Towards Dense Air Quality Monitoring: Time-Dependent Statistical Gas Distribution Modelling and Sensor Planning
To my parents

for their endless love, support, and encouragement.
Towards Dense Air Quality Monitoring:
Time-Dependent Statistical Gas Distribution
Modelling and Sensor Planning
Abstract

This thesis addresses the problem of gas distribution modelling for gas monitoring and gas detection. The presented research is particularly focused on the methods that are suitable for uncontrolled environments. In such environments, gas source locations and the physical properties of the environment, such as humidity and temperature may be unknown or only sparse noisy local measurements are available. Example applications include air pollution monitoring, leakage detection, and search and rescue operations.

This thesis addresses how to efficiently obtain and compute predictive models that accurately represent spatio-temporal gas distribution.

Most statistical gas distribution modelling methods assume that gas dispersion can be modelled as a time-constant random process. While this assumption may hold in some situations, it is necessary to model variations over time in order to enable applications of gas distribution modelling for a wider range of realistic scenarios.

This thesis proposes two time-dependent gas distribution modelling methods. In the first method, a temporal (sub-)sampling strategy is introduced. In the second method, a time-dependent gas distribution modelling approach is presented, which introduces a recency weight that relates measurement to prediction time. These contributions are presented and evaluated as an extension of a previously proposed method called Kernel DM+V using several simulation and real-world experiments. The results of comparing the proposed time-dependent gas distribution modelling approaches to the time-independent version Kernel DM+V indicate a consistent improvement in the prediction of unseen measurements, particularly in dynamic scenarios under the condition that there is a sufficient spatial coverage. Dynamic scenarios are often defined as environments where strong fluctuations and gas plume development are present.

For mobile robot olfaction, we are interested in sampling strategies that provide accurate gas distribution models given a small number of samples in a limited time span. Correspondingly, this thesis addresses the problem of selecting the most informative locations to acquire the next samples.
As a further contribution, this thesis proposes a novel adaptive sensor planning method. This method is based on a modified artificial potential field, which selects the next sampling location based on the currently predicted gas distribution and the spatial distribution of previously collected samples. In particular, three objectives are used that direct the sensor towards areas of (1) high predictive mean and (2) high predictive variance, while (3) maximising the coverage area. The relative weight of these objectives corresponds to a trade-off between exploration and exploitation in the sampling strategy. This thesis discusses the weights or importance factors and evaluates the performance of the proposed sampling strategy. The results of the simulation experiments indicate an improved quality of the gas distribution models when using the proposed sensor planning method compared to commonly used methods, such as random sampling and sampling along a predefined sweeping trajectory. In this thesis, we show that applying a locality constraint on the proposed sampling method decreases the travelling distance, which makes the proposed sensor planning approach suitable for real-world applications where limited resources and time are available. As a real-world use-case, we applied the proposed sensor planning approach on a micro-drone in outdoor experiments.

Finally, this thesis discusses the potential of using gas distribution modelling and sensor planning in large-scale outdoor real-world applications. We integrated the proposed methods in a framework for decision-making in hazardous incidents where gas leakage is involved and applied the gas distribution modelling in two real-world use-cases. Our investigation indicates that the proposed sensor planning and gas distribution modelling approaches can be used to inform experts both about the gas plume and the distribution of gas in order to improve the assessment of an incident.

Keywords: mobile robot olfaction; time-dependent gas distribution modelling; temporal sub-sampling; sensor planning; artificial potential field; gas monitoring.
Acknowledgements

Writing the final pages of the dissertation gives an opportunity to look back and reflect on the PhD years. In thinking about all moments of research, demo preparations, excitement, life crises, deadline nights, and wrapping up the dissertation, reaching to this point would not have been possible without the support and presence of many people through these years.

First, I would like to express my gratitude to my supervisor Prof. Achim Lilienthal for giving me the opportunity to join the Mobile Robot Olfaction Lab (MRO) at AASS Research Centre as a PhD student. The extent of your knowledge and expertise and your thorough feedback have been invaluable throughout my PhD study years. I would also like to thank my co-advisor Prof. Amy Loutfi for valuable discussions and feedback, support and encouragement, and more importantly, believing in me.

I would like to thank my opponent and committee for accepting to review this thesis. This thesis was partially funded by the EU FP7 project, Diadem. I would like to acknowledge my colleagues from the industrial and academic partners in this project. This project was a unique experience and gave me the opportunity to work on a real-world application. I would like to thank my colleagues at AASS (in alphabetic order): Han, Matteo, Sepideh, and Victor, my colleagues at the MRO lab, and Patrick (research visitor from BAM at the time) for their great collaboration on research projects and great scientific discussions.

I would like to thank to my friends who contributed to reviewing my work in different stages.

I would like to thank PhD students past and present, especially Marco, Marios, Marcello, Matteo, Karol, Krzysztof, Robert, and Todor for making the labs atmosphere friendly and warm and for great conversation during and after work hours. Thank you to the colleagues in the lab who offered their help in meeting a deadline by sharing computational resources. I would also like to thank senior researchers past and present in the lab, especially Federico, Lars, Mathias, and Mitko for great discussions over coffee breaks.

I am grateful to my amazing "Örebro" friends (in alphabetic order) Athanasia, Francesca, Hadi, Houssam, Iran, Marjan, Sepideh, Stella, and Yannis. Most
of the highlights of these years would not have happened without you. Special thanks go to Sepideh for not only being a great colleague but also for being a great friend and flatmate in all these days. Thank you for the night shifts close to deadlines, great discussions, and for being my fellow adventurer.

Working at Meltwater and Spotify, I have gotten to know many amazing people. I have learned from them how to utilize what I have learned at university to solve real-world challenges. Thank you (in alphabetic order) Bhaskar, Boxun, Elaine, Fredrik, Giuliano, Magnus, and Thuy, not just for interesting technical conversations, hours of whiteboard discussions, and building new solutions together, but also for patiently listening to my thesis stories and inspiring me along the way.

I am thankful (in alphabetic order) to Babak, Gabor, Galina, and Oliver for inspiration and encouragement. Thank you for your invaluable advice along the way.

I would like to thank my lifelong friends who, despite the distance, have always supported me; whom I could call wherever and whenever to share my thoughts, feelings, or events in my life (in alphabetic order) Bahareh, Marzieh, Shohreh, Tara, and Zahra.

I am grateful to my wonderful family: my parents, Peyman, Zoya, Keyvan, Mandana, Rojan, and Kian. Thank you for being by my side along the way despite the distance, for your encouragement, inspiration, and for your support.

This dissertation would not have been possible without you, thank you!
3.2.1 Simulated 2D Gas Dispersal Experiments .......................... 36
3.2.2 Simulated 3D Gas Dispersal Experiments .......................... 37

4 GDM Approaches With Predictive Variance ......................... 41
4.1 Kernel DM+V ................................................................... 42
  4.1.1 Extensions of Kernel DM+V ...................................... 44
4.2 Gaussian Processes ....................................................... 46
4.3 Gaussian Process Mixture Model ..................................... 48
  4.3.1 Initialisation of the Mixture Components ...................... 49
  4.3.2 Iterative Learning via Expectation-Maximisation Method .... 49
  4.3.3 Learning the Gating Function for Unseen Test Points ....... 50
4.4 Meta-parameter Selection and Evaluation Method ................. 52
  4.4.1 Evaluation Measure ................................................ 52
  4.4.2 Learning of Meta-Parameters .................................... 52
4.5 Experimental Comparison .............................................. 53
  4.5.1 Discussion of Kernel DM+V and the GPM Approach ........... 53
4.6 Discussion and Conclusion ............................................. 54

5 Time-dependent GDM ..................................................... 57
5.1 Problem Statement ...................................................... 58
5.2 Related Work ........................................................... 59
5.3 Meta-parameter Selection and Evaluation Method .................. 60
5.4 Temporal Sub-sampling ................................................ 61
  5.4.1 Evaluation and Results ............................................ 61
5.5 TD Kernel DM+V ......................................................... 63
  5.5.1 Experiments and Results, Basic Scenarios ....................... 64
  5.5.2 Experiments and Results, More Complex Scenarios ............ 67
  5.5.3 Recency Function .................................................. 68
5.6 TD Kernel DM+V versus Temporal Sub-sampling .................... 71
5.7 Model Selection and Meta-parameters ................................ 72
5.8 Experiments with Fully Developed Plumes ......................... 77
5.9 Dependence on Target Time .......................................... 78
5.10 Summary, Conclusions and Future Work ............................ 80

6 Multi Objective Sensor Planning ....................................... 83
6.1 Problem Statement ...................................................... 84
6.2 Related Work ........................................................... 85
  6.2.1 Sensor Planning Methods in Environmental Monitoring .......... 86
  6.2.2 Utility Functions Used in Sensor Planning Methods ............. 87
  6.2.3 Artificial Potential Field based Path Planning .................. 88
6.3 Method ........................................................................ 89
  6.3.1 Artificial Potential Field based Sensor Planning ................. 89
6.4 Evaluation Metrics ..................................................... 91
  6.4.1 Distribution Similarity ............................................. 92
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.4.2</td>
<td>Coverage</td>
<td>92</td>
</tr>
<tr>
<td>6.4.3</td>
<td>Plume Coverage</td>
<td>92</td>
</tr>
<tr>
<td>6.4.4</td>
<td>Travelling Distance</td>
<td>93</td>
</tr>
<tr>
<td>6.4.5</td>
<td>Distance to Gas Source Location</td>
<td>93</td>
</tr>
<tr>
<td>6.5</td>
<td>Experiments</td>
<td>93</td>
</tr>
<tr>
<td>6.5.1</td>
<td>Simulation</td>
<td>93</td>
</tr>
<tr>
<td>6.5.2</td>
<td>Real-world Experiments</td>
<td>105</td>
</tr>
<tr>
<td>6.6</td>
<td>Discussion and Conclusion</td>
<td>106</td>
</tr>
<tr>
<td>7</td>
<td>A Large-scale Pollution Monitoring Application</td>
<td>109</td>
</tr>
<tr>
<td>7.1</td>
<td>Diadem Project</td>
<td>110</td>
</tr>
<tr>
<td>7.1.1</td>
<td>Gas Monitoring</td>
<td>112</td>
</tr>
<tr>
<td>7.1.2</td>
<td>Distributed Gas Detection and Gas Source Localisation</td>
<td>112</td>
</tr>
<tr>
<td>7.1.3</td>
<td>ARGOS</td>
<td>113</td>
</tr>
<tr>
<td>7.2</td>
<td>GDM Module: Comprehensive Overview Tool</td>
<td>113</td>
</tr>
<tr>
<td>7.2.1</td>
<td>DCMR and Diadem Use-case</td>
<td>114</td>
</tr>
<tr>
<td>7.2.2</td>
<td>GDM in Combination with Other Solutions</td>
<td>115</td>
</tr>
<tr>
<td>7.2.3</td>
<td>GDM with Uncalibrated Data</td>
<td>115</td>
</tr>
<tr>
<td>7.2.4</td>
<td>Applications and Limitations of Existing GDM Methods</td>
<td>115</td>
</tr>
<tr>
<td>7.3</td>
<td>Sensor Planning Module</td>
<td>118</td>
</tr>
<tr>
<td>7.4</td>
<td>Gas Monitoring Component in Diadem</td>
<td>119</td>
</tr>
<tr>
<td>7.4.1</td>
<td>The GDM Module</td>
<td>119</td>
</tr>
<tr>
<td>7.4.2</td>
<td>The Sensor Planning Module</td>
<td>120</td>
</tr>
<tr>
<td>7.4.3</td>
<td>Integration in the Diadem Framework</td>
<td>121</td>
</tr>
<tr>
<td>7.4.4</td>
<td>Integration with ARGOS</td>
<td>122</td>
</tr>
<tr>
<td>7.4.5</td>
<td>Evaluation</td>
<td>122</td>
</tr>
<tr>
<td>7.5</td>
<td>Discussion and Conclusion</td>
<td>127</td>
</tr>
<tr>
<td>7.5.1</td>
<td>Diadem</td>
<td>127</td>
</tr>
<tr>
<td>7.5.2</td>
<td>GDM and Sensor Planning in Diadem</td>
<td>129</td>
</tr>
<tr>
<td>8</td>
<td>Conclusion</td>
<td>131</td>
</tr>
<tr>
<td>8.1</td>
<td>Contributions</td>
<td>131</td>
</tr>
<tr>
<td>8.1.1</td>
<td>Gas Distribution Modelling</td>
<td>131</td>
</tr>
<tr>
<td>8.1.2</td>
<td>Sensor Planning</td>
<td>132</td>
</tr>
<tr>
<td>8.1.3</td>
<td>Real-world Application</td>
<td>133</td>
</tr>
<tr>
<td>8.2</td>
<td>Limitations</td>
<td>133</td>
</tr>
<tr>
<td>8.2.1</td>
<td>Gas Distribution Modelling</td>
<td>133</td>
</tr>
<tr>
<td>8.2.2</td>
<td>Sensor Planning</td>
<td>134</td>
</tr>
<tr>
<td>8.2.3</td>
<td>Real-world Application</td>
<td>134</td>
</tr>
<tr>
<td>8.3</td>
<td>Ethics</td>
<td>134</td>
</tr>
<tr>
<td>8.4</td>
<td>Future Work</td>
<td>135</td>
</tr>
<tr>
<td>8.4.1</td>
<td>Gas Distribution Modelling</td>
<td>135</td>
</tr>
<tr>
<td>8.4.2</td>
<td>Sensor Planning</td>
<td>136</td>
</tr>
<tr>
<td>8.4.3</td>
<td>Real-world Application</td>
<td>137</td>
</tr>
</tbody>
</table>

References 139
# List of Figures

1.1 Pollution Monitoring. ............................................. 2

2.1 Visualisation of GDM problem in 1D. ......................... 12
2.2 Simple solutions to GDM problem in 1D. .................... 13
2.3 A simple GDM which averages measurements of equidistant sensors. ............................................. 14
2.4 A simple GDM using peak measurements of equidistant sensors. ............................................. 14
2.5 GDM using interpolation on a mobile sensor’s collected samples. ............................................. 15
2.6 Discretisation of a Gaussian kernel onto a grid. ............ 18
2.7 Impact of kernel width and cell size in Kernel DM – larger cell size. ............................................. 19
2.8 Impact of kernel width and cell size in Kernel DM – smaller cell size. ............................................. 20
2.9 Modelling the Spatial Distribution of Gas Detection Events. ............................................. 21

3.1 Wind tunnel experiment: setup. .................................. 25
3.2 Wind tunnel experiment: plume at the presence of laminar flow and obstacle. ............................................. 26
3.3 Small sensor network in controlled indoor experiment. .. 28
3.4 Terrain robot used for data collection in uncontrolled indoor and outdoor experiments. ......................... 29
3.5 Uncontrolled indoor "3-Room" experiment. .................. 30
3.6 Uncontrolled indoor "Corridor" experiment. .............. 31
3.7 Uncontrolled "Outdoor" experiment. ......................... 32
3.8 Outdoor experiment with micro-drone: Robotics Platform.. 32
3.9 Outdoor experiment at AASS using micro-drone: experiment setup. ............................................. 33
3.10 Outdoor experiment using gas sensor network in the first experiment: Diadem project. ......................... 34
3.11 Outdoor experiment using gas sensor network in the second experiment: Diadem project. ......................... 35
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.12</td>
<td>Simulation of gas dispersion in a wind tunnel with no obstacle.</td>
</tr>
<tr>
<td>3.13</td>
<td>Simulation of gas dispersion in a wind tunnel at the presence of an obstacle.</td>
</tr>
<tr>
<td>3.14</td>
<td>A 2D snapshot from the 3D ROS gas dispersion simulation.</td>
</tr>
<tr>
<td>4.1</td>
<td>GDM confidence map.</td>
</tr>
<tr>
<td>4.2</td>
<td>Visualisation of the kernel in Kernel DM+V at the presence of wind and absence of wind.</td>
</tr>
<tr>
<td>4.3</td>
<td>Visualisation of GDM with 3D Kernel DM+V.</td>
</tr>
<tr>
<td>4.4</td>
<td>GP initialisation.</td>
</tr>
<tr>
<td>4.5</td>
<td>Components in different iterations of the learning using EM algorithm.</td>
</tr>
<tr>
<td>4.6</td>
<td>Learned gate function and the resulting distribution of the GPM.</td>
</tr>
<tr>
<td>5.1</td>
<td>Locations at which samples were collected in the SmallNet and Corridor Experiments.</td>
</tr>
<tr>
<td>5.2</td>
<td>Predictive mean maps created for the &quot;Corridor&quot; experiments.</td>
</tr>
<tr>
<td>5.3</td>
<td>Predictive mean maps created for the &quot;SmallNet&quot; experiments.</td>
</tr>
<tr>
<td>5.4</td>
<td>Predictive mean maps created for the &quot;Sim-No-Obstacle&quot; experiments.</td>
</tr>
<tr>
<td>5.5</td>
<td>Predictive mean maps created for the &quot;Sim-With-Obstacle&quot; experiments.</td>
</tr>
<tr>
<td>5.6</td>
<td>Predictive mean and predictive variance maps created for experiments with the ROS 3D gas dispersial simulation engine – cs1.</td>
</tr>
<tr>
<td>5.7</td>
<td>Predictive mean and predictive variance maps created for experiments with the ROS 3D gas dispersial simulation engine – cs2.</td>
</tr>
<tr>
<td>5.8</td>
<td>Predictive mean and predictive variance maps created for experiments with the ROS 3D gas dispersial simulation engine – cs3.</td>
</tr>
<tr>
<td>5.9</td>
<td>Predictive mean and predictive variance maps created for experiments with the ROS 3D gas dispersial simulation engine – cs4.</td>
</tr>
<tr>
<td>5.10</td>
<td>NLPD comparison of TD Kernel DM+V using different recency functions.</td>
</tr>
<tr>
<td>5.11</td>
<td>NLPD color maps for c − σ, c − β, and σ − β in the &quot;Corridor&quot; and &quot;SmallNet&quot; experiments.</td>
</tr>
<tr>
<td>5.12</td>
<td>NLPD color maps for c − σ, c − β, and σ − β in the basic simulation experiments.</td>
</tr>
<tr>
<td>5.13</td>
<td>NLPD color maps for c − σ, c − β, and σ − β in the more complex simulation experiments.</td>
</tr>
<tr>
<td>5.14</td>
<td>NLPD comparison of TD Kernel DM+V and Kernel DM+V in experiments with fully developed plumes.</td>
</tr>
</tbody>
</table>
5.15 Performance of TD Kernel DM+V when test sets and target times vary over time. ........................................ 80

6.1 The predictive mean map created using ground truth data in the "Sim-ROS" c4 experiment. ................................. 94
6.2 Trajectory of sweeping sampling strategies in simulated experiments. ............................................................. 95
6.3 Impact of parameter selection on the distribution similarity when collecting 1200 samples. ............................ 96
6.4 Impact of parameter selection on the distribution similarity for different sample sizes. ................................. 98
6.5 Impact of parameter selection on the total travelling distance when collecting 1200 samples. .......................... 99
6.6 Impact of parameter selection on the coverage when collecting 1200 samples. .............................................. 99
6.7 Impact of parameter selection on the plume coverage when collecting 1200 samples. ................................. 100
6.8 Impact of parameter selection on the samples distance from gas source location in different simulated experiments. 100
6.9 KL-distance comparison of predictive mean maps created using APFSP sampling strategy using different GDM methods. 104
6.10 Sampling trajectory of the sensor planning algorithm in "QC-AASS" experiment. ............................................ 105
6.11 Sampling trajectory of the sensor planning algorithm in "QC-AASS" experiment. ............................................ 106

7.1 Schema of Diadem process. .................................................. 112
7.2 Gas Monitoring in the Diadem framework. ............................... 120
7.3 Integration of GDM in DPIF. .................................................. 121
7.4 An example of GDM output visualised in ARGOS. ....................... 122
7.5 Integration of GDM with ARGOS. ............................................ 123
7.6 Relative concentration map overlaid on the Google Earth map of the investigated area in the "Diadem-1" experiment. 124
7.7 Predictive mean map overlaid on Google Earth map of the area using measurements in scenario A. ....................... 125
7.8 The consecutive predictive mean and variance maps and the corresponding difference maps of each two consecutive maps using measurements in scenario A. .......................... 126
7.9 Predictive mean map overlaid on Google Earth map of the area using measurements in scenario B. ....................... 127
7.10 The consecutive predictive mean and variance maps and the corresponding difference maps of each two consecutive maps using measurements in scenario B. .......................... 128
List of Tables

3.1 List of real-world datasets and their specifications. 24
3.2 Sensor specifications of sensor arrays used in controlled indoor experiments. 27
3.3 Sensor specifications in uncontrolled indoor experiments. 29
4.1 Comparison of standard GP, GPM, and Kernel DM+V. 53
5.1 NLPD comparison of models created with Kernel DM+V using different temporal (sub-)sampling. 63
5.2 Comparison of basic Kernel DM+V and TD Kernel DM+V in terms of NLPD and RMSE. 65
5.3 Performance comparison of basic Kernel DM+V and TD Kernel DM+V in ROS Simulation experiments. 68
5.4 NLPD comparison of TD Kernel DM+V and temporal (sub-)sampling. 72
5.5 NLPD comparison of TD Kernel DM+V and Kernel DM+V with constant cell size in real-world and basic simulation experiments. 76
5.6 NLPD comparison of TD Kernel DM+V and Kernel DM+V with constant cell size in the more complex simulation experiments. 77
5.7 Performance of TD Kernel DM+V gas distribution models for test sets sampled from the basic simulation at increasing times in the future. 79
6.1 Performance comparison of sensor planning methods. 101
6.2 Performance comparison of APFSP and SPPAM-A. 102
6.3 Performance comparison of APFSP sampling strategies using different gas distribution models. 104
Chapter 1

Introduction

1.1 Motivation

Environmental concerns, especially in urban areas with their critical impact on the quality of human life, have become increasingly important and are therefore of key interests to different scientific communities. To assess environmental quality and make effective decisions, a thorough knowledge of the environment is required. This knowledge is acquired using sensors.

In environmental monitoring scenarios, such as air or water pollution monitoring where local measurements are required, it is not economically efficient to collect samples continuously over a large area. In this domain, it is a challenge to plan where to acquire additional samples and how to build a model from sparse samples.

Mobile robotics and artificial olfaction can help to address pressing environmental problems that involve gas emission. Typical tasks include leakage detection, air pollution monitoring, and search and rescue operations. Sensor networks and mobile robots equipped with gas sensors can be used for example in air pollution monitoring (Figure 1.1).

In air pollution monitoring, we are interested in data-driven statistical gas distribution models that can provide comprehensive information about a large number of gas measurements, highlighting, for example, areas of unusual gas accumulation, assisting in locating gas sources and planning where future measurements have to be acquired [1].

Several publications have addressed the problem of sensor placement and data inference in environmental monitoring in recent years. Different domains that address this issue include sensor network, experimental design, spatial statistics, machine learning, decision theory, and operation research [2]. Meanwhile, research in machine learning and robotics have seen significant developments to take on problems with large quantities of data, and to allow them to expand their potential to address real-world problems [2]. The great de-
Figure 1.1: Top left: In pollution monitoring and management of incidents caused by different pollution sources, gas sensors are used. Gas sensors collect information from the environment to build a model of the observed phenomena. Top right: These sensors can be placed as stationary sensors or mounted on mobile terrain robots, quad-copters, or manually carried by a human operator. Bottom: Data-driven statistical gas distribution modelling approaches can build gas distribution maps using these sensor measurements. The red areas indicate relative high accumulation of gas and the blue colour indicate areas with lower accumulation of gas.

Recent advances in mobile robot olfaction enabled prototype systems to explore practical applications in real-world environments. Examples include Reggente et al. [3], where a mobile sensor collected gas measurements in urban public places while collecting garbage bags, and a dedicated mobile robot to monitor landfill and biogas production sites (Hernandez Bennetts et al. [4]).
1.2. CHALLENGES

As in the latter example, gas sampling with an autonomous robot is especially beneficial in scenarios where humans could be directly exposed to harmful gases [5, 6]. The practical applications mentioned highlight as a common requirement the ability to derive a truthful gas distribution model, which allows highlighting areas with unusual gas accumulation, estimation of the location of gas sources, or detection of other anomalies in the “gas space” [5].

One particular motivation for research presented in this dissertation came from the EC project Diadem (Distributed Information Acquisition and Decision-making for Environmental Management [5]). The particular use-case in Diadem was the emission of hazardous chemical gases due to chemical incidents. A large target area of approximately 56 km$^2$ at the Rotterdam harbour in the Netherlands was considered, a densely inhabited industrial area where several refineries and oil factories are located. In this area, gas measurements were continuously collected using a very sparse stationary sensor network. One goal of the Diadem project was to investigate gas distribution modelling as a means for environmental monitoring and to plan where to collect further measurements (which could then be sampled with mobile sensors carried by field operators). While the sensor network proved to be too sparse and a solid ground truth evaluation was not possible at such a large scale, gas distribution mapping nevertheless turned out to be useful for the experts who monitor the gas sensors in the Rotterdam area in a control room. One particularly relevant conclusion in this project was that when looking at a sequence of gas distribution maps, it was easier for them to identify areas with high gas accumulation, potential gas source locations and to predict further development of the gas distribution than looking at a spatially unconnected visualisation of the measurement data (which is what they used at the time). This observation further encouraged us to focus on temporal and spatial sampling methods that allow creating better gas distribution models [7, 8].

1.2 Challenges

Statistical gas distribution modelling (GDM) aims at deriving a truthful representation of the environment from a set of spatially and temporally sparse measurements [9]. Modelling spatial distribution of gas is very challenging mainly because of the chaotic nature of gas dispersal. The complex interaction of the gas with its surroundings is dominated by three physical effects. First, on a long time scale, diffusion mixes the gas with the surrounding atmosphere to achieve a homogeneous mixture of both. Second, turbulent air flow fragments the gas emanating from a source into intermittent patches of high concentration with steep gradients at their edges [10]. Third, advective flow moves these patches. Due to the effects of turbulence and advective flow, it is possible to observe high concentrations in locations distant from the source location. In controlled environments with laminar flow, patches of gas particles are mainly distributed along the wind direction in a plume. In environments
where wind and temperature are not controlled, advective flow and turbulence have a stronger effect on gas dispersion which makes the prediction of movement of gas patches more complex.

Besides the physics of gas dispersal, limitations of gas sensors also make gas distribution modelling difficult. The commercially available and inexpensive sensors, which are widely used in large-scale pollution monitoring applications, provide information about a small spatial region close to sensor’s surface only since the measurements require direct interaction between the sensor surface and the analyte molecules [11]. Therefore, instantaneous gas measurements over a large field would require a dense grid of sensors which is usually not a viable solution due to high cost and lack of flexibility [1].

### 1.3 Problem Statement

The research presented in this dissertation mainly addresses the problem of gas distribution modelling (GDM) with the purpose of air quality monitoring, and gas detection in uncontrolled environments where gas source locations and physical properties of the environment such as humidity and temperature, are unknown and sparse noisy local measurements are available. In this line of research, temporal and spatial sampling methods are explored to create accurate gas distribution models.

The focus of this research is on gas distribution modelling in various environmental conditions from controlled indoor to uncontrolled outdoor environments. In addition, this dissertation provides sensor planning approaches which enhance the quality of gas distribution given a small number of samples and considers the real-world gas distribution modelling challenges such as resource constraints and sensing limitation of gas sensors.

To evaluate gas distribution models and sensor planning methods presented in this dissertation, several experiments are performed both in simulation and real-world environments with a mobile robot and stationary sensor networks. These experiments are carried out using existing datasets contributed by collaborators from Applied Autonomous Sensor Systems (AASS) research centre, and other research centres are used. Detailed information about these datasets is presented in Chapter 3.

The primary objectives of this research are

- to obtain a truthful representation of gas distribution having sparse measurements with limited knowledge available about the physical properties of the environment in an uncontrolled environment, and

- to plan where to acquire the next sample to gain more information about gas distribution, considering resource constraints.
1.4 Contributions

This dissertation provides solutions for the two aforementioned objectives. In particular, this research makes the following contributions:

- Analysis and comparison of state-of-the-art GDM methods on various real-world datasets [1].

- A novel time-dependent statistical gas distribution modelling method to derive a more accurate estimation of gas distribution is proposed. This method introduces time-dependency and a relation to a time-scale in generating the gas distribution model either by sub-sampling or by introducing a recency weight that relates measurement and prediction time. This contribution is an extension of a previously existing method called Kernel DM+V [9, 1].

- The proposed time-dependent GDM method has been carefully evaluated in experiments performed in (1) two real environments: a stationary sensor networks in a small controlled environment and a mobile sensors collecting data in a corridor, and (2) several simulated experiments: in the presence of predominantly laminar flow and turbulence caused by obstacles.

- An adaptive multi-criteria sensor planning method is proposed. This method is based on a modified artificial potential field which selects the next sampling point based on the spatial information of collected samples and the information from GDM. This sensor planning approach is suitable for our domain where we take into account resource constraints, maximum coverage, and the properties of the created gas distribution model.

- The proposed sensor planning approach has been validated in a real-world online sampling using a micro-drone in an uncontrolled outdoor environment to detect a carbon monoxide smoke gas source.

- Extensive evaluation of the proposed sensor planning in gas distribution modelling is carried out in a simulated environment using various performance measures.

- The proposed gas distribution modelling and sensor planning approaches were developed and integrated in the Diadem EC project for decision-making in hazardous incidents where gas leakage is involved. These modules were applied in a large-scale outdoor scenario in the Rotterdam port using a stationary sensor network. In this case study, there was no ground truth available; however, qualitative analysis on usability was carried out.
Although the focus of this dissertation is to provide more efficient solutions for gas distribution modelling, the proposed solutions can be used in other domains such as gamma radiation dispersion monitoring, gas source detection [12], and oceanographic studies too [13].

1.5 Outline

This dissertation is organised as follows:

Chapter 2 explains the problem of gas distribution modelling and presents a review of state-of-the-art studies.

Chapter 3 describes data sets and various experimental setups used to study gas distribution models and sensor planning methods. The focus of this dissertation is not on data collection and creating experimental setups; therefore, available data sets are used. The main part of the text in this chapter is taken from the reference publications related to the corresponding data sets.

Chapter 4 presents a review on a particular group of statistical distribution modelling approaches which estimate variance in addition to mean when modelling gas distribution. This chapter provides an extensive comparison of these approaches. The content of this chapter is based on my publication [1].

Chapter 5 introduces methods to build time-dependent gas distribution models. Evaluation on simulated and real-world data are provided [14, 15, 7]. The content of this chapter is based on my publication [7].

Chapter 6 presents a new approach to address sensor planning in the study of gas dispersion. Evaluation and analysis on simulated and real-world data are provided. The developed algorithm is presented in joint publications with Patrick Neumann in [8] and [12], who applied this method in sample selection for a micro-drone. In addition to evaluation on simulation data performed by the author, evaluation results provided by Patrick Neumann within this research collaboration are also presented in this chapter to discuss better and explain performance of the proposed sensor planning method.

Chapter 7 gives an example of using gas distribution modelling in a real-world case study. This chapter summarises work carried out to use the developed methods in a real-world application within the EU-FP7 project Diadem [5].

Chapter 8 concludes the addressed gas dispersion study problem by summarising contributions to solve time-dependent gas distribution modelling and
sensor planning, discussing limitations of the presented methods, and outlining potential research directions for future work.

### 1.6 Publications

The work presented in this dissertation is partly published in a number of journal and conference papers. Note that some of the publications are the result of my collaboration with other researchers. The following list of publications presents a declaration of contribution where applicable.

