

# Automatic Optical Depth Parametrization in Radiative Transfer Model RTTOV v13 via LASSO-Induced Sparsity for Satellite Data Assimilation

Franklin Vargas Jiménez<sup>1,2</sup> and Juan Carlos De los Reyes<sup>1,2</sup>

<sup>1</sup>MODEMAT Research Center in Mathematical Modelling and Optimization, Quito, Ecuador.

<sup>2</sup>Department of Mathematics, Escuela Politécnica Nacional, Quito, Ecuador.

**Correspondence:** Juan Carlos De los Reyes (juan.delosreyes@epn.edu.ec)

**Abstract.** The assimilation of satellite spectral sounder data requires fast and accurate radiative transfer models. This study proposes a novel methodology to automatically parameterize atmospheric optical depths within the Radiative Transfer for TOVS (RTTOV) version 13 scheme using statistical thresholds across pressure levels and Least Absolute Shrinkage and Selection Operator (LASSO) regression to induce sparsity. Numerical experiments with Visible Infrared Imaging Radiometer Suite (VIIRS) infrared channels demonstrate that this approach significantly reduces computational costs while maintaining accuracy. The sparsity also facilitates the automatic selection of absorbing gases and predictors by channel and pressure level, making it particularly effective for multispectral instruments with numerous atmospheric variables. These findings highlight the potential of sparse regression methods to enhance the efficiency of radiative transfer models for satellite data assimilation.

## 1 Introduction

In satellite data assimilation and remote sensing retrievals, as well as their applications in numerical weather prediction (NWP), the radiative transfer equation (RT) is the main model used to retrieve global atmospheric variables, such as temperature and trace gases concentrations, including water vapor, ozone, carbon dioxide, and other atmospheric constituents. This is achieved by utilizing top of the atmosphere (TOA) radiance measurements from satellite sounders operating across different channels of the electromagnetic spectrum. The numerical implementation of the RT equation as a forward model can primarily be carried out using two approaches: Line-by-Line (LBL) Radiative Transfer models and Fast Radiative Transfer models (Fast-RT).

Line-by-line models simulate satellite radiance by rigorously integrating atmospheric physics and chemical phenomena. These models are highly accurate in replicating the precision of modern instruments, such as hyperspectral sounders like AIRS, CrIS and IASI. However, they are characterized by significant computational demands in terms of CPU time and memory, making them impractical for use in operational data assimilation. Some of the most well-known models in this category include: LBLRTM, developed at Atmospheric and Environmental Research, Inc. (AER) (Clough et al., 1992; Clough and Iacono, 1995; Clough et al., 2005); AMSUTRAN, developed at the Met Office (UK) (Turner et al., 2019); and GENLN2, developed at the National Center for Atmospheric Research (NCAR) (Edwards, 1992). A comparison between LBLRTM and GENLN2 is presented in (Matricardi, 2007). Another software worth mentioning is kCARTA (DeSouza-Machado et al., 2020),

a pseudo Line-by-Line model that uses precomputed and compressed physically intensive processes in RT model to compute  
25 radiances more quickly while maintaining accuracy.

On the other hand, the most common Fast-RT models estimate the expected radiance in a channel (what a sensor actually measures) and are typically based on statistical approaches. In these models, the complex and computationally costly physical processes of RT modeling, the calculation of atmospheric transmittances, are parameterized using statistical models and trained with output from Line-by-Line software on real atmospheric profile databases. The parameters are adjusted using standard  
30 linear regression models or other machine learning techniques. While these methods sacrifice a small degree of accuracy, they significantly reduce computational costs, making them practical for use in operational data assimilation. Some of the most well-known models in this category include: OPTRAN, developed by the NESDIS-NCEP community (McMillin et al., 1995; Kleespies et al., 2004; McMillin et al., 2006); The Joint Center for Satellite Data Assimilation (JCSDA) Community Radiative Transfer Model (CRTM) (Han et al., 2006; Chen et al., 2008); and the RTTOV model, see (Saunders et al., 2018)  
35 and the references therein. Other studies using statistical approaches include (Matricardi, 2010), which incorporates principal component analysis in RTTOV, as well as (Liu et al., 2006; Krishnan et al., 2012; Cao et al., 2021; Stegmann et al., 2022; Mauzeri et al., 2022; Su et al., 2023), which apply machine learning techniques for parametrization, feature reduction, and sampling strategies.

Even though RTTOV is more efficient than line-by-line models, it remains prohibitively expensive for operational use cases<sup>1</sup>.  
40 Indeed, in current Fast RT models based on linear regression, such as OPTRAN and RTTOV, training is performed separately for each gas type and pressure level, resulting in an over-parametrization of the RT model. To reduce the number of parameters and make the evaluation of the trained RT model less computationally expensive, it is essential to carefully select the most significant gases for each spectral channel of each instrument type, reduce the number of pressure levels, and implement other ad hoc strategies. These decisions must account for the large number of possible combinations and trade-offs, and are typically  
45 made by expert teams.

One promising approach to reducing the number of parameters without relying on expert committees is the use of optimization methods that induce sparsity in the parameters. In particular, the use of LASSO regression, a regularization method that penalizes the regression coefficients with the  $\ell_1$ -norm, has proven effective for variable selection and model complexity reduction in various large-scale applications (see, e.g., Heilemann et al., 2024; Pak et al., 2025). In the context of radiative  
50 transfer, LASSO regression was applied by (Cardall et al., 2023) to estimate water quality parameters such as clarity, temperature, and chlorophyll-a, based on correlations with in-situ measurements and near-coincident Landsat spectral data, with a focus on model explainability. In (Li et al., 2020), the authors proposed an algorithm for detecting hazardous clouds using passive infrared remote sensing technology with variable selection. Other studies that combine or compare LASSO with machine learning methods for remote sensing include: the removal of redundant features in PolSAR and optical images (Hong  
55 and Kong, 2021); estimation of aboveground forest biomass with variable selection (Wang et al., 2022a); identification of important environmental variables for retrieving soil moisture content (Wang et al., 2022b); evaluation of the accuracy and

---

<sup>1</sup>This is the case for Ecuador's METEO operational system, which currently relies on an HPC with only 700 cores.

generalization capacity of grassland models (Smith et al., 2023); and a comparison of different machine learning methods for predicting soybean yield (Joshi et al., 2023).

Building on this approach, in this paper we target the automatic selection of gases and optical depth predictors in Fast RT models by inducing sparsity in the weight predictors using LASSO regression. We propose a parametrization of transmittances based on statistical thresholds to automatically select the appropriate gases by channel and pressure level, and to induce sparsity in the parameters by replacing the classical regression problem with a LASSO problem within the RTTOV framework. The proposed methodology is tested with VIIRS infrared channels, and the results are compared with the standard RTTOV model. To the best of the authors’ knowledge, this is the first time that LASSO regression has been applied to the RTTOV model to automate the selection of gases and parameters.

One of the key aspects in LASSO models is the choice of the regularization weight in front of the  $\ell_1$ -norm. This weight controls the trade-off between fitting the training data well and keeping the model simple by reducing the number of non-zero coefficients. In our context, selecting an appropriate regularization weight is crucial for effectively identifying the most relevant gases and optical depth predictors while avoiding overfitting.

To establish a rigorous criterion for choosing this parameter—rather than relying on a tedious trial-and-error process—we propose a bilevel optimization approach (see, e.g., De los Reyes and Villacís (2022); De los Reyes (2023)). The idea is to formulate an upper-level optimization problem that encodes a model quality criterion, while the LASSO problem serves as the lower-level constraint. In this article, we successfully test two types of loss functions: the first, based on an  $\ell_0$  seminorm that prescribes the number of non-zero predictors; and the second, inspired by a Bayesian Information Criterion-type objective.

This manuscript is organized as follows: Section 2 outlines the theoretical framework for the RT equation in Line-By-Line models and details the general scheme of Fast-RT methods, focusing on RTTOV. Section 3 introduces the proposed transmittance parametrization using statistical inference and LASSO regression model, as well as the bilevel optimization approach for selecting the regularization weight. Section 4 presents the experimental settings and numerical results comparing RTTOV with the proposed method. Finally, Section 5 offers conclusions of the performance of the proposed approach.

