

16 June 2022

Institute of Geodesy and Geoinformatics  
The Faculty of Environmental Engineering and Geodesy  
Wroclaw University of Environmental and Life Sciences  
Wroclaw, Poland

**Report on the doctoral dissertation manuscript submitted by Mr Kamil Smolak on**

***“Assessing the impact of data processing methods on bias in human mobility science”***

**in fulfilment of the requirements of the Degree of Doctor of Philosophy at the The Faculty of Environmental Engineering and Geodesy, Wroclaw University of Environmental and Life Sciences.**

The research resulting in the PhD thesis of Mr Kamil Smolak has been undertaken under the supervision of Prof Witold Rohm and co-supervised by Dr Katarzyna Sifa-Nowicka. The dissertation of Mr Smolak investigates the effects of the Modifiable Areal Unit Problem (MAUP) and Modifiable Temporal Unit Problem (MTUP) – both significant, long-standing analytical problems in GIScience - on the interpretation of spatio-temporal movement data, with a special focus on human mobility prediction tasks.

The domain of large spatio-temporal mobility data analysis is now thriving and rapidly growing. MAUP and MTUP remain, unfortunately, neglected aspects of mobility analysis. This dissertation has thus managed to identify an important knowledge gap. Research addressed in this dissertation can therefore significantly support critical interpretation of trajectory data and their inherent uncertainty.

Smolak presents valuable new methodological advances that will significantly inform the ability to model, predict, and possibly simulate more realistic movement patterns. Mr Smolak does so in a dissertation structured along three main content chapters (Chapters 3.1, 3.2 and 3.3), bookmarked with an Introduction, Investigation chapter, and a Conclusions chapter at the end.

The three main content Chapters are based on two published journal papers (Computers, Environment and Urban Systems, Scientific Reports) and one submitted journal paper (EPJ Data Science). The successfully targeted high-impact journal venues attest to the high ambitions of this

---

**Assoc Prof Martin Tomko, PhD**  
**Department of Infrastructure Engineering**  
**The University of Melbourne**  
**3031 VIC**  
**Australia**

**Phone:**  
**+61 457 779 908**  
**tomkom@unimelb.edu.au**

research. The research is also following recent trends encouraging research reproducibility, by providing access to a computational implementation of the developed methods, by publishing a QGIS plugin (implementing the 3W framework) and a Python library.

Smolak states his main problem as:

*“Processing methods and spatio-temporal resolution of human mobility have a significant impact on observed mobility, creating a bias influencing the retrieved output of analyses.”*

The dissertation addresses this problem through a sequence of three main works, summarised in the dissertation in Chapter 3.

## **Evaluation of the detailed technical contribution of the dissertation**

### **Chapter 1: Introduction**

In this chapter, Smolak introduces the motivation of the problem and grounds his research, and provides a literature review enabling the reader to acquire a terminology to follow the reading of this dissertation (Section 1.1).

Smolak identifies a significant gap in the knowledge in mobility prediction, which requires a data preparation process that significantly influences the outcomes. This is, the temporal and spatial resolution of the aggregated data (aggregation to locations and detection of significant locations) influences the way trajectories are able to capture mobility. Smolak links these problems to the well-known problems of MAUT and MTUP, and shows that depending on data processing, the results from mobility analyses will be inconsistent. Smolak tracks this effect to a set of well known papers, that all have conclusions that suffer from this problem.

The Theoretical Background Section of this dissertation provides a brief summary of the literature in the field, enough to enable the reader to understand the contributions later in the dissertation. This is a useful, brief, and to the point Section of the first Chapter.

More traditionally (from my, biased, experience), the Introduction chapter would only include what Smolak includes in the first part of the Introduction, possibly linking to Figure 2.1 in the Investigation, and explicitly outlining the gaps in knowledge. This would also mean to – explicitly – avoid any references to the author’s own work here, I have also somehow missed a well identifiable place where Smolak would include the main research question/hypothesis of this dissertation. This then occurs much later in the dissertation. That said, I am fine with the structure as presented, it is an individual choice.

### **Chapter 2: Investigations and Chapter 3: Content of Publications**

Chapter 2 presents the overall workflow of the work that constitutes the dissertation (Figure 2.1), and then summarises the main contributions as presented in the papers that constitute the main contribution of this dissertation. Chapter three serves as a brief summary of the three publications that constitute this dissertation. While together the two chapters fulfill their aim and synthesize the research well, I find the content in Chapter 3.1, 3.2 and 3.3 in some way unnecessary – it is already summarised in the Investigations chapter, and then the journal papers themselves are available. This is fine, however, if this is the usual format of the dissertations at this University. This comment has absolutely no bearing on the quality of the dissertation and of the research done.

