Predictability of extreme weather events in lead time of two weeks using JMA's Global Ensemble Prediction System

Ryosuke Hoshi (Japan Meteorological Agency) Shingo Ito (Japan Meteorological Agency)

1. Introduction

- Extreme weather events such as heat waves, cold air outbreaks, and heavy snowfall have significant impacts on people's lives and economic activities, and they sometimes cause many mortalities.
  - ⇒ Early preparation for such weather events is needed, and for the purpose, more accurate forecast information of longer lead time is required.
- In general, the longer the lead time, the more the error grows in model forecasting. So it is necessary to construct forecast information that takes the error growth into account.
  - ⇒ In order to capture extreme weather events, Rivoire et al. (2023) examined spatial and temporal aggregation which allows for small predictive errors in the location and time.
- Statistical postprocessing techniques are an essential component to improve the quality of forecasts.
  - ⇒ Currently, many postprocessing techniques have been developed, ranging from simple bias corrections to very sophisticated methods using state-of-the-art machine learning techniques (Vannitsem et al., 2021).

This Study's Purpose

This study investigates the predictability of extreme weather events in lead time of two weeks using a variety of postprocessing techniques and aggregating the forecast information in the location and time.

2. Data

- Reforecast Data (JMA's Global Ensemble Prediction System; GEPS)

Table 1: Configuration of GEPS reforecast data.

Resolution	Ensemble size	Initial dates
0.375º(lon) x 0.375º(lat)	13 members	15th and the last day of month from 1991 to 2020

- Observation Data
  Surface meteorological observation data for temperature (145 sites) and Automated Meteorological Data Acquisition System (AMeDaS) data for wind and snow (approximately 930 sites) in Japan (Fig.1)

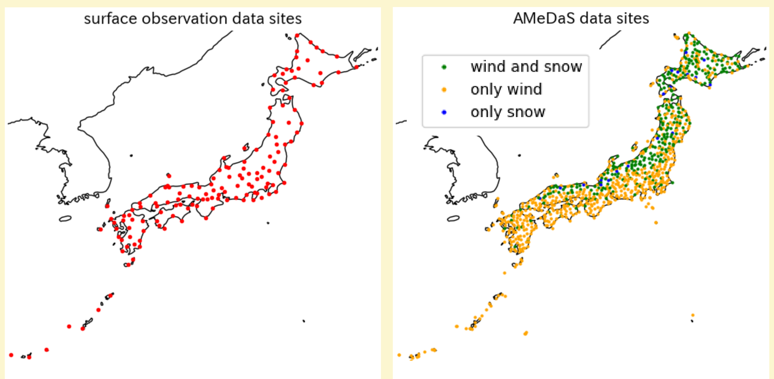


Fig.1: Sites of observation data.

3. The predictability of extreme temperature

- We investigated the predictability of extreme temperature. Extreme high / low temperature was defined as daily 90th / 10th percentile temperature.
- Fig.2 shows a comparison of the forecast skills of various postprocessing methods. As we used the standardized anomalies of observations and predictions, seasonal specific characteristics were largely removed (Dabernig et al., 2017) and the parameters of the postprocessing methods could be fitted for all seasons. The verifications were conducted for all surface observation points and all seasons.

Table 2: Overview of the different postprocessing methods.

Methods	Predictors (ensemble mean)	Fitting of parameters
Linear Regression (LR)	T_{sfc}	individual for every station
Multiple Linear Regression (MLR)	T_{sfc}, $TTD_{850/925hPa}$, 3h FRR, U_{sfc}, V_{sfc}, CLM	individual for every station
Natural Gradient Boosting (NGB)	T_{sfc}, $TTD_{850/925hPa}$, 3h FRR, U_{sfc}, V_{sfc}, CLM, lat, lon, altitude	simultaneous for all stations
Quantile Regression Forests (QRF)	T_{sfc}, T_{max}, T_{min}, T_{Q75}, T_{Q50}, T_{Q25}, $TTD_{850/925hPa}$, 3h FRR, U_{sfc}, V_{sfc}, CLM, lat, lon, altitude	simultaneous for all stations
Support Vector Regression (SVR)	T_{sfc}	individual for every station

※ T: temperature, TTD: dew-point depression, FRR: precipitation, U: zonal wind speed, V: meridian wind speed, CLM: cloud cover in low and medium levels, Q: quantiles

※ For methods other than QRF, the predictive distribution was assumed to be normal.

