Overall an interesting paper on an important topic with a storm that has become a wonderful test case for coastal ocean storm interactions. The study is well formed and remains largely focused on storm induced inertial currents. Some minor additions and edits are required, including a more detailed and distinct methods section for the observational data utilized. While the data was generally publicly available, more details on how the authors treated the data for QAQC, or what default QAQC if any they used from the downloaded data is required.

Specific line-by-line Comments:

Line 32 - 35 - while an interesting comment it is disconnected from the current article.

Line 96 - Caroline should be Carolina

Line 99 - What was the vertical gradient in temperature? This is likely more important than the surface/bottom temperature difference.

Line 102 - Schofield et al., 2010 is a reference for Slocum gliders generally, however there are multiple references for the Hurricane Irene specifically (Glenn et al., 2016) being the most prominent.

There is no clear methods/data section, with some of the observational data described within what looks like results sections.
Line 234 - Why is the effective depth assumed to be 2.4m? Is there a reference for this?

Line 236 - Is the accuracy of HF Radar here referring to this dataset in particular or more generally? A more recent publication from Roarty et al., on HF Radar in the region can be found here: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2020JC016368

Line 239 - Were tides removed from the HF Radar fields and model current fields? I believe later in the paper they were, but it’s not clear what was done for this spatial analysis.

Line 266 - Please comment on data QAQC, Glider setup details, or where this information can be found e.g. previous publications or where it was downloaded. I’m assuming it was from the IOOS Glider DAC? Also an additional paper on Hurricane Irene mixing from glider and ROMS data was detailed here https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2017JC012756

And a detailed exploration of pre-storm mixing was carried out by Watkins and Whitt here: https://journals.ametsoc.org/view/journals/phoc/50/12/jpo-d-20-0134.1.xml

Line 290 - How did the maximum N-squared values compare between the observations and glider? It appears in Figure 3c and 3d that the observed N-squared was significantly greater than in the model ahead of the deepening and mixing, but similar during the deepening event?

Line 314 - I’m not clear on the use of the Zhang reference here. Is this referring to tropical cyclone shallow water mixing generally?

Line 315 - Caroline should be Carolina

Line 315 - Over what time-scale did the SST recover to pre-hurricane levels? Off North Carolina there was likely very little Cold Pool water, thus mixing should result in very little cooling. Plots of bottom. Temperature pre-storm from the model will likely show this.

Line 323 - I agree that the model/data mismatches are largely not too critical for the process investigations presented here. I think the strength of stratification is likely the
most important model feature to validate as it can affect the vertical mixing and generation/dissipation of NIC.

Line 342 - 343 - Were data dropouts documented, or could there be dynamical reason that the NIC are in poorer agreement offshore? The HF Radar data should include quality flags to identify missing or low quality data.

Line 378 - 381 - Is the 75m D3 location the beginning of the shelf-break front, a mesoscale feature impinging on the shelf, or simply too far from the main track? Adding the reference lines to additional spatial figures would be helpful for interpretation rather than needing to flip back to figure 1 for the reader.