Authors' response to the Anonymous Referee #2

We thank the reviewer for their time and comments that helped us to improve the paper. Detailed responses to individual comments are provided below point-by-point. We also paste the text to reflect the changes to the manuscript. In summary, two major changes have been made in line with the reviewer's suggestions:

- We have turned off the wind stress parametrization and repeated training and analyses with wind-only input data. Results with raw wind input is indeed better than with the wind stress transformation. We now use raw wind input during training. The manuscript has been modified accordingly.
- We have performed another sensitivity study to check the information redundancy in sea level pressure / wind input data. The HIDRA model was re-trained with wind-only and pressure-only inputs. The best results are obtained when both inputs are provided to HIDRA indicating complementarity of the two inputs and ability of HIDRA to compensate for any potential redundancy.

Comment 1: This manuscript proposed a model named HIDRA based on a deep learning network to forecast the sea level in the Adriatic. This is a good try and shows the deep-learning-based model has a good future in ocean environment research and forecasting. This work is worthwhile and the manuscript is well-written in general.

Response: We thank the reviewer for the encouraging comment.

Comment 2: However, there are still several critical issues to be clarified. The Large and Pond parameterization is suitable for calculating the wind stress over the open sea with deep water. In other words, this scheme cannot use in this study. As the data is the basic and core of machine learning, the authors should find another scheme to redo this work.

Response: We thank the reviewer for this insight. Indeed, Large and Pond formulation has its limits, but is often applied in oceanography as a tradeoff between complexity and accuracy. In our case it is worth pointing out that this particular parametrization was not chosen to most concisely represent the vertical momentum flux at the sea surface (which would admittedly require more complex schemes and more data), but to merely introduce the nonlinear wind stress dependence on the wind. Thus our reasoning was that this would make learning the relationships between the atmospheric conditions and tide gauge sea level easier for the deep network.

However, following the referee's insight, we have performed a number of new experiments. In particular, HIDRA was re-trained with using the raw wind instead of Large and Pond parametrization. The prediction performance did not change overall, which means that the network is capable of modeling, at some level, how wind translates into vertical momentum flux. We did notice that the prediction performance on storm surges did improve on average by approximately 1 cm. These results will be added to the revised manuscript as a separate subsection in the "Hidra Architecture Analysis" Section 4.1. For convenience we include the new subsection below:

"We proceed to inspect the impact of Large and Pond parametrization which might oversimplify the wind stress dependence on the wind. To this end we consider another variant of HIDRA, which uses raw wind instead of wind stress. Results in Table 7 show that, overall, the performance between the two wind-input variants is indistinguishable. However, on storm surges, using raw wind reduces the RMSE by approximately 1 cm when compared to the setup which uses wind stress. It appears that HIDRA is capable of extracting the information important for sea level prediction during storm surges also directly from the raw wind."

Table 7. Performance of HIDRA variants with different wind and pressure input configurations. $HIDRA_0$ uses the wind stress, $HIDRA_{wnd}$ uses the raw wind, $HIDRA_{no\_wnd}$ does not use wind inputs and $HIDRA_{no\_prs}$ uses the wind stress, but not the air pressure. Performance of the astronomic tide is provided for reference.

	MAE [cm]	RMSE [cm]	Bias [cm]	Likelihood
Overall				
$HIDRA_0$	**4.9**	**6.4**	-0.4	**0.0470**
$HIDRA_{wnd}$	**4.9**	**6.4**	**0.2**	0.0465
$HIDRA_{no\_wnd}$	5.3	7.1	-0.3	0.0434
$HIDRA_{no\_prs}$	5.3	7.0	-0.1	0.0451
Reference (tide)	12.1	15.7	-2.4	-
Storm surge events				
$HIDRA_0$	10.3	12.9	-9.3	0.0253
$HIDRA_{wnd}$	**9.3**	**11.6**	**-7.8**	**0.0274**
$HIDRA_{no\_wnd}$	13.7	16.5	-12.9	0.0192
$HIDRA_{no\_prs}$	11.6	14.0	-10.8	0.0225
Reference (tide)	49.6	50.4	-49.6	-

Comment 3: In fact, the wind is associated with sea level pressure and latitude (Coriolis force).For the Adriatic, the difference of wind in different locations due to Coriolis force can be ignored, which means the wind is almost determined by sea level pressure. So, wind stress has included information on sea level pressure. I think it's double-counted when the authors used wind stress as well as sea level pressure.

