Round Robin: Final selection and ranking of algorithms

v1.1 20 January 2020

Author	Approved	Signature	Date
Fabrice Ardhuin, Guillaume Dodet, Ellis Ash	Fabrice Ardhuin		20/01/2020
ESA Acceptance	Digitally signed by Craig Donlon Date: 2020.01.20 16:00:36 +01'00'		

Issue	Date	Comments
1.0	15/01/2020	First version for approval by ESA
1.1	20/01/2020	Increased precision of results to ensure score always spans full range. Minor corrections.

Contents

1. Introduction

This document describes the final evaluation and selection of algorithms considered during the altimetry Round Robin process of the Sea State CCI project.

In order to be able to keep at least two algorithms for each category (Low Rate Measurement Mode, LRM and Synthetic Aperture Radar Mode, SARM), the final choice between the two has been made by consensus between the project management (Fabrice, Guillaume and Ellis), ESA (Craig & Paolo), and the production team (Ifremer), based on a selection of qualitative and quantitative criteria listed below. The reason for selecting two algorithms is to keep some flexibility for taking into account other constraints (processing time, portability in the production environment).

As exposed in the CCI Sea State User Requirement Document, users of the CCI dataset are strongly concerned by:

- stability of estimates across instruments (eventually from TOPEX to Jason 6 and beyond, especially across the LRM to SARM boundary);
- accuracy and stability of high sea state values;
- accuracy at the coast where sea state and sea level may be combined into a total sea level (note that extremes are the most important in this context, and applications probably require the combination of models and EO data to arrive a the necessary sampling).

We acknowledge that the data will also be used for many other applications (such as defining the sea state climate for engineering projects, again a question for which extremes are dominant). We shall thus follow the requirements laid out in the CCI-Sea State “User Requirement Document”, and in particular the section 5.1 ‘top level requirements’.

Among these the concern for coastal applications and high resolution are the most relevant for selecting retracking algorithms. Indeed, the CCI-Sea State User questionnaire shows that a significant fraction of sea state data users are ocean and coastal engineers (around 30% among the 184 participants to the survey). One of the first motivation to use satellite sea state was the study of extreme events (to the question “your interest in satellite data concerns?”, more than 120 answered “extreme event”). Most participants to the survey were interested in long-term significant wave height data at high-resolution (~10 km) for the study of extreme sea state conditions and the impact on the coast. This was confirmed by the recommendations of the UCM to produce consistent long-term sea state data records in order to better characterize extreme sea states and trends.

2. Criteria for LRM non-filtered data and for SARM data

2.1. Qualitative criteria on resolved signal

Because denoising techniques can be capable to separate noise from signal (e.g. Quilfen and Chapron 2019) we have set a lower limit on the spectral level at 50% of the expected Power Spectral Density (PSD) level, based on the denoised CCI-v1 data (such as shown in Quilfen and Chapron 2019 for the Agulhas current region). The logic is that if the retracked data is below that level some important signal is certainly missing in the data.

A first estimate of this global mean PSD is shown below, as provided by Y. Quilfen, using Jason 2 data only.

