Anomaly Detection In Onboard-Recorded Flight Data Using Cluster Analysis

Lishuai Li, PhD Candidate, MIT (lishuai@mit.edu)

Maxime Gariel, Postdoc, MIT (maxime.gariel@gmail.com)

R. John Hansman, Professor, MIT (rjhans@mit.edu)

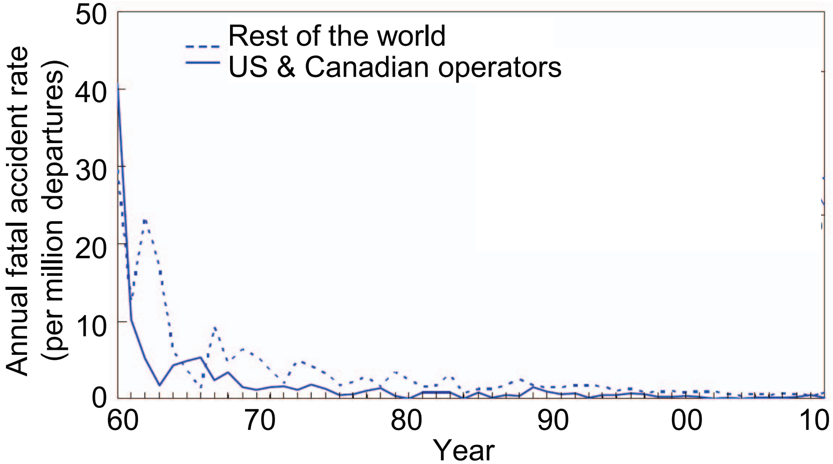
Rafael Palacios, Professor, Comillas Pontifical University, Spain
(rafael.palacios@iit.upcomillas.es)

30th DASC

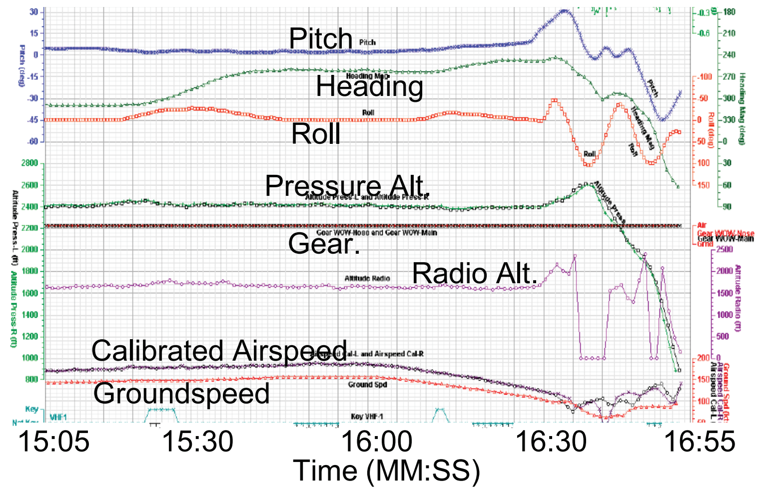
October 18, 2011

Motivation

- **Aviation safety management**
  - Past: learn from accidents
  - Now: proactively identify safety hazards during routine operations
- **Massive data collected during routine flights**
  - Can be used to improve safety and operations
  - Data recorded onboard by Flight Data Recorder (FDR) or Quick Access Recorder (QAR)
  - No. of parameters: 88-2000
  - Sample rate: 0.25-8 Hz



Fatal Accident Rate – Commercial Jet Fleet[1]



Onboard Recorded Flight Data Example[2]

[1] Boeing, 2010 Statistical Summary, June 2011. [2] NTSB, FDR Group Chairmen's Factual Report, DCA09MA027, 2009

Current Data Analysis

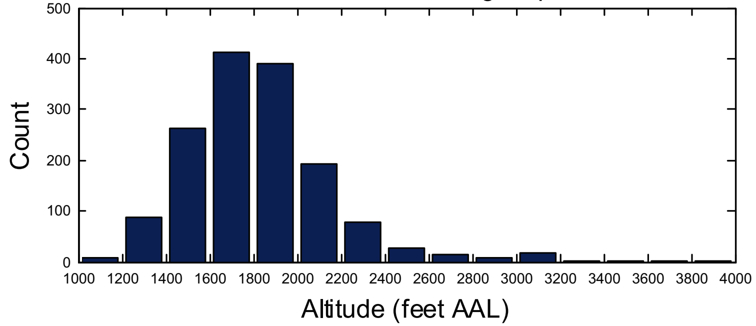
- **Current analysis method commonly used in Flight Operational Quality Assurance (FOQA) programs**
  - Exceedance analysis
    - ✓ Detect anomalous events by pre-specified limits

Examples of Exceedance Events[1]

Event	Parameters and safety limits
Takeoff Climb Speed High	HAT > x ft, HAA < x ft, CAS > V2 + x kts
Excessive Power Increase	HAT < 500 ft, Δ of N_1 > x
Operation Below Glideslope	Glide Slope Deviation Low > x dots, HAT < x ft

  - Statistical analysis on specific queries

Altitude At Which Landing Flap Is Set [2]



- ***Limitation: Only known issues are examined***

[1] Federal Aviation Administration. 2004. Advisory Circular. 120-82 Flight Operational Quality Assurance [2] Campbell, Neil. 2003. Flight Data Analysis – An Airline Perspective

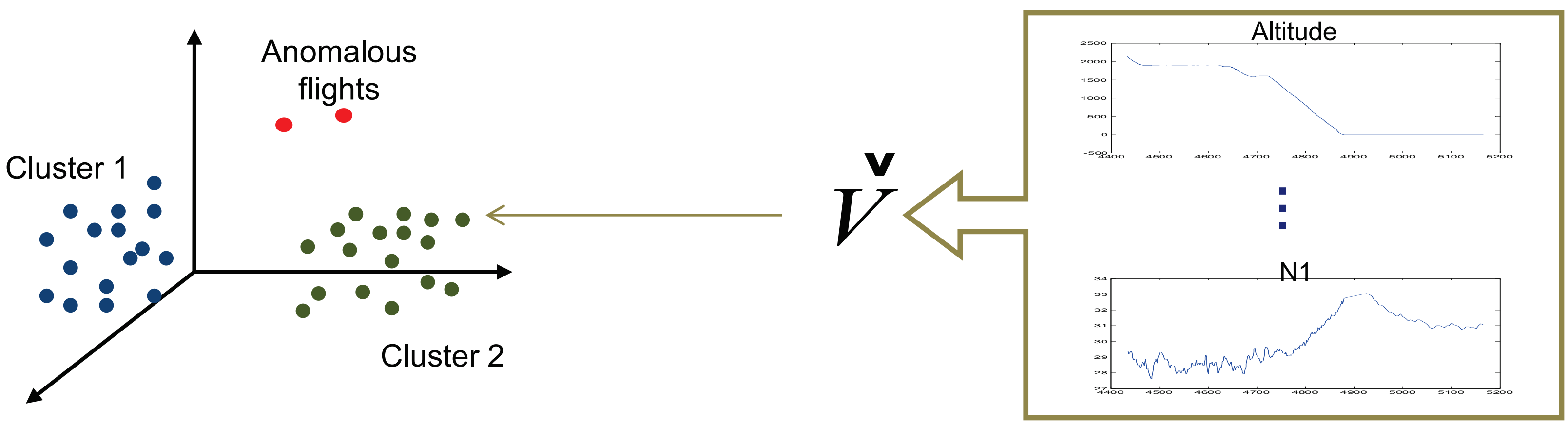
Objective

- **Develop a method to detect anomalous flights**
  - From routine FDR/QAR data
  - Without specification of what parameters to watch and their thresholds
- **Characterize data patterns of multivariate time series**
  - Identify nominal cases and abnormal cases
- **Identify anomalous flights, which are then referred to domain experts for further review**

Approach

- **Multivariate Cluster Analysis**
  - Identify clusters of similar flights based on multiple flight parameters

