jsmetrics v0.2.0: a Python package for metrics and algorithms used to identify or characterise atmospheric jet-streams.

Response to Editor

Dear Editor,

We are pleased to see that Reviewer #2 felt that our revised manuscript was much improved and that the online documentation was now sufficient. They have raised a series of issues/comments, and we respond to each of them below.

We feel we have dealt with all of the additional minor changes suggested by Reviewer #2 in their latest review. There are some suggestions that are beyond the scope of the software package in its current form, and having established a dialogue with Dr Gloria Manney, we hope to implement some of these suggestions in subsequent releases.

Response to Reviewer #2

We once again thank Dr. Gloria Manney for the help and advice she has provided in curating this manuscript. We are extreme grateful for all the time and effort that she has invested in understanding software package and helping us refine the text.

Please find our response to your comments below.

Recommendation:

Reviewer: *The revised manuscript and code / documentation are all greatly improved and should be suitable for publication in GMD pending some further clarifications in the text.*
Author Response: Thank you, and we hope that we have now made these further clarifications in the text.

General Comments:

Reviewer: *The online documentation of the jsmetrics software has been vastly improved; it is now straightforward to run any of the metrics, and there is a helpful set of examples for doing so. The manuscript is also much improved, and I believe this work is now sufficiently mature for publication in GMD. I do have quite a number of comments on issues or language that I feel still needs some clarification, but while these may be somewhat extensive, they are all IMO in the nature of "minor" revisions. In a few cases that will be noted below I have already discussed appropriate modifications with the lead author; these concerns are included here to keep the online record of the review / revision process complete.*

In general, I still think that the distinction drawn between "jet statistics" and "jet core algorithms" is too strong, since there can be (depending on the algorithm and the application of it) a large overlap in the information they provide. Nevertheless, as long this overlap is acknowledged and the

capabilities of and the primary outputs of each group of algorithms are clearly described, this choice does not materially impact use of the package or interpretation of the results. Several of my comments below focus on further clarification of this issue.
Author Response: We have made changes to the manuscript based on your comments about this issue (detailed throughout this response), and have written new descriptions for jet statistics and jet core algorithms, that we hope will clarify the distinction of these two categories. The new descriptions are:

"Jet statistics — Statistics for isolating individual quantities synonymous with the jet stream from upper-level wind speed within a given time window" (lines 26-27).

"Jet core algorithms — Methods that return a mask of coordinates related to the jet location, e.g., identifying the maximum wind speed throughout the horizontal and/or vertical plane within a given time window" (lines 30-31).

We do accept that methods from these two categories can offer the same kind of information (e.g. you could extract the latitude of jet cores provide by the jet core algorithms), but we hope to have clarified that different approaches are required to process and analyse the outputs of the jet statistics and jet core algorithms from jsmetrics. It felt appropriate for this work to distinguish categories based on types of outputs, and in this process, also describe what the software does and can be used for. In my own thesis (TK), I will treat the distinction between different jet stream metrics quite differently, because I will focusing on the scientific results provided by a particular use case of jsmetrics. We hope that this latest distinction between the categories will be clear for the reader.

Reviewer: *In the same vein, I still question the choice not to include jet core windspeed in the outputs for jet core algorithms (such as M11) that identify the jet core locations using windspeed (since that information is already available as it is essential to using the algorithm). In this case, while this would not have been my choice, the authors now include examples showing how to get this from the "mask" of jet cores – so, as above, this does seriously impact use of the package.*
Author Response: We recognise your concern here, but we would like to clarify that this decision was made for the software's efficiency and speed. We wanted to avoid including jet core wind speeds as well as a jet core mask by default, because it was not proven to scale well and we found it to be sub optimal for storage and computing resources. Including only a mask instead of the derived outputs of a mask (like jet core speed, latitude, altitude) means that we can save memory. We hope that offloading some of the procedures to extract other variables from the mask to the online examples will cover the broader uses of the package.

A Couple of Comments Re the Author Responses:
Reviewer: *Response to my general point (2), regarding the same issue mentioned just above: Part of my concern is that the original discussion made it sound like the "jet statistics" algorithms provided more information than the jet core algorithms, when in fact the opposite is typically the case. I think the revised discussion does improve this, but there are a couple of places where this could be clarified further, noted below; in particular, see my comment on Fig. 9 and the discussion thereof.*
Author Response: We hope that the new distinction of the two type of methods, mentioned above, is clear. As we mention later in this review, the discussion of the method used to create Figure 9 has also edited for clarification (lines 396-402).

Reviewer: *Regarding the authors' response about M11 implementation and the jupyter notebook detailing that (which was indeed very helpful; it would be excellent to be able to see something like this for other algorithms, though I appreciate that that may be too much work in cases where you haven't done something similar to this already), to paraphrase my exchange with the lead author (denoted as TK) on this implementation:*
In[23]: I noted that JETPAC {the formal acronym for the package described by M11} has an undocumented feature in that, for each longitude slice without windspeeds > core_threshold (currently 40m/s), if there are regions with windspeed > edge_threshold, it catalogs those regions and the location / value of the max windspeed within them; if there are no windspeeds > edge_threshold, it catalogs a single maximum windspeed location (location, windspeed, other characteristics). It appears that if you wanted to allow that feature (at least the core > max > edge option) in jsmetrics, you have that information here before you do the down selecting to remove the regions with no cores.

TK responded that he would look into this.
Author Response: This feature of M11 is now included in the latest version of jsmetrics, and if possible, we would be very keen to continue our dialogue with you to verify M11.

Reviewer: *In[26]: I questioned why even do this {downsample contour found for edge of jet region to get only the points above, below, equatorward, poleward of the jet core} when what you've got (the full region mask) provides more information than the original (above/below, poleward/equatorward)?! … I believe (if I've followed everything correctly) that you do have the option to retain and save this full mask, is that right?*

TK responded that the full mask is, indeed, included in the outputs.
Author Response: Yes, having this direct dialogue was really helpful.

Reviewer: *In[36]: It was not obvious to me in trying to go through this that the largest of the local maxima is always the one selected when one is / some {that are in the same "jet region"} are eliminated?*

TK answered that the function currently did not do this, but that he would correct it to do so.
Author Response: The function has now been corrected, so this is now included in the latest version of jsmetrics.

Reviewer: *Regarding the "alternative with diagonals checked for jet cores", I had already realized that not checking the diagonals was something that may pose a big inefficiency in the original JETPAC implementation – though it does work as intended in the end, I guessed that there is a lot of unnecessary checking of multiple local maxima because of that (which could impact the speed).*

TK's response indicates that the option to check the diagonals is / will be included in jsmetrics.
Author Response: This diagonal checking is now included in the latest version of the M11 implementation in jsmetrics. This makes a small change to the jet cores shown by M11 in Figure 3.

Specific Comments On the Manuscript (in order of appearance in the manuscript, not importance; line numbers are from tracked changes version):

Reviewer: *Lines 12–13 (Abstract): Suggest rewording, something like "We classify the methods for characterising jet streams in the literature into three broad strategies: statistics...", since it really is a choice you made in how to group them in the package as opposed to a "proposal" for how*

they should be thought of in general.
Author Response: We have reworded this part of the abstract, and so it now reads:
"We classify the methods for characterising jet streams in the literature into three broad strategies: statistics that isolate individual values from the wind speed profile (jet statistics), methods for quantifying the sinuosity of the upper air (waviness metrics), and algorithms that identify a mask related to the coordinates of fast flowing wind throughout the horizontal and/or vertical plane (jet core algorithms)" (lines 9-12).

Reviewer: *Lines 28–29: Suggest "We divide these common approaches into three broad types:"*
Author response: Thanks for the suggestion, we have updated the text (line 25).

