ANNALS OF GEOPHYSICS, 69, 1, A101, 2026; doi:10.4401/ag-9424
OPEN ACCESS

Encoding Gridded Atmospheric Data with Classical and Quantum Methods

David Sládek*,1

(1) Department of Military Geography and Meteorology, University of Defence, Kounicova 65, Brno, Czechia

Article history: received September 11, 2025; accepted January 7, 2026

Abstract

Accurate encoding of spatial meteorological data is critical for applying quantum machine learning (QML) in climatology and atmospheric sciences – important domains for geospatial methods. This study explores quantum encoding techniques adapted to the constraints of limited qubit space, focusing on gridded numerical weather prediction (NWP) outputs related to low-visibility events at multiple Czech airports. Beyond quantum encoding, we investigate the role of dimensionality reduction (PCA, t-SNE, UMAP, Isomap) and its integration with quantum amplitude and angle encoding schemes. We assess their capacity to represent visibility transitions using fidelity-based measures. Results indicate spatial heterogeneity in encoding effectiveness, with no single method dominating across all locations. To better preserve spatial and physical structure, we introduce expert-informed groupwise embeddings applied separately on meteorological clusters, rather than across the entire dataset. This approach improves the physical relevance and continuity of spatial patterns. Results suggest that temporal features (diurnal cycles), complicate fidelity assessment and emphasize the importance of temporal segmentation and data curation. Our findings demonstrate that combining classical dimensionality reduction, quantum encoding, and domain expertise offers a promising path toward effectively representing complex spatial-temporal atmospheric patterns. This work supports future development of quantum-assisted weather forecasting systems.

Keywords: Dimensionality reduction; Fidelity; Grib; Gridded data; Isomap; PCA; Quantum encoding; t-SNE; UMAP

1. Introduction

The efficient encoding of spatial data for quantum machine learning (QML) is pivotal for quantum computing and processing complex, high-dimensional datasets. A key challenge lies in the data-embedding process, which often becomes a computational bottleneck. The goal of this research is to investigate encoding strategies that allow vast, gridded spatial datasets – particularly those representing meteorological variables – to be represented within the limited qubit resources of current quantum hardware.

Recent developments in quantum data encoding have introduced promising solutions. For instance, Gan et al. (2022) propose a bosonic encoding method that enhances the expressivity of QML models by improving gate efficiency, offering a capable path through the encoding bottleneck. Similarly, Nirala et al. (2023) investigate encoding

information in the spatial correlations of entangled twin beams. By modulating the angular spectrum of the pump beam, they control the spatial structure of the entanglement, effectively using the high-dimensional Hilbert space associated with spatial modes for information encoding. Notably, their approach demonstrates that the encoded information can only be retrieved through joint spatial measurements of both beams – highlighting the nonlocal character of the spatial encoding. Importantly, they show that temporal quantum correlations are preserved despite the spatial modulation, which is significant for quantum networks that require simultaneous spatial expressivity and temporal coherence. Their work underscores the feasibility of encoding structured spatial information in quantum systems, aligning with the broader goal of preserving both spatial structure and temporal dependencies in geospatial data representations – an aspect also central to our study of encoding meteorological fields using QML frameworks.

A critical requirement in spatial encoding is the ability to capture symmetries and spatial relationships efficiently. Many types of spatial data, such as point clouds or gridded environmental data, exhibit permutation symmetry – where the order of points does not affect the underlying structure. Encoding this symmetry directly into quantum states, for example by superposing all possible point orderings, improves generalization and model performance, particularly for 3D spatial data (Heredge et al., 2023).

Permutation-invariant encodings, quantum autoencoders, and other advanced data embedding techniques have been proposed as strategies to handle these challenges. Quantum autoencoders, for example, compress high-dimensional data into lower-dimensional quantum states, reducing qubit requirements while retaining key information. They have been successfully applied to compress and reconstruct high-dimensional quantum states with high fidelity (Romero et al., 2016; Du and Tao, 2021; Pepper et al., 2018; Zhang et al., 2022).

In general, several encoding schemes are available for mapping classical spatial data to quantum states. Common approaches include basis encoding, angle encoding, and amplitude encoding – each with different trade-offs in terms of circuit depth, expressivity, and computational cost (Rath and Date, 2023; Shin et al., 2022; Huang et al., 2021). More recent strategies, such as exponential encoding, aim to represent complex functions efficiently on quantum hardware. Randomized measurement-based encoding has also emerged as a scalable method for processing high-dimensional data, with circuit depth that scales linearly with input size (Shin et al., 2022; Haque et al., 2025). These techniques can be grouped into three main categories, which are summarized in Table 1.

According to the literature, the most effective encoding strategies depend on the structure of the data and the nature of the task. For unordered or highly symmetric data, permutation-invariant encoding and quantum autoencoders are beneficial (Romero et al., 2016). For continuous or structured spatial features, amplitude or angle encoding provides better physical interpretability and supports downstream analysis (Hernández and Amigó, 2021; Mai et al., 2022).

However, not all existing methods are well-suited for structured spatial datasets. While permutation-invariant encodings are powerful for unstructured or unordered data, they are less applicable to datasets with strong spatial or temporal dependencies, such as gridded meteorological fields. Similarly, quantum autoencoders, although effective for compression, do not preserve geometric relationships or distance metrics – both of which are essential for interpreting meteorological variables (Bouquin et al., 2023). Basis encoding, while compact, scales poorly with dimensionality and lacks the expressiveness required for continuous-valued atmospheric data.

Table 1. Encoding methods explored in the literature, their key benefits, and suitable data.

Encoding Method	Key Benefit	Suitable For	Source
Quantum autoencoder	Generalization for unordered spatial data	Point clouds, 3D images	(Heredge et al., 2023)
Permutation-invariant encoding	Compression, resource efficiency	High-dimensional spatial data	(Romero et al., 2016; Du and Tao, 2021; Pepper et al., 2018; Zhang et al., 2022).
Exponential/amplitude encoding	Hardware efficiency, expressivity	Complex spatial features	(Rath and Date, 2023; Shin et al., 2022; Huang et al., 2021).

The motivation for transitioning toward quantum representations in meteorology stems from the inherent limitations of classical exploratory analysis in capturing the high-dimensional, non-linear dependencies. As Numerical Weather Prediction (NWP) models move toward higher resolutions, the complexity of feature interactions – particularly for stochastic variables like visibility – requires more expressive computational spaces. Unlike classical kernels that map data to a fixed-dimensional space, quantum encodings project meteorological variables into a massive Hilbert space. This allows for the representation of complex 'multi-variable correlations' – such as the interplay between moisture, radiation, and aerosol density – that are mathematically difficult to resolve using classical linear methods. By improving the representation of these sub-grid scale processes, QML-based encodings offer a pathway to modernize NWP parameterization schemes. Replacing simplified empirical formulas with quantum-enhanced emulators could lead to a more physically grounded representation of atmospheric boundary layer transitions, ultimately enhancing the precision of local forecasting tasks.

Despite recent advances, encoding strategies that preserve both the spatial continuity and physical interpretability of structured meteorological data – such as gridded visibility fields – remain underexplored. To the best of the author's knowledge, this is the first study to apply and compare quantum amplitude and angle encoding techniques to dimensionally reduced meteorological visibility data, assessing their performance via fidelity measures and their alignment with real-world visibility transitions. This can help directly estimate the anticipated performance of QML models.

Accordingly, this study focuses on amplitude and angle encoding schemes, which are more appropriate for structured environmental datasets. Combined with classical dimensionality reduction techniques, these encodings support an efficient and interpretable quantum representation of key meteorological features. This hybrid approach aims to facilitate the processing of spatial weather data within QML frameworks, offering a practical path for applying QML to meteorological forecasting tasks.

2. Data and Methods

In this study, we follow a structured approach consisting of three main stages: (1) data preprocessing, (2) dimensionality reduction, and (3) quantum encoding with evaluation of encoding quality (Fig. 1).

