AIAA GNC Conference Session: 84- GNC-49

Paper: AIAA-2009-6079 “Effectiveness of Pairing Flights When Evaluating The Accuracy of a Conflict Probe”

Presented to: AIAA GNC, Chicago, Illinois

Authors:

Mike Paglione, Federal Aviation Administration

Robert Oaks of General Dynamics IT

Date: August 12, 2009

Presentation Outline

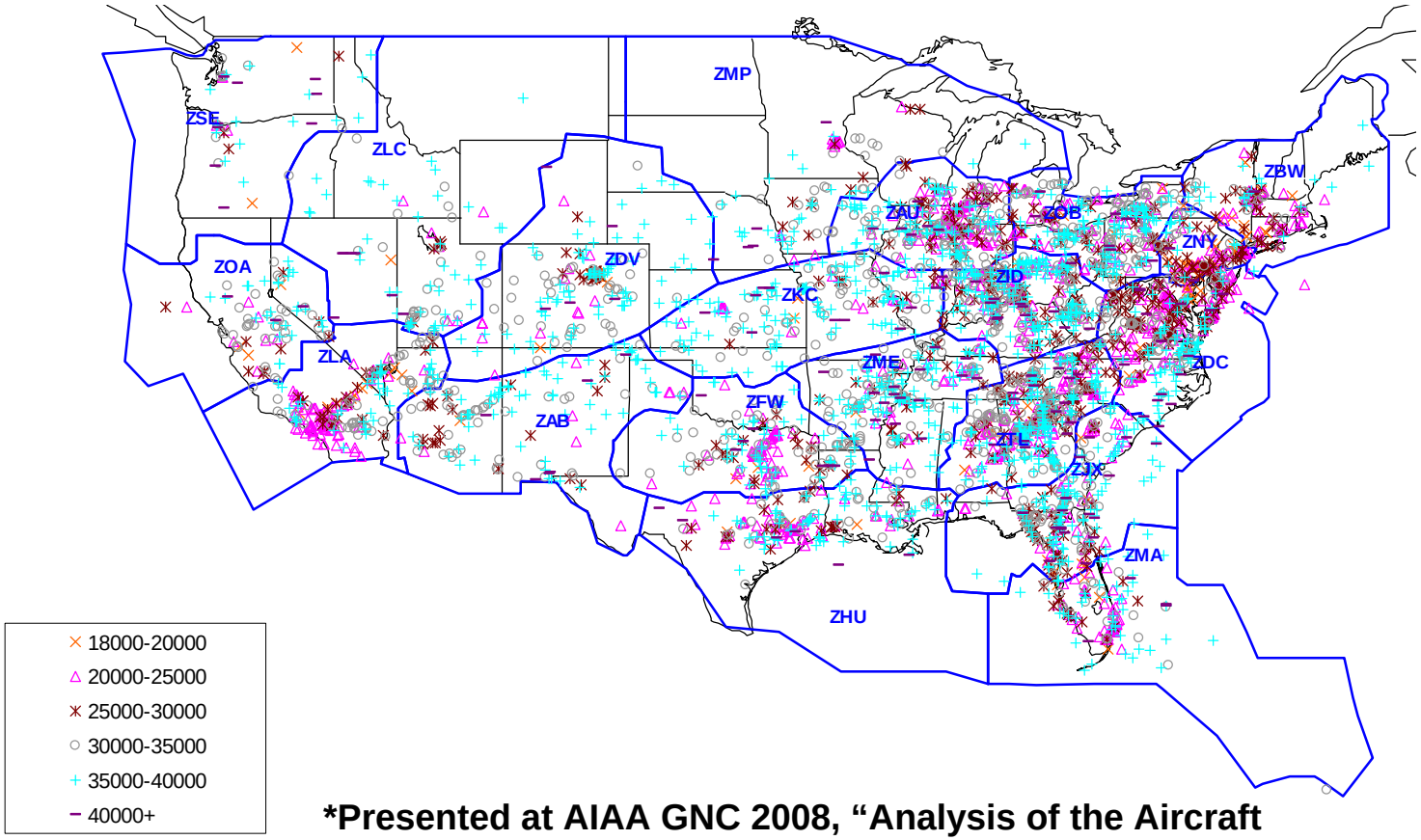
- **Motivation**
- **PDSA Evaluation Process**
- **Analysis & Example**
  - Measure Trajectory Predictor (TP)
    - Standard Two Sample Mean Test
    - Paired Data Test
    - Non-parametric Paired Data Test
  - Measure Conflict Probe (CP)
  - Correlate Results
- **Conclusions**

Motivation of the Study

- **Primary function of Air Traffic Control (ATC)**
  - Safely separate aircraft to prevent collisions
  - Organize/expedite flow of traffic
  - Provide info to pilots when possible
- **Joint Planning Development Office (JPDO)**
  - Forecasts 1.4 to 3 times traffic increase by 2025
  - Next Generation (NextGen) Program
    - Separation Management Project -- Improve separation management automation support
- **Conflict Probe Automation**
  - Predict aircraft-to-aircraft conflicts and recommend resolutions
  - To be used – <u>MUST be accurate and timely</u>

