

Application of Machine Learning Techniques to Aviation Operations: Promises and Challenges

Dr. Banavar Sridhar

Universities Space Research Association (USRA)

@NASA Ames Research Center

Moffett Field, CA 94035

Ecole Nationale de l'Aviation Civile (ENAC)

Toulouse, France

15-18 October 2019

What is this talk about?

- Increasing interest in applying Machine Learning Techniques (MLT) to solve problems in Aviation Operations (AO)
- Review simulation and analysis methods in AO
- Promises and challenges of applying MLT to AO problems
- Compare physics-based modeling and data driven Modeling using examples from recent literature
- Concluding remarks

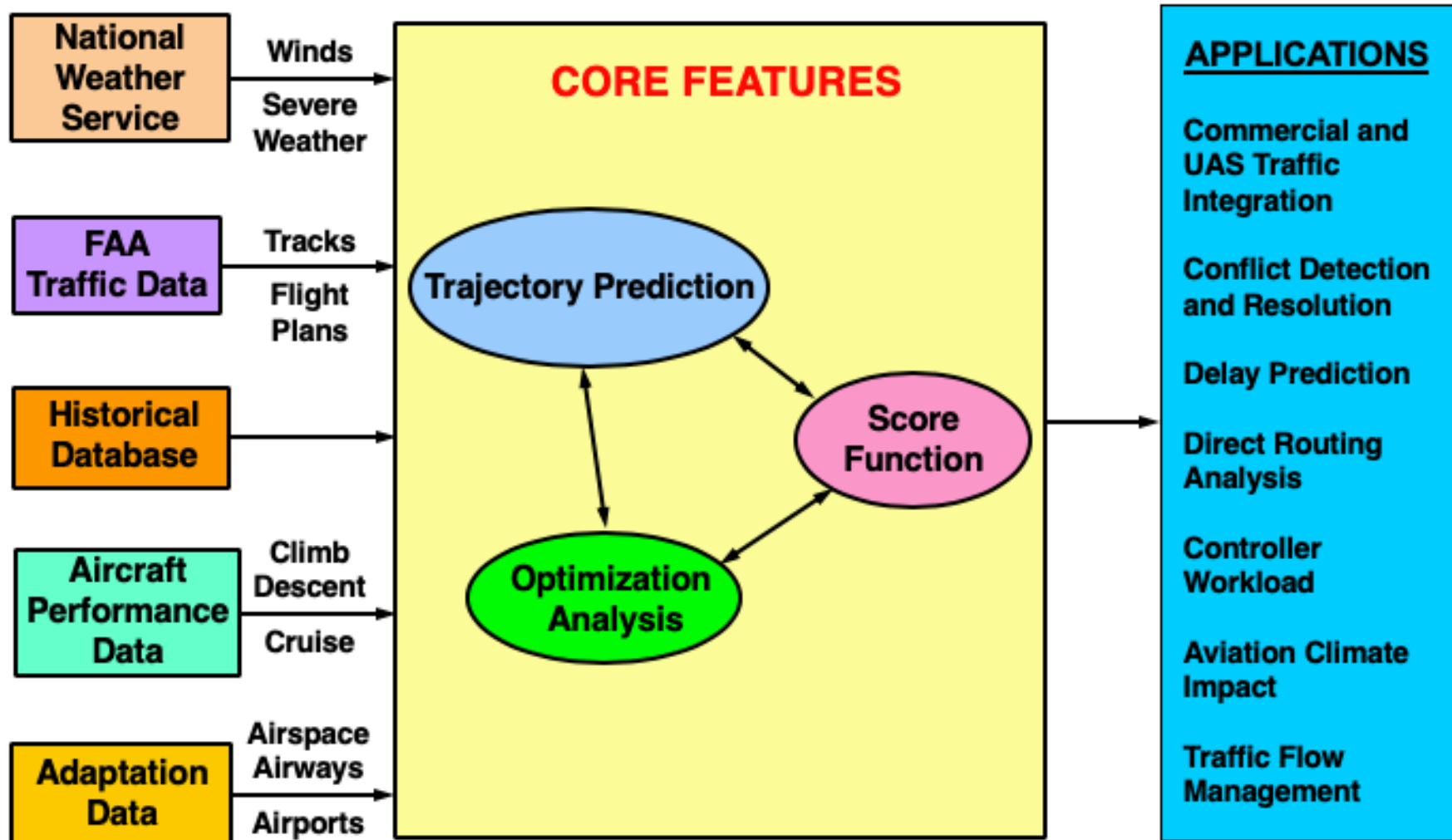
Outline

- Simulation and analysis
- Data sources
- MLT
- Applications
 - Detailed example
 - Different areas
- Conclusions

Simulation and Analysis

- Problems in AO range widely in spatio-temporal scale
 - Conflict detection involving two aircraft (local, seconds)
 - Controlling traffic in a sector (many aircraft, minutes)
 - Traffic Flow Management (Large number of aircraft, hours)
 - Impact on climate (Global and several decades)
- Modeling approach
 - Task
 - Data/Information
 - Problem formulation
 - Types of models
 - Criterion for success

Air Traffic Simulation Model



Problem Formulation

- Most problems in Air Traffic Management (ATM) can be formulated as

$$\frac{dx}{dt} = f(x, u, w, \theta)$$

$$y = g(x, u, w, \theta)$$

x (state), u (control), w (disturbance/uncertainty), y (output)

- Select u such that y is close to y_d by minimizing

$$\text{Min}_u \int_{t_0}^{t_f} (y_d - y)^2 + u^2 dt$$

$$\text{Min}_u \sum_{t_0}^{t_f} (y_d - y)^2 + u^2 dt$$

- Prediction: Given $y(t)$, for $t \leq 0$, find $y(t)$ for $t > 0$
- In classification problem, it may be necessary to divide y into several groups ($y_1, y_2, y_3, \dots, y_p$)
- Data-driven models derive *f* and *g* using data

Systems, Feedback and Neuro-Dynamic Programming

- Markov Chains

$$x(k+1) = A(k)x(k)$$

- Transition from current state $x(k)$ to new state $x(k+1)$ depends on the transition probability matrix $A(k)$
- Transition cost (**reward**) to go from $x(k)$ to $x(k+1)$: $c(k,u,k+1)$
- Multi-stage optimization: Costs satisfy Bellman's equations

$$J^*(k) = \underset{u}{\text{Min}} E[c(k, u, k + 1) + J^*(k + 1)|k, u] \text{ for all } k$$

- Approximate the cost-to-go by

$$J^*(k + 1) = J^{\sim}(k + 1, r)$$

- Minimization provides the feedback (**agent**) policy
- Neural Network provides the approximation architecture and calculation of r to minimize the error between J^* and J^{\sim}

Characteristics of Physics-based models

- Choice of state variables and their relationship to the physical quantities, dimensionality
- Model reduction
- Low order unbiased minimum variance models
- Feature Selection

Metrics for Evaluation

- Maximum Absolute Square Error (MASE), Root Mean Squared Error (RMSE)

	Actual: Positive	Actual: Negative	
Predicted: YES	TP	FP	TPR=TP/(TP+FN) TNR=TN/(TN+FP)
Predicted: NO	FN	TN	FNR=FN/(FN+NP) FPR=FP/(FP+TN)

- True Positive Rate (TPR), True Negative Rate (TNR), False Negative Rate (FNR) and False Positive Rate (FPR)
- Precision: $TP/(TP+FP)$, Recall: $TP/(TP+FN)$
- F_1 -score: Harmonic mean between precision and recall
$$F1 = 2 (\text{precision} \times \text{recall}) / (\text{precision} + \text{recall})$$
- Cross-correlation
- Receiver Operating Characteristics (ROC) curve

Data Sources: FAA

- **Operations System Network (OPSNET)**
 - From 1990; 45 airports; Different types of daily delays
- **Aviation System Performance Metrics (ASPM)**
 - Available from 2000; 77 airports in US; Every 15 minutes; provides the airport specific data, runway configuration and the local meteorological conditions at each airport. Hourly values of wind speed, visibility, ceiling, Instrument Meteorological Conditions (IMC), scheduled arrivals and departures, Airport hourly delays and airport arrival rates (AAR)
- **Terminal Area Forecast (TAF)**
 - Database used by the FAA for planning purposes and covers airports in the US; Historical (1990-2017) and forecast data (2018-2045) for enplanements, airport operations, TRACON operations, and based aircraft
 - Covers 264 FAA towered airports, 254 Federal contract tower airports, 30 radar approach control facilities and 2850 non-FAA airports.
- **Meteorological Aviation Weather Report (METAR)**
 - airport identifier, time of observation, wind, visibility, runway visual range, present weather phenomena, sky conditions, temperature, dew point, and altimeter setting.

