Chapter 6
Statistical Gas Distribution Modeling Using Kernel Methods

Sahar Asadi
Örebro University, Sweden

Matteo Reggente
Örebro University, Sweden

Cyrill Stachniss
University of Freiburg, Germany

Christian Plagemann
Stanford University, USA

Achim J. Lilienthal
Örebro University, Sweden

ABSTRACT

Gas distribution models can provide comprehensive information about a large number of gas concentration measurements, highlighting, for example, areas of unusual gas accumulation. They can also help to locate gas sources and to plan where future measurements should be carried out. Current physical modeling methods, however, are computationally expensive and not applicable for real world scenarios with real-time and high resolution demands. This chapter reviews kernel methods that statistically model gas distribution. Gas measurements are treated as random variables, and the gas distribution is predicted at unseen locations either using a kernel density estimation or a kernel regression approach. The resulting statistical

DOI: 10.4018/978-1-61520-915-6.ch006

models do not make strong assumptions about the functional form of the gas distribution, such as the number or locations of gas sources, for example. The major focus of this chapter is on two-dimensional models that provide estimates for the means and predictive variances of the distribution. Furthermore, three extensions to the presented kernel density estimation algorithm are described, which allow to include wind information, to extend the model to three dimensions, and to reflect time-dependent changes of the random process that generates the gas distribution measurements. All methods are discussed based on experimental validation using real sensor data.

INTRODUCTION

Modeling the distribution of gas in an environment aims at deriving a truthful representation of the observed gas distribution from a set of spatially and temporally distributed measurements of relevant variables, foremost gas concentration, but also wind, pressure, and temperature (Lilienthal et al., 2009b). The task of building gas distribution models is challenging mainly because of the chaotic nature of gas dispersal. The complex interaction of gas with its surrounding 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 in the long run. Second, turbulent air flow fragments the gas emanating from a source into intermittent patches of high concentration with steep gradients at their edges (Robert & Webster, 2002). 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. These effects are especially important in uncontrolled environments.

Besides the physics of gas dispersal, limitations of gas sensors also make gas distribution modeling difficult. Gas sensors provide information about a small spatial region only since the measurements require direct interaction between the sensor surface and the analyte molecules. Therefore, instantaneous measurements of gas concentration 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.

Gas distribution modeling (GDM) methods can be categorized as model-based and model-free. Model-based approaches infer the parameters of an analytical gas distribution model from the measurements. In principle, Computational Fluid Dynamics (CFD) models can be applied to solve the governing set of equations numerically. 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. Many other model-based ap-

proaches were developed for atmospheric dispersion (Olesen et al., 2005). Such models typically also cannot efficiently incorporate sensor information efficiently on the fly and do not provide a sufficient level of detail.

Model-free approaches do not make strong assumptions about a particular functional form of the gas distribution. They typically treat gas sensor measurements as random variables and derive a statistical model of the observed gas dispersion from the measurements.

In this chapter, we introduce and review kernel methods for statistical, model-free gas distribution modeling and present results in both indoor and outdoor environments. Section "Background" presents an overview on previous statistical gas distribution modeling methods. In section "Kernel Extrapolation Distribution Mapping", the kernel extrapolation approach (Kernel DM) will be described. Sections "Kernel DM+V" and –"Gaussian Process Mixture Model" review the Kernel DM+V algorithm and a Gaussian Processes Mixture approach as two statistical gas distribution modeling approaches to estimate both mean and variance of the distribution. A comparison of these two approaches is presented in section" Evaluation and Comparison of Gas Distribution Models". Compared to indoor environments, outdoor environments present additional challenges such as a strong and unstable wind field. Section " Kernel DM+V/W – Using Wind Information for Gas Distribution Mapping" describes a modification of the basic Kernel DM+V algorithm, which allows including wind information during model creation. Sections " Three-Dimensional Statistical Gas Distribution Mapping" and " Time-dependent Gas Distribution Modeling, TD Kernel DM+V" describe further modifications of the basic Kernel DM + V algorithm that extend the models to three dimensions and include a time-dependent component, respectively. The chapter ends with conclusions and suggestions for future work.

BACKGROUND

Several methods for statistical gas distribution modeling have been published. A straightforward solution to obtain a model of the time-averaged gas distribution is to use a dense grid of sensors. In (Ishida et al., 1998), a gas distribution model is created from measurements collected by a grid of sensors. These measurements are averaged over a prolonged period of time and discretized to a grid that represents the topology of the sensor network. A similar method is presented in (Purnamadjaja & Russell, 2005), where maximum values of the measurement interval are mapped instead of average concentrations. An alternative to a network of stationary sensors is to use a single mobile sensor. While a network of sensors has advantages in terms of coverage and sampling time, using a mobile sensor avoids calibration issues and allows for adaptive sampling of the environment. In practice, it is beneficial

to combine stationary sensor networks with autonomous mobile sensors. Pyk et al. (2006) create a gas distribution map by using a single sensor, which collects measurements consecutively instead of the parallel acquisition of readings in a sensor network. This method has been applied in an experiment in a wind tunnel. At each pre-specified measurement 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. The drawback of such interpolating methods is that there is no means of averaging out instantaneous response fluctuations at measurement locations. This leads to increasingly jagged distribution maps. In (Hayes et al., 2002), a group of mobile robots equipped with conducting polymer sensors were used to create a histogram representation of the distribution of water vapor created by a hot water pan behind a fan. The histogram bins collect the number of odor hits received by all robots in the corresponding area while they performed a random walk behavior. Odor 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 by Lilienthal and Duckett (2004), introduced in the following section, 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 a Parzen window.

KERNEL EXTRAPOLATION DISTRIBUTION MAPPING

The Kernel DM method by Lilienthal and Duckett (2004) 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 belief 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 readings since the sensor response represents interactions at the surface of the sensors, i.e. at a very small area. A key idea of Kernel DM is therefore

how to extrapolate the sensor readings beyond this small area by reasoning over the decreasing information content a single sensor reading provides about the mean estimate at locations in a certain distance from the sensor surface.

Kernel DM discretizes 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 reading 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.

