

Advancing understanding of indoor conditions using artificial intelligence methods

Nicholas Christakis (Νικόλαος Χρηστάκης),^{1,2} Dimitris Drikakis (Δημήτρης Δρικάκης),^{1,a)} and Ioannis W. Kokkinakis (Ιωάννης Κοκκινάκης)¹

¹⁾*Institute for Advanced Modeling and Simulation, University of Nicosia, Nicosia, CY-2417, Cyprus*

²⁾*Laboratory of Applied Mathematics, University of Crete, Heraklion GR-70013, Greece*

This study presents a novel methodology for optimizing probe placement in indoor air-conditioned environments by integrating computational fluid dynamics simulations with artificial intelligence techniques in an unsupervised learning framework. The 'Reduce Uncertainty and Increase Confidence' algorithm identified spatially distinct thermal and velocity clusters based on temperature and velocity magnitude distributions. Optimization of probe positions within these clusters, guided by sequential least squares programming, resulted in an effective strategy to minimize probe redundancy while maximizing spatial coverage. The methodology highlights the interplay between temperature, relative humidity, velocity, and turbulence intensity, revealing critical insights into airflow behavior and its implications for occupant comfort. The findings of the presented study underscore the potential for targeted probe placement to provide a robust framework for advanced indoor climate control.

I. INTRODUCTION

Indoor thermal comfort significantly influences human health, productivity, and living standards. Modern buildings utilize Heating, Ventilation, and Air Conditioning (HVAC) systems to maintain thermal comfort.^{1,2} Various studies have explored how ventilation affects indoor thermal comfort using methods like sensor technology, yet these can be time-consuming or costly.³ Experimental methods have been used,⁴ but require substantial time and resources, presenting a challenge when multiple scenarios must be evaluated.

Computational models have been utilized to predict thermal comfort indoors.^{5,6} Yet, these models often face challenges due to computational costs and the intricate nature of turbulent flows, which encompasses uncertainties inherent in turbulent dynamics and the logistical complexity of large-scale measurements and computations.⁷⁻⁹

Artificial Intelligence (AI) provides a robust framework for uncovering hidden structures within noisy and incomplete data sets.¹⁰ In particular, unsupervised learning (UL) has demonstrated exceptional potential in tackling complex scientific problems, especially in the grouping and classifying of data sets. Applications of UL span diverse fields, including unravelling genetic complexities¹¹, advancing natural language processing¹², particle dynamics¹³, and astronomical phenomena.¹⁴

One of the key advantages of UL is its ability to analyze unlabeled data, such as experimental observations, computational outputs, or field measurements, with minimal human intervention. This feature is especially critical in areas where the underlying physics is not yet fully understood.^{15,16} Applying UL in engineering, physics, and biomedical sciences often encounters higher levels of uncertainty, given the intricate and poorly understood physical processes involved.¹⁷ Unlike general data in marketing or economics, clustering and prediction in fluid mechanics and engineering, in general, must

align with physical principles and remain explainable. Recent advances in UL algorithms, such as the Reduce Uncertainty and Increase Confidence (RUN-ICON) algorithm¹⁶, have demonstrated increased effectiveness in clustering and identifying regions of interest in datasets with high complexity.

This study aims to integrate CFD simulations with UL to explore comfort zones in a room where an HVAC supplies colder air at lower temperature and relative humidity, and the evolution of parameters, such as air velocity, temperature, and relative humidity, is recorded at 728 locations in the room. The study is structured into the following stages: (i) Data preprocessing: Trends and correlations in velocity, temperature, and relative humidity data obtained by computational fluid dynamics (CFD) are identified after the turbulent flow has evolved for some time. (ii) Clustering using RUN-ICON: Later flow stages are analyzed to uncover dominant clusters corresponding to flow regions of interest. (iii) Non-linear programming: Key spatial coordinates representing comfort zones are determined, leveraging optimization techniques to refine the accuracy of the clustering results. (iv) Turbulence analysis: Detailed studies of turbulence characteristics are carried out in these locations to understand their role in comfort and airflow dynamics.

This research highlights the potential benefits of integrating Unsupervised Learning (UL) with Computational Fluid Dynamics (CFD) to overcome challenges in designing efficient HVAC systems for turbulent indoor environments. The study offers practical strategies to enhance indoor air quality and thermal comfort, setting the stage for more efficient and adaptable air conditioning designs.

In Section II, we introduce the issue of identifying comfort zones in air-conditioned spaces, including a detailed explanation of the data collection process facilitated by CFD modeling. Section III provides an overview of the RUN-ICON algorithm, while Section IV presents the clustering and turbulence analysis results using non-linear programming techniques.

Finally, Section V summarizes the study's key findings and proposes future directions for integrating Artificial Intelligence (AI) with CFD to advance HVAC system design further.

^{a)}Electronic mail: Author to whom correspondence should be addressed: drikakis.d@unic.ac.cy.

II. FLOW AND ENVIRONMENTAL DATA INDOORS

The temperature of a closed rectangular room, measuring $6.0 \times 2.4 \times 2.8$ meters (length $\times$ height $\times$ width), is controlled by an air conditioning (AC) unit. The unit is placed in the center of the ceiling and comprises a square outlet (48 cm^2) and four smaller rectangular shaped inlets ($48 \times 4 \text{ cm}^2$ each) that eject cold air at 18°C at a 45° degree angle relative to the ceiling plane centerline. Figure 1 illustrates the flowfield that develops in the rectangular room – about 1 minute after the AC unit is turned on. Typically, in suddenly expanded flow, an effective diameter can be estimated such that the area of the circle is equal to that of the rectangular plane. Thus, based on the AC inlet(s) effective diameter and the velocity of the ejected cold air (4 m/s), the inlet Reynolds number is just above $Re_j \simeq 40,000$. Figure 2 provides further information on the position and dimension of the commercial four-way square cassette AC unit considered that is commonly encountered.

The initial room temperature is 25°C with a 60% relative humidity. The simulation considers a binary multi-component flow. The first component is dry air (0% moisture) with an adiabatic index of $\gamma_{da} = 7/5$ and molar mass of $\mathcal{M}_{da} = 28.964 \text{ kg/kmol}$. The second component is water vapor with an adiabatic index of $\gamma_{wv} = 4/3$ and molar mass of $\mathcal{M}_{wv} = 18.015 \text{ kg/kmol}$. The velocity is initially assumed to be equal to zero everywhere in the room. The pressure varies in the normal (y) direction due to gravity, i.e.,

$$p(y) = p_0 + \rho_m g_y y, \quad (1)$$

where the Earth's gravitational acceleration is taken as $g_y = -9.81 \text{ m/s}^2$. The internal energy, e_i is calculated according to:

$$e_i = (p_{da} + p_{wv}) / (\gamma - 1) \quad (2)$$

where p_{da} is the room pressure of the dry air component and p_{wv} is the water vapor pressure. The former is obtained according to $p_{da} = p_0 - p_{wv}$, where p_0 is taken here as the stagnation pressure at ground/sea level, i.e., $p_0 = 101,325 \text{ Pa}$. The water vapor pressure is calculated according to $p_{wv} = x_{wv} p_0$, where the molar-fraction is $x_{wv} = w_{wv} (\mathcal{M} / \mathcal{M}_{wv})$, and the total molar mass of a miscible mixture is obtained according to:

$$\mathcal{M} = \frac{1}{\sum_i (w_i / \mathcal{M}_i)} \quad (3)$$

where w_i is the mass-fraction of the i -th component. The initial air mixture density, ρ_m , is taken to be constant and is calculated based on the prescribed relative humidity, i.e., the fraction of water vapour and dry air:

$$\rho_m = \rho_{da} + \rho_{wv} \quad (4)$$

where the individual component density is calculated from $\rho_i = p_i / (R_i^* T_0)$. Here, R_i^* is the specific gas constant of the i -th component, while T_0 is the initial (stagnation) room temperature.