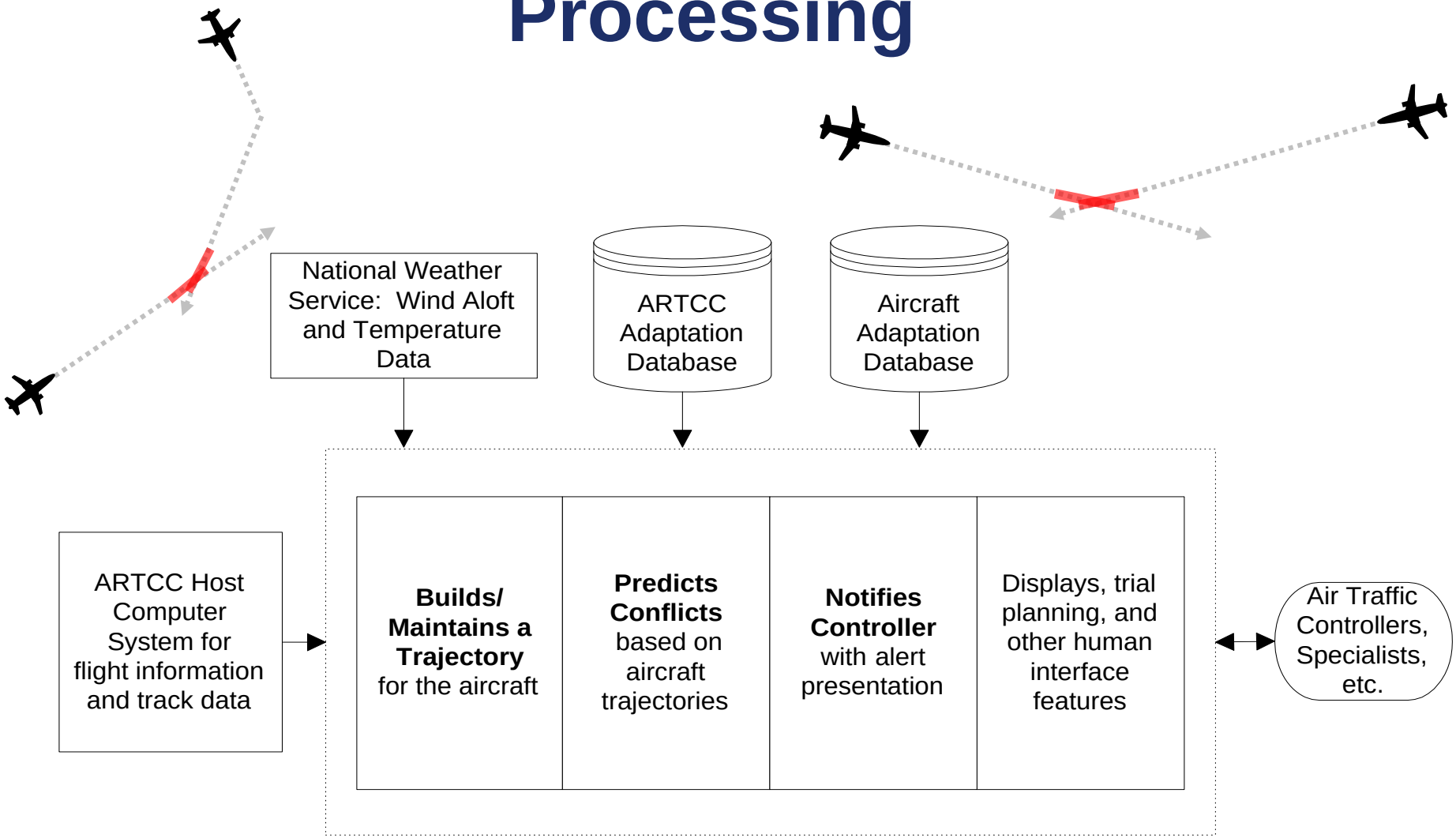
Horizontal View of Conflicts

Data Source: Hour 6 Collected on April-3-2008

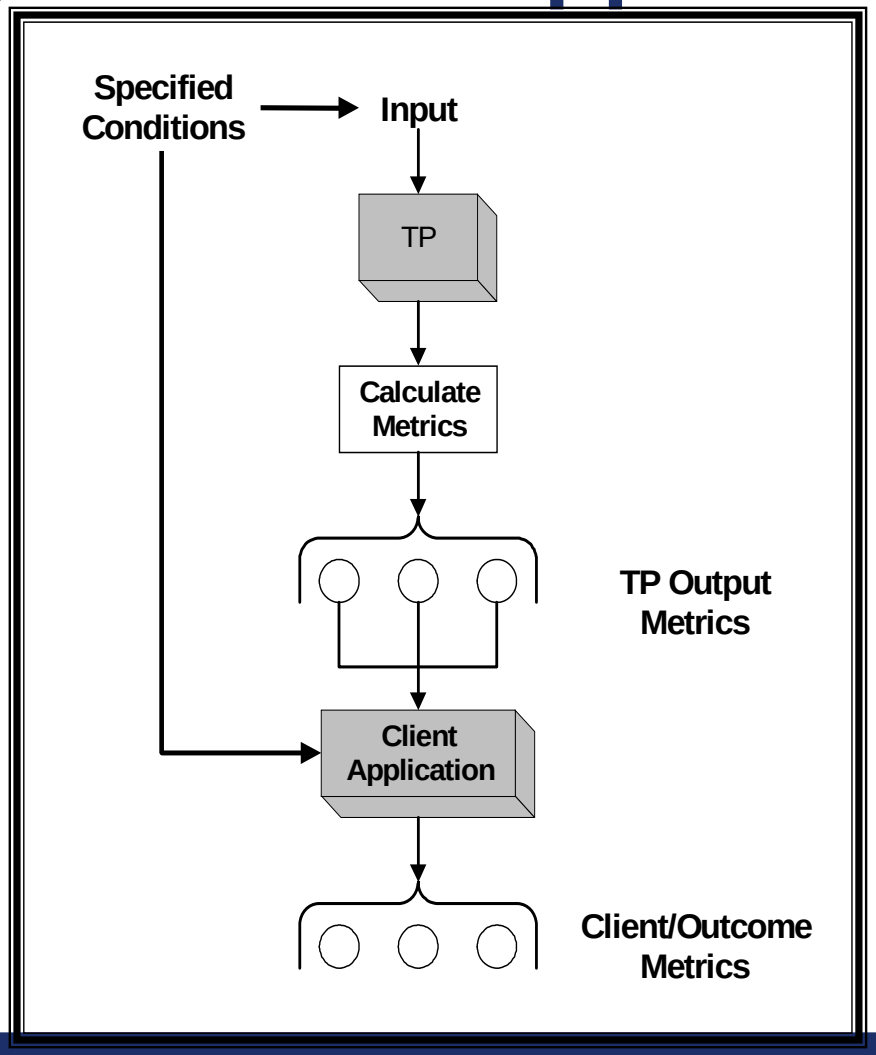


***Presented at AIAA GNC 2008, "Analysis of the Aircraft to Aircraft Conflict Properties in the NAS"**

Components of Conflict Probe (CP) Processing



Trajectory Prediction Metrics to Higher Order Application Metrics



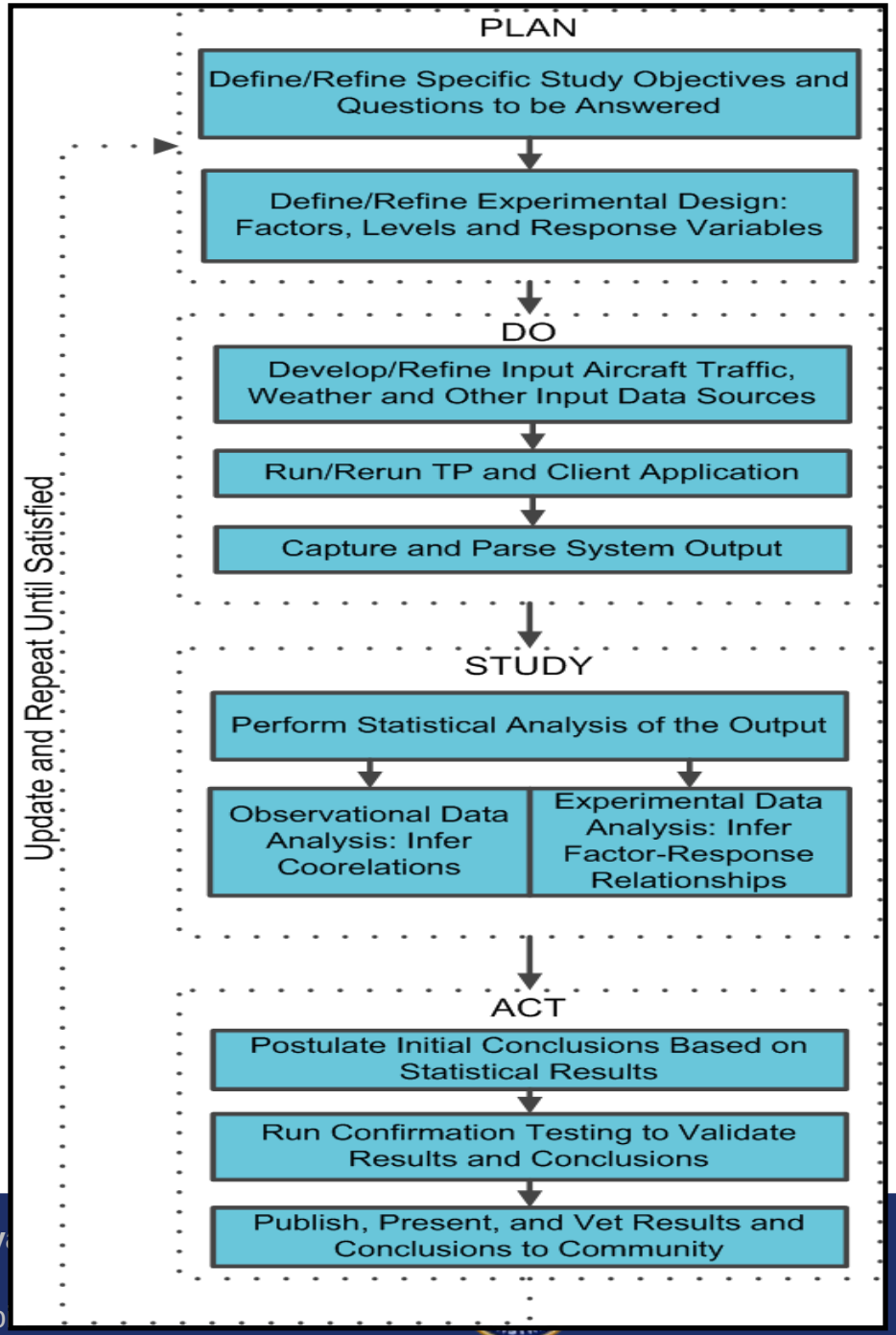
Trajectory Predictor (TP)