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 $(1\leq i\leq n)$, 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 $\delta_i^{(k)} = \left|x^{(k)} - x_i\right|$. Using this distance, the value of the kernel for a given measurement is computed for each cell by

$$\omega_i^{(k)} = \mathrm{N}(\delta_i^{(k)}, \sigma), \tag{1}$$

where σ is the kernel width. To enhance the performance in practice, the evaluation of Equation (1) 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

$$\Omega^{(k)} = \sum_{i=1}^{n} \omega_i^{(k)}, \tag{2}$$

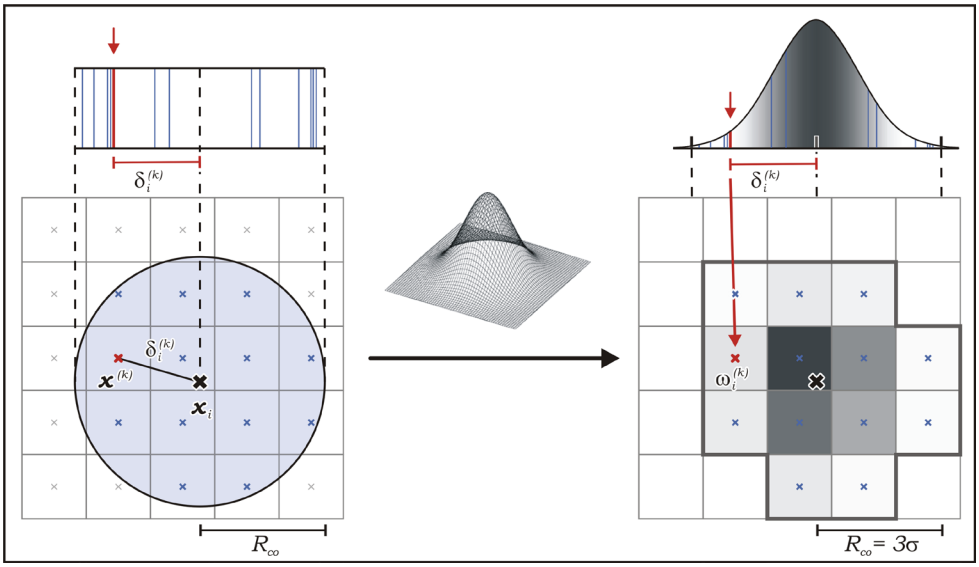
$$R^{(k)} = \sum_{i=1}^{n} \omega_i^{(k)} r_i. \tag{3}$$

Finally, the mean estimate is computed as

$$R^{(k)} = R^{(k)} / \Omega^{(k)} : \Omega^{(k)} \geq \Omega_{\min}. \tag{4}$$

The threshold $\Omega_{\min}$ makes sure that estimates are based on a sufficient amount of measurements collected in the vicinity of the estimate location. The Kernel DM algorithm depends on three parameters: the width σ of the Gaussian function, the

Figure 1. Discretization of the Gaussian kernel onto the grid



cut-off radius R_{co} and the threshold Ω_{min}. Only choosing an appropriate value of σ 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.

Figure 1 shows how a single reading is convolved onto a 5×5 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 1 by a surrounding strong border. The weightings for these cells are determined by evaluating the Gaussian function for the displacement values. These weights are represented in Figure 1 by shadings of grey. Darker shadings indicate higher weights, which corresponds to a stronger contribution of the measurement value r_i in the calculation of the average concentration value for a particular cell.

Lilienthal and Duckett (2004) applied the Kernel DM method to experiments in an indoor environment where data were collected with an array of sensors mounted on a mobile robot. The experimental setup includes a single target gas, but if a gas discrimination component is included gas distribution mapping can be extended to the case of multiple different gas sources, for example as described in (Loutfi et al., 2008). In the experiment presented in Lilienthal and Duckett (2004), 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 SLAM framework as described in (Lilienthal et al., 2007).

KERNEL DM+V

Kernel DM computes an estimate of the mean gas distribution. However, also 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 the gas source. Kernel DM+V proposed by Lilienthal et al. (2009b) is a statistical approach which models mean and variance distribution. As it will be discussed in Sec. " Evaluation and Comparison of Gas Distribution Models", estimating the predictive variance also enables better quantitative evaluation of alternative statistical models and provides the base to learn meta-parameters. The predictive variance can also be beneficial for sensor planning based on the current model.

Like in Kernel DM, the distribution is represented as a grid map in Kernel DM+V and a univariate 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 and $R^{(k)}$ by integrating weighted readings

$$\omega_i^{(k)} = \mathrm{N}\left(\left|x_i - x^{(k)}\right|, \sigma\right), \tag{5}$$

$$\Omega^{(k)} = \sum_{i=1}^{n} \omega_i^{(k)}, \tag{6}$$

$$R^{(k)} = \sum_{i=1}^{n} \omega_i^{(k)} r_i. \tag{7}$$

Here, $x^{(k)}$ denotes the centre of the cell k and the kernel width σ is a parameter of the algorithm. The integrated weights $\Omega^{(k)}$ are used to normalize the weighted readings $R^{(k)}$ and to compute a further map $\alpha^{(k)}$, which represents the confidence in the obtained estimates.

$$\alpha^{(k)} = 1 - e^{-\left(\Omega^{(k)2}/\sigma_\Omega^2\right)}, \tag{8}$$

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

$$r^{(k)} = \alpha^{(k)} \frac{R^{(k)}}{\Omega^{(k)}} + \left\{1 - \alpha^{(k)}\right\} r_0 \tag{9}$$

where r_0 represents the mean concentration estimate for cells for which there is not sufficient information available from nearby readings indicated by a low value of $\alpha^{(k)}$. In (Lilienthal et al, 2009b), 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 σ_Ω^2, 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 $V^{(k)}$

$$V^{(k)} = \sum_{i=1}^{n} \mathrm{N}\left(\left|x_i - x^{(k)}\right|, \sigma\right)\left(r_i - r^{(k(i))}\right)^2, \tag{10}$$

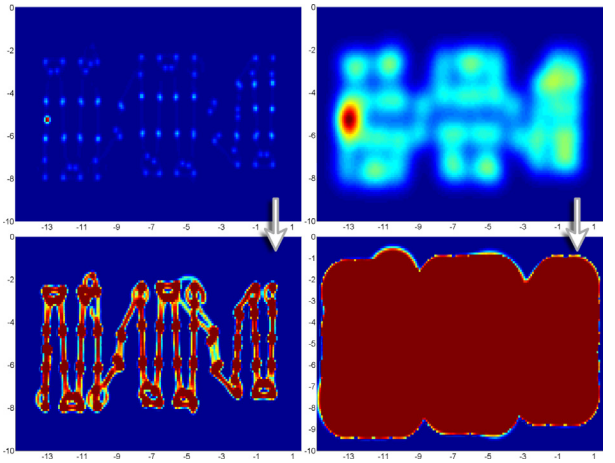
$$v^{(k)} = \alpha^{(k)} \frac{V^{(k)}}{\Omega^{(k)}} + \left\{1 - \alpha^{(k)}\right\} v_0, \tag{11}$$

where v_0 is an estimate of the distribution variance in regions far from measurement points. In (Lilienthal et al, 2009b) v_0 is computed as the average over all variance contributions.

