PURDUE UNIVERSITY | The Data Mine

FLIGHT TRAJECTORY STITCHING

Sandia National Laboratories

Contributors: Maxwell Low, Annie Jancaric, Kuba Bal, Bryce Castle, Rishee Gupta, Max Martinez, Chirayu Marwaha, Ali Paloba, Chaithanya Podishetti, Jabez Williams

Definitions

Trajectory - *set of time & position points that define a flight*
Tracklet – *a segment of a flight's trajectory*
Features - *a value that summarizes a set of points, examples include average velocity or duration*
Fréchet Distance - *used to compare two curves in a trajectory to check how similar they may be.*
Transponder - *aircraft device that sends positional data*
Kinematics – *physical attributions of the aircraft*

Introduction

The track stitching problem attempts to match together fragments of tracklets (i.e. collection of points each with latitude, longitude, and timestamps), so that matched tracklet pairings represent complete motion paths.

An effective track stitching solution allows what would otherwise be incomplete, disjointed motion datasets to be closer to a useful dataset. With ever-increasing amounts of data being generated by moving objects, effective solutions to the track stitching problems are of increased importance. This poster presents two different approaches to this problem.

Motivation

- Individuals may tamper with transponder data stream
- Data collectors are spare in some geographical locations
- Flight's transponders are often turned off before landing

Getting started

- Summarize current state of trajectory research
- Create a ground truth dataset by removing data from full flights, this allows us to check our work later
- Primary methods of stitching trajectories to create a predicted flight path
- Create a range of features for single flight

Acknowledgments

We would like to thank everyone at the Data Mine for their constant support and help during our research.

We would also like to thank our mentors, Dr. Katrina Ward and Dr. Andy Wilson at Sandia National Laboratories for their guidance and support throughout this year. They have not only aided in our success in this research, but also supported data science knowledge and personal growth. Additional thanks to Dr. Benjamin Newton, who provided the team with a background presentation and inspiration.

Research Methodology

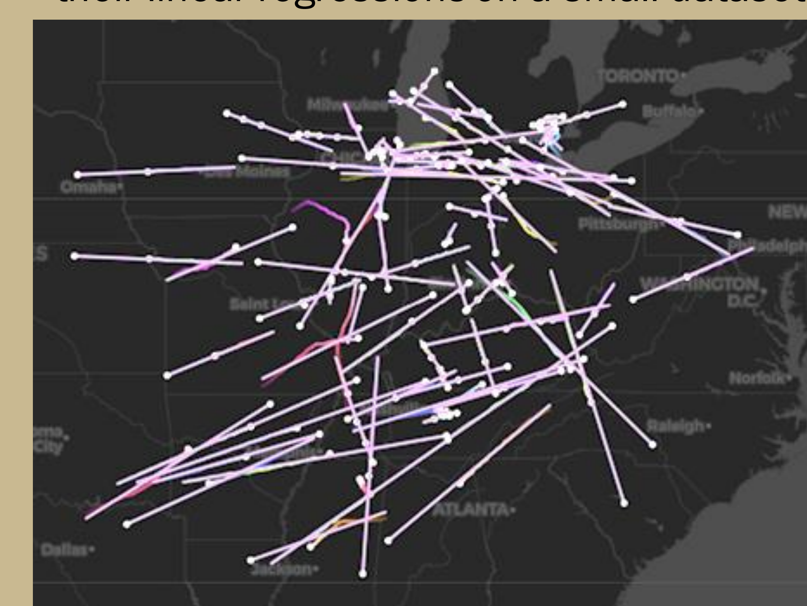
Regression Analysis

► *A plane's trajectory can be predicted with a linear regression.* ◄

Our steps:

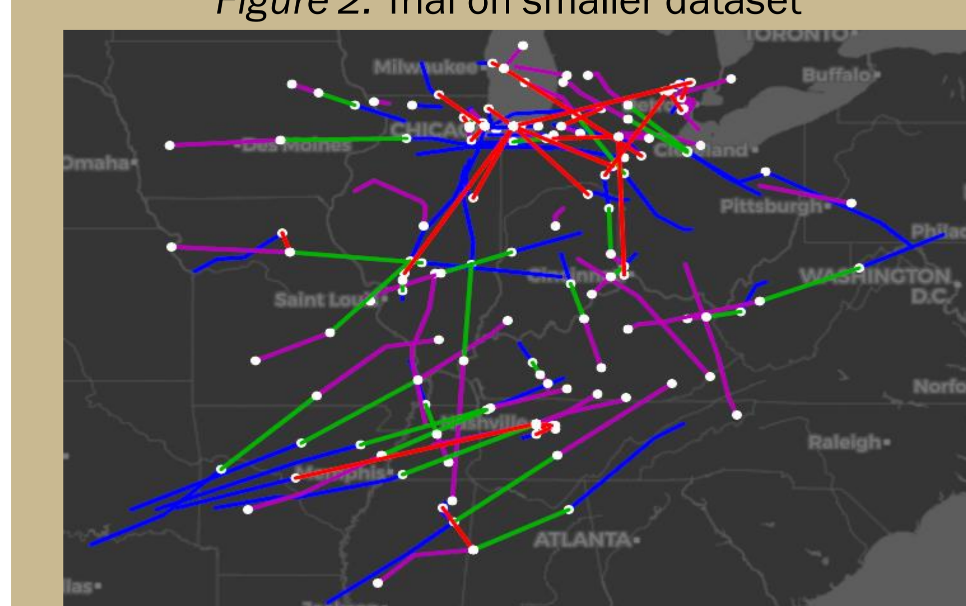
1. Eliminate nonsensical flight paths.
   - Examples include time travel, supersonic flights, a trajectory connecting to itself.
2. Make a linear regression of the second half of each flight.
   - We use only the second half to perform better when the flight is curved.
   - A point's longitude is its x-coordinate and its latitude is its y-coordinate.
   - The second half of the flight predicts the trajectory's path better compared to the first half or the whole trajectory
3. Project the linear regression out to the next tracklet and measure how similar they are.
   - We measure similarity by checking the distance between the regression and the flight at several points.
4. The tracklet which matches the linear projection most closely is judged to be the best match

Figure 1. Visualization of tracklets and their linear regressions on a small dataset.



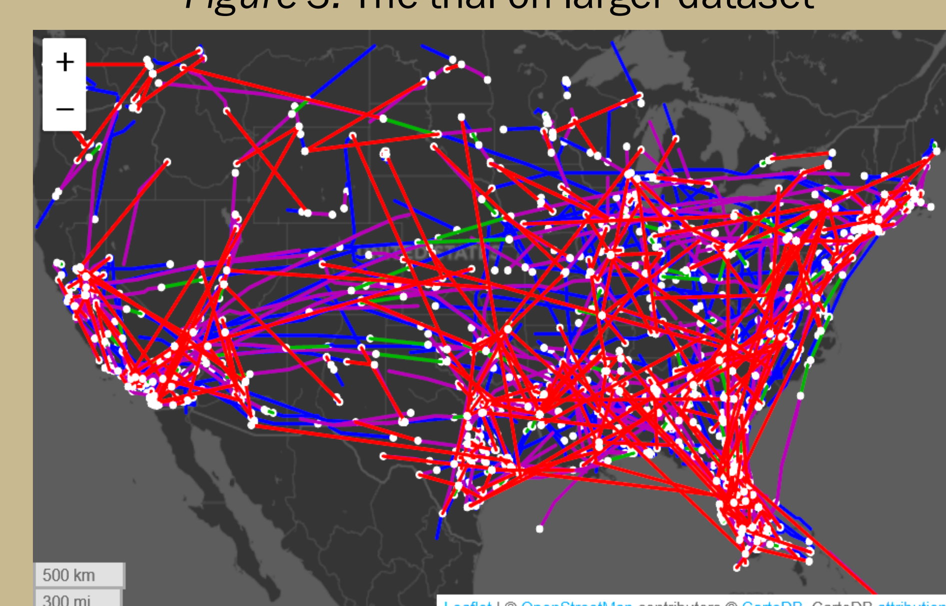
Purple trajectories are linear regressions, other colors are tracklets

Figure 2. Trial on smaller dataset



Green lines represent correct pairings and red lines represent incorrect ones.

Figure 3. The trial on larger dataset



Green lines represent correct pairings and red lines represent incorrect ones.

Future Research & Improvements:

- Methods to determine if a plane landed at an airport at the end of its trajectory.
- Methods to detect if the turning radius of a connection is too tight.
- A regression which accounts for the curvature of the Earth.

