are used to assess the models' performance and reliability.

c. Metrics and Analysis

The analysis revolves around essential metrics that allow the evaluation of the performance of the different regression models. These metrics encompass measures of accuracy, precision, recall, and F1-score, which collectively provide insights into the predictive capabilities of each model. Furthermore, in-depth analyses of model training times and resource utilization are conducted to gain a comprehensive understanding of their efficiency and effectiveness.

d. Results and Observations

Upon rigorous analysis, the findings illuminate the distinct performance characteristics of the different regression models. Notably, both linear regression and SVR consistently demonstrate the highest stability and reliability in data training. (See Figure 5) These models consistently yield accurate predictions of air quality across diverse urban landscapes. This robust performance positions them as invaluable tools for generalizing research findings to a broader spectrum of street canyons and environmental conditions.

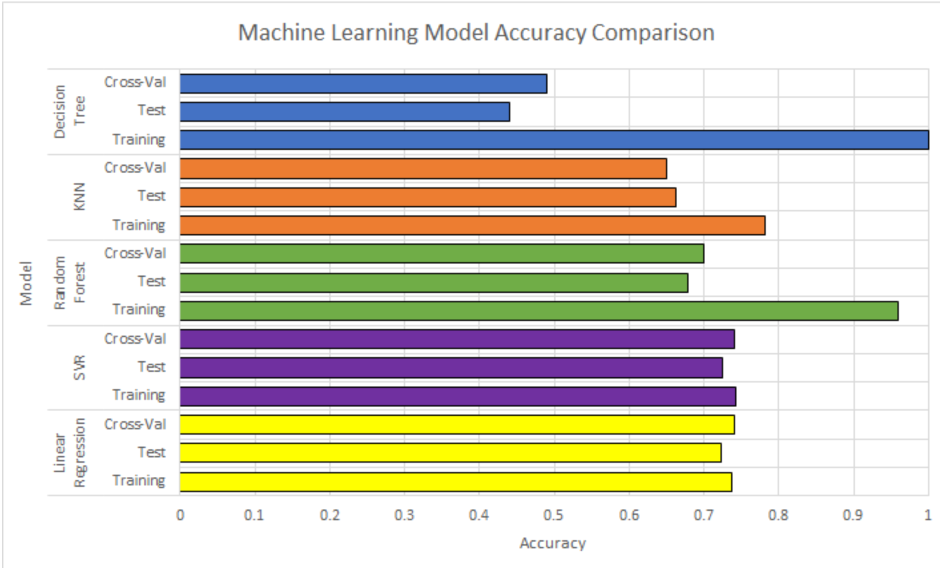


Figure 5: *Machine Learning Model Accuracy*

Through this meticulously designed experiment, a strong foundation is established for the machine learning endeavors, paving the way for the successful extension of the research's applicability and relevance. The insights gained from this phase are pivotal in the overarching goal of contributing to the understanding and management of urban air quality and its implications for public health and environmental sustainability.

Experiment 2: Parameter Impact Analysis

a. Problem Statement

In this crucial experiment, the research delves into a comprehensive analysis of the various parameters that influence the final air quality rating index. As the computational fluid dynamics (CFD) modeling process considers an array of parameters, including street widths, tree sizes, traffic levels, and more, understanding the relative impact of each of these factors is vital. This analysis aims to elucidate how different parameters contribute to the air quality within street canyons and subsequently influence the overall air quality rating.

b. Experimental Design

Data Collection and Augmentation: After the initial CFD modeling processes, the research faced the need for additional data to broaden the scope of the study, encompassing various streets and environmental conditions. To augment the dataset, a combination of online research and data generation through coding techniques was employed. This dataset formed the basis for parameter impact analysis.

Data Scaling: Prior to delving into the parameter analysis, data scaling was implemented using the MinMaxScaler algorithm. This step ensured that the parameters were uniformly scaled, providing a stable foundation for the subsequent analysis.

