

Learning to Land Anywhere: Transferable Generative Models for Aircraft Trajectories

Olav Finne Præsteng Larsen², Massimiliano Ruocco^{1,2}, Michail Spitieris¹, Abdulmajid Murad¹, Martina Ragosta¹

¹Department of Software Engineering, Safety and Security
SINTEF Digital
Trondheim, Norway
{abdulmajid.murad, massimiliano.ruocco, michail.spitieris,
martina.ragosta}@sintef.no

²Department of Computer Science
Norwegian University of Science and Technology
Trondheim, Norway
massimiliano.ruocco@ntnu.no

Abstract—Access to trajectory data is a key requirement for developing and validating Air Traffic Management (ATM) solutions, yet many secondary and regional airports face severe data scarcity. This limits the applicability of machine learning methods and the ability to perform large-scale simulations or “what-if” analyses. In this paper, we investigate whether generative models trained on data-rich airports can be efficiently adapted to data-scarce airports using transfer learning. We adapt state-of-the-art diffusion- and flow-matching-based architectures to the aviation domain and evaluate their transferability between Zürich (source) and Dublin (target) landing trajectory datasets. Models are pretrained on Zürich and fine-tuned on Dublin with varying amounts of local data, ranging from 0% to 100%. Results show that diffusion-based models achieve competitive performance with as little as 5% of the Dublin data and reach baseline-level performance around 20%, consistently outperforming models trained from scratch across metrics and visual inspections. Latent Flow Matching and Latent Diffusion models also benefit from pretraining, though with more variable gains, while Flow Matching models show weaker generalization. Despite challenges in capturing rare trajectory patterns, these findings demonstrate the potential of transfer learning to substantially reduce data requirements for trajectory generation in ATM, enabling realistic synthetic data generation even in environments with limited historical records.

Index Terms—Air traffic management, embeddings, variational autoencoders, operational analytics, trajectory clustering, outlier detection, synthetic data, generative models

I. INTRODUCTION

The development of robust Air Traffic Management (ATM) systems increasingly relies on the availability of high-quality trajectory data. Such data underpins the design of machine learning algorithms, the validation of new operational procedures, and large-scale simulation studies. However, while major hub airports typically record large volumes of flight data, many secondary and regional airports exhibit substantial gaps or inconsistencies in historical surveillance data, often due to limited sensor coverage, fewer daily operations, incomplete archival records, or aircraft-type-specific sparsity. While scarcity is not uniform across all airports, these conditions create meaningful barriers for training and validating data-intensive ATM models. This imbalance limits the ability to test and validate new ATM solutions across diverse operational contexts. Moreover, the exploration of rare events, edge cases,

and “what-if” scenarios requires datasets far larger and more varied than what is available from real-world operations.

Recent advances in generative modeling have shown promise in addressing this challenge by synthesizing realistic flight trajectories. Techniques such as Variational Autoencoders (VAEs) [1], Generative Adversarial Networks (GANs) [2], and Diffusion Models [3] have demonstrated their ability to capture the distributional characteristics of real flight data and generate plausible synthetic samples. Nevertheless, these approaches typically assume access to a sufficient amount of local training data. In practice, this assumption often fails for airports with limited operations or incomplete historical records, restricting the applicability of generative methods where they are needed most.

In this paper, we investigate whether transfer learning can mitigate these challenges. Specifically, we explore whether generative models trained on a data-rich airport can be effectively adapted to airports with scarce data. By pretraining on a large landing trajectory dataset from Zürich Airport and fine-tuning on limited data from Dublin Airport, we evaluate the transferability of state-of-the-art generative architectures, including Diffusion Models, Flow Matching, and Latent Flow Matching.

The main contributions of this work are as follows:

- We present, to the best of our knowledge, the first systematic study of transfer learning for generative trajectory models in the ATM domain.
- We demonstrate that diffusion-based models can achieve competitive performance with as little as 5% of the target airport’s data and reach baseline-level performance around 20%, significantly reducing local data requirements.
- We provide a comparative analysis of different generative approaches, highlighting the strengths of diffusion-based architectures and the limitations of flow-matching methods for transfer learning.
- We discuss the implications of transfer learning for enabling realistic trajectory generation in data-scarce environments, thereby broadening the applicability of synthetic data in ATM research and development.



By showing that generative models can transfer knowledge across airports, this work paves the way toward scalable synthetic trajectory generation, supporting the development of data-driven ATM systems even in operational contexts where historical records are limited.

II. STATE OF THE ART

Research on synthetic flight landing trajectories lies at the intersection of deep generative modeling and aviation operations. Progress in deep generative models has been rapid, spanning Variational Autoencoders (VAE) [1], Generative Adversarial Network (GAN) [2], Diffusion Models (DM) [3, 4, 5], and, most recently, Flow Matching (FM) [6]. Latent variants such as Latent Diffusion Models (LDM) [7] and latent flow matching (LFM) [8] combine strong sample quality with improved efficiency by shifting generation to a compressed space. While most architectural innovation has emerged in computer vision, a growing body of work adapts these ideas to spatio-temporal trajectories, where temporal dependencies, multi-modality, and physical constraints are central. In particular, this work focuses primarily on advances in trajectory generation in the ATM domain. This is currently a growing research area aiming to address challenges related to data accessibility, privacy, and the complexity of flight data [9].

Aircraft trajectory generation can be formulated as the problem of generating a set of fully timestamped coordinates for an aircraft that reflect physical behavior and the statistical properties of the spatio-temporal bounds [10]. The existing literature on generative flight trajectory models divides them into two categories: model-driven and data-driven approaches [10, 11, 12]. Model-driven approaches rely on mathematical models of flight, ensuring that the trajectories follow the laws of aerodynamics and other operational constraints [10].

Data-driven methods use machine learning techniques to model and generate flight trajectories directly from data [10]. These methods, such as TimeGAN [11], TCVAE and VampPrior[13], have shown promise in capturing temporal dependencies and generating realistic landing trajectories and [12] that applied the proposed TimeVQVAE architecture [14] to generate complete flight trajectories and introduced new evaluation methods tailored to trajectory generation. Interestingly, [13] generates trajectories by groundspeed and track. In contrast, [11] uses longitude and latitude, which is the most common approach in other domains, demonstrating that there are different approaches to modeling the data to be learned.

