Leveraging Federated Learning for Generative Precipitation Nowcasting in Diverse U.S. Climate Regions

Zhe Xu, University of Massachusetts Amherst, Amherst, MA; and T. Mehboob, A. Rutkovskii, and M. Zink

MOTIVATION

Short-term precipitation nowcasting is challenged by massive data volumes and strong regional climate heterogeneity. Training generative nowcasting models traditionally requires transferring large volumes of radar data to a centralized compute node, which can introduce significant network overhead and limit scalability.

Federated learning (FL) offers an alternative paradigm by training a shared model across distributed datasets through aggregation of locally trained updates, rather than transferring raw data to a central server. This approach is particularly well suited for precipitation nowcasting, where radar observations are naturally distributed across geographically separated regions.

By keeping raw data local while exchanging only model parameters, FL reduces data transfer costs and enables efficient large-scale training. At the same time, it supports privacy-preserving collaboration across institutions or regions, making it possible to jointly train models on geographically diverse radar networks without requiring centralized data sharing.

GOALS & METHODS

We construct seven radar-centered subdomains from Multi-Radar/Multi-Sensor System (MRMS) data source representing diverse U.S. precipitation regimes as federated learning (FL) clients and train the Deep Generative Model of Radar (DGMR) using Federated Averaging (FedAvg) with Flower. Model performance is evaluated across different training-history lengths (1–48 months) using a standardized event-driven benchmark and unified multi-metric evaluation.

Goals. Determine whether federated learning can match or exceed centralized DGMR training across climate-diverse radar regions, and quantify how model skill evolves as the available training history increases under a fixed training budget.

Methods

- **Data & clients.** NOAA MRMS PrecipRate data (2-min cadence, ~1 km grid) are partitioned into seven radar-centered subdomains (~333×333 km; 3°×3° windows), each representing an FL client (**Fig.1**). Preprocessing converts GRIB2→NetCDF, filters precipitation-informative samples, and extracts 256×256 training crops (**Fig.2**).
- **Prediction task & models.** Each sample contains 4 input frames and 12 forecast frames (2-min intervals). DGMR is used as the primary generative nowcasting model. As a classical baseline we use the STEPS algorithm implemented in PySTEPS, a widely used extrapolation-based radar nowcasting method.
- **Training protocols.** Centralized DGMR (pooled data) is compared with federated DGMR trained using FedAvg. Both use matched training budgets (100 epochs/ rounds), full client participation, and identical validation procedures. An example output see **Fig.3.**
- **Evaluation.** Models are assessed using a fixed event-driven benchmark built from authoritative precipitation-event catalogs. Metrics (e.g., CSI, POD, CRPS, PSNR, spectral scores, **Fig.4**) are averaged across lead times, and an equal-weight TOPSIS score summarizes multi-metric performance (**Fig.3**).

CONCLUSIONS

- Federated DGMR achieves performance comparable to centralized training under the same training budget, and can outperform it when training data are limited.
- This advantage suggests that federated learning provides a useful regularization effect in small or heterogeneous data regimes.
- As the available training history increases, centralized training gradually closes the gap and the two paradigms converge.
- A standardized event-driven benchmark enables fair cross-model evaluation beyond "easy" no-rain cases.
- Overall, these results indicate that federated learning offers a practical approach for training generative precipitation nowcasting models across climate-diverse radar networks without requiring centralized data aggregation.

REFERENCES

- Xu, Z., et al.(2025): X-band Radar Data Case Study. 105th Annual AMS Meeting.
- Zhang, J., et al. (2016): MRMS QPE overview. *BAMS*.
- Ravuri, S., et al. (2021): DGMR nowcasting. *Nature*.
- McMahan, H. B., et al. (2017): FedAvg. *AISTATS*.
- Beutel, D. J., et al. (2020): Flower FL framework. *arXiv*.
- Pulkkinen, S., et al. (2019): PySTEPS. *GMD*.
- Hwang, C.-L., & Yoon, K. (1981): TOPSIS. *Springer*.

