Data Driven Modeling for the Simulation of Converging Runway Operations

Adric Eckstein
The MITRE Corporation
Center for Advanced Aviation Systems Development
Mclean, VA, USA
aeckstein@mitre.org

***Abstract*—A novel methodology is presented for generating data driven models for the general application of modeling and simulation. This approach relies on the use of principle component analysis to decompose a given data set into a basis of linearly uncorrelated modes. Data-driven models are then constructed from radar track data in order to develop models for a Monte Carlo simulation to evaluate the collision risk of converging runway operations.**

Keywords - data driven models, converging runway operations, principle component analysis, modeling and simulation

I. Introduction

Conventional modeling approaches are often derived from performance specifications in order to infer conclusions from modeling and simulation. While this approach is sufficient, it frequently requires supplementary data for validation and may rely heavily on assumptions or overly simplistic models.

This study introduces a novel modeling approach in which data-driven models are generated to capture the operational variation within a given data set. This approach utilizes principle component analysis to reduce a data set to a series of linear models. This modeling technique is used to drive Monte Carlo simulations of converging and intersecting runway operations. This simulation generates collision risk factors used to define the operational range for improved airport operations.

II. Converging Runway Operations

With increasing airport operations, new concepts are required in order to enhance throughput while maintaining a high level of safety. One such concept is reducing inter-operation time for dependent converging and intersecting runway operations. Specifically, this paper focuses on the arrival-departure converging runway operation. Fig. 1 shows an example of this operation at Chicago O'Hare International airport (ORD).

Conventional operations require "landing assured" for arrivals to 27R in order to release the departure from 32L, which limits the departure efficiency to that of arrivals on 27R. Alternatively, a "no-go box" defines a specified region prior to the runway threshold in which it is unsafe to release a departure. Provided there is no arriving aircraft in this region, the departures would be cleared, increasing the departure efficiency. The goal is to define the smallest size and location of the "no-go box" to ensure safe operations while allowing the maximum throughput.

In this instance, we want to measure the risk associated with releasing a departure, given an arriving aircraft at a specified distance from threshold. The risk of the arriving aircraft initiating a missed approach and successively colliding with the departing aircraft should be sufficiently small to meet safety requirements. For this study, a collision was defined only as a function of the lateral trajectory, independent of altitude separation. Some of the factors that influence this risk are as follows:

- *runway geometries and locations*
- *fleet mix*
- *missed approach rate*
- *missed approach initiation height*

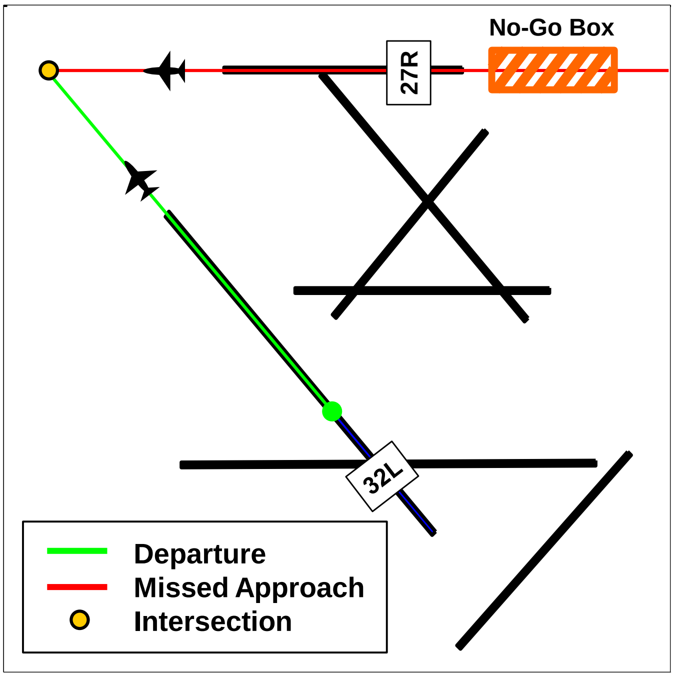

Figure 1. Converging runway operations at ORD.

- *departure clear-to-roll time*
- *aircraft trajectories*

A Monte Carlo approach has been adopted in order to compute the collision risk, where the airport operation is simulated using models which describe the variability in each component. Previous studies have adopted a similar methodology to assess the risk of these operations[1]. Simulations of aircraft pairs are allowed to evolve independently, such that additional risk mitigations factors such as controller intervention or TCAS alerting are not requirements for the safety of the operation.

In this paper, the modeling approach for the aircraft trajectories is described. Since the aircraft trajectories define the separation time when the arrival/departure aircraft tracks cross, this is an essential component to modeling the risk. Each of the other factors is treated separately to determine the overall risk of a given operation, and is beyond the scope of this paper.

A. Radar Track Data

In order to capture the operational performance of aircraft, historical high-update radar track data is examined. The departure tracks were grouped by aircraft type for over a year of track data at 14 airports to generate enough statistics for a wide variety of aircraft types.