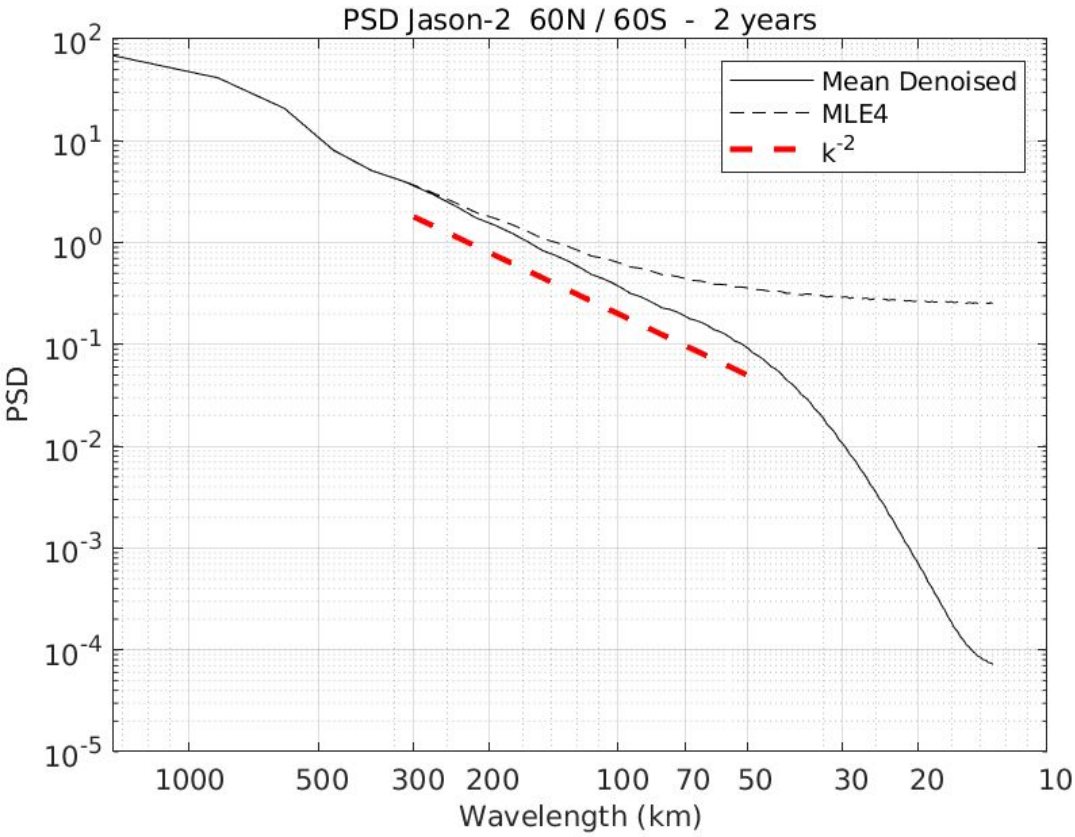


Figure 1: Average power spectral density from the de-noised Jason 2 data over the globe. This was obtained using 256-point segment from the 1 Hz data (original MLE4 or denoised data).

Note that we have not enforced an upper limit to the PSD because potentially the noise and signal can be separated (e.g. using the kind of denoising applied to CCI-v1), although it may be more difficult for higher spectral levels.

The average global PSD of Hs should be around 0.4 m2/cycle/ km at 100 km wavelength and 0.1 m2/cycle/ km at 50 km wavelength. The general shape of the Hs spectra estimated by Quilfen and Chapron (2019) is supported by numerical modelling (Ardhuin et al. 2017) and theoretical analysis (Villas-Boas and Young, submitted). In general the spectral shape follows the surface kinetic energy (KE) spectrum because the variability of Hs is dominated by current-induced refraction, and this KE spectrum may be further linked to the SSH spectrum via the surface quasi-geostrophic theory (e.g. Klein et al. 2008).

The spatial variability along the tracks contains signal and noise. We thus expect the data to have higher variance than the signal alone. This excess of variance due to noise may be removed by adequate filtering (e.g. Quilfen et al. 2018). However, if the variance is lower, it means that some signal must have been removed in the processing. For this reason we have defined the following criteria for the spatial scales of interest to most users (50-100km wavelengths). These are also the scales for which we expect the new algorithms to provide useful data:

Spectral level at 100 km (qualitative pass / fail): the global average PSD should be above 0.2 m^2/ (cycle/km).

Spectral level at 50 km (qualitative pass / fail): the global average PSD should be above 0.05 m^2/ (cycle/km).

2.2 Quantitative criteria to arrive at Combined score for ranking:

We propose to estimate a combined score as follows, the lowest score being the best:

0.3 * Accuracy against models for global areas (std in m)

+ 0.3 * Accuracy against coastal buoys (Standard deviation of the differences (SDD) in m for all buoys < 20 km from shore):

+ 0.1* Accuracy (compared to models, SDD in m) for large Hs (5 m)

+ 0.1* Accuracy (compared to models, SDD in m) for very large Hs (10 m) :

+ 0.1* Intrinsic noise level (SDD in m)

+ 0.1* Intrinsic noise level (SDD in m) for d2c < 20 km

Alternatively we may give 1 point for the algorithm with the highest score on a given criterion, 2 points for the second … and add the points with the same weights as defined above. The final ranking is obtained by ranking the score in inverse order. Presumably the result will be the same.

3. Evaluation results

3.1 Qualitative assessment of spectrum

As mentioned above, the spatial variability along the tracks contains signal and noise. In the figure below, the STARv2 retracker stands out as having a PSD at 100 km wavelength that is 4 times lower than all the other retrackers and that appears too low compared to the analysis discussed above. As a result STARv2 is excluded from the final selection.

