We thank both referees for their insightful comments and remarkable editorial work. In the following document, we hope to respond to the main comments of both referees as well as all the minor points. The manuscript has also been thoroughly revised following the many comments about its clarity. Some of these revisions are still ongoing as we publish this response text, therefore the line numbers that we cite in this document might not exactly match the final manuscript.
In addition to the changes prompted by the referee's comments, the manuscript has also received changes due to an update in the methods, altering the results quantitatively thanks to more precise and reproducible diagnostics, but keeping all the conclusions unchanged. These changes are detailed at the end of this document and might also be mentioned in the responses to individual comments.

Referee 1

Summary and major comments

This study presents two diagnostic tools applicable to jets over the North Atlantic: (1) a SOM trained on upper-level flow patterns from ERA5 output across several decades, and (2), a jet identification scheme that locates tracks individual jet axes. Each jet axis is categorized as either an instance of the polar jet/eddy-driven jet (EDJ) or the subtropical jet (STJ). The authors have developed an impressive set of methods to generate these tools and provide many well-motivated avenues for the use of these tools in future research. The authors are also very attentive to recognizing some of the limitations of their methodology throughout the manuscript. Overall, I find this paper to be scientifically significant and I think many of the applied methods and conclusions are very valid. However, I am concerned about the validity of the methods used to categorize each jet feature as either an EDJ or STJ. There are also several moments where the writing in the paper is not as clear as I think it could be (I have pointed out these moments in my line-by-line comments).

I suspect that the approach used to distinguish the EDJ and STJ using jet frequencies in mean latitude–mean longitude–pressure level space may not be as physical and therefore may not be as robust as an approach that examines how the jets are distributed in wind speed–potential temperatures space. For example, the authors show a histogram of jets in mean latitude and mean longitude in panel (a) of Figure 4. Although there are two peaks in the distribution, I wonder if the interpretation that one peak belongs to the EDJ and one peak belongs to the STJ is really appropriate. How do we know that both types of jets do not contribute to each of these peaks (i.e., these peaks are just peaks in jet occurrence in general, not peaks mostly in one type of jet or the other)? Additionally, there appears to be no bimodal distribution in the histogram shown in panel (b) of Figure 4. How is it possible to draw a line separating the EDJ and STJ from each other?

Furthermore, there is often only one jet present during the summer (which is clearly reflected in the double jet index time series from Figure 10). The unimodal maximum in frequency observed along the pressure level dimension (panel (b) in Figure 4) suggests that maybe this single jet is best described as a combination of both the EDJ and STJ. Alternatively, I think one could argue that such a jet could be better represented by the physical processes supporting the STJ as opposed to the EDJ, since the Hadley Cell remains during the summer, but baroclinicity over the mid-latitudes is drastically reduced.

To address these concerns, I suggest examining seasonal histograms of the jet frequencies for bins of potential temperature in the vertical direction to confirm the validity of the current delineation between the EDJs and STJs. I suspect that the distribution may still, however, be unimodal. In that case, I think a sensitivity test is warranted to determine how the choice of a cutoff in potential temperature or pressure level between the categories along the vertical dimension impacts the frequencies of each category in total and geographically. If the categories are very sensitive to this choice, then maybe a third category of jet is required to describe the summer jet, especially when the double jet index is relatively low.
We thank the referee for their comment. Your concern about the performance of the jet categorization is very valid, and more work has been done to explore this potential problem, which has led us to a better categorization method.

Firstly, the jet categorization now assigns a continuous score, between 0 and 1, to each jet point, as opposed to a hard 0 or 1 score to the jet as a whole. A score near 0 means the point is a STJ point, a score near 1 is a EDJ point, and scores near the middle (fairly rare) can be seen as hybrid or miscategorized. The average of these scores across a jet core object is also between 0 and 1, typically close to either extrema. Jets with a score close to the middle can be seen either as hybrid or as misdetected, as can happen on composites for example. Secondly, the new categorization method uses two variables: potential temperature and jet vertical extent / baroclinicity proxy, instead of the spatial longitude - latitude - pressure level 3D space we were using before. The potential temperature is interpolated from pressure level temperature data, and the baroclinicity proxy is defined as the ratio between lower level wind speed magnitude, at 500 hPa, to high level wind speed magnitude, the vertical maximum of U on pressure levels ranging from 350 to 175 hPa, that is the original wind speed magnitude field used in the jet detection. This second variable has been used by Koch et al. (2006) to create their two categories of jet events, although they didn't name them EDJ and STJ.

The method is applied independently to every month, and the results can be seen in Figure 1.

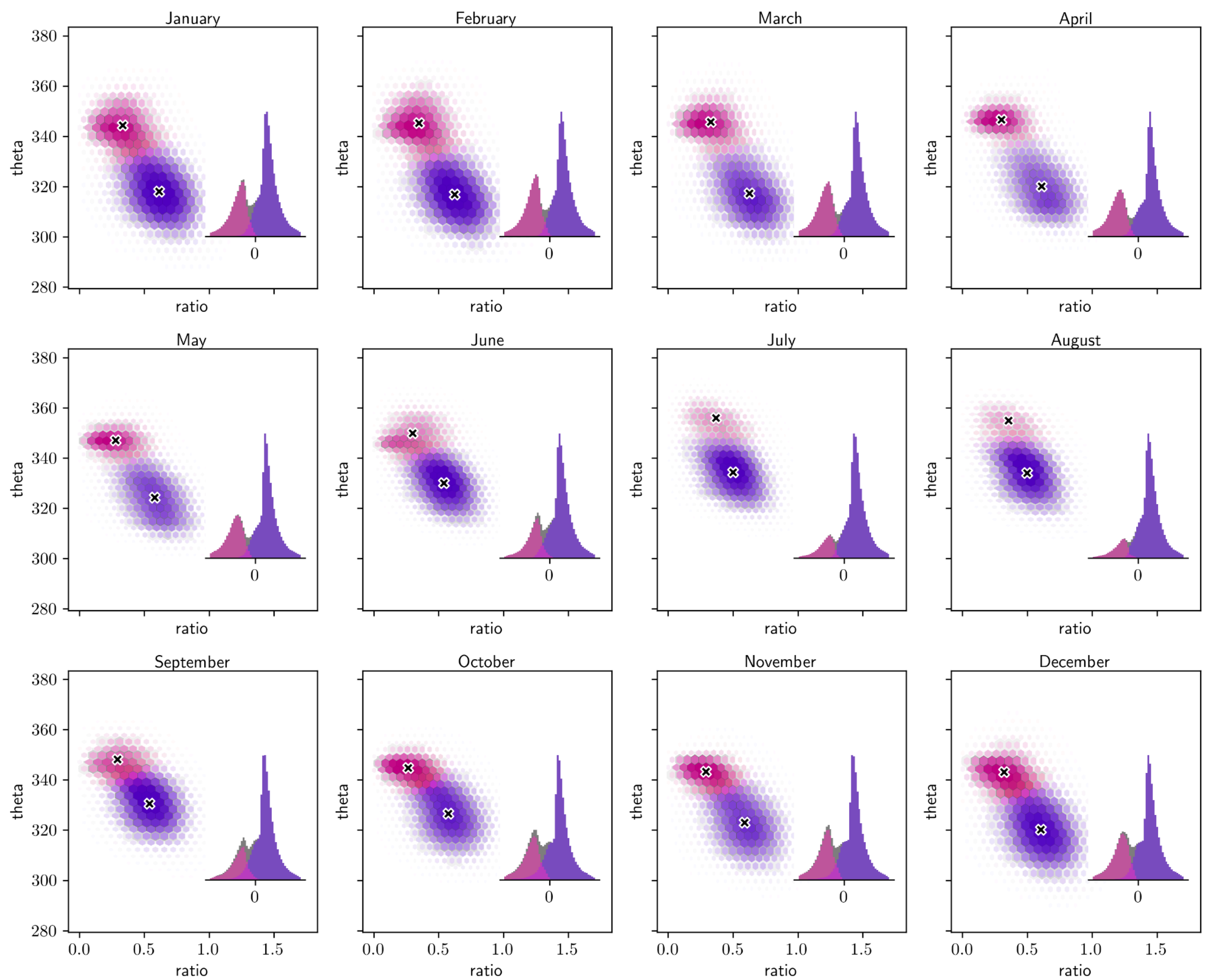

Figure 1: Results of the new jet categorization for each month of the year. On the left of each box, the jet points are binned in the 2D space potential temperature - vertical extent proxy, in order to illustrate the underlying distribution in this 2D space. The size and lightness of the hexagonal bins indicate their height. The crosses indicate the centers of either discovered Gaussian, x_1 *and* x_2*. On the right of each box, the a histogram of the quantity* $\log \frac{||x - x_1||}{||x - x_2||}$*, where* x *is a 2D vector containing a jet point's vertical extend proxy and potential temperature, and* $||y||$ *is the 2-norm of the vector* y*. This quantity is not the final score and just serves illustrative purposes. The colors on both sides of each box indicate which Gaussian component is dominant, from pink for STJ to purple for EDJ.*

To move further in the analysis and produce, for example, figures 10 and 11, each jet needs to be assigned a category, and a cutoff in jet-mean STJ / EDJ cluster belonging needs to be defined. Here, we show that the results are largely insensitive to this choice, by reproducing figure 10 (albeit with a reduced choice of variables) with a cutoff at 0.1, 0.5 and 0.9 for the EDJ. The results can be seen on the first 3 rows of figure 2.

Additionally, we ask what happens if we accept that jets far away from the extrema, for example with EDJ cluster scores between 0.1 and 0.9, are assigned to a third, “hybrid” category instead. The results can be seen on the fourth row of figure 2.

