Response to Comments from Reviewer 2

This study presents an evaluation of the TanSat-2 satellite's capabilities for dual-band solar-induced chlorophyll fluorescence (SIF) retrieval through spectral simulations and end-to-end orbit simulations. The research demonstrates a well-structured approach, rigorous methodological design, and scientifically valuable findings. It is recommended for acceptance after addressing the following revisions:

Thanks a lot for your encouraging words and helpful comments. We have carefully revised the manuscript according to your comments and suggestions, especially on the analysis of the potential biases or regional variations in retrieval errors and the descriptions of the data preprocessing steps. The responses are in blue font, and the relevant revised parts of the manuscript are attached in purple font.

Comments 1. Line numbers should be consistent in the whole manuscript not start with 1 for each page.

Thank you for bringing to our attention the need for consistent line numbering. We have applied uniform line numbering throughout the entire manuscript to ensure clarity and ease of reference. We appreciate your feedback and hope that this adjustment enhances readability.

Comments 2. For section 4.2: The manuscript lacks detailed descriptions of data preprocessing steps, particularly in the end-to-end orbit simulations. For example, cloud contamination were not explicitly modeled, and the impact of cloud screening on SIF retrieval accuracy remains unclear. It is recommended to supplement details on cloud screening strategies and their implications for retrieval performance.

Thank you for your valuable comments. We have provided a more detailed description of the data preprocessing steps in the simulation experiments section, specifically in Sections 2.2.1 and 2.2.3. The issue of cloud interference is indeed of significant importance. Cloud presence can substantially affect radiative transfer processes and, consequently, SIF retrievals. In our study, we did not explicitly model cloud contamination or use cloud cover products for screening. The atmospheric radiative transfer framework based on MODTRAN is not well-suited for accurately simulating the effects of cloud contamination. Our primary objective is to highlight the advantages of the TanSat-2 sensor's enhanced spectral capabilities and its high temporal and spatial resolution for SIF retrieval. As such, our end-to-end orbital simulation assumes clear-sky conditions without incorporating cloud effects. Furthermore, the CAPHI instrument onboard TanSat-2 will provide aerosol optical depth (AOD) and cloud coverage information, which will enhance our understanding of atmospheric conditions and support future SIF retrievals with TanSat-2. These considerations are also described in Section 2.2.3.

In Section 2.2.1:

The Soil Canopy Observation Photochemistry and Energy Flux (SCOPE) model (van der Tol et al., 2009) is capable of simulating vegetation canopy reflectance spectra and SIF under diverse canopy structures and leaf biochemical conditions,

including leaf optical properties (e.g., chlorophyll content, dry matter) and canopy structural parameters (e.g., leaf area index, canopy height). Atmospheric radiative transfer functions were derived from the Moderate-resolution Atmospheric TRANsmission model (MODTRAN5; Berk et al., 1998, 2000) to generate TOA radiance. Critical processing steps included: (1) using the MODTRAN Interrogation Technique (MIT) (Verhoef and Bach, 2012; Verhoef et al., 2018) to extract 18 spectral transfer functions parameterized by aerosol optical depth, water vapor content, and observation geometry from MODTRAN5 outputs; (2) using the RTMo module in SCOPE, which dynamically couples these functions with bidirectional reflectance distribution (BRDF) and SIF emission spectra to resolve surface-atmosphere interactions. The simulated atmospheric parameters, canopy reflectance, and SIF signals were then integrated using the radiative transfer operator defined in Equation 1 to generate the TOA radiance spectra across the 640–850 nm range.

