Generating Synthetic Flight Tracks for Collision Risk Analysis Using Artificial Intelligence: Case Study of Chicago O'Hare Runway 27C

Shahab Aref*, John Shortle, Lance Sherry
Center for Air Transportation System Research/Systems Engineering and Operations Research
George Mason University, Fairfax, Virginia, US, 22030
Email: saref2@gmu.edu

Total Number of Pages: 15

Submitted [Submission Date]
*Corresponding Author

ABSTRACT

Objectives: Collision Risk Analysis (CRA) is required before the deployment of new air traffic management procedures and navigation concepts. To achieve the acceptable confidence interval on the estimation of the rare-event (e.g. 1E-9), requires a very large number of representative flight trajectories. The required sample size is prohibitive through means of traditional simulation. Previous studies have demonstrated the feasibility of generating synthetic flight tracks using AI/ML algorithms such as Variational Autoencoders (VAEs). This paper describes the end-to-end software framework that operationalizes this workflow.

Methods: The CRA Tool suite described in this paper includes three complementary software tools: The Pre/Post Processing Tool (3PT) converts flight tracks back-and-forth between Geographic Coordinate System (GCS) and Cartesian Coordinate System (CCS). CCS is required for training the VAE. The Variational Autoencoder Training Tool (VTT) trains airport- and procedure-specific VAEs using configurable neural network architectures, training hyperparameters, and normalization statistics. The Synthetic Flight Track Generation Tool (SFTGT) generates synthetic flight tracks from trained VAEs and user-selected seed tracks, evaluates their realism using physics-based and statistical metrics, and exports the validated trajectories for collision risk analysis.

Findings: The three tools operate as an integrated workflow to transform raw ADS-B data into synthetic flight tracks. The synthetic flight tracks are aerodynamically realistic and reflect air traffic control procedures. On the order of a million flight tracks can be generated in an hour or two depending on the complexity of the navigation procedure. The tool suite is demonstrated on a case study for arrivals to Chicago O'Hare International Airport (ORD) Runway 27C.

Novelty: Operationalizing AI-based synthetic flight track generation through user-friendly software tools. The tool suite bridges the gap between research algorithms and their practical application in aviation safety studies.

Practical Applications: The CRA Tool Suite enables aviation safety analysts to efficiently generate and validate synthetic flight tracks for collision risk analysis of existing and proposed terminal-area procedures. The tool suite specifically provides a practical method for generating the very large sample sizes (e.g., on the order of 3×10^7 flight tracks or flight-track pairs, depending on the collision risk analysis) required for collision risk analysis.

INTRODUCTION

Collision Risk Analysis (CRA) for Terminal Area operations is a safety assessment methodology required prior to the deployment of new air traffic management procedures and navigation concepts (International Civil Aviation Organization 2010). It functions as a formal evaluation framework to quantify the probability of separation violations and potential mid-air conflicts within highly congested terminal airspace. By systematically analyzing flight trajectories, CRA provides aviation authorities and system designers with a mathematical assurance of safety, ensuring that operational modifications do not introduce unacceptable risks into the airspace.

The necessity of CRA stems from the requirement to validate safety against extremely rare catastrophic events, typically on the order of 1E-9 (International Civil Aviation Organization 2018). Achieving a statistically acceptable confidence interval when estimating such low-probability rare events requires an exceptionally large sample size of representative flight trajectories. Traditional simulation methodologies, such as (Lankford and Ladecky 2003) fall short, as the prohibitive sample sizes required to accurately model these rare encounters make conventional brute-force simulation computationally impractical.

Modern safety assessments approaches to CRA can bridge this gap, by ensuring that confidence intervals are significantly minimized through massive, statistically appropriate sample sizes of flight tracks (Aref et.al. 2024a, 2024b).

The paper describes a CRA tool suite to generate a large sample size of flight tracks. The process includes an integrated end-to-end framework starting with the Pre/Post Processing Tool (3PT) to transform and normalize raw Automatic Dependent Surveillance-Broadcast (ADS-B) telemetry data. These standardized datasets are then ingested by Variational Autoencoder Training Tool (VTT) to train airport- and procedure-specific generative models. Finally, Synthetic Flight Track Generation Tool (SFTGT) uses the trained models alongside seed tracks to produce millions of realistic trajectories, evaluating their validity via physics-based and statistical metrics to supply the dense data needed for rigorous collision risk quantification

This paper is organized as follows: The Methods section describes the CRA tool suite. The Case Study section describes the application of the CRA tool suite to generate a large sample size of flight tracks for Chicago O'Hare airport Runway 27C. The Conclusion section concludes with a discussion of the implications of this capability, limitations, and future work.

METHODS

For CRA, the three tools operate together as an integrated workflow to process raw ADS-B data, train a Variational Autoencoder (VAE), generate synthetic flight tracks, and validate the generated data to ensure their suitability for collision risk analysis (Figure 1). Starting on the left, 3PT preprocesses raw ADS-B data by performing the necessary processing and standardization required for VAE training. The processed dataset is then used by the VTT to train a VAE model. Once the VAE has been trained, the SFTGT uses the trained model together with seed flight tracks to generate synthetic flight tracks. In addition to generation, SFTGT provides visualization and evaluation capabilities to assess the realism of the synthetic tracks by computing a range of physics-based and statistical metrics.

