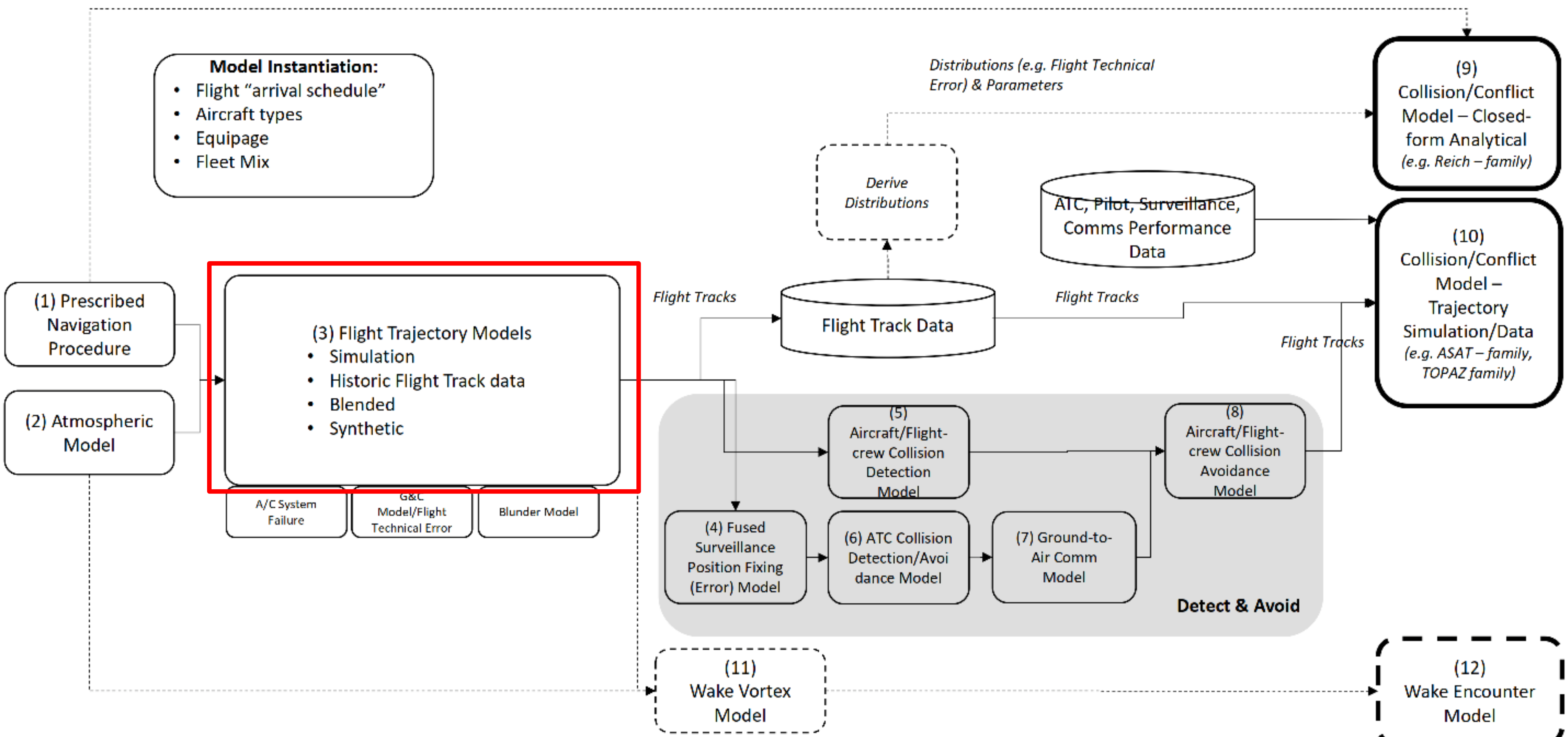
Using AI to Generate Synthetic Flight Tracks for Collision Risk Analysis

Shahab Aref, John Shortle, Lance Sherry
George Mason University

AI4SE & SE4AI Workshop
October 12, 2023

GEORGE MASON UNIVERSITY

Superset of Collision Risk Functions



Thapa, A., J. Shortle, L. Sherry. 2023. Air-to-Air Collision Risk Models (CRM) in the Terminal Airspace. 2023 Integrated Communication, Navigation and Surveillance Conference (ICNS)..

Flight Trajectory Sources

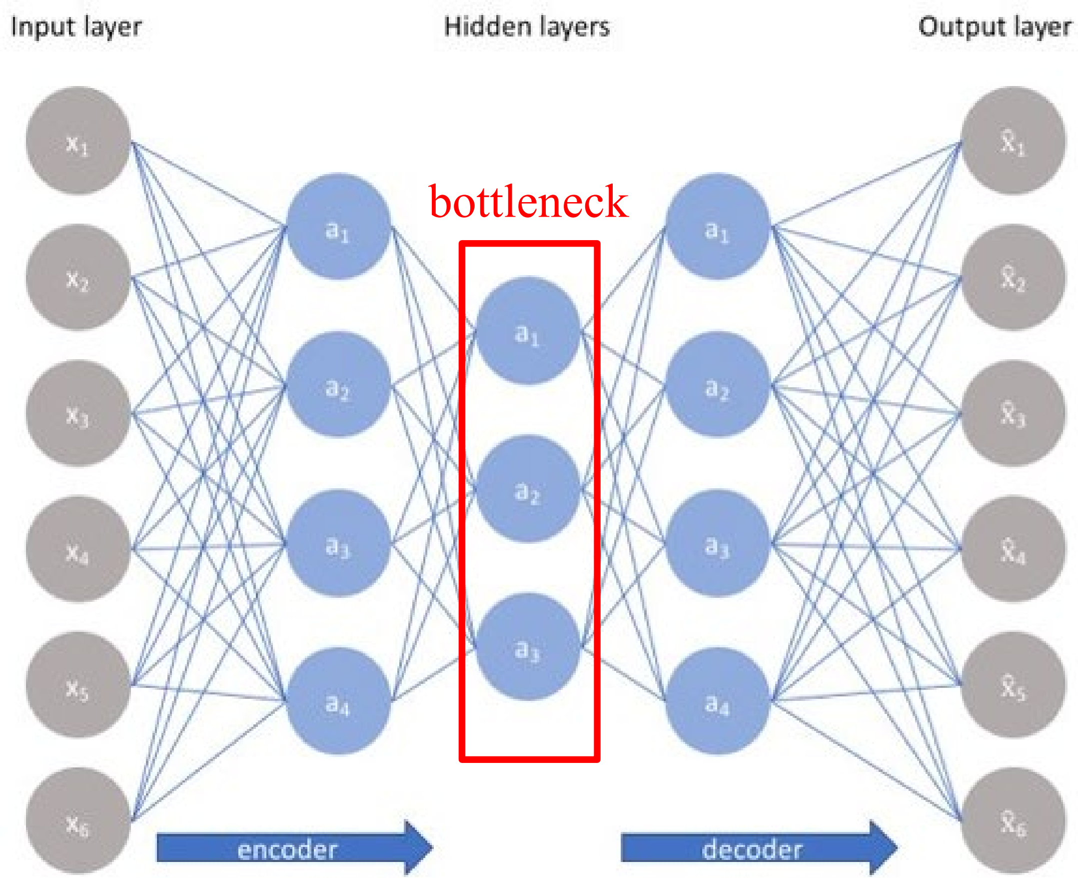
- Historic tracks
  - Limited data to observe / analyze rare events
- Simulated tracks
  - Model may not capture full realism
- Synthetic tracks
  - Potential to generate new (never been seen) tracks that are realistic and follow distributions of historic tracks (Krauth et al. 2023)

Krauth, Timothé, Adrien Lafage, Jérôme Morio, Xavier Olive, and Manuel Waltert. 2023. Deep generative modelling of aircraft trajectories in terminal maneuvering areas. *Machine Learning with Applications*, 11, 10046.

