

A New Approach to Inversion of Multi-spectral Data with Applications to FUV Remote Sensing

Matthew LeDuc¹, Tomoko Matsuo^{2,1}, and William Kleiber¹

¹Department of Applied Mathematics, University of Colorado Boulder, Boulder, CO, 80309

²Ann & H.J. Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder, Boulder, CO, 80309

Correspondence: Matthew LeDuc (matthew.leduc@colorado.edu)

Abstract. Many atmospheric measurement techniques involve inversion of photon counts detected by multi-spectral sensors spanning the X-ray to microwave regions of the electromagnetic spectrum. Although photon counts follow Poisson statistics, commonly used inversion techniques often rely on statistical assumptions that disregard the Poisson nature of the sensor data, limiting the scientific utility of datasets. Motivated to overcome this limiting assumption, this study focuses on retrieval techniques that involve the ratio of counts received in different sub-bands and introduces a new computationally efficient and robust approach to this type of inverse problem that respects the underlying count statistics. The method assumes that the received photon counts in each channel are a realization of a binned point process, allowing the ratio of the channel intensities to be modeled within a hierarchical Bayesian framework. This allows us to directly incorporate correlation between the bins via the prior that is modeled using a permanental process. It further enables more accurate uncertainty quantification without costly sampling procedures common in Bayesian inversion methods. The method is verified and validated on thermospheric column-integrated neutral temperature retrievals from simulated top-of-atmosphere far-ultraviolet (FUV) disk emission data corresponding to 2-8 November 2018, which includes a minor geomagnetic storm. The sub-bands associated with the N₂ Lyman-Birge-Hopfield (2,0) transition are used for the ratio calculation. The method is also demonstrated on calibrated photon counts data from the NASA Global-scale Observations of the Limb and Disk (GOLD) mission from the same time period and from 11 May 2024 during a severe geomagnetic storm. The study demonstrates the method's ability to accurately recover neutral temperature in a variety of geophysical conditions, attesting to its potential to extend the fidelity of neutral temperature retrievals over a broader range of solar zenith angles (0-100°), compared to the current limit of (0-80°) available with existing techniques.

1 Introduction

In recent decades, far-ultraviolet (FUV) remote sensing has played a key role in advancing space physics by providing valuable measurements of space plasma and neutral species. In the upper atmosphere, these emissions are primarily caused by various physical processes such as photoionization, photoexcitation, charge exchange and recombination that depend on thermospheric composition and temperature and ionospheric plasma density. The fact that these FUV emissions are observable from space without contending with significant background emissions makes FUV remote sensing attractive for upper atmosphere research

25 (Paxton et al., 2017). In particular, the Lyman-Birge-Hopfield (LBH) bands, which arise 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 absorbed by O_2 , so their observed emissions 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, can be used to retrieve the thermospheric temperature. Temperature and compositions are fundamental physical properties of the upper atmosphere and are important parameters for space weather research due to their relationship to aerodynamic drag experienced by objects in low Earth orbit (e.g., Aksnes et al., 2006; Doornbos and Klinkrad, 2006; Krywonos et al., 2012; Leonard et al., 2012; Emmert, 2015; Zesta and Huang, 2016; Mehta et al., 2023). Current inverse modeling techniques, however, often make assumptions on the statistical nature of photon count data that are especially problematic when photon counts are low, such as when solar zenith angle (SZA) is large. 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 et al., 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.

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., Cantrall and Matsuo, 2021; Zhang et al., 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 et al., 1995; Zhang et al., 2004; Meier, 2021; Correia et al., 2021) as well as estimates of ionospheric structure (Yin et al. (2023)). Outside of FUV sensing, these techniques are also used in LiDAR profiling (Jia and Yi (2014)), spectrometry (Coath et al. (2013)), x-ray astronomy (Park et al. (2006); Jin et al. (2006); Wang et al. (2024)), astrochemistry (Boersma et al. (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.

Following the work by Jin et al. (2006) and Park et al. (2006), we reformulate the problem of estimating a quantity of interest from ratios of disjoint channels as a hierarchical Bayesian inference problem. Specifically, the photon count data are used to infer the ratio of Poisson means, which we consider the geophysically relevant variable. This formulation allows us to rigorously characterize the effect of shot noise on the estimate of channel means, leading to a better understanding of the uncertainties in the estimates due to the shot noise. 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. (2024b)). Additionally, this new formulation

facilitates explicitly introducing spatial structure into the estimation by treating the received photon count data as a realization
60 of a Poisson point process (Streit, 2010).

Among a variety of methods proposed to estimate the Poisson intensity from point process data, including the work by
Nowak and Kolaczyk (2000); Antoniadis and Bigot (2006); De Oliveira (2013), we choose to model the intensity with a
permanental process model (McCullagh and Møller, 2006), allowing us to use the theory of reproducing kernel Hilbert spaces
to recover the intensity and perform uncertainty quantification (Flaxman et al., 2017; Walder and Bishop, 2017). Once the ratio
65 of Poisson means Z is estimated, the effective neutral temperature is then recovered using the approximate linear relationship
(Cantrall and Matsuo, 2021)

$$Z \approx mT_{eff} + z_0 \tag{1}$$

where T_{eff} is the effective neutral temperature and m and z_0 are slope and intercept parameters that depend on the specific
LBH band and the observation geometry. These parameters are fitted using the Budzien vibrational-rotational model (Budzien
70 et al., 2001). The method is verified and validated with simulated GOLD disk emission photon count data generated using an
instrument simulator (Cantrall, 2022) and the NOAA Whole Atmosphere Model (WAM, Akmaev (2011)) for 2-8 November
2018. The method is furthermore demonstrated on calibrated, geolocated photon count data collected by GOLD for the same
time period, as well as data from the Gannon storm in May 2024 (Grandin et al., 2024). 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
75 during geomagnetically quiet and severely disrupted periods, and attest to the potential for extending temperature retrievals
with uncertainty quantification based on the LBH band emission to higher SZAs than currently made available. In addition, the
method yields robust results with a computational cost that is feasible as an operational algorithm. Specifically, it achieves a full
characterization of the posterior distribution of the neutral temperature without computationally intense sampling procedures
that are common in Bayesian inversion methods. The method is implemented in an R package (R Core Team, 2025) that is
80 publicly available (LeDuc and Matsuo, 2026).

In Sect. 2 we introduce the statistical modeling framework for temperature inference. In Sect. 3 and Sect. 4, we illustrate the
new inversion procedures on simulated and actual GOLD disk emission data, respectively. Section 5 contains discussion of the
computational performance, as well as directions for future research. Lastly, in Sect. 6 we provide the conclusion.

2 Hierarchical Bayesian Inverse Method

85 In this section, we derive the inversion procedure as a hierarchical Bayesian method and discuss potential sources of error in the
procedure. Since this work is motivated by estimation of thermospheric column-integrated neutral temperature from channel
ratio data, we focus on that particular application. However, we believe that this method is more generally applicable due to
the other areas where channel ratio data are used in inverse problems described previously.

2.1 Building on Previous Work: The Forward Model

90 We seek to estimate the column-integrated thermospheric temperature from the top-of-atmosphere LBH band emission sensed by the GOLD instrument. Specifically, photon counts from radiometrically calibrated, geolocated LIC GOLD science data products are considered (McClintock et al., 2020). The column-integrated (effective) temperature T_{eff} can be given as a function of the wavelength λ as

$$T_{eff}(\lambda) = \frac{\int_0^\infty V(r, \lambda) \exp(-\tau(r, \lambda)) T_n(r) dr}{\int_0^\infty V(r, \lambda) \exp(-\tau(r, \lambda)) dr} \quad (2)$$

95 where r (cm) is the slant path distance, V (photons \times s⁻¹ cm⁻³) is the LBH volume emission rate, τ the optical depth, and T_{eff} (K) the neutral temperature (Evans et al., 2024b). This represents a weighted average of the neutral temperature along the line of sight, with the weight determined by the product $V(r, \lambda) \exp(-\tau(r, \lambda))$. The LBH volume emission rate depends on neutral temperature, collisional quenching by O₂ and other species, and the local excitation rate from electron impact or solar radiation (Solomon, 2017; Cantrall and Matsuo, 2021). Since LBH emissions derive from the same electronic transition of N₂ 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. (2024b), 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., Cantrall et al., 2019; Cantrall and Matsuo, 2021; Evans et al., 2024b). Despite this limitation, the NASA GOLD mission data product of column-integrated temperature (TDISK) has broad scientific applications and has
105 been widely used in research (Evans et al., 2024b).

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 N₂ and neutral temperature. For the purposes of the study, we use the populations in Ajello et al. (2020), which are derived from GOLD data. The modeled spectra are convolved with the spectral point-spread function of the instrument to approximate the intensities as seen by the detector. Principal
110 component analysis of simulated GOLD data has shown that under typical geophysical conditions the intensity in the upper portion of the LBH (2,0) band, from wavelength 138.56-139.2 nm, is positively correlated with temperature, while the intensity in the lower portion from 138-138.56 nm is negatively correlated with temperature (Cantrall and Matsuo, 2021). A similar structure is shown to hold in the (1,1) and (2,3) bands as well (Cantrall et al., 2019). So, as the column-integrated temperature increases, the observed photon counts in the upper portion of the band increases while the observed counts in the lower portion
115 of the band decrease. This leads to the approximately linear relationship between the ratio of the long wavelength to short wavelength portions of the band and the neutral temperature (Cantrall and Matsuo, 2021), which is used in the study. As noted in Zhang et al. (2019) and 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
120 across the whole band, rather just a small section of it (Cantrall and Matsuo (2021)).

2.2 Modeling the Intensity Ratio

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 et al. (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

$$\begin{aligned} \{a_i\}_{i=1}^{n_a} | \Lambda_a &\sim \text{Poisson}(\Lambda_a) \\ \{b_i\}_{i=1}^{n_b} | \Lambda_b &\sim \text{Poisson}(\Lambda_b) \end{aligned} \quad (3)$$

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

$$p_Z(z) = \frac{p}{qB(\alpha, \beta)} \frac{(z/q)^{\alpha p - 1}}{(1 + (z/q)^p)^{\alpha + \beta}} \quad (4)$$

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 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). 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).

We extend this to incorporate spatial information by utilizing Poisson point processes (Streit, 2010). 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

$$\mathbb{E}(N(R)) = \Lambda(R) = \int_R \lambda(\mathbf{s}) dw(\mathbf{s}) \quad (5)$$

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. These are extended by the Cox process model, where $\lambda(\mathbf{s})$ is itself a random function (Cox, 1955). We are interested in a special kind of Cox process known as the permanental process, where the assumption is that $\lambda(\mathbf{s}) = \frac{1}{2} c f(\mathbf{s})^2$, with $f(\mathbf{s})$ a Gaussian process (Møller, 2005; McCullagh and Møller, 2006). The estimation of $\lambda(\mathbf{s})$ then becomes a problem of estimating the latent process $f(\mathbf{s})$. In our specific case, the underlying point process has been binned, losing knowledge of the emission latitude and longitude. For this reason, instead of estimating $\lambda(\mathbf{s})$ we estimate $\Lambda(R_i)$ for each bin. However, our model can be generalized to the model of Walder and Bishop (2017) if, instead of binned data, the raw point process data are available.

