



*The AFIT of Today is the Air Force of Tomorrow.*

# **45TH WS ELECTRIC FIELD MILL LIGHTNING PREDICTION THRESHOLD ANALYSIS**

**Maj Charles A. Skrovan**

**Thesis Advisor: Lt Col Geyer**

**Reader: Dr. Hill**

15 Jan 2020

*Air University: The Intellectual and Leadership Center of the Air Force*

*Aim High...Fly - Fight - Win*



## Abstract



*The AFIT of Today is the Air Force of Tomorrow.*

**This study seeks to determine a single electric field mill reading threshold for lightning onset and a separate single EFM reading threshold for lightning cessation. A regression model analysis and threshold analysis of time series data is planned to be used to determine thresholds for 20 minutes to 30 minutes, using 5-minute increments before the first total lightning detection and 15 minutes after the last total lightning detection in the vicinity of an EFM. Threshold values for each EFM voltage reading will also be considered from 100 volts/meter to 4000 volts/meter in 100 volts/meter increments for each detection time.**

*Air University: The Intellectual and Leadership Center of the Air Force*

*Aim High...Fly - Fight - Win*

2



## Purpose



*The AFIT of Today is the Air Force of Tomorrow.*

- **Determine a threshold Electric Field Mill reading value to predict lightning onset (30 minutes prior to first lightning strike) and succession (15 minutes after last lightning strike)**
  - **Lightning strikes can have potential negative affect on space vehicles and hardware**
  - **Stop work for certain processing operations when lightning detected w/in 5NM of CCAFS leads to loss of production**

*Air University: The Intellectual and Leadership Center of the Air Force*

*Aim High...Fly - Fight - Win*



## Data Preparation



*The AFIT of Today is the Air Force of Tomorrow.*

- Convert RAW data to R data frame combined 20,000+ separate files
- Reduce & filter noise EFM data from 50 Hz readings to 1-minute averages for 31 sensors
  - $50 \text{ Hz} * 60 \text{ Seconds} * 60 \text{ Minutes} * 24 \text{ Hours} * 30 \text{ Days} * 5 \text{ Months} * 4 \text{ Years} = \sim 2.6 \text{ billion rows}$
  - Converting to minute data and missing reduced to less than 741,000 rows by 31 sensor columns
- Convert LDAR distances from meters to lat/long then to 5 NM radius for each of 31 sensors
  - (consider total lightning: cloud-to-ground and cloud-to-cloud)
- Combine EFM and LDAR data sets into single data frame
- Create additional statistical columns
  - Absolute values of EFM readings
  - Absolute mean of EFM readings
  - Geometric mean of EFM readings
  - Binary lightning strike/storm windows
  - Determine storm length and time-to-end

*Air University: The Intellectual and Leadership Center of the Air Force*

*Aim High...Fly - Fight - Win*

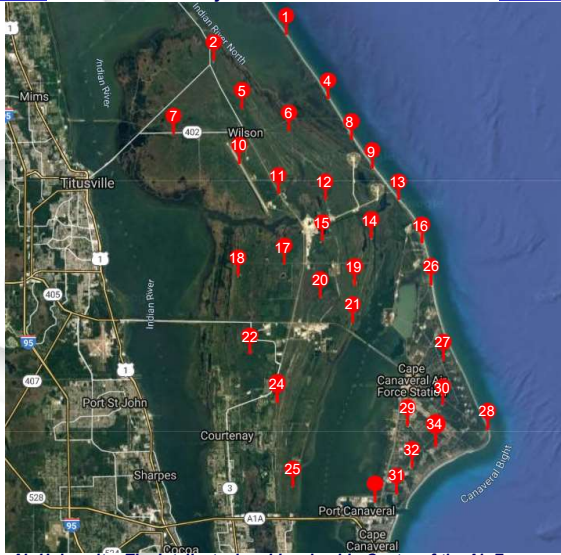
4



# EFM Locations



*The AFIT of Today is the Air Force of Tomorrow.*



*Air University: The Intellectual and Leadership Center of the Air Force*

*Aim High...Fly - Fight - Win*

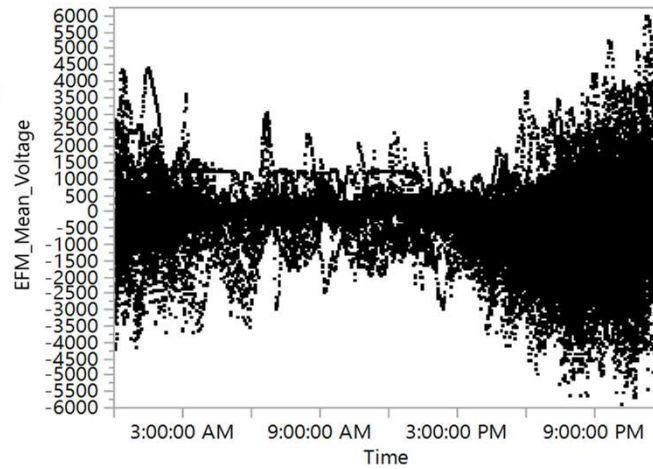


# EFM Mean Voltage



The AFIT of Today is the Air Force of Tomorrow.

