Clustering technique considering temporal coherence of ensemble members

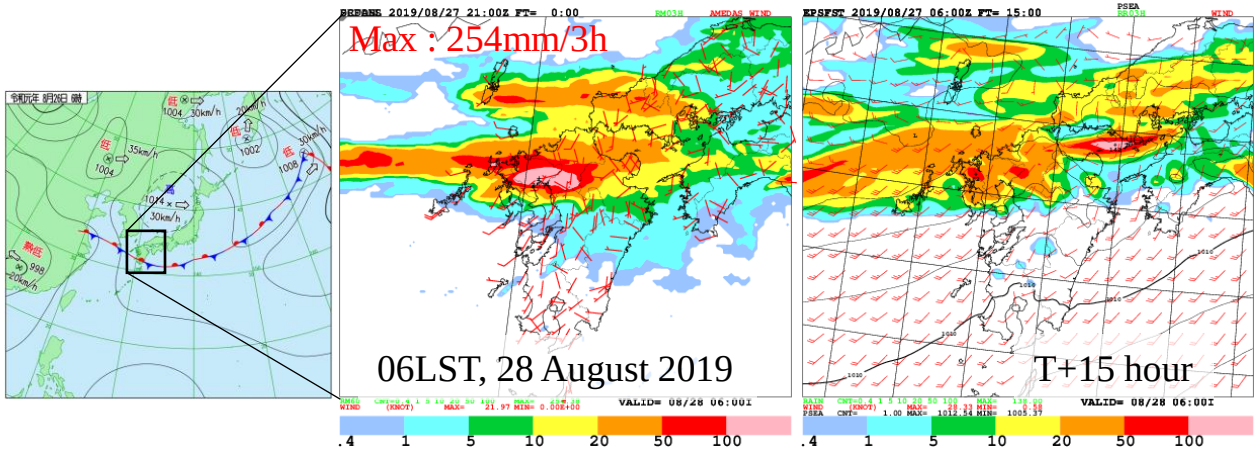
Kosuke Ono

Meteorological research institute, Japan meteorological agency

Regional model and EPS at JMA

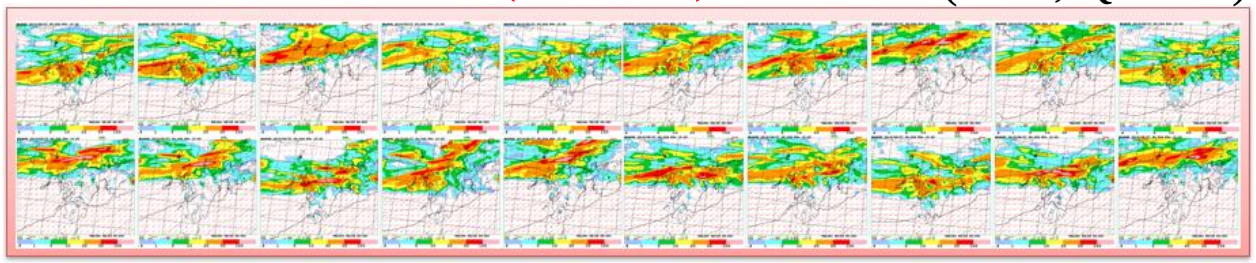
for short range forecast (~1 or 2 days ahead)

Observation **MSM**



- Initialized 8 times/day by 4DVAR
- Forecast length：78 or 39 hours
- dx = 5 km, 96 layers

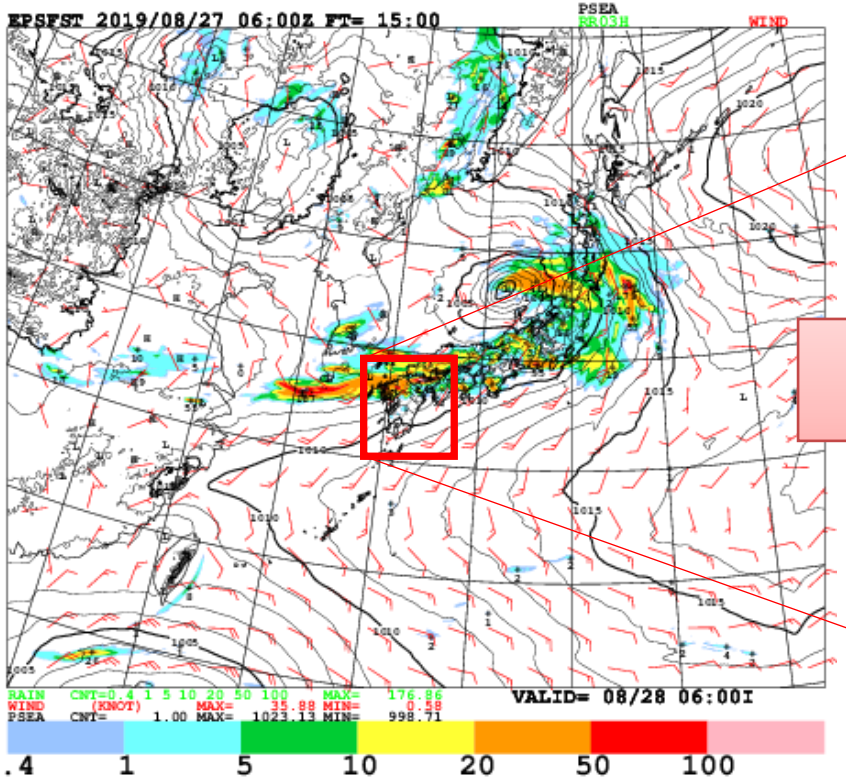
Mesoscale EPS (MEPS) Ono et al. (2021, QJRMS)