Conflict Probe (CP)

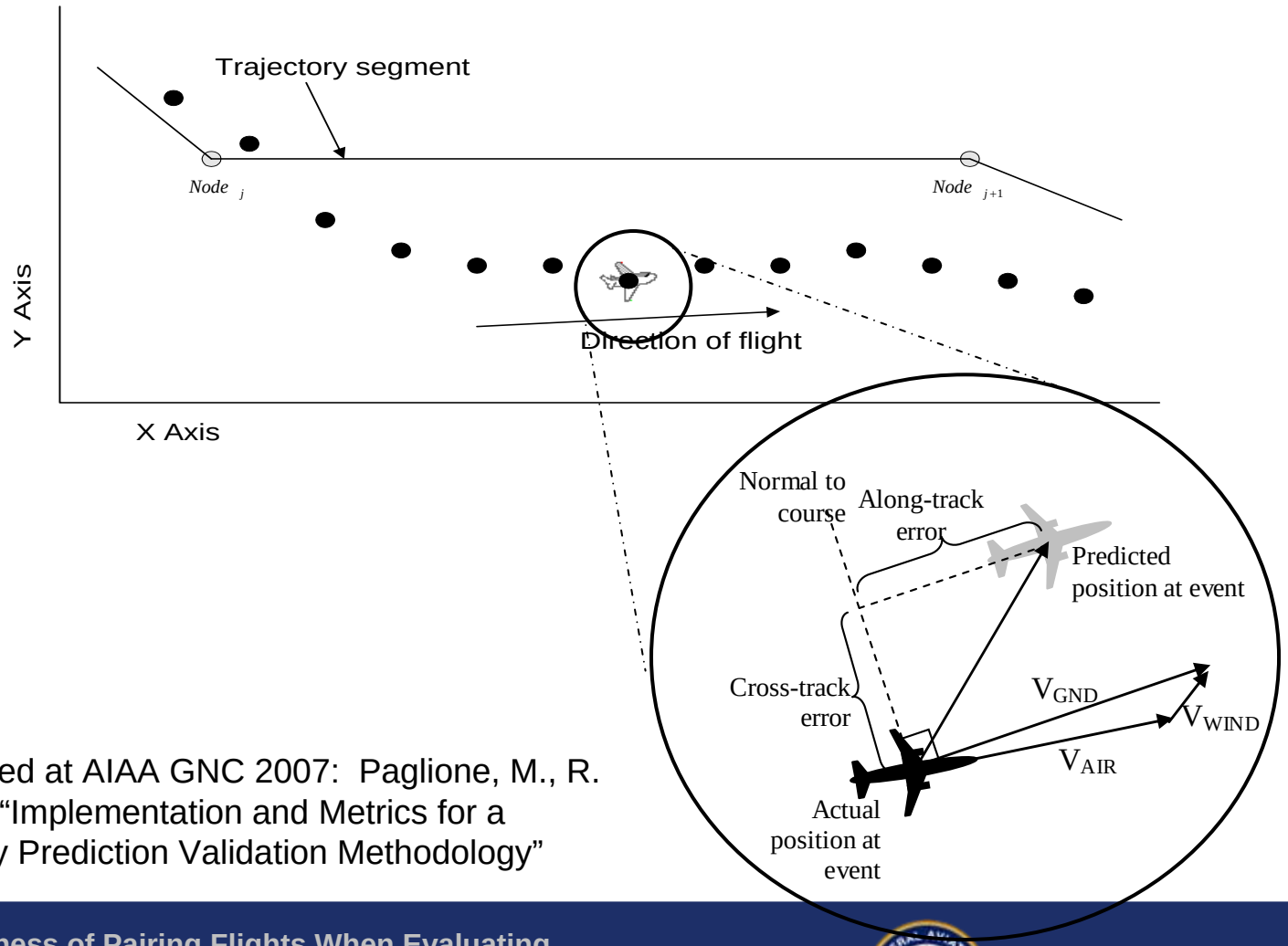
*Presented at DASC 2005: Mondoloni, et.al., "Assessing Trajectory Prediction Performance: Metrics Definition"

PDSA Evaluation Process

- Plan
- Do
- Study
- Act



Trajectory Accuracy Metrics



*Presented at AIAA GNC 2007: Paglione, M., R. D. Oaks, "Implementation and Metrics for a Trajectory Prediction Validation Methodology"

Definitions for Conflict Prediction Accuracy

- **Compare**
  - Pseudo conflicts using time-shifted recorded traffic data
  - CP predicts potential conflicts called alerts