A block-structured hexahedral orthogonal Cartesian mesh is employed. The computational cells have an edge length

of 2 cm ($\Delta x^+ \simeq 700$), resulting in a mesh resolution of $300 \times 120 \times 140$, giving a total of 5,040,000 cells; the cell size is the same in all directions: $\Delta x = \Delta y = \Delta z = 2 \text{ cm}$. We performed numerical experiments with coarser meshes containing hexahedral cells with double the edge length (i.e., 4 and 8 cm) and a finer mesh with an edge length of 1 cm. For 90 seconds of actual flow time, the results on the finer mesh exhibited less than $< 1.3\%$ difference compared to the elected mesh (edge length of 2 cm).¹⁸

A. Governing equations

The three-dimensional Navier-Stokes Equations (NSE) for a Newtonian viscous air mixture are considered. For a finite control volume, V , the multi-component NSE can be written in the fully-conservative form as follows¹⁹:

$$\frac{\partial}{\partial t} \iiint_V \rho dV = \oint_A -\rho \mathbf{u} \cdot \hat{\mathbf{n}} dA \quad (5)$$

$$\begin{aligned} \frac{\partial}{\partial t} \iiint_V \rho \mathbf{u} dV = & \oint_A [-\rho \mathbf{u} \mathbf{u} - p \mathbf{I} + \boldsymbol{\tau}] \cdot \hat{\mathbf{n}} dA \\ & + \iiint_V \rho \mathbf{f}_b dV \end{aligned} \quad (6)$$

$$\begin{aligned} \frac{\partial}{\partial t} \iiint_V \rho e_t dV = & \oint_A [-\rho h_t \mathbf{u} + \mathbf{u} \cdot \boldsymbol{\tau} - \mathbf{q}_c - \mathbf{q}_d] \cdot \hat{\mathbf{n}} dA \\ & + \iiint_V \rho [\mathbf{f}_b \cdot \mathbf{u} + \dot{h}] dV \end{aligned} \quad (7)$$

$$\begin{aligned} \frac{\partial}{\partial t} \iiint_V \rho y_j dV = & \oint_A [-\rho y_j \mathbf{u} + \mathbf{J}_j] \cdot \hat{\mathbf{n}} dA \\ & + \iiint_V \rho \dot{y}_j dV, \end{aligned} \quad (8)$$

where ρ is the density; $\mathbf{u}$ is the velocity vector; p is the static pressure; $\hat{\mathbf{n}}$ is the outward pointing unit normal of a surface element dA of the closed finite control volume dV ; $\mathbf{f}_b$ is an external body force; $h_t = e_t + p/\rho$ is the total specific enthalpy (per unit mass); $e_t = e + \mathbf{u} \cdot \mathbf{u}/2$ is the total specific energy; $e = c_v T$ is the specific internal energy; y_j is the mass-fraction of the j -th species. We denote T the temperature, c_v the specific heat capacity at constant volume, and $\gamma = c_p/c_v$ the heat capacity ratio where c_p and c_v are the specific heat capacity at constant pressure and volume, respectively.

For a Newtonian fluid, the shear stress tensor is given by:

$$\boldsymbol{\tau} = \lambda (\nabla \cdot \mathbf{u}) \mathbf{I} + \mu [\nabla \otimes \mathbf{u} + (\nabla \otimes \mathbf{u})^T], \quad (9)$$

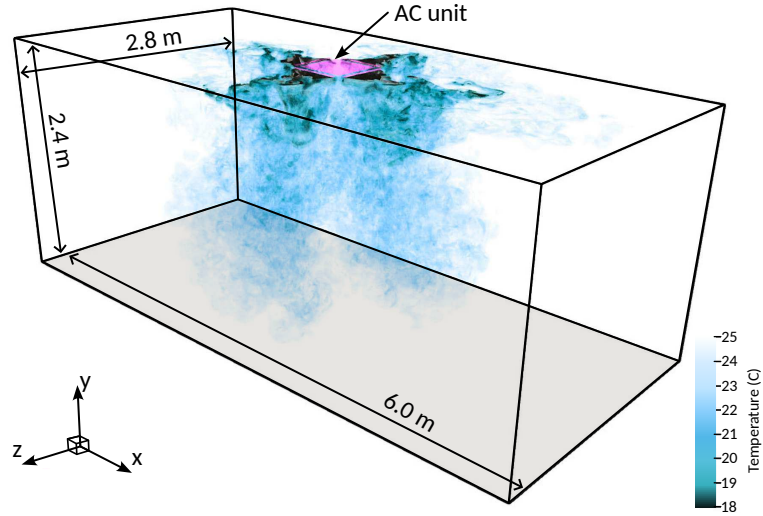


FIG. 1. Illustration of the computational domain. As shown, the room's length is 6 m, its height 2.4 m, and its width 2.8 m. The air conditioning unit is centrally located on the ceiling of the room. The (x: 0 m, y: 0 m, z: 0 m) point is located at the bottom left inner corner of the room.

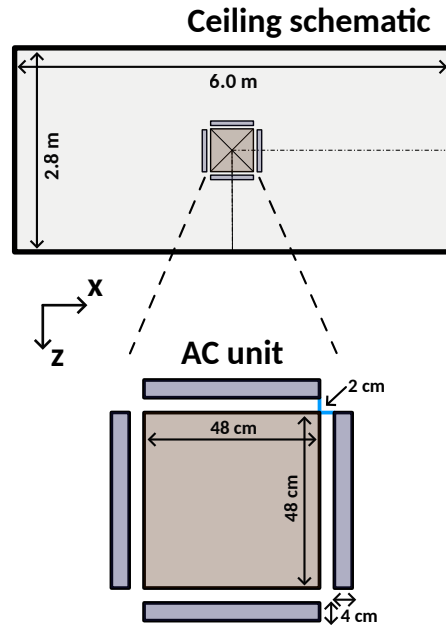


FIG. 2. Sketch illustrating the size and location of the commercial four-way square cassette AC unit considered.

where I is the identity tensor; $\lambda = -2\mu/3$ is the second viscosity coefficient given by the Stokes hypothesis; μ is the dynamic viscosity. The heat flux is calculated according to Fourier's Law of heat conduction, i.e., $\mathbf{q}_e = -\kappa \nabla T$, where $\kappa = c_p \mu / Pr$ is the thermal conductivity and Pr is the Prandtl number. The species diffusional fluxes are commonly computed via the Fickian (gradient) diffusion approximation, i.e., $\mathbf{J}_j = \rho D_j \nabla y_j$, where $D_j = \mu_j / (\rho Sc_j)$ is the j -th species mass diffusivity and Sc is the Schmidt number. The energy equation includes the inter-diffusional enthalpy flux arising from the species mixing²⁰, i.e., $\mathbf{q}_d = -\sum_{i=1}^{N_{sp}} h_i \mathbf{J}_i$, where the enthalpy of each i -th species is defined by $h_i = e_i + p_i / \rho_i$. No heat or mass sources, i.e., $\dot{h}$ and $\dot{y}$, are considered in the present study.

B. Computational method

The computational data are obtained using the block-structured CFD code CNS3D (Compressible Navier-Stokes Solver in three-dimensions).^{9,21} The code solves the advective terms using the Godunov-type (upwind) method, whose intercell numerical fluxes are calculated by solving the Riemann problem using the reconstructed values of the primitive variables at the cell interfaces. A one-dimensional swept unidirectional stencil is used for spatial reconstruction. It includes high-resolution methods of up to 11th order of accuracy in space and 4th order of accuracy in time.

CNS3D can be used to simulate turbulent flows using a varied range of approaches, such as (i) Reynolds-Averaged Navier-Stokes (RANS), (ii) implicit Large Eddy Simulations (ILES), and (iii) Direct Numerical Simulations (DNS).^{9,21}

However, the focus of present simulations is on coarse-grained ILES, which provides the best approach between accuracy and computational cost for any engineering geometry. In classical Large Eddy Simulations (LES), the smallest length scales, which are the most computationally expensive to resolve, are removed via low-pass filtering of the Navier–Stokes equations. The unresolved scales of turbulence are then modeled using subgrid-scale models. In ILES, the computational grid filters the small, unresolved scales. The modeling of the unresolved scales is then achieved *implicitly* through the nonlinear dissipation embedded in the high-resolution, high-order numerical schemes used to discretize the convective terms. There is a significant body of work in the literature, both theoretical and numerical, explaining ILES methods and demonstrating their accuracy in turbulent flows.²²

The present study uses the Harten–Lax–van Leer contact (HLLC) approximate Riemann solver with a modified 11th-order weighted-essentially-non-oscillatory (WENO) reconstruction scheme.^{21,23} High-order WENO schemes have proven resilient to the low-Mach number dissipation associated with compressible solvers²⁴ in contrast to Monotonic Upstream-centered Scheme for Conservation Laws (MUSCL)²⁵, which generally require a low-Mach number treatment such as that proposed in the references.²⁶ Moreover, the phase/component variables are reconstructed according to²⁷ to avoid spurious numerical oscillations from occurring at interfaces between fluids of different heat capacity ratios ($\gamma_i \neq \gamma_j$) due to the use of the four-equation model or otherwise, the conservative mass-fraction transport equation (Eq. (8) is considered a diffuse-interface method).¹⁹

Although the computational cost increases with the order of the scheme for the same grid resolution, the computational cost incurred by high-order methods is offset by the increase in accuracy obtained.²⁴ The above studies have shown that lower-order methods require much finer grids to attain a similar precision as high-order methods. Moreover, high-order schemes scale more favorably over large numbers of cores and nodes, further narrowing the difference in the apparent computational expense.