- **S. Asadi, H. Fan, V. Hernandez Bennetts and A. Lilienthal, “Time-dependent gas distribution modelling,”** in Special Issue of the Robotics and Autonomous Systems (RAS) journal on the 7th European Conference on Mobile Robots (ECMR’15), vol. 69, pp. 157–170, Oct 2017 [7]. This article extends the work presented in [14] by comparing the performance of the two introduced time-dependent gas distribution modelling approaches, discussing the influence of meta-parameters on the quality of gas distribution models, investigating the quality of gas distribution models in different stages of plume development, analysing the impact of the choice of target time on the quality of gas distribution models, using additional evaluation measures, and presenting evaluations of gas distribution models in more realistic simulation experiments. In addition to the simulation datasets used in [14], in this paper we use a new 3D gas dispersion simulator in a larger simulated environment that contains physical obstacles in order to have a better analysis in simulation experiments. In this work, H. Fan and Dr. V. H. Bennetts contributed by building up a pipeline to experiment on different simulation setups and providing 3D simulation data sets. The content of this chapter covers the major part of Chapter 5.

- **S. Asadi and A. Lilienthal, “Approaches to time-dependent gas distribution modelling,”** in Mobile Robots (ECMR), 2015 European Conference on, pp. 1–6, Sept 2015 [14]. This paper presents two approaches that explicitly consider the measurement time in gas distribution modelling, either by sub-sampling according to a given time-scale or by introducing a recency weight that relates measurement and prediction time. I evaluated the performance of these two approaches using existing simulation and real-world data sets. The content of this paper is mostly presented in Chapter 5.

- **P. Neumann, S. Asadi, V. Hernandez Bennetts, A. J. Lilienthal, and M. Bartholmai, “Monitoring of CCS Areas using Micro Unmanned Aerial Vehicles (MUAVs),”** Energy Procedia, vol. 37, pp. 4182–4190, 2013 [16]. This paper presents results of using a micro-drone in greenhouse gas monitoring specifically CO₂. My contribution to this publication includes pro-
viding the adaptive sensor planning solution used in the paper for CO₂ monitoring. Moreover, I contributed to data analysis and evaluation.

- P. Neumann, S. Asadi, A. J. Lilienthal, M. Bartholmai, and J. H. Schiller, “Autonomous gas-sensitive microdrone: Wind vector estimation and gas distribution mapping,” Robotics Automation Magazine, IEEE, vol. 19, pp. 50–61, march 2012 [8]. This paper is the main publication of my collaborative research with Patrick Neumann, at that time a PhD student from Berlin visiting the AASS research centre. My main contributions are the theoretical formulation of sensor placement, performing simulation experiments, and optimisation of meta parameters. Moreover, I contributed in the evaluation of experiments. Patrick Neumanns contributions within this collaboration were designing the robotic platform, data collection with the micro-drone, sensor calibration, introducing a method to obtain wind measurement, and optimising the sensor planning algorithm for a single gas-sensitive micro-drone by adding locality constraints. The content of this publication is presented mostly in Chapter 6.

- P. Neumann, S. Asadi, J. H. Schiller, A. J. Lilienthal, and M. Bartholmai, “An artificial potential field based sampling strategy for a gas-sensitive micro-drone,” in Proceedings of the IROS 2011 Workshop on Robotics for Environmental Monitoring (WREM), (San Francisco, CA), pp. 34–38, 2011 [12]. This paper is a joint work with Patrick Neumann during his visit to AASS. In this paper, a sampling strategy for a micro-drone is proposed, and evaluation results in an outdoor environment are presented. I worked closely with Patrick Neumann on this publication developing the theoretical sensor planning method. Patrick Neumann developed the robotic platform and built the experimental setup. Furthermore, he adapted the proposed sensor planning approach to enhance the micro-drone path planning by defining a locality constraint. The experimental analysis was a collaborative effort. The content of this paper is discussed and explained in detail in Chapter 6.

- S. Asadi, S. Pashami, A. Loutfi, and A. J. Lilienthal, “TD Kernel DM+V: Time-dependent statistical gas distribution modelling on simulated measurements,” in AIP Conference Proceedings Volume 1362: Olfaction and Electronic Nose - Proceedings of the 14th International Symposium on Olfaction and Electronic Nose (ISOEN), pp. 281–283, 2011 [15]. In this paper, a time-dependent model to estimate gas distribution is proposed. The proposed idea has been analysed and evaluated in a simulated wind tunnel with a gas source. The gas dispersion simulation engine was developed by Sepideh Pashami. In this work, Sepideh Pashami contributed by providing simulation data and building up a simulation pipeline for time-dependent gas distribution modelling. As the main contributor of
this paper, I introduced a time-dependent statistical model and carried out experiments and evaluation in a simulated environment. Detailed results are presented in Chapter 5.

- S. Asadi, M. Reggente, C. Stachniss, C. Plagemann, and A. J. Lilienthal, “Statistical gas distribution modelling using Kernel methods,” in Intelligent Systems for Machine Olfaction: Tools and Methodologies (E. L. Hines and M. S. Leeson, eds.), ch. 6, pp. 153–179, IGI Global, 2011. [1]. This publication is a book chapter which presents a survey of statistical gas distribution modelling methods. As the main contributor to this paper, I carried out a thorough qualitative and comparative analysis of state-of-the-art methods and explored potential lines of research based on these methods. As a result, a time-dependent extension of Kernel DM+V is proposed. The content of this chapter covers the major part of Chapters 2 and 4.

- S. Asadi, C. Badica, T. Comes, C. Conrado, V. Evers, F. Groen, S. Ilie, J. S. Jensen, A. Lilienthal, B. Milan, T. Neidhart, K. Nieuwenhuis, S. Pashami, G. Pavlin, J. Pehrsson, R. Pinchuk, M. Scafes, L. Schou-Jensen, F. Schultmann, and N. Wijngaards, “ICT solutions supporting collaborative information acquisition, situation assessment and decision making in contemporary environmental management problems: the Diadem approach,” in Proceedings of the International Conference on Innovations in Sharing Environmental Observation and Information (EnviroInfo) (P. S. E. W. Pillmann, S. Schade, Ed.), pp. 920–931, Shaker Verlag, 2011 [5]. This paper is the final paper of the FP7 EU project Diadem. In the context of this project, the AASS research centre researched gas distribution modelling and sensor planning where only sparse measurements are available in an outdoor environment. My responsibility in this project was to build these two modules and integrate them with other components in the developed ICT solution. My contribution in this publication is presented in the sections related to these two modules and their integration. The content of this paper is partially presented in Chapter 7.

- S. Pashami, S. Asadi and A. J. Lilienthal, “Integration of OpenFOAM flow simulation and filament-based gas propagation models for gas dispersion simulation,” in Proceedings of the Open Source CFD International Conference, 2010 [17]. Sepideh Pashami developed a simulation engine to model gas dispersion. This work was part of a collaboration in the AASS machine olfaction research group. As a partial result of this collaboration, Sepideh Pashami modelled gas dispersion in a wind tunnel in the presence of an obstacle and with no obstacle. Throughout this research, I used this simulated environment for time-dependent gas dispersion study and sensor planning. As my contribution to this publi-
cation, I carried out gas distribution modelling using simulation data and
developed a pipeline to acquire snapshot information as well as building
weight-recency models. The content of this paper is used in Chapters 3, 5
and 6.

  nce for statistical gas distribution modelling,” in AIP Conference Pro-
  ceedings Volume 1137: Olfaction and Electronic Nose - Proceedings of
  the 13th International Symposium on Olfaction and Electronic Nose
  (ISOEN), pp. 65–68, 2009 [18]. In this paper, we investigated poten-
  tial improvement in gas distribution modelling using predictive variance.
  We proposed a statistical measure based on predictive variance to eval-
  uate gas distribution models and to estimate meta-parameters of gas
  distribution models. As my main contribution, I explored utilising sen-
  sor planning and time-dependent gas distribution modelling using this
  measure, which is widely used in evaluations presented in this disserta-
  tion (Chapters 4, 5 and 6).
Chapter 2

Background

Chapter 1 explained the problems that are addressed in the research presented in this dissertation. In particular, gas distribution modelling was explained and the challenges to create accurate gas distribution models were discussed. This chapter provides a review of state-of-the-art studies in gas distribution modelling. Gas distribution modelling methods can be categorised as model-based and model-free. Model-based approaches infer the parameters of an analytical gas distribution model from the measurements. Examples of model-based approaches are Gaussian plume models, Gaussian puff models, Lagrangian particle models, and Computational Fluid Dynamics (CFD) [19, 20]. In principle, CFD models can be applied to solve the governing set of equations numerically. However, current CFD methods are computationally expensive and not suitable for realistic scenarios in which a high resolution is required and the model needs to be updated with new measurements in real time [1, 9]. In addition, approaches such as CFD models depend on accurate knowledge of the environment state (boundary conditions), while in real-world scenarios, usually these types of prior knowledge are not available.

Other model-based approaches such as Gaussian plume models rely on simplified assumptions such as certain plume shape. These models are applicable only when the assumptions hold and can provide only coarse grain information about gas distribution.

Model-free approaches, on the other hand, do not make strong assumptions about a particular functional form of the gas distribution such as gas source location or plume shape. The focus of the research in this dissertation is on a class of model-free approaches that treat gas sensor measurements as random variables and derive a statistical model of the observed gas dispersion from those measurements. Model-free gas distribution modelling is often interpreted as spatial regression problem [19]. This chapter presents an overview on the existing model-free statistical gas distribution modelling methods. The content of this chapter is mainly based on a book chapter that we published on statistical gas distribution modelling [1].
2.1 Problem Statement: Statistical Gas Distribution Modelling

Statistical gas distribution modelling (GDM) aims to estimate the gas distribution from a set of existing measurements collected in a target area. More formally, in the environment $A$, GDM aims at providing an accurate representation of the observed phenomenon to predict the observation $r_*$ at the unseen location $x_*$ from a set of observations $D$ as

$$
\hat{r} = p(r_*|x_*, (x_i, r_i) \in D, 1 \leq i \leq |D|),
$$

(2.1)

where the pair $(x_i, r_i)$ denotes the measurement collected at location $x_i$ with the measurement value of $r_i$.

Figure 2.1 visualises an example of the GDM problem in one dimension (1D). In this example, a set of 20 samples is collected from an observed phenomenon. There is no measurement available at location $x_*$. The actual value of the observed phenomenon at this location is $r_*$ indicated by a red cross in this figure. GDM provides an estimate of this unseen measurement. One simple solution is to estimate the unseen measurement with the measurement value of the nearest sample (see Figure 2.2(a)). This approach fails if there is a high variance in the measurement values of the neighbouring samples. Another solution is to interpolate the measurement value from neighbouring samples (see Figure 2.2(b)). Similarly, this approach has low accuracy when measurements are sparse. In this chapter, a review of existing GDM approaches in the state-of-the-art is presented.

Figure 2.1: In GDM, one is interested in estimation of a measurement at an unseen location $x_*$ where no measurement is recorded. The circles filled in blue indicate collected samples. The red cross represents the actual value of the observed phenomenon at $x_*$. 
2.2. RELATED WORK – GDM APPROACHES WITHOUT PREDICTIVE VARIANCE

Figure 2.2: Estimation of measurement at unseen location $x_*$ by (a) using the measurement value at the nearest sampling location and (b) interpolation on the neighbouring samples. The red cross indicates the actual value at $x_*$ and the square filled in red indicates the estimated value.

2.2 Related Work – GDM Approaches Without Predictive Variance

2.2.1 Simultaneous Measurements – Using a Dense Sensor Grid

Several GDM methods have been published in recent years. A straightforward solution to obtain a model of the time-averaged gas distribution is to use a dense grid of sensors. In [21], a gas distribution model is created from measurements collected by a grid of sensors in a wind tunnel. These measurements are averaged over a prolonged period of time and discretised to a grid that repre-
sents the topology of the sensor network. An example of a map created with this method is presented in Figure 2.3. A similar method is presented in [22], where maximum values of the measurement intervals are mapped instead of average concentrations (see Figure 2.4).

Figure 2.3: Gas distribution model created by averaging sensor over 5 minutes. Sample are collected at equidistant locations. This figure is reproduced from [21].

Figure 2.4: Gas distribution model created by using maximum values of measurements in the experiment’s time span. Sample are collected at equidistant locations. This figure is reproduced from [22].
A network of stationary sensors has the advantage of reducing the required time to create a gas distribution map, but it requires considerable effort in sensors calibration and maintenance [20]. An alternative to a network of stationary sensors is to use mobile sensors. In general, mobile sensors allow adaptive sampling. A single mobile sensor can avoid some calibration issues. Pyk et al. [23] create a gas distribution map by using a single sensor, which collects measurements consecutively instead of the parallel acquisition of measurements in a sensor network. In an experiment, Pyk et al. applied this method for gas distribution modelling in a wind tunnel. At each pre-specified sampling location, the sensor was exposed to the gas distribution for two minutes. At locations other than the measurement points, the map was interpolated using bi-cubic or triangle-based cubic filtering, depending on whether the measurement locations formed an equidistant grid or not. Figure 2.5 shows an example of estimated gas distribution in this experiment. The drawback of such interpolating methods is that there is no means of averaging instantaneous response fluctuations at sampling locations. This leads to increasingly jagged distribution maps (see Figure 2.5(b)).

Figure 2.5: Gas distribution model created by (a) bi-cubic interpolation and (b) triangular-based cubic filtering. These graphs are reproduced from [23].

### 2.2.3 Histogram-based GDM

In practice, it is beneficial to combine sensor networks with autonomous mobile sensors. In [24], a group of mobile robots equipped with conducting polymer sensors were used to create a histogram representation of the distribution of water vapour created by a hot water pan behind a fan. The histogram bins collect the number of odour hits received by all robots in the corresponding
area while they performed a random walk behaviour. Odour hits were counted whenever the response level of the gas sensors exceeded a predefined threshold. A potential problem with this method is that it uses only binary information from the measurements. In this way, useful information may be discarded. In addition, the proposed histogram method depends strongly on bin size and the predefined threshold. Another disadvantage of this approach is that it requires even coverage of the environment.

The Kernel extrapolation Distribution Mapping (Kernel DM) algorithm, proposed by Lilienthal and Duckett [25], can be regarded as a refinement of a histogram-based approach that does not rely on the assumption of an even coverage of the environment. Kernel DM is inspired by non-parametric estimation of density functions using Nadaraya-Watson estimator. In this approach, simultaneous measurements in a sensor network or mobile sensor samples can be used. This approach discretises the area into a grid and extrapolates gas distribution over this grid using a Gaussian kernel. Section 2.2.4 explains Kernel DM in detail.

2.2.4 Kernel DM

The Kernel DM method proposed by Lilienthal and Duckett [25] represents a gas distribution in the form of a grid map. This gas distribution grid map resembles occupancy grid maps in mobile robotics. However, there are important differences. Each cell in an occupancy grid map represents the likelihood of whether the area covered by the cell is occupied or not. Cells in the Kernel DM grid map represent an estimate of the distribution mean at the location of the particular cell. Occupancy grid maps are typically built based on input from a laser scanner or sonar sensors. With each measurement, these range sensors cover a wide area, and there is typically a substantial overlap between individual readings. The problem is different for grid maps based on gas sensor measurements because the sensor response represents interactions at the surface of the sensor, i.e. at a very small area. Therefore, a key idea of Kernel DM is how to extrapolate the sensor measurements beyond this small area by incorporating the decreasing information content that a single sensor measurement provides about the mean estimate at locations in a certain distance from the sensor surface.

Kernel DM discretises the available space into grid cells and computes an estimate of the distribution mean for each cell by using a symmetric Gaussian kernel N. The Gaussian kernel models the decreasing information content of a single sensor measurement depending on the distance of the measurement location from the respective cell centre. In other words, the Gaussian kernel represents a weight function, which indicates the likelihood that the measurement represents the average concentration at a given distance from the point of measurement (see Figure 2.6).
Assuming that $D = \{(x_1, r_1), (x_2, r_2), \ldots, (x_n, r_n)\}$ is the set of measurements $r_i$ collected at locations $x_i$ (1 ≤ $i$ ≤ n), the following steps are performed to predict the distribution mean at each location in the field. First, the displacement of the measurement location $x_i$ from the centre of the grid cell $k$ is computed as $\delta_{i}^{(k)} = |x^{(k)} - x_i|$. Using this distance, the value of the kernel for a given measurement is computed for each cell by

$$\omega_{i}^{(k)} = N(\delta_{i}^{(k)}, \sigma),$$  \hspace{1cm} (2.2)$$

where $\sigma$ is the kernel width. To enhance the performance, in practice, the computation of Equation 2.2 is restricted to cells within a certain cut-off radius $R_{co}$. In the next step, two temporary maps, $\Omega^{(k)}$ and $R^{(k)}$, are created that represent integrated weights (corresponding to the density of measurements) and integrated weighted readings as follows

$$\Omega^{(k)} = \sum_{i=1}^{n} \omega_{i}^{(k)}, \hspace{1cm} (2.3)$$

$$R^{(k)} = \sum_{i=1}^{n} \omega_{i}^{(k)}r_i. \hspace{1cm} (2.4)$$

Finally, the mean estimate is computed as

$$r^{(k)} = \frac{R^{(k)}}{\Omega^{(k)}} : \Omega^{(k)} \geq \Omega_{\text{min}}. \hspace{1cm} (2.5)$$

The threshold $\Omega_{\text{min}}$ ensures that estimates are based on a sufficient amount of measurements collected in the vicinity of the estimate location.

Figure 2.6 shows how a single measurement is convolved onto a $5 \times 5$ grid. On the left side, thirteen cells are found to have a distance of less than the cut-off radius (here $R_{co} = 3\sigma$) from the point of measurement. These cells are indicated on the right side of Figure 2.6 by a surrounding strong border. The weights in these cells are determined by evaluating the Gaussian function for the displacement values. These weights are represented in Figure 2.6 by shadings of grey. Darker shadings indicate higher weights, which correspond to a stronger contribution of the measurement value $r_i$ in the calculation of the average concentration value for a particular cell.

The Kernel DM algorithm depends on three parameters: the width $\sigma$ of the Gaussian function, the cut-off radius $R_{co}$, and the threshold $\Omega_{\text{min}}$. Only choosing an appropriate value of $\sigma$ is critical for the performance of the method. The kernel width should be wide enough to have sufficient extrapolation. On the other hand, it should be narrow enough to preserve the fine details of the mapped structure.

Figures 2.7 and 2.8 show the impact of cell size and kernel width on the estimation of predictive mean on an example of one-dimensional dataset. Applying kernels with very large width is almost like estimating predictive mean
in each cell as the averaging of all measurements (see Figure 2.7(a)). On the other hand, applying a very small kernel width can lead to over-fitting (see Figure 2.7(d). Cell size has impact on the resolution of our estimation. When the resolution gets lower (large cell sizes), the curve of the estimated mean becomes less smooth (see Figure 2.8).

Lilienthal and Duckett [25] applied the Kernel DM method to experiments in an indoor environment. Samples were collected with an array of sensors mounted on a mobile robot. In the experiment presented by Lilienthal and Duckett [25], perfect knowledge about the position of the sensors at the time of the measurement is assumed. To account also for the uncertainty about the sensor position, one can integrate gas distribution mapping into a Simultaneous Localization and Mapping (SLAM) framework as described in the work of Lilienthal et al. [26]. The experimental setup included a single target gas. The work of Bennetts et al. [27] extends the gas distribution mapping to consider multiple types of gas sources by including a gas discrimination component.

### 2.2.5 Kalman Filtering for GDM

Blanco et al. [28] presented a Bayesian approach to build gas distribution maps. This method applies a derivation of Kalman Filtering (KF) to create gas distribution model. Similar to Kernel DM, this method discretises the environment into cells and updates the gas distribution map by adding measurements sequentially. In each step, KF allows to update the value of neighbouring cells in a window around the given measurement.
2.2. RELATED WORK – GDM APPROACHES WITHOUT PREDICTIVE VARIANCE

Figure 2.7: Impact of kernel width $\sigma$ on the predictive mean using Kernel DM in a one-dimensional example dataset. In this figure, the red line indicates the predictive mean value created using Kernel DM. Cell size is $c=0.10$ m in all of the experiments. The kernel width is set to: (a) $\sigma=5.00$ m, (b) $\sigma=1.00$ m, (a) $\sigma=0.50$ m, and (d) $\sigma=0.10$ m.
Figure 2.8: Impact of cell size $c$ on the predictive mean using Kernel DM in a one-dimensional example dataset. In this figure, the red line indicates the predictive mean value created using Kernel DM. The kernel width is $\sigma = 0.50$ m in all of the experiments. The cell size is set to: (a) $c = 5.00$ m, (b) $c = 1.00$ m, (c) $c = 0.50$ m and (d) $c = 0.10$ m.
2.2. RELATED WORK – GDM APPROACHES WITHOUT PREDICTIVE VARIANCE

2.2.6 Bayesian Spatial Event Distribution

All the methods presented in this chapter so far are based on the assumption that information is collected using a single type of sensor. In practice, these sensors can have differences in their measurement values due to several factors such as fabrication, age, gas saturation, and humidity. This makes comparison and integration of collected measurements difficult. Moreover, in real-world applications such as incident management, information about the gas dispersion can be available from various sources such as different type of sensors, human reports, and environmental expert reports. To combine information from different sources and get a better interpretation of gas sensor readings, Schaffernicht et al. [19] proposed the Bayesian Spatial Event Distribution (BASED) method. BASED is an event-based method in which detection and non-detection of gas are defined as events. Similar to Kernel DM, this method discretises the explored area into a grid. Detection of an event at each cell is defined by a Bernoulli distribution. Schaffernicht et al. used Beta distribution to derive maximum likelihood of detection of an event at each cell from collected samples. The meta-parameters in the Beta distribution implicitly count the number of events observed in each cell. These meta-parameters are computed using information from neighbouring cells using a Gaussian weight similar to the Kernel DM (see Figure 2.9).

![Figure 2.9](image)

Figure 2.9: To model distribution of gas based on events, the area is discretised into cells. (a) In each cell a prior distribution, Bernoulli distribution, is computed to model likelihood of gas detection events. (b) Meta-parameters of Bernoulli distribution are defined using reported events in the neighbouring cells using a Gaussian kernel.

Schaffernicht et al. [19] applied the BASED method in an indoor controlled experiment where samples we collected using both a small stationary sensor network and a mobile sensor. The qualitative evaluation of BASED in is presented in [19]; however, no quantitative result is available to compare with other methods. The qualitative results in [19], together with the fact that this approach can use information from various sources and is independent of sensor calibration, makes BASED potentially a good candidate to use in real-world scenarios such as the use-case investigated in the Diadem project [29].
Gas distribution modelling in real-world applications is further discussed in Chapter 7.

2.3 Discussion and Conclusion

This chapter presented several statistical gas distribution modelling methods. The resulting model-free statistical gas distribution models make no strong assumptions about the functional form of the gas distribution, such as the number of gas sources or their locations, for example. The discussed methods treat gas measurements as random variables and predict the gas distribution at unseen locations.

Both the mentioned Bayesian based method and Kernel based methods can estimate variance of predicted gas measurements at each location. Most of the works in the state-of-the-art in recent years are applying Kernel based statistical modelling in gas distribution modelling problem. Chapter 4 presents research carried out to enhance Kernel based gas distribution modelling by capturing variance of gas concentration fluctuations in addition to estimating mean of gas measurements. Furthermore, it presents an evaluation metric for gas distribution modelling using predictive variance and presents evaluation results in both indoor and outdoor environments.
Chapter 3

Experimental Setup

I used several datasets to evaluate gas distribution modelling and sensor planning methods presented in this dissertation. These datasets are in two main categories: (1) simulation and (2) real-world. Collecting data and setting up experiments are time-consuming and are not the focus of this dissertation; therefore, available gas dispersion datasets are used. This has been possible through the great effort of researchers from Applied Autonomous Sensor Systems Research Center (AASS) at Örebro University [30, 17, 31, 32] and their collaboration with researchers from DISAL Lab at École Polytechnique Fédérale de Lausanne (EPFL) [33], the Federal Institute for Materials Research and Testing [Bundesanstalt für Materialforschung und -prüfung (BAM)] [8], and Tokyo University of Agriculture and Technology [34].

This chapter describes datasets that are used in this dissertation. First, Section 3.1 explains the experimental setup, data collection, and sensors used for each real-world dataset. This is followed by a brief description of the 2D and 3D simulation engines and the simulation datasets in Section 3.2.

3.1 Datasets Collected in Real-world Experiments

The statistical gas distribution modelling methods and sensor planning approaches proposed in this dissertation are applied to real-world datasets, where samples are collected by either a mobile sensor or a stationary sensor network in environments under both controlled and uncontrolled conditions. Table 3.1 presents information about these real-world experiments.

3.1.1 Controlled Indoor: Experiments in a Wind Tunnel Using a Cartesian Sweeping Arm

Lochmatter et al. [33] performed a set of experiments in a wind tunnel with the area of $18 \times 4 \text{ m}^2$ and height of 1.9 m at the DISAL Lab, EPFL, Lausanne. The experiments were carried out at room temperature with a wind flow of
Table 3.1: List of real-world datasets and their specifications.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Area</th>
<th>Sampling System</th>
<th>Environment</th>
<th>Courtesy of</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPFL</td>
<td>$18 \times 4 \times 1.9$ (m$^3$)</td>
<td>Cartesian arm</td>
<td>indoor, controlled</td>
<td>T. Lochmatter [33]</td>
</tr>
<tr>
<td>Corridor</td>
<td>$16 \times 2$ (m$^2$)</td>
<td>mobile sensor</td>
<td>indoor, uncontrolled</td>
<td>M. Trincavelli, M. Reggentte [35]</td>
</tr>
<tr>
<td>3-Room</td>
<td>$16 \times 4$ (m$^2$)</td>
<td></td>
<td>indoor, uncontrolled</td>
<td>M. Trincavelli, M. Reggentte [31]</td>
</tr>
<tr>
<td>Outdoor</td>
<td>$8 \times 8$ (m$^2$)</td>
<td></td>
<td>outdoor</td>
<td>M. Trincavelli, M. Reggentte [36]</td>
</tr>
<tr>
<td>QC-AASS</td>
<td>$8 \times 6$ (m$^2$)</td>
<td>quadcopter micro-drone</td>
<td>outdoor</td>
<td>P. Neumann, V. Bennetts [12, 16]</td>
</tr>
<tr>
<td>SmallNet</td>
<td>$3 \times 3 \times 4$ (m$^3$)</td>
<td>stationary sensor network</td>
<td>indoor, controlled</td>
<td>M. Trincavelli, Y. Wada [34]</td>
</tr>
<tr>
<td>Diadem-I</td>
<td>$30 \times 20$ (km$^2$)</td>
<td></td>
<td>outdoor</td>
<td>Diadem, DCMR [5]</td>
</tr>
<tr>
<td>Diadem-II</td>
<td>$5 \times 6$ (km$^2$)</td>
<td></td>
<td>outdoor</td>
<td>Diadem, DCMR [5]</td>
</tr>
</tbody>
</table>

approximately 1 m/s, from left to right, within the tunnel. The length of the area that is practically used is 14 m. A single Ethanol gas source was located on the center line of the tunnel at a distance of 1.5 m from the left side of the tunnel. A sketch of the experimental setup is shown in Figure 3.1. This dataset is referred as "EPFL" in the rest of this dissertation.

Robotic Platform

Measurements were collected with a metal oxide gas sensor from E2V Technologies (MiCS-5521) [37] to detect volatile organic compounds (VOC). The sensor was mounted on a Cartesian robot arm. The traversing system was programmed to scan the target area 0.10 m above the ground with the source at lattice points at a distance of 0.12 m. At each point, 50 measurements were recorded with a frequency of approximately 5 Hz. The robotic apparatus takes approximately 3 seconds to reach the next measurement point. An experiment where the whole area is scanned once takes between 10 to 12 hours [33].
Experimental Setup

To collect samples, two configurations were used: (1) experiments in the presence of a laminar flow (see Figure 3.1(a)) and (2) experiments in the presence of turbulence caused by placing an obstacle in the tunnel. As an obstacle, a tall hexagonal prism (placed 4 m downwind from the gas source location) is used. Figure 3.1(b) illustrates the experimental setup.

![Figure 3.1: A sketch of the experimental setup used in the wind tunnel experiments (a) with no obstacle and laminar flow and (b) in the presence of an obstacle: a tall hexagonal obstacle located at (A) 4 m downwind from the source location (Adapted from [33]).](image)

Figure 3.2 shows the odour profile in the wind tunnel with laminar flow and with the presence of an obstacle. The odour profile is measured using a metal oxide sensor (MiCS-5521), from E2V Technologies [37], mounted on a Khepera III robot. The robot traverses in a grid with the resolution of 0.30 m at the length of the tunnel and 0.05 m in its width. Each measurement is recorded as an average over approximately 20 s. The odour profile indicates a slight drift of the wind direction towards one side of the wind tunnel. In [33], Lochmatter reported that this is most likely caused by the fan, and this drift does not significantly affect the results.
In addition to sample collection using a single mobile sensor, the performance of the proposed gas distribution modelling approaches are assessed with sample collection using stationary sensor networks.

Wada et al. [34] developed an experimental setup to perform gas source tracking in a controlled, small-scale indoor environment. The experiments were performed in a $4.85 \times 3.42 \times 2.10 \text{ m}^3$ closed room. Artificial airflow was generated by two rows of fans mounted on the floor and the walls (see Figure 3.3(a)). The fans created a circulating airflow field mimicking natural convection. The airflow velocity near the floor was approximately $0.15 \text{ m/s}$, and was almost uniform over the entire floor. Ethanol and 2-propanol vapours were used as gas source. The gas source was emitting with a constant flow rate ($0.2 \text{ l/min}$). The gas source outlet was placed on the floor approximately in the center of the target area at $0.10 \text{ m}$ from the floor.

A stationary sensor network of 10 metal oxide Figaro TGS 2620 sensors was mounted on the floor of the room (see Figure 3.3(b)). The sampling rate was $4 \text{ Hz}$. In addition, one photoionisation gas detector sensor (ppbRAE 3000, RAE Systems) and two arrays of gas sensors, one comprised of metal-oxide sensors from Figaro Engineering [38] and the other from E2V Technologies [37], were mounted on a Pioneer 3D-DX mobile robot equipped with an LMS2000 SICK laser range scanner and a two-dimensional anemometer (WindSonic,
3.1. DATASETS COLLECTED IN REAL-WORLD EXPERIMENTS

Gill). Table 3.2 presents the list of sensors used in the two sensor arrays. Figure 3.3(c) shows the mobile platform with sensors that were mounted on it.

Table 3.2: Sensor specifications of sensor arrays used in controlled indoor experiments.

<table>
<thead>
<tr>
<th>Sensor Model</th>
<th>Number of Sensors</th>
<th>Target Gas</th>
</tr>
</thead>
<tbody>
<tr>
<td>e2V MiCS 2610</td>
<td>1</td>
<td>Ozone (O_3)</td>
</tr>
<tr>
<td>e2V MiCS 2710</td>
<td>1</td>
<td>Nitrogen Dioxide (NO_2)</td>
</tr>
<tr>
<td>e2V MiCS 5521</td>
<td>2</td>
<td>Carbon Monoxide (CO), Hydrocarbons, VOCs</td>
</tr>
<tr>
<td>e2V MiCS 5121</td>
<td>1</td>
<td>Carbon Monoxide (CO), Hydrocarbons, VOCs</td>
</tr>
<tr>
<td>e2V MiCS 5135</td>
<td>1</td>
<td>Carbon Monoxide, Hydrocarbons, VOCs</td>
</tr>
<tr>
<td>Figaro TGS 2600</td>
<td>2</td>
<td>Hydrogen (H_2), CO</td>
</tr>
<tr>
<td>Figaro TGS 2611</td>
<td>1</td>
<td>Methane (CH_4)</td>
</tr>
<tr>
<td>Figaro TGS 2620</td>
<td>1</td>
<td>Carbon Monoxide (CO), Organic Solvents</td>
</tr>
<tr>
<td>Figaro TGS 2602</td>
<td>1</td>
<td>Ammonia (NH_3), Hydrogen Sulfide (H_2S), VOCs</td>
</tr>
</tbody>
</table>

In this dissertation, we mainly use sensor network data from this set of experiments. Also, we selected experiments where Ethanol was used as the gas source. This dataset will be referred as "SmallNet" in the rest of this dissertation.

