Research in Industrial Projects for Students

Sponsor

AECOM

Final Report

Air Ventilation Assessment

Student Members

Owen LEVIN (Project Manager), *University of Minnesota - Twin Cities*
levin453@umn.edu

Sajant ANAND, *Wake Forest University*

Ke Cheng HSU, *Hong Kong University of Science and Technology*

Ho Nam TSE, *Hong Kong University of Science and Technology*

Acadmeic Mentor

Dr. Chun Yin DY, cydy@connect.ust.hk

Sponsoring Mentor

Dr. Karl AN, Karl.An@aecom.com

Date: August 12, 2018

This project was jointly supported by AECOM, NSF Grant OISE-1460018 and the sponsorship from Dr. Lee Shiu and Dr. Jennie Mui Lee.

Abstract

Air Ventilation Assessment (AVA) in Hong Kong is a set of guidelines for weak wind conditions and is used to evaluate the impact a potential development scheme may have the air ventilation in the vicinity. Whenever a new development is begun, AVA is required to choose the scheme that maximizes pedestrian airflow availability. AVA is done by selecting test points in the schemes and evaluating some quantities related to wind velocity using the values at these points. We show that AVA results can be altered by selecting different sets of points for evaluation, which demonstrates that this method isn't objective. In this report, three novel approaches to solve this problem are discussed. The method of Bias Optimization seeks to obtain an optimal set of test points such that the quantities evaluated are near their "true" values. The next method of Vector Field Similarity identifies the scheme that is the most similar to the one of best air ventilation through wind field topology. The last method of Particle Tracking identifies the scheme that most efficiently disperses particles from the urban area. We intend to combine these three approaches to create a single, unified method that air quality experts in Hong Kong can use when evaluating potential development sites. Once complete, this work will be published in an engineering or environmental science journal.

Acknowledgments

We would like to express our deep gratitude to Professor Tim Leung of the Hong Kong University of Science and Technology (HKUST) and Professor and Associate Director Jorge Balbás of University of California, Los Angeles (UCLA) and the Institute of Pure Math and Applied Mathematics (IPAM) for their collaborative effort to make this wonderful program happen. We would also like to thank our academic mentor Dr. David Dy (HKUST) for his patient guidance, enthusiastic encouragement and useful critiques of this work. We also would like to thank our industry mentors Dr. Karl An (AECOM) and Simon Wong Sze Ming (AECOM) for their help in generating CFD data for air ventilation assessment and providing insightful advice. Finally, we would like to offer special thanks to AECOM, the National Science Foundation (grant number OISE-1460018), IPAM, HKUST Math Department, Dr Lee Shiu, and Dr. Jennie Mui Lee for their financial support.

Contents

Abstract **3**

Acknowledgments **5**

1 Introduction **13**
1.1 Sponsoring Organization 13
1.2 Hong Kong's Air Ventilation Assessment 13
1.3 Problem description 16
1.4 Our data 17

2 AVA Bias Minimization **21**
2.1 Introduction 21
2.2 Methodology 24
2.3 Results 26
2.4 Conclusion and future work 31

3 Vector Field Similarity Comparison **33**
3.1 Introduction 33
3.2 Methodology 33
3.3 Results 38
3.4 Future Work 39

4 Particle Tracking **41**
4.1 Introduction 41
4.2 Particle Tracking in *Fluent* 41
4.3 Results 45
4.4 Conclusions 52

5 Conclusion **55**

REFERENCES

Selected Bibliography Including Cited Works **57**

List of Figures

1.1 Project area and assessment area for the proposed scheme 15
1.2 Sampling of 120 particles for proposed and full scheme. 17
1.3 A histogram of the difference between LVR on the full scheme and LVR on the proposed scheme over 2000 trials of random test point sets. 18
1.4 Base, Full, Empty, and Proposed schemes. 19

2.1 Base scheme distribution of LVR and SVR. 27
2.2 Proposed scheme distribution of LVR and SVR. 28
2.3 The average LVR as minimal spacing changes. 29
2.4 The average SVR as minimal spacing changes. 30

3.1 Particle tracking of 400 particles. 34
3.2 Histograms of *Mean of Curvature* against *Euclidean Distance*. 37
3.3 Histograms of *Velocity Dot Product* against *Euclidean Distance*. 37
3.4 Values of metrics in histogram comparisons, each ran for 10 times. 38

4.1 Setup of particle tracking in Ansys *Fluent*. 45
4.2 Initial and final states of each of the four schemes for linear injection.. . . . 47
4.3 Boxplots for time spent and steps taken, linear injection. 48
4.4 Boxplots for time spent and steps taken excluding trapped particles, linear injection. 48
4.5 Particles vs. time for all four schemes with linear injection. 49
4.6 Initial and final states of each of the four schemes for urban circle injection. 50
4.7 Boxplots for time spent and steps taken, urban circle injection. 51
4.8 Boxplots for time spent and steps taken excluding trapped particles, urban circle injection. 52
4.9 Particles vs. time for all four schemes with urban circle injection. 53

5.1 Prototype of pipeline for scheme evaluation. 56

List of Tables

2.1 Velocity ratios estimated for both the distribution of V and for optimized test points. 31

3.1 Means of the metrics for *Mean of Curvature*. 38
3.2 Means of the metrics for *Velocity Dot Product*. 38

4.1 Particle fates for linear injection trajectories. 46
4.2 Comparison metrics for linear injection. Minimum of each quantity among the four schemes is shown in in red. 47
4.3 Comparison metrics for linear injection, excluding trapped particles. Minimum of each quantity among the four schemes is shown in in red. 48
4.4 Particle fates for urban circle injection trajectories. 50
4.5 Comparison metrics for urban injection. Minimum of each quantity among the four schemes is shown in in red. 51
4.6 Comparison metrics for urban circle injection, excluding trapped particles. Minimum of each quantity among the four schemes is shown in in red. . . . 52

Chapter 1

Introduction

1.1 Sponsoring Organization

AECOM is a multinational engineering company that provides design, consulting, construction, and management services to a wide range of clients. Ranked #161 on the 2017 Fortune 500 list, **AECOM** was named as one of the World's Most Admired Companies for the fourth consecutive year by Fortune. In Hong Kong, **AECOM Asia Company Ltd.** was commissioned by the Planning Department (PlanD) to undertake Air Ventilation Assessments (AVA) for potential improvement or mitigation measures in building design.

1.2 Hong Kong's Air Ventilation Assessment

1.2.1 Background

After the SARS outbreak in 2003, the Hong Kong government became worried about the negative effects on air-flow induced by the density and topology of high-rise buildings in Hong Kong [7]. It was widely believed that such factors contribute the the spread of diseases like SARS and could lead to future epidemics if not properly addressed.

For the purpose of sustainable development and monitoring air flow around urban blocks, the Hong Kong Government commissioned the "Feasibility Study on Establishment of Air Ventilation Assessment" [8]. The study was completed in 2006 and endorsed by the Committee on Planning and Land Development on 7 June 2005. The Committee agreed that government will take the lead to apply Air Ventilation Assessments (AVA) to all major government projects which may have major impacts on the macroscopic wind environment. These include public housing projects, planning studies for new building development, comprehensive redevelopments of pre-existing city regions, preparations of new town plans, and major revision to town plans.

The AVA guide is followed by project proponents to assess the impacts of proposed building developments on the wind environment in their surrounding vicinity at pedestrian level, To pass the assessment, aggregate wind flow must be analyzed on the perimeter of the building, in the surrounding area, and at special points of interest.

1.2.2 Methodology

the AVA is carried out in three stages: expert evaluation, initial study, and detailed study. This section will focus on initial study.

Expert evaluation

Expert Evaluation provides a preliminary examination on design and facilitates the identification of the problems. It is useful and cost effective to perform expert evaluation in the first stage to determine specific wind feature of the site. In this stage, an expert gives a qualitative estimate of wind direction and magnitude based on data obtained from the Hong Kong Observatory on Waglan Island or nearby weather stations [11]. Expert evaluation should identify obvious problem areas of the proposed development and define the focuses and methodology in the following initial and detail study.

Initial study

Following the expert evaluation, the initial study provides a rough quantitative assessment of wind performance at pedestrian level (2 meters), v_p. To do so, v_p is compared to the wind speed at 500 meters, v_∞, where the wind is roughly unaffected by building topology. This is considered the maximum wind available, and is thus the benchmark to which the pedestrian-level wind is compared. The initial study may be performed with wind speed data from computational Fluid Dynamics (CFD) simulation or experimental tests such as wind tunnel, but usually the former is more feasible and widely used for this stage of the assessment.

In the final detailed study, comprehensive wind data from 16 directions must be considered, e.g. North, North-North-East, North-East, East-North-East, East, etc. For the initial study however, it may be appropriate to consider only prevailing wind direction such that the aggregate probability of reduced set of wind directions exceed 75%. The probability of each wind direction is captured in terms of a *wind rose*. A wind rose specifies the wind profile, wind turbulence intensity, and profile of the site. This information can be obtained by conducting large-scale topographical experiments. In this case, the Hong Kong Observatory on Waglan Island wind data should be referenced to for the experimental study.

The developing area should be divided into the project area, the assessment area, and the surrounding area. The project area includes all open areas within project site boundary, while the assessment area and surrounding area include the area up to a distance of H and $2H$ respectively surrounding of project area, where H is the tallest building on site (See Figure 1.1). Occasionally, it is necessary to enlarge the assessment or surrounding area to include some prominent features such as tall buildings or large and bulky obstructions.

After classifying the developing area into three different types, test points are then placed according to the category. Three types of test points may be placed to assess the wind velocity at each site: perimeter, overall, and special.

Perimeter test points are positioned at the project area boundary and are used to measure the immediate effect around the proposed building. Typically, perimeter test points are located 10m to 50m center to center and 30-50 of them are sufficient to measure the wind environment around the scheme.

Overall test points are evenly distributed at the open space of project and assessment area. For practical reasons, around 50 to 80 test points may be adequate for typical development sites.

Using the perimeter and overall test points, one calculates the Site Velocity Ratio (SVR) defined by Equation 1.2 and Local velocity Ratio (LVR) defined by Equation 1.3 as indicators of the wind environment of a site.

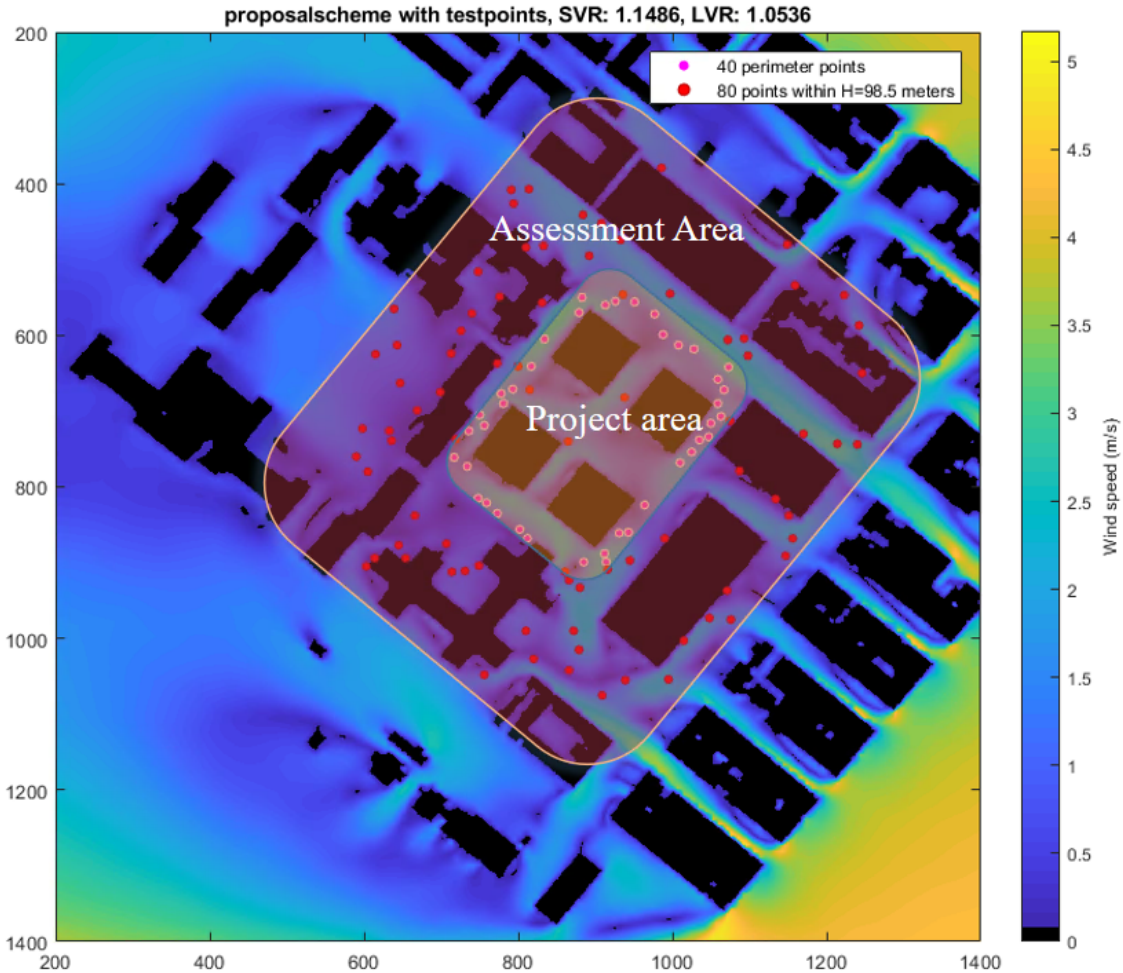


Figure 1.1: Project area and assessment area for proposed scheme. Here, the wind flows from the east and the color indicates the speed.

Definition 1.1 (Velocity Ratio)**.** The Velocity Ratio of a point q is given by

$$\mathrm{VR}(q) = \sum_{d \in D} \left[P(d) \cdot \frac{v_{\mathrm{p}}(q)}{v_{\infty}(q)} \right], \tag{1.1}$$

where D is the set of 16 wind directions being considered, and $v_p(q)$ and $v_\infty(q)$ are the wind speeds at 2 meters and 500 meters high at point q, respectively. $P(d)$ is the percentage of the time that the wind will flow from direction d.

Definition 1.2 (Site Velocity Ratio)**.** The Site Velocity Ratio of a point q is given by

$$\mathrm{SVR}(q) = \frac{1}{|P|} \sum_{q \in P} \mathrm{VR}(q), \tag{1.2}$$

where P is the set of perimeter test point.

Definition 1.3 (Local Velocity Ratio)**.** The Local Velocity Ratio of a point q is given by

$$\mathrm{LVR}(q) = \frac{1}{|P \cup O|} \sum_{q \in P \cup O} \mathrm{VR}(q), \tag{1.3}$$

where P and O are the set of perimeter test points and overall points respectively.

SVR and LVR serve as indicators of the overall wind environment for the comparison of different schemes. SVR measures the immediate effects on the wind environment in the vicinity of a development. On the other hand, LVR gives a hint of influences on the local environment in the overall region of the city where the proposed development is placed.

Due to the limited airflow conditions in Hong Kong, the higher the SVR and LVR, the better the air ventilation. Usually, the development is approved if the LVR and SVR are higher for the proposed scheme. The local air ventilation considerations should always take precedence over the site specific air ventilation considerations.

Finally, special test points are positioned at positions where special wind features are likely to happen such as wind gusts or wind tunnel effects. Unlike perimeter and overall test points which are used to calculate SVR and LVR, special test points are not included in the site and local ventilation assessment as they may distort the average velocity ratios. Special test points independently provide additional information regarding localized problems.

Detailed Study

With or without the Initial Study, the Detailed Study concludes the AVA. Detailed study quantitatively compares the design using indicators mentioned in initial study and provide a summary of identified problems and solutions, if any. Detailed Study is essential for more complex sites and developments, and where key air ventilation concerns have been reviewed and identified in the Expert Evaluation/Initial Study. There are majorly two differences between initial and detailed study. Firstly, it is mandatory to acquire wind availability data experimentally for detailed study. Secondly, we need to take wind profiles from 16 directions when calculating Velocity Ratio.

1.3 Problem description

The current AVA practices are not fully objective because the result can change based on the choice of test points. To demonstrate the problem, we picked 120 random test points (40 perimeter points and 80 overall test points) for the proposed and full schemes 2000 times. An example of a test point configuration can be seen in Figure 1.2. It turns out that over 100 times the LVR of the proposed scheme was worse than that of the full scheme. Similarly, the full scheme SVR was higher than the proposed scheme's SVR over 200 times. Yet, the majority of the time, the proposed scheme wins out as seen in Figure 1.3, a histogram of the experiment's results.