- ✓ 20 perturbed runs and control run (MSM)
- ✓ Initial and lateral boundary perturbations : SVs
  - ✓ SPPT has installed since March 2023
- ✓ Same model settings as MSM
  - ✓ dx=5km, 39 h forecasts, 4 time/day
- ✓ In operation since June 2019

JMA's Operational short-range forecast

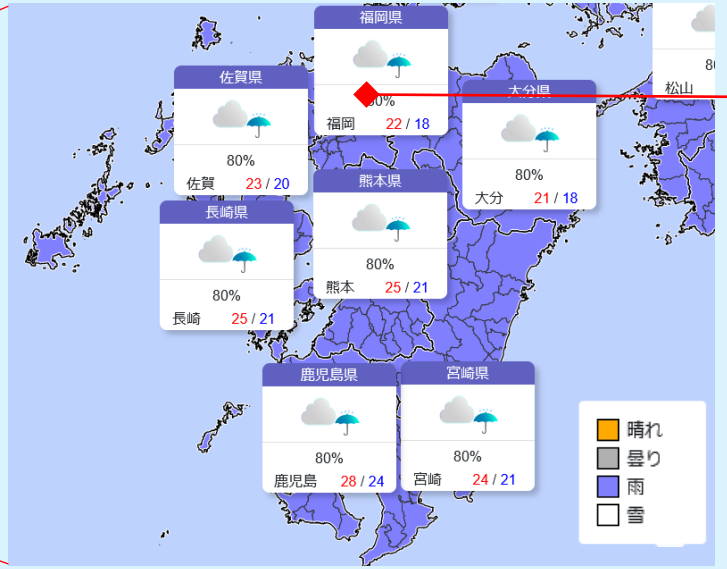
MSM forecast



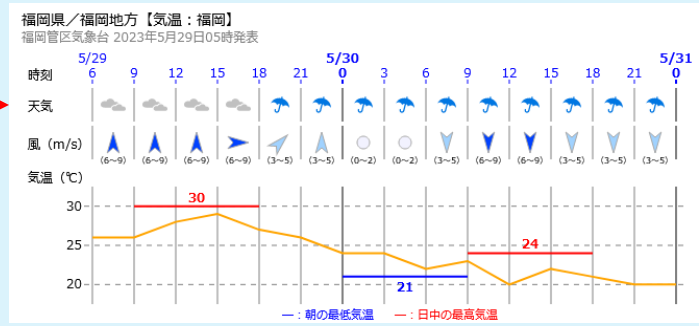
T+15, Initialized 06UTC 27 Aug. 2019

Operational short-range forecasts

Regional forecasts



Station forecasts

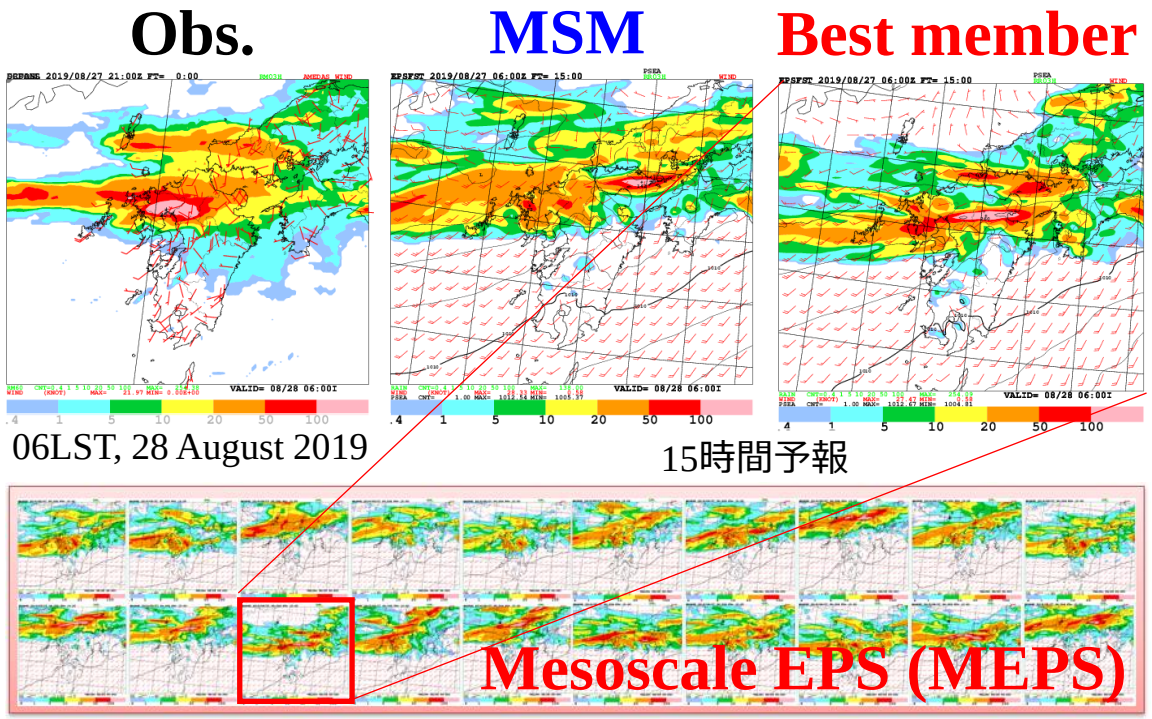


A specific weather secnario

Weather scenarios from MEPS > probabilities

Forecasters : Responsible for each prefecture (**Fixed region**)

