Mobile Environmental Sensing System Across Grid Environments

Imperial College London

Newcastle University

University of Cambridge

University of Leeds

University of Southampton

Final Report to Department for Transport

December 2009

Executive Summary

Understanding the generation, distribution and impacts of urban and regional air pollution is a major scientific challenge. Academic and practical work is limited by a lack of data of adequate temporal and spatial granularity. Current trends in sensor, communications and computing technologies create opportunities for pervasive, high-resolution data capture from portable devices. The MESSAGE consortium brought together research groups in eScience, transport, sensors and communications technologies from Imperial, Cambridge, Leeds, Newcastle and Southampton. The project addressed the challenge of how to capitalise on simultaneous improvements in the quality and availability of data.

Three types of sensors measuring multiple types of pollution were deployed. Sensors have been built based on ultraviolet absorption spectroscopy technology, which uses ultraviolet light to detect up to five traffic pollutants simultaneously, including nitrogen oxides and sulphur dioxides This can take measurements at less than 5 second intervals, fast enough to allow deployment on moving vehicles. The measurements and the location of each mobile sensor can be displayed in real time on a web based graphical interface. The second type of sensor uses an electro-chemical technique. The innovation is to deploy multiple devices tuned to different pollutants on a compact, battery powered and wireless connected platform. This can be carried by individuals to measure the pollution that they are exposed to. The sensors are small enough to fit into a pocket and utilize the wearer's mobile phone to transmit the pollution levels to a remote database. Electro chemical sensors linked to a very low power consumption platform created a device that could easily be deployed on street furniture. Low power wireless networking enables dense networks of such sensors to be deployed to provide 24/7 pollution measurements. MESSAGE used of the state-of-the-art Instrumented City facility at Leeds for the validation of sensors. More than 100 sensors were deployed in London, Leicester, Gateshead and Cambridge to analyse the link between traffic congestion and levels of pollution.

Such systems produce large quantities of data to be communicated, processed and stored. A significant result was the design and implementation of a data architecture that supported large numbers of the different types of sensors and allowed the system to be dynamically scaled up as more sensors are added. Data can be stored in multiple databases incorporating the Urban Traffic Management and Control specification. The data can be coupled with data from other sources, such as weather and traffic flow and fed into applications such as emission and pollution dispersion models. User applications can interrogate the data contained in multiple databases using data "federation" middleware, known as OGSA-DAI and OGSA-DQP. This eScience middleware insulates the users and applications from the underlying complexity.

Researchers can now measure and model air quality in unprecedented detail to improve their understanding about pollution hotspots and analyse the factors such as bad urban design. The MESSAGE data infrastructure provide the means to collect, process and manage the large scale data sets needed to understand how air pollution forms, lingers and dissipates in high emission zones. Potential users were consulted and a set of "use cases" developed.
The overall architecture and data management processes can be used in other domains where there is a need for large scale, remote data collection, linkage with other data sets, multiple data processing tasks and long term data curation.

Moving from the laboratory to deployment is a major challenge and to demonstrate a complete system operating as an integrated whole is a considerable achievement. MESSAGE has created a pool of human capital and experience of how to achieve real progress in a complex organisational situation.

Information regarding the project can be found at www.message-project.org

Contents

Project participants

Principal Investigator

Professor John Polak, Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London

Co-Investigators

Imperial College London
Professor J Darlington, Department of Computing
Professor Y Guo, Department of Computing
Dr J Hassard, Department of Physics
Professor K Leung, Department of Electrical and Electronic Engineering
Professor W Ochieng, Centre for Transport Studies, Department of Civil and Environmental Engineering

Newcastle University
Professor P T Blythe, Department of Civil Engineering and Geosciences
Professor M Bell, Department of Civil Engineering and Geosciences
Professor P Watson, Director, North East Regional e-Science Centre, School of Computing Science
Professor B Sharif, School of Electrical, Electronic and Computer Engineering

University of Cambridge
Professor P V Landshoff, Department of Applied Mathematics and Theoretical Physics,
Professor R L Jones, Depratment of Chemistry, University of Cambridge
Professor J Bacon, Computer Laboratory, University of Cambridge

University of Southampton
Professor M McDonald, School of Civil Engineering and the Environment,
Dr B J Waterson, School of Civil Engineering and the Environment

University of Leeds
Dr H Chen, Institute for Transport Studies

Project Co-ordination

Prof N Hoose, Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London

Project Objectives

The project has developed and demonstrated the potential of diverse, low cost sensors to provide data for the planning, management and control of the environmental impacts of transport activity at urban, regional and national level. This includes their implementation on vehicles and people to act as mobile, real-time environmental probes, sensing transport and non-transport related pollutants and hazards.

The overall aims of this project are:

- To harness the potential of diverse, low cost and ubiquitous environmental sensors to provide data to address key scientific challenges in the field of transport and environmental monitoring and modelling and analysis.
- To develop a flexible and reusable e-Science infrastructure to support a wide range of scientific, policy-related and commercial uses and applications for the resultant data and to demonstrate the operation and utility of this infrastructure in diverse case study applications.

