Response to Reviewer 1

We would like to thank the reviewer for their valuable time and feedback. The reviewer's comments are in ***Times (italic and bold)***, our response is in Red Arial.

Summary

This paper describes the methodological implementation of a regime-specific quality control standards for the extremely rare limit (ERL) climatological test used to filter anomalous data from BSRN data. In its current implementation the ERL is applied globally to all sites in the same manner to remove outliers outside of the range of physical expectations. The authors argue that this test is far too conservative and develop a method for deriving regime dependent limits that more closely follow the data.

We sincerely thank the reviewer for their careful reading of our manuscript and for their accurate summary of our work.

Overall Feedback

I think that this paper is well written and worthy of publication with only minor revisions and response to some of my following questions. Most of the feedback below consists of recommendations about places in the manuscript to refine and be more numerically descriptive around figures and methodology.

We again thank the reviewer for the positive overall assessment of our work. We have carefully studied and fully addressed all the comments below.

At a pretty basic level, I need some clarification regarding why the ERL criterion on the plots are displayed as regions. The test, as I read the inequalities in eq. 1-3, appears to define a fixed negative lower bound and a SZA dependent upper bound limit. However, in figure 1 these are plotted as regions rather than a single upper-bounded functional limit. Could you please explain more directly in your response and in the text how these inequalities produce this bounding region? Does the area of the region come from the joint solar zenith and the sun-earth distance dependence, and thus there is a distribution at each SZA for diVerent times of year (changing R)?

Thank you for this question. The upper bound of the ERL test is not only dependent on the solar zenith angle but also explicitly incorporates the orbital eccentricity correction factor, which accounts for the Earth's elliptical orbit. Because the Sun–Earth distance varies continuously throughout the year, the top-of-atmosphere extraterrestrial solar radiation undergoes periodic fluctuations. Consequently, for any fixed SZA, the calculated ERL upper bound varies slightly depending on the time of year. When we plot the annual data and the corresponding QC limits on a scatter plot with SZA as the *x*-axis, this seasonal fluctuation effectively broadens what would be a single functional curve into a two-dimensional bounding region. To prevent similar confusion for future readers, we have added a brief explanation to the caption of Fig.1 of the revised, explicitly stating that this boundary region arises from the joint dependence on SZA and the time-varying Sun–Earth distance.

Another domain specific clarification – and I'm sure this is likely a conventional nomenclature thing, but I found the symbolic transition from irradiances to transmittances

kind of confusing initially. Some of it follows from the irradiances: Bn becomes kb (beam), and Dh becomes kd (diffuse) – so why does Gh not become kg but instead is written as kt (total?/transmittance?)? Perhaps a brief response is all that's necessary, but it could easily be addressed with a short sentence in-text I suppose.

Thank you for your careful review. Indeed, the notation is a bit confusing. Yet, this is actually a standard usage in solar energy meteorology. Therefore, we have added in the revised manuscript "the notation for the transmittances follows the widely accepted convention in solar energy meteorology (Shen et al., 2026; Ma and Yang, 2026)." In other words, we cited two recent references to elaborate our choice of symbols.

Could you potentially comment on why the northern "bar" and "tik" BSRN sites don't find their way into the category 2 despite sharing higher latitude seasonality of some of the other polarward clustered sites. Are tik and bar cloudier than the other members of group 2 that are likely very dry? The ena site is very cloudy, but is it really comparable to these higher latitude group members? Is this a sign that diffuse, either from clouds or surface reflection is playing a common role for these sites? In figure 5 I can see two rows that appear to have greater distance in some features – are these bar and tik? It is my understanding that some BSRN sites have information regarding upwelling measurements as well – is there perhaps something that can be learned here from surface reflectance in addition to transmitted radiation? It seems to me that these two cases might be distinguished by whether or not diffuse radiation is traced back to surface reflection or from scattering within clouds.

