We sincerely thank the Reviewer for the thorough and constructive assessment of our manuscript and for recognizing the importance of developing attribution frameworks for aviation turbulence. Below we provide a detailed point-by-point reply to the three main comments. Reviewer's comments are in normal font, while our replies are in bold font.

General comments

This manuscript introduced an analogue-based attribution framework, TurboMeter, which had been developed to assess relative contributions of anthropogenic climate change and natural variability on aviation turbulence in changing climate. The key procedures to the framework are, first, select turbulence-prone days in an independent time frame (a one year period of 2024 in this study) using automated in-situ turbulence observations from commercial aircraft; second, characterize the synoptic-scale patterns of the selected turbulence days using the 500-hPa geopotential height in ERA5 as a single metric; third, for each of the selected turbulence days, identify analogues in both past (1950–1986) and present (1987–2023) climates that have similar 500-hPa geopotential height patterns with the selected turbulence event; and then last, compute and compare a single turbulence index (TI1) and its components (vertical wind shear (VWS) and deformation (DEF)) to assess if there are any changes between the past- and present-climate analogues despite the similar 500-hPa geopotential height patterns.

I agree to the necessity of assessing the attribution of anthropogenic climate change to aviation turbulence, separately from the attribution of natural variability. However, I have several major concerns about the algorithms on which the TurboMeter framework is built on, i.e., the characterization of the synoptic-scale patterns and the case selection, in particular. In brief, I doubt if using the 500-hPa geopotential height alone is sufficient to characterize synoptic-scale patterns and identify analogues, especially characterizing the synoptic conditions that are critical to generation, maintenance, and intensification of aviation turbulence; in addition, the case selection was made based on the EDR threshold only, without separating sources of turbulence (e.g., clear-air turbulence, mountain-wave turbulence, convectively induced turbulence), which seems to make it even harder to justify using the 500-hPa geopotential height alone as a single metric to identify synoptic-scale patterns responsible for the selected turbulent days.

Besides my concerns on the TurboMeter framework, the analysis of the TurboMeter output needs to be improved to provide more in-depth explanation of physical mechanisms supporting the output. The result plots in the manuscript focused on 500-hPa geopotential height, TI1, VWS and DEF. However, the main texts mentioned and relied heavily on changes in and patterns of the jet stream in explaining the TurboMeter output, while the jet stream and other synoptic-scale patterns were never shown or compared in the manuscript. I think the analysis needs to be improved by explicitly showing the differences and similarities in jets, troughs, and other possible sources of turbulence between the past and current climate analogues.

While I agree to the motivation, objective, and necessity of this study, I have major concerns about the methodology that would need significant revisions and could lead to significant changes in the results, as well as lack of in-depth analysis on physical mechanisms explaining the TurboMeter output. Therefore, my recommendation is to reject this manuscript to allow the authors sufficient time to work on their revisions but strongly encourage to re-submit the manuscript.

We appreciate the concerns raised regarding both the analogue-selection methodology and the physical interpretation of the results. We agree that the manuscript would benefit from a more explicit validation of the analogue framework, a stronger presentation of the upper-level dynamical environment associated with the selected analogues, and a clearer discussion of the scope and limitations of TurboMeter. In a revised version of the manuscript, we plan to address these points by (i) providing additional evidence of the consistency of the selected analogues, (ii) including explicit diagnostics of upper-level wind and jet-stream structure, and (iii) clarifying that TurboMeter is intended as a circulation-constrained attribution framework rather than a process-specific turbulence classification system.

Major comments

1. This study used a single metric of 500-hPa geopotential height to characterize synoptic-scale patterns of turbulence-prone days in 2024 and then to identify analogues (i.e. days with similar synoptic-scale patterns) in past (1950–1986) and present (1987–2023) climates. Then, based on the comparison between the past and present climate analogues, the contribution of anthropogenic climate change was determined; that is, if the characteristics of TI1 are found differently between the past- and present-climate analogues despite similar synoptic-scale patterns, their differences can be attributed to anthropogenic climate change.

For this analogue-based attribution framework works, it should be validated that the 500-hPa geopotential height is a good metric to characterize synoptic-scale patterns that generate, maintain, and intensity turbulence in the selected turbulent days. However, this study does not provide that critical validation of the metric (500-hPa geopotential height). This validation can be done by comparing the weather patterns between the analogues for each case. Do the past and climate analogues selected based on the 500-hPa geopotential height have similar synoptic weather patterns (i.e., jet, trough, baroclinicity, convection, etc.)? Or do they just happen to have similar 500-hPa geopotential height patterns but show very different weather conditions among the analogues? Without such a validation of the use of 500-hPa geopotential height as a metric to characterize synoptic-scale patterns, it is hard to justify the outcomes of the attribution method based on the metric.

