This manuscript presents a new deep-learning storm surge forecasting (HIDRA2) for the Northern Adriatic which is an “updated” version of a previous deep learning architecture HIDRA1. The authors show that HIDRA2 outperforms its previous version HIDRA1 and it also performs better that the Copernicus Mediterranean Reanalysis. The paper is well written, and I think it must be published after some minor changes.

**Major comment:**

The authors compare their new HIDRA2 with CMEMS Reanalysis, that is a 3D model in a regular horizontal grid. I recommend to the authors to compare the performance of HIDRA2 against numerical simulations specifically design to determine the sea surface elevation. As examples, the authors may use the global simulation by Muis et al. (2020) which has, in the Mediterranean, a 1.25 km of coastal resolution; or they can use the newest available simulation performed by Toomey et al. (2022) specifically designed for the Mediterranean Sea with a coastal resolution of 200 m. The latest hindcast includes the wave setup component that maybe relevant when computing the total elevation in the studied area.

**Minor comments:**

- The authors use the term storm surge to refer to the combined effect of atmospheric pressure, winds, and astronomic tides. I strongly recommend the authors to follow the terminology detailed in Gregory et al. (2019). In that paper, they write: “N9 Storm surge: The elevation or depression of the sea surface with respect to the predicted tide during a storm.” And they also write: “Sea-surface height (SSH) can be greatly elevated during a storm by a storm surge, and the consequent extreme sea level is sometimes called a storm tide”. So, the combination of storm surge + astronomical tide should be referred as
storm tide, if we decide to follow the definitions in Gregory et al. (2019).

- Line 17: this line needs, at least, two references where I have included (Ref. XXXX): “Global mean sea level rise, related to anthropogenic climate change (Ref. XXXX), is causing a worldwide increase in coastal flooding frequency and is leading to a myriad of negative consequences for coastal communities, civil safety and economies (Ref. XXXX).”

- Fig. 1: Please change the title of the figure to “Adriatic Sea topo-bathymetry”.

- Line 35-36: put the citation to Medvedev between parentheses.

- Line 48: “HIDRA1 ensemble (Žust et al., 2021) is a million times faster than the operational numerical ocean model ensemble based on NEMO engine”. Although it maybe true, it is not fair to compare the computational time of a 3D model that computes water dynamics (currents, temperatures, salinities, ...) in several vertical layers with HIDRA1 or HIDRA2 that gives SSH information only and in a single point. The authors should compare the computational time with, for example Muis et al. (2020) or Toomey et al. (2022), and multiply their computational time by the number of nodes that the other two studies are computing.

- Fig. 10: it is difficult to appreciate the differences between the different datasets. I would recommend the authors to apply a filter to the spectrum. There may be a newer reference but I usually do it following the Chapter 5- Time-series Analysis Methods (https://www.sciencedirect.com/science/article/pii/B978044450756350006X) from Data Analysis Methods in Physical Oceanography by Emery and Thomson.

Some thoughts:

1) Could this system be scaled up and applied to estimate SSH values at a Mediterranean scale?

2) How does the predictability change if the system is fed with the first predicted day?

References:

Gregory, J.M., Griffies, S.M., Hughes, C.W. et al. Concepts and Terminology for Sea Level:
