The study analyses the occurrence frequency of subcritical Richardson numbers, in a comprehensive near-global multi-year HVRRS radiosonde data set, as well as in the co-located vertical profiles in the ECWMF ERA5 reanalysis. Based on several comparability criteria like the seasonality, vertical distribution, and correlation coefficients in different climate zones, an adjusted Richardson number threshold of $R_i = 1$ is identified which exhibits characteristics most suited to diagnose regions in the ERA5 data which might be subjected to the occurrence of Kelvin Helmholtz instability.

I have two main concerns (see main comments below), one is more of a criticism on the wording, but it concerns the interpretation of the results at several points, and if the authors agree with the criticism, the wording and some aspects of the discussion throughout the manuscript have to be reworked.

My second main concern is that the interpretation of correlations and correlation coefficients in some of the analysis sections are too strong, which might therefore either need some reworking, a more detailed and broader description of the results leaving room for interpretation, or a reconsideration if these results are suited for publication.

Despite these main concerns, I found the results on the capability of the ERA5 to resolve subcritical Richardson numbers in comparison with the HVRRS interesting, and I think these along with the climatologies as key results could be suited for publication, however, only after consideration of the comments below.

I think the sections on the orographic and dynamical environment of regions of small Richardson numbers need clarification, mainly but not only in the context of the two main comments.

**Main comments:**

01) I am uncertain to what amount the identification of vertical segments with $R_i < R_i$ (with $R_i = 1/4$ in HVRRS and/or $R_i = 1$ in ERA5) is equatable with the identification of KHI.

   1. Subcritical Richardson numbers are a necessary but not sufficient criterion for the occurrence of dynamic instability and KHI.
   2. Negative values for $N^2$ indicate convective instability. If the data was not filtered, I would expect convection to contribute significantly to the statistics of subcritical Richardson numbers.
      - **HVRRS:**
        - In Fig. 15, the secondary occurrence frequency maximum $OF(R_i < 1/4)$ below vertical wind shears of 10 m/s/km could be indicative of such convective contributions. But this is hard to say for sure because all values are vertical averages.
        - At p.11 L.304 you include convection in the PBL in the interpretation.
      - **ERA5:** Negative static stability in the model is (as I understand it) quickly resolved through the cloud parametrisation, and therefore rare in the model output. Still, regions of low static stability ($\rightarrow$ small $R_i$) could be indicative of regions prone to convection, particularly for the adjusted threshold of $R_i = 1$.
        - At p.13 L.361 you mention both KHI and convection in the PBL (diurnal cycle?).
        - At p.15 L.418 you write the $OF(KHI)$ maximum is related to convective heating.

I realize that Fritts et al. (2014) (who you cite in the introduction) discuss how the distinction between dynamic and convective instability can be “challenging an misleading”, but I am not sure how this discussion for Lidar measurements in the MLT region is applicable for the troposphere.
Would you describe convection in the troposphere generally as weak?

I would suggest to replace OF(KHI) in the manuscript by OF(Ri<\text{Ri}_c), and either expand the discussion and/or leave it to the readers interpretation which process dominates in each region and season. See also the more specific comments further below.

02)
I am not sure if the interpretation of the data based on the derived correlation coefficients is always justified:

1. The paragraph beginning at p.14 L.399: While I understand the idea behind the analysis and the approach, I am not sure that the interpretation is strongly supported by the data. From the plot it is not really possible to connect the scattered data with the derived linear correlation in Fig. 11. I would assume that a 2d-histogram would reveal that the highest data density is clustered at low SDOR, and that the moderate-to-large SDOR values are negligible for the estimation of the correlation coefficient. Therefore, I would say the interpretation "over mountainous areas a high low-level OF(Ri<1) would be expected" is too strong of a statement based on the data. But of course this is open for discussion.

Furthermore, could you point out which regions in Fig. 8 you mean with the first sentence at p.14 L.399?

2. The paragraph beginning at p.15 L.425: Again, I understand the approach of the analysis and the idea behind it, but I would not agree that the conclusion is justified. The correlation is very weak. It is likely dominated by the point cloud at highest data density at low near-surface-winds. I would not agree that this supports the interpretation at the end of the paragraph.

3. For comparison: At p.17. L.473 you write: "[The ERA5 determined occurrence frequency of Ri<1/4] is poorly correlated with HVRRS-determined ones at all heights and over all climate zones."

If this refers to Fig. 7, it might be too strong of a statement considering the overall interpretation of correlation coefficients in the manuscript.