The Kernel DM+V algorithm depends on three parameters: the kernel width σ, which governs the amount of extrapolation on individual sensor measurement (and the complexity of the model); the cell size c that determines the resolution at which different predictions can be made, and the scaling parameter σ_Ω, 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 σ_Ω 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 σ_Ω is not critical as long as it is of the right scale. Kernel width and cell size can be learned from the collected measurements. Learning meta-parameters of Kernel DM+V is discussed in Sec. ” Evaluation and Comparison of Gas Distribution Models”.

Figure 2 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 σ_Ω (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

Figure 2. Weight map $\Omega^{(k)}$ (top row) and the corresponding confidence map $\alpha^{(k)}$ (bottom row) obtained using parameters σ=0.10m, $\sigma_\Omega = 5.0 \cdot N(0, 0.10) \approx 20.0$ (left column) and σ=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.



close to the actual measurement. For wider kernels or smaller values of σ_Ω (right column), the area for which predictions based on extrapolation are made is larger.

GAUSSIAN PROCESS MIXTURE MODEL

Another approach to estimate the predictive variance in gas distribution modeling is the Gaussian Process Mixture model (GPM) proposed by Stachniss et al. (2009). Gaussian Processes (GPs) are non-parametric probabilistic models for solving regression problems. 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 neighboring data points. More formally, assuming that $\{(x_i, f_i)\}_{i=1}^{n}$ are samples of a Gaussian process, $f = (f_1, f_2, \ldots, f_n)^T$ has a Gaussian distribution

$$f_i = f(x_i) := \mathrm{N}(\boldsymbol{\mu}, \mathbf{K}),\ \boldsymbol{\mu} \in \mathrm{R}^n,\ \mathbf{K} \in \mathrm{R}^{n \times n} \tag{12}$$

where $\boldsymbol{\mu}$ is the mean and $\mathbf{K}$ corresponds to the covariance matrix which is constructed using a so-called covariance function k. For reasons of simplicity, one often assumes that $\boldsymbol{\mu}$=0. A standard choice for the covariance is the exponential function

$$[\mathbf{K}]_{ij} = k(x_i, x_j) = \sigma_f^2 \exp\left(-\frac{1}{2}\frac{\left|x_i - x_j\right|^2}{l^2}\right), \tag{13}$$

in which l is the length-scale parameter that influences the smoothness of the function f and σ_f^2 is the signal variance parameter. The parameter σ_n^2 represents the global noise. The parameters l, σ_f^2, and σ_n^2, termed the "hyper parameters" of the GP model, have to be learned. Other popular choices for the covariance function are Matern kernels or neural network kernel functions.

Let $\mathbf{X}=[x_1;\ldots;x_n]^T$ and $\mathbf{R}=[r_1;\ldots;r_n]$ be the set of locations and the corresponding measurements respectively. An arbitrary test set $\mathbf{X}_*$ is chosen from the input data set. As described in Section "Kernel Extrapolation Distribution Mapping", 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 modeling is to predict accurately the value of r_* at an unseen location $x_* \in \mathbf{X}_*$.

To predict the value of the measurement at x_*, predictive mean and variance are estimated as

$$\overline{f}(\mathbf{X}_*) \coloneqq \mathrm{E}[f(\mathbf{X}_*)] = k(\mathbf{X}_*, \mathbf{X})[k(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathrm{I}]^{-1}\mathbf{r}\,, \tag{14}$$

and respectively,

$$V[f(\mathbf{X}_*)] = k(\mathbf{X}_*, \mathbf{X}_*) - k(\mathbf{X}_*, \mathbf{X})[k(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathrm{I}]^{-1} k(\mathbf{X}, \mathbf{X}_*)\,. \tag{15}$$

The main drawback of GPs is their computational complexity. To estimate the predictive mean in Equation (14) and the predictive variance in Equation (15), the inversion of $[k(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathrm{I}]^{-1}$ has to be computed. This inversion has $O(n^3)$ time complexity where n is the size of the training data set. Another limitation of using GPs in the context of gas distribution modeling is that standard GPs create a unimodal distribution for any given location while 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.

One solution to overcome unimodality is using Gaussian Process Mixture models. A GPM model is a locally weighted sum of GP models. Let $\{GP_i\}_{i=1}^m$ be a set of GPs as individual components of the mixture model. To predict the value of r_* at the location x_*, first, it has to be estimated that to which component this measurement belongs. Then, using the distribution model of the corresponding component

the value of the measurement r_* can be predicted. A *gating function* $\mathrm{P}(z(x_*)=i)$ is defined as the probability that x_* is associated with the component GP_i. The likelihood of observing r_* is achieved by

$$h(x_*) := p(r_* \mid x_*) = \sum_{i=1}^{m} \mathrm{P}(z(x_*) = i)\mathrm{N}_i(r_*; x_*), \tag{16}$$

where $\mathrm{N}_i(r;x)$ denotes the Gaussian distribution function with mean $\overline{f}(x)$ and the corresponding variance $V[f_i(x)] + \sigma_n^2$. To create the GPM model, the predictive mean and variance are obtained by

$$\overline{h}(x_*) := \mathrm{E}[h(x_*)] = \sum_{i=1}^{m} \mathrm{P}(z(x_*) = i)\overline{f}_i(x_*), \tag{17}$$

$$V[h(x_*)] = \sum_{i=1}^{m} \mathrm{P}(z(x_*) = i)\left(V[f_i(x_*)] + (\overline{f}_i(x_*) - \overline{h}(x_*))^2\right). \tag{18}$$

The predictive model of components and the gating function are learned using the Expectation Maximization (EM) method. The following describes in detail the algorithm to build the GPM model for a given data set.

Initialization 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 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, an error function GP_Δ is defined. This error function is learned using data from D excluding the training set n_1. GP_Δ captures the absolute difference between a set of target values and the prediction of GP_1. The next component is initialized from the data which indicated higher error in the created model. The sub sampling 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 modeling problem. The two components of the GPM represent a background signal and peaks.

Iterative Learning via Expectation-Maximization Method

The Expectation-Maximization (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 ($x_j \in \mathbf{X}$) corresponds to model component i. This is done by re-estimating the gating function for each data point j as

$$\mathrm{P}(z(x_j)=i) \leftarrow \frac{\mathrm{P}(z(x_j)=i)\mathrm{N}_i(r_j;x_j)}{\sum_{k=1}^{m}\mathrm{P}(z(x_j)=k)\mathrm{N}_k(r_j;x_j)}. \tag{19}$$

Then, in the next step, the components are updated based on the new estimate of $\mathrm{P}(z(x_j)=i)$. To update the component i, the predictive mean and variance of the mixture model are computed. This is achieved by modifying Equation (14) to

$$\bar{f}_i(\mathbf{X}_*) = k(\mathbf{X}_*,\mathbf{X})[k(\mathbf{X},\mathbf{X}) + \ddot{} \ ^{i}]^{-1}\mathbf{r}\,, \tag{20}$$

where

$$[\ddot{} \ ^{i}]_{jj} = \frac{\sigma_n^2}{\mathrm{P}(z(x_j)=i)} \tag{21}$$

The mixture model has three hyper parameters $\{\sigma_f, l, \sigma_n\}$ which have to be estimated. A common way is to first initialize these values using a heuristic and then optimizing the values. A reasonable heuristic is the one proposed by (Snelson and Ghahramani, 2006), which is also used in (Stachniss et al, 2009):

$$l \leftarrow \max_{x_j} \mathrm{P}(z(x_j)=i)\left\|x_j - \bar{x}\right\|, \tag{22}$$

$$\sigma_f^2 \leftarrow \frac{\sum_{j=1}^{n}\mathrm{P}(z(x_j)=i)(r_j - \mathrm{E}[r])^2}{\sum_{j=1}^{n}\mathrm{P}(z(x_j)=i)}, \tag{23}$$

$$\sigma_n^2 \leftarrow \frac{1}{4}\sigma_f^2, \tag{24}$$

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

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. (2009) 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

$$\mathrm{P}(z(x_*) = i) = \frac{\exp(\bar{f}_i^{\,z}(x_*))}{\sum_{j=1}^{m} \exp(\bar{f}_j^{\,z}(x_*))}. \tag{25}$$

Here $\bar{f}_i^{\,z}(x)$ denotes the prediction of *z* for GP_i computed in the EM algorithm.

To illustrate this procedure, we consider the following one-dimensional toy example according to (Stachniss et al., 2009). 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 3 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 initialized and used together with the first component as the input to the EM algorithm. The individual images in Figure 4 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

Figure 3. Left: The standard GP used to initialize the first mixture component. Right: The error GP used to initialize the second mixture component.

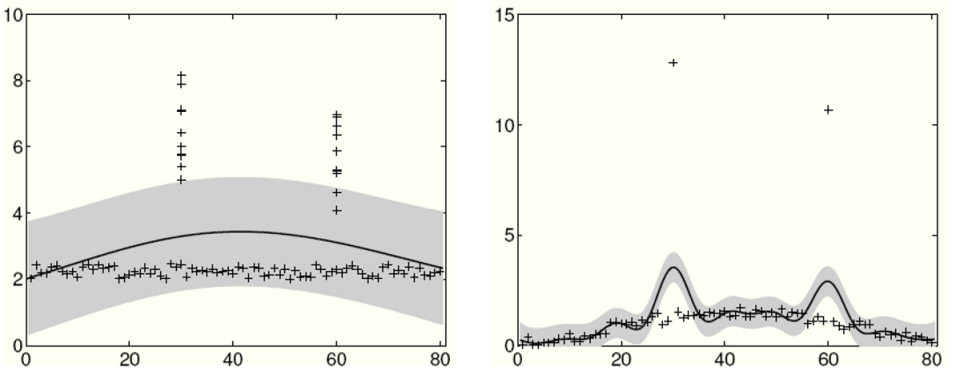
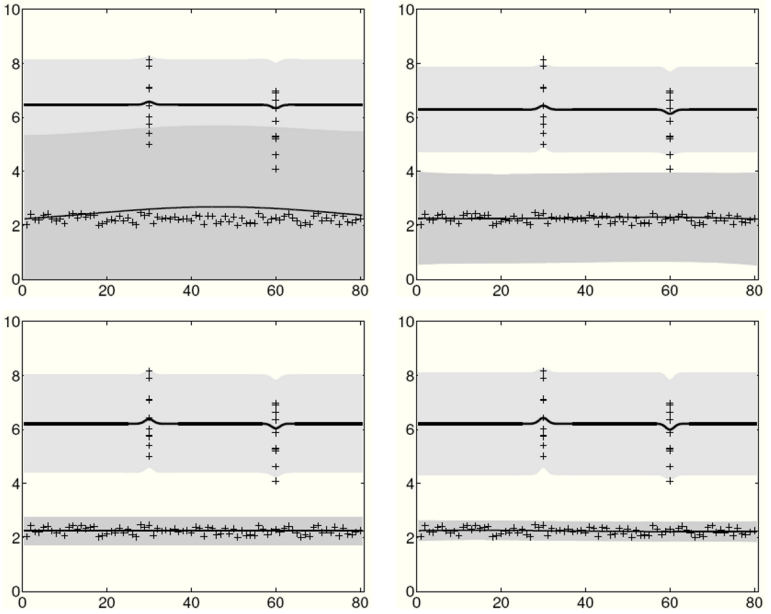
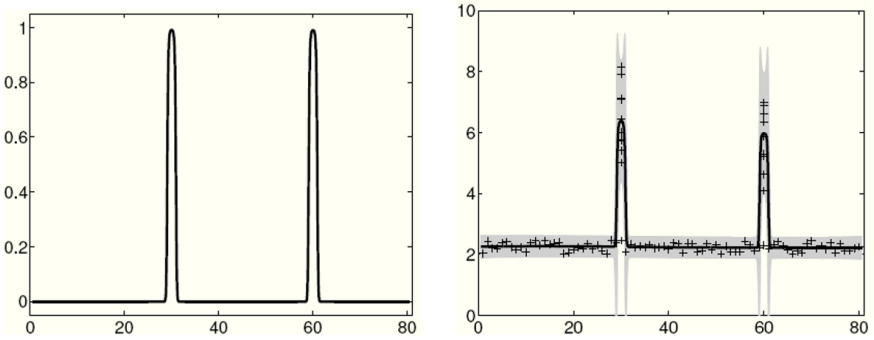