Hyperspace **Flight Parameters** of Flight X

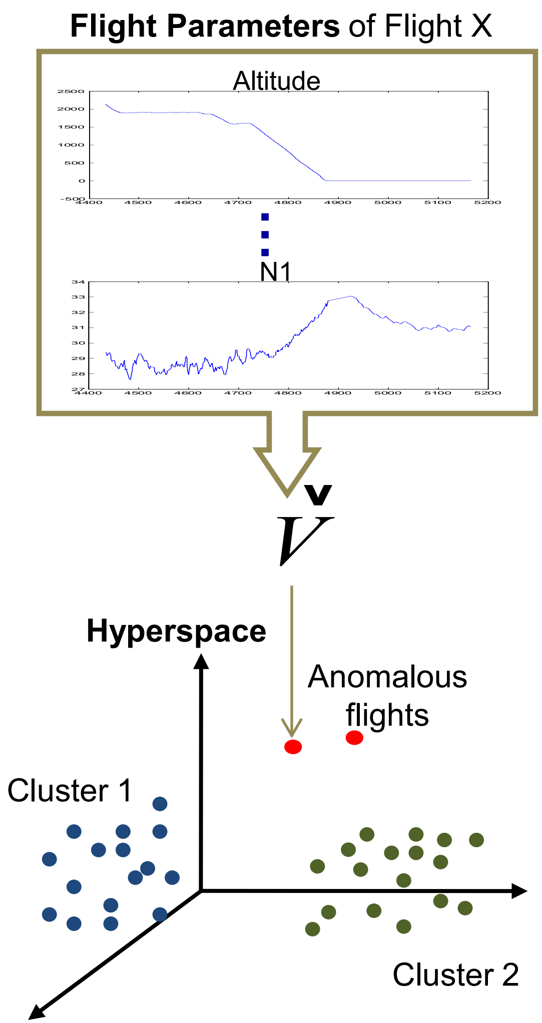


- **Perform cluster analysis in a hyperspace where each data point represents a flight**
  - Identify nominal clusters
  - Detect outliers

- **Need to transform data recordings to a form applicable for cluster analysis**
  - Convert all time series to a vector for each flight

Three-Step Process

- **General Approach**



- **Detailed Steps**

Time Series

Step 1 - Time Series to Vectors

Vectors of High Dimensions

Step 2 - Dimension Reduction

Only when the number of data points is smaller than vector dimensionality

Vectors of Reduced Dimensions

Step 3 - Cluster Analysis

- Clusters
- Anomalous Flights

Method Evaluation

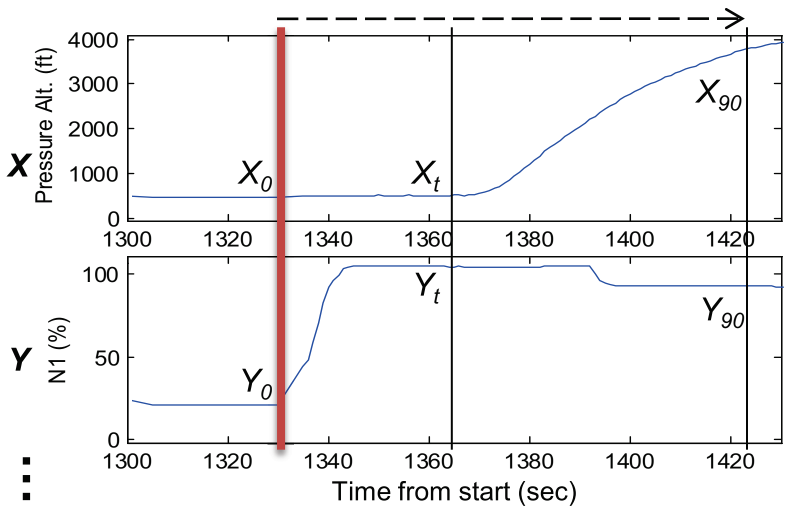
- **A limited set of FDR data was available for this research**
  - 2881 flights
  - 7 aircraft types (13 model variants)
- **Subsets by aircraft model were separately evaluated**
  - Available flight parameters vary by aircraft model
  - Results of B777 subset were presented in this paper
    - ✓ Number of flights: 365
    - ✓ Number of flight parameters: 69
- **Applied the multivariate cluster analysis approach**
  - Three-step process
- **Evaluated results from cluster analysis**
  - Anomalous flights
  - Nominal clusters

Step 1: Time-series to Vectors

- **Obtain samples referring to *a specific event***

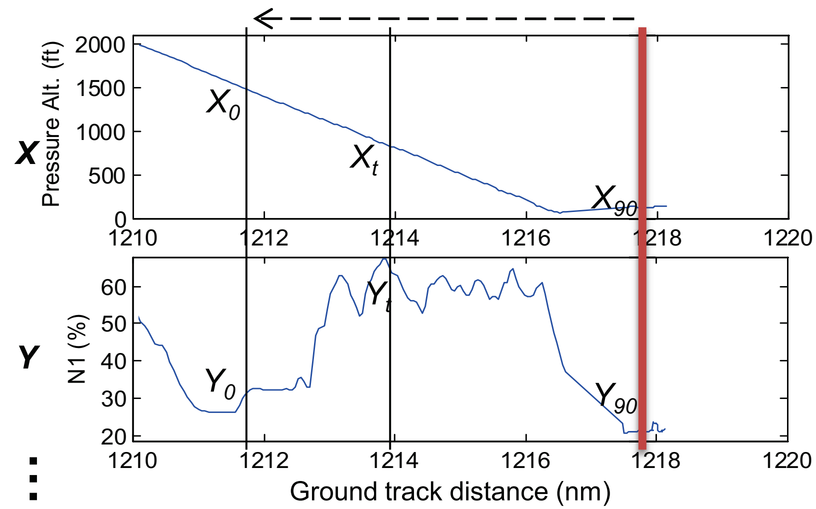
Takeoff Phase

Track samples *from* ***applying takeoff power*** to 90 sec after takeoff



Approach Phase

Back track samples from ***touchdown*** to 6nm before touchdown



- **Form a vector for each flight:**

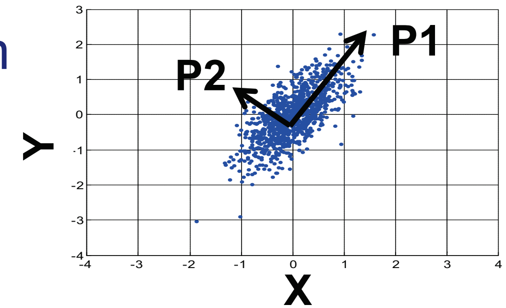
$$\vec{V} = \underbrace{[\underbrace{X_0, X_1, \cdots, X_{90}}_{\text{91 samples per parameter}}, Y_0, Y_1, \cdots, Y_{90}, \cdots]}_{\substack{\text{68 parameters for takeoff} \\ \text{69 parameters for approach ("Radio Height" only available during approach)}}}$$

Step 2: Dimension Reduction

- **Due to the sparseness of this dataset, need to reduce dimensions to increase data density for cluster analysis**

Number of data points	**<<**	**Number of dimensions**
365 flights		6279 for approach; 6188 for takeoff

- **Principal Component Analysis (PCA)**
  - Transform data to a new coordinate system
    - ✓ The 1st PC captures the greatest variance by any projection of the data; the 2nd PC captures the second greatest variance; and so on



- **Keep the first *K* PCs that explain more than 90% variance**

$$[X_0, X_1, \mathrm{L}\ , Y_0, Y_1, \mathrm{L}\ , \mathrm{L}\ ]$$

PCA

$$[P_1, P_2, \mathrm{L}\ , P_K]$$

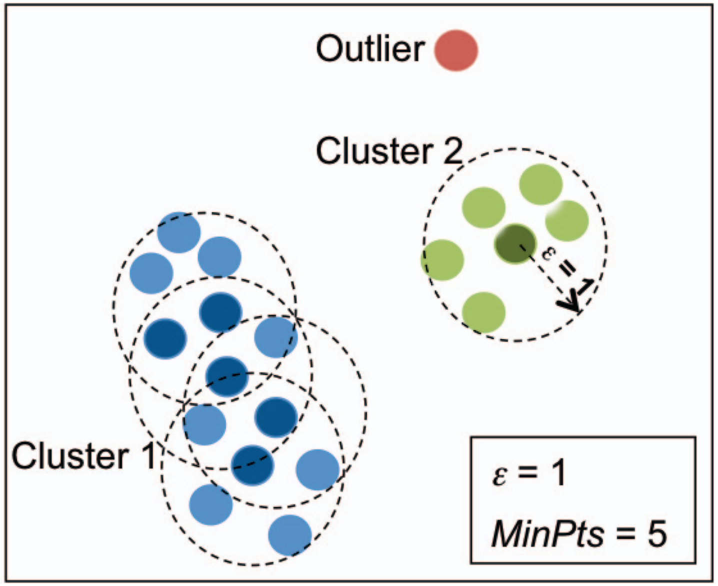
Number of dimensions

Takeoff	Approach
6188	6279
77	95

Step 3: Cluster Analysis

- **Density-based clustering algorithm: DBSCAN**
  - Find density-connected points progressively to form a cluster
  - Advantages:
    - ✓ Handle outliers in the data
    - ✓ No need to know how many clusters in advance
    - ✓ Discover clusters of arbitrary shape

