Spatial correlation of the observation data and its application to validate the regional climate model

Noriko Ishizaki, Izuru Takayabu, and
Masao Kanamitsu

Introduction

- Higher resolution model have been developed.
  ⇒How can we compare the model skill?
- The model value is representative of each grid. On the other hand, the observational stations have different spatial extent of representation.

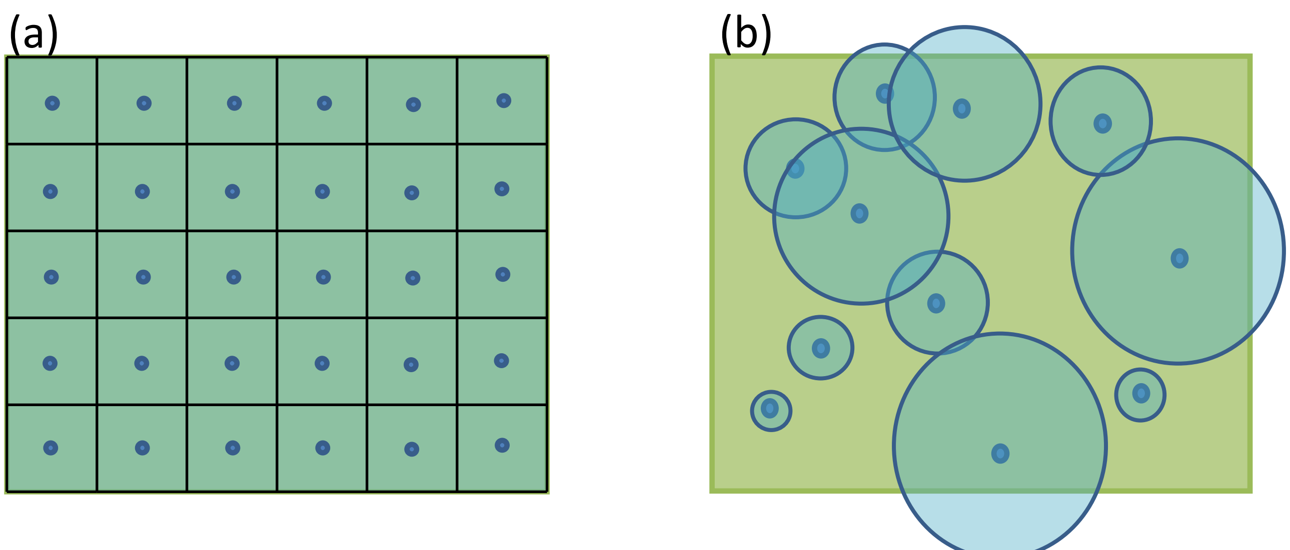
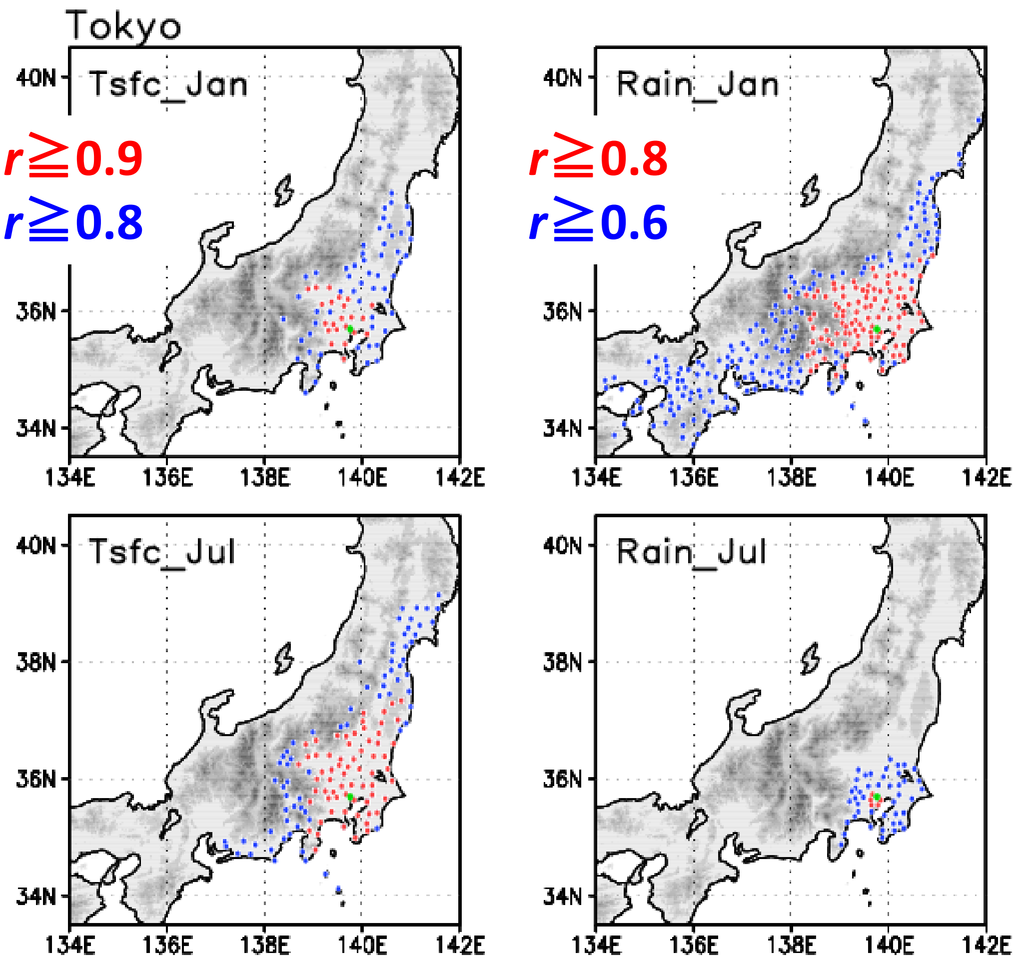


Fig.1: Representative area at each grid in the model (a) or station (b).

Anisotropy of representativeness

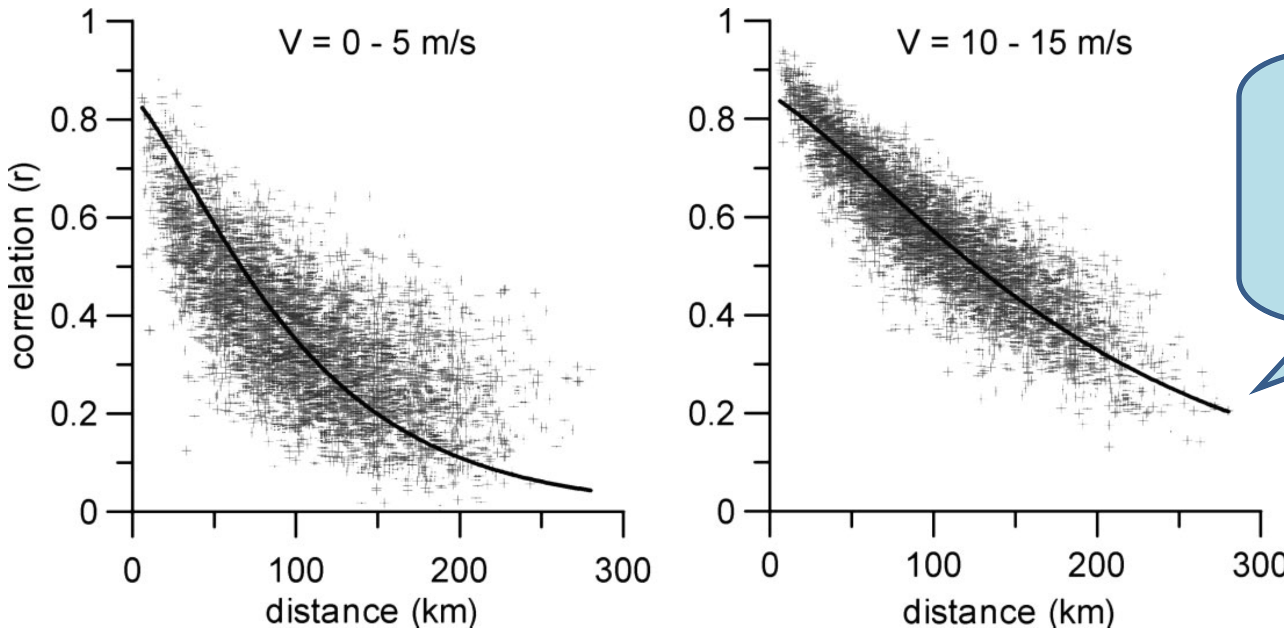


Spatial correlation of meteorological variables may be one indicator of the representative area.

The shapes of high correlation area are **not isotropic**.
They have different distributions between temperature and precipitation, and seasons.

Fig.2: Distribution of the stations which have high correlation coefficient of the surface temperature (left) and precipitation at Tokyo AMeDAS station. Upper panels are for January, and lower panels are for July.

Correlation decay distance (CDD)



Precipitation is less widely spread at low wind speeds.

Figure 5 of Johansson, B. and D. Chen(2003): Scatter plot of correlation between daily precipitation versus distance between stations.

CDD depends on…

- **Wind speed** (e.g., Johansson and Chen 2003)
- **Geographical location** (e.g., Jones et al. 1997, Hulme 1997)
- **Season** (e.g., Alexander et al. 2006)

Station to grid (objective analysis)

- **Distance weighting**
  - APHRO_JP (Kamiguchi et al. 2010)
  - Hofstra, N. and M. New (2009)
- **Regression**
  - PRISM (Daly et al. 1994)
  - Mesh_Clim
- **Kriging / Optimal interpolation**
  - JRA-25 (surface temperature)
  - ERA15

Distance, Altitude, Slope, Slope direction, Facet, Wind speed, Latitude, Longitude, Distance to ocean…

Although there are various interpolation method, a reliability or representativeness of a point have not been fully discussed so far.

Purpose of this study

- How can we compare model skill with different horizontal resolution?

□ **What is the parameter** which characterize the representativeness?

□ **How should we quantify and use** the information of differences of representativeness error for validation?

- We used the spatial correlation to see the feature of representativeness.