A fourth-order central scheme is used for the viscous terms. The solution is advanced in time using a five-stage (fourth-order accurate) optimal strong-stability-preserving Runge–Kutta method.²⁸ Further details about the code can be found in Refs.^{9,21} and references therein.

C. Data curation

The light-blue shaded regions are the AC unit inlets, pushing cold ($\sim 18^\circ\text{C}$) air in at 4 m/s with a relative humidity of 40% and an angle of 45° degrees outward (towards the room walls). The total inlet mass-flux is (counter-) balanced by the light-yellow shaded outlet located in the middle. The initial room temperature is 25°C with a 60% relative humidity.

Probes are positioned every half meter in the x direction and every 40 centimetres in the y —and z directions for 728 probes.

The 13 positions in the x -direction are:

$$x(i) = [0.01, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 5.99]^T$$

the seven positions in the y -direction:

$$y(j) = [0.01, 0.4, 0.8, 1.2, 1.6, 2.0, 2.39]^T$$

and the eight positions in z -direction:

$$z(k) = [0.01, 0.4, 0.8, 1.2, 1.6, 2.0, 2.4, 2.79]^T$$

Note that the ± 0.01 meters at the start/end of the above coordinate vectors are essentially the cell center locations of the first computational cells from the room boundaries (walls). During the 90 seconds of the simulated flow, air properties were recorded at a regular interval at a sampling rate of almost ~ 111 samples per second, or, in other words, a sampling frequency of 111 Hz.

Figure 3 plots the temporal variation of the temperature and relative humidity as measured at three probes. The three probe locations were chosen due to their importance, which will be discussed later in the present study.

Of the three probes, probe two is positioned closest to the AC unit and, as a result, is affected not only the most (in terms of changes in magnitude) but also the earliest. Moreover, the magnitude of the oscillations indicates the much greater turbulent state of the air currents in the room that are closer to the AC unit outlets. By the end of the first 90 s, the air around probe 2 is almost 1°C cooler and approximately 1.5% drier. Importantly, Fig. 3 clearly illustrates the significant differences in the air properties between different locations inside the room—at least during the first minute and a half after the AC is switched on, for the present room and AC configuration considered.

III. CLUSTERING USING UNSUPERVISED LEARNING

A. Unsupervised Learning Method

Unsupervised learning (UL) represents a core paradigm in machine learning that has remained consistently relevant in a wide array of research fields and practical applications, including image analysis^{29,30}, sleep stage categorization³¹, and detection of mechanical damage.³² Despite their versatility, UL algorithms encounter specific obstacles in particle clustering tasks, such as sensitivity to hyperparameters, challenges with feature selection and dimensionality, and limitations in scalability, among other issues.¹⁷ As implemented in this study, the RUN-ICON algorithm addresses these challenges by enhancing confidence and minimizing uncertainty in selecting dominant clusters.¹⁶

A notable advantage of RUN-ICON over other UL approaches lies in its systematic approach to identifying the optimal number of clusters, reducing the need for intuitive or subjective decision-making. Unlike other methods that rely on

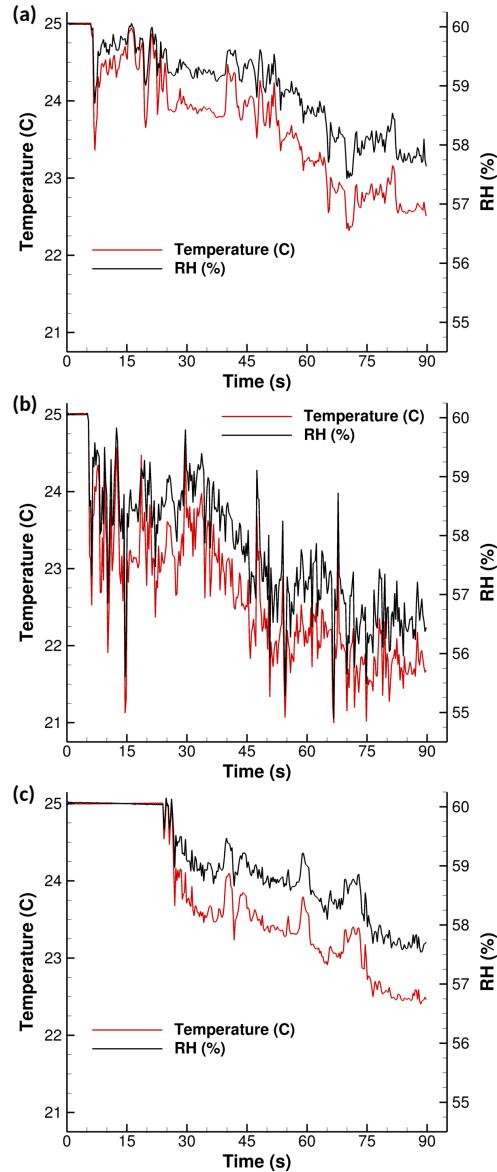


FIG. 3. Temperature ($^{\circ}\text{C}$) and relative humidity (RH%) over time as measured at three probe locations: (a) Probe 1: $[x, y, z] = [1.5, 0.8, 1.2]$ m, (b) Probe 2: $[x, y, z] = [3.0, 1.2, 1.6]$ m, and (c) Probe 3: $[x, y, z] = [5.0, 1.2, 1.6]$ m.

subjective criteria to determine the best cluster count, RUN-ICON establishes a structured process to select the optimal cluster number, ensuring robust and repeatable results. The algorithm identifies dominant cluster centers across multiple iterations of the K-means++ method without dependence on the Sum of Squared Errors. It introduces novel metrics, namely the Clustering Dominance Index (CDI) and Uncertainty. CDI reflects the likelihood of a specific clustering pattern based on its frequency across multiple attempts at a given cluster count. Uncertainty is measured by the relative difference between the upper and lower CDI limits for a given configuration, representing the maximum variance from the average for that particular setup. The algorithm thus focuses on identifying stable clusters that reveal the genuine structure within the data, enhancing reliability by reducing the impact of outliers or noise. This approach results in clustering outcomes that are more robust and interpretable. The steps of the algorithm are as follows:

1. Execute the K-means++ clustering algorithm 100 times for a specified number of clusters in each iteration.
2. After each execution, determine the coordinates of the cluster centers.
3. Compare the cluster center coordinates from all 100 runs to pinpoint those coordinates that occur most frequently.
4. Select the cluster centers with the highest frequency as the dominant centers for the specified number of clusters and compute the CDI.
5. Repeat steps 1-4 nine additional times, resulting in ten sets of dominant cluster centers.
6. Calculate the average frequency of occurrence and CDI for these dominant centers over the ten repetitions, ensuring consistency across clustering centers. Determine the upper and lower bounds of these metrics and examine the variance.
7. Repeat steps 1-6 for different cluster counts, starting from 3 and progressing up to 10 clusters.
8. Identify the cluster count with the highest average CDI across all clustering attempts (from 3 to 10 clusters), select the corresponding dominant cluster centers, and ensure that the variance between the upper and lower bounds remains below 30%.
9. Choose the cluster count with the highest average CDI and minimal variance as the optimal number of clusters for RUN-ICON.

The advantages of the RUN-ICON algorithm compared to other UL techniques were demonstrated in a prior study.¹⁶ Its reliability and interpretability were validated across various computational datasets where the intended clustering patterns were predefined. The algorithm achieved approximately 97% accuracy in detecting the expected clusters, compared to

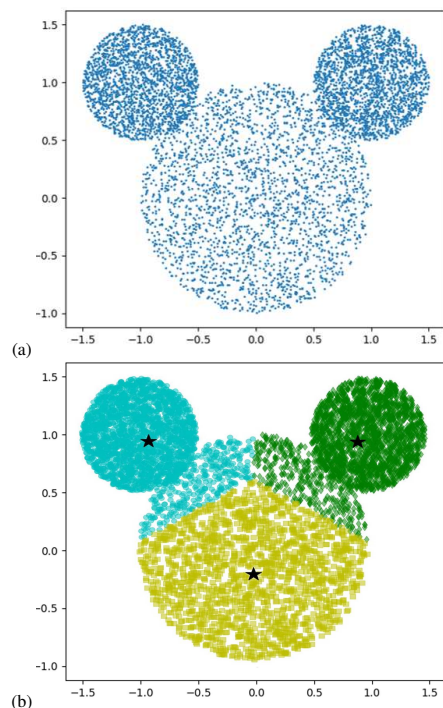


FIG. 4. The Mickey Mouse dataset: (a) "face" and "ears", as generated in three circles with random points within every circle, (b) predicted dominant clustering by RUN-ICON algorithm. Each colour indicates points belonging to the same cluster, and the stars indicate the cluster centers.