The high-update radar source data utilized reports aircraft position at 1 second intervals. The position and time of these track reports were passed through a series of least squares filters[2] from which the ground speed was estimated. Next, each departure was normalized to its throttle-up time, estimated for each speed profile. Each track was then reduced to the first 180 seconds of flight. The ground speed data set for each aircraft type then consists of 180 ground speed measurements for ***N*** departures. Fig. 2 shows a two-dimensional probability density function (pdf) plot of this data for the set of Airbus A319 aircraft. Each slice through the x-axis would represent a histogram of the track speeds at a given time. The mean as well as the 95% upper and lower confidence bounds on the speeds at each time is also shown.

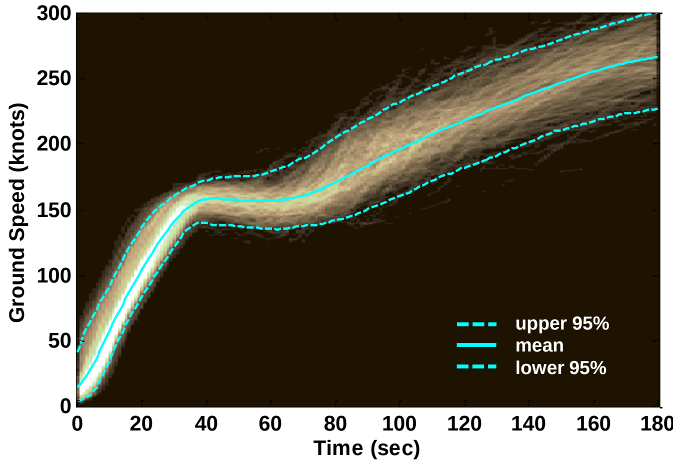

Figure 2. Two-dimensional pdf of A319 departure ground speed trajectory from track data.

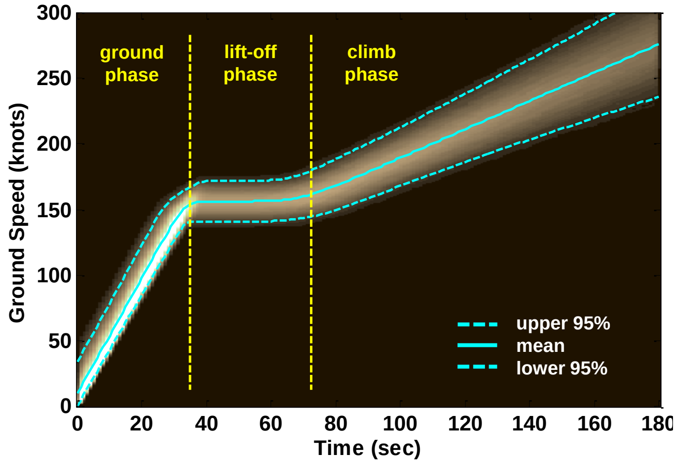

Figure 3. Two dimensinoal pdf of A319 constant acceleration trajectory model.

III. Data Driven Models

A. Conventional Approaches

Trajectory modeling, in this context, relies upon defining a speed profile for a given aircraft type. Using the runway location and heading, the speed profile can then be integrated in order to define a lateral trajectory for modeling and simulation.

Numerous trajectory models exist for various aircraft types. These models often rely on engine performance specifications or simplistic empirical models. However, such models often lack the operational variance in performance that would be seen in day-to-day operations. Variability may be introduced in the parameters which define such models, but are often a poor match when validated against data.

Alternatively, one can define the model directly from the given data. Conventional approaches to data-driven models might simplify the profile from Fig. 2 into simple linear segments with constant acceleration. Such an example is shown in Fig. 3. The speed profile consists of 3 segments. The first is a high acceleration ground roll phase, ending at the rotation point; followed by a constant speed take-off segment (during which gear and flaps are retracted); ending with a climb acceleration segment. Each segment is a simple linear model with coefficients drawn from known infinite distributions, based on fits to the given data.

While the bulk features of the speed profile are captured, there are apparent features in the actual speed profiles which indicate non-constant accelerations through these segments. In addition, not all speed profiles demonstrated this basic 3-segment behavior, as shown by other aircraft types in Fig. 4. There are clearly different signatures in each speed profile as well as distinct differences in the variability, which cannot be captured from such a simple linear model.

B. Principle Component Analysis

An attractive alternative for data-driven models is the use of Principle Component Analysis (PCA). Previous studies have utilized PCA as part of data driven modeling to cluster data groups[3] or deriving filter coefficients[4].