Data Sources: Bureau of Transportation Statistics (BTS)

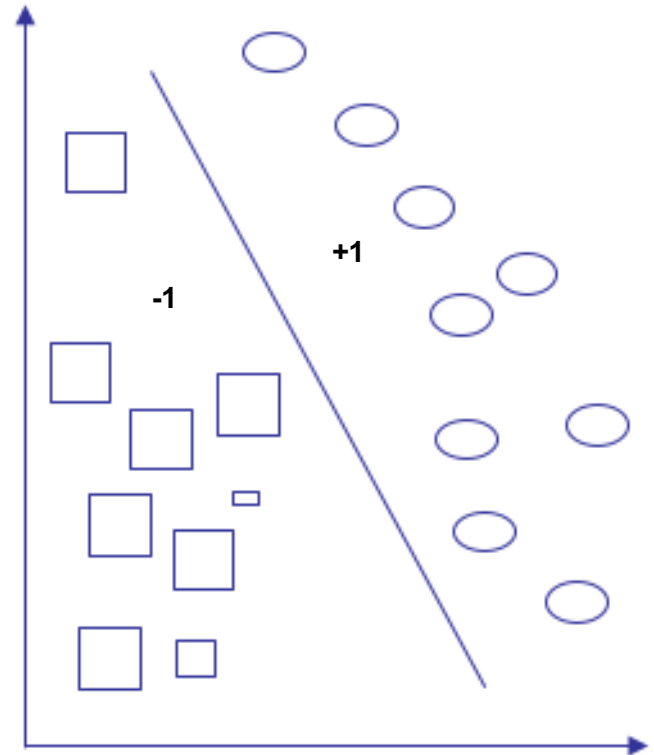
- Data provided by air carriers that have more than 0.5 percent of total domestic scheduled-service passenger revenue
 - Airlines report causes of delays in five broad categories:
 - (a) Air Carrier Delays: cancellation or delay due to circumstances within the airline's control (e.g. aircraft maintenance or crew problems)
 - (b) Extreme Weather conditions such as tornado, blizzard or hurricane that delays or prevents the operation of a flight such as tornado, blizzard or hurricane
 - (c) National Aviation System (NAS): Delays and cancellations attributable to the national aviation system to manage traffic safely during non-extreme weather conditions, airport operations and heavy traffic volume
 - (d) Late-arriving aircraft: Flight delayed due to aircraft arriving late from a previous flight
 - (e) Security: Delays or cancellations for maintaining security of aviation such as caused by evacuation of a terminal or inoperative screening equipment.

Machine Learning Techniques (MLT)

- Major concepts in MLT originate from Pattern Recognition, Computer Vision, Text Processing and Voice Recognition (sparse or repetitive data)
- Define terminology and characteristics to provide background to review applications
- Techniques selected based on the frequency of application in ATM
- Methods
 - Classification
 - Support Vector Machine (SVM)
 - Decision Trees
 - Neural Networks
 - Reinforcement Learning (RL)

Support Vector Machine (SVM)

- SVM classifies data using Linear Discriminant Function (LDF) to minimize the error in classification of training samples
 - Computational simplicity
 - Gradient procedures used to speed up computation of the hyperplane
 - Used for both regression and classification
 - Robust performance under limited, sparse, noisy data
- SVM performs classification of non-linear decision functions by transforming inputs using kernel functions
 - Gaussian radial basis function (RBF)

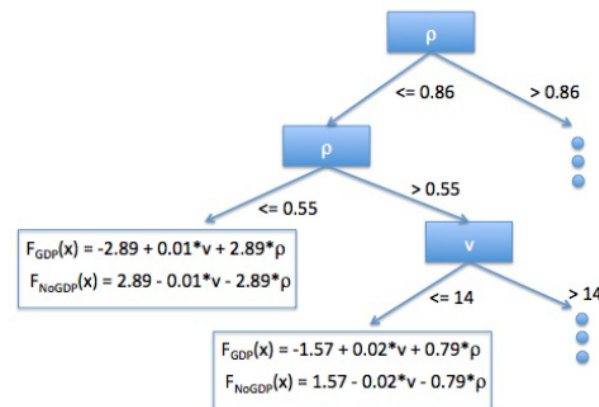


$$y = w^T x + b$$

If $w^T x_i + b \geq 0$, then $y_i = +1$
If $w^T x_i + b < 0$, then $y_i = -1$

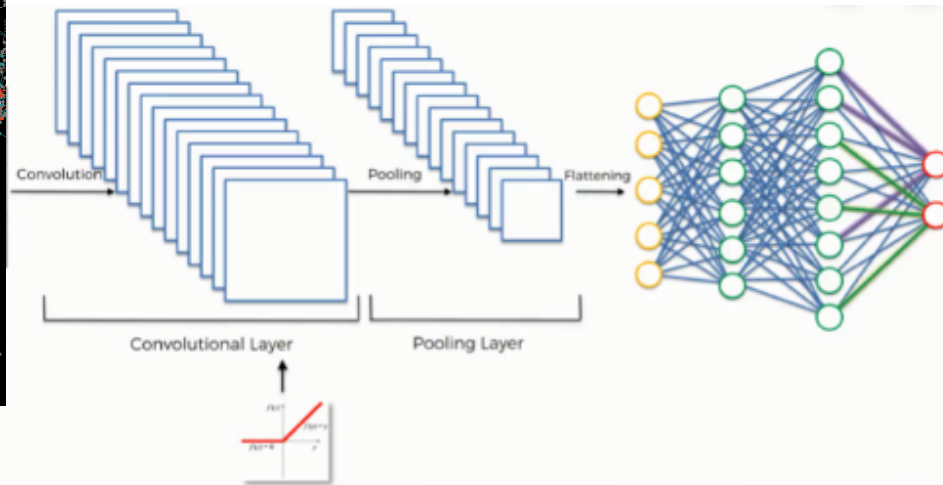
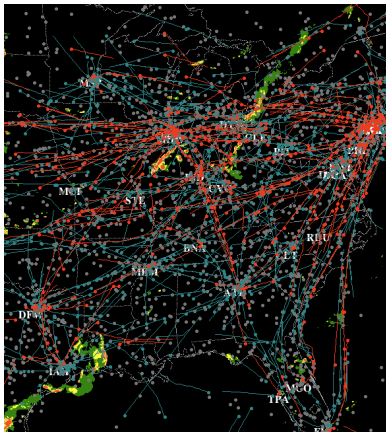
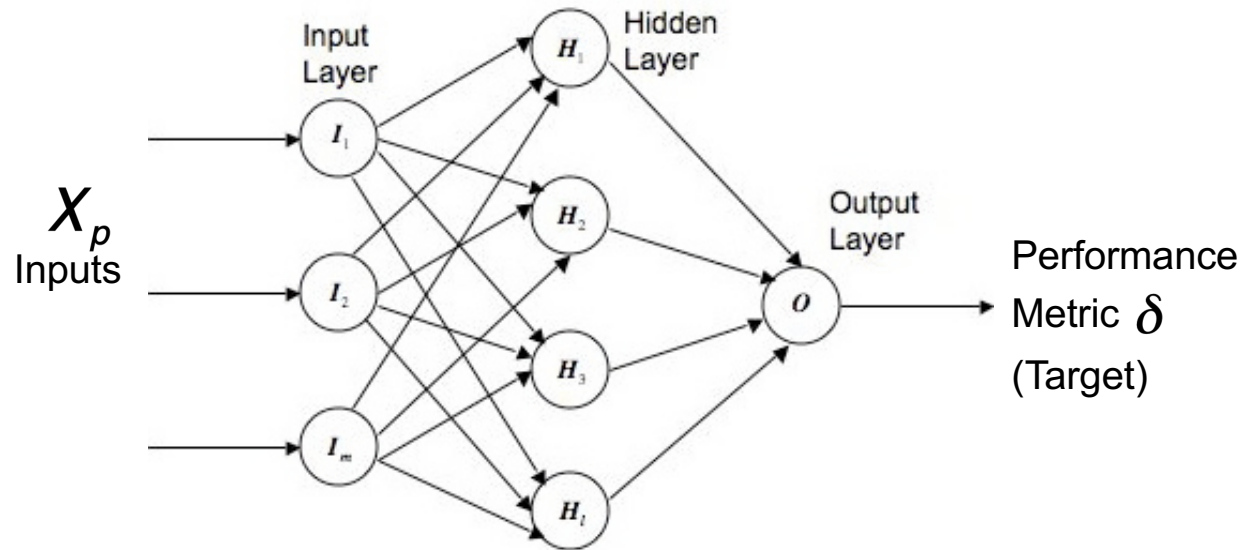
Decision Trees

