

Monte Carlo Simulations for Evaluating the Accuracy of GLM Detection Efficiency and False Alarm Rate Retrievals

Katrina Virts

University of Alabama in Huntsville (UAH)

William Koshak

National Aeronautics and Space Administration (NASA)

GLM Science Team Meeting

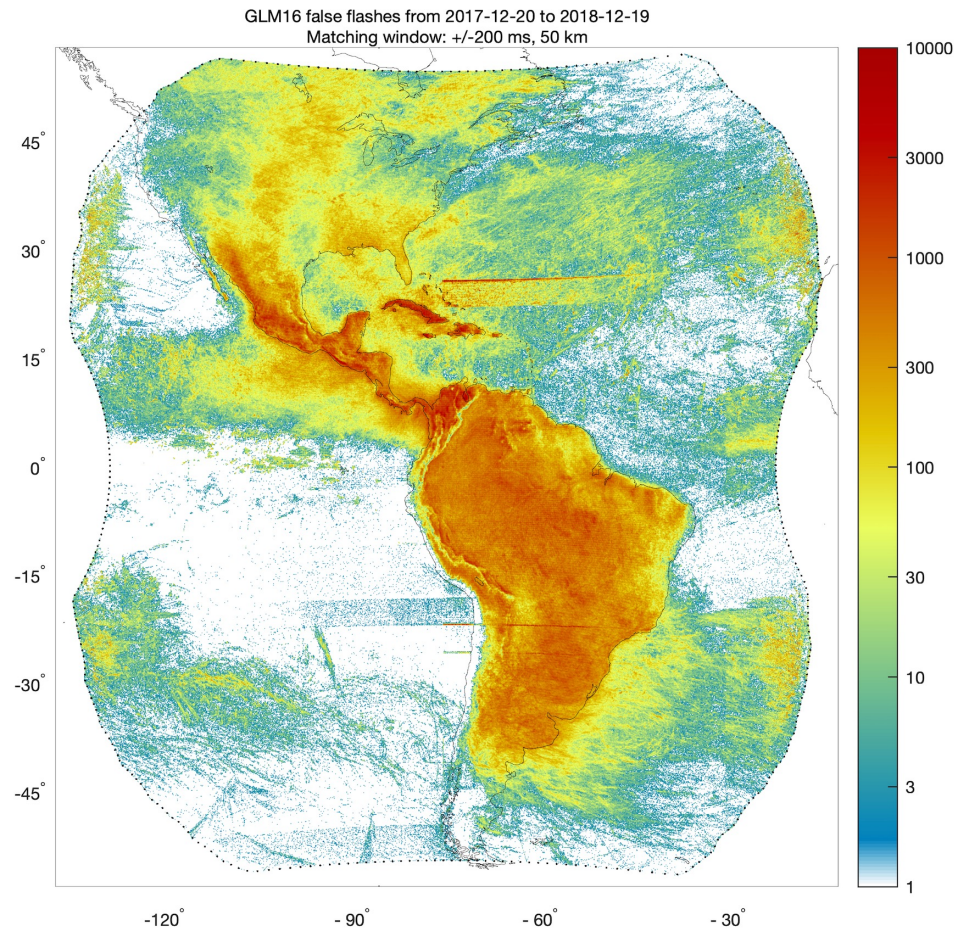
13 September 2022

Motivation

- Assessing GLM's detection performance throughout its FOV is complicated by the fact that the detection performance of the available reference networks is both **imperfect and imperfectly known**
- Two key performance metrics are:
 - **Detection efficiency (DE)** = % of true lightning flashes detected by GLM
 - **False alarm rate (FAR)** = % of GLM flashes that are not true lightning
- If GLM detects a flash that the ground networks don't, GLM is incorrectly assumed to be false alarming