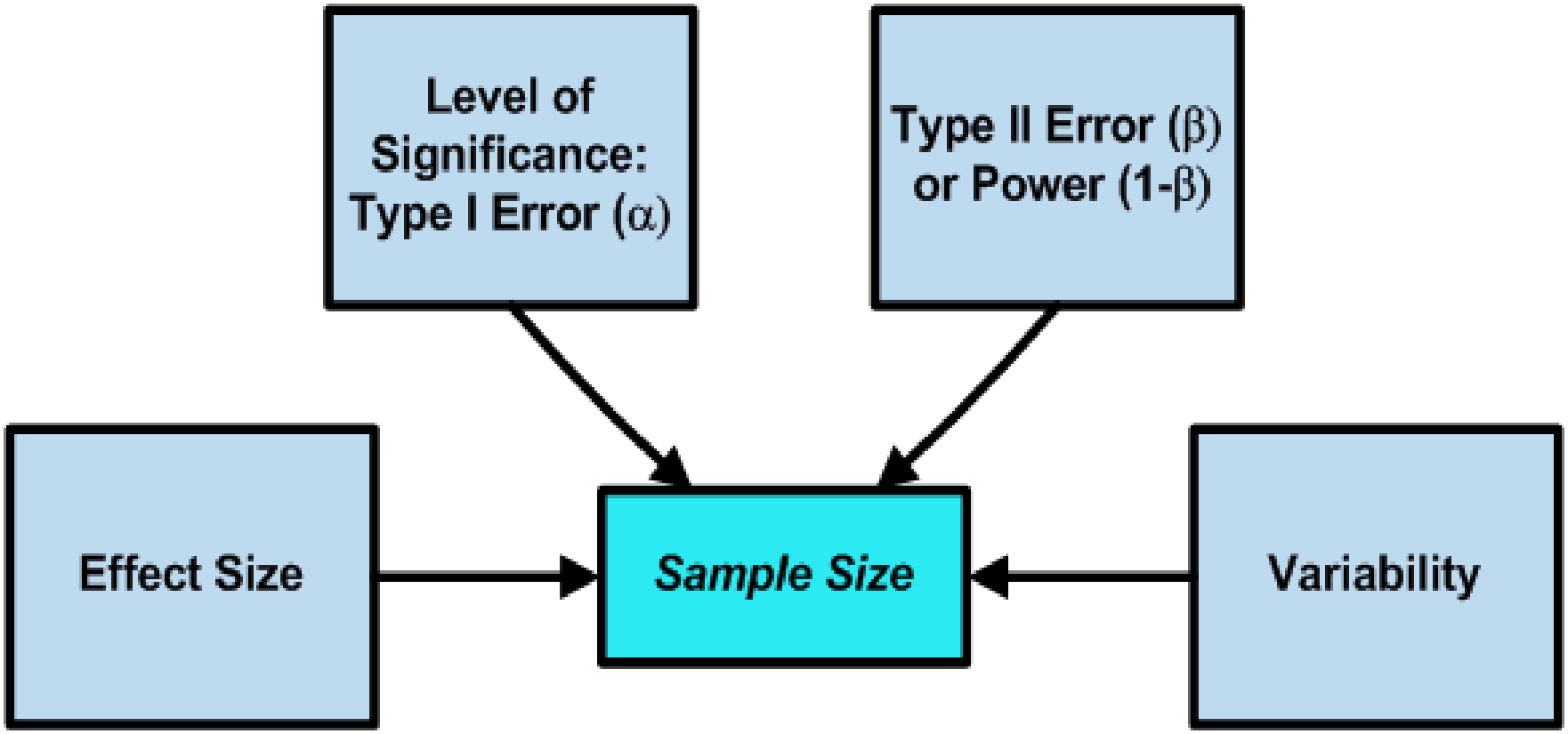
	CONFLICT OCCURS	**CONFLICT DOES NOT OCCUR**
ALERT	CP predicts conflict and it occurs **VA** – valid alert	CP predicts conflict and it does not occur **FA** -- false alert
NO ALERT	CP does not predict conflict and it occurs **MA** -- missed alert	CP does not predict conflict and it does not occur **NC** - correct no-calls
	Total Number of Conflicts	Total Number of Non-Conflicts Encounters without conflicts

- **To ensure timeliness of alerts, CP required to present conflict prediction with some lead-time or actual warning time**
  - Minimum Warning Time (MWT) ~ 5 minutes if non-popup event

PDSA: Research Object = Improve CP

- **Plan: Research questions**
  - How to improve CP? Improve TP
  - How to improve TP? Measure errors then isolate sources of errors and reduce.
- **Plan: Define metrics (e.g. cross track error)**
- **Do: Take measurements and sample**
- **Study: Hypothesis Test – Two Sample t-Test**
  - Compare two systems, same system with upgrade, etc.
  - Is new system equal to baseline? $H_o : \mu_b - \mu_n = 0$
  - Balance two errors
    - Type I (α) – failing the null hypothesis when true
    - Type II (β) – failing to reject the null hypothesis when it is false

Relationships with Sample Size



Example – Two Sample t-Test Applied

- **Compare baseline and new system**
- **Metric – sample mean cross track error (nm)**
- **236 flights, 118 per system**
- **Test statistic:**

$$t = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{s_b^2}{m} + \frac{s_n^2}{n}}}$$

- **Reject null hypothesis if**

$$t \geq t_{\alpha/2,\upsilon} \quad \textbf{or} \quad t \leq -t_{\alpha/2,\upsilon}$$

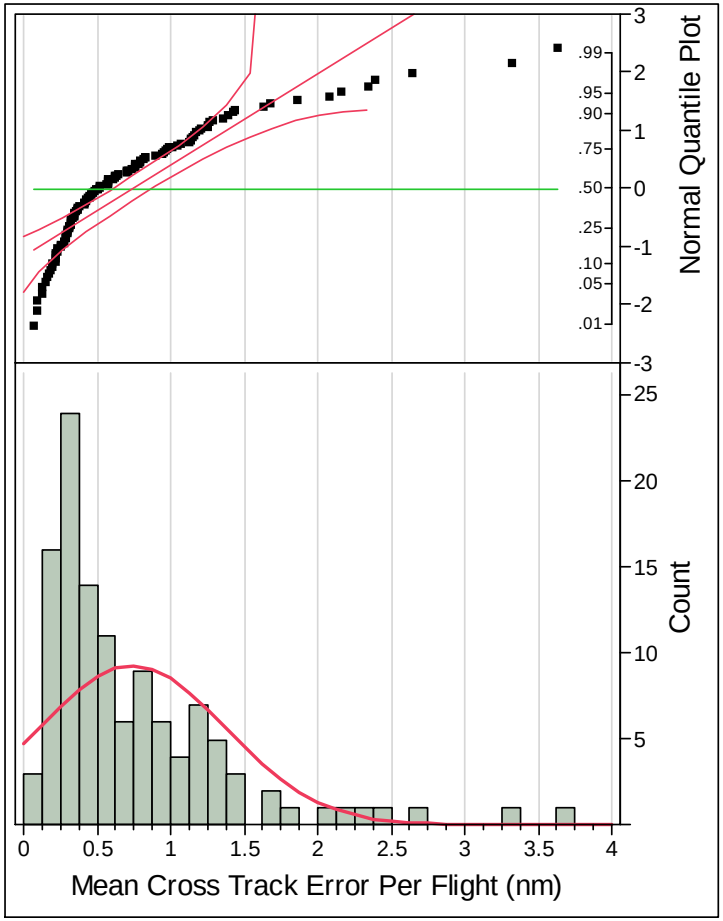
- **Two key assumptions – normally distribution and independent samples**