- Classification and Regression Tree (CART)
 - Used both for classification and regression
 - Easy to interpret and see importance of feature based on its location
 - Sensitive to inputs



- Ensemble decision trees: Random Forest
 - Build several classifiers (trees) by sampling data (forest) and combine the outcome of samples to produce final result

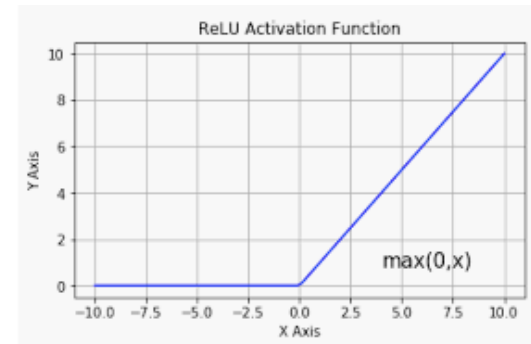
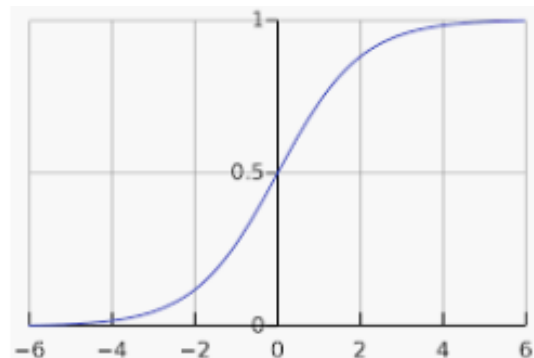
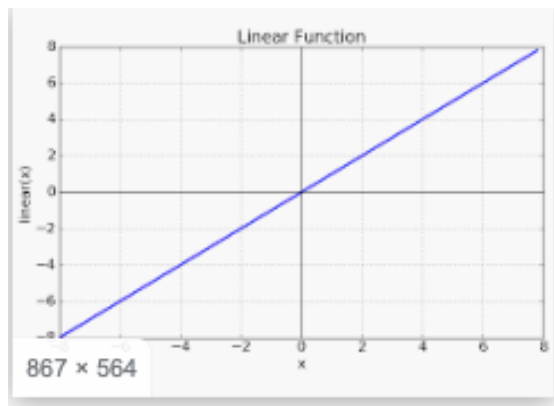
Neural Networks



- Good general purpose modeling approach
 - Ability to model large class of input/output relations
 - Generalization capability

Neural Networks: Design Considerations

- Architecture: input layer, # of hidden layers, output layers
- Activation Functions: Identity, Sigmoid, ReLU (Rectified Linear Units)

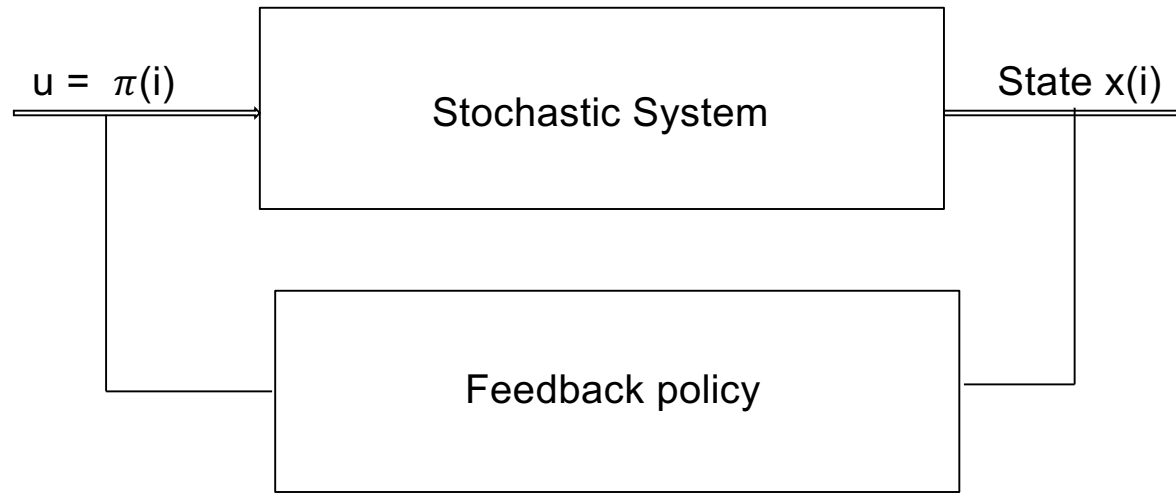


- Training: Learning rate

$$f(x + \Delta x) = f(x) + \left[\frac{\delta f}{\delta x} \right]^T * \Delta x + 0.5 * \frac{\delta^2 f}{\delta x^2} * \Delta x^2$$

$$\Delta x = -2.0 * H^{-1} * g^T$$

Reinforcement Learning



- Learning environment generates and presents scenarios to the agent
- Based on the feedback policy, agent tries to solve the problem
- Receives positive reward for successful performance and negative reward or penalty for poor performance

$$R^{n+1}(x, a) = (1 - \alpha)R^n(x, a) + \alpha[r(x, a, j) + \beta \underset{b \in A}{\text{Min}} R^n(j, b)]$$

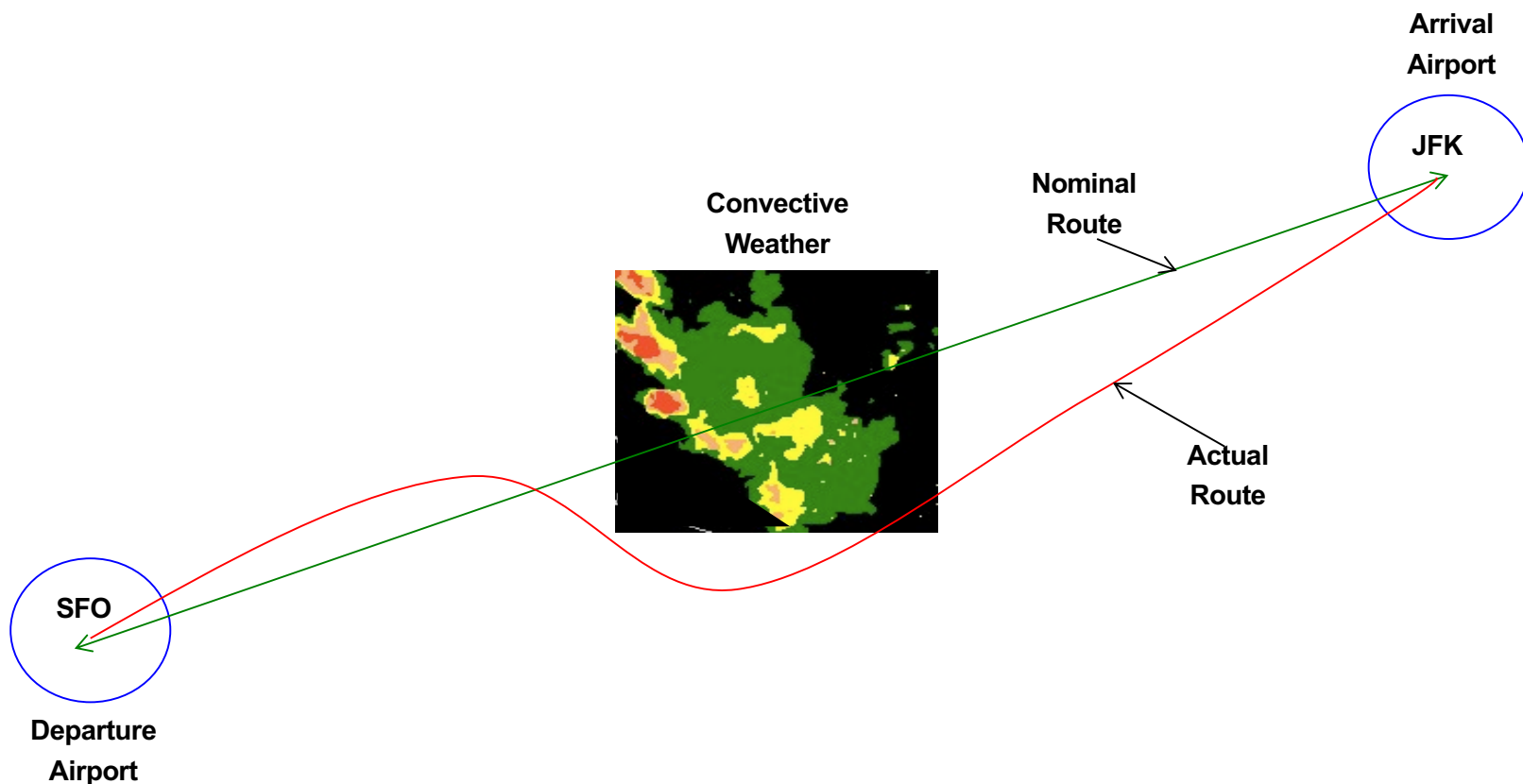
- Training objective is to maximize the reward function and the policy is consistently correct for new scenarios

