A comparative study on the creation of multi-satellite SST ensemble using OI and BMA.

Department of Spatial Information Engineering
Pukyong National Univ.(South Korea)
The doctor's course student
Kwangjin Kim

PUKYONG NATIONAL UNIVERSITY

GeoComputing Pukyong National University

Contents

1. **Background & Purpose**

2. **Methods**
   - BMA
   - EM Algorithm

3. **Ensemble test**
   - Ensemble with BMA
   - Ensemble with OI(ongoing)

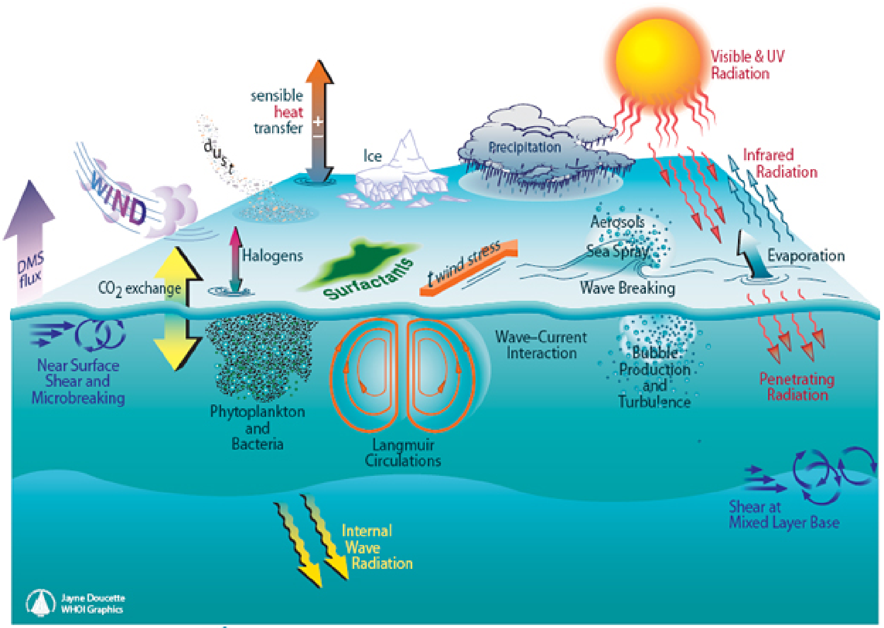
4. **Summary & future work**

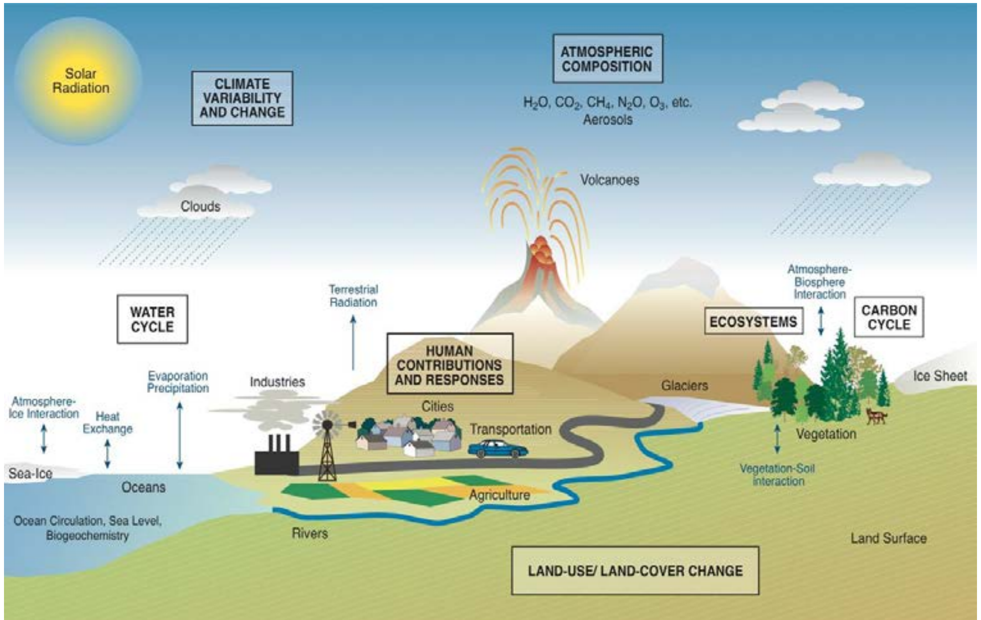
PUKYONG NATIONAL UNIVERSITY

GeoComputing Pukyong National University

1. Background & Purpose

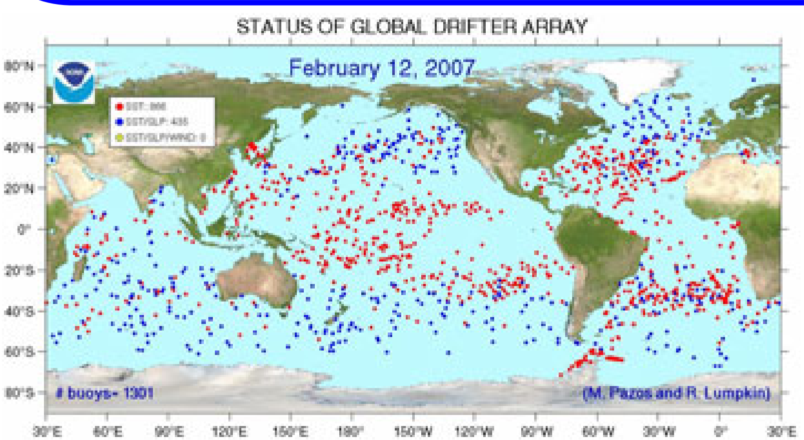
- Sea surface temperature (SST) is an important parameter in understanding atmosphere–ocean circulation processes and monitoring global climate change
- And it also has many applications in marine science and numerical weather prediction.
- So, measuring accurate SST is demanded.

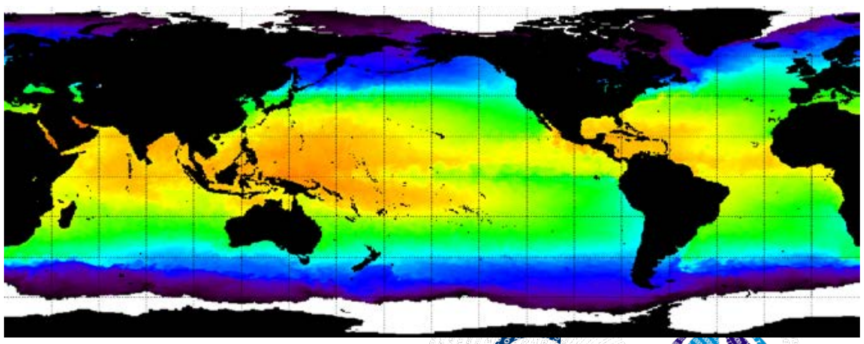




1. Background & Purpose

- While in situ measurements of SST are generally more accurate than irregular point-based observations, satellite remote sensing can provide a spatially continuous and consistent dataset.
- The satellite SST product has uncertainty, because the value is calculated through sensor and algorithm.
- In addition, different satellite products have different amount of uncertainty in the use of different sensors and algorithms.
- To reduce the uncertainty, many studies on climate forecast model and assimilation have been used model ensemble.