## 2 Radiative Transfer Equation

The monochromatic radiative transfer equation for the upwelling radiance in a clear sky, without solar radiation contribution, for a non-scattering atmosphere and in local thermodynamic equilibrium, is given by:

$$I(\nu, \theta) = \tau_s(\nu, \theta) \epsilon_s(\nu, \theta) B(\nu, T_s) + \int_{\tau_s}^1 B(\nu, T(p)) d\tau + (1 - \epsilon_s(\nu, \theta)) \tau_s^2(\nu, \theta) \int_{\tau_s}^1 \frac{B(\nu, T(p))}{\tau^2} d\tau, \quad (1)$$

where  $I(\nu, \theta)$  is the monochromatic TOA radiance at wavenumber  $\nu$  and satellite zenith angle  $\theta$ ;  $B(\nu, T)$  is the Planck function at temperature  $T$ ;  $\tau(\nu, \theta, p, T, q)$  denotes the layer-to-space transmittance dependent on pressure  $p$ , temperature  $T$ , and gas concentration  $q$ . Here,  $T_s$ ,  $\epsilon_s$ , and  $\tau_s$  represent surface skin temperature, emissivity, and transmittance respectively, (Weinreb et al., 1981).

The terms correspond to surface emission, upward atmospheric emission, and downward atmospheric emission reflected at the surface (assuming specular reflection). Surface emissivity can be close to 1 for  $\nu$  between 714-1250  $cm^{-1}$  and for surfaces  
90 such as bodies of water, ice and healthy plant leaves, carbon powder, allowing the last term to be discarded.

The model described above, computed for each wavenumber  $\nu$  is called the Line-by-Line model, and the resulting radiance is monochromatic.

Satellite-measured radiance is polychromatic, simulated by convolving Eq. (1) with the instrument's Normalized Spectral Response Function (NSRF):

$$95 \quad \bar{I}(\nu^*, \theta) = \int_{\nu_a}^{\nu_b} \phi(\nu, \nu^*) I(\nu, \theta) d\nu, \quad (2)$$

where  $\phi(\nu, \nu^*)$  is the NSRF, representing the sensitivity to radiance within the spectral channel  $[\nu_a, \nu_b]$ , with  $\nu^*$  representing the centroid of the response. Using the expression (2) in Eq. (1), the polychromatic radiance for the spectral channel identified with  $\nu^*$ , assuming  $\epsilon_s = 1$ , can be written as (see Weinreb et al., 1981):

$$\bar{I}(\nu^*, \theta) = \bar{\tau}_s(\nu^*, \theta) B(\nu, T_{es}) + \int_{\tau_s}^1 B(\nu^*, T_e(p)) d\bar{\tau}, \quad (3)$$

100 where  $T_{es}$  and  $T_e$  are empirical effective temperatures obtained via regression. The polychromatic transmittance is given by:

$$\bar{\tau}(\nu^*, \theta, p, T, q) = \int_{\nu_a}^{\nu_b} \phi(\nu, \nu^*) \tau(\nu, \theta, p, T, q) d\nu. \quad (4)$$

Transmittance follows Beer-Lambert law  $\tau = e^{-d}$ , with optical depth  $d(\nu, \theta, p, T, q)$  accounting for absorption by gases (e.g.,  $H_2O$ ,  $O_3$ ,  $CO_2$ ,  $CH_4$ ) and continuum effects. The monochromatic optical depth for a set of gases  $\mathbf{g}_1, \dots, \mathbf{g}_s$  is:

$$d(\nu, \theta, p, T, q) = -\frac{\sec(\theta)}{g} \sum_{l=1}^s \int_0^p K^{\mathbf{g}_l}(\nu, p', T(p')) q^{\mathbf{g}_l}(p') dp', \quad (5)$$

105 where  $g$  is gravitational acceleration,  $K^{\mathbf{g}_l}$  is the absorption function modeled via Voigt profiles (see Lavrentieva et al., 2011).

## 2.1 Fast Radiative Transfer Model

Fast RT models discretize the atmosphere into  $L$  layers:

$$p_0 < p_1 < \dots < p_L,$$

where  $p_0$  is the top-of-atmosphere and  $p_L$  the surface pressure. Polychromatic radiance Eq. (3) is computed numerically,  
110 requiring parameterization of polychromatic transmittance to reduce computational cost. In Fast-RT models, the polychromatic optical depth is parameterized and fitted via linear regression to approximate Eq. (5), following ideas from McMillin and Fleming (McMillin and Fleming, 1976; Fleming and McMillin, 1977; McMillin et al., 1979).

The polychromatic optical depth from layer  $i$  to the top of the atmosphere, for a single channel and gas  $\mathbf{g}_l$ , is:

$$d_i^{\mathbf{g}_l} = d_{i-1}^{\mathbf{g}_l} + \sum_{j=1}^{m_l} w_{ij}^{\mathbf{g}_l} X_{ij}^{\mathbf{g}_l}, \quad d_0^{\mathbf{g}_l} = 0, \quad i = 1, \dots, L, \quad (6)$$

115 where  $X_{ij}^{\mathbf{g}_l}$  are predictors depending on view angle, temperature, and gas concentration. The parameters  $w_{ij}^{\mathbf{g}_l}$  define the model. Appendix B provides details on the RTTOV v13 predictors, and further information can be found in Saunders et al. (2017).

This parametrization includes a fixed gas mixture—whose spatio-temporal variations minimally affect radiance—and variable gases, primarily  $\text{H}_2\text{O}$ , optionally including  $\text{O}_3$ ,  $\text{CO}_2$ ,  $\text{N}_2\text{O}$ ,  $\text{CO}$ ,  $\text{CH}_4$ , and  $\text{SO}_2$ , varying by channel. Water vapor absorption may be split into line and continuum components.

120 The polychromatic transmittance of layer  $i$  for gas  $\mathbf{g}_l$  is approximated by:

$$\tau_{(i,0)}^{\mathbf{g}_l} = \exp(-d_i^{\mathbf{g}_l}), \quad (7)$$

and total transmittance approximated as:

$$\tau_{(i,0)}^{TOT} = \prod_{l=1}^s \tau_{(i,0)}^{\mathbf{g}_l}. \quad (8)$$

Parameters are fitted using a database of  $M$  vertical atmospheric profiles:

$$125 \quad (p_i, T_{ij}, q_{ij}^{\mathbf{g}_1}, \dots, q_{ij}^{\mathbf{g}_s}), \quad i = 0, \dots, L, \quad j = 1, \dots, M,$$

with polychromatic transmittances computed with Line-by-Line software for  $N$  view angles  $\theta_k$ :

$$(\tau_{ijk}^{\mathbf{g}_1}, \dots, \tau_{ijk}^{\mathbf{g}_s}), \quad i = 1, \dots, L, \quad j = 1, \dots, M, \quad k = 1, \dots, N. \quad (9)$$

Since total polychromatic transmittance is not simply the product of individual gases transmittances (unlike the monochromatic case), data (9) are corrected following (Xiong and McMillin, 2005; McMillin et al., 2006), as in RTTOV v13 (Hocking  
130 et al., 2021), by introducing a corrective term  $\tau_{(i,0)}^{COR}$ :

$$\tau_{(i,0)}^{TOT} = \tau_{(i,0)}^{COR} \prod_{l=1}^s \tau_{(i,0)}^{\mathbf{g}_l}, \quad (10)$$

which is parameterized similarly to Eq. (6) and (7). The corrective transmittance for training is:

$$\tau_{ijk}^{COR} = \frac{\tau_{ijk}^{TOT}}{\hat{\tau}_{ijk}^{TOT}},$$

where  $\tau_{ijk}^{TOT}$  is the Line-by-Line polychromatic transmittance including all absorbers, and  $\hat{\tau}_{ijk}^{TOT}$  is the modeled transmittance  
135 from Eq. (8).