Figure 4. Components in different iterations of the learning using EM algorithm



is depicted in the left image of Figure 5 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 3.

Figure 5. Left: The learned gating function. Right: Resulting distribution of the GP mixture model.



EVALUATION AND COMPARISON OF GAS DISTRIBUTION MODELS

Evaluation and Learning of Meta Parameters

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 accurately predicts unseen observations.

To evaluate a gas distribution model, some ground truth information is required. An indirect evaluation 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 left out sensors. However, when the sensor network is too sparse or there is only a single mobile sensor, this solution is not feasible. In (Stachniss et al., 2009) and Lilienthal et al. (2009a) 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 $p(r_i|x_i)$, the NLPD of unseen measurements $\{r_1,r_2,\ldots,r_n\}$ acquired at locations $\{x_1,x_2,\ldots,x_n\}$ is computed as

$$NLPD = \frac{1}{2n}\sum_{\{(x_i,r_i)\}\in D}\left\{\log \hat{v}(x_i) + \frac{(r_i - \hat{r}(x_i))^2}{\hat{v}(x_i)}\right\} + \frac{1}{2}\log(2\pi), \tag{26}$$

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

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

Table 1. Comparison of GPM and Kernel DM+V

Data	NLPD (GPM)	NLPD (Kernel DM+V)
3-rooms	-1.54	-1.44
corridor	-1.60	-1.81
outdoor	-1.80	-1.75

Experimental Comparison of the Kernel DM+V and the GPM Approach

In the same way as a fixed distribution model is evaluated depending on its meta parameters, different gas distribution modeling approaches can be compared by their respective NLPD for unseen measurements. Since the goal is to maximize the likelihood of unseen data, better models minimize the NLPD. Table 1 shows an NLPD comparison of the GPM model (see Sec. " Gaussian Process Mixture Model") with the model obtained from the Kernel DM+V algorithm (Sec. " Kernel DM+V"). The comparison is based on data sets 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, GPM and Kernel DM+V exhibit a comparable performance for gas distribution modeling in the tested environments.

Discussion of Kernel DM+V and the GPM Approach

Kernel DM+V discretizes the space into grid cells and estimates a variance in addition to the mean of the gas distribution. Gas distribution modeling 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 weighs 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 modeling is treated as a regression problem. A kernel describes the covariance matrix, which models the influence of measurements at neighboring data points on the estimation. The approach proposed by Stachniss et al. (2009) 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 response. At each location, these two components are mixed in different proportions. The GPM approach proposed by

Stachniss et al. (2009) 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 mean and a predictive variance.

In GPM, both learning the model and making predictions require matrix inversion, which has 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 4σ around the measurement location. Therefore, the effective computational complexity is $O[n \cdot (\sigma/c)2]$. The complexity of computing the maps $\alpha^{(k)}$, $r^{(k)}$, and $v^{(k)}$ is $O[(D/c)2]$ and computing $V^{(k)}$ requires one pass through the data ($O[n]$). The overall complexity is $O[n \cdot (\sigma/c)2]$. 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 Sec. "Experimental Comparison of the Kernel DM+V and the GPM Approach"). 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 mixture of distribution models.

An advantage that comes with the straightforward interpretation of the Kernel DM+V algorithm is that it lends itself to a variety of extensions. Three such extensions, inclusion of wind information for gas distribution modeling (Kernel DM+V/W), 3D modeling (3D Kernel DM+V) and modeling time-dependency are described in the following section. 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.

EXTENSIONS OF KERNEL DM+V

Kernel DM+V/W: Using Wind Information for Gas Distribution Mapping

The statistical models presented in the previous sections have been used to model gas dispersion in indoor or controlled environments with relatively weak air flow. An important aspect for gas distribution modeling especially in outdoor applications is to consider wind information when building the gas distribution model. The local airflow is in general an important parameter due to the strong influence of advective

transport on gas dispersal. In (Reggente & Lilienthal, 2009b), the Kernel DM+V/W algorithm is proposed. Kernel DM+V/W is an extension of 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 bivariate 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. When reliable wind information is available, the symmetric univariate Gaussian in Equation (5) is replaced by an elliptic, bivariate Gaussian with the semi-major axis oriented along the wind direction. The 2×2 covariance matrix Σ is computed according to the current measurement of the local airflow $\vec{v}$ at the sensor location as follows.

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

$$\pi\sigma^2=\pi ab. \tag{27}$$

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

$$a = \sigma + \gamma \left|\vec{v}\right|. \tag{28}$$

By combining Equation (27) and Equation (28), the length of the semi-minor axis is

$$b = \frac{\sigma}{1 + \gamma \left|\vec{v}\right| / \sigma}. \tag{29}$$

Equations (28) and (29) describe the relation between the Eigenvalues of the covariance matrix and the wind intensity. The parameter γis related to the duration over which the wind vector $\vec{v}$ is assumed to be constant. For the initial experiments described in (Reggente & Lilienthal, 2009b), γ was set heuristically to 1. γ can also be learned from the input data.

