Digital Typhoon Dataset: 40+ years of satellite images of tropical cyclones for machine learning research

Asanobu Kitamoto *1*2, Jared Hwang *3*1, Bastien Vuillod *4*1, Lucas Gautier *5*1

*1 National Institute of Informatics

*2 **Typhoon Science and Technology Research Center (TRC), Yokohama National University**

*3 University of Southern California

*4 Grenoble-INP, Ensimag

*5 Université Clermont Auvergne, ISIMA

Self introduction (Asanobu KITAMOTO)

https://researchmap.jp/kitamoto/

@kitamotoasanobu

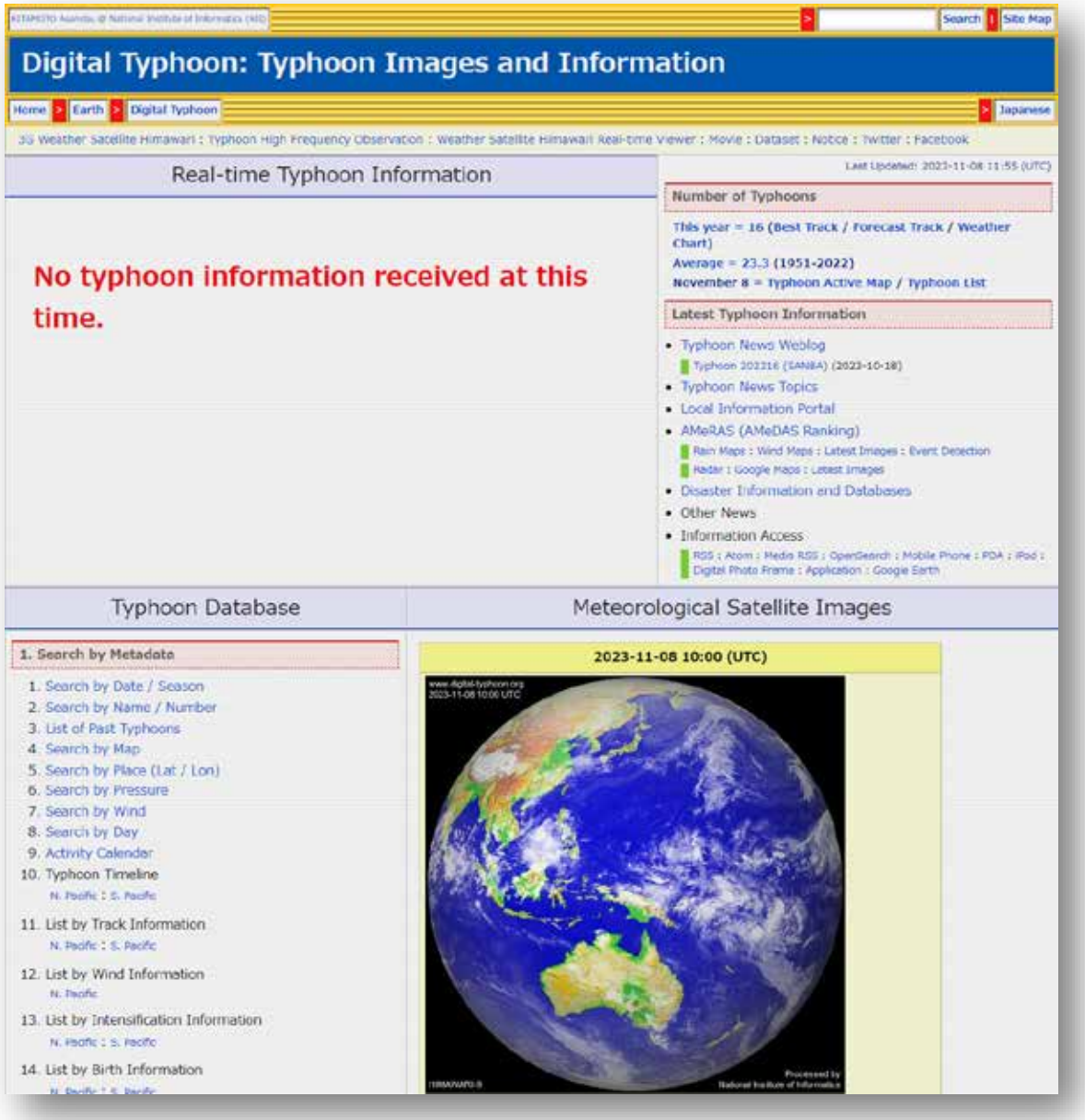
Certified weather forecaster
No. 5439

- Professor, National Institute of Informatics
- Typhoon Data Science Lab @ TRC-YNU
- **Background**: Informatics and Computer science
- **Research Area**:
  - Digital humanities
  - Data-driven science for earth environment (especially meteorology and disasters)
  - Open science