Physics-based and Data-driven Models

Property	Physics-based Model	Data-driven Models
Model	Linear, Non-Linear, Dynamic, Static, Queueing	Black-Box
Interpretation	Easy to explain results in terms of physical quantities	Hard to interpret and gain trust in the system
Model-Building	Expensive and requires lot of application expertise	Availability of quantity and quality of data
Suitability	Availability of well-defined physical models	Ideal for building causal relationship between inputs and outputs when good physics-based models are non-existent or expensive to build
Feature Selection	Defined by the model and various methods to reduce dimensions (Aggregation, time and space separation)	Major issue to reduce the dimension in complex problems
Size	Various methods to determine minimal order unbiased minimal variance models	Efforts to balance over-fitting and under-fitting by cross-validation, regularization and other methods

Applications

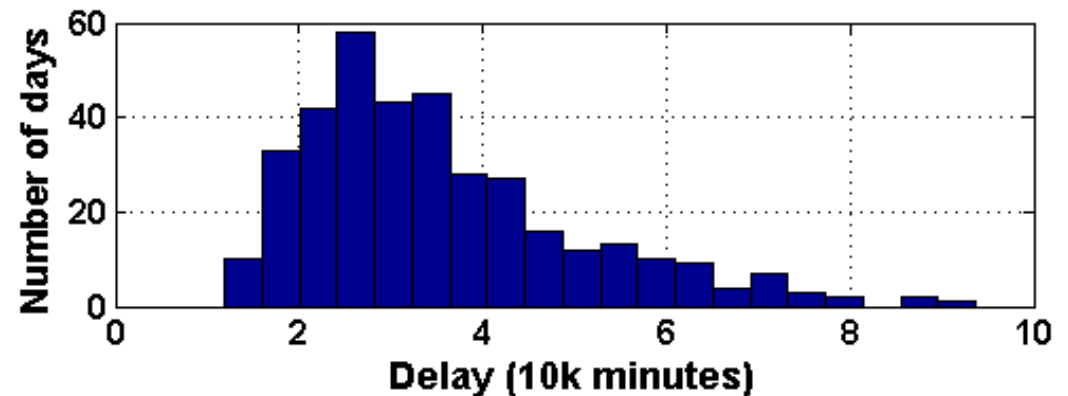
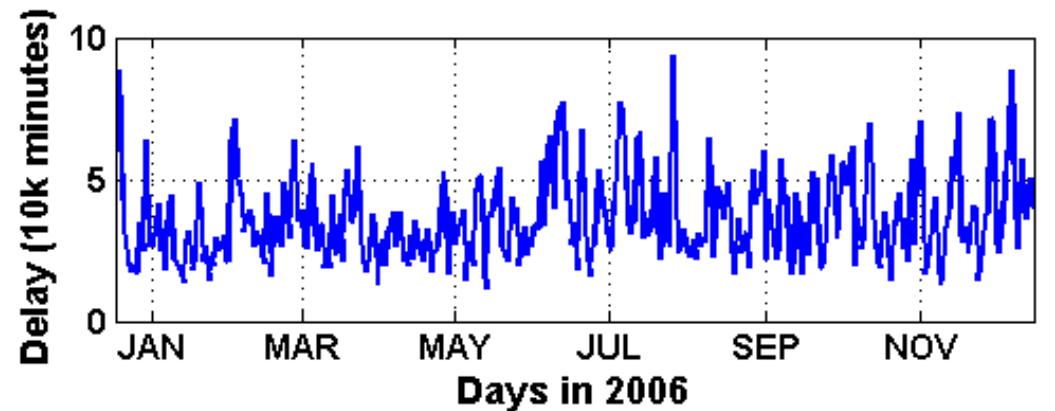
Task: Modeling flight delays and cancellations due to Weather



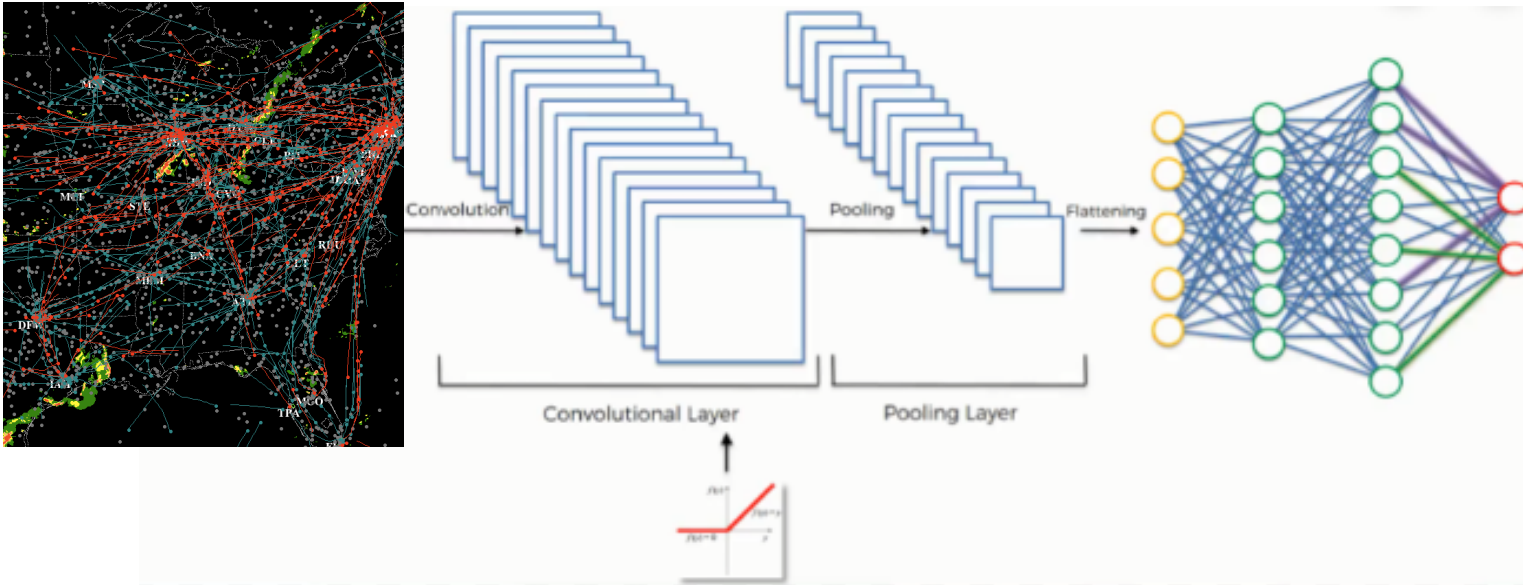
- Weather is the major cause of delay in the National Airspace System (NAS)
- Relate delay, cancellations and other NAS performance metrics to the weather conditions

Factors in modeling delay

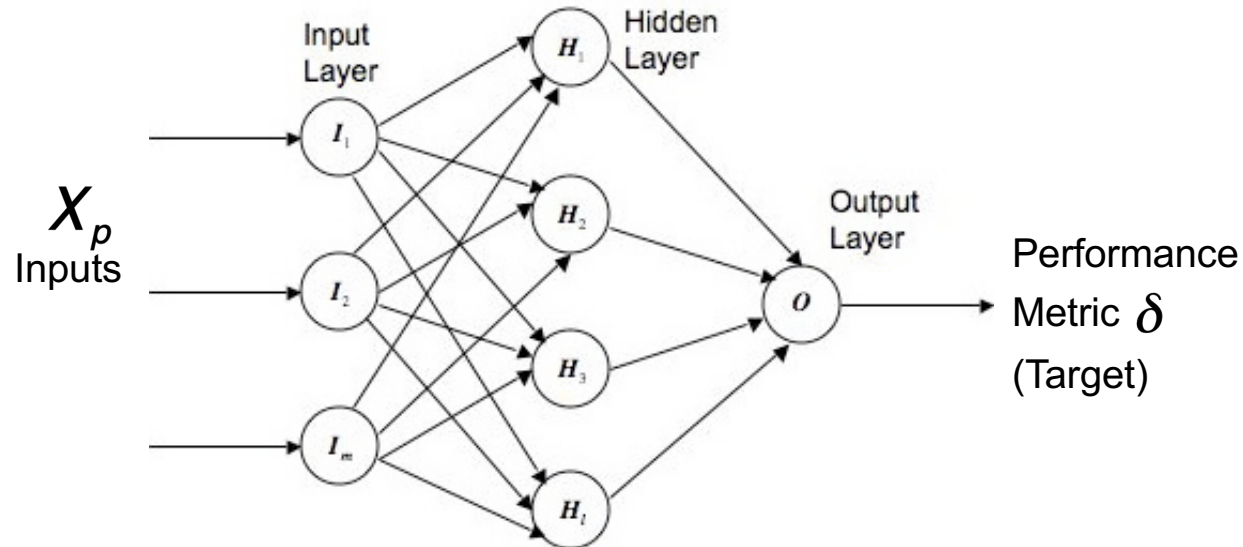
- Data
 - OPSNET, ASPM
 - Traffic Data (2004-2008)
 - Convective Weather, Wind data
- Choice of nominal traffic
- Feature selection
- Modeling approach



