International Transport Forum
Global dialogue for better transport

The use of big data in transport modelling

Luis (Pilo) Willumsen
Nommon Solutions and Technologies

01 OLD DATA AND NEW SOURCES

02 ANALYTICS AND QUALITY

03 MODELLING REQUIREMENTS

04 CONCLUSIONS AND RECOMMENDATIONS

OLD DATA AND NEW SOURCES

We use transport models for many applications

Urban Transport Planning

National and Regional Transport Plans

Transport concession studies

Traffic Management Studies

Competition Research

Accessibility and catchment area studies

Preparing for events on specific days

Conventional data collection

- Is not as good as our models
- Is a **mixture of different surveys** and observations from different days
- **Small sample, assumed to represent an "average day of neutral month"**
- **Very Expensive and quality not always good**
- **Facing increasing resistance to answer yet another questionnaire**, omission or simplification of journeys
- Data quality is variable: **errors, sampling bias, etc**.
- **Long elapsed time from specification to practical use**

New sources

Mobile phone data

Door-to-door trips, high sample size, high representativity

Smart cards

Public transport demand, stop-to-stop trips

Apps and Apps aggregators

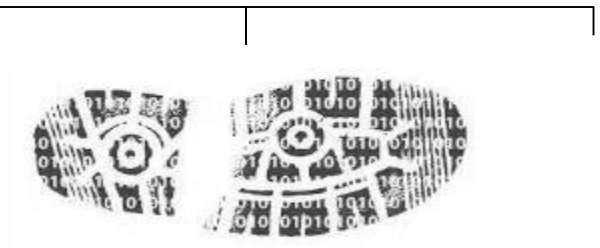
Door-to-door trips, high spacio-temporal resolution, bias

GPS Navigation

Road network speed profiles

ANPR, Bluetooth, WiFi

Speed profiles
Local OD matrices



Digital traces from commonly used devices and technologies

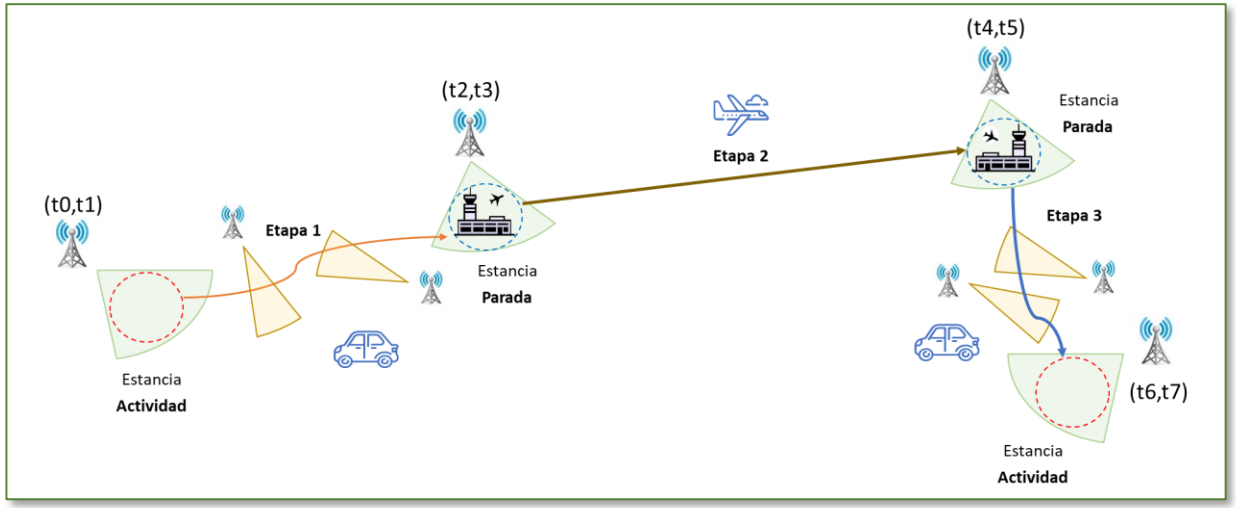
Currently accepted use in the USA, Spain and the United Kingdom

