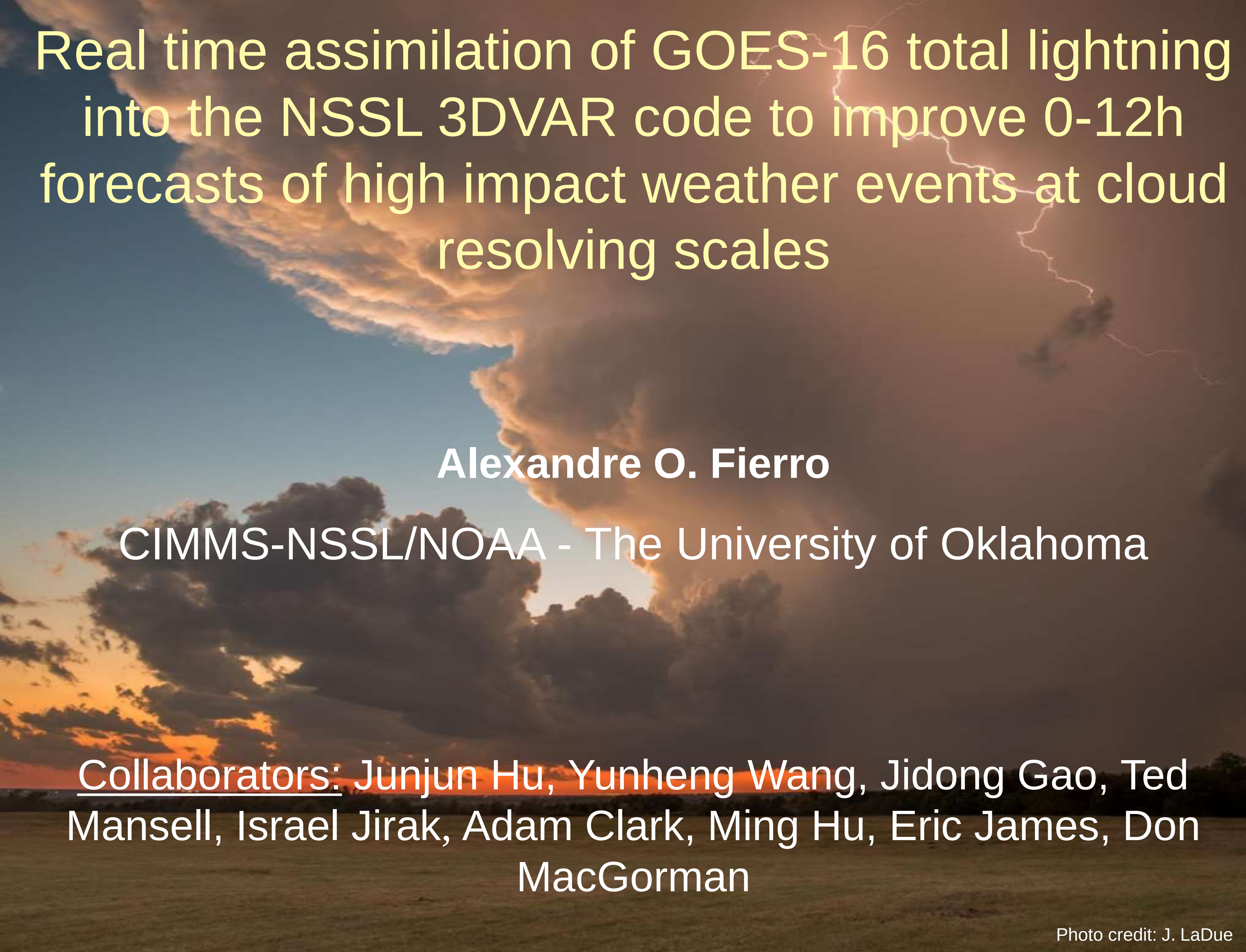


Real time assimilation of GOES-16 total lightning into the NSSL 3DVAR code to improve 0-12h forecasts of high impact weather events at cloud resolving scales

Alexandre O. Fierro

CIMMS-NSSL/NOAA - The University of Oklahoma

Collaborators: Junjun Hu, Yunheng Wang, Jidong Gao, Ted Mansell, Israel Jirak, Adam Clark, Ming Hu, Eric James, Don MacGorman

Photo credit: J. LaDue

Four main methods of lightning data assimilation were investigated by CIMMS-NSSL thus far:

1. Using lightning to force convection initiation by nudging qv where lightning is observed but convection is absent in the model. Forcing is maintained for 10s of minutes to achieve a model response to sustain storms.

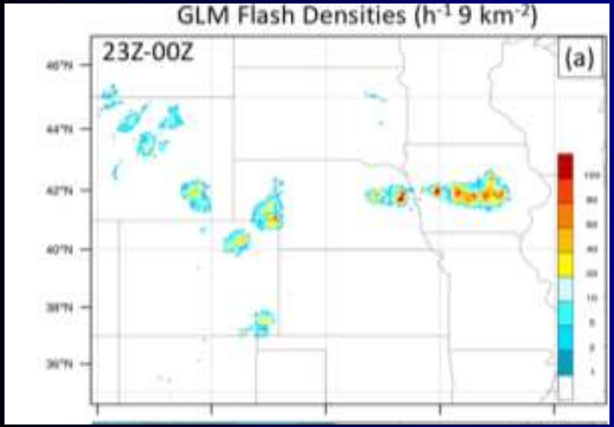
2. Variational (3DVAR) assimilation with high frequency (<= 15 min) successive cycling also using Qv as pseudo observations (proxy) for lightning. [used in this project for real time SFE during HWT]

3. Ensemble Kalman Filter to modulate convection (e.g., strengthen or weaken) in the ensemble members. Ensemble covariances provide adjustments to all state variables (e.g., temperature, water vapor, winds, liquid water and ice particles). Data introduced on 1-5 minutes intervals.

4. Hybrids: EnKF-VAR or ensemble of 3DEnVARs with high frequency (<= 15 min) successive cycling.

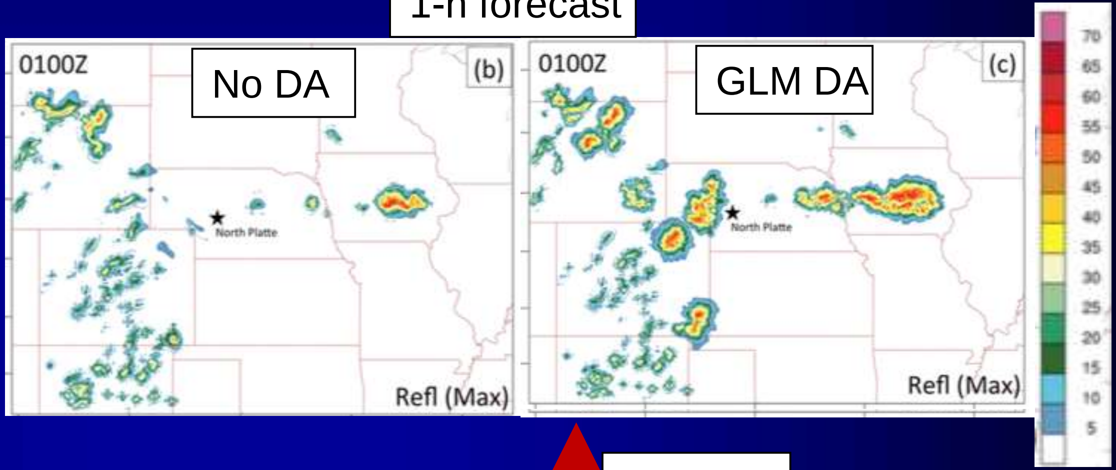
Lightning DA used in real time during the SFE in a nutshell:

Boosts thermal buoyancy via Qv adjustments toward water saturation (RH=95%) between LCL and LCL+3km.