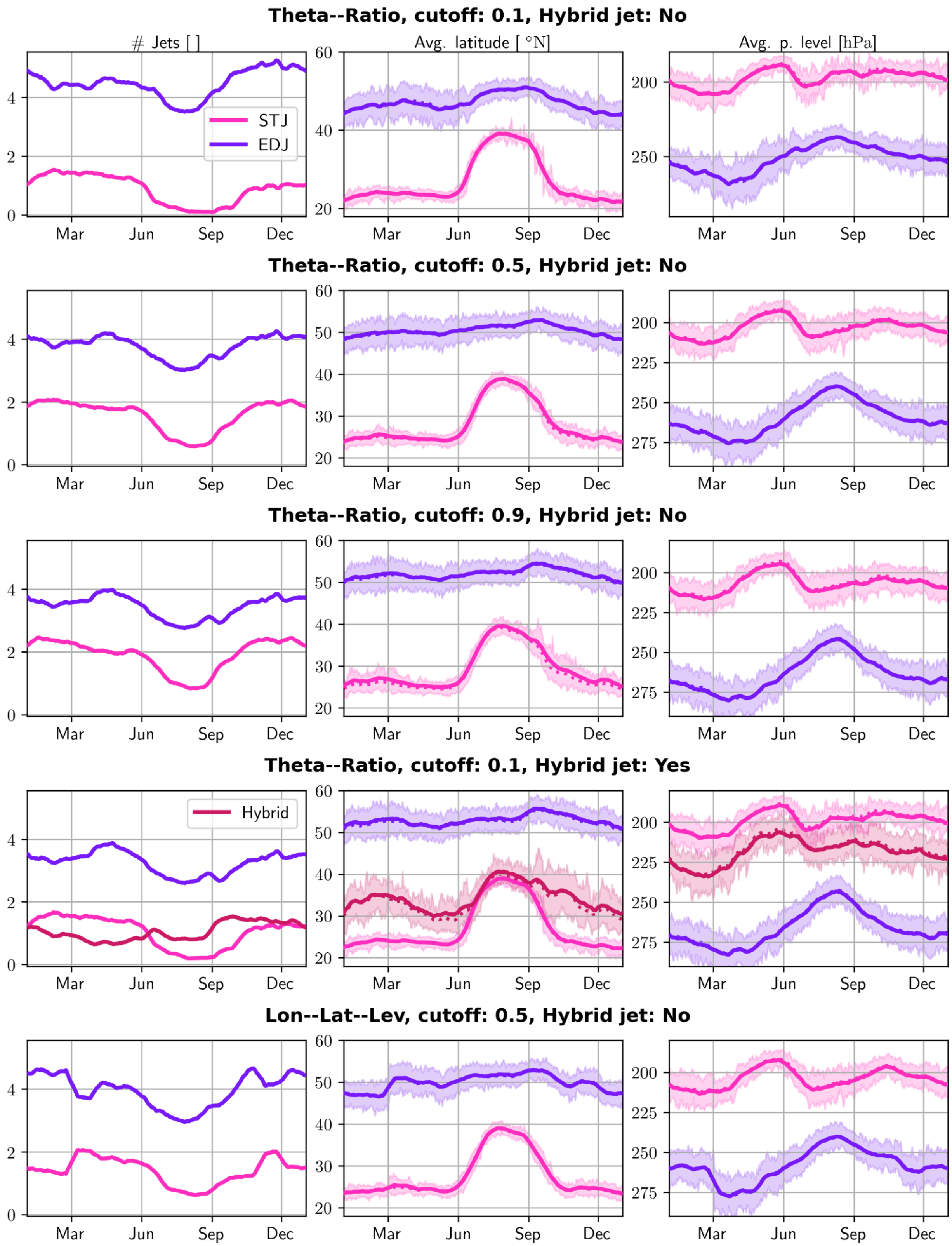

Figure 2: Reproduction of the main text's figure 10 with only three jet properties: mean latitude, mean pressure level and maximum speed. First row: The jets are assigned to the category EDJ

if the jet mean score for the EDJ cluster is above 0.1. Otherwise, the jet is assigned to the category STJ. Second row: same but with a cutoff of 0.5. Third row: same but with a cutoff of 0.9. Fourth row: the jets whose score is between 0.1 and 0.9 are assigned to a third "Hybrid" category, in red. Fifth row: the scores are assigned with respect to clusters defined in the longitude - latitude - pressure level subspace, as was the case in the original figure 10, and the cutoff is 0.5.

The overall conclusion we draw from this study is that very few jets belong to the hybrid category. We therefore decided against introducing this third category in the paper. Furthermore, in its seasonal cycle at least, the hybrid jet behaves very similarly to the STJ.

We would interpret these findings as follows. In summer, the subtropical jet is shifted North with the Hadley cell and interacts with extratropical eddies more, making it lose more momentum (Martius 2014) and potentially making it more baroclinic. This makes the distinction more fuzzy, so there are more jets that don't fall cleanly on either Gaussian, i.e. more jets with a score very different from either 0 or 1. These jets still seem to behave more like subtropical jets than EDJs, potentially because most of their momentum still comes from the (sub) tropics and is conserved.

For comparison with our earlier categorization method, we reproduce figure 10 one last time, going through the new categorization method but with the previous choice of variables: that is longitude - latitude - pressure level. The results can be seen on the fifth and last row of figure 5.

We also show the differences in maps of jet occurence frequencies. The results can be seen on figure 3 (for the new method) and 4 (for the new method with the old choice of variables)

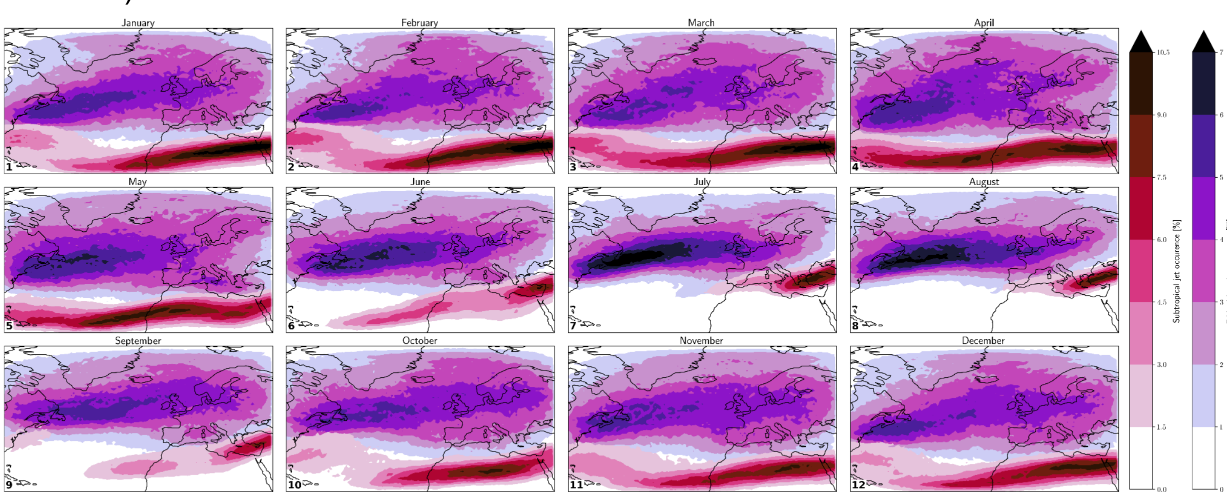

Figure 3: Jet occurence frequencies, in % of timesteps, for each jet category: pink for STJ and purple for EDJ. Following the new method and a cutoff at 0.5.

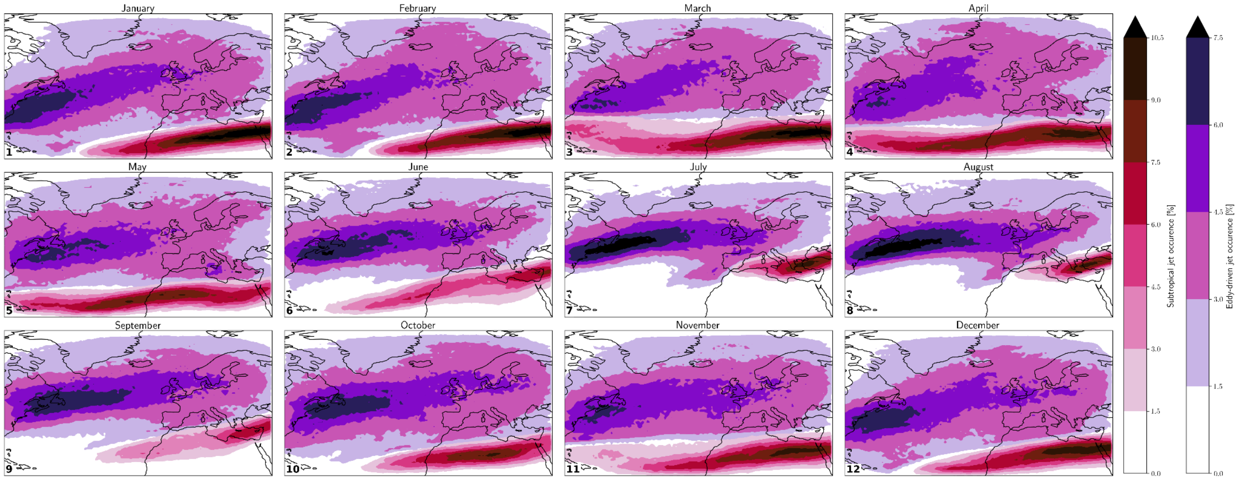

Figure 4: Jet occurrence frequencies, in % of timesteps, for each jet category: pink for STJ and purple for EDJ. Following the new method but with the earlier choice of variables: longitude, latitude and pressure levels. The cutoff between EDJ and STJ is at 0.5.

Specific Comments

Line 49: I am not quite sure what the authors mean by “double jet state.” Maybe this is referring to the simultaneous presence of both the EDJ and STJ? If so, then maybe the phrase “a persistent double jet state” could be replaced with something a little clearer, along the lines of “the persistent simultaneous presence of the EDJ and STJ”.
Yes, this is what we meant. The manuscript was modified (lines 52-54) to clarify this point.

Line 72: I think this review of past research in the introduction, especially on projected changes in the EDJ and STJ in warmer climates, is extremely thorough and thoughtful.
Thanks!

Line 110: I think it would be useful to mention what climatology these geopotential height anomalies are computed relative to. Maybe the authors used a daily climatology that includes all year from 1959 to 2022, for example?
You are right. The manuscript was modified (lines 117-120) to clarify this point. Note: instead of plotting the anomalies in an Appendix, we now compute weather regimes from them, and plot instead the overlap between SOM cluster occurrence and weather regime occurrence, to provide a more quantitative discussion on the links between SOM and weather regimes.

Lines 115-138: speaking as someone who is less familiar with SOMs than other researchers, I found the explanation of how SOM clustering works in this section to be very clear and appropriately concise.
Thank you!

Lines 119 - 120: I think a more intuitive phrase could be used instead of “arraying the clusters on the nodes of a regular 2D grid.” For example, the authors could write “The SOM adds another layer to this algorithm, by organizing the clusters into nodes within a regular 2D grid…”
We agree with this comment. The manuscript was modified (line 128) to clarify this point.

Lines 165-166: I am confused by what this sentence is meant to convey. Maybe the authors are just trying to say that the transition matrices allow a metric to be computed that communicates how predictable a SOM node is following the occurrence of a prior, different SOM node? Not sure, but I think this sentence needs to be reworded for clarity. The subsequent sentences in this paragraph, however, make perfect sense to me.
This sentence was indeed unclear. The whole section 2.2.3 has undergone a rework with slight modifications in persistence metrics definition. Please refer to our response to reviewer 2 for more details.

Lines 188-189: I think it is worth rewording this sentence that starts with “As a first difference to S17…” to be a little more clear and to make it obvious that vertical maxima in the wind speed U, specifically, is used to identify the presence of a jet. Maybe instead, the authors could write something like, “Whereas S17 identifies the presence of a jet using the vertical maximum in U on the 2PVU surface, we use vertical maxima over several high-altitude pressure levels.” I also recommend adding a brief clarification that “PVU” refers to potential vorticity units, where 1 PVU = 10^{-6} K kg^{-1} m^2 s^{-1}. I have observed that making this clarification is customary in peer-reviewed literature that references potential vorticity, even minimally.
We have reworked this section (2.3.1) heavily to accommodate for the few changes in the algorithm since the original submission, as well as to clarify the several points raised by referee 1.
Changes were made to the manuscript on line 224 to incorporate the comment about PVU.

Line 195: I am not sure what a “contour library” is, and I think this term might not be commonly used in scientific literature focused on synoptic meteorology. Does this refer to a publicly-available coding package for Python users, for example? I think some text

to briefly clarify what this term refers to should be added, or maybe there is a more widely used term to replace it with.
It was indeed referring to an open-source python package (contourpy). We are now using “routine” since we are really using one function, the package’s implementation of the marching square algorithm in C with python bindings. The word “algorithm” would have been a repetition in that particular sentence.
Changes were made to the manuscript on line 228 to incorporate this comment.

Lines 204-207: I find the text here to be confusing, especially because the criteria described here sound very similar to the criteria introduced in the previous paragraph (lines 197-203). Are the criteria described in these lines applied in addition to the criteria in the previous paragraph? If so, the statement that “jets are defined” with the application of this second set of criteria seems to contradict the statement that “jets are… extracted” using the criteria outlined in the previous paragraph. It sounds like the jets that are extracted in lines 197-203 may not officially be “jets” yet. In light of all this confusion, I think it is worth modifying the text on these lines (and possibly on lines 197-203) to make the step-by-step process here clearer.
These paragraphs needed a rework we agree. The three criteria are introduced first at the end of the second paragraph: two pointwise and one sequence / potential jet-wise. Then the step-by-step process is detailed in the two following paragraphs. Changes were made to the manuscript on lines 222-235 to incorporate this comment.

Lines 212 – 214: I feel as if saying “the two differences” here makes it seem as if these are the only two differences between this algorithm and S17, whereas there is at least one other difference I see mentioned in Section 2.3.1 (the use of vertical wind speed maxima over several high-altitude levels instead of just on the 2PVU surface). I think it would be easiest for readers to follow these differences if all of them were listed together in the same paragraph after the step-by-step description of the jet identification algorithm has been completed. For example, the authors could write something along the lines of “The difference in results from our algorithm and the S17 algorithm are due to three important differences in algorithms themselves.”, and then list each of the differences (including the departure from the use of the 2PVU surface) following this sentence.
This is a bit more difficult we believe. The sentence was written this way originally because we did not treat the choice of 2D field as a change in the algorithm per se, more as a “best practice”. Indeed, the S17 algorithm functions perfectly well when applied to, for instance, wind speed fields on a single pressure level. For clarity, we have reworded all comparisons with S17. These changes are visible on lines 245-255.

Line 222: “low altitudes” may be a typing error (i.e., the authors may have meant to write “high altitudes” instead). If this is not a typing error, then why is the averaging performed at low altitudes instead of high altitudes, given that we know jet cores reside at high altitudes? I think adding some text briefly explain this might be helpful.
We did mean low altitudes here. The authors of the article cited here wanted to isolate the EDJ, that is associated with vertically deep baroclinicity and is present at low altitudes, from the STJ, that is much more localized at high altitudes. You are right that this is not obvious and should be explained more. This has been done in the updated version of the manuscript, lines 257.

Lines 270-271: I am struggling to understand what the sentence starting with “One instead has to rely…” despite being very familiar with the Winters and Martin 2017 study. Part of my confusion comes from the fact that I do not know what variable is used to compute the “depth” that this sentence refers to – is this the depth of the drops in the tropopause height that accompany the presence of a jet? If so, why do low-level winds and latitude impact the use of these drops in tropopause height to categorize the jet? I think this sentence in the manuscript could be reworded or some specificity could be added to provide more clarity.
We agree that the term “depth” can be misleading here. We meant “jet vertical extent” or simply its baroclinicity, which is now the phrase we use. The manuscript has been modified line 307.

Lines 276-277: could the authors offer some explanation as to why they chose to use jet frequencies in mean latitude, mean longitude, and pressure level as opposed to jet frequencies in wind speed and potential temperature? This may help clear up some of my concern mentioned in the “General Comments” section of this document, which highlights my concern toward using the mean latitude, mean longitude, and pressure level approach. I know that the authors acknowledge that there may be some mis-categorization of jets using their method, but I worry this mis-categorization could potentially be too large using this method.
We hope to have addressed this comment with our response to this referee’s main comment.

Line 277: I think most readers will be familiar with the fitting of a unimodal Gaussian distribution to a 3D dataset, but I think many would be unfamiliar with what a “two-component Gaussian mixture” is refers to. Does this just involve fitting two overlapping Gaussian distributions to the 3D data? Some elaboration on what this fitting technique involves could be very useful.
You are correct. Along with the major rewrite of section 2.3.3, we have added two sentences to explain this. Lines 321-328:

…fits a two-component Gaussian Mixture model to facilitate the discovery of the two regions. A two-component Gaussian mixture model assumes that the data are bimodal and tries to fit their empirical distribution as a sum of two Gaussian distributions. Each Gaussian is defined by its mean and covariance matrix, and these are the parameters the model fits to the data. The density in the EDJ Gaussian component at each point, computed using the standardized distance to this component's center, is then used as a continuous score and not a hard assignment.

Line 279: I feel that text is missing from this paragraph which connects the construction of the histograms to the categorization of individual jet objects. Could a sentence or two be added that explains how the delineation between the EDJ and STJ regimes in the histograms are determined, and that an individual jet object is assigned either the STJ or EDJ classification depending on which regime it falls within?
We hope to have made this clearer with our response to the previous comment and the added sentences.

Line 321: I would argue that the EDJ appears fairly wavy in the top most row (nodes 1-6) as well, and I think this is worth noting. Unless the authors intended to use the word “columns” instead?
You’re right. It’s not possible to make such a broad statement about EDJ waviness here, so we just removed the sentence.

Lines 321- 323: I recommend either adding the exhaustive list of clusters indicating each weather regime, or adding “e.g.” prior to the listing of each clusters in the sets of parentheses to make it clear that these are not the only clusters featuring each weather regime.
You are right. With a similar comment raised by referee 2, we have added a quantitative comparison to the summer weather regimes and have reworded this section (3.1) to be more precise.

Line 330: I wonder if the authors could offer some intuition in the manuscript as to why cluster 8 is so common, especially since this is such a stark result from Figure 6. Please ignore this comment, however, if this topic is already discussed at a later point in the manuscript (I may have missed it). Additionally, I do wonder if cluster 8 could be seen as sort of a “catch all” cluster for any upper-level flow patterns that do not fit into other nodes, especially given that Figure A1 shows that composited geopotential height anomalies for this cluster are relatively weak compared to other clusters. I imagine that a Sammon map for the SOM may help indicate whether cluster 8 is indeed a “catch-all” cluster?
Cluster 8 in the original manuscript, now clusters 10, 11, 16, 17 and 23 with the more reproducible SOM (see details at the end) are catch all for early June. We have reworded this paragraph to remind the reader of this.

Sammon maps do not understand periodic boundary conditions, and I would not know how to modify them such that they do, since the size of the domain is hard to predict in advance and is meaningful in SMs, as far as I understand.

We think that figure 7 of the main text, plus this additional figure (5) showing the mean projection error on each cluster, that was added as an additional panel of the main text's figure 6, should support our claim that these clusters are indeed catch all for early June.

Finally, with a great addition from referee 2, we have altered slightly our definition of persistence. Persistence is still defined as the length of a stay on a SOM node before leaving it, but the loosened definition allows for jumps from the start cluster to another cluster as long as the distance between the two is below a certain threshold. Before, we based this on the grid distance, and set a threshold of 1. Now, we use the Euclidean distance between the cluster weights, and use as threshold the 10th percentile of inter-cluster distances.

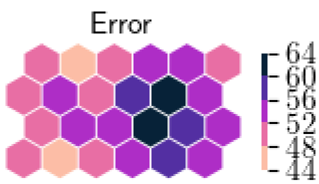

Figure 5: RMSE error, i.e. mean distance between instantaneous snapshots and the weight matrix of its best matching cluster, averaged by cluster. This is now a panel on the main text's figure 6.

Lines 367 - 370: I am confused on what information led the authors to the distinction made between the interpretation of the high residence times for clusters 2, 8, and 13 and the interpretation of the high residence times for cluster 17 and 18. I may be missing something critical that was mentioned earlier in the transcript, but I think some text could be added here to clarify why clusters 2, 8, and 13 are not representative of true state persistence.
Cluster 8 is a catch all for early June, and 2 and 13 fill a very similar role. We emphasize this point more in the updated version of the manuscript.

Lines 403-404: I think the text which reads, "All the results are split by jet category and always colored in the same way: pink for the STJ and purple for the EDJ. The double jet index is colored in black." would be the kind of descriptive text that belongs in the caption for Figure 10 as opposed to in the main text of the manuscript, since most readers will look to the caption first to understand the use of color in the figure.
You are right. This is changed in the updated version of the manuscript, where this figure has moved to position 8.

Line 489-491: I wonder if the authors could briefly offer a hypothesis or speculation in the manuscript text as to why the two methods of assessing persistence often disagree.
That is a difficult question with only pieces of answers. The three measures of persistence have several differences:

- Jet lifetime and SOM persistence are both nonlocal. One only knows about the lifetime of a jet when it has stopped being tracked, typically after being advected out of the domain. Similarly, one can only determine the length of a SOM stay after it's ended.
- Only the jet COM speed can be assessed locally in time.
- The SOM persistence can be influenced by mean wind variability, which is not the case for the two jet persistence metrics.
- The two jet related metrics only assess the persistence of one jet at a time, making them harder to interpret.

With a similar comment from referee 2, we explore this question more in the new section 3.4, focused on this question

Line 515: I think "waveguidability" is not a commonly-used word in the literature, and I myself am not quite sure what that means. Could the authors add some brief text to clarify what this term refers to?
It is indeed a more niche term than we first imagined. The term was discarded in favor of "the ability of the [..] jets to carry and guide Rossby waves" as per Wirth and Polster (2021), line 578.

Line 546: I really appreciate that the authors added a discussion on methods they tried and eventually rejected for jet feature extraction, and felt that it was a very thoughtful addition. I think the knowledge presented here will be highly useful for readers who may be interested in developing similar datasets involving feature extraction.
Thanks for your kind comments.