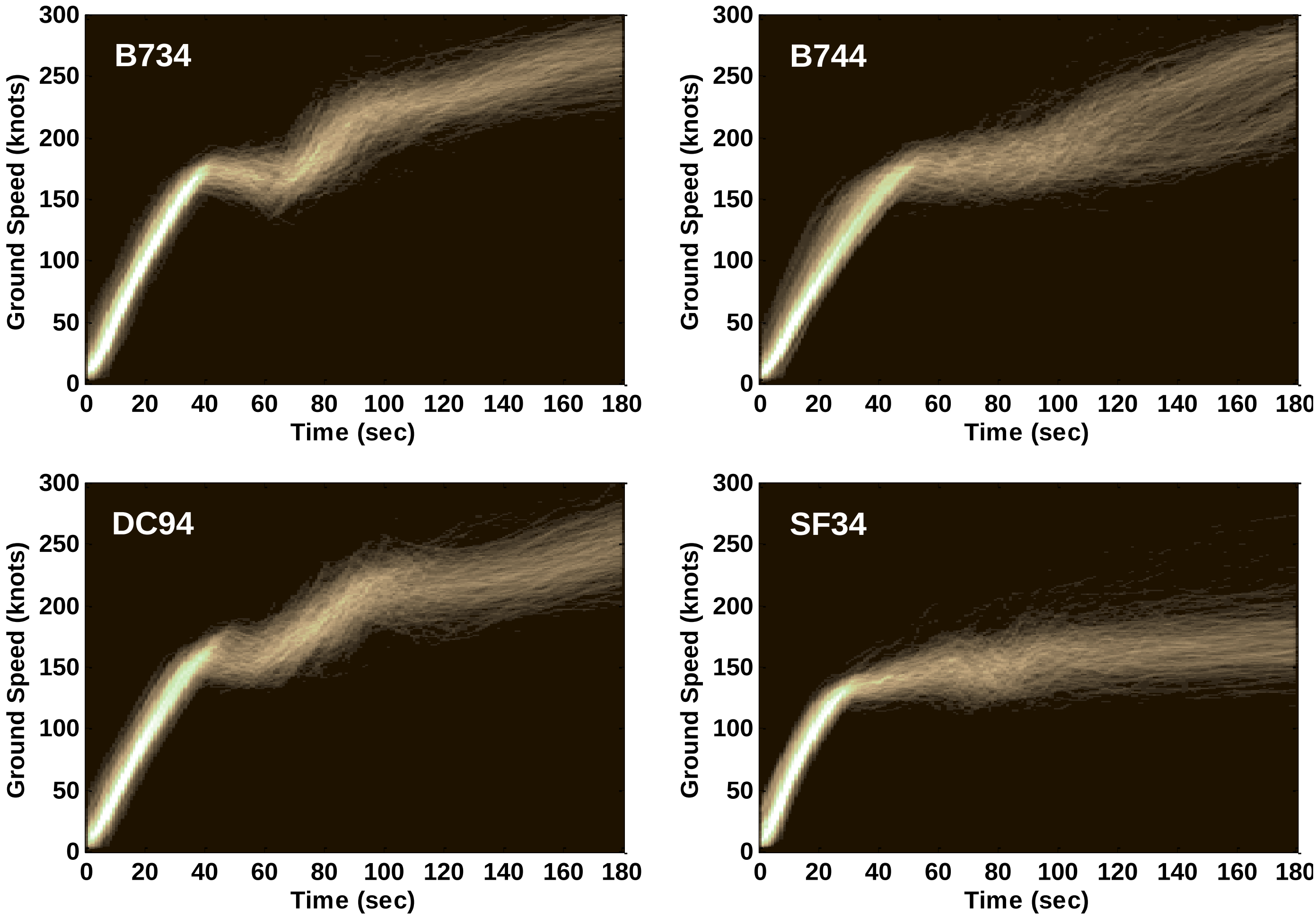

Figure 4. Two-dimensional pdf of departure ground speed trajectory for various aircraft types from radar track data.

PCA is a mathematical decomposition of a given data set through the use of eigenvalue decomposition[5]. Through PCA, a data set is decomposed into a set of uncorrelated principle components, which define a special basis set for the given data. PCA can also be described through the Proper Orthogonal Decomposition[6] (POD) of a given data set:

$$\mathbf{X} = \left[\vec{x}_1 \quad \vec{x}_2 \quad \dots \quad \vec{x}_n\right] = \mathbf{CM} = \left[\vec{c}_1 \quad \vec{c}_2 \quad \dots \quad \vec{c}_k\right]\left[\vec{m}_1 \quad \vec{m}_2 \quad \dots \quad \vec{m}_k\right]^T \tag{1}$$

The data set, $\boldsymbol{X}$, is defined by a series of random variables, $\boldsymbol{x_i}$. In the context of this study, each random variable is the ground speed at a given time (giving 180 random variables). The principle components (modes), $\boldsymbol{m_i}$, are defined by the eigenvectors of the covariance matrix, $\boldsymbol{X^T X}$. Similarly, the variance of each coefficient, $\boldsymbol{c_i}$, is the eigenvalue of the covariance matrix. The distributions of these coefficients will be critical to the definition of the model.

The modes and coefficients represent a linear decomposition of the data set similar to a Fourier decomposition. However, each principle component in the PCA basis accounts for the maximum variability in the data. As such, the PCA basis will have less error than any other basis set when reconstructing using a subset of the components. Another essential property of the PCA basis is that each principle component is linearly uncorrelated with all other components.