While the application of data-driven generative models in aviation remains underdeveloped, significant advances have been made in other fields, particularly urban traffic, where state-of-the-art generative models such as VAEs [15], GANs [16], and DMs [17, 18] have successfully learned complex trajectory data distributions. Recent advancements in DMs [17, 19, 20] have demonstrated significant potential for spatio-temporal data generation. Although aviation and urban traffic differ due to the three-dimensional and sparse nature of airspaces compared to the strict topological constraints

of urban traffic [18], the DMs' ability to handle complex, high-dimensional distributions [4, 5] suggests they can overcome the challenges posed by flight trajectories. These are challenges such as the need for physical realism, airspace restrictions, and various weather conditions [10, 21].

While newer DMs have been proposed in urban traffic [18, 20], these are specifically tailored to the urban traffic domain by incorporating road constraint information into the models to improve performance. Therefore, DiffTraj [17] is the chosen diffusion architecture to be adapted to the aviation domain, as it has good performance while not implementing strict domain-specific constraints, potentially making it well-suited for generating flight trajectories. Additionally, newer advancements have been made in general deep generative models with the introduction of FM models [6], an alternative to DMs.

The use of DMs and FMs also opens up possibilities for other methodologies, such as latent diffusion models (LDM) [7] and latent flow matching (LFM) [8], which are unexplored in the field of flight trajectory generation. The proven TCVAE from [13] can be used to represent trajectories in a latent space, and the DiffTraj architecture can then be used to generate the latent representations to be decoded into trajectories.

Finally, while diffusion has also been explored for *prediction* in terminal airspace [22], our emphasis remains on *generation* rather than prediction, i.e., learning the underlying distribution to produce diverse, plausible samples [10, 23].

Although there is literature on trajectory prediction, and emerging work on generative modeling of trajectories, relatively few studies consider transfer learning of generative models in the ATM (or broader mobility) domain. For instance [24] explores pre-training on one airport and applying to another, albeit in a predictive (rather than generative) setting. In urban mobility, works such as TrajGAIL [25] frameworks to transfer from abundant to scarce data regimes. What sets our work apart is that, to our knowledge, it is the first study in the ATM domain to apply diffusion and flow-matching (including latent) generative architectures in a transfer-learning framework for landing-trajectory generation, conducting a systematic assessment of how much target-domain data is required when fine-tuning from a data-rich source.

Building on prior work in deep generative models, our work focuses on the under-explored question of transfer learning: we systematically test whether state-of-the-art deep generative models pretrained on a data-rich source can be efficiently fine-tuned on a data-scarce target to match or surpass target-only baselines.

III. MODEL ARCHITECTURES

We investigate four families of generative models for aircraft landing trajectories: *Diffusion Models (DMs)*, *Flow Matching (FM)*, *Latent Diffusion Models (LDMs)*, and *Latent Flow Matching Models (LFMs)*. All architectures are adapted from recent advances in spatio-temporal generative modeling and extended to the aviation domain.



A. Diffusion Models (DMs)

Diffusion probabilistic models [3, 4, 5] construct a forward process that gradually corrupts data with Gaussian noise, and a reverse process trained to denoise step by step. For trajectory synthesis, we adapt the *DiffTraj* architecture [17], which employs a UNet backbone with residual and attention blocks, together with wide-and-deep embeddings for conditional inputs (airport/runway). Unlike urban traffic where strict road-network constraints can be enforced, aviation requires flexibility in sparse three-dimensional airspaces. By removing road-specific constraints while retaining the robust denoising framework, this work represents one of the first systematic applications of diffusion models to aircraft landing trajectories.

B. Flow Matching (FM)

Flow Matching (FM) [6] is a recently introduced generative paradigm that treats synthesis as a continuous-time transport problem, learning velocity fields that map a simple prior distribution to the target data distribution. Compared to diffusion, FM eliminates the need for discrete noise schedules and sampling schedulers, potentially offering faster generation. To the best of our knowledge, this is the *first study to apply FM to trajectory generation in aviation*, extending a cutting-edge generative approach from computer vision into ATM research.

C. Latent Variants (LDM and LFM)

Both diffusion and FM models are combined with a Temporal Convolutional Variational Autoencoder (TCVAE) [13], which compresses trajectories into a structured latent representation. Operating in this latent space reduces dimensionality, accelerates training and sampling, and enables more compact generative modeling.

In the *Latent Diffusion Model (LDM)* [7], the denoising process occurs in latent space, and the decoded outputs preserve the temporal dependencies of flight dynamics. In the *Latent Flow Matching (LFM)* [8], the continuous transport dynamics are applied directly to latent embeddings, allowing FM to exploit compact trajectory representations.

To our knowledge, this work represents the *first use of LDMs and LFM for flight trajectory generation*, contributing novel architectures to the ATM domain.

D. Conditional Generation and Representation

The models incorporate contextual information through conditional generation. Each trajectory is associated with an airport identity token, encoded as a learned embedding and injected into the UNet backbone via the wide-and-deep conditioning module. At Zurich (LSZH), where the landing dataset includes runway identifiers, we additionally condition on the runway label. This allows the model to distinguish approach geometries that correspond to different runway directions, downwind legs, and base-turn structures.

Runway conditioning at LSZH is essential because the airport exhibits distinct multimodal approach patterns determined by the active runway. Without this conditioning, a generative model tends to average across these modes, producing

unrealistic mixtures or suppressing low-frequency approach corridors. Explicit conditioning separates these modalities and helps preserve the full diversity of LSZH approaches during training.

For Dublin (EIDW), the dataset includes a “runway” field, but all entries take the same value (“Dublin”), effectively meaning that runway identity is not distinguished in the data. As a result, all EIDW trajectories share a single categorical condition, and intra-airport variability is learned exclusively through their kinematic sequences. This asymmetry is intentional: pretrained LSZH models learn a rich set of runway-specific modes in a labeled environment, while fine-tuning at EIDW adapts these learned representations to the aggregated, unlabeled structure of the Dublin arrival flow.

All conditional inputs (airport token, and runway token when available) are embedded into learned vectors and applied at all stages of the UNet through feature-wise modulation. This mechanism is shared across the diffusion, flow-matching, latent-diffusion, and latent-flow-matching variants, ensuring a uniform conditioning strategy across architectures.

