

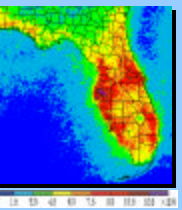


A Statistical Procedure to Forecast Warm Season Lightning over Portions of the Florida Peninsula

Phillip E. Shafer and Henry E. Fuelberg
The Florida State University, Tallahassee, Florida



1. Introduction



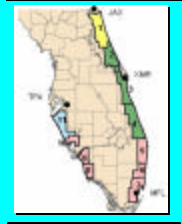
- Florida leads the nation in cloud to ground (CG) lightning strikes
- Atlantic and Gulf of Mexico sea breezes influence lightning patterns
- Lightning causes power disruptions - problematic for power companies such as Florida Power and Light Corporation (FP&L)
- Accurate forecasts of the location and amount of lightning can reduce manpower and repair costs

Study Objectives:

- Develop a statistical scheme that provides improved forecast guidance for warm season CG lightning within eleven coastal areas of the Florida Peninsula that are serviced by FP&L
- Non-midnight time period
- Warm season months: June through August
- Logistic regression equations predict whether at least one flash will occur, and the amount of lightning conditional on at least one flash occurring

2. Data

Study Areas:



Radiosonde Data:

- 1200 UTC sounding data for JAX, XMR, MFL, and TPA
- 597 parameters calculated to serve as candidate predictors
- Layer averaged wind direction and speed, stability, moisture
- Assumptions: 1) Conditions will not change significantly from the sounding time through the end of the forecast period. 2) No spatial variations within each radiosonde domain

Lightning Data:

- National Lightning Detection Network (NLDN)
- 16 years (1989-2004)
- Daily noon - midnight and 6am - noon counts tabulated for each area

3. Equation Development

Predictands:

- To develop equations that predict whether at least one flash will occur, binary lightning indicators were assigned to each day in the dataset, i.e., "1" if at least one flash was observed, or a "0" otherwise
- To predict the amount of lightning, flash counts in each area first were subdivided into 4 quartile groups:

| Forecast Area | Q1 | Q2 | Q3 | Q4 |
|---------------------------|------|--------|---------|---------|
| Miami-Dade | 1-11 | 12-22 | 23-36 | 37-50 |
| Broward | 1-7 | 8-12 | 13-19 | 20-25 |
| Palm Beach | 1-11 | 12-27 | 28-35 | 36-43 |
| Manatee/St. Lucie/Ind. R. | 1-18 | 19-113 | 114-278 | 279-378 |
| Brevard | 1-11 | 12-27 | 28-35 | 36-43 |
| Volsia | 1-16 | 17-29 | 30-34 | 35-44 |
| Flagler/St. Johns | 1-18 | 19-115 | 116-264 | 265-444 |
| Manatee/Sarasota | 1-20 | 21-68 | 69-209 | 210-339 |
| Charlotte | 1-6 | 7-31 | 32-103 | 104-113 |
| Lee | 1-12 | 13-46 | 47-107 | 108-117 |
| Collier | 1-13 | 14-46 | 47-124 | 125-134 |

Binary Logistic Regression:

- Best results were achieved having separate equations to determine:
 1. Probability of a Q1 event
 2. Probability of = Q3 event
 3. Probability of a Q4 event
- One logistic regression model for each binary outcome:

$$\ln \left[\frac{p}{1-p} \right] = f(X_n)$$

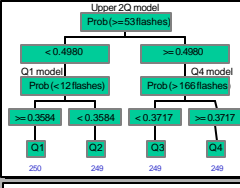
$$f(X_n) = b_0 + b_1 X_1 + \dots + b_k X_k$$

$$p = \exp(f(X_n)) / [1 + \exp(f(X_n))]$$

Probability bounded between 0-1

- Less stringent assumptions than linear regression methods
- Statistical Package for the Social Sciences (SPSS) used
- Forward stepwise selection
- Test for backward elimination
- Decision trees constructed to determine the most likely quartile based on probability thresholds for the three logistic equations
- Thresholds chosen so that the scheme would not have a prediction bias toward any one quartile

Quartile Scheme: Miami-Dade



| Obs. | Q1 | Q2 | Q3 | Q4 | hit rate | within 10 |
|--------------|-----|----|----|-----|----------|-----------|
| Q1 | 117 | 77 | 34 | 22 | 47% | 78% |
| Q2 | 74 | 82 | 61 | 33 | 33% | 87% |
| Q3 | 39 | 57 | 77 | 73 | 31% | 84% |
| Q4 | 20 | 33 | 77 | 121 | 48% | 79% |
| Total: | | | | | 40% | 82% |
| Persistence: | | | | | 36% | 75% |
| Skill score: | | | | | 6.4% | 28.7% |

- Scheme best forecasts Q1 & Q4 events, and predicting to within one quartile of the observed
- Scheme more skillful than persistence (% improvement over persistence is positive)