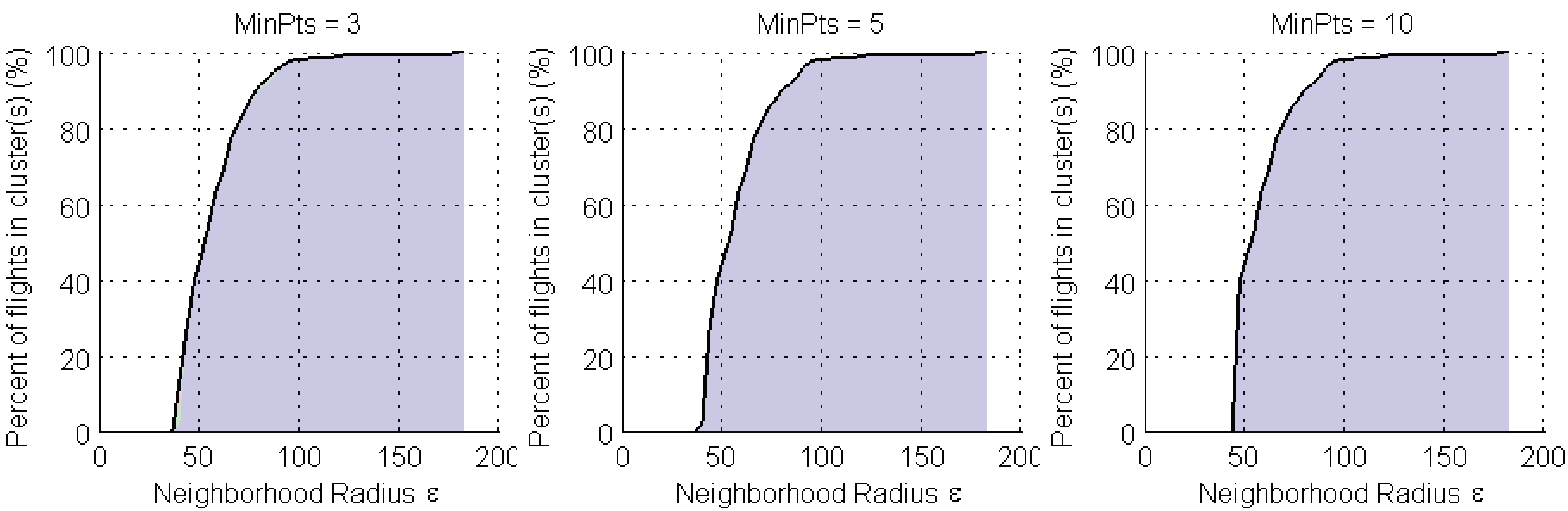
- **Two global parameters:**
  - ε
    - Max. radius of the neighborhood
    - Distance between points
  - *MinPts*
    - Min. number of points in an ε-neighborhood



- **A cluster is created if at least *MinPts* lie within a ball of radius ε centered at a point**

Sensitivity Analysis of DBSCAN Parameters

Sensitivity Analysis of ε and *MinPts on B777 Approach*



- **Sensitivity analysis of *MinPts* and ε was conducted**
  - Results were not sensitive to *MinPts on the B777 dataset*
    MinPts = 5 was used
  - Results were sensitive to ε *on the B777 dataset*
    ε was set to find the top 1%, 3% or 5% outliers

Method Evaluation Results

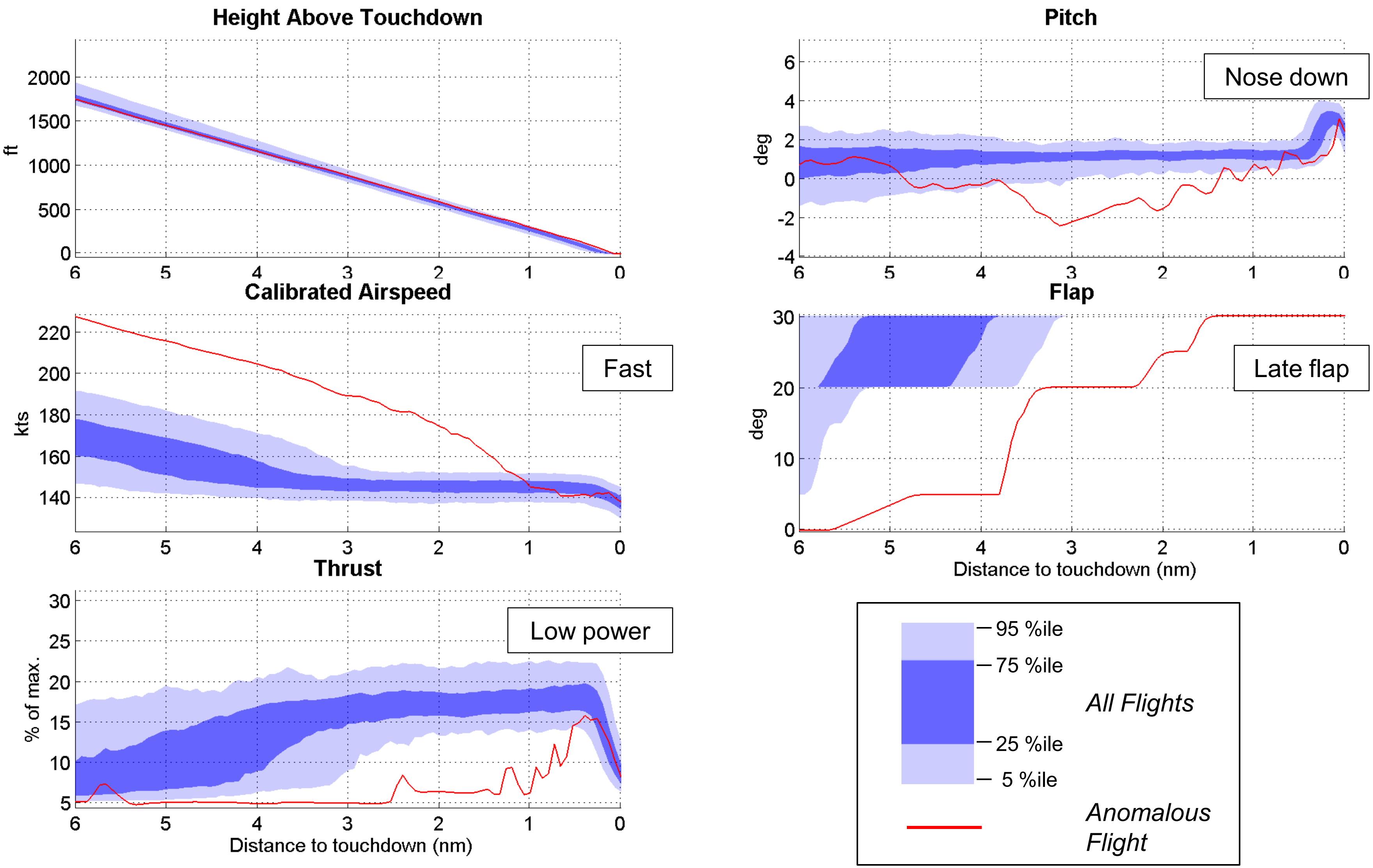
- **Evaluation dataset**
  - Aircraft type: B777
  - 365 flights
  - 14 airports as origin or destination
  - 69 flight parameters
- **Evaluation procedure**
  - Separate analysis for approach phase and takeoff phase
  - Identified top 1%, 3% and 5% outliers
  - Inspected all anomalous flights for abnormal behaviors
  - Examined commonalities for flights in nominal clusters