These aims lead to a number of specific objectives:

1. To develop the capability for suitably equipped vehicles and individuals to act as mobile, real-time environmental probes, sensing transport and non- transport related pollutants and hazards.
2. To extend existing e-Science, sensor, communication and modelling technologies to enable the integrating of data from heterogeneous fixed and mobile environmental sensors grids in real time to provide dynamic estimates of pollutant and hazard concentrations.
3. To demonstrate how these data can be usefully correlated with a wide range of other complementary dynamic data on, for example, weather conditions, transport network performance, vehicle mix and performance, driver behaviour, travel demand, pollutant exposure and health outcomes.
4. To implement relevant e-Science tool sets and (fixed and mobile) sensor and communication system in a number of selected real-world case study applications, involving close collaboration with business and the public sector, and to thereby demonstrate their value to the research and policy community.
5. To actively disseminate the results of the research to the wider research and policy community

Advances in sensor technology

Newcastle University developed a "mote" device designed to enable dense deployments in urban areas to allow continuous, minute-by-minute monitoring of the pollution distribution over an area. The devices are battery powered to facilitate simple installation and use a low-power, ZigBee-based communication protocol to route data between nodes and to a central gateway for backhaul to the data management system. Typically a network of sensors is deployed with sensors attached to street furniture (e.g. lamp columns and railings) with up to 80 metres between nodes. Each local network has a special "gateway" communications node to route communications to the wider world. The gateway can be mounted at a high level on a lamp column to gain access to mains power from the lantern and uses GPRS to provide wireless communication via the Internet to the remote server. A variety of low power sensors can be mounted on the basic mote platform. Within the project devices sense CO, NO2 and noise and have onboard GPS for positioning. Use of ultrasonic sensors to count vehicles was also investigated.

University of Cambridge developed four broad groups of sensors – types A-D. Type A devices sense CO, NO and NO_2, have onboard GPS and connect to a mobile phone via Bluetooth to transmit data over the mobile phone network. Type B devices sense CO, total volatile organic compounds (VOC), e.g. benzene,) and CO_2 and have onboard GPS and GPRS capabilities. Type C devices are similar to type As but with onboard GPRS for direct data transmission without requiring connection to a mobile phone. Type D devices sense CO, NO_2 and temperature and connect to a mobile phone. A microcontroller on the type D devices records pollution and temperature data making use of the analogue microphone input on the phone to pass sensor data to it for transmission over the mobile phone network.

Duvas Technologies Ltd, a spin-out from Imperial College London, developed a high-performance multi-species mobile sensor, based on ultraviolet differential spectroscopy, designed for vehicle mounting or being carried by hand. The developed devices employed a variant on Differential Ultraviolet Absorption Spectroscopy (DUVAS), itself being a derivative of Differential Optical Absorption Spectroscopy (DOAS). DOAS originally stems from atmospheric physics research, where the analysis of narrow-band absorption features overcomes the need for calibration and provides an absolute measure [6]. Moreover a number of key pollution species absorb in the UV, and the instantaneous spectroscopic technique provides a high degree of selectivity. The sensors provide high resolution real-time measurements of a range of species including oxides of nitrogen (NO/NO_2), sulphur dioxide (SO_2), ozone (O_3) and benzene (C_6H_6) at part per billion levels. The field units were developed for mobile deployment [4] with the objective of providing high time/spatial resolution measurements across a wide area. The Duvas Technologies/Imperial devices have the ability to send data using 3G/GPRS via the mobile phone network enabling greater autonomy and reliability in data transmission but also have support for multi-hop WiFi data transmission with a store and forward configuration enabling data to be buffered locally on the device when no connectivity is available to transmit data. Server-side proxies allowed for the transmission of data to several databases. Tests were conducted to investigate the potential for ad hoc wireless communications between sensors and infrastructure nodes, in particular multi-hop data buffering techniques during intermittent communication regimes. This was found to be effective, although the vehicle dynamics and the consequent node topology have a strong bearing on the connectivity of the network.

Use of wireless communications

The three sensor platforms developed as part of the project also required wireless communications to support the real-time transfer of data into the remote databases and data processing systems.

Cambridge investigated the potential for personal devices (mobile phones) to support a sensing system utilising either the wireless "bluetooth" connection that is more or less ubiquitous on mobile phones or linking to the phone via its "audio"port. Data can then be transferred using the General Packet Radio Service (GPRS) available on most mobile phones and networks. GPRS provides the ability for the device to continually stream data whilst on the move. The type C device obviated the need to use the mobile phone of the person carrying the sensor by building the communications electronics and software into the sensor device.

Newcastle University have developed a "smart-dust" network using Zigbee (IEEE 802.15.4) to connect together the sensor motes. Data can be "hopped" from mote to mote until it reaches the "gateway" node, effectively a special mote with a GPRS connection. The low power requirements for Zigbee communications meant that sensor motes can be operated for several months from D Cell batteries. The same is not true for GPRS which consumes significantly more power and therefore the gateway node must have a permanent power supply. Motes can be several tens of metres apart provided there is a good line of sight to the next mote in the network. This means careful survey and design of the locations of motes and ensuring the the antenna is not being masked by, for example, the lamp post where the device is mounted. A certain amount of trial and error is inevitable and some motes may need to be re-positioned. The team successfully installed networks of more than 50 motes with more than 90% successful data retrieval.

