

OPUS

Optimising the use of Partial information in Urban and regional Systems

Project IST-2001-32471

IST Programme

Title : **Evaluation**

Author(s) : Axhausen, K.W. (IVT, ETH Zürich)

Deliverable No. : D12.2
Version : 1

Date : Final

Dissemination Level : LI — Limited to programme participants
Deliverable Nature : RE — Report
Deliverable Type : PD — Programme Deliverable

Project Coordinator : Imperial College London
Contact Person : Prof. John Polak
Address : Centre for Transport Studies
Department of Civil and Environmental Engineering
Imperial College London
London SW7 2AZ
United Kingdom

Telephone : +44-(0)20-7594.6089
Fax : +44-(0)20-7594.6102
e-mail : j.polak@imperial.ac.uk

Consortium : CTS, DEPH, TfL, KATALYSIS, ETHZ, FUNDP, PTV,
SYSTEMATICA, WHO.
MINNERVA, SURVEY AND STATISTICAL
COMPUTING

TABLE OF CONTENTS

TECHNICAL ABSTRACT	1
EXECUTIVE SUMMARY	2
1. PURPOSE	3
2. QUALITATIVE ASSESSMENT OF THE CASE STUDIES.....	4
2.1 Feasibility study: Lombardy	4
2.2 Feasibility: Agent generation	4
2.3 Feasibility study: Health impact assessment.....	4
2.4 Feasibility study: Epidemiology	5
2.5 Case study: origin-destination updating	5
2.6 Case study: Leisure excursions	5
2.7 IVT: Updating the regional origin-destination matrix using multiple data sources	6
3. INITIAL ASSESSMENT OF THE PROJECT IMPACTS	7
4. SUMMARY AND OUTLOOK.....	8
5. LITERATURE	9

TECHNICAL ABSTRACT

The deliverable D12.2 evaluates the progress made by the OPUS project, in particular through the case- and feasibility studies, on the basis of the evaluation plan (Deliverable D12.1), as far this is possible.

EXECUTIVE SUMMARY

This Evaluation is Deliverable D12.2 of the Fifth-Framework project OPUS. It implements the evaluation plan (Deliverable D12.1), as far as this is possible, given the work undertaken and reported in case and feasibility studies (Deliverables D11.3; D10.3; D8.2 and D9.2).

1. PURPOSE

The OPUS project has been developing a general statistical framework to improve the combination of complex spatial and temporal data from survey and non-survey sources. The approach is Bayesian in principle and acknowledges the structural relationship between the variables of interest while accounting sampling and non-sampling errors. The framework has been applied in a set of case and feasibility studies of differing complexity drawn from the transport and health fields.

The technical annex describes the specific objective of this work package and therefore deliverable as:

- This work package will provide an independent evaluation of the effectiveness of the methods in a practical application. It will concentrate on validation of results and examine the extent to which the results represent an improvement on previous methods based on input from all project partners.

The changes in the timetable and the results obtained since the start of the project have led to a readjustment of these objectives, so as to maximise the contribution of this report to the project. The deliverable will focus now on:

- Sketch of the case and feasibility studies
- A qualitative assessment of individual contributions
- An initial assessment of the project to the wider practise

The next section will assess the case and feasibility studies, while also describing them. The concluding section will offer an initial assessment of the project to the wider practise in the fields, on which it has focussed.

2. QUALITATIVE ASSESSMENT OF THE CASE STUDIES

Evaluation is not context independent. While there are general objectives, an evaluation procedure should account for the specific case to be able to capture the essence of it. The discussion below will therefore include a description of each of the cases and feasibility studies. Due to the time pressure at the end of the project, in turn due to software issues, none of the case studies was able to perform comparative tests originally intended. The assessment has to be qualitative.

2.1 Feasibility study: Lombardy

The feasibility of a large scale enrichment and spatial disaggregation of an existing origin-destination matrix is discussed for the Lombardy feasibility study in the context of a new river crossing. The study clearly sets out the available and needed data. Using the conceptual tools of the OPUS project it specifies the tasks and data flows using the graphical model. It is clear, how this application has benefited from the conceptual OPUS tools through an increased awareness of Bayesian Belief Networks, but specifically through the tools advanced in the project.