Representativeness of AMeDAS station

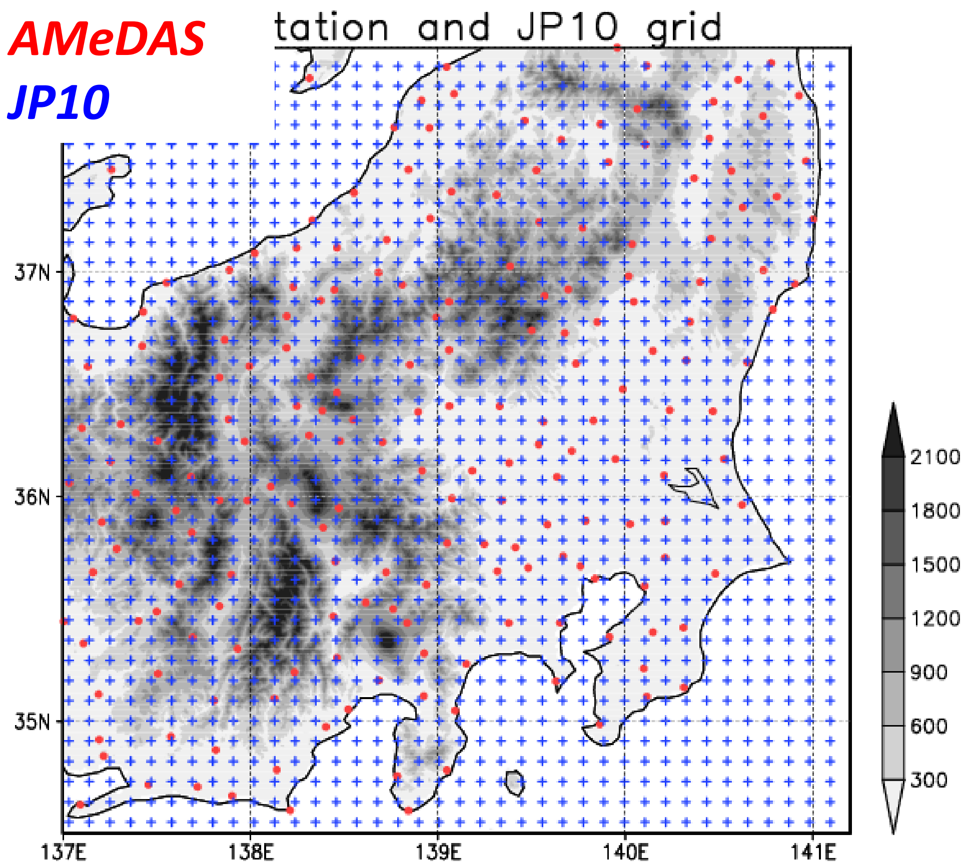


Fig. 3: Comparison of horizontal distribution for AMeDAS stations (red dots) and 20km-NHRCM grid (blue dots). Gray shade denotes topography with unit [m].

Observation:

- AMeDAS (~17km)
- 768 stations over Japan
- Daily temperature and precipitation
- 1980-2004

Topography:

- GTOPO30

Model results:

- JRA-25 (~100km)
- NHRCM (20km)
- JP10 (10km)

Spread direction of representativeness

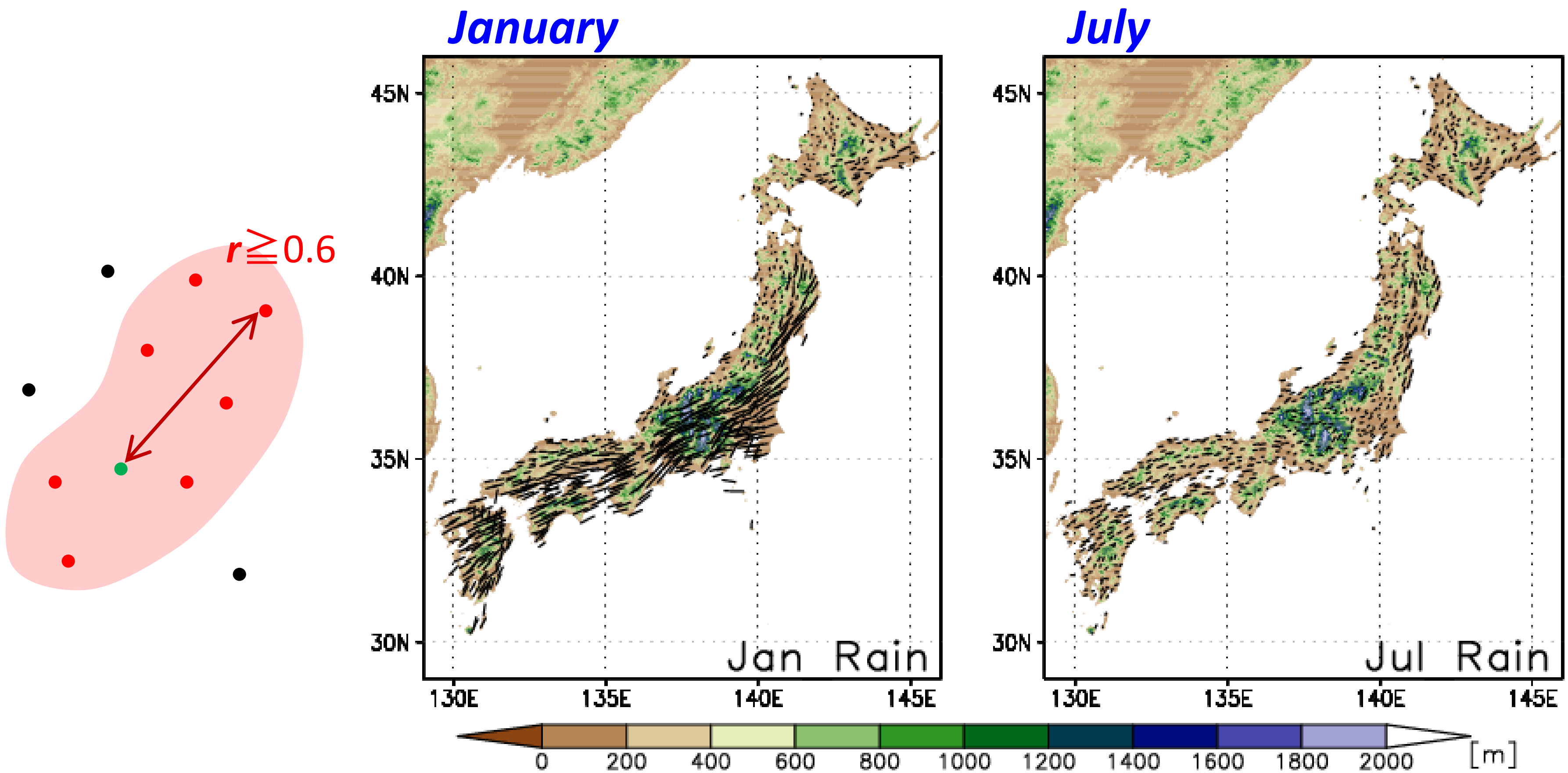
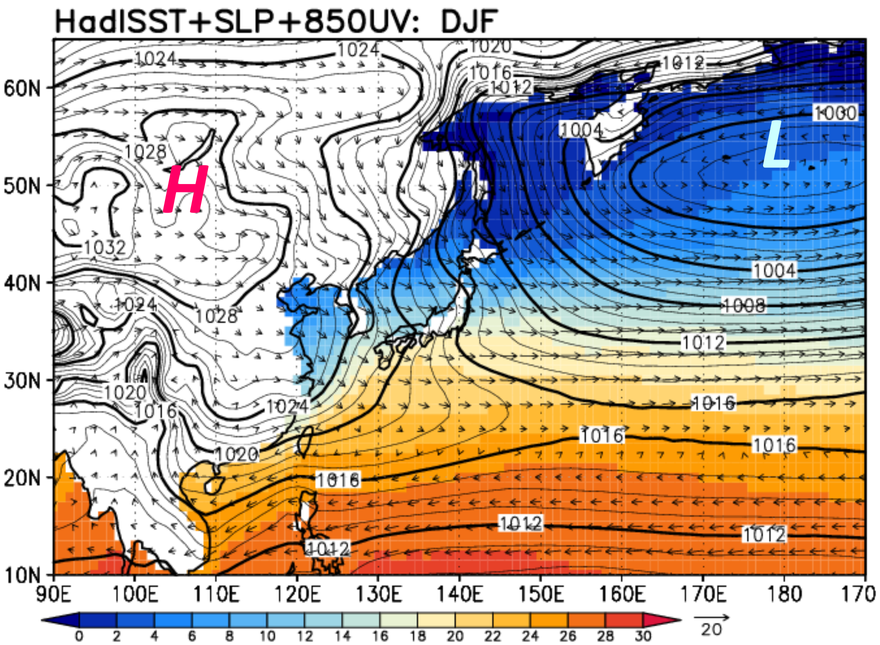


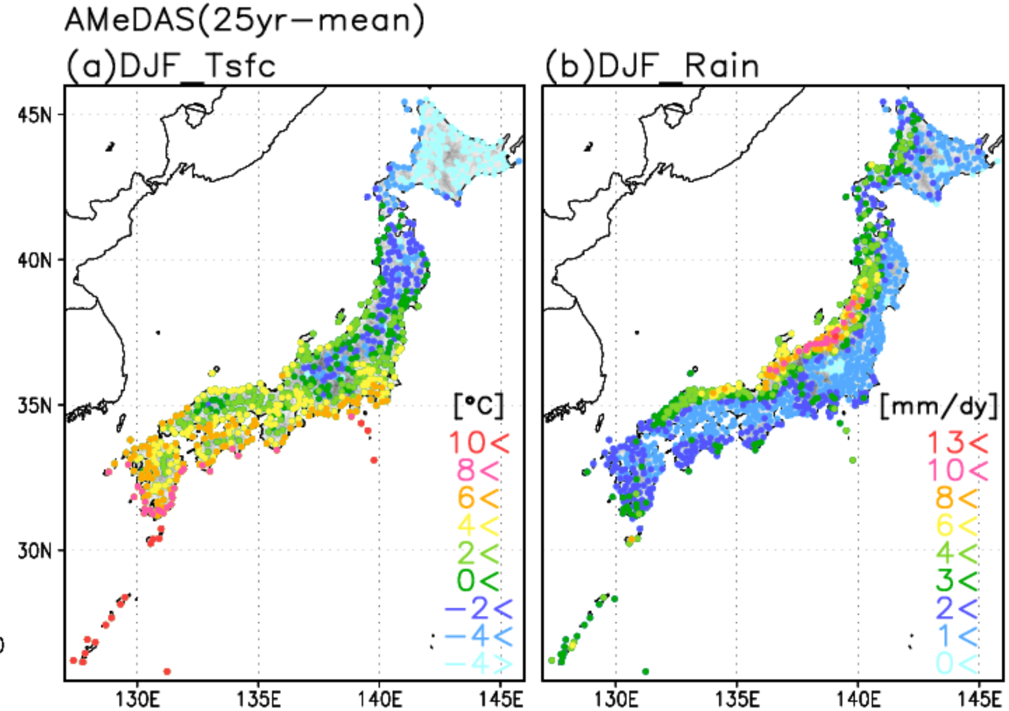
Fig. 4: The direction and distance toward farthest station with correlation of daily precipitation larger than 0.6.

