Manuscript Title: A New Approach to Inversion of Multi-spectral Data with Applications to FUV Remote Sensing
Manuscript Number: egusphere-2025-5570

Dear Editor and Reviewers

Thank you very much for the detailed and thoughtful reviews. In the following text, reviewers' comments are in bold and our response is in plain text. Comments are referred to by the line number in the annotated preprint where applicable. Responses to the reviewers are provided in numerical order.

1 Response to Reviewer 1

1.1 Summary comments

Comment: The approach presented is interesting and may be useful, but the more information on the context for the present work is needed. For example, the relationship of the errors and uncertainties in the present and previous approaches is not sufficiently clear. While some of the figures suggest much smaller errors (apparently 10K) than in the current GOLD data products (30-40K near solar max) the temperatures retrieved for the May 2024 storm deviate dramatically from the released data which show unambiguous consistency between the temperature and composition fields, an indication that the temperature structure is real but is not apparent when the approach described in the paper is applied. The geophysically interesting structure observed during the storm seems to be obscured by the approach presented, possibly indicating biases inherent in the proposed approach when analyzing rare, atypical observations.

Response: With respect to the estimated errors being referred to (e.g., Figures 2 and 3 in the preprint), we would first like to point out that the results in Figures 2 and 3 show the continuous rank probability score (CRPS), not a raw error. The CRPS, given by Eq. 14 in the annotated preprint (Eq. 16 in the current version), is a measure of the accuracy of the posterior distribution that takes into account not only the location of the distribution, but the shape as well (e.g. Gneiting and Rafferty 2007). For this reason, it is commonly used for Bayesian models; however, care should be taken when comparing it to other error metrics. For example, given a true neutral temperature of 820K, an estimate of 800K, and a one-σ uncertainty of 30K under a Gaussian model, the CRPS is about 12K, which is comparable to our results on simulated data.

To mitigate potential misunderstanding and facilitate comparison to previous work, we decided to put our results in the context of C. Cantrall and Matsuo 2021, which is not a Bayesian method; we have altered both Figures 1 and 3. The new Figure 1 is shown in Fig. 1, and includes along the top row the neutral temperatures derived from that method. In order to facilitate direct comparison between the two methods, the results in the new Figure 3 are now in terms of percent error, and include the results for the method in C. Cantrall and Matsuo 2021. That figure is reported here as Fig. 2, and shows errors on the order of 3-5% for our method and 8% or more using the method of C. Cantrall and Matsuo 2021.

Comment: As noted by the authors, the approach presented may be useful for data with low signal to noise. The assumption of a normal distribution in GOLD's current temperature retrieval, while appropriate for most of the dayside where the difference between normal and Poisson statistics are negligible, is not appropriate for all solar zenith angles. However, the authors mischaracterize the current solar zenith angle restrictions in the GOLD data products. Retrievals are possible with lower signal to noise data, but the current retrieval introduces a cold bias (as discussed by Evans et al. 2024). While the figures presented by the authors show that the proposed approach also becomes subject to biases at the larger solar zenith angles where signal to noise is lower (Figure 3), it may be productive to explore the use of Poisson statistics for data with low signal to noise. The use of data at large solar zenith angles is further complicated by the increases in peak emission altitude that occur at large SZA (as discussed in Evans et al. 2024).

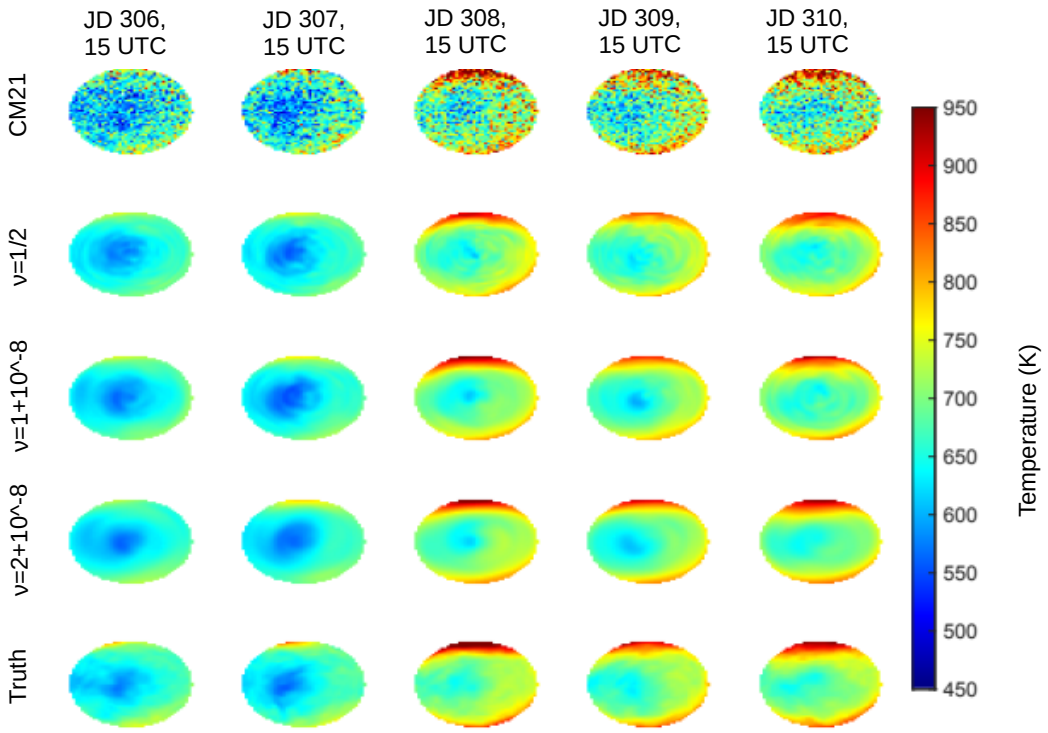

Figure 1: To replace Figure 1 in the manuscript.

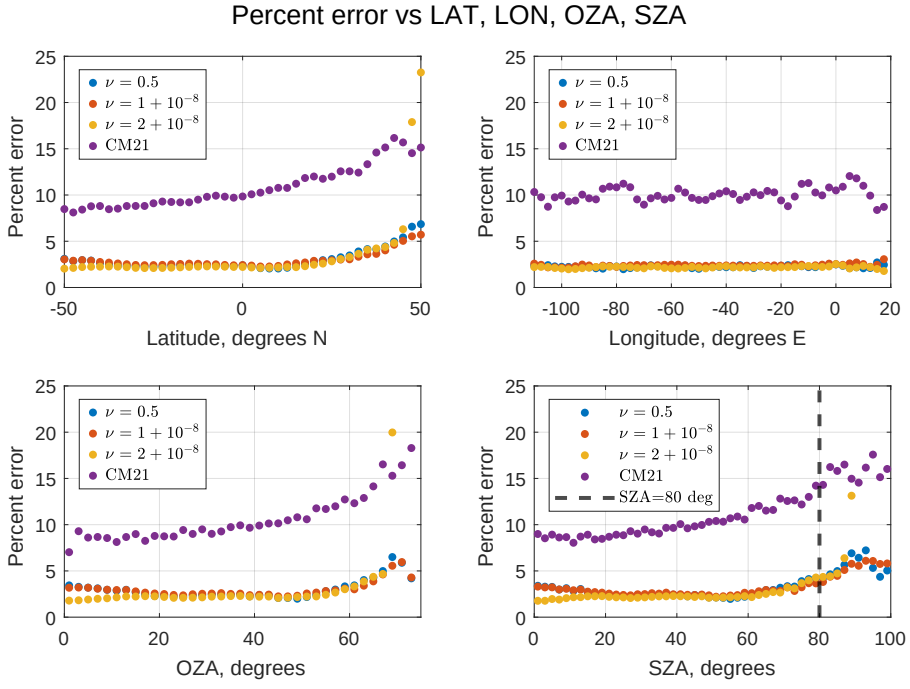

Figure 2: To replace Figure 3 in the manuscript.

Response: Thank you for the additional context regarding the GOLD temperature retrieval. We have addressed this in response to comments in the annotated preprint. We hope that our responses are able to address your concerns.

Comment: The paper has multiple assumptions in the work presented and in comments on previous work that need to be clarified or corrected. An example is the comment in line 98 that a fully calibrated instrument model. Such a model is unnecessary for any of the approached that researchers have used, but a relative calibration is necessary for all the approaches, especially when using multiple emission bands. Other cases are noted in the annotated .pdf file.

Response: We address this comment in the next section. Again, we thank you for your attention to detail and thorough review of the work.

Comment: A related issue is the possible geophysical implications of assumptions made, e.g., cap harmonics, by the authors. There is an underlying asymmetry in the morning versus evening or the northern latitudes versus the southern latitudes distorted by the symmetry within the harmonics. Another example is the significance of being dependent on the posterior distribution, e.g., in the May 2024 storm results.

Response: With respect to the use of spherical cap harmonics, the underlying asymmetry can be characterized by our method, as shown in, for example, Figure 3. This is because we do not attempt to estimate the field in terms of the spherical cap harmonics, rather we use them to describe the prior covariance of the latent field that we wish to estimate. This potentially leads to not being able to accurately estimate the field where there are rapid changes in the gradient, especially if the prior smoothness is too large. We also want to emphasize that our method does not depend on this specific kernel, any appropriate covariance kernel would have worked. The optimal one to use in a given situation is determined by the user. We have added a discussion on the importance of the choice of prior in various places, most notably a new section, Section 2.4, devoted entirely to sources of error. With respect to the choice of kernel, this section reads

> "The most prominent source of model misspecification error in our method could come from choosing the kernel matrix $\mathbf{K}$. Our method works by determining the emission intensities that best fit the model jointly over the entire domain, rather than at each location independently. This allows the method to mitigate the effects of shot noise, which is assumed uncorrelated across space. This effectively increases the SNR at each location. However, it is possible that small-scale structures, such as traveling ionospheric disturbances, can be smoothed over by this approach. This may be an acceptable tradeoff, particularly in low-SNR estimation scenarios, however its effects should be understood when interpreting the estimation results. The best kernel for a given application likely varies, and in the future a careful study of the effects of kernel choice on estimation is desirable. Note that this is a limitation associated with all Bayesian methods, as inclusion of a prior introduces some bias into estimates that must be understood in scientific and operational applications."

Comment: The limitations of the Cantrall are noted very late in the paper (line 312?), after the extensive discussions that are likely to be unfamiliar to most readers. Seems to deflate the significance of the work. If introduced earlier, could you compare the present results with the previous Cantrall Matsuo results and include some discussion of the role of Poisson statistics in the differences? That could clarify the significance and possible limitations of each new fundamental assumption in your retrieval, Poisson statistics and Bayesian probability.

Response: We have included discussion of sources of error in the newly written Section 2.4. We have also added a comparison to the results of C. Cantrall and Matsuo 2021 to the simulated data study.

1.2 Response to comments in the annotated preprint

- **Typo in the abstract**

  This has been fixed. Thank you for bringing it to our attention.

- **Line 3: "This assumption is valid if the count rate is high"**

  While this is a valid approximation when count rates are high, the count rates are highly variable in space and time, and so this approximation is not always appropriate. This is especially noticeable at high solar zenith angles, which is the motivation for this study.

- **Line 23: "You should provide the transition here (upper and lower states)"**

  We included this as a note on the history of these sorts of estimation techniques. Because of this and the comments on line 33, we have changed this portion of the introduction to state

  > "In particular, the Lyman-Birge-Hopfield (LBH) bands, which arises from the transition $N_2(a^1\Pi_g, v') \rightarrow N_2(X^1\Sigma_g^+, v'')$, where (v', v'') denote the vibrational levels of the upper (excited electronic) and lower (ground electronic) states, are some of the most prominent emission features in the FUV spectrum. Short-wavelength N_2 LBH bands, such as $(3, 0)$ and $(2, 0)$, are strongly absorbed by O_2, so their observed emission decrease with increasing O_2 density. Comparing these absorbed bands to less-absorbed bands allows retrieval of the O/N_2 composition ratio in the thermosphere. The rotational structure of N_2 LBH bands, arising from transitions between rotational levels within the upper $a^1\Pi_g$ and lower $X^1\Sigma_g^+$ electronic states, is used to retrieve the thermospheric temperature. Temperature and compositions are fundamental physical properties of the upper atmosphere and are important parameters for ..."

- **Line 24: ""Column integrated" is a more accurate term, but not necessarily for all techniques that have been published."**

  We have made this change.

- **Lines 25-26, Re: references about neutral temperature and satellite drag: "This list is incomplete and should be expanded"**

  We have added Doornbos and Klinkrad 2006; Krywonos et al. 2012; Leonard, Forbes, and Born 2012; Emmert 2015; Zesta and Huang 2016; Mehta et al. 2023 to this portion of the paper.

- **Line 28: "The count rate is dependent on more parameters than just SZA. Instrument performance matters, of course. If solar activity is high the count rate may follow normal statistics even at large SZA."**

  This is true. However, high solar zenith angle provides the motivating example for the current study, and stated here as an example.

