On the Feasibility of Localising Smart Devices using Air Pressure

Shubham Aggarwal
Department of Mathematics and Statistics,
Maynooth University,
Kildare, Ireland.

David Malone
Hamilton Institute / Department of Mathematics and Statistics,
Maynooth University,
Kildare, Ireland.

2021 32nd Irish Signals and Systems Conference (ISSC) | 978-1-6654-3429-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/ISSC52156.2021.9467882

***Abstract*—Many devices, such as phones and fitness trackers, contain barometers to measure air pressure, allowing the tracking of air pressure over time. While such changes can be used to identify changes in altitude, if a device is stationary, changes could also be used to identity a person's location in conjunction with meteorological information, which potentially makes pressure data sensitive information. We will use the meteorological data to see how effective device localisation can be when matching air pressure profiles to different locations in Ireland.**

***Index Terms*—localisation; smart devices; air pressure; barometers**

I. Introduction

A growing number of smart devices are used by people today. In this paper, we are going to consider smart devices with three common data sources: a barometer, an accelerometer and a clock. Devices such as smart phones, smart watches and other fitness trackers commonly have all three of these. An accelerometer is used for activity monitoring and gesture identification, a barometer used to identify changes in altitude (to count, say, stairs climbed), and a clock is necessary for many purposes, such as data recording. While many smart devices also include GPS, we are going to study the feasibility of approximately localising a device using a time series of pressure measurements, without the use of GPS.

Suppose a smart device with an accelerometer can determine when it is stationary. While stationary, the clock and barometer can be used to form a time series of air pressure readings. In this paper, we are interested in how much information these air pressure readings reveals about the device's location. Of course, to link a time series of air pressure measurements to a location, one must have records of air pressure values for different locations. Fortunately, air pressure at different locations is of considerable interest to meteorologists, and records of such values are available.

In Section II we review related work. In Section III we describe our source of meteorological data and the performance of the barometers in smart devices. In Section IV we describe two methods for combining these data sources to estimate a device's location. The results of using these methods is described in Section V, and we conclude in Section VI.

II. Related Work

Barometers have been used to track a variety of human activity including augmenting GPS accuracy with altitude information, determining changes of floor within buildings and even detecting sudden altitude changes associated with falls [1]. While we have not seen information on localising humans with barometer measurements, people have proposed using pressure measurements to localise vehicles moving on roads, where the elevation changes of the roads have been mapped in advance [2], [3].

Smart devices have accidently leaked information about location. For example, the Strava exercise tracking app provided a heat-map of locations where people engaged in exercise. As the app was used by a significant number of US military personnel, it accidently revealed the location of US military bases [4], [5].

The idea of deriving sensitive information from apparently innocuous data has been of interest in recent years. In 1997, Latanya Sweeney showed that is was possible to re-identify publicly accessible US medical records using voter records [6]. Data from the Netflix competition to rank films was partially de-anonymised by combining it with public data from the Internet Movie Database [7]. Sometimes re-identification is possible even without an additional data source [8]. We consider the use of air pressure information to identify location to be in a similar vein.

III. Data Sources

A. Met Éireann MÉRA data

For a source of meteorological data, we use the Met Éireann Re-Analysis (MÉRA) data set [9]. This data is produced by combining historical observations and a physical weather model to generate a consistent reconstruction of past weather. The reanalysis period extends from 1981 to 2015, and provides estimates of many meteorological quantities on a grid spaced by approximately 2.5km. In a sense, the reanalysis provides a best reconstruction of actual weather conditions, given the observations available.

Data is available in GRIB (GRIdded Binary) format, and we requested surface pressure values for the month of June 2015. The file contains time values every three hours, with three intermediate steps, effectively providing hourly data.