IV. EXPERIMENTAL SETTINGS

These four architectures form the basis for our transfer-learning study, in which models are pretrained on a data-rich source airport (Zürich, LSZH) and subsequently fine-tuned on a data-scarce target airport (Dublin, EIDW) under varying amounts of available local data. By comparing pretrained models against from-scratch baselines, we systematically evaluate how architectural choices impact cross-airport transferability.

A. Dataset

We use publicly available *landing-trajectory* datasets for Zürich (LSZH) and Dublin (EIDW) from the `traffic` library [26], originally derived from OpenSky [27]. LSZH is employed as the data-rich *source* domain due to its widespread use in prior trajectory-generation studies [11, 13]; EIDW serves as the *target* domain with comparable scale. To ensure comparability and computational efficiency, all trajectories are resampled to a fixed length of $T=200$ time steps using `traffic`, following established practice for sequence models [17]. Details about the data are summarized in Table I.

TABLE I. LANDING DATASETS USED IN THIS STUDY.

Airport	Period	Samples	Runways	ICAO
Zürich	2019-10–2019-11	~19,000	2	LSZH
Dublin	2019-10–2019-11	~20,000	4	EIDW

All datasets are standardized, with *separate* scalers for trajectory variables and conditional inputs. Scalers are fit on the training split and applied to validation/test to avoid leakage. For EIDW, runway labels are unavailable; during conditioning. We therefore use a single airport token (“EIDW”) instead of the runway identity, while LSZH includes runway-specific labels.



B. Evaluation Metrics

Evaluating generative models for flight trajectories is a nontrivial task: no single metric jointly captures statistical fidelity, geometric plausibility, and operational realism. Existing frameworks range from expert assessments and physics-aware simulation (e.g., BlueSky simulator [28]) to compact statistical criteria [10]. The former offer operational insight but are resource intensive; the latter support reproducible model selection but do not certify operational validity, because these metrics evaluate only statistical alignment and geometric similarity. They do not account for procedure design, aircraft performance envelopes, separation constraints, or controller-feasible behavior, which typically require expert assessment or physics-aware simulation.

Given our focus on transfer learning and comparative analysis across models and data budgets, we adopt a lightweight quantitative protocol. Following prior work [11, 13], metrics are computed in the lateral plane using longitude and latitude only—excluding altitude—to emphasize spatial alignment. When models are trained in a kinematic representation (track and groundspeed), geographic footprints are reconstructed under the same convention for evaluation.

We assess *similarity* to quantify distributional alignment between real and synthetic trajectories using quantitative, reproducible metrics that capture both statistical and structural properties. As in [11, 13], we report energy distance (e-distance), a Euclidean-distance-based measure of distributional discrepancy [29], together with Maximum Mean Discrepancy (MMD) [30] and Dynamic Time Warping (DTW) [10] to capture, respectively, global distributional differences and local shape-level deviations. Where informative, we also compute Kullback–Leibler divergence (KL) and Jensen–Shannon divergence (JSD), inspired by urban-traffic evaluation [17], which are effective for comparing spatial density maps (trajectory heatmaps) and provide a complementary view of how well the synthetic distribution approximates the real one.

For e-distance and MMD, trajectories are flattened over the temporal dimension and compared using pairwise Euclidean distances or Gaussian-kernel similarities, following the standard empirical estimators used in prior trajectory-generation work. DTW is applied to the *sequential* latitude–longitude representation of each trajectory and computes the minimum-cost alignment between a synthetic path and its corresponding real path under the same conditioning. This alignment allows local time shifts, enabling comparison of shapes even when trajectories differ slightly in timing or speed. We report the mean DTW cost over $N = 100$ synthetic–real pairs per model.

We evaluate *diversity* via dimensionality-reduction diagnostics. Principal Component Analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) visualize coverage, cluster structure, and outliers by projecting real and synthetic sets into a low-dimensional space. Following [16], trajectories are flattened along the temporal dimension before projection to enable a like-for-like comparison of distributional spread.

We also perform qualitative analysis with visualization,

plotting synthetic flight paths alongside real ones to assess whether shapes and patterns appear realistic. Such visual inspection, common in the evaluation of generative models [23], can reveal implausible paths or insufficient variation that aggregate metrics may obscure.

V. RESULTS AND DISCUSSION

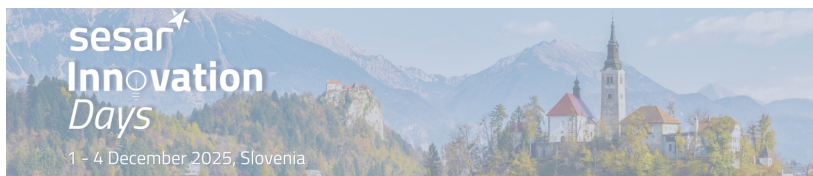
A. Transfer learning: rationale and setup

We investigate whether knowledge learned at a data-rich airport transfers to a data-scarce airport, thereby reducing target-domain data requirements for realistic trajectory generation. This experiment follows the transfer-learning paradigm—pretraining on a source domain and fine-tuning on a target domain—to test cross-airport generalization of our generative models in the aviation setting (17). The objective is to quantify how much pretraining improves similarity to the target distribution when only limited target data are available. Models are first *pretrained* on the Zürich (LSZH) landing dataset using the same configuration as in the baselines; they are then *fine-tuned* on Dublin (EIDW) with data fractions $s \in \{0, 5, 20, 50, 100\}\%$ of the EIDW training split (fixed train/validation/test partitions and identical conditioning across runs). As runway labels are unavailable for Dublin, the runway condition is replaced by a single airport token (“EIDW”). For each split s , we train for 100 epochs (batch size 32), generate $N=100$ condition-matched samples, and evaluate against the held-out EIDW test set.

The Dublin baseline for each architecture is trained from scratch using 100% of the EIDW training data only, with no exposure to LSZH data during initialization or training. These Dublin-only models serve as reference baselines against which all pretrained-and-fine-tuned models are compared.

For each model family we compare a *pretrained* model fine-tuned at $s \in \{0, 5, 20, 50, 100\}\%$ *against a baseline trained from scratch on 100% of EIDW*. This design probes how close a pretrained, low-data model can get to a strong full-data baseline at the target airport, thereby providing complementary evidence of transferability beyond local paired comparisons. Quantitative evaluation reports energy distance (e-distance), Maximum Mean Discrepancy (MMD), and Dynamic Time Warping (DTW), capturing distributional alignment and lateral shape agreement under consistent conditions. Metrics are computed on unsmoothed data; a secondary EWMA smoothing pass is used only for visualization in the longitude/latitude setting.