In January, the direction extend parallel to the mountain ranges in many regions. However, the length of the shape in the Japan Sea side gives remarkable contrast to the Pacific side.

January in Japan



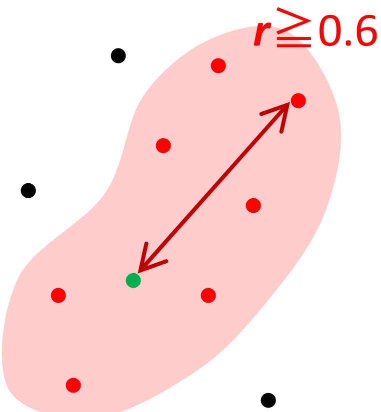
SST, Sea level pressure and flow pattern on 850 hPa level.
[18-year HadISST, 26-year mean JRA-25]



Surface observation of the surface temperature and rainfalls in DJF. [25-year mean AMeDAS]

Cold air outbreak from Siberian high brings heavy snow to Japan Sea side region during boreal winter. It is in contrast to Pacific side which have less rain. The snow area corresponds to small CDD region.

Spread direction of representativeness



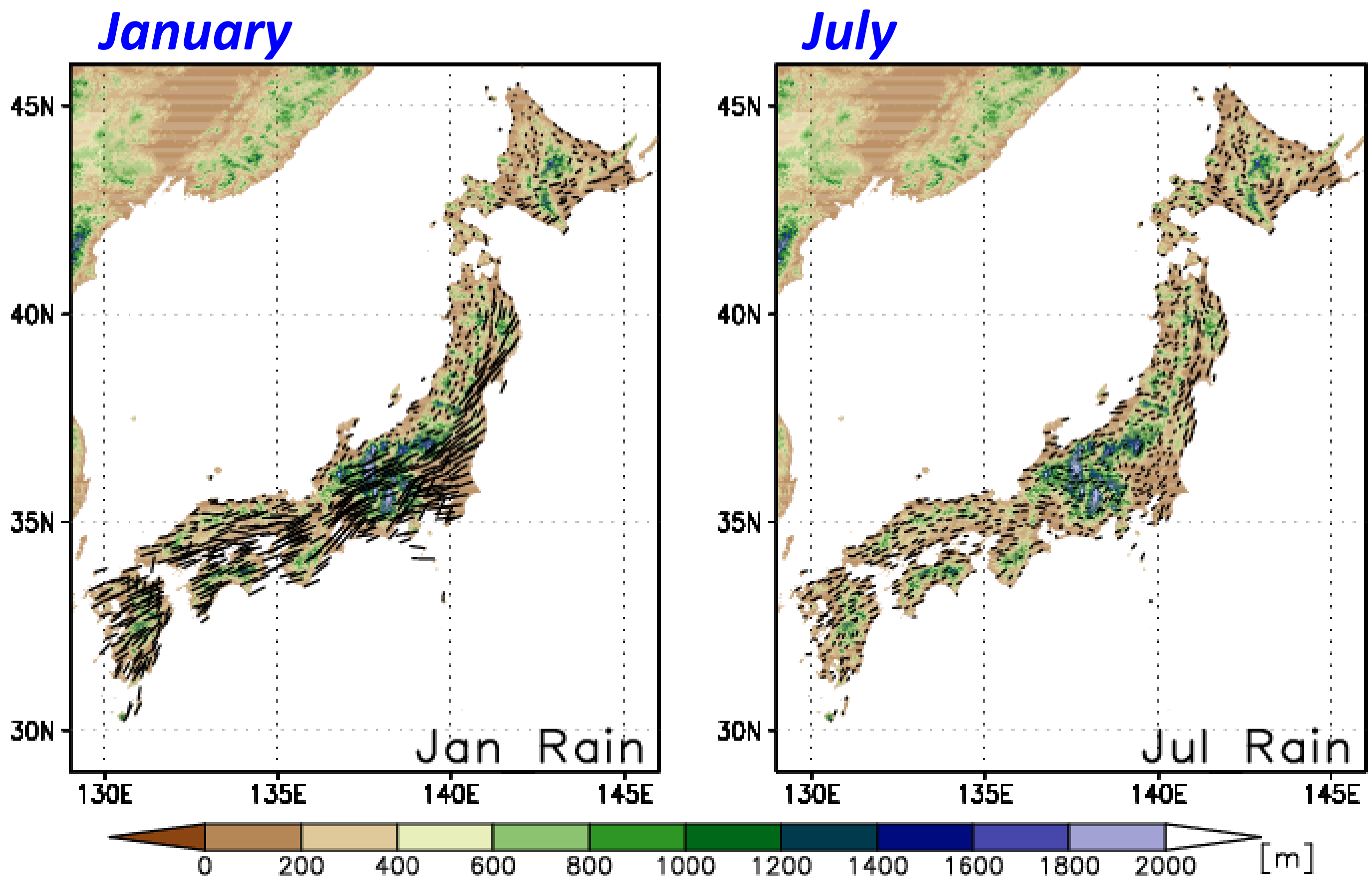
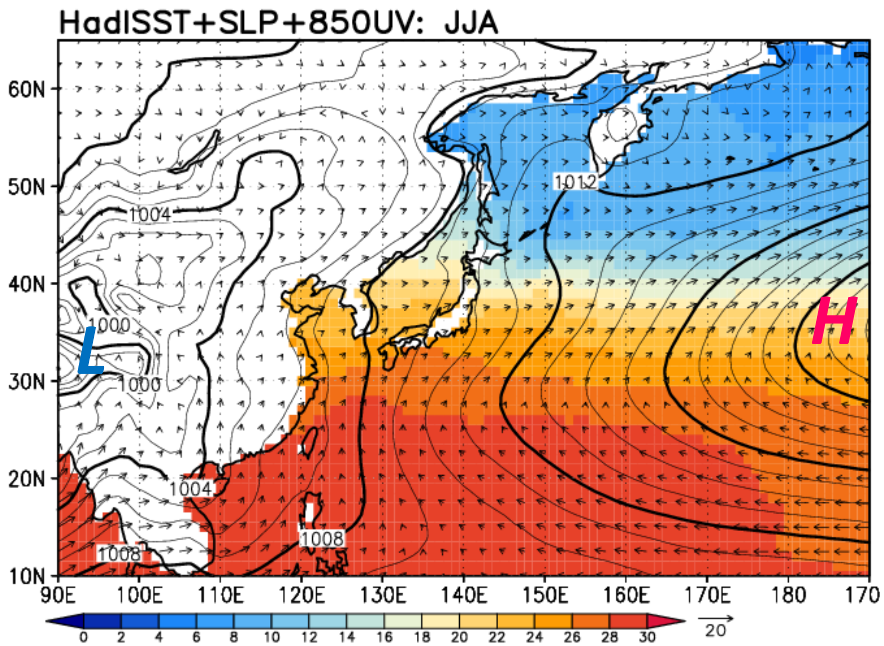


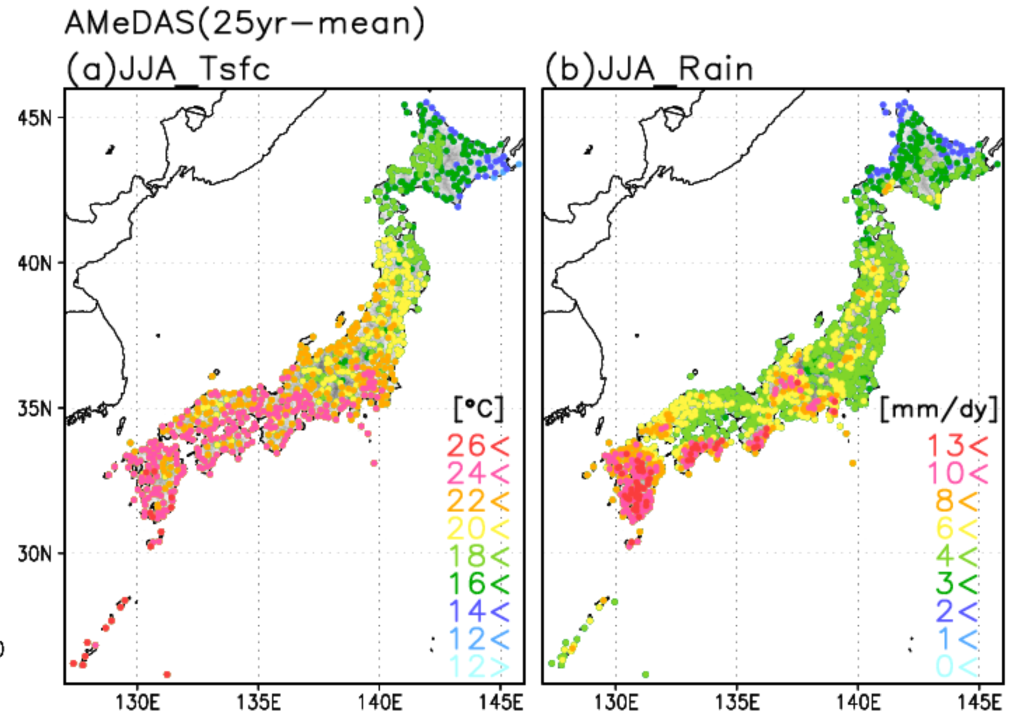
Fig. 4: The direction and distance toward farthest station with correlation of daily precipitation larger than 0.6.

In July, CDD is relatively small in whole Japan compared to winter season.