Objective

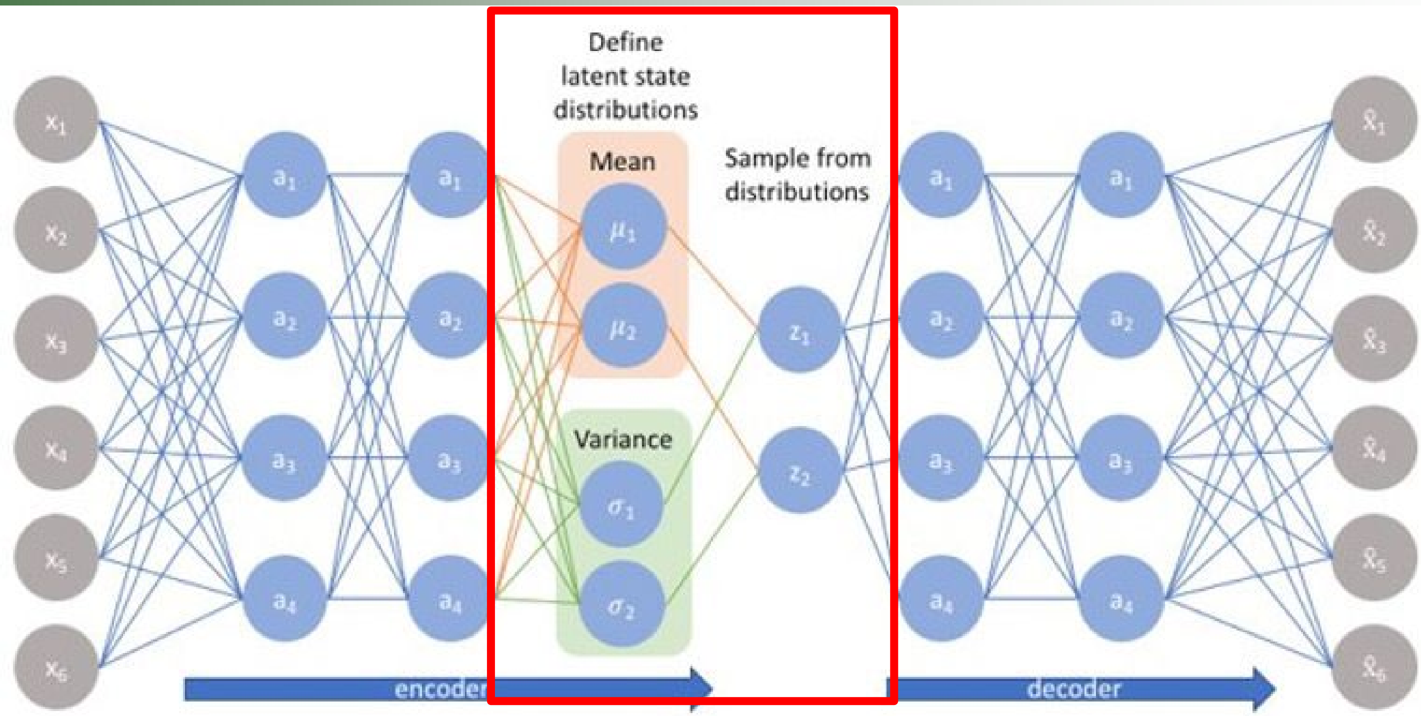
- Investigate the feasibility of generating synthetic flight tracks using variational autoencoder and related methods

Autoencoder

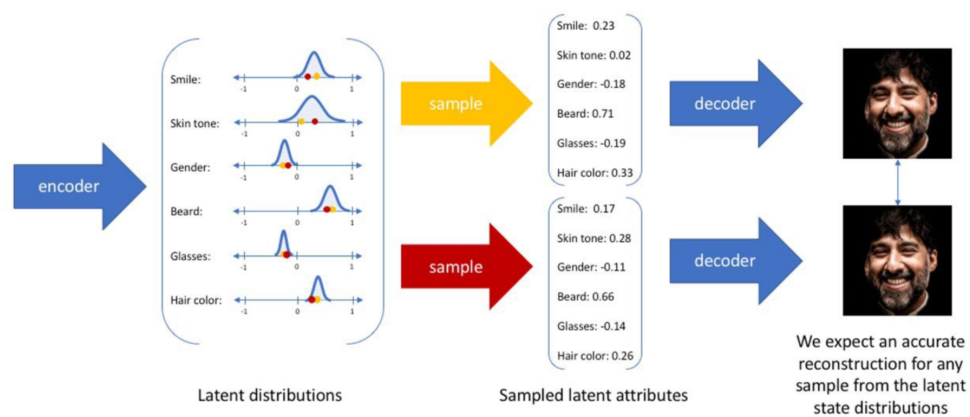


The bottleneck constrains the amount of information that can traverse the full network, forcing a learned compression of the input data.

Variational Autoencoder

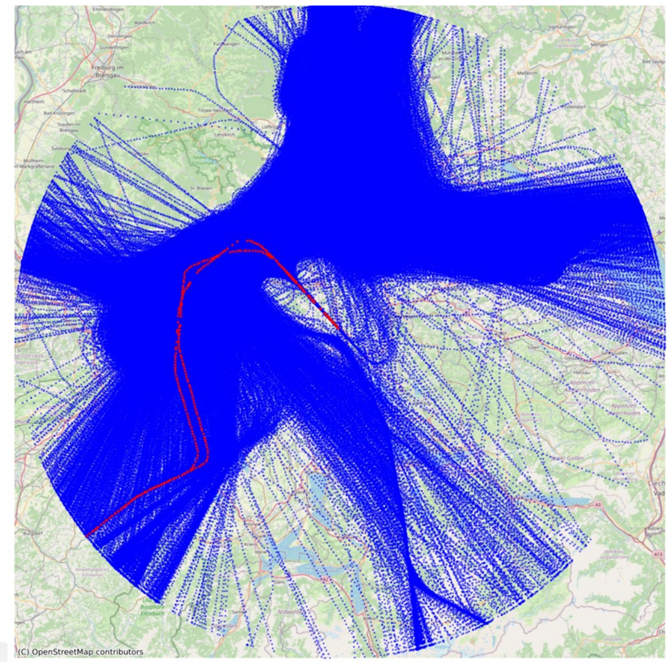


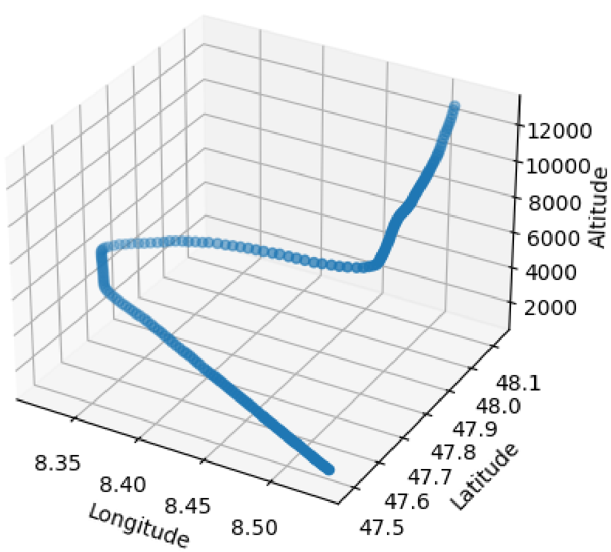
Provides a probabilistic manner for describing an observation in latent space

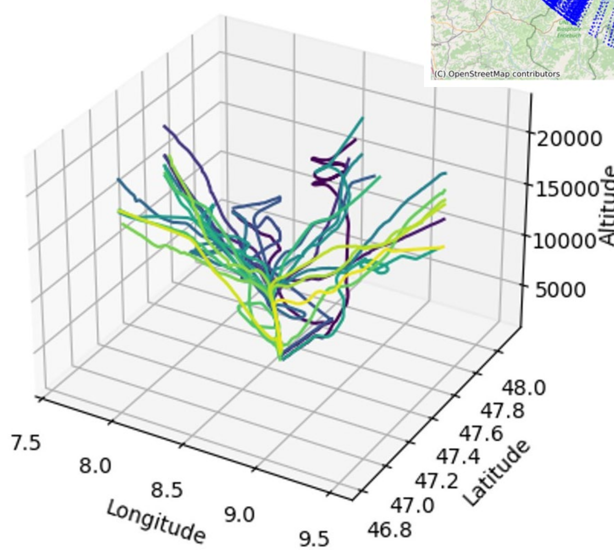


Data Summary

- Location: Zurich Airport (LSZH)
- Year: 2019, 2 months
- Runway: 14
- Number of flights: 14,441
- Source: OpenSky Network