We thank the Reviewer for this comment. The rationale behind the use of Z500 is that it provides a robust and widely used representation of the large-scale circulation state, allowing us to identify atmospheric configurations that share similar synoptic-scale characteristics (Faranda et al., 2024; Faranda and Alberti, 2026). We would like to clarify that the objective of TurboMeter is not to reproduce the complete atmospheric state associated with each event, nor to identify analogues that are identical in every dynamical (and thermodynamical) aspect. Rather, the framework aims, as usual in attribution studies (e.g., Faranda et al., 2024) to constrain the large-scale circulation regime and subsequently assess whether turbulence-relevant diagnostics differ between past and present climates under comparable circulation patterns. This is also the main reason to consider the full state-space (i.e., the complete Z500 field) instead of its projections onto fixed orthogonal basis which project the system dynamics over a low-dimensional space. While this is certainly useful to identify the weather regimes beyond physical processes, it can affect the results of the identified analogues, being constrained over similar weather regimes rather than similar large-scale synoptic conditions, including more processes acting together (Faranda et al., 2017).

Nevertheless, we acknowledge that the current version of the manuscript does not sufficiently demonstrate whether the selected analogues exhibit consistency. In a revised version of the manuscript, we therefore plan to include further diagnostics, as in ClimaMeter, to evaluate the quality of the selected analogues, as well as, we will include other fields (e.g., upper-level wind speed, total column water vapour, among others) to detect the further characterize jet stream variability and the possible presence of different mechanisms relevant to turbulence generation. Furthermore, we will at least show the Z500 maps for the detected analogues and we are also planning to investigate, a posteriori, the similarities between the weather regimes among the analogues. We believe that these additional analyses will help clarify the physical meaning of the analogue framework and provide a more comprehensive assessment of its robustness.

2. In this study, turbulence-prone days were selected based on automated in-flight turbulence reports using an EDR threshold (EDR ≥ 0.2), without distinguishing the primary mechanisms and/or sources of turbulence. Turbulence can be generated by diverse sources, which include but are not limited to clear-air turbulence in the vicinity of jet, trough, tropopause etc. in the absence of clouds; mountain-wave turbulence due to breaking waves; convectively induced turbulence associated with moist instabilities, intense updrafts/downdrafts, outflows etc. This can be especially problematic as relative contributions of turbulence sources may differ among the four flight corridors (i.e., North Atlantic, North Pacific, East Asia, and CONUS), therefore may impact the differences in the TurboMeter outcomes among them. I think it is needed to validate that the selected turbulence days in 2024 and their past and present analogues are associated with the same source of turbulence; for example, if a selected case

in 2024 is clear air turbulence, their analogues should be clear-air turbulence days as well. Also, the turbulence index TI1 and its components seem more appropriate for clear-air turbulence than for mountain-wave turbulence or convectively induced turbulence.

Such complexity in the possible source of turbulence is another reason why I don't think the 500-hPa geopotential height alone is a good metric to identify similar synoptic patterns and analogues. Do the analogies have similar synoptic patterns in terms of jets and troughs, background atmospheric conditions for wave breaking, and initiation and development of convection, which all are synoptic conditions for turbulence generation?

We thank the Reviewer for highlighting the important issue of turbulence-generation mechanisms and their potential influence on the interpretation of the results. We clearly agree that aviation turbulence can arise from a variety of physical processes, including clear-air turbulence, mountain-wave turbulence, convectively induced turbulence, and so on, that may also vary among flight corridors. As mentioned before, the primary objective of TurboMeter, however, is to detect changes in turbulence-related metrics under similar large-scale synoptic conditions and not to attribute changes in individual turbulence-generation mechanisms. For this reason, turbulence-prone days were identified using EDR observations, which provide an objective and operationally relevant measure of turbulence intensity regardless of the underlying generation mechanism. We acknowledge that the manuscript did not sufficiently clarify this aspect, suggesting that the framework seeks to isolate specific turbulence processes, whereas its intended purpose is to assess changes in the broader atmospheric environments associated with observed turbulence occurrence.

In a revised manuscript, we will expand the discussion regarding the presence of other possible turbulence sources, clarify that TurboMeter is not a turbulence-classification framework, and discuss how the coexistence of different mechanisms should be considered when interpreting the results, also including additional fields (see also our reply to the previous comment).

3. Overall, I think the result section needs revisions in terms of plots and texts both.

The plots presented focused on the TurboMeter outputs, including 500-hPa geopotential height, TI1, VWS, and DEF. However, neither the main text nor the figure caption mentioned what vertical levels the TI1, VWS, and DEF plots are for. If those turbulence parameters are presented for a 500-hPa level, I think it needs to be updated to an upper level, for example 300 hPa, which is in typical cursing altitude range and is of great interest in aviation turbulence.