### **Chapter 3.1:**

This Chapter contains a first methodological contribution of this dissertation, presenting the 3W model, demonstrating how the consideration of the profile of the population (WHO), characterised by the Home-Work groupings, related to the footprint of these groups (WHERE) and the daily rhythm of their activities (WHEN) constitutes a blueprint for mobility analysis and prediction.

The framework is well demonstrated and constitutes a very valuable blueprint for other mobility analyses. The application of the framework has been demonstrated by achieving highly competitive results, reducing the error rate on synthetic data. This is a very worthy, well tested contribution, accompanied by a software release on Zenodo.

### **Chapter 3.2:**

In this manuscript, Smolak focuses on the application of data processing (change in data scales) on movement predictability. He produces a range of valuable results demonstrating how different spatial and temporal aggregation methods (grid vs clustering) lead to very divergent analytical results. This is a very valuable contribution to any analyst who decides how to design their analytical workflow. Again, the conclusions are supported by thorough, and well presented statistical analyses. I really enjoyed reading this paper.

### **Chapter 3.3:**

In this last manuscript presents a method that applies pattern matching on symbolic sequences on predictions. The issue of extended stays in the same location increase predictability, as identified in this work (and in a very recent work of Lici Amichi too - reference below). This manuscript puts into question recent claims about the predictability of movement sequences, and relativizes this statement. Smolak proposes alternative metrics to measure predictability, in particular ISA and ESR. This is a valuable contribution to the field, tested on an interesting set of data. These metrics will now need to be further tested to evaluate their usefulness by other researchers.

Amichi, Licia, Aline Carneiro Viana, Mark Crovella, and Antonio Loureiro. "Impacts of novelty seeking on predictability in human mobility." PhD diss., Inria, 2021.

### **Chapter 4: Conclusion**

This dissertation is meaningfully, and briefly summarised in this last chapter. I recapitulates the findings concisely, and hints at future work. Maybe a longer elaboration on limitations could have been included too.

### **Minor comments:**

1. In the glossary, the entry for MAUP includes an unfortunate typo (aeral -> areal)
2. Chapter 1 – I would suggest to lift the content of Section 1.1.x by a level, or even make it a separate Chapter (Theoretical background, or literature review). There is no Section 1.2, so this makes the content imbalanced, and critically, the main introduction that contains own contributions and research questions, is blended with background.
3. MTUP has (although not under this term) been studied in movement science before, although this dissertation is focused only on symbolic trajectories. Yet, it could be worth to include in references and reflect on the work below:

- o Laube, Patrick, and Ross S. Purves. "How fast is a cow? Cross-scale analysis of movement data." *Transactions in GIS* 15, no. 3 (2011): 401-418.

## Overall conclusion

I really enjoyed reading this dissertation. It is based on three journal publications (one under review) presented to a core audience in the community, complemented by released code to undertake analysis. The dissertation addresses an important problem, and adds to the pool of evidence about how the problem impacts on mobility modelling, analysis and simulation.

This dissertation provides a substantial and creative contribution to literature, with a wealth of new methodological advances. It methodologically contributes computational and analytical advances that are of potential high interest to the mobility analysis community. The work does not present a solution to the core problem – but that is expected, as the problem of MAUP and MTUP is not tractable. The dissertation contribution and the methodological advancements is therefore at the level of understanding of the problem behaviour, and offers tools for analysts to refine their analytical approach.

This dissertation confirms the maturity of Mr Kamil Smolak as a researcher, who is making an original contribution to the field of mobility analysis.

Therefore, without any reserve, I believe that the dissertation of Mr Kamil Smolak fulfills the requirements for a doctoral degree in particular under Article 13 of the Act of March 14, 2003 Ustawa o stopniach naukowych I tytule naukowym oraz o stopniach I tytulew zakresie sztuki (Dz. U. 2003 Nr 65 poz. 595z późn. zm.).

*I, the undersigned, confirm that I do not have any conflict of interest to declare in relation with this dissertation.*



**Assoc Prof Martin Tomko, PhD**  
The University of Melbourne