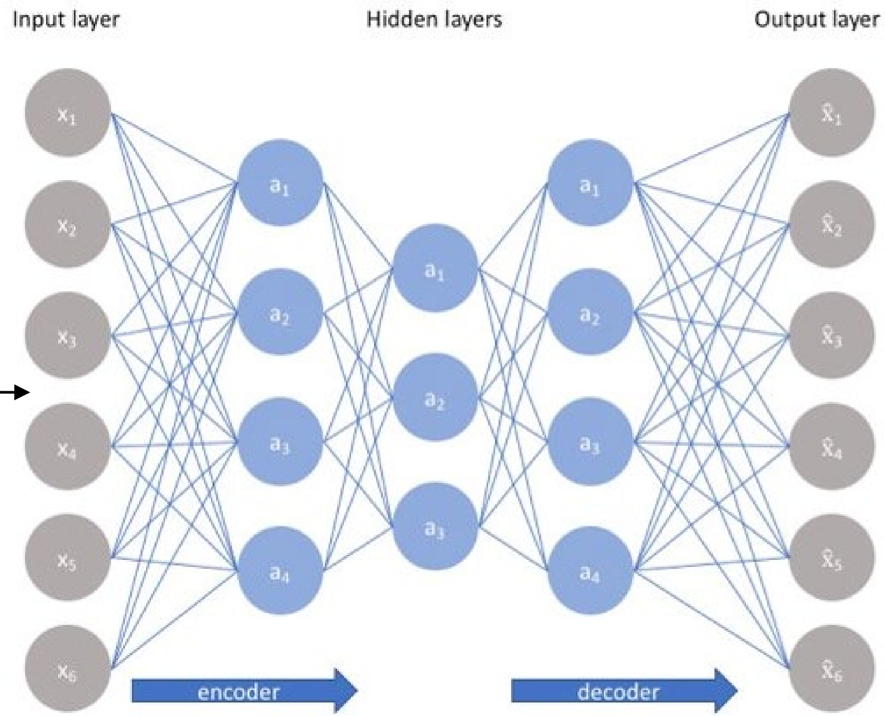




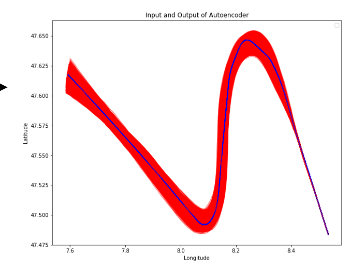


Conceptual Summary

Historic and/or simulated flight tracks



Synthetic tracks



Encoder breaks tracks into segments / features and captures their variability

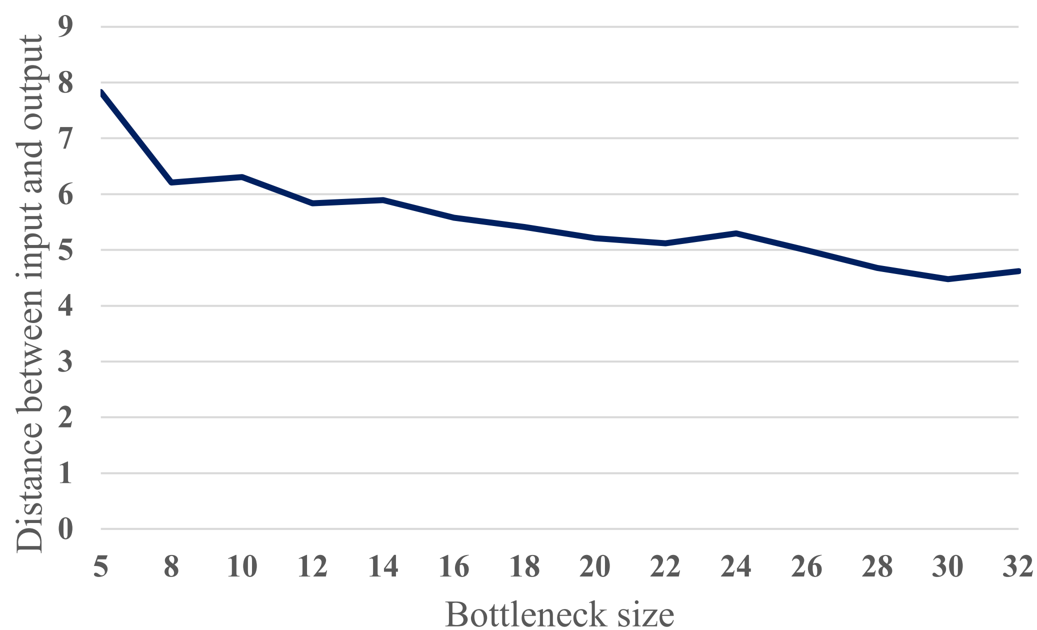
Fundamental track building blocks (distance, speed, headings, segments, turn radii, etc.)

Decoder combines segments with different variability in a way that is compatible

Training Parameters

- Architecture: Encoder and Decoder: 3 Layers (input, 128, 64, Bottleneck)
- Input = 600
- Bottleneck = 30
- ReLU
- Optimizer: Adam
- Batch size = 32
- Learning rate = 0.001
- Epoch = 1000
- Recon loss factor = 2.8
- KL loss factor = 0.001
- Ave. distance between real and synthetic test data = 0.5277 km

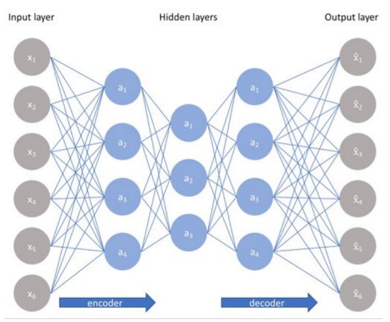
Autoencoder error as a function of bottleneck size



Autoencoder Sample Output

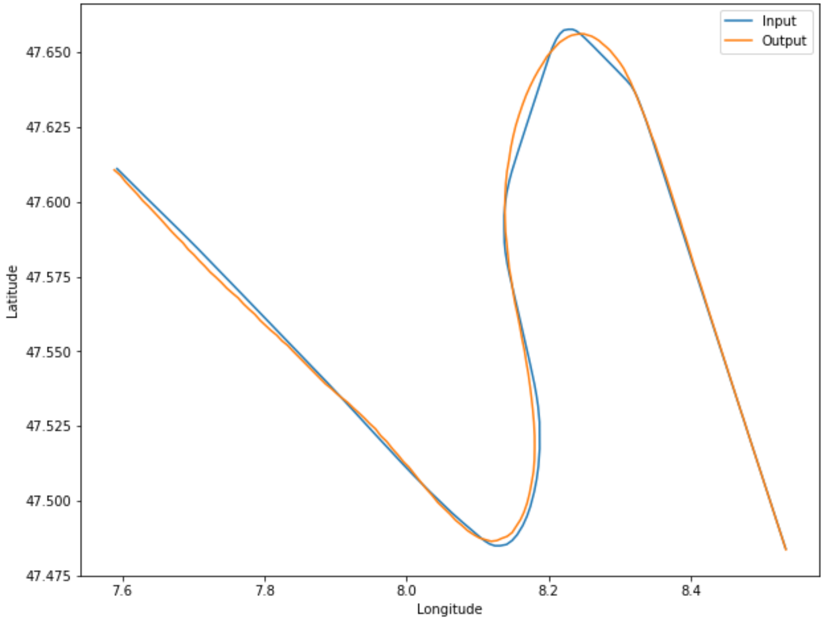
Trained autoencoder

Test track
(not in training set)