In Section 2.2.3:

It should be noted that we did not explicitly simulate cloud contamination or utilize cloud cover products for screening purposes. As such, our end-to-end orbital simulation operates under the assumption of clear sky conditions, deliberately excluding the potential impacts of cloud cover. Furthermore, the CAPHI instrument aboard the satellite will supply data on AOD and cloud coverage, enhancing our understanding of atmospheric conditions and cloud dynamics, while also providing supplementary data for future SIF retrieval efforts with TanSat-2. Furthermore, our simulations exclude rotational Raman scattering (RRS) effects. The RRS effects will be relatively small in the spectral range of red and far-red band (Vasilkov et al., 2013). Fig. 4 presents pseudo-color composites using the near-infrared, red, and green bands of MCD43C4, comparing surface reflectance reconstructions with simulated TOA radiances in seven geomorphologically distinct regions, including desert, boreal forest, and tropical rainforest ecosystems.

Comments 3. For page 16: The manuscript mentions that differences between retrieved SIF and GOSIF inputs were generally within 0.15 mW m^{-2} sr^{-1} nm^{-1}, yet it does not explore potential biases or regional variations in retrieval errors. A more detailed discussion on possible sources of discrepancies (e.g., land cover types, atmospheric effects) would strengthen the validity of the results.

Thank you for your comments. We have provided a more detailed description in the current version of the manuscript regarding potential biases in SIF retrieval errors, their regional variations, and possible sources of these errors (Section 4.2).

In Section 4.2:

The retrieval uncertainties manifest predominantly as spectrally structured noise coupled with pronounced sensitivity to atmospheric scattering processes. The former primarily varies with scene-specific radiance magnitudes within retrieval windows, while systematic SIF underestimation in high aerosol-loading

regimes arises from unaccounted scattering effects within our forward model. In order to disentangle the error sources, we conducted a statistical analysis of the far-red SIF retrieval errors (results for the red band are similar) in relation to AOD, the albedo within the fitting window, SZA, and VZA, as shown in Fig. 15. The results indicate that RMSE increases significantly with the increase in surface albedo and the decrease in SZA. Both changes enhance the background radiance, leading to a noticeable rise in retrieval uncertainty due to an increase in signal noise. Spatially, these retrieval errors dominate in bright surface areas, such as the Sahara Desert and the Congo Basin, as shown in Fig. 13. Meanwhile, retrieval bias exhibits substantial amplification under higher AOD and larger VZA. This phenomenon arises because the higher AOD strengthens atmospheric scattering efficiency, while larger VZA values extend the effective radiative path length. These factors collectively amplify atmospheric scattering effects, resulting in progressively larger underestimation of SIF. The spatial pattern of this bias prominently features regions with high aerosol, particularly Central and South Asia, as depicted in Fig. 13. Notably, red-SIF retrieval over aquatic environments necessitates distinct processing techniques compared to terrestrial environments. This technical disparity manifests as substantially exaggerated SIF estimates within specific watersheds, particularly those in northern/western Russia and the Great Lakes region of North America.