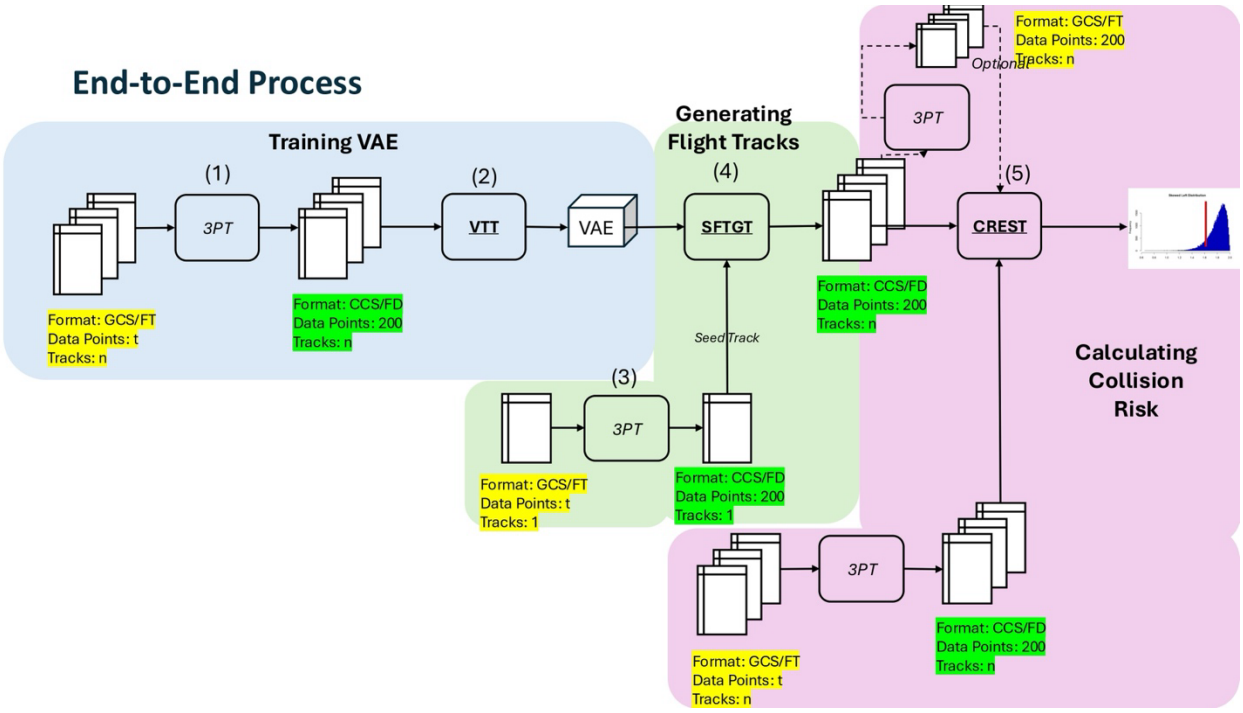


Figure 1. The end-to-end process of collision risk analysis

This section describes the end-to-end workflow for generating synthetic flight tracks, from processing raw ADS-B data to preparing datasets for collision risk analysis.

Step 1: Data preprocessing (3PT)

The first step is preprocessing the ADS-B data using the 3PT. The raw ADS-B data which is in GCS are uploaded into the tool, which groups the data based on the callsign, aircraft identifier, and other available attributes to identify individual flight tracks. Next, the longitude, latitude, altitude, and ground speed of each flight track are standardized using the method described in Section III, Step 1 of (Aref et al. 2024a). This process converts every flight track into a predefined number of equally spaced data points in CCS. For example, each trajectory may be represented by 200 fixed-distance (FD) data points. Consequently, each flight track is represented as a 200 × 4 matrix, where the four columns correspond to longitude, latitude, altitude, and ground speed. This standardization is necessary because the input layer of the VAE requires a fixed-size input.

Step 2: VAE Training (VTT)

The standardized flight tracks generated by 3PT are used by the VTT to train the VAE. Before training, the flight track data are normalized using the method described in Section III, Step 2 of (Aref et al. 2024a). The mean and standard deviation of each feature (longitude, latitude, altitude, and ground speed) are computed over the training dataset. Each feature is then normalized by subtracting its mean and dividing by its standard deviation. This normalization improves training stability and accelerates convergence.

The VTT allows users to configure various training hyperparameters, including the number of epochs, batch size, learning rate, latent dimension, and loss-function coefficients. The total training loss consists of three components: Reconstruction Loss, Kullback-Leibler (KL) Divergence Loss, and Monotonicity Loss, and is defined as

$$\text{Loss} = \alpha \times (\text{Reconstruction loss}) + \beta \times (\text{KL loss}) + \gamma \times (\text{Monotonicity loss}), \quad (1)$$

where α, β, and γ are user-defined weighting coefficients. Adjusting these coefficients influences the characteristics of the learned latent space and the variability of the generated synthetic flight tracks (Aref et al. 2026a). The output of VTT is a trained VAE model together with the normalization parameters, which are subsequently used by the SFTGT.

Step 3: Seed Flight Track Preparation (3PT)

To generate synthetic flight tracks, a seed flight track is repeatedly provided as input to the trained VAE. By sampling from the latent space, the VAE generates multiple variations of the seed flight track. Since the VAE requires a fixed-size input, the seed flight track must contain a fixed number of data points. Therefore, the seed flight track is uploaded to 3PT, where FD standardization uses interpolation to convert the trajectory into the required number of equally spaced data points for VAE input.

Step 4: Generate Synthetic Flight Tracks (SFTGT)

This tool is used to generate synthetic data and also evaluate their realism using different metrics. Synthetic flight track generation requires two inputs: (1) a trained VAE and (2) one or more seed flight tracks (Aref et.al. 2024a).

A seed flight track represents the desired trajectory from which synthetic variations are generated. The seed track may correspond to a proposed navigation procedure, a simulated trajectory, a flight simulator trajectory, or a historic flight track. The trained VAE is obtained from the previous step, while the seed track can be prepared using 3PT. After an ADS-B flight track is processed and converted into the standardized FD format, it can be used as a seed track. The trained VAE and the selected seed track are then loaded into SFTGT. After specifying the desired number of synthetic flight tracks, the tool repeatedly inputs the seed track into the trained VAE to generate multiple realistic variations while preserving the overall characteristics of the original flight trajectory.

In addition to synthetic flight track generation, SFTGT provides visualization and evaluation capabilities to assess the realism of the generated trajectories using a range of physics-based and statistical metrics before they are used for collision risk analysis. After synthetic flight tracks are generated, they must be evaluated to ensure that they are realistic for use in collision risk analysis. The SFTGT performs this evaluation by comparing the generated synthetic flight tracks with historic flight tracks at both the individual flight track level and the flight track set level.