The linear regression fitting problem for gas  $\mathbf{g}_l$  and layer  $i$  is:

$$(LS_{\mathbf{g}_l}) \quad \min_{\mathbf{w}_i^{\mathbf{g}_l} \in \mathbb{R}^{m_l}} \frac{1}{2MN} \|A_i^{\mathbf{g}_l} \mathbf{w}_i^{\mathbf{g}_l} - Y_i^{\mathbf{g}_l}\|_2^2, \quad (11)$$

where  $A_i^{\mathbf{g}^l} \in \mathbb{R}^{MN \times m_l}$  contains predictors  $X_{ij}^{\mathbf{g}^l}$  for angles, temperatures, and concentrations across profiles, and  $Y_i^{\mathbf{g}^l} \in \mathbb{R}^{MN}$  the corresponding optical depths.

140 In RTTOV v13, parameter counts per channel reach nearly 11,000, considering variable and fixed gases, layers, and corrections. Reduction is achieved by expert-based gas selection, layer thinning, and thresholding (see Saunders et al., 2017).

### 3 A Sparse Parametrization of Optical Depths

In this section, we present a methodology to significantly reduce the number of parameters used in optical depth parametrization within the RTTOV v13 framework. The methodology involves automatically selecting absorbing gases per channel and pressure  
145 level, as well as identifying the most important predictors for each atmospheric layer. This approach induces sparsity in the regression parameters by combining two tools: statistical inference to determine whether a given gas at a particular layer requires no parametrization, a parametrization with a single predictor, or a more complex parametrization as described in Eq. (6). In the latter case, the classic linear regression problem is replaced with a LASSO regression problem to select predictors and induce sparsity in the parameter vectors.

#### 150 3.1 Parametrization Based on Statistical Inference

The aim here is to preprocess the data of the polychromatic transmittances in a channel to determine which atmospheric layers require optical depth parametrization and to automatically exclude gases that do not significantly contribute to the radiance absorption in that channel. To achieve this, we will use confidence intervals to estimate the true polychromatic transmittances.

For a gas  $\mathbf{g}_l$  or correction term in a fixed layer  $i$ , we construct a confidence interval for the mean of the polychromatic  
155 transmittances of the layer  $i$ . This is given by:

$$[\bar{\tau}_i^{\mathbf{g}^l} - E_i^{\mathbf{g}^l}, \bar{\tau}_i^{\mathbf{g}^l} + E_i^{\mathbf{g}^l}]$$

where

$$E_i^{\mathbf{g}^l} = Z_{1-\frac{\alpha}{2}} \frac{s_i^{\mathbf{g}^l}}{\sqrt{NM}},$$

$\bar{\tau}_i^{\mathbf{g}^l}$  is the mean polychromatic transmittance for layer  $i$ , considering  $N$  angles and  $M$  atmospheric profiles,  $s_i^{\mathbf{g}^l}$  is the corre-  
160 sponding standard deviation, and  $Z_{1-\frac{\alpha}{2}}$  is the critical value of a distribution for a confidence level of  $1 - \alpha$ . Given that the number of data points in each layer is  $NM$ , which is usually sufficiently large (in our experiments, for  $N = 6$  angles and  $M = 83$  profiles,  $NM = 498$ ), the standard normal distribution is used to obtain the critical value. Thus, the absolute error in approximating the true value of the polychromatic transmittance of gas  $\mathbf{g}_l$  in layer  $i$  with  $\bar{\tau}_i^{\mathbf{g}^l}$  is at most  $E_i^{\mathbf{g}^l}$ , with a probability of  $\alpha$  that the absolute error exceeds this value. In our case, the confidence level is set to  $\alpha = 10^{-6}$ .

165 Based on the above, the following statistical thresholds for optical depth parametrizations are proposed. Let  $\epsilon_1$  and  $\epsilon_2$  be positive and sufficiently small values, these will be used as thresholds to determine whether  $\bar{\tau}_i^{\mathbf{g}^l}$  is close to the true value or close to 1. Define the mean optical depth for layer  $i$  as  $\bar{d}_i^{\mathbf{g}^l} = -\ln(\bar{\tau}_i^{\mathbf{g}^l})$ , and consider the following three cases:

- *Case I*: If  $E_i^{\mathbf{g}_l} > \epsilon_1$ , the polychromatic transmittance due to gas  $\mathbf{g}_l$  in layer  $i$  has high variability with respect to the value of the atmospheric variables in that layer. In this case, the optical depth parametrization follows as in Eq. (6) for layer  $i$ .
- *Case II*: If  $E_i^{\mathbf{g}_l} \leq \epsilon_1$  and  $\bar{d}_i^{\mathbf{g}_l} > \epsilon_2$ , unlike the previous case, the polychromatic transmittance due to gas  $\mathbf{g}_l$  in layer  $i$  has low variability with respect to the value of the atmospheric variables in that layer, and can be estimated by  $\bar{\tau}_i^{\mathbf{g}_l}$ , but is not close to 1. Thus, the optical depth can be parameterized with a single predictor as follows:

$$d_i^{\mathbf{g}_l} = d_{i-1}^{\mathbf{g}_l} + w_{i0}^{\mathbf{g}_l} X_{i0}^{\mathbf{g}_l},$$

where,  $X_{0i} = 1$  and  $w_{i0}^{\mathbf{g}_l} = \bar{d}_i^{\mathbf{g}_l}$ . If this occurs in all layers, and since the parametrization does not depend on atmospheric variables, the gas  $\mathbf{g}_l$  can be included with fixed gases.

- *Case III*: If  $E_i^{\mathbf{g}_l} \leq \epsilon_1$  and  $\bar{d}_i^{\mathbf{g}_l} \leq \epsilon_2$ , the polychromatic transmittance in layer  $i$  can not only be estimated by  $\bar{\tau}_i^{\mathbf{g}_l}$  but is also close to 1, meaning that gas  $\mathbf{g}_l$  does not cause significant absorbance in this layer. The relative error of approximating  $\bar{\tau}_i^{\mathbf{g}_l}$  with 1 is given by:

$$\frac{1 - \bar{\tau}_i}{\bar{\tau}_i} = e^{\bar{d}_i^{\mathbf{g}_l}} - 1 = \bar{d}_i^{\mathbf{g}_l} e^{\xi} \leq \epsilon_2 e^{\epsilon_2} \approx \epsilon_2,$$

for some  $\xi \in (0, \bar{d}_i^{\mathbf{g}_l})$ . If this condition is met for all layers, then gas  $\mathbf{g}_l$  is automatically discarded.

To summarize the above, the parametrization of optical depths based on statistical thresholds is as follows:

$$d_0^{\mathbf{g}_l} = 0$$

$$d_i^{\mathbf{g}_l} = d_{i-1}^{\mathbf{g}_l} + \begin{cases} \sum_{j=1}^{m_l} w_{ij}^{\mathbf{g}_l} X_{ij}^{\mathbf{g}_l}, & E_i^{\mathbf{g}_l} > \epsilon_1, \\ \bar{d}_i^{\mathbf{g}_l}, & E_i^{\mathbf{g}_l} \leq \epsilon_1 \text{ and } \bar{d}_i^{\mathbf{g}_l} > \epsilon_2, \\ 0, & \text{otherwise,} \end{cases} \quad (12)$$

for  $i = 1, 2, \dots, L$ . The transmittances from layer  $i$  to the top of the atmosphere are still calculated using (7).

The statistical threshold tolerances  $\epsilon_1$  and  $\epsilon_2$  should be sufficiently small. In our experiments, we set  $\epsilon_2 = \epsilon_1$  and evaluate the model performance for different small values of  $\epsilon_1$ .

### 3.2 LASSO Regression and optimal choice of regularization parameter

After discarding parameter groups using the previous statistical approach with Eq. (12), in *Case I*, the remaining parameters are typically estimated by solving an ordinary least squares (OLS) problem, which involves a large number of parameters.

To reduce the number of parameters, we propose to induce sparsity in the parameter vector  $\mathbf{w}_i^{\mathbf{g}_l}$  by solving the LASSO problem. This is done by replacing the OLS problem (11) with the following optimization problem:

$$(LASSO_{\mathbf{g}_l}) \quad \mathbf{w}_i^{\mathbf{g}_l} = \arg \min_{\mathbf{w} \in \mathbb{R}^{m_l}} \mathcal{L}_\lambda(\mathbf{w}) \quad (13)$$

where

$$\mathcal{L}_\lambda(\mathbf{w}) = \frac{1}{MN} \|A_i^{\mathbf{g}_l} \mathbf{w} - Y_i^{\mathbf{g}_l}\|_2^2 + \lambda \|\mathbf{w}\|_1$$

195 and  $\lambda \geq 0$  is the regularization parameter. As  $\lambda \rightarrow +\infty$ , high sparsity is induced, and as  $\lambda \rightarrow 0$ , sparsity is low. Specifically, if  $\lambda = 0$ , the problem reduces to the least squares problem (11).

