

Editor guidance, and replies

As you saw, there is some misunderstanding between you and the two reviewers regarding your line of arguments. Regretfully, I find a key part of your responses to the comments not very clear, and hope that you can provide a convincing line of thoughts in a revised version of the manuscript.

- The author thanks the editor for his assistance in understanding feedback from reviewers.
- A lot of the confusion stems from the scope of this manuscript. The aim is to focus the manuscript on the AA forcing of European storminess, rather than a study of all forcings of the climate. The fourth and fifth paragraphs of the revised Introduction provide explanations on why a scope limited to AA forcing is valuable, and why a study of the full forcing of multidecadal storminess is of diminished value at this time.
- The revised Introduction also has less description of other forcings, as requested by the editor and reviewers, and consistent with the targeted scope of this study. It is hoped that the Title, Abstract, Introduction and Conclusions of the submitted manuscript clarify the scope of this study.
- Note that the responses to Referees uploaded on 13th April 2026 have now been substantially revised in the new file upload.

Maybe my view of your manuscript helps:

The abstract references the observed time series of windstorm-related loss in Europe and its high values during the 1980s and 1990s, asking about a potential cause. It tells that you could assign a share of about 45% in the increase from preindustrial time to the late 20th century to the (changes in) AA forcing. This is, of course, based on the DAMIP results. You contradict this view in your responses – in particular with respect to the demand for a comparison with a full transient forcing experiment by reviewer 2 .

- Thank you, your view helps to clarify some of the issues.
- One source of confusion seems to be the mention of high values of losses in the 1980s and 1990s in the first sentence of the Abstract, and first paragraph of the Introduction. These sentences serve to provide context, on the importance of multidecadal loss variations. However, the main aim is to describe impacts of AA forcing, and assess the quality of model results (Section 4), rather than provide a full explanation of the high values in the 1980s and 1990s. For example, the Abstract describes how DAMIP models indicate European wind losses were boosted by 45% in the late 20th century relative to preindustrial times. It does not find that 45% of the increase from preindustrial to the late 20th century is due to AA forcing.
- The Introduction has been revised to discuss the value of studying AA-forcing alone, in the fourth and fifth paragraphs. Also, this new submission contains revisions to my reply to the reviewer's suggestion to study full-forcing experiments.

Of course, AA forcing is just one of the many forcing agents affecting the atmosphere, and additional internal variability occurring because of the climate system properties. All these factors are influencing the occurrence and intensity of windstorms and eventually European storm damage simultaneously. You will be aware of the fact that AA forcing can result in different systematic spatio-temporal anomalies when it is the only factor, compared to its effect with a combined forcing with other varying forcing agents. This is, to my understanding, why the reviewer suggested also to show the results from both the full transient forcing and variable AA forcing only runs. Thus, you should either include results from a full transient forcing run (what I would find appropriate), or clearly explain why you do not want to show it, and in how far the results are still valuable.

- The main concern about full-forcing experiments is that they have documented weaknesses in their simulation of volcanic and solar forcing, and internal variability too. The reviewer stated that the full-forcing simulations should be compared to observed historical data, so that confidence in model simulations can be established. However, this is a poor validation framework. For example, the full-forcing experiment may produce different storm anomalies from observed, but this does not mean that modelled AA forcing is different from observed, since the model full-forcing errors may be due to its simulation of volcanic or solar forcing, or internal variability. Instead, this study presents results from AA-forced simulations, then assesses the quality of these model simulations in Section 4, to provide the reader with valuable information on AA forcing.
- There are many published studies that focus on one particular forcing of climate anomalies, or one particular climate process, and they are valuable. The ability of climate models to isolate forcings, or processes, is regarded as one of their great benefits.
- The reasons why the scope of this study is confined to AA forcing, and the value of such a scope, is discussed in the revised Introduction.

Your Fig. 2a confirms an increase of the loss towards the end of the 20th century in the DAMIP AA only simulation, with just two of the models responsible for the mean result. Of course, the question remains open why it is these two models that react with a damage signal to AA forcing. I think you could discuss if this is related to a simulated anomaly of the geopotential height gradient at a location which is not relevant for European windstorms, or if there is no signal produced by a model.

