Available online at www.sciencedirect.com
ScienceDirect
Procedia Chemistry 00 (2009) 000–000
Procedia Chemistry
www.elsevier.com/locate/procedia

Proceedings of the Eurosensors XXIII conference

Statistical Evaluation of the Kernel DM+V/W Algorithm for Building Gas Distribution Maps in Uncontrolled Environments

Matteo Reggente* and Achim J. Lilienthal

AASS Research Center – Örebro University - Sweden

Abstract

In this paper we present a statistical evaluation of the Kernel DM+V/W algorithm to build two-dimensional gas distribution maps with a mobile robot. In addition to gas sensor measurements from an "e-nose" the Kernel DM+V/W algorithm also takes into account wind information received from an ultrasonic anemometer. We evaluate the method based on real measurements in three uncontrolled environments with very different properties. As a measure for the model quality we compute how well unseen measurements are predicted in terms of the data likelihood. A paired Wilcoxon signed rank test shows a significant improvement (at a confidence level of 95%) of the model quality when using wind information.

Keywords: gas distribution; e-nose; gas sensing; mobile robots; kernel density estimation; model evaluation.

1. Introduction

In urban environments and especially where population and traffic density are relatively high, human exposure to hazardous substances often exceeds air quality standards[1]. Pollution is typically measured with monitoring stations, typically several stationary units in a city. The total number of air quality monitoring stations in a city is limited by practical constraints accordingly the selection of sampling locations is a crucial issue. Monitoring stations are often placed at critical sites, for example near busy traffic axis, rather than in urban locations or parks away from road traffic, where humans are present[2]. This example illustrates that there is a need for refining the monitoring scale. On one hand, this need motivates the development of small inexpensive gas sensors for air pollution monitoring, in large stationary sensor networks[3]. A higher resolution can also be achieved, on the other hand, by using gas sensors carried by a robot. As a "mobile-nose" (m-nose) such robot can act as a mobile node in a sensors network. They offer several advantages for environmental monitoring including: higher and adaptive monitoring resolution, source tracking, integration into existing applications, first aid and cleanup of hazardous or radioactive waste sites, compensation for inactive sensors, and adaption to the dynamic changes of the environment. The data collected with a m-nose can be used to compute gas distribution map of the environment[4]. However, a difficulty lies in the fluctuating nature of the gas distribution. In natural environments, advective flow dominates gas dispersal compared to slow molecular diffusion. Since the airflow we encounter is almost always turbulent, the gas distribution becomes patchy and meandering. Few publications on gas distribution monitoring consider the influence of turbulent wind.

* Corresponding author. Tel.: +46-193-013-55; fax: +46-193-034-63.
E-mail address: matteo.reggente@oru.se

Addison[5] et al. propose an integrated method for predicting the spatial pollutant distribution within a street canyon. This method was based on a Lagrangian stochastic particle model superimposed on a velocity and turbulence field, which was assumed to be known. Statistical gas distribution modeling avoids making strong assumption about the environmental conditions. Moreover, even under conditions that can be modeled by a stationary random process, gas concentration has to be measured for a prolonged time (in the order of minutes) at each location in order to obtain a reliable measure of the concentration mean[6]. Compared to the corresponding demand on sampling density (in time and space), sampling will be always sparse in realistic applications. In order to obtain nevertheless a truthful statistical representation of the gas distribution, several approaches have been proposed in the area of mobile olfaction. Again, analytical models have been used that describe the time-averaged effect of turbulence as diffusion (eddy diffusivity) made the assumption of a stable uniform airflow field[6]. An alternative are interpolation methods[4,7] that provide a statistical representation of the gas distribution without assuming a pre-defined functional form of the distribution. In this paper, our approach is based on kernel extrapolation that treats sensors measurements as random variables[8]. We present an extension of the Kernel DM+V algorithm that uses information about the local airflow, in addition to gas sensors measurements, to compute the statistical gas distribution model (Kernel DM+V/W).
Most experiments in the domain of airbone chemical sensing with mobile robots were carried out in small controlled environments. In most of the cases, uniform strong airflow fields were artificially created. Otherwise, small areas in larger rooms were carefully chosen to have constant airflow. The major contribution of this paper is a statistical evaluation of the Kernel DM+V/W algorithm[9] based on measurements in three uncontrolled environments with very different properties: an enclosed indoor area, a part of a long corridor with open ends and a high ceiling, and an outdoor scenario, obtained with a mobile robot. As a measure for the model quality we compute how well unseen measurements are predicted in terms of the data likelihood and then a statistical test is performed to compare the Kernel DM+V/W algorithm with the Kernel DM+V algorithm that does not consider wind information.