Neural Networks

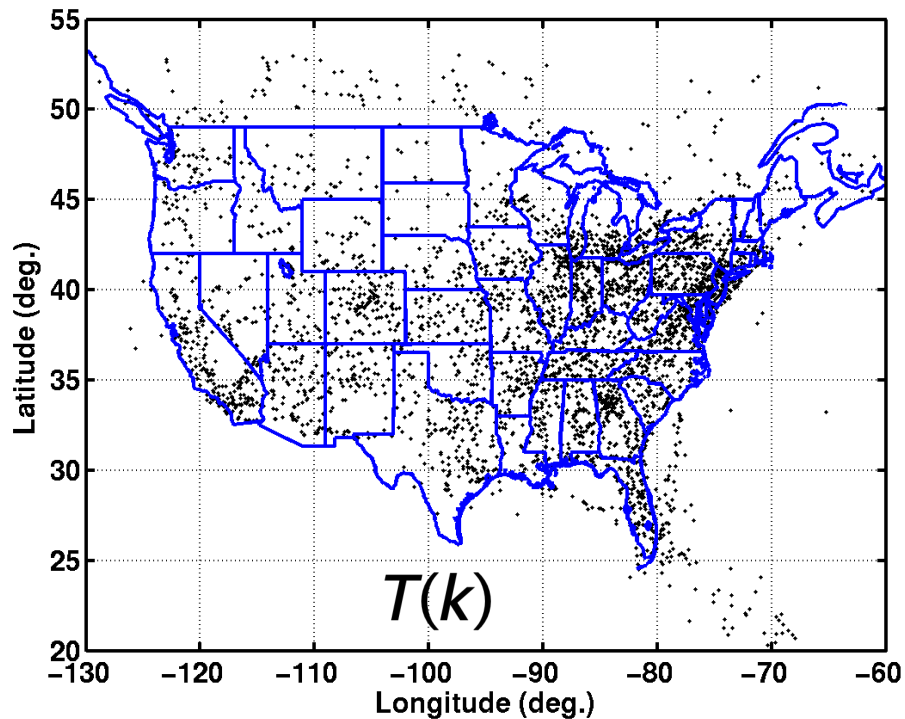


- Selection of WITI reduces the input images of traffic and convective weather and simplifies the convolutional and pooling layers

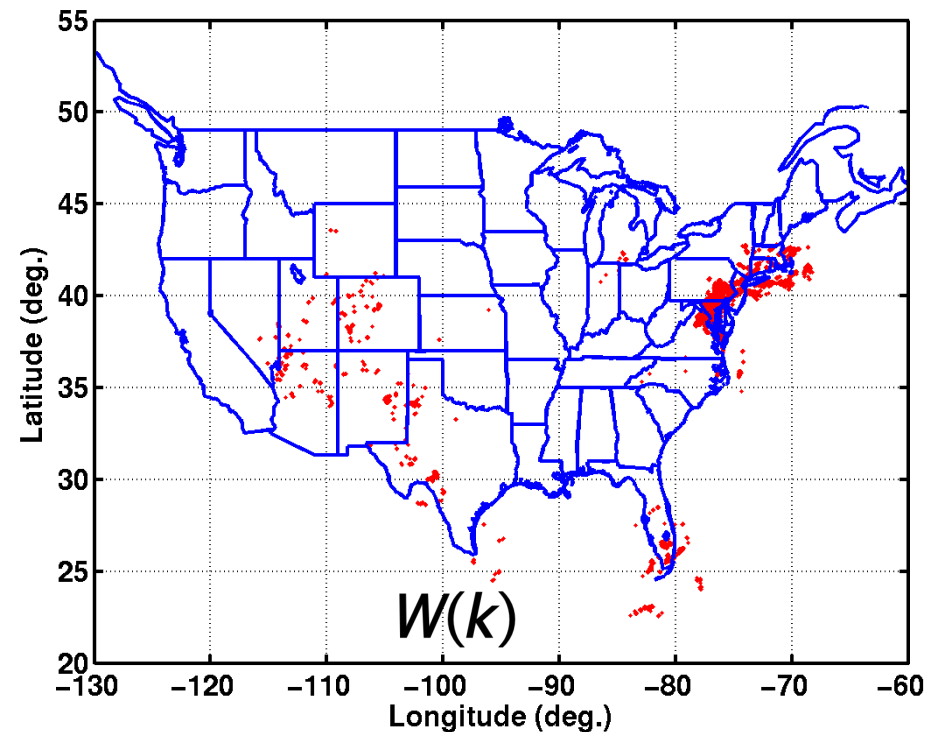


Feature Selection : Weather Impacted Traffic Index (WITI)

Aircraft positions

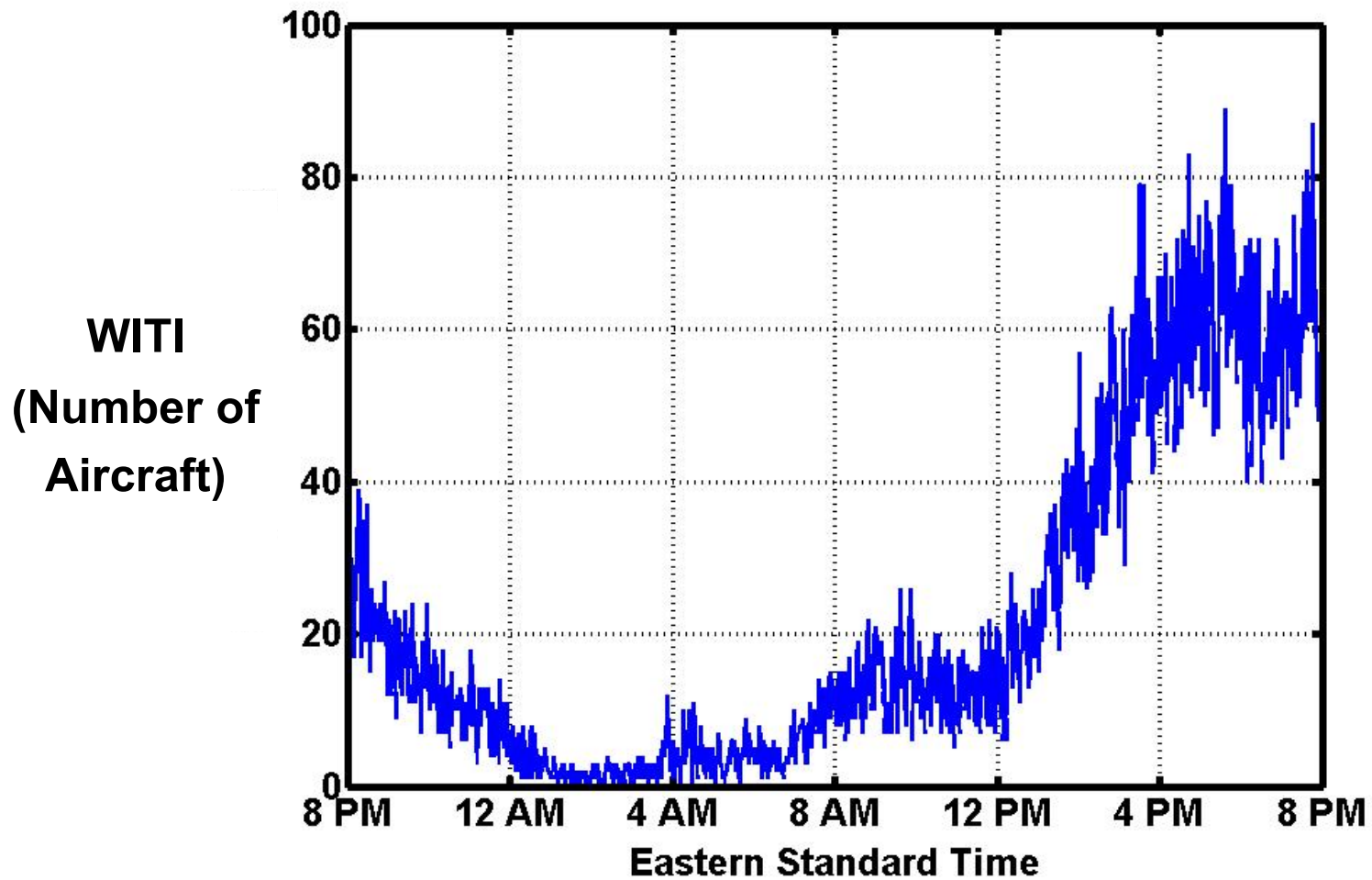


Severe weather



$$WITI(k) = \sum_{1 \leq j \leq m} \sum_{1 \leq i \leq n} T_{i,j}(k) W_{i,j}(k)$$