Imperial investigated the use of WiFi (IEEE 802.11.g) and WiMax (IEEE 802.16) technologies for communications and positioning of the DUVAS sensors. Use of WiFi would have enabled ad-hoc networks to be created between sensors on different vehicles and between them and fixed WiFi or WiMax nodes. Given the potential for the coordinated deployment of network nodes on centrally-managed fleet vehicles, the project modelled the formation of vehicle ad-hoc networks (VANETS) featuring vehicles of different types, with systematically different behavior patterns. The connectivity of VANETs that consist of buses moving in urban area, was studies.. Buses have a unique set of behavior characteristics, such as fixed routes, schedules, bus stops, etc., which gives rise to distinct impact on node connectivity in the communication network. Through extensive simulations based on real bus routes in central London, model results were generated showing the impact of the locations of stops and the prevailing traffic patterns on node connectivity and investigate the fundamental characteristics of the contact duration and inter-contact time between buses. It became clear that witht eh number f DUVAS devices that the project could afford (6) that a VANET approach was not possible so GPRS was also utilised. However for a wider implementation there are some characteristics of WiFi that can be exploited.

The Global Positioning System(GPS) is widely used to provide spatial awareness. However, it has limited performance in dense urban areas due to signal attenuation and blockage. Different positioning systems have been developed to address GPS limitations. However, most suffer from either low accuracy, short range, the effect due to mobility or a combination of some of these factors. The part MESSAGE project addressed these issues through the exploitation of the Wi-Fi signals that are used to transfer data among sensor nodes and to the infrastructure network. The IEEE802.11g WLAN standard has been selected differential *time difference of arrival* (DTDOA) has been applied to avoid the need for precise synchronisation between reference nodes and a two-stage algorithm has been proposed to accurately estimate the arrival time of signals at the reference nodes. Oversampling at the reference nodes has been applied to improve time resolution. Simulation results show that the two-stage arrival time estimation method can measure distances to within 1m range error under reasonable signal to noise values.

System Architecture

The MESSAGE e-Science architecture was developed to address the technical challenges facing the full project pipeline, from the sensor devices themselves, through the communications layers, to the real-time storage and long-term data warehousing components. The architecture also supports a potentially unlimited range of higher-level applications and algorithms and makes provision for application specific "data marts" to support them. Integration of third-party data is also possible to support various algorithms and data analysis that may be undertaken. Communication back to the sensors allows for sensor mode operation changes in the event of localised clustering.
A high-level view of the e-Science architecture is shown in Error: Reference source not found and the various elements of the architecture are discussed below.

