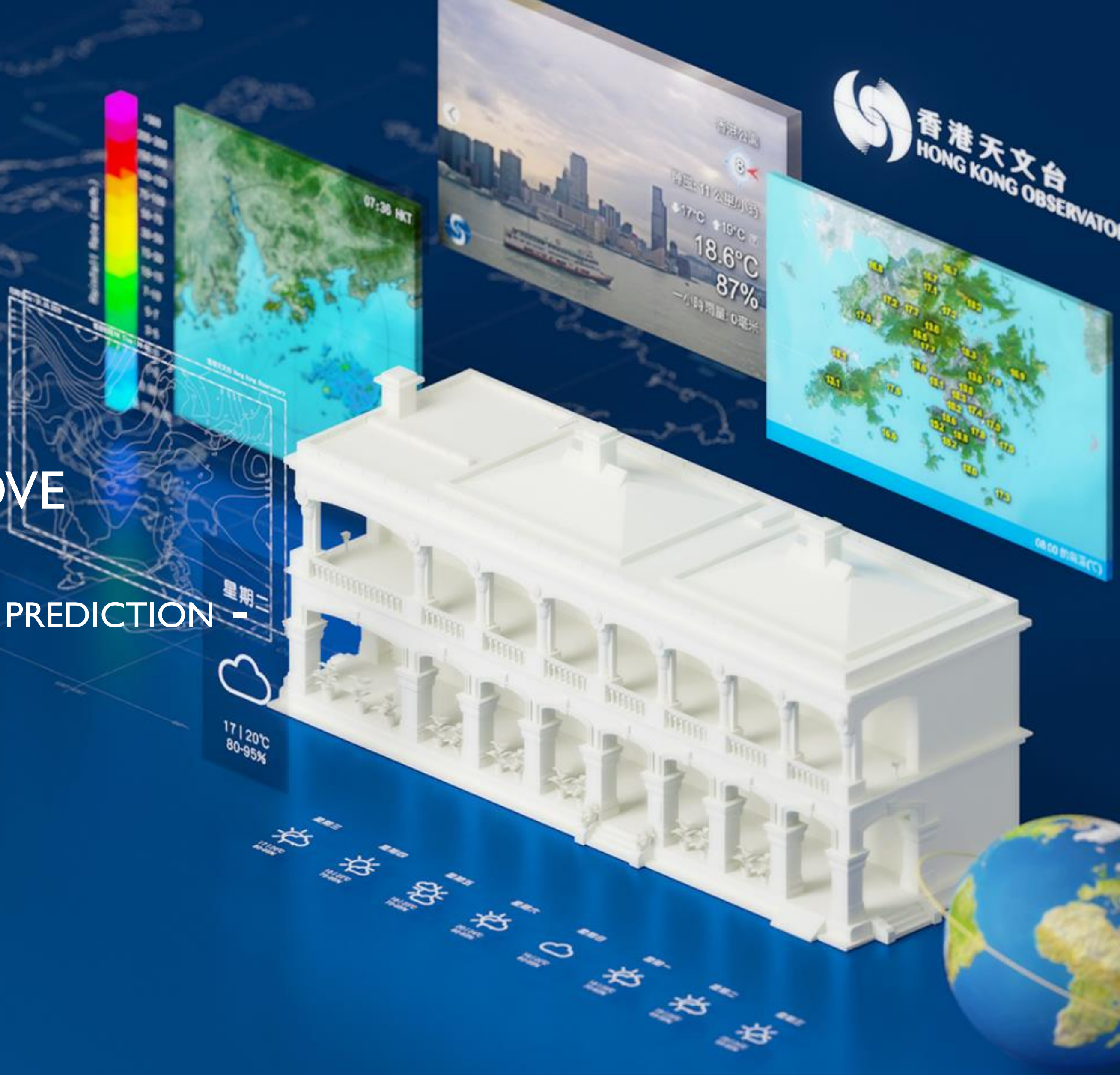
香港天文台
HONG KONG OBSERVATORY

HOW MACHINE LEARNING CAN IMPROVE TROPICAL CYCLONE FORECASTS

PERSPECTIVES FROM RAINFALL NOWCAST AND INTENSITY PREDICTION –



TRCG Special Session

17th Integrated Workshop (Online)

30 November 2022

Wai-Kin Wong

Chairman of TRCG, ESCAP/WMO Typhoon Committee

Senior Scientific Officer, Forecast Development Division, HKO

E-mail: wkwong@hko.gov.hk

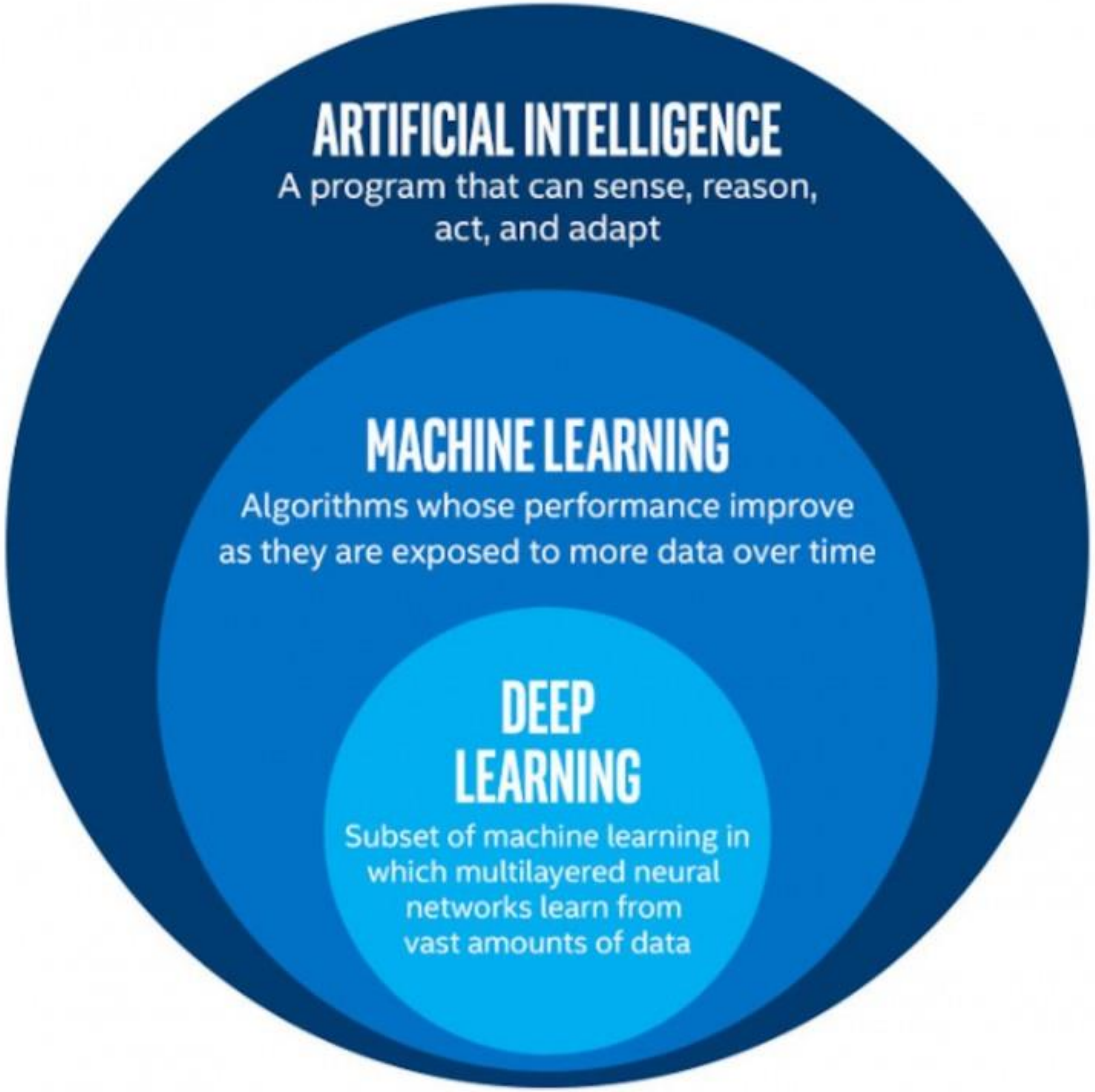
From Artificial Intelligence (AI) to Machine Learning (ML) and Deep Learning

Main categories of ML algorithms:

(a) Unsupervised
- Unlabeled data in training

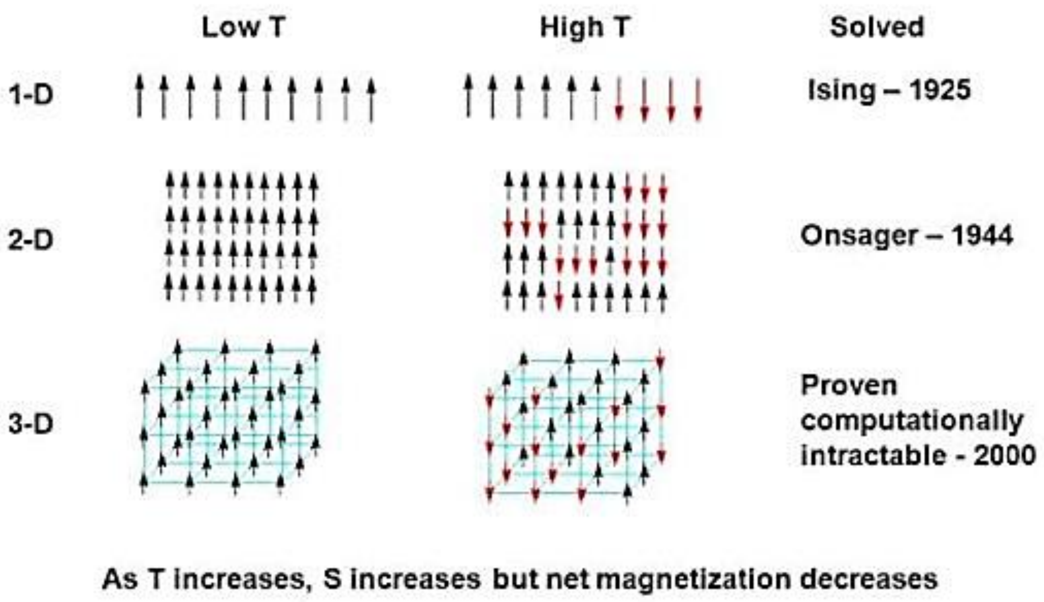
(b) Supervised
- Learning from labeled data

(c) Reinforcement
- Works on a basic principle of positive and negative feedback

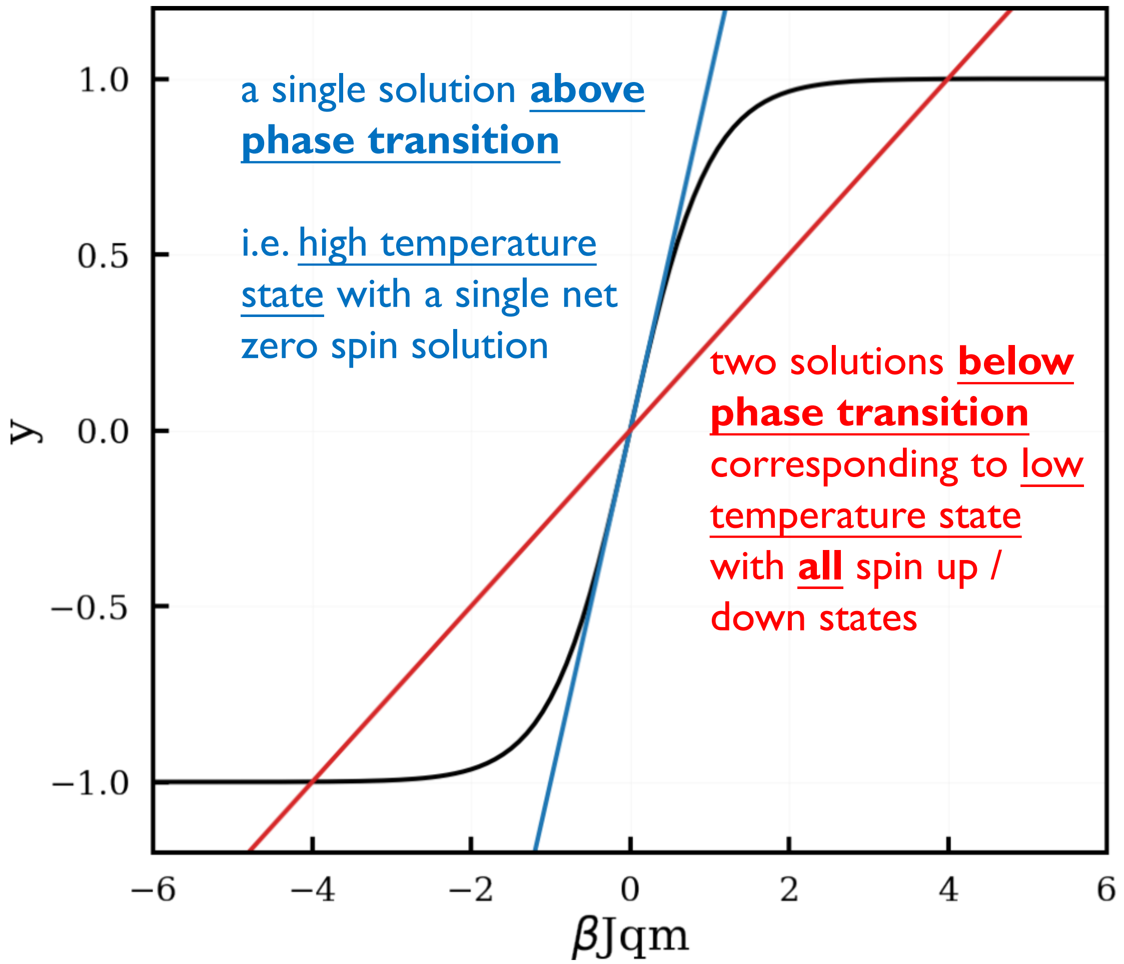


Intuition from physics in understanding basic principles of AI / ML

- Ising Model – ferromagnetism model where atoms in a solid lattice can be either spin-up or spin-down



Solution of magnetization order parameter (m) under nil background magnetic field



Machine Learning using Feed-forward Neural Network (FFNN)

- NN – mapping high dimensional space to a smaller space
- A simple example – linear NN:
  - Note: multiple layers of linear NN is still linear

Weight matrix

Output

bias

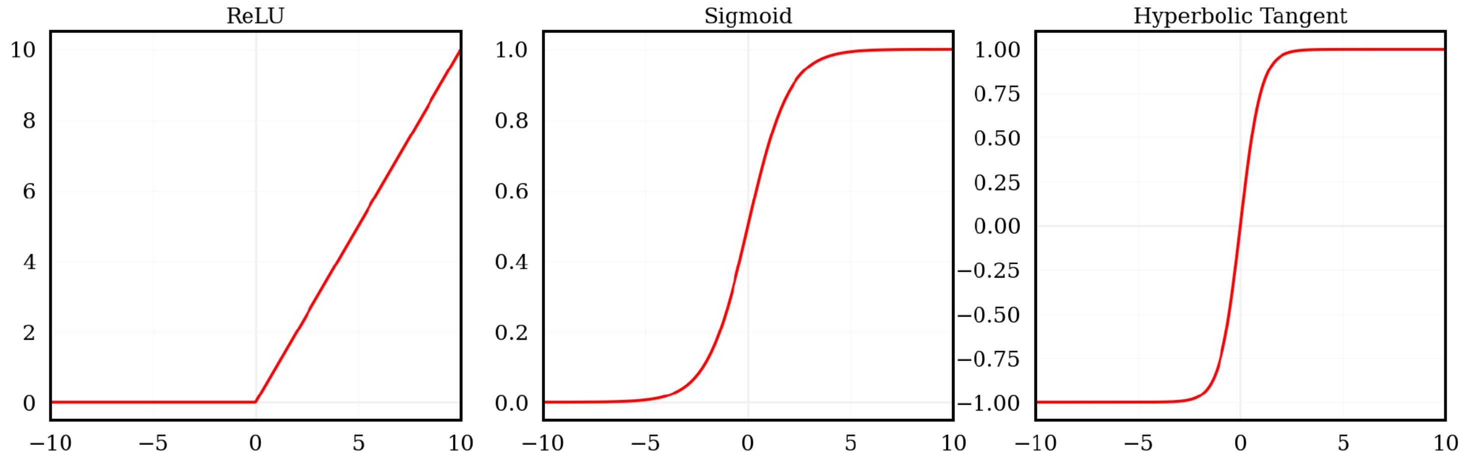
$$\mathbf{y} = \mathbf{W}\ \mathbf{x} + \mathbf{b}$$

Input
(e.g. image)

- Non-linearity in mapping between layers
  - Nonlinear activation function $\sigma(z)$

$$\mathbf{y} = \sigma(\mathbf{W}\ \mathbf{x} + \mathbf{b})$$

ReLU –
Rectified Linear Unit

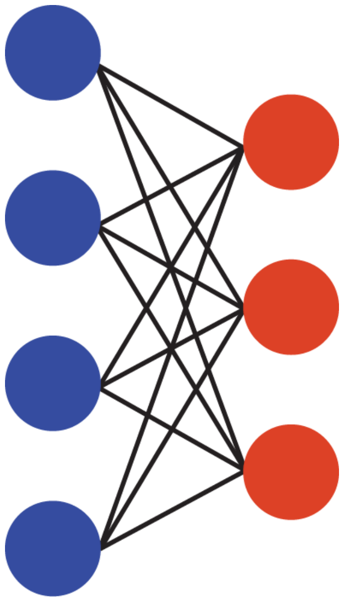


Stacking multiple layers of non-linear activation function in NN

- Analogous to restricted Boltzmann machine (RBM)
  - Additional nodes (and layer) to learn a representation of higher dimensional features from given data set
- Performance is (hopefully) improved by increasing depth of NN
  - Complexity exceeding (degrees of freedom in) data set would result in overfitting
- Relevant degrees of freedom propagate while those irrelevant are integrated out under the mapping through training
  - analogus to renormalization group (RG) flow

Schematic of RBM

Blue – visible layer
Orange – hidden layer



- Generative
  - "create" data / information through learning
- Stochastic
  - through learned probability distribution
- Fully connected
  - eliminate interaction between nodes of the same group (hidden or visible)

Machine Learning using Feed-forward Neural Network (FFNN)

Feed-forward neural network constructs a mapping $\mathbf{Y} = f(\mathbf{X};\theta)$ by stacking various basic blocks such as the fully-connected layer, the convolution layer, the deconvolution layer, and the activation layer.