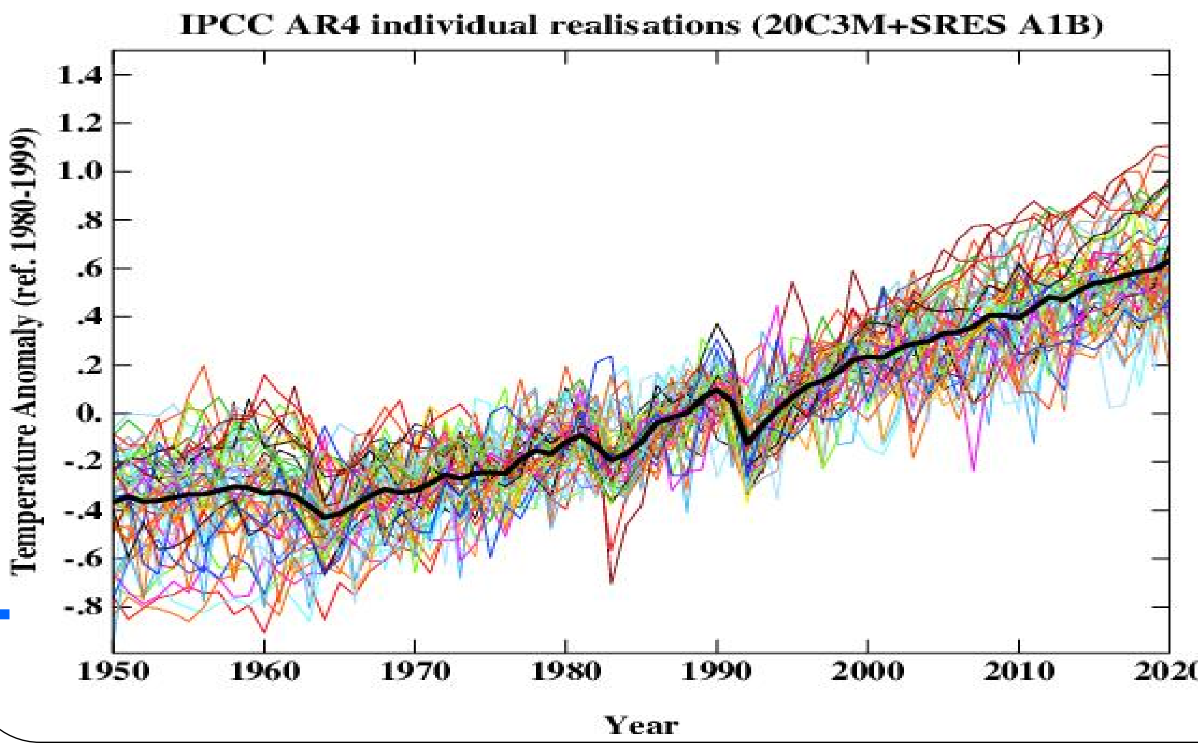


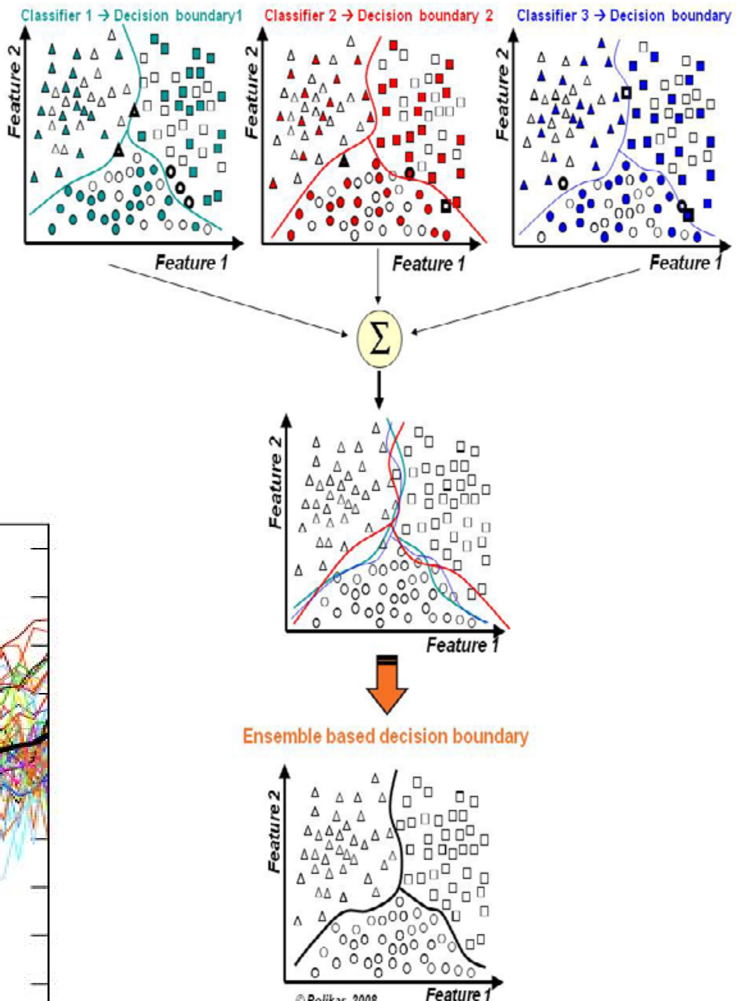


1. Background & Purpose

What is the advantage of the ensemble?

- Reliability of the data
  Multi- models or –products can improve the reliability of the data than when using only one data.
- Improve data accuracy
  It is possible to improve the accuracy through applying conventional statistical techniques.





Pukyong National University

1. Background & Purpose

- General ensemble method can not consider the difference in the size of the uncertainty about reference data.
- So, ensemble method considered about the different size of uncertainty is demanded.
- Bayesian model averaging (BMA) can be an alternative to the uncertainty problem by conditioning each ensemble member using the posterior probability for weighting scheme.
- Many BMA applications have been developed for use in climate forecasting by general circulation models (GCMs)
- But the BMA ensemble has not yet been applied to any satellite products, including SST.
- Optimal interpolation has been used with formal synthesis method in many organization. And the ability has been known as good.

Purpose of study

- The match-up database of reference data, satellite data(radiometer as MODIS and AVHRR and micrometer as AMSR) will be stacked with same temporal-spatial.

- The BMA ensemble and OI are carried out using the constructed match-up DB.

- The result of each method will be compared and validated various aspect like as bias statistics, bias spatial distribution and the cost for calculation.

Ensemble product using BMA

VS

Ensemble product using OI

Bayesian Model Averaging(BMA)

- Bayesian Model Averaging(BMA) is one of the model averaging methods using the probability density function(PDF).
- BMA, the fundamentals of which are provided in the literature, provides a coherent mechanism to account for such uncertainty.
- In the BMA, posterior probability is used as weight.
- However, the calculation of the posterior PDF is so difficult.

The law of total probability

$$\mathrm{p(y)} = \sum_{k=1}^{K} p(y|M_k)\, p(M_k|D)$$

$$\mathrm{P}(\mathrm{y}|f_1, \dots, f_K) = \sum_{k=1}^{K} w_k\, g(y|\widetilde{f_k})$$

$$\mathrm{E}[y_{st}|f_{1st}, \dots, f_{Kst}] = \sum_{K=1}^{K} w_k \widetilde{f_{Kst}}$$

$\sim \mathrm{N}(a_k + b_k M_k, \sigma^2)$

- The distribution of the errors for SST is known to follow normal distribution → The erorr of bias-corrected prediction is also following normal distribution.
- Due to the difficulty of estimating the variance of PDF, it is nearly impossible to estimate the posterior probability for the entire data.

EM(Expectation-Maximization) Algorithm

- EM algorithm estimate the information of hidden or missing data using the Maximum Likelihood Estimator.
- In this study, EM algorithm is used to estimate the hidden data(PDF of prediction). That is, mean $(a_k + b_k f_k)$ and variance(σ^2) are estimated.
- The mean and variance are updated through iterated process of E step and M step.

Expectation step

J th estimated PDF

$$\hat{z}_{kst}^{(j)} = \frac{w_k^{(j-1)} g(y_{st} \mid \tilde{f}_{kst}, \sigma_k^{2(j-1)})}{\sum_{i=1}^{K} w_i^{(j-1)} g(y_{st} \mid \tilde{f}_{ist}, \sigma_i^{2(j-1)})}$$

Iteration

Maximization step

weight

$$w_k^{(j)} = \frac{1}{n} \sum_{s,t} \hat{z}_{kst}^{(j)}$$

variance

$$\sigma_k^{2(j)} = \frac{\sum_{s,t} \hat{z}_{kst}^{(j)} (y_{st} - \tilde{f}_{kst})^2}{\sum_{s,t} \hat{z}_{kst}^{(j)}}$$

2. Methods

Input : Cluster number k, a database, Stopping tolerance $\varepsilon (> 0)$

Output : A set of k clusters with weight that maximize Log-likelihood function.

(1) Expectation Step

For each database record x,

Compute the membership probability of x in each cluster h = 1,…,k.