Because the full scheme is essentially a large wind block, it is expected to be a clear loser each time. Since the results of a single AVA test vary depending on where the test points are placed, the current AVA practice of choosing a single set of test points are inconsistent and allow for accidental bias. If only a single set of test points were used, one could end up with either scheme appearing to have better wind availability despite the full scheme having a worse wind environment.

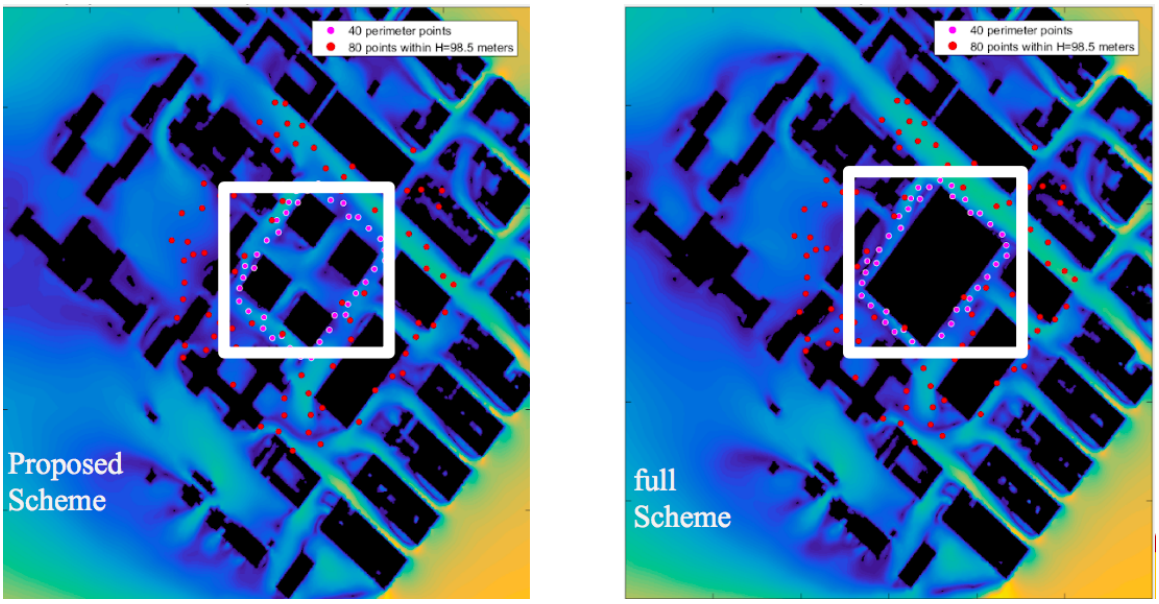


Figure 1.2: Sampling of 120 particles for proposed and full scheme.

Through our work, we will demonstrate methods for both choosing more objective sets of test points for the AVA, and to find good wind availability through methods outside of the current AVA practices.

1.4 Our data

In Figure 1.4 we see the wind speed data visualized for four different schemes which our group used to carry out our analyses. The Base scheme is the set of buildings currently standing in Hong Kong at the development site. The Proposed scheme is an proposed new set of buildings which would replace the Base scheme's buildings on the development site.

The Full and Empty schemes are benchmarks for comparison. The Empty scheme has nothing built on the development site, while the Full scheme has one large building taking up the entire development site.

The wind velocity vector field data for each of those four schemes was generated and provided to us by AECOM. The generation process involved a CDF model which approximated a time-independent steady-state solution to the Navier-Stokes equations.

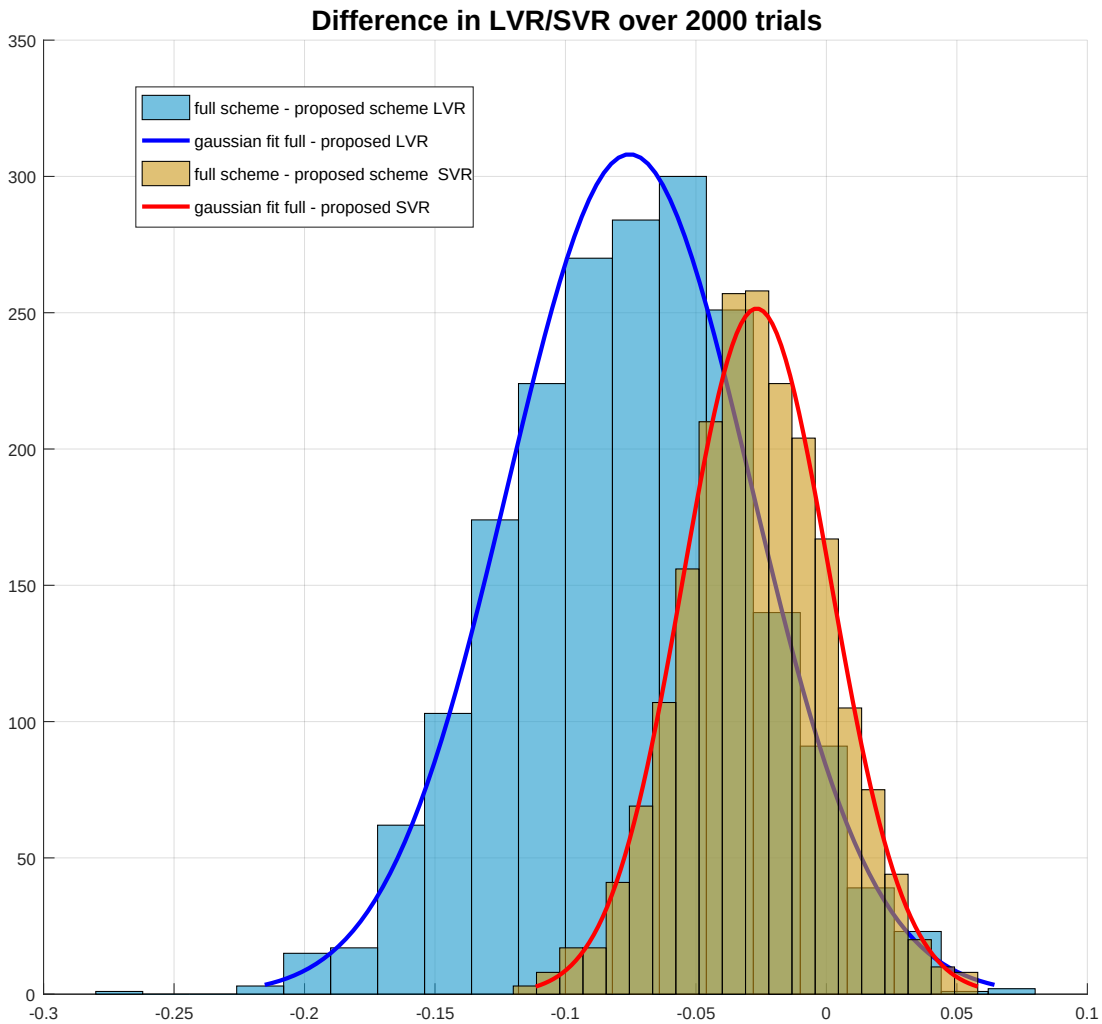


Figure 1.3: A histogram of the difference between LVR on the full scheme and LVR on the proposed scheme over 2000 trials of random test point sets. For each trial, the test points used were the same between the two building schemes.

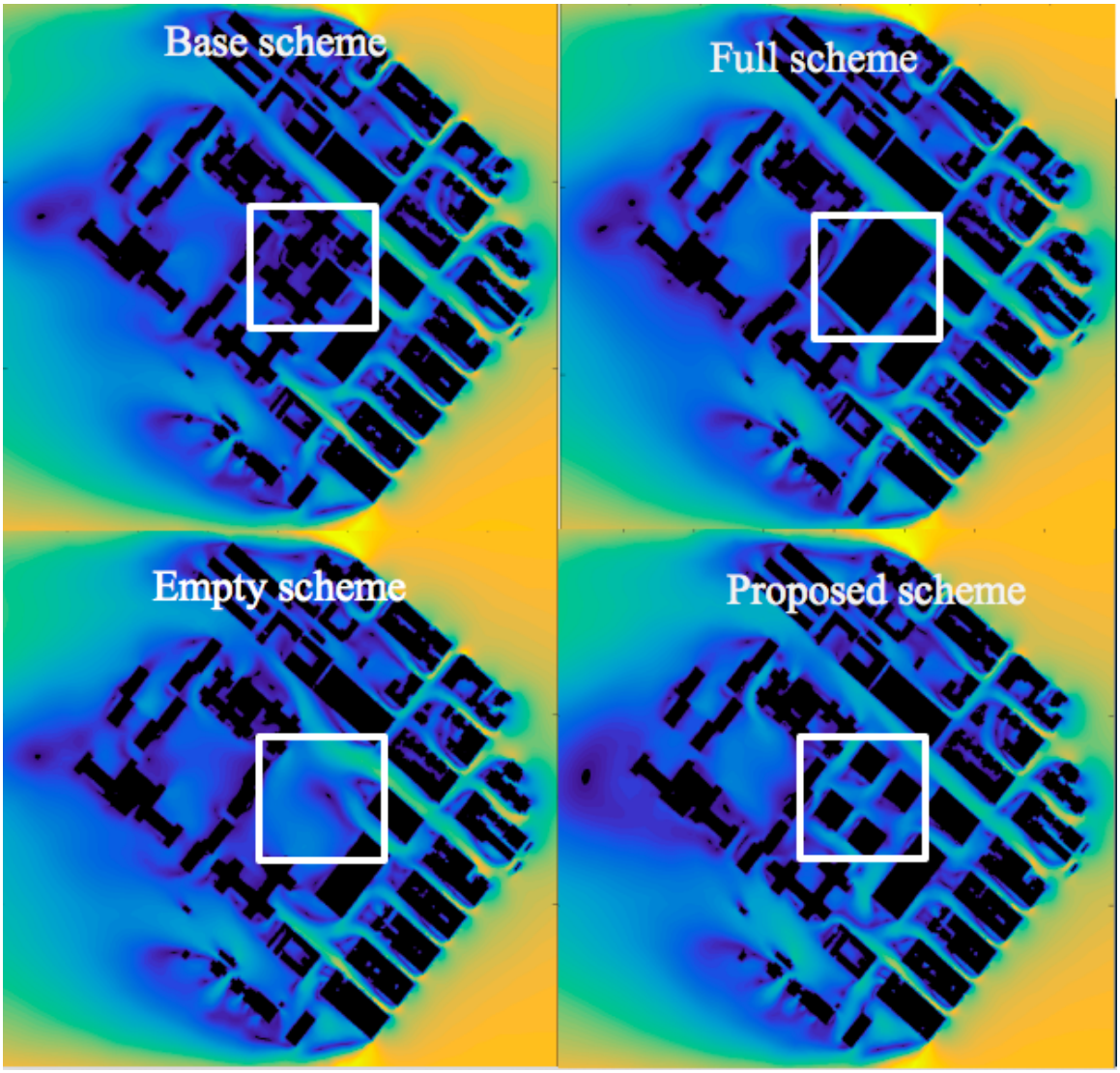


Figure 1.4: Base, Full, Empty, and Proposed schemes.

Chapter 2

AVA Bias Minimization

2.1 Introduction

Even when a biased result is identified, the current practices of the AVA still require statistics evaluated on a set of test points in the final detailed study. In this chapter, we describe an attempt to identify sets of test points that maximize the objectivity of the initial and detailed studies in the AVA. The method can be broken down into two parts.

First, sample the LVR and SVR on both the base and proposed schemes with many possible sets of test points. These samples will converge by the Central Limit Theorem (Theorem 2.7) to a multivariate normal distribution whose mean and covariances can be closely estimated.

The second step is to find the set of test points with velocity ratios aligning most closely with the sampled statistics. Assuming that the average LVR and SVR represent the typical/truest values for a given development scheme, test points aligning well with these averages can be assumed to represent the values of LVR and SVR objectively. On the other hand, a set of test points with LVR or SVR in the tails of the sampled distribution of velocity ratios seems to be inherently picking values that do not coincide with real values. This would lead to a biased assessment.

After discussing some mathematical preliminaries, we shall discuss the method proposed more rigorously.

2.1.1 Mathematical preliminaries

The main background required to understand the method presented in this chapter are the basics of probability theory such as *sample spaces, events, independence, identical in distribution, expectation, probability density functions,* and *cumulative density functions.* We shall also introduce a few extra terms in order to rigorously define the method we propose.

In the AVA, the main quantities of interest are the SVR and LVR of a set of test points on the base and proposed development schemes. These values can succinctly be recorded in a single vector, $V(T)$

Definition 2.1 (VR vector)**.** In particular, given a set T of test points, we will consider

the vector V of aggregate velocity ratios given by

$$V(T) = \begin{bmatrix} \mathrm{SVR}_{\mathrm{base}}(T) \\ \mathrm{LVR}_{\mathrm{base}}(T) \\ \mathrm{SVR}_{\mathrm{prop}}(T) \\ \mathrm{LVR}_{\mathrm{prop}}(T) \end{bmatrix}.$$

If one picks a set of test points uniformly at random from the set of all test points, then one could consider V to be a vector-valued random variable distributed as the velocity ratios of random test points. Our goal is to estimate the distribution of V, then pick a set of test points that approximates the mean of the distribution.

Before we discuss the distribution itself, we should specify a few relevant definitions. A typical measure to see how two given random variables coincide is their *covariance*.

Definition 2.2 (Covariance)**.** The *covariance* of two random variables X and Y is given by

$$\mathrm{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y,$$

where $\mu_X = E[X]$ and $\mu_Y = E[Y]$.

Similarly, the *variance* can measure how close to its average value a random variable usually is.

Definition 2.3. The *variance* of a random variable X is given by

$$\mathrm{var}(X) = \mathrm{cov}(X, X) = E[X^2] - \mu_X^2,$$

However, variance and covariance are each only defined for one dimensional variables. If we have a vector-valued random variable such as V, we instead use a higher dimensional analogue of the variance called a *covariance matrix*.

Definition 2.4 (Covariance matrix)**.** Given a vector $X = (X_1, X_2, \ldots, X_n)$ where each entry is a finite variance random variable with $E[X_i] = \mu_i$ for $1 \leq i \leq n$, then we can define a covariance matrix S as the matrix where the (i, j) entry is is $\mathrm{cov}(X_i, X_j)$. Equivalently,

$$S = E[(X - E[X])(X - E[X])^\top].$$

Consequently, the covariance matrix is a positive, semi-definite, symmetric matrix and the diagonal entries are the variances of each X_i.

Given that, we can define the distribution of repeated samples of V.

Definition 2.5 (Multivariate normal distribution)**.** A mean $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$ and a corresponding covariance matrix S uniquely define a *multivariate normal distribution* denoted $\mathcal{N}(\mu, S)$ with probability density function given by

$$f(x) = \frac{1}{(2\pi)^{n/2}(\det S)^{1/2}} \exp\left[-\frac{1}{2}(x - \mu)^\top S^{-1}(x - \mu)\right].$$

To see that sampling V will converge to a Multivariate normal distribution, we will first specify what we mean by "converge."

Definition 2.6 (Convergence in distribution)**.** Given a random variable Y, let $F_Y(x)$ denote the cumulative distribution function (CDF) of Y. Then a sequence of random variables $X_1, X_2, X_3, \ldots$ is said to *converge in distribution* to a random variable X if $\lim_{n\to\infty} F_{X_n}(x) \to F_X(x)$ on all x where $F_X(x)$ is continuous.

In particular, the CDF of the average of many samples of V will converge to the CDF of a 4-variate normal since we have four aggregate velocities as the components of V. To see this, we will use the following theorem.

Theorem 2.7 (Central Limit Theorem)**.** *Let $V_1, V_2, \ldots$ be independent and identically distributed random vectors with expectation μ and finite covariances*

$$S_{ij} = E[(V_{1,i} - \mu_i)(V_{1,j} - \mu_j)].$$

Then if $X_n = V_1 + V_2 + \cdots + V_n$, we have $(X_n - n\mu)/\sqrt{n}$ converges in distribution to the multivariate normal distribution $\mathcal{N}(0, S)$.

For further variants and proofs of the Central Limit Theorem, one can read most standard probability texts, e.g. Durrett's *Probability Theory and Examples* [6].

Proposition 2.8. *Samples of V satisfy Theorem 2.7.*

We will prove Proposition 2.8 after describing our sampling methodology in the next section. To estimate what μ and S are in the convergent distribution, we use a technique called Maximum Likelihood Estimation. It won't be necessary to rigorously define it here, but the motivating idea is that we know the distribution is multivariate normal by Proposition 2.8. To find the parameters, we pick the μ and S which maximize probability we observed all of the samples we saw. For a more detailed treatment, see any number of statistics, estimation theory, or machine learning text books, e.g. [4]. A standard exercise in estimation is to show that the MLE estimates for μ and S are

$$\mu_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n} V_i$$

and

$$S_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n}(V_i - \mu_{\text{MLE}})(V_i - \mu_{\text{MLE}})^\top.$$

At this point, we can estimate a specific distribution for the samples of V. However, we have no guarantee that any particular set of test points is near μ_{MLE} in every component. this is where the optimization comes in. There is a metric specifically used to measure the distance of a sample from a normal distribution to its mean.