2.2 Feasibility: Agent generation

The generation of agents for micro-simulation models has been a prominent concern for transport modelling since the development of the Dutch National Model in the late 1970's, but has become so through the wide-spread adoption of the agent-based paradigm in recent years. The feasibility study explores how the Dominici samplers could be used for this purpose in the context of a Belgian study. It carefully constructs the general a-priori model (GAPM) and the Bayesian Belief Network derived from it. It concludes that the application should be possible, but also that the scaling of the samplers with regards to speed and stability will be crucial. The study makes clear, how the OPUS concepts improve the structuring and development of solution approaches to complex data fusion problems.

2.3 Feasibility study: Health impact assessment

The World Health Organization discussed the potential areas of application for OPUS methods within the health domain and identified as an area of relevance to the development and utilization of OPUS, which is particularly important to the health sector, the implementation of health impact assessment models in a Bayesian perspective

addressing all the form of uncertainty. It has taken this further by implementing a model of the health impacts of particulate matter exposure in terms of attributable deaths and years of life lost. The deliverable reports the results of the application of the methodology for a set of Italian cities in substantial detail. The stress, which OPUS has laid on the presentation of the uncertainty of any results, is clearly visible in the reporting of the results.

2.4 Feasibility study: Epidemiology

The Department of Epidemiology and Public Health at Imperial College has integrated transport, time activity data with airborne pollution and hospital admissions data to improve estimates of the impact of transport use on cardiovascular disease in Sheffield. This successful practical application has shown how interdisciplinary work can benefit from the OPUS concepts, which give each data source its own weight and uncertainty. The results show once again, how important it is to keep track of the inherent variances of our model estimate.

2.5 Case study: origin-destination updating

This case study is actually more than just a case study, as it develops an approach to overcome the challenge of the scale generated by the problem, which is to estimate the demand matrix for a subarea of central London. Guided by the GAPM and the associated Bayesian Belief Network the case restructures the overall model by introducing an Expectation Maximisation approach and by subdividing the overall problem. These two steps allow the estimation of the matrix of interest. The authors are able to show that the approach is able to generate credible estimates integrating two different data sources. The exercise is repeated employing the Dominici approach, again confirming the overall experience. Again, the OPUS concepts help to improve the treatment of a complex and large scale problem. It also shows that the concepts lead to practically useful model structures and results.

2.6 Case study: Leisure excursions

This case study addresses the combination of various data sources to obtain a consistent and richer estimate of the total amount of travel. Such improved and joint estimates are urgently needed on the one hand to benefit from multiple data collection efforts, but also on the other hand to achieve consensus on the loads of the transport system. In the Swiss case three sources provide information about the number of leisure excursions and day trips: the national travel survey (Mikrozensus Verkehr 2000), the national income and expenditure survey (Einkommens- und Verbrauchserhebung 2000) and the privately funded survey Schweizer Reisemarkt (Swiss Travel Market 2000). These three are inconsistent in their scopes, partially in their object definitions and their sampling processes. Combining a model of the trip numbers, based on the consistent parts of the

data, with a model of the different sampling procedures the case study was able to obtain a coherent estimate of the total trip making, which is clearly much more than just a weighted average. This is one of the first applications of Bayesian Belief Networks in this domain. It will alert other practitioners and researchers to the range of possibilities available to them. In keeping with OPUS the case study highlights the gains obtainable from additional information through the estimation of partial models, but here on the three data sets simultaneously.

2.7 IVT: Updating the regional origin-destination matrix using multiple data sources

The IVT has access to a range of data which provide information about the traffic volumes in the Zürich area. These sources look at different aspects of the traffic. The automated count data of the Federal Office for Roads (ASTRA) and the automatic and manual counts of the Canton Zürich provide estimates of volumes in each direction at a cross section. At some locations, these are supplemented by speed measurements. The third data source are speed estimates for each link in the regional network derived from a spatial-lag regression model (Hackney, Bernard, Bindra, Axhausen, 2006) of floating car data collected in 2003 by the IVT.

The case study carefully prepared the data following the OPUS concepts. Unfortunately, the algorithms did not scale, in spite of intensive work on the code. This scale of problem will have to wait until the algorithms have been improved further.