- **Line 31: "Not sure what you're refering to here. Maybe the PSF?"**

  Due to the physical processes that govern the emission of photons in the thermosphere, the emission rates are spatially correlated and this correlation over space carries over into the temperature field. This is what we refer to when we discuss spatial structure. Our goal is to leverage this spatial correlation in the field to develop a method that is more resistant to the effects of shot noise.

- **Line 33, comments 1 and 2: "Is this technique actually widely used? This sort of statement should cite appropriate references." and "Composition is not retrieved from the LBH bands alone, which is how this comment could be interpreted. While O2 column densities are derived from differences in the band brightnesses that does not appear to be relevant nor is it discussed in this paper."**

  We agree that this is unclear, and that the use of the word *widely* in reference to band ratio based estimates of the neutral temperature is misleading. We intended this statement to be about the use of band ratio techniques in FUV sensing of the thermosphere in general, not just neutral temperature and not just with LBH bands. We have modified this statement to increase clarity. This portion now reads

  > "One class of inversion methods used to estimate properties of the thermosphere from FUV emissions is based on ratios of total observed counts in sub-bands of the received spectra. For example, changes in the shape of the LBH bands are correlated with the changes in ambient temperature (e.g., Aksnes et al. 2006), and the ratios of some of these channels have been

shown to have an approximately linear relationship with temperature under typical geophysical conditions (e.g., C. Cantrall and Matsuo 2021; Yongliang Zhang, Larry J. Paxton, and Schaefer 2019). Ratio-based techniques are widely used for $\Sigma O/N_2$ and QEUV data products using the ratio of atomic oxygen and LBH emissions (e.g., Strickland, Evans, and L. J. Paxton 1995; Y. Zhang et al. 2004; Meier 2021; Correira et al. 2021) as well as estimates of ionospheric structure (Yin, Qin, and Larry J. Paxton 2023). Outside of FUV sensing, these techniques are also used in LiDAR profiling (Jia and Yi 2014), spectrometry (Coath, Steele, and Lunnon 2013), x-ray astronomy (Park et al. 2006; Jin, S. Zhang, and Wu 2006; Wang et al. 2024), astrochemistry (Boersma, Rubin, and Allamandola 2012), and synthetic aperture radar (Gallardo i Peres et al. 2024), among others. The relative simplicity of the two-channel ratio inversion techniques makes them amenable to integration of spatial models."

- **Line 34-35: "This sounds a bit off as it is stated. The rotational distribution varies with temperature. In principle, all the rotational energy level contribute, throughout the temperature range that occurs in the atmosphere. Regardless of the rotational energy levels, the LBH band transitions are still between the same vibrational energy levels. Since the LBH bands are transitions between vibrational levels, the transitions are the same for all portions of a band. The amount of emission observed in the different portions of the band does change (is correlated) with temperature and is correlated with the temperature of the atmosphere."**

  This has been rewritten as follows:

  > "For example, changes in the shape of the LBH bands are correlated with the changes in ambient temperature (e.g., Aksnes et al. 2006)"

- **Line 37: "This should be qualified. I believe their work focused on only some of the temperature range expected in Earth's atmosphere**

  The text has been edited as follows:

  > "One class of inversion methods used to estimate properties of the thermosphere from FUV emissions is based on ratios of total observed counts in sub-bands of the received spectra. For example, changes in the shape of the LBH bands are correlated with the changes in ambient temperature (e.g., Aksnes et al. 2006), and the ratios of some of these channels have been shown to have an approximately linear relationship with temperature in typical geophysical conditions (e.g., C. Cantrall and Matsuo 2021; Yongliang Zhang, Larry J. Paxton, and Schaefer 2019)."

- **Line 42: "I'm not sure what you mean by spatial models"**

  We discuss what we mean by spatial models in response to the comment on line 31. These sorts of methods can be computationally intense since they attempt to recover the joint distribution of the quantity of interest at every location, however since band ratio models only involve taking ratios of counts they provide minimal extra computational load on top of that, making them natural candidates for more complex models like the one we describe.

- **Lines 48-49, comment 1: "I don't think such a statement was made in the TDISK paper."**

  We are referring to the following statement in the TDISK paper, Section 4.3: "The largest source of uncertainty in derived neutral temperatures is random noise in the disk scan observations, since random noise rather than systematic error constrains the fit of the forward model to the observed spectra". A similar statement is found in McClintock et al. 2020 Section 3.1 in reference to radiance measurements, from which the other GOLD measurements are derived: "Random errors in equation (2) arise primarily from the Poisson statistical uncertainties (variance = number of observed detector counts) in C and the sum S' + D + B ".

  We have modified the text in question as:

> "It is particularly relevant to upper atmosphere FUV remote sensing as shot noise effects are a major source of uncertainty in the radiance measurements from which the GOLD mission data products are retrieved, including the neutral temperature (McClintock et al. 2020; Evans et al. 2024)."

- **Lines 48-49, comment 2: "is this always true? How significant are the expected differences? at what signal to noise are the differences expected to be significant? What are the cases where one should be using Poisson statistics?"**

  Depending on the measure of fit used and the count rate, it may be significant. While to our knowledge there is no specific analysis of this for FUV remote sensing, there has been work done on this in the context of x-ray sensing. The work in Humphrey, Liu, and Buote 2009 showed that there are consistent biases of between 0.5-1 counts per sampling period derived from minimizing relative least-squares error to fit Poisson distributed data. This seems small, but it is meaningful at low SNR. They also found that the same bias persisted even at high SNR, and that the induced bias was the same magnitude as the statistical uncertainty unless the square root of the total counts in the dataset greatly exceeded the number of spectral bins. They argue that to eliminate the bias, the data in *every* bin used in the fit must be approximately Gaussian. Additionally, Yamada et al. 2019 found that the width of emission lines could be underfit by as much as 10% using Gaussian statistics even when photon counts were in the hundreds, and Nicolaou et al. 2024 found similar sized biases in the estimates of certain plasma parameters from x-ray data. It is not currently known whether a similar effect can be observed in the context of FUV remote sensing.

  The following has been added to the introduction:

  > "Even when photon counts are moderate to high, inversion techniques assuming Gaussian statistics have been observed to introduce bias into retrievals that can be on the order of statistical uncertainty (e.g., Humphrey, Liu, and Buote 2009; Yamada et al. 2019; Nicolaou et al. 2024). Additionally, commonly used inversion techniques do not take into account the spatial structure of the photon count sensor data, limiting their ability to recover the underlying spatially correlated physical parameters."

- **Line 56: "Are you saying that the relationship is fundamentally linear or that a linear relationship is assumed in that work?"**

  We are saying that C. Cantrall and Matsuo 2021 assumes that the relationship is linear in their analysis. The linear model comes from a fit to the Budzien model (Budzien et al. 2001) and is very accurate under typical conditions (in this study, there is a 3 K RMSE between 400 and 1000 K due to the linearization on simulated data, which is why we neglected these errors). The relationship used in the simulation study was calculated using the vibrational populations from Ajello et al. 2020, which were used in C. Cantrall and Matsuo 2021.

- **Line 67: "The SZA limit for the TDISK algorithm is conservative and is imposed to produce operational data products. The GOLD data could be processed to higher SZA as long as the data are filtered to make sure the SNR is higher than the threshold where a bias is known to occur."**

  Thanks for pointing this out. The text in question has been modified as

  > "These examples show that the proposed new approach is able to accurately retrieve the column-integrated temperature in a wide variety of geophysical conditions, both during geomagnetically quiet and severely disrupted periods, and attest to the potential for extending temperature retrievals with uncertanty quantification based on the LBH band emission to higher SZAs than currently made available"

  Similar statements elsewhere in the text have been edited in a manner to emphasize that the SZA/OZA limits discussed are limits at which the GOLD mission does not estimate TDISK. For example, text in Section 3.2 has been modified to state

"Currently, the GOLD mission does not report TDISK values at an observing zenith angle (OZA) greater than 75° or SZA greater than 80°, which are conservative limits designed to avoid biases in the data products that may arise in these situations (Evans et al. 2024). This partially motivates our investigation, as we believe that including spatial structure can allow us to obtain accurate estimates of the temperature even at high SZAs."

- **Line 68-69: "Since the use of Bayseian statistics depends on knowledge of prior events, how should the user identify and use it for atypical events which may not be represented in the previous data. some of the most interesting observations are atypical, the Gannon storm for example. The results shown in Figure 6 appear to lose much of the structure in the temperatures."**

  This is a good question, and relevant to Bayesian inversion methods broadly. The only prior information required for this algorithm is a prior covariance for the latent vector $\vec{f}$ through the combination of K and γ. One way to handle this is to relax the prior by allowing large prior marginal variances and minimal restrictions on the spatial fields, for example by setting γ to be small. A similar approach is used in the GOLD TDISK algorithm, which uses a very broad prior distribution. Indeed, it is common in the statistical literature to test robustness of methods to prior specifications, but it is important to highlight that flexibility in the prior is an important feature of this approach in that different modeling groups may use differing priors within the same framework.

  Another way to handle this would involve doing a cross validation procedure or similar for a wide range of geophysical conditions with a simulated instrument model and selecting the priors based on that. For example, one could determine the best prior parameters for estimation as a function of K_p, $QEUV$, D_{st}, or some other geophysical parameter(s), and use the appropriate one in given geophysical conditions. The best way to proceed depends on the specific aims of the mission and the analysis to be performed. Some discussion of this is contained in the newly added Section 2.4.

- **Lines 86-87: "Are you sure? The version I'm familiar with (and have used) calculates the rotational energy structure of each vibrational level of the LBH bands as a function of temperature. The vibrational populations are specified to the model, but they are independent of temperature." and "The model assumes Franck-Condon factors unless the user supplies the vibrational populations, and the populations are not temperature dependent. The sentence needs to be rephrased for clarity."**

  The version that we have is a lookup table of spectra at 1K resolution in temperature for each vibrational population contained in a netCDF file. The received spectrum is estimated by weighting the modeled spectra in accordance with the vibrational populations. This version of the model is the same as was used in C. Cantrall and Matsuo 2021 (see Section 2.1 therein). We have edited this section of the manuscript to avoid confusion:

  The forward modeling of the neutral temperature is based on the vibrational-rotational band model (Budzien et al. 2001), which supplies laboratory LBH spectra given vibrational populations of N_2 and neutral temperature.

- **Lines 94-96: "'approximately' - for the range of temperatures likely in the Earth's atmosphere, if I'm remembering Evan et al 2024b correctly. The Evans reference covers this issue more thoroughly and is worthy of mention regarding this.**

  We have added additional discussion of this in several places, most notably the newly added Section 2.4. The relevant portion from that section reads

  "Our method does not rigorously propagate error in the forward model. This error is assumed negligible, however in certain scenarios it may not be. For example, the linear relationship between temperature and channel ratio assumed in C. Cantrall and Matsuo 2021 was derived under typical geophysical conditions, and becomes less accurate when a larger range of conditions must be accounted for (Evans et al. 2024; Eastes et al. 2025). This motivates the inclusion of models of the form $Z = (mT + z_0)^p$, which can reduce errors relative to the linear forward model. However, these errors are still not accounted for. The RMS error in the fit

can be added in quadrature to the error estimates derived from our techniques for an approximation of the total error. Other sources of error in the forward model that are more specific to neutral temperature estimation are described in Section 5."

- **Lines 98-99: "A fully calibrated instrument model is unnecessary in the approach being used for GOLD data products. Only a relative calibration is used in the T retrieval, and that is not a fundamental requirement, especially when using a single emission band." and "Since the prominent advantage is not really an advantage over alternative algorithms that work in count space (such as the GOLD TDISK algorithm), what are other advantages?".**

  This is meant as a comment on band ratio techniques in general. We acknowledge that the GOLD algorithm does not require absolute calibration, just a stable relative calibration across the wavelengths observed. Other advantages, such as the very fast calculation of a band ratio relative to fitting a full spectral model, are also relevant. The text in question has been edited:

  " As noted in Yongliang Zhang, Larry J. Paxton, and Schaefer 2019 and C. Cantrall and Matsuo 2021, two-channel ratio approaches have several advantages. One of them is that the absolute, calibrated intensity is not required. Other advantages include the ease of calculating a ratio compared to fitting a full spectral model and the reduced uncertainty by not requiring knowledge of variation in instrument performance across the whole band, rather just a small section of it (C. Cantrall and Matsuo 2021)."

- **Lines 101-104 Comment 1: "The count rates in the two channels are certainly correlated and derive from the same parent population. If the count rate goes up in one channel it should go down in the other. The authors acknowledge this in the previous paragraph."**

  Due to their derivation from the same physical process, the count rates are related as geophysical variables. When we discuss the independence, we are speaking here of the independence of the count data given the rates. Given this independence, which comes from the fact that the two bands are disjoint, the estimates of the parameters Λ_a and Λ_b are independent of each other. This is what allows us to derive the distribution of their ratio as a BP distribution.

