1. Model hyperparameters and additional results

Some additional results related to metrics and evaluation are presented below, such as HSS and CSI scores for the city of Toronto. Prediction samples in both GOES-16 and IMERG data are also included.

1.1. Model Hyperparameters

The hyperparameters for Earthformer and TUPANN (both VED and MaxViT) are shown below. Hyperparameters Evolution Network/NowcastNet were the ones selected in the original paper (Zhang et al., 2023), trained up to 20 epochs each. Since CasCast uses a Denoising Transformer (DiT) model trained with a different image size (384x384), slight adaptations were made to support the 256x256 images used in this work. The hyperparameters modified were input_size to 32 and hidden_size to 512; other hyperparameters were chosen as in their original paper (Gong et al., 2024). CasCast was trained for up to 300 epochs and the chosen configuration was the one that minimized the L2 distance between predicted and ground truth noise in the validation set. Excluding CasCast, all remaining hyperparameters were selected by maximizing the mean CSI value in the validation set during training. The selected model checkpoint was also chosen by maximizing the same CSI value in the validation set for these models.

Table 1. *Model hyperparameters for VED, Earthformer, and MaxViT.*

(a) VED

Parameter	Value
n_epochs	150
batch_size	8
learning_rate	0.0001
channels	128
embed_dim	4
reduc_factor	4
λ_{cos}	0.00165
λ_{KL}	1.0e-06
λ_{motion}	0.0033
λ_{int}	0.995
dropout	0.2

(b) Earthformer

Parameter	Value
n_epochs	50
batch_size	4
learning_rate	0.0001
num_global_vectors	6
num_heads	2
base_units	64

(c) MaxViT

Parameter	Value
n_epochs	20
batch_size	8
learning_rate	0.0001
MaxViT_depth	4
MaxViT_dim	64

1.2. HSS results

Table 2 summarizes aggregated HSS metrics for the four cities using GOES-16 data. TUPANN achieves the best scores for high thresholds and remains competitive for lower thresholds, mirroring the CSI behaviour.

1.3. Visual inspection of model predictions

Figure 1 includes predictions for several models at a given moment in time. While TUPANN and EarthFormer show relatively blurred predictions, note NowcastNet and CasCast add several artifacts to its predictions. Overall, TUPANN achieves a reasonable trade-off between good evaluation metrics and

Table 2. *Aggregated HSS metrics for GOES-16 data across cities. Bold values denote the best, underlined values the second best. TUPANN excels at high thresholds and is second or first at lower thresholds.*

Model	HSS-M ↑	HSS_4 ↑	HSS_8 ↑	HSS_{16} ↑	HSS_{32} ↑	HSS_{64} ↑
	Rio de Janeiro					
Earthformer	0.360	0.473	0.438	0.488	0.382	0.017
NowcastNet	0.373	0.454	0.429	0.478	0.394	0.112
PySTEPS (LK)	0.266	0.369	0.341	0.365	0.247	0.010
PySTEPS (DARTS)	0.268	0.356	0.347	0.369	0.244	0.025
CasCast	0.266	0.438	0.321	0.271	0.270	0.032
TUPANN (ours)	**0.393**	**0.473**	**0.439**	**0.492**	**0.428**	**0.135**
	Miami					
Earthformer	0.230	**0.412**	0.299	0.265	0.176	0.000
NowcastNet	0.224	0.368	0.277	0.223	0.177	0.076
PySTEPS (LK)	0.195	0.299	0.227	0.213	0.147	0.087
PySTEPS (DARTS)	0.192	0.301	0.231	0.210	0.135	0.083
CasCast	0.237	0.388	0.304	0.247	0.208	0.039
TUPANN (ours)	**0.277**	0.398	**0.309**	**0.298**	**0.237**	**0.146**
	Manaus					
Earthformer	0.417	**0.502**	**0.476**	**0.472**	0.414	0.220
NowcastNet	0.384	0.457	0.437	0.429	0.372	0.228
PySTEPS (LK)	0.319	0.385	0.369	0.349	0.271	0.222
PySTEPS (DARTS)	0.314	0.386	0.371	0.350	0.269	0.195
CasCast	0.401	0.486	0.447	0.444	0.406	0.222
TUPANN (ours)	**0.435**	0.479	0.464	0.469	**0.430**	**0.333**
	La Paz					
Earthformer	0.454	**0.487**	**0.486**	**0.523**	0.485	0.285
NowcastNet	0.439	0.472	0.451	0.479	0.458	0.335
PySTEPS (LK)	0.340	0.378	0.381	0.390	0.326	0.223
PySTEPS (DARTS)	0.356	0.398	0.405	0.412	0.339	0.225
CasCast	0.351	0.441	0.381	0.381	0.370	0.183
TUPANN (ours)	**0.468**	0.482	0.481	0.512	**0.490**	**0.376**