July in Japan



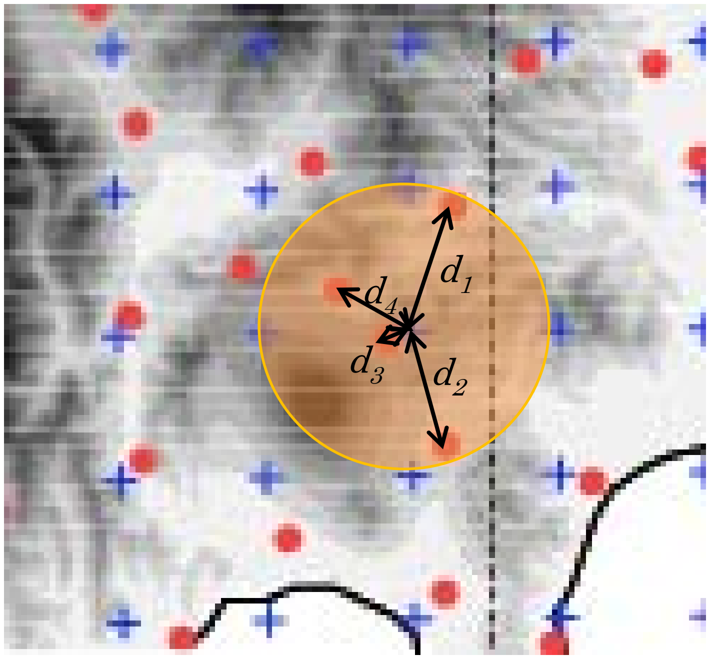
SST, Sea level pressure and flow patter on 850 hPa level.
[18-year HadISST, 26-year mean JRA-25]



Surface observation of the surface temperature and rainfalls in JJA. [25-year mean AMeDAS]

Precipitation amount generally becomes larger than winter season. Much of rain is brought by convective rainfall events. This can explain relatively small CDD in summer season in Japan.

Simple interpolation for validation



AMeDAS
Model

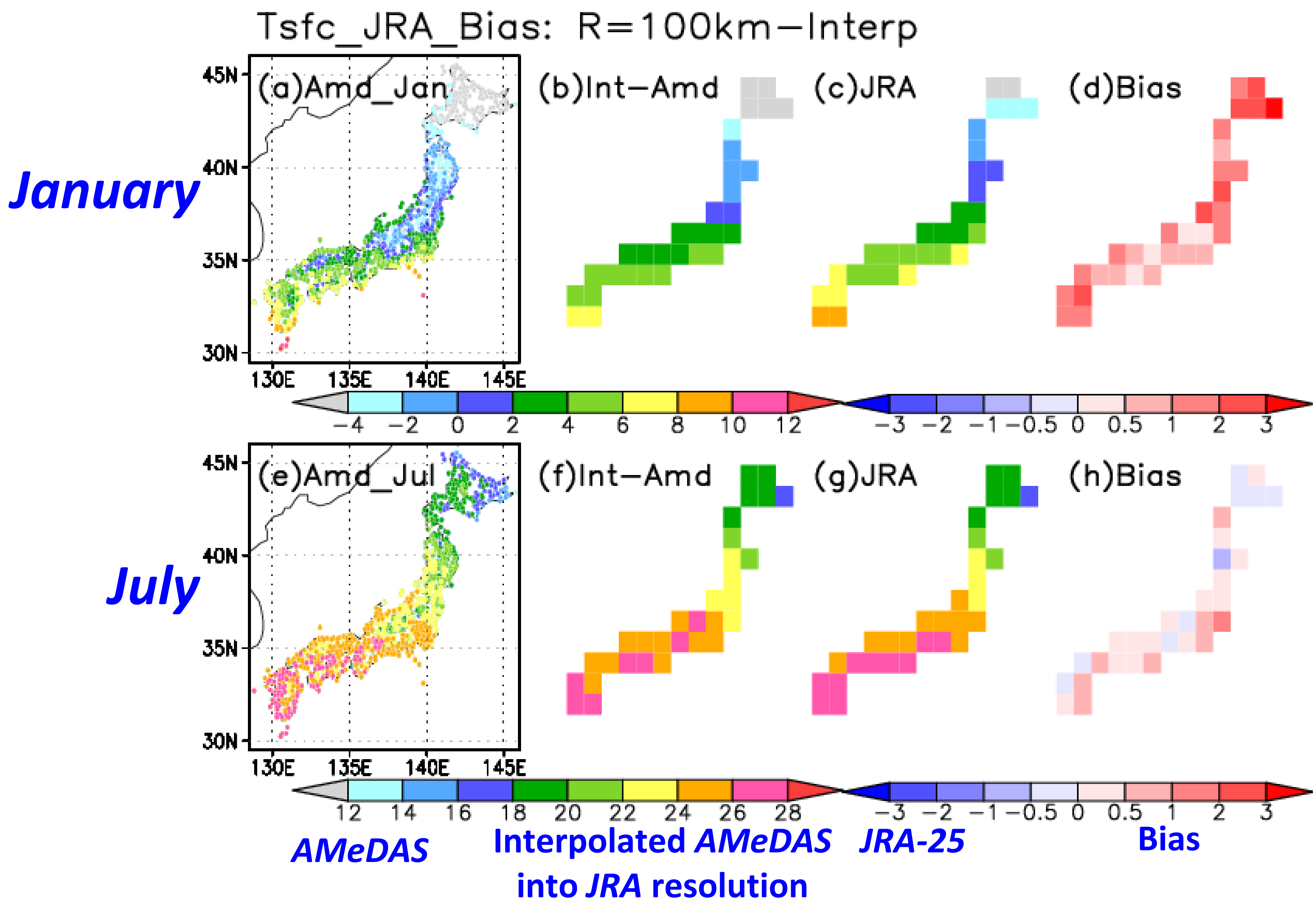
Radius $\begin{cases} \text{=100km (JRA-25)} \\ \text{=20km (NHRCM)} \\ \text{=10km (JP10)} \end{cases}$

$$O_{interp} = \frac{\sum_{k=1}^{n} \frac{1}{d_k^2} O_k}{\sum_{k=1}^{n} \frac{1}{d_k^2}}$$

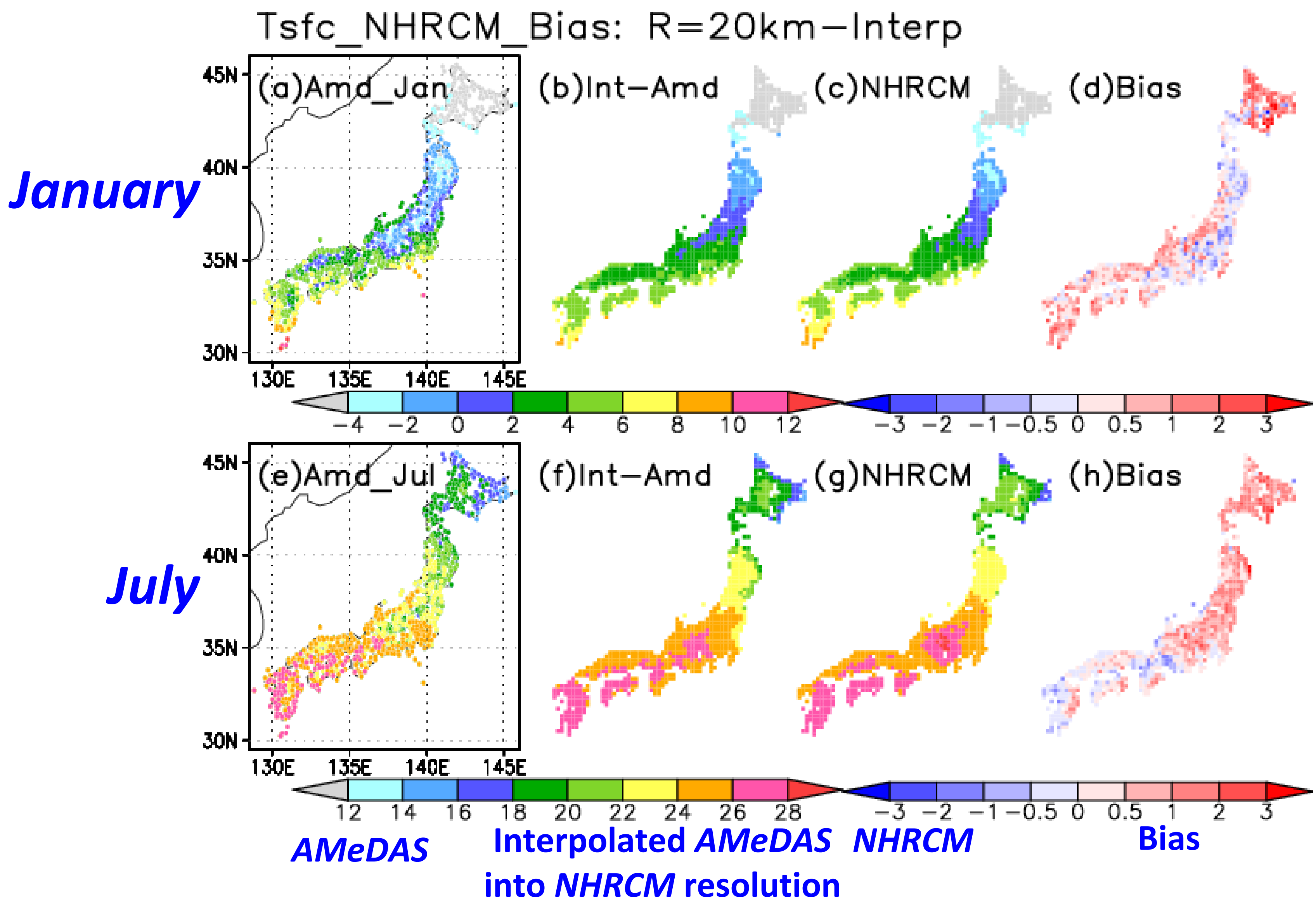
As a first step, we apply the distance from stations to the target grid as “weight” in order to prepare the gridded observation data set.

We use all stations inside of a circle with radius the model resolution.