Reviewer: *Lines 30–31: This sounds like each of these algorithms returns one and only one of latitude, speed, or width. I think some of them return more than one, right? So this wording should be modified.*
Author response: We agree, and so we have changed the description of jet statistics to:
"Statistics for isolating individual quantities synonymous with the jet stream from upper-level wind speed within a given time window (e.g. latitude, speed, width)" (lines 26-27).
We hope the change of wording from *single* to *individual* is more applicable and better describes this category i.e., you can have a collection of individual quantities (latitude, speed, width, etc.).

Reviewer: *Lines 32–33: The wording makes it sound rather like sinuosity is the only measure of waviness, when in fact there are many – again, greater care with the wording would be helpful.*
Author response: Yes, we agree and have removed 'sinuosity' from the sentence here to reduce confusion. The new description for waviness metrics is:
"Statistics and algorithms for determining the `waviness' of upper-level mean flow within a given time window. These metrics only have meaning at an integrated global scale" (lines 28-29).

Reviewer: *Lines 34–35: May want to say something about identifying the maximum windspeed and/or the region around that maximum, since that is the definition of "jet core".*
Author response: We have refined our description of jet core algorithms in lines 30-31 to better account for all aspects of this category. This change should also account for your later comment about K14 (in reference to *lines 365–369*) to distinguish that method as a jet core algorithm which returns a mask of coordinates.

Reviewer: *Lines 40–41: These references don't really seem like the best choices here, since, while they are mainly review papers, they generally cite few (most of them none) of the results from papers using metrics implemented in jsmetrics – that is, they don't demonstrate that the metrics you are implementing provide conflicting or confusing information.*
Author response: We agree and have removed references from this sentence and reworded it to:
"The differences between these types of approaches could lead to confusion about the trends shown in the planet's jet streams across a range of modelling and observational studies" (line 32-33).

Reviewer: *Lines 61–64: This is a bit of a moving target, but this paper:*
Spensberger et al 2023, DOI: 10.1175/JCLI-D-23-0080.1, published (early online release) in J. Clim. since the original version of the jsmetrics manuscript is a very good choice that could be added for discussion of thermally / eddy driven jets; it also demonstrates a new way of identifying them using potential temperature (something that might easily be implemented in jsmetrics in the future). There are a few other places below where I also suggest citing it.
Author response: Thank you for suggesting this reference, we agree it is a very good choice as a citation for this manuscript and have now included it on lines 56, 60, 179-180, and 295-296.

Reviewer: *Lines 63–64: Suggest something like "...tropospheric jet streams but diagnostics included may identify either or both of the "primary" types...." to make it clear that one or both of eddy or thermally driven jets may be identified – i.e., that this statement simply says you are not identifying stratospheric jets.*
Author response: Thank you, we have reworded this sentence to.
"Tropospheric jet streams in observations often exist in "merged states", especially across the mid-latitudes (Stendel et al., 2021), but diagnostics included in this package are not yet able to disaggregate the two "primary" types of jets" (lines 56-58).

Reviewer: *Line 71: I don't think "synonymous" is what you mean here (that would say that these jets are cold waves, heat waves, etc). Perhaps something like "..directly involved in {development | evolution} of…"*
Author response: Thank you. This has been changed to: *"directly involved in the development of cold waves…"* (line 65).

Reviewer: *Line 75–78: This seems out of place here. Suggest joining this with the paragraph at the end of the previous section, then starting this section with something like "Despite their importance to climate studies, features of…"*
Author response: Okay, we have moved this content to the end of the introduction (lines 67-69), and the start of section 2 begins with *"Despite their importance…"*, like you suggest.

