Western University

Scholarship@Western

Electronic Thesis and Dissertation Repository

11-24-2022 2:10 PM

Pressure Drop and Heat Transfer in Flow Over an Array of Blocks of Varying Heights: A Statistical and AI Analysis on the Effect of Block Height Variation

Ali Navidi, *The University of Western Ontario*

Supervisor: Ogden, Kelly, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Mechanical and Materials Engineering

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Heat Transfer, Combustion Commons

Recommended Citation

Abstract

The presence of a stiff obstruction in the path of fluid causes the creation of a boundary layer over and around the obstruction. The flow over an idealized, two-dimensional series of blocks is numerically investigated to determine how statistical blocks height variation, such as standard deviation, mean, and skewness, influence pressure drop and heat flux. These data sets serve as a foundation for developing models for estimating the heat transfer coefficient of each block using machine learning (ML) methods. The results show that the pressure drop increased by 60% when the standard deviation of heights of blocks increased from 0.1 to 0.4 due to promoting turbulent mixing over the blocks, hence boosting pressure drop and heat flux. Furthermore, the ML model has great potential for predicting the Convective heat transfer coefficient (CHTC) of an individual block given the heights of a few nearby obstacles.

Keywords

Summary for Lay Audience

In nature, the presence of obstacles in the path of a fluid, such as a group of nearby plants in the way of water flow, a forest, or a group of buildings in the way of wind, results in a distinct fluid behaviour over and around the obstacle, with the flow going over the obstacle and slowing down near it. Due to the extensive industrial applications and environmental impacts of barriers on flow behaviour, it is vital to understand the flow near obstructions. This investigation uses idealized simulations to examine the influence of height variation within a single cluster of obstacles. The obstruction might symbolize a vegetation canopy, an urban canopy, or urban structures. Previous research on this topic focuses on groups of blocks that follow a pattern. In this study, the effect of random variation of height is investigated. The height of the 14 blocks for each set of simulations is assumed to be random, and they are generated by defining the average height and the variation of heights. Each of these parameters' effect on the total heat flux and pressure drop are investigated. Additionally, the effect of the qualitative flow regimes is explored. Using simulation-generated data, a machine learning model for predicting the heat transfer coefficient of blocks has been developed. The feature importance analysis in machine learning indicates that only data for a few neighbouring blocks are required to calculate the heat transfer coefficient.

Acknowledgments

I'd like to convey my profound appreciation to my supervisor, Professor Ogden, for providing me with this chance and for her advice, encouragement, and support throughout this incredible research experience. I shall be eternally appreciative for this. For my personal development at Western and the accomplishment of this thesis, her patience, freedom, and direction have been indispensable.

I would also like to extend my deepest gratitude to the outstanding instructors and staff at Western University for their assistance and for providing such a valuable learning experience through this program, as well as to the Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Research Development for their financial support..

I would like to thank all my friends and coworkers who have provided support and encouragement along this trip. The invaluable contacts I've formed at Western University are irreplaceable. I wish them nothing but success in the future.

Lastly, I would like to express my inexhaustible admiration and gratitude to my family, particularly my parents. Without the sacrifices they have made and the support they have offered throughout my life, none of this would have been possible. There are insufficient words to express my gratitude to them.

Table of Contents

List of Tables

List of Figures

Nomenclature

Abbreviations

MRE	Mean relative error
MSE	Mean squared error
RMSE	Root mean square error
SHTC	Surface heat transfer Coefficient
CHTC	Convective heat transfer Coefficient
STD	Standard deviation
SK	Skewness
H1	Height of First Block
LES	Large Eddy Simulation
RANS	Reynolds-averaged Navier-Stokes simulations
CFD	Computational fluid dynamics
SST	Shear stress transport
2D/3D	Two/Three Dimensional
BES	Building energy simulation

Ti	Turbulence intensity
RSM	Reynolds stress turbulence model
CNN	Convolutional Neural Networks
LSTM	Long Short-Term Memory layer
ML	Machine learning
AI	Artificial intelligence
ANN	Artificial Neural Network
DL	Deep Learning
GBRT	Gradient Boosting Regression Trees
SVR	Support Vector Regression
SVM	Support Vector Machines

Subscripts

avg	Stands for the average value of results
exp	Stands for experimental results
sim	Stands for simulation results
Min	Stands for the minimum amount
max	Stands for the maximum amount
Ref	Stands for reference
Wall	Stands for wall surface

Symbols

$$Re = \frac{UL}{\nu}$$ Reynolds number

U   Velocity of the field

L   Characteristic length

P Pressure

u   Velocity in x direction

v   Velocity in y direction

k   Turbulent kinetic energy

S   Spacing between adjacent buildings/cubes

H   Height of blocks

q   Cross-diffusion term

S_t Strain rate

F_2 Blending functions

Greek letters

ω Turbulent dissipation rate

ν kinematic viscosity

$\widetilde{G}_K$ Generation of kinetic energy associated with turbulence caused by mean velocity gradients

G_ω	Generation of ω
Γ_k	Effective diffusivities of K
Γ_ω	Effective diffusivities of ω
Y_K	Dissipation of k
Y_ω	Dissipation of ω
S_k	User-defined source term for k
S_ω	User-defined source term for ω
σ_k	Turbulent Prandtl numbers for k
σ_ω	Turbulent Prandtl numbers for ω
μ_t	Turbulent viscosity

List of Appendices

Chapter 1

1   Introduction

1.1   Motivation

The presence of a porous surface or rigid obstacle results in the formation of a boundary layer in flow over and around an obstacle. Numerous examples of this phenomenon exist in nature and engineering, including flows past aquatic and terrestrial vegetation, atmospheric boundary layers forming over forests, tall buildings in cities, and wind/tidal farms, which consist of arrays of wind/tidal turbines. The formation of a boundary layer over an urban area and a submerged canopy is shown in Figure 1-1 and Figure 1-2. These impediments have a substantial impact on the flow momentum and energy transfer process near and downstream of the obstacles. Aquatic vegetation, for instance, offers shielding [1] and impacts sediment production, erosion [2]–[5], and carbon sequestration in coastal habitats [6]. Urban Buildings and canopies cause multi-scale turbulent motions in the atmospheric boundary layer [7], [8], which governs flow mixing, momentum, and heat and mass exchange within the layer, including pollutants and other chemical compounds [9], [10].

Considering the vast industrial uses and environmental effects of obstacles on flow behaviour, it is necessary to comprehend the flow near impediments. This study carries out idealized simulations that will enable us to study the effect of height variation within a single cluster of obstacles. The obstacle may represent a vegetation canopy, urban canopy or urban buildings. The Introduction chapter is organized as follows. First, the significance of studying the flow over the vegetation canopy will be discussed. Then, an introduction to atmospheric flow and its applications on buildings is provided.
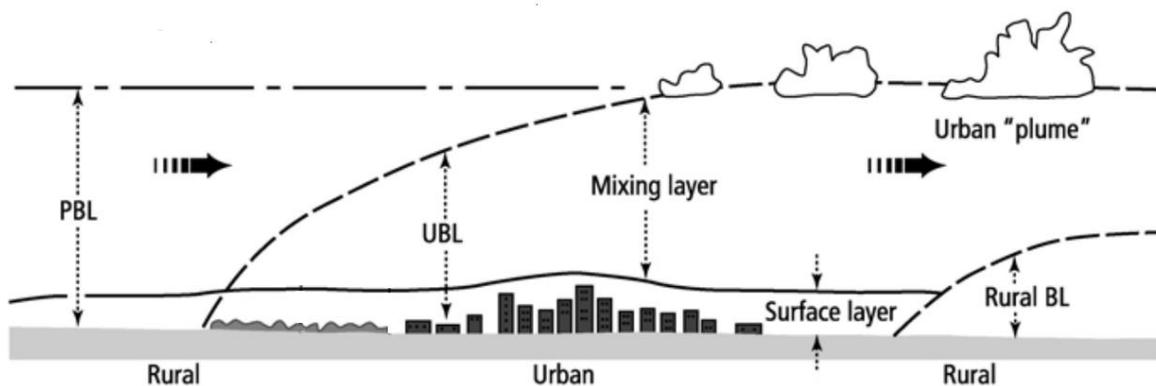
Figure 1-1: Schematic of the urban boundary layers (adapted from Piringer et al. 2002) (Planetary Boundary Layer, Urban Boundary Layer, Urban Canopy Layer)
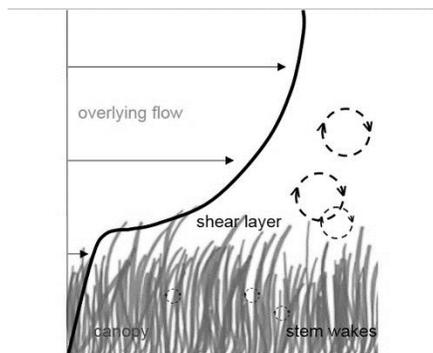
Figure 1-2: schematic of submerged canopy boundary layer (adapted from Beudin[11]**)**

1.2   Canopy Literature Review

Coastal regions are crucial for economic development, international connectivity, and political communication [1]. In addition, they are one of the most frequent human activity zones. This creates significant demand for beach development for housing, recreational amenities, and coast protection from storms and overflowing waves. Waves, tides, and storm surges have a tremendous influence on the construction of ports, canals, and coastal structures and are essential in determining the coastline's geometry. Due to the inherent benefits of utilizing natural protection measures to safeguard coastlines, protection strategies are shifting from structural reinforcement to natural ones. Hirashi and Harada [2] conducted an experimental study to determine how well the coastal green belt protected against tsunami waves using observational data from the occurred tsunami. They discovered that vegetation outperformed structural obstacles in terms of both cost and force reduction. Coastal vegetation is regarded as a kind of biological control and plays a significant role in developing and preserving ecosystems. Vegetation generates an extra drag force that alters the mean and turbulent velocity profiles [3]–[6], which in turn affects sediment transport and channel morphology [7]–[9]. Aquatic canopies can be classed as submerged, emergent, or suspended based on their growth characteristics. In this study, the idealized simulation best represents the submerged canopy.
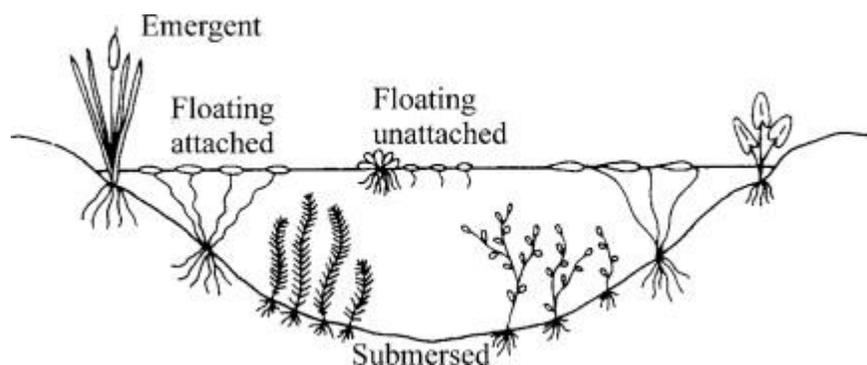
Figure 1-3: Canopy classification showing emergent, submerged and suspended (floating) canopies (adapted from Walter [12]**)**

Additionally, the vegetation canopy dampens waves, stabilizes the seafloor, and improves local water quality by filtering nutrients from the water. In particular, velocity dissipation inside the canopy region results in sediment deposition [10], greater light transmission [11], and nutrient retention, which creates a favourable habitat for invertebrate larvae and other species [12]–[14]. The flood plain vegetation can lower the local velocity by 20–44 percent compared to the main channel, resulting in less erosion and better sediment retention [15], [16].

Wave attenuation by coastal forests depends on plant properties such as geometry and structure, submergence ratio, density, stiffness, and local arrangement, as well as wave circumstances, including incident wave height, period, and direction [1]. Numerous studies have investigated vegetation characteristics' effect on various properties, including wave attenuation, pressure drop, etc. Etminan [17] analyzes the mechanisms that influence canopy flow resistance, such as canopy drag, including blockage, sheltering, and delayed separation, and presents a new model for estimating canopy drag coefficients within emerging canopies using the LES method in OpenFoam. Etminan discovered that sheltering and delayed separation fractionally reduce the drag of extremely sparse canopies, whereas blockage significantly increases the drag of dense canopies. He modelled canopies as arrays of rigid circular cylinders, which is standard practice for approximating an aquatic canopy due to the complexity of modelling the geometry of real natural canopies [18], [19]. Chen [20] presents the transition from open channel flow to flow through submerged vegetation in a two-box model using velocity measurements. In Chen's work, wooden arrays of rigid circular cylinders of varied densities and depths of submergence in the staggered pattern were used to study submerged canopies and flow adjustment experimentally. He discusses the steps of flow adjustment and concludes that canopy length has no effect on flow adjustment. Thang [21] developed empirical formulas for the drag coefficient of an isolated cylinder and an array of cylinders, as well as the vertically distributed or local drag coefficient of the submerged canopy in an open channel flow, based on the experimental results for a wide range of Reynolds number. Using a model of a mangrove forest, Shan [22] experimentally evaluated the influence of tree layout (random versus aligned) and tree density on drag forces and velocity. In each tree model, an aerial root system was included. He found that a tree within a randomly

distributed forest may feel a larger or smaller force than a tree at the leading edge, and the variance in drag force increases with tree density, whereas for a tree within an aligned distributed forest, the force on an individual tree within the forest was always less compared to a tree at the leading edge. Using the finite element approach and quadratic shape functions, Sabokrouhiyeh [23] used a two-dimensional depth-averaged model to simulate flow, mass transport, and contaminant removal in a conceptual free-water surface with heterogeneous vegetation patterns in order to investigate the effect of various vegetation patterns on reducing pollutant load and identify optimal vegetation distributions that maximize contaminant removal. Simulations demonstrate that concentration reduction effectiveness improves monotonically with average stem density, but mass removal peaks at an intermediate average stem density value. Xu [24] studied velocity and force on individual plants inside an emergent canopy with real plant shapes and developed a model to predict vertical profiles of velocity and turbulent kinetic energy. By describing the link between the integral length scale and plant shape, he adapted a turbulence model for random arrays of rigid cylinders in order to forecast both the vertical distribution and the channel average of TKE.

 Lou [25] conducted an experiment and mathematical studies to examine the effects of submerged rigid vegetation on sediment suspension under waves, currents, and combined wave-current flows. Lou imitated a canopy from the wooden cylinder in three configurations: sparse, dense, and density that varies vertically. He discovered that lower velocity and more turbulence increase sediment suspension, demonstrating the significance of vegetation-induced turbulence in sediment suspension.

Density and varied vegetation heights exhibited substantial impacts on flow velocity and turbulence, which altered bed shear stress and turbulent diffusivity. For instance, Lou [25] considers the effect of various heights of vegetation in natural ecosystems with four specific heights. However, there is a need to research the effect of random vegetation height on the flow. The result of this study indicates whether the height variation of the vegetation within a single canopy may have a significant effect on the flow.

1.3   Urban Building Literature Review

The rising trend in urbanization [13], [14], energy consumption [15]–[19] and climate change [20], [21] have become one of the most significant issues facing society. Increased human-made urban structures have drastically altered the surface energy balance of the metropolitan area, resulting in a distinct urban climate different from adjacent rural areas [22], [23] in terms of ventilation and warmer ambient temperature known as an urban heat island [24]–[28]. The urban heat island effect causes urban areas to have higher temperatures and slower wind speeds due to the wind shielding effect. In addition, climate change puts lots of species in danger. Extreme heat has increased in frequency and intensity over most land regions, which is identified as one of the main contributors to heat-related fatalities [29]–[34].

Moreover, more than 70% of the world's population is anticipated to reside in cities by 2050; therefore, various research has been done in the field of urban microclimates. It is crucial to understand the urban microclimate since it is essential for a variety of applications. For instance, accurate results for building energy models must be obtained, which requires reliable microclimatic data at the construction site. City planners also require finely resolved climate data in order to improve the thermal comfort of existing urban areas or to construct new urban zones with a high level of thermal comfort. Numerous studies have found that the urban microclimate environment has a major impact on building energy consumption patterns [35], [36]. For example, raising the ambient temperature by 1 K may lead to an increase in building energy usage for summer cooling by 5–10% [37]. Buildings provide a wind sheltering and shading effect in an actual urban context, which can reduce local wind speed patterns and sometimes lower building surface temperatures due to the shading effect [38], [39], which becomes a key factor influencing building energy consumption. According to Liu et al. [40], the local microclimate, measured by exterior surface convective heat transfer coefficients (CHTCs), has a 4 percent impact on total cooling energy usage. Mirsadeghi [41] examined the external CHTC models used by building energy simulation (BES) programs. They demonstrated that the uncertainties in the different correlations might result in more than 30% deviations in the yearly cooling energy demand and 14% in the hourly peak cooling energy demand of an

isolated cubic structure. An essential metric for determining the flow and energy exchange between a street unit and the surrounding air at the microscale is the convective heat transfer coefficient (CHTC) of buildings [42], [43],

$$CHTC = \frac{q}{T_{Wall} - T_{Ref}} \tag{1-1}$$

In the equation 1-1, q is the convective heat transfer normal to the wall surfaces, T_{Ref} is the reference temperature of the air far from the buildings, and T_{Wall} is the wall surface temperature. The CHTC is used to characterize the thermal resistance of the fluid layer near the surface. In most cases, it is determined by means of controlled experiments or numerical simulations with a heat flux and temperature difference that are explicitly stated.

In the literature, convective heat transfer has been studied using analytical solutions [44], [45], outdoor full-scale measurements [46]–[48], wind tunnel [49]–[55] and water tank experiments [56], and numerical simulations (Computational Fluid Dynamic(CFD)) [40], [57]–[66], [66]–[75]. Analytical solutions are usually performed on simple geometry such as flat plate and cylinder [44], [76]–[78], whereas other methods usually investigate CHTC on building facades. The CFD approach, which is a numerical method, is becoming an increasingly attractive way of analysis as a result of the difficulties associated with the conduct of experiments as well as the recent developments in computer resources.

Understanding the relationships between urban buildings and the urban thermal environment is critical for reducing urban heat islands and creating a sustainable and healthy urban constructed environment [79]. Generally, studies for generically distributed building blocks are focused on the effect of different urban geometries (e.g. orientation, density) [80]–[85], urban vegetation patterns [10], [86], [87], building materials [88], building forms [89], [90], air characteristic(e.g. wind speed, wind direction and surface to air temperature differences) [65], [91]. In terms of wind speed, in the summer, every 1 m/s increase in wind speed can prevent a 2 C increase in the temperature of urban air [92].

Iousef [41] examines the influence of several CHTC models on the estimated energy consumption of three types of structures: buildings with more height than width, buildings with less height than width, and buildings with equal height. He proved that the use of various CHTC models might result in deviations of up to 14.5% for the annual heating demand and +42.0% for the annual cooling demand when compared to the conventional CHTC model. Montazeri and Blocken [70] created generalized equations for the average CHTC at all building faces using three-dimensional CFD simulations considering the buildings' reference wind velocity, width, and height. They apply the realizable k-ε model for high Reynolds flow and generate an equation based on the results of 81 separate isolated building simulations using a high-resolution hybrid grid with over two million prismatic and hexahedral cells. Liu [40] examined the effect of plan area densities on CHTC distributions on the windward, leeward, lateral, and top faces of buildings using a numerical simulation of regular arrays of cubic buildings. Buildings are modelled to examine CHTC in an urban environment using Large Eddy Simulations (LES). These buildings are placed according to varied plan area densities to simulate different urban neighborhoods, and they are used to characterize different flow regimes in urban regions. The finding demonstrates that the Reynolds number has little effect on CHTC distributions. To accurately anticipate a building's entire energy balance and energy consumption, the urban density of its surroundings must be taken into account. Liu [93] used a variety of geometrical models to numerically investigate the effect of nearby structures on the wind flow around the target building with Reynolds-averaged Navier-Stokes (RANS) equations with a realizable k-ε turbulence model. The models vary in terms of the building's architectural elements. They considered the intricate building structures around the target building in their computational domain, while the remaining regions were handled as surface roughness. The results demonstrated that the sheltering and channelling effects of the surrounding buildings had a significant impact on the wind flow around the target building. This study determined that the region within a 3L radius of the building must be modelled with precise building structures in order to produce realistic wind flow and pressure distributions around the building. Awol [94] quantitatively examined the effect of constructed area density on the CHTC of buildings using twelve distinct packing densities of regular arrays of cubical structures. He used CFD simulation and the second-moment

closure method, the Reynolds stress turbulence model (RSM) to study the buildings. His findings suggest that the characteristics of convective heat transfer change depending on the flow regime. In the isolated flow regimes, the CHTC trends are characterized by abrupt changes with density. On the other hand, in the interference and skimming flow regimes, the CHTC gradually declines with the growth of building area density.

Studies on CHTC have focused on wind speed, spacing, and wind angle effects. In the majority of studies using generic building layouts, every structure has the same height, and the impact of having a variety of building heights within the same array of buildings is not thoroughly examined. Meanwhile, a small number of studies analyzed the effect of different heights. For example, Tominaga [95] analyzed the ventilation efficiency in a non-uniform building height metropolitan region. Hang et al. [96] investigated the effect of building height variation on urban ventilation. Pillai et al. [97] and Pillai and Yoshie [97] analyzed the heat removal from typical metropolitan regions with uniform and non-uniform building heights using CFD models. Allegrini [98] studied the impact of buoyancy on the mean and local air temperatures in non-uniform building heights. In all these studies, building heights were not randomly selected and followed a pattern.

Moreover, Oke [99] discovered three distinct flow regimes with variable flow topologies based on the proximity of various obstacles. The first is when the distance between buildings is substantially greater than their height, in which case the flow downstream of obstruction is rarely affected by the wake from other obstacles. In this circumstance, known as an isolated regime ($h/s<0.3$), the flow field gradually begins to diverge around a barrier that is completely isolated. The second flow regime is the wake interference flow regime ($0.3<h/s<0.65$). In this instance, there is a significant interaction between the wake of an obstacle and an obstacle downstream. In the third regime, a counter-current vortex is entrapped in the spaces between the obstructions while the flow above the obstacles skims over them ($h/s>0.65$). Due to the varying block heights in this project, the criteria for these regimes have been revised.

The likelihood of tall structures being present is quite high when the height is random, particularly when the standard deviation is high. Within and above the urban canopy,

momentum and scalar exchange are affected by the height of urban structures. Due to the fact that wake effects may be significant over long distances, they are essential for urban-flow modeling at all spatial scales. In addition, the presence of a tall building will change the mean flow, turbulence statistics and instantaneous flow structure of the wake region. On the other hand, the presence of lower structures in the near-wake diffuses the center of the recirculation zone higher, hence decreasing the vertical depth over which flow reversal occurs. Numerous studies have analyzed tall-building aerodynamics to define building design standards [100], [101], wind comfort and safety for pedestrians [102], [103], and/or urban ventilation [104]. The impact of the surrounding environment on tall structures [105], [106] as well as the effect of a tall building's wake on low-lying surroundings, have been studied previously [107].

Daniels [108] investigated the impacts of the inflow turbulence intensity and integral length scales using LES of the flow over tall buildings to construct a model for the vortex-induced vibration and stochastic response of a tall structure in strong non-synoptic wind regimes. Le [109] used LES to examine the aerodynamic behavior of tall structures in two configurations: an isolated building and a building surrounded by complex buildings. The surrounding arrangement has much lower mean pressure values than the isolated setups, by at least 50 percent. The along-wind and torsional responses of the surrounding design are found to be 30% less than those of the isolated configuration. This demonstrates the need of considering the surrounding impacts when studying the pressure distributions and reactions of a tall structure. Cheng [110] used steady state incompressible RANS algorithms and the same exact urban morphology manufactured via 3D printing (at a smaller scale) for wind tunnel studies to confirm the conclusion. There are local maxima of turbulence kinetic energy (TKE) and shear stress at the roof level of these tall structures. In addition, he discovered that the presence of high-rise structures had a substantial effect on the urban boundary layer.

1.4   Machine Learning Background

Numerous Machine Learning (ML) methods exist, including supervised ML[111], semi-supervised ML, and unsupervised ML[112]. In supervised ML, labelled datasets are used to train algorithms that properly categorize data or predict outcomes. The model modifies

its weights as input data are fed into it until the model is well fitted. Data would then need to be divided into two groups called labels and features (A feature is one column of the input set's data. For instance, if you're attempting to predict the price of houses, your input features could include the number of bedrooms, the home's location, the house's age, etc. The label is the price of the house). Unsupervised ML refers to a family of machine learning algorithms that discover patterns in data without anticipating results, meaning that none of the input is labelled.

In building energy assessment, unsupervised machine learning, in particular clustering, may be used to construct data groups with comparable energy characteristics in buildings, which can be used to find anomalies in building energy data[113]. Alternatively, supervised learning algorithms may construct statistical energy models to estimate building energy consumption by modelling complicated connections between inputs and energy performance using simulation results or measurement data. This study aims to assess ML regression models for predicting the surface heat transfer coefficient of a building under random combinations of building height with constant average height, specific standard deviation, and skewness.

1.4.1   Data Preprocessing

Data may be collected from innumerable sources. The raw data must be preprocessed for machines to comprehend it. This technique is known as data preprocessing. Raw data can be in any format, including text, image, etc., and are often incoherent, insufficient, and contain missing values. The raw data deficiency must be addressed to train a model adequately.

The initial phase of preprocessing is to evaluate the consistency and relevance of the overall data to the specific project. The second phase is data cleansing, consisting of inserting missing data and correcting or removing irrelevant data. By discretization and normalization, the data are converted to a suitable format in the third stage. In this project, the first two steps were skipped because the data was collected from a CFD simulation and then inspected for any missing or irrelevant information.

1.4.2   Machine Learning Models

The overall architecture of the machine learning algorithm is depicted in Figure 1-4; the input data are used to train the model, which then generates an output. To investigate the effect of height variation on the convective heat transfer coefficient, a comprehensive analysis utilizing computational fluid dynamics (CFD) and ML algorithms is performed. Several regression algorithms, which will be described in this section, have been used to solve the regression problem in this project; they are described in the following sections.
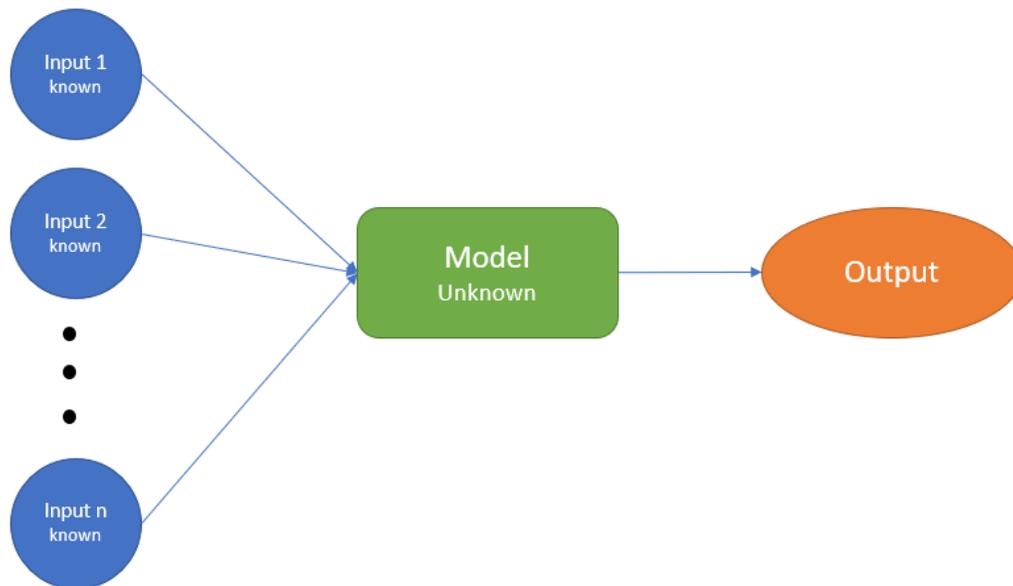
Figure 1-4: Generic Machine Learning Model Architecture

1.4.2.1   Random forest

The Random forest algorithm created by Breiman [114], is an ensemble regression approach. The ensemble's fundamental components are tree-structured predictors, and as each is generated by injecting randomness, the approach is referred to as random forests [115]. The core principle of random forests is to learn a variety of independent decision trees and then apply a consensus approach to forecast the unknown samples. This method combines the bagging sampling strategy and the random selection of features, established independently by Ho [116], Ho [117] and Amit and Geman [118] in order to generate a set of decision trees with a controlled variation. In other words, Random forests generate

numerous decision trees using bootstrapped datasets of the original data and randomly picking a subset of variables at each branch of the decision tree. The model then picks the mode of each decision tree's predictions, so reducing the chance of mistakes associated with a certain tree. The structure of the Random Forest model is illustrated in Figure 1-5.
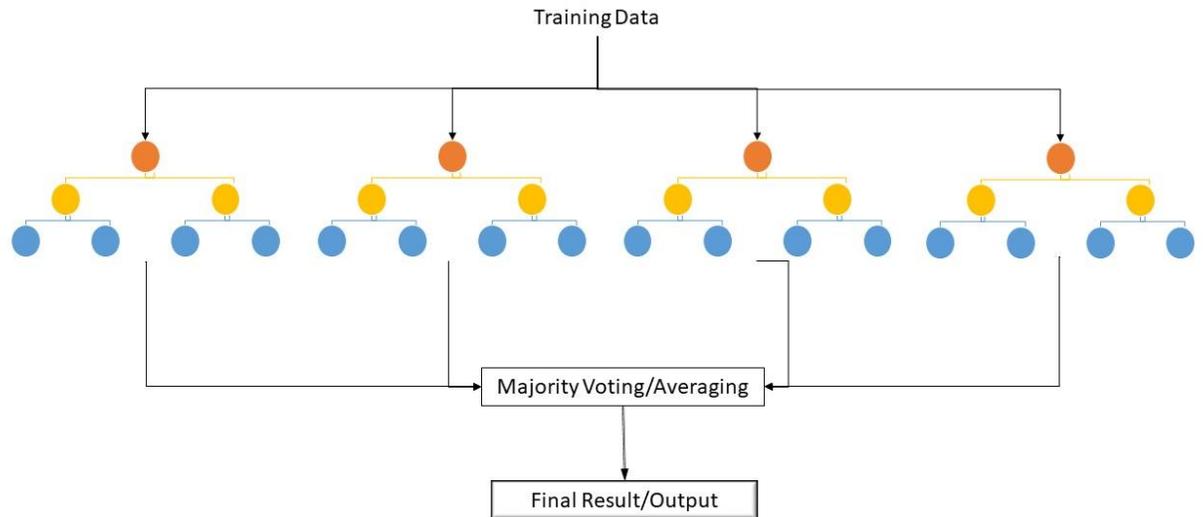
Figure 1-5: Structure of Random Forest Model (Different trees have been created and run in parallel with no interaction, with random forest producing the mean of the classes (or majority vote) as the prediction for all trees.)

1.4.2.2   SVR

It is possible to adapt the notions of support vector machines (SVM), which are used for classification problems, so that they may be applied to regression problems. This method is known as support vector regression (SVR)[119]. SVR has been demonstrated to be an effective technique in real-value function estimation, despite the fact that it is not as widely used as SVM [120]. In the realm of fluid dynamics, the SVM has been used for turbulence modelling [121] and reduced-order modelling [122].

SVMs tackle binary classification issues by recasting them as convex optimization issues [123]. Finding the largest margin separating the hyperplane while properly categorizing as many training points as feasible is the optimization challenge. This ideal hyperplane is represented by support vectors in SVMs. The SVM's sparse solution and strong generalizability facilitate its adaption to regression situations.

To better understand the concept behind it, we should first understand how SVM works. To classify data in the figure below into two groups, we only consider the components around the borders called the support vector. To prevent misclassification, it is necessary to optimize the distance between the border and adjacent components, which is the maximum margin problem. The same techniques can also be applied to regression problems. The regression problem deals with continuous value. The approximation of the continuous-valued function can be represented as equation 1-2.

$$y = \sum_{j=1}^{M} \omega_j x_i + b \qquad y, b \in \mathbb{R}, \qquad x, \omega \in \mathbb{R}^M \tag{1-2}$$

In the above equation, M is the polynomial order used to approximate a function. ω is the weighting coefficient which acts as a regularizing term, and b is bias. SVR formulates this function approximation issue as an optimization problem that minimizes the distance between the expected and desired outputs while searching for the narrowest tube centred on the surface. The objective of the optimization is to minimize the square magnitude of the approximated normal vector to the surface, ||w||, as shown in equation 1-3.

$$\min \frac{1}{2} \|w\|^2 \tag{1-3}$$

SVR employs an ε-insensitive loss function, punishing predictions that are further than ε from the desired output. The width of the tube is determined by the value of ε (for example, for two dimension classification problem ε is represented as the maximum margin in Figure

1-6); a smaller value implies a lower error tolerance and also influences the number of support vectors and, consequently, the sparsity of the solution.
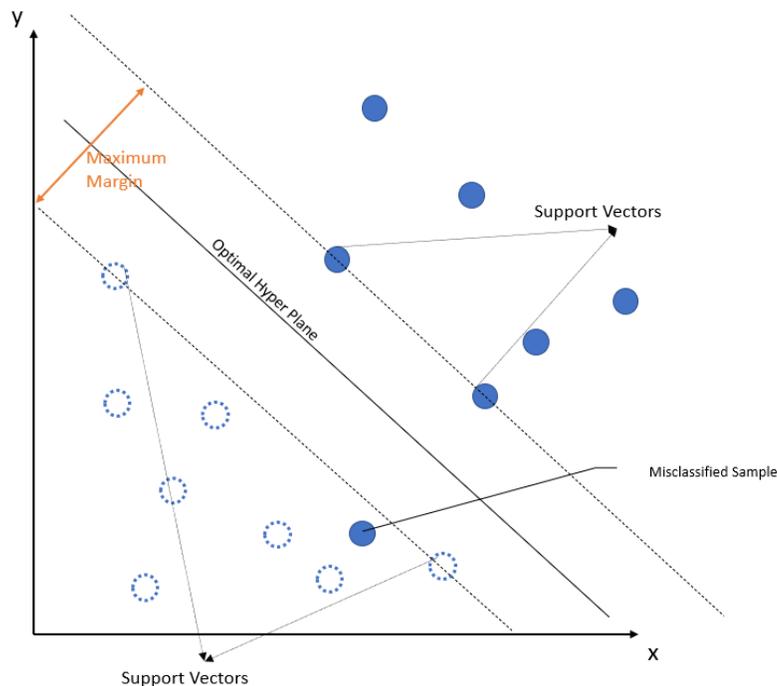
Figure 1-6: SVR Architecture Example. The purpose of the Support Vector is to locate an N-dimensional hyperplane capable of classifying data points. Support Vectors are the data points located on the margin and closest to the Hyperplane. SVR, or Support Vector Regression, may be used to identify and forecast dependent variables when the majority of the data is inside the optimal margins on both sides of the hyperplane.

1.4.2.3   Gradient Boosting Regression Trees (GBRT)

The GBRT technique combines several weak learners using a boosting strategy in which more trees are sequentially added without modifying model parameters in order to minimize the loss function. In the GBRT model, the number of trees, the learning rate, and the maximum depth of the tree are among the most important hyperparameters that significantly impact the model's prediction accuracy. A larger number of trees improves the accuracy of the model's predictions; however, an excessive number of trees may result

in overfitting. The learning rate, on the other hand, governs the contribution of each tree to the predictions, whilst the maximum depth reflects the complexity of each tree.

XGBoost is an efficient and scalable implementation of the gradient boosting framework described by Friedman [124], which contains an efficient linear model solver and tree learning algorithm which was initially developed for classification problems, and then implemented for numerous objective functions, such as regression, classification, and ranking [125]. XGboost is a variation of Gradient Boosted Decision Trees (GBDT) that prevents overfitting by including a regularization term. In addition, XGBoost employs a second-order Taylor series loss function rather than the first-order derivative used in GBDT [126]. Execution speed and model performance are the two primary reasons to employ XGBoost.

1.4.2.4   Ensemble Learning

Ensemble learning is a paradigm for ML that involves combining several different base models in order to achieve better learning results [127]. The primary concept behind this is that by methodically combining the foundational models, an ensemble can produce superior results to those generated by any individual models. There are two ensemble approaches for generating ML learners: the parallel ensemble method and the sequential ensemble method [128]. In the first case, ML learners are generated in parallel, whereas in the latter, they are generated sequentially.

1.4.2.5   Deep learning

Deep learning (DL) is a robust machine learning method that was initially developed by Hinton [129]. The idea of accomplishment of human-level performance is one of the most important milestones of the contemporary ML era[130]. Based on learning theory, ML approaches, enable the generalization of pattern recognition on unseen data[123]. Deep multi-layer architectures and DL approaches improved ML performance by learning high-level data representation[131].

The Artificial Neural Network, or ANN, possibly the most well-known deep learning technique for supervised learning, is a fundamental approximator of nonlinear functions in

a wide range of scientific fields [132]. According to the universal approximation theorem, a sufficiently deep network may approximate any function[133]. A neural network's structure consists of interconnected layers, with each layer containing a distinct number of neurons. The first layer, referred to as the input layer, contains input data that is connected to one or more hidden layers in order to generate a prediction in the output layer. The architecture of a neural network with two hidden layers, five inputs, and three outputs is shown in Figure 1-7. Information is sent between neurons in successive hidden layers through weighted connections, which are then averaged with a bias and scaled using an activation function before being delivered to the next layer. Hyperparameters are values that affect the way the model learns.

ANN architecture can be defined by the number of inputs, outputs, the number of hidden layers, and the number of neurons in each layer.

ML neurons, the fundamental building block of NN, are a simplification of neurons in the human brain. This leads to a modular structure in NN, which provides NNs with strength and adaptability. Each neuron gets an input, processes it using an activation function, and then generates an output. Different neuronal combinations result in distinct NN architectures. Feedforward networks are among the most prevalent architectures; they consist of layers of neurons with weighted outputs serving as inputs to subsequent levels[134]. In this project, feedforward networks have been implemented.
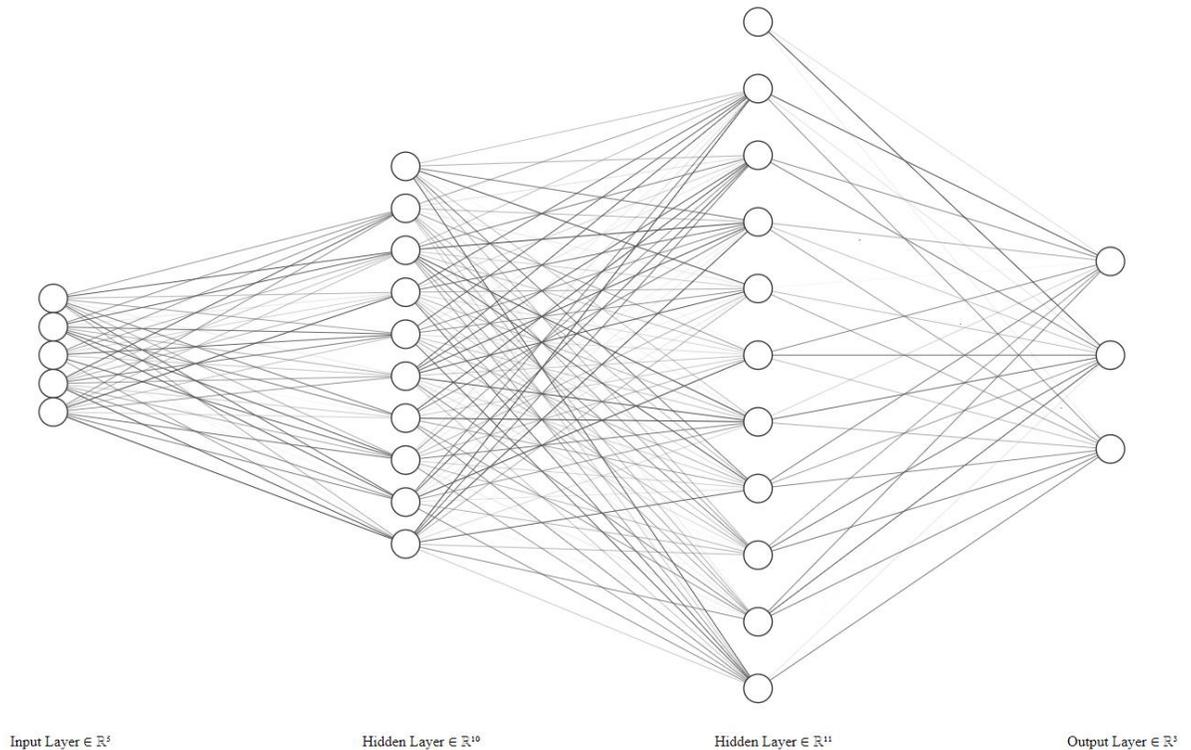
Figure 1-7: An example of neural network architecture with five features in the input layer and two hidden layers with nine and ten neurons in each with a bias parameter in each hidden layer. The neuron on top of the figure that has no connection with the layer before.

1.4.2.5.1   Activation function

The activation function, also known as the transfer function, is a function that receives an input and transforms it in order to produce the desired output. The activation function can be categorized as linear or nonlinear. The linear activation function is unaffected by the complexity and other features of the data supplied to neural networks. In addition, it does not restrict the function's output to a certain range. Nonlinear activation functions are increasingly used owing to their capacity to reduce the model complexity by generalizing and adapting the data with varying outputs.

Simply put, to avoid linearity, activation functions are required. Without them, data would pass across the network's nodes and layers using just linear functions (a*x+b). No matter

how many layers the input passes through, the output is always the result of a linear function since the composite of these linear functions is another linear function.
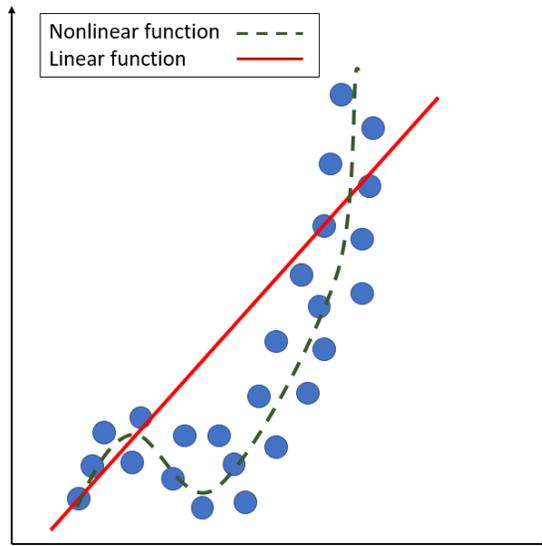
Figure 1-8: An illustration of the advantages of nonlinear functions for data model fitting

Sigmoid, tanh, and ReLu are the three most common nonlinear activation functions, which will be described below.

Sigmoid given in equation 1-4 is an S-shaped plot that ranges from 0 to 1. This function is monotonic and differentiable.

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{1-4}$$

Tanh given in equation   , also known as a hyperbolic tangent, is an S-shaped activation function with a range of -1 to 1. This function is mostly used for classifying problems. One of the benefits of the Tanh function over the Sigmoid function is that negative inputs are mapped strongly negative, and zero inputs are mapped close to zero. Note that when utilizing the Tanh hidden layer function, Xavier Normal or Xavier Uniform weight initialization is suggested.

$$\sigma(z) = \tanh(z) = \frac{2}{1 + e^{-2x}} - 1 \qquad (1\text{-}5)$$

ReLU, which is short for (Rectified Linear Unit) is the most popular activation function and is virtually always employed in convolutional neural networks and deep learning with a range from zero to infinity. One of the downsides of this approach is that it does not effectively map negative outcomes and converts all negative numbers to zero. ReLu is easier to optimize, and it has a computational simplicity advantage.

$$\sigma(z) = \begin{cases} z & z > 0 \\ 0 & z \leq 0 \end{cases} \qquad (1\text{-}6)$$
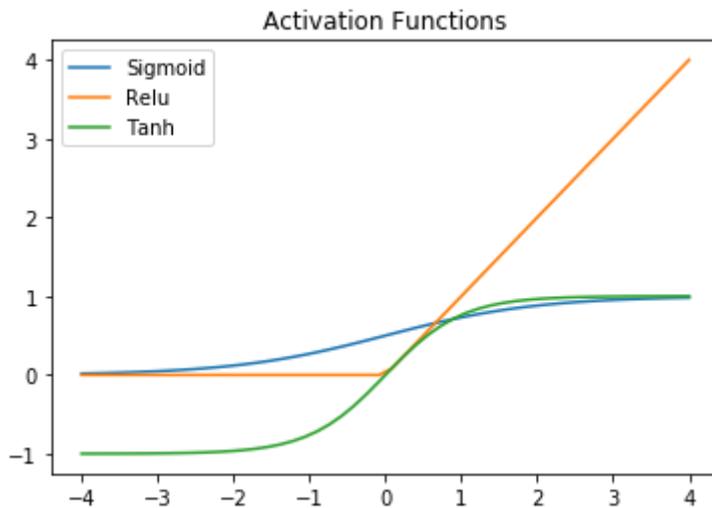
Figure 1-9: Activation functions range

1.4.3   Machine learning Concepts

This section introduces machine learning concepts that have been utilized throughout this project.

1.4.3.1   Errors

In a machine learning algorithm, we should define an error metric as a loss function to assess the results of the model [135]. Mean squared error, MSE, is an important loss function for regression problems. The MSE is computed as the squared difference between predicted and actual values in a dataset given in equation   . The sign is eliminated by squaring these two numbers, resulting in a positive error value. In addition, the impact of squaring is to exacerbate or amplify larger errors. This indicates that the squared positive error increases as the difference between the expected and actual numbers increases. As a result, MSE penalizes models more for greater mistakes as a loss function. The root mean squared error, RMSE, may be viewed as an extension of the squared mean error. The RMSE is the square root of the MSE; therefore, the units are identical to the actual goal value. MAE is linear and intuitive in comparison. The Mean Absolute Error, MAE, is determined by averaging the absolute error values. MSE and RMSE penalize more serious errors than small ones, hence increasing the mean error score. The MAE does not apply different weights to distinct sorts of errors; rather, scores grow linearly as the number of errors increases. MAE and RMSE are calculated as follows (equation 1-7):

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Predicted\ Value_i - Actual\ Value_i)^2 \qquad (1\text{-}7\text{-}1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |Predicted\ Value_i - Actual\ Value_i| \qquad (1\text{-}7\text{-}2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Predicted\ Value_i - Actual\ Value_i)^2}{n}} \qquad (1\text{-}7\text{-}3)$$

Higher values of $RMSE$ and MAE reflect greater error between values. MSE and RMSE are more prevalent in the literature and have been utilized more frequently. The primary objective of this project was to minimize MSE error for each model.

1.4.3.2   Drop out

The Dropout layer reduces overfitting by randomly setting input units to 0 with a rated frequency at each training step, as shown in Figure 1-10, which compares a model with dropout to a model without dropout. Inputs not set to 0 are scaled up by the factor in the equation to maintain the same total sum.

$$scale = \frac{1}{1 - \eta} \tag{1-8}$$

Where η is the dropout rate which is a user-defined constant whose weights are initially scaled based on the selected dropout rate.

| (a) Standard Neural Net | (b) After applying drop out |
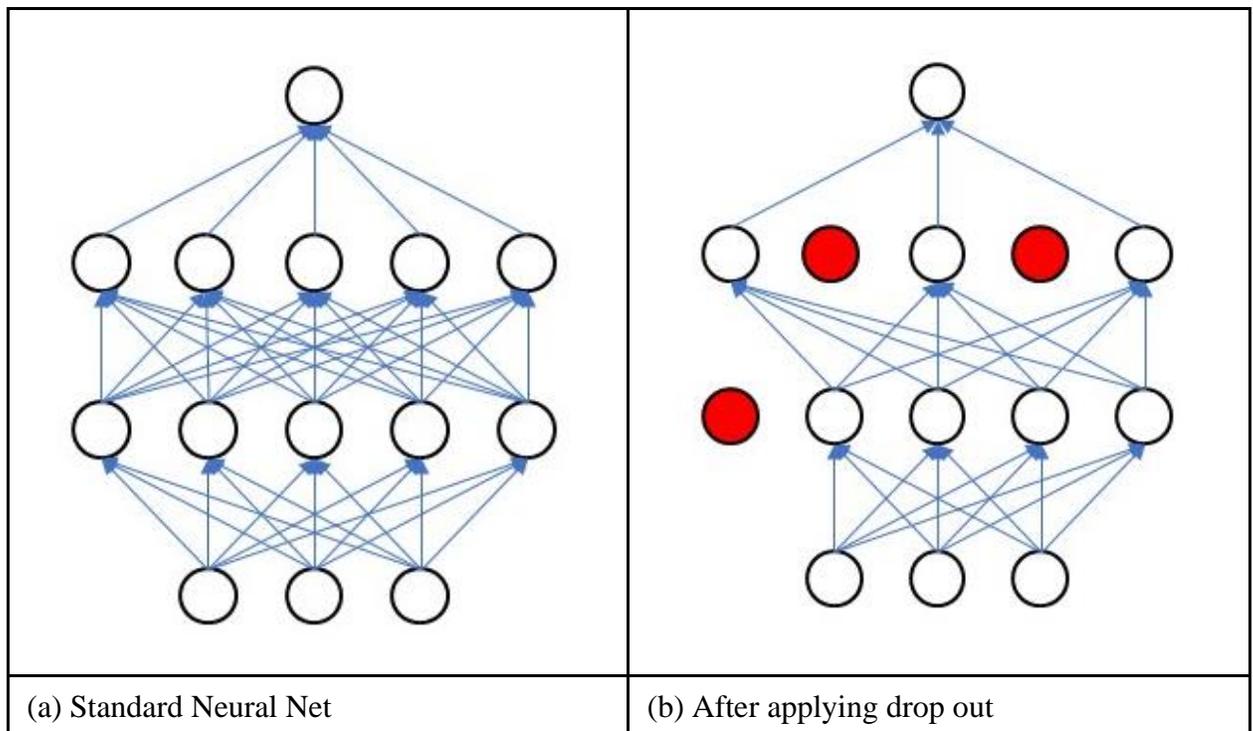
Figure 1-10: Comparison of the architecture of a) standard model and b) model with drop out

1.4.3.3   Feature Importance

Feature importance is a core principle in machine learning that refers to the relative importance of each feature in the training data, i.e., which feature has the most influence on the prediction of the target value. Rogers [136] demonstrated that irrelevant features can impede error convergence which demonstrates the importance of feature importance analysis. Common forms and sources of feature significance scores include statistical correlation scores, coefficients derived from models such as linear models, decision trees, and permutation importance scores. Feature selection within a Random Forest framework has been utilized in this project. Random Forest [114] utilizes the randomization concept [137] to improve the variety of a dataset. This algorithm's basic learners are decision trees that employ information gain as the criteria for splitting each node. In order to determine the ideal binary split for each node, these trees often search through a huge number of potential binary splits for every feature. Due to the random exploration of features, Random Forest is well-suited for feature selection, and the preferred metric of feature relevance is the average information acquisition throughout forest construction. In other words, the variable relevance in the random forest model is determined by permuting each input variable to get the mean squared errors for each tree. The influence of each variable on the model's performance is recognized as a significant factor of that variable [138].

1.4.3.4   Hyperparameter tuning

To develop accurate ML models, it is essential to perform Hyperparameter tuning on of the data, which reduces over-fitting and improves the model's precision [139]. Typically, hyperparameters are user-defined parameters that set machine learning models or are selected based on the hyperparameters used in the same study field. In artificial neural networks (ANN), a hyperparameter may be the number of hidden neurons; in support vector machines (SVM), it would be the regularization parameter [140], [141]. These parameters should be optimized for predicted accuracy. There is no particular mathematical method for optimizing hyperparameters that yields accurate predictions. For example, in the case of ANN, there is no specified procedure for determining the right number of hidden neurons or hidden layers, and they must be chosen based on the particular data set and model. Another crucial aspect of hyperparameter tuning is that the parameters must be

adjusted to prevent overfitting during the training process. The process of hyperparameter tuning involves executing several trials using training data with provided values for the hyperparameters, set within the defined limitations for each hyperparameter. When the task is complete, a summary of all the trials and the optimal hyperparameter configuration are derived based on the selected criteria (for example, minimizing validation set error).

1.5   Machine Learning Literature Review

Artificial intelligence (AI) approaches have lately attracted a great deal of interest because of their exceptional ability to solve a variety of difficult challenges. The goal of AI is to create computing systems capable of rational action or thought [142]. ML is a division of artificial intelligence (AI) and computer science that extracts information from data. The most effective ML algorithms automate decision-making by generalizing from existing cases. Specifically, the algorithm can generate output for an input it has never seen before without human assistance. In the past few decades, numerous studies in the fields of image science, natural language processing, and recommendation systems have led to the commercial success of ML.