For individual flight tracks, aerodynamic and operational characteristics, including bank angle, course change, rate of descent, level segments, and descending segments, are calculated for both the historic and synthetic flight tracks. These metrics are then compared to determine whether the synthetic tracks exhibit unrealistic behavior. Significant deviations or outlier values indicate that the generated tracks may not accurately represent realistic aircraft operations.

At the flight track set level, the tool evaluates the statistical properties of the generated trajectories by calculating the Lateral Dispersion (LD), Vertical Dispersion (VD), and Groundspeed Dispersion (SD) distributions, as described in (Aref et.al. 2024b). These distributions are compared with those of the historic flight tracks to determine whether the synthetic flight tracks preserve the natural variability observed in real operations. Large discrepancies between the two distributions suggest that the generated flight tracks do not adequately represent the operational characteristics of the real flight procedure.

To quantitatively compare the historic and synthetic flight track distributions, SFTGT employs two statistical metrics: the Weitzman Overlap Coefficient (WOC) (Montoya et al. 2019) and the Kullback-Leibler (KL) Divergence (Hershey and Olsen 2007).

The Weitzman Overlap Coefficient measures the degree of overlap between two probability density functions (PDFs) and is defined as

$$WOC = \int_{-\infty}^{+\infty} \min\big(p(x), q(x)\big)\, dx, \tag{2}$$

where $p(x)$ and $q(x)$ are the probability density functions of the historic and synthetic flight track distributions, respectively. The WOC ranges from 0 to 1, where 0 indicates no overlap and 1 indicates identical distributions. A larger WOC therefore indicates greater overlap between the historic and synthetic data.

The Kullback-Leibler (KL) Divergence quantifies the difference between two probability distributions and is defined as

$$D_{KL}(p||q) = \int p(x) log \frac{p(x)}{q(x)} dx. \tag{3}$$

Similar to the WOC, $p(x)$ and $q(x)$ represent the probability density functions of the historic and synthetic flight track distributions. The KL Divergence is always non-negative, with a value of 0 indicating that the two distributions are identical.

The WOC and KL Divergence are calculated for the LD, VD, and SD distributions. Together, these metrics provide a quantitative assessment of how closely the synthetic flight tracks match the statistical characteristics of the historic flight tracks.

Finally, the lateral and vertical characteristics of the synthetic flight tracks are evaluated using the physics-based realism metrics presented in (Aref et al. 2025). Specifically, the distributions of course change, bank angle, level segments, and descending segments are compared between the historic and synthetic flight tracks. Similarity between these distributions indicates that the generated trajectories satisfy the expected aerodynamic and operational characteristics, whereas significant discrepancies suggest that the synthetic flight tracks are unsuitable for collision risk analysis.

In some applications, the flight tracks used to train the VAE differ from the seed flight track used for synthetic flight track generation. As a result, the generated synthetic flight tracks may not fully capture the characteristics of the seed track. To better preserve these characteristics, the seed flight track is used to fine-tune the trained VAE before synthetic flight track generation. The fine-tuning methodology is described in detail in (Aref et al. 2026b). The impact of VAE coefficients in the fine-tuning process is briefly explained in table 1.

$$\text{Loss} = \alpha\times(\text{Reconstruction loss}) + \beta\times(\text{KL loss}) + \gamma\times(\text{Monotonicity loss}), \tag{4}$$

Table 1. VAE loss coefficients

Parameter	Explanation
α	Weights the reconstruction loss, controlling how closely synthetic tracks match real tracks
β	Weights the KL loss, controlling the diversity (spread) of the generated tracks
γ	Weights the monotonicity loss, penalizing unrealistic altitude increases

After fine-tuning, the updated VAE is used to generate a new set of synthetic flight tracks. These generated tracks are then evaluated using the realism assessment capabilities of SFTGT to determine

whether they satisfy the required physics-based and statistical realism criteria and are suitable for use in collision risk analysis.

After generating synthetic flight tracks and verifying their realism, the tracks must be converted into a format suitable for collision risk analysis. By default, the generated synthetic flight tracks are represented in FD format, in which each trajectory consists of 200 equally spaced data points. Consequently, the time interval between consecutive data points varies from one flight track to another.

For collision risk analysis, however, all flight tracks are better to be represented in fixed-time (FT) format, where the time interval between consecutive data points is constant (e.g., 5 seconds). To perform this conversion, the synthetic flight tracks are uploaded into the 3PT. Using the FT Standardization function, each flight track is converted from FD format to FT format through linear interpolation. The last data point of each flight track, corresponding to the point closest to the runway threshold, is assigned a timestamp of t=0. Linear interpolation is then applied while traversing backward along the trajectory to reconstruct the flight track at uniformly spaced time intervals.

CASE STUDY: CHICAGO O'HARE (ORD) RUNWAY 27C

A case study generating arrival flight tracks for Runway 27C at Chicago O'Hare International Airport (ORD) is used to demonstrate the capability of the tools. From start to finish it takes less than an hour to train the VAE and generate 1 million flight tracks. The VAE was trained using VTT. The flight tracks were generated from the trained VAE on standard laptop.

Historic ADS-B flight trajectory data were processed using 3PT to identify individual flight tracks, standardize the trajectories into fixed-distance format, and prepare the training dataset. The VTT is used to train a VAE, and finally, SFTGT is used to generate and validate synthetic flight tracks. The detail of these steps is presented in the following steps:

The standardized data were then used by the VTT to train a runway-specific VAE. After training, the SFTGT generated synthetic flight tracks from user-selected seed tracks and evaluated their realism using both physics-based metrics (e.g., bank angle, course change, rate of descent, level segments, and descending segments) and statistical metrics (e.g., lateral, vertical, and groundspeed dispersion, Weitzman Overlap Coefficient, and Kullback-Leibler divergence). Finally, the generated synthetic flight tracks were converted to fixed-time format for use in collision risk analysis. This case study demonstrates the complete end-to-end workflow of the proposed framework and shows that the three tools can successfully transform raw ADS-B data into validated synthetic flight tracks suitable for aviation safety studies and collision risk analysis.