The regularization parameter  $\lambda$  has to be carefully selected to ensure that the approximation of the transmittance in layer  $i$  maintains a high level of accuracy relative to the least squares solution (11), while achieving a model with fewer parameters. Although standard techniques such as cross-validation exist for tuning  $\lambda$ , they may not always be appropriate, especially when  
200 alternative loss criteria are more relevant to the specific modeling goals. To address this choice, we adopt a bilevel optimization approach (see, e.g., (De los Reyes and Villacís, 2022; De los Reyes, 2023)), where the LASSO problem forms the lower-level constraint and the upper-level objective reflects a model quality criterion. This results in the following bilevel problem:

$$\begin{aligned} \min_{\lambda \in \mathbb{R}} \quad & F(\mathbf{w}(\lambda)) \\ \text{s.t.} \quad & \mathbf{w}(\lambda) = \arg \min_{\mathbf{w} \in \mathbb{R}^{m_l}} \mathcal{L}_\lambda(\mathbf{w}), \\ & 0 \leq \lambda \leq \lambda_0, \end{aligned} \tag{14}$$

where  $\lambda_0 > 0$  is a given upper bound. In the following, we show how to reduce this bilevel problem to a standard nonlinear  
205 optimization problem. For the sake of clarity, we omit the indices corresponding to gas and pressure level.

Under the assumption that matrix  $A$  is full rank, problem (13) has a unique solution for each  $\lambda \geq 0$ , denoted by  $\mathbf{w}(\lambda)$ . The collection of these solutions as  $\lambda$  varies over the positive real numbers, is called the *regularization path*  $\mathcal{P} = \{\mathbf{w}(\lambda) : \lambda > 0\}$ . A key structural property of the regularization path is that it is well-defined, unique, and continuous piecewise linear. Moreover, and it can be computed using the homotopy algorithm for the LASSO problem (Osborne et al., 2000), an algorithm with  
210 exponential complexity but low computational cost, that returns the vertices of the regularization path; both the properties and the algorithm are described in Mairal and Yu (2012). The algorithm constructs a finite, monotonically decreasing sequence of values  $\{\lambda_k\}_{k=0}^r$ , with  $\lambda_0 = \|A^T Y\|_\infty$  and  $\lambda_r = 0$ . For each  $\lambda_k$  in this sequence, the corresponding solution to the LASSO problem,  $\mathbf{w}(\lambda_k)$ , is a vertex of the regularization path  $\mathcal{P}$ , and it can be shown that  $\mathbf{w}(\lambda) = 0$ , for all  $\lambda \geq \lambda_0$ . In each line segment of this path, the sparsity pattern of  $\mathbf{w}(\lambda)$  does not change; that is, the support set  $\mathcal{S}(\mathbf{w}(\lambda)) = \{j \in \{1, 2, \dots, m_l\} : \mathbf{w}_j(\lambda) \neq 0\}$   
215 remains fixed for all  $\lambda \in (\lambda_{k+1}, \lambda_k]$ .

These properties of the regularization path allow the bilevel problem (14) to be reduced to a standard one-dimensional optimization problem with bound constraints:

$$\min_{\lambda \in [0, \lambda_0]} F(\mathbf{w}(\lambda)). \tag{15}$$

We still need to establish the upper-level loss function  $F$ , which serves as a model quality criterion for the LASSO regulariza-  
220 tion path. To this end, we propose two formulations for empirical comparison: the first is based on the optimal selection of the regularization parameter in the LASSO problem using an  $\ell_0$ -regression cost function; the second is based on a well-established statistical tool for optimal model selection, the Bayesian Information Criterion (BIC).

### 3.2.1 Bilevel formulation based on the $\ell_0$ regression

The best subset selection problem (Bertsimas et al., 2016; Miller, 2002) consists of solving a least squares formulation that allows explicit control of sparsity through the choice of the number of predictors, this is:

$$\begin{aligned} \min_{\mathbf{w} \in \mathbb{R}^{m_l}} \quad & \frac{1}{MN} \|\mathbf{A}\mathbf{w} - \mathbf{Y}\|_2^2 \\ \text{s.t.} \quad & \text{card}(\mathcal{S}(\mathbf{w})) \leq k, \end{aligned} \tag{16}$$

for  $k \in \{1, \dots, m_l\}$  given. As this problem is NP-hard, the computational loss can be prohibitive, especially when several subset sizes must be tested without prior knowledge of  $k$ . To mitigate this, more tractable relaxations have been proposed, such as the  $\ell_0$  regression, obtained from a penalized formulation of the problem (16):

$$\min_{\mathbf{w} \in \mathbb{R}^{m_l}} \frac{1}{NM} \|\mathbf{A}\mathbf{w} - \mathbf{Y}\|_2^2 + \gamma(\text{card}(\mathcal{S}(\mathbf{w})) - k),$$

where  $\gamma > 0$  is the penalty parameter. Motivated by this problem, we propose as the merit function

$$F_{\ell_0}(\mathbf{w}) = \frac{1}{N_v M_v} \|\mathbf{A}_v \mathbf{w} - \mathbf{Y}_v\|_2^2 + \gamma(\beta) \text{card}(\mathcal{S}(\mathbf{w})),$$

which is used in the bilevel problem (14) to assess LASSO solutions on validation data, balancing generalization and complexity through a weighting parameter  $\beta \in (0, 1]$ . As a reference,  $\mathbf{w}(0)$  (the OLS solution) achieves the best fit but maximum complexity ( $\beta = 1$ ), while  $\mathbf{w}(\lambda_0) = 0$  is the opposite ( $\beta = 0$ ). Consequently, the penalty is defined as

$$\gamma(\beta) = \frac{1}{NM} \left( \frac{1}{\beta} - 1 \right) (\|\mathbf{Y}\|_2^2 - \|\mathbf{A}\mathbf{w}(0) - \mathbf{Y}\|_2^2).$$

To prioritize model data fidelity over low complexity,  $\beta$  should be close to 1. In our experimental setting, we choose  $\beta = 1 - 10^{-4}$ .

With  $F_{\ell_0}(\mathbf{w}(\lambda))$  as the objective function of (15), it is a piecewise continuous objective function, smooth along each linear segment of the regularization path and with discontinuities at  $\{\lambda_k\}_{k=0}^r$ . Moreover,  $F_{\ell_0}(\mathbf{w}(\lambda))$  is a quadratic polynomial for  $\lambda \in (\lambda_{k+1}, \lambda_k]$ , since  $\text{card}(\mathcal{S}(\mathbf{w}(\lambda)))$  remains constant within this interval. If we denote  $\tilde{\lambda}_k$  as the minimizer of this polynomial over the closure of this interval, then problem (15) reduces to a discrete parameter optimization problem:

$$\min_{0 \leq k \leq r} F_{\ell_0}(\mathbf{w}(\tilde{\lambda}_k)). \tag{17}$$

### 3.2.2 Bilevel formulation based on the Bayesian Information Criterion

In this case, the choice of the loss function  $F(\mathbf{w})$  is inspired by the Bayesian Information Criterion for model selection (Schwarz, 1978). Similar to  $\ell_0$ -regression, it penalizes model complexity but does not require a tuning parameter. Given a collection  $\mathcal{P}$  of candidate models, and letting  $\sigma(\mathbf{w})$  denote the maximum likelihood under model  $\mathbf{w} \in \mathcal{P}$ , the BIC-based objective is given by

$$\min_{\mathbf{w} \in \mathcal{P}} n \ln(\sigma(\mathbf{w})) + \ln(n) K(\mathbf{w}),$$

250 where  $K(\mathbf{w})$  denotes the number of explanatory variables in model  $\mathbf{w}$  (or a measure of model complexity), and  $n$  is the number of data points used to construct model  $\mathbf{w}$ . If the “true model” belongs to  $\mathcal{P}$ , then the probability that BIC selects this model approaches 1 as the number of data points increases.