Reviewer: *Line 83: Should be "specific questions" and later in this line "and / or" since it is usually not just one characteristic if indeed they are developed for such a specific purpose. Which M11 (aka JETPAC) definitely was not, and I expect that is the case for some of the others as well (especially jet core algorithms that provide a wealth of information). JETPAC was developed (as previewed in M11) to be useful for many purposes; in addition to the papers (those cited here, along with another in which JETPAC diagnostics are correlated with Asian summer monsoon anticyclone characteristics, Manney et al, J Clim, 2021(b), DOI: 10.1175/JCLI-D-20-0729.1; and another in relation to the tropopause inversion layer, Peevey et al., 2014, JGR,doi:10.1002/2014JD021808) that use it to study climatology, variability, and trends in the jet streams and related phenomena, it is also being used and has been used in studies of transport and STE, and for analysing UTLS composition variability / trends (e.g., Olsen et al, 2019, JGR, https://doi.org/10.1029/2019JD030435; Millán et al, 2023, AMT, https://doi.org/10.5194/amt-16-2957-2023). My point is not that you should cite all these papers, but that some tropospheric jet diagnostics have been developed for very broad purposes.*
Author response: Thank you for pointing this out, we have corrected the text to "specific questions", and "and/or" (line 75).

Reviewer: *Line 96: Other methods have "evaluate latitudinal shifts, slowing or speeding up of the jet" as one of their primary purposes, e.g., M11 and PO13.*
Author response: We recognise this and agree, but we felt as though, for the remit of describing this software, it would be appropriate to describe what jet statistics are most useful for in this section and in isolation to the jet core algorithms uses.

Reviewer: *Line 124: Spensberger & Spengler's method is not currently implemented, right? And it doesn't IMO fit the category of "jet statistics" (see usage in Spensberger et al., 2023, mentioned above).*
Author response: It is not implemented, but we had made reference to this method in Table 5 under jet statistics originally. We agree with your opinion here and have place this method to jet core

algorithm in Table 5. We have also removed Spensberger & Spengler, 2020 reference from the text here (line 111).

Reviewer: *Line 136 and Table 2: A brief description of what each of these metrics actually does would be helpful (i.e., what is calculated from what to get the metric).*
Author response: Thank you for your suggestion here. We have now updated the text to include a brief overview of both of the waviness metrics included in jsmetrics (lines 123-126).

Reviewer: *Lines 169–170: Seems odd to use Manney et al (2014) in these two reference lists when the discussion is about the algorithms described in Manney et al (2011) and the other references given are all the "methods" papers for that technique.*
Author response: Yes, we agree and have replaced Manney et al. (2014) with Manney et al. (2011) (lines 157-158).

Reviewer: *Line 187: Suggest "a method based on latitude to distinguish" – that will help clarify why is doesn't work very well most of the time.*
Author response: Done (line 176).

Reviewer: *Line 195: Would be good to mention Spensberger et al., 2023 (see citation info above), as well as the method introduced by Christenson et al (2017, J Clim, DOI: 10.1175/JCLI-D-16-0565.1), which both (in different implementations) use jet core potential temperature to distinguish eddy from thermally driven jets.*
Author response: We have included a reference to the method in both of these papers and the updated text is:
"Although not currently implemented in jsmetrics, Christenson et al. (2017) and Spensberger et al. (2023) propose methods which use the potential temperature of jet cores to distinguish eddy from thermally driven jets." (lines 179-180).

Reviewer: *Lines 214–215: This is as good a place as any to mention that allowing different vertical coordinates (e.g., altitude, potential temperature, and, especially, model levels) should be a high priority for future jsmetrics development. E.g., JETPAC is typically run on reanalysis model levels because of the inadequacy of the "standard" pressure levels to capture the vertical structure (e.g., Manney et al, 2017, ACP); but is also sometimes run on pressure, altitude, or potential temperature levels when being used with other datasets to which those coordinates are native.*
Author response: We completely agree with your suggestion here, and have updated the text to include the following sentence:
"Whilst the current iteration of jsmetrics is only compatible with data with standard pressure levels (plev), for future development of the package, it is a priority to include compatibility with other vertical coordinate systems." (lines 199-200).

Reviewer: *Line 291: "capable of" is not the right wording here, since most of the jet core algorithms are "capable of" this and you are excluding all of them as well. You could just say something about showing metrics that look at lower tropospheric u-wind and leave it at that – that is sufficient to explain which methods you show here.*
Author response: We have removed "capable of" from this sentence (lines 278-279).