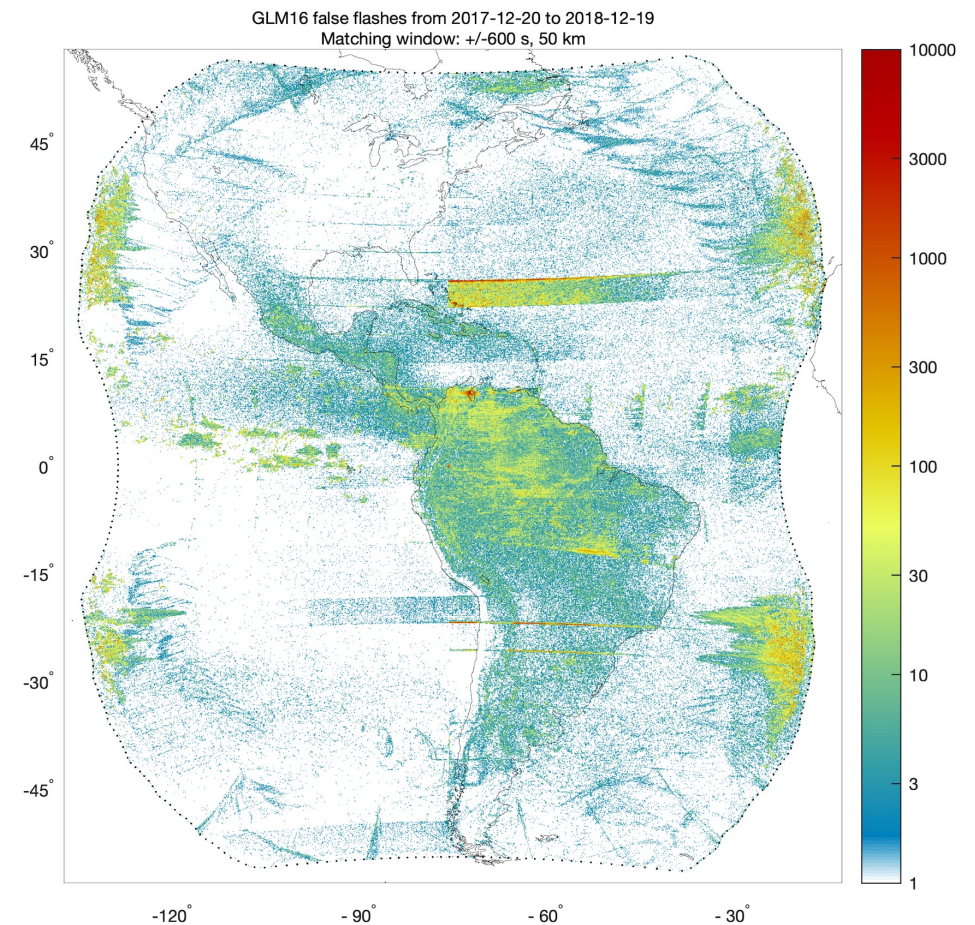
Motivation

- Example: construct a climatology of GLM artifacts (flashes that cannot be matched to any ground network detection)
- Matching criteria: ± 200 ms and 50 km between flash centroids
- Resembles the overall lightning climatology (and artifacts)



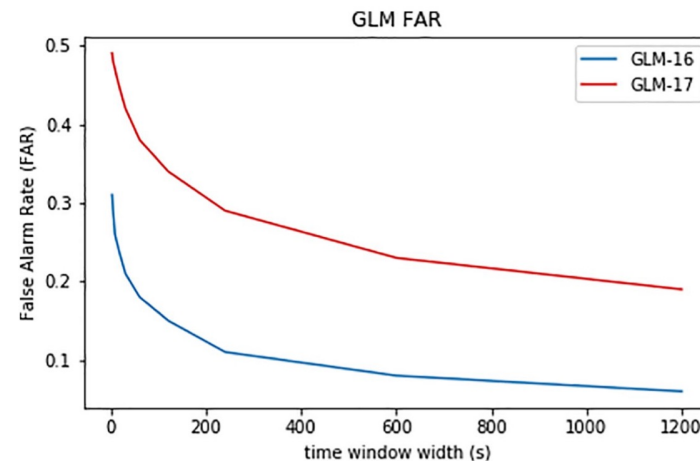
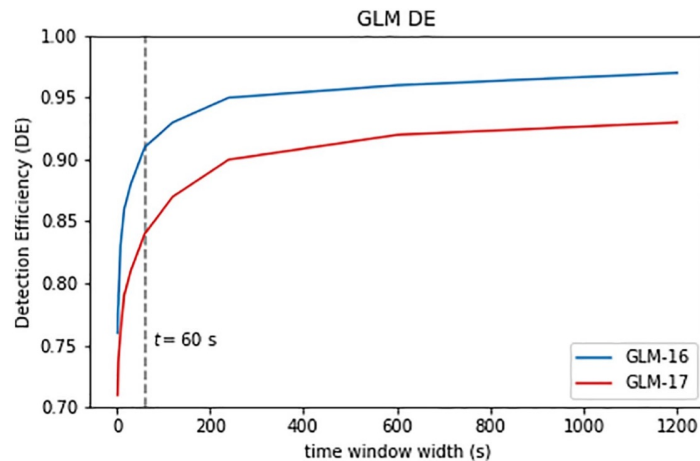
Motivation

- Example: construct a climatology of GLM artifacts (flashes that cannot be matched to any ground network detection)
- Matching criteria: ± 10 min and 50 km between flash centroids
- Recognizable pattern of artifacts (Bahama bar, blooming areas, etc.)