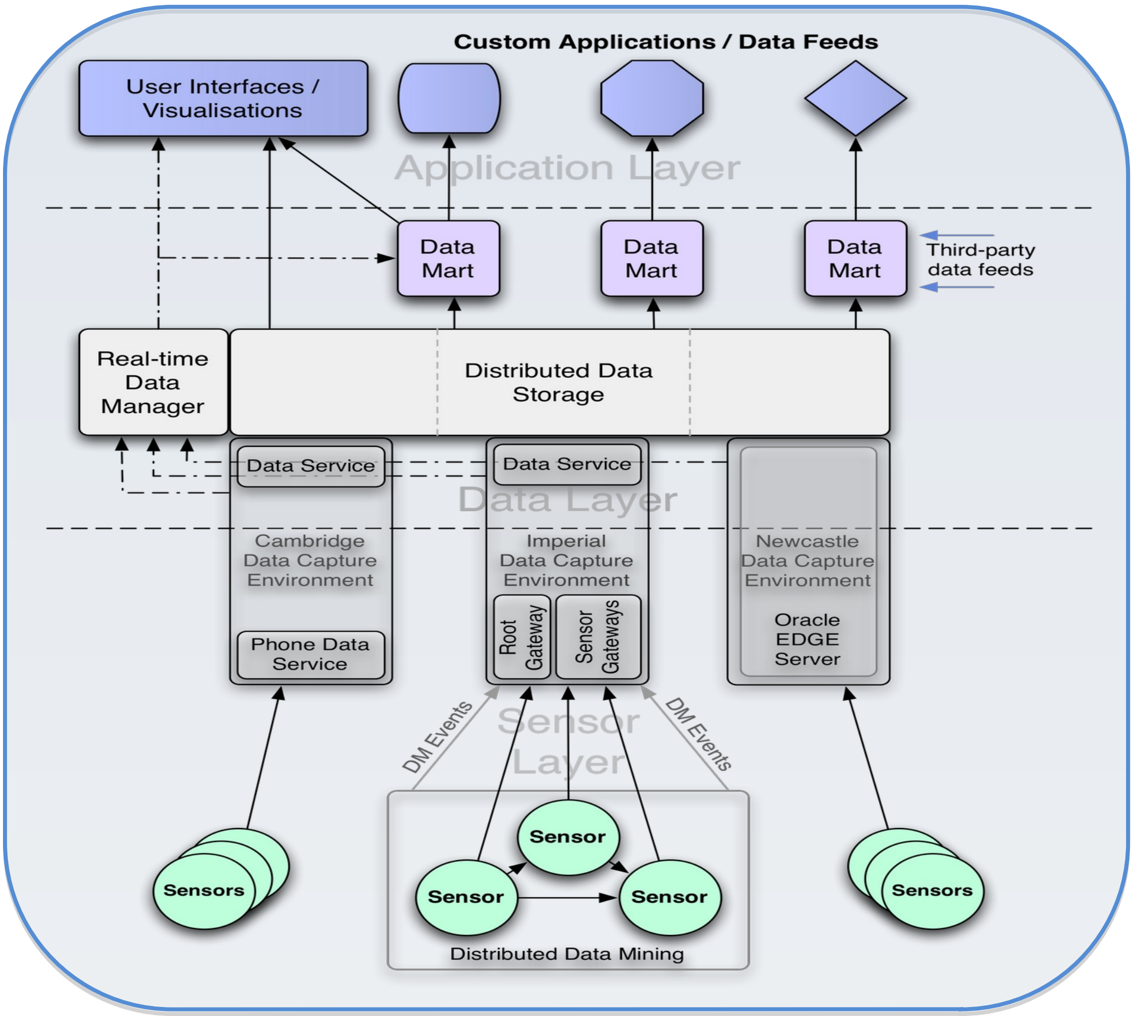


Figure 1: MESSAGE high-level e-Science Architecture

Sensors generate packets containing data that they have collected and send this data to an endpoint exposed by the MESSAGE data management infrastructure. The location that sensors send data to may be hard-coded within the sensors' onboard software or it may be resolved dynamically when a sensor device comes online. Due to the differences in sensor devices, their onboard computing power and technical capabilities, MESSAGE took the approach of developing data capture stacks closely coupled to sensor types in order to ensure maximum efficiency in development and operation. A common database schema was developed for the underlying databases used for data storage and a common interface was developed for data extraction.

Data Management

Imperial College London developed a dynamically scalable data capture infrastructure based on a set of Web Services – a Root Gateway, Sensor Gateways and a Data Storage Service. This infrastructure was developed to handle XML formatted data packets being transmitted by Duvas Technologies' sensors. Table segmentation was allowed to minimise database operation times on large datasets. The infrastructure was designed with a view to being able to use new, dynamically scalable, pay-per-use Cloud computing infrastructure to allow the number of computing resources to grow or shrink in proportion to the number of sensors active at any point in time. University of Cambridge sensors use onboard GPRS to send packets of data to a known endpoint. The server receiving the data is able to carry out pre-processing on the data, where required, and then handles storage of the data into a database. University of Newcastle's sensors generate efficiently packed binary packets of data and pass this data, via a gateway, to an Oracle Sensor Edge Server designed to handle feeds from multiple sensors and manages the pre-processing of data and storage into an underlying database.

University of Newcastle led the development of a common database schema for MESSAGE. This schema is based on a modified version of the Urban Traffic Management and Control (UTMC) [8] guidelines. Modifications were made in order to support additional data being generated by certain types of sensor that are not directly supported in the UTMC specification. As a result of legacy requirements the different groups were inevitably operating databases on different DBMS. In order to access multiple heterogeneous databases through a common interface, the OGSA-DAI (Data Access and Integration) [9] and DQP (Distributed Query Processing) [10] software, initially developed under the UK e-Science Core Programme and now a product of the UK Open Middleware Infrastructure Institute (OMII-UK), was selected. OGSA-DQP provides a single point of contact that can be used to specify queries referencing data across multiple sites. A simple web-based interface has been developed in collaboration with OMII-UK to enable SQL queries to be entered into a browser and addressed directly at different MESSAGE databases or at all databases via DQP. The results can be made available in a variety of different formats, currently CSV, XML and KML.

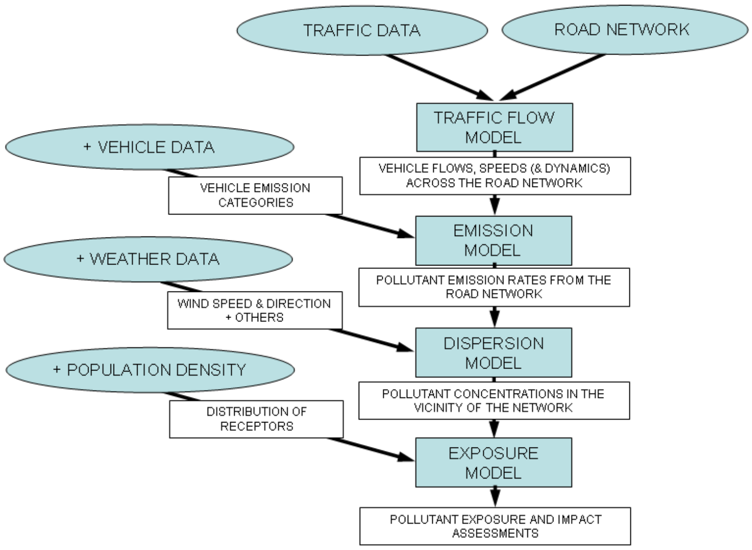
The top level of the MESSAGE architecture provides an environment for applications and associated data marts. A data mart is a pre-processed data store that is formed from processed sensor data that may have been combined with other third-party data. Data marts may be the result of running data through one or more complex algorithms and the process of building/maintaining the content may be time consuming and computationally intensive. Applications may take data directly from a data mart, or from the raw MESSAGE data store.

The real-time data management element of MESSAGE uses a special instance of a data-mart paired with a real-time data manager that takes a feed of raw sensor data from the data management layer and parses and pre-processes this data before pushing it directly to registered clients for display. It was recognised that real-time data processing could not be linked directly to the main data store due to the latency and inefficiency of storing data and then trying to read it back from the database immediately.

When analysis is time critical, there are more efficient approaches. MESSAGE took advantage of the potential for localised groups of sensor devices to make ad-hoc communications links by developing methods for in-network data mining where groups of sensor devices could collaborate on the analysis of data they have captured in order to identify problems such as the build up of a pollution hotspot. When such an event is identified, one or more sensors from the collaborating group send a trigger to the master platform highlighting the problem so that it can be logged and remedial action taken. The in-network distributed data mining process is based on the use of a k-means clustering algorithm that is designed to identify small clusters of readings that are outside of acceptable limits and are likely to represent a pollution 'event'.