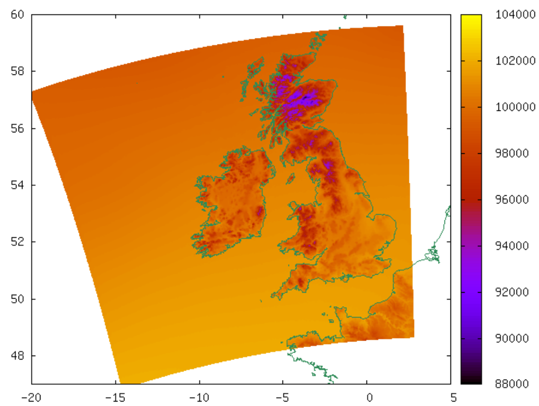

Fig. 1. MÉRA air pressure data (major land boundaries added).

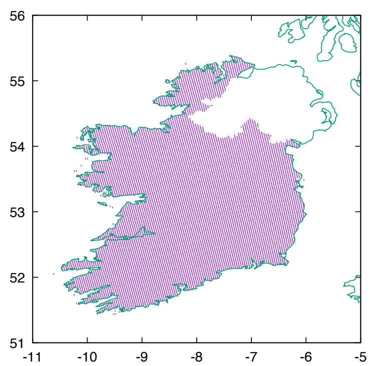

Fig. 2. Filtered MÉRA locations (major land boundaries added).

Each hour contains 258,681 data points at varying latitudes and longitudes. Data was extracted using ecCodes GRIB tools[1].

The data for the first hour is plotted in Figure 1. We can see that the area covered by the MÉRA data covers much of the UK, Ireland and even the Atlantic and continental Europe. Mountainous areas are clearly visible as areas of low pressure in the data. We confirmed that the data points are separated by approximately 2.5km using the haversine distance. We also checked the resolution of the pressure values; data values are separated by multiples of 0.25Pa.

As we are interested in localising stationary smart devices, we can restrict our interest to a subset of these locations. In practice, one might use other information to give a broad indication of the devices' location to filter these points. For example, though IP geolocation has accuracy challenges [10], [11], identifying the country that a device is in is often practical.

As an example, we use the 2011 Census Boundary data to filter locations within the provinces of the Republic of Ireland[2]. This reduced the number of locations to 11,205.

The resulting locations are shown in Figure 2. At this scale, we can see the individual points are visible, including points on islands. There are small discrepancies between the major land boundaries shown and census boundaries. This is because the land boundaries are lower resolution and also the census boundaries include some inland waterways.

[1] Available from https://confluence.ecmwf.int/display/ECC/ecCodes+Home.
[2] Available from https://data.gov.ie/dataset/census-2011-boundary-files.

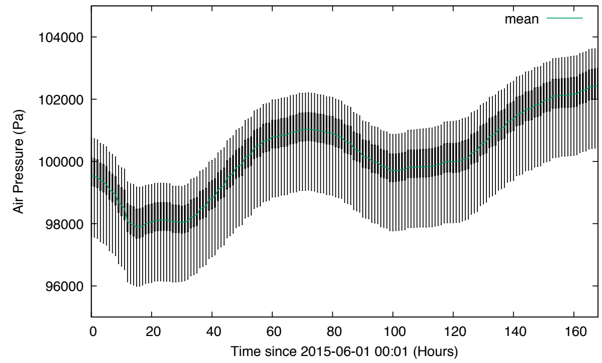

Fig. 3. Distribution of air pressure during the first week of June 2015. Boxes show IQR, whiskers show 5^{th} and 95^{th} percentiles.

Finally, any success in localising a device depends on the variability of air pressure values over both space and time. To give a preliminary indication of the variability, we show a box and whiskers plot of the hourly air pressure over the first week of June 2015 in Figure 3. Each box indicates the interquartile range, with the whiskers showing 5^{th} and 95^{th} percentiles over the filtered locations. The mean is also shown.

We see that the interquartile range for this week is typically around 1,000Pa. At resolutions between 0.25–1Pa, a single hour could potentially provide 1,000-4,000 distinct values to help localise the device at an area with pressure within the interquartile range. This suggests that combining multiple hours of data will be necessary to localise devices, as we have over 10,000 locations to distinguish. We also note that values shown move in a correlated way over time, so we should not expect changes to air pressure to be independent.