Consider a region S divided into disjoint bins R_i such that $S = \bigcup_{i=1}^d R_i$. Given a point process on S , our data become pairs of counts a_i in each bin R_i , $i = 1, \dots, d$. We assume each a_i is Poisson distributed with intensity $\Lambda(R_i) = \frac{c}{2} f_i^2$ and attempt to recover the latent vector $\mathbf{f} = (f_1, f_2, \dots, f_d)^T$. The log-likelihood of \mathbf{f} is given by

$$\ell(\mathbf{f} | \{a_i\}_{i=1}^d) = \sum_{i=1}^d a_i \log\left(\frac{c}{2} f_i^2\right) - \frac{c}{2} \langle \mathbf{f}, \mathbf{f} \rangle. \quad (6)$$

where $\langle \cdot, \cdot \rangle$ is the standard inner product on \mathbb{R}^d . Adding the prior $\mathbf{f} \sim \mathcal{N}(0, \frac{1}{\gamma}\mathbf{K})$ for some positive definite Hermitian matrix \mathbf{K} , we have that the log-posterior is

$$\ell(\mathbf{f}|\{a_i\}_{i=1}^d) = \sum_{i=1}^d a_i \log\left(\frac{c}{2}f_i^2\right) - \frac{c}{2}\langle \mathbf{f}, \mathbf{f} \rangle - \frac{\gamma}{2}\langle \mathbf{f}, \mathbf{K}^{-1}\mathbf{f} \rangle. \quad (7)$$

Since \mathbf{K} is a covariance matrix, it can be written as $\mathbf{K} = \Phi\mathbf{H}\Phi^T$ with Φ_{ij} the value of the j^{th} eigenvector at location i and $\mathbf{H} = \text{diag}(\eta_i)$, with η_i the eigenvalues of \mathbf{K} . Then the inner products can be combined to generate an equivalent norm as

$$\frac{c}{2}\langle \mathbf{f}, \mathbf{f} \rangle + \frac{\gamma}{2}\langle \mathbf{f}, \mathbf{K}^{-1}\mathbf{f} \rangle = \frac{1}{2}\langle \mathbf{f}, (c\mathbf{I} + \gamma\mathbf{K}^{-1})\mathbf{f} \rangle. \quad (8)$$

This new norm defines a norm on an equivalent kernel space with kernel $\tilde{\mathbf{K}}$ which we can write as

$$\begin{aligned} \tilde{\mathbf{K}} &= \Phi\tilde{\mathbf{H}}\Phi^T, \\ \tilde{\mathbf{H}} &= \text{diag}\left(\frac{\eta_i}{c\eta_i + \gamma}\right). \end{aligned} \quad (9)$$

As in the infinite-dimensional case, we can say that the solution has the form $\hat{\mathbf{f}} = \tilde{\mathbf{K}}\hat{\boldsymbol{\psi}}$ for some coefficients $\boldsymbol{\psi}^1$. This leads to the likelihood

$$\ell(\boldsymbol{\psi}|\{a_i\}_{i=1}^d) = \sum_{i=1}^d a_i \log\left(\frac{c}{2}\left(\tilde{\mathbf{K}}\boldsymbol{\psi}\right)_i^2\right) - \frac{1}{2}\boldsymbol{\psi}^T\tilde{\mathbf{K}}\boldsymbol{\psi} \quad (10)$$

with $\left(\tilde{\mathbf{K}}\boldsymbol{\psi}\right)_i$ the i^{th} entry of $\tilde{\mathbf{K}}\boldsymbol{\psi}$. This is the form of the log-posterior distribution of $\boldsymbol{\psi}$ with log-likelihood function given by the first term in Eq. (10) and prior distribution $\boldsymbol{\psi} \sim \mathcal{N}\left(0, \tilde{\mathbf{K}}^{-1}\right)$.

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 $\hat{\mathbf{f}} = \tilde{\mathbf{K}}\hat{\boldsymbol{\psi}}$, where $\hat{\boldsymbol{\psi}}$ is the maximum a-posteriori (MAP) estimate, and the predictive covariance of \mathbf{f} is approximated by

$$\begin{aligned} \hat{\boldsymbol{\Sigma}} &= \tilde{\mathbf{K}} - \tilde{\mathbf{K}}\left(\mathbf{D}^{-1} + \tilde{\mathbf{K}}\right)^{-1}\tilde{\mathbf{K}}, \\ \mathbf{D} &= \text{diag}\left(\frac{\hat{\psi}_i^2}{2a_i}\right). \end{aligned} \quad (11)$$

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

$$\hat{\Lambda}(R_i) \sim \Gamma(\hat{\alpha}, \hat{\beta}), \quad \text{with} \quad \hat{\alpha} = \frac{(\mu^2 + \sigma^2)^2}{2\sigma^2(2\mu^2 + \sigma^2)}, \quad \hat{\beta} = \frac{\mu^2 + \sigma^2}{\sigma^2 c(2\mu^2 + \sigma^2)} \quad (12)$$

¹This is not strictly necessary for the problem we are solving of recovering \mathbf{f} in a binned point process, however it is critical for the feasibility of the method in recovering a random function, which is the case with unbinned point process data. This allows the estimation to be reduced to a finite dimensional problem.

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).

175 With the above derivations, we can proceed as in the previous section to determine the parameters of the distribution of Z at a given spatial location. Once we know that $\hat{\Lambda}_a \sim \Gamma(\hat{\alpha}_a, \hat{\beta}_a)$ and $\hat{\Lambda}_b \sim \Gamma(\hat{\alpha}_b, \hat{\beta}_b)$ we have

$$Z = \frac{\hat{\Lambda}_a}{\hat{\Lambda}_b} \sim BP(\hat{\alpha}_a, \hat{\alpha}_b, 1, q), \quad (13)$$

where $q = \hat{\beta}_b/\hat{\beta}_a$, analogous to the pointwise model in Eq. (4). 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
180 resulting temperature field.

2.3 Modeling the Temperature

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

$$\begin{aligned} T_{eff} &\sim -\frac{z_0}{m} + BP\left(\hat{\alpha}_a, \hat{\alpha}_b, 1, \frac{q}{m}\right) \\ \hat{T}_{eff}^{(MAP)} &= -\frac{z_0}{m} + \begin{cases} \frac{q}{m} \frac{\hat{\alpha}_a - 1}{\hat{\alpha}_b + 1}, & \hat{\alpha}_a > 1 \\ 0, & else \end{cases} \\ \hat{T}_{eff}^{(PM)} &= -\frac{z_0}{m} + \begin{cases} \frac{q}{m} \frac{\hat{\alpha}_a}{\hat{\alpha}_b - 1}, & \hat{\alpha}_b > 1 \\ \infty, & else \end{cases}. \end{aligned} \quad (14)$$

185 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 $T_{eff} \sim -\frac{z_0}{m} + BP(\hat{\alpha}_a, \hat{\alpha}_b, p, \frac{1}{m}q^{1/p})$.

190 2.4 Sources of Error

Due to the construction following that of Cantrall and Matsuo (2021), this method inherits many of the sources of error described therein, and additionally in Evans et al. (2024b) and Eastes et al. (2025). We divide these errors into two types: measurement error, such as variations in detector performance, and model misspecification error in addition to the effects of shot noise which the model is designed to account for.

195 **2.4.1 Measurement Error**

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 \AA having been shown to lead to errors of 50K in temperature estimation, and resolution errors of 1 \AA leading to similar biases (Cantrall and Matsuo, 2021). We consider the
200 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 (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.

205 One source of error that is especially impactful in this estimation scheme is variation in the sensitivity of the instrument. We will see some consequences of this in Section 4, but due to the construction of this as a spatial model such spatially correlated biases are not removed, and can be made worse by spreading into unaffected areas of the detector. Variations in instrument sensitivity both spatially and in the frequency domain should be well-understood and accounted for before applying this technique.

210 **2.4.2 Model Misspecification Error**

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
215 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.

220 Our method, as currently described, contains no way to estimate the background contamination, including from sources such as braking radiation. This contamination is removed by GOLD L1C data, and so is considered negligible for this application so long as the high background flag in the GOLD L1C data is not activated. Discussion of how to account for this is left to Section 5.

Our method does not rigorously propagate error in the forward model. This error is assumed negligible, however in certain
225 scenarios it may not be. For example, the linear relationship between temperature and channel ratio assumed in 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., 2024b; 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. These include the variation of the cutoff wavelength with temperature (Evans et al., 2024b) and dependence on vibrational populations (Cantrall and Matsuo, 2021; Eastes et al., 2025).

3 Application to Simulated Data

To verify and validate the new approach, we apply the method developed in Sect. 2 to simulated GOLD LBH disk emission data generated using the NOAA National Weather Service Whole Atmosphere Model (WAM, Akmaev (2011)). The simulation study corresponds to Nov 2-8 2018. Synthetic GOLD emission data are generated as follows. WAM simulations conducted with realistic solar and magnetosphere forcing are used as input to the Global Airglow Model (GLOW, Solomon (2017)) which calculates the volume emission rates. These volume emission rates are passed into a GOLD instrument simulator developed in Cantrall (2022), which returns the slant column brightness (in units of counts per Angstrom) at each location on the detector. These brightnesses are convolved with the GOLD instrument point spread function to account for instrument effects and then used to simulate Poisson-distributed photon counts to include the effects of shot noise. Finally, these data are binned spatially, as done in the GOLD mission TDISK algorithm, to 250x250 km² resolution at satellite nadir, corresponding to about 1500 bins (Evans et al., 2024b). The emission location is assigned to the middle of the combined spatial bins. The column-integrated effective temperatures directly calculated from WAM temperature fields using Eq. (2) in the same resolution serves as the ground truth.

Each full-disk retrieval takes approximately 40 seconds and uses less than 2GB of memory on a desktop computer with 16 GB RAM and an Intel i7 CPU. Most of the computing time is spent solving the optimization problem in Eq. (10) twice, although for large problems generating the equivalent kernel matrix is also costly. Because we have a closed form approximation to the posterior given by Eq. (14), we are able to avoid costly sampling algorithms that are common in other Bayesian retrievals, making the computations relatively efficient on these datasets. See Section 5.1 for more discussion on computational performance.

3.1 Choice of Kernel

Since the process we are interested in is observed on a section of a spherical shell by GOLD, we choose a kernel given by

$$k(\mathbf{s}_1, \mathbf{s}_2) = \sum_{l,m} \frac{1}{l^\nu(1+l)^\nu} V_{lm}(\mathbf{s}_1) V_{lm}(\mathbf{s}_2) \quad (15)$$

where the V_{lm} are spherical cap harmonics (SCHAs), which are eigenfunctions of the Laplacian on the spherical cap (Haines, 1985). Then the kernel matrix is given by $\mathbf{K}_{i,j} = k(\mathbf{s}_i, \mathbf{s}_j)$ and \mathbf{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., Haines, 1988; Richmond and Kamide,

1988; Hwang and Chen, 1997). Further details on the construction of the spherical cap harmonics are included in 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, $f(\mathbf{s})$ is in $H^\nu(S)$ the Sobolev space of order ν when $\nu > 1$, and the Sobolev embedding theorem implies that $f(\mathbf{s})$ 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. For the purposes of the simulation study we chose $\gamma = 1$. Since there is no ground truth to compare to, this should be chosen via a cross-validation procedure or by kriging the latent vector \mathbf{f} onto held out spatial locations and selecting the value that minimizes some error measure.

3.2 Model Evaluation

We evaluate the model performance primarily using the continuous rank probability score (CRPS) and RMS error. The CRPS is given by

$$CRPS(F|y) = \int_{\mathbb{R}} (F(x) - H(x - y))^2 dx \quad (16)$$

where F is the predictive cumulative distribution function (CDF) of the neutral temperature, y the true neutral temperature, and H the Heaviside step function. This is a strictly proper scoring rule, meaning its unique minimizer is the deterministic CDF $H(x - y)$ (Gneiting and Raftery, 2007). Since it incorporates the entire distribution rather than just a point estimate it allows us to assess the quality of the posterior beyond the error in estimation. CRPS is negatively-oriented in that smaller values indicate superior predictive models.

As a first pass, we test how our model performs in estimating the neutral temperature at 15:00 UTC, or local noon at satellite nadir. At this time the largest portion of the instrument's viewing area is sunlit, meaning that the algorithm is expected to perform best in this situation. The retrieved fields from JD 306 - JD 310 are shown in Fig. 1, along with the estimates derived using the method of Cantrall and Matsuo (2021) (top row, CM21) and the "true" field from WAM simulations (bottom row). The algorithm closely tracks the true temperature field for all cases, with variations in ν leading to smoother or rougher estimates of the underlying field. It also appears to be much more robust to shot noise than the algorithm of Cantrall and Matsuo (2021).