80% accuracy from other methods. To offer another example of the algorithm's ability to identify dominant clusters, we generated for the present study a dataset of 7,500 points in (x, y) space, which represented the head of Mickey Mouse, i.e., one central circle representing the "face" and two smaller circles, symmetrically attached to the bigger one, representing the "ears". All data points within the three circles were generated randomly (Figure 4(a)). It is expected that RUN-ICON should predict three dominant clusters, thus separating the "face" from the "ears". Indeed, the algorithm predicted as dominant the separation in three clusters, as shown in Figure 4(b), with high confidence (97%) and low uncertainty (2%). The two smaller clusters resemble Mickey's "ears", positioned symmetrically above the larger central cluster that forms the "face". All other clustering configurations were predicted with confidence less than 40%.

Additionally, RUN-ICON was tested on mathematically simulated particle dispersions¹⁷, where particles were randomly distributed in space. RUN-ICON identified dominant clustering structures, unlike other UL techniques, which failed

to do so. Furthermore, the combination of CFD and RUN-ICON either independently³³ or in conjunction with Gaussian Mixture Models¹³, showcased the capabilities of the algorithm to identify meaningful particle clusters effectively. These successes highlight the algorithm's potential for application in complex flow scenarios and realistic contexts.

IV. ANALYSIS OF TEMPORAL FLOW DYNAMICS IN A CONTROLLED ROOM ENVIRONMENT

This work aims to identify distinct comfort zones within a room by applying the RUN-ICON algorithm for the flow described in Section II, driven by variations in temperature, humidity, and airflow patterns. By categorizing regions based on similar flow characteristics, we can define representative clusters that reflect different comfort levels experienced within the space. This clustering approach highlights specific comfort zones and helps pinpoint optimal, representative locations for future measurements. These identified positions will serve as reference points for similar studies in other rooms, enabling efficient and targeted data collection without needing to instrument each room extensively. This streamlined method provides a transferable framework for assessing comfort zones and air distribution in diverse indoor environments.

A. Data Pre-processing and Derived Metrics

We utilize computational data from 728 location probes throughout the room, recording measurements of the x , y , and z components of velocity and temperature and relative humidity. We analyze them at three key time points (40, 60, and 90 seconds). These time points capture instances as the turbulent flow evolves, influenced by the introduced cooler air. This setup enables a detailed examination of how air movement and parameter distributions change over time.

To distill the complexity of the collected data, we focus, apart from temperature and relative humidity, on two derived parameters that serve as robust indicators of flow structure and similarity across regions. The first parameter is the Euclidean distance (L_2 norm) from the room's origin (bottom left inner corner, as shown in Figure 1), which represents spatial proximity and allows us to account for potential regions of influence from the air conditioning source. This parameter provides a measure of spatial grouping critical for understanding how regions relate based on distance, potentially sharing similar flow characteristics.

The second parameter is the magnitude of the velocity vector, calculated from the x , y , and z velocity components. This measure serves as an indicator of the local intensity of airflow. By analyzing velocity magnitudes, we gain insights into higher and lower air momentum zones, which often reflect the degree of turbulence in different regions. The L_2 norm of distances and the velocity magnitudes offer complementary perspectives for delineating distinct regions within the room, potentially enabling the categorization of flow areas with UL based on spatial and dynamic flow attributes.

These four chosen parameters allow for a focused, interpretable analysis. They contribute to a better understanding of the underlying flow structure and enable meaningful categorization of flow regions influenced by the air conditioning dynamics and room geometry.

B. Statistical Analysis of Flow Variables

The Spearman rank correlation was calculated to analyze relationships between the four chosen variables (namely, L_2 norm, velocity magnitude, temperature and relative humidity) within the flow data and its corresponding p -value for the three-time points. The Spearman rank correlation r is a non-parametric measure that assesses the strength and direction of a monotonic relationship between two variables, regardless of the distribution or linearity of the data.³⁴ Spearman's correlation coefficient ranges from -1 to +1, where values close to +1 or -1 indicate a strong positive or negative association, respectively, and values near zero suggest no monotonic relationship. The p -value associated with each Spearman correlation quantifies the probability that the observed correlation could occur by random chance. A low p -value (typically below 0.05) indicates a statistically significant relationship, suggesting that the correlation is unlikely due to random variation and that there is a meaningful association between the variables analyzed.³⁵ This approach provides insight into variable interdependencies within the flow regions, guiding further analysis and interpretation of comfort zones. In Tables I-VI, the Spearman rank correlation matrices and their corresponding p -value matrices for the L_2 norm D , velocity v , temperature T and relative humidity RH at 40 s, 60 s and 90 s are presented:

TABLE I. Spearman Rank Correlation Matrix at 40 s for L_2 norm, velocity, temperature and relative humidity.

	D	v	T	RH
D	1.000	-0.030	-0.007	0.011
v	-0.030	1.000	-0.431	-0.432
T	-0.007	-0.431	1.000	0.978
RH	0.011	-0.432	0.978	1.000

TABLE II. p -values Matrix at 40 s for L_2 norm, velocity, temperature and relative humidity.

	D	v	T	RH
D	0.000	0.413	0.852	0.763
v	0.413	0.000	0.000	0.000
T	0.852	0.000	0.000	0.000
RH	0.763	0.000	0.000	0.000

Next, we discuss the key observations and interpret the correlations while considering the impact of the air conditioning unit's operation.

TABLE III. Spearman Rank Correlation Matrix at 60 s for L_2 norm, velocity, temperature and relative humidity.

	D	v	T	RH
D	1.000	-0.079	0.135	0.138
v	-0.079	1.000	-0.463	-0.452
T	0.135	-0.463	1.000	0.991
RH	0.138	-0.452	0.991	1.000

TABLE IV. p -values Matrix at 60 s for L_2 norm, velocity, temperature and relative humidity.

	D	v	T	RH
D	0.000	0.034	0.000	0.000
v	0.034	0.000	0.000	0.000
T	0.000	0.000	0.000	0.000
RH	0.000	0.000	0.000	0.000

TABLE V. Spearman Rank Correlation Matrix at 90 s for L_2 norm, velocity, temperature and relative humidity.

	D	v	T	RH
D	1.000	-0.031	0.073	0.077
v	-0.031	1.000	-0.436	-0.430
T	0.073	-0.436	1.000	0.992
RH	0.077	-0.430	0.992	1.000

TABLE VI. p -values Matrix at 90 s for L_2 norm, velocity, temperature and relative humidity.

	D	v	T	RH
D	0.000	0.406	0.048	0.039
v	0.406	0.000	0.000	0.000
T	0.048	0.000	0.000	0.000
RH	0.039	0.000	0.000	0.000

1. Correlation Between Temperature and Relative Humidity

In all time instances (40s, 60s, and 90s), there is a consistently high positive correlation between temperature and relative humidity with strong statistical significance. At 40 s, $r = 0.978$ with $p = 0.000$; At 60 s, $r = 0.991$ with $p = 0.000$; At 90 s, $r = 0.992$ with $p = 0.000$. This strong correlation is expected because temperature and humidity are often interrelated in a controlled environment. Initially, both temperature and humidity were higher in the room, likely due to external environmental factors and the room's insulation properties. As the air conditioner begins to function, it simultaneously reduces the temperature and relative humidity. The positive correlation between T and RH suggests that the cooling and drying effects of the air conditioner are occurring in tandem, maintaining a relationship between the two variables. Cooler air holds less moisture, resulting in a drop in temperature and humidity, reinforcing the close relationship between these two

variables.

2. Negative Correlation Between Velocity and Temperature/Relative Humidity

Another notable pattern in the matrices is the consistent negative correlation between velocity (v), temperature (T), and relative humidity (RH), again with strong statistical significance. At 40 s, the correlations are $r = -0.431$ for T with $p = 0.000$, and $r = -0.432$ for RH with $p = 0.000$. At 60 s, the correlations are $r = -0.463$ for T with $p = 0.000$, and $r = -0.452$ for RH with $p = 0.000$. At 90 s, the correlations are $r = -0.436$ for T with $p = 0.000$, and $r = -0.430$ for RH with $p = 0.000$. This negative correlation reflects that the HVAC introduces air at a higher velocity than the room's initial, relatively stagnant conditions. The incoming conditioned air is colder and drier than the existing room air. This negative correlation suggests that regions of the room experiencing higher air velocity are closer to the cooling source, where the air conditioner's actions have significantly reduced the temperature and humidity.