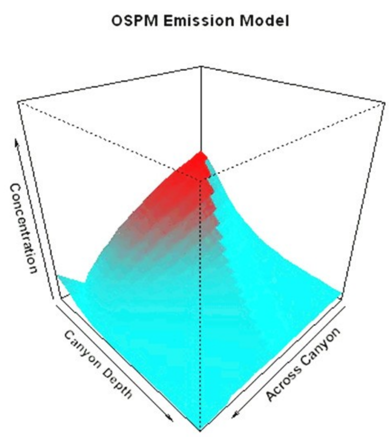
Data Modelling

Advances in pollution sensing networks are increasing our ability to detect hotspots, however, the appropriate course of remedial action is not necessarily obvious, particularly when considering complex urban road networks. In order to evaluate the impact of alternative strategies, it is necessary to employ high resolution, microscopic traffic and emission models to adequately represent the effects of changes in traffic behaviour on vehicle emissions (see Figure 4) and the effect on air quality. These models are computationally expensive, and require multiple runs for each scenario of interest in order to provide a robust output. Therefore on-demand utility computing resources are used to provide a scalable distributed computing environment for the evaluation of the emissions associated with alternative traffic management strategies. By making a model stack of traffic flow, the emissions caused by this traffic flow and the dispersion in the air, it is possible to analyse different traffic management scenarios and assess the impact on the environment in real time, such that the appropriate measure can be taken.



Modelling is used by the traffic manager to design an appropriate control strategy to alleviate pollution hotspots. Often several options are considered (such as queue cascading or re-location to less sensitive areas of the built environment or open spaces to facilitate dispersion) and the most suitable for the specific problem is implemented. Crucial to the accuracy of the air quality prediction is the validation of the traffic micro-simulation model. Parameters usually validated include, link journey time and flows as well as queue length.

Data from SCOOT, Split Cycle Offset Optimisation Technique (TRL, UK) is received as a real-time web service and processed (using data mining techniques) to generate the traffic flow regimes, congested, busy, smoothed and free flow. Maps provide assessments of the flow regimes, congested (red), busy (yellow), smoothed flow (blue) and free flow (green) across a SCOOT Region. This data provides an information platform for the decision maker. Data is passed to the warehouse where, using the UK emissions factors, the traffic emission (carbon monoxide, carbon dioxide and nitrogen dioxide) is estimated for each cycle. Finally the Canyon OSPM model is used to predict the roadside pollutant levels. Dispersion depends on wind direction, and strength, the height of the buildings and width of the road. The predicted pollutants are compared with those measured by the motes.



Data Visualisation

Visualisation of data provides many opportunities for the observation of key properties, patterns or events that may not be immediately obvious when looking at raw data values. In MESSAGE, various different options for visualisations were tested. A web browser-based visualisation (see Figure 4) showing near real-time sensor data and basic analysis of historic data is one of the means of visualisation resulting from the project. The interface is built using Adobe Flex and uses a Google Maps window on which to overlay sensor information. The OGSA-DAI/DQP web query interface described earlier allows data to be output in KML format and this can be imported directly into Google Earth/Maps and visualised. Other visualisation techniques were investigated during the MESSAGE project including using multiple map layers to overlay different pollution data layers on a map, real-time pollution trails, area-averaged grid tiling, and height-based wall maps.