Other scientific comments:

**ABSTRACT**

p.2 L.32 I would prefer to write "KHI is indicated by the critical value of dimensionless Richardson (Ri) number, which is predicted to be ¼ from linear stability analysis." or something along these lines.

Otherwise a reader without background knowledge might perceive the critical Richardson number as a purely empirical threshold.

p.2 L.34 and L.36 I would suggest to delete the brackets with "137 level" and "10 m", since these values are not directly comparable.

p.2 L.45 In the context of major comment 01, I would suggest to replace "The occurrence frequency of KHI..." with "The occurrence frequency of subcritical Richardson numbers."

p.2 L.45 From Fig. 9d and 10k I would argue that there is also a seasonality in the tropics. I see your point that it might overall be less pronounced, but I still would prefer to not exclude the tropics specifically. There are large scale flow features like the Summer Asian Monsoon and the tropical...
easterly jet or the variability of the Walker circulation (also in relation to ENSO) which could be associated with the tropical seasonality in subcritical Richardson numbers (Roja Raman et al. (2009), Sunilkumar et al. (2015), Kaluza et al. (2021)).

INTRODUCTION

p.3 L.70 Baumgarten and Fritts (2014) focus on the MLT region, and although the statement (or parts of it) might be applicable to the region analyzed here, maybe there is a more general source in literature which describes the environment in which KHI arises.

p.3 L.73 Just to clarify, with “the associated large gradient in jet stream” you mean the wind gradients? Or stability gradients? Maybe you could rephrase it to make it more clear.

p.3 L.74 “which may increase clear-air turbulence (Williams and Joshi, 2013).” Do you mean it could increase in a changing climate? If this is the case, maybe rephrase it in an individual sentence.

p.3 L.81 Maybe remove the sentence “Among others, KHI is one of the most common causes of turbulence throughout the atmosphere from Earth’s surface to the lower thermosphere (Fritts et al., 2011; Sharman et al., 2012).” This is somewhat repeated from the sentence before.

p.4 L.89 “Values of Ri close to zero favor strong instability, deep billows, and relatively intense turbulence, whereas values of Ri closer to ¼ favor weak instability, shallow billows (Fritts et al., 2011).” I am not sure if this information is necessary for this study.

p.4 L.91 “The threshold value of Ri can be potentially used an indicator of turbulence (for instance, Jaeger et al., 2007).” Suggestion for clarification: “The Richardson number criterion can be applied as a turbulence diagnostic in numerical model output (e.g. Sharman and Pearson, 2017), and it has been used as such in climatological studies on the occurrence of clear air turbulence (Jaeger and Sprenger, 2007).” Since you cite Jaeger et al. (2007), a more direct comparison with your results in the later sections could be interesting, since Jaeger et al. also presents a Richardson number climatology with an adjusted Ri\textsubscript{t} threshold. And it would put your results in context with the literature.

Kunkel et al. (2019) include a brief discussion on the capability of ECMWF models to resolve subcritical Richardson numbers, and argue as well that Ri\textsubscript{t}=1 might be a good proxy for observed KHI (although only in a case study).

A very recent study on this subject is by Lee at al. (2023), in their climatology on UTLS turbulence diagnostics. They also set the Richardson number threshold from 0-1.

p.4 L.101 “While in numerical models, turbulent dissipation rate, turbulent diffusivity and other parameters representing turbulent mixing efficiency are the most basic parameters, which need to be accurately parameterized to evaluate the impact of turbulent effect on matter and energy distribution (Gavrilov et al., 2005). For this reason, the indices of turbulence, such as large wind shear, small Ri, the negative squared Brunt-väisälä frequency, could be a great tool to characterize turbulence (Jaeger et al., 2007).” I am not sure how the two sentences and the two citations fit together. Can you maybe rephrase the key statement you want to make here?

p.5 L.121 Is AEOLUS data assimilated in ERA5? At all, or consistently for the time period analyzed?
Thus, our objectives are to: (1) Evaluate the performance of ERA5 at different heights and climate zones in estimating wind shear and small Richardson number occurrence frequencies, in comparison with a large high-resolution radiosonde dataset spanning the years from 2017 to 2022.

2.1 HIGH-RESOLUTION RADIOSONDE DATA SET

Are 115 million radiosonde profiles analyzed in this study? I was wondering because taking the upper limit of the color scale in Fig. S1 times the number of radiosonde stations 3000*434 equals about 1.3 million soundings. Maybe I am missing something.

2.2 ERA5 REANALYSIS AND THE COLLOCATION PROCEDURE

I would suggest to delete the part “...and 500 m in the lower stratosphere.” It is not really necessary nor supported by Fig. S2.

