Uncertainty Quantification and Global Sensitivity Analysis of Efficient Numerical Atmospheric Dispersion Models

Tasnia Shahid

Department of Mathematics and Statistics,
University of Strathclyde,
Glasgow, United Kingdom

A thesis submitted for the degree of
Doctor of Philosophy

September 2025

Copyright Declaration

Chapter 3 of this thesis contains material which has been submitted for publication, reference [41]. The results contained therein are the result of the author's original research in collaboration with the article's co-authors, Professor John A. Mackenzie, Dr Alison Ramage, and Professor Alistair B. Forbes.

Signed:

Date:

Acknowledgments

First and foremost, I would like to express my most sincere gratitude to my supervisors, John Mackenzie and Alison Ramage, for their invaluable guidance, collaboration, and support. Their encouragement and expertise have been fundamental to every stage of this work. I would also like to thank my industrial supervisor, Alistair Forbes, for his thoughtful insight and perspective, which provided crucial context and greatly strengthened the development of this research.

This work would not have been possible without the generous financial support of the National Physical Laboratory (NPL) and the Engineering and Physical Sciences Research Council (EPSRC), whose funding is gratefully acknowledged. I am also indebted to the Department of Mathematics and Statistics at the University of Strathclyde for providing an excellent research environment, and to colleagues and friends for their continual camaraderie, guidance, and assistance throughout this journey.

I am deeply thankful for the relentless support and encouragement of my brother, Shahnazam, whose continual care and faith has been my anchor through it all. I would also like to thank Ava for her friendship, reassurance, and unfaltering presence, which has been a constant source of comfort and motivation. I am equally thankful to my wider circle of friends who have offered joyful companionship over the years. Above all, I am profoundly grateful to my parents for their unwavering love, guidance, and belief in me throughout this journey. My father has inspired me all my life, and my mother has shown me the meaning of patience and dedication. Their enduring support has been the foundation of my strength in both challenging and rewarding times, and this thesis would not have been possible without them. All that I do is dedicated to honouring their devotion.

Abstract

Air pollution is one of the most pressing global challenges, with severe consequences for human health, ecosystems, and climate. Atmospheric pollutant dispersion is governed by turbulent transport and diffusion, and mathematical models are used to simulate concentration fields that represent these processes. However, the accuracy of such models typically depends on input parameters whose values may not be known exactly, making the use of uncertainty quantification (UQ) and sensitivity analysis tools essential. Conventional global sensitivity analysis (GSA) often collapses spatial information into scalar measures, limiting its ability to reveal how the influence of individual parameters varies across space. This thesis applies spatially resolved GSA to steady-state advection-diffusion systems for modelling pollutant dispersal. Specifically, Sobol indices are used to characterise how parameter influence varies across space, with confidence intervals constructed at each stage of analysis to assess the reliability of the sensitivity estimates. The investigation progresses from one-dimensional test problems, which establish a benchmark and confirm the variance in wind speed as the dominant source of output variability, to two-dimensional models that capture more realistic dispersion. The models use the Streamline Upwind Petrov-Galerkin (SUPG) finite element method to ensure robust and reliable discretisation of the problem. In the 2D case, computational demands are mitigated by the use of artificial neural network (ANN) surrogate models trained on numerical simulations, enabling efficient estimation of Sobol indices, which serve as the GSA measure in this work. This approach is then applied to an urban air quality model formulated as a full PDE system. The subsequent study demonstrates both the dominant role of wind speed and the regional significance of the stability exponent and emission rates, with ANNs again providing scalability while preserving key spatial sensitivity patterns. The findings show that spatially resolved GSA reveals parameter influence in ways inaccessible to scalar approaches, highlighting how sensitivities shift across space and how targeted reduction of uncertainty in key parameters can lower output variance. By combining numerical modelling, surrogate methods, and UQ, the thesis demonstrates that high-resolution spatial GSA is both feasible and informative for pollutant dispersal problems, and proposes a transferable methodology for other PDE-based models in which spatial variability is key.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Urban air quality is a pressing global concern, posing substantial threats to public health and environmental well-being. A joint statement issued by several public health organisations and scientific societies reflecting on the new World Health Organisation (WHO) Air Quality Guidelines published in 2021 includes the observations that "there is now broad expert consensus that air pollution is a major global public health risk factor and puts an enormous financial burden on societies" and "air pollution currently ranks fourth among the major risk factors for global disease and mortality', leading to millions of premature deaths on an annual basis [23,72]. Health conditions linked to air pollution include respiratory infections, chronic obstructive pulmonary disease (COPD), cardiovascular diseases, and neurodegenerative disorders [31, 36, 37]. The current state of global air quality is concerning as many regions worldwide face severe challenges. Air quality standards vary across countries, and many fall short of WHO thresholds. Despite some progress over recent decades several urban cities still exceed WHO guidelines for regulated pollutants. Notably, even low exposure levels have measurable impacts on respiratory and cardiovascular systems [3, 65].

1.1 Motivation for Air Quality Modelling

Air quality modelling has become an essential component of environmental research and is a critical tool for understanding and mitigating the environmental and health impacts of pollution in an increasingly urbanised and industrialised world. Air quality models help identify pollution sources, simulate pollutant dispersion, and evaluate exposure risks. Effective modelling of air quality problems helps form a scientific basis for informing regulatory frameworks.

Air pollution originates from diverse anthropogenic sources, including industrial processes, vehicular emissions, domestic combustion, and waste disposal. Key pollutants such as nitrogen oxides (NOx), sulfur dioxide (SO2), ozone (O3), carbon monoxide (CO), volatile organic compounds (VOCs), and particulate matter, including PM2.5 and PM10, contribute to a wide array of health effects [48, 61]. In addition, heavy metals such as zinc, often released through industrial discharges, accumulate in soils and water systems, leading to long-term ecological degradation and health issues for humans through contamination of water, air, and soil [33, 60].

The spatial and temporal variation in air quality is substantial, shaped by factors such as population density, industrial activity, land use, and the efficacy of environmental regulations. Urban areas typically exhibit poorer air quality due to dense traffic and industrial emissions, while rural areas experience variability linked to agriculture and biomass burning. These fluctuations are further modulated by weather patterns, seasonal variation, and localised human activity, necessitating high-resolution spatio-temporal monitoring systems [11, 72, 73].

Understanding pollutant dispersion dynamics is critical to accurately model air quality. The spatial distribution of pollutants is influenced by emission source types, intensity, meteorological conditions, and geographical layout. For instance, compact urban development does not inherently lead to higher pollution levels; the net effect depends on the interaction between source-specific emissions and atmospheric conditions [9, 53].

Air pollution research occupies a central role in applied mathematics, where it involves modelling nonlinear, spatio-temporal, and multivariable phenomena. Predicting pollution trends, evaluating exposure impacts, and designing intervention strategies demand sophisticated statistical and computational models [27, 29]. Advances in satellite remote sensing and distributed sensor networks are enhancing our ability to understand chemical interactions, emissions attribution, and pollutant dynamics at multiple scales, leading to more effective air quality management strategies [71].

The environmental impacts of air pollution also extend beyond human health. Ecological structures, including forests and agricultural systems, are increasingly stressed by airborne pollutants, especially ozone, which is transported across continents, affecting biodiversity and crop yields [2]. Long-term exposure has been linked to structural changes in forest ecosystems worldwide [12]. Furthermore, the interaction between air pollution and climate dynamics can intensify stress on regional water availability and ecosystem productivity. Regional chemistry–climate feedbacks can even offset expected climate change trends such as warming and wetting unless air pollution is explicitly incorporated into climate models [13].

Climate change is expected to amplify these effects, particularly by increasing ground-level ozone and particulate matter concentrations [65]. Pollutants such as aerosols and greenhouse gases alter the terrestrial hydrological cycle and contribute to radiative forcing, intensifying both atmospheric pollution and climate variability [2]. Understanding these interactions is essential, as they pose substantial risks to environmental sustainability and socio-economic stability. Mathematical modelling provides the tools to simulate how emissions affect atmospheric composition, climate processes, and their feedbacks. These models are critical for developing integrated policies that address air quality and climate change simultaneously [6].

1.2 Mathematical Models for Pollutant Dispersion

The National Physical Laboratory (NPL), as the UK's national metrology institute, has a strong interest in research that advances the accuracy and reliability of environmental measurements. This work's focus on spatially resolved sensitivity analysis directly supports the development of robust modelling tools for air quality assessment, which are essential for evidence-based regulation and standards. By improving methods for quantifying and managing uncertainty, this thesis aligns closely with NPL's mission to provide measurement science that underpins policy, industry, and public health.

Central to standard pollutant dispersal models is the challenge of accurately describing pollutant dispersion, the transport and spread of contaminants in the atmosphere driven by advection, diffusion, and, in some cases, chemical transformation. In this section, we review key developments in the mathematical modelling of pollutant dispersion, focusing on the governing equations, modelling assumptions, and numerical solution techniques.

Modelling Approaches

Air quality models are designed to simulate real-world pollution scenarios, yet their accuracy is inherently constrained by the complexity of atmospheric processes and the limited availability of complete observational data. Various modelling approaches can be employed to predict and assess pollutant concentrations in the atmosphere, from simple, steady-state formulations to more advanced time-dependent models capable of resolving atmospheric chemistry, meteorological interactions, and multi-scale transport dynamics.

Dispersion models are typically categorised into three main types: Gaussian, Eulerian, and Lagrangian. Each is characterised by distinct assumptions, numerical structures, and applica-

tion domains. Understanding their differences is essential for selecting an appropriate modelling approach based on the objectives of a given study, spatial scale, the pollutant type, and the computational resources available. Physical considerations such as turbulence, boundary conditions, and source terms, which describe emissions, are also integral components of accurate air quality modelling [32, 38, 66].

Gaussian plume models are among the earliest and most widely used dispersion modelling techniques. They assume pollutant concentration follows a Gaussian (normal) distribution as it moves away from the source. These models are most suitable for near-field applications (typically up to 50 km), especially when atmospheric conditions are relatively stable. Due to their low computational demands, Gaussian models are often used for regulatory purposes and for estimating concentrations from point, line, or area sources. However, they offer limited capability in representing time-dependent behaviour, complex terrain, or chemical transformations, which restricts their use in more dynamic or heterogeneous environments [32, 66].

Eulerian models adopt a fixed spatial grid and solve conservation equations for mass, momentum, and energy to simulate the evolution of pollutant concentrations. They are capable of handling multiple sources, complex chemical reactions, and both near-field and far-field transport. These models are widely used in regional-scale air quality assessments where the interplay between emissions, meteorology, and atmospheric chemistry is significant. The primary advantage of Eulerian frameworks lies in their ability to couple with meteorological models and to simulate chemical transformation processes over extended spatial domains. However, their computational cost can be substantial, particularly at high resolutions or when simulating a large number of chemical species [32, 47].

Lagrangian models, by contrast, track the movement of individual air parcels or particles through time and space, using a moving frame of reference. This allows for more detailed representation of point sources, complex terrain, and long-range pollutant transport. Lagrangian models typically require detailed meteorological input and can be computationally intensive, but their flexibility and adaptability make them well-suited for studies requiring high spatial and temporal precision [32].

A further category of dispersion modelling is Computational Fluid Dynamics (CFD) models, which are particularly valuable for small-scale applications involving complex geometries, such as urban environments or industrial sites. CFD is essentially a high-resolution Eulerian approach. These models numerically solve the Navier–Stokes equations, adapted to atmospheric flow, on a high-resolution computational mesh that captures intricate surface features.

CFD frameworks can incorporate different turbulence models, such as the Reynolds-averaged Navier–Stokes (RANS) approach, often paired with the widely used $k-\varepsilon$ model, or Large Eddy Simulation (LES), which resolves large turbulent structures [38]. While CFD provides physically detailed, high-resolution simulations, it is computationally demanding and thus generally constrained to limited spatial and temporal domains.

Each modelling approach presents trade-offs between complexity, resolution, and computational feasibility. The choice of model is influenced by several contextual factors, including geographic scale and topography, timescale of interest, land use characteristics, source configuration, and removal mechanisms such as deposition or chemical decay [32]. Advanced models, while more capable of simulating realistic scenarios, often require extensive environmental input data and a strong understanding of atmospheric dynamics.

Governing Equations and Physical Processes

Pollutant dispersion models are primarily based on advection-diffusion and reaction-diffusion equations, which describe how pollutants are transported by the airflow (advection), spread out due to turbulent and molecular mixing (diffusion), and potentially undergo chemical transformations (reaction). These equations form the mathematical foundation for most air quality models, capturing the spatial and temporal evolution of pollutant concentrations within the atmosphere [38, 62]. The underlying advection-diffusion and advection-reaction equations can be understood as extensions or implementations of the Eulerian and Lagrangian frameworks. In Eulerian models, these equations are solved at each grid point, producing spatially continuous concentration fields, while in Lagrangian models, the equations govern the stochastic movement of particles, ensuring that their collective behaviour reproduces the same transport and reaction dynamics.

To represent these equations meaningfully, models must incorporate the key physical processes that govern atmospheric pollutant behaviour. Advection describes the transport of pollutants by the wind profile, which can be characterised by wind velocity and direction. In many formulations, wind velocity is expressed as the sum of a deterministic (mean flow) and a random (turbulent fluctuation) component to capture both organised transport and stochastic mixing. Eddy diffusion represents the spreading due to turbulent mixing [4,62]. Turbulence accounts for processes such as the development of mixing layers and the influence of atmospheric stability on vertical and horizontal dispersion [38, 62]. The roughness length characterises the effect of surface features (vegetation, buildings, terrain) on wind profiles and turbulence [25]. Deposition

processes, both dry (gravitational settling, surface uptake) and wet (removal by rain, snow, or other forms of precipitation), remove pollutants from the atmosphere, while chemical reactions and transformations can alter pollutant species and concentrations over time [38].

A key challenge in solving these equations lies in the appropriate specification of boundary conditions, which define how the model interacts with its spatial domain. These conditions determine how pollutants enter, leave, or behave at the limits of the model, ensuring that solutions remain physically realistic and numerically stable. In dispersion modelling, boundaries may represent the limits of the computational domain, and their formulation can significantly affect predicted concentration outputs. Choices of boundary conditions often depend on the modelling approach, the spatial and temporal scale of interest, and the available observational or input data.

Numerical Methods

Solving the atmospheric transport equations that govern pollutant dispersion requires robust numerical methods capable of handling advection, diffusion, and reaction terms while maintaining numerical stability and accuracy. Among the most widely used techniques are Finite Difference Methods (FDMs), Finite Volume Methods (FVMs), Finite Element Methods (FEMs), and adaptive grid techniques, each offering distinct advantages and drawbacks.

FDMs use a straightforward approach to approximate derivatives in the governing equations using finite differences on a discretised grid. For explicit FDMs, the solution at the next time step is computed directly from known values at the current step, making them simple to implement but requiring very small time steps to ensure numerical stability. In implicit FDMs, the solution at the subsequent time step is obtained by solving a system of equations that simultaneously incorporates the dependent variables at that future time level. This approach permits the use of larger time steps while ensuring numerical stability, but increases computational effort [38, 49].

FVMs solve the governing equations by dividing the domain into small control volumes and evaluating fluxes of quantities across their boundaries. Instead of working at single grid points, FVMs calculatethe fluxes of quantities across the boundaries of these volumes, which makes them well suited for transport problems. They are widely used in air quality models because of their ability to handle complex flows and irregular geometries, though accurately computing fluxes can make them more computationally demanding [69].

FEMs offer a more flexible and mathematically rigorous framework by subdividing the com-

putational domain into elements and approximating the solution using basis functions defined locally on each element. This piecewise representation allows FEMs to effectively handle complex geometries, irregular grids, and heterogeneous conditions, which are common in urban or varied terrain environments. Among the common FEMs applied to air quality modelling are the Streamline Upwind Petrov–Galerkin (SUPG) method, which stabilises solutions to advection-dominated problems by introducing directional weighting to reduce numerical oscillations, and the discontinuous Galerkin (DG) method, which allows for discontinuities between elements, enhancing stability and reducing numerical diffusion in sharp gradient regions [15, 38]. Traditional Galerkin FEMs can produce non-physical oscillations and instabilities when simulating sharp concentration gradients or steep fronts common in atmospheric pollutant plumes. SUPG mitigates these issues by modifying the weighting functions in the variational formulation, introducing an upwind bias aligned with the direction of flow (streamlines). This selective addition of artificial diffusion along the flow direction effectively reduces oscillations, and it is known that this approach improves numerical stability and accuracy [15].

To further enhance solution accuracy while managing computational cost, adaptive grid techniques dynamically modify grid resolution based on evolving solution features. Three common approaches are h-refinement, which adds grid points to increase resolution in regions of high gradients, and r-refinement, which relocates existing grid points toward areas requiring enhanced detail, and p-refinement, which increases the degree of the polynomial basis functions used to approximate the solution within each element. Adaptive grid methods have been shown to significantly improve model fidelity for localised phenomena such as point source plumes without incurring the computational expense of uniformly fine meshes across the domain [38].

The choice among these numerical methods typically depends on the modelling objectives and constraints. FDMs are often good for large-scale, long-range dispersion models where computational efficiency is paramount. In contrast, FEMs and adaptive grids are better suited to urban-scale or complex terrain applications where accuracy in representing spatial heterogeneity and sharp gradients is critical, despite their higher computational cost. Across all methods, considerations of grid resolution remain central. Finer grids enable capturing intricate spatial variability and detailed dispersion features but increase computational demand. Conversely, coarser grids reduce computational load but may oversimplify the plume dynamics. To address these computational challenges, the use of parallel computing and high-performance computing architectures has become popular, allowing for efficient handling of large-scale models with fine spatial resolution and detailed physical parametrisations [38].

1.3 Uncertainty Quantification in Air Quality Modelling

In the context of mathematical modelling, uncertainty quantification (UQ) is often used for identifying and measuring uncertainties inherent in both input parameters and model formulations. In air quality simulations, these uncertainties can arise from incomplete or imprecise data, model structure limitations, and the inherent randomness of environmental systems, all of which can impact the reliability of model predictions [74]. By systematically managing and analysing these uncertainties, UQ can enhance confidence in simulation outcomes and inform robust decision-making in environmental policy and public health [42, 45].

Sources of Uncertainty

Air quality models are influenced by multiple sources of uncertainty that collectively contribute to variability in output predictions. One primary source is input data variability, where inaccuracies and inconsistencies in emission inventories and meteorological data propagate through simulations, leading to significant uncertainty in pollutant concentration estimates [24]. In this thesis, the input data are the samples generated from probability distribution functions that are used to describe model parameters. Another critical source is parameter uncertainty, which arises from the imprecise knowledge of model parameters such as emission factors and chemical reaction rates. These uncertainties can be amplified by nonlinear atmospheric processes, resulting in considerable variation in model outputs. Model structural errors also play a significant role; these uncertainties originate from the simplifications and assumptions underlying model design, with different conceptual frameworks often producing divergent results, underscoring the sensitivity of predictions to model formulation [26]. Lastly, numerical approximations introduced during the discretisation and approximation of governing equations contribute to uncertainty, particularly in cases dominated by advection or stiff chemical kinetics. Although generally smaller than errors from other sources, numerical errors remain an important factor in the overall uncertainty.

Impact of Uncertainty on Model Predictions

The presence of uncertainties significantly affects the accuracy and reliability of air quality predictions. Studies have shown that variability in meteorological inputs and physical model assumptions can lead to changes of up to 20–30% in pollutant concentrations within highly polluted regions, influencing both spatial distribution and transport pathways [19]. This level

of variability highlights the necessity for rigorous uncertainty propagation techniques, such as Monte Carlo (MC) simulations, which allow for the probabilistic treatment of predicted pollutant levels and provide confidence bounds around forecasts.

Uncertainty Quantification Techniques

Several methodologies exist for quantifying uncertainty, broadly classified into probabilistic and non-probabilistic approaches. Probabilistic methods, such as Bayesian inference, incorporate prior knowledge and observed data to probabilistically represent uncertainty within, e.g., model parameters. Alternatively, non-probabilistic techniques like interval analysis and fuzzy set theory provide frameworks for representing uncertainty when statistical information is scarce or incomplete [74].

Probabilistic methods form the foundation of many UQ studies by representing uncertainty through probability distributions. Such approaches are particularly advantageous when dealing with complicated nonlinear systems [74]. In the context of air quality modelling, probabilistic methods enable the prediction of pollutant concentration distributions rather than single output values, providing a richer understanding of uncertainty. Non-probabilistic methods, including interval analysis and fuzzy set theory, offer alternative strategies when precise probability distributions are difficult to establish due to limited data or inherent ambiguity. These frameworks represent uncertainty through bounded intervals or membership functions rather than exact probabilities, making them well-suited for scenarios with imprecise or incomplete information. Such methods provide valuable flexibility in representing epistemic uncertainty where subjective interpretation or data gaps are prevalent [74]. The choice between these frameworks depends on the nature of available data, the type of uncertainty encountered, and the specific goals of the modelling study.

Methods such as MC simulations can be used to address the numerical propagation of uncertainty through models by employing repeated random sampling to explore the variability of outputs arising from uncertain inputs, thereby offering a probabilistic approach that captures the full range of possible outcomes rather than a single deterministic result. Enhancements like Quasi-Monte Carlo (QMC) methods utilize low-discrepancy sequences to achieve more efficient and faster convergence compared to traditional MC sampling. Complementing these, surrogate models act as computationally inexpensive approximations of complex full-physics simulations, enabling extensive uncertainty analyses that would otherwise be prohibitively expensive.

1.3.1 Sensitivity Analysis in Air Quality Modelling

Sensitivity analysis (SA) is a fundamental tool in UQ that focuses specifically on understanding how variations in model inputs affect the model outputs. By identifying which parameters most influence model behaviour, SA helps prioritise efforts in data collection, model refinement, and resource allocation [1,59,68]. While UQ broadly addresses uncertainty sources and propagation, SA narrows the analysis to input-output relationships, providing critical insights into model dynamics and guiding targeted improvements.

SA techniques can be broadly categorised by their scope and computational approach. A key distinction lies between local and global methods. Local sensitivity analysis (LSA) examines the effect of infinitesimal changes around a fixed nominal input set and is best suited for linear or near-linear models where parameter interactions are minimal [28]. Global sensitivity analysis (GSA), in contrast, provides a more comprehensive evaluation by considering the full range of input variability and their potential interactions. GSA is thus better suited for complex, nonlinear systems in which understanding model behaviour across the entire parameter space is crucial [59]. This method captures a broader perspective of model dynamics and is indispensable when aiming to identify key drivers of uncertainty in realistic, multifactorial scenarios.

Among GSA methods, variance-based techniques, particularly Sobol indices, have gained prominence for their ability to decompose output variance into contributions from individual inputs and their interactions [17, 52, 58, 63]. This quantitative decomposition of variance yields interpretable metrics that reveal both main effects and higher-order interactions, aiding the prioritisation of influential parameters [5, 51, 63]. However, a Sobol indices approach can be computational expensive, especially for complex models with many parameters [68].

Screening methods are another approach to SA,design to identify efficiently non-influential variables at early modelling stages, thereby reducing dimensionality and computational burden. The Morris method is widely recognised for its computational efficiency and ability to provide qualitative rankings of input importance, making it valuable in preliminary analyses [7].

SA techniques can also be distinguished based on their computational strategy, particularly between derivative-based and variance-based methods. Derivative-based approaches assess sensitivity by calculating the partial derivatives of the model output with respect to each input parameter. These methods tend to be computationally efficient and are effective for smooth, differentiable models. However, they may fail to accurately capture the influence of interactions between variables or nonlinear responses [34]. On the other hand, variance-based techniques provide a robust statistical framework for decomposing the output variance into contributions

from individual inputs and their interactions. Among these, Sobol indices stand out for their ability to account for complex interdependencies across the entire input space, offering detailed and interpretable sensitivity metrics [58, 63].

The handling of spatial dependence in SA is a particularly significant issue given the spatially distributed nature of environmental phenomena. Traditional SA methods often rely on scalar outputs, such as domain-averaged pollutant concentrations, which may obscure critical spatial variability in parameter influence. One of the main challenges in employing Sobol indices in spatially distributed models is the computational demand due to the high dimensionality and complex interactions within such models.

To address computational challenges, surrogate methods have been increasingly employed to approximate complex air quality simulations efficiently. By enabling extensive sampling over high-dimensional input spaces, surrogate-based approaches facilitate the practical application of variance-based GSA methods, including Sobol indices, in otherwise intractable problems [67].

There are many techniques being developed to incorporate spatial variability directly into the analysis. One approach is to use spatial decompositions to handle data with spatial components effectively. This method allows one to perform SA on spatial maps of Sobol indices, capturing spatial variability and offering a more comprehensive view of the sensitivity across different spatial regions [44].

However, the treatment of spatial dependence in SA still poses challenges, and ongoing research aims to integrate spatial considerations more thoroughly to improve the reliability and interpretability of SA in spatially distributed systems. Despite these developments, there remains a clear gap in the literature concerning the systematic evaluation of Sobol indices with explicit spatial dependence, particularly within the context of air quality modelling. While methods such as spatial decomposition and surrogate modelling offer promising avenues, their application remains relatively limited in scope, particularly within the area of air quality modelling.

Spatial variability significantly impacts pollutant concentration levels because air pollution is subject to both spatial and temporal changes influenced by factors like population density, urbanisation, and industrial activities. Models that incorporate spatial relationships improve the accuracy of air quality predictions by capturing the dependencies not just between neighbouring locations but also between sites with similar functional characteristics or patterns [70]. This capacity to model diverse inter-station relationships enhances the predictive power of air quality assessments. Moreover, spatial variability affects pollutant concentrations by introducing het-

erogeneity in air pollution levels across different regions. For example, in urban environments, factors such as local emission sources, industrial zones, traffic patterns, street and building layouts, and meteorological conditions can create spatial variability. Research highlighted the use of Geographic Information Systems (GIS) and spatial regression models to quantify and map spatial variabilities, revealing how industrial activities and prevailing wind patterns contribute to such inconsistencies [57]. These analyses help identify pollution hotspots and inform strategies for air quality management. There is a growing recognition that in order to provide SA in realistic air quality predictions, there is a need to develop comprehensive frameworks that integrate spatial dependence more systematically into the calculation and interpretation of Sobol indices.

1.4 Thesis Overview

This thesis aims to contribute to the field of pollution dispersion modelling by the analysis and advancing the application of spatially-dependent analysis of Sobol indices tailored to air quality models. Incorporating spatially-dependent SA into air quality modelling provides a more nuanced understanding of how pollutants disperse and impact different spatial locations. By explicitly accounting for spatial variability, this work seeks to support the development of methodologies that enhance the understanding of input-output relationships across heterogeneous spatial domains. In doing so, it contributes a novel perspective to the field, demonstrating how spatially resolved Sobol indices can uncover nuanced patterns of sensitivity that traditional, spatially-independent techniques may overlook.

The central research question addressed in this thesis is: Which advection-diffusion model parameters are the most significant contributors to uncertainty in pollutant concentration across a spatial domain, and how does this influence vary regionally? This question is embedded within a broader inquiry into how spatially resolved SA can enhance our understanding of model behaviour in environmental systems, specifically air quality models. As environmental phenomena are inherently spatial, an aggregated perspective may limit the depth and reliability of uncertainty assessments. Here, a aggregated perspective refers to summarising model sensitivity using a single value over the certain regions of domain, which can obscure important spatial variations.

The significance of this research lies in its potential to inform more precise and geographically targeted air quality management strategies. Understanding how dominant parameters

vary spatially can help identify region-specific drivers of model uncertainty, guide monitoring efforts, and improve source attribution. This thesis focuses on the spatially resolved GSA of air quality models using Sobol indices.

To address this overarching question, the specific objectives of this research are as follows:

1. To identify and quantify the dominant contributors to uncertainty in pollutant concentrations across the spatial domain
2. To investigate how spatially distributed Sobol indices can provide more informative insights by assessing how these contributions vary in different regions.
3. To develop and apply surrogate modelling techniques for efficient estimation of Sobol indices and their associated confidence intervals, with a focus on preserving spatial patterns of sensitivity.
4. To assess the trade-offs between computational efficiency and fidelity introduced by surrogate approximations, evaluating their ability to retain spatial features of the sensitivity landscape while reducing computational burden.

Through these objectives, the thesis aims to advance the methodological foundations of spatially-dependent SA and contribute practical tools and insights for its application in air quality modelling. This spatial insight is particularly valuable for real-world applications such as optimising sensor placement for real-time monitoring, refining model calibration in regionally diverse environments, and designing location-specific emission control strategies.

The thesis is structured to progressively develop and apply a novel spatially-dependent SA framework. We proceed in Chapter 2 with a detailed account of the mathematical background necessary for what follows. This includes a discussion of the modelling assumptions underlying the steady-state advection-diffusion PDE, the numerical solution strategy based on the Streamline Upwind/Petrov–Galerkin (SUPG) method, and the implementation of QMC-based GSA. This chapter also elaborates on the methodological details of bootstrapping for confidence interval estimation. Chapter 3 introduces a one-dimensional (1D) test case designed to verify the spatially-dependent GSA framework within a simplified setting. Part of the material in this chapter has been published in [41]. Chapter 4 extends the methodology to a 2D domain to test computational performance and scalability in more realistic conditions. In this chapter, enhancements to Sobol index estimation are developed, including the introduction of neural network surrogate modelling. Chapter 5 applies the full framework to an idealised urban air quality model incorporating additional physical realism, such as turbulent eddy diffusion,

atmospheric stability, and mixing layer dynamics. Finally, we summarise the key findings of the thesis in the Chapter 6, reflecting on the methodological contributions and the insights gained from spatially resolved SA in air quality modelling. We also outline the limitations of the current study and suggests avenues for future work.

Chapter 2

Mathematical Background

In this chapter, we present a detailed review of key mathematical and computational background material required for this thesis. We describe a simplified model problem on given 1D and 2D domains in Section 2.1 and Section 2.3, respectively, based on steady-state advection-diffusion boundary value problems (BVPs). We then present the numerical methods employed to approximate the solutions of the BVPs under consideration. In particular, the Streamline Upwind Petrov–Galerkin (SUPG) finite element method is adopted due to its stabilising properties for advection-dominated problems. The formulation and implementation of the SUPG method are described in detail for the one-dimensional case in Section 2.2, and extended to two dimensions in Section 2.4. In this thesis, we study pollutant dispersion models as uncertainty quantification (UQ) problems and conduct global sensitivity analysis (GSA). In Section 2.5 we present a variance-based GSA measure called Sobol indices, which we use throughout the rest of this thesis. We also describe how bootstrapping is used to calculate confidence intervals (CIs) of estimates of Sobol indices in Section 2.6.

2.1 A Simplified 1D Pollutant Dispersion Model

Suppose we have a pollutant in the 1D domain $\Omega_1 = [0, b]$, which originates from sources at points $x_{s_1}, x_{s_2} \in \Omega_1$ with source rates s_1 and s_2, respectively. We assume the pollutant is transported through advection with a wind speed $a > 0$ and diffusion at a rate $D > 0$. With

these assumptions, steady-state pollutant dispersion can be modelled using the simple BVP

$$\begin{aligned} -D\frac{d^2u}{dx^2} + a\frac{du}{dx} &= s(x) \text{ in } \Omega_1, \\ u(0) = 0, \;\; u(b) &= 0 \end{aligned} \tag{2.1}$$

where

$$s(x) = \frac{s_1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x - x_{s_1})^2}{\sigma^2}\right\} + \frac{s_2}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x - x_{s_2})^2}{\sigma^2}\right\} \tag{2.2}$$

is a sum of continuous Gaussian source functions over a length scale σ, which governs the spatial extent of each source. Smaller σ values approximate a more localised source, while larger σ values produce a smoother and more spatially distributed release. This parameter directly influences both the peak pollutant concentration and its spatial dispersion within the domain. Gaussian sources are chosen to model a spatially distributed release of the pollutant, which can better represent realistic emission profiles than idealised point sources.

We assume that the boundary conditions are implemented sufficiently far from both source locations such that there is zero pollutant concentration at the boundaries. Here, "sufficiently far" implies that the distance between each source and the boundary is large enough to ensure that the concentration field decays to negligible levels before reaching the domain limits. Under these conditions, we have observed no evidence of interference with the interior flow or concentration fields, such as boundary layer distortion or oscillatory behaviour, when applying Dirichlet boundary conditions.

Other types of boundary conditions can also be considered depending on the problem setup. For example, zero Neumann ($\partial u/\partial n = 0$) conditions on outflow boundaries can represent open-domain behaviour by allowing pollutant transport to exit without imposing a concentration gradient. In this study, the Dirichlet boundaries are sufficient and do not introduce any noticeable numerical artefacts.

In (2.1), the wind speed, diffusivity, and source rates are presented as fixed parameters that quantify distinct physical processes within the model. In practice, these quantities are known, but subject to uncertainty. To address the inherent unpredictability, we adopt a probabilistic approach, where we assume that the parameters a, D, s_1, and s_2 are continuous random variables with associated probability distribution functions. This means that the pollutant concentration u is also continuous random variable. By sampling parameters from probability distribution functions, we can carry out a parametric study that enables us to interpret the model as a UQ problem. This probabilistic treatment allows for a realistic representation of the

uncertainty associated with the parameters, acknowledging their variable nature in practical scenarios.

The variability in wind speed, a pivotal parameter that influences pollutant advection, is modelled by a Weibull distribution, following [50]. The Weibull distribution is chosen for its adaptability in representing a variety of wind speed behaviours observed in natural settings. To account for uncertainties in diffusivity, we employ a lognormal distribution for modelling diffusion. This choice ensures non-negativity and positive skewness, which aligns with physical conditions governing diffusivity, offering a flexible means to model different levels of uncertainty while maintaining adherence to the model's physical constraints. We use this choice of diffusion to study the 1D BVP (2.1), and later in a 2D setting, described in Section (2.3). Finally, the uncertainties associated with source rates are modelled using a uniform distribution for simplicity, assuming equal likelihood for all values within a specified range. In summary, we characterise the uncertainty in the model parameters by sampling from probability distribution functions as follows:

$$\begin{aligned} a &\sim \text{Weibull}(\lambda, k), \qquad D \sim \text{Lognormal}(\mu_D, \sigma_D), \\ s_1 &\sim \text{Uniform}(\alpha_1, \beta_1), \quad s_2 \sim \text{Uniform}(\alpha_2, \beta_2), \end{aligned} \tag{2.3}$$

where λ is a scale parameter and k is a shape parameter, the lognormal mean is denoted by μ_D and the lognormal standard deviation is denoted by σ_D. Note that $\alpha_1, \alpha_2, \beta_1, \beta_2 > 0$. Examples of the probability distributions used are given in Figure 3.2.