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

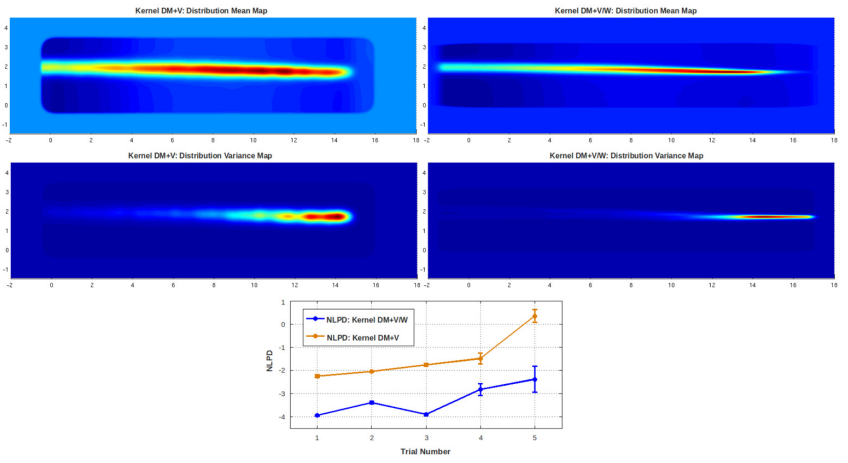
$$\Sigma_R^{-1} = R^{-1}\Sigma^{-1}R, \tag{30}$$

where R is the rotation matrix and Σ_R^{-1} is the inverse of the rotated covariance matrix.

By means of the adaptive kernel described by Equations (28), (29) and (30), Kernel DM+V/W models information about the trajectory of a sensed patch of gas, given a wind measurement. The bivariate Gaussian kernel describes a distribution over the locations where the sensed patch came from and where it tends to move to (here it is important to bear in mind that the resulting model describes the time-averaged gas distribution).

Reggente & Lilienthal (2009b) showed that applying the wind extension improves the performance of the Kernel DM+V approach. As an example of the application of Kernel DM+V/W, Figure 6 shows the predictive mean (top) and variance gas distribution (middle) for an experiment in a wind tunnel with laminar flow. In the maps on the left, the Kernel DM+V method is used, i.e. without considering wind information. The maps in the right of Figure 6 are obtained by using the Kernel DM+V/W algorithm. The NLPD comparison of using these two algorithms in Figure 6 (bottom) shows that Kernel DM+V/W can predict gas distribution dramatically more accurately than Kernel DM+V, especially in the presence of strong wind.

Figure 6. Mean gas distribution map (top) and its corresponding variance gas distribution map (middle) obtained from Kernel DM+V (left) and from Kernel DM+V/W using wind information (right), units in the plots are meters. The measurements are collected in a wind tunnel with laminar flow along the x-axis. The gas source is located at (15m, 2.0m). Comparison of NLPD from using Kernel DM+V (brown line) and Kernel DM+V/W (blue line) in this experiment is illustrated in the bottom of the figure.



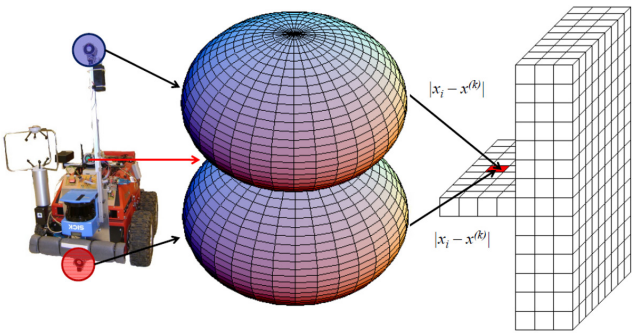
Three-Dimensional Statistical Gas Distribution Mapping

So far, all of the presented methods create two-dimensional models. Generally, however, gas dispersion has a three-dimensional structure and it can be helpful to model this 3D structure. In (Reggente & Lilienthal, 2009a), 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 the previous sub-section. In both cases, only the kernel is modified compared to the original algorithm. 3D Kernel DM+V uses kernel extrapolation with a trivariate Gaussian kernel to model the decreasing likelihood that a reading represents the true concentration with respect to the distance in three dimensions. The 3D Gaussian kernel (see Figure 7) is defined by a kernel width along the three axis σ_x, σ_y, and σ_z:

$$\omega_i^k(\sigma_x,\sigma_y,\sigma_z) = \mathrm{N}\left(\left|x_i - x^{(k)}\right|,\sigma_x,\sigma_y,\sigma_z\right). \tag{31}$$

As visualized in Figure 7, 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 readings, and weighted variance contributions are integrated and stored in temporary grid maps $\Omega^{(k)}$, $R^{(k)}$, and $V^{(k)}$, now using Equation (31) instead of Equation (5). Reggente and Lilienthal (2009a) used 3D Kernel DM+V to create gas distribution models in an uncontrolled indoor environment experiment. In this experiment, a mobile robot equipped with

Figure 7. Visualization of gas distribution modeling with 3D Kernel DM+V. Left side: The pollution monitoring prototype robot used to build gas distribution maps. Centre: three dimensional Gaussian kernel. Right side: the weight computed for a particular cell depends on the parameters of the 3D kernel and the distance between the centre of the cell and the measurement point.



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 2D 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. Compared to all other slices from the 3D map investigated, the slice at the height of the middle sensor was also found to have the minimum Kullback-Leibler distance. 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.

Time-Dependent Gas Distribution Modeling, TM Kernel DM+V

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 modeled by introducing a decreasing importance of measurements with increasing time between measurement and prediction.

Having a set of measurements $D= \{r_1, r_2,...,r_N\}$ collected at locations $\{x_1, x_2,...,x_N\}$ and times $\{t_1, t_2,...,t_N\}$, we want to create a model to estimate a snapshot of the gas distribution at time t^* with $t^*>t_1$. The spatial kernel is the same as in the basic Kernel DM+V algorithm, except from a time-dependency term φ, which is defined as follows:

$$\varphi(t^*, t_i) = e^{(-\beta\cdot(t^*-t_i))}, \tag{32}$$

where t_i is the time stamp of measurement r_i with $t_1 \le t_i \le t^*$ (i.e. we do not include future measurements into the prediction for t^*). β is a scaling factor for the recency of the measurements compared to time t^*. The total weight function that replaces Equation (5) then has the form

$$\omega_i^{(k)}(t^*, t_i) = \mathrm{N}\left(\left|x_i - x^{(k)}\right|, \sigma\right) \cdot \varphi(t^*, t_i). \tag{33}$$