Deterministic forecasts from MEPS

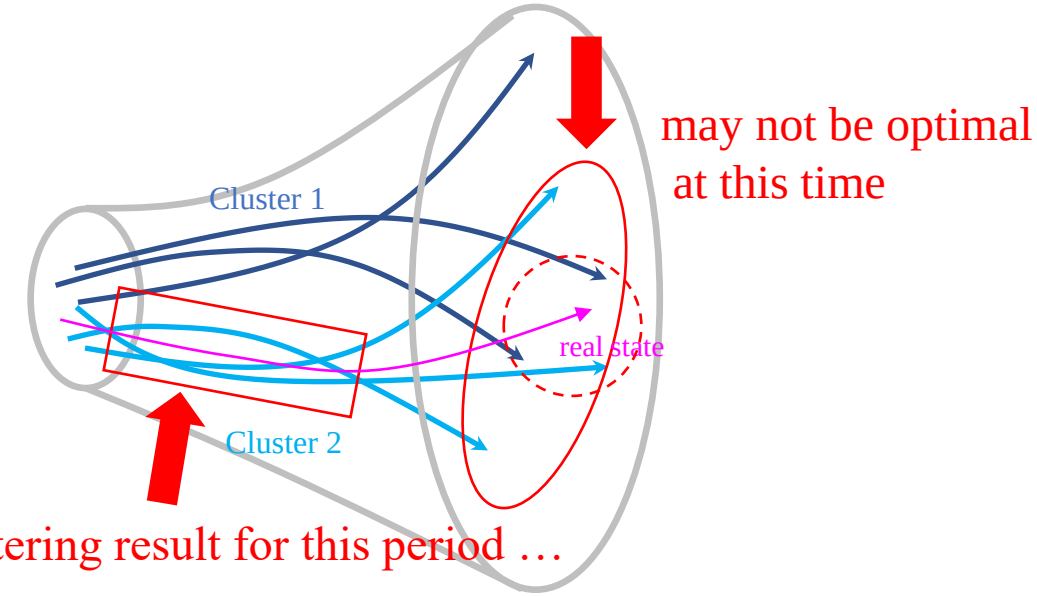


For operational forecasters

- 21 forecasts : too much information
- Reduce information : **Clustering**

Clusters for weather scenarios

- Intra-cluster members : change over time

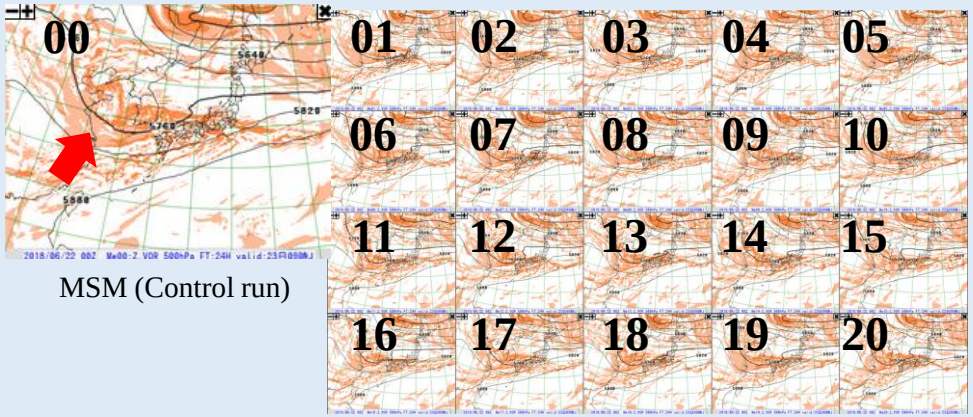


Purpose

- Summary of background
  - Forecasters are
    - ✓ Responsible for *a fixed domain and stations*
    - ✓ Issue forecasts as *a specific scenario*
      - ✓ Scenarios from MEPS
- Important points for clustering
  - ✓ Fixed domain during a forecast
  - ✓ Consider temporal coherency of intra-cluster members