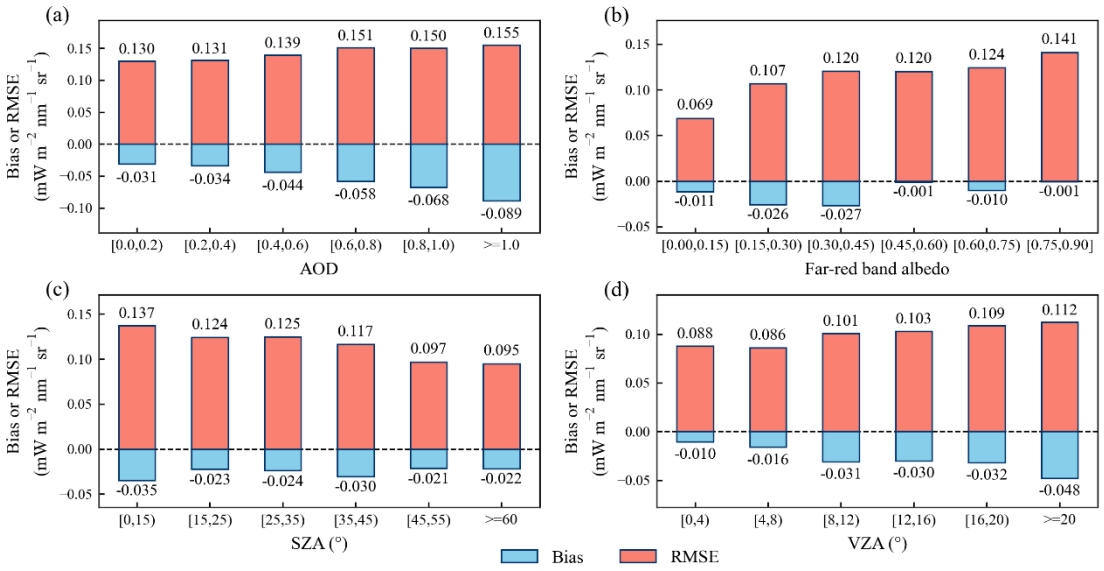

Figure 15. Statistical analysis of far-red SIF retrieval errors with respect to (a) AOD, (b) far-red band albedo, (c) SZA, and (d) VZA. The bias represents the mean value of ΔSIF (retrieved SIF - true SIF).

Comments 4. For appendix Figure A2 & A4: The mentioned figures showing 4-day and 8-day composites. They are mentioned in the results but are not adequately described in the main text. It is recommended to integrate a brief discussion of their significance and ensure that all supplementary figures are clearly referenced.

Thank you for your comments. The results of the 4-day and 8-day composites are described in more detail in Section 4.2 of the main text. Figures A1 to A4 have been combined into two figures, with clear references to them.

In Section 4.2:

Moreover, we evaluated the results of 4-day and 8-day composites, as shown in Fig A1 and A2. Compared to single-day observations, the global coverage is more comprehensive, and the composite method significantly reduces retrieval errors by increasing observation density and suppressing noise (the RMSE for the 4-day and 8-day composites decreased by 35% and 47%, respectively). This further demonstrates the performance of the instrument on the TanSat-2 satellite. However, several persistent systematic biases were detected: underestimation in high AOD regions (Central Asia/South Asia) due to enhanced atmospheric scattering, overestimation in bright surfaces (Sahara Desert, Congo Basin) caused by radiation saturation, and overestimation of red SIF near water bodies (Russian rivers, Great Lakes in North America) due to residual surface reflectance effects. Overall, the 4-day and 8-day composites achieved excellent accuracy, demonstrating TanSat-2 robust retrieval performance that balances spatial coverage and accuracy. The R^2 values for the two channels were 0.95 and 0.79, and 0.97 and 0.85, respectively, while the RMSE values were 0.053 and 0.041, and 0.043 and 0.034 mW m^{-2} sr^{-1} nm^{-1}. It should be noted that although the AOD product used partially accounts for cloud impacts, it does not explicitly model cloud contamination or apply cloud fraction products for screening. Consequently, our results assume clear-sky conditions. Additionally, these simulations do not incorporate RRS, as the filter only considers data with a SZA less than 70°, where the RRS effects are minimal.