Response: Thank you for the comment. Indeed, winds in the Adriatic basin have a substantial geostrophic component. But we would like to note that the two strongest winds, Bora and Scirocco, can also have a significant non-geostrophic contribution (orographic channeling, non-linear wave breaking, density-driven flows during weak Bora episodes etc., *e.g.* Pasarić et al., 2007, https://angeo.copernicus.org/articles/25/1263/2007/; Grisogono and Belušić, 2009, https://onlinelibrary.wiley.com/doi/10.1111/j.1600-0870.2008.00369.x). In such cases the wind field can exhibit cross-isobaric flow and reflects other constraints beyond the geostrophic equilibrium.

Thus, to explicitly verify the potential effect of the information redundancy between the wind and air pressure inputs, we trained several additional variants of HIDRA. One that used wind input but not pressure, and another that used the air pressure but not the wind. Results show that performance drops by approximately 10% when excluding either wind or air pressure, implying that the network accounts for potential redundancy and capitalizes on the complementary information encoded in both inputs. These results will be added in the revised manuscript as a separate subsection in the "Hidra Architecture Analysis" Section 4.1. For convenience we include the new subsection below:

"Bora and Scirocco characteristics in the Adriatic basin are often determined through an interplay of geostrophic, orographic and other influences (Pasarić et al., 2007; Grisogono and Belušić, 2009). At other times however, non-geostrophic effects may play a lesser role and the wind field is largely determined by the pressure field. To investigate potential information redundancy between the wind and pressure inputs, two HIDRA variants were trained: one which did not use the wind input and another which used the wind, but not the pressure. Results in Table 7 show that removing either wind or air pressure input leads to an approximately 9% increase of RMSE. HIDRA seems to compensate for potential redundancy in the inputs and capitalizes on the fact that wind in the basin is not entirely pressure driven. In any case using both inputs is preferred."

Table 7. Performance of HIDRA variants with different wind and pressure input configurations. HIDRA_0 uses the wind stress, $\mathrm{HIDRA}_{\mathrm{wnd}}$ uses the raw wind, $\mathrm{HIDRA}_{\mathrm{no\_wnd}}$ does not use wind inputs and $\mathrm{HIDRA}_{\mathrm{no\_prs}}$ uses the wind stress, but not the air pressure. Performance of the astronomic tide is provided for reference.

	MAE [cm]	RMSE [cm]	Bias [cm]	Likelihood
Overall				
HIDRA_0	**4.9**	**6.4**	-0.4	**0.0470**
$\mathrm{HIDRA}_{\mathrm{wnd}}$	**4.9**	**6.4**	**0.2**	0.0465
$\mathrm{HIDRA}_{\mathrm{no\_wnd}}$	5.3	7.1	-0.3	0.0434
$\mathrm{HIDRA}_{\mathrm{no\_prs}}$	5.3	7.0	-0.1	0.0451
Reference (tide)	12.1	15.7	-2.4	-
Storm surge events				
HIDRA_0	10.3	12.9	-9.3	0.0253
$\mathrm{HIDRA}_{\mathrm{wnd}}$	**9.3**	**11.6**	**-7.8**	**0.0274**
$\mathrm{HIDRA}_{\mathrm{no\_wnd}}$	13.7	16.5	-12.9	0.0192
$\mathrm{HIDRA}_{\mathrm{no\_prs}}$	11.6	14.0	-10.8	0.0225
Reference (tide)	49.6	50.4	-49.6	-

Comment 4: Topography is important for the sea level, besides the wind stress (or sea level pressure). Therefore, topography should also be considered in the HIDRA model.

Response: We agree that topography affects the sea level dynamics and is imperative for sea level prediction in physical models. We would nevertheless like to point out that topography is constant and does not change with time, thus it is not necessary to provide it explicitly as an input parameter to HIDRA. Note that HIDRA is not a dynamical model which would have to explicitly take into account the complex per-point topography interaction. Rather, interactions between dynamic spatially-varying and static elements (like topography), relevant for sea level prediction accuracy, are *learned implicitly* by the deep neural network from the vast amount of data. To achieve this in HIDRA, the spatial encoding is enforced by providing the normalized spatial coordinates as part of the input to the network (*x* and *y* input channels), thus allowing the learning algorithm to make such spatially-dependent relations.

Furthermore, HIDRA predicts the sea level for a single specific location, taking into account the past sea level measurements from that location (Koper tide gauge). These measurements already reflect Adriatic basin bathymetry: sea level experiences topographic amplification due to shallowness in the north, resonant amplification due to forcing frequency being close to the basin seiche frequency, and also reflection due to the closure of the basin in the north. All these bathymetry effects are already implicitly contained in the observations that the network receives as the input. In fact, our experiments show that HIDRA can respond in a manner that consistently reflects these bathymetric constraints, for example that it amplifies the signal in the basin seiche $(21.5\ \text{hr})^{-1}$ frequency band during a Scirocco event.

