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

Overall Assessment

The authors explore the statistical relationship and effective skill provided by lower stratospheric initial conditions over the Arctic in December (anomalies in 70 hPa temperature and wind, together with the 200 hPa geopotential height) upon storminess over the North Sea a month later (i.e., throughout the following January). They use an established machine learning method (random forest regression), after applying PCA as a means of dimensionality reduction, and compare this with output from a novel and more sophisticated weather model emulator (ACE2), trained on ERA5 over an extended period (1940-2024). The authors test other monthly pairs (always lagged by one month) during the course of the extended winter season but only found December-January to yield a notable statistical linkage.

The authors devise a storm index each year, simply by counting stormy days as those above the 90th percentile according to climatology, for a representative grid cell in the North Sea basin. Taking a small sample of the least stormy months, the stratospheric initial conditions on 1st December are perturbed for an ensemble of ACE2 forecasts, in order to demonstrate the role of a strengthened lower stratospheric polar vortex at the start of boreal winter upon North Sea January storminess.

Overall, I think the study is interesting, novel and important in helping to further our understanding of the role of the stratosphere in driving storminess and surface wind extremes over the North Sea region. However, I recommend a major revision prior to publication in *Weather and Climate Dynamics*. I believe the authors could do more to quantify the enhanced risk and shift in the number (possibly also intensity) of extreme windy days due to the early December initial conditions. I'm also unsure about some of the choices made such as selecting a single grid cell as opposed to multiple grid cells over a larger target region. The representativeness of the grid cell selected for the wider region seems justified to me, only over a relatively small region of the central-northern part of the basin, and fairly distant from land where most societal and economic impacts would occur? I suggest reconsideration of this approach in favour of a wider region.

I outline my comments below which I hope the authors find useful.

General Comments

December-January Linkage: It's interesting that this combination was the only pair found to yield a notable statistical linkage. This appears to be at odds with other studies applying alternative approaches such as nudged hindcast experiments for attribution purposes. For example, Williams et al. (2025) shows a robust linkage between stratospheric conditions in January 2022 and the notably stormy February 2022 affecting the UK and northern Europe. Although only performed for one such case study, many other examples exist within the ERA5 record in which late winter storminess might conceivably be linked to the stratosphere (e.g., February 2020).

Could the result the authors find therefore perhaps reflect some limitation of the ACE2 model in failing to capture such relationship, maybe owing to the training algorithm or the number of years of training data considered? It would be good for the authors to discuss their results within the context of earlier published literature using such alternative approaches.

Figure 6: Whilst this Figure is interesting, I think it would be interesting for the authors to probe further the output from these modelling experiments. I would be interested to know how shifted the top 10th percentile of the distribution is in the perturbed relative to the control case. Is there a shift in the PDF of all the days between perturbed and control?

Again, by taking multiple grid cells into consideration, such evaluation becomes more powerful due to the increased sample size, and the result can be deemed more robust. So, I strongly encourage the authors to consider multiple grid cells, particularly if considering such analysis which would help increase the importance of the study.

Using days 7-60, the difference in the number of days above 90th percentile according to the control case, between perturbed and control, could also be assessed. What is the risk enhancement of a 90th percentile or above day due to the effective strengthening of the lower stratospheric polar vortex? Spatial maps could be an option for presenting this.

ACE2 Experiments: If the experiments are fast and easy to run, would it have made sense to consider other initialisation dates besides December 1st? A lagged ensemble of dates could have been (or perhaps be considered in a future study) by maybe taking several of the initialisation dates either side of 1 December.

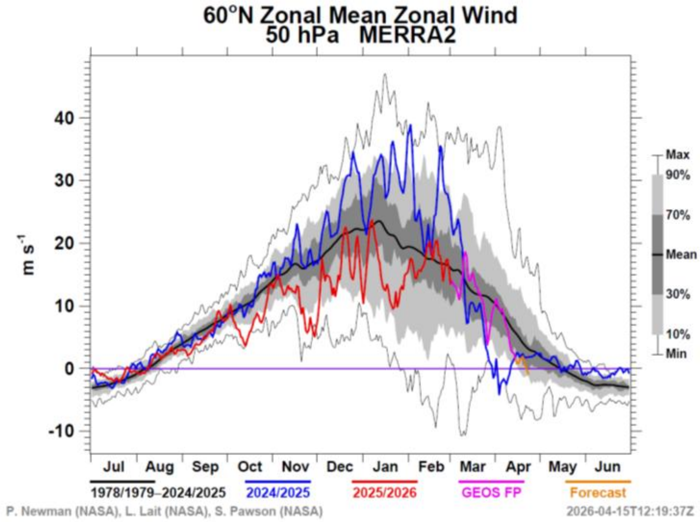
However, in doing so, the each initialisation date (ensemble member) cannot be considered fully independent?

