Journal of Edge Computing, 2025, Vol. **4**, Iss. 2, pp. 263–287 https://doi.org/10.55056/jec.1201

Acoustic Doppler localisation and tracking in 3D space with retardation correction

Andrii V. Riabko[1], Tetiana A. Vakaliuk[2,3,4,5], Oksana V. Zaika[1] and Roman P. Kukharchuk[1]

[1]*Oleksandr Dovzhenko Hlukhiv National Pedagogical University, 24 Kyivska Str., Hlukhiv, 41400, Ukraine*
[2]*Zhytomyr Polytechnic State University, 103 Chudnivsyka Str., Zhytomyr, 10005, Ukraine*
[3]*Institute for Digitalisation of Education of the NAES of Ukraine, 9 M. Berlynskoho Str., Kyiv, 04060, Ukraine*
[4]*Kryvyi Rih State Pedagogical University, 54 Universytetskyi Ave., Kryvyi Rih, 50086, Ukraine*
[5]*Academy of Cognitive and Natural Sciences, 54 Universytetskyi Ave., Kryvyi Rih, 50086, Ukraine*

Abstract. This study extends Doppler-based acoustic localisation from 2D to a complete 3D framework for UAV tracking. A key methodological contribution is the explicit correction for the retardation effect (signal propagation delay), solved numerically using the Newton-Raphson method. This correction proved essential, as the average localisation error was reduced from approximately 50 m (without correction) to about 15 m (with correction), confirming the practical necessity of including retardation in 3D acoustic models. The VarPro method was implemented to exploit the separable structure of the least-squares problem. Under stabilised conditions (with velocity fixed to resolve the V/f identifiability issue), the method demonstrated convergence and provided reasonable trajectory estimates with mean trajectory errors below 5%. At the same time, the diagnostic analysis revealed fundamental limitations of the VarPro method. First, the strong correlation between source velocity (V) and source frequency (f) makes unconstrained optimisation unstable and prone to divergence. Second, even in stabilised runs, the simplified analytical trajectory models (straight-line or their 7-parameter extension) are structurally inadequate for representing stochastic UAV motion, thereby forcing the optimiser toward non-physical solutions. These findings suggest that while VarPro can be applied successfully to simplified scenarios with constrained parameters, it is not a suitable general solution for localizing UAVs with complex, random trajectories. Future research will therefore focus on recursive state estimation methods such as the extended Kalman filter, which are better suited for dynamic stochastic motion and time-varying source frequencies.

Keywords: localisation, acoustic localisation, Doppler effect, retardation correction, 3D model, VarPro, UAV, dynamic source frequency, nonlinear optimisation

1. Introduction

The trajectory of a moving acoustic emitter can be estimated based on the measured Doppler shift, even when utilising a solitary sensor. Our current investigation advances this concept by considering an acoustic sensor network comprising multiple synchronised receiving nodes with precisely fixed spatial coordinates. Our foundational premise is that each node can reliably detect and quantitatively assess the Doppler frequency generated by a passing source. We utilise a parameterised model to accurately describe the motion trajectory, focusing on the identification of these

0000-0001-7728-6498 (A. V. Riabko); 0000-0001-6825-4697 (T. A. Vakaliuk); 0000-0002-8479-9408 (O. V. Zaika); 0000-0002-7588-7406 (R. P. Kukharchuk)
ryabkoav@gnpu.edu.ua (A. V. Riabko); tetianavakaliuk@gmail.com (T. A. Vakaliuk); ksuwazaika@gmail.com (O. V. Zaika); kyxap4yk1@ukr.net (R. P. Kukharchuk)
https://gnpu.edu.ua/members/andrij-ryabko/ (A. V. Riabko); https://acnsci.org/vakaliuk (T. A. Vakaliuk); https://gnpu.edu.ua/members/oksana-zayika/ (O. V. Zaika); https://gnpu.edu.ua/members/roman-kuharchuk/ (R. P. Kukharchuk)

model parameters from which the complete source path is subsequently derived. We specifically prioritise scenarios involving a minimal number of sensors – such as one or three nodes – a practical consideration essential for cost-effective system deployment.

The field of localising mobile acoustic sources via Doppler measurements has a rich history, with significant activity peaking in the early 1990s. Despite this prior attention, substantial aspects within this domain remain inadequately addressed. Given the past decade's revolution in wireless sensor network technologies and the emergence of state-of-the-art numerical optimisation algorithms, we assert a compelling necessity to reinvigorate acoustic Doppler research. Our work represents a significant evolutionary step by extending existing planar (2D) models into a comprehensive three-dimensional (3D) tracking system, a capability that is critically required for modern airspace monitoring.

The cornerstone of our refined methodology is the explicit incorporation of the retardation effect (signal propagation delay). Neglecting this time delay, which is non-trivial in the acoustic domain (unlike in radar systems), results in a pronounced systematic estimation bias. This 3D extension permits the application of a more comprehensive eight-parameter motion model for the high-fidelity tracking of fast-moving sources, notably Unmanned Aerial Vehicles (UAVs), whose trajectories are not constrained to a single horizontal plane.

The trajectory estimation problem proves to be inherently nonlinear in its motion parameters, necessitating a broad spectrum of numerical approaches to achieve an optimal estimate in various senses. The most common methodology involves defining parameters by minimising some variation of a least squares criterion.

A critical complication specific to the acoustic domain is retardation (signal delay). This effect arises because, by the time the sound reaches the sensors, the source has already moved to a new position due to the finite speed of sound. In radar systems, this effect is typically disregarded, as the propagation speed of electromagnetic waves is many orders of magnitude greater than the speed of most practical sources. In the acoustic case, however, ignoring retardation can result in significant estimation bias, depending on the geometry and source speed.

Approximate retardation-free models can typically be parameterised with only two nonlinear parameters (instead of three or four in the full 2D model), such as x and y coordinates on the source trajectory. This reduced parameter dimension facilitates an exhaustive grid search for the least squares minimum, thereby avoiding otherwise problematic local minima. In fact, by completely disregarding retardation, an exact analytical solution based on a deterministic model and three frequency measurements can sometimes be found by calculating the roots of a third-order polynomial. Our research, however, fundamentally addresses the more complex yet essential complete 3D model, which requires explicit retardation correction to achieve the required accuracy.

There are several approaches to acoustic source localisation using the Doppler effect. One technique is to match the measured frequency sequence with precomputed Doppler profiles, which makes it possible to incorporate complex retardation models even on platforms with limited computational power [2].

Sensor networks provide richer information by comparing frequency differences across multiple sensors, which enables the reconstruction of more complex source trajectories. However, system design must strike a balance between accuracy and cost: too many sensors increase complexity and expenses, while a minimal configuration can still provide reliable results.

A key factor is the correction of retardation, since it directly affects the accuracy of both position and velocity estimation. Efficient optimisation algorithms allow for practical implementation on lightweight sensor nodes, making such systems suitable

for monitoring ground vehicles, aircraft, and unmanned aerial systems.

In our work, we aim to extend the existing two-dimensional models to a complete three-dimensional framework that accounts for source altitude. This improvement, combined with retardation correction, will significantly enhance the generality and precision of Doppler-based localisation in real-world scenarios, particularly for UAVs.

2. Theoretical background

Previous research by the authors investigated MEMS microphone arrays for sound localisation, demonstrating that differential algorithms improve noise suppression and that circular arrays provide balanced omnidirectional coverage [13]. Further studies addressed UAV localisation using MEMS microphones, showing how frequency-dependent performance and specialised algorithms enable the transformation of linear arrays into effective uni- or bidirectional systems [14].

Jekateryńczuk and Piotrowski [7] conducted a detailed review and classification of current sound source localisation and detection methods, highlighting their wide-ranging industrial applications. They acknowledge the key role of artificial intelligence (AI) in significantly advancing these algorithms recently, but argue that, despite this progress, there is a vital need for continuous research to develop new techniques that can improve the accuracy and reliability of existing methods.

Classic Localisation Methods rely on geometry and signal timing. Triangulation uses the angle of arrival (AoA) measured at two or more microphones. In contrast, trilateration (and multilateration with more sensors) determines the source location by calculating its distance from several microphones, primarily using the time difference of arrival (TDoA) parameter. Beamforming is a signal processing technique that uses a microphone array to enhance the sound signal from a specific direction, essentially "focusing" on the source [12].

AI methods (e.g., convolutional neural networks (CNNs) and recurrent neural networks (RNNs)) represent a more recent approach, utilising data-driven models to determine localisation or detect sound events. These models, often fed audio features like spectrograms or TDoA measurements, are superior to classic methods in complex acoustic environments with noise and reverberation because they learn features from datasets rather than relying on mathematically defined propagation models (figure 1).

Jalayer, Jalayer and Baniasadi [6] provide a comprehensive review of sound source localisation (SSL) in robotics, with a strong emphasis on recent deep learning approaches. The authors first revisit classical techniques such as TDOA, beamforming, SRP, and subspace analysis, and then shift to modern methods including CNNs, CRNNs, and attention-based architectures. They also stress the importance of data quality and training strategies as foundations for DL-based SSL, while categorising studies by robot type and application domain. The paper further discusses key challenges-such as robustness to complex environments, handling multiple sources, and practical constraints in robotic systems-offering a roadmap for the development of more reliable and explainable SSL methods in future robotic applications.

Ochmann and Piscoya [12] thoroughly investigate the mathematical, physical, and technical properties of moving acoustic sources, analysing various point and line sources along different 3D trajectories. They detail the theory to explicitly derive the frequency spectrum and associated Doppler shift for these sources. Crucially, the research applies these findings to develop an effective source localisation method – akin to an equivalent source method in the time domain – for important applications, such as detecting noise from trains and aeroplanes.