2. Kernel DM+V/W Algorithm

In this section we briefly describe the basic ideas of the Kernel DM+V with Wind algorithm, a detailed description can be found in[9,8]. The gas distribution mapping problem addressed here is to learn a predictive two dimensional model $p(r|\mathbf{x},\mathbf{x}_{1:n},\mathbf{r}_{1:n})$ for the gas reading r at location $\mathbf{x}$, given the robot trajectory $\mathbf{x}_{1:n}$ and the corresponding measurement $\mathbf{r}_{1:n}$. The method proposed models the distribution mean $r^{(k)}$ and the corresponding variance $v^{(k)}$. To study how the gas is dispersed in the uncontrolled environment we consider the concentration readings from the "e-nose" and the anemometer readings. The central idea of kernel extrapolation methods is to understand gas distribution mapping as a density estimation problem addressed using convolution with a Gaussian kernel. The kernel can be interpreted as modelling the information content about the statistics of the gas distribution with respect to the point of measurement. By correlating the shape of the bivariate Gaussian kernel with the wind measurements we model the information content depending on the direction and magnitude of the wind. Intuitively a wind measurement provides information about where the dispersed patch of gas is likely to have come from and where it will tend to move to. In the case of zero wind (and also when no wind information is available) the contour of the Gaussian kernel is a circle. In the case of non-zero wind, we stretch the circular shape to an ellipse with the semi-major axis oriented along the wind direction.

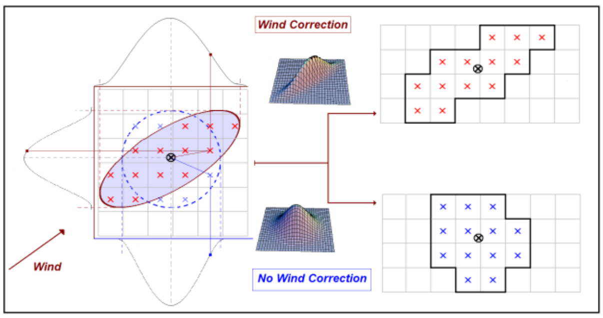

Fig. 1. Discretisation of the Gaussian kernel onto a grid. Left side: Model of the information content of a gas sensor reading (the sampling location is depicted in the center by a black "X") in the case of a radially symmetric Gaussian kernel and bivariate Gaussian kernel respectively. The blue-dashed circle represents the contour of the kernel in absence of wind and the red solid line ellipsoid for the case with wind correction. Right side: strongly affected cells are surrounded by a solid.

The bivariate normal distribution contour is an ellipses governed by a mean vector μ and covariance matrix ∑. An example that shows how a single reading is convolved onto a gridmap is given in Figure 1. Cells that are strongly affected by the measurement are indicated on the right side of the figure by a surrounding strong border. It is evident that the cells are affected differently in the two cases.

3. Experimental Setup

An ATRV-JR robot equipped with a SICK LMS 200 laser range scanner (for localization) and "electronic nose" was used for the monitoring experiments. The "electronic nose" comprise different Figaro 26xx gas sensors enclosed in an aluminum tube. The tubes are horizontally mounted at the front side of the robot and actively ventilated through a fan that creates a constant airflow towards the gas sensors. The 3D ultrasonic anemometer used to measure the airflow is a Young 81000 with a range from 2 cm/s up to 40 m/s and a resolution of 1 cm/s. The placement of the anemometer had to be a compromise between the desires to measure the airflow as close to the "e-nose" and as undisturbed as possible. It was finally placed above the top of the robot thus minimizing the influence of the fan of the "electronic noses" and the body of the robot itself. Alternative solutions would be to use smaller 2D anemometers mounted in the vicinity of the sensors[10].