Reviewer: *Figure 2: As I noted in my original review, you need to define in the caption what the width (top to bottom), length (side to side) of the shaded parts represent, what the thick lines near thecentre represent, and what the length of the thin lines means. Not everyone is familiar with a "violin plot" and the reader shouldn't have to stop and go look it up!*

Author response: We have updated the figure caption and in text description to describe what the violin plot show:
"The thicker black line in the centre of each violin plot indicates the interquartile range, and the thinner line indicates the 95% confidence interval. The white dot represents the median and the shading which forms the body of each violin is a Kernel Density Estimation, with wider sections representing a higher probability of occurence". (lines 274-277)

Reviewer: *Line 297: What feature in Figure 2 shows the "Interquartile Range"? Also put this in the Figure 2 caption.*
Author response: As above, we have changed the in-text description and figure caption to describe which feature of the violin shows the Interquartile range.

Reviewer: *Line 323: Add Spensberger et al (2023, citation above) to this reference list, it has very good discussion of this.*
Author response: Done (line 296-297).

Reviewer: *Lines 335–336: Several papers have shown (unlike the European and Asian CAOs during that winter) that the stratosphere / SSW didn't have a very big impact on this event, see, e.g., Davis et al, 2022, https://doi.org/10.1038/s41467-022-28836-1; Zhang et al 2022, https://doi.org/10.1029/2021GL096840; Bolinger et al 2022, https://doi.org/10.1016/j.wace.2022.100461; as well as others.*
Author response: Thanks for pointing this out, we were wrong with our assumption that a SSW had affected this event, as such we have removed the reference to a SSW in this part (lines 304-305).

Reviewer: *Lines 351–352: "only selecting cells of local 'maxima' " is not really "stricter conditions" if they are using the same wind speed threshold for a "core" to exist – it is just providing information only on the core itself rather than on both that and the surrounding "jet region".*
Author response: We agree, and the text has been edited to:
"Notably, S09, M11, P013, and K14 all use a 30 ms threshold (but not in the same way) and both S09 and PO13 select only cells of local 'maxima'; M11 and K14 also extract regions around each core/maxima" (lines 317-319).

Reviewer: *Line 365–369: Identifying "jet centers" in latitude and longitude makes "jet centers" a very different beast than "jet cores", which are defined as a maximum in a horizontal coordinate / vertical coordinate plane (the core implementation in M11, for example, doesn't care what those horizontal and vertical coordinates are, just that it is given one of each). As such, K14 really does not fit the "jet core algorithm" category. Nor the "jet statistics" category. Not suggesting that you change it, but that you clarify better the fundamental difference of this method.*
Author response: We hope that our new definition of jet core algorithms and jet statistics (lines 30-31 & 26-27) clarifies the distinction between jet core algorithms and other types. When considering that all the jet core algorithms do are isolate coordinates throughout a given plane, then this makes sense that these coordinates can be latitude-longitude as well as latitude-vertical.

Reviewer: *Line 399–400: This is consistent with very different jet behaviour in different broad latitude regions, and the fact that both the North Atlantic and North Pacific have complex / highly variable jet patterns, whereas the region over Europe / Asia / W Pacific has a strong persistent subtropical jet (e.g., Koch et al., 2006; Manney et al, 2014; Spensberger & Spengler, 2020).*
Author response: Yes, we agree, this is a sensible interpretation. We edited the sentence to reflect that this a real phenomena rather than estimation issue from the metrics:

"In Figure 5, the mean jet position varies more in the Northern Hemisphere (33.22-49.75◦N) than in the North Atlantic (44.62-49.55◦N), North Pacific (31.57-46.81◦N), or Southern Hemisphere (42.37-50.59◦S)" (lines 360-362).