The study begins by collecting environmental data from the Aladin numerical weather prediction model at ~2.3 km resolution. Key variables relevant to visibility forecasting – temperature, relative humidity, visibility, visibility in precipitation, shortwave radiation, and longwave radiation – are extracted across the Czech Republic. This data is spatially cropped to areas within ~10 km of selected airports (consistent with aviation standards) and organized into georeferenced DataFrames, where each row represents a single time step.

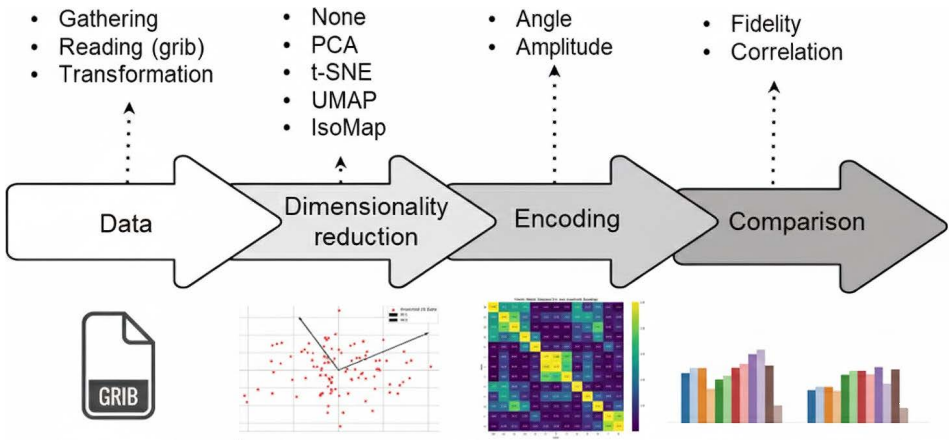

Figure 1. Flowchart of the experimental process. The pipeline begins with the collection of gridded, georeferenced environmental data, followed by preprocessing, cropping to regions of interest, and transformation into a structured DataFrame. Next, dimensionality reduction is applied – either linear or manifold-based – before encoding the data for quantum representation. Finally, the encoded outputs are compared across time steps and validated against observed visibility data.

Each airport-specific DataFrame includes approximately 5 × 5 spatial grid points, 6 variables, and 15 hourly time steps – yielding about 384 records per hour, or ~2,250 values per site. Given this structured spatiotemporal format, we apply a range of classical dimensionality reduction techniques to compress the data while retaining relevant meteorological patterns. Efficient reduction is essential for quantum processing, which is limited by qubit availability. In operational forecasting scenarios – or if the task were formulated as a native quantum problem – datasets would likely scale further to include vertical atmospheric layers and expanded spatial domains with millions of points, increasing the need for scalable reduction strategies.

Following dimensionality reduction, we apply two quantum encoding methods – amplitude encoding and angle encoding – to transform the reduced data into quantum states. For angle encoding, the data is first reduced to ten principal components using PCA to conform to typical quantum circuit size limits.

To assess encoding performance, we compute fidelity distance matrices for each transformation. These are compared to a reference "visibility difference" matrix, derived from absolute differences in visibility across time steps. To quantify how well each meth-od preserves the structure of the original data, we calculate Kendall's Tau correlation between the flattened fidelity matrix and the visibility difference matrix.

This evaluation framework enables both local and global comparison of dimensionality reduction and quantum encoding techniques, with a focus on preserving meaningful structure in meteorological time series – particularly during visibility disruptions.

2.1 Data and Preprocessing

For this study, we used data from the Aladin numerical weather prediction model at ~2.3 km resolution. We selected key variables relevant to visibility forecasting: temperature, relative humidity, visibility, visibility in precipitation, and shortwave and longwave radiation. Data were extracted from areas covering within ~10 km of each airport, aligning with the vicinity defined by aviation regulations.

We tested the low visibility episode on 15/04/2025 from 20:00 UTC to 16/04/2025 10:00 UTC, at four international airports (Fig. 2): Prague (ICAO code: LKPR), Prague-Kbely (LKKB), Pardubice (LKPD) and Brno (LKTB).

These airports were selected because they represent a range of visibility conditions during the event. Prague Ruzyne Airport (LKPR) experienced international flight cancellations due to low visibility, while Brno Airport (LKTB) reported no significant visibility reduction. Pardubice (LKPD) observed localized ground fog, serving as a borderline case of operational interest. Notably, visibility in Pardubice during the episode never dropped below five kilometers, which generally does not impact most commercial flights.

This setup provides a relevant context for setting up quantum computing problems using georeferenced gridded inputs, especially in challenging low visibility scenarios.

2.2 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a statistical technique used for dimensionality reduction, projecting high-dimensional data onto a lower-dimensional subspace while preserving as much variance as possible (Fig. 3) (Jolliffe and Cadima, 2016).

This is achieved by identifying the principal components – orthogonal directions of maximum variance in the data (Jolliffe and Cadima, 2016). Given a centered dataset, we:

- Compute Covariance Matrix;
- Perform Eigen Decomposition;
- Select top k Components;
- Form projection matrix using eigenvectors corresponding to the k largest eigenvalues.
- Project Data;

PCA is widely used for noise reduction, feature compression, and visualization. However, it is a linear method and may not capture complex non-linear patterns (van der Maaten and Hinton, 2008). The data was reduced to up to the first 10 principal components to provide a compatible input vector for subsequent quantum angle encoding, ensuring a balance between information retention and quantum computational efficiency.

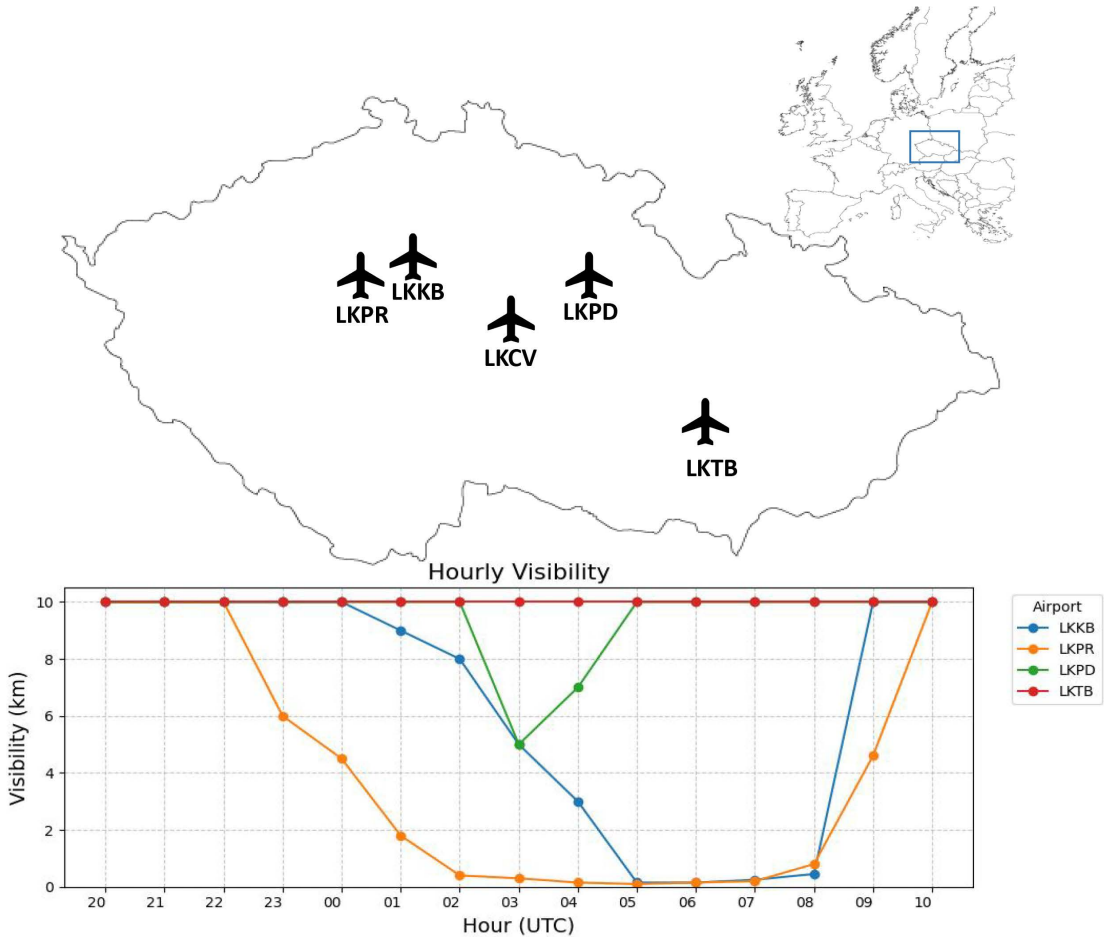

Figure 2. Schematic map of the studied area (Czechia), visibility at four studied airports from 15/04/2025 20:00 UTC to 16/04/2025 10:00 UTC.