3. Distance and Flow Dynamics

The matrices also show relatively weak correlations between distance parameter D and velocity v , temperature T or humidity RH . At 40 s for v , T , and RH , $r = -0.030$, $r = -0.007$, and $r = 0.011$ respectively, with corresponding p -values of 0.413, 0.852, and 0.763. At 60 s, D has correlations of $r = -0.079$ with v ($p = 0.034$), $r = 0.135$ with T ($p = 0.000$), and $r = 0.138$ with RH ($p = 0.000$). At 90 s, D shows correlations of $r = -0.031$ with v ($p = 0.406$), $r = 0.073$ with T ($p = 0.048$), and $r = 0.077$ with RH ($p = 0.039$). The small correlations between distance and velocity, temperature or humidity suggest that, over time, the spatial distribution of velocity, temperature and humidity in the room is becoming more uniform. In the initial moments, the air conditioner's effect is most pronounced near the source of conditioned air (closer to the AC unit), leading to a more noticeable difference in velocity, temperature and humidity at various distances. Over time, however, the air becomes more evenly distributed across the room, reducing their gradient with respect to distance.

The strong positive correlation between temperature and relative humidity at all time steps highlights the interdependent nature of these two variables in a room subjected to air conditioning. The consistently negative correlation between velocity and temperature and relative humidity reflects the role of the air conditioner in displacing warmer, moister air with cooler, drier air. These findings align with the expected physical behavior of the system. As higher-velocity air enters the room, it lowers both temperature and humidity, creating gradients that are eventually smoothed out over time as the room reaches a more uniform state.

C. Cluster-Based Analysis for Optimal Probe Placement

The study aims to identify areas in the room with similar temperature and velocity patterns, allowing monitoring with fewer probes. Clustering was carried out using the RUN-ICON algorithm, focusing on temperature and velocity relative to distance while excluding relative humidity because it heavily depends on temperature. The goal is to demonstrate that the dominant clusters share similar distances from the origin (measured by the L_2 norm) across equivalent spatial positions in all cases. For this reason, a radius of influence for each cluster center is determined by calculating the standard deviation of the spatial coordinates. This allows us to verify that the corresponding predicted centers fall within each other's radius of influence. As demonstrated, by limiting spatial representation to the L_2 norm of distances, clustering reliability increased, achieving 70-90% confidence levels across clusters.

Since predicted clusters have clear spatial boundaries, non-linear programming can determine the exact x , y , and z coordinates for flow monitoring at these locations. For this reason, the Sequential Least Squares Programming (SLSQP) algorithm was employed to convert the L_2 norms back to x , y , and z coordinates, constrained by each cluster's spatial range through the formulation of an objective function from the equality:

$$x^2 + y^2 + z^2 = L_2^2 \quad (10)$$

SLSQP is an iterative optimization algorithm commonly used for solving constrained optimization problems³⁶. It can handle both linear and non-linear constraints and bounds on variables. SLSQP operates by approximating the original problem as a series of quadratic programming subproblems and sequentially improves the solution's estimate using first-order derivatives of the objective function and constraints. This makes it a robust and efficient method for handling non-linear programming problems.

D. Results

The clustering analysis provided insights into the room's spatial and thermal dynamics at different times, influencing probe placement recommendations.

1. Time: 40 seconds

Temperature-distance and velocity-distance clustering revealed three distinct zones (see Fig. 5) with (confidence, uncertainty): (62.6%, 15%) and (97%, 2%), respectively. The three clusters are denoted by the indices (1, 2, 3), corresponding to the room's left, middle, and right sections, respectively, as illustrated in Fig. 1. The radii of influence (i.e., the standard deviation, in m) for the three clusters (1, 2, 3) were (0.65, 0.52, 0.57) and (0.64, 0.51, 0.58), respectively. Using SLSQP optimization, cluster positions were calculated within each cluster. The constraints in x , y , and z are the same for temperature

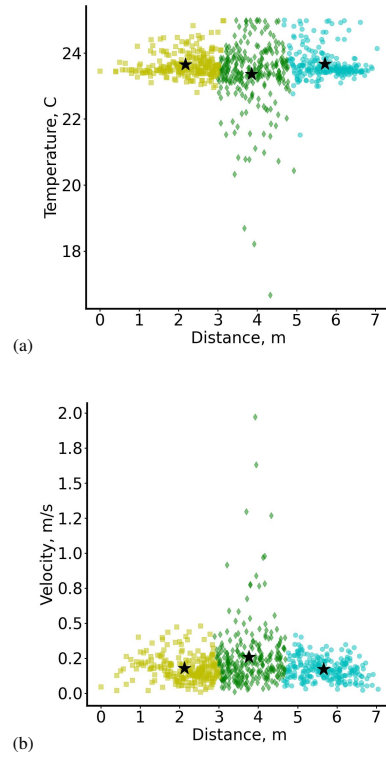


FIG. 5. Clustering at time 40 s from the start of the simulation. The x -axis is the distance from the origin (L_2 norm) in m, and the y -axis is either (a) Temperature (in °C) or (b) Velocity (in m/s). Each color indicates points of the room belonging to the same cluster and the stars indicate the cluster centers.

and velocity. They only differ in x , based on the cluster position, whether closer to the side walls or the center of the room. In this way, the constraints in the x -direction are given in Table VII, while the constraints in y and z are given in Table VIII. The objective functions for temperature and velocity result from Eq. (10) with L_2 as recorded in Table IX.

TABLE VII. Constraints in the x -direction for the three dominant clusters at 40 sec.

Cluster	Lower bound (m)	Upper bound (m)
1	0.5	2.9
2	2.5	4.5
3	3.1	5.5

In this way, we obtain the optimal cluster coordinates for

TABLE VIII. Constraints in the y - and z -directions for the three dominant clusters at 40 sec.

Axis	Lower bound (m)	Upper bound (m)
y	0.0	2.4
z	0.0	2.8

TABLE IX. L_2 values for the three dominant clusters at 40 sec.

Cluster	L_2 for temperature	L_2 for velocity
1	2.17	2.12
2	3.86	3.76
3	5.72	5.66

temperature as: cluster 1 $(x, y, z) = (1.59, 0.96, 1.12)$ m, cluster 2 $(x, y, z) = (3.28, 1.31, 1.53)$ m, and cluster 3 $(x, y, z) = (5.42, 1.18, 1.38)$ m. Similarly, the optimal cluster coordinates for velocity are predicted as: cluster 1 $(x, y, z) = (1.56, 0.94, 1.09)$ m, cluster 2 $(x, y, z) = (3.22, 1.29, 1.50)$ m, and cluster 3 $(x, y, z) = (5.37, 1.17, 1.37)$ m.

2. Time: 60 seconds

Temperature-distance clustering revealed, as before, three distinct zones (see Fig. 6(a)) with (confidence, uncertainty): (75%, 6%). At this time, velocity-distance clustering revealed an additional cluster, likely resulting from localized turbulence, which divided the central region of the room into two distinct areas, as shown in Fig. 6(b). The confidence and uncertainty for this clustering were (87%, 7%). However, the minor central cluster, characterized by higher velocities, contained only 17 points compared to the major central cluster's 255 points. Consequently, given the high confidence and low uncertainty associated with separating into three dominant clusters (73%, 14%), the minor cluster was disregarded for simplicity, and its points were merged into the major central cluster. The radii of influence for the three clusters (from left to right, as seen in Fig. 6) were (0.66, 0.49, 0.61) m and (0.62, 0.53, 0.57) m, for temperature and velocity, respectively. Using SLSQP optimization, cluster positions were calculated within each cluster. The constraints in x , y , and z are the same as in the 40 s case. The objective functions differ due to the distances predicted for the cluster centers' locations. The objective functions for temperature and velocity result from Eq. (10) with L_2 as recorded in Table X.

In this way, we obtain the optimal cluster coordinates for

TABLE X. L_2 values for the three dominant clusters at 60 s.