In our context, the model set  $\mathcal{P}$  consists of LASSO solutions for each  $\lambda \geq 0$ , built using  $n = NM$  data points, and a good approximation to the true model is given by the ordinary least squares solution  $\mathbf{w}(0)$ . We define:

$$255 \quad \sigma(\mathbf{w}) = \frac{1}{NM} \|A\mathbf{w} - Y\|_2^2 \quad \text{and} \quad K(\mathbf{w}) = \text{card}(\mathcal{S}(\mathbf{w})), \quad \text{for } \mathbf{w} \in \mathcal{P},$$

The BIC-based objective function is then defined as:

$$F_{BIC}(\mathbf{w}) = NM \ln \left( \frac{1}{NM} \|A\mathbf{w} - Y\|_2^2 \right) + \ln(NM) \text{card}(\mathcal{S}(\mathbf{w})).$$

This is a piecewise continuous objective function, smooth along each linear segment of the regularization path and with discontinuities at  $\{\lambda_k\}_{k=0}^r$ . It can be verified that  $F_{BIC}(\mathbf{w}(\lambda))$  is monotonically increasing in  $\lambda \in (\lambda_{k+1}, \lambda_k]$  since  $\text{card}(\mathcal{S}(\mathbf{w}(\lambda)))$  remains constant within this interval. Therefore, the solution of the bilevel problem (15) with the BIC-based merit function occurs at one of the discontinuity points  $\lambda_k$ . Consequently, problem (15) reduces to the discrete model selection problem:

$$\min_{0 \leq k \leq r} F_{BIC}(\mathbf{w}(\lambda_k)). \quad (18)$$

### 3.2.3 Post-LASSO for model bias reduction

An important characteristic of LASSO solutions is that they are biased toward zero whenever  $\lambda > 0$ . As a result, the mean squared error of  $\mathbf{w}(\lambda)$  may not accurately reflect the true likelihood of the model, particularly in the context of the BIC-based formulation. To address this, we use a post-penalized estimator, namely an ordinary least squares regression restricted to the set of predictors selected by LASSO (Belloni and Chernozhukov, 2011). This approach is known as the Post-LASSO problem. As a direct consequence of the predictor set remaining fixed within each line segment of the LASSO regularization path, the Post-LASSO problems can be formulated for each  $k = 0, 1, \dots, r$  as:

$$270 \quad \min_{\mathbf{w} \in \mathbb{R}^{m_t}} \frac{1}{NM} \|A\mathbf{w} - Y\|_2^2$$

$$\text{s.t. } \mathbf{w}_j = 0, \quad \text{for } j \notin \mathcal{S}(\mathbf{w}(\lambda_k)).$$

Let  $\{\mathbf{w}_{LS}^k\}_{k=0}^r$  denote the set of Post-LASSO solutions corresponding to the sequence  $\{\lambda_k\}_{k=0}^r$ . Instead of using solutions from the LASSO regularization path in the  $\ell_0$ -regression (17) or BIC-based (18) formulations, we employ the Post-LASSO solutions  $\mathbf{w}_{LS}^k$ , which provides an alternative model selector with reduced bias:

$$\min_{0 \leq k \leq r} F(\mathbf{w}_{LS}^k). \quad (19)$$

275 Finally, this formulation is used to select the weights for the optical depth parametrization for each gas and pressure level, using either the bilevel  $\ell_0$ +LASSO regression or the bilevel BIC+LASSO regression formulations.

## 4 Numerical Results

This section evaluates the performance of the proposed parametrization compared to the standard RTTOV v13. Specifically, it studies the level of sparsity achieved and its impact on accuracy relative to RTTOV v13 and Line-by-Line calculations using LBLRTM. Performance is measured via the root mean square error (RMSE) of the transmittances compared to Line-by-Line transmittances, and by assessing the brightness temperature (BT) approximation error from the Fast-RT models against Line-by-Line results. Additionally, the BT error is compared to the Noise-Equivalent Delta Temperature (NEdT) of the M-band VIIRS instrument to assess the proportion of profiles for which the model error remains below the instrument noise.

### 4.1 Experiment settings

For training the RTTOV parametrizations and the proposed sparse variants, six variable gases are considered: H<sub>2</sub>O, O<sub>3</sub>, CO<sub>2</sub>, N<sub>2</sub>O, CO, and CH<sub>4</sub>. The Fast-RT model can additionally consider SO<sub>2</sub> as a variable gas, but here it will be treated as a fixed gas among the total of 22 fixed gases considered. No distinction is made between water vapor absorption lines and continuum absorption. For the viewing angle, we consider 6 path secant angles from 1 to 2.25 with step 0.25 (from 0° to 63.61°).

#### 4.1.1 Spectral Response Functions of VIIRS M-bands:

The VIIRS is an instrument on NOAA’s Suomi NPP and NOAA-20 satellites, part of the Joint Polar Satellite System (JPSS). It features 16 moderate resolution bands (M-bands) that cover visible and infrared spectra. This study focuses on spectral response functions for bands M7 to M16, which cover the near (NIR), medium (MIR), and long (LIR) infrared ranges. In this study, we use the VIIRS SRF J2, which can be downloaded from the following link: <https://ncc.nesdis.noaa.gov/NOAA-21/index.php>. Details on the centers and spectral ranges of these bands can be found in Tables 1 and 2 in Cao et al. (2017).

For each channel, the wavenumber  $\nu$  and the corresponding Spectral Response Function (SRF) values are tabulated. The wavenumber tabulation typically covers a broader spectral range, denoted as  $[\nu_a, \nu_b]$ , with noisy SRF values at the extremes of this interval. Therefore, the SRF must be truncated to a smaller interval that retains most of the relevant SRF information. Instead of using Tables 1 and 2 from Cao et al. (2017) for our calculations, we utilize channels with a spectral range broader than those. These channels are defined as  $[\nu^* - \nu_l, \nu^* + \nu_u]$ , where  $\nu^*$  is the centroid of SRF in  $[\nu_a, \nu_b]$ ,  $\nu_l$  and  $\nu_u$  are the tabulated wavenumber values closest to  $\nu^*$  below and above, respectively, such that the relative truncation error does not exceed  $\epsilon = 9 \times 10^{-4}$ . Specifically:

$$(1 - \epsilon) \int_{\nu_a}^{\nu_b} \phi(\nu^*, \nu) d\nu \leq \int_{\nu^* - \nu_l}^{\nu^* + \nu_u} \phi(\nu^*, \nu) d\nu.$$

The integrals are calculated using the composite trapezoidal rule. The SRF data are then truncated and normalized within this new interval, and the centroid  $\nu^*$  is recalculated. The updated channels and centroids are presented in Table 1.

| Band       | Centroid ( $cm^{-1}$ ) | Spectral Range ( $cm^{-1}$ ) | IR  |
|------------|------------------------|------------------------------|-----|
| <b>M7</b>  | 11525.42               | 11070.96 – 12048.02          | NIR |
| <b>M8</b>  | 8056.98                | 7924.69 – 8170.62            | NIR |
| <b>M9</b>  | 7235.57                | 7134.59 – 7373.52            | NIR |
| <b>M10</b> | 6199.43                | 5853.32 – 6522.30            | NIR |
| <b>M11</b> | 4442.00                | 4342.01 – 4549.99            | NIR |
| <b>M12</b> | 2711.61                | 2545.18 – 2867.98            | MIR |
| <b>M13</b> | 2489.30                | 2354.64 – 2607.44            | MIR |
| <b>M14</b> | 1166.76                | 1111.73 – 1235.32            | LIR |
| <b>M15</b> | 939.82                 | 875.89 – 1008.36             | LIR |
| <b>M16</b> | 839.10                 | 782.35 – 896.29              | LIR |

**Table 1.** VIIRS IR M-bands (wavenumber)

305 By truncating the noisy tails of the SRF in this way, the resulting NSRF for each channel is interpolated using natural cubic splines to be used for calculating polychromatic transmittances with a much finer spectral resolution than the tabulated NSRF data. It can be shown that the error made by approximating the polychromatic transmittance with the truncated NSRF does not exceed  $\epsilon$ .

