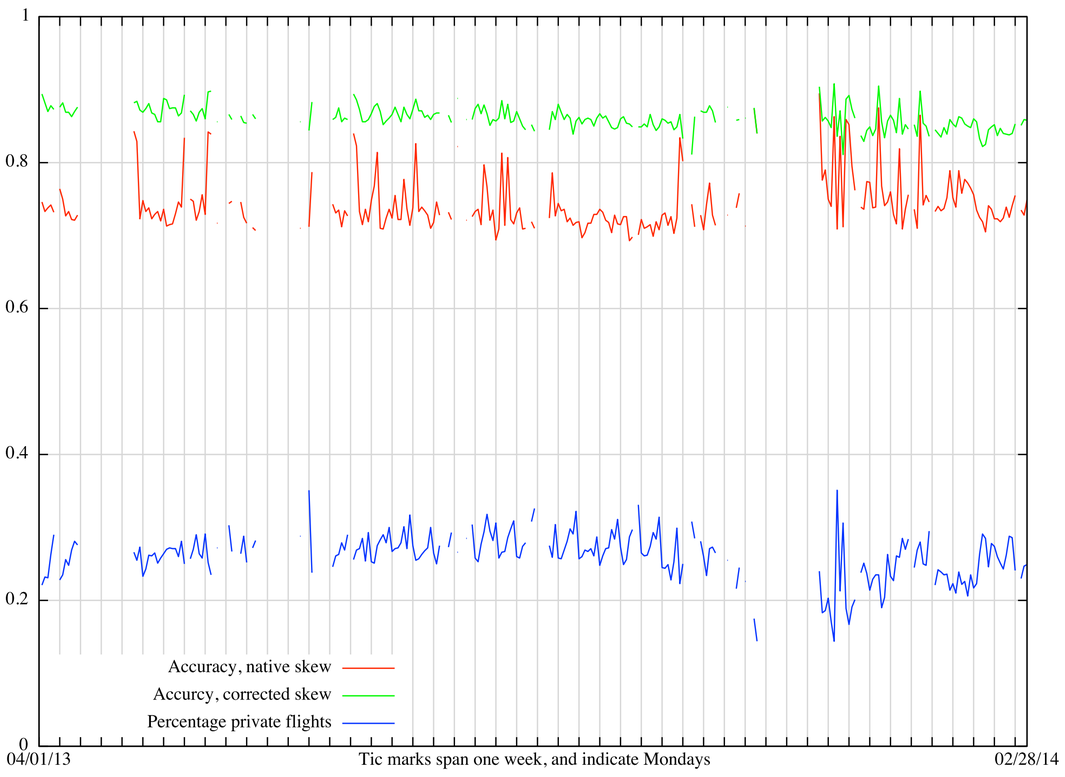
Characterizing and Detecting Aircraft Identity and Diversion

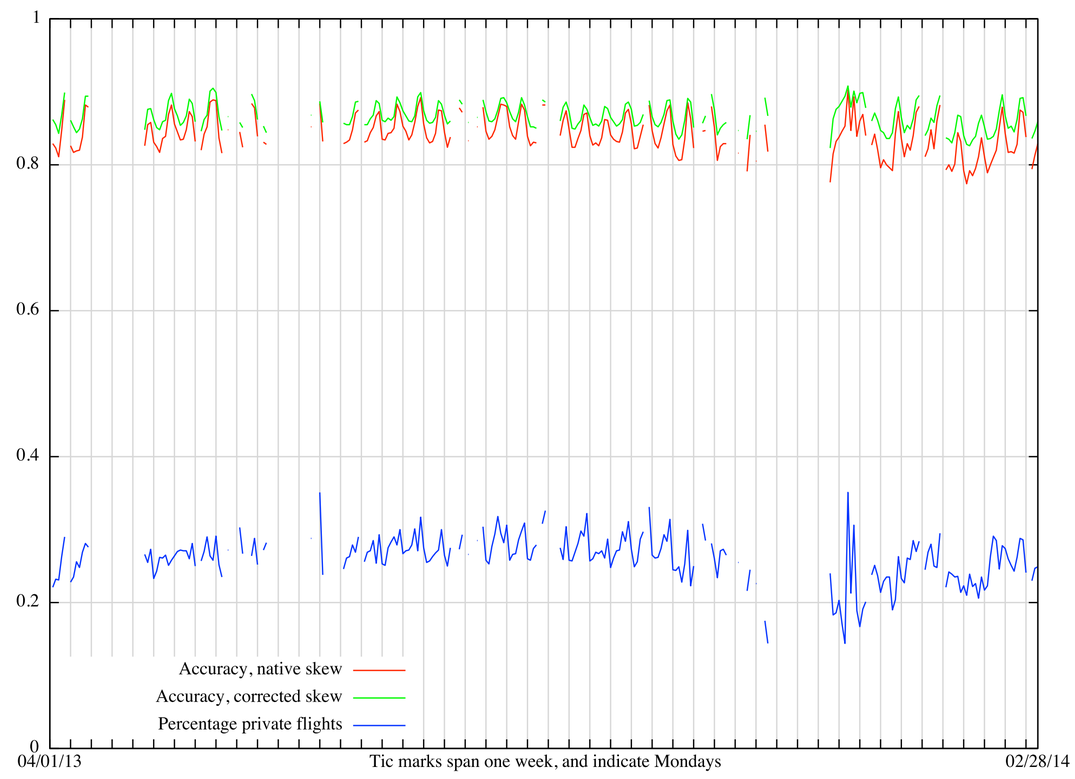
Authors: Philip Kegelmeyer (wpk@sandia.gov), Timothy M. Shead, Danny Rintoul, Andy Wilson. Sandia National Laboratories.