Bivariate Fit of EFM Mean Voltage By Time



Air University: The Intellectual and Leadership Center of the Air Force

Aim High...Fly - Fight - Win

The scatter plot shows that there is potentially some correlation for EFM voltages to time of day, however further analysis proves otherwise. This is due to the huge amount of data close to the origin. Diurnal variation.

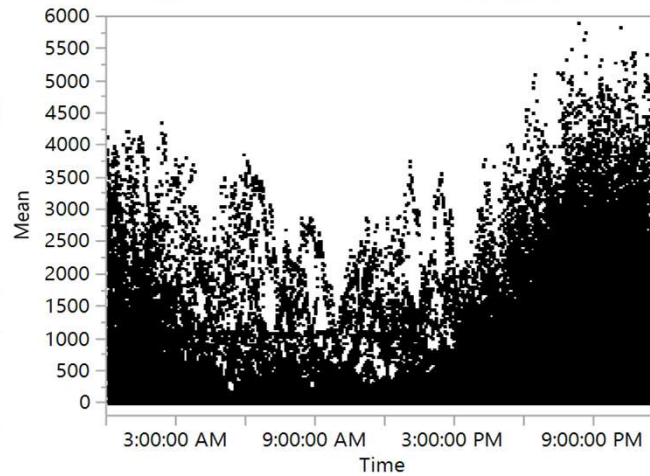


# EFM Absolute Mean Voltage



*The AFIT of Today is the Air Force of Tomorrow.*

**Bivariate Fit of EFM Absolute Mean Voltage By Time**



*Air University: The Intellectual and Leadership Center of the Air Force*

*Aim High...Fly - Fight - Win*

This is a scatter plot of all the EFM readings measured as the difference between the EFM volt/m and the non-storm average volt/m for that sensor.

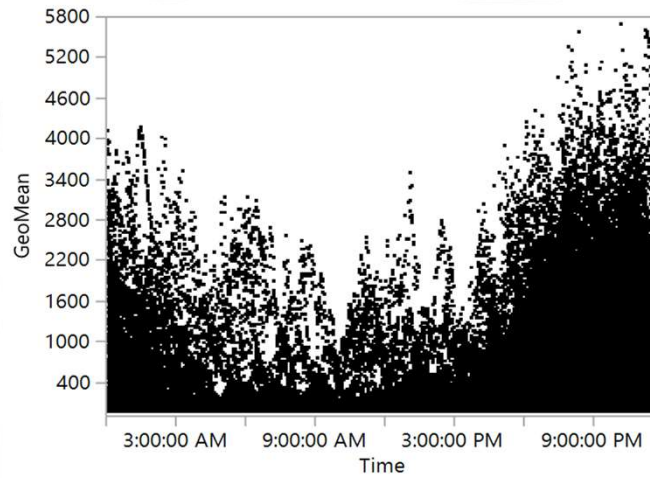


# EFM Geometric Mean Voltage



The AFIT of Today is the Air Force of Tomorrow.

Bivariate Fit of EFM Geometric Mean Voltage By Time



Air University: The Intellectual and Leadership Center of the Air Force

Aim High...Fly - Fight - Win

Geometric mean just shows tighter bands as extremely large values aren't as heavily weighted





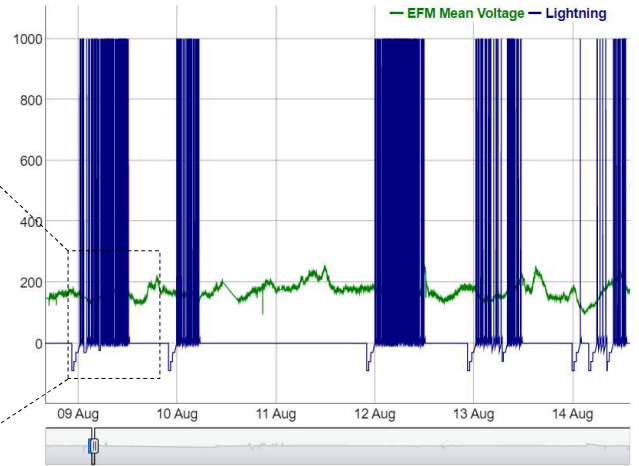
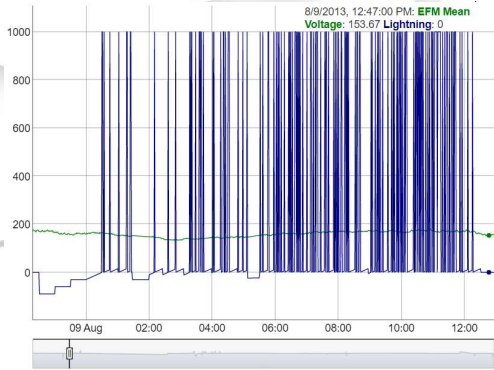
# EFM vs. Lightning Issues



The AFIT of Today is the Air Force of Tomorrow.

