Imperial College
London

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

Gaussian Processes for Optimal Sensor Position

Background & Progress Report

Author:
Adrian LÖWENSTEIN

Supervisor:
Rossella ARCUCCI
Miguel MOLINA-SOLANA

Submitted in partial fulfillment of the requirements for the MSc degree in Computing (Machine Learning) of Imperial College London

June 2019

Contents

Chapter 1

Introduction

1.1 Summary

Gaussian processes (GP) have been widely used since the 1970's in the fields of geostatistics and meteorology. Current applications are in diverse fields including sensor placement In this project, we propose the employment of a GP model to calculate the optimal spatial positioning of sensors to study and collect air pollution data in big cities. We will then validate the results by means of a data assimilation software with the data at the proposed positions.

1.2 Data

London South Bank University (LSBU) air pollution data (velocity, tracer)

Chapter 2

Background

In this chapter we will cover the literature and the theory that will be used throughout the project. First we will review the context of the project and how it fits into the **MAGIC** project. Then our focus goes to the definition of **Gaussian Processes** (GP) and how they are used in the context of geospatial data. Furthermore, the use of GP relies heavily on **Covariance** matrixes which needs to be estimated. Those tools enable us to create **optimisation** algorithms for the position of sensors. Finally we will quickly explore the concepts of **Data Assimilation** (DA) that will be used to validate the results of the optimisation.

2.1 The MAGIC Project

This work is done in the context of the **Managing Air for Green Inner Cities** project. This is a multidisciplinary project and has for objective to find solutions to the pollution and heating phenomenons in cities. Traditionally, urban environmental control relies on polluting and energy consuming heating, ventilation and cooling (HVAC) systems. The usage of the systems increases the heat and the pollution levels, inducing an increased need for the HVAC. The MAGIC project aims at breaking this vicious circle and has for objective to provide tools to make possible the design of cities acting as a natural HVAC system.

This has been extensively discussed by Song et al. (2018). For this purpose, integrated management and decision-support system is under development. It includes a variety of simulations for pollutants and temperature at different scales; a set of mathematical tools to allow fast computation in the context of real-time analysis; and cost-benefit models to asses the viability of the planning options and decisions.

As explained by Song et al. (2018), the test site which has been selected to conduct the study is a real urban area located in London South Bank University (LSBU) in Elephant and Castle, London. In order to investigate the effect of ventilation on the cities problem, researchers in the MAGIC project have created simulations and experiments both in outdoor and indoor conditions, on the test site. They used wind tunnel experiments and computational fluid dynamics (CFD) to simulate the outdoor environment. Further works include the development of reduced-order modelling (ROM) in order to make faster the simulations while keeping a high level of accuracy (Arcucci et al., 2018).

Another key research direction in the use Data Assimilation (DA) and more specifically Variational DA (VarDA) for assimilating measured data in real time and allowing better prediction of the model in the near future (Arcucci et al., 2018). The further use of those method would be the optimisation of the position of the sensors which provide information for the VarDA.

2.2 Data Assimilation

2.2.1 Definitions

2.2.2 DA in the MAGIC Project

2.3 Gaussian Process and Sensor Optimisation

2.3.1 Usage of Gaussian Processes

As explained by Rasmussen and Williams (2006), the history of Gaussian Processes Rasmussen and Williams (2006) goes back at least as far as the 1940s. A lot of usages were developped in various fields. For prediction in geostatistics with methods known as kringing, and metheorology for example. For spatial prediction or.

Gradually GP started to be be used in more general cases for regression. Works from ... present the usage of the GP for one dimensional regression problems.

2.3.2 Definitions

2.3.3 Usage in Sensor Optimisation

2.3.4 Definitions

2.4 Covariance

We have seen how GP could be used for the optimisation of the position of sensors. In order to have good results we need to have a good estimate of the kernel function between the points of

2.4.1 Stationnary Covariance

Isotropic kernels

2.4.2 Covariance Estimation

Sample variance = bad estimator in high dimensions

Chapter 3

Progress

3.1 Data Exploration

3.2 Covariance Estimation

3.3 Sensor Optimisation

Cressie (1991) Arcucci et al. (2018)

Chapter 4

Further Developements

Bibliography

Arcucci, R., Pain, C., and Guo, Y.-K. (2018). Effective variational data assimilation in air-pollution prediction. *Big Data Mining and Analytics*, 1(4):297–307. pages 3, 5

Cressie, N. A. C. (1991). *Statistics for spatial data*. Wiley series in probability and mathematical statistics. Wiley, New York. pages 5

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass. OCLC: ocm61285753. pages 3

Song, J., Fan, S., Lin, W., Mottet, L., Woodward, H., Davies Wykes, M., Arcucci, R., Xiao, D., Debay, J.-E., ApSimon, H., Aristodemou, E., Birch, D., Carpentieri, M., Fang, F., Herzog, M., Hunt, G. R., Jones, R. L., Pain, C., Pavlidis, D., Robins, A. G., Short, C. A., and Linden, P. F. (2018). Natural ventilation in cities: the implications of fluid mechanics. *Building Research & Information*, 46(8):809–828. pages 2