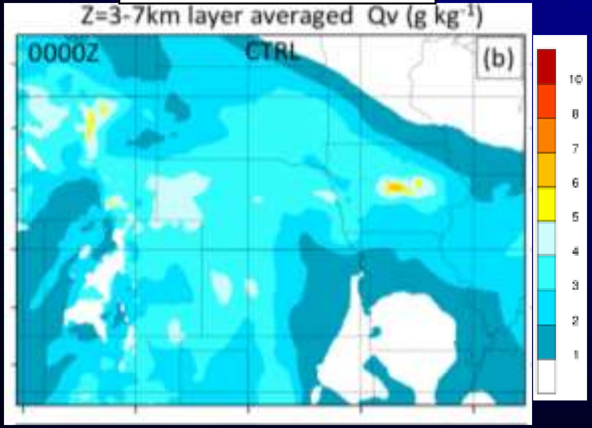
+

1-h forecast



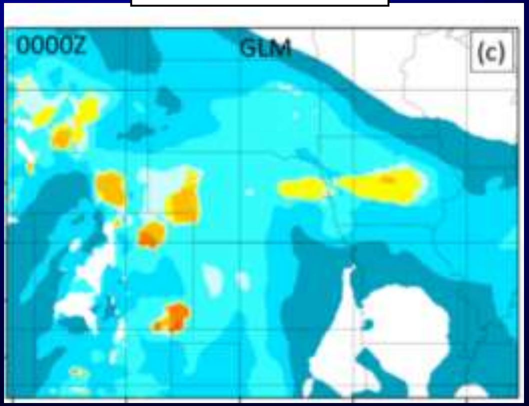
forecast

Background Qv



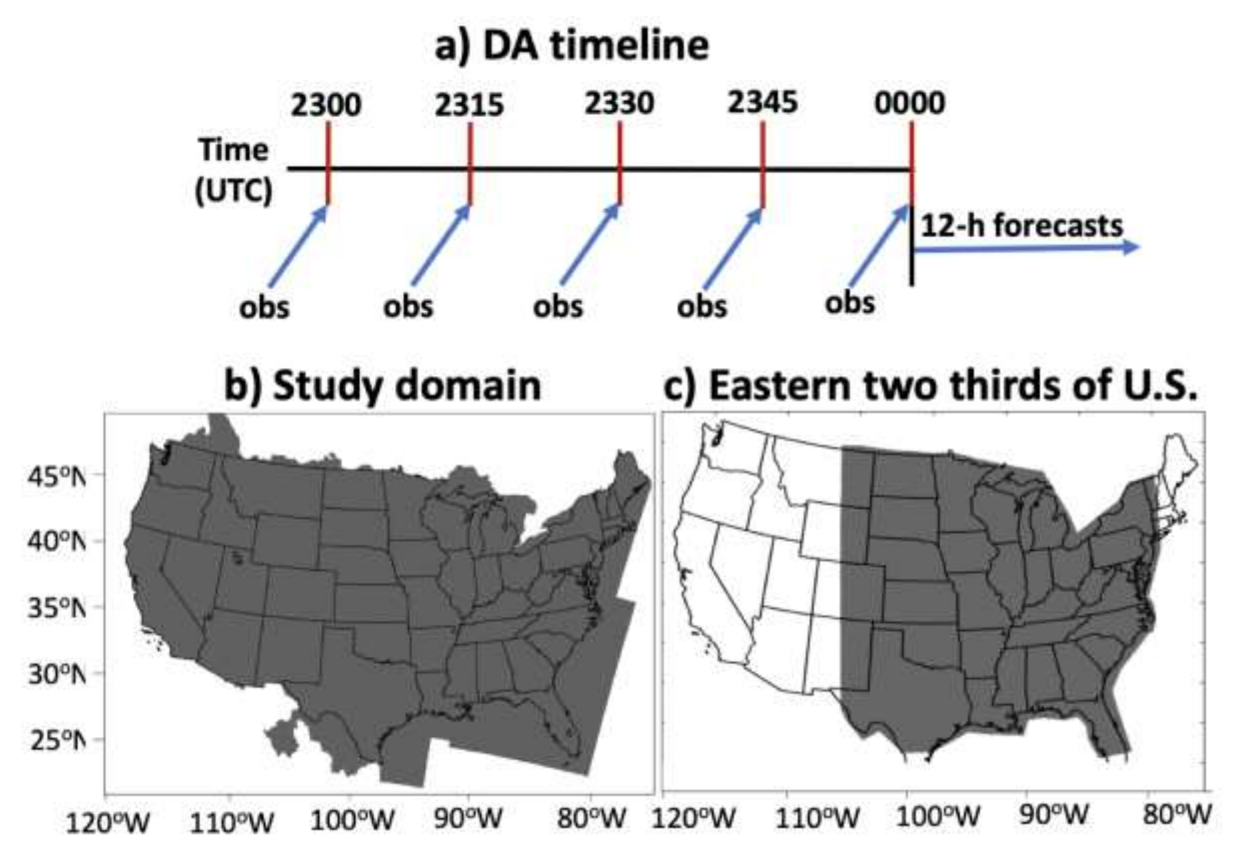
3DVAR analysis

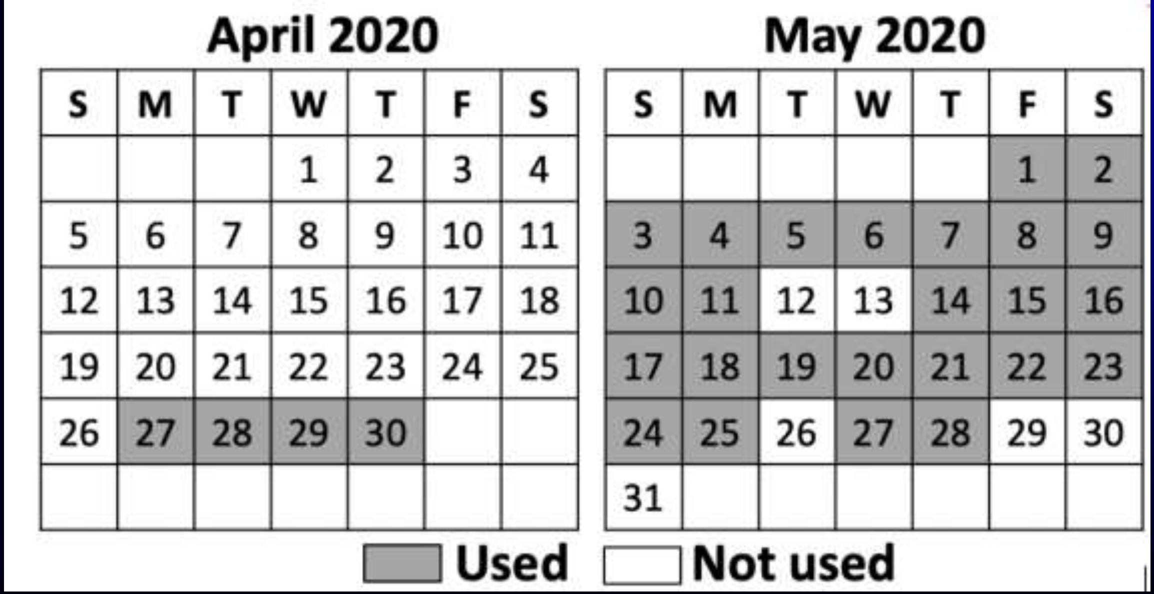
Adjusted Qv



Fierro et al. (2019, MWR)

Real time DA during the SFE:





April 2020

S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	▨ 27	▨ 28	▨ 29	▨ 30		

May 2020

S	M	T	W	T	F	S
					▨ 1	▨ 2
▨ 3	▨ 4	▨ 5	▨ 6	▨ 7	▨ 8	▨ 9
▨ 10	▨ 11	12	13	▨ 14	▨ 15	▨ 16
▨ 17	▨ 18	▨ 19	▨ 20	▨ 21	▨ 22	▨ 23
▨ 24	▨ 25	26	▨ 27	▨ 28	29	30
31						

▨ **Used** ☐ **Not used**

- Use CLUE domain (3-km, 4860 km x 3360 km) and HRRRv4 model with RAPv5 input data.

- DA performed daily between 23-00UTC.

- Use 15 min 3DVAR cycles with 10-min acc GLM data.

- At 00UTC, a 12 h deterministic fcst is launched.

- A sample of 29 fcst days was obtained.

- Analysis focus on precip; contrasts eastern 2/3 vs western 1/3 CONUS (good vs poor radar coverage areas).

SFE Experiments

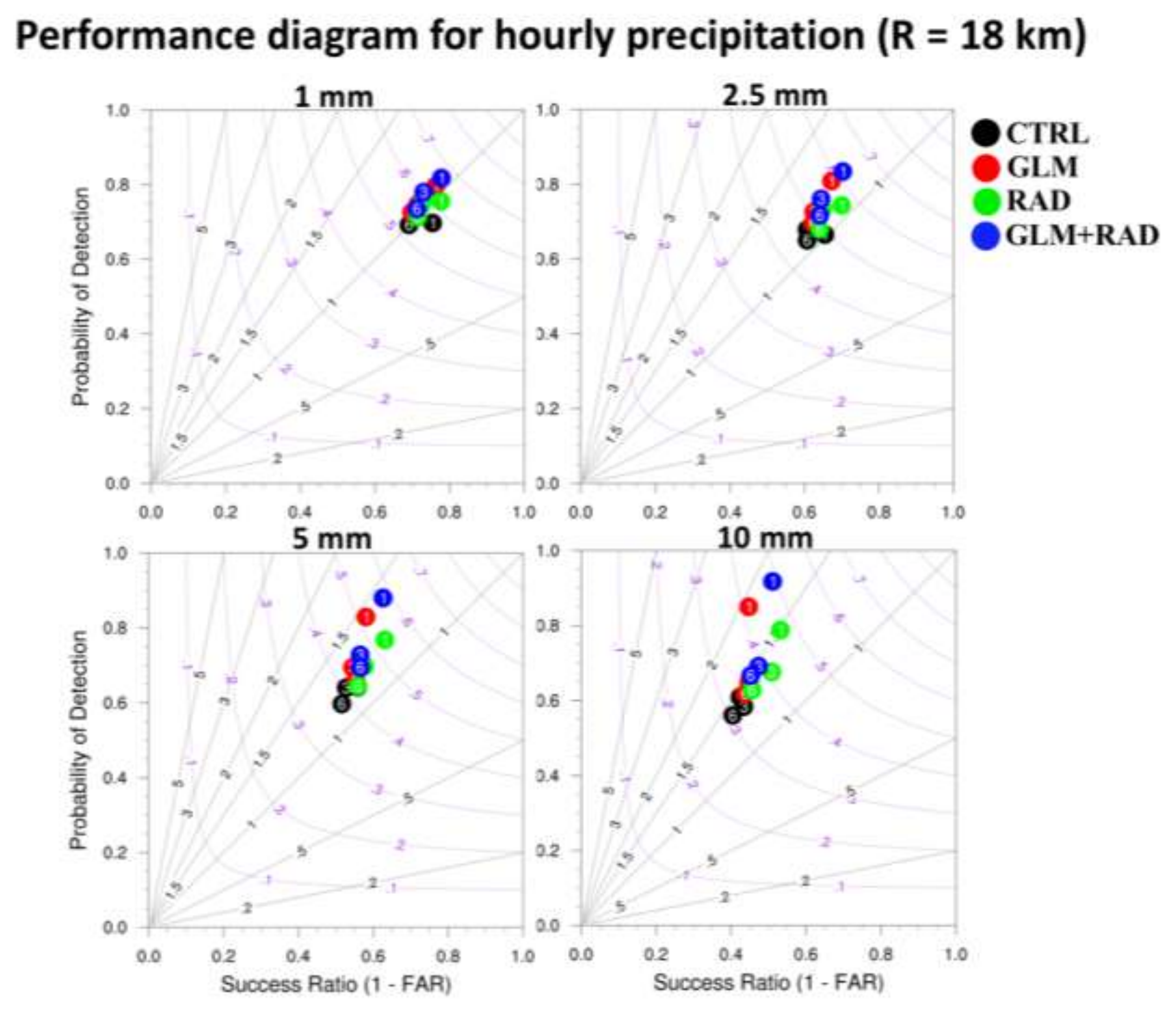
Experiments	**Description**	**Data assimilated variationally**	**Model variables adjusted**
CTRL	Control run	None	None
GLM	Lightning DA run.	GLM flash density rates.	q_v (LCL-3 km)
RAD	Radar DA run	Vr and dBZ	q_r, q_g, q_s, q_h, u, v, w, θ
RAD+GLM	Lightning + Radar DA run	GLM flash density rates, Vr and dBZ	q_v (LCL-3 km), q_r, q_g, q_s, q_h, u, v, w, θ

-Level II (Vr+ dBZ factor) data from 140+ radars were assimilated.

-SFE eval during HWT solely focused on RAD-based experiments over the western 1/3 CONUS to gauge added value of GLM in radar-data sparse areas.

-CTRL and GLM DA run performed and analyzed offline during SFE.