- 9 – 14 Aug 2013
- Issue: relatively flat EFM response during known storms

Maj Charles A. Skrovan  
11/13/2019



Air University: The Intellectual and Leadership Center of the Air Force

Aim High...Fly - Fight - Win

EFM readings stay near average when lightning storms are occurring.



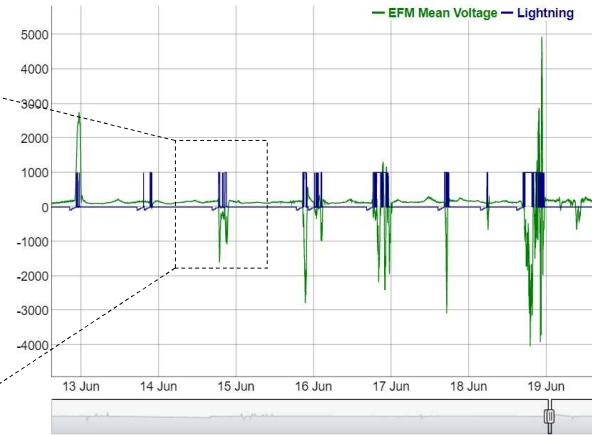
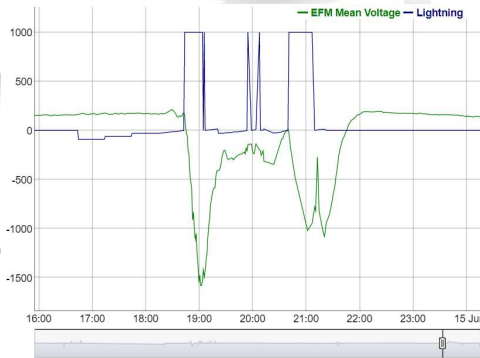
# EFM vs. Lightning Issues



The AFIT of Today is the Air Force of Tomorrow.

- 13 – 19 Jun 2016
- Issue: response occurs after storms begin

Maj Charles A. Skrovan  
11/13/2019



Air University: The Intellectual and Leadership Center of the Air Force

Aim High...Fly - Fight - Win

EFM readings spike AFTER first lightning strike occurs/storm begins.

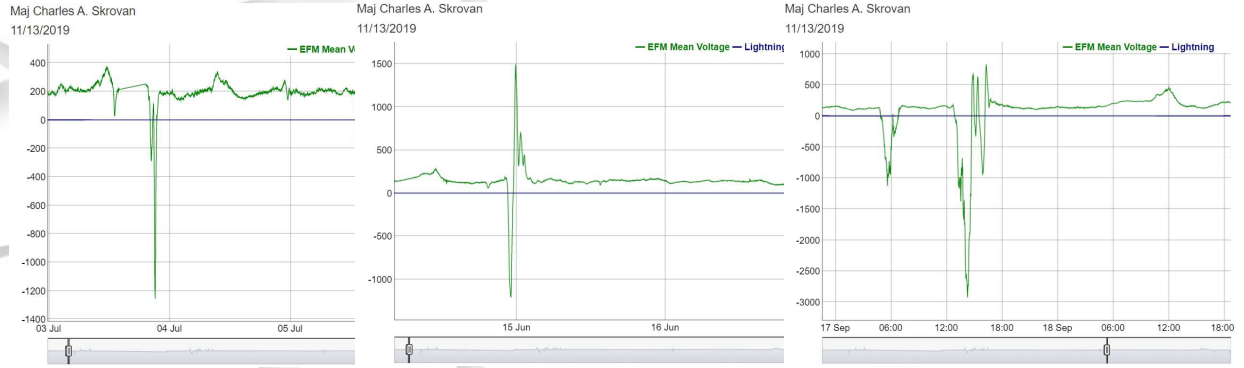


# EFM vs. Lightning Issues



The AFIT of Today is the Air Force of Tomorrow.

- 14 – 17 Jun, 3 – 7 Jul 2013, 17 – 19 Jun 2015
- Issue: EFM spikes when no lightning storms occur



Air University: The Intellectual and Leadership Center of the Air Force

Aim High...Fly - Fight - Win

EFM spikes but no storms are reported based on LDAR data.





# EFM Mean Voltage Least Squares Fit



The AFIT of Today is the Air Force of Tomorrow.

