Review of the manuscript "Intraseasonal prediction of monthly storminess in the North Sea with the ACE2 atmospheric emulator and Random Forests" by Aziz et al.

The present manuscript analyses the predictability of monthly storminess in the North Sea in winter using two different approaches: a machine learning algorithm (Random Forests Regression, RFR) and a climate model emulator (ACE2 trained on ERA5 data). The storminess metric is defined based on 10-m wind at a grid point located in the North Sea.
The authors identify the lower stratosphere in December and, in particular, the wind and temperature at 70hPa, as a primary driver of North Sea storminess in January. Indeed, in the ACE2 emulations, modifying the stratospheric conditions on December 1st to reflect a cooler, stronger polar vortex successfully increase the emulated January surface wind speeds in the North Sea. This result highlights the dependence of storminess on stratospheric polar vortex conditions. Similarly, the RFR, which considers polar temperature and zonal wind at 70hPa along with geopotential height at 200hPa in December, demonstrates a correlation of 0.55-0.60 between predicted and observed North Sea storminess in January. This suggests that large-scale stratospheric/upper tropospheric conditions explain roughly 30% of the observed variability in storms.
The authors also extend the RFR analysis to other consecutive month pairs, yielding lower predictability for storminess. They attribute this finding to the seasonal cycle of stratosphere-troposphere coupling.

Overall, the work is interesting for the scientific community. It employs very novel tools to assess the subseasonal-to-seasonal (S2S) predictability of storminess in the North Sea and identify new drivers. However, I have detected several important caveats in the analysis that should be addressed before the manuscript can be recommended for publication.

General comments

1) Identification of predictors of storminess: My first comment refers to the identification of predictors of storminess in the North Sea. First, the authors indicate that the spatial coherence of storm data is short, and they mention a global map of correlations between the number of January stormy days in the North Sea grid with other grid cells (but I could not find this map). Secondly, the temporal persistence of the monthly number of stormy days is shown to be very low. From these results, the authors infer then that SST anomalies (remote or in the North Sea) do not play a relevant role in the storminess due to their long persistence and large spatial scale. Although I agree with the large spatio-temporal scales of SST anomalies, I think it would be important to show SST plots confirming their lack of influence on North Sea storminess (for instance, correlation/regression maps of global SSTs (with different lags) on the storm index or composite maps of global SST anomalies for winters with a high or low number of storms (as in Figure 3),...). I think this is particularly important for ruling out the effects of remote SST anomalies, as they influence remote regional atmosphere through teleconnections that can change with time, even at intraseasonal time scales. For instance, ENSO influence on the North Atlantic circulation in early winter is different from that in late winter (e.g.: Ayarzaguena et al. 2018).
Another suggestion would be to include SST indices of specific well-known regions such the Tropical North Atlantic or El Niño 3.4 among others as predictors in the RFR model and see if results change significantly.

2) ACE2 experiments: I have a couple of comments of different kind regarding ACE2 experiments.
First, although the authors provide a long description of ACE2 experiments, some key aspects or results are still missing. For instance, ensemble members are mentioned in L223, but there is no information about the number of ensemble members that the authors use in this study (maybe only 6?). Moreover, the authors indicate in the abstract that apart from the emulations initialised on December 1st, they have also performed others initialised on November 1st that fail to produce a meaningful response in the ensuing months. I think this result should be also shown in the results section.

Another aspect related to ACE2 experiments concerns the variables used for the perturbated emulations. Based on Figure 5 and considering that the stratospheric circulation is highly zonal, I am not sure if the authors obtain much additional information when including the meridional wind at 70hPa. Indeed, in the case of the RFR model, only the zonal component is used. Have the authors tried geopotential height at 70hPa instead of the two wind components? I am also curious to see the results when using stratospheric variables at higher levels in both ACE2 experiments and the RFR model analysis. This could even provide a longer time window for predictability skill.

3) Storminess index: As previously mentioned, the North Sea storminess index is based on the wind in a single grid point. It would be important that the authors justify why they are selecting this specific point. Moreover, the index might not be very robust as it is based on a single point. Would it make sense to perform an average of the wind over the area? Or define stormy days when there is a specific number of grid points under stormy conditions? I think it is worth testing the sensitivity of the results to the definition of the storminess index.