Common types of FFNN:

(a) multi-layer perceptron (MLP), which stacks multiple fully-connected (FC) layers and nonlinear activations

(b) convolutional neural network (CNN) that stacks multiple convolution layers, pooling layers, deconvolution layers, FC layers, activation layers, normalization layers and other transformations.

The parameters of FFNN are estimated by minimizing the **loss function** plus **regularization terms**

$$\theta^\star = \arg\min_\theta \mathbb{E}_{\hat{p}_{\text{data}}} [l(\mathbf{Y}, f(\mathbf{X};\theta))] + \Omega(\theta)$$

$\Omega \sim L_1$ or L_2 loss

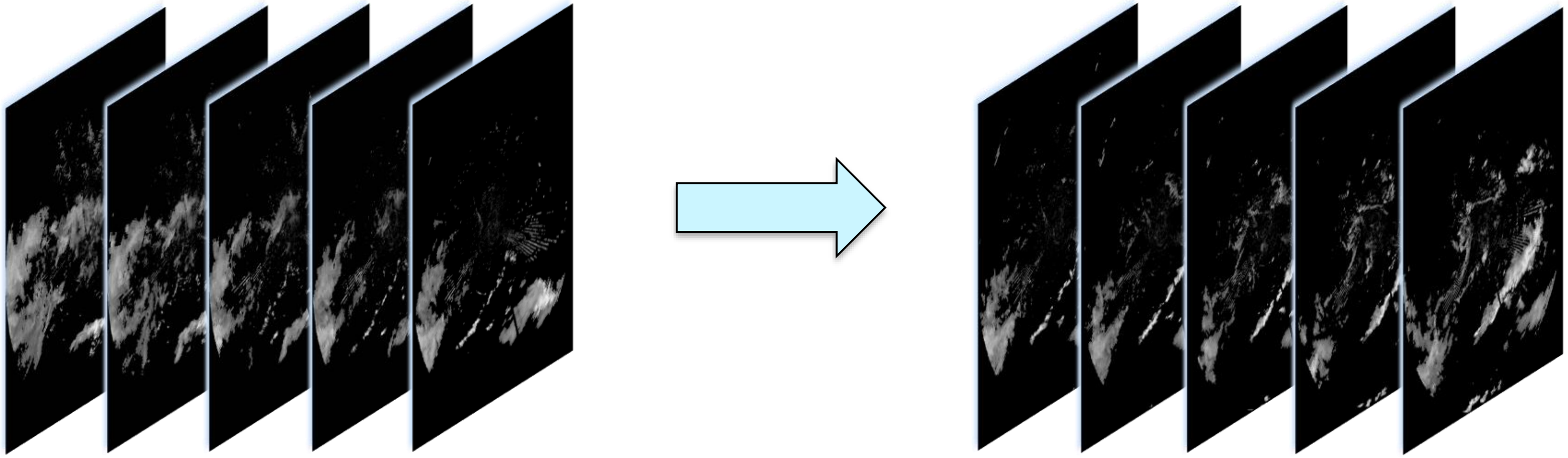
L_1 loss $= \sum_{i=1}^{n} |y_{true} - y_{predicted}|$

L_2 loss $= \sum_{i=1}^{n} (y_{true} - y_{predicted})^2$

Usually, the optimization problem is solved via stochastic-gradient-based methods in which the gradient is computed by **backpropagation**

Predicting evolution of weather radar image as a spatiotemporal sequence forecast

- **Input sequence**: observed radar maps up to current time step
- **Output sequence**: predicted radar maps for future time steps



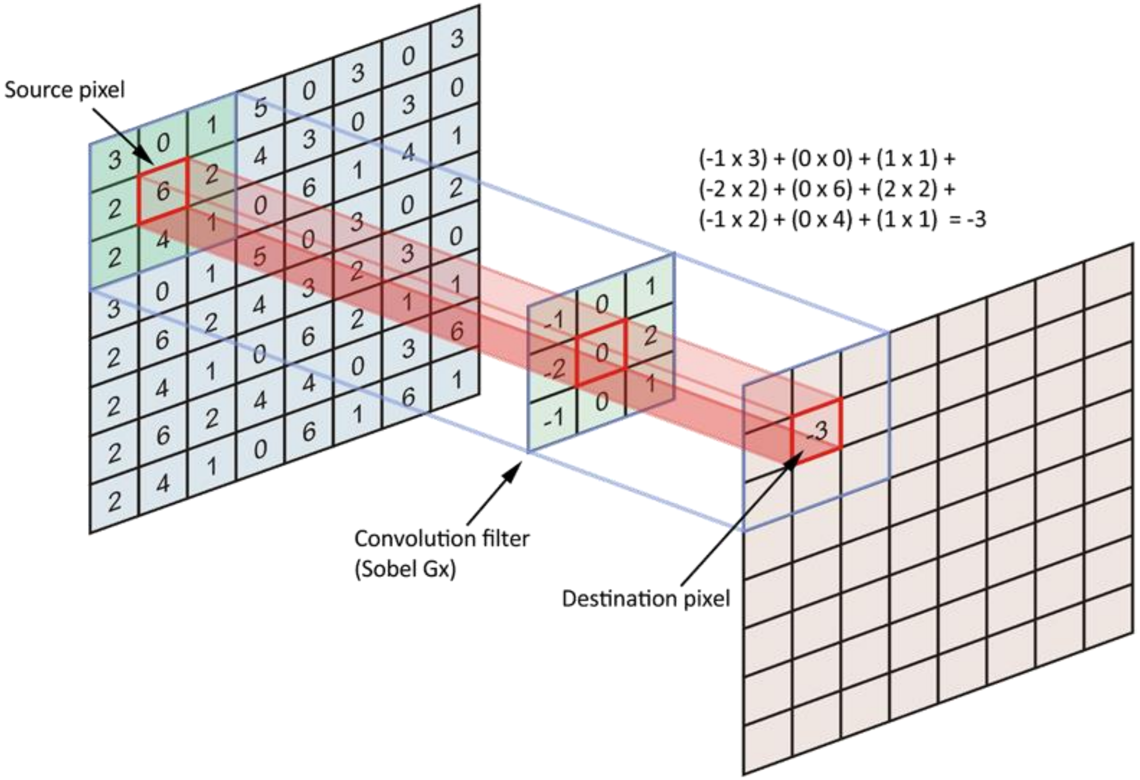
$$\tilde{\mathcal{X}}_{t+1}, \ldots, \tilde{\mathcal{X}}_{t+K} = \underset{\mathcal{X}_{t+1}, \ldots, \mathcal{X}_{t+K}}{\arg\max} \; p(\mathcal{X}_{t+1}, \ldots, \mathcal{X}_{t+K} \mid \hat{\mathcal{X}}_{t-J+1}, \hat{\mathcal{X}}_{t-J+2}, \ldots, \hat{\mathcal{X}}_t)$$

Maximize posterior pdf of echo sequence across K time levels based on previous J time levels of observations

Convolution Layer

A convolution layer takes advantage of the translational invariance property of image data that computes the output by scanning over the input and applying the same set of linear filters.

Although the input can have an arbitrary dimensionality, 2D convolution is commonly used for precipitation nowcasting for extracting the spatial correlation in meteorological images.



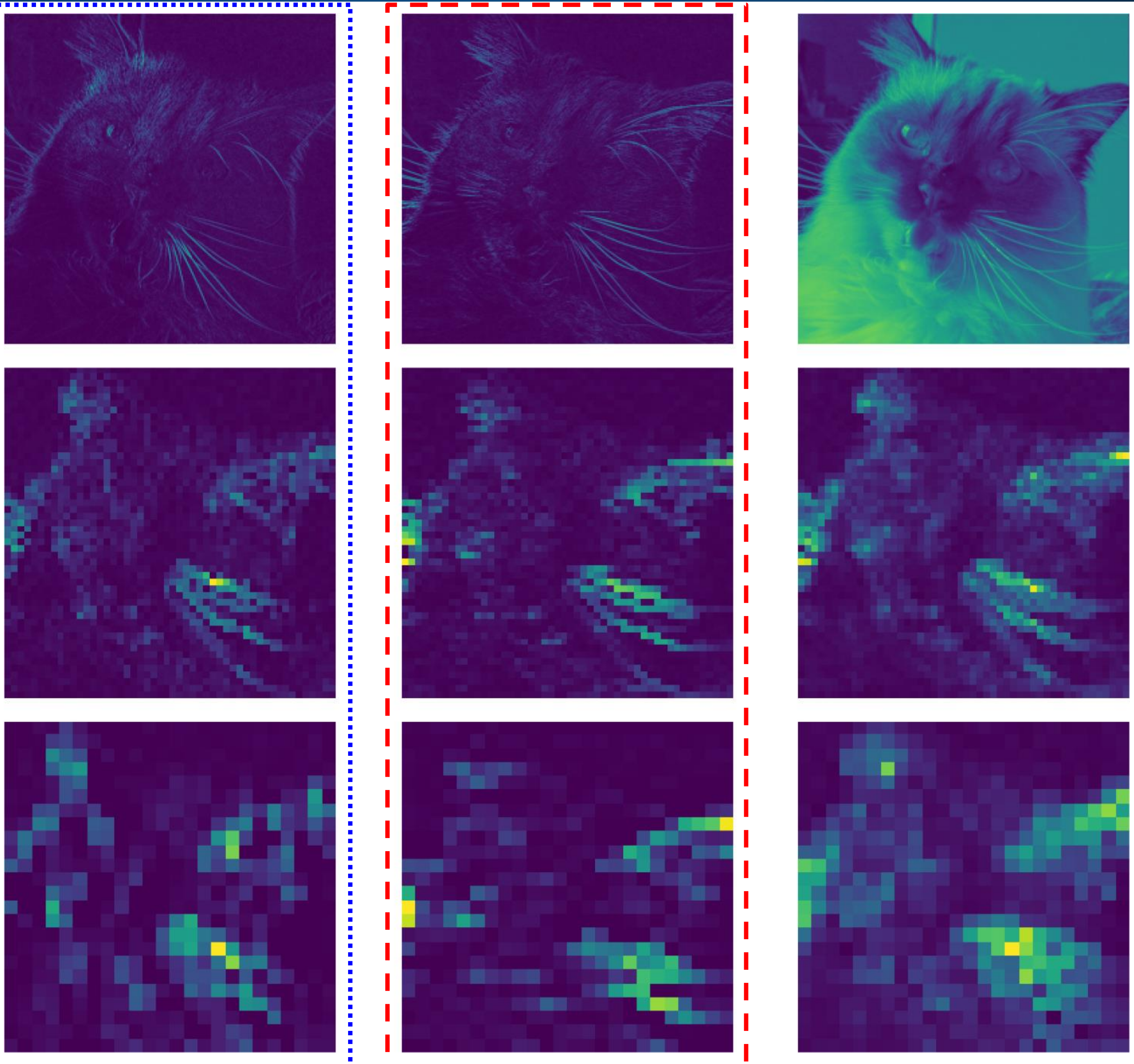
0	1	0
0	1	0
0	1	0

3x3 filter for detecting VERTICAL lines

0	0	0
1	1	1
0	0	0

3x3 filter for detecting HORIZONTAL lines

Convolve output from previous layer using horizontal filter



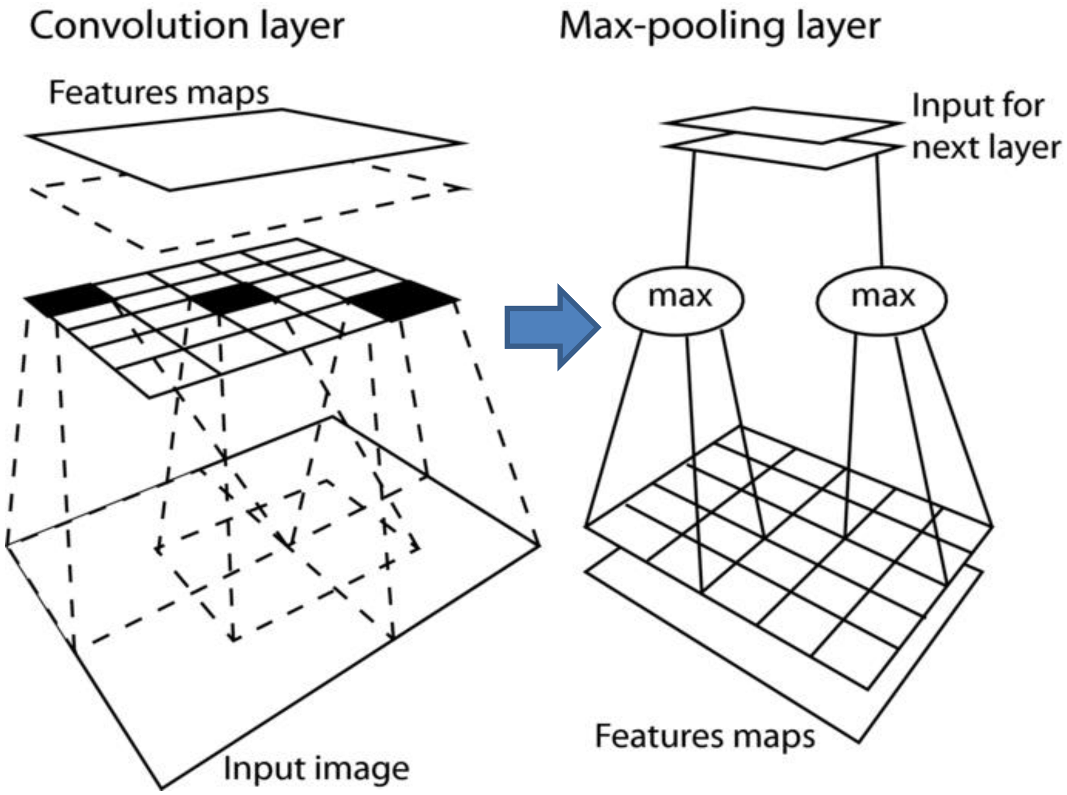
1	0	1
0	1	0
1	0	1

"X-shaped" filter

Convolutional and pooling layers

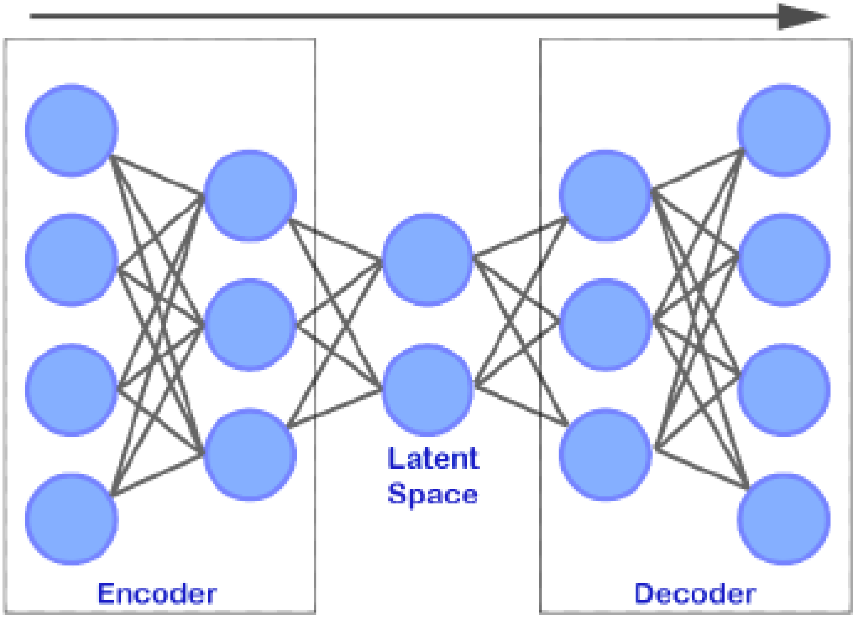
- Convolution: feature detector
- Max-pooling: local translation invariance