Step 1: Data preprocessing (3PT)

In the first step, raw ADS-B data in GCS format are uploaded to 3PT. The dataset consists of 429 arrival flight tracks to ORD Runway 27C in 2026, as shown in figure 2.

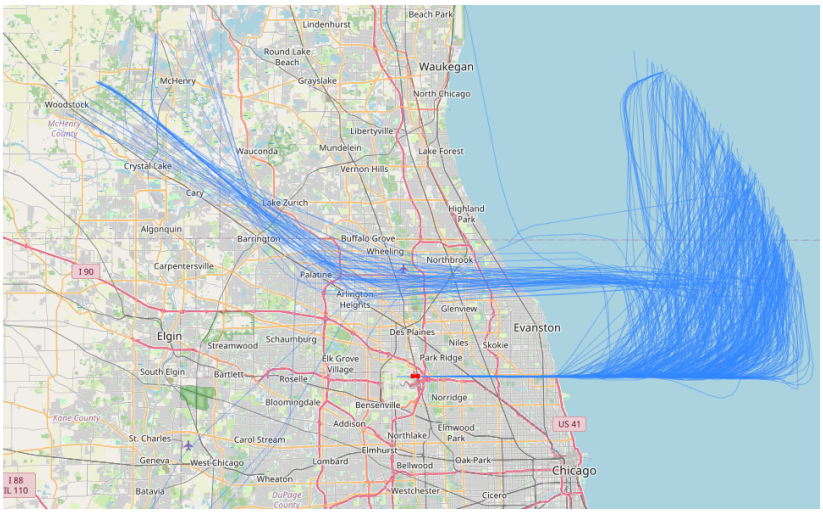
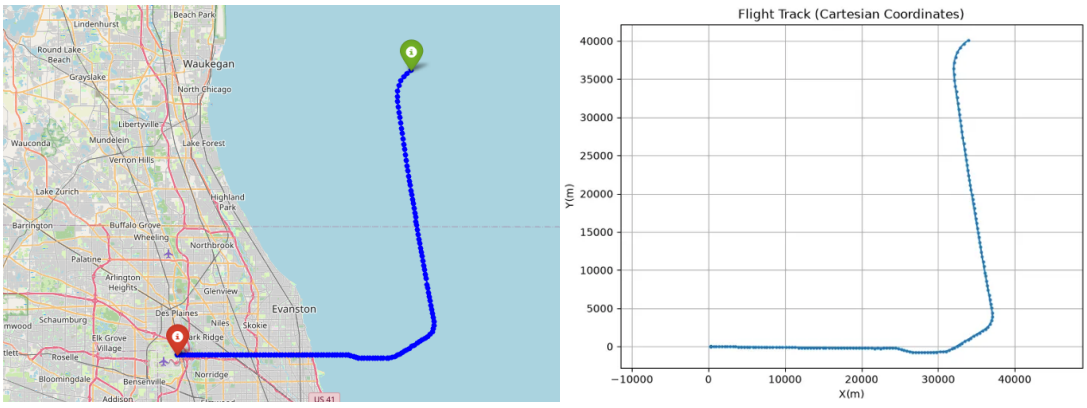


Figure 2. ORD 27C arrival flight tracks

The 3PT first groups the data using unique flight identifiers (e.g., callsign and aircraft identifier) to identify individual flight tracks. It then reports the total number of identified flight tracks and the corresponding number of data points for each track. To prepare the data for VAE training, 3PT converts the flight tracks from GCS to CCS format. To verify that the grouping and coordinate transformation have been performed correctly, the tool provides visualization of individual flight tracks. An example flight track in both GCS and CCS formats is shown in figure 3.



(a) (b)

Figure 3. Sample GCS and CCS flight track

After the coordinate transformation, all flight tracks are standardized into FD format by resampling each trajectory to a fixed number of equally spaced data points based on the traveled distance. In this case study, each flight track is standardized to 200 data points. Consequently, every flight track is represented by a 200 × 4 matrix, where the four columns correspond to longitude, latitude, altitude, and ground speed. Figure 4 illustrates an example of a flight track FD standardization, where the red markers represent the 200 equally spaced data points.

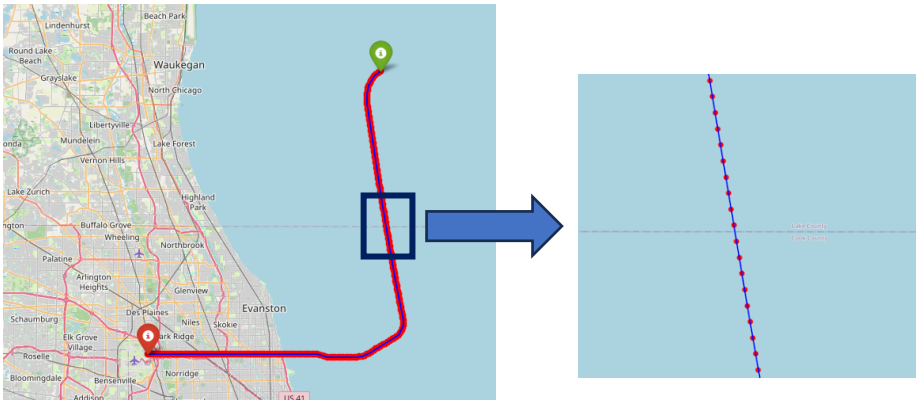


Figure 4. Fixed-distance standardized flight track

Step 2: VAE Training (VTT)

The standardized FD flight tracks generated by 3PT are uploaded to VTT to train a VAE for ORD Runway 27C. Before training, the flight tracks are normalized using the mean and standard deviation of the longitude, latitude, altitude, and ground speed computed from the training dataset. The normalization parameters are automatically stored together with the trained VAE and are later used during synthetic flight track generation.

VTT provides a graphical interface that enables users to configure the VAE architecture and training hyperparameters, including the number of epochs, batch size, learning rate, latent dimension, and the coefficients of the loss function. The total loss consists of three components which were presented under Method section.