## EFM Mean Voltage/Time vs. Lightning Prediction

### Summary of Fit

RSquare	0.141512
RSquare Adj	0.14151
Root Mean Square Error	211.1348
Mean of Response	49.37997
Observations (or Sum Wgts)	735811

### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	5406842121	2.7034e+9	60644.86
Error	735808	3.2801e+10	44577.906	Prob > F
C. Total	735810	3.8208e+10		<.0001*

### Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	45.217513	0.504074	89.70	<.0001*
EFM_Mean_Voltage	-0.217787	0.00065	-335.2	<.0001*
Time	0.0006772	9.885e-6	68.50	<.0001*

### Prediction Expression

45.21751263+ -0.217786895\*EFM\_Mean\_Voltage+ 0.0006771592\*Time

## EFM Abs Mean V./Time vs. Lightning Prediction

### Summary of Fit

RSquare	0.211176
RSquare Adj	0.211174
Root Mean Square Error	202.3867
Mean of Response	49.37967
Observations (or Sum Wgts)	735814

### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	8068539637	4.0343e+9	98492.00
Error	735811	3.0139e+10	40960.383	Prob > F
C. Total	735813	3.8208e+10		<.0001*

### Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	3.8571723	0.472209	8.17	<.0001*
AbsMean	0.2464774	0.00057	432.78	<.0001*
Time	0.0001819	9.605e-6	18.94	<.0001*

### Prediction Expression

3.8571722573+ 0.2464773585\*AbsMean+ 0.0001819149\*Time

## EFM Geo Mean V./Time vs. Lightning Prediction

### Summary of Fit

RSquare	0.192712
RSquare Adj	0.19271
Root Mean Square Error	204.7416
Mean of Response	49.37967
Observations (or Sum Wgts)	735814

### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	7363080395	3.6815e+9	87824.82
Error	735811	3.0845e+10	41919.133	Prob > F
C. Total	735813	3.8208e+10		<.0001*

### Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	6.50769	0.477587	13.63	<.0001*
GeoMean	0.2901371	0.000712	407.66	<.0001*
Time	0.0002731	9.693e-6	28.17	<.0001*

### Prediction Expression

6.5076899928+ 0.2901370729\*GeoMean+ 0.0002730918\*Time

Air University: The Intellectual and Leadership Center of the Air Force

Aim High...Fly - Fight - Win

Least Squares Linear Regression models comparing the Mean Voltage and Time, Absolute Mean Voltage from Sensor Mean and Time, and Geometric Mean Voltage from Sensor Mean, with Lightning as the response show how each input explains the amount of variance in the dataset. The best method is Absolute Mean Voltage from the Sensor Mean with Time, however, this only accounts for just over 20% of variation in the dataset.



# EFM All Sensors Least Squares Fit



The AFIT of Today is the Air Force of Tomorrow.

### Effect Summary

Source	LogWorth	PValue
KSC12	1167.053	0.00000
KSC7	1055.954	0.00000
KSC25	900.665	0.00000
KSC17	479.796	0.00000
KSC24	354.208	0.00000
KSC18	296.113	0.00000
KSC21	247.384	0.00000
KSC13	156.992	0.00000
KSC14	147.887	0.00000
KSC28	135.421	0.00000
KSC31	125.632	0.00000
KSC32	84.588	0.00000
KSC8	56.456	0.00000
KSC16	49.849	0.00000
KSC11	40.731	0.00000
KSC30	34.434	0.00000
KSC34	22.304	0.00000
KSC22	20.770	0.00000
Time	20.650	0.00000
KSC4	20.517	0.00000
KSC10	11.977	0.00000
KSC6	11.785	0.00000
KSC1	8.815	0.00000
KSC15	6.078	0.00000
KSC19	5.159	0.00001
KSC29	3.647	0.00023
KSG5	3.457	0.00035
KSC9	2.952	0.00112
KSC26	2.193	0.00641
KSC27	1.402	0.03964
KSC20	0.961	0.10931
KSC2	0.808	0.15548

### Summary of Fit

RSquare	0.24433
RSquare Adj	0.244297
Root Mean Square Error	198.0919
Mean of Response	49.37967
Observations (or Sum Wgts)	735814

### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	32	9335284932	291727654	7434.366
Error	735781	2.8872e+10	39240.419	Prob > F
C. Total	735813	3.8208e+10		<.0001*

### Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	3.8686377	0.464281	8.33	<.0001*
KSC1	0.0036001	0.000596	6.04	<.0001*
KSC2	-0.001285	0.000905	-1.42	0.1555
KSC4	0.008185	0.000865	9.46	<.0001*
KSC5	-0.004645	0.001299	-3.58	0.0003*
KSC6	0.0086125	0.00122	7.06	<.0001*
KSC7	0.0601687	0.000862	69.79	<.0001*
KSC8	0.0191534	0.001202	15.94	<.0001*
KSC9	-0.004007	0.00123	-3.26	0.0011*
KSC10	0.005815	0.000816	7.12	<.0001*
KSC11	0.0137313	0.001018	13.49	<.0001*
KSC12	-0.057067	0.000778	-73.38	<.0001*
KSC13	0.0257522	0.000962	26.76	<.0001*
KSC14	0.026971	0.001039	25.97	<.0001*
KSC15	-0.000983	0.000199	-4.93	<.0001*
KSC16	-0.011221	0.000801	-14.01	<.0001*
KSC17	0.0488653	0.001041	46.95	<.0001*
KSC18	0.0363352	0.000986	36.84	<.0001*
KSC19	0.005309	0.001181	4.50	<.0001*
KSC20	-0.001797	0.001122	-1.60	0.1093
KSC21	-0.029905	0.000889	-33.65	<.0001*
KSC22	-0.009189	0.000965	-9.52	<.0001*
KSC24	0.0368672	0.000915	40.31	<.0001*
KSC25	0.048271	0.000749	64.43	<.0001*
KSC26	0.002195	0.000805	2.73	0.0064*
KSC27	0.0021873	0.001063	2.06	0.0396*
KSC28	0.0232719	0.000937	24.84	<.0001*
KSC29	-0.003823	0.001036	-3.69	0.0002*
KSC30	0.0130781	0.001057	12.37	<.0001*
KSC31	-0.019096	0.000798	-23.92	<.0001*
KSC32	0.0259393	0.001325	19.58	<.0001*
KSC34	-0.009525	0.000964	-9.88	<.0001*
Time	8.9468e-5	9.424e-6	9.49	<.0001*

Air University: The Intellectual and Leadership Center of the Air Force

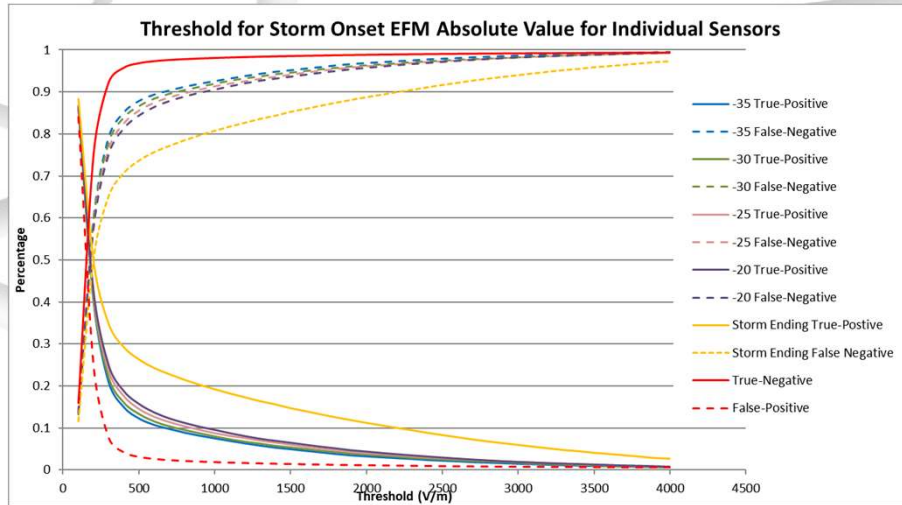
Aim High...Fly - Fight - Win

A Least Squares Fit regression model using individual sensor readings and time offer the most explanation of variation in data, but still only accounts for up to 25% of the variation. This is NOT a good fit for the data.



## Threshold Confusion Matrix for Individual Sensors

The AFIT of Today is the Air Force of Tomorrow.



- Inversely proportional: As True-Negative predictions get better, True-Positive predictions get worse

Air University: The Intellectual and Leadership Center of the Air Force

Aim High...Fly - Fight - Win

True-Negative measures voltage below threshold when there are no lightning storms presenting.

True-Positive measures voltage above threshold when lightning storms are occurring.

False-Positive measure voltage above threshold when there are no lightning storms presenting.