Size of state-to-state convolutional kernel for capturing of spatiotemporal motion patterns

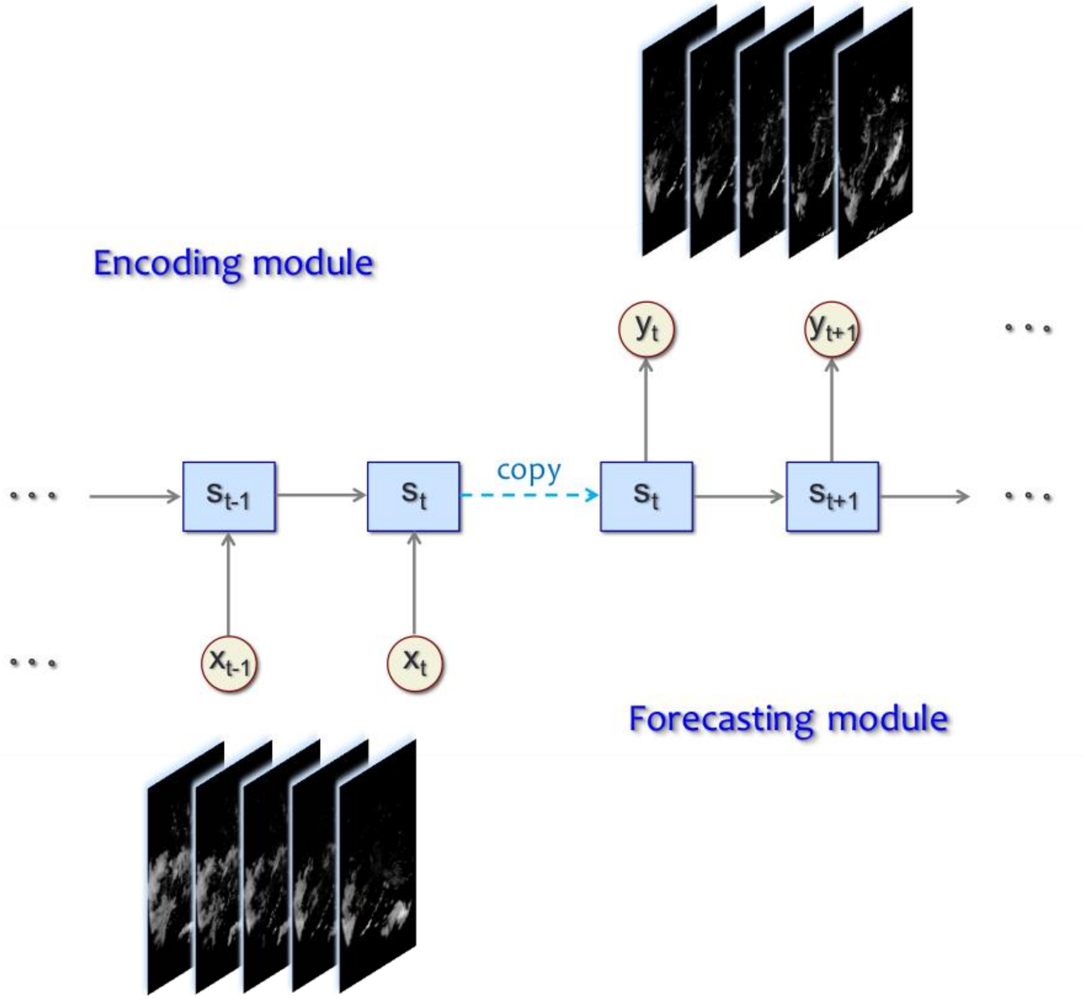


Autoencoders

- 3 components: encoder, decoder and latent space
- Force data sets through a compressed representation of data such that a minimal amount of information is lost
- Common applications: data noise reduction, generative prediction and anomaly detection
- Probability distribution in latent space can be learned via variational autoencoder approach



Spatiotemporal encoding-forecasting model



- Convolutional long short-term memory (ConvLSTM) model
  - X. Shi, Z. Chen, H. Wang, D.Y. Yeung, W.K. Wong and W.C. Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. NIPS 2015.

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY https://arxiv.org/abs/1506.04214

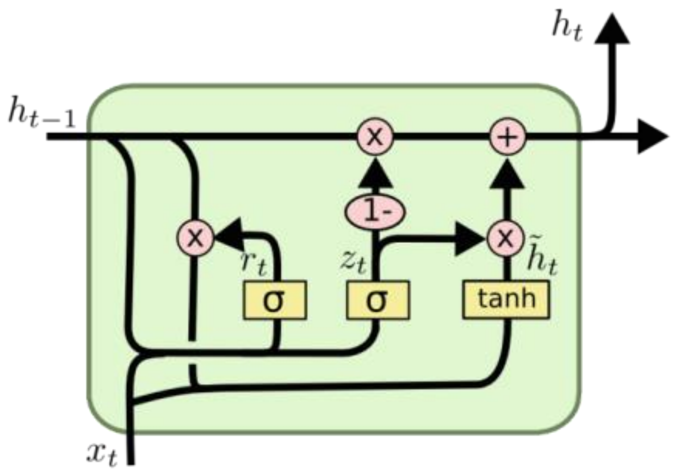
- Two key components:
  - Convolutional layers
  - Long short-term memory (LSTM) cells in recurrent neural network (RNN) model

Trajectory Gated Recurrence Unit (TrajGRU)

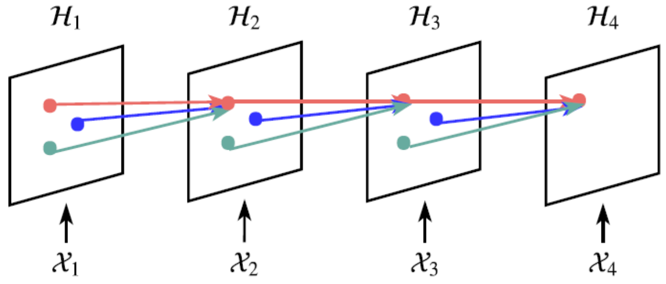
TraGRU replaces LSTM, introduces "Trajectory" and adopts weighted error function

Xingjian Shi, Zhihan Gao, Leonard Lausen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo, 2017: Deep learning for precipitation nowcasting: A benchmark and a new model.
https://arxiv.org/pdf/1706.03458.pdf

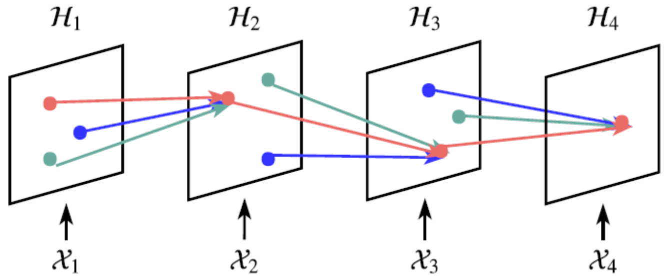
GRU (Gated Recurrent Unit) includes *reset gate* and *update gate*, similar to LSTM but more efficient.



Trajectory:
Recurrent connections are dynamically determined



(a) ConvRNN: Links are fixed over time/location.



(b) TrajRNN: Links are dynamically determined.

Weighted Error:
optimize performance in heavy rain

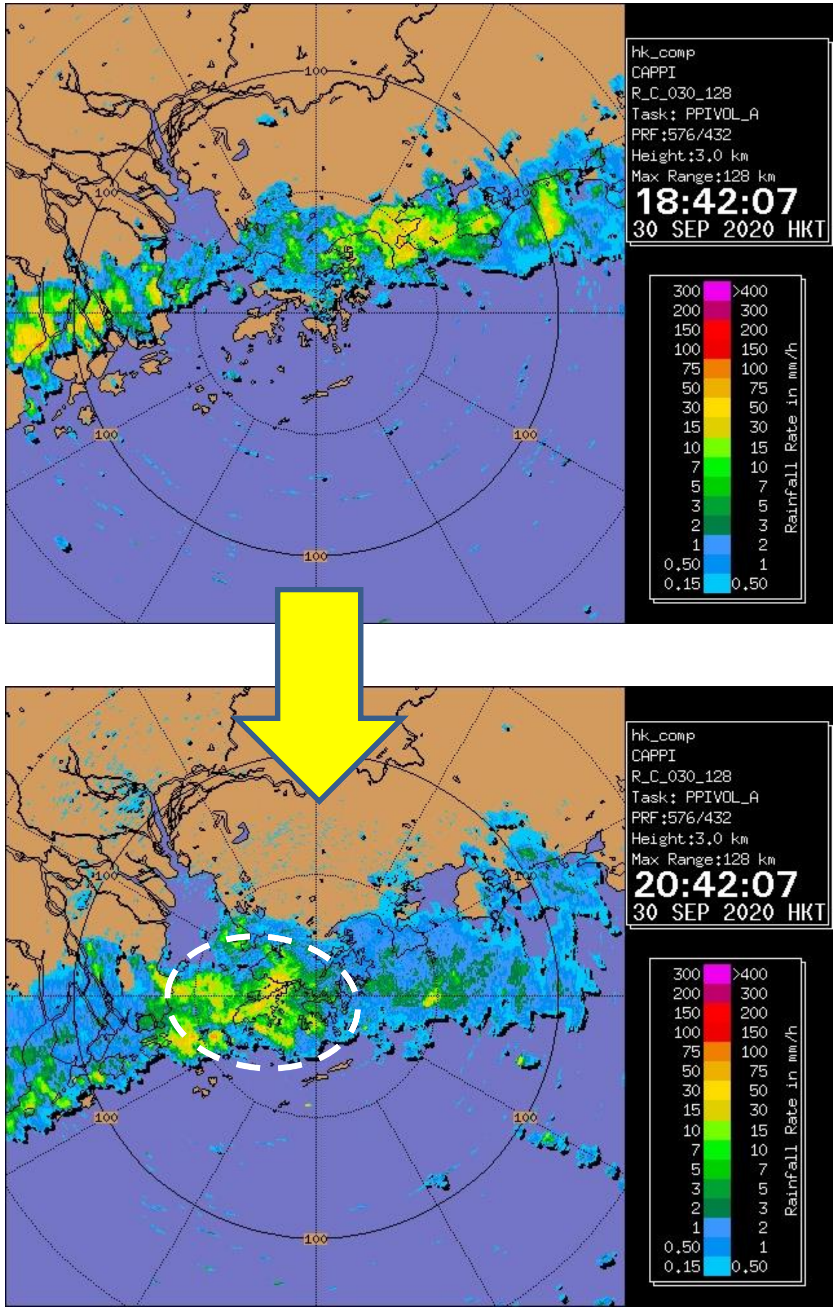
$w(x)$ to each pixel according to its rainfall intensity x: $w(x) = \begin{cases} 1, & x < 2 \\ 2, & 2 \leq x < 5 \\ 5, & 5 \leq x < 10 \\ 10, & 10 \leq x < 30 \\ 30, & x \geq 30 \end{cases}$. Also, the masked pixels have weight 0. The resulting B-MSE and B-MAE scores are computed as B-MSE $= \frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{480}\sum_{j=1}^{480} w_{n,i,j}(x_{n,i,j} - \hat{x}_{n,i,j})^2$ and B-MAE $= \frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{480}\sum_{j=1}^{480} w_{n,i,j}|x_{n,i,j} - \hat{x}_{n,i,j}|$, where N is the total number of frames and $w_{n,i,j}$ is the weight corresponding to the (i,j)th pixel in the nth frame. For the conventional MSE and MAE measures, we simply set all the weights to 1 except the masked points.

Table 3: HKO-7 benchmark result. We mark the best result within a specific setting with **bold face** and the second best result by underlining. Each cell contains the mean score of the 20 predicted frames. In the online setting, all algorithms have used the online learning strategy described in the paper. '↑' means that the score is higher the better while '↓' means that the score is lower the better. '$r \geq \tau$' means the skill score at the τmm/h rainfall threshold. For 2D CNN, 3D CNN, ConvGRU and TrajGRU models, we train the models with three different random seeds and report the mean scores.

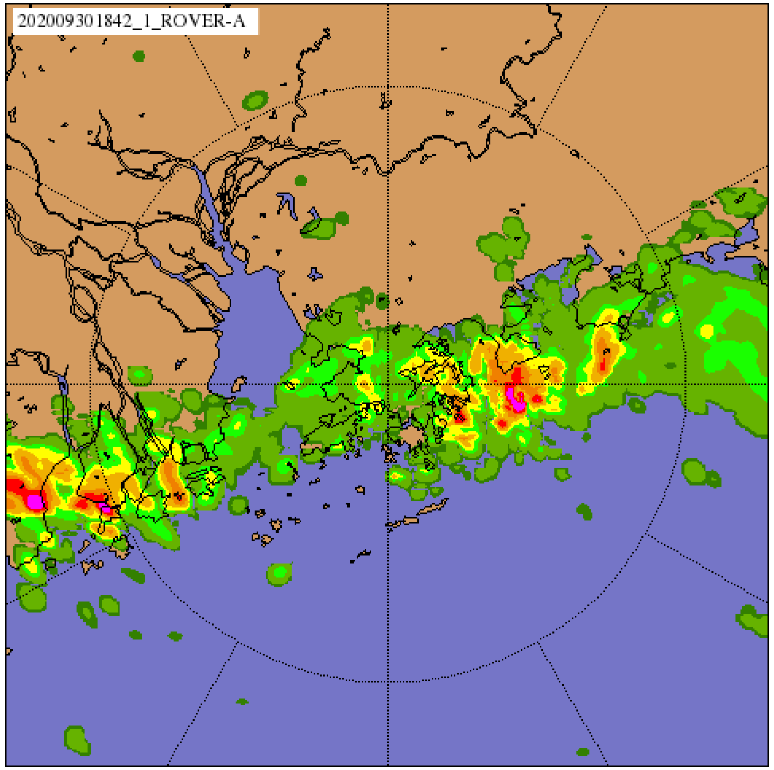
Algorithms	CSI ↑ $r \geq 0.5$	CSI ↑ $r \geq 2$	CSI ↑ $r \geq 5$	CSI ↑ $r \geq 10$	CSI ↑ $r \geq 30$	HSS ↑ $r \geq 0.5$	HSS ↑ $r \geq 2$	HSS ↑ $r \geq 5$	HSS ↑ $r \geq 10$	HSS ↑ $r \geq 30$	B-MSE ↓	B-MAE ↓
Offline Setting												
Last Frame	0.4022	0.3266	0.2401	0.1574	0.0692	0.5207	0.4531	0.3582	0.2512	0.1193	15274	28042
ROVER + Linear	0.4762	0.4089	0.3151	0.2146	0.1067	0.6038	0.5473	0.4516	0.3301	0.1762	11651	23437
ROVER + Non-linear	0.4655	0.4074	0.3226	0.2164	0.0951	0.5896	0.5436	0.4590	0.3318	0.1576	10945	22857
2D CNN	0.5095	0.4396	0.3406	0.2392	0.1093	0.6366	0.5809	0.4851	0.3690	0.1885	7332	18091
3D CNN	0.5109	0.4411	0.3415	0.2424	0.1185	0.6334	0.5825	0.4862	0.3734	0.2034	7202	17593
ConvGRU-nobal	0.5476	0.4661	0.3526	0.2138	0.0712	0.6756	0.6094	0.4981	0.3286	0.1160	9087	19642
ConvGRU	0.5489	0.4731	0.3720	0.2789	0.1776	0.6701	0.6104	0.5163	0.4159	0.2893	5951	15000
TrajGRU	**0.5528**	**0.4759**	**0.3751**	**0.2835**	**0.1856**	**0.6731**	**0.6126**	**0.5192**	**0.4207**	**0.2996**	**5816**	**14675**
Online Setting												
2D CNN	0.5112	0.4363	0.3364	0.2435	0.1263	0.6365	0.5756	0.4790	0.3744	0.2162	6654	17071
3D CNN	0.5106	0.4344	0.3345	0.2427	0.1299	0.6355	0.5736	0.4766	0.3733	0.2220	6690	16903
ConvGRU	0.5511	0.4737	0.3742	0.2843	0.1837	0.6712	0.6105	0.5183	0.4226	0.2981	5724	14772
TrajGRU	**0.5563**	**0.4798**	**0.3808**	**0.2914**	**0.1933**	**0.6760**	**0.6164**	**0.5253**	**0.4308**	**0.3111**	**5589**	**14465**