Definition 2.9 (Mahalanobis distance)**.** The *Mahalanobis distance* between an observation $X = (x_1, x_2, \ldots, x_n)$ and a set of samples with mean $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$ and covariance matrix S is given by

$$d(X, \mu) = \sqrt{(X - \mu)^\top S^{-1}(X - \mu)}.$$

For more information on the Mahalanobis distance, one can also check any of many standard statistics or machine learning texts, such as [4]. Roughly speaking, the Mahalanobis distance measures the (multi-dimensional analogue to the) number of standard deviations away a set of test points' velocity ratios $V(T)$ are from the sample average μ_{MLE}. To pick a set of test points, we will attempt to minimize the Mahalanobis distance between $V(T)$ and μ_{MLE} over the set of all test points, $\mathcal{T}$.

2.2 Methodology

Due to the lack of available data for this project, at each point we used the velocity at 2 meters instead of the the velocity ratio between 2 meters and 500 meters. This is essentially just a re-scaling of the same values and would not affect the results of the method presented herein. Hence, *Velocity* and *Velocity Ratio* will be used synonymously.

Moreover, we only had data from one direction instead of 16, so the values only give the velocity ratios over that data instead of the weighted average of the velocity ratios over all directions.

2.2.1 Estimating the sampling distribution of V

To explicitly define the sampling distribution, we must first describe which points can be sampled. Because we can only sample test points using simulation data, all of these methods assume a discrete, finite set of usable points. In an ideal scenario, one would have simulated the wind speed at 500 meters and at 2 meters high on a uniformly spaced grid of a fine enough resolution to capture detail of the buildings and structures in the city. Because we did not have such data, we interpolated the wind speed data we did have to fit a uniform 0.5 meters/unit grid covering 800 meters along each axis, or 1600 units.

With a uniform grid, one may view the speed data as intensities on an image and apply image processing techniques to implement the sampling. Letting $I = ([0, 1600] \cap \mathbb{Z})^2$, for our purposes, image is synonymous with a bounded function $f : I \to [0, +\infty)$.

To sample the points, we created binary masks which evaluate to 1 on pixels which are allowed to be a test point in a given sample and evaluate to 0 otherwise. We can store these efficiently as list of linear indices into the image, then sample a new test point by picking a random element of the list. After adding a point to our set of test points, we must exclude the possibility of picking any new points within a radius of 10 meters from it per the AVA guidelines. To do so, we simply update the mask to evaluate to 0 on pixels within the radius of any previously picked points before sampling a new point. This can be done through either a convolution with a disk or plotting a circle on the image.
The exact sampling procedure to pick one random sample of N test points is given below.

Step 1. Interpolate the speed data to a discrete uniformly spaced grid of pixels.

Step 2. Create a mask M of the valid pixels from which test points may be picked.

Step 3. Initialize the set of test points $T = \varnothing$.

Step 4. Choose a pixel $p \in M$ uniformly at random.

Step 5. Set $T := T \cup \{p\}$.

Step 6. Set $M := M \setminus \{\text{pixels within 10 meters of } p\}$.

Step 7. Repeat Steps 4-6 until T contains N pixels.

We can now prove Proposition 2.8.

Proof. Since no set of test points influences the choice of any point in another set, the samples of test points are independent. We will prove that each sample is identically distributed by induction. To pick a sample set of test points T, we initialize with T as an empty set. For our initial test point added to T, we choose uniformly at random from pixels

in M. So when the size of T is 0 or 1, the samples of V are independent and identically distributed. Moreover, conditioning on having previously sampled an identical set of test points, the distribution of possible points to add will be the same uniform distribution. As a consequence, each set of N test points is sampled from the same distribution since it is identical when there is a single or no points sampled, and each subsequent point is sampled with the same probability given the previous points sampled. Thus, each sample of V is *independent and identically distributed.*

Finally, the covariances are finite since there are a finite number of possible sets of test points we can sample on the discrete grid we use to sample. The covariance matrix for each sample is then also an independent and identically distributed random variable given by the covariance matrix of V if it exists. The covariance matrix will exist as long as there are average values of LVR and SVR for each scheme. We take this as a given when we are working within the AVA framework, since the alternative is that the AVA cannot give an unbiased result. Moreover, empirical analysis corroborates the assumption.

Under the empirically correct assumption that average values for SVR and LVR for each scheme exist, these characteristics prove that samples of V satisfy Theorem 2.7. □

After sampling sufficiently many samples of V, we can compute the estimated mean and covariance of the distribution of velocity ratios for random sets of test points, μ_{MLE} and S_{MLE}. Let n be the minimum number of pixels between points. Then empirically, at least n^2 samples will give a confidence interval of less than ± 0.04 meters per second in estimating μ_{MLE}.

Nevertheless, in most of our experiments, we sampled far more times. The AVA guidelines give a minimum spacing of 10 meters between points, which by the above finding suggests we use at least 400 samples. We always used at least 2000 samples. While this is sufficient for close estimates of the parameters for V's distribution, the sample test points closest to μ_{MLE} always seems to be at least 0.3 units away under the Mahalanobis metric. Even doing 5000 or 10000 samples usually does not give a much closer set of points.

Ideally, we wish to find a set of test points whose velocity ratios on the base and proposed schemes approximate μ_{MLE} nearly exactly. To accomplish this, we run an optimization procedure on the sample T with the smallest $d(V(T), \mu_{\text{MLE}})$. The procedure transforms T into T_{OPT}, a new point set with closer aggregate velocity ratios to μ_{MLE}.

2.2.2 Optimizing the test point set

Given μ_{MLE}, S_{MLE}, and a test point set T, the optimization is done via a simple randomized greedy algorithm. The exact optimization procedure to transform T into T_{OPT} is shown below

STEP 1. Initialize the final point set $T_{\text{OPT}} := \underset{\text{sample test point sets } T}{\operatorname{argmin}} \{d(V(T), \mu_{\text{MLE}})\}$.

STEP 2. Create a mask M of the valid pixels from which test points may be picked.

STEP 3. Pick a random test point $p \in T$.

STEP 4. Find $p' \in M$ minimizing $d(V((T_{\text{OPT}} \setminus \{p\}) \cup \{p'\}), \mu_{\text{MLE}})$

STEP 5. If $p' \neq p$, set $T_{\text{OPT}} := (T_{\text{OPT}} \setminus \{p\}) \cup \{p'\}$.

STEP 6. Repeat Steps 2-5 until there is no $p \in T_{\text{OPT}}$ that can be moved to decrease $d(T_{\text{OPT}}, \mu_{\text{MLE}})$.

A nice property of the above optimization algorithm is that points only move when the distance to the mean can decrease. One consequence is that usually the first few movements will decrease the distance to μ_{MLE} a lot, and afterward, only minor decreases occur until eventually no decreases can be made at all.

Due to the finite space of options, convergence to a local optimum whose distance is bounded above by the initial T is guaranteed. Convergence usually occurs quickly with only 20-30 points actually moved.

Empirically, $d(T_{\text{OPT}}, \mu_{\text{MLE}})$ is typically about 50-100 times smaller than $d(T, \mu_{\text{MLE}})$. How much smaller exactly and how quickly convergence occurs would depend on both the order of points moved and the exact initial conditions. That is, the set T used to initialize T_{OPT} and also the mask chosen for valid movements will affect the answer.

In our implementation, we use one of two masks. Either we use the whole valid sampling area for test points, or we restrict M to the intersection of the valid sampling area and a local area (disk of radius of 30 meters) around p. Both tend to give similar final Mahalanobis distances from μ_{MLE}. That said, the latter gives a resulting point set very similar to the input, T.

It is also possible to have an expert choose an approximate set of test points T giving good coverage of the area, then optimize T with local movements only. This would result in a set of points that both have good coverage of the whole region we care about and approximately reproduce the mean LVR and SVR values of μ_{MLE}.

2.3 Results

Overall, for any given sampling regime, this method successfully produces a set of test points with minimal bias that can be used in wind tunnel experiments for the AVA.

However, there is still freedom to choose the masks from which test points may be sampled. Altering this will inherently change the distribution of V. For example, when points may be sampled anywhere within $2H$ of the development site, we can see in Figures 2.3 and 2.4 that the minimal spacing between test points affects the LVR and SVR means.

In Figure 2.3 it is shown that depending on the minimum spacing allowed between points,the relative ordering of the Local Velocity Ratios of each scheme can be altered.

On the other hand, one can see that restricting the mask to be only within 30 meters of buildings can also change the result. This was the mask used for the LVR comparison in Figure 1.3. With that mask, the proposed scheme's LVR is on par with or above that of the full scheme, unlike the results of Figure 2.3 which used a mask allowing any non-building pixel that wasn't too close to another point as the sampling region.

When producing the data used in Figures 2.3 and 2.4, the maximum number of test points possible were used for each spacing. This sampling method is not representative of the test point distribution used for the AVA, which uses 30-50 perimeter points and 50-80 overall points. Nonetheless, it demonstrates the care that must be used in choosing how the test points are sampled if accurate results are desired. For this reason, we strongly recommend having an expert in the field design the mask choose the parameters used to sample random test points. With these steps taken, the resulting distribution is more likely to mirror the distribution of random test points that would actually be allowed for the AVA.

For a mask restricted to be within 30 meters of buildings, which seems to be closest to the AVA guidelines, Table 2.1 displays a set of optimized test points' velocity ratios for the base and proposed scheme experiments seen in Figures 2.1 and 2.2.

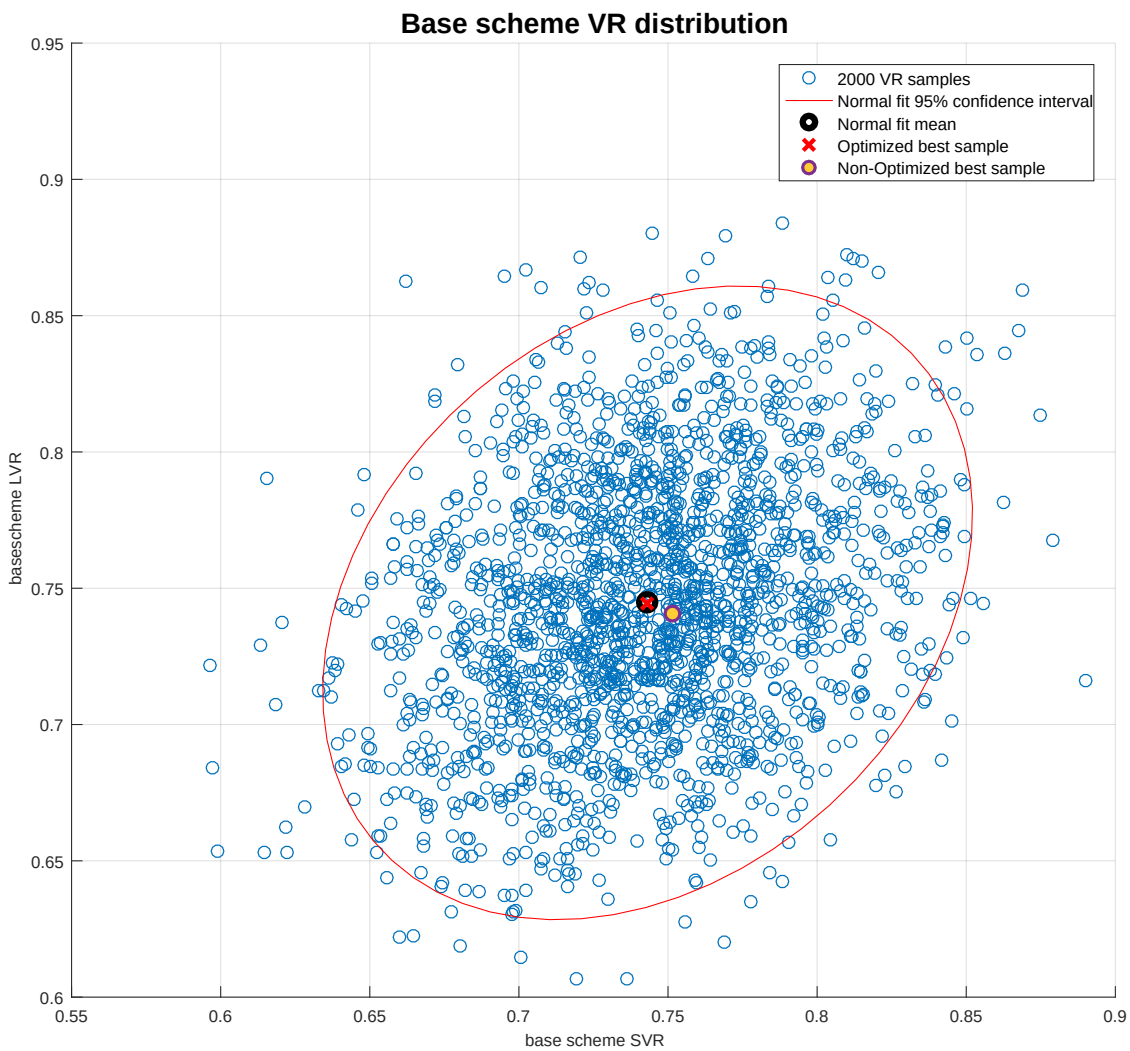


Figure 2.1: A projection of the sampling distribution of V and optimized sample to the base-scheme plane.

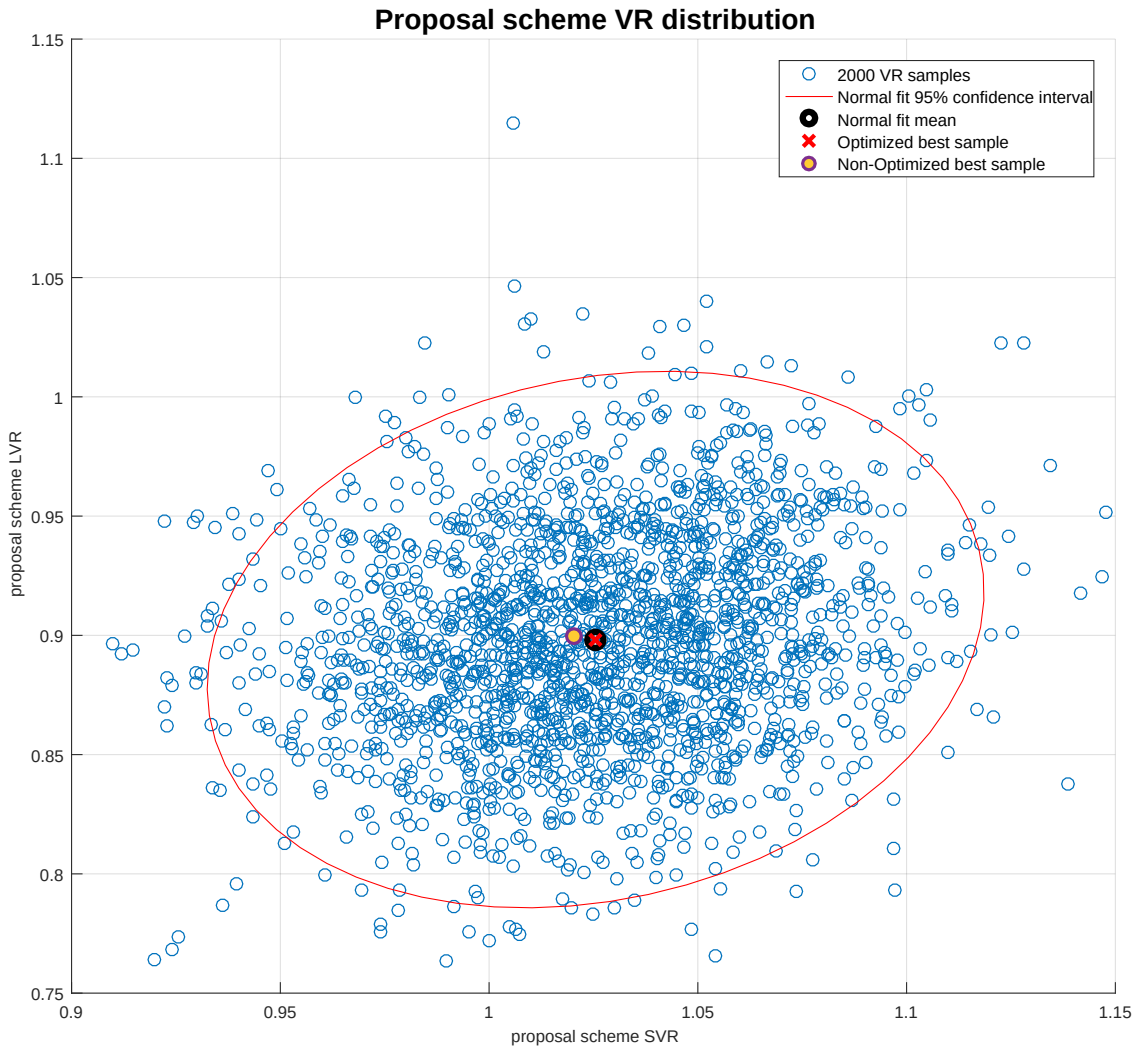


Figure 2.2: A projection of the sampling distribution of V and optimized sample to the proposed-scheme plane.