Anomalous Flights Identified in Approach Phase

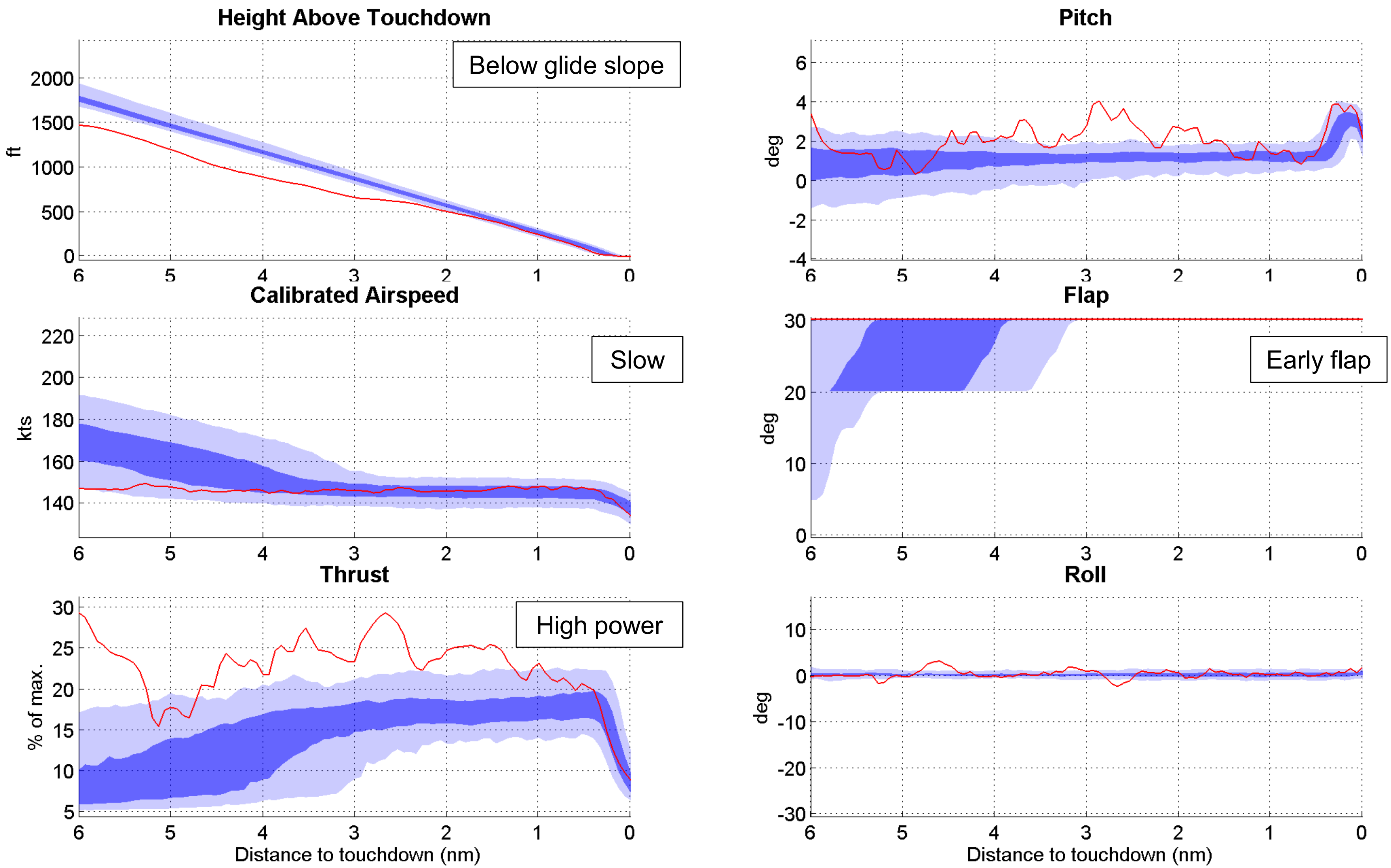
	Flight ID	1%	3%	5%	Abnormal Behaviors	Abnormal Type
✓	373547	x	x	x	Fast	High energy approach
	377844	x	x	x	High, line up late	High energy approach
	378688	x	x	x	Fast, unstable airspeed	High energy approach
	377288		x	x	Initially fast	High energy approach
✓	383780		x	x	Low, slow	Low energy approach
	375698		x	x	Low, high power	Low energy approach
	383270		x	x	Low, unusual yaw trim	Low energy approach
	372235		x	x	Hot weather	Weather effect
✓	371044		x	x	Abnormal high pitch	Other unusual operation
	371045		x	x	Line up late	Other unusual operation
	377860			x	Fast	High energy approach
	379685			x	Initially fast, then slow	High energy approach
	371040			x	Fast	High energy approach
✓	379665			x	Strong crosswind	Weather effect
	383285			x	Unusual flap setting	Other unusual operation
	384110			x	Unusual flap setting	Other unusual operation

- **Types of abnormal behaviors**
  - High energy approach, low energy approach, weather effect, other unusual operation
- **Example anomalous flights of each identified type are shown in the following slides**
  - Example flights are indicated by ✓

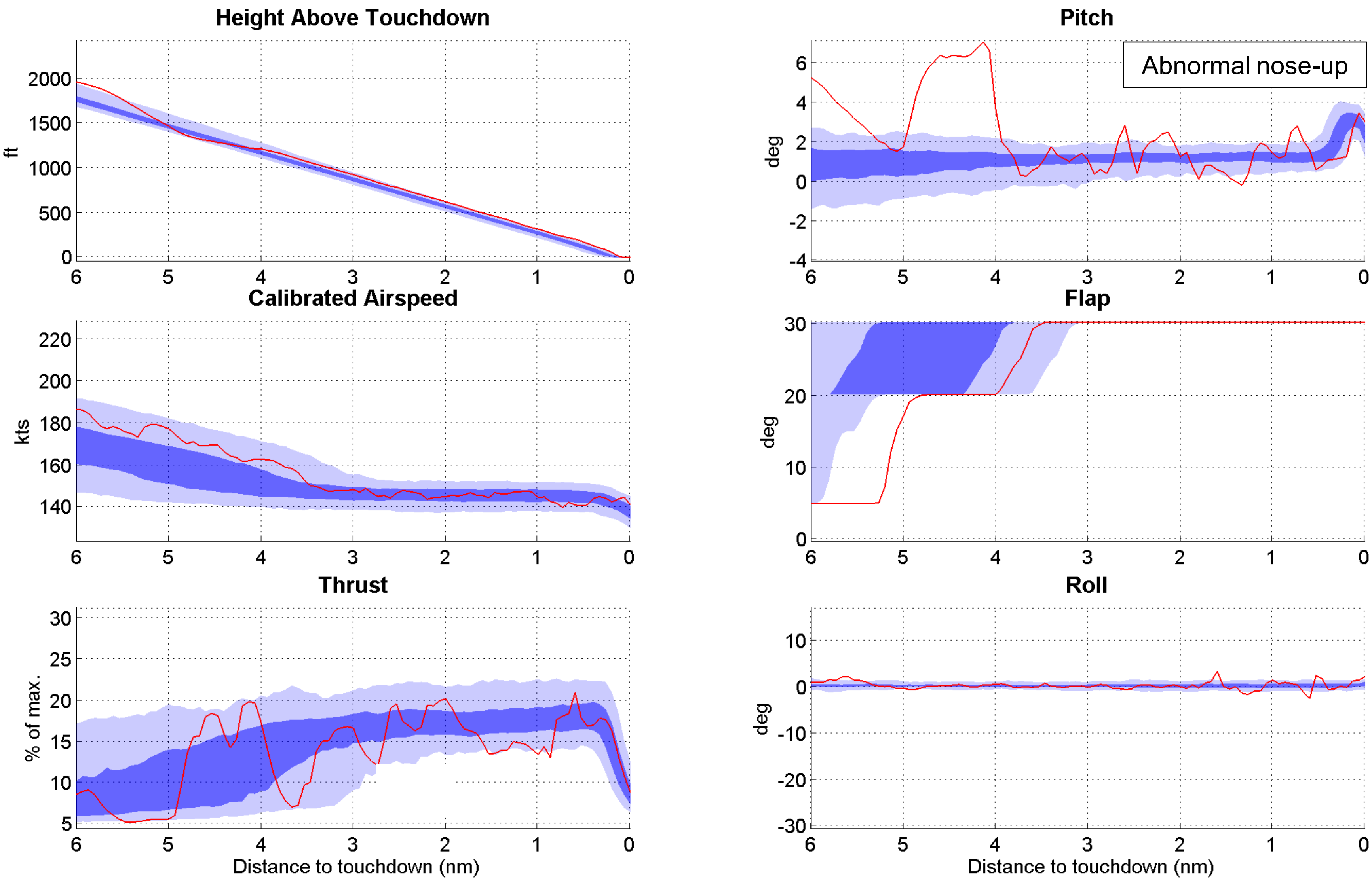
Anomalous Flight: 373547 – Fast Approach