  This construction was not clear in the highlighted text, so we have rewritten this section of the paper to emphasize the independence of the count data given the underlying means:

  "The first step is to model the two-channel intensity ratio as a random process. To motivate the choice of model we first consider the case where the inversion is done pointwise, as in Park et al. 2006 and Jin, S. Zhang, and Wu 2006. Throughout the remainder of the paper, the notation $X \sim F$ means that the random variable X has the distribution F. If we assume that the count data from the channels are independently Poisson distributed with different mean parameters Λ_a and Λ_b, so that

$$\begin{aligned} \{a_i\}_{i=1}^{n_a} | \Lambda_a &\sim Poisson(\Lambda_a) \\ \{b_i\}_{i=1}^{n_b} | \Lambda_b &\sim Poisson(\Lambda_b) \end{aligned} \tag{1}$$

  then the posterior distribution of the ratio $Z = \Lambda_a/\Lambda_b$ given the data is a generalized Beta-Prime distribution $BP(\alpha, \beta, p, q)$, which has a density of the form

$$p_Z(z) = \frac{p}{qB(\alpha,\beta)} \frac{(z/q)^{\alpha p - 1}}{(1 + (z/q)^p)^{\alpha+\beta}} \tag{2}$$

  if $z > 0$, and 0 elsewhere, where $\alpha > 0, \beta > 0, p > 0$ and $q > 0$ are parameters and $B(\alpha, \beta)$ is the Beta function. In this parameterization, α controls the behavior near 0, β near ∞, and p and q are respectively "peakedness" and location parameters (McDonald and Xu 1995). A proof of this is given in Appendix A, and further discussion of this distribution along with applications can be found in McDonald and Richards 1987 and McDonald and Xu 1995."

- **Lines 101-104 Comment 2: "Is the two "channels" being being independent necessary? Unclear that they are independent when both are described by the same vibrational energy equation?"**

  Given the underlying intensities, Poisson distributed counts observed in disjoint bands are independent. In principle, a model for a case where the bands overlapped and the count data were not independent may be possible using a compound Poisson distribution, but is beyond the scope of our investigation.

  The assumed independence of the means follows from the independence of the count data and the fact that we have not used a joint prior to model them. Using a joint prior is, in theory, possible, but it would likely require a different modeling approach and be more computationally expensive. Modeling both together would effectively be the same as doing one estimation rather than two, but with twice as many bins. Since the computational cost scales as $O(n^2)$ even if we avoid taking the eigendecomposition of $\mathbf{K}$, even in a best case scenario this will approximately double the time to estimate the intensities. The remainder of the model also relies on the independence of Λ_a and Λ_b, and so would also have to be redone to accommodate this. Lastly, this procedure would introduce additional bias into the estimates without what we consider to be a clear benefit.

- **Lines 101-104 Comment 3: "then you're assuming that the temperature changes you see have a dependence on the changes previously seen. Would the dependence change with the delay time between current and previous observations?"**

  The model being discussed in this portion of the paper comes from the papers Jin, S. Zhang, and Wu 2006 and Park et al. 2006. In those papers, estimation is done point by point without considering any data at other spatial locations or time points. This is what we mean when we call it a pointwise model.

  This comment does allude to some points that we do not discuss explicitly in the text however. The first point is that the GOLD instrument takes time to scan the disk, so there is also temporal variation between the bins. However, we believe that the time it takes to scan the disk is sufficiently short with respect to typical evolution times to neglect this for the purposes of our study. There are situations where it may be wise to include this, such as during solar eclipses where the motion of the moon causes more rapid changes in the thermospheric state than normal (e.g. Aryal et al. 2025). This is out of the scope of this particular study, but problems of this sort are of interest to us generally.

  The second point is that, while the model as presented in this paper is derived only with reference to spatial correlation, by adjusting the kernel choice, temporal correlation can be included as well. This could allow someone to use this to jointly estimate several disk images at once. For that, the strength of this correlation should decrease with lag time between observations as noted by the reviwer.

  We have modified the first portion of this paragraph to say

  > "The first step is to model the two-channel intensity ratio as a random process. To motivate the choice of model we first consider the case where the inversion is done without consideration to spatial or temporal correlation, as in Park et al. 2006 and Jin, S. Zhang, and Wu 2006."

- **Line 140: Is this derivation actual new? It looks similar to Rogers 2000.**

  This result has a similar form to the optimal estimation technique in Rodgers 2000 (Eqs. 2.27 and 2.30, 4.7 and 4.8 therein). This is in part due to the fact that both methods employ a Gaussian prior distribution for the quantity of interest. However, ours is derived under the assumption of a Poisson likelihood, while the technique in Rodgers 2000 is derived assuming a Gaussian likelihood and is only optimal when that is applicable. While this is usually a good approximation for these data, it is not always suitable. The biggest difference between the two is that the posterior covariance derived in Rodgers 2000 depends on neither the data nor the estimated field, while the posterior covariance in our model depends on both due to the Poisson likelihood used in the derivation.

- **Lines 149-150: "I'm still not clear what this spatial structure is supposed to represent."**

  This represents spatially correlated variations in the Poisson intensities. See our responses to other comments.

- **Equation 12: "How do you get from an approximation to an equality?"**

  The notation $T_{eff} \sim -\frac{z_0}{m} + BP(\alpha_a, \alpha_b, 1, q/m)$ means that the random variable T_{eff} has a Beta-Prime distribution with the given parameters shifted by the quantity $-z_0/m$, and the second line of the equation

  $$\hat{T}_{eff} = -\frac{z_0}{m} + \begin{cases} \frac{q}{m}\frac{\hat{\alpha}_a - 1}{\hat{\alpha}_b + 1}, & \hat{\alpha}_a > 1 \\ 0, & else \end{cases}$$

  gives the formula for the maximum a-posteriori (MAP) estimate of the effective temperature $\hat{T}_{eff}$ given that distribution. This notation denotes approximation in other contexts, and so we have added text to Sec. 2.2 to explicitly state this (see the response to comment 1 on lines 101-104 in the annotated preprint).

- **Line 156: "This may be true mathematically, but how is this relevant to geophysical retrieval of temperature?"**

  We have included this primarily because it is true mathematically, and it gives our model some degree of additional flexibility. However, we do want to point out that, as mentioned before, the linear model is only an approximation derived under "typical" (or at least moderate) geophysical conditions, and this model provides a better one if extreme conditions are of interest. We have addressed this in the newly added Section 2.4, see our response to the comment on lines 94-96.

- **Line 158: "I don't see any discussion of any spectral resolution limits of the technique. The GOLD data has relatively high spectral resolution and it is known that the TDISK algorithm works well with 4 pixel spectral binning of GOLD data. How well does this technique work on lower resolution data?"**

  As long as the necessary bands can be resolved, this algorithm is expected to work. Additionally, we intend this model to be used other applications than just estimation of the neutral temperature, so we considered discussion beyond what is included in C. Cantrall and Matsuo 2021 to be out of scope. For this reason, we did not look at spectral resolution restrictions. However, we have added the following text in Section 2.4:

  > "The first sources of measurement error are due to systematic biases caused by variations in spectral registration and resolution in the photon count data. In the case of neutral temperature estimation, these have been found to have a potentially significant effect on temperature estimation, with errors in spectral registration of 0.1 Å having been shown to lead to errors of 50K in temperature estimation, and resolution errors of 1 Å leading to similar biases (C. Cantrall and Matsuo 2021). We consider the GOLD L1C data to have sufficiently accounted for these sources of uncertainty for the purposes of our exploratory analyses, but they should be considered in any application of these methods and the potential biases quantified as accurately as possible. In the case of neutral temperature estimation it has been shown that including multiple independent measurements of the neutral temperature can decrease these biases (C. Cantrall 2022), but this is outside the scope of our current efforts, and such concurrent estimates may not always be available depending on the domain of application."

- **Line 164: "Do the slant column brightnesses include temperature dependence along the slant path for both the LBH band system and O2 absorption, which is also wavelength dependent? Do the slant column brightnesses allow for variable solar illumination along the slant path?"**

  The slant column brightnesses include the temperature dependence for the LBH band system, but not for the O_2 absorption. However, wavelength dependence on the O_2 absorption is considered in the instrument model. More details on the instrument model are given in Section 1.2 of C. Cantrall 2022, but the model relies on inputting simulated atmospheres into GLOW to calculate the emission rates. These are then used as inputs to the instrument model.

  Variation of the absorption cross section with wavelength is not considered in our inversion since there is only a 1.5% difference between the two channels considered, which is negligible compared to variation due to temperature (C. Cantrall and Matsuo 2021).

Line 166: "For each image?"

Yes, after the 2x2 binning and removal of bins that do not correspond to the disk there are about (usually slightly less than) 1500 locations where the temperature is estimated at local noon. This includes the combined Northern and Southern hemisphere scans.

- **Line 171: "This runtime is rather slow. Is this a single spectrum or a full GOLD disk scan (or pair of disk scans)?"**

  This is for a pair of scans containing observations of both hemispheres. It also includes taking eigendecompositions of the kernel matrices, which can be precomputed in an operational context, and which has a higher computational complexity than the rest of the algorithm. We test the algorithm this way because that is the implementation used in our R package that was developed for this application (LeDuc and Matsuo 2026), but if this is not done the computation time on GOLD data goes down to about 15 seconds or so. This is discussed in Section 5.1, and the time required to run with a precomputed decomposition is added now.

  > "Due to the eigenvalue decomposition necessary to determine $\tilde{\mathbf{K}}$ the algorithm scales asymptotically as $O(n^3)$, which can be mitigated by using a low-rank decomposition, sparse kernel matrices, or precomputing the eigendecomposition of the kernel. Additionally, if the kernel has an explicit Mercer expansion, the need for the eigenvalue decomposition can be avoided altogether. Precomputing the eigendecomposition reduces the runtime for problems involving 1500 bins to about 15 seconds. The rest of the algorithm does not require any matrix inversions and scales approximately as $O(n^2)$."

- **Caption, figure 1: "How is the accuracy quantified?"**

  In the original version of the document, accuracy was assessed by the CRPS. In addition to the CRPS, which is still used in Figure 2 of our work, we have updated Figure 3 of the manuscript to be a plot of percent error (Figure 2 here).

- **Line 200-201: "Why is it expected to perform better when the disk is fully illuminated? Is it the use of cap harmonics? while you've included some examination of cases where the disk is not fully illuminated at early and late local time, the effects are difficult to understand when geophysical variation are also present."**

  There are two reasons. The first is that there is higher average SNR over the disk. The second is that the spherical cap kernel can struggle when the gradient of the emission field is large if the smoothness parameter is set too high.

- **Line 202: "This is a subjective statement. What is the percent difference from truth (error) of the simulated values? How do the errors compare with the estimated uncertainties in the simulation?"**

  We have edited Figure 3 in the mansucript to include information about percent errors (see Figure 2 above) and added a table of 95% HPD interval sizes (Table 1).

- **Caption, figure 2: "I'm not familiar with this measure of performance, so I can't determine if the performance is good or bad. An old fashioned error estimate would help readers assess performance in a way that is more intuitive. The authors have previously published results for a set of data using the two band technique. How do those results compare with results using the technique described in this paper?"**

  We have added this in Figures 1 and 3. We want to keep one plot in terms of CRPS because it is a measure of accuracy of the entire distribution, which is not well-quantified by only the percent errors.

- **Line 210: "Presumably this is at high latitudes. If so, what is the emission angle for these cases?"**

  At local noon, this occurs at northern latitudes with reference observation angle of more than 65 degrees in our simulation.

- **Line 213: "So your approach may deliver lower uncertainties during quiet conditions but obscure some geophysical activity effects?"**

  Yes, we expect that if the covariance kernel is not properly chosen this could happen. We see this in the analysis of data from the Gannon storm, where the SCHA kernel oversmooths the field and obscures the vortices. This is a hazard of spatial methods, especially in cases where small-scale structure is of the greatest importance. One has to be careful to design the method in a way that rejects uncorrelated noise while still keeping small scale structure intact. This is discussed in Section 2.4 and the exact wording is presented in response to the last of the summary comments above.

- **Line 216: "But can't you say what is the primary driver for the cases that you have considered?"**

  We believe the primary driver in this case is the SZA, but we want to avoid being too definitive without having done a more in depth study. We have removed this sentence from the manuscript.

- **Line 219: "This statement needs to be qualified since, as stated above, the GOLD SZA and EMA limits are conservative because the data products are produced operationally. Some of the GOLD data can be safely processed at SZAs larger than 80 deg."**

  See our response to the comments on line 67.

- **Line 223, comment 1: "Is this the only source of noise in the Level 1 count data?"**

  No, it is not the only source of noise (there is also noise from the background and braking radiation, among others as described in McClintock et al. 2020, Evans et al. 2024, etc) but it is a major one and the one that this algorithm is designed to handle. Additional sources of noise in the measurement are discussed in Section 2.4 and Section 5.