In addition to performance at local noon, we are interested in the performance of our algorithm in low-light conditions. 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

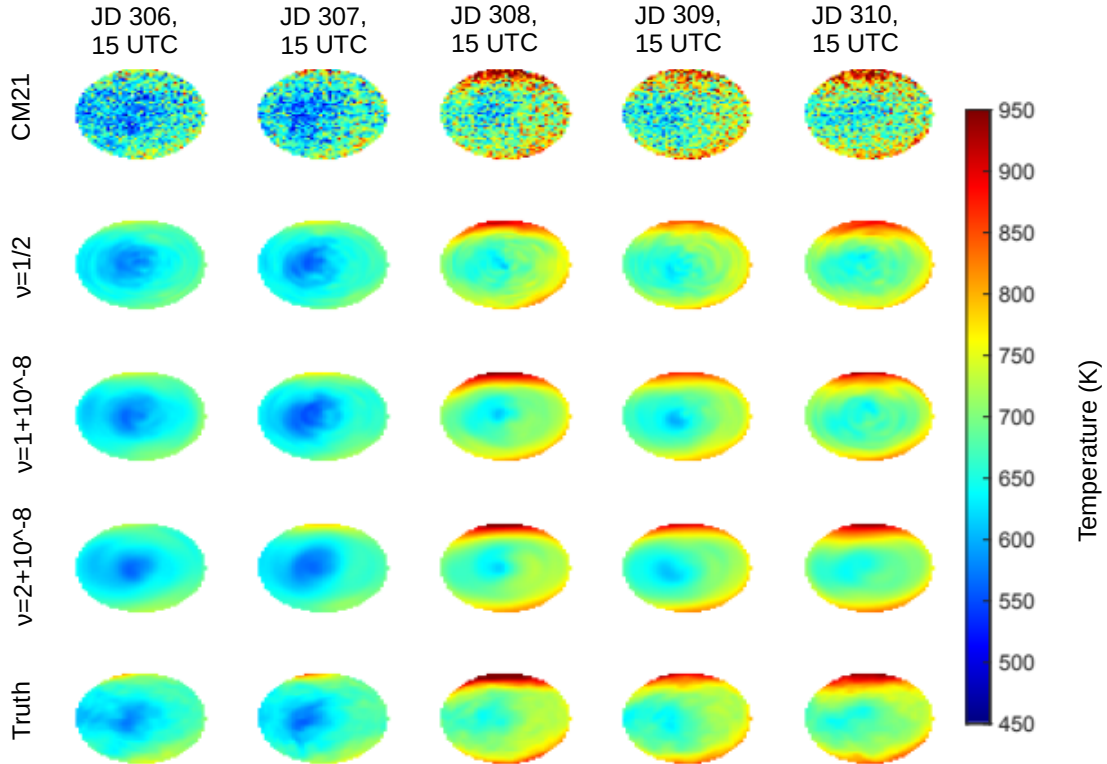


Figure 1. Estimated neutral temperature (first 4 rows) vs simulated field (last row), JD 306-310 at 15:00 UTC (Noon satellite local time) for the model of Cantrall and Matsuo (2021) (top row, CM21) and parameters $\nu = 0.5$ (2nd row), $1 + 10^{-8}$ (3rd row), and $2 + 10^{-8}$ (4th row). The algorithm is able to generate accurate estimates of the neutral temperature regardless of the smoothness parameter selected (see figures 2 and 3), although the parameter has a noticeable effect on the smoothness of the retrieved field.

(Evans et al. (2024b)). 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. To investigate this, we processed data from 12:00 and 18:00 UTC as well. In this situation, a portion of the viewing area has a solar zenith angle higher than the nominal observation boundary of 80° used by the GOLD mission, however there is still a large portion of the viewing area that is sunlit. At UTC
295 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% respectively. In this regime, the assumption that counts are Gaussian breaks down, introducing biases and inhibiting uncertainty quantification. Additionally, there are portions of the viewing area where the solar zenith angle exceeds 80° at local noon. These results for the average CRPS at each time are shown in Fig. 2 along with the K_p index. What we see is that the CRPS of the models for $\nu = 0.5$ and $\nu = 1 + 10^{-8}$ are
300 roughly constant with time and show only a slight increase with the K_p index, never exceeding a CRPS of 15 K. However, the CRPS of the model with $\nu = 2 + 10^{-8}$ increases significantly, especially in the afternoon, and exceeds 20 K at 18:00 UTC on

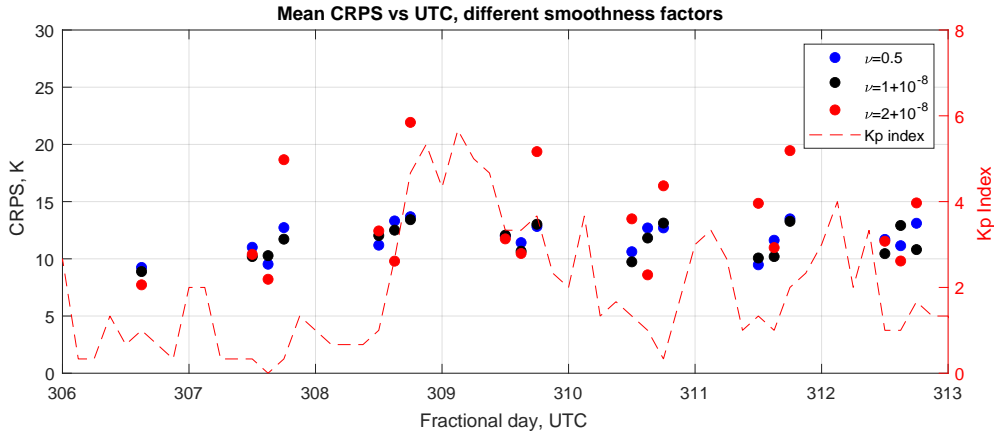


Figure 2. CRPS vs UTC for different kernel smoothness parameters, along with the K_p index from Matzka et al. (2021). The performance of the model depends strongly on local time and decreases slightly during storm time, especially in the case $\nu = 2 + 10^{-8}$.

JD 308. This is because the onset of even this relatively minor storm generates structure in the field that the model is unable to capture.

We have plotted the RMS error in estimation over latitude, longitude, OZA, and SZA in Figure 3. The permanental process
 305 model reduces error by half or more relative to the model of Cantrall and Matsuo (2021) and maintains errors less than 6 % up to nearly 90° SZA. This suggests that the method may be able to extend the scientific utility of FUV disk emission data by allowing analysis of data collected from areas with high SZA.

3.3 Uncertainty Quantification

Our model provides a full posterior distribution of the intensity ratio, and therefore the temperature. In this section, we examine
 310 how accurately our model captures uncertainty. This exercise is important to ensure that the model handles uncertainty induced by shot noise properly and that the estimates of the neutral temperature are reliable. We quantify the uncertainty captured by the model by determining the highest posterior density sets with coverage rate r (r -HPD set) for varying values of r . The r -HPD set is the set I_r such that $\int_{I_r} f(x)dx = r$ and that $\forall (x \in I_r, y \in I_r^C), f(x) \geq f(y)$ (Amaral Turkman et al., 2019). If the posterior distribution accurately quantifies the uncertainty, then the true parameter lies in the r -HPD set with probability
 315 r . To determine how well the posterior HPD matches the nominal coverage for varying coverage parameters r , we compare the probability that the true value is within I_r for varying values of r from 0.05 to 0.95. These results are shown in Fig. 4 for the SCHA kernel with $\nu = 0.5, 1 + 10^{-8}, 2 + 10^{-8}$, and have been divided into times where $K_p < 3$, indicating calmer conditions, and $K_p \geq 3$, indicating higher levels of geomagnetic disturbance.

The coverage probabilities are estimated from 10 retrievals for each simulated field, and then compiled according to K_p
 320 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

Percent error vs LAT, LON, OZA, SZA

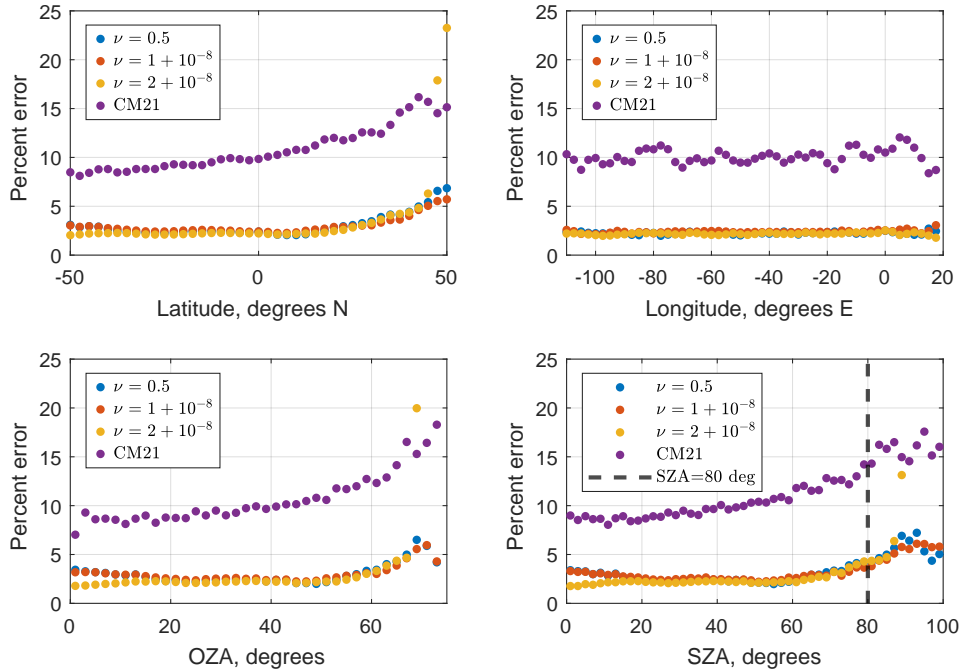


Figure 3. RMS error in percent for varying latitude, longitude, observing angle, and solar zenith angle for the simulated data study. All latitudes and longitudes are included when determining the RMSE as a function of latitude, observing angle, and solar zenith angle, while only latitudes between $\pm 10^\circ$ are considered when determining the RMSE as a function of longitude.

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 ± 27 K, 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.

| 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.

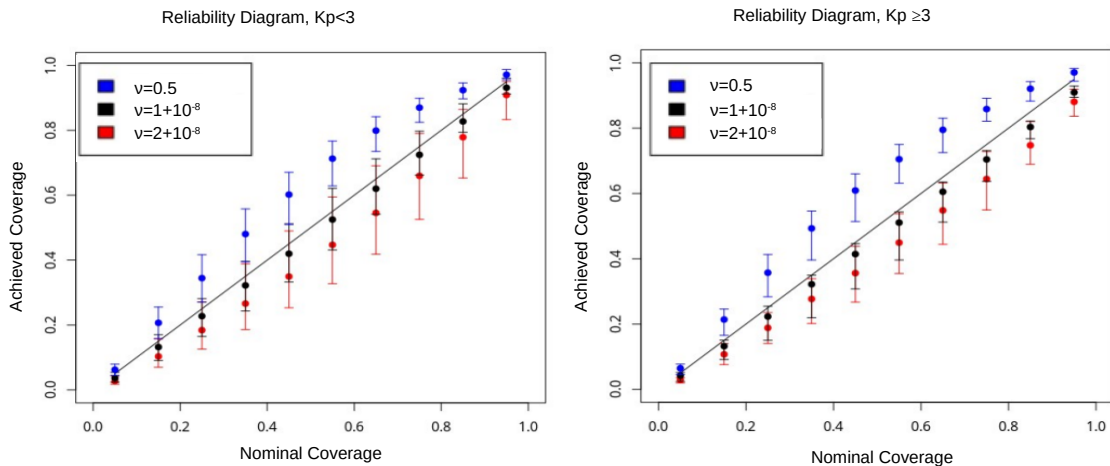


Figure 4. Reliability diagram for the model with varying smoothness parameters and K_p indices. Dots show median coverage probability of a highest-posterior density interval with nominal coverage probability along the x axis, error bars are 10th and 90th percentiles. The models perform worse at uncertainty quantification as K_p increases, however for calm conditions the $\nu = 1 + 10^{-8}$ model captures the uncertainty very well.

335 4 Application to GOLD Disk Emission Data

4.1 Nov 2018 Geomagnetic Storm

As a first demonstration of the method on real GOLD data, we apply it to background corrected GOLD L1C data from 2-6 November 2018. This is the same time period studied using WAM simulations in Section 3, and includes a minor geomagnetic storm on 4-5 November (JD 308-309).

340 Since the GOLD instrument scans each hemisphere independently, the inversion incorporates two consecutive scans into each retrieval. The start times of these scans are separated in time by 12 minutes, which is short enough relative to the nominal thermospheric timescales to ignore the time difference between the scans. 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., 2024b). Again we selected $\gamma = 1$.