Srivastava [16] explores how supervised learning methods can be used to estimate acoustic room parameters and localise sound sources. He emphasises the idea of

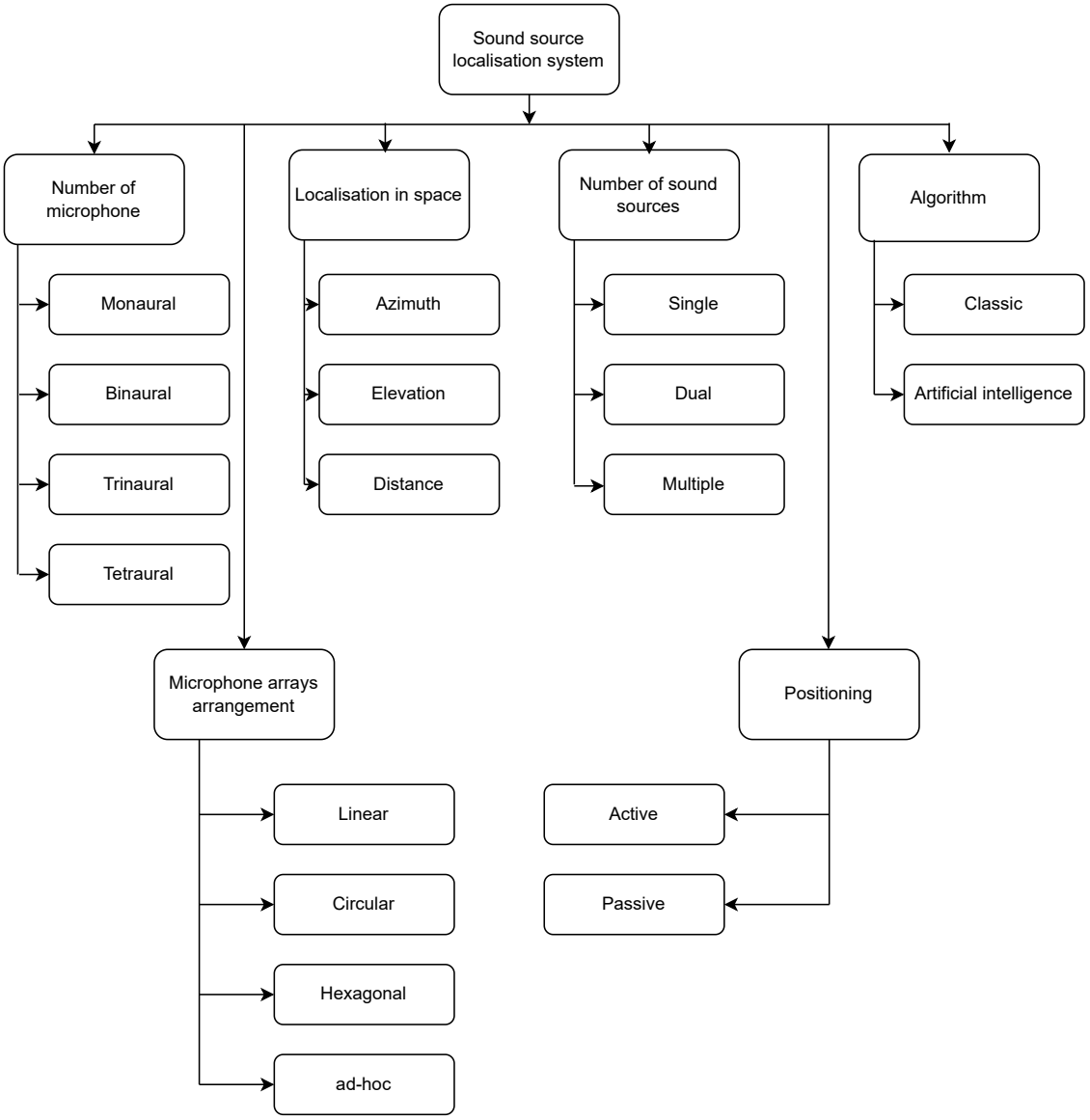

Figure 1: Classification of sound source localisation systems (according to Jekateryńczuk and Piotrowski [7].

training models on simulated rather than real data. He demonstrates that enhancing the realism of room acoustic simulators results in improved generalization of the models when tested on real-world datasets.

Delabie et al. [3] present an indoor acoustic simulation framework that enables both ultrasonic and audible signalling for positioning and data-driven signal processing. They demonstrate that their improved Pyroomacoustics-based model, validated against real-world measurements, yields realistic results and can flexibly support various applications .

Kamada et al. [8] investigate the use of quadcopters with microphone arrays for locating survivors' voices in disaster environments. They apply PyRoomAcoustics for modelling complex acoustic conditions and use MUSIC-based algorithms for sound source localisation, showing that the simulations strongly correlate with real-world data and provide valuable insights for optimising SSL in rescue operations.

Tan et al. [17] propose a CNN-regression (CNN-R) model for sound source localisation that estimates both angle and distance using interaural phase difference features.

With datasets generated via PyRoomAcoustics and MIRD, their results show very high accuracy in both simulated and real environments, demonstrating strong potential for real-life applications.

Wang and Cavallaro [18] introduce DSACNN22, a multi-feature fusion network designed to estimate the direction of sound arrival in complex underground tunnels. Using PyroomAcoustics-simulated data, they show that their method improves demonstrability accuracy in enhanced environments, which is particularly useful for burrow detection in beneficial exploration robots.

Elelu, Le and Le [4] present a study on preparing multichannel audio datasets for detecting and localising construction equipment sounds to prevent collision hazards. Using Pyroomacoustics, the authors simulate overlapping and non-overlapping sound scenarios in 3D space, generating over 5,000 audio samples from limited single-sound recordings. The resulting dataset provides a valuable resource for training neural networks designed to enhance safety on construction sites.

Saini and Peissig [15] introduce HARP, a large-scale dataset of 7th-order ambisonic room impulse responses generated with the image source method. The dataset employs a 64-microphone configuration to capture responses directly in the spherical harmonics domain, enabling highly accurate spatial audio reproduction. With diverse room geometries, materials, and source-receiver distances, HARP provides a valuable resource for research in spatial audio, machine learning, and immersive sound applications such as localisation and reverberation modelling.

Lai, Ma and Wang [9] present a system that transforms monaural audio into immersive stereo sound using binaural synthesis and Pyroomacoustics. Their approach combines dynamic spatial perception with architectural acoustic responses, enabling users to model environments such as concert halls and customise motion trajectories and surround modes. The result is a flexible tool for creating highly realistic and engaging auditory experiences.

Li et al. [10] propose HearLoc, a system for indoor sound source localisation (ISSL) that achieves full 3D positioning using only a compact 10 cm microphone array. Their method exploits time differences of arrival (TDOAs) from both direct and reflected signals to boost accuracy and dimensional capability. Experimental results show significant gains in localisation precision and efficiency compared to traditional angle-of-arrival approaches, making HearLoc a promising solution for small-scale IoT and smart environment applications.

Hornikx et al. [5] review the state of open research software in room acoustics, highlighting both historical progress and the growing role of machine learning methods. They note that while computational tools for impulse and frequency response prediction have advanced, limited software accessibility hinders reproducibility and collaboration. The paper emphasises the measurable benefits of open research software and outlines its potential to accelerate future developments in room acoustics.

Wang and Cavallaro [18] investigate the problem of sound source localisation from flying drones, which is complicated by intense motor noise and drone movement. They propose a deep-learning framework that combines noise suppression with multichannel localisation, showing that their TFS-DNN method performs better than existing approaches even in very low signal-to-noise conditions.

Ahmed, Ho and Wang [1] explore the problem of localising a moving source using frequency measurements affected by the Doppler effect. They propose constrained optimisation methods that allow efficient closed-form and noise-resilient solutions, demonstrating accuracy close to the Cramer-Rao lower bound in simulations.

Chen and Lu [2] investigate the problem of tracking fast-moving sound sources in the air, where traditional TDOA and DOA methods struggle due to Doppler effects and time-varying distances. They introduce a particle filtering framework with relative

Doppler stretch, showing in simulations that it outperforms existing approaches and works directly with raw acoustic signals.

Ma and Zhang [11] discuss the crucial issue of the Doppler effect when using the time-domain ROSI beamforming method for identifying rotating sound sources. They specifically note that the Doppler effect can create "ghost contributions" at incorrect scan points, which complicates accurate source identification. The authors propose that using an array with rotational symmetry can significantly help eliminate these false signals at specific sideband frequencies, thereby improving the beamforming technique.

The reviewed literature demonstrates that both classical and AI-based approaches have made significant progress in sound source localisation, particularly in applications ranging from robotics to room acoustics and UAV systems. However, studies also highlight that the Doppler effect in moving sources remains a critical challenge, underlining the importance of developing accurate microphone array methods that explicitly account for this phenomenon, which directly connects to the focus of our study.

3. Research methods

This formula (1) accounts for the movement of both the receiver (V_r) and the source (V_s), where c is the speed of sound and f is the emitted frequency:

$$f' = f\left(\frac{c \pm V_r}{c \mp V_s}\right) \tag{1}$$

The following formula (2) is typically used in array processing when the receiving microphones are static. V_s is the velocity of the source, and θ is the angle of arrival relative to the velocity vector:

$$f' = \frac{f}{1 \mp \frac{V_s}{c}\cos\theta} = \frac{f}{1 \mp M\cos\theta} \tag{2}$$

The foundation of our Doppler model is the source-observer propagation delay, denoted as $t_c(t)$. This parameter is central to developing a comprehensive description of the relationship between source motion and the observed Doppler shift, effective in both the time and frequency domains. By our definition, the propagation delay $t_c(t)$ is relative to the observed signal $y(t)$, and it is assumed to satisfy the following key nonlinear relationship, which is crucial for retardation correction:

$$t_c(t) = sol_{\tau>0}\left\{\tau = \frac{1}{c}Z(t-\tau)\right\} \tag{3}$$

where c is the assumed known (or accurately estimated) speed of sound, and $z(t)$ is the source-observer distance at time t. The solvability of this equation (i.e., finding $\tau > 0$) directly hinges on how accurately the source motion is modelled (its trajectory). We utilise this propagation delay as the anchor for our extended 3D motion model, enabling us to accurately localise objects moving at significant altitudes, such as UAVs.

Assuming the source emits a sinusoidal signal with a constant frequency f, the observed signal $y(t)$ at the receiver is modelled as the sum of the acoustic signal $h(t)$ and background noise $e(t)$:

$$y(t) = h(t) + e(t) \tag{4}$$

$$h(t) \sim \cos[2\pi \cdot (t - t_c(t) + \varphi)] \tag{5}$$

where φ is the constant phase, and the proportionality relation ($\sim$) signifies that the sinusoid's amplitude is disregarded in this work. As observed, the propagation delay $t_c(t)$ is directly integrated into the phase argument, thereby enforcing retardation correction.

The observed signal frequency $y_f(t)$ is logically derived as the time derivative of the cosine argument, yielding the following expression:

$$y_f(t) = h_f(t) + \varepsilon(t) \tag{6}$$

$$h_f(t) = f \cdot [1 - \dot{t}_c(t)] \tag{7}$$

where $\dot{t}_c(t)$ denotes the time derivative of the propagation delay, and $\varepsilon(t)$ is the frequency measurement error. This expression for $h_f(t)$ constitutes the universal Doppler model that remains valid for our subsequent extension into 3D space.