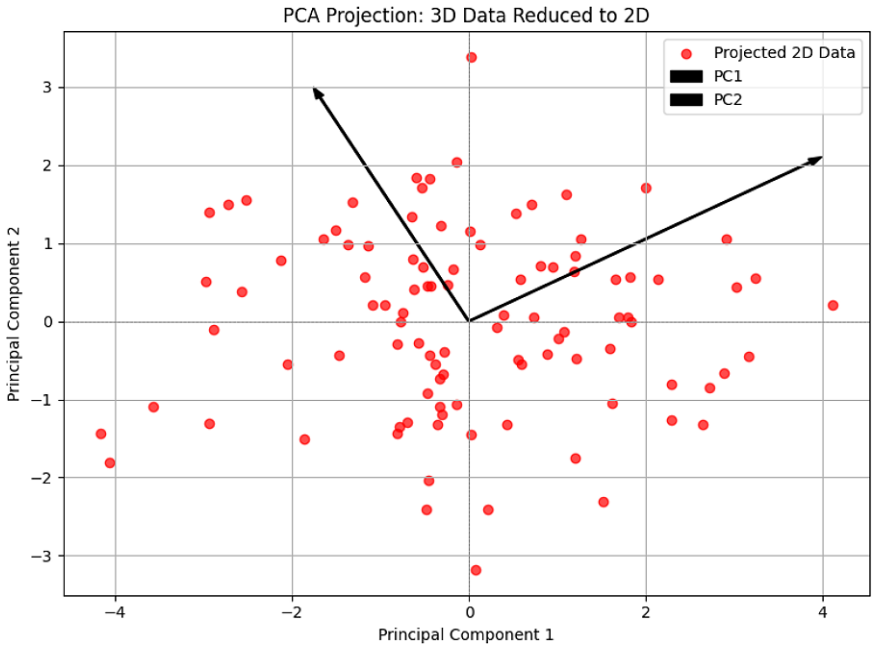

Figure 3. Example of Principal Component Analysis (PCA) reducing simulated 3D data to 2D. A scatterplot of the original 3D data projected onto a 2D plane defined by the first two principal components. The arrows represent the directions of maximum variance (principal components), along which the data has been reoriented.

2.3 t-distributed Stochastic Neighbor Embedding (t-SNE)

The t-distributed Stochastic Neighbor Embedding (t-SNE) method is a widely used non-linear dimensionality reduction technique, particularly effective for visualizing high-dimensional data in two or three dimensions. Unlike Principal Component Analysis (PCA), which emphasizes preserving global variance and structure, t-SNE focuses on maintaining local neighborhoods, making it especially powerful for revealing clusters and complex manifolds in the data (Cai, 2021; Kimura, 2021). As a result, it can produce significantly different projections compared to PCA, as illustrated in the Fig. 4.

The general workflow of t-SNE involves the following steps:

- Computing pairwise similarities between data points in the original high-dimensional space using Gaussian distributions (Cai, 2021);
- Modelling similarities in the lower-dimensional space with a Student's t-distribution, which has heavier tails to better handle crowding effects;
- Minimizing the Kullback-Leibler divergence between the two distributions via gradient descent (Linderman et al., 2019).

While t-SNE is highly effective for visualization, it does come with some limitations:

- It lacks interpretability in terms of feature contributions;
- It does not provide a straightforward mapping function for embedding new data;
- It is sensitive to hyperparameters, such as perplexity and learning rate.

Despite these drawbacks, t-SNE remains a popular tool across domains such as bioinformatics, natural language processing, and quantum feature visualization, due to its ability to uncover latent structures in complex datasets (Sainburg, 2020).

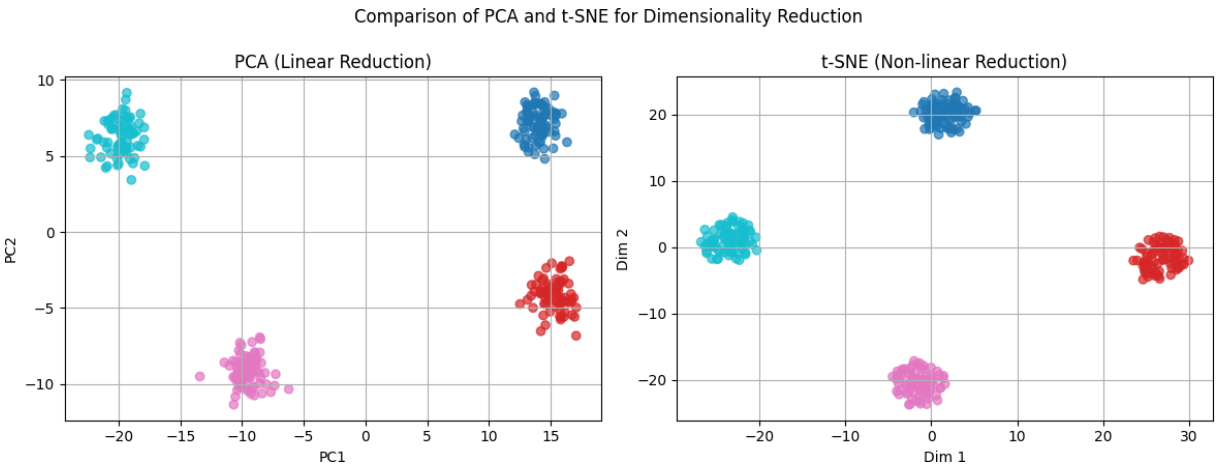

Figure 4. t-SNE projection of high-dimensional data into 2D. The method preserves local neighborhood structures, revealing well-defined clusters and non-linear relationships that can be missed by linear PCA.

2.4 Uniform Manifold Approximation and Projection (UMAP)

Uniform Manifold Approximation and Projection (UMAP) is a non-linear dimensionality reduction technique designed for the visualization and analysis of high-dimensional data (McInnes et al., 2018). It preserves both local neighborhood structures and aspects of global geometry, offering a scalable and efficient alternative to methods like t-SNE. UMAP is grounded in principles from manifold theory and topological data analysis. The UMAP algorithm consists of three main steps (Li et al., 2019):

1) Fuzzy Topological Representation: weighted graph is constructed using approximate nearest neighbors, followed by the computation of a fuzzy simplicial complex that captures the local relationships in the high-dimensional space.

2) Graph Optimization in Low Dimensions: A corresponding fuzzy graph is defined in a lower-dimensional space (typically 2D or 3D). The layout is optimized by minimizing a cross-entropy loss between the high- and low-dimensional fuzzy sets, thereby preserving neighborhood relationships.
3) Stochastic gradient descent (SGD) is used to refine the low-dimensional embedding, ensuring it correctly reflects the structure of the original high-dimensional manifold.

UMAP offers several key advantages. It is a non-linear, manifold-aware technique that effectively balances the preservation of local detail and global structure. Unlike t-SNE, it scales well to large datasets and operates significantly faster. UMAP also allows fine-tuning through parameters such as *n_neighbors* or *min_dist*, which control the locality of neighborhood preservation and the overall spread of the resulting embedding.