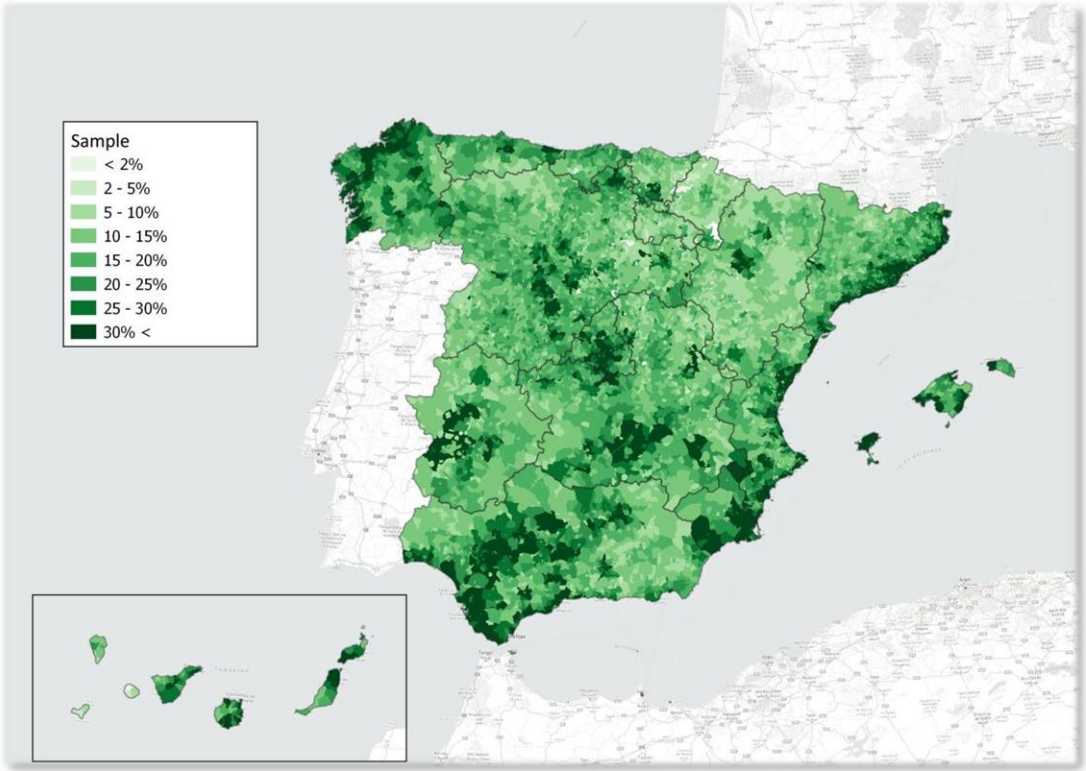
Mobile phone data

Is the only data that offers door-to-door traces

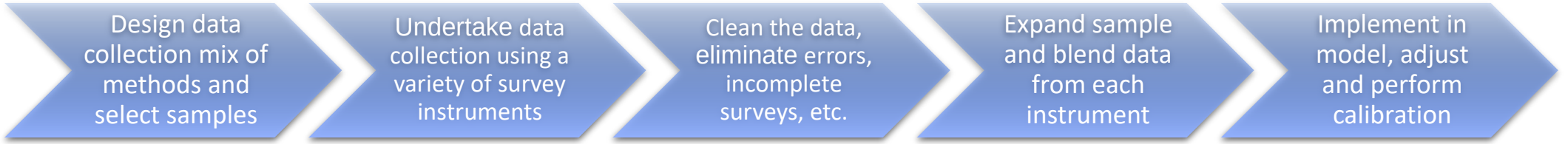
Can be extracted from **apps** that provide Location Based Services, or

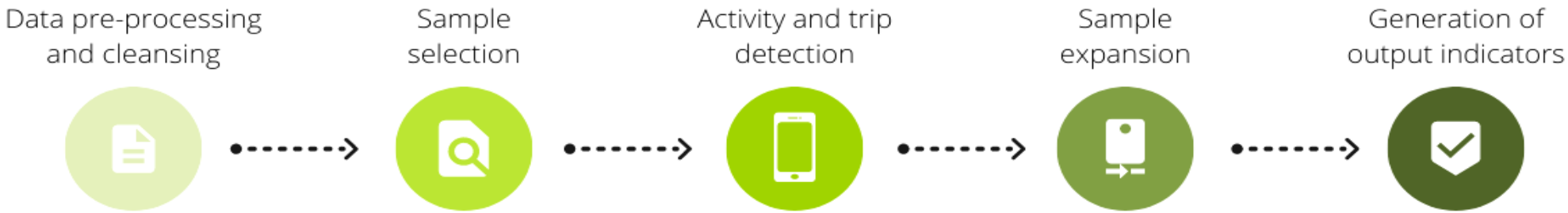
From the interaction between the phone and the **network** (BTS)





