Enhanced RF-based 3D UAV Outdoor Geolocation: from Trilateration to Machine Learning Approaches

Mariem Belhor, Anne Savard, Anthony Fleury, Patrick Sondi, Valeria Loscri

Enhanced RF-based 3D UAV Outdoor Geolocation: from Trilateration to Machine Learning Approaches

Mariem Belhor*, Anne Savard*, Anthony Fleury*, Patrick Sondi* and Valeria Loscri†

* IMT Nord Europe, Institut Mines Télécom, Centre for Digital Systems, F-59653 Villeneuve d'Ascq, France
† Inria Center at the University of Lille
Email: {firstname.lastname}@imt-nord-europe.fr, valeria.loscri@inria.fr,

***Abstract*—Recently the use of Unmanned Aerial Vehicles (UAVs), commonly referred to as drones, has exploded in several domains, leading to potential security issues. As such, estimating the exact position of those eventual malicious drones has become of crucial interest. However, computing an accurate and precise geolocation of these drones, especially in outdoor environments, remains challenging. This paper focuses on outdoor 3-dimensional (3D) drone geolocation techniques based on Radio Frequency (RF) signals. We first present a RF-based 3D drone geolocation dataset, and then apply and compare various geolocation techniques, ranging from geometrical-based to machine learning-based methods. We further propose a new hybrid method blending the two above categories of geolocation techniques, that achieves an average 3D error of the order of 11.7 meters within a search volume of about $520 \times 560 \times 115\,\mathrm{m}^3$, significantly below the one achieved with geometrical-based techniques, and with a reduced computational complexity compared to the regular machine-learning based techniques.**

***Index Terms*—3D drone geolocation, RF-based geolocation, Outdoor geolocation, Machine learning, UAV geolocation**

I. Introduction

The increasing misuse of UAVs has brought attention to critical issues in cybersecurity and national defense. As these UAVs become more accessible to the public while enhancing its technological capabilities, they are increasingly exploited for malicious purposes, including espionage, unauthorized data collection, and hacking into sensitive spaces. These drones, equipped with cameras, microphones, or hacking software, can be deployed stealthily, presenting unparalleled risks to the safety of critical infrastructures, military bases, and private organizations, especially when they violate protected airspace to collect confidential information or destabilize communications and data systems via cyber-attacks [1], [2].

As such, there is a pressing need for reliable systems able of detecting and geolocalizing an unintended presence of a drone in different environments and airspace. In the remaining, *localization* refers to the process of identifying a broader area, such as a city or country, where the object under consideration is active. On the other hand, *geolocation* is a more precise process, pinpointing the exact geographical coordinates of an object, device, or individual [3].

This work has been supported by ANR ASTRID DEPOSIA, an initiative for AI in defense and security operated by the French National Research Agency (ANR) and sponsored by the French Defence Innovation Agency (AID).

Two main paths regarding drone geolocalization can be found in the literature: using Global Positioning System (GPS) or exploiting RF-signals.

The use of GPS for drone geolocation poses challenges in cybersecurity because of signal interference, generating hence GPS noise which makes an accurate location tracking impossible [4]. As an alternative, RF signal-based geolocation, for which positions are determined by exploiting RF signals from cellular networks, as well as from Wi-Fi or Bluetooth beacons, is widely used. The authors of [5] assert that employing RF signals for drone localization offers many advantages that compensate for the GPS limitations, particularly in terms of security and susceptibility to interference, such as noise and signal disruption. GPS, while widely used, has vulnerabilities including signal jamming, spoofing, and reduced effectiveness in dense urban areas or indoors, where signals can be obstructed [3], [4]. In contrast, RF signals provide a more robust alternative for precise localization, maintaining functionality in areas where GPS signals are compromised, ensuring greater operational security and reliability for drones, especially in critical applications [5].

Due to the enhanced resilience and reliability of RF-based systems over traditional GPS-based ones, we focus on RF-based localization for the remainder of the paper. Within this context, multiple classifications of RF-based geolocation approaches have been identified and explored in the most recent literature [5], [6]. One prevalent classification splits geolocation methods into two primary categories: classical and Machine Learning (ML)-based approaches.

A. Classical geolocation methods

Classical geolocation methods encompass two different strategies: either measurement-based methods or exploiting typological information derived from the measurements. In the remaining of this paper, we focus only on measurement-based methods due to their higher accuracy [7], [8]. The latter can be further categorized given the type of exploited measurements. i) *Time-based measurement methods*, where distances are computed thanks either to absolute time information such as *Time of Arrival* (ToA) or *Time of Flight* (ToF), which require precise synchronization and are vulnerable to timing errors; or relative time-based metrics, such as *Time Difference of Arrival* (TDoA), which increases the robustness to timing errors while

reducing the latency and the energy consumption compared to ToA and ToF methods. [7].
ii) *Angle-based measurement methods*, where distances are computed thanks to *Angle of Arrival* (AoA), also referred to as *Direction of Arrival* (DoA). The latter require strategically placed sensors to capture signals and determine the transmitter's position through triangulation [9] and are often combined with other techniques such as TDoA or *Received Signal Strength Indicator* (RSSI) to enhance accuracy and reliability in determining positions [9], [10]. Compared to time-based methods, they are typically less susceptible to synchronization errors, but may require more complex angular measurement devices.
iii) *Channel-based measurement methods*, where distances are computed based on RF channel characteristics, such as for instance RSSI [11]. As the RSSI decreases as the distance between the transmitter and receiver increases, it can be used to estimate how far the receiver is [12]. Moreover, channel-based methods are often more precise and robust to environmental variations and noise disruptions, such as multi-path or signal weakening caused by weather or physical barriers, than time-based or angle-based methods [13].