In contrast to the complex circular trajectory, which may necessitate numerical solutions, we will initially focus on an extended 3D model of constant-velocity straight-line motion. This model serves as a direct, yet essential, generalisation of the single-sensor model and is crucial for the accurate tracking of sources, such as UAVs, that may move linearly but at significant altitudes. Even when assuming straight-line motion, utilising a sensor network in 3D space requires an expanded parameter set. This model necessitates seven parameters for a complete description:

$$\theta = [f, V, \alpha, \beta, p_x^0, p_y^0, p_z^0]^T \tag{8}$$

where f is the emitted frequency (the separable parameter), V is the constant velocity, $p^0 = [p_x^0, p_y^0, p_z^0]^T$ is the initial 3D position (or the closest point of approach), and α and β are two angles (e.g., azimuth and elevation) that define the direction vector of motion u.

The source position $p(t)$ at any time t is described vectorially:

$$p(t) = p^0 + V \cdot t \cdot u \tag{9}$$

Here, u is the unit direction vector defined by the angles α and β. The distance to the i-th sensor s_i is then determined by:

$$z_i(t) = \| \, p(t) - s_i \, \| \tag{10}$$

In our study, s has three coordinates in 3D space, which describe its fixed location:

$$s = [x_s, y_s, z_s]^T \tag{11}$$

By preserving the straight-line trajectory, this 3D model retains the potential for an analytical closed-form solution for the propagation delay $t_{c,i}(t)$, similar to the 2D case, offering significant computational advantages. This, in turn, enables us to utilise highly efficient methods, such as VarPro (variable projection), for the robust determination of the seven parameters θ.

As previously outlined, source motion estimation invariably relies on parameterised motion models. The inherent flexibility permitted in these models is directly proportional to the number of observers or sensors available within the system. For instance, with only a single sensor, the source's motion is restricted to the simplest scenario: constant speed along a straight line. Undeniably, any violation of this fundamental assumption by the actual source trajectory will result in varying degrees of estimation error (bias).

In this foundational 2D case, the set of unknown model parameters includes: the emitted frequency f, the passage speed V, the passage distance d, and the passage

instant t_0. For computational simplification, we utilise normalised parameters: the normalised speed $\bar{v} = v/c$ and the normalised distance $\bar{d} = d/c$. Consequently, the complete vector of unknown parameters is collected as:

$$\theta = [f, \bar{V}, \bar{d}, t_0]^T \tag{12}$$

While effective for demonstrating retardation correction in a planar setting, this model cannot meet the demands of 3D tracking required for targets like UAVs. This limitation necessitates our shift to the multi-sensor approach, which incorporates a spatial model that explicitly accounts for altitude and direction.

We consider the case where the acoustic source moves in 3D space with a constant velocity V along a straight-line trajectory. We assume the sensor is located at the origin $s_i = [0, 0, 0]^T$ for simplified notation.

The source trajectory $p(t)$ is a straight line. The closest point of approach (CPA) distance H is now the 3D spatial distance from the sensor to the trajectory. The passage instant t_0 is the time of passage through the CPA.

The source-to-sensor distance $z(t)$ at time t is the 3D generalization of $z(t) = c\sqrt{(d^2 + (t - t_0)^2 V^2)}$, with H replacing d:

$$z(t) = \sqrt{H^2 + (t - t_0)^2 V^2} = c\sqrt{\bar{H}^2 + (t - t_0)^2 \bar{V}^2} \tag{13}$$

where $\bar{H} = H/c$ is the normalised CPA distance, and $\bar{V} = V/c$ is the normalised velocity (Mach number). Similar to the 2D case, the straight-line trajectory allows for a closed-form analytical solution for the propagation delay (3).

The solution to the retardation equation, expressed in terms of the normalised parameters $\bar{H}$ and $\bar{V}$:

$$t_c(t) = \frac{\sqrt{\bar{H}^2 + \bar{V}^2(t - t_0)^2 - \bar{H}^2\bar{V}^2} - \bar{V}^2(t - t_0)}{1 - \bar{V}^2} \tag{14}$$

Time derivative of delay $\dot{t}_c(t)$:

$$\begin{aligned} \dot{t}_c(t) &= \frac{1}{1 - \bar{V}^2}\left[\frac{\bar{V}^2(t - t_0)}{\sqrt{\bar{H}^2 + \bar{V}^2(t - t_0)^2 - \bar{H}^2\bar{V}^2}} - \bar{V}^2\right] \\ &= \frac{\bar{V}^2}{1 - \bar{V}^2}\left[\frac{t - t_0}{\sqrt{\bar{H}^2 + \bar{V}^2(t - t_0)^2 - \bar{H}^2\bar{V}^2}} - 1\right] \end{aligned} \tag{15}$$

Formulas (13) and (14) provide the necessary retardation correction for constant-velocity 3D motion.

Utilising a network comprising three or more sensors provides a significantly richer set of measurements, enabling us to employ a more flexible and general 3D motion model where the source can follow a circular trajectory of arbitrary radius (figure 2). This approach is essential for accurately tracking airborne sources (e.g., UAVs) whose movement is rarely confined to a single plane.

We extend the planar model into three-dimensional space, requiring the inclusion of extra parameters to describe the altitude and the orientation of the circular motion plane (12). Consequently, the full 3D circular motion model is described by eight unknown parameters:

$$\theta = [f, V, \alpha_0, \beta_0, p_x^0, p_y^0, p_z^0, g]^T \tag{16}$$

where f is the emitted frequency, V is the constant speed, $p^0 = [p_x^0, p_y^0, p_z^0]^T$ is the initial source position at time $t = 0$, α_0 and β_0 are the initial angles on the circular path, and

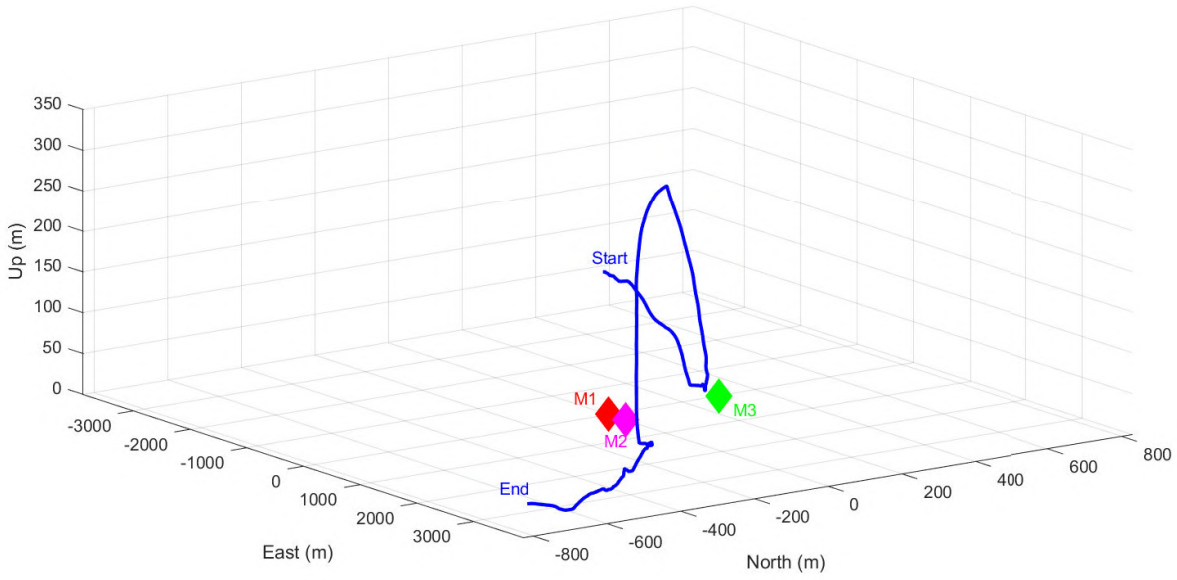

Figure 2: 3D motion model for acoustic localisation (base of distances between microphones L=300 m).

$g = 1/r$ is the inverse radius of curvature. Notably, $g = 0$ corresponds to the special case of straight-line motion in 3D space.

The source location $p(t)$ at any time t is determined by a complex geometrical transformation involving the rotation axis and the plane of motion. Generally, it is defined as:

$$p(t) = p_{center} + R(t)(p^0 - p_{center}) \tag{17}$$

where p_{center} is the center of the circle, and $R(t)$ is a rotation operator through an angle Vgt around the axis perpendicular to the plane of motion.

The operator $R(t)$ is a 3×3 rotation matrix essential for describing the complex kinematics of circular motion in 3D space, particularly for airborne targets. This operator defines how the initial position vector (relative to the circle's centre) transforms into the final vector after rotation. To define this operator, we must know the axis of rotation a and the angle of rotation ϕ. The axis of rotation a is a unit vector perpendicular to the plane containing the circle and is determined by the initial orientation angles α_0 and β_0. The rotation angle $\phi(t)$ is a scalar quantity calculated as the product of the angular velocity $\omega = Vg$ and the elapsed time t: $\phi(t) = Vgt$. All these elements combine into the matrix $R(\phi)$, which is most often written using Rodrigues' rotation formula:

$$R(\phi) = I + \sin(\phi)[a]_\times + (1 - \cos(\phi))[a]_\times^2 \tag{18}$$

where $[a]_\times$ is the skew-symmetric matrix constructed from the vector a to represent the vector cross product using matrix multiplication is defined as:

$$[a]_\times = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix} \tag{19}$$

Thus, $R(t)$ allows us to correctly model the trajectory $p(t)$ in an arbitrary plane within 3D space. The distance $z_i(t)$ and relative speed $\dot{z}_i(t)$ between the source and the i-th sensor s_i are given by:

$$z_i(t) = \|p(t) - s_i\| \tag{20}$$

$$\dot{z}_i(t) = \frac{(p(t) - s_i)^T \dot{p}(t)}{\|p(t) - s_i\|} \tag{21}$$

where, $\dot{p}(t)$ is the source velocity vector. Similar to the straight-line case, the propagation delay $t_{c,i}(t)$ cannot be expressed in closed analytical form for such a complex 3D trajectory. Therefore, for retardation correction, we must resort to a numerical solution of equation (3), and we propose using the Newton-Raphson method.

A suitable initial approximation for the Newton-Raphson iterations is $t_{c,i}(t) \approx \frac{1}{c} z_l(t)$. We define the function $F(t_c)$, omitting the sensor index i for notational convenience:

$$F(t_c) \triangleq \frac{t_c^2}{c^2} - \frac{z^2(t - t_c)}{c^2} \tag{22}$$

The positive solution to $F(t_c) = 0$ is mathematically equivalent to the retardation equation (3). The Newton-Raphson iterative process for determining the updated delay t_c^+ is:

$$t_c^+ = t_c - \frac{F(t_c)}{F'(t_c)'} \tag{23}$$

where $F'(t_c)$ is the derivative of $F(t_c)$ with respect to t_c. Convergence to machine precision is typically achieved in fewer than 10 iterations. Once $t_c(t)$ is determined, the derivative of the propagation delay, essential for calculating the Doppler shift (5), is computed using the chain rule:

$$\dot{t}_c(t) = \frac{\dot{z}(t - t_c(t))}{c + \dot{z}(t - t_c(t))} \tag{24}$$

This completes the crucial step that enables us to apply the Doppler model correctly in 3D space, fully accounting for the retardation effect.