C. Continuous Univariate Coefficient Distributions

For this example, the ground speed data set of the Airbus

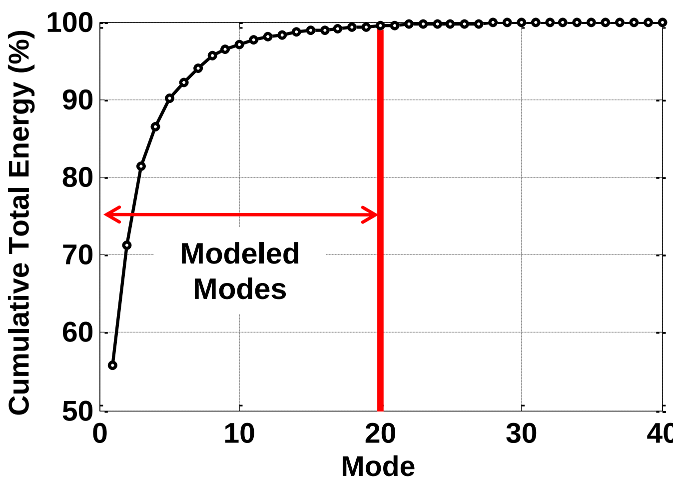

Figure 5. Cumulative total energy in the A319 PCA decomposition.

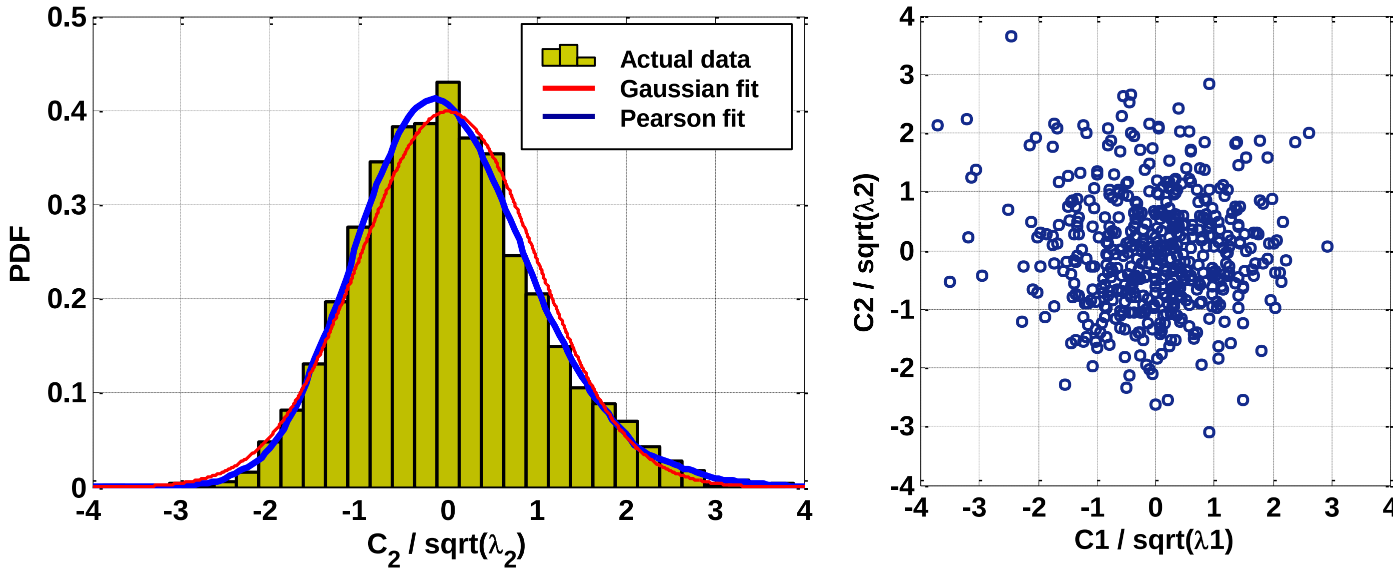

Figure 6. Continuous distribution fits to the second mode's coefficients (left) and scatter plot of first and second mode's coefficients (right).

A320 aircraft is examined. PCA is used to decompose the set into its principle components (the mean mode is subtracted off prior to decomposition to center the coefficients). As PCA represents an optimal decomposition with respect to the variance of the data, it serves as a perfect tool to reduce the order of the data. Fig. 5 shows that by retaining the first 20 modes, the model is able to account for 99.5% of the variation in the data. While more or fewer modes could easily be retained, 20 modes are hereafter modeled for demonstration of the methodology.

Next, a model for the distribution of each mode's coefficients is developed. Fig. 6 (left) gives a histogram of the first mode's coefficient. Two continuous distributions are shown overlaid with the data, a Gaussian distribution and a Pearson Distribution[7] (defined by mean, standard deviation, skew, and kurtosis). Relating this distribution back to the eigenvalue decomposition of the covariance matrix, the standard deviation of the Gaussian is equivalent to the square root of the eigenvalue. While the Gaussian is a good fit to the data, the Pearson fit better accounts for the skew in the data. This process is applied to each of the modes coefficients, giving an infinite distribution for each mode. Dominant outliers were removed prior to fitting the infinite distribution in order to better approximate the data.