Motivation

- Bateman et al. (2021) calculated GLM DE and FAR as a function of the temporal matching criteria:

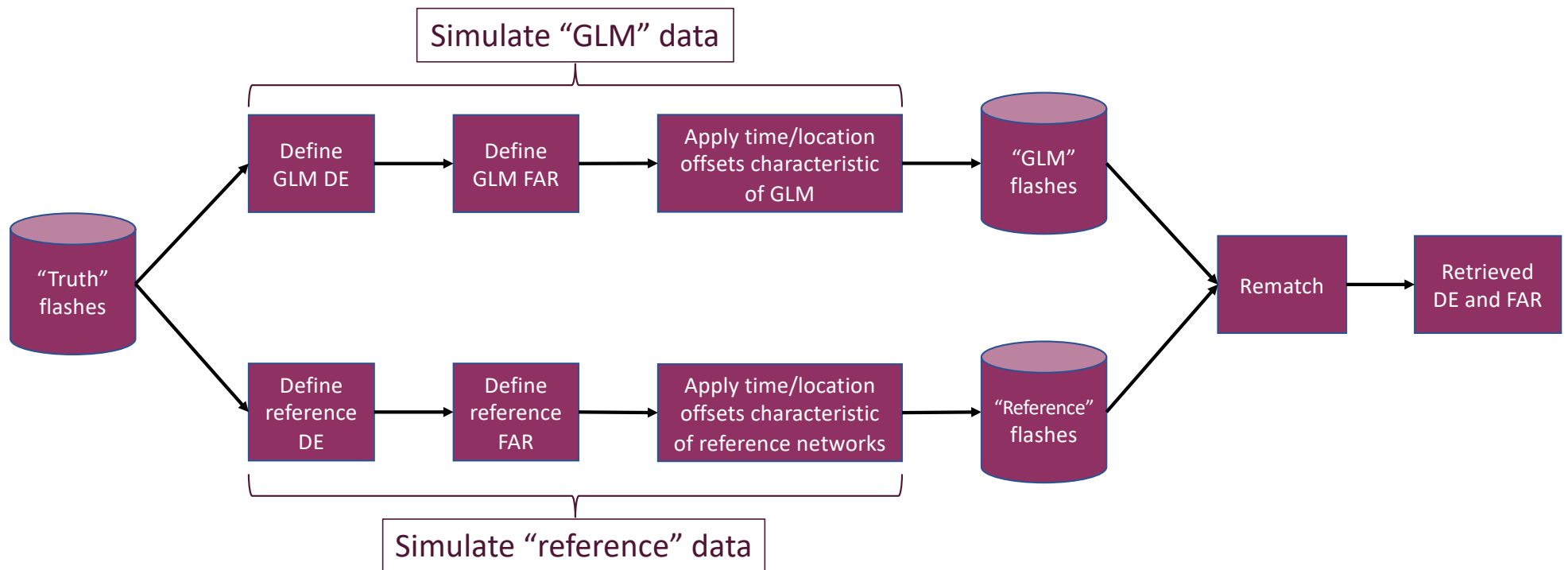


- DE was as high as >90% for both GLM-16 and GLM-17, and FAR as low as ~5% (GLM-16) to ~20% (GLM-17) when using a ± 10 min matching window
- Question: what matching criteria provide the most accurate estimate of GLM performance?

