GeoSeg: Path Loss Estimation Method for Outdoor LoRa LPWAN

Alimuddin Arriesgado[13], Marc Caesar Talampas[1245]
[1]Electrical and Electronics Engineering Institute, University of the Philippines, Quezon City, Philippines
[2]Space Technology Missions and Systems Bureau, Philippine Space Agency, Quezon City, Philippines
[3]asarriesgado@mcl.edu.ph, [4]marc.talampas@eee.upd.edu.ph, [5]marc.talampas@philsa.gov.ph

Accurate path loss modeling in low-power wide-area networks (LPWAN) benefits a variety of processes such as network design and analysis, and supports applications such as localization and tracking. Land cover information is indispensable in accurate path loss modeling; most of the path loss models employed in LoRa LPWAN studies utilize land cover information to create a regional description of the propagation environment (e.g., urban, semi urban, etc.). However, this regional approach to path loss modeling is incongruent with the anisotropic characteristic of propagation environments with long radio links and diverse land covers which is the case in outdoor LPWAN deployments. Recent LPWAN studies show that using per-link modeling with detailed land cover data from satellite images gives more accurate path loss estimates than regional modeling. Similarly, geographical clustering is another technique that addresses the issue related to regional modeling by identifying subregions within the propagation environment and developing per-subregion path loss model. This study introduces GeoSeg, an enhanced log-distance path loss model that integrates per-link modeling with geographical clustering. Developed and evaluated using an open-access dataset for LoRa, GeoSeg demonstrates 23% improved path loss estimation accuracy compared to the standard log-distance model.

***Index Terms*—Path loss model, Low power wide area networks, LoRa**

I. Introduction

The number of IoT connected devices has seen steady growth in recent years with an expected 29 billion connections by 2027 [1]. Outdoor IoT applications need technologies that transmit data over long distances with minimal power. LoRa is a long range, low power wireless communication technology that has become the de facto platform of IoT [2]. At only 14dBm transmission power, LoRa can transmit data of up to 25km with line-of-sight [3]. Hundreds of millions of LoRa devices are connected in various networks in more than 100 countries [2]; with these devices come a host of applications that will require or benefit from accurate path loss modeling.

Many LoRa studies use regional modeling for path loss estimation [4]–[8]. Regional modeling relies on regional environment information to obtain a general description of the whole propagation area. In regional modeling, after categorizing the propagation environment (e.g., urban, suburban, rural, etc.), all radio links in the propagation area are modeled using the same path loss formula or propagation parameters. For example, the path loss of all the radio links for a suburban area can be modeled using the semi-urban variant of the Okumura-Hata model or the log-distance path loss model with its path loss exponent value adjusted for a suburban environment. This regional approach to path loss modeling results in an isotropic representation of the gateway coverage.

Long radio links, such as LoRa, deployed in a diverse land-cover environment encounter different propagation environments depending on the direction of travel [9]. This leads to a non-isotropic gateway coverage making regional modeling incompatible. In contrast, a per-link approach can capture the non-isotropic characteristic of the gateway coverage by taking into account the diversity of the traversed land covers of each radio link.

Geographical clustering is another approach that addresses the issues related to regional models in diverse propagation environments. This approach partitions the propagation area into distinct subregions using unsupervised machine learning, with each subregion assumed to exhibit unique propagation characteristics. Per-subregion models are then developed to capture these variations.

The log-distance path loss model is an empirical analytical model widely used in LoRa studies, offering an environment-specific parameter: the path loss exponent (η). Calibrating the log-distance path loss model primarily involves determining η. Once η is known, the total path loss (PL) of a transmission at a distance (d) can be calculated using (1). Here, PL_0 represents the measured path loss at the reference distance d_0, and $X_\sigma \sim \mathcal{N}(0, \sigma^2)$ denotes the shadowing component modeled as a zero-mean Gaussian random variable with variance σ^2.

$$PL = PL_0 + 10\eta \log\left(\frac{d}{d_0}\right) + X_\sigma \tag{1}$$

This study introduces GeoSeg, a variant of the log-distance path loss model that integrates per-link modeling with geographical clustering. GeoSeg was developed using an open-access LoRaWAN dataset [10]. To evaluate the performance of our proposed method, its path loss estimation accuracy was compared to the standard log distance path loss model.

