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

Katrina Virts (UAH) and William Koshak (NASA)

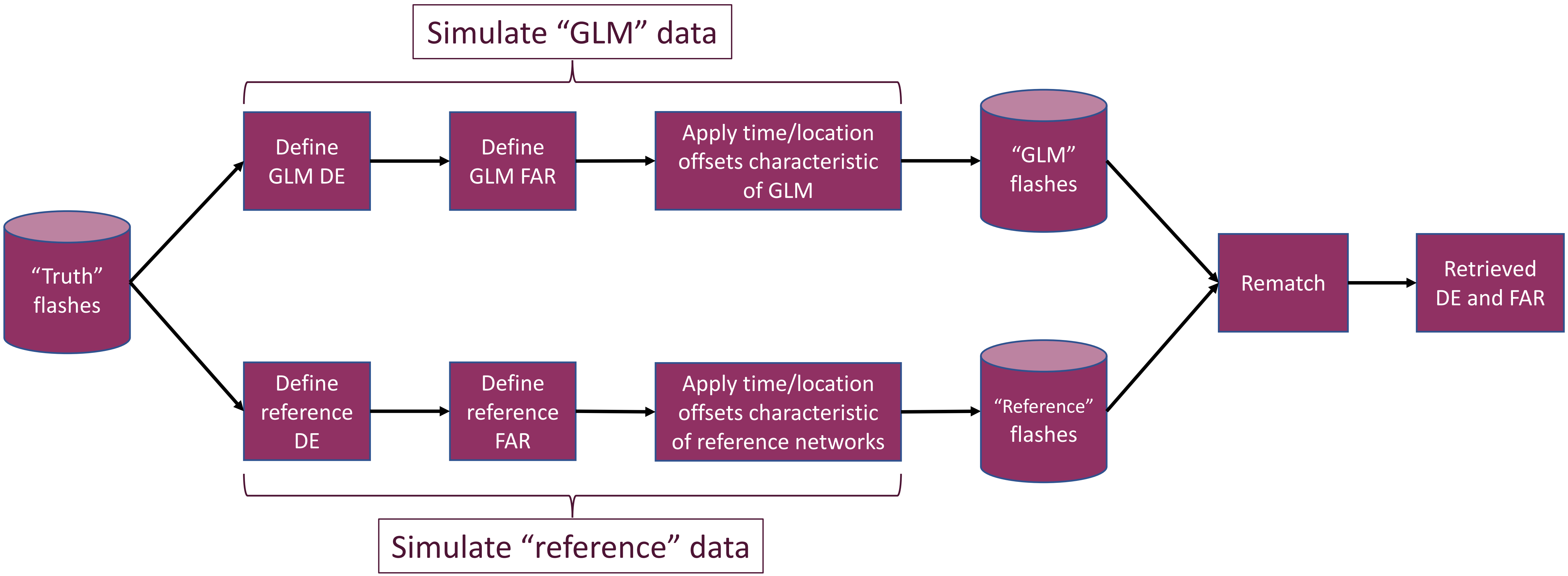
GLM Science Team Meeting

22 September 2021

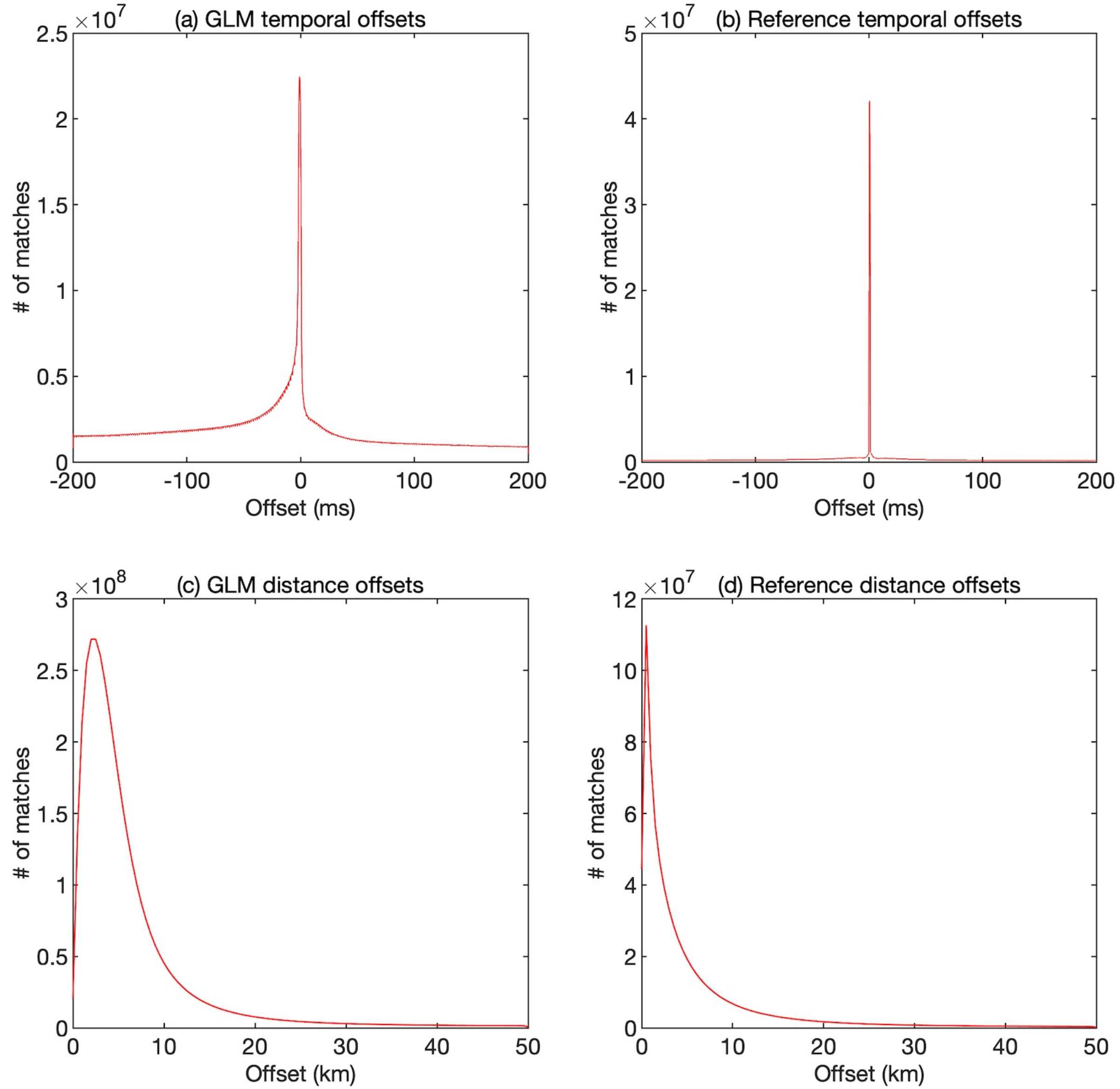
Estimating Errors in DE and FAR

- Analysis of GLM DE and FAR (Bateman et al. 2021; Lang et al. 2020; Murphy and Said 2020; Rutledge et al. 2020; Zhang and Cummins 2020; others in this section) has focused on GLM performance as a function of geography, storm type, cloud characteristics, time of day, flash size and duration, etc.
- Other factors that can affect the calculated GLM DE and FAR:
  - Reference network DE and FAR
  - Location and timing accuracy of GLM and reference networks
  - Geographic variations in reference network performance
  - Spatial and temporal matching criteria
- The challenge: Estimate uncertainty in the retrieved GLM DE and FAR as a function of these factors by means of Monte Carlo computer simulations

Monte Carlo Simulation Procedure



Characteristic Time/Location Offsets



Observed offsets derived from July 2019 – June 2020

GLM offsets derived by seeking “best reference match” for each GLM group

Reference offsets derived by matching reference pulses/strokes from two networks to each other

Analysis is “reference network-agnostic” – the simulated reference data contain characteristics of both real-world networks