To support cross-airport generalization and avoid geographic leakage, the primary experiments operate in the *track/groundspeed* parameterization (with altitude and Δt), reconstructing (lat, lon) only for evaluation. This representation is explicitly *airport-agnostic*: it encodes kinematics rather than absolute geography, requires no choice of map projection or local tangent plane, and is invariant to rigid transforms (translation/rotation) of the spatial frame. As such, it avoids geodesy/projection artifacts that arise with longitude/latitude when comparing distant airports and prevents leakage of



location-specific geometry into the model [13, 31]. The conditioning scheme is likewise agnostic to runway enumeration at the target: a single “EIDW” token replaces runway identity, while LSZH retains runway labels during pretraining. For each target-data split we generate $N=100$ condition-matched samples and report mean \pm standard deviation for e-distance, MMD, and DTW (lower is better). In the tables below, **bold** denotes a *significant* improvement over the full-data Dublin baseline (Welch’s t -test, $p < 0.05$). Metrics are computed on unsmoothed data; trajectory and t-SNE visualizations use the same plotting settings as in the thesis.

Diffusion (DM)

Table II summarizes DM transfer performance. Zero-shot transfer (0%) performs markedly worse than the Dublin baseline across all metrics, confirming that purely source-trained lateral structure does not carry to a new terminal area without *any* target supervision. With only 5% of Dublin labels, the pretrained DM already *surpasses* the baseline on DTW (**23.21** vs. 29.10) and reduces MMD while keeping e-distance within a narrow margin. At 20%, the DM essentially closes the e-distance gap and achieves a further DTW reduction (**22.94**), indicating that fine-tuning aligns both global distributional fit and local path geometry. As the data budget grows to 50–100%, DTW falls steeply (**18.43** and **15.88**), while e-distance remains comparable to baseline. The small MMD uptick at larger splits is consistent with finite-sample variability with $N=100$ generations and does not appear in the visual diagnostics.

Qualitative evidence in Figure 1 shows that 5–20% models already reproduce the dominant, densely flown corridors on the eastern side, with fewer artifacts at 20% than at 5%. Residual errors concentrate in infrequent approach sectors (e.g., south–west entries), a pattern that persists, albeit attenuated, even at 100%. The t-SNE overlays in Figure 2 echo this: overlap between synthetic and real clusters tightens monotonically with more Dublin data, while sparsely populated regions remain hardest to match. These findings are consistent with the thesis-wide observation that DM exhibits the strongest *sample efficiency* among the tested architectures.

TABLE II. DM TRANSFER RESULTS ON DUBLIN (TRACK/GROUNDSPEED). LOWER IS BETTER.

Split	e-distance	MMD	DTW
Baseline	0.662 \pm 0.170	0.113 \pm 0.169	29.099 \pm 0.106
0.00	1.228 \pm 0.445	0.746 \pm 0.207	118.222 \pm 0.352
0.05	0.693 \pm 0.183	0.097 \pm 0.139	23.211 \pm 0.084
0.20	0.681 \pm 0.178	0.084 \pm 0.135	22.941 \pm 0.081
0.50	0.679 \pm 0.175	0.101 \pm 0.131	18.428 \pm 0.067
1.00	0.668 \pm 0.171	0.134 \pm 0.142	15.876 \pm 0.060

Flow Matching (FM)

FM exhibits smaller and less consistent transfer gains than DM (Table III). The zero-shot model underperforms the Dublin baseline on all metrics, though it degrades far less than DM at 0%, indicating a modest degree of cross-airport

TABLE III. FM TRANSFER RESULTS ON DUBLIN (TRACK/GROUNDSPEED). LOWER IS BETTER.

Split	e-distance	MMD	DTW
Baseline	0.694 \pm 0.180	0.094 \pm 0.153	26.554 \pm 0.098
0.00	0.750 \pm 0.216	0.368 \pm 0.179	37.490 \pm 0.125
0.05	0.720 \pm 0.190	0.115 \pm 0.161	28.859 \pm 0.100
0.20	0.713 \pm 0.187	0.071 \pm 0.160	27.397 \pm 0.094
0.50	0.676 \pm 0.177	0.262 \pm 0.171	21.607 \pm 0.081
1.00	0.697 \pm 0.186	0.081 \pm 0.154	24.975 \pm 0.092

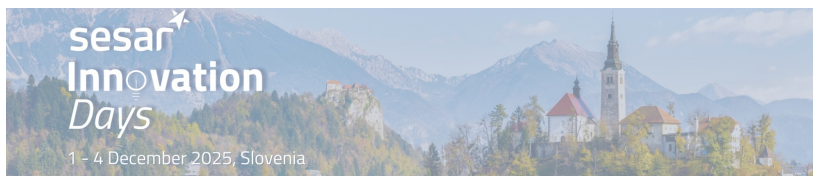
portability in the learned flow. With 5–20% target data, e-distance and MMD trend downward relative to 0%, but remain broadly comparable to baseline and present higher variability. The 50% split achieves a *large* DTW improvement (**21.61**), yet coincides with a spike in MMD, suggesting a sharper geometric fit along a subset of approach corridors at the expense of distributional coverage. At 100%, e-distance and MMD align with baseline while DTW remains significantly lower (**24.98**), indicating improved local path geometry but limited overall distributional gains.

Trajectory overlays in Figure 3 show that FM tends to favor specific high-density sectors (notably the upper-right approach), producing more outliers on the left-hand side of the terminal area. The 0.20 model spreads coverage more evenly but introduces noise, consistent with its low MMD yet modest DTW. The 1.00 model reduces artifacts and yields the visually cleanest paths, while still underrepresenting rarer west-facing approaches. The t-SNE plots in Figure 4 confirm the picture: partial cluster coverage at 0.05–0.20, improved but incomplete overlap at 1.00, and residual outliers. Overall, FM benefits from transfer, but its gains are smaller and more variable than DM’s, in line with the thesis baseline where FM lagged in raw performance.

Latent Diffusion (LDM)

LDM shows the same qualitative pattern as DM but with smaller effect sizes (Table IV). Without target data (0%), performance lags the baseline, reflecting the need to adapt the latent generative mapping to local arrival structure. With 5% data, LDM narrows the gap; at 20% it reaches baseline-level e-distance/MMD and achieves a *significant* DTW improvement (**21.40** vs. 22.78). Larger splits (50–100%) further reduce DTW (**19.64**, **20.04**); MMD decreases up to 20% then increases—consistent with higher sensitivity of kernel metrics in finite- N settings.