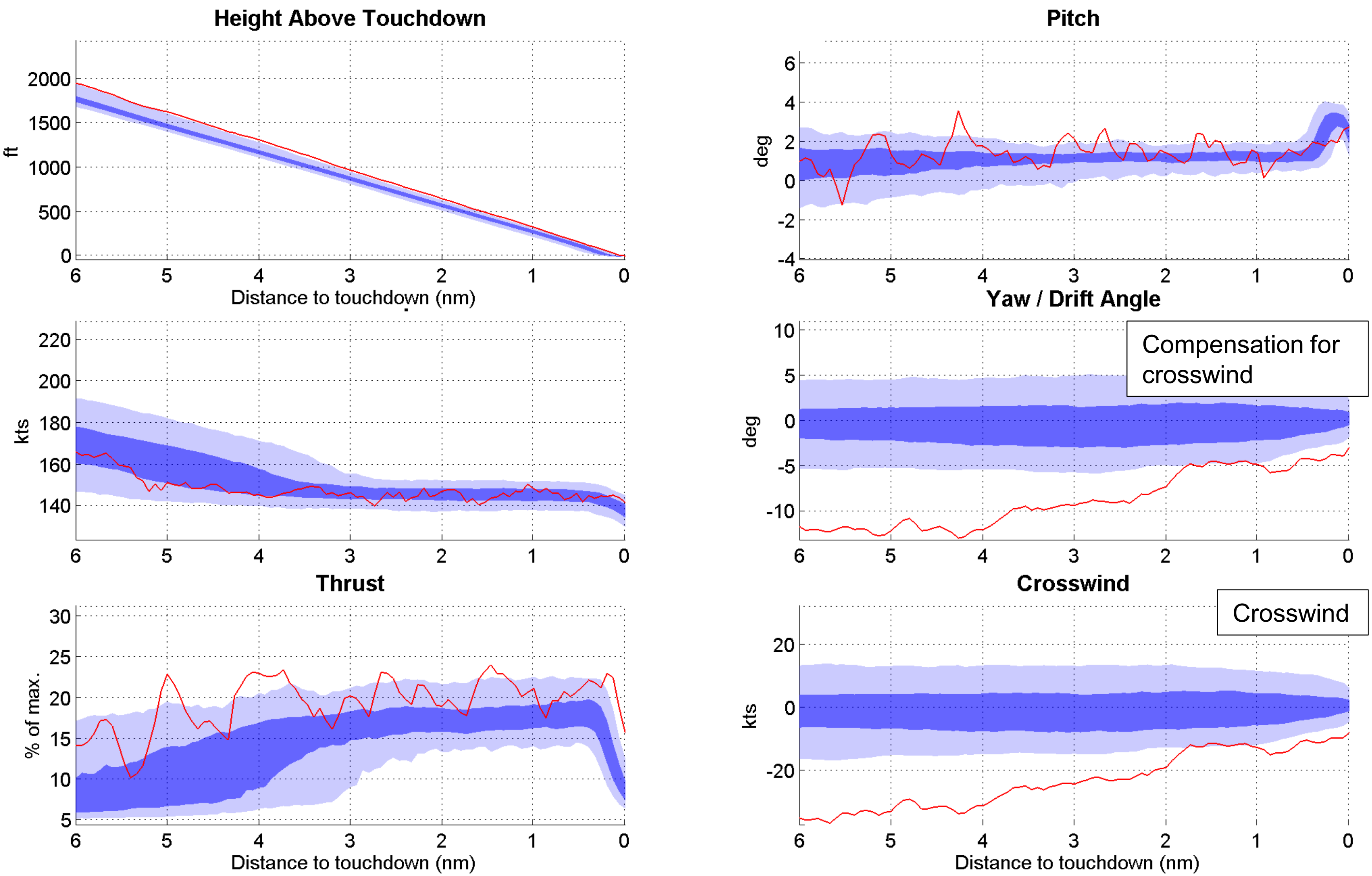
Anomalous Flight: 383780 – Low & Slow Approach