- **Line 223, comment 2: "Is it "statistically accurate" since active time results may be biased by the more abundant quiet time results?"**

  We separated the results by K_p index before calculating the reliability diagrams in Figure 4. Additionally, these statistics were calculated using 10 realizations of the retrieved spectra at each time step. This separation should remove the bias you mention and provide sufficient sample support for estimating the coverages during storm time.

- **Caption, figure 3: "It's evident from the plot that the performance of the technique starts to break down above an SZA of 80 deg. This is not surprising. It provides support for the GOLD SZA constraint of 80 deg. Your claim that the technique can be applied reliably for SZA $>$ 80 deg is not supported by the results."**

  Yes, it does start to break down as SZA increases. We only suggest that the method may be able to push the limit up. In the updated version of this plot we show that we can achieve RMS errors of approximately 6% at SZA of 90 degrees. While further research into exactly how much benefit can be gained is required, we believe that this shows promise.

- **Caption, figure 3, comment 2: "Why is this done? The latitude range is effective an EMA constraint. What do the results look like at larger EMA?"**

  This allows variation with observation angle to be slightly decoupled from the effects of SZA, since SZA is most extreme at high latitudes. We believe that this will help uncover any potential biases due solely to observation angle, which appear to be negligible here.

- **Line 238-239: "This is not all that reassuring. The selected v value may be the best uncertainty quantification of the three v values, but that doesn't mean the quantified uncertainty is meaningful."**

  We have added Table 1 to this part of the manuscript, where it appears as Table 1. It shows the average upper and lower tails of the 95% HPD interval derived from each model in units of percent of the estimated temperature for both calm and storm time conditions. The uncertainty derived from the $\nu = 1 + 10^{-8}$ model is over 5%, corresponding to a bit more than 30K given an estimate of 600K in calm conditions. We have also added discussion of this to Section 3.3 of the paper. The last part of that section now reads

ν	Calm conditions	Storm time
0.5	(-7.3,7.6)%	(-7.9,8.1)%
$1 + 10^{-8}$	(-5.4,5.5)%	(-5.8,6.0)%
$2 + 10^{-8}$	(-4.4,4.6)%	(-4.9,5.1)%

Table 1: Average length of 95% HPD intervals, in units of percent of the temperature estimate, for each value of ν in both calm ($K_p < 3$) and storm ($K_p \geq 3$) conditions.

> "The coverage probabilities are estimated from 10 retrievals for each simulated field, and then compiled according to K_p index. The plots in Fig. 4 show the median coverage probability over all applicable times and spatial locations from the simulations, and error bars show the 10th and 90th percentiles of coverages. The mean uncertainties at nominal coverage of 0.95 are shown in Table 1 in units of percent of the estimated temperature. Since the intervals are asymmetric, we provide upper and lower uncertainties. The model with $\nu = 0.5$ provides the widest uncertainty intervals, and we can see from Figure 4 that these intervals are too wide, since the achieved coverage greatly exceeds the nominal coverage in all conditions. Practically, this model will lead one to believe that estimates of T_{eff} are more uncertain than they actually are. In a data assimilation scheme, for example, these inflated errors can lead to errors in forecasts due to their effect on the estimated state, which will be artificially biased toward the model. In calm conditions with a true temperature of 600K, this model would lead to approximate uncertainties of approximately $(-44, +45)$K, while the $\nu = 1 + 10^{-8}$ model would report uncertainties of about $(-32, +33)$K and the $\nu = 2 + 10^{-8}$ model would report uncertainties of about $\pm$27K, which Figure 4 suggests are more accurate representations of the uncertainty due to shot noise. Generally, we see that the $\nu = 1+10^{-8}$ and $\nu = 2+10^{-8}$ models appear to do best at quantifying uncertainty, especially in geomagnetically calm conditions, and their performance degrades in geomagnetically active time periods. The $\nu = 2 + 10^{-8}$ model appears especially affected by this transition. This is due to the fact that setting ν higher makes the retrieved field smoother, so it cannot properly recover the smaller scale structures that appear during a storm."

- **Caption, figure 4: "I don't fully understand this figure."**

  The figure is meant to show how well the posterior distributions for those parameters captures the uncertainty induced by shot noise. Since the method gives us full posterior distributions of the data, we can determine regions where we ought to find the true effective temperature (as calculated using Eq. 2) with probability 0.1, 0.2, etc (the "nominal coverage", x axis) and compare how often the true effective temperature falls within these regions (the "achieved coverage", y axis). If the model accurately describes uncertainty due to shot noise, the nominal and achieved coverages should be the same on average. This plot shows that, under geomagnetically calm conditions, this is the case when $\nu = 1 + 10^{-8}$. However, when $\nu = 0.5$ we overestimate uncertainty due to shot noise, and underestimate it when $\nu = 2 + 10^{-8}$. During storm time, the ability of the model with $\nu = 1 + 10^{-8}$ degrades slightly and it begins to underestimate the uncertainty.

  These probabilities were estimated for 10 realizations of the retrieved spectra at each time step. The posterior sets were calculated at each spatial location for each realization of the field and then we calculated the probability that the ground truth T_{eff} lied within each interval. These probabilities were stored, and the plot shows the 10th, 50th, and 90th percentiles of these probabilities for the geomagnetically calm and storm time conditions.

- **Lines 249-250:"Again, a very subjective statement. My visual inspection arrives at a very different conclusion."**

  We have updated the figure and added a figure showing bias relative to TDISK (figures 3 and 4 here.) This portion of the text has been updated to read

  > "The results of this inversion are shown in Fig. 5. While we obtain similar results to TDISK at high latitudes, as well as on average over SZA and OZA (Figure 6), there is a visible discrepancy

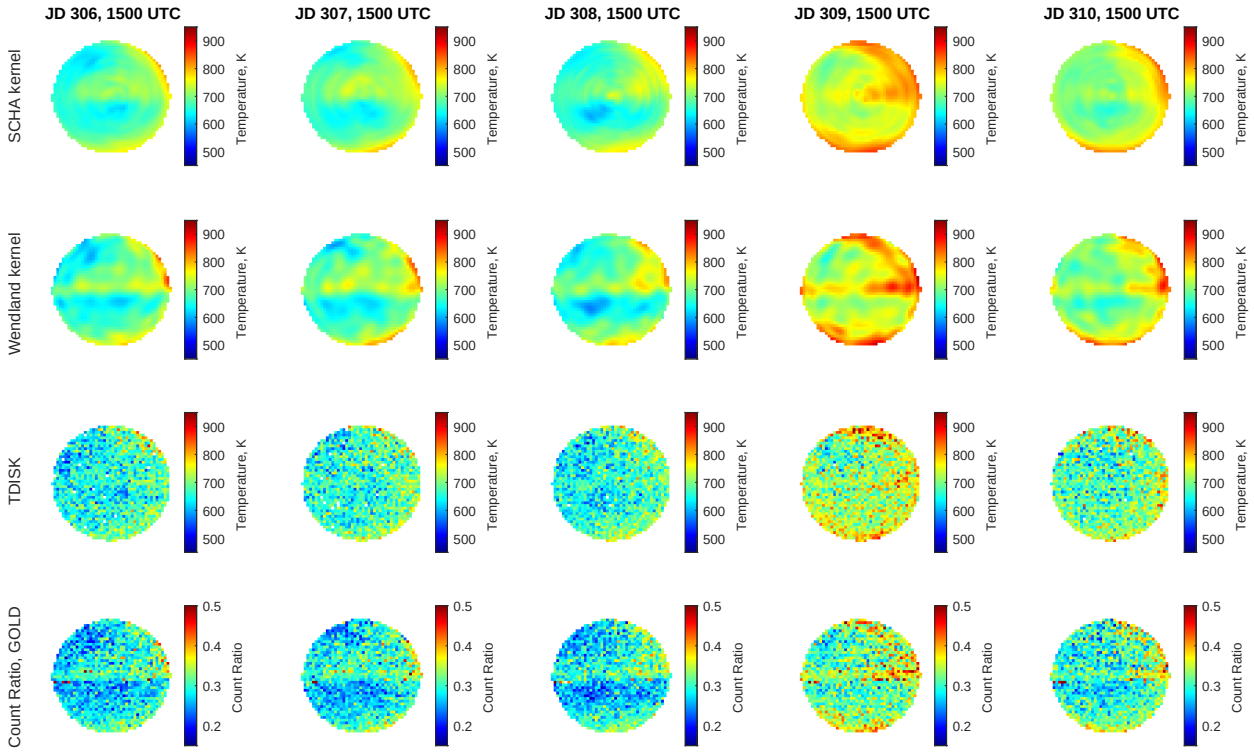

Figure 3: Estimated temperature, local noon, Nov. 2018 storm. Estimates with the SCHA and Wendland kernels are on the first two rows, the third rwo shows TDISK, and the last row shows the count ratio from the GOLD L1C data.

near the equator, corresponding to the lower portion of the scan of the northern hemisphere and the upper portion of the scan of the southern hemisphere. Upon examination of the L1C data, we can see a corresponding enhancement in the count ratio at the same location that is present in the scan of the northern hemisphere, but not the southern hemisphere. This, coupled with the fact that similar artifacts are not seen in the results on simulated data (Fig. 1) suggests that this variation is due to the varying sensitivity of the detector across the slit. Similar artifacts can also be found in earlier versions of the TDISK data (e.g. Figure 7 of C. Cantrall and Matsuo 2021). These artifacts lead to a significant bias in the estimates relative to TDISK at the equator (Figure 6). It is therefore important to have an understanding of the variations in the detector sensitivity and eliminate artifacts in the retrievals."

- **Line 254: "Across scans or along the slit? These are not equivalent".**

  This should say across the slit. Thank you for pointing this out. The passage has been updated to reflect this:

  > "This, coupled with the fact that similar artifacts are not seen in the results on simulated data (Fig. 1) suggests that this variation is due to the varying sensitivity of the detector across the slit."

  Additionally, we have changed the wording at the end of the previous paragraph. This now reads

  > "Since the scans overlap, we average the data in the overlap region, which reduces (but does not completely eliminate) the bias in the retrievals caused by the varying sensitivity of the detector across the slit (Evans et al. 2024)".

- **Line 267: ""somewhat"? they are quite extreme."**

  The term "somewhat" has been removed.

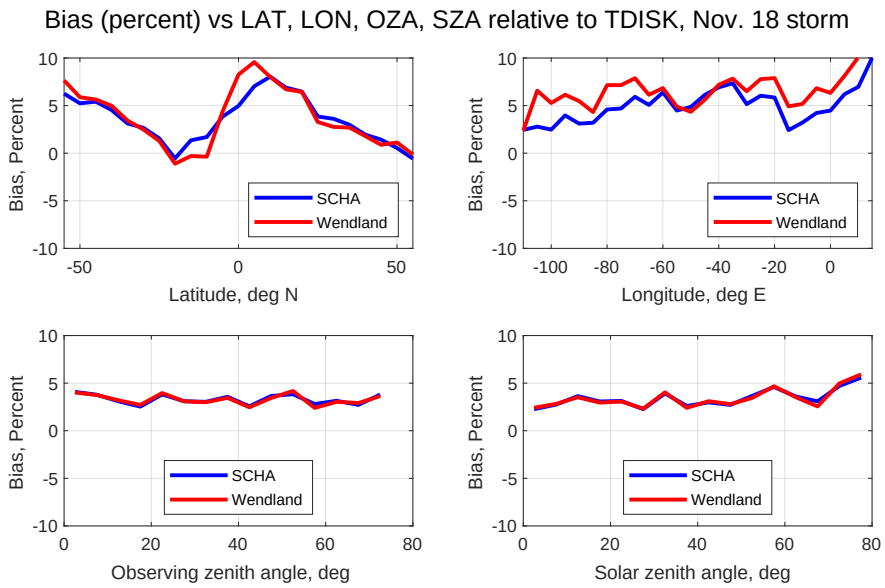

Figure 4: Bias (percent) vs latitude, longitude, OZA, SZA during the November 2018 storm.

- **Line 276: "This statement requires qualification. It's likely that the GOLD TDISK algorithm can also be run for the Gannon storm data without spatial binning. The released data is binned spatially to have consistency across different conditions but that does not mean spatially bin all of the data is necessary."**

  We would not be surprised to see that the current TDISK method can be run on the Gannon data without binning. We believe that this method would be suitable for this sort of estimation without binning under a wide range of conditions, but this has not been verified yet. This comment is better suited for the discussion. We have added a paragraph that discusses some of the advantages of spatial modeling, as well as the possibility of estimation without additional binning which remains to be studied. The following has been added:

  > "One significant advantage of spatial models is the ability to leverage the correlation structure for optimal estimation of the underlying field at unobserved locations, for example via kriging (Stein 1999). By kriging the latent vector $\vec{f}$ onto unobserved locations we are able to derive statistically optimal estimates of the neutral temperature at these locations given the model, as well as covariances between observed and unobserved locations. This would allow estimation of the quantity of interest at arbitrary resolution, and additionally allows simulation of spatial fields given observations via conditional simulation (Chiles and Lantuéjoul 2005). The spatial correlation may also allow estimates of the quantity of interest to be derived without having to perform binning, which is sometimes necessary due to low SNR. The conditions under which such estimation can be done with this model are yet to be studied, but provide a potential avenue for future research."