Three scenarios (Fig. 2) were selected for the gas distribution mapping experiments. First, an enclosed indoor area consisting of three rooms separated by protruding walls. The whole was monitored and the path of the robot is approximately 14×6 m^2. The second location chosen was a section of a long corridor with open ends and a high ceiling. The area covered by the trajectory of the robot is approximately 14×2. m^2. There were more disturbances in this scenario caused by people passing by and the opening of doors and windows during the experiments. Finally, an outdoor scenario was considered. Here, the experiments were carried out in an 8×8 m^2 region that is part of a much bigger open area. In all the experiments, the robot followed a predefined sweeping trajectory or performed a random walk, in the area of interest. Along its path, the robot stopped at predefined positions (waypoints) and carried out a sequence of measurements on the spot for 10 s. The reason for stopping the robot at each waypoint to collect measurements is due to the difficulty in compensating for the shaking of the anemometer when the robot is moving. This is particularly critical indoors where very weak airflow has to be measured.

Fig. 2. The prototype pollution monitoring robot during the mission in the experimental environments: an enclosed indoor area (left), a part of a long corridor with open ends and a high ceiling (center), and an outdoor scenario (right).

4. Model Evaluation and Results

The knowledge of the exact gas source position and the correlation with the maximum in the map has been considered as a way to evaluate the gas distribution model. However, the maximum of the gas distribution does not necessarily have to correspond to the true location of the gas source. An alternative way to evaluate the model quality is to evaluate how well unseen measurements are predicted by the distribution model. To do this we split the dataset D into disjoint sets D_{learn} and D_{eval}. We use D_{learn} to learn the hyper-parameters and the corresponding model and then compute the likelihood of unseen measurements as the average negative log predictive density (NLPD) over the n' samples in D_{eval}. Since we want to maximize the likelihood of the data points our goal is minimize the NLPD. In order to evaluate whether the Kernel DM+V/W algorithm improves the model, we have calculated the NLPD with and without wind correction in three different scenarios which correspond to different wind characteristics.

For the outdoor experiments and the trials carried out in the enclosed indoor area first, we used the first half of the dataset to learn the hyper parameters and the corresponding model and the second half to evaluate it, then the second half to train and the first to evaluate. In the corridor experiments we started with the first three quarters of each data set for training and the remaining quarter for evaluation. This procedure was repeated using the third, second, and finally the first quarter for evaluation (and the remaining quarters for training, respectively). The differences in splitting the data sets was motivated by the duration of the experiments (two hours for the outdoor area and the enclosed room; 4 hours for the corridor experiments), the robot's path (4 sweeps or random walk in the corridor compared to two sweeps in the other two scenarios). Due to the difficulty of measuring wind precisely when the robot is moving, we considered gas sensors and anemometer readings only when the robot is stopped and only gas sensor readings when the robot is driving.

The results are outlined in Figure 3a and 3b. The bar-plots show in black the number of times in which the Kernel DM+V/W algorithm performed better (in terms of a lower NLPD) Kernel DM+V without consideration of wind information. The orange part of the bar indicates the number of times in which Kernel DM+V/W performed worse. In the right side of figure 3a the absolute difference of the obtained NLPD ($\Delta(\text{NLPD}) = (\text{NLPD}_{\text{DM+V}}{-}\text{NLPD}_{\text{DM+V/W}})$) is plotted with a dashed red line. Positive values correspond to an improved model when the wind kernel is used. Figure 3a shows results obtained from data sets that contain periods where the robot was driving and periods where it was stopped (Drive & Stop trials). Figure 3b shows results from data sets that contain only those parts of the data where the robot was stopped (Stop trials). It is evident the performance is much better in the latter case. The results of a paired Wilcoxon signed-rank test confirm this observation at a confidence level of 95% (see Table 1). The statistical evidence in the case of the stop trials is significant even at a higher confidence level.