Optical flow

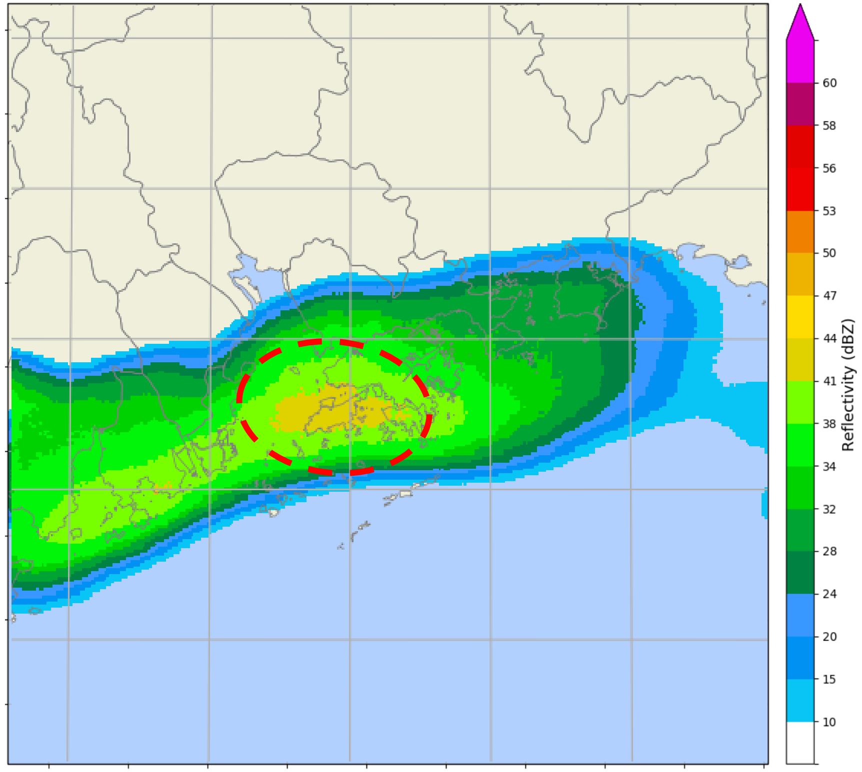
2-h nowcast of radar reflectivity from 2020/09/30 18:42H



Extrapolation using optical flow field

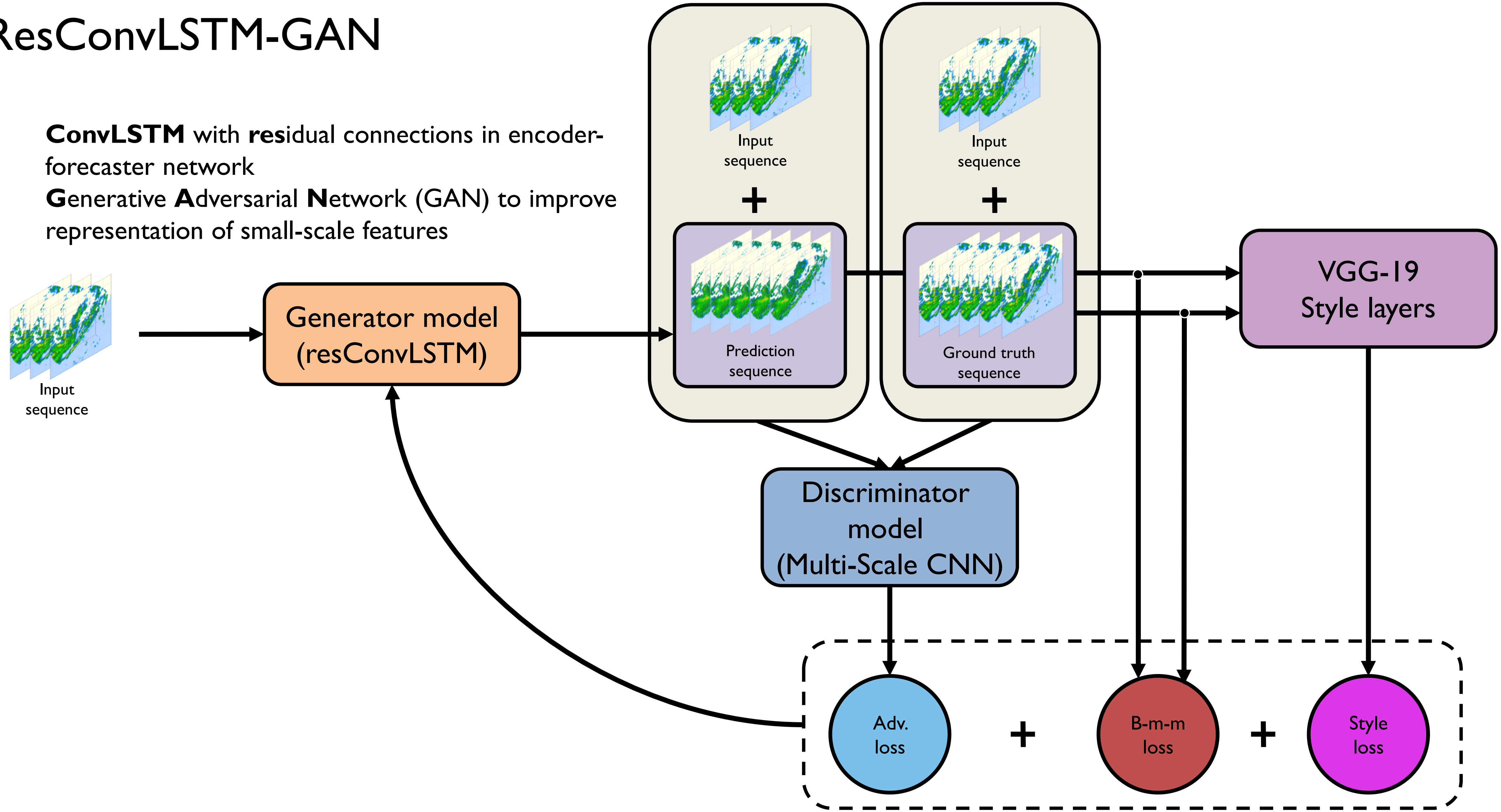


TrajGRU deep learning nowcast



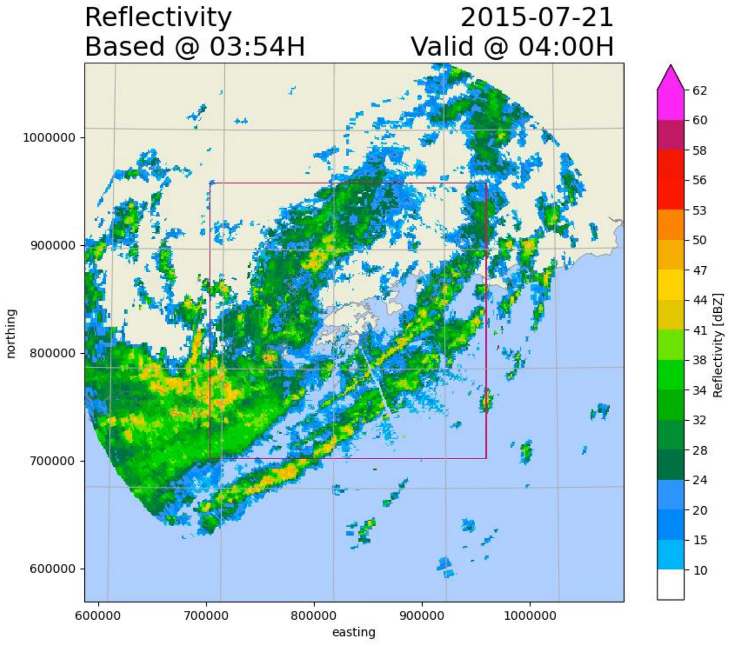
ResConvLSTM-GAN

- **ConvLSTM** with **res**idual connections in encoder-forecaster network
- **G**enerative **A**dversarial **N**etwork (GAN) to improve representation of small-scale features

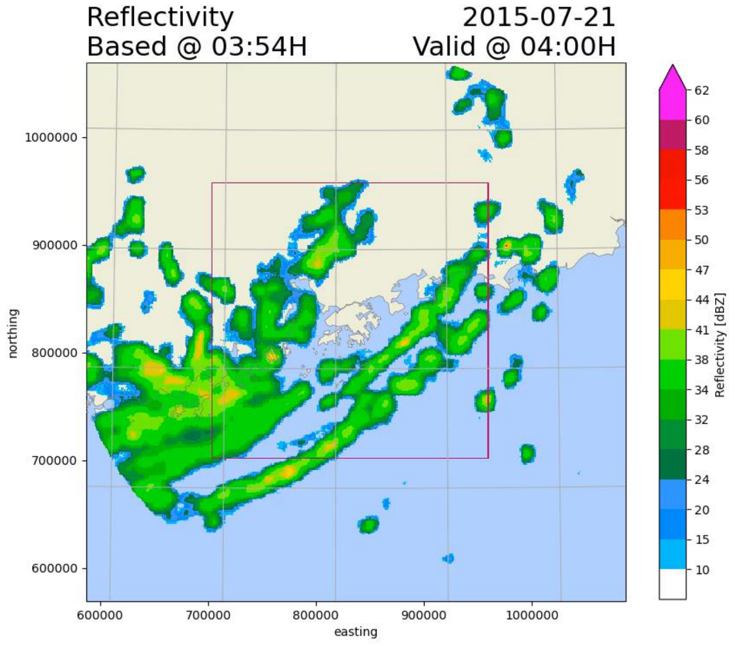


Rainfall Nowcast Using GAN – Example (1)