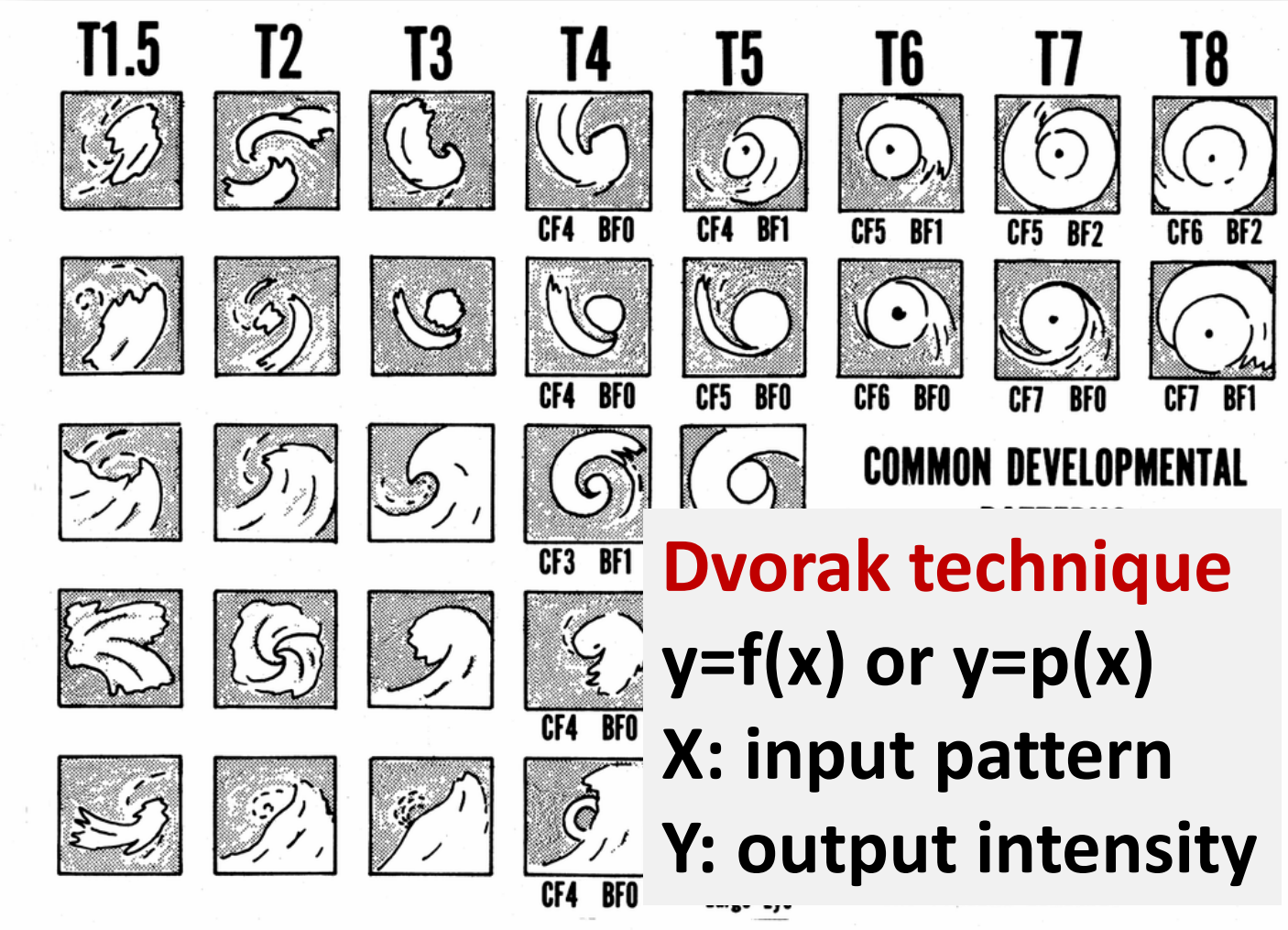
Digital Typhoon

http://agora.ex.nii.ac.jp/digital-typhoon/

- The project started in **1999**
- The website started in **2003**
- Organize the world's information about typhoon-related data and make it universally accessible and useful
- **Concept: seamlessly integrate current and past data**
- **20 million page views per year**



Data-Driven Typhoon Analysis



Asanobu KITAMOTO, "'Digital Typhoon' -- Typhoon Analysis based on Artificial Intelligence Approach“, IPSJ Tech Report, Vol. CVIM123-8, pp. 59-66, Sep **2000.**