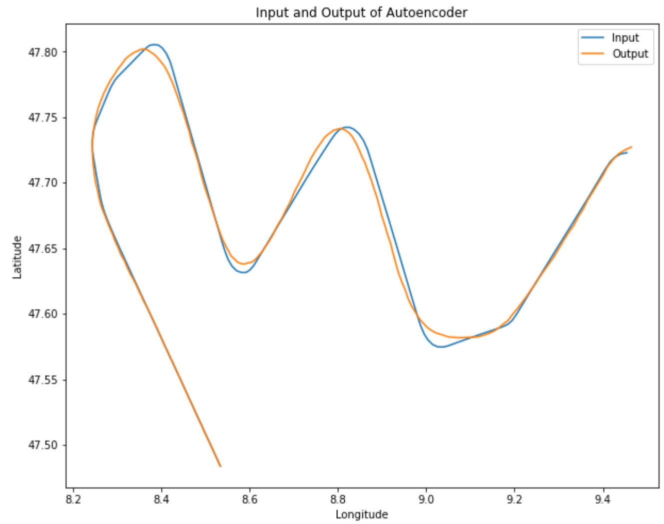
Output track

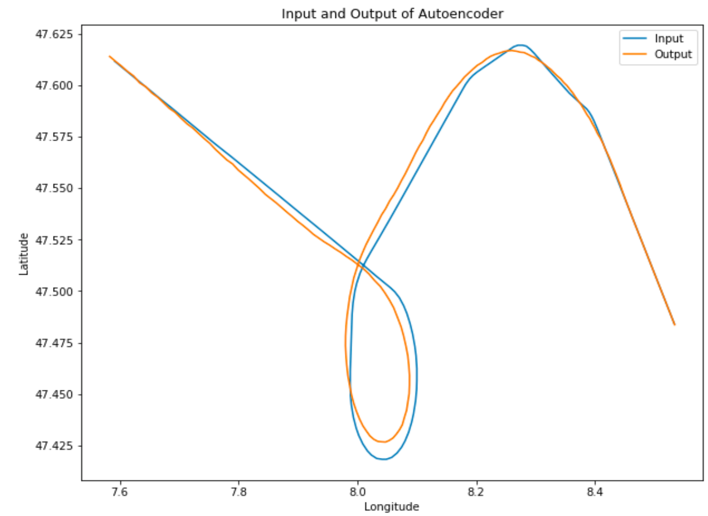
Example input / output

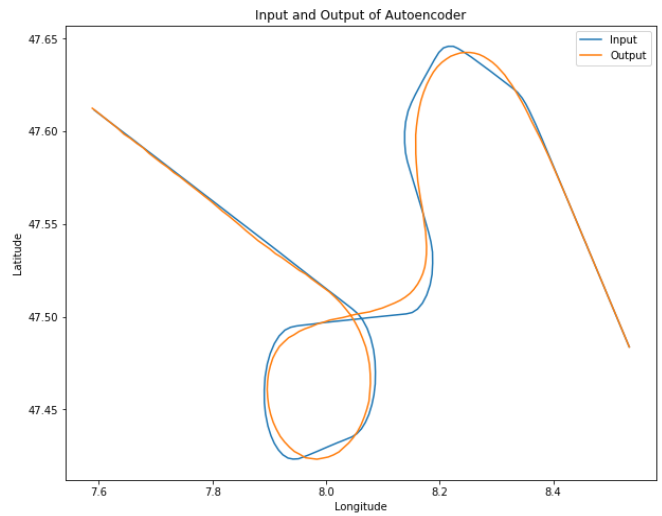


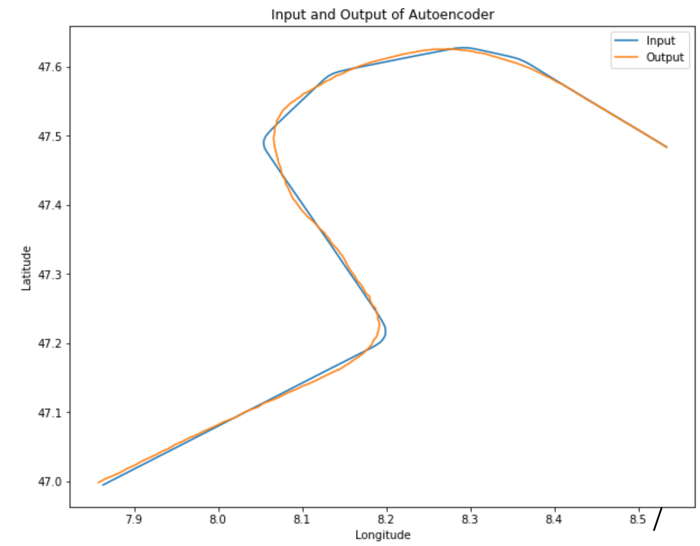
Example Results

Trained autoencoder can replicate complicated arrival tracks



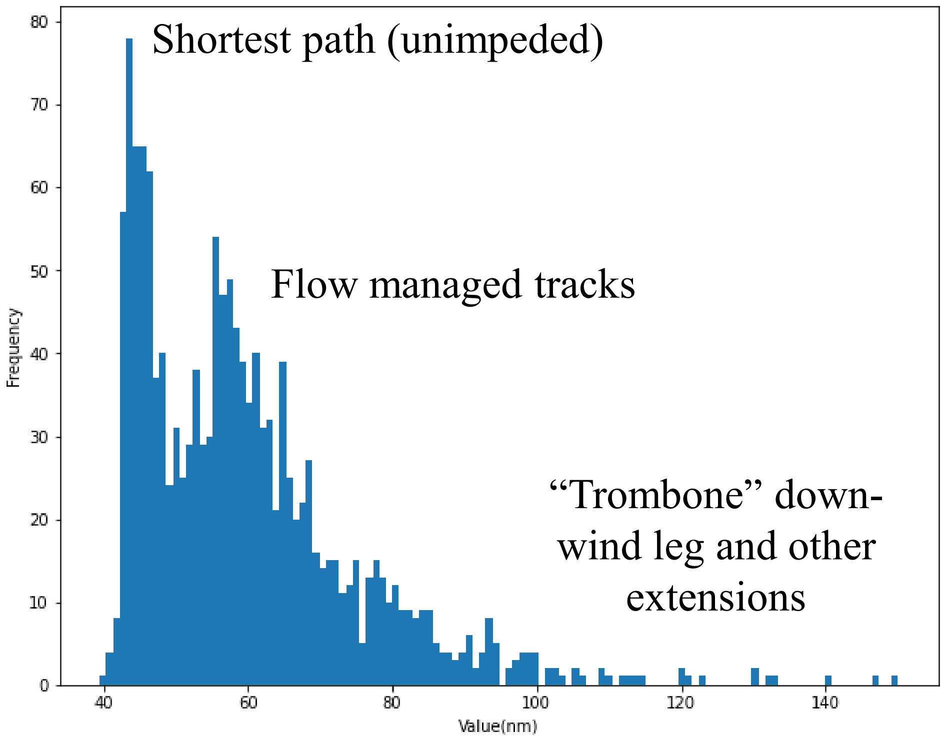






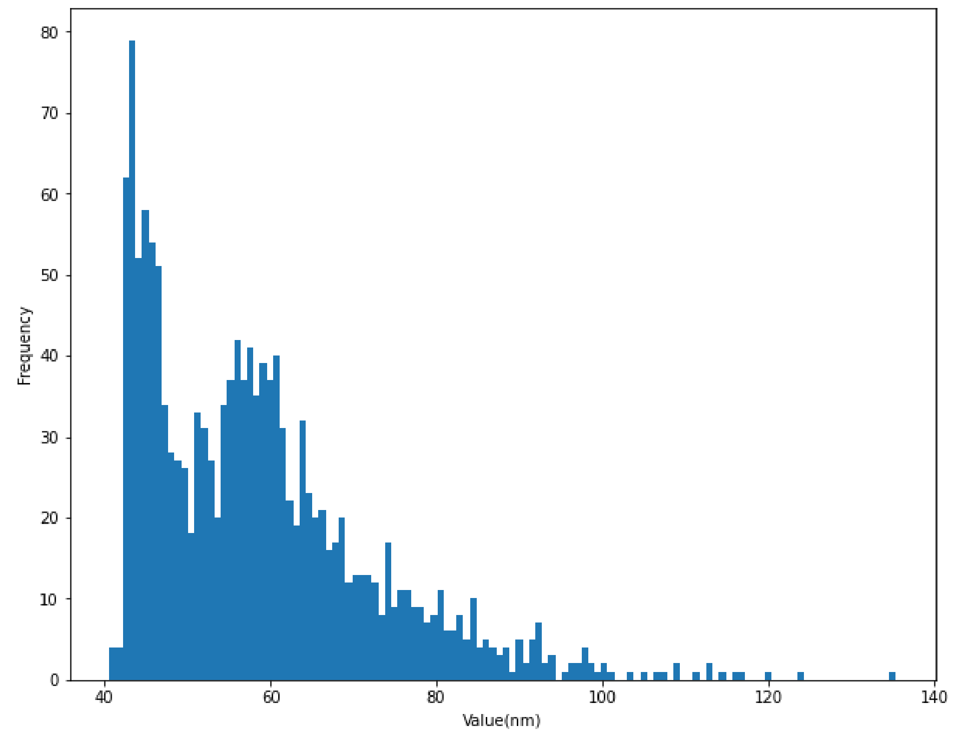
Comparing Along-Track Distance

Historic Flight Tracks



Along-track Distance (nm)

Synthetic Flight Tracks

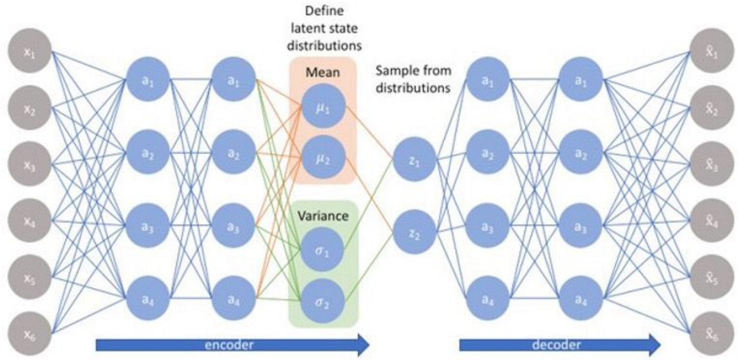


Along-track Distance (nm)

Variational Autoencoder Output

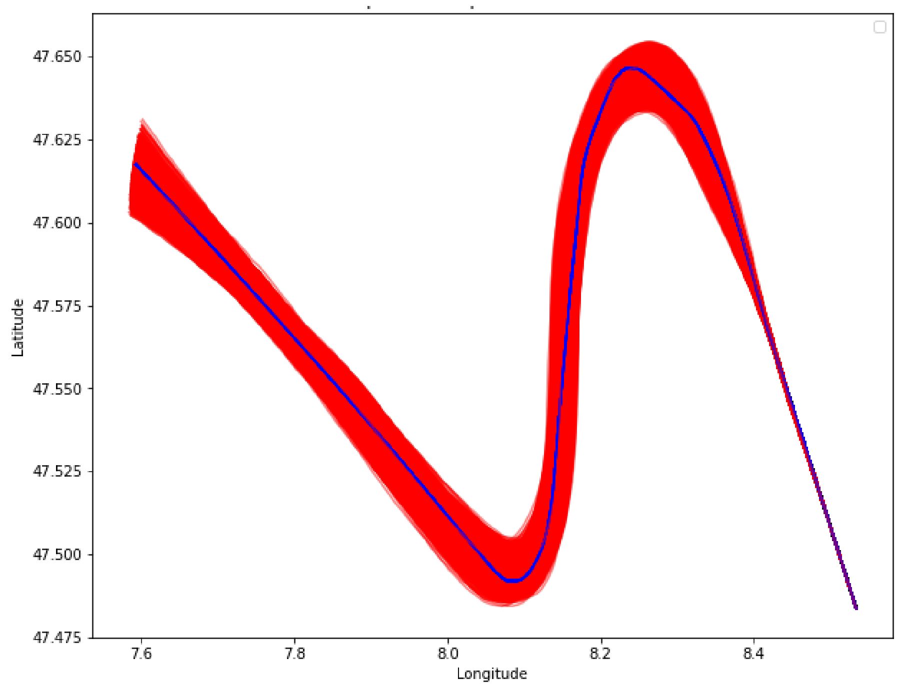
Trained variational autoencoder

Single track
(not in training set)