Example – Two Sample t-Test Applied Continued

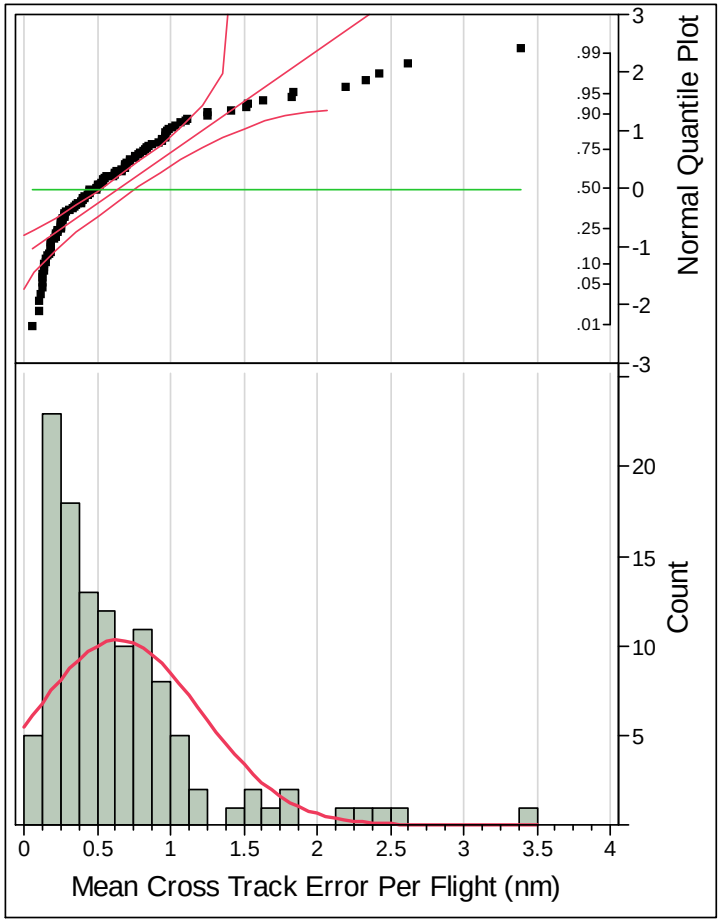
- **t = 1.209, p-value=0.113**
- **Reject if t>=1.97 or t<=-1.97**
- **Thus, do not reject but ….**

Parameter	Prospective	Retrospective
Type I error (α)	0.05	0.05
Standard deviation (nm)	0.6	0.64, 0.57
Mean of baseline system (nm)	0.7	0.74
Mean of new system (nm)	0.6	0.64
Total sample size from both runs	1600	236
Power (1 - β)	0.91	0.25

Not Normally Distributed Data!

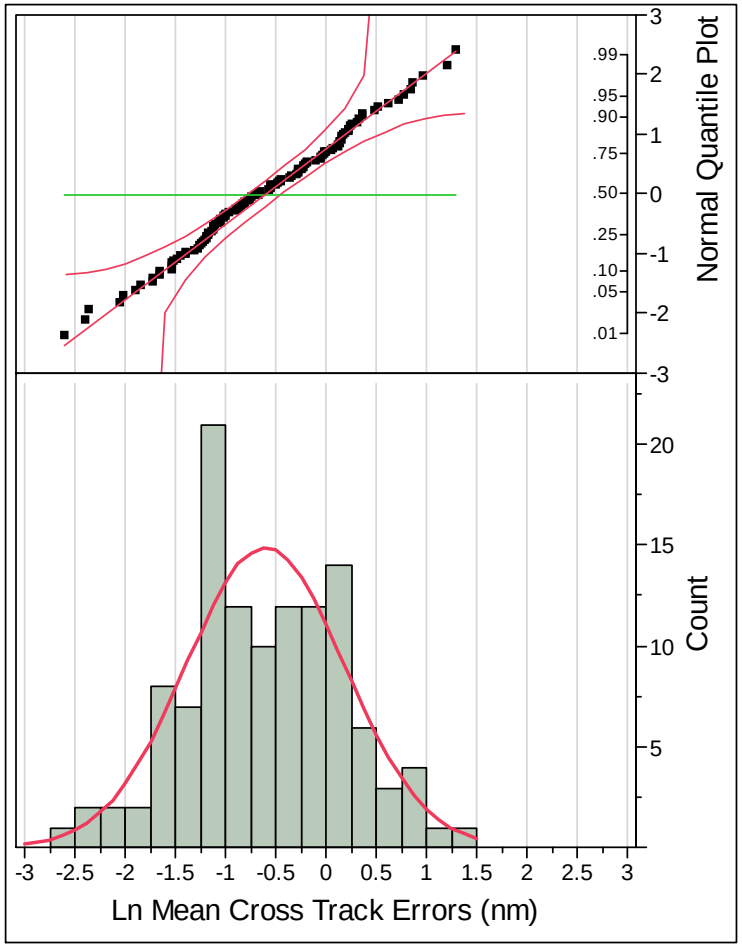


Baseline System

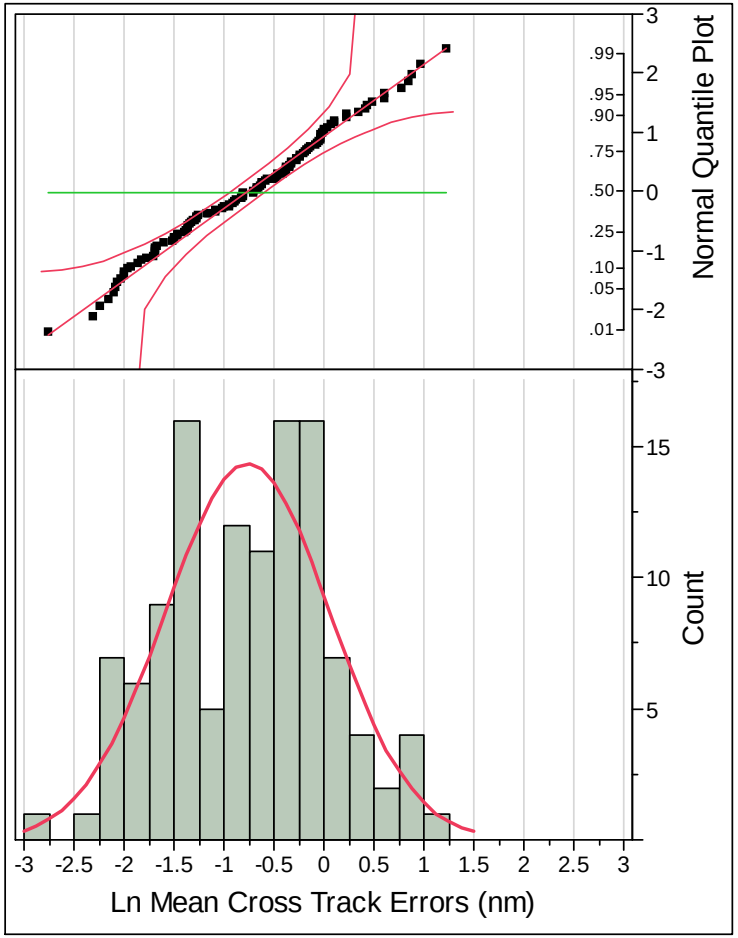


New System

Transform Data with Natural Log Now Normally Distributed Data.