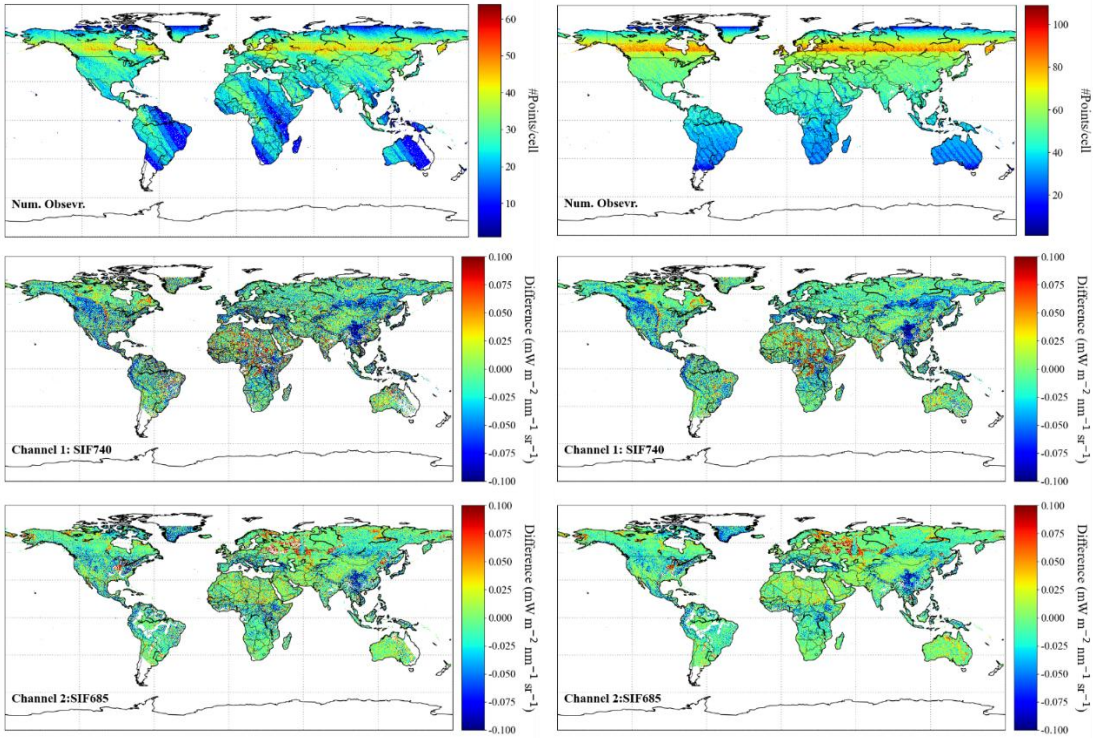

Figure A1. Global maps of TanSat-2 observation counts (top panels) and retrieval

errors (retrieved SIF minus reference SIF) for Channel l (middle panels) and Channel 2 (bottom panels). Results derive from 4-day (left column) and 8-day (right column) composites at 0.05° grid resolution.

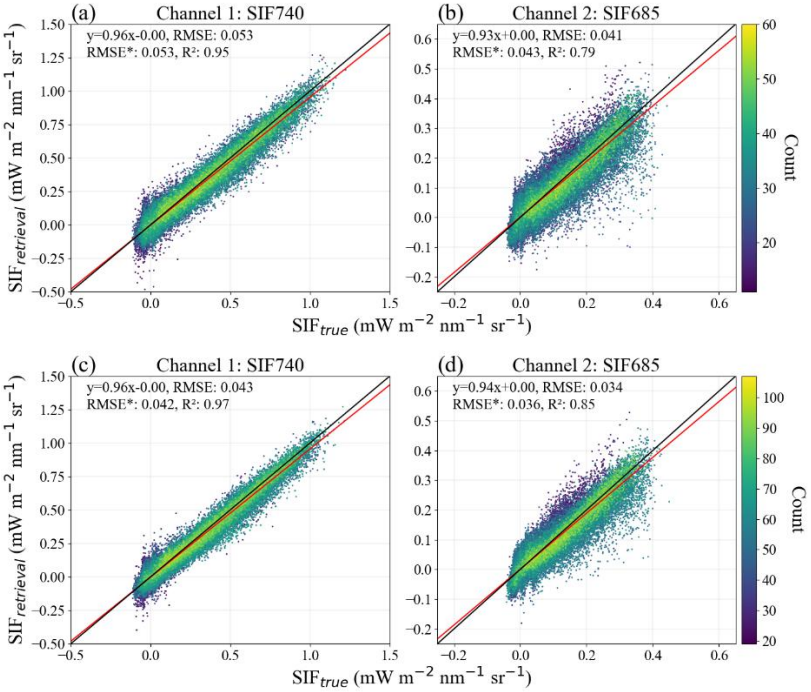

Figure A2. Retrieved versus reference SIF relationships for TanSat-2's two channels derived from 4-day composites (a, b) and 8-day composites (c, d). Black lines denote the 1:1 relationship, while red lines indicate linear regression fits.

