Lab 1 - GPS Error Analysis

Raj Harshit Srirangam

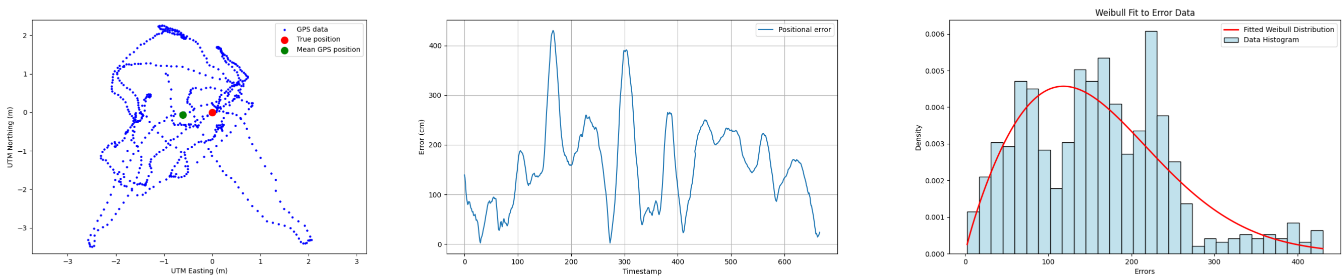
1 GPS Resolution

The GPS receiver used, a BU-353, outputs data in the GPGGA format with latitude and longitude reported in degrees and minutes, with a precision of 0.01 minutes. For latitude, this equates to a resolution of approximately 18.52m (since 1 arc minute equals 1852m). For longitude, the resolution varies by latitude and is calculated by multiplying 18.52m by the cosine of the latitude. At a latitude of 42.338°, the resolution is approximately 13.69m. Thus, the overall resolution is roughly 18.52m by 13.69m. Additionally, the sampling rate for the measurements is observed to be 1Hz. This level of precision sets a baseline for understanding the potential positional accuracy and error in the data collected.

2 Data Collection Methodology

The GPS data was collected under two conditions: stationary and walking. The main goal in data collection was to ensure the true GPS position was known for accurate error calculation. Recognizable location markers were selected, including a single point for the stationary data and the starting/ending points for the walking data. These markers were chosen because they could be easily identified from satellite imagery, allowing precise determination of the true locations. The stationary data was collected in the library quad, with the true GPS position selected as the intersection point of different sections of the mosaic paving. The walking data was collected on Columbus Avenue with the corner of a blister paving near EXP as the starting point and a light pole near Carter Playground as the ending point. Additionally, a compass was used to maintain a heading of 60°northeast, ensuring a straight-line path. The coordinates of the true stationary point are 42°20'20.0"N, 71°05'17.9"W. For the walking data, the true starting point is located at 42°20'11.6"N, 71°05'16.0"W, and the true ending point is at 42°20'20.6"N, 71°05'01.7"W. The true walking path followed can be approximated to be the straight line connecting the starting and ending points which comes out to a distance of 429m travelled over 7 minutes and 20 seconds. These location markers serve as the baseline for the error analysis in the following sections. The data was recorded using rosbag2 and converted into json files and then plotted using pyplot. The weather conditions during the data collection period were overcast/slightly windy. The altitude of all the location markers is approximately 3.3 m. Two sets of data was collected for each case and the best one was selected.

3 Analysis of Stationary Data



(a) Scatter Plot (b) Error over time (c) Histogram and fitted weibull

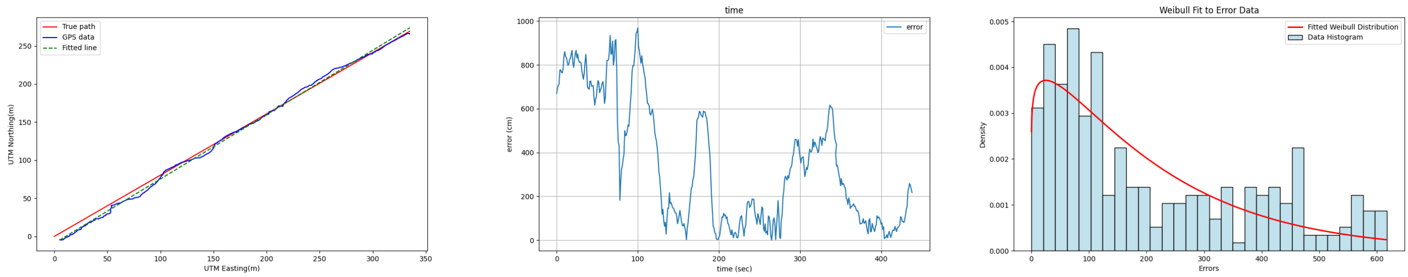
Figure 1: Plots for stationary data

The stationary dataset contains 669 samples. Figure 1a shows the UTM easting vs UTM northing as a scatter plot. Note that the data plotted is relative to the true GPS position [327968m, 4689516m] in zone 19T. The mean position of the GPS readings comes out to [-0.612m,0.065m] which is off by about 61.6cm. The sample mean is 10 times more accurate in the northing compared to the easting. The dispersion of the data can be inferred through the standard deviation which is 1.243m. Interestingly, looking at the standard deviations across the X and Y directions, which are 1.005m and 1.388m respectively, it can be observed that though the mean of the northing is extremely close to the true northing, it is more dispersed than the easting which is over 60cm off. Observing the scatter plot, it can be seen that the GPS keeps trying to correct the error in it values and keeps overadjusting the value going past the true position.