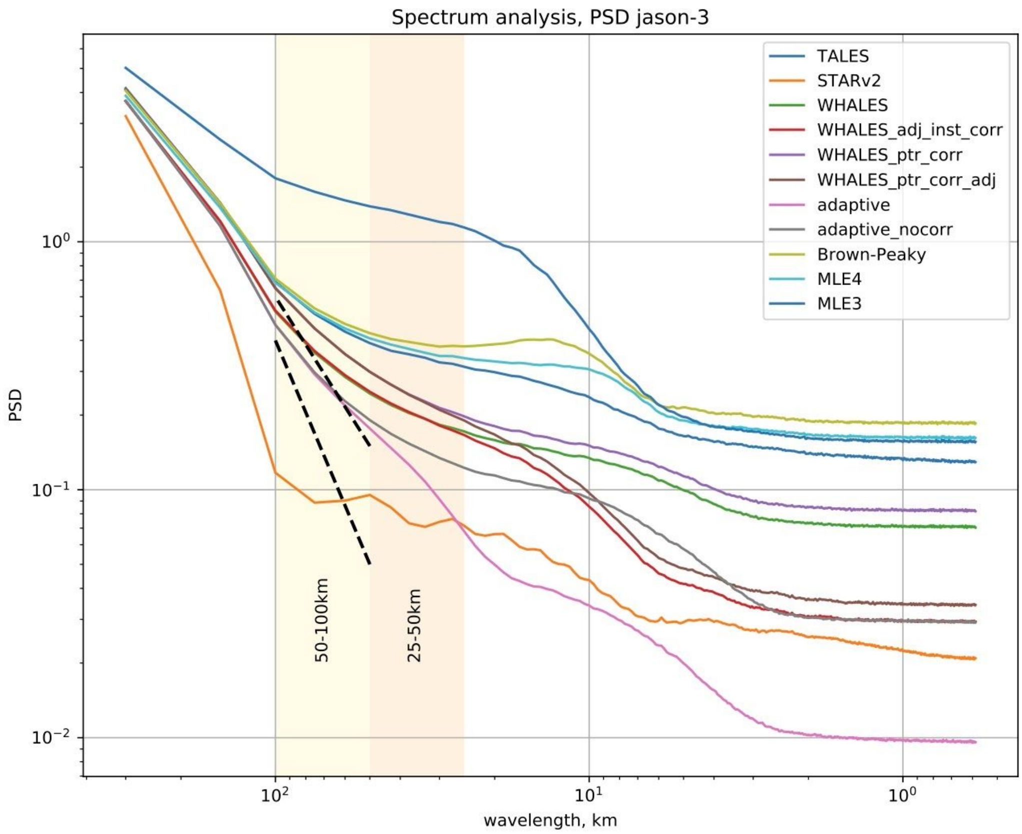
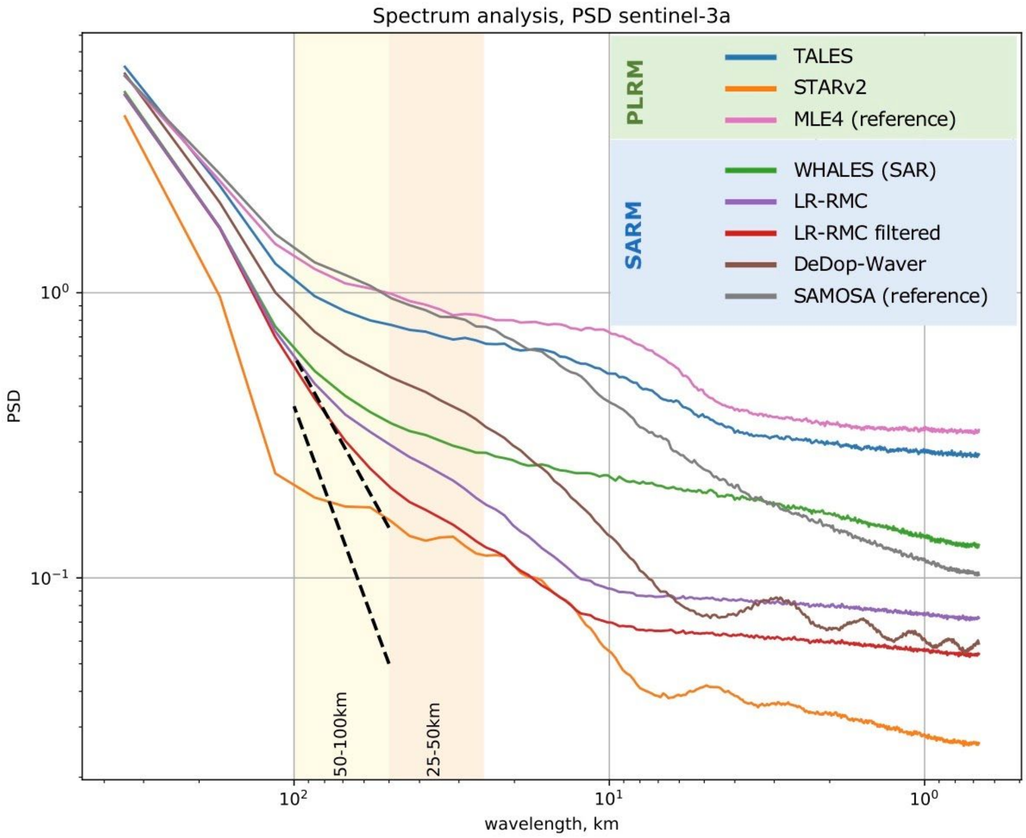