Fig. 6 (right) gives a plot of the second mode's coefficient against the first mode's coefficient (normalized to the square root of their eigenvalues). It is apparent by the random scatter of the data that the distributions are largely uncorrelated. This is again a result of the PCA, where each mode is linearly uncorrelated with other modes. This property enables sampling from each of the Pearson distributions independently, which can then be reconstructed into the speed trajectory by summing along each of the principle components. A sample trajectory is then given by:

$$\vec{v} = \sum_{i=1}^{20} s_i \cdot \vec{m}_i \qquad (2)$$

where s_i is a random sample of the Pearson distribution from the i^{th} mode. The sample trajectory, v, is then used as an input to the Monte Carlo simulation. With any other basis decomposition, multivariate sampling would be required, which is much more complex and can lead to several additional sources of error.

In practice, bounds were placed on each coefficient and the variance of the coefficients in order to prevent unrealistic trajectories caused from sampling from infinite distributions. These bounds were determined from the sample size and the extrema observed in the actual data.

Fig. 7 shows the two-dimensional pdf of the PCA model for the Airbus A319 speed profile. Since the PCA basis is defined from the data set, the model is able to capture all of the effects of the input data set (shown in Fig. 2), without any *a*

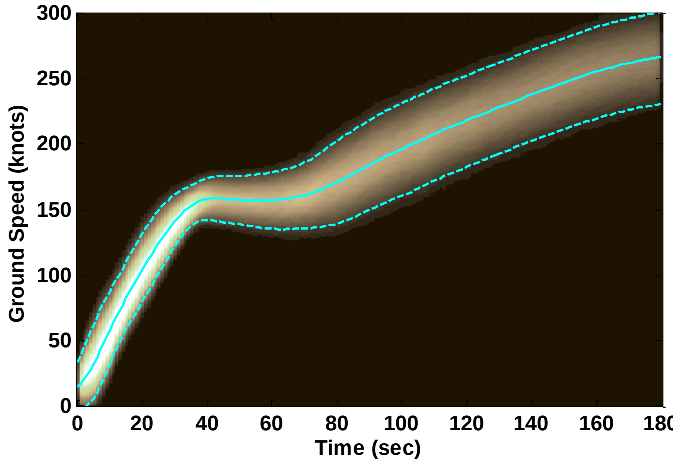

Figure 7. Two-dimensional pdf of A319 PCA decomposition trajectory model.