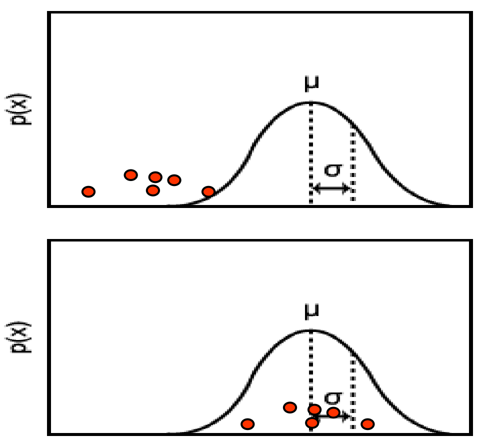
(2) Maximization Step

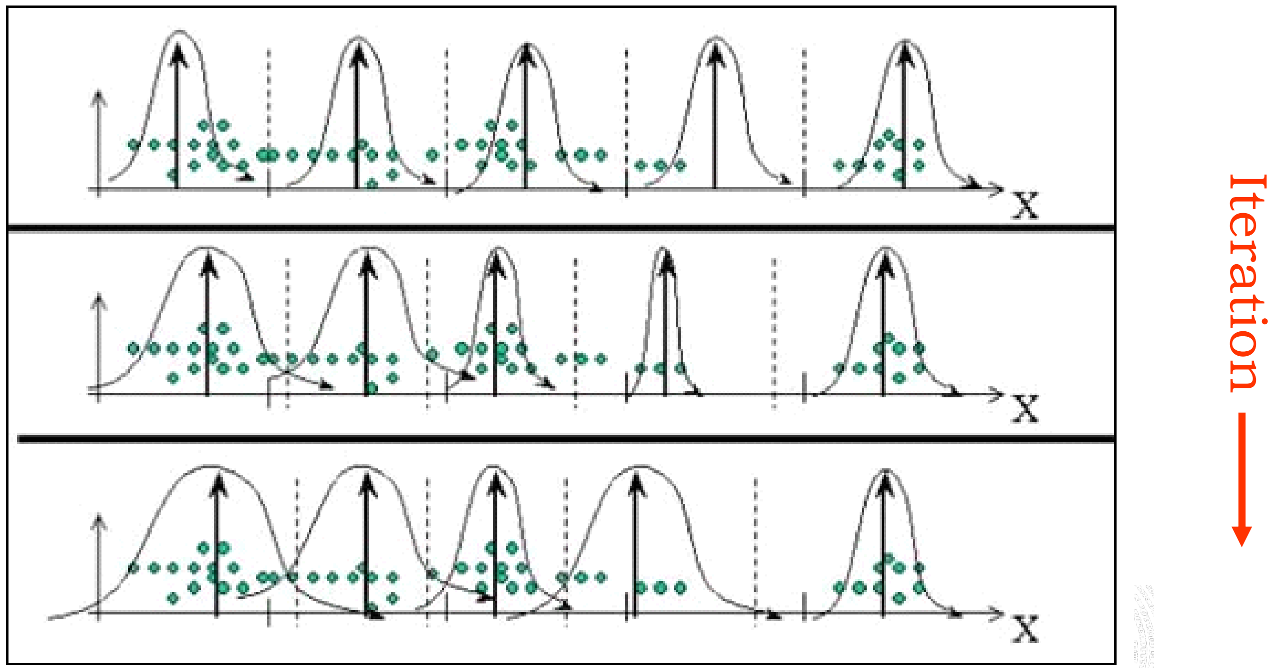
Update mixture model parameter (probability weight)

(3) Stopping criteria

If stop criteria is satisfied stop

Else set j = j+1 and goto (1)





2. Methods

Optimal Interpolation(OI)

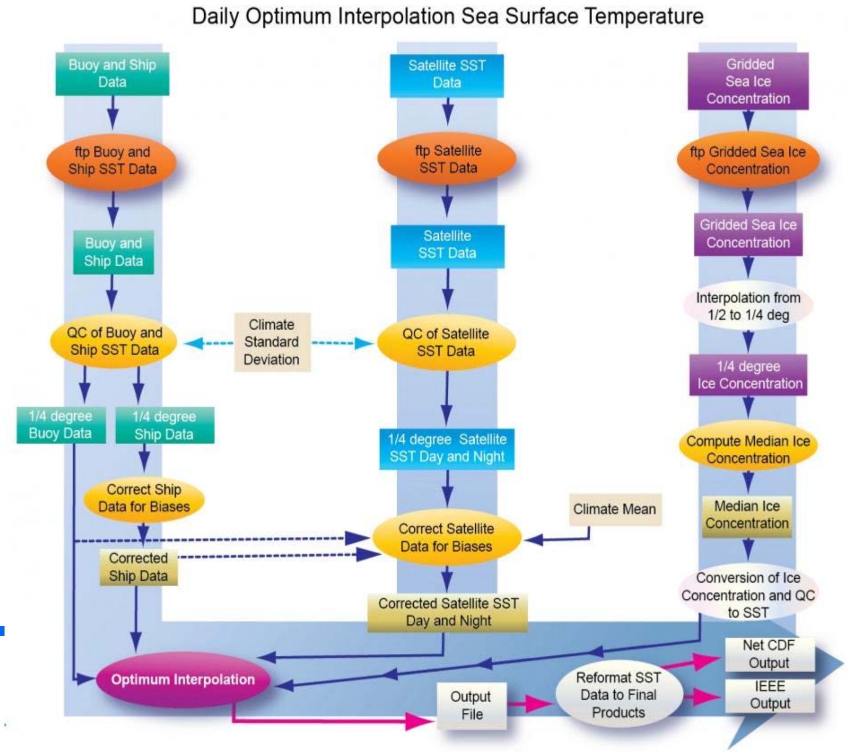
- Optimal Interpolation(OI) is a commonly used and fairly simple but powerful method of interpolation and data assimilation.
- The analysis is obtained in the form of the Best Linear Unbiased Estimator(BLUE).
- Because of powerful ability, OI has been used to SST in various countries and institutions.

$$r_k = \sum_{i=1}^{N} w_i q_i$$

$$\sum_{i=1}^{N} M_{ij} w_{ik} = < \pi_j \pi_k >$$

$$M_{ij} = < \pi_i \pi_j > + \epsilon_i \epsilon_j < \beta_i \beta_j >$$

$$< \pi_i \pi_j > = \exp \left[\frac{-(x_i - x_j)^2}{\lambda^2} \right]$$



3. Ensemble test – data for test

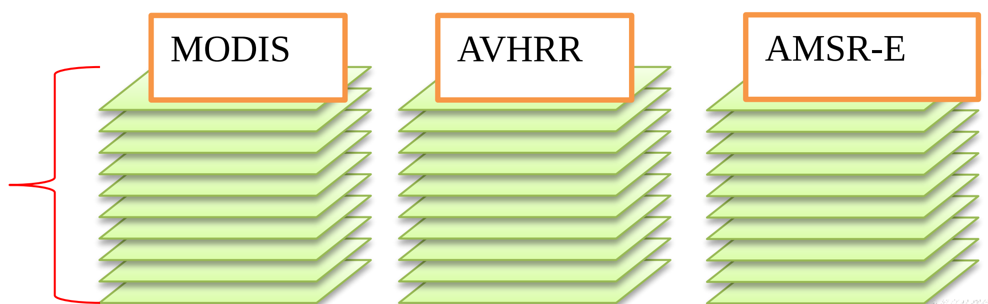
Product: Sea Surface Temperature(SST)

Reference data: AATSR/Envisat

Background: ECMWF reanalysis

Ensemble members: (1)MODIS/Aqua, (2) AVHRR/NOAA, (3)AMSR_E/Aqua

Study area : latitude -60° – 60° longitude -180° - 180°

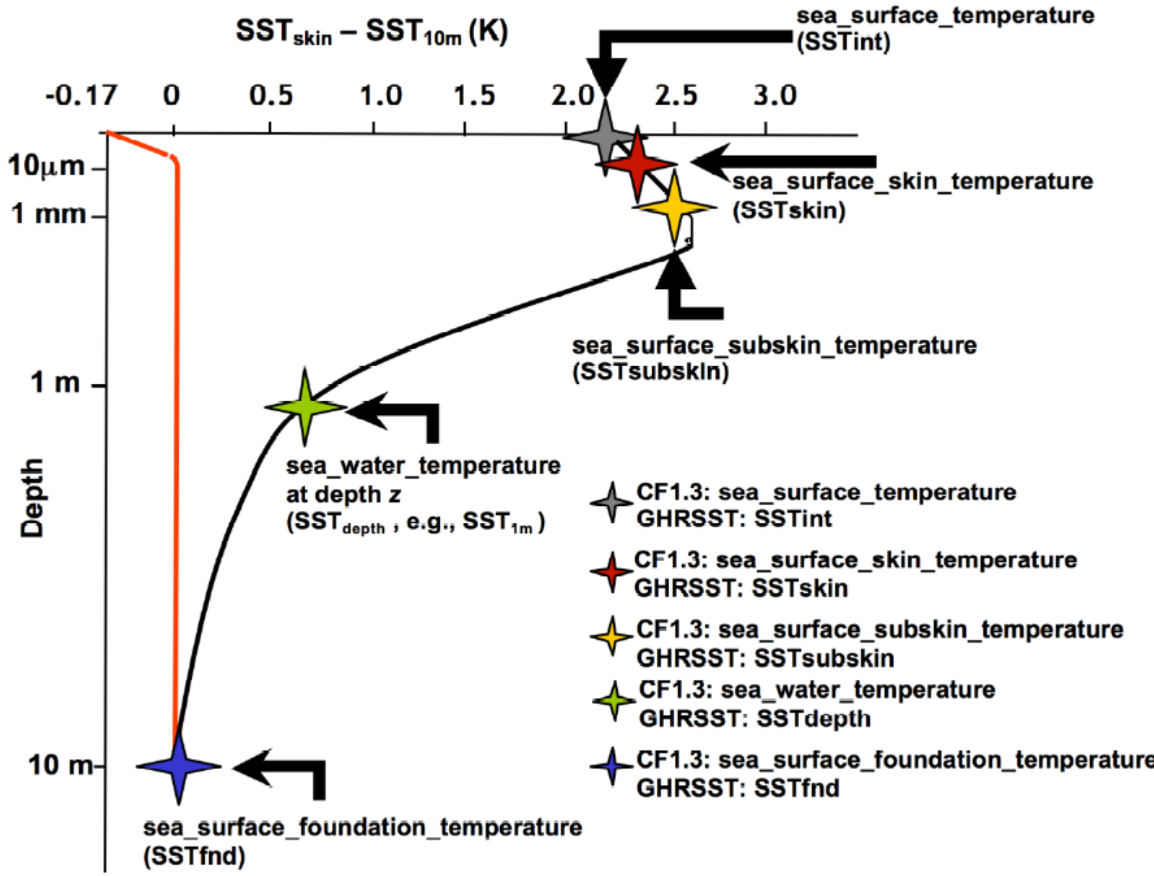


2006/1
~ 2008/12
(monthly daytime SST)

