GreenIoT: Cloud-Based Data Fusion for Smart Cities

Xiuming Liu[1], Abdul Rahman Khankan[1], Mohamad Alsioufi[1], Zhitao He[1,2], Edith C.H. Ngai[1]
1. Department of Information Technology, Uppsala University
2. Networked Embedded Systems Group, SICS Swedish ICT

UPPSALA UNIVERSITET

Introduction

In GreenIoT project [1], we develop an internet-of-things (IoT) testbed consisting of a network of low cost sensors and a cloud-based air quality data analysis service for smart cities. The GreenIoT project builds an ecosystem for smart cities with three layers: 1) Embedded sensors which are low cost and portable; 2) Cloud-based data fusion services, including storage, analysis, and visualisation; 3) Applications powered by the data from IoT and cloud. Those three layers are connected via machine-to-machine (M2M) communications (e.g., MQTT + 6LoWPAN) and mobile cellular networks.

System Architecture

The cloud-based data analytic service is one of the central components in the GreenIoT testbed, not only because it bridges sensor networks and applications which might be developed and operated by different organisations, but also because of discovering knowledge from noisy sensory data.

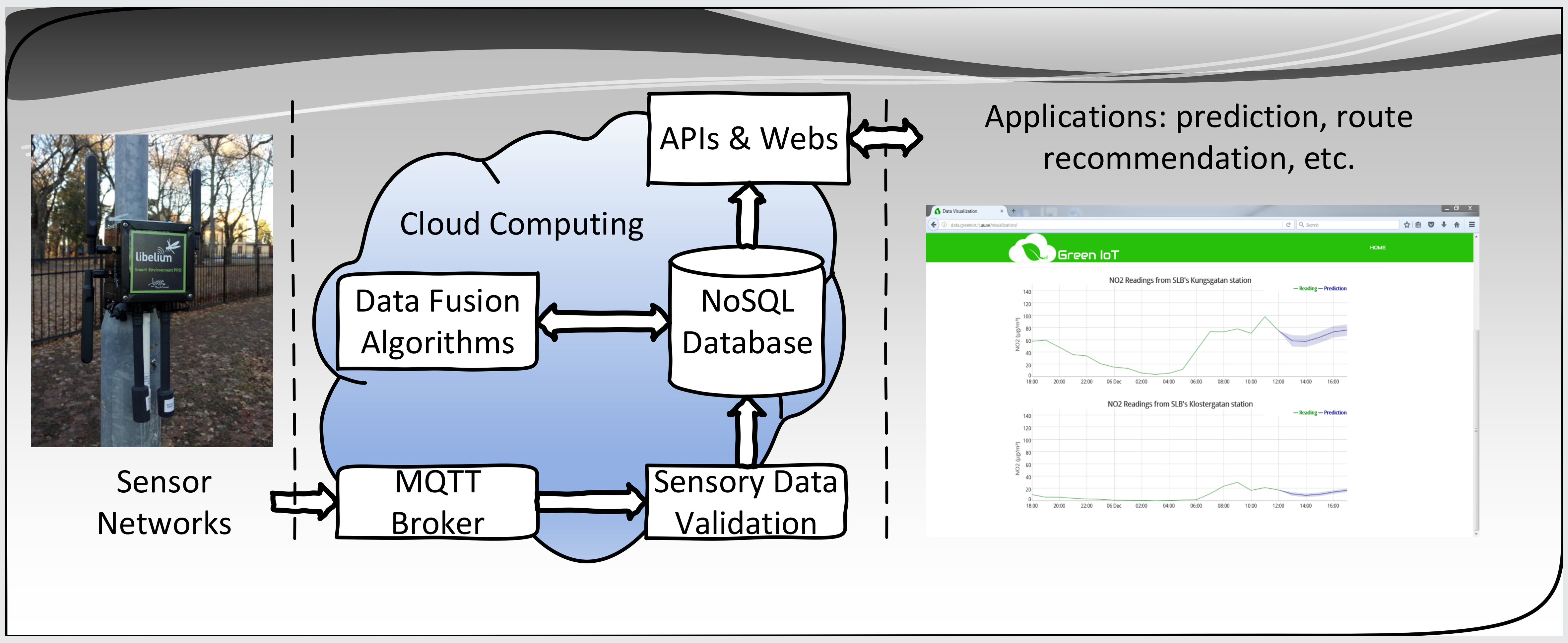


Figure 1: System architecture of the cloud-based services.

The cloud-based system consists of the following sub-components: communication interfaces, a sensory data validation mechanism, a document-oriented database, and a collection of data fusion algorithms.

Key Modules

Communication Interface

The M2M communication protocol MQTT is applied for the interface between sensors and the cloud. For the communication between cloud and applications, a set of web application programming interfaces (APIs) are implemented.

Data Validation

The data validation module prevents malformed sensory data from being stored and used.

Data Storage

A data storage system based on MongoDB is implemented for the system. The database stores all data (including both raw sensory data and data fusion results) in SenML JSON arrays.

Data Fusion Algorithms

The central task of the data fusion service is to process sensory data and provide useful information.