#### 4.1.2 Vertical profile database ECMWF83:

310 For training the optical depth parametrization, we use the ECMWF83 database, which includes 83 vertical profiles with temperature and gas concentrations for H<sub>2</sub>O, O<sub>3</sub>, CO<sub>2</sub>, N<sub>2</sub>O, CO and CH<sub>4</sub>, across 101 pressure levels, originally created to train RTTOV (Matricardi, 2008). A separate database with 22 vertical profiles covers fixed gases. These datasets are available from NWP SAF of EUMETSAT and can be downloaded at <https://nwp-saf.eumetsat.int/site/>.

#### 4.1.3 Line-by-Line Transmittances with LBLRTM:

315 In this study, LBLRTM v12.15.1 (February 2023) will be employed for Line-by-Line calculations. The software uses AER Continuum MT CKD v4.1.1. for continuum models of water vapor and other gases and the AER Line Parameter Database v3.8.1. for line parameters, which consolidates various line spectral databases, primarily HITRAN 2016 (Gordon et al., 2017).

The principal parameter in the LBLRTM calculation, to generate the optical depths for training and top-of-atmosphere radiances, are the following:

- 320
- The continuum absorption is not activated for isolated gases and fixed gases, nor when all gases are included.
  - The Voigt profile is chosen for the shape of spectral lines,

- The spectral resolution is set to  $d\nu = \bar{\alpha}_\nu/1.5$  where  $\bar{\alpha}_\nu$  is the average value of the Voigt halfwidth for the layer. Consequently, the spectral resolution is not homogeneous across channels, achieving an average spectral resolution from  $7.1 \times 10^{-3}$  for M7 to  $4.1 \times 10^{-4}$  for M16.
- The calculation of optical depths with the software is performed only for the observation point at nadir. For other angles, variations are made directly in the calculation of polychromatic transmittances.

#### 4.1.4 RTTOV v13 and Proposals Settings:

For short, we will abbreviate Fast-RT models as follows: RTTOV13 for the standard RTTOV v13; SI for RTTOV13 with statistical threshold and ordinary least squares for parameterization; BIC+L1 for RTTOV13 with statistical threshold and BIC+LASSO regression for parameterization; and L0+L1 for RTTOV13 with statistical threshold and  $\ell_0$ +LASSO regression for parameterization

We implemented the transmittance parametrization of RTTOV v13 as described in (Saunders et al., 2020), using the same predictors, except for the method of selecting gases per channel, which is detailed below.

In RTTOV v13 in the standard form, regression parameters are obtained by including only the gases that exhibit absorption lines in each channel, as shown in Table 2. In the proposed RTTOV variants, using statistical inference and LASSO regression, all gases are included in the training.

Additionally, there are other criteria for selecting predictors in the correction term and training data by level, which are listed below:

- **Threshold for gases correction term:** Predictors for fixed gases are always included in the correction term. For other gases, predictors for a specific gas in a layer are included only if any of the corresponding optical depths in the training profile for that layer exceed a threshold 0.01 for CH<sub>4</sub> and 0.005 for the other gases. As a result, for all the VIIRS channels studied, only predictors for fixed gases and water vapor are included in the correction term.
- **Threshold for Optical Depth Data Training:** Optical depth data in a layer for a gas is omitted if the corresponding transmittance from the layer to the surface is less than  $3 \times 10^{-6}$ . As a result, only channel M10 is affected by this selection criterion.

| Channels | Gases                                                                                      |
|----------|--------------------------------------------------------------------------------------------|
| M7       | H <sub>2</sub> O, CO <sub>2</sub> , CH <sub>4</sub>                                        |
| M8       | H <sub>2</sub> O, CO <sub>2</sub> , CO, CH <sub>4</sub>                                    |
| M9       | H <sub>2</sub> O, CO <sub>2</sub> , NO <sub>2</sub> , CH <sub>4</sub>                      |
| M10-M11  | H <sub>2</sub> O, O <sub>3</sub> , CO <sub>2</sub> , N <sub>2</sub> O, CO, CH <sub>4</sub> |
| M12-M16  | H <sub>2</sub> O, O <sub>3</sub> , CO <sub>2</sub> , N <sub>2</sub> O, CH <sub>4</sub>     |

**Table 2.** Gases considered in RTTOV v13 for VIIRS M-bands.

The performance of the three proposed models, SI, BIC+L1, and L0+LASSO, is evaluated using different statistical threshold parameters  $\epsilon_1 \in \{10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}\}$ . Since the L0+LASSO bilevel model is based on a validation data criterion for the

upper-level merit function, we split the  $NM$  data randomly in half, using one half for training the LASSO problems and the other half as validation data for evaluating model quality using the  $\ell_0$  regression.

## 350 4.2 Sparsity Pattern in the parametrization of optical depths

Table 3 summarizes the percentage of non-zero parameters (%NZ) out of a total of 11,000 parameters (worst-case scenario) for each type of optical depth model: RTTOV13, SI, BIC+L1, and L0+L1. Figures 1 and 2 show the percentage of parameter usage and computation time relative to RTTOV13. Tables 4, 5, 6, and 7 provide details on the number of non-zero parameters (NNZ) for each gas type and correction factor, for  $\epsilon_1 = 10^{-6}$ .

355 In Table 3, the increase in sparsity for the proposed parametrizations compared to the general RTTOV v13 scheme is evident. RTTOV v13 induces sparsity by manually selecting gases and applying optical depth thresholds to include predictors in the correction factor. Using  $\epsilon_1 = 10^{-6}$  as a reference, in the best-case scenario with channel M7, where greater sparsity is achieved with RTTOV13, the sparsity level of RTTOV13 (53.64%) increases to 93.66% for SI, 94.39% for BIC+L1, and 96.22% for L0+L1. Conversely, in the worst-case scenario with channels M10 and M11, where RTTOV13 achieves lower sparsity (20%),  
360 the levels increase to 76.49% and 77.13% for SI, 80.12% and 81.30% for BIC+L1, and 90.50% and 89.89% for L0+L1. As the statistical threshold tolerances decrease, sparsity levels also decrease; however, they remain higher than those of RTTOV13, suggesting that the computational cost benefits are preserved while achieving better sparsity results with the proposed L0+L1 model.

| Fast-RT |           | M7    | M8    | M9    | M10   | M11   | M12   | M13   | M14   | M15   | M16   |
|---------|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| RTTOV13 |           | 46.36 | 46.47 | 59.77 | 80.00 | 80.00 | 69.95 | 68.18 | 70.00 | 69.41 | 69.77 |
| SI      | $10^{-6}$ | 6.34  | 9.17  | 24.49 | 23.51 | 22.87 | 34.20 | 30.55 | 37.95 | 34.43 | 27.83 |
|         | $10^{-7}$ | 13.05 | 20.23 | 27.85 | 28.80 | 28.69 | 53.49 | 46.34 | 50.64 | 44.38 | 39.35 |
|         | $10^{-8}$ | 24.09 | 32.67 | 35.28 | 36.99 | 41.61 | 64.95 | 57.72 | 56.59 | 50.86 | 46.32 |
|         | $10^{-9}$ | 37.76 | 40.36 | 47.10 | 56.57 | 55.13 | 68.29 | 62.51 | 62.69 | 57.75 | 47.95 |
| BIC+L1  | $10^{-6}$ | 5.61  | 8.31  | 18.37 | 19.88 | 18.70 | 27.94 | 25.21 | 30.36 | 29.76 | 23.15 |
|         | $10^{-7}$ | 10.33 | 17.17 | 20.63 | 23.96 | 22.10 | 43.84 | 37.26 | 38.77 | 37.25 | 32.60 |
|         | $10^{-8}$ | 17.87 | 25.01 | 26.31 | 30.40 | 32.22 | 52.51 | 45.03 | 43.58 | 42.47 | 36.50 |
|         | $10^{-9}$ | 26.39 | 29.46 | 36.24 | 44.93 | 42.48 | 54.45 | 48.39 | 48.79 | 47.48 | 37.67 |
| L0+L1   | $10^{-6}$ | 3.78  | 6.14  | 12.18 | 9.50  | 10.11 | 16.85 | 15.68 | 19.63 | 23.87 | 16.71 |
|         | $10^{-7}$ | 7.20  | 11.39 | 13.79 | 11.99 | 12.99 | 22.07 | 21.39 | 24.55 | 29.61 | 23.50 |
|         | $10^{-8}$ | 11.00 | 15.73 | 15.93 | 15.34 | 16.96 | 24.59 | 23.53 | 27.63 | 33.95 | 27.38 |
|         | $10^{-9}$ | 14.98 | 18.78 | 18.51 | 20.76 | 21.03 | 25.70 | 25.41 | 31.38 | 37.59 | 28.35 |