SFE/HWT real-time experiment over CLUE domain; preliminary results

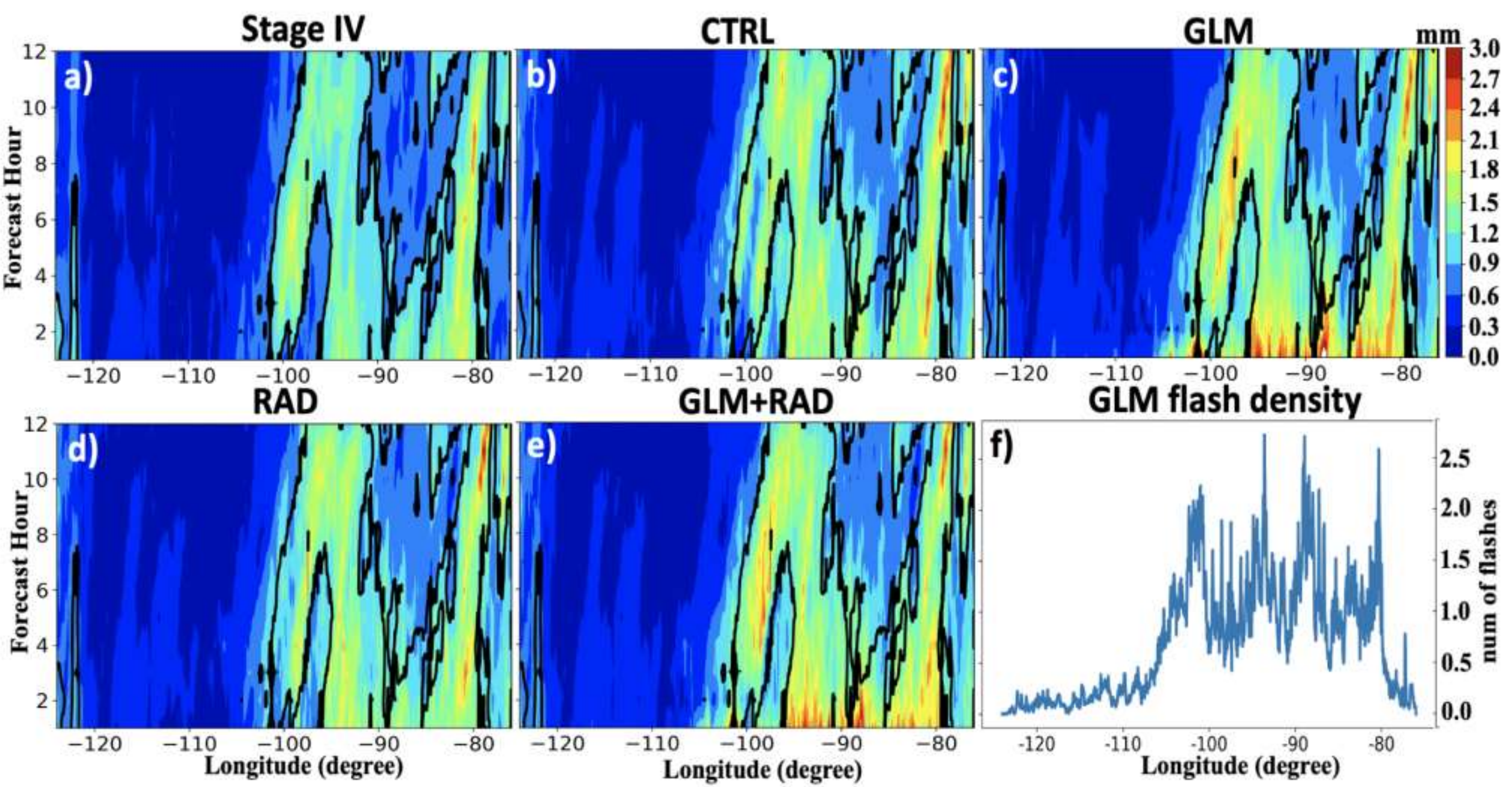


Excellent performance of CTRL is explained by RAPv5 data already blending info from a large arrays of obs, including lightning and radar.

Performance diagrams aggregated over all 29 forecast days over CONUS for 1, 3 and 6-h forecast show a general improvement in forecast skill over CTRL for all DA runs; with the best results obtained for GLM+RAD.

Individual cases reveal that assimilating GLM data showed benefit in radar-sparse areas such as the mountainous west, the Gulf of Mexico, East cast and the Sierra Madre in Mexico.

SFE/HWT real-time experiment; preliminary results



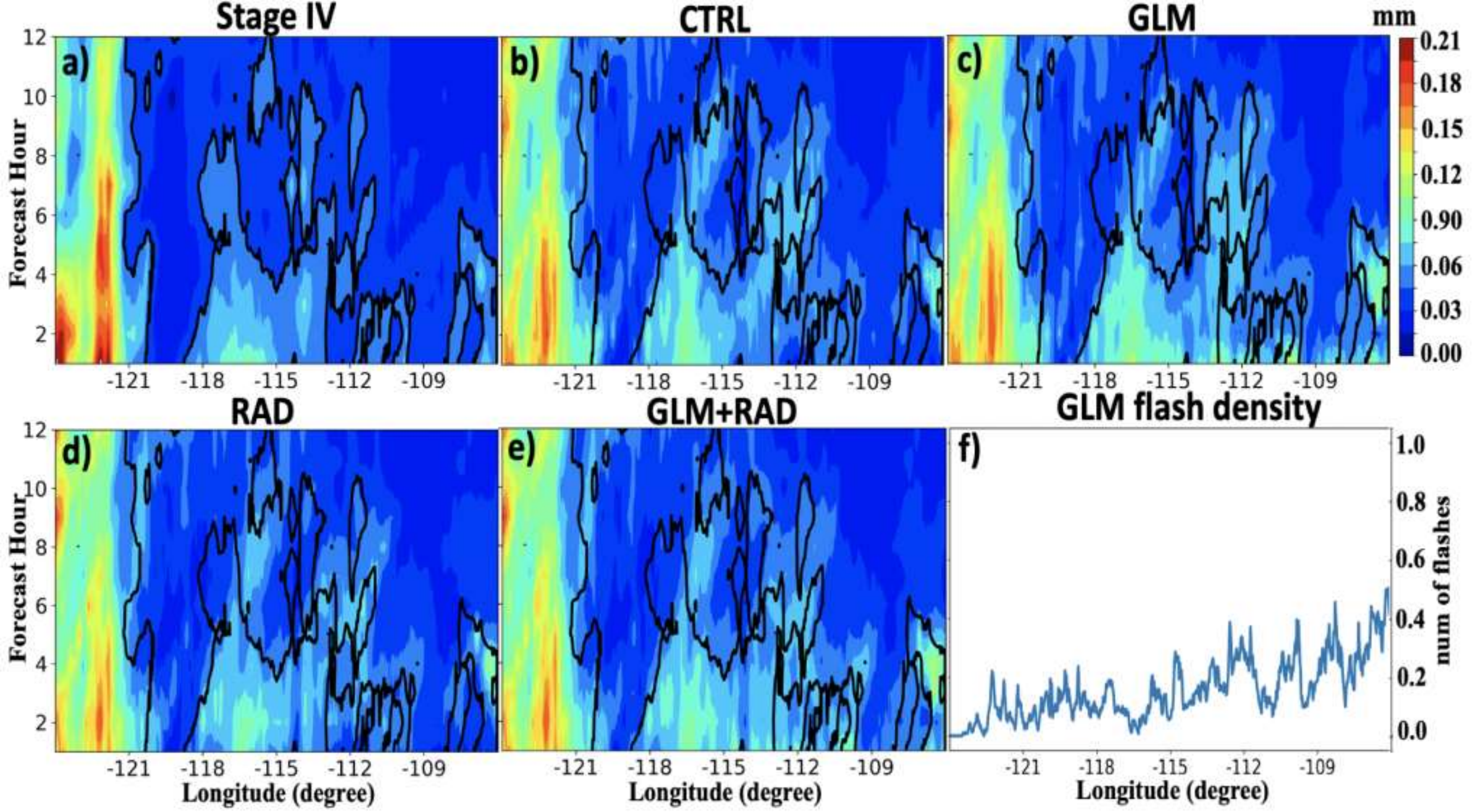
0-6h rainfall aggregated over 29 fcst days

GLM DA adds in more precipitation during the first 2-3h of forecast, especially over the eastern 2/3rd of CONUS where bulk of lightning occurs.

SFE/HWT real-time experiment; preliminary results

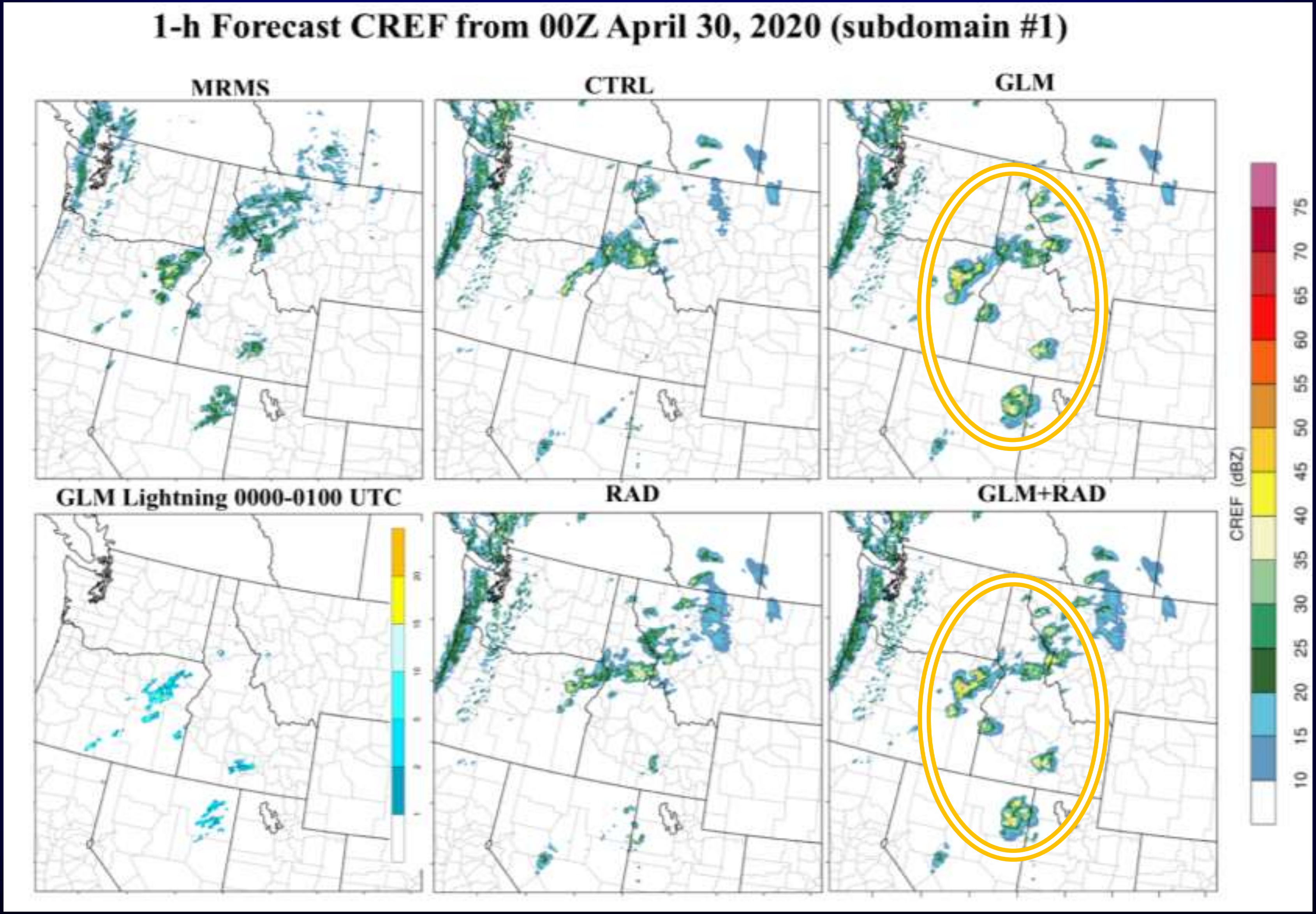
Western US: Relatively less rainfall is added overall by RAD or GLM DA owing to the weaker convective nature of the storms over the western CONUS (e.g., monsoon storms).