GeoSeg's development introduced several methodological innovations, which are summarized below:

1) Combined per-link modeling with geographical clustering to develop a per-link variant of the log-distance path loss model. This variant achieves improved path loss estimation accuracy compared to the standard log-distance path loss model.
2) Developed a context-aware embedding process that characterizes the 2D propagation environment between transceivers, incorporating land cover types, their spatial sequence, and positions relative to line-of-sight.

979-8-3315-6946-4/25/$31.00 ©2025 IEEE

3) Modified a Hidden Markov Model (HMM)-based technique for variable-length sequence clustering, to handle large-scale dataset processing.
4) Introduced a calibration process for the log-distance path loss model that is easily integrable to preexisting LPWAN deployments and/or datasets.

II. Related Work

A. Per-link Modeling

Recent LoRa-based studies have adopted two main approaches for per-link path loss estimation: creating modified per-link versions of existing propagation models [9], [11], or employing neural networks trained on per-link data [12]. All these approaches utilize multispectral satellite images to capture a more granular description of radio links.

One study proposed a per-link variant of the Okumura-Hata model [9], which was selected for its adaptable formulations for different environments (e.g., urban, semi-urban, open space). Satellite imagery was used to identify the dominant land cover along each radio link, allowing classification of the environment. The appropriate Okumura-Hata formulation was then applied to estimate path loss.

Another method, SateLoc, is a per-link variant of the log-distance path loss model that uses segmented links to compute total path loss [11]. This method represents the radio link as a straight line connecting the target node and the gateway. By leveraging land cover classification data, adjacent cells traversed by the radio link are grouped into segments based on similarity. Path loss is then computed sequentially for each segment, starting from the segment closest to the target node. SateLoc provides finer land cover granularity than the per-link Okumura-Hata model. However, SateLoc's representation of radio links as single-pixel-wide lines fails to account for environmental attenuation from surrounding areas. Additionally, its path loss estimates are highly sensitive to land cover misclassifications along these linear paths [12].

Unlike SateLoc, DeepLoRa [12] represents a radio link as a 2D rectangular region rather than a line. It models the propagation path as a continuous series of micro-links spanning from the target node to the base station, forming a rectangular area. This approach more effectively captures signal attenuation from surrounding land cover [12]. While DeepLoRa achieves higher accuracy than SateLoc, its neural network-based design requires extensive measurement data for training.

B. Geographical Clustering

The site-specific modeling employed through geographical clustering can lead to improved path loss estimation accuracy, as demonstrated by a study conducted on an LTE network [13]. Clustering algorithms rely on input feature vectors that encode propagation region information between transceivers. Therefore, a critical aspect of geographical clustering is how the propagation region is represented and transformed into meaningful feature vectors for clustering. Some studies used fix-sized rectangular regions around the transmitter and receiver to capture the effect of built structures in the propagation environment [14]–[17]. These rectangular regions are then rotated depending on the Tx-Rx direction. However, since these regions are fixed in size, the propagation representation does not fully cover the entire link region between the target node and the base station. A representation similar to that of DeepLoRa is desirable—i.e., a 2D region connecting the transceivers. However, DeepLoRa's embedding method does not account for the spatial arrangement of land covers relative to the line of sight within the link region.

C. Estimation of the Path Loss Exponent (η)

Possible values of η are available in the literature [18]. However, except for the free-space scenario, the path loss exponent values for different environments are typically provided as ranges. Therefore, site-specific model calibration remains necessary for improved model performance. Field measurements are usually required prior to the actual estimation of η (e.g., using Maximum Likelihood Estimation). The data collection methods used in previous studies require the transceivers to be co-located within the same land cover type area or environment for which η is being estimated [19][11]. This requirement may not be easily implementable with existing deployments or datasets where such conditions are not met.