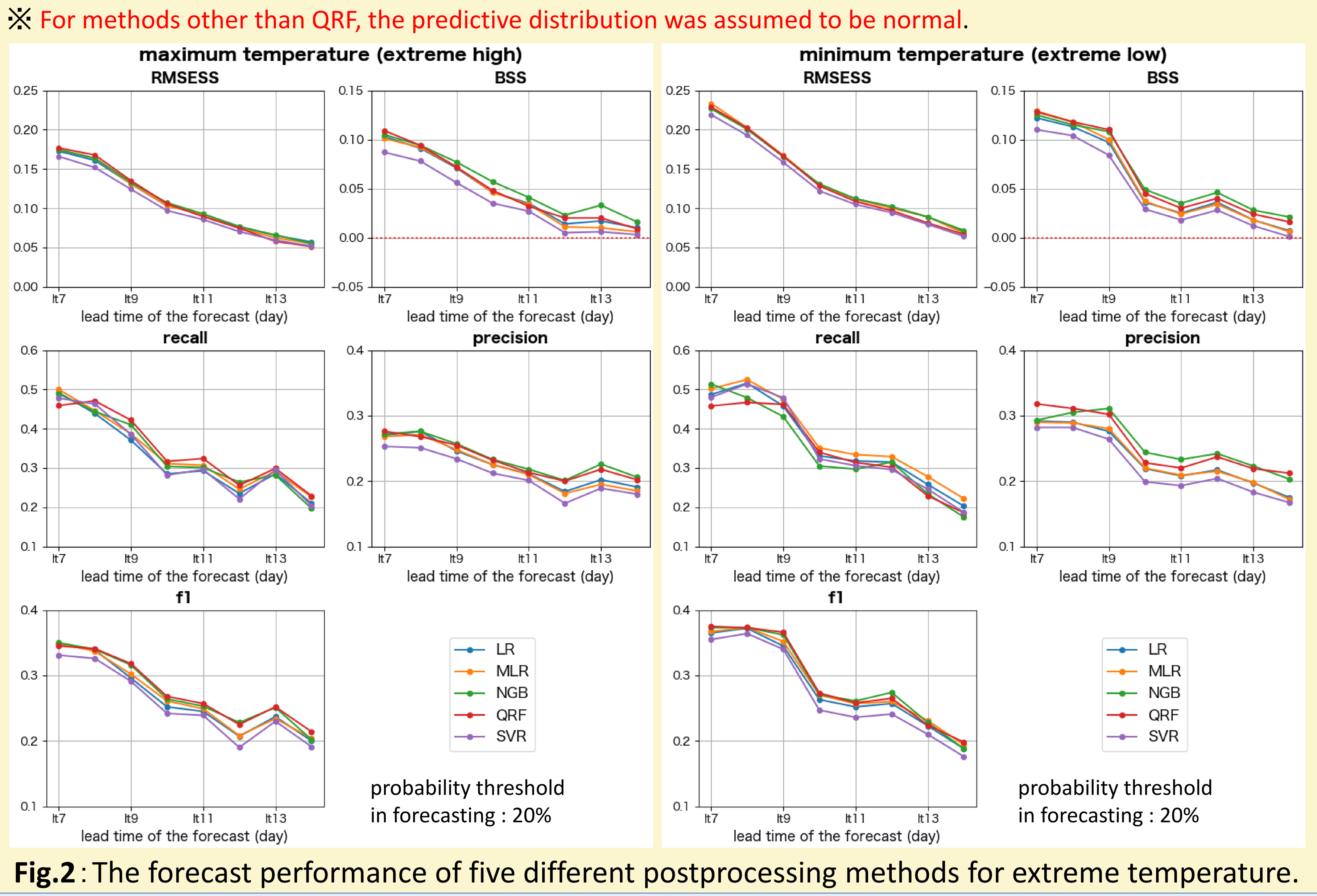


Fig.2: The forecast performance of five different postprocessing methods for extreme temperature.

◎ While there is little difference in RMSESS, **BSS and F1-score are higher in ensemble learning methods such as NGB and QRF.**

◎ In the second half of the second week, **both recall and precision are less than 0.3.**

4. Temporal aggregation

- In the second week, predictive errors in the location of synoptic scale phenomena such as anticyclones and cyclones are often observed. It is strict to predict extreme weather events on the exact same day. Thus, we analyzed extreme events that were temporally aggregated.
- We counted the number of ensemble members which could predict the extreme temperature (Fig.3). While in many events the number of ensemble members which exceeded the threshold of the extreme temperature on the exact same day is less than 1, predictable number of members increases when the forecast errors of 1 day before and after are allowed (3d window).

