

## ***Supporting Information:***

# **Approaching the Limit of Predictability in Human Mobility**

Xin Lu<sup>1,2,3,4,\*</sup>, Erik Wetter<sup>2,5</sup>, Nita Bharti<sup>6</sup>, Andy Tatem<sup>7,8</sup>, and Linus Bengtsson<sup>2,3</sup>

1, College of Information System and Management, National University of Defense

Technology, 410073 Changsha, China

2, Flowminder Foundation, 17177 Stockholm, Sweden

3, Department of Public Health Sciences, Karolinska Institutet, 17177 Stockholm, Sweden

4, Department of Sociology, Stockholm University, 17177 Stockholm, Sweden

5, Department of Management and Organization, Stockholm School of Economics, 11383 Stockholm, Sweden

6, Department of Biology, Center for Infectious Disease Dynamics and Huck Institute of Life Sciences, Penn State University, University Park, PA 16801, USA

7, Department of Geography and Environment, University of Southampton, Highfield, Southampton, United Kingdom

8, Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA

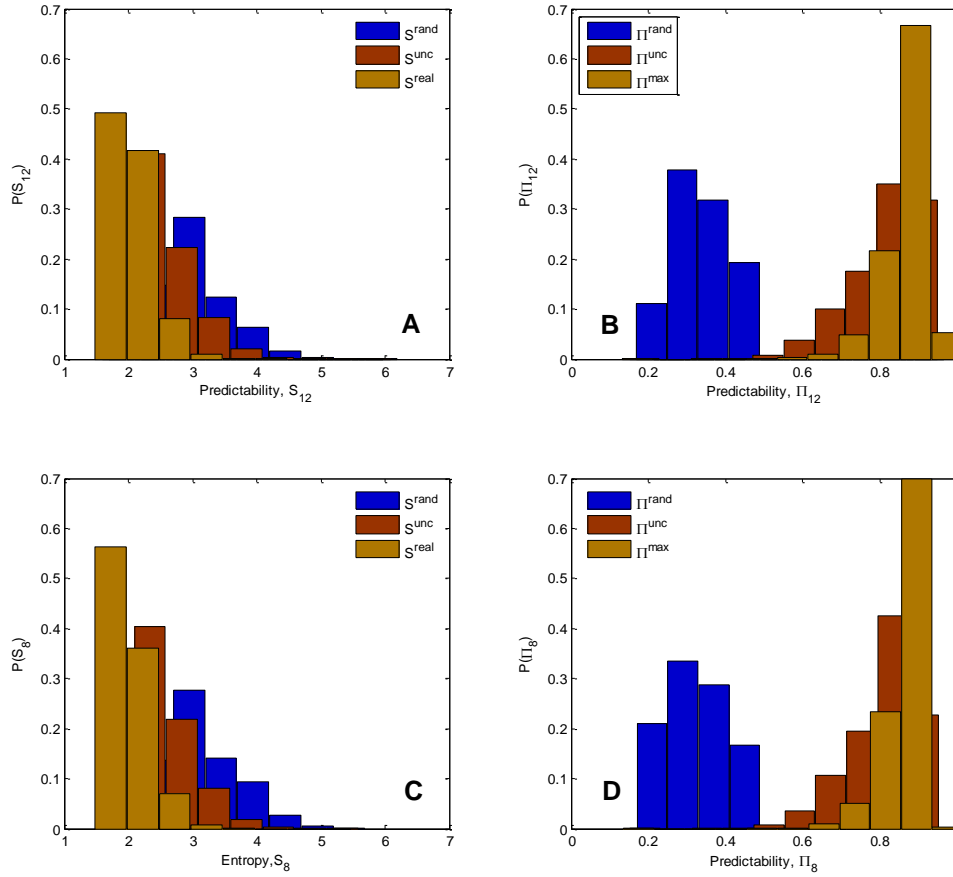
## **S1. Distribution of Entropy and Predictability with Increased Sampling Rates**

To examine the effect of sampling rates on the distributions of entropy and predictability, we have increased the sampling rates of trajectory from once per day to two times ( $S_{12}$  and  $\Pi_{12}$ ) and three times ( $S_8$  and  $\Pi_8$ ) per day. Specifically we construct a trajectory for each individual by his/her observed location every 12 hours (12:00, 0:00) and every 8 hours (8:00, 16:00, 0:00), respectively. Distribution of entropy and predictability for trajectories with more than 80% known locations are shown in **Fig. S1**.

We can see that, despite the different sampling rate, the distributions of  $S$  and  $\Pi$  are very similar to those shown in the main text, indicating that the sampling rate has little effect on our conclusion: that considering both the spatial and temporal correlation can reduce the uncertainties and improve predictability significantly.

---

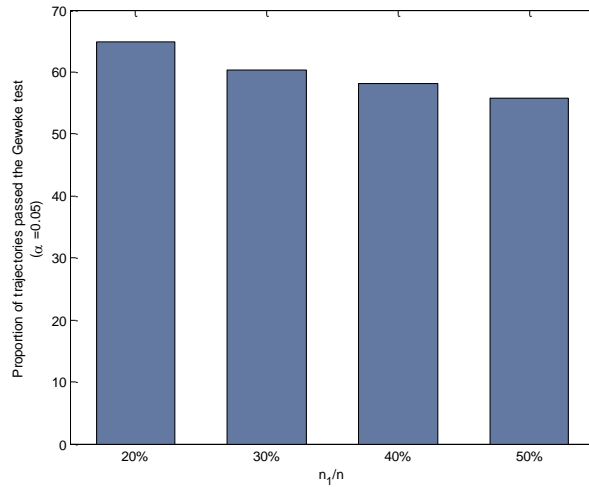
\* Email: xin.lu@flowminder.org



**Fig. S1. Distribution of entropy and predictability for trajectories sampled two times a day (A, B) and three times a day (C, D).** In comparison to the daily trajectory from the main text, which has the median entropy (predictability) of 2.0 (0.25), 0.90 (0.84) and 0.71 (0.89), the median entropy (predictability) with increased sampling rate is 2.32 (0.20), 0.97 (0.84), 0.69 (0.90) for trajectories sampled twice a day, and is 2.59 (0.17), 1.04 (0.83), 0.70 (0.90), respectively.

## S2. Geweke Diagnostic

Based on the Geweke Diagnostic method presented in the main text, the proportion of trajectories that passed the convergence test is presented in **Fig. S2**. We can see that when the length of tested chains  $X_i^1 = \{x_{i,t}^1 : t = 1, \dots, n_1\}$  varies from 20% to 50% of the length of the trajectory, the percentage of trajectories that passed the convergence test decreases from 65% to 56%. By the end, the percentage of trajectories passed all the tests (i.e., for  $n_1/n = \{0.2, 0.3, 0.4, 0.5\}$ ) is 49%.



**Fig. S2.** Proportion of individual trajectories that passed the Geweke test.