Thank you for your question and the short answer is "yes." Although BAR and TIK are high-latitude Arctic stations with a strong seasonal zenith-angle cycle similar to several Cluster 2 sites, they are assigned to Cluster 4 because the clustering is based on multivariate radiation-climatology features (13 extracted descriptors from the irradiance PDFs and their seasonal harmonics), not on latitude alone. In those features, BAR and TIK exhibit a cloud-dominated, diffuse-heavy, direct-beam-suppressed signature that aligns with Cluster 4 rather than with the drier, clearer-beam polar sites in Cluster 2. More specifically, see the table below. The features f3.kt.l, f5.kb.l, and f7.kd.l denote the fitted weights on the low-clearness components of the kt, kb, and kd mixture PDFs. Higher values indicate greater prevalence of attenuated, cloudy irradiance regimes. BAR has the highest f5.kb.l in the entire BSRN network (0.98) and an effectively negligible high-beam component (f4.kb.h ≈ 0.002). Its irradiance climatology is dominated by attenuated direct beam, consistent with persistent Arctic cloud/fog rather than a dry-polar radiative environment. TIK is also strongly cloud-biased in both global and beam statistics (f3.kt.l = 0.40; f5.kb.l = 0.76), closer to marine Cluster 4 sites such as LER (f5.kb.l = 0.86) than to dry Cluster 2 sites such as DOM or SPO. We have added some explanation in Section 3.1 in the revised manuscript.

Feature	Cluster 2 mean	Cluster 4 mean	BAR	TIK	Representative dry Cluster 2 sites
f3.kt.l (low global)	0.20	0.36	0.23	0.40	CAP 0.20, ALE 0.17
f5.kb.l (low direct beam)	0.51	0.83	0.98	0.76	DOM 0.17, SPO 0.37
f4.kb.h (high clear beam)	—	—	0.002	0.11	SPO 0.33, EUR 0.47

As for the second part of the question, indeed, some BSRN sites have albedo measurements, which could serve as additional features during clustering. However, insofar as the current structure of the paper is concern, we did not include those measurements for they are not available at all sites, hence making the clustering operation difficult. Similar arguments can be made for longwave measurements. Although all sites measure downwelling longwave, the characterization is very different from the shortwave components, e.g., how longwave clear-sky models behave. Hence, we added in the conclusion section a sentence, leaving that for future work. More specifically: “Additionally, incorporating auxiliary data that reflects site climatology, such as surface albedo or longwave radiation, could further enhance the robustness of the radiation climate classification. As climate classification remains an ongoing task, this work is intended to serve as a preliminary framework for future developments in the field.”

Specific Feedback

1. Figure 1: It would also be might also be interesting if figure 1 indicated the number and/or percentage of screened observations. It’s clearly very small by visual inspection, but it’s hard to tell with small scattered points in low-density regions – I think this change would bolster the numerical point the figure is making visually.

Thank you for your suggestion. We have added the number of screened observations into the figure.

2. Section 2.2.1: Could you further explain your selection of the harmonic order maximum of 25. In text you say that it was selected empirically – but by what metric? Is there perhaps a more solid physical motivation for the number of harmonics that are appropriate? If you divide 4 years into 25 intervals that’s something on the order of two months, was that the shortest oscillation period you wanted to be able to resolve? Could this reasoning be important to document for future studies? Furthermore, can you explain exactly how the two features are extracted (e.g., the maximum – minimum across all 4 years vs. average yearly maximum minus average yearly minimum)? Does this have consequences in how you consider annual and interannual variability in radiation climatology later?

We thank the reviewer for this insightful comment. The choice of m=25 was determined through an empirical sensitivity analysis aimed at balancing the goodness of the seasonal fit against the risk of overfitting. When testing lower harmonic orders (e.g., < 10), we observed that the seasonal fit was insufficient—effectively underfitting—and failed to adequately capture the true amplitude of the annual radiation cycle. Conversely, significantly higher orders introduced high-frequency artifacts that did not correspond to physical climatological signals. Regarding the extraction of the c_0 (mean irradiance) and harmonic-range features, these were derived from the intercept and the range (maximum minus minimum over the 4-year period) of the harmonic regression fit, respectively. Furthermore, as these two metrics are only a portion of the total feature set, minor deviations in these values are not expected to significantly impact the final classification results. We have updated the manuscript to clarify these methodological details and to document this reasoning for the benefit of future studies.

3. Line 131: Opening sentence of the paragraph at the start of section 2.2.2 is a little confusingly worded. I think it’s specifically because you’re listing the transmittance variables and later mentioning irradiance components and thus your use of “latter” appears to refer to a *list with more than a pair of variables (as is commonly how former and latter is used. Would this change maintain your meaning? :*

   a. *"The density-based features are extracted from ~~kt, kb, and kd~~ the transmittance components, rather than from the irradiance components themselves, because the latter random variables are "contaminated" by yearly and diurnal cycles and are thus unable to fully reflect sky conditions. "*

Thank you for your careful review. Replacing "kt, kb, and kd" with "the transmittances" would not change the meaning. We have revised accordingly.