In summary, a desirable path loss estimation model should:

1) Employ per-link modeling.
2) Incorporate the 2D spatial information of land covers into radio link calculations.
3) Feature a calibration process that's easily integrable with existing LPWAN deployments or datasets.

Our proposed method, GeoSeg, uses a segmented log-distance path loss model, with each segment corresponding to a traversed subregion. This approach embeds 2D spatial information into each segment of the radio link. Additionally, this study utilizes a per-subregion calibration process that does not require co-locating the transmitter and receiver within the same subregion. We hypothesize that this propagation model provides more accurate path loss estimations than the standard log-distance model in environments with diverse land cover types.

III. Methodology

The development of the proposed path loss model, GeoSeg, involved three main processes: land cover classification, geographical clustering, and model calibration. Multispectral satellite imagery and GIS data were used for land cover classification. For geographical clustering, this study developed an embedding method that generated features from each radio link's propagation region and applied a Hidden Markov Model (HMM)-based clustering approach. Finally, the model calibration step involved simultaneously determining the path loss exponent for each subregion without requiring the transmitter and receiver to be co-located within the same subregion of interest.

A. Land Cover Classification

The area of study is located in Antwerp, Belgium, where the dataset for the path loss model calibration was collected [10]. It is a 4.3 km × 2.7 km region bounded by the lower-left and upper-right coordinates ($51.192646°$N and $4.425558°$E) and ($51.231342°$N and $4.425558°$E), respectively. The area comprises five main land cover types: buildings, trees, open fields, roads, and water. The multi-spectral image was obtained from the Copernicus Sentinel-2 database and filtered for the period from January 1, 2019, to December 31, 2019. The bands used for the classification were B2, B3, B4, B5, B6, B7, and B8. Additionally, the NDVI (Normalized Difference Vegetation Index) was derived from the NIR and red bands (i.e., B8 and B4, respectively).

Google Earth Engine (GEE) was used for land cover classification. Training and validation cells were manually labeled, with the following distribution: 552 buildings, 513 trees, 504 fields, 517 roads, and 520 water, totaling 2,606 data points. Of these, 80% of the dataset was used for training, and the remaining 20% was reserved for validation. GEE offers four built-in classifiers: random forest, Naive Bayes, CART, and SVM. Among these, the random forest classifier achieved the highest overall classification accuracy of 83% for this study's dataset. The resulting land cover classification map has a resolution of 10 m × 10 m, matching the cell resolution of the multi-spectral image. This process generated a dataset comprising 181,451 entries, with each entry containing longitude, latitude, and land cover classification information.

B. Geographical Clustering

1) Embedding

The embedding process transformed each link region (width w and length n) into a vector $v \in \mathbb{R}^{wn}$. These vectors served as inputs for the clustering algorithm. To account for the influence of land cover type and its position relative to the line of sight of the link region, this study assigned a value ω to each cell in the propagation region. The value of ω depends on both the land cover type c and its relative position λ with respect to the link region's line of sight. The value of λ depends on the width w of the link region and its length n, both are expressed in terms of cell count. The value of ω and λ for a cell located in the i-th row and j-th column of the propagation region was computed using (2) and (3) respectively, where $i \in \{0, 1, ..., n-1\}$ and $j \in \{0, 1, ..., w-1\}$.

$$\omega_{ij} = \lambda_{ij} c_{ij} \tag{2}$$

$$\lambda_{ij} = \left\lfloor \frac{w}{2} \right\rfloor + 1 - \left| (j \bmod w) - \left\lfloor \frac{w}{2} \right\rfloor \right| \tag{3}$$

For this study, the land cover types field, road, and water were collectively categorized as open land covers and has $c = 0$. Buildings and trees have $c = 2$ and $c = 1$ respectively. Finally, the k-th element of the feature vector $v[k] = \omega_{ij}$ where $k = wi + j$. Fig. 1 shows a sample embedding process of a propagation region with $n = 2$ and $w = 7$.