reasonable precipitation plots. Figure 2 also includes precipitation plots for the IMERG dataset, where the temporal and spatial resolution of the image is coarser.

1.4. Results for Toronto

For completeness Table 3 reports CSI scores for the city of Toronto using GOES-16 data. This city was excluded from the main text once it did not present as many extreme precipitation events as the other selected alternatives. The multi-city TUPANN model performs best across most thresholds, while the single-city TUPANN model ranks second. Although extreme precipitation events are rare in Toronto, these results demonstrate that TUPANN generalizes to additional regions.

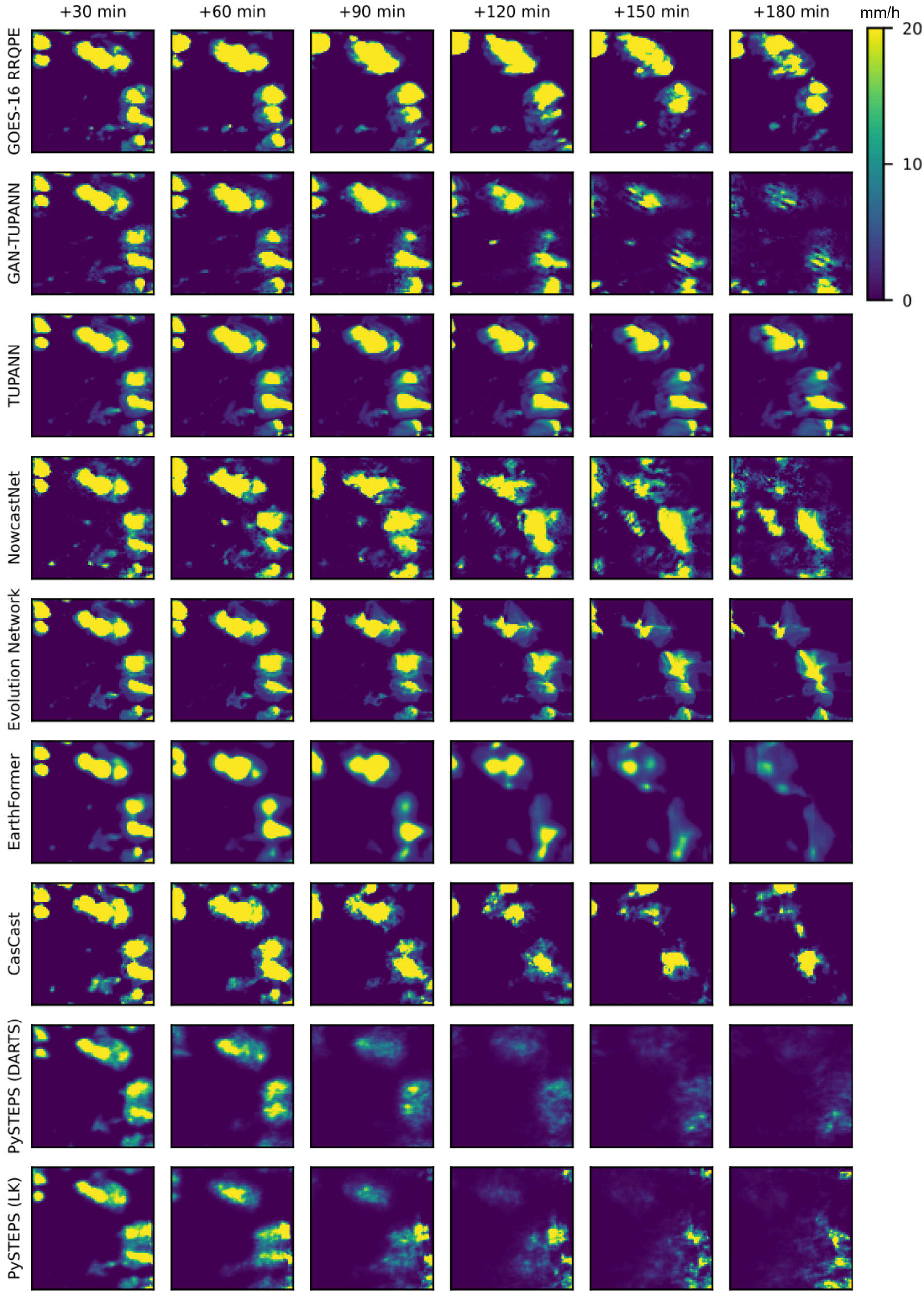

Figure 1. *Predictions up to 3h ahead for a rain event in Manaus, starting at 2021-11-20 22:00 UTC. Rows show GOES-16 RRQPE observations (ground truth), TUPANN, and other models; columns represent lead times. TUPANN displays greater skill in predicting the movement and intensity of the rain event, while generative models produce visually sharper but less accurate predictions.*