For the Weibull distribution, the shape parameter k controls the shape of the distribution, such that a higher value k indicates that the wind speed values are left skewed and a lower value k suggests that the wind speeds are right skewed. We note that when $k \approx 3$, the Weibull distribution approximates the normal distribution. The scale parameter λ reflects the variability within the wind speed data. Adjusting λ affects the spread of the distribution, with higher values of λ indicating a stretched distribution with a lower peak. Conversely, smaller values of λ indicate a narrower distribution with an increased peak. The Weibull distribution has mean $\lambda\Gamma(1+1/k)$ and variance $\lambda^2 \left[\Gamma(1+2/k) - (\Gamma(1+2/k))^2\right]$, where $\Gamma(\cdot)$ denotes the gamma function.

The lognormal distribution indicates that the logarithm of the diffusion data is normally distributed. Increasing μ_D shifts the distribution to the right, which means that the median diffusivity increases. Conversely, decreasing μ_D shifts it to the left, resulting in a lower median. Larger values of σ_D increase the spread of the lognormal distribution, while smaller values of σ_D

result in a more concentrated distribution. The lognormal distribution has mean $\exp(\mu_D+\sigma_D^2/2)$ and variance $\left[\exp(\sigma_D^2)-1\right]\exp(2\mu_D+\sigma_D^2)$.

The uniform distribution for source strength indicates that the sources rates between α_1 and β_1 are equally likely for s_1. The same can be said for the source rate s_2 evenly distributed between α_2 and β_2. The uniform distribution has mean $(\alpha_i+\beta_i)/2$ and variance $(\beta_i-\alpha_i)^2/12$, for $i=1,2$.

2.2 Streamline Upwind Petrov-Galerkin Approximation

A numerical approximation of the solution to the steady-state BVP (2.1) can be obtained using a Streamline Upwind Petrov-Galerkin (SUPG) finite element method (see, e.g. [15, §6.3.2]). The Péclet number is a dimensionless parameter that quantifies the relative importance of advection and diffusion in transport processes. Defined as the ratio of advective to diffusive transport rates, it indicates whether transport is more strongly driven by advection or by diffusion. In the discrete setting, this concept is localised to the computational mesh through the mesh Péclet number, which measures the balance between advection and diffusion at the element level. Let P_h denote the mesh Péclet number, where

$$P_h = \frac{ha}{2D}, \tag{2.4}$$

for $a>0$, where h is a measure of element size which we define below.

2.2.1 Weak Formulation in 1D

We begin by constructing the weak form of the model (2.1). The weak formulation of a differential equation involves multiplying the equation by a smooth test function and integrating by parts over the domain. This approach naturally leads to Sobolev spaces, which include functions whose derivatives exist in a weak sense and are integrable, allowing solutions that may not be classically differentiable to be treated rigorously in the weak formulation of PDEs. The weak form is useful because it broadens the set of admissible solutions and provides a foundation for stable, convergent numerical approximations like FEMs.

First, we define the space in which the weak solution exists. In order to do this, we introduce

the space of functions that are Lebesgue square integrable, given by

$$L_2(\Omega_1) = \left\{ u : \Omega_1 \Rightarrow \mathbb{R} \middle| \int_{\Omega_1} u^2 \, dx < \infty \right\}, \tag{2.5}$$

where the L_2 measure $\|u\| = \left(\int_{\Omega_1} u^2 \, dx \right)^{1/2}$ holds. Then the function space

$$H^1_{E_0}(\Omega_1) = \left\{ u : \Omega_1 \Rightarrow \mathbb{R} \middle| u, \frac{du}{dx} \in L_2(\Omega_1), u(0) = u(b) = 0 \right\}$$

is the space in which the solution of the weak form exists. That is, the weak solution is well defined if all first derivatives of u are in $L_2(\Omega_1)$ and the Dirichlet boundary conditions are satisfied [15].

The weak form is obtained by multiplying the pollutant dispersion model PDE in (2.1) by a test function $v = v(x) \in H^1_{E_0}(\Omega_1)$ and integrating over the domain Ω_1, where $s = s(x)$:

$$-D \int_{\Omega_1} \frac{d^2u}{dx^2} v \, dx + a \int_{\Omega_1} \frac{du}{dx} v \, dx = \int_{\Omega_1} sv \, dx.$$

Using integration by parts, we get

$$-D \left[\frac{du}{dx} v \right]_0^b + D \int_{\Omega_1} \frac{du}{dx} \frac{dv}{dx} \, dx + a \int_{\Omega_1} \frac{du}{dx} v \, dx = \int_{\Omega_1} sv \, dx.$$

Then, the weak form becomes: find $u \in H^1_{E_0}(\Omega_1)$ such that

$$D \int_{\Omega_1} \frac{du}{dx} \frac{dv}{dx} \, dx + a \int_{\Omega_1} \frac{du}{dx} v \, dx = \int_{\Omega_1} sv \, dx, \quad \text{for all } v \in H^1_{E_0}(\Omega_1). \tag{2.6}$$

2.2.2 Galerkin Finite Element Method in 1D

To approximate the solution of the weak formulation (2.6), we must discretise the continuous 1D domain Ω_1 into elements. We use a uniform mesh with N elements, where the element size is $h = b/N$.

We now define a finite-dimensional test space $S^h_0 \subset H^1_{E_0}(\Omega_1)$. The Galerkin method for the pollutant dispersion problem is: find $u_h \in S^h_0$ such that

$$D \int_{\Omega_1} \frac{du_h}{dx} \frac{dv_h}{dx} \, dx + a \int_{\Omega_1} \frac{du_h}{dx} v_h \, dx = \int_{\Omega_1} sv_h \, dx \quad \text{for all } v_h \in S^h_0.$$

Now assume that $S^h_0 \subset H^1_{E_0}(\Omega_1)$ is a finite n-dimensional vector space, spanned by linear

basis functions $\{\phi_1(x), \ldots, \phi_n(x)\}$. These basis functions are continuous and piecewise linear over the computational mesh, consistent with the standard finite element formulation used throughout this thesis. The finite element approximation $u_h \in S_0^h$ is then uniquely associated with the vector $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)^\top$ of real coefficients in the expansion

$$u_h = \sum_{j=1}^{n} u_j \phi_j. \tag{2.7}$$

For the Galerkin method, we choose n test functions equal to the basis functions ϕ_i, where $i = 1, \ldots, n$. Substituting u for u_h in the weak form then leads to the following problem: find u_j where $j = 1, \ldots, n$ such that

$$\sum_{j=1}^{n} u_j \left[D \int_{\Omega_1} \frac{d\phi_j}{dx} \frac{d\phi_i}{dx}\, dx + a \int_{\Omega_1} \frac{d\phi_j}{dx} \phi_i dx \right] = \int_{\Omega_1} s\phi_i\, dx$$

for $i = 1, \ldots, n$. This can be written in matrix form as the linear system of equations

$$[DK + aC]\boldsymbol{u} = \boldsymbol{s},$$

where K, C, and $\boldsymbol{s}$ have entries

$$k_{ij} = \int_{\Omega_1} \frac{d\phi_j}{dx} \frac{d\phi_i}{dx}\, dx, \quad c_{ij} = \int_{\Omega_1} \frac{d\phi_j}{dx} \phi_i\, dx, \quad s_i = \int_{\Omega_1} s\phi_i\, dx, \tag{2.8}$$

for $i, j = 1 \ldots n$, respectively, and the convection matrix C is skew-self-adjoint [15].

2.2.3 Streamline Upwind Petrov-Galerkin Method in 1D

As the parameters in (2.3) are sampled over a set of different probability distribution functions, there will be inherent variability in the mesh Péclet number for a fixed h. That is, here P_h is a function of random variables, and therefore it can take a wide range of values. Standard Galerkin finite element solutions of (2.1) are known to suffer from oscillations resulting in inaccurate and unstable solutions when $P_h > 1$. Hence, it is important to consider approximation methods that are robust when parameter variations can lead to a high degree of variability in the mesh Péclet number. Using a SUPG discretisation ensures that our numerical approximation technique is robust to the possibility of advection-dominated problems. In this section, we describe the 1D SUPG method used in this thesis to approximate the solution to the pollutant dispersion problem (2.1).

To address this issue of oscillations that arise in numerical solutions, we can incorporate additional diffusion in the direction of the wind if the mesh Péclet number $P_h > 1$. This can be done using a Petrov-Galerkin method, where we use a test space that differs from the trial space. For the BVP (2.1), we take the test space to be the space spanned by functions of the form

$$v + \tau a \frac{dv}{dx}, \tag{2.9}$$

where $v \in H^1_{E_0}(\Omega_1)$ and $\tau \in \mathbb{R}$ is a constant stabilisation parameter. This leads to the Streamline Upwind Petrov-Galerkin (SUPG) method.

Multiplying (2.1) by the test functions (2.9) gives

$$\begin{aligned}
&- D \int_{\Omega_1} \frac{d^2u}{dx^2} \left(v + \tau a \frac{dv}{dx} \right) dx + a \int_{\Omega_1} \frac{du}{dx} \left(v + \tau a \frac{dv}{dx} \right) dx = \int_{\Omega_1} s \left(v + \tau a \frac{dv}{dx} \right) dx \\
\Rightarrow &- D \int_{\Omega_1} \frac{d^2u}{dx^2} v \, dx + a \int_{\Omega_1} \frac{du}{dx} v \, dx - \tau a D \int_{\Omega_1} \frac{d^2u}{dx^2} \frac{dv}{dx} \, dx + \tau a^2 \int_{\Omega_1} \frac{du}{dx} \frac{dv}{dx} \, dx \\
&= \int_{\Omega_1} sv \, dx + \tau a \int_{\Omega_1} s \frac{dv}{dx} \, dx.
\end{aligned} \tag{2.10}$$

We note that the first two terms on the LHS (2.10) and the first term on the RHS (2.10) are exactly the same as in the Galerkin case. Therefore, by using integration by parts, we obtain

$$\begin{aligned}
&D \int_{\Omega_1} \frac{du}{dx} \frac{dv}{dx} \, dx + a \int_{\Omega_1} \frac{du}{dx} v \, dx - \tau a D \int_{\Omega_1} \frac{d^2u}{dx^2} \frac{dv}{dx} \, dx + \tau a^2 \int_{\Omega_1} \frac{du}{dx} \frac{dv}{dx} \, dx \\
&= \int_{\Omega_1} sv \, dx + \tau a \int_{\Omega_1} s \frac{dv}{dx} \, dx.
\end{aligned}$$

We let

$$-\tau a D \int_{\Omega_1} \frac{d^2u}{dx^2} \frac{dv}{dx} \, dx \equiv -\tau a D \sum_E \int_E \frac{d^2u}{dx^2} \frac{dv}{dx} \, dx$$

be a sum over all of the elements. In order to construct a discrete formulation, we replace u and v by u_h and v_h, respectively, in the same function spaces as before.

Using the expansion (2.7) as before with the same basis $\{\phi_1, \phi_2, \ldots, \phi_n\}$ and $v = \phi_i$, we obtain the problem: find u_j, where $j = 1, \ldots, n$ such that

$$\begin{aligned}
&\sum_{j=1}^{n} u_j \left[D \int_{\Omega_1} \frac{d\phi_j}{dx} \frac{d\phi_i}{dx} \, dx + a \int_{\Omega_1} \frac{d\phi_j}{dx} \phi_i \, dx + \tau a^2 \int_{\Omega_1} \frac{d\phi_j}{dx} \frac{d\phi_i}{dx} \, dx - \tau a D \sum_E \int_E \frac{d^2\phi_j}{dx^2} \frac{d\phi_i}{dx} \, dx \right] \\
&\qquad = \int_{\Omega_1} s\phi_i \, dx + \tau a \int_{\Omega_1} s \frac{d\phi_i}{dx} \, dx
\end{aligned}$$

for $i = 1, \ldots, n$.

In this thesis, we choose to use piecewise linear basis functions to construct the numerical solution of the BVP (2.1) and hence

$$\sum_E \int_E \frac{d^2\phi_j}{dx^2} \frac{d\phi_i}{dx}\, dx = 0$$

due to the second derivative term. Therefore, the final problem is: find u_j where $j = 1, \ldots, n$ such that

$$\sum_{j=1}^{n} u_j \left[D \int_{\Omega_1} \frac{d\phi_j}{dx} \frac{d\phi_i}{dx}\, dx + a \int_{\Omega_1} \frac{d\phi_j}{dx} \phi_i\, dx + \tau a^2 \int_{\Omega_1} \frac{d\phi_j}{dx} \frac{d\phi_i}{dx}\, dx \right] = \int_{\Omega_1} s\phi_i\, dx + \tau a \int_{\Omega_1} s \frac{d\phi_i}{dx}\, dx$$

for $i = 1, \ldots, n$.

We assume for now that the wind speed a is constant. In order to improve the computational efficiency when using the SUPG method applied to UQ problems and GSA, we write the linear system of equations arising from the SUPG method as

$$\left[(D + \tau a^2)K + aC\right] \boldsymbol{u} = s_1 \boldsymbol{l_1} + s_2 \boldsymbol{l_2}, \tag{2.11}$$

where the entries for K and C are given in (2.8) and

$$(\boldsymbol{l_1})_i = \int_{\Omega_1} \phi_i \exp\left\{ -\frac{(x - x_{s_1})^2}{\sigma^2} \right\} dx + \tau a \int_{\Omega_1} \exp\left\{ -\frac{(x - x_{s_1})^2}{\sigma^2} \right\} \frac{d\phi_i}{dx}\, dx,$$

$$(\boldsymbol{l_2})_i = \int_{\Omega_1} \phi_i \exp\left\{ -\frac{(x - x_{s_2})^2}{\sigma^2} \right\} dx + \tau a \int_{\Omega_1} \exp\left\{ -\frac{(x - x_{s_1})^2}{\sigma^2} \right\} \frac{d\phi_i}{dx}\, dx,$$

for $i = 1, \ldots, n$. We take τ take to be [49]

$$\tau = \begin{cases} \coth\left(\frac{ah}{2}\right) - \frac{2}{ah} & \text{for } P_h \geq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.12}$$

For advection-dominated regimes, where $P_h \geq 1$, (2.12) introduces numerical diffusion to suppress non-physical oscillations and smoothly interpolates between no stabilization and full upwind, depending on the local element Péclet number.

2.3 A Simplified 2D Pollutant Dispersion Model

In this section, we extend the BVP described in Section 2.1 to a 2D domain. We will study this problem in Chapter 4. Modelling in this way provides more realistic spatial properties. We introduce a pollutant to the 2D domain $\Omega_2 = \{(x, y) : 0 \leq x \leq b_x, 0 \leq y \leq b_y\}$ at the locations $\boldsymbol{\mu_{s_1}} = (\mu_{x_1}, \mu_{y_1})$ and $\boldsymbol{\mu_{s_2}} = (\mu_{x_2}, \mu_{y_2})$ such that

$$s(x,y) = \frac{s_1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-\left[(x-\mu_{x_1})^2 + (y-\mu_{y_1})^2\right]}{2\pi\sigma^2}\right\} + \frac{s_2}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-\left[(x-\mu_{x_2})^2 + (y-\mu_{y_2})^2\right]}{2\pi\sigma^2}\right\},$$

with scaling parameter σ. Wind speed is given by $a > 0$ in the direction $\boldsymbol{w} \in \mathbb{R}^2$ with $\|\boldsymbol{w}\| = 1$ and we assume $\operatorname{div} \boldsymbol{w} = 0$. Diffusion is modelled by a rate of $D > 0$. We therefore model steady-state pollutant dispersion with the BVP

$$\begin{aligned} -D\nabla^2 u + a\boldsymbol{w} \cdot \nabla u &= s(x,y) \text{ in } \Omega_2, \\ u &= 0 \text{ on } \partial\Omega_2, \end{aligned} \tag{2.13}$$

where $\partial\Omega_2$ denotes the boundary of the domain Ω_2. We assume that the pollutant sources are located sufficiently far from all domain boundaries such that Dirichlet boundary conditions hold.

Similarly to Section 2.1, we are interested in studying the BVP (2.13) as a UQ problem by modelling the quantities a, D, s_1 and s_2 as random variables to address the variability of these quantities in practical scenarios. We characterise the uncertainty in these quantities by sampling the parameters from the probability distribution functions given in (2.3). We conduct analysis of the UQ problem in Chapter 4.

2.4 Streamline Upwind Petrov-Galerkin Approximation

2.4.1 Weak Formulation in 2D

To construct the weak form of (2.13), we define the space in which the weak solution exists as

$$H^1_{E_0}(\Omega_2) = \left\{ u : \Omega_2 \Rightarrow \mathbb{R} \middle| u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y} \in L_2(\Omega_2), u = 0 \text{ on } \partial\Omega_2 \right\}, \tag{2.14}$$

where $L_2(\Omega_2)$ denotes the space of functions which are Lebesgue square integrable, given by (2.5). We obtain the weak form by multiplying (2.13) by a test function $v \in H^1_{E_0}(\Omega_2)$ and integrating over the domain:

$$-D\int_{\Omega_2}\left(\nabla^2 u\right)v\,d\Omega_2 + a\int_{\Omega_2}(\boldsymbol{w}\cdot\nabla u)\,v\,d\Omega_2 = \int_{\Omega_2} sv\,d\Omega_2, \tag{2.15}$$

where $s = s(x, y)$.

Given the identity

$$\left(\nabla^2 u\right)v = -\nabla u\cdot\nabla v + \operatorname{div}(v\nabla u),$$

we can write (2.15) as

$$\begin{aligned} &D\left[\int_{\Omega_2}\nabla u\cdot\nabla v\,d\Omega_2 - \int_{\Omega_2}\operatorname{div}(v\nabla u)\,d\Omega_2\right] + a\int_{\Omega_2}(\boldsymbol{w}\cdot\nabla u)v\,d\Omega_2 = \int_{\Omega_2} sv\,d\Omega_2 \\ \Rightarrow &D\int_{\Omega_2}\nabla u\cdot\nabla v\,d\Omega_2 + a\int_{\Omega_2}(\boldsymbol{w}\cdot\nabla u)v\,d\Omega_2 = \int_{\Omega_2} sv\,d\Omega_2 + D\int_{\Omega_2}\operatorname{div}(v\nabla u)\,d\Omega_2. \end{aligned} \tag{2.16}$$

From the Divergence Theorem, we have

$$\int_{\Omega_2}\operatorname{div}(v\nabla u)\,d\Omega_2 = \int_{\partial\Omega_2}(\nabla u\cdot n)\,v\,dl.$$

However, in the case of our Dirichlet problem, we have $v \in H^1_{E_0}(\Omega_2)$, so the final weak formulation is: find $u \in H^1_{E_0}(\Omega_2)$ such that

$$D\int_{\Omega_2}\nabla u\cdot\nabla v\,d\Omega_2 + a\int_{\Omega_2}(\boldsymbol{w}\cdot\nabla u)v\,d\Omega_2 = \int_{\Omega_2} sv\,d\Omega_2, \quad \text{for all } v \in H^1_{E_0}(\Omega_2). \tag{2.17}$$

2.4.2 Galerkin Method in 2D

A discrete weak formulation for the BVP (2.13) is constructed by discretising the domain Ω_2 using triangular partition of the domain, which we describe in Section 2.4.3. We define the finite-dimensional test space $S^h_0 \subset H^1_{E_0}(\Omega_2)$. We assume that $S^h_0 \subset H^1_{E_0}(\Omega_2)$ is a finite n-dimensional vector space spanned by standard linear basis functions $\{\phi_1(x, y), \phi_2(x, y), \ldots, \phi_n(x, y)\}$. Then, the Galerkin method for the pollutant dispersion problem in 2D is: find $u_h \in S^h_0$ such that

$$D\int_{\Omega_2}\nabla u_h\cdot\nabla v_h\,d\Omega_2 + a\int_{\Omega_2}(\boldsymbol{w}\cdot\nabla u_h)v_h\,d\Omega_2 = \int_{\Omega_2} sv_h\,d\Omega_2 \quad \text{for all } v_h \in S^h_0.$$

The finite element approximation $u_h \in S_0^h$ is then uniquely associated with the vector $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)^T$ of real coefficients in the expansion

$$u_h = \sum_{j=1}^{n} u_j \phi_j \quad \text{for } j = 1, \ldots, n.$$

Substituting for u_h in the weak form and using the n basis functions as test functions leads to the following problem: find $u_j, j = 1, \ldots, n$ such that

$$\sum_{j=1}^{n} u_j \left[D \int_{\Omega_2} \nabla\phi_j \cdot \nabla\phi_i \, d\Omega_2 + a \int_{\Omega_2} (\boldsymbol{w} \cdot \nabla\phi_j)\, \phi_i \, d\Omega_2 \right] = \int_{\Omega_2} s\phi_i \, d\Omega_2,$$

for $i = 1, \ldots, n$. This can be written in matrix form as the linear system of equations

$$[DK + aC]\boldsymbol{u} = \boldsymbol{s}$$

with

$$k_{ij} = \int_{\Omega_2} \nabla\phi_j \cdot \nabla\phi_i \, d\Omega_2, \;\; c_{ij} = \int_{\Omega_2} (\boldsymbol{w} \cdot \nabla\phi_j)\, \phi_i \, d\Omega_2, \;\; s_i = \int_{\Omega_2} s\phi_i \, d\Omega_2. \tag{2.18}$$

The stiffness matrix K is positive definite and symmetric. Let C be the convection matrix, and let $\boldsymbol{s}$ be the vector of source terms. In the case of the BVP (2.13), where we only have Dirichlet boundary conditions, the convection matrix C is skew-self-adjoint [15].

2.4.3 Domain Triangulation

In order to implement the Galerkin method, we discretise the domain into triangles. Suppose we have an element E with local nodes 1, 2 and 3 with (x, y) co-ordinates (x_1, y_1), (x_2, y_2), (x_3, y_3) and a reference triangle R with local nodes 1, 2 and 3 with (ξ, η) co-ordinates $(0, 0)$, $(0, 1)$, $(1, 1)$, where $h = b/N$ is the mesh size, b is the length the of the domain, and N is the number of nodes, assuming that we have some structured triangulation.

For each element E, we produce an element matrix. The element matrices are computed for each element separately and then assembled using a connectivity array to form the global matrix that represents the entire domain. To compute the element matrices, we define the

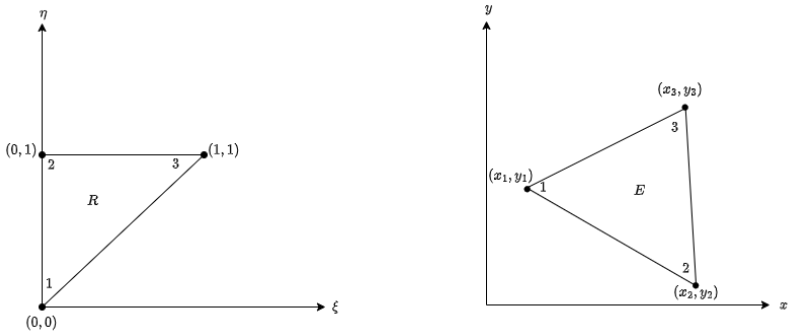


Figure 2.1: Mapping of the reference triangle R to a element E in the Galerkin method, with mesh size $h = b/N$.

element basis functions

$$\begin{aligned} \psi_1(\xi, \eta) &= 1 - \xi, \\ \psi_2(\xi, \eta) &= \xi - \eta, \\ \psi_3(\xi, \eta) &= \eta, \end{aligned} \tag{2.19}$$

on the reference element R. Then we can map to any general triangle with vertex coordinates (x_ν, y_ν), $\nu = 1, 2, 3$ by the change of variables

$$x(\xi, \eta) = \sum_{\nu=1}^{3} x_\nu \psi_\nu(\xi, \eta), \quad y(\xi, \eta) = \sum_{\nu=1}^{3} y_\nu \psi_\nu(\xi, \eta). \tag{2.20}$$

We note that the mappings (2.20) are differentiable. We define the Jacobian matrix

$$J = \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial x}{\partial \eta} \\ \frac{\partial y}{\partial \xi} & \frac{\partial y}{\partial \eta} \end{bmatrix} = \begin{bmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_2 & y_3 - y_2 \end{bmatrix}.$$

Then, we find the determinant of the Jacobian $|J|$ to be

$$|J| = \begin{vmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial x}{\partial \eta} \\ \frac{\partial y}{\partial \xi} & \frac{\partial y}{\partial \eta} \end{vmatrix} = \frac{\partial x}{\partial \xi}\frac{\partial y}{\partial \eta} - \frac{\partial x}{\partial \eta}\frac{\partial y}{\partial \xi} = (x_2 - x_1)(y_3 - y_2) - (x_3 - x_2)(y_2 - y_1).$$

We detail the construction of the element matrices K_e, C_e, and the load vector $\boldsymbol{s}_e$ below, followed by the assembly of the global matrices.

Stiffness Matrix

Consider now the element matrix entry

$$k_{ij} = \int_E \nabla\psi_i \cdot \nabla\psi_j \, dE = \int_E \left[\frac{\partial\psi_i}{\partial x}\frac{\partial\psi_j}{\partial x} + \frac{\partial\psi_i}{\partial y}\frac{\partial\psi_j}{\partial y} \right] dE,$$

where $i, j = 1, 2, 3$. We consider the terms in the integrand, where using the chain rule, we have

$$\frac{\partial\psi_i}{\partial x} = \frac{\partial\psi_i}{\partial\xi}\frac{\partial\xi}{\partial x} + \frac{\partial\psi_i}{\partial\eta}\frac{\partial\eta}{\partial x}, \quad \frac{\partial\psi_i}{\partial y} = \frac{\partial\psi_i}{\partial\xi}\frac{\partial\xi}{\partial y} + \frac{\partial\psi_i}{\partial\eta}\frac{\partial\eta}{\partial y}. \tag{2.21}$$

To find the derivatives $\partial\xi/\partial x$, $\partial\xi/\partial y$ and $\partial\eta/\partial x$, $\partial\eta/\partial x$, we use the relationship

$$\begin{bmatrix} \frac{\partial\xi}{\partial x} & \frac{\partial\xi}{\partial y} \\ \frac{\partial\eta}{\partial x} & \frac{\partial\eta}{\partial y} \end{bmatrix} = \begin{bmatrix} \frac{\partial x}{\partial\xi} & \frac{\partial x}{\partial\eta} \\ \frac{\partial y}{\partial\xi} & \frac{\partial y}{\partial\eta} \end{bmatrix}^{-1} = \frac{1}{|J|} \begin{bmatrix} \frac{\partial y}{\partial\eta} & -\frac{\partial x}{\partial\eta} \\ -\frac{\partial y}{\partial\xi} & \frac{\partial x}{\partial\xi} \end{bmatrix}$$

to obtain

$$\frac{\partial\xi}{\partial x} = \frac{1}{|J|}\frac{\partial y}{\partial\eta}, \quad \frac{\partial\xi}{\partial y} = -\frac{1}{|J|}\frac{\partial x}{\partial\eta}, \quad \frac{\partial\eta}{\partial x} = -\frac{1}{|J|}\frac{\partial y}{\partial\xi}, \quad \frac{\partial\eta}{\partial y} = \frac{1}{|J|}\frac{\partial x}{\partial\xi}. \tag{2.22}$$

Hence, we can write each integrand in terms of ξ and η. Then we obtain entries of the element stiffness matrix using the calculation

$$\begin{aligned} k_{ij} = \frac{1}{|J|}\int_0^1 \int_0^\xi \Bigg[&\left(\frac{\partial\psi_i}{\partial\xi}\frac{\partial y}{\partial\eta} - \frac{\partial\psi_i}{\partial\eta}\frac{\partial y}{\partial\xi} \right)\left(\frac{\partial\psi_j}{\partial\xi}\frac{\partial y}{\partial\eta} - \frac{\partial\psi_j}{\partial\eta}\frac{\partial y}{\partial\xi} \right) \\ &+ \left(-\frac{\partial\psi_i}{\partial\xi}\frac{\partial x}{\partial\eta} + \frac{\partial\psi_i}{\partial\eta}\frac{\partial x}{\partial\xi} \right)\left(-\frac{\partial\psi_j}{\partial\xi}\frac{\partial x}{\partial\eta} + \frac{\partial\psi_i}{\partial\eta}\frac{\partial x}{\partial\xi} \right) \Bigg] d\eta d\xi. \end{aligned} \tag{2.23}$$

Convection Matrix

Similarly, for the convection matrix C, we have

$$c_{ij} = \int_E (\boldsymbol{w} \cdot \nabla\psi_j)\,\psi_i \, dE = \int_E \left[w_x\psi_i\frac{\partial\psi_j}{\partial x} + w_y\psi_i\frac{\partial\psi_j}{\partial y} \right] dE,$$

where $i, j = 1, 2, 3$, and $\boldsymbol{w} = [w_x, w_y]$. Applying a similar approach to deriving k_{ij} by using the chain rule (2.21) to map onto reference element T and implementing the relationship (2.22), we obtain the entries of the element convection matrix

$$c_{ij} = \frac{1}{|J|}\int_0^1 \int_0^\xi \left[w_x\psi_i\left(\frac{\partial\psi_j}{\partial\xi}\frac{\partial y}{\partial\eta} - \frac{\partial\psi_j}{\partial\eta}\frac{\partial y}{\partial\xi} \right) + w_y\psi_i\left(-\frac{\partial\psi_j}{\partial\xi}\frac{\partial x}{\partial\eta} + \frac{\partial\psi_j}{\partial\eta}\frac{\partial x}{\partial\xi} \right) \right] d\eta d\xi. \tag{2.24}$$

Source Vector

To construct the entries of the element source vector, we need to evaluate the integral

$$\int_E s\psi_i \, dE, \quad i = 1, 2, 3.$$

This can be written as

$$\int_E s\psi_i \, dE = |J| \int_0^1 \int_0^\xi s\psi_i \, d\eta d\xi, \quad i = 1, 2, 3.$$

Assembled Matrices

For each element E, we now assemble the element matrices K_e, C_e and $\boldsymbol{s}_e$ constructed above into global matrices K, C and $\boldsymbol{s}$ using a connectivity array to map the local elements to the global domain. Before assembly, the global matrices are initialised to zero. The assembly is implemented in practice using nested for loops, shown by the following psuedocode.

```
for k=1:nels
   for i=1:3
      for j=1:3
         K(NN(k,i),NN(k,j))=K(NN(k,i),NN(k,j))+kel(i,j);
         C(NN(k,i),NN(k,j))=C(NN(k,i),NN(k,j))+cel(i,j);
      end
      S(NN(k,i))=S(NN(k,i))+sel(i);
   end
end
```

Here `NN(k,i)` denotes the global node number of element k with local node number i, `nels` denotes the number of elements, and `kel`, `cel`, and `sel` are the entries of the matrices K_e, C_e, and $\boldsymbol{s}_e$.

2.4.4 Streamline Upwind Petrov-Galerkin Method in 2D

As stated in Section 2.2.3, the numerical solutions with the Galerkin method of advection-diffusion problems can have oscillations if the mesh Péclet number is greater than one. This can occur in cases where the problem is advection-dominated or if the mesh refinement is too

coarse. We define the mesh Péclet number in the 2D case by

$$P_h = \frac{ah^{el}}{2D}, \tag{2.25}$$

where h is a measure of element size. In order to deal with such oscillations, we implement the Streamline Upwind Petrov-Galerkin (SUPG) method, where we use a distinct test space and trial space. For the pollutant dispersion problem (2.13), we take the test space to be spanned by the functions of the form

$$v + \tau a\boldsymbol{w} \cdot \nabla v, \tag{2.26}$$

where $v \in H^1_{E_0}(\Omega_2)$ and $\tau > 0$ is a constant stabilisation parameter, given by

$$\tau = \frac{\delta h^{el}}{a||\boldsymbol{w}||}, \tag{2.27}$$

where $\delta > 0$ is a parameter and $||\boldsymbol{w}|| = 1$. We choose δ to be

$$\delta = \max\left(0, \frac{1}{2} - \frac{2D}{ah^{el}}\right).$$

This choice of δ provides an element-wise stabilisation that correctly transitions between the Galerkin method for diffusion-dominated cases, where $P_h \leq 1$, and the optimal convective scaling for advection-dominated cases, $P_h > 1$, based on the approximate solution of the 1D boundary-layer (bubble) problem as detailed in [15, §6.3.2]. For the uniform mesh used in this thesis, P_h is constant across all elements for any given choice of a and D. Hence, δ is also constant across all elements.

The SUPG solution satisfies an *a priori* error estimate under this choice of τ given by

$$||u - u_h||_{sd} \leq Ch^{3/2}||\nabla^2 u||,$$

where C is a constant independent of the diffusion coefficient D, $||\nabla^2 u||$ measures the $\mathcal{H}^2$ regularity (see below), and $||\cdot||_{sd}$ is the streamline diffusion norm given by

$$||v||_{sd} := \left(D||\nabla v||^2 + \tau||a\boldsymbol{w} \cdot \nabla v||^2\right)^{1/2}.$$

A problem is defined to be $\mathcal{H}^2$ regular if, for any square-integrable input $f \in L_2(\Omega)$, its solution

u has square-integrable second derivatives and satisfies a bound of the form

$$\|u\|_{2,\Omega} \leq C_\Omega \|f\|.$$

For details of definition of the $\mathcal{H}^2$ norm and $\mathcal{H}^2$ regularity see [15, §1.5.1]. The above *a priori* estimate is valid under specific assumptions on the problem and discretisation. In particular, it holds for constant-coefficient convection–diffusion equations where both the advection field $a\boldsymbol{w}$ and diffusion coefficient D are uniform across the domain, and the finite element mesh is uniform with linear or bilinear basis functions. Moreover, the *a priori* estimate applies in the advection-dominated regime, characterised by a mesh size satisfying $h > 2D$ (or equivalently, a mesh Péclet number $P_h = h|\boldsymbol{w}|/(2D) > 1$). These conditions ensure stability and controlled error in the advection-dominated regime, and it motivates the choice of τ which enhances stability along streamlines without compromising accuracy. For further details and proof, see [15, Theorem 6.6], which is a special case of the general analysis in [21, pp. 302ff.].

In our problem, we hold the advective field and diffusivity constant in any given application of the SUPG method, which we formulated with linear basis functions. The conditions on τ ensure that in diffusion-dominated regimes, $\delta = 0$. Note that when $\delta = 0$, we reproduce the standard Galerkin method.