ML is a rapidly increasing research field that substantially impacts various disciplines of science and engineering. ML approaches have shown a vast application potential in several engineering disciplines, including wind energy [143], [144], building materials [145], [146], and structural health monitoring [147], [148]. ML has become an integral component of fluid dynamics' experimental, computational, and theoretical aspects. The methodologies available in ML are adequate for handling big data sets and describing the nonlinearity prevalent in fluid flow studies. Notably, it is crucial to promote the use of methodologies that are inspired by ML and data science when trying to solve issues in fluid dynamics. This is mostly important for problems that are difficult to tackle using traditional methods. Numerous objectives in fluid dynamics, including analysis, modelling, estimation, design optimization, and control, may be framed as optimization problems and handled with machine learning. The fact that fluids are nonlinear and may have multiple scales in space and time makes it difficult to find solutions to these problems, which can be seen as high-dimensional and non-convex optimization problems. Fortunately, advances

in ML are helping us improve our capacity to tackle optimization issues that were previously intractable [134].

In order to successfully implement an algorithm for ML, there are various stages that need to be followed. Each of these processes must be guided by a knowledgeable person, as the process is not automatic. The processes may include determining the project's goal, collecting data and preparing it to create a clear data set with adequate features (data preprocessing), picking a ML algorithm that best matches the data and the physics of the problem, and selecting loss functions to assess the performance of the algorithm [149].

In recent years, fluid mechanics has been the subject of extensive study aided by machine learning. For example, Hu [150] uses machine learning to forecast wind pressures around circular cylinders based on the results of published studies. Reynolds number (Re), turbulence intensity (Ti) of the incident wind, and circumferential angle of the cylinder, which significantly influence the wind pressure surrounding smooth circular cylinders, have been utilized as inputs for a machine learning algorithm. Mean and fluctuating pressures were predicted using three machine learning techniques, including decision tree regressor, random forest, and gradient boosting regression trees (GBRT). He demonstrated that the gradient boosting regression trees models offer an effective and affordable substitute to conventional wind tunnel experiments and computational fluid dynamic simulations for calculating wind pressures around two-dimensional smooth circular cylinders. Bre [151] employs an artificial neural network (ANN) to estimate the surface-average pressure coefficient for each wall of buildings(flat-, gable-, and hip-roofed surface of rectangular buildings) based on the building geometry and the wind angle. He used experimental data to create three different ANN models for each roof style. The outcome demonstrates that, compared to frequently employed parametric equations for the determination of pressure coefficients, the implemented ANN models were much more accurate. Kochkov [152] utilized end-to-end deep learning to improve approximations within computational fluid dynamics for modelling two-dimensional turbulent flows by introducing a method for calculating the accurate time evolution of solutions to nonlinear partial differential equations using an order of magnitude coarser grid than is traditionally required for the same accuracy. The suggested machine learning model is Equipped with

the same precision as a powerful numerical solver but with a resolution 8–10 times finer and 40–80 times faster. Hughes [153] used machine learning regression models in order to forecast the heat transfer coefficient and frictional pressure gradient during condensation of three zeotropic mixes in micro and macro channels. Between support vector regression (SVR), random forest regression (RFR), and gradient boost (GB) models, SVR minimizes mean absolute percent error for his data the most effectively. In addition, he conducted a feature importance analysis and determined that the Reynolds number of both phases and a dimensionless temperature gradient are the most essential features in heat transfer calculations. For estimating pressure drop, however, the Bond number, Weber number, and vapour-phase Reynolds number were the most crucial factors. Tiggeloven [154] uses artificial intelligence to forecast the surge component of sea-level fluctuation based on local atmospheric conditions in order to enhance coastal adaptation and management. To forecast hourly surges, he constructed ensembles of Neural Network (NN) models using Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), Long Short-Term Memory layer (LSTM), and a mixture of the latter two (ConvLSTM). He attempted to minimize the mean absolute error for the full-time series in order to make more accurate predictions. Li [155] examined the use of machine learning methods, such as ANNs and CNNs, to estimate surface water flood risks in urban settings. He took into account variables such as coordinates, elevation, slope gradient, imperviousness, land use, land cover, soil type, and normalized differential vegetation index. The machine learning models can create flood risk maps quickly and precisely, while hydrodynamic models, which are typically employed, are computationally costly to build. Tian [138] designed a model to discover energy trends for urban structures. Using ten machine learning methods, he investigated the energy characteristics of London, UK homes from three perspectives: the tuning process of the learning model, variable importance, and spatial analysis of model disagreement. The tuning process of machine learning models determined the intricacy of the correlations between input factors and urban building energy efficiency. The variable importance derived by machine learning models discovered factors of significance for effectively calculating the energy consumption of urban structures. The places with atypical energy consumption patterns have been identified using spatial analysis and

machine learning approaches. The combination of these three factors would significantly enhance our comprehension of the energy behaviors of urban structures.

1.6   Research Gap

Previous research investigating the behaviour of flow through an array of blocks has focused on arrays with identical blocks or a repeating pattern in the variation. No prior research has been done to explore the impacts of a random array or determine whether flow properties correlate well with the statistics of the array. Moreover, the use of ML techniques has been demonstrated to have a significant potential for application in a variety of engineering fields, including fluid flow. Numerous analytical and semi-analytical research have been carried out in order to quantify the heat transfer coefficient; however, to the best of the authors' knowledge, the heat transfer coefficient has never been computed using the ML technique.

1.7   Research Scope

The goals of this work are to:

- Identify correlations between the statistics of a field of blocks with random height and the pressure drop, total heat transfer, and heat transfer coefficient of the flow over the block array.
- Investigate qualitative flow features and how they are affected by the block height variation, including the traditional flow regimes.
- Assess the ability of machine learning models to predict the CHTC of specific blocks within the array based on the heights of a few surrounding blocks.

1.8   Research Outline

The first chapter provides an overview and literature review of the objectives of this research, as well as an introduction to the machine learning techniques used in this project. In chapter 2, a set of CFD simulations of flow over arrays of blocks with random heights is performed in order to examine the effect of heights statistics on pressure drop and heat flux. A model-scale experiment validates the simulation procedure. Chapter 3 discusses the machine learning models used to predict the heat transfer coefficient of blocks based on

the surrounding block heights. Chapter 4 concludes with a discussion of the study's general conclusion and potential research prospects.

References

[1] J. L. Kemp, D. M. Harper, and G. A. Crosa, "The habitat-scale ecohydraulics of rivers," *Ecol. Eng.*, vol. 16, no. 1, pp. 17–29, Oct. 2000.

[2] R. O. Tinoco and G. Coco, "A laboratory study on sediment resuspension within arrays of rigid cylinders," *Adv. Water Resour.*, vol. 92, pp. 1–9, Jun. 2016.

[3] E. de Langre, "Effects of wind on plants," *Annu. Rev. Fluid Mech.*, vol. 40, pp. 141–168, 2008.

[4] H. M. Nepf, "Flow and transport in regions with aquatic vegetation," *Annu. Rev. Fluid Mech.*, vol. 44, pp. 123–142, 2011.

[5] H. M. Nepf, "Hydrodynamics of vegetated channels," *J. Hydraul. Res.*, vol. 50, no. 3, pp. 262–279, Jun. 2012.

[6] J. Lei and H. Nepf, "Evolution of flow velocity from the leading edge of 2-D and 3-D submerged canopies," *J. Fluid Mech.*, vol. 916, 2021.

[7] J. Huang, M. Cassiani, and J. D. Albertson, "Analysis of Coherent Structures Within the Atmospheric Boundary Layer," *Boundary-Layer Meteorol. 2009 1312*, vol. 131, no. 2, pp. 147–171, Feb. 2009.

[8] G. Katul, G. Kuhn, J. Schieldge, and C. I. Hsieh, "THE EJECTION-SWEEP CHARACTER OF SCALAR FLUXES IN THE UNSTABLE SURFACE LAYER," *Boundary-Layer Meteorol. 1997 831*, vol. 83, no. 1, pp. 1–26, 1997.

[9] S. Iousef, H. Montazeri, B. Blocken, and P. van Wesemael, "Impact of exterior convective heat transfer coefficient models on the energy demand prediction of buildings with different geometry," *Build. Simul. 2019 125*, vol. 12, no. 5, pp. 797–816, Apr. 2019.

[10] B. Hong and B. Lin, "Numerical study of the influences of different patterns of the building and green space on micro-scale outdoor thermal comfort and indoor natural ventilation," *Build. Simul. 2014 75*, vol. 7, no. 5, pp. 525–536, Mar. 2014.

[11] A. Beudin, T. S. Kalra, N. K. Ganju, and J. C. Warner, "Development of a coupled wave-flow-vegetation interaction model," *Comput. Geosci.*, vol. 100, pp. 76–86, Mar. 2017.

[12] W. K. Dodds, "Microbes and Plants," in *Freshwater Ecology*, Elsevier, 2002, pp. 124–151.

[13] M. Makvandi, B. Li, M. Elsadek, Z. Khodabakhshi, and M. Ahmadi, "The Interactive Impact of Building Diversity on the Thermal Balance and Micro-Climate Change under the Influence of Rapid Urbanization," *Sustain. 2019, Vol.*

11, Page 1662, vol. 11, no. 6, p. 1662, Mar. 2019.

[14] A. Yiannakou and K. D. Salata, "Adaptation to Climate Change through Spatial Planning in Compact Urban Areas: A Case Study in the City of Thessaloniki," *Sustain. 2017, Vol. 9, Page 271*, vol. 9, no. 2, p. 271, Feb. 2017.

[15] Y. Deng, Z. Feng, J. Fang, and S. J. Cao, "Impact of ventilation rates on indoor thermal comfort and energy efficiency of ground-source heat pump system," *Sustain. Cities Soc.*, vol. 37, pp. 154–163, Feb. 2018.

[16] M. K. Kim, J. Liu, and S. J. Cao, "Energy analysis of a hybrid radiant cooling system under hot and humid climates: A case study at Shanghai in China," *Build. Environ.*, vol. 137, pp. 208–214, Jun. 2018.

[17] X. R. Kong, Y. Deng, L. Li, W. S. Gong, and S. J. Cao, "Experimental and numerical study on the thermal performance of ground source heat pump with a set of designed buried pipes," *Appl. Therm. Eng.*, vol. 114, pp. 110–117, Mar. 2017.

[18] J. Liu, X. Xie, F. Qin, S. Song, and D. Lv, "A case study of ground source direct cooling system integrated with water storage tank system," *Build. Simul.*, vol. 9, no. 6, p. 659, Dec. 2016.

[19] J. Liu, S. Zhu, M. K. Kim, and J. Srebric, "A Review of CFD Analysis Methods for Personalized Ventilation (PV) in Indoor Built Environments," *Sustain. 2019, Vol. 11, Page 4166*, vol. 11, no. 15, p. 4166, Aug. 2019.

[20] L. H. U. W. Abeydeera, J. W. Mesthrige, and T. I. Samarasinghalage, "Global Research on Carbon Emissions: A Scientometric Review," *Sustain. 2019, Vol. 11, Page 3972*, vol. 11, no. 14, p. 3972, Jul. 2019.

[21] Y. Choi, S. Lee, and H. Moon, "Urban Physical Environments and the Duration of High Air Temperature: Focusing on Solar Radiation Trapping Effects," *Sustain. 2018, Vol. 10, Page 4837*, vol. 10, no. 12, p. 4837, Dec. 2018.

[22] A. J. Arnfield, "Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island," *Int. J. Climatol.*, vol. 23, no. 1, pp. 1–26, Jan. 2003.

[23] H. J. S. Fernando, D. Zajic, S. Di Sabatino, R. Dimitrova, B. Hedquist, and A. Dallman, "Flow, turbulence, and pollutant dispersion in urban atmospheresa)," *Phys. Fluids*, vol. 22, no. 5, p. 051301, May 2010.

[24] T. Oke, G. Mills, A. Christen, and J. Voogt, *Urban climates*. Cambridge University Press, 2017.

[25] T. R. Oke, G. T. Johnson, D. G. Steyn, and I. D. Watson, "Simulation of surface urban heat islands under 'ideal' conditions at night part 2: Diagnosis of causation," *Boundary-Layer Meteorol.*, vol. 56, no. 4, pp. 339–358, Sep. 1991.

[26] L. Zhao, X. Lee, R. B. Smith, and K. Oleson, "Strong contributions of local background climate to urban heat islands," *Nat. 2014 5117508*, vol. 511, no. 7508, pp. 216–219, Jul. 2014.

[27] L. Chen, J. Hang, M. Sandberg, L. Claesson, S. Di Sabatino, and H. Wigo, "The impacts of building height variations and building packing densities on flow adjustment and city breathability in idealized urban models," *Build. Environ.*, vol. 118, pp. 344–361, Jun. 2017.

[28] J. Hang *et al.*, "City breathability in medium density urban-like geometries evaluated through the pollutant transport rate and the net escape velocity," *Build. Environ.*, vol. 94, no. P1, pp. 166–182, Dec. 2015.

[29] R. Garcia-Herrera, J. Díaz, R. M. Trigo, J. Luterbacher, and E. M. Fischer, "A Review of the European Summer Heat Wave of 2003," *http://dx.doi.org/10.1080/10643380802238137*, vol. 40, no. 4, pp. 267–306, 2010.

[30] C. Heaviside, S. Vardoulakis, and X. M. Cai, "Attribution of mortality to the urban heat island during heatwaves in the West Midlands, UK," *Environ. Heal. A Glob. Access Sci. Source*, vol. 15, no. 1, pp. 49–59, Mar. 2016.

[31] M. Nitschke, G. R. Tucker, A. L. Hansen, S. Williams, Y. Zhang, and P. Bi, "Impact of two recent extreme heat episodes on morbidity and mortality in Adelaide, South Australia: A case-series analysis," *Environ. Heal. A Glob. Access Sci. Source*, vol. 10, no. 1, 2011.

[32] P. J. Robinson, "On the definition of a heat wave," *J. Appl. Meteorol.*, vol. 40, no. 4, pp. 762–775, 2001.

[33] T. T. Smith, B. F. Zaitchik, and J. M. Gohlke, "Heat waves in the United States: Definitions, patterns and trends," *Clim. Change*, vol. 118, no. 3–4, pp. 811–825, Jun. 2013.

[34] C. Mora *et al.*, "Global risk of deadly heat," *Nat. Clim. Chang. 2017 77*, vol. 7, no. 7, pp. 501–506, Jun. 2017.

[35] N. Santillán-Soto, O. R. García-Cueto, A. A. Lambert-Arista, S. Ojeda-Benítez, and S. E. Cruz-Sotelo, "Comparative Analysis of Two Urban Microclimates: Energy Consumption and Greenhouse Gas Emissions," *Sustain. 2019, Vol. 11, Page 2045*, vol. 11, no. 7, p. 2045, Apr. 2019.

[36] K. Lundgren and T. Kjellstrom, "Sustainability Challenges from Climate Change and Air Conditioning Use in Urban Areas," *Sustain. 2013, Vol. 5, Pages 3116-3128*, vol. 5, no. 7, pp. 3116–3128, Jul. 2013.

[37] M. Kolokotroni, I. Giannitsaris, and R. Watkins, "The effect of the London urban heat island on building summer cooling demand and night ventilation strategies," *Sol. Energy*, vol. 80, no. 4, pp. 383–392, Apr. 2006.

[38] N. H. Wong *et al.*, "Evaluation of the impact of the surrounding urban morphology on building energy consumption," *Sol. Energy*, vol. 85, no. 1, pp. 57–71, Jan. 2011.

[39] J. Li *et al.*, "The Effect of Tree-Planting Patterns on the Microclimate within a Courtyard," *Sustain. 2019, Vol. 11, Page 1665*, vol. 11, no. 6, p. 1665, Mar. 2019.

[40] J. Liu, J. Srebric, and N. Yu, "Numerical simulation of convective heat transfer coefficients at the external surfaces of building arrays immersed in a turbulent boundary layer," *Int. J. Heat Mass Transf.*, vol. 61, no. 1, pp. 209–225, Jun. 2013.

[41] M. Mirsadeghi, D. Cóstola, B. Blocken, and J. L. M. Hensen, "Review of external convective heat transfer coefficient models in building energy simulation programs: Implementation and uncertainty," *Appl. Therm. Eng.*, vol. 56, no. 1–2, pp. 134–151, Jul. 2013.

[42] J. A. Palyvos, "A survey of wind convection coefficient correlations for building envelope energy systems' modeling," *Appl. Therm. Eng.*, vol. 28, no. 8–9, pp. 801–808, Jun. 2008.

[43] J. Chen, G. Augenbroe, Q. Wang, and X. Song, "Uncertainty analysis of thermal comfort in a prototypical naturally ventilated office building and its implications compared to deterministic simulation," *Energy Build.*, vol. 146, pp. 283–294, Jul. 2017.

[44] B. Tashtoush and Z. Kodah, "No slip boundary effects in non-Darcian mixed convection from a vertical wall in saturated porous media. Analytical solution," *Heat Mass Transf. und Stoffuebertragung*, vol. 34, no. 1, pp. 35–39, 1998.

[45] R. Cai, C. Gou, N. Zhang, R. Cai, C. Gou, and N. Zhang, "Explicit analytical solutions of the anisotropic Brinkman model for the natural convection in porous media (II)," *ScChG*, vol. 48, no. 4, pp. 422–430, 2005.

[46] R. D. Clear, L. Gartland, and F. C. Winkelmann, "An empirical correlation for the outside convective air-film coefficient for horizontal roofs," *Energy Build.*, vol. 35, no. 8, pp. 797–811, Sep. 2003.

[47] Y. Liu and D. J. Harris, "Full-scale measurements of convective coefficient on external surface of a low-rise building in sheltered conditions," *Build. Environ.*, vol. 42, no. 7, pp. 2718–2736, Jul. 2007.

[48] J. Shao, J. Liu, J. Zhao, W. Zhang, D. Sun, and Z. Fu, "A novel method for full-scale measurement of the external convective heat transfer coefficient for building horizontal roof," *Energy Build.*, vol. 41, no. 8, pp. 840–847, Aug. 2009.

[49] E. R. Meinders and K. Hanjalí, "Experimental study of the convective heat transfer from in-line and staggered configurations of two wall-mounted cubes," *Int. J. Heat Mass Transf.*, vol. 45, no. 3, pp. 465–482, Jan. 2002.

[50] F. Mallor, C. Sanmiguel Vila, A. Ianiro, and S. Discetti, "Wall-mounted perforated cubes in a boundary layer: Local heat transfer enhancement and control," *Int. J. Heat Mass Transf.*, vol. 117, pp. 498–507, Feb. 2018.

[51] H. Nakamura, T. Igarashi, and T. Tsutsui, "Local heat transfer around a wall-mounted cube in the turbulent boundary layer," *Int. J. Heat Mass Transf.*, vol. 44, no. 18, pp. 3385–3395, Sep. 2001.

[52] E. M. Sparrow, J. W. Ramsey, and E. A. Mass, "Effect of Finite Width on Heat Transfer and Fluid Flow about an Inclined Rectangular Plate," *J. Heat Transfer*, vol. 101, no. 2, pp. 199–204, May 1979.

[53] M. K. Cliyn and V. Natarajan, "Local Heat/Mass Transfer Distributions on the Surface of a Wall-Mounted Cube," *J. Heat Transfer*, vol. 113, no. 4, pp. 851–857, Nov. 1991.

[54] V. Natarajan and M. K. Chyu, "Effect of Flow Angle-of-Attack on the Local Heat/Mass Transfer From a Wall-Mounted Cube," *J. Heat Transfer*, vol. 116, no. 3, pp. 552–560, Aug. 1994.

[55] E. R. Meinders, K. Hanjalic, and R. J. Martinuzzi, "Experimental Study of the Local Convection Heat Transfer From a Wall-Mounted Cube in Turbulent Channel Flow," *J. Heat Transfer*, vol. 121, no. 3, pp. 564–573, Aug. 1999.

[56] Y. Fan, Q. Wang, S. Yin, and Y. Li, "Effect of city shape on urban wind patterns and convective heat transfer in calm and stable background conditions," *Build. Environ.*, vol. 162, p. 106288, Sep. 2019.

[57] A. Vallati, G. Galli, C. Colucci, and P. Oclon, "Influence of the geometrical parameters of urban canyons on the convective heat transfer coefficient," *Therm. Sci.*, vol. 23, no. 2, pp. 1211–1223, 2019.

[58] J. Allegrini, V. Dorer, T. Defraeye, and J. Carmeliet, "An adaptive temperature wall function for mixed convective flows at exterior surfaces of buildings in street canyons," *Build. Environ.*, vol. 49, no. 1, pp. 55–66, Mar. 2012.

[59] T. Defraeye, B. Blocken, and J. Carmeliet, "CFD simulation of heat transfer at surfaces of bluff bodies in turbulent boundary layers: Evaluation of a forced-convective temperature wall function for mixed convection," *J. Wind Eng. Ind. Aerodyn.*, vol. 104–106, pp. 439–446, May 2012.

[60] S. Saneinejad, P. Moonen, T. Defraeye, D. Derome, and J. Carmeliet, "Coupled CFD, radiation and porous media transport model for evaluating evaporative cooling in an urban environment," *J. Wind Eng. Ind. Aerodyn.*, vol. 104–106, pp. 455–463, May 2012.

[61] A. de L. Vollaro, G. Galli, and A. Vallati, "CFD Analysis of Convective Heat Transfer Coefficient on External Surfaces of Buildings," *Sustain. 2015, Vol. 7,*

Pages 9088-9099, vol. 7, no. 7, pp. 9088–9099, Jul. 2015.

[62] Y. He, C. Yuan, C. Ren, W. Wang, Y. Shi, and E. Ng, "Urban ventilation assessment with improved vertical wind profile in high-density cities – Investigations in nighttime extreme heat," *Build. Environ.*, vol. 216, p. 109018, May 2022.

[63] M. Hadavi and H. Pasdarshahri, "Investigating effects of urban configuration and density on urban climate and building systems energy consumption," *J. Build. Eng.*, vol. 44, p. 102710, Dec. 2021.

[64] M. G. Emmel, M. O. Abadie, and N. Mendes, "New external convective heat transfer coefficient correlations for isolated low-rise buildings," *Energy Build.*, vol. 39, no. 3, pp. 335–342, Mar. 2007.

[65] B. Blocken, T. Defraeye, D. Derome, and J. Carmeliet, "High-resolution CFD simulations for forced convective heat transfer coefficients at the facade of a low-rise building," *Build. Environ.*, vol. 44, no. 12, pp. 2396–2412, Dec. 2009.

[66] T. Defraeye, B. Blocken, and J. Carmeliet, "CFD analysis of convective heat transfer at the surfaces of a cube immersed in a turbulent boundary layer," *Int. J. Heat Mass Transf.*, vol. 53, no. 1–3, pp. 297–308, Jan. 2010.

[67] H. Montazeri, B. Blocken, D. Derome, J. Carmeliet, and J. L. M. Hensen, "CFD analysis of forced convective heat transfer coefficients at windward building facades: Influence of building geometry," *J. Wind Eng. Ind. Aerodyn.*, vol. 146, pp. 102–116, Nov. 2015.

[68] J. Liu, M. Heidarinejad, S. Gracik, and J. Srebric, "The impact of exterior surface convective heat transfer coefficients on the building energy consumption in urban neighborhoods with different plan area densities," *Energy Build.*, vol. 86, pp. 449–463, Jan. 2015.

[69] Q. Li, E. Bou-Zeid, W. Anderson, S. Grimmond, and M. Hultmark, "Quality and reliability of LES of convective scalar transfer at high Reynolds numbers," *Int. J. Heat Mass Transf.*, vol. 102, pp. 959–970, Nov. 2016.

[70] H. Montazeri and B. Blocken, "New generalized expressions for forced convective heat transfer coefficients at building facades and roofs," *Build. Environ.*, vol. 119, pp. 153–168, Jul. 2017.

[71] H. Montazeri and B. Blocken, "Extension of generalized forced convective heat transfer coefficient expressions for isolated buildings taking into account oblique wind directions," *Build. Environ.*, vol. 140, pp. 194–208, Aug. 2018.

[72] S. Saneinejad, P. Moonen, T. Defraeye, and J. Carmeliet, "Analysis of convective heat and mass transfer at the vertical walls of a street canyon," *J. Wind Eng. Ind. Aerodyn.*, vol. 99, no. 4, pp. 424–433, Apr. 2011.

[73] P. Karava, C. M. Jubayer, and E. Savory, "Numerical modelling of forced convective heat transfer from the inclined windward roof of an isolated low-rise building with application to photovoltaic/thermal systems," *Appl. Therm. Eng.*, vol. 31, no. 11–12, pp. 1950–1963, Aug. 2011.

[74] P. Karava, C. Mohammad Jubayer, E. Savory, and S. Li, "Effect of incident flow conditions on convective heat transfer from the inclined windward roof of a low-rise building with application to photovoltaic-thermal systems," *J. Wind Eng. Ind. Aerodyn.*, vol. 104–106, pp. 428–438, May 2012.

[75] J. Allegrini, V. Dorer, and J. Carmeliet, "Analysis of convective heat transfer at building façades in street canyons and its influence on the predictions of space cooling demand in buildings," *J. Wind Eng. Ind. Aerodyn.*, vol. 104–106, pp. 464–473, May 2012.

[76] R. Cai and N. Zhang, "Explicit analytical solutions of 2-D laminar natural convection," *Int. J. Heat Mass Transf.*, vol. 46, no. 5, pp. 931–934, Feb. 2003.

[77] Q. Sun, K. S. Choi, and X. Mao, "An analytical solution of convective heat transfer in microchannel or nanochannel," *Int. Commun. Heat Mass Transf.*, vol. 117, p. 104766, Oct. 2020.

[78] M. Massabó, R. Cianci, and O. Paladino, "Some analytical solutions for two-dimensional convection–dispersion equation in cylindrical geometry," *Environ. Model. Softw.*, vol. 21, no. 5, pp. 681–688, May 2006.

[79] G. Chen *et al.*, "Scaled outdoor experimental studies of urban thermal environment in street canyon models with various aspect ratios and thermal storage," *Sci. Total Environ.*, vol. 726, p. 138147, Jul. 2020.

[80] B. Lin, X. Li, Y. Zhu, and Y. Qin, "Numerical simulation studies of the different vegetation patterns' effects on outdoor pedestrian thermal comfort," *J. Wind Eng. Ind. Aerodyn.*, vol. 96, no. 10–11, pp. 1707–1718, Oct. 2008.

[81] X. Yang, L. Zhao, M. Bruse, and Q. Meng, "An integrated simulation method for building energy performance assessment in urban environments," *Energy Build.*, vol. 54, pp. 243–251, Nov. 2012.

[82] K. Perini and A. Magliocco, "Effects of vegetation, urban density, building height, and atmospheric conditions on local temperatures and thermal comfort," *Urban For. Urban Green.*, vol. 13, no. 3, pp. 495–506, Jan. 2014.

[83] M. Taleghani, L. Kleerekoper, M. Tenpierik, and A. Van Den Dobbelsteen, "Outdoor thermal comfort within five different urban forms in the Netherlands," *Build. Environ.*, vol. 83, pp. 65–78, Jan. 2015.

[84] S. Murakami, "Indoor/outdoor climate design by CFD based on the Software Platform," *Int. J. Heat Fluid Flow*, vol. 25, no. 5, pp. 849–863, Oct. 2004.

[85] S. Murakami, "Environmental design of outdoor climate based on CFD," *Fluid Dyn. Res.*, vol. 38, no. 2–3, pp. 108–126, Feb. 2006.

[86] A. Dimoudi and M. Nikolopoulou, "Vegetation in the urban environment: microclimatic analysis and benefits," *Energy Build.*, vol. 35, no. 1, pp. 69–76, Jan. 2003.

[87] M. W. Yahia and E. Johansson, "Landscape interventions in improving thermal comfort in the hot dry city of Damascus, Syria-The example of residential spaces with detached buildings," *Landsc. Urban Plan.*, vol. 125, pp. 1–16, 2014.

[88] P. A. Mirzaei and F. Haghighat, "A procedure to quantify the impact of mitigation techniques on the urban ventilation," *Build. Environ.*, vol. 47, no. 1, pp. 410–420, Jan. 2012.

[89] R. X. Lee, S. K. Jusuf, and N. H. Wong, "The study of height variation on outdoor ventilation for Singapore's high-rise residential housing estates," *Int. J. Low-Carbon Technol.*, vol. 10, no. 1, pp. 15–33, Mar. 2015.

[90] E. Johansson, J. Spangenberg, M. L. Gouvêa, and E. D. Freitas, "Scale-integrated atmospheric simulations to assess thermal comfort in different urban tissues in the warm humid summer of São Paulo, Brazil," *Urban Clim.*, vol. 6, pp. 24–43, Dec. 2013.

[91] I. Lienhard and V. Lienhard, *A Heat Transfer Textbook*, 5th ed. Phlogiston Press, 2020.

[92] V. Cheng, E. Ng, C. Chan, and B. Givoni, "Outdoor thermal comfort study in a sub-tropical climate: A longitudinal study based in Hong Kong," *Int. J. Biometeorol.*, vol. 56, no. 1, pp. 43–56, Jan. 2012.

[93] S. Liu *et al.*, "Influence of surrounding buildings on wind flow around a building predicted by CFD simulations," *Build. Environ.*, vol. 140, pp. 1–10, Aug. 2018.

[94] A. Awol, G. T. Bitsuamlak, and F. Tariku, "Numerical estimation of the external convective heat transfer coefficient for buildings in an urban-like setting," *Build. Environ.*, vol. 169, p. 106557, Feb. 2020.

[95] Y. Tominaga, "Visualization of city breathability based on CFD technique: Case study for urban blocks in Niigata City," *J. Vis.*, vol. 15, no. 3, pp. 269–276, Aug. 2012.

[96] J. Hang, Y. Li, M. Sandberg, R. Buccolieri, and S. Di Sabatino, "The influence of building height variability on pollutant dispersion and pedestrian ventilation in idealized high-rise urban areas," *Build. Environ.*, vol. 56, pp. 346–360, Oct. 2012.

[97] S. Subramania, R. Yoshie, and J. Chung, "Experimental and computational studies on heat transfer from urban canopy and its dependence on urban parameters," *Fifth

Int. Symp. Comput. Wind Eng., 2010.

[98] J. Allegrini and J. Carmeliet, "Coupled CFD and building energy simulations for studying the impacts of building height topology and buoyancy on local urban microclimates," *Urban Clim.*, vol. 21, pp. 278–305, Sep. 2017.

[99] T. R. Oke, "Street design and urban canopy layer climate," *Energy Build.*, vol. 11, no. 1–3, pp. 103–113, Mar. 1988.

[100] J. D. Holmes, "Along- and cross-wind response of a generic tall building: Comparison of wind-tunnel data with codes and standards," *J. Wind Eng. Ind. Aerodyn.*, vol. 132, pp. 136–141, Sep. 2014.

[101] D. K. Kwon and A. Kareem, "Comparative study of major international wind codes and standards for wind effects on tall buildings," *Eng. Struct.*, vol. 51, pp. 23–35, Jun. 2013.

[102] Y. Tominaga *et al.*, "AIJ guidelines for practical applications of CFD to pedestrian wind environment around buildings," *J. Wind Eng. Ind. Aerodyn.*, vol. 96, no. 10–11, pp. 1749–1761, Oct. 2008.

[103] B. Blocken, W. D. Janssen, and T. van Hooff, "CFD simulation for pedestrian wind comfort and wind safety in urban areas: General decision framework and case study for the Eindhoven University campus," *Environ. Model. Softw.*, vol. 30, pp. 15–34, Apr. 2012.

[104] E. Ng, "Policies and technical guidelines for urban planning of high-density cities – air ventilation assessment (AVA) of Hong Kong," *Build. Environ.*, vol. 44, no. 7, pp. 1478–1488, Jul. 2009.

[105] A. Elshaer, H. Aboshosha, G. Bitsuamlak, A. El Damatty, and A. Dagnew, "LES evaluation of wind-induced responses for an isolated and a surrounded tall building," *Eng. Struct.*, vol. 115, pp. 179–195, May 2016.

[106] T. H. Le and L. Caracoglia, "Modeling vortex-shedding effects for the stochastic response of tall buildings in non-synoptic winds," *J. Fluids Struct.*, vol. 61, pp. 461–491, Feb. 2016.

[107] D. Hertwig *et al.*, "Wake Characteristics of Tall Buildings in a Realistic Urban Canopy," *Boundary-Layer Meteorol.*, vol. 172, no. 2, pp. 239–270, Aug. 2019.

[108] S. J. Daniels, I. P. Castro, and Z. T. Xie, "Peak loading and surface pressure fluctuations of a tall model building," *J. Wind Eng. Ind. Aerodyn.*, vol. 120, pp. 19–28, Sep. 2013.

[109] T. H. Le and L. Caracoglia, "Modeling vortex-shedding effects for the stochastic response of tall buildings in non-synoptic winds," *J. Fluids Struct.*, vol. 61, pp. 461–491, Feb. 2016.

[110] W. C. Cheng *et al.*, "Turbulent flows over real heterogeneous urban surfaces: Wind tunnel experiments and Reynolds-averaged Navier-Stokes simulations," *Build. Simul. 2021 145*, vol. 14, no. 5, pp. 1345–1358, Jan. 2021.

[111] P. Saraswat, "Supervised Machine Learning Algorithm: A Review of Classification Techniques," in *Smart Innovation, Systems and Technologies*, 2022, vol. 273, no. 1, pp. 477–482.

[112] D. Ruppert, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction," *J. Am. Stat. Assoc.*, vol. 99, no. 466, pp. 567–567, Jun. 2004.

[113] C. Miller, Z. Nagy, and A. Schlueter, "A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 1365–1377, Jan. 2018.

[114] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[115] M. R. Segal, "Machine learning benchmarks and random forest regression," 2004.

[116] T. K. Ho, "Random decision forests," *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, vol. 1, pp. 278–282, 1995.

[117] Tin Kam Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, 1998.

[118] Y. Amit and D. Geman, "Shape Quantization and Recognition with Randomized Trees," *Neural Comput.*, vol. 9, no. 7, pp. 1545–1588, Oct. 1997.

[119] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput. 2004 143*, vol. 14, no. 3, pp. 199–222, Aug. 2004.

[120] M. Awad and R. Khanna, "Support Vector Regression," *Effic. Learn. Mach.*, pp. 67–80, 2015.

[121] J. Ling and J. Templeton, "Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier Stokes uncertainty," *Phys. Fluids*, vol. 27, no. 8, p. 085103, Aug. 2015.

[122] K. T. Carlberg, A. Jameson, M. J. Kochenderfer, J. Morton, L. Peng, and F. D. Witherden, "Recovering missing CFD data for high-order discretizations using deep neural networks and dynamics learning," *J. Comput. Phys.*, vol. 395, pp. 105–124, Oct. 2019.

[123] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.

[124] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.,"

Ann. Stat., vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[125] Y. Freund, R. E. Schapire, and N. Abe, "A brief introduction to boosting," *Journal-Japanese Soc. Artif. Intell.*, vol. 14, pp. 771–780, 1999.

[126] R. K. Vinayak and R. Gilad-Bachrach, "DART: Dropouts meet Multiple Additive Regression Trees," vol. 38. PMLR, pp. 489–497, 21-Feb-2015.

[127] A. L'Heureux, K. Grolinger, H. F. Elyamany, and M. A. M. Capretz, "Machine Learning with Big Data: Challenges and Approaches," *IEEE Access*, vol. 5, pp. 7776–7797, 2017.

[128] Z. H. Zhou, *Ensemble methods: Foundations and algorithms*. CRC press, 2012.

[129] T. G. Dietterich, "Ensemble methods in machine learning," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2000, vol. 1857 LNCS, pp. 1–15.

[130] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[131] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553. Nature Publishing Group, pp. 436–444, 27-May-2015.

[132] C. Shen, "A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists," *Water Resour. Res.*, vol. 54, no. 11, pp. 8558–8593, Nov. 2018.

[133] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, Jan. 1989.

[134] S. L. Brunton, M. S. Hemati, K. Taira, S. L. Brunton, M. S. Hemati, and K. Taira, "Special issue on machine learning and data-driven methods in fluid dynamics," *Theor. Comput. Fluid Dyn. 2020 344*, vol. 34, no. 4, pp. 333–337, Aug. 2020.

[135] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, pp. 1–24, Jul. 2021.

[136] J. Rogers and S. Gunn, "Identifying Feature Relevance Using a Random Forest," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3940 LNCS, Springer Verlag, 2006, pp. 173–184.

[137] T. G. Dietterich, "Experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization," *Mach. Learn.*, vol. 40, no. 2, pp. 139–157, Aug. 2000.

[138] W. Tian, C. Zhu, Y. Sun, Z. Li, and B. Yin, "Energy characteristics of urban

buildings: Assessment by machine learning," *Build. Simul.*, vol. 14, no. 1, pp. 179–193, Feb. 2021.

[139] J. Bergstra, D. Yamins, and D. D. Cox, "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures," in *30th International Conference on Machine Learning, ICML 2013*, 2013, vol. 28, no. PART 1, pp. 115–123.

[140] P. Tsirikoglou, S. Abraham, F. Contino, C. Lacor, and G. Ghorbaniasl, "A hyperparameters selection technique for support vector regression models," *Appl. Soft Comput. J.*, vol. 61, pp. 139–148, Dec. 2017.

[141] J. Bergstra, D. Yamins, and D. D. Cox, "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures," in *30th International Conference on Machine Learning, ICML 2013*, 2013, vol. 28, no. PART 1, pp. 115–123.

[142] S. Russell and P. Norvig, "A modern, agent-oriented approach to introductory artificial intelligence," *ACM SIGART Bull.*, vol. 6, no. 2, pp. 24–26, Apr. 1995.

[143] R. Becker and D. Thrän, "Completion of wind turbine data sets for wind integration studies applying random forests and k-nearest neighbors," *Appl. Energy*, vol. 208, pp. 252–262, Dec. 2017.

[144] J. Heinermann and O. Kramer, "Machine learning ensembles for wind power prediction," *Renew. Energy*, vol. 89, pp. 671–679, Apr. 2016.

[145] M. Sonebi, A. Cevik, S. Grünewald, and J. Walraven, "Modelling the fresh properties of self-compacting concrete using support vector machine approach," *Constr. Build. Mater.*, vol. 106, pp. 55–64, Mar. 2016.

[146] J. S. Chou, C. F. Tsai, A. D. Pham, and Y. H. Lu, "Machine learning in concrete strength simulations: Multi-nation data analytics," *Constr. Build. Mater.*, vol. 73, pp. 771–780, Dec. 2014.

[147] M. H. Rafiei and H. Adeli, "A novel unsupervised deep learning model for global and local health condition assessment of structures," *Eng. Struct.*, vol. 156, pp. 598–607, Feb. 2018.

[148] V. Alves, A. Cury, N. Roitman, C. Magluta, and C. Cremona, "Structural modification assessment using supervised learning methods applied to vibration data," *Eng. Struct.*, vol. 99, pp. 439–448, Sep. 2015.

[149] S. L. Brunton, "Applying Machine Learning to Study Fluid Mechanics," *Acta Mech. Sin.*, vol. 37, no. 12, pp. 1718–1726, Oct. 2021.

[150] G. Hu and K. C. S. Kwok, "Predicting wind pressures around circular cylinders using machine learning techniques," *J. Wind Eng. Ind. Aerodyn.*, vol. 198, no.

March 2019, p. 104099, Jan. 2019.

[151] F. Bre, J. M. Gimenez, and V. D. Fachinotti, "Prediction of wind pressure coefficients on building surfaces using artificial neural networks," *Energy Build.*, vol. 158, pp. 1429–1441, Jan. 2018.

[152] D. Kochkov, J. A. Smith, A. Alieva, Q. Wang, M. P. Brenner, and S. Hoyer, "Machine learning–accelerated computational fluid dynamics," *Proc. Natl. Acad. Sci.*, vol. 118, no. 21, May 2021.

[153] M. T. Hughes, S. M. Chen, and S. Garimella, "Machine-learning-based heat transfer and pressure drop model for internal flow condensation of binary mixtures," *Int. J. Heat Mass Transf.*, vol. 194, p. 123109, Sep. 2022.

[154] T. Tiggeloven, A. Couasnon, C. van Straaten, S. Muis, and P. J. Ward, "Exploring deep learning capabilities for surge predictions in coastal areas," *Sci. Rep.*, vol. 11, no. 1, p. 17224, Dec. 2021.

[155] Z. Li, H. Liu, C. Luo, and G. Fu, "Assessing surface water flood risks in urban areas using machine learning," *Water (Switzerland)*, vol. 13, no. 24, Dec. 2021.

Chapter 2

2   Fluid Dynamics Numerical Simulations

In order to study the effect of height variation on pressure drop and total heat flux of the block array, and the surface heat transfer coefficient of each obstacle, a set of idealized 2d simulations have been conducted. The obstacles in this study could represent a vegetation canopy or the buildings in an urban area. The simulations are highly idealized to investigate whether the random height variation of obstacle characteristics may significantly alter the flow. This chapter discusses the simulation setup, simulation validation, and simulation results. Qualitative flow regimes are also studied in this chapter.

2.1   Simulation Setup

2.1.1   Geometry, Model, and Boundary Conditions

An idealized row of two-dimensional (2D) rectangular blocks experiences an incoming flow, as shown in Figure 2-1. This is a highly simplified model similar to those frequently utilized in the literature to explore and elucidate the underlying governing mechanisms of multiple environmental fluid flow phenomena [1]–[4]. For example, Lin [5] conducted a numerical study utilizing an idealized two-dimensional street canyon and an experimental study on the Reynolds number independent flow regimes and Reynolds number independence criteria over a broad Reynolds number range in order to address various vortex-flow regimes with similar building configurations but at different scales. Su [6] employed idealized 2D simulation to examine the effects of real trees and their varied structures on flow behaviour.

This study investigates the surface heat transfer coefficient, total heat flux, and total pressure drop across 14 blocks representative of a canopy or buildings. Each block's height is chosen randomly so that an array of blocks has an average height of 1 and specified standard deviations and skewness. As a result, a flow between each pair of blocks may belong to one of three flow regimes: isolation, interference, or skimming. Each simulation includes a 2D array of 14 blocks that may represent a slice through a row of buildings or fully submerged plants. Although the flow characteristics would differ in 3D because the

flow could move around obstacles, a highly idealized 2D flow is used to investigate how significant the effect of random height variation might be in environmental flows.

Incompressible time-averaged continuity, momentum, and energy equations are solved using ANSYS Fluent's steady state coupled pressure-based solver, which is suited for incompressible flow. ANSYS Fluent uses the finite volumes method (FVM) to solve the partial differential equations that describe the conservation of energy, mass, and momentum. From the momentum equations, the velocity field is derived. Solving a pressure or pressure correction equation produced by manipulating continuity and momentum equations yields the pressure field. Note that the bouncy effect is neglected.
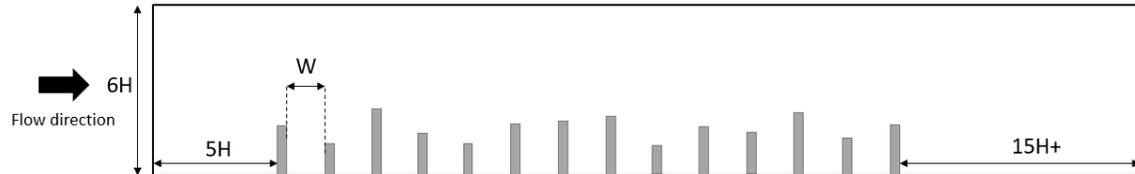
Figure 2-1: Computational Domain

The height of the computational domain is 6H (where H is the average height of the blocks) so that the upper boundary is far enough from the blocks that the flow near it is unaffected by the blocks, and a symmetrical boundary condition is applied to the top surface [7], [8]. According to Tominaga [9] and Franke [10], the domain length downstream of the last block should be stretched to at least 15H to ensure that flow redevelopment is attained behind the wake region. Moreover, a distance of 5H has been set between the inflow boundary and the first building to allow a boundary layer to develop upstream of the first block. The block surfaces are no-slip walls at a consistent temperature of 310K. A pressure outlet condition of zero gage pressure is imposed at the domain's outlet. The inlet velocity is set to 0.15 m/s with a turbulence intensity of 5 percent. The inlet temperature is 300K. The temperature difference of 10 degrees between the inlet and obstacles has been chosen according to the literature recommendation [11], [12]. Townsend [13] argued that when Re crosses a certain threshold, the flow regime will be non-dimensionally similar with Re for the same geometric characteristics. Lin [14] declared that having Re > 11,000 for the flow

inside the street canyon will satisfy Re independence. The Reynolds number in this study (equation 2-1) is close to this threshold.

$$Re = \frac{\rho U_r}{\mu} H = \frac{1.225 \left[\frac{kg}{m^3}\right] * 0.15 \left[\frac{m}{s}\right] * 1[m]}{1.7894e - 05[\frac{kg}{ms}]} = 10268 \qquad (2\text{-}1)$$

2.1.2   Gid Independence Test

A mesh was generated for the geometry described above is shown in Figure 2-2. It has one hundred fifty thousand polyhedral cells with a cell size of 0.1H and hexagonal cell dominance. The area around the blocks, their edges, and the wake zone are refined. Five inflation layers using prismatic cells with a growth rate of 1.2 on the bottom and block surface have been defined. A grid sensitivity test investigated various element sizes around each block, and the total pressure drop of each simulation has been compared to check the convergence to best balance simulation accuracy and computational cost. Due to the fact that the change in pressure drop for grid sizes smaller than 0.1 is less than 2%, the grid size of 0.1 was chosen for these simulations (shown in Figure 2-3).
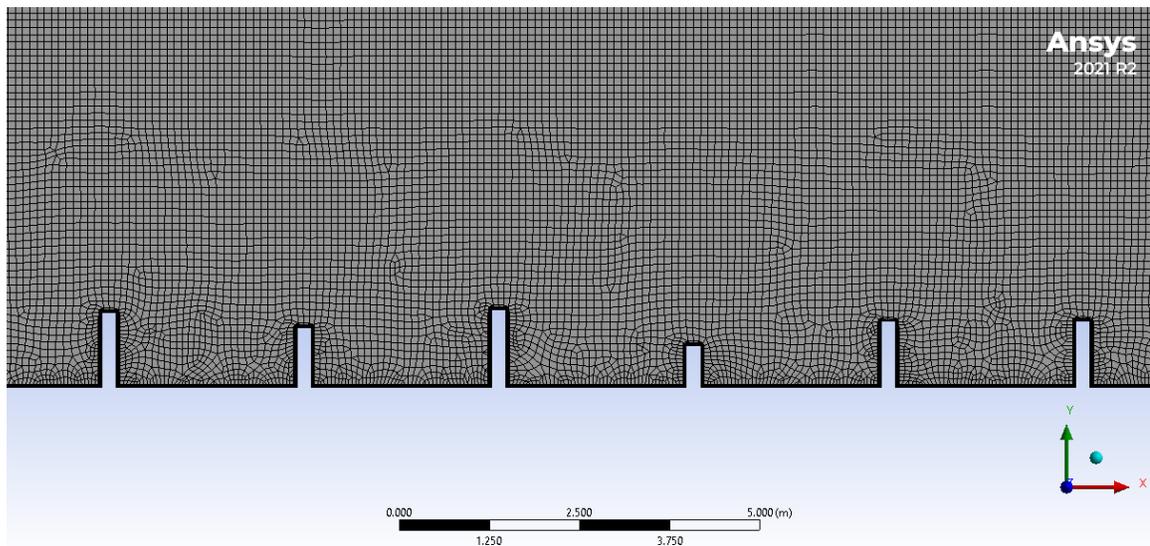
Figure 2-2: Generated mesh in ANSYS
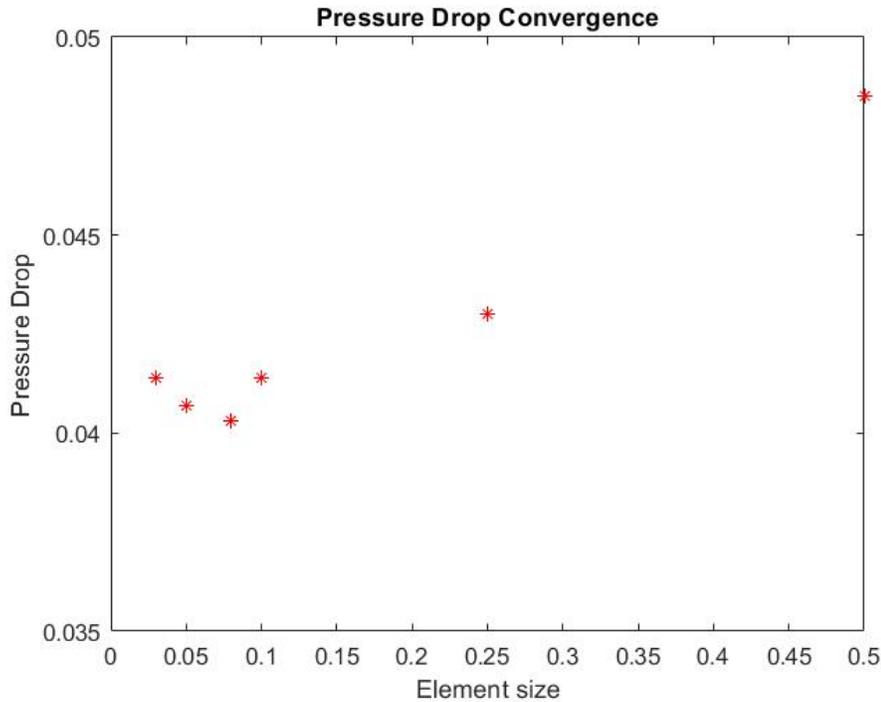
Figure 2-3: Pressure drop vs element size to check grid sensitivity

Two hundred and four simulations with different combinations of block heights have been conducted in order to investigate statistical trends and collect enough data to implement machine learning. To determine statistical trends, some statistical measurements should be held constant while others are altered to determine the impact of each variable on the final result. Simulations were conducted for random block heights with an imposed mean height value of 1, standard deviation values of 0.1, 0.2, 0.3, 0.4, skewness values of -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, and first block height values of 0.75, 0.90, 1.10, 1.20.

2.1.3   Turbulence model

CFD modelling is widely used to assess geophysical flows such as urban physics and microclimate [15]–[17] and vegetation canopies[18]. The two most common turbulence models for this flow are Large Eddy Simulation (LES) and Reynolds-averaged Navier-Stokes simulations (RANS). The computational cost of a microclimatic CFD simulation that links the temperature and velocity fields is significantly increased when LES is used over RANS. It is anticipated that the greater computing cost of LES and often adequate

accuracy of RANS [19]–[22] are the two primary reasons why the vast majority of research was conducted using RANS, even though LES is typically regarded as more accurate than RANS [16], [21], [23]–[25].

Two equation turbulence models are among the most prevalent RANS-based turbulence models. Models such as the k-ε model and the k- ω model are often used for the majority of engineering challenges. Since the standard k-ε model cannot be integrated all the way to the wall, it has a hard time predicting flows, especially when the creation of turbulence is substantially greater than its dissipation, such as in flow separation and impingement. The shear stress transport (SST) *k-ω* turbulence model presented by Menter [26] has advantages over both k-ω and k-ε. Moreover, The SST k-ω model is distinguished by an enhanced near-wall treatment that transitions automatically between the low and high Re number ω formulation [27]. In other words, the SST k-ω model activates the standard k-ω model close to the surface and the k-ε model farther away [28]. Pang [29] demonstrates that the SST k-ω model is able to properly forecast a high Reynolds number flow around a cylindrical bluff body compared to the other two-equation RANS models and that it can capture the impacts of surface roughness. In general, two-equation turbulence models cannot resolve all complicated flow field features around bluff bodies [9], [30], [31], [31]–[33] without a fine mesh. In this study, SST k-ω has been used with fine refinement and inflation layers near the walls. The numerical method is validated using previous experimental results in section 2.1.4.

The following (equation 2-2) are two advection-diffusion equations for turbulent kinetic energy, k, and turbulent frequency, ω [34]–[37], which are used in the SST k-ω:

$$\frac{\partial(\rho k)}{\partial t} + \frac{\partial}{\partial x_i}(U_i \rho k) = S_k - Y_K + \widetilde{G}_K + \frac{\partial}{\partial x_j}\left(\Gamma_k \frac{\partial}{\partial x_j} k\right)$$

$$\frac{\partial(\rho \omega)}{\partial t} + \frac{\partial}{\partial x_i}(U_i \rho \omega) = \frac{\partial}{\partial x_j}\left(\Gamma_\omega \frac{\partial}{\partial x_j} \omega\right) + G_\omega - Y_\omega + D_\omega + S_\omega$$

(2-2)

In equations (2-2), $\widetilde{G}_K$ stands for the generation of kinetic energy associated with turbulence caused by mean velocity gradients. G_ω represents the generation of ω. Γ_k and Γ_ω stand for k's and ω's respective effective diffusivities. Y_K and Y_ω represent the dissipation of k and ω due to turbulence. D_ω represents the cross-diffusion term. S_k and S_ω are user-defined source terms.

Menter defines the effective viscosities (kg m^{-1} s^{-1}) as follows (equation 2-3):

$$\Gamma_k = \mu + \mu_t \frac{1}{\sigma_k}$$

$$\Gamma_\omega = \mu + \mu_t \frac{1}{\sigma_\omega}$$

(2-3)

In the equation above, σ_k and σ_ω are the turbulent Prandtl numbers for k and ω. μ_t is turbulent viscosity, and it is calculated as follows (equation 2-4):

$$\mu_t = \frac{\rho k}{\omega} \frac{1}{\max\left(\frac{1}{\alpha^*}, \frac{S_t\, F_2}{\alpha_1 \omega}\right)}$$

(2-4)

S_t is the strain rate magnitude and F_2 is a blending function. Further information about the implementation of this model can be found in ANSYS documentation 4.5.2.