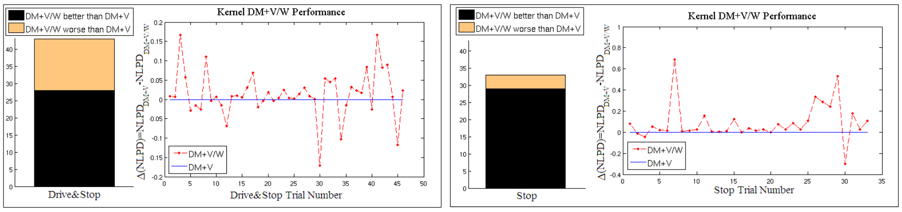

Table 1. Statistical evaluation

Trials Type	Wilcoxon Signed-Rank Test – *p* value
Drive & Stop	0.027
Stop	$5.6{\cdot}10^{-5}$

Fig. 3. a) Drive & Stop trials. b) Stop trials. Bar-plot: black = number of times in which Kernel DM+V/W performed better than Kernel DM+V (in terms of a lower NLPD); Orange worse. Graph: $\Delta(\text{NLPD}) = (\text{NLPD}_{\text{DM+V}} - \text{NLPD}_{\text{DM+V/W}})$. Positive values of the red line correspond an improvement of the model using Kernel DM+V/W.

5. Conclusion and Future Works

Gas distribution modelling with a mobile robot in an uncontrolled environment is a challenging field of research. This is mainly due to turbulent nature of gas dispersal in a real-world scenario. In this paper we present a statistical evaluation of the Kernel DM+V/W algorithm that uses wind information to build two-dimensional gas distribution maps. Experiments were performed with a mobile robot equipped with an "e-nose" and an ultrasonic anemometer in three uncontrolled environments. The proposed Kernel DM+V/W algorithm is compared to the Kernel DM+V algorithm, which does not use wind information for building the gas distribution model, in terms of the ability of the obtained model to predict unseen data. The results show a significant improvement when using wind information despite a sub-optimal sensor set-up. This work suggests modifying the hardware to compensate the anemometer shaking in order to have information of the wind also when the robot is moving. Another interesting extension is to integrate the proposed method with a 3D model obtained by data acquired from noses mounted at different heights.

References

1. Vardoulakis S. *Modelling air quality in street canyons: a review,* Atmospheric Environment, vol. 37, no. 2, pp. 155–182, 2003.
2. Fenger J. *Urban Air Quality*, Atmospheric Environment, vol. 33, pp. 4877–4900, 1999.
3. Maruo Y. Y, Ogawa S, Ichino T, Murao N, Uchiyama M. Atmospheric Environment, vol. 37, pp. 1065–1074, 2003.
4. Lilienthal A. and Tom D. *Building Gas Concentration Gridmaps with a Mobile Robot*, *Robotics and Auto. Syst.*, vol. 48, pp. 3–16, 2004.
5. Addison P.S, Currie J.I, Low D.J; McCann J.M. Environmental Monitoring and Assessment, no. 65, pp.333–342, 2000.
6. Ishida H, Nakamoto T.; Moriizumi T. *Remote Sensing of Gas/Odor Source Location and Concentration Distribution Using Mobile System*, Sensors and Actuators B, vol. 49, pp. 52-57 1998.
7. Stachniss C, Plagemann C, Lilienthal A. *Learning Gas Distribution Models using Sparse Gaussian Process Mixtures,* Autonomous Robots, 26:2-3, pp. 187-202, 2009.
8. Lilienthal A.J, Reggente M, Trincavelli M, Blanco J.L, Gonzalez J, *A Statistical Approach to Gas Distribution Modelling with Mobile Robots – The Kernel DM+V Algorithm*, Proc. IROS 2009, to appear.
9. Reggente M. And Lilienthal A. J. *Using Local Wind Information for Gas Distribution Mapping in Outdoor Environments with a Mobile Robot*. Proc. IEEE Sensors, 2009, to appear.
10. Lochmatter T; Raemy X; Matthey L; Indra S; Martinoli A. *A Comparison of Casting and Spiraling Algorithms for Odor Source Localization in Laminar Flow*. Proc. ICRA, pp. 138-1143, 2008.