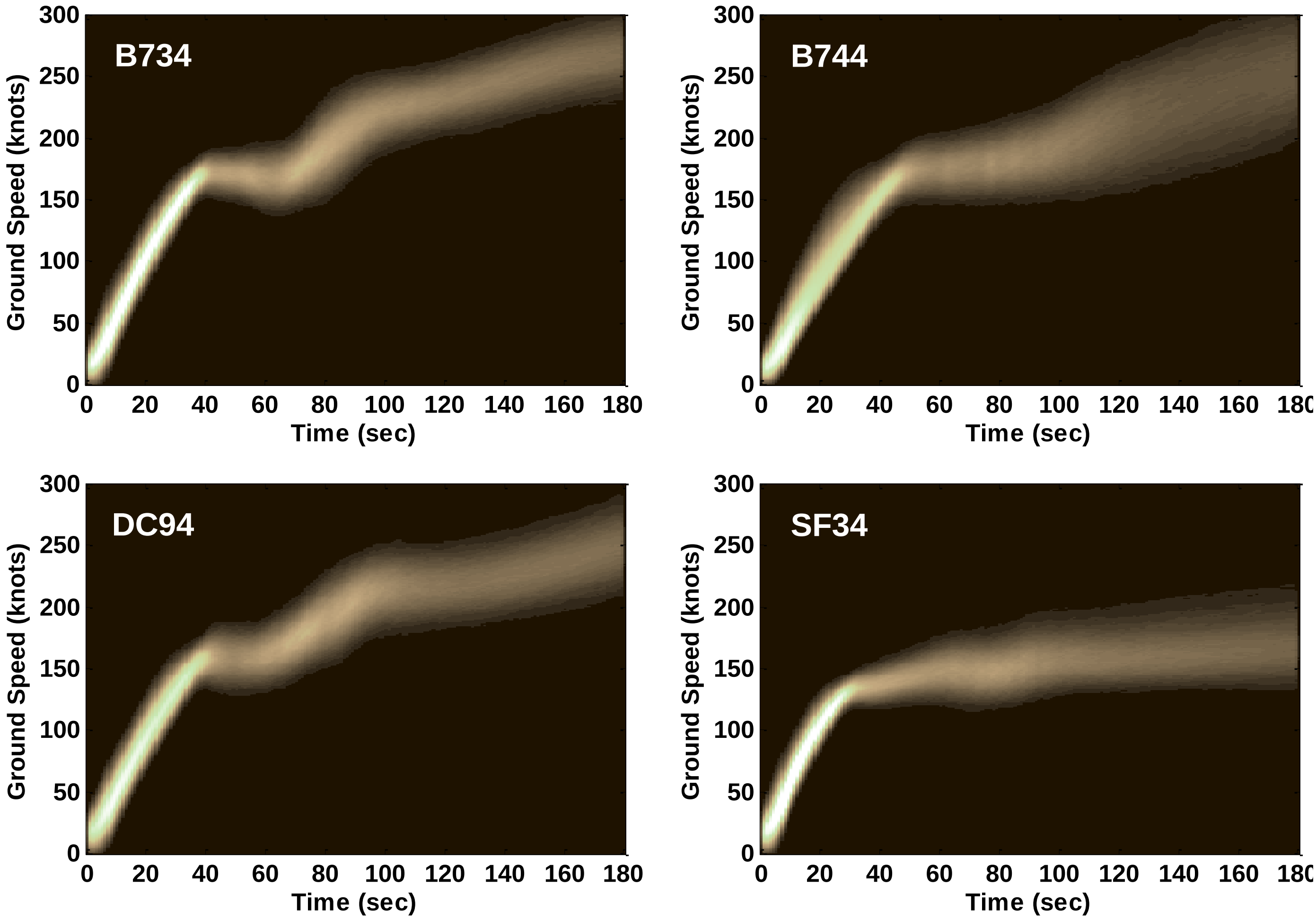

Figure 9. Two-dimensional pdf of PCA decomposition model for various aircraft types.

priori assumptions on the form of the model. This feature sets the PCA modeling approach apart from other conventional models. This model is also able to directly capture the correlation between speeds at different times, since each speed profile is a linear combination of the principle components of the original data set.

Because this approach does not rely on any *a priori* inputs in order to construct the model, we can follow the same methodology for other aircraft types. The models for each of the aircraft types from Fig. 4 are shown in Fig. 8. Each of the models is able to capture the distinct signatures and variability from its corresponding data set.

D. Nonlinear Correlation Errors

One of the key assumptions in this modeling approach is the independence of each principle component from the others. In terms of the model, this is based on the assumption that the underlying variation in trajectories is caused by independent random variations. Each mode, by definition, is linearly uncorrelated with all other modes. However, this does define independence, as higher order correlations may exist in the data or multivariate groupings of tracks.

Fig. 9 plots a two-dimensional pdf of the ground speeds for 500 McDonnell Douglas DC10 aircraft. It is apparent that there is a largely bimodal variation in the data during the climb phase of the departure. This is likely caused by variations in airspeed restrictions, region, or aliasing of aircraft types. Regardless, we are attempting to define a generic DC10

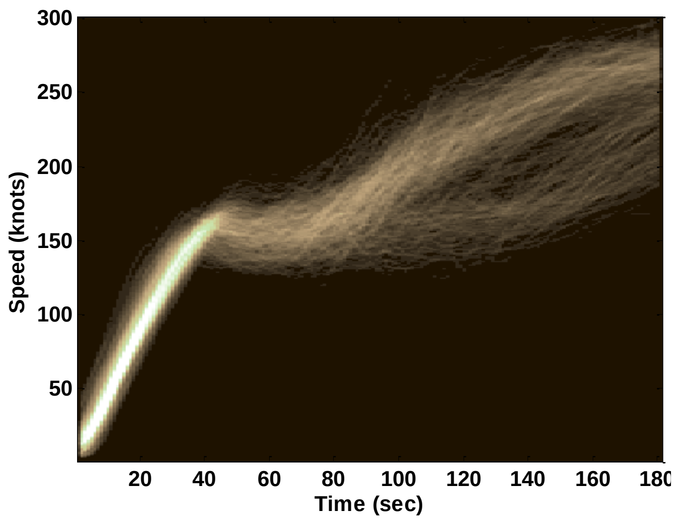

Figure 8. Two-dimensional pdf of DC10 departure ground speed trajectory from track data.

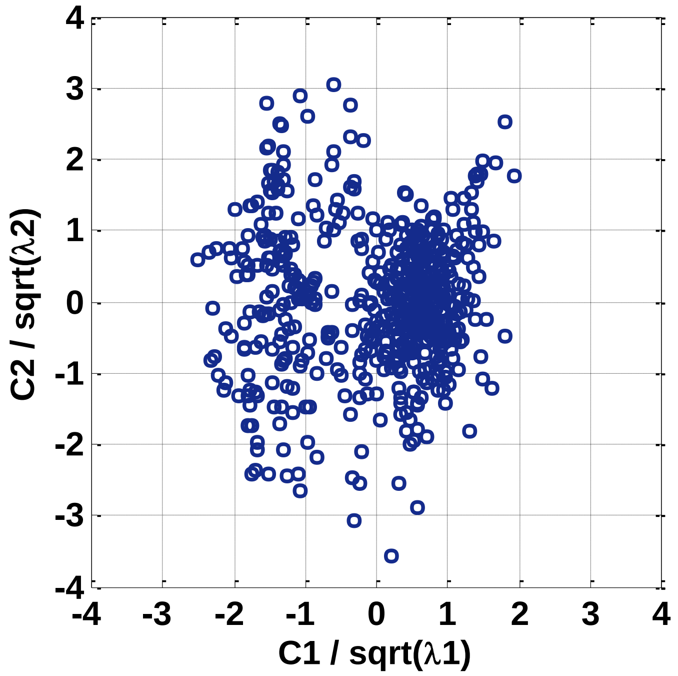

Figure 10. Scatter plot of DC10 first and second mode's coefficients.

trajectory model based upon the full data.