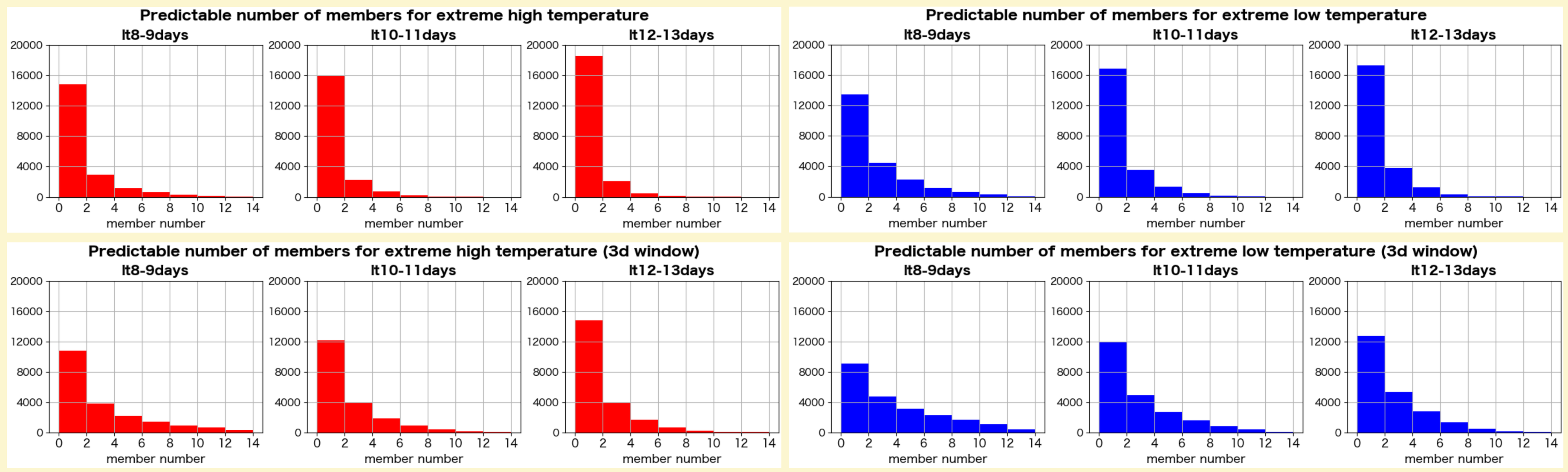
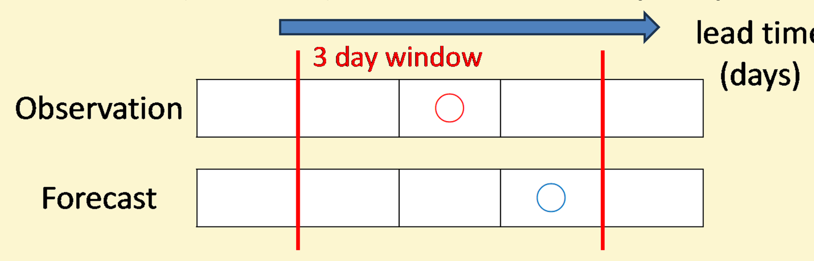


Fig.3: Histogram of the number of ensemble members which could predict extreme temperature on the exact same day (upper) and in the 3d window (bottom). LR was used as postprocessing.

- We examined the forecast skills on the verification "one event or more of extreme temperature occurred during 3 days" (Fig.4).