Application Examples

Forecasting the Time Series Data

The state-space model is used to represent the dynamics of air qualities:

$$\boldsymbol{x}_t = f(\boldsymbol{x}_{1:t-1}) + \boldsymbol{u}_t, \text{ and } \boldsymbol{y}_t = g(\boldsymbol{x}_t) + \boldsymbol{e}_t. \quad (1)$$

To model the dynamic function with GP [2], let

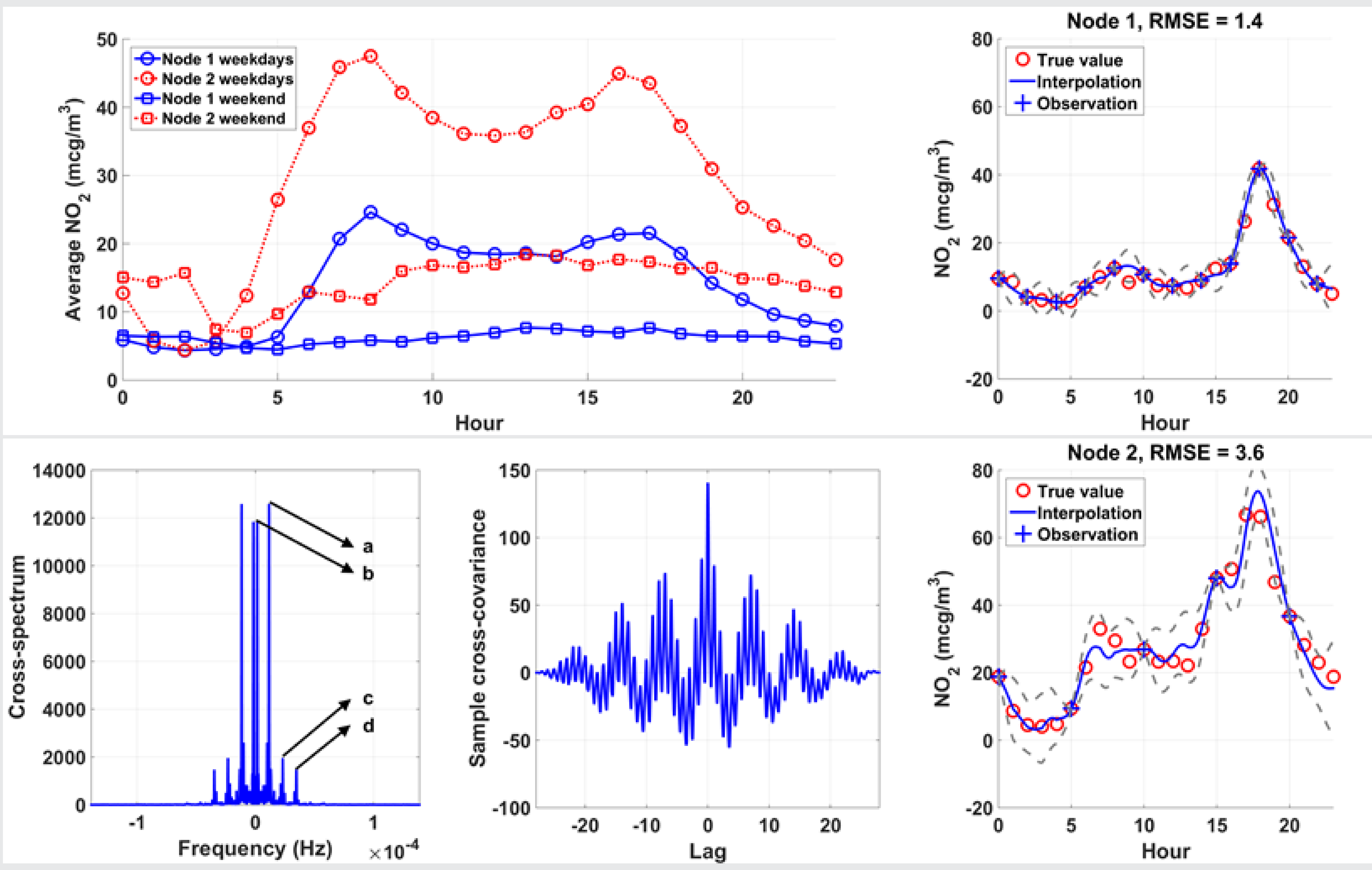
$$\boldsymbol{x}_t \sim \mathcal{GP}(m(t), k(t, t')) \quad (2)$$

where the mean function $m(t)$ and the kernel function $k(t, t')$ are learned from historical records [3].

Spatial-Temporal Interpolations

The inputs from a network of sensors are a set of spatial-temporal data. The number of sensors is usually much smaller than the size of the service area ($\text{size}(\boldsymbol{y}_t) \ll \text{size}(\boldsymbol{x}_t)$). Therefore, a spatial-temporal interpolation is required to provide a continuous air pollution map.

From Statistical Modelling to Regression For Sensory Data



Conclusions

An information system for smart cities should not only have the ability to gather data from IoT but also to abstract useful information from raw sensory data and visualize the knowledge in an intuitive way. A part of the GreenIoT project's goal is to explore research questions raised in handling spatial-temporal and heterogeneous sensory data in smart cities. For future works, the project aims to produce a real-time air pollution map and integrate data from other dimensions to further improve the quality of services.

References

[1] B. Ahlgren, M. Hidell, and E. Ngai, "Internet-of-things for smart cities: Interoperability and open data," *IEEE Internet Computing*, 2016.

[2] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning.* University Press Group Limited, 2006.

[3] X. Liu, T. Xi, and E. Ngai, "Data modelling with gaussian process in sensor networks for urban environmental monitoring," in *2016 IEEE 24th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS)*, Sept 2016, pp. 457–462.

Acknowledgements

This work is supported by the Vinnova GreenIoT project (2015-00347) in Sweden. Mohamad Alsioufi is supported by the Swedish Institute scholarship during his study at Uppsala University. These supports are gratefully acknowledged.