Figures courtesy Bateman et al. (2021)
<https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2020EA001237>

Simulation Procedure

- The challenge: Estimate uncertainty in the retrieved GLM performance metrics using Monte Carlo computer simulations:



Simulation Specifics

| Simulation Type | Vary temporal matching criteria and GLM performance | Vary temporal matching criteria and reference performance |
|----------------------|--|--|
| Simulation Specifics | <p>GLM DE = 0.3, 0.4, ... 1</p> <p>GLM FAR = 0, 0.1, 0.2</p> <p>Reference DE = 0.7</p> <p>Reference FAR = 0.05</p> <p>GLM and reference offsets = observed</p> <p>Distance matching criteria = standard (50 km)</p> <p>Temporal matching criteria = 0.3, 1, 3, 10, ..., 3000 x standard window of ± 200 ms</p> | <p>GLM DE = 0.7</p> <p>GLM FAR = 0.05</p> <p>Reference DE = 0.3, 0.4, ... 1</p> <p>Reference FAR = 0, 0.1, 0.2</p> <p>GLM and reference offsets = observed</p> <p>Distance matching criteria = standard (50 km)</p> <p>Temporal matching criteria = 0.3, 1, 3, 10, ..., 3000 x standard window of ± 200 ms</p> |

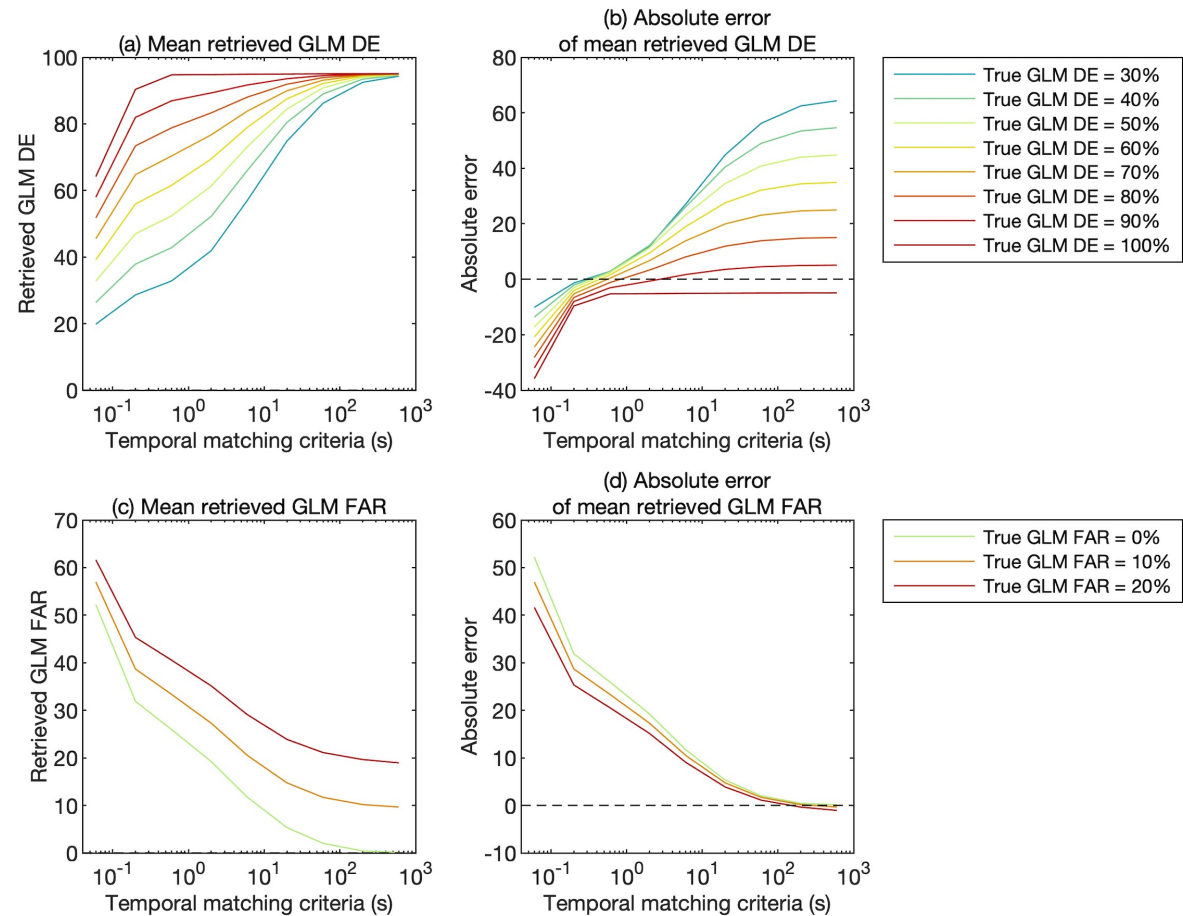
Vary Matching Criteria and GLM Performance

For DE:

- Retrieved DE converges to 95% for broad matching windows regardless of true GLM DE, reflecting the true reference FAR of 5%.
- Absolute errors of 0 occur for criteria ranging from ± 0.3 s (low true GLM DE) up to ± 3 s (high true GLM DE).

