Review of

Title: Assessing the coastal hazard of medicane Ianos through ensemble modelling
Author(s): Christian Ferrarin, Florian Pantillon, Silvio Davolio, Marco Bajo,
Mario Marcello Miglietta, Elenio Avolio, Diego S. Carrió, Ioannis Pytharoulis,
Claudio Sanchez, Platon Patlakas, Juan Jesús González-Alemán, and Emmanouil
Flaounas
MS No.: egusphere-2022-990
MS type: Research article
Iteration: Initial submission

The paper presents results of an ensemble of barotropic ocean models, coupled to a wave model and a number of atmospheric models, to give a probabilistic forecast of coastal sea levels in Greece during medicane Ianos.

The paper is clearly written and the Figures are clear. There are benefits (and pitfalls) of a large ensemble spread during an extreme event, but the most interesting takeaway to me seems to be that higher resolution is not automatically the holy grail of ocean modeling and will not, in itself, solve the prediction problem. I recommend publication after some minor revisions.

Specific remarks:

page10, L175: can the authors be more specific of what SLA product exactly they use. SLA netCDFs give sla_filtered variable, but according to the variable attribute, this is

sla_filtered:comment = "The sea level anomaly is the sea surface height above mean sea surface height; the uncorrected sla can be computed as follows: [uncorrected sla]=[sla from product]+[dac]+[ocean_tide]+[internal_tide]-[lwe]"

The authors state that they use sla_filtered, "uncorrected for the " dac. Can the authors please provide the exact arithmetic expression of what is included and what is subtracted in SLA that they use?

Page11, L185: perhaps it would be useful for readers to mention that this is a UTIDE package.

Figure 8: a-c: are red lines further filtered after subtraction of tides? I am surprised that UTIDE itself would make such a good detiding job…

page14, L237: yes, mean+stdev could provide a conservative estimate of risk – but they could also lead to so many false positives that the product would cease to be taken seriously by downstream stakeholders. Perhaps this could be mentioned as a downside of such conservative estimates. Figure would be a dramatic false positive if this conservative approach were to be used. There is an compromise to be found between model precision ()how many predicted floods occurred) and its recall (how many occurred floods were predicted). It is not obvious to me that simply adding mean+stdev leads to a good compromise.

Page15, L245: I suppose it would also be interesting to employ deep-learning classifiers instead of binary threshold based methods for this task...