For the main texts, the differences in turbulence between the past and present climates, as well as their difference among the four flight corridors, were mainly explained in terms of the jet dynamics. However, this manuscript never showed how the actual synoptic-scale patterns look like including jets, temperature patterns, etc., to support their arguments. I think the analysis needs to be improved by explicitly showing the differences and similarities in jets, troughs, and other possible sources of turbulence between the past and current climate analogues.

We appreciate the Reviewer's comments regarding the physical interpretation of the results and agree that the manuscript would benefit from a more explicit connection between the analogue-based diagnostics and the underlying atmospheric dynamics.

First of all we need to clarify that the diagnostics TI1, VWS, and DEF are obtained for the cruising altitude range, i.e., within 300 hPa and 200 hPa, and not for the 500 hPa level. We would make this more clear. While these diagnostics are physically linked to upper-level circulation changes, the corresponding jet-stream structure and upper-level wind fields are not explicitly shown in the manuscript. As correctly noted by the reviewer, this makes it more difficult for the reader to directly assess the dynamical mechanisms discussed throughout the text.

In a revised version, we therefore plan to complement the existing analyses with additional diagnostics of upper-level wind speed and jet-stream characteristics. These fields will be examined both for representative analogue cases and for the composite analyses used throughout the manuscript. This will allow a more direct comparison of the large-scale circulation environments associated with the past and present analogues and provide a clearer physical interpretation of the observed changes in TI1, VWS, and DEF. For the sake of clarity, we just show below two examples of additional metrics, including jet stream speed and total column water vapor for two cases shown in the manuscript, one over East Asia and one over North Atlantic.

The overall consistency of the analogue-selection procedure can be clearly observed. In both corridors, the past and present analogues reproduce the main large-scale circulation features of the target event, as indicated by the similarity of the Z500 patterns and the associated upper-level jet structures. This suggests that the analogue framework is successfully constraining the large-scale dynamical environment while allowing differences in turbulence-relevant atmospheric properties to emerge between climate periods.

For the East Asia corridor, the present-minus-past differences reveal a strengthening of the upper-level jet particularly along its core and equatorward flank, together with a modest increase in total column water vapour. While the overall circulation pattern remains comparable between the two analogue sets, the present-climate analogues are characterized by stronger upper-level winds and a slightly moister background

atmosphere. These changes are consistent with enhanced dynamical conditions favourable to increased turbulence occurrence.

For the North Atlantic corridor, the similarities between the event and the analogue composites are particularly evident in both the Z500 pattern and the position of the jet stream. The present-climate analogues exhibit a strengthened jet over large portions of the basin, accompanied by a general increase in atmospheric moisture. Although the large-scale circulation regime remains broadly unchanged, the present-climate environment appears dynamically more energetic and thermodynamically moister than its historical counterpart.

Overall, these results support the interpretation that, under comparable large-scale circulation regimes, present-climate analogues tend to be associated with stronger upper-level winds and enhanced moisture availability that may contribute to the higher values of turbulence-relevant diagnostics observed in our analysis.