“Truth” flashes = daily observed GLM-16 flashes from July – September 2020

Simulation: Vary GLM DE and FAR

Simulation specifics:

GLM DE = 0.4, 0.5, …, 1

GLM FAR = 0, 0.1, …, 0.4

Reference DE = 0.4, 0.7, 0.9

Reference FAR = 0.05

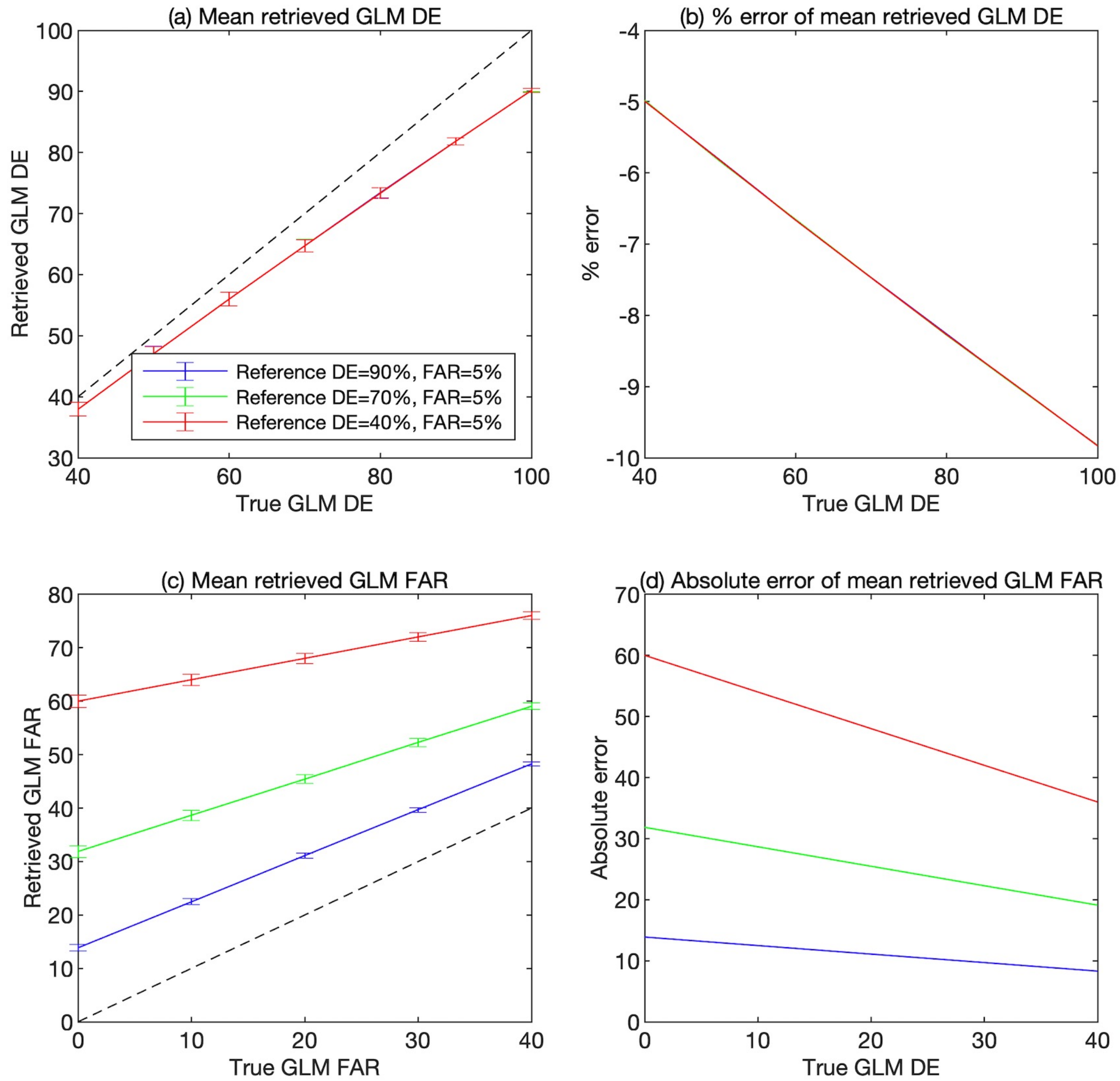
GLM offsets = observed

Reference offsets = observed

Matching criteria = standard

Results:

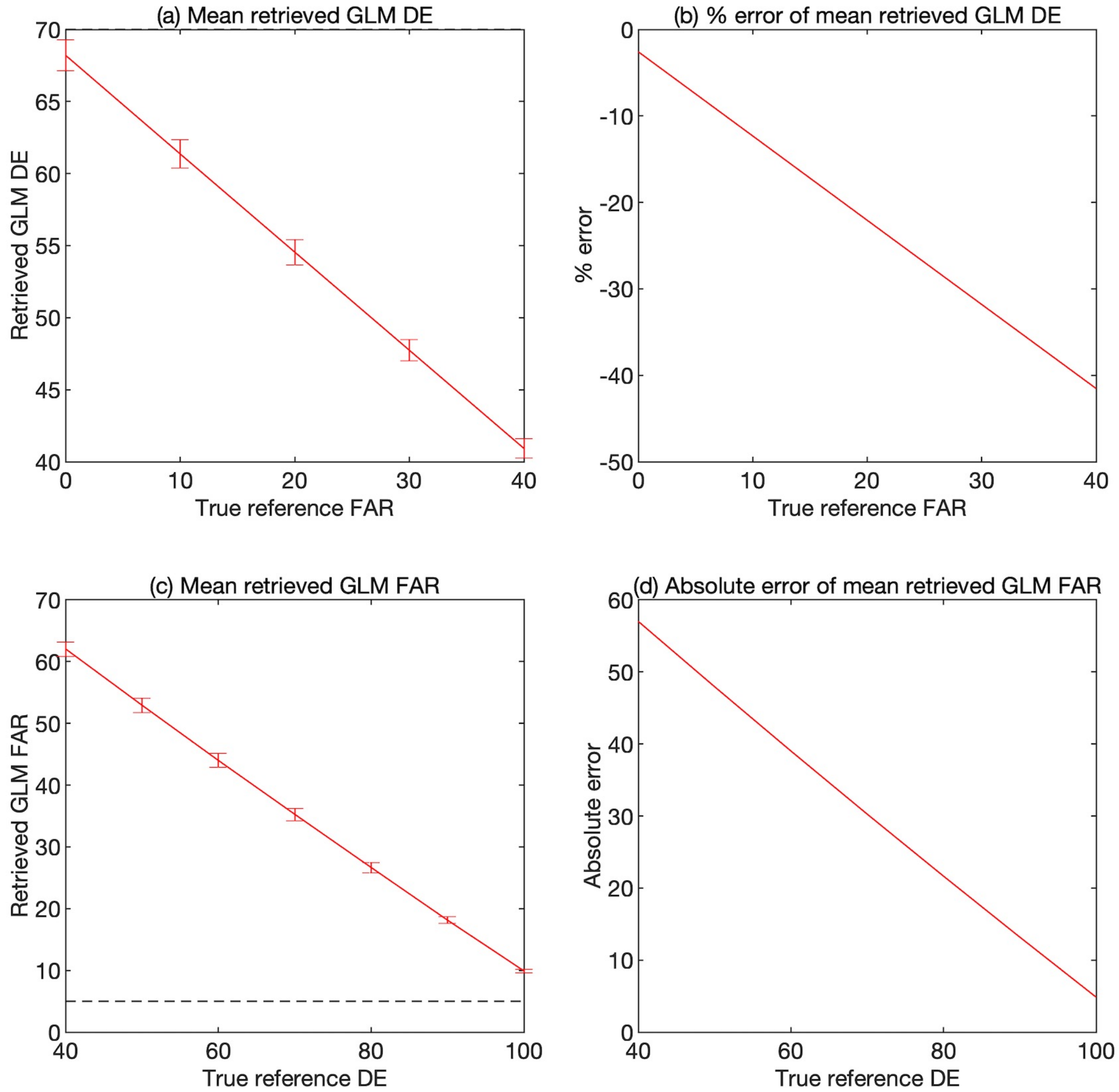
- Retrieved GLM DE is ~identical regardless of reference DE
- Retrieved GLM DE underestimates the true DE (% error from -5% to -9.8%)
- Retrieved GLM FAR varies with both true GLM FAR and true reference DE
- Absolute error of retrieved FAR ranges from 9-13% (when reference = great) to 35-60% (when reference = poor)



Simulation: Vary Reference DE and FAR

Simulation specifics:
GLM DE = 0.7 (spec)
GLM FAR = 0.05 (spec)
Reference DE = 0.4, 0.5, …, 1
Reference FAR = 0, 0.1, …, 0.4
GLM offsets = observed
Reference offsets = observed
Matching criteria = standard

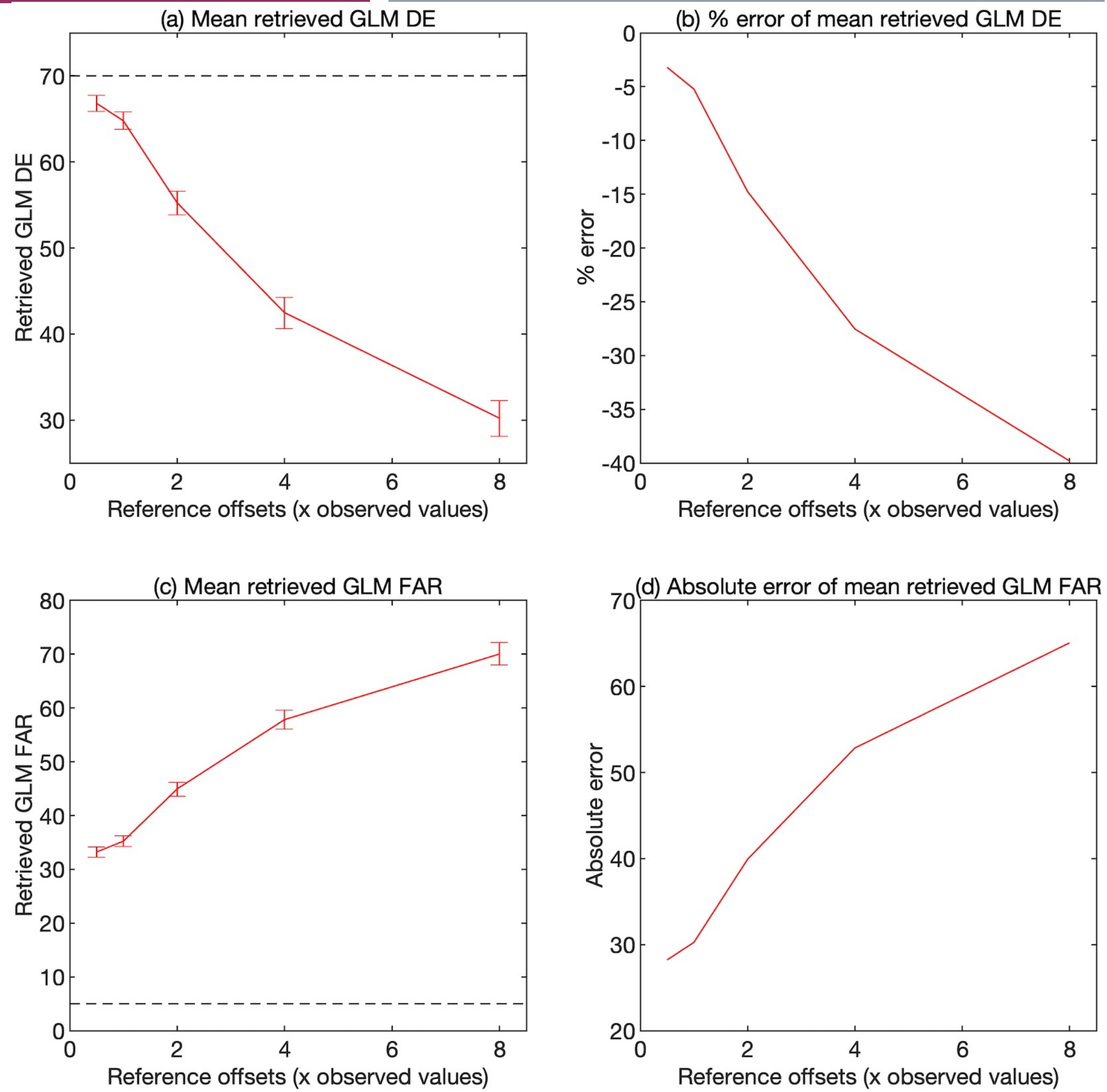
Results:
- Retrieved GLM DE underestimates the true DE (% error up to 41%)
- Absolute error of retrieved GLM FAR ~5% for true reference DE=100% but ranges up to ~57% for true reference DE=40%



