OPUS

Optimising the use of Partial Information in Urban and Regional Systems

IST-2001-32471

1953 EU/25 2003 EUROPE

Information Society Technologies

Coherent knowledge from disparate data sources

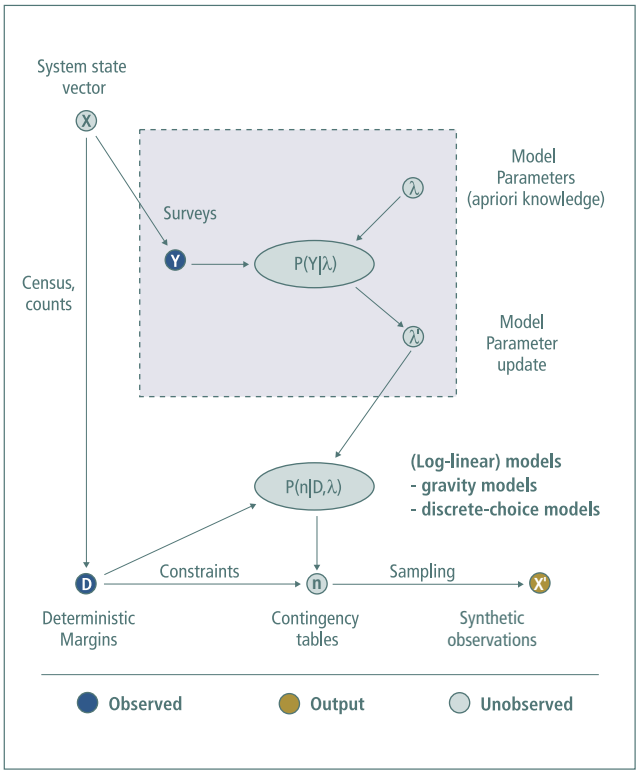
Developing a generic statistical framework for the optimal combination of complex spatial and temporal data from survey and non-survey sources.

OVERVIEW

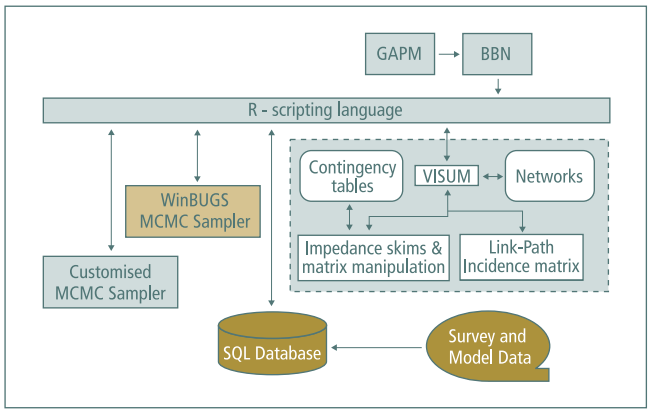
To meet the needs for comprehensive information on socio-economic systems such as urban and regional transport planning, and in the health services sector, data from multiple sources (e.g. conventional sample surveys, census records, operational data streams and data generated by ICT systems) must be combined. However, due to the lack of appropriate general purpose techniques, this process of data combination is often done in an ad hoc and unsatisfactory way.

The overall aim of the OPUS project is to develop a general methodology that enables the combination of information from multiple complex spatial, temporal and real time data sources in a statistically coherent fashion and to thereby contribute to improved decision making in the public and private sectors within Europe. This methodology draws heavily on recent developments in Bayesian statistical analysis and in metadata-based methods of information management.

Transport planning is a prominent example of an application domain that uses multiple sources of data, and will be the main test case for OPUS, but the generic nature of the approach is also being demonstrated through the inclusion of an application in the field of health information. The project is undertaking a number of transport-oriented study applications in London, Zürich, Milan, and on a national level in Belgium. Methods for extending the framework to aspects of the health domain are also being investigated.



METHODOLOGY FOR MULTI-DIMENSIONAL CONTINGENCY TABLE DATA



ARCHITECTURE OF THE OPUS SYSTEM

BENEFITS

- Improved estimation of detailed travel demand, using all available information;
- Avoidance of simplified combination of data that can give erroneous estimates;
- Indicators of data quality, to provide guidance for new data collection;
- A framework for managing data from rolling survey programmes;
- Better understanding of the role of variability and uncertainty in results and models;
- Avoidance of confusion from different, apparently conflicting, estimates of the same quantity;
- Greater user confidence and relevance of results;
- A generalised methodology for other domains of interest.

METHODOLOGY

The methodology is based on the well-known Bayesian statistical concept of using one set of data, represented in the form of conditional probabilities, to update some prior information to produce new, posterior estimates. OPUS builds on the developments in recent years in Bayesian modelling techniques, notably those using Monte Carlo Markov Chain (MCMC) sampling methods, which use computer simulation to provide posterior estimates involving multiple sets of data and forms of probability distribution functions that are not amenable to classical statistical methods.