- **Lines 303-305: "Whether the (1,1) and (2,3) bands have sufficient SNR for temperature retrievals depends on the method used."**

  This is true. The work in C. Cantrall and Matsuo 2021 and C. Cantrall 2022 showed that these bands have too low of SNR for temperature retrievals on their own using the band ratio method, however it was also demonstrated that including the (1,1) band improved neutral temperature estimation in C. Cantrall 2022. We have rewritten the first part of the opening paragraph in Section 5.2 as follows:

  > "The LBH (1,1) and (2,3) bands are not used in this study in part because they are not considered to have sufficient SNR for temperature retrievals on their own using the two channel

ratio method (C. Cantrall and Matsuo 2021; C. Cantrall 2022). However, both of these bands may be beneficial to consider in the future. While the (2,0) band is not isolated from other LBH emissions, with the lower portion of the band overlapping with the (5,2) band, the (1,1) and (2,3) bands are. This means that, in principle, retrievals using these bands are independent of the specified populations of molecular nitrogen, removing a significant source of model specification error C. Cantrall and Matsuo 2021."

- **Line 304: "This begs the question as to how you are dealing with the vibrational populations of the v'=2 and v'=5 bands. Presumably you are using populations derived from GOLD data, but this information should be provided."**

  We are using the populations from Ajello et al. 2020, as was done in the work that inspired this study. We have added this to Section 2.1.

- **Line 306: "using only one of the band, yes, but if both are used then there is a dependence on the populations."**

  This depends on the method used. If we use each to independently estimate T_{eff} via the band ratio method and then weight those measurements by the inverse variance to derive a new estimate, the result still has no dependence on the vibrational populations.

- **Line 309: "contribute"?**

  This change has been made.

- **Line 309-310: "This statement is confusing. Aren't you claiming that your technique has "advantages" over TDISK because it "works" for SZA $>$ 80 deg."**

  This statement is here because the previous work that motivated this work did not include those bands due to their lower SNR, but applying our technique to these bands can improve performance. Using them on their own would not sufficient. However, for example, using them to get individual estimates of T_{eff} and optimally combining them into a single estimate with lower uncertainty is possible. Because of these possibilities, we are claiming that our work has the potential to extend band ratio estimation techniques into low SNR scenarios.

  We have modified the text in question to say

  > "For this reason, extending these techniques in a way to allow estimates of the neutral temperature using the (1,1) and (2,3) bands, which have too low of SNR to get accurate T_{eff} estimates with the band ratio technique (C. Cantrall and Matsuo 2021; C. Cantrall 2022) ..."

- **Line 315: Typo**

  This has been fixed. Thank you for pointing it out.

- **Line 322-323: "Yes, if sufficient measurement of the background are not available. The algorithm used for the GOLD data does include a separate term for the background. GOLD obtains better results using coincident measurements of the background, which the instrument was designed to provide."**

  We use the background corrected measurements in the comparison with GOLD. This has been added to the start of Section 4. We mention this here because this technique should offer a straightforward way to estimate the background jointly with the signal (for example, by using an empirical Bayes method informed by regions where there is no signal) but this has not been done. Uncertainty in the background measurement will necessarily propagate forward into the estimation of the temperature, so being able to account for both is desirable, especially in low-SNR scenarios or other situations in which background contamination becomes significant.

  We have edited this section as follows:

  > "While the GOLD mission L1C data product used mitigates most of the effect of the background (McClintock et al. 2020), an algorithm that can account for it independently is desirable for further improvement as uncertainty in the background measurement will necessarily propagate forward into the estimation of the temperature."

- **Line 328: "That "may" disregard the real statistical properties of the sensor data. This is a function of the count rate and whether it is in the Poisson regime or can be approximated as a normal distribution."**

  The text has been changed.

- **Line 330: "Is the approach only suitable for use with two channels, is it more widely applicable?"**

  The approach can be applied to estimating the temperature using multiple band ratios in concert. This is done, for example, in C. Cantrall 2022 using the (2,0) and (1,1) bands as an extension to the work in C. Cantrall and Matsuo 2021 and was shown to give more accurate estimates of thermospheric neutral temperature. We expect that performing these statistically independent estimates of T_{eff} using each band ratio and combining them optimally (weighted by their inverse posterior variances) will lead to improved estimates of the field. However, there are two major drawbacks. As mentioned in Section 5.2, the posterior distribution of the estimate performed in this way with two different band ratios is related to the Lauricella D function, and we are not aware of any derivation of a posterior distribution in the case where more than 2 of these estimates are used (Pham-Gia and Turkkan 2002). Future work involves investigating suitable approximations to these distributions to allow a method like this to be applied, such as the Delta method (Oehlert 1992) or a moment-matching technique like what we used in Section 2.2. The second drawback is that now we have to solve the optimization problem again. There are speedups that can be implemented, such as precomputing $\tilde{\mathbf{K}}$, which we do not do in this paper, but this still adds potentially significant computational load and likely would further hinder operational use.

  In addition, suppose that, for example, for some problem there are 3 bands A, B, C such that $\frac{A}{B+C} \approx mT + z_0$. We expect that this approach can be extended to such a case by combining the observations B and C into a single dataset, in which case the algorithm is identical to the one presented here, or by considering them separately. This would require solving the optimization problem for 3 Gaussian processes $\vec{f}, \vec{g}$, and $\vec{h}$. Then the distributions of $\vec{f}^2$ and $\vec{g}^2 + \vec{h}^2$ can be approximated in the same manner as we outlined (i.e. by a Laplace Approximation and moment matching to a Gamma distribution) and then the procedure continues from there.

- **Line 334: "From the results for the May 2024 storm, this is doubtful in at least some cases. Some comments, somewhere in the paper, about when your approach would excel and when it would be problematic would be useful to readers."**

  We have added Section 2.4, which discusses potential sources of error in more detail and revamped the plots in Section 4. We have also added plots of bias relative to TDISK (see figures 3, 4, 5, and 6 in this document).

- **Line 339: "Subjective statement"**

  We have modified this line of the manuscript as follows:

  > "We show that the overall features of retrieved column-integrated temperature are generally consistent with the GOLD mission (TDISK v5) data product and demonstrate the potential for providing reliable uncertainty quantification, as well as enabling retrieval of thermospheric neutral temperature in low-SNR scenarios, including at extremely high solar zenith angles."

2 Response to Reviewer 2

2.1 Major comments

Comment: Line 55: The paper treats the ratio of Poisson intensities as the primary geophysical variable. While this is mathematically well justified, the physical meaning of the inferred ratio field should be discussed more explicitly. In particular, it would be helpful to clarify under what conditions the linear relationship $Z \approx mT_{eff} + z_0$ remains valid (e.g., dependence on viewing geometry, vibrational population assumptions, or background contamination). Although it may be further explained in Cantrall and Matsuo, 2021; it would be helpful to provide a brief description in this manuscript to provide sufficient context to the reader.

Response: The linear relationship was derived using the Budzien vibrational-rotational band model (Budzien et al. 2001) and WAM simulated upper atmospheric conditions. It is mostly valid under nominal geophysical conditions and holds independently of viewing geometry or vibrational population assumptions, although the exact values of the vibrational populations can change the exact fit depending on the band ratio being used, and the wavelength that divides the bins may move slightly based on the vibrational populations or temperature (e.g. C. Cantrall 2022; Evans et al. 2024). For older versions of the GOLD data the relationship also changed with observing angle, however this effect is (mostly) small in current versions of the data. We have added some discussion of this in various points in the paper, see the response to Reviewer 1's comments on lines 37 and 56 of the annotated preprint. We also discuss this in the newly added Section 2.4.

Comment: Line 180: The spherical cap harmonic (SCHA) kernel is well motivated, but the practical procedure for selecting the smoothness parameter ν is not fully specified. The manuscript concludes that $\nu = 1 + 10^{-8}$ performs best, but this appears empirical. A clearer description of how this choice would be made in an operational context (e.g., cross-validation, heuristic rules) would improve reproducibility.

Response: For the purpose of this study, we chose values of ν that covered a range of conditions imposed on the estimated intensity. The values $\nu = 0.5, 1 + 10^{-8}, 2 + 10^{-8}$ were chosen because they correspond to fields that are not continuous, continuous, and differentiable respectively. For someone who wishes to use this in an operational context, a cross-validation procedure would be most appropriate. This can be done as a function of K_p index, D_{st}, etc since we expect that the field will become rougher (meaning that ν should decrease) during storm time. This should also include cross-validation with other kernels, like the Wendland kernel. We saw that the Wendland kernel also handled the data quite well, and there are likely many others that will also do a good job.

We already have a brief discussion of this in Appendix C. However, we have edited Section 3.1 to discuss this explicitly:

> "Since the process we are interested in is observed on a section of a spherical shell by GOLD, we choose a kernel given by
>
> $$k(\vec{s}_1, \vec{s}_2) = \sum_{l,m} \frac{1}{l^\nu (1+l)^\nu} V_{lm}(\vec{s}_1) V_{lm}(\vec{s}_2)$$
>
> where the V_{lm} are spherical cap harmonics (SCHAs), which are eigenfunctions of the Laplacian on the spherical cap (G. V. Haines 1985). Then the kernel matrix is given by $\mathbf{K}_{i,j} = k(\vec{s}_i, \vec{s}_j)$ and $\vec{s}_i$ is the reference latitude and longitude of R_i. These functions are well-studied in geophysics, especially in inverse problems that involve estimating the gradient of a process from incomplete measurements, where only a small portion of the globe can be observed at a time (e.g., G. Haines 1988; Richmond and Kamide 1988; Hwang and Chen 1997). Further details on the construction of the spherical cap harmonics are included in G. V. Haines 1985 and Hwang and Chen 1997. The parameter ν, which we call the smoothness parameter, controls the decay of the coefficients in the eigenvector expansion, leading to smoother fields with suppressed high-frequency variation when ν is large. In the context of random function estimation, for example when the data come from a

realization of an unbinned point process, $f(\vec{s})$ is in $H^{\nu}(S)$ the Sobolev space of order ν when $\nu > 1$, and the Sobolev embedding theorem implies that $f(\vec{s})$ has n derivatives if $\nu > 1 + n$ (Hunter and Nachtergaele 2020). For this reason, we focus our analysis on the cases $\nu = 0.5, 1 + 10^{-8}, 2 + 10^{-8}$, with the latter two corresponding to a continuous and differentiable random field, respectively. When applying this method operationally, a cross-validation procedure should be used to select the best value of ν. This can be done as a function of K_p index, D_{st} index, solar activities, or other geophysical variables since we expect that the optimal value of ν will change during disturbed periods. It is also desirable to include other kernels in cross-validation analysis. A more in depth discussion of the SCHA kernel along with a comparison with other common choices is included in Appendix C."

Comment: Section 3.3: The reliability diagrams are a strong point of the paper. However, the text would benefit from a more explicit interpretation of what "overdispersed" means in this context and how this impacts scientific use (e.g., assimilation or model validation). It may also be useful to report typical uncertainty magnitudes in Kelvin for representative conditions.

Response: We have added Table 1, which shows uncertainty magnitudes for the 95% HPD in both storm time and calm conditions, in units of percentage of the estimated temperature. We also report, as an example, representative uncertainties in K for a background temperature of 600K in calm conditions.

In addition to adding the table to Section 3.3, we have rewritten the last paragraph to discuss potential effects of errors in uncertainty quantification on a data assimilation scheme, and removed the term overdispersed.

"The coverage probabilities are estimated from 10 retrievals for each simulated field, and then compiled according to K_p index. The plots in Fig. 4 show the median coverage probability over all applicable times and spatial locations from the simulations, and error bars show the 10th and 90th percentiles of coverages. The mean uncertainties at nominal coverage of 0.95 are shown in Table 1 in units of percent of the estimated temperature. Since the intervals are asymmetric, we provide upper and lower uncertainties. The model with $\nu = 0.5$ provides the widest uncertainty intervals, and we can see from Figure 4 that these intervals are too wide, since the achieved coverage greatly exceeds the nominal coverage in all conditions. Practically, this model leads one to believe that estimates of T_{eff} are more uncertain than they actually are. In a data assimilation scheme, for example, these inflated observation errors can cause the estimated state to artificially biased toward the model, leading to poorly constrained model forecasts. In calm conditions corresponding to a true temperature of 600K, this model would lead to approximate uncertainties of approximately $(-44, +45)$K, while the $\nu = 1 + 10^{-8}$ model would report uncertainties of about $(-32, +33)$K and the $\nu = 2 + 10^{-8}$ model would report uncertainties of about $\pm$27K, which Figure 4 suggests are more accurate representations of the uncertainty due to shot noise. Generally, we see that the $\nu = 1 + 10^{-8}$ and $\nu = 2 + 10^{-8}$ models appear to be the best at quantifying uncertainty, especially in geomagnetically calm conditions, and their performance degrades in geomagnetically active time periods. The $\nu = 2 + 10^{-8}$ model appears especially affected by this transition. This is due to the fact that setting ν higher makes the retrieved field smoother, so it cannot properly recover the structured temperature fields due to geomagnetic disturbances."

