The following publication K. -H. Lam, C. -C. Cheung and W. -C. Lee, "New RSSI-Based LoRa Localization Algorithms for Very Noisy Outdoor Environment," 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), Tokyo, Japan, 2018, pp. 794-799 is available at https://doi.org/10.1109/COMPSAC.2018.10340.

New RSSI-based LoRa localization Algorithms for Very Noisy Outdoor Environment

Ka-Ho Lam, Chi-Chung Cheung and Wah-Ching Lee
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University
Hong Kong, China
Email: {Ka.Ho.LAM,encccl,enwclee}@connect.polyu.hk

***Abstract*—LoRa networks have been proven suitable for localization in outdoor environments [1] [2]. Performance investigations involving simulation models and real experiments have shown that the RSSI-based LoRa localization algorithms in [1] perform compatible to the Global Positioning System (GPS), the most popular outdoor localization system. However, the algorithms' performance degrades significantly in very noisy outdoor environments because the effect of noisy nodes (anchor nodes that are highly affected by noise) cannot be totally avoided during localization. Based on this observation, this paper proposes two new RSSI-based LoRa localization algorithms to further improve the accuracy of the localization in very noisy outdoor environments. One new algorithm iteratively removes all noisy nodes and uses the remaining anchor nodes to process the localization, while the other uses density clustering to provide the best estimation. Our performance investigation shows that the proposed algorithms significantly outperform the algorithms in [1] in terms of the localization error if the outdoor environment is very noisy.**

***Keywords*-localization system; smart environment; urban networking; LoRa technology;**

I. Introduction

Low Power Wide Area Networks (LPWANs) have recently become more and more popular because of their attractive features: extremely low power consumption, long range communication and low installation cost. Different wireless technologies have been developed to support LPWANs, wherein LoRa technology is the earliest and most popular [3] [4] [5]. Since the features of LoRa technology are suitable for localization, some studies have investigated the possibility to using this technology for localization in outdoor environments [1] [2]. Researchers are interested in outdoor environments because different wireless technologies are curently available to support localization in indoor environment (e.g., Bluetooth, Zeebee and WiFi) but the satellite-based technology is the only popular solution for outdoor environments. Furthermore, LoRa technology is attractive for localization because it can support both indoor and outdoor environments while satellite-based technology cannot.

The research work in [2] used TDOA (the Time Difference Of Arrival) to do the localization and carried out some real experiments in rural areas. The coverage area was a four-sided polygon around 2 to 3 km. The localization error (the distance between the estimated location and the real location of the target node) was not small (over 1 km in some cases) and the overall processing time to collect readings (data measurement) was too long, so that the overall procedure does not seem fit for real-time applications. [1] proposed two Received Signal Strength Indicator (RSSI)-based LoRa localization algorithms for outdoor environments. They handle non-Gaussian noise like blocking and multi-path, which are the most important problems in LoRa localization. Note that Gaussian noise (e.g., background noise) is handled by the traditional optimization algorithm called Linear Least Square (LLS) model (described later in Section II). The performance investigation in [1] showed that the performance of the proposed algorithms are compatible with the Global Positioning System (GPS), one of the most popular outdoor localization technologies, in terms of the localization error.

However, their performance degrades significantly if the outdoor environment is very noisy. One of the reasons is that the proposed algorithms are capable of reducing non-Gaussian noise (e.g., blocking and multi-path) if the number of noisy nodes (anchor nodes that are highly affected by noise) is small (one or two). If the number is large (more than two), the proposed algorithms may involve noisy nodes in localization and thus the accuracy of the estimation is significantly reduced. Based on this observation, we propose two new algorithms to reduce non-Gaussian noise efficiently if the outdoor environment is very noisy. One new algorithm iteratively removes all noisy nodes and uses the remaining nodes to process the localization, while the other uses density clustering to get the best estimation. Our performance investigation shows that the proposed algorithms significantly outperform the algorithms in [1] if the outdoor environment is very noisy.

The rest of this paper is organized as follows: Section II describes the related work including the system model, the well-known LLS optimization model and the algorithms in [1]. The algorithms in [1] are described here because our new algorithms are developed based on them so a brief description of the algorithms helps reader understanding. Section III presents the new algorithms. Section IV shows the simulation model and describes the real experiments carried out to investigate the new algorithms' performance. Then the performance comparisons between our new algorithms and the algorithms in [1] are presented. Section V contains the conclusions and future works.