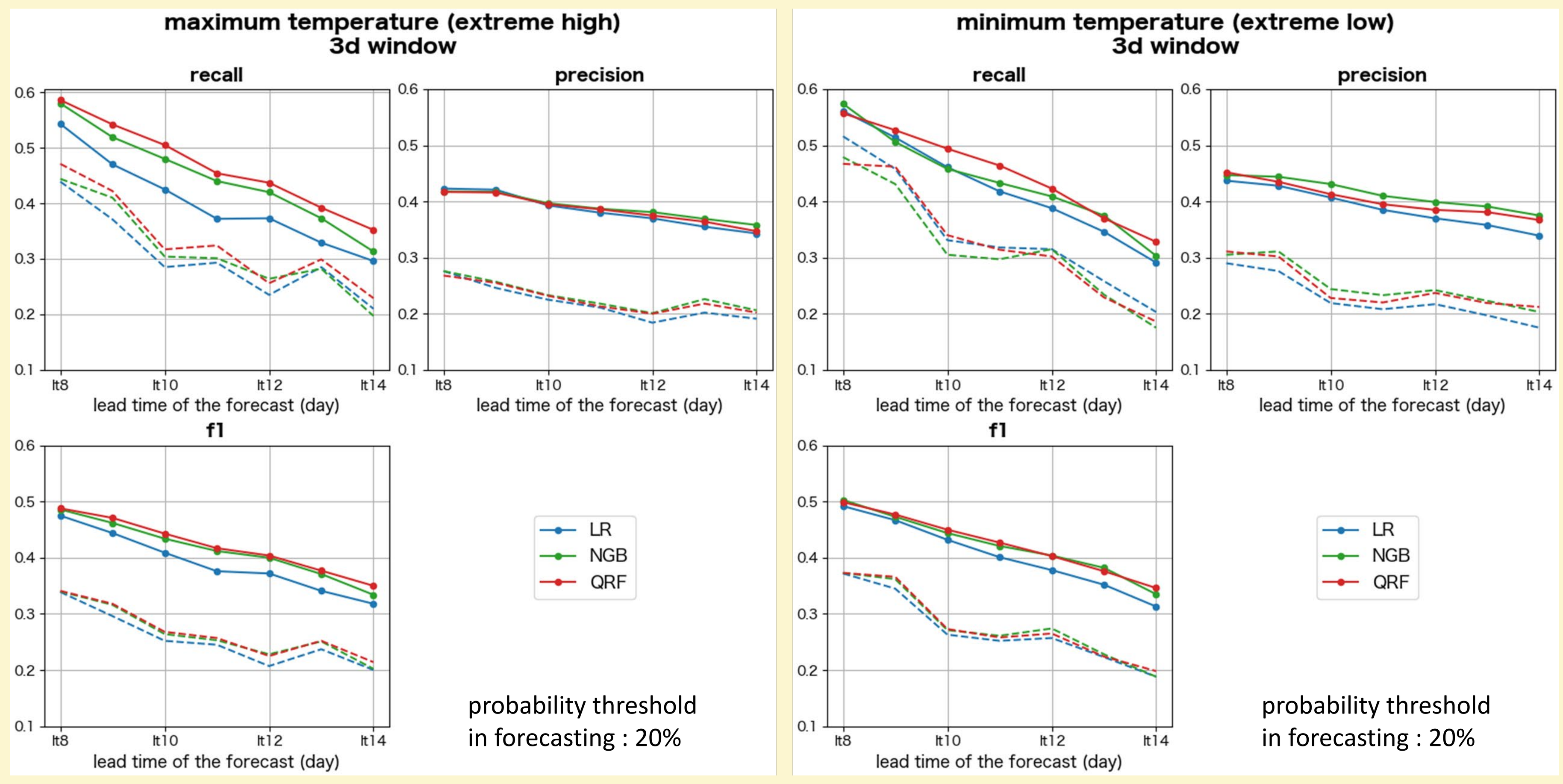


Fig.4: The forecast score for a minimum of one extreme temperature event in the 3d window (solid line). The dashed lines indicate the score required to be predicted on the exact same day, same as the results of Fig.2.

◎ **The skill for the temporally aggregated extreme temperature is much larger.**
In the second half of the second week, **both recall and precision are about 0.4.**

◎ NGB and QRF are more accurate than LR.

5. Various other phenomena

- We also investigated the predictability of storms and heavy snowfall, focusing on the winter season (November-March) when the phenomena occurs more frequently.

Table 3: Overview of storms and heavy snowfall surveys.

Target phenomena	Methods	Predictors (ensemble mean)
Storms (90th percentile wind speed)	Non-homogeneous Regression (NR)	Surface wind
Heavy Snowfall (95th percentile snowfall)	Gamma Regression	Precipitation, (Temperature)

※ Temporal window: 1day, 3days Spatial window: focus on the maximum value in the prefecture (about 5000km²)

※ Surface temperature was used to determine the type of precipitation.