Despite the extensive available literature on classical approaches for drone localization, to the best of our knowledge, an accurate three-dimensional (3D) outdoor geolocation is still lacking. Indeed, achieving precise geolocation in 3D outdoor environments, poses significant technical challenges, such as accurately processing signal interference in diverse environments as well as being robust to arbitrary environmental variations. Furthermore, the complexity of 3D geolocation as compared to 2D, provides an additional level of challenge since it requires to also estimate the altitude, an especially difficult task in urban areas with high-rise buildings and diverse topographies. These challenges, as well as the low level of accuracy achieved by classical methods, pave the transition to ML-based geolocalization techniques, that can handle such limitations and outperform classical methods, as we will see in the remaining of this paper. Nonetheless, these aforementioned classical methods are still often used and improved [7], [8] and serve as baseline approaches to compare with more recent ML-based ones.

B. ML-based geolocation methods

Recently, ML-based techniques for both detecting and geolocating malicious drones have gain momentum, especially due to their high achieved accuracy level, their ability to process and analyse varied datasets and their flexibility to operate in unfamiliar scenarios without needing significant system adjustments [14]. These ML-based methods can be further divided into two principal research paths:
i) determining the general area where the drone is currently active. In such a case, the localization problem under study reduces to a classification task, often concerning broader zones rather than precise geographical positions [15], [16]. Most of the existing literature on 3D drone localization focuses currently on such problems.
ii) determining the precise geolocation, targeting the exact pinpointing of a drone's geographical position. Currently, most of the research carried out on exact geolocation addresses 2D geolocation problem, mainly in indoor setups. In contrast, RF-based outdoor 3D drone geolocation is a relatively unexplored area. The closest works to our are [17] and [18]. However, [17] primarily focuses on 2D geolocation using extended Kalman filtering with RSSI and TDoA. To the best of our knowledge, very few works focus on RF-based outdoor 3D drone geolocation. For instance, [18] addresses this issue by modeling the distances estimated through RSSI as noisy approximations of the actual ones, hence no real RSSI measurements are exploited.

C. Main contributions

In this paper, we focus exclusively on RF-based outdoor 3D drone geolocation. Elaborating on the lack of both outdoor 3D drone geolocation techniques and available datasets for RF-based geolocation, our main contributions are three-fold:
i) *Deriving a new dataset for outdoor RF-Signal based UAV geolocation:* following the lack of available open-source dataset for outdoor RF-based geolocation, we propose to adapt an existing dataset, that was originally intended for RF signal data collection with a drone in an outdoor environment. We implement a significant preprocessing to adapt it for outdoor geolocation purposes. This dataset, which can be used to compare various RF-based geolocation approaches, will be made available via GitHub upon publication.
ii) Comparing various techniques from the literature: We here exploit both a classical 3D channel-based measurement method, as well as well-known ML-based regression methods achieving good performance in various regression problems. Note that these methods have not been used so far for the 3D drone geolocation problem. All these methods are then used as benchmarks for our numerical simulations.
iii) *Proposing a new enhanced geolocation technique:* We propose a new blending of geometrical-based and ML-based methods, yielding hence an hybrid approach that enhances the accuracy of the 3D position estimation.
To summarize, this paper is among the first ones to consider RF-based 3D drone outdoor geolocation. We here provide a set of possible techniques to obtain a precise 3D position estimation of the drone while exploiting only RSSI measurements. All these techniques are then compared one to another both in terms of mean 3D error, mean absolute error and root mean square error, allowing to compare the precision of the proposed solutions, as well as in terms of CPU, to assess the computational complexity of the aforementioned methods.
Remarkably, the ML-based methods significantly enhance the accuracy of the estimated 3D position compared to the considered geometrical-based method, but some of the considered ML-based methods may be more computationally complex. Finally, our proposed hybrid method allows to benefit from the very high accuracy achieved by the ML-based methods while reducing, for almost all considered ML-based methods, the complexity compared to their non-hybrid counterpart. Our

overall contribution is summerized in Fig. 1.

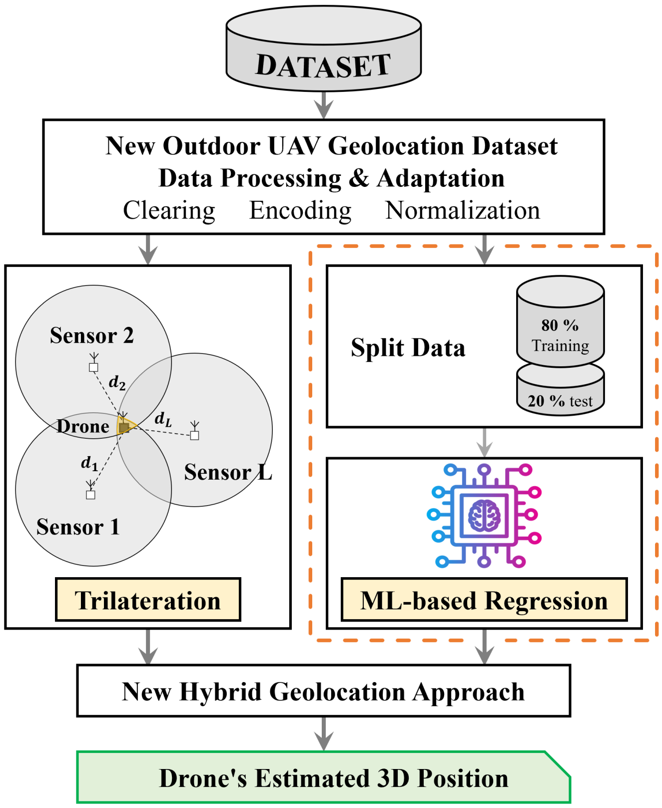


Fig. 1: Summary of our main contribution: After a prior dataset adaptation, we provide a set of possible techniques, encompassing both a geometrical approach, i.e. trilateration, ML-based methods as well as a new proposed hybrid approach blending the two latter, to estimate the 3D drone position in an outdoor environment based solely on RSSI measurements.