PUKYONG NATIONAL UNIVERSITY

GeoComputing Pukyong National University

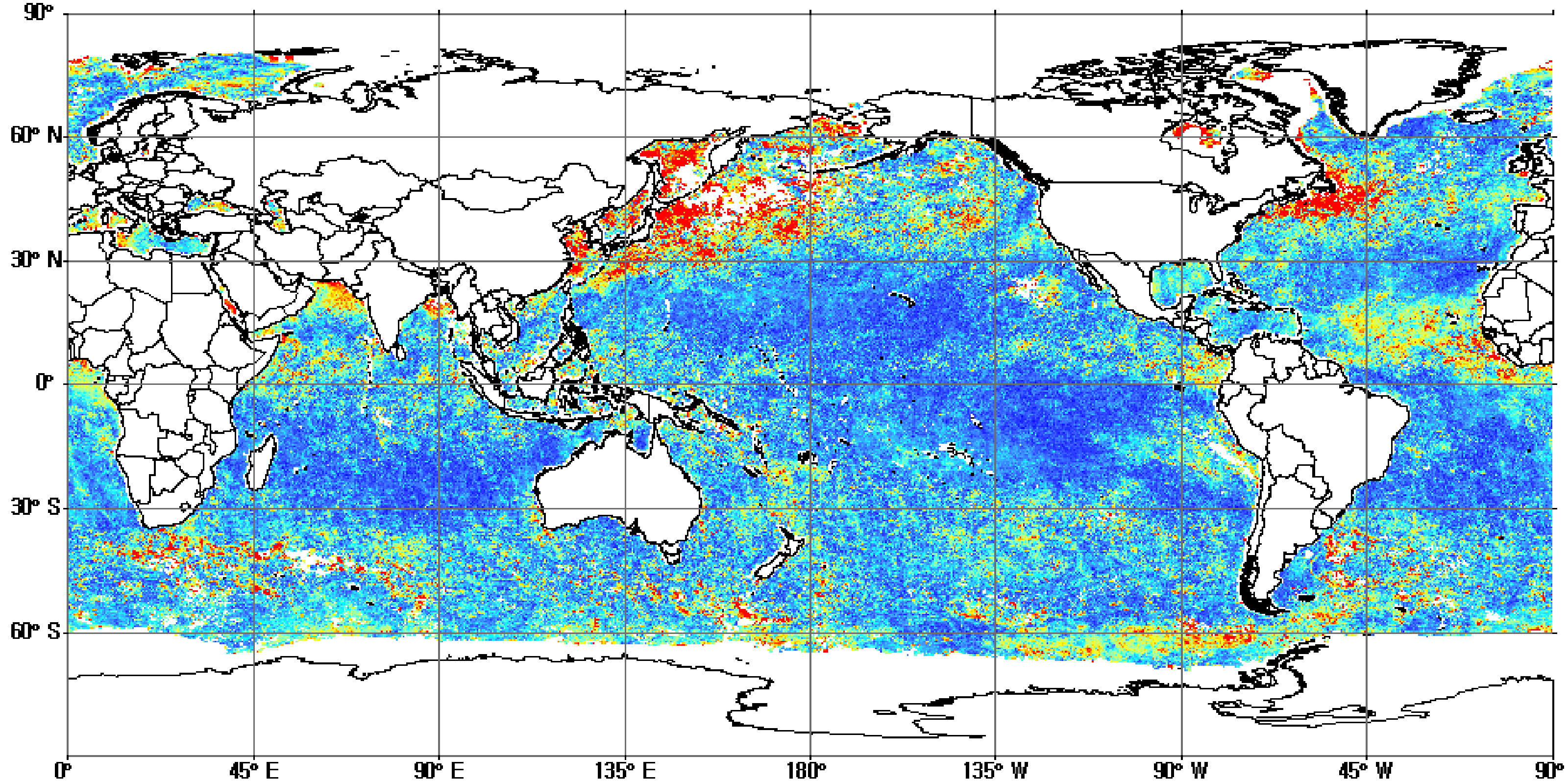
Ensemble member data → MODIS, AVHRR, AMSR-E



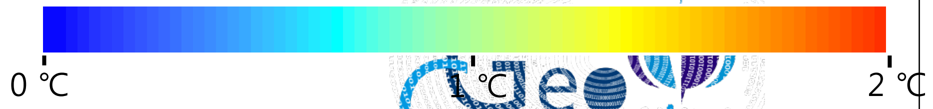
1. Night-time or strong winds profile in red
2. Daytime, strong solar radiation, and light winds profile in black

- SST measured by the Radiometer is called with skin SST.
- SST measured by the microwave is called with sub-skin SST.
- Skin temperature products and sub-skin temperature was used together with ensemble member for BMA ensemble. Because there is little difference in the depth and temperature.
- In this study, the SST product of MODIS/Aqua, AVHRR/NOAA and AMSR-E/Aqua was used as ensemble member.

Difference in ensemble members in same month(June 2006)



MAX(MODIS, AVHRR, AMSR-E) - MIN(MODIS, AVHRR, AMSR-E)

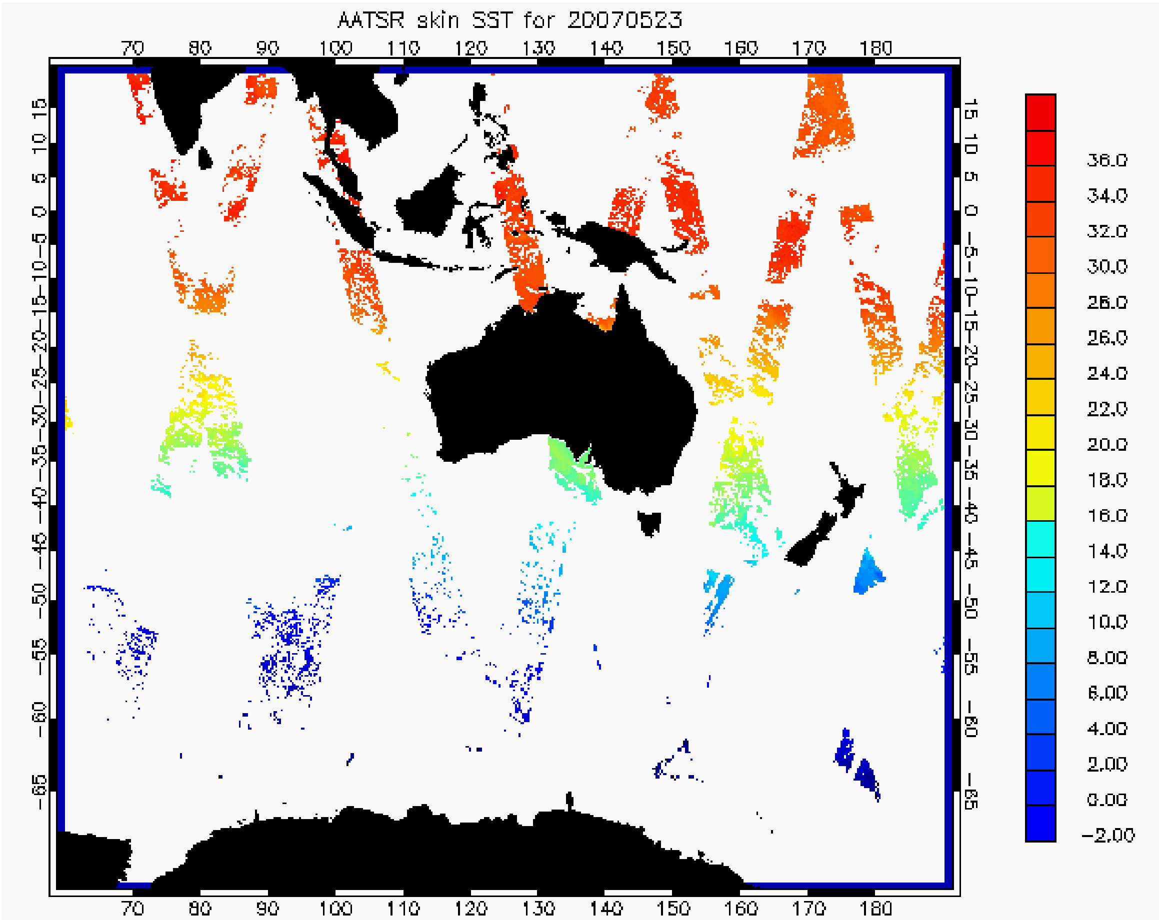


PUKYONG NATIONAL UNIVERSITY

Geo Computing Pukyong National University

3. Ensemble test – data for test

Reference data → AATSR



- It is difficult to obtain global in-situ SST. And for using in-situ, data like as buoy, transportation is needed because it detect the foundation SST.