If the orographic gravity wave dispersion is an ECMWF product, it should be listed here as well.

2.3 THE OCCURRENCE FREQUENCY OF KHI AND ITS UNCERTAINTY

I am missing an introduction of the critical Richardson number as a quantity which can be derived from linear theory. I personally would include this information in a study which focuses on the Richardson number.

Please give the definition of the Brunt-Väisälä frequency and the vertical wind shear.

“The occurrence frequency of KHI is defined as the ratio of Ri<1/4 relative to all Ri calculations at a specified time period or height interval.” I don’t think this sentence is necessary, and it can be misleading since you vary the threshold for the Richardson number.

“In Eq.(1), the length scale of overbar could potentially impact the value of Ri, and eventually, the occurrence frequency of KHI. In addition, the critical value of Ri and the vertical resolution of archived radiosonde could also cause the change in Ri values.” I would suggest to rephrase these two sentences to make the key point more clear. Suggestion: “The Richardson number calculated from Eq.(1) depends on the vertical resolution of the underlying data, as well as on the averaging interval. Ultimately, this influences the estimated occurrence frequency for subcritical Richardson numbers as a proxy for KHI. We resample the HVRRS data...”

2.4 GRAVITY WAVE ENERGY

“Only the perturbations with vertical wavelengths of 0.3–6.9 km are considered as GWs (Wang and Geller, 2003).”
For clarification, is this due to the vertical limitation from 2-8.9 km (i.e. 6.9 km as an upper limit) and the effective vertical resolution of 150m (i.e., half a wavelength) in Wang and Geller (2003)? Or do I misread this information? How does this apply to your data? Maybe this information should be provided after the description of the “tropospheric segment”.

3 RESULTS AND DISCUSSION

p.9 L.250 please provide a link to Table 1 if the value 4m/s/km is taken from this table. If it is not the case, please indicate how the value was derived.

p.9 L.251 “However, the oceanic shear is hard to be quantitatively assessed by a large number of in-situ radiosonde stations, with this aspect likely being evaluated by the ship-based radiosonde.” I am not sure if this sentence is necessary, from the data description it is clear that radiosonde soundings over the oceans are sparse.

p.9 L.253 “Over the tropical oceans, Savazzi et al. (2022) found the wind bias between EUREC4A field campaign and the ERA5 reanalysis varies greatly from day to day, attributing to the bias in wind forecasting in the ERA5 reanalysis.” Again just a suggestion, but I personally would remove this sentence as well, since Savazzi et al. (2022) has been cited in the introduction, and the statement made here is not directly linked to the analysis. It might increase the reading flow if the manuscript is a bit more compact.

p.10 L.258 Maybe provide a link to Fig. 2b at the end of the sentence.

p.10 L.259 Suggestion: “It is noteworthy that shear in the ERA5 reanalysis at heights of 10–15 km a.s.l. is significantly underestimated compared to the HVRRS, especially at middle latitudes, with a mean absolute error for all stations of about 8 m/s/km (Table 1).” (Assuming the value 8m/s/km is taken from Table 1)

p.10 L.265 Suggestion: “Following Houchi et al. (2010), the monthly shears over seven typical climate zones are separately investigated (Fig. 3), which are defined as follows: polar (70°–90°), mid latitudes (40°–70°), subtropics (20°–40°), and tropics (20°S–20°N). Over the polar region HVRRS-based shears are exceptionally strong in the stratosphere (Fig.3a, g), which could be attributed to the stratospheric polar jet.” I think it is clear from the data description and the plots that both Hemispheres are analyzed separately.

p.10 L.274 Is the value 16 m/s/km taken from Table 1? If this is the case, please provide a link in the text. If it is not the case, please indicate how the value was derived.

p.10 L.275 Maybe remove “in the Northern/Southern Hemisphere”, it is again implied.

p.10 L.277 and Fig.3 h-n: Could you please comment on the vertically aligned wind shear structures in the ERA5 derived wind shear in summer 2017 and beginning of 2020? Are these real or artefacts due to the data processing? They look suspicious particularly since the figure shows monthly averages. Furthermore, why do the regions of missing values in Fig. 3g and 3n differ (mid 2017 and end 2022)?

p. 10 L.279: “…which is about 3 km lower than that in the HVRRS.”
Just out of interest, would you say this is the same shear layer shifted downward, i.e., a 3 km altitude bias in the dynamic structure of the UTLS? Or are the shear magnitudes at the altitude of the maximum in ERA5 comparable to the ones in the HVRRS and the model fails to resolve the further increasing shear in the stratosphere?

p.10 L.283 I would suggest to remove the second part of the sentence, and just write: “The comparison between HVRRS-based and ERA5-based shears at three typical regimes are tabulated in Table 1.”