The remainder of this paper is organized as follows: Section 2 delves into the system model. Section 3 presents the proposed geolocation dataset. Section 4 provides an overview of our conducted experiments and presents a thorough analysis of the findings. Section 5 concludes the paper.

II. Geolocation System Model

The aim of this paper is to determine the exact drone 3D position in outdoor environments exploiting RF signals. The system under study is composed by R receivers, whose positions are fixed and known as $\mathrm{Pos}_r = (X_r, Y_r, Z_r)$, $r \in \{1, \ldots, R\}$, as well as by a drone, whose position $\mathrm{Pos} = (X, Y, Z)$ needs to be estimated solely given the R received RSSI and does not require the knowledge of the previous drone position, i.e. there is no drone tracking process in our proposed method. Since we seek for a 3D geolocation, we need at least four RSSI measurements to estimate the drone position; the latter is assumed to be met in the following and will be insured via the preprocessing of the considered dataset as detailed in Section III. For instance, the scenario considered in Fig. 2 involves an outdoor environment where four receivers are deployed to estimate the drone 3D position.

Notations: The following notations will be used: RSSI_r denotes the RSSI measured by the r-th ground-based receiver, $\widehat{\mathrm{Pos}} = (\widehat{X}, \widehat{Y}, \widehat{Z})$ denotes the estimated position of the drone. $||x, y||$ represents the Euclidean distance in a 3D space between the points x and y.

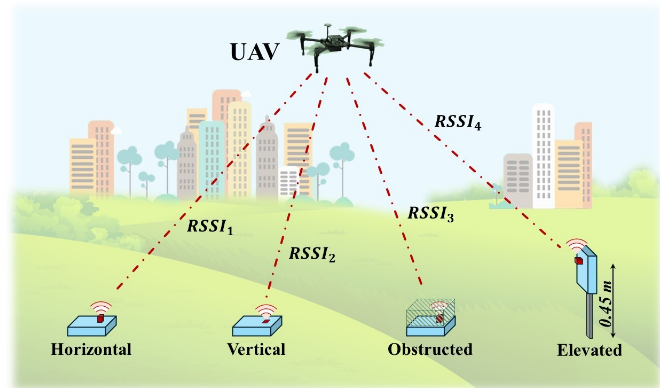


Fig. 2: Environment setup for outdoor drone geolocation based on four RSSI measurements

A. Overall problem formulation

The goal of this paper is, given at least four RSSI measured by the ground-based receivers, to accurately estimate the exact 3D position of the drone. As such, we aim at minimizing the empirical average error between the actual and estimated drone position over the entire dataset, which is expressed as

$$\begin{aligned} \textbf{(OP)} \quad \min_{\widehat{\mathrm{Pos}} \in \mathbb{R}^3} \quad & \frac{1}{n} \sum_{i=1}^{n} \left|\left|\mathrm{Pos}(i), \widehat{\mathrm{Pos}}(i)\right|\right|^2 \\ \text{s.t. } & \mathrm{Lat}_{\min} \leq \widehat{X} \leq \mathrm{Lat}_{\max}, \\ & \mathrm{Lon}_{\min} \leq \widehat{Y} \leq \mathrm{Lon}_{\max}, \\ & \mathrm{Alt}_{\min} \leq \widehat{Z} \leq \mathrm{Alt}_{\max}, \end{aligned}$$

where n denotes the number of entries in the dataset, each one being associated with four RSSI measurements. The three constraints restrict the drone position estimation to take place in a restricted area in order to enhance its accuracy. The specific dimension of this search space will be presented in Section IV.

B. Algorithms for drone geolocation

We now present the two considered 3D RSSI-based geolocation methods, namely a classical one based on distances extracted from the RSSI and a ML-based one. We then blend the two latter to propose a new hybrid approach, where the baseline position estimate achieved with the geometrical approach is fine-tuned using the ML-driven regression approach.

a) RSSI-based 3D geolocation using trilateration: This 3D geolocation method belongs to the classical method framework, that solely relies on the computations of distances from the received RSSI. Under trilateration, the 3D drone position is computed given its distance to at least four fixed points in space [12], which are here our four deployed ground-based receivers. Let in the following denote by d_r the Euclidean distance between the drone and the r-th receiver. Note that these distances can be easily estimated from the RSSI, by exploiting the commonly used logarithmic law describing the signal attenuation as a function of distance as

$$\mathrm{RSSI}(d) = \mathrm{RSSI}_0 - 10\alpha \log_{10}\left(\frac{d}{d_0}\right), \tag{1}$$

where RSSI_0 is the signal strength at a reference distance of d_0 m and α is the path loss exponent. Hence, the distance between any transmitter and receiver pair writes as

$$d = f_{\mathrm{RSSI}\to d}(\mathrm{RSSI}) \triangleq d_0 10^{-(\mathrm{RSSI}-\mathrm{RSSI}_0)/(10\alpha)}. \tag{2}$$