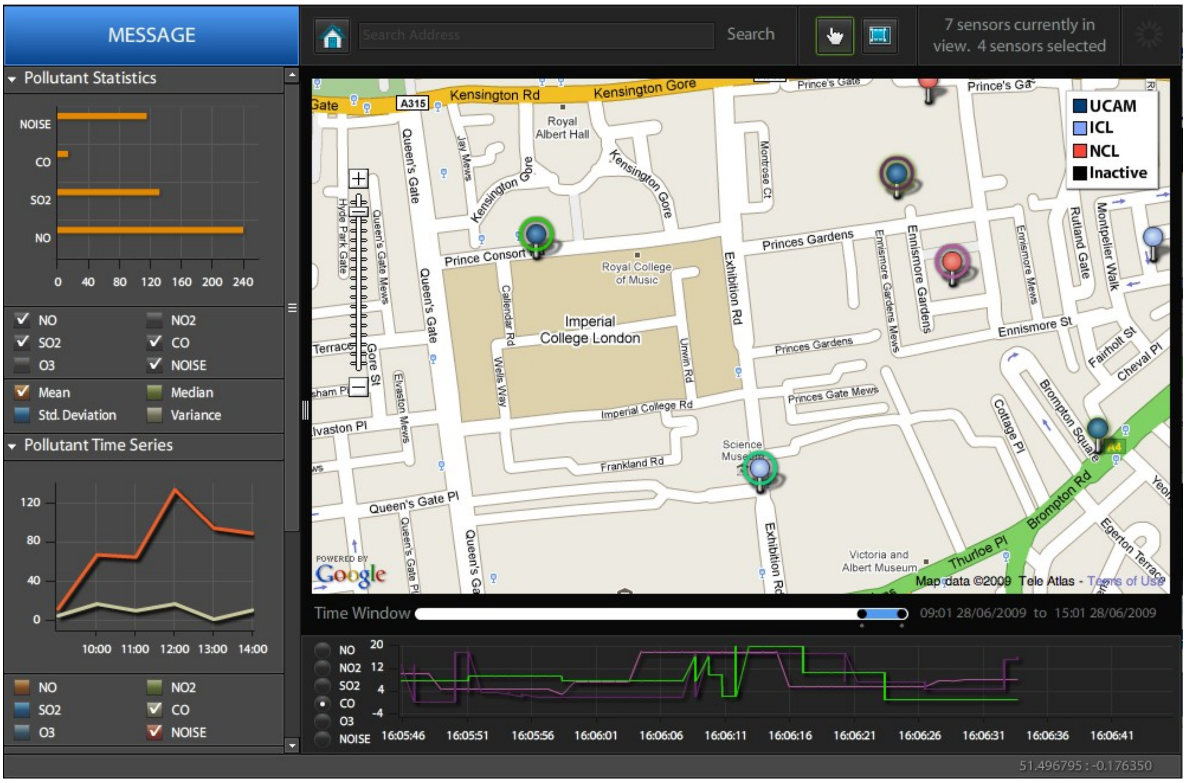


Figure 2: The MESSAGE web-based user interface (Google Maps display Powered by Google, Map data ©2009 Tele Atlas)

Deployments

Newcastle University developed a "mote" device designed to enable dense deployments in urban areas to allow continuous, minute-by-minute monitoring of the pollution distribution over an area. During the trial deployments described, Newcastle University tested approximately 130 sensors and the robustness and reliability of the Newcastle MESSAGE system architecture. Sensor operation has been verified over the temperature range -9 to +38°C.

Typically a network of sensors is deployed with sensors attached to street furniture (e.g. lamp columns and railings) with up to 80 metres between nodes. The first deployment of forty static and ten dynamic sensors was carried out in Gateshead during August to December 2008 [3]. In parallel with these deployments, reference measurements were undertaken in conjunction with the University of Leeds Instrumented City facilities. The communications, software and sensor performance was fully tested. As a result of four months of deployment and evaluation, a new generation of the hardware and software was produced.

The transferability and scalability of the system was demonstrated in other UK cities. In Leicester, the largest array of 50 motes was deployed in the period February – September 2009 across an area of about 1km$^{2.}$ Leicester was chosen as a demonstration site because the wealth of historic data collected since 1987 enables the features of the MESSAGE data warehouse to be demonstrated. Also, it has an air quality management system, Airviro, which uses both real-time traffic data (from SCOOT) and meteorological conditions to produce "nowcasts" and "forecasts". The street canyon pollutant predictions from the Airviro model for the 50 mote locations are brought to the server as a web service to enable correlation with measurements.

A high degree of operational stability and scalability has been demonstrated with further deployments in Newcastle, in London in South Kensington close to the Natural History Museum and at three sites in Palermo (Italy). In Newcastle, a network of 30 motes was operated for nearly 4 months; in Gateshead, 30 static motes were operated for 5 months and supplemented by 10 mobile motes for a two week survey; in London, a network of 12 motes operated for 3 months in conjunction with the co-ordinated system trials; in Leicester, 50 motes operated for 7 months allowing investigation of the impacts of changes in traffic levels and in Palermo, a set of 3 motes were installed at two different air pollution monitoring cabins for two weeks during each of Winter, Spring and Summer.

The UCAM air quality sensor units have been deployed on a number of occasions in Cambridge, London, Valencia and at Cranfield airport. The most recent deployment in Cambridge was composed of approximately 40 sensors deployed multi modally for approximately 3 hours in August 2009. The sensors were a mix of A, B, C and D sensors deployed using pedestrians, cyclists, motorists and at static sites alongside Cambridge Council monitoring sites. The coverage for the deployment is shown in Figure 4, a plot illustrating nitric oxide (NO) and NO_2 concentrations with height as a proxy for concentration.

Figure 4: NO (Blue) and NO_2 (Green) concentrations variation across Cambridge. (Google Earth screenshot content ©2009 Google - Imagery ©2009 Infoterra Ltd & Bluesky)

These plots are a "snapshot" of the levels of the measured air quality gasses in the deployment areas. The Cambridge mass deployment dataset alone incorporates approximately 100,000 measurements . The sensors were also deployed using a different method at Cranfield airport. The aim of the experiment was to capture data related to the aircraft engine exhaust plumes as the BAe-146 took off. Eight take off episodes are clearly marked by increased NO_2 concentrations. A total of over 50 sensors have been deployed and tested. These deployments have demonstrated how large numbers of low cost sensors can be deployed in a co-operative manner to capture a snapshot of the pollutant distributions across an area.

Imperial College led on the development of a novel, compact design, mobile/portable multi-species gas analyser to complement the low-cost electrochemical sensors used in the Cambridge and Newcastle units. Initial deployments in December 2008 demonstrated the potential to gather data from probe vehicles across long trajectories and in the process to identify localised hotspots. A set of further tests were then conducted around a test site in South Kensington where the same circuit was driven repeatedly using two vehicles to identify whether mobile probes could be used to capture the temporal variability in pollution levels at a given location. These results suggested that while there is a high degree of variability between individual trajectories, the mean values over the course of an hour, and the average spatial distribution of pollution measured by the two vehicles was very similar.

A set of DUVAS units were adapted for use as portable reference units running on batteries and were carried around the test site by pedestrians and cyclists [7]. Extended sampling at a single location allowed the temporal variation to be captured. This is shown in Figure 9 where a unit was deployed near the centre of a busy intersection. The periodicity in the measured values is seen to correspond to the traffic signal timings (~48 seconds per cycle) with the relationship between NO and NO_2 also clearly observable. Figure 10 illustrates a comparison between two identical units carried along a main street and junction. Units were carried approximately 1m apart for the initial phase, and then separated at curb-side after a street crossing to make the synchronised comparison between curb and street-side approximately 10m apart.