Example of 100km-JRA: Temperature



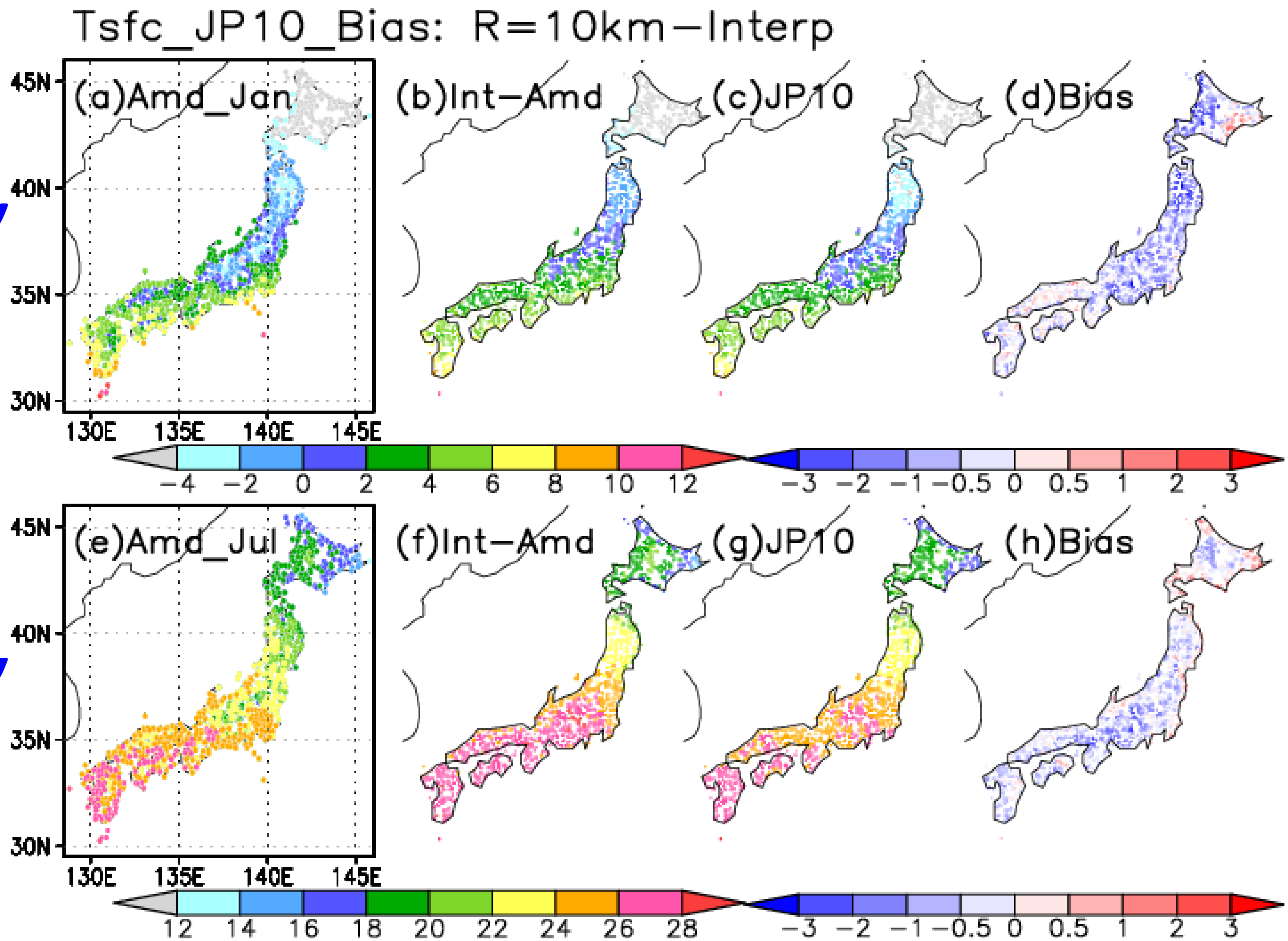
Example of 20km-NHRCM: Temperature



Example of 10km-JP10: Temperature

January

July



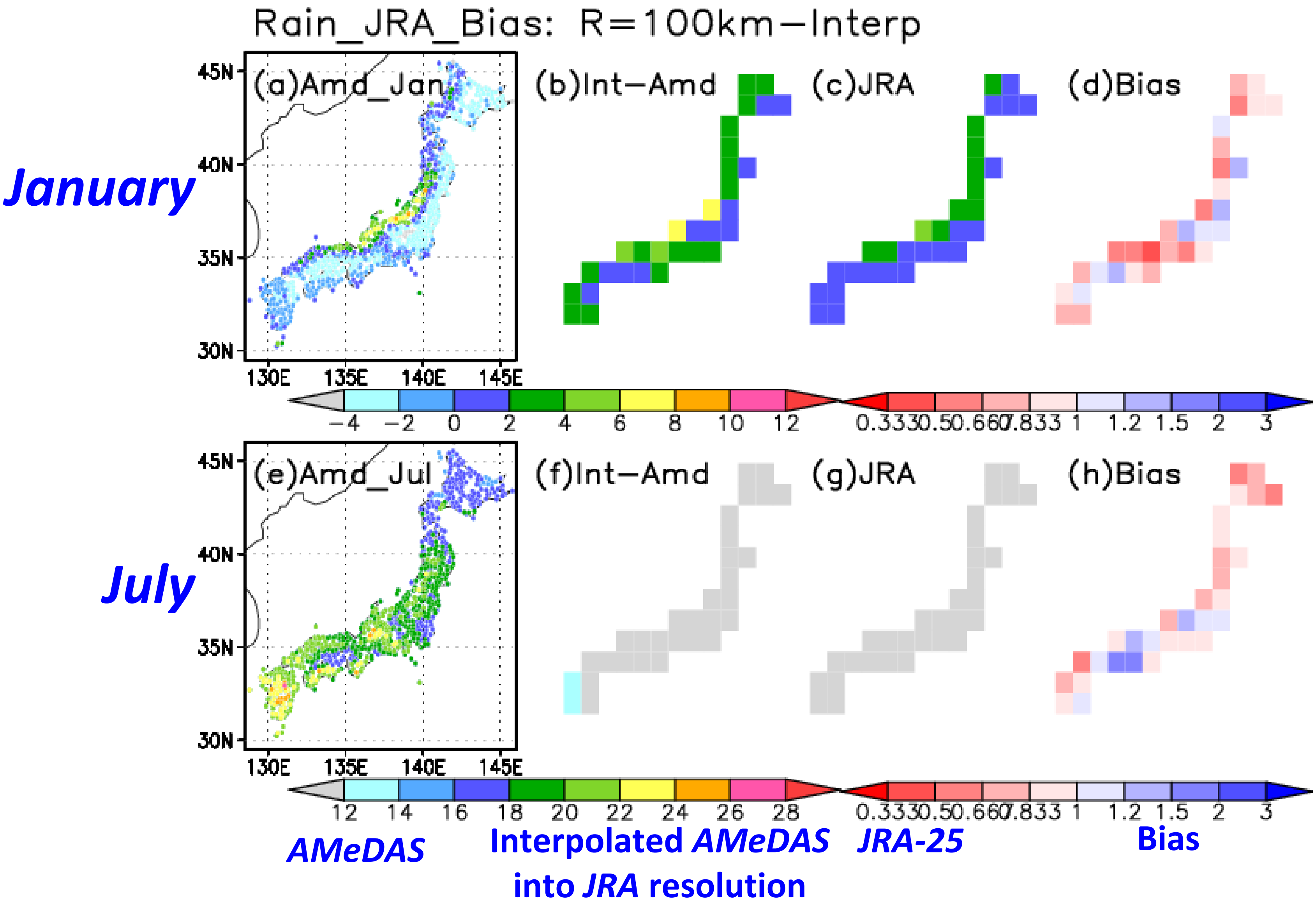
AMeDAS

Interpolated *AMeDAS* into *JP10* resolution

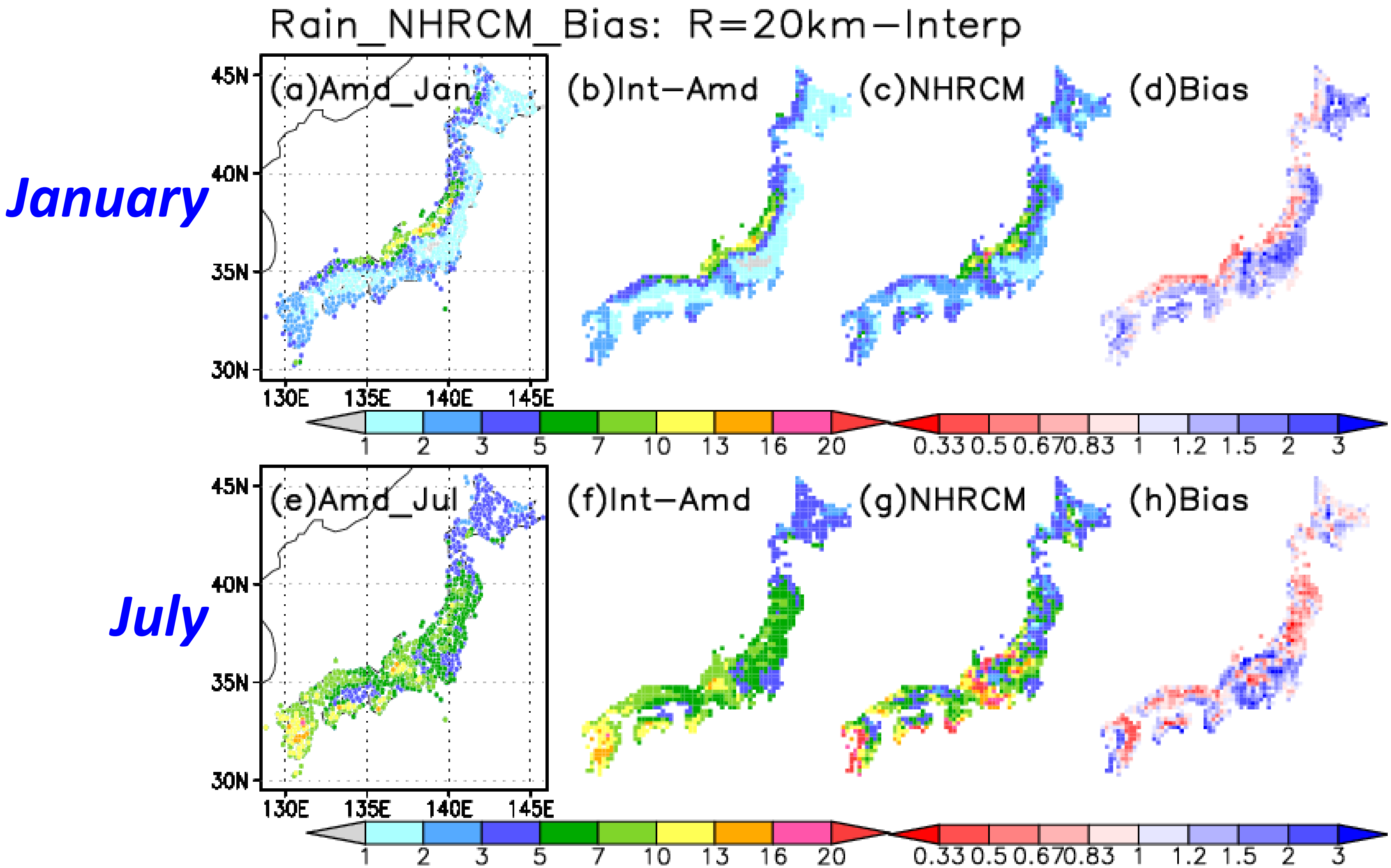
JP10

Bias

Example of 100km-JRA: Precipitation



Example of 20km-NHRCM: Precipitation



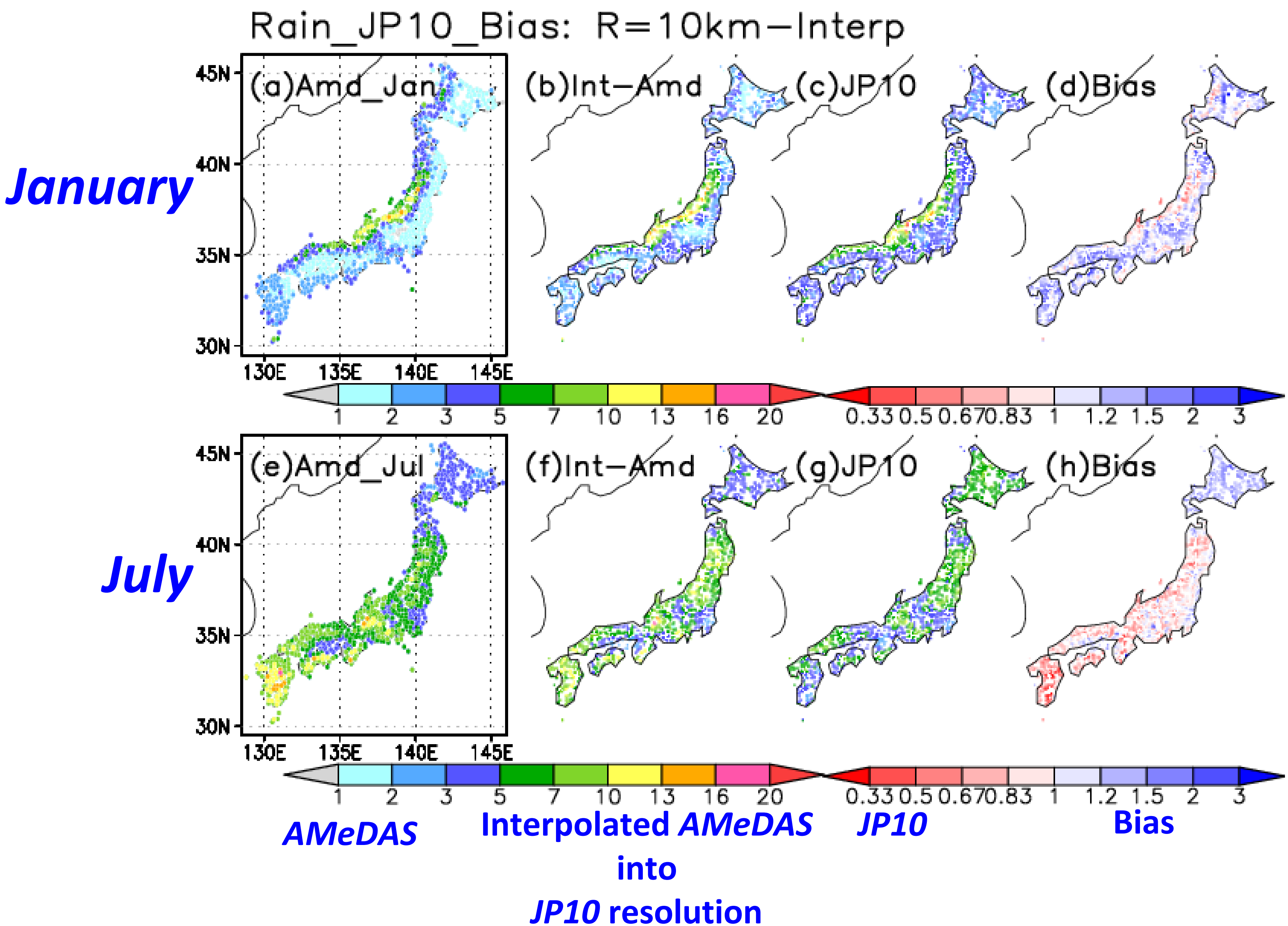
AMeDAS

Interpolated *AMeDAS* into *NHRCM* resolution

NHRCM

Bias

Example of 10km-JP10: Precipitation



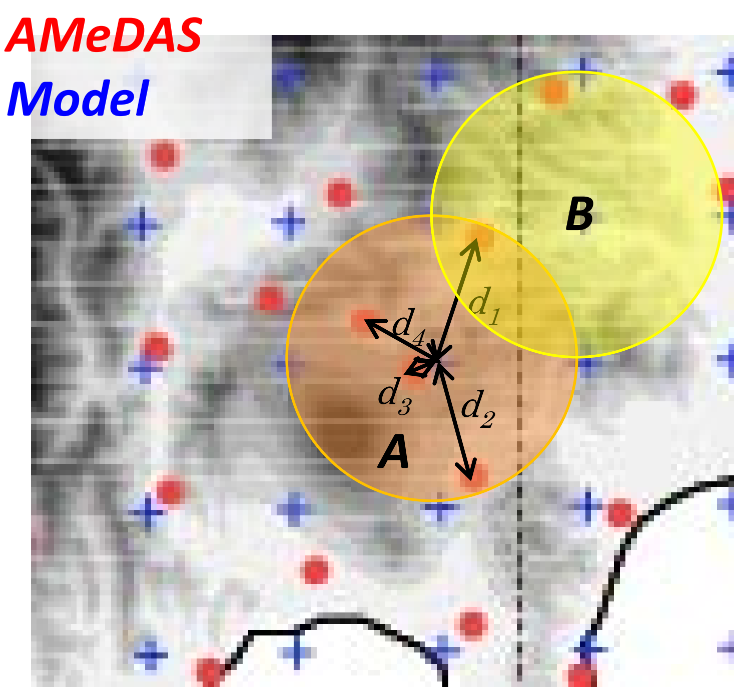