Baseline System



New System

Example – Two Sample t-Test Applied Transform Data Using Natural Log

- **t = 1.510, p-value=0.066**
- **Reject if t>=1.97 or t<=-1.97**
- **Thus, closer still do not reject but ….**

Parameter	Prospective	Retrospective
Type I error (α)	0.05	0.05
Standard deviation (nm)	0.8	0.79, 0.82
Mean of baseline system (nm)	-0.35	-0.61
Mean of new system (nm)	-0.50	-0.77
Total sample size from both runs	1600	236
Power (1 - β)	0.96	0.30

Example – Paired Data t-Test Applied

- **Compare baseline and new system**
- **Metric – sample mean cross track error (nm)**
- **118 flights per system but pair flights:** $D_i = x_i - y_i$
- **Now test statistic:**

$$t = \frac{\bar{d}}{s_D / \sqrt{n}}$$

- **Reject null hypothesis if**

$$t \geq t_{\alpha/2, n-1} \quad \textbf{or} \quad t \leq -t_{\alpha/2, n-1}$$

- **Key assumption – normally distributed differences**

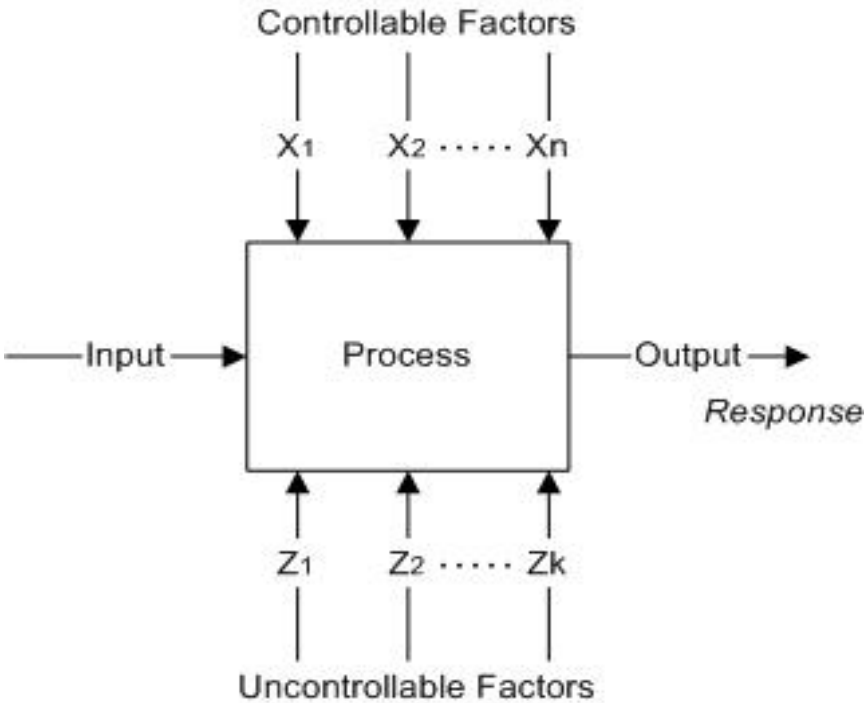
Example – Paired Data t-Test Applied Continued

- **t = 3.26, p-value=0.0015**
- **Reject if t>=2.27 or t<=-2.27**
- **Thus, reject hypothesis…and**

Parameter	Prospective	Retrospective
Type I error (α)	0.05	0.05
Standard deviation (nm)	0.3	0.317
Mean of baseline system (nm)	0.7	0.74
Mean of new system (nm)	0.6	0.64
Total sample size from both runs	800	118
Power (1 - β)	1.0	0.92

Why so much better?

- **Data not independent**
- **Data positively correlated**
- **Blocking out extraneous factors…**



$$V(\overline{X} - \overline{Y}) = \frac{V(D_i)}{n} = \frac{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}{n}$$

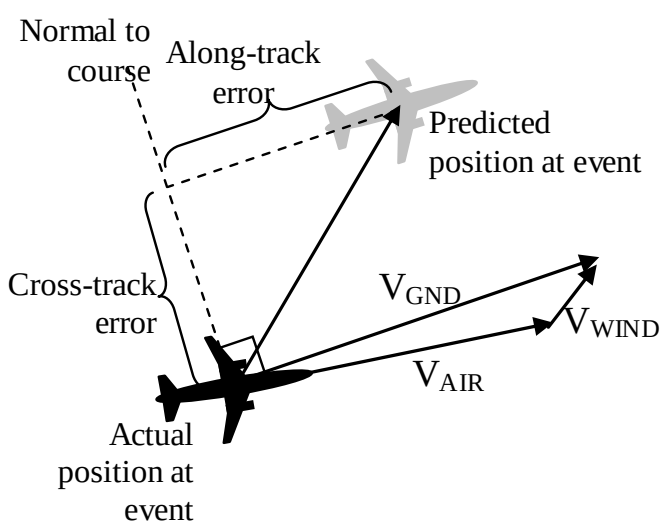
Example – Non-Parametric Paired Test

- **Wilcoxon Signed-Rank Test**
- **Data not required to be normally distributed**
- **Steps**
  - Absolute values of paired differences calculated and ranked in ascending order
  - Next, ranks of the positive measurements are summed, referred to as S+.
  - Test statistic: $$Z = \frac{S_+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

- **Reject null hypothesis if** $Z \geq z_{\alpha/2}$ or $Z \leq -z_{\alpha/2}$
- **From same example**
  - Z=1396.5
  - p-value=0.0001

Observational Data TP and CP Results

- **Combine errors by calculating ratio of**



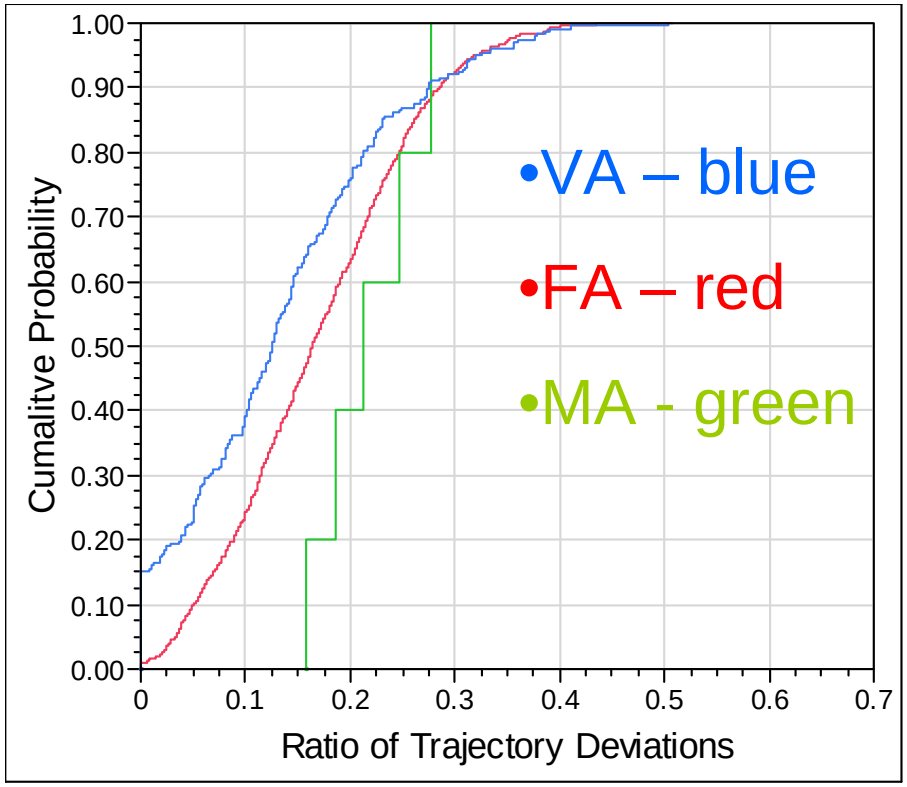
Number of traj errors within time window beyond threshold