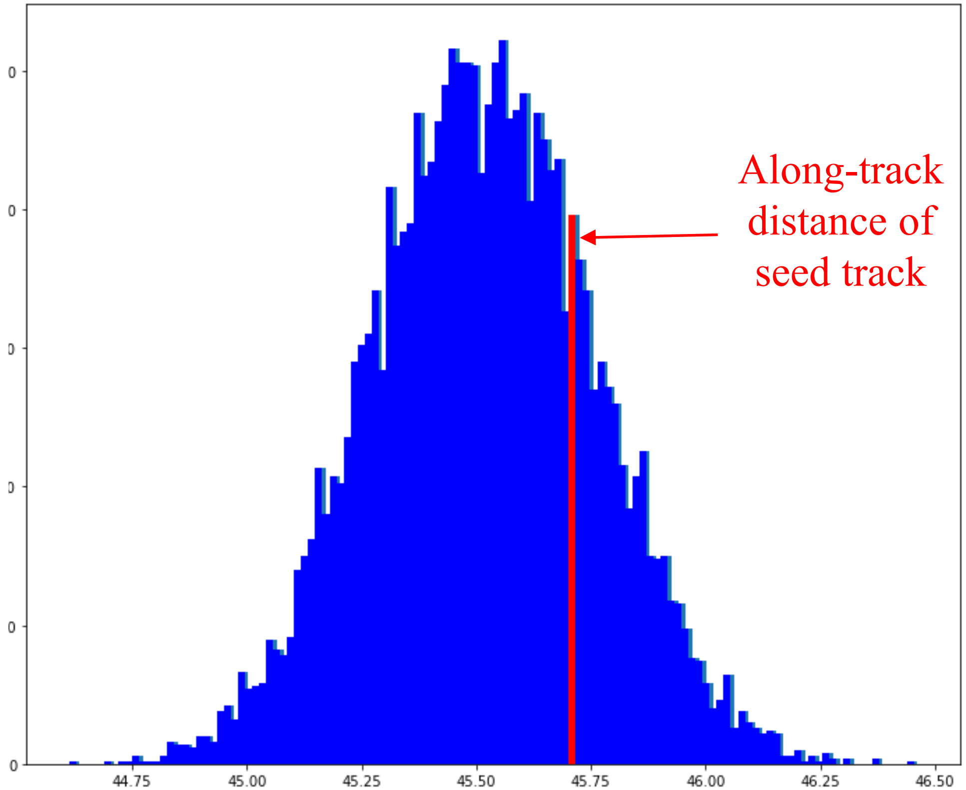
Multiple random tracks

100,000 generated flight tracks from 1 seed track

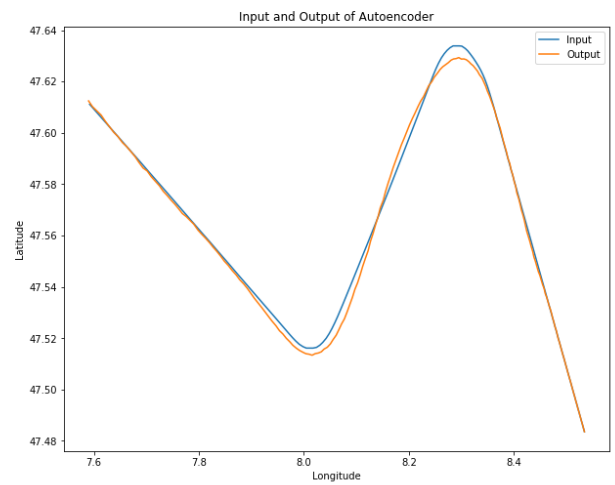


Distribution of Synthetic Tracks

10000 synthetic tracks generated

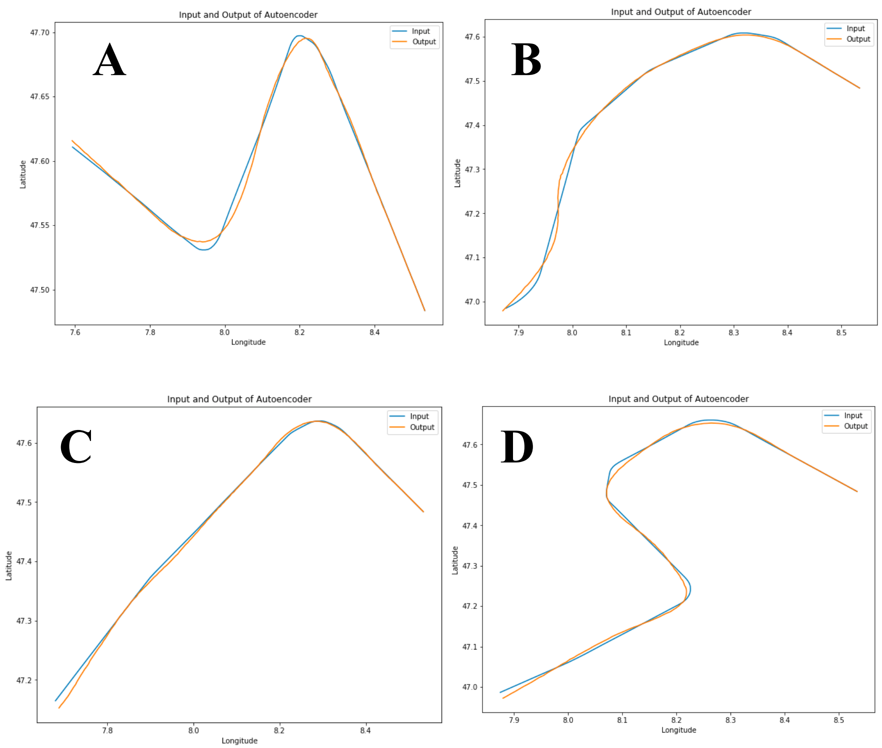


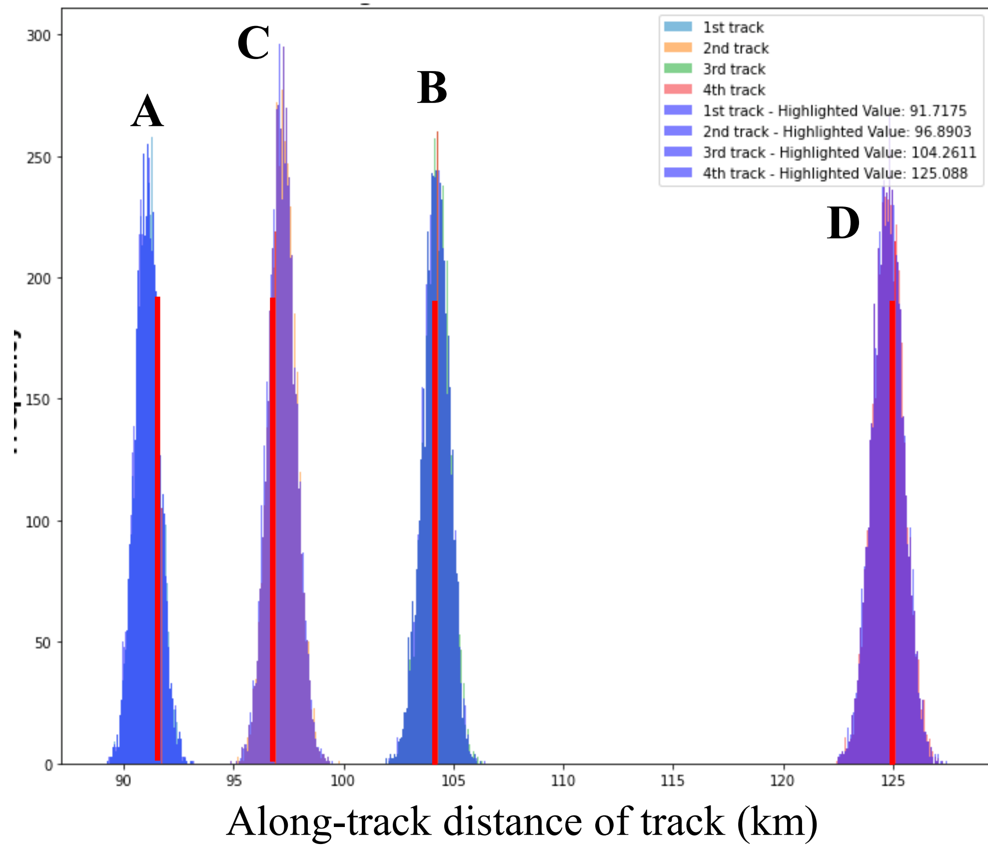
Along-track distance of track (nm)