Cluster	L_2 for temperature	L_2 for velocity
1	2.19	2.08
2	3.80	3.71
3	5.63	5.64

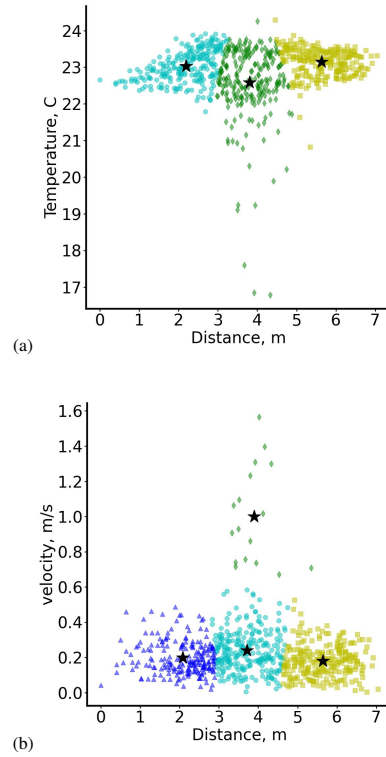


FIG. 6. Clustering at time 60 s from the start of the simulation. The x -axis is the distance from the origin (L_2 norm) in m, and the y -axis is either (a) Temperature (in $^{\circ}\text{C}$) or (b) Velocity (in m/s). Each color indicates points of the room belonging to the same cluster and the stars indicate the cluster centers.

temperature as cluster 1 $(x, y, z) = (1.69, 0.81, 0.95)$ m, cluster 2 $(x, y, z) = (3.06, 1.47, 1.71)$ m, and cluster 3 $(x, y, z) = (4.88, 1.17, 1.37)$ m. Similarly, the optimal cluster coordinates for velocity are predicted as cluster 1 $(x, y, z) = (1.54, 0.93, 1.08)$ m, cluster 2 $(x, y, z) = (3.15, 1.26, 1.47)$ m, and cluster 3 $(x, y, z) = (5.25, 1.26, 1.47)$ m.

3. Time: 90 seconds

Higher confidence (82% for temperature and 97% for velocity) and lower uncertainty levels (9% for temperature and 2% for velocity) validated the clustering results, establishing three optimal clusters, consistent with the earlier clustering results (see Fig. 7). The radii of influence (in m) for the three clusters (from left to right, as seen in Fig. 7) were (0.64, 0.51,

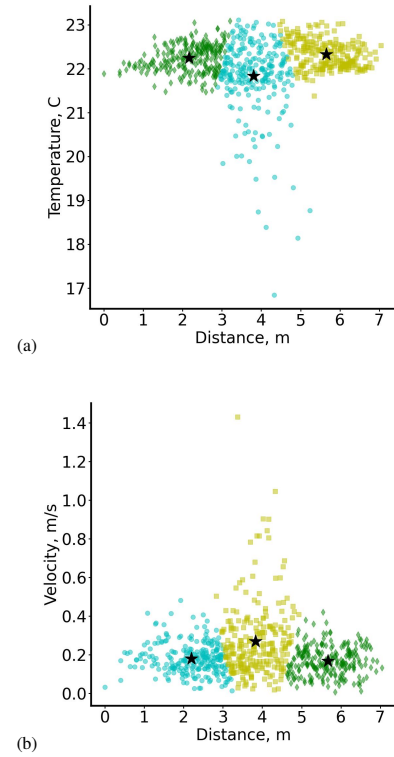


FIG. 7. Clustering at time 90 s from the start of the simulation. The x -axis is the distance from the origin (L_2 norm) in m, and the y -axis is either (a) Temperature (in $^{\circ}\text{C}$) or (b) Velocity (in m/s). Each color indicates points of the room belonging to the same cluster and the stars indicate the cluster centers.

0.58) and (0.63, 0.52, 0.55) for temperature and velocity, respectively. Using SLSQP optimization, cluster positions were calculated within each cluster. The constraints in x , y , and z are the same as in the 40 and 60 s case. Only the objective functions differ due to the different distances predicted for the cluster center's location. The objective functions for temperature and velocity result from Eq.(10) with L_2 as recorded in Table XI.

TABLE XI. L_2 values for the three dominant clusters at 90 s.

Cluster	L_2 for temperature	L_2 for velocity
1	2.18	2.20
2	3.80	3.83
3	5.65	5.67

In this way, we obtain the optimal cluster coordinates for temperature as: cluster 1 $(x, y, z) = (1.54, 0.93, 1.08)$ m, cluster 2 $(x, y, z) = (3.00, 1.52, 1.77)$ m, and cluster 3 $(x, y, z) = (5.25, 1.26, 1.47)$ m. Similarly, the optimal cluster coordinates for velocity are predicted as: cluster 1 $(x, y, z) = (1.55, 0.92, 1.08)$ m, cluster 2 $(x, y, z) = (3.24, 1.29, 1.51)$ m, and cluster 3 $(x, y, z) = (5.26, 1.26, 1.47)$ m.

4. Probe group clustering identification

We identified three distinct locations within the room that can adequately characterize the airflow around the air conditioning unit and help define comfort zones. These locations were determined based on their ability to represent the flow regions at 40 s, 60 s, and 90 s after allowing sufficient time for turbulence to develop (90 s). Table XII gives the minimum-maximum values from the above analyses for each coordinate (x, y, z) and a radius of influence. The radii of influence were calculated as the mean of all standard deviations for temperature and velocity for each cluster center.

TABLE XII. Main cluster location bounds and radii of influence

Cluster	x (m)	y (m)	z (m)	radius of influence (m)
1	1.54–1.69	0.81–0.96	0.95–1.12	0.62
2	3.00–3.28	1.26–1.52	1.47–1.77	0.52
3	4.88–5.42	1.17–1.26	1.37–1.47	0.58

The maximum and minimum coordinates shown in Table XII represent the intervals within which the clusters may be identified. Notably, these intervals all fall within the calculated radius of influence for their corresponding cluster center. This confirms the stability and reliability of the predicted locations as representative points for both temperature and velocity distributions. In this way, we may select the mean values for each location's coordinates as representative positions. For the given room dimensions, these are given in Table XIII. Figure 8 illustrates the main clusters' positions as centers of spheres with radii equal to the corresponding radii of influence, as given in Table XII.

TABLE XIII. Main cluster locations

Location	x (m)	y (m)	z (m)
1	1.58	0.92	1.10
2	3.16	1.36	1.58
3	5.24	1.22	1.42

In our data, the closest probe positions corresponding to these locations are given in Table XIV.

These selected positions provide an effective and simplified representation of airflow within the room, particularly for defining comfort zones under the influence of the air conditioning unit.

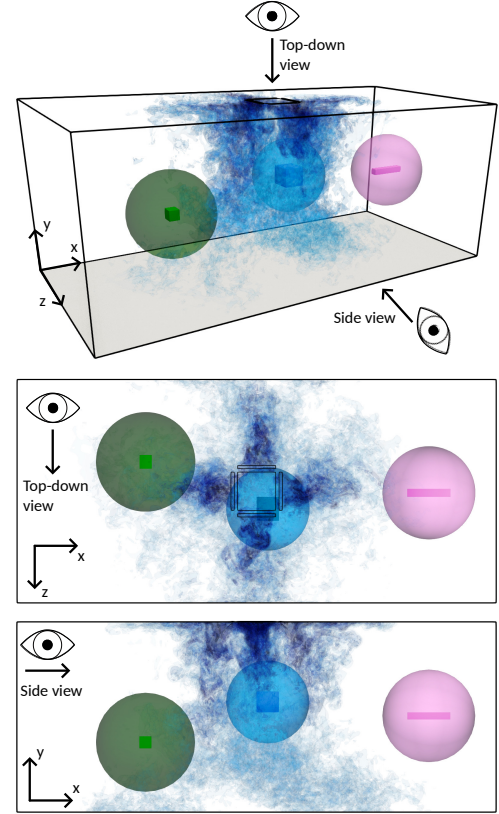


FIG. 8. Illustration of the three clusters' positions as centers of spheres with radii equal to the corresponding radii of influence. Cluster 1: green coloured, Cluster 2: blue coloured, and Cluster 3: pink coloured shapes.

TABLE XIV. Closest probe positions

Probe positions	x (m)	y (m)	z (m)
1	1.5	0.8	1.2
2	3.0	1.2	1.6
3	5.0	1.2	1.6

E. Turbulence analysis at probe positions

At the identified probe points, turbulent characteristics were evaluated through Reynolds number, Re , and Turbulence Intensity, TI . The velocity magnitude time series, from 0 to 90 s, was retrieved from the data set for each of the three probes. Using these velocity data, the mean and standard deviation were determined. Then, Re and TI were calculated as

follows³⁷:

$$Re = \rho \bar{u} L / \mu \quad (11)$$

where $\rho = 1.2 \text{ kg/m}^3$ is the density of air; $\bar{u}$ is the mean of the velocity magnitude (m/s); $L = 3 \text{ m}$ is the characteristic length, chosen as half the x -dimension of the room (comparable to the room's height and width); $\mu = 1.81 \times 10^{-5} \text{ Pa}\cdot\text{s}$ is the dynamic viscosity of the air.

