Ocean Sci. Discuss., 8, C438–C442, 2011
www.ocean-sci-discuss.net/8/C438/2011/

Ocean Science
Discussions

Interactive comment on "Extraction of spatial-temporal rules from mesoscale eddies in the South China Sea Based on rough set theory" *by* Y. Du et al.

S. Royston (Referee)

samyst@noc.ac.uk

Received and published: 28 July 2011

General comments

The manuscript discusses a method of obtaining information about relationships in data, with the application of extracting 'rules' relating to the origin and characteristics of mesoscale eddies in the South China Sea. The motivation for this data-driven approach is to complement existing techniques used to investigate the spatio-temporal evolution of eddies (cruise observations, remote sensing and hydrodynamic modelling). Several rules are derived from remote sensing and operational hydrodynamic model data which are consistent with the mechanisms of eddy generation and evolution iden-

tified by previous studies in this region.

To my knowledge, there are very few publications on the application of rule-based methods to oceanographic science and in these works, there has been a focus on the accuracy of prediction. Therefore, this manuscript presents a major step forward by successfully extracting information about the characteristics of eddies from available data. This information on relationships within the data has the potential to drive the focus of future research. The methodology is sufficiently described and the introduction and reference list provide appropriate background information on both eddy characteristics in the South China Sea and the rough set theory methodology.

Although I find the writing clear and informative, in my opinion the manuscript could include more detail for the traceability of results and there are a number of typographical and formatting errors. Therefore, I recommend revision prior to final publication.

Specific comments

Section 2.1: Rought set theory could be given a more low-level introduction. For example, it would be useful to visualise the partitioning and sets in the attribute space, such as Figure 1 in Pawlak (1982), showing how the upper and lower bounds encompass the partitions which are identical (wholly within the rough set boundary) or consistent (partly within the rough set boundary) respectively. In this way, uncertainty in both the attributes and the concept is taken into account. It might also be useful to explain how rough set theory uses a decision system or decision table to determine equivalence classes in the decision attribute, D, from the available conditional attributes, C.

Section 2.1.2: The "certainty" factor could be explained linguistically for clarity. For example, the "certainty" factor is the proportion of instances in the database following both the conditonal attributes *and* the equivalence class of the decision attribute (i.e. for the same region of the conditional attribute space, how many of the instances in the database lead to the same class in the decision attribute?).

Section 2.2: Boolean reasoning. It is indicated that the Boolean reasoning algorithm is

used to discretize continuous data. A brief description of the method would be useful, both for clarity and in case one wishes to replicate the work. If my understanding of what the authors have done is correct, it should be made clear that the discretisation algorithm is applied anew to the conditional attributes for each decision table, based on the optimal cuts for the decision attribute chosen (of the three discussed in this manuscript).

Section 2.2: Description and selection of attributes. It is not clear to me how the 76 eddies identified in the study lead to the 391 typical states (as stated in Section 3.2). Are these 391 states identified at different times in the evolution of the eddies? For example, In Table 6 and Figure 7(b), Rule 10 indicates the eddy is generated in Zone 4, but one typical state is displayed located in Zone 2. Can the authors clarify how often the typical states are identified?

Section 3.1: NLOM vs observation data. It is stated that rough set theory can represent knowledge from incomplete datasets with uncertainty (Section 1). In this case, have the authors attempted to take into account the uncertainties in the NLOM current speed data, for example by using fuzzy rather than crisp partitions in the resulting vorticity attribute? Also, the sentence "So we consider the NLOM SSHA acceptable in the SCS" should include the accuracy of the NLOM current speed.

Section 3.3: Just a comment. A number of the rules could be simplified and there might be more scientifically informative rules by allowing confidence $<$ 100%. For example, for the decision attribute of 'generating season' (Table 6), Rules 1, 2 and 3 suggest that given the 'zone of origin' is Z_2 *and* the 'central temperature' $\in [-\infty, 26.22)$ °C, there are distinct attributes such as 'depth' and 'type' that differentiate winter generated eddies from those generated in other seasons.

The use of rough set theory to forecast eddy characteristics is aluded to in the abstract, discussion and conclusion. Perhaps a sentence could be added in the conclusion explaining how the authors would use rough set theory to make forecasts, as part of

possible future work.

Technical corrections
Notation:

- Pg. 1266 equation (1), line 18: Should "U/D" read "$U|D$" (i.e. "U given D")?
- Pg. 1266 line 19: "CX is ...", should be $\underline{C}X$.
- Pg. 1267 equation (3): The symbol ϕ should be the emptyset symbol $\emptyset$.

Grammatical and typographical:

- Pg. 1262 line 13: Change "2000m-deeper" to "$\geq$2000m" or "2000m and deeper".
- Pg. 1267 line 2: Should read "Once the reducts *are* found ...".
- Pg. 1269 line 20: Typographical error "ware" should read "are".
- Pg. 1277 lines 27-29: Figure 8(b) does not show any eddy states following Rules 1, 4 *and* 5. Perhaps this statement should read "following Rules 1 and 5 or following Rules 4 and 5".
- Pg. 1278 line 10: "Table 6 ..." should read "Table 7 ...".

Formatting:

- Pg. 1270 line 17 to pg. 1271 line 8: The numbering should be restarted at 7.
- References: Wang (2001) is missing. Pg 1284 lines 10-11 and 12-13 are the wrong way around.

- Figures: The captions to figures 5-7 are incorrect. It would be useful if the colour scale in Figure 4(c) could be amended to better highlight the eddy in the NLOM SST field.

Other:

- Pg. 1269 line 12: The citation is not given in the reference list. If it relates to the derivation of 'certainty' and 'coverage' factors, the citation should be given in Section 2.1.2.
- Pg. 1274 lines 1-10: Figure 6 implies that, in Zone 1, there is one eddy following Rule 1 only, one eddy following Rule 2 only and the remainder following both. Similarly, in Zone 3, the Figure implies there is one eddy following Rule 3 only and one eddy following Rule 4 only, with the remainder following both.
- Pg. 1274 line 27: It would be useful to also give the western name for the Xisha Islands (Paracel Islands) and Dongsha Island (Pratas Islands). It would be useful to identify the locations of these islands in Figure 1.
- Pg. 1278 line 23: The statement "no cold eddy occurs" should be clarified, for those 'typical states' identified with rule certainty = 100% and rule coverage $>$ 10%. .

References

Pawlak, Z.: Rough sets, Int. J. Parallel Prog., 11(5), 341-356, 1982.

Interactive comment on Ocean Sci. Discuss., 8, 1261, 2011.