Mask



SFE/HWT real-time experiment; Individual cases

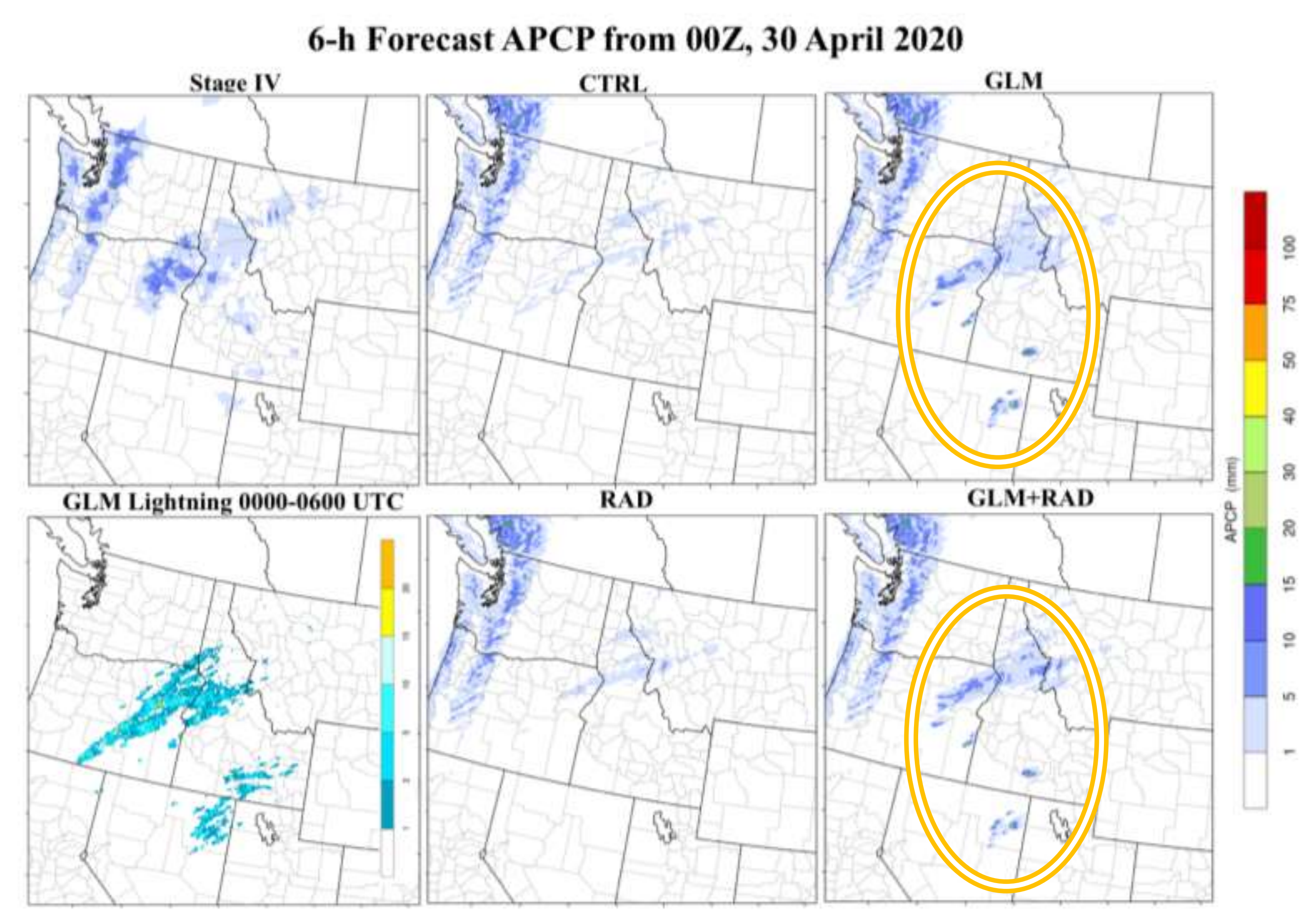
Western US: example of GLM DA improvements over areas characterized by poor radar coverage



1-h Forecast CREF from 00Z April 30, 2020 (subdomain #1)

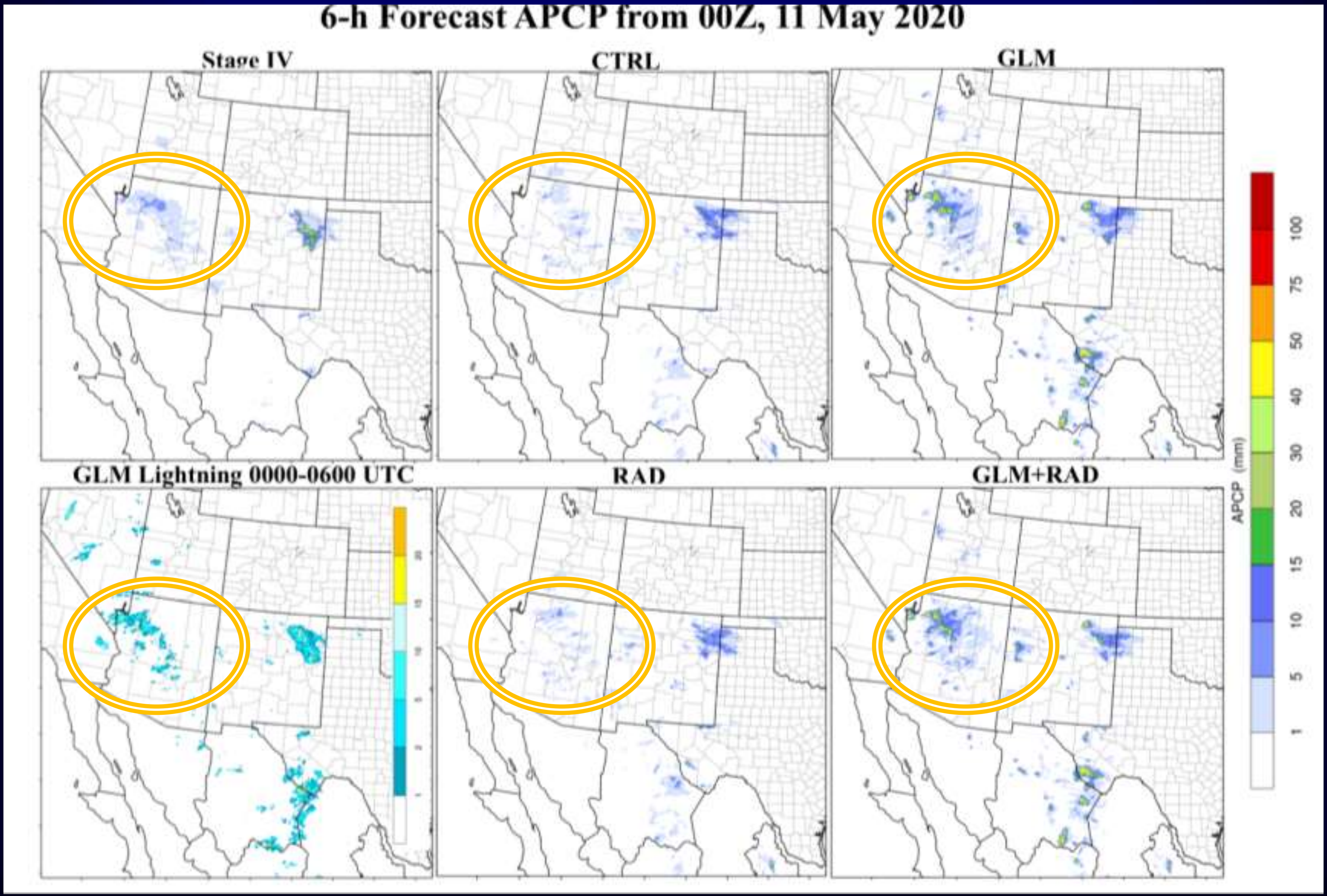
SFE/HWT real-time experiment; Individual cases

Western US: example of GLM DA improvements over areas characterized by poor radar coverage (rainfall)