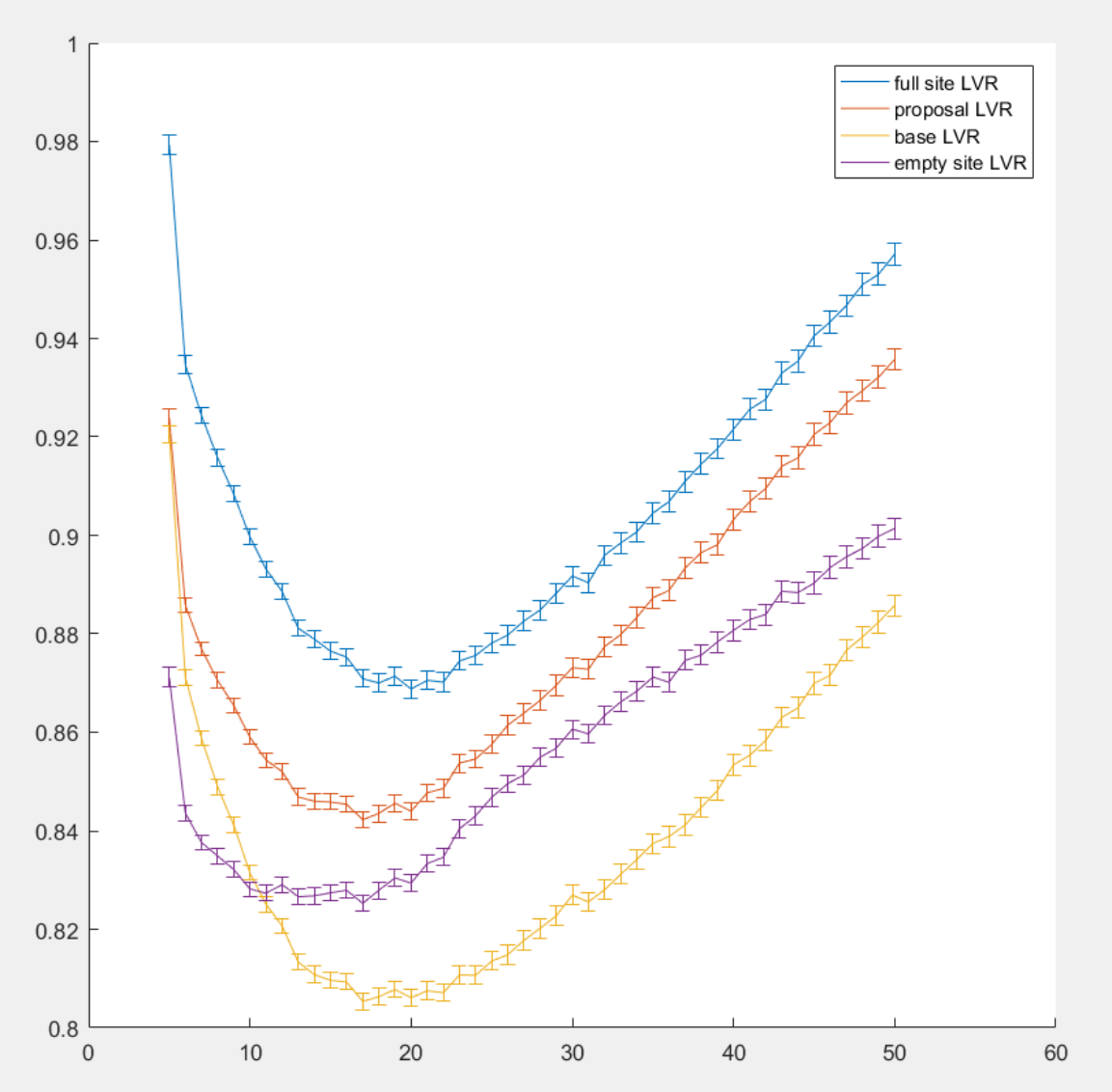


Figure 2.3: The MLE for LVR (y-axis with 95% confidence intervals) of each development scheme as the minimal spacing (x-axis in meters) between points changes

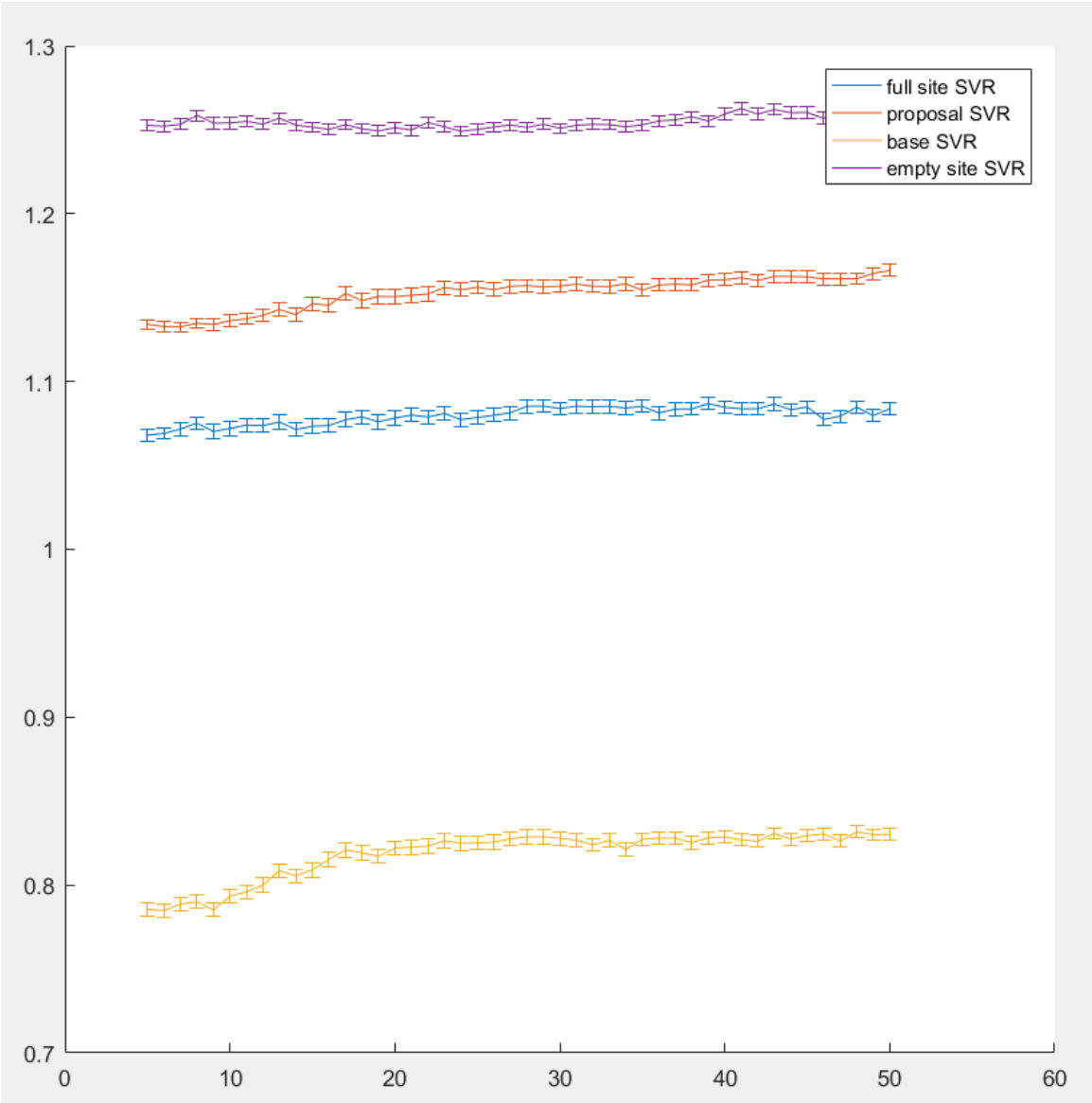


Figure 2.4: The MLE for SVR (y-axis with 95% confidence intervals) of each development scheme as the minimal spacing (x-axis in meters) between points changes.

	Base SVR	Base LVR	Proposed SVR	Proposed LVR	$d(*,\ \mu_{\mathrm{MLE}})$
μ_{MLE}	0.7431	0.7447	1.0255	0.8982	0
Best sample test points	0.7516	0.7408	1.0204	0.8996	0.2558
Optimized test points	0.7430	0.7446	1.0253	0.8982	0.0057

Table 2.1: Velocity ratios estimated for both the distribution of V and for optimized test points.

2.4 Conclusion and future work

In this section we developed a method which given a sample space of valid point sets will pick a set of unbiased test points closely reproducing the average values of LVR and SVR for any set of schemes. The set of test points can then be used in the wind tunnel experiments during the final step of the AVA.

In the future, we will evaluate the results of our methods using the wind data from all 16 of the wind directions used in the AVA. Further, we will utilize the expert guidance of AECOM to pick sampling regimes most closely mirroring the distribution of valid human-picked test points.

Chapter 3

Vector Field Similarity Comparison

3.1 Introduction

One method to reduce the bias in the current AVA practice is to concern the whole wind flow, instead of some particular points. Since wind fields are vector fields, if we could compare the vector fields of different schemes to the field with the best wind ventilation, by finding the most similar one, we then can conclude that scheme has a better ventilation.

However, it is difficult to compare vector fields directly since there are a lot of information one could extract from the fields. Therefore, usually vector fields are compared indirectly, by comparing certain features from the fields. One of indirect comparison methods include [10], which requires identifying the critical points of the vector fields, and finding the transition of critical points between the two fields. As mentioned by Dinh *et al.* [5], locating critical points are computationally intensive. Some natural occurring vector fields do not contain sufficient critical points which could also make this method obsolete.

In the following sections, the methodology of comparing wind fields and thus wind ventilation is explained. Results from this method is also discussed.

3.2 Methodology

The method chosen for this approach is introduced in [5], which is also an indirect method to compare vector fields but through topology. Since we wish not to compare the wind fields directly, we hope to sample the *geometric* distributions in forms of histograms of some selected features from the vector fields. These geometric features are chosen to reflect the shape of the vector fields. Then, we can compare the similarity of the histograms to the histogram of the benchmark wind field by some common metrics. Intuitively, among several schemes, the one that is the most similar in shape comparing to the benchmark field can be considered the best. These geometric distributions are robust to noise due to its statistical nature [5].

3.2.1 Extraction of vector field

From CFD software *Fluent*, 1,000,000 uniformly distributed massless particles were released near the city plot (See Figure 3.1). The paths of the particles follow the streamlines of the vector fields, from which the information is extracted. The detailed setup in *Fluent* of

particle tracking used in this approach is listed below. Readers may also refer to 4 for more details on particle tracking theory in *Fluent*.

- Number of timesteps: 5000
- Injection type: Surface
  - Plane, bounded and with sample points
  - Sample points: 1000 by 1000
  - Points: $(4800, 11100, 0)$, $(4800, 11750, 0)$, $(4800, 11750, 500)$
- Filter z-coordinate: Inside $0 - 2$ meters

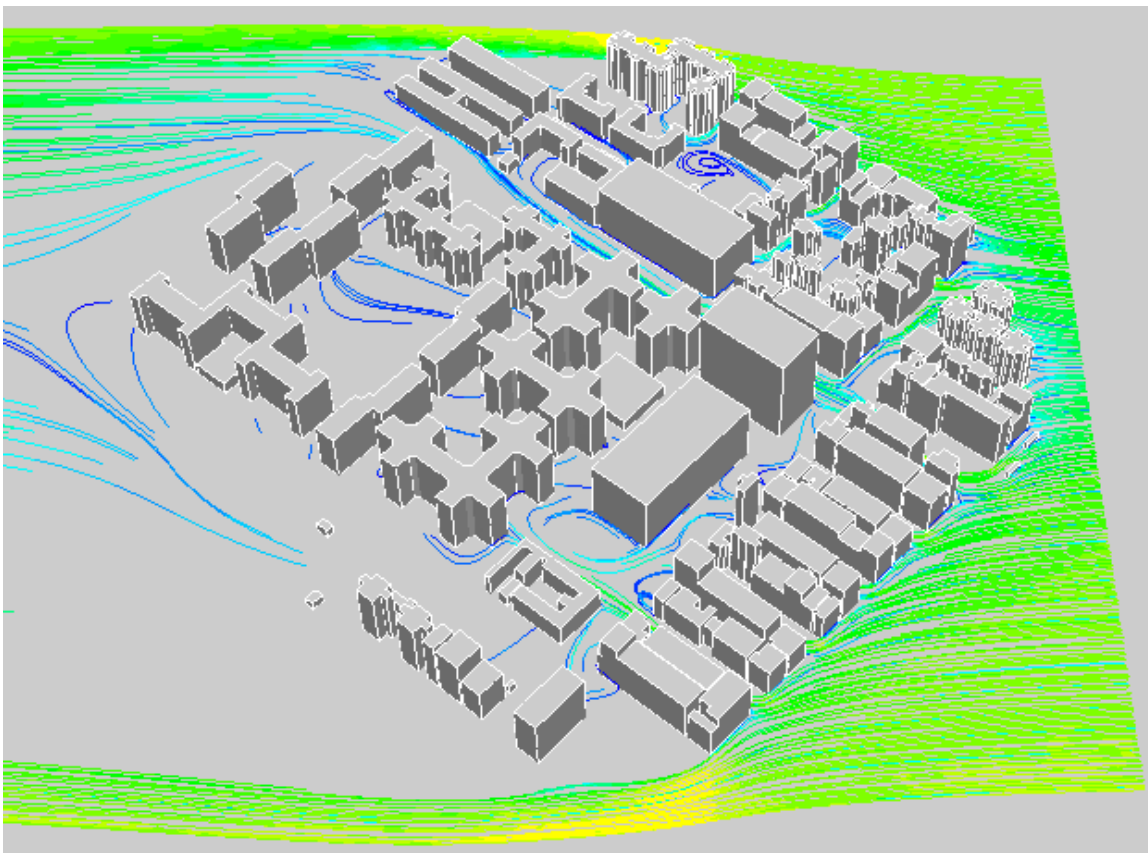


Figure 3.1: Particle tracking of 400 particles as an illustration.

3.2.2 Sampling geometric distributions

Two features are chosen for this approach: *velocity dot product* and *mean of curvature* of pairs of points on the same streamline. These features are chosen because they represent certain geometric properties of the fields. *Curvature* at a particular point measures the degree that the streamline deviates from a straight line and is given geometrically by the reciprocal radius of the circle that best approximates the curve near that point. The *velocity dot product* includes information of wind speeds and wind angles, which is useful in determining the topology of the vector field.

We extract data on streamlines to record the structural information of the field implicitly. The intuition is that they provide data with stronger geometric relations than sampling points randomly from the field. Here we give the definitions of mathematical definition of curvature.

Definition 3.1. The curvature of a parametrized curve $\boldsymbol{\gamma}(t)$ is defined by

$$\kappa := \frac{\|\boldsymbol{\gamma}' \times \boldsymbol{\gamma}''\|}{\|\boldsymbol{\gamma}'\|^3}. \tag{3.1}$$

Remark. $\boldsymbol{\gamma}'$ is provided by the software *Fluent* as velocity, and $\boldsymbol{\gamma}''$ is calculated by the method of finite difference.

Here we also define some commonly used terminologies associated with histograms.

Definition 3.2. A histogram is a function f that counts the number of observations that fall into each of the disjoint categories (or bins), such that if there are n bins with m total observations, we have

$$m = \sum_{i=1}^{n} f[i].$$

Definition 3.3. A normalized histogram is a function f that finds the probability of each bin from the samples, such that if there are n bins, we have

$$1 = \sum_{i=1}^{n} f[i].$$

Remark. f is essentially a discrete probability distribution.

The procedures of sampling are described below. Denote q as some quantity to be sampled.

