

# Cluster Results

*Robert Schlegel*

*30 March 2017*

## **Concepts that need to be investigated**

Much has been done thus far in regards to the research, and a clear picture of what is further required now exists. The work with SOMs is likely sound, but does require that a few more variables be tested. Furthermore, it is not yet certain that SOMs will be the best clustering technique. To that end, K-means clustering and hierarchical clustering have also been identified as alternative techniques. This now gives rise to the possibility that two papers could emerge from this work. One on the resultant clustering of synoptic air-sea states during coastal MHWs, and another paper that discusses the strengths of these various clustering techniques.

## **The metrics for each MHW in each cluster**

This requires that once the different events have been clustered, regardless of the technique used, or the variables controlled for within (see below), a summary of the event metrics must also be provided. These then will allow for the second more meaningful round of the interpretation of the results.

## **Effect of pixel resolution on clustering**

The more dimensions/ variables one introduces to a cluster analysis, the more stress will exist in the results. As large stress values are generally considered to be a negative aspect of clustering, it is best to attempt to reduce it where possible. For this research that means reducing the pixel resolution of the reanalysis products. There are two reasons that this cannot simply be done out of hand. The first is that the reduction in resolution may affect the clustering of the events. So this must be documented. The other problem this faces is that the reduction of pixel resolution would require that any results produced be shown at this same reduced resolution. And because the goal is to show meso-scale forcing on the coast, higher pixel resolutions would be preferable. Regardless, the ERA-Interim data are at a resolution of 0.5 degrees, which requires that the BRAN data be reduced to this same resolution for appropriate cluster comparison.

## **Effect of lat/ lon extent on clustering**

With more traditional cluster analyses, the values being compared would have far fewer dimensions. In this regard one would endeavour to only include variables that seem relevant to the question being asked. For example, if clustering different rock pools by the species found within them, one would likely create better results by not including any anomalous species finds in the results. In regards to this work, it is best to include only the pixels that are likely relevant to the meso-scale features that may be impacting the coast. More specifically, cutting out the Agulhas retroflection above the Southern Ocean will prevent any behaviour there from affecting the clustering of events that are occurring along the coastline of South Africa.

## **Effect of running air and sea variables separately**

It may be that air and sea values work in tandem with one another to force MHWs, but it is more likely that they do not. Therefore it is necessary to run all clustering techniques on air-sea values combined, as well as separately.

## **Do normal days cluster**

The idea here is to include the 366 daily synoptic climatology values in with the synoptic MHW values to see if they cluster differently.

## **Hierarchical clustering**

Dendrograms

K-means clustering

Ordiplots

SOMs

SOM nodes

References