Comment 5: Line 327, the description is inaccurate. Usually, the regional ocean model, especially the coastal model, can simulate the sea level including the tidal directly by using the water level boundary condition.

Response: Thank you for pointing this out. The description will be updated by including NEMO configuration namelist into the paper supplement (the parameter space of the NEMO model is too large to be completely covered in the descriptive text). We also agree that regional ocean models can simulate sea level (including tides) via open boundary conditions.

As pointed out by the referee, tides are not included in the forcing of the current NEMO setup. This decision is partly based on the fact that we have a tide-gauge in Koper and we can analyze the tidal constituents for Koper directly from the local observations, which seemed to be the most straightforward way of obtaining tides in Koper. We have however in the past compared full (with tides on open boundaries) NEMO sea-levels to the setup presented in this paper. The main result (unpublished) was that sea level from NEMO with tides at the open boundary offers comparable, but somewhat worse representation of observed total sea levels than the non-tidal setup of NEMO with tides computed on the Koper tide gauge.

Furthermore, as explained in the paper, HIDRA is using tidal sea-levels and residuals obtained from the Koper tide-gauge. Allowing NEMO and HIDRA to use the same tidal signal allows for more consistent comparisons of their performance. We have therefore chosen not to include tidal forcing at NEMO open boundary (in the Ionian Sea) in this study, but to obtain the tidal part of the sea level signal from local observations.

Comment 6: Moreover, the authors cannot claim HIDRA is better than NEMO because they only compared with results from only one NEMO configuration they used. If the NEMO is tuned carefully, maybe the results are better. In fact, the NEMO in the storm surge events seems better than HIDRA.

Response: We thank the reviewer for this comment. We agree -- there certainly exists a possibility that a better tuned setup of NEMO would produce better results. For example, assimilation of sea level data might be of substantial benefit and our current setup does not have it. Therefore, we must certainly clarify that we do not claim that HIDRA is better than NEMO in general - but rather that HIDRA does compare favorably with the only specific operational setup of NEMO at our disposal.

To make the manuscript more consistent with the referee's arguments, we have amended the manuscript to be very specific that we are referring to the specific operational setup of NEMO at Slovenian Environment Agency. We however fear that a stand-alone NEMO setup sensitivity study would diverge too far from the scope of our paper, which is to present a numerically cheap machine learning architecture which can compete with (and most often outperform) a much more complex and numerically demanding general circulation model.

Minor comments:

Comment 7: How to deal with the land points in this study is missed.

Response: HIDRA does not distinguish between wet and dry points - it focuses on the synoptic pattern of surface meteorological fields. It does however receive spatial encoding of atmospheric fields (x and y input channels). Again, as in topography, ECMWF land/sea mask is a constant field and as such cannot profoundly impact gradient descent during the learning process. To make this clearer, the following sentence was added to the manuscript:

"Atmospheric fields over land and sea are treated in the same manner, i.e. while HIDRA does receive an explicit spatial encoding of atmospheric fields (Section 3.1.1), it does not employ a land/sea mask."

Comment 8: Why did the authors select the 29x37 for the atmospheric tensor?

Response: The spatial size is specific to the input ECMWF ensemble grid. The ECMWF ensemble forecast over the domain of the study contains 73×57 grid points. As described in Section 2.3, we further downscale the data by a factor of two and end up with a grid of dimensions 37×29. Following the standard convention, spatial maps are represented as height-first tensors, thus 29×37. To clarify the origin of the atmospheric tensor dimensions we changed a line in Section 2.2 to

"… In this study, the following forecast fields were subset to the Adriatic basin, represented by a 73×57 spatial grid (see Figure 2) …"

and a line in Section 2.3 to

"The data is standardised and global average pooling is used to reduce the dimensionality of the atmospheric data -- spatial dimension of the data in samples is reduced in half, from 73 x 57 points to 37 x 29 points, and the temporal dimension is reduced by a factor of 4."

Comment 9: It's better to give a table for the HIDRA and NEMO configuration

Response: As pointed out in our response to Reviewer's Comment 5, we now include the NEMO configuration namelist in the supplementary material. Apart from the architectural design of HIDRA specified in Figure 3, the only free parameters are the learning rates and batch sizes summarized in the last paragraph of Section 3.1.4.