Figure 4 and Figure 4 caption:
I think the word "arrayed" is not commonly used and I worry that it will not be intuitive for many readers. Could a different word or phrase be used here to describe the display of jet frequencies in each hexagonal bin?
We have changed to projected and binned instead of arrayed and binned.

I am confused on why the units are arbitrary in this figure. Most plots of this nature have frequency units. For example, does panel (a) not show the count of identified summer jets per hexagonal bin? My reason for asking this is that if possible, I think adding a colorbar (or colorbars) to quantify the magnitudes of the shading here would be helpful.

I would certainly be very interested in seeing the magnitude of the frequencies here.
We have modified the figure to add two colorbars, with units of %.

The cutoff between EDJs and STJs is drawn through a very high-density region of the histogram in panel (b), which makes me feel concerned that classifications of jets as either EDJ or STJ could be very sensitive to where this cutoff is drawn. Could a sensitivity test be performed to illuminate how much the choice of this cutoff impacts the frequencies of each category both overall and a geographic map (i.e., in latitude/longitude space)?
We hope to have addressed this comment with our response to this referee's main comment.

Figure 5: I feel that some nodes in this figure further demonstrate my concern over how physical the delineation between the EDJ and STJ is. For example, the switch from the EDJ to the STJ categorization assigned to the relatively zonal jet in nodes 4 and 5 feels a bit arbitrary. I wonder if it could be more physically reasonable to think of most of the zonal extent of this jet as being more like a STJ in nature.
There is a subtlety here that is not explained in the paper. The SOM clusters are represented by the composites of horizontal wind speed fields of all timesteps belonging to them, where each snapshot is already the wind at the pressure level of maximum wind speed. To detect and categorize jets on SOM centers, we composite all the fields that we need on this horizontal surface (u, v, U, theta and pressure level). The long jets that switches, apparently arbitrarily, on clusters 4 and 5, did so because of a sharp jump in pressure level. A similar thing happens with the updated methods on cluster 1, because of a sharp jump in theta level in the composite.

Figure 7: I think it would be useful to know whether the population of each cluster is in units of days or timesteps. Could a phrase be added to the Figure 7 caption that mentions the units, and could the units be added to the colorbar in the figure as well?
It is in unit of timesteps. This was added in the new version.

Note for all figures: I really appreciate how clean and aesthetically pleasing all of the figures are. I found them very easy to read.
Thank you!

Technical Corrections

Line 19: missing "of" after "instead"

Line 21: missing "a" after "towards"

Line 36: missing connecting word after “clear”; maybe this should be modified to read “… always clear, since both sources of momentum…”

Line 47: I suggest using “favor” instead of “favorise”

Line 57: I suggest replacing “The signal is however weak…” with “However, the signal is weak…”

Line 89: I think the authors meant to write “stationarity” instead of “stationary”

Line 95: I suggest replacing “works” with “work”

Line 140: I suggest inserting “(June, July August)” after “JJA” to clearly define the acronym for readers

Line 151: misspelled “matrix”

Line 174: I recommend replacing “With a large SOM with sometimes…” with “With a large SOM that sometimes has…”

Line 182: I believe a reference to some other section of the paper and a closing parentheses bracket might be missing here

Lines 182-183: I recommend rewording “This is why we apply it to the full dataset…” to instead read “Therefore, we apply our detection method to all seasons within our dataset…”

Line 188: I think the authors may have meant to write something along the lines of “As done prior to training the SOM,” instead of “As for the SOM”

Line 212: I suggest replacing “difference” with “differences”

Line 273: I suggest replacing “categorization bins and counts” with “categorization involving binning and counting”

Line 283: misspelled “characterized”

Line 286: I think the word “algorithm” may be missing after “A straight forward feature tracking”

Line 303: I recommend spelling out the word “hour” in “6H-timesteps” to follow the convention used in previous sentences

Line 308: I suggest replacing “result” with “results”

Line 317: I recommend replacing “high-zonal-overlap double jet states” with “the double jets with high zonal overlap” to improve sentence flow here

Line 319: I recommend replacing “centre right” with “center-right” to follow the convention used in the previous sentence, which uses “center-left”

Line 328: I recommend replacing “honeycomb” with “hexagonal” to follow conventions used in previous paragraphs

Line 332: I recommend replacing “panel b” with “panel (b),” and following this convention at other similar places in the manuscript text

Line 342: missing “in” after “resulting”

Line 353: misspelled “occurrence”; I recommend spelling out “JA” as “June and August,” or clarifying that “JA” refers to “July and August” in order to establish the meaning of the acronym

Line 354: I recommend replacing “…in early June and cluster 2…” with “…in early June, cluster 2…”; I also recommend replacing “…early June, cluster 8…” with “…early June, and cluster 8…”

Line 363: I recommend replacing “to flow” with “the large-scale flow pattern” or “flow pattern”

Lines 363-364: I recommend replacing “…into an next state and the 95th percentiles…” with “…into the next state, while the 95th percentiles…”

Lines 397: I recommend replacing “way” with “ways”

Line 400: I recommend replacing “width’s” with “width”

Line 412: I recommend replacing “…between STJ latitude and Hadley cell edge…” with “…between STJ latitude and the Hadley cell edge…”

Line 420: I recommend replacing “questions” with “question”

Line 433 – 434: I recommend rewriting “Some trends. like for the double jet index, even change signs between summer and all-year.” as “Some trends exhibit different signs when comparing the summer period to the full year, such as for the double jet index.” to improve the sentence flow and fix punctuation errors

Line 435: I suggest replacing “accord” with “agreement”

Line 449: misspelled “across”

Line 450: “June” was not capitalized

Line 481: I think the word “known” is not supposed to be here

Line 478: I recommend adding “in upper-level flow patterns” after “seasonal shift” to be more a little more precise

Line 485: I recommend defining STJ and EDJ as referring to the subtropical jet and eddy-driven jet prior to their first use in Section 4: Discussion and summary

Lines 493-494: I recommend rewriting “Computing the properties the jet features of every time step” as “Computing properties of the jet features at every time step”

Line 539: misspelled “expectedly”

All the technical comments were addressed with the updated manuscript. Thanks a lot for this in-depth editorial work.

Referee 2:

Summary:

The authors assess trends and intra-seasonal variability of jet streams over the Euro-Atlantic region in summer. A wide ensemble of subtropical and eddy-driven jet streams’ properties is analyzed, with the intention to utilize the know-how in follow-up studies into summer extreme weather. To better understand the spatial structure of jet streams, a self-organizing map of wind fields is trained. I consider the study a worthwhile addition to the effort of understanding atmospheric circulation. However, there are several major issues that the authors should address.

We thank referee 2 for their comments and thorough reviews. In the following, we hope to address all points raised by them.

Major Comments

The majority of the paper focuses on summer jets. While some results are presented on an annual scale, which effectively puts the summer findings into perspective, I believe the primary focus on summer should be more explicitly stated in the title and abstract.

Yes, there is a focus on the summer jets, however, in the second section of the paper we also discuss year-round jet metrics. We prefer not to change the title since it would become overly complex. However, we add a comment to the abstract to highlight the focus on the summer season.

First, we apply the self-organizing map (SOM) clustering algorithm to create a 2D distance-preserving discrete feature space to the tropopause-level summer wind field over the North Atlantic. The dynamics of the tropopause-level summer wind can then be described by the time series of visited SOM clusters, in which a long stay in a given cluster relates to a persistent state and a rapid transition between clusters that are far apart relates to a sudden considerable shift in the configuration of upper-level flow. Second, we adapt and apply a jet axis detection and tracking algorithm to extract individual jets and classify them in the canonical categories of eddy-driven and subtropical jets (EDJ and STJ, respectively). Then, we compute a wide range of jet indices on each jet for the entire year to provide easily interpretable scalar time series representing upper-tropospheric dynamics.

This work will exclusively focus on the characterization of historical trends, seasonal cycles, and other statistical properties of the jet stream dynamics, while ongoing and future work will use the tools presented here and apply them to the study of connections between jet dynamics and extreme weather. The SOM allows the identification of specific summer jet configurations, each one representative of a large number of days in historical time series, whose frequency or persistence had increased or decreased in the last decades. Detecting and categorizing jets adds a layer of interpretability and precision to previously and newly defined jet properties, allowing for a finer characterization of their trends and seasonal signals.

Given that jet streams are highly localized features, I find it potentially problematic to train the SOMs on largescale wind fields. It is unclear to what extent the properties of the SOM clusters reflect the jets themselves versus circulation variability far from the individual jets. I wonder how the SOM (and your subsequent results) would differ if only the identified jets were used as input instead of the entire wind fields.

The classification of a wind field into a particular SOM cluster strongly depends on the field's average wind speed. This dependence can easily introduce unwanted artifacts in any classification based on Euclidean distance—specifically, the field might be classified with the centroid that has the most similar average wind speed rather than the most similar pattern. One way to address this issue is to use a different measure of similarity, such as pattern correlation, although this is not feasible with SOMs. An alternative approach would be to remove the fields' means before training the SOM.

Was the input dataset weighted by latitude before training the SOMs?

This is a very important question that we hope to respond to satisfyingly here.

Firstly, in this basin, and with our method to create fields by maximum wind speed in the vertical, jets take a lot of space in each snapshot and are virtually the only high wind speed areas. With the coarsening to 1.5°, most smaller scale features are erased anyways. Jets are also known to explain most of the variability in the Atlantic basin (Athanasiadis et al. 2010, for example). Jets are high intensity features, so they make up most of the Euclidean distance between two nodes anyways. We can show that jets are picked up by the SOM, and not the smaller-scale variability. We plot composites of frequency of identified jets for each SOM cluster and compare them against the jets identified in the SOM cluster centers. They match very closely on most of the 24 clusters, as one can see on figure 6. This figure was added to the updated version of the manuscript as figure 11.

We standardize the data before training and weigh them by the cosine of latitude. We added this information and more such technical details lines 162-165.

Training SOMs with the jets as input is a good idea that we hope to explore further in future work. However, a good distance metric would have to be found, because Euclidean or SSIM won't give satisfying results. Since jet cores are spatially very sparse, two timesteps with the same jet configuration but with a constant latitudinal shift of, say, 1° in the EDJ core in the second time step would result in high Euclidean distance even though the original training with wind speed fields would find a lower one, and that latter behaviour is closer to the what we want and expect. Any Minkowski distance will have the same problem. Furthermore, we wanted the two methods to be mostly independent from each other, so that potential shortcomings of one (e.g. not all jets are found) does not alter the performances of the other. This is an arbitrary choice granted.

