Ocean Sci. Discuss., 9, C1425–C1429, 2013
www.ocean-sci-discuss.net/9/C1425/2013/

Ocean Science
Discussions

Interactive comment on "A clustering analysis of eddies' spatial distribution in the South China Sea" *by* J. Yi et al.

Anonymous Referee #2

Received and published: 1 January 2013

General comments

This study makes use of a clustering analysis to establish whether eddies in the South China Sea (SCS) are generated randomly or in preferred geographic areas. The data set used comes from a state-of-the-art numerical model assimilating along-track data from 3 satellite altimeters, which has been validated by comparisons with gridded altimeter data (AVISO) and cruise measurements of currents. Eddies are identified using an algorithm adapted from Chelton et al. (2011) and briefly described in Appendix A. The clustering analysis is performed with the K-means approach, where K is the specified number of clusters. Some aspects of clustering analysis are briefly described in Appendices B and C, which is very useful for readers (such as myself) not familiar with

this type of analysis. An issue with clustering analysis is specifying the "right" number of clusters, K. Some parameters are used to select a reasonable value, but there is not a unique best value. The authors first describe results for K=4, which yield clusters reminiscent of the 4 geographic zones defined in Wang et al. (2003), although with some significant differences. The authors then describe results for K=10 for anticyclonic eddies (AEs) and K=6 for cyclonic eddies (CEs), providing a finer description of eddies preferred generation areas in the SCS. The different areas are associated with different generation mechanisms by referring to the literature on eddies in the SCS.

It is the first time I have seen such an analysis applied to physical oceanographic data, so it appears to me as a novelty. While the generation locations of eddies in the SCS reported in Figure 1 appear quite random to the eye, the statistical analysis clearly proves that they are not random, and the results of the clustering analysis can be meaningfully compared with different generation mechanisms characterizing different areas reported in the literature. Therefore I believe that this paper makes an original and useful contribution to the analysis of mesoscale eddies in the SCS, providing incentive to use the method for different parts of the ocean and for different physical properties such as eddy trajectories. I recommend publication in Ocean Science after minor revisions.

Specific comments

1. T-test analysis: although I am not a statistician, I think there are possibly some issues with the analysis. First, I don't understand the column in Table 1 entitled "Sig. of Levene's for Normal Distribution". Is the normality of the distributions tested here? A quick reading about Levene's test tells me that this test is designed to assess the equality of variances in different samples, not the normality of a distribution. If it is used for the latter, some explanations should be given. According to the results, AEs may not be normally-distributed (small significance

of 0.061), which would be a problem for using a t-test. This should be discussed. Second, the t-test results for AEs and CEs are computed under the assumption of equal variance (footnote "a"), whereas the results of Levene's Test for Equality of Variance show that the variances of AEs and random data differ (small significance of 0.046), as well as the variances of CEs and random data (small significance of 0.044). Therefore, either there is a typo in footnote "a" ("equal variance" should be replaced by "unequal variances"), or the t-tests should be recomputed under the assumption of unequal variances. Finally, the significance of t-tests for Equality of Means is less than 0.0005 in all cases, which is much smaller than the value of 0.05 mentioned in the text (p. 3457, l. 4). Why not using the actual values in the text?

2. p. 3458, l. 6-9: there is some inconsistency in first stating that K=9 for all eddies "is discarded for the separation might be too fine", and then stating that "K=10 is retained for AEs". I suggest removing the discussion about K=9 for all eddies and simply stating that results are shown only for the next local maximum in each SC curve after the K=4 maximum.

3. "improper clustering": on p. 3459, l. 16-17 and on p. 3464, l. 10-11, the authors state that "4-cluster separation improperly groups Z1 and eastern Z2 where the mechanisms are supposed to be different". I would avoid using the term "improperly", because the clustering algorithm is not based on physical mechanisms of eddy generation, but on spatial clustering tendency. Therefore the results of the 4-cluster analysis are not improperly grouping two zones together, but rather failing to distinguish between two zones that are geographically close but that feature different eddy generation mechanisms. This illustrates a general drawback of the clustering analysis if it is intended to uncover different eddy generation mechanisms: there may well exist some geographic areas where different generation mechanisms coexist, and clustering analysis would not be suited to discriminate between the different generation mechanisms in such areas. Perhaps this limita-

tion should be pointed out in the conclusions. A good place would be just after the statement that "different generation mechanisms are expected to be reflected by proper geographic partitions" (p. 3465, l. 2-3), which may not always be the case.

Technical corrections

- Geographic names like "Nansha trough", "Dongsha Islands", "Taiwan Island", "Hainan Island", that are mentioned in the text should be indicated in Figure 1 to help the reader unfamiliar with the area locate them. To avoid making Figure 1 unreadable, they could be indicated by their first letter only, and the correspondence between the letter and the full name given in the figure caption.

- p. 3454, l. 15: "More details can be referred to Du et al. (2011)". There are numerous instances throughout the paper where parentheses for references are not properly placed, so I will not list them all here and rather let the authors carefully review and correct them.

- p. 3455, l. 2: (Tan et al., 2006).

- p. 3456, l. 7: I would replace "to answer the question (a): if eddies' distribution has a clustering tendency or not. Second, correct cluster number is discussed [...]" by "to answer the following questions. (a) Has the eddies' distribution a clustering tendency? (b) What is the correct cluster number? The latter is discussed [...]".

- p. 3459, l. 1: add "of Wang et al. (2003)" after "Z4".

- p. 3461, l. 1 (starting on previous page): "By principal component analysis of SLA and wind stress curl, Shaw et al. (1999) found that the generation of this

dipole structure": it is not clear which dipole structure is being referred to here.

- p. 3463, l. 15: "AEs are preponderant in southwesterly monsoon while CEs are preponderant in northeasterly monsoon".
- p. 3464, l. 15: "patterns of AEs are relatively more aggregated than those of CEs".
- p. 3471, l. 24-25: proper reference is "Wang, G. H. (2004), Discussion on the movement of mesoscale eddies in the South China Sea (in Chinese), Ph.D. dissertation, Sch. of Phys. Oceanogr., Ocean Univ. of China, Qingdao, China.

Interactive comment on Ocean Sci. Discuss., 9, 3451, 2012.