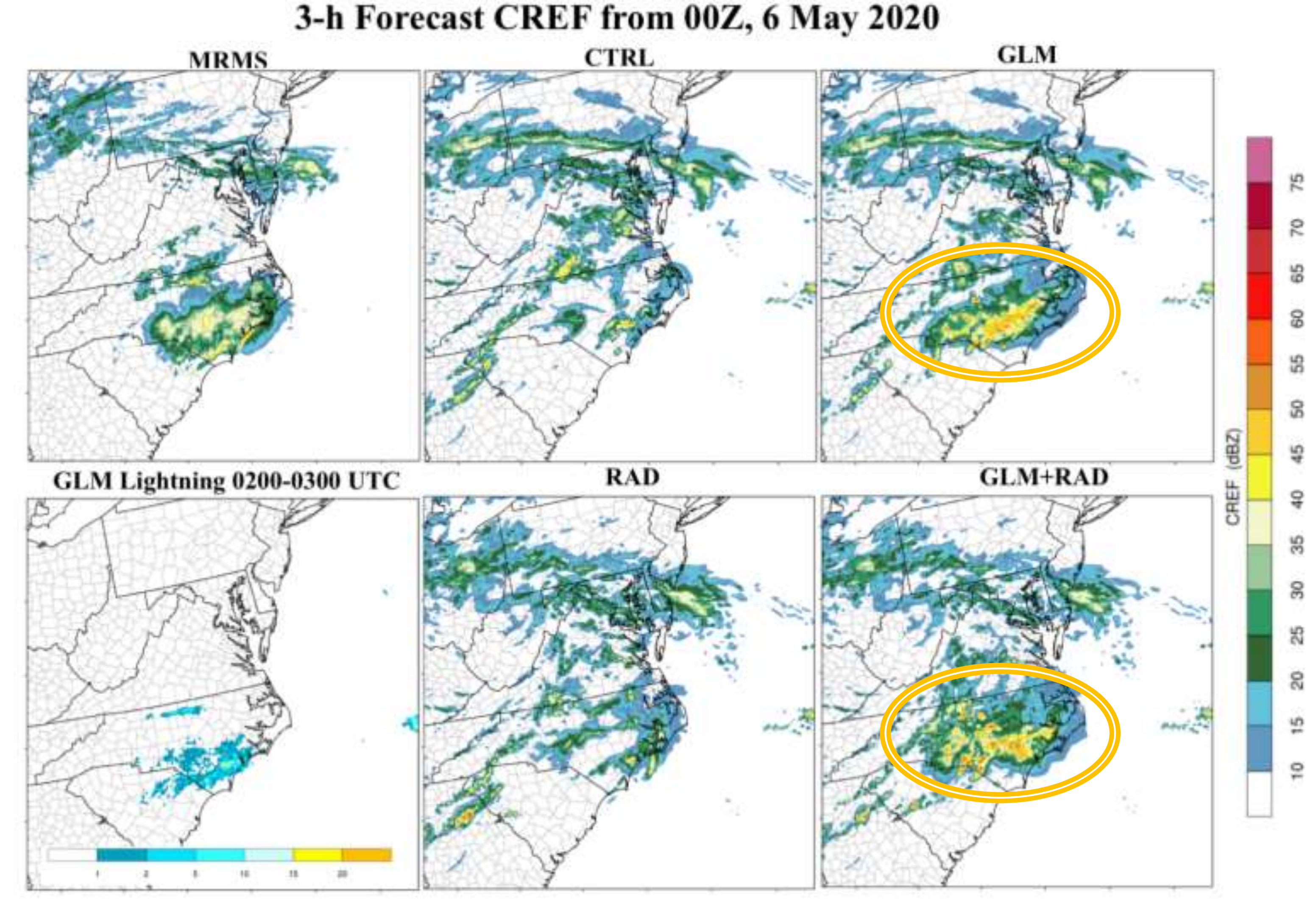
SFE/HWT real-time experiment; Individual cases

Western US: Other example of GLM DA improvements over areas characterized by poor radar coverage (rainfall)

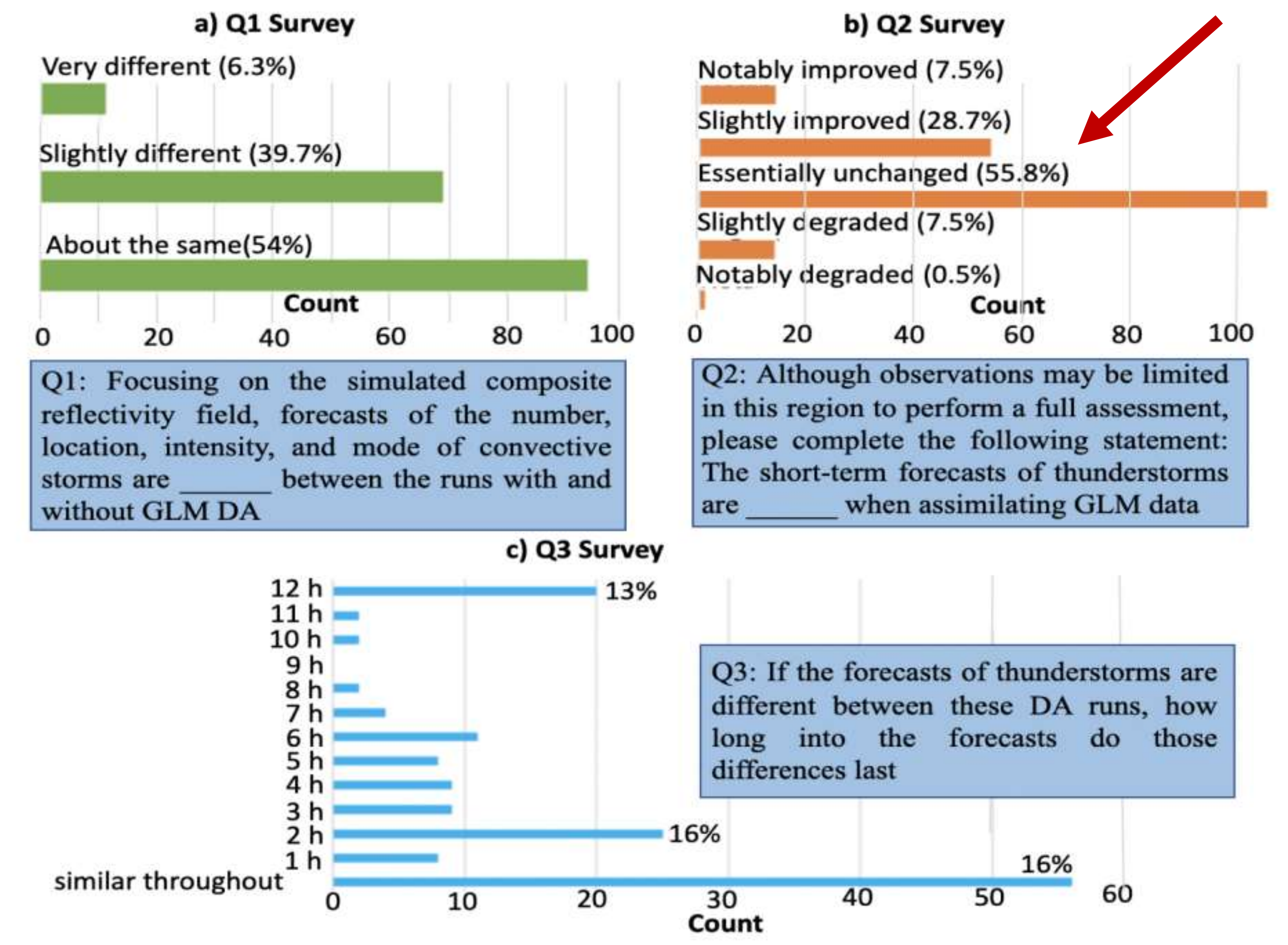


SFE/HWT real-time experiment; Individual cases

Eastern 2/3rd US: example of GLM DA improvements over areas characterized by good radar coverage



SFE/HWT participant survey analysis:



Q1: Focusing on the simulated composite reflectivity field, forecasts of the number, location, intensity, and mode of convective storms are _______ between the runs with and without GLM DA

Q2: Although observations may be limited in this region to perform a full assessment, please complete the following statement: The short-term forecasts of thunderstorms are _______ when assimilating GLM data

Q3: If the forecasts of thunderstorms are different between these DA runs, how long into the forecasts do those differences last

Ongoing and future work involving GLM DA

- Complete analysis of HRRRv4 real time radar +/ GLM DA runs conducted during the SFE of Spring 2020.
- Combine Radar and GLM DA with sfc obs such as Mesonet and/or satellite products.
- More systematic usage of hybrid VAR-EnKF implementation for GLM lightning using Qv- or RH-based operators.
- Parallel work also evaluating GSI-EnKF hybrids DA of GLM FED data using Qg-based obs operators.
- Evaluate FOD vs FED assimilation.
- Until JEDI ready for research applications, couple NSSL-VAR with FV3 / FV3-SAR core.

Lightning DA manuscripts from 2019 onward:

Hu J., **A. O. Fierro**, Y. Wang, J. Gao, E. R. Mansell, A. J. Clark, I. Jirak and M. Hu, 2020: Assessment of storm-scale real time assimilation of GOES-16 GLM lightning-derived water vapor mass and radar data on short term precipitation forecasts during the 2020 Spring forecast experiment. Submitted to Monthly Weather Review.

Hu J., **A. O. Fierro**, Y. Wang, J. Gao and E. R Mansell, 2020: Exploring the Assimilation of GLM-Derived Water Vapor Mass in a Cycled 3DVAR Framework for the Short-Term Forecasts of High-Impact Convective Events. Monthly Weather Review. Volume 148, 1005-1028.

Hu J., J. Gao, Y. Wang, S. Pan, **A. O. Fierro**, P. Skinner, K. Knopfmeier, E. R. Mansell and P. Heiselman, 2020: Evaluation of a Warn-on-Forecast 3DVAR analysis and forecast system on quasi- real time short-term forecasts of high impact weather events. Submitted to Quarterly J. Royal. Metr. Soc.

Fierro, A. O., Wang. Y, Hu J., Gao J., and E. R. Mansell, 2020: Proof-of-concept evaluation of ensemble of 3DEnVARs assimilation (ENH3DA) of GLM-observed total lightning data for the 1 May 2018 tornado outbreak. *Submitted to Monthly Weather Review.*

Fierro, A. O., Wang. Y, Gao J., and E. R. Mansell, 2019: Variational assimilation of radar data and GLM-lightning derived water vapor for the short-term forecasts of high-impact convective events. *Monthly Weather Review.* Volume **147**, 4045-4069.

Rong K., M. Xue, **A. O. Fierro**, Y. Jung, C. Liu, E. R. Mansell and D. R. MacGorman, 2020: Assimilation of GLM-observed Flash Extent Density in GSI EnKF: Proof-of-concept for the Analysis and short-term Forecast of the 13 July 2018 Mesoscale Convective System. *Monthly Weather Review. Volume* **148**, 2111-2133.

