A Bayesian approach to modeling GPS errors for comparing forensic evidence

Nika Molan
nm83087@student.uni-lj.si
Faculty of Computer and Information science
University of Ljubljana
Večna pot 113
SI-1000 Ljubljana, Slovenija

Ema Leila Grošelj
eg61487@student.uni-lj.si
Faculty of Computer and Information science
University of Ljubljana
Večna pot 113
SI-1000 Ljubljana, Slovenija

Klemen Vovk
kv4582@student.uni-lj.si
Faculty of Computer and Information science
University of Ljubljana
Večna pot 113
SI-1000 Ljubljana, Slovenija

ABSTRACT

We present a Bayesian approach for evaluating single-point GPS evidence in forensic investigations. We introduce a probabilistic model implemented in Stan that uses Markov chain Monte Carlo sampling to estimate the data-generating processes of GPS measurements from different proposed locations. Our method transforms geographical coordinates to polar coordinates, modeling both distance and directional errors, to compare which proposed location most likely generated the evidence point. We validate our approach using three datasets, including two newly collected sets from Ljubljana and Novo mesto. The results demonstrate the model's effectiveness in distinguishing between proposed locations and quantifying uncertainty through likelihood ratios.

KEYWORDS

device geolocation as evidence, MCMC, digital forensics, likelihood ratio, Stan

1 INTRODUCTION

Our aim is to provide a Bayesian approach for the evaluation of a single GPS localization in light of two proposed locations. Concretely, we want to provide a framework that can answer the following question: *given a single localization point E (latitude, longitude) retrieved from a device D, and two candidate locations P1 and P2 (also latitude, longitude), is it more likely that D generated E from P1 or from P2).* This is a non-exhaustive, contrastive approach. GPS evidence has been proven problematic in court, as it has often been dismissed or not presented at all [1] due to the danger of wrong judgment.

1.1 Motivating problem

The forensic investigation yielded a single localization (evidence point *E* that has been extracted from a seized device D. Due to other evidence from the crime scene, we have been provided with three proposed locations P1, P2 of where the device could have been when *E* was generated (due to GPS error, physical limitations - e.g. *E* is in an unreachable place). Our approach assumes that these three proposals are the only three possible and this has to be stated when it is used to give reasoning under uncertainty.

1.2 Our contributions

The contributions of this paper are summarized as follows:

- A Bayesian statistics approach utilizing Markov chain Monte Carlo sampling to estimate the data generating processes (DGPs) of GPS measurements taken from different locations to compare which DGP most likely generated a single GPS measurement obtained as forensic evidence with uncertainty quantification.
- We release two GPS measurement datasets that we collected in Ljubljana and Novo mesto to aid further research
- We make the full implementation code, along with the data, results, MCMC diagnostics, and visualizations for all three datasets publicly available at our Github repository [1].

.

2 RELATED WORK

In [3] the authors computed a likelihood ratio to compare two candidate locations (P1, P2) in light of the evidence *E* and interpret it with the verbal table shown in Figure 1. They also considered that errors of geolocations may not be equal in all directions (the horizontal error is dependent on the direction) which resulted in high computational complexity due to sample dependence.

Further research into GPS localization errors has been done (see [4] where the authors report a localization error of up to 5 meters in low-cost phones) while others (see [3] have measurements varying up to 100 meters taken from a single point.

3 DATASETS

For all used datasets we provide raw data in the input format for our approach in the form of *datasets/raw/DATASETLABEL/measurements.csv* and *datasets/raw/DATASETLABEL/gt_points.csv*. The scripts used to produce them are also included along with the transformed version of the datasets that were used by the Stan model. Find these transformed values under *datasets/raw/DATASETLABEL/transformed.csv* in the provided Github repository.

3.1 University of Lausanne dataset (UNIL)

We obtain this dataset from the public implementation of [3]. It consists of 699 images that were taken from two proposal points as reference measurements on the University of Lausanne campus. For further details refer to the original paper in [3]. We provide a visualization of the points in Figure 1.

[1] https://github.com/KlemenVovk/gps_evaluation

Figure 1: A geographical visualization of the UNIL dataset from [3]. Evidence point E is the single localization that was recovered as forensic evidence. Then proposal points P1 and P2 are proposed by the authors and the same device is used to take multiple reference measurements at both proposal points. Only deduplicated measurements are shown.