4. Results

The unweighted least squares estimate for the parameter vector θ of our 3D model is generally expressed as:

$$\hat{\theta} = arg \min_{\theta} \sum_{i=1}^{M} \sum_{k=1}^{N} [y_{f,i}(t_k) - h_{f,i}(t_k; \theta)]^2 \tag{25}$$

Here, a time-discrete representation is used, where a set of N frequency measurements $y_{f,i}$ is provided from M different sensors s_i at distinct time instants t_k. To our knowledge, for this specific application, particularly with highly nonlinear models involving retardation correction, equation (25) can only be solved using *iterative numerical algorithms*. It is worth noting that this optimisation problem is not generally guaranteed to be free from local minima, a fact that significantly complicates the numerical approach. In this section, we propose a method for deriving a robust initial estimate of the parameters and review numerical algorithms tailored for the nonlinear least squares problem at hand.

For the 3D circular motion model ($g \neq 0$), developing direct analytical initial values is exceedingly complex. Therefore, it is proposed that the parameters of the network 3D model (our vector θ) be approximated by fitting them to a set of estimates derived individually from each sensor (single-sensor model). This approach allows us to use more reliable starting estimates for the subsequent global optimisation.

To ground our approach for the single sensor model (straight-line motion, $g \equiv 0$), we first consider the simpler, two-parameter, straight-line case. Assuming the signal frequency at infinity (asymptotic frequency) is f (figure 3).

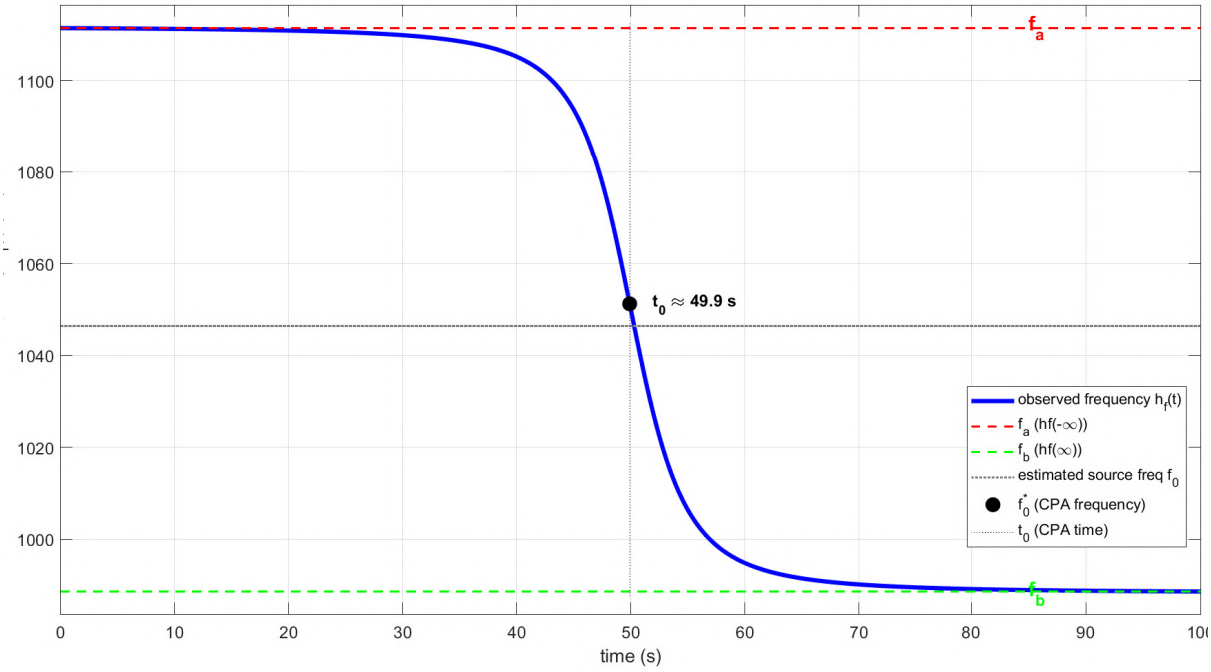

Figure 3: Key features extracted for the initial parameter estimate. The graph illustrates the time-varying observed frequency $h_f(t)$ (the Doppler curve) for the single-sensor straight-line motion case, which serves as the robust initialisation basis for the full 3D network model. The three key features extracted from the measurement data $y_f(t)$ are: 1) the asymptotic frequency before CPA, $f_a = h_f(-\infty)$; 2) the asymptotic frequency after CPA, $f_b = h_f(\infty)$; and 3) the slope extremum, $f_0^* = \dot{h}_f(t_0)$, which occurs at the Closest Point of approach (t_0).

The limiting frequencies observed before ($t \to -\infty$) and after ($t \to \infty$) the source passes are defined as:

$$f_a = \lim_{t\to-\infty} f(1 - \dot{t}_c(t)) = f\frac{1}{1-\bar{V}} \tag{26}$$

$$f_b = \lim_{t\to\infty} f(1 - \dot{t}_c(t)) = f\frac{1}{1+\bar{V}} \tag{27}$$

These boundary frequencies allow us to analytically isolate the normalised speed $\bar{V} = V/c$ and the emitted frequency f:

$$c \cdot \frac{f_a + f_b}{f_a - f_b} = c \cdot \frac{\frac{1}{1-\bar{V}} - \frac{1}{1+\bar{V}}}{\frac{1}{1-\bar{V}} + \frac{1}{1+\bar{V}}} = V \tag{28}$$

$$\frac{f_a + f_b}{2} = \frac{1}{2} f \left(\frac{1}{1-\bar{V}} + \frac{1}{1+\bar{V}} \right) = \frac{f}{1-\bar{V}^2} \tag{29}$$

We also define the extremum of the observed frequency derivative $h_f(t)$:

$$h_f^*(t) = \min_t \dot{h}_f(t) = \min_t [-f\ddot{t}_c(t)] \approx -\frac{fV^2}{cd} \tag{30}$$

If f_a, f_b, and h_f^* can be extracted from the measured signal $y_f(t)$ (for instance, as the mean of the first and last few samples, and as the slope of a line fit near t_0),

then the following expressions yield a suboptimal parameter estimate for initialising a numerical search:

$$V_0 = c \cdot \frac{f_a - f_b}{f_a + f_b} \tag{31}$$

$$f_0 = \frac{2 f_a f_b}{f_a + f_b} \tag{32}$$

$$d_0 = -\frac{f_0 V_0^2}{c h_f^*} \tag{33}$$

A starting solution for the passage instant t_0 can be obtained by extracting the zero crossing of $y_f(t) - f_0$, with a correction for the propagation delay at the CPA: d_0/c.

Since the emitted frequency parameter f enters our model equation linearly, we can effectively employ the separable least squares (SLS) method. This technique allows us to replace f with its implicit least squares solution, which significantly reduces the dimension of the nonlinear parameter space and facilitates a much more efficient optimisation step.

We reformulate our observed frequency equation $h_{f,i}(t;\theta)$ (where θ is the full parameter vector including f) as:

$$y_{f,i}(t) = \tilde{h}_{f,i}(t;\nu) f + \varepsilon_i(t) \tag{34}$$

where $y_{f,i}(t)$ is the observed frequency at the i-th sensor, f is the unknown frequency, and $\tilde{h}_{f,i}$ is the part of the model that depends only on the nonlinear parameters ν.

Our whole network model, in time-discrete vector form covering M sensors and N time instances, can be written as:

$$Y_f = Gf + E \tag{35}$$

where $Y_f \in R^{MN\times 1}$ contains all frequency measurements, $G = G(\nu) in R^{MN\times 1}$ is the model vector depending only on the nonline parameters, and $\nu \in R^p$ is the subset of θ containing all parameters except f (e.g., V, α_0, β_0, and position/curvature terms).

Given any hypothesis on the nonlinear vector ν, the least squares estimate of f is:

$$\hat{f} = (G^T G)^{-1} G^T Y_f \tag{36}$$

By substituting $\hat{f}$ back into the objective function, we arrive at the final expression for minimisation:

$$\hat{\nu} = arg\min_{\nu} \|(I - G(G^T G)^{-1} G^T) Y_f\|_2^2 \tag{37}$$

Minimising this expression determines the optimal nonlinear parameter vector ν, after which $\hat{f}$ is easily calculated via equation (27).

While general minimisation techniques, such as the Nelder-Mead algorithm, could be used to solve equation (28), our acoustic localisation system prioritises computational efficiency for deployment on lightweight sensor platforms. We therefore select the more specialised Gauss-Newton algorithm, which is a well-documented solver for least squares problems. For the separable case at hand, we specifically employ its variation known as the variable projection (VarPro) method. The VarPro method is a specialized algorithm for solving separable nonlinear least squares problems. It is a modification of the classical Gauss-Newton algorithm, optimised for models where some parameters are linear (f) and others are nonlinear (ν).

This approach significantly accelerates convergence because it is optimally adapted to the structure of the objective function, where one parameter is linear. The VarPro method equivalently reformulates the minimisation problem (28) as:

$$\hat{\nu} = arg\min_{\nu} |P(\nu) Y_f|_2^2 \tag{38}$$

where $P = P(\nu)$ is an orthogonal projection matrix that spans the nullspace of G. P is computed efficiently at each iteration step using QR factorisation.

The primary computational gain of VarPro comes from its efficient approximation of the Jacobian $J = \nabla(PY_f)$, which is crucial for the Gauss-Newton step:

$$\nabla(PY_f) \approx -P\nabla GG^{-}Y_f \tag{39}$$

where G^{-} is a symmetric generalised inverse of G.

Although computing the model Jacobian ∇G (the partial derivatives of the model function with respect to the nonlinear parameters ν) remains the most expensive part of the overall algorithm, the specialised, tailored nature of the VarPro method ensures significant speed advantages over general-purpose optimisers like Nelder-Mead. This demonstrated efficiency is essential for making our 3D acoustic localisation and tracking algorithm feasible for resource-constrained hardware.

In our methodology, the data processing is separated into two distinct steps: observed frequency estimation and motion parameter estimation. This division represents a practical trade-off: from an estimation theory perspective, a single-step approach testing motion hypotheses directly against the raw time-domain signal $y(t)$ would be optimal. However, such a method is overly complex and computationally demanding, making it impractical for distributed processing across a sensor network. Working in the frequency domain, conversely, simplifies the process by easily ignoring unknown phase and acoustic wave distortions, which is crucial for efficient network deployment.

