available at www.sciencedirect.com
SCIENCE DIRECT®
journal homepage: www.elsevier.com/locate/ecolmodel

Grid-based analysis of air pollution data

M. Richards, M. Ghanem*, M. Osmond, Y. Guo, J. Hassard

Imperial College London, 180 Queens Gate, London SW7 2BW, United Kingdom

ARTICLE INFO

Article history:
Available online 7 February 2006

Keywords:
Sensor network
Sensor grid
Air pollution monitoring
Data mining
e-Science

ABSTRACT

In this paper, we present a distributed infrastructure based on Grid computing technology and data integration and mining tools to discuss the main informatics challenges that arise when a high-throughput sensor network is constructed to address real-time urban air pollution monitoring and mapping. We present the background to urban air pollution monitoring and modelling and describe the high-throughput sensors developed within this project to tackle the problem. We present a solution that addresses the informatics challenges based on the integration of distributed sensors, Grid technologies, data integration, data mining and GIS systems. We also present a case study for examining the effectiveness of visual and automated methods developed for the analysis of generated data sets.

1. Introduction and overview

As many cities around the world become more congested, concerns increase over the level of urban air pollution being generated and in particular its impact on localised human health effects such as asthma or bronchitis. The more this relationship is understood, the better chance there is of controlling and ultimately minimising such effects. In the majority of the developed world, legislation has already been introduced to the extent that local authorities are required by law to conduct regular Local Air Quality Reviews of key urban pollutants such as benzene, SO_2, NO_x or ozone—produced by industrial activity and/or road transport (UK Department of Trade and Industry, 2000). In order to implement this however, pollutant concentrations must be monitored accurately and ideally in situ so that sources may be identified quickly and the atmospheric dynamics of the process understood. Furthermore, such data would lend itself to real-time environmental decision-making capabilities as a result of hazardous levels being rapidly identified.

The Discovery Net project (Curcin et al., 2002) is developing grid-based methods for the integration and analysis of data generated from distributed high-throughput devices in a variety of application areas including life science, environmental science and remote sensing. The goal is to develop an advanced generic computing infrastructure that supports real-time processing, interpretation, integration, visualisation and mining of massive amounts of time-critical data generated from such devices. One of the main application areas of the project is the analysis of data generated by the GUSTO high-throughput pollution monitoring sensors (see Section 4).

Deploying a sensor network over a target region, such as a heavily industrialised or densely populated area, creates a wealth of data allowing new types of analyses to be conducted. These include the analysis and visualisation of the spatiotemporal variation of multiple pollutants with respect to one another, and their correlation with third-party data, such as weather, health or traffic data. Such analysis can also provide valuable clues as to how local health effects (e.g. aggravated respiratory illnesses) occur. However, modern sensor technologies, e.g. GUSTO, which measure pollutants at a high level of accuracy and throughput can generate up to 8 GB data each day per sensor. This raises many informatics challenges with respect to managing and analysing the collected data.

* *Corresponding author*. Tel.: +44 20 7594 8357; fax: +44 20 7594 8294.
E-mail addresses: mark.richards@imperial.ac.uk (M. Richards), m.ghanem@imperial.ac.uk (M. Ghanem), michelle.osmond@imperial.ac.uk (M. Osmond), y.guo@imperial.ac.uk (Y. Guo), j.hassard@imperial.ac.uk (J. Hassard).
0304-3800/$ – see front matter
doi:10.1016/j.ecolmodel.2005.10.042

In the remainder of this paper, we describe the motivations for the development of the high-throughput GUSTO sensors within the project as well as the sensors themselves. We also discuss the main informatics challenges that arise when a high-throughput grid is constructed based on such sensors, present a snapshot of the infrastructure developed to address these problems and discuss the various data visualisation and analysis scenarios for which the platform has been designed.