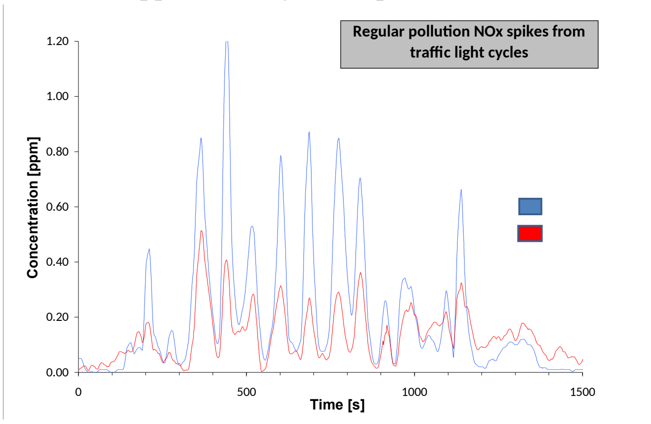


Figure 9: Periodic peaks in pollution levels in the middle of a busy arterial road from a static unit [7]

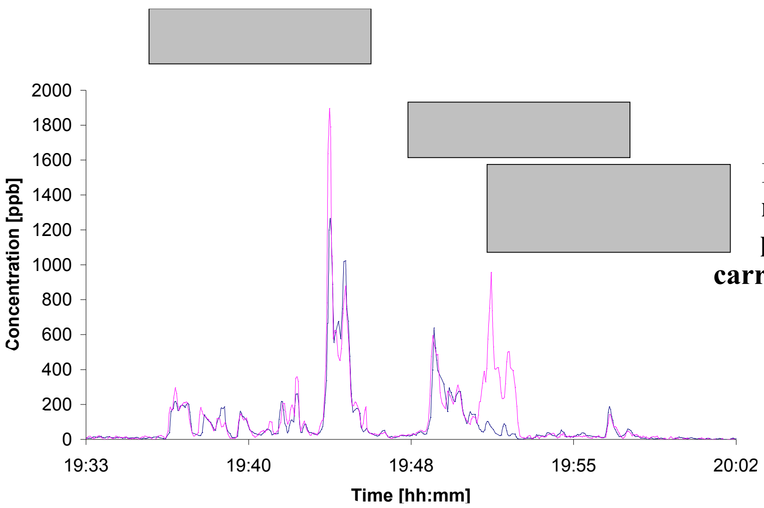


Figure 10 Comparison of measurement of NO using two portable Duvas units being carried along street [7]

The current datasets indicate relatively large differences over a short distance/time periods and therefore both higher spatial and temporal resolution could provide more information and insight of the nature of pollution and exposure. The technologies developed in MESSAGE can further provide the data necessary to improve management of transport and reduce its environmental impacts through quantifying the real world impact of different interventions. These technologies also open a range of opportunities in other application areas.

Potential uses for MESSAGE infrastructure

One of the initial stages of the project comprised a wide ranging review of potential applications, including both understanding the needs of individual travellers, service providers and transport system managers for air pollution data [4] and considering potential applications of the MESSAGE system beyond air pollution monitoring. One of the principle outcomes from this phase of the project was the creation of a range of 'technical scenarios', designed to stretch the development of the MESSAGE components and system infrastructure in different directions.

Theme 1 – Traffic Management

The area of traffic and environmental management is a theme which arose from the desires of local authorities to better understand the spatial and temporal spread of air pollution and to begin to be able to both respond to short term pollution episodes (for example around individual junctions – Figure 1) and understand the detailed patterns within longer term problem areas. Technical scenarios in this theme were therefore focused on examining how increased spatial and temporal resolution in air pollution data can be used for better traffic network understanding control.

Theme 2 – Health and Exposure

The importance of monitoring and understanding how the exposure of individuals to air pollution varies spatially and temporarily was identified by all groups within the initial review, both for travellers in general and specifically for those individuals with identified respiratory conditions/vulnerabilities (e.g. asthmatics or young children). This need is supported by the recognition from epidemiological studies that the health impacts associated with poor air quality appear often to be the result of acute exposures, rather than chronically poor air quality of the same mean level. Technical scenarios in this area were therefore focused on understanding variations in pollutant concentrations (especially in pedestrian areas such as pavements) and on detecting and reporting particularly high pollutant concentration episodes.

Key challenges for the MESSAGE system for these technical scenarios have therefore been in the identification of short-term and transient high pollutant concentration episodes (where the number of available sensors may be very limited), alongside integration of pollutant concentrations over typical pedestrian and cyclist paths to estimate long-term exposure levels.

Theme 3 – Traveller Information

Identified as a key driver within the user needs assessment, demonstrating the provision of real-time and high spatial and temporal resolution air pollution data directly to individual travellers has been one of the most challenging scenarios for the MESSAGE system. Central to this challenge is the need to aggregate sensor data from across widespread spatial and temporal boundaries and yet to make that information reflect the location and journey of individual users.

Key challenges for the MESSAGE system in this theme have therefore been the need to aggregate data from across a wide spatial area in real time; the need to establish and maintain historical profiles for different levels of spatial detail and combine these with real-time data to provide comparison information; and the need to allow for short-term forecasts of air quality data (e.g. from dispersion models) to be included within the collated information.