Skill Scores for All Areas

| Forecast Area | Quartile Hit Rate | Quartile Persistence | Skill Score |
|---------------------------|-------------------|----------------------|-------------|
| Flagler/St. Johns | 0.46 | 0.33 | 20.0% |
| Brevard | 0.40 | 0.30 | 14.4% |
| Manatee/St. Lucie/Ind. R. | 0.40 | 0.31 | 13.3% |
| Manatee/Sarasota | 0.40 | 0.32 | 11.2% |
| Collier | 0.40 | 0.32 | 11.2% |
| Lee | 0.38 | 0.31 | 10.9% |
| Volsia | 0.38 | 0.31 | 9.2% |
| Palm Beach | 0.40 | 0.34 | 8.1% |
| Broward | 0.38 | 0.32 | 8.1% |
| MiamiDade | 0.40 | 0.36 | 6.4% |
| Charlotte | 0.35 | 0.31 | 6.4% |

K-fold Cross-Validation

- Results for dependent data do not assess how well the equations will predict cases that were not involved in model development
- A k-fold cross-validation was performed for the Flagler/St. Johns and Charlotte forecast areas
- Results showed only a slight reduction in skill compared to the dependent data, and still were more skillful than persistence!

4. Results

Final Prediction Equations Example: Miami-Dade Area

Probability of at least one CG flash:

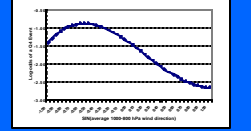
| Predictor | Coefficient | P-value |
|----------------------|-------------|---------|
| SSI | -0.170 | 0.000 |
| Sin(wdir) | -1.001 | 0.000 |
| MNSPD | -0.094 | 0.000 |
| MNSPD ² | 0.020 | 0.000 |
| Morn yes/no | 1.240 | 0.000 |
| PD yes/no | 0.919 | 0.000 |
| (PD yes/no) x (T925) | 0.297 | 0.004 |
| Constant | 1.325 | 0.000 |

Conditional probability of a Q4 event:

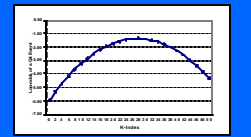
| Predictor | Coefficient | P-value |
|-------------------------|-------------|---------|
| KI | -0.310 | 0.007 |
| Sin(wdir) | -1.707 | 0.000 |
| Sin(wdir) ² | 0.340 | 0.000 |
| PD quartile | 0.263 | 0.000 |
| (PD quartile) x (SSI) | -0.104 | 0.019 |
| (PD quartile) x (MNSPD) | -0.131 | 0.000 |
| Constant | 1.390 | 0.000 |

- All predictors statistically significant at = 95% level (p-value = 0.05)
- Sin(wdir): sine of the layer averaged wind direction has strongest influence
- Showalter stability index (SSI) and K-index (KI) are important
- Non-linear and interaction effects also were selected

Non-Linear Relationships:



- Log-odds of a Q4 event is maximized for sin(wdir) between -0.65 and -0.45, or wind directions between 205° -220° (SW)
- Opposing flow results in greater convergence along sea breeze front



- Likelihood of Q4 event peaks for KI values between 25 and 30° C, then declines for larger KI

Results for Dependent Data:

Yes/No Equation: Miami-Dade

| Model Forecast | Model Performance |
|--------------------------|------------------------------|
| Observed Yes | Yes: 915, No: 82, CSI: 0.77 |
| Observed No | Yes: 190, No: 183, FAR: 0.17 |
| POD: 0.92 | |
| Bias: 1.10 | |
| Persistence CSI: 0.69 | |
| Hit rate: 0.80 | |
| Model skill score: 25.4% | |

Skill scores for all eleven areas

| Forecast Area | Model CSI | Persistence CSI | Skill score |
|---------------------------|-----------|-----------------|-------------|
| Volsia | 0.75 | 0.64 | 32.0% |
| Flagler/St. Johns | 0.77 | 0.66 | 31.6% |
| Brevard | 0.78 | 0.69 | 27.0% |
| Collier | 0.77 | 0.69 | 25.7% |
| MiamiDade | 0.77 | 0.69 | 25.4% |
| Palm Beach | 0.76 | 0.68 | 25.3% |
| Broward | 0.74 | 0.65 | 24.5% |
| Manatee/Sarasota | 0.77 | 0.71 | 23.6% |
| Lee | 0.75 | 0.68 | 21.4% |
| Manatee/St. Lucie/Ind. R. | 0.80 | 0.75 | 20.5% |
| Charlotte | 0.72 | 0.66 | 15.6% |

5. Conclusions & Future Work

SUMMARY & CONCLUSIONS

- Speed and direction of low-level wind was found to be the dominant factor contributing to the likelihood of lightning in each area
- Guidance for all eleven areas superior to persistence
- Considerable skill predicting whether at least one CG flash will occur
- Quartile scheme does best predicting to within one quartile of the observed
- Also does well distinguishing upper two quartile events from lower two quartile events
- Guidance is currently being used operationally by FP&L during the warm season

FUTURE WORK

- Use of morning soundings was based on several assumptions that are not always valid
- Interactions with smaller scale circulations (i.e., lake and river breezes) and cloud microphysics were not considered
- Future work will utilize mesoscale model output to create spatial forecast fields of lightning probability and amount for the entire Florida Peninsula
- Model data will be more location and time specific than a static morning sounding
- Incorporation of model-derived cloud microphysics hopefully can be related to charging mechanisms