REVIEWER #2:

This manuscript uses an extensive set of observations on the southeast Australian continental shelf to estimate the scales of variability of various factors. As discussed in the manuscript, knowledge of such scales is critical to designing observational and modeling systems that resolve key processes. I find no major faults in the manuscript, but have a number of questions and comments that the authors should address to improve the manuscript.

We thank the reviewer for his/her interest and relevant questions that helped us improving the manuscript.

1. The description of gliders and the sampling (page 20104) is a bit too vague, and at times somewhat inaccurate. A citation to a general glider reference (e.g., Davis et al. 2003 or Rudnick, et al. 2004) would be helpful for the reader. The statement that gliders 'use seawater to change their buoyancy' is not particularly accurate; each type of glider changes its volume (by either moving oil between internal and external bladders or displacing seawater), thereby changing its buoyancy to rise and fall. This vertical motion is translated into forward motion by wings and controlling the glider's pitch, resulting in a sawtooth path through the water. [I'm sure the authors know this, but they should include it for the sake of unfamiliar readers.] Stating that the 'average horizontal displacement between two dives is around 200 m' is probably true, but somewhat misleading; shallower dives have closer horizontal (and temporal spacing) and so there are more of them, biasing the 'average horizontal displacement' small. Dives to 100 m should be separated by ~500 m in calm water; dives to 200 m by ~1000 m; and so on. Over the deeper part of the shelf, resolution is much less than the 200 m average reported, so I suggest the authors clarify this point.

The glider description has been improved and detailed: "Ocean gliders are autonomous underwater vehicles which change their buoyancy to dive up through the water column. Without propulsion, this vertical motion is transformed into horizontal momentum using the vehicle's wings, while its pitch controls the forward motion. During the resulting vertical sawtooth pattern through the water column, a wealth of scientific observations are recorded and analyzed here."

The reviewer is right, the distance travelled over ground between dives directly depends on the dive depth, which is now clarified in the manuscript: "The horizontal displacement between two dives increases with the depth of the dive, with median over ground distances from 130m (for dives in 25 - 50m of water) to 1100m (in 150 - 200m of water)."

2. Are salinity measurements from pumped or unpumped CTDs? If unpumped, how significant is salinity spiking in areas of large temperature gradients? How does this affect the scale analysis?

As the large dataset includes deployments from 2008, some of the vehicles were equipped with an unpumped CTD. However, a salinity spike correction is routinely performed in the quality control procedure. We therefore do not expect this common issue to affect the scale analysis presented. It is now specified: "including a salinity spike correction due to the use of unpumped CTDs in early deployments."

3. The definition of the structure function (Eq. 1) could be me clearly presented as 1/2 the mean

square difference between values at a given separation. The empirical formulation for the structure function (Eq. 2) needs more description, particularly the empirical constants.

More explicit descriptions on equation 2 was also requested by Referee 1. We thank the reviewer for the useful suggestion.

For equation 1, variance of [Z(x)-Z(x+h)] has been changed into $\frac{1}{2}\frac{1}{N}\sum(Z(x)-Z(x+h))^2$ *and described as "half the mean square difference between values at a given separation h" following the reviewers' suggestion.*

Equation 2 is now further described in the manuscript:

"In this equation, the power ½ comes from a fourth-root of [Z(x)-Z(x+h)]^2 that reduces the skewness in the distribution, thereby approaching a Gaussian process. The 4th square acts to correct the scale and returns the same units as equation 1, while the denominator adjusts the bias resulting from the whole transformation. This estimate is more robust statistically in the sense that the mean can be applied to the new distribution. Compared to equation 1, the semivariogram is only slightly modified for the highest lags when using equation 2, but the parameters (sill, range and nugget that are investigated in section 3 remain very similar."

4. Page 20105, lines 4-5: Why pairs within 0.1 degrees? Perhaps give the distance in kilometers for clarity.

In order to investigate anisotropy, data have to be constrained in the opposed direction (meridional / along-shelf when analyzing zonal / cross-shelf semivariogram and inversely). We chose 0.1 degrees, as it still allows a sufficient number of measurement pairs (a minimum of 30 valid data pairs per lag was suggested by Journel and Huijbregts, 1978). The manuscript now specifies "0.1^o (~10 km)".

REFERENCE:

Journel, A. G., Huijbregts, C.J., 1978. Mining Geostatistics, Academic Press Inc, London, UK

5. Regarding homogeneity of the statistics: Lumping observations together to calculate structure functions assumes homogeneity in the statistics. I would expect that there is a change in scales ffrom the inner to outer shelf that could perhaps be diagnosed from these observations. Lack of homogeneity in the vertical is more concerning; surely statistics in the mixed layer differ (vertical scale ~ mixed layer depth?) from those in the thermocline (small vertical scale?) and from those below the thermocline (longer vertical scale?).

Homogeneity in the statistics can indeed be issue, in particular in the water column which is characterized by multiple spatial scales. Semivariograms do not identify multiple scales but only the dominant scales, which is why we do not expect to resolve the smallest vertical scales through the thermocline. Considering the good vertical resolution of the dataset, this could probably be addressed using autocorrelation functions, but will require further investigation.

6. There is a good bit of flipping back and forth between 'semivariogram' and ' structure function'; best to pick one and stick with it.

This has been modified throughout the manuscript, keeping the term “semivariogram”.

7. I find the terms ’sill’, ’range’, and ’nugget’ difficult to follow, though the authors make a good effort to clarify them. ’Range’ is particularly troublesome in usages like (page 20108, Line 25) ’mean temperature ranges...’ since range typical means the difference between minimum and maximum value of a variable. Consider not using these particular terms.

Unfortunately “sill”, “range”, and “nugget” are the standard terms used when describing semivariograms. We have clarified the term ‘range’ throughout the manuscript, by replacing it by “decorrelation range”, “semivariogram range” or “length scale“.

8. Page 20113, lines 12-13: this is not a complete sentence.

We thank the reviewer for pointing out this mistake. It now reads: “The length scales calculated here can be used to guide the design of ocean observing systems, in particular to answer questions related to the observation density needed to resolve along and across shore variability in both the physical and biological parameters.“