B. *Barometers in Smart Devices*

Given that one of the common applications of these devices is to measure a change of floor in a building, we expect that these devices will all be able to comfortably measure a change of around 2m. Near sea level, assuming hydrostasis, the change in air pressure per meter is approximately ρg, where ρ is the density of air and g is the acceleration due to gravity. This gives a value of about 12Pa/m.

The barometers in smart devices are of surprisingly high quality. For example, the iPhone 6 and 7 are rumoured to use a custom Bosch BMP280 sensor. The data sheet for this sensor reports that it can measure pressure ranges from 3,000–11,000Pa, with a relative accuracy of 12Pa and an absolute accuracy of 100Pa [12], however, this seems quite pessimistic compared with some reports [1]. The RMS noise is given as 1.3Pa and the resolution 0.16Pa in the data sheet.

More recent sensors are likely to have even better capabilities, with devices such as drones making use of barometers to maintain constant altitude, to roughly 0.1m. For example, the

BMP085 is available to hobbyist at a price of a few euro and reports an absolute accuracy of better than 3Pa [13].

The Apple and Fitbit barometer APIs[3] report pressure in units of Pascals, suggesting a resolution of at least 1Pa. In fact, both APIs seem to report results as floating point numbers, so better resolution is possible. The Fitbit API also indicates that the barometer can be read at rates from 1Hz to 40Hz, so averaging to reduce noise is certainly practical.

Testing an iPhone 6s and Xs with a barometer application, we found that it could report the pressure to a resolution 1Pa and that measuring at height differences of 1m reported a change 11–13Pa. The expected change is around 12Pa, which suggests the device is capable of measuring to the nearest Pascal. Quick tests indicated that the Apple Watch had comparable performance and the Fitbit Vista reports a change in altitude at the resolution of 1ft.

As the meteorological data we have is historical, we choose to simulate our smart device barometer data. We will do this by taking choosing a location from the meteorological data and then adding noise values to the pressure values from the meteorological time series. The noise will represent both the difference between (1) the meteorological data and ground truth and (2) between ground truth and the the values recorded by the smart device.

IV. Method

Suppose that we determine from the accelerometer that a device has been stationary. This might correspond to a period where a watch or phone has been left on a bedside locker overnight, or a period where no steps have been taken, such as someone sitting at a desk. Air pressure measurements from this period can be filtered to produce a higher quality time series of air pressure values on the same time scale as our meteorological data. We call this series $\hat{P}_t$, where the t takes values at discrete (e.g. hourly) values in some set T.

At the same time, we have our meteorological data $P_t(\mathrm{lat}, \mathrm{long})$, where the location, $(\mathrm{lat}, \mathrm{long})$, is from our filtered set of locations of interest, F. An obvious approach for matching locations to the observed time series is to use least squares

$$\underset{(\mathrm{lat},\mathrm{long})\in F}{\arg\min} \sum_{t\in T} \left(P_t(\mathrm{lat}, \mathrm{long}) - \hat{P}_t \right)^2 .$$

This minimisation can be performed by exhaustively searching our set of locations F. It can be calculated in a single pass over the meteorological data by accumulating the sum of squared error (SSE) for each location. Note, there is a possibility that the location minimising the SSE is not unique.

When we consider the possible sources of errors in our data, least squares has some attractive features. Our measured $\hat{P}_t$ will have errors relative to the actual air pressure, however a common assumption is to treat errors as normally distributed. Likewise, the meteorological data is reconstructed from observations, so normally distributed residual errors could be used as an approximation. This means that the difference between $\hat{P}_t$ and $P_t(\mathrm{lat}, \mathrm{long})$ will be the sum of two normals, and so also normally distributed. Consequently, a least squares estimate would also correspond to a maximum likelihood estimate.

We may also have systematic factors to accommodate. The data sheets for the smart device barometers indicate that there may be a larger absolute error than relative error, suggesting that the values may have some constant offset. Second, our meteorological data is for surface pressure, however the person may be stationary, but not at the expected surface height. In this case, we can introduce an extra parameter, δP, to represent the unknown constant offset due to these systematic factors. We can then estimate δP by minimising,

$$\underset{\substack{(\mathrm{lat},\mathrm{long})\in F \\ \delta P}}{\arg\min} \sum_{t\in T} \left(P_t(\mathrm{lat}, \mathrm{long}) - \hat{P}_t - \delta P \right)^2 .$$

Note that we can do the minimisation with respect to δP explicitly by differentiating, and find that at each $(\mathrm{lat}, \mathrm{long}) \in F$:

$$\delta P_{(\mathrm{lat},\mathrm{long})} = \frac{1}{|T|} \sum_{t\in T} P_t(\mathrm{lat}, \mathrm{long}) - \hat{P}_t .$$

Once $\delta P_{(\mathrm{lat},\mathrm{long})}$ is known, an exhaustive search of our locations F is possible. We could regard $\delta P_{(\mathrm{lat},\mathrm{long})}$ as an adjustment so that the mean value of $P_t(\mathrm{lat}, \mathrm{long})$ matches that of $\hat{P}_t$. In some sense, we are using one degree of freedom to estimate δP and then perform least squares matching, where the means have been adjusted to match. This also tells us that in the case of $|T| = 1$, all locations will have zero SSE.

In this case, minimisation can also be achieved in a single pass over the meteorological data by maintaining the SSE and the sum of $P_t(\mathrm{lat}, \mathrm{long}) - \hat{P}_t$ per location. If some bounds are known on the size of the systematic factors, the range of δP could be constrained (e.g. based on range of building height and absolute barometer error). Alternatively, if a prior distribution was available for δP, more sophisticated estimates could be made.

V. Results

A. Initial Tests

As an initial test of our method, we try to locate a simulated device at a location near Maynooth (53.366°,-6.612°), at ground level. We generate our pressure measurements using,

$$\hat{P}_t = P_t(53.366°, -6.612°) + n_t ,$$

where n_t are i.i.d. normal samples with mean zero and variance σ^2. We use four values for the noise: $\sigma = 0, 1.25, 5, 20$Pa. No noise corresponds to artificially good matching; 1.25Pa corresponds to an optimistic view of smart device barometer performance where the noise is about the same size as the precision in our data sources; 5Pa is a slightly pessimistic view on a smart device barometer; and 20Pa represents quite poor pressure measurements. Note that as time points are an hour apart, there should be time to make multiple barometer readings on the smart device and denoise them.

[3] See https://dev.fitbit.com/build/guides/sensors/barometer/ and https://developer.apple.com/documentation/coremotion/cmaltitudedata.