II. Related Work

A. The System Model

The following notations are defined for future use. $(\cdot)^T$ is the transpose operation of matrix. Let Γ, $N(\Gamma)$ and AN_i be a set of anchor nodes (ANs), the total number of anchor nodes and the i-th AN in Γ respectively. A two dimensional network is used to represent the environment during localization. There is a target node (TN) with an unknown coordinate $\theta = [x, y]^T$ where $\theta \in \Re^2$, which will be estimated by a localization system. Moreover, the coordinate of AN_i is $\theta_i = [x_i, y_i]^T$ where $\theta_i \in \Re^2$ for $i = 1, 2, ..., N(\Gamma)$. Let

$$d_i = d(\theta, \theta_i) = \sqrt{(x - x_i)^2 + (y - y_i)^2} \tag{1}$$

be the distance between the TN and AN_i. According to [6], the Received Signal Strength Indicator (RSSI) between the TN and AN_i is defined as

$$Z_i = Z_0 - 10\alpha \log_{10} d_i + \omega, \tag{2}$$

where α is the path-loss exponent and ω is a zero-mean Gaussian random variable representing the background noise. RSSI is used for location estimation in this research work.

Let $\tilde{Z}(\Gamma) = \{\tilde{Z}_1, \tilde{Z}_2, ..., \tilde{Z}_{N(\Gamma)}, \tilde{Z}_0, \alpha\}$ be a set of measured RSSI values where $\tilde{Z}_i$ is the measured RSSI value between the TN and AN_i, and $\tilde{Z}_0$ is the referenced RSSI value.

B. Linear Least Squares Algorithm

One of the most popular mathematical models for localization is Linear Least Squares (LLS) algorithm [7–10]. It is widely used for localization if RSSI is used for location estimation. From (1) and (2), the system of equations can be rewritten into a matrix form,

$$\mathbf{A}\theta' = \mathbf{b}, \tag{3}$$

where

$$\mathbf{A} = \begin{bmatrix} -2x_1 & -2y_1 & 1 \\ -2x_2 & -2y_2 & 1 \\ \vdots & \vdots & \vdots \\ -2x_N & -2y_N & 1 \end{bmatrix}, \theta' = \begin{bmatrix} x \\ y \\ R \end{bmatrix} \tag{4}$$

and

$$\mathbf{b} = \begin{bmatrix} 10^{\frac{2}{\alpha}(Z_0 - Z_1)} - R_1 \\ 10^{\frac{2}{\alpha}(Z_0 - Z_2)} - R_2 \\ \vdots \\ 10^{\frac{2}{\alpha}(Z_0 - Z_N)} - R_N \end{bmatrix}. \tag{5}$$

Note that $R_i = x_i^2 + y_i^2$. The solution is given by [9],

$$\hat{\theta} = 0.5\mathbf{A}^{\dagger}\mathbf{b}, \tag{6}$$

where

$$\hat{\theta} = \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{R} \end{bmatrix} \text{ and } \mathbf{A}^{\dagger} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T. \tag{7}$$

Note that $\hat{x}$ and $\hat{y}$ are the estimated coordinates of the TN and $\hat{R} = \hat{x}^2 + \hat{y}^2$. LLS can eliminate Gaussian noise properly in localization but it is not designed to eliminate non-Gaussian noise. The algorithm in [1] designed to eliminate non-Gaussian noise is described later. To distinguish this traditional algorithm from the algorithms in [1] and our proposed algorithms, we call it as L3M (LoRa-based Localization Linear Model) for LoRa-based localization.

C. Estimated RSSI Error

L3M is widely used to estimate the location of the TN. However, since the exact location of the TN is unknown, the accuracy of estimation is also unknown. In [1], a heuristic approach called Estimated RSSI Error (ERE) is proposed to identify the accuracy of the estimated location of the TN. The concept of ERE is simple: it is expected that if the estimated location is close to the real location of the TN, the calculated RSSI value by using the estimated location and AN_i should be close to the measured RSSI value from AN_i. Thus, heuristically, if the calculated RSSI value is close to the measured RSSI value, the estimated location of the TN is close to the real location.