### 3.1.3 Uncontrolled Indoor: Experiments Using a Mobile Sensor

Trincavelli et al. [31] and Reggente et al. [35] performed experiments in an uncontrolled indoor environment for the study of pollution monitoring. Two sets of experiments were carried out: one in a corridor and the other in a tri-partitioned room. In the rest of this dissertation, these two datasets are referred as "Corridor" and "3-Room", respectively.

**Robotic Platform**

The experiments were carried out using an ATRV-JR robot. Samples were collected in a predefined sweeping path. The mobile platform included a SICK LMS 200 sensor to localise the robot and build an occupancy grid map of
Figure 3.3: (a) Sketch of the experimental setup. (b) Location of stationary sensors (marked with red circles) and predefined sampling locations for the mobile robot (marked with solid blue line) and (c) the robot used in the "SmallNet" experiments [34].
the environment, an ultrasonic anemometer (Young 81000) to collect wind measurements and a sensor array to collect gas measurements. Table 3.3 lists the metal oxide sensors that this sensor array is comprised of.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Target Gas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figaro TGS 2600</td>
<td>Hydrogen (H₂) and Carbon Monoxide (CO)</td>
</tr>
<tr>
<td>Figaro TGS 2602</td>
<td>Ammonia (NH₃), Hydrogen Sulfide (H₂S), Volatile Organic Compounds (VOCs)</td>
</tr>
<tr>
<td>Figaro TGS 2611</td>
<td>Methane (CH₄)</td>
</tr>
<tr>
<td>Figaro TGS 2620</td>
<td>Organic Solvents</td>
</tr>
<tr>
<td>Figaro TGS 4161</td>
<td>Carbon Dioxide (CO₂)</td>
</tr>
</tbody>
</table>

The electronic nose is enclosed in an aluminium tube mounted horizontally on the robot (see Figure 3.4). The electronic nose is actively ventilated through a fan that creates a constant airflow toward the gas sensors [9].

Figure 3.4: The terrain robot, "Rasmus", used to collect data in uncontrolled indoor and outdoor experiments carried out at AASS (Adapted from [9]).

Experimental Setup

The "3-Room" experiment is carried out in an enclosed area of 16 × 4 m² which is partitioned by protruding walls into three rooms. The robot moves on a predefined sweeping trajectory. As the gas source, a box filled with Ethanol was placed on the floor, approximately in the middle of the room. Samples
were collected with the gas sensors mounted at the height of 0.34 m above the floor. Figure 3.5 shows the experimental setup and the sampling trajectories.

Figure 3.5: The pollution monitoring in the "3-Room" experiment with the "Rasmus" robot. A box containing Ethanol was used as the gas source [9].

In the "Corridor" experiment, an area of $14 \times 4 \text{ m}^2$ in a long corridor with a high ceiling was explored, where a box filled with Ethanol as gas source was placed on the floor close to the center of the target area. Figure 3.6 shows this experimental setup. The mobile robot collected gas measurements in a random walk trajectory [35] with metal oxide sensors mounted on the robot at different heights. In the "Corridor" experiment, we only use samples collected with the gas sensors mounted lowest at 0.10 m above the floor to have samples approximately at the same height as the gas source height. Since Ethanol is heavier than air, the lowest sensors are expected to contain most information about the gas distribution.

3.1.4 Outdoor: Experiments Using a Mobile Sensor

To evaluate the performance of gas distribution modelling in more realistic scenarios, we carried out experiments in outdoor environments. The same robotic platform used for the "Corridor" and "3-Room" experiments was also used to carry out experiments in an open area on the campus of Örebro University [36], where the environmental conditions such as wind, temperature, and humidity are uncontrolled.

The mobile robot moved in a predefined sweeping path to collect gas measurements in an area of $8 \times 8 \text{ m}^2$ where a box containing Ethanol was located
at the center of the area as the gas source (see Figure 3.7). This experiment is referred as "Outdoor" in the rest of this dissertation.

### 3.1.5 Outdoor: Experiments Using a Micro-Drone

A small-scale outdoor experiment was performed using a micro-drone. This experiment was carried out on the campus of Örebro University by P. Neumann from Federal Institute for Materials Research and Testing (BAM), and V. Hernandez and myself from AASS [12, 30]. To collect samples, a sensor planning method was developed within this dissertation for the gas source localisation task using a micro-drone. Datasets created from this experiment are referred as "QC-AASS".

**Robotic Platform**

The mobile robotic measurement system has been developed by BAM, in cooperation with AirRobot GmbH [39]. This system includes a gas-sensitive sensor module (approximately 200 g) for the AirRobot drone AR100-B (see Figure 3.8). The drone can be flown by the line of sight, via on-board video camera and video goggles, as well as by autonomous waypoint tracking. In
Figure 3.7: Pollution monitoring in an open area at the university campus ("Outdoor" experiment) with the "Rasmus" robot. A box filled with Ethanol was used as gas source in the center of the investigation area [36].

the experiments presented in this dissertation, autonomous waypoint tracking mode was used.

Figure 3.8: (a) The AirRobot AR100-B micro-drone and (b) the gas-sensitive payload [8].

The Inertial Measurement Unit (IMU) provides the basis for flight control and wind vector estimation during operation. Magnetic field sensors and GPS are used to improve the accuracy of the IMU and to compensate for sensor drift. The IMU of the drone also contains a barometric pressure sensor to control the drone altitude [8].

A commercially available gas detector (Dräger X-am 5600) is the base unit of the gas-sensitive payload. Depending on the scenario, it can measure
many combustible gases and vapours with the catalytic (Cat.) sensor (thermal sensors) as well as different (toxic) gases, e.g. O\(_2\), CO, H\(_2\)S, NH\(_3\), CO\(_2\), SO\(_2\), PH\(_3\), HCN, NO\(_2\), and C\(_{12}\) with electrochemical and infra-red sensors. The "QC-AASS" dataset includes samples from CO concentration in the explored area.

Experimental Setup

The "QC-AASS" experiment was carried out in an outdoor environment with the area of \(8 \times 12\) m\(^2\). The robotic platform was a micro-drone equipped with electrochemical CO sensors (Dräger Sensors XXS) [8]. Gas concentration and wind measurements were recorded with a frequency of 1 Hz. Measurements were taken at each measurement position for about 20 seconds. At each measurement position, the average wind vector was used for all individual gas sensor measurements acquired at this measurement position.

The flight speed of the drone between measurement positions was set to 1 m/s. Because of the low flight height of about 1 m, the height of the drone was controlled manually during the experiments. Each run took around 14 to 19 minutes to complete. A barbecue filled with burning coal and fresh damp wood was used as a pollution source (see Figure 3.9) and was placed approximately in the middle of the target area (at approximately (6.3, 3.8) m from the bottom left corner). The drone was set to autonomous waypoint mode directly after take-off, which started the experiment.

Figure 3.9: Pollution source and micro-drone during one of the experiments at AASS [8].
3.1.6 Large-scale Outdoor: Experiments Using a Stationary Sensor Network

As part of the evaluation and case study in the Diadem project [29], one of the industrial partners, DCMR, provided two datasets using a network of stationary sensors in the Rijnmond area of the Rotterdam port in the Netherlands. DCMR is the regional environmental protection agency of the local and regional authorities operating in Rijnmond, the larger Port of Rotterdam area in the Netherlands [40]. This area is below sea level and polluted because of refinery factories, chemical plants, waste dump sites, etc. The stationary sensors are mounted in this area to monitor gas pollution and plan for further investigation of leakage sources and potential evacuation plans accordingly. For this purpose, two sets of data were collected: "Diadem-I" and "Diadem-II".

"Diadem-I" was collected by a sparse network of MOX (metal oxide) sensors in a 2D configuration (all sensors were at approximately equal height). The sensor network included 10 e-noses which were placed in a large area (approximately 56 km²) in the Rotterdam port (see Figure 3.10). Each e-nose includes four MOX gas sensors (TGS 2602, TGS 2620, TGS 2611 and TGS 2600), which have been steady-state calibrated once installed. These sensors recorded the gas measurements at 3 minute intervals. Coarse wind information was also available via five wind measurement stations, which were, however, not located at the same place and height as the e-noses.

Figure 3.10: Google Earth maps from the Rijnmond area where the sensors are mounted to collect the "Diadem-I" dataset. The white circles indicate e-nose positions, and the green/red/orange/blue bars illustrate the relative value of the respective sensor response of the four gas sensors in each e-nose. The white triangles indicate the meteorological stations [5].
Next, a denser sensor network was mounted in the Rotterdam port. In "Diadem-II", the target area was smaller (approximately $6.8 \times 5.2 \text{ km}^2$) than the one in "Diadem-I" setup and the sensor density was higher (network with 30 e-noses). Similar to "Diadem-I", the dataset was collected by DCMR with a sparse array of MOX (metal oxide) semiconductor sensors in a 2D configuration, which were steady-state calibrated once installed. Gas measurements were recorded at 3 minute intervals. The measurements are available for a period of 12 months in 2009 and 2010. The wind information from five meteorological stations (with varying height) was available. The sensor locations are shown as an overlay of the Google Earth map in Figure 3.11.

Figure 3.11: Google Earth map from the Rijnmond area where the "Diadem-II" dataset was collected. The green pins indicate e-nose positions. The blue icons indicate meteorological stations [5].

### 3.2 Datasets Collected in Simulation Experiments

Evaluation of gas distribution modelling approaches requires a sufficient number of samples with a good coverage of the observation area over a certain period and reliable ground truth information. Because of the difficulty in collecting concentration measurements and the respective ground truth data in real-world experiments, it is instructive to complement real-world measurements with data from realistic gas dispersal simulations. Apart from the availability of ground truth information, simulation of gas dispersal has additional advantages: it enables repetition of experiments under identical conditions, allows relatively effortless testing in various scenarios, and avoids intricate calibration issues which typically occur with gas sensors used in real-world experiments. For this purpose, we use simulation experiments.
3.2.1 Simulated 2D Gas Dispersal Experiments

To study and evaluate different statistical gas distribution modelling and sensor planning methods, we used data from a 2D gas dispersion simulation engine developed by Pashami et al. [17]. This gas dispersal simulation package integrates OpenFOAM [41] fluid flow simulation and the filament-based gas propagation model proposed in [42].

OpenFOAM is an open source CFD simulation tool, which solves the governing equations of fluid flow numerically. OpenFOAM offers different CFD algorithms to model laminar or turbulent flow. The wind tunnel environment was selected to match available data from real experiments performed in a wind-tunnel at the DISAL Lab, EPFL, Lausanne [33]. Detailed description of the original experiment is presented in Section 3.1.1. In the real wind tunnel experiment samples were collected using a sweeping movement of a 2D arm over several hours and no ground truth data was available. Therefore, we used the real data only to validate the gas dispersal simulation [17] and used simulated wind tunnel data for the study in this dissertation.

Two simulation experiments were performed in a wind tunnel environment of $16 \times 4 \, \text{m}^2$ with an inlet of approximately 1 m/s on the left and an Ethanol gas source at $(x,y) = (1.2) \, \text{m}$. In one set of experiments, we simulated gas dispersion under the effect of predominantly laminar air flow in an empty wind tunnel ("Sim-No-Obstacle" data set). In the second set of experiments, we introduced turbulent flow by adding a second inlet of 2 m/s at the bottom of the tunnel and placing an obstacle in the tunnel ("Sim-With-Obstacle" dataset).

Measurements in both experiments were collected at random locations with a sampling frequency of 1 Hz. To capture the temporal variety of an evolving gas plume, we selected data from the start of the odour release until some time after the gas plume reached the end of the tunnel (20 s) when the plume was expected to be fully evolved. Figure 3.12 shows a snapshot that illustrate gas dispersion at $t = 20 \, \text{s}$ in the "Sim-No-Obstacle" experiment. In the simulation, the flow model is built for a wind tunnel with an inlet at the left and outlet at the right; however, in the filament propagation, a wall is simulated at the right end of the tunnel. Therefore, in Figure 3.12, a patch of accumulated filaments is observed close to the wall at the right end of the tunnel. In the evaluations and comparisons presented in next chapters, to make the simulation comparable with real-world wind tunnel experiments, 0.2 m of the right end of the simulated wind tunnel is discarded.

To build the flow model for the "Sim-With-Obstacle" dataset, a rectangular obstacle of width 0.8 m and length 2.0 m was placed below the horizontal center line of the tunnel and in front of the second inlet. The obstacle was centred at $(7,1.6) \, \text{m}$. Figure 3.13 shows snapshots of the gas dispersion, the corresponding wind intensity and wind direction map at $t=20 \, \text{s}$ in the "Sim-With-Obstacle" experiment. Measurement values are available at every 0.1 m.
3.2. DATASETS COLLECTED IN SIMULATION EXPERIMENTS

3.2.2 Simulated 3D Gas Dispersal Experiments

In addition to the two simulated wind tunnel data sets, we analyse gas distribution modelling and sensor planning with a new 3D gas dispersion simulator [32]. As [17], the new gas dispersion simulation engine also uses a combination of OpenFOAM to compute fluid flow and the filament-based gas propagation model proposed in [42]. However, the simulation engine is integrated into the Robotic Operation System (ROS) and offers additional models corresponding to the metal oxide sensors used in the real-world experiments.

The 3D simulation experiments were performed in a large simulated environment \( (60 \times 20 \times 5 \text{ m}^3) \), where two obstacles introduce additional turbulence. The flow direction is from left to the right of the tunnel with an inlet of 1 m/s on the left. An Ethanol gas source is placed close to the floor at a height of 0.10 m to match real-world experiments. The simulation resolution is 0.5 m. From the 3D data, we used simulated measurements collected at the lowest available height, 0.25 m. The concentration at 0.25 m is the concentration inside a cube located on the height 0.0 to 0.5 m. The sampling location is the center of such a cube. Therefore, in this resolution measurements at 0.25 m and gas source location (0.10 m) are both in the simulation cells with the lowers height.

Samples are collected with a frequency of 1 Hz. In this environment, it takes approximately 50 s for the plume to reach the end of the tunnel and we analysed samples for the first 100 s after the gas source was activated. Figure 3.14 shows a snapshot of gas dispersion in this simulated environment where the gas source is located at \( (5,10,0.10) \text{m} \). The two large blue cubes indicate obstacles. High concentrations of gas – which tend to appear downstream of the obstacles – are indicated by shades of red while areas with low concentrations appear with blue colours. This dataset is referred as "Sim-ROS" in the rest of this dissertation.
In real-world scenarios a range of complex situations and changes can occur. In order to capture some of these more complex scenarios, further simulation experiments were performed with the 3D ROS gas dispersion simulation and considered the following situations:

**cs1/2 Two different gas source locations:** We first carried out experiments with the gas source at different places: once in the center of the target area, $P_A = (5.0, 10.0, 0.10)$ m, and once in a non-central position $P_B = (5.0, 15.0, 0.10)$ m. In the corresponding first two experiments the
3.2. DATASETS COLLECTED IN SIMULATION EXPERIMENTS

Figure 3.14: A 2D Snapshot from the 3D ROS gas dispersion simulation. The blue filled cubes indicate obstacles. The coloured dots represent patches of gas: shades of blue indicate low concentrations while red colors indicate higher gas concentration levels. The gas source location is shown with a white filled circle.

gas release rate was kept constant. While in the experiment with $P_A$ as gas source location three spots of increased gas concentration formed, upwind of the two obstacles and the outlet, only one spot of increased gas concentration appeared upwind of an obstacle in the experiment with $P_B$.

cs3 Constant versus variable gas release rate: In a third experiment with the gas source at $P_A$, we increased the gas release rate in random steps depending on the current release rate (with a frequency of 0.2 Hz we chose an increment randomly from 0–100% of the number of molecules of the current release rate). On average, the release rate was thus increased exponentially. We did not use a fixed increment to avoid a negligible increase of the release rate.

cs4 Ideal versus slow sensor response model: In the real-world experiments presented in this dissertation, mostly metal oxide sensors are used. Therefore, we also carried out an experiment in which measurements of a metal oxide sensor were simulated using a sensor model for the metal-oxide (MOX) Figaro TGS-2620 sensor. In all other experiments, we used an ideal sensor model, i.e. that the sensor values reflect the concentration directly, approximating e.g. the behaviour of photoionisation detections (PIDs). The gas source was located at $P_A$ in this experiment, and the gas release rate was kept constant.
Chapter 4

GDM Approaches With Predictive Variance

Chapter 2 presented an overview of the state-of-the-art in statistical gas distribution modelling. Compared to controlled indoor environments, outdoor environments present additional challenges such as stronger changes in wind direction and intensity, air pressure and temperature. Therefore, in practical applications, we are interested in approaches that can estimate the variance of the gas concentration fluctuations.

In this chapter, different Kernel methods that estimate both the mean and variance of gas distribution are discussed. Section 4.1 presents Kernel DM+V, which is an extension of the Kernel DM method discussed in Chapter 2. In addition, this section discusses methods in the literature that are derived from Kernel DM+V for a better prediction of gas distribution in real-world applications (see Section 4.1.1). Section 4.2 explains methods which use Gaussian Processes to create gas distribution models; particularly, it discusses GPMM which uses a mixture of Gaussian Processes in GDM (see Section 4.3).

Section 4.4 presents an evaluation measure to evaluate performance of different GDM methods. A comparison of the mentioned the Kernel approaches is presented in Section 4.5. As it will be discussed in Section 4.4, estimating the predictive variance enables better quantitative evaluation of alternative statistical models and provides the basis to learn meta-parameters. The predictive variance can also be beneficial for sensor planning based on the current model. The chapter ends with suggestions, considering the previous works, to enhance gas distribution modelling in real-world applications and to use the information from gas distribution models to select the next sampling locations (Section 4.6).

The content of this chapter is mainly based on a book chapter that we published on statistical gas distribution modelling using Kernel methods [1].
4.1 Kernel DM+V

Kernel DM computes an estimate of the mean gas distribution. However, the spatial structure of the distribution variance can provide important information about the gas distribution by highlighting areas of high fluctuation, which are often found in close vicinity to a gas source. Kernel DM+V proposed by Lilienthal et al. [9] is a statistical modelling approach which models mean and variance distributions.

As in Kernel DM, the distribution is represented as a grid map in Kernel DM+V, and a uni-variate Gaussian kernel \( N \) is used to represent the importance of measurements \( r_i \) obtained at the location \( x_i \) for the statistical model at grid cell \( k \). In the first step of Kernel DM+V, two temporary grid maps are computed: \( \Omega^{(k)} \) by integrating spatial importance weights as in Equation 2.3 and \( R^{(k)} \) by integrating weighted readings as in Equation 2.4. The integrated weights \( \Omega^{(k)} \) are used to normalize the weighted measurements \( R^{(k)} \) and to compute a further map \( \alpha^{(k)} \), which represents the confidence in the obtained estimates

\[
\alpha^{(k)} = 1 - \exp\left(-\frac{(\Omega^{(k)})^2}{\sigma^2_{\Omega}}\right). \tag{4.1}
\]

The confidence map is used to compute the mean concentration estimate \( r^{(k)} \) as

\[
r^{(k)} = \alpha^{(k)} \frac{R^{(k)}}{\Omega^{(k)}} + \{1 - \alpha^{(k)}\} r_0 \tag{4.2}
\]

where \( r_0 \) represents the mean concentration estimate for cells for which there is not sufficient information available from nearby measurements indicated by a low value of \( \alpha^{(k)} \). In the work of Lilienthal et al. [9], \( r_0 \) is set to be the average over all sensor readings. Whether \( \alpha^{(k)} \) is considered low or not is determined by the scaling parameter \( \sigma^2_{\Omega} \), which defines a soft margin for values of \( \Omega^{(k)} \).

The variance estimate for cell \( k \) depends on the density of measurements in the vicinity of a cell and the true variance of measurements at the location of the cell, computed from variance contributions \( (r_i - r^{(k(i))})^2 \), where \( r^{(k(i))} \) is the predicted mean for the cell \( k(i) \) closest to the measurement point \( x_i \).

The variance map \( V^{(k)} \) is computed from the variance contributions integrated in a further temporary map \( V^{(k)} \)

\[
V^{(k)} = \sum_{i=1}^{n} \omega^{(k)}_{i} (r_i - r^{(k(i))})^2, \tag{4.3}
\]

and

\[
v^{(k)} = \alpha^{(k)} \frac{V^{(k)}}{\Omega^{(k)}} + \{1 - \alpha^{(k)}\} v_0, \tag{4.4}
\]
where $v_0$ is an estimate of the distribution variance in regions far from measurement points. In the work of Lilienthal et al. [9], $v_0$ is computed as the average over all variance contributions.

The Kernel DM+V algorithm depends on three parameters: (1) the kernel width $\sigma$, which governs the amount of extrapolation on individual sensor measurement; (2) the cell size $c$ that determines the resolution at which different predictions can be made; and (3) the scaling parameter $\sigma_{\Omega}$, which defines a soft threshold that decides whether it is assumed that sufficient knowledge from nearby measurements is available to make predictions. Smaller values of $\sigma_{\Omega}$ entail a lower threshold on $\Omega^{(k)}$, i.e. an increasing tendency to trust the distribution estimate obtained from extrapolation on local measurements. The exact value of $\sigma_{\Omega}$ is not critical as long as it is of the right scale. Kernel width and cell size can be learned from the collected measurements. The meta-parameter selection for Kernel DM+V is discussed in Section 4.4.

Figure 4.1 shows an example of a weight map $\Omega^{(k)}$ (top row) and the corresponding confidence map $\alpha^{(k)}$ (bottom row). For narrow kernels and large values of $\sigma_{\Omega}$ (left column), one can see the trajectory of the gas sensor carried by the robot, indicating that the predictions from extrapolation will only be considered trustworthy at locations close to the actual measurement. For wider kernels or smaller values of $\sigma_{\Omega}$ (right column), the area for which predictions are made is larger.

Figure 4.1: Weight map $\Omega^{(k)}$ (top row) and the corresponding confidence map $\alpha^{(k)}$ (bottom row) obtained using parameters $\sigma=0.1m$, $\sigma_{\Omega} = 5.0 \cdot N(0, 0.10) \approx 20.0$ (left column) and $\sigma=0.5m$, $\sigma_{\Omega} = 1.0 \cdot N(0, 0.50) \approx 0.8$ (right column) on a grid with cell size $c=0.05m$. This figure is reproduced from [9].
4.1.1 Extensions of Kernel DM+V

Using Wind Information for Gas Distribution Mapping

The statistical models, discussed earlier in this dissertation, have been used to model gas dispersion in indoor or controlled environments with relatively weak air flow. An important aspect for gas distribution modelling, especially in outdoor applications, is to consider wind information when building the gas distribution model. The local airflow is in general an important parameter because of the strong influence of advective transport on gas dispersal. Regente et al. [36] proposed the Kernel DM+V/W algorithm. Kernel DM+V/W is an extension of the Kernel DM+V. It considers wind information in order to compute the gas distribution model. Spatial integration of the point measurements is carried out using a bi-variate Gaussian kernel. By selecting the shape of the kernel based on the local wind measurement, the Kernel DM+V/W algorithm models the information content depending on the direction and intensity of the wind (see Figure 4.2). When reliable wind information is available, the symmetric uni-variate Gaussian kernel in Equation 2.2 is replaced by an elliptic, bi-variate Gaussian kernel with the semi-major axis oriented along the wind direction. The $2 \times 2$ covariance matrix $\Sigma$ is computed according to the current measurement of the local airflow $\vec{v}$ at the sensor location as described below.

First, we assume that the total information content is the same for each measurement. This assumption results in the constraint that the area of the covariance ellipse remains constant, that is

$$\pi a^2 = \pi ab.$$  \hspace{1cm} (4.5)

The semi-major axis $a$ is then stretched according to the wind intensity assuming a linear dependency as

$$a = \sigma + \gamma |\vec{v}|.$$  \hspace{1cm} (4.6)

By combining Equation 4.5 and Equation 4.6, the length of the semi-minor axis is

$$b = \frac{\sigma}{1 + \gamma |\vec{v}|/\sigma}.$$  \hspace{1cm} (4.7)

Equations 4.6 and 4.7 describe the relation between the Eigenvalues of the covariance matrix and the wind intensity. The parameter $\gamma$ is related to the duration over which the wind vector $\vec{v}$ is assumed to be constant. For the initial experiments described in [36], $\gamma$ was heuristically set to 1. $\gamma$ can also be learned from the input data [20].

Finally, the covariance matrix is rotated so that the semi-major axis is aligned with the wind direction

$$\Sigma^{-1} = R^{-1} \Sigma^{-1} R.$$  \hspace{1cm} (4.8)
where $R$ is the rotation matrix, and $\Sigma^{-1}_R$ is the inverse of the rotated covariance matrix. By means of the adaptive kernel described by Equations 4.6, 4.7 and 4.8, Kernel DM+V/W models information about the trajectory of a sensed patch of gas, given a wind measurement. The bi-variate Gaussian kernel describes a distribution over the locations where the sensed patch came from and where it tends to move (here it is important to bear in mind that the resulting model describes the time-averaged gas distribution). Reggente and Lilienthal [36] performed experiments in wind tunnel and outdoor environments. The results indicated that applying the wind extension improves the performance of the Kernel DM+V approach.

![Figure 4.2: Visualisation of the kernel in Kernel DM+V/W at the presence of wind and absence of wind. Left: For a sample marked by a black $\otimes$, the blue dashed line indicates the contour of the kernel in the absence of wind and the red solid line shows the elliptic contour of the kernel at the presence of wind. Right: the cells affected by the kernel are marked by cross [36].](image)

### Three-Dimensional Gas Distribution Mapping

Kernel DM+V models gas distribution in two dimensions. In outdoor environments, changes in physical conditions such as changes in temperature and pressure are uncontrolled and have stronger impact on gas dispersion in three dimensions. Therefore, we are interested in models which can incorporate measurements from different heights and predict gas distribution in three dimensions. In the work of Reggente and Lilienthal [43], a modification of the two-dimensional Kernel DM+V algorithm to build three-dimensional gas distribution maps is introduced. The idea of 3D Kernel DM+V is similar to the basic idea of Kernel DM+V/W discussed in Section 4.1.1. In both cases, only the kernel is modified compared to the original algorithm. 3D Kernel DM+V uses kernel
extrapolation with a tri-variate Gaussian kernel to model the decreasing likelihood that a sensor reading represents the true concentration with respect to the distance in three dimensions. The 3D Gaussian kernel (see Figure 4.3) is defined by a kernel width along the three axis \( \sigma_x, \sigma_y, \) and \( \sigma_z \):

\[
\omega_i^{(k)}(\sigma_x, \sigma_y, \sigma_z) = N(\delta_i^{(k)}, \sigma_x, \sigma_y, \sigma_z). \tag{4.9}
\]

As visualised in Figure 4.3, weights are evaluated at the distance between the location of the measurement \( x_i \) and the center \( x^{(k)} \) of the cell \( k \). In the same way as in the two-dimensional version, weights, weighted sensor measurements, and weighted variance contributions are integrated and stored in temporary grid maps \( \Omega^{(k)}, R^{(k)}, \) and \( V^{(k)} \), now using Equation 4.9 instead of Equation 2.2.

Reggente and Lilienthal [43] used 3D Kernel DM+V to create gas distribution models in an uncontrolled indoor environment experiment. In this experiment, a mobile robot equipped with three gas sensor arrays mounted at different heights performed random walks to collect data. The 2D model created by the measurements from the sensor in the middle was used to evaluate the 3D map created from measurements taken from the lower and upper sensors as follows: First the data collected by the lower and upper sensors are used to build the 3D model of gas distribution in the environment. Then, this model is sliced into layers with different heights. To evaluate 3D Kernel DM+V, the model created for the middle layer is compared with the model created from real measurements collected by the middle sensor using the two-dimensional Kernel DM+V algorithm. The comparison between the slice of the 3D map and the 2D map at the middle sensor shows clear structural similarities. These initial results demonstrate that 3D Kernel DM+V enables creating maps in three dimensions which provide useful information at heights different from the height of the sensors.

## 4.2 Gaussian Processes

Gaussian Processes (GPs) are non-parametric probabilistic models for solving regression problems [44]. GPs can provide an estimate both on the mean and an variance distributions. GP is easy to implement and it has been widely used in different robotic problems [45]. In the work of Stachniss et al. [11], Gaussian Processes are applied to model gas distribution.

One can view GPs as an infinite dimensional Gaussian distribution defined by a mean and a covariance function. A covariance matrix models the influence of neighbouring data points. More formally, assuming that \( \{(x_i, r_i)\}_{i=1}^n \) are samples of a Gaussian process, \( f = (r_1, r_2, ..., r_n)^T \) has a Gaussian distribution

\[
r_i = f(x_i) := \mathcal{N}(\mu, K), \mu \in \mathbb{R}^n, K \in \mathbb{R}^{n \times n} \tag{4.10}
\]
where $\mu$ is the mean and $K$ corresponds to the covariance matrix that is constructed using a so-called covariance function $k$. For simplicity of notation, one can assume that $\mu = 0$, since the expectation is a linear operator and, thus, for any deterministic mean function $m(x)$, the Gaussian process over $f(x) := f(x) - m(x)$ has zero mean [11, 44]. A standard choice for the covariance is the squared exponential function

$$[K]_{ij} = k(x_i, x_j) = \sigma_f^2 \exp\left(-\frac{1}{2} \frac{|x_i - x_j|^2}{l^2}\right) \tag{4.11}$$

in which $l$ is the length-scale parameter that influences the smoothness of the function $f$, and $\sigma_f^2$ is the signal variance parameter. The parameter $\sigma_n^2$ represents the global noise. The parameters $l$, $\sigma_f^2$, and $\sigma_n^2$, termed the meta-parameters of the GP model, have to be learned.

Let $X = [x_1; x_2; \ldots; x_n]^T$ and $R = [r_1; r_2; \ldots; r_n]^T$ be the set of locations and the corresponding measurements, respectively. An arbitrary test set $X_*$ is chosen from the input data set. As described in Section 2.2.4, assuming that $D = \{(x_i, r_i)\}_{i=1}^n$ is the set of measurements $r_i$ collected at locations $x_i$, the goal of gas distribution modelling is to predict accurately the value of $r_*$ at an unseen location $x_* \in X_*$. To predict the value of the measurement at $x_*$, predictive mean and variance are estimated as

$$\tilde{f}(X_*) := E[f(X_*)] = k(X_*, X)[k(X, X) + \sigma_n^2 I]^{-1} r \tag{4.12}$$

and

$$V[f(X_*)] = k(X_*, X_*) - k(X_*, X)[k(X, X) + \sigma_n^2 I]^{-1} r,$$  

$$\tag{4.13}$$
respectively. The main drawback of GPs is their computational complexity. To estimate the predictive mean in Equation 4.12 and the predictive variance in Equation 4.13, the inversion of $[k(X, X) + \sigma_n^2 I]$ has to be computed. This inversion has $O(n^3)$ time complexity where $n$ is the size of the training data set. Standard GPs create a uni-modal distribution for any given location. This is a limitation of GPs when it comes to the context of gas distribution modelling since the histogram of the observations of a gas distribution over time, typically shows at least two modes corresponding to noisy measurements of concentration peaks and the background.

4.3 Gaussian Process Mixture Model

One solution to overcome uni-modality is using Gaussian Process Mixture models (GPM). Stachniss et al. [11] proposed a method based on GPM to estimate the predictive variance in gas distribution modelling. A GPM model is a locally weighted sum of GP models. Let $\{\text{GP}_i\}_{i=1}^m$ be a set of GPs as individual components of the mixture model. To predict the value of $r_*$ at the location $x_*$, we must first estimate that to which component – peaks or background – this measurement belongs. Then, using the distribution model of the corresponding component, the value of the measurement $r_*$ can be predicted.