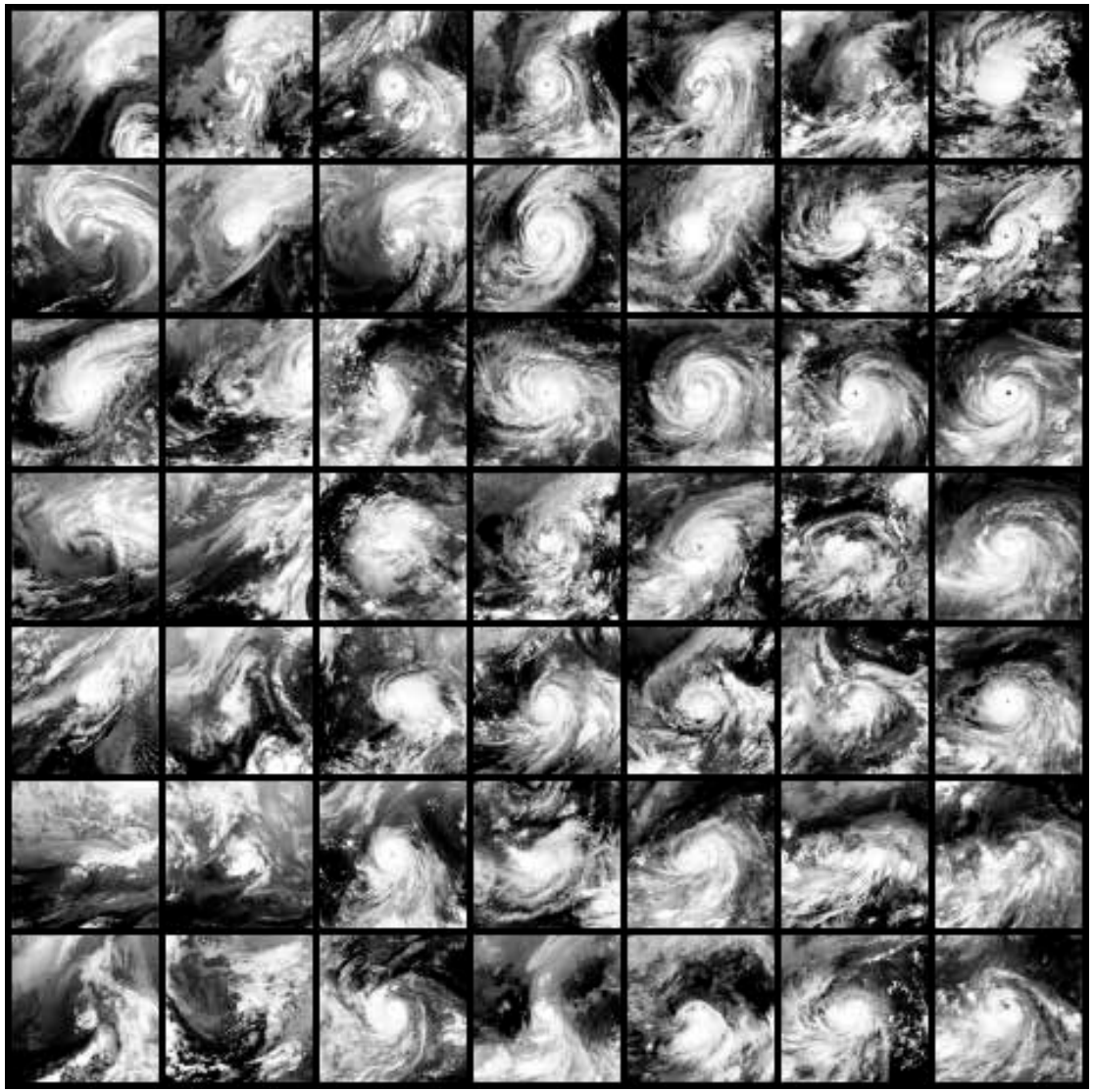
Asanobu KITAMOTO, "Interpretation of Typhoon Cloud Patterns by Holistic Analysis“, IEICE Tech Report, Vol. PRMU2000-240, pp. 129-136, Mar **2001.**

Multimodal Datasets

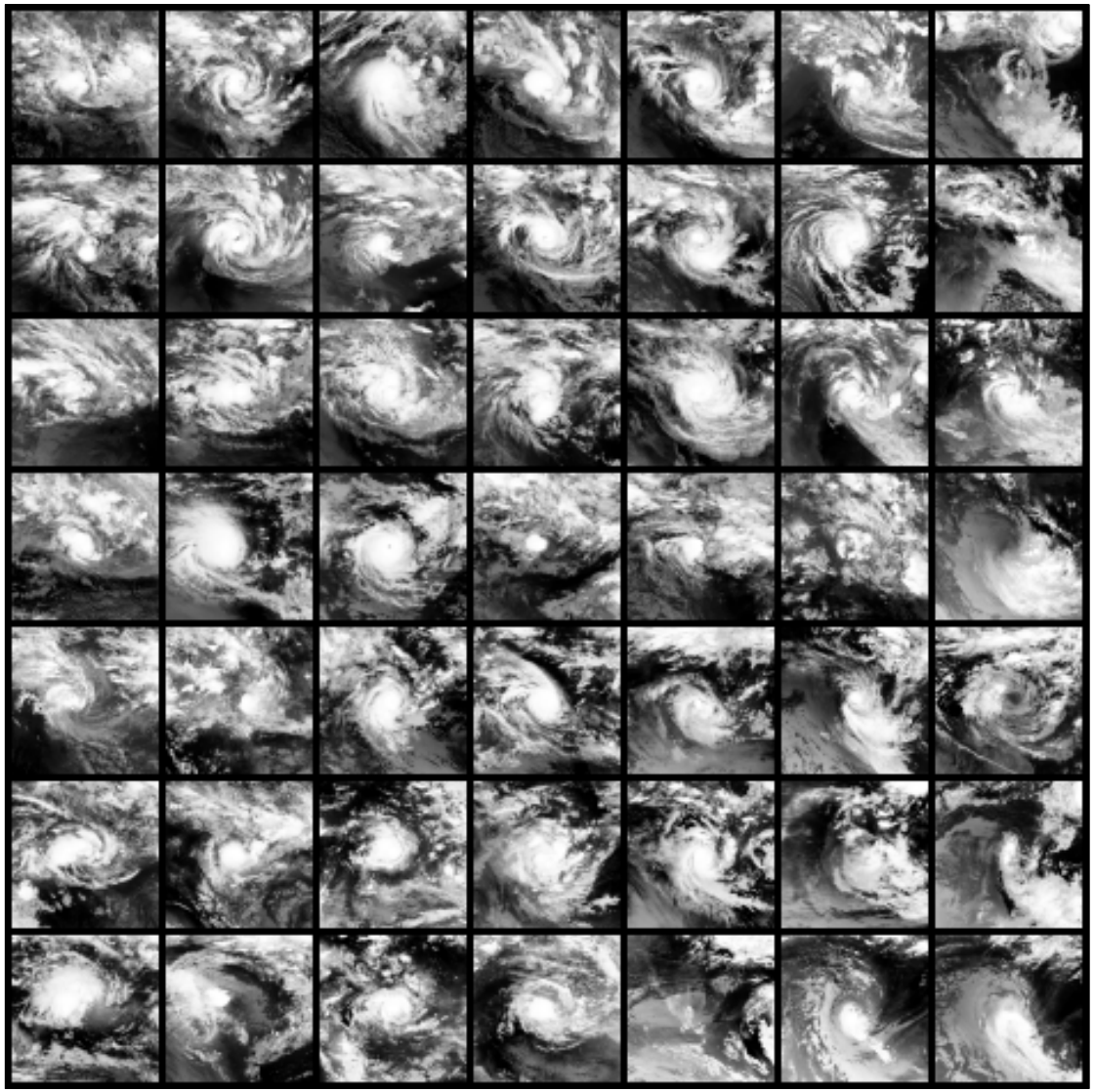
http://agora.ex.nii.ac.jp/digital-typhoon/summary/wnp/s/201919.html.en