Old and New data need processing





Analytics and quality

Mobile phone data (apps and network) have limitations

They do not provide some key data

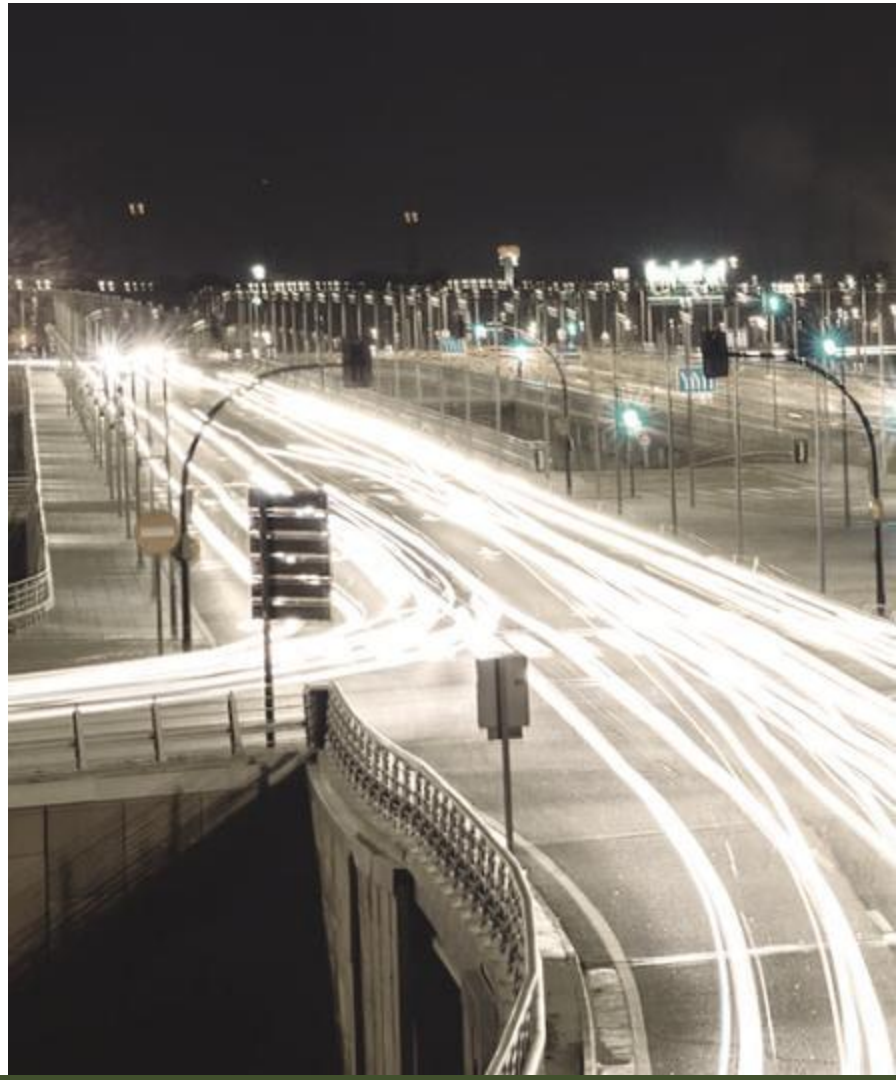
- Vehicle ownership and driving licence holding, for example
- Only a few activities and trip purposes can be identified
- Cannot help with subjective information, like that obtained with Stated Preference and Attitudinal surveys

Insufficient spatial resolution (network) and discontinuities in the data (apps) for some applications