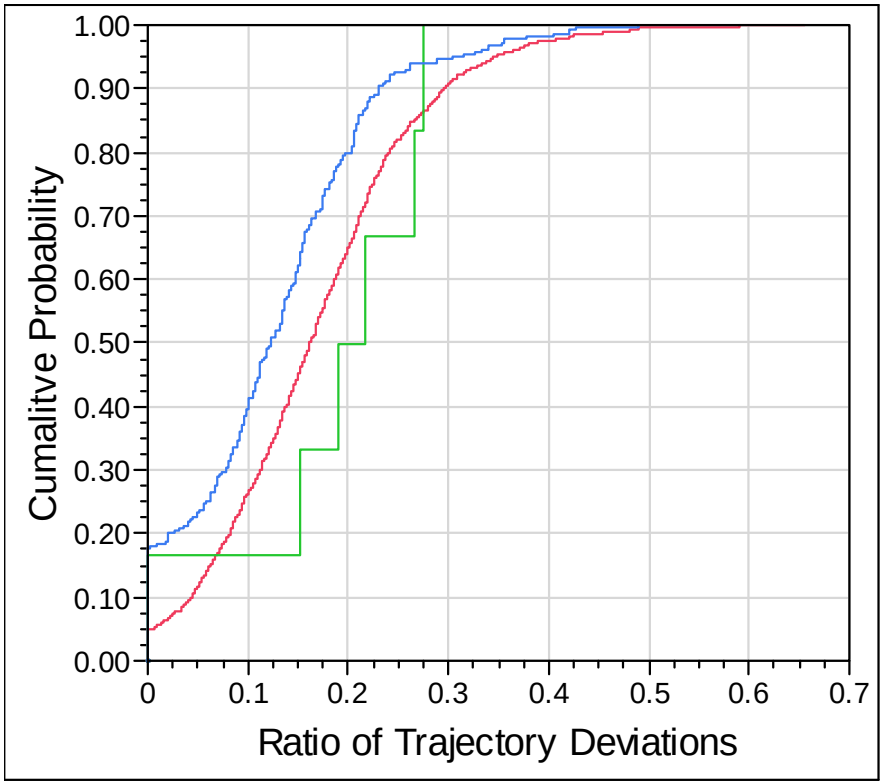
- Cross track error, > 1.5 nm
- Along track error, >2.5 nm
- Vertical error

>500 ft

- **For each conflict prediction event VA, MA, FA, or even NC, ratio calculated from (PCST, PCST-1200s)**