Specific Comments

L11-12: If emulated mean January surface wind speeds increase following modified 'stratospheric initial conditions on 1st December', what exactly is this in response to? Decrease in 70 hPa air temperature and/or increase (decrease) in zonal (meridional) wind?

Or is this just an upper bound concerning possible surface wind speed increase that could result? I think it should be clarified.

L23-24: The intensity of the polar vortex most commonly attains maximum strength in January but interannual variability is largest from January to March (e.g., below).



So, I'm not sure this sentence is quite correct. Downward coupling of stratospheric anomalies to the surface tends to maximise in mid-to late-winter in accordance with larger swings in the strength of the polar vortex?

L54-55: 'Scaife et al. (2014) used a new long-range forecast system…' → Can it be considered that 'new' given the study is now over 10 years old?

L72-76: Citations missing here. I suggest the authors expand more upon the complexity of the role of ENSO on European winter weather, including its evolution from late autumn into late winter/early spring.

For instance, Thornton et al. (2022) which discusses the importance of an early winter teleconnection via a tropospheric pathway that project strongly onto the East Atlantic Pattern, whilst the stratospheric teleconnection pathway is relatively inactive (Ineson and Scaife, 2009). The role of ENSO changes over winter (in part due to the increasing role of the stratospheric pathway), which projects more strongly onto the NAO compared with the EAP into winter.

L73-74: This may be a true result according to reanalyses such as ERA5, but it has been shown that this result might be affected by a limited sample of years. See Ineson et al. (2024) for example, where they provide evidence using hindcast experiments that this finding might be due to chance.

With a bigger sample of years, a difference in SSW frequency between El Niño and La Niña states would perhaps emerge, in accordance with the known ENSO impact on the boreal wintertime polar vortex strength.

L112-114: Here, other such approaches could be mentioned, including stratospheric nudging protocols, such as SNAPSI, for assessment of predictability and attribution (Hitchcock et al., 2022).

L116-120: The content of some but not all sections are explicitly outlined here. What can the reader expect in the rest of Section 4 and Section 5?

L133-134: Some further information of the 90th percentile definition would be beneficial. For instance, is this estimate daily evolving, or calculated using all days across the month? If daily evolving, did the authors experiment in anyway by applying moving average or filters to optimise this? If not, was it tested that the difference between the 90th percentile from day to day is somewhat gradual in its evolution (if it jumps about a lot from day to day then this would suggest artefacts due to sampling could be influential)?

Although I appreciate the results are unlikely sensitive to the exact choice picked, given the relatively large sample size considered?

L135-138: I am curious as to why the authors decide to evaluate one grid cell only and not consider multiple grid cells across a wider area. Was this opted for due to computational limitations? It is not clear what the justification is for doing this.

L147-149: Shown in Fig. 1b? Mentioning 'global map' confused me at first.

I have a slight concern that the short 'coherence length' implies the result reported in this study is not necessarily applicable to the whole North Sea basin but rather only a limited central northern part of the North Sea? Or is this unfair?

L161: '…configurations of SSWs…' → What is meant by this? Different configurations of vortex morphology or perhaps certain types of events (e.g., split v displacement or non-downward/downward propagating etc.)?

L163-164: It would be good for the authors to expand on this. Dependency on prior weather regimes driven by intraseasonal modes of influence such as the MJO? MJO is just one example but a good one to mention in this context as studies suggest stratosphere-troposphere coupling is strongly MJO dependent (e.g., Yadav et al., 2024).

L171-173: I think it would be good to expand by acknowledging however that the spatial pattern of the anomalies differs markedly between across each of the top 5 stormiest and no-storm Januaries.