Kong R., M. Xue, C. Liu, **A. O. Fierro**, E. R. Mansell, and D. R. MacGorman, 2020: Assimilation of GOES-R Geostationary Lightning Mapper Flash Extent Density data in GSI 3DVar, EnKF, and Hybrid En3DVar for the Analysis and Short-Term Forecast of a Supercell Storm Case**.** *Submitted to Monthly Weather Review.*

Prat A. C., S. Federico, R. C. Torcasio, **A. O. Fierro**, Stefano Dietrich, 2020: Lightning data assimilation in the WRF-ARW model for short-term rainfall forecasts of three severe storm cases in Italy. *Submitted to Atmos Res.*

Extra slides for questions

Types of lightning data assimilated:

(1) Ground based networks divided in 3 categories
- VLF (global/intl: WWLLN, GLD360, ZEUS). 3–30 kHz
- Broadband (intl: ENTLN). 1 Hz–12 MHz
- VHF (regional: LMA). 30–300 MHz

(2) Spaceborne optical instruments
- Low Earth Orbit (TRMM-LIS)
- Near Polar Orbit (OTD-Microlab-1)
- *Geostationary (GEOSR GLM / FenYun4 LMI/ MTG)*

Each of these technologies sees or detects different physical aspects of lightning flashes (photon emission versus dE/dt pulses or sferics), which must be accounted for during DA exercises.

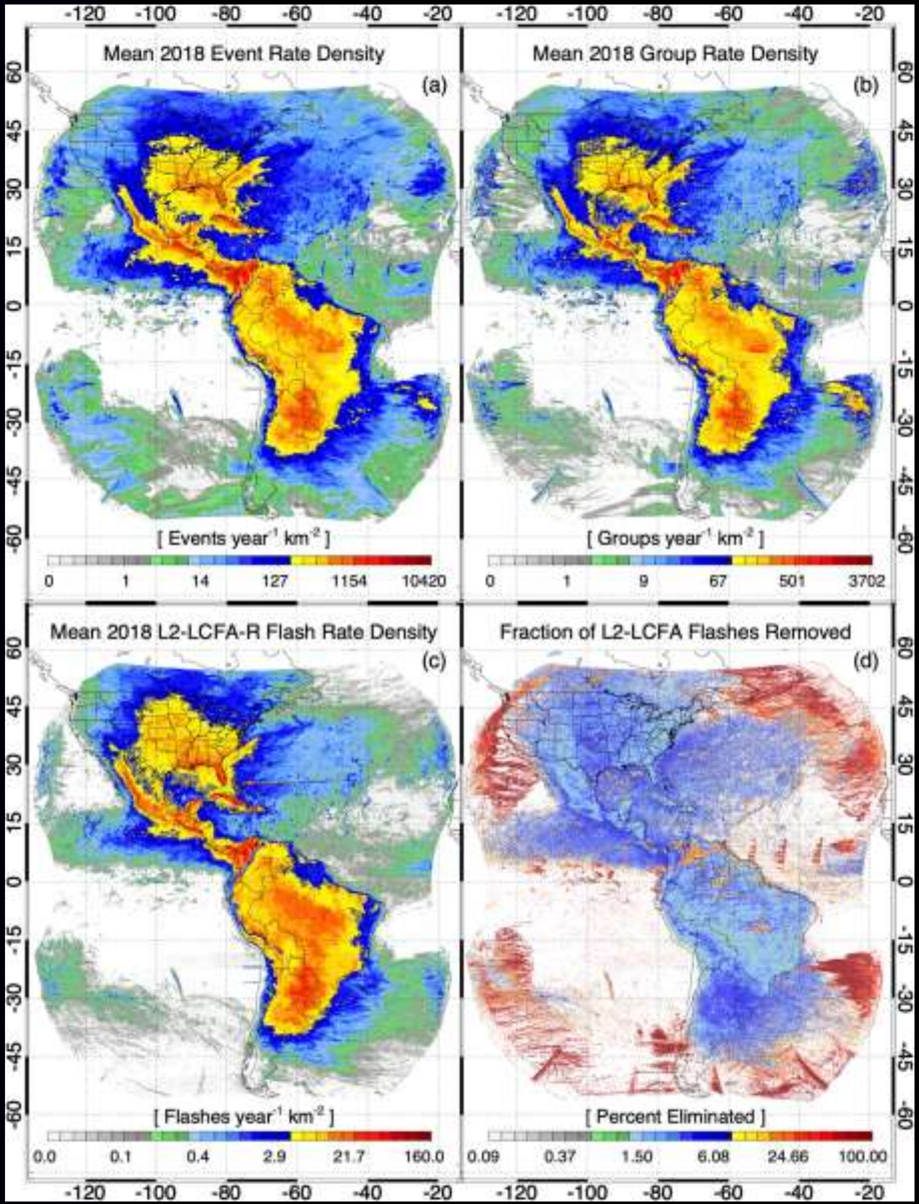
Spaceborne optical instruments: GLM

Characteristics

- Staring CCD imager (optical)
  (1372x1300 pixels)
- Near uniform spatial resolution
  8 km nadir, 12 km edge fov
- Coverage up to 52 deg lat
- 70-90% flash detection day and night
- 2 ms frame rate
- < 20 sec product latency
- Delivers three primary "lightning variables" related by parent-to-child relationships based on fixed time & spatial thresholds: Events, groups and flashes.

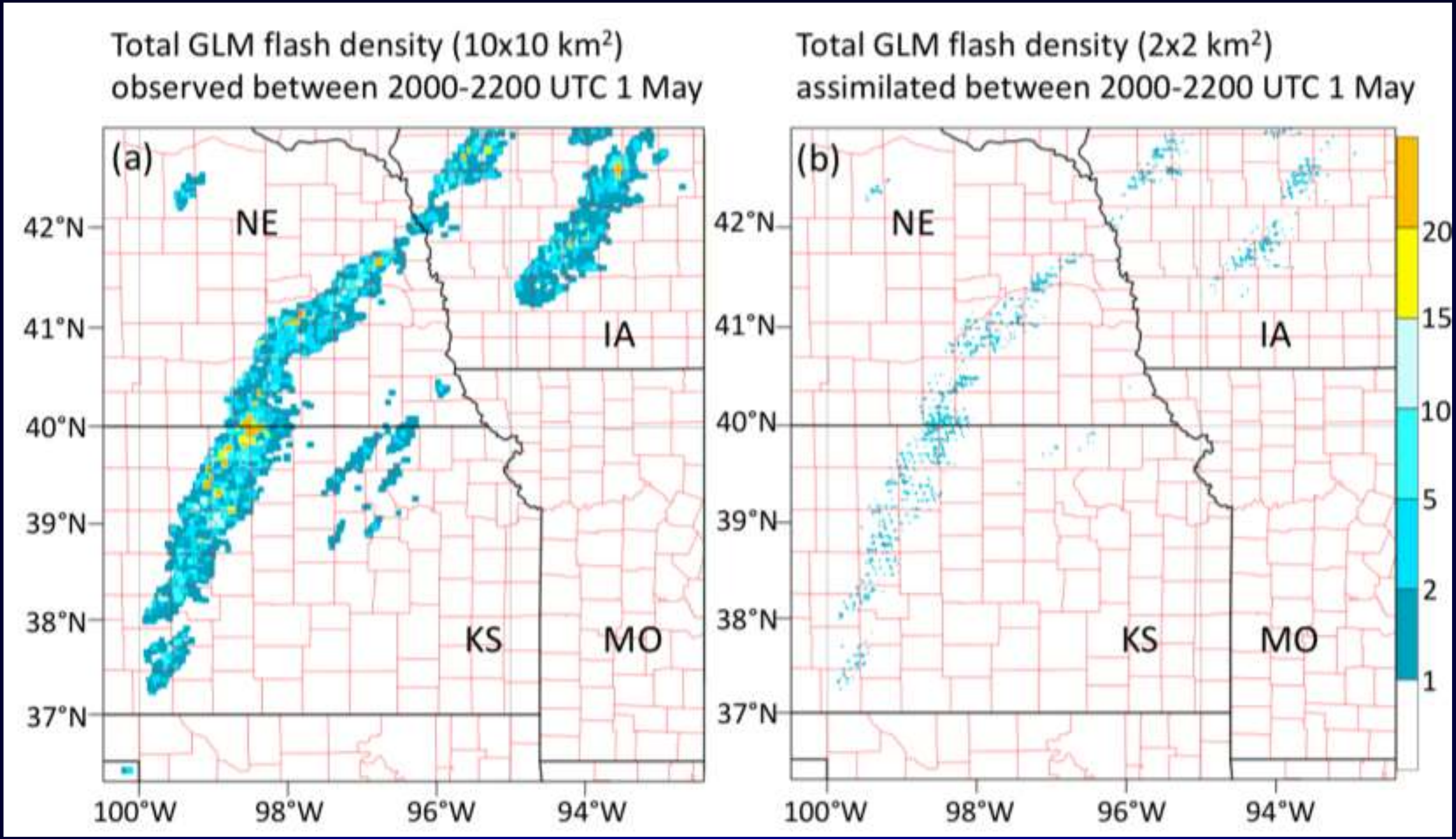
Different instruments observe same phenomenon differently → challenge for DA applications!