2.1.4 Validation

Validation with measured data is an important step in using CFD for numerical investigations. Meinders's [38] experimental data, an experiment with cube arrays, similar to the setup we used here, was used to validate the numerical methods used in this study. Khan and Saha [39], Awol [40], Montazeri and Blocken (2018), Liu (2013) and others have also used this experimental data to validate their simulation. Meinders' (1998) experiment was carried out in a wind tunnel with a 500 mm $\times$ 50 mm test section. Nine identical cubes were positioned along the longitudinal flow direction at half of the vertical channel height.

The cube size and spacing between the cubes are both 15 mm (aspect ratio W/H =1). The outside layer of the cube was a 1.5 mm epoxy shell, while the inside part of the cube had a copper core of 12 mm. The copper core was kept at a constant 75 °C. Meanwhile, the conductivity of the epoxy substance was around 0.24 W/m.K. The schematic of this experiment can be found in Figure 2-4.
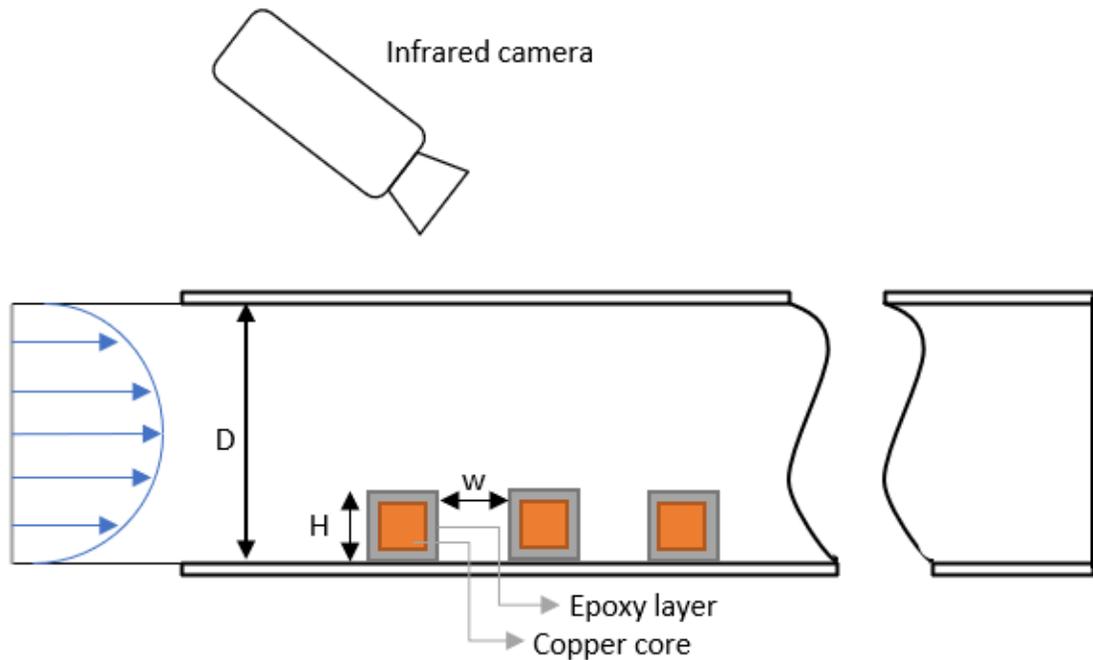
Figure 2-4: Schematic of experimental configuration and measurement cube.

After accounting for radiative losses from supply flux, an infrared camera was utilized to scan the external surfaces of the cubes to obtain the temperature distribution necessary to determine the convective heat transfer coefficient. This experiment was carried out with a bulk velocity of 5.1 m/s, yielding a Reynolds number of around 5065 based on the cube height. The incoming airflow temperature was set at 21 degrees Celsius, which was also used as the far field air temperature to calculate the CHTCs.

The simulation model was a one-to-one scale copy of the tunnel section with a downstream fetch size of 15H to verify that the downstream wake effects were fully included within the computational domain. Following the recommendation of Tominaga [9] and Franke et al.

[10], an upstream fetch of 5H was used. Moreover, the width of the air tunnel is set to 3.3H, consistent with Liu[42] and Awol[43].
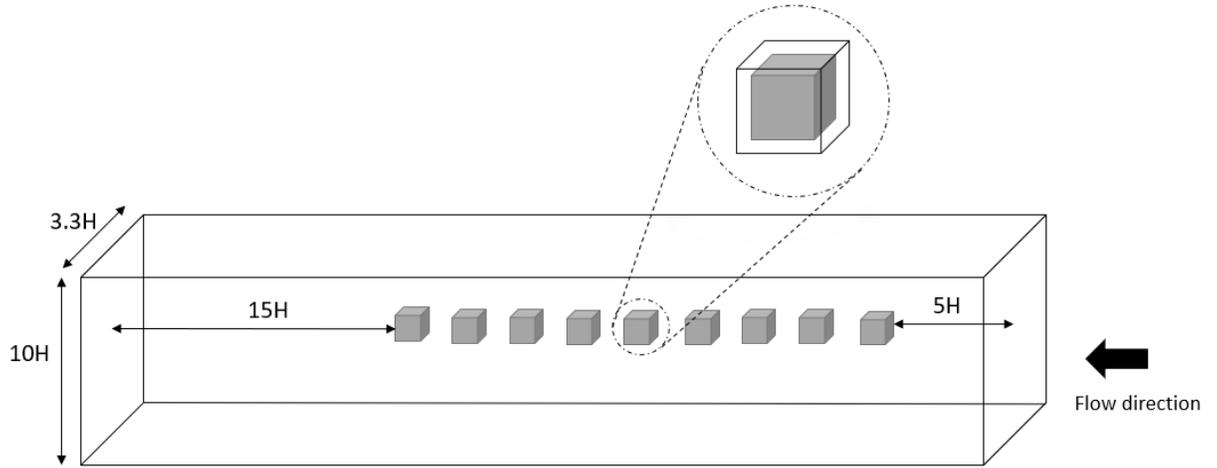
Figure 2-5: The computational domain

To match the conditions in the experiment, the wall on which the cubes are placed is oriented vertically. The epoxy-copper interface inside the cube set is set to 75 °C. The incoming airflow is heated to 21 degrees Celsius. The outlet pressure is kept constant. The computational domain's lateral, top, and bottom faces are all considered adiabatic no-slip walls. The cubes' outer surfaces are configured as no-slip walls with non-adiabatic environmental conditions.

The material properties used in this simulation are in Table 2-1:

Table 2-1: Material property of the experiment

	Air	Epoxy
Density (kg/m^3)	1.225	1668
Specific Heat Capacity ($\frac{J}{kg.K}$)	1006.4	1150
Thermal Conductivity ($\frac{W}{mK}$))	0.0257	0.24
Viscosity ($\frac{Kg}{m.s}$)	2.197e-5	

The inlet velocity profile is specified to match the velocity from the experiment. The lower near logarithmic and middle uniform portions of the inlet velocity profiles are immediately applied from Meinders' digitized database of experimental inputs. The upper laminar component of the profile is derived from the work of Awol[43] (2020), which was driven by the recycling approach. An 8-order polynomial has been fit to the data, shown in Figure 2-6, and used as the inlet velocity in the simulation.

The generated equation is as follows (equation 2-5):

$$
\begin{aligned}
\frac{U}{U_\infty} = &-0.079649 \times \left(\frac{Z}{Z_h}\right)^8 + 1.003 \times \left(\frac{Z}{Z_h}\right)^7 - 5.2009 \times \left(\frac{Z}{Z_h}\right)^6 \\
&+ 14.3266 \times \left(\frac{Z}{Z_h}\right)^5 - 22.5728 \times \left(\frac{Z}{Z_h}\right)^4 \\
&+ 20.4461 \times \left(\frac{Z}{Z_h}\right)^3 - 10.3134 \times \left(\frac{Z}{Z_h}\right)^2 \\
&+ 2.9312 \times \left(\frac{Z}{Z_h}\right) + 0.41711
\end{aligned}
\tag{2-5}
$$
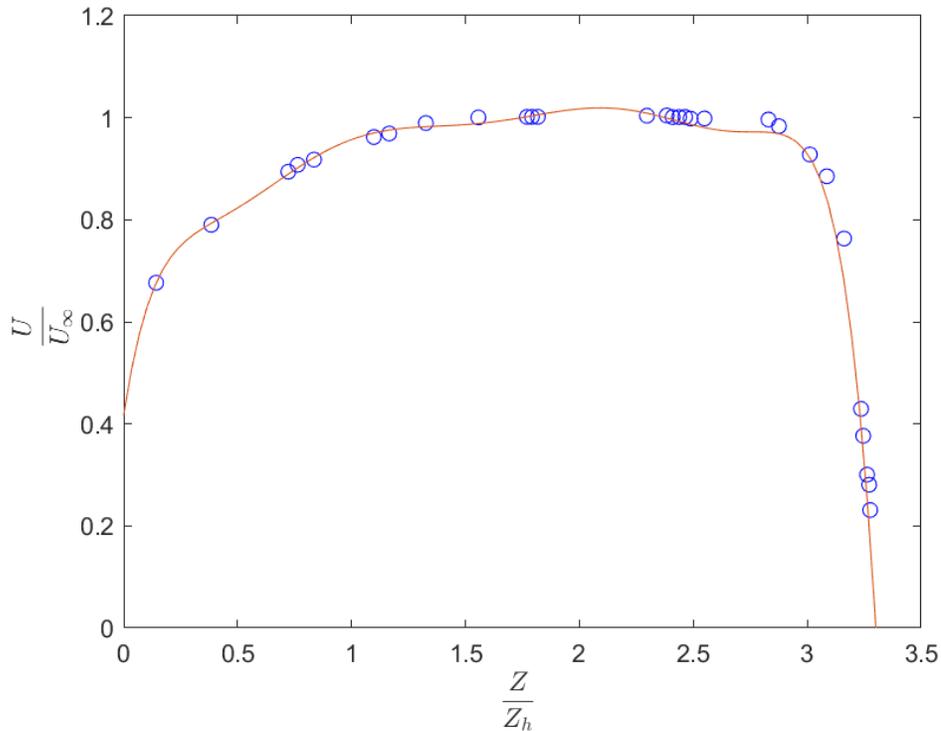
Figure 2-6: The adopted mean inlet velocity profile

In the experiment, turbulence intensity has also been defined using a profile. However, in ANSYS software, the turbulence intensity cannot be specified as a profile and must be constant. As a result, we analyze the effect of different turbulence intensities as well as the influence of varying bulk velocities (U_∞) to ensure the accuracy of the velocity profile and the best turbulence intensity.

The surface temperature distribution of the fifth cube in a row is compared between the experiment and simulation. The results are collected at the intersection points (AB, BC, and CD) of the cube and a vertical plane running longitudinally across the cube, as well as the horizontal mid-line on the windward (EF), lateral (FG, HE), and leeward (GH) surfaces (See Figure 2-7).
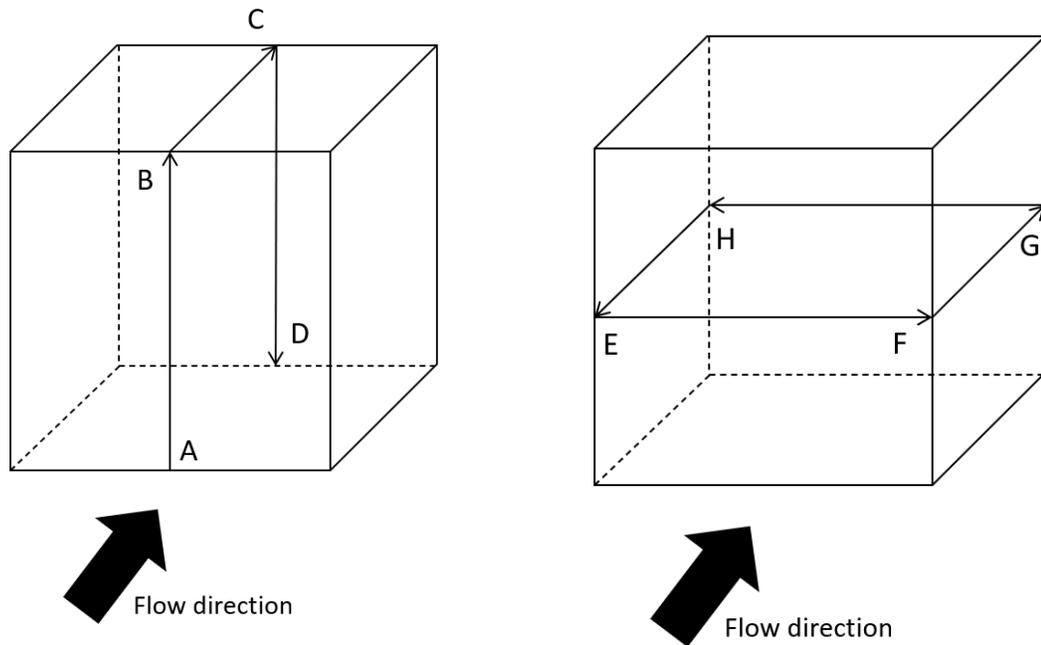
Figure 2-7: Schematic showing the paths along which temperature is recorded and compared between the experiment and simulations

The effect of conduction heat transmission and temperature on the cube's exterior surface was investigated in two ways. The first method, used by Liu (2013), involves modelling and meshing the epoxy shell of the cubes to examine conduction heat transfer. The second technique, similar to Awol (2020), determines the temperature on the outer surface of the cubes without modelling the interior section, using the shell model in ANSYS. The effect of the cube modelling approach, as well as several other modelling options, are discussed in the following sections to quantify how well the numerical methods replicate the experimental results.

A mesh with 8 million polyhedral cells with refinement around the cubes was chosen to assure computational correctness, and a grid sensitivity test was performed by changing the element size around the cube in the range of 0.01, 0.005, 0.001, and 0.0005. The temperature distribution along line AB has no significant change in element size of 0.001 and lower. It's worth noting that refinement was chosen in order to get a low y+ value for cells that are close to the walls. The grid shape is illustrated in Figure 2-8.

a)

b)
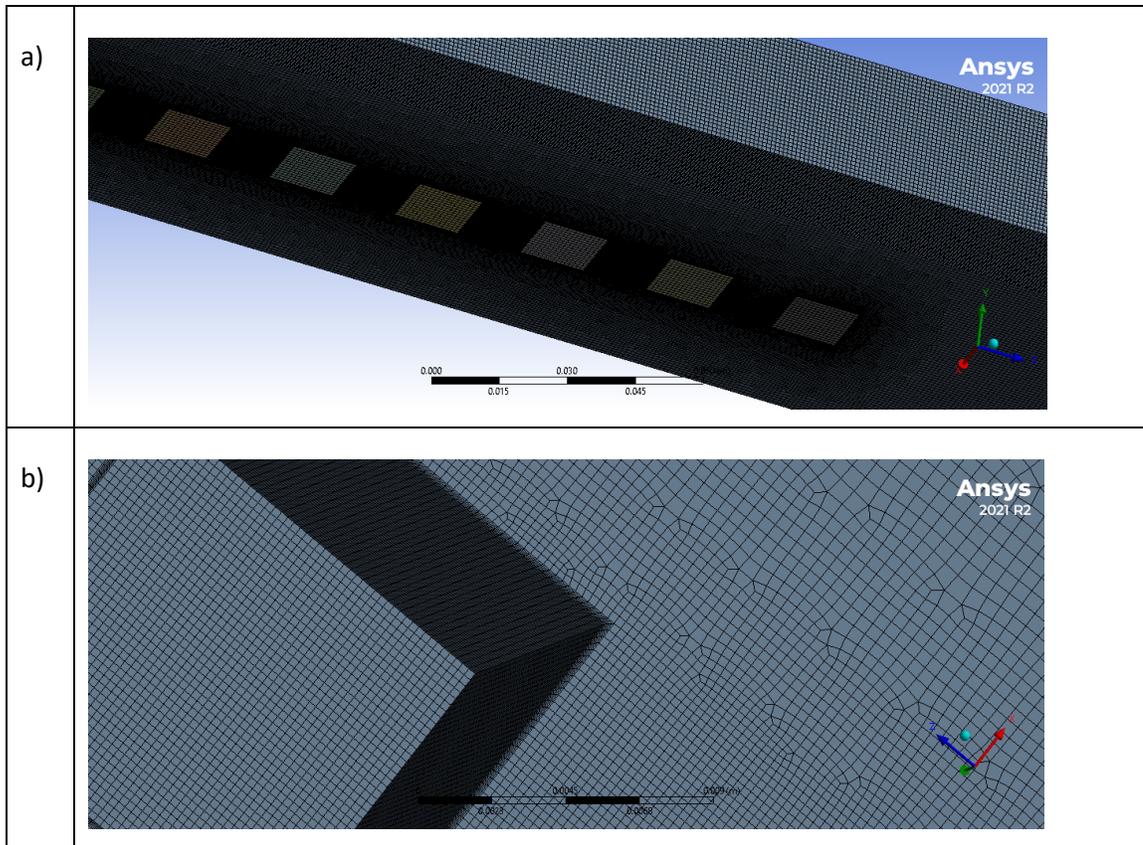
Figure 2-8: Mesh shape for validation

2.1.4.1　Comparison of modelling cubes versus using shell property

Temperature and conduction boundary conditions can be applied to cubes in two ways. To calculate the temperature on the outer surface, one method is to create geometry and mesh for the epoxy layer of the cubes and define the material properties. The other option is to define properties in ANSYS using the shell properties model. The second method is more computationally efficient because it eliminates the need to mesh the small cubes. The outcomes of these two approaches are compared using the same boundary conditions; the results are shown in Figure 2-9.

a)

b)
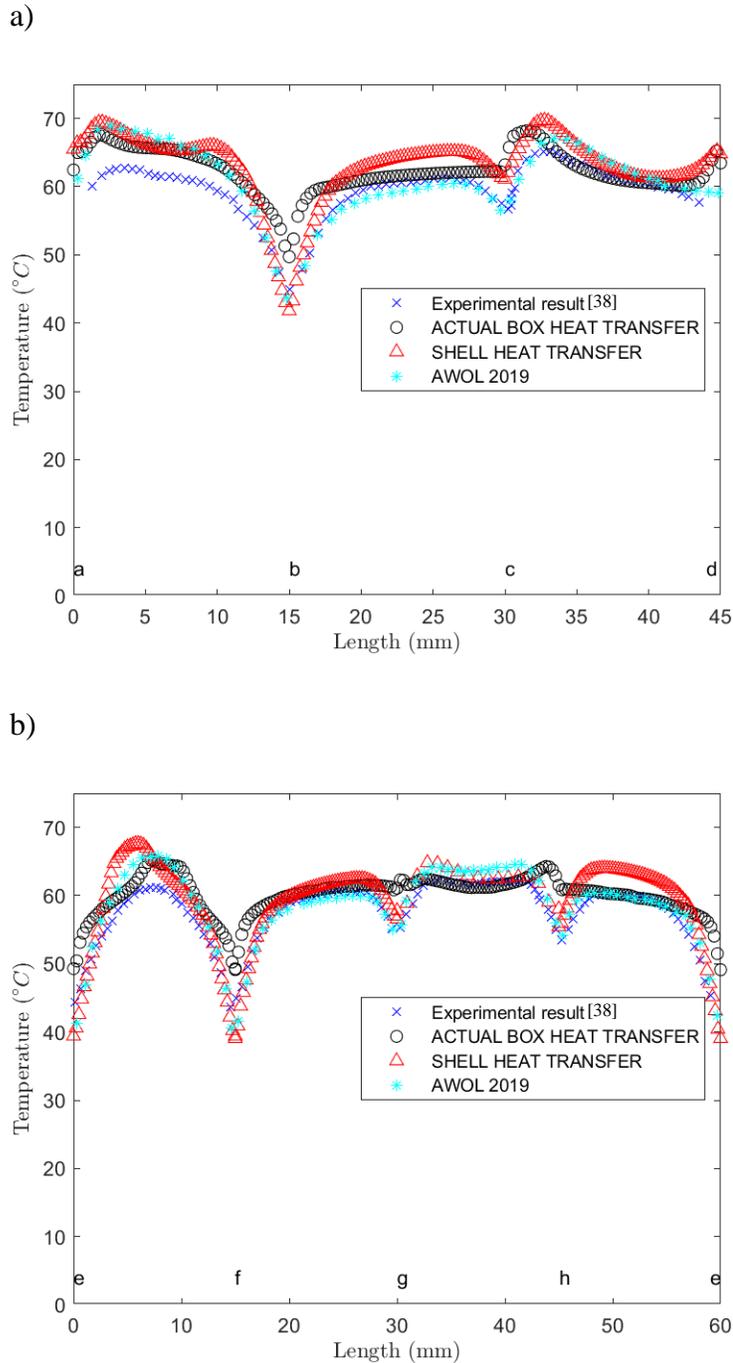
Figure 2-9: Comparison of the model using shell properties versus the model defining the actual cubes along the paths: a) ABCD b) EFGHE

The mean relative error (MRE) of the temperature along the two paths of these two simulations, calculated using equation 2-6, is nearly identical, as shown in Table 2-2. In the formula, avg, exp and sim stand for average value, experiment and simulation,

respectively. The average value refers to the average temperature of the lines; for example, the MRE of ab-bc-cd is the average temperature of lines ab, bc, and cd. The model that used the shell property predicted the cube's corner better but had lower accuracy in the middle of each face. As a result, we cannot say with certainty which model is superior.

$$\text{MRE} = \frac{T_{avg,exp} - T_{avg,sim}}{T_{avg,exp}} \times 100 \qquad \text{(2-6)}$$

Table 2-2: Mean relative error of different models

	ERROR ab-bc-cd	ERROR ef-fg-gh-he
Shell	5.01	4.88
Solid	6.11	4.49

2.1.4.2 Effect of different turbulence intensities using shell property

In the experiment and Awol(2020) work, the inlet turbulence intensity is defined as a profile rather than a constant value. In this project, the turbulence intensity was unable to be defined as a profile due to the limitations of the software and computation power. Therefore, a series of simulations with varying turbulence intensities have been conducted to evaluate the effect of turbulence intensity and to determine the optimal stimulation for validation purposes. In this section, simulations have been done using shell properties instead of meshing the epoxy layer of the cube. The velocity has been defined using the velocity profile in AWOL work while assuming $U_{inf} = 5.1 \text{ m/s}$. The turbulence intensity has been assigned three different magnitudes of 2%, 5% and 8%, and the results are compared in Figure 2-10.

a)
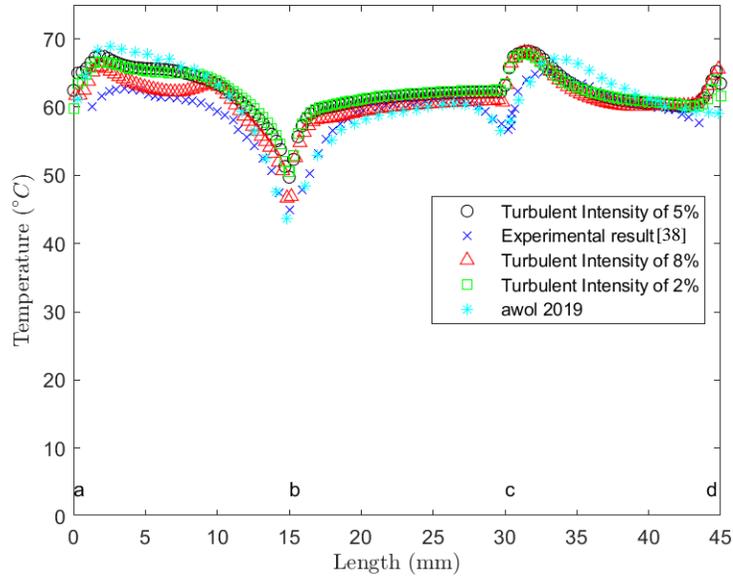
b)
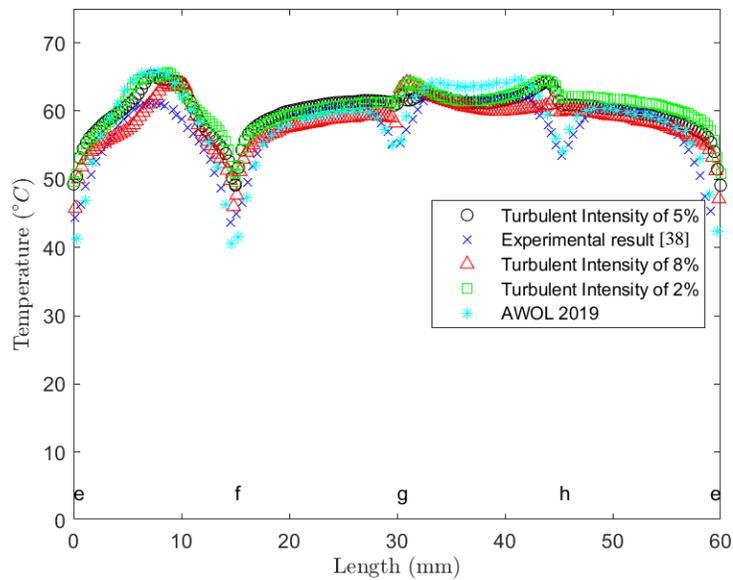
Figure 2-10: Comparison of different turbulence intensity results using shell properties along the paths: a) ABCD b) EFGHE

The MRE, given in equation 2-6, of different turbulence intensities compared with the experimental results has been calculated (Table 2-3).

Table 2-3: Mean relative error of different turbulence intensity results using shell properties

	ERROR ab-bc-cd	ERROR ef-fg-gh-he
TKE2	5.27	5.83
TKE5	5.01	4.88
TKE8	3.30	3.94
Awol	3.54	2.92

The result shows that the turbulence intensity of 8% has the lowest relative error using the shell property. As discussed earlier, turbulence intensity in the experiment follows a profile and is not a constant value; however, in this simulation, using a value of 8% would result in an acceptable error. The MRE of the simulation in this study is larger than the MRE of the Awol simulation. However, the error is still within an acceptable range.

2.1.4.3   Effect of bulk velocity

In this section, the cubes' epoxy layer was modelled instead of using shell properties, and the epoxy-copper interface temperature boundary condition was specified. Using the velocity profile from the Awol work, the bulk velocity or U_{inf} has been varied to explore the effect of bulk velocity on simulation results and ensure the accuracy of the velocity profile. Three different magnitudes of 4, 5.1 and 8 m/s have been set for bulk velocity, while the inlet bulk velocity of the experiment was 5.1 m/s. The results shown in Figure 2-11 indicate that a bulk velocity of 8 m/s best matches the experimental data, although all simulations provide reasonable results.

a)

b)
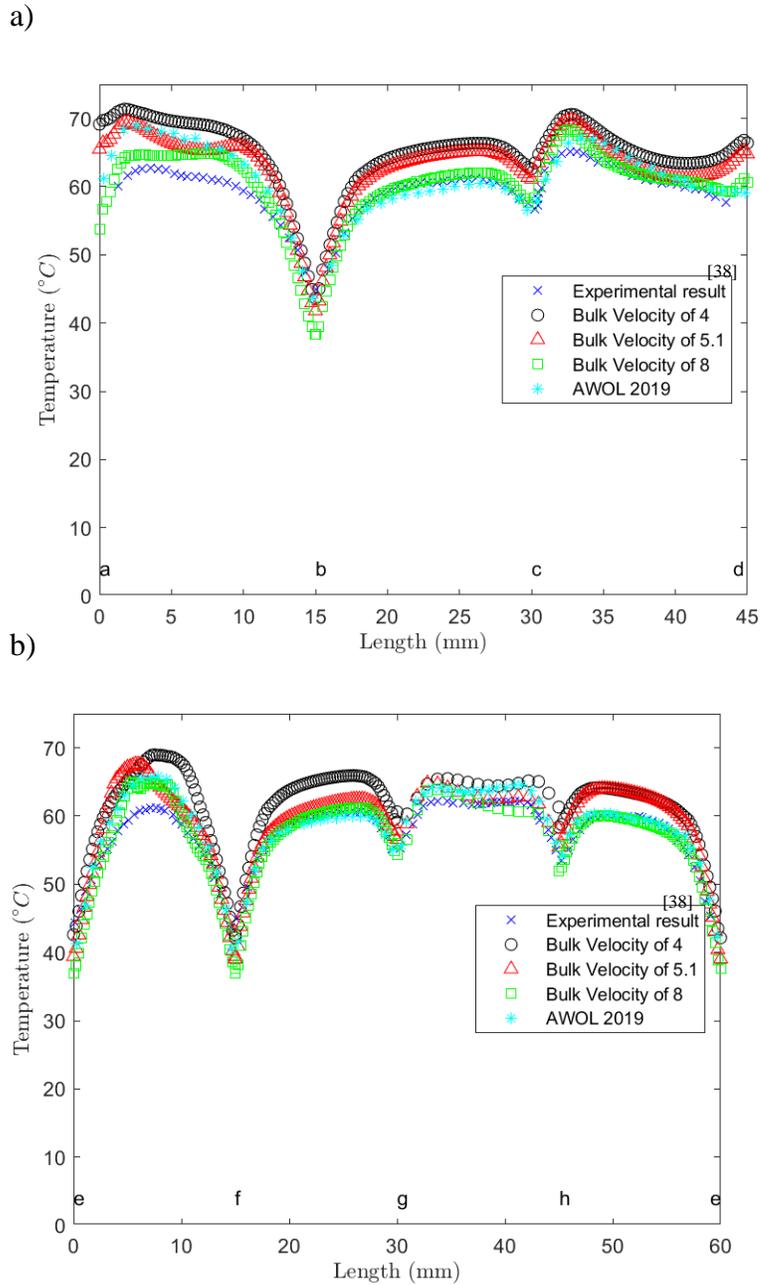
Figure 2-11: Comparison of different turbulence intensity results using shell properties along the paths: a) ABCD b) EFGHE

The mean relative error of all three simulations compared with experimental data has been provided in Table 2-4:

Table 2-4: Mean relative error of different bulk velocity

	ERROR ab-bc-cd	ERROR ef-fg-gh-he
V4	8.82	7.73
V5.1	6.11	4.50
V8	2.68	2.27
Awol	3.54	2.92

Comparing the error of Awol's work to the error of these simulations indicates that the fitted velocity profile may be inaccurate. Using a bulk velocity of 5.1 as in the experiment and a bulk velocity of 8 both result in an acceptable error range. It should be noted that assuming a bulk velocity of 8 m/s would result in greater agreement with the experiments than Awol's work exhibits.

2.1.4.4 Effect of different turbulence intensity modelling cube's interior

As minor differences were observed in meshing cubes and using shell properties, the effect of turbulence intensity using meshed cubes has been investigated. The turbulence intensity is set to 2%, 5%, and 8%, and Figure 2-12 shows the result.

a)

b)
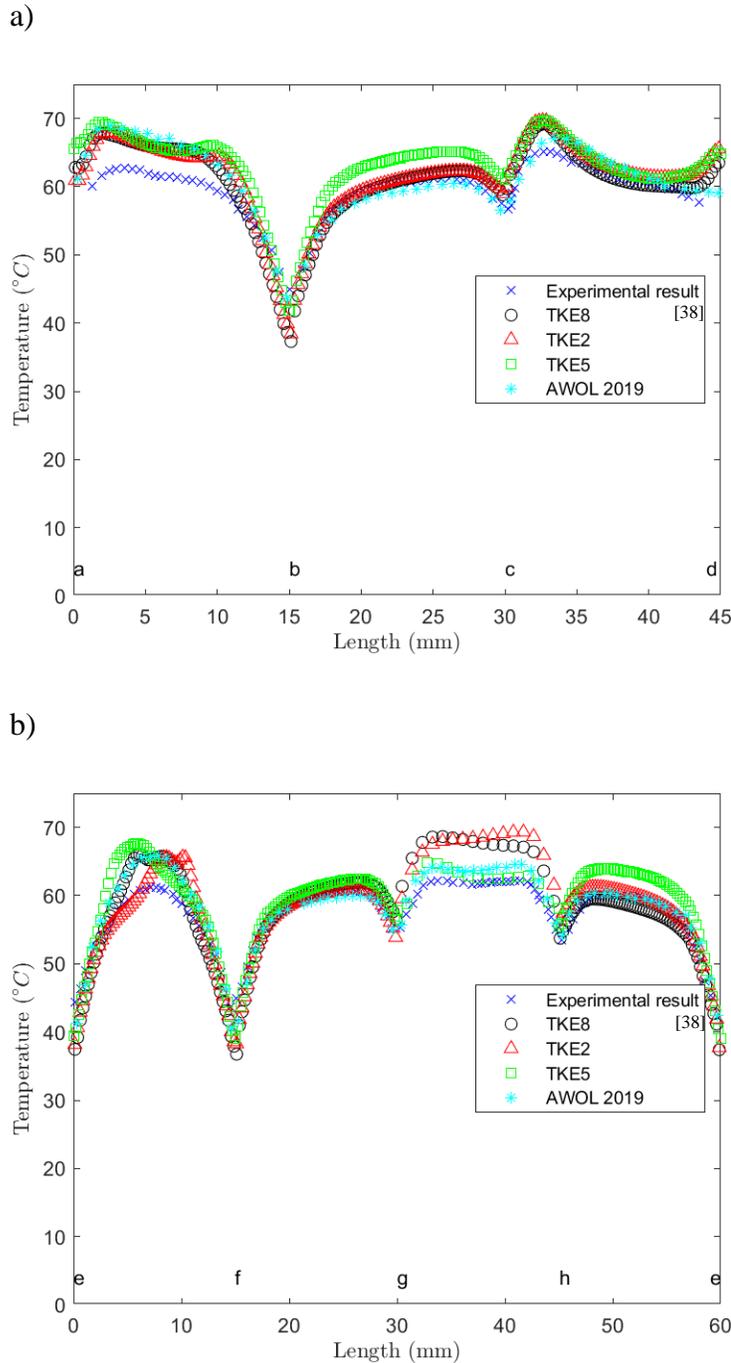
Figure 2-12: Comparison of different turbulence intensity results using actual cubes along the paths: a) ABCD b) EFGHE

In contrast to the model with shell properties, lower turbulence intensity results in a lower relative error when the epoxy cubes are modelled and meshed (calculated and shown in Table 2-5). Therefore, the turbulence intensity of 2 percent has the lowest relative error

modelling epoxy layer of cubes, which is slightly higher than the relative error of the Awol simulation.

Table 2-5: Mean relative error of different turbulence intensity results modelling epoxy layer of cubes

	ERROR ab-bc-cd	ERROR ef-fg-gh-he
TKE2	3.99	4.51
TKE5	6.11	4.49
TKE8	3.53	8.87
Awol	3.54	2.92

2.1.4.5   Conclusion

For validation, comprehensive measured data is compared to simulation findings surrounding the fifth cube, which is located in the center of the cube array. The validation findings for the temperature distributions in the vertical and horizontal planes at the fifth cube's surfaces have been discussed.

In Liu's simulations, shown in Figure 2-13, the SST k-ω model has around a 10% error. Therefore, Liu implemented a more computationally expensive turbulent model instead, which has a 3-4% error. However, in this work, the average difference between experimental data and findings produced with the SST k– ω turbulence model is roughly 6%, so that SST k-ω turbulence model, as well as the discretization schemes that were used, are acceptable. The comparison of temperature distribution of different works is summarized in Figure 2-13.

Varying the bulk velocity, turbulence intensity, and how the epoxy layers of the cubes are modelled shows that the error can be further reduced within reasonable parameters. Furthermore, it should be noted that some differences between the simulated and experimental results are attributable to the limited resolution of the infrared camera used to capture temperatures at the cube borders and the conductive heat loss from the heated

cube to the base plate in the experiment. Therefore, the results show that the numerical methods used in this work are valid.

Moreover, conduction heat transfer within a solid is not employed in the remainder of this project, and the simulation results will be used primarily to study the overall trend of the effect of cube height on the heat transfer and pressure drop, so the values themselves are not used precisely.

a)

b)
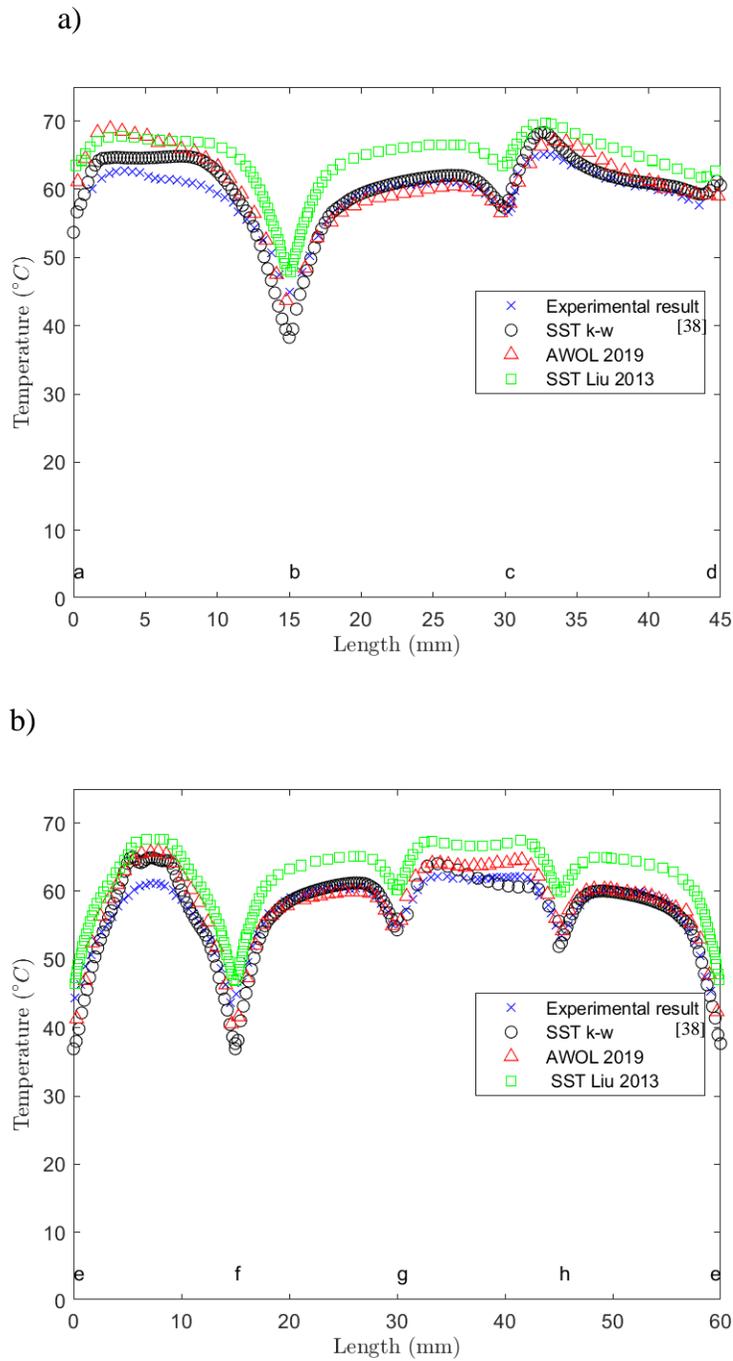
Figure 2-13: Temperature distributions on the surface around the fifth cube along the path: (a) ABCD in the vertical plane and (b) the path EFGH in the horizontal plane (b).

2.2 Results and analysis of statistical trends

The results of 204 simulations have been collected and categorized according to their block height statistics. For instance, simulations with the same standard deviation of heights have been grouped together to analyze the effect of skewness and height of the first block. The results of each group are analyzed to determine the trend in heat flux, pressure drop, and surface heat transfer with the height statistics that have been defined earlier.

2.2.1 Total Pressure Drop

In this section, the influence of various statistical values on the total pressure drop across the row of blocks is examined. The total pressure drop is calculated by subtracting the average pressure at the inlet from the pressure at the outlet. The pressure drop is then normalized by the pressure drop value (PD0) corresponding to the simulation with the lowest pressure drop and heat flux.

2.2.1.1 Effect of the height of the first block

Due to the sheltering effect and important role of the first block in developing the boundary layer, the effect of the height of the first block is studied. The data is divided into groups with similar standard deviations and skewness to isolate the impact of the height of the first block. The data have been plotted with a linear trend line so that the slope may be determined as a correlation metric. Figure 2-14, Figure 2-15, Figure 2-16, and Figure 2-17 illustrate the correlation of normalized pressure drop with the height of the first block having standard deviations of 0.1, 0.2, 0.3, and 0.4, respectively. Note that for each subplot, the skewness is also held constant.
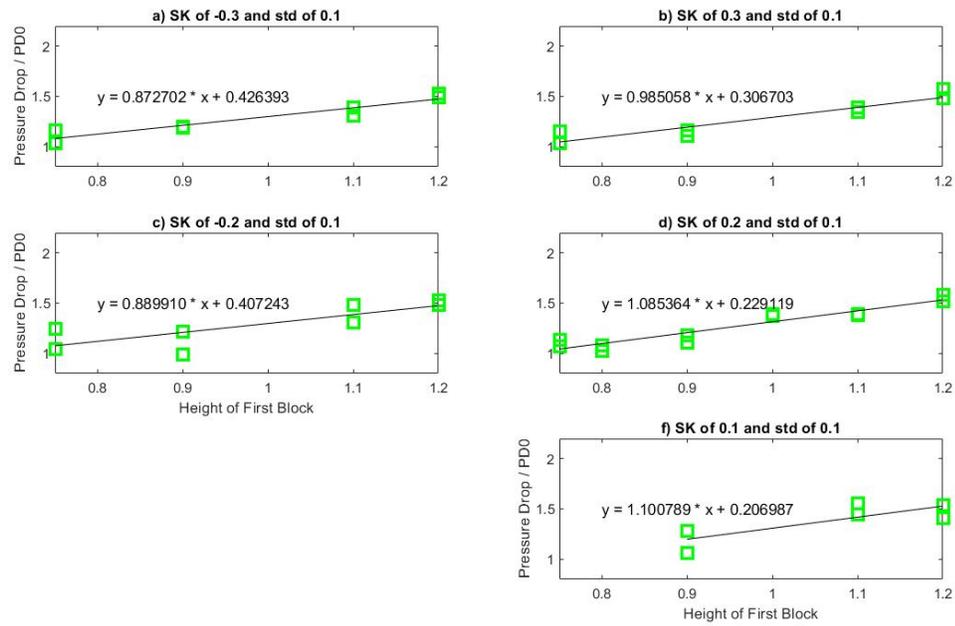
Figure 2-14: Variation of normalized pressure drop with respect to the height of the first block having a constant standard deviation of 0.1 and skewness a) -0.3 b) 0.3 c) -0.2 d) 0.2 f) 0.1
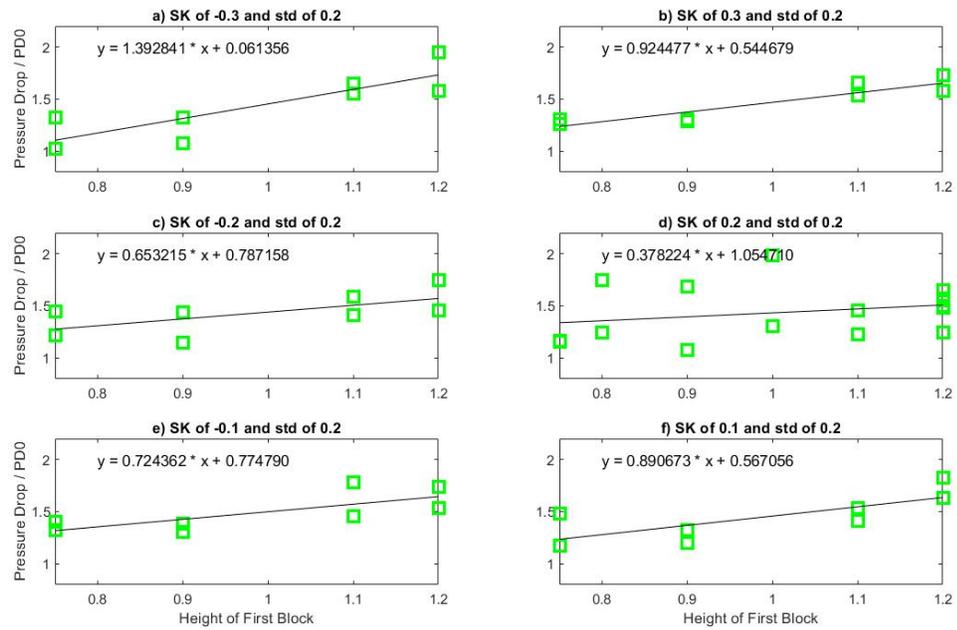
Figure 2-15: Variation of normalized pressure drop with respect to the height of the first block having a constant standard deviation of 0.2 and skewness a) -0.3 b) 0.3 c) -0.2 d) 0.2 e) -0.1 f) 0.1
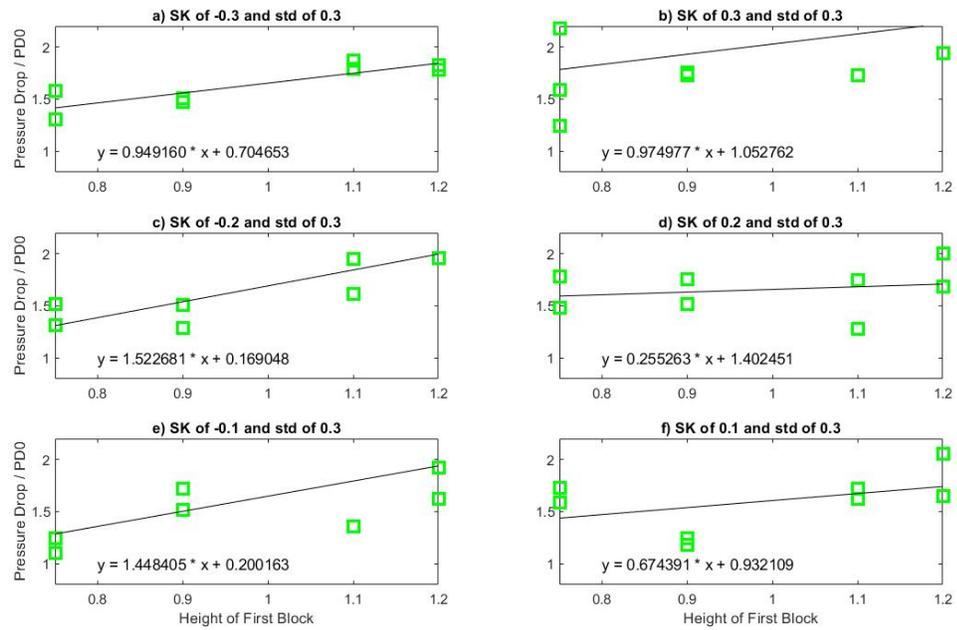
Figure 2-16: Variation of normalized pressure drop with respect to the height of the first block having a constant standard deviation of 0.3 and skewness a) -0.3 b) 0.3 c) - 0.2 d) 0.2 e) -0.1 f) 0.1
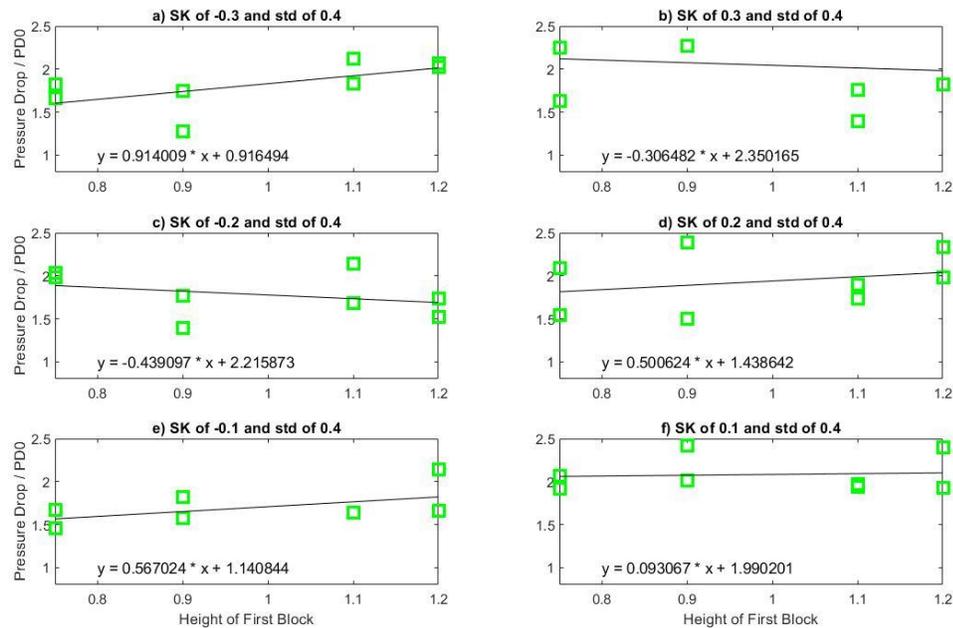
Figure 2-17: Variation of normalized pressure drop with respect to the height of the first block having a constant standard deviation of 0.4 and skewness a) -0.3 b) 0.3 c) -0.2 d) 0.2 e) -0.1 f) 0.1

The height of the first block has a significant effect on the structure of the wake; a taller first block forces a more abrupt transition in the flow when it encounters the array of blocks. The slopes of each subplot from Figure 2-14 to Figure 2-17 are shown in Table 2-6 and the p-value of the sets of slopes is 1e-6, which indicates that there is a statistically significant correlation between the height of the first block and the pressure drop. There is a 1E-4% chance that height of first block and pressure drop are uncorrelated when standard deviation and skewness are fixed. Most slopes are positive, and the few slopes that are negative have much smaller magnitudes. The negative magnitude of slopes occurs for simulations with higher standard deviation; in these cases, the larger variation of subsequent blocks becomes more significant and may overshadow the effect of the height of the first block.