Comment: Section 4.2: The comparison with TDISK is mainly qualitative. Quantitative metrics (e.g., mean bias, RMS difference) between the new retrieval and TDISK, in addition to CRPS relative to WAM truth, would make the evaluation more convincing. This is particularly important for the May 2024 storm case.

Response: We have added plots of bias relative to TDISK for both the Nov. 18 and May 24 cases. We have also updated the plots of our estimates. These plots are shown in this document, both in the response to Reviewer 1 and below (Figs 3,4,5, and 6).

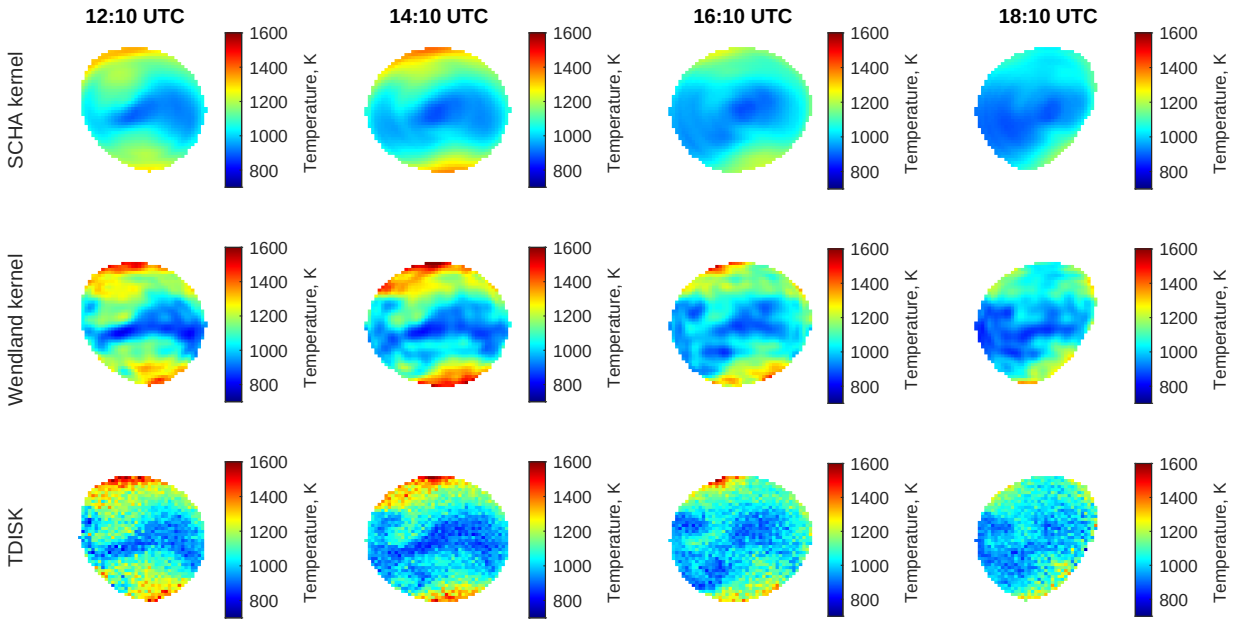

Figure 5: Retrieved temperatures for 12:10-18:10 UTC on 11 May 2024 using both a spherical cap harmonic and Wendland kernel (top two rows) compared with TDISK (bottom row).

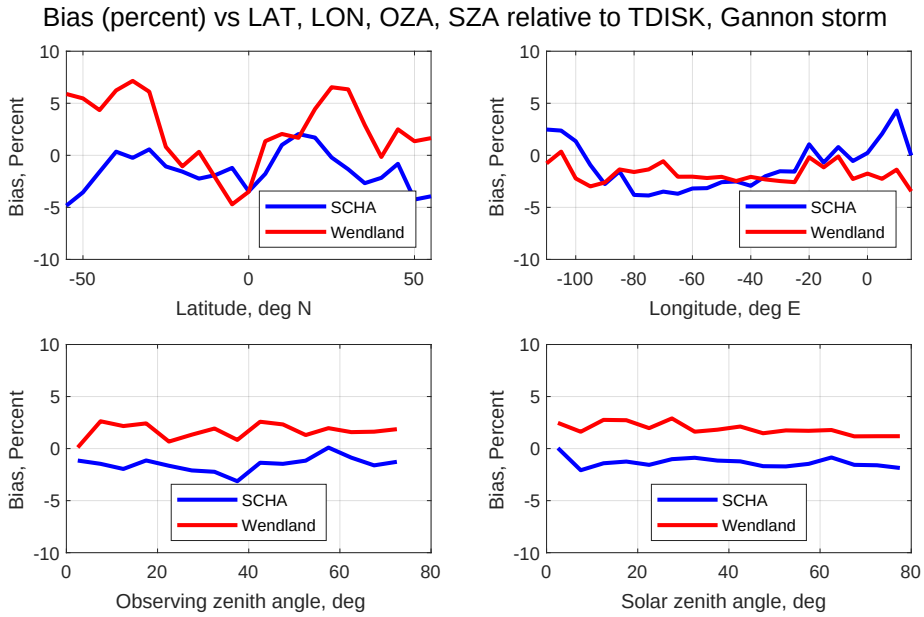

Figure 6: Bias (percent) of estimated temperatures relative to TDISK as a function of latitude, longitude, observing zenith angle, and solar zenith angle during the Gannon storm. Our estimates generally differ from TDISK by less than 5%, well within the TDISK random uncertainty.

2.2 Minor comments

Comment: Several symbols are introduced quickly (e.g., alpha, beta, q, p in the beta-prime distribution). A short table summarizing distribution parameters and their meaning would help readability.

Response: We have added the following text to the paragraph immediately following the introduction of the beta-prime pdf:

> "In this parameterization, α and β are shape parameters that control behavior near 0 and ∞ respectively, and p and q are respectively "peakedness" and location parameters (McDonald and Xu, 1995)."

In addition, to avoid overloading of parameter definitions we are now using the notation $\vec{\psi}$ to refer to kernel coefficients instead of $\vec{\alpha}$.

Comment: Figure 1 would benefit from explicitly labeling the local time and storm phase in the caption. Additionally, I suggest to plot the difference (in K or %) between the reconstructed temperature and the ground truth. It would provide the reader additional information regarding the spatial variability of the error.

Response: The updated figure has the time in UTC in the figure itself, and the local time is in the caption. All plots shown are from local noon at satellite nadir. Figure 1 has been modified to include the estimates of C. Cantrall and Matsuo 2021 (See Figure 1 in this document). We have also updated Figure 3 in the original document, which now is a plot of relative error and includes the Cantrall and Matsuo model (Fig. 2).

Comment: Line 10: There is an equation (E(N(R)) with no number. Also, could you add a general description of variable s (bold s) here (e.g., does it represent a 3-D dimension?). If I understood correctly, it is linked to the bold s used in the kernel, and there is a definition for bold s here.

Response: We have added a number to the equation, thank you for noticing this. The variable $\mathbf{s}$ represents a point in the space S where the point process is defined, in this case a spatial location. We have added an explicit definition of the variable $\mathbf{s}$ in Section 2.2. This part of the manuscript (corresponding to lines 110-114 in the original manuscript) now states

> "A Poisson point process is a random measure defined on a space S, which for our application becomes a spherical cap domain, such that, for all measurable subsets R of S, we have that $N(R)$ the number of points in R is Poisson distributed with
>
> $$\mathbb{E}\left(N(R)\right) = \Lambda(R) = \int_R \lambda(\mathbf{s}) \mathrm{d}w(\mathbf{s}) \tag{3}$$
>
> for some intensity $\lambda(\mathbf{s})$, where $w(\mathbf{s})$ is the ambient measure on S and $\mathbf{s}$ is a point in S, for example a spatial location."

We have additionally replaced the s in Eq. 2.2 representing the slant path distance with r.

Comment: In Eq. (12), the condition $\alpha_a > 1$ for the MAP estimate should be highlighted earlier, since it affects interpretability when counts are very small.

Response: We did not highlight this earlier because across all simulations the minimal value of $\hat{\alpha}_a$ was approximately 12.3. However, in more photon limited applications than we examined in the study this may become an obstacle. We have edited the text in Section 2.3 and added an expression for the posterior mean.

"From Eq. 1 we know that $Z \approx mT_{eff} + z_0$. Since $Z \sim BP(\hat{\alpha}_a, \hat{\alpha}_b, 1, q)$, the temperature $T_{eff} = \frac{1}{m}(Z - z_0)$ has the distribution and MAP and posterior mean estimates

$$\begin{aligned} T_{eff} &\sim -\frac{z_0}{m} + BP\left(\hat{\alpha}_a, \hat{\alpha}_b, 1, \frac{q}{m}\right) \\ \hat{T}_{eff}^{(MAP)} &= -\frac{z_0}{m} + \begin{cases} \frac{q}{m}\frac{\hat{\alpha}_a - 1}{\hat{\alpha}_b + 1}, & \hat{\alpha}_a > 1 \\ 0, & else \end{cases} \\ \hat{T}_{eff}^{(PM)} &= -\frac{z_0}{m} + \begin{cases} \frac{q}{m}\frac{\hat{\alpha}_a}{\hat{\alpha}_b - 1}, & \hat{\alpha}_b > 1 \\ \infty, & else \end{cases}. \end{aligned}$$

The MAP and posterior mean estimates rely on the estimated parameters $\hat{\alpha}_a$ and $\hat{\alpha}_b$ respectively being greater than 1. For all cases examined in the study, the minimal estimated values are considerably larger than this limit, suggesting that it does not present an obstacle for either estimate. This allows estimation of the posterior of the neutral temperature without any sampling, leading to fast inference once the intensities are known. It also allows extension to problems where $Z \approx (mT_{eff} + z_0)^p$ for some exponent p, in which case $T_{eff} \sim -\frac{z_0}{m} + BP\left(\hat{\alpha}_a, \hat{\alpha}_b, p, \frac{1}{m}q^{1/p}\right)$."

Comment: The Laplace approximation introduced in Section 2.2 plays a central role in enabling the computational efficiency of the method, as it leads to a closed-form posterior and avoids the need for sampling-based Bayesian inference. While this connection is implied later in Sections 3 and 5.1, it may be helpful to state this link more explicitly when the approximation is first introduced.

Response: Thank you for pointing this out. The suggestion has been fully incorporated. The paragraph starting at line 137 in the original manuscript has been replaced by

"The posterior distribution here has no closed form, and typically requires a sampling algorithm such as Markov Chain Monte Carlo to estimate it. However, using a Laplace approximation (Rasmussen and Williams 2005) for the posterior we can avoid the sampling, which is computationally intense for problems of this size. Using this approximation we find that the predictive mean is given by $\hat{\vec{f}} = \tilde{\mathbf{K}}\hat{\vec{\psi}}$, where $\hat{\vec{\psi}}$ is the maximum a-posteriori (MAP) estimate, and the predictive covariance of $\vec{f}$ is approximated by

$$\begin{aligned} \hat{\mathbf{\Sigma}} &= \tilde{\mathbf{K}} - \tilde{\mathbf{K}}\left(\mathbf{D}^{-1} + \tilde{\mathbf{K}}\right)^{-1}\tilde{\mathbf{K}}, \\ \mathbf{D} &= \operatorname{diag}\left(\frac{\hat{\psi}_i^2}{2a_i}\right). \end{aligned} \tag{4}$$

See Appendix B for the derivation. Then the distribution of $\hat{\Lambda}(R_i) = \frac{c}{2}\hat{f}_i^2$ given the data is well-approximated by

$$\hat{\Lambda}(R_i) \sim \Gamma(\hat{\alpha}, \hat{\beta}), \quad \text{with} \quad \hat{\alpha} = \frac{\left(\mu^2 + \sigma^2\right)^2}{2\sigma^2\left(2\mu^2 + \sigma^2\right)}, \quad \hat{\beta} = \frac{\mu^2 + \sigma^2}{\sigma^2 c\left(2\mu^2 + \sigma^2\right)} \tag{5}$$

where μ and σ are the posterior mean and standard deviation of the estimate of $\hat{f}_i$ from the Laplace approximation and $\hat{\alpha}$ and $\hat{\beta}$ are the shape and rate parameters of the gamma distribution (analogous to Walder and Bishop 2017 Sect. 4.1.5, which uses the shape/scale parameterization)."

3 Response to Reviewer 3

Comment: The 1st half of the abstract below to introduction. The term "temperature" is not clearly defined in the abstract. Does it refer to a temperature at a fixed altitude, pressure level or height integrated (effective) temperature? There is no definition of specific bands for the ratio calculation. The last statement is on a broader solar zenith angle (SZA). What is the SZA range that this manuscript deals with? Most of the statements in the abstract do not provide specific information.