Anomalous Flight:
371044 – Abnormal Pitch



Anomalous Flight: 379665 – Crosswind Approach

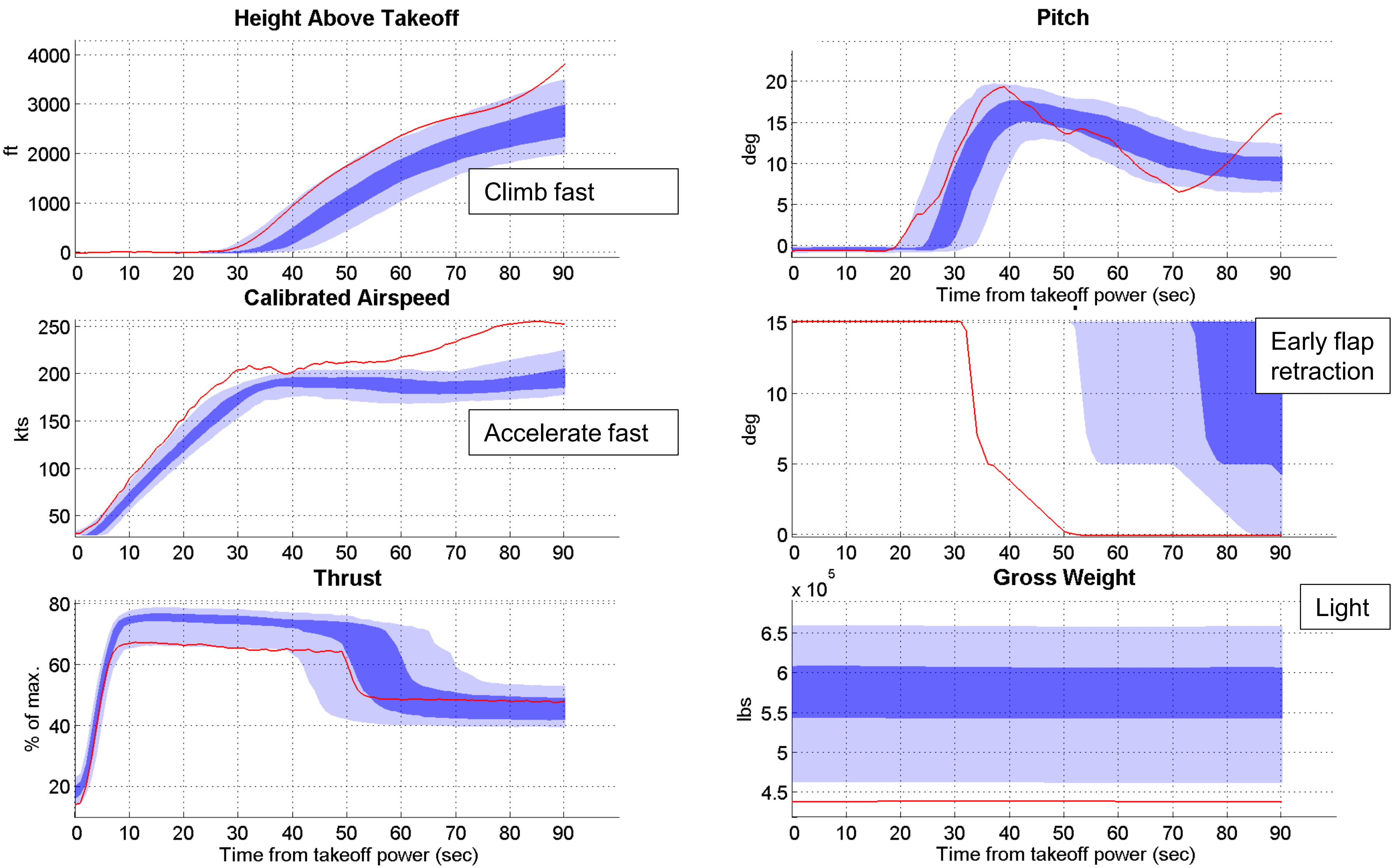


Anomalous Flights Identified in Takeoff Phase

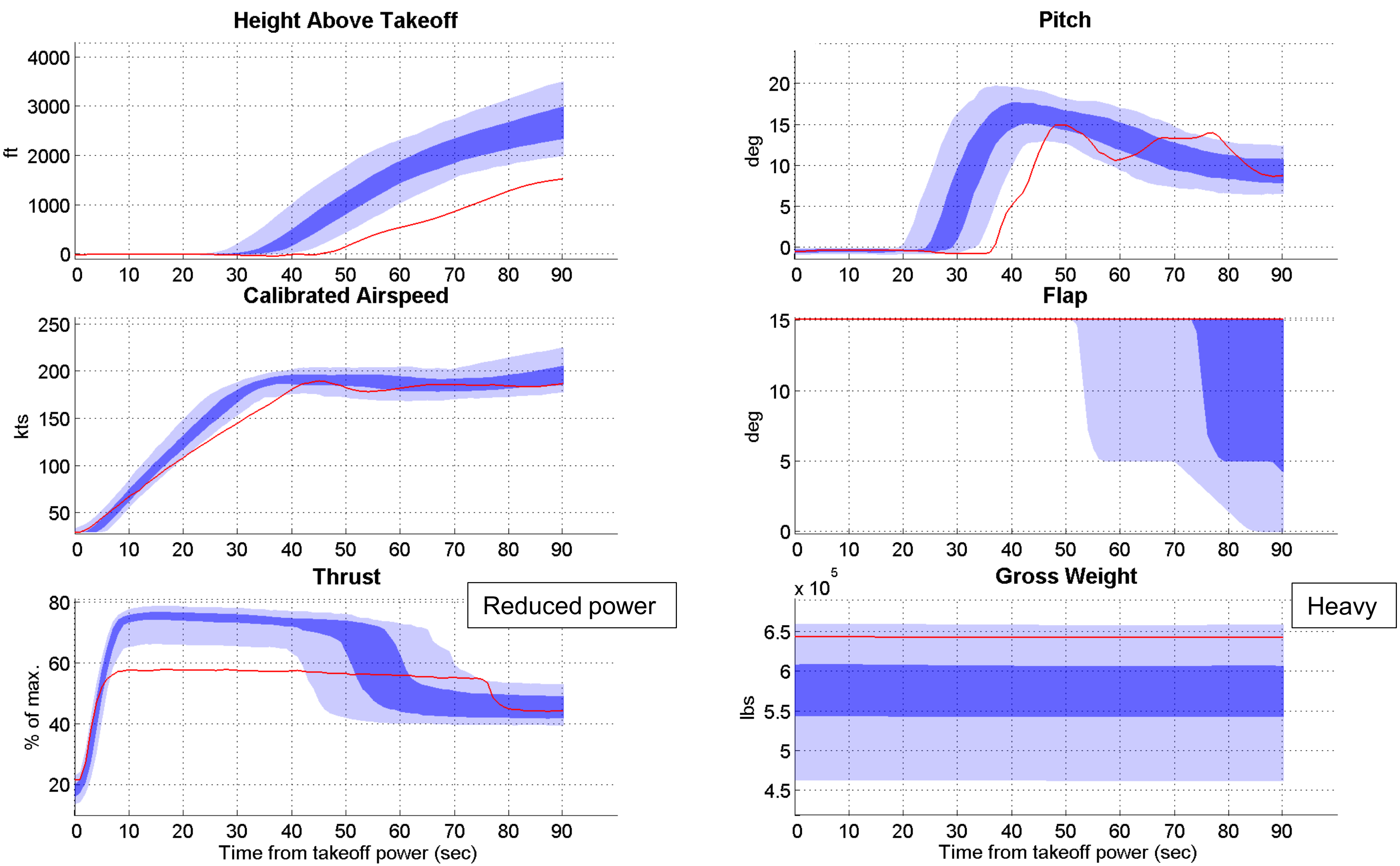
	Flight ID	1%	3%	5%	Abnormal Behaviors	Type
✓	370715	x	x	x	Accelerate fast, climb fast, light	High power takeoff
	380219	x	x	x	Early rotation, crosswind	High power takeoff
	371045	x	x	x	Reduced takeoff power, climb slow, accelerate slow	Low power takeoff
	371046	x	x	x	Reduced takeoff power, climb slow, accelerate slow	Low power takeoff
	377862		x	x	Early rotation, climb out high and fast, light	High power takeoff
	385702		x	x	Climb out high, high normal load, high pitch attitude	High power takeoff
	378692		x	x	Extended period of applying takeoff power	Other unusual operation
	386369		x	x	Early turn after takeoff	Other unusual operation
	370723		x	x	Spoiler raise, strong wind	Other unusual operation
	380217			x	Early rotation, early turn	High power takeoff
	383285			x	Early rotation, early turn, light	High power takeoff
	384110			x	Climb out high, early turn, light	High power takeoff
	385160			x	Climb out high, high pitch rotation, light	High power takeoff
✓	379636			x	Reduced takeoff power, climb slow, accelerate slow	Low power takeoff
	369755			x	Reduced takeoff power, climb slow, accelerate slow	Low power takeoff
	385444			x	Reduced takeoff power, climb slow, accelerate slow	Low power takeoff
	370019			x	Reduced takeoff power, climb slow, accelerate slow	Low power takeoff
	368486			x	Reduced takeoff power, climb slow, accelerate slow	Low power takeoff
	368487			x	Reduced takeoff power, climb slow, accelerate slow	Low power takeoff
	372209			x	Start with reduced power then switch to normal power, climb slow, accelerate slow	Low power takeoff
✓	373921			x	Double-rotation, early turn	Other unusual operation
	369204			x	Excessive reduction of power after takeoff	Other unusual operation

- **Types of abnormal behaviors**
  - High power takeoff, low power takeoff, other unusual operation
- **Example anomalous flights of each identified type are shown in the following slides**
  - Example flights are indicated by ✓

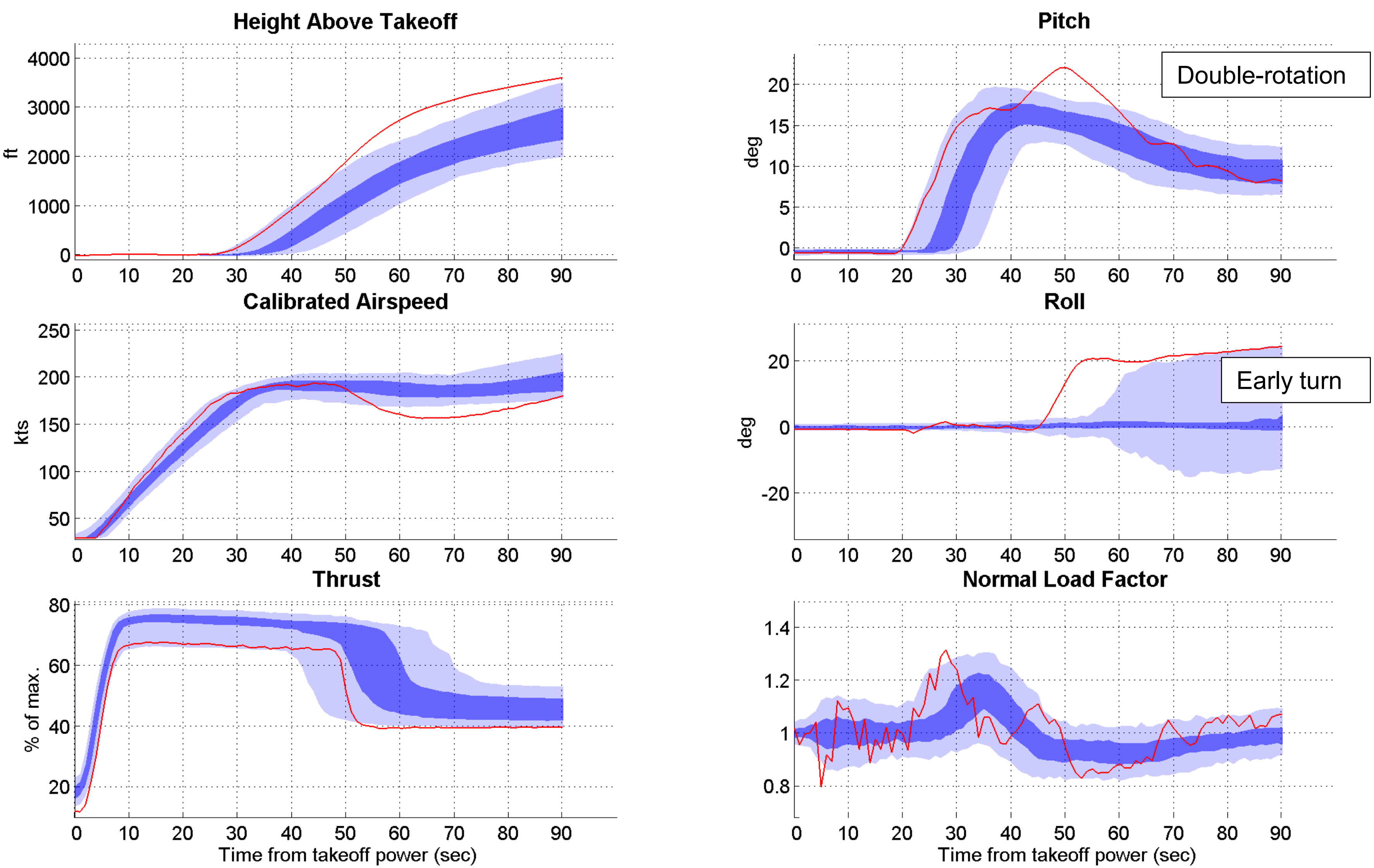
Anomalous Flight: 370715 – Light & Fast Takeoff