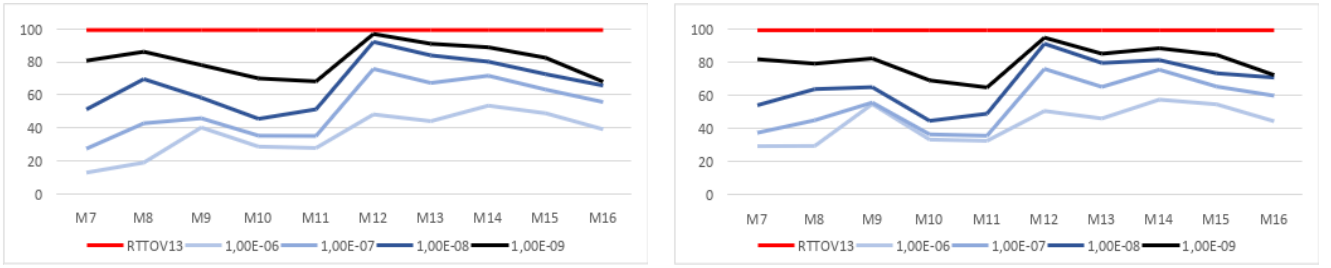
**Table 3.** Percentage of nonzero parameters in RTTOV v13 for each channel, for the standard configuration, SI with OLS regression, BIC+LASSO regression, and  $\ell_0$ +LASSO regression. The second column represents the different statistical thresholds  $\epsilon_1$  used for the proposed RTTOV v13 variants.

In Figures 1 and 2, we present the percentage of parameter usage in the proposed optical depth approximations within  
365 RTTOV, relative to the number of parameters used in the standard RTTOV configuration, and the percentage of runtime re-

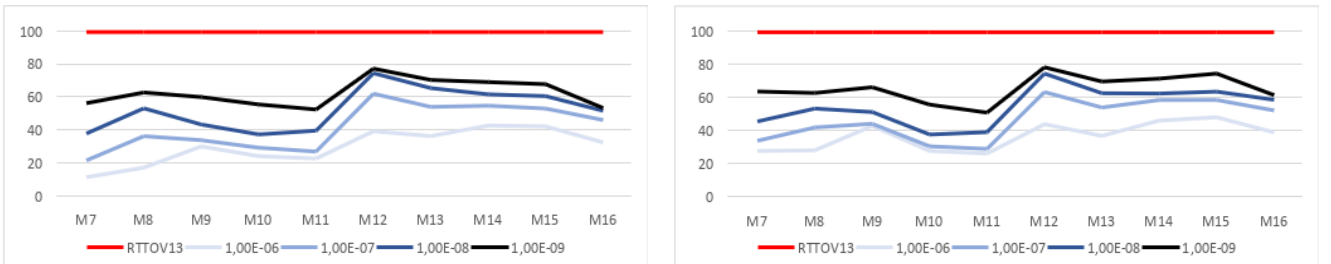
quired by the proposed schemes compared to standard RTTOV. The measured runtime corresponds to the average time of 200 evaluations of the parameterized function used to compute approximate transmittances for the 83 atmospheric profiles with 6 different viewing angles. For the following comparisons, we use  $\epsilon_1 = 10^{-6}$  as a reference. For the SI configuration, parameter usage across all channels ranges from 13.67% to 54.21% relative to standard RTTOV, corresponding to a runtime ranging from 29.99% to 58.32%; for the BIC+L1 configuration, usage ranges from 12.10% to 43.38%, with runtime from 26.75% to 48.64%; and for the L0+L1 configuration, usage ranges from 8.16% to 34.39%, with runtime from 13.77% to 41.39%. These results suggest that the computational cost of evaluating parameterized transmittances is significantly and proportionally reduced with the proposed parametrizations.

Although the absolute runtime difference is small for this limited number of profiles, in practical scenarios where transmittance functions must be evaluated for hundreds of thousands of atmospheric profiles, as required in satellite data retrieval applications, the reduction in computational time becomes highly significant for the efficiency of the retrieval process.

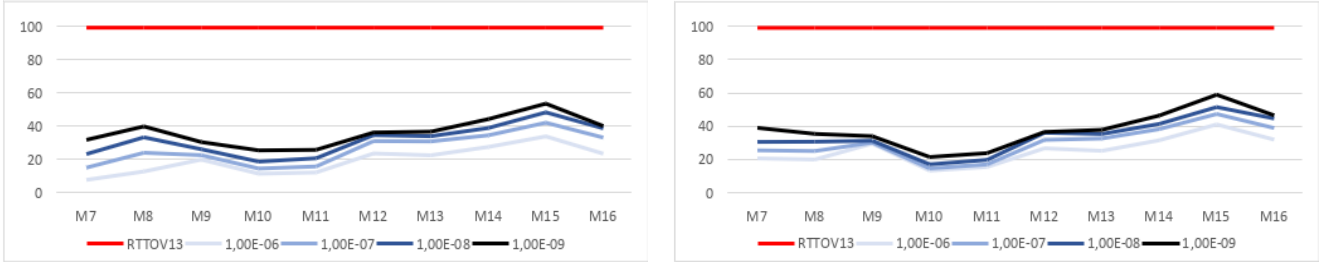
As an illustrative example, from Figure 2, for channel M15, for each 100 time units required to compute transmittances with the RTTOV13 model, the L0+L1 model takes only 41.69 time units with  $\epsilon_1 = 10^{-6}$ , and 59.63 time units with  $\epsilon_1 = 10^{-9}$  (worst-case), representing a significant reduction in runtime.



**Figure 1.** Parameter usage (left) and runtime (right) of the SI method, expressed as percentages relative to those of RTTOV v13 (fixed at 100%) for different values of  $\epsilon_1$ .



**Figure 2.** Parameter usage (left) and runtime (right) of the BIC+LASSO method, expressed as percentages relative to those of RTTOV v13 (fixed at 100%) for different values of  $\epsilon_1$ .



**Figure 3.** Parameter usage (left) and runtime (right) of the L0+LASSO method, expressed as percentages relative to those of RTTOV v13 (fixed at 100%) for different values of  $\epsilon_1$ .

380 In Tables 5, 6, and 7, the effectiveness of introducing statistical thresholds to discard irrelevant gases by channel is evident compared to Table 4. A number of non-zero parameters below 100 for a specific gas corresponds to *Case II* of the statistical threshold parameterization, suggesting that the corresponding gas can be included with the fixed gases.

| Gas | M7   | M8   | M9   | M10  | M11  | M12  | M13  | M14  | M15  | M16  |
|-----|------|------|------|------|------|------|------|------|------|------|
| FIX | 900  | 900  | 900  | 900  | 900  | 900  | 900  | 900  | 900  | 900  |
| H2O | 1400 | 1400 | 1400 | 1400 | 1400 | 1400 | 1400 | 1400 | 1400 | 1400 |
| O3  | 0    | 0    | 0    | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 |
| CO2 | 1300 | 0    | 1300 | 1300 | 1300 | 1300 | 1300 | 1300 | 1300 | 1300 |
| N2O | 0    | 12   | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 |
| CO  | 0    | 1300 | 0    | 1300 | 1300 | 0    | 0    | 0    | 0    | 0    |
| CH4 | 1100 | 1100 | 1100 | 1100 | 1100 | 1100 | 1100 | 1100 | 1100 | 1100 |
| COR | 400  | 400  | 675  | 400  | 400  | 595  | 400  | 600  | 535  | 575  |