2. Pollutant sources and health effects

Human beings breathe in and out approximately once every 4 s, which equates to over 8 million times a year. Urban air pollution is therefore one of the most important environmental issues that may be considered due to its direct effect on human health. It is known that exposure to high concentrations of air pollution over short periods of time (usually seconds) is far more harmful to human health than long term exposure to lower concentrations (The World Bank, 1994). Thus, when monitoring pollution in order to assess the true impact on human health, a temporal resolution of a few seconds is vital. However, legislative directives are based on running mean average concentrations of at least 15 min. This implies that large spikes (elevated levels) of short duration may be shrouded within the mean value, particularly if that value is close to the legal limit.

Air pollution mainly results from anthropogenic (human) activities and has diverse causes and sources. “Stationary sources”, such as factories, power plants, and smelters; “area sources”, which are smaller sources such as dry cleaners and degreasing operations; “mobile sources”, such as cars, buses, planes, and “natural sources”, such as windblown dust and wildfires, all contribute to air pollution.

The primary airborne pollutants covered by European and other legislation are: SO_2, NO_x (NO/NO_2), benzene, ozone, CO/CO_2, and Particulate Matter ($PM_{10}/PM_{2.5}$) (UK Environment Agency, 2004). Currently GUSTO sensors are optimised to monitor SO_2, NO, NO_2, benzene and ozone—primarily due to the fact that all of these compounds have measurable absorption signatures in a fairly narrow part of the UV spectrum, (see Section 4.1). However, further optimisation of the sensor is anticipated (with the addition of infrared capability) that will lead to the inclusion of CO/CO_2 and particulate matter, thus covering the whole suite. A summary of the key pollutants (covered by GUSTO), likely sources and resulting health effects are summarised in Table 1, and discussed in more detail below.

2.1. *Sulphur dioxide (SO_2)*

SO_2 is prevalent in most industrial raw materials, including crude oil, coal, and common ores like aluminium, copper, zinc, lead, and iron. Sulphur gases are produced when fuel, such as oil and especially coal, is burnt, during mining and industrial processes, e.g. when petrol is extracted from crude oil and naturally from volcanic eruptions.

Health effects of SO_2 gas are irritation to the eyes and respiratory system, reduced pulmonary functions and aggravation to respiratory diseases such as asthma, chronic bronchitis and emphysema. Exposure to extremely high concentrations will cause permanent damage to the respiratory system as well as extreme irritation to the eyes (due to production of dilute sulphuric acid around the eyes). When SO_2 reacts with other chemicals in the air to form tiny sulphate particles, these may also be inhaled in which case they gather in the lungs and are associated with increased respiratory symptoms and disease, difficulty in breathing, and premature death (US Environmental Protection Agency, 2004).

2.2. *Benzene (C_6H_6)*

Benzene is the most common of a group of compounds referred to as Volatile Organic Compounds (VOCs). Benzene is a minor constituent of petrol. Generally, VOCs are produced as fuel by-products in a combustion process.

Benzene is a known carcinogen, however the main health hazard arises from its role in the production of ground level ozone.

2.3. *Ozone (O_3)*

Ozone (O_3) is a colourless gas formed at ground level by reactions involving VOCs and nitrogen oxides. Ground level ozone

Table 1 – Summary of pollutants and health effects

Pollutant	Formula	Source	Health effects	Legal limit (ppbv)	Averaging time
Sulphur dioxide	SO_2	Petroleum refineries/coal powered power stations	Irritation to eyes and respiratory system. Reduced pulmonary functions	100	15 min
Benzene	C_6H_6	Transport/industry unburned fuel products	Known carcinogen. Also plays role in formation of ground level ozone	5	Running annual mean
Nitric oxide	NO	High temperature combustion processes/road transport	Can increase incidences of acute respiratory disease in children	16	Annual mean
Nitrogen dioxide	NO_2	High temperature combustion processes/road transport	Irritation to lungs and lowered resistance to respiratory infections such as influenza	105	1 h mean
Ozone	O_3	Ground level reactions involving NO_x and VOCs	Respiratory infection, lung inflammation, aggravation of asthma	50	Running 8 h mean