A gating function decides to which mixture component a given measurement belongs. More formally, a gating function $P(z(x_*) = i)$ is defined as the probability that $x_*$ is associated with the component GP$_i$. The likelihood of observing $r_*$ is achieved by

$$h(x_*) := p(r_* | x_*) = \sum_{i=1}^m P(z(x_*) = i) \cdot N_i(r_*; x_*) \quad (4.14)$$

where $N_i(r_*; x_*)$ denotes the Gaussian distribution function with mean $\bar{x}_i$ and the corresponding variance $\text{V}[f_i(x)] + \sigma_n^2$. To create the GPM model, the predictive mean and variance are obtained by

$$\bar{h}(x_*) := E[h(x_*)] = \sum_{i=1}^m P(z(x_*) = i) \cdot \bar{f}_i(x_*) \quad (4.15)$$

and

$$\text{V}[h(x_*)] = \sum_{i=1}^m P(z(x_*) = i) \cdot \left( \text{V}[f_i(x_*)] + \left(\bar{f}_i(x_*) - \bar{h}(x_*)\right)^2 \right). \quad (4.16)$$

In the work of Stachniss et al. [11], the gating function and meta-parameters of the mixture models are learned using Expectation Maximisation (EM). EM is an iterative process to maximise the likelihood estimates of the parameters of the predictive model. EM consists of two steps: first, (E) step which estimates
the probability of a measurement belonging to a mixture component, and then (M) step which computes meta-parameters maximising the expected log-likelihood found on the E step. The following section describes in detail the algorithm to build GDM using GPM.

### 4.3.1 Initialisation of the Mixture Components

To learn the parameters of the first component $GP_1$, a set of $n_1$ data is randomly selected from the data set $\mathcal{D}$. $GP_1$ is used as the initial estimate of the gas distribution. This initial component usually represents areas with unusual gas accumulation poorly. To improve the accuracy of the model, we define an error function $GP_{\Delta}$. This error function is learned using data from $\mathcal{D}$ excluding the training set $n_1$. $GP_{\Delta}$ captures the absolute difference between a set of target values and the prediction of $GP_1$. The next component is initialised from the data, which indicated higher error in the created model. The subsampling procedure reduces the size of the data set and the time complexity, respectively, to learn the model. This method recognizes two GP components for the gas distribution modelling problem. The two components of the GPM represent a background signal and peaks.

### 4.3.2 Iterative Learning via Expectation-Maximisation Method

The Expectation-Maximisation (EM) algorithm is an iterative method to obtain the maximum likelihood estimate of parameters of the predictive model. The EM algorithm consists of two steps: First, it estimates the probability that data point $j$, $x_j \in X$, corresponds to model component $i$. This is done by re-estimating the gating function for each data point $j$ as

$$P(z(x_j) = i) \leftarrow \frac{P(z(x_j) = i)N_i(r_j; x_j)}{\sum_{k=1}^{m} P(z(x_j) = k)N_k(r_j; x_j)}.$$  \hspace{1cm} (4.17)

In the next step, the components are updated based on the new estimate of $P(z(x_j) = i)$. To update the component $i$, the predictive mean and variance of the mixture model are computed. This is achieved by modifying Equation 4.12 to

$$\tilde{f}_i = k(X_*, X)[k(X, X) + \Psi^i]^{-1}r$$  \hspace{1cm} (4.18)

where

$$[\Psi^i]_{jj} = \frac{\sigma^2_n}{P(z(x_j) = i)}.$$  \hspace{1cm} (4.19)

The mixture model has three meta-parameters $\{\sigma_r, l, \sigma_n\}$ which have to be estimated. A common way is to first initialise these values using a heuristic and then optimising the values. A reasonable heuristic is the one proposed by
Snelson and Ghahramani [46], which is also used in the work of Stachniss et al [11]:

\[
l \leftarrow \max_{x_j} P(z(x_j) = i) \|x_j - \overline{x}\|,
\]

\[
\sigma^2_t \leftarrow \frac{\sum_{j=1}^{n} P(z(x_j) = i)(r_j - E[r])^2}{\sum_{j=1}^{n} P(z(x_j) = i)},
\]

\[
\sigma^2_n \leftarrow \frac{1}{4} \sigma^2_t,
\]

where \( \overline{x} \) is the weighted mean of the inputs, each \( x_j \) having a weight of \( P(z(x_j) = i) \).

### 4.3.3 Learning the Gating Function for Unseen Test Points

In the EM algorithm, for each data point \( j \) from the training set, the probability of being assigned to the component \( GP_i \) is determined by maximizing the cross validation data likelihood. To generalize this to unseen locations, another GP is used to model a proper gating function. Stachniss et al. [11] use a gating GP for each component \( i \), that uses the \( x_j \) as input and outputs \( z(x_j) \) according to the EM output. In this way, a gating function is obtained that models the assignment probability for an arbitrary location as

\[
P(z(x_*) = i) = \frac{\exp(f_{z_i}(x_*))}{\sum_{j=1}^{m} \exp(f_{z_j}(x_*))},
\]

Here \( f_{z_i}(x) \) denotes the prediction of \( z \) for \( GP_i \) computed in the EM algorithm.

To illustrate this procedure, consider the following one-dimensional toy example presented in the work of Stachniss et al. [11]. The first part of the data points were uniformly distributed around a \( y \) value of 2, while the second part was generated with higher noise at two distinct locations. The left image of Figure 4.4 depicts the standard GP learned from the input data points, and the right one shows the resulting error GP. Based on the error GP, the second mixture component is initialised and used together with the first component as the input to the EM algorithm. The individual images in Figure 4.5 illustrate the iterations of the EM algorithm (to be read from left to right and from top to bottom). They depict the two components of the mixture model. The learned gating function after the convergence of the algorithm is depicted in the left image of Figure 4.6 and the final GP mixture model is shown in the right image. It is obvious that this model is a better representation of the input data set than the standard GP model shown in the left image of Figure 4.4.
4.3. GAUSSIAN PROCESS MIXTURE MODEL

Figure 4.4: Left: The standard GP used to initialise the first mixture component. Right: The error GP used to initialise the second mixture component [1, 11].

Figure 4.5: Components in different iterations of the learning using EM algorithm [1, 11].

Figure 4.6: Left: The learned gating function. Right: Resulting distribution of the GP mixture model [1, 11].
4.4 Meta-parameter Selection and Evaluation Method

4.4.1 Evaluation Measure

After creating a model, it is important to know whether the model is good for a specific application or not, and to compare its performance with other models. The criterion applied here for a good model is that

- it explains the training observations and
- it accurately predicts unseen observations.

To evaluate a gas distribution model, some ground truth information is required. An indirect approach is to quantify the capability of the model to identify hidden parameters, for example, the location of the gas source. However, local gas accumulation peaks do not necessarily correspond to the gas source location. If there is a dense grid of stationary sensors available, then one simple solution to evaluate the created model is to compare the model derived from all but one or a few sensors with the measurements of the excluded sensors. However, when the sensor network is too sparse or there is only a single mobile sensor, this solution is not feasible. In the work of Stachniss et al. [11] and Lilienthal et al. [18], a similar method is applied and model predictions are compared to unseen measurements. A gas distribution model represents the time-averaged concentration and the expected fluctuations. The mean and variance of the distribution are both considered by the average Negative Log Predictive Density (NLPD), which is a standard criterion to evaluate distribution models. Under the assumption of a Gaussian posterior \( p(r_i \mid x_i) \), the NLPD of unseen measurements \( \{r_1, r_2, ..., r_n\} \) acquired at locations \( \{x_1, x_2, ..., x_n\} \) is computed as

\[
\text{NLPD} = \frac{1}{2n} \sum_{(x_i, r_i) \in \mathcal{D}} \left\{ \log \hat{v}(x_i) + \frac{(r_i - \hat{r}(x_i))^2}{\hat{v}(x_i)} \right\} + \frac{1}{2} \log(2\pi),
\]

(4.24)

where \( \hat{v}(x_i) \) and \( \hat{r}_i \) are estimates of the predictive variance and the mean, respectively.

4.4.2 Learning of Meta-Parameters

Kernel methods rely mainly on the two meta-parameters, kernel width \( \sigma \), and cell size \( c \). Since an estimate of the variance is available, the meta-parameters can be learned using the NPLD. This can be done, for example, by dividing the sample set \( \mathcal{D} = \{(x_i, r_i)\}_{i=1}^n \) into disjoint sets \( \mathcal{D}_{\text{train}} \) and \( \mathcal{D}_{\text{test}} \), and determining optimal values of the model parameters by cross-validation on \( \mathcal{D}_{\text{train}} \), keeping \( \mathcal{D}_{\text{test}} \) for evaluation.
4.5 Experimental Comparison

In the same way as a fixed distribution model is evaluated depending on its meta-parameters, different gas distribution modelling approaches can be compared by their respective NLPD for unseen measurements. Since the goal is to maximise the likelihood of unseen measurements, better models minimise the NLPD. Table 4.1 shows an NLPD comparison of the models created using GPs (see Section 4.2) and the GPM (see Section 4.3) with the model obtained from the Kernel DM+V algorithm (Section 4.1). The comparison is based on datasets from three different environments in which a robot carried out a sweeping movement consisting of two full sweeps. The first sweep was used for training and the second sweep (in opposite direction) for testing. As a preliminary result from this investigation, GDM models created using GPs show a poor performance among the three compared approaches, while GPM and Kernel DM+V exhibit a comparable performance for gas distribution modelling in the tested environments. In Section 4.5.1, we discuss and compare further the two approaches, GPM and Kernel DM+V.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>NLPD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GP</td>
</tr>
<tr>
<td>3-Room</td>
<td>-1.22</td>
</tr>
<tr>
<td>Corridor</td>
<td>-0.98</td>
</tr>
<tr>
<td>Outdoor</td>
<td>-1.11</td>
</tr>
</tbody>
</table>

4.5.1 Discussion of Kernel DM+V and the GPM Approach

Kernel DM+V discretises the space into grid cells and estimates a variance in addition to the mean of the gas distribution. Gas distribution modelling is treated as a density estimation problem. The predictive variance depends on the true variance of the measurements at a certain location and the distance of the measurements to this location. Kernel DM+V weights the importance of each data point for the estimate at a certain grid cell by a kernel function that depends on the distance of the measurement location from the respective centre of the cell.

The GPM approach models gas distributions as a locally weighted sum of Gaussian Process models. Gas distribution modelling is treated as a regression problem. A kernel describes the covariance matrix, which models the influence of measurements at neighbouring data points on the estimation. The approach proposed by Stachniss et al. [11] assumes two mixture components. These components represent peak areas, where sensors tend to respond strongly, and background, which models the signal in the absence of a distinct sensor re-
sponse. At each location, these two components are mixed in different proportions. The GPM approach proposed by Stachniss et al. [11], learns the separate components and a gating function that determines the probability with which a location is described by the components. The basic representation of the model is the collected data set. The GPM approach also models the mean and a predictive variance.

In GPM, both learning the model and making predictions require matrix inversion, which has the time complexity of \( O(n^3) \). The complexity of computing a distribution map with Kernel DM+V is generally \( O(n \cdot (D_c)^2) \), where \( n \) is the number of training samples, \( D \) is the dimension of the environment and \( c \) is the cell size. In practice, it is not necessary to evaluate the Gaussian weighting function \( N \) for all cells. Thus, the region for which the weights are computed is limited to a circle of radius \( 4\sigma \) around the measurement location. Therefore, the effective computational complexity is \( O(n \cdot (\frac{D_c}{\sigma})^2) \). The complexity of computing the maps \( a^{(k)}, r^{(k)}, \) and \( v^{(k)} \) is \( O((\frac{D_c}{\sigma})^2) \), and computing \( V^{(k)} \) requires one pass through the data \( (O(n)) \). The overall complexity is \( O(n \cdot (\frac{D_c}{\sigma})^2) \). Making predictions requires only a look-up of the values in the map.

The performance of GPM was experimentally found to be similar to the performance of Kernel DM+V (see Section 4.5). The time complexity, and also the fact that the learning procedure is simpler, give Kernel DM+V an edge over the GPM method. On the other hand, Kernel DM+V makes stronger assumptions about the distribution at a particular grid cell than GPM, which estimates the model as a mixture of distribution models.

An advantage of the straightforward interpretation of the Kernel DM+V algorithm is that it lends itself to a variety of extensions. Three such extensions are inclusion of wind information for gas distribution modelling (Kernel DM+V/W), three-dimensional modelling (3D Kernel DM+V), and modelling time-dependency which will be discussed in Chapter 5. This is not to say that extensions of the GPM approach into these domains are not possible or would perform worse, they simply have not been developed so far to the best of our knowledge.

**4.6 Discussion and Conclusion**

An important recent development was the introduction of modelling approaches to estimate the distribution variance in addition to the distribution mean. The two basic approaches, Kernel DM+V and GPM, were compared in this chapter. The modelling accuracy of the two approaches was found to be similar. However, Kernel DM+V has advantages over GPM in terms of computational complexity, a more straightforward interpretation, a simpler learning approach, and its flexibility with respect to modifications of the basic algorithm. On the other hand, Kernel DM+V makes stronger assumptions about the distribution at a particular grid cell than GPM, which estimates the model as mixture of distribution models.
The gas distribution models presented, all make Gaussian assumptions. In GPM, these assumptions are not as strong as in Kernel DM+V because the GPM approach considers a mixture of distributions. Nevertheless, it is an important direction for future work to relax the Gaussian assumptions made. Another possible research direction is to investigate sub-sampling strategies (as applied in the GPM approach) for the Kernel DM+V algorithm.

In addition to the basic approaches, two modifications of the Kernel DM+V algorithm were described in this chapter: (1) Kernel DM+V/W, which adapts the kernel shape according to wind measurements, and (2) 3D Kernel DM+V, which extends the distribution model to three dimensions using a 3D kernel. While a quantitative comparison is difficult in the case of the 3D Kernel DM+V algorithm due to calibration issues with the gas sensors, the quantitative comparison of the two-dimensional model Kernel DM+V/W with the basic version of the Kernel DM+V algorithm found an improved model quality for Kernel DM+V/W.

Considering the results and analysis of different statistical kernel methods for gas distribution modelling that are presented in this chapter, we chose Kernel DM+V as the baseline method in the research presented in this dissertation. The following chapters present research carried out to enhance gas distribution modelling by introducing time-dependency in creating gas distribution models and providing a better sampling strategy to select future sampling locations.

So far, all the presented methods made the assumption that the gas distribution is generated by a time-invariant random process. This assumption allows to average over measurements independent of the time when they were recorded. It is reasonable, however, to assume that more recent sensor readings carry more important information in order to predict future measurements. This can be modelled by introducing a decreasing importance of measurements with increasing time between measurement and prediction. Chapter 5 presents research carried out to build time-dependent gas distribution models. Moreover, it introduces a novel approach as a time-dependent extension of Kernel DM+V.

A statistical distribution model can be considered good or truthful if it explains the measurements, which were used to build the model, and predicts new observations well. To obtain a truthful representation, we need to consider a sufficient number of measurements. To quantify this requirement, we can use the NLPD to compute internal consistency (how well are training data explained) and predictive power of the model (how well are unseen measurements predicted). A related question is at which locations the next measurements should be carried out in order to obtain a good model in minimum time. Again, the NLPD can be used to compare different sensing strategies regarding their suitability for gas distribution modelling. Furthermore, the predictive variance is an important ingredient for techniques that suggest new measurement locations based on the current model. The process of selection of
future measurements, or sensor planning, is addressed in details in Chapter 6. Chapter 6 introduces a new sampling method as well. The proposed sensor planning method employs a utility function, which prioritises measurements in areas with high uncertainty, high concentration, or high variance.
Chapter 5

Time-dependent GDM

Chapter 2 and Chapter 4 discussed several statistical gas distribution modelling approaches to study the spatial distribution of gas. A crucial assumption in all presented methods is that the gas distribution is generated by a time-invariant random process. While this assumption approximately holds in some situations, it is necessary to model variations over time in order to enable applications of gas distribution modelling in a wider range of realistic scenarios.

This chapter discusses approaches to improve gas distribution modelling in realistic scenarios where environmental conditions such as gas source emission rates, as well as wind direction and intensity tend to change over time. The main contribution presented in this chapter is the introduction and analysis of two time-dependent gas distribution modelling approaches. Additional contributions include the analysis of the influence of the meta-parameters and the choice of the target time on the quality of the gas distribution models, and evaluation of the presented approaches in real-world experiments and several simulation experiments.

In the remainder of this chapter, first, the impact of temporal (sub-)sampling strategies on the quality of the resulting gas distribution models are investigated; i.e. for a given number of samples, we assess if it is better to select samples from recent measurements or distributed over all measurements. As expected, the results indicate that more recent measurements are more informative to derive an estimate of the current gas distribution. Second, as one of the main contributions of this dissertation, a gas distribution modelling approach is introduced which considers temporal importance to create gas distribution models. We include temporal importance in the form of a recency function and thus extend the state-of-the-art GDM approach, Kernel DM+V, to a time-dependent version, TD Kernel DM+V. The recency function corresponds to a time-scale that relates the age of measurements to their validity for the gas distribution model. This time-scale negotiates a compromise between two conflicting requirements to obtain accurate gas distribution models: (1)
using as many measurements as possible and (2) using only very recent measurements. In the approach proposed in this chapter, the time-scale parameter is learned from available measurements.

The problem statement and related work are presented in Section 5.1 and Section 5.2, respectively. To evaluate time-dependent GDM, we performed experiments both in simulation and real-world environments with a mobile robot and a stationary sensor network. In Section 5.3, evaluation methods and the selection of meta-parameters to create gas distribution models using the evaluation measures are discussed. Next, we describe and evaluate the two proposed time-dependent GDM methods: temporal sub-sampling in Section 5.4 and TD Kernel DM+V in Section 5.5. This is followed by a comparison of temporal sub-sampling using only most recent measurements with TD Kernel DM+V in Section 5.6. TD Kernel DM+V depends on three meta-parameters that are learned in the model selection. In Section 5.7, the impact of each meta-parameter in the model selection is analysed. In the simulation experiments, the time-dependent models are compared to time-invariant ones when the plume is evolving. In Section 5.8, we investigate the quality of the models built using TD Kernel DM+V in situations where the plume is evolving compared to fully developed, quasi-stationary plumes. Finally, we evaluate the dependence on the target time for which the gas distribution model is created in Section 5.9.

The performance of TD Kernel DM+V in the simulation experiments demonstrates that TD Kernel DM+V improves the prediction quality of the obtained gas distribution models. This represents an important step forward for statistical gas distribution modelling in real-world scenarios, especially in dynamic situations.

The results presented in this chapter are mostly derived from my publication [7].

5.1 Problem Statement

Given a set of collected observations $D = \{(x_i, r_i, t_i)\}_{i=1}^N$, statistical distribution modelling methods aim at providing a truthful representation of the observed phenomena to predict an unseen value of $r_s$ at location $x_s$ and time $t_s$ as

$$p(r_s | x_s, t_s, (x_i, r_i, t_i), 1 \leq i \leq N).$$  \hspace{1cm} (5.1)

Time-invariant distribution modelling approaches discussed in Chapter 4 assume that the underlying random process estimated does not change over time and thus the observed phenomena can be modelled without considering sampling times:

$$p(r_s | x_s, (x_i, r_i), 1 \leq i \leq N).$$  \hspace{1cm} (5.2)
5.2 Related Work

Most publications in GDM utilise only the spatial information to estimate gas distribution models (Equation 5.2 instead of Equation 5.1). A crucial assumption in these methods is that the gas distribution is generated by a time-invariant random process. This assumption allows us to average over measurements independent of their collection time. This means that the corresponding GDM approaches model a time-invariant mean (and some of them also a time-invariant variance that describes fluctuations around the mean). However, in many situations, a time-invariant random process does not allow an accurate representation of the current gas distribution. In such cases, it is reasonable to use time-dependent models that aim to model changes in the underlying random process. This temporal aspect of gas distribution modelling has been addressed only in a few recent publications.

The most related works in time-dependent GDM are [47] and [48], both published after author’s publication, [1], where the idea of using exponentially decreasing weights to create time-dependent gas distribution models was introduced.

In [47], Monroy proposed a time-dependent GDM approach applying a Gaussian Markov Random Field. Similar to TD Kernel DM+V, this method estimates gas distribution over a grid by taking into account temporal weights in addition to the Gaussian spatial weight of measurements. In [47], the proposed method is compared with Kernel DM+V in a controlled indoor environment and its corresponding simulated environment. Both qualitative and quantitative results of this comparison indicate a better performance of the time-dependent method. Note that in addition to time-dependency, the method proposed in [47], considers obstacles in creating a statistical model. Therefore, the improvements in the result may stem from both enhancements. The work in [47] does not provide information on how the meta-parameters are selected.

In [48], Marjovi et al. presented a time-dependent gas distribution mapping method. Similar to Kernel DM+V, this method estimates gas distribution over a grid. The estimated mean was calculated as the weighted sum of the measurements. Measurements were weighted using a combination of their Euclidean spatial distance and their temporal difference to the current time when the gas distribution map was created. The temporal and spatial weightings were controlled by two parameters to determine the distance limit to consider neighbouring samples. Marjovi et al. applied a time series to estimate the gas concentration at 5 time steps further in future. In this approach, meta-parameters were defined heuristically. Although in [48], the impact of different values of one of the parameters were discussed, but the parameters were not optimised when creating models. The presented results are limited to experiments that were carried out with multi robots in a small controlled indoor environment, and evaluations in more realistic scenarios are not presented.
In [1], we presented the idea of using an exponentially decreasing recency weight to incorporate temporal extrapolation to Kernel DM+V. In this work, the temporal weight was defined with a heuristic timescale factor and experiments were performed in controlled and uncontrolled indoor environments. In the work presented in [14], we introduced two different solutions to incorporate time-dependency in GDM and presented results both in the simulation and real-world experiments. The timescale factor in the simulation experiments is learned together with spatial meta-parameters when generating the gas distribution model. The content of this chapter is based on our recent article [7] which presented evaluations of our proposed time-dependent GDM in more realistic simulation experiments. Furthermore, the meta-parameter selection method used in the work presented in [14] was sensitive to the particular initialisation values of meta-parameters and got sometimes “trapped” in local minima leading to overfitting in some cases. These issues were addressed in [7].

This chapter focuses on exploring the temporal aspect of GDM and presents two solutions: (1) temporal (sub-)sampling and (2) extending Kernel DM+V to capture both spatial and temporal dispersion of gas. In addition, we discuss how the performance of these time-dependent solutions changes with prediction time.

5.3 Meta-parameter Selection and Evaluation Method

A common metric to measure the performance of a predictive model is the root mean square error (RMSE). For unseen measurements \( \{r_1, r_2, \ldots, r_n\} \) acquired at locations \( \{x_1, x_2, \ldots, x_n\} \), the RMSE is defined as \( \sqrt{\text{MSE}} \) where

\[
\text{MSE} = \frac{1}{n} \sum_{(x_i, r_i) \in \mathcal{D}} (r_i - \hat{r}(x_i))^2, \tag{5.3}
\]

and \( \hat{r}(x_i) \) is the prediction of the measurement at \( x_i \).

When it comes to gas distribution modelling, the magnitude of fluctuation is an important feature that a gas distribution model should predict. However, RMSE does not capture this aspect of gas distribution modelling. A standard criterion that takes predictive variance as a measure of the magnitude of the expected fluctuations into account is NLPD. The evaluation method, NLPD, is explained in detail in Chapter 4.

To compare different gas distribution models, we use mainly NLPD in this chapter. We also present results using the conventional model evaluation measure, RMSE.

In the experiments presented in this chapter, we selected meta-parameters using the first 80% of recorded samples as tv set, holding back the last 20% of samples as test set. The tv set was divided into training (the first 60%) and validation set (the last 40%). In order to avoid a dependency on the
optimisation routine, we use a grid search to learn the meta-parameters in this thesis.

In time-dependent GDM, we are interested in evaluating the quality of a gas distribution model for the prediction of unseen measurements at certain target times. The target time is assumed to be more recent than the samples from which the gas distribution model was created. Section 5.9 investigates different ways to set the target time.

5.4 Temporal Sub-sampling

Most of the existing gas distribution modelling approaches assume an underlying stationary random process. In this section, we stick to this assumption and investigate different temporal (sub-)sampling strategies to model slow changes in the underlying random process. We apply the different sampling strategies to the Kernel DM+V algorithm, which assumes a stationary random process. It is intuitive that sample selection from more recent measurements should capture more accurate information about the current behaviour of gas distribution. However, selecting only more recent samples also means to discard measurements, which can result in decreased model quality.

5.4.1 Evaluation and Results

In this evaluation, first, samples are sorted according to their collection time. The last 20% of measurements are selected as the test set. To create sub-sampling models, we sub-sampled X percent of the measurements from the remaining 80%. The X percent of measurements are either selected from the whole sampling interval or from the first X percent, the second, ... and the most recent X percent of collected samples. We created gas distribution models for each of these subsets and compared the models in terms of their NLPD over the test set.

We evaluated temporal sub-sampling in the real-world experiments, "Corridor" and "SmallNet", as well as, in the basic simulation experiments, "Sim-No-Obstacle" and "Sim-With-Obstacle" (see Chapter 3). When selecting the sub-sampling intervals, it is important that the datasets have a similar spatial distribution as the whole dataset, they should cover the target area fully. In "SmallNet", this constraint always holds since samples are collected using a set of stationary sensors. Considering the actual trajectories in the "Corridor" dataset (see Figure 5.1), we chose X = \{33\%, 50\%\}.

Results are shown in Table 5.1. As expected, the results indicate that sample selection from more recent measurements improves the performance of Kernel DM+V, as long as a sufficient spatial coverage is given. In the two basic simulations, the NLPD values of the models created using the first X percent of samples are positive. In the basic simulation experiments it takes at least 50\% of the sampling time for the plume to reach the end of the tunnel.
Therefore, models created using the first $X$ percent can not predict unseen measurements of a developed plume well. In addition, we observed that in the basic simulation experiments, the gas distribution models created using the most recent $X = 33\%$ results in lower NLPD values compared to the corresponding models created using $X = 50\%$ of samples. This effect is more visible in the results from the "Sim-With-Obstacle" experiment. This can be potentially explained by noiseless measurements and the presence of a narrow plume with steep concentrations at the edges in these experiments. Our investigation show that both effects lead to very small non-zero prediction mean and variance values in areas where the confidence values of the predicted model is low, which contributes to low values of NLPD. This affect is more pronounced when a smaller number of samples are available (lower measurement weights and larger areas with low confidence).
Table 5.1: NLPD comparison of models created with Kernel DM+V using different temporal sub-sampling. Random samplings experiments are repeated 10 times. In all experiments, models created using the most recent partition of data (the 2\textsuperscript{nd} partition in X = 50\% and the 3\textsuperscript{rd} partition in X = 33\%), result in lower NLPD, and thus perform better.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Sub-sampling</th>
<th>NLPD</th>
<th>Sub-sampling</th>
<th>NLPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corridor</td>
<td>1\textsuperscript{st}</td>
<td>-1.629</td>
<td>1\textsuperscript{st}</td>
<td>-1.553</td>
</tr>
<tr>
<td></td>
<td>2\textsuperscript{nd}</td>
<td>-1.837</td>
<td>2\textsuperscript{nd}</td>
<td>-1.758</td>
</tr>
<tr>
<td></td>
<td>Rand</td>
<td>-1.748 ± 0.009</td>
<td>3\textsuperscript{rd}</td>
<td>-1.800</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rand</td>
<td>-1.747 ± 0.016</td>
</tr>
<tr>
<td>SmallNet</td>
<td>1\textsuperscript{st}</td>
<td>-0.825</td>
<td>1\textsuperscript{st}</td>
<td>-0.296</td>
</tr>
<tr>
<td></td>
<td>2\textsuperscript{nd}</td>
<td>-1.983</td>
<td>2\textsuperscript{nd}</td>
<td>-1.292</td>
</tr>
<tr>
<td></td>
<td>Rand</td>
<td>-1.613 ± 0.017</td>
<td>3\textsuperscript{rd}</td>
<td>-2.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rand</td>
<td>-1.619 ± 0.010</td>
</tr>
<tr>
<td>Sim-No-Obstacle</td>
<td>1\textsuperscript{st}</td>
<td>&gt; 0</td>
<td>1\textsuperscript{st}</td>
<td>&gt; 0</td>
</tr>
<tr>
<td></td>
<td>2\textsuperscript{nd}</td>
<td>-4.488</td>
<td>2\textsuperscript{nd}</td>
<td>&gt; 0</td>
</tr>
<tr>
<td></td>
<td>Rand</td>
<td>-2.748 ± 0.013</td>
<td>3\textsuperscript{rd}</td>
<td>-4.452</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rand</td>
<td>-2.736 ± 0.041</td>
</tr>
<tr>
<td>Sim-With-Obstacle</td>
<td>1\textsuperscript{st}</td>
<td>-2.283</td>
<td>1\textsuperscript{st}</td>
<td>&gt; 0</td>
</tr>
<tr>
<td></td>
<td>2\textsuperscript{nd}</td>
<td>-5.067</td>
<td>2\textsuperscript{nd}</td>
<td>-4.820</td>
</tr>
<tr>
<td></td>
<td>Rand</td>
<td>-4.963 ± 0.071</td>
<td>3\textsuperscript{rd}</td>
<td>-6.164</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rand</td>
<td>-5.492 ± 0.518</td>
</tr>
</tbody>
</table>

5.5 TD Kernel DM+V

The TD Kernel DM+V approach extends the Kernel DM+V algorithm by considering the temporal importance of measurements through a time-dependent recency weight. The recency weight corresponds to a time-scale, which determines how quickly measurements are considered irrelevant for the gas distribution model at a given (future) target time.

TD Kernel DM+V operates on a set of samples \((r_i, x_i, t_i)\) and creates a model to estimate gas distribution at target time \(t_*\) (see Equation 5.1) – for Kernel DM+V, on the other hand, the sample times \(t_i\) are not considered. In TD Kernel DM+V, the spatial kernel is the same as in the Kernel DM+V algorithm, but the following temporal term \(\varphi\) is introduced:

\[
\varphi(t_*, t_i) = \exp^{-\beta(t_* - t_i)}
\]

where \(t_i\) is the time when measurement \(r_i\) is recorded with \(t_1 \leq t_i \leq t_*\) (i.e. we do not include future measurements into the prediction for \(t_*\)). \(t_1\) is the first
measurement time, $t_\ast - t_i$ denotes the temporal distance of the $i^{th}$ measurement from the prediction time and $\beta$ is the timescale factor that determines recency of the measurements. When building gas distribution models, the kernel as a whole now assigns a strong influence on those measurements that were collected close in space to the current grid cell $i$ and close to the target time $t_\ast$:

$$\omega^{(k)}_i = N(|x_i - x^{(k)}|, \sigma)\varphi(t_\ast, t_i). \quad (5.5)$$

$N(|x_i - x^{(k)}|, \sigma)$ is the Gaussian spatial weight function used in Kernel DM+V. $|x_i - x^{(k)}|$ is the spatial distance of measurement $i$ from the cell center $k$, and $\sigma$ is the spatial kernel width.

### 5.5.1 Experiments and Results, Basic Scenarios

We investigated the performance of TD Kernel DM+V first with simulation and real-world data from relatively simple, basic scenarios with a central source location and constant release rates. More complex scenarios are discussed in Section 5.5.2. To create gas distribution models, we used the 80% of the samples that were recorded first, and evaluated on 20% of the samples at the end of the dataset, as mentioned in Section 5.3. The models were created for the target time $t_\ast$ at which the last measurement in the test set was collected. The meta-parameters, timescale $\beta$, kernel width $\sigma$ and cell size $c$, are learned in the training phase as described in Section 5.3.

In the "Corridor" and "SmallNet" experiments, the values to search in the meta-parameter space were set to $0.20 \leq \sigma \leq 1.00 \text{ m}$ for the kernel width, and $0.05 \leq m \leq 0.50 \text{ m}$ for the cell size, both sampled with the resolution of $0.05 \text{ m}$. For the timescale factor, $0 \leq \beta \leq 3 \times 10^{-3} \text{ s}^{-1}$ with the resolution of $2.5 \times 10^{-4} \text{ s}^{-1}$ in the "Corridor" experiment, and $0 \leq \beta \leq 1.50 \times 10^{-2} \text{ s}^{-1}$ with the resolution of $5.0 \times 10^{-4} \text{ s}^{-1}$ in the "SmallNet" experiments were used. In the basic simulation experiments, "Sim-No-Obstacle" and "Sim-With-Obstacle", $0.20 \leq \sigma \leq 0.40 \text{ m}$ and $0.05 \leq m \leq 0.25 \text{ m}$, both with the resolution of $0.05 \text{ m}$, and $0 \leq \beta \leq 5.00 \times 10^{-3} \text{ s}^{-1}$ with the resolution of $0.02 \text{ s}^{-1}$ were used.