- Purpose of this study
  - ➢ Procedures
  - ➢ Advantages

Clustering method

Target : **Geopotential height forecasts at 500 hPa**

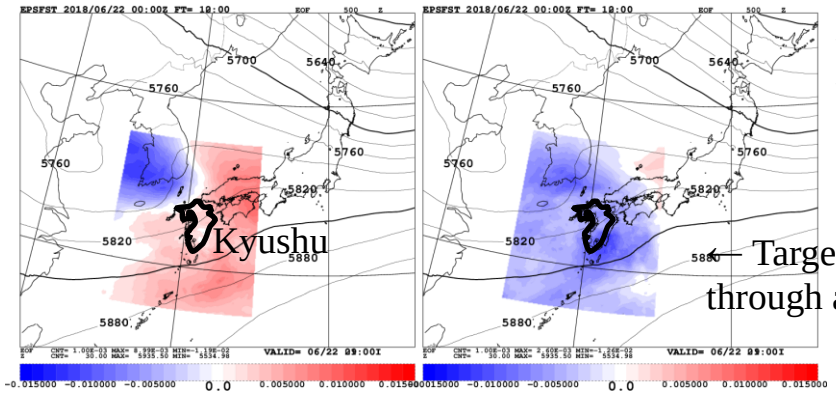


MSM (Control run)

Perturbed runs

T+36 (Initialized at 00 UTC 22 June, 2018)

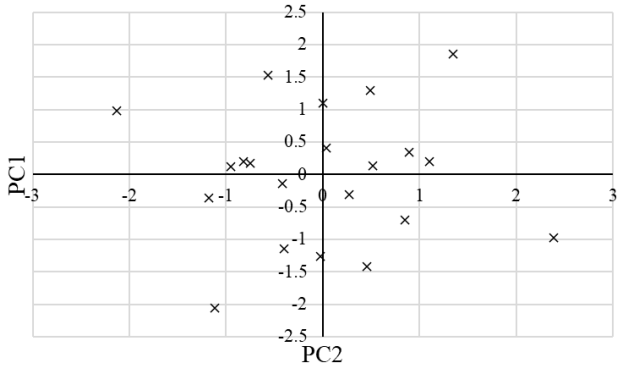
1. Principal component analysis



- ✓ Calculate covariance matrix
- ✓ Extract first two principal components

← Targeted region is fixed through a forecast (39 hours)

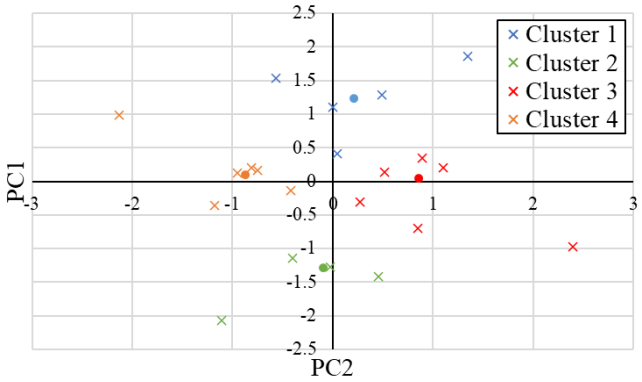
2. Project all members onto 2-D phase space



×: Each member

3. Clustering

Fuzzy c5means method (similar to k-means)