By replacing Equation (5) with Equation (33), the predictive mean and variance can be computed as in Equation (9) and Equation (11) if the computation of the measurement average r_0 and the total variance of sensor measurements v_0 are also modified with the temporal weight function in Equation (32) so that more recent contributions have a higher influence:

$$r_0(t^*) = \frac{\sum_{t_i=t_1}^{t^*} \varphi(t^*, t_i) \cdot r_i}{\sum_{t_i=t_1}^{t^*} \varphi(t^*, t_i)}, \tag{34}$$

$$v_0(t^*) = \frac{\sum_{t_i=t_1}^{t^*} \varphi(t^*, t_i) \cdot \left(r_i - r^{(k(i))}\right)^2}{\sum_{t_i=t_1}^{t^*} \varphi(t^*, t_i)}. \tag{35}$$

Introducing time dependency in the proposed way was found to improve the modeling accuracy in initial experiments. Tab. 2 presents an NLPD comparison of time-dependent Kernel DM+V (TD Kernel DM+V) with the basic, not time-dependent version. The comparison is based on data sets from two different experiments: one experiment in a corridor, in which a single mobile sensor (carried by a robot) randomly moves and collects data; and a second experiment performed in a small room (4.9 × 3.4 m^2) where a network of 10 stationary sensors collected the measurements. The model was created in both cases using the 80% of the data that were recorded first, and evaluated on 20% of the data that were recorded afterwards. While the meta-parameter of the temporal weight function was set heuristically to a fixed value ($\beta = 0.0072s^{-1}$ for the corridor experiment and $\beta = 0.0016s^{-1}$ for the network experiment), the width of the spatial kernel and the optimal cell size were learned by optimizing the 5-fold cross-validation NLPD over the training set.

Table 2. Comparison of models created with TD Kernel DM+V and the basic version of the Kernel DM+V algorithm. Cell and kernel size are learned in each case based on the respective data sets.

Data	NLPD (Kernel DM+V)	NLPD (TD Kernel DM+V)	Kernel size (m)	cell size (m)
Network	-1.778	-2.180	0.410	0.101
	-1.694	-1.772	0.500	0.105
Corridor	-0.793	-1.068	0.420	0.100
	-0.828	-1.092	0.600	0.105

The preliminary results presented in Tab. 2 show that TD Kernel DM+V performs better than Kernel DM+V. The TD Kernel DM+V snapshots created for the time of the last measurement in the training set indeed do improve the model accuracy by assuming an increased importance of more recent readings.

CONCLUSION

This chapter introduced several kernel methods for statistical gas distribution modeling. 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 location, for example. The discussed methods treat gas measurements as random variables and predict the gas distribution at unseen locations.

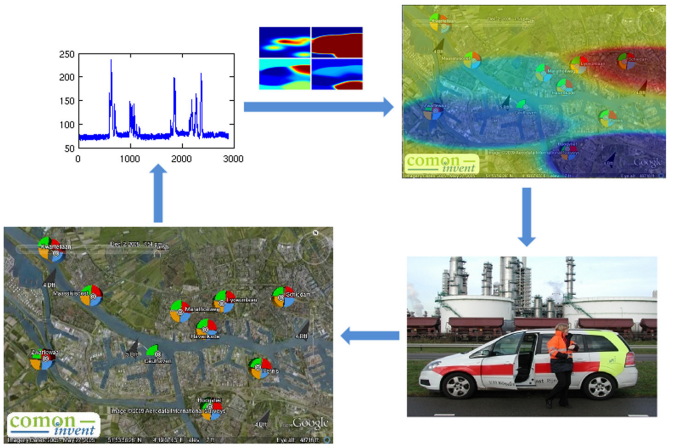
An important recent development was the introduction of modeling approaches to estimate the distribution variance in addition to the distribution mean. The two basic approaches, Kernel DM+V and GPM, were compared in this chapter. The modeling accuracy of the two approaches was found to be similar. However, 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.

In addition to the basic approaches, three modifications of the Kernel DM+V algorithm were described in this chapter: Kernel DM+V/W, which adapts the kernel shape according to wind measurements; 3D Kernel DM+V, which extends the distribution model to three dimensions using a 3D kernel and TD Kernel DM+V, which introduces a temporal weight function to acknowledge that more recent measurements carry more important information to create a model for predicting future measurements. The presented evaluations using real sensor data demonstrate the potential of all three modifications. 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 comparisons of the two-dimensional models, Kernel DM+V/W and TD Kernel DM+V, with the basic version of the Kernel DM+V algorithm found an improved model quality for both Kernel DM+V/W and TD Kernel DM+V.

FUTURE RESEARCH DIRECTIONS

The results presented in this chapter are based on real sensor data collected in uncontrolled indoor and small-scale outdoor environments. A next step is to test

Figure 8. Kernel DM+V in a large-scale outdoor environment. Snapshot obtained using Kernel DM+V for gas distribution modeling on a sensor network deployed over the Rijmond area.



the discussed kernel methods in large scale, real world applications. This line of research is pursued in the EC project Diadem (FP7-ICT 224318, 2008). Diadem addresses Distributed Information Acquisition and Decision-making for Environmental Management, particularly in cases that involve chemical incidents that cause the emission of hazardous chemical gases. The test scenario is an area of approximately 50 km^2 at the Rotterdam port in the Netherlands, a densely inhabited industrial area. The statistical gas distribution models can provide a comprehensive view on a large number of gas measurements to the decision makers and allow identifying areas of high gas accumulation for further inspection. In Figure 8, an experiment with a network of 10 sensors over the Rijmond area of Rotterdam port is illustrated.

Since the wind measurements available in Diadem are not obtained at the same position as the gas sensor measurements, future work includes to study methods that model the wind field based on sparse measurements and estimate the local wind at each sensor position, so as to be able to apply Kernel DM+V to compute the gas distribution.

Due to the large size of the environment and the sparseness of stationary sensors, the sensor network is combined with mobile sensors carried by field operators. This raises the question where to make measurements with these mobile sensors (the problem of sensor planning). The goal is to dispatch future measurements to areas about which relatively little information is available, or which have been identified as places with potentially high gas accumulation. The statistical gas distribution model can be used for sensor planning in the considered domain. The involved cost

function to evaluate will have to consider preferably areas of high gas accumulation or high predictive variance and regions in which measurements with maximum information gain are expected.