Anomalous Flight: 379636 – Low Power Takeoff



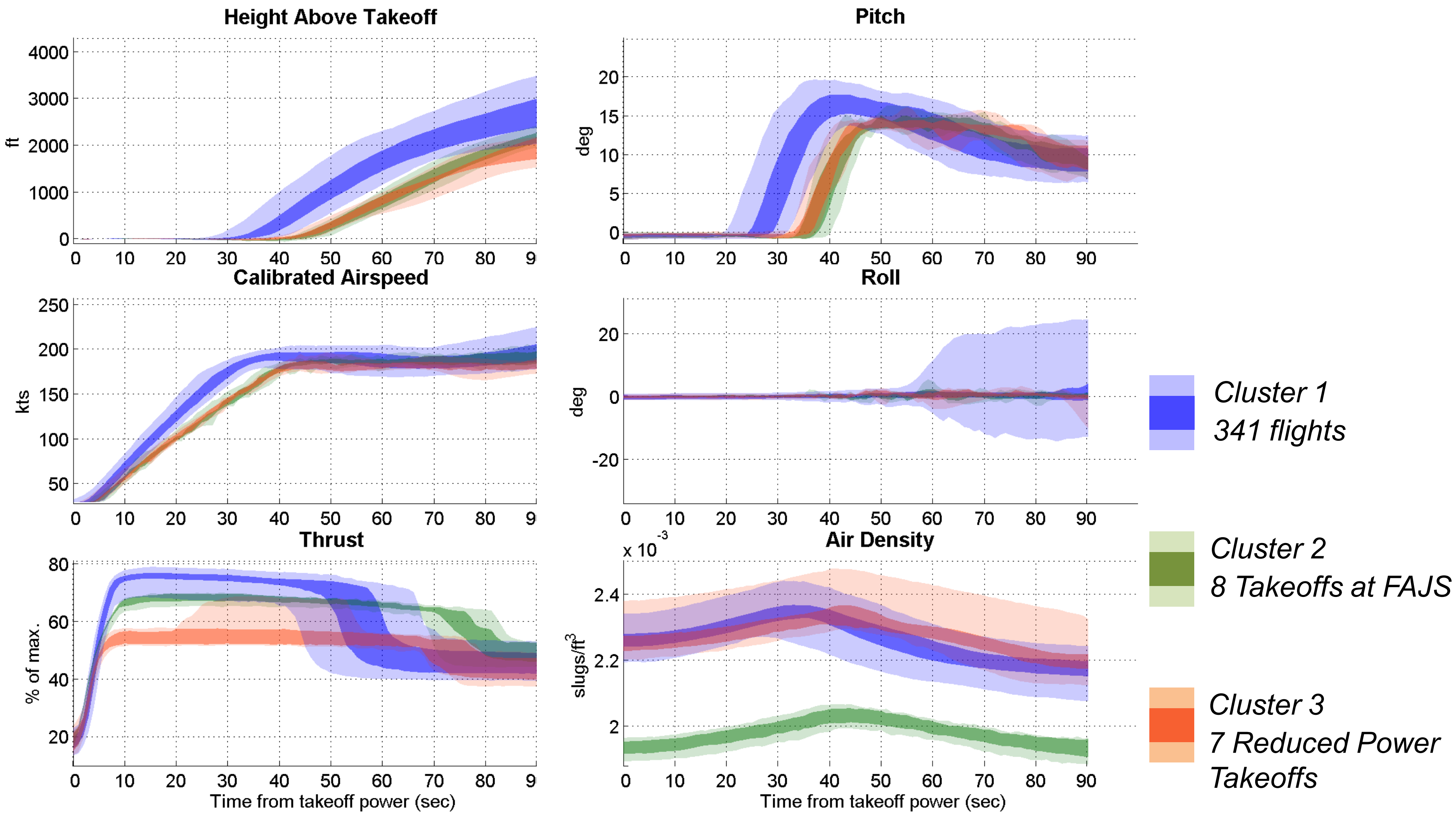
Anomalous Flight: 373921 – Double-Rotation



Multiple Nominal Clusters

- **Multiple clusters were identified in the takeoff phase for B777**
  - Cluster 1 - Most common flights
    - ✓ 341 flights
  - Cluster 2 - Takeoffs from a specific high altitude airport, OR Tambo International Airport (ICAO: FAJS), near the city of Johannesburg, South Africa
    - ✓ 8 flights
  - Cluster 3 - Takeoffs with reduced takeoff power setting
    - ✓ 7 flights
- ***Multiple nominal data patterns can be classified by cluster analysis***

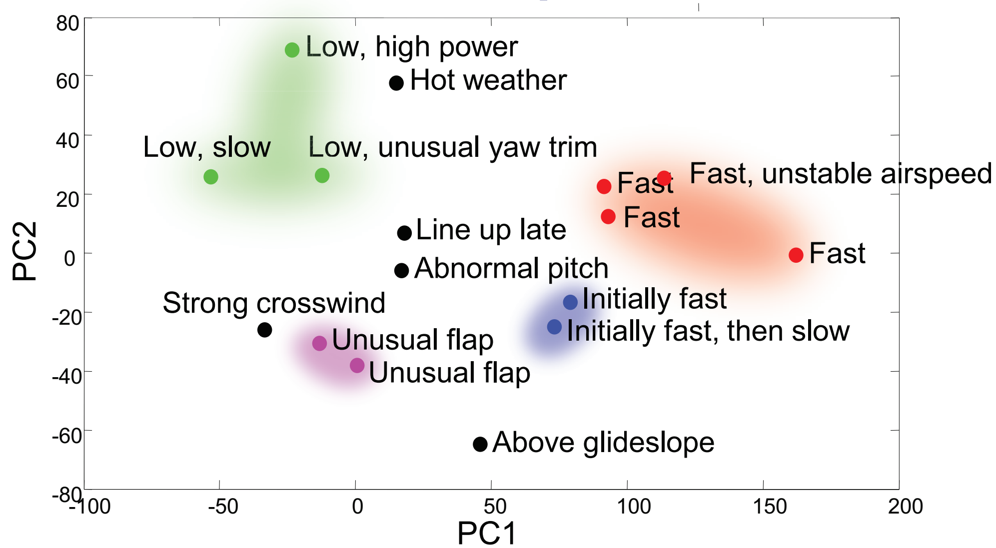
Multiple Nominal Clusters



Practical Issue

- **How many outlier flights should be identified?**
  - 3400 flights per day at a major airline (e.g. AA, Southwest)
  - 1% outliers → 30+ flights need to be reviewed per day
  - Tradeoff between workload of reviewing and chance of identifying emerging risks

- **Evaluating whether abnormal types (repetitive abnormal cases) can be identified in the cluster space**



Anomalous flights (identified in B777 approach) plotted by the first two principal components

Conclusions

- **Developed a method to detect anomalous flights using cluster analysis**
  - Advantages
    - ✓ Minimum prior knowledge of data required
    - ✓ Identify multiple nominal data patterns
  - Limitation
    - ✓ Current transformation method is only applicable to flight phases start or end with a specific event
- **Applied the proposed method on a dataset of B777 flights for takeoff phase and approach phase**
- **Initial evaluation indicates that cluster analysis is a promising approach for anomaly detection in FDR/QAR data**

Next Steps

- **Method testing**
  - Apply the method on a larger set of airline data
- **Method evaluation**
  - Challenge: ground truth not available
  - Evaluate results with domain experts
  - Cross-validate by comparing with other methods
    - ✓ Candidate method: NASA MKAD method (Das et al. 2010*)
- **Method development**
  - Auto-classification of repetitive anomalous cases
  - Tools to support expert review of anomalous flights
    - ✓ Specifically identify anomalies in time and flight parameters when identified anomalous flights are reviewed