Figure 2: Average power spectral density from the Jason 3 data over the selected tracks.

Figure 3: Average power spectral density from the Sentinel-3A data over the selected tracks.



3.2 Quantitative assessment and ranking

The following tables apply the criteria of section 2.2 to the Round Robin results to arrive at a relative score for each metric and overall ranking. The rows corresponding to WHALES_adj, WHALES_realPTR_adj, Adaptive HFA and LR-RMC_HFA are shaded in grey as they are excluded from the selection since they correspond to algorithms using a filtering techniques that need further assessment and can be applied a posteriori. Full results are presented in Annex A.

Metric score and overall ranking for Low Rate Measurement mode (LRM) (Jason-3 retracking)							
Algorithm	**swh_sdd ocean**	**swh_sdd coast**	**swh_sdd >5m**	**swh_sdd >10m**	**noise ocean**	**noise coast**	**overall ranking a,b***
MLE-3	1	3	1	4	3	2	11,11
MLE-4	3	8	4	6	2	3	9,8

Brown-Peaky	4	9	3	1	1	1	10,7
WHALES	5	7	5	5	6	6	6,4
WHALES_adj	7	2	7	10	9	8	4,9
WHALES_realPTR	6	6	6	3	5	5	7,5
WHALES_realPTR_adj	9	1	9	8	7	7	5,10
Adaptive	8	5	8	7	8	10	3,3
Adaptive_HFA	10	4	10	9	11	11	2,2
TALES	2	10	2	2	4	4	8,6
STARv2	11	11	11	11	10	9	1,1

*The overall ranking is calculated using two methods: a) based on weighted metric scores; b) based on weighted metric results.

Metric score and overall ranking for SAR mode and PLRM (Sentinel-3 retracking)							
Algorithm	**swh_sdd ocean**	**swh_sdd coast**	**swh_sdd >5m**	**swh_sdd >10m**	**noise ocean**	**noise coast**	**overall ranking a,b***
SAMOSA-2.5	1	1	1	2	3	3	8,8
WHALES-SAR	5	4	6	3	4	5	4,4
DeDop-Waver	2	5	2	4	6	7	5,5
LR-RMC	6	8	7	7	5	4	2,2
LR-RMC_HFA	7	7	8	8	7	6	1,1
MLE-4-PLRM	3	2	5	5	1	2	7,7
TALES-PLRM	4	3	3	6	2	1	6,6
STARv2-PLRM	8	6	4	1	8	8	3,3

4. Conclusion

Based on the evaluation in section 3 we select two algorithms from each category (Low Rate Measurement Mode, LRM and Synthetic Aperture Radar Mode, SARM) for further development and implementation in the production of the v2 dataset during the current phase of Sea State CCI:

LRM final selection

1) Adaptive
2) WHALES

SARM final selection

1) LR-RMC
2) WHALES-SAR

We recommend that a further round robin be conducted early in a phase 2 of Sea State CCI focussing on the re-evaluation of SARM algorithms that will have benefitted from further development during the intervening period.