Table 2-6: Slope of pressure drop changes with respect to the height of the first block

STD\SK	-0.3	-0.2	-0.1	0.1	0.2	0.3	avg
0.1	0.873	0.890	NAN	0.985	1.085	1.101	0.987
0.2	1.393	0.653	0.724	0.924	0.378	0.891	0.827
0.3	0.949	1.523	1.448	0.975	0.255	0.674	0.971
0.4	0.914	-0.439	0.567	-0.305	0.050	0.093	0.147
avg	1.032	0.657	0.913	0.645	0.442	0.690	0.733

Figure 2-18 demonstrates that as the standard deviation increases, the effect of the first block's height decreases. Similarly, Figure 2-19 illustrates that increasing skewness also decreased the effect of the first block height's on the pressure drop. The effect of the first block's height is greater in sets with negative skewness than in sets with positive skewness; this is because, in a negatively skewed distribution with a tall first block, there is a larger height difference between the first block and the few blocks that immediately follows.

$$y = -2.434953 * x + 1.346401$$
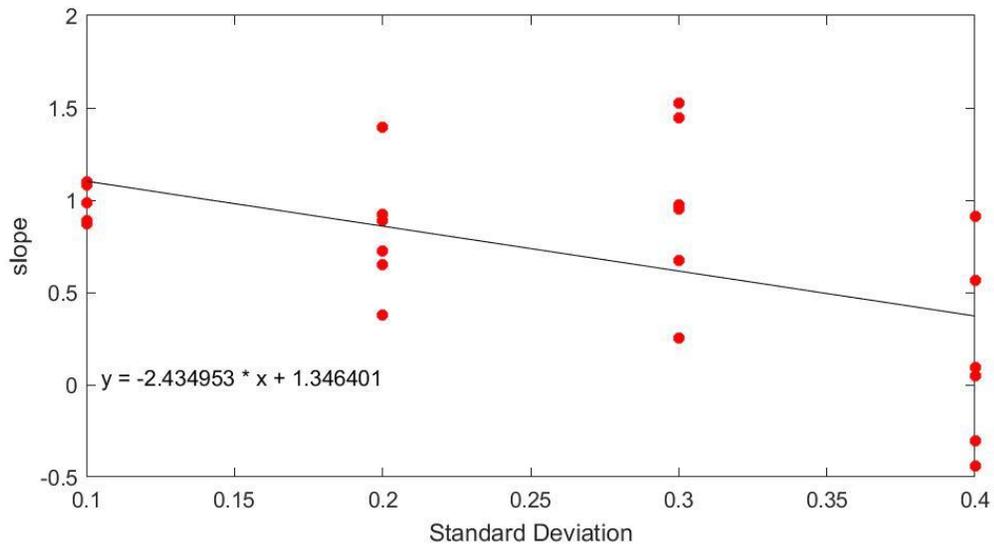
Figure 2-18: the slope of the effect of height of the first block on pressure drop with respect to standard deviation
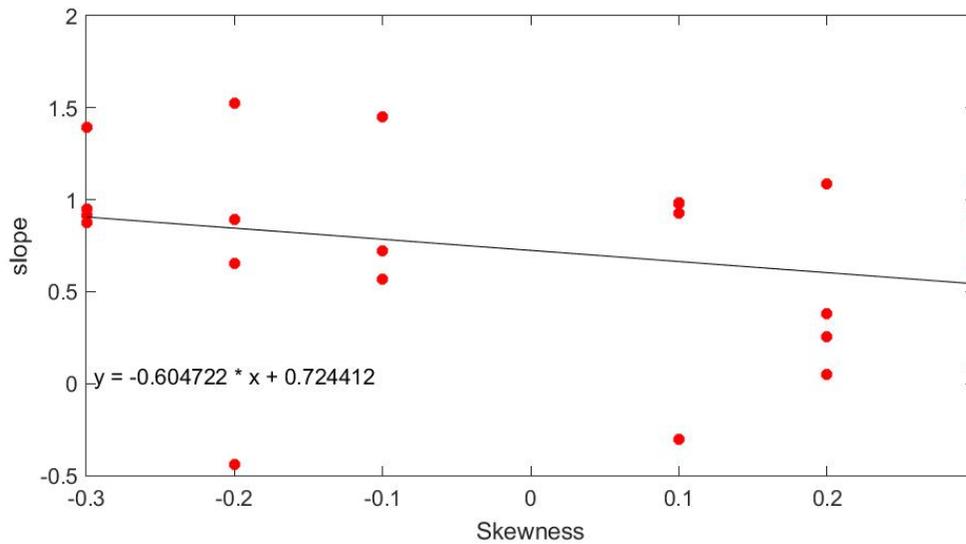
Figure 2-19: the slope of the effect of height of the first block on pressure drop with respect to skewness

2.2.1.2   Effect of standard deviation

The effect of standard deviation on the total pressure drop has been studied by holding the skewness and height of the first block constant for each subset of the simulations shown in Figure 2-20, Figure 2-21, Figure 2-22, and Figure 2-23 with a constant height of the first block of 0.9, 1.1, 1.2, and 0.75 respectively. The standard deviation is anticipated to be the most influential statistical variable on the flow. The standard deviation may operate

similarly to surface roughness, promoting turbulent mixing around the tops of the blocks, hence boosting pressure drop.
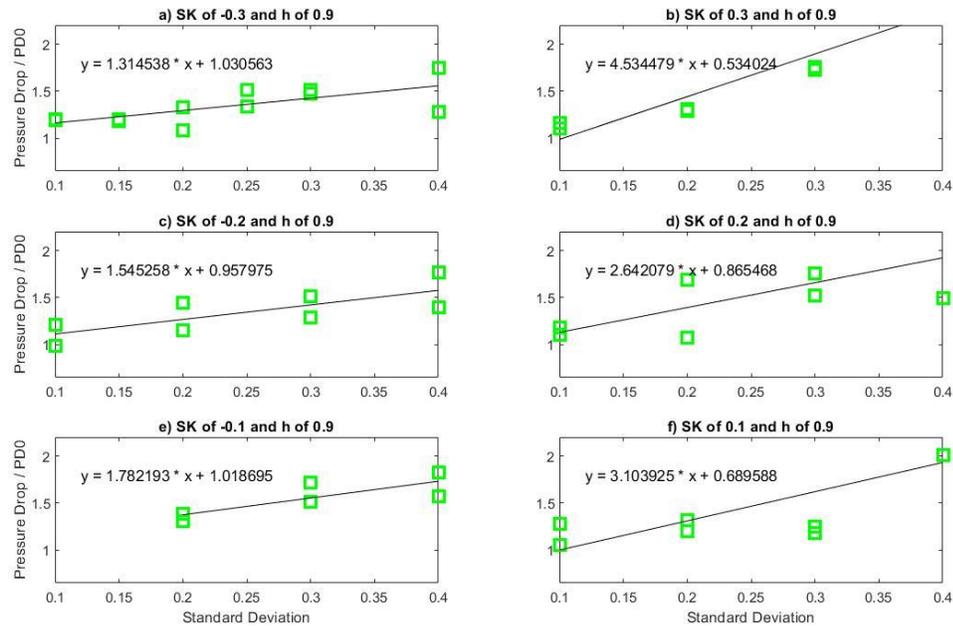
Figure 2-20: Variation of normalized pressure drop with respect to standard deviation having constant first block height of 0.9 and skewness a) -0.3 b) 0.3 c) -0.2 d) 0.2 e) -0.1 f) 0.1
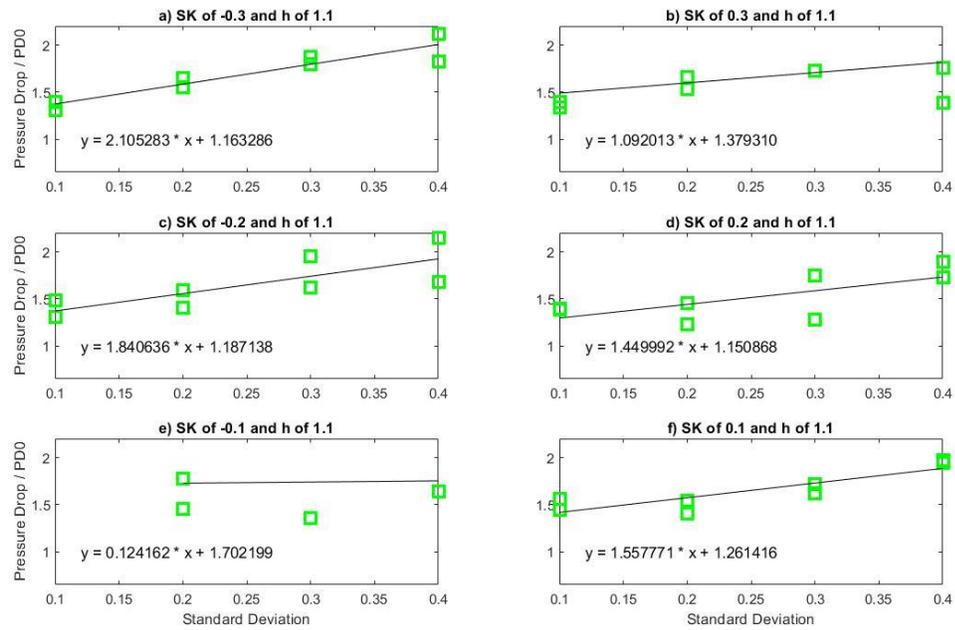
Figure 2-21: Variation of normalized pressure drop with respect to standard deviation having constant first block height of 1.1 and skewness a) -0.3 b) 0.3 c) -0.2 d) 0.2 e) -0.1 f) 0.1
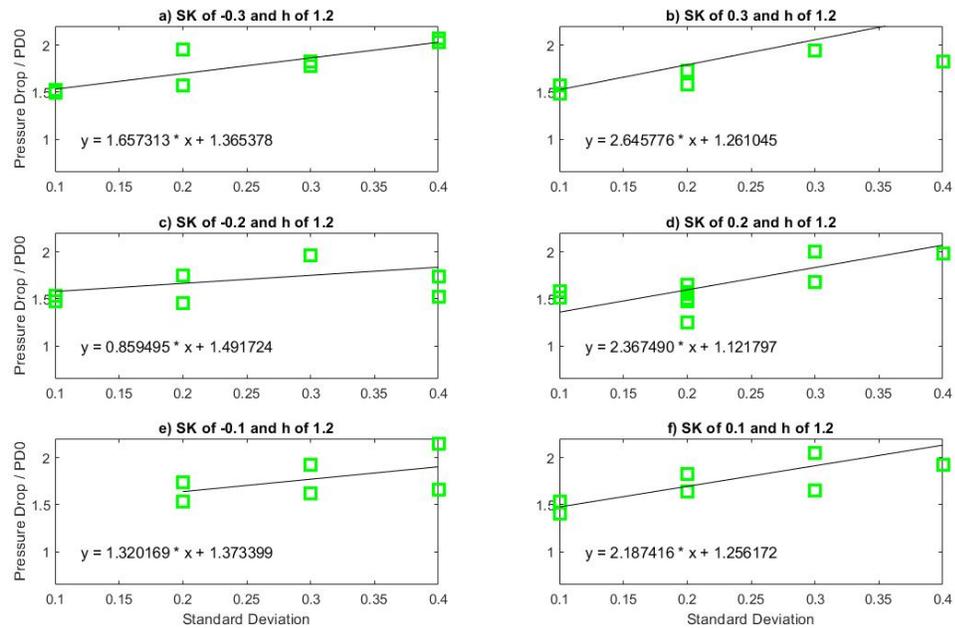
Figure 2-22: Variation of normalized pressure drop with respect to standard deviation having constant first block height of 1.2 and skewness a) -0.3 b) 0.3 c) -0.2 d) 0.2 e) -0.1 f) 0.1
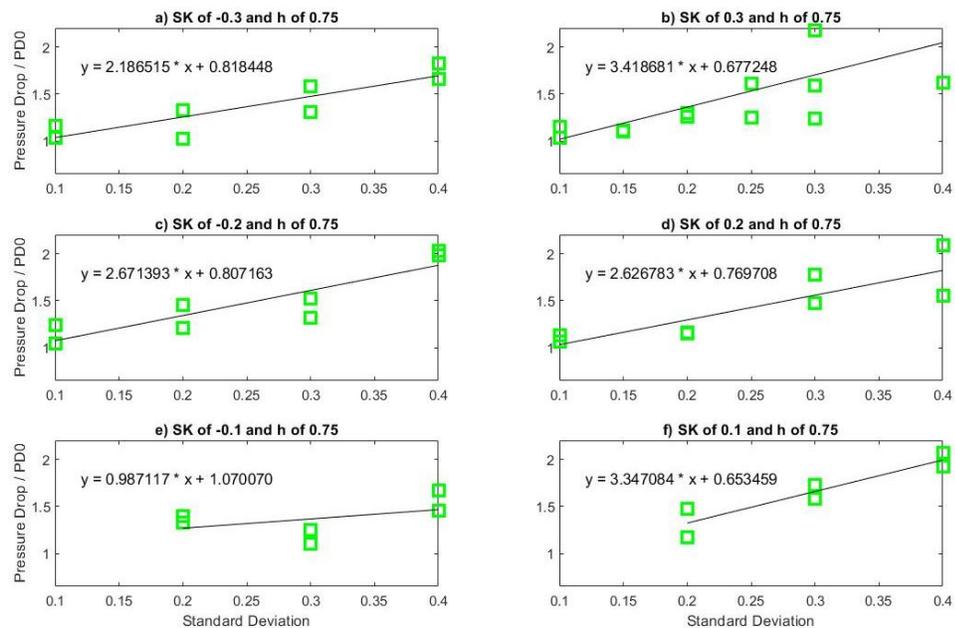
Figure 2-23: Variation of normalized pressure drop with respect to standard deviation having constant skewness and first block height of 0.75

The results indicate a direct relationship between the standard deviation and the overall pressure drop. Compared to the other statistical characteristics considered in this study, the standard deviation is the most important moment other than the average height in predicting pressure drop. The slope values of all subplots from Figure 2-20 to Figure 2-23 are shown in Table 2-7, and they are all positive and with p-value of 1E-11 suggesting that it is statistically significant, indicating there is less than a 2E-9% chance that the pressure dop is not correlated with the standard deviation. When the skewness is -0.1, and the height of the first block is 1.1, the value of the slope is the smallest, which could be due to the small amount of data in the subset.

Table 2-7: Slope of pressure drop changes with respect to standard deviation

h\SK	-0.3	-0.2	-0.1	0.1	0.2	0.3	avg
0.75	2.187	2.671	0.987	3.419	2.627	3.347	2.540
0.90	1.315	1.545	1.782	4.534	2.642	2.642	2.410
1.10	2.105	1.841	0.124	1.092	1.450	1.558	1.362
1.20	1.657	2.646	0.859	1.320	2.367	2.187	1.840
avg	1.816	2.176	0.938	2.591	2.272	2.434	2.038

Increasing the height of the first block reduces the effect of standard deviation on pressure drop, as shown in Figure 2-24. Figure 2-25 demonstrates that as skewness increases, the effect of standard deviation increases. This is because, with positive skewness, there is more likely to be a taller first block, which disrupts the flow more and results in a larger pressure drop.
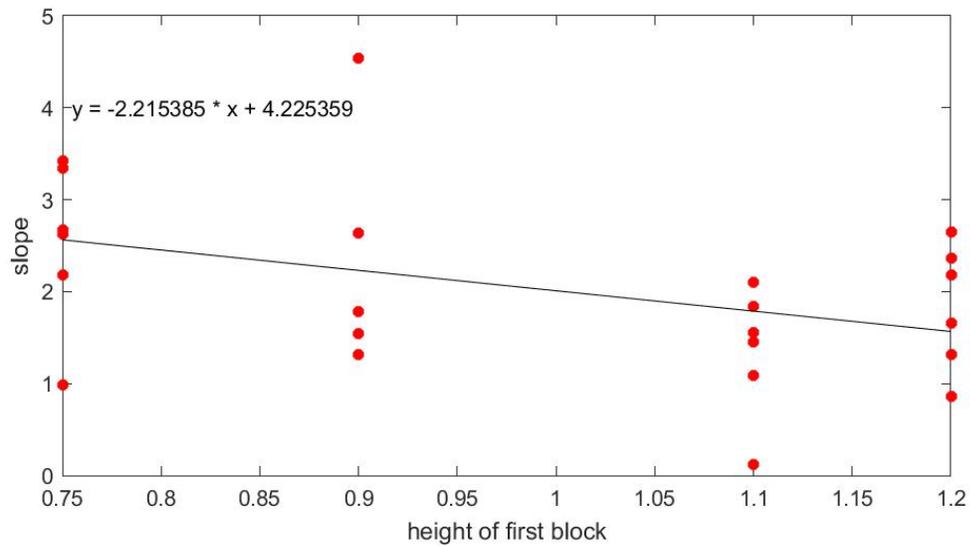
Figure 2-24: the slope of the effect of standard deviation on pressure drop with respect to the height of the first block
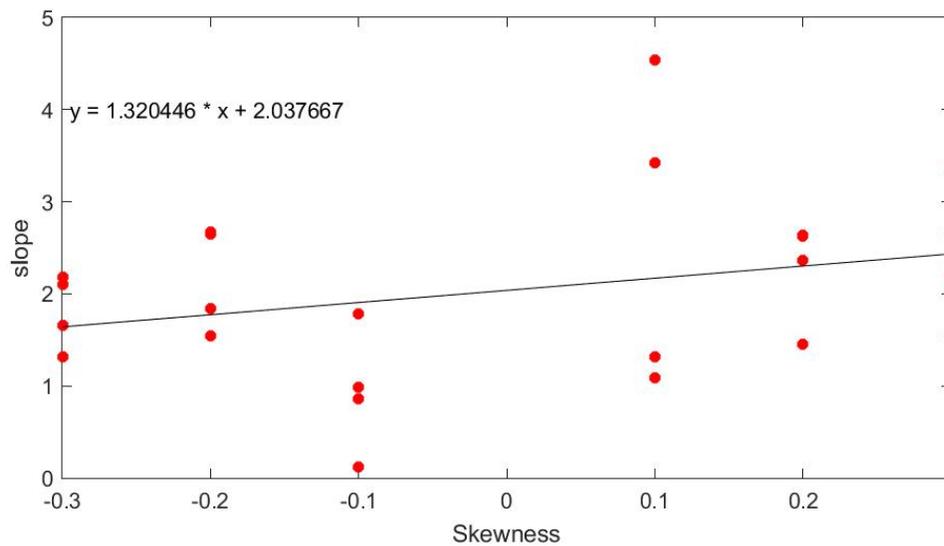
Figure 2-25: the slope of the effect of standard deviation on pressure drop with respect to skewness

2.2.1.3   Effect of skewness

This section examines the influence of skewness on pressure drop by holding the standard deviation and height of the first block constant. Figure 2-26 illustrates the relationship

between pressure drop and skewness when the starting block height is 0.75, 0.9, 1.1, and 1.2 and the standard deviation is 0.1. Figure 2-27, Figure 2-28, and Figure 2-29 depict the influence of skewness on normalized pressure drop with varying heights of the first block and constant standard deviations of 0.1, 0.2, 0.3, and 0.4, respectively. Negative skewness indicates that the taller buildings are positioned more toward the end of the array, which may not increase turbulence as much as when the taller buildings are at the beginning of the array of blocks.
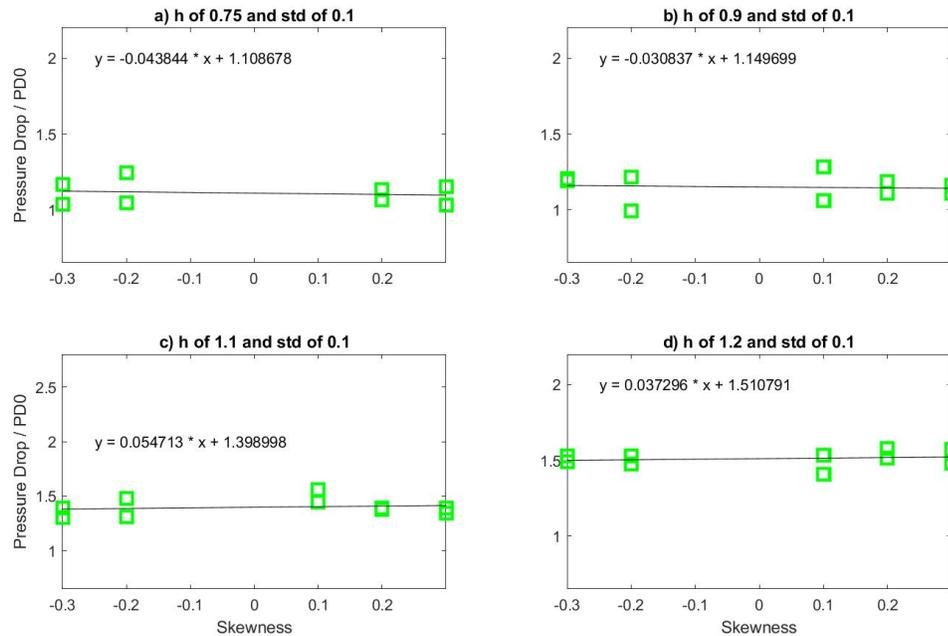
Figure 2-26: Variation of normalized pressure drop with respect to skewness having a constant standard deviation of 0.1 and first block height of a) 0.75 b) 0.9 c) 1.1 d) 1.2
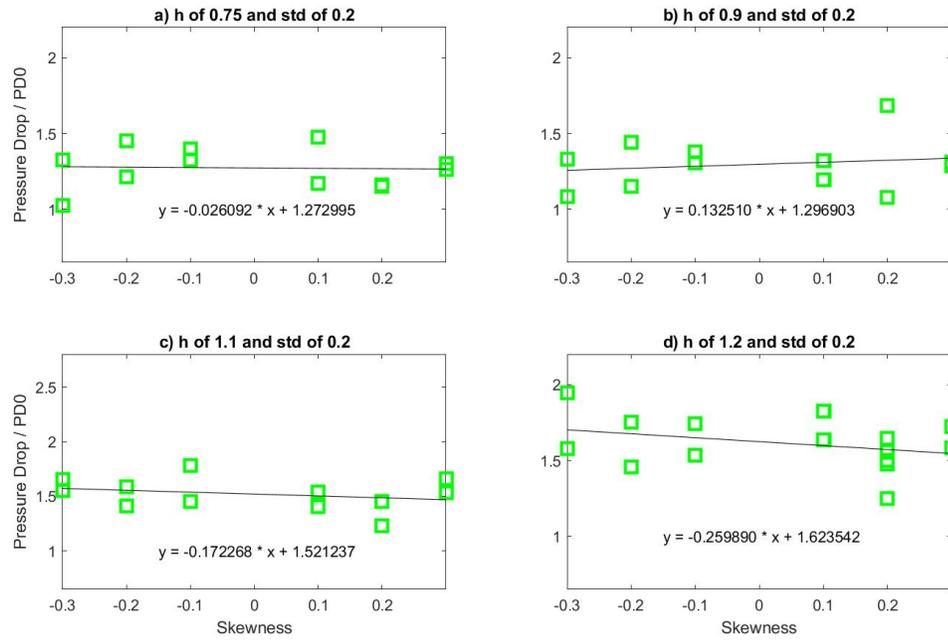
Figure 2-27: Variation of pressure drop with respect to skewness having a constant standard deviation of 0.2 and first block height of a) 0.75 b) 0.9 c) 1.1 d) 1.2
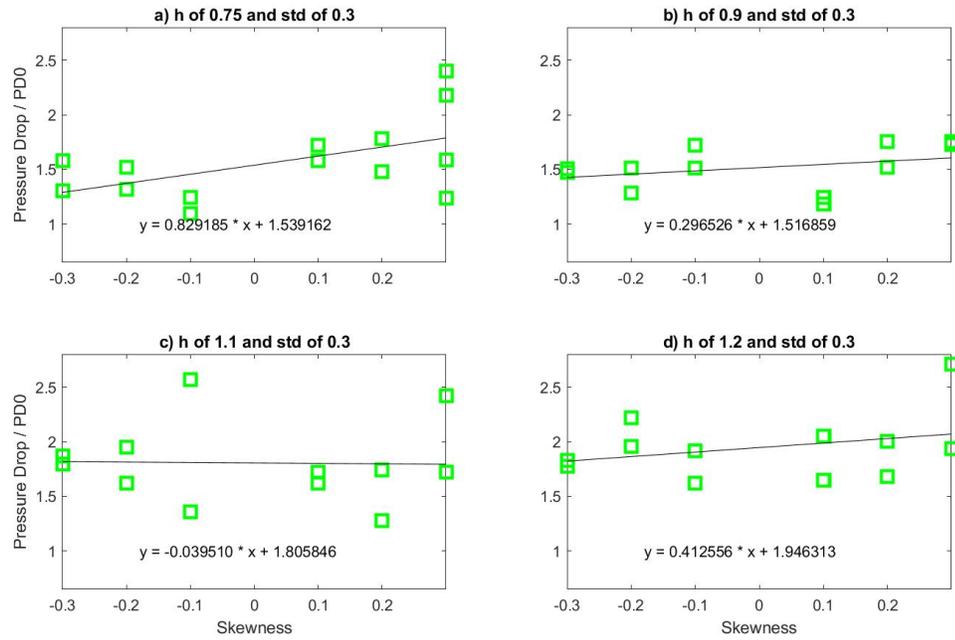
Figure 2-28: Variation of normalized pressure drop with respect to skewness having a constant standard deviation of 0.3 and first block height of a) 0.75 b) 0.9 c) 1.1 d) 1.2
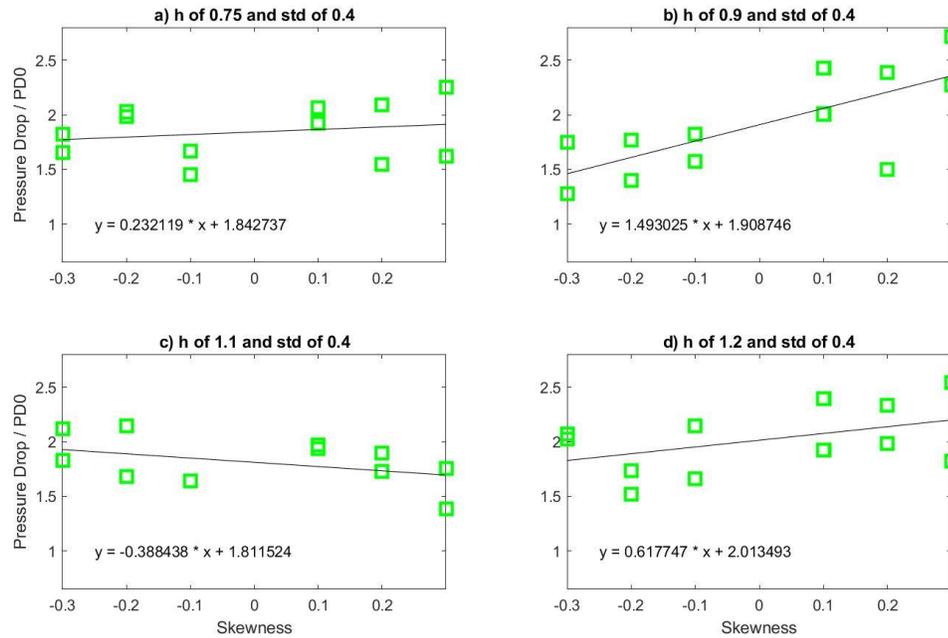
Figure 2-29: Variation of normalized pressure drop with respect to skewness having a constant standard deviation of 0.4 and first block height of a) 0.75 b) 0.9 c) 1.1 d) 1.2

Table 2-8 displays the relationship between pressure drop and skewness for each combination of standard deviation and first block height. Because the slope might be positive or negative, with p-value of 11%, which indicates that there is a 11% chance that the trend is not statistically significant. However, the slopes themselves or the average of slopes are often positive, suggesting that positive skewness or taller blocks at the leading end of the array typically result in greater pressure drop rates.

Table 2-8: Slope of pressure drop change with respect to skewness

STD\h	0.75	0.90	1.10	1.20	avg
0.1	-0.044	-0.031	0.055	0.037	0.004
0.2	-0.026	0.133	-0.172	-0.260	-0.081
0.3	0.829	0.297	-0.040	0.413	0.375
0.4	0.232	1.493	-0.388	0.618	0.489
avg	0.248	0.473	-0.136	0.202	0.197

In Figure 2-30, a cubic polynomial has been fitted to all of the simulated data, showing how the pressure drop varies with standard deviation and skewness. The result illustrates that the standard deviation has the largest effect on the pressure drop. The effect of skewness is smaller than the effect of standard deviation, but the effect of skewness increases as the standard deviation increases.
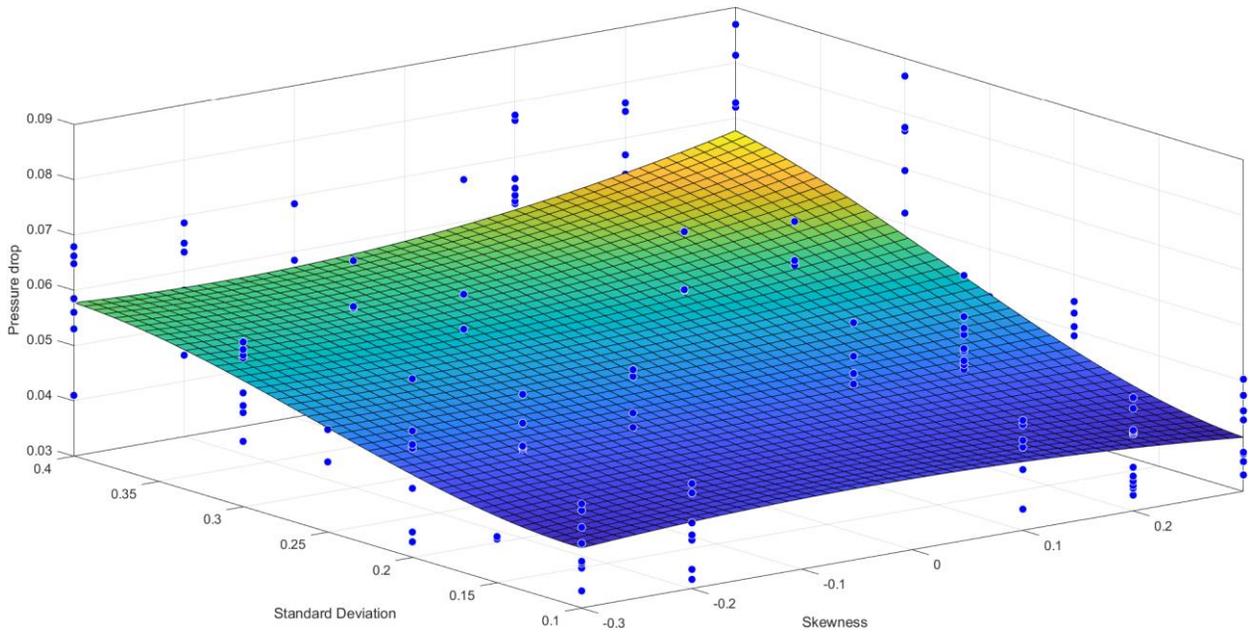
Figure 2-30: 3D plot of pressure drop with respect to skewness and standard deviation

2.2.2   Heat flux

More variation in the heights of the blocks generates more turbulence in the flow, resulting in a greater pressure drop and heat flux. The same physical effects that cause the flow to generate more turbulence will also increase the heat flux and pressure drop. In this section, we will examine whether the effects of statistics on heat flux are similar to those on pressure drop.

Flux is the transfer of some variable per unit area per unit of time. Typically, for kinematic heat flux, the heat value is divided by the density multiplied by the specific heat to get units of $\frac{\frac{J}{m^2s}}{(\frac{Kg}{m^3})(\frac{J}{Kg.K})} = \frac{Km}{s}$. This is just a temperature multiplied by a velocity. In this way, the calculated heat flux is independent of density and specific heat coefficient. To determine the effect on the total heat flux from the array of blocks, the heat flux at the outlet of the domain is subtracted from the inlet heat flux. This gives the net heat flux from the blocks to the fluid. Then, the value of heat flux is normalized by the outcome of the same simulation used to normalize the pressure drop. In the next subsections, the effect of first block height, standard deviation, and skewness on the total heat flux is studied.

2.2.2.1   Effect of the height of the first Block

Due to the fact that the height of the first block plays a significant role in the formation of the boundary layer and can shield the following blocks if it is very tall, the effect of the height of the first block is also being investigated. Note that early results suggested that the height of the initial block could be an important parameter.

In this part, the influence of heat flux in relation to the first block's height is examined. The data is separated into sets with identical standard deviations and skewness to examine the influence of first height alone. The data have been displayed with a linear trend line to determine the slope as a correlation metric of heat flux change and height of the first block. Figure 2-31 shows the effect of the height of the first block on the heat flux, having a constant standard deviation of 0.1 and different skewness. Following the same trend, Figure

2-32, Figure 2-33, and Figure 2-34 show the effect of the height of the first block on the heat flux with constant standard deviation values of 0.2, 0.3, and 0.4, respectively.
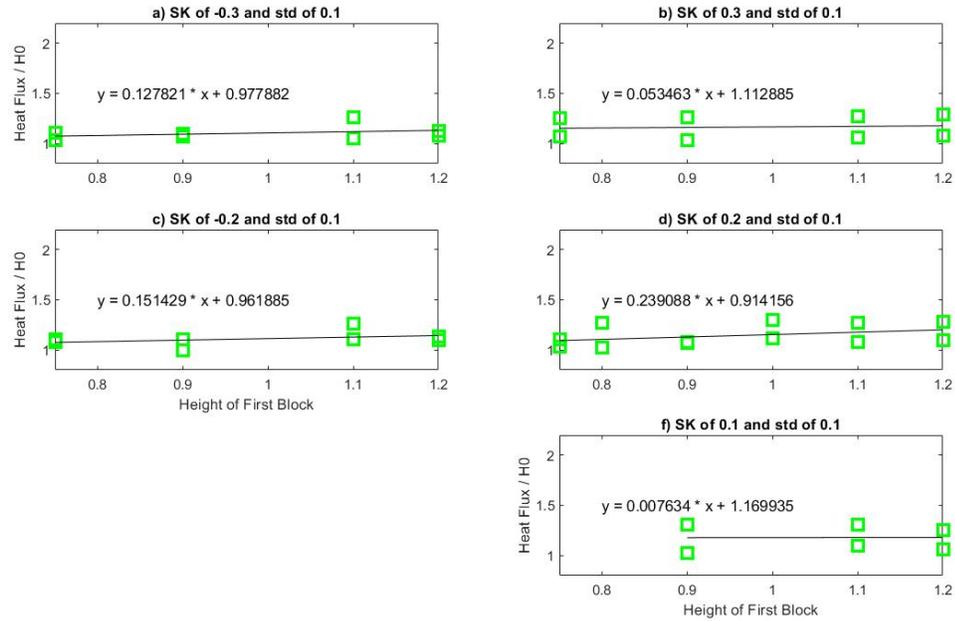
Figure 2-31: Variation of normalized heat flux with respect to the height of the first block having a constant standard deviation of 0.1 and skewness of a) -0.3 b) 0.3 c) -0.2 d) 0.2 f) 0.1
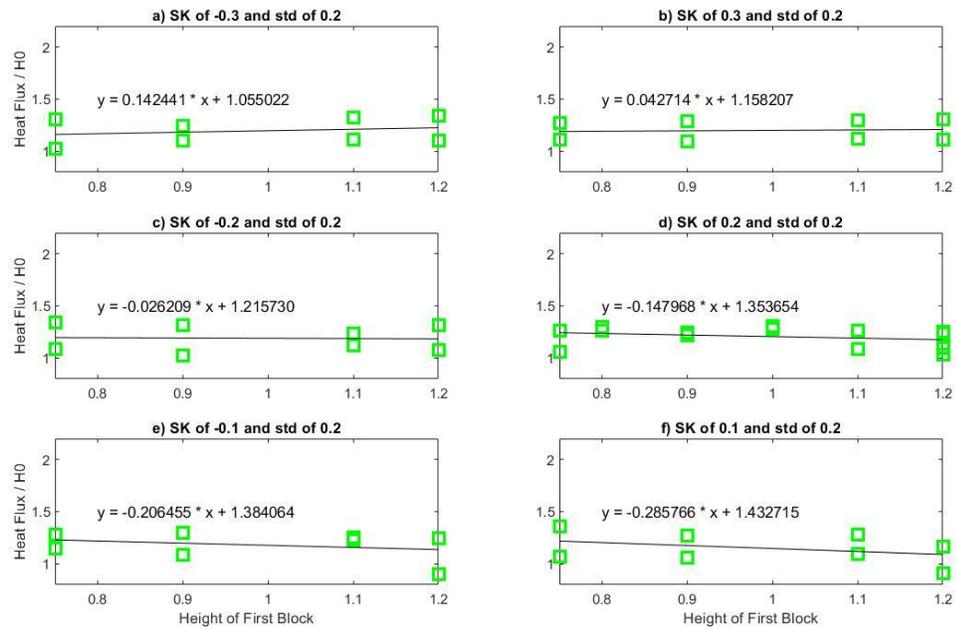
Figure 2-32: Variation of normalized heat flux with respect to the height of the first block having a constant standard deviation of 0.2 and skewness of a) -0.3 b) 0.3 c) -0.2 d) 0.2 e) -0.1 f) 0.1
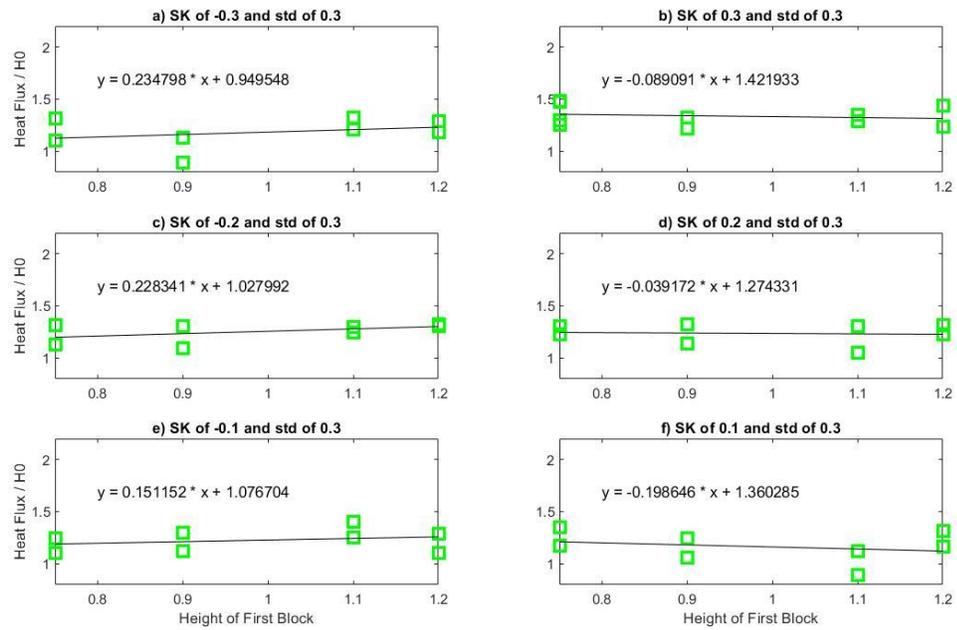
Figure 2-33: Variation of normalized heat flux with respect to the height of the first block having a constant standard deviation of 0.3 and skewness of a) -0.3 b) 0.3 c) -0.2 d) 0.2 e) -0.1 f) 0.1
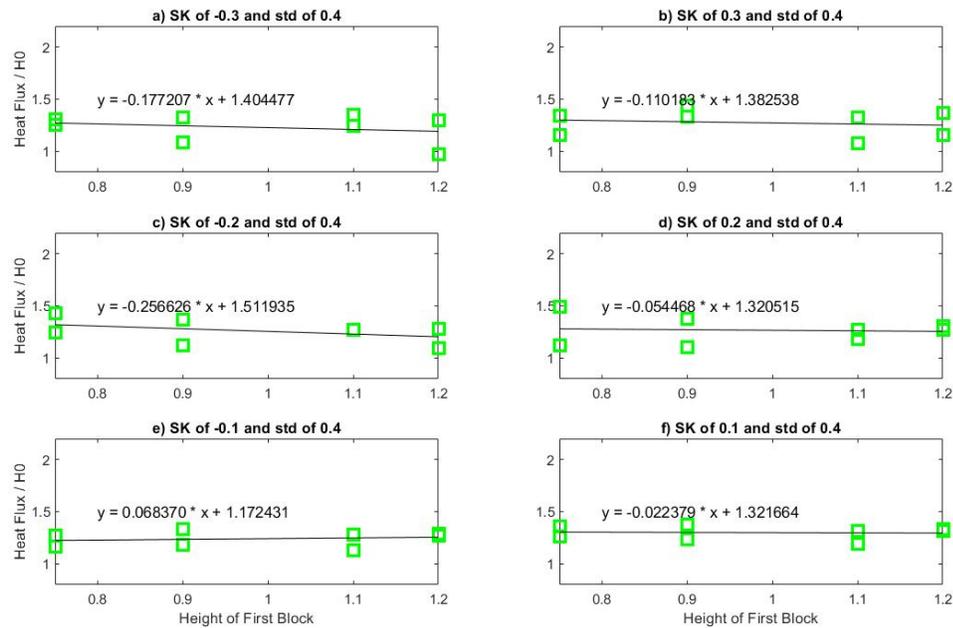
Figure 2-34: Variation of normalized heat flux with respect to the height of the first block having a constant standard deviation of 0.4 and skewness of a) -0.3 b) 0.3 c) -0.2 d) 0.2 e) -0.1 f) 0.1

The line's slope is utilized as the correlation coefficient. Table 2-9 displays the magnitude of each slope. Since generating the height with a small standard deviation and skewness is computationally costly, we did not generate simulations for skewness of -0.1 and a standard deviation of 0.1 in this project. Consequently, the plots and tables for this combination of statistics are empty or NAN.

Table 2-9: Slope of heat flux changes with respect to the height of the first block

STD\SK	-0.3	-0.2	-0.1	0.1	0.2	0.3	avg
0.1	0.128	0.151	NAN	0.053	0.239	0.008	0.116
0.2	0.142	-0.026	-0.206	0.043	-0.148	-0.286	-0.080
0.3	0.023	0.228	0.151	-0.089	-0.039	-0.199	0.013
0.4	-0.177	-0.257	0.068	-0.110	-0.054	-0.022	-0.092
avg	0.029	0.024	0.004	-0.026	-0.001	-0.125	-0.016

As evidenced by the fact that certain slopes are positive and others are negative, the height of the first block and the heat flux are not strongly correlated, while results show a direct correlation between the height of the first block and pressure drop. The slopes value with respect to standard deviation and skewness is plotted in Figure 2-35 and Figure 2-36. The results show that having negative skewness, meaning the block height at the end of the street is larger, the height of the first block has a more robust direct relation with heat flux and having positive skewness height of the first block has an inverse relation with heat flux.
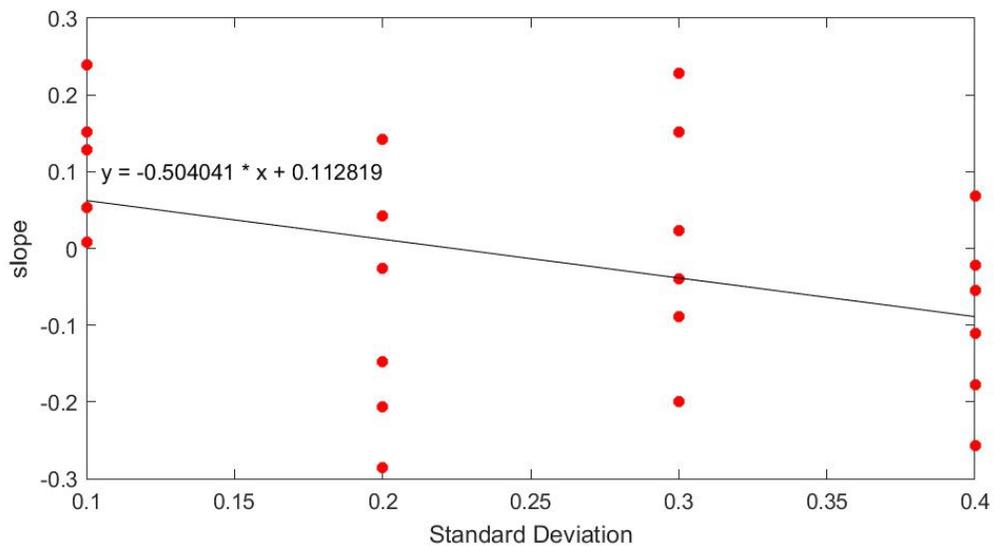
Figure 2-35: the slope of the effect of height of the first block on heat flux with respect to standard deviation
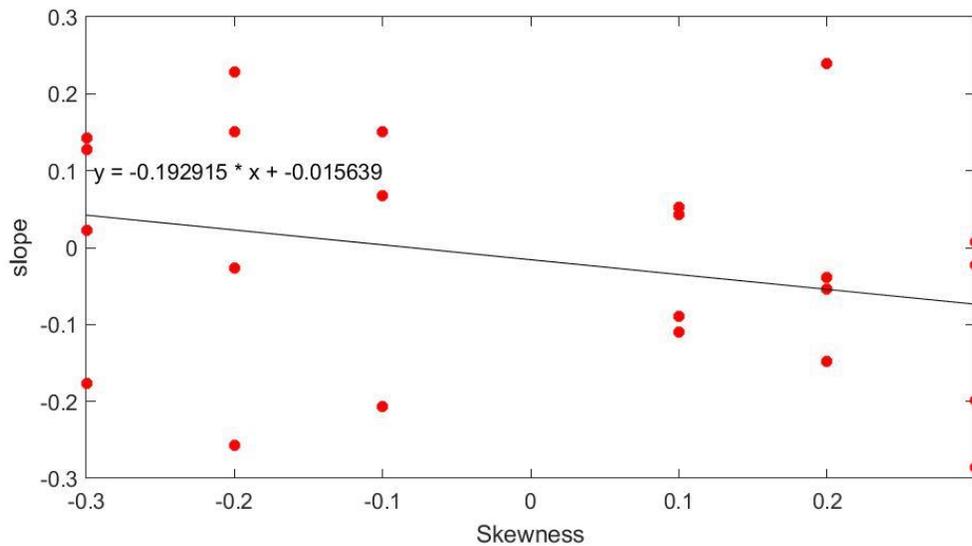
Figure 2-36: the slope of the effect of height of the first block on heat flux with respect to skewness

2.2.2.2   Effect of standard deviation

Standard deviation is expected to be the most important statistical parameter affecting the flow because it is the lowest moment other than the mean heights of blocks, which is constant for all simulations, and also it was found to be the most important parameter affecting the pressure drop. The standard deviation may behave similarly to surface roughness, enhancing turbulent mixing around the tops of the blocks and increasing the heat transfer and pressure drag.

The effect of standard deviation on heat flux is studied while holding the first block's skewness and height constant. The value of the standard deviation varied from 0.1, 0.2, 0.3, and 0.4. The results are shown in Figure 2-37, which shows how heat flux varies with standard deviation for various skewness and first block height 0.9. Figure 2-38, Figure

2-39, and Figure 2-40 depict the variation of heat flux with standard deviation for heights of 1.1, 1.2, and 0.75 for the first block, respectively.
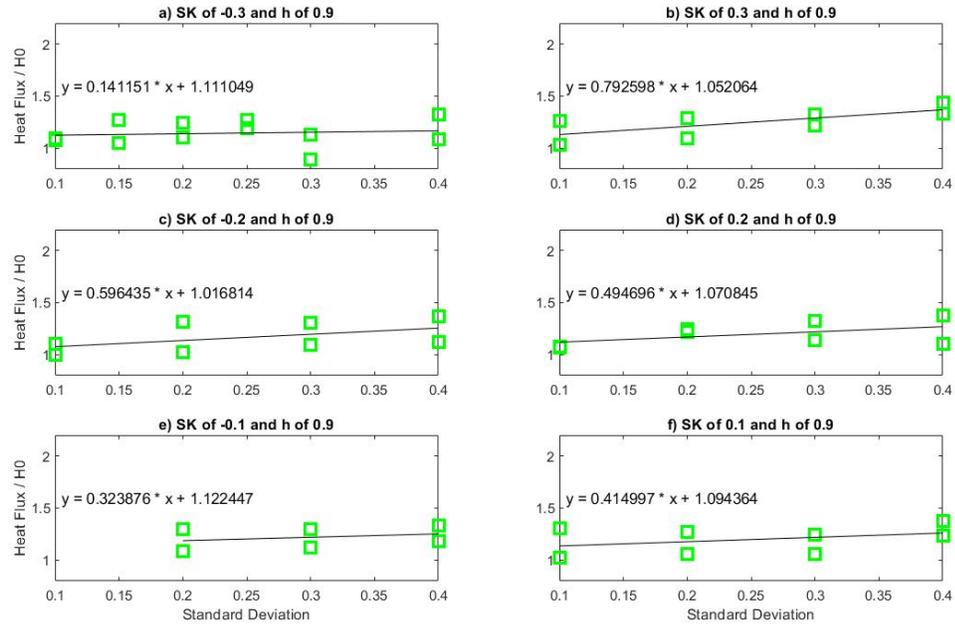
Figure 2-37: Variation of normalized heat flux with respect to standard deviation having constant first block height of 0.9 and skewness of a) -0.3 b) 0.3 c) -0.2 d) 0.2 e) -0.1 f) 0.1
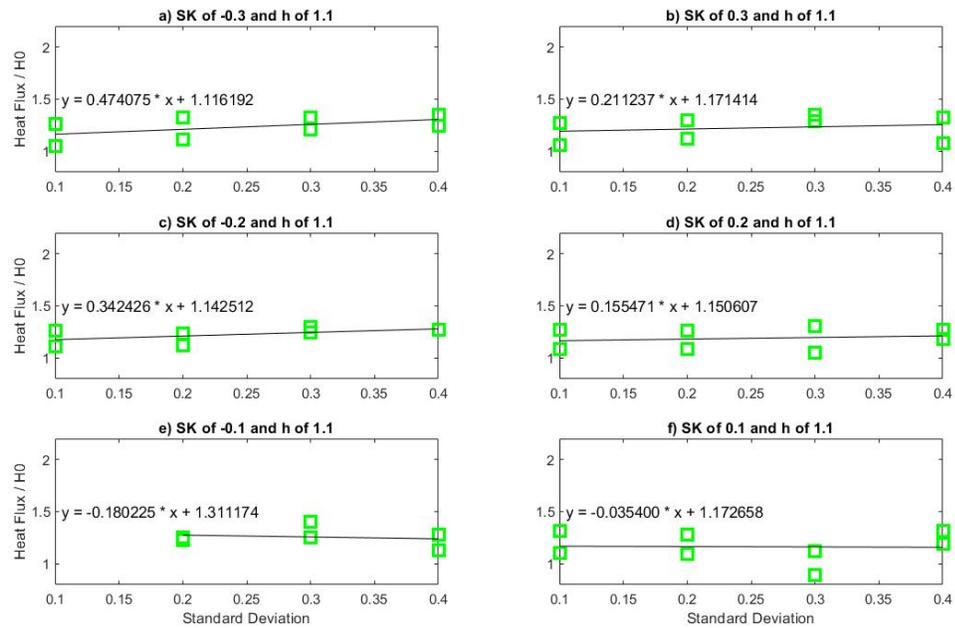
Figure 2-38: Variation of normalized heat flux with respect to standard deviation having constant first block height of 1.1 and skewness of a) -0.3 b) 0.3 c) -0.2 d) 0.2 e) -0.1 f) 0.1
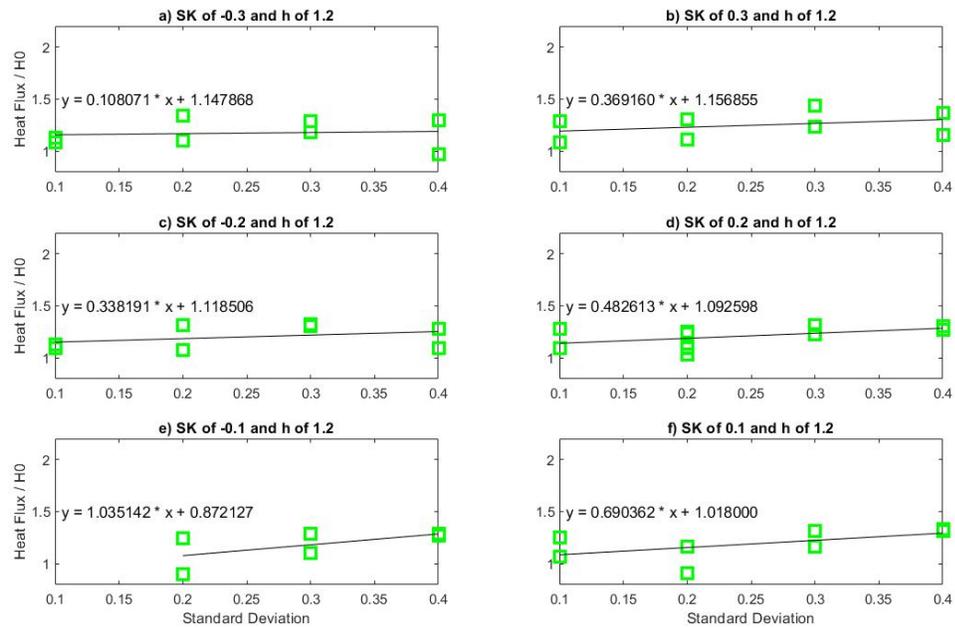
Figure 2-39: Variation of normalized heat flux with respect to standard deviation having constant first block height of 1.2 and skewness of a) -0.3 b) 0.3 c) -0.2 d) 0.2 e) -0.1 f) 0.1
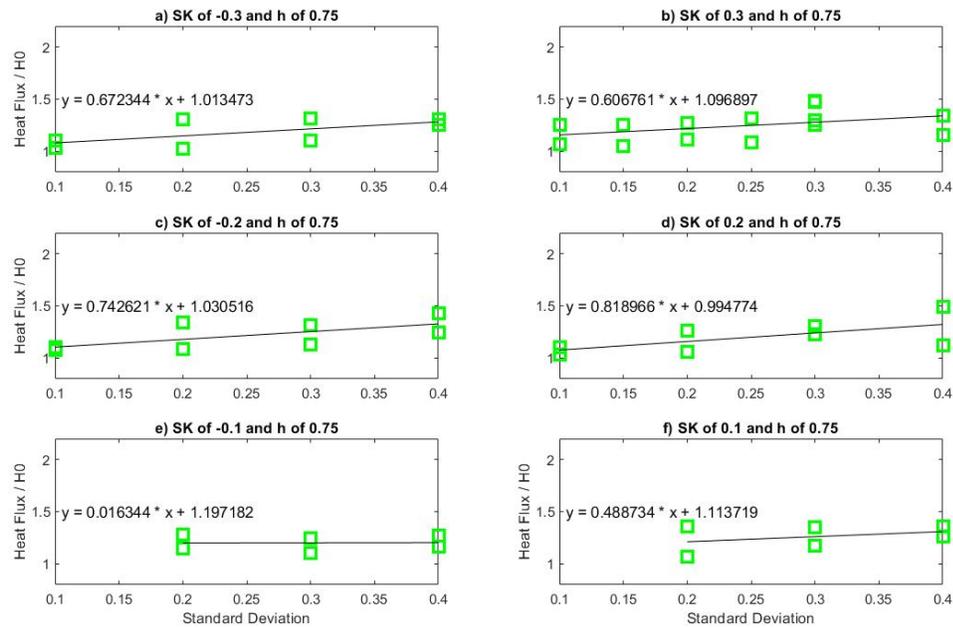
Figure 2-40: Variation of normalized heat flux with respect to standard deviation having constant skewness and first block height of 0.75

The standard deviation exhibits a strong correlation, and almost all the slopes are positive, demonstrating the significance of height standard deviation on total heat flux as it was for pressure drop. This conclusion and trend are statistically significant, with a p-value of 4.75E-7, which show the probability that the heat flux is not associated with the standard deviation is less than 4E-5%. Increasing the standard deviation, which corresponds to a larger variation in heights, increases the amount of heat flux. Two anomalies in the slope data occur for the height of the first block of 1.1 and skewness of -0.1 and 0.1. For a skewness of -0.1, the slope is calculated from a smaller number of simulations compared to the other sets. There is also a significant amount of scattering in the data because each simulation has a unique set of block heights. The two factors combined explain why a simulation set may occasionally exhibit the opposite trend from what is expected. However, most of the slopes are positive and have much larger magnitudes than the two cases with negative slopes. The slopes are shown in Table 2-10.