1. Satellite images
2. Tracking chart
3. Pressure chart
4. Ground observations
5. Forecast history
6. Weather warnings
7. Weather charts
8. Mass media news
9. Social media text

Visual data
Numerical data
Grid data
Textual data
Structured data



NH: **Typhoon 192,548** images

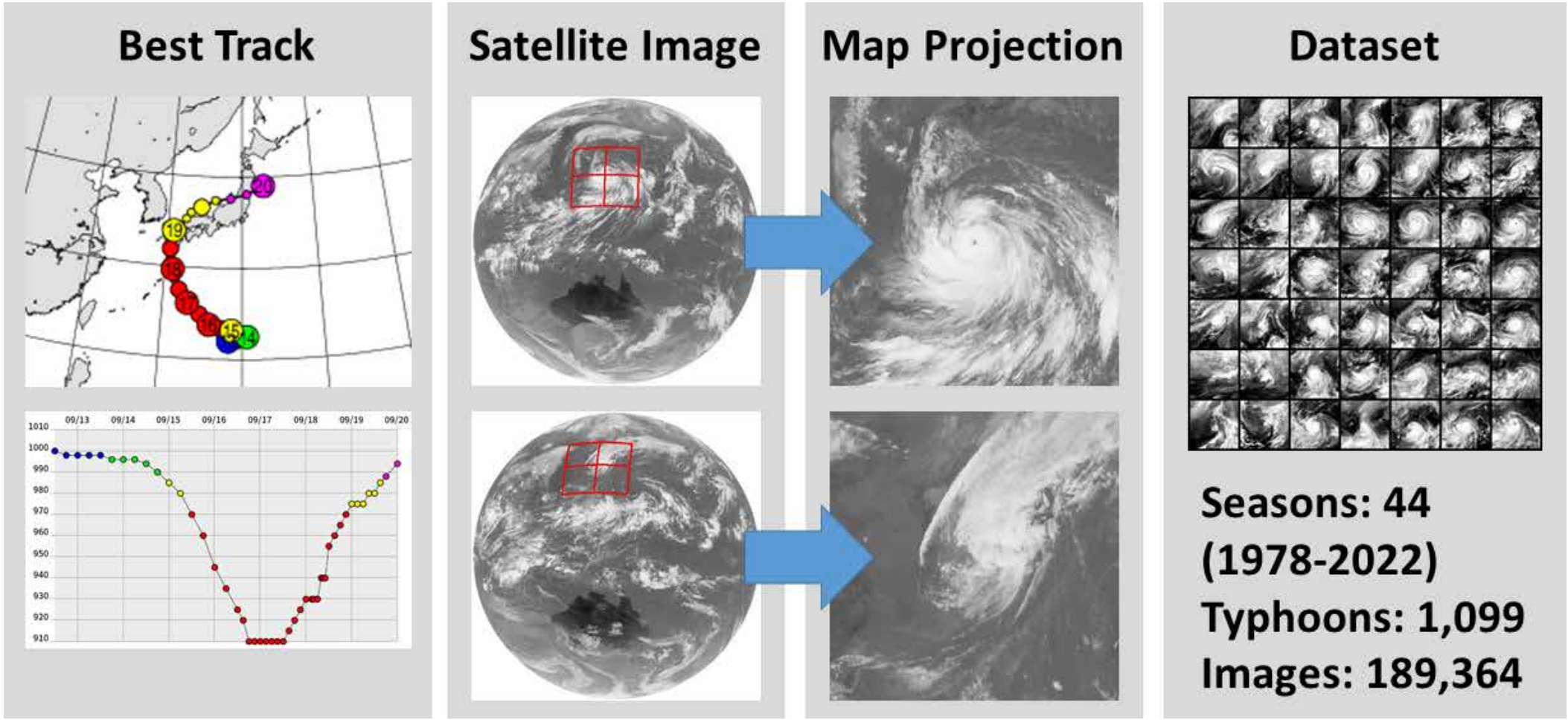


SH: **Cyclone 68,515** images

Digital Typhoon Dataset

http://agora.ex.nii.ac.jp/digital-typhoon/dataset/

Released on November 7, 2023



Details of the Dataset

- **Basin**: Western North Pacific
- **Season**: 1978-2022 (1978, 1979, 1980 are not complete) = the **longest** typhoon image dataset
- **Images**: 189,364
- **Typhoons**: 1,099
- **Satellites**: Himawari1-9+GOES9
- **Channel**: Infrared (IR1)
- **Best track:** JMA
- **Dataset Size**: 54GB

- **Image size**: 512x512 pixels
- **Scene size**: 2500x2500 km
- **Spatial resolution**: 4.88 km
- **Temporal resolution**: 1 hour
- **Map projection**: Lambert azimuthal equal-area projection
- **Sensor calibration**: JMA recalibration method is applied to brightness temperature.