Figure 3 caption: A good idea to state that these anomalies correspond to high and low storm number Januaries for the region of the North Sea of focus in this study.

L176-178: Figs. 4 and 5 are for both zonal and meridional wind but only zonal is mentioned?

It seems that the meridional wind component anomaly pattern is less important in terms of difference between most and less stormiest Januarys, which is not surprising. Although it is interesting a negative region tends to emerge in the North Atlantic sector (anomalous northerly flow and greater tendency towards cyclonic pattern over Scandinavia)?

L179: 'We also include the December monthly mean of geopotential height anomalies at 200 hPa…' → I don't see it anywhere. Did the authors mean to include it?

L223: '…ensemble members…' → How many were considered?

L230-235: Again, I'm left wondering how many members are considered for each of the "control" and "perturbed" ensemble. Is it 6 for each, taking the Decembers shown in Figures 3-5 on the bottom row? There is only 5 for such cases shown, however.

Or were the initial conditions for each December from ERA5 also perturbed (if so, it is not explained)?

L267: '…years with missing data…' → What missing data? Surely ERA5 is continuous for the predictor variables considered over the full record (1940-2024)?

L278: '…1948, 1949, 1950, 2000, and 2020…' → How were these 6 Decembers (also including 2001) decided upon for analysis? I note that they are different apart from 1950, as the identified five least stormy Januaries, shown in the bottom row of Figures 3-5.

Its not clear to me why different Decembers were used to those identified earlier. The authors need to add more information to explain their reasoning.

L282: 'December 2001' → In what way is this a slightly different situation to the other five? More storms occurred in January 2002? Needs explaining.

Unless the authors mean to highlight this case as an example of which the perturbation applied to a December preceding a January largely absent of storms just so happens to result in a decrease in mean surface wind (perhaps pointing to the role of internal variability which can still occasionally result in the opposite-signed response?). In any case, it's unclear to me.

L283-284: 'We only see a modest increase, mainly south of Norway in the North Sea.' → But more widely a decrease of ~0.5 to 1 m s^{-1}?

L285: 'There are also other contributing factors, but they are not addressed in this study.' → This sentence reads as though the authors know what these were. I think its not so straightforward to necessarily know what these were, but examples of suspected factors could be highlighted (e.g., did an active phase of the MJO occur prior to January 2002 that would counteract the role of the strong polar vortex?).

Figure 6: Why is there a difference in domain considered compared to Fig. 2? I think the rationale needs explaining.

I'm not sure how lacking in storm activity the subsequent Januaries are apart from following December 1950 (which was earlier stated as containing zero throughout the month). If alternative Decembers to those shown in Figs. 3-5 were picked, it would be good to state the counts of storms for these (are they non-zero, albeit a low number?).

Perhaps the biggest wind speed increase is therefore an unsurprising result in the perturbed case following Dec 1950 which had zero storms?

Figure 7: Would it not be meaningful to display what years 0 to 25 actually correspond to on the x-axis?

Figures 7-8: It is explained that Fig. 8 is for the test set of data (30 %). Given the correlation is the same for both, I assume Fig. 7 is also for this same test set of data. It is not explained for Fig. 7.

L305-309: It would be helpful to describe each row in turn or re-order the rows in Fig. 10 in accordance with the text.

Figure 10: I would suggest inverting the colour scale to match the direction of the PC loading according to its physical representation. In other words, strong vortex favours negative 200 hPa GH (blue), positive U at 70 hPa (red) and negative TA at 70 hPa (blue).

L311-312: I suggest making it clearer that this is a typical response over northern Europe, but not necessarily for central and southern Europe. For instance, Afargan-Gerstman et al. (2024) showed increased storm activity following SSWs here, using subseasonal reforecast experiments (e.g., their Fig. 3).

However, the footprint of strongest wind speeds tends to occur on the equatorward side of the main storm track. So, following weak polar vortex (or SSW) events, regions such as Iberia are at increased risk of surface wind extremes (e.g., Ayarzagüena et al., 2018; Utrabo-Carazo et al., 2024).

Table 2: The correlation values should always be closer to 1 or -1 than R^2. The values only seem to match up for Nov-Dec and Dec-Jan? The correlation for Nov-Dec is not expressed as a range. Needs fixing.