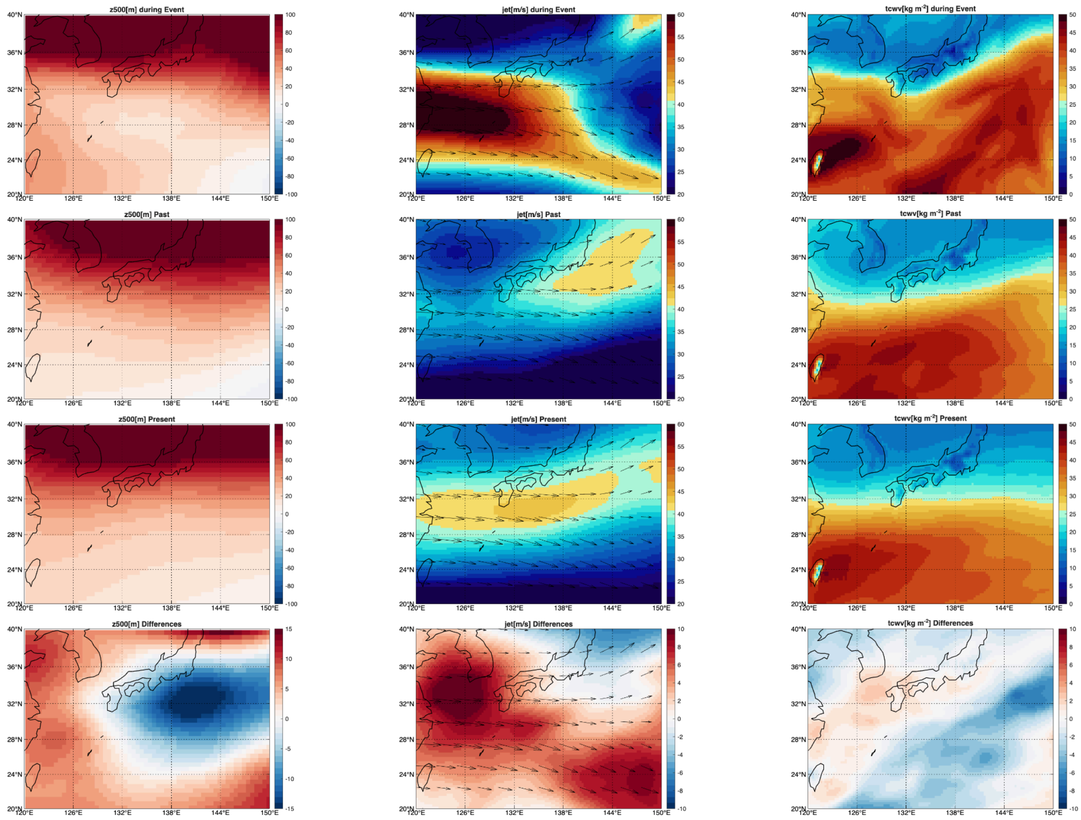

Figure R1. Composite fields for the East Asia corridor (20°–40°N, 120°–150°E). Columns show (from left to right) 500-hPa geopotential height (Z500; m), 250-hPa wind speed (m s^{-1}; shaded) and horizontal wind vectors, and total column water vapour (TCWV; kg m^{-2}). Rows show (from top to bottom) the 2024 turbulence event, past-climate analogues (1950–1986), present-climate analogues (1987–2023), and present-minus-past differences.

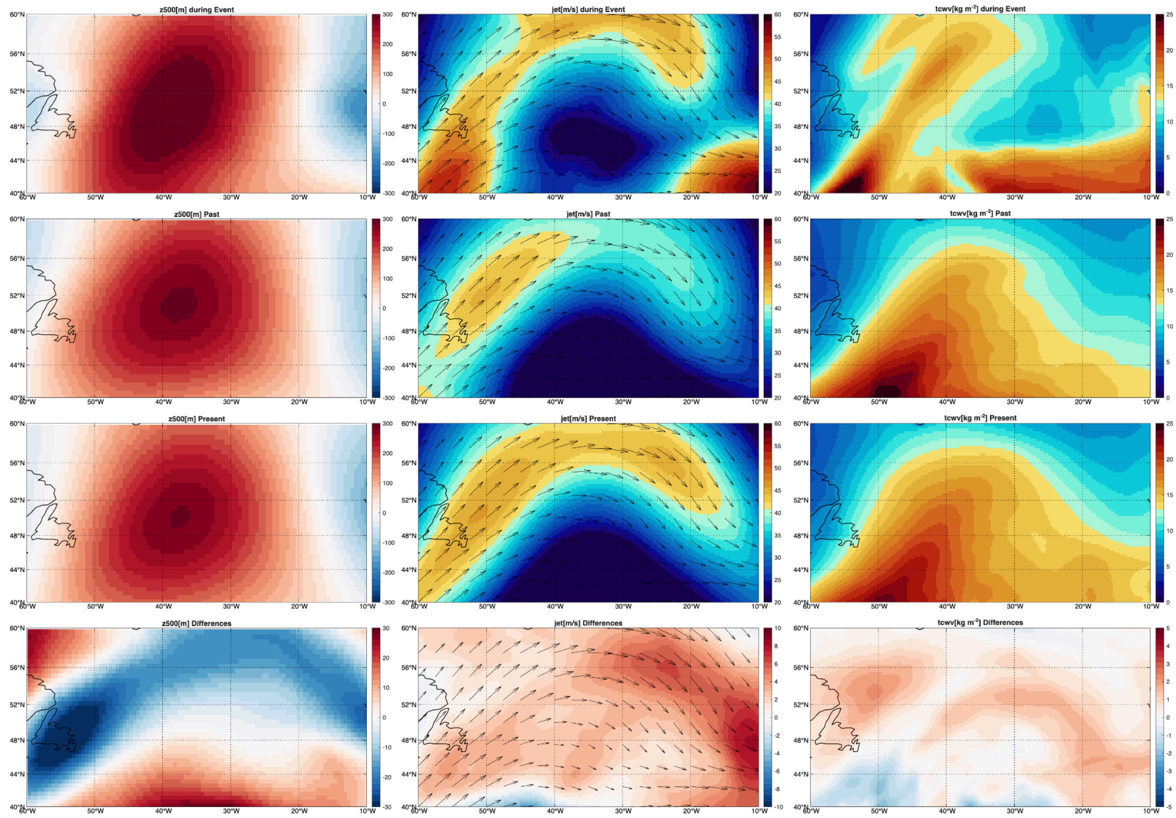

Figure R2. Composite fields for the North Atlantic corridor (40°–60°N, 60°W–10°W) as in Figure R1.