The results presented in Table 5.2 show that TD Kernel DM+V performed better than Kernel DM+V in our experiments. Corresponding predictive mean maps created using Kernel DM+V (left) and TD Kernel DM+V (right) on data from "Corridor" and "SmallNet" data are shown in Figure 5.2 and Figure 5.3, respectively. Predictive mean maps for the simulation experiments "Sim-No-Obstacle" and "Sim-With-Obstacle" are presented in Figure 5.4 and Figure 5.5, respectively. In all these figures, red indicates larger values and blue represents smaller values in the model and colour bars are used to show the scale and relative differences in the created maps in arbitrary units.

Although the maps look qualitatively similar, the predictive mean maps created using TD Kernel DM+V tend to represent the gas distribution better than Kernel DM+V mean maps. This can be seen in the NLPD comparison in
Table 5.2: Comparison of basic Kernel DM+V and TD Kernel DM+V in terms of NLPD. The NLPD comparison indicates, by smaller values of NLPD and RMSE, that TD Kernel DM+V consistently performs better than Kernel DM+V in the evaluated experiments.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>GDM</th>
<th>NLPD</th>
<th>RMSE</th>
<th>(\sigma) (m)</th>
<th>(c) (m)</th>
<th>(\beta) (s(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corridor</td>
<td>Kernel DM+V</td>
<td>-1.74</td>
<td>0.191</td>
<td>0.50</td>
<td>0.45</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>TD Kernel DM+V</td>
<td>-1.80</td>
<td>0.186</td>
<td>0.90</td>
<td>0.35</td>
<td>(1.35 \times 10^{-3})</td>
</tr>
<tr>
<td>SmallNet</td>
<td>Kernel DM+V</td>
<td>-1.62</td>
<td>0.250</td>
<td>0.40</td>
<td>0.05</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>TD Kernel DM+V</td>
<td>-2.59</td>
<td>0.149</td>
<td>0.25</td>
<td>0.10</td>
<td>(8.25 \times 10^{-3})</td>
</tr>
<tr>
<td>Sim-No-Obstacle</td>
<td>Kernel DM+V</td>
<td>-2.76</td>
<td>0.258</td>
<td>0.40</td>
<td>0.10</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>TD Kernel DM+V</td>
<td>-5.13</td>
<td>0.212</td>
<td>0.20</td>
<td>0.05</td>
<td>(3.60 \times 10^{-1})</td>
</tr>
<tr>
<td>Sim-With-Obstacle</td>
<td>Kernel DM+V</td>
<td>-2.00</td>
<td>0.251</td>
<td>0.20</td>
<td>0.15</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>TD Kernel DM+V</td>
<td>-4.93</td>
<td>0.248</td>
<td>0.20</td>
<td>0.05</td>
<td>(6.00 \times 10^{-2})</td>
</tr>
</tbody>
</table>

Figure 5.2: Predictive mean maps created for the "Corridor" experiments using (a) Kernel DM+V (\(\sigma = 0.50\)m, \(c = 0.45\)m) and (b) TD Kernel DM+V (\(\sigma = 0.90\)m, \(c = 0.35\)m, \(\beta = 1.35 \times 10^{-3}\)s\(^{-1}\)). The gas source location is shown with a white filled circle.

Table 5.2, which particularly shows a substantial improvement in created gas distribution models for simulation experiments when using TD Kernel DM+V.
Figure 5.3: Predictive mean maps created for the "SmallNet" experiments using (a) Kernel DM+V ($\sigma = 0.40\text{m}, c = 0.05\text{m}$) and (b) TD Kernel DM+V ($\sigma = 0.25\text{m}, c = 0.10\text{m}, \beta = 8.25 \times 10^{-3}\text{s}^{-1}$). The gas source location is shown with a white filled circle.

Figure 5.4: Predictive mean maps created for the "Sim-No-Obstacle" experiments using: (a) Kernel DM+V ($\sigma = 0.40\text{m}, c = 0.10\text{m}$) and (b) TD Kernel DM+V ($\sigma = 0.20\text{m}, c = 0.05\text{m}, \beta = 3.6 \times 10^{-1}\text{s}^{-1}$). The gas source location is shown with a white filled circle.

Figure 5.5: Predictive mean maps created for the "Sim-With-Obstacle" experiments using (a) Kernel DM+V ($\sigma = 0.20\text{m}, c = 0.15\text{m}$) and (b) TD Kernel DM+V ($\sigma = 0.20\text{m}, c = 0.05\text{m}, \beta = 6.0 \times 10^{-2}\text{s}^{-1}$). The gas source location is shown with a white filled circle.
5.5.2 Experiments and Results, More Complex Scenarios

In real-world scenarios, a range of complex situations and changes can occur. In order to capture some of these more complex scenarios, we performed further simulation experiments with the ROS 3D gas dispersion simulation, "Sim-ROS", and considered the following situations:

cs1/2 Two different gas source locations: We first carried out and compared experiments with the gas source at different places: once in the center of the target area, $P_A = (5.0,10.0,0.10) \text{ m}$, and once in a non-central position $P_B = (5.0,15.0,0.10) \text{ m}$. In the corresponding first two experiments the gas release rate was kept constant. While in the experiment with $P_A$ as gas source location three spots of increased gas concentration formed, upwind of the two obstacles and the outlet (see predictive mean maps in Figure 5.6), only one spot of increased gas concentration appeared upwind of an obstacle in the experiment with $P_B$ (see predictive mean maps in Figure 5.7).

cs3 Constant versus variable gas release rate: In a third experiment with the gas source at $P_A$, we increased the gas release rate in random steps depending on the current release rate (with a frequency of 0.2 Hz we chose an increment randomly from 0—100% of the number of molecules of the current release rate). On average, the release rate was thus increased exponentially. We did not use a fixed increment to avoid a negligible increase of the release rate.

cs4 Ideal versus slow sensor response model: In the real-world experiments presented in this chapter, metal oxide sensors are used. We therefore also carried out a fourth experiment in which measurements of a metal oxide sensor were simulated using a sensor model for the metal-oxide (MOX) Figaro TGS-2620 sensor. In all other experiments we used an ideal sensor model, i.e. that the sensor values reflect the concentration directly, approximating e.g. the behaviour of photoionisation detections (PIDs). The gas source was located at $P_A$ in this experiment and the gas release rate was kept constant.

Detailed information about "Sim-ROS" dataset is available in Chapter 3.

In the same way as described in Section 5.3, we created gas distribution models using the first 80% of the samples and evaluated them on the last 20% of the samples. The values to search in the meta-parameter space were $0.25 \text{ m} \leq \sigma \leq 2.50 \text{ m}$, and $0.25 \text{ m} \leq c \leq 2.00 \text{ m}$, both with the resolution of $0.25 \text{ m}$, and $0 \leq \beta \leq 0.30 \text{ s}^{-1}$ with the resolution of $0.05 \text{ s}^{-1}$.

Table 5.3 compares the results obtained with Kernel DM+V and TD Kernel DM+V in experiments cs1 – cs4. We find that TD Kernel DM+V outperforms Kernel DM+V consistently in experiments cs1 – cs4. This also indicates a
better performance of TD Kernel DM+V compared to Kernel DM+V when the physical properties of the environment do not change in the experiment.

The Figures 5.6, 5.7, 5.8, and 5.9 compare predictive mean maps from Kernel DM+V (left) and TD Kernel DM+V (right) for the experiments cs1 – cs4. In each experiment, the predictive mean maps created using Kernel DM+V and TD Kernel DM+V overall look similar. However, the TD Kernel DM+V seems to capture better details at the areas with relatively higher gas accumulation, i.e. close to the edges of the obstacles. This is more pronounced in cs1, cs3, and cs4 where the gas source is located at $P_A$.

Table 5.3: Performance comparison of basic Kernel DM+V and TD Kernel DM+V in ROS Simulation experiments. The NLPD comparison indicates, by smaller values of NLPD and RMSE, that TD Kernel DM+V consistently performs better than Kernel DM+V in the evaluated experiments.

<table>
<thead>
<tr>
<th>Exp.</th>
<th>Source, Sensor Model, Release Rate</th>
<th>GDM</th>
<th>NLPD</th>
<th>RMSE</th>
<th>$\sigma$ (m)</th>
<th>$c$ (m)</th>
<th>$\beta$ (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cs1</td>
<td>$P_A$, Ideal, Constant</td>
<td>Kernel DM+V</td>
<td>-2.97</td>
<td>0.263</td>
<td>1.75</td>
<td>0.25</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TD Kernel DM+V</td>
<td>-3.59</td>
<td>0.211</td>
<td>1.50</td>
<td>0.25</td>
<td>0.10</td>
</tr>
<tr>
<td>cs2</td>
<td>$P_B$, Ideal, Constant</td>
<td>Kernel DM+V</td>
<td>-3.50</td>
<td>0.224</td>
<td>2.00</td>
<td>1.00</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TD Kernel DM+V</td>
<td>-4.89</td>
<td>0.193</td>
<td>1.75</td>
<td>0.50</td>
<td>0.10</td>
</tr>
<tr>
<td>cs3</td>
<td>$P_A$, Ideal, Rand Increase</td>
<td>Kernel DM+V</td>
<td>-2.26</td>
<td>0.286</td>
<td>2.00</td>
<td>0.50</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TD Kernel DM+V</td>
<td>-3.33</td>
<td>0.240</td>
<td>1.50</td>
<td>0.25</td>
<td>0.10</td>
</tr>
<tr>
<td>cs4</td>
<td>$P_A$, MOX, Constant</td>
<td>Kernel DM+V</td>
<td>-1.42</td>
<td>0.309</td>
<td>1.25</td>
<td>0.25</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TD Kernel DM+V</td>
<td>-3.15</td>
<td>0.211</td>
<td>0.75</td>
<td>0.50</td>
<td>0.10</td>
</tr>
</tbody>
</table>

5.5.3 Recency Function

To define the temporal importance in TD Kernel DM+V, an exponential function was chosen as recency weight because exponential functions govern many decay processes. Intuitively, a recency function must be a monotonically increasing function, having non negative first derivation. The recency function output values should be between zero and one. In this section, we compare the choice of exponential function with two alternatives: linear and sigmoid.
Figure 5.6: Predictive mean maps created for experiments with the ROS 3D gas dispersal simulation engine – cs1: (a) Kernel DM+V ($\sigma = 1.75\text{m}, c = 0.25\text{m}$) and (b) TD Kernel DM+V ($\sigma = 1.50\text{m}, c = 0.25\text{m}, \beta = 0.10\text{s}^{-1}$). The gas source is located at $P_A$ (shown with a white filled circle) and has a constant gas release rate. The obstacles are shown with white squares.

Figure 5.7: Predictive mean created for experiments with the ROS 3D gas dispersal simulation engine – cs2: (a) Kernel DM+V ($\sigma = 2.00\text{m}, c = 1.00\text{m}$) and (b) TD Kernel DM+V ($\sigma = 1.75\text{m}, c = 0.50\text{m}, \beta = 0.10\text{s}^{-1}$). The gas source is located at $P_B$ (shown with a white filled circle) and has a constant gas release rate. The obstacles are shown with white squares.

Linear Recency Function

The linear recency weight for each measurement $t_i$ is defined as the ratio of the recency of this measurement at the prediction time $t_*$ to the age of the measurement:

$$\varphi(t_*, t_i) = 1 - \frac{t_* - t_i}{t_* - t_1} \quad (5.6)$$

where $t_1$ is the first measurement time. Note that here no timescale is introduced as meta-parameter in the linear recency function.
Figure 5.8: Predictive mean maps created for experiments with the ROS 3D gas dispersal simulation engine – cs3: (a) Kernel DM+V ($\sigma = 2.00\text{m}, c = 0.50\text{m}$) and (b) TD Kernel DM+V ($\sigma = 1.50\text{m}, c = 0.25\text{m}, \beta = 0.10\text{ s}^{-1}$). The gas source is located at $P_A$ (shown with a white filled circle) and the gas release rate is increased exponentially (on average). The obstacles are shown with white squares.

Figure 5.9: Predictive mean maps created for experiments with the ROS 3D gas dispersal simulation engine – cs4: (a) Basic Kernel DM+V ($\sigma = 1.25\text{m}, c = 0.25\text{m}$) and (b) TD Kernel DM+V ($\sigma = 0.75\text{m}, c = 0.50\text{m}, \beta = 0.10\text{ s}^{-1}$). Samples are collected using a MOX sensor model. The gas source is located at $P_A$ (shown with a white filled circle) and has a constant gas release rate. The obstacles are shown with white squares.

Sigmoid Recency Function

As an example of sigmoid function, we define the recency weight for measurement $t_i$ using a tangent hyperbolic function:

$$\varphi(t_*, t_i) = \frac{\exp{\beta(t_* - t_i)} - \exp{-\beta(t_* - t_i)}}{\exp{\beta(t_* - t_i)} + \exp{-\beta(t_* - t_i)}}$$

(5.7)

The recency weight depends on the timescale factor $\beta$ which will be selected by optimising the NLPD values when building the model.
Comparison

We compared the performance of TD Kernel DM+V using different choices of recency function in the real-world experiments and the two basic simulation experiments. Figure 5.10 presents the NLPD comparison of gas distribution models created using TD Kernel DM+V with different recency functions. The results indicate that the gas distribution models created using TD Kernel DM+V with exponential recency weights consistently perform better than the ones with linear and sigmoid recency weight functions. In the rest of this chapter, the exponential recency weight function is used to build TD Kernel DM+V models.

Figure 5.10: NLPD comparison of gas distribution models created using TD Kernel DM+V with different recency functions: linear (bars in blue), sigmoid (bars in red), and exponential (bars in orange). The exponential recency weight function consistently perform better than linear and sigmoid recency functions in both real-world and basic simulation experiments.

5.6 TD Kernel DM+V versus Temporal Sub-sampling

To compare the performance of the two presented time-dependent GDM methods, TD Kernel DM+V and temporal (sub-)sampling, we carried out experiments using both real-world ("SmallNet" and "Corridor") and basic simulation ("Sim-No-Obstacle" and "Sim-With-Obstacle") datasets. We used the exact same experimental setup as in the evaluation of temporal sub-sampling in Section 5.4.1.

The results presented in Section 5.4.1 showed that when applying the Kernel DM+V method, models created using the most recent X percent of measurements perform better than models created using X percent of samples which are randomly collected from the whole sampling interval. Section 5.5 showed that TD Kernel DM+V performs better than Kernel DM+V when
both are given the exact same samples. In this section, we compare the performance of the two approaches – TD Kernel DM+V and temporal (sub-)sampling by (1) collecting the most recent $X$ percent of samples to create a model using Kernel DM+V, and (2) collecting randomly $X$ percent of samples from the whole sampling interval to create a model using TD Kernel DM+V.

Table 5.4 presents an NLPD comparison of models created applying Kernel DM+V on only the most recent samples to models created applying TD Kernel DM+V on samples randomly selected from the whole sampling interval. For the same sample size, in "Corridor" and "Sim-No-Obstacle" experiments models created using the most recent samples perform better, while, in the rest of experiments, TD Kernel DM+V performs better. Further investigations are needed to identify under which gas dispersion parameters and environmental conditions sub-sampling outperforms recency weighting.

<table>
<thead>
<tr>
<th>Exp.</th>
<th>GDM</th>
<th>Subsampling</th>
<th>NLPD</th>
<th>Subsampling</th>
<th>NLPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corridor</td>
<td>Kernel DM+V</td>
<td>2nd</td>
<td>-1.837</td>
<td>3rd</td>
<td>-1.800</td>
</tr>
<tr>
<td></td>
<td>TD Kernel DM+V</td>
<td>Rand</td>
<td>-1.821±0.051</td>
<td>Rand</td>
<td>-1.797±0.011</td>
</tr>
<tr>
<td>SmallNet</td>
<td>Kernel DM+V</td>
<td>2nd</td>
<td>-1.983</td>
<td>3rd</td>
<td>-2.001</td>
</tr>
<tr>
<td></td>
<td>TD Kernel DM+V</td>
<td>Rand</td>
<td>-2.594±0.073</td>
<td>Rand</td>
<td>-2.490±0.019</td>
</tr>
<tr>
<td>Sim-No-Obstacle</td>
<td>Kernel DM+V</td>
<td>2nd</td>
<td>-4.488</td>
<td>3rd</td>
<td>-4.452</td>
</tr>
<tr>
<td></td>
<td>TD Kernel DM+V</td>
<td>Rand</td>
<td>-4.077±0.096</td>
<td>Rand</td>
<td>-3.900±0.080</td>
</tr>
<tr>
<td>Sim-With-Obstacle</td>
<td>Kernel DM+V</td>
<td>2nd</td>
<td>-5.067</td>
<td>3rd</td>
<td>-6.164</td>
</tr>
<tr>
<td></td>
<td>TD Kernel DM+V</td>
<td>Rand</td>
<td>-6.072±0.873</td>
<td>Rand</td>
<td>-7.520±0.862</td>
</tr>
</tbody>
</table>

### 5.7 Model Selection and Meta-parameters

The meta-parameters of the time-dependent GDM method TD Kernel DM+V, kernel width $\sigma$, cell size $c$, and timescale factor $\beta$, are selected by minimising the NLPD value in the training phase. In this section, the impact of these
meta-parameters on the quality of a model are discussed. Figure 5.11–5.13 present two-dimensional graphs of pairs of meta-parameters coloured with the corresponding NLPD values: $c - \sigma$ on the left, $c - \beta$ in the middle, and $\sigma - \beta$ on the right. In these figures, colour bars are used to show the scale and relative differences in the NLPD values. Dark blue indicates lower (better) NLPD values and yellow indicates larger NLPD values. For a better distinction of areas with lower NLPD values, in the visualisations, the NLPD values larger than 10.0 are treated as NLPD = 10. Areas for which no model has been created (namely where $c \leq \sigma$) are coloured in white.

The results indicate that the choice of the cell size is often not critical, while the choice of the timescale is. The selection of the kernel width also seems to be quite critical. Selection of smaller cell sizes increases the computational complexity but also allows creating maps with finer resolution. The sparsity of the available samples in real-world experiments can explain the larger optimal values of kernel width in the "Corridor" and "SmallNet" experiments (see Figure 5.11) when optimising NLPD compared to its optimal values in the basic simulations (see Figure 5.12). In addition, in the basic simulations (see Figure 5.12), we observe smaller values of kernel width compared to the ones in a model created for more complex simulation experiments (see Figure 5.13). The small kernel widths in the basic simulations can potentially be explained by the density of the samples and the presence of a narrow plume with steep concentrations at the edges in those experiments.

Figure 5.11: NLPD color maps from left to right for $c - \sigma$, $c - \beta$, and $\sigma - \beta$ in the real-world experiments: "Corridor" (top) and "SmallNet" (bottom) experiments. The colors indicate NLPD values with blue as more negative and yellow as more positive values of NLPD.
The cell size determines the resolution at which different predictions can be made [9]. The lower sensitivity to the precise value of \( c \) enables fast, practical optimisation schemes in which first \( \beta \) and \( \sigma \) are found at a course resolution (with a large \( c \)) and then refined at a finer resolution (at a smaller value of \( c \)). In further experiments, we therefore examined optimising \( \sigma \) and, for TD Kernel DM+V, \( \beta \) given a constant cell size. In these further experiments, we see a similar trend with a superior performance of TD Kernel DM+V in comparison to basic Kernel DM+V.

In an experiment, we investigated the sensitivity of meta-parameter optimisation to the cell size. Table 5.5 presents the NLPD and RMSE comparison of TD Kernel DM+V and Kernel DM+V in the real-world and basic simulation experiments using constant cell size \( c = \{0.05\text{m}, 0.10\text{m}, 0.15\text{m}, 0.25\text{m}\} \). In these experiments, the kernel width and timescale factor were selected by optimising the NLPD values for each constant cell size. The results presented in Table 5.5 show a better performance of TD Kernel DM+V. In addition, the results indicate low sensitivity of the meta-parameter selections on the cell size; i.e. for different values of constant cell size, the same values of kernel width (\( \sigma \)) and roughly the same timescale factor (\( \beta \)) are selected when optimising NLPD values to build gas distribution models.

In another experiment, we compared the performance of TD Kernel DM+V with Kernel DM+V in the more complex simulation experiments, "Sim-ROS" using constant cell size \( c = 0.10\text{m} \). The kernel width \( \sigma \) and for the TD Kernel
Figure 5.13: NLPD color maps from left to right for $c - \sigma$, $c - \beta$, and $\sigma - \beta$ in the more complex simulation experiments: $cs_1$, $cs_2$, $cs_3$, and $cs_4$ from top to the bottom. The colors indicate NLPD values with blue as more negative and yellow as more positive values of NLPD.
Table 5.5: NLPD comparison of TD Kernel DM+V and Kernel DM+V in real-world and basic simulation experiments. All models are built with a constant cell size $c = \{0.05 \text{ m}, 0.10 \text{ m}, 0.15 \text{ m}, 0.25 \text{ m}\}$; kernel width and timescales are optimised. The NLPD comparison indicates that by smaller values of NLPD and RMSE, TD Kernel DM+V consistently performs better than Kernel DM+V in the evaluated experiments.

<table>
<thead>
<tr>
<th>Exp.</th>
<th>c (m)</th>
<th>Kernel DM+V</th>
<th>TD Kernel DM+V</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>σ (m)</td>
<td>NLPD</td>
<td>RMSE</td>
</tr>
<tr>
<td>Corridor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>1.00</td>
<td>-1.74</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>1.00</td>
<td>-1.74</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>1.00</td>
<td>-1.74</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>1.00</td>
<td>-1.74</td>
</tr>
<tr>
<td>SmallNet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.40</td>
<td>-1.62</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.40</td>
<td>-1.62</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.40</td>
<td>-1.83</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.40</td>
<td>-1.88</td>
</tr>
<tr>
<td>Sim-No-Obstacle</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.40</td>
<td>-2.76</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.40</td>
<td>-2.76</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.40</td>
<td>-2.73</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.30</td>
<td>-2.70</td>
</tr>
<tr>
<td>Sim-With-Obstacle</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.20</td>
<td>-4.46</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.20</td>
<td>-3.42</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.20</td>
<td>-2.00</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.30</td>
<td>-3.46</td>
</tr>
</tbody>
</table>
5.8. EXPERIMENTS WITH FULLY DEVELOPED PLUMES

DM+V, timescale factor $\beta$, were selected by optimising the NLPD value. The results are presented in Table 5.6. Same as in the real-world and basic simulation experiments, we see consistently a better performance of TD Kernel DM+V in terms of NLPD and RMSE compared to Kernel DM+V.

Table 5.6: NLPD comparison of TD Kernel DM+V and Kernel DM+V in the complex simulation experiments, "Sim-ROS". All models are built with a constant cell size $c = 0.10$ m; kernel width and timescale are optimised. The NLPD comparison indicates that by smaller values of NLPD and RMSE, TD Kernel DM+V consistently performs better than Kernel DM+V in the evaluated experiments.

| Exp. | $c$ (m) | Kernel DM+V | TD Kernel DM+V | | | |
|------|--------|-------------|----------------|----|----------------|
|      |        | $\sigma$ (m) | NLPD | RMSE | $\beta$ (s$^{-1}$) | NLPD | RMSE |
| cs1  | 0.10   | 1.75        | -3.04 | 0.263 | 1.50 | 0.10 | -3.68 | 0.210 |
| c2   | 0.10   | 2.50        | -3.33 | 0.227 | 1.75 | 0.10 | -4.89 | 0.193 |
| c3   | 0.10   | 2.00        | -2.26 | 0.286 | 1.75 | 0.10 | -3.17 | 0.245 |
| c4   | 0.10   | 1.50        | -1.54 | 0.290 | 0.75 | 0.10 | -3.29 | 0.196 |

5.8 Experiments with Fully Developed Plumes

In the "Corridor" and "SmallNet" experiments, sampling started after the gas was released for a while and the plume was already fully developed. This is similar to the real-world scenarios considered in this thesis. In the simulation experiments, since we were interested in creating a more dynamic scenario where stronger fluctuations and plume development are present, we considered the measurements from the moment gas emission started. In this setup, as shown in Section 5.5, models created using TD Kernel DM+V performed substantially better than models created using Kernel DM+V in predicting unseen measurements. In this section we assess whether or not TD Kernel DM+V performs consistently better in simulation in situations where the plume is already fully developed (reached the end of the tunnel) when sampling collection started.

We created models using the 3D simulation data and compared models created using

- Exp. (1): Samples collected at $t = 1 - 80$s at which the plume is expected not to be fully developed. We evaluated models on samples collected at $t = 81 - 100$s (the target time was set to $t_* = 100$s).

- Exp. (2): Samples collected at $t = 101 - 180$s at which the plume is expected to be fully developed. We evaluated models on samples collected at $t = 181 - 200$s (the target time was set to $t_* = 200$s).
Figure 5.14 presents the NLPD comparison of the models created using TD Kernel DM+V to the ones created using Kernel DM+V in Exp. (1) on the left and in Exp. (2) on the right. The lower values of NLPD for TD Kernel DM+V show that TD Kernel DM+V consistently performs better than Kernel DM+V in both cases when the plume is developing, and when the plume is fully developed.

Figure 5.14: NLPD comparison of TD Kernel DM+V (bars in red) and Kernel DM+V (bars in blue) in (a) Exp. (1) with $1 \leq t \leq 100$ and (b) Exp. (2) with $101 \leq t \leq 200$. In both experiments, TD Kernel DM+V performs better than Kernel DM+V. cs1–cs4 on the x-axis correspond to experiments cs1–cs4 described in Section 3.2.2.

5.9 Dependence on Target Time

The results in Section 5.5.1 and Section 5.5.2 indicate that the introduction of a recency weight in TD Kernel DM+V improves the performance of Kernel DM+V. In the experiments presented, the gas distribution models were computed for a single target time – the maximum sampling time in the test set. In this section, we investigate the performance of TD Kernel DM+V for different target times.

We performed two experiments: (1) a comparison of gas distribution models on test sets sampled at increasing times in the future, and (2) optimising the meta-parameters of TD Kernel DM+V for the time at which samples in each test set were collected.

In the first experiment, we used the meta-parameters shown in Table 5.2 for the "Sim-No-Obstacle" and "Sim-With-Obstacle" dataset and computed gas distribution models for the maximum time ($t_* = 20s$) in the test set. We then evaluated the models on test sets that include only samples recorded at one particular time in the future. The results of experiment (1) presented in
5.9. DEPENDENCE ON TARGET TIME

Table 5.7 show that for a given model, it gets harder to predict measurements that are further in the future.

This trend is expected and much stronger in the "Sim-With-Obstacle" experiment, most likely since in this situation it takes longer for the plume to fully develop and a stationary situation is not reached after \( t = 16s \).

In Section 5.5.1, the test set for simulation data was collected from data between \( t = 17s \) and \( t = 20s \), and the target time was set to \( t_\ast = 20s \) – the maximum recording time in the test set. To analyse the impact of change in target time, in the second experiment, we created models for target time \( t_\ast = \{17s,18s,19s,20s\} \). The results are also presented in Table 5.7. The last row represents the same setup and therefore shows the same results as the evaluation of TD Kernel DM+V in Section 5.5.1.

Table 5.7: Performance of TD Kernel DM+V gas distribution models (created from data between \( t = 1s \) and \( t = 16s \) for different target times in the future) for test sets sampled from the simulation at increasing times in the future.

<table>
<thead>
<tr>
<th>Exp.</th>
<th>Test set from ( t_{test}(s) )</th>
<th>( t_\ast ) for training (s)</th>
<th>Sim-No-Obstacle</th>
<th>Sim-With-Obstacle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NLPD</td>
<td>( \sigma ) (m)</td>
<td>c (m)</td>
</tr>
<tr>
<td>(1)</td>
<td>17</td>
<td>20</td>
<td>-6.05</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>20</td>
<td>-6.14</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>20</td>
<td>-5.68</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>20</td>
<td>-5.22</td>
<td>0.20</td>
</tr>
<tr>
<td>(2)</td>
<td>17-20</td>
<td>17</td>
<td>-6.01</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>17-20</td>
<td>18</td>
<td>-5.65</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>17-20</td>
<td>19</td>
<td>-5.40</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>17-20</td>
<td>20</td>
<td>-5.13</td>
<td>0.20</td>
</tr>
</tbody>
</table>

In the evaluation of the second experiment, we again considered test sets sampled at a single time from the "Sim-No-Obstacle" and "Sim-With-Obstacle" datasets. However, we now (re-)computed the models (including learning of the meta-parameters) for the exact target time of each test set. Figure 5.15 shows the NLPD values of the different test sets (left) and the corresponding learned timescale values \( \beta \) (right). We observe a monotonic decrease of the parameter \( \beta \) corresponding to an increase in the time-scale. The models thus increasingly rely on all (not just the most recent) training data to predict increasingly future measurements. Especially in the case of the "Sim-No-Obstacle" data, this preserved the quality of the gas distribution model.
Figure 5.15: Performance of TD Kernel DM+V when test sets and target times vary over time. The two datasets used are "Sim-No-Obstacle" (blue line with filled square marks) and "Sim-With-Obstacle" (red line with filled circle marks). (a) NLPD value for increasing target and test time. (b) The timescale parameter $\beta$ learned for each target time.

5.10 Summary, Conclusions and Future Work

In dynamic situations a time-invariant random process often can not describe an observed gas distribution well: examples include an evolving gas plume or major changes in gas dispersion due to a sudden change of the environmental conditions. This chapter discussed approaches to gas distribution modelling, which introduce a time-dependency and a relation to a timescale in generating gas distribution models either by sub-sampling or by introducing a recency weight that relates measurement and prediction time.

First, we investigated the impact of sub-sampling strategies on the quality of the resulting gas distribution models. As expected, the results in Section 5.4 indicate that more recent measurements are more informative to derive an estimate of the current gas distribution.

Second, we presented an approach which considers temporal importance to create gas distribution models. We include temporal importance in the form of a recency function to extend the state-of-the-art gas distribution modelling approach, Kernel DM+V, to a time-dependent version, TD Kernel DM+V. The recency function corresponds to a time-scale that relates the age of measurements to their validity for the gas distribution model. To define the temporal importance in TD Kernel DM+V, we compared the performance of TD Kernel DM+V using different recency functions in this chapter. The results indicated a better performance of the exponential function as recency function. Therefore, an exponential function was chosen as recency weight in the rest of analysis presented in this chapter.

We compared TD Kernel DM+V to the established Kernel DM+V algorithm in Section 5.5 using a number of real-world and simulated gas distri-
bution data and found a consistent improvement in the prediction of unseen measurements, expressed in terms of the NLPD (negative log predictive density) and RMSE (root mean square error). While this may be a consequence of an additional meta-parameter of TD Kernel DM+V, we find a substantial improvement especially in dynamic scenarios.

In Section 5.6, we compared Kernel DM+V with temporal (sub-)sampling and TD Kernel DM+V to create gas distribution models. For the same number of samples, the results indicate a better performance of gas distribution modelling in some of the experiments when using only the most recent samples, while in other experiments, models created using TD Kernel DM+V on samples selected from the whole sampling time interval outperformed models created using only the most recent samples. Future work will include further investigations on gas dispersion parameters and environmental conditions under which sub-sampling outperforms recency weighting.

TD Kernel DM+V depends on three meta-parameters; kernel width $\sigma$, cell size $c$, and timescale factor $\beta$. The analysis presented in Section 5.7 indicate that the critical meta-parameters are $\sigma$ and $\beta$ when optimising the NLPD, while cell size $c$, which decides the resolution at which predictions are made, is less critical. We carried out experiments where models were created given a constant cell size. The results indicate a consistent improvement in the quality of GDM in TD Kernel DM+V compared to Kernel DM+V.