3.2 our Ljubljana dataset (LJ)

Consists of 4 predefined points (evidence E, and three proposal points $P1$, $P2$, and $P3$, each point is specified by latitude and longitude) and a total of 450 images captured while standing on those proposal points, while also noting if the iPhone camera app had permission for precise location for every image. A visualization of the dataset is shown in Figure 2.

3.3 our Novo mesto dataset (NM)

Consists of 4 predefined points (evidence E, and three proposal points $P1$, $P2$, and $P3$, each point is specified by latitude and longitude) and a total of 429 images captured while standing on those proposal points, while also noting if the iPhone camera app had permission to precise location for every image. A visualization of the dataset is shown in Figure 3.

4 METHODS

Unless specified, everything in this section applies to all used datasets (UNIL, LJ, and NM).

4.1 LJ and NM dataset collection

To more clearly understand what affects the accuracy of GPS evidence as well as error patterns, we collect two additional datasets with the same device. The distance between the predefined points was around 100 meters in Ljubljana and 10 meters in Novo mesto. To take images an iPhone 11 Pro was used with the default camera app. We also note granting and denying permission to precise locations to the camera app for each image. To obtain localizations from images, we extract the latitude, longitude, and time of capture from the EXIF data of each image, and note the corresponding

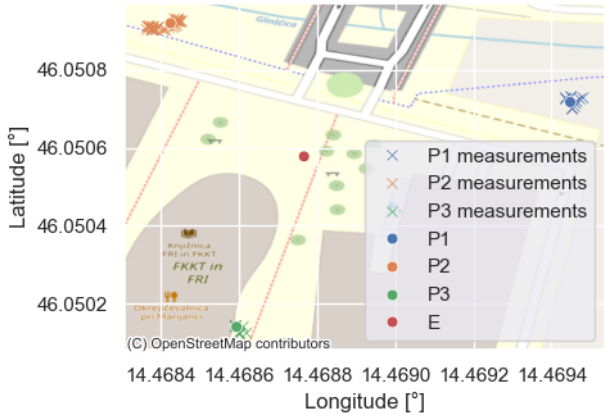

Figure 2: A geographical visualization of our LJ dataset. More specifically, on the campus of the Faculty of Computer and Information Science, University of Ljubljana. Only deduplicated measurements are shown.

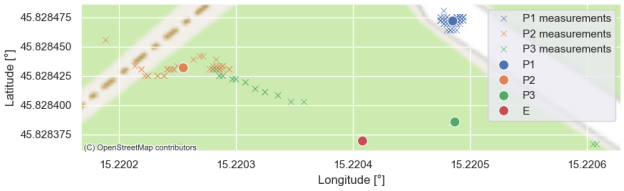

Figure 3: A geographical visualization of our NM dataset. Only deduplicated measurements are shown.

proposal point it was taken from and if precise location permission was granted.

4.2 Dataset preprocessing

Each measurement is defined by time, latitude, longitude, and the label of the proposal it was taken from. We keep only measurements that were retrieved when precise location permission was given to the iPhone camera app when it took the picture (indicated by the *precise* column in the dataset).

We remove consecutive duplicate GPS measurements per proposal by sorting all corresponding measurements ascending by date and time, then removing all consecutive duplicates based on latitude

and longitude. This is done because consecutive duplicates could be due to caching and/or rate-limiting to GPS queries. Consequently, if we try to model distance and angle errors of GPS measurements, some angles/distances will have artificially more probability mass due to duplicates, even though these duplicates are obtained from the same GPS measurement.

4.3 Transformation of measurements from geographical to polar coordinates

To simplify the modeling and interpretability of GPS errors, we convert each measurement from (latitude, and longitude) coordinates to distance (in meters) and angle (azimuth from the north, in radians) from the ground truth point (proposal) it was taken from. To illustrate the concept we show the UNIL dataset converted to polar coordinates in Figure 4. We aim to model the distance and directionality of GPS errors taken from proposal points to estimate from which proposal point we are more likely to observe the retrieved evidence point E GPS coordinates.