- Figure 4 of the revised manuscript shows maps of geopotential height anomalies for each climate model ensemble, and the associated text mentions how loss changes are consistent with gradients in geopotential height anomalies. Specifically, the four models with positive signals (significantly different from zero at the 0.1% level) also have steeper gradients of geopotential height anomalies over Europe, and the two models with non-significant loss signals do not have steeper gradients.
- The reader can get insight into the variability of storm response to AA forcing between climate models in the revised Section 4. In brief, based on the findings of published research, AA forcing caused a modified gradient of net radiation at the top-of-atmosphere, which is compensated by anomalies in both ocean and atmosphere heat transports. At northern mid-latitudes, the ocean heat transport anomalies mainly concern AMOC strength, while the atmosphere heat transport anomalies are via storm track intensity. Essentially, models with more responsive storm tracks have smaller AMOC changes, and vice-versa (see Table 2). The revised Section 4 proceeds to evaluate the AMOC responses to AA forcing, to identify those models with more realistic storm track changes.

With respect to the potential role of AMOC, you should clearly take the comments from reviewer 1 into account, in particular pointing at potential delays of an AMOC reaction to forcings.

- The discussion of AMOC has been removed from the revised Introduction, following the advice from the editor and both reviewers to shorten discussion of all forcings.

Finally, you clearly should obey the wish for clarity with respect to definitions, and considerably shorten and structure the section on other forcing agents.

- The discussion of other forcing agents has been considerably shortened in the revised Introduction, and now reflects the focus on AA forcing in this study. The manuscript has been reviewed for clarity of definitions, and revised appropriately.

Referee 1

Summary: The manuscript analyses ensembles of simulations from five general circulation models covering the period 1950-present, driven solely by changes in anthropogenic aerosol (AA) forcings. The goal is to identify the AA's signature in the economic losses from high winds across Europe. These simulations are compared with control simulations, with constant external forcing set at preindustrial levels.

The main conclusion is that the AA signal in wind-related losses is clear, although the exact mechanisms are hampered by high model spread.

Recommendation: In my opinion, the tesseract question is interesting, and the results could be useful, despite the large model spreads. However, I also see several aspects in the manuscript that need clear improvement. My main concerns include the statistical analysis and several paragraphs concerning the description of internal and externally forced variability and the role of the AMOC, which, in my opinion, are quite unclear.

- The author thanks Referee 1 for feedback on the manuscript. Their suggestions have improved the quality of the manuscript, in terms of its structure and content. In particular, the descriptions of multidecadal forcings have been greatly condensed in the revised Introduction, in line with feedback from editor and both reviewers, and the descriptions of statistical methods are now more detailed, and gathered into a new subsection in the revised manuscript.
- On a minor note, this study analysed results from six climate models ensembles of simulations covering the 1850-2014 period (see Table 1 of manuscript).

Main points

1) Basically, the only statistical analysis conducted to identify the impact of AA on wind-related losses is a t-test to assess changes in the mean between two periods. I have several issues with this approach:

- The AA is not constant in time, and so the impact of AA on wind-related losses should display a comparable time evolution. I am aware that the complexity of the mechanisms involved, including the purported impact of AA on the AMOC and its feedback on storminess may involve some temporal lag, but from the model time series shown in Figure 2b only one model shows a time evolution can be roughly compared to the time evolution of the short-wave downwelling radiative forcing shown, for instance, in Hassan et al. The SWR displays a very clear maximum in the last 2 decades of the 20th century, but the time series of losses shown in this manuscript (Figure 2b) does not show, by far, this type of behaviour, with the possible exception of the model CanESM5.

Therefore, I think a more sophisticated statistical methodology should be applied to identify the AA signal in wind-related losses. A stationary t-test is not enough

- The reviewer expected the model-mean European storm loss anomalies in Fig. 2b to co-vary with the multimodel-mean shortwave radiative forcing in the northernmost Atlantic shown in Hassan et al. (2021; hereafter H21). A lack of consistency leads the reviewer to conclude that the significant forcing reported by t-tests may be misleading, which in turn motivates their request for a more sophisticated statistical test.
- The lack of consistency between these two quantities is to be expected, since they are fundamentally different quantities:

- H21 presented SWR anomalies over a limited area of the northernmost North Atlantic, and such a quantity has an uncertain relation to European storm losses, since anomalies in other parts of the climate system will exert influence on European storminess
 - The quantities shown in Fig. 2b represent individual model-mean signals, hence they contain larger sampling variability than the multi-model means used in H21. Therefore, internal climate variability has smaller amplitude in the multimodel-mean used in H21
 - Different sets of climate models are analysed in H21, compared to this study
- As a result, it is difficult to draw any conclusions based on comparing the windstorm loss anomalies in this study to the net downwelling shortwave radiation at the surface over the northernmost Atlantic in H21
 - Nevertheless, a review of the two types of statistical testing used in the manuscript has been conducted, inspired by the reviewer's interest in this aspect of the study.
 - The multimodel mean signal (Fig. 2a) was assessed statistically using its standard error. It was found that the multimodel loss anomaly is more than ten standard errors above zero (Fig. 2a), indicating very high confidence that the multimodel mean loss anomaly due to AA forcing is significantly different from zero. A more sophisticated test will not materially change this result.
 - The t-test was described as stationary by the reviewer, which implies it has no time dependence. However, it definitely contains time information, since it tests a time period of maximum AA forcing against a null hypothesis with no AA forcing. The t-test was designed to deliver robust results by considering the size of the time period of AA forcing: a compromise is needed between using larger numbers of years for robustness, versus inclusion of years with weaker AA forcing hence smaller signals and less robust results. The original manuscript used the 40-year period 1970-2009, and this is lengthened to 1960-2009 in the revised manuscript, to include strong AA-forcing in the 1960s (e.g. Lund et al., 2019; Needham and Randall, 2023).
 - The reviewer suggests a more sophisticated test involving time evolution of AA forcing. Such a test includes a model of the temporal character of anomalies as the basis of testing. Given the complex time evolution of signals in the climate system (noted by the reviewer), it is not clear whether a more sophisticated test would be more accurate than a t-test which implicitly includes time information via the use of the 1960-2009 period corresponding to maximum AA forcing.
 - The manuscript has been revised in response to the reviewer's questions. The revisions include a general discussion of the considerations used in the design of the t-test, to ensure its results were as robust as possible, as well as a specific change to the sample of years (1970-2009 increased to 1960-2009). These changes were included in the new subsection 2.3, and have improved the quality of the manuscript. The author is grateful to the reviewer for drawing attention to statistical testing.

2) The way to conduct the t-test is unclear. The data description states that the time series of wind-related losses has been smoothed by a -20year Butterworth filter 'to highlight the decadal variations in the results' (line 125). Does this mean that the t-test has been applied to the smoothed data? Also, the ensuing paragraph seems to indicate that the standard deviations of the series, needed to estimate the 'separation' of the mean values, have been calculated using the smoothed values. If this is the case, then the calculation is not correct, because the number of degrees of freedom (the $\sqrt{N}$ by which the standard deviation is divided) is not the number of time steps, but much less. Otherwise, one could artificially inflate the precision of the mean estimate, but smoothing the original data will. This needs to be clearly explained.

- Thank you for highlighting how the description of statistical testing is not sufficiently detailed to prevent concerns about quality of analysis methods. The manuscript has been revised to provide a

more detailed description of the statistical tests (standard errors, and t-tests) in the new subsection 2.3. Specifically, the second paragraph of subsection 2.3 explicitly states the use of unfiltered annual means in standard error calculations: ‘...using the entire timeseries of unfiltered annual means from its corresponding Control integration ...’. The final paragraph of the new subsection 2.3 describes the t-test, and also explicitly states the use of unfiltered data: ‘... unfiltered ensemble-mean loss anomalies for each year in the 1960-2009 period...’.

Particular points

3) The explanation of what constitutes internal variability is not correct and should be thoroughly revised. For instance, the text states that ‘Climate models consistently produce slower climate variations in the North Atlantic sector in the absence of all external forcings, which is referred to as internal variability. This sentence is misleading. First, internal climate variations are produced by models forced by constant forcing (not the absence of forcing); secondly, internal variability occurs at all time scales (from second-scale turbulence to millennial-scale ocean currents).

‘This multidecadal driver can be broken down into two different components.’.