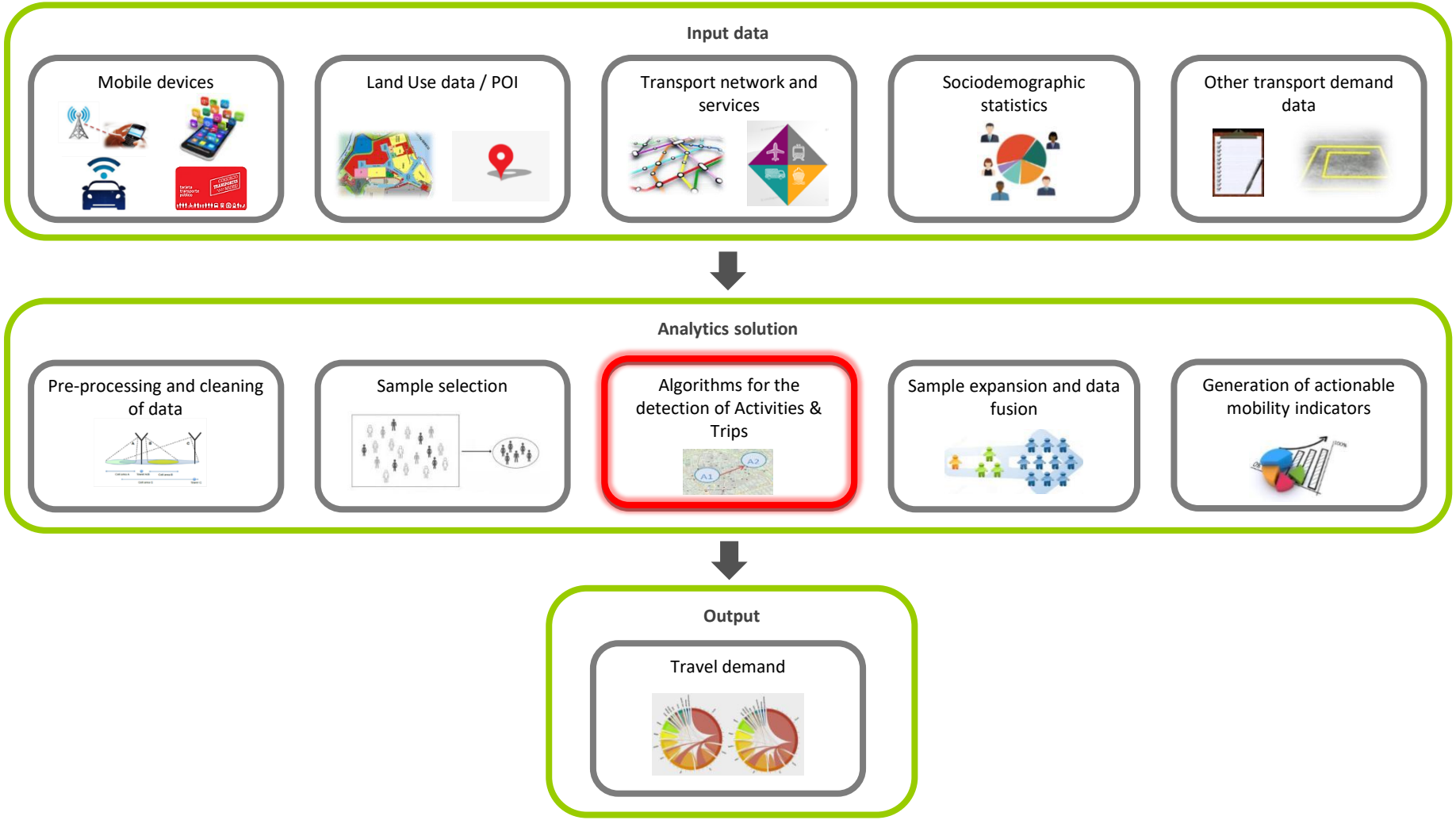
- Difficulties in identifying the mode of transport in urban areas
- Difficult to discriminate car from bus and motorcycle
- Loss of short (intracell) trips

→ Data fusion is needed to overcome these limitations

The algorithms for data cleaning, sample selection, analysis, data fusion and sample expansion are more important than the quality of the raw data



General approach



Data fusion enhances the value of mobile phone data

Essential Context data

- Census data
- Map data
- Land use data
- Point of Interest (POI)

There are no standard data fusion solutions;

Some good complementary sources:

- Traffic counts, some classified
- Public Transport smartcard, ticketing and service data
- Surveys: Household Travel and Intercept Trip data, Supply side surveys
- Tourist data
- Some client (end-user) data

