

# Calling for Validation

## Demonstrating the use of mobile phone data to validate integrated land use transportation models

**Shan JIANG<sup>1</sup>; Laura VIÑA-ARIAS<sup>2</sup>; Christopher ZEGRAS<sup>1,3</sup>; Joseph FERREIRA<sup>1</sup>; Marta GONZALEZ<sup>2</sup>**

<sup>1</sup>Dept. of Urban Studies and Planning  
Massachusetts Institute of Technology

Room 10-485, 77 Massachusetts Avenue, Cambridge, MA 02139, USA  
+1-617-452-2433, shanjiang@mit.edu, czegras@mit.edu, jf@mit.edu

<sup>3</sup>Corresponding author

<sup>2</sup>Dept. of Civil and Environmental Engineering  
Massachusetts Institute of Technology

Room 1-153, 77 Massachusetts Avenue, Cambridge, MA 02139, USA  
+1-617-715-4140, laurav@mit.edu, martag@mit.edu

**Keywords:** integrated models, validation, travel demand, mobile phone data

### Introduction

Paradoxically, understanding the dynamics of the modern metropolis has grown both increasingly possible and increasingly complex in the information age [1, 2]. On the one hand, the interaction of advanced information and communication technologies (ICTs) with “traditional” elements of the metropolitan accessibility system – i.e., land use and transportation – poses new questions, regarding for example how such technologies might substitute for traditional mobility, enhance certain modes, and/or induce fundamentally new behaviors (e.g. see [3, 4]). On the other hand, the digital traces that such technologies produce – such as theoretically trackable mobile devices – hold great promise in generating new data sources to improve model estimation and validation [5, 6, 7]; while increased computational power more generally allows for more complex and sophisticated model development and deployment [8].

In this paper we demonstrate the use of a particular source of ICT data, mobile phones, generated and provided by a mobile phone service provider, to help validate an integrated land use transportation (LUT) model calibrated for the Lisbon, Portugal, metropolitan area (hereafter LMA). Specifically, we use 1 month of anonymous data provided by a private cell phone network operator. These data allowed us to identify, for each phone, all phone activity, localized to the nearest cellular telephone tower (601 towers in LMA). We use the cellular phone towers to generate analysis zones consistent with existing statistical and administrative boundaries (i.e., census blocks and civil parishes) (see Figure 1). We also infer, for each cell phone tower analysis zone, the number of phones that “reside” and/or “work” in that zone, based on the phone activity profile generated over the observed month.

These data are combined with population data from the census, disaggregated employment data estimated from various sources [9] (including points-of-interest data extracted from online Portuguese yellow pages, and an administrative survey data on economic activities in LMA), and other information sources to generate a full OD scaling weight matrix (residents and employment by zone). We then generate an observed cell phone OD matrix for all travel (see Figure 2) and estimate journey-to-work cell phone OD matrix and scale the latter to approximate the population, using the scaling weight matrix, and also estimate traffic assignment to the roadway network based on the same scaled ODs.

