# Trajectory Characterization Using Tracktable

Tom Neafus, Nathan Lieu, Chris Iglesias, Tarun Vellanki, Yoav Dekel, Abhishek Jani

PURDUE UNIVERSITY

Sandia National Laboratories

SAND2020-3932D

## Abstract

Through the use of Python and object-oriented programming, the Sandia National Laboratories Corporate Partners Learning Community focused on categorizing and identifying patterns in airplane flight data. Through using R-Tree sorting, clustering, and distance geometry, trajectories were analyzed based on various characteristics. These characteristics will be used to find correlations between various components in a trajectory to ultimately predict features of a trajectory based on early conditions.

## Background

The goal of this project is to look for trends in large sets of flight data and use these trends to characterize flight trajectories. These characteristics can then be used to predict the trajectories of other flights based on early-flight conditions. There are over 10 million flights each year in the United States alone, and by assessing the common characteristics of flights, it should be possible to identify and correct anomalies before they turn into more serious issues.

The main tool used in this project is Tracktable, a Python module written by researchers at Sandia National Laboratories for the purpose of analyzing trajectory data.

Figure 1: Flights filtered by R-tree query.

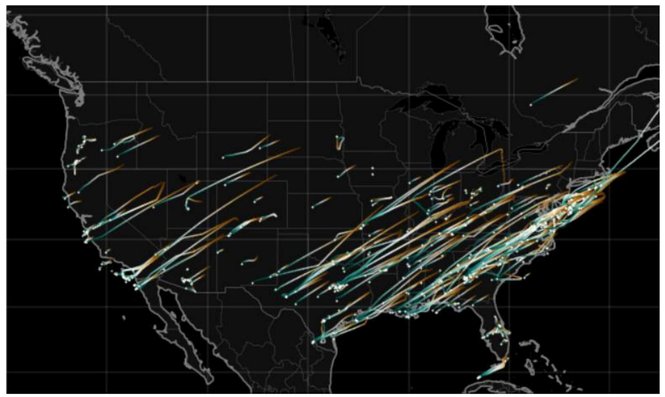

Figure 2: A cluster of flights, clustered by bearing.

Figure 3: Flights with similar distance geometry.

## R-Tree Sorting

Using R-Tree sorting, a data tree structure used for sorting multi-dimensional information, we can group together data. The R-Tree finds points that would fall into a “box” given the value of two corners. In our case, flight paths can be filtered through an R-Tree by using their feature vectors, which allows us to group a set a flights based on many things, like similar length and shape. In Figure 1, flights are sorted based on their similar value for straightness ratio.

## Clustering

Using DBSCAN, an implementation of the k-nearest neighbor algorithm, flights were sorted into clusters based on various features. Figure 2 is an example using with bearings. Each flight in this cluster has an overall bearing within .2 degrees of at least four other flights. Clustering based on different features allows us to characterize groups of trajectories and look for which features tend to correlate.

## Distance Geometry

Distance geometry is a way to measure a curve, by breaking it into smaller parts. It gives information bends and wiggles in curve and how big they are
The algorithm generalizes 3 parameters: a trajectory, depth, and normalization. The depth is used to determine the fidelity of the measurement. Higher depth levels include more information about shape. Normalization determines whether information about size is included.

## Results

Using the methods of R-tree sorting, clustering, and distance geometry, flights were characterized by a wide range of characteristics. These characteristics will be used to find correlations between different aspects of trajectories. By leveraging these correlations, we aim to make accurate predictions of future flight conditions based on early characteristics.

## Acknowledgements

We would like to thank Sandia National Laboratories for use of the Tracktable software and for their guidance and mentorship throughout this project. We would also like to thank Dr. Ward and the Data Mine staff for facilitating this collaboration between Purdue University and Sandia National Laboratories.

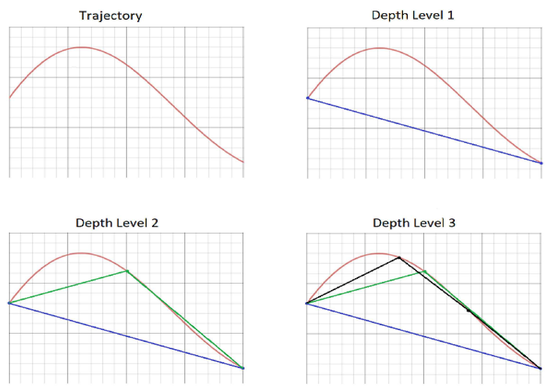


Figure 4: An example of distance geometry.