The reconstruction loss measures the difference between the input flight track and the reconstructed output produced by the decoder. Minimizing this loss enables the VAE to accurately reproduce the spatial and operational characteristics of the training flight tracks. The KL divergence loss regularizes the latent space by pushing the learned latent distribution to remain close to a standard normal distribution. This regularization promotes a smooth latent space, allowing the VAE to generate diverse yet realistic synthetic flight tracks through latent-space sampling. The monotonicity loss encourages a physically realistic descent profile by penalizing unrealistic increases in altitude along the approach trajectory.

The number of training epochs determines how many times the complete training dataset is presented to the VAE. Increasing the number of epochs generally improves convergence but may lead to overfitting if training continues after the validation performance stabilizes. The batch size specifies the number of flight tracks used to update the model parameters during each optimization step. Smaller batch sizes introduce greater randomness into the optimization process and may improve generalization, whereas larger batch sizes typically provide more stable gradient estimates and faster computation. The learning rate controls the magnitude of the parameter updates during optimization. A learning rate that is too large may prevent convergence or cause unstable training, while a learning rate that is too small can significantly increase the training time and slow convergence. For the ORD Runway 27C case study, the VAE was trained for 2,000 epochs using a batch size of 8 and a learning rate of 0.001. The weighting coefficients for the reconstruction, KL divergence, and monotonicity loss terms were set to 1.5, 0.005, and 0.5, respectively. Encoder has linear layers with 800, 256, 128, and 64 neurons. Latent space includes 60 neurons.

During training, VTT displays the training progress, including the current epoch, loss values, and loss history. After training is complete, the trained VAE model and its associated normalization parameters are exported for use in the subsequent synthetic flight track generation step. The VAE loss during training in VTT for the ORD Runway 27C converge during the training with a decreasing trend (Figure 5).

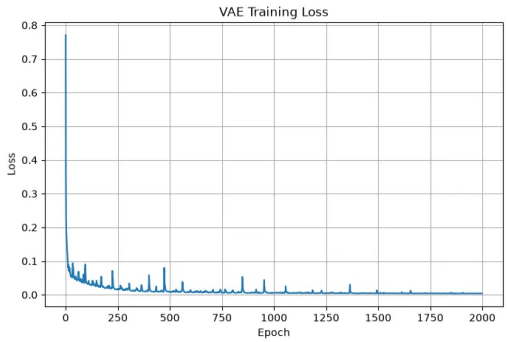
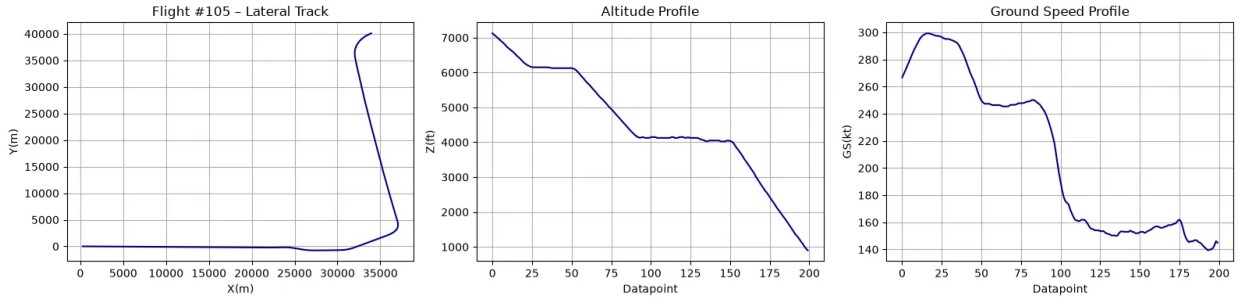


Figure 5. VAE training loss for the ORD Runway 27C case study

Step 3: Seed Flight Track Preparation (3PT)

To generate synthetic flight tracks for ORD Runway 27C, a seed flight track is required in addition to the trained VAE. The seed flight track is repeatedly provided as input to the VAE, which generates multiple variations by sampling from the latent space. The seed flight track may represent a proposed flight procedure, a computer-simulated flight track, a flight simulator trajectory, or a historic flight track. In this case study, a historic ADS-B flight track corresponding to an arrival on ORD Runway 27C was selected as the seed flight track (Figure 6). The track was uploaded to 3PT and standardized into FD format consisting of 200 equally spaced data points.

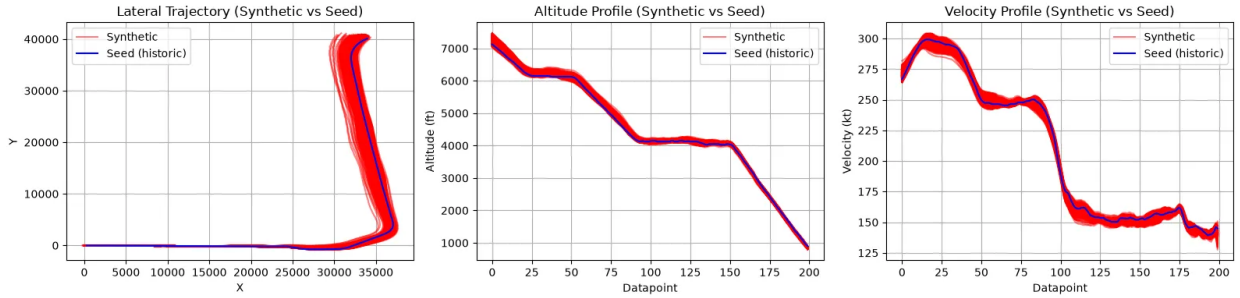


(a) **(b)** **(c)**

Figure 6. Seed flight track. (a) Lateral trajectory (b) Vertical profile (c) Groundspeed profile

Step 4: Generate Synthetic Flight Tracks (SFTGT)

After the VAE has been trained and the seed flight track has been prepared, both are uploaded to SFTGT to generate synthetic flight tracks. In this case study, 1,000 synthetic flight tracks were generated from the seed flight track. Figure 7 shows the lateral trajectories, vertical profiles, and ground speed profiles of the seed flight track and the generated synthetic flight tracks. As shown in the figure, the synthetic flight tracks successfully capture the primary spatial and operational characteristics of the seed flight track while exhibiting realistic variations.