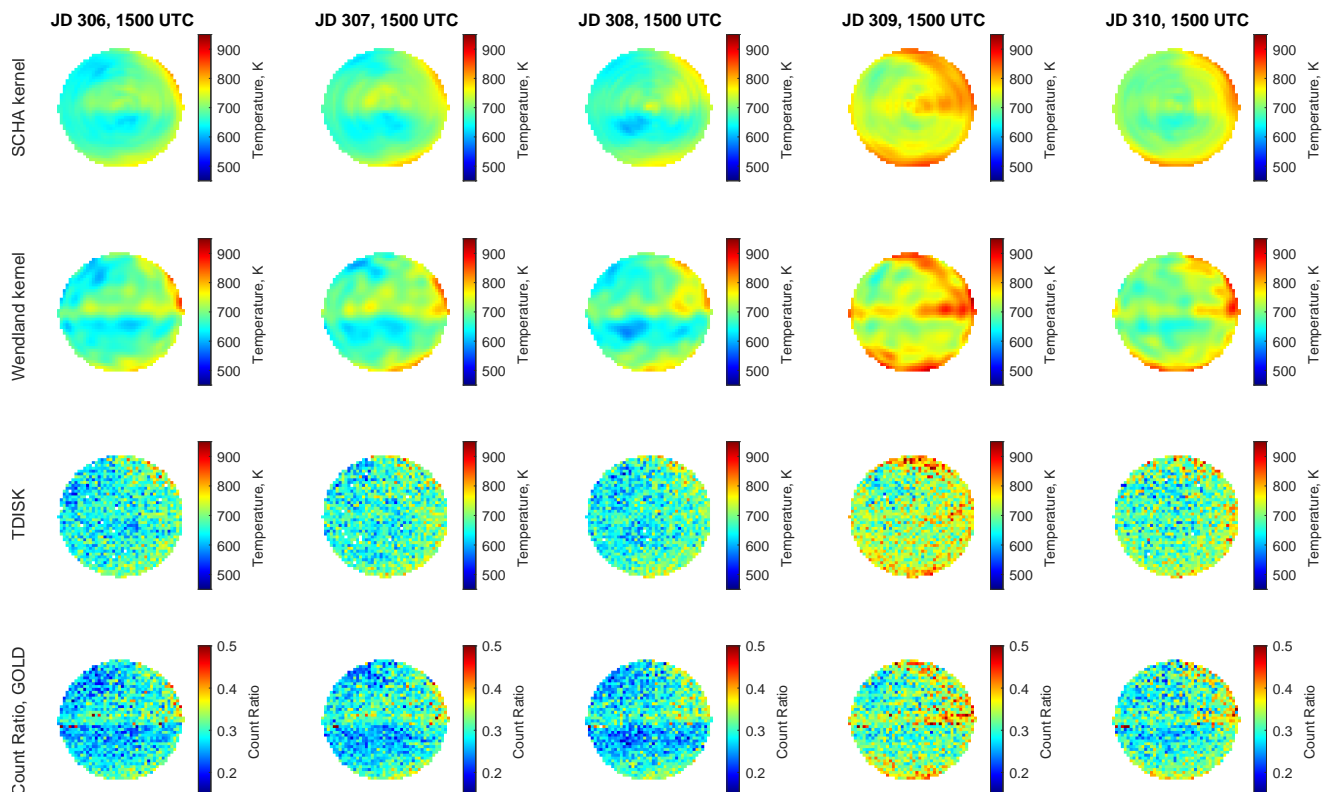


Figure 5. Retrieved temperature using a spherical cap and Wendland kernel (top two rows) compared to TDISK (third row) and the ratio of counts in the upper and lower portion of the LBH (2,0) band (4th row) for JD 306-310, 2018 (2-6 November 2018) at around 1500 UTC, local noon. The retrieved temperatures contain nonphysical structure near the equator inherited from biases in the data due to differences in the sensitivity of the instrument in the different scans of the equatorial region. Similar artifacts can be seen in older versions of the TDISK data.

345 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 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 LIC 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

350 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 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.

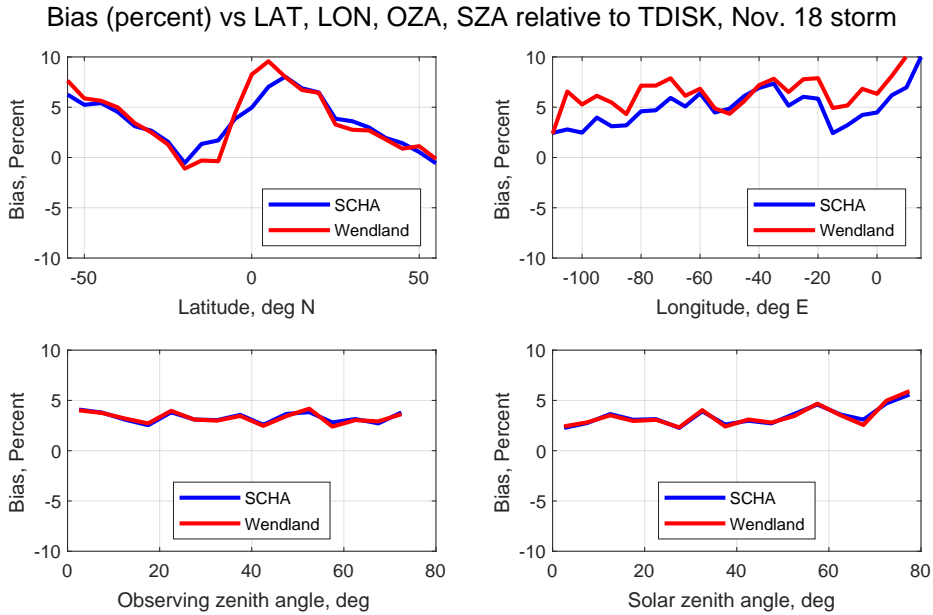


Figure 6. Bias (percent) of estimated temperatures relative to TDISK as a function of latitude, longitude, observing zenith angle, and solar zenith angle during the period of Nov 2-6 2018. Our estimates are routinely biased high, in large part due to the variation in detector sensitivity that is not accounted for.

4.2 May 2024 Geomagnetic Storm

355 The second demonstration example includes a major geomagnetic storm that occurred in May 2024 (also known as the Gannon Storm, Grandin et al. (2024)). With the Dst index reaching below -400 nT and a peak K_p index of 9, this storm resulted in a major expansion of the upper atmosphere causing the neutral temperature to elevate, leading to orbital decay of low Earth orbit satellites due to increased atmospheric drag.

The GOLD mission detected previously unseen structure in the thermosphere during this event, simultaneously seen in both
 360 LBH and 135.6 nm radiances, as well as retrievals of neutral temperature, atmospheric composition, and total electron content. The detected equator-to-pole differences in neutral temperature of over 400K, with high-latitude temperatures in excess of 1400K (Evans et al., 2024a) are well beyond a typical FUV observation and retrieval scenario under nominal conditions. Similar structures have since been observed in other storms as well (Correira et al., 2025). To investigate the performance of our algorithm with the corresponding GOLD L1C data under extreme conditions, we compare our results to the TDISK data
 365 product using both the SCHA kernel and a Wendland kernel function given by (Wendland, 1995)

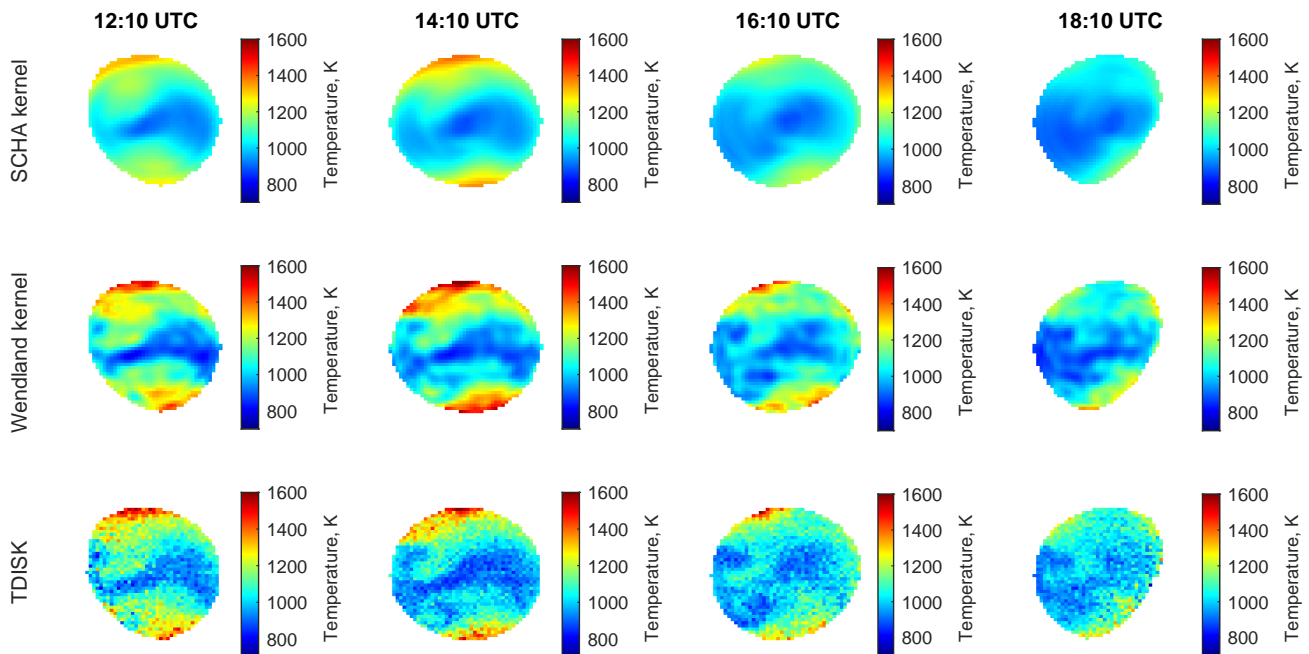


Figure 7. Retrieved temperatures for 12:10-18:10 UTC on 11 May 2024 using both a spherical cap harmonic and Wendland kernel (top two rows) compared with TDISK (bottom row).

$$K(d; r) = \begin{cases} \frac{1}{3} \left(1 - \left|\frac{d}{r}\right|\right)^6 \left(35 \left|\frac{d}{r}\right|^2 + 18 \left|\frac{d}{r}\right| + 3\right), & |d| < r \\ 0, & |d| \geq r \end{cases} \quad (17)$$

Note that in the model evaluation with simulated data for November 2018 the Wendland kernel is found to perform similarly to the SCHA kernel, albeit with slightly worse uncertainty quantification (see Appendix C for details). The results of the retrieval for May 2024 are shown in Fig. 7 using $r = 1/3$ and $\nu = 1 + 10^{-8}$ in the Wendland and SCHA kernels respectively. Due to the extreme nature of the storm, we selected a slightly smaller $\gamma = 1/5$, corresponding to a less informative prior distribution. We see that both models are able to recover spatial structure in the field, especially earlier in the day, however the SCHA kernel tends to oversmooth the estimated field. The retrievals can be improved by more careful selection of the parameters of the kernel, allowing capture of smaller scale structure such as the vortices that are apparent in TDISK, or traveling ionospheric disturbances in more typical scenarios. Our method primarily improves upon Cantrall and Matsuo (2021) by allowing rejection of uncorrelated noise, however small spatial structures can be washed out as well if kernel parameters are not chosen carefully.

We have also included a plot of the relative biases between our two estimates and TDISK during the Gannon storm (Figure 8). We see that the Wendland kernel results are biased slightly high on average over OZA and SZA, with varying bias with latitude

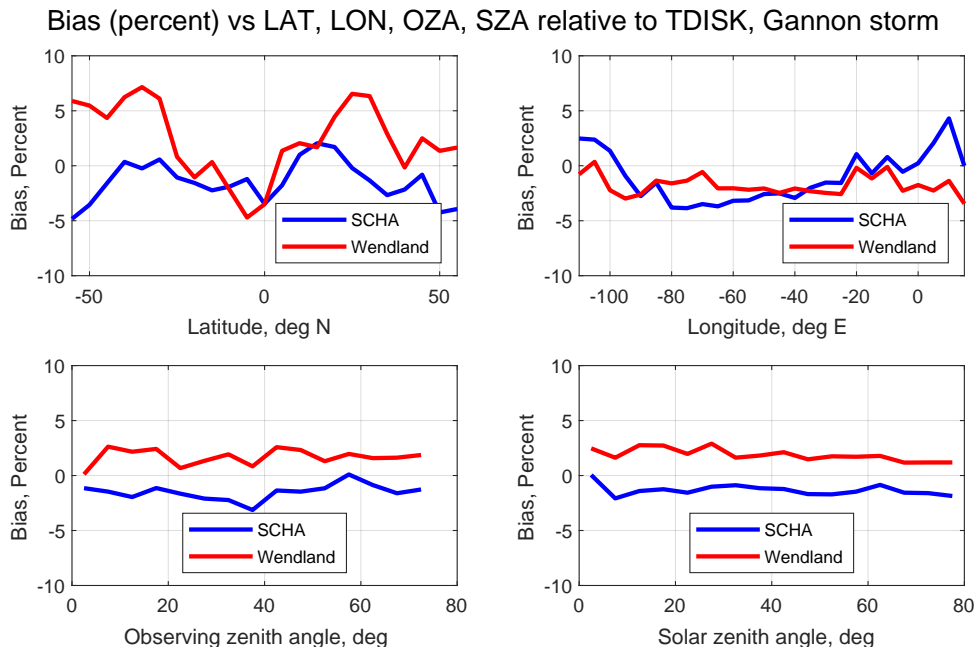


Figure 8. Bias (percent) of estimated temperatures relative to TDISK as a function of latitude, longitude, observing zenith angle, and solar zenith angle during the Gannon storm. Our estimates generally differ from TDISK by less than 5%, well within the TDISK random uncertainty.

and a low bias at the equator. The SCHA result is generally biased low. The largest bias is less than 8%, and the average 95% random uncertainties in TDISK across the plotted estimates are approximately 8.6%, 7.6%, 7.8%, and 8.7%. The uncertainties in the SCHA and Wendland estimates due to shot noise are approximately $\pm 3\%$ and $\pm 6\%$ respectively. The uncertainties from the SCHA appear to be too low, as seen in Figure 4 when the prior is too smooth. However, using the Wendland kernel leads to uncertainties due to shot noise that are more in line with the estimates provided by TDISK, which incorporate other sources of uncertainty as well (Evans et al., 2024b).