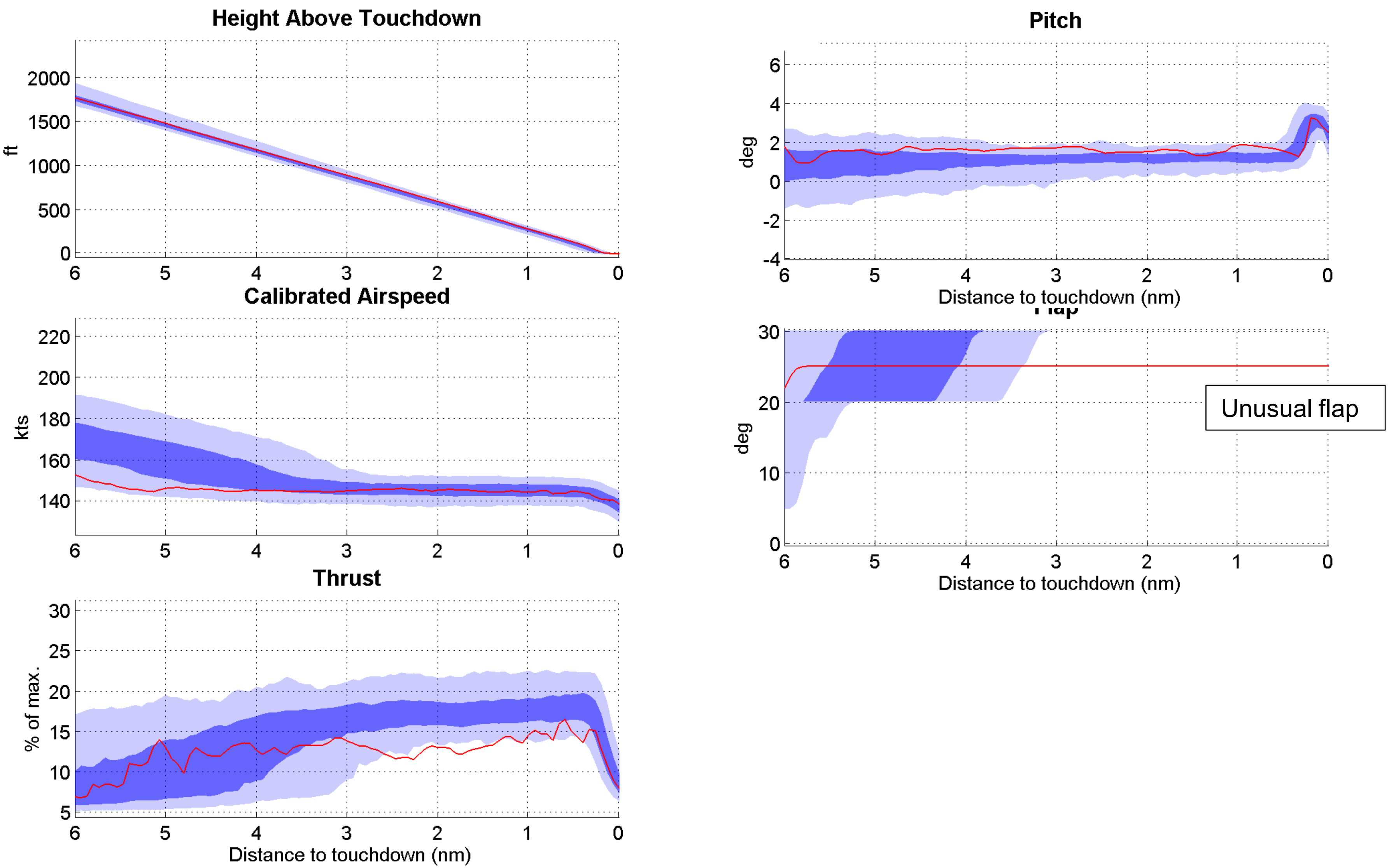
* Das, S., Matthews, B. L., Srivastava, A. N., & Oza, N. C. (2010). Multiple kernel learning for heterogeneous anomaly detection: algorithm and aviation safety case study. *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 47–56). ACM.

Thank you!

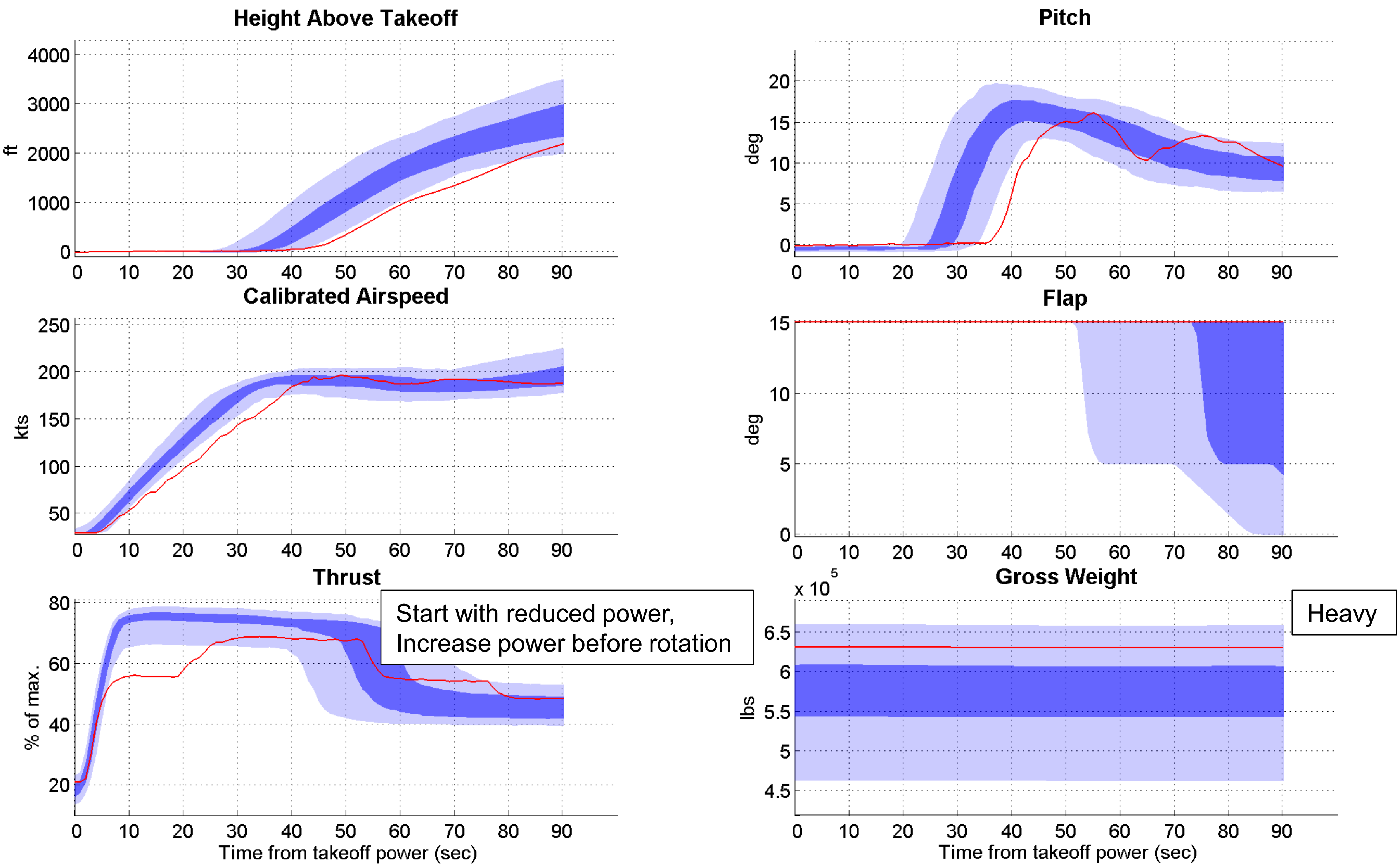
Questions?

Backup Slides

Anomalous Flight: 383285 – Unusual Flap Setting



Anomalous Flight: 372209 – Power Change Before Rotation



Sensitivity Analysis of DBSCAN Parameters

Sensitivity Analysis of ε and *MinPts on B777 Takeoff*