Thus, since the estimated drone position is computed as in [12] based on the four estimated distances d_r, the goal is to minimize the error, averaged over the four RSSI measurements, between the 3D estimated distances d_r and the real one, i.e. d_r, which writes as

$$\min \quad \frac{1}{4}\sum_{r=1}^{4}\left(f_{\mathrm{RSSI}\to d}(\mathrm{RSSI_r}) - d_r\right)^2. \tag{3}$$

b) ML-driven regression for RSSI-based 3D geolocation: The main limitation of the previous method comes from the need to compute distances. Instead, we propose to exploit ML-driven regression that can learn complex and non-linear relationships between the RSSI and the drone coordinates, eliminating thus the need for direct distance computation. ML-based methods learn to minimize a so-called *loss function*, which in our case corresponds to the mean squared error, i.e. the objective function of the optimization problem **(OP)**. Note that we here consider supervised learning approaches, where during the training phase, one can access the real 3D drone position.

c) Hybrid RSSI-based 3D geolocation blending trilateration and ML-driven regression: In order to improve the accuracy and reliability of the drone 3D geolocation, we propose to mix both approaches: as such, the trilateration method provides an initial geometry-based estimate, while the ML technique offers an adaptable approach that can account for more complex factors, as well as being more robust to the inherent presence of noise throughout the acquisition process. We here propose to minimize the average gap between the real position of the drone Pos and a weighted combination of positions computed by the trilateration, denoted as $\mathrm{Pos}_{\mathrm{Tri}}$, and by the ML-based method, denoted as $\mathrm{Pos}_{\mathrm{ML}}$:

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\left((1-\beta)\mathrm{Pos}_{\mathrm{Tri}}(i) + \beta\mathrm{Pos}_{\mathrm{ML}}(i)\right), \mathrm{Pos}(i)\right\|^2, \tag{4}$$

where the weighting parameter $\beta \in [0,1]$ allows to switch between a trilateration-driven or a ML-driven approach.

III. New dataset for outdoor UAV RF signal-based geolocation

We now detail our proposed dataset for drone RF-based geolocation. We start by presenting an open-source dataset containing RF measurements, originally designed in a setup where a drone collects some data transmitted by four sensors that were deployed on the ground. Due to the lack of open-source datasets for RF-based geolocation in outdoor, we propose to exploit this dataset and to adapt it to the geolocation problem under study.

A. Open-source dataset adaptation

To the best of our knowledge, there is no open-source available dataset[1] for outdoor drone 3D geolocation based on RF-measurements. Nonetheless, we were able to find an open-source dataset used for environment data collection, where real measurements of RSSI where performed outdoor in a rather different setup than the one under study. In fact, the goal was not to geolocate the drone but to ensure some connectivity in terms of received RSSI between four sensors deployed on the ground and a drone that was collecting all the data [19]. Despite the difference in both objectives, i.e. geolocating a drone vs. ensuring connectivity with a drone, the measurements are highly relevant to our setup, where we also need four ground-deployed transmitters and a drone flying in an outdoor area. Indeed, this dataset encompasses RSSI measurements performed in the 2.4 GHz frequency band, as well as some crucial data to validate the accuracy of our proposed geolocation method, such as the 3D drone position. Note that this dataset also include many other parameters related to transmitters and receiver, enriching the evaluation and analysis context; as well as different deployment configurations, especially with the presence of obstacle on the sensors, enabling thus a robust and comprehensive evaluation of the geolocation methods.

As such, based on channel reciprocity, experimentally validated under the used ZigBee protocol [20], we propose to adapt this dataset by inverting the role of the ground-deployed sensors and the drone: the ground-deployed sensors will now play the role of the receivers, instead of the originally transmitters, whereas the drone will now be the transmitter. The new obtained dataset will hence be particularly relevant for outdoor 3D drone geolocation based on RF-measurements and will be used to validate the previously presented geolocation techniques.

B. Dataset Adaptation and preprocessing

In order to build a dataset for RF-based 3D drone geolocation, we propose to only keep the following relevant features for each of the 121,503 records in the original dataset: timestamp, test location (road, grassy field or hilly field), 3D coordinates of the sensors, RSSI measurement and 3D drone coordinates. After this first step, we propose to group all acquisitions given a specific time duration of one second.

[1] Original Dataset: https://datadryad.org/stash/dataset/doi:10.25349/D9KS3W

As such, all received RSSI measurements during this time window are averaged for each of the receivers, yielding 4 RSSI values each one second, the same holding also for the drone 3D position. Note that in some cases, we may lack one or more RSSI values, such entries will be removed from our new geolocation dataset, since an accurate 3D position can only be achieved with at least 4 RSSI measurements. This meticulous process yields a dataset of 5,093 records, where each entry is composed by an averaged timestamp, 4 averaged RSSI measurements, the averaged 3D drone position and the 3D coordinates of the sensors. This systematic approach guarantees efficient and accurate data management for subsequent drone signal analysis.

IV. Experimental Analysis

A. Implementation and comparison metrics

Regarding the ML-based geolocation, we propose to exploit several well-known algorithms from the literature, that were shown to achieve high accuracy for regression problems, the general framework in which the considered geolocation problem falls into [21]. More precisely, we consider *K-Nearest Neighbors* (KNN), *Random Forest Regression* (RFR), *XGBoost* (XGBST), and *Gradient Boosting* (GB). To evaluate those ML algorithms, 80% of our dataset is used for the training, whereas the remaining 20% are used for the testing.

