The authors developed a coupled atmosphere-ocean model system in the Adriatic Sea. They evaluated the ocean part performance of the coupled model in this manuscript (MS).

It is challenging to develop an ocean model in a dynamic region with complex bathymetry and applied it to a long-term simulation. In MS, the authors implemented the coupled model system for a 31-year simulation. Model simulated SSH, SST, temperature, salinity, as well as current, were validated by the satellite measurements and in-situ observations. Methods they chose for the validation, such as Taylor diagram, MAD, T-S diagrams and so on, are widely used in skill assessment. As the result, the model can reproduce dynamical properties and the general pattern of the variations.

Response: Thank you very much for your review.

Here are comments and suggestions:

- Line 90: Liu et al. (2021) wasn’t listed in the “References”.

Response: Accepted. Will be added to the manuscript.

- Section 2.1: Is the nesting 2-way or 1-way?

Response: The ocean grids are one way nested. The sentence in the text will be modified as follow:

“... (2) the complex coastal Adriatic Sea dynamics with a one-way nested 1-km grid (676 x 730).”
Section 3.1.1: Looks like there is a conspicuous difference in the EOF1 amplitude. Did the authors calculate correlation coefficients between the amplitude of observation and amplitude of the model?

Figure R1: Time variations of the amplitude of EOF1 for both observation and model.

Response: The correlation coefficient between observed and modelled EOF1 is 0.65 and the normalized standardized deviation is 1.19 which shows that the model amplifies the seasonal signal by comparison to the observations (see Fig. R1 below). The manuscript will be changed as follow:

“Overall, it can clearly be seen that, for both Adriatic and northern Ionian Seas (Fig. 2), the first EOF component (EOF1) represents the seasonal variability of both AdriSC ROMS 3-km and JPL_MEASURES results with spatial signal and amplitudes slightly stronger in the model (i.e. 81.2% of the total signal with amplitudes varying between ±8.0; Fig. 3) than in the observations (i.e. 74.5% of the total signal with amplitudes ranging between ±6.0; Fig. 3). Additionally, the correlation coefficient between the time variations of the observed and modelled EOF1 is only 0.65 associated with a normalized standard deviation of 1.19.”

Section 3.1.2: Comparing to the reference, the standard deviation of salinity is quite low (~0.25, Figure 6b). This should be mentioned in MS.

Response: The authors think the reviewer means Section 3.2.1 and not 3.1.2 which only discusses the SST and SSH variabilities. Additionally, the misrepresentation of the salinity is already discussed in the manuscript in section 3.2.1:

“... but do not properly capture the observed salinity (i.e. correlations around 0.7 and normalized standardized deviations between 0.3 and 0.5).”

Section 3.2.1: Median temperature bias reaches almost four degrees in the subsurface (30m, Figure 8d). Can the authors explain why the model has such a large bias in the subsurface?

Response: The authors believe that the observed biases found around the thermocline and/or halocline can be attributed to smooth vertical gradients near the mixed layer base. However, to be fair, as the authors did not systematically tested different parametrizations for the vertical mixing and diffusivity in the ROMS models, they cannot discriminate whether the obtained biases are due to the ROMS model setup itself or the MEDSEA
forcing. It should also be noted that for this specific case, the number of measurements is largely decreased where the extreme bias occurs. This means that, if unrealistic measurements were not flagged during the quality check process, they may weight in far much than when more observations are available. The following paragraph will be included in the text:

"Additionally, independently of the subdomains, the analysis of the vertical profiles shows that the temperature and salinity biases often present a peak in the vicinity of the thermocline/halocline depth which can probably be linked to an inaccurate representation of vertical diffusivity and vertical mixing in the AdriSC ROMS models. However, more in-depth work should be done to discriminate whether the vertical biases are linked to the AdriSC ROMS model set-up per se or to the MEDSEA fields used to force the initial and boundary conditions."

- Section 3.3.1: The correlation coefficient of direction looks very poor in the Taylor diagram. However, in the Q-Q plot, the modeled direction matches observation very well. Can the author explain it?

Response: The difference between Q-Q plot and correlation coefficient can easily be explained by the fact that Q-Q plots compare the distributions of the variable not their value at each time like with the correlation coefficient or the scatter plot. Generally speaking, it is good to keep in mind that it is easier to obtain a matching Q-Q plot than for example a matching scatter plot as illustrated in Figure 11 of the article. In this case the authors attribute the time-phase differences to the tidal representation and the fact that the time of the archived data was not necessarily provided in UTC:

"However, due to the already mentioned lack of synchronization, modelled current speeds and most especially modelled current directions can be extremely spread compared to the observations. Despite the inherent difficulties to reproduce the ocean dynamics at the hourly scale, the scattering of the AdriSC ROMS 1-km results can also result from the uncertainties linked to the observational dataset time references. Indeed, due to the lack of metadata availability for a certain number of datasets, some observations which may have been provided in local time have been compared with model results in Universal Time Coordinated (UTC)."