Multiplying (2.13) by the test functions (2.26) gives

$$\begin{aligned}
&- D \int_{\Omega_2} \left(\nabla^2 u\right) (v + \tau a \boldsymbol{w} \cdot \nabla v)\, d\Omega_2 + a \int_{\Omega_2} (\boldsymbol{w} \cdot \nabla u)(v + \tau \boldsymbol{w} \cdot \nabla v)\, d\Omega_2 \\
&= \int_{\Omega_2} s(v + \tau a \boldsymbol{w} \cdot \nabla v)\, d\Omega_2 \\
\Rightarrow &- D \int_{\Omega_2} \left(\nabla^2 u\right) v\, d\Omega_2 + a \int_{\Omega_2} (\boldsymbol{w} \cdot \nabla u) v\, d\Omega_2 \\
&- aD \int_{\Omega_2} \left(\nabla^2 u\right) (\tau \boldsymbol{w} \cdot \nabla v)\, d\Omega_2 + a \int_{\Omega_2} (\boldsymbol{w} \cdot \nabla u)(\tau \boldsymbol{w} \cdot \nabla v)\, d\Omega_2 \\
&= \int_{\Omega_2} sv\, d\Omega_2 + \int_{\Omega_2} s(\tau a \boldsymbol{w} \cdot \nabla v)\, d\Omega_2. \qquad (2.28)
\end{aligned}$$

The first two LHS terms of (2.28) and the first RHS term of (2.28) here are exactly the same as in the Galerkin case, which is shown in (2.16). So this can be treated in the same way as in

Section 2.4.1 to obtain

$$\begin{aligned}
&D\int_{\Omega_2}\nabla u\cdot\nabla v\,d\Omega_2 + a\int_{\Omega_2}(\boldsymbol{w}\cdot\nabla u)v\,d\Omega_2\\
&-aD\int_{\Omega_2}\left(\nabla^2 u\right)(\tau\boldsymbol{w}\cdot\nabla v)\,d\Omega_2 + a\int_{\Omega_2}(\boldsymbol{w}\cdot\nabla u)(\tau\boldsymbol{w}\cdot\nabla v)\,d\Omega_2\\
&=\int_{\Omega_2} sv\,d\Omega_2 + D\int_{\partial\Omega_2}\frac{\partial u}{\partial n}v\,dl + a\int_{\Omega_2} s(\tau\boldsymbol{w}\cdot\nabla v)\,d\Omega_2.
\end{aligned}$$

As in the Galerkin case, we assume Dirichlet boundary conditions on u. Therefore, $v = 0$ on $\partial\Omega_2$. Hence, this reduces to

$$\begin{aligned}
&D\int_{\Omega_2}\nabla u\cdot\nabla v\,d\Omega_2 + a\int_{\Omega_2}(\boldsymbol{w}\cdot\nabla u)v\,d\Omega_2\\
&-aD\int_{\Omega_2}\left(\nabla^2 u\right)(\tau\boldsymbol{w}\cdot\nabla v)\,d\Omega_2 + a\int_{\Omega_2}(\boldsymbol{w}\cdot\nabla u)(\tau\boldsymbol{w}\cdot\nabla v)\,d\Omega_2\\
&=\int_{\Omega_2} sv\,d\Omega_2 + a\int_{\Omega_2} s(\tau\boldsymbol{w}\cdot\nabla v)\,d\Omega_2.
\end{aligned}\tag{2.29}$$

We now require a discrete formulation of (2.29). This is done by replacing u and v by u_h and v_h. We now need to consider the terms:

$$-aD\int_{\Omega_2}\left(\nabla^2 u_h\right)(\tau\boldsymbol{w}\cdot\nabla v_h)\;d\Omega_2,\quad a\int_{\Omega_2}(\boldsymbol{w}\cdot\nabla u_h)(\tau\boldsymbol{w}\cdot\nabla v_h)\;d\Omega_2,$$
$$\text{and}\quad a\int_{\Omega_2} s\,(\tau\boldsymbol{w}\cdot\nabla v_h)\;d\Omega_2.$$

Since τ depends only on quantities defined at the element level, such as h^{el}, a, D, and since we employ a uniform mesh in which the element size h^{el} is the same for all elements, these quantities are fixed within each element. Consequently, the stabilisation parameter τ is constant over each element. This allows the first integral to be replaced by a sum over all elements such that we get

$$-aD\int_{\Omega_2}\left(\nabla^2 u_h\right)(\tau\boldsymbol{w}\cdot\nabla v_h)\;d\Omega_2 \equiv -\tau aD\sum_E\int_E\left(\nabla^2 u_h\right)(\boldsymbol{w}\cdot\nabla v_h)\;dE = 0,$$

as we are using piecewise linear basis functions. Using the same basis $\{\phi_1,\phi_2,\ldots,\phi_n\}$ as before, we can use the expansion

$$u_h = \sum_{j=1}^{n} u_j\phi_j$$

where $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)^T$ is a vector of the real coefficients. We obtain n equations by letting $v = \phi_i$ for $i = 1,\ldots,n$. This leads to the following problem: find u_j, where $j = 1,\ldots,n$ such

that

$$\sum_{j=1}^{n} u_j \left[D \int_{\Omega_2} \nabla\phi_j \cdot \nabla\phi_i \, d\Omega_2 + a \int_{\Omega_2} (\boldsymbol{w} \cdot \nabla\phi_j)\, \phi_i \, d\Omega_2 \right.$$
$$\left. + \tau a \int_{\Omega_2} (\boldsymbol{w} \cdot \nabla\phi_j)\, (\boldsymbol{w} \cdot \nabla\phi_i)\ d\Omega_2 \right]$$
$$= \int_{\Omega_2} s\phi_i \, d\Omega_2 + \tau a \int_{\Omega_2} s\, (\boldsymbol{w} \cdot \nabla\phi_i)\ d\Omega_2,$$

for $i = 1, \ldots, n$. This can be written in matrix form as the linear system of equations

$$[DK + aC + \tau aT]\boldsymbol{u} = \boldsymbol{s}. \tag{2.30}$$

with K and C as defined in (2.18), upwind transport matrix T with entries defined as

$$t_{ij} = \int_{\Omega_2} (\boldsymbol{w} \cdot \nabla\phi_j)\, (\boldsymbol{w} \cdot \nabla\phi_i)\ d\Omega_2$$

and

$$s_i = \int_{\Omega_2} s\phi_i \, d\Omega_2 + \tau a \int_{\Omega_2} s\, (\boldsymbol{w}.\nabla\phi_i)\ d\Omega_2. \tag{2.31}$$

Note that T is a symmetric and positive definite matrix [15].

The element stiffness and convection matrices are the same as in (2.23) and (2.24), respectively. To assemble the element matrix T, we compute

$$t_{ij} = \int_E (\boldsymbol{w} \cdot \nabla\psi_j)\, (\boldsymbol{w} \cdot \nabla\psi_i)\ dE = \int_E \left[w_x \frac{\partial\psi_j}{\partial x} + w_y \frac{\partial\psi_j}{\partial y} \right] \left[w_x \frac{\partial\psi_i}{\partial x} + w_y \frac{\partial\psi_i}{\partial y} \right] dE,$$

where $i, j = 1, 2, 3$, ψ_i are the element basis functions given by (2.19), and $\boldsymbol{w} = (w_x, w_y)$. This can be transformed to reference element R, as in Section 2.4.3. Using the chain rule (2.21) and the identity (2.22), we get

$$t_{ij} = \frac{1}{|J|} \int_0^1 \int_0^\xi \left[w_x \left(\frac{\partial\psi_j}{\partial\xi}\frac{\partial y}{\partial\eta} - \frac{\partial\psi_j}{\partial\eta}\frac{\partial y}{\partial\xi} \right) + w_y \left(\frac{\partial\psi_j}{\partial\eta}\frac{\partial x}{\partial\xi} - \frac{\partial\psi_j}{\partial\xi}\frac{\partial x}{\partial\eta} \right) \right]$$
$$\left[w_x \left(\frac{\partial\psi_i}{\partial\xi}\frac{\partial y}{\partial\eta} - \frac{\partial\psi_i}{\partial\eta}\frac{\partial y}{\partial\xi} \right) + w_y \left(\frac{\partial\psi_i}{\partial\eta}\frac{\partial x}{\partial\xi} - \frac{\partial\psi_i}{\partial\xi}\frac{\partial x}{\partial\eta} \right) \right] d\eta d\xi.$$

Similarly, the additional term on the RHS of (2.31) is assembled by the integral

$$\int_E s\, (\boldsymbol{w} \cdot \nabla\psi_i)\ dE = \int_E s \left[w_x \frac{\partial\psi_i}{\partial x} + w_y \frac{\partial\psi_i}{\partial y} \right] dE,$$
$$\Rightarrow \frac{1}{|J|} \int_0^1 \int_0^\xi s \left[w_x \left(\frac{\partial\psi_i}{\partial\xi}\frac{\partial y}{\partial\eta} - \frac{\partial\psi_i}{\partial\eta}\frac{\partial y}{\partial\xi} \right) + w_y \left(\frac{\partial\psi_i}{\partial\eta}\frac{\partial x}{\partial\xi} - \frac{\partial\psi_i}{\partial\xi}\frac{\partial x}{\partial\eta} \right) \right] d\eta d\xi.$$

2.5 Global Sensitivity Analysis

The pollutant dispersion problems that we study in this thesis involve a number of model parameters that impact the pollutant concentration. Sensitivity analysis (SA) is a statistical tool that can be used to characterise the uncertainty that a parameter contributes to the quantity of interest. It plays a key role in mathematical and computational modelling by providing valuable information about how the parameters in a model affect its output. Global sensitivity analysis (GSA) measures can be used to identify which parameters have the most significant influence on the quantity of interest [52, 54, 59].

In the remainder of this thesis, we are interested in investigating the relative influence of the model parameters on the variability of the pollutant concentration $u(x, y)$. This is done using variance-based GSA measures known as Sobol indices [1, 64]. Sobol indices quantify the proportion of the total variance in the quantity of interest that can be attributed to each parameter under consideration. Importantly, Sobol indices can also be used to analyse how interactions between parameters may affect the output variance.

2.5.1 Sobol Indices

In this section, we detail the construction of two different Sobol sensitivity measures, first-order Sobol indices and total effect Sobol indices. First-order Sobol indices provide a measure of the variance contributed by individual parameters, whereas the total effect measures the total variance contributed by a parameter, including the variance contributed by its interaction with other parameters.

In the 1D case, we consider the SUPG solution of the BVP at the point x as a black box function $f_x : \mathbb{R}^k \to \mathbb{R}$, and denote the map from the inputs from a vector of parameters $\boldsymbol{P} = (P_1, P_2, \ldots, P_k)$ to the pollutant concentration at the point x by $Y_x = f_x(\boldsymbol{P})$. The variance in the pollutant concentration at the point x is denoted by $\mathbb{V}(f_x(\boldsymbol{P}))$. We will use $\boldsymbol{P}_{\tilde{j}}$ to denote all the parameters except the jth parameter. By the law of total variance [8],

$$\mathbb{V}(Y_x(\boldsymbol{P})) = \mathbb{E}_{\boldsymbol{P}_{\tilde{j}}}(\mathbb{V}_{P_j}(Y_x|\boldsymbol{P}_{\tilde{j}})) + \mathbb{V}_{P_j}(\mathbb{E}_{\boldsymbol{P}_{\tilde{j}}}(Y_x|P_j)),$$

where $\mathbb{E}_{\boldsymbol{P}_{\tilde{j}}}(\mathbb{V}_{P_j}(Y_x|\boldsymbol{P}_{\tilde{j}}))$ is the expected conditional variance of Y_x given $\boldsymbol{P}_{\tilde{j}}$, and $\mathbb{V}_{P_j}(\mathbb{E}_{\boldsymbol{P}_{\tilde{j}}}(Y_x|P_j))$ is the variance of the conditional expectation of Y_x given P_j, for $j = 1, 2, \ldots, k$.

First-order Sobol indices can be computed at point x for parameter j by taking the ratio

between the variance of an individual parameter and the total variance of the output. Thus, the first-order Sobol index at point x for parameter j is given by

$$S_{x,j} = \frac{\mathbb{V}_{P_j}(\mathbb{E}_{\boldsymbol{P}_{\tilde{j}}}(Y_x|P_j))}{\mathbb{V}(Y_x(\boldsymbol{P}))}, \quad j = 1, \ldots, k. \tag{2.32}$$

An interpretation of $\mathbb{V}_{P_j}(\mathbb{E}_{\boldsymbol{P}_{\tilde{j}}}(Y_x|P_j))$ is that it is the expected reduction in variance given $\boldsymbol{P}_{\tilde{j}}$ [58] .

The total effect Sobol index at point x for parameter j is given by

$$S^t_{x,j} = \frac{\mathbb{E}_{\boldsymbol{P}_{\tilde{j}}}(\mathbb{V}_{P_j}(Y_x|\boldsymbol{P}_{\tilde{j}}))}{\mathbb{V}(Y_x(\boldsymbol{P}))}, \quad j = 1, \ldots, k, \tag{2.33}$$

where $\mathbb{E}_{\boldsymbol{P}_{\tilde{j}}}(\mathbb{V}_{P_j}(Y_x|\boldsymbol{P}_{\tilde{j}}))$ denotes the remaining expected variance given $\boldsymbol{P}_{\tilde{j}}$.

Sobol indices can be defined for the 2D case by considering the SUPG solution at a point $\boldsymbol{x} = (x, y)$. At each spatial location $\boldsymbol{x} \in \Omega_2$, we define a black-box function $f_{\boldsymbol{x}} : \mathbb{R}^k \to \mathbb{R}$ mapping the input parameters $\boldsymbol{P} = (P_1, \ldots, P_k)$ to the pollutant concentration at that point, so that $Y_{\boldsymbol{x}} = f_{\boldsymbol{x}}(\boldsymbol{P})$. The Sobol indices are then evaluated pointwise over the domain by applying the same variance decomposition as in (2.32) and (2.33) at each spatial location $\boldsymbol{x}$, yielding spatial maps of first-order and total-effect Sobol indices, $S_{\boldsymbol{x},j}$ and $S^t_{\boldsymbol{x},j}$, respectively. This pointwise formulation allows for a straightforward extension to the 2D domain, enabling sensitivity indices to be computed independently at each location.

2.5.2 Quasi-Monte Carlo Method

In order to study the relative impact of the parameters (2.3) on the pollutant dispersion problem, we estimate the Sobol indices by numerical approximation. Sobol indices are typically estimated using Monte Carlo methods. The Monte Carlo method is a computational technique that uses repeated random sampling in order to estimate numerical results. In this study, the Monte Carlo method is employed to approximate pollutant concentration. Specifically, wind speed, diffusivity, and source rate parameters are sampled from their respective probability distributions, and the resulting simulations of the BVP are computed to estimate the pollutant concentration levels. The Monte Carlo samples of the input parameters are taken to be independent of each other and its rate of convergence is $\mathcal{O}(m^{-1/2})$, where m is the sample size. The rate of convergence describes how quickly the error in the estimator decreases as the number of samples increases Here we choose to implement a Quasi-Monte Carlo approach as this method converges at an approximate rate of $\mathcal{O}(m^{-1})$ [35].

A Quasi-Monte Carlo method is a low-discrepancy sampling method that generates samples of $\boldsymbol{P}$ such that the points sampled are distributed as uniformly as possible over the unit hypercube $[0,1]^k$, where k is the dimension of $\boldsymbol{P}$ [58]. Low-discrepancy sampling ensures minimal gaps or clusters within the distribution of points in the sampling space. Unlike the classic Monte Carlo method, we require that successive sampling points are not independent in order to achieve uniformity. We have to progressively fill the space using ordered digital sequences. These sequences are constructed recursively using a base-b system. A base-b system is a positional number system in which any non-negative integer i is represented using powers of an integer base $b \geq 2$ as $i = \sum_{l \geq 0} a_l b^l$ with digits $a_l \in 0, \ldots, b-1$. In the construction of low-discrepancy sequences, this is used to systematically generate each new point in an order that fills the largest gaps left by previous points, promoting uniformity across all dimensions. Quasi-Monte Carlo sampling can be achieved by low-discrepancy sequences, such as Sobol sequences, Halton sequences, or Faure sequences [39]. These are a deterministic set of points designed to achieve a more uniform and structured coverage of a given space compared to purely random samples.

In order to use Quasi-Monte Carlo method to sample the parameters given in (2.3), we map the unit hypercube $[0,1]^k$ to different ranges using the inverse cumulative distribution functions (CDFs), also known as the quantile function, with respect to each probability distribution function. For a given random variable X with CDF $F_X(x)$, the inverse CDF $F_X^{-1}(y)$ maps a value $y \in [0,1]$ to a sample in the domain of X, such that $P(X \leq F_X^{-1}(y)) = y$. This transformation allows uniform samples in $[0,1]^k$ to be transformed into samples from any desired marginal distribution [20].

In this thesis, we will generate Quasi-Monte Carlo samples from the probability distributions (2.3) using Sobol sequences. We choose to use Sobol sequences due to their equidistribution property of sampling, and because they provide good results in high dimensions [5]. In order to get the right probability distribution for each parameter, we apply the inverse cumulative distribution function to each set of Quasi-Monte Carlo points.

2.6 Confidence Intervals

Confidence intervals (CIs) provide a range of plausible values for an unknown parameter based on observed data $(X_1, X_2, \ldots, X_m)$. To draw meaningful analysis and conclusions about the point estimates of the Sobol indices (2.32) and (2.33), we need a measure of accuracy and

reliability. In this thesis, we will compute CIs of these Sobol estimates in Chapters 3, 4 and 5.

For a parameter θ, the probability that the true value lies within the $(1-\alpha)$ CI is defined by

$$P(L \leq \theta \leq U) = 1 - \alpha,$$

where L is the lower confidence limit, U the upper confidence limit, and α is the significance level [55, §5.4]. For example, a 95% CI corresponds to $\alpha = 0.05$, meaning that the true value of a population parameter or statistical estimator should be captured within the interval 95% of the time.

In the classical approach, the limits are derived from the sampling distribution of the estimator $\hat{\theta}$. If this distribution is known or can be well-approximated , for example by a normal distribution, the CI is obtained by

$$\hat{\theta} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{m}}$$

This method is effective when the distribution is known and the standard error can be calculated directly. In this work, we use the bootstrap method instead, as it allows us to estimate confidence intervals directly from the data without requiring explicit formulas for the standard error. Bootstrapping is a computationally inexpensive way to generate a distribution of Sobol indices estimates.

Bootstrapping is a resampling process with replacement, in which data from a single dataset can be used to generate new estimates of Sobol indices and hence obtain a distribution of Sobol indices. Bootstrapping refers to the process of resampling in which we draw B samples of the original dataset with replacement [55]. That is, suppose that we have a dataset of m observations given by

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_m),$$

where each X_i represents the pollutant concentration at a fixed spatial location $(x, y) \in \Omega$, obtained from numerical solutions of the governing equations, (2.1) in 1D and (2.13) in 2D, using the SUPG method, with uncertain input parameters defined in (2.3). For $b = 1, \ldots, B$, where B is the total number of bootstrap replicates, we generate a resampled dataset $\boldsymbol{X}_b^* = \left(X_{b,1}^*, X_{b,2}^*, \ldots, X_{b,m}^*\right)$ from $\mathbf{X}$. For each resample $\boldsymbol{X}_b^*$, compute the estimator of interest $\hat{\theta}_b^*$. In this study, $\hat{\theta}_b^*$ is the bootstrap estimate of a Sobol sensitivity index.

The bias-corrected (BC) percentile approach is used to account for bias in the bootstrap distribution by finding the α and $(1-\alpha)$ percentiles of the distribution of the bootstrap estimators

$\hat{\theta}_b^*$. The bias-correction factor is computed as [14, §13]

$$z_0 = \Phi^{-1}\left(\frac{\sum_{i=1}^{B}(\hat{\theta}_b^* \leq \theta)}{B}\right),$$

where the inverse normal Φ^{-1} computes the proportion of bootstrap estimates which are less than or equal to the observed Sobol index $\hat{\theta}_b^* \leq \theta$, where θ is the observed Sobol index. The lower and upper percentiles of the CI are

$$lb = 2z_0 + z_\alpha, \quad ub = 2z_0 + z_{(1-\alpha)},$$

where $z_\alpha = \Phi^{-1}(\alpha)$ is the α percentile point of the standard normal distribution. We compute the quantiles of the bootstrap statistic θ_b^* such that they correspond to the lower and upper percentiles lb and ub, respectively. The CI for the Sobol index θ is then given by

$$\theta \in [q(\theta_b^*, lb),\ q(\theta_b^*, ub)],$$

where $q(\theta_b^*, lb)$ is the lower quantile and $q(\theta_b^*, ub)$ is the upper quantile.

2.7 Artificial Neural Networks

Artificial Neural Networks (ANNs) are computational models inspired by the human brain's biological neural networks. They are composed of interconnected processing elements called nodes, working together to solve specific problems. ANNs can approximate any continuous function on a compact domain with arbitrary accuracy, provided they have a sufficient number of neurons and layers. This property, known as the Universal Approximation Theorem, underlies their ability to represent complex, nonlinear mappings in data and models. The structure of an ANN typically consists of multiple layers of nodes, including input and output layers as well as some "hidden layers" between them [10].

In this thesis, we focus on using feedforward neural networks (NNs), in which information is fed forward between nodes from one layer to the next. The nodes in a NN use activation functions, such as sigmoid, hyperbolic tangent, or rectified linear unit functions, to transform their inputs into outputs. The sigmoid function scales inputs between 0 and 1 using an S-shaped curve, given by $f(x) = 1/(1+e^{-x})$. The hyperbolic tangent function is similar, but scales inputs between -1 and 1 using an S-shaped curve, where $f(x) = \tanh(x)$. The Rectified Linear Unit

(ReLU) activation function is defined as $f(x) = \max(0, x)$. These nonlinear activation functions allow the NN to learn complex, nonlinear relationships between the input features and the target variable being predicted [10].

Training a NN is necessary to enable the network to learn underlying patterns and generalise accurately to unseen data. This is achieved by iteratively adjusting weights and biases based on input data and corresponding outputs to minimise a loss function. In this work, we define the loss function as the mean-squared error (MSE),

$$\mathcal{L} = \frac{1}{m_{train}} \sum_{i=1}^{m_{train}} (\hat{y}_i - y_i)^2,$$

which calculates the average squared differences between the predicted outcomes denoted by $\hat{y}_i$, and the target value denoted by y_i. The number of samples used is denoted by m_{train}. During training, the network performs a forward pass to compute predictions, then the loss function evaluates the discrepancy. Back-propagation is used to compute the gradients of the loss function with respect to the weights and biases, which are updated using a gradient descent method [22].

Chapter 3

Sensitivity Analysis of a One-Dimensional Pollutant Dispersion Model

In this chapter, we study a parametric differential equation as a UQ problem by assigning probability distribution functions to model uncertainties associated with the parameters. We are interested in assessing how the variance in pollutant concentration in a simple pollutant dispersal model can be attributed to the variance in model input parameters such as wind speed, diffusivity, and source strengths. A spatial analysis of Sobol indices is performed to better understand how the relative influence of different parameters varies throughout the spatial domain. We focus on the simple 1D pollutant dispersion model introduced in Section 2.1, which retains the key features of a more complex underlying physical problem and therefore is still useful for gaining insight into sensitivity analysis of parameters. Part of the material in this chapter has been published in [41].

3.1 Implementation and Validation

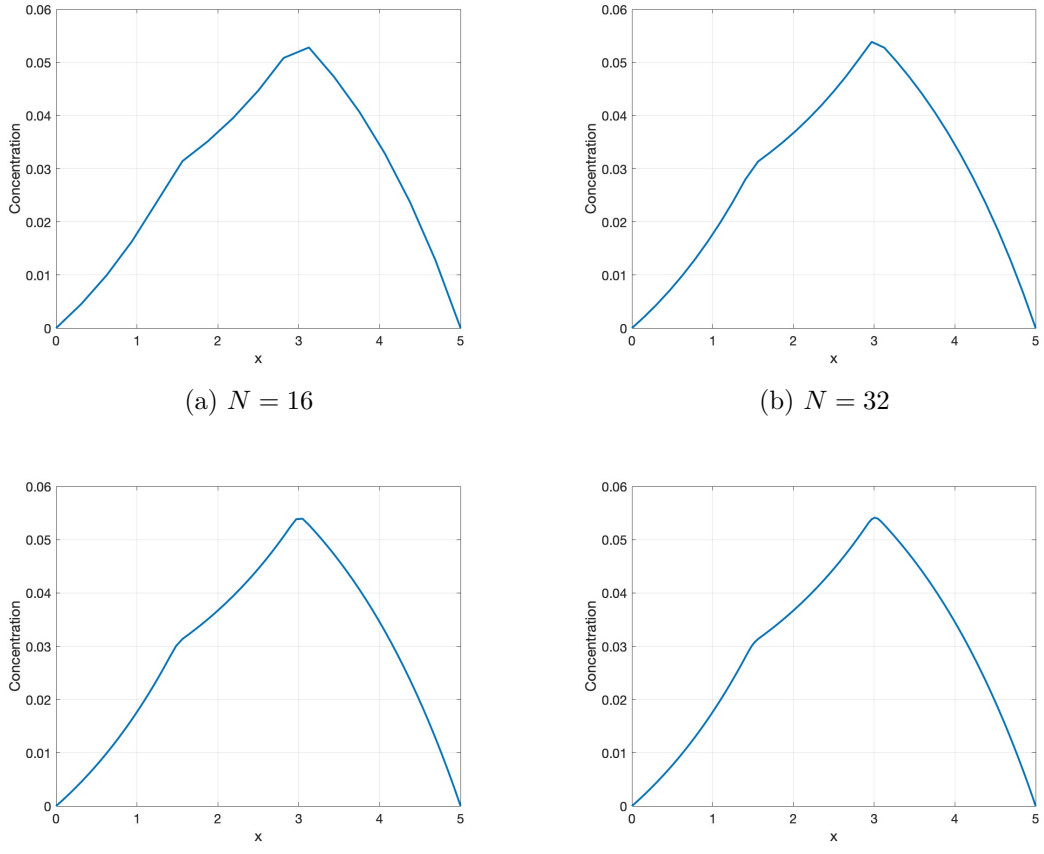
Recall the 1D BVP (2.1), which we solve using the SUPG method, detailed in Section 2.2. In what follows, we focus on an example in which the sources are located at $x_{s_1} = 1.5$ and $x_{s_2} = 3$ in the domain $\Omega_1 = [0, 5]$. The units for spatial dimensions are arbitrary here, as we are using a non-dimensional formulation.

The numerical implementation of the SUPG formulation was carried out in MATLAB using a structured, 1D mesh comprising of N uniformly spaced elements, where the mesh spacing is given by $h = 5/N$. Sparse matrix storage was employed to assemble the stiffness and convective matrices, K and C, efficiently, which significantly reduced memory overhead and improved computational performance. Forcing functions were evaluated using MATLAB's numerical integration function `integral` to ensure accurate representation of source terms within each element. MATLAB does this using adaptive quadrature and we set the relative error tolerance to `1e-12`. The linear system (2.11) was solved using MATLAB's backslash operator, which provides a robust and efficient direct solution method for sparse systems. Dirichlet boundary conditions were applied consistently across all simulations by explicitly enforcing zero values at the boundary nodes of the 1D mesh after solving the SUPG linear system. The implementation allows straightforward variation of the mesh resolution, enabling systematic mesh refinement studies.

To ensure numerical stability and reliability, the Péclet number and associated stabilisation parameter τ, which was formulated in (2.12), were monitored to confirm appropriate scaling with wind speed a, diffusivity D, and mesh resolution h. For small mesh sizes, the assembled matrices were examined for expected sparsity patterns and consistency between K and C. The resulting numerical solutions were visually inspected to ensure physical smoothness and to confirm that any potential oscillations in convection-dominated regimes were effectively suppressed by the streamline upwind stabilisation. These internal checks established confidence in the correctness of the implementation and provided the foundation for a quantitative mesh refinement study to support the selection of an appropriate mesh resolution.

To assess the effect of spatial resolution on the numerical results, a mesh refinement study was performed with $N = 16$, $N = 32$, $N = 64$ and $N = 128$ uniformly spaced elements, while the model parameters were held fixed. For the purposes of this study, we take these to be $a = 4.4062$, $D = 1.6487$, $s_1 = 1$, and $s_2 = 2$. Figure 3.1 shows that across all mesh sizes, the overall pollutant concentration levels and the position of the concentration peaks remained relatively consistent, indicating that the global solution characteristics were not sensitive to the mesh resolution. However, the spatial representation of these peaks was visibly sharper on coarser meshes, which is primarily a consequence of interpolation between a smaller number of spatial points rather than improved resolution of the underlying solution, as seen in Figures 3.1a and 3.1b. The concentration fields became progressively smoother, and while $N = 64$ elements does provide better resolution, Figure 3.1c still does not fully resolve the peak smoothly. This

is resolved by taking $N = 128$ elements, as seen in Figure 3.1d. We achieve a concentration field that is continuous and physically realistic without spurious oscillations. Table 3.1 presents the elapsed computation time for each mesh, recorded using MATLAB's built-in stopwatch function, `tic` and `toc`, to illustrate the trade-off between resolution and computational cost. We observe that while the time required to solve the linear system increases with finer meshes, the absolute solve times remain low. In particular, the use of $N = 128$ elements provides a clear improvement in spatial resolution that justifies the additional computational cost. The consistency of pollutant concentration values across all meshes, together with the smoothness achieved with $N = 128$ elements, supports that the numerical solution has reached practical mesh independence. Accordingly, $N = 128$ was adopted for all subsequent uncertainty analyses as a suitable compromise between accuracy and efficiency.



(a) $N = 16$ (b) $N = 32$

(c) $N = 64$ (d) $N = 128$

Figure 3.1: Mesh refinement study of SUPG finite element discretisation in $\Omega = [0, 5]$ with fixed input parameters.

N	Elapsed Time (seconds)
16	0.047528
32	0.068080
64	0.073038
128	0.173974

Table 3.1: Elapsed time taken to run SUPG solver for varying mesh sizes, run on Intel Core i5 Macbook Pro.

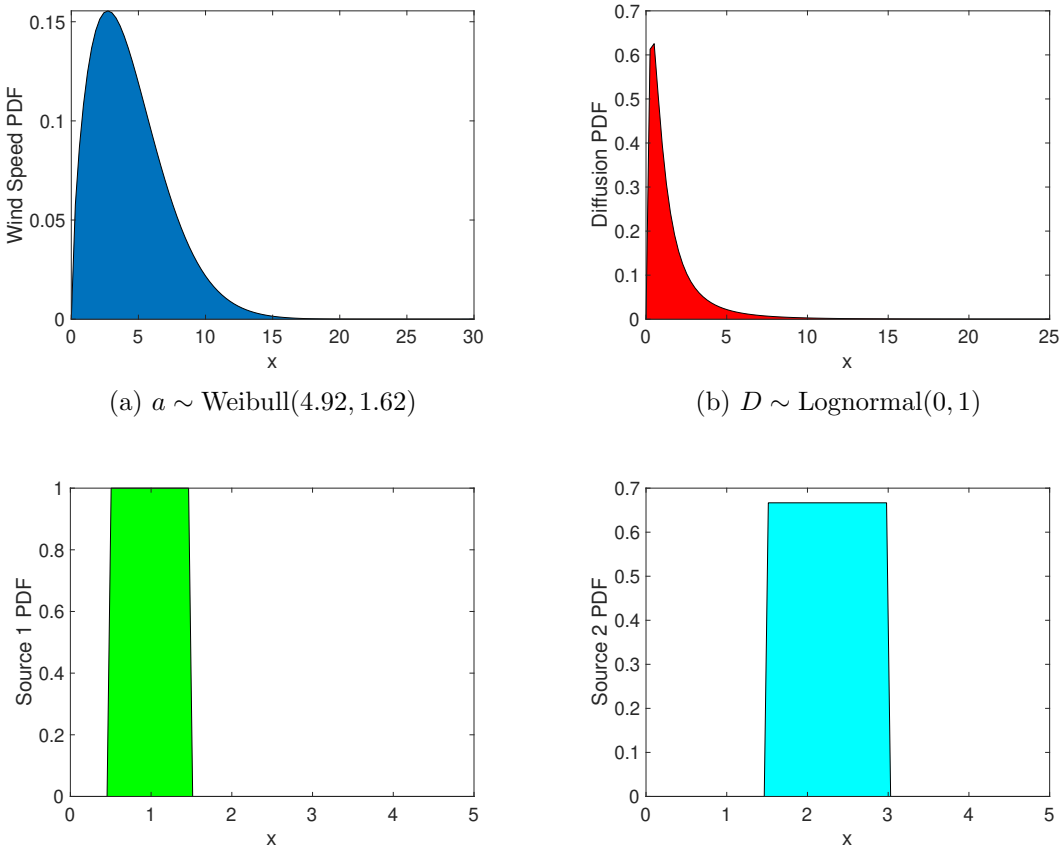
3.2 Uncertainty Quantification of Pollutant Concentration

We are able to evaluate the pollutant dispersion model using the SUPG method (2.11) as a UQ problem by performing a parametric study in which we sample the parameters a, D, s_1, and s_2 from the probability distribution functions (2.3). The parameters are sampled using a Quasi-Monte Carlo approach, as detailed in Section 2.5.2, with sample size $m = 10^4$, which was selected to ensure sufficient accuracy and convergence of the Sobol index estimates while maintaining computational feasibility for the parametric study using the SUPG method. The choice of this sample size is supported by an additional set of Sobol index estimates computed for the pollutant dispersion model using an increased sample size of $m = 10^5$, as presented in Section 3.3.2, to show visually indistinguishable results, hence motivating $m = 10^4$ as sufficient.

The parameter distributions and their respective ranges were selected to capture a representative range of mesh Péclet numbers, allowing exploration of different regimes of advection and diffusion dominated pollutant dispersion problems. This choice enables assessment of how variations in these physical parameters influence the Sobol sensitivity indices, discussed in Section 3.3.2. A summary of the parameter distributions, expected values, and variances is given in Table 3.2. Figure 3.2 shows the plots of these probability distribution functions. We fix the source length scale in (2.1) to $\sigma = 0.1$.

Parameter	Distribution	Expected Value	Variance
a	Weibull(4.92,1.62)	4.4062	7.7715
D	Lognormal(0,1)	1.6487	4.6708
s_1	Uniform(0.5,1.5)	1	0.0833
s_2	Uniform(1,3)	2	0.3333

Table 3.2: Summary of parameter distributions, expected values and variances.



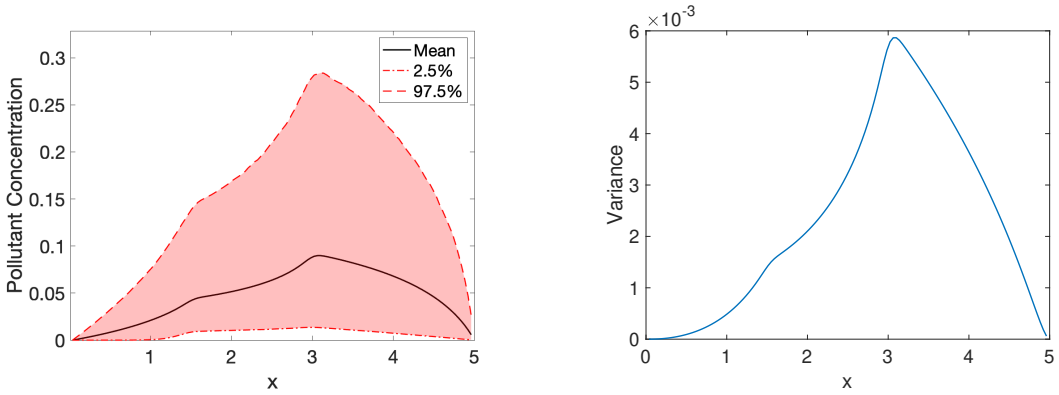
(a) $a \sim \text{Weibull}(4.92, 1.62)$ (b) $D \sim \text{Lognormal}(0, 1)$

(c) $s_1 \sim \text{Uniform}(0.5, 1.5)$ (d) $s_2 \sim \text{Uniform}(1, 3)$

Figure 3.2: Sampling distribution of pollutant dispersion parameters.

