Re-evaluation of “jsmetrics v0.2.0: a Python package for metrics and algorithms used to identify or characterise atmospheric jet-streams”, by Keel, et al.
Reviewed by Gloria Manney

Recommendation: 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.

General Comments: 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.

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.

A Couple of Comments Re the Author Responses:

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.

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.

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.

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.

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.

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

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.

Lines 28–29: Suggest "We divide these common approaches into three broad types:"

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.

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.

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".

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.

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.

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.

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…"

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…"

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.

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.

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).

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).

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.

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.

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.

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.

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.

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 the

centre 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!

Line 297: What feature in Figure 2 shows the "Interquartile Range"? Also put this in the Figure 2 caption.

Line 323: Add Spensberger et al (2023, citation above) to this reference list, it has very good discussion of this.

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.

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".

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.

Line 399–400: This is consistent with very different jet behavior in different broad latitude regions, and the fact that both the North Atlantic and North Pacific have complex / highy 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).

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.

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).

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).
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.

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.

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

Line 39: Suggest instead of "a confusing" either "an unclear message" or "apparently conflicting messages"

Line 45: "from" should be "using" or "upon"

Line 59: Either "positions…are" or "position…is"

Line 79: Suggest "identify and characterise" insead of "detect and then characterise"

Line 81: Add "e.g.," before references, since there are lots of others that discuss aspects of this.

Lines 84–85: Suggest "This initial set of metrics was included based on, first, ... and, second, the frequency of their usage in the literature."

Line 86: "of" should be "in".

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").

Line 107: Delete “any”

Line 110: Again, “ ‘latitude’ and ‘speed’ “ is not a “single” metric!

Line 113: “While each jet statistic”

Line 193: Replace “which introduces a physical-based” with “by introducing a physically-based”.

Line 210: Suggest “...due in part to…” (since this is by no means the only advantage of using xarray!!)

Line 224: Replace “nor” with “and / or” and delete comma.

Line 225: This is a less-than-obvious case, but needs to be either “methods’ docstrings” or “method’s docstring”.

Line 228: Add comma after “i.e.,”

Line 235: Delete “of” at end of line.

Line 257: “figure” should be “Figure”.

Line 259: “rework and refactor” seems a bit redundant (since refactoring is a kind of reworking).

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.”

Line 279: Should be “...from the ERA5…”

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”.

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.)

Line 329: There should not be a comma.

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.)

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).

Line 338: Delete colon after "levels".

LIne 341: Suggest "...some of the methods…"

Line 343: "...of the figures…"

Line 353: "P13" should be "PO13".

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."

Line 364: "...cores in each may…"

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

Lines 373, 375, 376: See comment re lines 333–334 just above.

Line 397: "figure" should be "figures"

Line 403: Should be "...strongest and most variable…"

Line 418: Change "which all centre" to "all centred on"

Line 425: Delete "to" and delete comma after "that".

Line 436: "which" should be "that".

Line 462: Need a comma after "metric".

Line 482: Should be "...inputs, i.e., with…"

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.

Line 506: “which” should be “that”.

Line 511: “data is” should be “data are”.

(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.)