Question: Can we, solely from ASDI metadata, characterize diverted vs undiverted flights? Commercial vs private flights? Well enough to label them? How confidently? What features are most useful for doing so?

Accuracy in Diverted/Undiverted Labeling
(Class averaged accuracy: compare to 0.5)



Accuracy in Private/Commercial Labeling
(Class averaged accuracy: compare to 0.5)



Data Description

- "Aircraft Situation Display to Industry" (ASDI) data
- Each aircraft pinged every 5-60 seconds.
- Each ping generates a vector of information called a "point".
- A point is status, heading, speed, etc; 17 fields of data.
- Points are assembled into "trajectories", one per flight.

Data Statistics

- Total storage: 303 Gbytes
- Total number of points: 1.3 billion
- Time to convert points to trajectories: 177 minutes (parallel)
- Total number of trajectories: 9.5 million

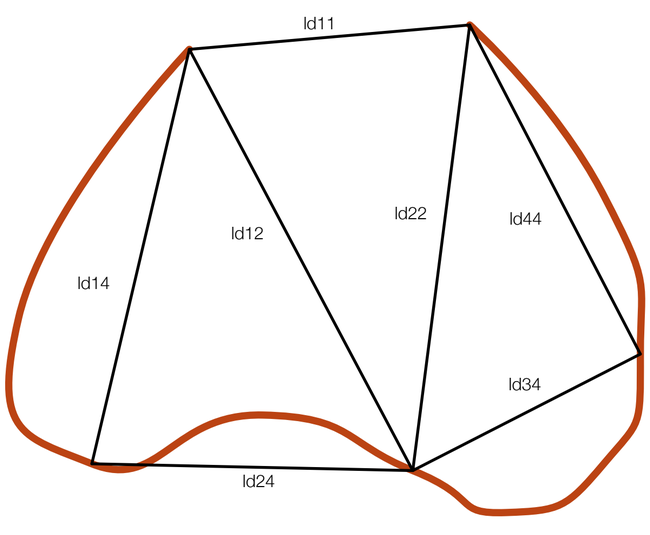
- Date range: April 1, 2013 to February 28, 2014.
- So: 334 days of data, but …
- 48 days with no data (and others with less data)