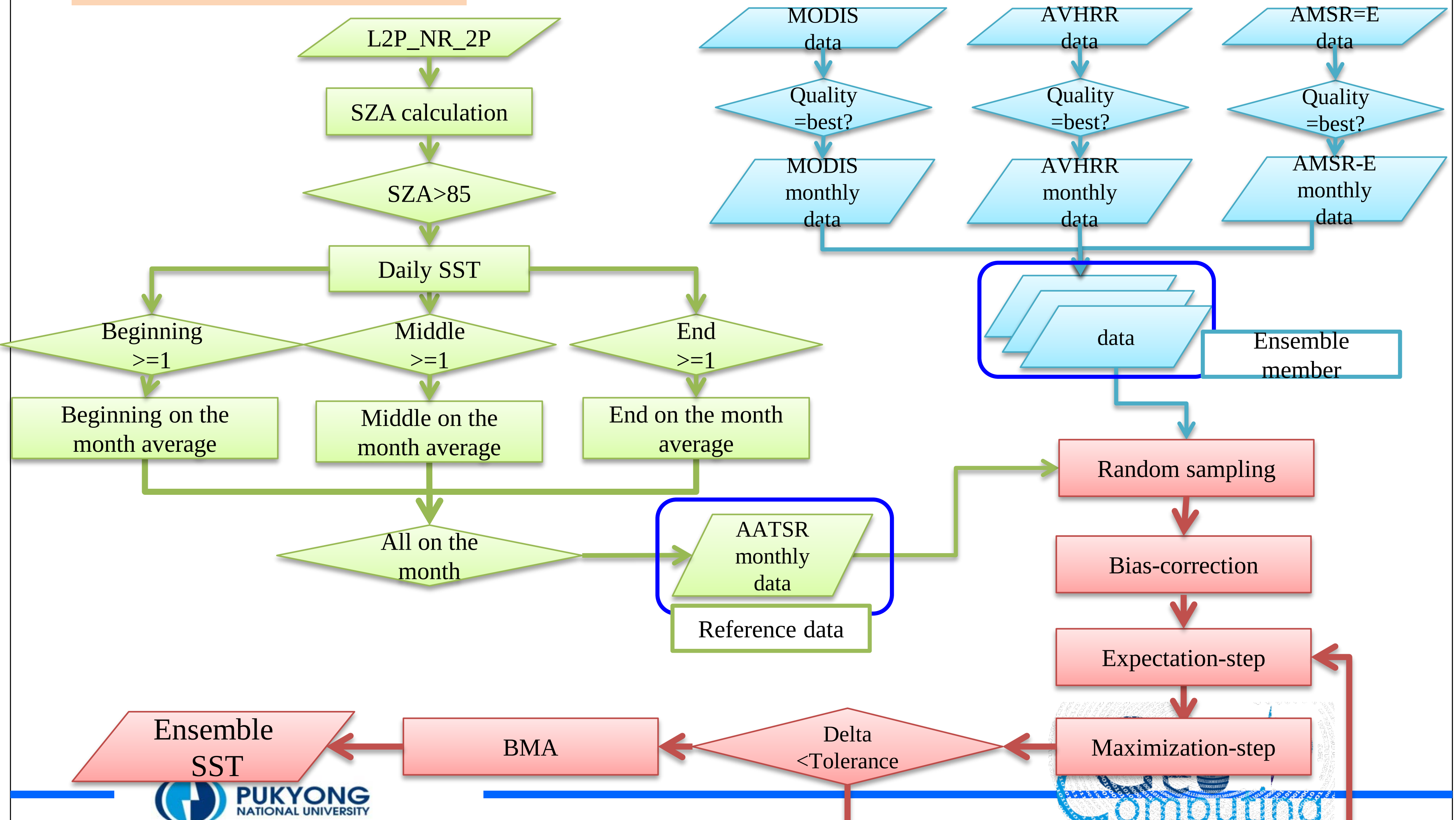
- AATSR/Envisat SST product was used as the reference data. Because the product is famous for accurate data and detect also skin SST. The error of the AATSR SST product is less than it detected with Buoy.

PUKYONG NATIONAL UNIVERSITY

GeoComputing Pukyong National University

3. Ensemble test – flow for ensemble BMA

Flow cart for BMA test



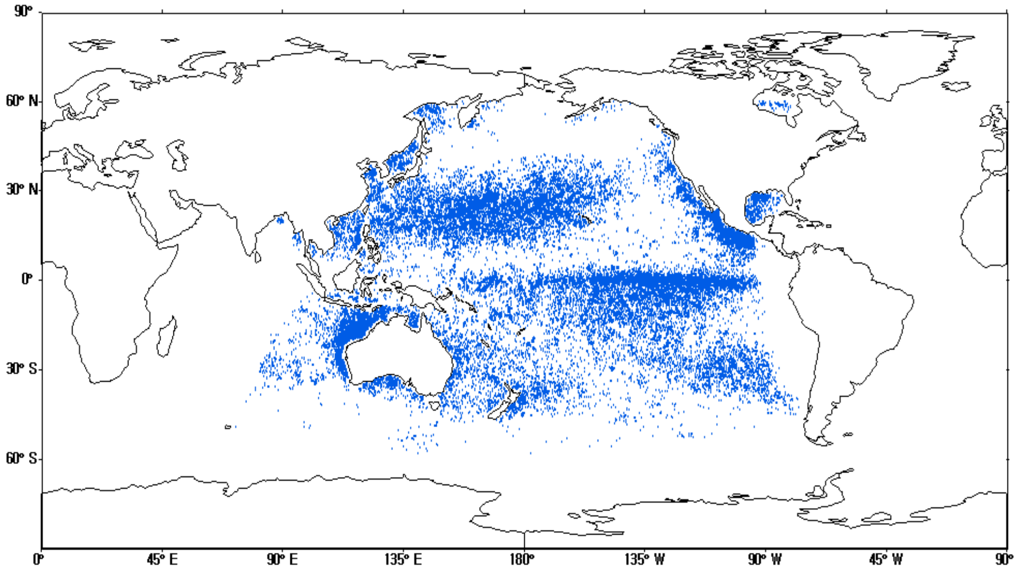
PUKYONG NATIONAL UNIVERSITY

Pukyong National University

3. Ensemble test – process in BMA

The location of randomly extracted match-up sample data

- Match up
  - Match-up database was constituted with ensemble members(MODIS, AVHRR, AMSR-E) and reference data(AATSR) of same month and location from January 2006 to December 2008 for 36 months.
- Random sampling
  - 1000 sample data per a month were extracted from the match-up database → total 36000(1000*36month)



Bias-correction

Three regression equations for MODIS, AVHRR, and AMSR-E with regard to the AATSR measurements were derived from the matchup database for use in the bias correction

$$SST_{AATSR} = 0.991 \times SST_{MODIS} + 0.1221$$

$$SST_{AATSR} = 0.992 \times SST_{AVHRR} - 0.0265$$

$$SST_{AATSR} = 0.984 \times SST_{AMSR-E} + 0.1265$$

PUKYONG NATIONAL UNIVERSITY

Geo Computing Pukyong National University

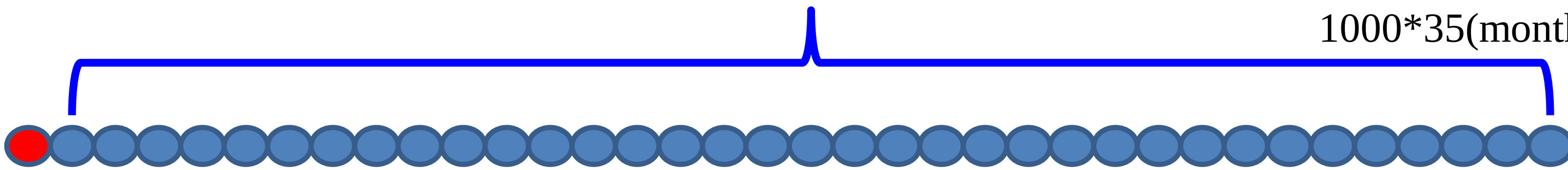
3. Ensemble test – ensemble BMA calculation &validation

	MODIS	AVHRR	AMSR-E	Mean Ensemble	Median Ensemble	BMA Ensemble
Mean bias	-0.104	-0.230	-0.288	-0.207	-0.201	0.001
RMSE	0.394	0.490	0.488	0.400	0.401	0.338