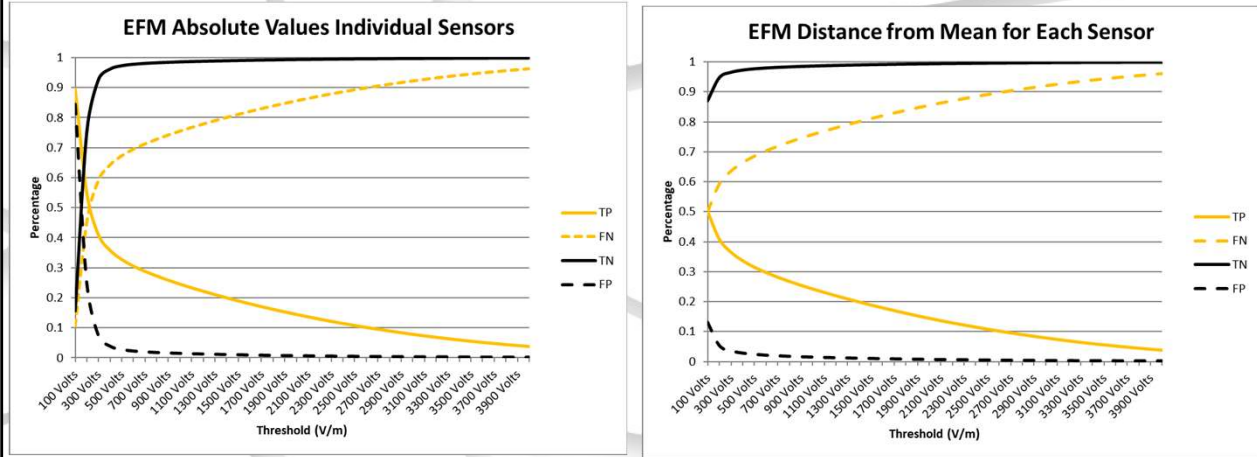
False-Negative measure voltage below threshold when lightning storms are occurring.



## Confusion Matrices for Individual Sensor Locations



The AFIT of Today is the Air Force of Tomorrow.



- No-Storm thresholds below 400 V/m absolute value or 300 V/m from sensor mean gives <95% “True” result
- Storm threshold gives best “True” result at 100 V/m for individual sensors (not useful)
- False-Positive threshold is less than 1% at 1500 V/m

Air University: The Intellectual and Leadership Center of the Air Force

Aim High...Fly - Fight - Win

Each individual sensor is tested to determine if the voltage reading exceeds the threshold value from 100 V/m to 4000 V/m.

True Positive is considered when a storm is occurring within 5 NM of the sensor and EFM voltage exceeds threshold.

False Positive is considered when no-storm is occurring within 5 NM of the sensor, but EFM voltage exceeds threshold.

True Negative is considered when no-storm is occurring within 5 NM of the sensor and EFM voltage is below threshold.

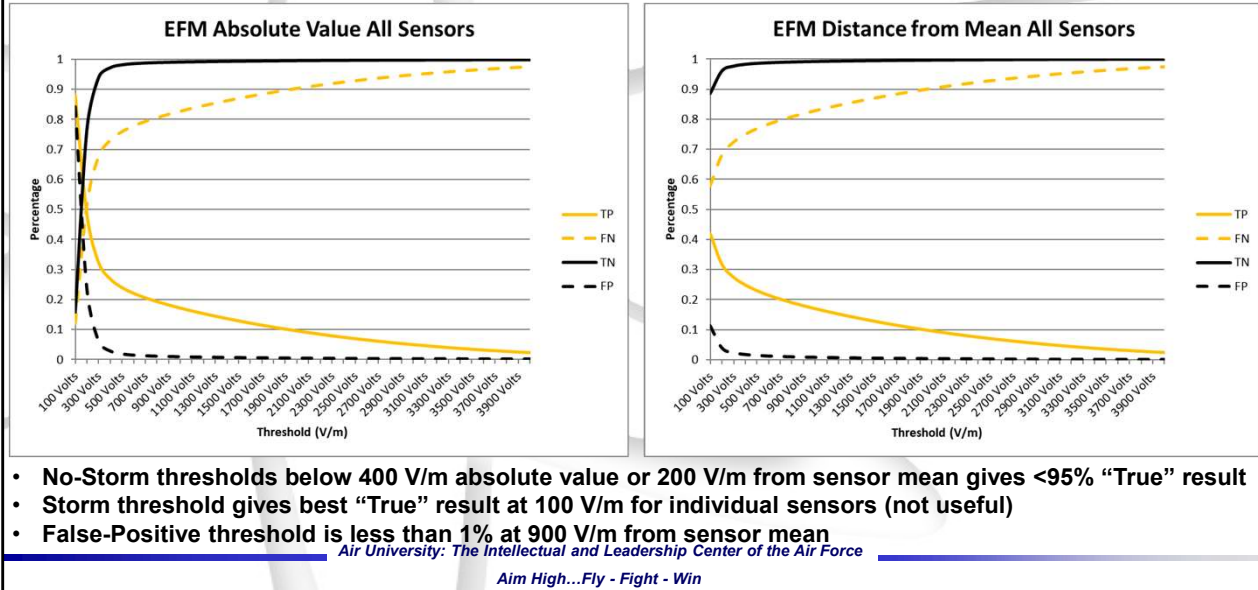
False Negative is considered when a storm is occurring within 5 NM of the sensor, but EFM voltage is below threshold.





## Confusion Matrices for Entire Sensor Area

The AFIT of Today is the Air Force of Tomorrow.



Each individual sensor is tested to determine if the voltage reading exceeds the threshold value from 100 V/m to 4000 V/m.

True Positive is considered when a storm is occurring anywhere in the vicinity of CCAFS and EFM voltage exceeds threshold.