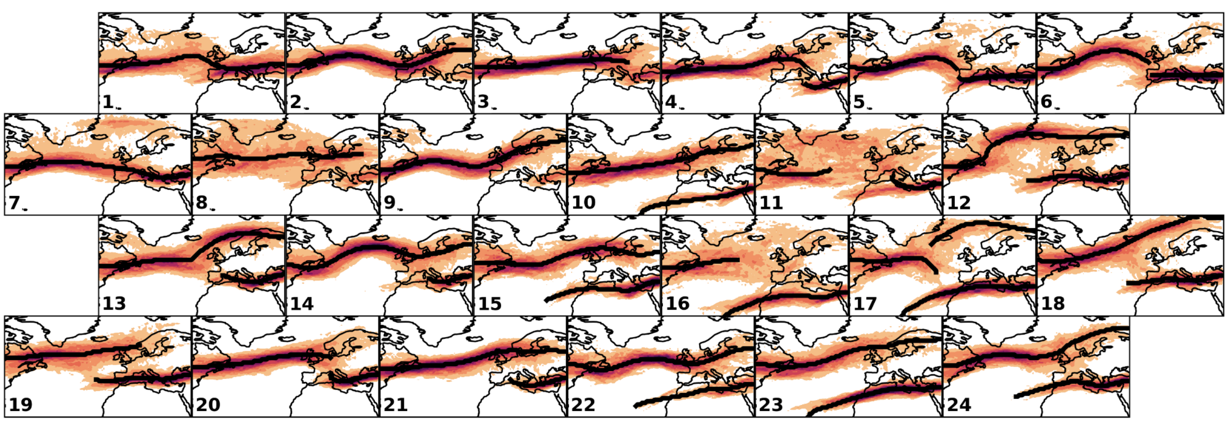

Figure 6: Composites of per-cluster detected jet frequency (shading) and jets detected on the cluster wind speed composites (black lines).

Section 3.4, where jet properties are projected onto SOMs, appears too brief. This section integrates the two perspectives and brings new insights, but it lacks depth. For instance, a quantitative comparison between independent jet properties projected onto SOMs and those calculated from SOM centroid patterns is missing. In line 454, the authors claim that there is a good match but do not provide quantitative evidence. Can jet properties be calculated from SOM clusters (e.g., the patterns in Figure 5), and can the SOM's skill be quantified? Figure 13 highlights several cases where there is a poor match, but these are not discussed in the text. For example, why is the COM speed of the EDJ so much greater in cluster 1 than in cluster 21? Similarly, why is R16 of the EDJ identical for clusters 5 and 16? This section should also serve as the foundation for a critical assessment of the application of SOMs to analyzing jets, which is missing from the current Discussion. Finally, the chosen color scales in Figure 5 are not ideal. While the blue and red monochromatic scales are understandable, using the scale applied for the Double Jet Index would likely be more informative."

Thank you for your suggestion. We had scrapped some of this discussion given the length of the paper already, but you are correct that it is now too short. We have added a new figure, mirroring figure 13 (now figure 12 in the updated version of the manuscript) but with properties computed directly on the SOM cluster centers. This is figure 8 in this document and figure C1 on the updated manuscript. We have also lengthened the discussion in this section a lot, based on these two figures and the discussion on persistence metrics that we elaborate on in a response to another one of your comments and another from referee 1.

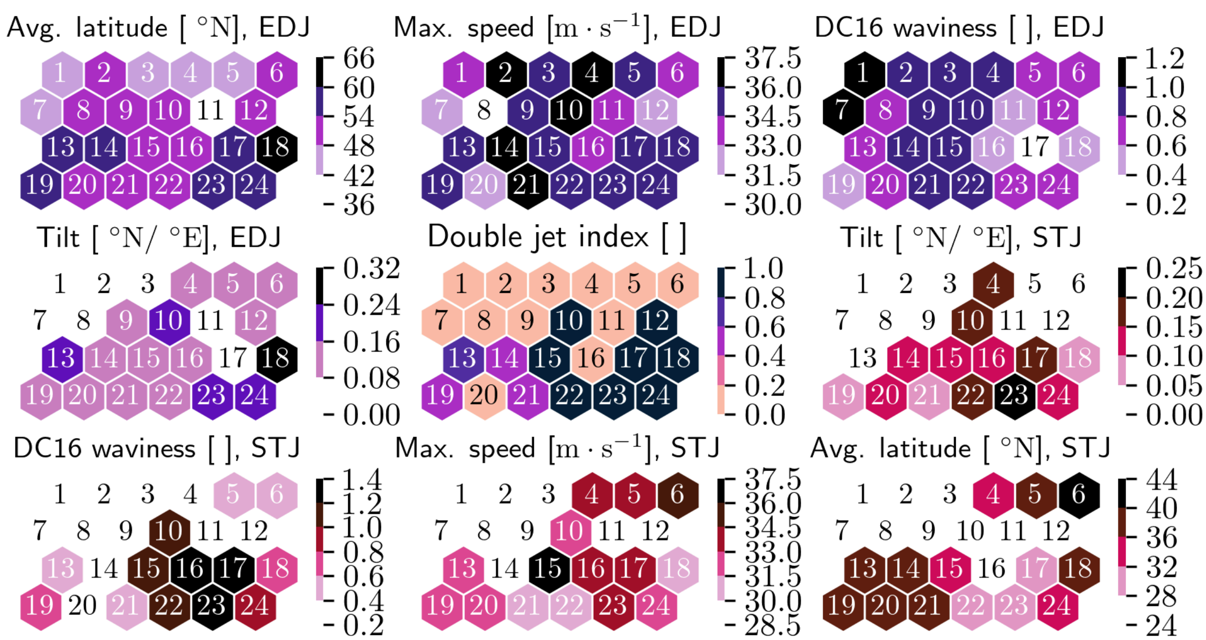

Figure 7: Jet properties, separated by jet category when applicable, projected on the SOM clusters. Shades of purple corresponds to EDJ properties and shades of pink to STJ.

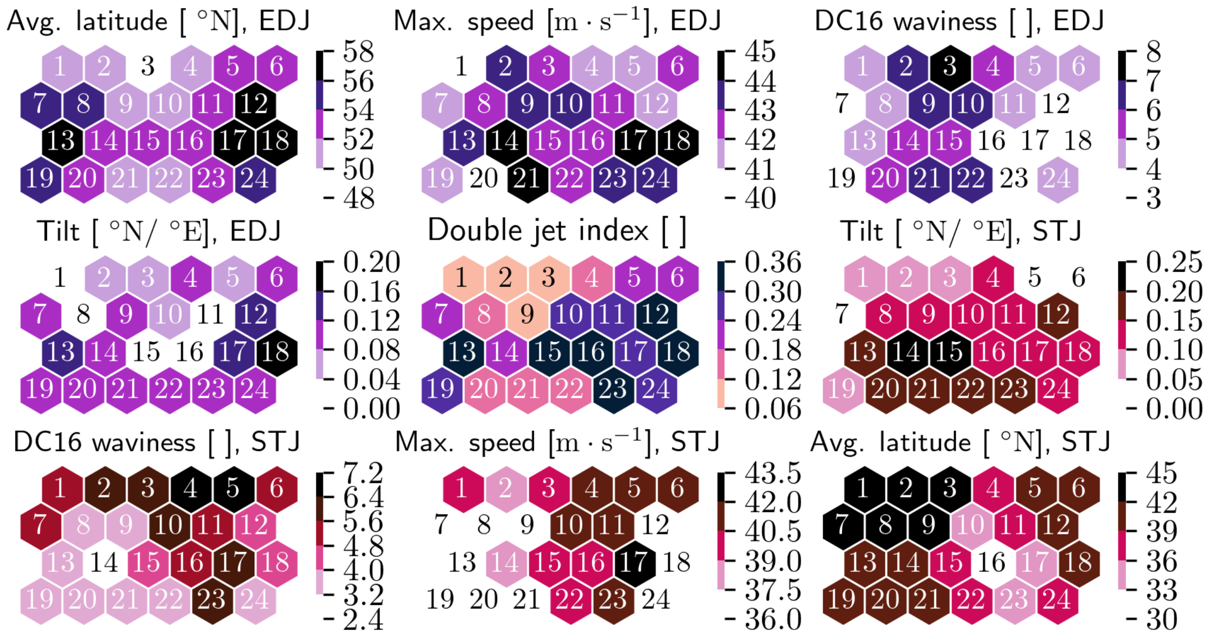

Figure 8: Jet properties, separated by jet category when applicable, computed from the jets detected on the SOM cluster centers. Shades of purple corresponds to EDJ properties and shades of pink to ST

It is not clear how the persistence based on SOM nodes is calculated. The authors mention a threshold of one, allowing departures from a given node without breaking a current episode. First, defining the threshold based on the highly idealized regular 2D grid is possible but it does not reflect the real distances between the nodes (cluster centroids). Defining the threshold for instance as the median Euclidean distance between neighboring nodes in the original high-dimensional data space may be a simple was to deal with the intra-summer variability of wind fields. Second, it is not clear how the individual episodes are found. For instance, imagine one has the following sequence of clusters: 15 1 8 8 8 2 8 8 8 15. Your description suggests that the following episodes are identified: a) 15, length=1; b) 1, length=8; c) 8, length=7; d) 2, length=4; e) possibly another 8, length=3 (?); f) 15m length=1. Is this correct? Please clarify in Sect. 2.2.3.

First point: this is an excellent idea. Figure 8 and methods were updated to use this threshold as it makes a lot more sense.

Second: We will clarify this in the text. We see how this interpretation can be made from the text. The algorithm does not double count and assigns the stay to the most visited cluster. For your example it would output : a) 15, length 1, then b) 8 (majority vote), length 8, and finally c) 15, length 1.

The manuscript was updated, lines 189-195:

…, this second definition with a small distance can be a more realistic measure of persistence. To account for varying degrees of similarity between neighboring clusters, we do not use the discrete grid distance between clusters but instead the Euclidean distance between SOM cluster weights. With a given

Euclidean distance threshold, say the 10th percentile of inter-cluster weight distances, the definition of persistence can therefore. At the start of summer, the first stay starts on whichever cluster is populated on the first of June, cluster i_0. As long as $\mathrm{BMU}(t)$ is on i_0 or any other cluster whose weight matrix is at a low enough distance from that of i_0, the stay continues. Once $\mathrm{BMU}(t)$ has left this neighborhood to arrive at another cluster i_1, a new stay starts with the same condition for it to terminate. The persistent stay is associated to the most visited cluster during the stay, which is not necessarily the starting one.

Several times throughout the paper (see minor comments below), the similarity of SOM patterns with modes of variability and weather regimes is mentioned. However, aside from the NAOI, this is presented only for selected patterns and not quantitatively. Linking the findings to established modes and/or regimes is an excellent idea that should be developed in greater detail.