One-leave-out-cross validation

The number of sample data for calculating the weight of a set :
1000*35(month)=35000

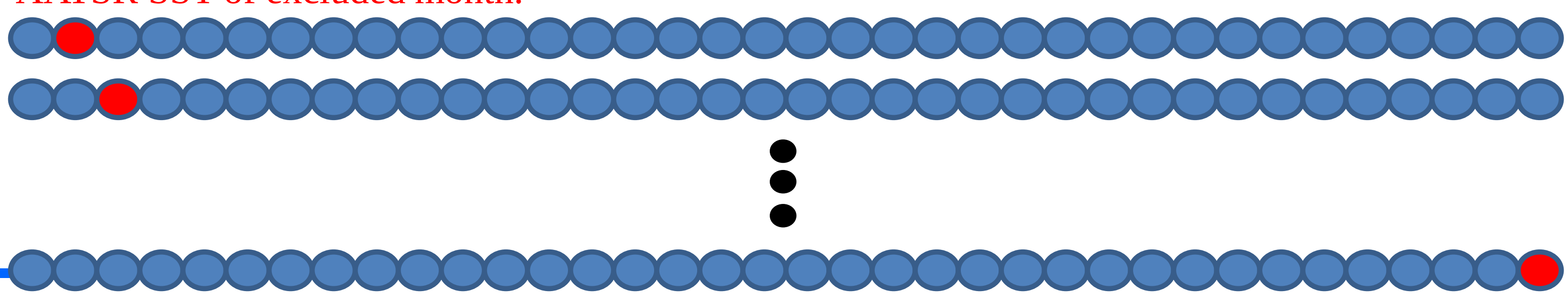
Used 35 months for calculating the weight.



: The calculation period

: The validation period

BMA ensemble is created with weight obtained by calculation period and is compared with AATSR SST of excluded month.



Validation of total 36 set
35000*36(set)

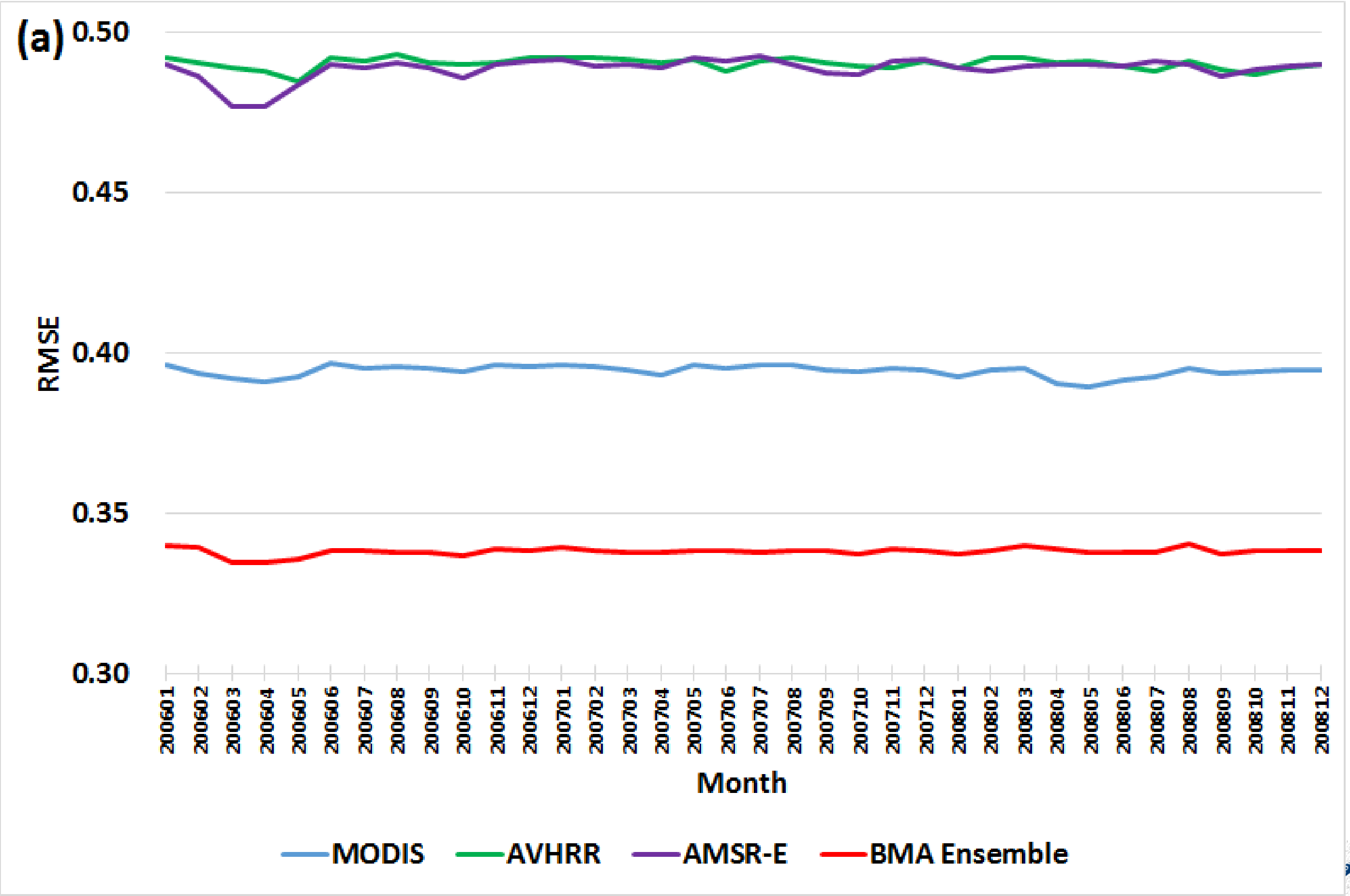
3. Ensemble test – ensemble BMA calculation &validation

Weight of ensemble members about 36 sets

Month	Weighting			Month	Weighting		
	MODIS	AVHRR	AMSR-E		MODIS	AVHRR	AMSR-E
200601	0.383	0.207	0.410	200707	0.384	0.204	0.412
200602	0.386	0.203	0.412	200708	0.384	0.204	0.412
200603	0.372	0.192	0.436	200709	0.383	0.204	0.412
200604	0.378	0.193	0.429	200710	0.385	0.202	0.413
200605	0.378	0.212	0.410	200711	0.391	0.211	0.398
200606	0.381	0.204	0.415	200712	0.396	0.206	0.398
200607	0.383	0.207	0.410	200801	0.397	0.204	0.399
200608	0.385	0.206	0.409	200802	0.388	0.201	0.411
200609	0.382	0.206	0.412	200803	0.393	0.205	0.402
200610	0.386	0.200	0.414	200804	0.406	0.202	0.392
200611	0.386	0.208	0.406	200805	0.411	0.197	0.392
200612	0.390	0.206	0.404	200806	0.402	0.204	0.394
200701	0.385	0.213	0.403	200807	0.393	0.206	0.401
200702	0.385	0.207	0.408	200808	0.389	0.206	0.404
200703	0.388	0.208	0.404	200809	0.386	0.208	0.406
200704	0.392	0.202	0.405	200810	0.380	0.212	0.407
200705	0.383	0.206	0.411	200811	0.381	0.210	0.409
200706	0.384	0.205	0.410	200812	0.391	0.206	0.403

3. Ensemble test – ensemble BMA calculation &validation

Compare with BMA ensemble and ensemble members



1set
= 1000(pixel)*35(month)
= 35000(pixel)