Let $\hat{Z}_i$ be the calculated RSSI value from the estimated location $\hat{\theta}$ to AN_i where

$$\hat{Z}_i = \tilde{Z}_0 - 10\alpha \log_{10} \hat{d}_i \tag{8}$$

and

$$\hat{d}_i = d(\hat{\theta}, \theta_i) = \sqrt{(\hat{x} - x_i)^2 + (\hat{y} - y_i)^2}. \tag{9}$$

Then, we define

$$ERE(\hat{\theta}, \Gamma, \tilde{Z}(\Gamma)) = \frac{1}{N(\Gamma)} \sum_i |\tilde{Z}_i - \hat{Z}_i|. \tag{10}$$

If $ERE(\hat{\theta}, \Gamma, \tilde{Z}(\Gamma))$ is close to zero, this indicates that $\hat{\theta}$ may be close to the real location. In a localization algorithm, when more than one estimated location of the TN is obtained, the one with the smallest ERE value is selected as it is expected that this one should have the smallest localization error (i.e., the distance between the TN and this estimated location should be the shortest one).

D. Existing LoRa-based Localization Algorithms

By using the above heuristic approach, two algorithms were proposed in [1] to eliminate non-Gaussian noise and hence improve the accuracy of localization.

One is L3M-C (LoRa-based Localization Linear Model with Clustering) which uses K-mean clustering to identify a noisy node (an anchor node that may be highly affected by noise) from all anchor nodes. At the beginning, it gets some estimated locations of the TN by applying L3M in different sets of anchor nodes. Then, it uses K-mean clustering to locate the best cluster by examining the ERE value of the center of each cluster. After that, by excluding the best cluster and counting the number of the occurrences of each anchor node in the rest of the clusters, it locates the noisy node and processes L3M again for the rest of anchor nodes (i.e., excludes the noisy node in localization).

The other algorithm is L3M-MRE (LoRa-based Localization Linear Model with Minimum RSSI Error). Like L3M-C, it gets some estimated locations of the TN by applying L3M in different sets of anchor nodes. Then it selects the one with the smallest ERE value among them.

The performance investigation in [1] showed that both algorithms can significantly improve the localization error significantly compared with L3M for outdoor environments. L3M-C can effectively remove a noisy node and process the localization with the rest of anchor nodes; while L3M-MRE can effectively select non-noisy nodes to do the localization. However, if the outdoor environment is very noisy and there are too many noisy nodes, neither algorithm cannot work properly. For L3M-C, it can remove one noisy node only. Other noisy nodes may still be involved into the localization and thus the localization error cannot be reduced significantly. For L3M-MRE, however, it is possible that all different sets of anchor nodes include noisy nodes and thus there is no chance to avoid noisy nodes when localization is processed thus its localization error also cannot be reduced significantly.

III. New algorithms for Very Noisy Outdoor Environment

Based on the above limitations, we propose two new LoRa-based localization algorithms for very noisy outdoor environments. We know that in very noisy outdoor environments there may be many noise nodes and we need to either eliminate or avoid all of them to process the localization. A new algorithm called LoRa-based Localization Linear Model with Iterative Elimination (L3M-IE) is proposed to iteratively eliminate all possible noisy nodes during localization. Another new algorithm called LoRa-based Localization Linear Model with Density-based Clustering (L3M-DC) is proposed to select non-noisy nodes in localization.

A. L3M-IE

L3M-C uses K-mean clustering to eliminate a noisy node in localization [1]. However, L3M-C works properly if there is at most one noisy node. If there is more than one noisy node, its performance will be significantly degraded because some noisy nodes may still be involved in localization. To address this limitation, we propose L3M-IE to further improve localization performance. It repeatedly applies L3M-C until all noisy nodes are eliminated. The L3M-IE algorithm is shown below:

- **Step 1**: Set Γ as the set of anchor nodes for localization.
- **Step 2**: Perform L3M-C to the set of anchor nodes. Get the estimated location of the TN and the noisy node.
- **Step 3**: Remove the noisy node from the set of anchor nodes and Go to Step 2 until the number of anchor nodes in the set is reduced to three.
- **Step 4**: Select the best estimated location from all estimated locations found in Step 2 by examining their ERE values.