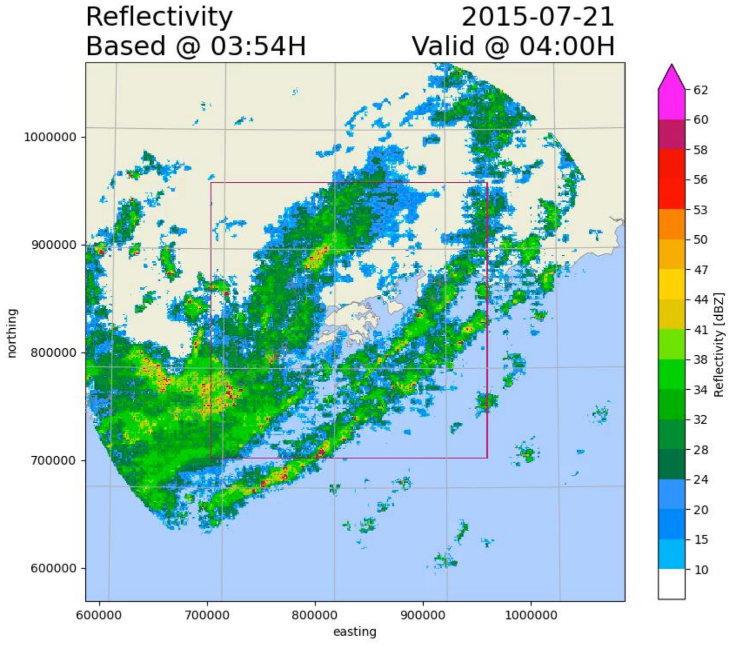
Actual



TrajGRU



ResConvLSTM_GAN

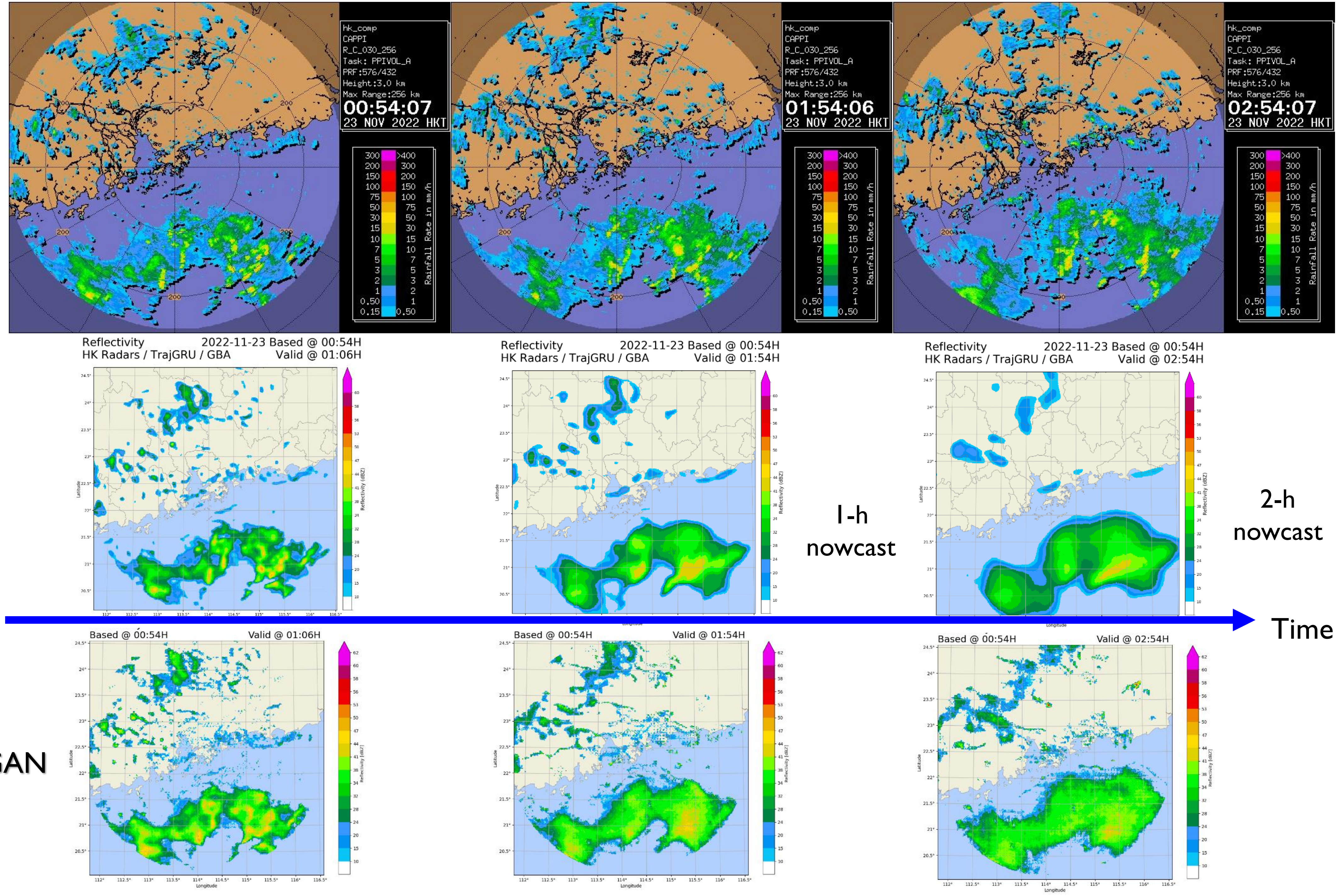


HONG KONG OBSERVATORY 香港天文台

Example from real-time trial

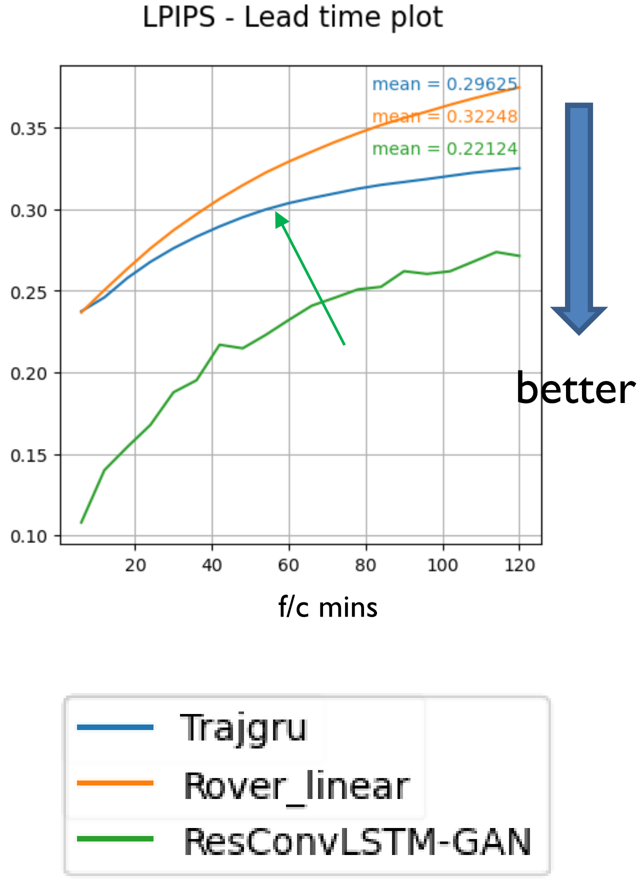
TrajGRU

ResConvLSTM-GAN

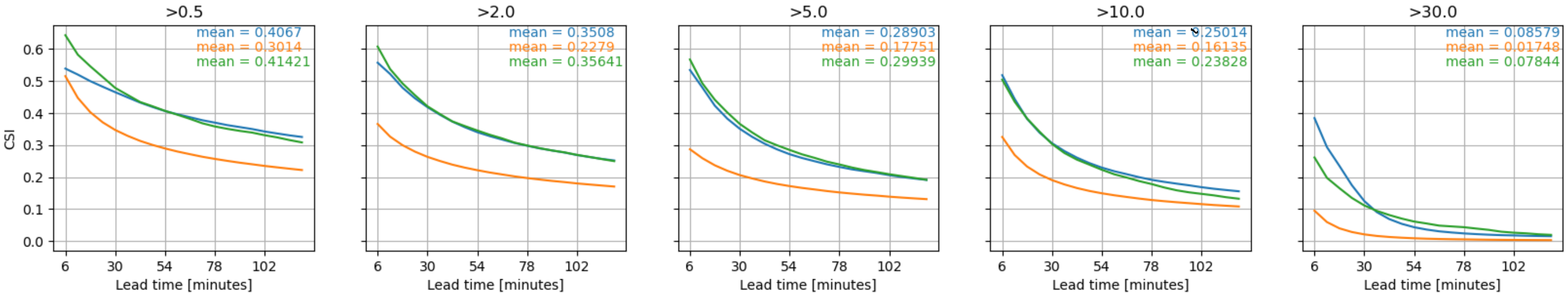


Verification of ResConvLSTM-GAN

- Learned Perceptual Image Patch Similarity (LPIPS)
  - perceptual similarity between two images
  - a low LPIPS score means the images are perceptual similar



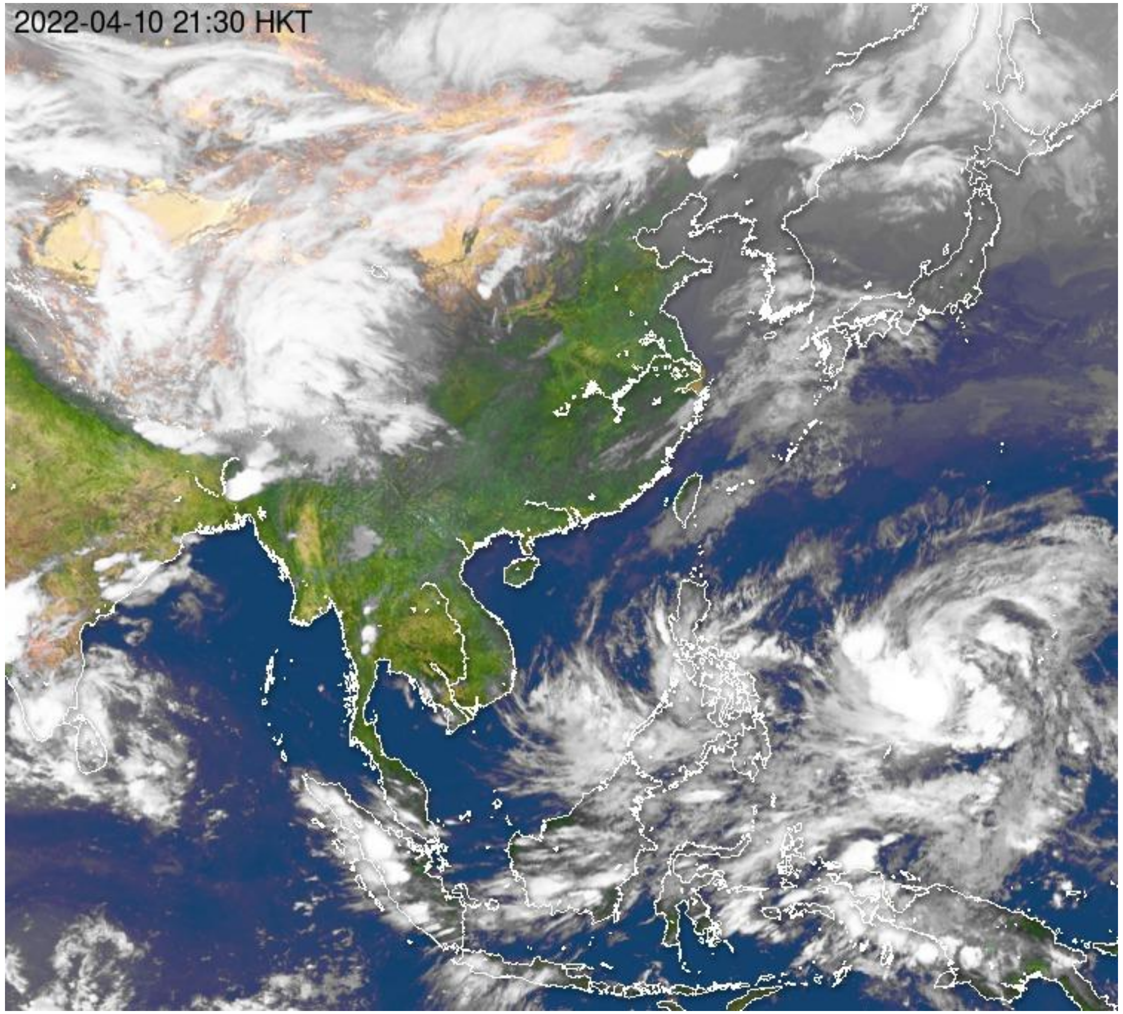
- Both TrajGRU and ResConvLSTM-GAN have higher CSI across all rainfall intensities

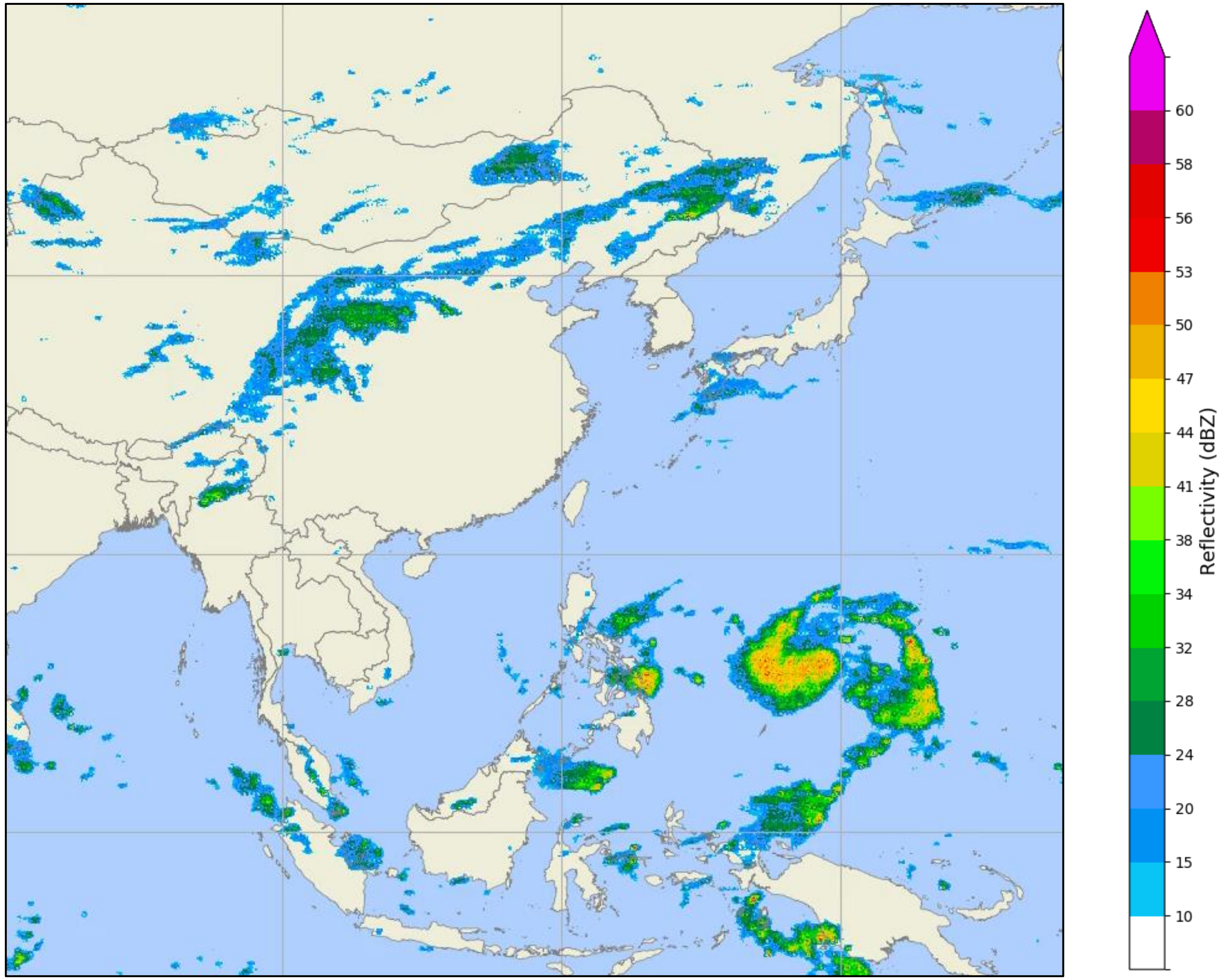


- Further fine-tuning on ResConvLSTM-GAN will be conducted to enhance its performance, especially in heavy rain scenarios

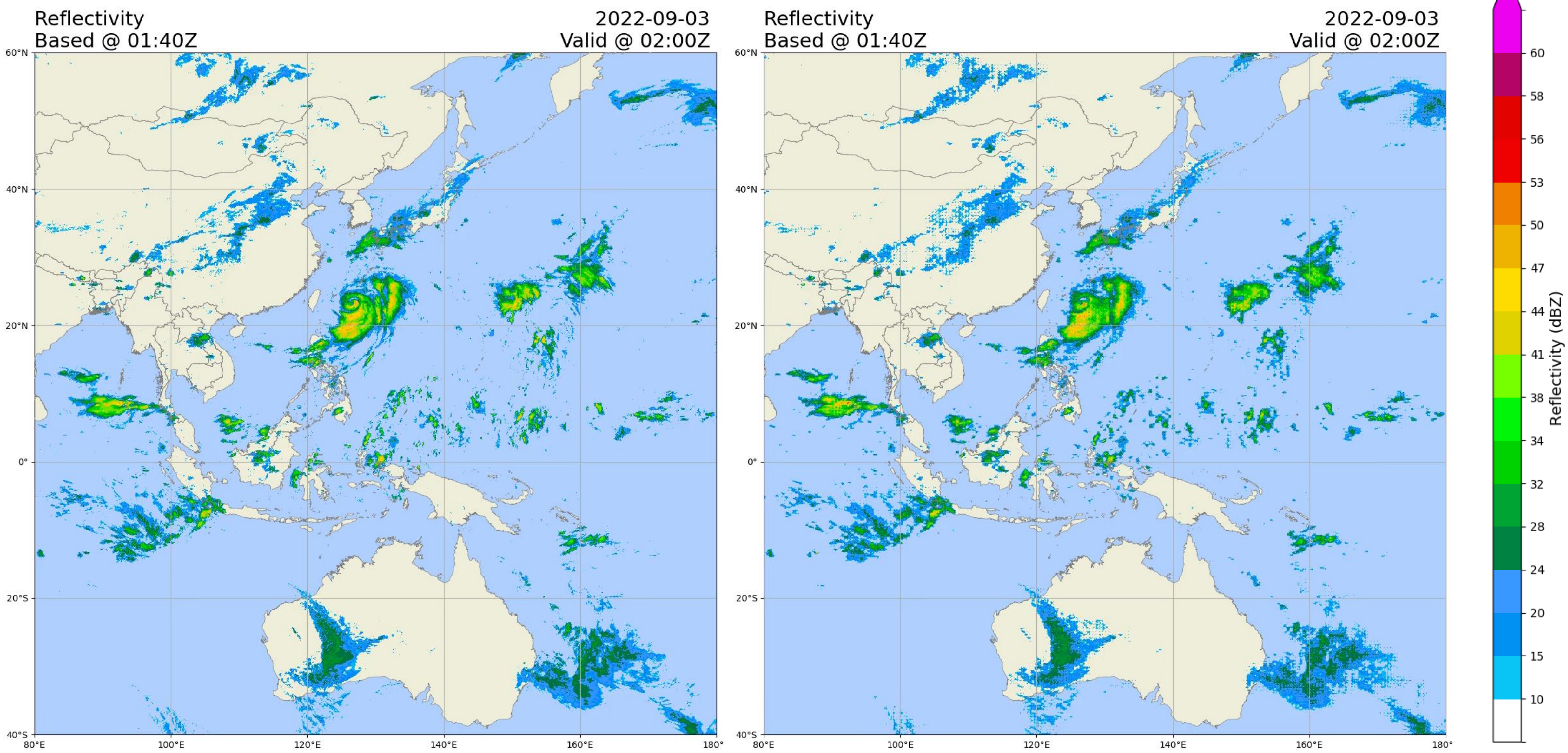
ResConvLSTM-GAN applied to Himawari-8 and GK2A blended imagery

- Development underway to blended satellite with radar and rainfall observation (rain-gauge) / estimate (e.g. GPM) to improve regional QPE / QPF