National WITI



Feature Selection Influenced by Data

- Number of aircraft affected by weather (X)
- Number of aircraft affected by weather in each Center or Airport (X_p)
- Performance metric: Delay or cancellation (δ)

- Models

- Linear Regression (LR) $\delta = \alpha X + \beta$

- Multiple Linear Regression (MLR) $\delta = \sum_{p=1}^{20} \alpha_p X_p + \beta_p$

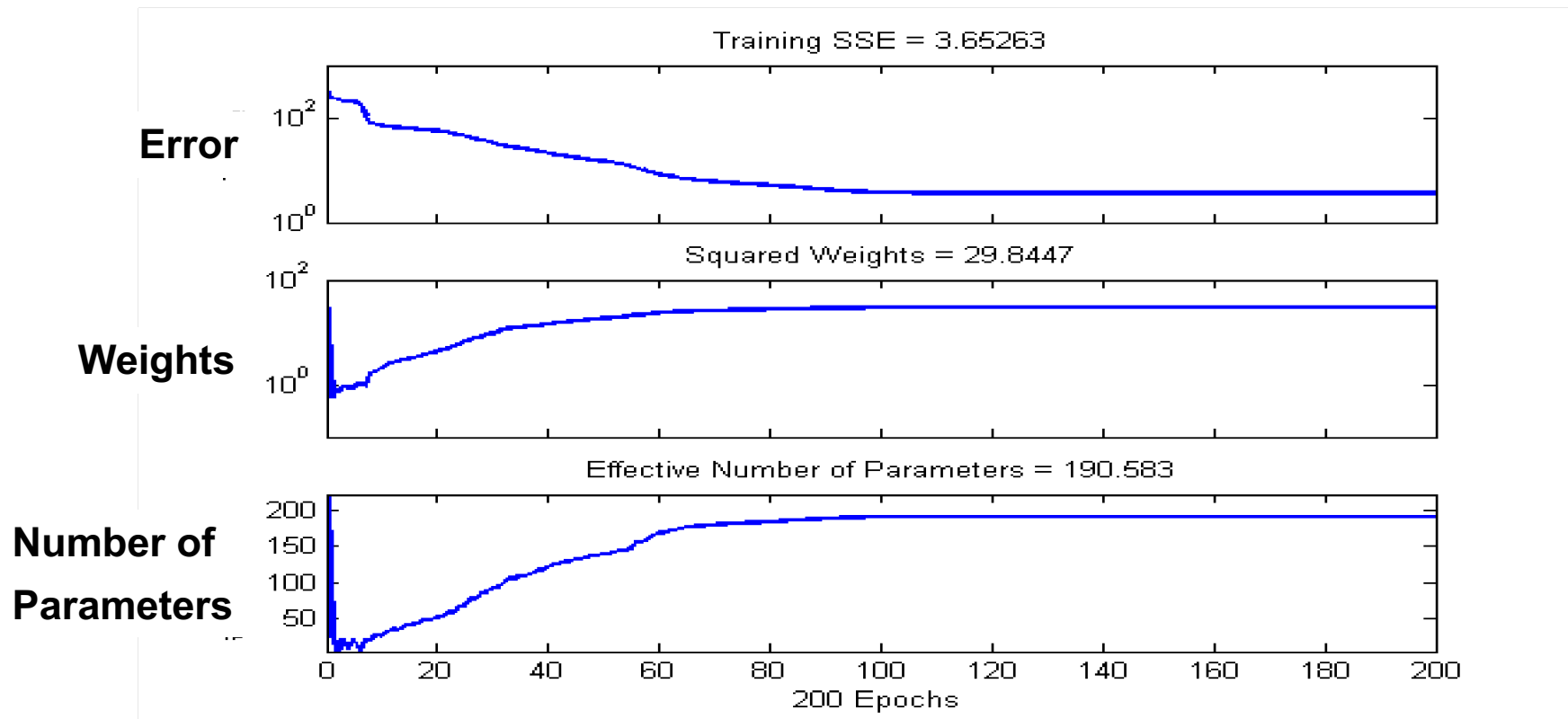
- Neural Networks $\delta = f(X_p)$

- Dynamic Models $\delta(t) = f(X_p(t-k), \dots, X_p(t-1), X_p(t), X_p(t+1), \dots, X_p(t+r))$

Neural Network Design Considerations

- Number of hidden layer
 - One was found adequate
- Number of neurons
 - Different values were used and errors for each were estimated
 - Neurons in the range $2/3$ (sum of number of inputs and outputs) provided consistent results and error
- Activation function
 - Nonlinear sigmoid function
- Data includes pairs of inputs and desired outputs
- $m(l+1)$ weights
- Weights updated using a gradient procedure until the sum of squares error (SSE) between the neural network output and the desired output is minimized
- Balance between over fitting and under fitting

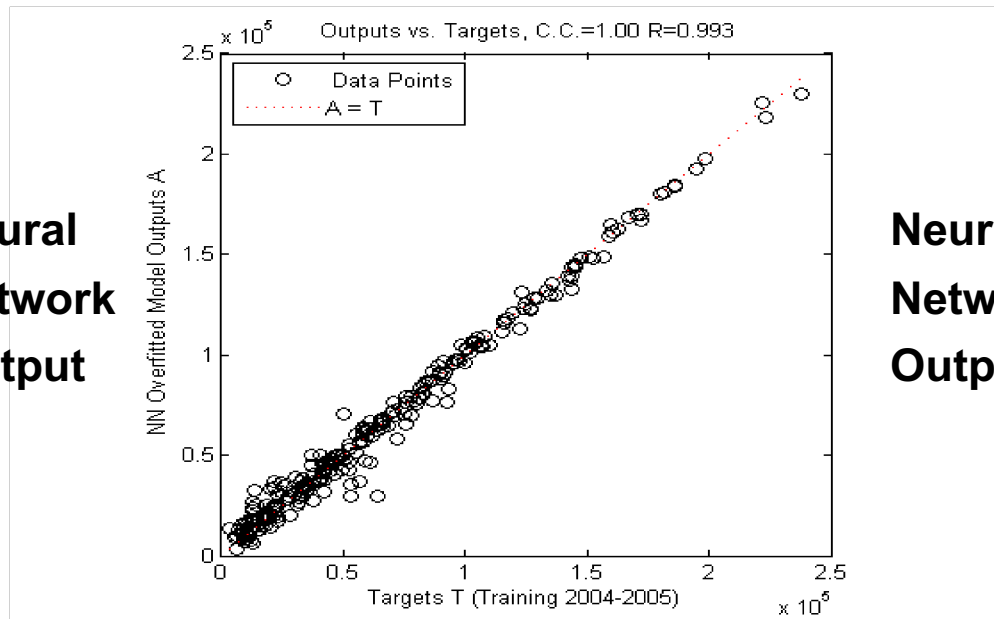
Variation of Training Error



- Result of training after 200 epochs
- Neural network represents total delay training data extremely well