Figure 3.3 illustrates the variability in the pollutant concentration u over the whole spatial domain using these PDFs for the input parameters. Figure 3.3a shows the empirical 2.5 and 97.5 percentiles of the solution profiles, with the empirical mean shown in black. The empirical variance of pollutant concentration is shown in Figure 3.3b. We observe that pollutant concentrations tend to be higher as we move downstream of the sources due to transport effects, causing an uneven distribution where some combinations of parameters can lead to a high pollutant concentration compared to the mean value.

To study how the distribution of the model parameters affects the pollutant concentration, we now consider the parameter values that give the highest and lowest pollutant concentrations. Figure 3.4 shows boxplots of the parameters that result in pollutant concentrations below the 2.5 percentile and above the 97.5 percentile at points $x = 1$, $x = 2$ and $x = 4$. We can see that the distribution of wind speed is significant to the level of pollutant concentration. Figure 3.4a shows



(a) Variability in pollutant concentration

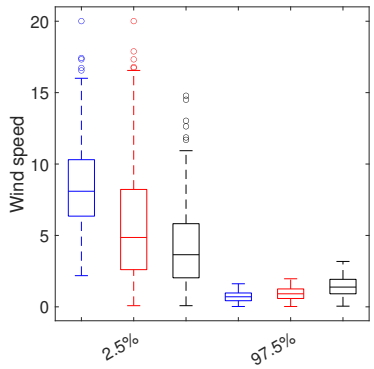
(b) Variance of Pollutant Concentration

Figure 3.3: Pollutant concentration using SUPG finite element discretisation with $N = 128$ elements and sampling with $m = 10^4$ Quasi-Monte Carlo samples.

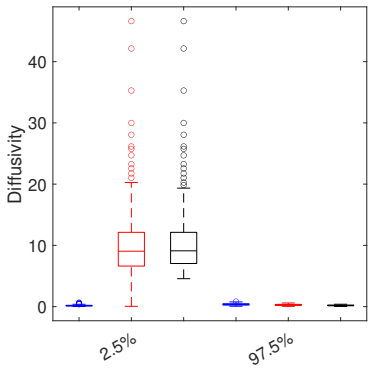
that large wind speeds correspond to low pollutant concentrations throughout the domain, and conversely, low wind speeds correspond to high pollutant concentrations. Figure 3.4b highlights the importance of the distribution of diffusivity, where we observe that at $x = 2$ and $x = 4$ a larger diffusivity results in lower pollutant concentration. Small diffusivity corresponds to high pollutant concentrations throughout the domain. We note that at $x = 1$, small diffusivity corresponds to pollutants both above the 97.5 percentile and below the 2.5 percentile. This indicates that diffusivity is not a prominent parameter at $x = 1$. Figures 3.4c and 3.4d show that greater source strengths also result in high levels of pollutant concentration. These boxplots indicate that the variability in the model parameters affects the level of pollutant concentration, and that the influence of certain parameters changes throughout the domain. These observations align well with our physical intuition about pollutant dispersion.

3.3 Spatially-Dependent Global Sensitivity Analysis

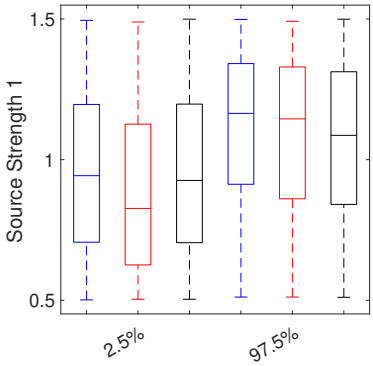
Having observed the variability in the concentration across the domain, we would like to be able to quantify the contribution of the individual input parameters to this uncertainty. We do this by using Sobol indices, as described in Section 2.5.1, as a measure of model output sensitivity to uncertainty in the input parameters.



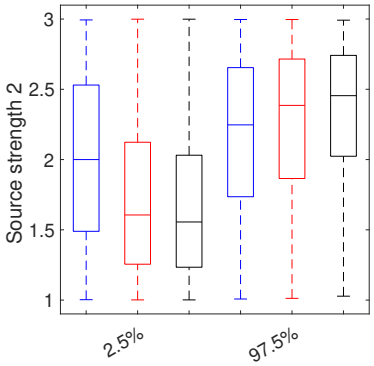
(a) Wind speed a



(b) Diffusivity D



(c) Source strength s_1



(d) Source strength s_2

Figure 3.4: Boxplots of wind speed, diffusivity, and source strength values that result in pollutant concentrations below the 2.5 percentile and above the 97.5 percentile: $x = 1$ (blue), $x = 2$ (red) and $x = 4$ (black).

3.3.1 Estimation of Sobol Indices

In order to use Sobol indices to analyse the sensitivities associated with the pollutant dispersion model (2.1), we require a method to estimate quantities (2.32) and (2.33). We follow the proposed best practice from the literature, using [30] to estimate the first-order indices and [58] to estimate the total effect indices as follows.

Let Z_1 and Z_2 denote two $m \times k$ design matrices, where m is the sample size and k is the number of parameters. For each spatial point x, we generate

$$\boldsymbol{y}_x^{(l)} = f_x\left(Z_l\right), \quad l = 1, 2,$$

a vector of pollutant concentrations obtained by solving the BVP on a fixed uniform grid, where $\overline{y_x^{(l)}} = \left[\sum_{i=1}^m (\boldsymbol{y}_x^{(l)})_i\right]/m$ is the sample mean concentration at the point x.

We then generate a resampled matrix $Z_{\sim j}$, obtained from Z_2 with the entries of the jth column fixed to the values in Z_1, and use the BVP to calculate $\boldsymbol{y}_{x,\sim j} = f_x(Z_{\sim j})$. This has mean concentration $\overline{y_{x,\sim j}} = \left[\sum_{i=1}^m (\boldsymbol{y}_{x,\sim j})_i\right]/m$. The first-order indices for each parameter are calculated according to [30]

$$\hat{S}_{x,j} = \frac{\sum_{i=1}^m \left[\left(\boldsymbol{y}_x^{(1)}\right)_i - \overline{y_x^{(1)}}\right]\left[\left(\boldsymbol{y}_{x,\sim j}\right)_i - \overline{y_{x,\sim j}}\right]}{\sum_{i=1}^m \left[\left(\boldsymbol{y}_x^{(1)}\right)_i - \overline{y_x^{(1)}}\right]^2}, \quad j = 1, \ldots, k. \tag{3.1}$$

For the estimation of the total effects, we generate another resampled matrix $Z^t_{\sim j}$, which is obtained from Z_1 with entries of the jth column taken from Z_2, then construct $\boldsymbol{y}^t_{x,\sim j} = f_x(Z^t_{\sim j})$ which contains the solution to the BVP corresponding to the resampled inputs $Z^t_{\sim j}$. The total effect indices for each parameter are calculated according to [58]

$$\hat{S}^t_{x,j} = \frac{1}{2}\frac{\sum_{i=1}^m \left[\left(\boldsymbol{y}_x^{(1)}\right)_i - \left(\boldsymbol{y}^t_{x,\sim j}\right)_i\right]^2}{\sum_{i=1}^m \left[\left(\boldsymbol{y}_x^{(1)}\right)_i - \overline{y_x^{(1)}}\right]^2}, \quad j = 1, \ldots, k. \tag{3.2}$$

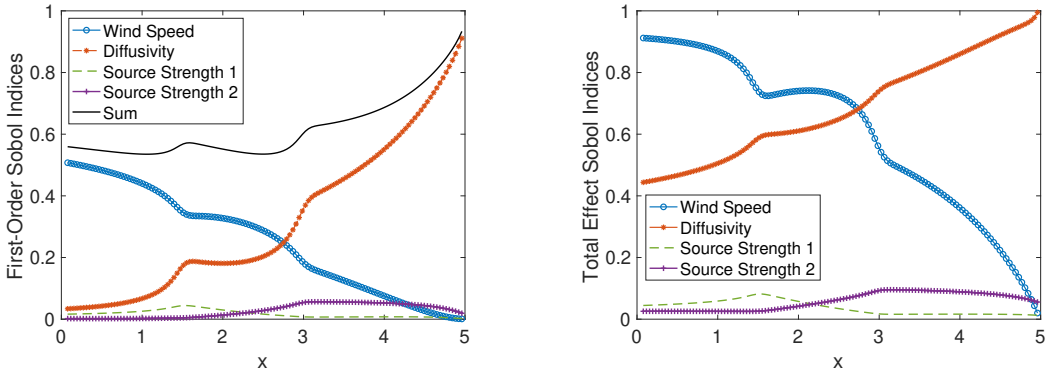
To calculate the computational cost of estimating Sobol indices (3.1) and (3.2), we consider the number of function evaluations required to estimate the indices, where each function evaluation is the SUPG solution of an $N \times N$ linear system. The estimates require the m function evaluation to populate $\boldsymbol{y}_x^{(1)}$. We require a further km function evaluations for $\boldsymbol{y}_{x,\sim j}$ and km function evaluations for $\boldsymbol{y}^t_{x,\sim j}$. The total computational cost of estimating Sobol indices at all spatial points is therefore $(1 + 2k)m$.

%labeltab:ishigami

3.3.2 Spatial Dependence of Sobol Indices

We are interested in using Sobol indices to conduct a spatial analysis of the sensitivity of the model to study the processes that most affect the uncertainty in pollutant concentration throughout the domain. Figure 3.5 shows the spatial dependence of the first-order and total effect Sobol indices across the domain, with the constant wind blowing from left ($x = 0$) to right ($x = 5$).

Recall that the first-order Sobol indices in Figure 3.5a indicate the proportion of variance



(a) First-order indices

(b) Total effect indices

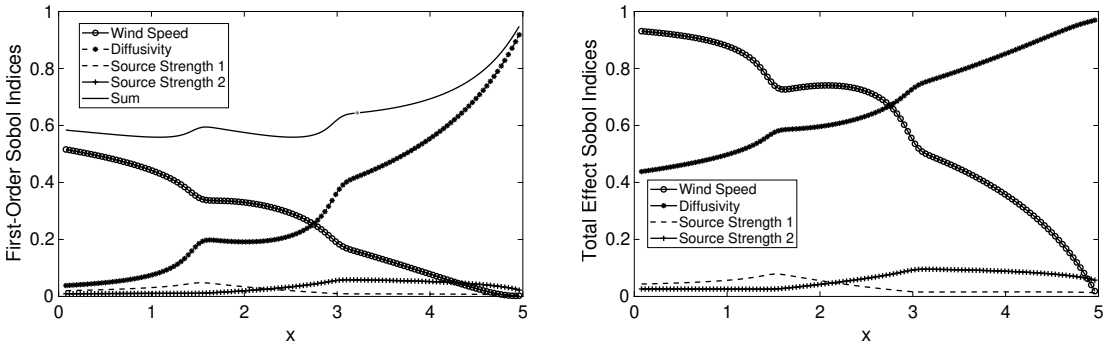
Figure 3.5: Spatial distribution of Sobol indices.

that each individual parameter contributes to the overall variance in the pollutant concentration. We see that upstream of the first source, at $x_{s_1} = 1.5$, the wind speed is the most influential parameter. As we move further downstream, we note that the relative influence of the wind speed begins to decrease, while the influence of diffusivity increases. After the second source, at $x_{s_2} = 3$, the diffusivity becomes the dominant factor that contributes uncertainty to the pollutant concentration. We also observe that the variance contributed by the source strength s_1 is localised about the source location x_{s_1}, and there is a similar effect for s_2 about x_{s_2}, which is not surprising.

The black line in Figure 3.5a shows that the sum of all first-order indices is less than one. This indicates that the first-order indices alone do not capture all the variance contribution towards the pollutant concentration. The total effect indices shown in Figure 3.5b highlight the interactions between the parameters. We note that while Figure 3.5a suggests that diffusivity has little variance contribution upstream of x_{s_1}, the total effects show that diffusivity plays an important role in the overall contribution of variance to pollutant concentration upstream and downstream of x_{s_1} when interacting with other factors. We also note that the total effect indices for the wind speed are larger than the first-order indices, indicating that the variance in wind speed further contributes to the pollutant dispersion problem when interacting with other parameters. A small increase in the total effect indices for s_1 and s_2 is observed compared to their first-order indices, again reflecting the influence of interaction with wind speed and diffusivity.

In order to support the choice of the sample size $m = 10^4$, an additional set of Sobol index estimates was computed for the pollutant dispersion model using an increased sample size of

$m = 10^5$, which are presented in Figure 3.6. By comparing Figure 3.6 with Figure 3.5, we see that sensitivity patterns and relative parameter contributions are visually indistinguishable, indicating that increasing the sample size by an order of magnitude does not lead to any appreciable change in the estimated Sobol indices. This comparison suggests that the Sobol index estimates are stable with respect to the Quasi-Monte Carlo sampling for the pollutant dispersion problem considered here. On this basis, $m = 10^4$ is used throughout the remainder of the parametric study as a suitable balance between numerical reliability and computational cost.



(a) First-order indices

(b) Total effect indices

Figure 3.6: Spatial distribution of Sobol indices using $m = 10^5$ Quasi-Monte Carlo samples.

The results presented in Figure 3.5 indicate that variability in the wind speed and the diffusivity are key factors in the overall variability in pollutant concentration throughout the domain. Hence, it is desirable to know more about how changes in the choice of probability distribution representing these parameters affect the variability in pollutant concentration. First, we repeat the experiment above with the mean value for wind speed fixed as $\mathbb{E}(a) = 4.4062$, but we reduce the variance by a factor of two by adjusting the shape and scale parameters of the Weibull distribution accordingly. Specifically, we use $a_{new} \sim \text{Weibull}(4.97, 2.38)$ with $\mathbb{E}(a_{new}) = 4.4062$ as before and $\mathbb{V}(a_{new}) = 3.8857$. Figure 3.7 compares the probability density functions of the two distributions a and a_{new}.

Figure 3.8 shows the variability in the pollutant concentration using the new wind speed distribution, with all other parameter distributions unchanged. Compared with Figure 3.3b, we see that the variance of the pollutant concentration has reduced by a third, as seen in Figure 3.8b. In addition, Figure 3.8a shows that the 97.5 percentile value is slightly lower, particularly upstream of the first source location x_{s_1}. As seen in Figure 3.7, the probability density function

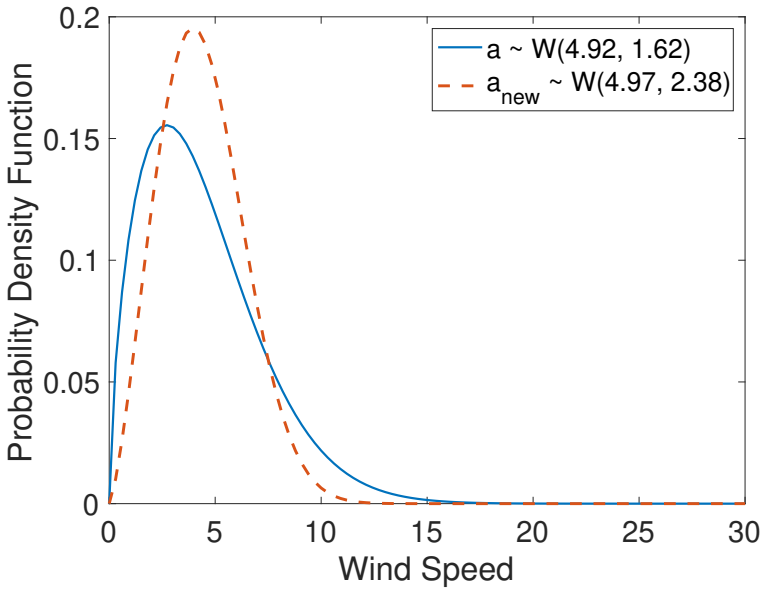
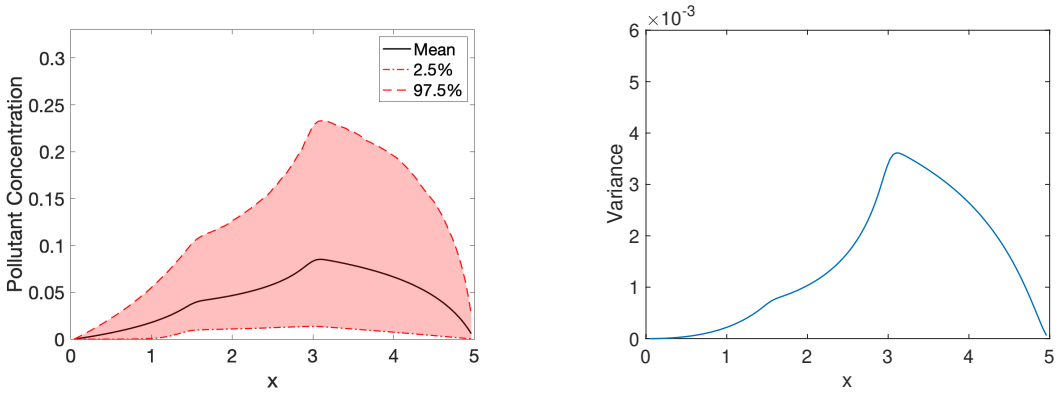


Figure 3.7: Probability density functions of Weibull distributions for wind speed.



(a) Variability in pollutant concentration

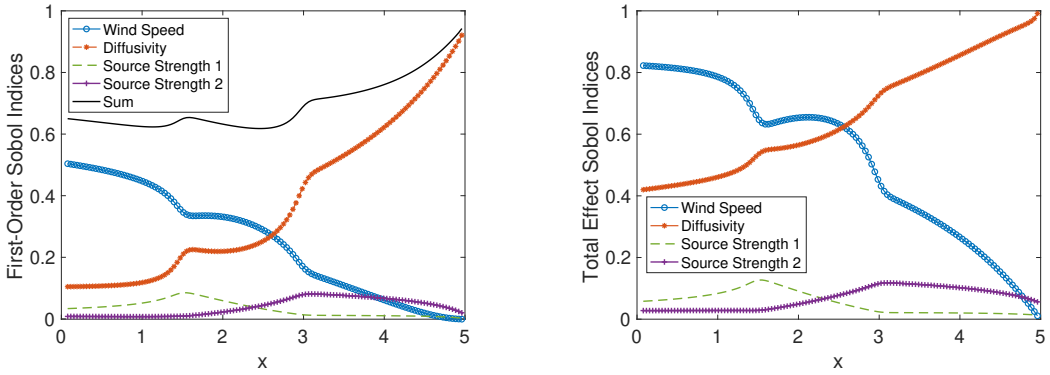
(b) Variance of Pollutant Concentration

Figure 3.8: Variability in pollutant concentration using $a_{new} \sim \text{Weibull}(4.97, 2.38)$.

of the new wind speed has fewer low wind speeds, and this change is responsible for the global reduction in the 97.5 percentile pollution level.

Having seen that reducing the variation in the wind speed has resulted in an overall reduction in the variance of the concentration, we now consider whether the relative contribution of the parameters to the new variance has also changed. Figure 3.9 shows the spatial dependence of the new first-order and total effect Sobol indices across the domain.

Compared with Figure 3.5, we see that diffusivity and source strengths now play a more



(a) First-order indices

(b) Total effect indices

Figure 3.9: Spatial distribution of Sobol indices using $a_{new} \sim \text{Weibull}(4.97, 2.38)$.

prominent role in the overall variance of the concentration upstream of $x_{s_1} = 1.5$. However, we note that, despite reducing the variance of the wind speed, it is clearly still the most important factor in this region of the domain. Furthermore, we observe that the sum of first-order indices in Figure 3.5a is greater than in Figure 3.9a. This also indicates that the variance contributed by individual parameters is greater after reducing the variance in the wind speed distribution.

We now repeat the same experiment, considering the effect of reducing the variance in the distribution of the diffusion parameter while all other parameters remain unchanged. Here, we sample the wind speed as $a \sim \text{Weibull}(4.92, 1.62)$ as before. Again, we reduce the variance by a factor of two by adjusting the lognormal mean and standard deviation parameters accordingly. Specifically, we sample $D_{new} \sim \text{Lognormal}(0.19, 0.79)$ with $\mathbb{E}(D_{new}) = 1.6487$ as before and $\mathbb{V}(D_{new}) = 2.3354$. Figure 3.10 compares the probability density functions of the two distributions D and D_{new}.

Figure 3.11 shows the variability in the pollutant concentration using the new distribution for diffusivity, with all other parameter distributions unchanged. As before, comparing with Figure 3.3b, we see that the variance of the pollutant concentration has now been reduced by a factor of two, as seen in Figure 3.11b. Figure 3.11a also shows that the 97.5 percentile value is slightly lower, particularly downstream of the first source location x_{s_1}. Similarly to the case where a change was made to the Weibull distribution, in Figure 3.10 we observe that the probability density function of the new diffusivity has fewer small diffusion rates, hence reducing the 97.5 percentile pollution concentration level.

Looking at the Sobol indices for this case in Figure 3.12a, we observe that the first-order indices for diffusivity increase at a steeper rate as compared with Figure 3.5a. Additionally,

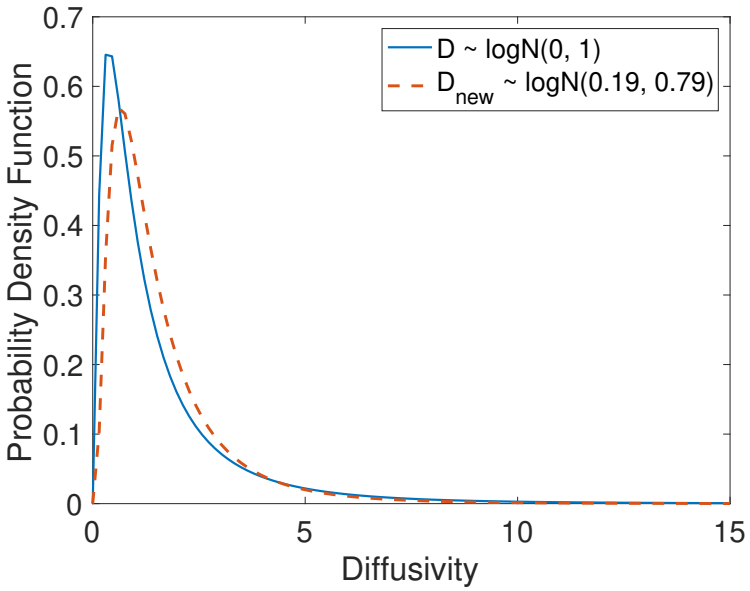
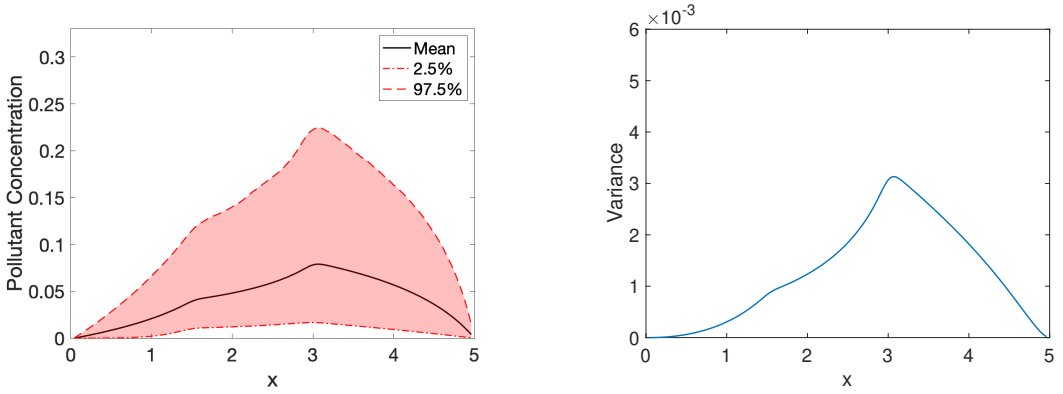


Figure 3.10: Probability density functions of Lognormal distributions for diffusivity.

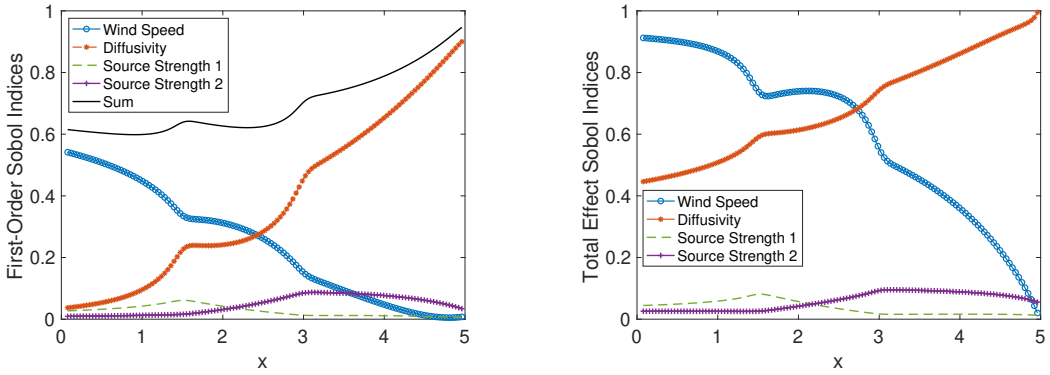


(a) Variability in pollutant concentration

(b) Variance of Pollutant Concentration

Figure 3.11: Variability in pollutant concentration using $D_{new} \sim \text{Lognormal}(0.19, 0.79)$.

the first-order indices for wind speed decrease at a slightly steeper rate too. This indicates that although we have decreased the variance in diffusion parameter distribution, diffusion actually becomes more prominent earlier in the domain. We note that the crossover from wind speed being the most influential parameter to diffusion also occurs slightly farther upstream of x_{s_2}. Although the rate at which diffusion gains prominence increases, we see in Figure 3.12b that the total effect indices of diffusivity are slightly lower than in Figure 3.5b, indicating that the



(a) First-order indices

(b) Total effect indices

Figure 3.12: Spatial distribution of Sobol indices using $D_{new} \sim \text{Lognormal}(0.19, 0.79)$.

overall variance contribution of the diffusivity has gone down. We also note that for both the first-order and total effect indices, we see that the influence of the source strengths has increased in Figure 3.12 as compared with Figure 3.5, both in terms of the variance contributions by s_1 and s_2, and also the area about x_{s_1} and x_{s_2} where the increased Sobol indices are observed. This is because reducing the variance in diffusivity lowers its overall contribution to uncertainty, thereby increasing the relative influence of the source strengths.

3.3.3 Confidence Intervals for Sobol Indices Point Estimates

In order to validate our results and make meaningful comments, we computed 95% CIs throughout the domain for both the first-order and total effect indices, for each parameter. The results presented in Section 3.3.2 are point estimates for the Sobol indices of each parameter throughout the domain. We now consider CIs to ensure that the results we have are in fact reliable.

We use the bootstrapping technique described in Chapter 2.6 by taking the function evaluations that are used to compute the estimates of the Sobol indices and resampling with replacement $B = 10^3$ times. This choice provides a reliable approximation of the sampling distribution of the Sobol indices and yields stable CI estimates, while keeping the computational cost manageable. Further increasing B leads to negligible changes in the resulting CIs, indicating that $B = 10^3$ is sufficient for the present analysis.

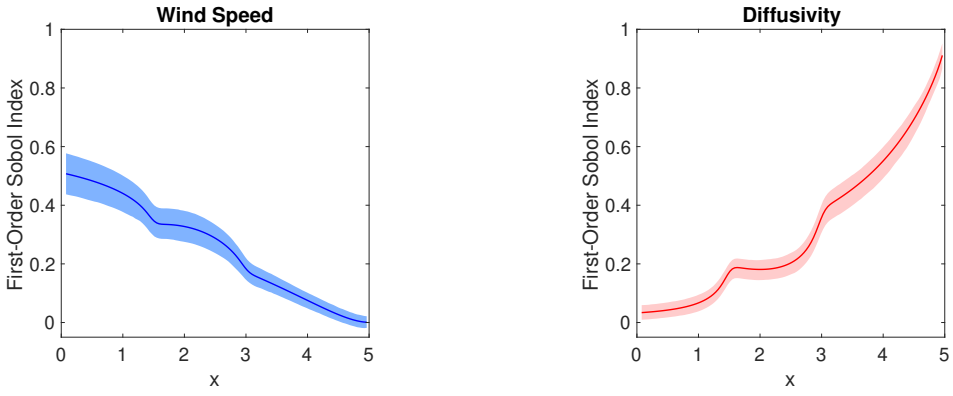
We compute a vector of pollutant concentrations $\boldsymbol{y}_x^{(l)} = f_x(Z_l)$ as in Section 3.3.1 by solving the pollutant dispersion problem (2.1) using the SUPG method presented in Section 2.2 with $N = 128$. A sampling distribution of pollutant concentrations is generated by resampling with replacement $B = 10^3$ times. Then, for each sampling distribution, we compute the sample

mean and variance, and replicate the sensitivity indices $S_i[b]$ and $S_{t_i}[b]$, following the estimation method in Section 3.3.1. The CIs are then computed using the BC percentile method presented in Section (2.6).

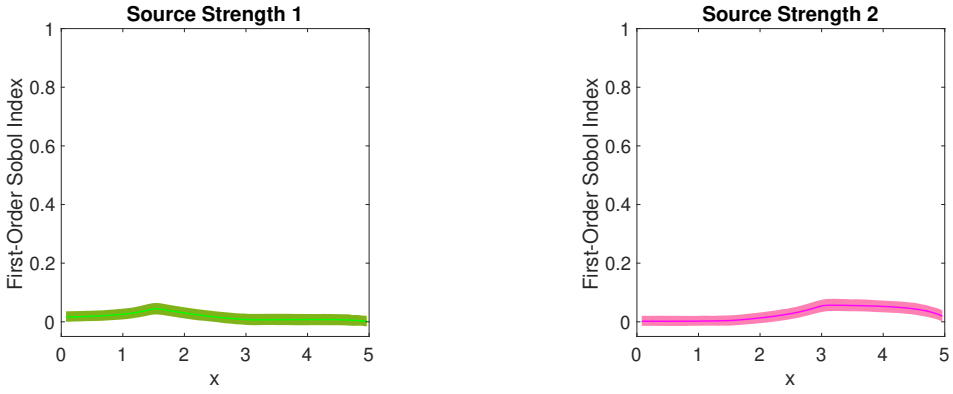
Figure 3.13 shows the 95% CIs for each first-order index throughout the domain using $m = 10^4$ samples. All plots are shown on the same scale for comparability reasons. The CIs calculated show that the first-order index for wind speed has the widest CI. However, the range of the wind speed CI still indicates that wind speed would be the most influential parameter upstream of x_{s_2}. The CI is widest upstream of the first source, located at $x_{s_1} = 1.5$, where in the first-order case, only the variance in the wind speed is contributing variance to the solution. In comparison, we see that all the other parameters have relatively narrow CIs, indicating that the point estimates of first-order Sobol indices for the remaining parameters are accurate. In particular, the CIs for the source strengths show that the point estimates of the Sobol indices are highly accurate, indicating that the variance in the source strengths alone have little impact on the variance of the model output. The CI for the first-order index for the diffusivity shows that there may be some more variability in the point estimates in the middle of the domain, downstream of the location of the first source at $x_{s_1} = 1.5$. However, near the boundaries, we see that the point estimates are very accurate.

These results indicate that the proportion of variance in the pollutant concentration attributed to the variance in the wind speed not only plays a significant role but is also the most difficult to accurately determine. We further analyse this by looking at the 95% confidence intervals for the total effect indices. As we know from our study of Sobol indices, the total effect indices show that interactions play a significant role in the uncertainty of the model output.

In Figure 3.14, we see that the 95% confidence intervals for the point estimates of the total effect indices of wind speed and diffusivity are wider than in the first-order case. We see that both these confidence intervals show the most variability upstream of the first source and, while they do get slightly narrower after the first source is introduced, they still indicate considerable variability in the point estimates of the total effect indices. This suggests that for the total effect indices, we need more samples to be able to accurately estimate the "true" values of the total effect indices of the wind speed and diffusivity. The 95% confidence intervals of the point estimates of the total effect indices of the sources are the only total effect indices that we can claim to be accurate. The confidence intervals for the two source strengths further confirm the intuition that the variance in the source strengths, even with interaction to other parameters, does not significantly impact the variance in the solution. In particular, the variance in s_1 is



(a) $a \sim \text{Weibull}(4.96, 1.63)$ (b) $D \sim \text{Lognormal}(0, 1)$



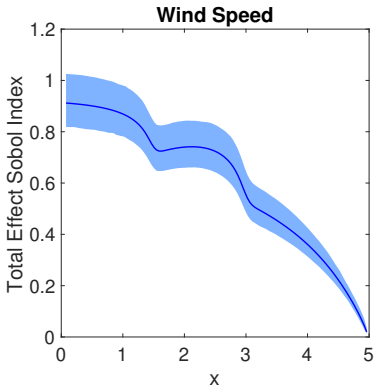
(c) $s_1 \sim \text{Uniform}(0.5, 1.5)$ (d) $s_2 \sim \text{Uniform}(1, 3)$

Figure 3.13: 95% confidence intervals of spatial distribution of first-order Sobol indices, computed using $m = 10^4$ Quasi-Monte Carlo samples.

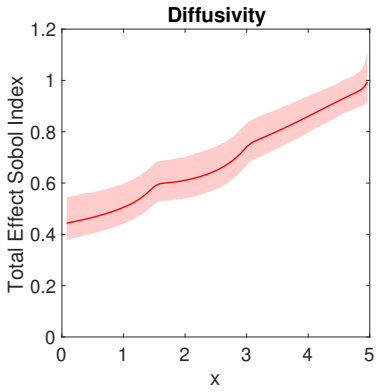
considered to have a relatively minor impact on the overall uncertainty in the model output. The confidence interval for s_2 appears to indicate some variability upstream of $x_{s_2} = 3$, suggesting that the variance in s_2, with interaction, could lead to some variance in the output. Therefore, although in general the variance in the source strength contributes very little to possibly no variance in the pollutant concentration, there is the possibility that some variance could be attributed to the stronger source.