Combining mobile data with map matching

The steps that determine the quality of the results

Validated and actionable outputs

Generation of indicators

Analytics and modelling

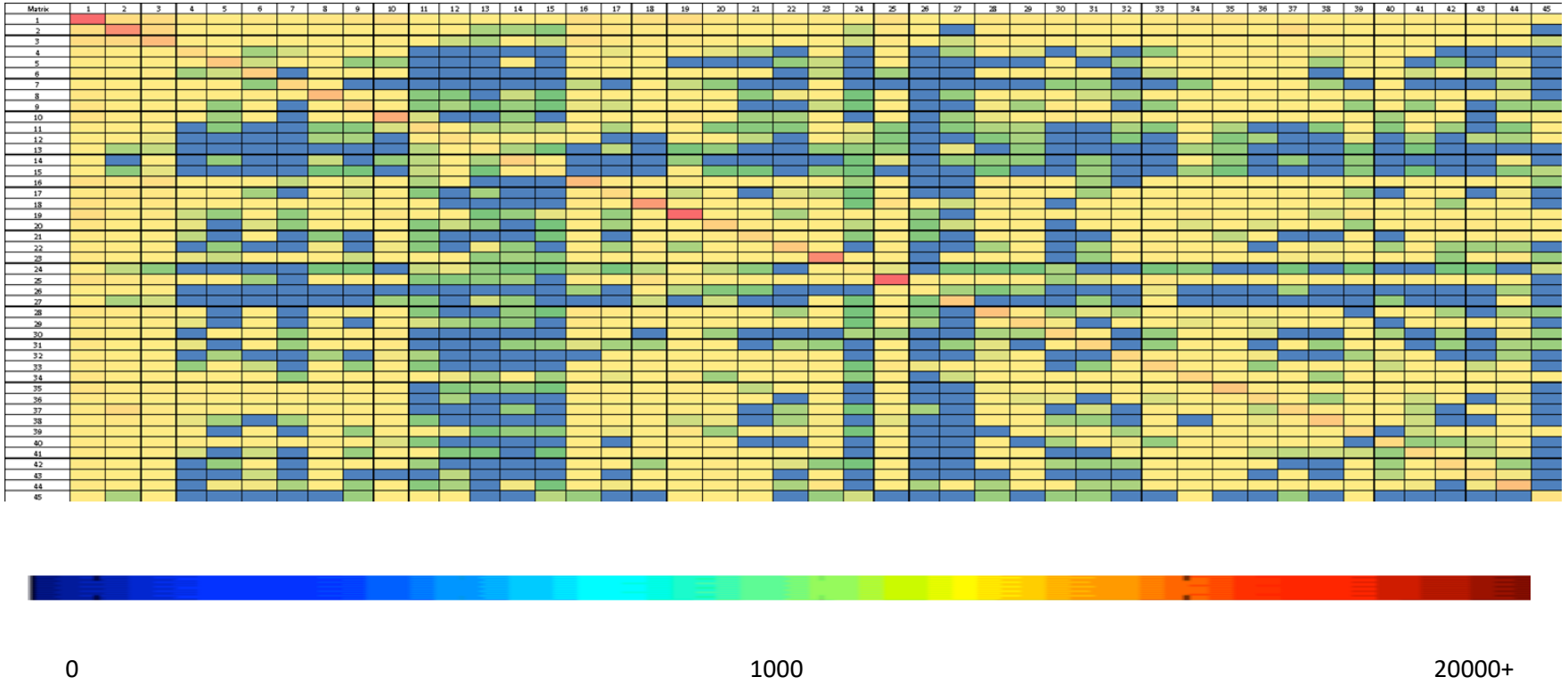
Cleaning and error correction

Storage and data management

Raw data acquisition

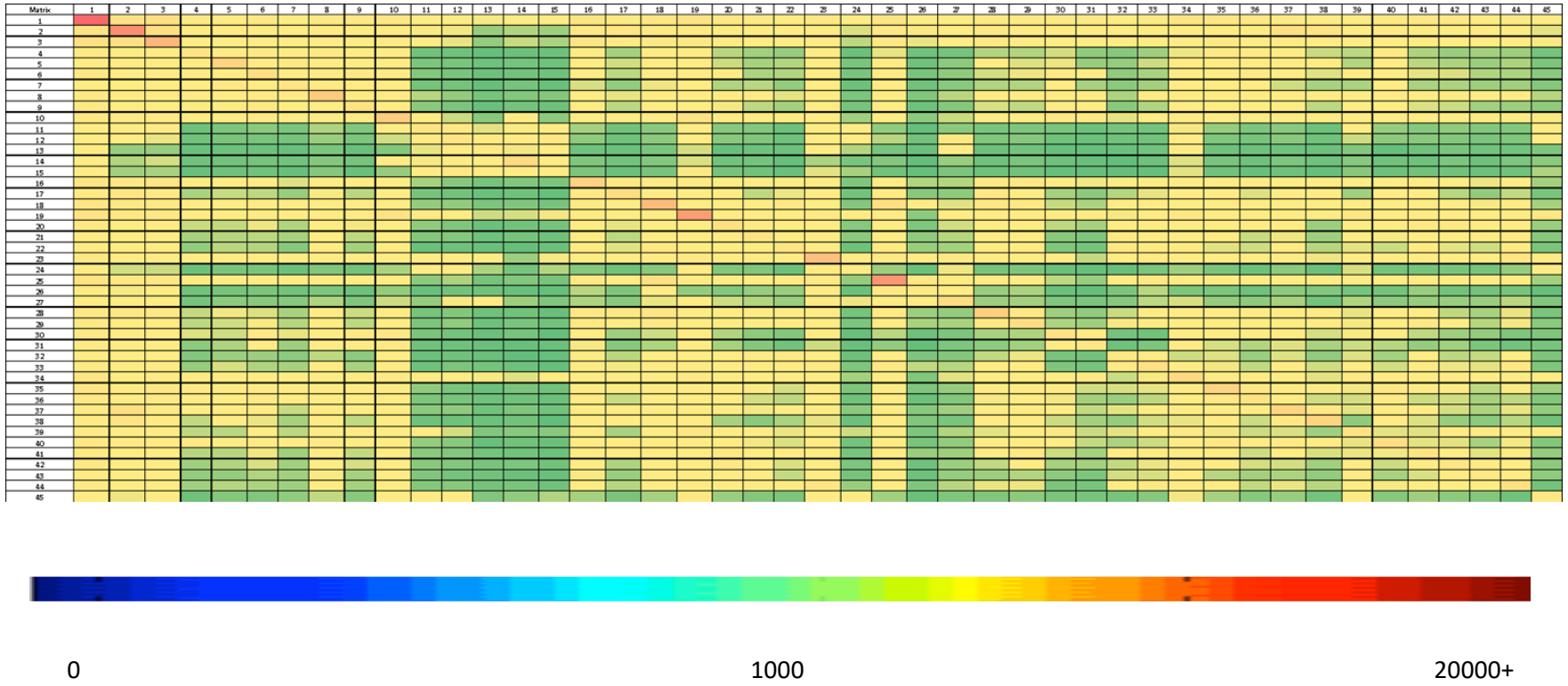
Validation: Comparing trip matrices from conventional and new data

Daily Trip Matrices from Household Travel Surveys for Santiago, Chile aggregated to 45 macro-zones



Validation: Comparing trip matrices from conventional and new data

Daily Trip Matrices from mobile phone network data for Santiago, at the same 45 macro-zones