The turbulence intensity, TI , can be obtained—in terms of percentage—according to:

$$TI = \sigma / \bar{u} \times 100 \quad (12)$$

where σ is the standard deviation of the velocity (m/s).

The analysis revealed distinct flow characteristics at each probe location: (i) The left side of the room: The Reynolds number found is $Re = 40,375$, with a turbulence intensity of $TI = 58\%$; (ii) The right side of the room: The Reynolds number is $Re = 28,492$, with a turbulence intensity of $TI = 54\%$; and (iii) Intermediate region, near the air conditioner: The Reynolds number is much higher at $Re = 70,460$, while the turbulence intensity is slightly lower at $TI = 47\%$. These results indicate varying flow characteristics across the room. On the left side, the higher turbulence intensity reflects more pronounced velocity fluctuations. The Reynolds number and turbulence intensity are lower on the right side, suggesting a more stable flow. The flow exhibits a much higher Reynolds number in the intermediate region, driven by increased velocities, indicating the dominance of inertial forces over viscous forces. Despite this, the lower turbulence intensity suggests reduced velocity fluctuations, marking a transition towards laminar behavior and smoother flow.

F. Discussion

The findings from this study offer significant insights into probe placement optimization within an air-conditioned environment where turbulent airflow is modelled. By applying clustering techniques and CFD simulations, we could map the room's temperature, humidity, and velocity distributions, identifying regions with distinct thermal and flow characteristics. This section explores the practical implications of these results, the effectiveness of the methodology, and potential avenues for further research and application.

The study demonstrates that airflow in an enclosed room environment, especially one with centrally located air conditioning, does not distribute uniformly. Instead, due to the turbulent flow generated by the temperature and velocity differential of the incoming air, pockets of varying temperature and velocity develop within the space. For HVAC system design, this information is essential, as it highlights the non-uniform cooling distribution and potential zones where additional airflow control may be necessary to maintain occupant comfort. A significant outcome is the relationship between the Reynolds number and Turbulence Intensity at the selected probe locations. Higher Re values near the air-conditioning outlet indicate strong, inertial-dominated airflow with lower

TI , suggesting that air in these zones is less affected by viscous forces and maintains a stable, directional flow. Such areas offer a smoother and more consistent airflow, which generally translates to higher thermal comfort for occupants. On the other hand, lower Re values and higher TI levels were recorded closer to the room's boundaries, where airflow is more disrupted. This suggests that edge areas of the room might experience drafts or fluctuating air speeds, potentially causing discomfort. For HVAC systems, understanding these distinctions in airflow behavior can be pivotal in developing strategies that mitigate uncomfortable drafts and optimize cooling patterns.

The RUN-ICON clustering approach applied in this study proved effective in identifying probe positions that adequately capture the spatial and thermal variations within the room. By categorizing regions based on temperature-velocity profiles, the clustering approach reduced the number of probes needed while allowing comprehensive data collection across the entire space. Notably, using the L_2 norm effectively reduced spatial dimensionality and focused on meaningful positional relationships relative to the room's origin. Sequential Least Squares Programming (SLSQP) for determining probe coordinates within each cluster further contributed to the optimization by aligning probe placement with calculated cluster centers while maintaining a margin of uncertainty around each probe position. This refinement ensures that the selected positions provide adequate coverage and account for spatial variability within each cluster, which is vital in turbulent flow conditions. This clustering-based approach to probe placement has practical implications for various fields beyond HVAC. In industrial settings, for example, where monitoring airflow, temperature, or humidity is crucial, optimized probe placement can reduce installation and maintenance costs by focusing probes on key areas. The methodology applied here also holds potential for environments where space constraints limit the number of probes that can be installed, as it provides a robust, data-driven method for choosing probe locations with maximum efficiency.

RUN-ICON ensures robustness by iteratively identifying stable clusters and employing novel metrics such as the Clustering Dominance Index (CDI) and Uncertainty to evaluate the consistency of clustering outcomes. These features minimize the impact of noise and outliers while maintaining reliable performance across diverse scenarios. Furthermore, the integration of CFD simulations enhances the robustness of the methodology by providing high-fidelity data that capture the intricate dynamics of turbulent airflow. The high accuracy of the algorithm has been demonstrated in previous works, where the algorithm was compared to other UL methods and metrics.^{16,17}

Interpretability is a key strength of this approach. RUN-ICON aligns clustering outcomes with fundamental physics by focusing on physically meaningful variables, such as velocity, temperature, and spatial proximity. This alignment allows clear insights into airflow patterns, facilitating the optimization of probe placements. Additionally, the structured outputs of the clustering algorithm, including confidence intervals and cluster boundaries, enhance its accessibility for

engineering applications.

The scalability of this methodology is demonstrated not only in its ability to handle large datasets, such as those from 728 probes but also in its adaptability to spaces of different dimensions. Table XV provides a general formula for scaling the positions of the three identified cluster centers based on room dimensions. Using this proportional approach, we demonstrate its application to a new room with dimensions of $10.0 \text{ m} \times 3.0 \text{ m} \times 4.0 \text{ m}$ in the example following the table.

TABLE XV. Generic formula for scaling cluster positions based on room dimensions.

Probe positions	x (m)	y (m)	z (m)
1	$\frac{1.58}{6.0} \times \text{Length}$	$\frac{0.92}{2.4} \times \text{Height}$	$\frac{1.10}{2.8} \times \text{Width}$
2	$\frac{3.16}{6.0} \times \text{Length}$	$\frac{1.36}{2.4} \times \text{Height}$	$\frac{1.58}{2.8} \times \text{Width}$
3	$\frac{5.24}{6.0} \times \text{Length}$	$\frac{1.22}{2.4} \times \text{Height}$	$\frac{1.42}{2.8} \times \text{Width}$

For a room measuring $(x,y,z) = (10.0, 3.0, 4.0) \text{ m}$, the scaled positions of the three probes are calculated as: probe 1 $(x,y,z) = (2.63, 1.15, 1.57) \text{ m}$, probe 2 $(x,y,z) = (5.27, 1.70, 2.26) \text{ m}$, probe 3 $(x,y,z) = (8.72, 1.53, 2.03) \text{ m}$.

To further validate the scalability of the proposed methodology, future work will focus on testing these proportional cluster placements in rooms of various dimensions and configurations. These tests will help determine whether the scaled positions reliably capture airflow and thermal characteristics in different settings, ensuring the robustness and versatility of the approach for real-world applications.

While the clustering and optimization techniques applied here demonstrate significant potential, we may identify areas where further consideration should be given. Firstly, this study assumes a static setup of the air-conditioning unit and room boundaries. In real-world scenarios, factors such as occupant movement, furniture placement, and other obstructions can alter airflow patterns, affecting temperature and velocity distributions. Future studies could incorporate dynamic factors or real-time adjustments to the clustering analysis based on updated flow data, potentially making the probe placement even more responsive to environmental changes. Additionally, although the CFD simulations provide a high-fidelity representation of turbulent flow, certain assumptions—such as uniform air properties and idealized boundary conditions—may differ from actual conditions. Including a broader range of boundary conditions or using experimental validation with simulations could further refine these findings. Lastly, although excluding relative humidity (RH) from the clustering criteria was justified due to its correlation with temperature, there may be situations where RH changes independently of temperature (e.g., in humid or damp environments). Future work could explore how additional environmental factors impact clustering confidence and probe placement recommendations.

V. CONCLUSIONS

This study presents an innovative approach for optimizing probe placement in air-conditioned spaces by integrating computational fluid dynamics (CFD) simulations with the RUN-ICON clustering algorithm. The method involves clustering distinct thermal and velocity zones and strategically positioning probes based on temperature and velocity distributions. This strategy minimizes redundancy and maximizes spatial coverage, offering a practical solution for monitoring indoor environments where airflow and temperature control are critical, particularly in HVAC systems that enhance occupant comfort and energy efficiency.

The findings reveal the non-uniform nature of airflow in enclosed spaces, influenced by turbulent flow patterns and thermal gradients. Higher Reynolds numbers (Re) near air-conditioning outlets are associated with stable airflow and low turbulence intensity (TI), favoring thermal comfort. In contrast, lower Re and higher TI near boundaries can disrupt airflow, potentially causing drafts and discomfort. These insights are crucial for HVAC design, enabling engineers to target specific airflow characteristics, such as maintaining high Re in seating areas for stability and minimizing high TI near occupants to reduce drafts, enhancing comfort and efficiency.