×: Each member

●: Cluster centroids

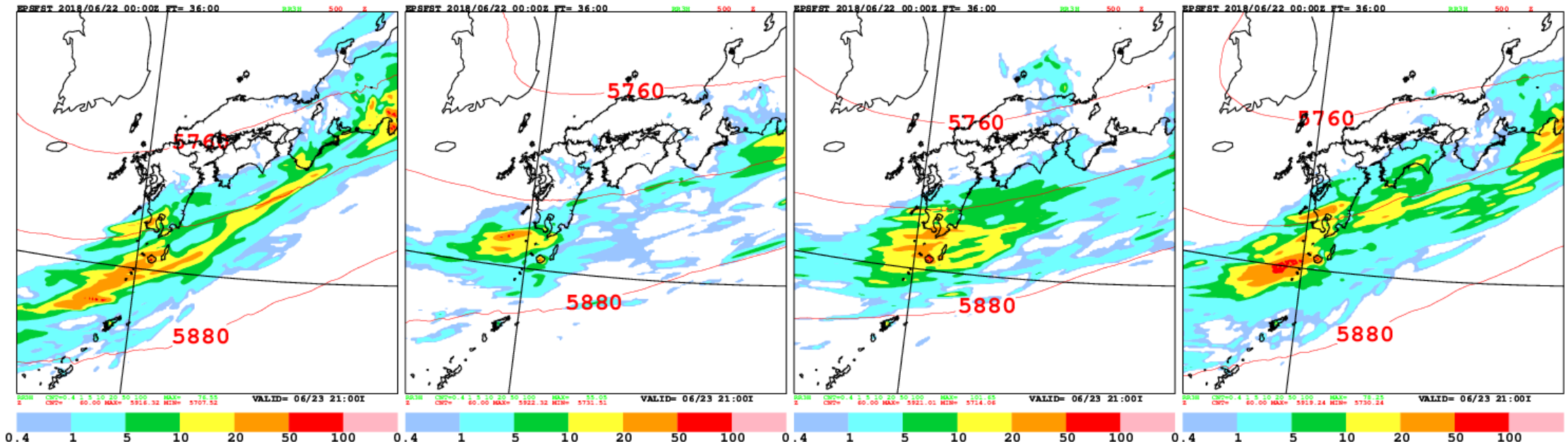
✓4 clusters

✓Cluster-mean used as a scenario

Clustering results at T+36

Initialized at 00 UTC 22 June, 2018

Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4



Geopotential height : Cluster mean
3 hour precipitation : Cluster mean with probability matching

- ✓ Different pressure trough scenarios
- ✓ Corresponding to trough, different precipitation scenarios forecasted

Clustering results every 3 hours

Geopotential height at 500 hPa (NO temporal connection)

Forecast lead time	Members in Cluster 1	Members in Cluster 2	Members in Cluster 3	Members in Cluster 4
0	01 06	11 16	00 03 04 05 07 09 10 12 18	02 08 13 14 15 17 19 20
3	00 01 05 09 12 18 20	03 04 17	02 08 10 11 15 16 19	06 07 13 14
6	00 02 07 13 14	08 10 11 15 16 19	03 04 05 12 17	01 06 09 18 20
9	01 05 06 09 17 18 20	00 02 07 13 14	03 04 12	08 10 11 15 16 19
12	00 06 07 13 18	01 02 04 05 09 20	10 11 12 14 15 19	03 08 16 17
15	08 15 16	00 02 05 06 10 18	07 11 12 13 14 19	01 03 04 09 17 20
18	05 06	07 11 12 13 14 18 19	00 01 02 03 04 08 09 10 17 20	15 16
21	00 02 07 08 11 12 13 14 18 19 20	01 03 09 10	15 16	04 05 06 17
24	00 02 11 12 13 18 19 20	01 03 08 09 10	04 05 06 17	07 14 15 16
27	01 02 03 08 09 10	07 11 13 18 19 20	14 15 16	00 04 05 06 12 17
30	02 14 15 16	00 04 05 06 12	07 11 13 18 19 20	01 03 08 09 10 17
33	02 14 15	04 05 19	07 11 13 16 18 20	00 01 03 06 08 09 10 12 17
36	04 05 19	09 14 15	02 07 11 13 16 18 20	00 01 03 06 08 10 12 17
39	09 10 14 15 17	00 01 03 06 08 12	04 05 19 20	02 07 11 13 16 18

temporally correlated

Note : Cluster numbers assigned randomly

Link clusters: Similarity of intra-cluster members

1. Compare clusters

Forecast lead time	Members in Cluster 1	Members in Cluster 2	Members in Cluster 3	Members in Cluster 4
21	00 02 07 08 11 12 13 14 18 19 20	01 03 09 10	15 16	04 05 06 17
24	00 02 11 12 13 18 19 20	01 03 08 09 10	04 05 06 17	07 14 15 16