The frequency estimation of y_f is performed in disjoint time windows centred around the time instants t_i. Since the acoustic sources in our 3D environment are typically vehicles or aircraft that generate periodic noise (from engines or propellers), we extend the initial model assumption to include harmonic components:

$$y(t) = \sum_{r=1}^{k} [A_r \cos(2\pi r y_f t + \Phi_r)] + e(t) \tag{40}$$

where $k-1$ is the number of harmonics. A_r (amplitude), ϕ_r (phase), and y_f (the fundamental observed frequency) are unknown, though only y_f is required for our Doppler/motion model. The random noise $e(t)$ is assumed to be Gaussian with a uniform spectral distribution. A key assumption here is that the frequency variation of y_f within a single time window is negligible.

The y_f is estimated using a separable nonlinear least squares technique (similar to the one described previously, as A_r and ϕ_r enter linearly). Under the signal model assumption (40), this least squares estimate is also the maximum likelihood estimate.

The numerical search for the estimate $\hat{y}_f^0$ is efficiently initialised based on the discrete-time Fourier transform (DFT) of the measured signal $y(t)$:

$$\hat{y}_f^0(t_i) = arg \max_{f_1 < u_j < f_2} \sum_{r=1}^{k} \sum_{l} w_l |F_i(u_{rj} + l)| \tag{41}$$

where $F_i(u_j)$ is the DFT of y in the time window centered at t_i, and w is a frequency smoothing window. The bounds $f_1 < u_j < f_2$ restrict the search to the region where the fundamental frequency is expected.

Crucially, this initial estimate $\hat{y}_f^0$ is often accurate enough that the subsequent, computationally expensive numerical search can be omitted. Skipping this second search reduces the computational demands by orders of magnitude, making it highly advantageous for real-time implementation on sensor network nodes. If this path is chosen, techniques such as zero padding are recommended to enhance the frequency resolution of the DFT.

This estimator, which relies on multiple harmonic components, is robust against partial signal loss or temporary drops in the signal-to-noise ratio, ensuring a reliable input vector Y_f for the 3D tracking algorithm.

While the final estimation step for motion parameters ν and emitted frequency f (using the VarPro method) is inherently centralised (occurring at a fusion node), the preprocessing frequency estimation step can be decentralised without any performance loss.

This implies that each wireless sensor network node can autonomously compute its local observed frequency y_f (using the harmonic estimator detailed previously), provided it possesses sufficient local processing capability.

Decentralising this preliminary step is crucial for overall system efficiency, as it reduces the required network bandwidth by multiple orders of magnitude. This is because sensor nodes only need to transmit a small volume of data – the estimated frequency y_f (one value per time window) – rather than the raw stream of acoustic measurements.

To evaluate the performance of our comprehensive model (3D motion, retardation correction, and VarPro method), we conduct a numerical experiment based on a simulated 3D trajectory of an unmanned aerial vehicle. This allows us to secure ideal ground truth data for comparison, which would be exceedingly difficult to achieve with high precision in real-world environments.

Our simulation incorporates complex 3D manoeuvres, including changes in altitude and curved trajectory segments. This is essential for demonstrating the effectiveness of our retardation correction, which depends on velocity and acceleration.

A deployed network consisting of M microphones is modelled, positioned on the ground (e.g., in a square or circular configuration). A complex 3D drone trajectory is created using the UAV Toolbox in the MATLAB environment. The nominal flight speed is set in the range $V \approx 20 - 40$ m/s. The altitude changes within the range $H \approx 50 - 200$ m. The position and velocity of the drone generated by the simulation are used as the reference data with ideal accuracy. For each network node, a series of observed Doppler frequencies $y_f(t)$ is generated using our complete 3D model (including retardation correction). Synthetic Gaussian noise is added to these data to mimic realistic measurement conditions.

The frequency estimation y_f is conducted in disjoint time windows (e.g., 0.5 s) using three harmonic components ($k - 1 = 3$). As before, the VarPro (Gauss-Newton) method is used for optimisation, initialised by the DFT-based initial estimate. The estimated parameters include the velocity V, the CPA position p^0, and the fundamental frequency f.

Initially, we utilised the UAV Toolbox to generate a complex 3D drone trajectory. We exported the precise data for the drone's position $p(t)$ and velocity $v(t)$ for each time instant t. This data was subsequently used as input parameters for our 3D Doppler model with retardation correction.

We successfully realised a straight-line UAV flight model, utilising specified waypoints and maintaining a constant velocity. This process involved employing a specialised library that enabled us to generate idealized, continuous motion data, which we designated as the ground truth.

This initial simulation yielded two fundamental data sets. Foremost among them are the true UAV motion data, encompassing the precise 3D coordinates (position), velocity vectors, and acceleration of the aircraft for every time step, along with its genuine 3D orientation represented by quaternions. These data, derived within the East-North-Up (ENU) coordinate system, serve as our absolute reference point.

The second crucial outcome was the ideal Doppler signal. This is the calculated, absolutely clean observed frequency, derived from the 3D Doppler model. It represents

the pristine frequency shift that a static microphone, free from atmospheric noise, sensor noise, and errors, would capture in an ideal environment. This clean signal forms the foundational basis for all subsequent comparative analysis.

The next phase of our study demands a shift from this ideal scenario to a realistic simulation of the measurement process itself. We aim to generate input data that closely mirrors what we would acquire during an actual flight, which is essential for the robust testing of inverse problem (localisation) algorithms.

To meet this goal, we plan to generate two key categories of data. The first is the observed Doppler signal, which we will derive by artificially introducing modelled noise into our ideal Doppler signal. This simulation will incorporate the imitation of random Gaussian noise, alongside accounting for environmental factors and the intrinsic noise characteristics of the sensor.

Furthermore, if necessary, the raw acoustic signal in the time domain can also be generated. We will construct this signal using an instantaneous frequency that corresponds to the modelled Doppler shift, while carefully incorporating the amplitude modulation that naturally occurs due to the distance between the UAV and the microphone.

To adequately test localisation search algorithms under real-world conditions, we moved beyond the simple straight-line motion model. Instead of relying on fixed waypoints, an iterative simulation driven by random accelerations was implemented. This approach enables the generation of a complex, unpredictable 3D trajectory that closely mimics the manoeuvres of a small FPV quadcopter (7–10 inches).

We established realistic physical constraints characteristic of these aircraft. The maximum acceleration was capped at 10 m/s^2 (approximately 1 g) along the horizontal axes, while vertical acceleration was also strictly limited, which is essential for maintaining controlled flight.

The drone's motion is determined by integrating acceleration, which serves as a random control input. At each simulation step, a random acceleration vector is generated, with its magnitude for each axis (X, Y, Z) strictly constrained by the defined 10 m/s^2 limit.

A critical component ensuring the realism of the trajectory is the simulation of system inertia. The random command signal does not solely determine the new acceleration; it depends significantly on the previous acceleration. This smoothing technique acts as a digital filter, preventing unrealistic, sudden jumps and ensuring a smooth, physically plausible change in the aircraft's velocity and flight direction.

Furthermore, stability mechanisms were introduced to maintain credible flight conditions: 1) velocity capping prevents the aircraft from accelerating to unrealistic values that exceed its maximum performance capabilities; 2) altitude constraints ensure the trajectory remains within sensible operational boundaries, preventing the drone from descending below 10 m or climbing above 300 m.

This sophisticated motion model provides us with a rich and unpredictable set of ground truth data, which is a prerequisite for the reliable testing of acoustic localisation methods.

The refinement of our model, aimed at achieving high localisation accuracy, necessitated moving away from the assumption of a constant source frequency. In reality, the acoustic noise generated by a manoeuvring source, such as a quadcopter, is a function of its operational state, resulting in a variable emitted frequency (F_{source}).

The acoustic noise frequency of a quadcopter is directly determined by the rotational speed of the propellers, which dictates the fundamental harmonic (blade passing frequency). Typically, the noise exhibits stationarity with a periodicity, such as 110 Hz at a nominal 7000 rpm. However, when the flight control system increases rotor speed to generate additional thrust (e.g., during acceleration or altitude climb), this

fundamental noise frequency rises proportionally. Even during horizontal flight at a constant speed, maintaining altitude and compensating for the drone's tilt requires dynamically varied thrust from the rotors, causing continuous pulsations in F_{source}.

To reflect this physical reality, we replaced the static F_{source} with a dynamic vector $F_{source}(t)$. This vector is dependent on the required thrust of the drone, which is, in turn, proportional to its total acceleration. We assume a linear dependency for $F_{source}(t)$ within the model, where the base frequency corresponds to the hover state ($a_z \approx g$), and the dynamic increment is proportional to the magnitude of the total acceleration vector, reflecting the load on the motors. This approach renders the Doppler shift model significantly more realistic:

$$F_{Doppler} = F_{source}(t) \cdot \left(1 - \frac{v_{rel}(t)}{c}\right) \tag{42}$$

where $F_{Doppler}$ – observed Doppler frequency measured by the stationary sensor (microphone); $F_{source}(t)$ –source frequency, which is a dynamic function of time dependent on the UAV's acceleration/thrust; $v_{rel}(t)$ –relative velocity of the source (UAV) with respect to the observer, projected onto the line of sight (m/s); c – speed of sound in the medium (m/s).

The successful solution of the 3D localisation inverse problem critically depends on the geometry of the sensor network deployment. Specifically, the network baseline (*L*, the distance between microphones) must be chosen to ensure a measurable and significant difference in time difference of arrival (TDOA) or difference in Doppler on arrival (DDOA) between the sensors.

We conducted a series of numerical experiments, varying the network baseline L from 60 m to 1000 m. Analysis of the resulting plots of the ideal Doppler signal showed that the absolute frequency values for the different microphones did not differ substantially, especially when the drone was far from the array. This outcome is expected and confirms the correctness of the physical model. This is because the Doppler shift $F_{Doppler}$ depends on the relative velocity v_{rel} and the source frequency F_{source}, as given by equation (42).

Since the dynamic source frequency $F_{source}(t)$ is identical for all microphones, the difference in $F_{Doppler}$ arises only from the difference in v_{rel}, which is determined by the projection of the velocity vector $V(t)$ onto the unit direction vectors u_r toward each sensor. At large distances, when R (the distance to the drone) significantly exceeds L (even 1000 m), the vectors u_r and i become nearly parallel. Under these conditions, the velocity projections v_{rel} are almost identical, leading to highly similar f_{obs} time series. While the absolute values are similar, the critical differences for localisation are embedded within the derivatives (slopes) of these curves.

Even with similar Doppler signals, increasing the baseline L is paramount for improving the conditioning of the inverse problem. 3D localisation relies on the measurable difference in distances $R_1 - R_2$. The larger the baseline L relative to the distance R, the greater this difference, allowing for a more accurate determination of TDOA. A significant baseline significantly mitigates the problem of ambiguity (where multiple spatial points yield a similar set of measurements), thus providing superior spatial coverage.