3.4 Summary

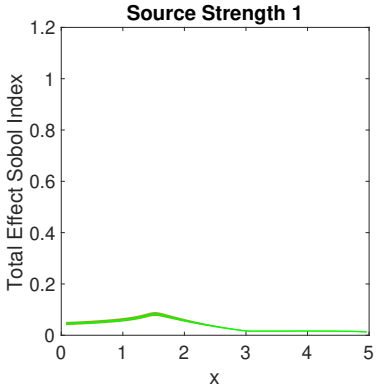
In this chapter, we investigated the variability in pollutant concentration within pollutant dispersal BVP by applying Sobol indices as a GSA tool. We have implemented numerical methods



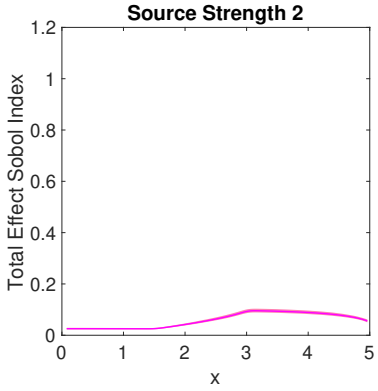
(a) $a \sim \text{Weibull}(4.96, 1.63)$



(b) $D \sim \text{Lognormal}(0, 1)$



(c) $s_1 \sim \text{Uniform}(0.5, 1.5)$



(d) $s_2 \sim \text{Uniform}(1, 3)$

Figure 3.14: 95% confidence intervals of spatial distribution of total effect Sobol indices, computed using $m = 10^4$ Quasi-Monte Carlo samples.

to estimate both the first-order and total-effect Sobol indices for model input parameters defined by probability distributions.

Our results demonstrate that the variance in pollutant concentration is spatially heterogeneous, with the influence of individual parameters varying significantly across the domain. Upstream of the first source location x_{s_1}, wind speed emerges as the dominant contributor of variance to model output, while further downstream, after the second source location x_{s_2}, diffusivity assumes greater relative importance. Around each source location, the variance contribution from source strengths is strongly localised. Moreover, total-effect indices highlight the significance of parameter interactions, particularly between wind speed and diffusivity, in driving variance even in regions where their first-order effects are minimal.

Importantly, our analysis also demonstrates that reducing the uncertainty in wind speed and diffusivity distributions can substantially decrease the overall variance in pollutant concentration. These findings show the value of spatially-resolved sensitivity measures for improving the reliability and predictive capability of pollutant dispersal models.

CIs were computed to quantify the uncertainty in the Sobol index estimates. The 95% CIs show that the wind speed parameter has the widest uncertainty, especially upstream of x_{s_1} indicating greater difficulty in accurately estimating its sensitivity in this region. The source strength parameters have relatively narrow CIs, reflecting more reliable and precise sensitivity estimates.

Chapter 4

Sensitivity Analysis of Two-Dimensional Pollutant Dispersion Model

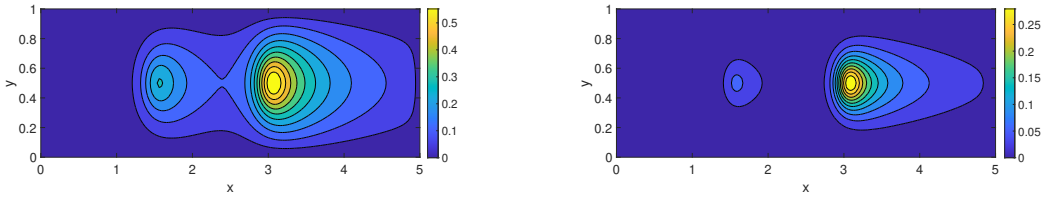
While Chapter 3 established the foundations for studying a simple pollutant dispersion problem as a UQ problem and is useful to gain insight into the sensitivity of the parameters, the application of a one-dimensional model is limited in its description of real-world problems. In this chapter, we extend the UQ problem established in Chapter 3 to a two-dimensional domain in order to better understand how the pollutant concentration varies throughout a more realistic spatial domain. Furthermore, because this leads to a significant increase in computational cost, we compare two different approaches to estimating Sobol indices and the computational efficiency of each approach.

We now study the 2D pollutant dispersion BVP (2.13), which we solve using the SUPG method detailed in Section 2.4. Similarly to Chapter 3, we focus on an example in which we assume the wind speed $a \sim \text{Weibull}(4.92, 1.62)$, the diffusivity $D \sim \text{Lognormal}(0, 1)$, and the source strengths $s_1 \sim \text{Uniform}(0.5, 1.5)$ and $s2 \sim \text{Uniform}(1, 3)$. Figure 3.2 shows the above probability distribution functions, and further details of the expected value and variance for each of these distributions can be found in Table 3.2. We focus on an example with source locations fixed at $\boldsymbol{x}_{s_1} = (1.5, 0.5)$ and $\boldsymbol{x}_{s_2} = (3, 0.5)$ in the domain $\Omega_2 = \{(x, y) : 0 \leq x \leq 5, 0 \leq y \leq 1\}$. We fix the wind direction $\boldsymbol{w} = (1, 0)$ and the source length scale $\sigma = 0.1$.

We interpret the pollutant dispersion model (2.13) as a UQ problem by conducting a para-

metric study of the model, in which we sample the parameters as described in Table 3.2 using a Quasi-Monte Carlo approach, detailed in Section 2.5.2, with sample size $m = 10^4$. We solve this using the SUPG method discussed in Section 2.4 with grid size $h = 1/64$, on a triangulation that is symmetric with respect to both coordinate axes.

Figure 4.1 illustrates the empirical mean and variance observed in the pollutant concentration. We observe that the distribution of the pollutant concentration is characterised by a peak about each source location and spread downstream of $\boldsymbol{x_{s_2}} = (3, 0.5)$, reflecting advection-dominated transport. Note that the mean and variance are symmetric about $y = 0.5$. Figure 4.1b indicates that most of the variance is concentrated downstream of $\boldsymbol{x_{s_2}}$.



(a) Mean pollutant concentration

(b) Variance of pollutant concentration

Figure 4.1: Pollutant concentration using SUPG finite element discretisation on a 320×64 mesh with $m = 10^4$ Quasi-Monte Carlo samples.

We use Sobol indices to quantify the uncertainty in the pollutant concentration by analysing the proportion of variance contributed by each parameter. Recall that Sobol indices are defined in Section 2.5.1. The first-order indices indicate the proportion of variance contributed by individual parameters, and the total effect indices account for the proportion of variance contributed by interactions between different parameters.

4.1 Estimation of Sobol Indices

In Section 3.3.1, we presented an approach to estimating Sobol indices, following the literature [30] to estimate the first-order indices (3.1) and a method presented in [58] to estimate the total effect indices (3.2). In this section, we describe an alternative approach to estimating the first-order Sobol indices for 2D space.

As before, we let Z_1 and Z_2 denote $m \times k$ design matrices, where m is the sample size and k is the number of parameters. For each spatial point $\boldsymbol{x} = (x, y)$, the vector of pollutant

concentrations obtained by solving the BVP on a given grid for the sampled parameters is

$$\boldsymbol{y}_{\boldsymbol{x}}^{(l)} = f_{\boldsymbol{x}}(Z_l), \quad l = 1, 2,$$

where $\overline{\boldsymbol{y}_{\boldsymbol{x}}^{(l)}} = \left[\sum_{i=1}^{m} (\boldsymbol{y}_{\boldsymbol{x}}^{l})_i\right] / m$ is the sample mean concentration at the point $\boldsymbol{x}$.

In Chapter 3, we generated two resampled matrices, $Z_{\sim j}$ to estimate the first-order indices and $Z_{\sim j}^{t}$ to estimate the total effect indices. However, we see that instead of generating two resampled matrices, we can use one resampled matrix and make use of the design matrix Z_2. Previously, Z_2 was only used in the generation of the resampled matrices.

Here, we take the same overall approach as in Section 3.3.1, but take Z_2 as the base matrix instead of Z_1. Specifically, we only resample the matrix $Z_{\sim j}^{t}$, which is obtained from Z_1 with entries in the jth column taken from Z_2.

This modification yields identical approximations of the Sobol index results while achieving a computational saving: the same resampled matrix can now be used for both first-order and total effect indices, avoiding redundant function evaluations and reducing the total number of model runs by km function evaluations. As before, the BVP is solved such that $\boldsymbol{y}_{\boldsymbol{x},\sim j}^{t} = f_{\boldsymbol{x}}(Z_{\sim j}^{t})$, with mean concentration $\overline{\boldsymbol{y}_{\boldsymbol{x},\sim j}^{t}} = \left[\sum_{i=1}^{m} (\boldsymbol{y}_{\boldsymbol{x},\sim j}^{t})_i\right] / m$. We now estimate the first-order indices for each parameter as

$$\hat{S}_{\boldsymbol{x},j} = \frac{\sum_{i=1}^{m} \left[\left(\boldsymbol{y}_{\boldsymbol{x}}^{(2)}\right)_i - \overline{\boldsymbol{y}_{\boldsymbol{x}}^{(2)}}\right] \left[\left(\boldsymbol{y}_{\boldsymbol{x},\sim j}^{t}\right)_i - \overline{\boldsymbol{y}_{\boldsymbol{x},\sim j}^{t}}\right]}{\sum_{i=1}^{m} \left[\left(\boldsymbol{y}_{\boldsymbol{x}}^{(2)}\right)_i - \overline{\boldsymbol{y}_{\boldsymbol{x}}^{(2)}}\right]^2}, \quad j = 1, \ldots, k. \tag{4.1}$$

The method of estimation for the total effect indices is the same as before. The total effect indices are estimated as in (3.2)

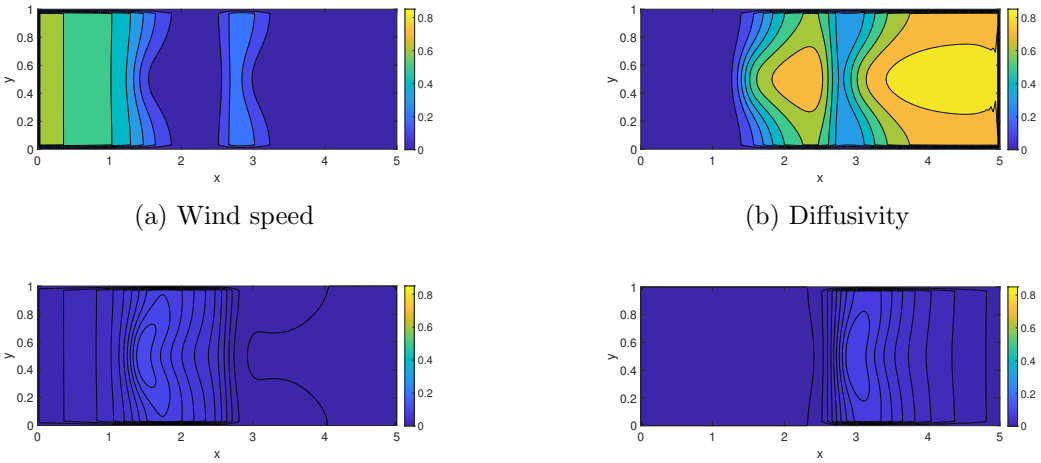
Let each function evaluation correspond to one SUPG solve of the linear system arising from the discretisation over the spatial domain $\Omega_2 = (0, 5) \times (0, 1)$. The computation of $\boldsymbol{y}_{\boldsymbol{x}}^{(1)}$ and $\boldsymbol{y}_{\boldsymbol{x}}^{(2)}$ each requires m function evaluations, while $\boldsymbol{y}_{\boldsymbol{x},\sim j}^{t}$ requires km evaluations. Therefore, the total computational cost of estimating both the first-order and total-effect Sobol indices at each spatial location using the above method is $(2 + k)m$ function evaluations.

In comparison, the method described in Section 3.3.1 incurs a higher cost of $(1 + 2k)m$ function evaluations. The revised approach in this chapter therefore achieves a reduction of $(k - 1)m$ function evaluations, representing a substantial computational saving, particularly in light of the high cost per evaluation in the domain Ω_2. This efficiency gain is largely due to requiring function evaluations of only a single reassembled matrix $Z_{\sim j}^{t}$, as opposed to two in

the previous method. Estimating the spatial distribution of the Sobol indices is significantly more expensive in the domain Ω_2 due to the increased number of spatial points to be evaluated.

4.2 Study of Spatially-Dependent Sobol Indices

Using the method described in Section 4.1, we calculate the first-order and total effect indices at each point on the finite element grid to study the variance contribution of each parameter. Figure 4.2 shows the spatial dependence of the first-order Sobol indices throughout the domain, with parameter distributions described in Table 3.2 and fixed wind direction $\boldsymbol{w} = [1, 0]$. Recall that the first-order indices presented in Figure 4.2 show the proportional variance contributed by each individual parameter to the overall variance shown in Figure 4.1b. We see that the



(a) Wind speed

(b) Diffusivity

(c) Source strength 1

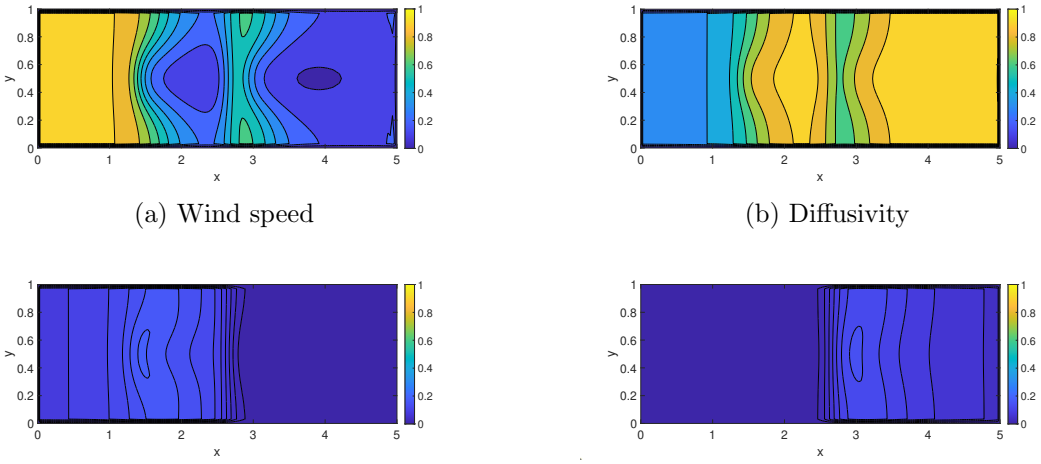
(d) Source strength 2

Figure 4.2: Spatial distribution of first-order Sobol indices.

most influential parameters are the wind speed and diffusivity. In comparison, the proportion of variance contributed by the source strengths throughout the domain is very low. We present the Sobol indices for the source strengths on a zoomed in scale in order to study their impact more closely in Figure 4.4. Upstream of the first source, at $\boldsymbol{x}_{s_1} = (1.5, 0.5)$, we observe that wind speed is more significant and as we move further downstream, we observe that the relative influence of diffusivity begins to overtake that of wind speed. We see that downstream of $\boldsymbol{x}_{s_1}$ the diffusivity becomes the dominant factor, but we note that about $\boldsymbol{x}_{s_2} = (3, 0.5)$ wind speed briefly gains prominence. Downstream of $\boldsymbol{x}_{s_2}$ diffusivity becomes the most influential parameter again. We observe in Figures 4.2c and 4.2d that the variance contributed by the source strength

s_1 is localised about the location of the source $\boldsymbol{x_{s_1}}$ and there is a similar effect for s_2 about $\boldsymbol{x}_{s_2}$.

We note that all first-order indices are less than one. This indicates that the first-order indices alone do not capture all the variance contribution towards the pollutant concentration. There is therefore interest in considering the total effect indices. This is due to the direction of the wind $\boldsymbol{w} = [1, 0]$. Figure 4.3 shows the spatial dependence of the total effect Sobol indices throughout the domain. The total effect indices of wind speed and diffusivity clearly show that



(a) Wind speed

(b) Diffusivity

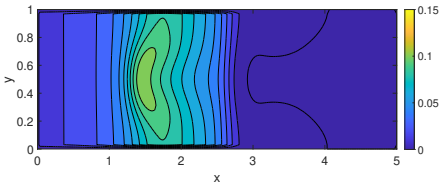
(c) Source strength 1

(d) Source strength 2

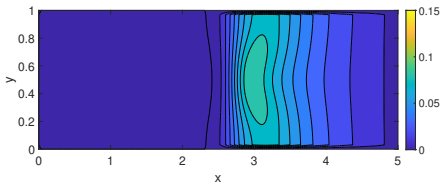
Figure 4.3: Spatial distribution of total effect Sobol indices.

interaction plays a significant role in the variance contributed by these parameters. Figure 4.2b suggests that diffusivity has very little variance contribution upstream of $\boldsymbol{x}_{s_1}$, but the total effect indices in Figure 4.3b show that diffusivity plays an important role in the overall contribution of variance to pollutant concentration upstream and downstream of x_{s_1} when interacting with other factors. A similar pattern can be observed for wind speed downstream of $\boldsymbol{x}_{s_2}$. Figure 4.2a shows very little variance contribution, however, Figure 4.3a shows that the wind speed does have an impact downstream of $\boldsymbol{x}_{s_2}$ when we consider interactions with other parameters. The total effect indices for s_1 and s_2, also shown in Figure 4.4, highlight that the variance contributed by these parameters is also heightened by interaction. In particular, we note that Figure 4.4c shows that the variance contributed by s_1 extends upstream of $\boldsymbol{x}_{s_1}$, and Figure 4.4d shows that the variance contributed by s_2 extends downstream of $\boldsymbol{x}_{s_2}$. We also note the symmetry of all of the Sobol indices about $y = 0.5$.

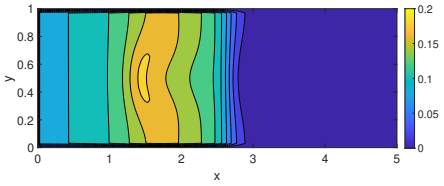
The findings shown in Figures 4.2 and 4.3 highlight that variability in the wind speed and diffusivity constitute a significant portion of the overall variance in pollutant concentration



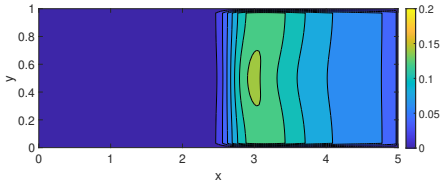
(a) First-order indices of s_1



(b) First-order indices of s_2



(c) Total effect indices of s_1

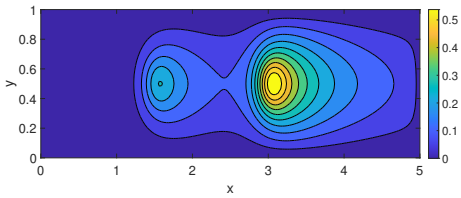


(d) Total effect indices of s_2

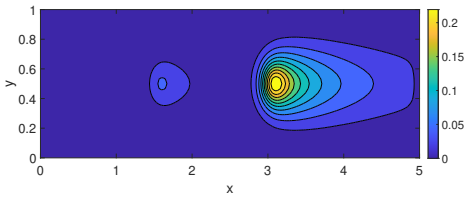
Figure 4.4: Spatial distribution of Sobol indices of source strengths.

across the spatial domain. We therefore now conduct further investigation into how changes in wind speed and diffusivity affect variability in pollutant concentration. This is done by repeating the pollutant dispersion simulations with the mean wind speed and diffusivity fixed as before but reducing the variance by half, similarly to the experiments carried out in 1D in Section 3.3.2. We look at this on a case-by-case basis. First, we reduce the variance of the wind speed by adjusting the shape and scale parameters of the Weibull distribution accordingly. Specifically, we use $a_{new} \sim \text{Weibull}(4.97, 2.38)$ with $\mathbb{E}(a_{new}) = 4.4062$ as before and $\mathbb{V}(a_{new}) = 3.8857$. A comparison of the probability density functions of the two distributions a and a_{new} can be found in Figure 3.7.

Figure 4.5 shows the pollutant concentration using the new wind speed distribution, with all other parameter distributions unchanged. By comparing Figure 4.5b directly with Figure



(a) Mean pollutant concentration

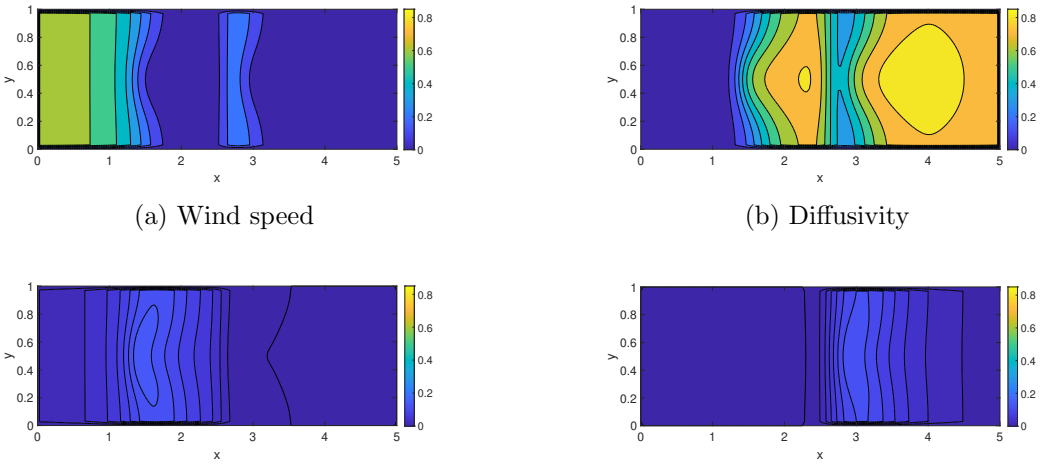


(b) Variance of pollutant concentration

Figure 4.5: Pollutant concentration using $a_{new} \sim \text{Weibull}(4.97, 2.38)$.

4.1b, we see that the variance of the pollutant concentration has reduced by 20%. We also note the reduced pollutant concentration shown in Figure 4.5a, with greater concentration dispersion

about the source locations. We now consider how the relative impact of each parameter has changed given the new wind speed distribution and the overall variance reduction. Figure 4.6



(a) Wind speed

(b) Diffusivity

(c) Source strength 1

(d) Source strength 2

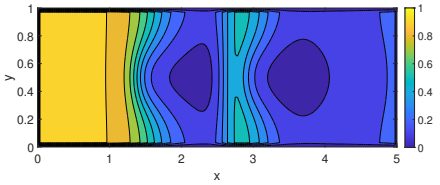
Figure 4.6: Spatial distribution of first-order Sobol indices using $a_{new} \sim \text{Weibull}(4.97, 2.38)$.

shows the first-order indices for the pollutant dispersion problem using the new wind speed a_{new}. We observe by comparison with Figure 4.2 that the relative impact of diffusivity has increased upstream of $\boldsymbol{x}_{s_1}$. This is due to the slight reduction in the impact of the wind speed upstream of $\boldsymbol{x}_{s_1}$, although it still remains the most influential parameter in this region.

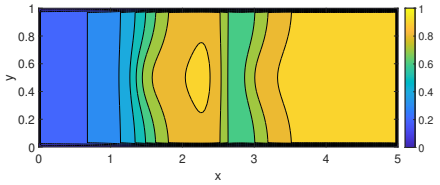
Figure 4.7 shows the total effect indices using the new wind speed a_{new}. By comparing directly with Figure 4.3, we note the total effect of wind speed downstream of $\boldsymbol{x}_{s_1}$ has decreased. We see that in this region, the total effect of diffusivity has increase slightly, with peaks spread further along the y-axis. Figure 4.8 shows that for both first-order and total effect indices, the relative impact of the source strengths has increased as a result of the reduction in wind speed variance.

Now we study the case where we reduce the variance of the diffusivity, keeping the wind speed and sources fixed as before. Specifically, we draw samples from $D_{new} \sim \text{Lognormal}(0.19, 0.79)$. This distribution maintains the previously established expectation $\mathbb{E}(D_{new}) = 1.6487$, but with the reduced variance $\mathbb{V}(D_{new}) = 2.3354$. Figure 3.10 provides a comparison of the probability density functions of the two distributions D and D_{new}.

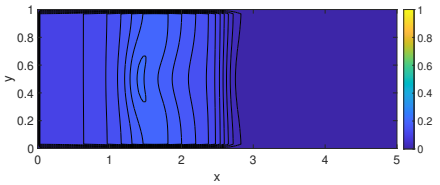
Figure 4.9 shows the pollutant concentration using the new diffusivity distribution. We observe a variance reduction of 40% in Figure 4.9b as compared with Figure 4.1b. We also note the reduced pollutant concentration shown in Figure 4.5a, with greater concentration dispersion



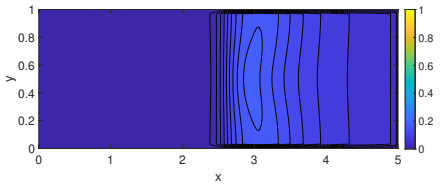
(a) Wind speed



(b) Diffusivity

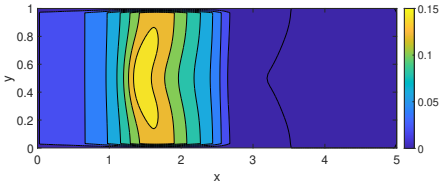


(c) Source strength 1

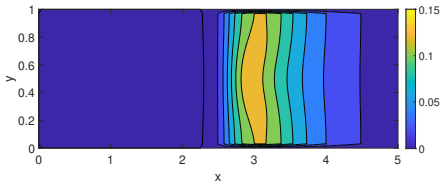


(d) Source strength 2

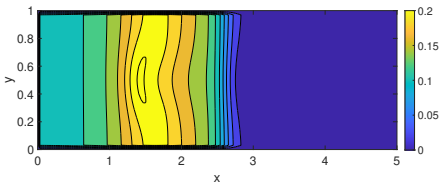
Figure 4.7: Spatial distribution of total effect Sobol indices using $a_{new} \sim \text{Weibull}(4.97, 2.38)$.



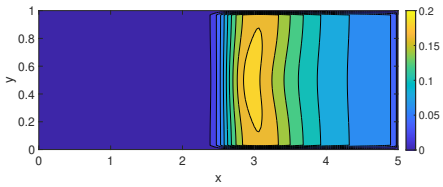
(a) First-order indices of s_1



(b) First-order indices of s_2



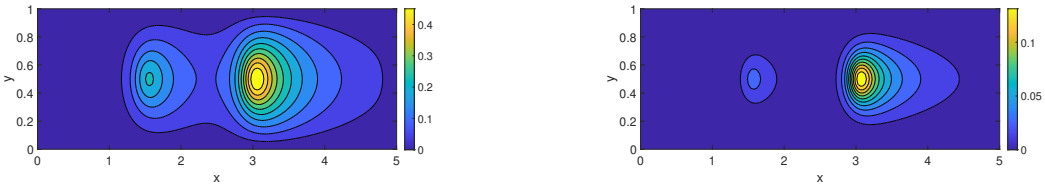
(c) Total effect indices of s_1



(d) Total effect indices of s_2

Figure 4.8: Spatial distribution of Sobol indices of source strength using $a_{new} \sim \text{Weibull}(4.97, 2.38)$.

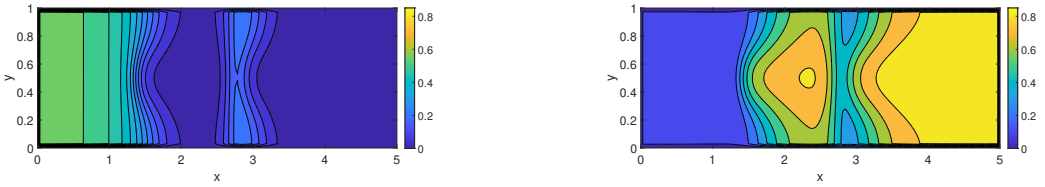
about the source locations. We next consider how the relative impact of each parameter has changed given the new diffusivity distribution and the overall variance reduction.



(a) Mean pollutant concentration

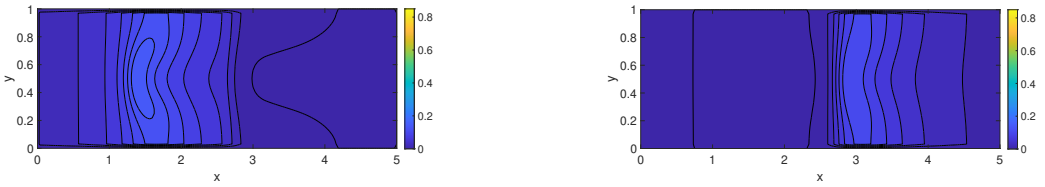
(b) Variance of Pollutant Concentration

Figure 4.9: Pollutant concentration using $D_{new} \sim \text{Lognormal}(0.19, 0.79)$



(a) Wind speed

(b) Diffusivity



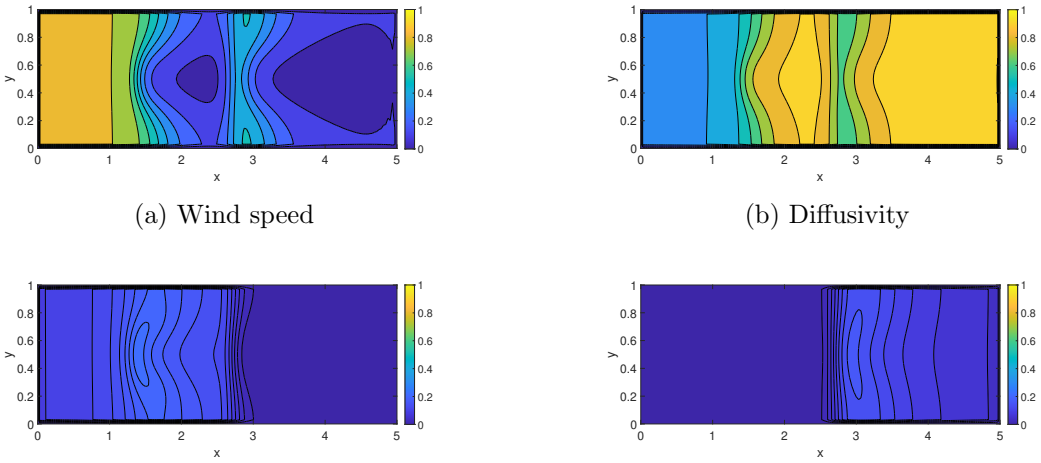
(c) Source strength 1

(d) Source strength 2

Figure 4.10: Spatial distribution of first-order Sobol indices using $D_{new} \sim \text{Lognormal}(0.19, 0.79)$.

Figure 4.10 shows the first-order indices using the new wind speed D_{new}. We observe by comparison with Figure 4.2 that the relative impact of diffusivity is very similar to Figure 4.2b. However, we note the slight difference in the spread of the first-order indices of diffusivity. In particular, the region between $\boldsymbol{x}_{s_1}$ and $\boldsymbol{x}_{s_2}$ has a steeper peak, indicating that the impact of diffusivity in this region has decreased slightly. We also note that upstream of $\boldsymbol{x}_{s_1}$, the wind speed has gained prominence.

Figure 4.11 shows the total effect indices for the pollutant dispersion problem using the new wind speed D_{new}. By comparing directly with Figure 4.3, we see that the impact of wind speed has actually increased throughout the domain as a result of interaction with other parameters. The total effect of diffusivity is also similar to Figure 4.3b, but we note the change in the



(a) Wind speed

(b) Diffusivity

(c) Source strength 1

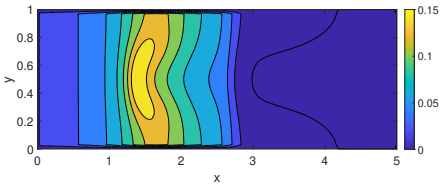
(d) Source strength 2

Figure 4.11: Spatial distribution of total effect Sobol indices using $D_{new} \sim \text{Lognormal}(0.19, 0.79)$.

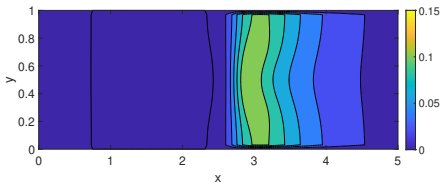
dispersion again. We also note that the total effect of diffusivity upstream of $\boldsymbol{x}_{s_1}$ has decreased. Figure 4.12 shows slight changes for both first-order and total effect indices. In particular, Figure 4.12a highlights increased impact of s_1 upstream of $\boldsymbol{x}_{s_1}$. The first-order and total effect indices of the sources are similar to Figure 4.4, but we note that the Sobol indices are now less dispersed than before. Across all figures of Sobol indices, we note that there is significantly less variations along the y-axis compared to the horizontal changes. Generally we observe slightly higher values toward the centre of the domain and lower values toward the top and bottom boundaries. All spatial sensitivity maps exhibit strong symmetry about the midline $y = 0.5$.

4.3 Confidence Intervals for Sobol Indices Point Estimates

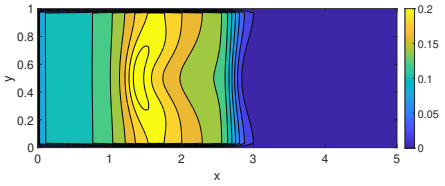
The results presented in Section 4.2 indicate that variability in diffusivity is a key factor in the overall variability in pollutant concentration throughout the domain. However, similar to Chapter 3, these results are point estimates for the Sobol indices of each parameter. We now compute the 95% CIs for the Sobol indices to accurately make observations about the variance contributed by each parameter. In this section, we select a few key spatial locations at which we observed a significant variance contribution by each parameter. In the 2D domain, computing CIs throughout the entire domain can be very costly. Choosing a few key points to compute the CIs at is more cost-effective. We implement this using the bootstrapping technique described in Section 2.6 by taking the function evaluations that are used to compute the estimates of



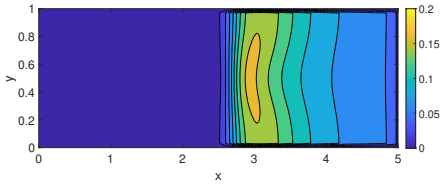
(a) First-order indices of s_1



(b) First-order indices of s_2



(c) Total effect indices of s_1



(d) Total effect indices of s_2

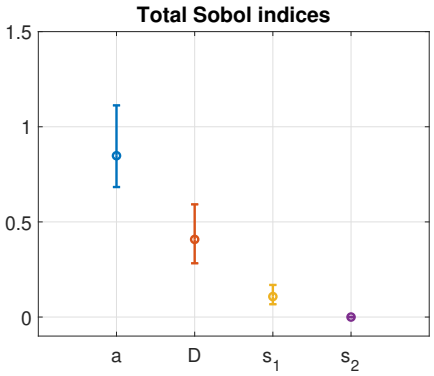
Figure 4.12: Spatial distribution of Sobol indices of source strength using $D_{new} \sim$ Lognormal$(0.19, 0.79)$.

the Sobol indices and resampling with replacement $B = 10^3$ times, using the same method described in Section 2.6.