Table 2-10: Slope of heat flux changes with respect to standard deviation

h\SK	-0.3	-0.2	-0.1	0.1	0.2	0.3	avg
0.75	0.672	0.743	0.016	0.607	0.819	0.489	0.558
0.90	0.141	0.596	0.324	0.793	0.495	0.415	0.461
1.10	0.474	0.342	-0.180	0.211	0.155	-0.035	0.161
1.20	0.108	0.338	1.035	0.369	0.483	0.690	0.504
avg	0.349	0.505	0.299	0.495	0.488	0.390	0.421

Figure 2-41 shows the effect of the height of the first block on the slopes in Table 2-10. It can be concluded that increasing the height of the first block reduces the impact of the standard deviation on heat flux. Although Figure 2-42 displays a negative correlation between skewness and slopes, it cannot be concluded that skewness has an effect on slopes due to the figure's data distribution.
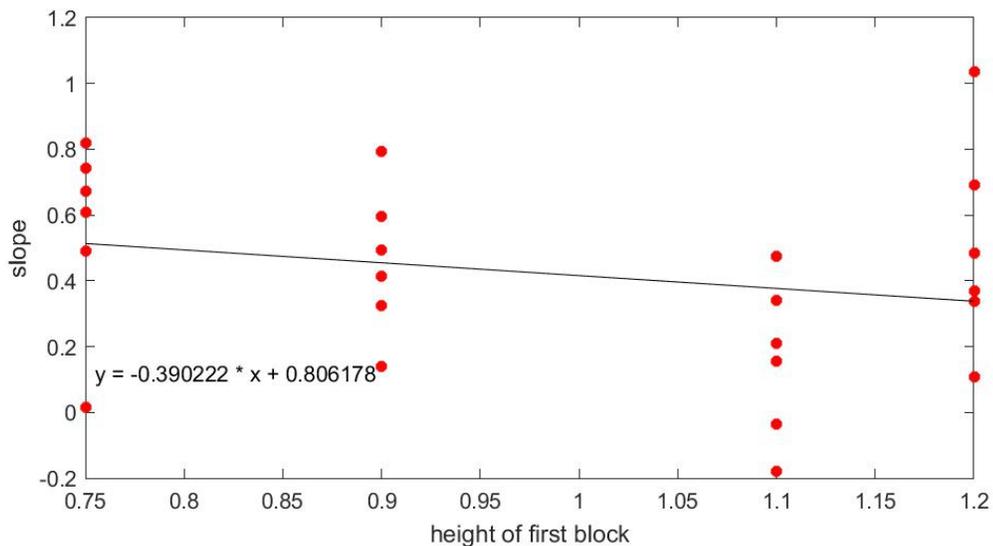
Figure 2-41: the slope of the effect of standard deviation on heat flux with respect to the height of the first block
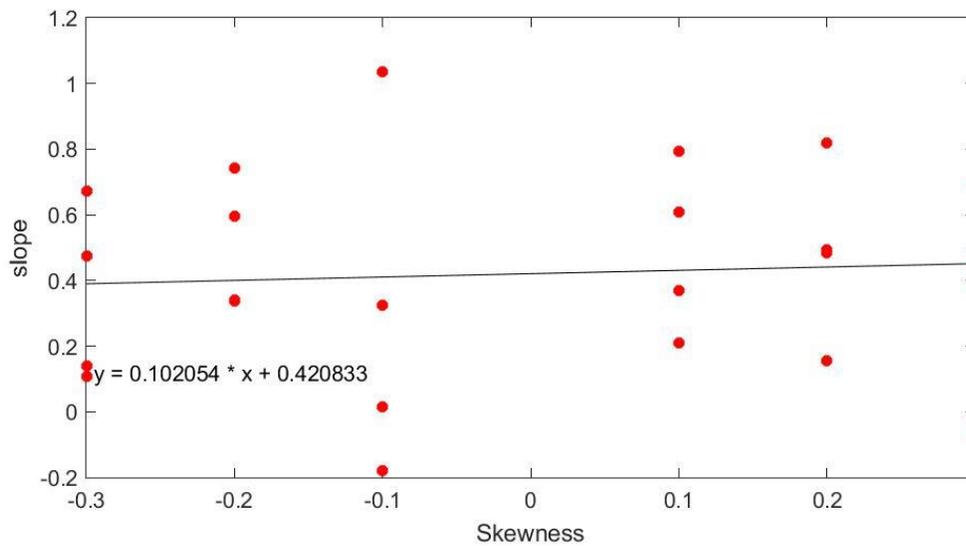
Figure 2-42: the slope of the effect of standard deviation on heat flux with respect to skewness

2.2.2.3   Effect of skewness

This section examines the influence of skewness on the total heat flux. A positive skewness may be expected to increase the heat flux because higher blocks early in the array create a more abrupt transition for the flow, which may increase the turbulence and, therefore, the heat flux. Each figure illustrates data with the same standard deviation and height of the first block. A linear line was used to illustrate the relationship between the change in heat flux and skewness. Figure 2-43 shows how heat flux varies with skewness when the initial block height is 1.1, with standard deviations of 0.1, 0.2, 0.3, and 0.4. Figure 2-44, Figure

2-45, and Figure 2-46 illustrate the effect of skewness on heat flux with different standard deviations and constant height of the first block of 1.2, 075, and 0.9, respectively.
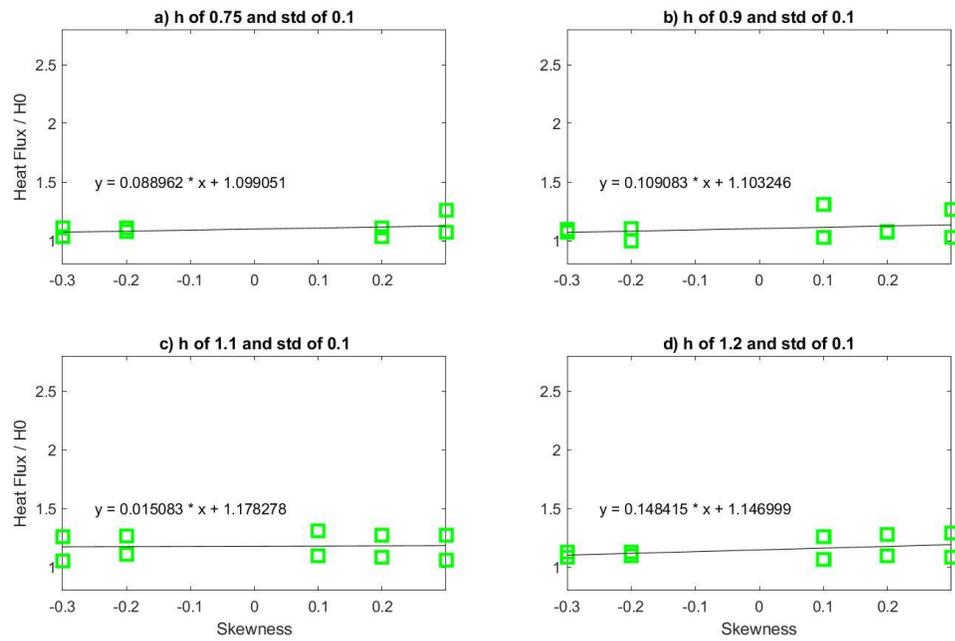
Figure 2-43: Variation of normalized heat flux with respect to skewness having a constant standard deviation of 0.1 and first block height of a) 0.75 b) 0.9 c) 1.1 d) 1.2
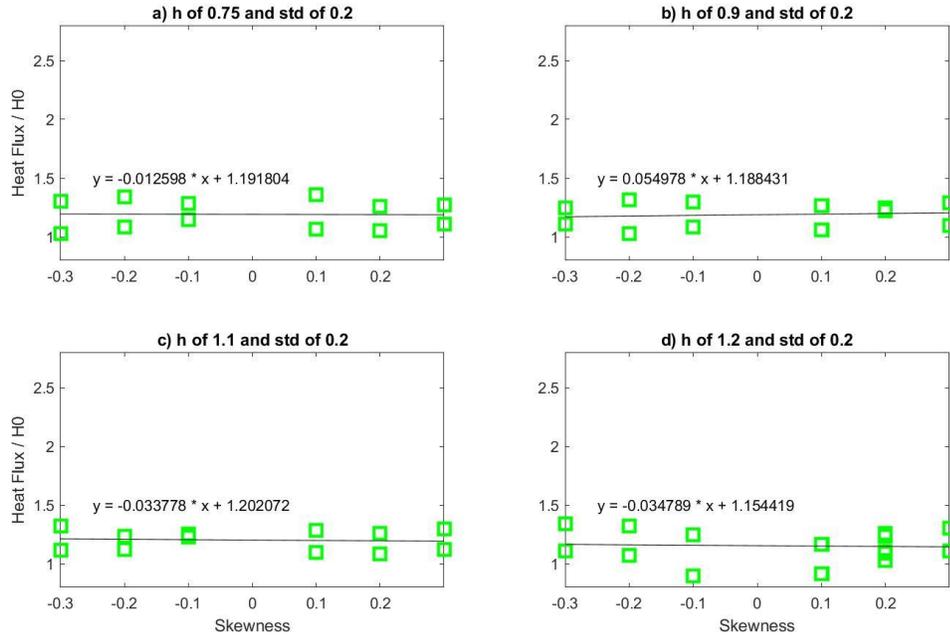
Figure 2-44: Variation of normalized heat flux with respect to skewness having a constant standard deviation of 0.2 and first block height of a) 0.75 b) 0.9 c) 1.1 d) 1.2
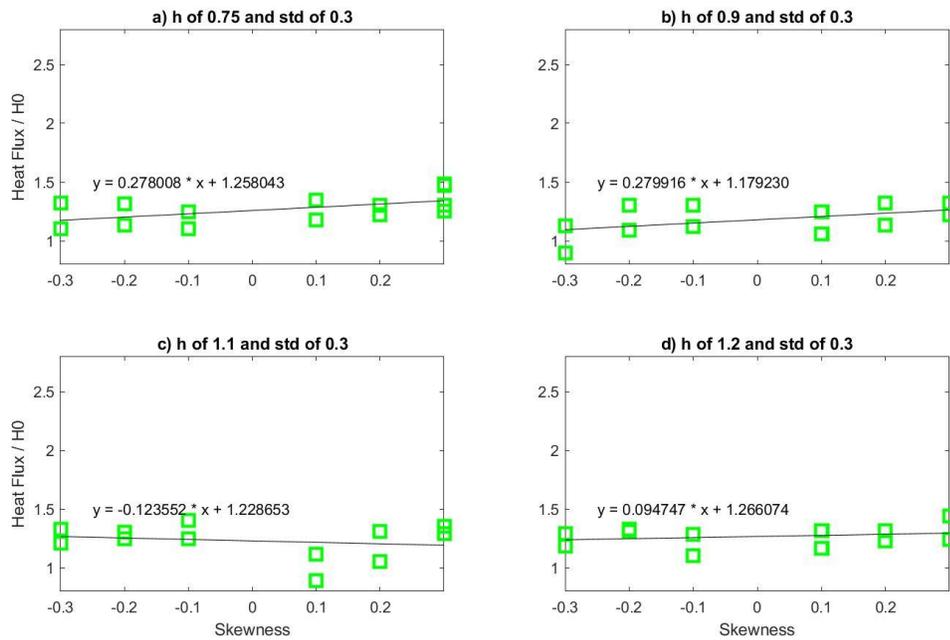
Figure 2-45: Variation of normalized heat flux with respect to skewness having a constant standard deviation of 0.3 and first block height of a) 0.75 b) 0.9 c) 1.1 d) 1.2
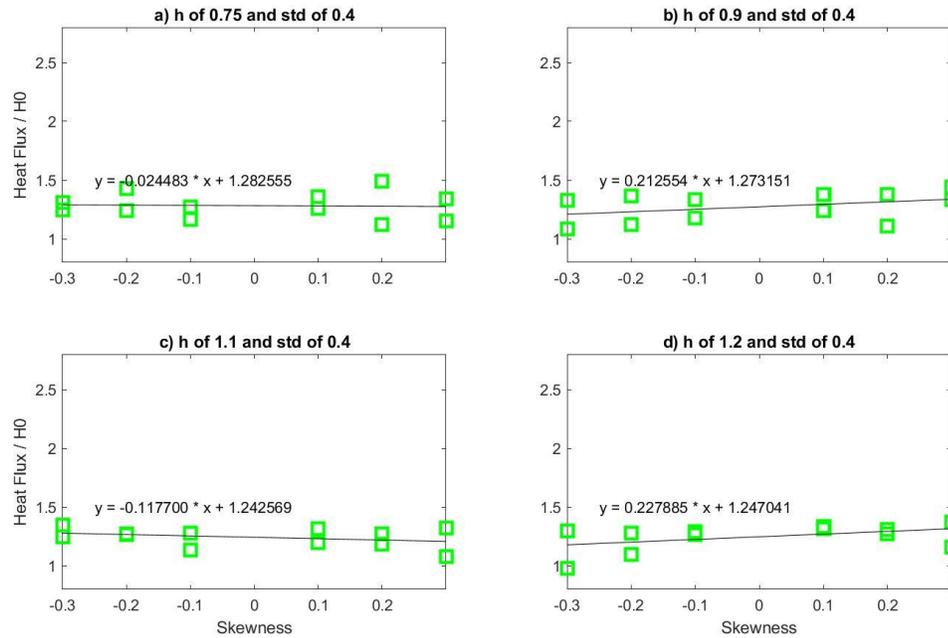
Figure 2-46: Variation of normalized heat flux with respect to skewness having a constant standard deviation of 0.4 and first block height of a) 0.75 b) 0.9 c) 1.1 d) 1.2

Table 2-11 summarizes the correlation between heat transfer and skewness for each combination of standard deviation and height of the first block. The result shows that there is not a strong association that has been established because the slope can be positive or negative, as it was concluded for pressure drop. However, the slopes are more often positive, indicating that positive skewness or taller blocks close to the leading end of the array typically result in higher rates of heat transfer. This is a weaker effect than the effect of standard deviation.

Table 2-11: Slope of heat flux change with respect to skewness

STD\h	0.75	0.90	1.10	1.20	avg
0.1	0.890	0.109	0.015	0.148	0.291
0.2	-0.013	0.055	-0.034	-0.035	-0.007
0.3	0.278	0.280	-0.124	0.095	0.132
0.4	-0.024	0.213	-0.118	0.228	0.075
avg	0.283	0.164	-0.065	0.109	0.123

2.2.3 Flow Regimes

The qualitative flow structures that occur between two adjacent blocks can be categorized into distinct regimes. Oke[44] classifies the flow regime in an idealized street canyon as follows: isolated (h/b < 0.2), wake interference (0.2 < h/b < 0.65) and skimming flow (h/b > 0.65). Chang[45] conducted a wind tunnel experiment and discovered that a 3D urban street canyon also exhibited skimming flow (h/b < 0.83), wake interference (0.83 < h/b < 0.2) and isolated roughness (h/b < 0.2). Most subsequent research describing the dynamics of urban street canyons were then based on these regimes. Since we are analyzing a 2D idealized street canyon with a constant inlet velocity, the range of each flow's ratio will vary. In this study, eight simulations with a constant height of 1 and varying spacing values of 1, 1.5, 3.5, 4.5, 5.5, 7.5,10, and 13 were conducted to identify the different regimes. The height with respect to spacing for each of these simulations is in Table 2-12. These regimes are then compared to the flow structures found in simulations with varying block heights.

Table 2-12: Height and spacing of block in simulations to study different regime

Height	Spacing	Height/ Spacing (H/S)
1	1	1.00
1	1.5	0.67
1	3.5	0.29
1	4.5	0.22
1	5.5	0.18
1	7.5	0.13
1	10	0.10
1	13	0.08

According to Oke, the simulations with H/S =1 and H/S=0.67 should be in the skimming regime, while those with H/S = 0.18, 0.13, 0.10, 0.08 should be in the isolated regime. A velocity contour plot for each simulation is shown in Figure 2-47, which illustrates the flow structures that occur for each case.
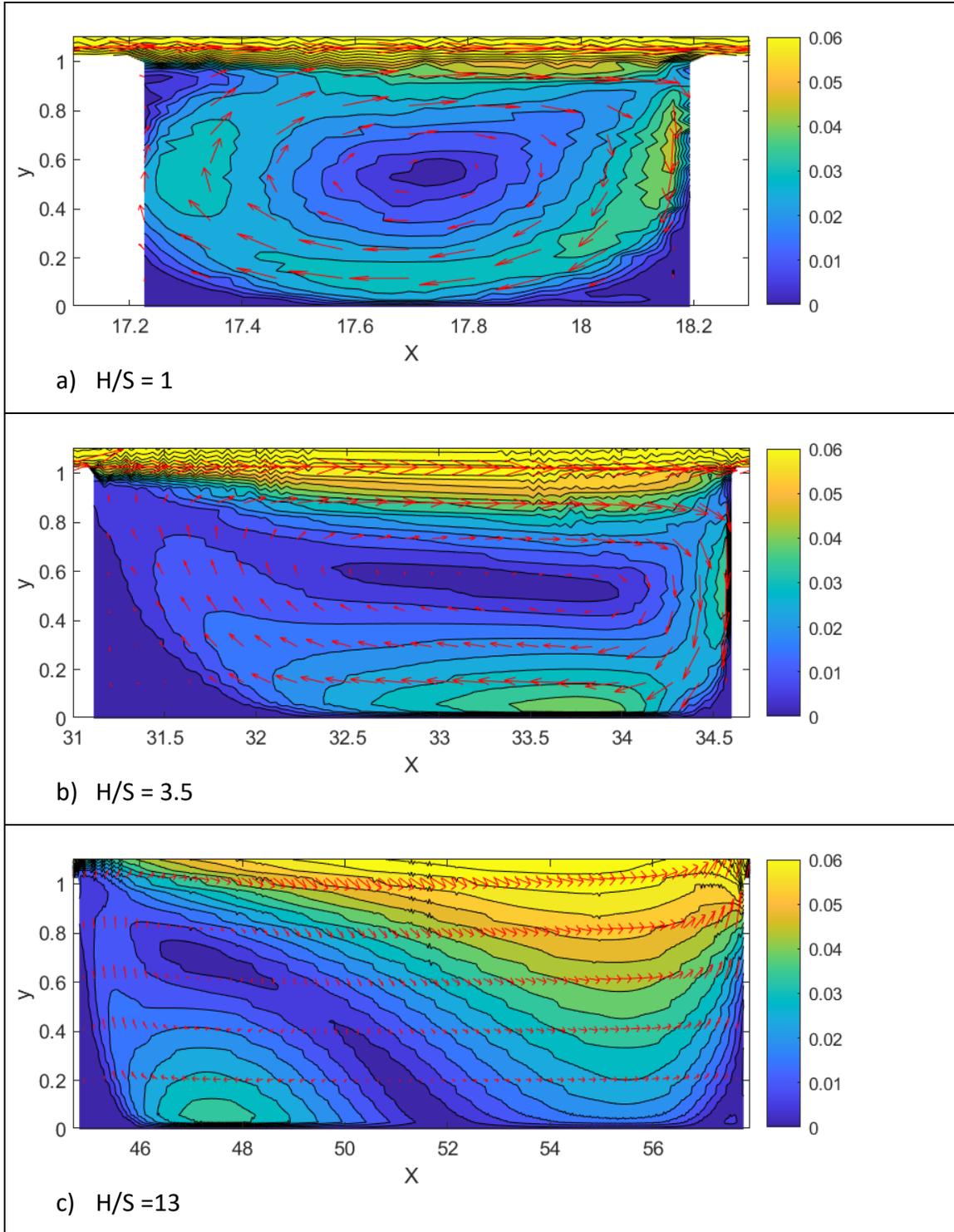
Figure 2-47: Velocity contour plot with spacing a) 1 b) 3.5 c) 13
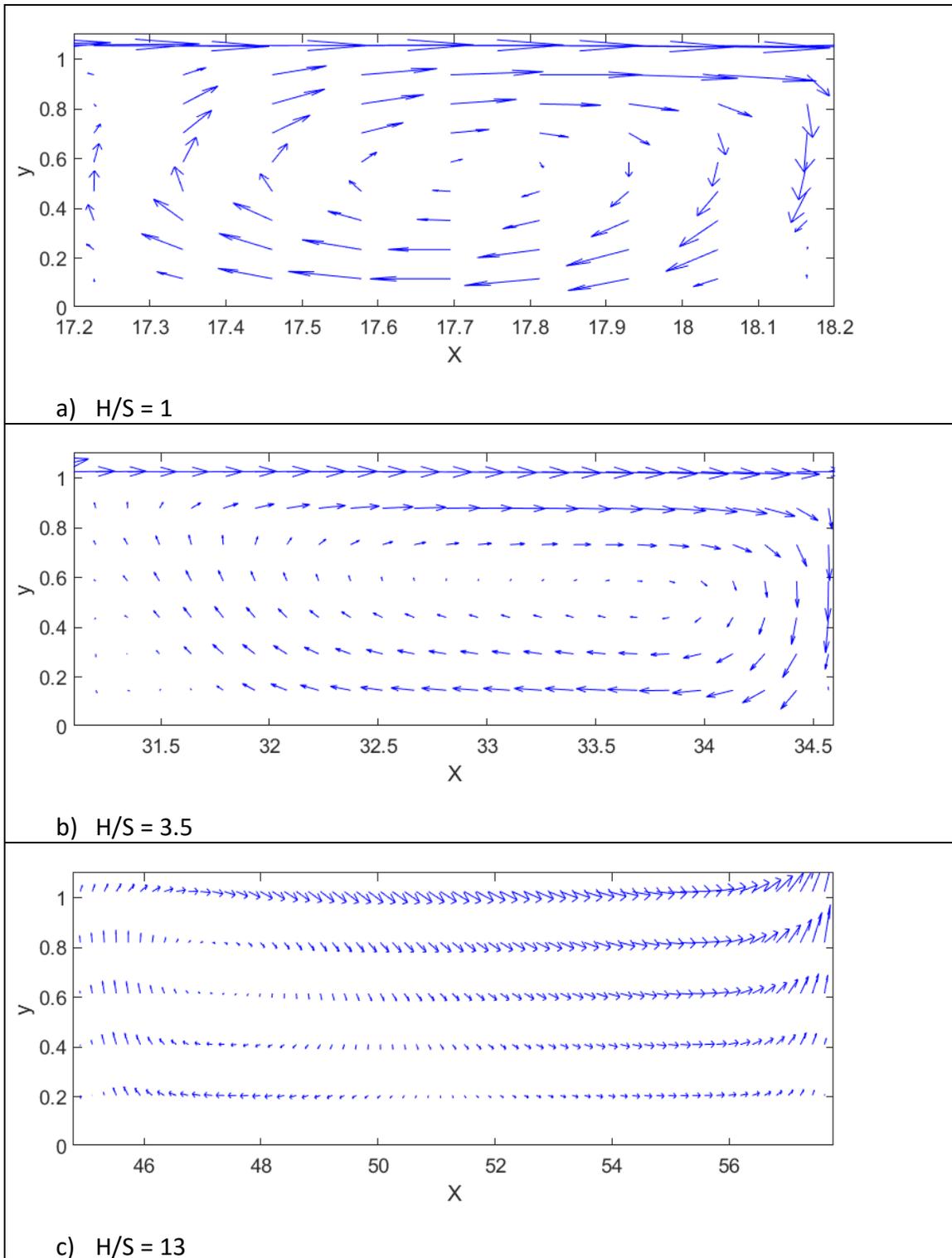
Figure 2-48: Vector plot with spacing a) 1 b) 3.5 c)13

Figure 2-47 and Figure 2-48 show that flows with S/H >0.65 are in the skimming regime, while flows with H/S < 0.1 are in the isolated regime. The isolated regime is achieved if a reattached velocity profile develops between two blocks, whereas a skimming regime develops when a single-centred vortex develops between two blocks (schematic of idealized regime shape is shown in Table 2-13). This is similar to the categorization proposed by Oke, with a slightly lower ratio required to achieve the isolated regime.

Table 2-13: schematic description of different flow regimes

Skimming	
Wake Interference	
Isolated	
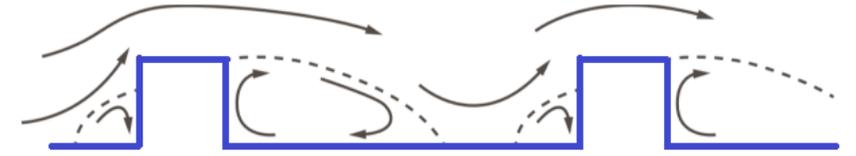
2.2.3.1   Effect of different Regimes

According to the result from section 2.2.12.2.2, the pressure difference and heat flux are correlated with the standard deviation and weakly with skewness. In some simulations, the statistical values are identical, but the pressure drop and heat flux are quite different because the individual block heights are different from any two sets of simulations. To better understand the effect of individual block heights, qualitative details of the flow are examined, including the effects of various regimes. The H/S ratio was calculated by

dividing the height of each block by its spacing. Simulations can be categorized into three distinct groups. The first is the "Only Wake Interference Regime". In this category, all of the blocks fall under the wake interference regime. The second category, "Contain all Wake Interference, Skimming, and Isolated Regime," includes simulations that cover the entire spectrum of H/S. The majority of blocks are in the regime of wake interferences, while only a few are in the regime of skimming or isolation. The final category, "Only Isolated and Wake Interferences regime" is similar to the second, with the exception that there are no blocks in the skimming regime. Notably, categories three and two do not overlap. Pressure drop with respect to standard deviation has been plotted in the figures below, with the category of each simulation indicated:
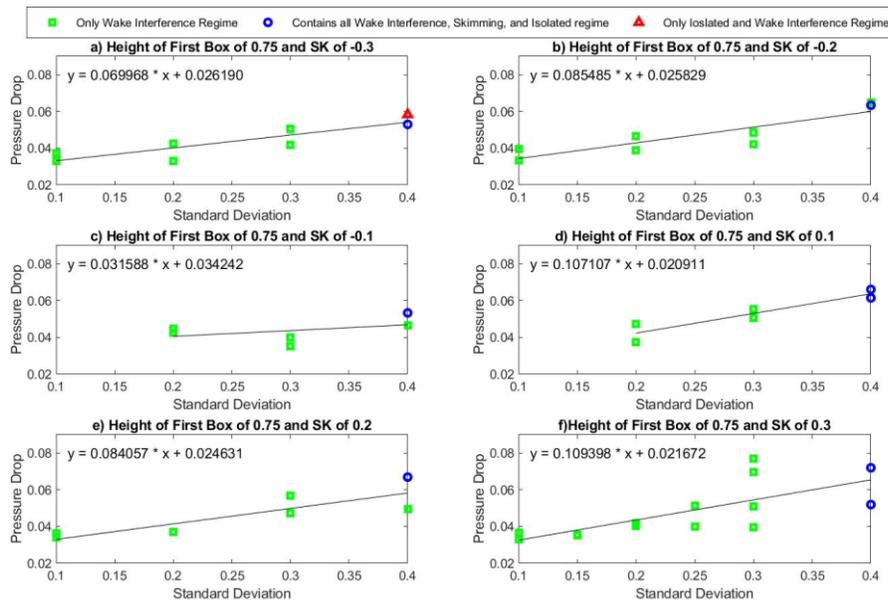
Figure 2-49: Pressure drop (Pa) with respect to standard deviation in different flow regimes with the height of the first block of 0.75 and skewness of a) -0.3 b) -0.2 c) -0.1 d) 0.1 e) 0.2 f) 0.3
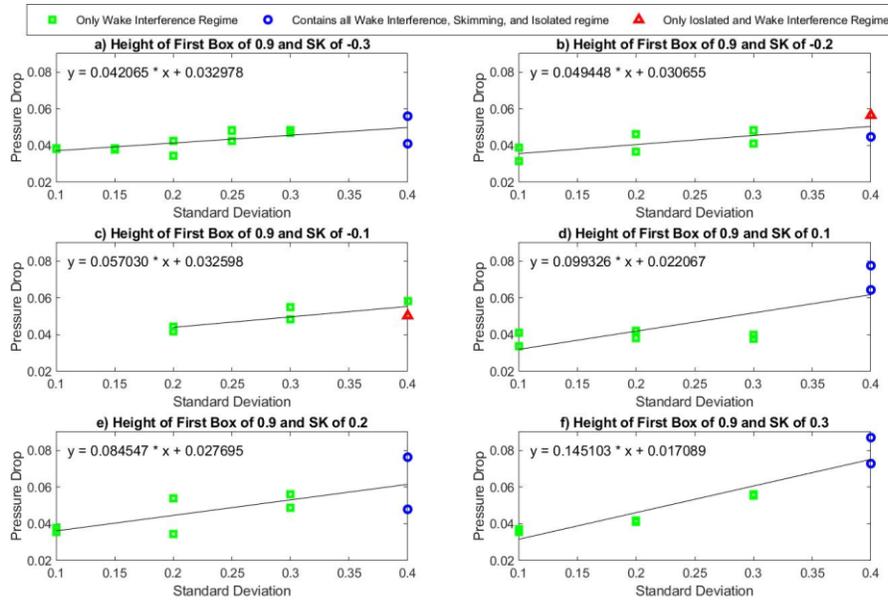
Figure 2-50: Pressure drop (Pa) with respect to standard deviation in different flow regimes with the height of the first block of 0.9 and skewness of a) -0.3 b) -0.2 c) -0.1 d) 0.1 e) 0.2 f) 0.3
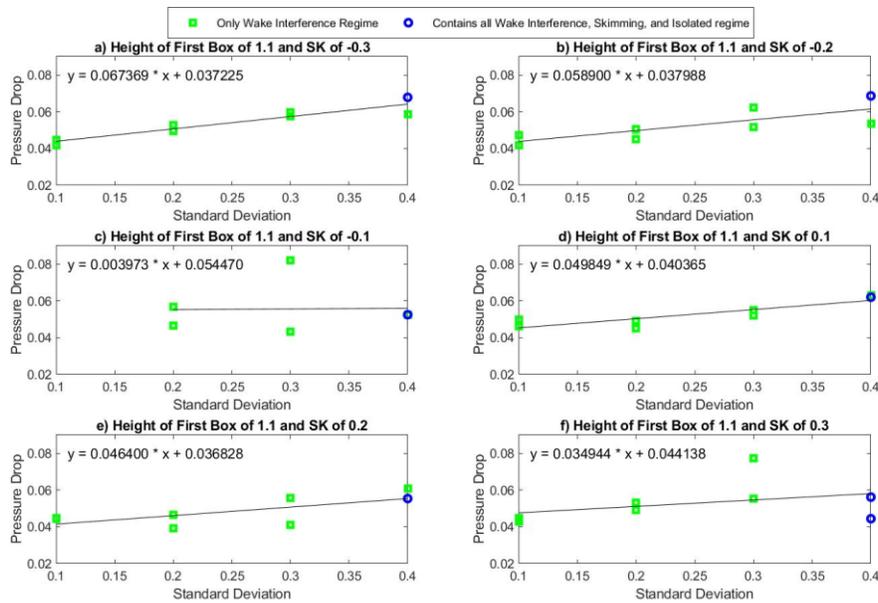
Figure 2-51: Pressure drop (Pa) with respect to standard deviation in different flow regimes with the height of the first block of 1.1 and skewness of a) -0.3 b) -0.2 c) -0.1 d) 0.1 e) 0.2 f) 0.3
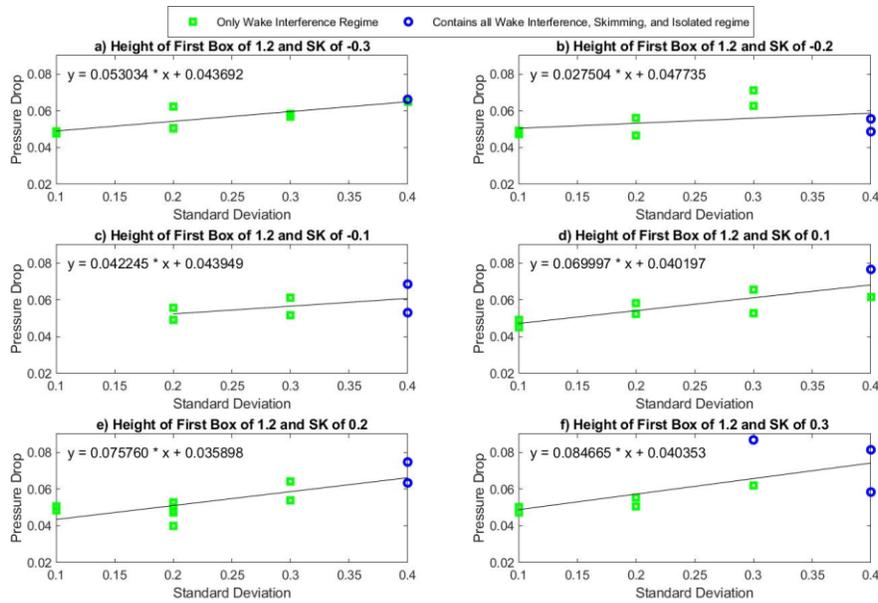
Figure 2-52: Pressure drop (Pa) with respect to standard deviation in different flow regimes with the height of the first block of 1.2 and skewness of a) -0.3 b) -0.2 c) -0.1 d) 0.1 e) 0.2 f) 0.3

To investigate whether the relationship between pressure drop and standard deviation may exhibit a different pattern depending on the flow category, two distinct linear trend lines have been plotted: one for simulations in the wake interference region and the other for all simulations. In Figure 2-53, simulations with noticeable trend lines have been shown:
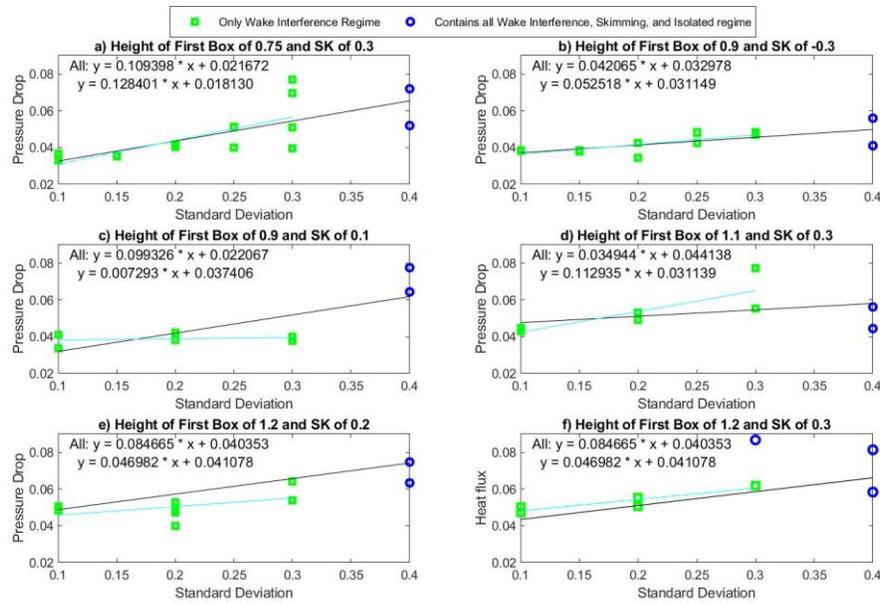
Figure 2-53: Different trend lines in different regimes a)H1 = 0.75, SK = -0.3 b) H1 = 0.9, SK = -0.3 c) H1 = 0.9, SK = 0.1 d) H1 = 1.1, SK = 0.3 e) H1 = 1.2, SK = 0.2 f) H1 =1.2, SK =0.3

The results demonstrate that the type of simulation regime has a small influence on the correlation between the standard deviation of block heights and pressure drop. The trend of the pressure drop increasing with increasing standard deviation is consistent whether simulations include all regimes or only a subset of the regimes.

2.2.3.2   Effect of the taller block on flow structure

Due to the presence of significant variation in the heights of blocks in each simulation set, there are additional flow structures that do not fall into the previously mentioned regimes and have a distinct appearance. For instance, a tall block's wake region may influence several subsequent blocks, creating different regime structures. Figure 2-54 illustrates the situation in which a small block exists between two larger blocks. The flow passes in the opposite direction over the middle blocks while generating a vortex between the first and second blocks.
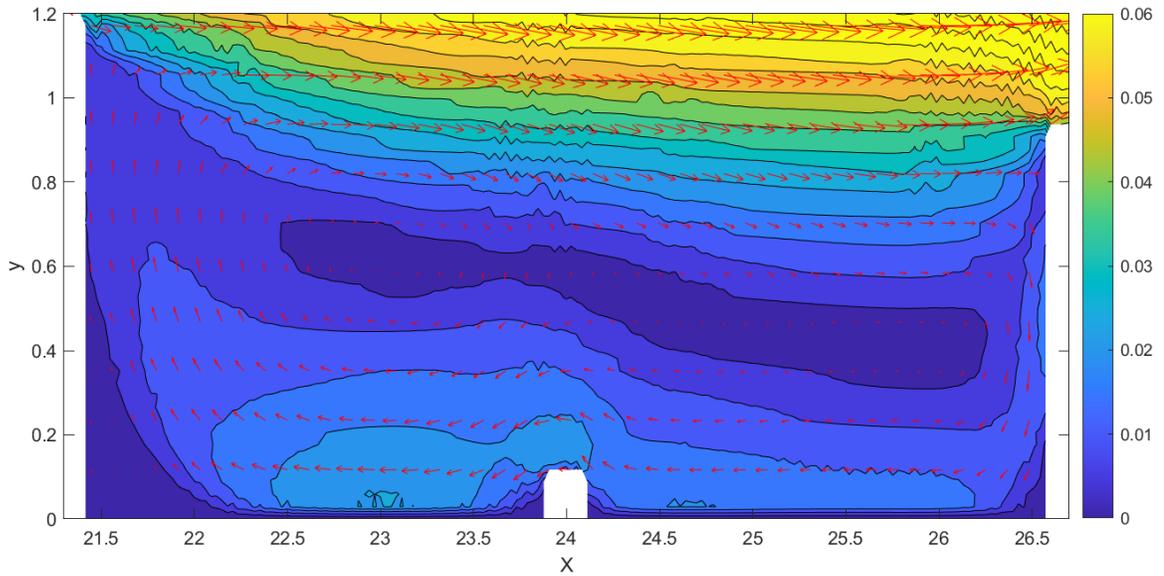
Figure 2-54: flow structure having one small block between two larger blocks

A tall block is followed by two smaller blocks in Figure 2-55. The wake of taller blocks influences the flow structure surrounding smaller blocks and creates a flow in the opposite direction above the smaller blocks.
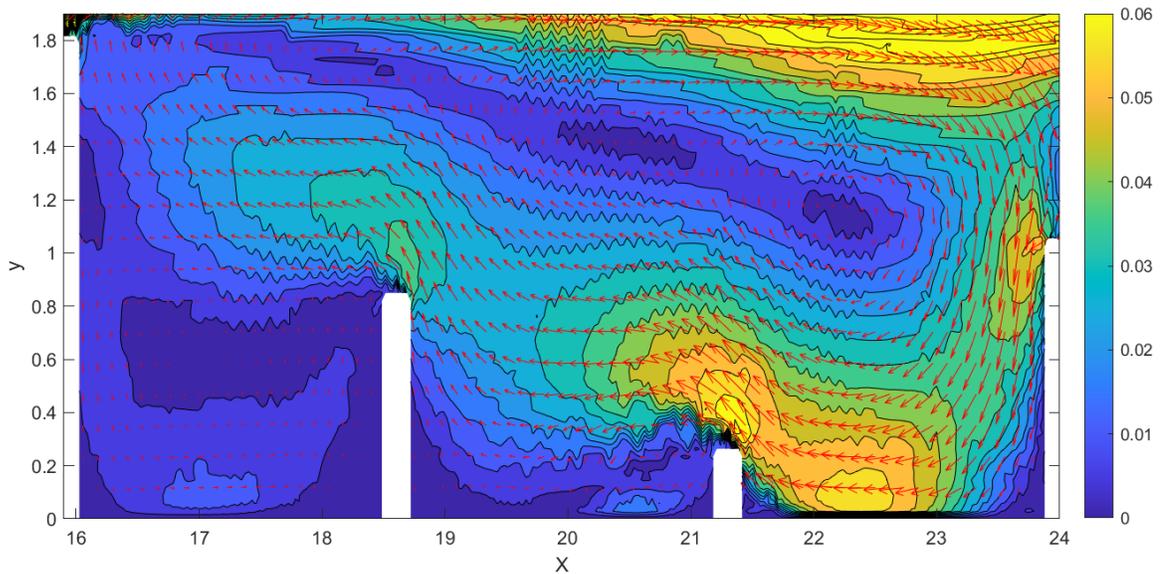
Figure 2-55: flow structure having one tall block followed by two smaller blocks

2.2.4   Different Height Average

To further investigate the effect of the various flow regimes, simulations with standard deviations ranging from 0.2 to 1.5 and height averages of 2, 2.5, and 3 are introduced. Simulations with larger heights are necessary for blocks in the skimming regime to occur within a simulation. Noting that the simulation domain for this set of simulations is entirely distinct from the previous set, these simulations are not directly comparable to the previous set; the boundary condition remains the same, but since H is being altered, the downstream space, height, and upstream space differ from previous simulations.

Figure 2-56 to Figure 2-59 illustrate how the heat flux and pressure drop depend on standard deviation. As predicted, pressure drop and heat flux followed the trend of increasing as the standard deviation increased, which supports the conclusion reached in the preceding sections. Figure 2-56 and Figure 2-57 also depict the regime of each simulation. "Only skimming regime" refers to the simulation in which all heights relative to spacing fall within the range of the skimming regime. "Skimming regime and wake interference regime" refers to the simulation in which the majority of blocks are in the

skimming regime, and only a few meet the wake interference criteria. Finally, in "contain all the regimes," all three distinct regimes appear within the simulation.
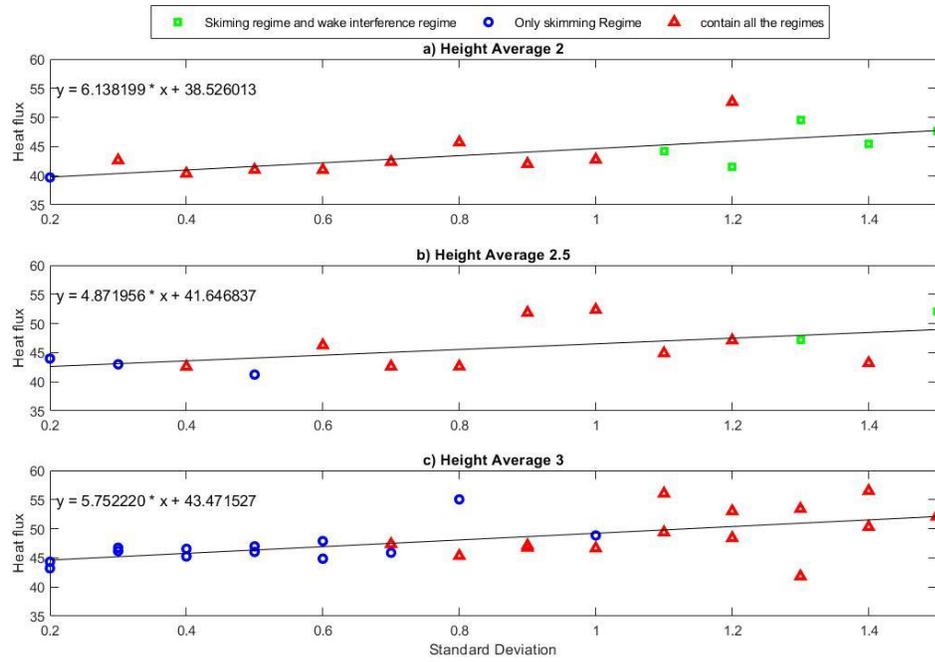
Figure 2-56: Heat flux with respect to a standard deviation with a height average of a) 2 b) 2.5 c) 3
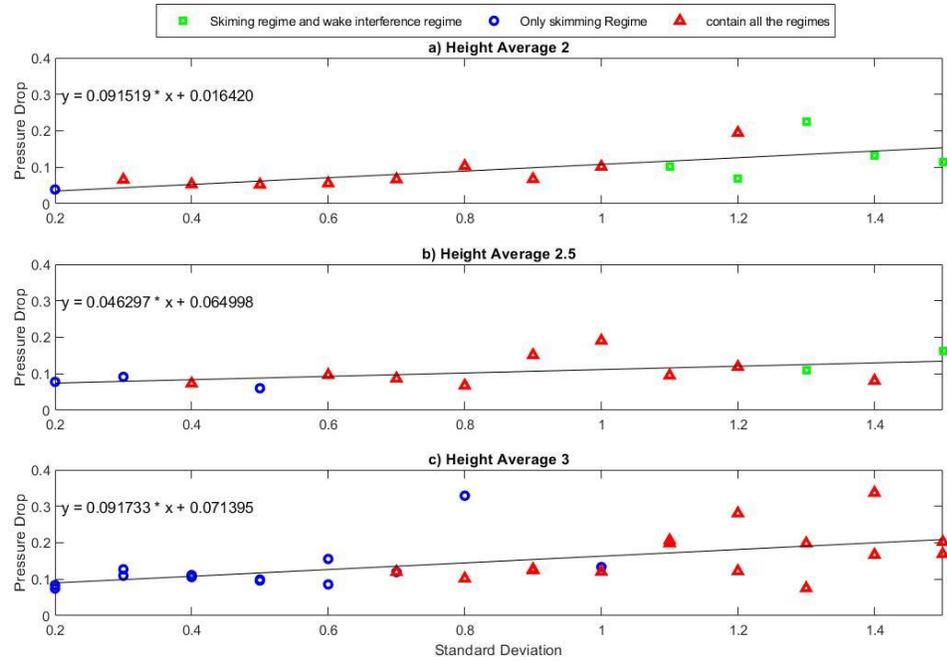
Figure 2-57: Pressure drop with respect to a standard deviation with a height average of a) 2 b) 2.5 c) 3

The heat flux and pressure drop will increase with increasing average height as shown in Figure 2-58 and Figure 2-59. However, the arrangement and height of each obstacle are of

greater importance since the average slope of the effect of standard deviation is greater than the average height effect.
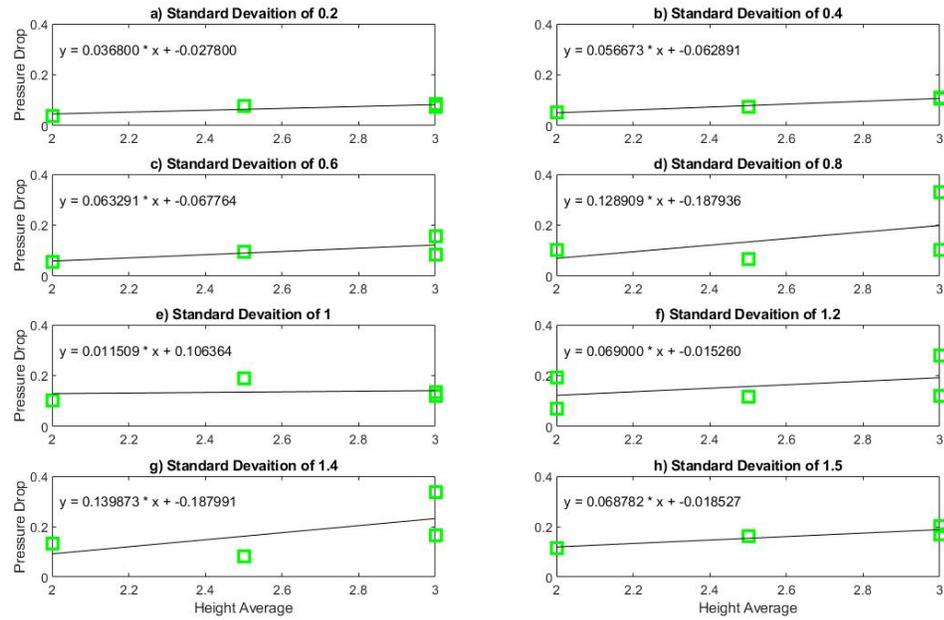
Figure 2-58: Pressure drop with respect to height average with a standard deviation of a) 0.2 b) 0.4 c) 0.6 d) 0.8 e) 1 f) 1.2 g)1.4 h)1.5
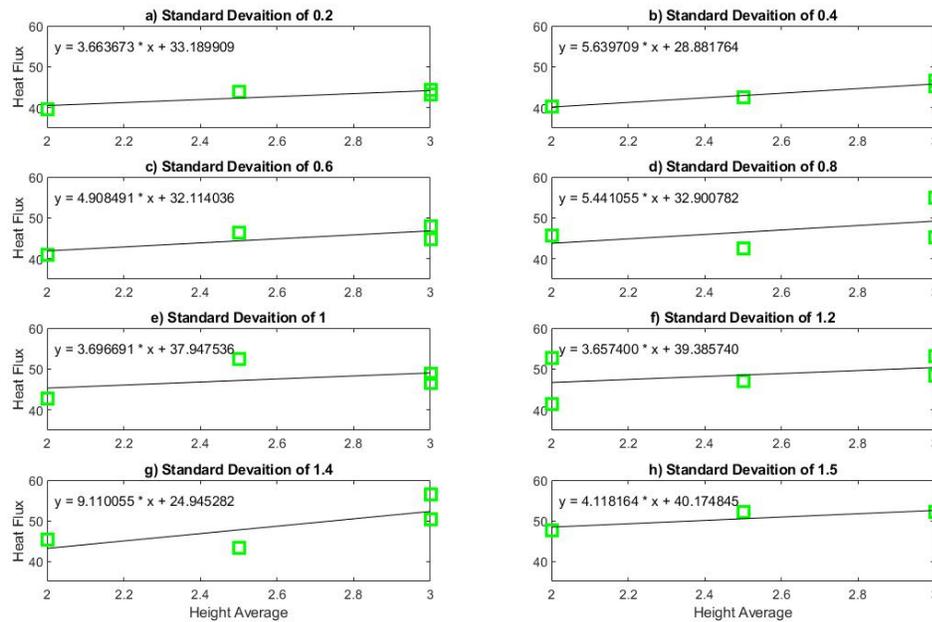
Figure 2-59: Heat flux with respect to height average with a standard deviation of a) 0.2 b) 0.4 c) 0.6 d) 0.8 e) 1 f) 1.2 g)1.4 h)1.5

2.2.5   Effect of Arrangement of blocks

The previous section investigates the effect of statistical measures and flow regimes on pressure drop and heat flux. In this section, six sets of simulations involving 14 blocks with heights that have the same standard deviation, skewness, first block height, and overall height average were conducted to investigate the qualitative flow behavior that results in different pressure drop and heat flux for simulations with identical values for the statistical parameters considered. The heights and related information are shown in Table 2-14:

Table 2-14: Heights, Standard deviation, Skewness, and Kurtosis of 6 simulations to explore the overall effect of the block's height

	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6
H1	1.20	1.20	1.20	1.20	1.20	1.20
H2	1.45	1.03	0.83	0.80	0.83	1.33
H3	1.14	1.12	1.01	0.68	1.06	0.96
H4	0.94	0.80	1.22	1.13	0.97	0.97
H5	1.05	1.09	1.02	1.40	0.96	0.67
H6	0.84	1.05	1.09	0.75	0.93	0.97
H7	1.09	1.13	0.76	1.15	0.86	1.18
H8	0.59	1.43	1.01	1.13	0.82	1.33
H9	0.92	1.00	1.40	0.85	1.23	1.03
H10	0.93	0.98	1.02	1.16	1.08	0.89
H11	0.83	0.73	0.70	0.86	0.68	0.76
H12	1.04	0.77	1.13	0.88	1.29	0.79
H13	1.04	0.96	0.75	1.02	1.33	0.98
H14	0.95	0.67	0.85	0.97	0.78	1.00
Standard deviation	0.199	0.204	0.203	0.204	0.202	0.199
skewness	0.201	0.204	0.200	0.196	0.198	0.197
mean	1.000	0.997	1.001	0.998	1.002	1.005
first height	1.200	1.195	1.205	1.197	1.202	1.202
kurtosis (Not equal for all cases)	3.822	2.753	2.243	2.194	1.873	2.272

The flow is in the wake interference regime for all blocks in all of these simulations. The following are the total pressure drop and heat flux for these six simulations (Table 2-15):

Table 2-15: Heat flux and pressure drop for six simulations with different statistical measures

	SET 1 (A)	SET 2 (B)	SET 3 (C)	SET 4 (D)	SET 5 (E)	SET 6 (F)
HEAT FLUX	34.528	33.348	34.658	34.353	34.833	34.143
PRESSURE DROP	0.0407	0.0472	0.0514	0.0529	0.0518	0.0478

All simulations have heat fluxes within 5 % of each other, but the pressure drops vary more significantly. Simulation 4 has the greatest pressure drop, while simulation 1 has the lowest, and the difference is about 30%, despite the fact that all simulations have the same standard deviation and average height. The velocity, turbulent kinetic energy, and vorticity profile are examined in order to determine the cause of the varied pressure drop in these simulations.

