

Evaluation Report for Dissertation

Title of the thesis:

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

submitted by: **Kamil Smolak**

PhD candidate

Institute of Geodesy and Geoinformatics

The Faculty of Environmental Engineering and Geodesy

Wrocław University of Environmental and Life Sciences

I confirm that this dissertation is suitable for defense at the Wrocław University of Environmental and Life Sciences.

Luca Pappalardo



This PhD thesis is about timely research: understanding the effects of Modifiable Areal Unit Problem (MAUP) and Modifiable Temporal Unit Problem (MTUP) on mobility data, specifically focusing on the next-location prediction task. The candidate addresses this topic by adopting a data science approach, producing important pieces of research within the field of human mobility.

The candidate focuses on three main themes that are connected in many ways: 1) Human mobility modelling, with the design of a model to generate synthetic mobility trajectories; 2) Studying MAUP and MTUP in the important mobility task of next-location prediction; 3) Assessing inconsistencies between predictability of human whereabouts and accuracy in next-location prediction.

The candidate demonstrates familiarity and deep knowledge of the state of the art. Methodologically, the combination of the analysis of mobility data and the application of network and data science tools is adequate. Regarding the content, all parts of the thesis deal with an essential aspect of human mobility science.

The research is of high quality and certainly reaches the level required by a promising PhD candidate. He has shown competence in a variety of techniques, he has acquired the knowledge necessary to carry out interdisciplinary research, and he was able to present the results in three papers (two accepted and one under evaluation). The unveiling of the discrepancies between the predictability limit and prediction accuracy is perhaps the most important result, which is expected to initiate a new wave of activity in measuring the predictability of human displacements. I confirm that this dissertation is suitable for the defense at the Wroclaw University of Environmental and Life Sciences.

The presentation of the thesis is of good quality, too. The manuscript is well-written and clear, demonstrating care dedicated to writing. I found very few typos/mistakes.

A handwritten signature in black ink, appearing to read "Andrzej Rzepka". The signature is fluid and cursive, with a long horizontal stroke at the end.

I have a few remarks, which I list below. They do not question the overall quality of the work, but they offer suggestions for improving the thesis manuscript.

1. Chapter 1, Section 1.1.3, pages 9-10

Only the accuracy metric is introduced to evaluate the goodness of next-location prediction algorithms. However, in the literature, there are other frequently-used metrics (e.g., see the survey paper at <https://dl.acm.org/doi/abs/10.1145/3485125>, Appendix D.2), such as k-accuracy, recall, precision, and AUC. In particular, The k-accuracy (ACC@k) is often used instead of the accuracy in the literature. The candidate should also discuss these other metrics and motivate better the choice of accuracy.

2. Chapter 1, Section 1.1.4.1, page 10

Since mobility modeling is an essential contribution of the candidate, the definition of trajectory generation and human mobility models should be more precise and rigorous. Moreover, the principles and metrics used to evaluate the realism of mobility models should be introduced and discussed in detail. In general, Chapter 1 (background) is crucial, and the concepts presented here should be as detailed as possible.

3. Chapter 2, page 12

From the manuscript: *"Following the Open Science principle and noticing, at the time, the lack of tools dedicated to human mobility data processing ..."*. This is not correctly true. Since 2019, some Python libraries have been dedicated to human mobility analysis (e.g., scikit-mobility, movingpandas, trackintel). Although they provide just the most basic and standard preprocessing methods, they are in stable versions now. The candidate should clarify that HuMobi, an important open science contribution, is not the only existing library dedicated to human mobility analysis. Also, it would be helpful to motivate why a new library was needed, and just the integration of the developed code with an existing one was insufficient.



4. Chapter 2, Section 2.2.1, pages 14-15

The theory of predictability is an essential concept in the thesis. For this reason, Section 2.2.1 should be introduced earlier, somewhere in Chapter 1 under Section 1.1. In general, while essential concepts are well detailed (like predictability), others are barely mentioned (see my previous comments). I suggest using the same degree of detail to introduce all the critical concepts in the thesis.

5. Chapter 3, Section 3.1

The WHO-WHERE-WHEN (3W) model is an important contribution to this thesis. However, this model is pretty similar to existing approaches that also use information about circadian rhythm (e.g., [TimeGeo](#) and [DITRAS](#)), but there is no comparison with them in terms of realism of the synthetic trajectories. The candidate should motivate this and highlight better the difference between 3W and other models using circadian rhythm (or similar tool). Moreover, the evaluation of 3W is made with respect to a few mobility metrics, while others are not considered (such as radius of gyration and k-radius, net capacity, and mobility entropy). This should be better motivated and discussed in the manuscript.