Following the methodology described above, the PCA decomposition is applied to the DC10 tracks. Fig. 10 gives a plot of the first two coefficients, normalized to the square root of their respective eigenvalues (similar to the plot in Fig. 6, right). Although the two coefficients are linearly uncorrelated, there is clearly a higher-order correlation between the two. This correlation is directly related to the bimodal grouping apparent from Fig. 9. Fig. 11 (left) shows the end result when the model is applied to the bimodal distribution of tracks. The variability between the two groups is smeared out across the range of speeds, producing an inaccurate model. While a dominant source of this error is the approximation of the C1 coefficient by a unimodal distribution, even a perfect match to the bimodal pdf would still produce an erroneous model, since the principle components contain higher order correlations.

Alternatively, we can start by dividing the set of DC10 trajectories by clustering the first couple modes' coefficients, from Fig. 10. Each of the groups is then independently modeled, creating two models for the DC10 aircraft. Random trajectories are then sampled from each of the two models, respective to their proportions in the data. This bimodal modal is shown in Fig. 11 (right), which is a much better model for the initial data set.

IV. Validating the Time To Intersection

The time which it takes a departure to cross the missed approach trajectory is a key factor when applying the PCA model to the Monte Carlo simulations. The time it takes the departure to reach the intersection point is found by integrating the speed profiles until reaching a distance of 14,400 ft (the distance from the T10 taxiway of runway 32L to the intersection point). These times are determined for the recorded tracks and the PCA model, shown in Fig. 12 for the Airbus A319 aircraft. There is clearly a better match for the PCA model than the simpler constant acceleration approach.

V. Extending the Model to Additional Data

Models are often required which can take a broader input in order to capture the relation between different parameters. In the context of the converging runway operations, the atmospheric conditions will have a substantial impact upon the performance of the aircraft, namely through the temperature and winds.

We can extend the PCA modeling concept to include these ancillary parameters as additional random variables in the PCA decomposition given in Eq. 1. The temperature and winds, measured on the ground at the airport during the release of the departure, are supplemented to each trajectory. Considering the wind a 2-component parameter, and the temperature a scalar, we now have an additional 3 random variables in the model. However, the data set, $\boldsymbol{X}$, now consists of random variables with different units. Depending upon the scale of the data, the PCA will produce different results.

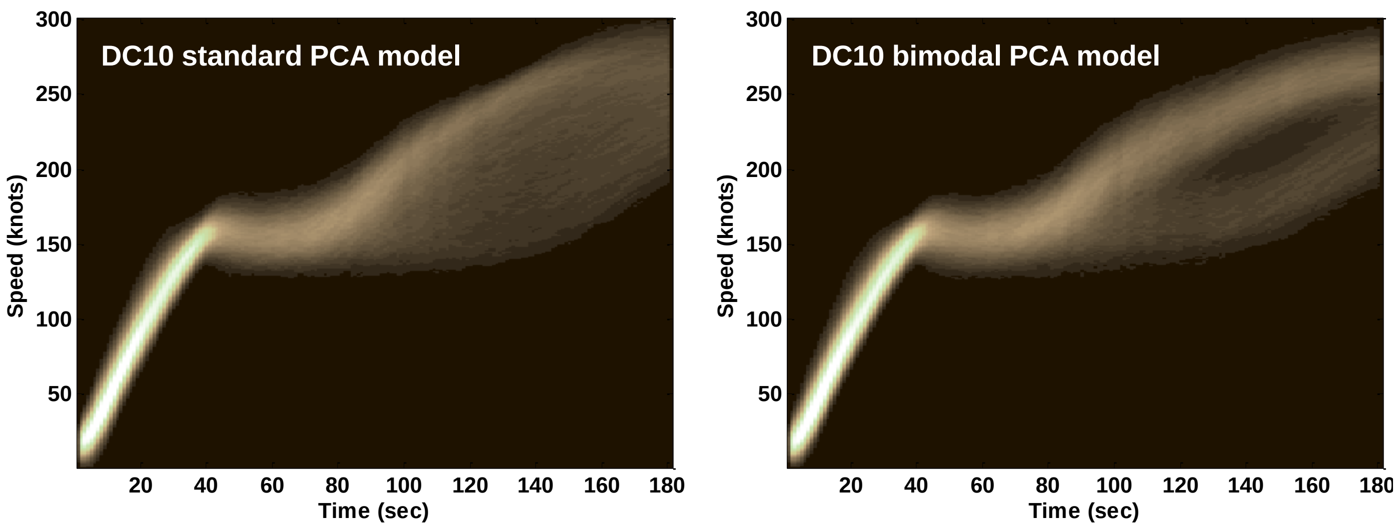

Figure 11. Standard PCA model of DC10 speeds (left) and bimodal PCA model of DC10 speeds (right).

Each data set, ***X***, can be normalized with respect to the variance of its input parameters. This is similar to the PCA obtained by standardized variables**Error! Bookmark not defined.**, where each variable centered around its mean and normalized to the square root of its variance. In this context, we normalize groups of variables, such that the 180 ground speeds are normalized to their average variance, the 2 wind speeds are normalized with respect to their variance, and the temperature normalized to its variance:

$$\mathbf{X} = \left[\frac{\vec{v}_1}{\sigma_v} \quad \frac{\vec{v}_2}{\sigma_v} \quad \cdots \quad \frac{\vec{v}_{180}}{\sigma_v} \quad \frac{\vec{w}_1}{\sigma_w} \quad \frac{\vec{w}_2}{\sigma_w} \quad \frac{t_1}{\sigma_t} \right]$$

The PCA modeling approach can then be applied to X where the last 3 elements of the resulting modes will contain the proportions of the principle components dedicated to the winds and temperatures. The modes are then parsed by their parameters and the normalization removed in order to return to the initial units of the data.