Visually (Figure 5), fine-tuning reduces outliers and stabilizes path geometry, particularly on the high-density eastern approaches; west-facing trajectories remain comparatively underrepresented, consistent with their lower frequency in the data. The t-SNE overlays (Figure 6) show steadily improving coverage of real clusters from 5% to 100%, with residual gaps confined to sparse regions. LDM thus benefits from pretraining and target adaptation, delivering robust DTW gains at moderate data budgets, though it does not match DM’s sample efficiency.



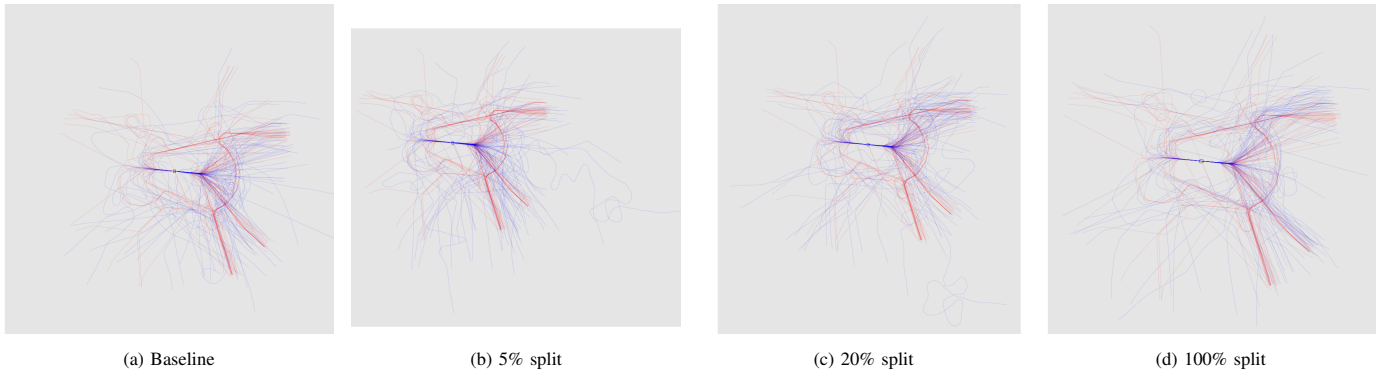


Figure 1. DM: real (red) vs. generated (blue) trajectory overlays across Dublin data fractions.

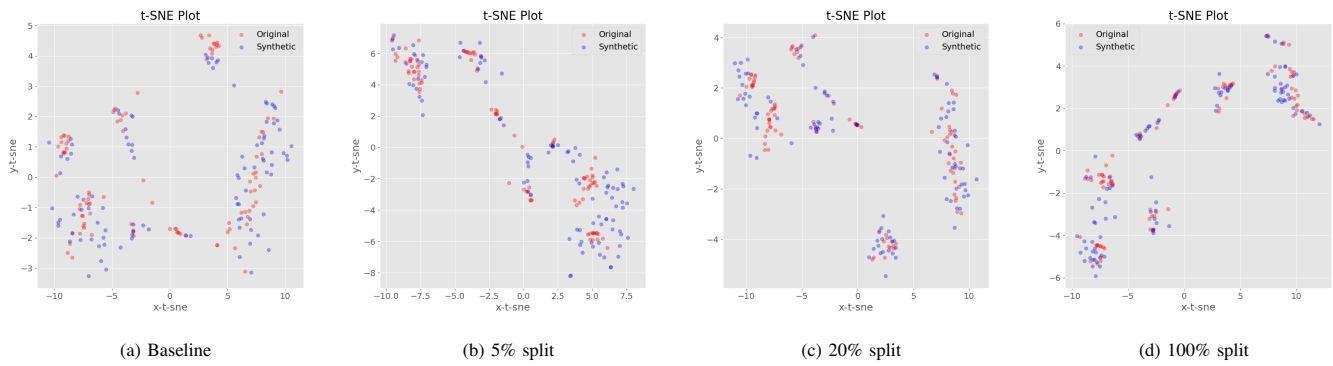


Figure 2. DM: t-SNE overlays (real in red, generated in blue) across Dublin data fractions.

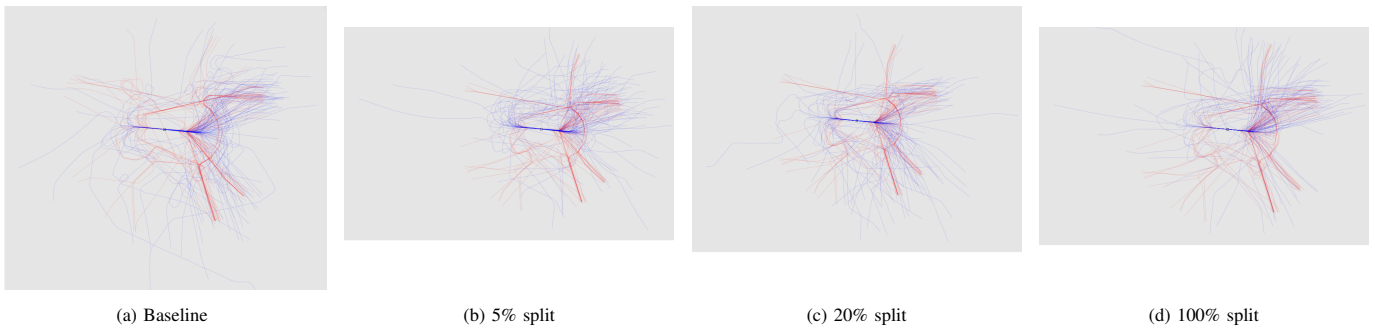


Figure 3. FM: real (red) vs. generated (blue) trajectory overlays across Dublin data fractions.