Intercomparison of model bias with different horizontal resolution

$$\cancel{w = \sum_{k=1}^{n} \frac{1}{d_k^2}}$$

$$w = \sum_{k=1}^{n} \frac{1}{(d_k + 1)^2}$$

The summation of the weight in each grid can be regarded as the index of reliability of the interpolated value.

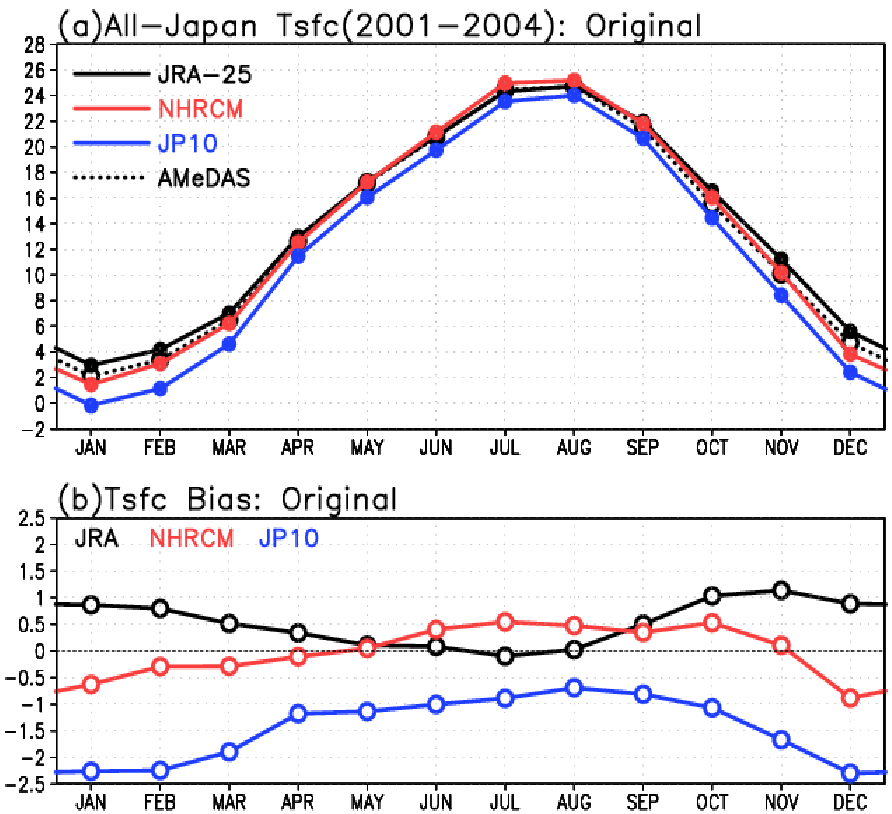
To avoid concentration of weight at few grids, the weight function is changed.



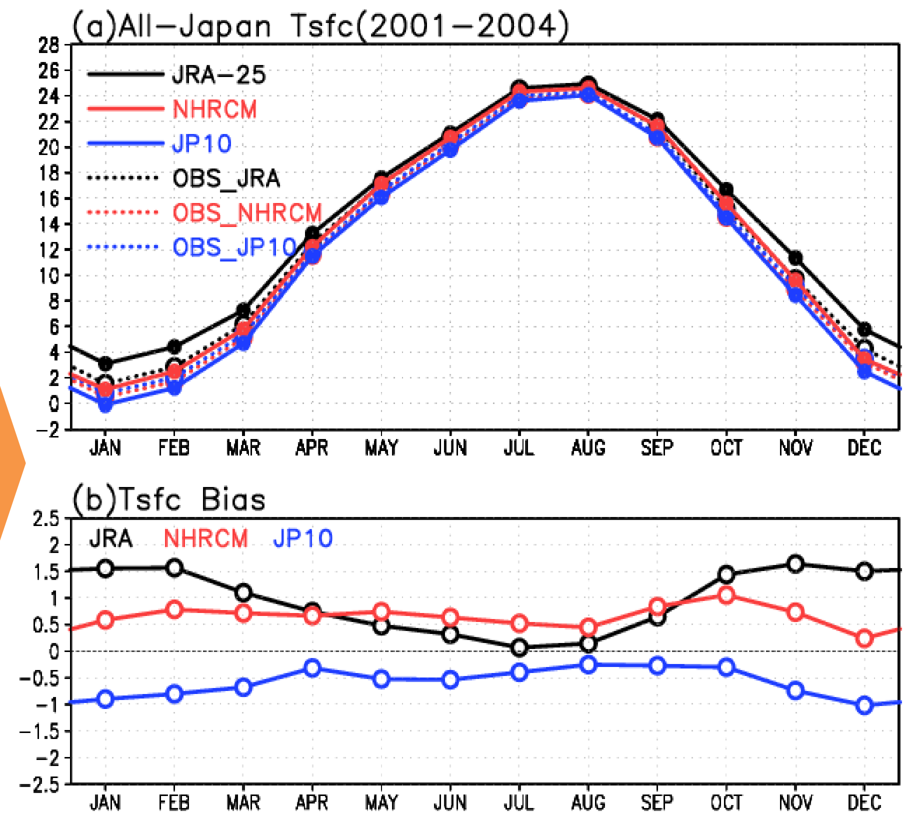
←A grid has a station which is mush closer to the grid compared to B grid. Thus, we consider the representativeness error in A is relatively small.