The challenges to modelling and forecasting

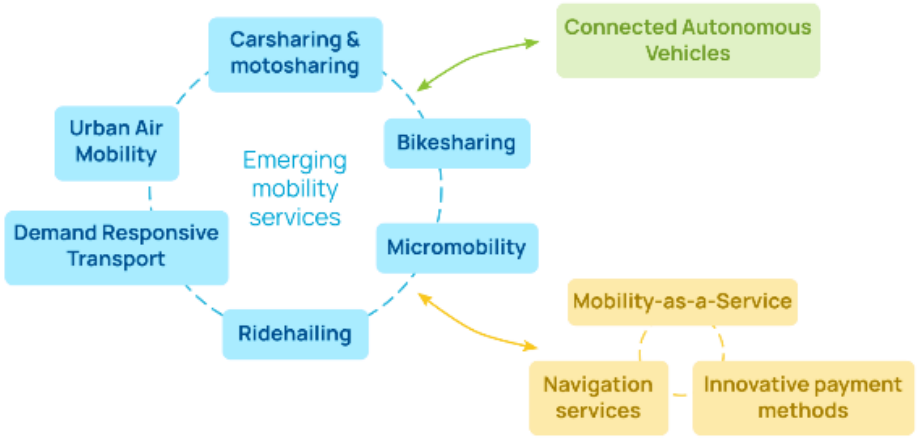
Challenges for travel demand forecasts

Three main challenges for travel forecasts

A. **Equity and Environmental impacts**

B. **New mobility technologies**

C. **Uncertainty**



Mobile phone data can help:

A. Providing a richer identification of impacted communities

B. Supporting the new models required to deal with demand responsive modes

C. Supporting better decision making under uncertainty

Modelling New Mobility Services

Only coarse simplifications of Demand Responsive modes can be achieved with the classic 4 stage approach