However, UMAP also presents some limitations. Its results are non-deterministic unless a fixed random seed is set, which can affect reproducibility. Additionally, it may not accurately capture structure in datasets characterized by sparse sampling or high noise, potentially leading to misleading embeddings in such cases.

2.5 Isomap (Isometric Mapping)

Isomap is a non-linear dimensionality reduction technique that extends classical Multidimensional Scaling (MDS) by preserving geodesic distances – i.e., the shortest paths along a curved manifold – rather than relying on straight-line Euclidean distances (Feng et al., 2024). This approach makes Isomap particularly effective for unfolding complex, non-linear manifolds embedded in high-dimensional spaces (Izenman, 2012).

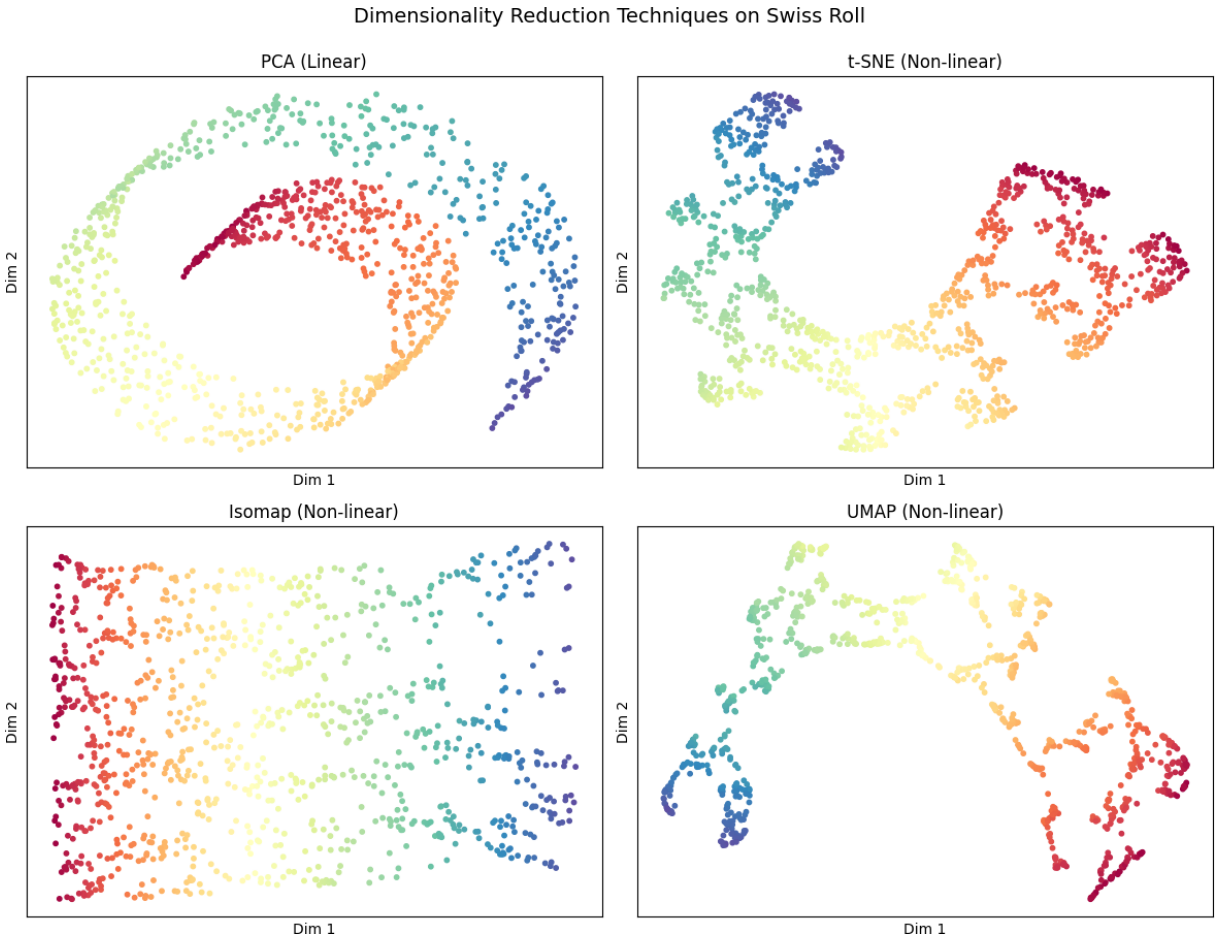

Figure 5. Comparison of dimensionality reduction techniques (PCA, t-SNE, Isomap, UMAP) applied to the Swiss Roll dataset. Isomap successfully unrolls the manifold structure, in contrast to methods like PCA or t-SNE, which focus more on linear variance or local clusters, respectively.

A classic demonstration of Isomap's capabilities is the Swiss Roll dataset, a synthetic example commonly used in manifold learning research. It consists of a 2D surface (a flat sheet) embedded in a 3D space that has been "rolled up." Traditional linear methods fail to capture the intrinsic geometry of such datasets, while Isomap successfully unrolls them by preserving their manifold structure (Zhang et al., 2016). Isomap operates in three main steps.

- Constructs a neighborhood graph by connecting each data point to its k nearest neighbors (or points within a fixed radius).
- Estimates geodesic distances between all data pairs using the shortest path along the graph.
- Applies classical MDS to this geodesic distance matrix, projecting the data into a lower-dimensional space while retaining the manifold's global geometry.

Unlike techniques like t-SNE, which emphasize local clustering, Isomap captures both local and global structure, making it especially useful for datasets where preserving the overall shape is critical. It can "unfold" curved structures while maintaining the relationships between distant points.

Isomap's main advantages lie in its ability to reveal global geometric features of the data and in its interpretability. However, the method is sensitive to noise and to the choice of neighborhood size (k), which can significantly impact the quality of the embedding. It also performs poorly on disconnected or sparsely sampled manifolds, where the estimated geodesic distances may be unreliable (Feng et al., 2024).

Isomap is commonly applied in fields such as image processing, robotics, and quantum feature space analysis, where reducing high-dimensional data while preserving its intrinsic shape is essential (Feng et al., 2024; Ranga et al., 2024).

2.6 Dimensionality reduction assumptions

The goal of dimensionality reduction here is to convert data from the Grib format into a DataFrame and reduce its dimensionality while preserving as much relevant information as possible. For illustration, the Swiss Roll dataset was used to compare techniques based on how well they preserve relationships between points, especially the smooth colour gradient representing intrinsic structure.

PCA, a linear method, projects data along directions of maximum variance but assumes linearity and orthogonality. It fails to capture the Swiss Roll's non-linear manifold, producing a distorted and flattened spiral where layers collapse. PCA is thus suited mainly for linear or simple compression tasks.

A t-SNE method focuses on preserving local neighbourhoods, effectively revealing clusters and local structure. It unrolls the Swiss Roll better than PCA, maintaining the colour gradient and layer separation, but often distorts global geometry and can be computationally expensive. It is best for exploratory visualization of local groupings rather than global embeddings.

Isomap assumes data lies on a smooth manifold and preserves global geodesic distances by using shortest paths on a neighbourhood graph before applying classical scaling. It partially unrolls the Swiss Roll, but results can be stretched or fragmented due to sensitivity to sampling density and neighbour selection. It works well when global manifold structure is critical and data is well-sampled.

UMAP combines manifold theory and topological data analysis to preserve both local and global structure. It produces the clearest and smoothest unrolling of the Swiss Roll with an accurate continuous gradient, outperforming t-SNE in global coherence and scalability. UMAP is robust and efficient for a wide range of complex, high-dimensional datasets.

In summary, UMAP provides the most faithful representation of the Swiss Roll manifold, followed by t-SNE for local detail, Isomap for global geometry under ideal sampling, and PCA being limited to linear cases. The choice of method depends on the data structure and analysis goals: PCA for linear data, t-SNE for local cluster exploration, Isomap for global manifold preservation, and UMAP as a versatile option balancing both local and global features efficiently.