The presented gas distribution models all make Gaussian assumptions. In GPM these assumptions are not as strong as in Kernel DM+V since the GPM approach considers a mixture of distributions. Nevertheless, it is an important direction for future work to relaxing 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. Alternative, efficient methods to find optimal meta-parameters will also be investigated.

In this chapter, preliminary results with a time-dependent extension of Kernel DM+V are shown. There are many possibilities to improve the presented TD Kernel DM+V algorithm. One possibility is to learn the meta-parameter of the temporal weight function rather than setting it heuristically as it is done at the moment. It also needs to be studied whether the time-dependent models enable improved sensor planning compared to time-independent distribution models.

ACKNOWLEDGMENT

This work has partly been supported by the EC under contract number FP7-224318-DIADEM: Distributed Information Acquisition and Decision-Making for Environmental Management.

REFERENCES

Diadem FP7-ICT. (2008). *Diadem–distributed information acquisition decision-making for environmental management*. Retrieved from http://www.pdc.dk/diadem/

Hayes, A. T., Martinoli, A., & Goodman, R. M. (2002). Distributed odor source localization. *IEEE Sensor Journal*. *Special Issue on Electronic Nose Technologies*, *2*(3), 260–271. doi:.doi:10.1109/JSEN.2002.800682

Ishida, H., Nakamoto, T., & Moriizumi, T. (1998). Remote sensing of gas/odor source location and concentration distribution using mobile system. *Sensors and Actuators. B, Chemical*, *49*(1-2), 52–57..doi:10.1016/S0925-4005(98)00036-7

Lilienthal, A., & Duckett, T. (2004). Building gas concentration Grid maps with a mobile robot. *Robotics and Autonomous Systems*, *48*(1), 3–16..doi:10.1016/j.robot.2004.05.002

Lilienthal, A. J., Asadi, S., & Reggente, M. (2009a). Estimating predictive variance for statistical gas distribution modelling. In *AIP Conference Proceedings Volume 1137: Olfaction and Electronic Nose - Proceedings of the 13th International Symposium on Olfaction and Electronic Nose* (pp. 65-68). doi:10.1063/1.3156628

Lilienthal, A. J., Loutfi, A., Blanco, J. L., Galindo, C., & Gonzalez, J. (2007). A Rao-Blackwellisation approach to GDM-SLAM–integrating SLAM and gas distribution mapping. In *Proceedings of the 3rd European Conference on Mobile Robots* (pp. 126-131).

Lilienthal, A. J., Loutfi, A., & Duckett, T. (2006). Airborne chemical sensing with mobile robots. *Sensors (Basel, Switzerland)*, *6*(11), 1616–1678..doi:10.3390/s6111616

Lilienthal, A. J., Reggente, M., Trincavelli, M., Blanco, J. L., & Gonzalez, J. (2009b). A statistical approach to gas distribution modelling with mobile robots–the Kernel DM+V algorithm. In *Proceeding of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 570-576). doi: 10.1109/IROS.2009.5354304

Loutfi, A., Coradeschi, S., Lilienthal, A. J., & Gonzalez, J. (2008). Gas distribution mapping of multiple odour sources using a mobile robot. *Robotica*, *27*(2), 311–319.. doi:10.1017/S0263574708004694

Olesen, H. R., Løfstrøm, P., Berkowicz, R., & Ketzel, M. (2005). *Regulatory odour model development: Survey of modelling tools and datasets with focus on building effects*. (NERI Technical Report No. 541).

Purnamadjaja, A., & Russell, R. (2005). Congregation behaviour in a robot swarm using pheromone communication. In *Proceedings of the Australian Conference on Robotics and Automation*.

Pyk, P. (2006). An artificial moth: Chemical source localization using robot based neuronal model of moth optomotor anemotactic search. *Autonomous Robots*, *20*(3), 197–213..doi:10.1007/s10514-006-7101-4

Reggente, M., & Lilienthal, A. J. (2009a). Three-dimensional statistical gas distribution mapping in an uncontrolled indoor environment. In *AIP Conference Proceedings Volume 1137: Olfaction and Electronic Nose - Proceedings* of *the 13th International Symposium on Olfaction and Electronic Nose* (pp. 109-112). doi: 10.1063/1.3156484

Reggente, M., & Lilienthal, A. J. (2009b). Using local wind information for gas distribution mapping in outdoor environments with a mobile robot. In *Proceedings of IEEE Sensors 2009 Conference* (pp. 1712-1720). doi: 10.1109/ICSENS.2009.5398498

Roberts, P. J. W., & Webster, D. R. (2002). Turbulent diffusion. In Shen, H., Cheng, A., Wang, K.-H., Teng, M. H., & Liu, C. (Eds.), *Environmental fluid mechanics-theories and application* (pp. 7–47). Reston, VA: ASCE Press.

Schölkopf & J. Platt (Ed.), *Advances in neural information processing systems* (pp. 1257–1264). Cambridge, MA: MIT Press.

Snelson, E., & Ghahramani, Z. (2006). Sparse Gaussian processes using pseudo-inputs. In Weiss, Y. (Ed.), *B*.

Stachniss, C., Plagemann, C., & Lilienthal, A. J. (2009). Learning gas distribution models using sparse Gaussian process mixtures. *Autonomous Robots*, *26*(2-3), 187–202..doi:10.1007/s10514-009-9111-5

ADDITIONAL READING

Mobile Robot Olfaction

Lochmatter, T. (2010). Bio-inspired probabilistic algorithms for distributed odor source localization using mobile robots. (Doctoral dissertation, EPFL, 2010), *Dissertation Abstracts*, 4628.

Gas Sensors

Nakamoto, T., & Ishida, H. (2008). Chemical Sensing in Spatial/Temporal Domains. *Chemical Review Journal*, *108*(2), 680–704. doi:10.1021/cr068117e

Environmental Monitoring

Gilbert, R. O. (1987). *Statistical Methods for Environmental Pollution Monitoring*. Wiley.

Machine Learning

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Plagemann, C. (2008). Gaussian Processes for Flexible Robot Learning. PhD Thesis, University of Freiburg, Germany.

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MA: The MIT Press.