The velocity profile for each of these simulations in the x direction, u, is shown in Figure 2-60.

a)

b)

c)

d)

e)
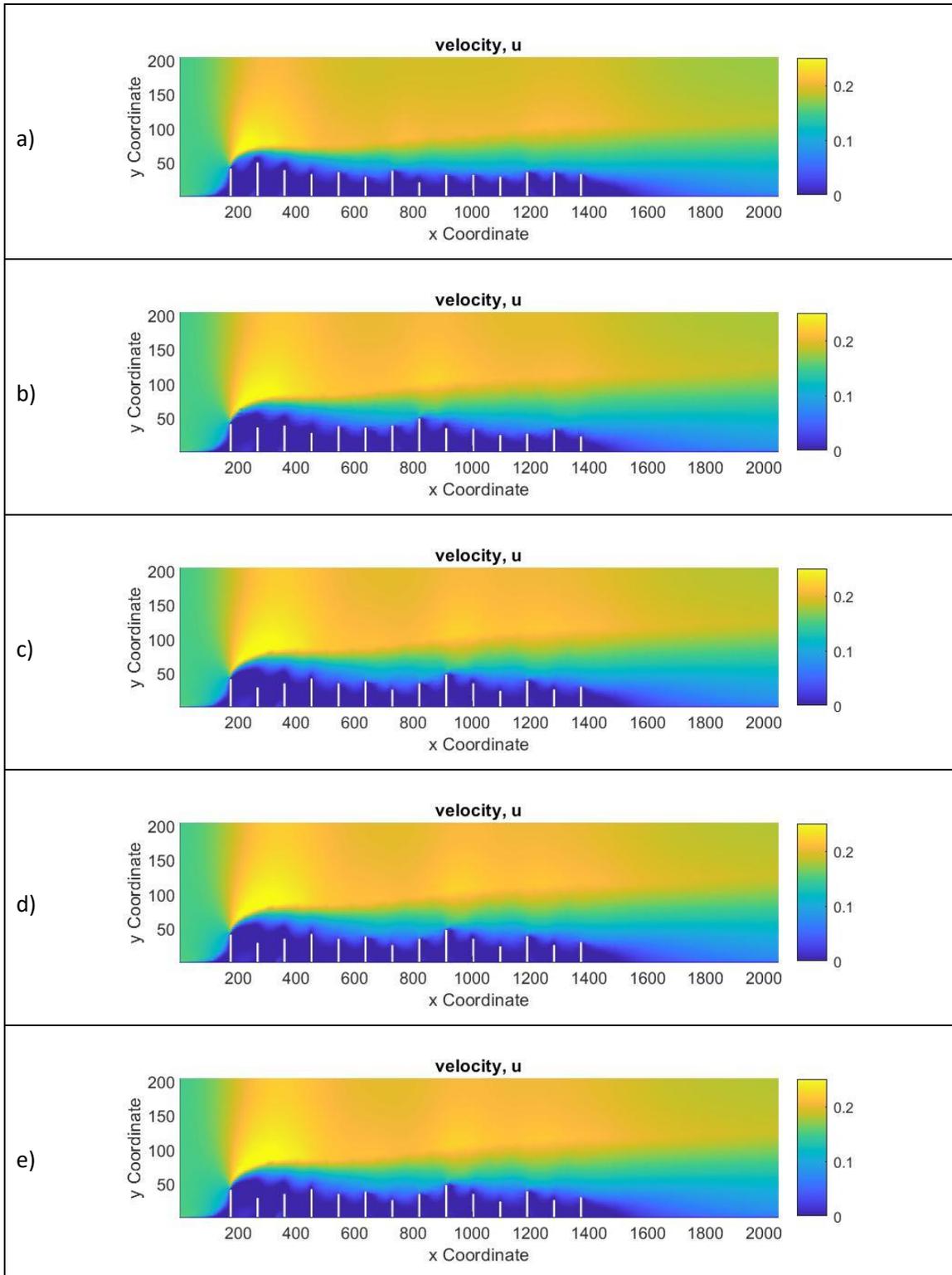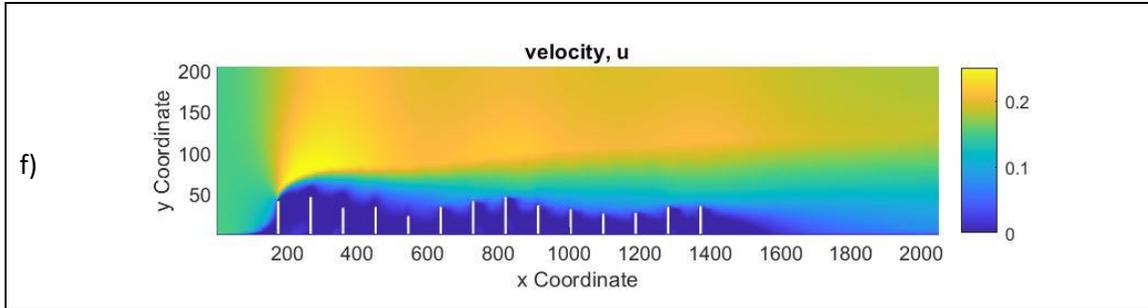
Figure 2-60: Velocity profile (m/s) along x direction for. a) simulation 1 b) simulation 2 c) simulation 3 d) simulation 4 e) simulation 5 f) simulation 6

The Turbulent kinetic energy plots are as follows (Figure 2-61):
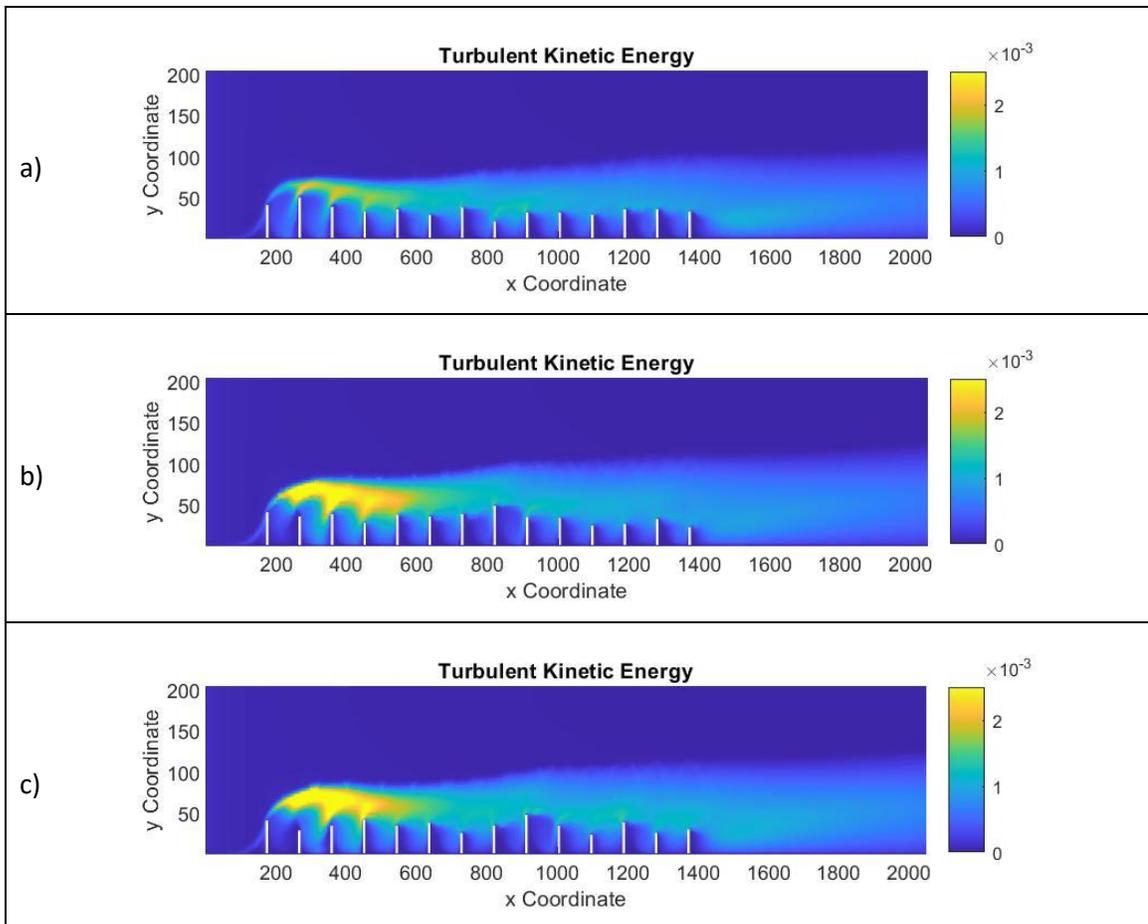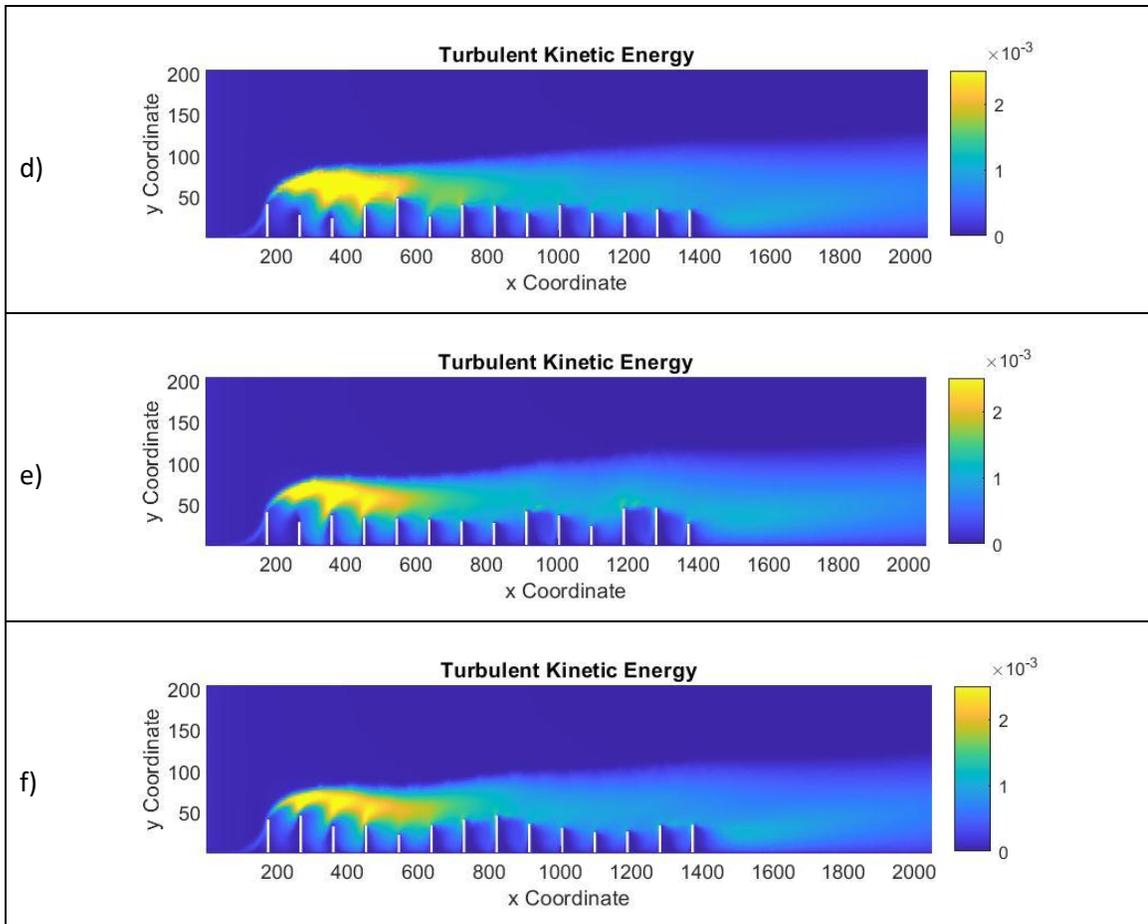
Figure 2-61: Turbulent kinetic energy profile (J/kg) for a) simulation 1 b) simulation 2 c) simulation 3 d) simulation 4 e) simulation 5 f) simulation 6

The vorticity profile is in Figure 2-62:
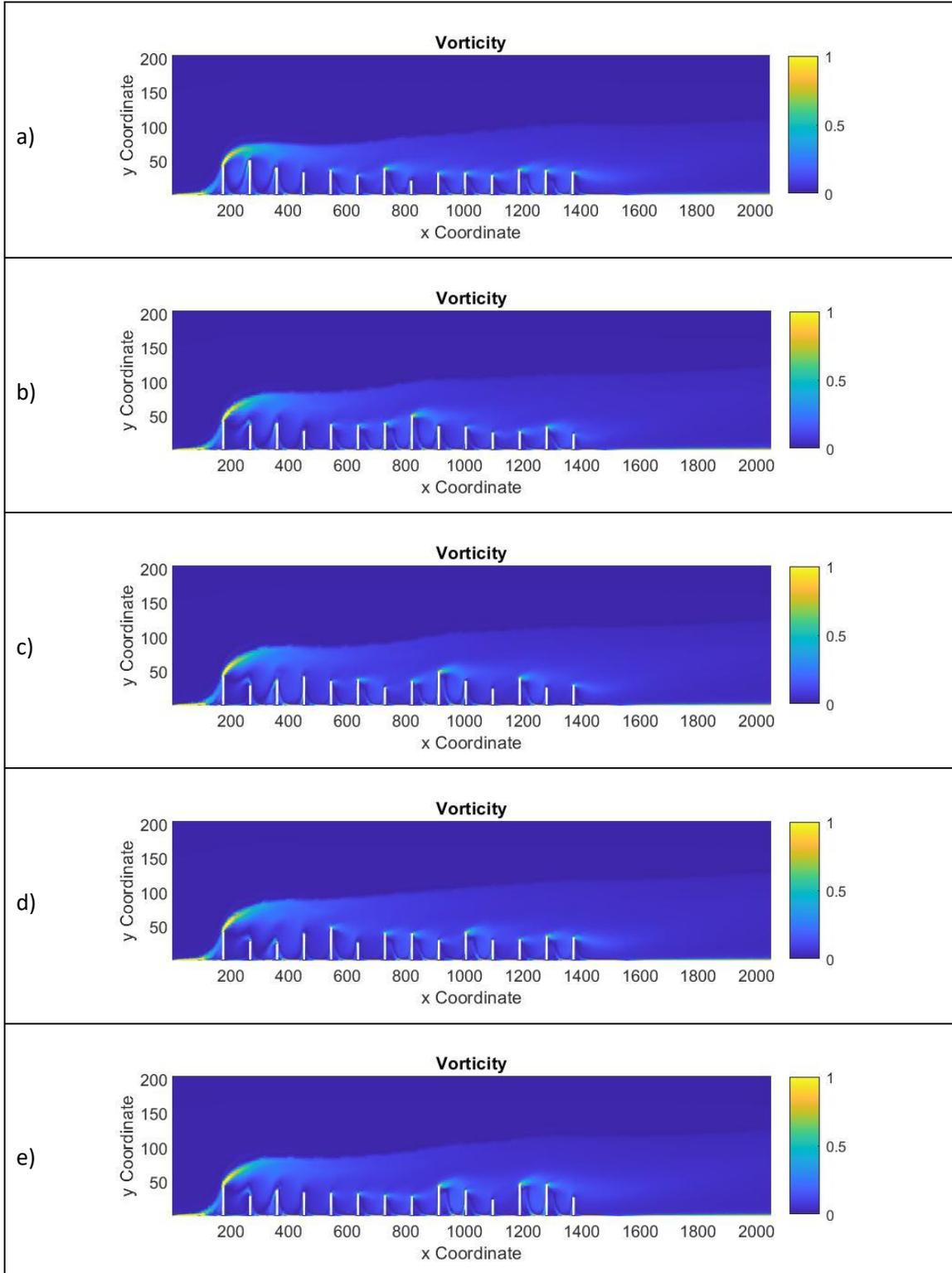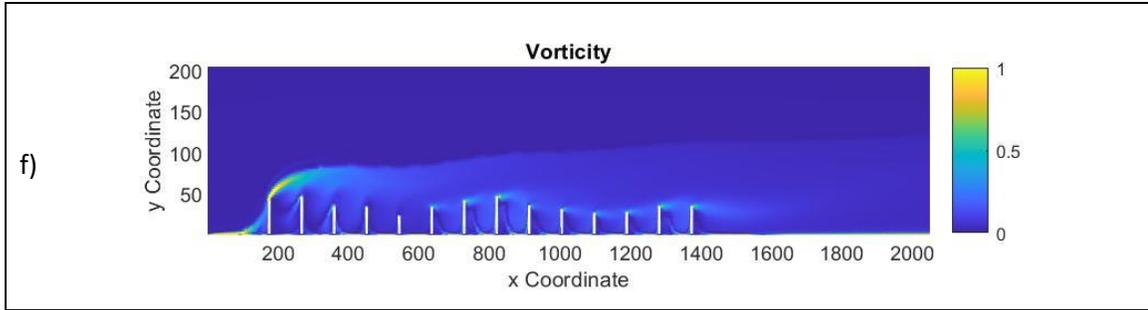
Figure 2-62: Vorticity profile for. a) simulation 1 b) simulation 2 c) simulation 3 d) simulation 4 e) simulation 5 f) simulation 6 in Table 13.

The static pressure profile is as follows (shown in Figure 2-63):
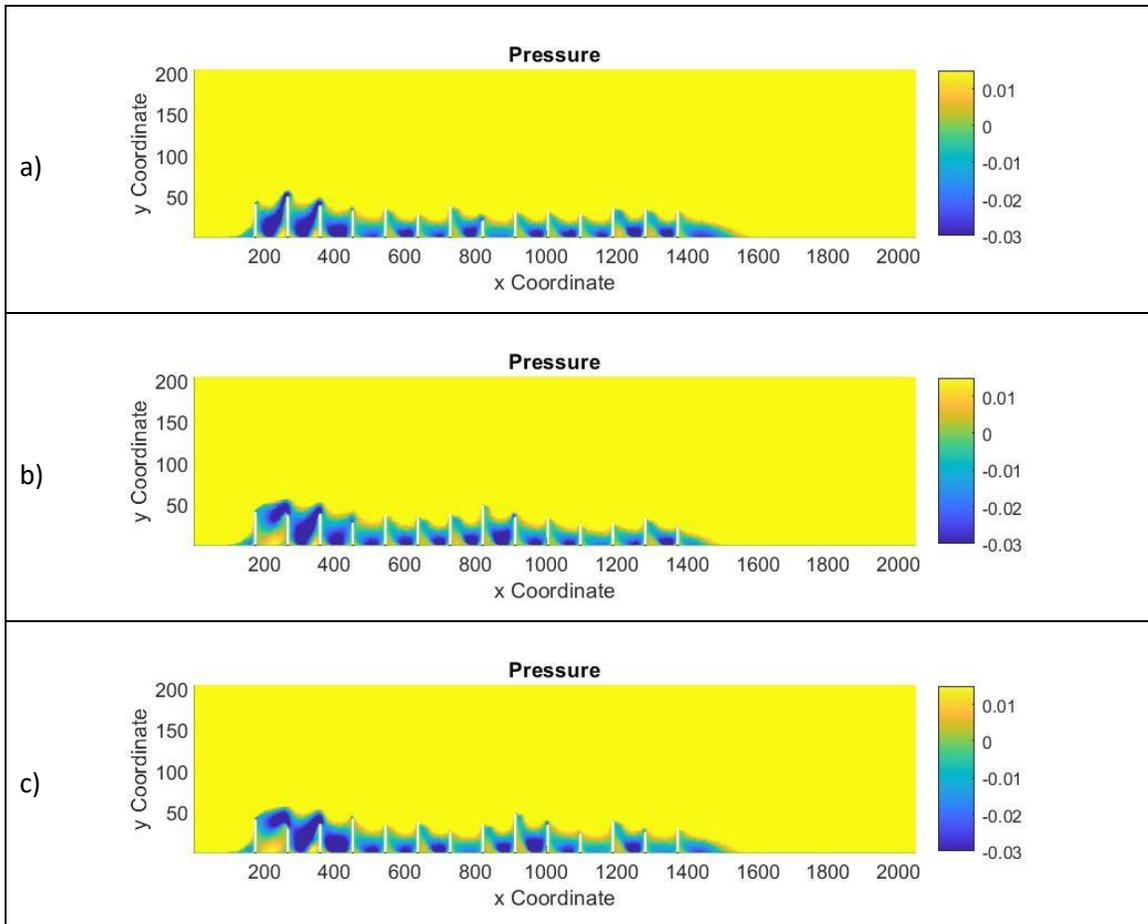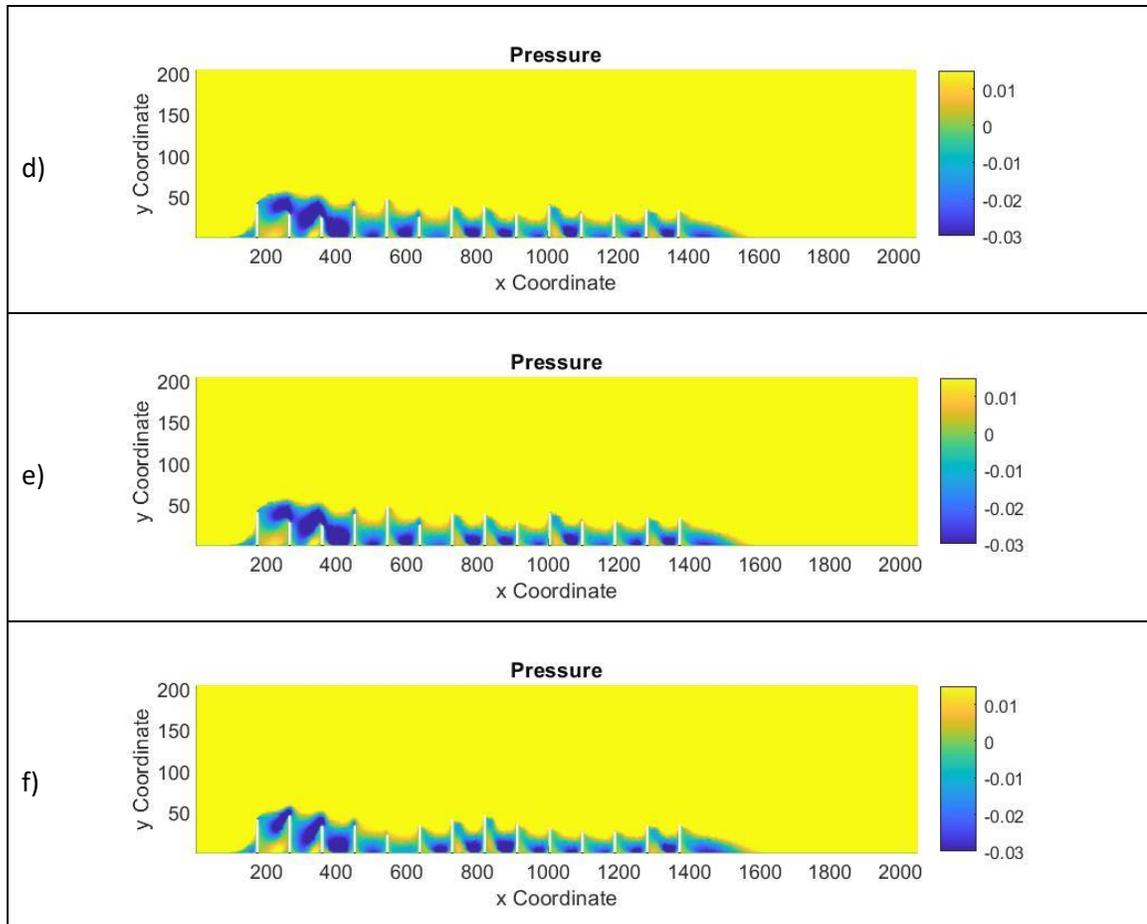
Figure 2-63: Pressure profile (Pa) for. a) simulation 1 b) simulation 2 c) simulation 3 d) simulation 4 e) simulation 5 f) simulation 6 in Table 13.

The simulations with a greater pressure drop (c, d, e) exhibit more intense turbulent kinetic energy at the top of the initial few blocks. These simulations also have the largest height difference h_1 - h_2, indicating that the height difference between the first two blocks in the wind path has the greatest effect on the total heat flux and pressure drop for setups that are otherwise statistically similar. The first case, which has the lowest pressure drop, also has the lowest turbulent kinetic energy and different pressure distribution between the first two blocks. This provides some insight into how the qualitative details of the flow are not always captured by the first few statistical moments.

2.2.6 Effect of the height difference between the first two blocks

The height difference between the first two blocks has the greatest impact on the total pressure drop and heat flux for simulations with equivalent mean, standard deviation, and skewness, as concluded in the previous section. The graphs below illustrate how this height difference affects the pressure drop. The height difference is calculated by subtracting the height of the second block from the height of the first block (h_1 - h_2). A line was fit to the data to represent the relationship between the pressure drop and height difference. The results are categorized based on the regimes defined in section 2.2.3.
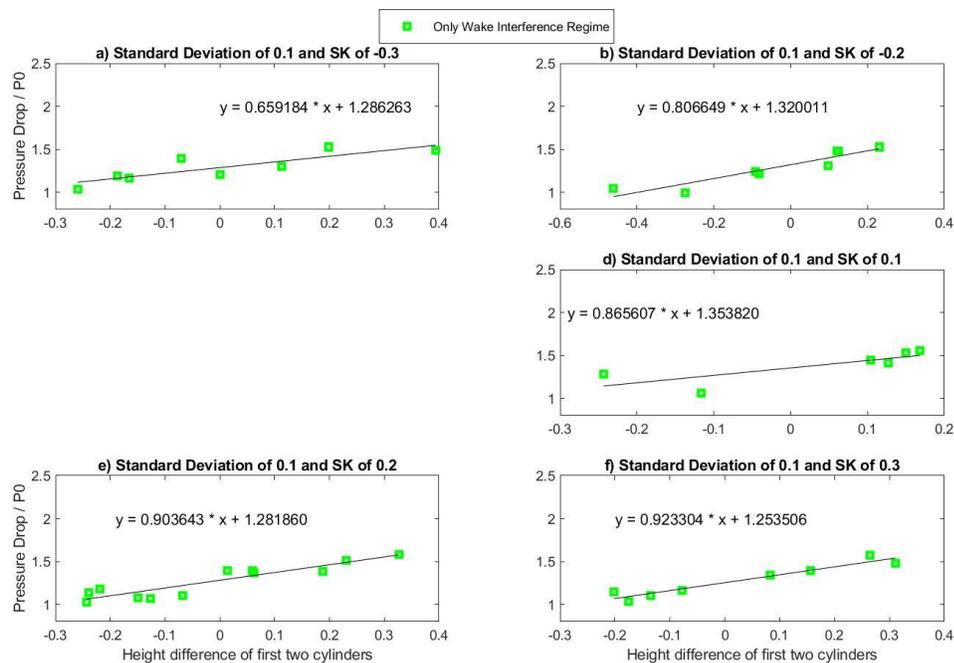
Figure 2-64: Pressure drop with respect to height difference having a standard deviation of 0.1 and skewness of a) -0.3 b) -0.2 d) 0.1 e) 0.2 f) 0.3
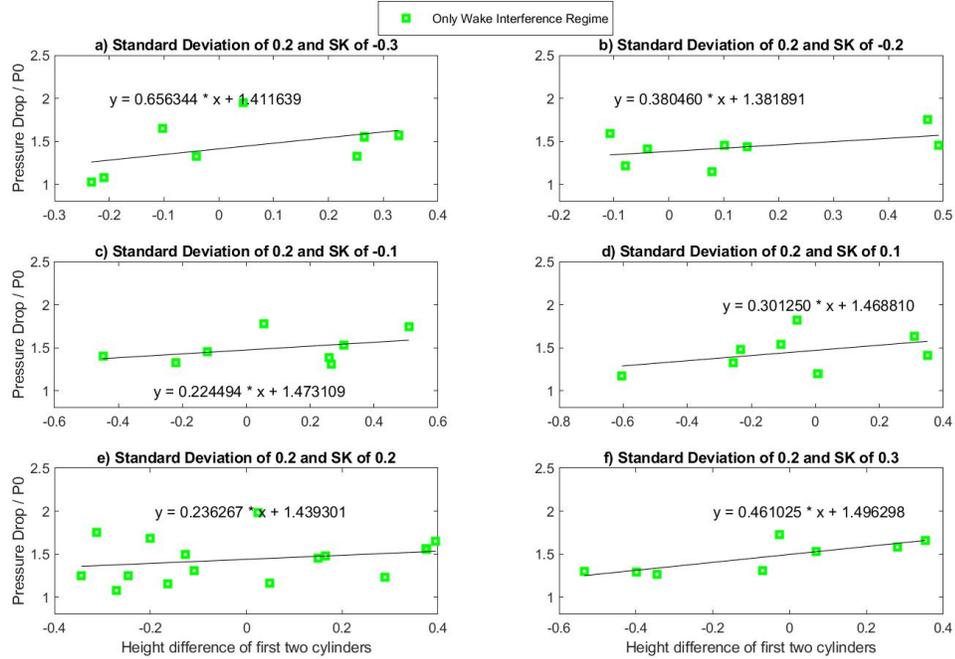
Figure 2-65: Pressure drop with respect to height difference having a standard deviation of 0.2 and skewness of a) -0.3 b) -0.2 c)-0.1 d) 0.1 e) 0.2 f) 0.3
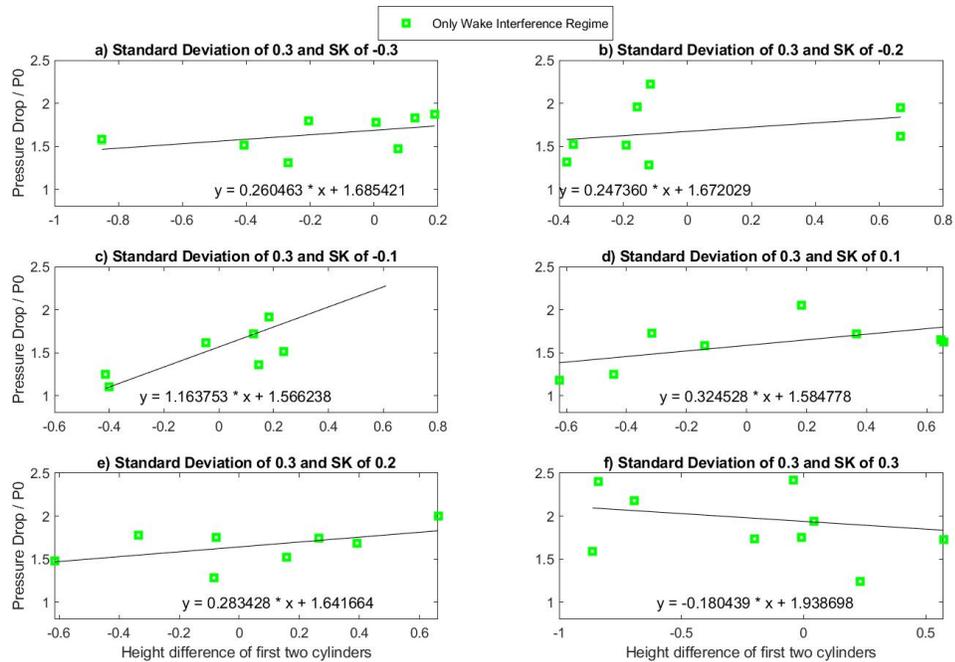
Figure 2-66: Pressure drop with respect to height difference having a standard deviation of 0.3 and skewness of a) -0.3 b) -0.2 c)-0.1 d) 0.1 e) 0.2 f) 0.3
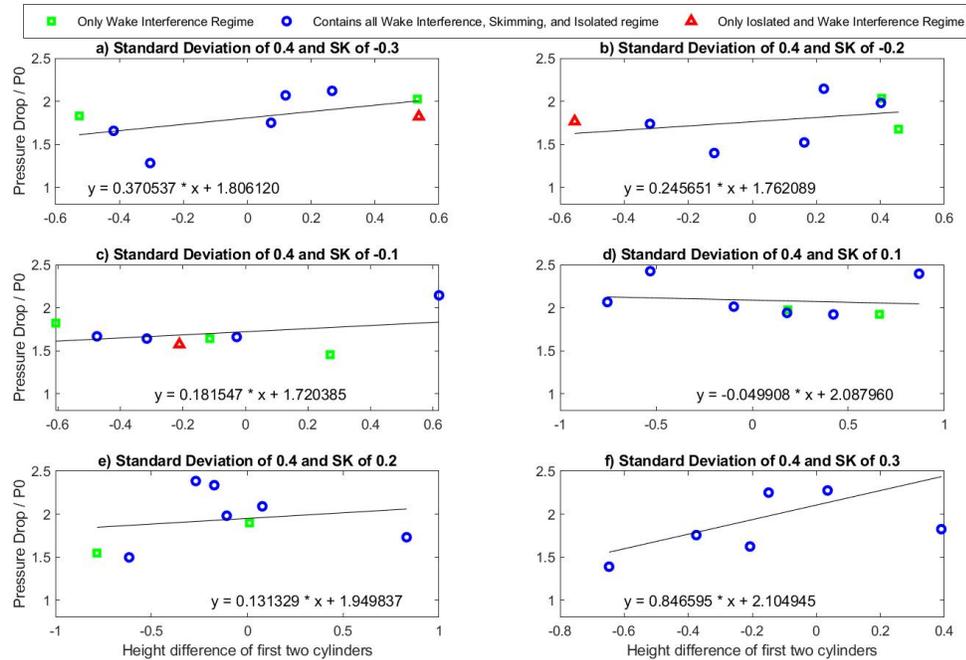
Figure 2-67: Pressure drop with respect to height difference having a standard deviation of 0.4 and skewness of a) -0.3 b) -0.2 c)-0.1 d) 0.1 e) 0.2 f) 0.3

The lower standard deviation of a set of heights means less fluctuation in height differences within a set. Figure 2-64 depicts the same slope for the simulation with a lower standard deviation, indicating that the relationship between the height difference of the first two blocks and pressure drop is constant for simulations with a lower standard deviation. At the same time, as the standard deviation increased, the slope value fluctuated, and the average slope value decreased. Note that in simulations that include all the flow regimes, the fluctuation in pressure drop is higher than the one that only has one regime.

A cubic polynomial has been fitted to the simulated data in Figure 2-68. The result suggests that when the second obstacle is larger than the first one, the standard deviation highly affects the total pressure drop compared with the case where the second obstacle is smaller. By increasing the standard deviation, the pressure drop also increases. The highest Pressure drop value is obtained at a combination of high standard deviation and a larger first obstacle. When the height of the first block is greater than that of the second, the pressure drop is often greater.
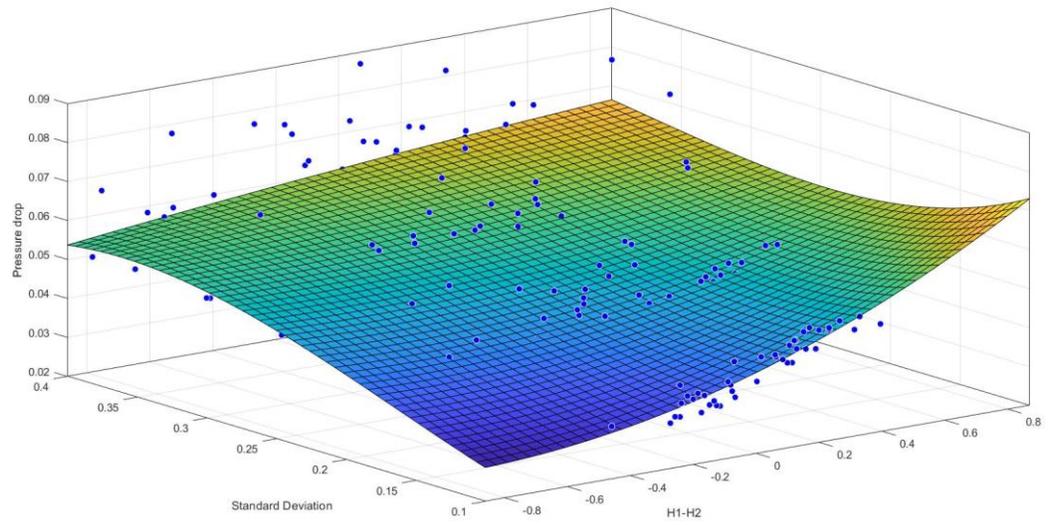
Figure 2-68: 3D plot of pressure drop with respect to height difference and standard deviation

2.2.7 Surface heat transfer Coefficient (SHTC)

The investigation of simulations with similar statistics indicates that the same statistics do not always imply equal pressure drop, heat flux, and heat transfer and that the arrangement of height is crucial. The most important variable in this analysis was the height difference between the first two blocks, followed by the height differences between the remaining blocks. Due to the nonlinear effect of the height difference between the adjacent blocks, the notion of testing machine learning's capacity to forecast variables such as the heat transfer coefficient has developed.

The convective heat transfer coefficient (CHTC) is an indispensable metric for determining the microscale flow and energy exchange between a block unit and its environment. Statistical trends on the heat transfer coefficient have been analyzed to determine the possibility of a correlation, and in the following chapter, a machine learning model is implemented to capture the correlation between block height and heat transfer coefficient.

Using the results of 204 simulations with random block heights, the surface heat transfer coefficient for each block surface was calculated. Note that surface heat transfer coefficient and convective heat transfer coefficient are equivalent, whereas in Ansys, the term SHTC has been used, so this study will continue to use this terminology. The average surface heat transfer coefficient for each block is considered. This section examines the impact of different statistical values on the average coefficient value of blocks 12 and 13. Therefore, in this section, SHTC is referred to as the average SHTC value of blocks 12th and 13th.

The effect of the first block's height on the average surface heat transfer coefficient was studied, and a line was fit to the data. The slopes are illustrated in Table 2-16.

Table 2-16: Slope of SHTC changes in blocks 12 to 13 with respect to the height of the first block

STD/SK	-0.3	-0.2	-0.1	0.1	0.2	0.3	AVG
0.1	0.125	0.068	NAN	0.098	0.034	0.109	0.087
0.2	0.121	0.278	-0.083	0.139	0.084	-0.18	0.06
0.3	0.097	0.053	0.152	0.064	0.121	0.024	0.085
0.4	0.333	-0.016	-0.237	-0.134	-0.215	-0.116	-0.064
AVG	0.169	0.096	-0.056	0.042	0.006	-0.041	0.04

Table 2 17 illustrates the effect of standard deviation.

Table 2-17: Slope of SHTC changes in blocks 12 to 13 with respect to standard deviation

SK/h	0.75	0.9	1.1	1.2	AVG
-0.3	-0.453	-0.261	-0.177	-0.186	-0.269
-0.2	0.076	0.12	-0.134	-0.071	-0.002
-0.1	0.255	-0.429	-0.447	-0.097	-0.18
0.1	0.06	0.071	-0.305	-0.449	-0.156
0.2	-0.148	-0.299	0.222	-0.766	-0.248
0.3	0.172	0.016	0.008	-0.01	0.024
AVG	-0.006	-0.13	-0.139	-0.278	-0.138

Table 2 18 depicts the influence of skewness.

Table 2-18: Slope of SHTC change in blocks 12 to 13 with respect to skewness

STD/h	0.75	0.9	1.1	1.2	AVG
0.1	0.015	0.014	-0.014	-0.007	0.002
0.2	0.042	-0.038	-0.08	-0.969	-0.261
0.3	0.115	-0.036	0.141	-0.005	0.054
0.4	0.118	0.058	0.062	-0.213	0.006
AVG	0.073	0	0.027	-0.299	-0.05

We do not expect to find a strong correlation between the heat transfer coefficient of blocks and defined statistics, as the heat transfer coefficient is determined based on what is occurring locally and not on broad statistics of block sets. The slopes show lots of variation, so it can be concluded that there is not a strong correlation between the average heat transfer coefficient of blocks of 12th and 13th with standard deviation, skewness and height of the first block.

2.3   Conclusion

Applications in urban canopies, vegetation canopies, and urban structures prompted us to conduct a generalized idealized study on the effect of height variation on total heat flux, total pressure drop, and surface heat transfer coefficient. A set of CFD simulations with 14 blocks, each of which may represent a building or canopy with random heights, were conducted. Using the experimental data available in the literature, the Reynold average two-equation SST k-omega model has been validated and used. Using the results of 204 simulations, the correlation between the standard deviation and skewness of the block heights and height of the first block on the total heat flux, total pressure drop, and surface heat transfer coefficient has been determined.

The results demonstrate that the standard deviation of heights and the average height of the blocks have the greatest influence on pressure drop, heat flux, and heat transfer. Skewness, however, has a small positive correlation. The three different regimes of isolated, wake interference, and skimming have all been observed, and a slight modification of the boundaries between the regimes is suggested for this specific flow configuration.

The study of simulations with identical statistics demonstrates that having identical means, standard deviation, skewness and height of the first block does not necessarily imply identical pressure drop, heat flux, and heat transfer and that the arrangement of heights is important. In this study, the height difference between the first two blocks was the most influential variable, followed by the height difference between the remaining adjacent blocks. Due to the nonlinearity of the height difference between the blocks, the concept of utilizing machine learning to capture this nonlinearity and test its ability to predict variables

such as the heat transfer coefficient has emerged which will be discussed in the next chapter.

References

[1]   Y. Zhang, Z. Gu, and C. W. Yu, "Review on numerical simulation of airflow and pollutant dispersion in urban street canyons under natural background wind condition," *Aerosol Air Qual. Res.*, vol. 18, no. 3, pp. 780–789, Mar. 2018.

[2]   Z. L. Gu, Y. W. Zhang, Y. Cheng, and S. C. Lee, "Effect of uneven building layout on air flow and pollutant dispersion in non-uniform street canyons," *Build. Environ.*, vol. 46, no. 12, pp. 2657–2665, Dec. 2011.

[3]   A. Dallman, S. Magnusson, R. Britter, L. Norford, D. Entekhabi, and H. J. S. Fernando, "Conditions for thermal circulation in urban street canyons," *Build. Environ.*, vol. 80, pp. 184–191, Oct. 2014.

[4]   J. Hang, Z. Luo, X. Wang, L. He, B. Wang, and W. Zhu, "The influence of street layouts and viaduct settings on daily carbon monoxide exposure and intake fraction in idealized urban canyons," *Environ. Pollut.*, vol. 220, no. Pt A, pp. 72–86, Jan. 2017.

[5]   Y. Lin *et al.*, "Investigation of the Reynolds number independence of cavity flow in 2D street canyons by wind tunnel experiments and numerical simulations," *Build. Environ.*, vol. 201, no. April, p. 107965, 2021.

[6]   J. Su, L. Wang, Z. Gu, M. Song, and Z. Cao, "Effects of real trees and their structure on pollutant dispersion and flow field in an idealized street canyon," *Atmos. Pollut. Res.*, vol. 10, no. 6, pp. 1699–1710, Nov. 2019.

[7]   R. X. Lee, S. K. Jusuf, and N. H. Wong, "The study of height variation on outdoor ventilation for Singapore's high-rise residential housing estates," *Int. J. Low-Carbon Technol.*, vol. 10, no. 1, pp. 15–33, Mar. 2015.

[8]   E. Johansson, J. Spangenberg, M. L. Gouvêa, and E. D. Freitas, "Scale-integrated atmospheric simulations to assess thermal comfort in different urban tissues in the warm humid summer of São Paulo, Brazil," *Urban Clim.*, vol. 6, pp. 24–43, Dec. 2013.

[9]   Y. Tominaga *et al.*, "AIJ guidelines for practical applications of CFD to pedestrian wind environment around buildings," *J. Wind Eng. Ind. Aerodyn.*, vol. 96, no. 10–11, pp. 1749–1761, Oct. 2008.

[10]   J. Franke, A. Hellsten, K. H. Schlünzen, and B. Carissimo, "The COST 732 Best Practice Guideline for CFD simulation of flows in the urban environment: A summary," *Int. J. Environ. Pollut.*, vol. 44, no. 1–4, pp. 419–427, 2011.

[11]   A. Hagishima and J. Tanimoto, "Field measurements for estimating the convective heat transfer coefficient at building surfaces," *Build. Environ.*, vol. 38, no. 7, pp. 873–881, Jul. 2003.

[12]   M. G. Emmel, M. O. Abadie, and N. Mendes, "New external convective heat transfer coefficient correlations for isolated low-rise buildings," *Energy Build.*, vol. 39, no. 3, pp. 335–342, Mar. 2007.

[13]   A. Townsend, *The structure of turbulent shear flow*. Cambridge university press, 1980.

[14]   Y. Lin *et al.*, "Investigation of the Reynolds number independence of cavity flow in 2D street canyons by wind tunnel experiments and numerical simulations," *Build. Environ.*, vol. 201, p. 107965, Aug. 2021.

[15]   B. Blocken, "LES over RANS in building simulation for outdoor and indoor applications: A foregone conclusion?," *Build. Simul. 2018 115*, vol. 11, no. 5, pp. 821–870, Jul. 2018.

[16]   B. Blocken, "Computational Fluid Dynamics for urban physics: Importance, scales, possibilities, limitations and ten tips and tricks towards accurate and reliable simulations," *Build. Environ.*, vol. 91, pp. 219–245, Sep. 2015.

[17]   Y. Toparlar, B. Blocken, B. Maiheu, and G. J. F. van Heijst, "A review on the CFD analysis of urban microclimate," *Renew. Sustain. Energy Rev.*, vol. 80, pp. 1613–1640, Dec. 2017.

[18]   V. Etminan, R. J. Lowe, and M. Ghisalberti, "A new model for predicting the drag exerted by vegetation canopies," *Water Resour. Res.*, vol. 53, no. 4, pp. 3179–3196, Apr. 2017.

[19]   R. Yoshie *et al.*, "Cooperative project for CFD prediction of pedestrian wind environment in the Architectural Institute of Japan," *J. Wind Eng. Ind. Aerodyn.*, vol. 95, no. 9–11, pp. 1551–1578, Oct. 2007.

[20]   K. Hanjalic, "Will RANS Survive LES? A View of Perspectives," *J. Fluids Eng.*, vol. 127, no. 5, pp. 831–839, Sep. 2005.

[21]   B. Blocken, T. Stathopoulos, and J. P. A. J. van Beeck, "Pedestrian-level wind conditions around buildings: Review of wind-tunnel and CFD techniques and their accuracy for wind comfort assessment," *Build. Environ.*, vol. 100, pp. 50–81, May 2016.

[22]   Y. Toparlar *et al.*, "CFD simulation and validation of urban microclimate: A case study for Bergpolder Zuid, Rotterdam," *Build. Environ.*, vol. 83, pp. 79–90, Jan. 2015.

[23]   S. Murakami, "Current status and future trends in computational wind engineering," *J. Wind Eng. Ind. Aerodyn.*, vol. 67–68, pp. 3–34, Apr. 1997.

[24]   Y. Tominaga and T. Stathopoulos, "CFD simulation of near-field pollutant dispersion in the urban environment: A review of current modeling techniques,"

Atmos. Environ., vol. 79, pp. 716–730, Nov. 2013.

[25] C. J. Baker, "Wind engineering—Past, present and future," *J. Wind Eng. Ind. Aerodyn.*, vol. 95, no. 9–11, pp. 843–870, Oct. 2007.

[26] F. R. Menter, "Zonal two equation κ-ω turbulence models for aerodynamic flows," *AIAA 23rd Fluid Dyn. Plasmadynamics, Lasers Conf. 1993*, 1993.

[27] H. Yu and J. Thé, "Validation and optimization of SST k-ω turbulence model for pollutant dispersion within a building array," *Atmos. Environ.*, vol. 145, pp. 225–238, Nov. 2016.

[28] E. Sagol, M. Reggio, and A. Ilinca, "Assessment of Two-Equation Turbulence Models and Validation of the Performance Characteristics of an Experimental Wind Turbine by CFD," *ISRN Mech. Eng.*, vol. 2012, pp. 1–10, Feb. 2012.

[29] A. L. J. Pang, M. Skote, and S. Y. Lim, "Modelling high Re flow around a 2D cylindrical bluff body using the k-ω (SST) turbulence model," *Prog. Comput. Fluid Dyn.*, vol. 16, no. 1, pp. 48–57, 2016.

[30] J. Shao, J. Liu, and J. Zhao, "Evaluation of various non-linear k–ε models for predicting wind flow around an isolated high-rise building within the surface boundary layer," *Build. Environ.*, vol. 57, pp. 145–155, Nov. 2012.

[31] Y. Tominaga, A. Mochida, S. Murakami, and S. Sawaki, "Comparison of various revised k–ε models and LES applied to flow around a high-rise building model with 1:1:2 shape placed within the surface boundary layer," *J. Wind Eng. Ind. Aerodyn.*, vol. 96, no. 4, pp. 389–411, Apr. 2008.

[32] S. Murakami, R. Ooka, A. Mochida, S. Yoshida, and S. Kim, "CFD analysis of wind climate from human scale to urban scale," *J. Wind Eng. Ind. Aerodyn.*, vol. 81, no. 1–3, pp. 57–81, May 1999.

[33] A. Mochida and I. Y. F. Lun, "Prediction of wind environment and thermal comfort at pedestrian level in urban area," *J. Wind Eng. Ind. Aerodyn.*, vol. 96, no. 10–11, pp. 1498–1527, Oct. 2008.

[34] F. Menter and T. Esch, "Elements of industrial heat transfer predictions," *16th Brazilian Congr. Mech. Eng.*, vol. 109, 2001.

[35] H. K. Versteeg and W. Malalasekera, *An introduction to computational fluid dynamics: the finite volume method*. Pearson education, 2007.

[36] F. R. Menter, M. Kuntz, and R. Langtry, "Ten years of industrial experience with the SST turbulence model," *Turbul. heat mass Transf.*, vol. 4, no. 1, pp. 625–632, 2003.

[37] P. A. C. Rocha, H. H. B. Rocha, F. O. M. Carneiro, M. E. Vieira da Silva, and A.

V. Bueno, "k–ω SST (shear stress transport) turbulence model calibration: A case study on a small scale horizontal axis wind turbine," *Energy*, vol. 65, pp. 412–418, Feb. 2014.

[38] E. R. Meinders, "Experimental study of heat transfer in turbulent flows over wall-mounted cubes.," 1998.

[39] B. A. Khan and A. K. Saha, "Direct numerical simulation of turbulent flow and heat transfer over a heated cube placed in a matrix of unheated cubes," *Int. J. Heat Mass Transf.*, vol. 171, p. 121052, Jun. 2021.

[40] A. Awol, "Numerical and Semi-Analytical Estimation of Convective Heat Transfer Coefficient for Buildings in an Urban-Like Setting," *Electron. Thesis Diss. Repos.*, Dec. 2019.

[41] H. Montazeri and B. Blocken, "Extension of generalized forced convective heat transfer coefficient expressions for isolated buildings taking into account oblique wind directions," *Build. Environ.*, vol. 140, pp. 194–208, Aug. 2018.

[42] J. Liu, J. Srebric, and N. Yu, "Numerical simulation of convective heat transfer coefficients at the external surfaces of building arrays immersed in a turbulent boundary layer," *Int. J. Heat Mass Transf.*, vol. 61, no. 1, pp. 209–225, Jun. 2013.

[43] A. Awol, G. T. Bitsuamlak, and F. Tariku, "Numerical estimation of the external convective heat transfer coefficient for buildings in an urban-like setting," *Build. Environ.*, vol. 169, p. 106557, Feb. 2020.

[44] T. R. Oke, "Street design and urban canopy layer climate," *Energy Build.*, vol. 11, no. 1–3, pp. 103–113, Mar. 1988.

[45] C. H. Chang and R. N. Meroney, "Concentration and flow distributions in urban street canyons: Wind tunnel and computational data," *J. Wind Eng. Ind. Aerodyn.*, vol. 91, no. 9, pp. 1141–1154, 2003.

Chapter 3

3   Machine learning

3.1   Dataset

Data is the lifeblood of machine learning, and our ability to construct successful models is dependent on the quality and amount of available data. The previous section describes a CFD analysis of 14 blocks with random heights and specified standard deviation, skewness, and total height average. In this section, the average surface heat transfer coefficient for each block has been determined using CFD simulations, and the height of 14 blocks and the mean surface heat transfer coefficient of the twelfth and thirteenth blocks were used to generate a data set to analyze with ML algorithms. The heights of 14 blocks are considered as input data, and the average surface heat transfer coefficient is the output of the ML model. Table 3-1 depicts the first ten simulations of the data set, which include a total of 204 simulations.