(a) **(b)** **(c)**

Figure 7. Seed and 1000 synthetic flight tracks. (a) Lateral trajectory (b) Vertical profile (c) Groundspeed profile

In the next step, the LD, VD, and SD of the synthetic tracks and the seed track are calculated. Figure 8 shows LD, VD, and SD of 1,000 synthetic flight tracks, the seed flight track, and the corresponding expected distributions at 9 NM from the runway threshold. The expected distributions serve as references for evaluating whether the synthetic flight tracks exhibit realistic behavior and the anticipated variability. As shown in the figure, the synthetic flight tracks are centered around the seed flight track, and their distributions are generally consistent with the expected distributions, although they do not fully overlap. Two approaches can be used to increase variability and improve the overlap: fine-tuning the VAE using different loss coefficients or generating a portion of the synthetic tracks using another seed track, an approach referred to as multi-seed synthetic track generation.

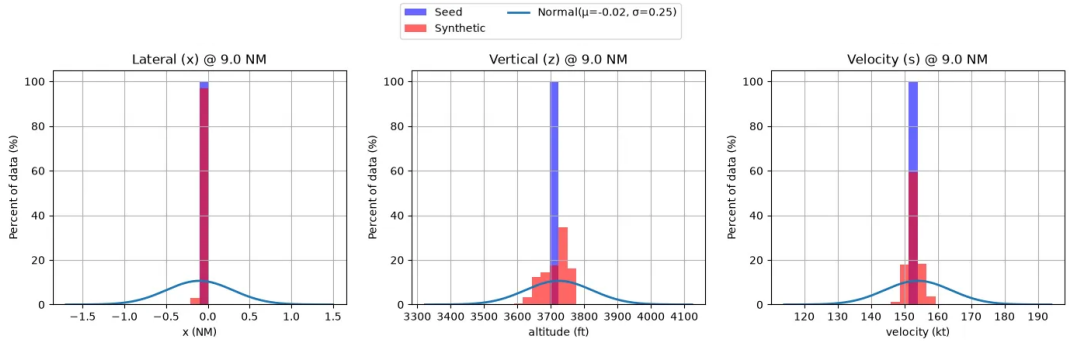


Figure 8. LD, VD, and SD of 1,000 synthetic flight tracks, the seed track, and expected tracks

Table 2 presents the WOC values between the distributions of the synthetic flight tracks and the corresponding expected distributions for LD, VD, and SD at multiple distances from the runway threshold. The expected distributions were represented by 3,000 randomly generated samples drawn from the corresponding normal distributions defined by the specified mean and standard deviation at each distance. The notation 3000/1000 indicates that the WOC was computed by comparing 3,000 samples from the expected distribution with 1,000 samples obtained from the generated synthetic flight tracks. The results show that the WOC between the synthetic and expected speed distributions (SD) ranges from 0.28 to 0.87 across the evaluated distances from the runway threshold. The VD also exhibits substantial agreement with the expected distribution, achieving a WOC of 0.76 at 4 NM. In contrast, the LD has relatively low WOC values because the generated synthetic flight tracks are concentrated near the runway centerline, whereas the expected LD is broader and includes aircraft located farther from the runway centerline. Consequently, the lower WOC values for LD primarily reflect the difference in the widths of the two distributions rather than unrealistic synthetic flight track behavior.

Table 2. WOC of LD, VD, and SD of synthetic tracks and expected distribution

WOC results

Distance (NM)	A	B	WOC lateral x	WOC vertical z	WOC velocity s	N A (x/z/s)	N B (x/z/s)
2	Normal	Synthetic	0.02	0.48	0.53	3000/3000/3000	1000/1000/1000
3	Normal	Synthetic	0.02	0.35	0.28	3000/3000/3000	1000/1000/1000
3.5	Normal	Synthetic	0.04	0.25	0.87	3000/3000/3000	1000/1000/1000
4	Normal	Synthetic	0.02	0.76	0.64	3000/3000/3000	1000/1000/1000
6	Normal	Synthetic	0.02	0.57	0.33	3000/3000/3000	1000/1000/1000
9	Normal	Synthetic	0.06	0.56	0.32	3000/3000/3000	1000/1000/1000

Table 3 presents the KL divergence between the expected and synthetic distributions for LD, VD, and SD at multiple distances from the runway threshold. Smaller KL divergence values indicate that the synthetic flight tracks more closely approximate the corresponding expected distributions. As shown in the table, SD at 3.5 NM has the lowest KL divergence (0.27) among all cases, indicating the closest similarity between the synthetic and expected distributions at this location. The VD also exhibits relatively low KL divergence values, particularly at 3.5 NM (0.52), indicating that the synthetic flight tracks accurately reproduce the expected vertical dispersion near the runway. In contrast, the LD has consistently larger KL divergence values. This is primarily because the generated synthetic flight tracks remain concentrated near the runway centerline, whereas the expected LD is intentionally broader and includes aircraft farther from the centerline. Consequently, the larger KL divergence values for LD indicates the difference in the spread of the two distributions rather than unrealistic synthetic flight track behavior.

Table 3. KL divergence between the synthetic tracks and expected distribution

KL results

Distance (NM)	p	q	N p (x/z/s)	N q (x/z/s)	KL(p||q) lateral x	KL(p||q) vertical z	KL(p||q) velocity s
2	Normal	Synthetic	3000/3000/3000	1000/1000/1000	26.65	1.58	3.44
3	Normal	Synthetic	3000/3000/3000	1000/1000/1000	26.8	4.75	7.13
3.5	Normal	Synthetic	3000/3000/3000	1000/1000/1000	25.94	5.62	0.27
4	Normal	Synthetic	3000/3000/3000	1000/1000/1000	26.88	0.59	2.73
6	Normal	Synthetic	3000/3000/3000	1000/1000/1000	26.71	3.9	10.18
9	Normal	Synthetic	3000/3000/3000	1000/1000/1000	24.63	4.18	10.07

Figure 9 shows the course change count and course change magnitude for the synthetic flight tracks and the seed flight track. A 10-degree threshold was used to identify course changes; therefore, the results are dependent on the selected threshold. As shown in the figure, most synthetic flight tracks exhibit one major course change corresponding to the turn onto the final approach. Instances with more than four detected course changes are primarily caused by variability in the flight segments preceding the final turn, where small trajectory deviations may exceed the detection threshold. Overall, the distribution of the course change count indicate that the synthetic flight tracks reproduce realistic turning behavior that is consistent with the seed flight track.