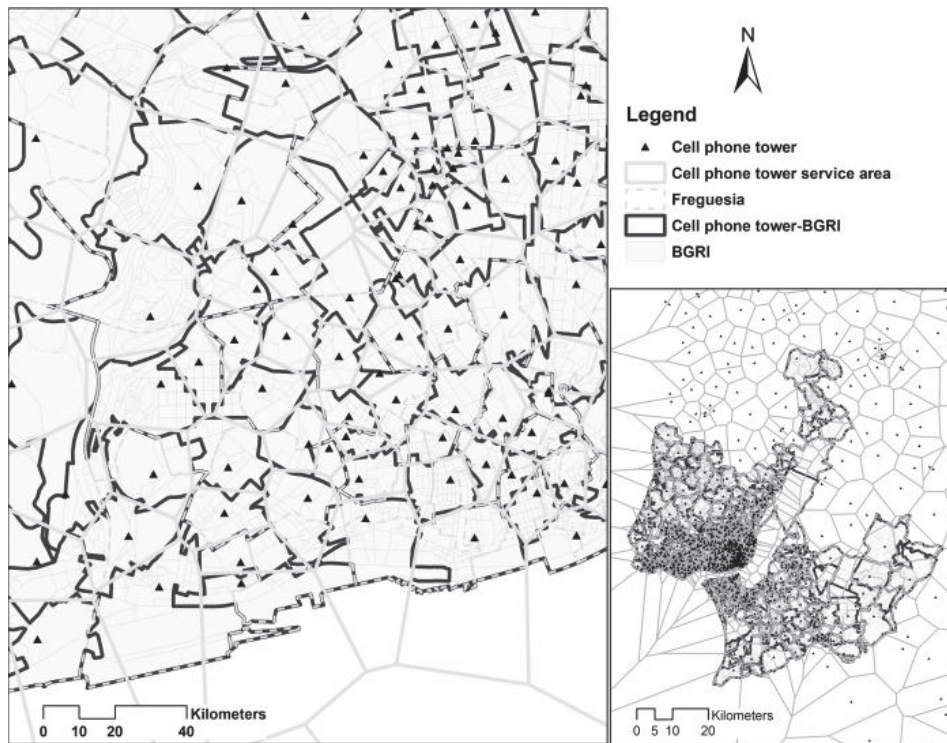


Figure 1: Lisbon's Spatial Analysis Units: Blocks (BGRIs), Cell phone Tower Service Area, Tower-BGRIs, and Parishes (*Freguesia*)

This information is then used to validate travel demand and travel times from two models. The first validation is for a traditional four-step travel demand model developed in TransCAD, which we use in combination with a land development and household/firm location choice model, UrbanSim [10] – an integrated simulation system calibrated for the LMA using a range of data sources from different years. The second is a cell phone travel flow estimation model.

The results show the potential for using mobile phone data to improve complex integrated urban land use and transportation models and demonstrate a viable approach for doing so. In combination with other increasingly available “real-time” data arising from the ad-hoc sensor networks that modern ICT systems provide in our cities, we believe these data sources will fulfill the promise of practical, scalable, and useful integrated LUT modeling for improved urban system management.



Figure 2: Average Daily ODs for all Trip Purposes in LMA as Estimated by Cell Phone Data

### Acknowledgements

This work was made possible by the generous support of the Government of Portugal through the Portuguese Foundation for International Cooperation in Science, Technology and Higher Education and was undertaken as part of the MIT-Portugal Program. We are grateful for the cooperation of the private cellular phone network operator who provided us with anonymous call logs, recorded by the network operator as required by law and for billing purposes, not for the purpose of this research. It was impossible to connect the phone records with any individual. We also gratefully acknowledge the contributions of Pu Wang, Weifeng Li, Yi Zhu, and Jae Seung Lee at MIT, USA; and Filipe Rodrigues and Francisco Câmara Pereira at the Universidade de Coimbra, Portugal.

## References

- [1] **Guhathakurta, S. Eds.** (2003). *Integrated land use and environmental models: a survey of current applications and research*, Springer.
- [2] **Reggiani, A. and Nijkamp, P. Eds.** (2009). *Complexity and Spatial Networks: In Search of Simplicity*, Springer.
- [3] **Ferreira, J. et al.** (2010). "Information Infrastructure for Research Collaboration in Land Use, Transportation, and Environmental Planning." *Transportation Research Record: Journal of the Transportation Research Board* 2183(-1): 85-93.
- [4] **Zegras, C. et al.** (2008). *Data Fusion for Travel Demand Management: State of the Practice and Prospects*. 4th International Symposium on Travel Demand Management. Vienna - Semmering, Austria.
- [5] **Kim, M., Kotz, D., and Kim, S.** (2006). *Extracting a mobility model from real user traces*. IEEE INFOCOM'06, Barcelona, Spain.
- [6] **Gonzalez, M.C., Hidalgo, C.A., and Barabasi, A.-L.** (2008). "Understanding individual human mobility patterns." *Nature* 453(7196): 779-782.
- [7] **Gudmundsson, J., Laube, P., and Wolle, T.** (2008). *Movement Patterns in Spatio-temporal Data*. Encyclopedia of GIS. S. Shekhar and H. Xiong, Springer.
- [8] **Kargupta, H. and J. Han, Eds.** (2009). *Next generation of data mining*, CRC Press.
- [9] **Jiang, S. et al.** (2010). *Towards an Activity-Based Approach for Estimating Travel Destinations*. Proceedings of the 12th World Conference on Transport Research, Lisbon, Portugal.
- [10] **Waddell, P.** (2002). "UrbanSim: Modeling Urban Development for Land Use, Transportation and Environmental Planning." *Journal of the American Planning Association* 68(3): 297-314.