The ML-based methods are then compared to the trilateration method presented in Section II-B as well as to our proposed hybrid method that blends all the aforementioned ML-based method with the trilateration one as presented in the same Section. The hybrid methods are thereafter denoted as HKNN, HRFR, HXGBST and HGB.
All geolocation methods are implemented using Python, within the Visual Studio Code IDE, and run on a 12th Gen Intel® Core™ i7-12700H CPU computer under Ubuntu 22.04.3 LTS with 32 GB of RAM.
In order to implement the trilateration method, the path loss exponent was set to $\alpha \in \{2, 3, 4\}$ corresponding respectively to the measurements performed on the road ($\alpha = 2$), the grassy field ($\alpha = 3$) and the hilly one ($\alpha = 4$). Further, the reference distance d_0 was set to 1m, for which the corresponding reference RSSI was recorded as $\mathrm{RSSI}_0 = -70$ dBm.
Regarding hybrid methods, the weighting parameter was set to $\beta = 0.9$, putting hence more emphasis on the ML-based method than on the trilateration one. This parameter was obtained using an exhaustive search method as described in [22]. All the other parameters used for the ML-based geolocation methods were obtained using the GridSearchCV technique [23] on our dataset, and are summarized in Table I.

Comparison metrics:
In the following, we will compare each of the geolocation method in terms of *Mean 3D error* (ξ_{3D}), *Mean Absolute Error* (MAE) and *Root Mean Square Error* (RMSE) expressed below. All these metrics rely on the gap between the estimated position of the drone $(\widehat{X}, \widehat{Y}, \widehat{Z})$ and its true position (X, Y, Z) and are average over the number of entries in the dataset n.

ML algorithms	Parameters	Best Value
KNN/ HKNN	Optimal number of neighbors	3
XGBST/ HXGBST	Learning rate	0.3
	Max depth	9
	N estimators	200
	Subsample	1.0
RFR/ HRFR	Max depth	50
	Max features	Sqrt
	Min samples leaf	1
	Min samples split	2
	N estimators	50
GB/ HGB	Learning rate	0.1
	Max depth	5
	N estimators	200
	Subsample	0.9

TABLE I: Best Parameters for ML Algorithms

To simplify the presentation of the aforementioned metrics, let $\delta_X = X - \widehat{X}$, $\delta_Y = Y - \widehat{Y}$ and $\delta_Z = Z - \widehat{Z}$ denote the error in terms of the latitude, longitude and altitude respectively. Note that since the latitude and longitude coordinates are expressed in degrees and that the altitude is expressed in meter, we need to convert the latitude and longitude coordinates in meters prior to any error computation, the latter been performed thanks to the Haversine formula [24]. Exploiting these notations, the comparison metrics write as

$$\xi_{3D} = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\delta_X^2(i) + \delta_Y^2(i) + \delta_Z^2(i)},$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\delta_X(i)|+|\delta_Y(i)|+|\delta_Z(i)|,$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\delta_X^2(i)+\delta_Y^2(i)+\delta_Z^2(i)}.$$

B. Experimental Results and Analysis

All the measurements were performed in a specific geographical area near Santa Barbara, California, USA. The area is geographically constrained as follows: the latitude ranges from $\mathrm{Lat}_{\min} = 34.41259703$ to $\mathrm{Lat}_{\max} = 34.41728644$, the longitude from $\mathrm{Lon}_{\min} = -119.88242$ to $\mathrm{Lon}_{\max} = -119.8763092$, and the altitude from $\mathrm{Alt}_{\min} = 9.258136171$ m to $\mathrm{Alt}_{\min} = 123.8962822$ m, yielding hence a search are of about $520 \times 560 \times 115\,\mathrm{m}^3$.

Table II compares all the presented 3D geolocation methods in terms of mean 3D error, MAE and RMSE, as well as in terms of CPU time in ms. First, one can note that the trilateration method yields the poorest results in terms of mean 3D error, MAE and RMSE, while being rather computationally complex. Regarding the standard ML-approaches from the literature, one can note that Gradient Boosting yields the best performance in terms of RMSE and mean 3D error, but at the cost of computational complexity. Regarding MAE, XGBoost yields the best performance, but it is less efficient in terms of mean 3D error and RMSE than Gradient Boosting. Finally, KNN is the less computationally complex method, but

Method	MAE (m)	RMSE (m)	ξ_{3D} (m)	CPU (ms)
Trilateration	53.69	62.76	46.62	18,835.45
KNN	1.08	2.69	23.36	**5,712.70**
XGBST	**0.16**	0.65	26.33	25,170.90
RFR	1.02	2.82	22.92	10,010.00
GB	0.19	**0.52**	**11.89**	31,449.60
HKNN	1.07	2.72	22.81	**4,364.81**
HXGBST	**0.19**	0.52	27.68	15,413.98
HRFR	1.02	2.78	20.30	19,919.15
HGB	**0.19**	**0.52**	**11.73**	20,067.01

TABLE II: Comparison of geolocation methods

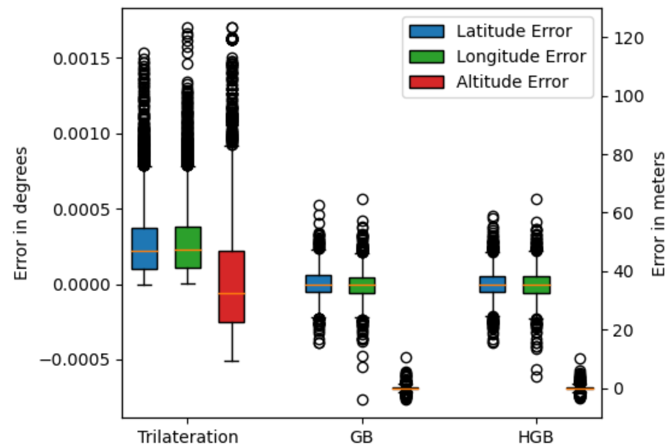


Fig. 3: Comparison of geolocation error per coordinate for the classical method, i.e. trilateration, the best ML one, i.e. Gradient Boosting, and the best proposed hybrid one, i.e. Hybrid Gradient Boosting.