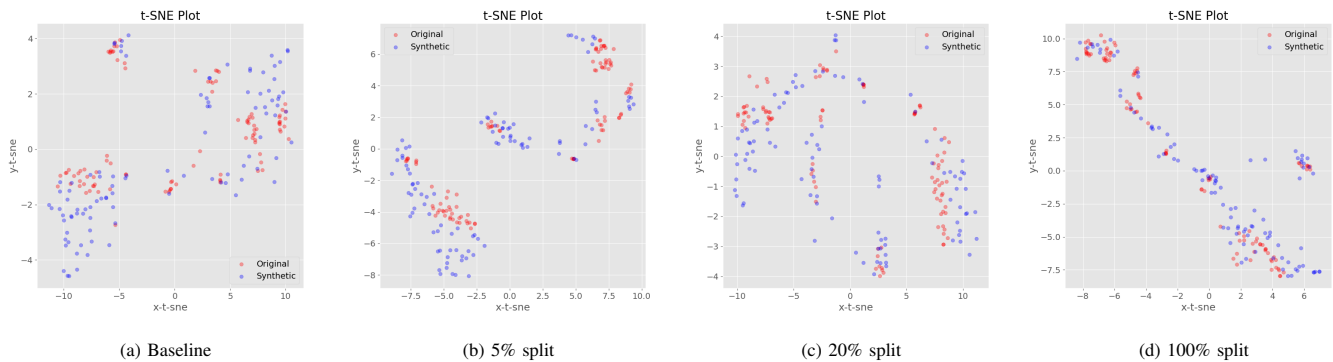


Figure 4. FM: t-SNE overlays (real in red, generated in blue) across Dublin data fractions.

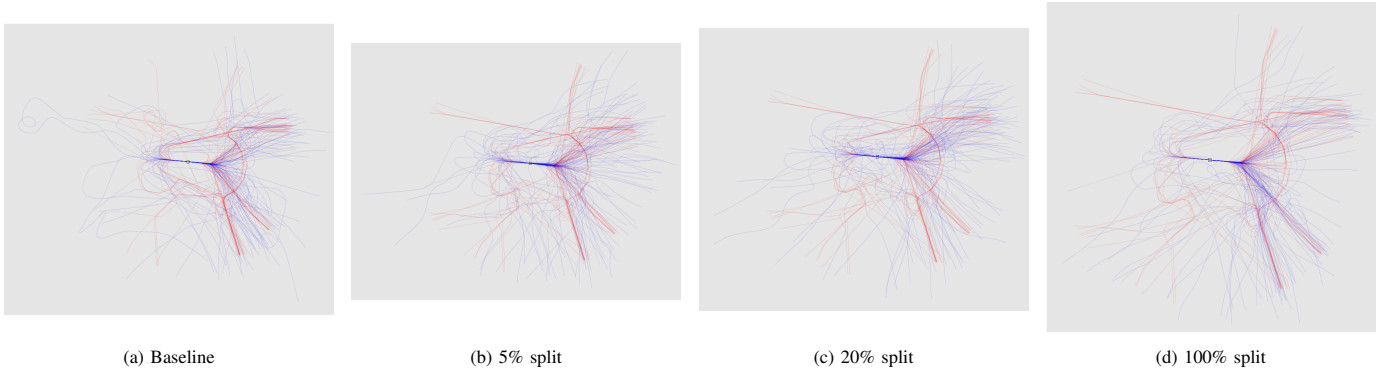


Figure 5. LDM: real (red) vs. generated (blue) trajectory overlays across Dublin data fractions.

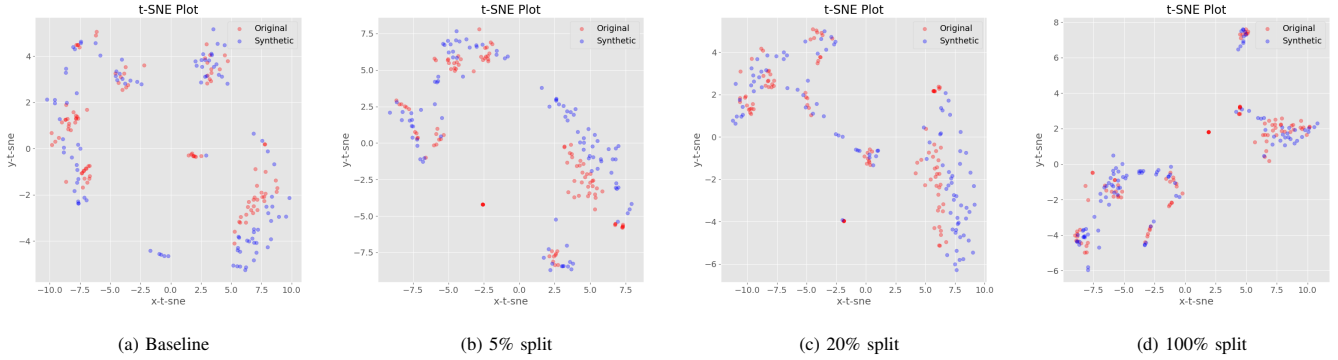


Figure 6. LDM: t-SNE overlays (real in red, generated in blue) across Dublin data fractions.

TABLE IV. LDM TRANSFER RESULTS ON DUBLIN (TRACK/GROUNDSPEED). LOWER IS BETTER.

Split	e-distance	MMD	DTW
Baseline	0.683 ± 0.173	0.128 ± 0.126	22.779 ± 0.084
0.00	0.743 ± 0.227	0.201 ± 0.108	34.800 ± 0.122
0.05	0.691 ± 0.179	0.162 ± 0.142	24.908 ± 0.088
0.20	0.683 ± 0.178	0.133 ± 0.144	21.401 ± 0.076
0.50	0.669 ± 0.174	0.188 ± 0.160	19.640 ± 0.074
1.00	0.667 ± 0.172	0.234 ± 0.144	20.038 ± 0.078

Latent Flow Matching (LFM)

Among latent models, LFM starts from the weakest Dublin baseline but gains the most from transfer (Table V). With 5% data, both e-distance and DTW significantly improve over the baseline, indicating that a small amount of target supervision suffices to tune the latent flow to local structure. At 20% the improvements strengthen across *all* metrics, including a pronounced MMD drop (from 0.168 to **0.081**), which suggests better coverage of the high-density regions. As the budget grows to 50–100%, DTW continues to decrease (**34.46** and **27.43**), and MMD reaches **0.060** at 100%. These results are consistent with the thesis insight that LFM benefits disproportionately from exposure to *diverse* target examples.