An 'OPUS methodology' has been developed that defines a procedure for expressing a domain of interest – 'transport' and 'health' in the case of OPUS – and converting this to a Bayesian Belief Network (BBN) that represents the structural characteristics and interactions of different elements of the domain. Once in this form the new estimates may, in principle, be derived using relatively standard MCMC sampling methods, but OPUS has provided a number of distinctive approaches to address the difficulties presented by large-scale problems and that are characteristic of individual domains.

METADATA IN OPUS

The project is developing a system to keep a complete audit trail of the statistical modelling, using metadata. This keeps track of the specifications used for the statistical model, and all the model fitting steps (where evidence about the model is extracted from datasets). This information is linked to results obtained from the model (such as synthetic data), so that a user can review the assumptions and processes used to fit the model (its Provenance) and can explore the quality of the model fit (its Reliability). The information about particular fitted models is stored in XML documents, and the structure and semantics of these documents has been specified using UML. A web application is being developed to display the Provenance and Reliability information.

OPUS & TRANSPORT

OPUS aims to provide transport planning agencies with a process that enhances transport planning and policy making by providing new, synthesised data which offers a more informative and accurate picture of transport making and travel patterns than is feasible by observation surveys alone. As transport modelling is one of the sources of information used by transport planners and policy makers, the ability of OPUS to incorporate data not only from a range of surveys but also from transport models is of value. The characteristic spatial qualities of transport data, such as transport networks and zonal tables of travel patterns (OD matrices) provide particular challenges that the OPUS method has addressed through the application of specialist MCMC samplers.

OPUS & HEALTH

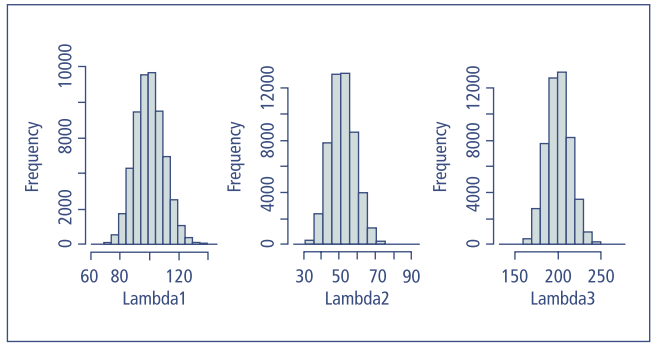
The World Health Organization has defined potential areas of application for OPUS methods within the health domain. An area of relevance to the development and utilisation of OPUS, that is particularly important to the health sector, is the modelling of spatio-temporal exposure to health data.

Health Impact Assessment studies were identified as a potentially suitable application where the OPUS methodology is likely to be beneficial. These kinds of studies translate the clinical and epidemiological evidence provided by scientific publications into impacts on exposed population and are of growing interest for policy-makers in Europe.

FEASIBILITY STUDIES

The demonstration of the OPUS methodology to the transport domain has been conducted through two sets of Case and Feasibility studies. The Case Studies, respectively for the cities of London and Zürich, use the extensive sets of data that are available to generate new information on travel demand patterns. A common point of interest is to reconcile differences in the data from different surveys, as well as providing a way of allowing new surveys to update information that is degrading through the passage of time. A focus for the Zürich Case Study is the use of data from networks, such as counts and speeds, while the London Case Study demonstrates the integration of diverse data, which is subject to various limitations induced by the difficulties of conducting travel surveys, to produce information that supports sub-regional studies within London.

Detailed 'desk studies' will also illustrate the potential of the OPUS method for the Lombardy Region in Italy and for Belgium. The Belgium Feasibility Study compares the OPUS methodology with recent work on developing a national set of transport demand data for the country. The focus of the Italian Feasibility Study is to demonstrate how the OPUS methodology can be used to exploit extensive data collection recently undertaken across Lombardy, together with other data such as Census information, to derive more disaggregated sets of travel demand information.



POSTERIOR DISTRIBUTIONS FOR OUTPUTS (eg TRAVEL COUNTS)

OPUS PARTICIPANTS

Research Organisations

CTS (Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London), United Kingdom - Lead Partner
DEPH (Department of Epidemiology and Public Health, Imperial College London), United Kingdom
ETH Zürich (Institut für Verkehrsplanung und Transportsysteme), Switzerland
FUNDP, Transport Research Group (Facultés Universitaires Notre-Dame de la Paix), Belgium

Practitioners

Minnerva Ltd., United Kingdom
Survey and Statistical Computing, United Kingdom
Katalysis Ltd., United Kingdom
PTV AG, Germany
Systematica SpA, Italy

Public Bodies

Transport for London (TfL), United Kingdom
World Health Organization, European Centre for Health and Environment, Italy

Internal reviewer

Oxford Systematics, Australia

Contact: Prof. John Polak at info@opus-project.org
Further information on OPUS is at
www.opus-project.org
where you can also join the User or Discussion Forums