Even though seed track is not the mean, the VAE generates synthetic tracks according to the trained data set

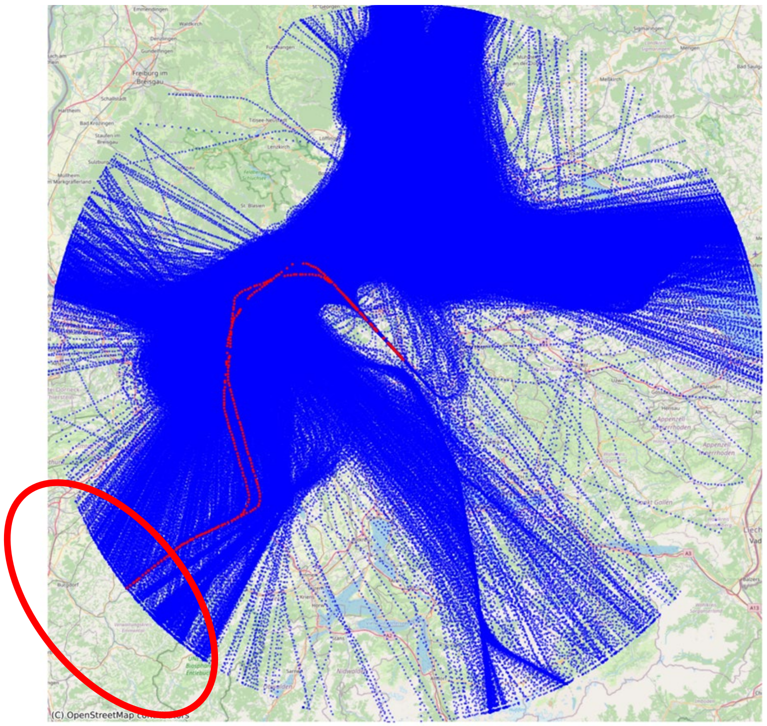
Distribution of Synthetic Tracks



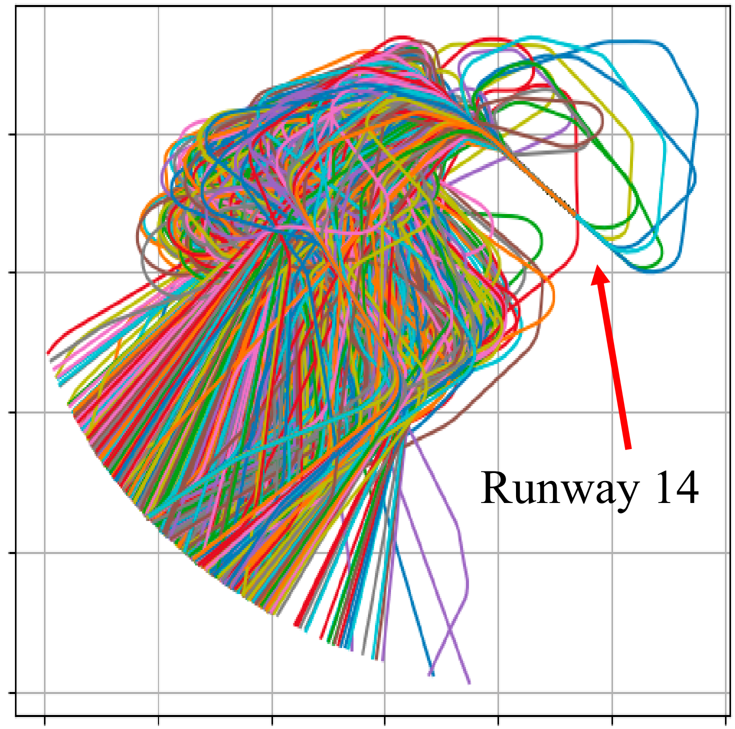


Distribution learned by VAE may be centered / left-of / right-of seed track

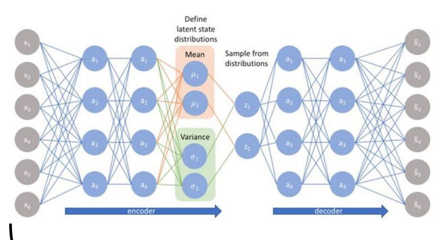
Training VAE on Flights from Southwest



2,052 flight tracks from southwest



Test ~200
Training ~1,800



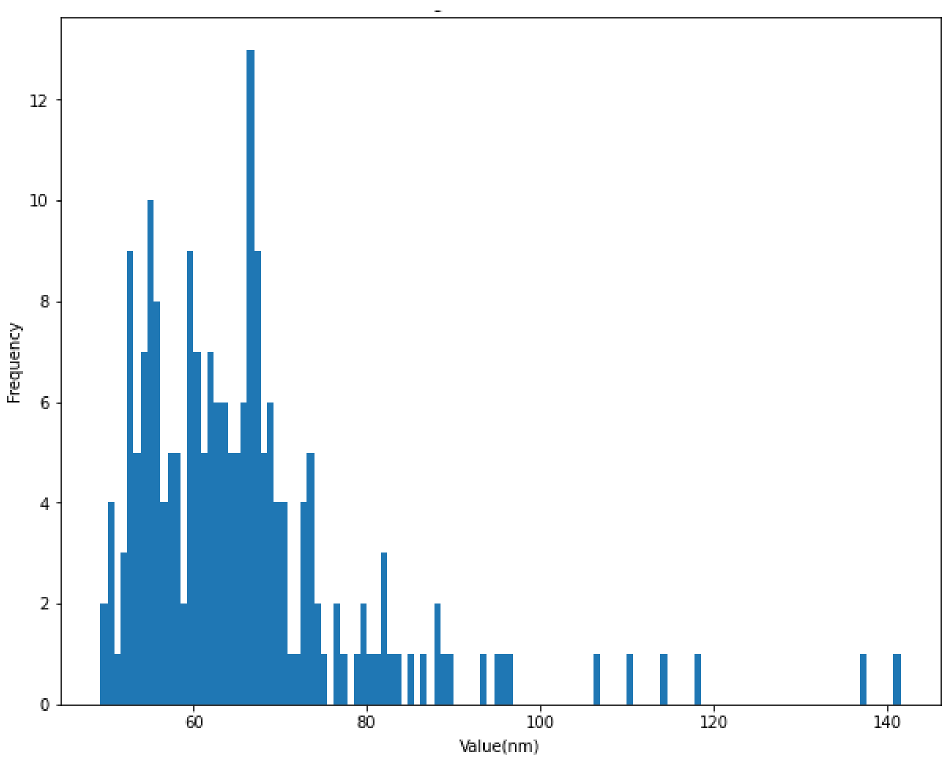