The trajectory overlays in Figure 7 show a clear reduction of noise and spurious paths from 5% through 100%, with realistic reproduction of the main eastern corridors and

persistent difficulty in the rare west-facing approaches. The t-SNE visualizations in Figure 8 confirm the progression: poor baseline coverage, improved cluster overlap at 20%, and substantially better alignment at 100%, albeit with residual outliers. While LFM does not match DM’s sample efficiency, its late-stage performance becomes competitive once sufficient target data are available.

TABLE V. LFM TRANSFER RESULTS ON DUBLIN (TRACK/GROUNDSPEED). LOWER IS BETTER.

Split	e-distance	MMD	DTW
Baseline	0.809 ± 0.212	0.168 ± 0.142	37.980 ± 0.116
0.00	0.775 ± 0.215	0.277 ± 0.167	48.915 ± 0.166
0.05	0.732 ± 0.186	0.147 ± 0.142	35.604 ± 0.127
0.20	0.705 ± 0.182	0.081 ± 0.151	32.555 ± 0.119
0.50	0.704 ± 0.183	0.174 ± 0.182	34.462 ± 0.116
1.00	0.677 ± 0.177	0.060 ± 0.160	27.433 ± 0.104

Summary across models

Pretraining consistently enhances *sample efficiency* across architectures. DM shows the strongest transfer: with only 5–20% target data it reaches baseline-level e-distance/MMD and *significantly* improves DTW. LDM benefits similarly, though gains are smaller. LFM begins from a weaker baseline but exhibits large relative improvements as more target data become available, culminating in strong DTW/MMD at 100%. FM benefits from transfer as well, but its gains are smaller



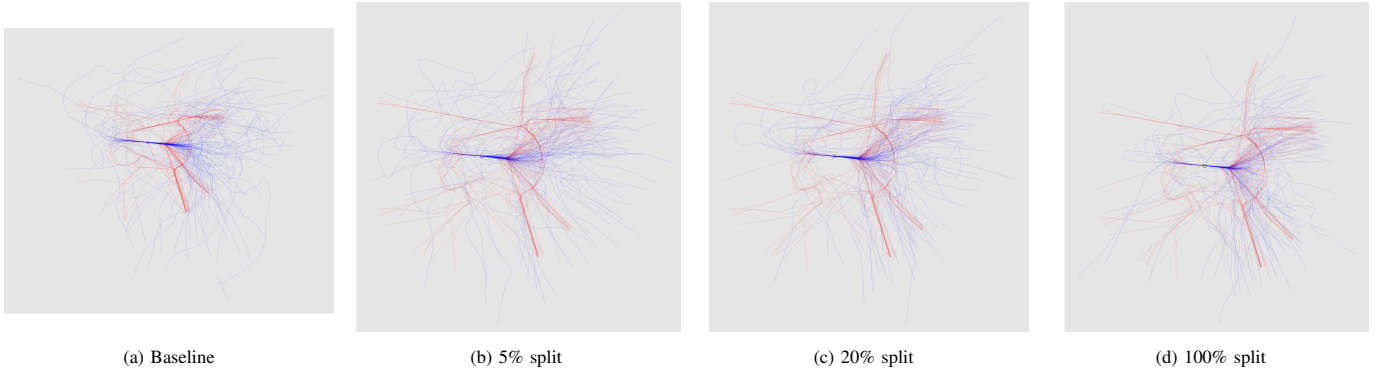


Figure 7. LFM: real (red) vs. generated (blue) trajectory overlays across Dublin data fractions.

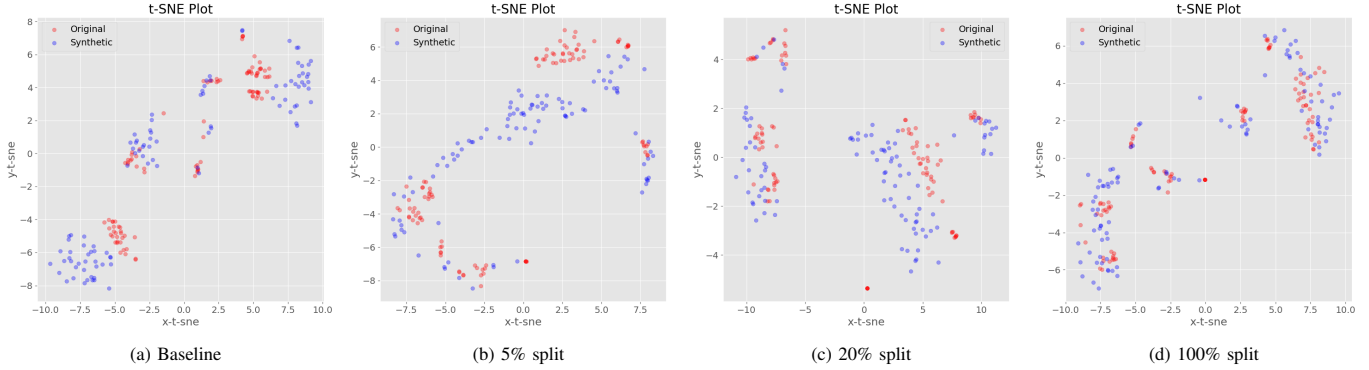


Figure 8. LFM: t-SNE overlays (real in red, generated in blue) across Dublin data fractions.

and more variable; it achieves competitive DTW at larger data budgets while lagging on distributional coverage in some splits. Across models, dominant approach corridors are reproduced reliably after transfer, while rarer runway usages (e.g., west-facing arrivals) remain underrepresented—highlighting opportunities for targeted augmentation or curriculum-based fine-tuning to emphasize rare modes.

VI. CONCLUSION

This paper presented, to the best of our knowledge, the first systematic study of transfer learning for deep generative models of aircraft landing trajectories. We evaluated four generative model families on transfer from a data-rich source airport (Zürich, LSZH) to a data-scarce target airport (Dublin, EIDW). Our findings show that pretraining consistently improves sample efficiency across architectures. Diffusion-based models exhibited the strongest transfer, achieving competitive performance with as little as 5% of the target data and matching or surpassing full-data baselines around 20%. Latent variants also benefited from pretraining: LDM followed the same trend as DM with smaller effect sizes, while LFM, despite its weaker baseline, gained disproportionately from additional target data and became competitive at larger splits. FM models also benefited from transfer, though their improvements were smaller and more variable. Across all models, transfer learning enabled realistic reproduction of dominant approach corridors while rare modes, such as west-facing

arrivals at Dublin, remained underrepresented. These results demonstrate that generative knowledge can be effectively transferred between airports, reducing the amount of local data required for high-quality trajectory synthesis. Future work should explore strategies to better capture rare trajectory patterns, including targeted augmentation, curriculum-based fine-tuning, and multimodal conditioning (e.g., weather). By establishing transfer learning as a viable approach, this study broadens the applicability of synthetic data generation in ATM, supporting scalable data-driven solutions even in operational contexts where historical records are limited.