The error of a sample in this case is calculated by measuring the euclidean distance from the sample to the true position. The positional error is plotted against time in Figure 1b. The range of errors was from 2.3cm(negligible since accuracy of the true position should be around 10cm) to 430.2cm. The mean and standard deviation of the error are 159.5cm and 88cm respectively. However, a more representative measure of the GPS's accuracy is the Root Mean Square Error (RMSE), which in this case is 182.2 cm.

Since the sample size is relatively small, it is not advised to generalize these statistics directly as accurate error estimates. The errors can instead be fit into a standard probability distribution. Multiple distributions were tested including Gaussian, log-normal, skew-normal and gamma but the closest fit turned out to be a Weibull distribution. After fitting, the parameters are Shape: 1.856, Location: 0, Scale: 179.22. Figure 1c shows the histogram and the fitted weibull distribution of the errors. To verify if this is a good fit, the goodness of fit was tested using the Kolmogorov-Smirnov Test which gave a p-value of 0.0208 which though below than threshold of 0.05 is a sufficiently good fit. A good estimate of the error would be the RMSE of the distribution which is 185.529cm. The error bound is the 95th percentile, which is 323.64 cm. This indicates that 95% of observed errors fall within this range, effectively filtering out extreme outliers.

4 Analysis of Walking Data



(a) Position Plot (b) Error over time (c) Histogram and fitted weibull

Figure 2: Plots for walking data

The walking dataset contains 441 samples. Figure 2a shows the UTM easting vs UTM northing and the true path followed along with the line that best fits the data. Note that the data plotted is relative to the true starting point [328005.03m, 4689256.03] in zone 19T. To get a general sense of the overall accuracy of the data, lets compare the angle of inclination of the true line to the line that best fits the data. The true line has a angle of inclination of 38.8064°and the fitted line has an angle of 40.187°along with an y-intercept of 9.06m. This means that the data is off by 1.381°(along with a little offset). From the plot, it can be noticed that similar to the stationary case, most of the error is in the UTM easting compared to the UTM northing. It is also clear that most of the error is in the initial half, possibly due to the GPS taking that time to correct the position and heading.

In this case, let us measure the positional error as the perpendicular distance of each sample point from the true line. Although this approach may yield less precise errors compared to stationary data, it is preferable than distance from line of best fit for this analysis. Figure 2bshows the error plot over time. The largest error seen is at 968.69cm and the mean and standard deviation are 345.842cm and 283.1576cm. The standard deviation being lower than the mean signifies the data is pretty consistent except for the outliers. The RMSE though, is 446.97cm which is concerning. For the sake of a fair comparison with the stationary data, let us consider just the consistent part of the data i.e the latter half. For this part, the mean, SD and RMSE are 209.2915cm, 175.9968cm and 273.4555cm respectively.

For distribution fitting, only the latter half is considered due to its predictability. An argument can be made that the initial error before the position is properly corrected would have been the same regardless of how much data was collected. So, ignoring that as an outlier would give a better model for the rest of the data regardless of how big it is and is thus more generalized. The closest fit is a Weibull distribution as shown in Figure 2c with parameters - Shape: 1.1025, Location: 0, Scale: 216.7911. Testing the goodness of fit using the KS test gives a p-value of 0.042 which is almost at the threshold. For reference, fitting the whole dataset into a distribution and using the KS test gives a p-value of 0.0045 showing the predictability of the half dataset compared to the whole. The RMSE of the distribution is 367.79cm and the 95th percentile is 461.94cm, both significantly higher in error than the stationary case.

5 Discussion

The error analysis in both the cases tells us that the GPS performs within acceptable error margins excluding the few outliers with the stationary reading being generally more accurate. The stationary data kept overadjusting and oscillating around the true position whereas the walking data was initially off by around 9m and reduced the error over a time of 160 seconds. The heading was fairly stable and accurate but the error adjustment was much slower compared to the stationary data. This is most likely because the GPS uses a Kalman filter (or something similar) to predict the next position based on your movement. These filters smooth the data over time, but they rely on a model of your motion, which means the adjustments to errors happen gradually as more data is collected resulting in more consistent data. When stationary, the GPS has no movement data to help stabilize the position, so small fluctuations in signal or noise cause more noticeable jumps in position estimates. Additionally, the walking data had extreme outliers possibly due to deviations of the true path from the ideal straight line which was not accounted for and since these deviations took over 30 seconds to adjust due to the slow error smoothing, the error estimates were magnified. Another reason for the outliers could be the multipath errors caused from the signals bouncing off the buildings or atmospheric conditions. Based on all of this it can be concluded that GPS navigation by itself is not very reliable to be used in real time navigation without additional correction techniques.