Response: The abstract has been updated with suggested specific information. The temperature referred to in the abstract is a column-integrated (effective) temperature. The sub-bands associated with the N_2 Lyman-Birge-Hopfield (2,0) transition are used for the ratio calculation. The specific wavelengths used for the ratio calculation are given in Section 2.1. The study places no restrictions directly on the solar zenith angle during estimation. The results shown include estimates of the effective temperature up to a solar zenith angle of 100 deg (Figure 3), thus extending the retrieval to a broader range of solar zenith angles (0-100°) compared to what is provided with existing techniques (0-80°). However, we caution that it comes with an increase in errors, which are discussed in the main text.

Comment: Equation (2). The column-integrated or effective thermospheric temperature depends on wavelength of FUV emissions. Will the method give the same Teff, regardless the wavelength used in the temperature retrieval?

Response: The column-integrated (effective) temperature inferred from LBH bands is largely, but not strictly, wavelength independent. The LBH bands originate from the same electronic transition system of N_2, so their emission generally peaks at similar altitudes and follows a similar vertical structure. The radiative transfer of LBH emissions is relatively straightforward in the upper atmosphere, where absorption is limited and emissions can be approximated as nearly optically thin. Their extinction properties however differ due to wavelength-dependent O_2 absorption, so it shows weak wavelength dependence due to altitude-dependent absorption effects. C. Cantrall 2022 provides more details.

We have added the following to Section 2.1:

> " The LBH volume emission rate depends on neutral temperature, collisional quenching by O_2 and other species, and the local excitation rate from electron impact or solar radiation (Solomon 2017; C. Cantrall and Matsuo 2021). Since LBH emissions derive from the same electronic transitino of N_2 and there is only a weak wavelength dependence in absorption, effective temperature is nearly independent of wavelength in the LBH system. As shown in Figure 4 of Evans et al. 2024, the most weight is assigned to altitudes around 120-200 km. Since these observations are integrated quantities, they cannot be assigned to a specific altitude without a-priori knowledge of the temperature structure (e.g., C. E. Cantrall, Matsuo, and Solomon 2019; C. Cantrall and Matsuo 2021; Evans et al. 2024). Despite this limitation, the NASA GOLD mission data product of column-integrated temperature (TDISK) has broad scientific applications and has been widely used in research (Evans et al. 2024)."

Comment: How do readers use the Teff to refer the thermospheric condition? Since T increases with altitude continuously in the thermosphere, the Teff is expected to be equal to the T at a specific reference altitude. Can this reference altitude be estimated? If there is no such information of the reference altitude, how Teff is useful for understanding the thermospheric dynamics and space weather application?

Response: The effective temperature is used often as a proxy for the temperature near 150-160 km altitude. The thermosphere, especially between 100-300 km altitude, lacks any observations that can constrain its state on global scales (discussed in Jones Jr et al. 2022). There are ground-based remote sensing measurements, but they are sparse in space and time. NASA ICON's limb sounder measurements provide the altitude temperature profiles but they only go up to about 130 km altitude. This motivates our study to maximize the utility of

this unique measurement from GOLD. The paper Evans et al. 2024 gives a comprehensive overview of the GOLD TDISK data product and its scientific applications. The GOLD TDISK data are extremely valuable for upper atmosphere science and space weather applications. GOLD TDISK has been shown to be useful in data assimilation (e.g. He, Yue, and Ren 2021; F. Laskar et al. 2021; Fazlul I Laskar et al. 2022) and has been used to study the effects of geophysical events like eclipses on the thermosphere (Aryal et al. 2025) among others. The scientific utility of TDISK is now mentioned in the beginning of Section 2.1.

Comment: Equation (2). The volume emission rate and O2 observation and Tn.

Response: Yes, the LBH volume emission rate depends on neutral temperature, collisional quenching by O_2 and other species (Solomon 2017), which is noted in the updated manuscript. For the purposes of the simulation study, the LBH volume emission rate is calculated by the Global Airglow Model (GLOW, Solomon 2017) with O_2 and T_n provided by the Whole Atmosphere Model (WAM, Akmaev 2011). Since we have access to these in the simulation we can use them to determine the ground truth T_{eff}.

Comment: The manuscript lists many equations with different approximations (e.g. Eqs 3-12). However, there are no data-based plots to support the approximations.

Response: We believe that this comment refers to the notation $X \sim F$ which is used throughout the paper, and especially in the equations mentioned. This notation means that the random variable X has distribution F. We have edited the first part of Section 2.2 to make this explicit:

> "Throughout the remainder of the paper, the notation $X \sim F$ means that the random variable X has the distribution F."

In addition, we have edited Eq. 3, which gave the PDF of a generalized Beta-prime variable, so that it now is

$$p_Z(z) = \frac{p}{qB(\alpha,\beta)} \frac{(z/q)^{\alpha p-1}}{(1+(z/q)^p)^{\alpha+\beta}} \tag{6}$$

rather than its previous form, which said that it is proportional to $\frac{z^{\alpha p-1}}{(1+(z/q)^p)^{\alpha+\beta}}$. We hope that these changes will sufficiently clarify the notation used in the manuscript.

Comment: Lines 165-166. Could a plot of the simulated Poisson-distributed photon counts included in the manuscript?

Response: Due to the current length of the manuscript we have left those plots out but have reproduced them here for the data from JD 306 (Fig 7). The interior of the green contour is everywhere where the Poisson intensity is less than 50 counts/bin, and the red contour denotes the limit for 25 counts/bin. At UTC 1200/1800 approximately 30% of the viewing area sees intensities less than 50 in the upper band, and about 14 percent sees intensities less than 25. For the lower band, these values are about 15 and 10% respectively. For the purposes of uncertainty quantification, we would want to have at a minimum 25 counts per bin using a Gaussian model since then the mean is 5 standard deviations away from zero. While there are large portions of the viewing area that have large numbers of photon counts, especially at local noon/1500 UTC, even moving just 3 hours off local noon means that there is a large portion of the viewing area where photon counts are greatly reduced. There are also biases that can appear in retrievals even outside photon limited situations. We discuss this in more detail in response to the comment from Reviewer 1 about the content on lines 48-49, and have added a discussion of this to the introduction.

We have also added the following text to Section 3.2:

> "At UTC 12:00 and 18:00 approximately 30 % of the viewing area has less than 50 expected counts in the upper band, and about 14% less than 25. In the lower band, these numbers are about 15% and 10%. In this regime, the assumption that counts are Gaussian breaks down, introducing biases and inhibiting uncertainty quantification."

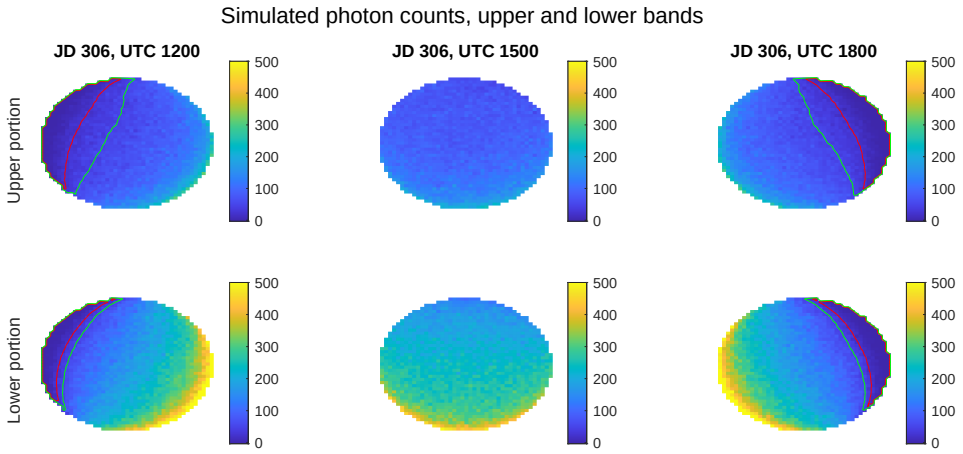

Figure 7: Photon counts, simulated data, JD 306 2018. The interior of the green contuors denotes areas where the expected photon count is less than 50, and the interior of the red contour denotes the area where it is less than 25.

Comment: Lines 149-150. Due to the construction of the intensities as spatial processes, the parameters of the marginal posteriors are calculated using spatial information and thus endow spatial structure on the resulting temperature field. Does this mean that the potential improvement in Teff estimation is associated with a lower spatial resolution? If this is true, the method introduces a bias where localized small scale temperature fluctuations or enhancements (e.g. TAD) is added to the background temperature. Smoothing of TDISK in Figure 5 could result in the retrieved temperature in Figure 5."

Response: The expected improvement in estimation of the temperature comes from the method's ability to improve the rejection of shot noise, which is uncorrelated spatially. By allowing the process to be constructed with an underlying spatial correlation, the estimates of the emission intensities can effectively borrow information from nearby points in an effort to determine the emission intensities that minimize errors over the entire domain rather than just at a given location. However, as you mentioned, care must be taken to ensure that small scale fluctuations in the temperature field are not removed by the algorithm. Using this algorithm in a scientific context will require a careful analysis of the choices of kernel. The kernel chosen determines a class of functions that the estimated field belongs to, and an appropriate choice of kernel should be able to reproduce the behavior seen in practice. Part of the goal of this exercise is in identifying an appropriate kernel that is supported by the data, which we would then expect to capture small scale resolution features.

Comment: Figure 1. The Teff estimation depends on values of γ. How is the γ selected when there is no ground truth data to validate it?

Response: We have added discussion of this to Section 3.1. Since there is no ground truth to compare the value of γ to, one way to pick the value would be to use something like cross-validation on simulated data. The user can select a set of values of γ and fit the model to a spatially subselected set of the data, then krige $\vec{f}$ onto the unobserved data and select the value of γ that gives the best estimates of the held out data based on likelihood, CRPS, RMSE, etc. Another way to do it would be to fit to the full dataset, then select the value of γ that maximizes the likelihood of the data, minimizes the errors, maximizes some information criterion, or some other scoring method. Care should be taken in this estimation, since the value of γ that fits best will likely depend on the kernel used.

A similar method is recommended for selecting values of ν, or other parameters for a given kernel function. This is discussed in more detail in the response to Reviewer 2's comment on line 180 above.

Comment: Figure 1. Is the center of the Teff maps for the nadir looking direction? If this is true, the estimated Teff apparently depends on the look angle and/or solar zenith angle. This

makes the Teff map difficult to use for science investigation.

Response: The center of the maps is the nadir look direction. While there is an effect due to the look direction, the effect is believed to be small relative to the effect of the solar zenith angle (Evans et al. 2024). While the users of these data products do need to be aware of these biases, they can be explicitly accounted for, for example, in data assimilation applications. Another application using the TDISK data, for example, involves investigating the temporal change in TDISK during specific events, as done in Aryal et al. 2025. This allows biases to be effectively dealt with.

Comment: Figure 5. The new T map shows a smooth plot while there many localized fluctuation in the Tdisk map. This is hard to judge that the smooth map is better than the Tdisk as many local temperature fluctuations. Furthermore, it is also better to compare Teff and Tdisk over an entire storm (including quiet time before and after the storm).

Response: These plots have been added as Fig. 5 and Fig. 6 in the updated manuscript (see Figures 3 and 4 in this document). This portion of the manuscript is intended to focus on the effect of unknown biases in the detector, which is why we originally did not include more detail.

Comment: Figure 6. The TDISK in the 2nd column shows a nice vortex in enhanced temperature in the southern hemisphere which nicely reveals the Coriolis effect. However, such a vortex structure is not seen in the Teff.

Response: We made some improvements to this plot, see Figure 5. We believe that the new plot shows that we are able to recover these vortices in TDISK. We have also shown a plot of relative difference between our estimates and TDISK (Figure 6).

Comment: The manuscript extends the work by Clayton and Matsuo (2021) with a different approach to derive column integrated temperature. However, there is no comparison between the Teff.

Response: We have modified several plots in the manuscript to address this. The first one, Figure 1 here, will replace Fig. 1 in the manuscript. We have added a row showing the derived effective temperature using the method of C. Cantrall and Matsuo 2021 (therein, CM21) which shows the effect of shot noise on the estimate.

The second figure, Figure 2, will replace Fig. 3 in the manuscript. It changes the plot to be a plot of percent error rather than CRPS, and we have added the error statistics for the method of C. Cantrall and Matsuo 2021. The plot shows a significant reduction in error across the board from adopting our technique.

Comment: The algorithm is partially validated by using simulated FUV spectra based on WAM neutral profiles. How are the Teff validated using observed GOLD spectra?

Response: Due to the aforementioned relative sparsity of observations, there is no scientifically validated global temperature dataset to compare to other than TDISK. For this reason, we are adding a plot of percent bias between our estimates and TDISK, shown in Figure 6.

We sincerely thank the reviewers for their thoughtful comments and detailed discussion. We believe the updated manuscript has been improved with your feedback. We hope that the revised version meets your expectations and is suitable for publication in Atmospheric Measurement Techniques.