Training

Synthetic tracks

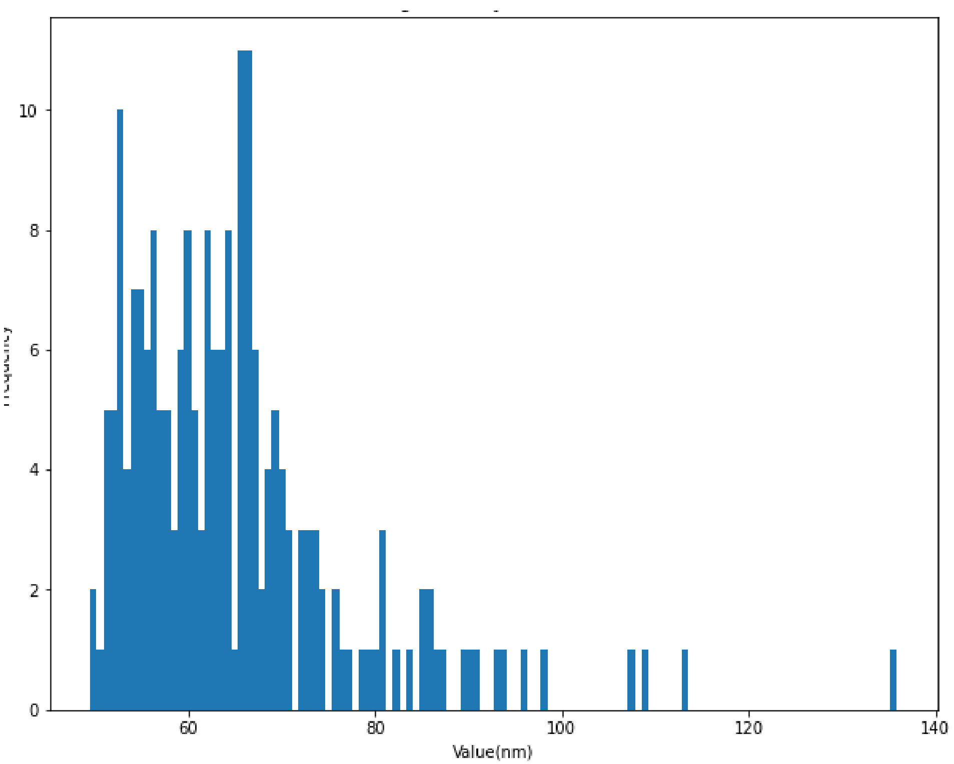
Historic tracks

Comparison of Along-track Distance

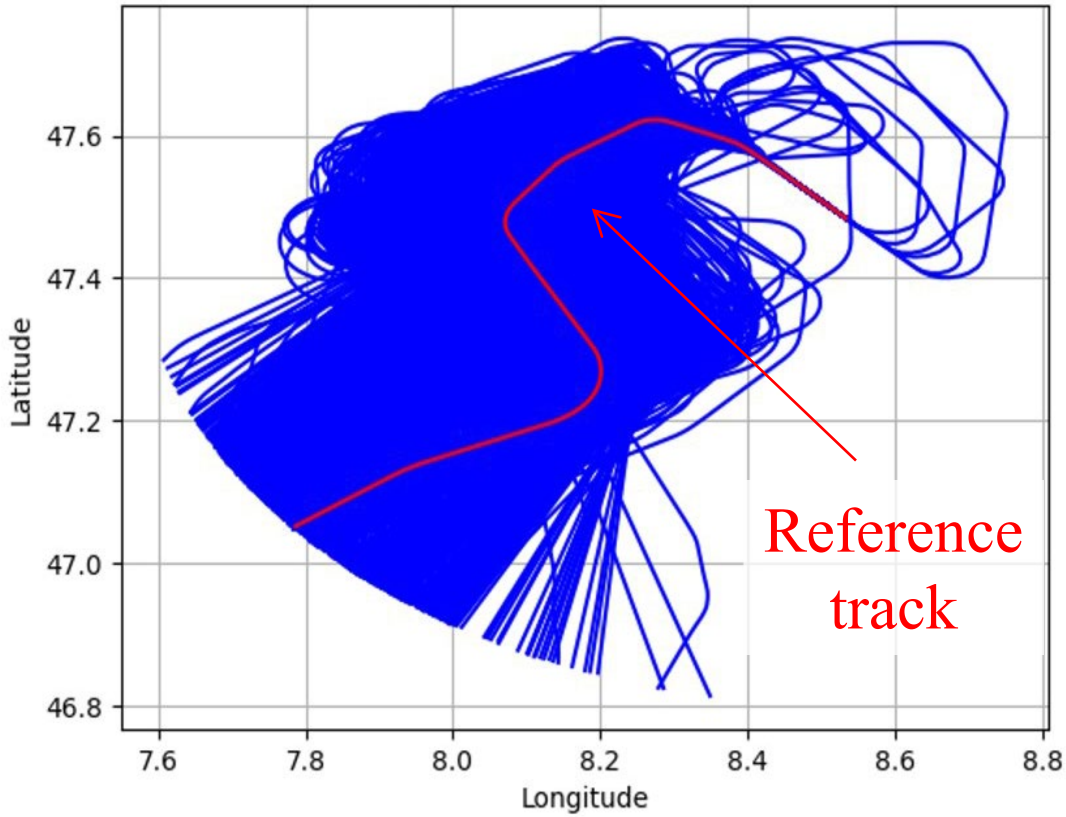
Test Tracks



Synthetic Tracks



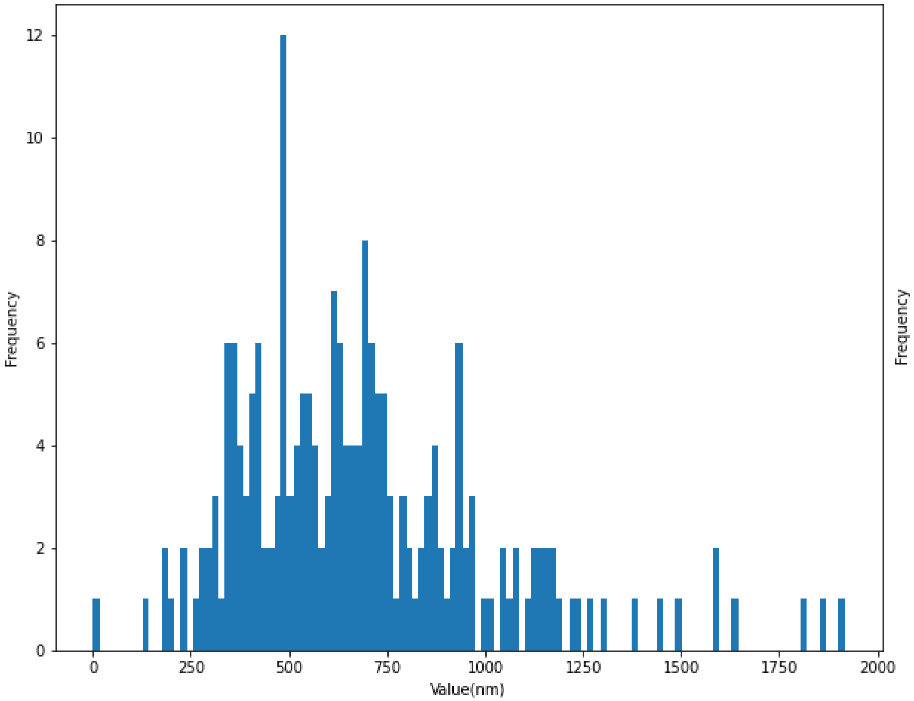
Cross-Track Distance



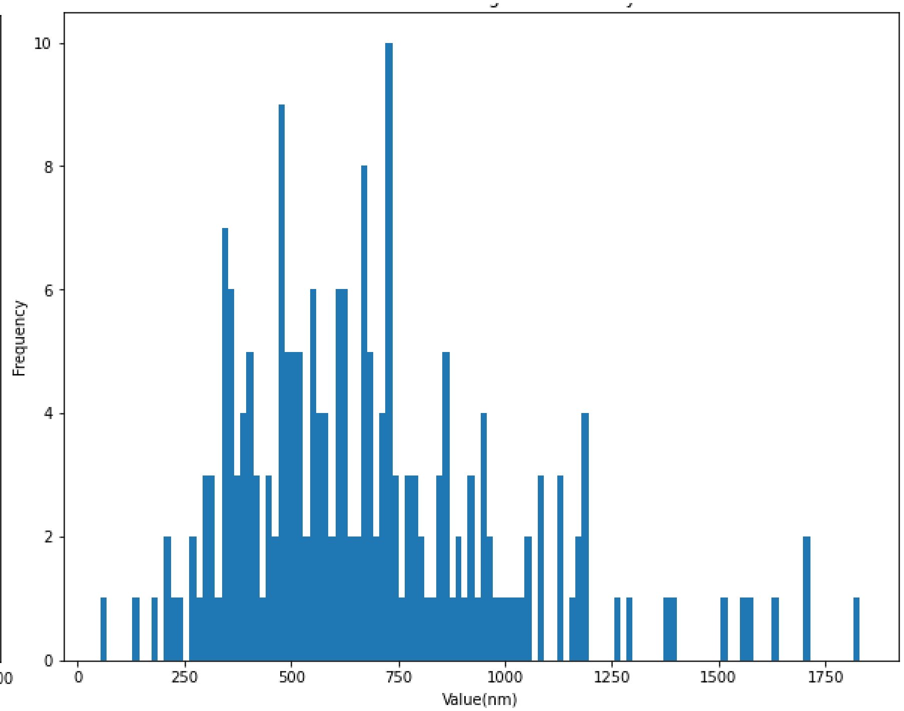
- Total cross-track distance = sum of distances (point by point) from a given track to reference track
- Reference track = track that minimizes distance to other tracks