Surface temperature bias over Japan

Average of all grids/stations



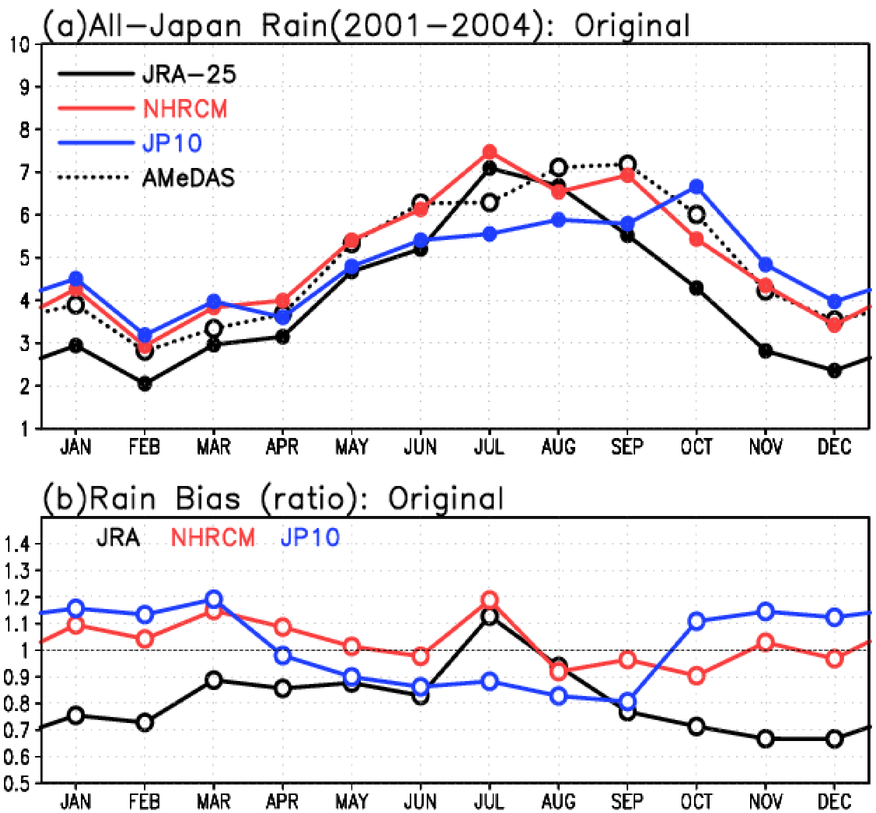
Average using weight



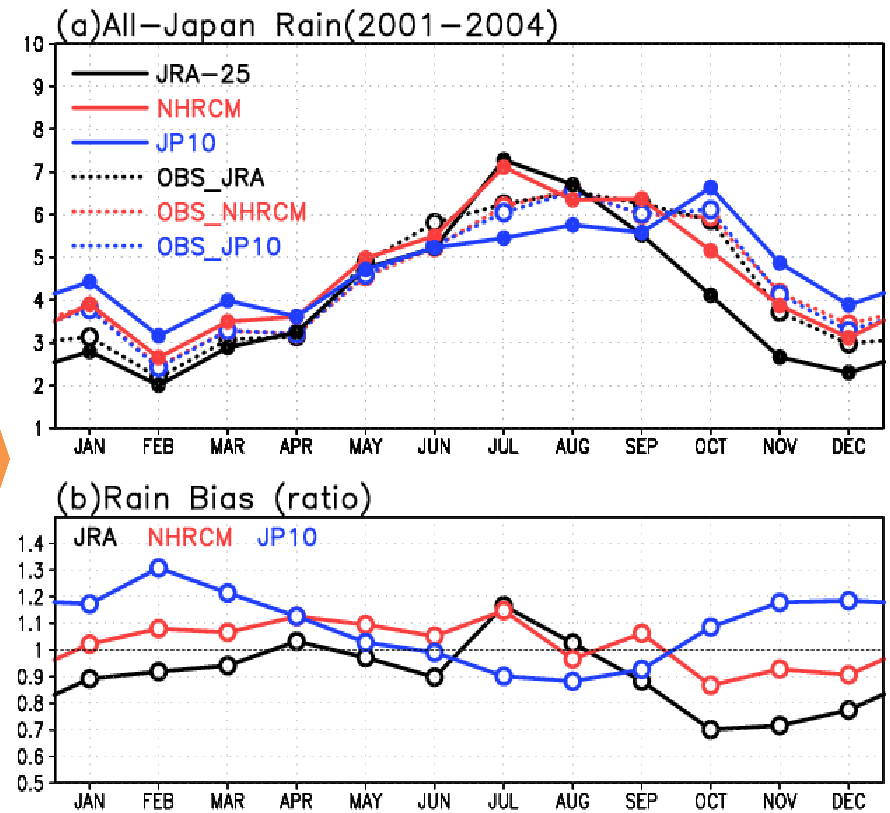
Cold bias in winter of JP10 is mainly due to the heterogeneity of observation (e.g., lack of data around mountainous region) and incorrect treatment of the representativeness.

Precipitation bias over Japan

Average of all grids/stations



Average using weight



Simple average of all grids/stations or inappropriate weight function may lead misreading of the model skill.

Problems

- The regional mean highly depends on small number of grids.
- The effect of the wind fields or local geographical factor is not still considered.
- If we regard the weight of observation as a index of representativeness error, we should include the potential factors (which contribute to the variation of meteorological field) into the weight.

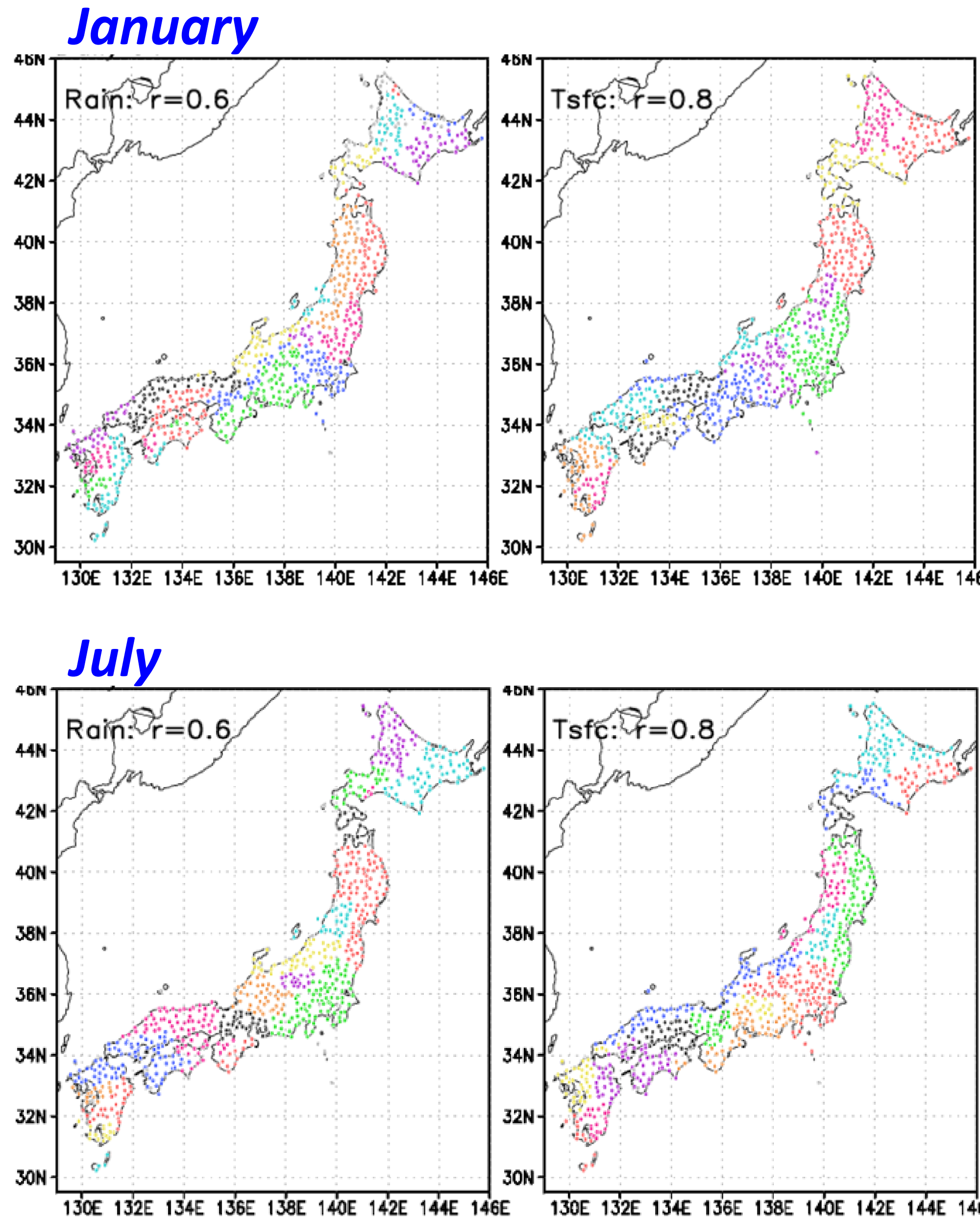
Future Plan

- The shape of the high correlation area is affected by the large-scale circulations, as much as the local geographical features.
- In order to estimate the representativeness error for validation, we should know **the potential factors in which contribute to the shape of anisotropy of high correlation area** in Japan.