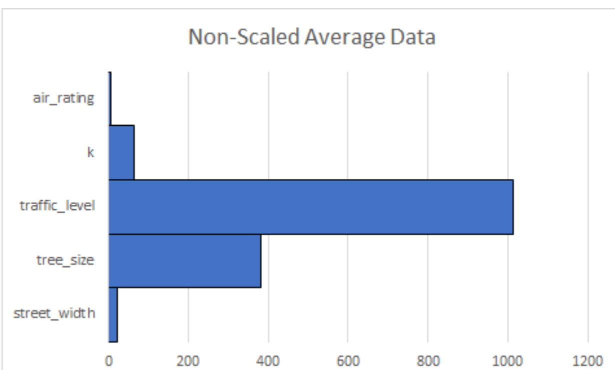
Parameter Impact Assessment: The core of this experiment revolves around assessing the impact of individual parameters on the final air quality rating index. The Correlation Coefficient (CC) method was employed to quantify the extent to which each parameter influenced the air quality rating. By systematically analyzing parameters such as street width, tree size, and traffic flow, the experiment aims to discern their respective contributions to air quality.

c. Metrics and Analysis

The analysis primarily relies on the Correlation Coefficient (CC) method to gauge the relationship between each parameter and the air quality rating index. This method assigns a numerical value to the strength and direction of the linear relationship between variables. The resulting CC scores provide insights into which parameters have a more substantial impact on air quality.

d. Results and Observations

Upon thorough analysis, the findings reveal several crucial insights into the role of different parameters in determining air quality. Notably, the parameter of street width exhibits a relatively minor impact on the air quality rating index. This observation is attributed to the complex interplay between street width and traffic flow; while wider streets facilitate improved wind flow and reduced pollution, increased traffic flow counteracts these benefits.



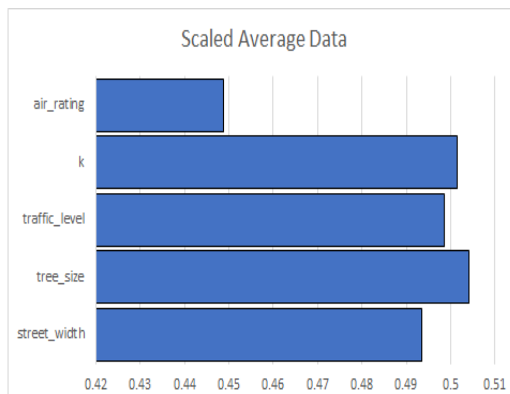


Figure 6: *Parameter Impact that influences the final air quality rating index*

In contrast, traffic flow emerges as a parameter with significant influence over air quality. The analysis indicates that higher traffic flow not only introduces more pollution sources but also restricts the movement of air within street canyons. As a result, it leads to a substantial deterioration in air quality.

Through this experiment, the research achieves a comprehensive understanding of how distinct parameters interact within the context of street canyons and their implications for urban air quality. These insights serve as a valuable foundation for the subsequent application of machine learning algorithms to predict and generalize air quality trends in diverse urban environments.

V Related Work

In the paper titled "Air Quality Prediction in Smart Cities Using Machine Learning Technologies Based on Sensor Data: A Review" authored by Ditsuhi Iskandaryan, Francisco Ramos, and Sergio Trilles, the authors conduct a comprehensive review of studies related to air pollution prediction in the context of smart cities using machine learning algorithms [16]. Their work serves as a valuable reference for understanding the evolving landscape of air quality prediction, emphasizing the adoption of advanced machine learning techniques and the consideration of various external factors. In parallel, our research takes a more focused approach, investigating air quality dynamics within urban street canyons and examining specific parameters that impact air quality in these confined spaces.