5 Discussion

385 5.1 Computational Performance

One of the drawbacks of Bayesian methods as practical retrieval algorithms is their computational load. Often, these methods require generating samples from the posterior distribution using an algorithm such as Markov Chain Monte Carlo, which can be computationally expensive. This work overcomes this problem with the use of a closed form posterior distribution, shifting the computational load to maximizing the likelihood in Eq. (10). Other options, such as a variational inference approach, are

390 possible (e.g. Lloyd et al. (2015)); however, as shown in Walder and Bishop (2017), variational inference is much less efficient than the techniques described in this paper.

The algorithm presented in the previous section is timed on a small runtime analysis study to recover the intensity function:

$$Z(x) = \frac{25 \sin\left(\frac{\pi}{2}x\right)^2 + 10}{8 \cos\left(\frac{\pi}{2}x\right)^2 + 10}, \quad -1 \leq x \leq 1 \quad (18)$$

a plot of which is shown in the upper portion of Fig. 9, using logarithmically spaced numbers of bins from 10 to 5600. 395 A Wendland kernel with $r = 3/4$ is used for the timing study. The minimization is performed with the implementation of conjugate gradient in the `optim` function in base R using the analytic gradient of Eq. (10) as an input.

With 1500 bins, the average time to retrieve the ratio function is about 40 seconds on a desktop computer with 16 GB RAM and an Intel i7 CPU. For problems involving less than 500 bins, the algorithm converges in an average of about 3 seconds, and in about 30 minutes for problems involving 5600 bins. This suggests that the model can be made fast enough 400 to run in an operational capacity if desired, especially for problems involving data sets in the low to mid thousands of points or less. 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. 405 The rest of the algorithm does not require any matrix inversions and scales approximately as $O(n^2)$. For problems with 1500 bins the peak usage is approximately 2 GB, making it easily small enough to run on commercially available laptops.

5.2 Future Work

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 (Cantrall and Matsuo (2021); Cantrall (2022)). 410 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 (Cantrall and Matsuo, 2021). Although several works, such as Aryal et al. (2022), Ajello et al. (2020), and Nicholls (1961), have attempted to determine these populations, any errors in the relative 415 populations of the $v' = 2$ and $v' = 5$ species will contribute to errors in the temperature estimation. 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 (Cantrall and Matsuo, 2021; Cantrall, 2022), or developing a technique that incorporates uncertainty in the vibrational populations may be of interest to practitioners.

One drawback of the two channel ratio method is that it is not robust to wavelength specification errors (Cantrall and Mat- 420 suo, 2021; Eastes et al., 2025). This can be partially overcome by using multiple bands, such as the (1,1) and/or (2,3) bands in conjunction with the (2,0) band, to do estimation and weighting observations by their inverse variance, as observed in Cantrall

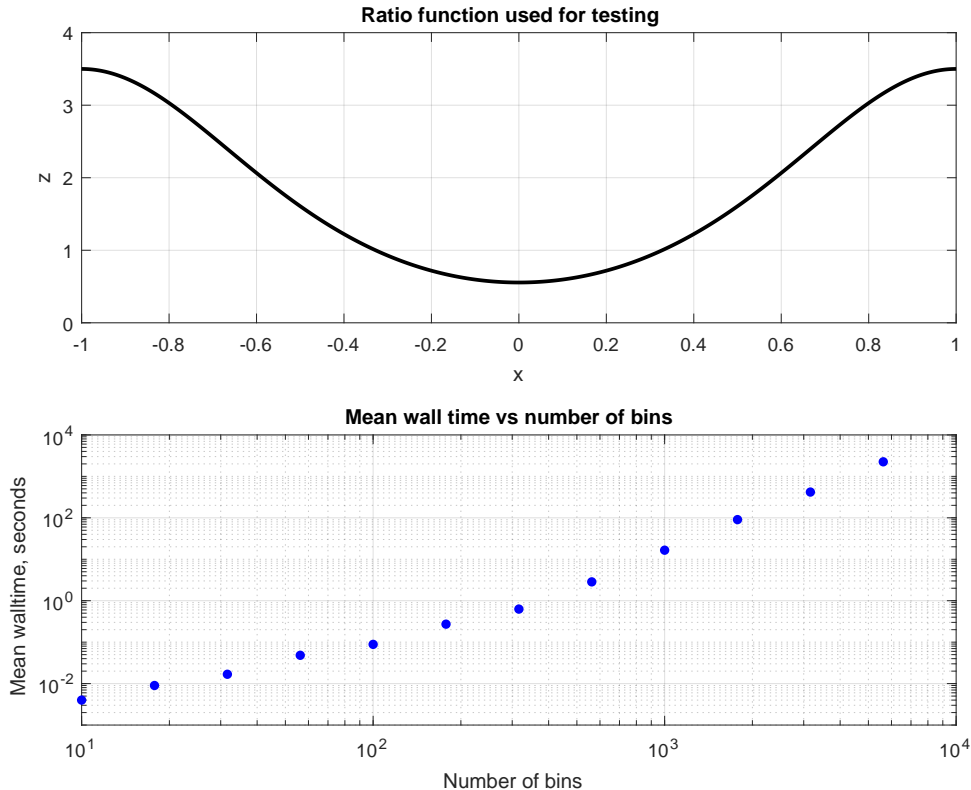


Figure 9. On the top: The ratio function $Z(x)$ estimated in the timing study. On the bottom: timing study results. Times shown are the mean wall times (in seconds) for the code to execute on a desktop computer over 5 repetitions.

(2022). Under our model the posterior distribution no longer takes on a closed form, being given by the Lauricella D function (Pham-Gia and Turkkan, 2002). One avenue of future research is to determine whether there are suitable approximations to these distributions that would allow incorporation of additional spectral information into the retrieval approach while main-
 425 taining efficient computational performance.

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 \mathbf{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
 430 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.

While we have developed this method with the assumption that the data contain counts generated by a single process, that is not necessarily true. Photons from other sources inhibit retrieval, especially when the source counts are low, and the GOLD instrument's geostationary orbit means it can be subjected to high energy particles from outside sources (Evans et al., 2024b). 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 and application to other domains, as uncertainty in the background measurement will necessarily propagate forward into the estimation of the quantity of interest. For instance, the model of Park et al. (2006) handled this by adding a separate term for the background count intensity and estimating both jointly. It is feasible to adopt such an approach to extend this work in the future.

6 Conclusions

While Poisson-distributed multi-spectral sensor data are common in many atmospheric remote sensing applications, current inversion methods used for retrieval of physical parameters often rely on statistical assumptions that may disregard the real statistical properties of sensor data, limiting the scientific utility of datasets. To address this limitation, we introduce a retrieval approach that is based on ratios of photon counts in non-overlapping spectral bands with a focus on FUV upper atmospheric remote sensing applications. This two-channel intensity ratio approach facilitates the development of a computationally efficient and robust method that respects the underlying statistics of the sensor data and naturally incorporates spatial structure. Specifically, the method uses a Poisson point process model with the intensity for each channel independently modeled with a squared Gaussian prior distribution. The new inverse modeling approach is shown to accurately recover spatial structure observed in the true fields and provide a complete posterior distribution of the physical parameter of interest that more faithfully captures variability of underlying geophysical process. The approach is demonstrated on thermospheric neutral temperature retrievals from simulated top-of-atmosphere FUV disk emission data during a minor geomagnetic storm period in November 2018, and from actual calibrated disk emission (L1C) data from the GOLD mission for the same period as well as a major geomagnetic storm that occurred in May 2024. 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.

Code and data availability. The R package developed for this study is available on github at <https://github.com/mfleduc/PoissonRatioUQ> and detailed in LeDuc and Matsuo (2026). GOLD data is available at <https://gold.cs.ucf.edu/data/search/> courtesy of NASA/GOLD and the mission science team.

Appendix A: Derivation of Eq. (4)

In Eq. (4), we present without proof the posterior distribution of the intensity ratio in the case of pointwise inversion with no background, as done in Jin et al. (2006) and Park et al. (2006), as a $BP(\alpha, \beta, p, q)$ distribution. This model contains as a special case priors of the form x^{-k_i} , equivalent to $\alpha_i = 1 - k_i$ and $\beta_i = 0$. The proofs for the cases $n_a = n_b = 1$ with $k_a = k_b = 0$ and $k_a = k_b = 0.5$ are given in Jin et al. (2006), but the distributions are not named as the $BP(a + 1 - k_a, b + 1 - k_b, 1, 1)$ distribution.

Assume that the count data $\{a_i\}_{i=1}^{n_a}$ from the upper channel and $\{b_i\}_{i=1}^{n_b}$ from the lower channel are independent with

$$a_i | \Lambda_a \stackrel{\text{iid}}{\sim} \text{Poisson}(\Lambda_a)$$

$$b_i | \Lambda_b \stackrel{\text{iid}}{\sim} \text{Poisson}(\Lambda_b)$$

for all i , where n_a and n_b are the number of independent observations of the total counts in each channel. In the case of GOLD data where we do 2x2 spatial binning prior to processing, $n_a = n_b = 4$. Now consider the prior $\Lambda_i \sim \Gamma(\alpha_i, \beta_i)$ for each Λ_i . Let $a = \sum_{i=1}^{n_a} a_i$ and $b = \sum_{i=1}^{n_b} b_i$. Then since the Gamma distribution is the conjugate prior of Λ we have that $\Lambda_a \sim \Gamma(a + \alpha_a, \beta_a + n_a)$ and $\Lambda_b \sim \Gamma(b + \alpha_b, \beta_b + n_b)$.

So, the *intensity ratio* $Z = \frac{\Lambda_a}{\Lambda_b}$ has the distribution

$$\begin{aligned} \mathbb{P}[Z \leq z] &= \mathbb{P}\left[\frac{\Lambda_a}{\Lambda_b} \leq z\right] = \int_0^\infty \mathbb{P}\left[\frac{\Lambda_a}{\Lambda_b} \leq z | \Lambda_b = x\right] d\mathbb{P}[\Lambda_b = x] \\ &= \int_0^\infty \mathbb{P}[\Lambda_a \leq z\Lambda_b | \Lambda_b = x] d\mathbb{P}[\Lambda_b = x] \\ &= \frac{(\beta_a + n_a)^{a+\alpha_a} (\beta_b + n_b)^{b+\alpha_b}}{\Gamma(a + \alpha_a) \Gamma(b + \alpha_b)} \int_0^\infty \left[\int_0^{zx} y^{a+\alpha_a-1} e^{-(\beta_a+n_a)y} dy \right] x^{b+\alpha_b-1} e^{-(\beta_b+n_b)x} dx \end{aligned} \quad (\text{A1})$$

The integrand is differentiable in z and is an \mathcal{L}^1 function in x , so by Lebesgue's Dominated Convergence Theorem we can pass derivatives into the integral. Then the probability density is given by

$$\begin{aligned} p_Z(z) &= \frac{(\beta_a + n_a)^{a+\alpha_a} (\beta_b + n_b)^{b+\alpha_b}}{\Gamma(a + \alpha_a) \Gamma(b + \alpha_b)} \int_0^\infty \frac{\partial}{\partial z} \left[\int_0^{zx} y^{a+\alpha_a-1} e^{-(\beta_a+n_a)y} dy \right] x^{b+\alpha_b-1} e^{-(\beta_b+n_b)x} dx \\ &= \frac{(\beta_a + n_a)^{a+\alpha_a} (\beta_b + n_b)^{b+\alpha_b}}{\Gamma(a + \alpha_a) \Gamma(b + \alpha_b)} \int_0^\infty x^{a+b+\alpha_a+\alpha_b-1} e^{-(\beta_b+n_b+(\beta_a+n_a)z)x} dx \end{aligned} \quad (\text{A2})$$

Letting $t = (\beta_b + n_b + (\beta_a + n_a)z)x$ we see that

$$\begin{aligned}
p_Z(z) &= \frac{(\beta_a + n_a)^{a+\alpha_a} (\beta_b + n_b)^{b+\alpha_b}}{\Gamma(a + \alpha_a) \Gamma(b + \alpha_b)} \int_0^\infty x^{a+b+\alpha_a+\alpha_b-1} e^{-(\beta_b+n_b+(\beta_a+n_a)z)x} dx \\
&= \frac{1}{B(a + \alpha_a, b + \beta_b)} \frac{\left(\frac{\beta_a+n_a}{\beta_b+n_b}\right)^{a+\alpha_a} z^{a+\alpha_a-1}}{\left(1 + \frac{\beta_a+n_a}{\beta_b+n_b} z\right)^{a+b+\alpha_a+\alpha_b}}
\end{aligned} \tag{A3}$$

which is a $BP(a + \alpha_a, b + \alpha_b, 1, \frac{\beta_b+n_b}{\beta_a+n_a})$ distribution as desired.