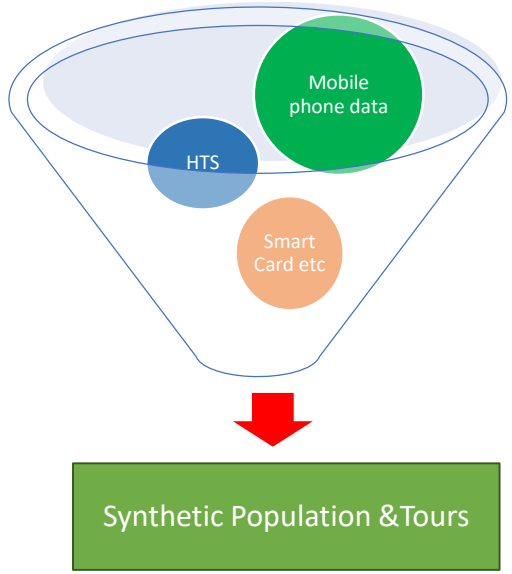
Agent Based Modelling is probably required to represent Demand Responsive modes and the timely allocation of vehicles to users

This requires the generation of a synthetic version of the travelling population and their tours

The ITF work on Shared Mobility (based on Household Travel Surveys) is a good example of this approach

Mobile phone data provides a much larger sample than HTS but a more limited set of identifiable activities

Developing an approach that delivers the optimal combination of the two data sources will be key to successful modelling



Dealing with uncertainty

- We must abandon the illusion of forecasting for a Central Case and finding projects with a good Benefit Cost Ratio
- We need a better understanding of the variability and evolution of demand, recently increased with Teleworking and eCommerce
- Must consider possible Future Scenarios
- Decision making must value the adaptability of projects and policies to changing conditions (Real Options)
- Evolving conditions should be detected early through good monitoring systems using new data sources
- This will facilitate the dynamic adaptation of policies, plans and projects