4. Figure 3: For your beam transmittance fits with power-sn-sn mixture of distributions of the beam transmittance (kb), are you getting densities that exceed zero above the upper maximum of the distribution? Similarly there is also a poor fit near zero – presumably these are both features driven by a linear constant being added to your distribution. Could these discrepancies exceeding maximum and near minimum values of your distribution cause issues later?

We thank the reviewer for identifying these discrepancies in the beam transmittance k_b distributions. The slight density exceeding the theoretical maximum and the fit near zero are consequences of our parametric model structure, specifically the unbounded support of the skew-normal (SN) components and the mathematical behavior of the power-law component near zero, rather than the addition of a linear constant. These minor fitting artifacts do not impact our results, as our analysis relies on the mixture weights—rather than the probability density function tails—to characterize site-specific radiation climates. Furthermore, the quality control limits are derived independently from these mixture models using a zenith-angle-dependent approach, and our pre-filtering of the k_b range ensures that the maximum likelihood estimation is driven by the distribution's core rather than its boundaries. We have clarified these model characteristics within the revised manuscript.

5. Line 357: change "distinct regime physics" to "distinct physical regimes"

Thank you for your suggestion. We have followed your suggested change here.

6. Figure 5: Could you provide some quantification of the average intragroup distance (e.g., group 2 has greater distances within its own group, indicating greater diversity) and intergroup distances (e.g., how similar is group 2 to 4, or group 1 to 5?). For example it looks to me like group 1 and group 5 have a smaller WD to one another than other groups – given that the explanation for these group is different, then why? Their distributions in figure 7 also look similar (though with less low SZA sampling in group 5).

We thank the reviewer for this suggestion to quantify the cluster characteristics. To address this, we have included a new table (Table 2) in the revised manuscript providing the mean pairwise Euclidean distances in the PCA. In this table, the diagonal entries represent the mean within-cluster distances, while the off-diagonal entries represent the mean inter-cluster distances. The results confirm that each cluster maintains a distinct identity, as evidenced by the within-cluster distances (diagonal) being consistently smaller than the inter-cluster distances (off-diagonal). Regarding the reviewer's specific observation about Group 1 and Group 5, the distance matrix shows an inter-cluster distance of 3.929, which is indeed among the smaller inter-cluster values. As the reviewer noted, their distributions in Figure 7 are similar, which explains their proximity in PCA space. While these groups are distinct in their specific climatological drivers, they share commonalities in their radiation signatures—likely due to overlapping conditions in their respective regional climates. We believe this quantification enhances the transparency of our classification and validates the distinctiveness of the defined radiation climate regimes.

7. Figure A6: It is notable that both GAN and GUR seem to have anomalous dense sampling of high SZA – is there some issue with the temporal sampling at these sites? There's a relatively high sampling of observations above SZA~80 and a sharp cut-off below that doesn't appear in the comparable sites you selected. Is the SZA sampling itself anomalous compared to other sites? What does that mean for this regime cluster? Is this just a data collection artifact? The dispersion in BNI seems like a reasonable explanation for why these groups cluster, but what causes this unique SZA dependent sampling? I didn't notice this in figure 7, why?

This is an excellent question! To clarify, it is important to note that our study uses raw data directly from the BSRN archive, which have already undergone preliminary QC by station scientists prior to submission. The anomalous density of high-SZA observations and the sharp cut-off at 75 degs are artifacts of the standard BSRN QC protocols applied during data ingestion. Specifically, the BSRN guidelines apply distinct thresholds based on the solar zenith angle (SZA). For example, while the closure test and diffuse ratio tests have consistent requirements, the thresholds are adjusted for high-SZA conditions Z > 75, often requiring a GHI > 50 W/m^2 to maintain validity. At these Indian sites, which are historically prone to lower data quality, the station scientists implement stringent removal of flagged points before archiving. Consequently, the data points visible in the high-SZA regime are those that survived these specific, localized QC checks, leading to the observed 'sharp cut' and dense sampling density compared to sites with less rigorous or different local filtering. These patterns are not reflective of the underlying radiation regime itself, but rather the manifestation of site-specific pre-processing. We have updated the manuscript to clarify that these clusters reflect the filtered state of the BSRN archive at those specific locations and that such sampling artifacts are inherent to the data provenance.