The study's optimized probe placements provide a framework for efficient environmental monitoring, reducing the probes needed while maintaining comprehensive data collection. Using the L_2 norm and Sequential Least Squares Programming facilitated the precise alignment of probes with cluster centers, accounting for spatial variability. These advancements can cut costs and improve maintenance efficiency in HVAC systems. Beyond HVAC, the methodology can be applied in industrial settings, allowing effective monitoring of temperature, velocity, and humidity with fewer resources. Further research incorporating dynamic factors, real-time validation, and the inclusion of relative humidity (RH) could refine the approach, expanding its applicability to various environmental conditions and enhancing its responsiveness to real-world scenarios.

VI. ACKNOWLEDGEMENTS

This paper is supported by the European Union's Horizon Europe Research and Innovation Actions programme under grant agreement No 101069937, project name: HS4U (HEALTHY SHIP 4U). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure, and Environment Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

VII. AUTHOR DECLARATIONS

The authors have no conflicts to disclose.

VIII. DATA AVAILABILITY

The authors' data supporting this study's findings are available upon reasonable request. The RUN-ICON algorithm in Python is freely available at <https://github.com/ASI-UNIC/RUN-ICON/>

REFERENCES

- ¹E. Ono, K. Mihara, K. Lam, and A. Chong, "The effects of a mismatch between thermal comfort modeling and hvac controls from an occupancy perspective," *Building and Environment* **220**, Article 109255 (2022).
- ²J. Allen, P. MacNaughton, J. Laurent, S. Flanagan, E. Eitland, and J. Spengler, "Green buildings and health," *Current Environmental Health Reports* **2**, 250–258 (2015).
- ³X. Zhao, Y. Yin, Z. He, and Z. Deng, "State-of-the-art, challenges and new perspectives of thermal comfort demand law for on-demand intelligent control of heating, ventilation, and air conditioning systems," *Energy and Buildings* **295**, Article 113325 (2023).
- ⁴M. Balbis-Morejón, J. M. Rey-Hernández, C. Amaris-Castilla, E. Velasco-Gómez, J. F. San José-Alonso, and F. J. Rey-Martínez, "Experimental study and analysis of thermal comfort in a university campus building in tropical climate," *Sustainability* **12**, 8886 (2020).
- ⁵X. Duan, S. Yu, J. Chu, D. Chen, and Y. Chen, "Evaluation of indoor thermal environments using a novel predicted mean vote model based on artificial neural networks," *Buildings* **12**, 1880 (2022).
- ⁶T. Wang, X. Li, Y. Lu, L. Dong, F. Shi, and Z. Lin, "An efficient thermal comfort prediction method for indoor airflow environment using a cfd-based deep learning model," *Building and Environment* **267**, 112246 (2025).
- ⁷P. L. Johnson and M. Wilczek, "Multiscale velocity gradients in turbulence," *Annual Review of Fluid Mechanics* **56**, 463–490 (2024).
- ⁸P. Yeung, X. Zhai, and K. R. Sreenivasan, "Extreme events in computational turbulence," *Proceedings of the National Academy of Sciences* **112**, 12633–12638 (2015).
- ⁹D. Drikakis, M. Hahn, A. Mosedale, and B. Thornber, "Large eddy simulation using high-resolution and high-order methods," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **367**, 2985–2997 (2009).
- ¹⁰K. Poulinakis, D. Drikakis, I. W. Kokkinakis, and S. M. Spottswood, "Machine-learning methods on noisy and sparse data," *Mathematics* **11** (2023), 10.3390/math11010236.
- ¹¹S. Quazi, "Artificial intelligence and machine learning in precision and genomic medicine," *Medical Oncology* **39**, 120 (2022).
- ¹²T. P. Nagarhalli, V. Vaze, and N. Rana, "Impact of machine learning in natural language processing: A review," in *2021 third international conference on intelligent communication technologies and virtual mobile networks (ICICV)* (IEEE, 2021) pp. 1529–1534.
- ¹³N. Christakis and D. Drikakis, "On particle dispersion statistics using unsupervised learning and gaussian mixture models," *Physics of Fluids* **36** (2024), 10.1063/5.0229111.
- ¹⁴S. Fotopoulou, "A review of unsupervised learning in astronomy," *Astronomy and Computing* , 100851 (2024).
- ¹⁵M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, and A. J. Aljaaf, "A systematic review on supervised and unsupervised machine learning algorithms for data science," in *Supervised and Unsupervised Learning for Data Science*, edited by M. W. Berry, A. Mohamed, and B. W. Yap (Springer International Publishing, Cham, 2020) pp. 3–21.
- ¹⁶N. Christakis and D. Drikakis, "Reducing uncertainty and increasing confidence in unsupervised learning," *Mathematics* **11** (2023), 10.3390/math11143063.
- ¹⁷N. Christakis and D. Drikakis, "Unsupervised learning of particles dispersion," *Mathematics* **11** (2023), 10.3390/math11173637.
- ¹⁸D. Drikakis, I. W. Kokkinakis, and P. Turchas, "Sparsity and mixing effects in deep learning predictions of temperature and humidity," *Physics of Fluids* **36**, 085191 (2024).
- ¹⁹G. Allaire, S. Clerc, and S. Kokh, "A five-equation model for the simulation of interfaces between compressible fluids," *Journal of Computational Physics* **181**, 577–616 (2002).
- ²⁰A. W. Cook, "Enthalpy diffusion in multicomponent flows," *Physics of Fluids* **21**, 055109 (2009).
- ²¹I. W. Kokkinakis and D. Drikakis, "Implicit large eddy simulation of weakly-compressible turbulent channel flow," *Computer Methods in Applied Mechanics and Engineering* **287**, 229 – 261 (2015).
- ²²F. F. Grinstein, L. G. Margolin, and W. J. Rider, *Implicit Large Eddy Simulation: Computing Turbulent Fluid Dynamics* (Cambridge University Press, 2007) p. 578.
- ²³E. F. Toro, M. Spruce, and W. Speares, "Restoration of the contact surface in the HLL-Riemann solver," *Shock Waves* **4**, 25 – 34 (1994).
- ²⁴A. Mosedale and D. Drikakis, "Assessment of very high order of accuracy in implicit les models," *Journal of Fluids Engineering* **129**, 1497–1503 (2007).
- ²⁵B. van Leer, "Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method," *Journal of Computational Physics* **32**, 101–136 (1979).
- ²⁶B. J. R. Thornber and D. Drikakis, "Numerical dissipation of upwind schemes in low mach flow," *International Journal for Numerical Methods in Fluids* **56**, 1535–1541 (2008).
- ²⁷I. Kokkinakis, D. Drikakis, D. Youngs, and R. Williams, "Two-equation and multi-fluid turbulence models for Rayleigh–Taylor mixing," *International Journal of Heat and Fluid Flow* **56**, 233–250 (2015).
- ²⁸R. Spiteri and S. Ruuth, "A new class of optimal high-order strong-stability-preserving time discretization methods," *SIAM Journal on Numerical Analysis* **40**, 469–491 (2002).
- ²⁹J. Lee and G. Lee, "Feature alignment by uncertainty and self-training for source-free unsupervised domain adaptation," *Neural Networks* **161**, 682–692 (2023).
- ³⁰J. Lee and G. Lee, "Unsupervised domain adaptation based on the predictive uncertainty of models," *Neurocomputing* **520**, 183–193 (2023).
- ³¹Z. Mousavi, T. Yousefi Rezaii, S. Sheykhivand, A. Farzamnia, and S. Razavi, "Deep convolutional neural network for classification of sleep stages from single-channel EEG signals," *Journal of Neuroscience Methods* **324**, 108312 (2019).
- ³²K. Eltoumy, M. Goma, and X. Liang, "Unsupervised learning methods for data-driven vibration-based structural health monitoring: a review," *Sensors* **23**, 3290 (2023).
- ³³N. Christakis, D. Drikakis, K. Ritos, and I. W. Kokkinakis, "Unsupervised machine learning of virus dispersion indoors," *Physics of Fluids* **36** (2024), 10.1063/5.0188610.
- ³⁴P. Sedgwick, "Spearman's rank correlation coefficient," *BMJ* **349** (2014), 10.1136/bmj.g7327.
- ³⁵S. Greenland, S. J. Senn, K. J. Rothman, J. B. Carlin, C. Poole, S. N. Goodman, and D. G. Altman, "Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations," *European journal of Epidemiology* **31**, 337–350 (2016).
- ³⁶J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer New York, NY, 1999).
- ³⁷J. L. Lumley and H. Tennekes, *A First Course in Turbulence* (The MIT Press, 1972).