Conclusion:

- Accuracy of Small dataset: 56%
- Accuracy of Large dataset: 27.6%
- Solution doesn't work well on denser datasets, from the increased possibility of false positive matches.
- Used only linear regression as higher-dimensional regressions led to overfitting easily

Feature Similarity

► *Pieces from the same flight should share similar features.* ◄

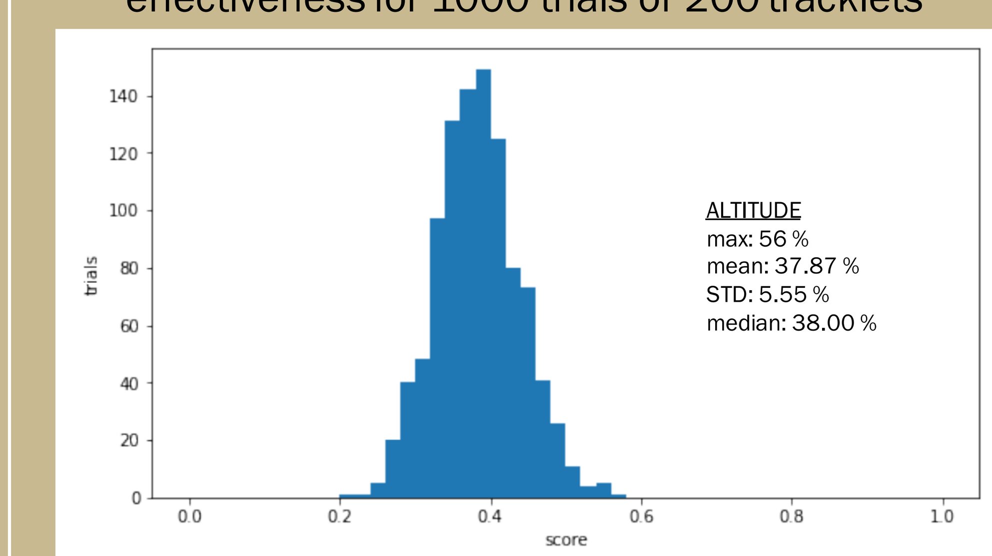
Our methodology:

1. Create features for each flight

altitude	heading	distance	duration
average speed	unsigned heading change	signed heading change	heading ratio
loitering ratio	climb rate	speed between	distance between

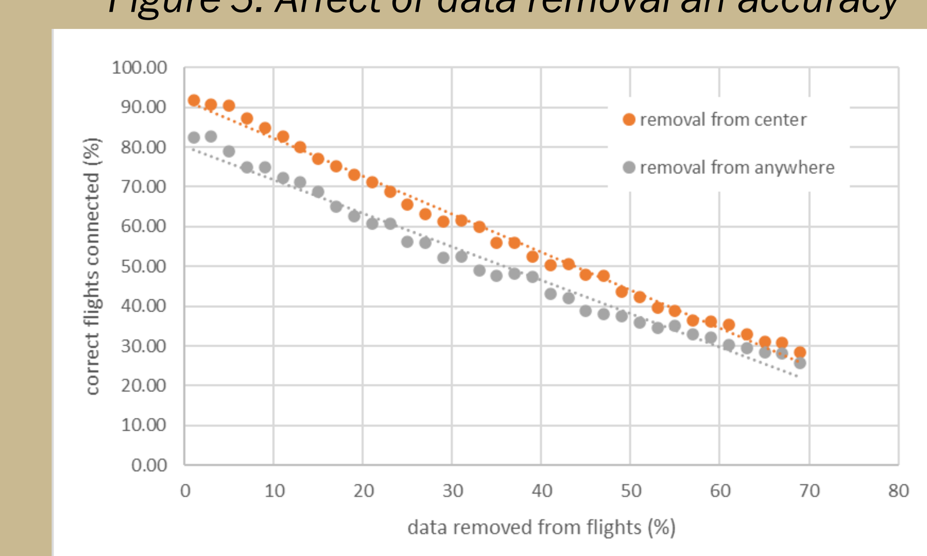
2. Create time possibility matrix by removing connections that require time travel
3. Compare feature of one flight to that of another
4. Normalize each matrix to prepare it for weighting
5. Combine multiple matrices together using weighting. Weights are found using Monte Carlo sampling & gradient descent
6. Select matches by selecting the largest similarity score

Figure 4. Termination Altitude matching solo effectiveness for 1000 trials of 200 tracklets



Altitude feature scores for flights missing 10%

Figure 5. Affect of data removal an accuracy



How removal percent affects solution effectiveness

Future Research & Improvements :

- PCA dimensional reduction to reduce the number of features processed (helps speed up runtime for larger datasets).
- Incorporating different features to see which ones work best.

Conclusion:

- Was able to correctly match 87% of pairs from a set of 1,000
- Altitude and heading are the most impactful features, contributing 38% and 15% respectively to our matching algorithm.
- Duration was the least impactful, contributing 0.2%