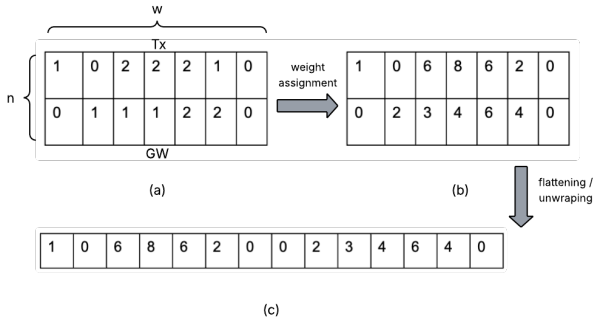

Fig. 1. Propagation region with $w = 7$ and $n = 2$. (a) Raw, unweighted cells containing only land cover type information. (b) Weighted cells after applying weight assignment formula. (c) Flattened propagation region matrix to create feature vector.

2) Clustering

Due to the variable lengths of propagation regions, their feature vectors exhibited differing dimensionalities. This necessitated the use of a clustering method capable of processing variable-length inputs—a common challenge in sequential data analysis [20]. To address this challenge, this study applied a Hidden Markov Model (HMM) based clustering method.

Consider a dataset D with N sequences $\{S_1, ..., S_N\}$ of varying lengths. The objective is to discover the inherent clustering that partitions these sequences into K distinct groups. The assumption is that each distinct group is generated by its own HMM. For this work, a clustering process [21] was modified to reduce the computational burden for large datasets. The modified clustering procedure proceeds as follows:

1) Randomly sample n sequences from the dataset D to create a reduced-sized dataset d of size n.
2) Fit an HMM m_i for each sequence $s_i \in d$, $1 \leq i \leq n$.
3) For each fitted model m_i and sequence s_j, evaluate the log-likelihood $L(s_j|m_i)$, $1 \leq i, j \leq n$. The log-likelihoods are used to define pairwise distances of sequence s_i and s_j [21].
4) Create a distance matrix A. The elements a_{ij} of A are equal to $\frac{1}{2}(L(s_j|m_i) + L(s_i|m_j))$.
5) Apply a hierarchical clustering algorithm to generate K clusters from the distance matrix A.
6) Having clustered the sequences into K groups, fit K HMMs $\{M_1, ..., M_k\}$, one to each cluster.
7) For each sequence $S_l \in D$, assign its cluster by solving $k = \underset{M_k}{\operatorname{argmax}}\ L(S_l|M_k)$.

C. Model calibration per subregion

1) Field measurement

Field measurement data were obtained from an open-access LoRaWAN dataset collected in an urban area in Antwerp, Belgium [10]. Only a subset of the dataset was used, selected by focusing on a gateway located in an

area with heterogeneous land cover. The reference ID of the gateway used for this study is FF01753E located near 51.1974874°N and 4.406706°E. For simplicity, only transmissions with spreading factor 7 (SF7) were included, as these were the most numerous. Only the relevant metadata were retained as features, namely longitude and latitude coordinates, and RSSI values. The distance between each transmission location and the gateway was derived and added as a feature. Local coordinates x and y were also added as features. Finally, each transmission location was assigned to the cluster of the nearest cell in the clustering map. This initial feature selection process resulted in a new dataset comprising 22,912 instances and the following features: latitude, longitude, x, y, RSSI, distance, and cluster.

RSSI measurements are inherently noisy; consequently, data preprocessing was applied to mitigate noise effects. The filtered dataset was derived from the original dataset through local averaging.

Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ be the original set of datapoints where each $\mathbf{x}_i$ represents a feature vector.

For each datapoint $\mathbf{x}_i$, using the longitude and latitude information of each dataset entry, identify the set of neighbor datapoints $\mathcal{N}_i$ whose elements are $x_j \in \mathbf{X}$ within a radius r:

$$\mathcal{N}_i = \{\mathbf{x}_j \in \mathbf{X} \mid \|\mathbf{x}_j - \mathbf{x}_i\| \leq r\} \tag{4}$$

An outlier is removed and is identified when it has less than N neighbors:

$$|\mathcal{N}_i| < N \tag{5}$$

For datapoints with at least N neighbors, their feature values are updated by taking the average values of all their neighbors.

$$\mathbf{x}_i = \frac{1}{|\mathcal{N}_i|} \sum_{\mathbf{x}_j \in \mathcal{N}_i} \mathbf{x}_j \tag{6}$$

Finally, a filtered dataset $\mathbf{X}'$ is created using the updated values. In this study, we used $r = 50m$ and $N = 30$. For simplicity, we did not explore alternative values for the radius and neighbor count. Future work could investigate the potential impact of varying these parameters on path loss estimation accuracy.