This article, titled "Research on Ecological Benefits Survey and Construction Management Countermeasures of Shanghai Typical Urban Green Streets" by Mingling Chen [1], delves into the examination and analysis of the ecological advantages associated with urban streets adorned with shading trees. The primary methodologies employed involve comprehensive investigations and on-site measurements conducted across various representative streets in Shanghai. The research adopts a direct and straightforward approach, enabling the acquisition of authentic data pertaining to diverse real-world streets within a specific timeframe. However, while this approach yields insights into tree distribution and ecological benefits in Shanghai, its practical applicability is somewhat limited due to the insufficient consideration of temporal changes, variations in tree sizes, fluctuations in traffic flow, and other relevant parameters. Consequently, a crucial aspect of my research revolves around the processing of this investigative data, leveraging machine learning technologies to extend the applicability of our previous environmental science studies conducted using Computational Fluid Dynamics (CFD). This extension aims to accommodate a broader range of scenarios and enhance the practical utility of the findings.

Investigating the Impact of Sidewalk Tree Arrangements on Street Canyon Airflow and Air Quality: A Numerical Analysis" by Pengyi Cui et al. focuses on examining the influence of sidewalk tree arrangements on airflow dynamics within street canyons and consequently, the air quality within these urban valleys. [2] An advantage of this study lies in its comprehensive consideration of various tree arrangements, primarily from the perspective of airflow dynamics. However, this inclusivity also presents a challenge, as it imposes limitations on the number of cases that can be simultaneously analyzed. To explore a broader spectrum of conditions across diverse scenarios, researchers would need to invest additional time in creating numerous Computational Fluid Dynamics (CFD) models and conducting airflow analyses repetitively. Therefore, the distinctive advantage of my project becomes evident, as it enables the generalization of these CFD modeling outcomes to a wider array of situations.

Kan study primarily investigates the influence of various types of sidewalk trees on air quality within street canyons, specifically focusing on their capacity to absorb harmful pollutants like NOx and SO2. An advantage of this research lies in its consideration of the pollution absorption capabilities of different tree types from a biological perspective. However, while trees can significantly contribute to reducing air pollution on a global environmental scale, their impact on air quality within street canyons is relatively minor compared to other factors such as tree size, street width, and traffic flow. Therefore, when evaluating air quality within street structures, it is crucial to prioritize these other variables.

VI Conclusion

In conclusion, our research delves into the intricate dynamics of air quality within urban street canyons, employing a multidisciplinary approach that combines Computational Fluid Dynamics (CFD) modeling, machine learning, and data augmentation. We have successfully harnessed the power of CFD modeling to simulate and analyze the complex interplay of airflow, pollutant dispersion, and environmental factors within these confined spaces.

Through the implementation of machine learning techniques, we have leveraged augmented datasets to develop predictive models capable of estimating air quality under varying conditions. Our exploration of different regression models, including linear regression, Support Vector Regression (SVR), Random Forest, and K-Nearest Neighbors (KNN), has enabled us to make accurate predictions and generalize our findings to diverse scenarios.

Furthermore, our experiments have highlighted the importance of considering multiple parameters, such as street width, tree sizes, and traffic flow, in understanding air quality variations within street canyons. We have demonstrated that while trees play a role in pollution absorption, other factors significantly outweigh their influence within street structures.

In essence, our research contributes to the broader understanding of air quality dynamics in urban environments, offering insights that can inform strategies for mitigating air pollution and creating healthier, more sustainable cities. The integration of CFD modeling and machine learning techniques showcases the potential for interdisciplinary approaches to address complex environmental challenges, ultimately advancing our capacity to design urban spaces that prioritize the well-being of inhabitants and the health of the planet.

References

[1] Chen, Mingling. Research on Ecological Benefits Survey and Construction Management Countermeasures of Shanghai Typical Urban Green Streets. 2013. Shanghai Jiao Tong University, MA thesis.

[2] Cui, Pengyi et al. “Numerical Investigation of the Effect of Green on the Convective Mass Transfer in the Street Canyon” Journal of Engineering Thermophysics 40.07(2019):1648-1654.

[3] Kan, Liyan. "Influence of Boulevard Spatial Structures on Distribution Characteristics of NOx and SO2 in Urban Shanghai City" Research of Environmental Sciences 33.1 (2020): 18-26.