In Section 5.8, we assessed the performance TD Kernel DM+V compared to Kernel DM+V in the conditions where the gas plume is evolving and where the gas plume has fully developed. In these experiments, TD Kernel DM+V consistently outperformed Kernel DM+V in both conditions.

TD Kernel DM+V requires to set a target time when computing a gas distribution model. In Section 5.9, we investigated the performance of TD Kernel DM+V for different target times. The results indicate that it may not pay out to compute models with the precise target times of the data in the test set. We even observed that the models were better when we used the maximum sampling time in the test set for all test data. Therefore, in the rest of experiments presented in this chapter, we used the maximum sampling time in the test set.

In another experiment in Section 5.9, for a given training dataset, we created models to predict gas distribution for test datasets collected at different times. We observed that for the prediction times further in future, the timescale factor $\beta$ increases when creating gas distribution models. This means that the models increasingly rely on all (not just the most recent) training data to predict increasingly future measurements. Future would should consider to analyse a larger set of data in order to understand this behaviour better. Studying information that may become visible in difference maps computed from a sequence of TD Kernel DM+V maps that are derived from consecutive intervals in time is an area for future work.
Chapter 6

Multi Objective Sensor Planning

Up to this point, we have discussed statistical gas distribution modelling. In real-world applications, to obtain a truthful representation of the gas distribution, one has to cope with resource constraints and sensing limitations of gas sensors. Thus, it is crucial to employ an efficient strategy which meets these constraints to plan sampling locations. In the mobile robot olfaction application domain, we are interested in sampling strategies which can enable providing accurate gas distribution models given a small number of samples in a limited time span by selecting the most informative locations to acquire next samples.

There are a number of methods for sensor planning in the environmental monitoring literature, but only a few are suitable for gas distribution modelling and gas source localisation. However, sensor planning for gas distribution modelling can be inspired by related work in other application domains. In the general context of environmental monitoring, sensor planning methods rely on the same principles, i.e. to gain information about an observed phenomenon in the environment. Existing solutions mainly deal with sensor deployment [49, 2, 50], whereas are interested in adaptive sampling. Furthermore, many works are tailored to sensor networks (mostly, stationary ones), whereas we deal with a single mobile sensor. However, it is possible to generalise a developed sensor planning method for a single mobile sensor to be applied to a network of sensors in which there is limited communication between the nodes.

This chapter addresses the sensor planning problem in the case of gas distribution modelling, which is the focus of this dissertation. In this chapter, we present a new sensor planning method which introduces a utility function based on the created gas distribution model and the spatial distribution of previously recorded sampling locations. The proposed function utilises a trade-off between exploration and exploitation considering the nature of the application, but it is not limited to a specific application. In this line of research, we introduced the utility function by a trade-off between exploration
in areas where the certainty of the created model is low and exploitation in areas with high values of predictive mean and variance of the gas distribution. High values of predictive mean and variance can indicate areas with high gas accumulation and potential gas source locations, respectively. This trade-off between exploration and exploitation can be mapped to a potential field where pseudo electric charges are placed at locations where samples have been previously recorded. On the other hand, distribution maps, predictive mean and variance act as potential fields created with pseudo electric charges with opposite charge sign. In this chapter, the proposed sensor planning method is described, and its evaluation in both simulation and real-world experiments is presented. Simulation experiments are performed in the ROS simulated environment using a MOX sensor model for the task of gas distribution modelling (see "Sim-ROS" cs4 in Chapter 3). Real-world experiments in this chapter include the "QC-AASS" experiment carried out with a micro-done for the task of gas distribution mapping and gas source localisation.

The results presented in this chapter are partially derived from my collaboration with Patrick Neumann from Federal Institute for Materials Research and Testing (BAM) [51] and are published in [12], [8], and [16]. I declare my contribution in Chapter 1.

6.1 Problem Statement

As mentioned earlier in this chapter, proper selection of sampling locations is important in gas distribution modelling. In fact, we are interested in a sampling strategy to collect observations in the environment in order to obtain a truthful representation of an observed phenomenon in the environment.

In Chapters 2 and 5, formal definition of statistical gas distribution modelling was presented. In this chapter, we recall the definition of statistical gas distribution modelling and adapt it to the task of sensor planning. We denote the environment $\mathcal{A}$, by $\mathcal{A} = (X_A, R_A, T_A)$ where $X_A$ is the set of locations in the environment, and $R_A$ and $T_A$ are the sets of corresponding measurement values and times, respectively. In this environment, a given set of collected observations is denoted by $D = (X, R, T)$ where $D \subseteq \mathcal{A}$. Distribution modelling aims to provide a truthful representation of the observed phenomenon in environment $\mathcal{A}$ to predict the observation $r_*$ at the location $x_*$ and time $t_*$ from the set of available observations from the environment as

$$\hat{r} = p(r_*|x_*, t_*, \{x_i, r_i, t_i\} \in D, 1 \leq i \leq |D|).$$  \hfill (6.1)

Time-invariant gas distribution modelling approaches assume that gas distribution can be modelled as time-constant random process and the average behaviour of the observed phenomenon is modelled by simplifying Equation 6.1 to

$$\hat{r} = p(r_*|x_*, x_i \in X, r_i \in R, 1 \leq i \leq |D|).$$  \hfill (6.2)
Sensor planning is the task of selecting locations $X_S$ in the sequence of $T_S$ to acquire next observations. The sensor planning task can be formalised as to find a subset $S$, where

$$S = \{(x_s, t_s) | (x_i, r_i, t_i) \in A, t_s > t_i, \forall t_i \in T\}. \quad (6.3)$$

If $|S| = 1$, the sensor planning selects the sampling location for only the immediately following step ahead. In this case, sensor planning is called "myopic"; otherwise, it is "non-myopic" [52].

If sampling locations are selected while the observations are collected, the sensor planning is called "on-line", otherwise it is "off-line" [53]. A common off-line sensor planning approach is to sample in sweep trajectories. A straightforward on-line sampling method is random sampling, which is not efficient for environmental monitoring applications.

In monitoring problems, $|D|$ is much smaller than $|A|$ due to sensing and resource limitations. To consider resource and sensing constraints, an alternative to predefined sampling strategies is to adaptively select sampling locations using information about previous sampling locations and measurement values.

Then, the challenge that sensor planning is facing is to select the most informative sampling locations with respect to a target or criteria. Sensor planning can be defined as an optimisation problem by formulating the criteria in a utility function which needs to be maximised. The utility function is dependent on the application domain, sensing characteristics, and resource constraints. An example of a criterion used in a utility function is the quality of the gas distribution model. In the literature, utility functions are also discussed as objective functions, target functions, criteria to meet, and policies [2].

If sensor locations are selected off-line and sensors are deployed in the environment prior to sampling, sensor planning is called sensor deployment. Sensor deployment is usually explored in stationary sensor networks and typically addresses the question of how to place the least number of sensors in a field to achieve desired coverage and communication connectivity properties. In the research presented in this dissertation, we are not dealing with sensor deployment. We are applying sensor planning on a mobile sensor which can be extended to a mobile sensor network. This chapter introduces an on-line myopic adaptive sampling method in which the utility function includes multiple criteria. The application domains are gas distribution modelling and gas source localisation. Specially, we are interested in applying this sensor planning method in uncontrolled environments.

### 6.2 Related Work

In general, sensor planning is the task of determining from the current situation where to acquire the next sample. Sensor planning is particularly relevant for environmental monitoring tasks as there are limited resources and sparse
measures available, and it is desired to guide sampling as efficiently as possible in order to extract the most information about the environment with given resources. In this section, a review of existing related work in sensor planning is presented. The methods discussed in this section are not limited to the ones previously presented in the mobile robot olfaction community for two reasons: first, there are only a few sampling methods available in this domain and second, the proposed sensor planning method in this dissertation is generic and not limited to only gas distribution modelling application.

6.2.1 Sensor Planning Methods in Environmental Monitoring

Sensor planning is an active field. The choice of sensor planning method is highly dependent on the application targets (e.g. rescue, monitoring, leakage detection), sensor limitations, sensing mobility, sensing communication properties and resource constraints such as time and power. The core of sensor planning is in the definition of the target function one wishes to employ. A target function is the utility function which determines the criteria for the strategy to select the next sampling location. Often the optimisation is defined as maximising this utility function.

To explore the area with a mobile sensor, where prior information about distribution and environment is not available, a straightforward sampling strategy is random walking; sampling along a predefined weeping trajectory is another simple alternative. In [8], Neumann et al. applied sweeping on a predefined trajectory as sampling method for gas distribution modelling in an experiment in an geochemically active area in the Tuscany region, Italy. In [31] and [30], predefined sweeping and spiral paths were used to collect samples in the study of gas distribution.

In environmental monitoring problems, in one hand we do not have prior knowledge about the environment and gas source location; on the other hand, our sampling is constrained by battery capacity. Moreover, in hazardous incidents in which gas emission is involved, time limitation is a hard constraint. Therefore, it is crucial to select the best subset of measurement locations which meets constraints and gives a truthful knowledge about the environment. This selection problem is computationally expensive [54]. As a simple method with a prior of the area, in the chapter, we consider a randomised sweeping (i.e. selecting the next sampling location randomly from a planned trajectory) for comparison in our evaluations.

Most of the previous work on informative path planning have either dealt with non-adaptive approximation algorithms [55], which plan and commit to the path before any observations are made, or with adaptive (often myopic) heuristics, which update and re-plan as new information is collected [56].

There are planning techniques that can compute optimal plans, fore example, by maintaining a belief over possible states of the world and by computing the strategy that is optimal with respect to that belief. One solution to this is
the Partially Observable Markov Decision Process, also known as POMDP [57]. The major disadvantage of the POMDP is that it is only practical in applications with a small number of states [56].

For efficiently observing these environments, Singh et al. [54] extended a non-myopic approach to perform adaptive informative path planning by addressing a trade-off between exploration (collecting information about the environment) and exploitation (using the current belief about the state of the environment most effectively). In [54], Singh et al. introduced a near-optimal algorithm for non-adaptive path planning grounded on the submodularity property. The submodularity property is as an intuitive diminishing returns property described as “adding a sensor to a small deployment helps more than adding it to a large deployment” [49, 2]. Based on an iterative Bayesian updating and replanning approach, Singh et al. extended this method to a Non-myopic Adaptive Informative path planning (NAIVE) algorithm for a single mobile robot [54]. NAIVE selects sampling locations in a trade-off between exploration and exploitation. The trade-off is tuned by an adaptivity gap. Adaptivity gap is the ratio of the performance of the optimal policy divided by the performance of the best non-adaptive path. Similarly, in this chapter, we propose an adaptive sensor planning method that considers a trade-off between exploration and exploitation. The presented work in this dissertation is, however, myopic and plans to collect samples at each step based on the collected observations from the environment and its defined objectives.

6.2.2 Utility Functions Used in Sensor Planning Methods

Having a limited number of sensors, and time and power constraints to monitor spatial phenomena, it is critical to plan sampling locations that provide the most informative samples to build an accurate model of the environment. Several publications on spatial monitoring address this problem. The selection of sampling points is typically based on different criteria such as the expected accuracy of estimated measurements and resource constraints, e.g., mobility, time, power, and wireless communication. This section explains some examples of utility functions used in previous works for the task of sensor planning.

A variety of approaches for sensor placement have been applied to environmental monitoring. Guestrin et al. [49] proposed Mutual Information (MI) as a criterion for sensor placement where Gaussian Processes (GPs) were used to model the observed spatial phenomena using collected samples. Guestrin et al. showed that finding a configuration of sensors which optimises MI is NP-complete. Thus, they used a polynomial-time approximation algorithm by using the submodularity property of MI function. Although their work addresses selection of sensor placement configuration in the sensor deployment problem, it is a base for online, adaptive algorithms for sensor planning as in [55]. Singh et al. [55] applied this method to monitor the ecological condition of a river using single and multiple mobile sensors and to plan the path for mobile sensors.
To monitor spatial phenomena with a sensor network, Krause et al. [58] applied GPs and used MI between explored and unexplored areas to optimise sensor placements. Based on temperature and precipitation datasets, Krause et al. showed that maximising MI between explored and unexplored areas performs better (in terms of prediction accuracy) than entropy or geometrical criteria [58].

The closest work to the research presented in this chapter is the work of Marchant et al. [59] that introduced an adaptive sampling method in the environmental monitoring domain. In this approach, the next sampling location was selected by introducing a utility function called acquisition function. This function aims to improve the predicted model, directs attention to spots with high predicted values, and decreases the total travelled distance by the mobile robot. The predictive model is created using GPs. To select the next sampling location, Marchant et al. [59] introduced the upper confidence bound discrete (UCBD) acquisition function in which mean, variance of the predicted model, and distance of measurements were considered. Details on parameter selections for UCBD and generated models are not discussed in [59].

The sensor planning approach presented in this chapter is an adaptive multi-criteria sensor planning technique that accomplishes a trade-off between exploration and exploitation by using an Artificial Potential Field (APF) based approach. In the proposed approach, the utility function combines three criteria to select the next sampling locations: maximising the coverage area by using information about previous sampling locations to explore areas where there is high uncertainty, targeting areas with high predictive mean and variance in current statistical gas distribution model. The gas distribution models are created using the Kernel DM+V method explained in Chapter 4. We compare this approach to random sampling, sweeping in a predefined trajectory, and randomised sweeping. The performance of the sensor planning is assessed using different measures such as quality of the predicted model and the travelling distance.

6.2.3 Artificial Potential Field based Path Planning

APF methods are used in a number of robotic applications including local navigation and obstacle avoidance. APF approaches have also been used for the spatial formation of a set of sensors, spatial monitoring, and coverage problems [60, 61].

Bamberger et al. [62] developed a stigmergic potential field based approach to coordinate movements, transient acts such as speed and direction, and task allocation among cooperating Unmanned Aerial Vehicles (UAVs) for coverage in searching a target in the given area. Sensors communicate their beliefs from their observations. In [62], the utility functions are not explained in detail, and quantitative evaluation are not presented. As an example, Bamberger et al. presented a simulated experiment with a single UAV were a gas release area
was given as a prior to the sensor and the UAV was deployed to search and characterise the spatial extent of the plume.

Schwager et al. [61] treated the multi-robot coverage problem as an optimisation problem and combined different existing strategies for deploying groups of robots in an environment including artificial potential fields, Voronoi-based coverage and probabilistic minimum variance objectives.

To direct sampling locations in gas distribution modelling, this thesis proposes an APF-based approach that combines the repulsive and attractive forces on a single mobile sensor. The attractive term directs a sensor toward areas with high values of predictive mean and variance, and the repulsive term applies a repulsion force to previous sampling locations. Samples are collected consecutively, and the APF changes at each iteration based on the measurement values and locations; therefore, there is no unique equilibrium state to which the mobile sensor tends to achieve in contrast to the method introduced in [60]. The gas distribution modelling approach used in the work presented in this chapter does not use any prior knowledge about the physical properties of the environment such as gas source location and presence of obstacles.

6.3 Method

6.3.1 Artificial Potential Field based Sensor Planning

To derive a truthful representation of gas distribution using mobile robots with limited power, I proposed an adaptive sampling strategy for sensor planning: Artificial Potential Field based Sensor Planning (APFSP). In collaboration with Patrick Neumann [51], we explored the potential of this approach to direct micro-drone sampling in real-world experiments.

The initial idea of APFSP was presented in [12]. APFSP uses information about the target area, previous sampling locations, and the current statistical gas distribution model. It combines several objectives in an artificial potential field approach. In particular, three objectives direct the sensor toward areas of (1) high predictive mean and (2) high predictive variance, while (3) maximising the coverage area. In this approach, we first discretise the area into a grid and estimate gas distribution mean and variance at each grid cell. The first two objectives implement exploitation of the information in the gas distribution model. They are realised with an attractive potential generated by charges placed in each grid cell center k. The strength of these charges is determined by the corresponding predictive mean and variance. Accordingly, two APF contributions, \( \text{APF}_M^{(k)} \) and \( \text{APF}_V^{(k)} \), are computed for each cell k as

\[
\text{APF}_M^{(k)} = \hat{r}^{(k)} \tag{6.4}
\]

and

\[
\text{APF}_V^{(k)} = \hat{v}^{(k)}, \tag{6.5}
\]
where \( \hat{r}^{(k)} \) and \( \hat{v}^{(k)} \) are predictive mean and variance values at cell center \( k \), respectively. The third objective corresponds to exploration. This objective is implemented by a repulsive potential which is generated by placing charges at each cell center \( k \) resulting in the third APF contribution \( \text{APF}_{C}^{(k)} \) as

\[
\text{APF}_{C}^{(k)} = \hat{\alpha}^{(k)} \tag{6.6}
\]

where \( \alpha^{(k)} \) is the confidence values for the cell center \( k \) computed by gas distribution modelling method Kernel DM+V.

In [12] and [8], where we first presented the idea of APF-based sensor planning, I introduced the repulsive potential by placing charges at all previous measurement locations \( X \), resulting in a third APF contribution \( \text{APF}_{R}^{(k)} \) as

\[
\text{APF}_{R}^{(k)} = q \sum_{i=1}^{|D|} \exp \left( -\frac{\|x_i - x^{(k)}\|_2}{d\sigma} \right) . \tag{6.7}
\]

In Equation 6.7, \( x_i \in X \) is the location of the recorded sample \( i \) and \( x^{(k)} \) is the location of cell center \( k \). The scaling distance factor, \( d\sigma \), is set to 1 in the implementation presented in [8]. Furthermore, in this implementation, the same repulsive force is assigned to all previous measurements, hence \( q = -1 \).

Performance analysis of different experiments indicated that using the confidence map information and directing sampling locations towards areas with low certainty in the predicted model as exploration factor is a better choice compared to \( \text{APF}_{R}^{(k)} \). Therefore, we define the utility function using the APF contributors, \( \text{APF}_{M}^{(k)} \), \( \text{APF}_{V}^{(k)} \), and \( \text{APF}_{C}^{(k)} \) as

\[
\text{APF}^{(k)} = \beta_{M} \times \text{APF}_{M}^{(k)} + \beta_{V} \times \text{APF}_{V}^{(k)} + \beta_{C} \times \text{APF}_{C}^{(k)} . \tag{6.8}
\]

where \( \beta_{M} \), \( \beta_{V} \), and \( \beta_{C} \) are the importance factors for \( \text{APF}_{M} \), \( \text{APF}_{V} \), and \( \text{APF}_{C} \), respectively and \( \beta_{M} + \beta_{V} + \beta_{C} = 1 \).

Ideally, every time that a sample is collected, we build an updated model and find the next sampling location as the position of cell \( k \) in which \( \text{APF}^{(k)} \) is maximised. In practice, when this process is applied in sensor planning of a mobile sensor in a real-world application where there are time and resource constraints, we need to avoid large movements of the mobile robot and thus, apply a locality constraint on the selection of the next sampling point. Neumann [8] introduced the locality constraint function to select the next sampling point for a micro-drone from \( n_{sp} \) suggested measurement locations at which the total APF takes its maximum. The adaptive sensor planning with imposing locality constraint is called SPPAM [51]. Locality constraints and SPPAM will be further explained later in this section.

Theoretically, it could happen that the attractive forces toward an increased mean and an increased variance in the opposite direction cancel themselves out.
In practice, it is unlikely that the attractive forces are completely balanced at the position of the sensor. Even if they were, the sensor would be directed toward one of the directions, and in the next step, the symmetry would be broken.

In the proposed sensor planning approach, APFSP, I do not use any form of gradient descent to navigate; therefore, the mobile sensor does not get trapped in local minima as in common potential field approaches to local obstacle avoidance. In addition, there are repulsive terms in this approach at all previous sampling locations. These terms help the sensor to avoid previously visited locations and thus to escape local minima.

Locality Constraint

As mentioned earlier, the APFSP approach distributes its suggestions over the target area without any spatial order. Moving the mobile gas sensor directly to these locations tends to create a see-saw movement over the target area, which empties the batteries sooner, resulting in fewer measurements. Therefore, we added a locality constraint by imposing a restriction on the selection of the next sampling point from the $n_{sp}$ suggestions of APFSP. Neumann explored different selection criteria to impose locality constraints in sensor planning for micro-drones [51]. One straightforward solution is to select the closest measurement location (SPPAM-A). This choice preserves the battery power compared to pure APFSP; however, it does not allow exploration of the areas with higher predictive mean and variance if they are not close-by. To solve this issue, one alternative solution is to select the most often suggested close-by measurement location (SPPAM-B). This step is implemented by a matrix $S$ that has the same discretisation as the gas distribution model. For each grid cell $k$, $S^{(k)}$ counts how often the cell was suggested since it was visited the last time. The next measurement point is ultimately selected as the one with the highest ratio $\frac{S^{(k)}}{d^{(k)}}$, where $d^{(k)}$ is the distance between the current position of the sensor and grid cell $k$. Thus, a location far away from the current position will only be selected if it was suggested frequently. In [51], Neumann compared these two locality constraints in a simulation experiment. The results indicated better performance of SPPAM-A measured as travelling distance and gas distribution modelling quality. Therefore, in the simulation experiments presented in this chapter, we use SPPAM-A as the locality constraint approach.

6.4 Evaluation Metrics

To analyse the performance of APFSP and its suitability for the study of gas dispersion, I applied several evaluation metrics. This section elaborates on the applied evaluation metrics.
6.4.1 Distribution Similarity

The main questions that the evaluation presented in this chapter investigates is which sensor planning approach leads to a better estimate of gas distribution in the target area given limited time and resources. For this purpose, the similarity of the created gas distribution models are compared to the model generated by the ground truth data. As mentioned in Chapter 3, simulated gas dispersion experiments provide a high-resolution ground truth that enables quantitative comparison and evaluation of performance of different sensor planning strategies. Therefore, to compare different sampling strategies, I use simulated gas distribution experiments and compare the quality of created gas distribution models with the available ground truth in these experiments. We use the Kullback-Leibler (KL) divergence or relative entropy for probability as a measure of distribution similarity. KL-divergence is defined as

\[ \text{KL}(p|q) = \int p(x) \ln \frac{q(x)}{p(x)} \, dx, \]  

(6.9)

where \( p(x) \) is the unknown distribution (the ground truth) and \( q(x) \) is the modelled distribution. To have a symmetric measure, we use KL-distance measure defined as

\[ \text{KL}(p, q) = \frac{1}{2} \text{KL}(p|q) + \frac{1}{2} \text{KL}(q|p). \]  

(6.10)

6.4.2 Coverage

To measure the coverage of a sampling strategy, the environment is discretised into a grid. The coverage is defined as the number of unique cells visited in sample collection. In the experiments presented in this chapter, the grid cell size is set to 0.25 m, the same as the cell size of the applied gas distribution models. With this definition, the maximum value of coverage is equal to the number of grid cells in the area divided to the total collected samples. Larger values of coverage indicate better coverage (exploration) of the environment while smaller values indicate less coverage. Coverage can be a key criterion in environmental monitoring applications, for example, in scenarios where no prior information about the gas dispersion and the environment is available.

6.4.3 Plume Coverage

Plume coverage can be a criterion in different applications such as incident management where gas emission is involved. Sample collection inside the plume improves the quality of the estimated gas distribution model. As a measure to evaluate the performance of APFSP, plume coverage is computed as the percentage of samples which measure gas concentrations larger than a threshold. Here, we set the threshold to 1% of the maximum concentration measured.
6.4.4 Travelling Distance

In real-world mobile robot olfaction applications, the sampling strategy should consider time and power consumption constraints. To investigate usability of the proposed sensor planning approach, we compared the total travelling distance to collect samples, computed as the sum of the Euclidean distances between consecutive samples.

6.4.5 Distance to Gas Source Location

APFSP imposes a trade-off between exploration and exploitation. This trade-off is introduced in APFSP using $\beta_C$, $\beta_M$, and $\beta_V$ which weight the importance of each criterion in the APFSP sensor planning. At the beginning of the sampling when the knowledge about the environment has low confidence, it is reasonable to focus on exploration in the environment over exploitation. As the number of collected samples increases, the predictive gas distribution model will have higher confidence values, and it is reasonable to expect the sampling strategy to focus more on the exploitation of areas with higher mean and variance. In use-cases where sensor planning is used for gas source localisation, it is expected that as more samples are collected, the robot explores the area close to the gas source. Therefore, we calculate the distance of the latest collected sample from the gas source as one of the evaluation measures in the experiments presented in chapter.

6.5 Experiments

6.5.1 Simulation

To evaluate sensor planning approaches in a more realistic simulation, we carried out experiments in the ROS simulated environment, "Sim-ROS", using MOX sensor model (cs4). The sensor planning collects samples on the two dimensional $x - y$ plane at the height of 0.25 m above the ground. In this experiment, the gas source is located at $(x,y,z) = (5 \text{ m}, 10 \text{ m}, 0.10 \text{ m})$ from the bottom left corner of a wind tunnel of size $60 \times 20 \times 5 \text{ m}^3$ where two obstacles introduce additional turbulence. The samples are collected at $t = 1 - 100$ seconds with the frequency of 1 Hz. The experimental setup is explained in detail in Chapter 3.

The ground truth includes all available measurements in the simulated experiment. Figure 6.1 presents the predictive mean map created using the ground truth data in "Sim-ROS" cs4. The gas distribution models are created using basic Kernel DM+V with the kernel width $\sigma = 1.25 \text{ m}$ and grid cell size $c = 0.25 \text{ m}$. The cell size and kernel width are selected by optimising the NLPD value (see Chapter 5). In each experiment, we perform sensor planning to collect 1200 samples which (approximately) correspond to one sample per square meter if samples were collected simultaneously at equal distance in this
simulated environment. Samples are collected iteratively over the simulation time span (i.e. 12 samples per second).

![Figure 6.1: The predictive mean map created using ground truth data, corresponding to data from simulated experiment "Sim-ROS" cs4. In cs4, the gas source location, P_A, is indicated with a filled white circle. In this experiment, a MOX sensor model is used to collect samples. All maps are created using Kernel DM+V with σ=1.25 m and c=0.25 m.](image)

The quality of the predictive maps create by applying APFSP is compared to those created by applying common sampling strategies:

- random sampling,
- vertical sweeping: sampling along a predefine sweeping trajectory perpendicular to the plume,
- horizontal sweeping: sampling along a predefine sweeping trajectory in parallel to the plume, and
- random sweeping: where samples are collected randomly from a predefined vertical sweeping trajectory.

In each experiment, "random sampling", "random sweeping", and APFSP approaches are performed 10 times with random initial sampling locations. Figure 6.2 shows an example of sampling trajectories of the sweeping sampling strategies. To compare equivalent trajectories, all examined sensor planning methods select the same number of measurements, 1200 samples.

Parameter Selection

In APFSP sampling strategy, the parameters $\beta_M$, $\beta_V$, and $\beta_C$ weigh the importance of objectives which direct sample collection towards areas with higher predictive mean, higher predictive variance, and lower confidence, respectively (see Section 6.3.1). The importance factors or meta-parameters are selected such that $\beta_M + \beta_V + \beta_C = 1$. This section investigates the impact of the
Figure 6.2: Trajectory of sampling locations in (a) sweeping perpendicular to the plume noted as "vertical sweeping", (b) sweeping in parallel to the plume noted as "horizontal sweeping", and (c) a random sampling from a predefined sweep perpendicular to the plume noted as "random sweeping".
selection of this meta-parameter set on the performance of sensor planning. Evaluation metrics that are applied to analyse the performance of sensor planning include KL-distance, area coverage, plume coverage, total travelling distance, and distance of the latest measurement point from the gas source (see Section 6.4). Evaluations are performed on the simulated experiment "Sim-ROS" cs4. Each experiment is performed 10 times. In each trial, the location of the first sample is randomly selected. In each experiment, 1200 samples are collected consecutively.

To visualise the evaluation results, ternary plots are used. Each graph presents the mean values of the corresponding evaluation metric reported from 10 trials. The blue color in Figure 6.3, 6.5, 6.6, 6.7, and 6.8 indicates smaller values of the evaluation metric, and yellow indicates larger values. In these ternary plots, we are interested in the set of meta-parameters which optimise evaluation metrics.

KL-distance evaluation of predictive mean maps created using APFSP with different meta-parameter sets is presented in Figure 6.3. Smaller values of KL-distance indicate a better quality of the gas distribution models. Hence, in Figure 6.3, we are interested in the blue areas (i.e. set of meta-parameters which lead to smaller values of KL-distance). This pattern is visible in areas with high values of $\beta_C$ and smaller values of $\beta_M$ or $\beta_V$. This result implies that, as expected, exploring areas where high values of uncertainty or lower confidence values are observed, leads to a better prediction of the gas distribution.

![Figure 6.3](image)

Figure 6.3: The ternary graph of KL-distance of predictive mean maps created by APFSP using different meta-parameter sets in the simulated experiment "Sim-ROS" cs4. Smaller values of KL-distance are indicated in blue, and larger values of KL-distance are indicated in yellow. The graph presents the mean of values reported from 10 trials.
In addition, we investigated the performance of APFSP sampling strategy in terms of KL-distance for different sample sizes. Figure 6.4 shows the KL-distance evaluation of the predictive mean maps created using APFSP sampling strategy to collect 200, 400, 600, 800, 1000, and 1200 samples, presented in order from the top left to the bottom right, respectively. In small number of samples (200 samples), combinations of non-zero values of $\beta_C$ and low to medium values (smaller than 0.8) of $\beta_M$ are good candidates. As the number of samples increases the importance of exploring in areas where gas distribution model shows lower confidence (i.e. larger values of $\beta_C$) becomes more important to derive better predictive mean maps.

In Figure 6.5, travelling distance evaluation of APFSP sampling strategy for different sets of meta-parameters is presented. Larger values of total travelling distance mean more power and time consumption for a mobile sensor in real-world applications. Therefore, we are interested in areas with smaller values of travelling distance (i.e. areas in blue and green color) in Figure 6.5. Results indicate that small values of $\beta_C$, with large values of $\beta_M$ or $\beta_V$ result in a shorter travelling distance.

Figure 6.6 illustrates the result of area coverage evaluation of APFSP sampling strategy for different sets of meta-parameters. In this visualisation, we are interested in areas in yellow (i.e., a set of parameters which leads to larger coverage of the area). As shown in Figure 6.6, the coverage of the investigation area is larger when large values of $\beta_C$ are selected.

Figure 6.7 shows the result of plume coverage evaluation for different sets of meta-parameters. In this graph, we are interested in areas in yellow (i.e., set of parameters which leads to having more samples collected inside the gas plume). Results indicate a better plume coverage when $\beta_C$ is relatively small and $\beta_M$ is large (which intuitively means a combination of exploration and exploitation to detect the plume and to exploit inside the plume).

The result of measuring the distance of the last sampling location selected by the APFSP from the gas source location is presented in Figure 6.8. As more samples are collected from the area, we are interested in collecting more information about the areas with higher gas accumulation, especially close to the gas source. Considering this, in Figure 6.8, we are interested in areas in dark blue (i.e., set of parameters which leads to having shorter distance of the last measurement from the gas source). Results indicate that large values of $\beta_V$ combined with relatively small values of $\beta_M$ leads to a better performance of APFSP in terms of distance to the gas source.

Comparison

In the next step, we compared the performance of APFSP to existing methods; "horizontal sweeping", "vertical sweeping", "random sweeping", and "random sampling". In this set of experiments, for each evaluation metric, the APFSP with the best performing meta-parameter set is used. Each experiment is car-
Figure 6.4: Ternary graphs of KL-distance of predictive mean maps created by APFSP using different meta-parameter sets in the simulated experiment "Sim-ROS" cs4. Smaller values of KL-distance are indicated in blue, and larger values of KL-distance are indicated in yellow. The graphs are created after collecting: (a) 200, (b) 400, (c) 600, (d) 800, (e) 1000, and (f) 1200 samples. Each graph presents the mean of values reported from 10 trials.
Figure 6.5: The ternary graph of travelling distance of APFSP using different meta-parameter sets in the simulated experiment "Sim-ROS" cs4. Blue color indicates shorter travelled distance and yellow indicates longer travelled distance in the area. The travelling distance values are in meter. The graph presents the mean of values reported from 10 trials.