2) Formulation of the problem

The log-distance path loss model (1) served as the foundational model for this study. This empirical analytical model is the most widely adopted path loss model in LoRa studies, offering environment-specific parameters that can be calibrated. Calibrating the log-distance path loss model involves, primarily, the determination of the path loss exponent (η). In (1), only one η needs to be identified since it is assumed that the radio link traverses a homogeneous propagation environment.

GeoSeg represents the radio link between the transmitter and receiver as a series of segments where each segment is a subregion. This variation in the representation of the radio link entails a modification in the formulation of the standard log distance path loss model. Fig. 2a illustrates a simplified scenario that was used to formulate the problem.

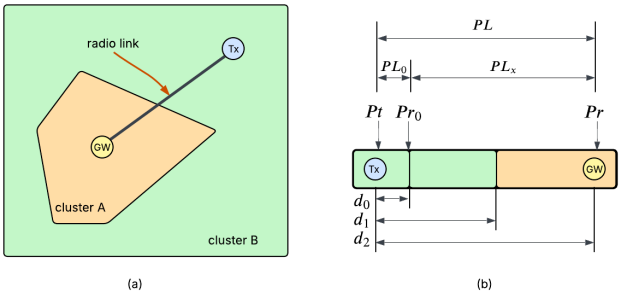

Fig. 2. Simple scenario where the propagation region traverses two subregions.

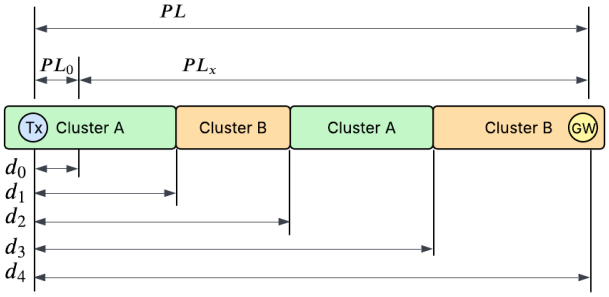

Fig. 3. A more complicated case where subregions are traversed multiple times.

3) Estimating η

The simple scenario presented in Fig. 2 has two segments (each segment representing a traversed subregion) and the PL_x is simply:

$$PL_x = 10\eta_A \log\left(\frac{d_1}{d_0}\right) + 10\eta_B \log\left(\frac{d_2}{d_1}\right) \tag{7}$$

A more complicated case is shown in Fig. 3 where there are four segments with repeating clusters. In In this case PL_x becomes:

$$\begin{aligned} PL_x =& 10\eta_A \log\left(\frac{d_1}{d_0}\right) + 10\eta_B \log\left(\frac{d_2}{d_1}\right) + \cdots \\ & \cdots 10\eta_A \log\left(\frac{d_3}{d_2}\right) + 10\eta_B \log\left(\frac{d_4}{d_3}\right) \\ =& 10\eta_A \log\left(\frac{d_1 d_3}{d_0 d_2}\right) + 10\eta_B \log\left(\frac{d_2 d_4}{d_1 d_3}\right) \end{aligned} \tag{8}$$

Eq. 9 is the general formulation for PL_x that has n segments with M clusters in $C = \{c_1, c_2, ..., c_M\}$. $D_1, D_2, \ldots, D_M$ are distance ratio features computed for each transmission by tracing the radio link from the transmitter to the gateway using Bresenham's line algorithm and analyzing how it traverses different geographical clusters. For each segment where the cluster changes, the distance ratio is

updated relative to a reference distance d_0, which is then set to the current distance within the same cluster.