Comments 5. While the manuscript is well-structured, some sections contain lengthy technical descriptions that could benefit from clearer segmentation. Additionally, minor grammatical inconsistencies exist, particularly in the descriptions of retrieval equations. Comprehensive language polishing is advised to improve readability and ensure consistency in technical terminology.

Thank you for your comments. We have reorganized the lengthy technical descriptions into more coherent sections and performed a thorough language revision throughout the manuscript. The following are examples of our revisions in Sections 2.2.3 and 3.1.

In Section 2.2.3:

Before revisions:

Reflectance spectra calculations utilized the green, red, and near-infrared bands from the Moderate Resolution Imaging Spectroradiometer (MODIS) MCD43C4 product (Huete et al., 2002). To accurately reflect real-world conditions, different simulation strategies for vegetation and non-vegetation surfaces were implemented. Global screening employed NDVI data from the MODIS MOD13C1 product (Huete et al., 2002), using a threshold of 0.2 to differentiate between non-green and green vegetation. Non-vegetation reflectance spectra were fitted using quadratic polynomials. Meanwhile, vegetation reflectance spectra, due to their complexity in these bands, were simulated using the data from Section 2.1. A singular value decomposition (SVD) was performed, and the first two basis

vectors were extracted to fit the spectral data, as illustrated in Fig. 3, with interpolation across the 640-800 nm range. For SIF spectra, the global OCO-2 SIF dataset (GOSIF) (Li et al., 2018) was used to map the global SIF distribution, filling data gaps through interpolation and depicting SIF spectral shape with a typical profile shown in Fig. 1.

After revisions:

The spectral reflectance characterization leveraged multi-band observations from the Moderate Resolution Imaging Spectroradiometer (MODIS) MCD43C4 product (Huete et al., 2002), incorporating green (555 nm), red (659 nm), and near-infrared (858 nm) spectral bands. To account for surface heterogeneity, distinct spectral simulation approaches were developed for vegetated and non-vegetated surfaces. Vegetation coverage was delineated using the MODIS MOD13C1 Normalized Difference Vegetation Index (NDVI) product (Huete et al., 2002), with a NDVI threshold of 0.2 discriminating non-photosynthetic vegetation from active canopies.

Non-vegetated surface reflectance spectra were parameterized through quadratic polynomial fitting, while vegetation spectra required advanced spectral decomposition due to their complex radiative interactions. As demonstrated in Fig. 3, vegetation reflectance characteristics between 640-800 nm were reconstructed through singular value decomposition (SVD) of the reference spectra described in Section 2.1, retaining the first two orthogonal basis vectors that collectively explained >95% of spectral variance. SIF spatial patterns were derived from the global OCO-2-based SIF product (GOSIF v2, Li et al., 2018), with spectral continuity achieved through adaptive gap-filling interpolation and depicting SIF spectral shape with a typical profile shown in Fig. 1.

Before revisions:

Datasets representing the global distribution of atmospheric conditions were assembled, incorporating water vapor data sourced from the ERA5, the fifth generation reanalysis of global climate and weather by the European Centre for Medium-Range Weather Forecasts (ECMWF) (Hersbach et al., 2020), and aerosol information from ECMWF's Atmospheric Composition Reanalysis 4 (EAC4), provided by the Copernicus Atmosphere Monitoring Service (CAMS) (Inness et al., 2019). Additionally, the 60-arcsecond resolution DEM data from Earth TOPOgraphy (ETOPO), furnished by the National Oceanic and Atmospheric Administration (NOAA) (Amante and Eakins, 2009), was resampled to a spatial resolution of 0.02 degrees to align with the satellite payload specifications. To accommodate the atmospheric conditions data, we expanded the range of atmospheric parameters outlined in Table 4, with further details provided in Table 5. These data are essential for calculating parameters such as ρ_0, S, $T_{\downarrow\uparrow}$, and $T_{\uparrow}$ in Equation 1, which are crucial for simulating TOA radiance. To optimize simulation times, we eschewed the use of MODTRAN 5 for per-pixel atmospheric parameter simulation, opting instead for a random forest model. We sampled 10% of the atmospheric conditions from Table 5—a total of 7,680 data points—and input them into MODTRAN 5 to simulate key atmospheric parameters. These parameters were then used to train the random forest

model, which was subsequently employed to simulate the atmospheric parameters for each pixel. These simulations were incorporated into Equation 1 to compute TOA radiance. Additional processes, including convolution and noise addition, were applied. Fig. 4 illustrates a pseudo-color image synthesized using the near-infrared, red, and green bands of MCD43C4, displaying the fitted reflectance and TOA radiances across seven representative geomorphic areas.

After revisions:

The atmospheric parameterization framework was constructed using a synthesis of global datasets to characterize key environmental variables. Total column water vapor data were obtained from ERA5 (the fifth generation ECMWF reanalysis; Hersbach et al., 2020), while aerosol information was acquired from the EAC4 reanalysis (Copernicus Atmosphere Monitoring Service; Inness et al., 2019). Topographic elevation data were derived from the ETOPO digital elevation model (NOAA; Amante and Eakins, 2009). Atmospheric temperature profiles were partitioned according to latitude. The atmospheric parameter space outlined in Table 4 was expanded (see Table 5) to compute critical radiative transfer variables (ρ_0, S, $T_{\downarrow\uparrow}$, and $T_{\uparrow}$ in Equation 1) required for simulating TOA radiance. To address computational constraints associated with full-resolution MODTRAN 5 radiative transfer modeling, a machine learning surrogate strategy was implemented. A representative subset of 7,680 atmospheric scenarios (10% of parameter combinations in Table 5) was simulated using MODTRAN 5. These simulations trained a random forest regressor to predict atmospheric parameters across the global domain at 0.02° resolution. These simulations were incorporated into Equation 1 to compute TOA radiance. Additional processes, including convolution and noise addition, were applied.

It should be noted that we did not explicitly simulate cloud contamination or utilize cloud cover products for screening purposes. As such, our end-to-end orbital simulation operates under the assumption of clear sky conditions, deliberately excluding the potential impacts of cloud cover. Furthermore, the CAPHI instrument aboard the satellite will supply data on AOD and cloud coverage, enhancing our understanding of atmospheric conditions and cloud dynamics, while also providing supplementary data for future SIF retrieval efforts with TanSat-2. Furthermore, our simulations exclude rotational Raman scattering (RRS) effects. The RRS effects will be relatively small in the spectral range of red and far-red band (Vasilkov et al., 2013). Fig. 4 presents pseudo-color composites using the near-infrared, red, and green bands of MCD43C4, comparing surface reflectance reconstructions with simulated TOA radiances in seven geomorphologically distinct regions, including desert, boreal forest, and tropical rainforest ecosystems.

In Section 3.1:

Before revisions:

For spectral characterization of SIF, the parameters λ_0 and σ_h for the far-red band are set at 740 nm and 21 nm, respectively. The red band utilizes a combination of two