Step 1. Choose sufficient random pairs of data in a scheme, such that the field is well-covered.

Step 2. For each pair of points, evaluate q and their Euclidean distance.

Step 3. Plot the 2D histograms of q against Euclidean distance.

The setups of plotting the histograms are listed below. For the details of the Python scripts used, please refer to [1].

- *Mean of curvature*
  - Number of bins: 50 by 50
  - Bin scale: Linear
  - Range of data sampled: 0th to 65th percentile for mean of curvature; 0th to 90th percentile for distance
- *Velocity dot product*
  - Number of bins: 50 by 50
  - Bin scale: Linear
  - Range of data sampled: 10th to 90th percentile for velocity dot product; 0th to 90th percentile for distance

Note the histograms have the following property.

Proposition 3.4. *The histograms are invariant to translation and rotation of wind fields.*

Proof. This is because the quantities *velocity dot product*, *mean of curvature* and *Euclidean distance* of pairs of points are invariant to translation and rotation of fields. □

Remark. This is a nice property to have. For example, let F be a wind field and F' be any rotated wind field of F. We would expect the histograms of F and F' to be identical since they are the same wind field with the same degree of ventilation.

3.2.3 Measuring the similarity of histograms

After sampling the histograms from all schemes, we then can compare them by some metrics. Three metrics are used and tested in this approach. These metrics are commonly used to compare discrete probability distributions, which are represented by the histograms.

Definition 3.5 (χ^2 distance)**.** The χ^2 distance between two n-bin normalized histograms f and g is defined to be

$$D_1(f,g) := \sum_{i=1}^{n} \frac{(f[i]-g[i])^2}{f[i]}. \tag{3.2}$$

Definition 3.6 (Alternative χ^2 distance)**.** The alternative χ^2 distance between two n-bin normalized histograms f and g are defined to be

$$D_2(f,g) := \frac{1}{2}\sum_{i=1}^{n} \frac{(f[i]-g[i])^2}{f[i]+g[i]}. \tag{3.3}$$

Remark. This alternative form is adapted from [9], which is used in texture comparison. The coefficient, however, is changed from 2 to $\frac{1}{2}$ to normalized this metric to be bounded in $[0,1]$.

Definition 3.7 (Hellinger distance)**.** The Hellinger distance between two n-bin normalized histograms f and g is defined to be

$$D_3(f,g) := \sqrt{1-\sum_{i=1}^{n}\sqrt{f[i]\cdot g[i]}}. \tag{3.4}$$

Remark. D_3 is also bounded by $[0,1]$.

In this approach, we assume the Empty Site Scheme has the best ventilation. Intuitively, with no buildings in site, the wind is blocked and hence it would not affect the air ventilation in the vicinity. Hence, the distances between other schemes and Empty Site Scheme are calculated and compared. We then conclude the scheme with the smallest distance as the one with best ventilation.

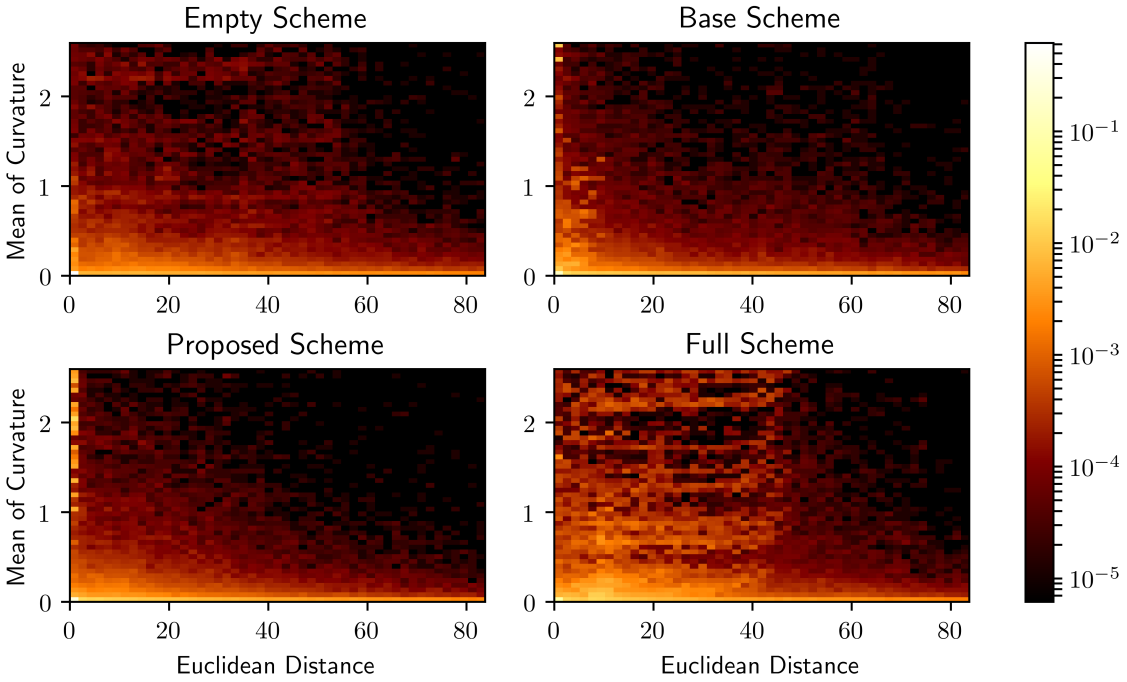


Figure 3.2: Histograms of *Mean of Curvature* against *Euclidean Distance*.

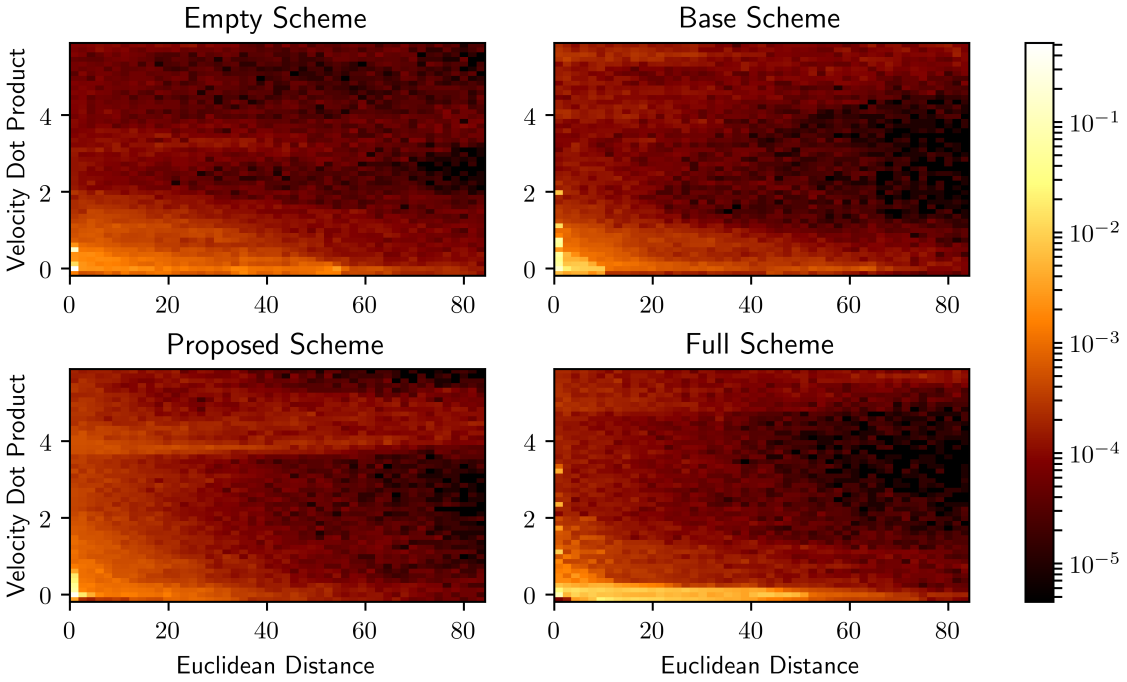
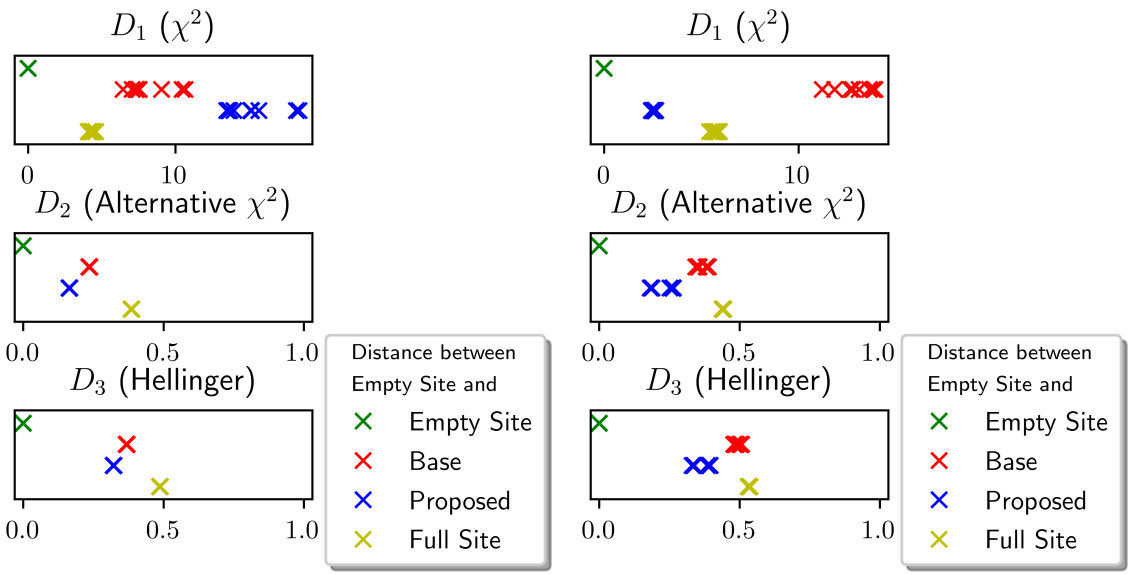


Figure 3.3: Histograms of *Velocity Dot Product* against *Euclidean Distance*.



(a) Mean of Curvature (b) Velocity Dot Product

Figure 3.4: Values of metrics in histogram comparisons, each ran for 10 times.

3.3 Results

The two sets of histograms are plotted (Figure 3.2, 3.3) and compared using the three metrics. From the figures, we can observe they retain similar general shape for histograms in the same category, while capturing some specific details from each scheme. For example, in the histograms of *mean of curvature*, some horizontal strokes are captured in Empty Site Scheme and Full Site Scheme. They may represent circular vortexes in the fields, or that the streamlines are generally curvy. From the histograms of *velocity dot product*, we can observe that Base Scheme and Full Site Scheme have more locations with lower velocity dot product. It could be either they generally have lower speed, or they have more wind velocity vectors in perpendicular directions. They show these histograms can capture specific geometric properties of the fields.

To test the performances of the metrics, comparisons in each setup are ran 10 times and their results are plotted (Figure 3.4). The tables 3.1 and 3.2 quantifies the averages of the metrics to show the relationship among the four schemes clearly.

	Empty Site	Base	Proposed	Full Site
D_2(Empty Site, $*$)	0.000	0.235	0.164	0.385
D_3(Empty Site, $*$)	0.000	0.368	0.321	0.486

Table 3.1: Means of the metrics for *Mean of Curvature*.

	Empty Site	Base	Proposed	Full Site
D_2(Empty Site, $*$)	0.000	0.366	0.214	0.439
D_3(Empty Site, $*$)	0.000	0.493	0.356	0.532

Table 3.2: Means of the metrics for *Velocity Dot Product*.

It can be observed that the standard χ^2 metric does not converge as well as other metrics. From this, we can conclude that this metric is not appropriate for our approach. We can

also see that the convergence in *velocity dot product* is not as good as *mean of curvature*. However, from alternative χ^2 metric and Hellinger metric, it can still be observed that the relation is

$$\text{Empty Site Scheme} < \text{Proposed Scheme} < \text{Base Scheme} < \text{Full Site Scheme},$$

where the smaller the distance, we can conclude the better the air ventilation. Hence, we conclude from this approach that the Proposed Scheme is better than Base Scheme.

3.4 Future Work

Currently, only a test case in one wind direction is used. To further validate this method, more test cases should be run to test its correctness. We may also verify our assumption on Empty Site Scheme being the best in air ventilation, and investigate a better benchmark than Empty Site Scheme if possible.

We might even look into methods without comparing to any vector field benchmarks since simulating an extra set of wind fields is time-consuming. Some brief ideas at the moment is to consider the optimal histograms for *mean of curvature* and *velocity dot product*. To have the wind completely flow through the city block, we may hypothesize the streamlines of the field should be straight with no reduce in velocity. Therefore, we might want the schemes that are being compared to have small *curvature* and large *velocity dot product*. We then can quantify a score for each scheme in its performance in *curvature* and *velocity dot product*, and rank the schemes with the scores.

Chapter 4

Particle Tracking

4.1 Introduction

The goal of AVA is to identify a building scheme that produces a LVR and SVR that is equal to or larger than that of the existing scheme. This is done to produce the highest quality pedestrian airflow possible in dense urban city blocks. As discussed previously, however, current methods are highly dependent on choice of test points. In this chapter, we introduce and develop particle tracking, a physical method for determining the scheme that best promotes airflow in the city. This method does not involve the choice of test points at all and thus avoids entirely the biasing pitfalls inherent in AVA.

In particle tracking, a large number of particles are released from specific locations, termed an injection. Their trajectory through the vector field is tracked until either they satisfy some exit condition, such as hitting the boundary wall of the simulation grid, or the maximum number of simulation steps is reached. With these discrete trajectories, we then calculate metrics, such as the average time spent by particles in the city and the number of unique particles that enter the city, to identify the scheme(s) that best flush out particles out of the area around the development cite.

The physical nature of this method can be visualized by imagining the particles as car exhaust or a gas leak. We hope that the wind flow quickly and efficiently moves these particles out of the development area, the segment of the city in which we are interested while doing this study. As air borne diseases, including Measles, Chickenpox, and Tuberculosis, are both deadly and increasingly prominent, an urban landscape that facilitates the rapid removal of toxic and harmful particles will provide residents with an increased quality of life.

All particle tracking trajectories are simulate using Ansys *Fluent*, the same computational fluid dynamics software used to generate the wind vector fields used in AVA. In this chapter, we will first discuss how particle tracking is done in this software, both in theory and in practice. We will discuss our analysis of the particle trajectories and will present the results from applying this method to the four different development schemes shown in Figure 1.4.

4.2 Particle Tracking in *Fluent*

Fluent is a computational fluid dynamics software produced by Ansys as a part of a suite of tools that facilitate any engineering simulation a design process may require. Using this

software, we release a large number of particles from specified locations and track their movement through the wind velocity vector field.

4.2.1 Theory

Discrete particle tracking in *Fluent* follows the Euler-Lagrange approach. The fluid in which the particles flow, the air in our simulations, are simulated using the Navier-Stokes equations, while the particle locations are solved in an iterative fashion starting from Newton's second equation of motion. A central assumption in this formulation is that the volume fraction of the particle phase among the fluid phase is less than 10%. Additionally this model works best for fluids with a well-defined entrance and exit condition, such as air blowing in from the east and out to the west, as is the case in our simulations. These conditions prevent particles from being indefinitely suspended in the fluid, a condition that this model doesn't sufficiently handle [3]. While the particles can interact with the fluid, exchanging mass, momentum, and energy, they do not interact with one another, and thus their trajectories are independent of one another.

Although *Fluent* supports both complex numerical integration methods and physical particle parameters, in this work we have explored only the simplest of both, steady-state particle tracking with no continuous phase interaction and mass-less, diameter-less particles. In steady-state simulations, the fluid flow is calculated first, and particles are later allowed to move in the resulting steady-state vector field. Unsteady tracking updates the fluid and particle phases in tandem, allowing for interaction between the phases. With no continuous phase interaction (default for the steady-state treatment), the discrete particle phase can have no effect on the continuous fluid phase. Given our choice of mass-less and diameter less particles, our particles have neither momentum nor energy and thus can have no effect on the fluid. With these combinations of parameters, particle tracking follows these steps:

Step 1. Solve the continuous flow.

Step 2. Place discrete particles at specified initial locations (injection).

Step 3. Track particle motions through air vector field and report trajectory.