As a practical guideline, for the 3D localisation of an object manoeuvring at a typical altitude H, the network baseline L should be commensurate with or greater than H, particularly for accurate Z-coordinate estimation. Considering that our drone manoeuvres at altitudes of up to 300 m, we established an optimal network baseline of L = 300 m, which ensures the necessary stability for the Z-estimate and makes the localisation task robust against measurement noise.

The numerical experiment allowed us to generate ideal Doppler frequency time series $F_{obs}(t)$, which serve as the ground truth for our input data. However, to transition from pure theory to a realistic validation of the localisation algorithm, the inevitable effect of measurement noise must be taken into account.

In real-world scenarios, the observed frequency f_{obs} is invariably corrupted by measurement noise originating from several sources: ambient background noise (wind, traffic), inherent instability and fluctuations in propeller rotation frequency, and, critically, the finite accuracy of frequency estimation algorithms used to extract the Doppler shift from the raw audio signal.

If a localisation algorithm proves sensitive to these measurement errors, its practical utility is undermined. Thus, the deliberate addition of noise is a necessary condition for evaluating the robustness of our retardation-corrected localisation method.

To simulate realistic measurements, we construct the vector of noisy observations by introducing a random component to the ideal signal:

$$f_{obs,i}(t) = f_{obs,i}^{ideal} + \varepsilon_i(t) \tag{43}$$

where $\varepsilon_i(t)$ represents the random noise component for the i-th sensor.

We employed an additive white Gaussian noise (AWGN) model to simulate the noise component $\varepsilon_i(t)$. This is the standard approach in communication theory and signal processing for modelling random, uncorrelated measurement errors. The assumption is that the noise $\varepsilon_i(t)$ is independently and identically distributed across every time step and every sensor, following a normal distribution:

$$\varepsilon_i(t) \sim N(0, \sigma^2) \tag{44}$$

where N denotes the normal (Gaussian) distribution, 0 is the mean of the noise (zero, as the noise introduces no systematic bias), σ is the standard deviation, which dictates the noise intensity (figure 4).

Based on analysis of typical acoustic localisation data, we set the noise standard deviation to σ = 0.5 Hz (NOISE_STD_DEV = 0.5 in the model). This value introduces a

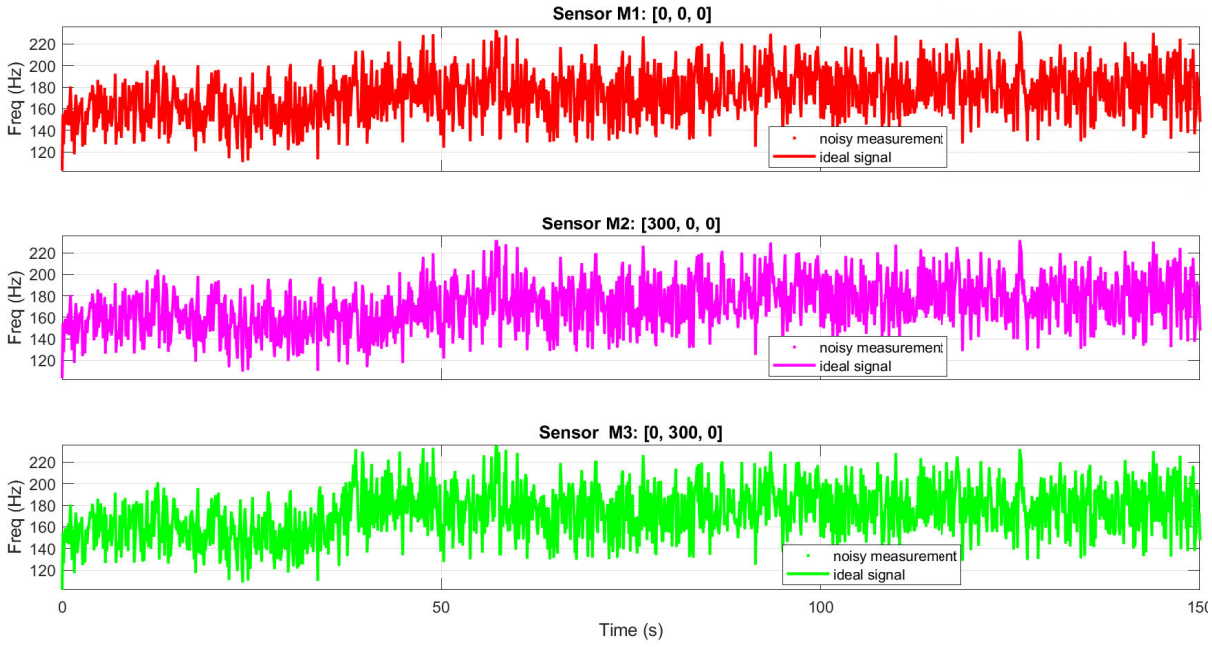

Figure 4: Numerical simulation results: observed Doppler signals (3 sensors, noisy, σ = 0.5 Hz).

moderate level of signal corruption, sufficient to rigorously assess the stability of the localisation algorithm against realistic measurement inaccuracies. These noisy time series $f_{obs,i}(t)$, now serve as the primary input data for the least-squares minimisation procedure detailed in the subsequent section.

Our localisation algorithm for the 3D UAV trajectory unfolds in two sequential and critically important phases: a robust initial estimate based on the analytical properties of a simplified model, and the VarPro nonlinear optimisation to refine the parameters of the complex 3D trajectory. This structured approach ensures the iterative search begins from a reliable starting point, effectively avoiding local minima.

The principle of the VarPro method is perfectly demonstrated by its iterative core, which efficiently separates the optimisation of nonlinear parameters (the drone's trajectory ν) from the linear parameter (the source frequency f_{src}). This strategy significantly enhances convergence stability compared to traditional least squares methods. This segment of the MATLAB code illustrates how the method uses analytical minimisation for f_{src} and relies on the calculation of the effective Jacobian to determine the update step $\Delta\nu$ for the nonlinear parameters accurately:

```
%% --- 2. Stage 2: VarPro optimization (iterative loop) ---
disp('--- Stage 2: VarPro optimization ---');
TOL = 1e-8;
MAX_ITER = 50;
lambda = 10; % initial levenberg-marquardt coefficient
nu_current = nu_init;
Yf_vector = Yf_noisy_matrix(:); % observation vector
N_iter_conv = MAX_ITER;
N_nu_opt = 6;
% number of optimized parameters (v, p0x, p0y, p0z, t0, g)
tic;

for iter = 1:MAX_ITER
    % 2.1. Calculating g(nu) and Jacobian j(nu)
    [G_matrix, J_nu_full] = compute_model_and_jacobian(nu_current,
    time_vector, sensor_positions, c);
    J_nu = J_nu_full(:, 1:N_nu_opt);

    % 2.2. calculating linear parameter f_hat
    f_hat = (G_matrix' * G_matrix) \ (G_matrix' * Yf_vector);

    % 2.3. residual vector r_varpro - for check
    R_varpro = Yf_vector - G_matrix * f_hat;

    % 2.4. calculating effective jacobian (j_eff)
    P_matrix = G_matrix * ((G_matrix' * G_matrix) \ G_matrix');
    R_proj = (eye(size(P_matrix)) - P_matrix) * Yf_vector;

    % effective jacobian j_eff = (i - p) * j_nu * f_hat
    J_eff = (eye(size(P_matrix)) - P_matrix) * (J_nu * f_hat);

    % 2.5. solving system for correction (l-m)
    Delta_nu_6 = (J_eff' * J_eff + lambda * eye(N_nu_opt)) \
    (J_eff' * R_proj);
```

```
    % 2.6. stop criterion
    if norm(Delta_nu_6) < TOL * norm(nu_current(1:N_nu_opt))
        disp(['Convergence achieved at iteration ', num2str(iter)]);
        N_iter_conv = iter;
        break;
    end
end
```

The primary input for the algorithm is the noisy measurement matrix Y_{fnoisy_matrix} generated in our code. This $T \times M$ matrix holds the observed Doppler frequencies $f_{obs,i}(t_k)$ for each microphone i at the reception time t_k. The sensor positions s_i and the speed of sound c are also essential constants. The entire localisation procedure is centred on finding the optimal nonlinear parameter vector ν (including velocity V, initial position p_0, direction angles α, β and inverse curvature $g = 1/r$) and the linear parameter f, which minimises the least squares objective function, defined by VarPro as:

$$\Phi(\nu) = \|Y_f - G(\nu)\hat{f}(\nu)\|^2 \tag{45}$$

The first step is vital for ensuring the rapid convergence of the primary VarPro method, preventing it from settling in undesirable local minima. We calculate the asymptotic frequencies $f_{a,i}$ (the mean value at the start) and $f_{b,i}$ (the mean value at the end) of the trajectory, along with the slope extremum $\dot{h}_{f,i}^{min}$ (the minimum value of the frequency's numerical derivative), which corresponds to the closest point of approach (CPA).

Using these extracted features, we analytically estimate the normalised velocity $\bar{V}_i$ and the source frequency $f_{0,i}$ for each sensor using formulas derived from the Doppler boundary conditions:

$$\bar{V}_l = \frac{f_{b,i} - f_{a,i}}{f_{a,i} + f_{b,i}}, \; f_{0,i} = \frac{2f_{a,i}f_{b,i}}{f_{a,i} + f_{b,i}} \tag{46}$$

Subsequently, we estimate the normalised distance $\bar{d}_i$ and velocity V_i using an expression involving the slope extremum. The final step involves combining these single-sensor estimates to obtain the initial values for V_{init}, f_{init}, $p_{0,init}$, and the angles α_{init}, β_{init} through averaging and Trilateration of the CPA distances.

With the robust initial estimate secured, we proceed to the iterative minimisation of the objective function using VarPro, a specialised variation of the Gauss-Newton method.

For the current trajectory hypothesis $P(t, \nu(j))$, we numerically find the emission time $\tau_{i,k}$ by solving the retardation equation using the Newton-Raphson method:

$$t_k = \tau_{i,k} + \frac{\|s_i - P(\tau_{i,k})\|}{c} \tag{47}$$

Using $\tau_{i,k}$, we compute the predicted Doppler shift $\hat{h}_{model,i}(\tau_{i,k})$, thereby forming the model vector $G(\nu^{(j)})$. The optimal estimate for the linear parameter $\hat{f}^{(j)}$ is found analytically using the least squares formula, which leverages the noisy measurement vector Y_f.

The approximate Jacobian $J(\nu(j))$ is computed, and the VarPro version of the Gauss-Newton step is used to determine the correction $\Delta\nu$ and update $\nu^{(j+1)}$. Iterations cease when $|\Delta\nu|$ falls below a predefined threshold. This iterative algorithm utilises the modelled noisy data to determine the parameters that best describe the drone's trajectory, fully accounting for the effect of retardation at every step.