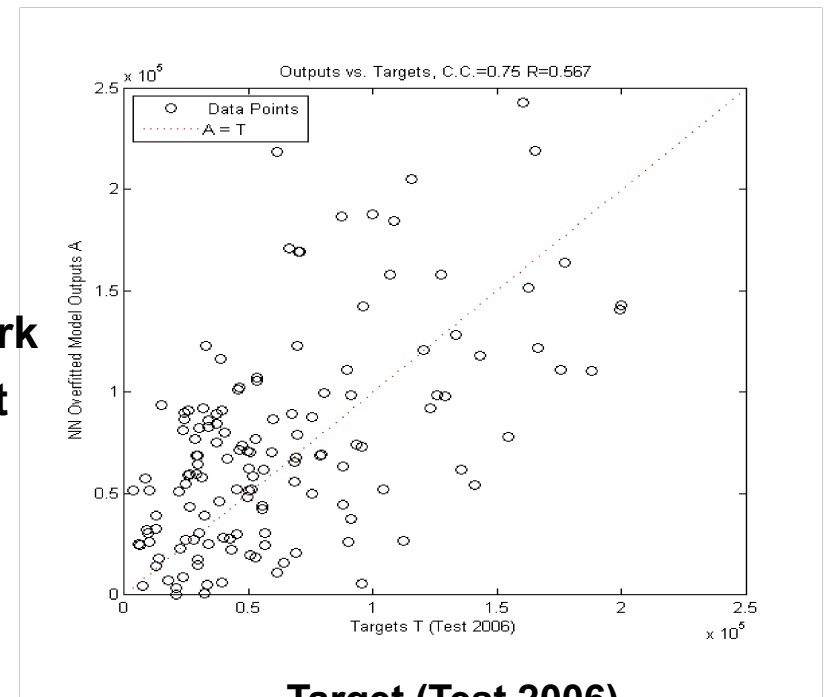
Over fitted Neural Network

Neural
Network
Output



Target (Training 2004-2005)

Neural
Network
Output



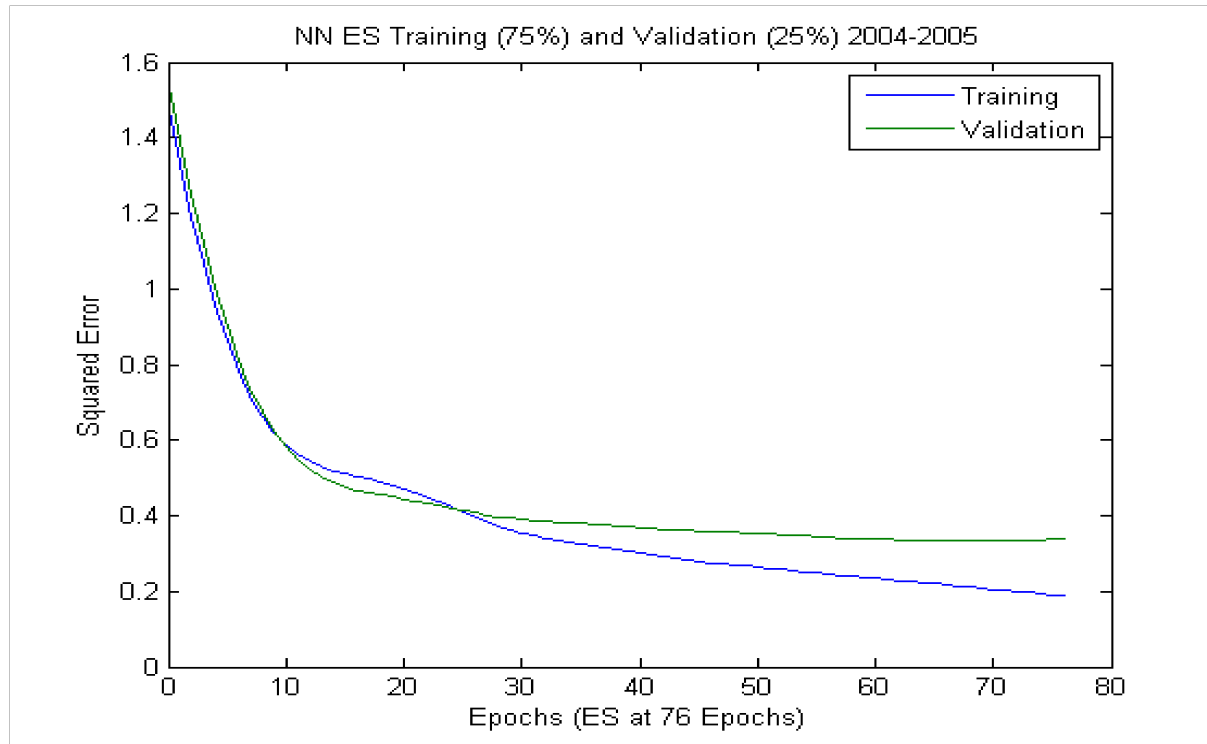
Target (Test 2006)

- Model fits the total delay training data (2004-2005) well, but does not generalize (correlation coefficient for test set is significantly less)

Methods for Good Design

- Training data should be sufficiently large and statistically representative
- Overly complex models should be avoided
- Methods to reduce complexity
 - Early Stopping (ES)
 - Principal Component Analysis (PCA)
 - Stepwise Regression (SR)
 - Bayesian Regularization (BR)

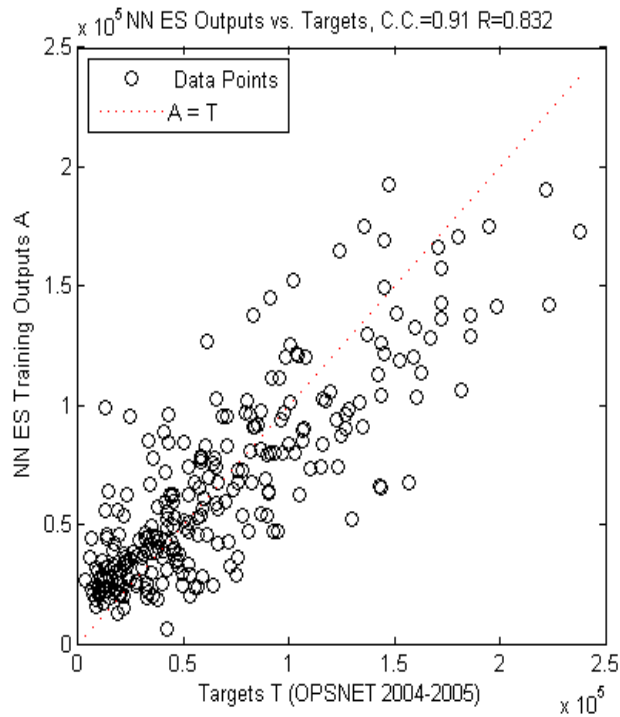
Early Stopping



- Training data (2004-2005) divided into two parts
 - Training set (80%) used to update weights
 - Validation set (20%) used for stopping criterion

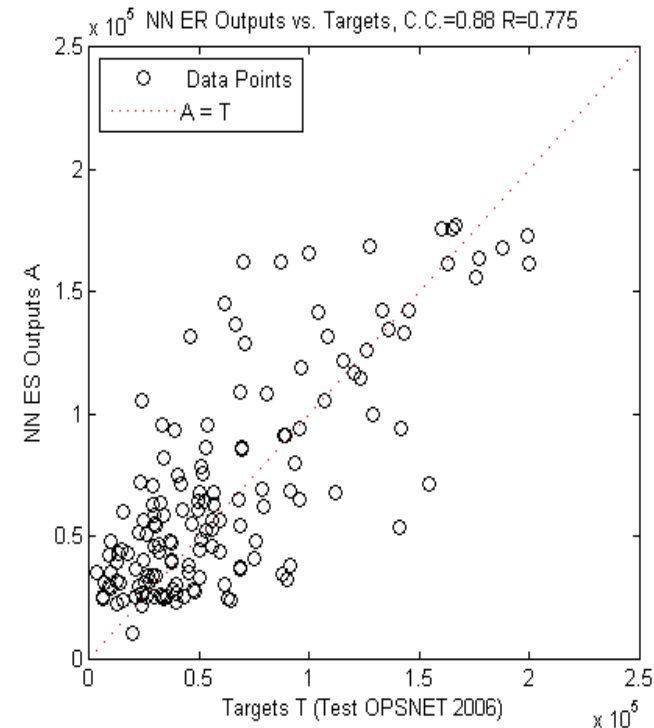
Performance with Early Stopping

**Neural
Network
Output**



Target (Training 2004-2005)

**Neural
Network
Output**



Target (Test 2006)

- 2004-2005 total delay training data with Early Stopping produces a more balanced model with better generalization capability
- N-fold cross-validation using 2004-2006 data

Computational Results

OPSNET total delay

Method	CC	RMSE	MAE
LR	0.71	32700	26600
MLR	0.77	31200	24500
BR	0.88	30000	23300
ES	0.88	30900	23200
PCA	0.88	30100	23100
SR	0.88	29600	22300

ASPM Scheduled delay

Method	CC	RMSE	MAE
LR	0.75	99200	74300
MLR	0.76	97600	72900
BR	0.88	95800	74300
ES	0.88	94200	70800
PCA	0.88	91700	68600
SR	0.87	99100	73800