2. If two or more candidates … Small members in previous time

Forecast lead time	Members in Cluster 1	Members in Cluster 2	Members in Cluster 3	Members in Cluster 4
21	00 02 07 08 11 12 13 14 18 19 20	01 03 09 10	15 16	04 05 06 17
24	00 02 11 12 13 18 19 20	01 03 08 09 10	04 05 06 17	07 14 15 16

3. Rearrange each cluster

Forecast lead time	Members in cluster 1	Members in cluster 2	Members in cluster 3	Members in cluster 4
21	04 05 06 17	15 16	01 03 09 10	00 02 07 08 11 12 13 14 18 19 20
24	04 05 06 17	07 14 15 16	01 03 08 09 10	00 02 11 12 13 18 19 20

These procedures conducted from (T+3 and 0) to (T+39 and T+36)

Clusters with temporal connection

Forecast lead time	Members in cluster 1	Members in cluster 2	Members in cluster 3	Members in cluster 4
0	01 06	11 16	00 03 04 05 07 09 10 12 18	02 08 13 14 15 17 19 20
3	06 07 13 14	03 04 17	00 01 05 09 12 18 20	02 08 10 11 15 16 19
6	00 02 07 13 14	03 04 05 12 17	01 06 09 18 20	08 10 11 15 16 19
9	00 02 07 13 14	03 04 12	01 05 06 09 17 18 20	08 10 11 15 16 19
12	00 06 07 13 18	03 08 16 17	01 02 04 05 09 20	10 11 12 14 15 19
15	00 02 05 06 10 18	08 15 16	01 03 04 09 17 20	07 11 12 13 14 19
18	05 06	15 16	00 01 02 03 04 08 09 10 17 20	07 11 12 13 14 18 19
21	04 05 06 17	15 16	01 03 09 10	00 02 07 08 11 12 13 14 18 19 20
24	04 05 06 17	07 14 15 16	01 03 08 09 10	00 02 11 12 13 18 19 20
27	00 04 05 06 12 17	14 15 16	01 02 03 08 09 10	07 11 13 18 19 20
30	00 04 05 06 12	02 14 15 16	01 03 08 09 10 17	07 11 13 18 19 20
33	04 05 19	02 14 15	00 01 03 06 08 09 10 12 17	07 11 13 16 18 20
36	04 05 19	09 14 15	00 01 03 06 08 10 12 17	02 07 11 13 16 18 20
39	04 05 19 20	09 10 14 15 17	00 01 03 06 08 12	02 07 11 13 16 18

Characteristics

- ✓ Same members compose same cluster for a certain period
- ✓ Number of intra-cluster members changes with time

Provide robustness of scenarios

Forecast lead time	Members in cluster 1	Members in cluster 2	Members in cluster 3	Members in cluster 4	**Sum** of number of members which moved between clusters
0	01 06	11 16	00 03 04 05 07 09 10 12 18	02 08 13 14 15 17 19 20	
3	06 07 13 14	03 04 17	00 01 05 09 12 18 20	02 08 10 11 15 16 19	11
6	00 02 07 13 14	03 04 05 12 17	01 06 09 18 20	08 10 11 15 16 19	5
9	00 02 07 13 14	03 04 12	01 05 06 09 17 18 20	08 10 11 15 16 19	2
12	00 06 07 13 18	03 08 16 17	01 02 04 05 09 20	10 11 12 14 15 19	9
15	00 02 05 06 10 18	08 15 16	01 03 04 09 17 20	07 11 12 13 14 19	8
18	05 06	15 16	00 01 02 03 04 08 09 10 17 20	07 11 12 13 14 18 19	5
21	04 05 06 17	15 16	01 03 09 10	00 02 07 08 11 12 13 14 18 19 20	6
24	04 05 06 17	07 14 15 16	01 03 08 09 10	00 02 11 12 13 18 19 20	3
27	00 04 05 06 12 17	14 15 16	01 02 03 08 09 10	07 11 13 18 19 20	4
30	00 04 05 06 12	02 14 15 16	01 03 08 09 10 17	07 11 13 18 19 20	2
33	04 05 19	02 14 15	00 01 03 06 08 09 10 12 17	07 11 13 16 18 20	5
36	04 05 19	09 14 15	00 01 03 06 08 10 12 17	02 07 11 13 16 18 20	2
39	04 05 19 20	09 10 14 15 17	00 01 03 06 08 12	02 07 11 13 16 18	3

Large sum

Small sum

Sums	Intra-cluster members	Scenarios (cluster-mean)
Small sums	Coherent	Robust with time
Large sums	Diversive	NOT robust with time