L328-330: I'm not sure this reflects the finding in Kidston et al. (2015). I would argue their Fig 1a implies the most pronounced surface impacts occurs in January-February (i.e., towards late winter). This fits with the relatively inactive stratospheric pathway in early winter reported upon, at least associated with ENSO (Ineson and Scaife, 2009).

L371: It would be good to elucidate other likely factors. Were there any particular drivers relevant during the early 2001/2002 winter that might have favoured more quiescent conditions in the North Sea region? The importance of internal variability, which is inherently large, should also be stressed.

Technical Remarks

L33: '…perhaps easier…' → '…more skilful…' perhaps?

L62: 'SSW…' → 'SSWs…'

L69-70: 'That study…' → 'Kidston et al. (2015)…'?

L71: '…making an accurate representation of the stratosphere essential…' → 'making an accurate representation of the stratosphere also essential…'

L74: 'SSW…' → 'SSWs…'

L140: 'e.g.' → Suggest this and rest of sentence is put in parentheses for readability.

L159: 'SSW…' → 'SSWs…'

Figure 10 caption: Suggest reordering mention of each variable in turn down the page (i.e., GH, U and TA).

L316: '…introduction 1' → Delete '1'.

L363: '?' → Should there be a missing citation that needs adding here? If not, delete.

Additional References

Afargan-Gerstman, H., Büeler, D., Wulff, C. O., Sprenger, M., & Domeisen, D. I. (2024). Stratospheric influence on the winter North Atlantic storm track in subseasonal reforecasts. *Weather and Climate Dynamics*, *5*(1), 231-249, https://doi.org/10.5194/wcd-5-231-2024.

Ayarzagüena, B., Barriopedro, D., Garrido-Perez, J. M., Abalos, M., De La Cámara, A., García-Herrera, R., … & Ordóñez, C. (2018). Stratospheric connection to the abrupt end of the 2016/2017 Iberian drought. *Geophysical Research Letters*, *45*(22), 12-639, https://doi.org/10.1029/2018GL079802.

Ineson, S., Dunstone, N. J., Scaife, A. A., Andrews, M. B., Lockwood, J. F., & Pang, B. (2024). Statistics of sudden stratospheric warmings using a large model ensemble. *Atmospheric Science Letters*, *25*(3), e1202, https://doi.org/10.1002/asl.1202.

Ineson, S., & Scaife, A. A. (2009). The role of the stratosphere in the European climate response to El Niño. *Nature Geoscience*, *2*(1), 32-36, https://doi.org/10.1038/ngeo381.

Hitchcock, P., Butler, A., Charlton-Perez, A., Garfinkel, C. I., Stockdale, T., Anstey, J., … & Hendon, H. (2022). Stratospheric Nudging And Predictable Surface Impacts (SNAPSI): a protocol for investigating the role of stratospheric polar vortex disturbances in subseasonal to seasonal forecasts. *Geoscientific Model Development*, *15*(13), 5073-5092, https://doi.org/10.5194/gmd-15-5073-2022.

Thornton, H. E., Smith, D. M., Scaife, A. A., & Dunstone, N. J. (2023). Seasonal predictability of the East Atlantic pattern in late autumn and early winter. *Geophysical Research Letters*, *50*(1), e2022GL100712, https://doi.org/10.1029/2022GL100712.

Utrabo-Carazo, E., Lockwood, J. F., Dunn, R. J., Minola, L., Aguilar, E., & Azorin-Molina, C. (2024). Effects of extreme stratospheric polar vortex events on near-surface wind gusts across Europe. *Environmental Research Letters*, *19*(9), 094044, https://doi.org/10.1088/1748-9326/ad67f4.

Williams, R. S., Maycock, A. C., Charnay, V., Knight, J., & Polichtchouk, I. (2025). Strong polar vortex favoured intense Northern European storminess in February 2022. *Communications Earth & Environment*, *6*(1), 226, https://doi.org/10.1038/s43247-025-02175-7.

Yadav, P., Garfinkel, C. I., & Domeisen, D. I. (2024). The role of the stratosphere in teleconnections arising from fast and slow MJO episodes. *Geophysical Research Letters*, *51*(1), e2023GL104826, https://doi.org/10.1029/2023GL104826.