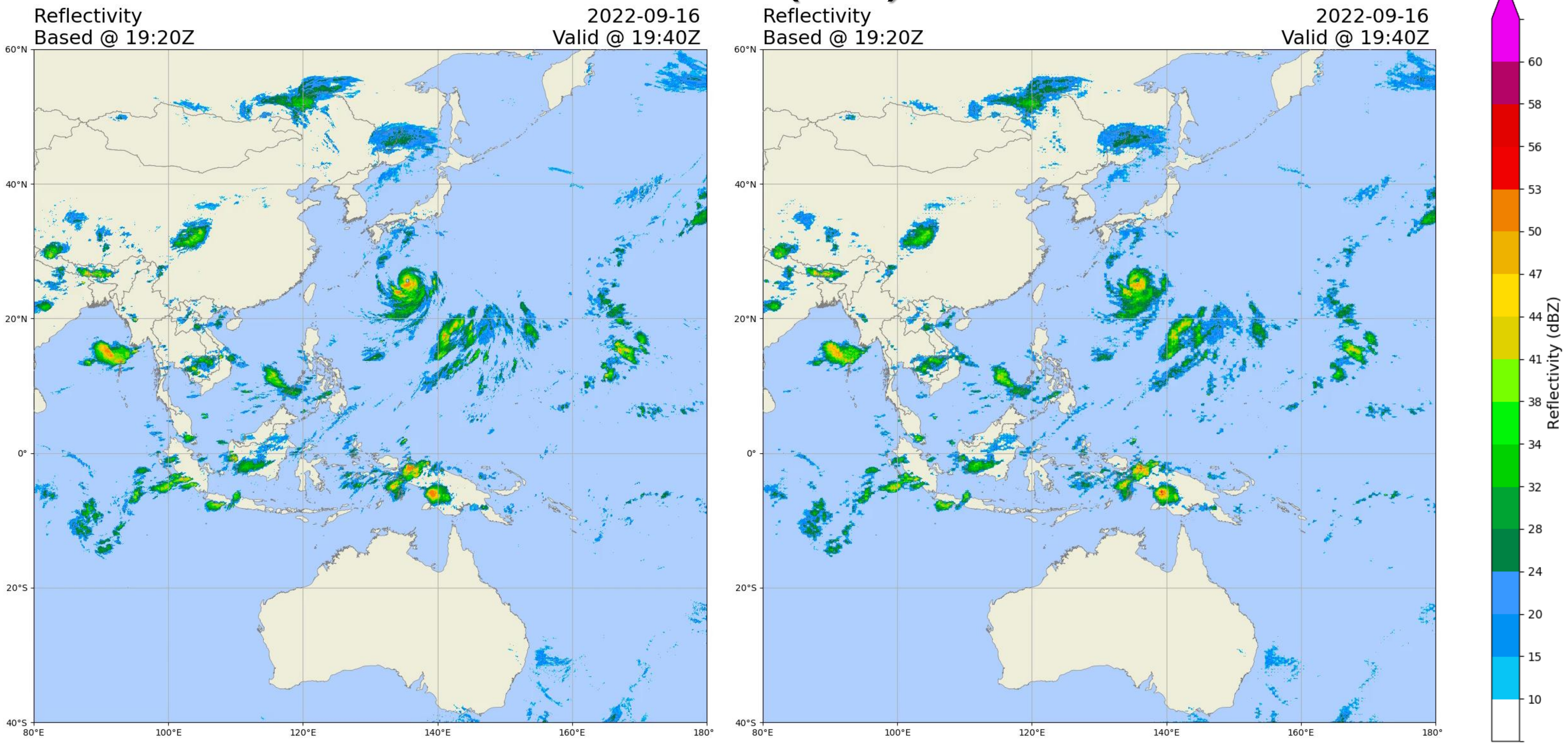




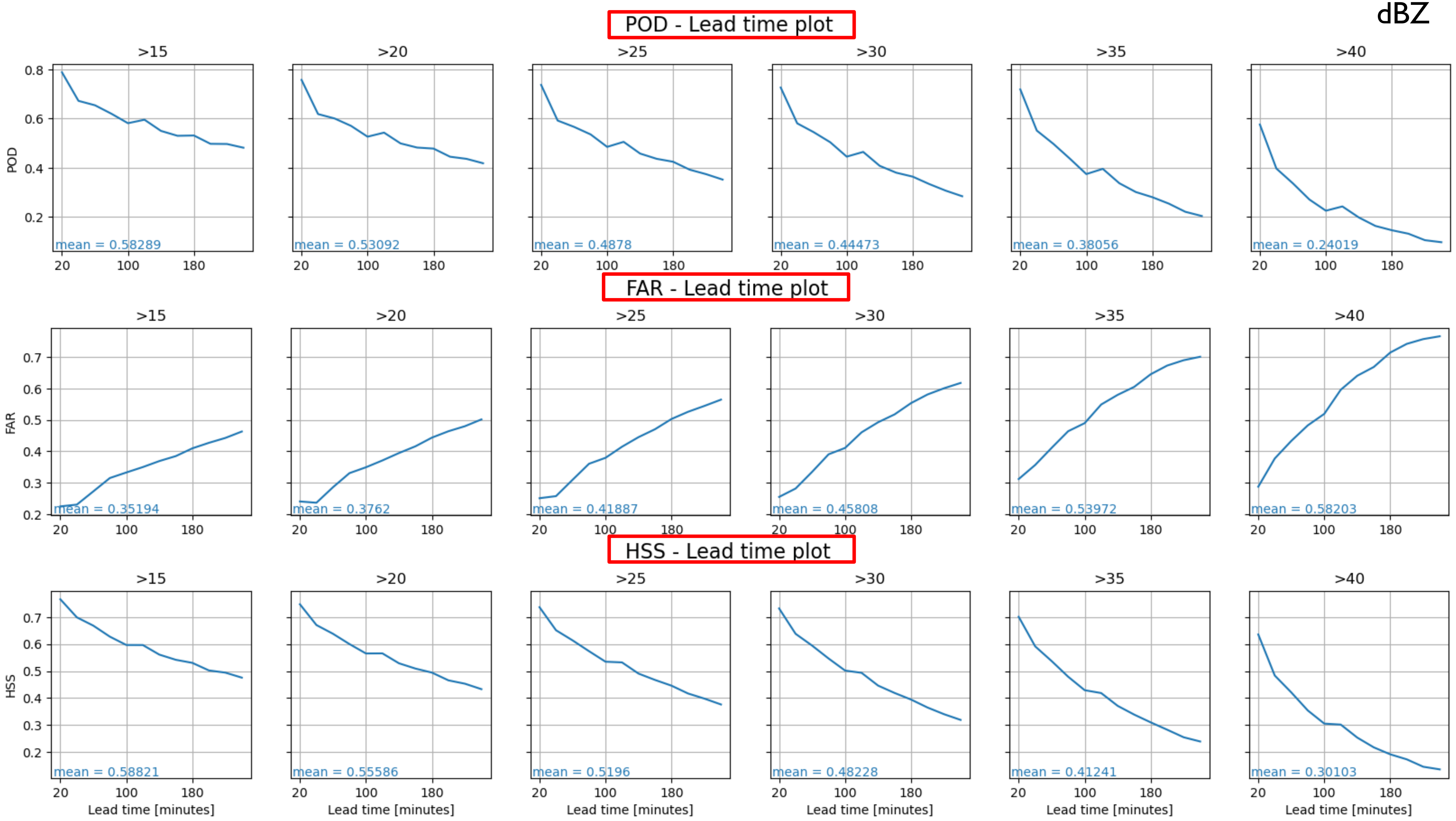
Hinnamnor (2222)



Nanmadol (2224)



dBZ


POD - Lead time plot
>15
>20
>25
>30
>35
>40
POD
mean = 0.58289
mean = 0.53092
mean = 0.4878
mean = 0.44473
mean = 0.38056
mean = 0.24019
FAR - Lead time plot
FAR
mean = 0.35194
mean = 0.3762
mean = 0.41887
mean = 0.45808
mean = 0.53972
mean = 0.58203
HSS - Lead time plot
HSS
mean = 0.58821
mean = 0.55586
mean = 0.5196
mean = 0.48228
mean = 0.41241
mean = 0.30103
Lead time [minutes]

Machine Learning in TC Intensity Forecasting

Using ML in post-processing HWRF model outputs for TC Intensity Prediction

A Feed Forward Neural Network Based on Model Output Statistics for Short-Term Hurricane Intensity Prediction

https://journals.ametsoc.org/doi/pdf/10.1175/WAF-D-18-0173.1

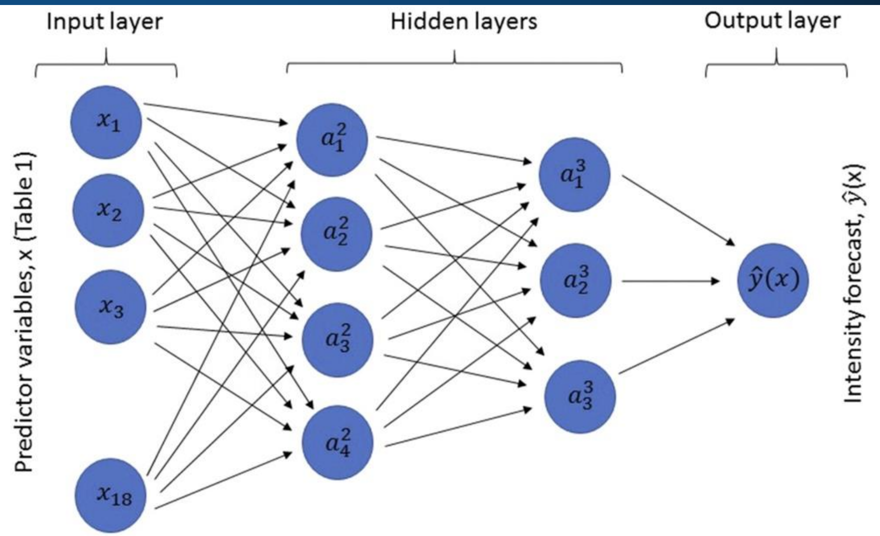
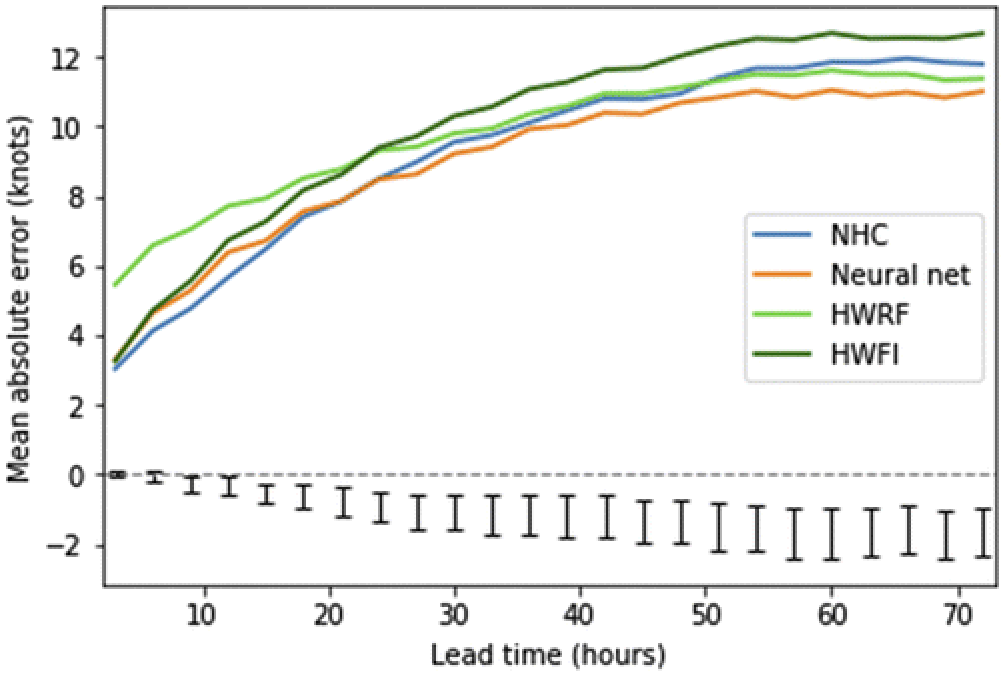


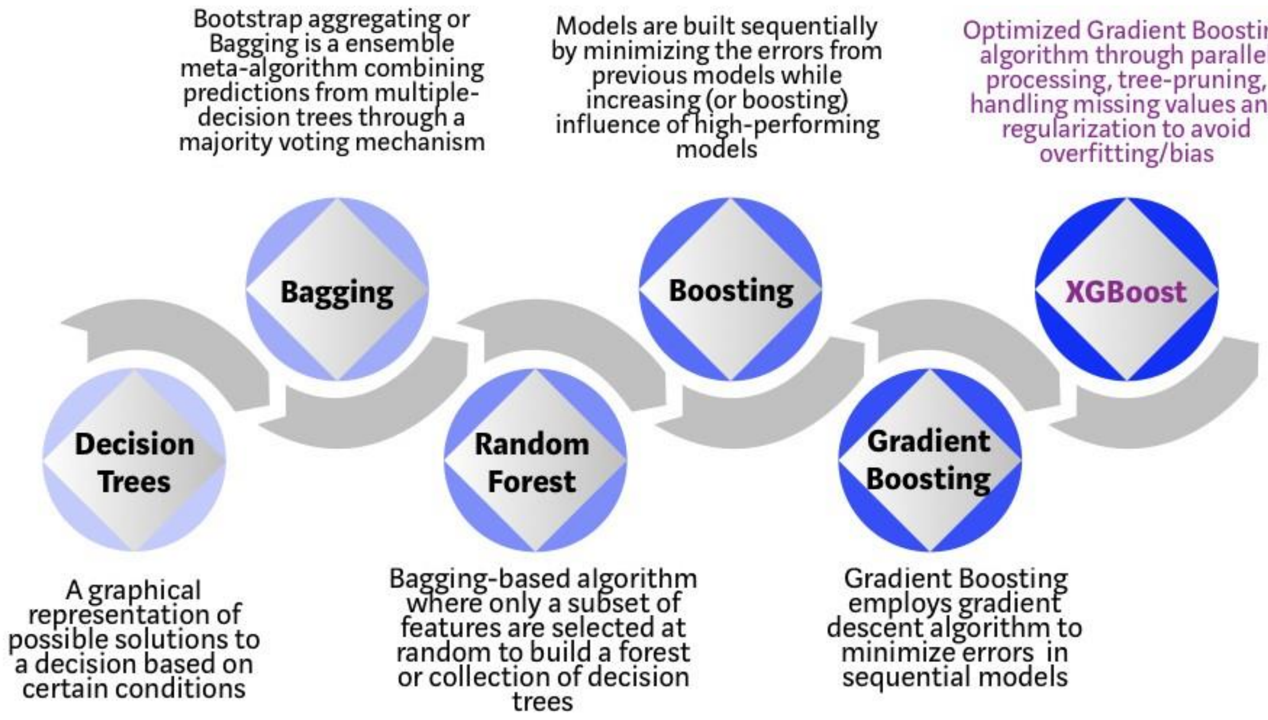
TABLE 1. A complete list of all 18 predictors incorporated into the neural network. All predictors are derived from HWRF output unless stated otherwise. Below, r, p, and T represent radius, pressure, and temperature, respectively. The mean and standard deviation of the predictors are provided in the two rightmost columns as well.