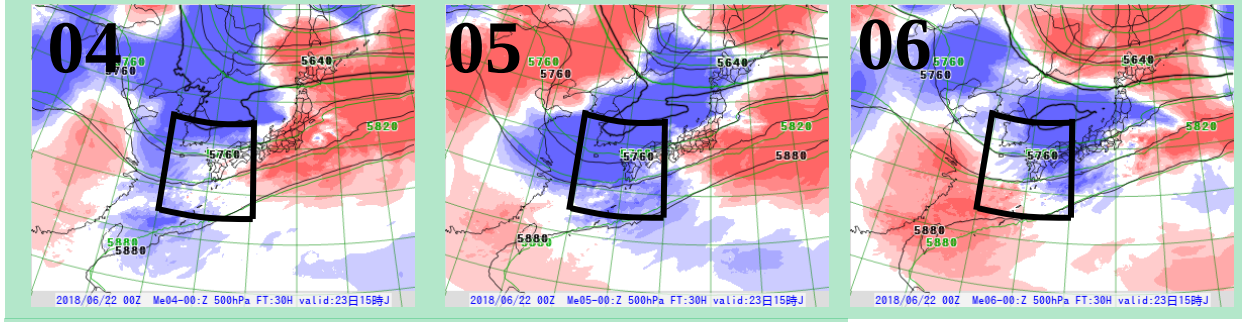
Useful for forecasters to use clusters

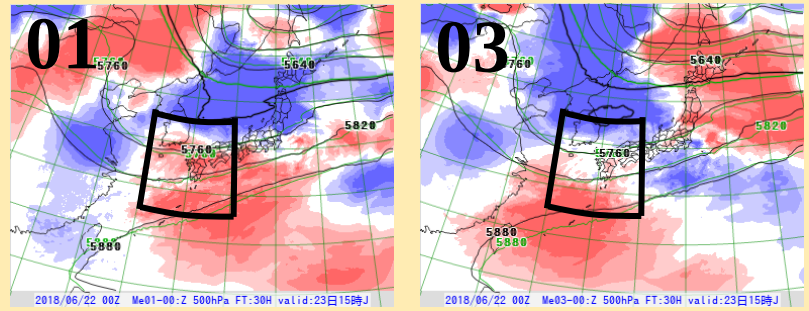
Optimal intra-cluster members

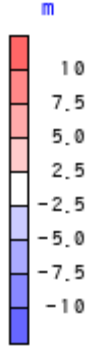
Perturbations (Perturbed run – control run) for geopotential height at 500 hPa