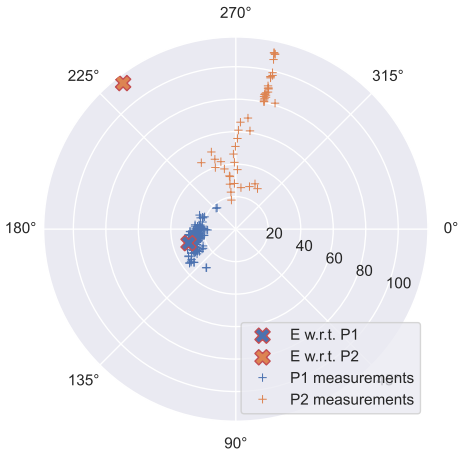

Figure 4: A visualization of the UNIL dataset from [3] after coordinate transformation (distance error in meters and angle - azimuth from north). Note that while we only have one evidence point E, we transform it w.r.t. each proposal point separately. Predefined proposals P1 and P2 are at the center of the polar plot and measurements show the directionality (angle) and the magnitude (in meters) of the GPS errors. Visually, the DGP that generated P1 measurements behaves more similarly to the DGP that generated the evidence point, than the DGP that generated P2 measurements.

4.4 Probabilitic model of GPS errors

To formalize the concept described above, we define a probabilistic model and estimate its parameters with MCMC sampling. We implement the model in the Stan probabilistic programming language. The model file is available in our Github repository as *bivariate_normal.stan*.

Let $M_{ij} = (d_j, \phi_j)$ be the $i-th$ measurement measured from proposal P_j where d_j is the distance in meters from P_j to M_{ij} and ϕ_j the azimuth (in radians) of the line between P_j and M_{ij}. Analogously, let M_{ik} be the $i-th$ reference measurement measured from proposal P_k. We model the set of measurements for each proposal as a bivariate normal distribution:

$$
\begin{aligned}
M_{ij} &\sim \text{MultiNormal}(\mu_j, \Sigma_j) \\
M_{ik} &\sim \text{MultiNormal}(\mu_k, \Sigma_k) \\
\mu_j &= [\mu_{dist_j}, \mu_{angle_j}] \\
\mu_k &= [\mu_{dist_k}, \mu_{angle_k}] \\
\Sigma_j &\in \mathbb{R}^{2x2} \\
\Sigma_k &\in \mathbb{R}^{2x2}
\end{aligned}
$$

where μ_j is a mean vector of distance (in meters) and angle (in radians) w.r.t. proposal P_j. Σ_j is the covariance matrix[2]. Analogously for μ_k and Σ_k w.r.t. proposal P_k. We utilize Stan's default, non-informative priors for all parameters.

To compare under which proposal (P_j or P_k) is the evidence point E more likely, we compute the likelihood of E under each of the models and compute the likelihood ratio. To enhance numerical stability without loss of expressiveness we use logarithms of likelihoods which can later be exponentiated back. We implement this in Stan's *generated quantities* block[3]:

$$
\begin{aligned}
\log L_j &= \log P(E_j | \mu_j, \Sigma_j) \\
\log L_k &= \log P(E_k | \mu_k, \Sigma_k) \\
\log LR &= \log L_j - \log L_k \\
LR &= \exp(\log LR)
\end{aligned}
$$

where E_j and E_k denote the evidence point E trasformed w.r.t. proposals P_j and P_k respectively.

To assess if the estimated bivariate normal models fit our data (reference measurements) we perform a posterior predictive check by randomly sampling points from our estimated bivariate normal models to create replicate datasets (this is also done in the *generated quantities* block in Stan). The idea is that if an estimated model fits input data well, we should be able to generate *similar*, synthetic data by randomly sampling from it. In other words, if the estimated model managed to capture the behavior of distance and angle errors in our reference measurements, it should be able to generate new, synthetic, measurements that resemble the same distance and angle errors. To visualize this, we overlay the generated synthetic data over the reference measurements (input data).

5 RESULTS

Due to the length limit of the paper we only show the full results for the UNIL dataset. However, all three datasets with all results, visualizations, and MCMC diagnostics are available in our public Github repository in three separate Jupyter notebooks corresponding to the three datasets.

We perform MCMC sampling with 4 chains of 4000 samples each, standard MCMC diagnostics (trace plots, effective sample

[2]We use the Cholesky parametrization of the Multivariate normal, which is natively implemented in Stan, so $\Sigma_j = L_j L_j'$ for efficiency and numerical stability during MCMC sampling, but omit it here for brevity

[3]Everything in the generated quantities block can be computed outside of Stan (e.g. in Python) as it is performed on the posterior draws after the MCMC sampling is done, we do it in Stan for clarity.