We agree and have made these discussions of similarity more thorough and quantitative in the updated version of the manuscript. Figure 9, which will be incorporated as panels to figure 6 in the main text, quantifies these similarities to the summer regimes, calculated following a methodology similar to Cassou et al. (2005) and Grams et al. (2017). 40% of summer days are not assigned to any of the 4 clusters, if we follow the criterion of Grams et al. (2017).

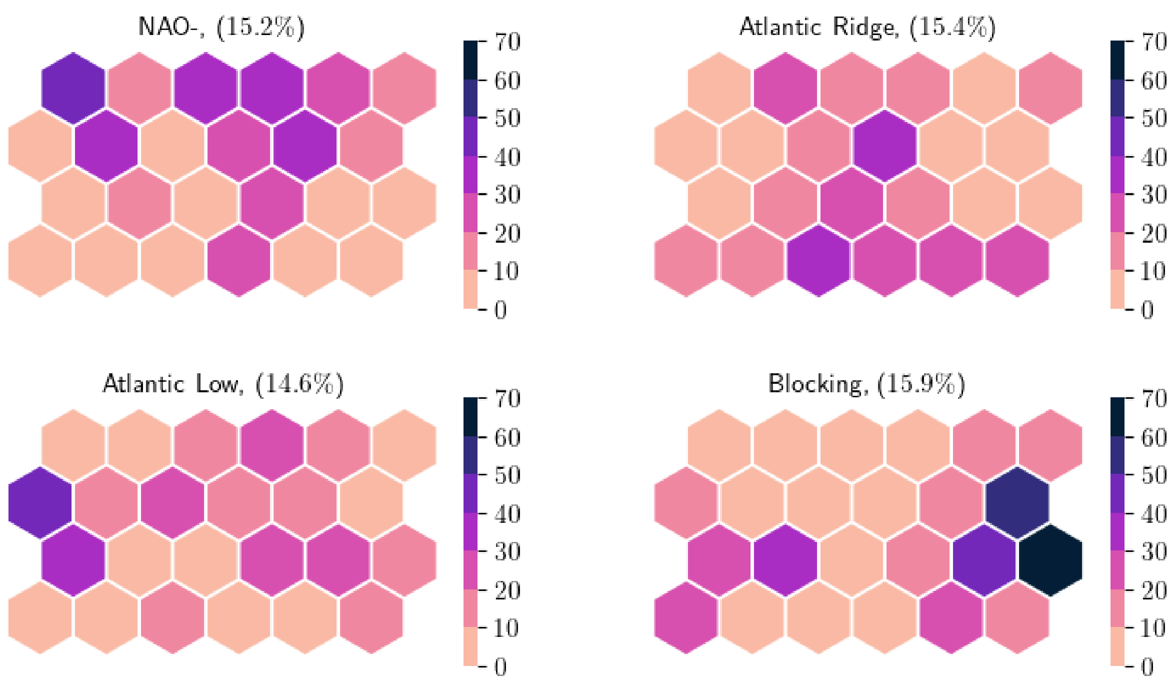

Figure 9: Percentage of timesteps associated to a summer weather regime for each SOM cluster. These are now panels of the updated manuscript's figure 6.

The quality of the Discussion and Summary section could be improved. Parts of Section 4 are overly vague and difficult to follow, and the section lacks a clear summary of key results. Additionally, portions of the Discussion appear only in earlier sections (e.g., the

description of annual results in lines 343–441), which disrupts the flow and focus. Consider restructuring or splitting the section to enhance clarity and coherence.
We have reworked the various result subsections and the Discussion and Summary section to improve the structure and clarity of the manuscript. Please refer to the end of this document where these changes are detailed.

Minor comments

Line 13: The term ”seasonal cycles” is not clear to me; I suppose that the authors mean intra-seasonal variability? This appears several times in the paper.
We will use intra-seasonal variability. It is more standard even though we believe seasonal cycle was more expressive.

Line 19: abbr. 2PVU not defined, and I think that the term “flattened pressure fields” may also be unknown to many readers
The term is introduced in data, but you are right it is highly nonstandard. All subsequent uses of the term were reworded.

Line 24: “sudden flow transition in June” seems unclear to me at this point
Thanks. We have reworded this sentence in the manuscript, lines 25-26.

Line 52: both references are missing from the references list
Thanks. This is fixed in the new version of the manuscript.

Lines 77–78: The term 'data-driven approaches' does not seem appropriate to me. Perhaps terms like 'complex,' 'blackbox-like,' or 'dimension-reduction-based' could be more fitting for these approaches. Neither of the approaches you use is strictly objective or subjective; they both require user/expert decisions and involve many parameters that need to be defined, which can potentially have a significant impact on the quality of the results. In the context of using SOMs in general, and specifically to study the link between circulation and extremes, I recommend referring to studies by Gibson et al. (10.1002/2016JD026256) and Stryhal et al. (10.1002/joc.7996).
You are correct that this term is not the most appropriate here. We will use “statistical” instead and put it in relation to expert defined features as before, lines 82-83.

Line 82: It is not clear to me which interactions you are referring to here
Thanks. We have reworded this sentence in the manuscript, lines 87-88. We were referring to interactions mentioned three paragraphs above.

Line 89: “tool to study stationary and recurrence”

Thanks. We fixed this typo in the manuscript.

Line 89: Please change your reference from preprint to the final paper (https://esd.copernicus.org/articles/14/955/2023/)
Fixed.

Lines 93-98: the use of language in this paragraph needs to be reviewed
We have reworded this paragraph (see next comment).

Lines 93-98: assess the seasonal cycle vs focus on summer … please reword
We have reworded this paragraph to clarify our focus, lines 98-103:
After presenting both techniques in detail, we demonstrate their capabilities on reanalysis data. This work focuses more on summer than the rest of the year. This season receives less attention when designing methods to characterize the circulation, and is yet very important for extreme events and which presents interesting, different trends compared to the rest of the year (Harvey et al., 2023). The SOM will only be trained on summer days, but the jets are detected on year-round data to provide more context to the summer results. We assess the seasonal variability of the upper level circulation, find the trends, or lack thereof, in various metrics, and study to the circulation persistence under different aspects.

Line 102: ERA5 provides much more that that; please reword
You are right, this was reworded lines 106-108.

Line 104: I suggest mentioning which levels are used exactly
You are right. This is done in the updated version of the manuscript, line 109.

Line 110: “geopotential”
Typo corrected.

Line 110: What kind of anomalies do you use?
Day-of-year, smoothed climatology. This was added in the methods lines 115-116 you are correct it was missing,.

Line 129: I am not sure that I understand what the “scale parameter” means. It is not used in Kohonen (2013) or the very well-known and cited papers on SOMs in atmospheric science, Hewitson and Crane (https://www.intres.com/abstracts/cr/v22/n1/p13-26/) or Sheridan and Lee (10.1177/0309133310397582). On the other hand, parameters such as (learning) rate and (neighborhood) radius, which considerably affect the resulting map, are not mentioned.
Learning rate is irrelevant in the “batch” training algorithm that we use (see kohonen 2013).

We used “scale parameter” for sigma as it is fairly standard way to describe a distribution’s spread. We reworded to “neighborhood radius” line 133, as it is indeed more standard in SOM literature. Thank you.

Lines 133-134: This claim would benefit from a reference. Furthermore, instead of 'allowing,' I would suggest using 'forcing.' Arguably, it is not the similarity of nodes that causes this, but rather the tendency of SOMs to overrepresent the center of the data space and under-represent its margins (i.e., extreme fields). This is compounded by the 2D constraint, where nodes lie on a plane even for multi-modal data, or more generally, in a space with much lower dimensionality than that of the data.
We didn’t spend too much time justifying this claim, but a similar one is made by Gibson et al. (2017) when the authors compare SOM-p (nonzero radius) to SOM-c (zero radius). From eq. 8 in Kohonen 2013, each cluster weight is defined by a weighted mean over all the data points, not just the ones that are closest to it, and the weights are the outputs of neighborhood function. This means that if the neighborhood radius is bigger than 0 at the end (or 1 in the nomenclature of Gibson et al. 2017), then each cluster bleeds into all the others, as its members have a nonzero contribution to all the others. We have added a reference to Gibson et al. (2017) and rephrased this claim to be more conservative.

Forcing is indeed a better choice of word.

We agree that the over-representation of central points is a problem with non-periodic boundaries, but not in our case. With periodic boundaries and nonzero radius at the end of training, then our justification above holds.

The last point is also valid and could be added to the discussion.

Line 135: “projected trajectory” is not clear
Reworded to “the trajectory when expressed as a succession of cluster visits”, line 142.

Line 135: I do not think that *x* was defined
That’s correct thank you. We have explicited this quantity lines 144-145.

Line 144 + Figure 1: Many readers may not be familiar with periodic boundaries, as planar topology is typically used in atmospheric science. Please provide more details on the toroidal SOM topology employed and explain why it was utilized instead.
The sentence directly after explains the only consequence of PBCs: the top and bottom rows are a distance one away from each other, and the same is true of the left and right columns. We added a sentence justifying why we used it: to avoid an artificial

over-representation of the center clusters, which would make our persistence metrics less relevant. Lines 155-156.

Line 165: reword

This paragraph was also pointed out by referee 1 and was reworded.

Line 182: Is there a reference missing?

A silent latex syntax error erased the hyperlink to the “next section”. Thanks for pointing it out.

Figure 4: “a, b) For each season, here JJA, …” is a bit confusing, since the demonstration is only for JJA in these panels

Figure 4 and its caption were modified, following an update in the jet categorization method after a comment from referee 1

Line 268: “this framework…” is not clear, please reword the sentence

We reworded to “this work” (line 295).

Lines 300-301: This is not clear to me

This is an almost irrelevant technicality that we simply removed in the updated version of the manuscript. We meant that if the tracking algorithm receives as input only daily summer data for instance, then a flag should not be carried over from 31-08-2020 to 01-06-2021, even though it is the next timestep in the input data. We never use this feature anymore anyways, since the jets are found and tracked in year-round data.

Line 304: Please explain why COM speed is considered a measure of persistence? I am not familiar with this metric and it seems a little bit counterintuitive.

COM speed helps characterize persistence because a low COM speed represents a jet not moving much between timesteps.

Line 315: phase > phase space

Thanks. We fixed this typo in the manuscript.

Line 321: maybe “edge” instead of “extremal”?

That is a better choice indeed, thanks.

Line 322+onwards: – It is not clear to me how the SOM nodes were associated with regimes; an objective analysis is carried out only for NAOI. For example, why was node2 associated with Scandinavian blocking and node 3 was not? Why was node 22? If you feel that this is an important addition to your study, an objective analysis involving

lower-order modes of variability (EOFs) and/or weather regimes should be carried out. Otherwise, one may doubt the reliability of interpretations, such as that in line 336-338
We have added a more quantitative description of the similarities to weather regimes in this section (3.1), including panels in figure 6.