Since numerical weather prediction outputs are derived from mathematical differential equations, the resulting data tends to be continuous and structured. This continuity can mitigate some of the limitations of PCA related to linearity. Therefore, if the data is not overly noisy, PCA may perform well. However, it may still struggle with more complex, non-linear structures – such as those seen in the Swiss Roll dataset.

2.7 Angle encoding

This technique (also known as qubit encoding) maps classical data features directly to the rotation angles of quantum gates, typically single-qubit rotation gates like the Pauli-X rotation (*RX*), Pauli-Y rotation (*RY*), or Pauli-Z rotation (*RZ*) (Huang et al., 2021). For instance, an RX gate, parameterized by an angle θ, rotates the state of a qubit around the x-axis of the Bloch sphere. In angle encoding, each classical data point is scaled and normalized to fit within a specific range, often $[0, 2\pi]$, and then used as the rotation angle for a corresponding qubit. Having *N* features in the classical data, would generally result into use of *N* qubits, with each feature controlling the rotation of its respective qubit (Huang et al., 2021).

Specifically, we employ Pauli-X rotation (RX) gates for this purpose. The encoding process begins by normalizing the input classical data features to the range $[0, 2\pi]$. This preprocessing step ensures that the full range of possible rotation angles on the Bloch sphere is utilized, maximizing the expressivity of the quantum state. Each normalized feature is then used as the parameter for an RX gate applied to a dedicated qubit. For an N-dimensional classical data point, this translates to an N-qubit quantum circuit where each qubit *i* undergoes an $RX(\theta_i)$ rotation, with θ_i being the normalized value of the *i-th* feature. This method results in a product state, where each qubit's state independently reflects a single feature of the classical data. This straightforward, one-to-one mapping between classical features and qubit rotations provides an interpretable and efficient way to embed classical information into the quantum computational space, forming the foundational input for subsequent quantum machine learning algorithm (Schuld and Killoran, 2019; Santini and Jain, 1999).

Given a classical feature vector:

$$x = [x_1, x_2, \ldots x_n] \in X^n,$$

the quantum state prepared via angle encoding can be represented as the tensor product of individually rotated qubits:

$$|\mathrm{x}\rangle = \otimes(\cos(\mathrm{x}_i)|0\rangle + \sin(\mathrm{x}_i)|1\rangle) \tag{1}$$

Each qubit undergoes a parameterized unitary rotation U_i defined as:

$$U_i(x_i) = \begin{bmatrix} \cos x_i & \sin x_i \\ -\sin x_i & \cos x_i \end{bmatrix} \tag{2}$$

The full quantum circuit for preparing the encoded state can then be written as:

$$S_x = \otimes_{i=1}^{N} U_i(x_i) \tag{3}$$

where S_x is the overall state preparation unitary applied to the initial $|0\rangle^{\otimes \mathrm{N}}$ state (Huang et al., 2021).

In this work, we specifically constrain the input to $\mathrm{N} = 10$ features derived from the PCA preprocessing (Section 2.2). This choice of 10 qubits serves a dual purpose: it captures the dominant variance of the meteorological visibility fields while maintaining a state-vector dimensionality ($2^{10} = 1024$ complex coefficients) that is computationally tractable for high-fidelity simulation.

2.8 Amplitude encoding

Amplitude encoding also was employed to represent the 2D gridded meteorological variables within this study. This method involves mapping the numerical values of the meteorological fields directly to the amplitudes of

a corresponding signal or data structure (Javadi Abhari et al., 2024). In essence, for a given gridded meteorological variable $V(x,y)$ at a specific grid point (x,y), its numerical value is encoded as the amplitude $A(x,y)$ of a representation. This can be conceptually expressed as:

$$A(x,y) = f\big(V(x,y)\big)$$

where f is a direct mapping function, often linear or identity, meaning that higher values of the meteorological variable correspond to larger amplitudes. Each grid point in the 2D field is thus assigned a unique amplitude value, thereby preserving the spatial distribution and relative magnitudes of the original meteorological data. This approach facilitates the direct translation of the gridded data into a format suitable for subsequent processing and analysis, particularly when the downstream operations are sensitive to variations in signal intensity.

2.9 Fidelity and Similarity

To assess the structural preservation of various data embeddings, we compute a fidelity matrix for each method. This adaptive approach ensures a consistent measure of similarity, irrespective of the embedding type (Javadi Abhari et al., 2024).

For quantum embeddings, which yield complex-valued data representations, we directly calculate the quantum state fidelity using the Qiskit *state_fidelity* function (Javadi Abhari et al., 2024). This measures the overlap between two quantum states, with a value of 1 indicating identical states and 0 indicating orthogonal states (Masot Llima and Garcia Saez, 2025; Sierra Sosa et al., 2020).

In contrast, for classical dimensionality reduction techniques producing real-valued embeddings, we derive similarity from Euclidean distances. Specifically, the pairwise Euclidean distance matrix is computed, then transformed into a similarity score by normalizing these distances and subtracting them from 1. This results in a fidelity value between 0 and 1, where 1 signifies maximum similarity (zero distance) and 0 represents minimum similarity (maximum distance within the dataset). This dual approach allows for a direct and fair comparison of how well each embedding method, whether classical or quantum, captures the inherent relationships within the data.

To quantify the alignment between the physical atmospheric state and the resulting embeddings, we calculate the correlation between the classical visibility difference matrix and the generated fidelity matrices. We utilize Kendall's Tau (τ), a non-parametric rank-based correlation, for this evaluation. This metric is specifically chosen because it assesses the monotonic relationship between the two spaces, measuring whether pairs of observations that are physically similar (low visibility difference) consistently result in higher embedding similarity (high fidelity). By focusing on the preservation of similarity rankings rather than linear values, Kendall's Tau provides a robust measure of how effectively the quantum and classical feature maps represent the underlying structural transitions of the meteorological data.

3. Results

This research investigated several approaches to quantum encoding of meteorological georeferenced values that had previously undergone dimensionality reduction. We focused on complex visibility data – likely to become a quantum relevance problem in future studies – as well as the modeling of low level clouds.

A range of dimensionality reduction techniques was tested, including both linear (PCA) and nonlinear methods (t-SNE, UMAP, IsoMap), with comparisons made using data from key airports in Prague – LKPR (Ruzyne) and LKKB (Kbely). For the period from 15 April to 16 April 2025, different patterns emerged across the methods (Fig. 6).

We used a matrix of hourly visibility differences as a proxy for similarity and compared it to the fidelity of individual quantum states.

The Kendall Tau correlations between embedding-based fidelity matrices and ground-truth visibility drops reveal alignment with observed visibility patterns at each airport. At LKPR, visibility steadily declines from 10 km to 0.1 km, with low visibility sustained over several hours. In contrast, LKKB experiences a sharp dip from 10 km to below 0.5 km during its peak low-visibility period. These substantial changes are reflected to varying degrees across