Table 3-1: Data set

Simulation number	Simulation 1	Simulation 2	Simulation 3	Simulation 4	Simulation 5	Simulation 6	Simulation 7	Simulation 8	Simulation 9	Simulation 10
H1	0.900	0.766	1.197	0.906	0.779	1.194	0.900	0.776	1.191	0.894
H2	0.900	1.026	0.803	1.180	0.870	1.075	1.119	0.903	0.961	1.029
H3	1.038	1.073	1.002	0.950	1.009	0.956	0.982	0.997	0.905	0.955
H4	0.998	1.115	0.995	1.067	0.934	1.007	0.947	1.099	0.904	1.094
H5	0.962	0.958	0.962	1.028	0.991	1.090	0.833	1.023	0.967	0.859
H6	1.031	1.056	1.024	1.048	0.939	1.011	1.037	1.016	0.967	0.930
H7	0.811	1.022	0.840	1.040	1.108	0.977	1.132	0.992	1.019	1.114
H8	0.995	1.019	1.054	0.929	0.995	0.951	1.065	0.954	1.068	0.946
H9	0.887	0.970	1.035	1.089	1.007	0.805	1.172	1.243	1.086	0.936
H10	1.124	0.958	1.097	1.058	1.124	0.869	0.902	0.978	1.134	1.166
H11	1.152	1.201	1.058	0.821	1.056	0.929	0.952	1.043	0.941	0.892
H12	1.049	0.968	0.943	1.009	0.962	1.053	0.937	1.026	0.818	1.128
H13	1.076	0.957	1.055	0.993	0.998	1.078	0.917	1.008	1.037	1.111
H14	1.107	0.935	0.965	0.848	1.182	1.064	1.099	0.951	0.996	0.948
SHTC AVG_12_13	0.901	0.851	0.944	0.903	0.900	0.965	0.928	0.918	0.925	0.895

Prior to training the machine learning models, the data was normalized using the MinMaxScaler, a function in the Scikit-learn package of Python [1], which scales and translates each feature independently to be between zero and one; for example, the entire height of the second block is scaled together. This scaling's advantages include its resistance to extremely tiny standard deviations of features and its ability to preserve zero entries in data. The scaled data is calculated as follows (equation 3-1):

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{3-1}$$

x_{max} represents the maximum value within the features while x_{min} presents the minimum. For example, for feature H3, which is the height of the third block, x_{max} is the maximum height of the third block, across all the simulations.

Random selections of this data were used to construct a model in order to generate probabilistic estimates of the surface heat transfer coefficient. In other words, the data set is ordered by decreasing standard deviation; to prevent the model from overfitting to this pattern, the data were shuffled. Note that the feature values in each row have been preserved while their positions have been randomized. The data is then separated into train and test sets. 90% of the data were assigned randomly to the training set, while the remaining 10% were assigned to the test set. On the chosen models, a grid search study with three-fold cross-validation [2] is conducted to determine the optimal hyperparameters for each model. The following describes the project's pipeline:
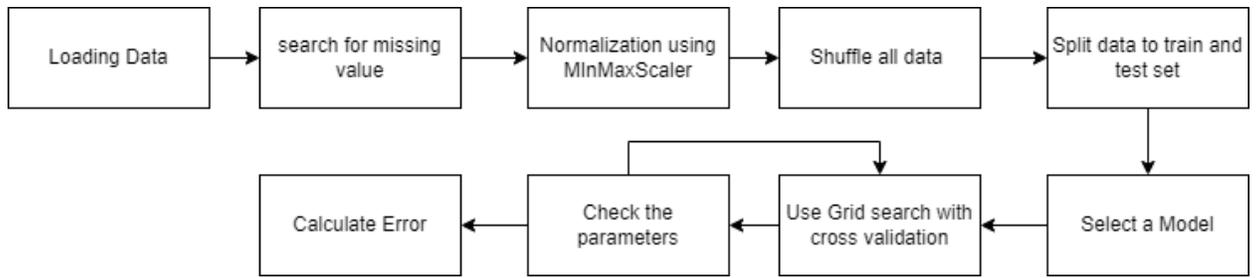
Figure 3-1: Pipeline of Project

3.2   Training Models

Several regression algorithms have been applied to the data using their default hyperparameters. The statistical model values, such as MSE, indicate the performance of the various models. The table below details the accuracy of each model in the training and testing sets. Similar levels of error between the train set and test set indicate a more general model. However, lower MSE values indicate a model with more accuracy. The RMSE and MSE of the train set and test set are shown for each regression algorithm in Table 3-2:

Table 3-2: Error of models in preliminary analysis to find the most promising model

Model	RMSE	MSE_Test	MSE_Train
Gradient Boosting Regressor	0.075	0.014	0.002
XGBoost Regressor	0.081	0.016	0.000
SVR	0.086	0.018	0.007
Bagging Regressor	0.087	0.018	0.004
Random Forest Regressor	0.092	0.020	0.003
AdaBoost Regressor	0.093	0.021	0.011
Ridge Regressor	0.096	0.022	0.019
Linear Regressor	0.099	0.024	0.018
Kneighbors Regressor	0.105	0.027	0.014

Based on the results of the preliminary studies, summarized in Table 3-2, a few models have been chosen to be optimized in order to obtain a more accurate prediction. The models

were selected based on the error and their capability to predict the data. Moreover, it is crucial to acquire the highest possible performance from the training data set when developing ML models for real issues. Consequently, optimizing the hyperparameters of ML algorithms is required during model training. Various methods have been proposed to automate the selection of hyperparameters to eliminate manual tuning, including grid search and random search hyperparameter optimization[3]. This work employed a grid research method with three-fold cross-validation to investigate all potential hyperparameter values inside a predefined domain. The number of cross-validation has been selected based on the computational power available for this project. The Hyperparameters of each of the selected models have been optimized to minimize MSE. The errors after hyperparameter tuning are as follows (Table 3-3):

Table 3-3: MSE after Hyperparameter tuning

Model	MSE on Train Set after hyperparameter tunning	MSE on Test Set after hyperparameter tunning
Random Forest Regression	0.013	0.019
Support Vector Regression	0.016	0.019
Gradient Boosting	0.009	0.019
XGBoost	0.002	0.015
Voting Regressor	0.009	0.018

In Random forest regression (RFR), the number of decision trees, the maximum depth of each tree, the minimum number of samples required to be at a leaf node, the minimum weighted fraction of the sum total of weights (of all the input samples) required to be at a leaf node, the maximum number of leaf nodes, and the maximum number of features were varied within the range specified in Table 3-4.

Table 3-4: Hyperparameter tuning range for Random Forest

Feature Name	Tuning Range	Tuned value
max_depth	2, 3, 4, 5, 6, 7	4
n_estimators	2, 10, 30, 40, 50, 100	10
min_samples_leaf	1, 2, 3, 4	3
min_weight_fraction_leaf	0.01, 0.02, 0.05, 0.1, 0.2, 0.3	0.01
max_feature	"auto", "log2", "sqrt"	"log2"
max_leaf_nodes	2, 5, 10, 20, 30	10

After hyperparameter tuning, the tuned parameter is shown in the third column of Table 3-4.

Table 3-5 summarizes the hyperparameters tuned for the Gradient boost model. They include the learning rate (Learning rate reduces the contribution of each tree by its value), the number of decision trees (n_estimators), the maximum depth of each tree, and the maximum number of features. The tuned parameter after hyperparameter tuning is listed in Table 3-5.

Table 3-5: Hyperparameter tuning range for Gradient Boost

Feature Name	Tuning Range	Tuned value
learning_rate	0.001, 0.01, 0.1	0.01
max_depth	4, 6, 8, 10, 12	8
max_features	1, 2, 3, 4, 5	5
n_estimators	10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120	70

Support vector regression requires generating a nonlinear curve using a kernel function to forecast outputs within an error margin. As shown in Table 3-6, the SVR hyperparameters that were adjusted were the error margin (epsilon), a regularization parameter (C) that

determines deviation weights, the kernel function, the degree of the polynomial kernel function (degree), and the kernel coefficient. Linear, polynomial, radial basis function (RBF), and sigmoid were the kernel functions that were considered for this study.

Table 3-6: Hyperparameter tuning range for Support Vector Regression

Feature Name	Tuning Range	Tuned value
kernel	"linear", "poly", "rbf", "sigmoid"	"rbf"
degree	3, 4, 5, 6	3
gamma	1,0.1,0.01,0.001	1
epsilon	0.005, 0.01, 0.05, 0.1, 0.5	0.1
C	0.1,1, 2, 10, 100	10

XGBoost is an effective method for constructing supervised regression models. As shown in Table 3-7, the variable XGBoost hyperparameters included the number of estimators, the maximum depth of each tree, the learning rate, step size shrinkage used in the update to prevent overfitting (eta), subsample ratio of the training instances, which aids in preventing overfitting, Minimum sum of instance weight required in a child (min child weight), and the booster type. The optimized hyperparameters are listed in Table 3-7.

Table 3-7: Hyperparameter tuning range for XGBoost

Feature Name	Tuning Range	Tuned value
max_depth	6, 9, 12	12
learning_rate	0.05, 0.1, 0.2	0.05
min_child_weight	2, 3, 4	3
subsample	0.9, 1, 1.1	0.9
booster	'gbtree', 'dart', 'gblinear'	'dart'
eta	0.05, 0.01, 0.005	0.05
n_estimators	5, 10, 15 ,20, 50, 100	50

VotingRegressor's purpose is to combine conceptually distinct machine learning regressors and deliver the average of their projected values. Such a regressor may be effective for balancing the individual shortcomings of a group of models with comparable performance. After combining the tuned versions of all models, the Voting Regressor was applied with the default values.

The plot of the Expected value versus the Predicted value of the average surface heat transfer coefficient of blocks 12 and 13 for the test set is shown in Figure 3-2:
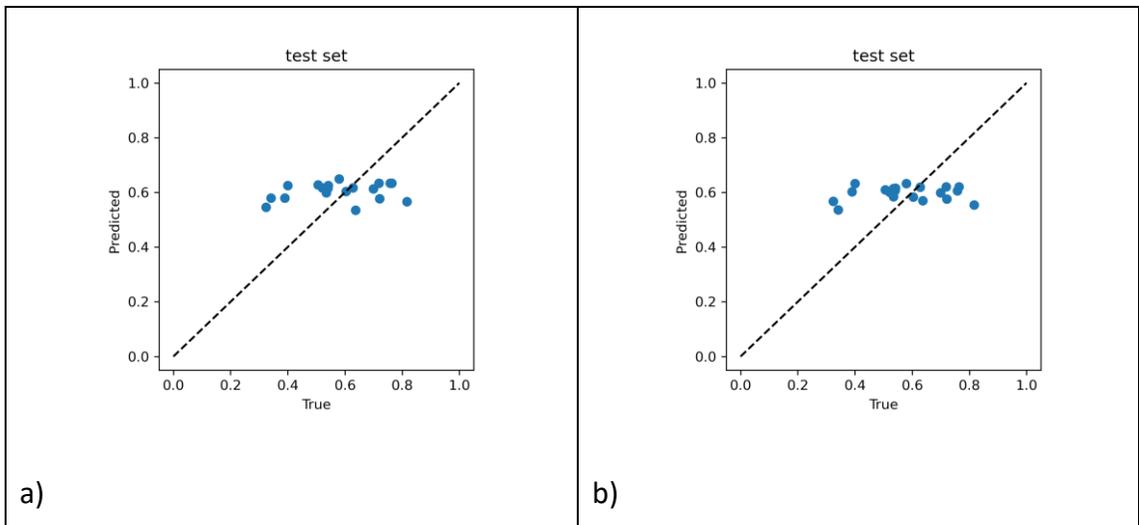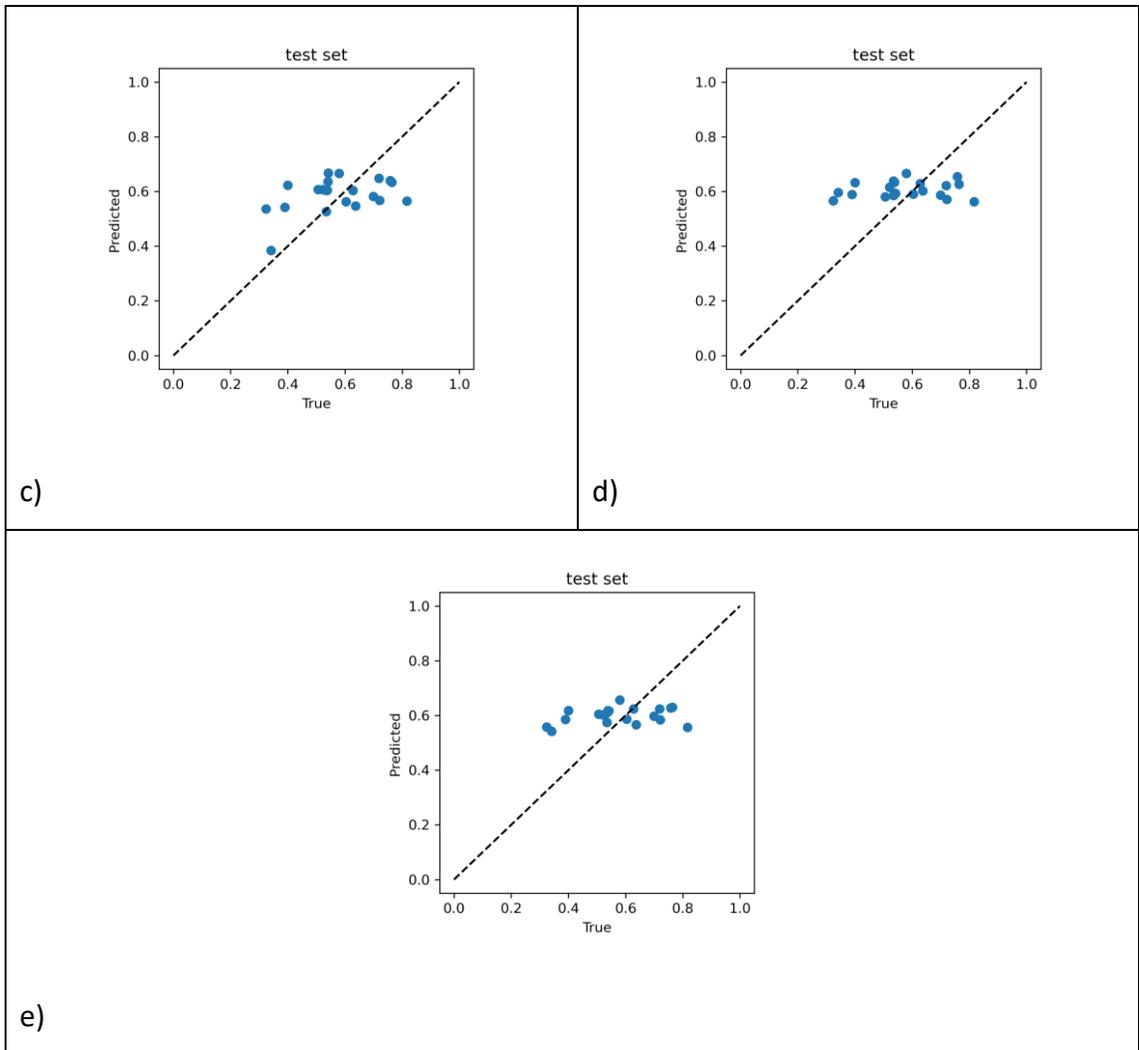
Figure 3-2: Predicted versus True value of test set in a) random forest b) gradient boost c) XGBoost d) SVR e) Voting regression for the set of all 14 heights as an input

Despite the fact that the model performance on the test set is unpromising, the model performance on the train set is strong, as depicted in Figure 3-3 for tuned XGBoost. This shows that the models are suffering from overfitting. Training machine learning models with insufficient data and many features can result in overfitting problems. Reducing the number of features and eliminating irrelevant features can aid in preventing model

overfitting. This motivates the work described in section 3.4. Note that some algorithms, like SVR, did not perform well on the input data.
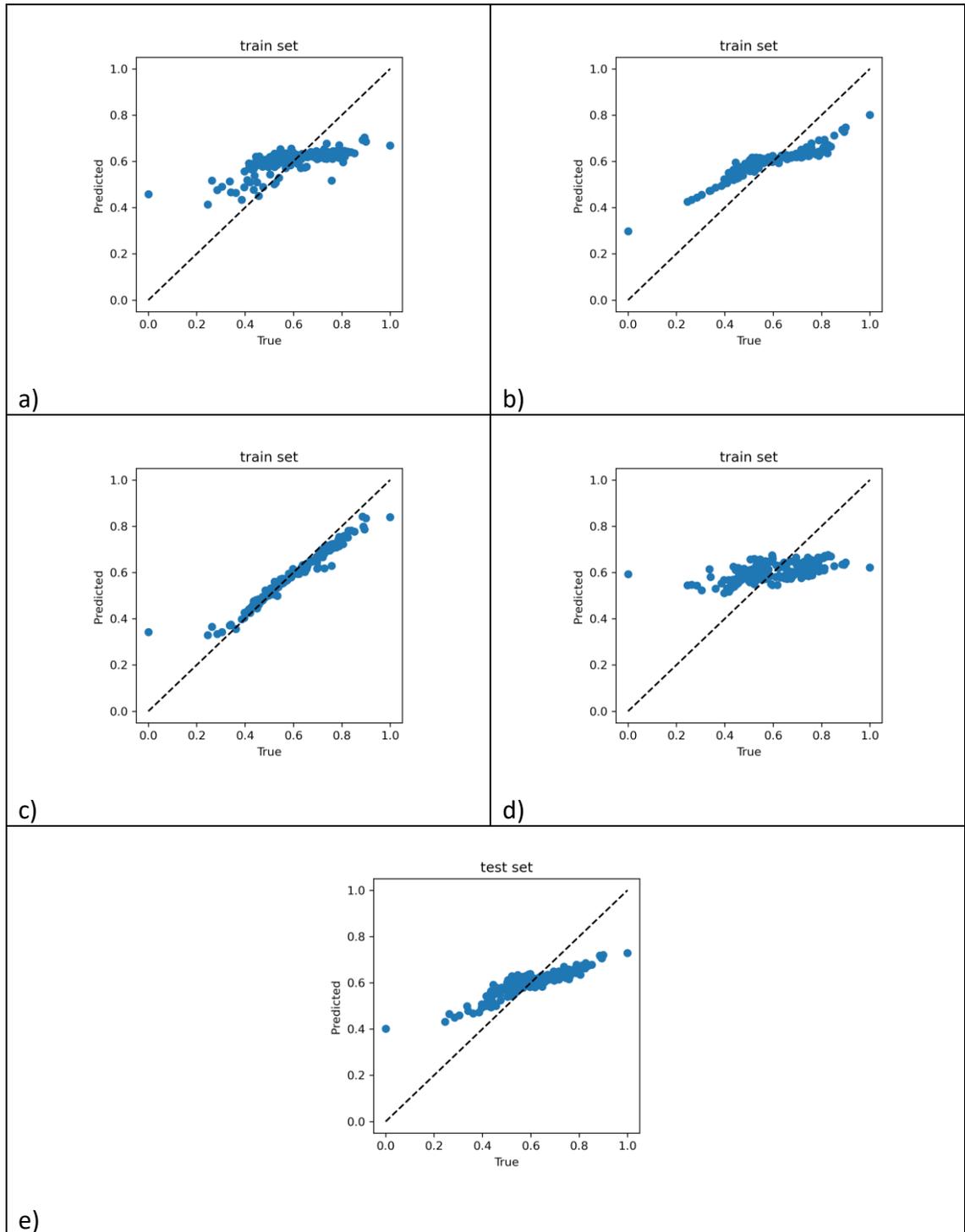
Figure 3-3: Predicted versus True value of train set in a) random forest b) gradient boost c) XGBoost d) SVR e) Voting regression for the set of all 14 heights as an input

3.3   Feature Importance Analysis

This study includes 14 distinct input features, which are the heights of 14 blocks in a single simulation. Including features that may have relatively minor effects on the target variable in training increases computation time and degrades the performance of the model. Consequently, finding relevant features using feature selection techniques is essential for reducing the dimensionality of data, eliminating irrelevant data, simplifying the resulting model, and accelerating the learning process [4]. Importance scores are one method for identifying the most relevant features. The relevance of an input parameter to a target variable is reflected by its importance score. Using the ".feature importances_" attribute, decision tree, and XGBoodt models from the Python scikit-learn module provide a straightforward approach to retrieve these scores. In order to reduce the number of features used as input and find the most important features, the feature important analysis is conducted on top of the dataset. The score of each feature corresponding to different models is shown in Figure 3-4.
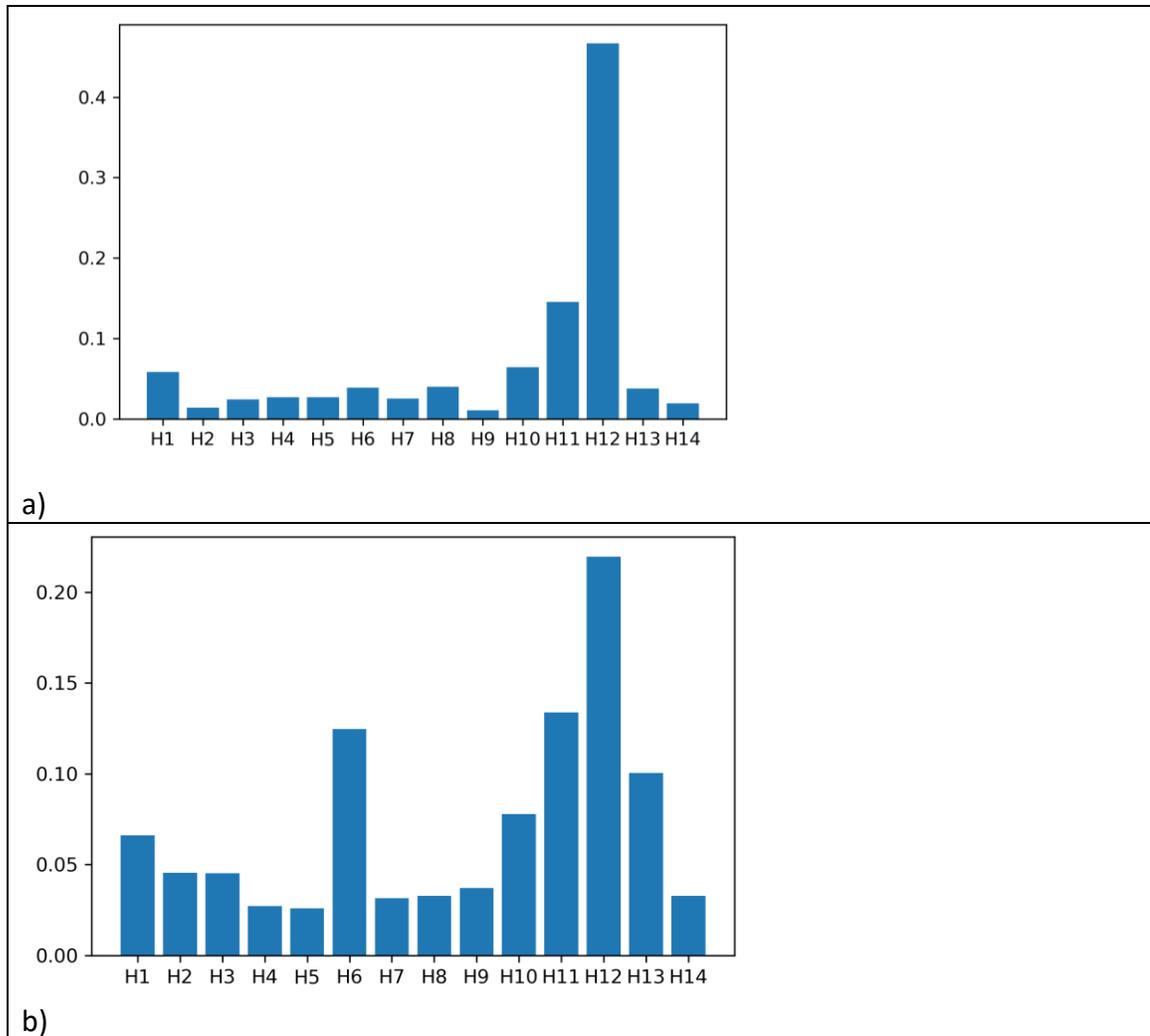
Figure 3-4: Feature importance analysis using a) Random Forest b) XGBoost

The goal of this machine learning model was to predict the average surface heat transfer coefficient of blocks 12 and 13. As expected from the physics of the problem, the heights of the blocks close to the target blocks are the most important. The height of the first block is also important because it has a significant effect on the initial development of the boundary layer. The score values of Xgboost and Random Forest are distinct due to the distinct model structures, and the value itself cannot be compared. In both cases, particularly for XGBoost, the height of the sixth block has a high score, which was unexpected.

3.4   Dropping irrelevant features

A modified data set is used to train the models to address the threat of overfitting caused by using all the features. Based on the result of the feature importance analysis, the six most important features (H1, H6, H10, H11, H12, H13) have been selected, and a training procedure have been done using a new modified dataset. The data set looks as follows (Table 3-8):

Table 3-8: Data set after dropping irrelevant features

Simulation number	H1	H6	H10	H11	H12	H13	SHTC AVG_12_13
Simulation 1	0.900	1.031	1.124	1.152	1.049	1.076	0.901
Simulation 2	0.766	1.056	0.958	1.201	0.968	0.957	0.851
Simulation 3	1.197	1.024	1.097	1.058	0.943	1.055	0.944
Simulation 4	0.906	1.048	1.058	0.821	1.009	0.993	0.903
Simulation 5	0.779	0.939	1.124	1.056	0.962	0.998	0.900
Simulation 6	1.194	1.011	0.869	0.929	1.053	1.078	0.965
Simulation 7	0.900	1.037	0.902	0.952	0.937	0.917	0.928
Simulation 8	0.776	1.016	0.978	1.043	1.026	1.008	0.918
Simulation 9	1.191	0.967	1.134	0.941	0.818	1.037	0.925
Simulation 10	0.894	0.930	1.166	0.892	1.128	1.111	0.895

A preliminary analysis has been conducted to determine the most promising models on which to perform hyperparameter tuning, and the error of each model is shown in Table 3-9:

Table 3-9: Error of models in preliminary analysis to find the most promising model using modified dataset after dropping irrelevant features

Model	Test Set		Train Set
	RMSE	MSE	MSE
SVR	0.111	0.121	0.029
AdaBoost Regressor	0.110	0.122	0.035
KNeighbors Regressor	0.112	0.126	0.042
Gradient Boosting Regressor	0.115	0.138	0.006
XGBoost Regressor	0.117	0.140	0.000
Random Forest Regressor	0.118	0.143	0.007
Bagging Regressor	0.120	0.146	0.011
Ridge Regresion	0.125	0.149	0.056
Linear Regression	0.128	0.155	0.055

Five distinct models were selected from a preliminary analysis of models in order to perform hyperparameter tuning. The error of each model after hyperparameter tuning is shown in Table 3-10:

Table 3-10: Error of selected models using modified dataset after dropping irrelevant features

Model	MSE Test Set	MSE Train Set
SVR	0.036	0.011
Random Forest Regressor	0.040	0.011
XGBoost Regressor	0.035	0.003
Kneighbors Regressor	0.034	0.016
Gradient Boosting Regressor	0.035	0.002
Voting Regressor	0.034	0.007

The tuned hyperparameter for each model is as follows (Table 3-11):

Table 3-11: Value of tuned hyperparameter of selected models using modified dataset after dropping irrelevant features

Model	Hyperparameter
SVR	C = 2
	degree = 1
	epsilon = 0.1
	gamma =1
	kernel = "rbf"
Random Forest Regressor	max_depth = 5
	max_features = log2
	max_leaf_nodes = 12
	min_samples_leaf = 1
	min_weight_fraction_leaf = 0.05
	n_estimators = 12
XGBoost Regressor	booster = gbtree
	eta = 0.05
	learning_rate = 0.05
	max_depth = 7
	min_child_weight = 1
	n_estimators = 50
	subsample = 0.9
KNeighbors Regressor	
	algorithm = 'auto'
	leaf_size = 1

	n_neighbors = 14
	p = 1
Gradient Boosting Regressor	learning_rate = 0.2
	max_depth = 10
	max_features = 2
	n_estimators = 8

The results show that reducing the number of features does not improve accuracy. This is likely due to the small size of the data set. The true value versus the predicted value with the model is illustrated in Figure 3-5.

a)

b)

c)
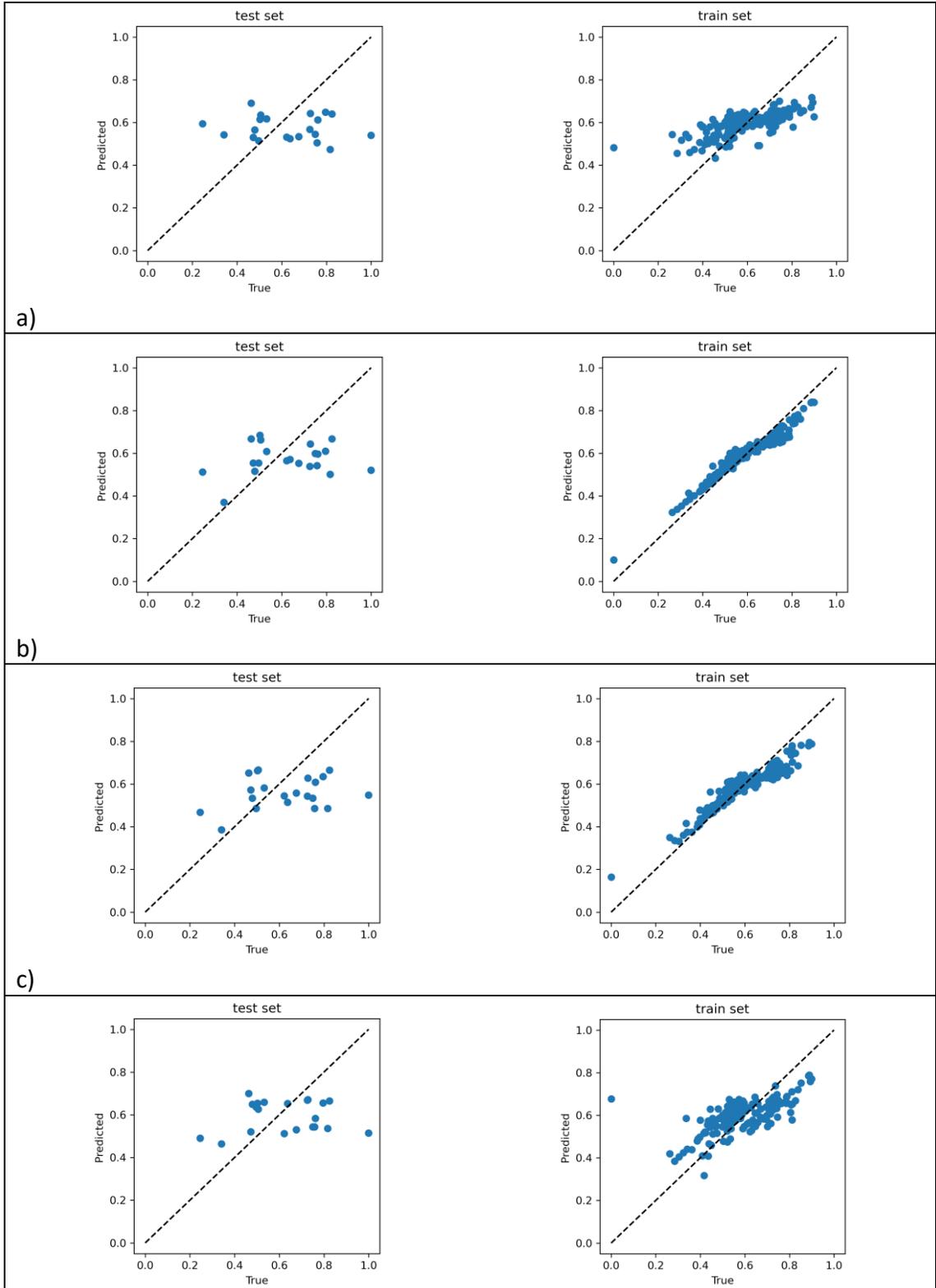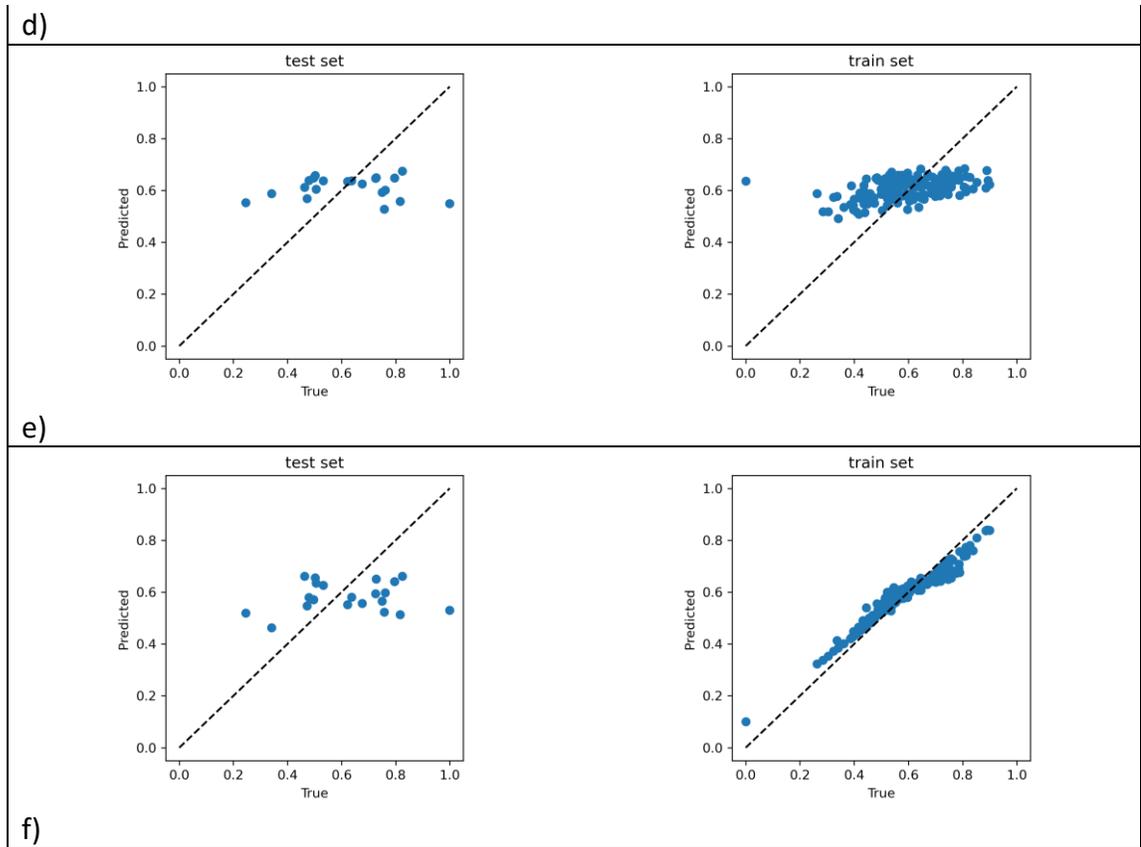
Figure 3-5: Predicted versus True value of test and train set in a) random forest b) gradient boost c) XGBoost d) SVR e) KNeighbors f) Voting regression for the set of all 14 heights as an input

Although the accuracy of the majority of models improved when compared to models that did not eliminate less important features from the data set, the models performed poorly on the test set, indicating that they are not generalizable and do not perform well with unseen data. This may be due to the small sample size. This motivates the work described in section3.5, in which a model predicts the surface heat transfer coefficient of a target building using only the heights of adjacent blocks.

3.5   Sets of 5 blocks data set

The results of feature importance showed that the height of 2-3 blocks before the target block has the greatest effect on the target value prediction. Therefore, to solve the problem of the limited data set, we examine sets of 5 blocks, excluding the first three blocks in each simulation. In other words, as illustrated in Figure 3-6, sets of 5 blocks are considered, and

the surface heat transfer coefficient of the fourth block is calculated. Multiple unique sets of five can be considered from each simulation, as shown in Figure 3-6. This approach is also consistent with the physics and goal of the problem: heat loss from a single building is important when designing a building, and the heat loss is expected to depend most strongly on the immediate surroundings.
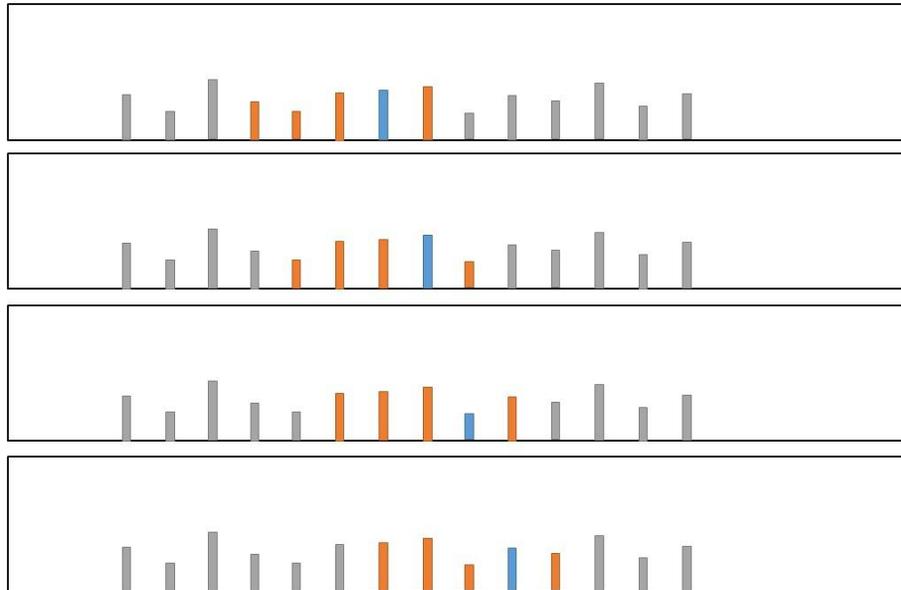
Figure 3-6: schematic of considering sets of five blocks in each simulation (The surface heat transfer coefficient of the fourth block in each set of five blocks is calculated.)

Using this strategy, the database grows substantially in size to include 1360 datasets. 90% of the data was used to train the model, while 10% was used to validate it. Several regression algorithms were trained on the dataset without hyperparameter tuning in order to identify the most promising ones. The MSE and RMSE of each model's test set are calculated (Table 3-12).

Table 3-12: Error of models in preliminary analysis to find the most promising model using sets of 5 blocks dataset

Model	RMSE	MSE
Linear Regression	0.075	0.048
Ridge Regression	0.075	0.048
Random Forest Regression	0.066	0.039
AdaBoost Regression	0.072	0.043
Bagging Regression	0.068	0.041
SVR	0.064	0.036
Kneighbors Regression	0.067	0.040
XGBoost Regression	0.072	0.046
Gradient Boosting Regression	0.067	0.039

Based on the result of the preliminary analysis, SVR, Random Forest, XGBoost, and Gradient Boost have been selected to be tuned on the new data set. The errors after hyperparameter tuning are as follows (Table 3-13):

Table 3-13: Error of selected models after hyperparameter tunning using sets of five blocs dataset

Model	RMSE	MSE
Random Forest Regression	0.107	0.011
Support Vector Regression	0.105	0.011
Gradient Boosting	0.103	0.011
XGBoost	0.105	0.011
Voting Regressor	0.103	0.011

The tunned hyperparameters for each model are in Table 3-14:

Table 3-14: Value of tuned hyperparameter of selected models using sets of 5 blocks dataset

Model	HyperParmeter
SVR	C = 10
	degree = 3
	epsilon = 0.1
	gamma = 1
	kernel = 'rbf'
Random Forest Regressor	max_depth = 7
	max_features = 'auto'
	max_leaf_nodes = 20
	min_samples_leaf = 1
	min_weight_fraction_leaf = 0.005
XGBoost Regressor	booster = 'gbtree'
	eta = 0.05
	learning_rate = 0.05
	max_depth = 6
	min_child_weight = 2
	subsample = 0.9
Gradient Boosting Regressor	learning_rate = 0.01
	max_depth = 6
	max_features = 2
	n_estimators = 240

The prediction versus the true value of each model is illustrated in Figure 3-7:
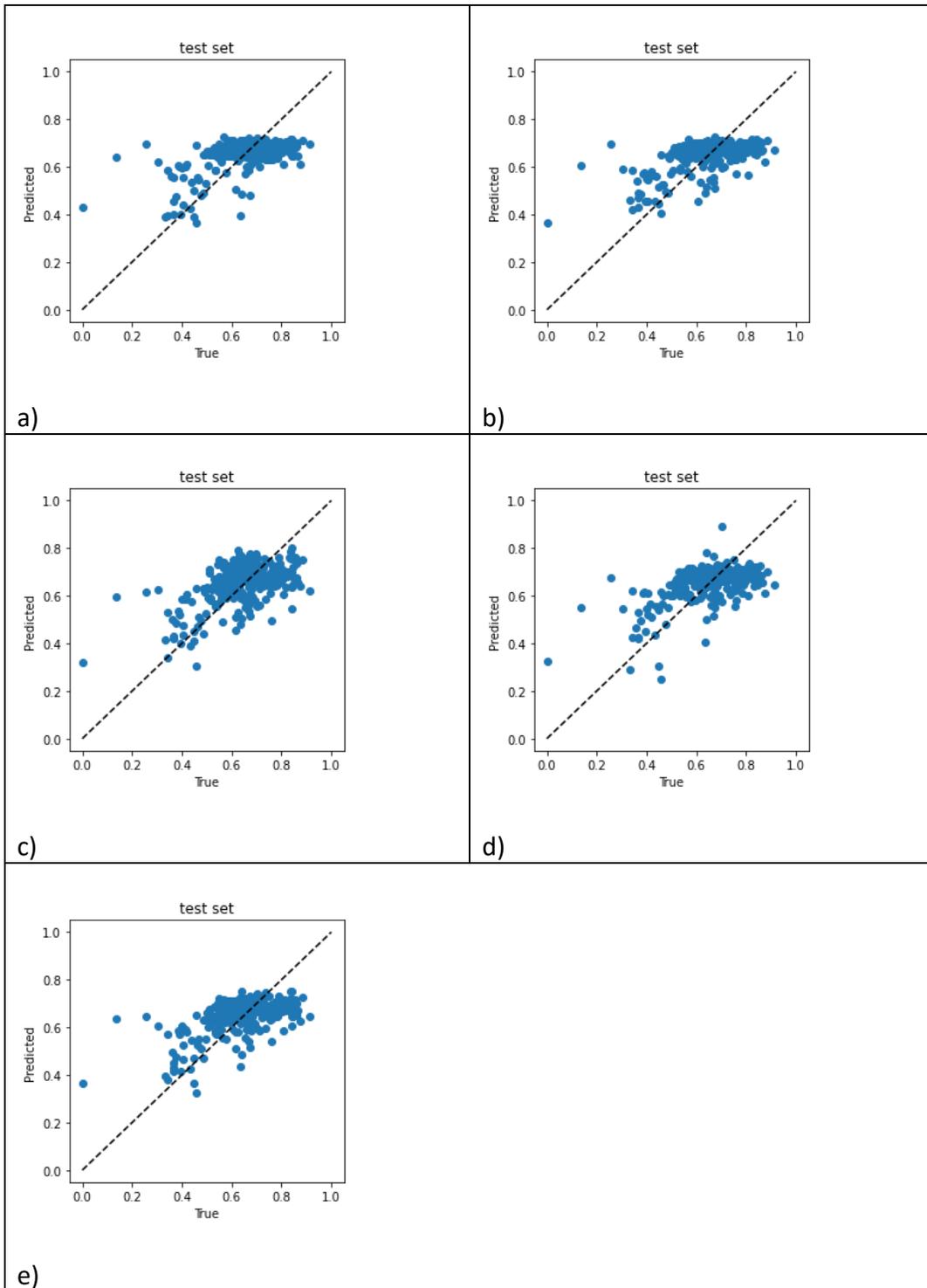
Figure 3-7: Predicted versus True value in a) random forest b) gradient boost c) XGBoost d) SVR e) Voting regression

Even though there is substantial non-linearity in the new data set, the model predictions, particularly XGBoost, are reasonable. Non-linearity in new data sets occurs due to the fact that we only consider a small portion of the simulation, despite knowing that the surrounding environment will affect the outcome. For instance, as depicted in Figure 3-8, there may be a situation in which the heights of five adjacent blocks are identical, but the heights of the surrounding blocks are different in each simulation, resulting in different outcomes. In this case, the ML algorithm will provide one output for the set of input values, and this output cannot match the correct value for all the cases illustrated in Figure 3-8.
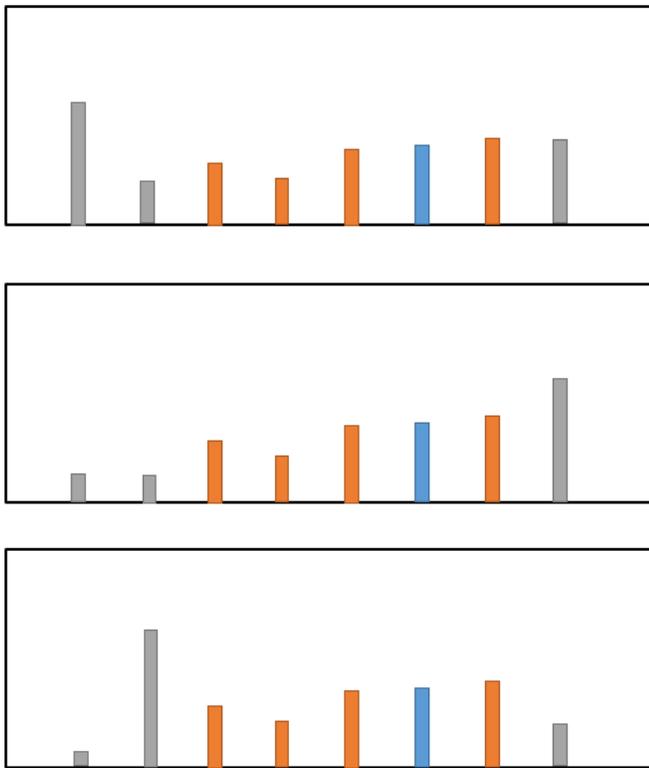
Figure 3-8: Example of possible non-linearity in data set of 5 blocks

The sets of 5 blocks have undergone a feature importance analysis for this approach, where the objective is to determine the surface heat transfer coefficient of the fourth block, and the relative importance of each block height in the input set is illustrated in Figure 3-8. As anticipated, the height of the block just upstream of the target block in the direction of the wind has the greatest impact on the heat transfer, followed by the height of the target block

itself. The importance of H2 and H5 is one-quarter that of the height of the block preceding the target block. In both cases, the feature important analysis of et of all the heights in the simulation and sets of five heights, the height of the block just upstream of the target building and the height of the target building are among the most important features in each simulation. Since the upstream block has a shielding effect on the target block, we expect the upstream block to have an effect on the heat transfer. The downstream block also affects the heat transfer because it affects the wake region structure.
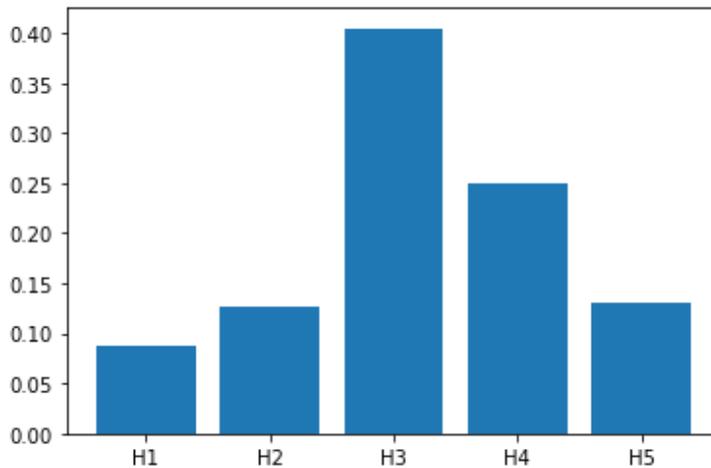
Figure 3-9: Feature importance analysis result on the sets of 5 blocks

For estimating the heat transfer coefficient of a building, it is not necessary to model the entire city; only the heights of a few nearby buildings are required to have a more accurate prediction.

3.6   Comparing model accuracy in different data sets

In this project, three distinct data sets were utilized. The first uses all of the data and features generated by simulation. To mitigate the problem of overfitting in models trained with the first data set, a second data set has been defined. The second set is created by excluding the less important features from the first dataset. Removing less relevant features was ineffective because the data set was too small, so a new data set was generated in order to increase the training set using the same number of simulations as previously. In this data

set, the sets of five blocks in each simulation have been analyzed; multiple sets from each simulation can be included. The error of the different tuned models is shown in Table 3-15.

Table 3-15: Error of different data sets after hyperparameter tuning (test set error)

	MSE		
Model	**All data**	**Data set after dropping less important features**	**sets of 5**
Random Forest Regression	0.019	0.035	0.011
Support Vector Regression	0.019	0.018	0.011
Gradient Boosting	0.019	0.032	0.011
XGBoost	0.015	0.038	0.011
Voting Regressor	0.018	0.033	0.011

In the majority of algorithms, dropping import features does not help reduce error or generate a more robust model. However, the model is improved by examining the sets of 5 blocks; despite the fact that there will be a great deal of non-linearity in the data because we do not capture the entire simulation and are only considering a portion of each simulation, the improvement is due to the larger data set. It is also a practical approach because the prediction is made with less information about the building's surroundings. To enhance the accuracy of the prediction, a deep learning analysis will be performed on the data.

3.7   Neural Network

Using the approach of only considering a few surrounding blocks to estimate the surface heat transfer coefficient, the number of data sets has exploded to 1360. With this quantity of data, we can employ more advanced machine learning algorithms, such as deep learning and particularly neural networks. The objective of using a Neural network is to generate a

model that is more accurate and general than the models generated in the preceding section in order to test the effectiveness of deep learning in estimating the heat transfer coefficient using a subset of the available data.

In ANN models, information propagation occurs through connections that accept information from a processing element (neuron) and transmit it to the subsequent neurons. Each piece of information is modified by a weight that reflects the importance of input variables to each layer's outputs. The pipeline for this project is depicted in the figure below. It resembles the pipeline described previously with a distinct optimization process.
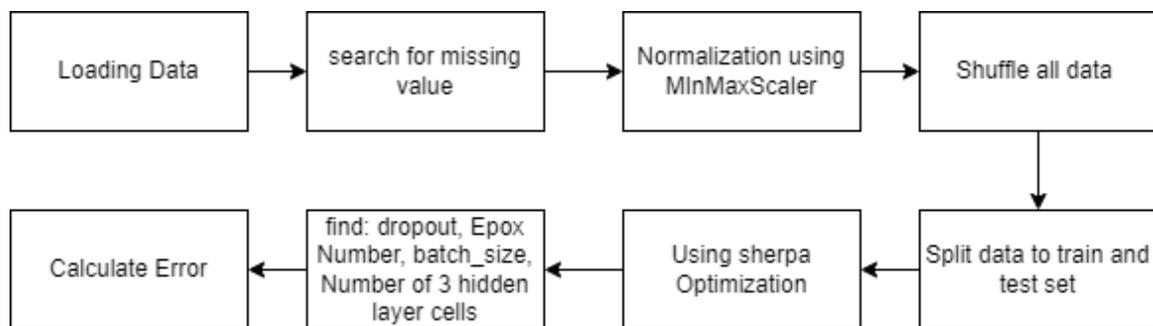
Figure 3-10: Pipeline of deep learning analysis

The Python hyperparameter optimization library SHERPA has been utilized for neural network hyperparameter optimization[5]. In this project, the number of neurons in each hidden layer, the dropout rate, the learning rate, the Epoch number, and the batch size were optimized to achieve the lowest MSE between the predicted and actual values using the SHERPA library. To do this, we use a random search optimization with a maximum of one thousand iterations for each neural network (NN). The default parameters for the NN models are determined by the values that result in the lowest loss. The ReLu activation function is used to activate all of the NN models[6]. The last hidden layer is a fully connected layer with a dropout. The Adam optimizer technique was chosen to train the NN model in order to minimize the MSE, the chosen loss function, between observed and predicted values. Only one node is present in the output layer, which represents the surface heat transfer coefficient of the target block. The implementation of NN model types is performed using the Python package Keras[7], which employs the Python package Tensorflow[8] as a backend. Compared to deep neural networks, a shallow network has a

more straightforward structure. Simple designs with few hidden layers and neurons are computationally cheaper and less susceptible to overfitting than complicated layouts. Increasing the number of hidden layers will dramatically increase the number of network connections and, as a result, the number of weights and biases that must be tuned. In this study, three distinct network architectures with 2, 3, and 4 hidden layers were investigated and optimized.