Because the altitude range from 10-15 km is not the middle and upper troposphere, particularly at high latitudes.

Furthermore, if the mean absolute errors at page 9 and 10 (4m/s/km, 8m/s/km and 16m/s/km) refer to Table 1, then it would be better to introduce Table 1 beforehand.

p.11 L.292 (Fig. S3): Just a general thought, it could be interesting to see this comparison for the troposphere and the stratosphere separately, or for different altitude intervals.

p.11 L.304 “The high occurrence frequency in the PBL regime could be likely related to the convective activity that leads to a negative N².”

Please see major comment 01.

p.11 L.307 “In addition, an obvious seasonal cycle of occurrence frequencies is revealed by HVRRS in the middle and upper troposphere and has a maximal in the spring season (March–April–May), which is consistent with the finding in Zhang et al. (2019).”

The maximum in the middle troposphere is during spring, the upper tropospheric maximum is also evident in winter I would say.

p.11 L.313 What does the value 8% refer to? Is it an estimation from Fig.4, or was it calculated from the data? If the latter is the case, which data exactly?

p.12 L.319 In Fig. 5 it says “10-15 km”.

Figure 6: Maybe line plots would increase the readability compared to scatter plots. Or smaller markers.

p.13 L.345 Please see major comment 01. Of course this is open to discussion, but I would suggest to replace OF(KHI) in the following analyses with OF(RiERAS<1) or something similar. I don’t think this lessens the impact of the results.

p.13 L.358 I am not sure if I can follow this argumentation. Could you point out exemplary regions for the spatial consistency? How meaningful is it to compare multi-year averages for these two highly variable quantities?

p.13 L.361 See major comment 01.

p.13 L.366 Suggestion: “In the free troposphere the spatial-temporal variability of OF(RiERAS<1) keeps high consistency with OF(RiHVRRS<1/4) over all climate zones.”

I would refrain from specifying the altitude range up to 30km because of the stratospheric polar maxima in the HVRRS.

Figure 9:

• What are the vertically aligned structures in the ERA5 analysis at the beginning of 2020?
• Why is the ERA5 data here (and in Fig. 4) cropped at 29 km?
• Why are there missing data patches in the beginning of 2022 in the ERA5 data across all climate zones?
• Why are contour lines included in the ERA5 plots but not in the HVRRS? Maybe keep them similar for better comparison.

p.14 L.372 I would suggest to remove the sentence “For regions without high-resolved wind and temperature measurements, the ERA5 model product could be a good choice to represent the thermodynamic instability of background atmosphere.”
This has been introduced in the motivation for the study.

p.14 L.377 Suggestion: “The seasonal variation of \( \text{OF}(R_i < R_i^t) \) with \( R_i^{HVRRS} = 1/4 \) and \( R_i^{ERA5} = 1 \) for all climate zones is further analyzed in Figure 10. In the mid latitudes and subtropics, the \( \text{OF}(R_i < R_i) \) exhibits maximum values in the PBL, as well as a local minimum in the middle troposphere and a local maximum at altitudes around 9 km. In the stratosphere the occurrence frequencies decrease to values of the order of 1% (Fig.10b,c,e,f).”
There is a significant variability in the vertical profiles across the different data sets, seasons and climate zones, and the single-value average occurrence frequencies in the current version of this paragraph might be confusing.

Figure 10: I would suggest to remove the gray shaded areas, I don’t think they are necessary for the analysis.
Please adjust the x-axis range in Fig. 10k and 10l so that the JJA maxima are not cut off.

p.14 L.387 I would suggest to delete the whole paragraph down to line 392, I don’t think it is important for the analysis, and kind of redundant with the paragraph before. I believe an overall reduction of the length would improve the manuscript.

p. 14. L.399 to L.406 Concerning the whole paragraph, please see major comment 02.

Figure 11: Could you plot the data as a binned 2d-histogram? What is the unit at the x-axis?

p.15 L.416 See major comment 01.

p.15 L.422 See major comment 01.

p.15 L.425 to L.431 Concerning the whole paragraph, please see major comment 02.
Provided that you agree with my criticism concerning the results of the analysis, I don’t know what would be the best way for improvement:
You could either go without this part of the analysis (I don’t think it would hurt the overall impact of the manuscript too much).
Or you could present the result without such a strong interpretation and leave more room for interpretation. I am not sure how meaningful this would be.
Or you could go into a more detailed analysis to try and sharpen the results. If this is within the scope of the study.