Figure 6: Correct “b” and “c” descriptors
The figure and its caption were updated, thank you.

Lines 332-334: While I find this description mostly acceptable, there are some misconceptions and inaccuracies: 1. What would constitute a perfect alignment of SOM axes with the NAO? 2. You suggest that the SOM you are using (which is technically a 3D lattice organized on the surface of a torus and then unwrapped onto a 2D surface for visualization) should align with the leading two modes of variability (note that the second mode is not analyzed at all in your paper). Can you support this with a reference? I found a study
(10.1029/2023JD039183) suggesting that SOMs behave this way only in very specific cases, but only planar grids were tested in that study. 3. It is unclear how relevant your suggestion regarding the PCs is, given that your object of interest is jets.
1. I would have expected NAO to align with the long axis of the SOM, as it is the main mode of variability. The fact that it does not goes in the direction of your other sub-point here.
2. We thought it intuitive that forcing the variability of a complex high-dimensional system on a 2D space would result in something related to PCs and didn’t spend too much time investigating this. With the findings of the study you mention, it is clear that the story is not as simple, and we have reworded this part.

I do not understand your side comment about a 3D lattice. Even though it has PBCs and can therefore be thought of as lying on the surface of a torus, the lattice is 2D.

Lines 339-342: Is this “closer inspection” rooted in any of the presented analyses? I cannot find anything that would support this. Please clarify.
The whole paragraph was reworked to accommodate the new figure 6. This wording referred to a closer inspection of the composite maps on figure 5.

Figure 7: The frequency of occurrence of cluster 8 in Week 1 is striking. Do you have an explanation for this? Averaged over almost 65 summers, how is it possible that this cluster is so much more likely in Week 1 than in Week2? One would expect that the pathway will be much more gradual.
We think it is a conjunction of two factors. First, the STJ is physically moving poleward fast in this period, according to figure 10, and with little noise around the mean. Second,

the random initialization that we were using in the first submitted version of the manuscript happened to concentrate all of the first week of June variability into one cluster. In the new version, the SOM reproducibly spreads it out more into 3 to 5 clusters, and the pathway looks more gradual.

Line 349: “Now we add a temporal dimension” is vague and inaccurate; was not the temporal dimension included before?
This was an awkward sentence you are right. We have reworded it line 387. The previous plots were all temporal aggregates (means, and slopes of regressions), so the temporal dimension had been removed before the plots.

Lines 351-352: Figure 7 does not show empty clusters –is it possible that all clusters are visited during the whole summer? Please reword.
We now use a more prudent wording.

Line 352: “left to right” is not clear
We have reworded this slightly, line 391.

Line 352 I do not think “JA” is very clear
We have reworded this, thanks also to a comment from referee 1.

Line 353: “early June”is unclear; Do you speak about the first week only (based on Week 1 in Figure 7)? Moreover, the frequency of clusters varies a lot in June, so it is not clear why NAO is described only for these three particular clusters. Additionally, your current wording suggests that NAO in clusters varies week to week.
This is a leftover sentence meant to transition into a paragraph talking about the “intransivity” hypothesis. It is very interesting to us that the first week of June can basically only sit on 3-5 clusters out of 24, whose average NAO is either negative, positive, or neutral. On a given year, typically only one of these 3-5 clusters is visited during early June before the transition to the mid-to-late summer clusters, and never again that year. Years that start summer on cluster, say, 13, which is NAO-, have a very different late summer cluster distribution (and associated weather impacts) than the years that started by spending the first week of June on cluster 8, which is NAO neutral. This part of the paper was scrapped to explore this hypothesis in a lot more details, so this sentence feels disconnected and useless. Sorry about this oversight. We removed it in the new version of the manuscript.

Line 355: “these few weeks” is unclear. Previous paragraph described only the first week

You are right. The statement holds for the second and maybe the third but to a lesser degree and makes the sentence less clear. We have reworded it to “first week”.

Line 356: What kind of error do you refer to?
Root mean square error (RMSE) between cluster center and cluster members. We have made this sentence more precise and added RMSE to figure 6 as panel c.

Lines 357-8: Do the days have free will in how they choose clusters? Consider rewording.
This whole paragraph was reworked with more prudent wording.

Lines 350-359: It seems to me that in June there might still be a considerable decrease in wind speed across the whole region. Since you do not subtract this seasonal change (e.g. by removing each field’s average wind speed), the SOM badly represents early June wind patterns. A few nodes represent a significant part of the data space (in terms of its volume). Consequently, the cluster centroids are very poor models of the fields classified with the clusters, and, in turn, very likely also of the jets’ properties. This is a very common issue if one uses classifications based on the Euclidean distance, which can easily lead to misinterpretation. For instance, the high persistence of cluster 8 might be an artifact of this issue; or, the positive trend in its occurrence might be unrelated to changes in jets, but may rather reflect a minor trend in wind speed, possibly even in regions far from jets’ occurrence. Consider additional sensitivity studies to make sure that statistical artifacts do not affect your results, or at least mention in the Discussion.
We added a map showing that the SOMs do indeed mostly capture the jets and not so much other features of the wind speed fields. Still, the error is greater in early June than the rest of summer, although in the same order of magnitude, and only for about 10 days out of 90. If the SOM is to remain independent of the jet identification method, which is what we want so that comparing their results remain meaningful, then any metric we use is bound to have such problem, as far as I am aware.

Line 361: I would suggest referring to the specific section given the length of the Methods section
You are right.

Line 363: “an next state”
This typo was fixed, thank you.

Lines 366-367: I disagree with your explanation of hotspots in the corners. Given the grid topology and the values for clusters 1, 7, and 24, it seems inaccurate. I observe one hotspot in the cross-section of the SOM torus, specifically at 2-8-13-14-20 (which

could represent early summer patterns), and another around 17-18 (corresponding to mid-to-late summer patterns). I suggest visualizing the data space and the SOM in 3D, for example using Sammon mapping. This is a useful tool for understanding the structure of the data and how the SOM represents it. It may also be a better alternative to your Figure 1.
Thank you for the suggestion. We investigated this but couldn't find a satisfactory way to do this, given the periodic boundary conditions of the SOM grid, so we have left this for future work

370 I do not understand the term “true state persistence”. Please reword.
Persistence on cluster 8 was artificial because it was the default cluster for early June but usually a bad representation of the daily snapshots that compose it. Persistence on most other clusters that do not have such a large projection error were a lot more meaningful. With the new methods and your suggestion on a better characterization of persistence, this spurious persistence does not show anymore.
The persistence discussion was moved to its own section, 3.4, to give it more space and discuss similarities and differences between SOM state and episodic persistence, and persistence of the jet objects in the domain.

Line 370: To what extent is persistence only (the liner) function of the frequency of occurrence?
Trends in yearly persistence are almost entirely a linear function of the yearly trends in frequency of occurrence. In the updated version of figure 8, now figure 13, we have discarded the trends since they can hardly be interpreted as anything else than artifacts of the trends in population.

Line 375: Rather than linear trends, one may utilize the SOM grid and visualize the seasonal frequencies and persistence of each node as x,y plots
Thanks for the suggestion. The result of doing this for cluster frequencies (in timesteps per year) can be seen on figure 9. I am not sure this adds enough value to the discussion (over linear trends) for the space it takes. For linear trends in residence times, we have simply removed them as we deemed them not meaningful, since as we discussed in the original manuscript they follow linear trends in frequency and their significance is dubious .

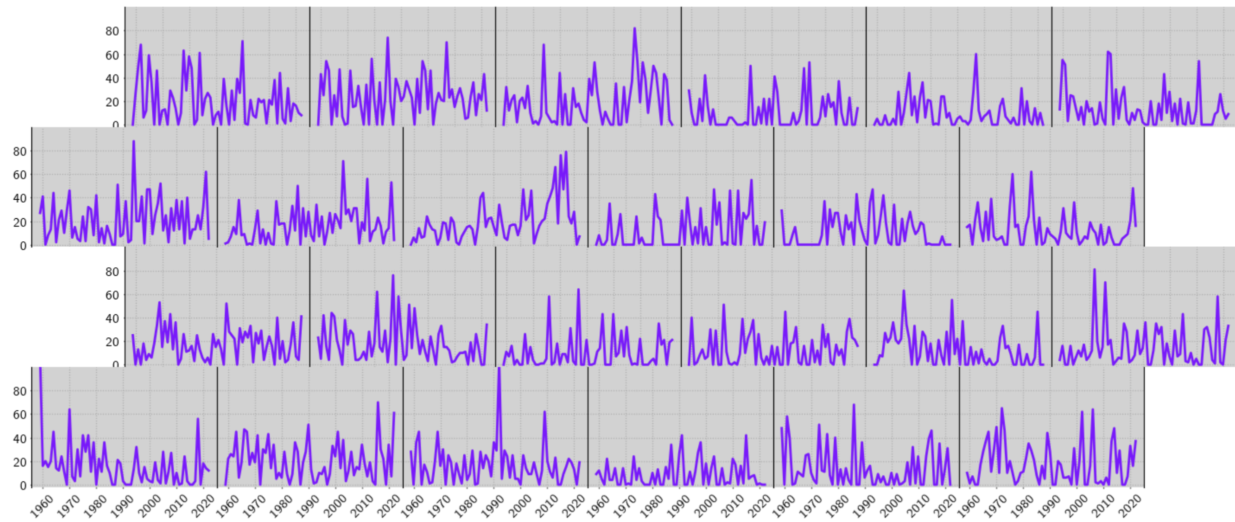

Figure 9: Yearly cluster frequencies, in time steps per year. Trends in populations are significant for four clusters: 1 (-), 3 (+), 9 (+) and 19 (-).

Line 382: Are you sure that cluster 6 resembles Greenland blocking? It does not seem obvious from Fig. 5. Maybe including the z500 patterns currently in Supplements would help.

Will reword this section more carefully and back it up with numbers. With the new figure 6 showing the association with weather regimes, we choose to leave the z500 pattern plot in the Appendix.

Line 384: Which clusters do you refer to?

We were referring to clusters 20, 5, 16 and 23. We have reworded this sentence with the updated plot (with the new SOM), and spelling out July-August.

Line 395: JLI and JSI were already defined in 2.3.2; R16 technically wasn't defined yet; abbreviations are not used consistently in the whole section. I would suggest not using them outside of pictures to make the text clearer. "this property" is not clear

We wanted to shortly re define the six properties defined. We have reworded this section with more careful uses of abbreviations, and use "departure from zonality" as another word for waviness that avoids repetition.