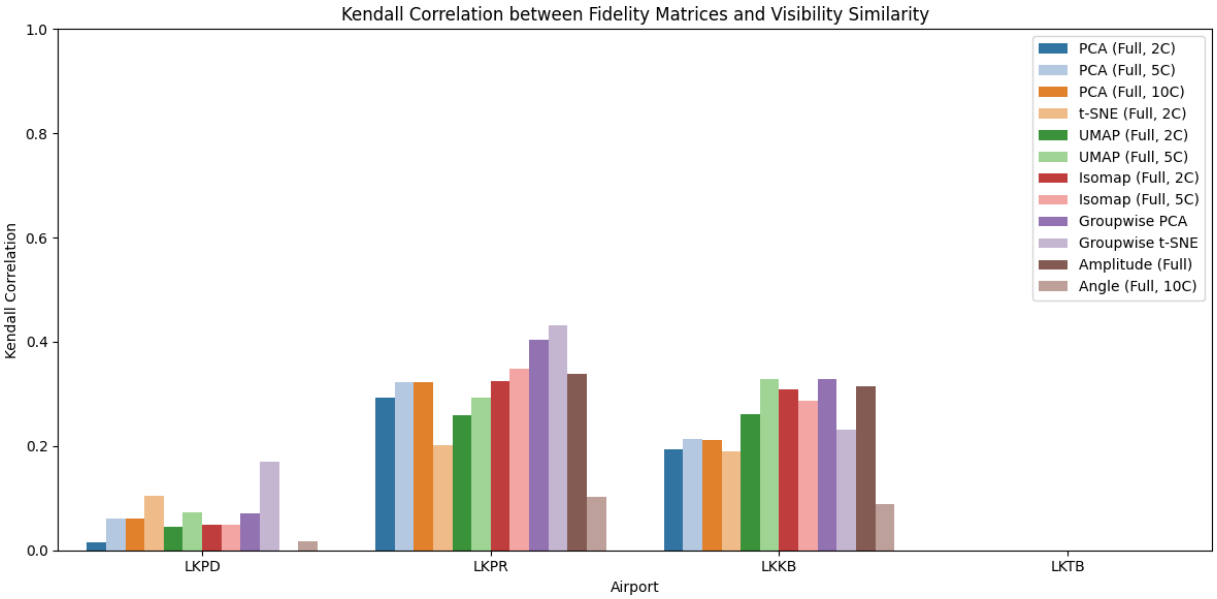

Figure 6. Kendall correlation of Fidelity matrices and visibility similarity matrices, defined as drops of visibility between hours.

the dimensionality reduction techniques. Nonlinear and groupwise methods demonstrate moderate correlations above 0.3 – reaching up to 0.43 for Groupwise t-SNE at LKPR – suggesting these approaches are better at capturing complex visibility transitions. Quantum amplitude encoding also performs comparably, indicating potential for representing such dynamics.

At LKPR, most methods yield moderate positive correlations between 0.3 and 0.43. Groupwise t-SNE (0.43) and Groupwise PCA (0.40) are among the best performers, indicating their relative strength in preserving the underlying visibility structure. Classical methods such as PCA and IsoMap also maintain consistent correlations around 0.3. Quantum amplitude encoding at LKPR produces a notable correlation of 0.34. At LKKB, the overall correlations are somewhat lower but still positive. UMAP, IsoMap, and Groupwise PCA perform best (around 0.3), with amplitude encoding again ranking well at 0.32. These results suggest nonlinear manifold learning methods may better capture subtle visibility dynamics at LKKB than simpler linear techniques.

In contrast, LKPD shows generally weak correlations, rarely exceeding 0.17. Groupwise t-SNE again performs best here (0.17), but quantum amplitude and angle encodings show near-zero or slightly negative correlations, suggesting a more complex or noisier data profile at this site that diminishes embedding effectiveness. At LKTB, all correlations are returning error values (NaN), due to zero variance in visibility observations during the examined period. As a result, this site serves effectively as an example case, providing a control where no meaningful structure or relationships are expected. Consequently, it was excluded from further interpretation.

Groupwise embeddings – based on variables such as temperature, radiation, and visibility – consistently outperform standard PCA, particularly when multiple components are included. These findings indicate that embedding methods tailored to local data properties, including quantum encodings, can better capture visibility changes in specific meteorological contexts.

The choice of the two Prague airports – LKPR and LKKB – was motivated by the presence of fog at both locations. LKKB, the more easterly Kbely airport, exhibited a later onset and slower development of fog. At LKPR (Ruzyne), the transition in visibility was more abrupt and better captured by the t-SNE method, although the visibility reduction was represented faithfully in the processed data across both sites. Fidelity remained stable across homogeneous time windows before fog onset, during fog, and after dissipation (Fig. 7).

Comparing two airports within the same city offers a useful basis for future research, especially in the context of incorporating static geographic models (e.g., terrain, land use, building type). The visibility and fidelity timelines for each method provide additional granularity (Figs. 8 and 9). Even some methods (e.g. t-SNE, 2C at LKPR) that performed poorly on average were still able to capture visibility drops effectively. However, they may be downgraded by the other factors.

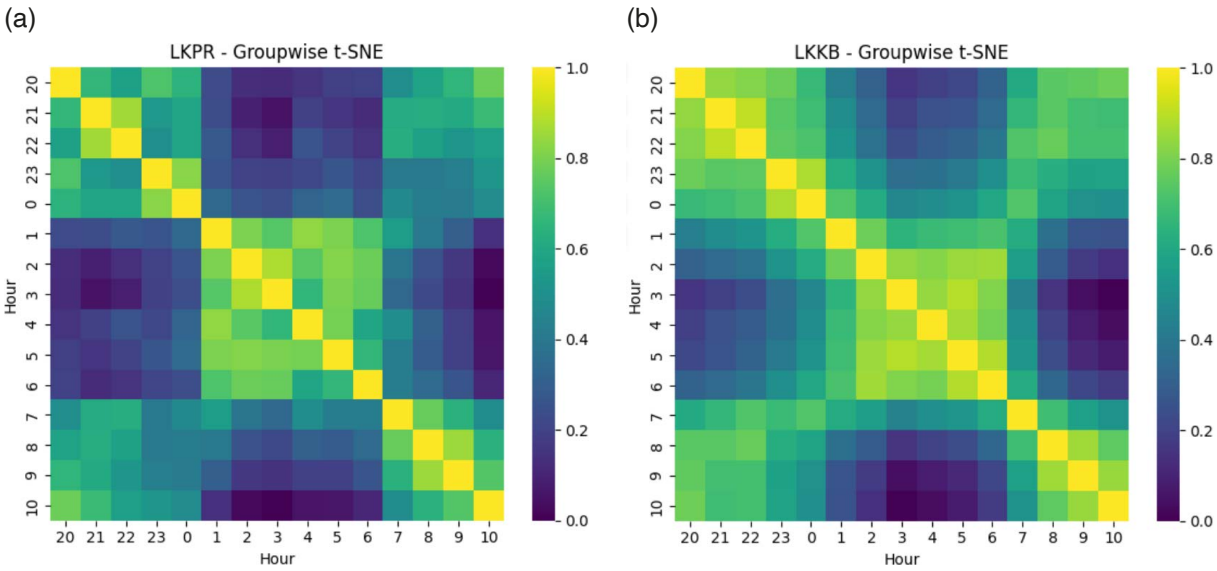

Figure 7. (a) Fidelity matrix of Ruzyne (LKPR); (b) Fidelity matrix of Kbely (LKKB), both using the Groupwise t-SNE method.

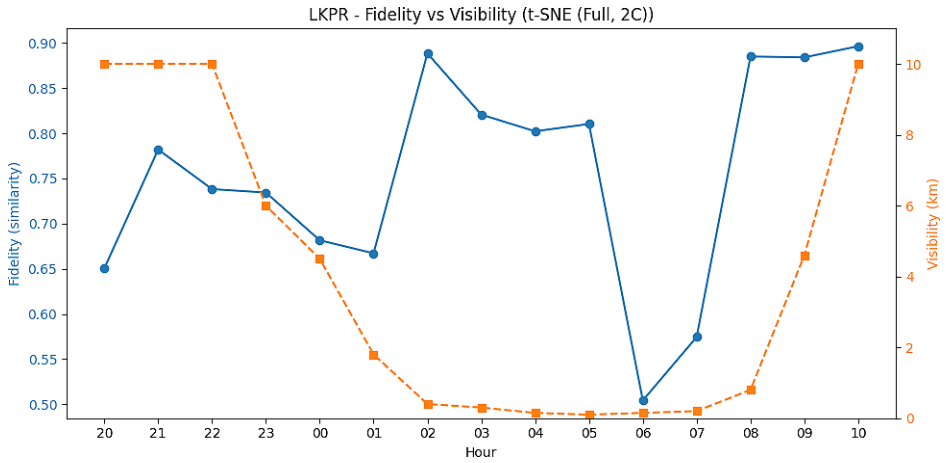

Figure 8. Inter-hourly Fidelity compared to observed visibility at LKPR. A drop is observed between 5 and 6 a.m., likely due to a sunrise-related data disruption.