5. References

Ardhuin, F., Gille, S.T., Menemenlis, D., Rocha, C.B., Rascle, N., Chapron, B., Gula, J., Molemaker, J., 2017. Small-scale open ocean currents have large effects on wind wave heights. Journal of Geophysical Research: Oceans 122, 4500–4517. https://doi.org/10.1002/2016JC012413

Klein, P., Hua, B.L., Lapeyre, G., Capet, X., Le Gentil, S., Sasaki, H., 2008. Upper Ocean Turbulence from High-Resolution 3D Simulations. J. Phys. Oceanogr. 38, 1748–1763. https://doi.org/10.1175/2007JPO3773.1

Quilfen, Y., Yurovskaya, M., Chapron, B., Ardhuin, F., 2018. Storm waves focusing and steepening in the Agulhas current: Satellite observations and modeling. Remote Sensing Of Environment 216, 561–571. https://doi.org/10.1016/j.rse.2018.07.020

Quilfen, Y., Chapron, B., Submitted. On denoising satellite altimeter measurements for high-resolution geophysical signal analysis. Advances in Space Research.

Villas Bôas, A.B., Young, W.R., Submitted. Diffusion of surface gravity wave action by mesoscale turbulence at the sea surface. J. Fluid Mech.

Annex A: Full results

Jason-3 LRM algorithm evaluation table

Metrics	Weight	MLE-3	MLE-4	Brown-Peaky	WHALES	WHALES_adj	WHALES_realPTR	WHALES_realPTR_adj	Adaptive	Adaptive_HFA	TALES	STARv2
Comparison against model (cy46r1): swh_sdd (d2c > 20 km)	0.3	0.675	0.308	0.2793	0.2216	0.2172	0.221	0.2154	0.2158	0.2103	0.3166	0.1923
Ranking		1	3	4	5	7	6	9	8	10	2	11
Weighted ranking		0.3	0.9	1.2	1.5	2.1	1.8	2.7	2.4	3	0.6	3.3
Weighted metrics		0.202	0.092	0.084	0.066	0.065	0.066	0.065	0.065	0.063	0.095	0.058
Comparison against buoys: swh_sdd (d2c <= 20 km)	0.3	1.323	0.657	0.6506	0.6935	1.4202	0.6992	1.4366	0.7221	0.7231	0.6399	0.4352
Ranking		3	8	9	7	2	6	1	5	4	10	11
Weighted ranking		0.9	2.4	2.7	2.1	0.6	1.8	0.3	1.5	1.2	3	3.3
Weighted metrics		0.3970	0.1971	0.1952	0.2081	0.4261	0.2098	0.4310	0.2166	0.2169	0.1920	0.1306
Comparison against model (cy46r1): swh_sdd (SWH > 5 m)	0.1	0.414	0.365	0.3688	0.3548	0.3448	0.3466	0.3367	0.3445	0.3356	0.375	0.3118
Ranking		1	4	3	5	7	6	9	8	10	2	11
Weighted ranking		0.1	0.4	0.3	0.5	0.7	0.6	0.9	0.8	1	0.2	1.1
Weighted metrics		0.0414	0.0365	0.0369	0.0355	0.0345	0.0347	0.0337	0.0345	0.0336	0.0375	0.0312
Comparison against model (cy46r1): swh_sdd (SWH > 10 m)	0.1	0.541	0.531	0.5557	0.5382	0.5024	0.5464	0.5086	0.5163	0.5043	0.5509	0.5017
Ranking		4	6	1	5	10	3	8	7	9	2	11
Weighted ranking		0.4	0.6	0.1	0.5	1	0.3	0.8	0.7	0.9	0.2	1.1
Weighted metrics		0.0541	0.0531	0.0556	0.0538	0.0502	0.0546	0.0509	0.0516	0.0504	0.0551	0.0502
Noise: swh (d2c > 20 km)	0.1	0.505	0.519	0.5513	0.3396	0.2171	0.3656	0.2335	0.2248	0.1199	0.4581	0.1729
Ranking		3	2	1	6	9	5	7	8	11	4	10
Weighted ranking		0.3	0.2	0.1	0.6	0.9	0.5	0.7	0.8	1.1	0.4	1
Weighted metrics		0.0505	0.0519	0.0551	0.0340	0.0217	0.0366	0.0234	0.0225	0.0120	0.0458	0.0173
Noise: swh (d2c <= 20 km)	0.1	0.572	0.554	0.5872	0.3734	0.2596	0.3934	0.2902	0.2276	0.1308	0.5214	0.2372
Ranking		2	3	1	6	8	5	7	10	11	4	9
Weighted ranking		0.2	0.3	0.1	0.6	0.8	0.5	0.7	1	1.1	0.4	0.9
Weighted metrics		0.0572	0.0554	0.0587	0.0373	0.0260	0.0393	0.0290	0.0228	0.0131	0.0521	0.0237
Method 1: Total weighted ranking		2.2	4.8	4.5	5.8	6.1	5.5	6.1	7.2	8.3	4.8	10.7
Method 2: Total Weighted metrics		0.803	0.486	0.48527	0.43513	0.62361	0.44126	0.6325	0.41269	0.38908	0.4775	0.3106
Ranking (method 1)		11	9	10	6	4	7	5	3	2	8	1
Ranking (method 2)		11	8	7	4	9	5	10	3	2	6	1