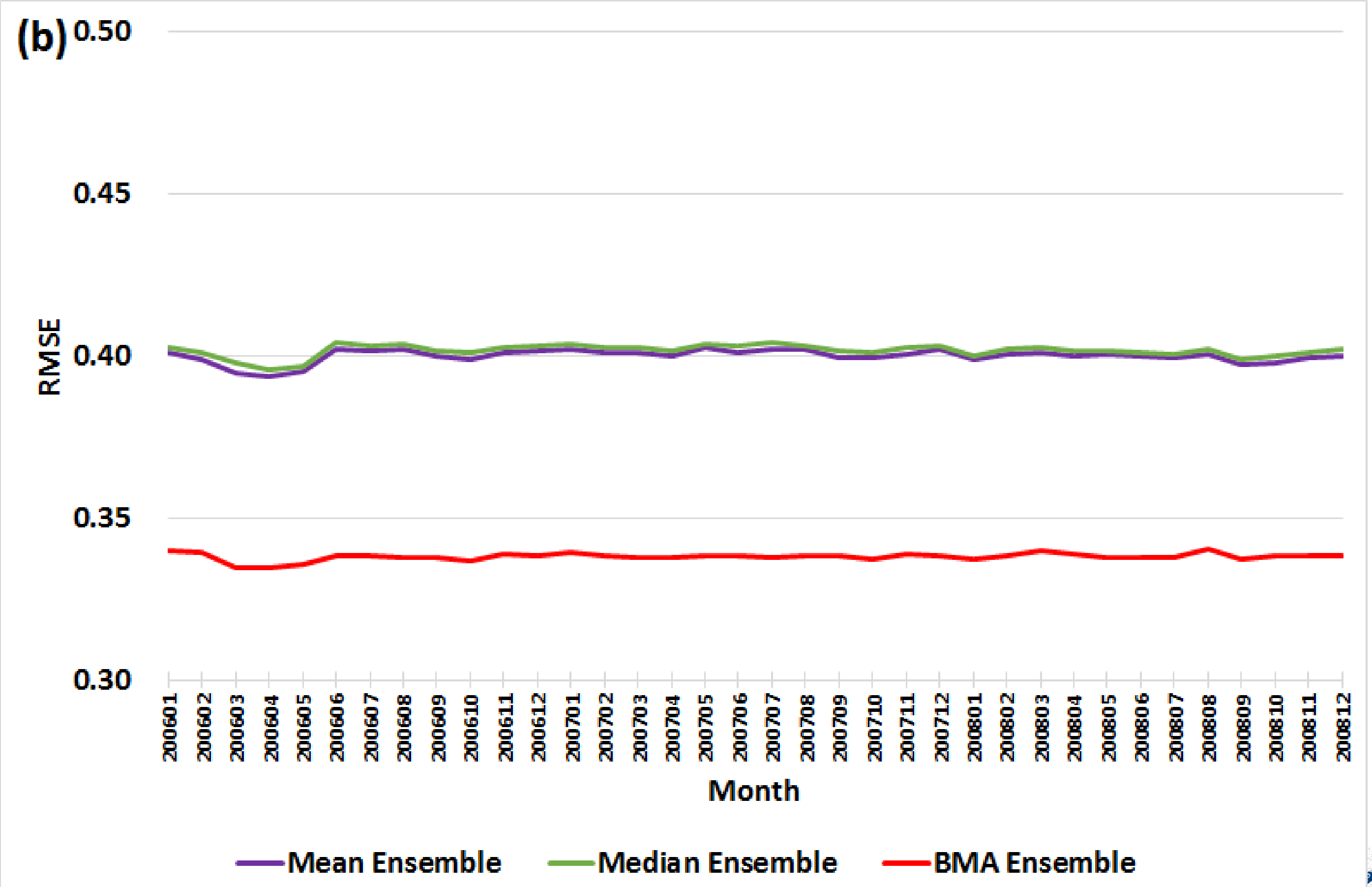
(Set number)

PUKYONG NATIONAL UNIVERSITY

GeoComputing Pukyong National University

3. Ensemble test – ensemble BMA calculation &validation

Compare with BMA ensemble and others ensemble(mean and median ensemble)



1set
= 1000(pixel)*35(month)
= 35000(pixel)

(Set number)

PUKYONG NATIONAL UNIVERSITY

GeoComputing Pukyong National University

3. Ensemble test – ensemble BMA calculation &validation

Distributions of error

- The spatial and temporal distributions of the errors of BMA ensembles were also examined.
- We separated all of the matchups for sample locations according to their latitude in 10-degree intervals.
- The BMA errors appeared to be randomly distributed irrespective of the latitude and month

Durbin-Watson test

- If there are temporal-spatial autocorrelation, that can not be referred to the appropriate statistical techniques.
- Using Durbin-Watson test, time-spatial autocorrelation inquire out.
- The Durbin–Watson test for locational or spatial dependence showed that approximately 70% of the error values were not autocorrelated.

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

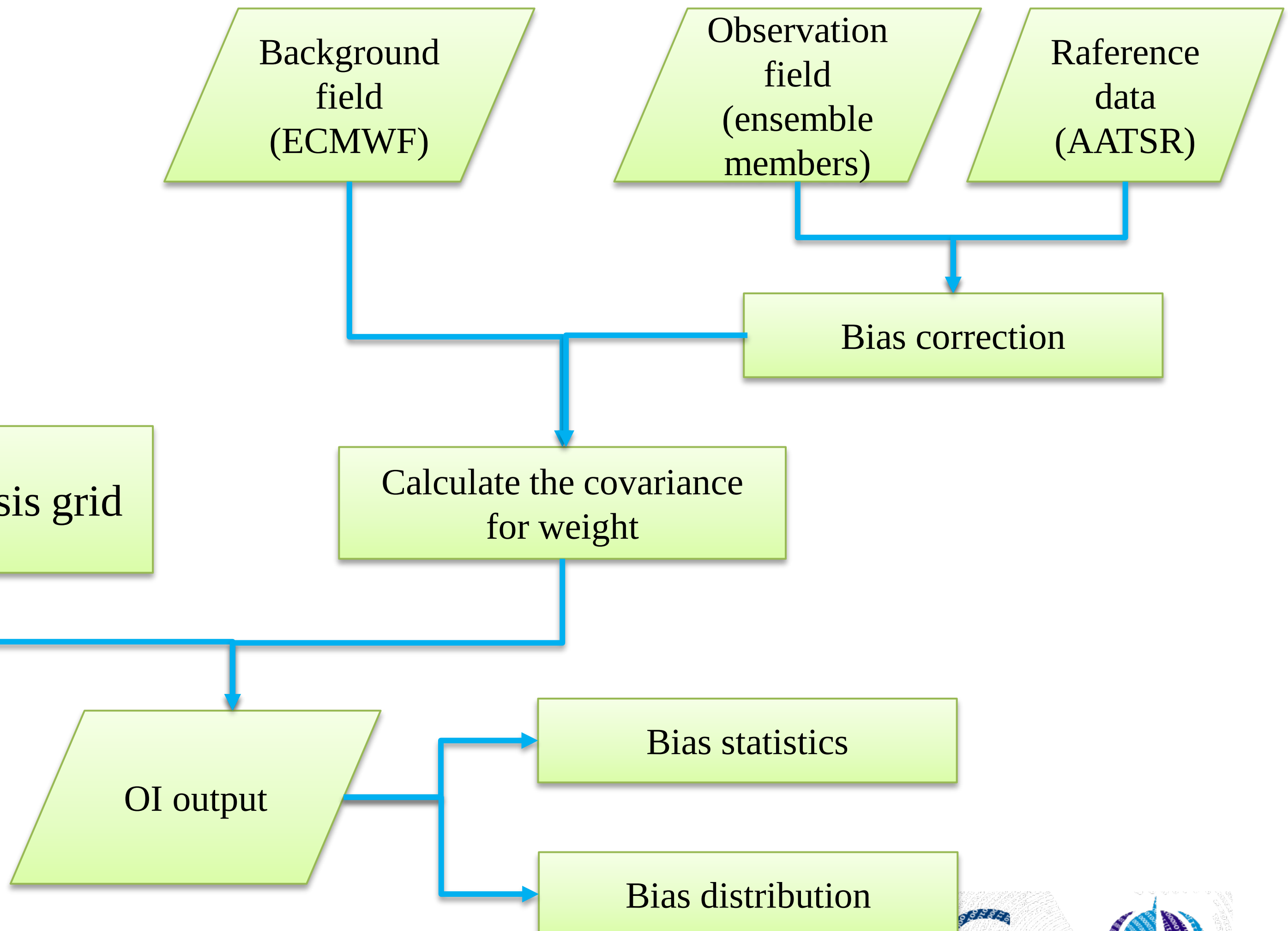
(a) 2006	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Mean	SD
60°N~50°N		0.017				0.330				0.038		-0.007	0.095	0.158
50°N~40°N	0.016	-0.025	-0.066	0.062	0.368	0.224	0.386	0.236	0.026	0.114	-0.078	-0.365	0.075	0.210
40°N~30°N	-0.041	-0.145	-0.021	0.108	0.447	0.145	0.161	0.199	-0.143	-0.040	-0.182	-0.137	0.029	0.187
30N~20N	-0.233	-0.312	0.187	0.126	0.190	0.217	0.072	0.133	-0.057	0.078	-0.018	-0.070	0.026	0.169
20°N~10°N	0.193	-0.008	-0.313	-0.135	0.064	0.238	0.169	0.153	-0.250	-0.257	-0.119	-0.100	-0.030	0.192
10°N~0	-0.070	-0.149	-0.054	-0.443	0.112	0.188	0.134	0.204	0.165	0.111	0.152	0.167	0.043	0.191
0~10°S	-0.106	-0.031	-0.183	-0.489	-0.024	0.119	-0.001	0.188	0.118	0.055	-0.054	-0.063	-0.039	0.176
10°S~20°S	-0.084	-0.072	-0.409	-0.370	-0.275	0.079	0.084	0.256	-0.021	0.122	0.118	0.433	-0.012	0.250
20°S~30°S	-0.111	-0.257	0.116	-0.104	-0.293	-0.068	0.033	0.043	0.178	0.211	0.027	0.132	-0.008	0.162
30°S~40°S	-0.157	-0.330	-0.018	0.193	-0.279	-0.054	0.021	0.091	0.022	-0.065	0.115	0.008	-0.038	0.154
40°S~50°S			0.661	-0.265		0.097	-0.215	0.010	-0.113	0.282	0.708		0.146	0.375
50°S~60°S			0.007	-0.213	-0.848	-0.883	-0.294	-0.057	-0.184	0.162			-0.289	0.383
Mean	-0.066	-0.131	-0.008	-0.139	-0.054	0.053	0.050	0.133	-0.023	0.068	0.067	0.000		
SD	0.120	0.129	0.281	0.239	0.383	0.317	0.184	0.100	0.142	0.141	0.250	0.212		