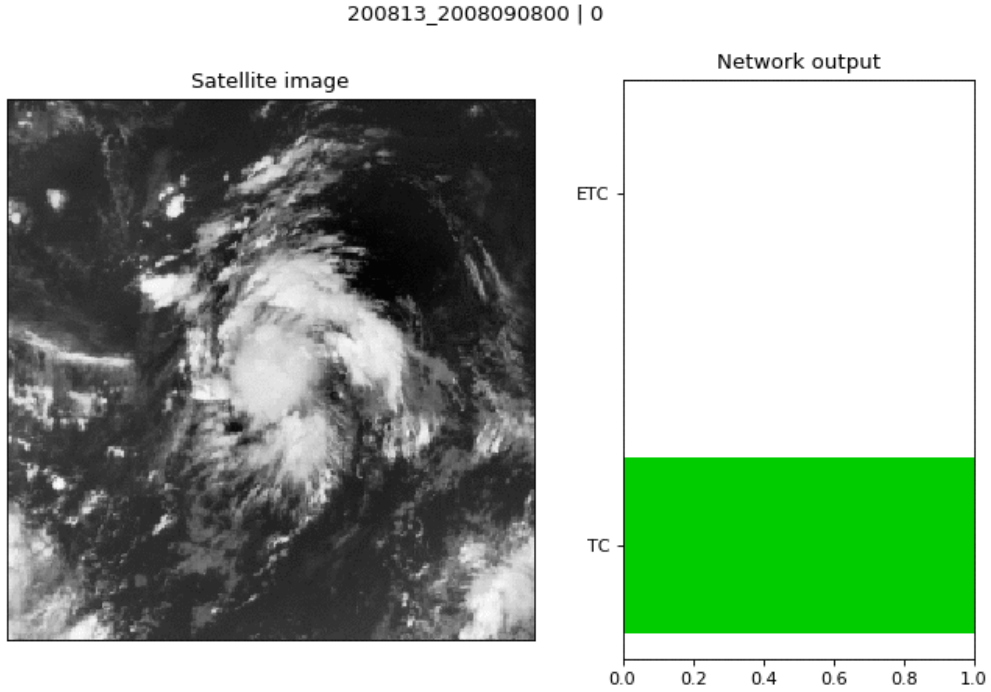
Features of the Dataset

1. Machine learning dataset comes with annotated training data, typical tasks, and performance metric
2. It comes with software library to manage the dataset (**pyphoon2**)
3. It is distributed with an open license (CC BY) so that anyone can use it freely for their own research
4. It offers the common ground for performance comparison of machine learning algorithms

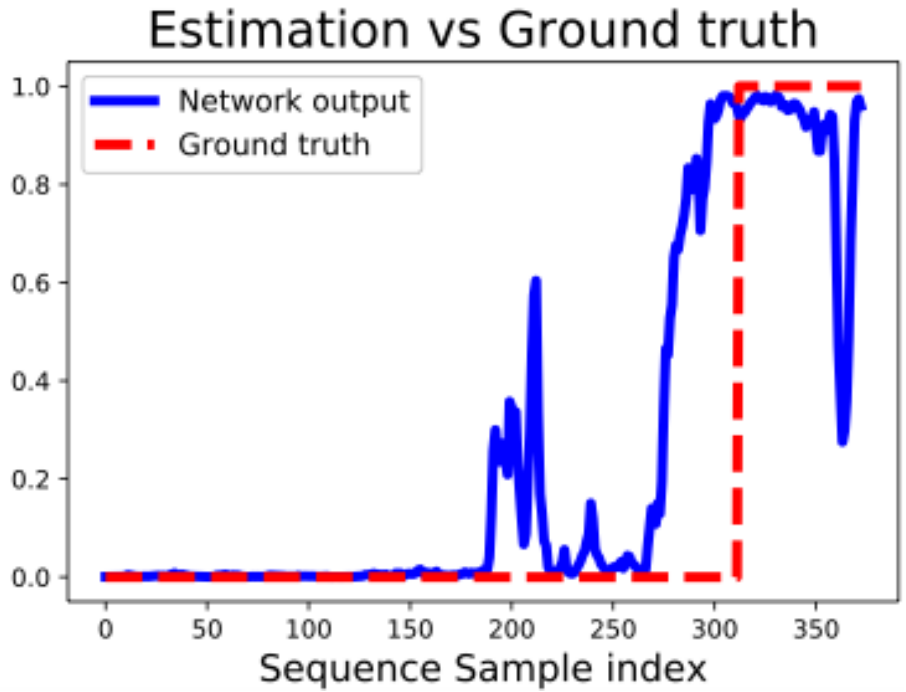
Targets of Inference

1. **Intensity**
   - Inferences on the strength and size of a tropical cyclone
2. **Track**
   - Inferences on the geographical location of a tropical cyclone
3. **Formation**
   - Inferences on the birth of a tropical cyclone
4. **Transition**
   - Inferences on the transition from a tropical cyclone to an extra-tropical cyclone