Sentinel-3 SARM and PLRM algorithm evaluation table

Metrics	Weight	SAMOSA-2.5	WHALES-SAR	DeDop-Waver	LR-RMC	LR-RMC_HFA	MLE-4-PLRM	TALES-PLRM	STARv2-PLRM
Comparison against model (cy46r1): swh_sdd (d2c > 20 km)	0.3	0.4834	0.263	0.3752	0.2512	0.2444	0.3248	0.314	0.2146
Ranking		1	5	2	6	7	3	4	8
Weighted ranking		0.3	1.5	0.6	1.8	2.1	0.9	1.2	2.4
Weighted metricss		0.1450	0.0789	0.1126	0.0754	0.0733	0.0974	0.0942	0.0644
Comparison against buoys: swh_sdd (d2c <= 20 km)	0.3	2.1175	0.5684	0.5540	0.4638	0.4795	1.4450	0.6029	0.4924
Ranking		1	4	5	8	7	2	3	6
Weighted ranking		0.3	1.2	1.5	2.4	2.1	0.6	0.9	1.8
Weighted metricss		0.6353	0.1705	0.1662	0.1391	0.1439	0.4335	0.1809	0.1477
Comparison against model (cy46r1): swh_sdd (SWH > 5 m)	0.1	0.6949	0.3502	0.432	0.3311	0.3248	0.3695	0.3848	0.3778
Ranking		1	6	2	7	8	5	3	4
Weighted ranking		0.1	0.6	0.2	0.7	0.8	0.5	0.3	0.4
Weighted metricss		0.0695	0.0350	0.0432	0.0331	0.0325	0.0370	0.0385	0.0378
Comparison against model (cy46r1): swh_sdd (SWH > 10 m)	0.1	0.8423	0.6523	0.6414	0.4404	0.4257	0.5935	0.528	1.1889
Ranking		2	3	4	7	8	5	6	1
Weighted ranking		0.2	0.3	0.4	0.7	0.8	0.5	0.6	0.1
Weighted metricss		0.0842	0.0652	0.0641	0.0440	0.0426	0.0594	0.0528	0.1189
Noise: swh (d2c > 20 km)	0.1	0.398	0.3732	0.278	0.321	0.2749	0.6742	0.6245	0.1725
Ranking		3	4	6	5	7	1	2	8
Weighted ranking		0.3	0.4	0.6	0.5	0.7	0.1	0.2	0.8
Weighted metricss		0.0398	0.0373	0.0278	0.0321	0.0275	0.0674	0.0625	0.0173
Noise: swh (d2c <= 20 km)	0.1	0.4042	0.3359	0.2837	0.3603	0.3058	0.6815	0.7046	0.2537
Ranking		3	5	7	4	6	2	1	8
Weighted ranking		0.3	0.5	0.7	0.4	0.6	0.2	0.1	0.8
Weighted metricss		0.0404	0.0336	0.0284	0.0360	0.0306	0.0682	0.0705	0.0254
Method 1: Total weighted ranking		1.5	4.5	4.0	6.5	7.1	2.8	3.3	6.3
Method 2: Total Weighted metrics		1.0142	0.4206	0.4423	0.3598	0.3503	0.7628	0.4993	0.4114
Ranking (method 1)		8	4	5	2	1	7	6	3
Ranking (method 2)		8	4	5	2	1	7	6	3