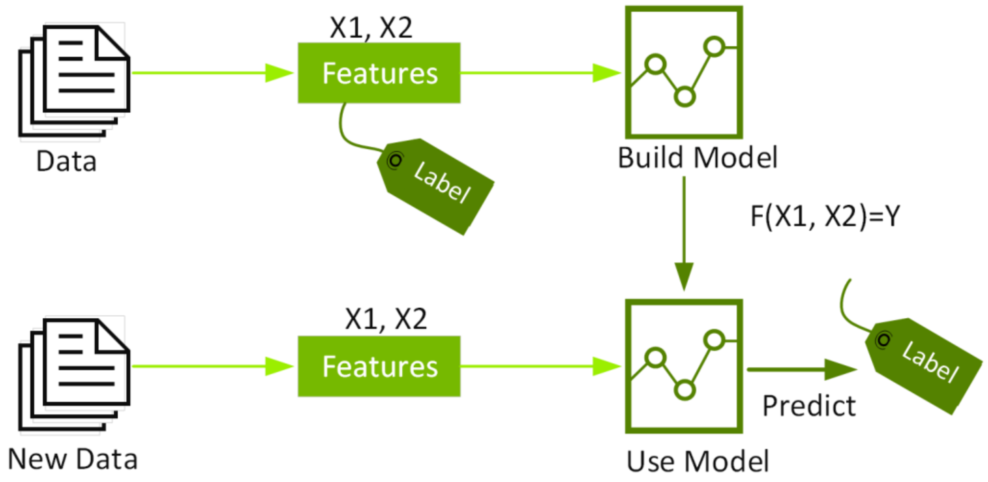
Description of predictor class	No. of predictors for this class	Mean	Std dev
Latitude of storm center (°N)	1	24.0	9.9
Longitude of storm center (°W)	1	107.1	3.9
"Interpolated" maximum 1-min 10-m wind speed (HWFI) (kt)	1	49.7	23.9
Minimum sea level pressure (hPa)	1	991.9	19.9
850–200-hPa vertical wind shear magnitude averaged over $0 \leq r < 500$ km (kt)	1	20.7	13.6
Storm translation speed (kt)	1	11.1	6.7
Sea surface T averaged over $0 \leq r < 50$ km (K)	1	296.7	9.1
Relative humidity ($200 \leq r < 800$ km) averaged over the layer $850 \leq p < 700$ hPa (%)	1	66.7	8.8
Convective available potential energy averaged over $0 \leq r < 100$ or $200 \leq r < 500$ km (J kg^{-1})	2	1132.0/1059.9	836.8/681.9
Surface turbulent sensible heat fluxes averaged over $100 \leq r < 200$ km (W m^{-2})	1	8.6	16.4
Total condensate averaged over $100 \leq r < 250$ km (g kg^{-1})	1	12.3	9.8
Two inertial stability-based parameters averaged over $850 \leq p < 500$ hPa and $0 \leq r < 100$ or $100 \leq r < 250$ km (10^{-6} s^{-2})	2	4.2/0.2	5.2/0.2
Symmetry parameter (as defined in Miyamoto and Takemi 2013) for total condensate and the coupling of inertial stability/vertical motion, over $850 \leq p < 500$ hPa and $0 \leq r < 100$ and $100 \leq r < 250$ km, respectively (%)	2	31.9/48.8	14.9/21.8
Operational estimate of the maximum 1-min 10-m wind speed at the initial time (kt)	1	57.6	28.8
A binary indicator specifying whether a storm is in the Atlantic or eastern Pacific basin	1	—	—



XGBoost

- XGBoost is a <u>supervised</u>, <u>decision-tree</u>-based <u>ensemble learning algorithm</u> based on gradient boosting framework for <u>regression</u> and <u>classification</u>





Source: Nvidia

Calibrating tropical cyclone intensity forecast of ECMWF EPS using XGBoost

Machine learning in calibrating tropical cyclone intensity forecast of ECMWF EPS
https://rmets.onlinelibrary.wiley.com/doi/full/10.1002/met.2041

- Predictors
  - Persistence
  - Current Intensity
  - TCHP
  - 200 hPa divergence
  - 500 – 300 hPa average RH
  - 850 – 200 hPa vertical wind shear
  - Selected percentiles of TC maximum wind and minimum pressure from EPS members
  - Forecast hours
  - Latitude and longitude of ensemble mean

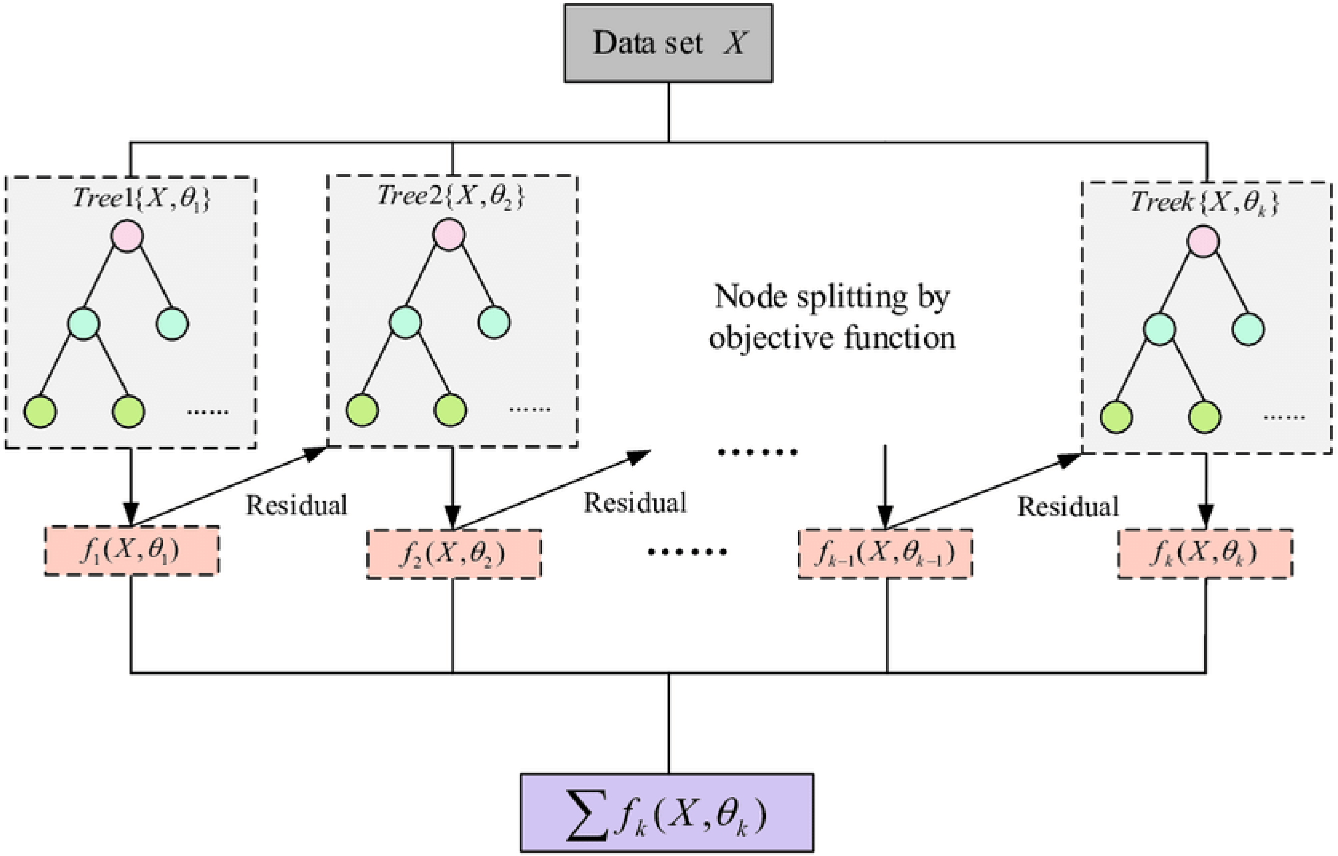
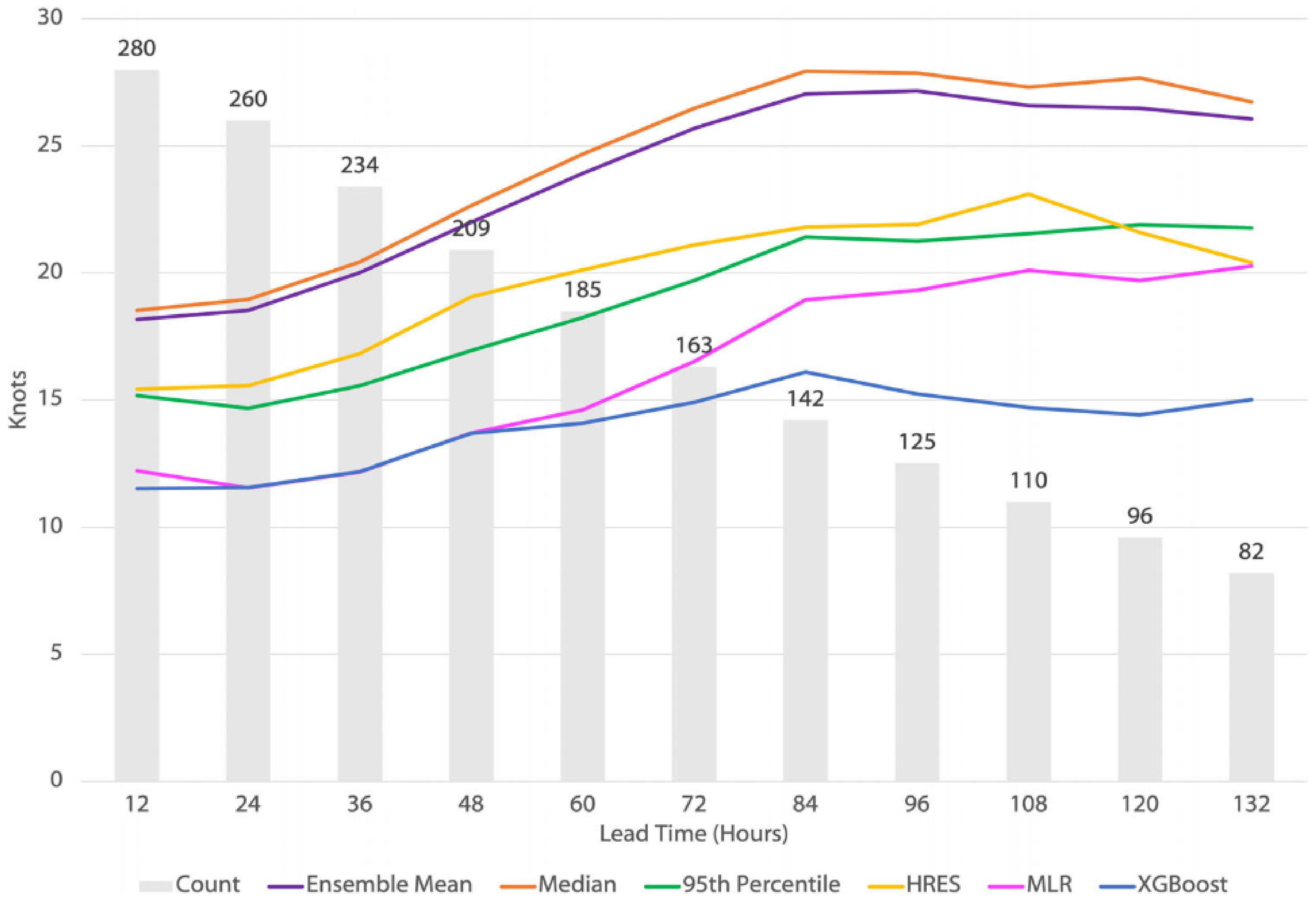
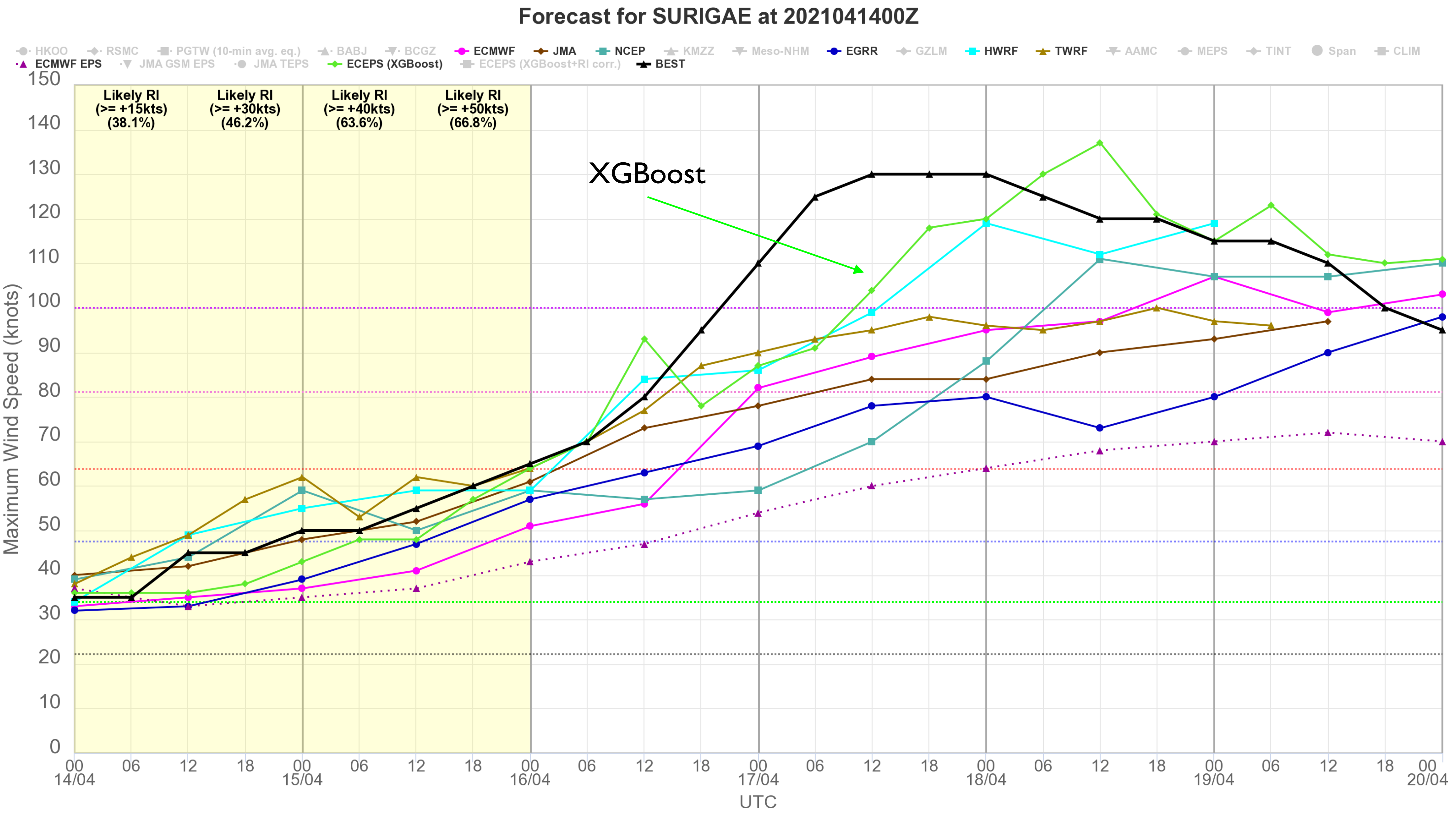
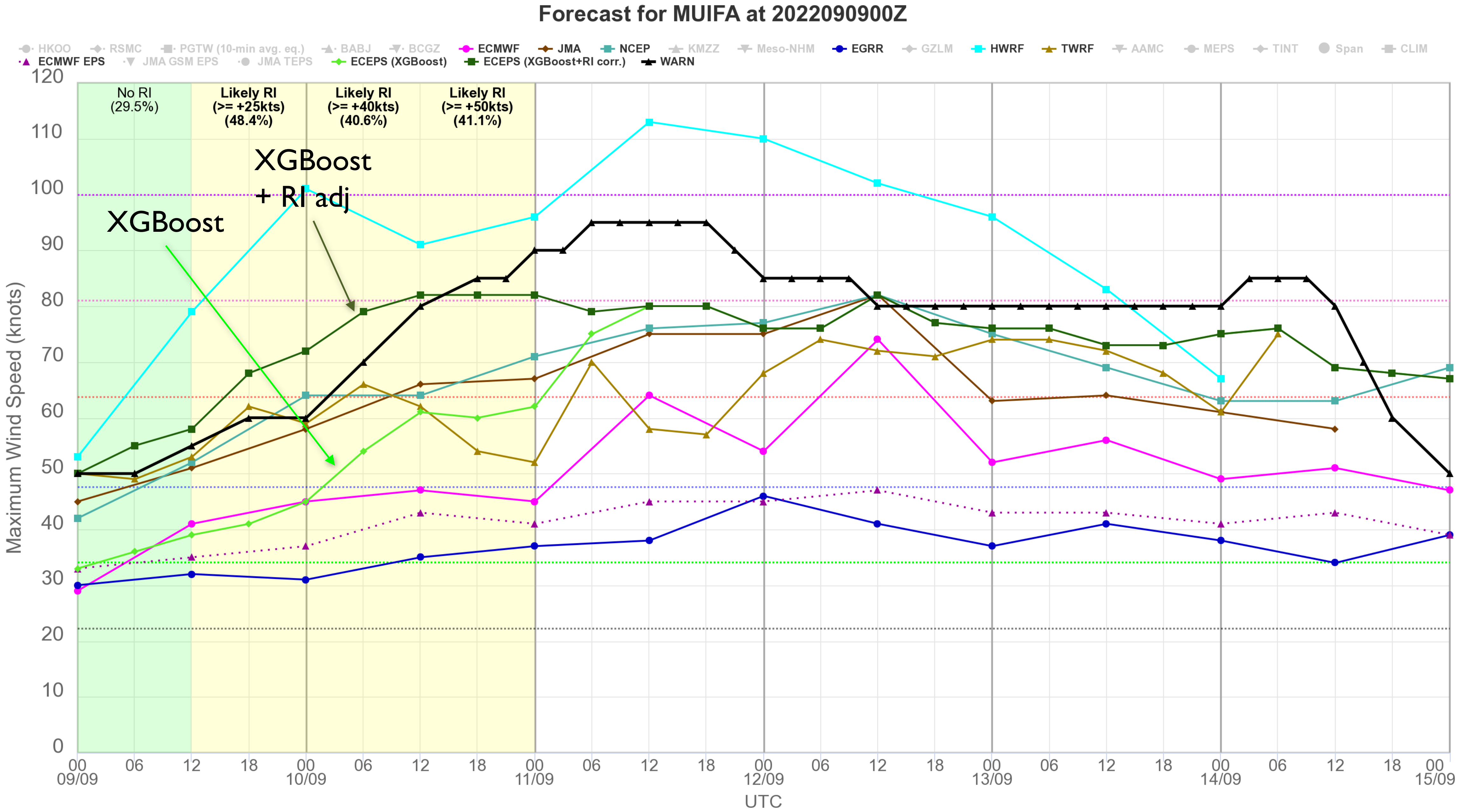