Based on Figure 4.2 and Figure 4.3, we choose the query points $\boldsymbol{qp}_1 = (1.0, 0.5)$, $\boldsymbol{qp}_2 = (1.75, 0.5)$ $\boldsymbol{qp}_3 = (4.5, 0.5)$ to evaluate the CIs upstream of the first source location $\boldsymbol{x}_{s_1}$, between the two sources, and downstream of the second source location $\boldsymbol{x}_{s_2}$, respectively. The CIs at the query points are illustrated in Figure 4.13. All plots are shown on the same scale for comparability reasons.

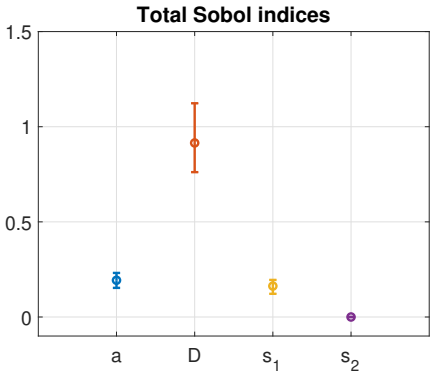
The CIs shows the influence of wind speed and diffusivity throughout the domain. At $\boldsymbol{qp}_1$, wind speed clearly has the greatest impact despite also having the widest CI. Figure 4.13a shows that the range of the wind speed CI compared with the CIs for diffusivity and the source strengths still indicate that wind speed is the most influential parameter at $\boldsymbol{qp}_1$. However, we observe that the impact of diffusivity at $\boldsymbol{qp}_1$ is made more prominent by interactions, as we see in Figure 4.13b. We also note that the CI for the total effect of s_1 indicates that it is the more influential source strength at $\boldsymbol{qp}_1$. At the query points $\boldsymbol{qp}_2$ and $\boldsymbol{qp}_3$, we observe that the diffusivity has a much larger impact than all other other variables for both first-order and total effect indices. Although the width of the CIs for the Sobol indices of diffusivity are notably the largest, the impact of the other variables is very small in comparison to the impact of diffusivity. These results indicate that as we move further downstream of the sources, the wind speed loses prominence very quickly in both the first-order and total effect indices.

(a) $\boldsymbol{qp}_1 = (1.0, 0.5)$



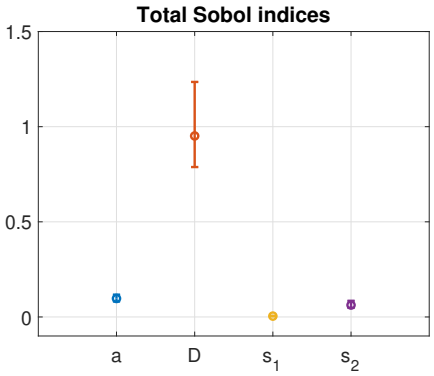
(b) $\boldsymbol{qp}_1 = (1.0, 0.5)$

(c) $\boldsymbol{qp}_2 = (1.75, 0.5)$



(d) $\boldsymbol{qp}_2 = (1.75, 0.5)$

(e) $\boldsymbol{qp}_3 = (4.5, 0.5)$



(f) $\boldsymbol{qp}_3 = (4.5, 0.5)$

Figure 4.13: Confidence intervals of Sobol indices at query points with $x_{s_1} = 1.5$ and $x_{s_2} = 3$. Computed using Quasi-Monte Carlo method with sample size $m = 10^4$.

4.4 Estimating Spatially-Dependent Sobol Indices using a Surrogate Model

While our previous simulations provide valuable insight into the pollutant concentration throughout the domain, computational challenges arise as estimating Sobol indices typically requires a large number of model evaluations to achieve convergence. Here this means that we have CIs narrow enough to accurately assess the relative influence of each parameter. Each model evaluation involves numerically solving a PDE, leading to a computationally expensive process. This is where surrogate modelling becomes indispensable.

Surrogate models are approximations of computational models used to reduce computational costs and enable efficient analysis of complex systems. The motivation for employing a surrogate model stems from the need for efficient global sensitivity analysis and uncertainty quantification. We use the MATLAB Regression Learner App to train and validate multiple regression models. The Regression Learner App enables us to build predictive models from simulated pollutant concentration data, making it particularly useful for developing surrogate models. Figure 4.14 shows the typical process of training regression models using the Regression Learner App. We select the generated training data and all models to be evaluated in parallel, as well as selecting the validation approach. The Regression Learner App automatically trains all available models and assess model performance based on root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE) and the R-squared statistics, such that we are able to easily compare and select the "best" regression model for our data [46]. Here we focus on the MSE of the predicted data compared with the simulated pollutant concentration data, in order to select the best model. Then we can export the best model and test the surrogate model's ability to capture the pollutant dispersion model uncertainty.

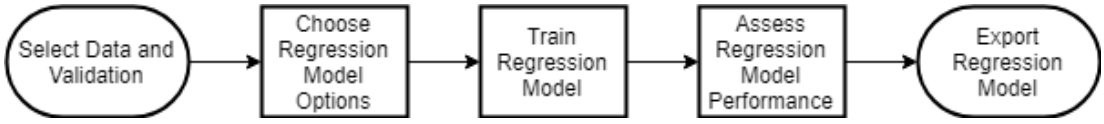


Figure 4.14: Workflow for training regression models using the MATLAB Regression Leaner App [46].

4.4.1 Surrogate Approximation using an Artificial Neural Network

We use an feedforward neural network (NN) as a surrogate model for the pollutant dispersion problem because it achieved the lowest MSE of validation data compared with other methods tested. These included linear regression models, regression trees, support vector machines, Gaussian process regression, kernel-based models, and ensembles of regression trees. Within the neural network category, a range of architectures was explored, including narrow, medium, and wide networks, as well as multi-layered designs, with the chosen narrow NN providing the best performance [46]

Pollutant dispersion processes often exhibit nonlinear dependencies on input parameters and NNs are well-suited to capture such complex, nonlinear relationships without requiring explicit analytical formulations. The NN is trained using a dataset of simulated pollutant concentrations at selected spatial locations. The inputs to the model consist of a design matrix with columns corresponding to the parameters a, D, s_1 and s_2, while the outputs are the pollutant concentrations at the selected spatial locations. A separate model is trained for each query point, and no data pre-processing, such as scaling or normalisation, is applied. A Quasi-Monte Carlo sampling approach was selected to provide a space filling design. This ensures a more comprehensive exploration of the input domain, which is particularly important in high-dimensional uncertainty quantification tasks. Additionally, a well-distributed training set leads to lower approximation error in the NN surrogate model.

The NN used consists of a single fully connected hidden layer with a layer size of 10 nodes. Each node uses the ReLU activation function. The output layer, which produces the prediction of the continous response variable, has a single node with no activation function. During training, no regularisation techniques were implemented to prevent overfitting for simplicity and computational efficiency. However, the Regression Learner App protects against over-fitting by applying 5-fold cross-validation. Prior to applying cross-validation, 10% of the training dataset is held back for testing.

Cross-validation is a model validation technique used to evaluate the predictive performance of a model by partitioning the dataset into k-folds. Here, the data is divided into $k = 5$ disjoint subsets, and in each iteration, one fold is set aside as the validation set while the remaining folds are used for training. The model is trained on the training folds and then evaluated on the validation fold. This process is repeated for all folds. The average validation error across all iterations is then computed, providing a robust estimate of the model's predictive accuracy when trained on the full data set. Although cross-validation requires fitting the model multiple

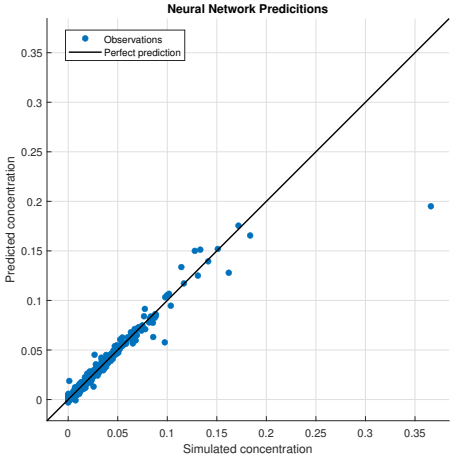
times, it maximises the use of available data.

To assess how the training sample size, m_{train}, influences the predictive performance of the NN, we conduct an experiment to explore the relationship between the training sample size m_{train} and the MSE at the same three query points as above: $\boldsymbol{qp}_1 = (1, 0.5)$, $\boldsymbol{qp}_2 = (1.75, 0.5)$, and $\boldsymbol{qp}_3 = (4.5, 0.5)$, using input parameters sampled from the distributions specified in (2.3). The MSE is evaluated on an independent validation set. Table 4.1 shows that the MSE decreases as we increase the sample size for each query point.

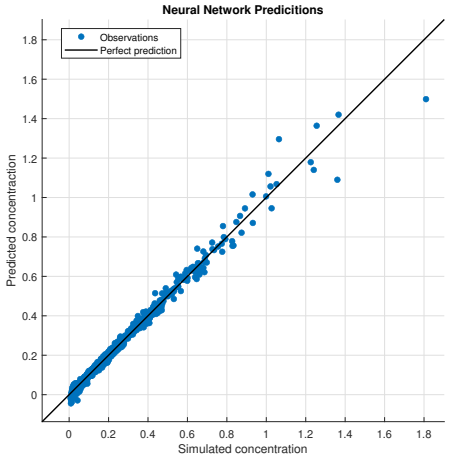
	MSE		
m_{train}	$\boldsymbol{qp}_1$	$\boldsymbol{qp}_2$	$\boldsymbol{qp}_3$
50	4.9023E-04	1.1011E-02	1.1329E-02
100	1.9356E-04	2.5038E-03	6.5185E-03
500	4.8198E-05	8.1870E-04	6.4670E-04
1000	4.5320E-05	5.1189E-04	4.9856E-04
5000	4.4471E-05	3.1216E-04	1.1223E-04
10000	4.1086E-05	2.8140E-04	9.9705E-05

Table 4.1: MSE of the pollutant concentration using feedforward NNs with respect to number of training samples.

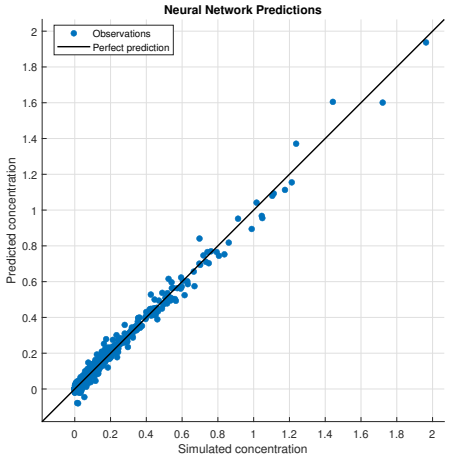
We implement the NN surrogate model using a training sample size of $m_{train} = 10^3$, which provides a balance between sufficient data for learning the complex input–output mapping and keeping the computational cost manageable. Figure 4.15 shows response plots comparing the simulated pollutant concentration values with the predicted NN output at each query point using training data. We observe that there is strong correlation between the observed and predicted pollution concentration levels. Hence, the overall prediction capabilities of the NN are strong. Furthermore, the cost of using the NN to compute Sobol indices is negligible in comparison to computing Sobol indices using function evaluations with the SUPG method. While the NNs incur a one-time upfront cost of training, which takes approximately 20 seconds, using MATLAB's stopwatch, we find that estimating Sobol indices takes approximately 58 seconds using the trained NN. In comparison, estimating Sobol indices using the full PDE-based model takes approximately 1,352 seconds. Therefore, we see a reduction of almost two orders of magnitude in evaluation time using the feedforward NN. We study the CIs of the Sobol estimates computed using the NN surrogate model at the same query points as in Section 4.3 for ease of comparison.



(a) $\boldsymbol{qp}_1 = (1, 0.5)$



(b) $\boldsymbol{qp}_2 = (1.75, 0.5)$



(c) $\boldsymbol{qp}_3 = (4.5, 0.5)$

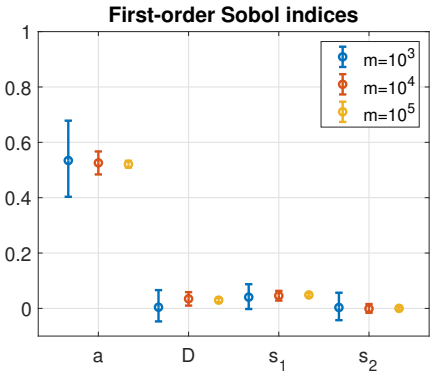
Figure 4.15: Response plots comparing the pollutant concentration simulations with the predicted feedforward NN output at specific query points, generated using a training dataset of $m_{train} = 10^3$ samples.

4.4.2 Study of Sobol Indices using a Surrogate Model

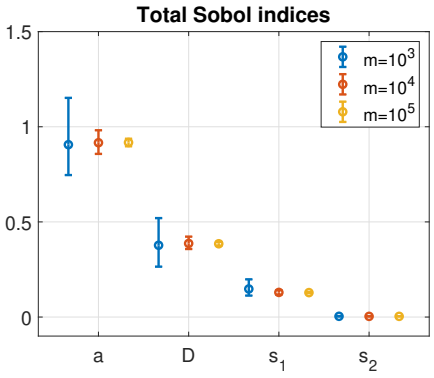
We compute the first-order and total effect Sobol indices as detailed in Section 4.1, where the parameters of interest are taken from the probability distributions given in Table 3.2, as before. As the computational cost of the surrogate model is cheaper than numerically solving the BVP (2.13), we can perform the sensitivity analysis using larger sample sizes. We are interested in studying the CIs of the Sobol indices. This is done by performing bootstrapping on the predicted concentration levels obtained by Quasi-Monte Carlo sampling using the feedforward NN. Figure 4.16 shows the point estimates and the 95% CIs at the query points $\boldsymbol{qp}_1 = (1.0, 0.5)$, $\boldsymbol{qp}_2 = (1.75, 0.5)$, and $\boldsymbol{qp}_3 = (4.5, 0.5)$ using increasing sample sizes. We observe that as the sample sizes increase, the estimates of the Sobol indices converge rapidly. This allows us to better approximate the true value for the first-order and total effect Sobol indices at the query points. In particular, we consider the Sobol indices with $m = 10^5$ samples. At $\boldsymbol{qp}_1$, the variance in the wind speed a contributes substantially to the output variance, with non-negligible interactions involving the diffusivity D. At $\boldsymbol{qp}_2$, D becomes the dominant factor, and we note that the impact of interactions diminish significantly. Meanwhile, at $\boldsymbol{qp}_3$, we deduce that D alone accounts for nearly all variance and the impact of a becomes negligible. While the indices for s_1 and s_2 remain small across all scenarios, we note that at $\boldsymbol{qp}_1$ and $\boldsymbol{qp}_2$, the variance in s_1 also contributes uncertainty to the model. This is because $\boldsymbol{qp}_1$ is located upstream of $\boldsymbol{x}_{s_1}$, and $\boldsymbol{qp}_2$ is located downstream. Both these locations pick up the impact of s_1. This is further highlighted by the total effect indices of s_1 and $\boldsymbol{qp}_1$ and $\boldsymbol{qp}_2$, which show the impact that s_1 has when we consider how it interacts with other parameters. We also see a similar effect for s_2 at $\boldsymbol{qp}_3$, which is located downstream of $\boldsymbol{x}_{s_2}$.

4.5 Summary

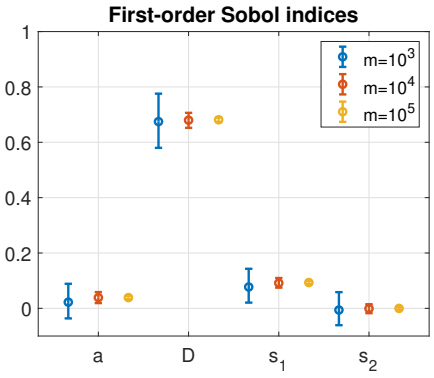
In this chapter, we have studied a simple BVP pollutant dispersal model in a 2D domain and described a more efficient method of estimating Sobol indices using existing samples. This method use km fewer function evaluations compared to the method we used in Chapter 3. The results of our spatial study of Sobol indices show that the uncertainty in wind speed and diffusivity have a significant impact on the variability in pollutant concentration throughout the domain. In addition, we calculated CIs at selected query points to ensure the reliability of our findings in order to draw accurate conclusions. Furthermore, the experiments demonstrate that a reduction in the variance of the input parameter distributions for wind speed and dif-



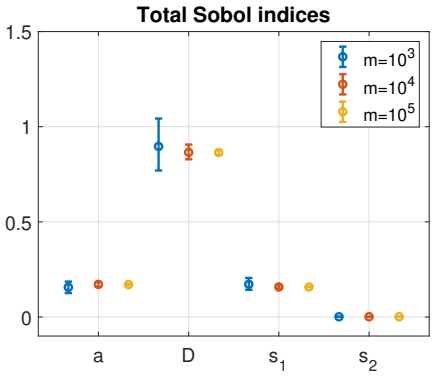
(a) $\boldsymbol{qp}_1 = (1.0, 0.5)$



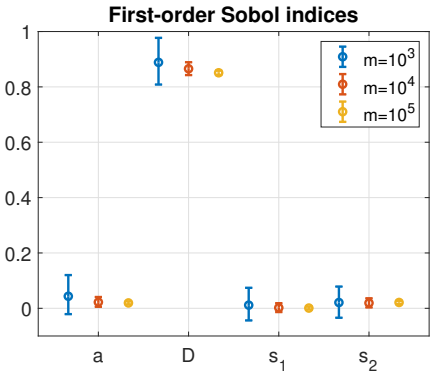
(b) $\boldsymbol{qp}_1 = (1.0, 0.5)$



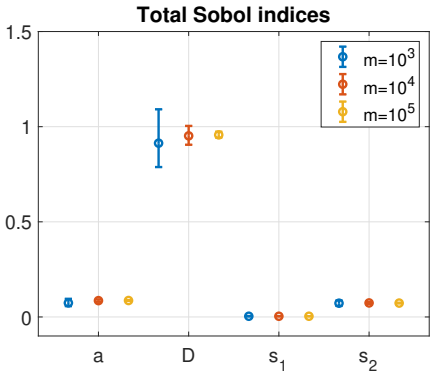
(c) $\boldsymbol{qp}_2 = (1.75, 0.5)$



(d) $\boldsymbol{qp}_2 = (1.75, 0.5)$



(e) $\boldsymbol{qp}_3 = (4.5, 0.5)$



(f) $\boldsymbol{qp}_3 = (4.5, 0.5)$

Figure 4.16: Confidence intervals of Sobol indices at query points, computed with a feedforward neural network.

fusivity can lead to a substantial decrease in the variance observed in the resulting pollutant concentration levels.

We also trained a surrogate model using using simulated pollutant concentration data to carry out further analysis of the CIs of the first-order and total effect Sobol indices at a lower computational cost with larger sample sizes. The mesh size $h = 1/32$ was chosen by testing different mesh sizes to ensure that the solution was sufficiently resolved while remaining computationally feasible, following the 1D mesh refinement study presented in Section 3.1. We observe that the results of the sensitivity analysis converge rapidly. Such results show the potential usefulness of surrogate models in estimating sensitivity measures for informing and improving pollutant dispersal models.

Chapter 5

Sensitivity Analysis for an Atmospheric Dispersion Model

The pollutant dispersion model presented in Section 2.3 serves as a good basis for understanding how Sobol indices can be applied to an advection-diffusion BVP. However, in this chapter we are interested in developing a model which more realistically reflects a wide range of atmospheric physics properties. We do this by adapting the 2D steady-state advection-diffusion model (2.13) to incorporate the effects of turbulent eddy diffusion and allowing the impact of the atmospheric boundary layer to be modelled within the wind speed. This approach to modelling an atmospheric dispersion problem attempts to address atmospheric turbulence which is responsible for most turbulent transports.

5.1 An Atmospheric Dispersion Model

We begin by considering the domain $\Omega_2 = \{(x, y) : 0 \leq x \leq b_1, 0 \leq y \leq b_2\}$ to be a horizontal cross section at a fixed height z_r above the surface at which the pollutant is emitted. We make the assumption that, at this height, variations in ground topography are negligible, therefore the wind flow is assumed to be horizontal. Furthermore, as we want to focus on the steady-state pollutant dispersion model, all physical properties are considered to be time invariant.

Atmospheric turbulence is an important property in atmospheric dispersion modelling. A key characteristic of turbulent flows is that velocity components may vary randomly at any location in the domain. Although sensor data can measure wind speeds, the presence of such

random fluctuations are difficult to measure in practice. To account for this, the wind velocity can be expressed as the sum of a deterministic and a random component [4, 62],

$$\boldsymbol{a} = \bar{\boldsymbol{a}} + \boldsymbol{a}', \tag{5.1}$$

where $\bar{\boldsymbol{a}}$ is the mean wind velocity and $\boldsymbol{a}'$ is a random perturbation with zero mean. That is, the wind velocity $\boldsymbol{a}$ is a random variable with mean $\bar{\boldsymbol{a}}$. Let

$$\boldsymbol{u} = \bar{\boldsymbol{u}} + \boldsymbol{u}',$$

where $\bar{\boldsymbol{u}}$ is a mean pollutant concentration and $\boldsymbol{u}'$ is the random perturbations in the pollutant concentration as a result of randomness in wind speed. Assuming that wind velocity and pollutant concentration are continuous functions in the domain Ω_2, we are interested in modelling the mean pollutant concentration.

We focus on short-range particulate transport over distances on the order of a few kilometres. This is discussed further in Section 5.1.1. On this length scale, we assume that molecular diffusion of the pollutant is negligible. Therefore, the diffusivity in the model is due to the presence of turbulent eddies. Eddy diffusion is a function of turbulence, modelled here by the wind velocity field (5.1), that describes the mixing of atmospheric fluids and masses. Prandtl's mixing-length theory describes the distance travelled by a particle at which it blends with surrounding atmospheric fluids [4]. This leads to the eddy diffusivity tensor

$$\boldsymbol{D} = \begin{bmatrix} D_{xx} & 0 \\ 0 & D_{yy} \end{bmatrix}. \tag{5.2}$$

In this study, we make the simplifying assumption that $D_{xx} = D_{yy} = D$. The use of a diffusion equation implies that the length scale of the transport is much smaller than the characteristic length over which the mean concentration is modelled [62].

In an urban setting, there are thousands of individual sources of pollutant emission. To simplify the model setting we will consider large sources, such as industrial facilities, modelled here as point sources. We consider a pollutant that is emitted at rates s_1 and s_2 from sources

at locations $\boldsymbol{\mu_{s_1}} = (\mu_{x_1}, \mu_{y_1})$ and $\boldsymbol{\mu_{s_2}} = (\mu_{x_2}, \mu_{y_2})$, respectively, such that

$$\begin{aligned} s(x,y) =& \frac{s_1}{\sigma\sqrt{2\pi}} \exp\left\{ \frac{-\left[(x-\mu_{x_1})^2 + (y-\mu_{y_1})^2\right]}{2\pi\sigma^2} \right\} \\ &+ \frac{s_2}{\sigma\sqrt{2\pi}} \exp\left\{ \frac{-\left[(x-\mu_{x_2})^2 + (y-\mu_{y_2})^2\right]}{2\pi\sigma^2} \right\}, \end{aligned}$$

with the spatial scaling parameter σ. Both source rates s_1 and s_2 are modelled at the same height z. The following BVP (5.3) describes the horizontal transport of the pollutant, such that

$$\begin{aligned} -D\nabla^2 u + a\mathbf{w}\cdot\nabla u = s(x,y) \quad \text{in} \quad \Omega_2, \\ u = 0 \quad \text{on} \quad \partial\Omega_2, \end{aligned} \tag{5.3}$$

where $D = D_{xx} = D_{yy}$, u is the mean concentration, and the wind velocity field $\boldsymbol{a}$ is decomposed into the mean wind speed, denoted a, and wind direction, denoted $\boldsymbol{w}$. We do this so that we are able to study the impact of wind speed independent of wind direction. Here, we assume constant wind direction.

The wind speed is modelled by the power-law correlation

$$a = a_r \left(\frac{z}{z_r}\right)^p, \tag{5.4}$$

where a_r denotes a reference wind speed at z_r, the height (in metres) at which we model the dispersion of the pollutant, and p is a fitting parameter which depends on atmospheric stability and surface roughness [25]. We note that mean transport wind speeds are typically assumed to be uniform in the horizontal direction within an urban boundary layer [4].

The diffusivity of the pollutant is no longer considered an independent parameter in (5.3). We now model the diffusivity as a function of the reference wind speed a_r to reflect the properties of (horizontal) turbulent eddy diffusion, which incorporates the turbulent mixing of pollutants, such that [25, 62]

$$D \simeq 0.1 a_* z_i^{4/3} (-\kappa L)^{-1/3}, \tag{5.5}$$

where $a_* = \kappa a_r / \ln(z_r/z_0)$ is the friction velocity, expressed as a function of the roughness length z_0 and the von Karman constant κ. We note that [62] contains a typographical error in the exponent of z_i in the expression for eddy diffusivity, where it is incorrectly given as $3/4$. If we substitute this exponent into the expression (5.5), where a_* has units $\mathrm{m\,s^{-1}}$ and both z_i and L have units m, we obtain $D \sim a_* z_i^{3/4} L^{-1/3}$, which has units $\mathrm{m^{17/12}\,s^{-1}}$. This is not

dimensionally consistent with eddy diffusivity, which must have units of $\mathrm{m}^2\,\mathrm{s}^{-1}$. By correcting the exponent to 4/3, as used in (5.5), we obtain the correct units and ensure dimensional consistency. This error in the expression of (5.5) has been replicated in [25].

The height of the mixing layer is denoted by z_i in (5.5), where $z_i > z_r > z$. The Monin-Obukhov length L is estimated as

$$\frac{1}{L} = c + d\log_{10}(z_0), \tag{5.6}$$

where the parameters c and d are determined by the Pasquill stability class. The Pasquill stability classes are defined on the basis of routine observations, including surface wind speed (at 10 m), solar radiation, and night-time cloud cover [62]. Table 5.1 shows the Monin-Obukhov parameters for different stability classes.

Pasquill stability class	c	d
A (Extremely unstable)	-0.096	0.029
B (Moderately unstable)	-0.037	0.029
C (Slightly unstable)	-0.002	0.018
D (Neutral)	0	0
E (Slightly stable)	0.004	-0.018
F (Moderately stable)	0.035	-0.036

Table 5.1: Monin-Obukhov length parameters for different stability classes, taken from [25].

5.1.1 Uncertainty Quantification of an Atmospheric Dispersion Model

The components of the atmospheric dispersion model (5.3) discussed in the previous section have inherent uncertainty, stemming from variable environmental and physical conditions. To investigate this, we again perform a parametric study by treating key model parameters as random variables drawn from suitable probability distributions.

We aim to explore model behaviour across a representative range of environmental conditions. Table 5.2 shows the observed ranges taken by the parameters of the atmospheric dispersion model. The ranges for the Monin-Obukhov length L are computed using (5.6) with c and d taken from the unstable Pasquill class A, B, and C, as seen in Table 5.1. In our study, we use stability class B and a roughness length of $z_0 = 0.1$ to compute the value for L. We take the height of the mixing layer to be $z_i = 10^3$ m, and the height at which we model emission to be $z = 100$ m. The height at which we model the pollutant dispersion model is $z_r = 10$ m.

Parameter (units)	Symbol	Range
Wind speed exponent	p	[0.1, 0.4]
Monin-Obukhov length (m)	L	[-15, 100]
Height of mixing layer (m)	z_i	$[10^2, 3 \times 10^3]$

Table 5.2: Parameters of the atmospheric dispersion model (5.3) and their accepted ranges, taken from [62].

The von Karman constant $\kappa = 0.4$.

To incorporate these parameters into a stochastic framework, we assign probability distributions based on their physical characteristics and empirical evidence. All parameters are converted into consistent units, with the spatial domain defined in metres: $\Omega_2 = \{(x, y) : 0 \text{ m} \leq x \leq 1000 \text{ m}, 0 \text{ m} \leq y \leq 200 \text{ m}\}$. The spatial scaling parameter $\sigma = 20$.

The reference wind speed, denoted a_r, plays a central role in pollutant transport. Empirical studies show that wind speeds are well described by the Weibull distribution due to its flexibility and good fit to observational wind data [50]. Accordingly, we assume $a_r \sim \text{Weibull}(\lambda_r, k_r)$, where λ_r is a scale parameters and k_r is a shape parameter. We note that the specific values for λ_r and k_r are taken from [50], with wind speeds expressed in ms^{-1}. Full details can be found in Table 5.3.

The wind profile exponent p characterises how wind speed varies with height. This exponent is directly related to atmospheric stability, with $p = 0.1$ suggesting very stable and $p = 0.4$ suggesting very unstable atmospheric conditions [62]. In the absence of further information, we assume all values within this range are equally likely, and hence we use $p \sim \text{Uniform}(0.1, 0.4)$. This range is based on the typical bounds for p listed in Table 5.2.

To model the source strengths s_1 and s_2, we select the gamma distribution as it is continuous and has a strictly positive support. Hence we use

$$s_1 \sim \text{Gamma}(\alpha_1, \beta_1), \;\; s_2 \sim \text{Gamma}(\alpha_2, \beta_2),$$

where α_1, α_2 are shape parameters, and β_1, β_2 are rate parameters. To determine these hyper-parameters, we use engineering estimates derived from an inverse Gaussian plume model, using a linear least squares method [40]. The Gaussian plume model is a well-established pollutant emission model, which solves the advection-diffusion equation with ground-level deposition for known emission sources [4, 62, 66]. This forward solution is integrated into an inverse problem framework, where a linear least-squares approach is applied. By minimising the discrepancy

between predicted and observed pollutant concentrations, the emission rates are iteratively refined to achieve the best fit with the observed data. The emission rates are initially given in tons per year and converted to grams per second as follows:

$$s_{eng,1} = \frac{30 \times 10^6}{3.154 \times 10^7}, \quad s_{eng,2} = \frac{85 \times 10^6}{3.154 \times 10^7}.$$

We solve for α_j, β_j by computing the inverse cumulative distribution function, such that the mode of the gamma function is equal to the engineering estimates $s_{eng,j}$ of the point sources:

$$\frac{\alpha_j - 1}{\beta_j} = s_{\text{eng },j},$$
$$\text{qgamma}\,(0.99, \alpha_j, \beta_j) = 3 s_{\text{eng },j}.$$

where $j = 1, 2$ and qgamma is the quantile function for the gamma distribution. We fit the distribution of the source strength in this way by assuming that the 99th percentile of the gamma distribution will be 3 times its mode, following [18, 66].

In summary, we consider the parameters modelled as random variables with the following probability distribution functions:

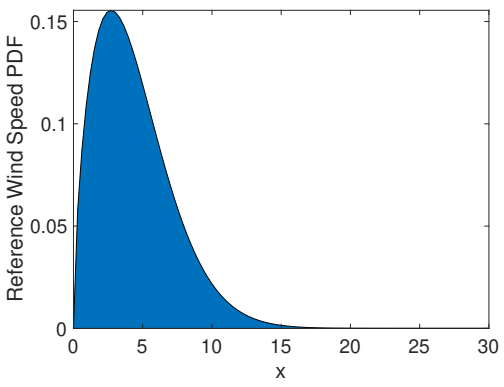
$$\begin{aligned} &a_r \sim \text{Weibull}(\lambda_r, k_r), \quad p \sim \text{Uniform}(0.1, 0.4), \\ &s_1 \sim \Gamma(4.7299, 0.2550), \qquad s_2 \sim \Gamma(4.7299, 0.7225). \end{aligned} \tag{5.7}$$

Table 5.3 provides the full specification of the sampled parameters, including their expected values and variances. The corresponding probability density functions are shown in Figure 5.1. All other parameters are kept constant initially, for the purposes of the sensitivity analysis conducted in Section 5.2. However, we note that other parameters may be considered in the parametric study by similarly sampling from relevant probability distributions.

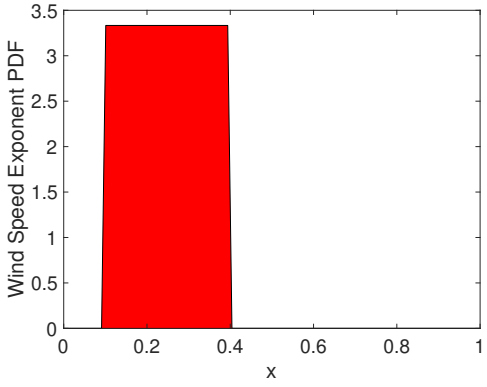
Parameter	Distribution	Expected Value	Variance
a_r	Weibull(4.92,1.62)	4.4062	7.7715
p	Uniform(0.1,0.4)	0.25	0.0075
s_1	Gamma(4.7299,0.2550)	1.2061	0.3076
s_2	Gamma(4.7299, 0.7225)	3.4174	2.4690

Table 5.3: Summary of parameter distributions, expected values and variances.

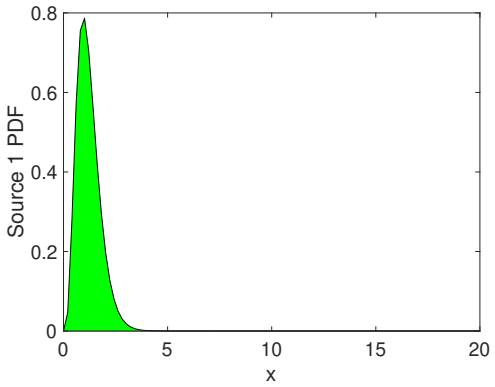
We are now able to conduct a parametric study of the atmospheric dispersion model (5.3). We approximate the solution of (5.3) using the SUPG method detailed in Section 2.4. Here,



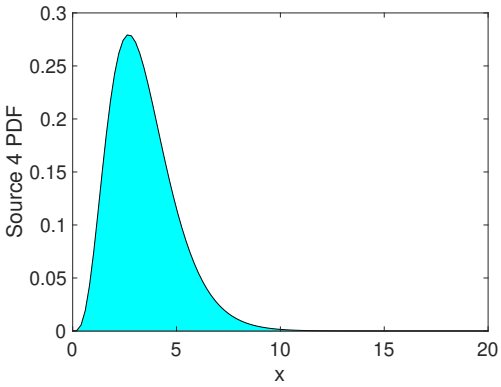
(a) $a_r \sim \text{Weibull}(4.92, 1.62)$



(b) $p \sim \text{Uniform}(0.1, 0.4)$



(c) $s_1 \sim \text{Gamma}(4.7299, 0.2550)$

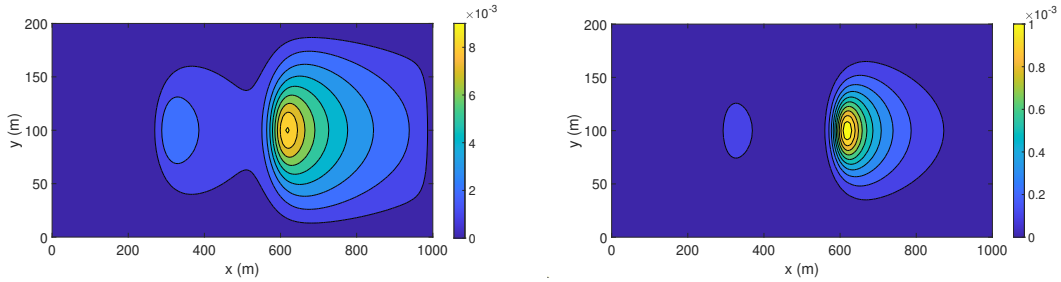


(d) $s_2 \sim \text{Gamma}(4.7299, 0.7225)$

Figure 5.1: Sampling distribution of pollutant dispersion parameters.

we used a uniform mesh of triangular elements with mesh size $h = 1/32$ to ensure a sufficiently resolved solution while maintaining computational feasibility. Dirichlet boundary conditions were implemented, and the MATLAB backslash operator was used as a direct solver to solve the linear system given by (2.30), with constant stabilisation parameter τ given by (2.27).