Transition Task

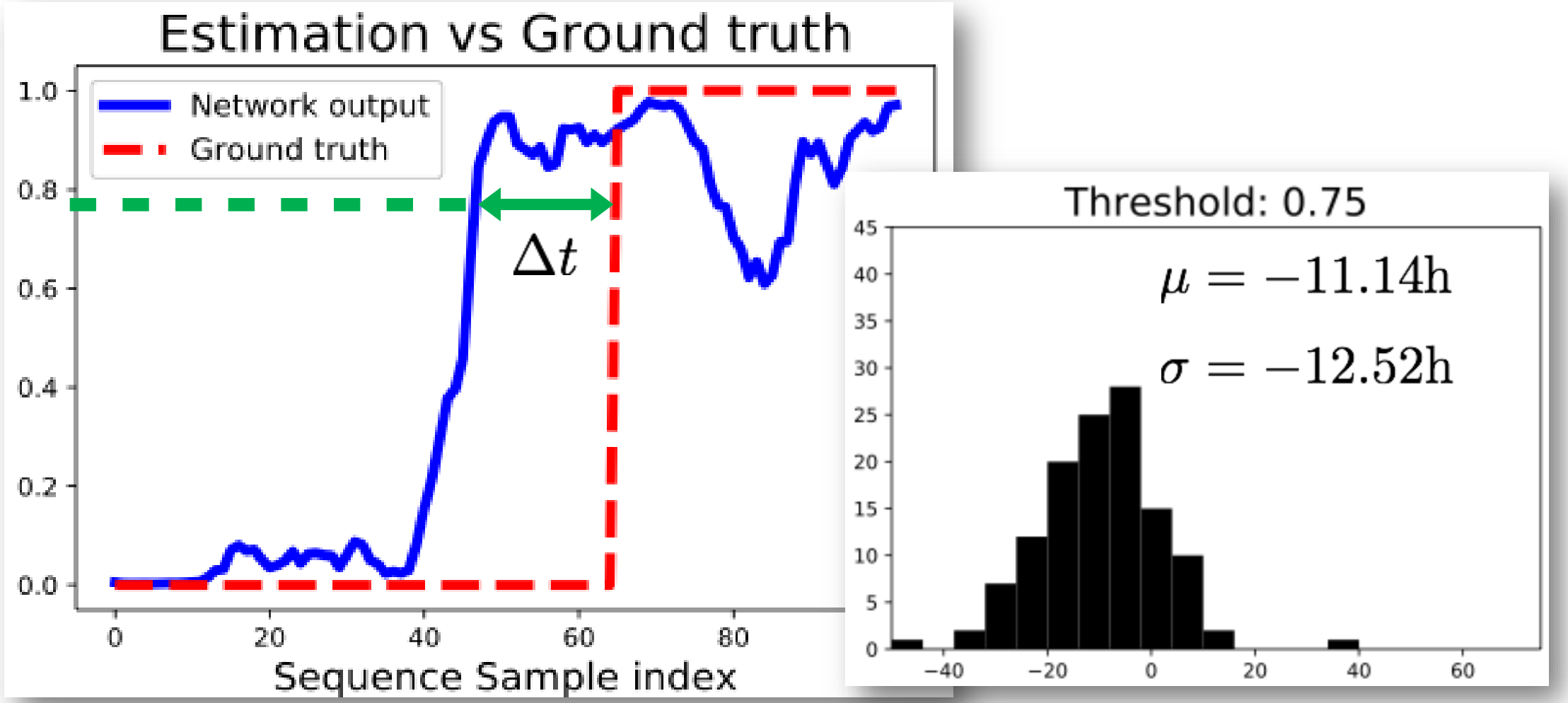


For Typhoon 200813, we modeled the transition from 0 (tropical cyclone) to 1 (extra-tropical cyclone)



Asanobu KITAMOTO, "Analysis of Typhoon Cloud Patterns and Transition to Extra Tropical Cyclones Using Machine Learning", 2018 Autumn Meeting of The Meteorological Society of Japan, 2018