Cluster analysis

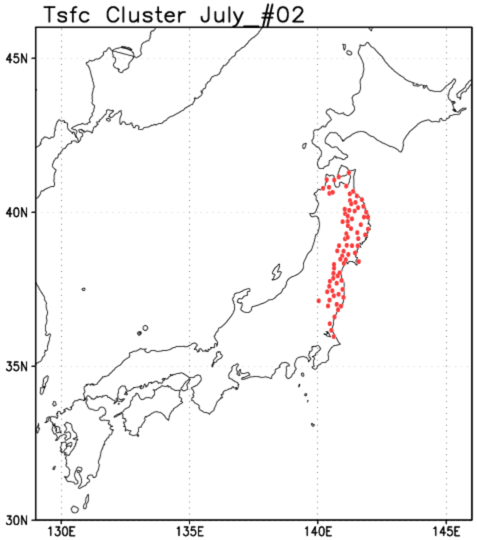
Using the correlation coefficient of AMeDAS, we divide Japan into some clusters. It is considered that the stations within the same cluster show similar daily variation.

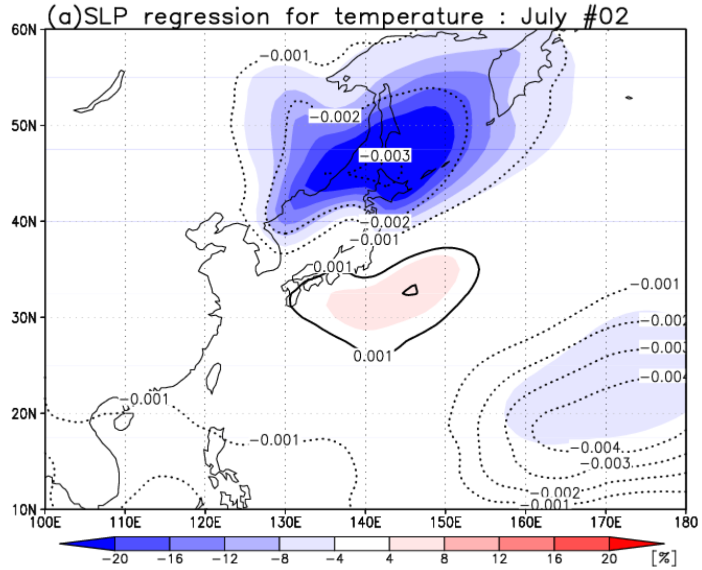
Generally, the correlation coefficient of the temperature is much larger than that of precipitation.



Linear regression: Ex.1 Tsfc #2

Relation with synoptic scale circulation

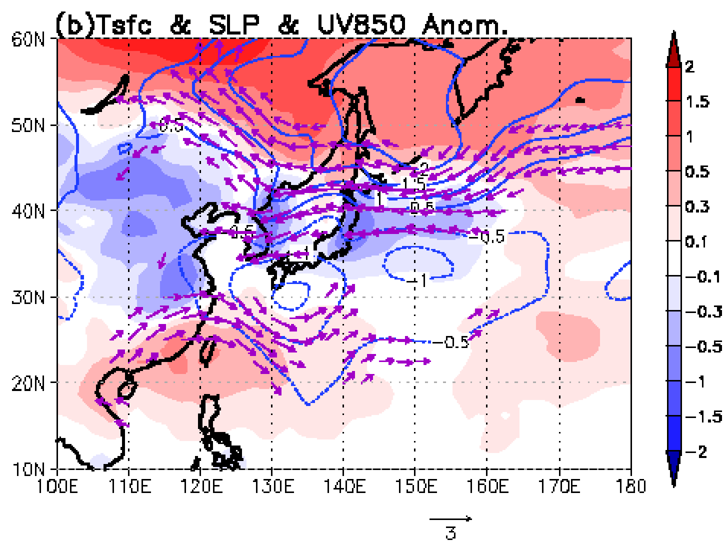




The regression pattern with the sea level pressure demonstrates negative correlation with temperature over the Okhotsk Sea.

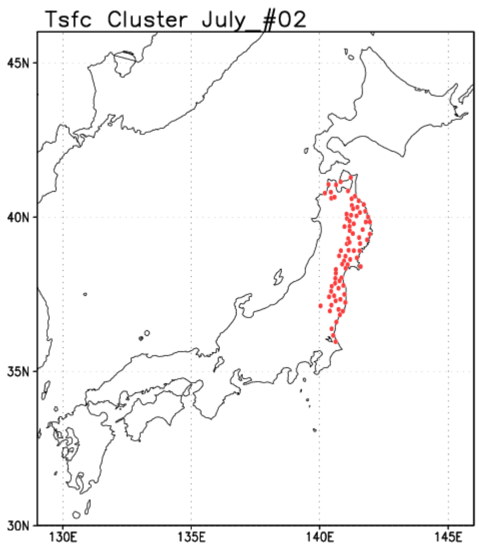
↗SLP regression with the surface temperature. Color indicate the contribution rate.

→The anomaly of the surface temperature (color) and SLP(contour) during Yamase.



This pattern is quite similar to the Yamase situation, which is characterized by the cold climate at Pacific side in association with appearance of Okhotsk High.

Linear regression: Ex.1 Tsfc #2



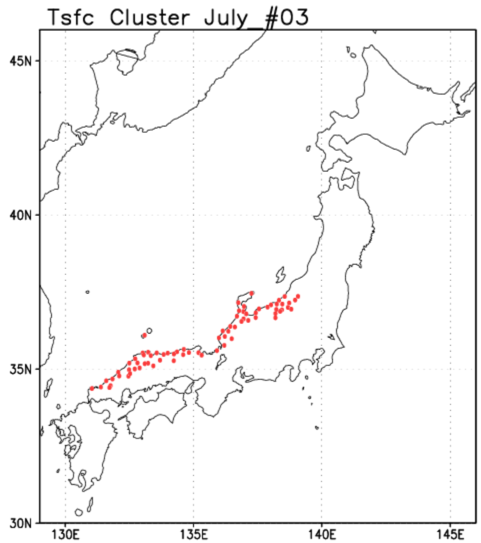
Relation with local geographical factors

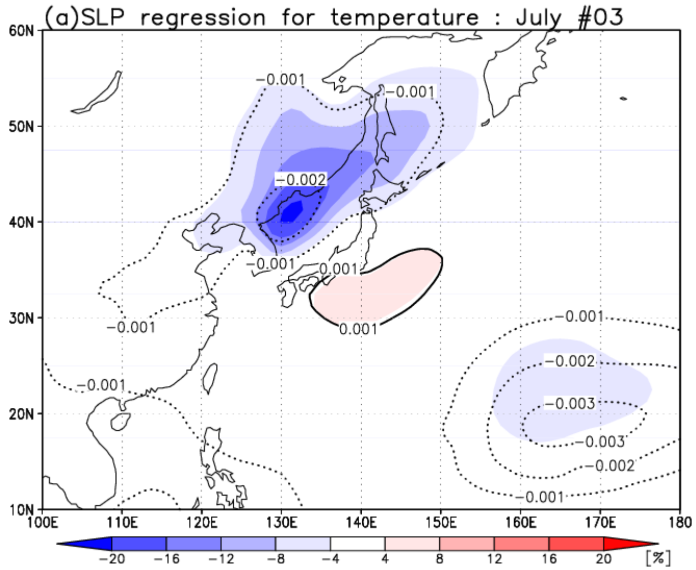
factors	Contribution[%]
Latitude	44.1
Longitude	26.2
Altitude	6.1
Distance to ocean	0.9

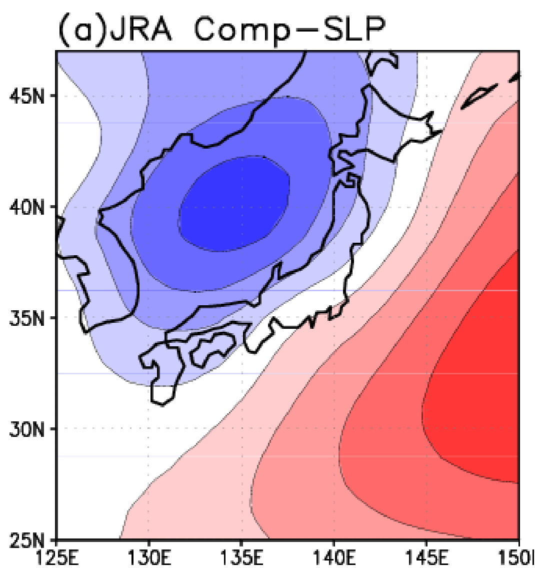
As other studies have mentioned, the several geographical factors show no small relation with the temperature as much as the large-scale circulation.

Linear regression: Ex.2 Tsfc #3

Relation with synoptic scale circulation







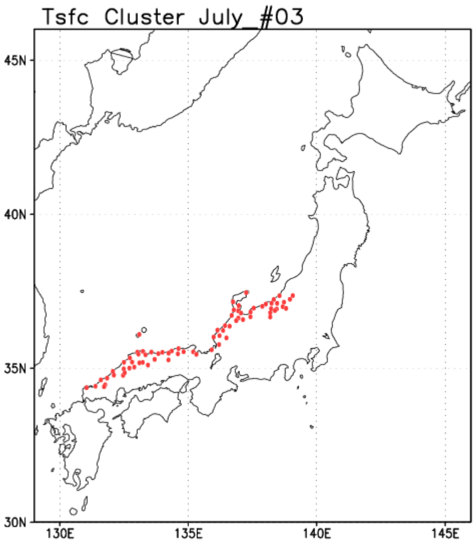
↗SLP regression with the surface temperature. Color indicate the contribution rate.

→The SLP pattern during foehn over the Toyama Plain.

The SLP regression pattern indicates the relation with the low pressure over the Japan Sea and warming at the stations of the cluster.

Similar pattern is often seen during foehn events over the coast of Japan Sea.

Linear regression: Ex.2 Tsfc #3



Relation with local geographical factors

factors	Contribution[%]
Latitude	9.5
Longitude	11.8
Altitude	71.7
Distance to ocean	0.3

The contribution rate of the relation between altitude is remarkably high in this case.

The relation with the variation of precipitation is more complicated…

Summary

- The variations of meteorological variables (temperature or precipitation) are affected by the various local geographical factors as much as the large-scale circulation. In association with this, the anisotropy of correlation distribution will change.

- It is preferable to consider the representativeness error when we compare the model results with different horizontal resolution. As one way of thinking, we consider the summation of weight in each grid as the index of representativeness error.

- Using model results and representativeness error, it is possible to construct the analysis with greater accuracy.

Thank you

nishizak@mri-jma.go.jp

Acknowledgement

This work was supported by the Global Environment Research Fund (S-5) of the Ministry of the Environment, Japan.