Reviewer: *Lines 409–410: This result may or may not be realistic, because of the very coarse vertical resolution of the date used in the example and the different ways each algorithm interacts with that resolution – there is a huge amount of real atmospheric variability between regions, so "more consistent" isn't necessarily expected or realistic.*
Author response: We have simplified this sentence to avoid confusion about 'consistency'. New sentence reads:
"In particular, Figure 5 shows that some metrics show more variation in their estimates across multiple regions than others" (line 370)

Reviewer: *Line 431: One person's "unimportant feature" could possibly be another person's primary research question! Define "unimportant feature" (and / or choose a more precise wording).*
Author response: We have chosen to reword this sentence and use the term 'noise'. The new sentence is: *"The comparison (Figure 8) demonstrates the losses and gains of time averaging: some features are diluted using the mean, while counts show more detail but can also include more noise"* (lines 387-389)

Reviewer: *Lines 434–449: As I said in my original review, M11 and PO13 (as well as some of the other "jet core" algorithms) are "purpose-built" to extract jet latitude – jet core latitude and altitude (or other "height" coordinate), along with windspeed, are the first and foremost outputs of these methods, and the ones that have been used most in following work with these methods. Further, it is disingenuous to say you have not implemented finding the latitude in these methods, since you have a point flagged as the jet core (and have in fact used this to plot those latitudes in Figure 3!), thus all you have to do is to find the index into the latitude coordinate from that core "mask" and extract the latitude – perhaps one or two lines of code (not at all different in principle from extracting the core wind speed, which you have done in (one of the) examples in the online documentation). Further, the method you use to get the latitudes for these (far more complex than what I just suggested) will inherently reduce any appearance of bi-modality since it doesn't allow multiple jets at a given longitude and involves more averaging than the simpler procedure I suggested – thus I don't think you can say anything about bi-modality given the method you have used to get the latitude from the jet core algorithms (as long as they actually identify a core location, I don't think using a wind speed threshold should have much to do with it; though looking at upper troposphere vs lower troposphere could definitely be a factor).*
Author response: We have updated the text to describe more specifically how we are creating an estimate for jet latitude for all the jet core algorithms, exclusively for the remit of this case study. We also have added a note that future version of jsmetrics could contain procedures that translate the outputs of jet core algorithms to jet statistics (lines 402-403).

The updated text on lines 396-403 is:
"To create an estimate for jet latitude from the jet core algorithms, we first compute the estimation of jet cores using a given algorithm and use these locations as a mask to extract wind speed values for each day. Using these values, we then extract the zonally-averaged maximum wind speed and define the associated latitude as the jet latitude value at the native resolution. For consistency's sake, we use a single method to extract the latitude from the multidimensional field returned by the algorithms in this case study. This is the latitude of the maximum wind in the region (despite other options being available to do this for the multidimensional fields, e.g. Manney et al. (2011) would select all the indexes of returned jet cores). Future versions of jsmetrics could contain a variety of procedures that process the outputs of jet core algorithms into jet statistics"

Also in these lines: In line 447, should be more specific and say that by "altitude of the methodology" you mean the jet core algorithms are looking at the upper troposphere and the jet statistics algorithms are looking at the lower troposphere – meaning a totally different definition of jets and looking at wind speeds in a totally different region, which clearly affects all of their characteristics.
Author response: Text changed, now reads:
"This is most likely due to the altitude of the methodology, as the jet core algorithms are looking at the upper troposphere and the jet statistics algorithms are looking at the lower troposphere (Tables 1 & 3)". (lines 405-407).

Reviewer: *Line 465: You might want to note here that Manney et al (2017, 2021a, b), Manney & Hegglin (2018), and PO13 all included extensive comparisons of different input reanalysis (which I assume is what you mean by "observational" since we don't have nice 3d gridded fields of actual observed winds!) datasets.*
Author response: We have changed "observational dataset" to "input reanalysis datasets" and included a note of the papers you mention here (lines 425-426).