Internal variability is not a ‘driver’, and this sentence may lead to confounding internal variability with external climate drivers.

‘The first concerns the Atlantic Meridional Overturning Circulation (AMOC)’.

The AMOC is a term that describes a three-dimensional climate system. It is not ‘internal’ or ‘external’ variability per se. Actually, the manuscript later describes the impact of an external forcing (AA) on the AMOC. This paragraph is really not well-structured and sounds superficial.

‘The second component is atmosphere-based, and consists of shorter-timescale extremes of sufficient magnitude to alter multidecadal averages of storminess’.

Again, the atmospheric variability may be internal or externally forced, at all time scales, not only the slow variations. Although high-frequency internal variability is also reflected in low-frequency variability (through sampling variability), some atmospheric processes generate intrinsic low-frequency internal variability.

Next, the manuscript cites EL Niño as an example of atmospheric internal variability, whereas it is very well known that ENSO arises through the coupling of the Tropical Ocean with the Tropical atmosphere.

There are many other aspects in this discretion that need revision. I really hate to be harsh, but this part of the introduction, in my opinion, needs a complete revision.

In addition, this part of the introduction is really not necessary for the manuscript's goal. I interpret the author as suggesting that AA may directly affect the atmospheric circulation, and that the AMOC, in turn, modifies the meridional heat transport and thus the atmospheric circulation. Whereas the direct path is very rapid, the AMOC path may display a delay of several years, due to the sluggish response of the deeper ocean. But this is only my interpretation of what the author wish to say after reading the whole manuscript. The issue of internal or externally forced variability is, in my opinion, just steering the flow of the text from the relevant messages, even if it were properly phrased.

- The author is grateful for this feedback. Their final point is the most material, on whether the manuscript should contain such a level of detail on processes that are not central to the study.
- Given that the study is focussed on AA forcing, the Introduction of the manuscript has been modified to contain a condensed description of internal variability, and other sources of multidecadal storm variations.

4) *'Observed multidecadal changes in European storm losses align with the AA-forced signals'*

Is there a reference for this? I really doubt that this sentence, as written, can be correct, since wind-related losses would be primarily affected by GDP and/or population growth. Does the author mean 'normalised' losses? Beyond socio-economic factors, is there really evidence that AA is the main physical factor affecting wind-related losses, given the large internal variability of the atmosphere? If the sentence were correct, it would imply that atmospheric variations are almost entirely driven by aerosol forcing, which is not true. This sentence really requires a very solid backup and very careful phrasing (what does 'align' mean exactly?).

- The reviewer states that the sentence *'Observed multidecadal changes in European storm losses align with the AA-forced signals'* implies that *'atmospheric variations are almost entirely driven by aerosol forcing'*. This link is not the intended message, and the author did not anticipate such a link being made in the context of an Introduction with descriptions of various forcings of storm changes.
- To prevent such a misunderstanding, the statement *'Observed multidecadal changes in European storm losses align with the AA-forced signals'* has been replaced by the following sentence in the revised manuscript: *'Further, a reconstruction of windstorm losses (with socio-economic effects removed) in Cusack (2023) contains anomalies that co-vary with storm track changes over the past few decades.'*

5) *In brief, sulphates have produced the largest radiative forcing in the industrial period, growing from a relatively small amount at the start of the 20th century to peaks in the 1980s and 1990s'*

*I guess the author means the largest *changes* in short-wave anthropogenic forcings. The largest radiative forcing is by far the sun. This is an example of inaccurate writing that can be found elsewhere in the manuscript.*

- Thanks. This inaccuracy is corrected in the revised manuscript: *'... sulphates have caused the largest multidecadal anomalies of shortwave radiative forcing in the industrial period...'*, and the author conducted a wider search to find and correct similar mistakes.

6) *The contribution of socioeconomic factors to the wind-related losses is taken into account by multiplying the physical (wind) factors by the population in a particular grid-cell (equation on page 4, equations are not numbered!) . It seems the population is considered constant over time (?). This would be quite unrealistic, as population and GDP growth over time would, I think, increase the exposure to wind extremes and thus dominate any trends in wind-driven losses. It would be rather easy to include population or GDP growth. Is there a reason not to do so? This point should at least be discussed.*

- A new paragraph was inserted in the revised version of subsection 2.2 to explain why socio-economic causes of loss trends are removed from storm loss estimates: *'While populations evolve over time, we follow the common practice of using population data fixed to a common year when applying Eq. (2). This ensures all modelled loss anomalies are caused by changes in storm climate, in isolation from socio-economic factors that drive changes in nominal losses.'*
- Equations are numbered in the revised manuscript.