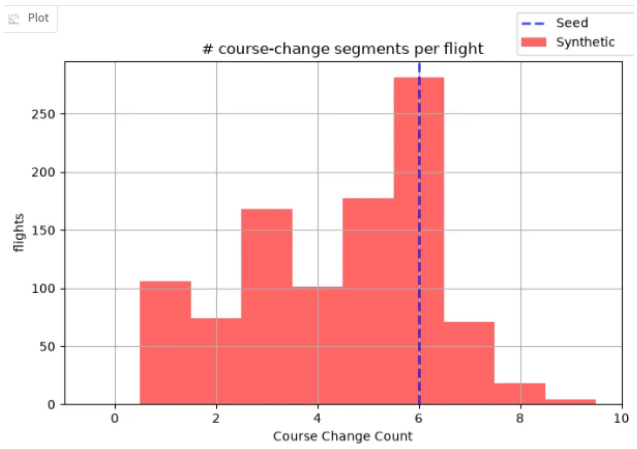


Figure 9. Course change count of the synthetic tracks and the seed track

Figure 10 shows the level segment count and the corresponding level segment distances for the synthetic flight tracks and the seed flight track. The seed flight track contains three level segments, whereas the synthetic flight tracks exhibit between zero and four level segments, indicating that the VAE is capable of generating a wide range of realistic descent profiles, from continuous descent operations to trajectories with multiple level segments. The level segment distances of the synthetic flight tracks range from approximately 2 NM to 6.5 NM, which is close to the distances in the seed flight track. These results indicate that the synthetic flight tracks reproduce realistic level flight during the approach.

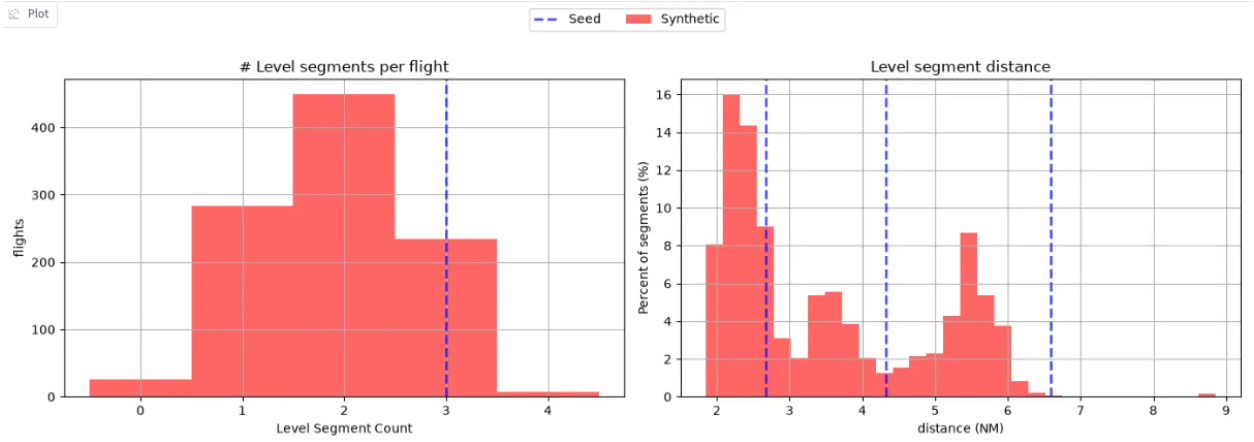


Figure 10. Level segment count and distance of the seed and synthetic tracks

Figure 11 shows the descending segment count and the corresponding descending segment distances for the seed flight track and the synthetic flight tracks. The seed flight track contains four descending segments, whereas the synthetic flight tracks exhibit between three and approximately ten descending segments. The additional descending segments are primarily caused by small variations in the generated altitude profiles that divide a long descent into multiple shorter descending segments. The histogram of descending segment distances shows that the synthetic flight tracks produce distributions centered around the distances observed in the seed flight track, particularly near 0 NM, 5 NM, and 9–10 NM from the runway threshold. These results indicate that the synthetic flight tracks reproduce realistic descent behavior while preserving the overall descent structure of the seed flight track.

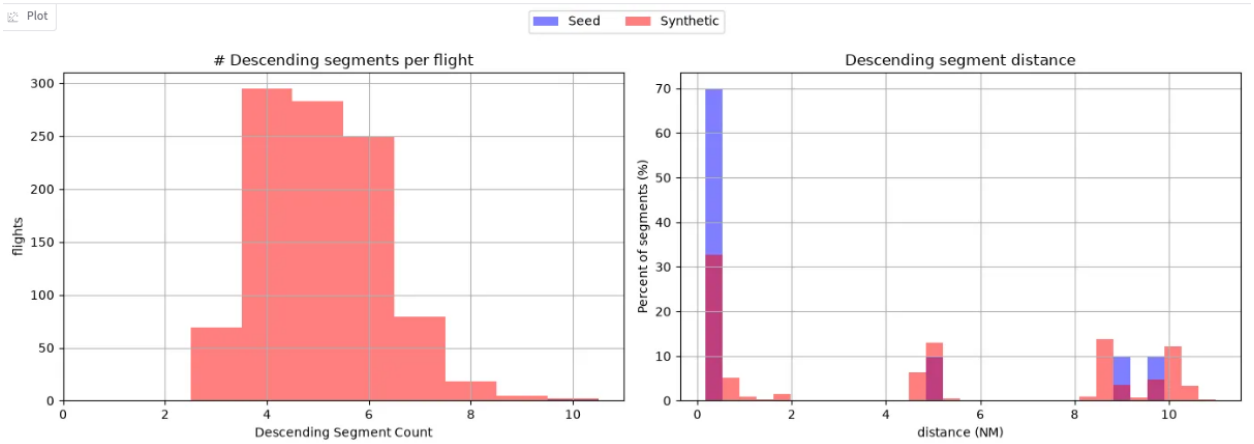


Figure 11. Descending segment of the seed and synthetic tracks

CONCLUSIONS

The Collision Risk Analysis (CRA) tool suite, enabled by AI, successfully bridges the gap between the need for a large sample sizes of flight track data for rare event analysis and the availability of these data sets. By systematically integrating data normalization, VAE training, and synthetic track generation, the framework provides a statistically rigorous alternative to computationally prohibitive brute-force simulations. This localized, data-driven approach yields the dense datasets required to model rare-event separation violations without sacrificing the physical or statistical integrity of actual operations.

Ultimately, the CRA tool suite offers aviation authorities and system designers a validated, scalable methodology to verify safety prior to deploying new air traffic procedures. By reducing uncertainty margins and minimizing confidence intervals through targeted synthetic trajectories, the suite serves as an essential framework for safety assurance.

The CRA tool suite's outputs remain highly sensitive to manual parameter tuning, specifically the arbitrary thresholds established for course changes along approach segments, which can artificially skew synthetic trajectory distributions and introduce systemic bias into the finalized risk quantification. Future operational validation remains dependent on precise parameter tuning, particularly regarding course change thresholds on approach segments.

ACKNOWLEDGMENTS

The authors acknowledge the technical contributions of Somil Shah, Rany Azzi, Sherri Shearon, Jiang Hong, Cody Nichols, Lacey Thompson, Luke Wood, Jacob Powers (FAA). Dr. Jie, Dr. Ran Ji, Dr. Chen, Ashim Thapa, Erfan Ashouri, Fahimeh Ghorbani, Andrew Kirk, and Jon West (CATSR/SEOR/GMU). This research was funded by the Federal Aviation Administration (FAA) under the NEXTOR IDIQ Contract No.: 693KA9-20-D00004, Delivery Order No. 693KA9-22-F00004

AUTHOR CONTRIBUTIONS

DECLARATION OF CONFLICTING INTERESTS

None

FUNDING

This research was funded by the Federal Aviation Administration (FAA) under the NEXTOR IDIQ Contract No.: 693KA9-20-D00004, Delivery Order No. 693KA9-22-F00004

REFERENCES

International Civil Aviation Organization (ICAO). 2010. *Manual on Implementation of a 300 m (1,000 ft) Vertical Separation Minimum Between FL 290 and FL 410 Inclusive*. 2nd ed. Doc 9574. Montreal: International Civil Aviation Organization.

International Civil Aviation Organization (ICAO). *Annex 11 to the Convention on International Civil Aviation: Air Traffic Services*, 15th ed. International Civil Aviation Organization, Montreal, Quebec, Canada, 2018.

Lankford, David N., and Shahar Ladecky. 2003. "Airspace Simulation and Analysis for TERPS (ASAT)." Presentation, Federal Aviation Administration, AFS-440, Oklahoma City, OK, November.

Aref, Shahab, John Shortle, and Lance Sherry. 2024a. "Generating Synthetic Flight Tracks for Collision Risk Safety Analysis: Variational Autoencoders with a Single Seed Track." 2024 Integrated Communications, Navigation and Surveillance Conference (ICNS), Herndon, VA, April 23–25. https://doi.org/10.1109/ICNS60906.2024.10550543

Aref, Shahab, John Shortle, and Lance Sherry. 2024b. "Using Variational Autoencoders to Generate 4D Synthetic Flight Tracks for Collision Risk Safety Analysis." Paper presented at the 2024 AIAA DATC/IEEE 43rd Digital Avionics Systems Conference (DASC), San Diego, CA, September 29–October 3. https://doi.org/10.1109/DASC62030.2024.10749485

Aref, Shahab, John Shortle, and Lance Sherry. 2026a. "Synthetic Flight Track Generation for Collision Risk Analysis: An Iterative Multi-Seed Approach." Paper presented at the 23rd Annual Conference on Systems Engineering Research (CSER), Arlington, VA, April 6–9.

Montoya, José A., Gudelia Figueroa P., and David González-Sánchez. 2019. "Statistical Inference for the Weitzman Overlapping Coefficient in a Family of Distributions." Applied Mathematical Modelling 71: 558–568. https://doi.org/10.1016/j.apm.2019.02.036

Hershey, John R., and Peder A. Olsen. 2007. "Approximating the Kullback–Leibler Divergence Between Gaussian Mixture Models." Paper presented at the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Honolulu, HI, April 15–20, IV-317–IV-320. https://doi.org/10.1109/ICASSP.2007.366913

Aref, Shahab, John Shortle, and Lance Sherry. 2025. "Physics-Based 'Realism Checks' for AI-Generated Synthetic Flight Tracks." Paper presented at the 2025 AIAA DATC/IEEE 44th Digital Avionics Systems Conference (DASC), Montreal, QC, Canada, September 14–18, 1–10. https://doi.org/10.1109/DASC66011.2025.11257394

Aref, Shahab, John Shortle, and Lance Sherry. 2026b. "Synthetic Flight Track Generation for Collision Risk Analysis: Los Angeles Runway 24R Downwind RNP." Paper presented at the 2026 Integrated Communications, Navigation and Surveillance Conference (ICNS), Herndon, VA, April 14–16. https://doi.org/10.1109/ICNS69853.2026.11570543

Ashouri, Erfan, and Lance Sherry. 2026. "High altitude icing condition identification: A lightweight deep segmentation framework for detecting aircraft-induced clouds in ground-based sky photography." In 2026 Integrated Communications, Navigation and Surveillance Conference (ICNS). IEEE.