Utilizing this approach to capture the surface wind and temperature, arrival-departure pairs can be sampled and matched according to the atmospheric conditions. Additional modes are included to account for the convergence of the energy of each parameter (similar to Fig. 5).

Conventional approaches might try to adjust the speed profiles or performance models based on the recorded surface measurements in order to account for these factors. However, such an approach adds a substantial number of assumptions which begin to degrade the accuracy of the model. By adding these components to the PCA model, no assumptions are made. Rather, aircraft pairs are selected which were shown to have similar surface conditions when they were flown.

VI. Conclusions

This paper presents a novel methodology for generating data driven models for the Monte Carlo simulation of airport operations. Specifically, this approach is applied to a collision risk analysis of converging runway operations for arrival-departure aircraft pairs. In order to estimate the collision risk, it is necessary to have precise models of each aircraft type's speed trajectory, which are obtained through data driven models from radar tracks.

Data sets were assembled using an aggregate of radar tracks from each aircraft type. Each data set was then decomposed using principle component analysis to obtain an empirical basis set of uncorrelated modes. The coefficients of each mode were then modeled using a Pearson distribution. Sample trajectories were then obtained by randomly sampling the Pearson distribution and summing over each of the modes in the PCA basis set. The resulting trajectories were shown to very closely match those of the input data set, subject to the assumption that the underlying data is governed by random independent fluctuations. Furthermore, it was shown that this approach can be extended to include additional parameters such as winds and temperature, which can be utilized to match the surface conditions of simulated aircraft pairs.

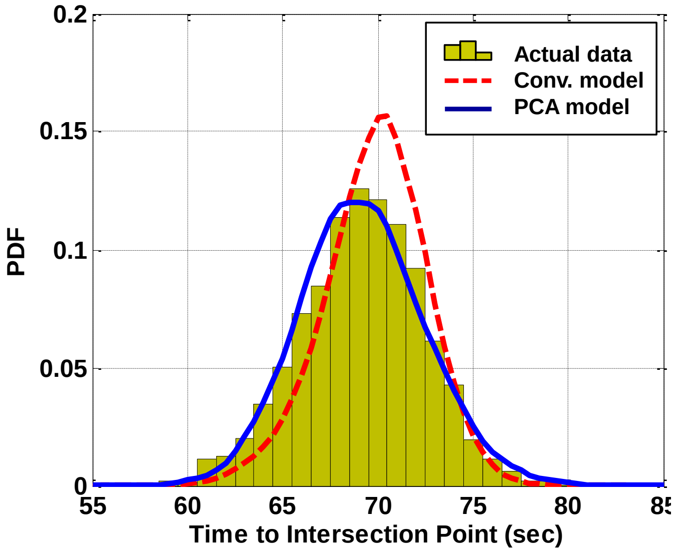

Figure 12. Departure time from T10-RW32L to the intersection point.

While this approach is shown to be successfully applied to the Monte Carlo simulation of converging runway operations, the methodology extends to a more general approach to modeling and simulation. The use of data driven models is able to integrate the validation process by using all available data to drive the generation of a given model, with only basic assumptions about the input data set.

Acknowledgment

The author would like to thank Joseph Clements for his reviews and contributions to the technical approach as well as the design of the Monte Carlo simulation methodology.

References

[1] H. A. P. Bloom, M. B. Klompstra, G. J. Bakker, "Accident Assessment of Simultaneous Converging Instrument Approaches," Air Traffic Quarterly, 2003, Vol 11(2), pp. 123-155.

[2] A. Eckstein, "Automated Flight Track Taxonomy for Measuring Benefits from Performance Based Navigation," Proceedings of the 9th Integrated Communications Navigation and Survaillence (ICNS) Conference, 2009, Arlington, VA.

[3] W. Matsutik, H. Pfister, M. Brand, and L. Mcmillian, "A Data Driven Reflectance Model," ACM Transactions on Graphics, Vol 22, No 3, July 2003.

[4] J. Hung, H. Wang, and L. Lee, " Comparative Analysis for Data-Driven Temporal Filters Obtained Via Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) In Speech Recognition," in Proceedings of ICASSP, 2002.

[5] R. A. Johnson and D. W. Wichern, "Applied Multivariate Statistical Analysis 5th Ed.," Upper Saddle River, NJ, Prentice Hall, 2002.

[6] Y. C. Liang, 2002, "Proper Orthogonal Decomposition and its Applications - Part I: Theory," Journal of Sound and Vibration, vol. 252, no. 3, pp.527-544.

[7] K. Pearson, "Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material," Philosophical Transactions of the Royal Society of London. A, Vol. 186 (1895), pp. 343- 414.