Referee 2

Summary

This study investigates whether variations in anthropogenic aerosol (AA) forcing may explain the observed multidecadal variations in property losses caused by European winter storms during the past decades. It is motivated by the result of a previous published study of the author, showing that European windstorm losses exhibited a notable positive anomaly during the 1980s and 1990s. The present study uses six CMIP6 climate models (with several ensemble members per model, in total 62 simulations) to isolate the effect of AA forcing on European windstorm losses, by comparing control simulations (with all external forcings set to preindustrial levels) to simulations with historical (1850-2014) AA forcing. A simple but well-established diagnostic for windstorm losses is applied to the wind data generated by the models, and differences between multidecadal averages of the losses ('AA forced' minus 'control') are investigated. The manuscript concludes, that AA forcing could have been a major driver of the observed positive anomaly of European windstorm losses during the 1980s and 1990s. To my understanding, this is the main conclusion of the manuscript.

Although the motivation and overall approach for this study are meaningful and interesting, the manuscript does, from my perspective, not present sufficient evidence for the main conclusion. Moreover, the Introduction section needs to be improved regarding terminology and structure, and also the description of the methodology requires some clarification. Details are given by the following comments.

- The author thanks Referee 2 for their comments. They have improved the revised manuscript.
- The feedback suggested that the scope of this manuscript needed to be more clearly defined. The main aim of this study is to measure the AA forcing of European windstorm losses using climate models. In contrast, Referee 2 writes that the present study is focussed on the extent to which AA forcing explains the total forcing of storminess, and that the manuscript requires analysis of the modelling of total forcing to answer their question. It is hoped that the revised manuscript and responses below clarify the scope of this work is confined to AA forcing of European windstorm losses.
- The text in the Introduction and the section on Data and Methods has been rewritten and reorganised, to improve the clarity of the revised manuscript.

Major comments

(1) The main question to be answered by the present study is: Does AA forcing explain the observed positive anomaly in windstorm losses during the 1980s and 1990s? (Or at least: to which extent does it explain the anomaly?) However, this question can only be answered with a climate model that does actually simulate the observed anomaly. In particular, since the observed anomaly occurred under the full historical forcing (not just AA forcing), the above question can only be answered with a climate model that simulates the anomaly under full historical forcing (the so-called CMIP6 'historical' experiment). Then we can ask: Is this simulated anomaly driven by AA forcing or by another external forcing or may it have occurred by internal variability?

Unfortunately, the manuscript does not provide any information regarding this point. In my opinion, the most straight-forward, and probably easiest, way to address this issue, would be to repeat the analysis for the corresponding model simulations with full historical forcing. Then the obtained time series of the (low-pass filtered) windstorm losses could simply be added to those obtained with AA forcing only, already shown in Figure 2.

- The reviewer states that the study should measure the extent to which AA forcing explains the full observed anomaly in the 1980s and 1990s, then recommends analysing model simulations with full historical forcings to answer their question.
- In contrast, this study measures the impacts of AA forcing on storm losses, rather than all forcings. The amplitude of the storm peak in the 1980s and 1990s is mentioned briefly, to illustrate the materiality of the subject of multidecadal storm variations, but a full explanation of such a peak is outside the scope of this study.
- The fourth and fifth paragraphs of the revised Introduction provide reasons why a study of the full forcing of multidecadal storminess is of limited value at this time, and why a scope limited to AA forcing of European storm losses is useful to insurance.
- The author has read many valuable published studies which focus on one specific forcing of climate, rather than provide a full explanation of the climate system variability, and hopes this study of AA forcing provides enough value, and new information, to constitute an article.

I also suggest to compute the corresponding time series from the ERA5 reanalysis, as there the observed positive anomaly in windstorm losses should definitely be visible.