For FAR:

- Absolute errors converge to 0 for temporal windows broader than $\sim \pm 100$ s.



Vary Matching Criteria and Reference Performance

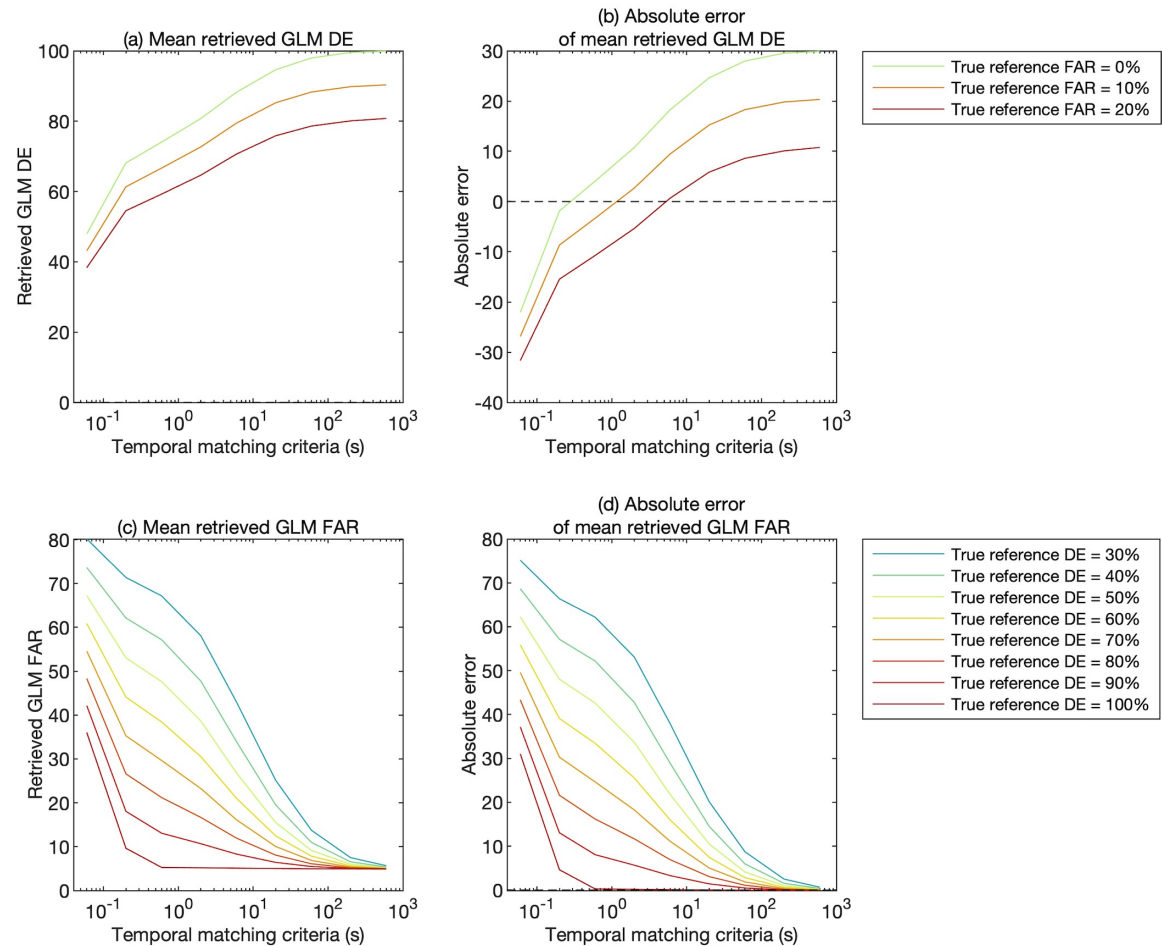
For DE:

- Absolute errors of 0 occur for criteria ranging from ± 0.3 s (low true reference FAR) up to ± 5 s (high true reference FAR).

For FAR:

- For all scenarios except unrealistically perfect reference DE, broad matching windows on the order of minutes are needed before absolute errors converge to 0.

Bottom line: **the optimal matching criteria are different for DE and FAR.** Broad matching criteria overestimate true GLM DE but are necessary to correctly estimate GLM FAR.