EXPERIMENTS & RESULTS

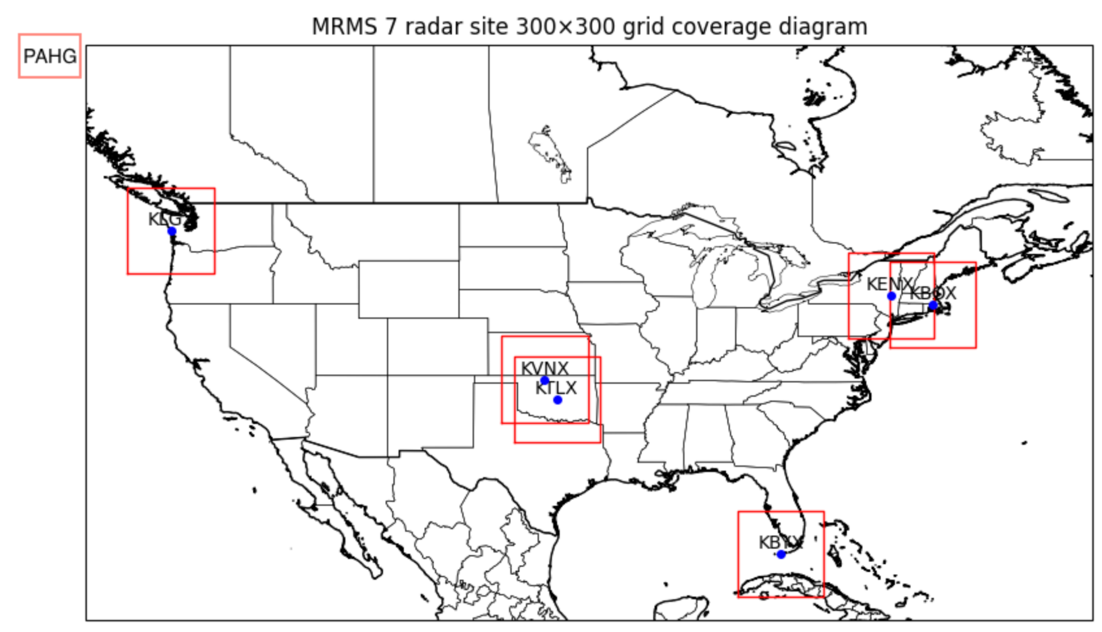

Fig.1, MRMS Data Region Splitting: 7 Regions in total

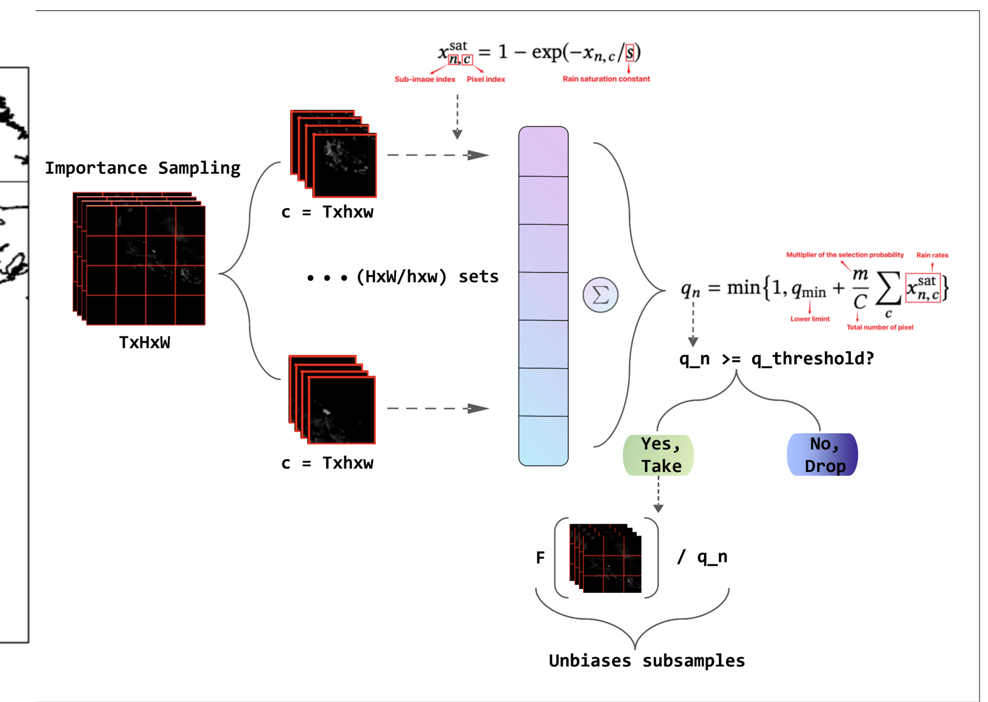