2018 GLM Data:



Petersen, (2019), JGR

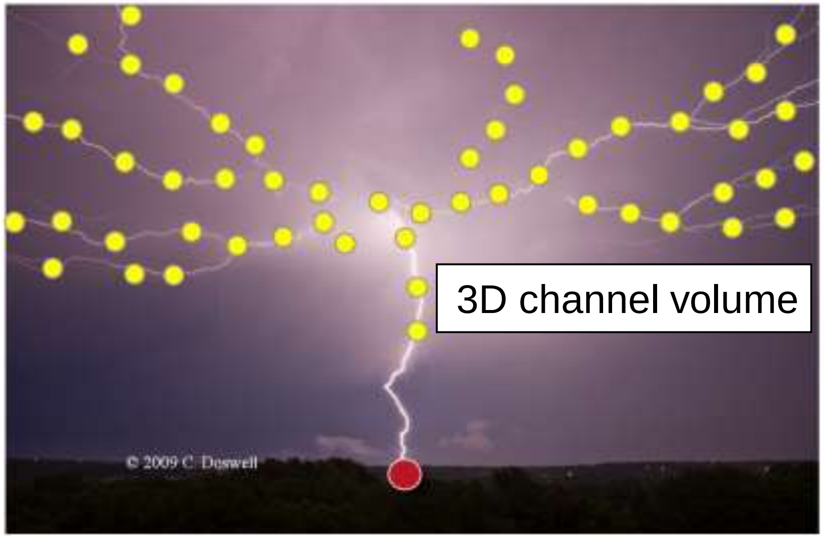
GLM gridding/prep for DA

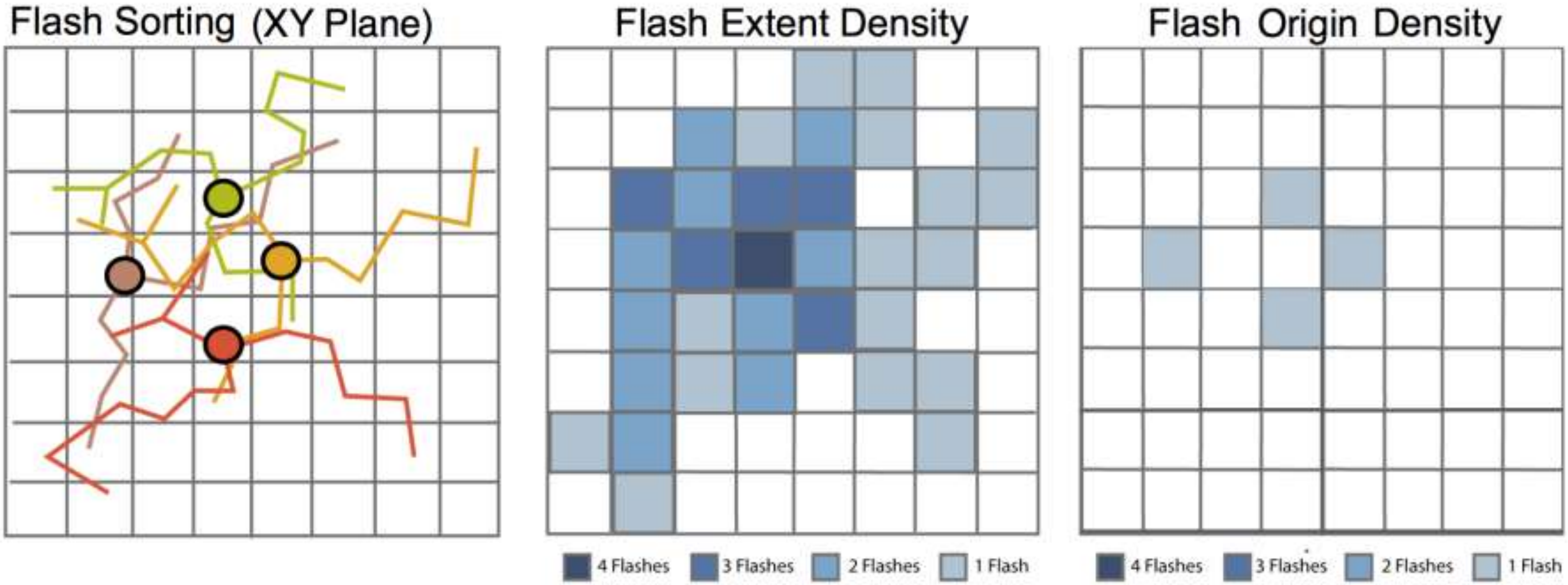


-8-12 km GLM pixel sizes purposively thinned down to 2km to reduce mass adjustments in the model.
-This was shown to help reduce wet bias potential while yielding similar improvements in forecast skill (Hu et al. 2020, MWR) .
-Method essentially equivalent to reducing horizontal length scale of the control variable Qv in 3DVAR analysis.

Flash metrics options for DA:

Use radiation points detected from ground-based lightning mapping arrays (LMAs)



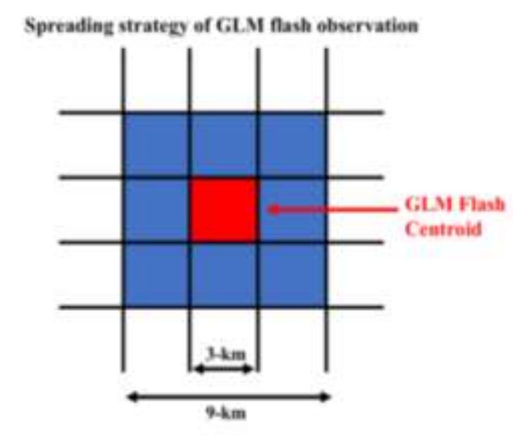


Flash Extent Density = Flash Footprint

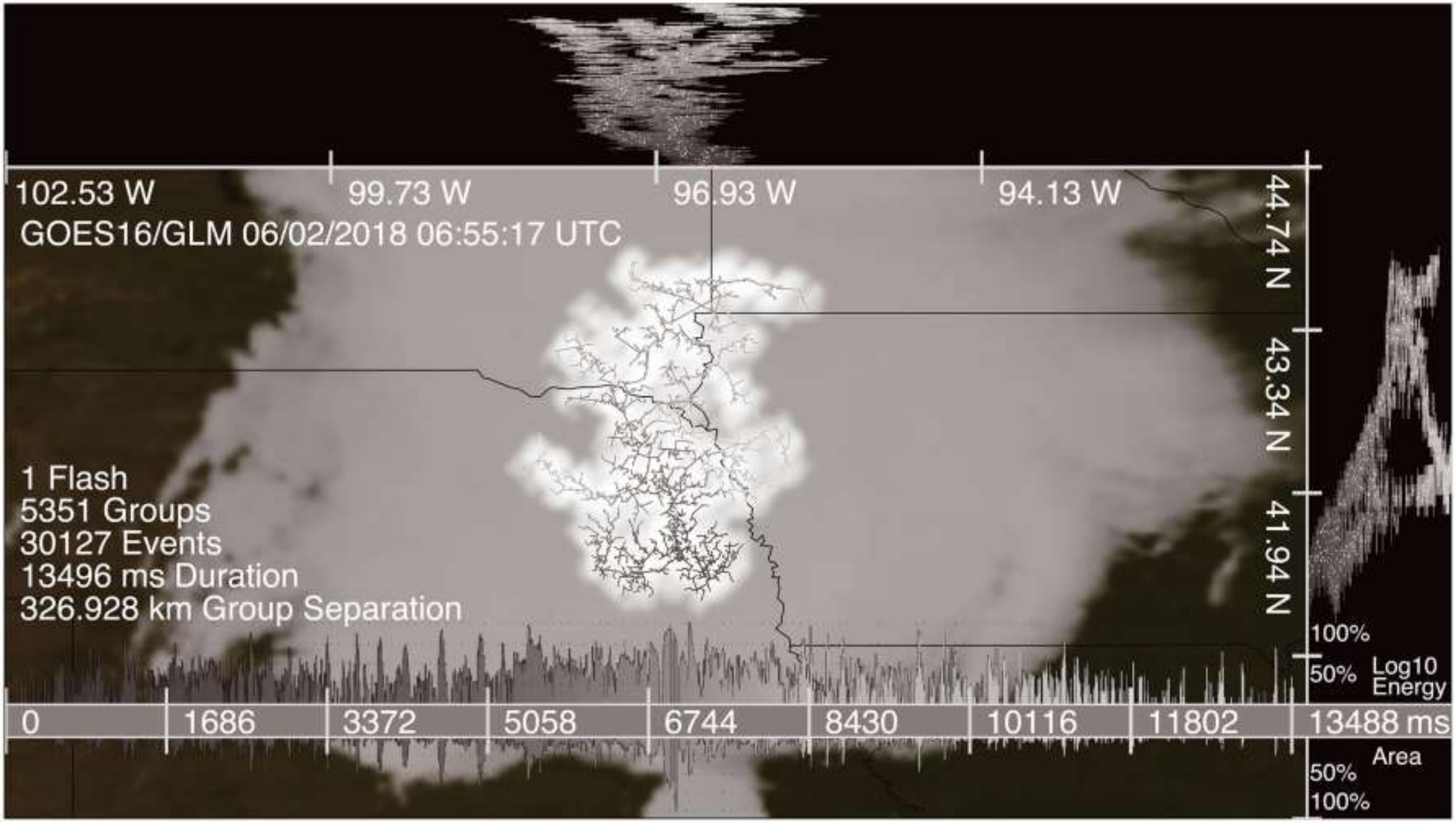
Lightning DA: create pseudo-observations

- Lightning not explicitly predicted in typical operational NWP → Find a proxy for lightning.
- Lightning → presence of moist mixed phase convection -> saturated (buoyant) ascent (wrt to water).
- Boost water vapor mass Q_v within a fixed layer above cloud base (LCL) towards Q_{sat_water}. Concomitantly, increasing Q_v at constant T boosts thermal buoyancy (via θ_v) and, ultimately, promotes updraft development.
- CAPE within 2-3 km layer above LCL most efficiently converted into KE + updrafts will be more systematically rooted in BL.
- Only applied whenever simulated RH= Q_v / Q_{sat_water} < fixed thresholds: i.e., if the model already is in the right direction don't adjust Qv.

- Project L2 GLM "flash" centroids onto CAM (dx= 3-km) grid.
- Account for ~9x9 km2 pixel resolution of the GLM by spreading footprint on 3-km grid as shown here →
- Create 3D pseudo Qv observations that are minimized in cost function.



Spaceborne optical instruments: GLM



Petersen, (2019), JGR: longest lasting 2018 GLM flash

VAR Lightning DA: Background

What is variational DA? Two main types: 3DVAR and 4DVAR.

- 3DVAR: Find the optimal analysis x=xa that minimizes a (scalar) cost function, proportional to the sum of the Euclidian distances between x and the background xb (initial guess), plus the distance between x to the observations yo, each weighted by a measure of their respective errors (covariance matrix).

Derivation:

- Assume that x=xa is a realization of a random process defined by the prior probability distribution function (given the background field) .
- Prob distributions are Gaussians [fully described by 1st and 2nd moments].
- Obs and background error are uncorrelated and unbiased
- Baye’s theorem yields to:

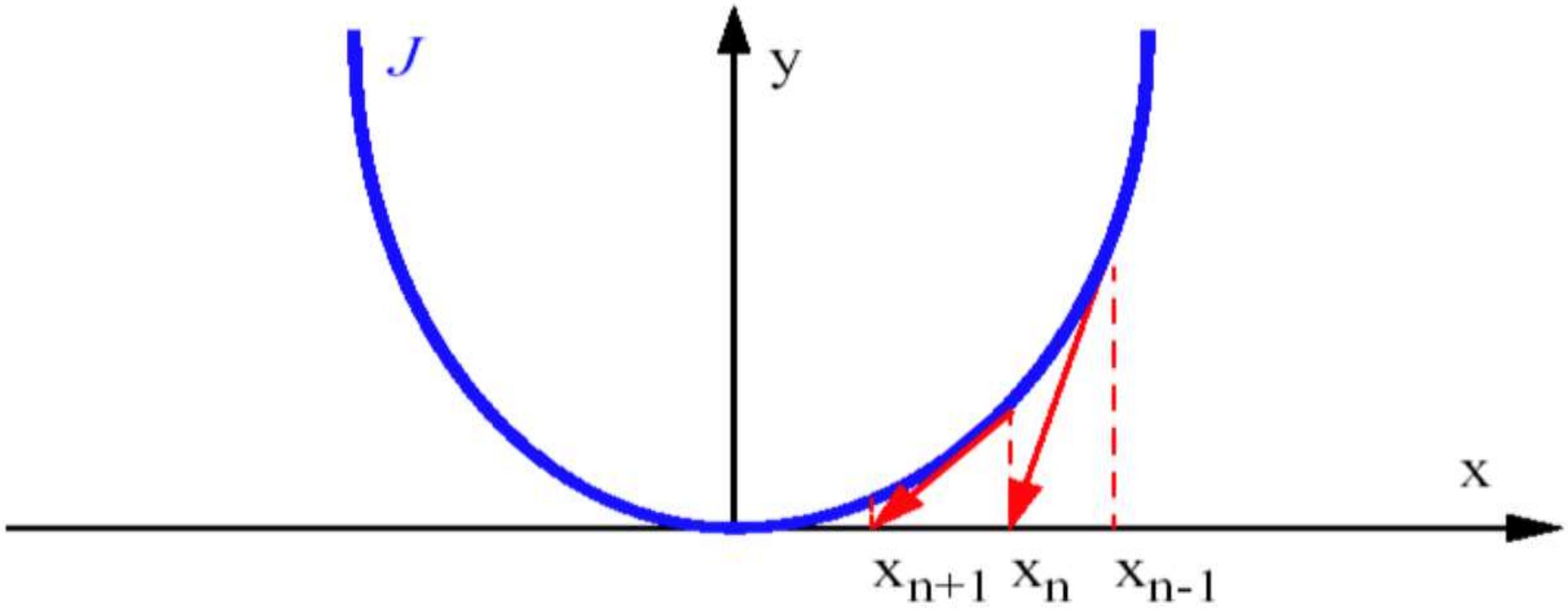
$$p(\mathbf{x}|\mathbf{y}_o) \propto \frac{1}{(2\pi)^{p/2}|\mathbf{R}|^{1/2}} \frac{1}{(2\pi)^{n/2}|\mathbf{B}|^{1/2}} \exp\left\{-\frac{1}{2}[(\mathbf{y}_o - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y}_o - H(\mathbf{x})) + (\mathbf{x}_b - \mathbf{x})^T \mathbf{B}^{-1}(\mathbf{x}_b - \mathbf{x})]\right\}$$

→The x that minimizes the minus of the exponent (cost function) is the x that maximizes the probability of the analysis.

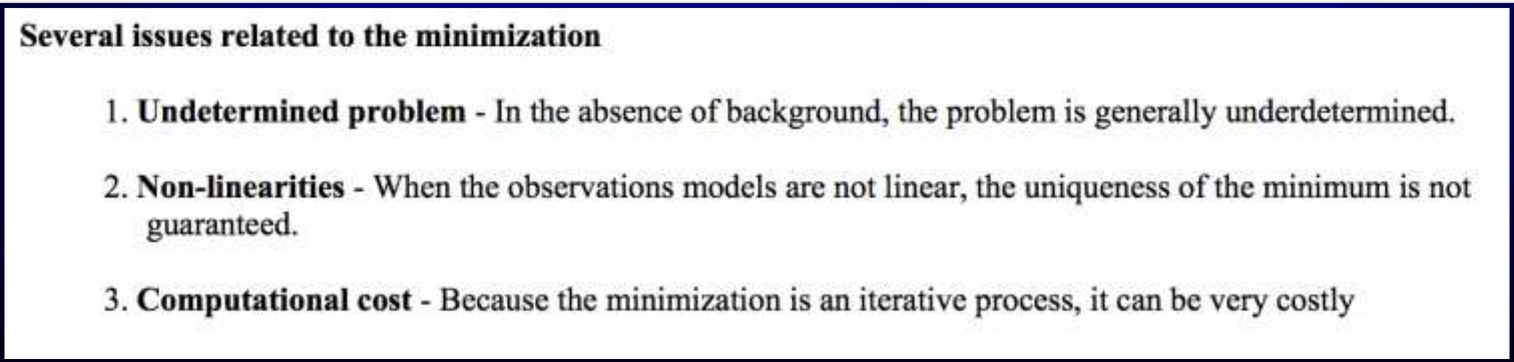
→3DVAR = optimal analysis estimator given the obs yo.

VAR Lightning DA: Background

The minimum of the quadratic cost function is found by an iterative procedure such as gradient descent; illustrated below for a linear model.



Linear model: J quadratic

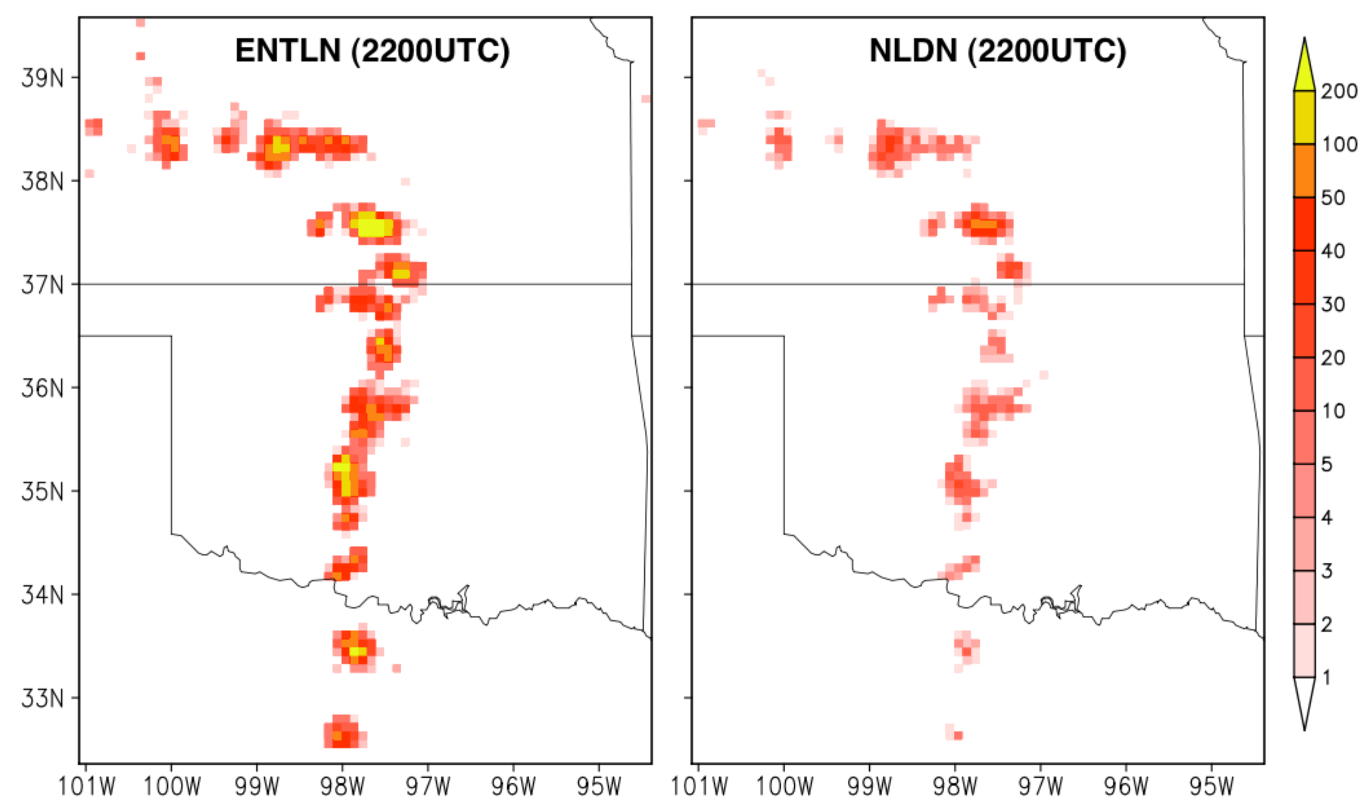


Several issues related to the minimization

1. **Undetermined problem** - In the absence of background, the problem is generally underdetermined.
2. **Non-linearities** - When the observations models are not linear, the uniqueness of the minimum is not guaranteed.
3. **Computational cost** - Because the minimization is an iterative process, it can be very costly

Ground based networks

Broadband vs VLF: ENTLN (CG+IC) versus NLDN (mainly CGs)

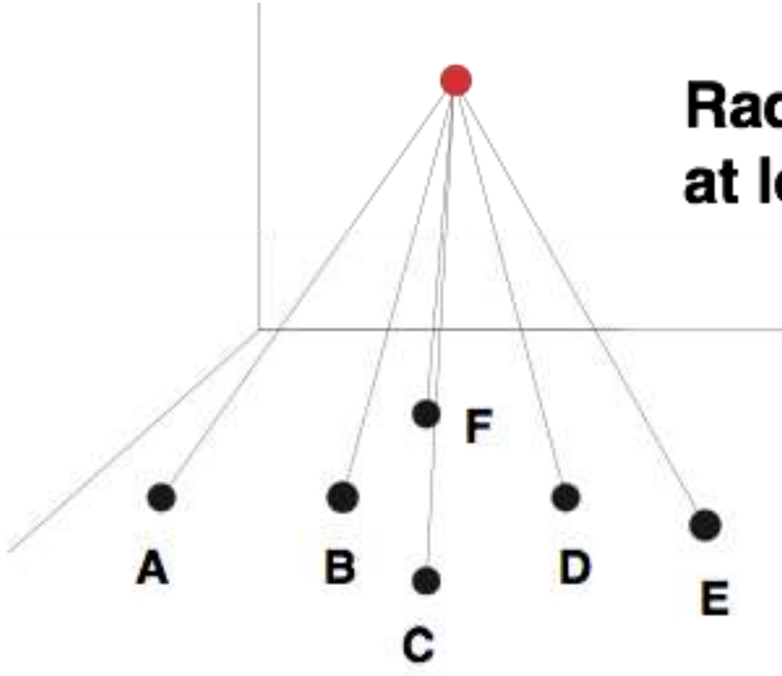


ENTLN/NLDN ≈ (IC+CG)/CG Ratio of 9x9-km 10-min gridded flash counts ranges from 2 to 10. IC+CG also spans a larger area. IC also better correlated with W and hence, timing of the convection.

Ground based networks: Time of Arrival

- •Measure time RF pulse arrives at multiple stations
- •Determine position and time of source

Locate 1-2 (VLF) to 1000+ (VHF) sources per flash



Radiation occurs at time t, at location (x, y, z)

Radiation arrives at station i at time t_i, location (x_i, y_i, z_i)

dr/dt=c → $t_i = t + \frac{\sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2}}{c}$

Nebraska storm (1-2 min exposure)


Credit: G. Takei

(IC+CG)/CG Ratio = 5,10 ?