Minor / Technical points (typos, grammar, wording, etc):

Reviewer: *Line 39: Suggest instead of "a confusing" either "an unclear message" or "apparently conflicting messages"*
Author response: We have reworded this sentence to *"The differences between these types of approaches could lead to confusion about the trends shown in the planet's jet streams across a range of modelling and observational studies."* (lines 32-33)

Reviewer: *Line 45: "from" should be "using" or "upon"*
Author response: done (line 38)

Reviewer: *Line 59: Either "positions…are" or "position…is"*
Author response: done, we went for 'position...is' (lines 51-52)

Reviewer: *Line 79: Suggest "identify and characterise" instead of "detect and then characterise"*
Author response: done (line 71).

Reviewer: *Line 81: Add "e.g.," before references, since there are lots of others that discuss aspects of this.*
Author response: done (line 74)

Reviewer: *Lines 84–85: Suggest "This initial set of metrics was included based on, first, ... and, second, the frequency of their usage in the literature."*
Author response: done (lines 76-78).

Reviewer: *Line 86: "of" should be "in".*
Author response: done (line 78).

Reviewer: *Line 90: "which" should be "that", and, to be consistent with the "single value" in line 88, "and" would need to be "or" (see comment above re the description of "jet statistics").*
Author response: we have replaced this with "and/or" (line 75).

Reviewer: *Line 107: Delete "any"*

Author response: done (removed from line 98).

Reviewer: *Line 110: Again, " 'latitude' and 'speed' " is not a "single" metric!*
Author response: We have changed to individual to be consistent with other references to jet statistics (line 100).

Reviewer: *Line 113: "While each jet statistic"*
Author response: done (line 103).

Reviewer: *Line 193: Replace "which introduces a physical-based" with "by introducing a physically-based".*
Author response: done (line 178).

Reviewer: *Line 210: Suggest "...due in part to…" (since this is by no means the only advantage of using xarray!!)*
Author response: Thank you, we have updated the text (lines 196-197).

Reviewer: *Line 224: Replace "nor" with "and / or" and delete comma.*
Author response: done (line 212).

Reviewer: *Line 225: This is a less-than-obvious case, but needs to be either "methods' docstrings" or "method's docstring".*
Author response: done, we went for 'method's docstring' (line 213).

Reviewer: *Line 228: Add comma after "i.e.,"*
Author response: done (line 216), and also changed on line 218.

Reviewer: *Line 235: Delete "of" at end of line.*
Author response: changed this to '...made a note of…' (line 223).

Reviewer: *Line 257: "figure" should be "Figure".*
Author response: done (line 245).

Reviewer: *Line 259: "rework and refactor" seems a bit redundant (since refactoring is a kind of reworking).*
Author response: changed to just 'refactor' (line 247).

Reviewer: *Lines 263–264: Now you are using "refactor" synonymously with "debug" (or "troubleshoot" or whatever term you prefer to use), whereas "refactor" is defined as "to improve internal code by making many small changes without altering the code's external behavior", which clearly implies that the code already produces the desired result. Also, the sentence structure has errors here, disregarding any content changes, it should be: "After which we either refactor the method further if it fails the validation, or write unit tests, finish the documentation, and integrate the metric into the jsmetrics package if it succeeds."*
Author response: We agree and have changed 'refactor' to 'debug', as this is closer to our meaning. Also, we have made the changes you suggest and the sentence now reads:
"After which we either debug the method further if it fails the validation, or write unit tests, finish the documentation, and integrate the metric into the jsmetrics package if it succeeds" (lines 251-252).

Reviewer: *Line 279: Should be "...from the ERA5…"*
Author response: Thanks, we have updated the text (line 266).

Reviewer: *Line 286: "is" should be "are" ("data" is plural); also, "details" (which, grammatically, should be "detail" since "data" is plural) isn't the right word here – perhaps use "comprise" or "consist of".*
Author response: Thank you, we have changed the text and we use 'consist of' (line 272).

Reviewer: *Line 326: Change "We hope to express that" to something like "The above example demonstrates that…" or "We hope the above example demonstrates that…" (I personally would leave out "We hope" since if you are publishing it you should express confidence in your results.)*
Author response: Thanks for your suggestion, we have updated the text to "The above example demonstrates that…" (line 299).