Line 412-413: Consider moving to Discussion

Thank you for your suggestion. We have moved this sentence about STJ latitude and Hadley cell to the discussion.

Lines 418-419: How? This should be discussed more and moved to Discussion

The relative positions of the jets alter the likely position and occurrence frequency of RWB events, discussed in Barnes and Hartmann (2012). We have fleshed out this comment more with this reference. This question is one of the many we hope to give more time to in future work. Thank you.

Line 469: “clear season signal” is too vague and could be easily mistaken for summer trends; consider intra-seasonal variability or similar instead
We have reworded to intra-seasonal variation, thanks for the suggestion.

Line 469: I am missing a section that would show and discuss the differences between the methods and their limitations. These paragraphs read a little bit too vague and one-sided.
Thank you for your suggestion. We have fleshed out this paragraph to include more discussion on the differences between methods and their results.

Lines 474-475: Did not you include the other group of nodes (that around cluster 8) in this description intentionally? If so, please discuss the reasoning?
Yes, because of their particular role as early June representative. With the new persistence metric, the early June clusters are not artificially persistent and this potential point of confusion does not have to be explained.

Lines 481-482: Reword + Which features?
We have reworded to “individual clusters”.

Line 485: It makes sense to define the abbreviations again in the summary section – but why so late?
That was a mistake after a last-minute edition of the first paragraph of conclusion. It’s fixed now, thank you.

Line 492: “subjective jet properties” is not clear since each method is an “objective” algorithm based on many subjective or expert choices
We have reworded to “pre-defined” and “statistical”, respectively.

Lines 494-495: Why was this not quantified?
They are different by nature, hard to compare against one another. Two are nonlocal in time (jet lifetime and SOM persistence) and one is local (jet COM speed). The two jet persistence properties can have multiple values per timestep, while the SOM persistence is going to be interpreted differently depending on which cluster it relates to: long stay on cluster 8 is not the same as a long stay in cluster 10. This is too broad a topic to cover properly in this paper, and is most probably the subject of our next work. Still, we have fleshed out this comment more in the updated version of the manuscript.

Lines 513-514: Change reference to the final paper (https://doi.org/10.5194/wcd-5-1269-2024).
Done, thank you.

Updates in methods:
An update in the methods was required to accommodate for large datasets (ensemble model output) and global data, in preparation for future work. The SOMs were also modified to be made almost entirely reproducible by changing the initialization step, which is a problem that was raised in later internal reviews. We have run the analyses of this work again with the updated methods to have these benefits already showcased. Here, we provide a short list of changes in the methods between the initial manuscript and the submission of this response, prompted either by the two referees or by feedback from colleagues:

- Internally, the backend of the jet feature extraction algorithm has changed to a python package called “polars”. It allows for significant performance boosts while keeping the algorithm identical. The algorithm has now been properly tested on global data and provides good results there too.
- The jet width algorithm can now be run on every point along the jet instead of every fifth point, thanks to these performance gains. The effects of this change are minimal for most jets but helps for extreme cases (wavy jets or jets close to the edges of the domain).
- We also changed the definition of the width to better represent the width of the STJ. Before, the width on either side of the core was defined as *the distance between the core jet point and the first point along the normal segment whose interpolated speed is below one half of the core speed*. This proportion has been changed to three quarters. The equatorward side of the STJ was often too close to the southern edge of our domain, therefore the algorithm would often not capture its width well.
- The jet categorization now runs on each jet point separately. It assigns to each jet point a score between 0 (STJ point) and 1 (EDJ point), based on a Gaussian Mixture model run independently for every week of the year (used to be every season) using the 2 quantities potential temperature and jet vertical extent as discriminants, following Spensberger et al. 2023. For more details about this change, please refer to our response to referee 1’s main comment. Looking at, for example, figure 2 and 3 of this document, the results seem to have changed by a small but noticeable amount. This in turns has modified the seasonal
- The SOMs are now initialized using the 2 first principal components of the training data, instead of randomly. This was the only source of stochasticity in the algorithm, since we use batch training and not iterative training. Now the SOMs are fully reproducible. Our earlier implementation was too slow to allow for it, and

it was dramatically improved.
This has altered the SOM training results in a specific way. Roughly the same 24 SOM clusters are found, when compared to the previous initialization method, but their absolute position on the grid has moved. For example, the region of the SOM highlighted as containing all of the variability of the first week of June, around clusters 1, 8 and 13 in the original manuscript, still exist in the new SOM but simply moved to another region of the SOM and are now around clusters 10, 17 and 23.

- SOM persistence is now defined as an uninterrupted stay on a single node, also allowing for jumps from the starting node to another node so long as the starting node and the jump destination are similar enough. This similarity used to be quantified by SOM grid distance, with a threshold of 1. It is now quantified using Euclidean distance between cluster weight matrices, with the 10th percentile of pairwise cluster weight matrices as threshold. This avoids creating a hotspot of artificial persistence for the group of high-RMSE and high-separatedness early June clusters.

Changes to organisation of the paper:
The organisation of section 3, the results section, was altered.

Section 3.1, *Atlantic summer SOM space,* doesn't present either predictability or persistence properties in the new version of the manuscript. The predictability aspect was entirely scrapped from the manuscript as our proxy for it was too sensitive to the seasonal variations of cluster occurrences to be meaningful. The persistence aspect now has its own section, section 3.4, *Jet persistence*. This section presents and discusses figure 13, an updated version of the original figure 8, which compares state persistence and jet object persistence.

In section 3.1, the paragraph discussing similarities between weather regimes and SOM z500 composites, that were presented as figure A1 in the Appendix, was reworked and figure A1 and the original Appendix A deleted. In the new version of the manuscript, this discussion now refers to SOM cluster and weather regime co-occurrence probabilities, presented as new panels e-g of figure 6.

Appendix A was added to discuss the possibility of a hybrid jet from our jet categorization method, presenting this document's figure 2.

The original Appendix A was scrapped in favour of quantitative discussions of similarities to weather regimes in section 3.1 and figure 6 in particular.

Changes to figures and tables:

- Figure 3: Changed the wind speed threshold from half the core speed to three quarters to represent the current definition of jet width.
- Figure 4: Changed From JJA 2D histograms (a and b) with seasonal spatial (c, d, e and f) maps to monthly 2D histograms, which is this document's first figure
- Figure 5: Updated with the new SOM with reproducible initialization.
- Figure 6: Updated with new SOM. Removed daily NAO (old panel b). Trends moved from panel c to panel b. Added RMSE and Separatedness (new panels c and d). Added Association to summer weather regimes (panels e through i)
- Figure 7: Updated to new SOM, updated colorbar label with unit.
- Figure 8, moved in the paper and is now figure 13: Updated to new SOM. Removed trends (was panel c). Moved mean residence time from panel a to panel b, moved 95th quantile of residence times from panel b to panel c. Added count of long stays (above 4 days) as the new panel a. Added mean decay time as panel d. Added projections of jet persistence properties as panels e through h. Panel e is STJ lifetime, panel f is EDJ lifetime, panel g is STJ COM speed and panel h is EDJ COM speed.
- Figure 9: Deleted.
- Figure 10, is now figure 8: reproduced with the new jet categorization method. Swapped R16 for DC16 waviness, as it shows more unique characteristics distinct from jet tilt. Is now figure 8.
- Figure 11, is now figure 9: reproduced with the new jet categorization method. Swapped R16 for DC16 waviness.
- Figure 12: deleted in favour of the new figure 10.
- Table 1: deleted in favor of the new figure 10.
- Figure 13, is now figure 12: Updated to new SOM and new jet categorization method. Updated colormaps to be more legible. Swapped speed of COM for tilt, since the persistence metrics are discussed in details in the new figure 13.
- NEW figure 10: An updated version of the original manuscript's figure B1, presenting yearly trends for every day of the year.
- NEW figure 11: A composite of the detected jet cores for every time step belonging to a cluster, overlaid with the jet cores detected on the wind speed composites (i.e. figure 5) to highlight the close match between mean position of detected jets and detected jets on the mean wind speed field.
- Figure A1 was scrapped in favor of a more quantitative discussion of weather regime overlap in the main text.
- NEW figure A1 is this document's figure 2, presenting the seasonal cycles of the jets' properties when the threshold between EDJ and STJ is altered, and also when allowing for a hybrid jet. A discussion, similar to that found at the start of this document when answering to reviewer 1's main comment, can be found.

- Figures B1 and B2 were updated to the results of the new jet categorization, and swapped to better match the order of their subset counterparts, respectively figures 8 and 10 of the updated manuscript.
- NEW figure C1: a mirror of figure 12 with jet categories relative to jets found on the SOM composite wind fields.

Finally, the title was changed from
Trends and seasonal signals in Atlantic feature-based jet stream characteristics and in weather types
to
Seasonal to decadal variability and persistence properties of the Euro-Atlantic jet streams characterized by complementary approaches.

References in this document (alphabetical order)

By the referees:
1. Gibson, P. B., Perkins-Kirkpatrick, S. E., Uotila, P., Pepler, A. S. & Alexander, L. V. On the use of self-organizing maps for studying climate extremes. Journal of Geophysical Research: Atmospheres 122, 3891–3903 (2017).
2. Stryhal, J., Beranová, R. & Huth, R. Representation of Modes of Atmospheric Circulation Variability by Self-Organizing Maps: A Study Using Synthetic Data. Journal of Geophysical Research: Atmospheres 128, e2023JD039183 (2023).
3. Stryhal, J. & Plavcová, E. On using self-organizing maps and discretized Sammon maps to study links between atmospheric circulation and weather extremes. International Journal of Climatology 43, 2678–2698 (2023).

By the authors:

4. Athanasiadis, P. J., Wallace, J. M. & Wettstein, J. J. Patterns of Wintertime Jet Stream Variability and Their Relation to the Storm Tracks. (2010) doi:10.1175/2009JAS3270.1.
5. Barnes, E. A. & Hartmann, D. L. Detection of Rossby wave breaking and its response to shifts of the midlatitude jet with climate change. Journal of Geophysical Research: Atmospheres 117, (2012).
6. Cassou, C., Terray, L. & Phillips, A. S. Tropical Atlantic Influence on European Heat Waves. Journal of Climate 18, 2805–2811 (2005).
7. Grams, C. M., Beerli, R., Pfenninger, S., Staffell, I. & Wernli, H. Balancing Europe's wind-power output through spatial deployment informed by weather regimes. Nature Clim Change 7, 557–562 (2017).
8. Koch, P., Wernli, H. & Davies, H. C. An event-based jet-stream climatology and typology. International Journal of Climatology 26, 283–301 (2006).
9. Wirth, V. & Polster, C. The Problem of Diagnosing Jet Waveguidability in the Presence of Large-Amplitude Eddies. Journal of the Atmospheric Sciences 78, 3137–3151 (2021).