Theme 4 – Pollution Modelling

The final and most technically demanding applications for the MESSAGE components and system are applications related to the calibration, validation and real-time use of detailed pollution dispersion models. These applications were identified as an area where high data accuracy was required at high levels of spatial and temporal resolution, and as such these technical scenarios were intended to represent the limits of what could be achieved by the MESSAGE components and system.

Key challenges for the MESSAGE project within this theme were therefore concerned with quantifying the numbers of sensors necessary to achieve desired levels of resolution, achieving sufficiently accurate positioning information for each sensor, handling the volumes of data produced at highest sensor resolutions and comparing the real-time data with simultaneous predictions from dispersion models.

Implications from overall system

The outcomes from MESSAGE are important in two major respects. Firstly, the project has brought together a mix of leading edge and legacy technologies in sensing, communications, data management, modelling and visualisation to demonstrate that it is possible to have a unified architecture that encapsulates a variety of heterogenous components at a variety of levels. For example, the project has developed three distinctive types of air pollution sensors, deployed them in the same location at the same time and fed the data into a set of independent databases that are linked in a "federated" manner. This allows data from other systems, e.g. weather, traffic data, to be integrated and still be accessed by a single user interface. This generic infrastructure has many potential applications where large scale sensing and data processing in real-time are needed. The first paper in out suite of 3 describes some of the key features of the "eScience" infrastructure.

Secondly, the project demonstrated that the infrastructure can be used to address a real-world issue, namely improving our understanding of the nature of air pollution from traffic. Substantial numbers of sensors have been, and are being, deployed at the roadside and carried by vehicles, pedestrians and cyclists. The resultant data sets are have a temporal and spatial richness that opens up all sorts of possibilities for techniques and strategies to mitigate the potential effects or pollutants such as, CO, NO_x, ozone, volatile organic compounds such as benzene and particulates. Noise pollution can also be monitored.

The identified need to increase the timeliness and spatial and temporal resolution of air quality monitoring data has, through the range of technical scenarios described here and others considered within the MESSAGE project, enabled the creation and demonstration of a system capable of collecting, transmitting, collating, analysing and storing air pollution information. The outcomes from the MESSAGE project will have significant impact in two ways. Firstly, the project has opened the way for a step change in the collection and analysis of air pollution. Secondly, and just as importantly, it has demonstrated a practical data management architecture that brings together mobile and fixed sensor data, external databases and data processing.

This architecture has potential application in domains beyond air pollution monitoring, and by working with application domain experts to further integrate complimentary datasets and real-time data feeds from other systems (for example meteorological data, urban topology data, etc.) and develop user orientated interfaces, the MESSAGE system can provide data to support a wide range of applications, from those focused on individuals to those focused on wide-scale management decisions. Perhaps the greatest asset of the system therefore is its flexibility. The continual consideration as the system has been developed was to facilitate the integration of unknown future sensor types, communication and positioning technology systems, end user applications and even domains which have not yet even been imagined. This has helped created a system capable of being implemented wherever sensors can report their position and their measurement, not just within the domain of air quality.

The MESSAGE project aimed to address many challenging problems and questions in the management of large quantities of data coming from distributed infrastructure using a variety of different types of connectivity. The project tackled the full stack from capture to use covering the difficult process of taking data from sensor nodes, pre-processing it where necessary and then storing it in an infrastructure capable of making it available for use in a wide range of different applications and processes. The display of data from sensors in near real-time was also covered and a web-based interface was developed to show this data and demonstrate a set of general analysis tasks. As well as demonstrating an infrastructure and prototypes of environments to manage and display data, MESSAGE also served to identify important areas where further work is required in developing and maintaining data management platforms and in making these systems efficient.

Next steps

We are only just at the starting point in interpreting data of this fine scale and the MESSAGE infrastructure greatly enhances our potential understanding about the concentration and dispersion of air pollution from traffic. It also enables us to monitor at the localised level the effects of changes in traffic patterns, perhaps as a result of a macro-level strategy in a different policy area such as increases in walking and cycling or reduction in carbon use. The current datasets indicate relatively large differences over a short distance/time periods and therefore both higher spatial and temporal resolution could provide more information and insight of the nature of pollution and exposure. The technologies developed in MESSAGE can further provide the data necessary to improve management of transport and reduce its environmental impacts through quantifying the real world impact of different interventions. These technologies also open a range of opportunities in other application areas.

In achieving the project outcomes a whole raft of innovations in sensor technology, wireless communications, positioning, database design, data mining, data management and data presentation were necessary, not withstanding the pool of existing research the project was able to draw on. Moving from the laboratory to deployment in the real world is a major challenge and to be able to demonstrate a complete system operating as a whole considerable achievement. It has taken a large amount of time and dedication from researchers and support staff at all levels. Multi-institution, multi disciplinary research is not the norm and everybody involved has to adapt to working effectively in such an environment. MESSAGE not only took forward the technical body of knowledge but has created a pool of human capital and a body of experience of how to achieve real progress in a complex organisational situation.

One of the objectives of MESSAGE was to disseminate the knowledge gained as widely as possible. There have been a wide range of publications with more to come, and the system has been demonstrated with press coverage in Gateshead, London and Cambridge. The project results can be used both to build system for immediate application and as an infrastructure to enable further research. Naturally, the research teams are keen to see the work exploited as widely as possible as well, and would welcome discussions with those who are able to build on or benefit from the MESSAGE results.

Publications

To be added