In the remainder of this chapter we use the source locations $\boldsymbol{x}_{s_1} = (300, 100)$ and $\boldsymbol{x}_{s_2} = (600, 100)$. We fix the wind direction $\boldsymbol{w} = [1, 0]$. Figure 5.2 shows the sample mean and variance of the pollutant concentration throughout the domain. We observe that the pollutant concentration tends to be higher as we move downstream of the sources due to transport effects, and note that the mean and variance of the pollutant concentration are symmetric about $y = 100$, as expected. Figure 5.2b indicates that most of the variance is concentrated downstream of $\boldsymbol{x}_{\boldsymbol{s_2}}$.



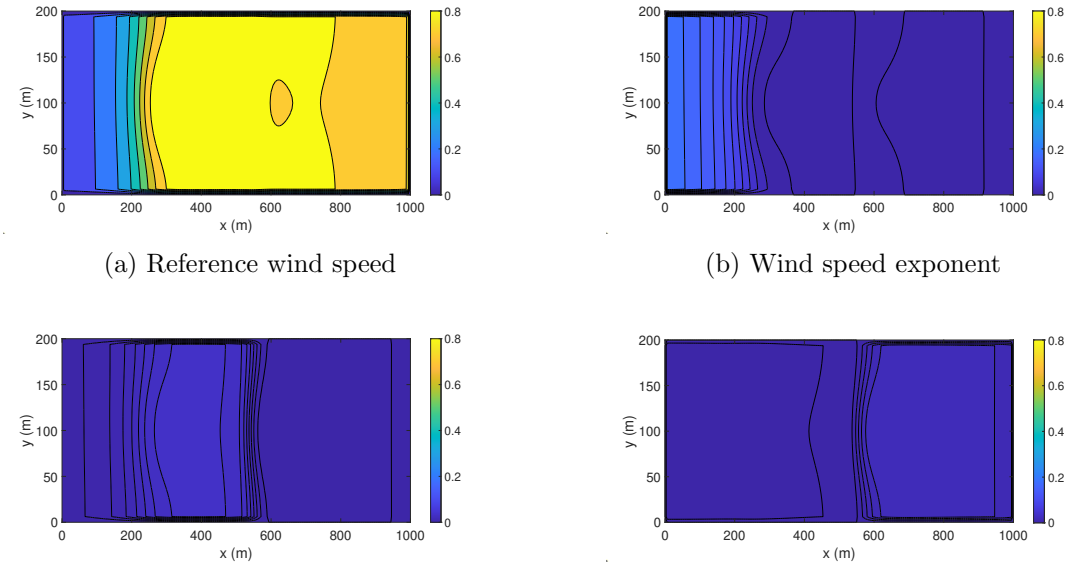
(a) Mean pollutant concentration

(b) Variance of pollutant Concentration

Figure 5.2: Pollutant concentrations were computed using the SUPG finite element discretisation on a uniform mesh with size $h = 1/32$, using $m = 10^4$ Quasi-Monte Carlo samples.

5.2 Study of Spatially-Dependent Sobol Indices

We now estimate the first-order and total effect Sobol indices to study the variance contribution of each parameter using the method described in Section 4.1. Figure 5.3 shows the spatial dependence of the first-order Sobol indices. Recall that these indices quantify the proportion of variance each individual parameter contributes to the overall variance in the pollutant concentration. We can clearly see that the most influential parameter is the reference wind speed a_r. In particular, we note that the relative impact of a_r increases as we move downstream in the domain. The wind speed exponent p upstream of $\boldsymbol{x}_{s_1}$ contributes a small proportion of uncertainty contributed by p in that region. However, in comparison to the first-order indices of a_r, the first-order indices of p and the source strengths s_1 and s_2 are very small. Figures



(a) Reference wind speed

(b) Wind speed exponent

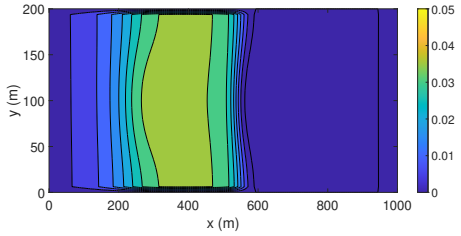
(c) Source strength 1

(d) Source strength 2

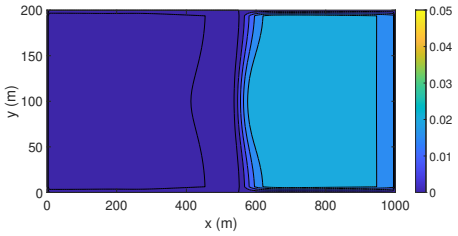
Figure 5.3: Spatial distribution of first-order Sobol indices.

5.4a and 5.4b provide a closer look at the first-order Sobol indices for source strengths. We see that the indices for s_1 are smaller than the indices for s_2, which aligns with our intuition since the variance of s_1 is also smaller than the variance of s_2. We can also see that the first-order indices of s_1 peak around $\boldsymbol{x}_{s_1}$, rapidly declining as we approach $\boldsymbol{x}_{s_2}$. The first-order indices of s_2 peak downstream of $\boldsymbol{x}_{s_2}$. Finally, we note that all the first-order indices are less than one, indicating that the first-order indices alone do not capture all the variance contribution towards the pollutant concentration.

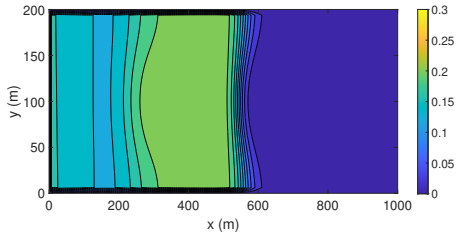
We now study the total effect indices to evaluate the relative impact that interactions between the parameters have on the variance of the pollutant concentration. The total effect indices presented in Figure 5.5 highlight that the uncertainty contributed by each parameter becomes more pronounced than in Figure 5.3, confirming that parameter interactions play a significant role in the variance of the pollutant concentration. The total effect indices of p show that the proportion of variance contributed by p upstream of $\boldsymbol{x}_{s_1}$ is in fact significant. This was not properly captured by the first-order indices. Furthermore, the relative impact of the source strengths is also increased with respect to their first-order indices, as seen in Figures 5.4c and 5.4d. In particular, we note that the impact of s_1 upstream of $\boldsymbol{x}_{s_1}$ is now much more significant than we observed for the first-order indices. However, despite the increased impact of p, s_1 and s_2, the most influential parameter throughout the domain remains a_r.



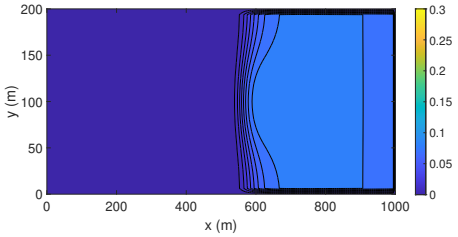
(a) First-order indices of s_1



(b) First-order indices of s_2

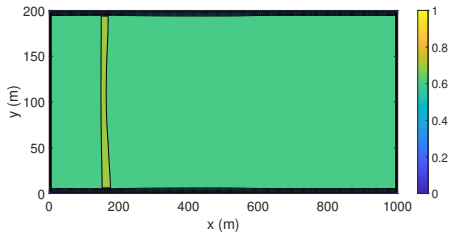


(c) Total effect indices of s_1

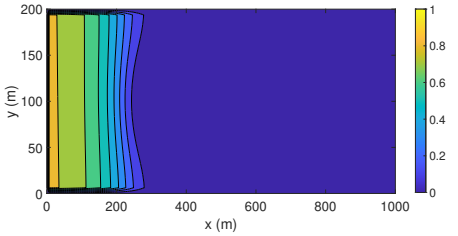


(d) Total effect indices of s_2

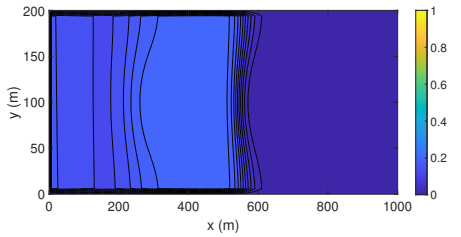
Figure 5.4: Spatial distribution of Sobol indices of source strengths.



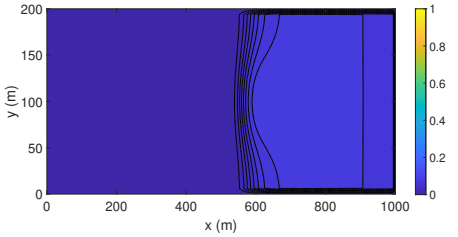
(a) Reference wind speed



(b) Wind speed exponent



(c) Source strength 1

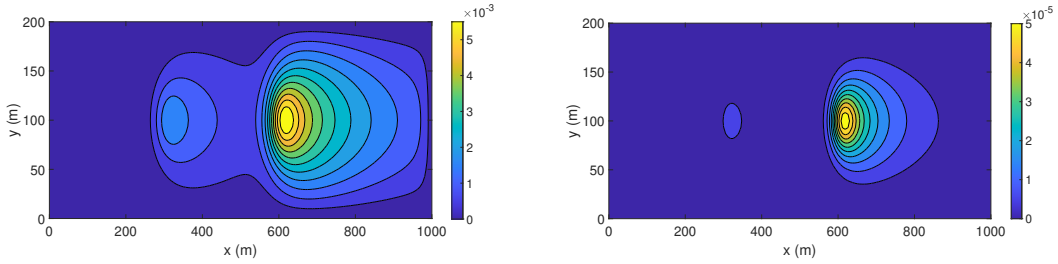


(d) Source strength 2

Figure 5.5: Spatial distribution of total effect Sobol indices.

Since the spatial study of Sobol indices suggests that the reference wind speed a_r is the most significant parameter contributing uncertainty to the overall pollutant concentration, we now conduct a further investigation into how changes in a_r affects variability in pollutant concentration. This is done by repeating the simulations (as in Section 3.3.2) with the mean reference wind speed fixed, but reducing the variance by half. Specifically, we use $a_{new} \sim$ Weibull$(4.97, 2.38)$ with $\mathbb{E}(a_{new}) = 4.4062$ and $\mathbb{V}(a_{r_2}) = 3.8857$, as before. A comparison of the probability density functions of the two distributions for a_r and a_{new} can be found in Figure 3.7.

Figure 5.6 shows the variability in the pollutant concentration using the new wind speed distribution, with all other parameter distributions unchanged. Direct comparison between



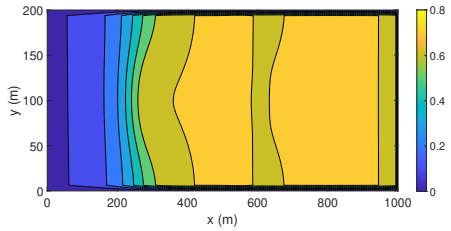
(a) Mean pollutant concentration (b) Variance of pollutant concentration

Figure 5.6: Variability in pollutant concentration using $a_{new} \sim$ Weibull$(4.97, 2.38)$.

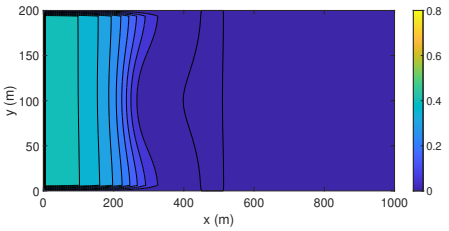
Figure 5.6b and Figure 5.2b shows a reduction in the variance of the pollutant concentration by two orders of magnitude. We now estimate the Sobol indices with respect to the new reference wind speed a_{new} so that we can study how the proportion of variance contributed by each parameter has changed.

Figure 5.7 shows the first-order indices using the new reference wind speed a_{new}. Comparing with Figure 5.3, we observe that there is a decrease in the first-order indices of a_{new}. In fact, we can see that as a result of the reduced variance in a_{new}, the relative impact of p, s_1 and s_2 has increased significantly. In particular, we note that the first-order indices of p becomes the dominant parameter contributing uncertainty to the pollutant concentration upstream of $\boldsymbol{x}_{s_1}$. Elsewhere in the domain, the new reference wind speed a_{new} remains the dominant contributor of uncertainty.

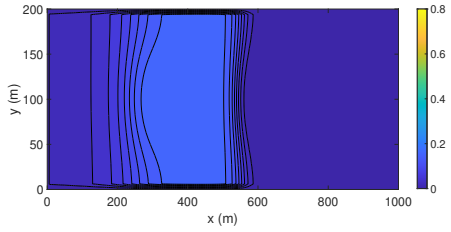
Figure 5.8 shows the total effect indices for the pollutant dispersion problem using the new reference wind speed a_{new}. We now see that the impact of uncertainty in all parameters has increased, highlighting the role that interaction between parameters plays with respect to the new reference wind speed. Upstream of $\boldsymbol{x}_{s_1}$, the most influential parameter is p, while down-



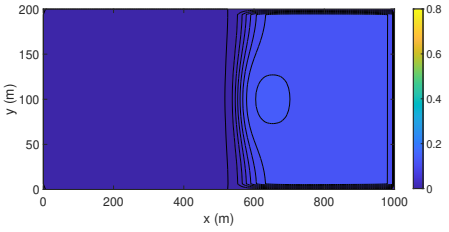
(a) Reference Wind Speed



(b) Wind Speed Exponent

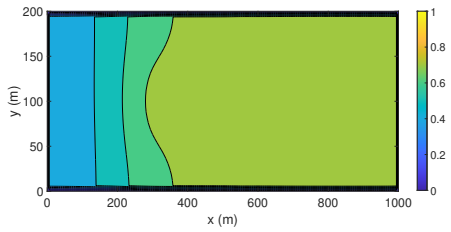


(c) Source Strength 1

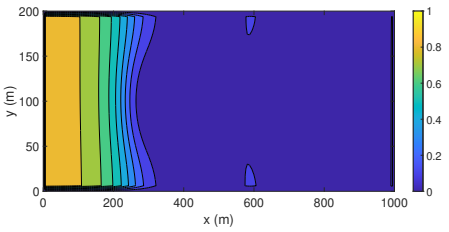


(d) Source Strength 2

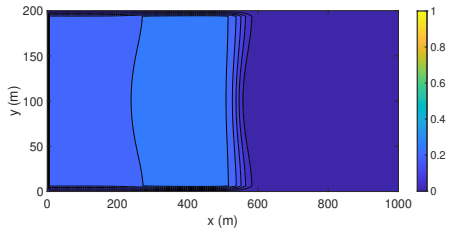
Figure 5.7: Spatial distribution of first-order Sobol indices using $a_{new} \sim \text{Weibull}(4.97, 2.38)$.



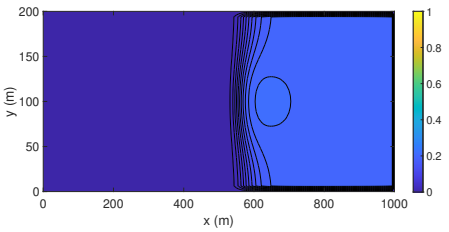
(a) Reference Wind Speed



(b) Wind Speed exponent



(c) Source Strength 1



(d) Source Strength 2

Figure 5.8: Spatial distribution of total effect Sobol indices using $a_{new} \sim \text{Weibull}(4.97, 2.38)$.

stream, the most influential parameter remains a_{new}. The relative impact of s_1 has increased in comparison to Figure 5.5c; however the impact of s_2 has remained similar to Figure 5.5d.

5.3 Confidence Intervals for Sobol Indices Point Estimates

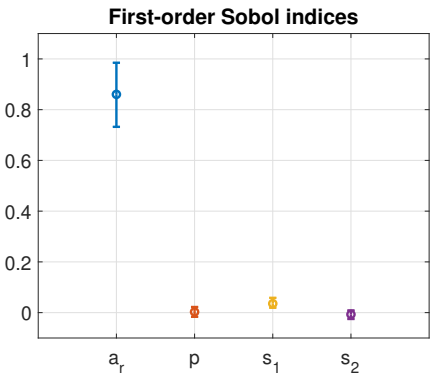
The results presented in Section 5.2 indicate that the variability in the reference wind speed a_r is a key factor in the overall variability in the pollutant concentration throughout the domain. We are interested in computing the 95% CIs for the Sobol indices to accurately make observations about the variance contributed by each parameter.

As in Section 4.3, in this section we select a few spatial locations at which we observe a significant variance contribution by each parameter in order to study the CIs in a cost-effective manner. We estimate the CIs using the bootstrapping technique described in Section 2.6 by taking the function evaluations that are used to compute the estimates of the Sobol indices and resampling with replacement $B = 10^3$ times (that is, using the same method described in Section 3.3.3).

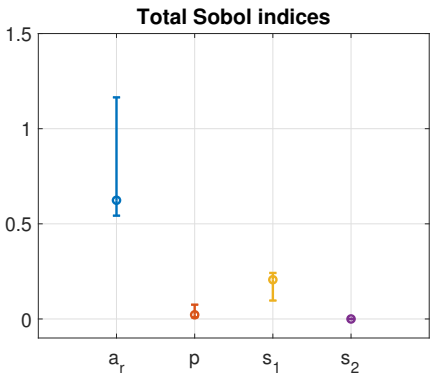
Based on Figure 5.3 and Figure 5.5, we choose the query points $\boldsymbol{qp}_1 = (275, 100)$, $\boldsymbol{qp}_2 = (450, 100)$, and $\boldsymbol{qp}_3 = (900, 100)$ to evaluate the CI. Not that these observation points lie upstream of the first source location $\boldsymbol{x}_{s_1}$, between the two sources, and downstream of the second source location $\boldsymbol{x}_{s_2}$, respectively. The CIs at the query points are illustrated by the point estimates with the CIs as error bars shown in Figure 5.9. All plots are shown on the same scale for comparability reasons.

We observe that across all query points, the reference wind speed a_r is clearly the most influential parameter, exhibiting the highest first-order and total effect indices throughout, despite having wider CIs. This indicates that uncertainty in the reference wind speed contributes most significantly to the overall model variance. We note that at $\boldsymbol{qp}_2$ the CI for the total effect index of s_1 is greater than that of p or s_2, highlight that downstream of the first source location $\boldsymbol{x}_{s_1}$, we see the impact that the uncertainty of s_1 has on the overall variance in the pollutant concentration. Similarly, at $\boldsymbol{qp}_3$, the CI for the total effect index of s_2 is greater than that of p or s_1, because the impact of s_2 is highlighted downstream of the second source location $\boldsymbol{x}_{s_2}$. These CIs confirm the overall trends observed in Figures 5.3 and 5.5.

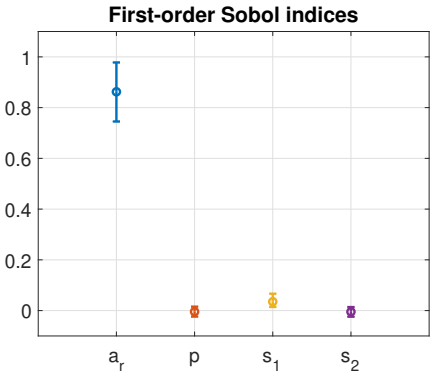
We now compare the CIs in Figure 5.9 to the CIs of the second experiment with $a_{new} \sim$ Weibull$(4.97, 2.38)$ in Figure 5.10. Here we observe that the widths of the CIs for the Sobol indices of a_{new} are smaller at all query points, due to the reduced variance in a_{new}. As a result of the narrower CIs, the relative influence of the model parameters becomes more apparent. Across all query points, the reference wind speed a_r remains the most influential parameter, exhibiting the largest Sobol indices. Figure 5.10 highlights that reducing the variance in a key



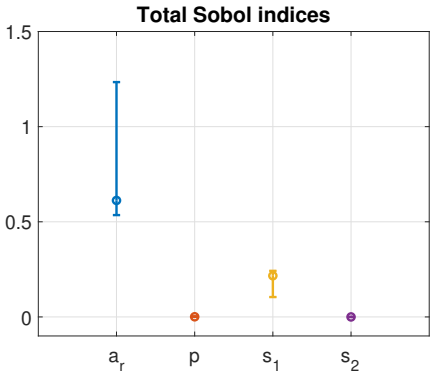
(a) $\boldsymbol{qp}_1 = (275, 100)$



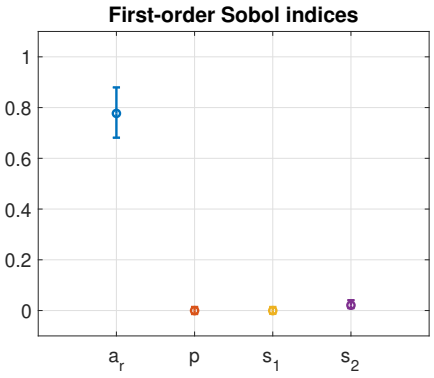
(b) $\boldsymbol{qp}_1 = (275, 100)$



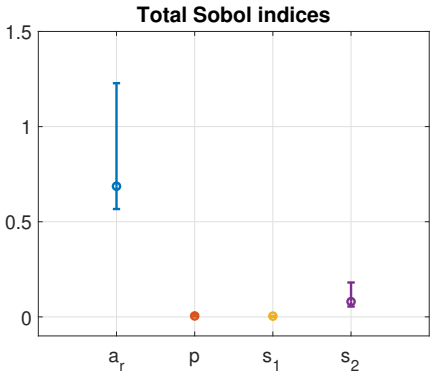
(c) $\boldsymbol{qp}_2 = (450, 100)$



(d) $\boldsymbol{qp}_2 = (450, 100)$

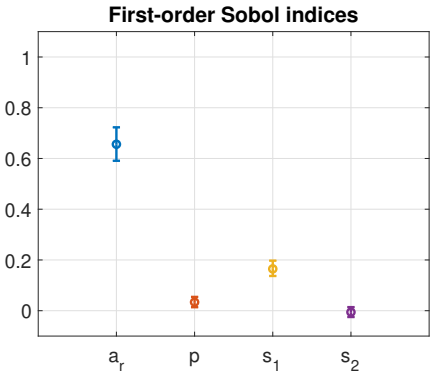


(e) $\boldsymbol{qp}_3 = (900, 100)$

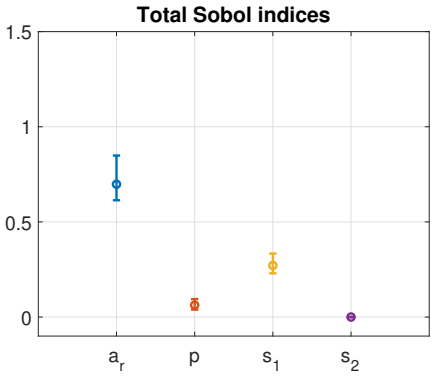


(f) $\boldsymbol{qp}_3 = (900, 100)$

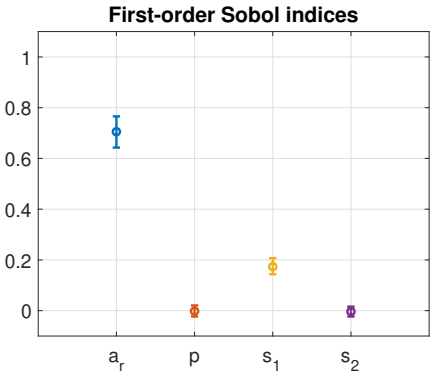
Figure 5.9: Point estimates and 95% confidence intervals of Sobol indices at three query points with $\boldsymbol{x}_{s_1} = (300, 100)$ and $\boldsymbol{x}_{s_2} = (600, 100)$, computed using Quasi-Monte Carlo method with sample size $m = 10^4$.



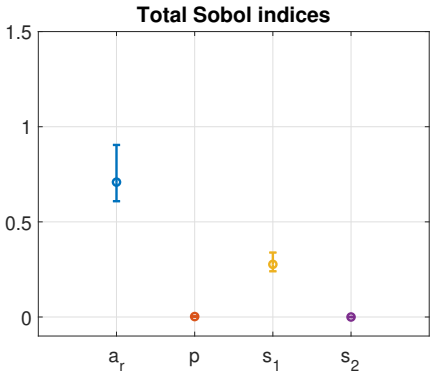
(a) $\boldsymbol{qp}_1 = (275, 100)$



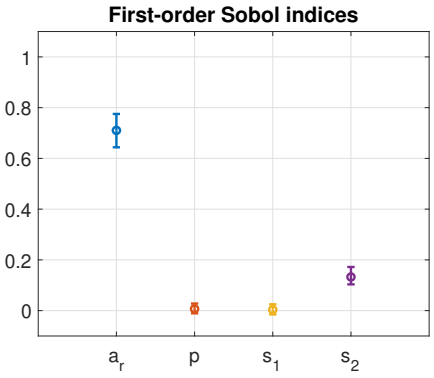
(b) $\boldsymbol{qp}_1 = (275, 100)$



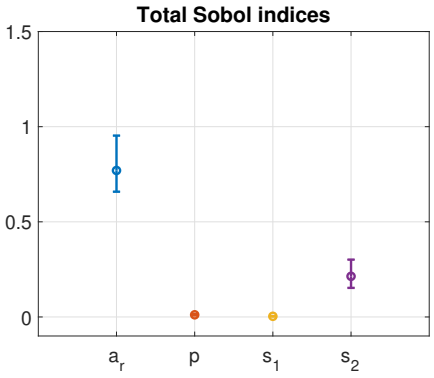
(c) $\boldsymbol{qp}_2 = (450, 100)$



(d) $\boldsymbol{qp}_2 = (450, 100)$



(e) $\boldsymbol{qp}_3 = (900, 100)$



(f) $\boldsymbol{qp}_3 = (900, 100)$

Figure 5.10: Point estimates and 95% confidence intervals of Sobol indices at three query points using $a_{new} \sim \text{Weibull}(4.97, 2.38)$.

parameter also reduces the CIs of the Sobol point estimates, making more reliable Sobol index estimation easier.

5.4 Estimating Spatially-Dependent Sobol Indices using a Surrogate Model

In this section, we extend the surrogate modelling approach developed in Chapter 4.4 to efficiently compute Sobol indices for the atmospheric dispersion model (5.3). In this study, we again use a feedforward neural network (NN) as the surrogate model and employ the MATLAB Regression Learner App to facilitate the training and validation of surrogate models. For details on the workflow of the Regression Learner App, see Figure 4.14. The evaluation criteria for selecting the surrogate model are discussed in Section 4.4.

5.4.1 Atmospheric Dispersion Surrogate Model

The architecture for the feedforward NN remains consistent with the model used in Section 4.4.1: a single hidden layer comprising 10 fully connected nodes activated by the ReLU function, followed by a output layer with one node. While the input layer differs in terms of the variables used, which now consists of the design matrix with columns representing samples of a_r, p, s_1 and s_2, the number of input nodes remains the same, and no data pre-processing such as scaling or normalisation was applied. The surrogate model is trained on a dataset of simulated pollutant concentration outputs generated via the SUPG method for the atmospheric dispersion model (5.3). This is done in a pointwise manner for specific spatial locations, and the NN is trained separately at each query point. In this section, we focus on the same query points as before, $\boldsymbol{qp}_1 = (275, 100)$, $\boldsymbol{qp}_2 = (450, 100)$, and $\boldsymbol{qp}_3 = (900, 100)$. As before, the input parameters are sampled using a Quasi-Monte Carlo approach to ensure a well-distributed training set. The dataset was divided such that 10% was reserved for independent testing, while the remaining data was used in a 5-fold cross-validation procedure to train and validate the model, as described in Section 4.4. We estimate the pollutant concentration values using the NN.

To assess how the training sample size, m_{train}, influences the predictive performance of the NNs in the context of the atmospheric dispersion model at each query point, we conduct a similar experiment to that presented in Section 4.4.1. Results are summarised in Table 5.4, where we observe that increasing the training sample size leads to improved surrogate accuracy.

m_{train}	MSE		
	$\boldsymbol{qp}_1$	$\boldsymbol{qp}_2$	$\boldsymbol{qp}_3$
50	5.7977E-05	2.4307E-05	2.9949E-04
100	1.6253E-05	2.2648E-05	9.2920E-05
500	5.3110E-06	9.4105E-06	2.1175E-05
1000	5.0111E-06	5.3820E-06	1.7246E-05
5000	4.6760E-06	4.4503E-06	1.6082E-05
10000	3.0810E-06	4.1664E-06	9.2111E-06

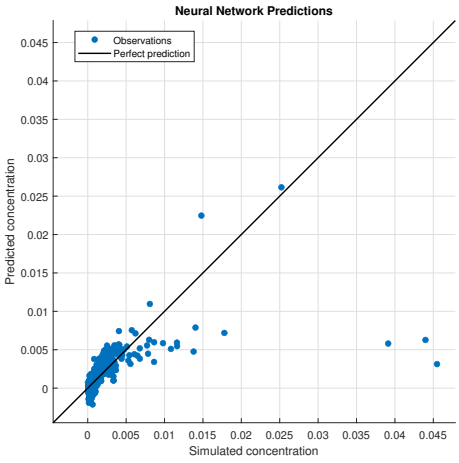
Table 5.4: MSE of the pollutant concentration prediction using feedforward NN with respect to size of training samples.

Figure 5.11 compares the predicted pollutant concentration levels of the surrogate model against the PDE-based pollutant concentration simulations at each query point. While the MSE values reported in Table 5.4 indicate low overall prediction error, the scatter plots in Figure 5.11 exhibit a more noticeable spread around the ideal prediction line. This apparent discrepancy arises because the neural network captures a moderate proportion of the variance in the simulated data, as indicated by the coefficient of determination, R^2. At all query points, the R^2 values range between 0.45 to 0.52. Nevertheless, we observe that the predicted values generally follow the trend of the simulated values indicating that the surrogate model demonstrates good predictive capability at the query points, capturing nonlinearities induced by the underlying turbulent dispersion processes. Thus the feedforward NNs demonstrate satisfactory overall performance, successfully learning the nonlinear mapping from input parameters to pollutant concentrations across most of the domain, and reliably capturing the dominant trends of the dispersion process.

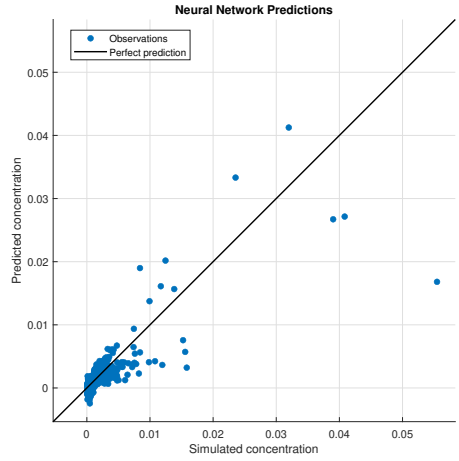
The trained NNs can now be used as a surrogate for computing Sobol indices. The negligible computational cost of evaluating the surrogate allows for extensive Quasi-Monte Carlo evaluations to estimate both first-order and total-effect indices. This enables an efficient sensitivity analysis of the atmospheric dispersion model. The results provide insight into the dominant drivers of variability within the atmospheric dispersion model (5.3), without the need for repeated high-fidelity PDE evaluations.

5.4.2 Study of Sobol Indices using a Surrogate Model

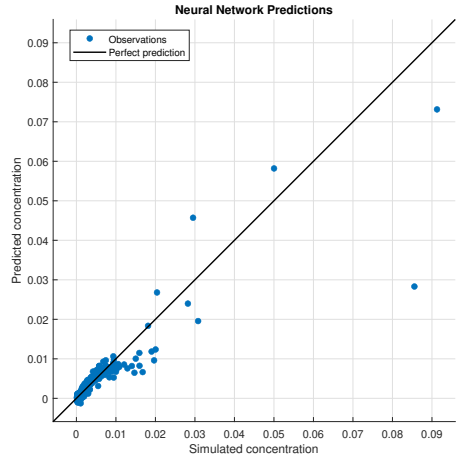
In this section, we compute Sobol indices using pollutant concentration predictions estimated from the feedforward NNs, with respect to the input parameter distributions defined in (5.7). The significantly lower computational cost of the surrogate model, compared to the full PDE-



(a) $\boldsymbol{qp}_1 = (275, 100)$



(b) $\boldsymbol{qp}_2 = (450, 100)$



(c) $\boldsymbol{qp}_3 = (900, 100)$

Figure 5.11: Response plots comparing the pollutant concentration simulations with the predicted feedforward NN output at specific query points, generated using a training dataset of $m_{train} = 10^3$ samples.

based pollutant dispersion solver, allows for the use of much larger sample sizes in the global sensitivity analysis. This enables a systematic investigation of how the CIs of the estimated Sobol indices behave with increasing sample size.

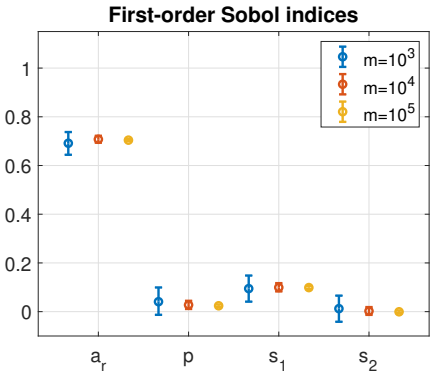
Using MATLAB's stopwatch, we find that the estimation of Sobol indices and CIs at selected query points required approximately between 46 seconds with the trained neural network surrogate, compared to approximately 5,832 seconds for the full PDE-based model. Although training the NNs incur an upfront cost of approximately 20 seconds each, the dramatic reduction in evaluation time, over two orders of magnitude, demonstrates the substantial computational efficiency gained by employing the surrogate model, enabling extensive global sensitivity analyses that would be otherwise prohibitively expensive.