Gaussian functions to depict a more intricate spectral shape, with λ_0 being 740 nm and 685 nm, and σ_h being 21 nm and 10 nm, respectively (Joiner et al., 2016; Zou et al., 2022). This spectral region encapsulates numerous solar Fraunhofer lines and atmospheric absorption features, enhancing the retrieval capabilities of SIF. Specifically, the spectral domain of TanSat-2 encompasses absorption lines such as the O_2-A at 758-772 nm and the O_2-B at 682-692 nm, which are integral for retrieving far-red and red SIF, respectively (Joiner et al., 2013; Guanter et al., 2015). Surrounding solar Fraunhofer and atmospheric absorption lines also play a crucial role in the SIF retrieval process in satellites with refined spectral resolution (Frankenberg et al., 2011; Joiner et al., 2011). As part of a semi-empirical approach, the performance of data-driven algorithms heavily relies on the empirical parameters used in the model. To optimize these parameters, different window settings were applied for each channel. The window settings for the far-red band were 747-758 nm, 759-772 nm, and 747-777 nm; for the red band, they were 672-686 nm, 682-697 nm, and 672-702 nm. Moreover, the permissible ranges for the parameters n_p (0-7) and n_v (1-50) were established for SIF retrieval (Köhler et al., 2015; Zou et al., 2022). Ultimately, only the parameters β_j, γ_k, and F_s remained as variables, with F_S determined through resolving the linear least squares problem.

SVD efficiently transforms a large set of correlated variables into a streamlined set of uncorrelated components, known as singular vectors. These vectors are strategically arranged such that each successive vector accounts for progressively less signal variability, enabling a hierarchical representation of data. By leveraging the principal singular vectors, we can reconstruct similar signals and effectively filter out noise. The implementation of SVD was carried out on the training dataset from spectral simulations. Fig. 5 illustrates the first six basis vectors for two distinct channels. Each subplot also quantifies the explained variance associated with each basis vector in the simulations. Predominantly, the spectral variations within the fitting window arise from Fraunhofer lines and atmospheric absorption features. It is evident that the initial set of singular vectors encapsulates the majority of the spectral variance across all simulations, while none of the vectors correspond to the SIF spectral shape.

After revisions:

The spectral parameterization of SIF employed Gaussian functions with band-specific configurations. For the far-red band, a single Gaussian function was adopted with λ_0 of 740 nm and σ_h of 21 nm. The red band required a dual Gaussian representation to capture its complex spectral features, characterized by λ_0 values of 685 nm and 740 nm with corresponding σ_h values of 10 nm and 21 nm, respectively (Joiner et al., 2016; Zou et al., 2022). These spectral regions encompass critical atmospheric absorption bands—specifically, the O_2-A band (758–772 nm) for far-red SIF and the O_2-B band (682–692 nm) for red SIF retrieval (Guanter et al., 2015; Joiner et al., 2013). Surrounding solar Fraunhofer and atmospheric absorption lines also play a crucial role in the SIF retrieval process in satellites with refined spectral resolution (Frankenberg et al., 2011; Joiner et al., 2011).

As a semi-empirical approach, the performance of data-driven algorithms heavily

depends on the empirical parameters employed in the model. To optimize these parameters, different retrieval windows were selected for each channel. Specifically, the far-red band employed window settings of 747–758 nm, 759–772 nm, and 747–777 nm, while the red band used 672–686 nm, 682–697 nm, and 672–702 nm. Additionally, the permissible ranges for the parameters n_p (0–7) and n_v (1–50) were defined for SIF retrieval (Köhler et al., 2015; Zou et al., 2022). Through variable reduction, the final retrievals retained only β_j, γ_k, and F_s, resolved via linear least squares optimization.

SVD was applied to spectral simulation datasets to disentangle dominant signal components from noise. These vectors are strategically arranged such that each successive vector accounts for progressively less signal variability, enabling a hierarchical representation of data. The first six basis vectors accounted for over 98% of cumulative spectral variance, as quantified in Fig. 5. Predominantly, the spectral variations within the fitting window arise from Fraunhofer lines and atmospheric absorption features. Notably, none of the basis vectors exhibited correlation with intrinsic SIF spectral shapes, confirming the method's capability to isolate fluorescence signals from background radiative processes.