One could perhaps create a figure with six panels, one for each model. Each panel could show the time series of: the individual ensemble members and of the ensemble mean under AA forcing only, the same (in another color) under full historical forcing, and the ERA5 time series.

This may help to judge the realism of the individual models in simulating the anomaly, to be explained by this study – in terms of both: magnitude (expressed as relative change) and timing.

- The revised Figure 2a contains a reconstruction of observed losses since 1950 from Cusack (2023). It is based on winds from ERA5, with an adjustment based on long-term records of gusts at weather stations to improve the validity of loss reconstructions. ERA5 data are also used in Figure 3.
- While full-forcing historical simulations are outside the scope of this manuscript (discussed above), the analysis of internal variability using individual ensemble members would be valuable, and a suitable avenue for future research into multiple forcings of storm losses.

It may turn out that some of the six models do not simulate the observed anomaly at all, in which case those models cannot be used to explain the occurrence of the anomaly, simply because it does not exist in those “model worlds”.

On the other hand, if all six models do actually simulate the observed anomaly to a sufficient degree of accuracy under full historical forcing, and if the AA forcing only simulations reproduce that anomaly (to some extent), then this would be an explicit indication for AA forcing being responsible for the observed anomaly (to that extent).

- While an analysis of full historical forcings is beyond the scope of this manuscript, the author wishes to add that the above approach to assessing the validity of AA forcing has limitations.
- Complex models often have compensating errors between component parts, hence the validity of anomalies from a full model can be a misleading guide to the validity of each component part. To be more specific, a valid full-forcing simulation can have large errors from individual forcings. (In my model development experience, this is the rule rather than the exception.)
- This study uses a different approach to assessing the validity of simulated AA forcing of storminess which is described in the revised Section 4. In brief, it uses a framework suggested by previous research, and evaluates the quality of modelled responses to AA forcing using observed data on

radiative forcing and poleward energy transports. The ability to study AA forcing in depth is one of the benefits of limiting the scope of this study to AA forcing.

(2) The terminology and structure of the Introduction section are rather confusing.

First, the four possible drivers (lines 23-24) are ordered in a counter-intuitive way. The order is: volcanic eruptions, internal variability, solar variations, AA forcing, so the first is an external forcing, then follows internal variability, and then two external forcings again. I think, it makes more sense to first list the possible external forcings and then mention internal variability. The subsequent paragraphs repeat this counter-intuitive order, so they should be reordered accordingly.

- The second paragraph of the revised Introduction contains a description of volcanic and solar forcing, then internal variability. This is followed by a discussion of AA forcing in the third paragraph. This order has been chosen both to isolate AA forcing, and to provide more information on this forcing in a full paragraph, because AA forcing is the focus of this study.

Moreover, the paragraph from line 34 to line 47 obviously is about internal variability, but line 45 mentions the solar cycle which is represents an external forcing.

- This is fixed in the revised Introduction, which contains a compressed description of internal variability, with no reference to the solar cycle.

In addition, it is not entirely clear to me how the concept of internal variability is used in this manuscript.

- Internal variability is described briefly in the revised Introduction, as a source of multidecadal storm variations, and mentioned again in Section 3 when discussing differences between AA-forced signals and those observed. Otherwise, the study is focused on AA forcing, and does not analyse internal variability. The revised manuscript presents results and uncertainties related to ensemble means, and the contribution of internal variability is relatively small (different phases of modes of internal variability in each ensemble member largely cancel when forming ensemble means).

Second, internal variability is called a 'driver' in the manuscript. This may make sense if properly defined, but as such a definition of the term 'driver' is not provided, it may be misunderstood as something similar to a forcing. In general, in my opinion, the terminology is sometimes imprecise throughout the manuscript and I really suggest to sharpen the terminology and wording.

- Referee 1 also raised this issue, and changes were made in the revised manuscript. Specifically, storm variations are described as emerging as a result of internal variability, rather than driven by internal variability.

Third, I am not sure whether all those rather lengthy explanations in the Introduction section are really needed. If condensed to what is really needed for this manuscript, some of my above caveats may already disappear.

- The revised Introduction has a condensed description of volcanic and solar forcings of storminess, and internal variability. This makes the Introduction consistent with the study's focus on AA forcing.