Simulation: Vary Reference Offsets

Simulation specifics:
GLM DE = 0.7 (spec)
GLM FAR = 0.05 (spec)
Reference DE = 0.7
Reference FAR = 0.05
GLM offsets = observed
Reference offsets = 0.5, 1, 2, 4, 8 x observed
Matching criteria = standard

Results:
- For offsets 50-100% of observed, % error of retrieved GLM DE is <5%. For offsets 800% of observed, retrieved GLM DE has a -40% error.
- Absolute error of retrieved GLM FAR ranges from ~30% for small reference offsets to ~65% for the extreme offset scenario.

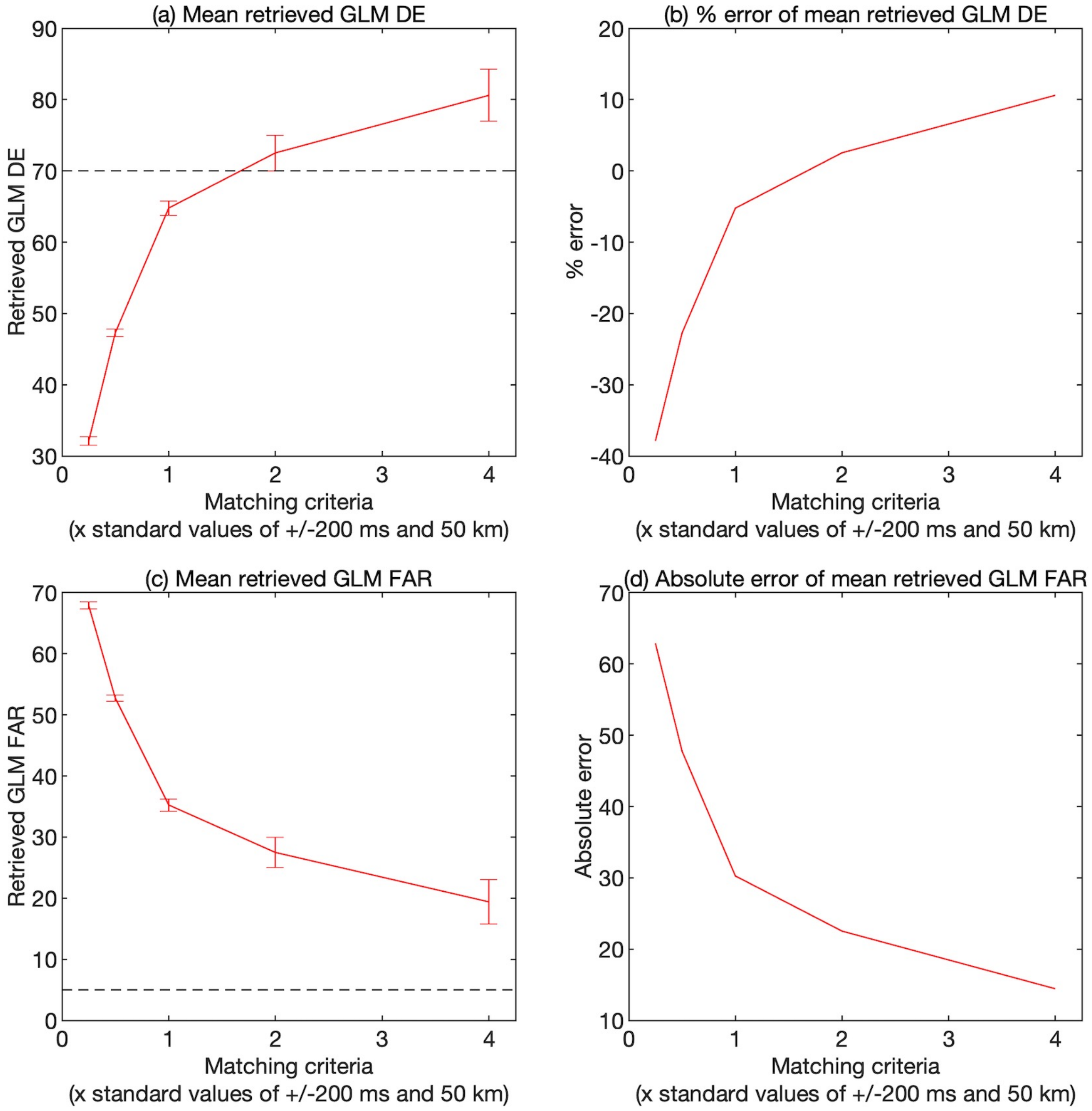


Simulation: Vary Matching Criteria

Simulation specifics:
GLM DE = 0.7 (spec)
GLM FAR = 0.05 (spec)
Reference DE = 0.7
Reference FAR = 0.05
GLM offsets = observed
Reference offsets = observed
Matching criteria = 0.25, 0.5, 1, 2, 4 x standard

Results:

- Strictest matching criteria underestimate true GLM DE by ~40%, with >60% absolute error in retrieved GLM FAR
- Broadest matching criteria overestimate true GLM DE by 10%, with 15% absolute error in retrieved GLM FAR
- The curve is steeper for more strict matching criteria
- Largest error bars are associated with the broadest matching criteria



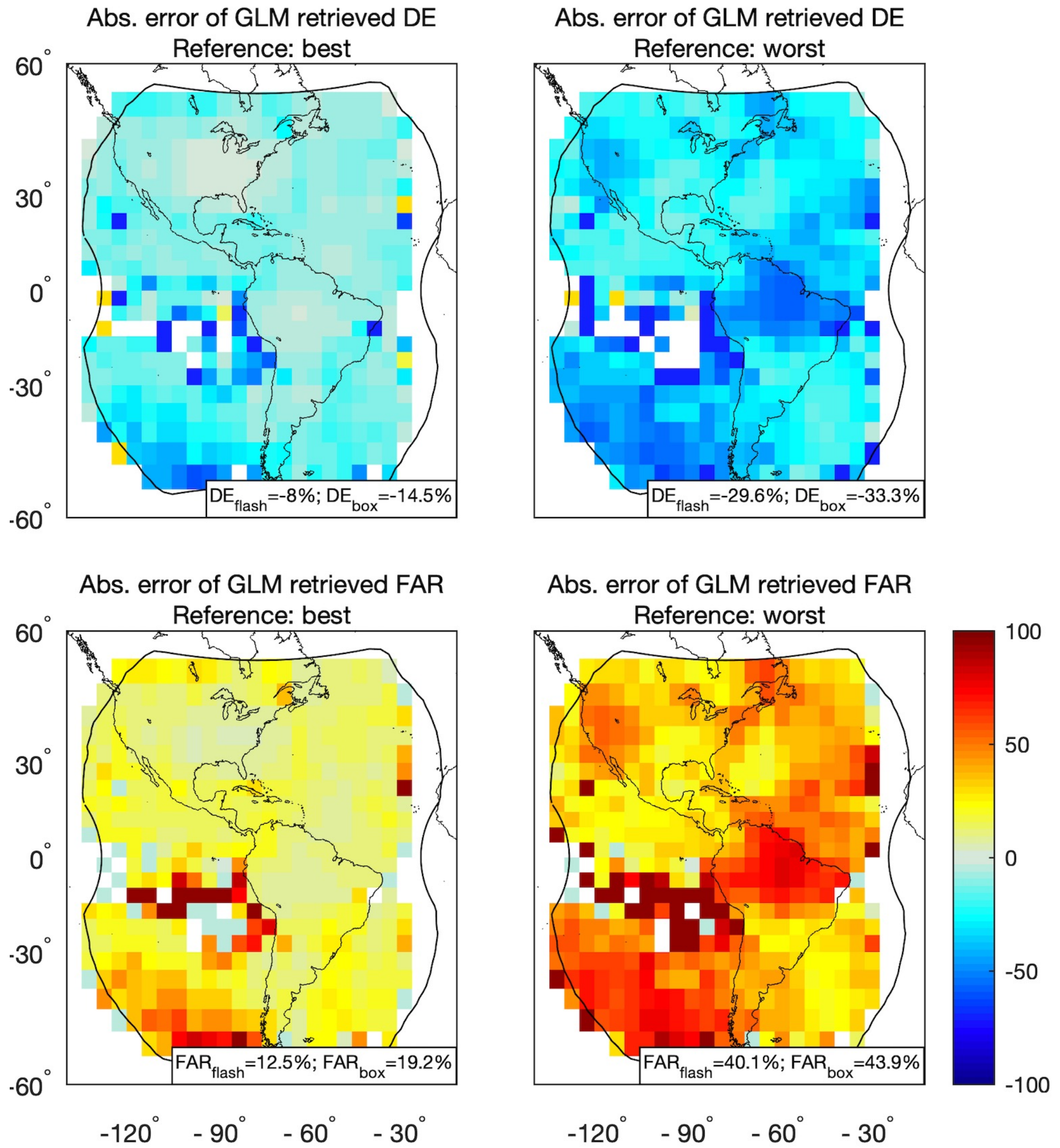
What About Geographic Variations?

- The simulations to this point have defined the "reference" network performance universally. In reality, this performance has a geographic dependence.
- Goal: Construct simulations in which the "reference" data performance varies in a geographically realistic manner
- Define a "best" and "worst" reference scenario at each grid point:
  "Best" = Higher DE and lower FAR
  "Worst" = Lower DE and higher FAR
- Analysis is "reference network-agnostic"
  - It does not matter which real-world network performs better in a given grid box
  - The important point is defining a realistic upper and lower bound on reference performance

Simulation: Vary Ref. Geographical Performance

Simulation specifics:
GLM DE = 0.7 everywhere (spec)
GLM FAR = 0.05 everywhere (spec)
Reference DE and FAR = "Best" and "worst" scenario at each grid box
GLM offsets = observed aggregate values
Reference offsets = observed at each grid box
Matching criteria = standard

Results:
- Retrieved GLM DE aggregated over all flashes ranges from 40% (reference = worst) to 62% (reference = best), implying absolute errors of -8 to -30%
- Retrieved GLM FAR aggregated over all flashes ranges from 17.5% (reference = best) to 45% (reference = worst), or absolute errors of 12.5% to 40%
- Geographic error patterns are evident, with GLM performance most severely underestimated in the Southern Hemisphere Pacific, where absolute errors **for the "best" reference scenario** range up to -30-60% (DE) and 30-70% (FAR)



Conclusions

- Monte Carlo simulations were conducted to analyze the impact upon retrieved GLM DE and FAR of:
  - Reference network performance (DE, FAR, location and timing errors)
  - Realistic geographic variations in reference network performance
  - Spatiotemporal matching criteria
- This analysis provides estimates of the error bars associated with retrieved GLM performance metrics
- Results illustrate that retrieved GLM performance is punished by imperfect reference network performance. In these simulations:
  - The retrieved DE for all flashes is anywhere from 8-30% lower than the true DE
  - The retrieved FAR for all flashes is 12-40% higher than the true FAR
  - GLM performance most severely underestimated in the S.H. Pacific