p.16 L.436 “Overall, large \( \text{OF}(KHI) \) always corresponds to strong GW activities and large wind shears, likely indicating that GW activity is crucial for the occurrence of KHI.”
Is this derived from Fig. 14b? If so please link the Figure in the text.
I also think that a more detailed description would be helpful. What do you define as “strong GW activity” in the plot? Where do you see the correlation between strong GW activity and \( \text{OF}(R_i < R_i) \)?
Maybe I am misreading the plot, but I find it hard to follow the conclusion.
Is the orographic gravity wave dissipation an ECMWF product? If this is the case please include it in the data description.
And just for clarification, is this the quantity derived from the parametrised gravity wave drag due to subgrid-scale orography? So you identify regions and time steps (months) of strong resolved gravity wave activity based on the parametrised non-resolved gravity waves?

Figure S5: According to the text (p.16 L.443) the correlation is based on monthly averaged values (of the gravity wave dissipation and the OF(Ri<Ri_t)). Please include this information in the description of Fig. S5.
Again just for clarification: What years are the data basis for this plot? 2017-2022? So about 60 data points are correlated at each grid point?
So during months with strong parametrised gravity wave activity, a strong activity of resolved gravity waves can be expected, which then modify the flow and stability parameters of the resolved flow, and result in an enhanced occurrence frequency for low Richardson numbers.
However, in regions where according to Fig. 8b small Richardson numbers are rare (e.g. over the Rocky Mountains, the Andes, Scandinavia, the Alps).
It would be interesting to see a time series of the monthly averaged gravity wave dissipation and the OF(Ri<Ri_t), for example over the Rocky Mountains. To get an impression of what this correlation means in terms of absolute values and the variability of OF(Ri<Ri_t). But this is maybe beyond the scope of this study, since it is already pretty comprehensive.

Figure 15: I am not sure if I read this plot correctly. The bin size for the filled color contour (i.e., the OF(RiHVRRS<1/4)) is apparently something of the order of 1 m/s in the x-axis, and 1 m/s/km in the y-axis. If this is the case, shouldn’t most of the filled contour display missing values, within 5x5 bins where only 1 or 2 matched profiles are located?

“The occurrence of KHI is potential crucial for many implications, such as aircraft, mass transfer, and climate change, just name a few, but it is very hard to be globally understood due to its fine structure.”
Please rephrase this sentence.
Aircraft → aircraft safety?
Climate change → how?

This study uses the ERA5 as the latest reanalysis product from the ECMWF as well as a comprehensive data set of HVRRS radiosonde soundings to globally characterize the distribution of low Richardson numbers as a proxy for the occurrence of KHI, for the years 2017 to 2022.

The underestimation therefore influences the performance of KHI analysis.

I would suggest to remove the sentence “The underestimation therefore influences the performance of KHI analysis.”

See major comment 02.

...especially in the middle and upper troposphere over midlatitude and subtropic regions in the Northern/Southern Hemisphere. "
Why especially in these regions?

The climatology of OF(Ri<Ri_t) exhibits significant seasonal cycles over all latitudes.”
When looking at Fig. 9 and 10 I don’t see why the tropics should be specifically excluded here.
This paragraph might need to be reworked based on how the manuscript was changed after the review.

Maybe this is a personal preference, but I would not write the final paragraph in subjunctive. For example (again just a suggestion): “Those findings are valuable for pointing out the performance of the ERA5 reanalysis in terms of resolving low Richardson numbers as a proxy for KHI, in comparison with a near-global high-resolution radiosonde measurement.” Same for the last sentence.

Please be more specific and include both data sets (ERA5 and HVRRS).

Technical corrections (typos etc.):

Throughout the manuscript: Maybe it should say “a.s.l.” instead of “a.s.l’’?

Remove the sentence “In addition, GW breaking has been identified as important sources of instability”, this is already stated one sentence earlier.

I would suggest to either write “The Richardson number is estimated by the…” or “Ri is estimated by the…”

“...in the supporting information.”

Suggestion: “Compared to ERA5, the HVRRS does not provide global seamless observations.”

“I would suggest to delete the sentence as it is already in the introduction.

“shear” instead of “shears”

“maximum” instead of “maximal”

Suggestion: “However, the ERA5 reanalysis does not provide such a seasonal cycle pattern..”

“underestimated” instead of “undervalued”. Same for “overvalued”.

remove “Among others”

I am not sure about the word “backbone”. Maybe replace it with “large scale structure” or something along those lines?

Replace “associated with El Niño Southern Oscillation (ENSO) events.” with “associated with the El Niño Southern Oscillation (ENSO).”

Typo: La Niña
References:


