

General comments:

Comment: It's nice to see Pikovnik et al.'s work comparing indices of Hadley cell strength. As a member of recent working groups on Hadley cell width, I think a paper like this is long overdue. I appreciate Pikovnik's diligence in comparing indices as a function of level and latitude and noting the impact of changing altitude or latitude on the trends from indices where the maximum value (i.e. of the stream function) is picked wherever it occurs. (An example of a paper that uses this metric is below.)

Response: We thank the Reviewer for their very detailed and constructive comments. We hope that all their comments and concerns are adequately addressed in our responses and that the text modifications in the revised manuscript are appropriate. Following the suggestion of Reviewer 1, we have added another metric of the HC strength based on the water vapour transport (Sohn and Park, 2013). Figures 2-5 were made more concise by showing stream-function-based metrics 100 hPa apart. Furthermore, to reduce redundancy, Figures 2 and 4 now show trends of seasonal-mean HC instead of monthly-mean HC, following the argument presented in Waugh et al. (2018). In order to follow the terminology used by other authors (e.g. Solomon et. al, 2016; David and Birner, 2018; Waugh et al., 2018), we have changed "indices" and "measures" into "metrics" throughout the text, as well as in the title.

References:

Davis, N., & Birner, T. (2017). On the Discrepancies in Tropical Belt Expansion between Reanalyses and Climate Models and among Tropical Belt Width Metrics, *Journal of Climate*, 30(4), 1211-1231.

Solomon, A., Polvani, L. M., Waugh, D. W., and Davis, S. M. (2016), Contrasting upper and lower atmospheric metrics of tropical expansion in the Southern Hemisphere, *Geophys. Res. Lett.*, 43, 10,496– 10,503, doi:10.1002/2016GL070917.

Sohn, B. J., and Park, S.-C. (2010), Strengthened tropical circulations in past three decades inferred from water vapor transport, *J. Geophys. Res.*, 115, D15112, doi:10.1029/2009JD013713.

Waugh, D. W., Grise, K. M., Seviour, W. J. M., Davis, S. M., Davis, N., Adam, O., Son, S.-W., Simpson, I. R., Staten, P. W., Maycock, A. C., Ummenhofer, C. C., Birner, T., & Ming, A. (2018). Revisiting the Relationship among Metrics of Tropical Expansion, *Journal of Climate*, 31(18), 7565-7581.

Comment: I am interested in the normal mode-based metric. I found the description in the paper inadequate; wording like "unbalanced flow" is foreign to those of us who haven't read up on the normal mode decomposition. In fact, the relationship between normal modes and the Hadley cell is not intuitively obvious; when I first read that there was a normal mode-based

metric, I thought, “this must be a usage of normal mode besides what I’m thinking.” But when I read a bit of Žagar & J. Tribbia (2020), I saw that the usage of normal modes was precisely what I was used to, and “unbalanced flow” describes flow that is not in thermal wind balance. Not surprisingly, the overturning one can deduce from the unbalanced flow – see Figures 12 and 14 of Žagar and Tribbia (2020) – looks very much like the overturning one can deduce from the divergent flow – see Figure 2 of Staten et al. (2019).

Response: We agree with the Reviewer that the description of the new metric derived from the normal-mode function decomposition was not given enough space and was not explained in sufficient detail. This is because the new metric has yet to be fully explored and tuned, and the paper focuses on discussing already established indices. While we hope to carry out more evaluation of the new metric in the future, we expanded its definition in Section 2 (lines 110-130) and in the result sections where needed. In particular, we refer to Žagar et al. (2017), who discuss the spectrum of the global flows projecting on the inertia-gravity eigensolutions of linearized primitive equations. Their figure 15 is enclosed here as Fig. 1. They distinguish three regimes projecting on non-Rossby modes. The large-scale regime (zonal wavenumber $k < 6$), where the Hadley cell and Walker circulation project, is denoted “unbalanced circulation”. Note however that Fig. 1 does not include the zonal-mean state, as $k=0$ is commonly not a part of the log-log energy spectrum. But its energy would be a continuation of the power-law of -1 associated with the large-scale flow.

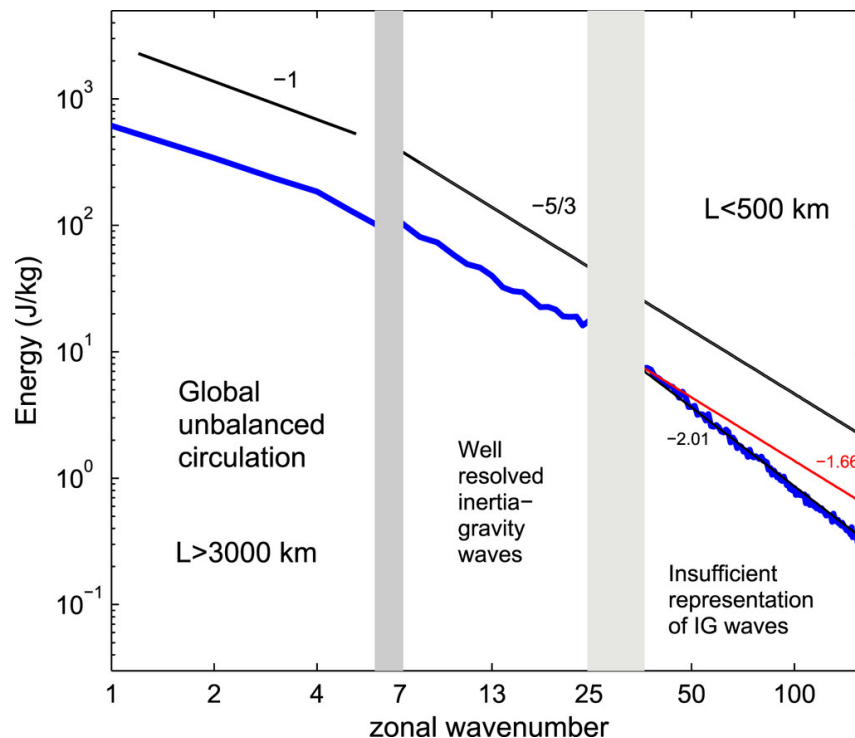


Figure 1: Three components of the circulation projecting on the inertia-gravity modes. Circulation projecting on inertia-gravity eigensolution of the linearized primitive equations at large scales is termed “unbalanced circulation”. In other words, this is circulation made of non-Rossby modes. From Žagar et al. (2017, J. Atmos. Sci.), their figure 15.

Furthermore, we have added the description of the “unbalanced” circulation at multiple instances within the manuscript, including in the Introduction (lines 54-55): “Here the term ‘unbalanced’ denotes circulation that projects on the inertia-gravity (or non-Rossby) eigensolutions of the primitive equations (Kasahara and Qian, 2000)”, and in lines 110-112: “Unbalanced circulation is derived using the normal-mode function decomposition and it corresponds to the circulation projecting onto the inertia-gravity eigensolutions (or IG modes) of the linearized primitive equations (Žagar et al., 2017).”

References:

Žagar, N., Jelić, D., Blaauw, M., & Bechtold, P. (2017). Energy Spectra and Inertia–Gravity Waves in Global Analyses, *Journal of the Atmospheric Sciences*, 74(8), 2447-2466

Comment: This regionality may help explain the shortcomings of metric 8 and its lack of correlation with the other metrics. I noticed that metric 8 uses only the energy from the $k=0$ unbalanced flow mode. This makes sense in as much as the Hadley circulation is zonally uniform. But in reality, the Hadley cell has centers of action, and there are even longitudes with counter-Hadley-cell-wise circulations. Much of the activity of interest when we study the Hadley circulation may be regionally focused. Perhaps the observed regional variations that are changing the zonal ψ -based metric are simply not captured by the changing energy of the $k=0$ unbalanced mode. I wonder whether including the energy from modes with $k \leq 4$ (in the tropics and subtropics) into a metric like metric 8 would help to capture the kinds of variations that are producing the temporal variations in ψ -based metrics.

It must be a little disappointing to create such an interesting and seemingly holistic metric...only for it to apparently underperform. It is a credit to the authors that they acknowledge its underperformance in this case. I can't help but wonder if there's more to it, though.

Response: We agree with the Reviewer that longitudinal inhomogeneity is an important characteristic of the Hadley circulation. However, all metrics of the global Hadley cell strength perform zonal-averaging by definition. Thus, we also opted for $k=0$. Besides, the zonal-mean state ($k=0$) modes also contain the majority of the total energy (kinetic plus available potential). Adding higher zonal wavenumbers ($k=1,2,3,4$ or even all zonal wavenumbers) barely affects the metric's variability and trend, as shown in Fig. 1 below (lines denoting IG ($k=0$) and IG ($k \leq 4$) largely overlap). This suggests that regional effects do not explain the discrepancy between the ψ -based metrics and the total unbalanced $k=0$ energy metric.

By default, our new metric based on the unbalanced circulation energy does not include the MRG waves that are obtained as the $n=0$ balanced mode by the numerical procedure of the K&S (1985) used by MODES. In principle, we would like to have it included among unbalanced or non-Rossby modes. We have tested whether adding the complete spectrum of MRG waves to the original metric affects the trends. Furthermore, we have also computed the complete zonal wavenumber spectrum of unbalanced circulation with and without the Kelvin waves, which

describe a large part of the tropical zonal circulation. None of that changed the variability or trends significantly. This may be because we use monthly mean fields to compute the energy metric even though the predominant variability in both the Kelvin and MRG waves is in intra-monthly periods, as can be seen in their real-time diagnostics

<http://modes.cen.uni-hamburg.de>. However, using daily data is computationally much more expensive, and thus not used here.

For the above reasons, we opted to continue using our original metric, the total energy contained in the zonal-mean unbalanced circulation, and we stress that a lack of correlations with other metrics does not necessarily reflect a discrepancy in physical mechanisms behind the processes making the Hadley cell and its trends.

We also note that the energy metric contains more than just the meridional wind signals, as the unbalanced circulation also represents the zonal wind and geopotential fields. Using the daily data (instead of monthly) and other refinements are necessary to improve our energy metric.

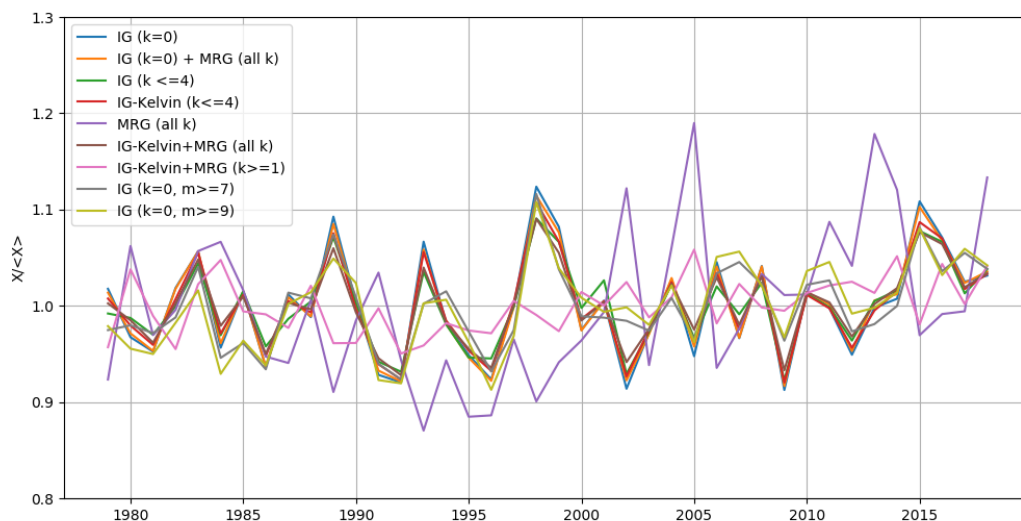
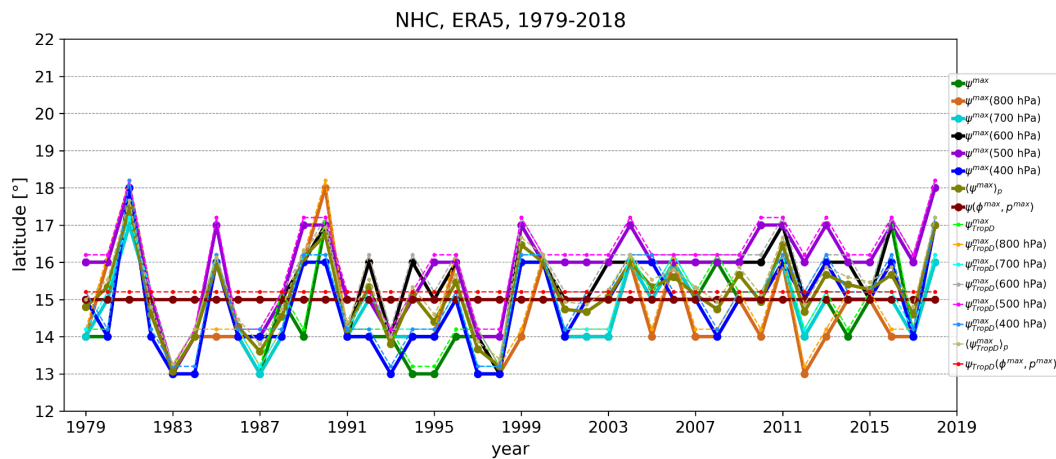


Figure 2: Sensitivity of the energy-based HC metric to the choice of zonal scales and inclusion of the Kelvin and MRG modes. Different lines are explained in the legend: IG refers to all inertia-gravity modes and no MRG wave, the inclusion of the Kelvin and MRG is indicated with + and -, and k stands for the zonal wavenumber.

Comment: Another outgrowth of Hadley cell width working group efforts that may be of use for the authors was the TropD software written by Ori Adam (see Adam et al., 2018). This software is built for detecting just the kind of metrics Pikovnik et al. use, and does so in a careful manner. I recommend trying the code and verifying that the metrics (or a subset of them) calculated by Pikovnik match those by Adam's software. The detection of extrema is particularly thoughtfully handled in TropD. It would also ensure that your results are intercomparable with those of other recent papers, such as Menzel & Waugh (2019).

Menzel & Waugh’s work is worth citing, I think, even though it is nominally about the subtropical jet, rather than Hadley cell intensity. In it, they find that the subtropical jet position is more closely related to Hadley cell intensity than Hadley cell width (and that Hadley cell width is more closely related to the speed of the subtropical jet than to its position). I think this could be cited in the introduction section, as a bit of motivation for studying Hadley cell intensity, as well as another example of a paper that uses metric 1. Also, it is interesting to note that the intensity of the zonal mean overturning (which is mathematically equivalent, I think, to the zonal mean of the overturning due to the “unbalanced” flow) would be so highly correlated with the subtropical jet, which is thought of as being in thermal wind balance.

Response: We thank the Reviewer for the suggestion. We have verified our results with TropD software by Adam et al. (2018). The results are shown in Fig. 3 below and show an excellent agreement with that of TropD software for large n , i.e. for the case with no smoothing applied (Adam et al., Section 2.3, their Eq. 1). In particular, this paper and Waugh et al. (2018), are very relevant as they show an extensive comparison of different metrics of HC width, thus we mentioned them in the Introduction: “Lately, multiple studies have compared different metrics of the tropical expansion (Solomon et al., 2016; Adam et al., 2018; Waugh et al., 2018). On contrary, the Hadley circulation strength has only been compared between different reanalyses and climate models (e.g. Stachnik et al., 2011; Chemke and Polvani, 2019), whereas no study (to our knowledge) has yet compared the metrics of the HC strength in the same dataset.”



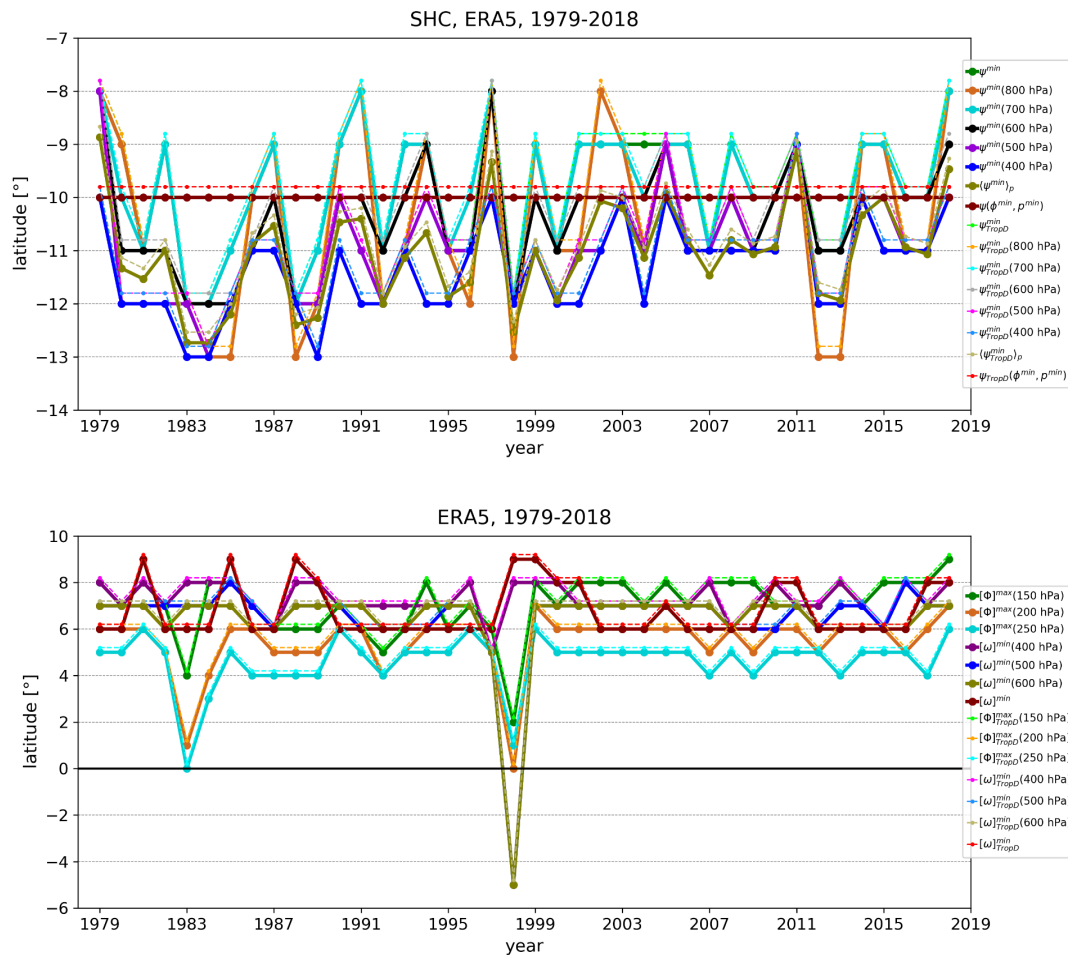


Fig. 3. Comparison of latitudes, where maximums of different HC strength metrics are obtained; produced by our methods (full lines) vs. TropD software (dashed lines) by Adam et al., 2018. Note that the dashed lines, obtained by TropD software, are increased by +0.2° for easier verification.

Comment: I am concerned about the use of the Mann-Kendall test for significance. This test requires data to have no autocorrelation. The authors say they use a “modified” version, but the paper does not say how or why it was modified, nor does it address the issue of auto-correlation in the time series. Figure 1 has large regions of weak signal (for example, during March in the lower troposphere over the SH) that are marked as significant. Much of this region is not marked as significant in ERA-Interim (Figure A1). Of course, differences in the trend from one dataset to another do not imply that the trend in one dataset is not significantly significant in that dataset. But perhaps the calculation of significance needs to be handled more circumspectly. That said, I appreciated how careful the authors were to distinguish between a trend that happens to be in the time series and trend related to forced climate change.

Response: We agree with the Reviewer that the “modified” MK test should have been explained better in the original manuscript. Indeed, we used the modified MK test using the pre-whitening method (Yue and Wang, 2002) that accounts for the autocorrelation of the time series. Thus, we have updated the text and changed the citation (lines 144-146): “The trends are considered significant if they pass the 95% threshold of the modified Mann-Kendall test using the trend-free pre-whitening method to eliminate the impact of serial autocorrelation (Yue and Wang, 2002).”

Note that we have also experimented with other derivations of Mann-Kendall test that account for autocorrelation, e.g. Hamed and Rao, 1998, and Yue and Wang, 2004. All tests yielded similar results on the significance of the trends.

Hamed, K. H., & Rao, A. R. (1998). A modified Mann-Kendall trend test for autocorrelated data. *Journal of hydrology*, 204(1-4), 182-196. doi:10.1016/S0022-1694(97)00125-X

Yue, S., & Wang, C. Y. (2002). Applicability of prewhitening to eliminate the influence of serial correlation on the Mann-Kendall test. *Water resources research*, 38(6), 4-1. doi:10.1029/2001WR000861

Yue, S., & Wang, C. (2004). The Mann-Kendall test modified by effective sample size to detect trend in serially correlated hydrological series. *Water resources management*, 18(3), 201-218. doi:10.1023/B:WARM.0000043140.61082.60

Little grammatical errors, odd wording, and writing problems are scattered through the paper. I'll mention just a couple, but the paper deserves some careful editing to fix these problems.

We have made a substantial effort (to the best of our ability) to improve the language in the revised manuscript.

The rest of my suggestions are mainly editorial.

Line 28: “minimum pressure velocity” is mathematically correct but also a bit confusing when what is being described is maximum ascent.

We appreciate the suggestion and have corrected the text accordingly.

Line 30: “the studies” is better as “studies” here, since a specific group of studies has not been referred to.

We have changed “the studies” to “studies” as the reviewer suggested.

Line 42: I can see why this sentence about Nguyen et al. appears in the same paragraph as the preceding list, but the topic of the sentence feels like it belongs in a new paragraph. It's also a little disappointing that the results from Nguyen et al. aren't described here; all that is said is that

they have relevant results. I suggest making this a new paragraph, and continuing by mentioning their relevant results. (And if they're not really relevant, the sentence could perhaps be deleted.)

We have decided to remove the statement that the metric by Nguyen et al. is the only study that addresses the vertical inhomogeneity of the HC strength and its trends. Besides, the pros and cons of different metrics are described in Section 2.2 below their definitions (lines 131-140).

Line 46: "we assess how sensitive are the trends" is awkward. "we assess how sensitive the trends are" is better.

We have reformulated the sentence as (lines 48-50): "For example, we assess the sensitivity of the trends derived from metrics based on the latitude-pressure stream-function profile ([\ref{eq:hc_strength}](#)) to the choice of the pressure level".

Line 50: It is taken for granted here that readers will understand what the "unbalanced" global circulation is, even though this is not a phrase

See also a response to Major comment 1. We have provided an additional explanation at multiple instances within the manuscript, including in the Introduction (lines 54-55): "Here the term 'unbalanced' denotes circulation that projects on the inertia-gravity (or non-Rossby) eigensolutions of the primitive equations (Kasahara and Qian, 2000)", and in lines 110-112: "Unbalanced circulation is derived using the normal-mode function decomposition and it corresponds to the circulation projecting onto the inertia-gravity eigensolutions (or IG modes) of the linearized primitive equations (Žagar et al., 2017)."

Line 60: While details are fine to be left in citations, some description or summary for the reader would be appreciated

We rearranged the data section, skipped the references to normal modes decomposition software, and rather expanded the description in section "Metrics of Hadley cell strength".

Line 69: "an extend" should be "the extent"

Corrected.

Line 70: "a part of the 40-year trends in the HC strength may be due to the multi-decadal variability." This is (virtually) given. Any time series can be decomposed into a trend + variability + error/noise. Say something more specific, or get rid of the sentence.

This should be true, but unfortunately, we have seen many manuscripts claiming a trend (e.g. climate change-related or model-error-related/bias-related), completely forgetting about internal variability. We merely want to point out that any 40-year trend is a combination of different factors, which we showed in a separate study as well (Zaplotnik et al., in review at J. Clim.). We think that a statement like this should be present to avoid misinterpretations of the trend by readers.

Line 74: I don't know if "feature" is the right word. It sounds like you are going to describe a particular anomaly. It might be better to say, simply, "Note that trends in..."

Corrected as suggested by the reviewer.

Line 81–82: as I mentioned above, in addition to the reliability of some climate model trends being called into question, I would question the reliability of the statistical significance test used here.

This was some uncareful wording on our side. We do not question the "climate model trends", but rather only the trends derived from the point metrics of HC intensity. We have thus removed the word "climate" from the sentence: "The presence of the inhomogeneities in the HC trends raises a question about the reliability of some of the trends derived from point metrics of the HC strength." See also related answers on the statistical significance tests above.

Lines 145–147: Of course the same feature (namely climatology) can be seen in Figure 1: Figure 1 is the climatology. Why not say, "Normalization accounts for some of these differences; normalized results are discussed near the end of Section 3.2 (see Figure 4)."

We thank the reviewer for an excellent suggestion. We have corrected the text accordingly.

Line 164: I don't know that "spurious" is the right word to describe point-wise trends. Maybe "non-representative" or "isolated." The point is not that the trends are wrong, but that they are probably not what the authors or readers are interested in.

We have removed the sentence in the revised manuscript, as the monthly-mean trends are replaced by seasonal-mean trends, where the described feature is not observed to such extent.

Line 175: "the measure of average HC strength" is vague. If it is a monthly average, any of these ψ -based metrics could be described this way.

Response: We agree that the terminology was dubious. Thus, we have changed the "average HC strength" to the "spatially-averaged HC strength" here and elsewhere in the text.

Line 243: "Insignificant correlations are not surprising as this index is largely different from all other indices." This is not the most satisfying or meaningful sentence. What does largely different mean? If the unbalanced flow is thought to be largely related to Hadley cell-wise circulation, isn't a higher correlation expected? I can't imagine the authors went through the effort to define this metric if they expected the correlations to be insignificant.

Response: We agree with the Reviewer that the sentence is unclear and it has been removed. There are multiple reasons for the lack of correlation. Figure 5 shows that metrics based on the same variable are correlated well; i.e. all stream function metrics are well correlated, including the new measure - spatially averaged stream function. The correlation of stream function metrics with the velocity potential and vertical velocity drops down to about 0.3-0.5. It is not surprising that a correlation with a very different metric based on a quadratic quantity that is associated with the stream function through the square of its vertical gradient is poor. It is hard to bring forward physical arguments behind the Ψ - ω correlation being about 0.4, and we do not have an explanation for the physical nature of correlations between the unbalanced energy and the stream function.

Line 246: “unimportant for the overall signal, but it may be important for the trends”. What is the difference between the signal and the trends? Often “signal” is used to describe “trends” and “noise”, is used to describe year-to-year or decade-to-decade “variability.” Did the authors mean “signal” or “climatology”?

We have now modified the text in lines 282-283 (former line 246) as: “The part of I_M from the extra-tropics and the stratosphere is however arguably unimportant for the overall magnitude of the metric.”

We have also alleviated the use of signal and noise in the revised manuscript to avoid misinterpretation.

Lines 250–252: Does this result imply that metric 8 is going to be sensitive to increasing tropopause height?

Metric 9 (former metric 8) as defined in the manuscript is not directly sensitive to increasing tropopause height, since the energy of the unbalanced motions remains steady. However, by imposing an artificial lid (e.g. at 100 hPa) like in the described experiment, the metric becomes sensitive to HC strength.

Line 265: “This was made evident by a new HC strength measure.” It is not clear what was made evident, or how it was made evident.

We agree with the Reviewer that the text is dubious. It is a remnant of an earlier manuscript version. Thus, we decided to remove the sentence as well as the preceding sentence.

Line 285: Based on my working group experience on Hadley cell width, I am skeptical of the idea of a “unified index.” Different metrics for Hadley cell width are of interest to different people for different reasons, even if their trends differ. If we created a unified index, there would be information missing from that index for each group. A unified index for Hadley cell strength would not capture simultaneously the differences between hemispheres, between regions, and in the deep tropical upwelling. Upper tropospheric circulation may be of interest to people studying the stratosphere or the tropical tropopause region, while mass is concentrated in the lower troposphere, so mass-weighted measures will leave them out.

We fully agree with the Reviewer on this one. What we meant was that studies like the IPCC report and similar other reports/studies that talk about a general strengthening/weakening of the HC should start using a “unified” HC metric, or to be more specific about why a certain metric is used over the other metric. Similarly, other studies should be specific about their use of the HC strength, and why they use it. As we point out earlier (lines 319-323) in the text, the use of the indices will depend on what we want to study (exactly what Reviewer suggests here). We have now rephrased the sentence to be more specific as to what we mean. See l. 319-326.