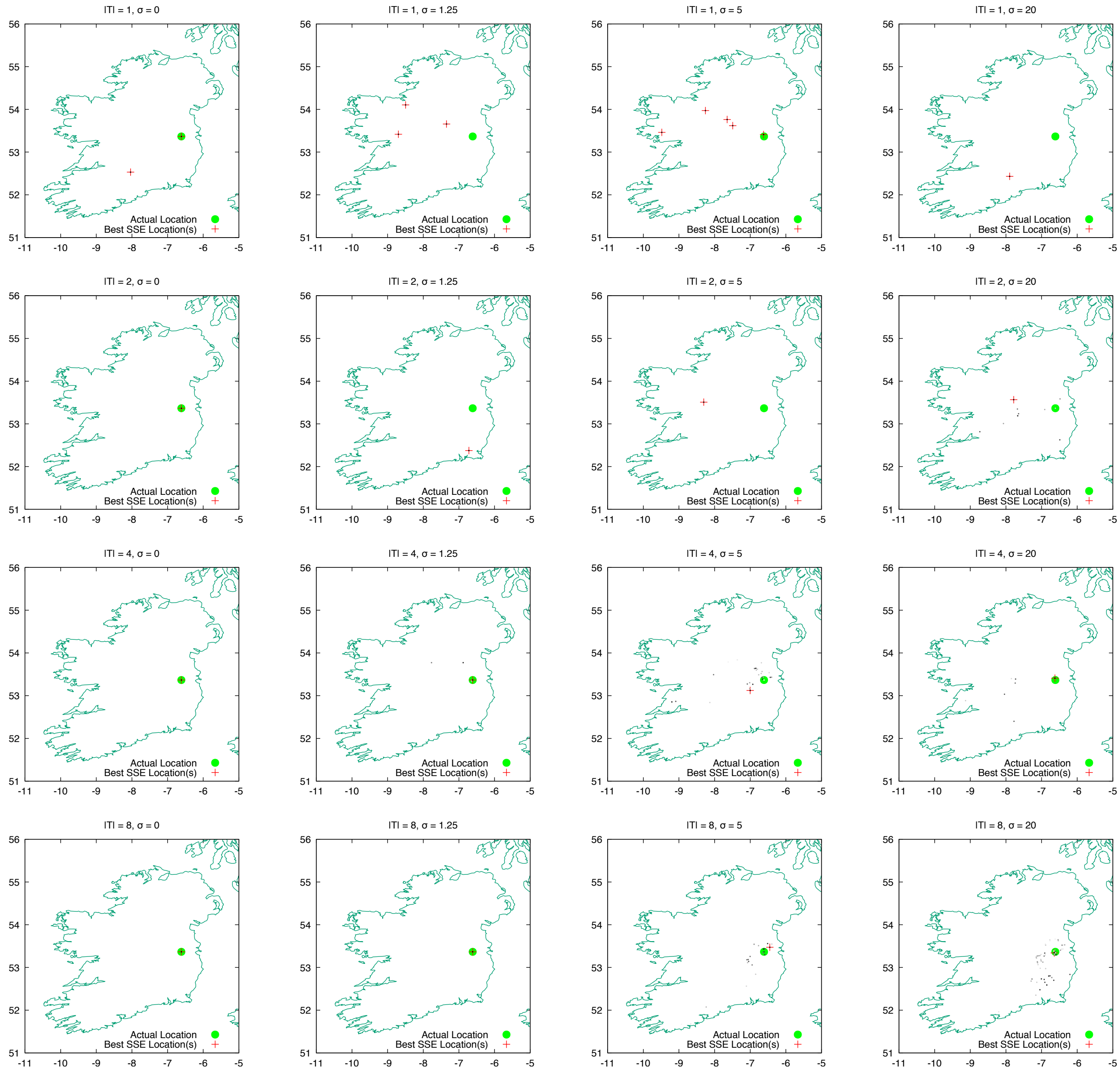

Fig. 4. Estimating the position of Maynooth. Rows represent attempts with $|T| = 1, 2, 4, 8$ points in time series. Columns represent noise with $\sigma = 0, 1.25, 5, 20$Pa. ● actual location, + locations with minimal SSE, gray dots indicate locations with SSE less than twice minimal.

We consider matching $|T| = 1, 2, 4, 8$ hourly-spaced time points, corresponding to someone being stationary for lunch, a cinema trip, spending an afternoon working at a desk or sleeping.

For each choice of σ and $|T|$ we pick a beginning time at random within our data, and simulate matching $(\hat{P}_t)_{t \in T}$ to our meteorological data using our first method. The results are shown in Figure 4. The correct location is shown as a green circle, and a red cross indicates locations with least square error. Rather than just show the location(s) with the least square error, we also show locations where the SSE is within a within a factor of two of the minimum.

Particularly, in the case where we use only one time point, we see that there may not be a unique location that is identified by our method. This is not surprising, given our observations about the spread of the meteorological air pressure data.

We also see that for $|T| = 1$ or 2 the estimated location may be quite far from the actual location, and increasing amounts of noise reduce the the ability to give estimates close to the right location. For $|T| = 4$, the locations estimated are all in the right part of the country and for $|T| = 8$ the estimates are all quite close, even in the case of high noise.

We make two other observations, based on running these tests. First, locations at different elevations to Maynooth

$\|T\|$	σ (Pa)	number of locations 1	2	3	4	5+
1	1.25	52.4%	24.1%	11.2%	5.9%	6.4%
1	5	51.8%	24.2%	12.5%	6.4%	5.0%
1	20	54.7%	22.8%	11.7%	6.5%	4.3%
2	1.25	99.2%	0.7%	0.1%	0.0%	0.0%
2	5	99.6%	0.4%	0.0%	0.0%	0.0%
2	20	99.8%	0.2%	0.0%	0.0%	0.0%
3+	> 0	100%	0.0%	0.0%	0.0%	0.0%

TABLE I
NUMBER OF MINIMAL SSE LOCATIONS WITHOUT SYSTEMATIC FACTOR ADJUSTMENT.

$\|T\|$	σ (Pa)	number of locations 1	2	3	4	5+
2	1.25	33.9%	17.6%	9.6%	7.1%	25.8%
2	5	41.2%	18.3%	10.1%	6.2%	24.2%
2	20	59.7%	14.3%	8.4%	3.4%	14.9%
3	1.25	96.4%	3.5%	0.1%	0.0%	0.0%
3	5	97.1%	2.6%	0.3%	0.0%	0.0%
3	20	99.0%	0.9%	0.1%	0.0%	0.0%
4+	> 0	100%	0.0%	0.0%	0.0%	0.0%

TABLE II
NUMBER OF MINIMAL SSE LOCATIONS ADJUSTING FOR SYSTEMATIC FACTORS.

typically have a large SSE, and so are rarely considered good matches. Second, each test runs quite rapidly, typically using less than 1s of clock time on a laptop.

B. Performance without Systematic Factors

The previous section gave us something of a feeling of the behaviour of the method for a single location. Clearly results may vary depending on the location and the measurement quality. There is also a potential issue of multiple locations minimising the SSE.

To study these problems, for each combination of numbers of time points and noise, we picked 1,000 random locations. Then for each location we conducted experiments as described in the previous section. We make two changes: first we omit $\sigma = 0$Pa as somewhat unrealistic; and we include $|T| = 3$, as there appears to be an interesting performance improvement between $|T| = 2$ and $|T| = 4$.

For each run, we first noted the number of locations solving our least squares problem. The results are summarised in Table I. We see that the problem of multiple optimal locations is largely confined to the case where we are only matching a single time point. Indeed, we didn't observe any situations where there were multiple optimal locations when matching more than two time points.

We next consider how good these optimal estimated locations are, relative to the location selected for each experiment. We noted the haversine distance between each estimated location and the actual location. In the case of multiple estimated locations, we take the distance averaged over the locations. The resulting cumulative distribution of the distances is shown in Figure 5.

In the case where $\sigma = 1.25$Pa and we have $|T| \geq 3$ time points, we achieve a location within about 2.5km more than 80% of the time. Even in the case where we have only two time points, we are still within 2.5km more than 50% of the time. However, the performance with just one time point is rather poor, only producing an answer within 100km around 40% of the time.

Naturally, it is more challenging to estimate locations when $\sigma = 5$Pa. Now we we need 8 time points to get a location within 2.5km 80% of the time. With 4 time points, we can get within 10km approximately 50% of the time. When $\sigma = 20$Pa even estimates using 8 time points have large distance errors much of the time.

C. Performance with Systematic Factors

In Section IV we outlined a method to handle situations where the smart device's barometer readings may have a systematic difference from the meteorological data due to differences due to height above ground level, errors in device measurements or other systematic factors. In this section, we will test this method for estimating the location.

As noted, our method to handle these systematic factors essentially adjusts the data so that the mean pressure measured by the smart device matches the mean in the meteorological data and then proceeds to estimate the location using least squares on the adjusted data. Consequently, we effectively loose 1 degree of independence in the data. For example, as predicted, when $|T| = 1$, we find that all locations match equally well, and so we get no useful location information. To account for this, we consider $|T| = 2, 3, 4, 5, 8, 9$ as we know we need an extra degree of freedom.

We also note that because the estimator makes the mean pressure values match, there is no need to choose a size for the systematic factors, as the estimate of δP will adjust for any constant offset at the first step. This means that we do not need to choose an offset size for our simulations, as all constant offsets will perform in the same way. Consequently, we use noise that is $N(0, \sigma^2)$, as before, with $\sigma = 1.25, 5, 20$Pa.

Again, for each combination of $|T|$ and σ, we ran 1,000 tests at randomly chosen locations. We recorded the estimated locations with the smallest SSE, when corrected for the estimated systematic factors. Table II shows how many locations were found to achieve the minimal SSE.

We see that with $|T| = 2$ time points we frequently find multiple locations with minimal SSE. However, with $|T| = 3$ we get a unique location in most situations, and with $|T| \geq 4$ we always get a unique location. If we compare this to Table I, it tallies well with our intuition that we have used a degree of freedom to estimate the systematic factors, and so need an extra time point in order to be able to identify locations.

Next we study the distribution of the distances between the actual locations and the estimated locations in Figure 6. As in Section V-B, we use the average distance if multiple locations have minimal SSE.

In the case where we have good results from our barometer ($\sigma = 1.25$Pa), with eight or nine time points, we can still achieve an estimated location within approximately 2.5km of the actual location over 80% of the time. However, the

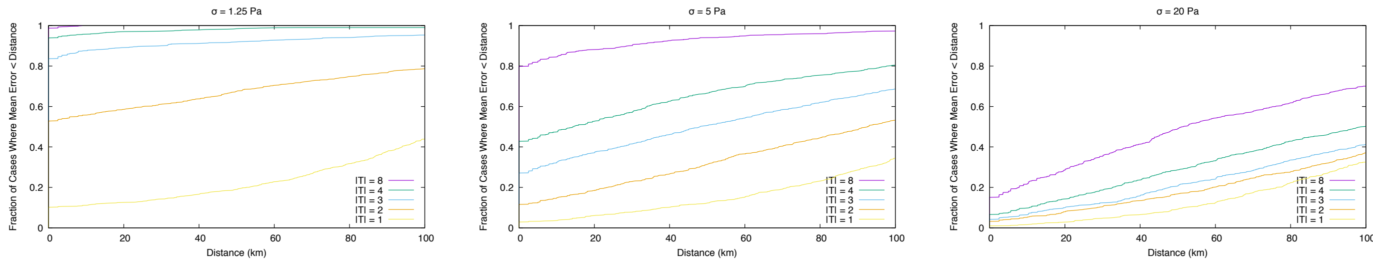

Fig. 5. Cumulative distribution function of the distance between the actual location and estimated location for $\sigma = 1.25, 5, 20$Pa.

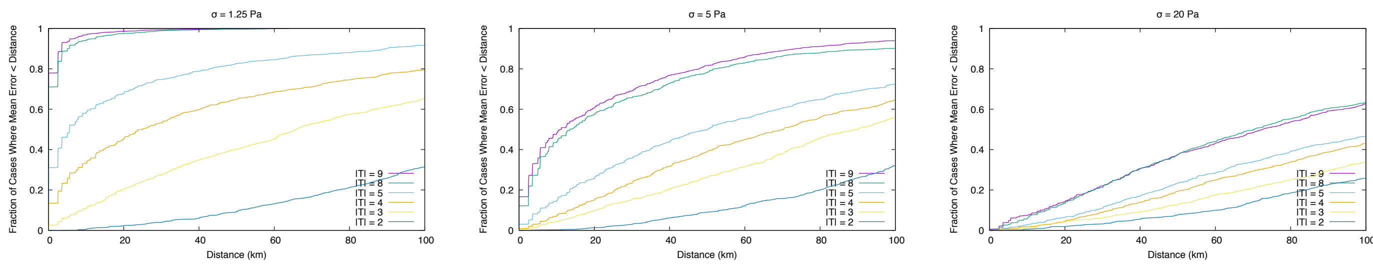

Fig. 6. Cumulative distribution function of the distance between the actual location and estimated location for $\sigma = 1.25, 5, 20$Pa adjusting for systematic factors.

performance in other cases ($|T| \leq 5$ or $\sigma = 5, 20$Pa) is considerably below that seen in Section V-B. This suggests that the absolute value of the air pressure is important in identifying the location. This is probably through its dependence on the elevation of the location, as we observed in Section V-A.

We conclude that the method still appears practical when accounting for systematic factors, however it requires more time points. This suggests that an improved location estimator that takes into account of restrictions on δP could be quite powerful.

VI. Conclusion

In this paper we looked at the feasibility of locating a stationary smart device via contemporary air pressure measurements. Our results show that this is feasible, particularly if the barometer is of good quality and a reasonable number of points can be matched (e.g. for a device that is stationary overnight). We see that it is possible to correct for systematic differences between meteorological and measured data, but that it makes the localisation task more challenging. Challenges for the method include the availability of contemporary meteorological information. Future work could include using additional sensor data (e.g. magnetometer), estimation of location using live data, testing if restricting the range for systematic differences can improve localisation, and the possibility of interpolating between locations.

Thanks to André Düsterhus, Daire Healy, Jack McDonnell and Eoin Whelan for advice on meteorological and geographic data and to Karl Stanley for testing an Apple Watch.

References

[1] A. Manivannan, W. C. B. Chin, A. Barrat, and R. Bouffanais, "On the challenges and potential of using barometric sensors to track human activity," *Sensors*, vol. 20, no. 23, p. 6786, 2020.

[2] B.-J. Ho, P. Martin, P. Swaminathan, and M. Srivastava, "From pressure to path: Barometer-based vehicle tracking," in *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, 2015, pp. 65–74.

[3] G. Chatip and F. Yilmaz, "Vehicle location estimation using air pressure measurements," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*. IEEE, 2018, pp. 1–4.

[4] J. Hsu, "The Strava heat map and the end of secrets," *Wired*, 2018.

[5] A. Hern, "Fitness tracking app Strava gives away location of secret US army bases," *The Guardian*, vol. 28, p. 7, 2018.

[6] L. Sweeney, "Only you, your doctor, and many others may know," *Technology Science*, vol. 2015092903, no. 9, p. 29, 2015.

[7] A. Narayanan and V. Shmatikov, "How to break anonymity of the Netflix prize dataset," *arXiv preprint cs/0610105*, 2006.

[8] M. Barbaro, T. Zeller, and S. Hansell, "A face is exposed for AOL searcher no. 4417749," *New York Times*, vol. 9, no. 2008, p. 8, 2006.

[9] E. Whelan, E. Gleeson, and J. Hanley, "An evaluation of MÉRA, a high-resolution mesoscale regional reanalysis," *Journal of Applied Meteorology and Climatology*, vol. 57, no. 9, pp. 2179–2196, 2018.

[10] I. Poese, S. Uhlig, M. A. Kaafar, B. Donnet, and B. Gueye, "IP geolocation databases: Unreliable?" *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 2, pp. 53–56, 2011.

[11] Y. Shavitt and N. Zilberman, "A geolocation databases study," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 10, pp. 2044–2056, 2011.

[12] *Digital Pressure Sensor*, Bosch, May 2015, rev. 1.14. [Online]. Available: https://cdn-shop.adafruit.com/datasheets/BST-BMP280-DS001-11.pdf

[13] *Precision Barometer and Altimeter Sensor*, Hopf, May 2013, rev. 2.0. [Online]. Available: https://cdn-shop.adafruit.com/datasheets/BST-BMP280-DS001-11.pdf