Figure 6.6: The ternary graph of coverage of APFSP using different meta-parameter sets in the simulated experiment "Sim-ROS" cs4. Smaller coverage are indicated in blue and yellow indicates larger coverage of the area. The coverage values are percentage of unique cells ($c = 0.25$) at which samples are collected. The graph presents the mean of values reported from 10 trials.
Figure 6.7: The ternary graphs of plume coverage of APFSP using different meta-parameter sets in the simulated experiment "Sim-ROS" cs4. Blue indicates smaller plume coverage and yellow indicates larger plume coverage. The values of plume coverage are presented in percentage. The graph presents the mean of values reported from 10 trials.

Figure 6.8: The ternary graph of the distance of the last sampling location from the gas source in APFSP using different meta-parameter sets in the simulated experiment "Sim-ROS" cs4. Blue color indicates smaller and yellow indicates larger distance from gas source. The values of distance to the gas source are in meter. The graph presents the mean of values reported from 10 trials.
ried out 10 times. The performance comparison of different approaches to APFSP is presented in Table 6.1. The results indicate that APFSP outperforms other examined methods in terms of KL-distance, plume coverage, and distance of the last measurement to the gas source. As expected, horizontal and vertical sweeping lead to a better area coverage after collection of all 1200 samples. Applying APFSP results in a larger travelling distance value compared to horizontal and vertical sweeping.

Table 6.1: Performance comparison of APFSP to other sensor planning methods using different metrics. For each evaluation metric, in APFSP, the results of the best performing meta-parameter set is reported. The corresponding meta-parameter sets \((\beta_M, \beta_V, \beta_C)\) for APFSP are presented in the last row. Each experiment is performed 10 times.

<table>
<thead>
<tr>
<th>Metric</th>
<th>KL-distance</th>
<th>Travelling distance (m)</th>
<th>Plume Coverage (%)</th>
<th>Area Coverage (%)</th>
<th>Distance to Gas Source (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweep</td>
<td>0.3175</td>
<td>1199</td>
<td>11.08</td>
<td>6.25</td>
<td>10.85</td>
</tr>
<tr>
<td>Vertical</td>
<td>0.2419</td>
<td>1199</td>
<td>8.67</td>
<td>6.25</td>
<td>10.85</td>
</tr>
<tr>
<td>Sweep</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random</td>
<td>0.0145</td>
<td>2.70\times10^4</td>
<td>11.18</td>
<td>4.05</td>
<td>30.38</td>
</tr>
<tr>
<td>Sweep</td>
<td>± 0.0033</td>
<td>±4.17\times10^2</td>
<td>±0.50</td>
<td>±0.047</td>
<td>±15.17</td>
</tr>
<tr>
<td>Random</td>
<td>0.0139</td>
<td>2.70\times10^4</td>
<td>11.58</td>
<td>5.56</td>
<td>29.26</td>
</tr>
<tr>
<td>Sweep</td>
<td>±0.0027</td>
<td>±5.86\times10^2</td>
<td>±0.91</td>
<td>±0.04</td>
<td>±17.74</td>
</tr>
<tr>
<td>APFSP</td>
<td>0.0097</td>
<td>1.54\times10^3</td>
<td>82.63</td>
<td>5.56</td>
<td>7.03</td>
</tr>
<tr>
<td>Param</td>
<td>±0.0019</td>
<td>±6.63\times10^2</td>
<td>±6.50</td>
<td>±0.05</td>
<td>±0.25</td>
</tr>
</tbody>
</table>

Imposing Locality Constraint

In real-world use-cases where there are time and resource constraints, we are interested in keeping the travelling distance as short as possible, while maximising the coverage of the area (e.g. in case of environmental monitoring), or maximising the plume coverage to collect as much information as possible about the gas plume and potentially identify the gas source location (in case of incident management). The parameter selection study in this section indicates that high importance factor for confidence \(\beta_C\) allows better coverage of the environment, but at the same time, it leads to a larger travelling distance. Imposing locality constraints reduces the travelling distance and see-saw movements.
This section investigates the impact of imposing locality constraint on the performance of APF-based sensor planning. In Section 6.3.1, two locality constraint logics for APF-based sensor planning were explained. For performance comparisons that are discussed in this section, we use SPPAM-A, the best performing method in the simulation results presented in Neumann’s work [51].

Table 6.2 presents the results of comparing APFSP and SPPAM-A using different evaluation metrics. Each experiment is carried out 10 times. The results confirm that the optimised meta-parameter set in SPPAM-A leads to a shorter travelling distance. APFSP, on the other hand, outperforms SPPAM-A when it comes to the quality of estimated gas distribution maps. APFSP performs slightly better in terms of plume coverage while SPPAM-A shows a slightly better performance in terms of distance of the latest measurement to the gas source location. However, the differences are not distinctive. Deciding on whether to use SPPAM-A or APFSP depends on the application. In applications where there are limited time and power available, SPPAM-A is a better choice while if constraints allow, APFSP is a better choice.

### Table 6.2: Performance comparison of APFSP to SPPAM-A using different metrics.

For each evaluation metric, the results of the best performing meta-parameter set is reported. The two rows of \((\beta_M,\beta_V,\beta_C)\) present the used meta-parameters corresponding to the results in APFSP and SPPAM-A. Each experiment is performed 10 times.

<table>
<thead>
<tr>
<th>Metric</th>
<th>KL-distance</th>
<th>Travelling distance (m)</th>
<th>Plume Coverage (%)</th>
<th>Area Coverage (%)</th>
<th>Distance to Gas Source (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>APFSP</td>
<td>0.0097</td>
<td>1.54x10^3 ± 0.0019</td>
<td>82.63 ± 6.50</td>
<td>5.56 ± 0.05</td>
<td>7.03 ± 0.25</td>
</tr>
<tr>
<td>Param</td>
<td>(0.0,1.0)</td>
<td>(0.8,0.2.0)</td>
<td>(0.3,0.7,0)</td>
<td>(0.0,1.0)</td>
<td>(0.0,0.6.0,4)</td>
</tr>
<tr>
<td>SPPAM-A</td>
<td>0.0135</td>
<td>5.81x10^2 ± 0.0001</td>
<td>75.78 ± 3.54</td>
<td>5.52 ± 0.06</td>
<td>7.31 ± 2.04</td>
</tr>
<tr>
<td>Param</td>
<td>(0.0,1.0)</td>
<td>(0.3,0.7,0)</td>
<td>(0.3,0.5,0.2)</td>
<td>(0.0,1.0)</td>
<td>(0.1,0,0)</td>
</tr>
</tbody>
</table>

**Time-dependent GDM in APFSP**

To analyse the performance of APFSP, so far we used Kernel DM+V to build gas distribution models. In another experiment, we compared the performance APFSP when using the time-dependent gas distribution modelling method, TD Kernel DM+V, compared to using its time-invariant version, Kernel DM+V. Kernel DM+V has two meta-parameters; kernel width \(\sigma\) and cell size \(c\). TD Kernel DM+V, in addition to these two meta-parameters, has a third meta-parameter, timescale factor \(\beta\), which weighs the importance of recency of measurement when creating gas distribution model. In this experiment, we set \(\sigma = 1.25 \text{ m}\) and \(c = 0.25 \text{ m}\), same as the meta-parameters in
Kernel DM+V. In the sampling process, every time a new sample collected, timescale factor $\beta$ is selected by minimising the NLPD value. The optimisation is carried out using the same data partitions as described in Chapter 5 (i.e., using the first 80% of collected samples in the order of time as the training set, and the last 20% as the validation set). Since at the beginning only a few samples are available, the meta-parameter selection is carried out only after at least 100 samples are collected. Prior to that we set $\beta$ to zero in TD Kernel DM+V which is equivalent to Kernel DM+V (see Chapter 5). TD Kernel DM+V requires a target time $t^*$. Here, we built gas distribution models based once considering the end time of the experiment. In the second experiment, we set the target time to the sampling time of the latest available measurement. The latter can be a good choice when the future prediction time is not decided, and the purpose is to get the best model for the next sampling time.

To compare the performance of APFSP with TD Kernel DM+V to APFSP with plain Kernel DM+V, we carried out APFSP with the meta-parameter set $(\beta_M, \beta_V, \beta_C) = (0.0, 0.0, 1.0)$ which is the best performing meta-parameter set in terms of minimising the KL-distance based on the analysis presented in Section 6.5.1. For each experiment, we carried out 10 trials.

Table 6.3 presents a performance comparison of APFSP using Kernel DM+V to TD Kernel DM+V. The results indicate that APSFP sampling strategy with TD Kernel DM+V performs slightly better in terms of KL-distance and leads to a shorter distance of the latest measurement to the gas source. When it comes to plume coverage and area coverage using Kernel DM+V leads to a better performance.

To analyse how the quality of the gas distribution model changes as more samples are collected in APFSP sampling strategy when using time-dependent GDM versus the time-invariant one, we computed KL-distance at each sampling point. Figure 6.9 presents the KL-distance comparison of predictive mean maps created by performing APFSP using Kernel DM+V (line in grey) to TD Kernel DM+V (line in red for $t^* = \max(t_i)$ and line blue for $t^* = 100$ s). The results presented in Figure 6.9 indicate that during the collection of the first 500 samples, KL-distance is lower when applying Kernel DM+V. However, as gradually more samples are collected, after 500 samples, TD Kernel DM+V, leads to lower KL-distance and thus is a better alternative. The KL-distance of sensor planning approaches applying the two methods converges around 1200 samples. At the beginning when there is only a small number of samples available, it makes sense to rely on all measurements when creating gas distribution model. The lower performance of TD Kernel DM+V at the beginning can be explained by the small number of sample available to optimise the timescale factor. Further investigation on when and how to use TD Kernel DM+V in sensor planning is an area for future work.
Table 6.3: Performance comparison of APFSP sampling strategies using TD Kernel DM+V versus using Kernel DM+V. APFSP meta-parameters are set to \((\beta_M, \beta_V, \beta_C) = (0.0, 0.0, 1.0)\). Maps are created with the parameters \((\sigma, c) = (1.25\,\text{m}, 0.25\,\text{m})\). In TD Kernel DM+V, the timescale factor \(\beta\) is learned by optimising the NLPD value after each sample collection. Each experiment is performed 10 times.

<table>
<thead>
<tr>
<th>GDM</th>
<th>Kernel DM+V</th>
<th>TD Kernel DM+V</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(t_* = \max(t_i))</td>
</tr>
<tr>
<td>KL-distance</td>
<td>0.010±0.002</td>
<td>0.008±0.002</td>
</tr>
<tr>
<td>Travelling distance (m)</td>
<td>2.77×10^4 ±4.41×10^2</td>
<td>2.88×10^4 ±4.80×10^2</td>
</tr>
<tr>
<td>Area Coverage (%)</td>
<td>5.56±0.05</td>
<td>5.26±0.03</td>
</tr>
<tr>
<td>Plume Coverage (%)</td>
<td>11.44±0.61</td>
<td>10.33±0.55</td>
</tr>
<tr>
<td>Distance to Gas Source (m)</td>
<td>34.25±14.99</td>
<td>18.62±11.86</td>
</tr>
</tbody>
</table>

Figure 6.9: KL-distance comparison of predictive mean maps created using APFSP sampling strategy with Kernel DM+V (line in black), TD Kernel DM+V with \(t_* = \max(t_i)\) (line in red), and TD Kernel DM+V with \(t_* = 100\) s (line in blue). Each experiment is performed 10 times. The shaded areas indicate confidence intervals and the solid lines show the mean of the trials. APFSP meta-parameters are set to \((\beta_M, \beta_V, \beta_C) = (0.0, 0.0, 1.0)\). Maps are created with the parameters \((\sigma, c) = (1.25\,\text{m}, 0.25\,\text{m})\). In TD Kernel DM+V, the timescale factor \(\beta\) is learned by optimising the NLPD value after each sample collection.
6.5.2 Real-world Experiments

We carried out real-world experiments "QC-AASS" with the gas-sensitive micro-drone as described in Chapter 3 using SPPAM-B to direct the micro-drone. The objectives of sensor planning in experiment were gas distribution mapping and gas source localisation. This experiment was carried out at the early stage of the collaboration on APF-based sensor planning when the exploration factor in APF was defined by using the repulsion from previous sampling locations (APF_R). Therefore, the APF was formulated as

\[ APF^{(k)} = \beta_M \times APF_M^{(k)} + \beta_V \times APF_V^{(k)} + \beta_R \times APF_R^{(k)}. \]  

(6.11)

Since wind information was available in these experiments, the Kernel DM+V/W algorithm was used to build the predictive gas distribution model, and the meta-parameters were set to \( (c = 0.15m, \sigma = 0.40m, \gamma = 0.2s) \). The SPPAM sensor planning algorithm was run with equal weight factors \( \beta_M = \beta_V = \beta_R \).

A total number of 16 runs were performed within this experiment. Each run took 14 to 19 minutes. Figure 6.10 demonstrate sampling trajectory of the mobile sensor in one of the trials (trial #3) using SPPAM sampling strategy.

Figure 6.10: Calculated sampling trajectory of the sensor planning algorithm (trial #3) with starting position \((x, y) = (11.0m, 7.0m)\) in the "QC-AASS" experiment. The gas source is approximately located at \((x, y) = (6.3m, 3.8m)\) [8].

Figure 6.11 shows the state of trial #3 after 31 measurement points when the last sample is collected. The top row shows the previously visited sampling locations (left) and the corresponding mean distribution map \( r^{(k)} \) (middle), as well as the variance distribution map \( v^{(k)} \) (right). The bottom row shows the visualisation of the APF map, the area with the next suggested sampling points and source location estimate, which are indicated by green and red
dots, respectively (middle), and the matrix $S$, where the white cells are the previously selected cells and the darkest cells correspond to the most often suggested ones. Figure 6.10 shows the trajectory produced by the sensor planning algorithm in run \#3 with starting position $(x, y) = (11.0 \text{m}, 7.0 \text{m})$.

![Image of sensor planning process](image)

Figure 6.11: Adaptive sampling process: From the previously visited sampling locations (top, left), predictive mean $r^{(k)}$ (top, middle) and variance maps $v^{(k)}$ (top, right) are created using the Kernel DM+V/W algorithm with $(\sigma = 0.4 \text{m}, c = 0.15 \text{m}, \gamma = 0.2 \text{s})$ as meta-parameters. The information from the previously visited sampling locations and these two maps are used to create the APF with $\beta_M = \beta_V = \beta_R$ (right). Then, a set of next suggested sampling points is selected. These points and the estimated source location are denoted by green and red dots, respectively (bottom, middle). The next sampling location is chosen based on the locality constraint sampling strategy, which is shown by the matrix $S$ (bottom, left), where the white cells are the previously selected cells, and the darkest cells correspond to the most often suggested ones. All plots were created after the last time step of the data collection (measurement 31 in the presented run) [8].

### 6.6 Discussion and Conclusion

The proposed APF-based approach balances objectives related to exploration and exploitation. The repulsive part prevents repeated measurements at the same locations and thus promotes exploration. This is incorporated using a confidence map created by statistical gas distribution modelling. The attractive part directs the attention to areas for which higher gas accumulation or higher variance in the predictive gas dispersion is predicted; in other words, this leads to further exploitation. The importance of these objectives is introduced in APF-based sensor planning using three meta-parameters: $\beta_M$, $\beta_V$, and $\beta_C$ corresponding to weights for predictive mean, predictive variance, and areas
with high uncertainty, respectively. This chapter presented an analysis of meta-
parameter selection in the performance of sensor planning in terms of gas
distribution quality, area coverage, plume coverage, total travelling distance
and distance of the latest measurement point to the gas source location. This
study provides insights on how to select meta-parameters depending on use-
case of sensor planning.

In the analysis presented in this chapter we observed that small values of
$\beta_C$ with relatively large values of $\beta_M$ weigh sampling locations for a better
plume coverage. Large values of $\beta_V$ together with small values of $\beta_M$ leads the
sample collection towards the areas with high predictive variance values which
can potentially indicate a gas source. Large values of $\beta_C$ leads to a better
quality of gas distribution maps and high area coverage; while small values of
$\beta_C$ with large values of $\beta_M$ or $\beta_V$ lead to a shorter total travelling distance.

In this chapter, we compared the performance of APFSP to commonly used
sampling strategies: random sampling, horizontal sampling, vertical sweeping
and random sweeping. Horizontal and vertical sweeping results in a better
area coverage but a lower plume coverage. In addition, the results showed that
APFSP leads to creating improved quality in gas distribution modelling.

Imposing locality constrain improves the required total travelling distance
in APF-based sampling, however, it results in worse KL-distance compared
to APFSP. The analysis suggest that for applications where there are limited
power and time available to use SPPAM-A, otherwise, APFSP.

With the introduction of the locality constraint, sensor planning compo-
nent can be used to plan suitable paths for a mobile gas sensor. The proposed
algorithm was tested both in simulation and in real-world experiments. The
results demonstrate that the proposed sensor planning approach allows obtaining
expressive gas distribution models in substantially shorter time as through
random or other forms of coverage sampling. Chapter 7 presents an exam-
ple application, where we investigate applying gas distribution modelling and
sensor planning in real-world scenarios.

In another experiment, we used time-dependent gas distribution modelling
instead of basic Kernel DM+V in sensor planning. This resulted in slightly
better quality of gas distribution models and a shorted distance of the last
measurement to the gas source location.

In the work presented in this chapter, we set the sensor planning meta-
parameters prior to the sampling. Future work will include learning the meta-
parameter set while sampling. In this chapter, a proof of concept to apply
time-dependent gas distribution modelling in integration with sensor planning
was explained. Online optimisation of the meta-parameters is a future work.
Chapter 7

A Large-scale Pollution Monitoring Application

Chapters 5 and 6 presented gas distribution modelling and sensor planning methods for dynamic environments, in particular when limited time and resources are available. The performance of these approaches has been evaluated using simulated experiments and real-world sensor measurements collected in small-scale indoor and outdoor environments. A next step is to apply the discussed kernel methods on a larger scale and in a real-world application. This line of research was pursued in the EC project Diadem (FP7-ICT 224318, 2008) [29]. Diadem addressed Distributed Information Acquisition and Decision-making for Environmental Management, particularly in cases that involved chemical incidents which caused emission of hazardous chemical gases.

The use-case scenario in the project is an area of approximately 50 km$^2$ at the Rotterdam port in the Netherlands, a densely inhabited industrial area where several refineries and oil factories are located.

The research and development was guided by real-world requirements provided by Diadem end user partners DCMR (Environmental Protection Agency, Port of Rotterdam area, The Netherlands) and DEMA (Danish Emergency Management Agency, Denmark) [5].

The Diadem framework consists of several components: to process large amounts of data available from heterogeneous sources of information, identify critical situations, assess the possible evolution of gas emissions, plan and evaluate preventive measures, and assist in decision-making.

In Diadem-like problems where gas emission is present, it is critical to have a good estimate of the gas distribution to highlight areas with potentially high gas accumulation, detect anomalies, and localise gas sources such as leakages. In fact, the gas distribution modelling and sensor planning methods discussed so far in this thesis work can be used to inform the experts about the gas plume and distribution of gas to improve their ability to assess the incident.
In the scenario addressed by Diadem, measurements are continuously collected by a network of gas sensors in a large-scale uncontrolled environment. These measurements need to be interpreted online to aid the overall assessment of the situation and to point out locations for additional measurements. The sensor planning and gas distribution modelling methods, proposed in Chapters 5 and 6, were used as solutions to improve the local assessment of the situation (i.e., early plume detection), identification of pollution sources, concentration estimates near the sources, and current impact on people’s health. For this purpose, we developed two modules: (1) Advanced Gas Distribution Models (GDM), and Sensor Planning. We integrated these two modules in the Diadem framework as the Gas Monitoring component. The contribution of this thesis work to the Diadem project was to develop this component and to integrate it with other components of Diadem.

This chapter first presents an overview of the Diadem project (see Section 7.1). To understand the challenges in a Diadem-like scenario and to identify where these modules can assist experts in the management of an environmental incident, we had field visits and meetings with experts from DEMA and DCMR, the industrial partners in Diadem. The feedback from these experts provided the key inputs in building the GDM and Sensor Planning modules and assessing their performance. Sections 7.2 and 7.3 summarise the results of these investigations on GDM and Sensor Planning, respectively. Section 7.4 presents the Diadem framework and explains how GDM and Sensor Planning modules are integrated into this framework.

The assessment of using GDM as a comprehensive tool in two examples, driven from the Diadem project, is elaborated in Section 7.4.5. The chapter ends with a discussion and conclusion regarding this assessment (Section 7.5). The overall result of this integration and the performance of the Gas Monitoring component in the defined use-cases is partially published in a publication of the Diadem project [5].

7.1 Diadem Project

Diadem is an EU FP7 research project that supports environmental management for chemical hazards in industrial areas, with an enhanced capacity to assess population exposure and health risks, to alert relevant groups, and to organise an efficient response. The emphasis is on advanced solutions which are economically feasible and which maximally exploit the existing communication, computing, and sensing resources. Diadem supports collaborative information acquisition, thus enabling efficient situation assessment and decision-making in complex environmental management applications. This is achieved through a combination of

1. a service-oriented approach which supports the integration of automated and human-based reasoning for large-scale collaborative sense making,
2. advanced approaches to gas detection and gas distribution modelling,

3. novel solutions that combine multi-criteria decision analysis and scenario-based reasoning, and

4. enhanced human-machine interfaces [5].

In Diadem, this thesis work contributed to the second point, gas detection and gas distribution modelling, by developing a gas monitoring component which is the focus of this chapter. In this section, first an overview of the Diadem is presented. This is followed by further discussion of the components that supported the gas detection and gas distribution modelling functionalities.

Figure 7.1 presents which components are involved in Diadem and how they work together in the Diadem Framework. To understand the current situation in an incident, information is collected in Diadem using heterogeneous sources including stationary sensor networks, mobile sensors carried by field investigators, and human reports. A human interaction component facilitates the acquisition of information through crowd sourcing by an interface where users can enter their queries on mobile phone and the internet. One of the industrial partners, DCMR, provided gas measurements collected by a sensor network. In the first step, two components enable an early assessment of the situation in the investigation area. These two components are gas detection system (DGDS) and gas monitoring. DGDS uses data collected from human reports and the measurements from gas sensors to detect gas plume at early stages and to localise potential gas pollution sources. The Gas Monitoring component uses gas sensor measurements to create gas distribution models which can identify areas of high gas accumulations. This is shown as GDM in Figure 7.1. To complement this local assessment of the situation, another component, ARGOS, predicts plume development at a larger scale using atmospheric gas distribution models and weather information. ARGOS uses the local assessments provided by DGDS and Gas Monitoring components to refine its estimates. In addition, ARGOS provides prognoses on incident development given the population and spatial information about the incident area. In the next step, the outputs from Gas Monitoring, DGDS, and ARGOS are visualised to help experts to assess the situation better. In addition, these outputs are sent to a situation based reasoning component (SBR), and a decision-making support system component (MCDA). The result will help the experts to make a better informed decision in the management of the incident. A framework, DPIF, facilitates the communication between different components and handles different requests.

In the rest of this section, the components involved in the detection of gas source and gas distribution modelling are explained. Section 7.4 elaborates on how Gas Monitoring is integrated in Diadem.
7.1.1 Gas Monitoring

The Gas Monitoring component provides a mapping of gas concentrations in the vicinity of the incident site using sensor measurements. This component enables assessment of the current situation (e.g., early plume detection). As a particular benefit, the potential of such a system to improve the effectiveness of the planning processes and the safety of the staff at the incident site has been acknowledged by experts at DCMR and DEMA (see Section 7.2). In Section 7.2, the development of Gas Monitoring component and its integration into the Diadem framework are explained in detail.

7.1.2 Distributed Gas Detection and Gas Source Localisation

The Distributed Gas Detection System (DGDS) is a component which enables fast detection of gas plumes and localisation of the pollution sources using heterogeneous data. The data sources are chemical sensors and reports from citizens collected at different locations and points in time. DGDS continuously collects data from these data sources and analyses the data for sensor measurements or complaints that are likely to be caused by the presence of some gas [5]. Given the wind direction, the system develops a set of hypotheses, each corresponding to a potential source. Observations obtained within the area associated with different hypotheses are used to rank the hypotheses according to the likelihood of the potential sources. The likelihood is determined by applying a decentralised Bayesian information fusion approach [63].
In Diadem, the two components DGDS and Gas Monitoring are complementing each other in the early assessment in the vicinity of the incident site. DGDS identifies likely gas source locations while Gas Monitoring provides a prediction of gas distribution which enables to detect plume and to identify areas with higher gas accumulation. In addition, Gas Monitoring based on the location of collected samples and the gas distribution suggests to the experts where to collect further measurements.

In the Diadem scenarios, no solid ground truth data is available to evaluate the quality of created gas distribution models. DGDS output can potentially be used to validate the quality of gas distribution models. On the other hand, the input from gas distribution modelling can be used as an additional input to increase the certainty in identifying likely gas source location(s). Improving the quality of gas distribution modelling using DGDS is a direction for future work.

7.1.3 ARGOS

One of the components in Diadem is ARGOS. ARGOS is a decision support tool which has its origins in nuclear emergency management [64]. The main application of ARGOS is to create a prognosis of how the situation will evolve, for instance which areas will be affected and when the areas will be affected.

ARGOS estimates the spatial distribution and concentration of gas plumes at different points in time using an atmospheric dispersion model and weather information. In addition, ARGOS can incorporate additional information such as the exposed population and infrastructure on a large scale, and allows the creation of prognoses of how the situation may evolve. Such knowledge is critical for short- as well as long-term decision-making, as it allows the estimation of the impact of various decision alternatives.

ARGOS is used when the location and time of a (gas) release into the air are known. DGDS provides information about the likely gas source locations to ARGOS. Gas Monitoring does not require prior knowledge about the gas source and provides gas distribution models in a finer resolution at the vicinity of the incident site. The results of Gas Monitoring enables ARGOS to improve its prognosis. ARGOS overlays the gas distribution models onto the map of the incident area for further analysis of the affected area.

7.2 GDM Module: Comprehensive Overview Tool

In the scenario addressed by Diadem, gas measurements are continuously collected by a network of metal oxide gas sensors. Wind information is available through reports from meteorological stations. These measurements have to be interpreted online.

Due to the working principle of the currently available affordable gas sensors, dense sensor networks can only be operated economically with sensors
that are, to a certain degree, unspecific (they respond to a range of chemicals) and that do not respond immediately to an increased gas concentration (the response time is in the order of seconds; recovery time is in the order of tens of seconds). As a consequence of the response characteristics, the absolute value reached by the sensors depends not only on the gas concentration but also on the time span that the sensors were exposed to the gas, which is a hidden variable. Under real-world conditions, gas sensors are virtually never exposed long enough to a constant concentration of a gas so as to reach a steady state response. However, standard gas sensor calibration assumes that the sensors are exposed under stable conditions to a single gas for a sufficiently long time to reach the steady state. In addition, in Diadem, samples are collected using a sensor network, and the cross-calibration between sensors in the network itself is an equally difficult and limiting factor.

In response to the issue that calibrated field data would not be available, we decided to investigate if applying statistical gas distribution modelling gas and presenting predictive maps as the relative concentration maps can be useful for the expert as a tool to provide a comprehensive view of the gas measurements. To this extent, the capability of the developed algorithms as an aid to summarise and interpret huge amounts of data from sensors, which are not calibrated (in the sense detailed above), are investigated. The developed gas distribution modelling algorithms are directly tested in the Diadem scenario with the data collected by DCMR as a tool to help human operators. However, the algorithms have been developed to be general so that they can be applied to Diadem-like or other scenarios in which fully calibrated sensors with the required specificity are available.

As part of this investigation, we created models using Diadem datasets and analysed the results. Moreover, we had close collaboration with experts from DCMR and DEMA to identify challenges in the real-world application and where Gas Monitoring can potentially assist them.

### 7.2.1 DCMR and Diadem Use-case

As use-cases for Diadem, DCMR provided two datasets collected by stationary sensor networks at the port of Rotterdam: "Diadem-I" and "Diadem-II" (see Chapter 3).

Rotterdam is geographically located below sea level. The wind field is strong with variable intensity and direction. There are several refineries and much petroleum-related industry. The pollution level is high (different smells can be detected with an unaided nose) and DCMR receives about five complaints per day. Companies are obliged to report incidents. When an incident is reported, field inspectors with gas sensor equipments travel to the site. In general, when field inspectors sniff a strange smell, or they receive a complaint call from citizens, they use photo-ionisation detectors (PID)
the air. Using gas coefficients from a handbook, the inspectors then calculate the approximate ppm concentration values for different types of gases.

DCMR experts in the following domains were particularly interested in using the outcome of the work on the Gas Monitoring component:

- Chemical advisers: Chemical advisers decide whether to send fire-fighters or field inspectors to the incident area for further investigations based on the received chemical reports on the amount of gas accumulation and gas source location.

- Control room experts: The experts in the control room analyse the received complaints and use the plume models to estimate the dispersion. They send field inspectors for further measurements to improve their estimates from the plume.

### 7.2.2 GDM in Combination with Other Solutions

In the meetings with DCMR and DEMA, the experts confirmed that the GDM module can assist them to make a better assessment of a crisis scenario. Representatives pointed out as an important advantage the module’s ease of use and its integration with other assistive tools which experts are using in this domain. To address this, we integrated the GDM module as a service in the Diadem framework. In Diadem, GDM provides input to the sensor planning and to ARGOS.

### 7.2.3 GDM with Uncalibrated Data

Given uncalibrated data, the gas distribution created by using normalisation of the individual responses can be interesting for DCMR to identify areas of unusual gas accumulation. In case such an area is detected, and there is a high risk of a hazardous gas, firefighters can be sent to the field with special equipment and protection. They then collect samples with Dräger tubes. In the case of negligible risk of hazardous gas, field inspectors are sent instead. Gas distribution maps created from uncalibrated measurements by GDM can help chemical advisers to identify areas where samples should be collected. This process can be automated using the sensor planning module.

### 7.2.4 Applications and Limitations of Existing GDM Methods

Selection of GDM Solution

To create gas distribution models, different solutions were studied. Considering the use-case specification and available sensor measurements, we chose Kernel DM+V as the solution in GDM module. This choice and potential extensions were discussed with experts from DCMR:
• Considering the wind in GDM: DCMR experts pointed out that the area sometimes has strong winds. Therefore, the wind has an impact on the dispersion of gas and consequently their decision-making plans. As explained in Chapter 4, Kernel DM+V does not take into account wind information in creating gas distribution models. There is, however, an extension of the Kernel DM+V approach called Kernel DM+V/W, that considers wind direction and intensity (see Chapter 4). A requirement for the Kernel DM+V/W algorithm is that wind information is collected at the same locations where the gas sensors are mounted. In a typical DCMR application, this is not the case; therefore, wind estimates at the gas sensor measurement points would be required. This step can be done in the future when wind information is available.

• Three dimensions in GDM: The applied solution creates gas distribution maps in two dimensions. However, there is a 3D extension of Kernel DM+V, which we agreed not to use in Diadem since the sensors are not mounted at different heights (see Chapter 4). DCMR experts confirmed that two-dimensional maps were sufficient in most cases.

• Capturing evolution of gas dispersion in GDM: TD Kernel DM+V, discussed in Chapter 5, considers the recency of measurements for the estimation of gas distribution models, and, to some degree, can capture the evolution of the gas dispersion. In the Diadem use-case, the required prediction is for a relatively long time in the future (e.g. the few hours). The analysis in Chapter 5 indicated that the gas distribution models increasingly rely on all (not just the most recent) data to predict measurements in a long time in future. In addition, the records of wind measurements from mounted sensors at the port of Rotterdam indicates that there is a strong wind intensity in this area, and the wind intensity and direction varies in short time intervals. This causes high fluctuations in the gas dispersion. Given the conditions of the Diadem use-case, relying on all historical measurements can be equally good or better choice than relying on only recent measurements. The analysis of using TD Kernel DM+V in Diadem-like scenarios is an area for future work.