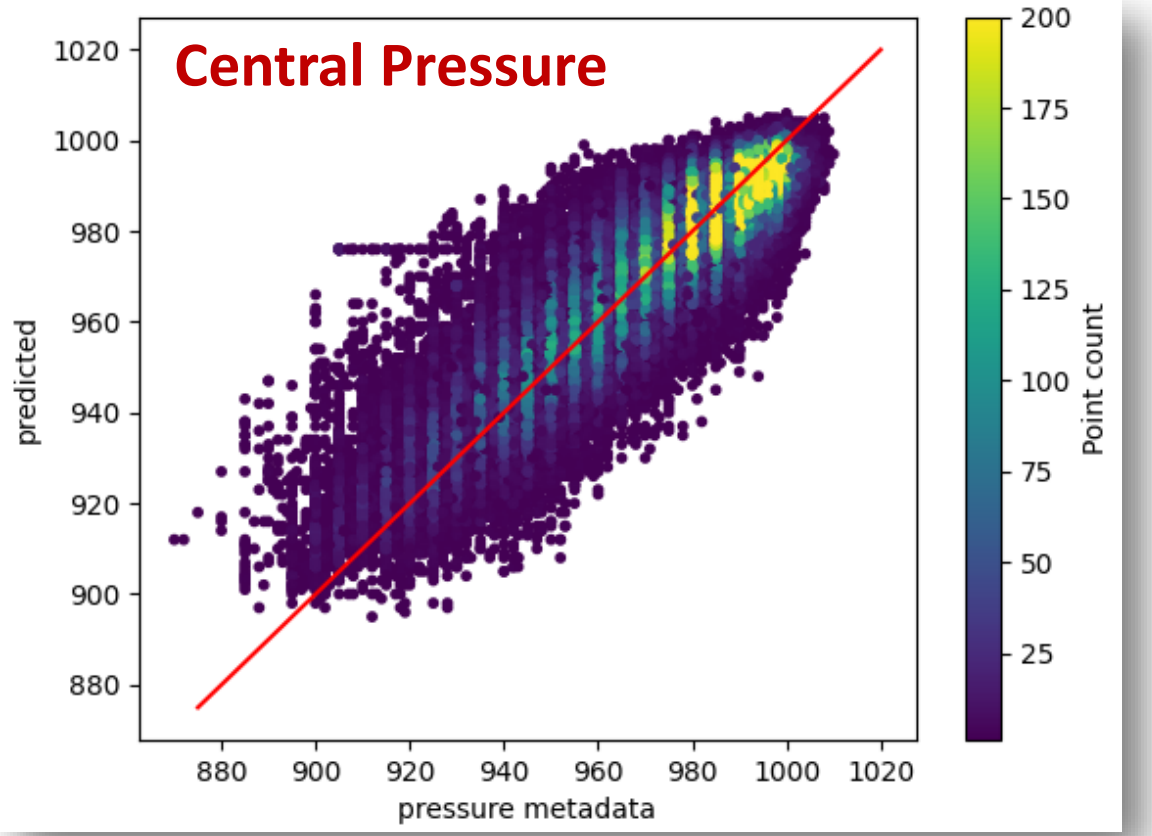
Time Difference from the Best Track



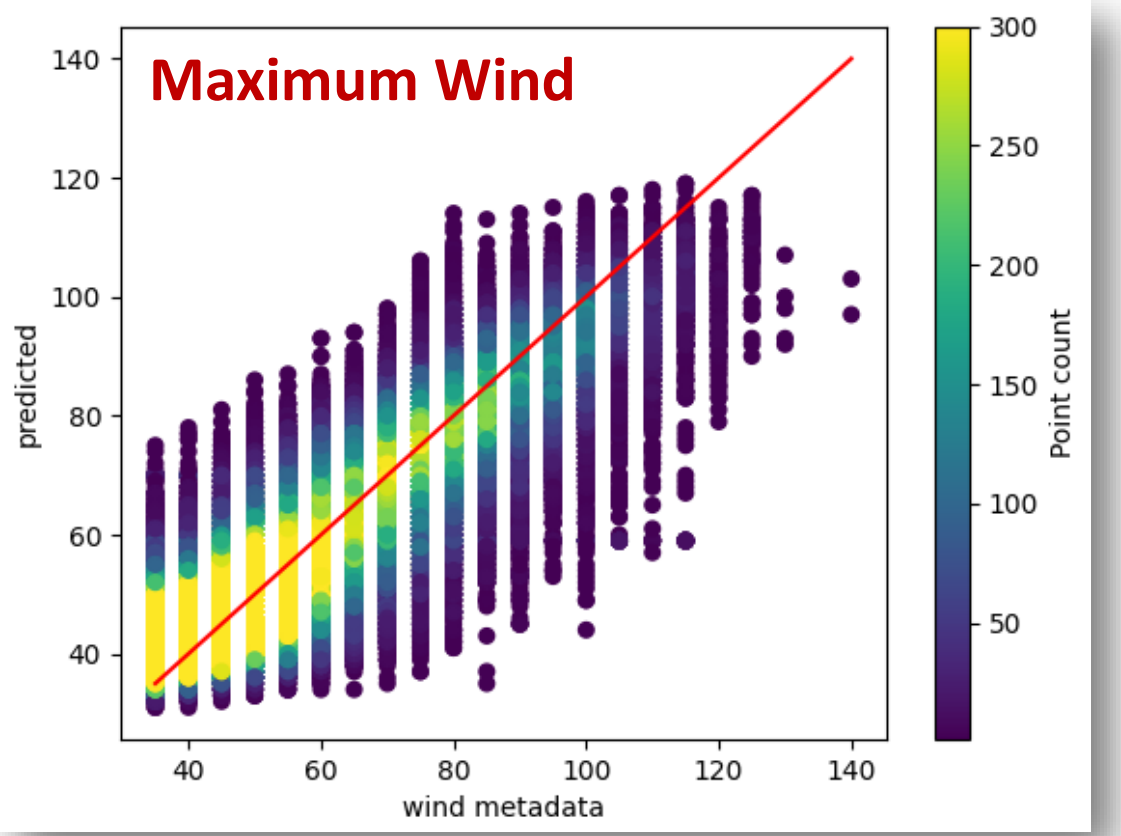
Types of Inference

1. **Analysis:** estimate current values using the current and past data
   - Supervised learning: classification tasks and regression tasks
   - Unsupervised learning: clustering, visualization, etc.
2. **Forecasting:** predict future values using the current and past data
   - Forecasting: mid-term forecasting for several days
   - Nowcasting: short-term forecasting for several hours
3. **Reanalysis:** review the whole sequence given all obtainable data
   - Analyze various biases in the long-term dataset
   - Create a homogeneous dataset to study climate change