The algorithm above eliminates noisy nodes one by one until anchor nodes cannot be further reduced (the minimum number of anchor nodes for localization is three). Moreover, we get the estimated location in each iteration and we expect that the best one (i.e., the one nearest to the real location of the TN) should be one of them. Thus, at the end, we use the heuristic approach ERE to locate the best one.

B. L3M-DC

L3M-MRE uses ERE to get the best estimation among all estimated locations [1]. However, when the environment is very noisy, the localization of all estimated locations may still involve noisy nodes and thus the overall performance is still degraded by noisy nodes. To address this limitation, we propose L3M-DC to further improve the performance by using Density-based spatial clustering of applications with noise (DBSCAN) to identify non-noisy nodes. DBSCAN groups sample points into high-density regions and such regions are separated by low-density regions [11] [12]. Furthermore, DBSCAN excludes outliers, if any, and they will not be found in high-density regions. We expect that if an anchor node is not greatly affected by the noise, all estimated locations involving this anchor node should be quite similar and they should be clustered in the same high-density region. Thus, heuristically, by using DBSCAN, non-noisy nodes can be found in high-density regions. The L3M-DC algorithm is shown below:

- **Step 1**: Get some estimated locations of the TN by applying L3M in different sets of anchor nodes.
- **Step 2**: Use DBSCAN to group them into different clusters.
- **Step 3**: Select three anchor nodes from each clusters by examining the number of occurrences.

- **Step 4**: Get some estimated locations from three selected anchor nodes of each clusters.
- **Step 5**: Select the best estimated location from all estimated locations found in Step 4 by examining their ERE values.

  In Step 3 of the above algorithm, we expect that the anchor nodes selected from each clusters are non-noisy nodes or the effect of noise is small. Thus, the final estimated location will not be greatly affected by the noise if we have enough non-noisy nodes in the whole set of anchor nodes.

IV. Performance Investigation

This section reports the performance of the proposed localization algorithms that were investigated using computer simulation and a real outdoor experiment.

A. Performance investigation of the localization algorithms by using simulation with real data

A simulation model was developed to investigate the performance of the proposed algorithms. The RSSI data used in the simulation were collected from a real experiment in Kai Tak (see Fig. 1). The distance information was measured using an infra-red distance meter with a 1.5 mm measurement error. Over 40 RSSI values were obtained at every measurement point. Then, these measured RSSI values were grouped as a dataset. The localization area in the simulation model was set to a 100 m diameter circle. All anchor nodes were evenly distributed on the circumference of the circle. To simulate the RSSI measurement in real situations, 20 RSSI values were randomly selected from the dataset (a reading in the dataset can be selected more than once). Finally, 2,000 simulations were run for each environment setting in the performance investigation process. All winners in the performance investigation are highlighted in bold.

To investigate different noise levels in a real environment, a measured RSSI value was randomly added a floating number between -20 to 20 and the probability that a measured RSSI value was modified in this way is called Noise Factor (NF). Note that all RSSI values from the dataset were measured from a real experiment. Therefore, even if NF was set to zero, noise (e.g. Gaussian noise and measurement error) can still be found in the measured RSSI value.

Table I shows the performance comparison among L3M, L3M-C, L3M-MRE, L3M-IE and L3M-DC when NF = 0. This environment is not noisy and L3M-MRE gave the best performance among all of them. Moreover, L3M-C and L3M-MRE performed similarly, and L3M-IE and L3M-DC performed sightly worse than L3M-C and L3M-MRE. This is because there were not too many noisy nodes and thus the advantages of L3M-IE and L3M-DC cannot be shown clearly. Finally, when the number of anchor nodes increased, the difference was small for all algorithms.

Figure 1. The location of Kai Tak in Google map.