Figure 5.12 presents the 95% CIs for the estimated Sobol indices at three spatial query points: $\boldsymbol{qp}_1 = (275, 100)$, $\boldsymbol{qp}_2 = (450, 100)$, and $\boldsymbol{qp}_3 = (900, 100)$. These CIs are again obtained by applying a bootstrapping procedure to the predicted concentration values generated using a Quasi-Monte Carlo sampling scheme applied to the surrogate model.

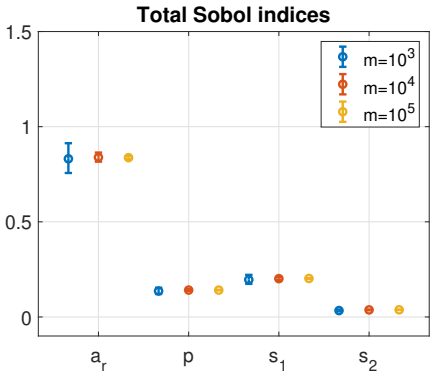
As the sample size increases, we observe that the Sobol index estimates converge rapidly, enabling more accurate approximation of the true values for both the first-order and total-effect indices at all specified query points. The CIs shown in Figure 5.12 are typically narrower than the intervals computed in Figure 5.9. We observe that the reference wind speed a_r is the most influential parameter contributing uncertainty to the surrogate model at all query points, which is consistent with the results presented in Figure 5.9.

In addition to the effect of reference wind speed, the CIs of first-order indices at $\boldsymbol{qp}_1$ indicate the impact of s_1 upstream of $\boldsymbol{x}_{s_1}$. We see a similar effect downstream of $\boldsymbol{x}_{s_1}$ at $\boldsymbol{qp}_2$ as well. At $\boldsymbol{qp}_3$, we observe that the first-order indices indicate that s_2 has a greater relative impact in this region, downstream of $\boldsymbol{x}_{s_2}$, than s_1. However, a_r remains the dominant contributor of uncertainty to the surrogate model.

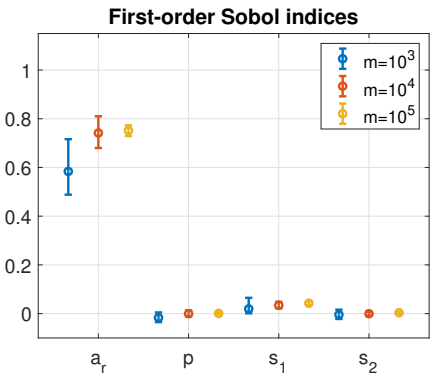
We note that the total effect indices at all the query points show that when parameter interactions are considered, the variance contributed by p, s_1 and s_2 to the surrogate model becomes more significant, whereas the first-order indices indicate that the effects of some of these parameters on their own can be considered negligible at certain query points. In particular, we note that at $\boldsymbol{qp}_1$, the total effect indices highlight the impact of s_1 and p, which tells us that the variance in the wind speed exponent, which describes atmospheric stability, has a greater impact on the uncertainty of the surrogate model upstream of $\boldsymbol{x}_{s_1}$. At $\boldsymbol{qp}_2$, the total effect indices highlight the impact of s_1 as sample size increases as $\boldsymbol{qp}_2$ is downstream of $\boldsymbol{x}_{s_1}$. Similarly,



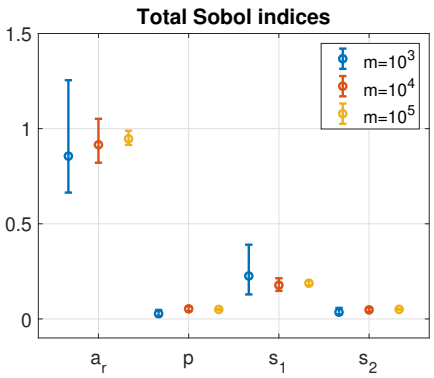
(a) $\boldsymbol{qp}_1 = (275, 100)$



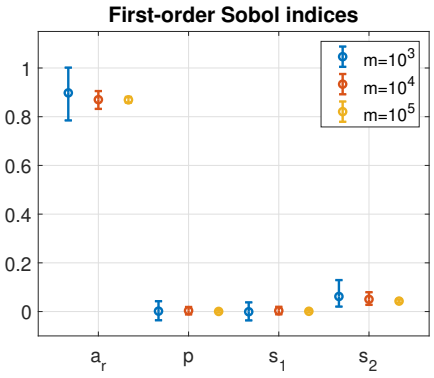
(b) $\boldsymbol{qp}_1 = (275, 100)$



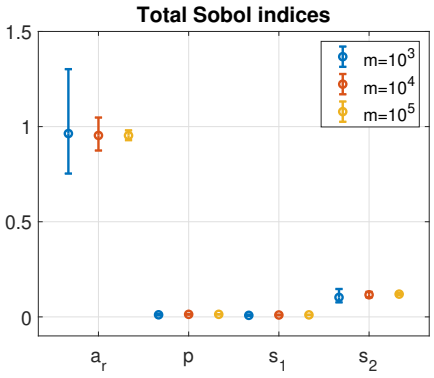
(c) $\boldsymbol{qp}_2 = (450, 100)$



(d) $\boldsymbol{qp}_2 = (450, 100)$



(e) $\boldsymbol{qp}_3 = (900, 100)$



(f) $\boldsymbol{qp}_3 = (900, 100)$

Figure 5.12: Point estimates and 95% confidence intervals of Sobol indices computed at selected spatial query points using neural network surrogate models trained at each query point respectively.

we note that at $\boldsymbol{qp}_3$, the total effect indices highlight the impact of s_2 as sample size increases as $\boldsymbol{qp}_3$ is downstream of $\boldsymbol{x}_{s_2}$. The total effect indices also support that the reference wind speed is the dominant contributor of uncertainty, with the impact of a_r also being heightened when we consider its interaction with other parameters.

The surrogate model provides an efficient means of capturing the sensitivity of the atmospheric dispersion model (5.3), despite the increased modelling complexity. It successfully reproduces the dominant trends and key sensitivity results, demonstrating its effectiveness for conducting sensitivity analysis in atmospheric modelling.

5.5 Summary

In this chapter, we have studied a more realistic pollutant dispersion problem in a 2D domain by modelling turbulent eddy diffusion. Here, we considered the parameters a_r, p, s_1 and s_2 when analysing the global sensitivity of the atmospheric dispersion model. The results of our spatial study of Sobol indices show that the uncertainty in reference wind speed a_r has the most significant impact on the variability in pollutant concentration throughout the domain. The CIs calculated at selected query points showed that despite wide intervals, a_r remained the most influential parameter, hence confirming the robustness of our conclusions. Further analysis demonstrated that reducing the variance in a_r leads to a significant reduction in the overall variance observed in the resulting pollutant concentration. Moreover, as the uncertainty in a_r decreases, the CIs become narrower, indicating increased accuracy in the estimates of the pollutant concentration. This is also observed in the variance for the pollutant concentration with the new reference wind speed. This finding highlights the critical importance of accurate measurements of input parameters such as wind speed when evaluating atmospheric dispersion.

To reduce the computational burden of full PDE evaluations, we also developed a surrogate model using a feedforward NN to approximate pollutant concentrations and facilitate efficient sensitivity analysis. The surrogate demonstrated good predictive accuracy at selected spatial locations, capturing the nonlinear dynamics of pollutant dispersion. By leveraging the surrogate model's low computational cost, we computed CIs of Sobol indices at selected query points. Results revealed spatial variability in dominant input parameters, such as the influence of s_1 and s_2 downstream of their respective emission sources and notable interaction effects of p upstream of $\boldsymbol{x}_{s_1}$. The NN supported the conclusion that uncertainty in a_r most significantly impacted the uncertainty in the surrogate model. Overall the surrogate effectively reproduced

the key sensitivity trends of the full model, providing a reliable and efficient approach for global sensitivity analysis in complex atmospheric pollutant dispersion modelling.

Chapter 6

Conclusions and Future Work

This thesis set out to advance the methodological foundations of spatially distributed global sensitivity analysis (GSA) for atmospheric models. Traditional GSA approaches often rely on scalar outputs, which can obscure critical spatial heterogeneities in model response. However, pollutant dispersion is inherently spatial in nature, and understanding how input uncertainty manifests across different regions of the domain is essential for a more complete interpretation of model behaviour.

The central aim of this research was to apply spatially-resolved Sobol index techniques to quantify how uncertainty in model inputs influences pollutant concentrations across a spatial domain. By explicitly incorporating spatial variability into the sensitivity analysis process, this work provides some informative insights on uncertainty quantification in PDE-based systems.

In this thesis, we have investigated spatially-resolved model uncertainties in steady-state air quality models by GSA techniques. In particular, our study focused on the use of Sobol indices to quantify how uncertainty in key input parameters propagates through advection-diffusion PDEs to affect pollutant concentrations throughout the spatial domains considered. The study progressed systematically from simple, idealised test cases to more realistic scenarios, with increasing physical complexity. Across all experiments, Quasi-Monte Carlo sampling was used to estimate first-order and total-effect Sobol indices, and non-parametric bootstrap resampling was employed to construct confidence intervals. To manage computational demands at larger scales, surrogate models based on artificial neural networks were employed, enabling efficient approximation of pollutant concentrations and scalable sensitivity analysis.

After a summary of key mathematical background is presented in Chapter 2, Chapter 3 introduces the underlying PDE problem in a one-dimensional setting, where the advection-

diffusion PDE is discretised using the Streamline Upwind Petrov–Galerkin (SUPG) finite element method. For details on the numerical method see Section 2.2. This controlled environment enables a detailed exploration of spatial sensitivity patterns resulting from uncertainty in the wind speed, diffusivity, and source strengths. Strong spatial variation in Sobol indices is revealed, with wind speed emerging as the dominant source of output uncertainty. These results establish a conceptual and computational benchmark for the rest of the study. Moreover, the experiments illustrate how targeted reduction in input uncertainty can significantly reduce output variance, providing a concrete example of the usefulness of spatially-dependent sensitivity information. Part of this work has been published in [41].

The experiment is extended to two-dimensional space in Chapter 4, where the computational burden associated with QMC-based sensitivity analysis increases substantially. To enable efficient spatially-resolved sensitivity analysis, we adapt the standard Sobol index estimation procedure to reduce computational cost while preserving accuracy. Instead of generating two separate resampled matrices for the first-order index calculation, as in Chapter 3, we reuse the original design matrices Z_1 and Z_2, and generate a single resampled matrix, formed by replacing the jth column of Z_1 with that of Z_2. Since each function evaluation corresponds to a full SUPG solve over the 2D domain, this revised approach significantly reduces the number of high-cost PDE solves required. resulting in a computational saving of $(k-1)m$ evaluations. This efficiency gain is especially valuable in high-resolution spatial domains where sensitivity indices are computed pointwise across a dense grid.

The Sobol indices computed are point estimates. We construct confidence intervals for the Sobol indices using bootstrap resampling at selected spatial points to provide insight into estimator reliability. To further reduce the computational burden associated with solving the full PDE model in Chapter 4, a feedforward neural network (NN) surrogate model is introduced and trained to approximate pollutant concentration outputs. NNs are chosen due to their ability to learn complex, nonlinear input–output relationships from limited training data. Once trained on a representative dataset of inputs and corresponding PDE solutions, the surrogate model serves as a fast, differentiable approximation of the forward model. Because the NN surrogate can produce significantly faster predictions than the PDE solver, it enables the computation of spatially-dependent Sobol indices across a broader range of sample sizes than would otherwise be manageable. The experimental results provide insight into the robustness of the surrogate-assisted sensitivity analysis and offer practical guidance on the sample sizes required to obtain reliable sensitivity indices in pollutant dispersion models.

In Chapter 5, we apply the above methodology to a physically motivated pollutant dispersion scenario with uncertain meteorological parameters, including the reference wind speed, atmospheric stability exponent, and emission rates. This model reflects more realistic atmospheric dynamics. The Sobol analysis reveals that while the wind speed remains dominant in most regions, the relative importance of other parameters shifts spatially, illustrating the value of location-specific sensitivity analysis. Once again, a feedforward NN surrogate model was trained on the full PDE solution to facilitate efficient computation of Sobol indices and their confidence intervals. The NN is employed to overcome the computational cost of large-scale uncertainty quantification, allowing us to examine the convergence and statistical reliability of the sensitivity estimates.

Overall, this thesis demonstrates that spatially-resolved GAS, supported by accurate surrogate modelling, provides a powerful framework for understanding how parametric uncertainty propagates through atmospheric dispersion models. By incrementally increasing model fidelity and numerical complexity, each chapter contributes to a coherent and rigorous exploration of spatial sensitivity in PDE-based environmental systems.

The implications of this work are relevant to both environmental modelling practice and methodological development. From a modelling perspective, the findings show that meaningful reductions in predictive uncertainty can be achieved by tightening distributions on key parameters, most notably wind speed, identified through spatial sensitivity patterns. This can be achieved on a practical level by deploying sensors to collect more accurate wind speed data. Targeted sensor placement, informed by spatial sensitivity maps, enables strategic data collection in regions where the model output is most sensitive to wind speed variability. This not only improves the reliability of model predictions but also ensures that limited observational resources are used most effectively.

From a methodological standpoint, this thesis demonstrates how surrogate modelling and uncertainty quantification techniques can be effectively combined to enable spatial GSA in complex environmental systems. By using a NN to approximate a computationally intensive air quality model, the analysis becomes feasible at the high sample sizes needed for reliable estimation of spatially dependent Sobol indices. The implementation of CI estimation further enhances the statistical rigour of the results, offering a practical way to quantify uncertainty in sensitivity metrics. Together, these elements form a scalable and generalisable approach that is well suited to models with high-dimensional input spaces and spatially distributed outputs. Beyond its specific application to atmospheric pollutant dispersion, the strategy extends more

broadly to other scientific domains where computational cost and spatial complexity limit the application of traditional uncertainty quantification methods. The use of surrogate models also opens avenues for multifidelity modelling approaches, where computationally cheap surrogates are combined with high-fidelity PDE solvers to balance accuracy and efficiency across scales of analysis.

While we demonstrate the effective use of NN surrogate models for spatial sensitivity analysis, the development and integration of multifidelity strategies were beyond the scope of this thesis. Incorporating multifidelity approaches typically requires thorough algorithmic frameworks to manage the interactions between models of differing resolutions and accuracies, as well as rigorous error control mechanisms. These requirements introduce additional methodological and computational challenges that that were beyond the focus of the current research. Nonetheless, multifidelity modelling represents a promising direction to extend the applicability of spatial SA to more computationally demanding and realistic atmospheric systems.

Improving the accuracy of surrogate models remains an important consideration in extending their applicability. In the air quality surrogate modelling study presented in Chapter 5, some minor discrepancies in the first-order Sobol indices were identified upstream of the first source location, highlighting localised challenges in capturing sensitivity information. These limitations, though not critical to the overall analysis, suggest that more advanced surrogate architectures may offer improved performance. Emerging research on physics-informed neural networks (PINNs) offers promising avenues to embed governing equations directly into surrogate training, potentially enhancing accuracy and robustness [43]. This remains an active area of investigation that could improve surrogate fidelity in future applications.

A further consideration is that this thesis did not incorporate observational data for model calibration or validation. All experiments were conducted using simulated outputs, limiting direct applicability to real-world scenarios. Integration of spatially-resolved sensitivity analysis with observational datasets, potentially through data assimilation, would greatly enhance the practical relevance and impact of the methodology developed here [16,56]. We also note that the input parameter spaces considered were deliberately limited to a small number of key uncertain parameters in each case. While this allowed for focused exploration of dominant variance contributors, real-world applications often involve higher-dimensional uncertainty with potentially complex parameter interactions. Investigating the scalability of the spatial sensitivity approach and surrogate modelling techniques to higher-dimensional parameter spaces remains an open challenge.

Other natural extensions of this work could include broadening the spatial GSA to encompass time-dependent and three-dimensional pollutant dispersion models. Addressing temporal evolution and vertical transport would better capture the complexities of atmospheric dynamics, offering a more comprehensive characterisation of uncertainty propagation. Additionally, the model could be extended to account for more complex source configurations, such as non-collinear sources or multiple sources with differing emission profiles, as well as wind fields that are not aligned with the centreline of the domain. Such an extension would necessitate enhanced surrogate modelling strategies capable of handling spatio-temporal outputs and increased computational demands, aligning with the challenges highlighted regarding multifidelity and surrogate accuracy.

Moving toward practical application, leveraging real-time sensor networks to iteratively inform and update spatial sensitivity analysis offers a promising avenue. Such a dynamic approach would enable continuous refinement of uncertainty characterisation, allowing models to adapt responsively to evolving environmental conditions. Employing targeted data acquisition strategies, such as adaptive sampling informed by current sensitivity patterns, can enhance the efficiency and accuracy of surrogate model training, particularly in regions exhibiting complex or poorly understood behaviour. This integration of data-driven refinement with spatial sensitivity analysis has the potential to transform static uncertainty assessments into agile tools for environmental monitoring and decision-making.

Applying spatially-resolved GSA beyond air quality to other environmental systems with spatially distributed processes, such as wildfire spread, groundwater contamination, or flood risk modelling, presents a promising avenue to extend the impact of this methodology. Each of these applications would benefit from spatially explicit insights into uncertainty propagation, potentially informing more targeted monitoring and mitigation strategies.

To summarise, this thesis has examined how spatially-resolved GSA can enhance the understanding of air quality models. It is hoped that the contributions presented will serve as a foundation for further work at the intersection of environmental modelling and efficient uncertainty quantification , ultimately supporting more informed and responsible decision-making for addressing one of the most pressing challenges in modern society.

Bibliography

[1] A. Allard and N. Fischer, *Sensitivity Analysis in Metrology: Study and Comparison on Different Indices for Measurement Uncertainty*, in Advanced Mathematical and Computational Tools for Metrology and Testing VIII, M. Bär and A. B. Forbes, eds., Singapore, 2009, World Scientific Publishing, pp. 1–6.

[2] H. Akimoto, *Global Air Quality and Pollution*, Science, 302 (2003), pp. 1716–1719.

[3] I. Annesi-Maesano, *The air of Europe: where are we going?*, European Respiratory Review, 26 (2017), p. 170024.

[4] S. P. Arya, *Air pollution meteorology and dispersion*, Oxford University Press, New York, 1999, pp. 287–299.

[5] E. Atanassov and S. Ivanovska, *On the Use of Sobol' Sequence for High Dimensional Simulation*, Springer International Publishing, 2022, pp. 646–652.

[6] D. A. Ayejoto, J. C. Agbasi, V. E. Nwazelibe, J. C. Egbueri, and J. O. Alao, *Understanding the connections between climate change, air pollution, and human health in Africa: Insights from a literature review*, Journal of Environmental Science and Health, Part C, 41 (2023), pp. 77–120.

[7] F. Campolongo, S. Tarantola, and A. Saltelli, *Tackling quantitatively large dimensionality problems*, Computer Physics Communications, 117 (1999), pp. 75–85.

[8] C. W. Champ and A. V. Sills, *The Generalized Laws of Total Variance and Total Covariance*, in Applied Mathematical Analysis and Computations I, D. Wanduku, S. Zheng, H. Zhou, Z. Chen, A. Sills, and E. Agyingi, eds., Cham, 2024, Springer Nature Switzerland, pp. 243–254.

[9] H.-S. Cho and M. Choi, *Effects of Compact Urban Development on Air Pollution: Empirical Evidence from Korea*, Sustainability, 6 (2014), pp. 5968–5982.

[10] R. Y. Choi, A. S. Coyner, J. Kalpathy-Cramer, M. F. Chiang, and J. P. Campbell, *Introduction to Machine Learning, Neural Networks, and Deep Learning.*, Translational Vision Science & Technology, 9 (2020), p. 14.

[11] J. E. Clougherty, I. Kheirbek, H. M. Eisl, Z. Ross, G. Pezeshki, J. E. Gorczynski, S. Johnson, S. Markowitz, D. Kass, and T. Matte, *Intra-urban spatial variability in wintertime street-level concentrations of multiple combustion-related air pollutants: The New York City Community Air Survey (NYCCAS)*, Journal of Exposure Science & Environmental Epidemiology, 23 (2013), pp. 232–240.

[12] A. De Marco, P. Sicard, Z. Feng, E. Agathokleous, R. Alonso, V. Araminiene, A. Augustatis, O. Badea, J. C. Beasley, C. Branquinho, V. J. Bruckman, A. Collalti, R. David-Schwartz, M. Domingos, E. Du, H. Garcia Gomez, S. Hashimoto, Y. Hoshika, T. Jakovljevic, S. McNulty, E. Oksanen, Y. Omidi Khaniabadi, A. Prescher, C. J. Saitanis, H. Sase, A. Schmitz, G. Voigt, M. Watanabe, M. D. Wood, M. V. Kozlov, and E. Paoletti, *Strategic roadmap to assess forest vulnerability under air pollution and climate change*, Global Change Biology, 28 (2022), pp. 5062–5085.

[13] K. Duan, G. Sun, Y. Zhang, K. Yahya, K. Wang, J. M. Madden, P. V. Caldwell, E. C. Cohen, and S. G. McNulty, *Impact of air pollution induced climate change on water availability and ecosystem productivity in the conterminous united states*, Climatic Change, 140 (2016), pp. 259–272.

[14] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall/CRC, May 1994.

[15] H. Elman, D. Silvester, and A. Wathen, *Finite Elements and Fast Iterative Solvers: With Applications in Incompressible Fluid Dynamics*, Oxford University Press, 2 ed., 2014, pp. 234–254.

[16] A. Franks, A. D'Amour, and A. Feller, *Flexible sensitivity analysis for observational studies without observable implications*, Journal of the American Statistical Association, 115 (2019), pp. 1730–1746.

[17] F. Gamboa, A. Janon, T. Klein, A. Lagnoux, and C. Prieur, *Statistical inference for Sobol pick-freeze Monte Carlo method*, Statistics, 50 (2015), pp. 881–902.

[18] J. G. García, B. Hosseini, and J. M. Stockie, *Simultaneous model calibration and source inversion in atmospheric dispersion models*, Pure and Applied Geophysics, 178 (2019), pp. 757–776.

[19] R. C. Gilliam, C. Hogrefe, J. M. Godowitch, S. Napelenok, R. Mathur, and S. T. Rao, *Impact of inherent meteorology uncertainty on air quality model predictions*, Journal of Geophysical Research: Atmospheres, 120 (2015).

[20] P. Glasserman, *Generating Random Numbers and Random Variables*, Springer New York, 2004, pp. 39–77.

[21] L. T. Hans-Görg Roos, Martin Stynes, *Robust Numerical Methods for Singularly Perturbed Differential Equations*, Springer Berlin Heidelberg, 2 ed., 2008.

[22] T. Hastie, R. Tibshirani, and J. Friedman, *Neural Networks*, Springer New York, Dec. 2008, pp. 389–416.

[23] B. Hoffmann, H. Boogaard, A. de Nazelle, Z. J. Andersen, M. Abramson, M. Brauer, B. Brunekreef, F. Forastiere, W. Huang, H. Kan, J. D. Kaufman, K. Katsouyanni, M. Krzyzanowski, N. Kuenzli, F. Laden, M. Nieuwenhuijsen, A. Mustapha, P. Powell, M. Rice, A. Roca-Barceló, C. J. Roscoe, A. Soares, K. Straif, and G. Thurston, *Who air quality guidelines 2021–aiming for healthier air for all: A joint statement by medical, public health, scientific societies and patient representative organisations*, International Journal of Public Health, 66 (2021).

[24] P. Holnicki and Z. Nahorski, *Emission Data Uncertainty in Urban Air Quality Modeling—Case Study*, Environmental Modeling & Assessment, 20 (2015), pp. 583–597.

[25] B. Hosseini and J. M. Stockie, *Estimating airborne particulate emissions using a finite-volume forward solver coupled with a Bayesian inversion approach*, Computers & Fluids, 154 (2017), pp. 27–43.

[26] A. L. Højberg and J. C. Refsgaard, *Model uncertainty–parameter uncertainty versus conceptual models.*, Water science and technology : a journal of the International Association on Water Pollution Research, 52 (2005), pp. 177–186.

[27] M. Imam, S. Adam, S. Dev, and N. Nesa, *Air quality monitoring using statistical learning models for sustainable environment*, Intelligent Systems with Applications, 22 (2024), p. 200333.

[28] B. Ingalls, *Sensitivity analysis: from model parameters to system behaviour*, Essays in Biochemistry, 45 (2008), pp. 177–194.

[29] S. C. Izah, L. Sylva, M. C. Ogwu, A. Shahsavani, S. Bazzazpour, and M. Rahmatinia, *Modeling and Statistical Approaches for Air Pollution Analysis*, Springer Nature Switzerland, 2024, pp. 425–456.

[30] A. Janon, T. Klein, A. Lagnoux, M. Nodet, and C. Prieur, *Asymptotic normality and efficiency of two Sobol index estimators*, ESAIM: Probability and Statistics, 18 (2014), pp. 342–364.

[31] R. R. Karri, G. Ravindran, N. M. Mubarak, B. Ambade, M. H. Dehghani, S. K. M. Hassan, and T. Vera, *Overview of diseases and health consequences due to air pollution*, Elsevier, 2024, pp. 1–26.

[32] S. Khan and Q. Hassan, *Review of developments in air quality modelling and air quality dispersion models*, Journal of Environmental Engineering and Science, 16 (2021), pp. 1–10.

[33] K. Khoiron, A. N. Probandari, W. Setyaningsih, H. S. Kasjono, R. H. Setyobudi, and O. Anne, *A review of environmental health impact from municipal solid waste (MSW) landfill*, Annals of Tropical Medicine and Public Health, 23 (2020), pp. 152–159.

[34] S. Kucherenko and S. Song, *Derivative-Based Global Sensitivity Measures and Their Link with Sobol' Sensitivity Indices*, Springer International Publishing, 2016, pp. 455–469.

[35] S. Kucherenko and S. Song, *Different numerical estimators for main effect global sensitivity indices*, Reliability Engineering & System Safety, 165 (2017), pp. 222–238.

[36] V. Kumar and C. Vellapandian, *Unraveling the Nexus Between Ambient Air Pollutants and Cardiovascular Morbidity: Mechanistic Insights and Therapeutic Horizons*, Cureus, (2024).

[37] H. H. Le, H. K. Ngo, L. S. Hieu, P. D. Sly, D. Phung, P. Le An, N. N. Vinh, T. N. Dang, D. T. H. Thuong, H. N. Thanh, and P. K. Thai, *Indoor air pollution is associated with respiratory symptoms in children in urban Vietnam*, Science of The Total Environment, 917 (2024), p. 170556.

[38] Á. Leelőssy, F. Molnár, F. Izsák, Á. Havasi, I. Lagzi, and R. Mészáros, *Dispersion modeling of air pollutants in the atmosphere: a review*, Open Geosciences, 6 (2014).

[39] C. Lemieux, *Monte Carlo and Quasi-Monte Carlo Sampling*, Springer Series in Statistics Ser., Springer New York, New York, 2009, pp. 347–368.

[40] E. Lushi and J. M. Stockie, *An inverse Gaussian plume approach for estimating atmospheric pollutant emissions from multiple point sources*, Atmospheric Environment, 44 (2010), pp. 1097–1107.

[41] J. A. Mackenzie, A. Ramage, T. Shahid, and A. B. Forbes, *Global Sensitivity Analysis of a Pollutant Dispersion Problem*, in Advanced Mathematical and Computational Tools for Metrology and Testing XIII, F. Pavese, A. Bosnjakovic, S. Eichstädt, A. B. Forbes, and J. A. Sousa, eds., Singapore, 2025, World Scientific, pp. 249–259.

[42] C. Malings, K. E. Knowland, N. Pavlovic, J. G. Coughlin, D. King, C. Keller, S. Cohn, and R. V. Martin, *Air quality estimation and forecasting via data fusion with uncertainty quantification: Theoretical framework and preliminary results*, Journal of Geophysical Research: Machine Learning and Computation, 1 (2024).

[43] S. Manavi, T. Becker, and E. Fattahi, *Enhanced surrogate modelling of heat conduction problems using physics-informed neural network framework*, International Communications in Heat and Mass Transfer, 142 (2023), p. 106662.

[44] A. Marrel, B. Iooss, M. Jullien, B. Laurent, and E. Volkova, *Global sensitivity analysis for models with spatially dependent outputs*, Environmetrics, 22 (2010), pp. 383–397.

[45] A. Marrel, N. Saint-Geours, and M. D. Lozzo, *Handbook of Uncertainty Quantification*, Springer International Publishing, 2017, ch. 39, pp. 1328–1355.

[46] MathWorks, *MATLAB Regression Learner App Online Documentation*. `https://uk.mathworks.com/help/stats/regression-learner-app.html`. [Accessed 27 July 2025].

[47] R. T. McNider and A. Pour-Biazar, *Meteorological modeling relevant to mesoscale and regional air quality applications: a review*, Journal of the Air & Waste Management Association, 70 (2019), pp. 2–43.

[48] M. R. Miller, M. Di Cesare, S. Rahimzadeh, M. Adeoye, P. Perel, S. Taylor, S. Shrikhande, K. Armstrong-Walenczak, A. S. V. Shah, C. D. Berenstein, R. Vedanthan, E. N. Achiri, S. Mehta, A. M. Adeoye, D. Piñeiro, and F. J. Pinto, *Clearing the air to address pollution & cardiovascular health crisis*, Global Heart, 19 (2024), p. 82.

[49] A. R. Mitchell and D. F. Griffiths, *The Finite Difference Method in Partial Differential Equations*, Wiley, 1980.

[50] E. C. Morgan, M. Lackner, R. M. Vogel, and L. G. Baise, *Probability distributions for offshore wind speeds*, Energy Conversion and Management, 52 (2011), pp. 15–26.

[51] A. B. Owen, *Monte Carlo theory, methods and examples*, `https://artowen.su.domains/mc/`, 2013.

[52] C. Prieur and S. Tarantola, *Variance-based sensitivity analysis: Theory and estimation algorithms*, in Handbook of Uncertainty Quantification, Springer International Publishing, 2015, pp. 1–23.

[53] C. Pénard-Morand and I. Annesi-Maesano, *Air pollution: from sources of emissions to health effects*, Breathe, 1 (2004), pp. 108–119.

[54] E. Qian, B. Peherstorfer, D. O'Malley, V. V. Vesselinov, and K. Willcox, *Multifidelity Monte Carlo Estimation of Variance and Sensitivity Indices*, SIAM/ASA Journal on Uncertainty Quantification, 6 (2018), pp. 683–706.

[55] K. M. Ramachandran and C. P. Tsokos, *Empirical methods*, Elsevier, 2021, pp. 531–568.

[56] E. T. R. Rosenman and A. B. Owen, *Designing experiments informed by observational studies*, Journal of Causal Inference, 9 (2021), pp. 147–171.

[57] R. Safarov, Z. Shomanova, Y. Nossenko, E. Kopishev, Z. Bexeitova, and R. Kamatov, *Spatial Analysis of Air Pollutants in an Industrial City Using GIS-Based Techniques: A Case Study of Pavlodar, Kazakhstan*, Sustainability, 16 (2024), p. 7834.

[58] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola, *Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index*, Computer Physics Communications, 181 (2010), pp. 259–270.

[59] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global Sensitivity Analysis. The Primer*, Wiley, dec 2007.

[60] A. Saravanan, Y. P. Ragini, P. S. Kumar, P. Thamarai, and G. Rangasamy, *A critical review on the removal of toxic pollutants from contaminated water using magnetic hybrids*, Environmental Science and Pollution Research, 30 (2023), pp. 105099–105118.

[61] P. Saxena and S. Sonwani, *Criteria Air Pollutants and their Impact on Environmental Health*, Springer Singapore, 2019.

[62] J. H. Seinfeld, *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, Wiley Interscience Publication, John Wiley & Sons Incorporated, Newark, 1st ed. ed., 2016. Description based on publisher supplied metadata and other sources.

[63] I. M. Sobol', *Sensitivity analysis for non-linear mathematical models*, Mathematical Modeling & Computational Experiment, (1993).

[64] ——, *Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates*, Mathematics and Computers in Simulation, 55 (2001), pp. 271–280.

[65] J. T. Spickett, H. Brown, and K. Rumchev, *Climate change and air quality: The potential impact on health*, Asia Pacific Journal of Public Health, 23 (2011), pp. 37S–45S.

[66] J. M. Stockie, *The mathematics of atmospheric dispersion modeling*, SIAM Review, 53 (2011), pp. 349–372.

[67] W. Tian, *Towards advanced uncertainty and sensitivity analysis of building energy performance using machine learning techniques*, Journal of Building Performance Simulation, 17 (2024), pp. 655–662.

[68] V. Todorov, I. Dimov, T. Ostromsky, S. Apostolov, R. Georgieva, Y. Dimitrov, and Z. Zlatev, *Advanced stochastic approaches for Sobol' sensitivity indices evaluation*, Neural Computing and Applications, 33 (2020), pp. 1999–2014.

[69] H. K. Versteeg, *An introduction to computational fluid dynamics*, Pearson Prentice Hall, Harlow, 2nd ed. ed., 2011.

[70] C. Wang, Y. Zhu, T. Zang, H. Liu, and J. Yu, *Modeling inter-station relationships with attentive temporal graph convolutional network for air quality prediction*, in Proceedings of the 14th ACM International Conference on Web Search and Data Mining, WSDM '21, ACM, Mar. 2021, pp. 616–634.

[71] J. J. West, A. Cohen, F. Dentener, B. Brunekreef, T. Zhu, B. Armstrong, M. L. Bell, M. Brauer, G. Carmichael, D. L. Costa, D. W. Dockery, M. Kleeman, M. Krzyzanowski, N. Künzli, C. Liousse, S.-C. C. Lung, R. V. Martin, U. Pöschl, C. A. Pope, J. M. Roberts, A. G. Russell, and C. Wiedinmyer, *"What We Breathe Impacts Our Health: Improving Understanding of the Link between Air Pollution and Health"*, Environmental Science & Technology, 50 (2016), pp. 4895–4904.

[72] M. J. Wolf, D. C. Esty, H. Kim, M. L. Bell, S. Brigham, Q. Nortonsmith, S. Zaharieva, Z. A. Wendling, A. de Sherbinin, and J. W. Emerson, *New insights*

for tracking global and local trends in exposure to air pollutants, Environmental Science & Technology, 56 (2022), pp. 3984–3996.

[73] I. C. Yadav and N. L. Devi, *Biomass Burning, Regional Air Quality, and Climate Change*, Elsevier, 2019, pp. 386–391.

[74] J. Zhang, J. Yin, and R. Wang, *Basic framework and main methods of uncertainty quantification*, Mathematical Problems in Engineering, 2020 (2020), pp. 1–18.