Specific comments

(3) Line 105: Is the 98th percentile computed from October to March only or from all months of the year? And is it computed from all years of the control simulation (i.e., from several hundreds of years for each simulation) or from a shorter reference period? And is it computed from each ensemble member separately or from all members together (for a specific model)?

- The 98th percentile of the wind is computed from all October to April windstorm seasons in the model's long Control integration (the Control for each model consists of a single run of hundreds of years). This is clarified in the second paragraph of subsection 2.2 in the revised manuscript : 'The first stage of processing defines the 98th percentile of the daily maximum wind in each grid cell ($v_{i,98}$). This quantity is defined uniquely for each model, based on all windstorm seasons in its Control simulation (see Table 1).'

(4) Line 109: What does "of up to three days" mean? This suggests the series may in some cases, that is, for some storm events be shorter than three days. Is that true and, if yes, in which cases does that happen?

- The manuscript was revised to clarify this definition: 'an event is defined as a period of up to three days centred on the days with peak values of D_d , though may be of shorter duration in practice when the maximum winds are below the 98th percentile throughout the domain'.

(5) Line 127: "...is computed using the full time series..." What does "full" time series mean? Does it mean "unfiltered" or does it mean "full length" (i.e., several hundreds of years) or does it mean both?

- The manuscript was revised to clarify this detail: '... using the entire timeseries of unfiltered annual means from each model's Control integration...'

(6) I am surprised that the t-test is explained in the Results section, rather than the Data and Methods section. Also, because the standard error is computed from the control simulations, which have constant external forcings, this error precisely represents/quantifies internal variability. This could be mentioned together with the description of the t-test.

- Practical details on the t-test and standard errors have been moved to a new subsection 2.3 in the revised manuscript. This reorganisation improves the clarity of the manuscript.
- The standard errors shown in the manuscript (Fig. 2a) are a measure of the sampling variability of the ensemble mean of 62 simulations. They do not represent the standard error of a single simulation, such as the Control. The text in subsection 2.3 of the revised manuscript explicitly states how the standard errors refer to the multimodel mean of AA-forced simulations, to avoid such a misunderstanding.
- Standard errors for single model simulations are readily available, but never discussed because the analysis concerns climate model ensembles of AA forcing.

(7) Line 217: "These tabulated values contain a pattern..." As I understand this, the "pattern" refers to the relation between the strength of the AMOC changes and the magnitude of European windstorm loss changes. And this relation is later used to construct some arguments. However: Is there any indication that this relation is either statistically significant or at least physically plausible (i.e., it does not appear just by chance)?

- Results are available for a sample of five models, which is too small for reliable statistical testing.
- However, previous research has revealed the physical framework connecting anomalies in AMOC and storm track intensity, whereby larger anomalies in one imply smaller changes in the other. In essence, Needham and Randall (2023) described how AA perturbed the latitudinal gradient of outgoing radiation at the top of atmosphere (TOA), and how the total poleward energy transport (PET) responds to this change. In turn, Needham et al. (2024) showed how the total PET anomalies are manifested as changes in two quantities in northern mid-latitudes, namely oceanic heat transport (largely due to AMOC variations) and atmosphere eddies (storm tracks). Therefore, the validity of the modelled change in storm tracks due to AA forcing can be inferred from assessments of the accuracy of historical anomalies in both simulated TOA radiation and AMOC.

- The author thanks the reviewer for drawing attention to the unclear description of this framework in the original version of the manuscript. The text in Section 4 has been substantially re-written to provide a clearer description of the reasoning, which leads to the conclusion that the reported bias in many CMIP6 models concerning an over-responsive AMOC to AA-forcing implies these models may underestimate storm track intensity anomalies in the late 20th century.

(8) Line 227: "...the magnitude of modelled AA forcing is smaller than best estimates of observed..."; versus line 231-232: "...it is possible that the models studied here have too strong AA forcing..." – which sounds contradictory to me.

- The sentence in the text was revised for clarity: ‘... although modelled magnitudes of AA forcing (around -1.1 Wm^{-2}) are smaller than the best estimate, Forster et al. (2021) indicate it is very likely that global ERF due to AA forcing in 2014 lies in a range from -0.6 to -2.0 Wm^{-2} , therefore, there is some chance that the AA forcings in these models are stronger than the true value.’