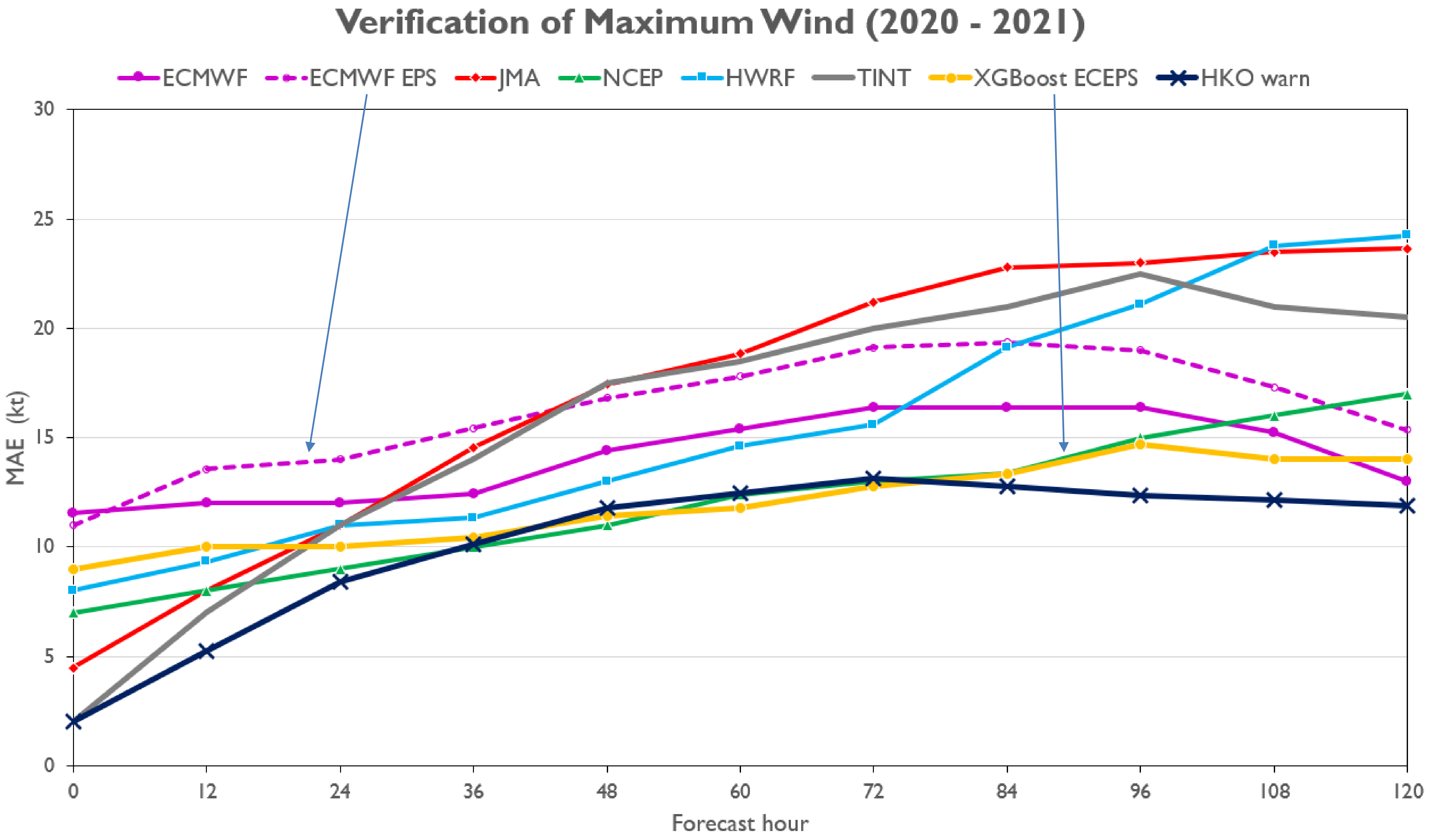
FIGURE 4 Forecast RMSE of maximum wind by XGBoost model (blue), ECMWF EPS 95th percentile (green), median (orange) and mean (purple) members, ECMWF HRES forecast (yellow) and the benchmark MLR model (magenta). Note that for weaker EPS percentile members the number of cases may be smaller since these members may have forecast the TC to dissipate



Ref: Machine learning in calibrating tropical cyclone intensity forecast of ECMWF EPS
https://rmets.onlinelibrary.wiley.com/doi/full/10.1002/met.2041


Forecast for SURIGAE at 2021041400Z
HKOO
RSMC
PGTW (10-min avg. eq.)
BABJ
BCGZ
ECMWF
JMA
NCEP
KMZZ
Meso-NHM
EGRR
GZLM
HWRF
TWRF
AAMC
MEPS
TINT
Span
CLIM
ECMWF EPS
JMA GSM EPS
JMA TEPS
ECEPS (XGBoost)
ECEPS (XGBoost+RI corr.)
BEST
Likely RI (>= +15kts) (38.1%)
Likely RI (>= +30kts) (46.2%)
Likely RI (>= +40kts) (63.6%)
Likely RI (>= +50kts) (66.8%)
XGBoost
Maximum Wind Speed (knots)
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
00 14/04
06
12
18
00 15/04
06
12
18
00 16/04
06
12
18
00 17/04
06
12
18
00 18/04
06
12
18
00 19/04
06
12
18
00 20/04
UTC


Forecast for MUIFA at 2022090900Z
HKOO
RSMC
PGTW (10-min avg. eq.)
BABJ
BCGZ
ECMWF
JMA
NCEP
KMZZ
Meso-NHM
EGRR
GZLM
HWRF
TWRF
AAMC
MEPS
TINT
Span
CLIM
ECMWF EPS
JMA GSM EPS
JMA TEPS
ECEPS (XGBoost)
ECEPS (XGBoost+RI corr.)
WARN
No RI (29.5%)
Likely RI (>= +25kts) (48.4%)
Likely RI (>= +40kts) (40.6%)
Likely RI (>= +50kts) (41.1%)
XGBoost + RI adj
XGBoost
Maximum Wind Speed (knots)
120
110
100
90
80
70
60
50
40
30
20
10
0
00 09/09
06
12
18
00 10/09
06
12
18
00 11/09
06
12
18
00 12/09
06
12
18
00 13/09
06
12
18
00 14/09
06
12
18
00 15/09
UTC


Verification of Maximum Wind (2020 - 2021)
ECMWF
ECMWF EPS
JMA
NCEP
HWRF
TINT
XGBoost ECEPS
HKO warn
MAE (kt)
0
5
10
15
20
25
30
0
12
24
36
48
60
72
84
96
108
120
Forecast hour

Next step on ML

- Physics-Informed Machine Learning (PIML) / Physics-Guided Neural Network (PGNN)
  - Apply physical knowledge to inform data prediction capability of ML models

- Enhanced nowcast ML with explicit dynamics constraint

$$\frac{dq}{dt} := \frac{\partial q}{\partial t} + \mathbf{u} \cdot \nabla q = S$$

$$\mathrm{loss}_{\mathrm{PHY}}(\hat{Y}, u, v, \tilde{S}) := \left(\frac{\partial \hat{Y}}{\partial t} + u\frac{\partial \hat{Y}}{\partial x} + v\frac{\partial \hat{Y}}{\partial y} - \tilde{S} \right)^2$$

- Additional loss function terms in minimization process of ML model (e.g.TrajGRU):

$$\lambda_{\mathrm{FSS}}\mathrm{loss}_{\mathrm{FSS}}(\hat{Y}, Y) + \lambda_{\mathrm{PHY}}\mathrm{loss}_{\mathrm{PHY}}(\hat{Y}, u, v, \tilde{S})$$

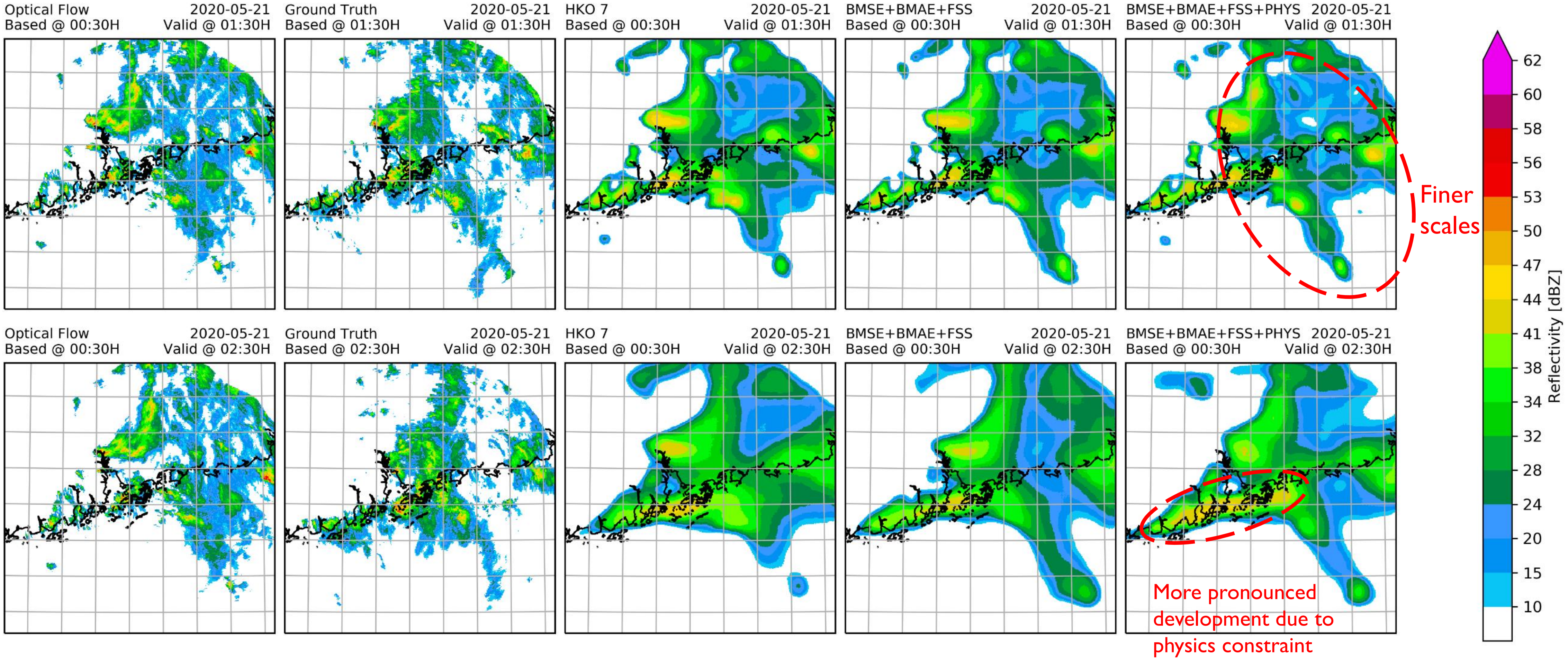
account for spatial and intensity errors

dynamical constraint

Optical flow extrapolation | Actual | TrajGRU | TrajGRU with spatio-intensity error constraint | TrajGRU with physics and spatio-intensity error constraint

1-hour nowcast

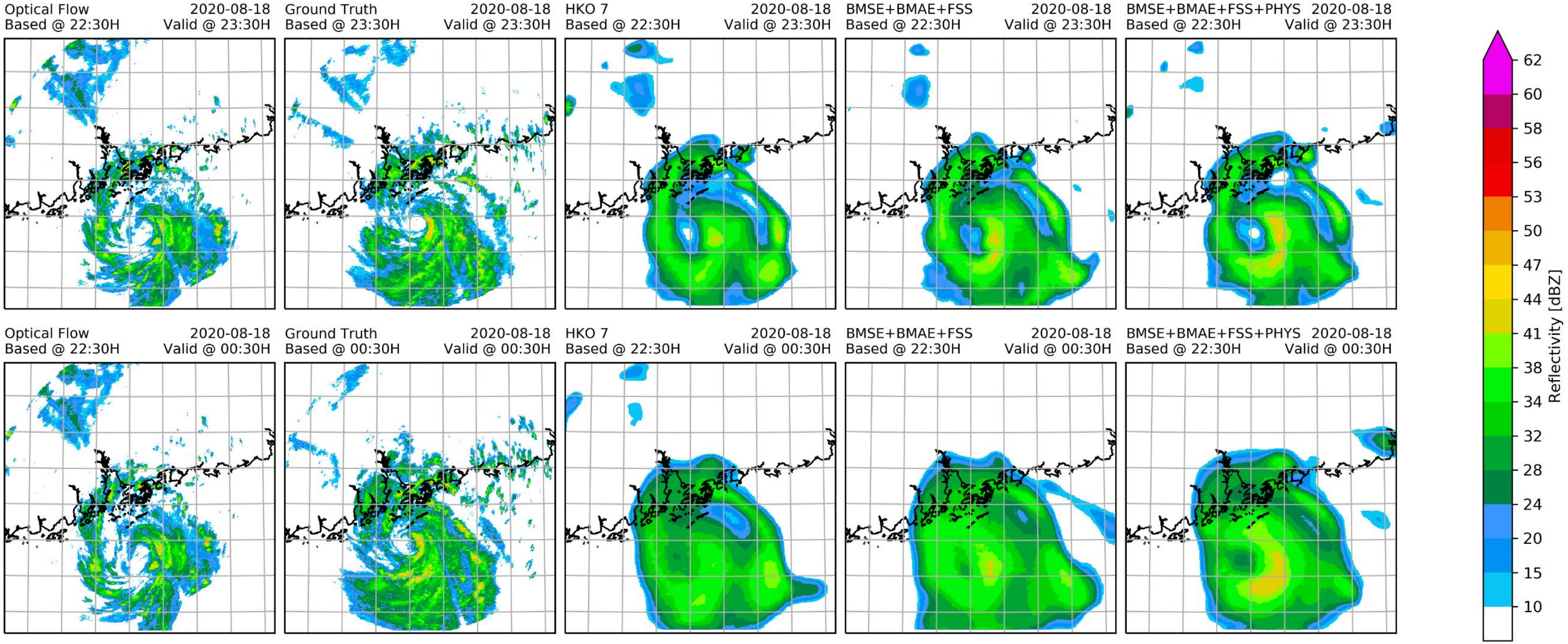
2-hour nowcast



Optical flow extrapolation | Actual | TrajGRU | TrajGRU with spatio-intensity error constraint | TrajGRU with physics and spatio-intensity error constraint

1-hour nowcast

2-hour nowcast



Enhanced asymmetry and intensity

Concluding remarks

- Machine learning (and deep learning) and intuition from principles of statistical physics
- Machine learning and deep learning have been evolving rapidly since development of rainfall nowcasting application
  - Powerful methods, yet rooms for improvement, in generating realistic "video" sequence of rainfall (and other meteorological patterns)
  - Choices of methods are more than science (and demanding efforts for trial and fine-tuning) – supervised vs unsupervised; convolutional, autoencoder, generative adversarial network, …
- More important to advance AI / ML in coming years:
  - clean data and how extreme / anomaly could be effectively detected or predicted
  - knowledge (from forecasters) and physics (explicit / parameterized / from full Earth system models)
  - data and computational scientists
  - applications in other scales of forecasts, re-analysis / hindcast and seamless prediction
- Training and collaborative development opportunities on AI / ML

香港天文台
HONG KONG OBSERVATORY

Thank you very much

Q&A