Reviewer: *Line 329: There should not be a comma.*
Author response: We have removed the comma after 'globe…' (line 301).

Reviewer: *Lines 333–334, and succeeding use: You use "North American Cold Wave" and other times "Texas Cold Wave". From what I've seen this event has been most frequently called the "Great Plains Cold Air Outbreak". Whichever term you choose, pick one and only one. (I also see no reason why it needs to be italicised.)*
Author response: Thanks for the suggestion, we have removed the italics and changed all mentions of this event to North American Cold Wave in the text.

Reviewer: *Line 334: Any of "...between 6 and 21 February…", "...from 6 through 21 February…", or "...from the 6th through the 21st of February…" would be correct (I favor the first as being most concise; whichever you use, try to be consistent in succeeding date range references).*
Author response: We have now chosen a consistent date range reference and changed all references to the same style. The style is like: *between 1200 UTC on 6th February and 1200 UTC on 21st February*. We will discuss further, as required, with the copy editor so it conforms to Copernicus style.

Reviewer: *Line 338: Delete colon after "levels".*
Author response: done (line 309).

Reviewer: *Line 341: Suggest "...some of the methods…"*
Author response: done (line 312).

Reviewer: *Line 343: "...of the figures…"*
Author response: done (lines 313-314).

Reviewer: *Line 353: "P13" should be "PO13".*
Author response: done (line 318).

Reviewer: *Line 360: Reword / correct: "...those cores; otherwise these jet cores in the same region will be considered part of the same core, at the location of the largest of the local wind speed maxima."*
Author response: This has been corrected (line 325).

Reviewer: *Line 364: "...cores in each may…"*
Author response: done (line 327).

Reviewer: *Figure 4. Say in the caption what exactly the "Standard North Pacific Region" is.*

Author response: Thank you for the suggestion, this has now been included.

Reviewer: *Lines 373, 375, 376: See comment re lines 333–334 just above.*
Author response: Thanks for the suggestion, we have removed the italics and changed all mentions of this event to North American Cold Wave in the text.

Reviewer: *Line 397: "figure" should be "figures"*
Author response: done (line 360).

Reviewer: *Line 403: Should be "...strongest and most variable…"*
Author response: changed (line 365).

Reviewer: *Line 418: Change "which all centre" to "all centred on"*
Author response: done (line 378).

Reviewer: *Line 425: Delete "to" and delete comma after "that".*
Author response: done (line 384).

Reviewer: *Line 436: "which" should be "that".*
Author response: done (line 395).

Reviewer: *Line 462: Need a comma after "metric".*
Author response: done (line 422).

Reviewer: *Line 482: Should be "...inputs, i.e., with…"*
Author response: done (line 442).

Reviewer: *Lines 490–494: The sentence structure here has problems, and the sentence is too long and complex to follow clearly. Suggest breaking it up into two or more sentences and restructuring.*
Author response: Thanks for the suggestion, we have broken this down into three sentences:

"Finally, we note that some metrics may be too complex for the remit of this package (e.g. Kern et al., 2018; Kern and Westermann, 2019; Bösiger et al., 2022). When developing the package, we avoided metrics that use variables describing different aspects of the upper-level flow synonymous with (characteristics of) jet streams, such as wind shear (e.g. Lee et al., 2019) and magnitude of atmospheric waves (e.g. Chemke and Ming, 2020). Similarly, we did not include any potential metrics that require a training element to run and those that are currently very computationally expensive(e.g. Limbach et al., 2012; Molnos et al., 2017)" (lines 450-455).

Reviewer: *Line 506: "which" should be "that".*
Author response: done (line 467).

Reviewer: *Line 511: "data is" should be "data are".*
Author response: done (line 472).

Reviewer: *(Note that I also found a number of typos and small errors similar to these in the online documentation, so would suggest more careful proofreading of that.)*
Author response: Thank you, we have also found a few and correct them since your review. We will also try to continually update the online docs with each iteration of the software.