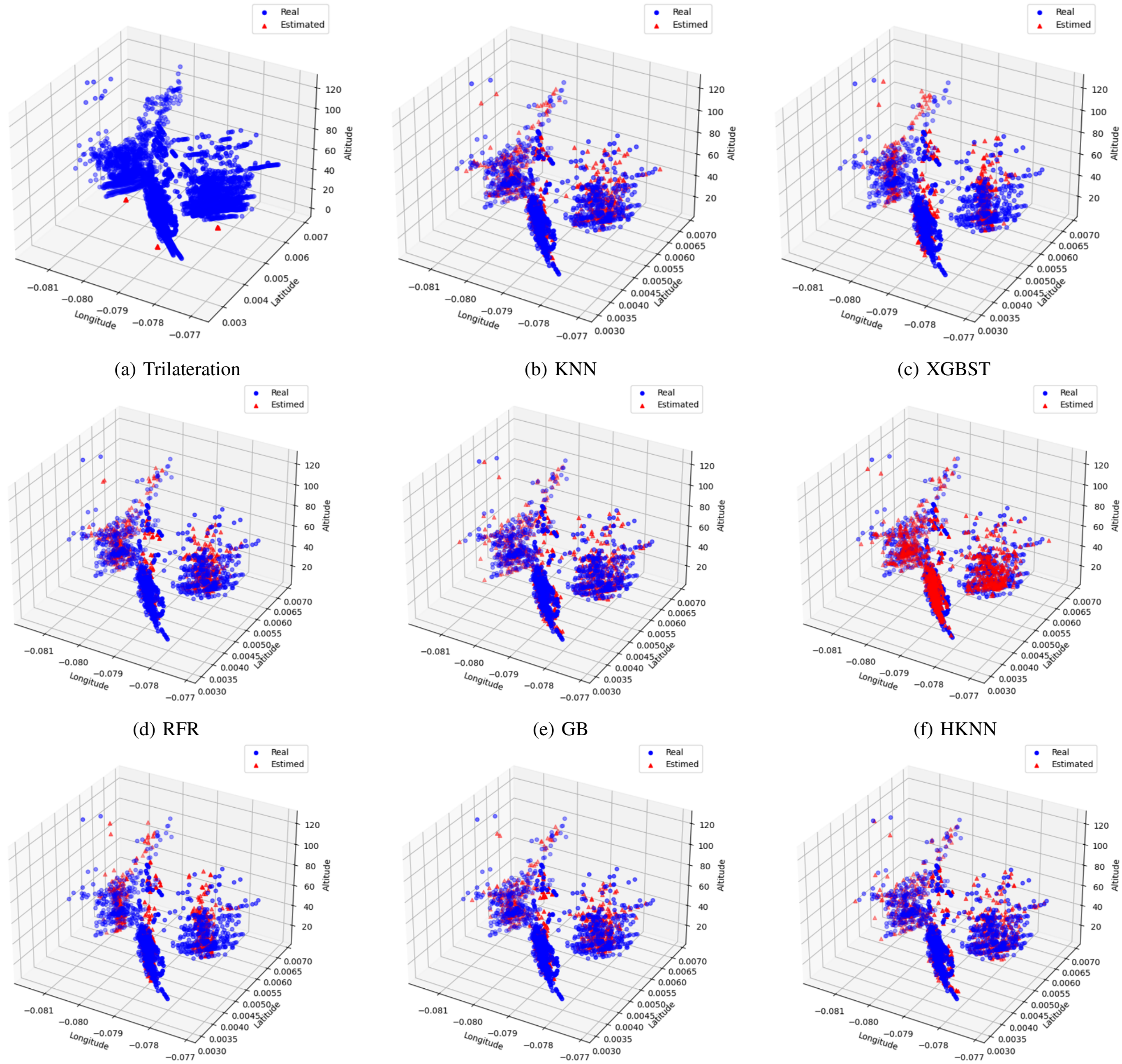
it is outperformed by the other methods regarding the error-based metrics. At last, all the above conclusions on ML-based methods carry over to our proposed hybrid ones: HKNN is the less computationally complex method; whereas HGB yields the best performance in terms of 3D position estimation, at the cost of its computationally complexity. Nonetheless, the hybrid HGB method achieves the same MAE and RMSE than its standard counterpart, while reducing both the mean 3D error and the computational complexity, showing hence the strength of our proposed hybrid approach. To conclude, our proposed HGB method yields a mean 3D error of 11.7 m in a search space of about $520 \times 560 \times 115\,\mathrm{m}^3$, which is very promising for practical applications, with a moderate computational complexity.

Let us now compare in more details the trilateration, the Gradient Boosting and our proposed Hybrid Gradient Boosting algorithms, as they yield respectively the best results for the classical, ML-based and hybrid approaches. Fig. 3 compares the average error of the three above methods obtained over the longitude, latitude and altitude component respectively. Note that for the longitude and latitude, the error is expressed in degree as indicated on the left-side of Fig. 3, whereas for the altitude, it is expressed in meter as indicated on the right-side of Fig. 3. One can see that ML-based geolocation methods clearly outperform the trilateration one with an average error on each of the coordinate close to zero with a reduced variance, which can also be observed in Fig. 4, that depicts for each of the nine considered methods the real drone position, indicated with blue circles, as well as the estimate one, indicated with red squares.