Cluster 1 **Cluster 3**

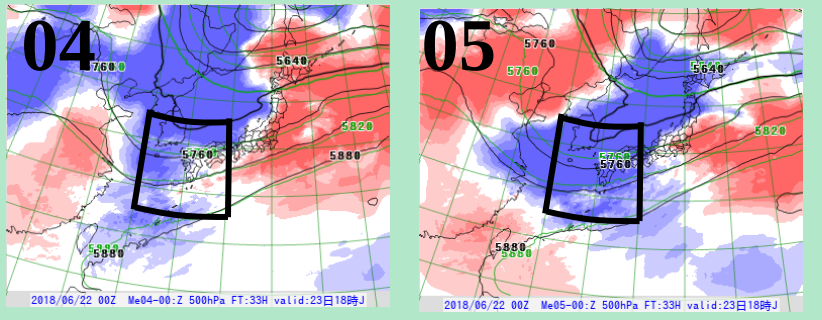
T+30h

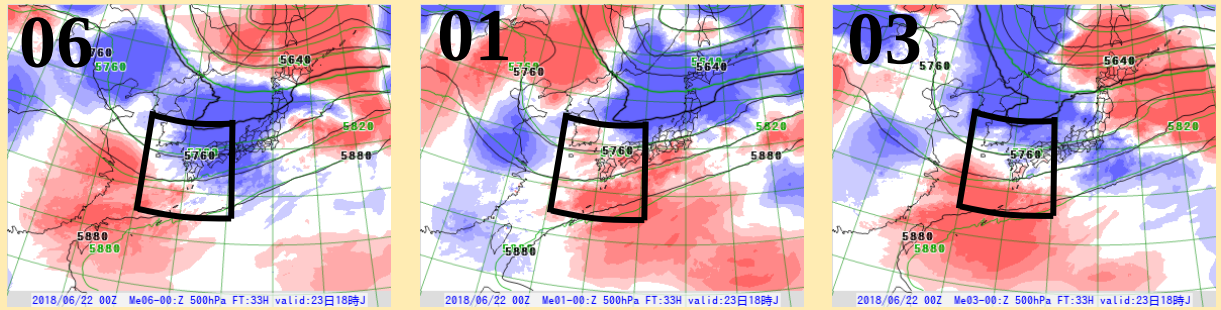






T+33h





negative pattern

negative and positive

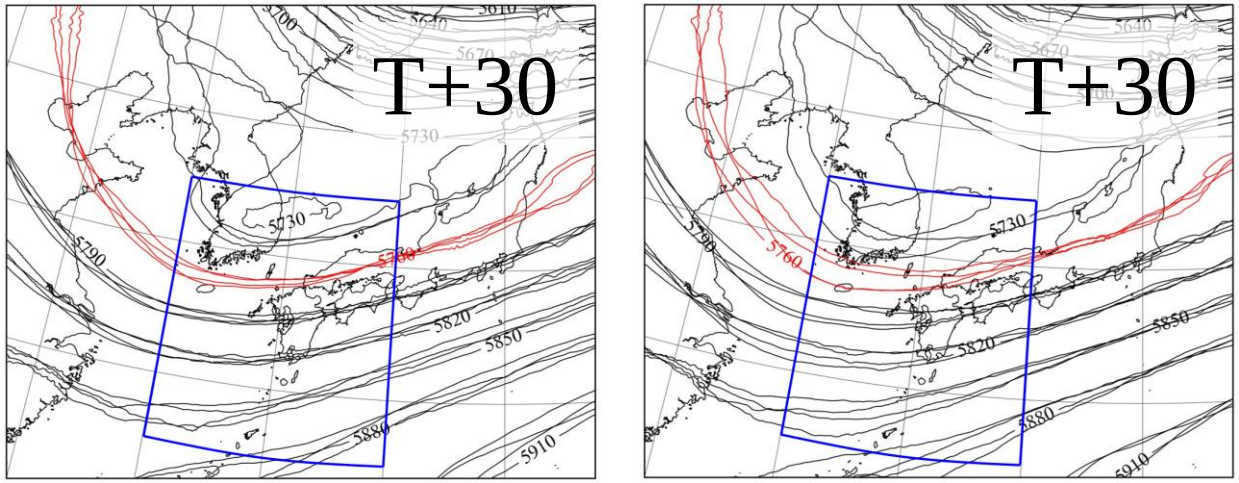
- ✓ Accommodate the time-varying perturbation structure,
- ✓ Provides **well-divided scenarios** for forecasters in charge of a fixed region

Well-divided scenarios

Forecast lead time	Members in cluster 1	Members in cluster 2	Members in cluster 3	Members in cluster 4
0	01 06	11 16	00 03 04 05 07 09 10 12 18	02 08 13 14 15 17 19 20
3	06 07 13 14	03 04 17	00 01 05 09 12 18 20	02 08 10 11 15 16 19
6	00 02 07 13 14	03 04 05 12 17	01 06 09 18 20	08 10 11 15 16 19
9	00 02 07 13 14	03 04 12	01 05 06 09 17 18 20	08 10 11 15 16 19
12	00 06 07 13 18	03 08 16 17	01 02 04 05 09 20	10 11 12 14 15 19
15	00 02 05 06 10 18	08 15 16	01 03 04 09 17 20	07 11 12 13 14 19
18	05 06	15 16	00 01 02 03 04 08 09 10 17 20	07 11 12 13 14 18 19
21	04 05 06 17	15 16	01 03 09 10	00 02 07 08 11 12 13 14 18 19 20
24	04 05 06 17	07 14 15 16	01 03 08 09 10	00 02 11 12 13 18 19 20
27	00 04 05 06 12 17	14 15 16	01 02 03 08 09 10	07 11 13 18 19 20
30	00 04 05 06 12	02 14 15 16	01 03 08 09 10 17	07 11 13 18 19 20
33	04 05 19	02 14 15	00 01 03 06 08 09 10 12 17	07 11 13 16 18 20
36	04 05 19	09 14 15	00 01 03 06 08 10 12 17	02 07 11 13 16 18 20
39	04 05 19 20	09 10 14 15 17	00 01 03 06 08 12	02 07 11 13 16 18

✓ Intra-cluster members : different between T+12 and 30

Spaghetti diagrams of 4 clusters for geopotential height at 500 hPa



Using clustering results at T+12

Summary

- **New clustering technique**
  - Forecasters' requests : Weather scenarios from MEPS
    - ➢Fixed region and 3-hour clustering
    - ➢Temporally connected
  - Provides
    - ➢Robustness of clustering scenarios
    - ➢Well divided scenarios
- **More details**
  - (Not shown) Improvement from MSM on RMSE for geopotential height
    - See Ono (2023, WAF)

Ongoing works and future tasks

- **Ongoing works**
  - Cluster using various elements for severe weather cases
    - Except for 500 hPa geopotential height
  - Select best scenario
    - Check my poster presentation
- **Future tasks**
  - Change the number of clusters with time
  - (Hopefully, select best member)