False Positive is considered when no-storm is occurring anywhere in the vicinity of CCAFS, but EFM voltage exceeds threshold.

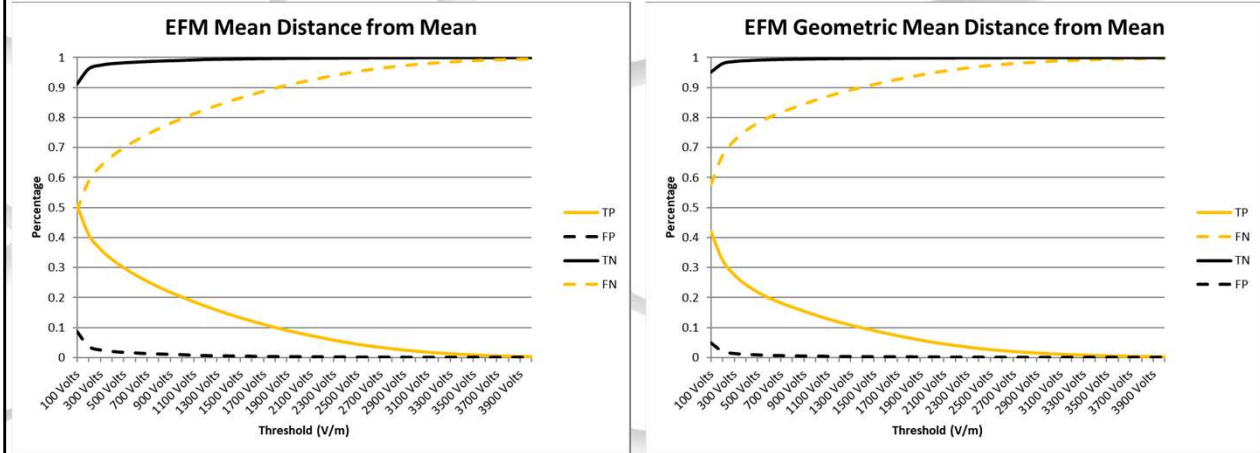
True Negative is considered when no-storm is occurring anywhere in the vicinity of CCAFS and EFM voltage is below threshold.

False Negative is considered when a storm is occurring anywhere in the vicinity of CCAFS, but EFM voltage is below threshold.



## Confusion Matrices for Means Over Entire Sensor Area

The AFIT of Today is the Air Force of Tomorrow.



- No-Storm thresholds below 200 V/m mean from sensor mean or 100 V/m geometric mean from sensor mean gives <95% "True" result
- Storm threshold gives best "True" result at 100 V/m for individual sensors (not useful)
- False-Positive threshold is less than 1% at 1500 V/m

Air University: The Intellectual and Leadership Center of the Air Force

Aim High...Fly - Fight - Win

The mean voltage and geometric mean voltage of all sensors are tested to determine if the voltage reading exceeds the threshold value from 100 V/m to 4000 V/m.

True Positive is considered when a storm is occurring anywhere in the vicinity of CCAFS and EFM mean/geometric mean voltage exceeds threshold.

False Positive is considered when no-storm is occurring anywhere in the vicinity of CCAFS, but EFM mean/geometric mean voltage exceeds threshold.

True Negative is considered when no-storm is occurring anywhere in the vicinity of CCAFS and EFM mean/geometric mean voltage is below threshold.

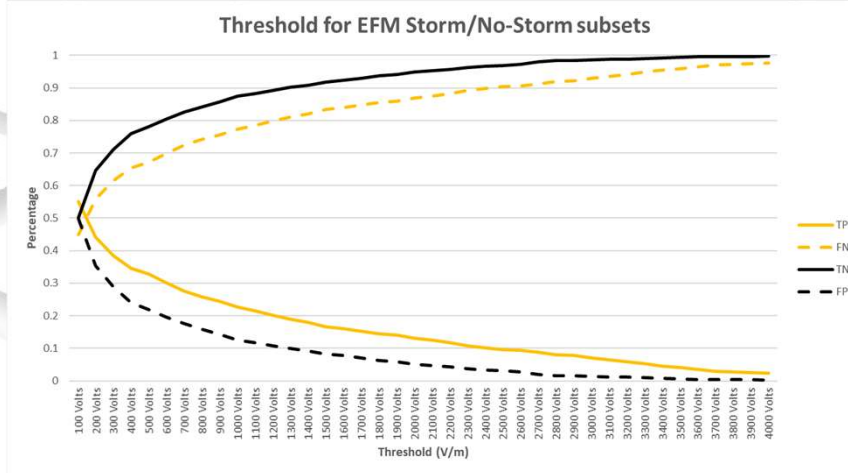
False Negative is considered when a storm is occurring anywhere in the vicinity of CCAFS, but EFM mean/geometric mean voltage is below threshold.