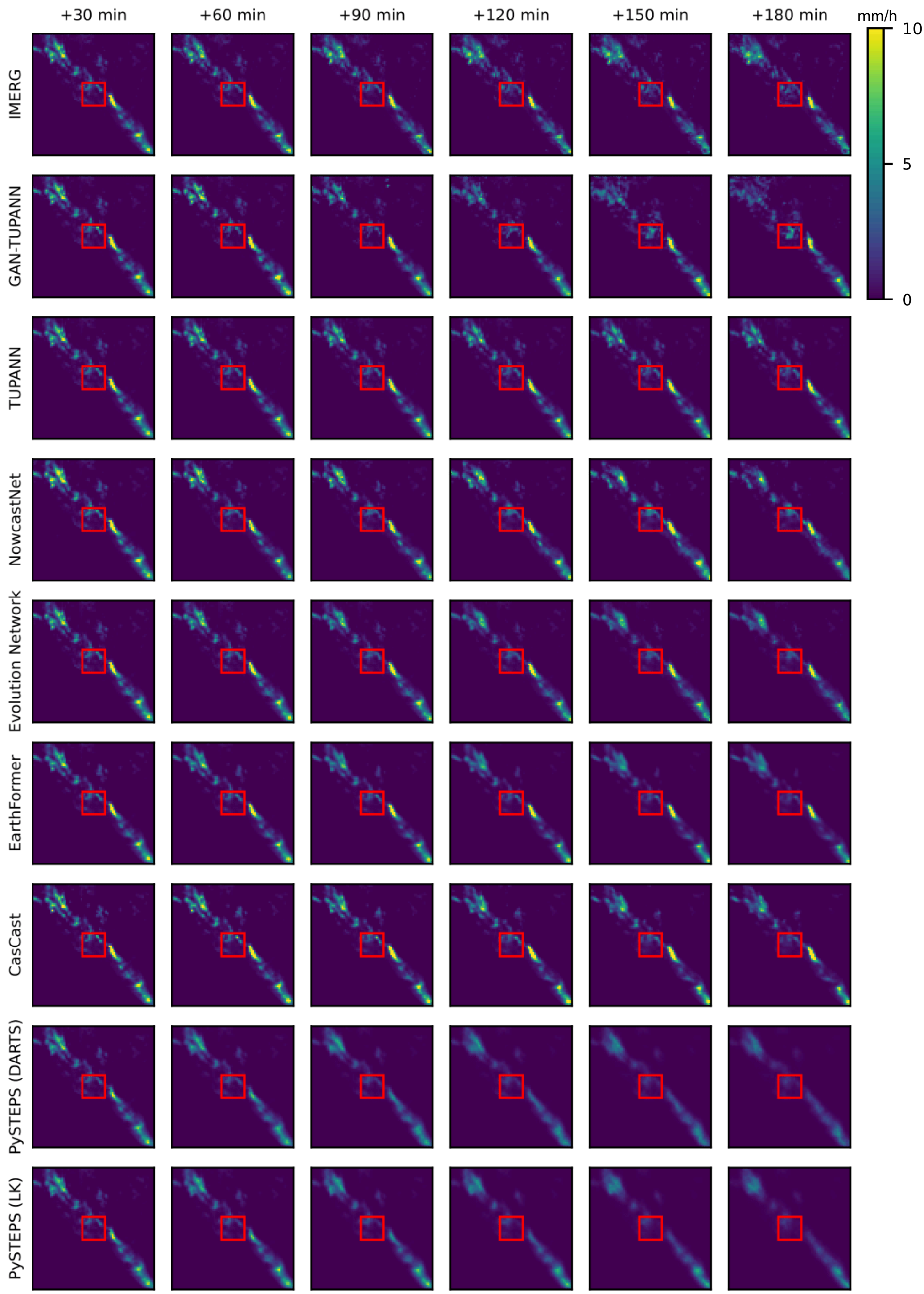

Figure 2. *Prediction sample from the IMERG test dataset centered in Rio de Janeiro, starting from 2023-01-08 07:30 UTC. The first row shows the IMERG observations (ground truth). The red square represents the same area as in the GOES-16 figures, demonstrating the difference in spatial resolution.*

Table 3. *Aggregated CSI metrics for Toronto with GOES-16 data. Bold denotes the best, underlined the second best.*

Model	CSI-M ↑		CSI_4 ↑		CSI_8 ↑		CSI_{16} ↑		CSI_{32} ↑		CSI_{64} ↑	
	POOL1	POOL4	POOL1	POOL4	POOL1	POOL4	POOL1	POOL4	POOL1	POOL4	POOL1	POOL4
Earthformer	0.209	0.187	0.288	0.280	0.233	0.203	0.154	0.133	0.102	0.084	0.000	0.000
NowcastNet	0.206	<u>0.216</u>	0.278	0.294	0.225	0.238	0.151	0.161	0.100	0.105	0.000	0.000
PySTEPS (LK)	0.179	0.173	0.250	0.240	0.200	0.186	0.130	0.120	0.078	0.070	0.000	0.000
PySTEPS (DARTS)	0.181	0.174	0.252	0.242	0.203	0.188	0.133	0.121	0.080	0.071	0.000	0.000
CasCast	0.192	0.203	0.270	<u>0.286</u>	0.217	<u>0.231</u>	0.145	<u>0.158</u>	0.096	<u>0.105</u>	0.000	0.000
TUPANN	<u>0.219</u>	<u>0.222</u>	<u>0.295</u>	0.295	<u>0.239</u>	<u>0.231</u>	<u>0.159</u>	<u>0.158</u>	<u>0.106</u>	<u>0.105</u>	0.000	0.000
TUPANN–Multicity	**0.229**	**0.230**	**0.305**	**0.298**	**0.250**	**0.241**	**0.170**	**0.165**	**0.114**	**0.107**	0.000	0.000

References

Gong, J., Bai, L., Ye, P., Xu, W., Liu, N., Dai, J., Yang, X., and Ouyang, W. (2024). Cascast: skillful high-resolution precipitation nowcasting via cascaded modelling. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org.

Zhang, Y., Long, M., Chen, K., Xing, L., Jin, R., Jordan, M. I., and Wang, J. (2023). Skilful nowcasting of extreme precipitation with nowcastnet. *Nature*, 619(7970):526–532.