The trajectory of a discrete phase particle is predicted by numerical integration of Newton's second law of motion in a Lagrangian reference frame, i.e. one in which the mesh representing the simulation body deforms with material deformation. With the choice of mass-less particles, the particle velocity u_p is exactly equal to the velocity of the fluid phase u at all points in the grid. Thus the trajectory of each particle x_p is predicted by

$$\frac{\mathrm{d}x_p}{\mathrm{d}t} = u. \tag{4.1}$$

The new particle location is calculated by a trapezoidal discretion of Equation 4.1, yielding the iterative formula

$$x_p^{n+1} = x_p^n + \frac{1}{2}\Delta t \left(u^n + u^{n+1}\right). \tag{4.2}$$

In this formula, x_p^n and is the location of particle p at simulation step n, and u^n is the velocity of the fluid at simulation step n at the location of particle p. The fluid velocity at the next time step is given by the classic forward Euler method:

$$u^{n+1} = u^n + \Delta t \left(u^n \cdot \nabla u^n \right) \tag{4.3}$$

∇u^n is the velocity gradient of the fluid velocity at the nth step, and as velocity itself is a vector, ∇u^n is a rank 2 tensor. The final term in parenthesis of Equation 4.3 is the derivative of fluid velocity with respect to time, i.e. the acceleration, as shown below. In the following work, u_i represents the velocity in the i_{th} coordinate. The superscript n is not written for clarity.

$$\begin{aligned}
u^n \cdot \nabla u^n &= (u_x, u_y, u_z) \cdot \left(\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial u}{\partial z} \right) \\
&= (u_x, u_y, u_z) \cdot \begin{bmatrix} \frac{\partial u_x}{\partial x} & \frac{\partial u_y}{\partial x} & \frac{\partial u_z}{\partial x} \\ \frac{\partial u_y}{\partial x} & \frac{\partial u_y}{\partial y} & \frac{\partial u_z}{\partial y} \\ \frac{\partial u_x}{\partial z} & \frac{\partial u_y}{\partial z} & \frac{\partial u_z}{\partial z} \end{bmatrix} \\
&= \left(u_x \frac{\partial u_x}{\partial x} + u_y \frac{\partial u_y}{\partial x} + u_z \frac{\partial u_z}{\partial x}, \ldots \right) \\
&= \left(\frac{\partial u_x}{\partial t} \frac{\partial x}{\partial x} + \frac{\partial u_y}{\partial t} \frac{\partial y}{\partial x} + \frac{\partial u_z}{\partial t} \frac{\partial z}{\partial x}, \ldots \right) \\
&= \left(\frac{\partial u_x}{\partial t}, \frac{\partial u_x}{\partial t}, \frac{\partial u_x}{\partial t} \right) \\
&= \frac{\partial u}{\partial t}
\end{aligned}$$

The integration time step Δt of each simulation step is determined by *Fluent* and thus varies for each particle and each step. This parameter is computed based on a user-chosen length scale L and the velocities of particle and continuous phase:

$$\Delta t = \frac{L}{u_p + u} = \frac{L}{2u}.$$

From this definition, it is apparent that L is the distance that a particle can travel before its motions of equations are resolved; thus a simulation with a smaller value of L will be finer and more precise but will require more steps to reach conclusion.

For more details about particle injection theory, including the equations of motion for particles with mass, interested readers are directed to the *Fluent* Theory Guide [3].

4.2.2 In Practice

In this subsection we will describe the process of tracking particles in *Fluent*. While this is simply an application of the software, we include this material so that our results can easily be reproduced by any interested parties.

We begin by selecting the discrete phase modelling methods to be used. As described above in Section 4.2.1, we employ steady-state tracking with no continuous phase interaction. We set the maximum number of simulation steps to be 5000 and specify a length scale $L = 0.01$ meters. This setup can be seen in Figure 4.1a.

Next we must define an injection, the set of initial conditions for the particles. This includes both physical parameters, such as mass and diameter distributions, and spatial parameters, such as initial conditions and velocity. With the wind blowing from east to west, we worked with two different injections, linear and urban circle. For the two other wind directions, we only conducting particle tracking with the second injection. In the first, we use a "group" injection to release 5000 particles from the line $x = 4800$, $y = 11000 \cdots 11800$, and $z = 2$. These N particles are spaced equally along this line by the formula

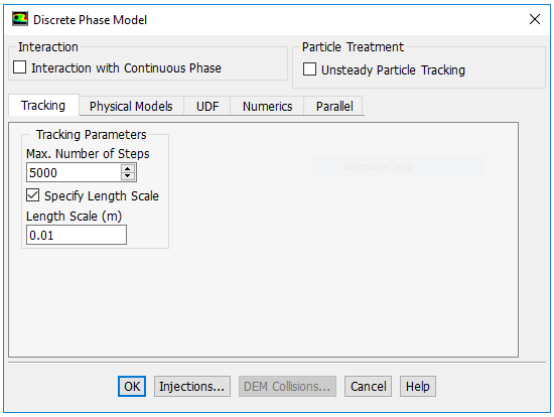
$$\phi_i = \phi_1 + \frac{\phi_N - \phi_1}{N - 1}(i - 1), \tag{4.4}$$

where ϕ_i, ϕ_1, and ϕ_N are the initial positions of the ith, 1st, and N^{th} particle, respectively. The setup for this injection can be seen in Figure 4.1b.

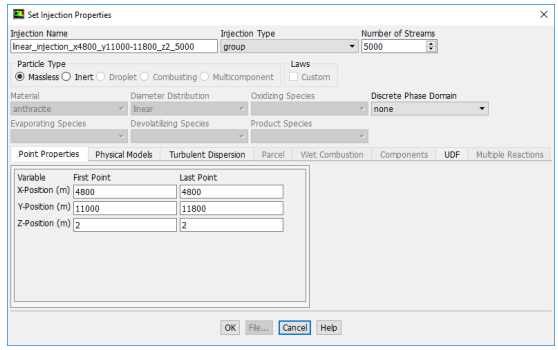
The second injection, the urban circle, is defined to be a circle of radius $2H$ centered ($x = 4447$, $y = 11440$) around the development site at a height of 2 meters, where $H = 98.5$ meters is the tallest building in all four schemes. We wish to place particles in the same initial positions for each of the four schemes. However, as the building patterns are different between schemes, the available locations for particles differ for each scheme. Thus, we use a "solid cone" injection to place 10000 particles in a circle of radius $2H$ around the development center of the Full site scheme, as can be seen in Figure 4.1c. We set the cone angle to $0°$, and the radius to $2H = 197$ meters. Of these 10000, only 4981 can be placed in area not occupied by a building. We then save the initial locations of these particles in the format specified by *Fluent* for a "file injection" [2]. This process is achieved using the the 'Create Initial Particle Location File' Jupyter notebook in the GitHub repository [1]. We then load this same injection into each of the four schemes, beginning with the Full scheme again. We run a simulation with a maximum of 50 steps and write the output file. This small number of steps is chosen as we only are interested in particle initial locations. Several particles will not be simulated, as the initial locations are not valid. We then create a new injection file of the initial conditions and repeat this process for the other three schemes. At the end, we will have an injection file of 4749 initial particle coordinates that are valid for all of the four schemes. This injection file can be used to specify initial particle locations regardless of the wind direction, as long as the scheme tested is one of the four used to generate the injection file.

Following this, we can generate the trajectories that we later analyze. We write the time of each simulation step, each of the three spatial coordinates, and the unique particle number to a file. The first four columns have 5 significant figures. When analyzing the data, a change in the particle number indicates that a new particle has been encountered. Figure 4.1d shows the setup necessary to track and write particle trajectories to file.

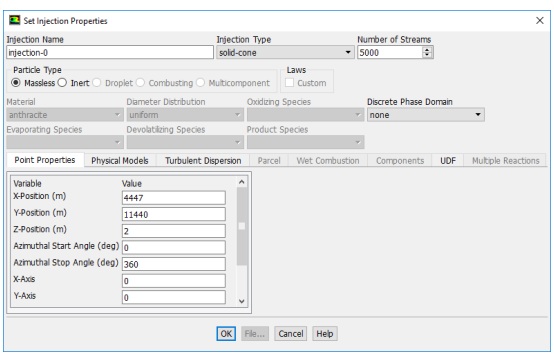
Particles can have several different fates when released and simulated. The most common state is "complete" in which the particle reaches the "outflow" wall in the simulation grid. When this happens, the particle escapes from the grid and is no longer tracked. In our simulations with wind blowing in from the east, the west wall is the outflow wall. The other significant state is "incomplete" in which the particle does not reach the outflow wall by the end of the simulation steps. With a small number of simulation steps, the particle may have simply ran out of steps while making its way to the outflow wall. However, with the large number of steps we used (5000), the particle is trapped in the urban area and is effectively stuck between the buildings.



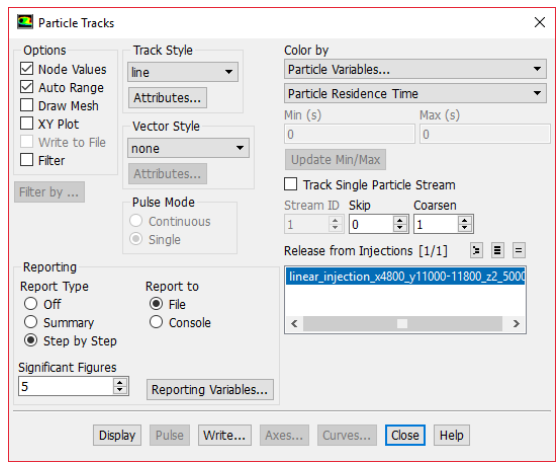
(a) Setup for Discrete Phase Model



(b) Setup for Linear Injection



(c) Setup for Urban Circle Injection



(d) Setup for Particle Tracking and writing to file

Figure 4.1: Setup of particle tracking in Ansys *Fluent*. Each subfigure is a screenshot of the program while preparing and generating particle trajectories.

4.3 Results

We analyze two sets of particle tracks. The first is the linear injection of 5000 particles to the east of the city with wind blowing from east to west. The second is the urban circle injection of 4749 particles with wind blowing from east to west. In each of these sets, we analyze data for all four building schemes: Empty Site, Base, Proposed, and Full.

The guiding principle in analyzing particle trajectory data is that the presence of particles is detrimental to resident health, and thus the scheme that both best reduces the number of particles in and quickly flushes these particles out of the assessment area is optimal. Air flow is assessed in the urban cylinder (UC) of radius $2H = 197$ meters and height 0.5 meters centered around the building site, i.e. the urban circle with a height of 0.5 meters. In agreement with current AVA practices and guidelines, the urban circle encompasses the area of the city whose wind flow should be affected by the potential building, while the height condition is chosen to cover the majority of pedestrians.

For each set of simulation data, we perform the following analyses. First we determine the number of particles that are in the urban cylinder at any point in time. In addition, we count the number of unique particles that ever enter this cylinder and accumulate both the

total time that these particles spend in this cylinder. Similarly we track the total number of steps each unique particle takes in the cylinder. We can then calculate the average and median time (steps) a particle spends (takes) in the UC. This gives us five metrics that we can use to compare schemes. The metrics are (1) number of unique particles that enter the cylinder, (2) average time spent by particles in the cylinder, (3) median time, (4) average steps, and (5) median steps.

All processing files and Jupyter notebooks with descriptions of results can be found in our group GitHub repository [1].

4.3.1 Linear Injection

This set of initial particle conditions simulates harmful particles being carried into the city by the wind. In Table 4.1, we list the number of particles released and the number of particles incomplete with both 5000 and 10000 steps for each of the four schemes. We see that for three of the four schemes, doubling the number of steps has no effect on the number of incomplete particles. Only for the Base scheme does doubling the number of steps decrease the number of incomplete particles by 1. Thus we feel confident simulating particle motion for 5000 steps.

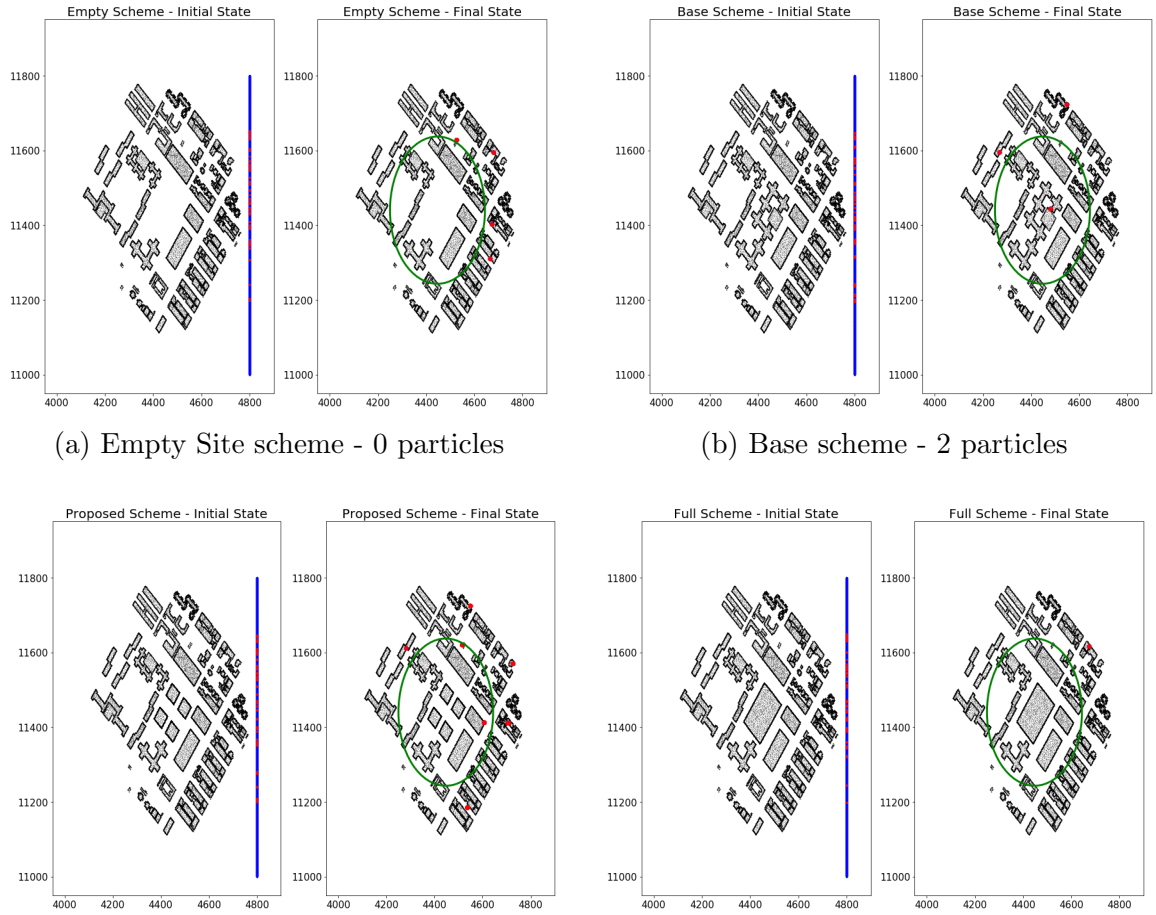
	Empty Site	Base	Proposed	Full Site
# Released	5000	5000	5000	5000
# Incomplete (5000 Steps)	166	206	261	178
# Incomplete (10000 Steps)	166	206	261	178

Table 4.1: Particle fates for linear injection trajectories.

In Figure 4.2, we show the initial and final states of each of the four schemes when releasing 5000 particles from the linear injection. The blue particles shown in the initial graph are the ones whose trajectory completes during the 5000 steps. No blue particles are seen in the final pane of each sub-figure, as particles can only be complete if they reach the outflow wall. The red particles are those with incomplete trajectories. We note that there are no clear pattern of red particles in the initial pane among all the four schemes, and thus incomplete particles do not originate from a single location. The green circle is the urban circle (radius of $2H = 197$ meters), and particles are only plotted if they have a height of $1.5 \leq z \leq 2$ meters, i.e. in the urban cylinder. In the final pane, the number of particles seen in the urban cylinder may not match with the number of particles that remain in the cylinder at the end of the simulation (seen in the caption); this is due to our choice of diameter-less particles. Particles can occupy the same physical location in our simulation. This is not physically unreasonable, however, as carbon monoxide, a air pollutant in automobile exhaust, has a kinetic diameter of 400 picometers. Our tracking has limited numerical precision and records particle location up to 0.01 meters.

After running our simulation, we analyzed the trajectories and calculated the five metrics used to rank the scheme. These metrics are shown in Table 4.2. The red values in each row denote the minimum value among the four schemes. We see that the Proposed scheme has the minimum number of unique particles in urban cylinder, median time spent by particles, and median steps taken. However the Proposed scheme has an order of magnitude higher average time and steps than the other three schemes. We believe these averages are skwed as the Proposed scheme captures more particles in the UC than the other schemes.