# Storm/No-Storm Threshold Confusion Matrix for Entire Sensor Area



The AFIT of Today is the Air Force of Tomorrow.



- No-Storm threshold below 2100 V/m mean from sensor mean gives <95% “True” result
- Storm threshold gives best “True” result at 100 V/m, just over 55% of all storms above threshold

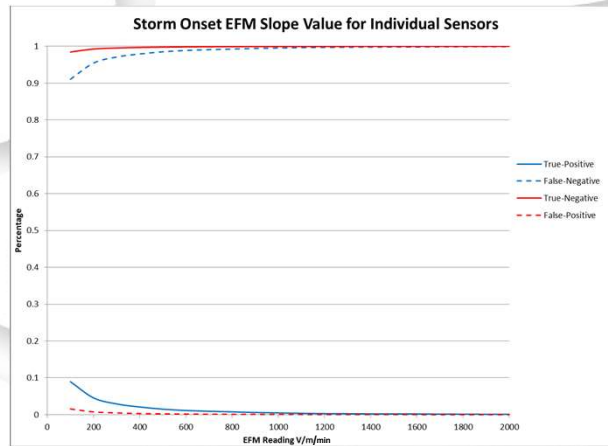
Air University: The Intellectual and Leadership Center of the Air Force

Aim High...Fly - Fight - Win



## Slope Threshold Confusion Matrix for Individual Sensors

The AFIT of Today is the Air Force of Tomorrow.



- Change in voltage from minute to minute is also not a good indicator for lightning onset
- 99% confidence that change in voltages below 500 V/m/min indicate no-storms

Air University: The Intellectual and Leadership Center of the Air Force

Aim High...Fly - Fight - Win



## Preliminary Results



*The AFIT of Today is the Air Force of Tomorrow.*

- **This is still largely a work in progress**
  - **Large timeseries data set reduced and noise-filtered using mean values**
  - **Least Squares Fit of EFM data only accounts for up to 25% of variance**
    - **Not a good fit to predict if lightning will or is occurring**
  - **Threshold analysis shows inverse relation w/True-Positive & True-Negative outcomes with increasing threshold voltages**
  - **EFM readings alone are NOT a useful tool for predicting lightning**
    - **Appear to be reactive, not predictive**

**Next phase: predict time to end of storm**

*Air University: The Intellectual and Leadership Center of the Air Force*

*Aim High...Fly - Fight - Win*



## Next Steps



*The AFIT of Today is the Air Force of Tomorrow.*

- **Setup lightning-ending prediction columns (complete)**
- **Subset data frame into storm-only datasets for individual storms**
- **Perform negative binomial regression analysis for each sensor on each storm**
  - **EFM readings vs. storm ending time**
  - **Lightning strikes vs. storm ending time**
- **Develop negative binomial regression model to compute time to storm end**

*Air University: The Intellectual and Leadership Center of the Air Force*

*Aim High...Fly - Fight - Win*



# Abstract



*The AFIT of Today is the Air Force of Tomorrow.*

# Questions?

*Air University: The Intellectual and Leadership Center of the Air Force*

*Aim High...Fly - Fight - Win*



## XTS Plot



*The AFIT of Today is the Air Force of Tomorrow.*

- EFM Mean Voltage and Lightning w/in 5 nm of any sensor



dygraphmeanall.html

*Air University: The Intellectual and Leadership Center of the Air Force*

*Aim High...Fly - Fight - Win*

24

Times of interest:

9 – 14 Aug 2013 – EFM reading close to average while lightning occurring

13 – 19 Jun 2016 – EFM readings spike after lightning has started

14 – 17 Jun 2013, 3 Jul – 7 Jul 2013 – EFM readings spike but no lightning occurring

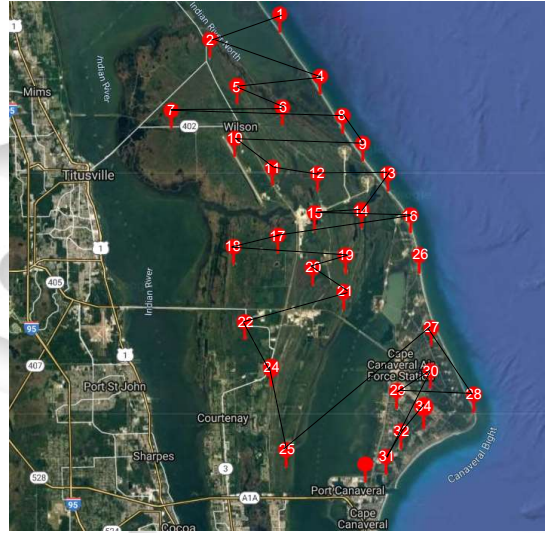








*The AFIT of Today is the Air Force of Tomorrow.*



*Air University: The Intellectual and Leadership Center of the Air Force*

*Aim High...Fly - Fight - Win*