Label Definitions

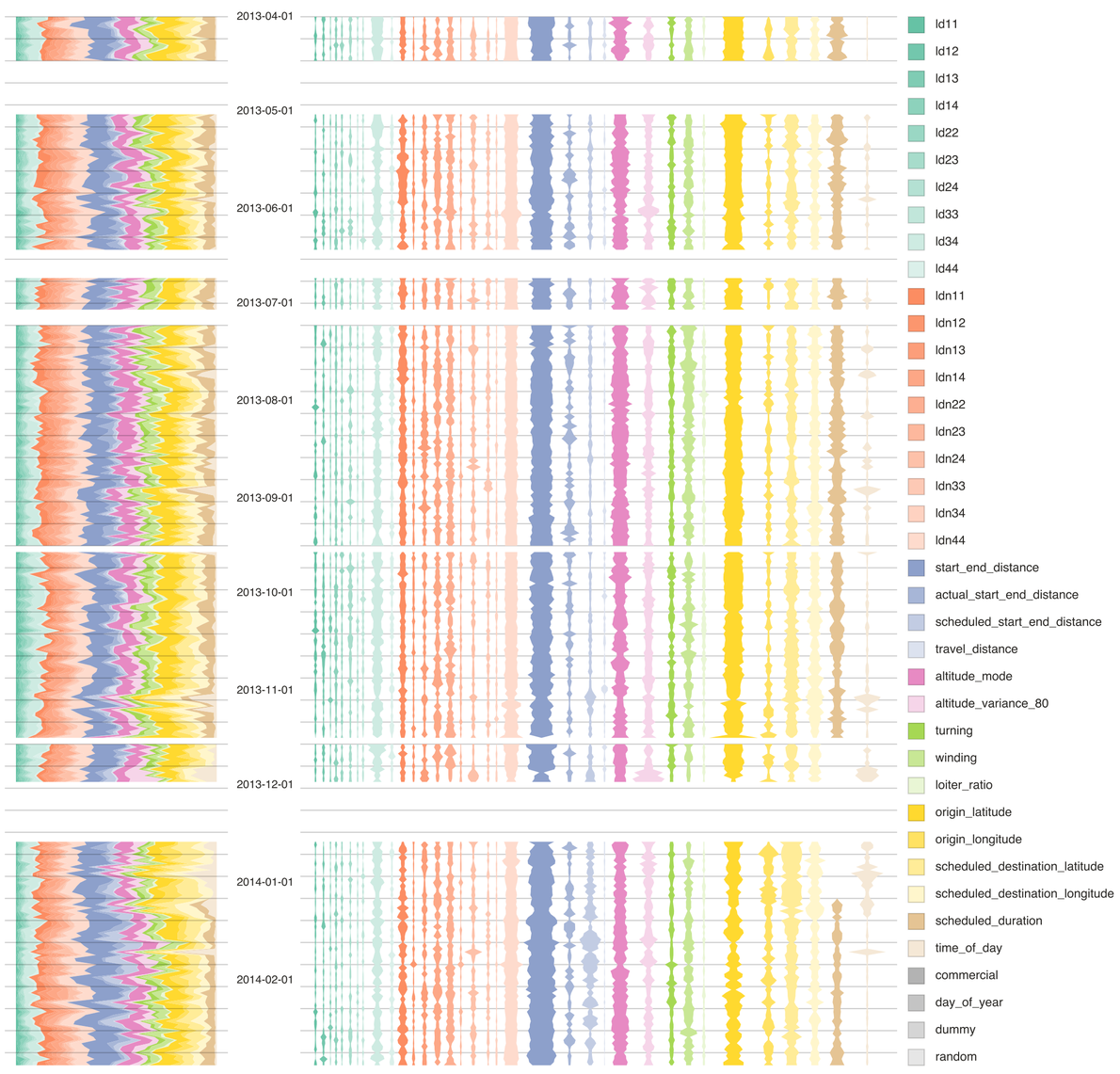
- Diverted: The "destination" field in the first and last points in the trajectory differ. Data is 7.2% diverted, overall.
- Tail number matches the regular expression "N[0-9]*". Data is 26.5% private, overall. But note the swings …

Feature Definitions, For Diversion

day_of_year	:
time_of_day	:
origin_latitude	:
origin_longitude	:
scheduled_destination_latitude	:
scheduled_destination_longitude	:
scheduled_duration	: delta between original scheduled departure / arrival times.
altitude_variance_80	: variance of altitude for middle 80% of trajectory.
altitude_mode	: mode of altitude for entire trajectory.
travel_distance	: actual path length.
scheduled_start_end_distance	: distance from origin airport to scheduled destination airport.
turning	: sum of the absolute values of the trajectory's turns.
winding	: sum of the signed values of the trajectory's turns.
loiter_ratio	: travel_distance divided by convex hull circumference.
ldNM	: segment N of M from an M-ary partition of the trajectory.
ldnNM	: ldNM divided by the actual path length.
commercial	: 0 for private, 1 for commercial.
random	: a random, irrelevant feature.
dummy	: a fixed convenience feature, not used for prediction



Feature Importance (for Diversion) as a Function of Time



Feature Definitions, For Identity

Delete …

commercial	: 0 for private, 1 for commercial.

Add…

start_end_distance	: distance from start trajectory point to end trajectory point.
actual_start_end_distance	: distance from origination airport to actual destination airport.
undiverted2	: 0 for diverted, 1 for undiverted

Sidebar: The Difference A Year Makes

The Probe Trajectories

Best matches when searching over just one day of data

Best matches when searching over 283 days of data

Time to sort 9.5 million trajectories with one probe: 15 minutes (serial).

This work was supported by the PANTHER LDRD program at Sandia National Laboratories. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Company, for the U.S. Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.