(b) 2007	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Mean	SD
60°N~50°N	-0.085	0.008	0.001						-0.138	-0.229	-0.108	-0.001	-0.079	0.088
50°N~40°N	-0.019	0.137	0.142	0.054	0.295	0.084	0.582	0.211	0.052	0.184	-0.065	0.015	0.139	0.173
40°N~30°N	-0.117	0.024	0.002	0.069	0.160	0.313	0.254	0.212	0.092	0.044	-0.110	-0.091	0.071	0.142
30N~20N	0.093	-0.107	0.040	0.120	0.261	0.182	0.073	0.032	0.060	0.065	-0.010	-0.037	0.064	0.097
20°N~10°N	-0.059	-0.135	0.190	0.062	0.148	0.295	0.228	0.066	-0.219	-0.220	-0.179	0.070	0.021	0.180
10°N~0	0.146	0.120	0.116	0.058	0.129	0.287	0.263	0.254	0.209	0.004	0.179	0.151	0.160	0.084
0~10°S	0.073	0.235	0.089	-0.076	0.006	0.150	0.239	-0.017	0.484	-0.045	0.003	0.288	0.119	0.166
10°S~20°S	0.167	-0.025	0.012	0.049	0.200	0.142	0.025	0.060	0.112	0.166	0.251	0.216	0.115	0.089
20°S~30°S	0.129	0.088	0.049	0.210	0.085	0.111	0.207	0.075	0.080	0.106	0.049	0.113	0.108	0.053
30°S~40°S	-0.071	0.112	0.080	-0.081	0.118	0.136	0.132	0.072	0.167	0.079	0.042	0.047	0.070	0.077
40°S~50°S	-0.311	0.144		-0.353		0.072	-0.179	-0.122	-0.192	-0.210	-0.204		-0.151	0.163
50°S~60°S					-0.909	-0.340	-0.020	-0.189	0.002	0.029	-0.324		-0.250	0.328
Mean	-0.005	0.055	0.072	0.011	0.049	0.130	0.164	0.059	0.059	-0.002	-0.040	0.077		
SD	0.144	0.113	0.063	0.153	0.347	0.178	0.196	0.136	0.191	0.145	0.162	0.117		

(c) 2008	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Mean	SD
60°N~50°N	-0.051	0.033	0.403								0.111		0.124	0.197
50°N~40°N	-0.104	-0.207	-0.030					0.106	0.113		-0.728	-0.281	-0.162	0.290
40°N~30°N	-0.239	-0.070	-0.012	-0.292	-0.179			-0.279	0.155	-0.296	-0.354	-0.151	-0.172	0.157
30N~20N	-0.180	-0.218	-0.018	-0.140	-0.182	-0.094	-0.012	-0.055	-0.113	-0.111	-0.113	0.017	-0.102	0.073
20°N~10°N	-0.284	0.279	-0.090	-0.154	-0.083	-0.115	-0.085	0.088	-0.412	-0.290	-0.029	-0.037	-0.101	0.181
10°N~0	0.164	-0.026	-0.071	0.020	0.140	0.077	0.077	0.022	-0.077	-0.001	0.014	-0.079	0.022	0.081
0~10°S	0.539	0.025	-0.076	-0.033	0.053	-0.132	0.043	0.061	0.119	0.036	0.184	-0.071	0.062	0.174
10°S~20°S	0.144	0.019	-0.024	0.078	0.181	0.153	-0.034	-0.019	0.097	0.210	-0.081	0.096	0.068	0.095
20°S~30°S	0.168	0.118	0.002	-0.163	0.144	0.060	0.126	0.027	-0.120	0.052	-0.006	0.214	0.052	0.113
30°S~40°S	0.260	0.145	-0.017	0.101	0.207	0.024	0.050	-0.083	-0.020	0.007	0.928		0.146	0.279
40°S~50°S	0.141		-0.128				-1.149	0.260	0.194	0.122			-0.094	0.534
50°S~60°S			-0.225	-0.492	-0.758		0.252	-0.516	0.219	-0.369	-0.194		-0.260	0.354
Mean	0.051	0.010	-0.024	-0.120	-0.053	-0.004	-0.081	-0.035	0.014	-0.064	-0.024	-0.037		
SD	0.247	0.153	0.149	0.189	0.304	0.110	0.412	0.208	0.186	0.195	0.402	0.150		

(d) All	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Mean	SD
60°N~50°N	-0.045	0.019	0.135			0.110			-0.046	-0.064	0.001	-0.003	0.013	0.073
50°N~40°N	-0.036	-0.032	0.015	0.039	0.221	0.103	0.323	0.184	0.064	0.099	-0.290	-0.210	0.040	0.172
40°N~30°N	-0.132	-0.064	-0.010	-0.038	0.143	0.153	0.138	0.044	0.035	-0.098	-0.215	-0.126	-0.014	0.120
30N~20N	-0.107	-0.212	0.070	0.035	0.090	0.102	0.044	0.037	-0.036	0.011	-0.047	-0.030	-0.004	0.090
20°N~10°N	-0.050	0.045	-0.071	-0.076	0.043	0.139	0.104	0.102	-0.293	-0.255	-0.109	-0.022	-0.037	0.137
10°N~0	0.080	-0.018	-0.003	-0.122	0.127	0.184	0.158	0.160	0.099	0.038	0.115	0.079	0.075	0.088
0~10°S	0.169	0.076	-0.057	-0.199	0.012	0.045	0.094	0.077	0.241	0.015	0.044	0.052	0.047	0.109
10°S~20°S	0.076	-0.026	-0.140	-0.081	0.036	0.124	0.025	0.099	0.063	0.166	0.096	0.248	0.057	0.106
20°S~30°S	0.062	-0.017	0.056	-0.019	-0.021	0.034	0.122	0.048	0.046	0.123	0.023	0.153	0.051	0.058
30°S~40°S	0.011	-0.024	0.015	0.071	0.015	0.036	0.068	0.027	0.056	0.007	0.362	0.018	0.055	0.100
40°S~50°S	-0.057	0.048	0.178	-0.206	0.000	0.056	-0.514	0.049	-0.037	0.065	0.168		-0.023	0.194
50°S~60°S			-0.072	-0.235	-0.838	-0.408	-0.021	-0.254	0.012	-0.060	-0.173		-0.228	0.265
Mean	-0.003	-0.019	0.010	-0.076	-0.016	0.057	0.049	0.052	0.017	0.004	-0.002	0.016		
SD	0.091	0.077	0.091	0.105	0.282	0.154	0.207	0.114	0.124	0.113	0.181	0.131		

3. Ensemble test – flow for OI process(ongoing)



3. Ensemble test – OI process detail

- Set the analysis grid (0.125 or 0.25)

- Set the Background field
  - 1979-2008 monthly SST using ECMWF reanalysis data.

- Bias-correction of members with AATSR
  - Using the best quality data of members and AATSR matching data.
  - Regression equation in process BMA will be used.

- Decorrelation scale set or calculation for calculating the covariance.
  - Fixed value? Value obtained from equation? (current problem)

- Calculate weight using the covariance of background-observation and observation-observation.

PUKYONG NATIONAL UNIVERSITY

GeoComputing Pukyong National University

Summary and Conclusion

- As the test, BMA was applied to mulit-satellite products.
- The weight used in BMA was estimated by EM algorithm.
- SST of MODIS, AVHRR, and AMSR-E as ensemble members and AATSR as reference data were used.
- For validation, 36 validation sets were created and the RMSE with reference data were compared with that of ensemble members and others ensembles
- As the result, BMA ensemble SST was shown the lowest RMSE in all of the validation sets.
- This is presumably because the weighting scheme was based on the posterior probability, namely the suitability of each member derived from training procedures.
- For comparison with the ensemble using OI, the process is ongoing.

Future work

- The result of OI will be carried out as BMA ensemble test procedure
- OI and BMA will be compared each other.
- SST data of Buoy is added as reference data for BMA weight training and bias correction.
- This method will be applying to geostationary satellites covering Asian regions, such as the Japanese Himawari-8/9, the Chinese FY-4, and the Korean GK-2A as ensemble members.

THANK YOU FOR YOUR ATTENTION

PUKYONG NATIONAL UNIVERSITY

GeoComputing Pukyong National University