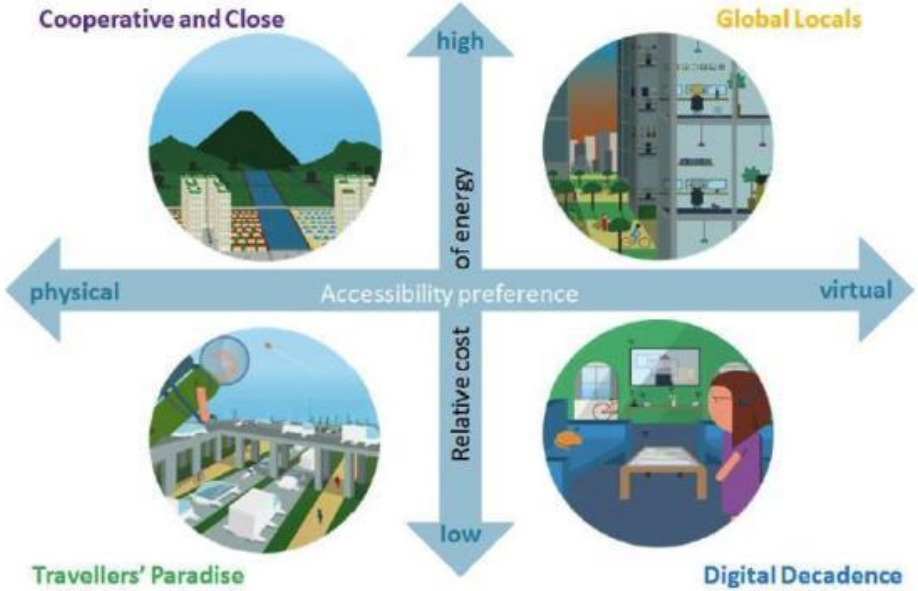


Figure 2: Pictorial summary of the four scenarios to 2042

Conclusions and Recommendations

Big data can contribute to transport modelling

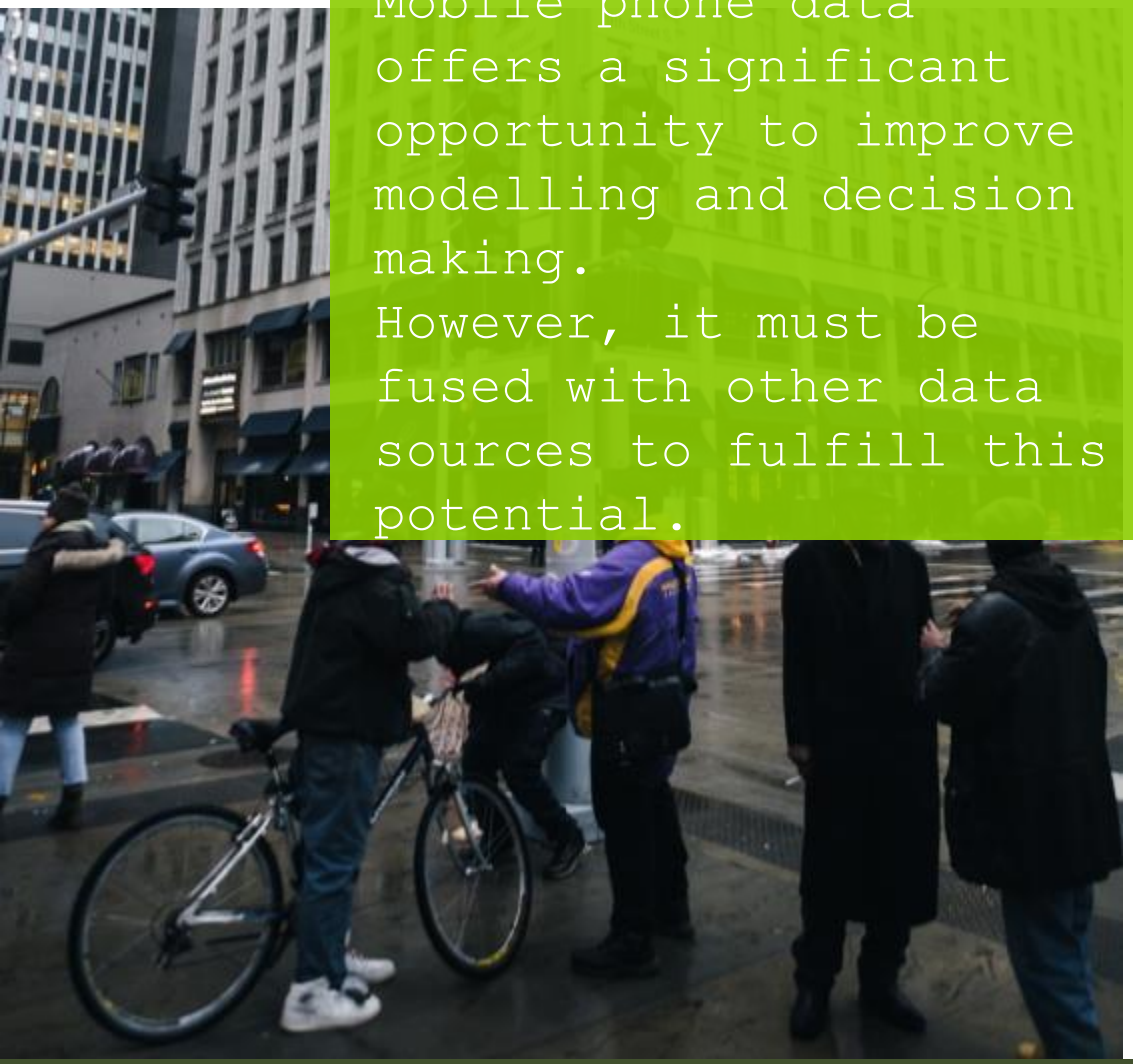
Descriptive analytics: monitoring travel demand

- Baseline diagnosis
- Early trend identification: ‘weak signals’ and transitions
- Ex-post impact assessment

Predictive analytics: supporting decisions

- Strategic planning: long-term demand predictions for the evaluation of infrastructures and transport services
- Estimating traffic, patronage and revenue projections for transport operations and concessions
- Tactical operation: short-term demand predictions for the optimisation of transport supply

Selected recommendations



Mobile phone data offers a significant opportunity to improve modelling and decision making.
However, it must be fused with other data sources to fulfill this potential.

- We need practical guidance on the use of mobile phone data emphasising the role of data fusion and validation of the results
- Research into a new cost-efficient combinations of old and new data sources is needed
- This combination would be critical to generate better synthetic populations and agent based models
- Early detection of changes in trends is a key contribution to flexible decision making in the face of uncertainty
- Mobile phone data can support learning from natural experiments offering new perspectives on mobility, for example the dynamics of recurrent and non-recurrent trips

THANKS

Luis (Pilo) Willumsen
mobility.analytics@nommon.es
www.nommon.es