Appendix B: Derivation of the Posterior Mean and Covariance from Eq. (10)

From Eq. (10) we know that

$$485 \quad \ell(\boldsymbol{\psi} | \{a_i\}_{i=1}^d) = \sum_{i=1}^d a_i \log \left(\frac{c}{2} \left(\tilde{\mathbf{K}} \boldsymbol{\psi} \right)_i^2 \right) - \frac{1}{2} \boldsymbol{\psi}^T \tilde{\mathbf{K}} \boldsymbol{\psi}$$

In order to derive the posterior mean and covariance, we will first write $\boldsymbol{f} = \boldsymbol{\Phi}^T \boldsymbol{\beta}$. Then the likelihood becomes

$$\ell(\boldsymbol{\alpha} | \{a_i\}_{i=1}^d) = \sum_{i=1}^d a_i \log \left(\frac{c}{2} \left(\boldsymbol{\Phi}^T \boldsymbol{\beta} \right)_i^2 \right) - \frac{1}{2} \langle \boldsymbol{\beta}, (c\mathbf{I} + \gamma\mathbf{H}^{-1}) \boldsymbol{\beta} \rangle \tag{B1}$$

where $\mathbf{H} = \text{diag}(\eta_i)$. From this we can see that the posterior mode estimate of $\boldsymbol{\beta}$ must satisfy

$$\hat{\boldsymbol{\beta}} = (c\mathbf{I} + \gamma\mathbf{H}^{-1})^{-1} \nabla_{\boldsymbol{\beta}} \left[\sum_{i=1}^d a_i \log \left(\frac{c}{2} \left(\boldsymbol{\Phi}^T \hat{\boldsymbol{\beta}} \right)_i^2 \right) \right] = (c\mathbf{I} + \gamma\mathbf{H}^{-1})^{-1} \sum_{i=1}^d \frac{2a_i \boldsymbol{\Phi}_{:,i}}{\left(\boldsymbol{\Phi}^T \hat{\boldsymbol{\beta}} \right)_i} \tag{B2}$$

490 where $\boldsymbol{\Phi}_{:,i}$ is the i^{th} column of $\boldsymbol{\Phi}$. Then the posterior mode estimate of \boldsymbol{f} is then given by

$$\hat{\boldsymbol{f}} = \boldsymbol{\Phi}^T \hat{\boldsymbol{\beta}} = \sum_{i=1}^d \left[\boldsymbol{\Phi}^T (c\mathbf{I} + \gamma\mathbf{H}^{-1})^{-1} \boldsymbol{\Phi}_{:,i} \right] \frac{2a_i}{\left(\boldsymbol{\Phi}^T \hat{\boldsymbol{\beta}} \right)_i} \tag{B3}$$

From this, we can see the posterior mode has the form $\hat{\boldsymbol{f}} = \tilde{\mathbf{K}} \hat{\boldsymbol{\psi}}$ by applying the eigenvector expansion of $\tilde{\mathbf{K}}$. This implies that, in our situation, $\hat{\boldsymbol{\psi}}_i = \frac{2a_i}{\boldsymbol{f}_i}$, which is analogous to Eq. (11) in Walder and Bishop (2017). This posterior mode is used as the posterior mean in the Laplace approximation of the posterior distribution.

495 The inverse of the posterior covariance of $\boldsymbol{\beta}$ is given by the Hessian matrix at the posterior mode, given by

$$\mathbf{Q}^{-1} = -\frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} L(\boldsymbol{\beta} | \{a_j\}_{j=1}^d) |_{\hat{\boldsymbol{\beta}}} = (c\mathbf{I} + \gamma\mathbf{H}^{-1}) + \sum_{i=1}^d \frac{2a_i}{\left(\boldsymbol{\Phi}^T \hat{\boldsymbol{\beta}} \right)_i^2} \boldsymbol{\Phi}_{:,i} \boldsymbol{\Phi}_{:,i}^T \tag{B4}$$

Now letting $\mathbf{Z} = (c\mathbf{I} + \gamma\mathbf{H}^{-1})$ and $\mathbf{D} = \text{diag}\left(\frac{\hat{\psi}_j^2}{2a_j}\right)$ we have that $\mathbf{Q}^{-1} = \mathbf{Z} + \mathbf{\Phi}\mathbf{D}\mathbf{\Phi}^T$ and can apply the Woodbury-Morrison formula to write that

$$\mathbf{Q} = \mathbf{Z}^{-1} - \mathbf{Z}^{-1}\mathbf{\Phi}(\mathbf{D}^{-1} + \mathbf{\Phi}^T\mathbf{Z}^{-1}\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{Z}^{-1} \quad (\text{B5})$$

500 The posterior covariance matrix of \mathbf{f} is then given by

$$\hat{\Sigma} = \mathbf{\Phi}^T\mathbf{Q}\mathbf{\Phi} \quad (\text{B6})$$

$$= \tilde{\mathbf{K}} - \tilde{\mathbf{K}}(\mathbf{D}^{-1} + \tilde{\mathbf{K}})^{-1}\tilde{\mathbf{K}} \quad (\text{B7})$$

as desired.

Appendix C: The kernel

505 Since we wish to retrieve a field whose domain is the surface of the sphere, one option for constructing the eigenvector expansion of the kernel would be to use spherical harmonics. However, since we are only observing a portion of the sphere this would mean that the underlying intensity field is assumed to be 0 outside the viewing area and so is not continuous. Thus, using spherical harmonics, which are nonlocalized analytic functions, to represent the field can lead to nonphysical ringing in the retrieved fields. Instead, we choose the expansion of the kernel to be given in terms of spherical cap harmonics V_{lm} (SCHAs, 510 Haines (1985)). These functions form an orthonormal basis for the Hilbert space of square integrable functions on the spherical cap, making them a natural basis for this problem. They have been used extensively in geoscience, for example in determining electrodynamics of the polar ionosphere (Richmond and Kamide, 1988), statistical analysis of TOPEX/POSEIDON data (Hwang and Chen, 1997), and mapping of total electron content over China and Iran (Liu et al., 2011; Razin and Voosoghi, 2017).

515 There are two families of spherical cap harmonics. One is the so-called even harmonics, which are calculated by solving the eigenvalue problem (Haines, 1985)

$$\begin{aligned} \Delta V_{lm}(\theta, \phi) &= \frac{1}{\eta_{lm}} V_{lm}(\theta, \phi) \\ V_{lm}(\theta, \phi) &= V_{lm}(\theta, \phi + 2\pi) \\ \frac{\partial}{\partial \phi} V_{lm}(\theta, \phi) &= \frac{\partial}{\partial \phi} V_{lm}(\theta, \phi + 2\pi) \\ \frac{\partial}{\partial \theta} V_{lm}(\theta_0, \phi) &= 0 \\ V_{lm}(0, \phi) &= 0, m = 0 \\ \frac{\partial}{\partial \theta} V_{lm}(0, \phi) &= 0, m \neq 0 \end{aligned} \quad (\text{C1})$$

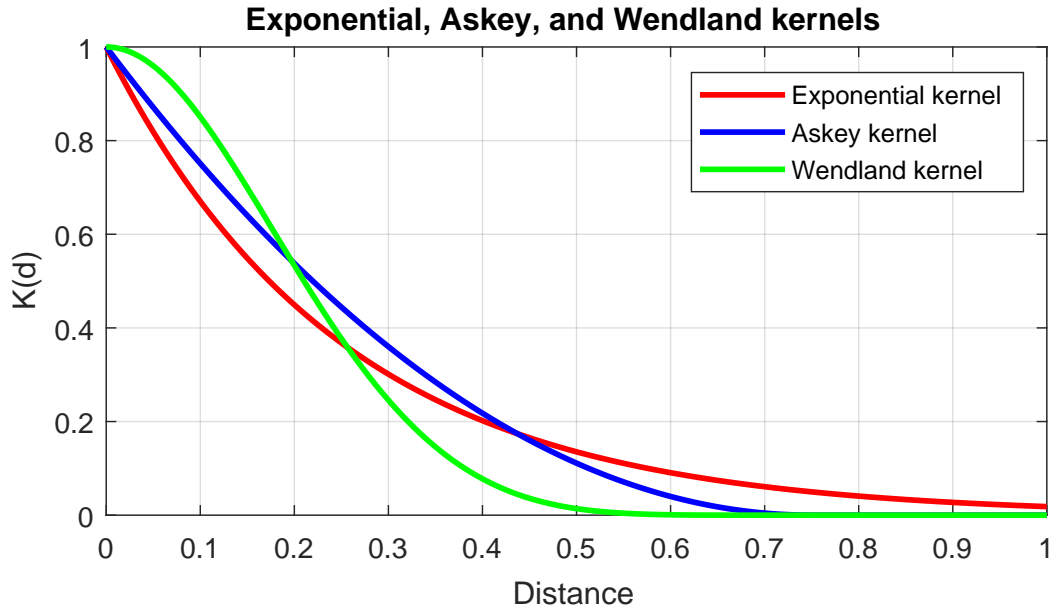


Figure C1. Askey (blue), exponential (red) and Wendland (green) kernels. The Askey and exponential kernels give rise to a rough process, while the Wendland kernel gives rise to a smoother process.

where θ_0 is the co-latitude of the cap boundary, taken for our purposes to be 64° . The odd harmonics are retrieved when the boundary condition $\frac{\partial}{\partial \theta} V_{lm}(\theta_0, \phi) = 0$ is replaced by $V_{lm}(\theta_0, \phi) = 0$. The even harmonics allow for the reconstruction of an arbitrary function on the boundary, and the odd harmonics allow reconstruction of a field with an arbitrary derivative at the boundary. Each of these sets of functions form an orthonormal basis of $\mathcal{L}^2(S)$, however they are not mutually orthogonal. Since we are concerned with reconstruction of the temperature field and not its derivatives, we choose to represent the field with the even harmonics. The solutions to this problem are the eigenvalue/eigenfunction pairs

$$V_{lm}(\theta, \phi) = e^{im\phi} P_{lm}(\cos(\theta))$$

$$\eta_{lm} = \frac{1}{l(l+1)} \quad (\text{C2})$$

where P_{lm} are the associated Legendre functions, and l need not be an integer, but $m \in \mathbb{Z}$ and $l > |m|$. The associated Legendre functions are not polynomials, but instead are related to the hypergeometric function (Hwang and Chen, 1997). Due to the computational difficulty of solving the eigenvalue problem for various l, m we only calculate the SCHAs up to $m = 20$, corresponding to 441 basis functions. Thus, the kernel \mathbf{K} is given by Eq. (15), and the equivalent kernel is then given by

$$\tilde{\mathbf{K}}_{ij} = \sum_{l,m} \frac{1}{c + \gamma l^\nu (1+l)^\nu} V_{lm}(\theta_i, \phi_i) V_{lm}(\theta_j, \phi_j) \quad (\text{C3})$$

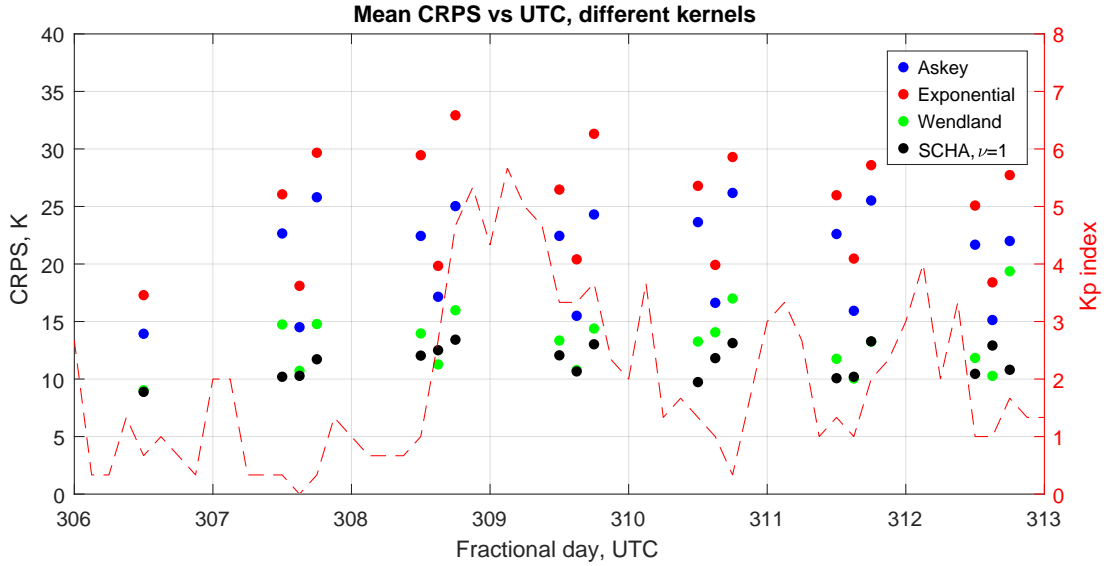


Figure C2. CRPS vs time for the Askey, exponential, Wendland, and SCHA kernels

530 As was shown in the body of the paper, the best of the SCHA kernels studied in terms of uncertainty quantification and retrieval accuracy was the one where $\nu = 1 + 10^{-8}$. We will compare this specific kernel with several other popular kernels that are positive definite on the sphere. A list of some possible kernels is included in Gneiting (2013). We choose to compare the SCHA kernel to the Askey kernel, given by

$$K(d; r) = \begin{cases} (1 - |d/r|)^2 & |d| < r \\ 0 & |d| \geq r \end{cases} \quad (\text{C4})$$

535 , the exponential kernel, given by

$$K(d; r) = \exp\left(-\left|\frac{d}{r}\right|\right) \quad (\text{C5})$$

, and the Wendland kernel given by Eq. (17). These kernels are plotted in Fig. C1. In all cases, the distance measure is the angular distance. From this figure, we see one feature that separates the Wendland from the other kernels: the number of derivatives that a kernel function has at $d = 0$ is related to the number of mean-square derivatives of the field (Stein, 1999), and the Askey and exponential kernels do not have any derivatives at $d = 0$, while the Wendland kernel does. This suggests that the Wendland kernel will generate smoother fields than the other two. Additionally, the Askey and Wendland kernels are compactly supported, meaning that they give rise to sparse kernel matrices. This property could make them useful in the analysis of large datasets, where correlations being zero outside a cutoff distance means that the algorithm can take advantage of sparsity in the resulting matrices. The exponential kernel is the covariance of an AR-1 process, leading to a sparse precision matrix for \mathbf{f} .

Reliability diagram, different kernels

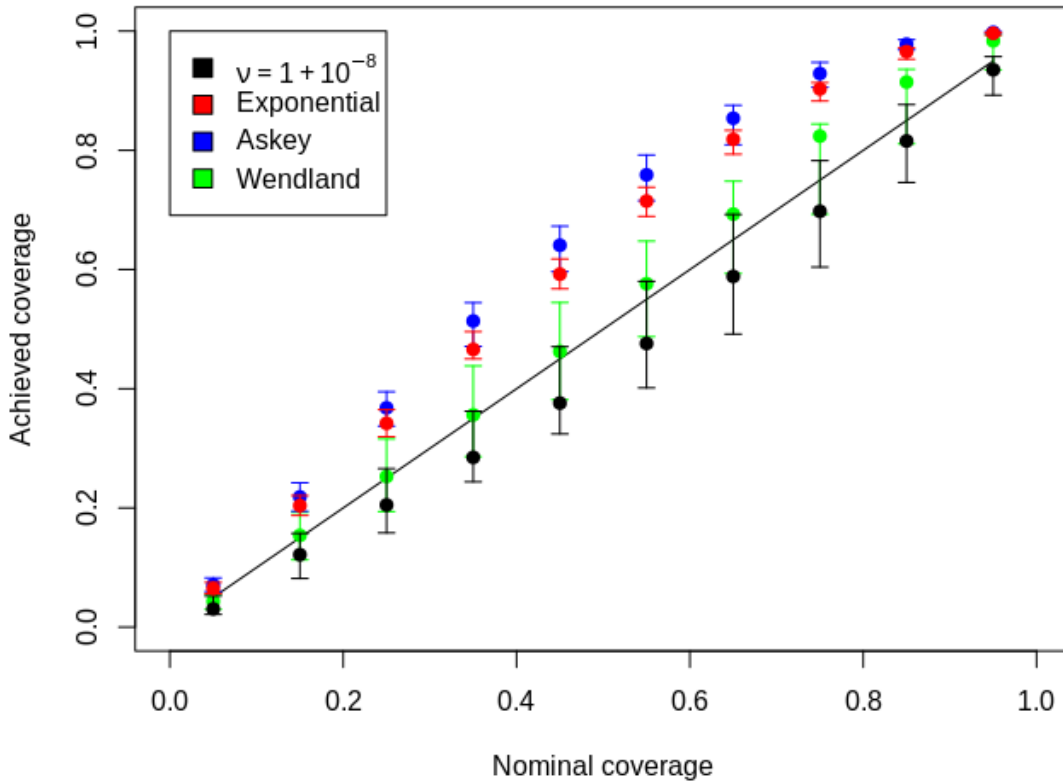


Figure C3. Reliability diagram for the different kernels, all K_p values.

545 We performed temperature retrievals on the simulated data examined in Sect. 3. In general, we found that the Wendland kernel had similar retrieval performance in terms of CRPS, with slightly worse uncertainty quantification as the nominal coverage increases. The other two kernels were not competitive in either metric. These results are shown in Figs. C2 and C3. We expect that this is due to the implied roughness of the fields from the Askey and exponential kernels. Since the retrieved field is an integrated quantity, we expect it to be relatively smooth, and from the data we see that in most conditions it is. This makes the Askey and exponential kernels less appropriate for the retrieval and less accurate in quantifying the uncertainty. The SCHA kernel with $\nu = 0.5$ showed similar levels of overdispersion in the posterior distribution, suggesting that incorrectly specifying the smoothness of the field is to blame.

550 While these results show that the SCHA kernel is the best option for our analysis on the simulated data, we want to stress that any operational choice of kernel should involve careful cross-validation and other thorough testing before use. This is especially important because the application of the SCHA kernel to time periods where most of the disk is not illuminated

could cause problems akin to the ringing issue experienced with spherical harmonics. In a situation like this, a Wendland kernel or another one not investigated here may be preferable. In fact, we see already in Section 4 that in the case of the Gannon storm a Wendland kernel is preferable to the SCHA kernel.

Author contributions. ML developed the technique and performed the analyses. TM and ML chose the datasets. TM and WK provided supervision and support. All authors provided interpretation and prepared the manuscript.

Competing interests. The authors declare no competing interests.

Acknowledgements. We would like to thank Dr. Clayton Cantrall for his help working with the curves of intensity ratio vs temperature generated for Cantrall and Matsuo (2021), which can be found in Cantrall (2021). We would like to acknowledge high-performance computing support from the Derecho system (doi:10.5065/qx9a-pg09) provided by the NSF National Center for Atmospheric Research (NCAR), sponsored by the National Science Foundation.

Financial support. This work was supported by NASA 80NSSC22K0175, UCAR SUBAWD006087, and NSF AGS-2231409 awards to the University of Colorado Boulder

References

- 570 Ajello, J. M., Evans, J. S., Veibell, V., Malone, C. P., Holsclaw, G. M., Hoskins, A. C., Lee, R. A., McClintock, W. E., Aryal, S., Eastes, R. W., and Schneider, N.: The UV Spectrum of the Lyman-Birge-Hopfield Band System of N₂ Induced by Cascading from Electron Impact, *Journal of Geophysical Research: Space Physics*, <https://doi.org/10.1029/2019JA027546>, 2020.
- Akmaev, R. A.: Whole Atmosphere Modeling: Connecting Terrestrial and Space Weather, *Reviews of Geophysics*, 49, <https://doi.org/10.1029/2011RG000364>, 2011.
- 575 Aksnes, A., Eastes, R., Budzien, S., and Dymond, K.: Neutral temperatures in the lower thermosphere from N₂ Lyman-Birge-Hopfield (LBH) band profiles, *Geophys. Res. Lett.*, 33, <https://doi.org/10.1029/2006GL026255>, 2006.
- Amaral Turkman, M. A., Paulino, C. D., and Müller, P.: *Computational Bayesian Statistics: An Introduction*, Institute of Mathematical Statistics Textbooks, Cambridge University Press, 2019.
- Antoniadis, A. and Bigot, J.: Poisson inverse problems, *The Annals of Statistics*, 34, 2132 – 2158, <https://doi.org/10.1214/009053606000000687>, 2006.
- 580 Aryal, S., Evans, J. S., Ajello, J. M., Solomon, S. C., Burns, A. W., Eastes, R. W., and McClintock, W. E.: Constraining the Upper Level Vibrational Populations of the N₂ Lyman-Birge-Hopfield Band System Using GOLD Mission’s Dayglow Observations, *Journal of Geophysical Research: Space Physics*, 127, <https://doi.org/10.1029/2021JA029869>, 2022.
- Boersma, C., Rubin, R., and Allamandola, L.: Spatial analysis of the polycyclic aromatic hydrocarbon features southeast of the Orion Bar, *The Astrophysical Journal*, 753, 168, <https://doi.org/10.1088/0004-637X/753/2/168>, 2012.
- 585 Budzien, S. A., Fortna, C. B., Dymond, K. F., Thonnard, S. E., Nicholas, A. C., McCoy, R. P., and Thomas, R. J.: Thermospheric temperature derived from ARGOS observations of N₂ Lyman-Birge-Hopfield emission, *Eos Trans. AGU*, 82, Spring Meet. Suppl., 2001.
- Cantrall, C.: Deriving column-integrated thermospheric temperature with the N₂ Lyman-Birge-Hopfield (2,0) band, <https://doi.org/10.17605/OSF.IO/KHNQ7>, 2021.
- 590 Cantrall, C.: New approaches for quantifying and understanding thermosphere temperature variability from far ultraviolet dayglow, Ph.D. thesis, University of Colorado-Boulder, 2022.
- Cantrall, C. and Matsuo, T.: Deriving column-integrated thermospheric temperature with the N₂ Lyman-Birge-Hopfield (2,0) band, *Atmospheric Measurement Techniques*, <https://doi.org/10.5194/amt-14-6917-2021>, 2021.
- Cantrall, C. E., Matsuo, T., and Solomon, S. C.: Upper Atmosphere Radiance Data Assimilation: A Feasibility Study for GOLD Far Ultraviolet Observations, *Journal of Geophysical Research: Space Physics*, 124, 8154–8164, <https://doi.org/10.1029/2019JA026910>, 2019.
- 595 Chiles, J.-P. and Lantuéjoul, C.: 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*, pp. 39–68, Springer, 2005.
- Coath, C. D., Steele, R. C. J., and Lunnon, W. F.: Statistical Bias in Isotope Ratios, *Journal of Analytical Atomic Spectrometry*, <https://doi.org/10.1039/C2JA10205F>, 2013.
- 600 Correia, J., Evans, J. S., Lumpe, J. D., Krywonos, A., Daniell, R., Veibell, V., McClintock, W. E., and Eastes, R. W.: Thermospheric Composition and Solar EUV Flux From the Global-Scale Observations of the Limb and Disk (GOLD) Mission, *Journal of Geophysical Research: Space Physics*, 126, <https://doi.org/10.1029/2021JA029517>, 2021.
- Correia, J., Evans, J. S., Lumpe, J. D., Eastes, R. W., Wang, W., Aryal, S., Krywonos, A., and McClintock, W. E.: Upper atmospheric vortices following strong geomagnetic storms, *Geophysical Research Letters*, 52, e2024GL113726, 2025.
- 605