Intensity Analysis Task

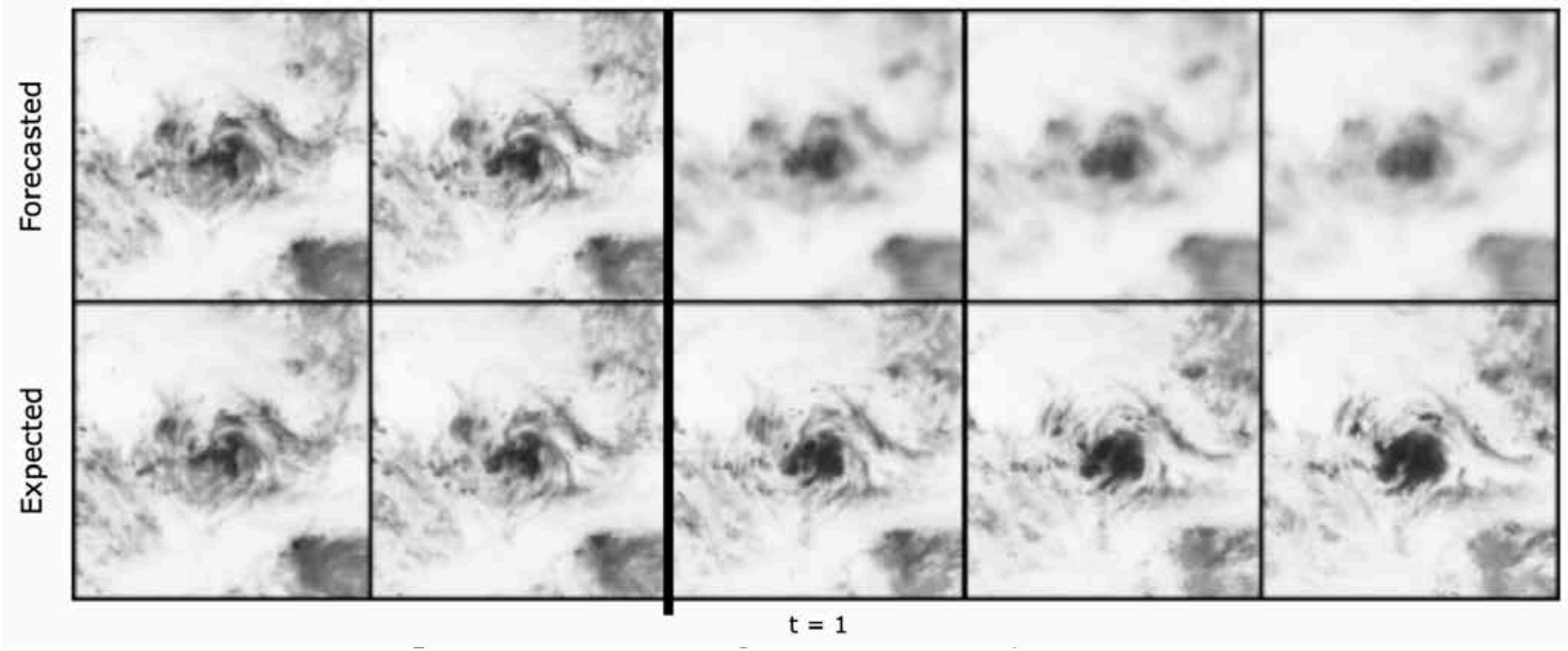


RMSE (ResNet18) 10.06 hPa
RMSE (ResNet50) 10.09 hPa



RMSE (ResNet18) 9.25 kt
RMSE (ResNet50) 9.13 kt

Intensity Forecasting Task



t	1	2	3	6	12
RMSE	10.24 ± 0.73	10.52 ± 0.79	11.00 ± 0.87	12.10 ± 0.85	14.69 ± 0.91

Intensity Reanalysis Task

RMSE	Train the First	Train the Second	Train the Third
Test the First	10.04 (±0.17)	9.92 (±0.09)	10.03 (±0.10)
Test the Second	12.80 (±0.19)	11.05 (±0.10)	11.17 (±0.10)
Test the Third	10.34 (±0.17)	10.03 (±0.08)	9.94 (±0.16)

1. Split the historical dataset by satellite generations (1, 2, 3)
2. Train on one generation and test for another generation
3. Check if the machine can learn similar image features

Other Tasks

1. Analyze tropical cyclones statistically, such as the distribution of eye's diameter and the symmetry of cloud patterns
2. Predict relevant events, such as rapid intensification, eyewall replacement, and overshooting cloud tops
3. Assess the risk of tropical cyclones to society by combining datasets in different modalities
4. Detect trends in long-term data due to technological advancement, human bias in annotation, and climate change

More Information

- Asanobu Kitamoto, Jared Hwang, Bastien Vuillod, Lucas Gautier, Yingtao Tian, Tarin Clanuwat, "Digital Typhoon: Long-term Satellite Image Dataset for the Spatio-Temporal Modeling of Tropical Cyclones", NeurIPS 2023 Datasets and Benchmarks (Spotlight), 2023. Also available as arXiv:2311.02665
- **Digital Typhoon Dataset**
  - http://agora.ex.nii.ac.jp/digital-typhoon/dataset/
- **kitamoto-lab/digital-typhoon @ GitHub**
  - https://github.com/kitamoto-lab/digital-typhoon
- **DIAS (Data Integration and Analysis System) Data Repository**
  - https://doi.org/10.20783/DIAS.664