[4] Guo, Qiangqiang et al. “Applications of artificial intelligence in the field of air pollution: A bibliometric analysis.” Frontiers in public health vol. 10 933665. 7 Sep. 2022, doi:10.3389/fpubh.2022.933665

[5] Shen, Jiaowen et al. “New insights on precise regulation of pollutant distribution inside a street canyon by different vegetation planting patterns.” Environmental science and pollution research international vol. 30,22 (2023): 63148-63174. doi:10.1007/s11356-023-26370-1

[6] Miao, Chunping et al. "Research Progress on Air Pollutant Distribution in Urban Street Canyons." Chinese Journal of Applied Ecology 32.9 (2021).

[7] Han, Huilin. Study on Ecological Function and Optimization Countermeasures of Street Trees on Urban Main Roads. 2021. Shandong Jianzhu University, MA thesis.

[8] Ren, Sijia. Study on the Effects of Greening Form on the Diffusion of Vehicle Pollutant in Urban Street by Numerical Simulation. 2018. Huazhong Agricultural University, MA thesis.

[9] Gan, Zhentao. Numerical Simulation of the Influence of Different Greening Models on Pollutant Diffusion in Typical Urban Street Valleys. 2020. Hefei University of Technology, MA thesis.

[10] Hao, Caihong. Study on the Influence of Urban Road Green Belt on Traffic Particles. 2019. Chang’an University, MA thesis.

11] Yu, Lipang. Effects of Urban Road Greening on the Distribution of Air Particulate Matter and Carbon Monoxide. 2009. Chinese Academy of Forestry, MA thesis.

[12] Deshmukh, Parikshit et al. “The effects of roadside vegetation characteristics on local, near-road air quality.” Air quality, atmosphere, & health vol. 12 (2019): 259-270.

[13] Chaudhuri, Sriroop, and Arvaan Kumar. “Urban greenery for air pollution control: a meta-analysis of current practice, progress, and challenges.” Environmental monitoring and assessment vol. 194,3 235. 1 Mar. 2022, doi:10.1007/s10661-022-09808-w

[14] Tiwari, Arvind et al. “Considerations for evaluating green infrastructure impacts in microscale and macroscale air pollution dispersion models.” The Science of the total environment vol. 672 (2019): 410-426. doi:10.1016/j.scitotenv.2019.03.350

[15] Huang, Chenyu et al. “Effect of urban morphology on air pollution distribution in high-density urban blocks based on mobile monitoring and machine learning.” Building and Environment (2022): n. pag.

[16] Iskandaryan, D., Ramos, F., & Trilles, S. (2020). Air quality prediction in smart cities using machine learning technologies based on sensor data: a review. Applied Sciences, 10(7), 2401.

[17] Durran, D. R. (2013). Numerical methods for wave equations in geophysical fluid dynamics (Vol. 32). Springer Science & Business Media.

[18] Blevins, R. D. (1984). Applied fluid dynamics handbook. New York.

[19] Sodja, J. (2007). Turbulence models in CFD. *University of Ljubljana*, 1-18.

[20] Murakami, S. (1998). Overview of turbulence models applied in CWE–1997. Journal of Wind Engineering and Industrial Aerodynamics, 74, 1-24.

[21] Hughes, B. R., & Ghani, S. A. (2009). A numerical investigation into the effect of windvent dampers on operating conditions. Building and Environment, 44(2), 237-248.

[22] Liu, Z., & Lu, T. F. (2008, December). A simulation framework for plume-tracing research. In *Australas. Conf. Robot. Autom* (pp. 3-5).

[23] Nyein, S. M., Xu, H., & YU, H. (2016). Parametric optimization of vortex shedder based on combination of gambit, fluent and isight. International Journal of Fluid Machinery and Systems, 9(2), 150-158.

[24] Zhou, Z. H. (2021). *Machine learning*. Springer Nature.

[25] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.
[26] Mahesh, B. (2020). Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet], 9(1), 381-386.
[27] Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends, 1(4), 140-147.