- Cox, D. R.: Some Statistical Methods Connected with Series of Events, *Journal of the Royal Statistical Society: Series B (Methodological)*, 17, 129–157, <https://doi.org/10.1111/j.2517-6161.1955.tb00188.x>, 1955.
- De Oliveira, V.: Hierarchical Poisson models for spatial count data, *Journal of Multivariate Analysis*, 122, 393–408, <https://doi.org/10.1016/j.jmva.2013.08.015>, 2013.
- 610 Doornbos, E. and Klinkrad, H.: Modelling of space weather effects on satellite drag, *Advances in Space Research*, 37, 1229–1239, 2006.
- Eastes, R., Evans, J. S., Gan, Q., McClintock, W., and Lumpe, J.: Remote sensing of lower-middle-thermosphere temperatures using the N2 Lyman–Birge–Hopfield (LBH) bands, *Atmospheric Measurement Techniques*, <https://doi.org/10.5194/amt-18-921-2025>, 2025.
- Emmert, J.: Thermospheric mass density: A review, *Advances in Space Research*, 56, 773–824, <https://doi.org/https://doi.org/10.1016/j.asr.2015.05.038>, 2015.
- 615 Evans, J. S., Correira, J., Lumpe, J. D., Eastes, R. W., Gan, Q., Laskar, F. I., Aryal, S., Wang, W., Burns, A. G., Beland, S., Cai, X., Codrescu, M., England, S., Greer, K., Krywonos, A., McClintock, W. E., Plummer, T., and Veibell, V.: GOLD Observations of the Thermospheric Response to the 10–12 May 2024 Gannon Superstorm, *Geophysical Research Letters*, 51, <https://doi.org/10.1029/2024GL110506>, 2024a.
- Evans, J. S., Lumpe, J. D., Eastes, R. W., Correira, J., Aryal, S., Laskar, F., Veibell, V., Krywonos, A., Plummer, T., and McClintock, W. E.: Disk Images of Neutral Temperature From the Global-Scale Observations of the Limb and Disk (GOLD) Mission, *Journal of Geophysical*
- 620 *Research: Space Physics*, 129, <https://doi.org/10.1029/2024JA032424>, 2024b.
- Flaxman, S., Teh, Y. W., and Sejdinovic, D.: Poisson intensity estimation with reproducing kernels, *Electronic Journal of Statistics*, 11, 5081–5104, <https://doi.org/10.1214/17-EJS1339SI>, 2017.
- Gallardo i Peres, G., Dall, J., Mason, P. J., Ghail, R., and Hensley, S.: A Generalized Beta Prime Distribution as the Ratio Probability Density Function for Change Detection Between Two SAR Intensity Images With Different Number of Looks, *IEEE Transactions on Geoscience*
- 625 *and Remote Sensing*, 62, 1–14, <https://doi.org/10.1109/TGRS.2024.3369509>, 2024.
- Gneiting, T.: Strictly and non-strictly positive definite functions on spheres, *Bernoulli*, 19, 1327–1349, <https://doi.org/10.3150/12-BEJSP06>, 2013.
- Gneiting, T. and Rafferty, A.: Strictly Proper Scoring Rules, Prediction, and Estimation, *Journal of the American Statistical Association*, <https://doi.org/10.1198/016214506000001437>, 2007.
- 630 Grandin, M., Bruus, E., Ledvina, V. E., Partamies, N., Barthelemy, M., Martinis, C., Dayton-Oxland, R., Gallardo-Lacourt, B., Nishimura, Y., Herlingshaw, K., Thomas, N., Karvinen, E., Lach, D., Spijkers, M., and Bergstrand, C.: The Gannon Storm: citizen science observations during the geomagnetic superstorm of 10 May 2024, *Geoscience Communication*, 7, 297–316, <https://doi.org/10.5194/gc-7-297-2024>, 2024.
- Haines, G.: Computer programs for spherical cap harmonic analysis of potential and general fields, *Computers & Geosciences*, 14, 413–447, <https://www.sciencedirect.com/science/article/pii/0098300488900271>, 1988.
- 635 Haines, G. V.: Spherical cap harmonic analysis, *Journal of Geophysical Research: Solid Earth*, 90, 2583–2591, <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JB090iB03p02583>, 1985.
- Humphrey, P. J., Liu, W., and Buote, D. A.: χ^2 and Poissonian data: Biases even in the high-count regime and how to avoid them, *The Astrophysical Journal*, 693, <https://doi.org/10.1088/0004-637X/693/1/822>, 2009.
- 640 Hunter, J. K. and Nachtergaele, B.: *Applied Analysis*, World Scientific, 2020.
- Hwang, C. and Chen, S.-K.: Fully normalized spherical cap harmonics: application to the analysis of sea-level data from TOPEX/POSEIDON and ERS-1, *Geophysical Journal International*, 129, 450–460, <https://doi.org/10.1111/j.1365-246X.1997.tb01595.x>, 1997.

- Jia, J. and Yi, F.: Atmospheric temperature measurements at altitudes of 5-30km with a double-grating-based pure rotational Raman lidar, *Appl. Opt.*, 53, 5330–5343, <https://doi.org/10.1364/AO.53.005330>, 2014.
- 645 Jin, Y., Zhang, S., and Wu, J.: Hardness Ratio estimation in Low Counting X-Ray Photometry, *The Astrophysical Journal*, <https://doi.org/10.1086/508677>, 2006.
- Krywonos, A., Murray, D., Eastes, R., Aksnes, A., Budzien, S., and Daniell, R.: Remote sensing of neutral temperatures in the Earth's thermosphere using the Lyman-Birge-Hopfield bands of N₂: Comparisons with satellite drag data, *Journal of Geophysical Research: Space Physics*, 117, 2012.
- 650 LeDuc, M. and Matsuo, T.: PoissonRatioUQ: An R package for band ratio uncertainty quantification, <https://arxiv.org/abs/2602.07165>, 2026.
- Leonard, J. M., Forbes, J. M., and Born, G. H.: Impact of tidal density variability on orbital and reentry predictions, *Space Weather*, 10, <https://doi.org/https://doi.org/10.1029/2012SW000842>, 2012.
- Liu, J., Chen, R., Wang, Z., and Zhang, H.: Spherical cap harmonic model for mapping and predicting regional TEC, *GPS Solutions*, 15, 109–119, [10.1007/s10291-010-0174-8](https://doi.org/10.1007/s10291-010-0174-8), 2011.
- 655 Lloyd, C., Gunter, T., Osborne, M., and Roberts, S.: Variational Inference for Gaussian Process Modulated Poisson Processes, in: *Proceedings of the 32nd International Conference on Machine Learning*, edited by Bach, F. and Blei, D., vol. 37 of *Proceedings of Machine Learning Research*, pp. 1814–1822, PMLR, <https://proceedings.mlr.press/v37/lloyd15.html>, 2015.
- Matzka, J., Bronkalla, O., Tornow, K., Elger, K., and Stolle, C.: Geomagnetic Kp index. V. 1.0., <https://doi.org/10.5880/Kp.0001>, 2021.
- McClintock, W. E., Eastes, R. W., Beland, S., Bryant, K. B., Burns, A. G., Correira, J., Daniell, R. E., Evans, J. S., Harper, C. S., Karan, 660 D. K., Krywonos, A., Lumpe, J. D., Plummer, T. M., Solomon, S. C., Vanier, B. A., and Veibel, V.: Global-Scale Observations of the Limb and Disk Mission Implementation: 2. Observations, Data Pipeline, and Level 1 Data Products, *Journal of Geophysical Research: Space Physics*, 125, <https://doi.org/10.1029/2020JA027809>, 2020.
- McCullagh, P. and Møller, J.: The Permanental Process, *Advances in Applied Probability*, 38, 873–888, <https://doi.org/10.1017/S0001867800001361>, 2006.
- 665 McDonald, J. B. and Richards, D. O.: Model selection: some generalized distributions, *Comm. Statist. Theory Methods*, 16, 1049–1074, <https://doi.org/10.1080/03610928708829422>, 1987.
- McDonald, J. B. and Xu, Y. J.: A generalization of the beta distribution with applications, *Journal of Econometrics*, [https://doi.org/10.1016/0304-4076\(94\)01612-4](https://doi.org/10.1016/0304-4076(94)01612-4), 1995.
- Mehta, P. M., Paul, S. N., Crisp, N. H., Sheridan, P. L., Siemes, C., March, G., and Bruinsma, S.: Satellite drag coefficient modeling for 670 thermosphere science and mission operations, *Advances in Space Research*, 72, 5443–5459, 2023.
- Meier, R. R.: The Thermospheric Column O/N₂ Ratio, *Journal of Geophysical Research: Space Physics*, 126, e2020JA029059, <https://doi.org/10.1029/2020JA029059>, 2021.
- Møller, J.: Properties of spatial Cox process models, *Journal of Statistical Research of Iran*, 2, 1–18, 2005.
- Nicholls, R.: Franck-Condon factors to high vibrational quantum numbers I: N₂ and N₂⁺, *Journal of Research of the National Bureau of Standards. Section A, Physics and Chemistry*, 65, 451, <https://doi.org/10.6028/jres.065A.047>, 1961.
- 675 Nicolaou, G., Livadiotis, G., Sarlis, N., and Ioannou, C.: Resolving velocity distribution function parameters from observations with significant Poisson statistical uncertainty, *RAS Techniques and Instruments*, 3, 874–878, <https://doi.org/10.1093/rasti/rzae059>, 2024.
- Nowak, R. D. and Kolaczyk, E. D.: A statistical multiscale framework for Poisson inverse problems., *IEEE Transactions on Information Theory*, 46, 1811–1825, <https://doi.org/10.1109/18.857793>, 2000.

- 680 Park, T., Kashyap, V. L., Siemiginowska, A., van Dyk, D. A., Zezas, A., Heinke, C., and Wargelin, B. J.: Bayesian Estimation of Hardness Ratios: Modeling and Computations, *The Astrophysical Journal*, 652, 610, <https://doi.org/10.1086/507406>, 2006.
- Paxton, L. J., Schaefer, R. K., Zhang, Y., and Kil, H.: Far ultraviolet instrument technology, *Journal of Geophysical Research: Space Physics*, 122, 2706–2733, <https://doi.org/10.1002/2016JA023578>, 2017.
- Pham-Gia, T. and Turkkan, N.: Operations on the Generalized F Variables and Applications, *Statistics*, 685 <https://doi.org/10.1080/02331880212855>, 2002.
- R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>, 2025.
- Rasmussen, C. E. and Williams, C. K. I.: Gaussian Processes for Machine Learning, The MIT Press, ISBN 9780262256834, <https://doi.org/10.7551/mitpress/3206.001.0001>, 2005.
- 690 Razin, M. and Voosoghi, B.: Regional ionosphere modeling using spherical cap harmonics and empirical orthogonal functions over Iran, *Acta Geodaetica et Geophysica*, 52, 19–33, <https://doi.org/10.1007/s40328-016-0162-8>, 2017.
- Richmond, A. D. and Kamide, Y.: Mapping electrodynamic features of the high-latitude ionosphere from localized observations: Technique, *Journal of Geophysical Research: Space Physics*, 93, 5741–5759, <https://doi.org/10.1029/JA093iA06p05741>, 1988.
- Solomon, S. C.: Global modeling of thermospheric airglow in the far ultraviolet, *JGR Space Physics*, 122, 7834–7848, 2017.
- 695 Stein, M. L.: Interpolation of Spatial Data: Some Theory for Kriging, Springer Series in Statistics, Springer, 1999.
- Streit, R. L.: Poisson Point Processes: Imaging, Tracking, and Sensing, Springer, <https://doi.org/10.1007/978-1-4419-6923-1>, 2010.
- Strickland, D. J., Evans, J. S., and Paxton, L. J.: Satellite remote sensing of thermospheric O/N₂ and solar EUV: 1. Theory, *Journal of Geophysical Research: Space Physics*, 100, 12 217–12 226, <https://doi.org/10.1029/95JA00574>, 1995.
- Walder, C. J. and Bishop, A. N.: Fast Bayesian Intensity Estimation for the Permanent Process, in: Proceedings of the 34th International Conference on Machine Learning, edited by Precup, D. and Teh, Y. W., vol. 70 of *Proceedings of Machine Learning Research*, pp. 3579–3588, PMLR, <https://proceedings.mlr.press/v70/walder17a.html>, 2017.
- 700 Wang, C., Liao, J.-Y., Guan, J., Liu, Y., Li, C.-K., Sai, N., Jin, J., and Zhang, S.-N.: Analysis of bright source hardness ratios in the 4 yr Insight-HXMT galactic plane scanning survey catalog, *Research in Astronomy and Astrophysics*, 24, 025 013, <https://doi.org/10.1088/1674-4527/ad18a3>, 2024.
- 705 Wendland, H.: Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree., *Adv Comput Math*, 4, 389–396, 1995.
- Yamada, S., Axelsson, M., Ishisaki, Y., Konami, S., Takemura, N., Kelley, R. L., Kilbourne, C. A., Leutenegger, M. A., Porter, F. S., Eckart, M. E., and Szymkowiak, A.: Poisson vs. Gaussian statistics for sparse X-ray data: Application to the soft X-ray spectrometer, *Publications of the Astronomical Society of Japan*, 71, <https://doi.org/10.1093/pasj/psz053>, 2019.
- 710 Yin, X., Qin, J., and Paxton, L. J.: A new strategy for ionospheric remote sensing using the 130.4/135.6 nm airglow intensity ratios, *Earth and Planetary Physics*, 7, 445–459, <https://doi.org/10.26464/epp2023042>, 2023.
- Zesta, E. and Huang, C. Y.: Satellite orbital drag, in: *Space weather fundamentals*, pp. 329–351, CRC Press, 2016.
- Zhang, Y., Paxton, L. J., Morrison, D., Wolven, B., Kil, H., Meng, C.-I., Mende, S. B., and Immel, T. J.: O/N₂ changes during 1–4 October 2002 storms: IMAGE SI-13 and TIMED/GUVI observations, *Journal of Geophysical Research: Space Physics*, 109, 715 <https://doi.org/https://doi.org/10.1029/2004JA010441>, 2004.

Zhang, Y., Paxton, L. J., and Schaefer, R. K.: Deriving Thermospheric Temperature From Observations by the Global Ultraviolet Imager on the Thermosphere Ionosphere Mesosphere Energetics and Dynamics Satellite, *Journal of Geophysical Research: Space Physics*, 124, 5848–5856, 2019.