3.7.1   Neural network with two hidden layers

It is vital to investigate the model setup and tune its hyperparameters in order to comprehend the performance of the applied method for a particular task. Hyperparameter optimization has been done for the model with two hidden layers having a learning rate ranging from 0.01 to 1 and a number of neurons ranging from 0 to 512. Using the Sherpa package, the optimized parameters are a learning rate of 0.01 and the first layer of 128 neurons and the second layer of 256 neurons. The MSE of the test set is 0.0091, whereas that of the train set is 0.0090. Figure 3-11 compares the actual and predicted outcomes of test and training data.
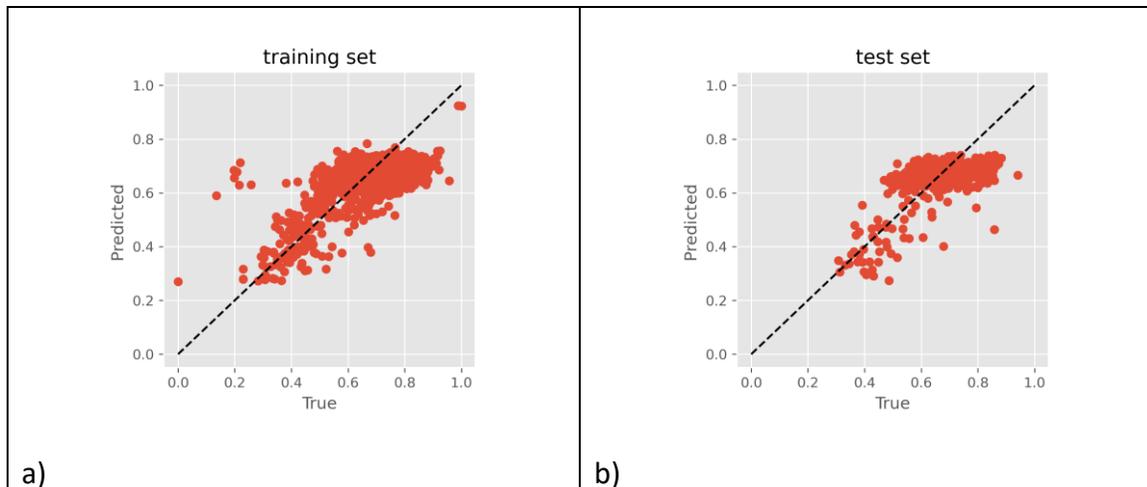
Figure 3-11: Actual vs. predicted values for a) training and b) testing set in 2 hidden layers deep learning model.

3.7.2   Neural network with three hidden layers

For a model with three hidden layers, hyperparameter tuning has been conducted. This resulted in the following hyperparameters: a dropout value of 0 (no dropout has been

suggested), batch size of 16, epochs number of 250, and the number of neurons in hidden layers are 20, 411, and 37. The MSE value of the test set is 0.0088 and for the train set is 0.00876. Figure 3-12 shows the actual versus the predicted result of the test and train data for the three hidden layer models.
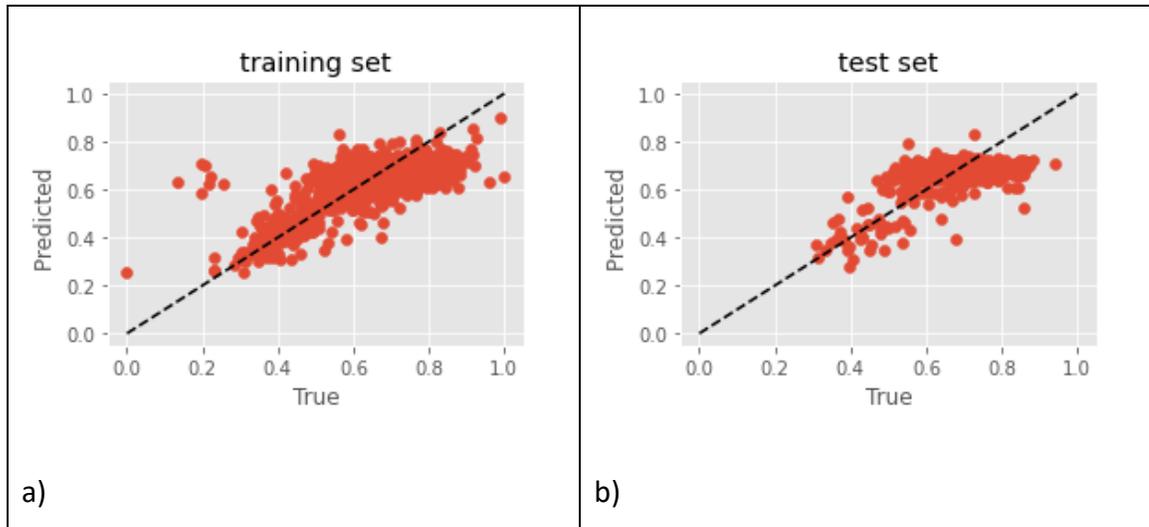
Figure 3-12: Actual vs. predicted values for a) training and b) testing set in 3 hidden layers deep learning model.

3.7.3   Neural network with four hidden layers

The model with four hidden layers has also been studied. The tuned hyperparameters are dropout of 0.1 prior to the output layer, and 352, 133, 97, and 127 neurons are in each hidden layer. The batch size is 16, the Epochs number is 81, and the validation split ratio is set to 0.2. The MSE of train error is 0.0090, and for the test set is 0.0093. Figure 3-13 shows the actual versus the predicted result of the test and train data, which illustrates the inaccuracy of the model while it has accepted error.
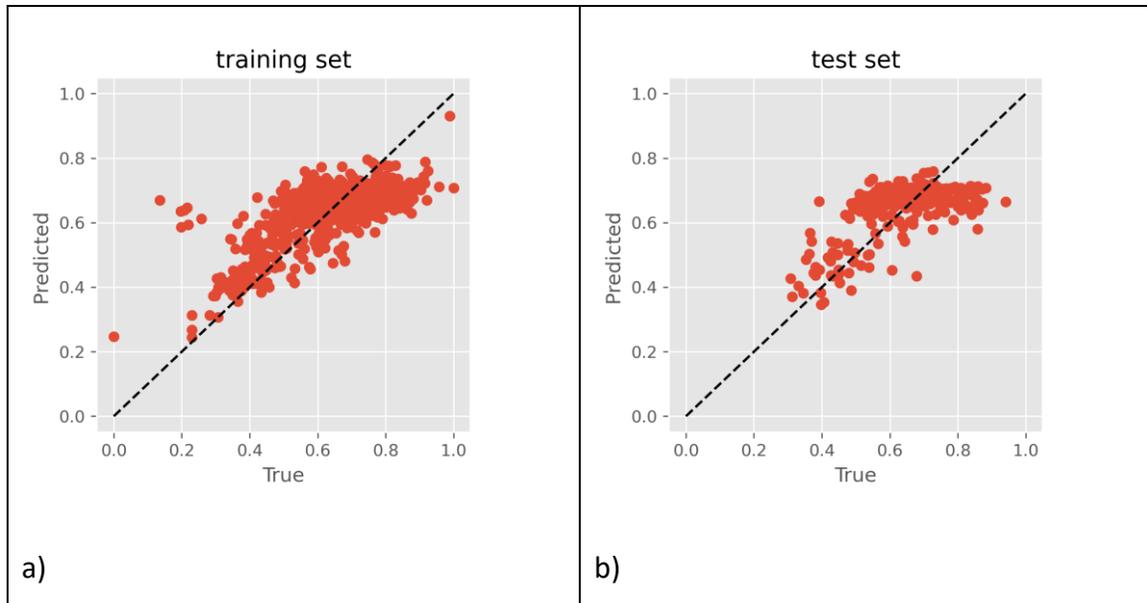
Figure 3-13: Actual vs. predicted values for a) training and b) testing set in 4 hidden layers deep learning model.

Although the model performs well on both train and test sets, the regularizer term has not been studied. In order to determine whether regularizer terms improve accuracy, a set of hyperparameters were manually chosen; the resulting values are shown in Table 3-16.

Table 3-16: selected hyperparameters and their accuracy of neural network with four hidden layers

Layer1	Layer2	Layer3	Layer4	DropOut	Regularizer	Epochs	Test Error	Train Error
352	133	97	127	0.2	l2	250	0.008	0.008
352	133	97	127	0.25	l2	250	0.009	0.008
32	64	64	32	0.2	l2	200	0.009	0.010
32	64	64	32	0.3	none	250	0.009	0.009
32	64	64	32	0.1	l2	200	0.009	0.010
32	64	64	32	0.3	l2	250	0.009	0.010
352	133	97	127	0.25	l2	150	0.009	0.009
352	133	97	127	0.1	l2	270	0.009	0.008
256	128	64	32	0.2	none	168	0.009	0.008
32	64	64	32	0.2	none	200	0.009	0.010

512	256	64	128	0.2	none	150	0.009	0.008
256	128	64	8	0.2	l2	350	0.009	0.009
512	256	64	128	0.3	none	100	0.009	0.008
352	133	97	127	0.1	l1_l2	250	0.009	0.011
32	64	64	32	0.4	none	200	0.010	0.010
32	64	64	32	0.4	l2	200	0.010	0.011
256	128	64	32	0.3	l2	500	0.010	0.007
256	128	64	8	0.2	none	350	0.010	0.007
352	133	97	127	0.25	none	1200	0.011	0.004
352	133	97	127	0.25	l2	1200	0.011	0.005
256	128	64	32	0.2	l2	1200	0.011	0.006
256	128	64	32	0.3	none	500	0.011	0.006
512	256	64	128	0.2	none	1200	0.012	0.003

By increasing the dropout value and adding a l2 regularizer term, the accuracy is improved. The predicted value versus the actual value is shown in Figure 3-14.
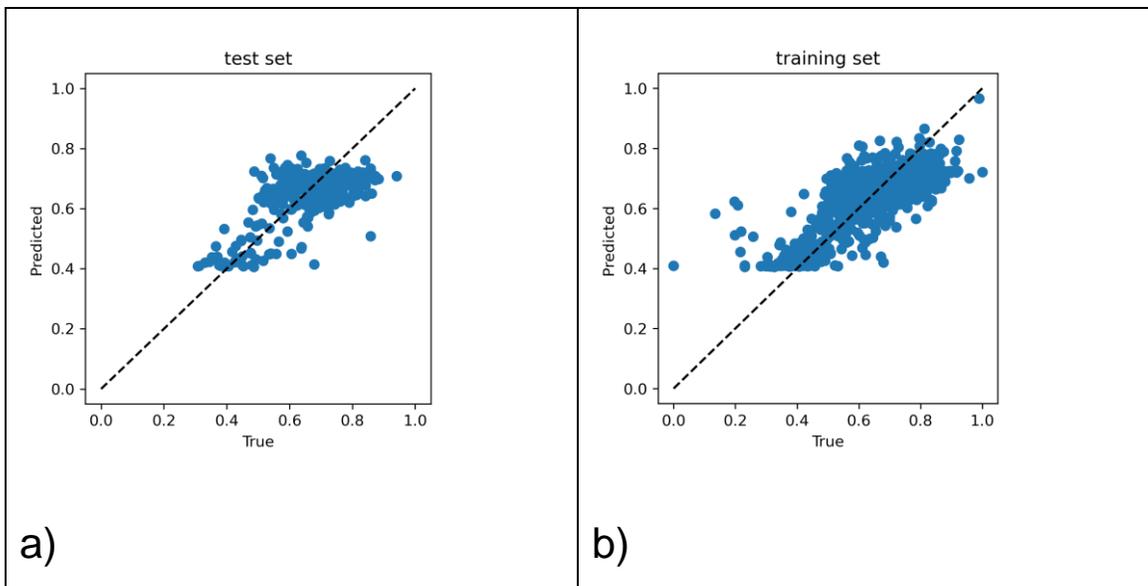
Figure 3-14: Predicted versus true value in a) test b) train set using 4 layers neural network with drop out of 0.2 and l2 regularizer

3.8   Conclusion

This study investigates the application of cutting-edge machine learning models to examine machine learning's capacity to forecast the surface heat transfer coefficient of a building. The results of 204 distinct simulations with the same domain and varied block sizes were used for this purpose. Several advanced machine learning regression models, including Random forest, SVR, XGBoost, Gradient Boosting, and deep learning, were tuned, trained, and evaluated on the dataset. K-fold cross-validation was employed during the tuning phase to ensure that generalized models were developed. The results demonstrate that the models accurately represented the underlying principles that contribute to the surface heat transfer coefficient of buildings. In addition, the diversity of the algorithms utilized in this study demonstrates the robustness of ML algorithms for data analysis despite the dataset's complexity and strong non-linearity. The prediction accuracy for the training set utilizing the entire dataset was quite high, whereas it was very poor for the testing set. For certain models, the inaccuracy of the test set was double that of the train set. This indicates that models for this dataset suffer from overfitting and lack of generalization to new, unknown data. Subsets of five blocks were then considered to increase the data set size. For this larger dataset, the three distinct NN models and machine learning models described previously were implemented. The root means squared error (RMSE) and mean square error (MSE) are tracked to evaluate each model's performance. The deep learning model's performance was superior to that of the machine learning model. The difference between the statistical errors of train and test data in deep learning models was less than in conventional machine learning models. Similar sized errors indicate good performance; there is a limit to how small the errors can get due to the non-linearity illustrated in Figure 3-8.

Moreover, In urban locations, the wind flow around a building is commonly affected by adjacent buildings owing to buffeting, channelling, and sheltering [9]–[11]. Tominaga[12] suggests modelling the surrounding buildings within a 1H-2H radius of the target building, plus at least one additional block in each direction to completely consider the effect of the surrounding. Tong [13]suggested explicitly modelling three layers of buildings in the surroundings (n=3) for regular street canyons (H/W=1, where H is the building height on the street side and W is the street width) under normal and oblique wind directions while

reducing the influence region to two layers (n=2) for wide canyons (H/W=1/3) and a high-rise configuration. These recommendations are consistent with the findings of this study's feature importance analysis, which indicate we must consider the height of two buildings in the wind direction before and after the target building. This result is very encouraging as it allows engineers to accurately estimate the surface heat transfer of each block while considering the surrounding environment with minimal computational costs. This work also demonstrates that for specified environmental conditions, ML could be used to estimate the surface heat transfer coefficient of a building based on the geometry of a few of the surrounding buildings.

References

[1]   F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[2]   T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning," 2009.

[3]   J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, 2012.

[4]   B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A Survey on Evolutionary Computation Approaches to Feature Selection," *IEEE Trans. Evol. Comput.*, vol. 20, no. 4, pp. 606–626, Aug. 2016.

[5]   L. Hertel, J. Collado, P. Sadowski, J. Ott, and P. Baldi, "Sherpa: Robust hyperparameter optimization for machine learning," *SoftwareX*, vol. 12, p. 100591, Jul. 2020.

[6]   M. A. Wani, F. A. Bhat, S. Afzal, and A. I. Khan, *Advances in Deep Learning*, vol. 57. Singapore: Springer Singapore, 2020.

[7]   F. Chollet, "Keras," *GitHub*. 2015.

[8]   M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," *Proc. 12th USENIX Symp. Oper. Syst. Des. Implementation, OSDI 2016*, pp. 265–283, May 2016.

[9]   T. G. Mara, T. C. E. Ho, and N. Isyumov, "Interference and Influence of nearby buildings: A discussion of the design approach," *Struct. Congr. 2012* , pp. 798–808, 2012.

[10]   S. Liu *et al.*, "Influence of surrounding buildings on wind flow around a building predicted by CFD simulations," *Build. Environ.*, vol. 140, pp. 1–10, 2018.

[11]   O. S. Asfour, "Prediction of wind environment in different grouping patterns of housing blocks," *Energy Build.*, vol. 42, no. 11, pp. 2061–2069, Nov. 2010.

[12]   Y. Tominaga *et al.*, "AIJ guidelines for practical applications of CFD to pedestrian wind environment around buildings," *J. Wind Eng. Ind. Aerodyn.*, vol. 96, no. 10–11, pp. 1749–1761, Oct. 2008.

[13]   Z. Tong, Y. Chen, and A. Malkawi, "Defining the Influence Region in neighborhood-scale CFD simulations for natural ventilation design," *Appl. Energy*, vol. 182, pp. 625–633, Nov. 2016.

Chapter 4

4   Conclusion

Several environmental flows, such as flow over a series of buildings and flow over a vegetation canopy, can be idealized as flow over an array of blocks. The blocks present a stiff barrier in the flow, which leads to the formation of a boundary layer above and around the obstruction. Due to the extensive industrial applications and environmental impacts of barriers on flow behaviour, it is vital to grasp the flow near obstructions. This block may represent an urban building; studying the flow around urban buildings can aid in the comprehension of pollution dissipation and the creation of more energy-efficient structures. Especially in the current era, when global warming poses a significant threat to humankind, any action that can conserve energy is crucial. On the other hand, this block may represent a submerged canopy, which has an effect on coastal habitats' shielding, sediment production and erosion, and carbon sequestration. In this research, a 2D idealized simulation was undertaken to examine the influence of height variation, and more specifically, the statistics of height variation, inside a single cluster of obstacles. Two hundred four simulations were performed for random block heights with a mean height value of 1 and standard deviation values of 0.1, 0.2, 0.3, 0.4, skewness values of -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, and first block height values of 0.75, 0.90, 1.10, 1.2.

The height of the first block has a considerable impact on the wake structure; a taller initial block causes a more abrupt shift in the flow when it hits the array of blocks. The height of the first block has a strong direct correlation with pressure drop; however, the height of the first block and heat flux are not strongly correlated, as shown by the large variety in slopes. Other than the average height, the standard deviation is the most important factor in determining heat flux and pressure drop. The amount of heat flux and pressure drop increases as the standard deviation, which corresponds to a big disparity in heights, grows. In addition, there was no clear link between heat flux and pressure drop and skewness. As a result of the large variation in heat flux values, we have less confidence in trends derived from heat flux. Quantitatively, when the standard deviation of block heights increased from 0.1 to 0.4, the pressure drop increased by 60%, while heat flux increased by 13%.

The distinct regimes of skimming, wake interference, and the isolated regime have been demonstrated and utilized in published works. In accordance with the conditions of this study, simulations with the same block height and different spacing are conducted to analyze different regimes and the metric used to determine which regime each block belongs to. Skimming flow ($h/s > 0.65$), wake interference ($0.65 > h/s > 0.1$), and isolated ($h/b > 0.1$) regimes are all observed in the simulations. The effect of the regime on the correlation between pressure drop and standard deviation has been investigated. The result shows that the association between the standard deviation of block heights and pressure drop is hardly influenced by simulation regime type. The pattern of rising pressure drop with increasing standard deviation is similar regardless of whether simulations with all regimes or simulations with just the wake interference regime are included.

Sets of simulations with varying average heights have been generated, and it has been demonstrated that increasing the average height of the blocks increases the total heat flux and pressure drop. These simulations also verified the conclusion made from the effect of the standard deviation of heights on the total pressure drop.

We cannot precisely predict the pressure drop, overall heat flux, or block heat transfer coefficient based on the block height statistics because having the same statistic does not guarantee the same result. Six simulations with the same first block height, standard deviation, and skewness but different block heights were analyzed to identify qualitative differences between simulations with the same statistical properties but different pressure drops and heat fluxes. Results show that the height difference between the first two blocks along the wind path has the greatest influence on the total heat flux and pressure drop, followed by the height differences of adjacent blocks along the path. This proves that the height arrangement of the blocks and how they affect each other also matter. For example, a taller downstream block will change the wake structure.

As the height difference between the first two blocks plays a significant role in determining the pressure drop, its influence has been studied in greater depth. In situations with a smaller standard deviation, the relationship between height difference and pressure drop is very direct. However, as the standard deviation increases, the relationship between pressure

drop and height difference becomes obscure. It is also proven that when the second obstacle is larger than the first one, the standard deviation has a higher effect on the total pressure drop compared with the case where the second obstacle is smaller. The largest pressure drop value is achieved by combining a high standard deviation with a greater first block than the second one.

A few statistical parameters are unable to fully capture the detailed flow behaviour. In order to comprehend the non-linearity and determine whether machine learning can perform well on this task, machine learning models were utilized to predict the heat transfer coefficient of each block. In this work, the simulation data were employed as input for the ML models. On the dataset, a number of advanced machine learning regression models, such as Random forest, SVR, XGBoost, Gradient Boosting, and deep learning, were tuned, trained, and evaluated. During the tuning phase, K-fold cross-validation was used to ensure the generalization of the models. The study indicates that machine learning is able to tackle underlying non-linearity and estimate the surface heat transfer coefficient of buildings with decent accuracy.

First, the data set contained the information of all 14 blocks within a single simulation. The accuracy of the model on the train set was much better than the accuracy of the model using the test set. For certain models, the inaccuracy of the test set was double that of the train set. This indicates that models for this dataset suffer from overfitting and lack of generalization to new, unknown data.

In order to tackle the overfitting problem, the inputs were reduced to the six most important block heights. In order to find the more important features, a feature importance analysis was conducted. This approach did not improve the accuracy, which may be due to the small size of the data set, which is 204.

Based on the result of feature importance and the work done in the literature, we know that only a few surrounding buildings of the target building are important to consider. Therefore, a new data set of groups of five consecutive blocks have been created to predict the heat transfer coefficient of the fourth block in the group, which significantly increased the size of the data set. For this larger dataset, the three distinct NN models and machine

learning models were implemented. The deep learning model's performance was superior to that of the machine learning model. The difference between the statistical errors of the train and test data set in deep learning models was less than in conventional machine learning models.

4.1　Future work

Due to the complexity of the flow studied in this project, we cannot assert that we have constructed a flawless analysis of height variation in this thesis study if such a model ever exists. There are several limits to this research, and there are various ways this issue might be further researched. First, the low amount of accessible data limits the predictive capabilities of the models in this investigation. Our model's effectiveness and the conclusiveness of our findings might be enhanced with further data. In addition, this research used idealized simulations. By including additional characteristics in our simulation, such as surface roughness, buoyancy effect and turbulence intensity, we may get more realistic predictions from machine learning models. On the other hand, the study was conducted in a 2D environment, which may not accurately represent the real world, and studying 3D simulation would unquestionably add a great deal to this topic. The data may also be confirmed by the experimental work, and the ML models can be checked utilizing the outcome of experimental data.

Appendices

Appendix A: Heights of blocks in each simulation

Simulation number	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14
Simulation 1	0.90	0.90	1.04	1.00	0.96	1.03	0.81	0.99	0.89	1.12	1.15	1.05	1.08	1.11
Simulation 2	0.77	1.03	1.07	1.12	0.96	1.06	1.02	1.02	0.97	0.96	1.20	0.97	0.96	0.94
Simulation 3	1.20	0.80	1.00	0.99	0.96	1.02	0.84	1.05	1.03	1.10	1.06	0.94	1.05	0.96
Simulation 4	0.91	1.18	0.95	1.07	1.03	1.05	1.04	0.93	1.09	1.06	0.82	1.01	0.99	0.85
Simulation 5	0.78	0.87	1.01	0.93	0.99	0.94	1.11	1.00	1.01	1.12	1.06	0.96	1.00	1.18
Simulation 6	1.19	1.07	0.96	1.01	1.09	1.01	0.98	0.95	0.80	0.87	0.93	1.05	1.08	1.06
Simulation 7	0.90	1.12	0.98	0.95	0.83	1.04	1.13	1.06	1.17	0.90	0.95	0.94	0.92	1.10
Simulation 8	0.78	0.90	1.00	1.10	1.02	1.02	0.99	0.95	1.24	0.98	1.04	1.03	1.01	0.95
Simulation 9	1.19	0.96	0.90	0.90	0.97	0.97	1.02	1.07	1.09	1.13	0.94	0.82	1.04	1.00
Simulation 10	0.89	1.03	0.96	1.09	0.86	0.93	1.11	0.95	0.94	1.17	0.89	1.13	1.11	0.95
Simulation 11	0.78	0.98	0.95	1.00	1.02	1.23	0.96	0.98	0.97	0.96	1.15	1.02	0.94	1.07
Simulation 12	1.20	0.89	1.03	1.03	1.00	0.92	1.03	1.12	0.97	0.83	0.96	0.89	1.14	1.00
Simulation 13	0.90	0.65	1.18	1.16	1.12	1.10	0.67	0.88	1.31	0.92	1.19	0.89	0.88	1.09
Simulation 14	0.75	0.98	1.17	0.86	1.00	0.97	1.12	0.66	1.19	1.28	1.18	1.20	0.93	0.75
Simulation 15	1.20	0.87	1.14	0.86	0.75	0.83	1.06	1.30	1.16	0.62	1.05	0.89	1.20	1.04
Simulation 16	0.90	0.82	1.17	1.23	1.20	1.01	0.58	0.93	1.35	0.95	0.85	1.07	1.07	0.87
Simulation 17	0.75	0.83	1.09	1.10	1.09	0.99	1.20	0.82	0.98	1.11	1.09	1.36	0.61	0.93
Simulation 18	1.20	0.71	1.09	1.09	1.01	1.27	1.27	1.02	0.93	1.16	0.72	0.72	0.95	0.86
Simulation 19	0.90	1.10	0.78	0.96	1.01	1.04	1.29	0.72	0.93	1.33	0.95	0.98	0.68	1.26
Simulation 20	0.75	0.91	0.70	1.04	1.35	0.93	1.35	1.04	1.02	1.01	0.70	1.14	1.07	0.96
Simulation 21	1.20	1.45	1.14	0.94	1.05	0.84	1.09	0.59	0.92	0.93	0.83	1.04	1.04	0.95
Simulation 22	0.90	0.97	0.92	1.00	0.83	0.99	0.71	0.71	1.24	0.81	1.39	1.15	1.15	1.15
Simulation 23	0.75	1.29	0.89	0.92	0.94	1.28	0.69	1.16	0.85	1.32	0.99	0.96	0.86	1.10
Simulation 24	1.20	0.92	0.78	1.39	1.09	0.98	0.73	1.01	0.99	1.04	0.76	1.25	0.75	1.05
Simulation 25	0.90	0.82	0.41	0.80	1.25	1.47	1.32	0.64	1.27	1.18	1.21	1.03	0.99	0.74
Simulation 26	0.75	1.02	0.78	0.51	1.02	0.85	0.74	1.22	1.24	1.33	1.00	1.39	0.39	1.08
Simulation 27	1.20	1.07	1.34	0.81	0.89	0.91	1.43	0.43	1.12	0.45	0.76	1.01	0.81	1.19
Simulation 28	0.90	1.02	1.21	0.45	0.98	0.71	1.16	1.51	1.34	1.26	0.58	1.03	0.80	1.04
Simulation 29	0.75	1.13	1.00	0.59	0.47	1.12	1.35	1.20	0.98	1.00	0.69	0.97	1.50	1.30
Simulation 30	1.20	1.32	0.87	0.78	1.06	0.47	1.32	0.87	0.93	0.89	0.53	1.06	1.23	1.50
Simulation 31	0.90	0.74	1.11	1.13	0.54	1.36	0.81	1.05	1.09	0.55	1.58	1.34	0.83	0.95
Simulation 32	0.75	1.08	1.32	1.35	0.74	0.89	1.49	0.79	0.70	0.83	0.53	1.26	0.95	1.35
Simulation 33	1.20	0.54	0.63	1.23	1.04	0.95	0.89	1.21	0.87	1.09	1.58	1.34	0.73	0.68
Simulation 34	0.90	1.10	1.00	1.42	0.60	1.60	0.94	1.17	1.20	0.90	0.63	0.79	0.59	1.21
Simulation 35	0.75	0.52	0.92	0.71	1.09	1.60	1.37	1.23	1.12	1.25	0.87	1.10	0.73	0.80
Simulation 36	1.20	1.15	0.63	1.16	0.60	0.83	1.16	1.04	0.71	0.82	0.69	0.99	1.48	1.54
Simulation 37	1.10	0.99	0.90	1.07	0.92	1.12	1.09	1.10	1.09	1.02	0.88	0.84	1.00	0.88

Simulation 38	1.10	0.98	0.99	0.97	0.94	1.08	0.97	1.15	0.99	0.80	1.06	0.89	0.97	1.13
Simulation 39	1.10	1.08	0.87	0.93	0.99	0.90	1.08	0.93	0.97	1.04	1.16	0.86	1.15	0.95
Simulation 40	1.10	1.02	1.03	0.94	1.16	0.87	1.17	0.88	1.05	0.93	0.86	1.03	1.01	0.94
Simulation 41	1.10	0.83	0.84	0.67	1.18	1.26	1.00	1.21	0.71	0.98	0.81	1.23	1.16	1.05
Simulation 42	1.10	1.21	0.97	1.07	0.65	1.20	0.99	1.02	0.90	1.26	0.75	0.83	1.27	0.77
Simulation 43	1.10	0.95	0.88	1.26	0.69	0.89	0.78	1.17	0.78	1.36	1.20	1.09	0.90	0.98
Simulation 44	1.10	1.03	0.91	0.74	0.91	0.77	0.78	0.97	1.04	1.33	1.26	1.07	1.30	0.78
Simulation 45	1.10	1.30	1.06	0.80	0.99	1.04	0.77	1.12	0.47	0.54	1.41	0.76	1.26	1.40
Simulation 46	1.10	0.43	0.91	0.68	1.05	0.61	0.87	1.09	1.40	0.85	1.29	0.97	1.26	1.46
Simulation 47	1.10	1.19	1.51	0.68	1.34	1.40	0.57	0.79	0.86	1.22	0.73	0.67	1.02	0.95
Simulation 48	1.10	1.14	0.54	1.10	0.68	1.08	0.94	1.60	1.03	0.59	0.83	1.08	0.81	1.42
Simulation 49	0.80	1.04	1.03	1.01	0.99	0.98	0.94	1.01	1.02	1.03	1.23	0.95	1.10	0.89
Simulation 50	1.00	0.94	1.02	0.86	0.88	0.96	1.12	1.05	1.16	0.94	1.03	0.85	1.02	1.17
Simulation 51	0.80	1.11	1.13	0.76	1.35	0.92	0.98	1.08	1.17	0.96	0.97	0.70	0.78	1.32
Simulation 52	1.00	0.97	1.07	0.98	0.91	0.70	0.67	1.24	1.17	0.89	1.05	1.08	0.84	1.41
Simulation 53	0.75	0.86	0.81	0.87	1.09	1.19	0.96	1.03	1.01	1.06	1.01	1.31	1.08	0.95
Simulation 54	0.75	0.64	0.97	1.07	0.60	0.85	1.00	1.05	1.22	1.13	1.55	1.23	1.07	0.90
Simulation 55	0.90	0.90	0.84	1.15	1.04	1.27	1.07	1.09	0.69	1.13	0.97	1.08	0.97	0.92
Simulation 56	0.90	1.34	1.34	1.05	1.21	0.98	0.74	1.00	1.19	0.94	0.51	0.86	0.70	1.23
Simulation 57	1.20	0.83	1.01	1.22	1.02	1.09	0.76	1.01	1.40	1.02	0.70	1.13	0.75	0.85
Simulation 58	1.20	0.83	1.06	0.97	0.96	0.93	0.86	0.82	1.23	1.08	0.68	1.29	1.33	0.78
Simulation 59	0.75	1.61	1.13	0.78	0.64	0.90	1.14	0.69	0.86	0.58	1.18	1.22	1.29	1.25
Simulation 60	0.90	1.20	1.09	1.53	1.75	0.72	1.17	0.11	0.92	0.87	1.20	0.56	1.00	1.00
Simulation 61	0.75	0.22	0.77	0.79	1.27	0.67	1.24	1.45	1.24	1.03	0.51	1.59	1.02	1.47
Simulation 62	1.10	0.83	0.82	0.27	0.30	1.70	1.17	0.68	1.13	1.01	1.33	1.05	1.49	1.06
Simulation 63	1.20	1.08	1.05	0.11	1.41	1.10	0.75	0.55	1.73	0.79	1.46	0.99	0.91	0.83
Simulation 64	0.90	1.02	1.25	1.68	1.61	0.82	1.21	0.81	1.04	1.23	0.20	0.86	0.42	0.98
Simulation 65	0.75	0.35	1.08	0.87	1.77	1.34	1.02	0.97	0.71	1.22	0.27	1.27	1.11	1.23
Simulation 66	1.10	0.88	0.82	1.64	0.97	0.73	1.53	0.21	1.00	0.81	0.50	1.36	1.10	1.41
Simulation 67	1.20	1.04	1.63	1.36	1.14	0.48	1.27	0.26	0.70	1.55	1.06	0.75	0.94	0.65
Simulation 68	0.90	1.52	1.37	1.18	1.73	0.43	0.81	1.35	0.89	0.66	0.39	1.12	0.87	0.73
Simulation 69	0.75	1.54	0.84	1.18	0.94	1.59	0.67	1.00	1.60	1.38	0.54	0.62	1.02	0.37
Simulation 70	1.10	1.08	1.56	0.92	0.93	1.56	0.82	0.66	0.62	1.62	0.76	0.31	1.39	0.69
Simulation 71	1.20	1.37	1.22	0.33	0.69	0.82	1.18	1.84	1.06	0.63	0.45	0.85	1.05	1.27
Simulation 72	0.90	0.83	0.93	0.95	0.60	1.46	0.62	0.52	0.43	1.48	1.21	1.08	1.76	1.29
Simulation 73	0.75	0.96	1.33	0.45	1.45	0.51	1.76	0.74	0.59	0.78	1.39	1.36	1.18	0.80
Simulation 74	1.10	1.75	0.93	0.51	0.74	0.85	0.93	1.35	1.60	0.88	0.30	1.32	0.94	0.75
Simulation 75	1.20	0.81	1.29	1.83	1.20	0.42	0.74	1.16	0.99	1.46	0.56	0.89	0.42	0.98
Simulation 76	0.90	1.00	1.16	1.34	0.87	0.90	0.76	1.79	0.35	1.39	0.34	0.98	1.42	0.78
Simulation 77	0.75	1.51	0.88	0.73	0.48	0.92	0.33	1.68	1.13	1.04	0.65	1.52	1.10	1.31
Simulation 78	1.10	0.92	0.61	1.37	1.07	0.50	0.81	1.79	0.85	1.47	0.99	1.12	1.09	0.25
Simulation 79	1.20	0.33	1.03	1.35	0.92	1.08	1.64	0.84	1.02	0.43	1.09	1.68	0.62	0.74
Simulation 80	0.90	1.34	0.86	0.83	1.34	0.52	0.98	1.46	0.71	1.25	1.31	1.04	0.79	0.67

Simulation 81	0.75	0.89	0.95	1.05	1.36	1.06	1.31	0.76	0.85	1.57	0.43	1.26	0.74	1.07
Simulation 82	1.10	0.44	0.67	0.97	1.05	1.53	1.42	1.00	1.25	0.83	0.83	1.27	0.74	0.84
Simulation 83	1.20	0.55	1.36	1.49	0.66	0.81	0.70	0.67	1.08	1.15	0.91	0.87	1.15	1.33
Simulation 84	0.90	0.90	0.69	1.12	0.74	0.90	1.30	1.21	1.09	1.23	1.27	0.95	0.88	0.87
Simulation 85	0.75	1.35	0.96	1.19	0.89	1.14	0.72	0.87	0.91	1.00	1.24	0.72	1.19	1.07
Simulation 86	1.10	0.75	1.18	1.40	0.96	1.13	1.10	0.74	1.09	0.98	0.69	1.02	0.82	1.02
Simulation 87	1.20	0.89	1.33	1.07	0.67	0.81	1.01	0.92	0.80	1.12	1.26	0.90	0.87	1.19
Simulation 88	0.90	1.01	1.13	0.97	0.86	0.83	1.04	1.00	1.01	1.08	1.03	1.19	1.00	0.92
Simulation 89	1.10	1.00	0.82	0.92	0.97	1.05	1.21	0.95	1.05	0.95	0.87	1.04	1.05	1.06
Simulation 90	1.20	1.08	1.01	1.07	1.04	0.83	0.99	0.92	0.98	0.88	1.09	1.01	1.01	0.89
Simulation 91	0.90	1.12	0.66	0.91	0.51	0.80	1.59	1.47	1.52	1.41	0.95	0.25	0.70	1.19
Simulation 92	0.75	1.22	1.42	1.30	1.80	0.75	0.86	1.10	1.33	0.70	0.85	0.88	0.13	0.96
Simulation 93	1.10	1.22	1.03	1.49	0.51	0.66	1.19	1.18	0.67	0.28	1.61	1.52	0.68	0.84
Simulation 94	1.20	0.58	1.42	0.87	0.98	0.56	1.01	0.25	1.11	1.32	0.96	1.39	1.72	0.65
Simulation 95	0.90	0.66	0.92	1.18	0.45	1.57	0.72	1.25	1.20	1.11	1.21	0.96	1.20	0.68
Simulation 96	0.75	1.17	0.92	0.75	0.52	0.93	0.57	1.19	1.36	1.45	1.11	1.14	1.36	0.81
Simulation 97	1.10	0.49	0.90	0.84	0.96	1.20	1.38	1.18	0.85	1.08	1.02	0.92	1.58	0.44
Simulation 98	1.20	1.25	1.10	0.98	0.52	1.02	0.64	1.27	0.58	1.56	0.83	1.21	1.12	0.78
Simulation 99	0.90	0.64	0.85	1.00	1.30	1.22	0.71	1.02	1.29	1.13	1.04	1.04	0.82	0.97
Simulation 100	0.75	1.20	0.69	1.30	0.75	0.94	1.01	0.80	1.19	0.87	1.14	1.12	0.98	1.20
Simulation 101	1.10	1.05	1.23	1.17	0.88	0.83	1.10	1.17	0.68	1.01	0.86	0.71	0.84	1.31
Simulation 102	1.20	0.89	1.15	0.80	0.68	1.24	0.93	0.89	0.81	1.02	1.21	1.23	1.17	0.75
Simulation 103	0.90	1.09	1.04	1.02	0.88	1.08	0.83	0.90	1.03	0.97	1.17	1.03	1.04	1.08
Simulation 104	0.75	0.91	1.07	0.98	1.02	1.02	1.02	1.01	1.22	0.98	0.96	1.01	0.97	1.05
Simulation 105	1.20	1.01	1.03	1.04	1.01	1.03	1.04	0.80	1.07	1.00	0.98	0.84	0.94	1.03
Simulation 106	0.90	0.98	1.01	0.79	0.87	1.04	1.04	0.88	1.04	1.02	1.06	1.03	1.20	1.10
Simulation 107	0.75	1.22	1.02	0.94	1.08	1.01	0.99	0.98	1.02	1.08	0.96	1.02	0.94	0.93
Simulation 108	1.20	0.97	0.98	1.05	1.00	1.10	1.03	1.06	0.87	1.02	0.79	1.07	0.90	0.95
Simulation 109	0.90	0.97	1.06	0.89	1.12	1.18	1.04	0.99	0.87	1.07	1.03	0.93	1.03	0.86
Simulation 110	0.75	0.99	1.03	0.92	1.09	1.01	1.00	0.99	0.99	0.93	0.93	0.98	1.25	1.03
Simulation 111	1.20	0.87	1.06	1.04	0.95	0.91	1.00	1.04	1.07	0.95	0.84	1.06	0.92	1.05
Simulation 112	0.90	0.98	0.93	0.81	1.07	0.96	1.22	1.03	1.03	1.03	1.09	0.97	1.07	0.92
Simulation 113	0.77	0.94	1.00	0.97	1.07	0.99	0.91	0.90	0.98	0.98	1.19	1.01	0.94	0.87
Simulation 114	1.20	0.93	1.04	1.17	1.05	0.91	1.05	0.99	0.98	1.00	0.99	0.90	0.82	1.05
Simulation 115	0.90	1.11	1.15	0.86	1.10	0.85	1.30	1.23	1.17	0.88	1.13	0.98	0.66	0.68
Simulation 116	0.75	0.80	1.16	1.17	1.28	1.21	1.21	0.80	0.84	1.23	0.76	1.06	0.99	1.15
Simulation 117	1.20	1.16	0.60	1.00	0.84	0.95	1.11	1.32	0.74	0.93	1.02	1.00	0.88	1.23
Simulation 118	0.90	0.76	0.98	0.99	1.01	1.01	1.18	1.21	1.22	0.59	0.89	0.94	1.37	1.05
Simulation 119	0.75	0.65	0.95	1.02	1.21	1.13	0.83	0.90	1.13	1.27	1.18	0.79	0.93	1.23
Simulation 120	1.20	0.73	1.17	0.65	0.84	0.80	0.82	1.10	1.15	1.04	0.98	1.00	1.23	1.31
Simulation 121	0.90	1.17	1.10	0.82	1.22	0.77	1.03	1.05	1.28	1.32	0.99	0.88	0.81	0.73
Simulation 122	0.75	0.70	1.01	0.98	1.04	0.97	1.17	1.42	0.72	1.22	1.03	1.02	1.09	0.92
Simulation 123	1.20	1.03	1.12	0.80	1.09	1.05	1.13	1.43	1.00	0.98	0.73	0.77	0.96	0.67

Simulation 124	0.90	1.30	1.16	1.34	0.97	0.75	1.03	1.07	0.86	1.22	0.80	0.77	0.77	1.04
Simulation 125	0.75	1.10	0.87	0.85	1.24	1.01	1.23	0.80	1.39	1.09	0.90	1.12	0.70	0.96
Simulation 126	1.20	1.22	0.96	1.07	0.83	0.86	0.87	0.89	0.86	0.80	1.32	1.27	0.72	1.14
Simulation 127	0.90	1.30	1.23	0.84	1.11	1.20	0.47	1.41	0.98	1.38	0.65	0.59	0.79	1.16
Simulation 128	0.75	1.60	1.12	1.03	0.58	1.00	1.02	0.88	0.37	1.22	1.26	1.07	0.95	1.20
Simulation 129	1.20	1.20	0.35	1.41	0.88	1.10	0.83	0.77	1.09	0.66	1.52	0.92	1.17	0.94
Simulation 130	0.90	1.09	1.30	1.34	1.22	0.65	1.39	0.77	0.94	0.81	0.46	0.75	1.37	1.07
Simulation 131	0.75	1.11	0.95	0.95	1.26	0.61	1.19	1.45	0.51	0.82	0.62	1.37	1.20	1.22
Simulation 132	1.20	1.35	1.08	0.50	0.82	0.80	1.41	1.14	1.13	1.06	0.62	0.89	1.38	0.64
Simulation 133	0.90	0.97	0.75	1.53	0.84	0.65	1.48	0.99	1.22	0.78	1.10	0.48	1.06	1.30
Simulation 134	0.75	1.37	1.06	0.57	1.21	0.75	1.52	0.80	1.29	0.83	0.81	0.60	1.29	1.10
Simulation 135	1.20	0.80	1.62	0.53	0.86	1.28	0.57	1.20	0.86	1.00	1.14	1.10	0.69	1.12
Simulation 136	0.90	0.91	1.14	1.05	1.17	0.80	0.73	1.50	0.46	0.94	0.90	1.22	1.54	0.76
Simulation 137	0.75	1.44	1.13	1.11	0.65	1.18	1.36	0.69	0.98	1.07	0.75	0.65	0.79	1.49
Simulation 138	1.20	1.09	1.09	0.95	0.53	0.89	1.16	0.98	0.51	1.15	1.03	1.00	1.68	0.69
Simulation 139	1.10	1.17	0.83	0.95	1.01	0.89	1.05	0.98	1.02	1.08	1.01	0.83	1.07	0.99
Simulation 140	1.10	1.01	1.13	1.13	0.96	0.92	1.09	0.82	0.96	1.04	0.96	1.05	0.87	0.93
Simulation 141	1.10	0.91	1.01	1.11	1.01	1.05	0.88	1.00	1.20	1.01	0.85	1.07	0.87	0.93
Simulation 142	1.10	0.94	0.85	1.11	1.04	0.97	0.98	0.96	0.99	1.14	0.95	0.94	1.17	0.86
Simulation 143	1.10	1.20	0.96	0.77	0.88	1.31	1.16	0.60	0.92	0.95	0.92	1.19	0.89	1.17
Simulation 144	1.10	1.14	0.98	0.69	1.01	1.14	0.97	1.06	1.39	1.17	0.63	0.99	0.95	0.82
Simulation 145	1.10	0.81	1.46	0.59	1.09	1.06	0.88	1.05	1.17	0.96	0.86	1.04	1.07	0.88
Simulation 146	1.10	0.74	0.96	0.93	1.22	1.05	0.89	1.31	0.69	1.13	0.81	1.35	0.88	0.94
Simulation 147	1.10	0.91	0.98	1.07	0.93	0.60	1.23	1.33	0.67	1.14	0.78	1.46	0.43	1.33
Simulation 148	1.10	0.43	0.99	1.09	0.87	0.75	1.19	1.56	0.92	1.34	0.93	0.53	1.14	1.10
Simulation 149	1.10	0.84	1.34	1.13	0.59	0.56	0.91	1.61	0.88	0.90	0.58	1.16	1.22	1.12
Simulation 150	1.10	0.53	1.11	1.53	0.86	1.08	1.17	0.87	0.80	1.53	0.80	1.20	0.60	0.89
Simulation 151	0.80	0.95	1.00	1.03	0.96	1.08	1.00	1.22	0.97	0.97	1.12	1.01	1.02	0.93
Simulation 152	1.00	0.94	1.07	0.92	0.92	0.82	0.95	1.04	0.95	1.15	1.00	1.01	1.06	1.18
Simulation 153	0.80	1.14	0.92	0.97	0.68	0.81	1.29	0.91	1.20	1.25	0.84	0.99	1.30	0.93
Simulation 154	1.00	1.11	1.45	0.86	0.97	0.90	0.84	0.97	0.90	1.18	0.58	1.02	1.09	1.16
Simulation 155	0.75	0.83	1.22	1.09	0.99	1.12	0.96	1.04	1.32	0.99	0.81	1.03	0.98	0.92
Simulation 156	0.75	1.07	1.21	0.87	1.54	0.81	0.54	0.94	1.03	0.85	1.09	0.92	1.15	1.28
Simulation 157	0.90	0.90	1.20	0.90	1.12	1.12	0.90	1.10	1.01	1.22	0.98	0.87	1.13	0.69
Simulation 158	0.90	0.67	0.81	1.26	0.72	0.53	1.08	1.34	1.25	0.93	1.30	1.15	1.11	0.93
Simulation 159	1.20	0.80	0.68	1.13	1.40	0.75	1.15	1.13	0.85	1.16	0.86	0.88	1.02	0.97
Simulation 160	1.20	1.33	0.96	0.97	0.67	0.97	1.18	1.33	1.03	0.89	0.76	0.79	0.98	1.00
Simulation 161	0.75	1.59	0.39	0.93	0.80	1.03	1.08	1.07	1.00	1.14	0.82	1.00	0.84	1.53
Simulation 162	0.90	0.83	0.33	0.47	1.25	0.88	0.42	1.16	1.30	1.00	1.51	1.15	1.64	1.27
Simulation 163	0.75	1.17	1.08	1.35	1.37	0.95	1.23	0.26	1.43	1.06	1.64	0.59	0.75	0.50
Simulation 164	1.10	1.62	0.92	1.12	0.75	1.42	0.89	0.46	0.33	0.47	1.42	1.29	1.02	1.31
Simulation 165	1.20	0.67	0.67	0.34	1.37	1.52	1.00	1.43	0.41	0.96	0.62	1.45	1.31	1.09
Simulation 166	0.90	1.46	0.61	1.62	1.39	0.55	1.05	0.24	1.40	0.67	1.03	0.95	0.90	1.28

Simulation 167	0.75	0.35	1.31	0.74	0.50	1.01	1.21	1.34	1.31	1.39	0.71	0.51	1.24	1.61
Simulation 168	1.10	0.65	1.51	1.31	0.50	0.71	1.62	0.32	0.99	1.32	1.08	1.21	1.11	0.51
Simulation 169	1.20	1.52	0.83	1.12	0.18	1.28	1.00	1.02	1.72	0.57	1.02	1.18	0.76	0.54
Simulation 170	0.90	1.17	0.97	1.64	1.09	0.82	1.17	1.16	0.79	0.31	1.75	0.43	1.08	0.66
Simulation 171	0.75	0.67	0.63	0.99	1.26	0.74	1.10	0.60	1.37	1.38	0.94	0.36	1.70	1.49
Simulation 172	1.10	0.27	0.94	1.29	1.76	1.10	1.62	0.76	0.63	1.27	0.80	1.04	0.57	0.86
Simulation 173	1.20	1.30	0.56	0.50	1.67	0.96	1.00	0.84	0.91	1.00	0.52	1.47	1.52	0.50
Simulation 174	0.90	0.87	0.41	1.37	0.33	0.90	1.88	0.98	1.14	1.19	0.71	0.80	1.21	1.29
Simulation 175	0.75	0.90	0.90	1.14	0.19	1.01	0.80	1.93	0.95	0.58	1.36	1.31	1.24	0.99
Simulation 176	1.10	1.47	0.49	0.80	1.11	0.74	1.13	0.83	0.90	1.54	0.47	1.74	0.54	1.24
Simulation 177	1.20	0.86	0.92	1.56	0.72	1.28	0.32	1.35	0.75	0.96	1.79	0.96	0.87	0.49
Simulation 178	0.90	1.43	1.05	0.82	1.79	0.82	0.25	1.05	0.69	1.32	0.89	1.11	0.53	1.37
Simulation 179	0.75	0.32	1.71	1.39	1.11	0.72	1.22	1.09	1.11	0.93	0.47	1.56	0.69	0.95
Simulation 180	1.10	0.91	0.43	1.59	1.30	1.13	0.80	1.62	0.61	1.18	0.43	1.43	0.72	0.80
Simulation 181	1.20	0.53	1.59	1.38	0.51	0.98	0.71	0.61	1.30	1.06	1.62	0.49	0.78	1.31
Simulation 182	0.90	1.53	1.23	0.87	1.17	1.24	1.44	0.83	0.62	0.71	0.99	0.96	1.07	0.50
Simulation 183	0.75	1.07	0.37	0.92	1.15	0.90	0.81	1.39	0.87	1.31	0.87	1.60	0.89	1.08
Simulation 184	1.10	0.74	0.83	1.05	0.96	1.62	1.05	0.89	0.38	1.23	1.39	1.05	0.96	0.79
Simulation 185	1.20	1.01	0.60	1.43	0.93	0.51	1.48	0.74	1.02	1.22	0.90	0.75	1.34	0.87
Simulation 186	0.90	1.15	0.59	0.99	0.76	1.41	1.06	0.93	1.00	1.00	1.25	1.02	0.88	0.99
Simulation 187	0.75	0.98	0.95	0.66	1.11	1.02	0.85	0.82	1.05	1.26	1.07	1.01	1.17	1.38
Simulation 188	1.10	1.20	0.74	0.91	1.15	0.75	0.70	1.05	1.38	0.95	0.82	1.17	1.00	1.04
Simulation 189	1.20	1.26	0.95	1.03	0.90	1.15	0.79	1.36	1.01	0.83	0.66	0.95	0.81	1.14
Simulation 190	0.90	1.14	1.06	0.86	1.06	1.02	0.98	0.92	1.11	0.85	0.92	1.02	0.98	1.16
Simulation 191	1.10	0.93	0.91	0.94	0.92	0.94	1.01	0.98	1.10	0.83	1.10	1.17	0.99	1.13
Simulation 192	1.20	1.05	0.98	0.95	0.94	0.90	0.85	1.03	1.04	1.07	1.05	1.09	0.81	0.93
Simulation 193	0.90	1.50	0.53	0.26	1.20	0.74	1.47	0.76	1.54	0.88	1.41	0.64	0.85	1.24
Simulation 194	0.75	0.48	1.39	0.57	1.55	0.66	0.87	1.33	0.39	1.29	1.51	1.34	1.06	0.87
Simulation 195	1.10	1.41	0.79	0.54	0.76	1.00	1.39	1.79	1.03	0.67	1.17	0.18	1.15	1.04
Simulation 196	1.20	1.23	1.65	1.27	1.56	0.35	0.50	0.90	0.89	0.42	1.28	0.80	0.91	1.02
Simulation 197	0.90	0.77	0.33	1.27	0.77	0.90	1.62	0.87	1.03	0.94	1.01	1.22	1.29	1.07
Simulation 198	0.75	1.15	1.01	0.80	1.14	0.59	0.96	1.48	1.39	1.28	1.12	1.16	0.70	0.53
Simulation 199	1.09	0.95	1.56	1.04	0.82	0.85	1.30	1.19	1.31	0.78	0.45	1.13	0.54	0.95
Simulation 200	1.20	1.02	0.38	0.72	1.28	0.97	0.71	1.14	1.58	0.83	1.21	0.84	1.25	0.85
Simulation 201	0.90	0.64	0.90	0.70	0.87	1.05	1.07	0.89	0.91	1.07	1.21	1.25	1.32	1.21
Simulation 202	0.75	0.97	0.95	1.39	0.93	1.01	1.18	0.89	0.89	1.19	0.59	1.00	1.10	1.13
Simulation 203	1.10	1.22	0.77	1.22	0.82	1.31	0.79	1.00	0.68	0.81	1.15	1.17	0.99	1.00
Simulation 204	1.20	0.69	1.07	0.77	0.74	1.08	1.25	0.91	1.25	1.15	0.89	0.77	1.24	0.95

Curriculum Vitae

Name: Ali Navidi

Post-secondary   Amirkabir University
Education and   Tehran, Tehran, Iran
Degrees: 2015-2020 B.A.

                               The University of Western Ontario
                               London, Ontario, Canada
                               2021-2022 M.A.

Honours and   Fred Groch Graduate Scholarship for master's student
Awards: 2022-2022

                               Full Scholarship for Bachelor's studies at Amirkabir University
                               2015-2020

Related Work   Teaching Assistant
Experience   The University of Western Ontario
                               2021-2022

                               Research Assistant
                               The University of Western Ontario
                               2021-2022

                               Teaching Assistant
                               Amirkabir University
                               2019-2020