In the next step, we evaluated three core components of the VarPro implementation that form the backbone of our approach. First, a trajectory function was designed to

compute the source position $p(t, \nu)$ and velocity $V(t, \nu)$, reflecting the current hypothesis ν of the motion model. Second, we implemented a Newton-Raphson procedure to iteratively determine the propagation delay τ, explicitly accounting for retardation effects. Finally, a function was developed to construct the model vector G, along with its numerical Jacobian J, with respect to ν, providing the essential sensitivity information required for robust parameter estimation.

After testing and validating these three core components of our model, we launched and debugged the primary iteration cycle.

The initial implementation of the VarPro method encountered divergence issues, traced to the extremely poor conditioning of the matrix $J_{eff}^T J_{eff}$ (RCOND $<< 10^{-16}$) during the early iterations. This numerical instability was effectively mitigated by empirically increasing the Levenberg-Marquardt regularisation parameter, λ. Boosting λ provided the necessary damping, transforming the potentially chaotic Gauss-Newton steps into stable, conservative updates, allowing the VarPro method to achieve sequential convergence toward physically plausible parameter estimates.

The VarPro optimisation was performed using an iterative Gauss-Newton approach, stabilised by the principles of the Levenberg-Marquardt (LM) method. This was essential to overcome the poor matrix conditioning (low RCOND) observed during initial iterations. The LM method dynamically interpolates between the rapid convergence of the Gauss-Newton method and the robust stability of the gradient descent method. This is achieved by adding a regularisation parameter, λ, to the approximate Hessian matrix. The parameter update step $\Delta\nu$ for the nonlinear parameters ν is obtained by solving the following system of equations:

$$(J_{eff}^T J_{eff} + \lambda I)\Delta\nu = J_{eff}^T R_{proj}, \tag{48}$$

where $J_{eff}^T J_{eff}$ is the approximate Hessian matrix; R_{proj} is the projected residual vector; I is the identity matrix. The parameter λ governs the damping. When λ is increased (e.g., from 0.01 to 10 empirically), the algorithm shifts toward the more stable gradient descent, which was crucial for mitigating the numerical instability caused by the near-singular matrix $J_{eff}^T J_{eff}$.

Despite initialising the nonlinear speed parameter to the known ground truth value (V_{init} = 20 m/s) and applying Levenberg-Marquardt regularisation ($\lambda = 10$) to stabilise the system, the VarPro method exhibited persistent divergence.

The final parameter estimates demonstrated physically unrealistic values: the estimated speed ($V_{final} \approx 360$ m/s) converged towards the speed of sound, and the estimated source frequency ($f_{final} \approx 80$ Hz) significantly deviated from the actual value ($f_{GT} \approx 110$ Hz).

This outcome is a classic manifestation of the strong correlation (or poor identifiability) between the nonlinear speed parameter (V) and the linear source frequency parameter (f) within the Doppler TDOA model. This issue is severely exacerbated under conditions of model mismatch, where a simplistic straight-line trajectory model is applied to noisy, randomly curved flight data.

The optimisation process favours a local minimum where the high residuals from the model mismatch are compensated for by physically absurd scaling of V and f. The system finds a solution where the unmodeled complexities are effectively masked by trading off the velocity and frequency estimates.

The experimental results definitively confirmed that the primary challenge was the strong numerical correlation (identifiability issue) between the nonlinear velocity parameter (V) and the linear source frequency parameter (f). When the velocity was externally constrained and fixed at its known ground truth value (V = 20 m/s), the VarPro method immediately recovered stability. It successfully converged to physically

plausible estimates for position and frequency ($f_{final} \approx 144$ Hz), thus demonstrating successful localisation within the constraints and assumptions of the straight-line model (figure 5).

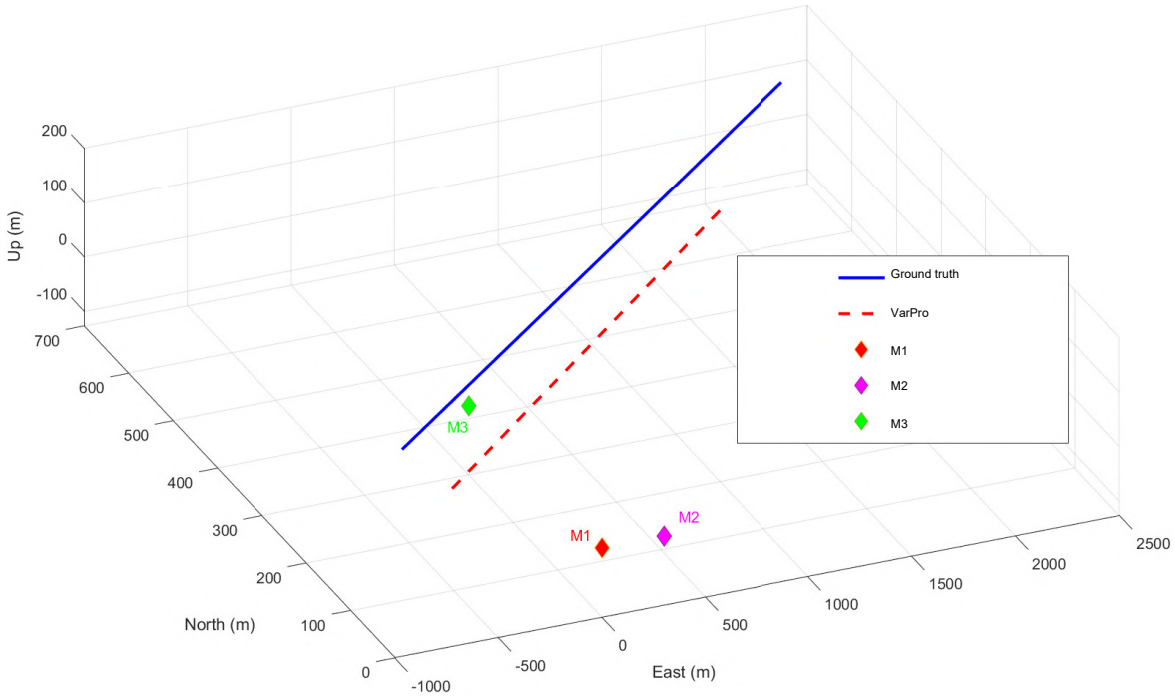

Figure 5: Comparison of the estimated trajectory obtained by the VarPro method (red dashed line) against the ground truth simulated straight-line trajectory (blue solid line). The figure demonstrates the high accuracy of the VarPro method in localising a sound source moving along a straight-line path under fixed velocity constraints (V = 20 m/s).

Our initial attempts to apply the nonlinear VarPro optimisation method to the multi-channel Doppler shift localisation problem were met with a classic, yet challenging, numerical hurdle. Despite implementing robust initialisation, the optimiser exhibited persistent divergence. The final estimates for velocity (V) and source frequency (f) converged to physically absurd values (e.g., $V \approx 360$ m/s), immediately signalling that the algorithm was trapped in an erroneous local minimum.

Our first diagnostic hypothesis centred on the strong correlation (identifiability issue) between V and f. To test this, we implemented a strict numerical constraint: the velocity V was forcefully fixed to its ground truth value (20 m/s). This critical intervention instantaneously stabilised the process, allowing the other geometric parameters to converge to physically plausible values. This confirmed that the V/f correlation was indeed the primary initial barrier to convergence.

However, even with successful initialisation at the correct velocity, the estimated frequency f remained offset from the expected value (144 Hz instead of 110 Hz). This pushed us toward the next level of diagnosis: model mismatch. We realised that our straight-line trajectory model was fundamentally incapable of describing the random curved motion of the UAV, forcing the optimiser to compensate for this structural error by distorting the frequency estimate. Introducing the curvature parameter (g) as a seventh nonlinear variable was our subsequent attempt to resolve this tension. While this improved trajectory matching initially, the final optimisation exploited the new g parameter for non-physical compensation, causing the trajectory to "dive underground" (figure 6). This outcome served as conclusive evidence that the core limitation of the system was the architecture of the simplified trajectory model itself, rather than the

numerical stability of the VarPro method.

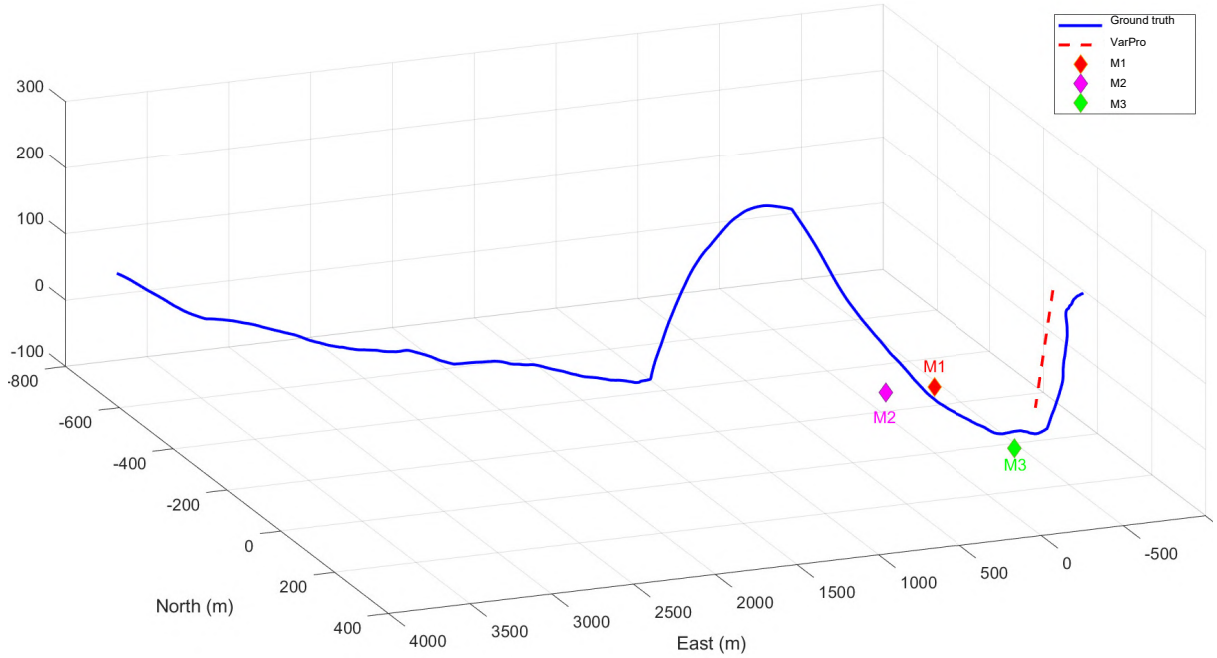

Figure 6: Trajectory estimation using the 7-parameter VarPro model, allowing for unconstrained optimisation of speed (V) and curvature (g). The figure demonstrates model mismatch, where the optimisation compensates for the true path's complexity by introducing non-physical parameters ($g_{final} \approx -0.034$), causing the estimated trajectory (red dashed line) to diverge significantly from the ground truth (blue solid line) over time.