Comparison of Total Cross Track Distance

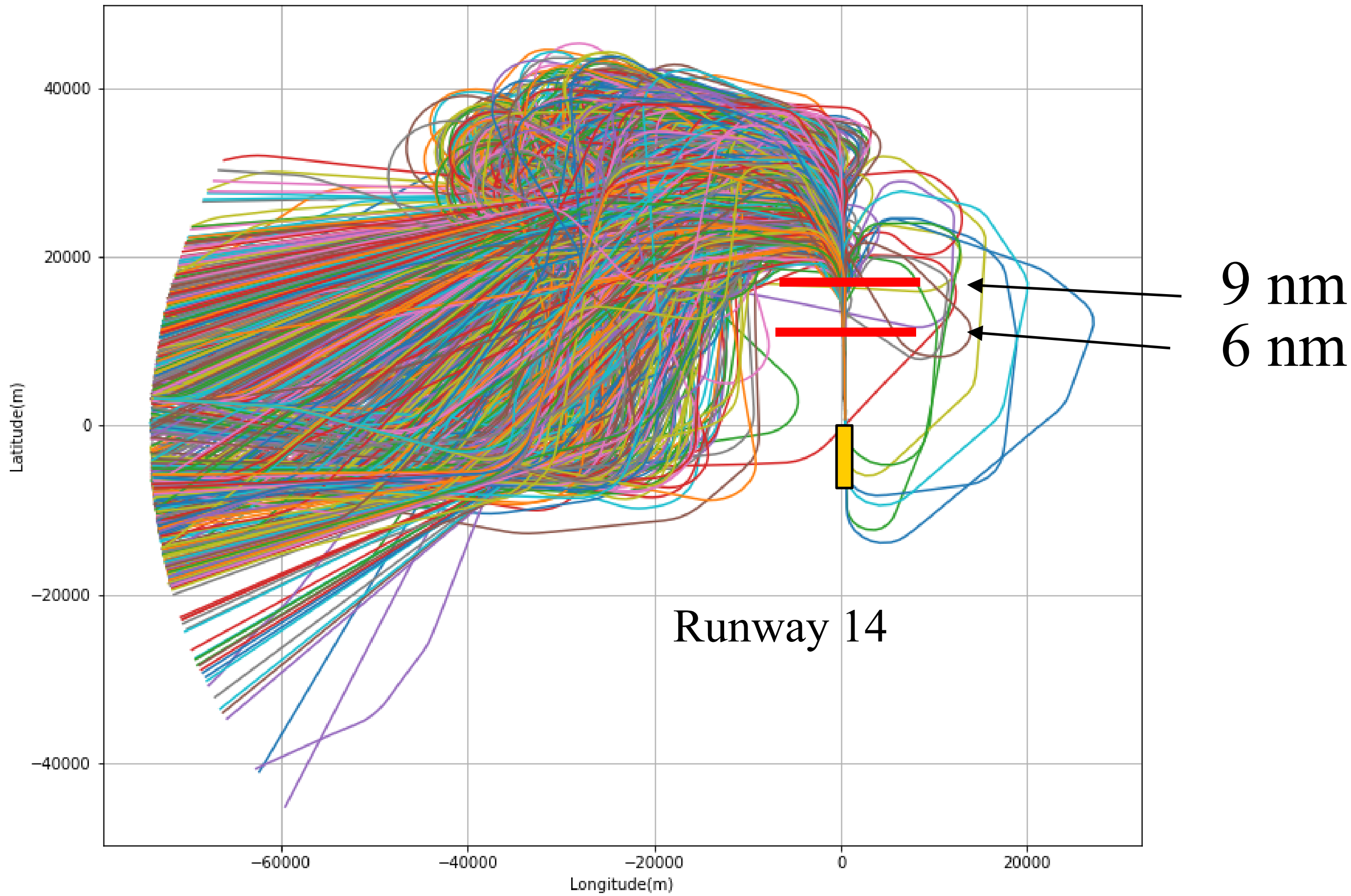
Distance between test tracks and reference track



Distance between synthetic tracks and reference track



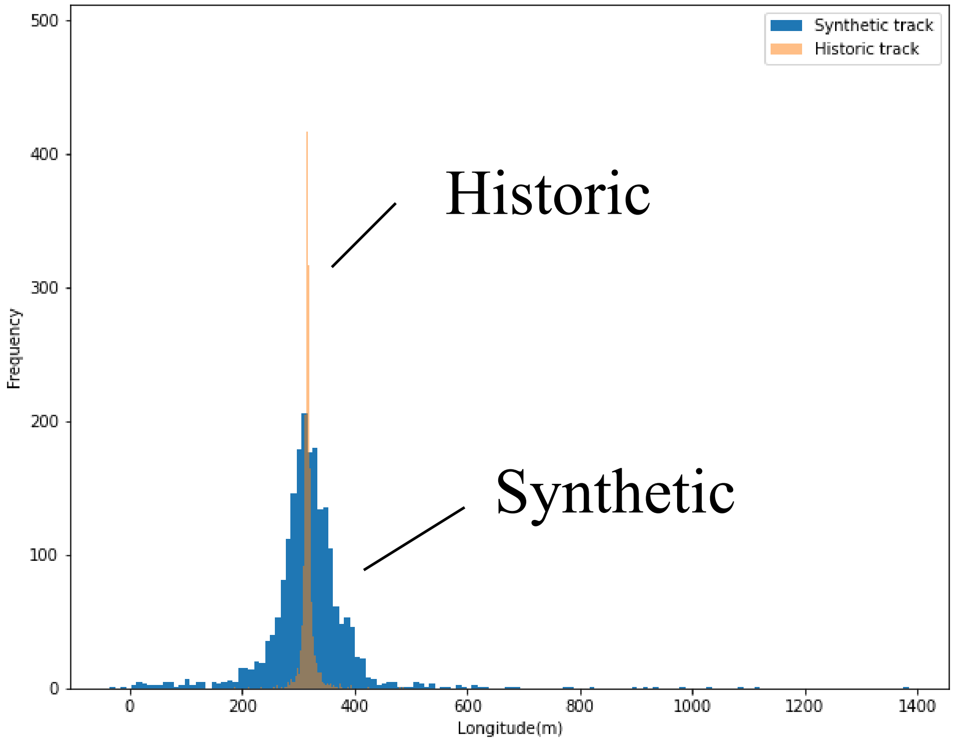
Lateral Dispersion



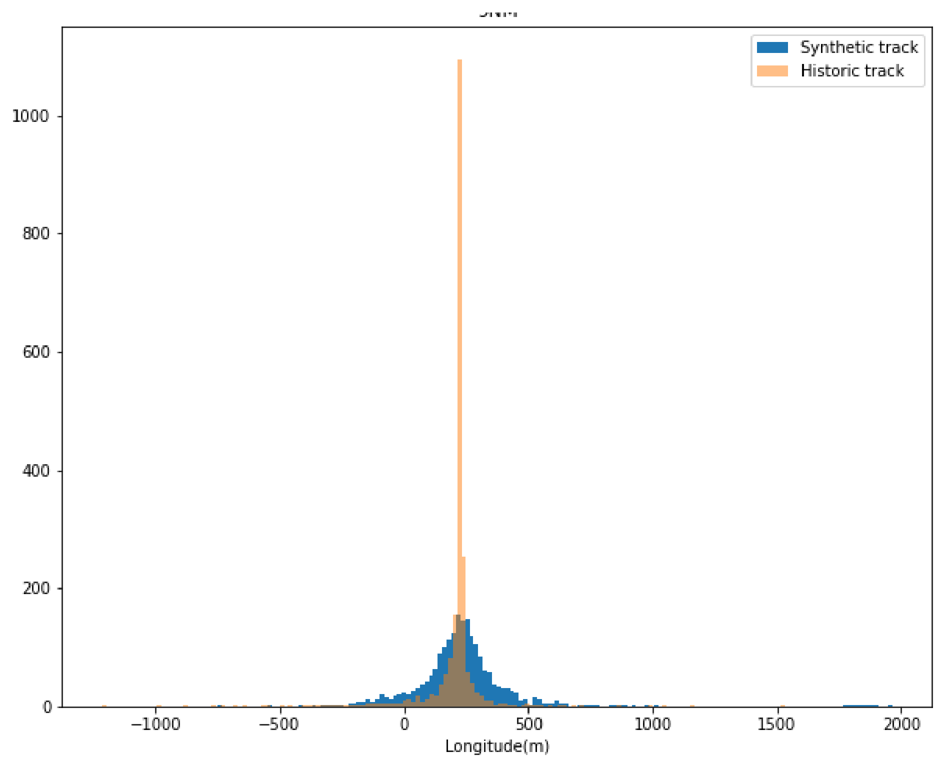
Rotated coordinate frame

Comparison of Lateral Dispersion

6 nm



9 nm



21

Summary

- Trained multiple types of autoencoders on Zurich arrival tracks
- Autoencoders can replicate complicated arrival patterns (holding patterns, etc.)
- Variational autoencoders can generate random variations from seed tracks
- Synthetic and historic tracks have similar distributions for some, but not all, metrics

Future Work and Challenges

- Comparison of collision risk metrics between synthetic and historical tracks
  - Do synthetic tracks have the same "tails" as the real system?
  - Can synthetic tracks effectively generate the edge cases (e.g., missed approaches)
  - Is it possible to quantify an effect track "multiplier"? (N historic tracks can be used to generate $K \times N$ synthetic track)
- Evaluation of track data at other airports