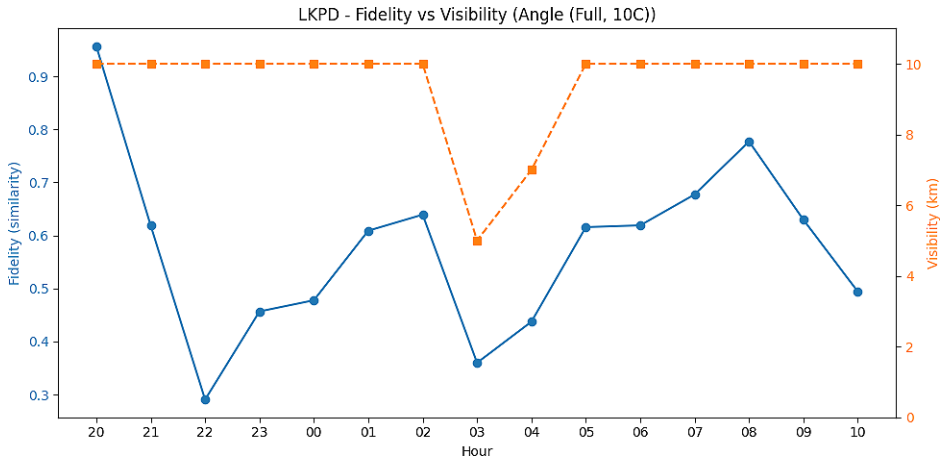

Figure 9. Inter-hourly Fidelity compared to observed visibility at LKPD. A drop is observed between 5 and 6 a.m., likely due to a sunrise-related data disruption.

At LKPR, this drop between 5-6 a.m. is notable, suggesting a model-related sunrise effect that did not translate into real visibility changes. Although the Angle encoding quantum method did not perform well overall at LKPD, it did visually represent the visibility drop during sunset. The first fidelity drop may again correspond to solar position effects reflected in the input NWP data.

At LKPD, visibility drops are also visible in several other methods. These patterns may result from sunlight emergence and associated temporal factors, which may not be fully captured in model values – especially not in visibility – yet still influence encoding outputs.

These observations underline the importance of visibility dynamics: sharper and more sustained drops are associated with stronger, more meaningful correlations between fidelity matrices and actual visibility. In contrast, at airports with stable conditions, the quantum or classical encoding outputs may appear noisy due to the absence of meaningful variance.

4. Discussion

Forecasting low-visibility events from structured meteorological data presents a particularly complex challenge – both due to the physical difficulty of the prediction task and the spatially heterogeneous nature of the underlying data. The visibility patterns examined in this study – derived from gridded weather fields and tied to specific observation points at multiple Prague-area airports – highlight the limitations of applying data-driven models without acknowledging the spatial structure embedded in atmospheric processes.

Previous studies have often focused on model cases or physical dependencies, focusing predominantly on temporal or purely statistical approaches. In contrast, this work explicitly incorporates the spatial continuity and locality of meteorological variables by combining classical dimensionality reduction with quantum encoding techniques tailored to structured spatial datasets. Our results demonstrate that preserving spatial coherence is crucial for capturing visibility transitions accurately, as reflected in the varying performance across airports with distinct local conditions.

The rationale for employing QML in NWP lies in the superior expressivity of quantum feature maps. While classical models often simplify high-dimensional meteorological dependencies, quantum encodings allow for the representation of these variables in a Hilbert space, where complex non-linear interactions – critical for resolving phenomena like fog – can be more naturally captured. By improving the representation of sub-grid scale processes, QML offers a pathway to enhance the precision of NWP parameterizations, moving beyond the limitations of classical statistical approximations.

A key factor influencing the outcomes is the bias introduced by the numerical weather prediction (NWP) model, which not only governs temporal evolution but also shapes the spatial distribution of input features. Variables like radiation and humidity are not uniformly distributed across space, and their gradients, particularly near surface layers, contribute significantly to fog development. However, these same variables can introduce noise into fidelity analysis when abrupt, model-driven transitions occur – such as around sunrise and sunset, where local surface radiation and its gradients can cause artificial jumps in the data. These are spatially dependent effects, which must be accounted for when encoding georeferenced meteorological fields for quantum or classical learning tasks.

Importantly, the results varied across the analyzed spatial locations (airports). Each airport demonstrated different visibility evolution patterns, and consequently, no single embedding or encoding method consistently outperformed the others. This underscores the site-specific nature of visibility conditions and the challenge of generalizing across heterogeneous spatial domains. While nonlinear embedding techniques like groupwise t-SNE showed moderate performance in detecting transitions at some locations (e.g., LKPR), their effectiveness diminished in others (e.g., LKPD), where the visibility signal was more complex or affected by noise.

These outcomes emphasize that spatial context cannot be abstracted away. For spatially structured fields like meteorological visibility, expert-informed grouping of variables based on physical proximity and relevance (e.g., separating visibility, radiation, moisture) enhances both classical and quantum embedding performance. Without such grouping, the data-driven encoding process can collapse under the weight of irrelevant or redundant spatial variables. Groupwise encodings – particularly those based on semantically meaningful spatial clusters – demonstrated better fidelity in capturing transitions during spatially homogeneous low-visibility events, further supporting the importance of preserving spatial organization in preprocessing.

Despite these challenges, most dimensionality reduction techniques succeeded in preserving key spatial structures in the reduced feature space. During episodes of stable fog conditions or persistent clear visibility, the resulting

quantum fidelity matrices reflected this homogeneity well, suggesting that quantum encodings can retain spatial continuity in georeferenced datasets. This capacity is particularly promising for applications in geoinformation science, where understanding transitions across both space and time is essential.

However, temporally autocorrelated spatial data introduces unique interpretation challenges. When no significant change occurs in visibility, fidelity fluctuations may reflect diurnal or annual cycles embedded in the model inputs, rather than actual spatial or temporal transitions. These baseline variations – driven by radiation or solar angle changes—must be accounted for through careful data curation and possibly detrending techniques, so that normal spatial variation is not misclassified as anomalous.

In conclusion, this study demonstrates the potential of combining classical dimensionality reduction with quantum encoding for analyzing geospatial meteorological datasets. However, the spatial dimension must be explicitly preserved and respected at every stage of preprocessing and encoding. Quantum amplitude and angle encodings show promise for capturing structured spatial phenomena like visibility reduction, but they must be grounded in physical domain knowledge and spatially coherent grouping strategies.

Our dimensionality reduction to encoding pipeline offers a practical and interpretable route for embedding meteorological fields into quantum circuits, particularly under current hardware constraints. By reducing dimensionality through PCA and other methods, we make it feasible to map structured geospatial data into limited qubit systems, enabling exploratory applications of QML in atmospheric science. However, this approach flattens complex spatial dependencies and treats temporal features only indirectly. In contrast, the work by Nirala et al. (2023) demonstrates the potential of encoding spatial correlations directly into entangled quantum states using the spatial degrees of freedom of light, while preserving temporal coherence. Although their method is rooted in quantum optics and currently less accessible for general-purpose QML, it points toward a more expressive and physically grounded form of quantum data representation. Bridging the gap between such native quantum encodings and practical machine learning pipelines may be essential for advancing geophysical applications of QML in the long term.

The constraint of $n = 10$ qubits was selected as a strategic balance between capturing the primary meteorological variance and maintaining the computational tractability of state-vector simulations (2^{10} complex coefficients). While this dimensionality is sufficient to resolve broad visibility transitions across the airport network, it serves as a baseline; future scaling to higher qubit counts will likely allow for the inclusion of finer, sub-grid scale features that were filtered out in this study. This 10-dimensional mapping confirms that essential atmospheric signals can be successfully compressed into a format suitable for current quantum architectures.