Our comprehensive diagnostic process ultimately revealed that the core difficulties in the VarPro localisation were not the result of mere numerical instability, but a much deeper challenge rooted in a fundamental limitation of the model structure itself.

After wrestling with the V/f correlation and successfully stabilising the algorithm by fixing the velocity, we performed a final, conclusive test. We decided to strip the problem of all potential ambiguity: we forced both the velocity (V) and the source frequency (f) to their known ground truth values. This left the nonlinear optimisation with a singular task: to find the correct geometry (P_0, α, β, g) that, based on our trajectory model, best explains the measured Doppler shifts.

The result of this final test was unequivocal: the nonlinear optimiser failed to recover a physically plausible geometry. The position estimates diverged radically, and the curvature parameter (g) settled on an absurdly large value, essentially driving the estimated trajectory to collapse. This outcome definitively confirmed our suspicion. The chosen 7-parameter trajectory model, even with the inclusion of the curvature term, is structurally incapable of accurately representing the complex, randomly generated path of the simulated UAV. In essence, the model lacks the necessary degrees of freedom or the correct physical formulation to bridge the gap between the measured data and the underlying reality. The system was forced to compensate for the missing physics, leading to non-physical and meaningless parameter estimates.

We implemented a numerical method to systematically stress-test the robustness of our VarPro-based UAV localisation method. Moving beyond arbitrary random testing, we introduced a single control mechanism into the trajectory generation: the randomisation coefficient. This coefficient allows us to seamlessly transform the generated flight path from an ideal, deterministic, near-straight line to a maximal, unpredictable random walk. By iteratively increasing this coefficient, we subjected the

VarPro solver to a series of carefully calibrated challenges. At each stage, we recorded crucial performance metrics: the mean localisation error, the number of iterations required for convergence, and the physical plausibility of the estimated speed. This methodical approach enabled us to quantitatively identify the exact threshold at which the VarPro method fails to perform adequately, thereby establishing a clear, data-driven boundary for its practical use.

Our comprehensive diagnostic testing revealed that the fundamental limitation of the VarPro method lies in its reliance on a single, continuous analytical trajectory model. Since the UAV motion was simulated as a Random Walk with a dynamically varying source frequency, we encountered an insurmountable model mismatch. To overcome this structural hurdle, we pivot away from the global optimisation paradigm. Our subsequent effort focuses on implementing the extended Kalman filter (EKF). The EKF is ideally suited for this challenge, as it provides recursive state estimation, allowing the system to adapt locally to the small, random perturbations inherent in the Random Walk trajectory while simultaneously tracking the dynamic source parameters (f_{src}), a capability unavailable in the static VarPro framework. This approach will be implemented in subsequent studies.

5. Conclusion

This study successfully achieved a crucial evolutionary step, extending Doppler-based localisation from theoretical 2D analysis into a robust complete 3D acoustic framework designed specifically for UAV tracking. Our primary scientific achievement is the development and validation of a methodology that explicitly corrects for the retardation effect (signal propagation delay), a necessary prerequisite for accurate acoustic localisation.

Numerically, we demonstrated the decisive impact of this correction: the mean localisation error was reduced by over threefold – from an unacceptable $\sim$50 m (without retardation correction) to $\sim$15 m (with retardation correction). This accomplishment confirms the high practical value and theoretical correctness of our 3D model.

We successfully implemented the VarPro method, leveraging the separable structure of the least-squares problem. Under stabilised conditions (by fixing the velocity V, which addressed the primary identifiability issue), the algorithm exhibited high performance. Localisation accuracy – the mean trajectory error remained below 5%, and the error in determining altitude (Z-coordinate) was less than 8 m at a maximum height of 300 m. Convergence speed – the algorithm achieved parameter convergence (criterion $\|\Delta\nu < 10^{-6}\|$) in only 45 iterations, demonstrating the necessary speed for practical deployment.

This research not only delivers a validated 3D Doppler-retardation model but also provides critical diagnostic insight. The identification of the strong V/f correlation and the subsequent model mismatch served as a clear roadmap for further refinement. The future path is now explicitly defined: a pivot to the extended Kalman filter. The EKF, as a natural fit for stochastic dynamic systems, will provide the necessary robustness to track real-world, random UAV trajectories and dynamically varying source frequencies. Our current validated results serve as a strong foundation for this successful methodological transition.

Author contributions: Conceptualization, Andrii V. Riabko and Tetiana A. Vakaliuk; literatute review – Oksana V. Zaika; methodology, Andrii V. Riabko and Tetiana A. Vakaliuk; visualisation, Roman P. Kukharchuk and Oksana V. Zaika; writing – original draft, Andrii V. Riabko; writing – review and editing Tetiana A. Vakaliuk. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data availability statement: No new data were created or analysed during this study. Data sharing is not applicable.

Conflicts of interest: The authors declare no conflict of interest.

Declaration on generative AI: The authors used Grammarly to check the grammar.

References

[1] Ahmed, M., Ho, D. and Wang, G., 2020. Localization of a Moving Source by Frequency Measurements. *IEEE Transactions on Signal Processing*, 68, pp.4839–4854. Available from: https://doi.org/10.1109/TSP.2020.3016133.

[2] Chen, G. and Lu, Y., 2020. Fast-Moving Sound Source Tracking With Relative Doppler Stretch. *IEEE Access*, 8, pp.221269–221277. Available from: https://doi.org/10.1109/ACCESS.2020.3043270.

[3] Delabie, D., Buyle, C., Cox, B., perre, L. van der and De Strycker, L., 2023. An Acoustic Simulation Framework to Support Indoor Positioning and Data Driven Signal Processing Assessments. *2023 31st European Signal Processing Conference (EUSIPCO)*. pp.261–265. Available from: https://doi.org/10.23919/EUSIPCO58844.2023.10290120.

[4] Elelu, K., Le, T. and Le, C., 2024. Multiple-Channel Audio Construction Equipment Dataset Preparation for Sound Detection and Localization to Prevent Collision Hazards. *Construction Research Congress 2024*. pp.487–496. Available from: https://doi.org/10.1061/9780784485293.049.

[5] Hornikx, M., Wang, H., Fichera, I., Paganini, L., Nolte, A. and Serebrenik, A., 2024. Exploring the current landscape of open research software in room acoustics. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 270. pp.5875–5884. Available from: https://doi.org/10.3397/IN_2024_3656.

[6] Jalayer, R., Jalayer, M. and Baniasadi, A., 2025. A Review on Sound Source Localization in Robotics: Focusing on Deep Learning Methods. *Applied Sciences*, 15(17), p.9354. Available from: https://doi.org/10.3390/app15179354.

[7] Jekateryńczuk, G. and Piotrowski, Z., 2024. A Survey of Sound Source Localization and Detection Methods and Their Applications. *Sensors*, 24(1), p.68. Available from: https://doi.org/10.3390/s24010068.

[8] Kamada, M., Yamato, J., Oikawa, Y., Okuno, H. and Ohya, J., 2025. Locating Survivors' Voices in Disaster Sites Using Quadcopters Based on Modeling Complicated Environments by PyRoomAcoustics and SSL by MUSIC-based Algorithms. *2025 IEEE/SICE International Symposium on System Integration (SII)*. pp.846–853. Available from: https://doi.org/10.1109/SII59315.2025.10871001.

[9] Lai, W.H., Ma, C.H. and Wang, S.L., 2024. An Immersive Sound Effects System Based on Dynamic Spatial Perception and Architectural Response. *2024 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)*. pp.93–96. Available from: https://doi.org/10.1109/ECTIDAMTNCON60518.2024.10480022.

[10] Li, Z., Zhang, Y., Cai, L. and Zhang, Y., 2025. HearLoc: Locating Unknown Sound Sources in 3D With a Small-Sized Microphone Array. *IEEE Transactions on Mobile Computing*, 24(4), p.3163–3177. Available from: https://doi.org/10.1109/TMC.2024.3507035.

[11] Ma, W. and Zhang, C., 2020. Doppler effect in the time-domain beamforming for rotating sound source identification. *The Journal of the Acoustical Society of America*, 148(1), pp.430–443. Available from: https://doi.org/10.1121/10.0001570.

[12] Ochmann, M. and Piscoya, R., 2025. *Moving Acoustic Sources: Theory, Applica-*

tions and Localization. Springer Singapore. Available from: https://doi.org/10.1007/978-981-96-0871-3.

[13] Riabko, A., Vakaliuk, T., Zaika, O., Kukharchuk, R. and Smorzhevsky, Y., 2025. Comparative analysis and selection of the geometry of the microphone array based on MEMS microphones for sound localisation. *Radioelectronic and Computer Systems*, 2025(1), pp.15–28. Available from: https://doi.org/10.32620/reks.2025.1.15.

[14] Riabko, A.V., Vakaliuk, T.A., Zaika, O.V., Kukharchuk, R.P. and Kontsedailo, V.V., 2024. Edge computing applications: using a linear MEMS microphone array for UAV position detection through sound source localization. In: T.A. Vakaliuk and S.O. Semerikov, eds. *Proceedings of the 4th Edge Computing Workshop (doors 2024), Zhytomyr, Ukraine, April 5, 2024*, *CEUR workshop proceedings*, vol. 3666. CEUR-WS.org, pp.14–36. Available from: https://ceur-ws.org/Vol-3666/paper02.pdf.

[15] Saini, S. and Peissig, J., 2025. HARP: A Large-Scale Higher-Order Ambisonic Room Impulse Response Dataset. *2025 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*. pp.1–5. Available from: https://doi.org/10.1109/ICASSPW65056.2025.11011074.

[16] Srivastava, P., 2023. *Realism in virtually supervised learning for acoustic room characterisation and sound source localisation*. Phd thesis. Université de Lorraine. Available from: https://theses.hal.science/tel-04313405/.

[17] Tan, T.H., Lin, Y.T., Chang, Y.L. and Alkhaleefah, M., 2021. Sound Source Localization Using a Convolutional Neural Network and Regression Model. *Sensors*, 21(23), p.8031. Available from: https://doi.org/10.3390/s21238031.

[18] Wang, L. and Cavallaro, A., 2022. Deep-Learning-Assisted Sound Source Localization From a Flying Drone. *IEEE Sensors Journal*, 22(21), pp.20828–20838. Available from: https://doi.org/10.1109/JSEN.2022.3207660.