Fig. 2, Data Pre-Processing: Importance Sampling

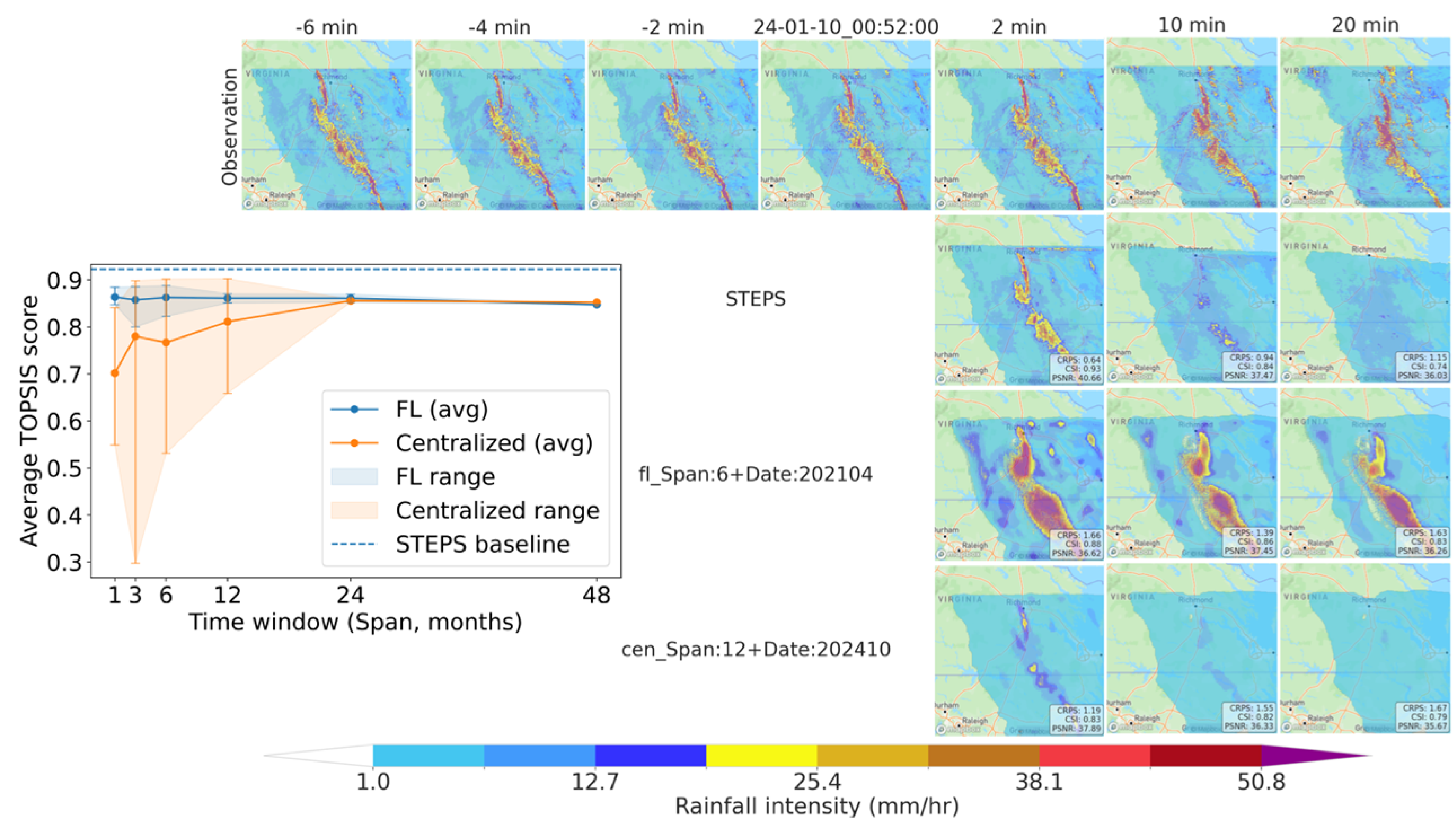

Fig.3, TOPSIS Trend & Example of the "Champion Models" (STEPS, Centralized and Federated trained)