$$\begin{aligned} PL_x =& 10\eta_{c\in C}\log\left(\frac{d_1}{d_0}\right) + 10\eta_{c\in C}\log\left(\frac{d_2}{d_1}\right) + \cdots \\ & \cdots + 10\eta_{c\in C}\log\left(\frac{d_n}{d_{n-1}}\right) \\ =& 10\eta_{c_1}\log\left(D_1\right) + 10\eta_{c_2}\log\left(D_2\right) + \cdots \\ & \cdots + 10\eta_{c_m}\log\left(D_M\right) \end{aligned} \tag{9}$$

Once the distance ratios are calculated, estimating the path loss exponent for each subregion becomes straightforward. Equation (9) can be rewritten as a linear regression model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \epsilon$, where ϵ is the residual vector. Given a training dataset of size p, the values of $\boldsymbol{y}$, $\boldsymbol{\beta}$, and $\boldsymbol{X}$ are:

$$\boldsymbol{y} = \begin{bmatrix} PL_X^{(1)} \\ PL_X^{(2)} \\ \vdots \\ PL_X^{(p)} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \eta_{c_1} \\ \eta_{c_2} \\ \vdots \\ \eta_{c_M} \end{bmatrix}$$

$$\boldsymbol{X} = \begin{bmatrix} 10\log(D_1^{(1)}) & 10\log(D_2^{(1)}) & \cdots & 10\log(D_M^{(1)}) \\ 10\log(D_1^{(2)}) & 10\log(D_2^{(2)}) & \cdots & 10\log(D_M^{(2)}) \\ \vdots & \vdots & \ddots & \vdots \\ 10\log(D_1^{(p)}) & 10\log(D_2^{(p)}) & \cdots & 10\log(D_M^{(p)}) \end{bmatrix}$$

Finally the estimates for the path loss exponents ($\hat{\boldsymbol{\beta}}$) are obtained using ordinary least squares (11).

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T y} \tag{11}$$

This segmented formulation of PL_x enables simultaneous estimation of the path loss exponent for each subregion. Notably, this approach not only eliminates the requirement for co-locating the gateway and transmitter within the same subregion but also removes the need for transmitters to be located within the subregion itself. As long as radio links between the gateway and transmitters traverse the subregion, its path loss exponent can be estimated.

4) Validation

Referring to Fig. 2b and Fig. 3, the total path loss PL is the sum of two path loss components: PL_0 and PL_x, where PL_x represents the path loss from Pr_0 to P_r. PL_x can be directly computed since P_r is explicitly available in the dataset, and Pr_0 is the received power at the designated reference distance d_0. In the dataset used, $d_0 = 27.39$ m and its corresponding $Pr_0 = -85.87$ dBm. On the other hand, PL_0 requires knowledge of the transmit power P_t, which is not provided in the dataset. This study focuses exclusively on estimating PL_x. Therefore, the ground truth values used for the validation process are given by:

$$PL_x = Pr_0 - P_r \tag{12}$$

The metric used for comparing the performances is the mean absolute error (MAE). Cross-validation was performed by training and validating the model using a randomly sampled 80-20 train-test split, repeating this process 30 times. The average MAE value was then recorded. This procedure was carried out for different number of clusters ($M \in \{1, 2, 3, 4\}$). To maintain cluster distribution integrity, stratified sampling was also employed.

IV. Results and Discussion

A. Prediction accuracy

The relationship between the number of clusters and path loss estimation accuracy is shown in Fig. 4. A single cluster ($M = 1$) treats the entire study area as a homogeneous propagation region, which is equivalent to the standard log-distance path loss model. As observed, adding clusters improves path loss estimation accuracy but with $M > 4$, the MAE increases. In this study, optimal number of clusters is $M = 4$.

Fig. 5 compares the performance of GeoSeg with the standard log-distance path loss model across all clusters. For this comparison the GeoSeg model that uses four clusters ($M = 4$) was used. GeoSeg demonstrates improved estimation accuracy for all clusters except cluster 2, but the performance difference in cluster 2 is less pronounced compared to other clusters. To evaluate whether the estimation errors of the two models are significantly different, hypothesis testing was conducted. Wilcoxon signed-rank test was employed and the computed p-value ($p << 0.05$) indicates a significant difference.