We can obtain a better understanding of the distribution of particle time (steps) spent (taken) in the urban cylinder for each of the four schemes through boxplots of these quan-



(a) Empty Site scheme - 0 particles

(b) Base scheme - 2 particles

(c) Proposed scheme - 7 particles

(d) Full scheme - 0 particles

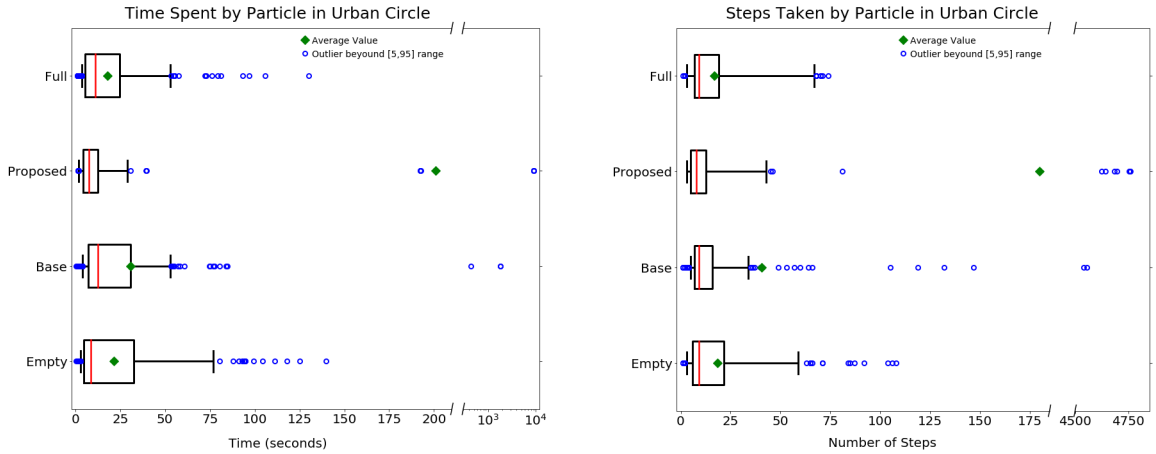
Figure 4.2: Initial and final states of each of the four schemes beginning with linear injection along with the number of particles remaining in the urban cylinder at the end of simulation.

	Empty Site	Base	Proposed	Full Site
# Urban Cylinder	250	343	194	257
Average Time (s)	21.3	30.9	200.9	18.0
Average Steps	18.3	40.4	179.8	16.8
Median Time (s)	8.7	12.7	7.4	11.0
Median Steps	9	9	8	9

Table 4.2: Comparison metrics for linear injection. Minimum of each quantity among the four schemes is shown in in red.

tities (Figure 4.3. In this, the boxplot extends from the 5th to the 95th percentile, while outliers outside of this range are represented as blue circles. The average of the quantity is shown by the green diamonds. We see that the Proposed scheme has both the minimum 25th and 75th percentile among the four schemes, so there is significant evidence that the Proposed scheme is most efficient at removing particles from the urban cylinder.

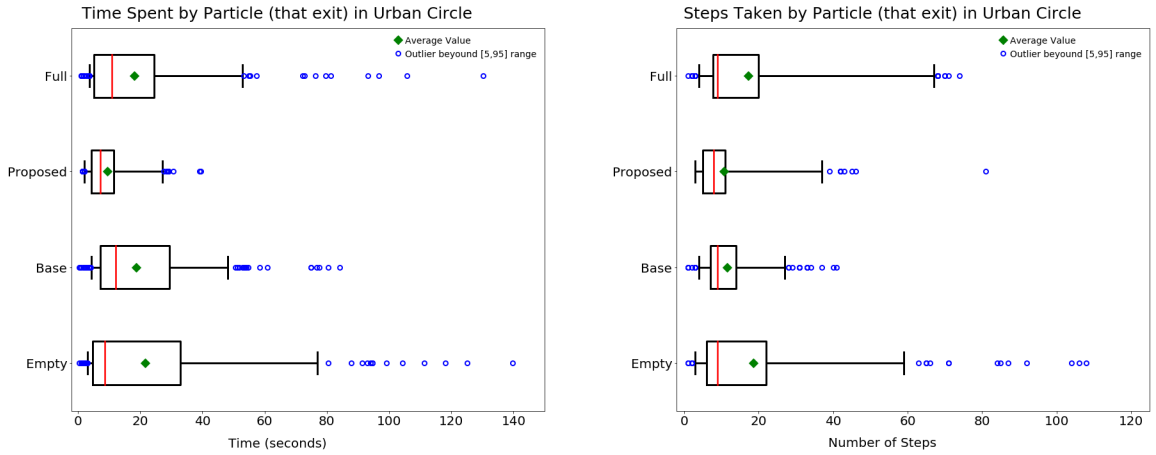
We then analyze the distributions of time spent and steps taken of the particles that both enter and eventually exit the urban cylinder. By excluding the incomplete particles from the distributions, as seen in Figure 4.4, we see drastic shifts in the averages for both time and steps of the Proposed scheme. With this analysis, we see that the Proposed scheme



(a) Time spent boxplot (b) Steps taken boxplot

Figure 4.3: Boxplots of time spend and steps taken in the urban city for the linear injection.

is the best of the four under all five metrics, as seen in Table 4.3. We see, however, that there is no clearly identifiable second-best scheme after the clearly best Proposed scheme.



(a) Time spent boxplot (b) Steps taken boxplot

Figure 4.4: Boxplots of time spend and steps taken in the urban city, excluding trapped particles, for the linear injection.

	Empty Site	Base	Proposed	Full Site
# Urban Cylinder	242	316	180	244
Average Time (s)	21.5	18.7	9.5	18.0
Average Steps	18.5	11.5	10.6	17.2
Median Time (s)	8.6	12.0	7.3	11.0
Median Steps	9	9	8	9

Table 4.3: Comparison metrics for linear injection, excluding trapped particles. Minimum of each quantity among the four schemes is shown in in red.

Finally, we analyze the number of particles in the urban cylinder at any point in time, as can be seen in Fig 4.5. We see that each scheme begins with 0 particles initially in the

cylinder, as expected from our knowledge of the injection scheme. We see that the Proposed scheme has a smaller initial spike in particles than the other three schemes but a increased number of particles in the cylinder as time progresses. This explains why the average time and steps of particles that enter the cylinder (Table 4.2) are significantly higher than those of the other three schemes.

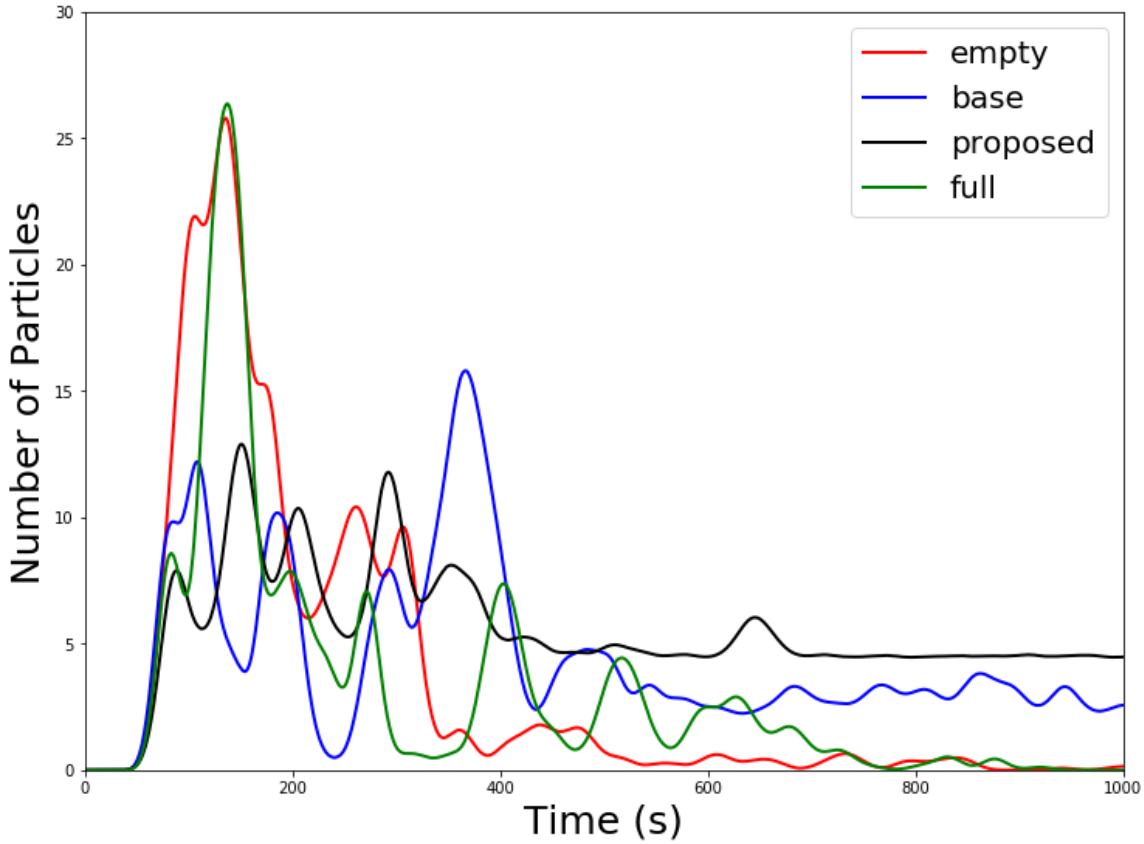


Figure 4.5: Particles vs. time for all four schemes with linear injection.

It is important to note that the analysis without trapped particles runs the risk of disregarding the issue of particles stuck in the city, but only around 250 of the 5000 released particles enter the urban cylinder in each scheme. To avoid the problems that this low a sample size of particles whose trajectory we analyze can introduce, we introduce our second injection scheme.

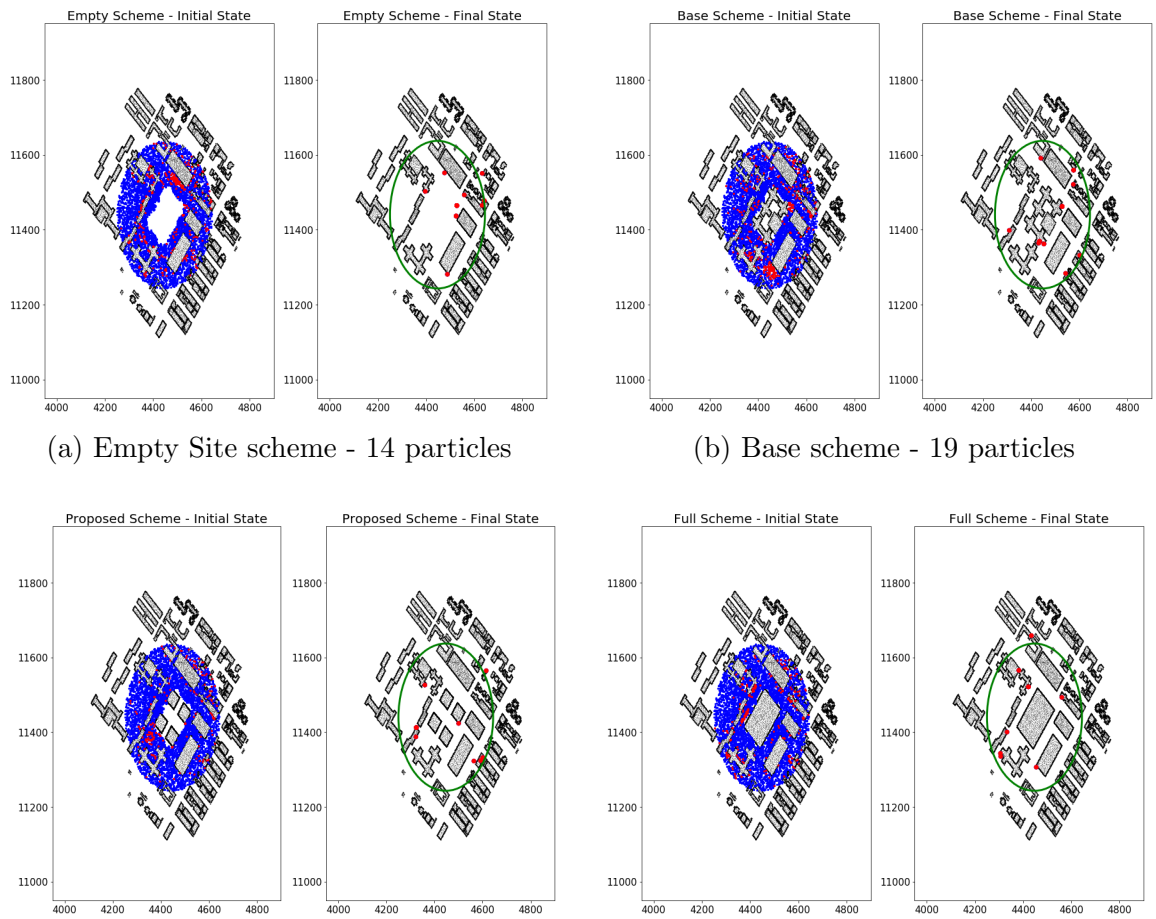
4.3.2 Urban Circle Injection

This set of initial particle conditions simulates harmful particles starting in the city, such as in a gas leak. An ideal scheme will quickly reduce the number of particles in the urban cylinder, as pedestrian airflow is negatively affected by the presence of these particles. In Table 4.4, we list the number of particles released and the number of particles incomplete with both 5000 and 10000 steps for each of the four schemes. We see that for two of the four schemes, doubling the number of steps has no effect on the number of incomplete particles. For the Base and Empty Site schemes, doubling the number of steps decrease the number of incomplete particles by 4 and 3, respectively. Thus we feel confident simulating particle motion for 5000 steps. This yields trajectories roughly half the size (200 MB) of those produced with 10000 steps.

	Empty Site	Base	Proposed	Full Site
# Released	4749	4749	4749	4749
# Incomplete (5000 Steps)	338	297	232	175
# Incomplete (10000 Steps)	334	294	232	175

Table 4.4: Particle fates for urban circle injection trajectories.

In Figure 4.6, we show the initial and final states of each of the four schemes when releasing 4749 particles from the urban circle injection. By design, the particles of all four schemes have the same initial conditions. The coloring of points in the two panes of each figure are the same as that from the linear injection analysis; the blue particles have complete trajectories, and the red particles are those with incomplete trajectories. Again, in the final pane, particles are only shown if final height is between 1.5 and 2 meters, while the green circle denotes the boundary of the urban cylinder. We note that there are no clear pattern of red particles in the initial pane among all the four schemes, and thus incomplete particles do not originate from a single location.



(a) Empty Site scheme - 14 particles

(b) Base scheme - 19 particles

(c) Proposed scheme - 11 particles

(d) Full scheme - 10 particles

Figure 4.6: Initial and final states of each of the four schemes beginning with urban circle injection along with the number of particles remaining in the urban cylinder at the end of simulation.

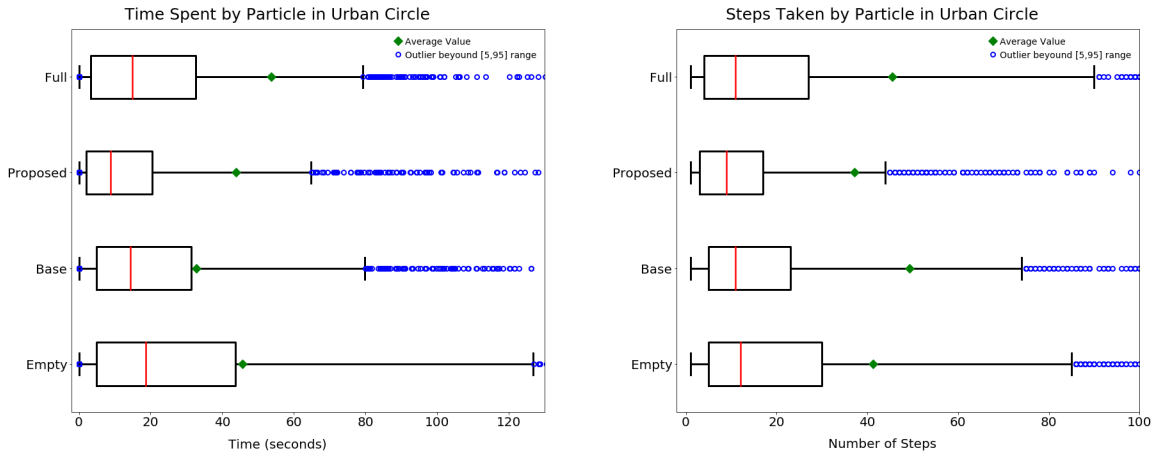
After running our simulation, we analyzed the trajectories and calculated the five metrics used to rank the scheme. These metrics are shown in Table 4.5. The red values in each

row denote the minimum value among the four schemes. We see that the Proposed scheme has the minimum number of unique particles enter urban cylinder, average steps taken by particles, median time spent, and median steps taken. The average time spent by particles for the Proposed scheme, while not the minimum among that of the four schemes, is not an order of magnitude larger, as was the case in the linear injection data.