ACKNOWLEDGMENT

This paper is based on research conducted within the SynthAIR project, which has received funding from the SESAR Joint Undertaking under the European Union’s Horizon Europe research and innovation program (grant agreement No. 101114847).

REFERENCES

- [1] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” Dec. 2022.
- [2] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” Jun. 2014.
- [3] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli, “Deep unsupervised learning using nonequilibrium thermodynamics,” Nov. 2015.



- [4] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” Dec. 2020.
- [5] J. Song, C. Meng, and S. Ermon, “Denoising diffusion implicit models,” Oct. 2022.
- [6] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Le, “Flow matching for generative modeling,” Feb. 2023.
- [7] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. (2021, Dec.) High-resolution image synthesis with latent diffusion models.
- [8] Q. Dao, H. Phung, B. Nguyen, and A. Tran, “Flow matching in latent space,” Jul. 2023.
- [9] EuroControl and Partners, “The FLY AI report demystifying and accelerating AI in aviation/ATM,” EUROCONTROL, p. 74, Mar. 2020.
- [10] X. Olive, J. Sun, M. Murça, and T. Krauth, “A framework to evaluate aircraft trajectory generation methods,” Oct. 2021.
- [11] S. Wijnands, A. Sharpanskykh, and K. Aly, “Generation of synthetic aircraft landing trajectories using generative adversarial networks,” *Conference: SESAR Innovation Days 2024*, Oct. 2024.
- [12] A. Murad and M. Ruocco, “Synthetic aircraft trajectory generation using time-based VQ-VAE,” Apr. 2025.
- [13] T. Krauth, A. Lafage, J. Morio, X. Olive, and M. Waltert, “Deep generative modelling of aircraft trajectories in terminal maneuvering areas,” *Machine Learning with Applications*, vol. 11, p. 100446, Mar. 2023.
- [14] D. Lee, S. Malacarne, and E. Aune, “Vector quantized time series generation with a bidirectional prior model,” Apr. 2023.
- [15] X. Chen, J. Xu, R. Zhou, W. Chen, J. Fang, and C. Liu, “TrajVAE: A variational AutoEncoder model for trajectory generation,” *Neurocomputing*, vol. 428, pp. 332–339, Mar. 2021.
- [16] J. Yoon, D. Jarrett, and M. van der Schaar, “Time-series generative adversarial networks,” in *Advances in Neural Information Processing Systems*, vol. 32. Curran Associates, Inc., 2019.
- [17] Y. Zhu, Y. Ye, S. Zhang, X. Zhao, and J. Yu, “DiffTraj: Generating GPS trajectory with diffusion probabilistic model,” *Advances in Neural Information Processing Systems*, vol. 36, pp. 65 168–65 188, Dec. 2023.
- [18] Y. Zhu, J. J. Yu, X. Zhao, Q. Liu, Y. Ye, W. Chen, Z. Zhang, X. Wei, and Y. Liang, “ControlTraj: Controllable trajectory generation with topology-constrained diffusion model,” in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, pp. 4676–4687.
- [19] Z. Kong, W. Ping, J. Huang, K. Zhao, and B. Catanzaro, “DiffWave: A versatile diffusion model for audio synthesis.”
- [20] T. Wei, Y. Lin, S. Guo, Y. Lin, Y. Huang, C. Xiang, Y. Bai, and H. Wan, “Diff-RNTraj: A structure-aware diffusion model for road network-constrained trajectory generation,” pp. 1–15, conference Name: IEEE Transactions on Knowledge and Data Engineering.
- [21] I. Gultepe, R. Sharman, P. D. Williams, B. Zhou, G. Ellrod, P. Minnis, S. Trier, S. Griffin, S. S. Yum, B. Gharabaghi, W. Feltz, M. Temimi, Z. Pu, L. N. Storer, P. Kneringer, M. J. Weston, H.-y. Chuang, L. Thobois, A. P. Dimri, S. J. Dietz, G. B. França, M. V. Almeida, and F. L. A. Neto, “A review of high impact weather for aviation meteorology,” vol. 176, no. 5, pp. 1869–1921.
- [22] Y. Yin, S. Zhang, Y. Zhang, Y. Zhang, and S. Xiang, “Aircraft trajectory prediction in terminal airspace with intentions derived from local history,” *Neurocomputing*, vol. 615, p. 128843, Jan. 2025.
- [23] K. P. Murphy, *Probabilistic Machine Learning: Advanced Topics*. MIT Press, Jan. 2023.
- [24] T. Kravaris and G. Vouros, “Transferable aircraft trajectory prediction with generative deep imitation learning,” *International Journal of Data Science and Analytics*, vol. 20, pp. 1977–1999, 06 2024.
- [25] S. Choi, J. Kim, and H. Yeo, “Trajgail: Generating urban vehicle trajectories using generative adversarial imitation learning,” *Transportation Research Part C: Emerging Technologies*, vol. 128, p. 103091, 2021.
- [26] X. Olive, “traffic, a toolbox for processing and analysing air traffic data,” *Journal of Open Source Software*, vol. 4, p. 1518, 2019.
- [27] M. Schäfer, M. Strohmeier, V. Lenders, I. Martinovic, and M. Wilhelm, “Bringing up OpenSky: A large-scale ADS-b sensor network for research,” in *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*, Apr. 2014, pp. 83–94.
- [28] J. M. Hoekstra and J. Ellerbroek, “BlueSky ATC simulator project: an open data and open source approach,” *Proceedings of the seventh International Conference for Research on Air Transport (ICRAT)*, Jun. 2016.
- [29] G. Szekely and M. Rizzo, “Testing for equal distributions in high dimension,” *InterStat*, vol. 5, Nov. 2004.
- [30] T. Viehmann, “Partial wasserstein and maximum mean discrepancy distances for bridging the gap between outlier detection and drift detection,” Jun. 2021.
- [31] J. P. Snyder, “Map projections: A working manual,” ISSN: 2330-7102 Publication Title: Professional Paper.