Sincerely,
Matthew LeDuc
University of Colorado Boulder

References

Ajello, Joseph M. et al. (2020). "The UV Spectrum of the Lyman-Birge-Hopfield Band System of N2 Induced by Cascading from Electron Impact". In: *Journal of Geophysical Research: Space Physics*. URL: https://doi.org/10.1029/2019JA027546.

Akmaev, R. A. (2011). "Whole Atmosphere Modeling: Connecting Terrestrial and Space Weather". In: *Reviews of Geophysics* 49.4. URL: https://doi.org/10.1029/2011RG000364.

Aksnes, A. et al. (2006). "Neutral temperatures in the lower thermosphere from N2 Lyman-Birge-Hopfield (LBH) band profiles". In: *Geophys. Res. Lett.* 33. DOI: 10.1029/2006GL026255.

Aryal, Saurav et al. (2025). "GOLD Observations of Thermospheric Neutral Temperature Variability During the 14 October 2023 Annular Solar Eclipse". In: *Geophysical Research Letters* 52.2, e2024GL110676. URL: https://doi.org/10.1029/2024GL110676.

Boersma, C, RH Rubin, and LJ Allamandola (2012). "Spatial analysis of the polycyclic aromatic hydrocarbon features southeast of the Orion Bar". In: *The Astrophysical Journal* 753.2, p. 168. DOI: 10.1088/0004-637X/753/2/168.

Budzien, S. A. et al. (2001). "Thermospheric temperature derived from ARGOS observations of N2 Lyman–Birge–Hopfield emission". In: *Eos Trans. AGU, 82, Spring Meet. Suppl.*

Cantrall, Clayton (2022). "New approaches for quantifying and understanding thermosphere temperature variability from far ultraviolet dayglow". PhD thesis. University of Colorado-Boulder.

Cantrall, Clayton and Tomoko Matsuo (2021). "Deriving column-integrated thermospheric temperature with the N2 Lyman–Birge–Hopfield (2,0) band". In: *Atmospheric Measurement Techniques*. DOI: 10.5194/amt-14-6917-2021.

Cantrall, Clayton E., Tomoko Matsuo, and Stanley C. Solomon (2019). "Upper Atmosphere Radiance Data Assimilation: A Feasibility Study for GOLD Far Ultraviolet Observations". In: *Journal of Geophysical Research: Space Physics* 124.10, pp. 8154–8164. DOI: 10.1029/2019JA026910.

Chiles, Jean-Paul and Christian Lantuéjoul (2005). "Prediction by conditional simulation: models and algorithms". In: *Space, structure and randomness: Contributions in honor of Georges Matheron in the field of geostatistics, random sets and mathematical morphology*. Springer, pp. 39–68.

Coath, Christopher D., Robert C. J. Steele, and W. Fred Lunnon (2013). "Statistical Bias in Isotope Ratios". In: *Journal of Analytical Atomic Spectrometry*. URL: https://doi.org/10.1039/C2JA10205F.

Correira, J. et al. (2021). "Thermospheric Composition and Solar EUV Flux From the Global-Scale Observations of the Limb and Disk (GOLD) Mission". In: *Journal of Geophysical Research: Space Physics* 126.12. URL: https://doi.org/10.1029/2021JA029517.

Doornbos, E and H Klinkrad (2006). "Modelling of space weather effects on satellite drag". In: *Advances in Space Research* 37.6, pp. 1229–1239.

Eastes, Richard et al. (2025). "Remote sensing of lower-middle-thermosphere temperatures using the N2 Lyman–Birge–Hopfield (LBH) bands". In: *Atmospheric Measurement Techniques*. URL: https://doi.org/10.5194/amt-18-921-2025.

Emmert, J.T. (2015). "Thermospheric mass density: A review". In: *Advances in Space Research* 56.5, pp. 773–824. ISSN: 0273-1177. DOI: https://doi.org/10.1016/j.asr.2015.05.038. URL: https://www.sciencedirect.com/science/article/pii/S0273117715003944.

Evans, J. S. et al. (2024). "Disk Images of Neutral Temperature From the Global-Scale Observations of the Limb and Disk (GOLD) Mission". In: *Journal of Geophysical Research: Space Physics* 129.6. DOI: 10.1029/2024JA032424.

Gallardo i Peres, Gerard et al. (2024). "A Generalized Beta Prime Distribution as the Ratio Probability Density Function for Change Detection Between Two SAR Intensity Images With Different Number of Looks". In: *IEEE Transactions on Geoscience and Remote Sensing* 62, pp. 1–14. DOI: 10.1109/TGRS.2024.3369509.

Gneiting, T. and A. Rafferty (2007). "Strictly Proper Scoring Rules, Prediction, and Estimation". In: *Journal of the American Statistical Association*. URL: https://doi.org/10.1198/016214506000001437.

Haines, G. V. (1985). "Spherical cap harmonic analysis". In: *Journal of Geophysical Research: Solid Earth* 90.B3, pp. 2583–2591. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JB090iB03p02583.

— (1988). "Computer programs for spherical cap harmonic analysis of potential and general fields". In: *Computers & Geosciences* 14.4, pp. 413–447. ISSN: 0098-3004. URL: https://www.sciencedirect.com/science/article/pii/0098300488900271.

He, Jianhui, Xinan Yue, and Zhipeng Ren (2021). "The impact of assimilating ionosphere and thermosphere observations on neutral temperature improvement: Observing system simulation experiments using EnKF". In: *Space Weather* 19.10, e2021SW002844.

Humphrey, Philip J., Wenhao Liu, and David A. Buote (2009). "χ^2 and Poissonian data: Biases even in the high-count regime and how to avoid them". In: *The Astrophysical Journal* 693.1. DOI: 10.1088/0004-637X/693/1/822.

Hunter, John K. and Bruno Nachtergaele (2020). *Applied Analysis*. World Scientific.

Hwang, Cheinway and Shin-Kuen Chen (1997). "Fully normalized spherical cap harmonics: application to the analysis of sea-level data from TOPEX/POSEIDON and ERS-1". In: *Geophysical Journal International* 129 (2), pp. 450–460. URL: https://doi.org/10.1111/j.1365-246X.1997.tb01595.x.

Jia, Jingyu and Fan Yi (Aug. 2014). "Atmospheric temperature measurements at altitudes of 5-30km with a double-grating-based pure rotational Raman lidar". In: *Appl. Opt.* 53.24, pp. 5330–5343. DOI: 10.1364/AO.53.005330.

Jin, Y.K., S.N. Zhang, and J.F. Wu (2006). "Hardness Ratio estimation in Low Counting X-Ray Photometry". In: *The Astrophysical Journal*. DOI: 10.1086/508677.

Jones Jr, McArthur et al. (2022). "On the importance of neutral composition and temperature measurements in the 100–200 km altitude region". In: *Frontiers in Astronomy and Space Sciences* 9, p. 1062967.

Krywonos, Andrey et al. (2012). "Remote sensing of neutral temperatures in the Earth's thermosphere using the Lyman-Birge-Hopfield bands of N2: Comparisons with satellite drag data". In: *Journal of Geophysical Research: Space Physics* 117.A9.

Laskar, Fazlul I et al. (2022). "Improving the thermosphere ionosphere in a whole atmosphere model by assimilating GOLD disk temperatures". In: *Journal of Geophysical Research: Space Physics* 127.3, e2021JA030045.

Laskar, FI et al. (2021). "Impact of GOLD retrieved thermospheric temperatures on a whole atmosphere data assimilation model". In: *Journal of Geophysical Research: Space Physics* 126.1, e2020JA028646.

LeDuc, Matthew and Tomoko Matsuo (2026). *PoissonRatioUQ: An R package for band ratio uncertainty quantification*. arXiv: 2602.07165 [stat.CO]. URL: https://arxiv.org/abs/2602.07165.

Leonard, J. M., J. M. Forbes, and G. H. Born (2012). "Impact of tidal density variability on orbital and reentry predictions". In: *Space Weather* 10.12. DOI: https://doi.org/10.1029/2012SW000842.

McClintock, William E. et al. (2020). "Global-Scale Observations of the Limb and Disk Mission Implementation: 2. Observations, Data Pipeline, and Level 1 Data Products". In: *Journal of Geophysical Research: Space Physics* 125.5. DOI: 10.1029/2020JA027809.

McDonald, James B. and Dale O. Richards (1987). "Model selection: some generalized distributions". In: *Comm. Statist. Theory Methods* 16.4, pp. 1049–1074. ISSN: 0361-0926,1532-415X. DOI: 10.1080/03610928708829422. URL: https://doi.org/10.1080/03610928708829422.

McDonald, James B. and Yexiao J. Xu (1995). "A generalization of the beta distribution with applications". In: *Journal of Econometrics*. URL: https://doi.org/10.1016/0304-4076(94)01612-4.

Mehta, Piyush M et al. (2023). "Satellite drag coefficient modeling for thermosphere science and mission operations". In: *Advances in Space Research* 72.12, pp. 5443–5459.

Meier, R. R. (2021). "The Thermospheric Column O/N2 Ratio". In: *Journal of Geophysical Research: Space Physics* 126.3, e2020JA029059. URL: https://doi.org/10.1029/2020JA029059.

Nicolaou, G et al. (2024). "Resolving velocity distribution function parameters from observations with significant Poisson statistical uncertainty". In: *RAS Techniques and Instruments* 3.1, pp. 874–878. DOI: 10.1093/rasti/rzae059.

Oehlert, G. W. (1992). "A Note on the Delta Method". In: *The American Statistician* 46.1, pp. 27–29. URL: https://doi.org/10.1080/00031305.1992.10475842.

Park, Taeyoung et al. (Nov. 2006). "Bayesian Estimation of Hardness Ratios: Modeling and Computations". In: *The Astrophysical Journal* 652.1, p. 610. DOI: 10.1086/507406.

Pham-Gia, T. and N. Turkkan (2002). "Operations on the Generalized F Variables and Applications". In: *Statistics*. DOI: 10.1080/02331880212855.

Rasmussen, Carl Edward and Christopher K. I. Williams (Nov. 2005). *Gaussian Processes for Machine Learning*. The MIT Press. ISBN: 9780262256834. DOI: 10.7551/mitpress/3206.001.0001.

Richmond, A. D. and Y. Kamide (1988). "Mapping electrodynamic features of the high-latitude ionosphere from localized observations: Technique". In: *Journal of Geophysical Research: Space Physics* 93.A6, pp. 5741–5759. DOI: 10.1029/JA093iA06p05741.

Rodgers, Clive D. (2000). *Inverse Methods for Atmospheric Sounding: Theory and Practice*. World Scientific.

Solomon, Stanley C. (2017). "Global modeling of thermospheric airglow in the far ultraviolet". In: *JGR Space Physics* 122, pp. 7834–7848.

Stein, Michael L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics. Springer.

Strickland, D. J., J. S. Evans, and L. J. Paxton (1995). "Satellite remote sensing of thermospheric O/N2 and solar EUV: 1. Theory". In: *Journal of Geophysical Research: Space Physics* 100.A7, pp. 12217–12226. URL: https://doi.org/10.1029/95JA00574.

Walder, Christian J. and Adrian N. Bishop (June 2017). "Fast Bayesian Intensity Estimation for the Permanental Process". In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by Doina Precup and Yee Whye Teh. Vol. 70. Proceedings of Machine Learning Research. PMLR, pp. 3579–3588. URL: https://proceedings.mlr.press/v70/walder17a.html.

Wang, Chen et al. (2024). "Analysis of bright source hardness ratios in the 4 yr Insight-HXMT galactic plane scanning survey catalog". In: *Research in Astronomy and Astrophysics* 24.2, p. 025013. DOI: 10.1088/1674-4527/ad18a3.

Yamada, Shinya et al. (2019). "Poisson vs. Gaussian statistics for sparse X-ray data: Application to the soft X-ray spectrometer". In: *Publications of the Astronomical Society of Japan* 71 (4). DOI: 10.1093/pasj/psz053.

Yin, XiaoHan, JianQi Qin, and Larry J. Paxton (2023). "A new strategy for ionospheric remote sensing using the 130.4/135.6 nm airglow intensity ratios". In: *Earth and Planetary Physics* 7.4, pp. 445–459. ISSN: 2096-3955. DOI: 10.26464/epp2023042.

Zesta, Eftyhia and Cheryl Y Huang (2016). "Satellite orbital drag". In: *Space weather fundamentals*. CRC Press, pp. 329–351.

Zhang, Y. et al. (2004). "O/N2 changes during 1–4 October 2002 storms: IMAGE SI-13 and TIMED/GUVI observations". In: *Journal of Geophysical Research: Space Physics* 109.A10. DOI: https://doi.org/10.1029/2004JA010441. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2004JA010441. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004JA010441.

Zhang, Yongliang, Larry J. Paxton, and Robert K. Schaefer (2019). "Deriving Thermospheric Temperature From Observations by the Global Ultraviolet Imager on the Thermosphere Ionosphere Mesosphere Energetics and Dynamics Satellite". In: *Journal of Geophysical Research: Space Physics* 124.7, pp. 5848–5856.