Table I

THE PERFORMANCE COMPARISON OF DIFFERENT LOCALIZATION ALGORITHMS IN AN OUTDOOR ENVIRONMENT (NF = 0.00)

$N(\Gamma)$	Localization error (m)				
	L3M	L3M-C	L3M-MRE	L3M-IE	L3M-DC
16	23.58	22.79	**18.93**	30.79	31.59
14	23.88	22.78	**19.86**	29.53	28.54
12	23.99	23.37	**19.64**	27.85	25.14
10	24.44	24.67	**20.19**	28.72	23.23

Table II

THE PERFORMANCE COMPARISON OF DIFFERENT LOCALIZATION ALGORITHMS IN A VERY NOISY OUTDOOR ENVIRONMENT (NF = 0.70)

$N(\Gamma)$	Localization error (m)				
	L3M	L3M-C	L3M-MRE	L3M-IE	L3M-DC
16	318.48	150.83	35.67	37.18	**33.08**
14	323.64	146.02	36.72	37.64	**34.66**
12	326.57	144.71	38.41	39.81	**37.97**
10	351.04	140.83	**39.80**	40.38	50.44

Table II shows the performance comparison among L3M, L3M-C, L3M-MRE, L3M-IE and L3M-DC when NF = 0.7. It can be considered a very noisy environment and L3M-DC outperformed all other algorithms when the number of anchor nodes was 12, 14 and 16 because it can identify non-noisy nodes effectively. L3M-MRE was the best when the number of anchor nodes was 10 because the number of anchor nodes is small and L3M-DC cannot locate non-noisy nodes properly by examining the number of occurrences. Moreover, the performance of L3M-IE was much better than L3M-C. Which means iterative elimination can effectively eliminate noisy nodes.

Table III shows the comparison between L3M-MRE and L3M-DC with different NF values. When the NF value is high, L3M-DC outperformed L3M-MRE, which is expected because L3M-DC can select non-noisy nodes properly.

Table IV shows the comparison between L3M-C and L3M-IE with different NF values. L3M-IE always outperformed L3M-C because its performance is always better during localization if we can eliminate all possible noisy nodes.

Table III
THE PERFORMANCE COMPARISON OF L3M-MRE AND L3M-DC WITH DIFFERENT NF VALUES

	Localization error (m)		
	$N(\Gamma) = 16$	$N(\Gamma) = 14$	$N(\Gamma) = 12$
L3M-MRE (NF=0.40)	**27.94**	29.05	30.82
L3M-MRE (NF=0.70)	**35.66**	36.72	38.41
L3M-MRE (NF=0.99)	**45.14**	47.15	49.15
L3M-DC (NF=0.40)	**30.10**	31.09	31.10
L3M-DC (NF=0.70)	**33.08**	34.66	37.97
L3M-DC (NF=0.99)	**37.06**	42.52	53.98

Table IV
THE PERFORMANCE COMPARISON OF L3M-C AND L3M-IE WITH DIFFERENT NF VALUES

	Localization error (m)		
	$N(\Gamma) = 16$	$N(\Gamma) = 14$	$N(\Gamma) = 12$
L3M-C (NF=0.40)	76.96	**74.12**	74.83
L3M-C (NF=0.70)	150.83	146.02	**144.71**
L3M-C (NF=0.99)	223.58	**219.81**	229.69
L3M-IE (NF=0.40)	**28.11**	29.00	31.10
L3M-IE (NF=0.70)	**37.18**	37.64	39.81
L3M-IE (NF=0.99)	**46.79**	48.56	51.14

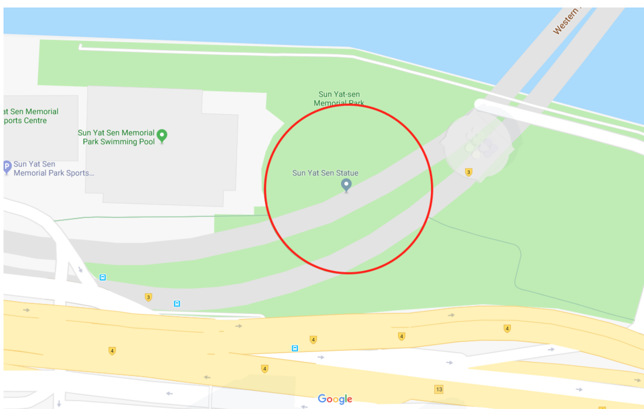

Figure 2. The location of Sun Yat Sen Memorial Park in Google map.