Fig. 5 compares the histograms of the averaged 3D error, given in meter, achieved by all our proposed hybrid ML-based methods, namely HKNN, XGBST, HRFR, and HGB. First, we can observe a significant variation in the spread of the average 3D error among the hybrid geolocation methods. For instance, HKNN and HRFR show a larger range of error magnitudes, indicating a less consistent geolocation performance compared to XGBST and HGB, for which the majority of the errors are of lower magnitude. Notably, HGB demonstrates a higher frequency of smaller errors that represents a more accurate geolocation performance compared to the other methods. Furthermore, a significant majority of the errors across all methods fall within the first 20-meter interval, indicating a strong concentration of lower magnitude errors.

V. Conclusion

In this paper, we focused on a RF-based outdoor drone 3D geolocation, for which we first compared a geometrical-based approach with several regression-relevant machine learning-based approaches from the literature on a dataset containing real RSSI measurements. The machine learning-based approaches highly improved the accuracy of the drone position estimation compared to a solely distance-based one. We further proposed an hybrid approach, where an initial drone position, estimated with the geometrical approach, is then fine-tuned with a machine learning method. The latter achieves the best performance in terms of accurate 3D position, with a mean 3D error of 11.7 m out of a search area of about $520 \times 560 \times 115\mathrm{m}^3$, while being less computationally complex than its standard counterpart. Such an accurate precision highlights the benefits of our proposed hybrid RF-based 3D geolocation approach for practical applications. These simulations results will also be consolidated via a new dataset of real measurements that will be acquired in the upcoming months, and that is specifically intended for RF-based outdoor 3D drone geolocation. Given the moderate size of the currently available datasets, we were not able to investigate the potential of deep learning based techniques, such as deep neural networks (DNN) or recurrent neural network (RNN), which will be leveraged through the aforementioned dataset under acquisition. Also, in order to achieve a highly accurate 3D position with a moderate computational complexity, a two



(a) Trilateration (b) KNN (c) XGBST

(d) RFR (e) GB (f) HKNN

(g) HXGBST (h) HRFR (i) HGB

Fig. 4: Comparison between the real and estimated positions of the drone for the classical method, i.e. trilateration, the four ML ones, i.e. KNN, XGBST, RFR, GB, and the four proposed hybrid ones, i.e. HKNN, HXGBST, HRFR, HGB

steps approach, where we first pinpoint the broader area of the drone activity and subsequently determine its exact location within that region, will be investigated.

References

[1] W. H. Adnan and M. F. Khamis, "Drone use in military and civilian application: Risk to national security," *Journal of Media and Inf. Warfare (JMIW)*, vol. 15, no. 1, pp. 60–70, 2022.

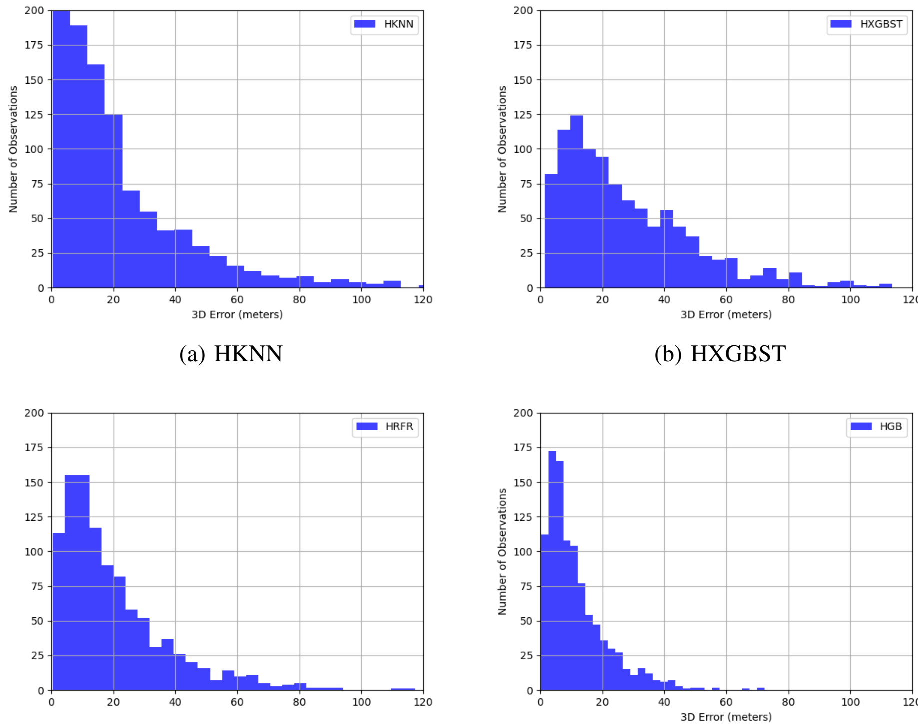
[2] W. BinSaeedan, A. Aldawsari, L. Alhussain, L. Alrushud, and L. Alfawzan, "Security Challenges for UAV Systems Communications: Potential Attacks and Countermeasures," in *Unmanned Aerial Vehicles Applications: Challenges and Trends*. Springer, 2023, pp. 269–288.

[3] P. Scerri, R. Glinton, S. Owens, D. Scerri, and K. Sycara, "Geolocation of rf emitters by many uavs," in *AIAA Infotech@ Aerospace 2007 Conference and Exhibit*, 2007, p. 2858.