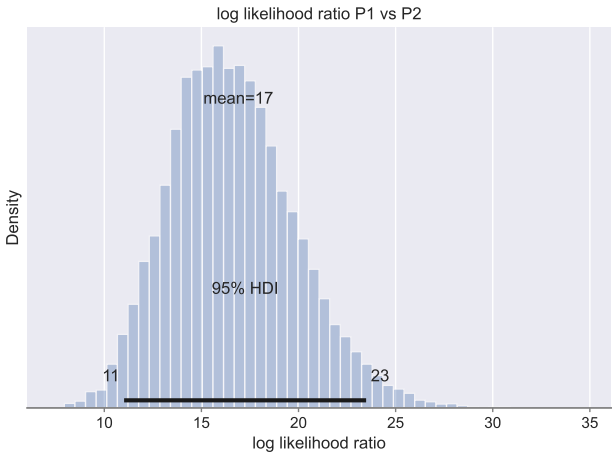

Figure 6: The posterior distribution of log-likelihood ratios for the UNIL dataset along with 95% high-density intervals to quantify uncertainty. These can be exponentiated back to regular likelihood ratios which are currently the standard used in courts as per [2] and [3] to compare forensic evidence.

Range of LR	Verbal Equivalent
1-3	In my opinion the observations are no more probable if [P1] rather than[P2] were true. Therefore, the observations do not assist in addressing which of the two propositions is true.
4-10	In my opinion the observations are slightly more probable if [P1] rather than [P2] were true.
10-100	In my opinion the observations are more probable if [P1] rather than [P2] were true.
100-1000	In my opinion the observations are much more probable if [P1] rather than [P2] were true.

Table 1: Table of LR verbal equivalent from [2] page 39.

sizes, R-hat values) do not indicate any issues in convergence. To further illustrate this we visualize a posterior predictive check of the UNIL dataset by overlaying a random replicate dataset over the real measurements in Figure 5.

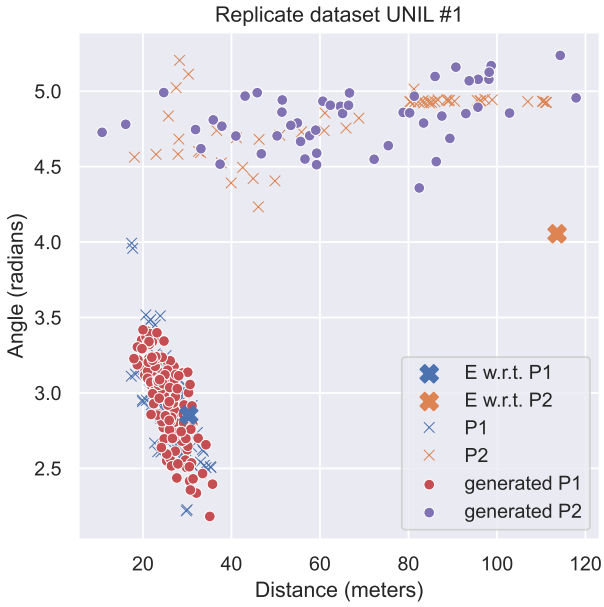

Figure 5: A posterior predictive check for the UNIL dataset. We observe that the synthetically generated measurements closely resemble the input measurements, indicating that the estimated bivariate normal models fit our data well.

Figure 6 depicts the posterior distribution of log-likelihood ratios for the UNIL dataset. Since these are logarithms, we can exponentiate them back to likelihood ratios to be able to provide the *court verbal equivalent* (see Table 1 as defined standard in [2].

6 DISCUSSION

This paper presents a novel Bayesian approach for modeling GPS errors in forensic evidence evaluation. By utilizing MCMC sampling to estimate data-generating processes of GPS measurements, we provide a statistical framework for comparing single-point device locations. Since our method results in posterior draws, it enables direct uncertainty quantification.

REFERENCES

[1] E. Casey, D.-O. Jaquet-Chiffelle, H. Spichiger, E. Ryser, and T. Souvignet. Structuring the evaluation of location-related mobile device evidence. *Forensic Science International: Digital Investigation*, 32:300928, 2020.
[2] F. S. Regulator. Development of evaluative opinions. Technical Report FSR-C-118, UK Forensic Science Regulator, Birmingham, 2021.
[3] H. Spichiger. A likelihood ratio approach for the evaluation of single point device locations. *Forensic Science International: Digital Investigation*, 2023.
[4] L. Wang, Z. Li, N. Wang, and Z. Wang. Real-time gnss precise point positioning for low-cost smart devices. *GPS Solutions*, 25:1–13, 2021.