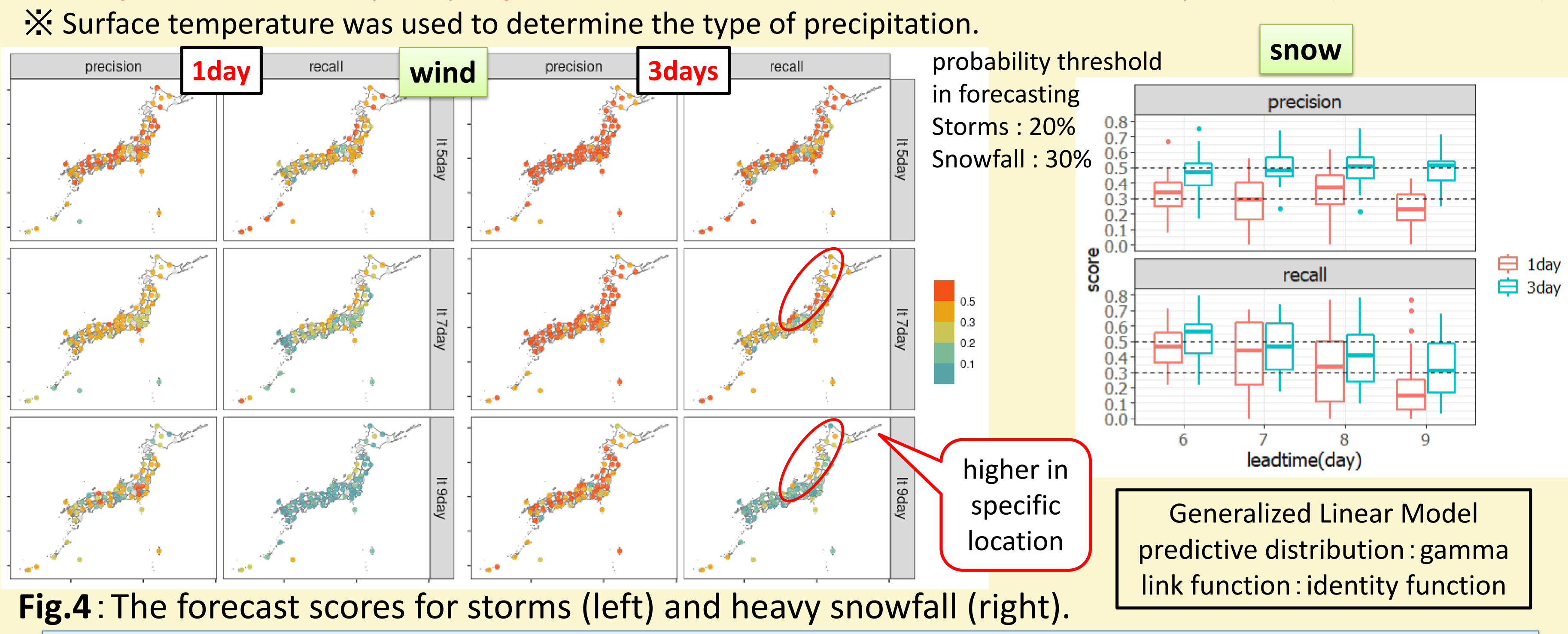


Fig.4: The forecast scores for storms (left) and heavy snowfall (right).

◎ The predictability of storms and heavy snowfall was confirmed until the first half of the second week when aggregating the extremes spatially and temporally.

6. Conclusion

- By considering spatial and temporal aggregation, we could find the predictability of extreme weather events in lead time of two weeks.
- Incorporating machine learning methods such as NGB and QRF could improve forecast accuracy (extreme temperature).

<References>

- Dabernig, M., Mayr, G. J., Messner, J. W., & Zeileis, A. (2017). Spatial ensemble post-processing with standardized anomalies. Quarterly Journal of the Royal Meteorological Society, 143(703), 909-916.
- Rivoire, P., Martius, O., Naveau, P., & Tuel, A. (2023). Assessment of subseasonal-to-seasonal (S2S) ensemble extreme precipitation forecast skill over Europe. Natural Hazards and Earth System Sciences, 23(8), 2857-2871.
- Vannitsem, S., Bremnes, J. B., Demaeyer, J., Evans, G. R., Flowerdew, J., Hemri, S., et al. (2021). Statistical postprocessing for weather forecasts: Review, challenges, and avenues in a big data world. Bulletin of the American Meteorological Society, 102(3), E681-E699.