A direct comparison of prediction accuracy between GeoSeg and other per-link methods—specifically SateLoc and the per-link Okumura-Hata model—was not performed in this study. These methods require data collection techniques or parameters that are either incompatible with or missing from the dataset used. For instance, SateLoc requires transceivers to be co-located within the same land cover type to estimate path loss exponents, which does not align with the current deployment and dataset. Similarly, the per-link Okumura-Hata model needs system parameters that were not available in the dataset. Using estimated values or

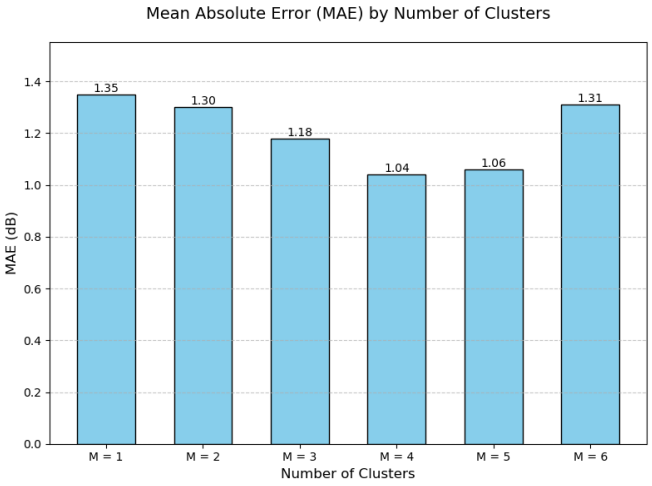

Fig. 4. Relationship between the number of clusters and path loss prediction accuracy.

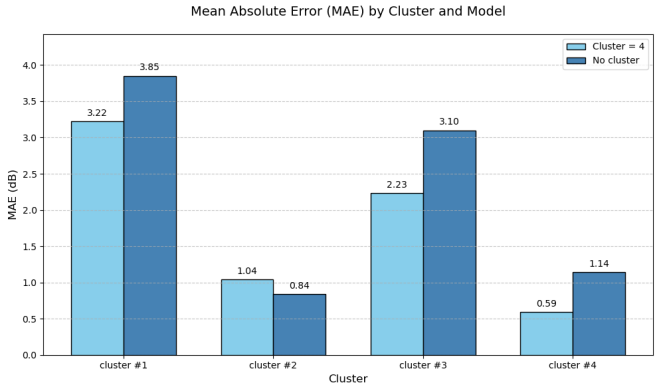

Fig. 5. Relationship between the number of clusters and path loss prediction accuracy.

adjusting the methods to fit the available data would reduce the reliability of the comparison. Therefore, GeoSeg was evaluated only against the standard long-distance path loss model. To enable a more thorough comparison, future work should collect data following the methodologies of these other studies and include the necessary system parameters.

V. Conclusion

It was shown that the diversity of land covers in an LP-WAN propagation environment necessitates a more granular approach to path loss modeling. Previous studies have used per-link modeling and geographical clustering techniques to improve the path loss estimation accuracy compared to regional models. This study developed a method that combines these two techniques to create an enhanced log-distance path loss model. Using real-world data, the proposed method achieved a 23% improvement in the MAE of path loss estimates compared to the standard log-distance path loss model.

References

[1] "State of IoT 2023: Number of connected IoT devices growing 16

[2] "LoRa PHY — Semtech — semtech.com," https://www.semtech.com/lora/what-is-lora, [Accessed 06-Jun-2023].

[3] H. Linka, M. Rademacher, O. G. Aliu, and K. Jonas, *Path loss models for low-power wide-area networks: Experimental results using LoRa*. VDE Verlag, 2018.

[4] M. Rademacher, H. Linka, T. Horstmann, and M. Henze, "Path loss in urban lora networks: A large-scale measurement study," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, 2021, pp. 1–6.