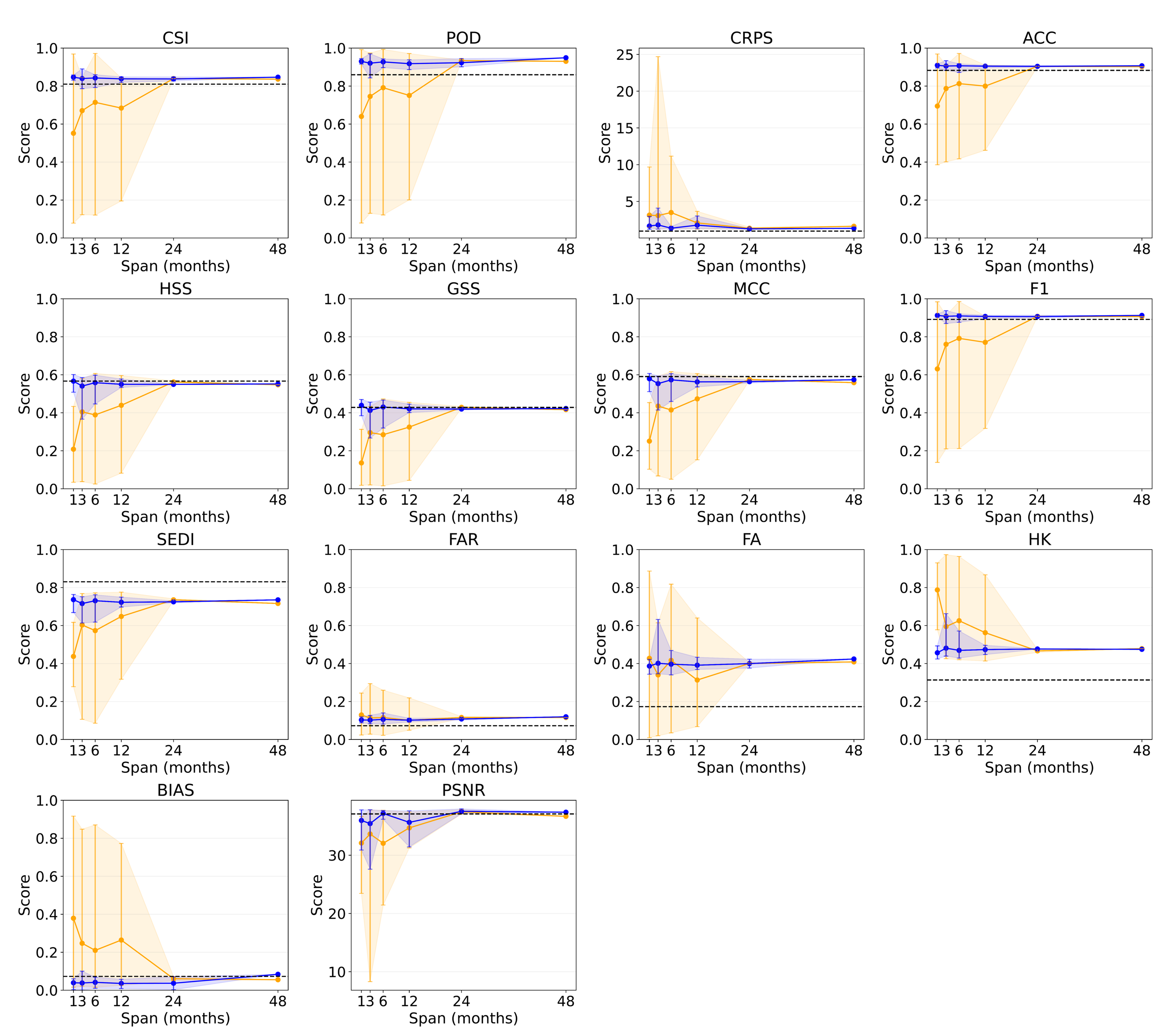

Fig.4, Model Performance Quantitive Metrics Results

- **CSI**(Critical Success Index)↑
- **POD**(Probability of Detection)↑
- **CRPS**(Continuous Ranked Probability Score)↓
- **ACC**(Accuracy)↑
- **HSS**(Heidke Skill Score)↑
- **GSS**(Gilbert Skill Score)↑
- **MCC**(Matthews Correlation Coefficient)↑
- **F1**(F1 Score)↑
- **SEDI**(Symmetric Extreme Dependency Index)↑
- **FAR**(False Alarm Ratio)↓
- **FA**(False Alarms)↓
- **HK**(Hanssen–Kuiper)↑
- **BIAS**(Frequency Bias)→1
- **PSNR**(Peak Signal-to-Noise Ratio)↑

ACKNOWLEDGEMENTS

This work has been partially supported by **Jerome M. and Linda L. Paros** whose generosity established the Paros Center for Atmospheric Research.