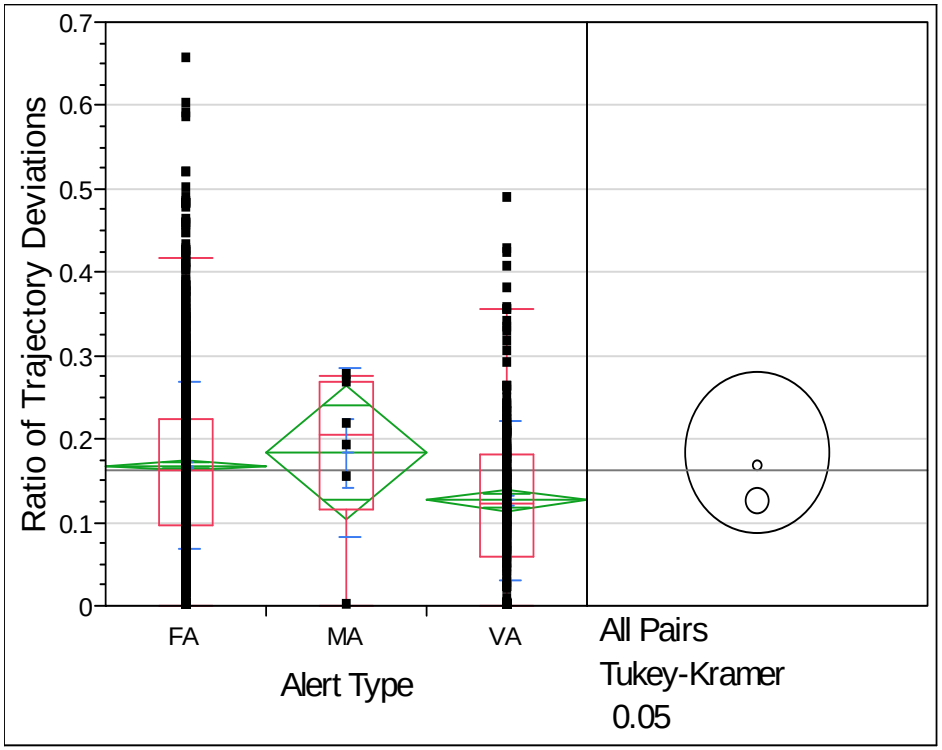
Observational Data: TP/CP Distribution

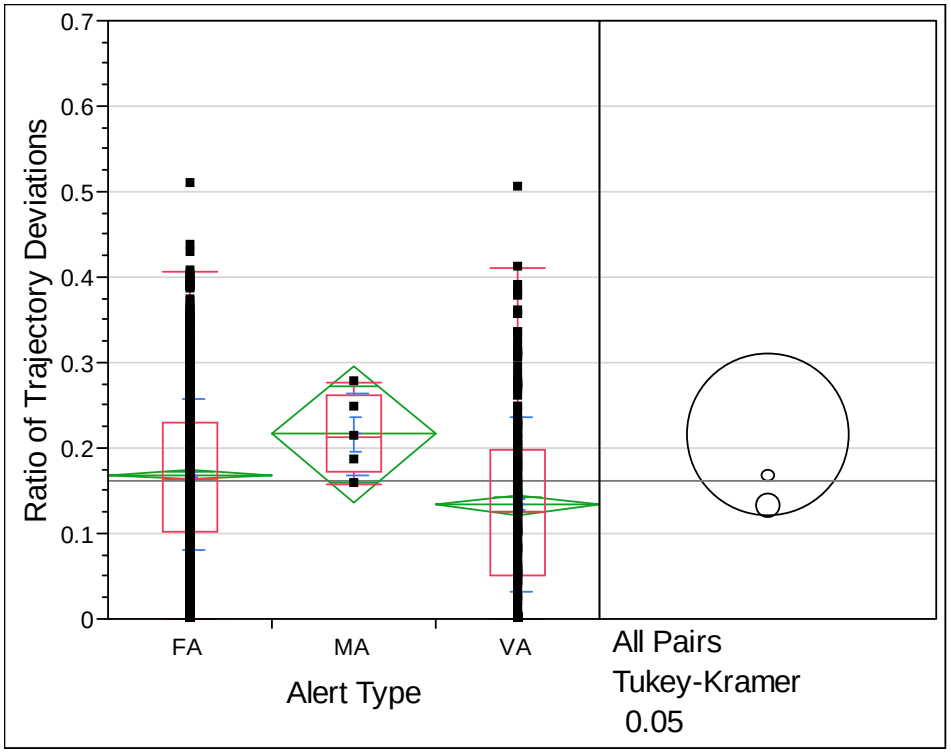




Baseline System **New System**

Observational Data: TP/CP Distribution Continued



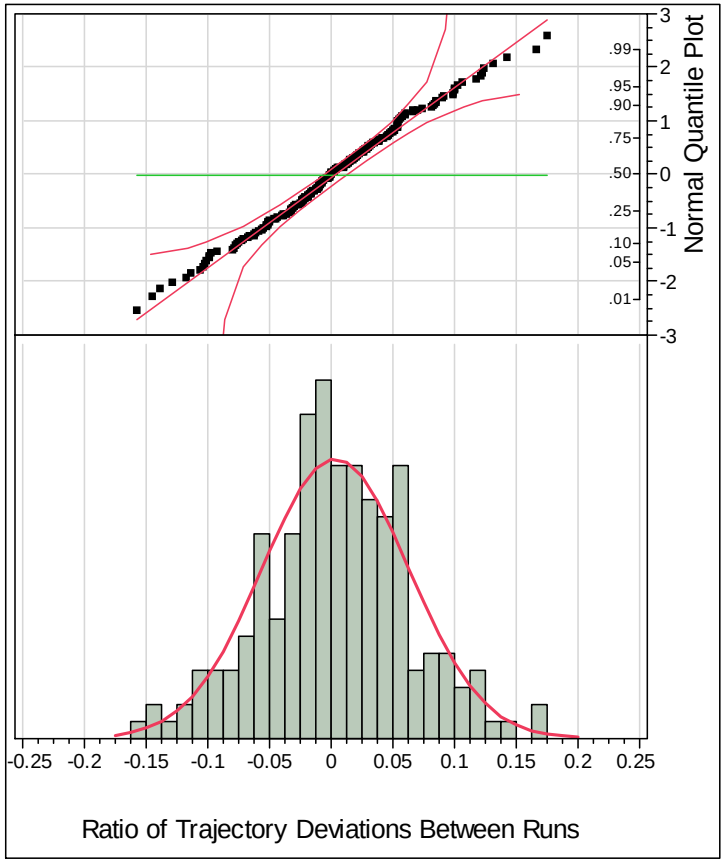


Baseline System **New System**

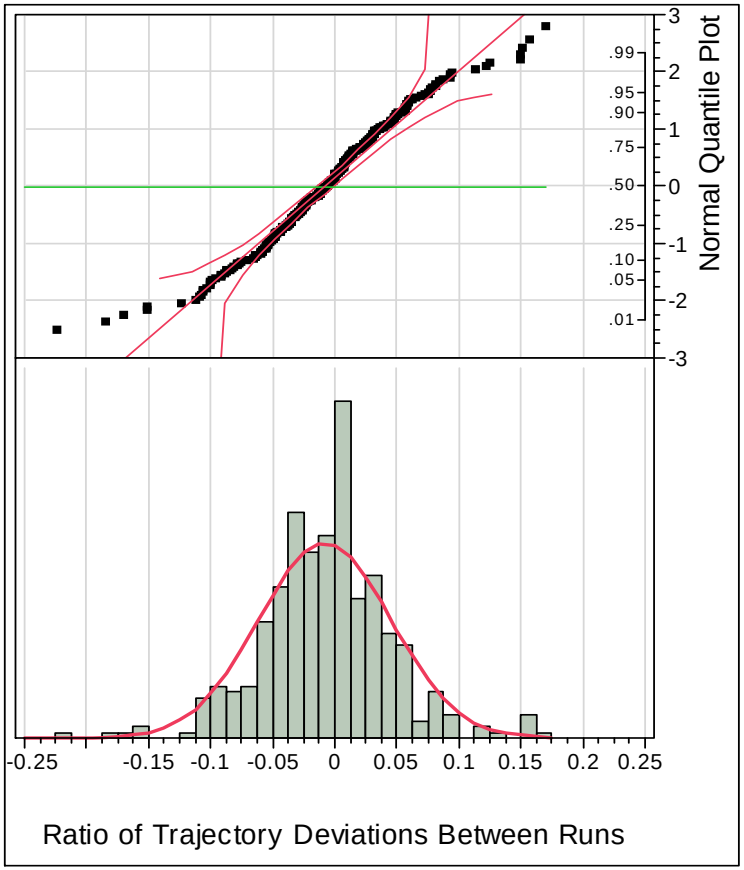
Pairing Conflict Predictions Between Runs

	CONFLICT OCCURS	CONFLICT DOES NOT OCCUR
ALERT by both Runs A and B	Both predicts conflict and it occurs (V_{A1}=V_{B1}-- valid alerts both) **SAME_VA**	Both predicts conflict and it does not occur (F_{A1}=F_{B1}-- false alert both) **SAME_FA**
ALERT by A and not B	A predicts conflict and it occurs (V_{A2} -- valid alerts by A only)	A predicts conflict and it does not occur (F_{A2} -- false alert by A only)
	B does not predict conflict and it occurs (M_{B2} -- missed alert by B only) **VA_MA**	B does not predict conflict and it does not occur (NC_B -- correct no-calls by B only) **FA_NC**
ALERT by B and not A	B predicts conflict and it occurs (V_{B2} -- valid alerts by B only)	B predicts conflict and it does not occur (F_{B2} -- false alert by B only)
	A does not predict conflict and it occurs (M_{A2} -- missed alert by A only) **MA_VA**	A does not predict conflict and it does not occur (NC_A -- correct no-calls by A only) **NC_FA**
NO ALERT by both Runs A and B	Both do not predict conflict and it occurs (M_{A1}= M_{B1} -- missed alert by both) **SAME_MA**	Both do not predict conflict and it does not occur (NC -- correct no-calls by both)
	Total Number of Conflicts *(Same for both Runs!)*	Total Number of Non-Conflicts (Encounters that did not have conflicts; *Same for both Runs!*)

Observational Data: Paired CP Data and Paired TP Distribution



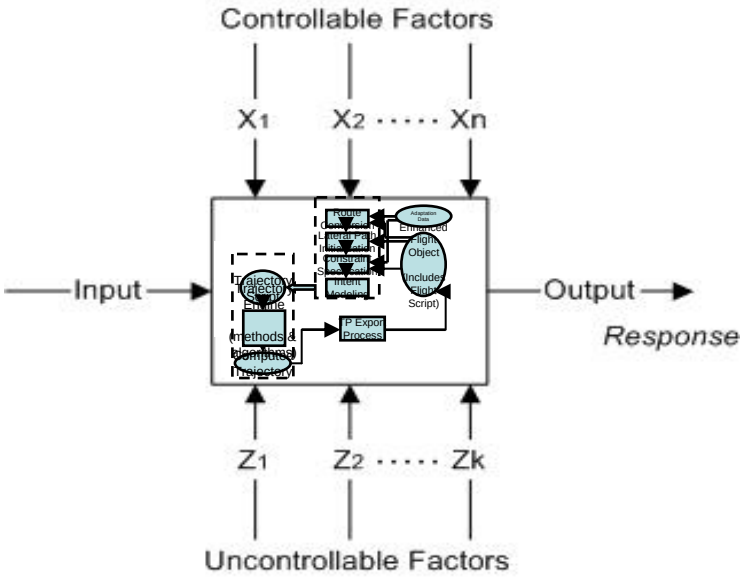
SAME_VA



FA_NC

Closing Remarks…

- **NextGen requires advances to CP and TP**
- **Plan-Do-Study-Act (PDSA) Evaluation Process**
  - Practical significance versus statistical significance
  - Remove uncontrollable factors – block by flight
  - Use non-statistical knowledge as well
  - Process is iterative & evolving
- **Example illustrates the techniques**
- **Future work**
  - Experimental design analysis
  - Not black box ~ white/gray box



Lead Author Contact Data

Mike M. Paglione
Federal Aviation Administration
Simulation & Analysis Group
mike.paglione@faa.gov
609-485-7926

Paper's Web Site:
http://acy.tc.faa.gov/cpat/docs/