Conclusions

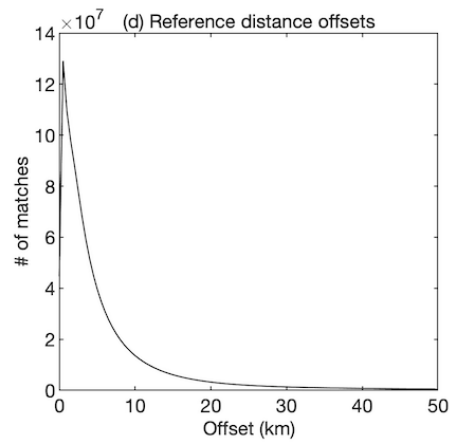
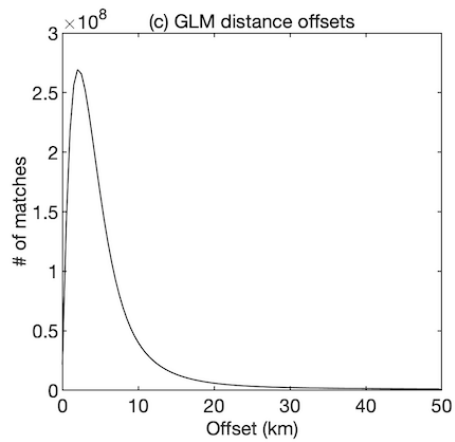
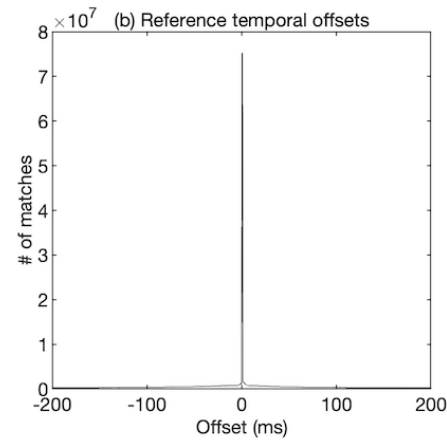
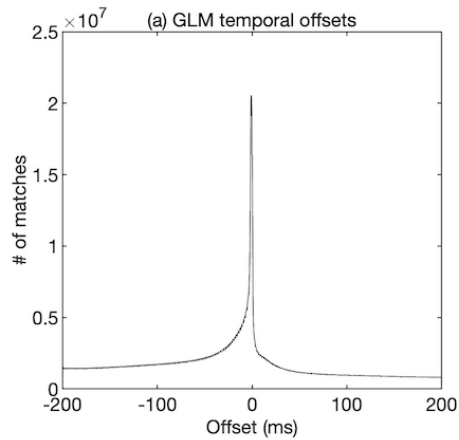
- Calculated GLM performance metrics (DE and FAR) depend on the temporal matching criteria applied
- Simulations provide a means of quantifying which matching criteria most accurately estimate the "true" GLM performance metrics
- The optimal matching criteria from these simulations appear to be:
 - For DE: on the order of ± 1 s
 - For FAR: on the order of minutes (the exact criteria selected are less important)
- Implications for the detection performance reported by Bateman et al. (2021) for matching criteria up to ± 10 minutes:
 - Reported DE values of $>90\%$ are almost certainly too high
 - Reported FAR values of $\sim 5\%$ (GLM-16) and $\sim 20\%$ (GLM-17) are more likely to be accurate

Thank You!

Katrina Virts

katrina.virts@uah.edu

Characteristic Time/Location Offsets



Observed offsets derived for 2021

For simulations, “truth” flashes = daily
observed GLM-16 flashes every 5 days
during 2020

Vary Temporal Matching Criteria

Simulation specifics:

GLM DE = 0.7 (spec)

GLM FAR = 0.05 (spec)

Reference DE = 0.7

Reference FAR = 0.05

GLM and reference offsets = observed

Distance matching criteria = standard (50 km)

Temporal matching criteria = 0.3, 1, 3, 10, ..., 3000 x standard window of ± 200 ms

Results:

- Steep curve for criteria near the standard values \rightarrow sensitivity of performance metrics when strict flash-to-flash matching is required
- Strictest matching criteria underestimate true GLM DE by $\sim 25\%$, with 50% error in retrieved GLM FAR
- Broadest matching criteria overestimate true GLM DE, asymptoting to 95% (reflecting the true reference FAR)
- Retrieved GLM FAR asymptotes to the true value of 5%

What about other GLM and reference performance scenarios?