[5] H. Vo, V. Hoang Long Nguyen, V. L. Tran, F. Ferrero, F.-Y. Lee, and M.-H. Tsai, "Advance path loss model for distance estimation using lorawan network's received signal strength indicator (rssi)," *IEEE Access*, vol. 12, pp. 83 205–83 216, 2024.

[6] G. M. Bianco, R. Giuliano, G. Marrocco, F. Mazzenga, and A. Mejia-Aguilar, "Lora system for search and rescue: Path-loss models and procedures in mountain scenarios," *IEEE Internet of Things Journal*, vol. 8, no. 3, pp. 1985–1999, 2021.

[7] R. Ballestrin, J. F. Feijó, M. Feldman, and I. Müller, "Exploring machine learning techniques for path loss prediction in lora networks," in *2024 19th International Symposium on Wireless Communication Systems (ISWCS)*, 2024, pp. 1–6.

[8] Y. Onykiienko, P. Popovych, R. Yaroshenko, A. Mitsukova, A. Beldyagina, and Y. Makarenko, "Using rssi data for lora network path loss modeling," in *2022 IEEE 41st International Conference on Electronics and Nanotechnology (ELNANO)*, 2022, pp. 576–580.

[9] S. Demetri, M. Zúñiga, G. P. Picco, F. Kuipers, L. Bruzzone, and T. Telkamp, "Automated estimation of link quality for lora: A remote sensing approach," in *Proceedings of the 18th International Conference on Information Processing in Sensor Networks*, 2019, pp. 145–156.

[10] M. Aernouts, R. Berkvens, K. Van Vlaenderen, and M. Weyn, "Sigfox and lorawan datasets for fingerprint localization in large urban and rural areas," *Data*, vol. 3, no. 2, 2018. [Online]. Available: https://www.mdpi.com/2306-5729/3/2/13

[11] Y. Lin, W. Dong, Y. Gao, and T. Gu, "Sateloc: A virtual fingerprinting approach to outdoor lora localization using satellite images," *ACM Transactions on Sensor Networks (TOSN)*, vol. 17, no. 4, pp. 1–28, 2021.

[12] L. Liu, Y. Yao, Z. Cao, and M. Zhang, "Deeplora: Learning accurate path loss model for long distance links in lpwan," in *INFOCOM*, 2021.

[13] T. Nagao and T. Hayashi, "Geographical clustering of path loss modeling for wireless emulation in various environments," in *2022 16th European Conference on Antennas and Propagation (EuCAP)*, 2022, pp. 1–5.

[14] ——, "Study on radio propagation prediction by machine learning using urban structure maps," in *2020 14th European Conference on Antennas and Propagation (EuCAP)*. IEEE, 2020, pp. 1–5.

[15] T. Imai, K. Kitao, and M. Inomata, "Radio propagation prediction model using convolutional neural networks by deep learning," in *2019 13th European Conference on Antennas and Propagation (EuCAP)*, 2019, pp. 1–5.

[16] T. Nagao and T. Hayashi, "Study on radio propagation prediction by machine learning using urban structure maps," in *2020 14th European Conference on Antennas and Propagation (EuCAP)*, 2020, pp. 1–5.

[17] ——, "A study on urban structure map extraction for radio propagation prediction using xgboost," in *2021 15th European Conference on Antennas and Propagation (EuCAP)*. IEEE, 2021, pp. 1–5.

[18] S. A. R. Zekavat, "Channel modeling and its impact on localization," in *Handbook of Position Location: Theory, Practice, and Advances*, S. A. R. Zekavat and R. M. Buehrer, Eds. New Jersey: John Wiley Sons, 2019, ch. 4, pp. 107–141.

[19] G. Callebaut and L. Van der Perre, "Characterization of lora point-to-point path loss: Measurement campaigns and modeling considering censored data," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1910–1918, 2020.

[20] F. Porikli, "Clustering variable length sequences by eigenvector decomposition using hmm," in *Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshops, SSPR 2004 and SPR 2004, Lisbon, Portugal, August 18-20, 2004. Proceedings*. Springer, 2004, pp. 352–360.

[21] P. Smyth, "Clustering sequences with hidden markov models," *Advances in neural information processing systems*, vol. 9, 1996.