3. INITIAL ASSESSMENT OF THE PROJECT IMPACTS

The OPUS project can be seen as part of two wider trends in the practise of modelling, or in all disciplines, which rely on statistical models for their advance. The first trend is the acknowledgement that point estimates are only the smaller part of the modelling results of interest. The second trend is the more systematic description and archiving of the model, its structure and estimates, and of the underlying data, which themselves might be model derived through imputation or external systems, such as transport models. While the methods associated with the first trends are well established in the health research domain, they are hardly so in transportation research. Through its papers, presentations and now case and feasibility studies, the OPUS project has helped to change this in the transport domain. While the adoption process will continue for some time, the OPUS project has opened the way. There is no need here in the final deliverable of the project to restate the benefits of the approach.

In addition to opening the way in general, the project has advanced two specific methods, the Tebaldi and Dominici samplers, from small scale to large scale application and solved the problems associated with this switch in scale. In the case of the Dominici sampler, the project has also developed a computationally faster Maximum likelihood interpretation, which will help to make the approach popular (Evers and Santapaola, 2007).

With the development of a systematic meta-data format, StatModel, for the description of statistical, in particular Bayesian-type, model, OPUS has demonstrated how one can close this important gap. StatModel, while functioning on its own, outlines how standard statistical tools could integrate the archiving of data manipulation, imputation and modelling in a comprehensive manner. The OPUS project is convinced, that this approach will see such integration in the near future, as it would improve the productivity of statistical work substantially.

4. SUMMARY AND OUTLOOK

The deliverable has set out the evaluation approach for the case and feasibility studies. It provides for both a quantitative and a qualitative approach, respecting the different tasks of the case and feasibility studies, as well as the differing levels of implementation, which were feasible during the project.

5. LITERATURE

Axhausen, K. W. and M. R. Wigan (2003) Public use of travel surveys: The metadata perspective, in P. Stopher and P. M. Jones (eds.) *Transport Survey Quality and Innovation*, 605-628, Pergamon, Oxford.

Chalasanani, V.S. and K.W. Axhausen (2005) Conceptual data model for the integrated travel survey and spatial data, in R. Khan, R. Banks, R. Cornelius, S. Evans and T. Manners (eds.) *Proceedings of ASC 2005 Maximising Data Value*, 123-135, ASC, Chesham.

Chalasanani, V. S., S. Schönfelder and K. W. Axhausen (2002) Archiving travel data: The Mobidrive example, *Arbeitsberichte Verkehrs- und Raumplanung*, **129**, IVT, ETH, Zürich.

Evers, L. and D. Santapaola (2007) On the use of the IPF algorithm for combining traffic count data with missing dimensions, paper presented at the paper presented to *86th Meeting of the Transportation Research Board*, Washington D.C.

Lindveld, C., and J.W. Polak (2004) Combining separate datasets in the OPUS project, paper presented to the 6th International Conference on Social Science Methodology, Amsterdam.

Lindveld, C., M. Logie and J.W. Polak (2006) A Bayesian framework for the integration of multiple large scale incomplete transport data source, paper presented at the *30th Annual Universities Transport Studies Group Conference*, Dublin.

Lindveld, C., M. Logie and J.W. Polak (2006) Using Bayesian belief networks and process metadata to address large scale data integration problems, paper presented at the *23rd European Transport Conference*, Strasbourg.

Lindveld, C., M. Logie and J.W. Polak (2007) The use of Bayesian Belief Networks for the integration heterogeneous transport data sources, paper to be presented at the *11th World Conference on Transport Research*, Berkeley.

Tebaldi, C. and M. West (1998) Bayesian inference on network traffic using link count data, *Journal of the American Statistical Association*, **93** (442) 557-576.

Westlake, A. (2004) Data integration through statistical modelling: The OPUS project, paper presented at the CompStat conference, Prague

Westlake, A. (2005) Combining data and knowledge in models, in R. Khan, R. Banks, R. Cornelius, S. Evans and T. Manners (eds.) *Proceedings of ASC 2005 Maximising Data Value*, ASC, Chesham.

Westlake, A. (2006) Managing metadata for statistical models, paper presented at the *Statistical and Scientific Database Management Workshop*, Vienna.

Westlake, A. and M. Wigan (2006) Integrating information about complex systems: The role of meta-data in the acceptability of results from models, paper presented to *85th Meeting of the Transportation Research Board*, Washington D.C.