	Empty Site	Base	Proposed	Full Site
# Urban Cylinder	2854	2998	2402	2729
Average Time (s)	45.6	32.9	43.8	53.7
Average Steps	41.3	49.4	37.1	45.4
Median Time (s)	18.7	14.3	8.8	14.8
Median Steps	12	11	9	11

Table 4.5: Comparison metrics for urban injection. Minimum of each quantity among the four schemes is shown in in red.

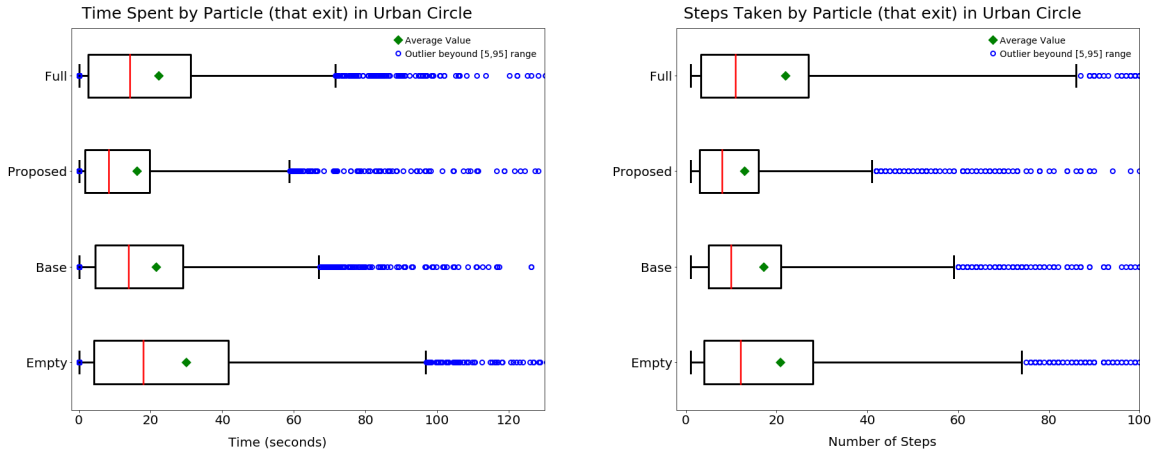
Again, we obtain a better understanding of the distribution of particle time (steps) spent (taken) in the urban cylinder for each of the four schemes through boxplots of these quantities (Figure 4.3. The boxplots have the same style as those in the linear injection, where the whiskers extend from the 5th to the 95th percentile, the outliers are shown as blue circles, and the average values are shown as green diamonds. In addition to being optimal in four of the five metrics listed in the table, we see that the Proposed scheme has both the minimum 25th and 75th percentile among the four schemes. Thus there is significant evidence that the Proposed scheme is most efficient at removing particles from the urban cylinder.



(a) Time spent boxplot (b) Steps taken boxplot

Figure 4.7: Boxplots of time spend and steps taken in the urban city for the urban circle injection.

We then analyze the distributions of time spent and steps taken of the particles that both enter and eventually exit the urban cylinder. By excluding the incomplete particles from the distributions, as seen in Figure 4.8, we see that the Proposed scheme is the best of the four under all five metrics, as seen in Table 4.6. As the Proposed scheme has the similar number of particles trapped in the urban cylinder as the other three schemes, as can be seen in captions of Figure 4.6, removing the trapped particles does not unfairly bias the results. Instead, we clearly find that Proposed scheme is most efficient at flushing particles out of the urban city, which is a preferable property for pedestrians.



(a) Time spent boxplot

(b) Steps taken boxplot

Figure 4.8: Boxplots of time spend and steps taken in the urban city, excluding trapped particles, for the urban circle injection.

	Empty Site	Base	Proposed	Full Site
# Urban Cylinder	2607	2715	2269	2538
Average Time (s)	29.8	21.5	16.2	22.3
Average Steps	20.8	17.0	12.9	21.9
Median Time (s)	18.0	13.8	8.4	14.2
Median Steps	12	10	8	11

Table 4.6: Comparison metrics for urban circle injection, excluding trapped particles. Minimum of each quantity among the four schemes is shown in in red.

Finally, we analyze the number of particles in the urban cylinder, as can be seen in Fig 4.9. We note that each scheme begins with 4749 particles initially in the cylinder, as expected from our knowledge of the injection scheme. We see that the particles leave the Proposed scheme much faster than in the other schemes, as at 200 seconds, the Proposed scheme holds half of the particles in the others.

4.4 Conclusions

The method of particle tracking allows us to obtain physically relevant quantitative observations about the ability of each scheme to remove harmful particles from around the development site. These particles can either enter the development site being carried by the wind, as in the linear injection, or begin near the development site, as in the urban circle injection. In both cases, we find that the Proposed scheme is the best of the four. We draw this conclusion from five quantities calculated – number of unique particles that enter the urban cylinder over the course of particle tracking, average time spent by particles in the cylinder, average steps taken, median time, and median steps – and the graph of number of particles in the cylinder versus time. The Proposed scheme wins in the majority of the five quantities, and when we confine our analysis to the particles that both enter the cylinder and finish their trajectories, the Proposed scheme wins in all five quantities.

This method is not without issues, which we plan to address moving forward. The most important is that while the Proposed scheme is clearly the best, there is no clear second-

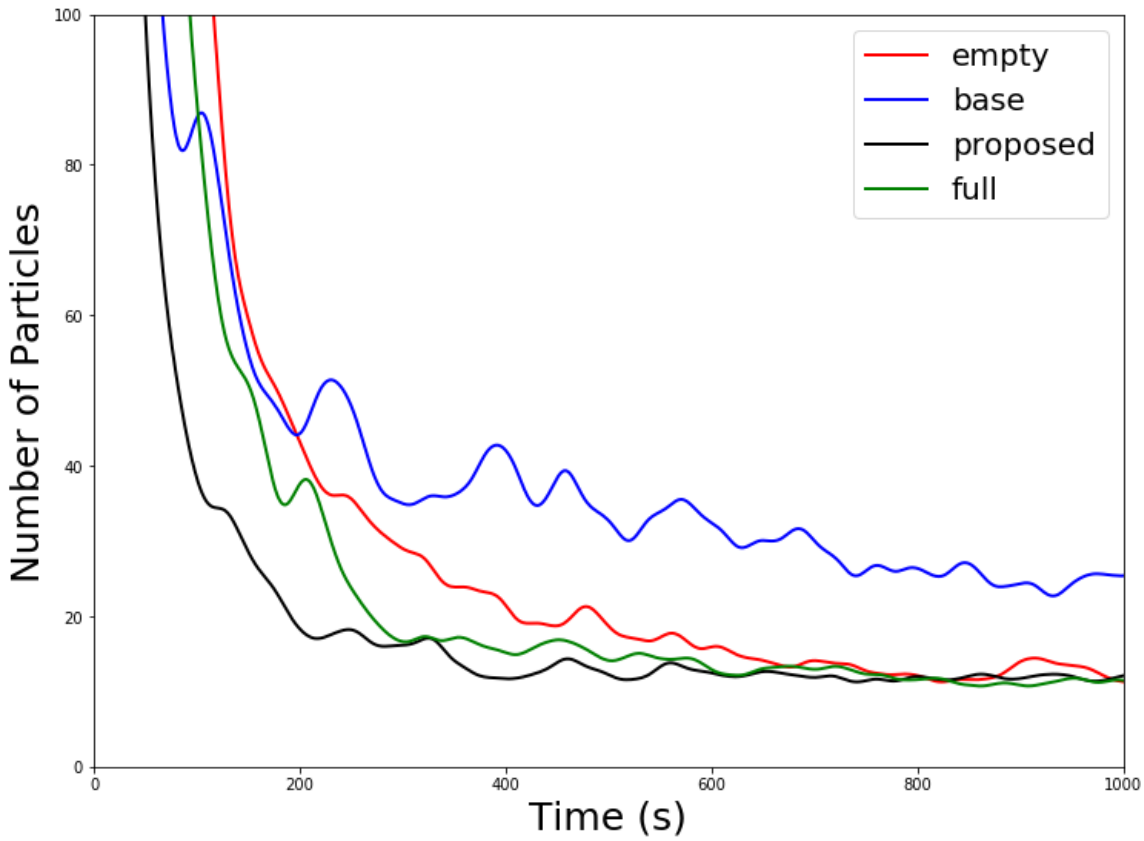


Figure 4.9: Particles vs. time for all four schemes with urban circle injection.

best scheme, which is required for particle tracking to be a robust ranking mechanism. With no clear ranking, it is difficult to compare this method to the current AVA procedures and the other approaches introduced in this report. This lack of a clear ranking is due, in part, to computing five separate quantities on the quality of airflow in each schemes. Ideally, we would derive a single numeric value for each scheme that encapsulates all possible information and use this value for ranking. Another issue in our analysis is our particles are massless and have diameters 0. To be more physically accurate, we'd like to model typical pollutants found in a densely populated city like Hong Kong. *Fluent* itself has numerous built-in particles that we intend to explore moving forward.

To address the need for a singular ranking factor, we propose a novel metric that produces a real number between 0 and 1 for each scheme. Let X be the set of particles released in the simulation, Y be the set of particles that ever enter the urban cylinder, and Z be the set of particles that enter the city and do not escape by the end of the simulation. The set Z is a subset of Y, and if we denote W the set of all particles that do not complete their trajectory (incomplete particles), we know that $Z = Y \cap W$ as the all particles that complete their trajectory end the simulation along the outflow wall. Given a development scheme b and wind direction d, our metric has the following definition:

$$M(b,d) = \frac{1}{2}\left(\frac{|X| - |Y|}{|X|} + \frac{|Y| - |Z|}{|Y|}\right)$$

This metric captures static information about the scheme, as sets Y and Z are compiled at the end of the simulation. Values of $M(b,d)$ closer to 1 indicate a scheme with better

ventilation properties, as fewer particles enter the urban cylinder, and of these particles that do enter, fewer are trapped in the cylinder. If $|Y| = 0$, we define $M(b, d) = 1$.

Chapter 5

Conclusion

Hong Kong, being one of the most populated cities in the world, faces great challenges in providing its residents with a high environmental standard of living. To address issues with air ventilation caused by densely packed high-rises, the Hong Kong Government established the Air Ventilation Assessment (AVA) system, a set of guidelines and methodologies used to evaluate the impact that potential urban development projects will have on pedestrian airflow. While allowing project developers to assess their designs, the results of AVA are derived from the placement of discrete test points around the development site. As we have demonstrated, however, the parameters calculated from the wind velocity at these test points are highly dependent on the specific locations of these points, and thus the ranking of different development schemes can vary with the choice of points. In fact, the results of the current practices can be skewed by deliberate choosing of particular points. AECOM, which performs environmental assessments of potential urban developments on behalf of its clients, seeks a more objective way to study the impact these developments will have on pedestrian airflow.

With this in mind, we have proposed and tested three novel methods for assessing airflow quality. The first method, Bias Optimization of Test Points, seeks to improve current AVA practices by identifying an objective set of test points that properly reflects the wind availability of the two schemes being compared. In this method, we sample a large number of sets of test points, compute the LVR and SVR values of each set of test points, and estimate the "true" (mean) value of the LVR and SVR for each scheme. Beginning from the sample test points that give values closest to the mean values, we then construct a set of test points that more closely approximates the true values by optimization. This approach provides us with a set of test points that yields a LVR and SVR within a few thousandths of a standard deviation, as measured by Malahanobis distance, from the true values.

The second and third methods compare wind schemes to identify those that are best for pedestrians. The Vector Field Similarity approach makes the physically intuitive assumption that the Empty Site scheme has the most desirable wind vector field and thus compares all other schemes to the Empty Site one. To do this analysis, we compute two geometric properties, velocity dot product and mean of curvature, from pairs of points on streamlines of the wind flow vector fields. We then plot these against Euclidean distance between the points as 2D histograms and calculate the distance between the histograms using three different measures. Two measures yield consistent results across multiple samples and give the following ranking in increasing distance from the Empty Site scheme: Proposed, Base, Full.

The final approach, Particle Tracking, seeks to extract physical observations on the ability of schemes to flush particles out of the area surrounding the development site. After releasing particles from one of two different initial conditions, we count the number of particles and calculate several different quantities of these particles, such as average time spent and steps taken, that are sufficiently close to the development site at a height that affects pedestrians. From this, we conclude that the Proposed scheme is best. We note however, that this method does not rank the Empty Site scheme as optimal, which contradicts the assumption made in the second approach.

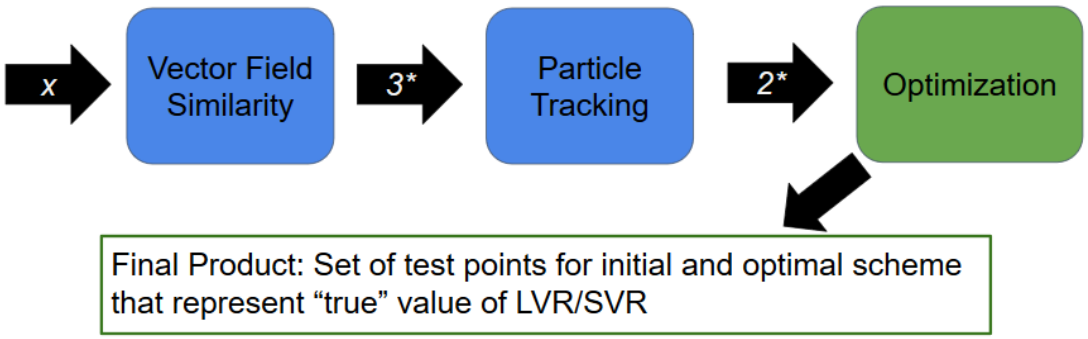


Figure 5.1: Prototype of pipeline that integrates all three methods proposed in this report to replace current AVA practice. The * denotes that the numbers shown are approximates.

Moving forward, there are features of each approach to be explored. Once each method has been proven to be robust for all wind directions and all ranges of simulation schemes, from simple geometries to intricate urban blocks, we envision a pipeline with a form similar to that of Figure 5.1 for evaluating the environmental impact potential developments will have on pedestrian airflow. Using a combination of the Vector Field Similarity and Particle Tracking methods, we can identify schemes that have favorable ventilation properties. Once we a small number of schemes to consider, we can use Bias Optimization to find a set of representative test points for the current development and each possible scheme that can be used in wind tunnel experiments. Ultimately, this process will allow AECOM and other air ventilation experts to objectively recommend building schemes with the pedestrians' best interests in mind.

Selected Bibliography Including Cited Works

[1] S. Anand, H. N. Tse, O. Levin, and K. C. Hsu, *Aecom-ava-assessment*. https://github.com/sajantanand/AECOM-AVA-Assessment, 2018.

[2] Ansys, Inc., *File injection format*. https://www.sharcnet.ca/Software/Fluent6/html/ug/node844.htm, 2006.

[3] ——, *Discrete phase*. http://www.afs.enea.it/project/neptunius/docs/fluent/html/th/node238.htm, 2009.

[4] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[5] H. Q. Dinh and L. Xu, *Measuring the similarity of vector fields using global distributions*, in Structural, Syntactic, and Statistical Pattern Recognition, N. da Vitoria Lobo, T. Kasparis, F. Roli, J. T. Kwok, M. Georgiopoulos, G. C. Anagnostopoulos, and M. Loog, eds., Berlin, Heidelberg, 2008, Springer Berlin Heidelberg, pp. 187–196.

[6] R. Durrett, *Probability: theory and examples*, Cambridge university press, 2010.

[7] E. Ng, *Policies and technical guidelines for urban planning of high-density cities – air ventilation assessment (AVA) of hong kong*, Building and Environment, 44 (2009), pp. 1478–1488.

[8] E. Ng, I. Tam, A. Ng, B. Givoni, L. Katzschner, K. Kwok, and V. Cheng, *Feasibility study for establishment of air ventilation assessment system–final report*, Department of Architecture & Chinese University of Hong Kong, (2005).

[9] J. Puzicha, T. Hofmann, and J. M. Buhmann, *Non-parametric similarity measures for unsupervised texture segmentation and image retrieval*, in Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on, IEEE, 1997, pp. 267–272.

[10] H. Theisel, C. Rössl, and H.-P. Seidel, *Using feature flow fields for topological comparison of vector fields.*, in VMV, 2003, pp. 521–528.

[11] W. Wang and E. Ng, *Air ventilation assessment under unstable atmospheric stratification — a comparative study for hong kong*, Building and Environment, 130 (2018), pp. 1–13.