Specific comments
L 49 and L 50 and later on: weaker/ stronger NAO → negative/positive phase of NAO.

L62: SSWs are triggered by the sustained dissipation of waves. Although gravity waves are also important, it is Rossby waves who play the most relevant role in the generation of SSWs.

L65: do the authors mean major SSWs?

L69-70: It is not clear the study that the authors are referring to here.

L73-75: Some studies such as Polvani et al. (2017) have shown that the high frequency of SSWs found during both phases of ENSO in observations was related to a short observational record. In model simulations with longer data record, only El Niño phase shows a higher frequency of SSWs than during the neutral phase.

L87: If it is the isobaric 70hPa surface, then please remove geopotential height level.

L112-115: Mouallem et al. (2025) is mostly focused on the generation of SSWs in an idealized model. Papers such as Kautz et al. (2020) or SNAPSI papers (Hitchcock et al. 2022, Day et al,

2025) fit better with the purposes of this study, as they use simulations where the stratospheric state is nudged to specific stratospheric conditions to assess their effects on the troposphere.

Table 1: I guess none of these values are statistically significant, but it would be good to indicate it.

L176: In Figures 4 and 5, the authors are not only showing zonal wind but also meridional wind. Following my second major comment, I would focus only on zonal wind and remove Figure 5.

Figure 6 shows a different area with respect to Figure 2 and since it does not have latitude and longitude labels, it is not easy for the reader to identify the region of study. I would suggest showing a larger area as in Figure 1.

L 308- 309: Westerly anomalies at 70hPa around 55º-70ºN are associated with a weak vortex. If the authors would like to refer to an intensified vortex, then I would suggest reversing the colorbar in Figure 10 to avoid confusion.

Figure 10 caption: Temperature at 70hPa is shown in the bottom row.

L329-330: The strongest downward coupling occurs from December to February. Kidston et al. (2015) mention November to January as a common period of active stratosphere-troposphere coupling in both hemispheres.

Technical comments

L63 and later on: Baldwin et al. (2020) → Baldwin et al. (2021)

Figures 3 and 4 caption: Please remove “temporal mean”

L236: cool → cold

References

Ayarzaguena B., Ineson, S., Dunstone, N. J., Baldwin, M. P. and Scaife A. A.: Intrasesonal effects of El Niño Southern Oscillation on North Atlantic Climate Journal of Climate, 31, 8861-8873, doi:10.1175/JCLI-D-18-0097.1, 2018

Dai, Y., Hitchcock, P., Butler, A. H., Garfinkel, C. I., and Seviour, W. J. M.: Assessing stratospheric contributions to subseasonal predictions of precipitation after the 2018 sudden stratospheric warming from the Stratospheric Nudging And Predictable Surface Impacts (SNAPSI) project, Weather Clim. Dynam., 6, 841–862, https://doi.org/10.5194/wcd-6-841-2025, 2025.

Hitchcock, P., Butler, A., Charlton-Perez, A., Garfinkel, C. I., Stockdale, T., Anstey, J., Mitchell, D., Domeisen, D. I. V., Wu, T., Lu, Y., Mastrangelo, D., Malguzzi, P., Lin, H., Muncaster, R., Merryfield, B., Sigmond, M., Xiang, B., Jia, L., Hyun, Y.-K., Oh, J., Specq, D., Simpson, I. R., Richter, J. H., Barton, C., Knight, J., Lim, E.-P., and Hendon, H.: Stratospheric Nudging And Predictable Surface Impacts (SNAPSI): a protocol for investigating the role of stratospheric polar vortex disturbances in subseasonal to seasonal forecasts, Geosci. Model Dev., 15, 5073–5092, https://doi.org/10.5194/gmd-15-5073-2022, 2022

Kautz, L.-A., Polichtchouk, I., Birner, T., Garny, H., and Pinto, J. G.: Enhanced extended-range predictability of the 2018 late-winter Eurasian cold spell due to the stratosphere, Q. J. R. Meteor. Soc., 146, 1040–1055, 2020.

Polvani, L. M., Sun L., Butler A. H., Richter J. H., and Deser C.: Distinguishing Stratospheric Sudden Warmings from ENSO as Key Drivers of Wintertime Climate Variability over the North Atlantic and Eurasia. J. Climate, 30, 1959–1969, https://doi.org/10.1175/JCLI-D-16-0277.1, 2017.