B. *Performance investigation of the localization of the target node in a real outdoor environment*

An outdoor experiment was carried out for this performance investigation. The location was Sun Yat Sen Memorial Park (see Fig. 2). Eight anchor nodes were used for localization and they were placed in a rough circle. The results are shown in Table V. All locations were measured using a GNSS receiver (Trimble R10) with a measurement error with ± 8 mm [13]. This experiment shows that L3M-IE outperformed L3M-C significantly and L3M-DC was the best among them all.

V. CONCLUSIONS AND FUTURE WORK

LoRa technology has been proven to be suitable for localization in outdoor environments because of its long communication range and low power consumption. Some algorithms that have been proposed for localization can handle Gaussian noise (e.g., background noise) and non-Gaussian noise (e.g., blocking and multi-path)

Table V
THE PERFORMANCE COMPARISON OF ALL LOCALIZATION ALGORITHMS IN THE REAL EXPERIMENT.

Test Point	Localization error (m)				
	LLS	L3M-C	L3M-IE	L3M-MRE	L3M-DC
P_1	14.92	27.17	13.06	**12.28**	13.72
P_2	35.04	10.04	8.94	8.92	**4.20**
P_3	36.09	26.21	15.60	7.93	**4.61**
Average	28.68	21.14	12.53	9.71	**7.51**

properly. However, when the environment is very noisy, their performance degrades significantly. To address this limitation, this paper proposes two new algorithms to eliminate noisy nodes and select non-noisy nodes properly for localization. Our performance investigation shows that the proposed algorithms significantly outperform existing algorithms in very noisy outdoor environment.

In the future, we will work on a more complicated algorithm to remove unreliable measurement(s) and focus on making use of reliable measurement(s) during computation. Additionally, we will work on more real experiments to investigate the performance of our localization algorithms in different outdoor environments.

ACKNOWLEDGMENT

The authors would like to thank Dr. Rex G. Sharman for his proof-reading. Furthermore, the authors would also like to thanks Cloudicty Co. Ltd. for their financial supports in building the LoRa-based localization system.

REFERENCES

[1] K. H. Lam, C. C. Cheung, and W. C. Lee, "Lora-based localization systems for noisy outdoor environment," in *2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2017, pp. 278–284.

[2] B. C. Fargas and M. N. Petersen, "Gps-free geolocation using lora in low-power wans," in *2017 Global Internet of Things Summit (GIoTS)*, June 2017, pp. 1–6.

[3] "Lora-alliance," https://www.lora-alliance.org/technology, accessed 31 December, 2017. [Online]. Available: https://www.lora-alliance.org/technology

[4] M. Centenaro, L. Vangelista, A. Zanella, and M. Zorzi, "Long-range communications in unlicensed bands: the rising stars in the iot and smart city scenarios," *IEEE Wireless Communications*, vol. 23, no. 5, pp. 60–67, 2016.

[5] L. Vangelista, "Frequency shift chirp modulation: the loratm modulation," *IEEE Signal Processing Letters*, vol. PP, no. 99, pp. 1–1, 2017.

[6] K. Pahlavan and A. H. Levesque, *Wireless information networks*. John Wiley & Sons, 2005, vol. 93.

[7] N. Salman, A. H. Kemp, and M. Ghogho, "Low complexity joint estimation of location and path-loss exponent," *IEEE Wireless Communications Letters*, vol. 1, no. 4, pp. 364–367, August 2012.

[8] I. Guvenc, S. Gezici, F. Watanabe, and H. Inamura, "Enhancements to linear least squares localization through reference selection and ml estimation," in *2008 IEEE Wireless Communications and Networking Conference*, 2008, pp. 284–289.

[9] K. Madsen, H. B. Nielsen, and O. Tingleff, *Methods for Non-Linear Least Squares Problems (2nd ed.)*. Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 2004.

[10] Y. Wang, "Linear least squares localization in sensor networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, p. 51, Mar 2015. [Online]. Available: https://doi.org/10.1186/s13638-015-0298-1

[11] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *Kdd*, vol. 96, no. 34, 1996, pp. 226–231.

[12] A. Ram, A. Sharma, A. S. Jalal, A. Agrawal, and R. Singh, "An enhanced density based spatial clustering of applications with noise," in *2009 IEEE International Advance Computing Conference*, March 2009, pp. 1475–1478.

[13] Trimble r10 gnss system datasheet. [Online]. Available: https://geospatial.trimble.com/products-and-solutions/r10