Five-fold cross-validation

Method	CC	RMSE	MAE
BR	0.88	29100	22000
ES	0.88	31500	23900
PCA	0.87	30500	23300
SR	0.89	29600	22500

Five-fold cross-validation

Method	CC	RMSE	MAE
BR	0.87	93700	70300
ES	0.87	95000	72000
PCA	0.87	94900	70800
SR	0.85	96100	73000

Traffic Delay Estimation

Problem	Flight Delay and Cancellation in US
Data	FAA OPSNET and ASPM data Convective weather 2005-2008.
Method	MLR and feed-forward NN with several stopping criterion
Feature Selection	Weather Influenced Traffic Index(WITI) at the Center, National and airport level
Method of Evaluation	MAE, RMSE and Correlation Coefficient
Remarks	For all metrics and seasons at all levels NN produced slightly better results the MLR

Traffic Delay Estimation

Problem	Network Delay Estimation in US
Data	BTS data 2011-2012 hourly different types of delays between 1107 origin-destination pairs and 158 airports
Method	Markov Jump Linear System CART and NN
Feature Selection	Local delay: Links between airports Global delay: time and type of day
Method of Evaluation	MAE, RMSE and Correlation Coefficient
Remarks	NN performed well in classifying links with high delays; MLJS performed better on estimating delay on individual links. Performs varies with problem and prediction horizon

Conflict Detection and Resolution

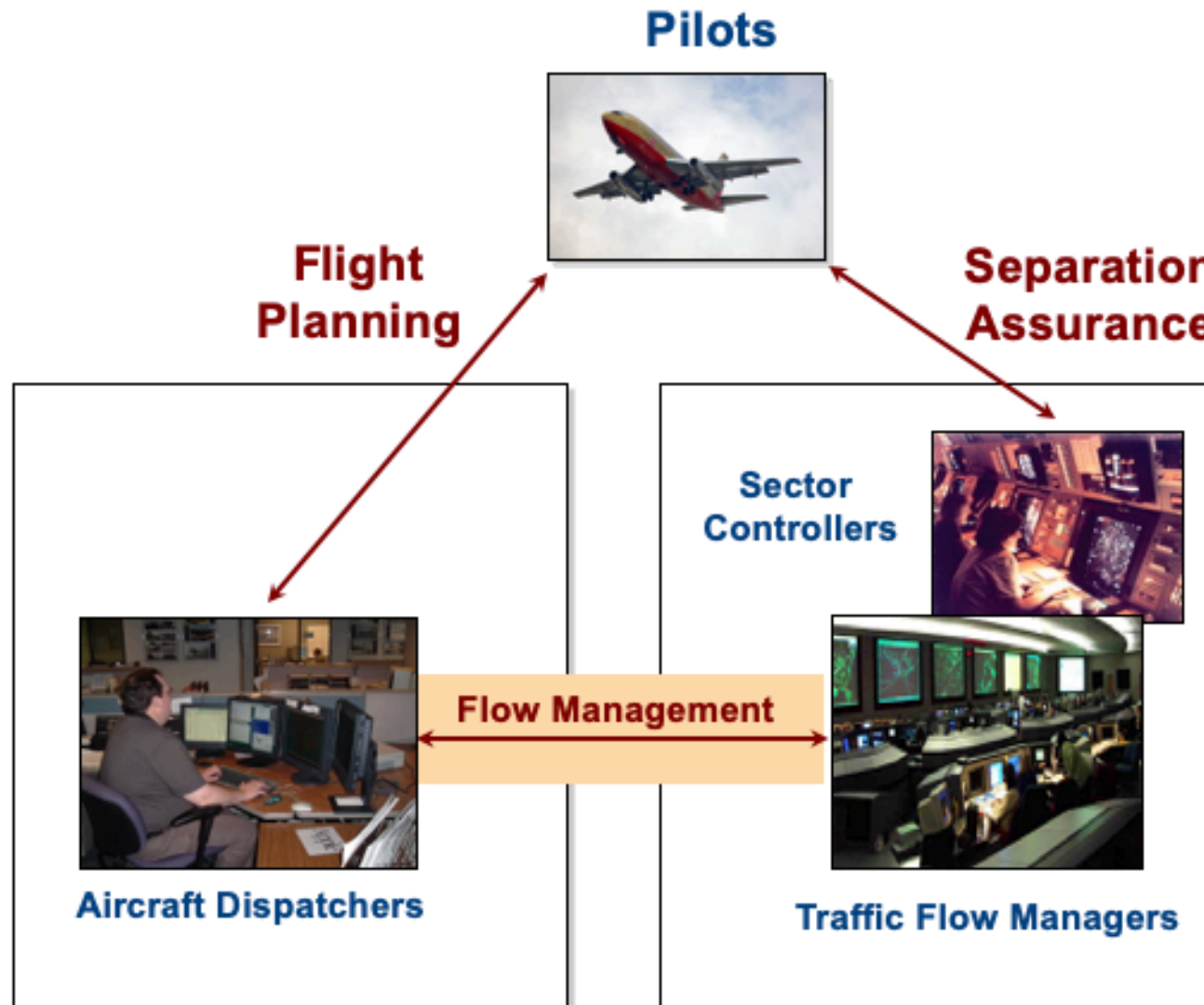
Conflict Detection and Resolution

Problem	Conflict detection and resolution
Data	Randomly generated conflicts in a circular area with radius of 50nm. Conflicts are detected when separation is $< 5\text{nm}$ and $t_{CPA} < 480 \text{ sec}$ Vehicle speed 400nm/hr and max 15 aircraft
Method	RL with Deep Deterministic Policy Gradient
Feature Selection	State and action of each aircraft represented by 8 parameters 73 parameters for 15 aircraft
Method of Evaluation	MAE, RMSE and Confusion Matrix
Remarks	AI agent had a success rate of 81% in the presence of uncertainty and 14 other aircraft

Conflict Detection and Resolution

Problem	Conflict detection and resolution
Data	Mode-S data covering France on Jan 20,2012 21,314 trajectories with flight number, time, position, ground speed, vertical speed, heading, wind speed and direction
Method	Linear (MLR),Non-Linear (SVM,FFNNs, KNN), Ensemble (GBM, and RF)
Feature Selection	Number of trajectories reduced to 88,217 by focusing on trajectories with t_{CPA} between 5 to 20 minutes
Method of Evaluation	MAE, RMSE and Confusion Matrix
Remarks	NN,GBM and RF performed better than the baseline model.GBM outperforms all other methods

Man-Machine Interaction



Man-Machine Interaction

Problem	Air Traffic Controller Workload
Data	Recorded air traffic data (position, velocity of aircraft at Dallas Fort-Worth Center Aug 10, 1998)
Method	Gradient-based Back-propagation Neural Network
Feature Selection	Spatial distribution of aircraft represented by a minimum spanning tree
Method of Evaluation	MAE, RMSE and Correlation Coefficient Confusion Matrix
Remarks	NN correctly identified 95% of low-workload cases, 82% medium-workload cases and was unable to identify high-workload cases due to limited high-workload samples

Man-Machine Interaction

Problem	Reroute Advisories
Data	Trial of DWR concept at American Airlines Accepted and rejected reroute advisories during May-September, 2014
Method	Logistic Regression, SVM, Decision Tree RF and Adaptive Boosting
Feature Selection	10 Features based on controller and pilot activity and expert opinion
Method of Evaluation	MAE, RMSE and Correlation Coefficient Confusion Matrix, F1 and ROC Ten-fold cross-validation
Remarks	RF and Adaboost performed best with F1 score 0.815 and 0.766 respectively better on estimating delay on individual links. Performs varies with problem and prediction horizon

Aviation Safety: Anomaly Detection

Aviation Safety: Anomaly Detection

Problem	Aviation Safety: Anomaly Detection
Data	Boeing 777 aircraft data, 365 flights between 14 airports Aug 10, 1998
Method	Density-based clustering algorithm
Feature Selection	69 flight parameters
Method of Evaluation	MAE, RMSE and Correlation Coefficient Confusion Matrix
Remarks	MLT detects anomaly without looking for violation of the range of parameters. Identified 1%, 3% and 5% outliers during take-off and landing. Poor sensitivity to short duration anomalies and latent features

Concluding Remarks

- Presented research on comparing different modeling approaches in aviation operations
- MLT provide a new class of complimentary tools
- Feature selection plays a key role
- Results on the application of MLT to aviation operations falls into three groups
 - Method of choice due to lack of physics-based models
 - MLT performs better than baseline techniques
 - MLT either marginally better than baseline technique or performs worse
- Task, prior knowledge, data: key to modeling approach