[4] S. Zekavat, R. M. Buehrer, G. D. Durgin, L. Lovisolo, Z. Wang, S. T. Goh, and A. Ghasemi, "An overview on position location: Past, present, future," *International J. of Wireless Inf. Netw.*, vol. 28, pp. 45–76, 2021.

[5] A. Gupta and X. Fernando, "Simultaneous localization and mapping (slam) and data fusion in unmanned aerial vehicles: Recent advances and challenges," *Drones*, vol. 6, no. 4, p. 85, 2022.

[6] C. Isaia and M. P. Michaelides, "A Review of Wireless Positioning



(a) HKNN (b) HXGBST (c) HRFR (d) HGB

Fig. 5: Comparison of error distributions for hybrid geolocation methods

Techniques and Technologies: From Smart Sensors to 6G," *Signals*, vol. 4, no. 1, pp. 90–136, 2023.

[7] Y. Sun, H. Fu, S. Abeywickrama, L. Jayasinghe, C. Yuen, and J. Chen, "Drone classification and localization using micro-doppler signature with low-frequency signal," in *IEEE ICCS*, 2018, pp. 413–417.

[8] H. Chen, L. Cao, and Y. Yue, "TDOA/AOA Hybrid Localization Based on Improved Dandelion Optimization Algorithm for Mobile Location Estimation Under NLOS Simulation Environment," *Wireless Personal Commun.*, pp. 1–26, 2023.

[9] S. M. M. Dehghan and H. Moradi, "A new approach for Simultaneous Localization of UAV and RF Sources (SLUS)," in *IEEE ICUAS*, 2014, pp. 744–749.

[10] J. Yin, Q. Wan, S. Yang, and K. Ho, "A simple and accurate TDOA-AOA localization method using two stations," *IEEE Signal Process. Lett.*, vol. 23, no. 1, pp. 144–148, 2015.

[11] T. Wattananavin, K. Sengchuai, N. Jindapetch, and A. Booranawong, "A comparative study of RSSI-based localization methods: RSSI variation caused by human presence and movement," *Sensing and Imaging*, vol. 21, pp. 1–20, 2020.

[12] P. Kumar, L. Reddy, and S. Varma, "Distance measurement and error estimation scheme for RSSI based localization in wireless sensor networks," in *IEEE WCSN*, 2009, pp. 1–4.

[13] B. Xiang, F. Yan, Y. Zhu, T. Wu, W. Xia, J. Pang, W. Liu, G. Heng, and L. Shen, "UAV assisted localization scheme of WSNs using RSSI and CSI information," in *IEEE ICCC*, 2020, pp. 718–722.

[14] C. Sandamini, M. W. P. Maduranga, V. Tilwari, J. Yahaya, F. Qamar, Q. N. Nguyen, and S. R. A. Ibrahim, "A Review of Indoor Positioning Systems for UAV Localization with Machine Learning Algorithms," *Electronics*, vol. 12, no. 7, p. 1533, 2023.

[15] S. Abeywickrama, L. Jayasinghe, H. Fu, S. Nissanka, and C. Yuen, "RF-based direction finding of UAVs using DNN," in *IEEE ICCS*, 2018, pp. 157–161.

[16] D. Shorten, A. Williamson, S. Srivastava, and J. C. Murray, "Localisation of drone controllers from RF signals using a deep learning approach," in *Proceedings of the international conference on pattern recognition and artificial intelligence*, 2018, pp. 89–97.

[17] U. Bhattacherjee, E. Ozturk, O. Ozdemir, I. Guvenc, M. L. Sichitiu, and H. Dai, "Experimental study of outdoor UAV localization and tracking using passive RF sensing," in *Proceedings of the 15th ACM Workshop on Wireless Network Testbeds, Experimental evaluation & CHaracterization*, 2022, pp. 31–38.

[18] G. Afifi and Y. Gadallah, "Unmanned Aerial Vehicles 3-D Autonomous Outdoor Localization: A Deep Learning Approach," in *IEEE WCNC*, 2022, pp. 908–913.

[19] M. Nekrasov, R. Allen, and E. Belding, "Aerial Measurements from Outdoor 2.4 GHz 802.15. 4 Network," *UC Santa Barbara: Santa Barbara, CA, USA*, 2019.

[20] J. Higuera, E. Kartsakli, J. L. Valenzuela, L. Alonso, A. Laya, R. Martínez, and A. Aguilar, "Experimental study of Bluetooth, ZigBee and IEEE 802.15. 4 technologies on board high-speed trains," in *IEEE 75th VTC Spring*, 2012, pp. 1–5.

[21] M. Fernández-Delgado, M. S. Sirsat, E. Cernadas, S. Alawadi, S. Barro, and M. Febrero-Bande, "An extensive experimental survey of regression methods," *Neural Netw.*, vol. 111, pp. 11–34, 2019.

[22] J. Nievergelt, "Exhaustive search, combinatorial optimization and enumeration: Exploring the potential of raw computing power," in *International Conference on Current Trends in Theory and Practice of Computer Science*. Springer, 2000, pp. 18–35.

[23] K. Jolly, *Machine learning with scikit-learn quick start guide: classification, regression, and clustering techniques in Python*. Packt Publishing Ltd, 2018.

[24] E. Maria, E. Budiman, M. Taruk *et al.*, "Measure distance locating nearest public facilities using Haversine and Euclidean Methods," in *Journal of Physics: Conference Series*, vol. 1450, no. 1. IOP Publishing, 2020, p. 012080.