Wind Information to Assess the Evolution of a Gas Distribution

According to the DCMR experts, it would be very helpful if the GDM maps would capture information about the evolution of the gas dispersion. One simple way would be to overlay the local wind information acquired at the meteorological stations in the form of arrows over the GDM maps. This approach is applied to the analysis of results provided in this chapter. More sophisticated solutions would first model a complete wind field and overlay the corresponding estimate. This development is identified as another potential direction for future work.
Plume Estimation

Two important challenges for DCMR are to predict where to carry out further measurements and to predict how the plume develops in the future (where there will be, for instance, a populated area involved). The method that is currently applied by DCMR requires a set of possible source locations as input. From this set, a number of possible plumes are estimated using information about complaints, wind direction, wind intensity, temperature, gas type, and gas dispersion characteristics.

DCMR experts have a set of plume models that are printed on transparencies. These plume models are computed using a simple atmospheric dispersion model that assumes, for example, a unidirectional wind field. To select the most likely hypothesis, these transparencies are put on a map of the field and an operator decides what the most likely hypothesis is. After selecting the most likely plume model, field inspectors are sent to the borders of the assumed plume to collect samples. With these new samples, the estimation of the plume is refined, and a new plume model will be chosen.

The experts who work in the DCMR control room were interested in GDM maps because these maps could potentially provide a more sophisticated estimate of the plume, especially in complex situations where the assumptions of the simple atmospheric dispersion model do not hold.

The area in which DCMR has mounted gas sensors often covers only part of the plume. The DCMR experts expressed their interest in information about the plume in a larger area. They asked particularly whether GDM can provide such an estimate. Applying extrapolation methods substantially beyond the polygon that includes all sensor locations will not provide meaningful results with reasonable accuracy. However, it is possible to develop additional approaches including analytical models or process information that might allow this request to be achieved – to a maximum of roughly the extensions of a plume. The development of such a method can be another area of future work.

One simplified proposed alternative was to analyse the evolution of gas plumes from successive maps of gas distribution created using Kernel DM+V. DCMR experts confirmed that this could help in obtaining estimates of the gas plume beyond the area covered by the sensors.

Plume in Combination with Source Location Information

Experts at DCMR pointed out that it would be interesting to investigate whether the gas distribution models created using GDM could be used to validate the plume detection to infer gas source locations in DGDS. We investigated this qualitatively and compared the result of GDM with the human reports as well as likely gas source location estimations from DGDS component on the same data set. The results are presented in Section 7.4.5. Further anal-
ysis and building an automatic inference system can be yet another promising area of future work.

7.3 Sensor Planning Module

In the Diadem use-case scenario, having a sensor planning module is crucial. This is because of the low density of the sensor measurements provided by the sparse sensor network. To increase the measurement density in critical areas, it is required to send field operators to these areas to collect additional measurements. As mentioned earlier, in the Diadem use-case scenario, measurements are collected continuously by the stationary sensors. These data are then combined with measurements collected by field operators for whom the Sensor Planning module suggests measurement locations.

Therefore, the aim of the Sensor Planning module is to identify places at which informative measurements can be collected, which are expected to minimise the number of additional measurements necessary to obtain a truthful representation of the gas distribution.

Parameters to be considered for exploration and exploitation in sensor planning were discussed with the experts in DEMA and DCMR. Areas with high gas concentration and the border of the plume are interesting targets to explore. This indicates that the proposed multi-objective sensor planning in Chapter 6 can be a good candidate to use in the Sensor Planning module in the Diadem framework.

The APF-based sensor planning approach, APFSP, that was presented in Chapter 6 directs the sampling attention to areas with high gas accumulation or high predictive variance. APF-based sensor planning uses gas distribution maps created by the GDM module to identify these areas. The objective function sets the importance factors in the sensor planning to have a higher plume coverage and a better quality of gas distribution models, which are the interests of the DCMR and DEMA experts. The Sensor Planning module uses these objectives to suggest the next sampling location based on the gas distribution model created by the GDM module and the record of previously selected sampling locations. Note that although the proposed sensor planning does not consider directing sampling towards the borders of the plume, having a good quality gas distribution model can lead to a better estimate of the plume and thus it can potentially provide information to experts to identify plume borders.

The proposed sensor planning solution can guide the process of sample collection as part of the Diadem framework. The objective function expresses a trade-off between exploration and exploitation, i.e. at the beginning of the sampling process, exploration will be more important because of the limited knowledge about the target area. When the model uncertainty decreases after a number of measurements are collected, exploitation in hot spot areas will become more important in the selection of sampling locations.
To plan further explorations and decide on further crisis management steps, DCMR experts consider additional parameters including geographical conditions such as rivers, buildings, etc. The current sensor planning approach does not consider these parameters; however, it allows for adding such objectives through an operator. This work can be implemented and evaluated in the future.

### 7.4 Gas Monitoring Component in Diadem

The gas monitoring component includes the GDM and Sensor Planning modules and aims to improve the assessment of the situation in Diadem-like scenario by providing information about the gas plume, and gas distribution in the investigation area, for example. Figure 7.2 illustrates the functionality of these two modules. GDM builds gas distribution maps using available samples. This map is communicated to the experts and relevant components through the Diadem framework. ARGOS visualises the map created by GDM. Furthermore, it enhances its physical models using this map. The Sensor Planning module suggests where to collect next samples using the outputs of GDM. This section explains the integration of the GDM and Sensor Planning modules in the Diadem framework.

#### 7.4.1 The GDM Module

Applying statistical gas distribution modelling approaches can provide a comprehensive view on a potentially large amount of data, helping experts to interpret the situation quickly. Because the approach makes only weak assumptions about a particular functional form of the gas distribution (e.g., it does not assume that source locations are known), complicated situations can be represented using the GDM module. In the Diadem use-cases, a network of gas sensors collects measurement data continuously. These data need to be interpreted online and in real-time to aid the overall situation assessment and point out locations for additional measurements.

As mentioned in Section 7.2.4, the Kernel DM+V method is used in the GDM module. In addition, the possibility to use a time-dependent method, TD Kernel DM+V, has been investigated. The timescale factor in the TD Kernel DM+V introduces recency weight for measurements based on their recency to create gas distribution model. In a highly fluctuating environment, like the one in the Diadem use-case scenario described earlier in this chapter, TD Kernel DM+V is not very successful in capturing time-variant properties of the gas distribution. Future work includes further investigation of using a time-dependent method in the GDM module.
Figure 7.2: Top: Network of stationary and mobile sensors collects measurements (left), which are sent to the GDM module (right). Bottom: The GDM module builds relative gas distribution maps which are sent to the experts and relevant components. ARGOS visualises these maps and possibly improves its physical model (right). The Sensor Planning module uses the output of GDM and suggests locations to acquire more samples (left).

### 7.4.2 The Sensor Planning Module

The Sensor Planning module provides suggestions to support experts in their decision regarding where to send field operators or firefighters to collect further measurements. Suggested locations are computed based on continuously updated gas distribution models obtained from the GDM module. The proposed sensor planning method introduced in Chapter 6, APFSP, is used in the Sensor Planning module. This method includes a number of objectives in sensor planning, such as even exploration, the tendency to investigate areas of increased gas accumulation, and unexplored areas. These areas are inferred from the gas distribution model. APFSP potentially allows for the experts to add additional objectives. The proposed sensor planning method does not include prior information about the environment such as obstacles and buildings. Extending the proposed approach to include spatial prior information such as obstacles is an area of study for future.
7.4.3 Integration in the Diadem Framework

The Diadem approach is a combination of complementary solutions which provide a mapping between gathered information, situation awareness, and decision-making. The integration of these solutions in the Diadem framework is addressed by using a service-oriented architecture named Dynamic Process Integration Framework (DPIF) [65]. DPIF provides a uniform service encapsulation, as well as an automated creation of information flows between the relevant experts and automated processes [65].

The Gas Monitoring component is integrated with other solutions in Diadem through DPIF wrapper agents. DPIF delivers to the Gas Monitoring component all the relevant information, and transports the output to relevant components of Diadem (see Figure 7.3). The developed GDM solution is running on a Web server and communicates with other components in Diadem through a DPIF agent. The input information is sent to the GDM as an XML query which includes field size and location, requested resolution, prediction time, and the value and location of the sensor measurements. These data are sent to GDM through DPIF to build a gas distribution model.

The output of GDM includes image files visualising predictive gas distribution mean and variance maps and an XML file which contains the corresponding estimated mean and variance values. These output files are uploaded to the GDM Web server and made accessible through DPIF to all interested parties and components that need the respective information. DPIF offers to download these output files to the relevant components.

Figure 7.3: Integration of the GDM module in DPIF with wrapping agents. DPIF allocates an agent with communication engine to each component. GDM sends and receives data from its web service to the agent allocated to GDM. DPIF communicates data between the GDM module and components or people that need the respective information through agents.
7.4.4 Integration with ARGOS

ARGOS uses the information from GDM to improve its initial assumptions, and thus provides better estimates of how the situation will evolve. In addition, ARGOS visualises the resulting relative concentration maps created by GDM in its graphical user interface to help experts to have a better assessment of gas dispersion in the vicinity of the incident site. This integration has been carried out both through DPIF and as a module directly in the ARGOS software. In the direct integration, GDM sends directly the relative concentration maps to ARGOS to be visualised there.

Figure 7.4 and 7.5 present two examples of this visualisation. The map, shown in Figure 7.4, was created based on the data from the "Diadem-II" experiment, collected on May 1, 2010, from 7:00 AM to 8:00 AM. The map presented in Figure 7.5(b) illustrates the visualisation of the GDM output in ARGOS in direction integration of GDM with ARGOS. This map was created using test measurements generated by ARGOS (Figure 7.5(a)).

![Figure 7.4](image1.png)

Figure 7.4: An example of GDM output visualised in ARGOS. Input data is from the "Diadem-II" dataset.

7.4.5 Evaluation

In the "Diadem-I" and "Diadem-II" datasets collected for the assessment of Diadem, there is no solid ground truth available, which makes a quantitative evaluation difficult. Therefore, we evaluated the consistency of the maps with reported gas sources and available wind information. A second means to evaluate the data is to ask expert users whether the produced maps correspond to their interpretation of the respective situation. This section presents the evaluation results.
7.4. GAS MONITORING COMPONENT IN DIADEM

Figure 7.5: (a) Collected measurements to be sent to GDM. (b) Overlay of GDM output in the graphical user interface of ARGOS.

Diadem-I

Figure 7.6 presents a relative concentration map created with the GDM module using part of the "Diadem-I" dataset collected on December 2, 2008, from 7:00 AM to 10:00 AM. The predictive mean map is then overlaid on the Google Earth map that marks the locations of sensors. In the studied period, the average wind direction was from west to east (left to right in Figure 7.6). The red region in Figure 7.6 indicates the area where the probability of high gas accumulation is the highest. This dataset does not have information about the gas source location(s) in the area. The DCMR experts could only confirm that the red region in Figure 7.6 corresponds to an area in which increased pollution is often found.

The sample maps were presented to DCMR experts. They found the gas distribution maps informative and confirmed that the provided maps can potentially assist them in interpretation of the crisis scenario. The DCMR experts pointed out that it would be even more informative to study a sequence of maps from successive time intervals. This was taken into account in the experiments with the "Diadem-II" dataset where the data collection is confined to a smaller area with a relatively large number of sensors (30 sensors). For detailed information about the experiment see Chapter 3.
CHAPTER 7. A LARGE-SCALE POLLUTION MONITORING APPLICATION

Diadem-II

From the "Diadem-II" dataset, two dates in which incidents were reported, were selected for analysis: scenario A (29th of June 2010), and scenario B (17th of May 2010). This section presents the results and analysis of using gas distribution mapping in these two scenarios.

In scenario A, we selected a subset collected on 29th of June 2010. This subset was selected because it included the data used in the Diadem demo (only the period from 6:30 AM – 7:15 AM was used) and results from the gas source detection component (DGDS) in the Diadem framework were available.

Figure 7.7 shows the predictive mean map overlaid on the corresponding Google Earth map of the area where sensors were located for the period from 6:30 AM to 7:15 AM in the scenario A. The minimum (blue) and maximum (red) in the mean map is 1.198 dB and 4.037 dB, respectively. The plume shape that is visible on the map is consistent with a likely gas source location that has been identified from the location of complaints, also shown in the figure by blue arrows.

Gas distribution maps were created for subsequent hours (00:00 AM to 1:00 AM, then 1:00 AM to 2:00 AM, etc.) and a video was compiled from these snapshots that showed the evolution of the gas distribution. The video was presented during the Diadem demo. Figure 7.8(a) and Figure 7.8(b) show the corresponding predictive mean and variance maps for the scenario A, and also the difference in between subsequent maps. These difference maps show that

Figure 7.6: Relative concentration map overlaid on the Google Earth map of the investigated area. Red indicates areas with a high probability of increased gas accumulation and the blue area indicates low gas accumulation. Black triangles (e.g. close to Geulhaven) illustrate wind directions.
the strongest change in the predictive mean of the gas distribution models was observed between the maps created for the two successive hours: 6:00 AM – 7:00 AM and 7:00 AM – 8:00 AM. In addition, the high values on the west sides of the difference maps in the predictive variance of the gas distribution model in these two time intervals aligns with the reported gas source location in this scenario (Figure 7.7). This observation is aligned with the reported incident at 6:30 AM - 7:15 AM.

In scenario B, a subset of the data collected on the 17th of May 2010 was selected for analysis because an incident happened at this day around 3:30 AM (at this time the first complaint was reported). A likely gas source location was also identified in this case (see the green target in Figure 7.9) and, according to the DCMR experts, the gas plume reached the area in which the measurement stations with the gas sensors are located at around 4:30 AM. The predictive mean map computed for the data collected from 4:00 AM to 5:00 AM is shown in Figure 7.9. Again, the statistical gas distribution model seems to be consistent with the likely gas source location.

To highlight changes in the gas distribution model over time, distribution models were computed from data collected by the sensor network for every hour in the period of 00:00 AM - 10:00 AM for the scenario B. Figure 7.10(a)
CHAPTER 7. A LARGE-SCALE POLLUTION MONITORING APPLICATION

(a) The consecutive generated predictive mean maps.

(b) The consecutive generated predictive variance maps.

Figure 7.8: (a) Top: The sequence of predictive mean maps. Bottom: The difference between successive predictive mean maps. (b) Top: The sequence of predictive variance maps. The difference between successive maps. The predictive mean and variance maps are created using samples from 00:00 AM to 10:00 AM in the scenario A. The maps are created from the collected samples after every hour (from MOX sensor #21 mounted on each node of the sensor network). The red rectangles indicate the the strongest difference between consecutive maps.

and Figure 7.10(b) show the corresponding predictive mean and variance maps and also the difference in between subsequent maps of the scenario B. These difference maps show that the strongest change was observed between the two successive hours: 3:00 AM – 4:00 AM and 4:00 AM – 5:00 AM, both in the predictive mean and the predictive variance of the gas distribution model. In addition, the difference maps show that the area in which the largest differences occur are detected at the upper boundary of the area that is covered by the sensor network. This is consistent with the recorded complaints, the location of the likely gas source, and the hypotheses about gas source locations identified by DGDS (Figure 7.9).

The results in the two discussed scenarios indicate that analysing the changes in the gas distribution over time can help in identifying incidents. In particular, in both scenarios looking at difference maps between consecutive predictive mean maps seemed promising in identifying incidents. In addition, the results suggest that analysing the changes in the gas distribution models over time can potentially be used as supporting information for detecting likely gas source locations. Investigating this further is an area for future work.
Figure 7.9: Predictive mean map overlaid on the corresponding Google Earth map of the area. The green target indicates a gas source of the likely associated case 10152632 in the scenario B. Bright blue arrows indicate complaints that are presumably related to the case. The other coloured arrows indicate gas source location hypotheses identified by the DGDS component in Diadem framework. Green pins mark sensor locations.

7.5 Discussion and Conclusion

7.5.1 Diadem

Prevention and mitigation of environmental conditions with potentially adverse effects on the population and ecosystems requires (1) identification of critical situations, (2) impact assessment taking into account possible evolutions of physical processes, (3) planning and evaluation of preventive and mitigation measures and (4) decision-making to select appropriate actions. This can only be achieved through adequate processing of large quantities of heterogeneous information based on rich expertise about diverse aspects of the physical world. To be able to process such quantities of information, automated solutions to processing must be used. However, the assessment and decision-making processes cannot be fully automated due to the domain complexity and lack of complete models. In addition, full automation might not be acceptable in many real-world organisations where people need to be in charge and be accountable. Instead, the Diadem approach supports evolutionary automation: automate wherever possible and keep the humans in the processing loop wherever automated solutions cannot be made reliable [5].
The technical solutions contributed by the Diadem project are not limited to the chosen use-cases. In fact, many of the Diadem solutions are generic and can be used in different contexts and different application domains. For instance, the GDM module, developed within the work on this dissertation, is based on a generic approach for estimating distributions based on measurements that are interpreted as samples generated by a random process. The developed algorithms can thus be used not only in the context of crisis management and airborne gases but also in other relevant contexts such as the assessment of ocean water salinity for example (see Chapter 1).

Overall, by integrating the aforementioned Diadem solutions, the reactivity, coverage and reliability of the environmental crisis management processes can significantly improve. As larger quantities of relevant information of different types and from larger areas can be obtained and processed in a timely manner, more reliable conclusions about the situation can be drawn in a shorter time, allowing better-informed decisions throughout the incident management process.

Results obtained in the Diadem project open up new opportunities in different domains such as security, knowledge management, collaborative problem
solving and complex decision-making. In these domains, important problems that can be addressed with the Diadem solutions include the detection of disease outbreaks, the planning and evaluation of governmental policies, financial decision-making, climate modelling, and the economic impact of terrorism counter-measures, to name a few.

7.5.2 GDM and Sensor Planning in Diadem

The Diadem project was an opportunity to assess the performance of statistical gas distribution modelling and sensor planning methods in a real-world application. Furthermore, it helped to identify the challenges in a real-world environment. More importantly, we came to understand the challenges that a customer would like a Gas Monitoring solution to solve for them, what can be approximated, and where the information has to be combined with other available information. Last but not least, it led us to find a potential direction of work in a real-world application.

In Diadem, gas measurements were collected using a sparse stationary network of sensors. These sensors were uncalibrated. Given this setup, GDM could only provide relative gas distribution maps. While the sensor network proved to be too sparse and a solid ground truth evaluation was not possible at such a large scale, the study in this chapter indicated that gas distribution mapping nevertheless turned out to be useful as a comprehensive overview tool for the experts in environmental management incident especially when the gas source location and type of gas are unknown. DEMA and DCMR experts highlighted during Diadem demos that GDM can improve the interpretation of the overall situation in case of a gas emission incident. One particularly relevant conclusion was that looking at the sequence of gas distribution maps can provide insightful information to the experts to detect situations in which larger changes occur and to help them to predict further development in the gas emission. Extending the GDM work in this thesis to analyse automatically the difference between consecutive maps and to detect larger changes in the situation is an area for future work. DCMR experts expressed that showing local wind information together with GDM would help them to estimate better the evolution of the gas dispersion. The wind information is however not available at sensor locations. Future work will investigate modelling wind fields and will use this information to estimate the development of plume.

The possibility of using the proposed sensor planning in this chapter was investigated and discussed with experts. Sensor planning can guide field investigators to choose where to acquire samples. One of the advantages of the proposed sensor planning is that it allows to gear the sensor planning based on the objectives to select better candidates for the next sampling point. In the Diadem demos, we could only demo the potential of the sensor planning to be integrated with other components through DPIF since mobile sensor data were not available. DCMR experts expressed their interest in the sensor planning
approach to provide suggestions on where field investigators should acquire the next measurements. The proposed sensor planning in this dissertation directs sample collection to have a higher coverage of the area and the areas with higher gas accumulation. Experts indicate that in addition to the areas with high gas accumulation, collecting more information from the plume borders is of interest. Future work should investigate how to incorporate sampling on the plume boundaries as an objective function in the sensor planning.
Chapter 8

Conclusion

This thesis addresses the problem of gas distribution modelling for gas monitoring and gas detection. In the following, first, contributions of this thesis are summarised. Then, limitations of the methods proposed in this thesis are discussed. Finally, directions for future work are outlined.

8.1 Contributions

This section presents contributions presented in this thesis on (time-invariant) gas distribution modelling, time-independent gas distribution modelling, sensor planning, and real-world applications.

8.1.1 Gas Distribution Modelling

In this thesis, first, a review of existing works in the field of statistical gas distribution modelling (GDM) is presented. Mainly, we focused on GDM methods which model distribution variance in addition to distribution mean of a gas distribution. The spatial structure of the distribution variance can provide important information about the gas distribution by highlighting areas of high fluctuation, which are often found in close vicinity to a gas source. We compared the performance of different GDM methods using real-world datasets collected in controlled and uncontrolled indoor environments. We used an evaluation measure called NLPD, to learn meta-parameters of gas distribution models and to evaluate the quality of models.

Most existing GDM methods assume that gas distribution is a time-constant random process. Such time-invariant approaches cannot model well evolving gas plumes, for example, or major changes in gas dispersion due to a sudden change of the conditions of the environment. As one of the main contributions of this thesis, two time-dependent GDM methods were proposed: (1) temporal (sub-)sampling and (2) a time-dependent gas distribution modelling method which relates recency of measurements to the prediction time. The analysis of
the first approach, temporal (sub-)sampling, showed that, models created using samples from a more recent time scale perform better than models created using samples from a larger time scale. The second proposed time-dependent gas distribution modelling method, TD Kernel DM+V, introduces a recency function that weighs samples based on their temporal difference to the prediction time to build a gas distribution model. To select a function that weighs the recency of measurements in TD Kernel DM+V, we compared the performance of TD Kernel DM+V using different recency functions. The results indicated an exponential function as the best choice among the tested recency functions. Therefore, an exponential function was chosen as recency weight in the rest of this thesis work. The exponential recency function is defined with a timescale parameter. The timescale is selected together with other meta-parameters by optimising NLPD. Evaluations were performed in two real-world experiments as well as several simulation experiments. The results showed that TD Kernel DM+V improves the quality of the created gas distribution models, especially in dynamic environments compared to existing time-invariant gas distribution modelling approaches. Furthermore, we compared the performance of TD Kernel DM+V with Kernel DM+V in the conditions where the gas plume is evolving and where the gas plume has fully developed. In these experiments, TD Kernel DM+V consistently outperformed Kernel DM+V in both conditions.

8.1.2 Sensor Planning

Another major contribution of this thesis is proposing an artificial potential based sensor planning method, APFSP. The introduced sensor planning method applies a modified artificial potential fields which balances objectives related exploration and exploitation. The potential field attracts towards areas with high values of predictive mean and variance. High values of predictive mean and variance can indicate areas with high gas accumulation and potential gas source locations, respectively. The repulsion prevents repeated measurements at the same locations, and thus, it promotes exploration in areas where the certainty of the created model is low.

The importance of these objectives is set by using three meta-parameters. In this thesis, we analyse the influence of the meta-parameters on the quality of the created gas distribution model, coverage of the investigation area, plume coverage (i.e. collecting samples inside the plume), travelling distance, and distance of the latest sampling location to the gas source location. The investigations are performed using simulation experiments that provide dense ground truth. The results indicate that a higher weight for exploring areas with higher uncertainty in the created gas distribution model leads to a better area coverage and better quality for GDM. On the other hand, a lower weight for exploring areas with high uncertainty together with a higher weight
for predictive mean or variance results in a better exploitation, e.g. a higher plume coverage or a shorter total travelling distance in the sampling period.

We compared the performance of APFSP with other commonly used sensor planning methods such as random sampling and sampling along a predefined sweeping trajectory. The results indicated that applying the APFSP sampling strategy performs better when it comes to the quality of created gas distribution models and plume coverage. Moreover, the evaluations presented in this thesis showed that applying a locality constraint decreases the travelling distance which makes the proposed sensor planning approach a suitable method for real-world applications where limited resources and time are available.

In this thesis, the APFSP method was analysed using Kernel DM+V to create gas distribution models. In an experiment, we replaced Kernel DM+V with the time-dependent GDM, TD Kernel DM+V. At the beginning of sampling when only a small number of samples are collected, gas distribution modelling relies on all measurements, and using Kernel DM+V leads to better gas distribution models. However, as more samples are collected, using TD Kernel DM+V begins to perform better. In addition, APFSP sampling using TD Kernel DM+V results in a shorter distance of the selected sampling location to the gas source as more samples are collected compared to using Kernel DM+V.

8.1.3 Real-world Application

We assessed the feasibility of using GDM and sensor planning in a real-world scenario. GDM and sensor planning solutions were developed and integrated into a framework for management of a crisis in which gas emission is involved. In this scenario, samples are very sparse. Therefore, GDM can only provide a low-resolution model. However, it was found that GDM can help experts as an assistive tool to understand the gas distribution and plan where to inspect further.

8.2 Limitations

This section discusses limitations of the presented gas distribution modelling and sensor planning methods, in particular when it comes to real-world applications.

8.2.1 Gas Distribution Modelling

The gas distribution models presented in this thesis reply on some assumptions. First, the presented GDM methods assumed that the sensors only pick up only the target gas and the sensors are calibrated.

Second, the proposed time-dependent GDM method, TD Kernel DM+V does not consider spatial information about the presence of obstacles when
creating grid cells and predicting mean and variance; this affects the quality of created models when an obstacle is present.

8.2.2 Sensor Planning

The proposed sensor planning provides a trade-off between exploration and exploitation. The meta-parameters weigh three terms corresponding to the predictive mean, predictive variance and confidence maps in the utility function. In this thesis, we analysed the effect of meta-parameter selection on the quality of gas distribution models, area coverage, plume coverage, travelling distance, and distance to the gas source location.

The proposed sensor planning does not consider spatial information about the environment such as obstacles.

8.2.3 Real-world Application

The proposed sensor planning and gas distribution modelling approaches were developed and integrated in the Diadem project for decision-making in hazardous incidents where gas leakage is involved. The presented solutions in Diadem were used in two real-world use-cases. We made a few assumptions to integrate gas distribution models in the crisis management framework in these use-cases. First, we assumed that all samples are collected at the same height, while this is not the case in most real-world scenarios. Gas distribution models are created in two dimensions, and in this thesis, we do not incorporate the height of the sampling locations when creating a gas distribution model.

Second, in the Diadem scenario, sensors are not calibrated; therefore, only a relative gas distribution model can be provided. Therefore, in practice, GDM can serve only as a comprehensive overview tool for the experts in environmental management incident especially when the gas source location and type of gas are unknown.

Finally, since no ground truth was available in the Diadem scenario, only qualitative evaluations could be carried out.

8.3 Ethics

This thesis presented temporal and spatial sampling methods to create accurate gas distribution models. These methods enable applying statistical gas distribution modelling in a wide range of applications. Example application include air pollution monitoring, leakage detection, and search and rescue operations. The focus of this thesis is on applying gas distribution modelling for air quality monitoring, hoping that this research leads to a healthier environment.
In this research, only gas sensor measurements are used, except in the Diadem use-cases where human reports are used as ground truth. All available human reports to this research were anonymised to protect privacy.

## 8.4 Future Work

This thesis presented novel gas distribution modelling and sensor planning methods to enable a wide range of applications. However, there are several future work directions which require further considerations. This section outlines future work on time-independent gas distribution modelling, sensor planning, and real-world applications.

### 8.4.1 Gas Distribution Modelling

The gas distribution modelling method presented in this thesis includes spatial information only about samples when creating gas distribution models. Moving to more realistic scenarios, incorporating spatial information about the environment, e.g. location of obstacles, to the gas distribution model should be considered in future research work.

In the presented simulation results, the gas source was located at a fixed height close to the ground, and one single type of gas was used as a gas source. The impact of changing gas source and gas compound on the performance of the GDM methods should be investigated in the future work.

Experiments in more realistic simulations with multiple gas sources and real-world environments with more complex spatial structure are required to assess further the performance of gas distribution models for real-world applications.

The time-dependent GDM study showed that selection of the time scale in sub-sampling affects the performance of kernel extrapolation based gas distribution modelling. Introducing a good temporal sub-sampling method which finds the best measurement subset is the next step that has to be addressed. One potential solution for sub-sampling is to select sampling intervals when larger fluctuations are observed in the gas distribution. One idea to identify these intervals is to divide the sampling period into intervals and create snapshots from gas distribution for each interval. Comparison of consecutive pairs of intervals can provide a measure to find a suitable subset of measurements. Future work should investigate further the selection of time span for sampling in a simulated experiment where a varying wind flow is present. To compare consecutive maps, one could use Kullback-Leibler distance. Another direction for future works is to study information that may become visible in difference maps computed from a sequence of TD Kernel DM+V maps that are derived from consecutive intervals in time.
This thesis presented the initial results of investigating dependency on the target time. However, further analysis with a wider range of time scales and more complex simulation experiments needs to be considered as next steps.

### 8.4.2 Sensor Planning

The presented sensor planning method, APFSP, uses information from the gas distribution model to plan the next sampling location. In the sensor planning, every time that a new sample is collected, the gas distribution model is updated, and the next sampling location is planned accordingly. In this thesis, the meta-parameters of the gas distribution model are set heuristically to the optimised values found in the study of gas distribution models with more samples. The values of the meta-parameters do not change during the sampling iterations. Future work will investigate how to select optimal values of meta-parameters for gas distribution model online.

In this dissertation, I discussed the impact of the meta-parameters on the performance of sensor planning using different measures. However, we set the meta-parameters in APFSP heuristically. Future work should include investigating the selection of optimal relative weights (meta-parameters) for each objective in the computation of sampling locations. Intuitively, one can consider setting the meta-parameters first to explore more and then, as more samples are collected, to exploit in areas with high predictive mean and variance as more samples are collected.

In this thesis, we investigated the performance of APFSP and the selection of the meta-parameters mainly in simulation experiments. As a next step, future work should investigate the performance of APFSP in more real-world experiments. Moreover, future work should assess the performance of APFSP in more complex simulations, e.g. where multiple gas sources and stronger fluctuations are present.

One way to take into account the temporal properties of gas dispersion in sensor planning is to use a time-dependent GDM method. In this thesis, we investigated the performance of APFSP when using TD Kernel DM+V to create gas distribution models. Another way to consider time-dependency is that the virtual charge which scales the strength of the repulsive potential exerted from previous measurement points should also be time-dependent; namely, it should be lower for earlier measurements.

With respect to real-world applications, future work should investigate how robust the proposed approach is to changes in wind direction and variations in wind intensity. Future work could for example use the idea of Kernel DM+V/W to extend TD Kernel DM+V and incorporate local wind information in the gas distribution model, which is used for planning the next sampling locations [20, 51].
Incorporating additional spatial information such as obstacles in sensor planning will be an inevitable step when moving to real-world experiments using mobile sensors.

### 8.4.3 Real-world Application

In the real-world use-case, Diadem, local wind measurements were not available at the same location where gas measurements were collected. Modelling the wind field to estimate wind intensity or direction at the gas measurement location should be considered in the future work.

In Diadem, we assumed that measurements are collected at the same height. A future research direction is to consider measurements from different heights. Using a 3D kernel similar to the idea of 3D Kernel DM+V [20] can be explored as a solution in future work.

A possible direction of future work is to incorporate events reported in Diadem-like scenarios to build a gas distribution model. Bayesian Spatial Event Distribution gas distribution mapping, presented in the work of Schaffernicht et al. [19] can be investigated as a possible solution to consider events in gas distribution models for Diadem-like scenarios.

In the Diadem scenario, no ground truth data were available to evaluate the quality of the created gas distribution models. Future work should investigate methods to consider reported events as a “week ground truth” to assess the performance of GDM.

Finally, future research should consider adding a gas discrimination component to GDM. One solution that can be is to use a gas distribution modelling method as presented in the work of Bennetts et al. [27], which considers multiple target gases.
Bibliography


Publications in the series
Örebro Studies in Technology