Future research should expand on these results by integrating topographical, land-use, and static geospatial factors, which may further enhance the interpretability and robustness of quantum machine learning applied to atmospheric spatial data. From the meteorological point of view, handling climatological assumptions is necessary.

The temporal evolution of meteorological variables such as temperature and radiation introduces cyclic patterns can confound the interpretation of fidelity and similarity measures. For instance, abrupt changes in radiation at sunrise and sunset manifest as shifts in encoded quantum states, which may not correspond directly to changes in visibility but rather reflect natural temporal rhythms. While this study focuses on a limited time window, this period represents a high-resolution case study of a complex, non-linear visibility transition. The use of high resolution NWP data allows for a granular evaluation of how quantum encodings capture rapid atmospheric shifts that are often lost in broader statistical averaging. Furthermore, the intuitivelly consistent performance across four distinct airport locations provides spatial cross-validation, confirming that the observed structural signals are robust characteristics of the encoding methodology rather than temporal artifacts. This focused approach establishes the representational fidelity required for future, larger-scale QML applications in atmospheric science.

To improve the robustness of encoding and subsequent analyses, future work should incorporate preprocessing techniques aimed at filtering or detrending these temporal cycles. Such filtering would help isolate meaningful spatial-temporal signals related to fog formation and dissipation from background temporal variability, thereby enhancing the fidelity of quantum and classical representations of meteorological data.

Acknowledgement. This research was funded by the project of the Czech Ministry of Defense, Military autonomous and robotic assets (DZRO-FVT22-VAROPS).

Data Availability Statement. Data used in these research are available as an open-source and will be provided upon reasonable request.

References

Bouquin, D., A. Trisovic, O. Bertuch and E. Colón Marrero (2023). Advancing software citation implementation (Software Citation Workshop 2022). Preprints, doi:10.1002/wics.101.

Cai, T. (2021). Theoretical foundations of t-SNE for visualizing high dimensional clustered data, arXiv preprint, https://arxiv.org/abs/2105.07536.

Du, Y. and D. Tao (2021). On exploring the potential of quantum auto encoder for learning quantum systems, IEEE Trans. Neural Netw. Learn. Sys., doi:10.1109/TNNLS.2024.3474793.

Feng, W., Z. Jiang, S. Liu, Z. Shi et al. (2024). Quantum Isomap algorithm for manifold learning, Phys. Rev. Appl., 22, 014049, doi:10.1103/PhysRevApplied.22.014049.

Gan, B., D. Leykam and D. Angelakis (2022). Fock state enhanced expressivity of quantum machine learning models, EPJ Quantum Technol., 9, 1, doi:10.1140/epjqt/s40507-022-00135-0.

Haque, E., M. Paul, F. Tohidi and A. Ulhaq (2025). An overview of quantum circuit design focusing on compression and representation, Electron., 14, 72, doi:10.3390/electronics14010072.

Heredge, J., C. Hill, L. Hollenberg and M. Sevior (2023). Permutation invariant encodings for quantum machine learning with point cloud data, Quantum Mach. Intell., 6, 25, doi:10.1007/s42484-024-00156-1.

Hernández, A. and J. M. Amigó (2021). Attention mechanisms and their applications to complex systems, Entropy, 23, 283, doi:10.3390/e23030283.

Huang, H., R. Kueng and J. Preskill (2021). Power of data in quantum machine learning, Nature Commu., 12, doi:10.1038/s41467-021-22539-9.

Izenman, A. J. (2012). Introduction to manifold learning, WIREs Computational Statistics, 4, 439-446., doi:10.1002/wics.1222.

Javadi Abhari, A., A. Amodio, D. Amaro and R. Ananthanarayanan (2024). Quantum computing with Qiskit. arXiv preprint. https://arxiv.org/abs/2405.08810.

Jolliffe, I. T. and J. Cadima (2016). Principal component analysis: A review and recent developments, Phil. Trans. R. Soc. A, 374, 20150202, doi:10.1098/rsta.2015.0202.

Kimura, M. (2021). Generalized t-SNE through the lens of information geometry, IEEE Access, 9, 129619-129625, doi:10.1109/ACCESS.2021.3113397.

Li, H. and H. Yang (2019). Implementation of Manifold Learning Algorithm Isometric Mapping, J. Compu. Commu., 7, doi:10.4236/jcc.2019.712002.

Llima, S. M. and A. G. Saez (2025). Advantages of Density in Tensor Network Geometries for Gradient Based Training. Algorithms, 18(70), doi:10.3390/a18020070.

Linderman, G. C., M. Rachh, J. G. Hoskins, S. Steinerberger et al. (2019). Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data, Nature Meth., 16, 3, 243-245, doi:10.1038/s41592-018-0308-4.

Mai, G., Z. Lin, J. Wu and Y. Zhang (2022). A review of location encoding for GeoAI: Methods and applications. Int. J. Geograph. Info. Sci., 36, 639-673, doi:10.1080/13658816.2021.2004602.

McInnes, L., J. Healy and J. Melville (2018). UMAP: Uniform Manifold Approximation and Projection, J. Open Source Softw., 3, 861, doi:10.21105/JOSS.00861.

Nirala, G., S. Pradyumna, A. Kumar and A. Marino (2023). Information encoding in the spatial correlations of entangled twin beams., Sci. Adv., 9, 22, doi:10.1126/sciadv.adf9161.

Pepper, A., N. Tischler and G. Pryde (2018). Experimental realization of a quantum autoencoder: The compression of qutrits via machine learning, Phys. Rev. Lett., 122, 060501, doi:10.1103/PhysRevLett.122.060501.

Ranga, D., P. Kumar and V. Singh (2024). Quantum Machine Learning: Exploring the Role of Data Encoding Techniques, Challenges, and Future Directions, Math., 12, 3318, doi:10.3390/math12213318.

Rath, M. and H. Date (2023). Quantum data encoding: A comparative analysis of classical-to-quantum mapping techniques and their impact on machine learning accuracy, arXiv preprint, doi:10.48550/arXiv.2311.10375.

Romero, J., J. Olson and A. Aspuru-Guzik (2016). Quantum autoencoders for efficient compression of quantum data. Quantum Science and Technology, 2. doi:10.1088/2058-9565/aa8072.

Santini, S. and R. Jain (1999). Similarity measures. IEEE Trans. Patt. Anal. Machine Intell., 21, 8, 871-883. doi:10.1109/34.790428.

Sainburg, T., L. McInnes and T. Q. Gentner (2020). Parametric UMAP: Learning embeddings with deep neural networks for representation and semi supervised learning, arXiv preprint, https://arxiv.org/abs/2009.12981.

Schuld, M. and N. Killoran (2019). Quantum machine learning in feature Hilbert spaces, Phys. Rev. Lett., 122, 040504 doi:10.1103/PhysRevLett.122.040504.

Shin, S., Y. Teo and H. Jeong (2022). Exponential data encoding for quantum supervised learning., Phys. Rev. A, 107, 012422 doi:10.1103/PhysRevA.107.012422.

Sierra Sosa, D., M. Telahun and A. Elmaghraby (2020). Tensorflow quantum: impacts of quantum state preparation on quantum machine learning performance. IEEE Access, 8, 215246-215255, doi:10.1109/ACCESS.2020.3040798.

van der Maaten, L. and G. Hinton (2008). Visualizing Data using t-SNE., J. Machine Learn. Res., 9, 2579-2605.

Zhang, H., J. Cui, C. Liu, Y. Zhang et al. (2022). Resource efficient high dimensional subspace teleportation with a quantum autoencoder, Sci. Adv., 8, doi:10.1126/sciadv.abn9783.

Zhang, Y., X. Wang, X., Y. Wang and D. Zhang (2016). Multi-manifold discriminant isomap for visualization and classification, Pattern Recogn., 55, 215-230, doi:10.1016/j.patcog.2016.02.001.

***CORRESPONDING AUTHOR: David, SLÁDEK,**
Department of Military Geography and Meteorology, University of Defence, Kounicova 65, Brno, Czechia
e-mail: david.sladek@unob.cz