**Table 4.** Number of nonzero parameters by gas type and channel in RTTOV13.

| Gas | M7  | M8  | M9   | M10  | M11 | M12 | M13  | M14  | M15  | M16 |
|-----|-----|-----|------|------|-----|-----|------|------|------|-----|
| FIX | 18  | 247 | 0    | 0    | 15  | 145 | 52   | 134  | 478  | 599 |
| H2O | 619 | 618 | 1374 | 618  | 604 | 775 | 576  | 802  | 687  | 716 |
| O3  | 0   | 0   | 0    | 0    | 0   | 644 | 19   | 1120 | 1096 | 466 |
| CO2 | 0   | 56  | 0    | 1142 | 0   | 212 | 724  | 0    | 1213 | 911 |
| N2O | 0   | 0   | 0    | 0    | 897 | 640 | 1024 | 1024 | 33   | 0   |
| CO  | 0   | 0   | 0    | 0    | 0   | 0   | 0    | 0    | 0    | 0   |
| CH4 | 47  | 0   | 778  | 768  | 995 | 964 | 819  | 678  | 0    | 0   |
| COR | 13  | 88  | 542  | 58   | 5   | 382 | 147  | 416  | 280  | 369 |

**Table 5.** Number of nonzero parameters by gas type and channel in SI for  $\epsilon_1 = 10^{-6}$ .

| Gas | M7  | M8  | M9   | M10  | M11 | M12 | M13 | M14 | M15  | M16 |
|-----|-----|-----|------|------|-----|-----|-----|-----|------|-----|
| FIX | 18  | 244 | 0    | 0    | 15  | 125 | 52  | 134 | 463  | 534 |
| H2O | 542 | 545 | 1102 | 535  | 531 | 704 | 475 | 732 | 653  | 670 |
| O3  | 0   | 0   | 0    | 0    | 0   | 531 | 19  | 969 | 904  | 382 |
| CO2 | 0   | 56  | 0    | 1024 | 0   | 196 | 648 | 0   | 1114 | 850 |
| N2O | 0   | 0   | 0    | 0    | 799 | 624 | 901 | 913 | 33   | 0   |
| CO  | 0   | 0   | 0    | 0    | 0   | 0   | 0   | 0   | 0    | 0   |
| CH4 | 47  | 0   | 567  | 576  | 707 | 759 | 559 | 402 | 0    | 0   |
| COR | 10  | 69  | 352  | 52   | 5   | 134 | 119 | 190 | 107  | 111 |

**Table 6.** Number of nonzero parameters by gas type and channel in BIC+L1 for  $\epsilon_1 = 10^{-6}$ .

| Gas | M7  | M8  | M9  | M10 | M11 | M12 | M13 | M14 | M15 | M16 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| FIX | 18  | 187 | 0   | 0   | 15  | 54  | 52  | 92  | 116 | 154 |
| H2O | 342 | 351 | 763 | 239 | 385 | 556 | 281 | 663 | 506 | 590 |
| O3  | 0   | 0   | 0   | 0   | 0   | 238 | 19  | 451 | 826 | 268 |
| CO2 | 0   | 56  | 0   | 394 | 0   | 96  | 497 | 0   | 977 | 661 |
| N2O | 0   | 0   | 0   | 0   | 244 | 268 | 411 | 348 | 33  | 0   |
| CO  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| CH4 | 47  | 0   | 203 | 355 | 463 | 443 | 317 | 353 | 0   | 0   |
| COR | 9   | 81  | 374 | 57  | 5   | 199 | 148 | 252 | 168 | 165 |

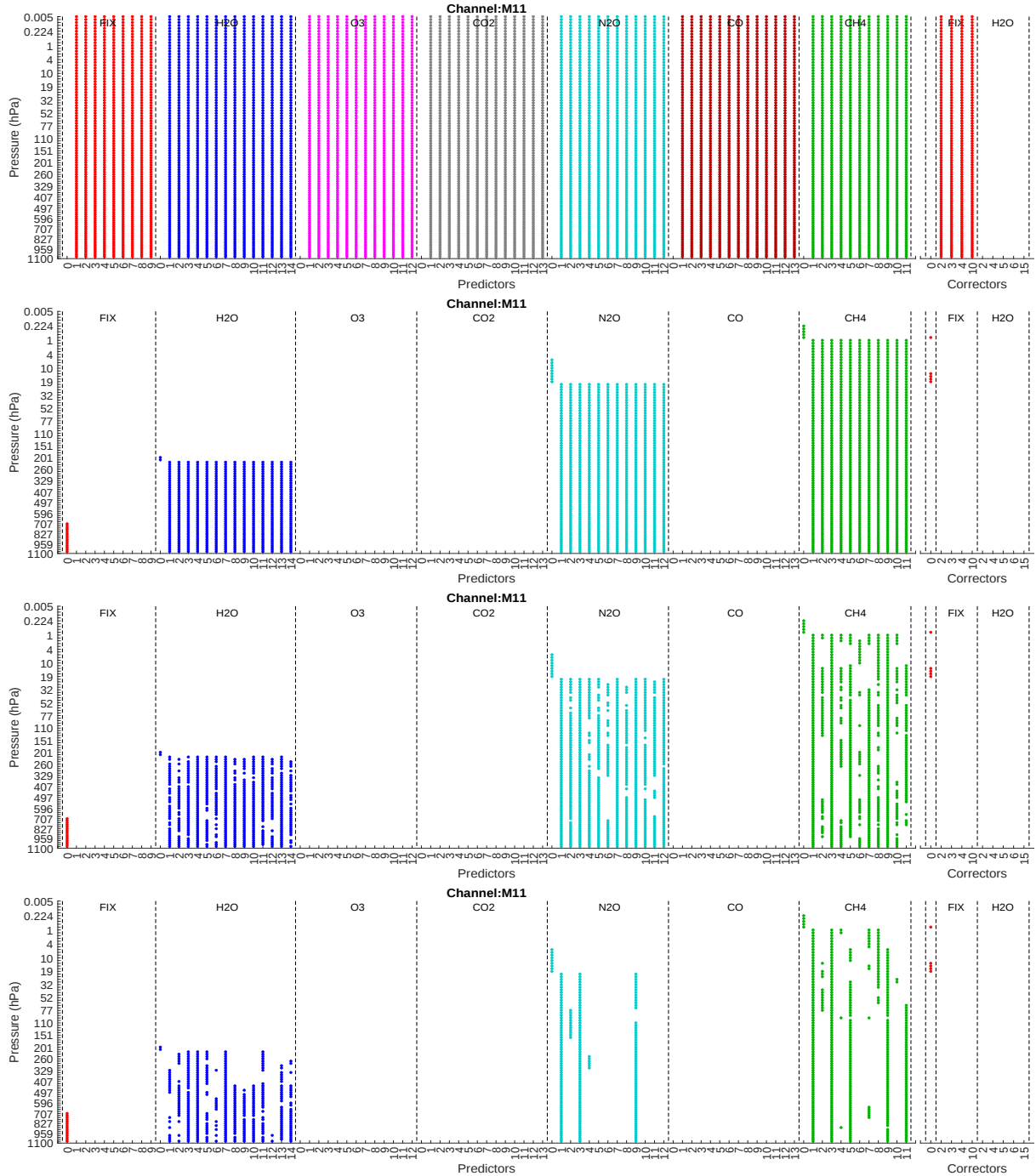
**Table 7.** Number of nonzero parameters by gas type and channel in L0+L1 for  $\epsilon_1 = 10^{-6}$ .

To illustrate in more detail, we reference channels M11 and M12 and compare the sparsity patterns in Figs. 4 and 5 among the four parameterizations using  $\epsilon_1 = 10^{-6}$  as a reference. For the L0+L1 model and the remaining channels, see Appendix A, Figs. A1 and A2. The numbering of predictors and correctors follows RTTOV v13 (Saunders et al., 2020), see Appendix B, except for predictor 0, which corresponds to the predictor in *Case II* of the statistical inference proposal. Each column represents the parameters of a predictor for each pressure level, and each point in a column represents a non-zero parameter associated with that predictor at the corresponding pressure level.

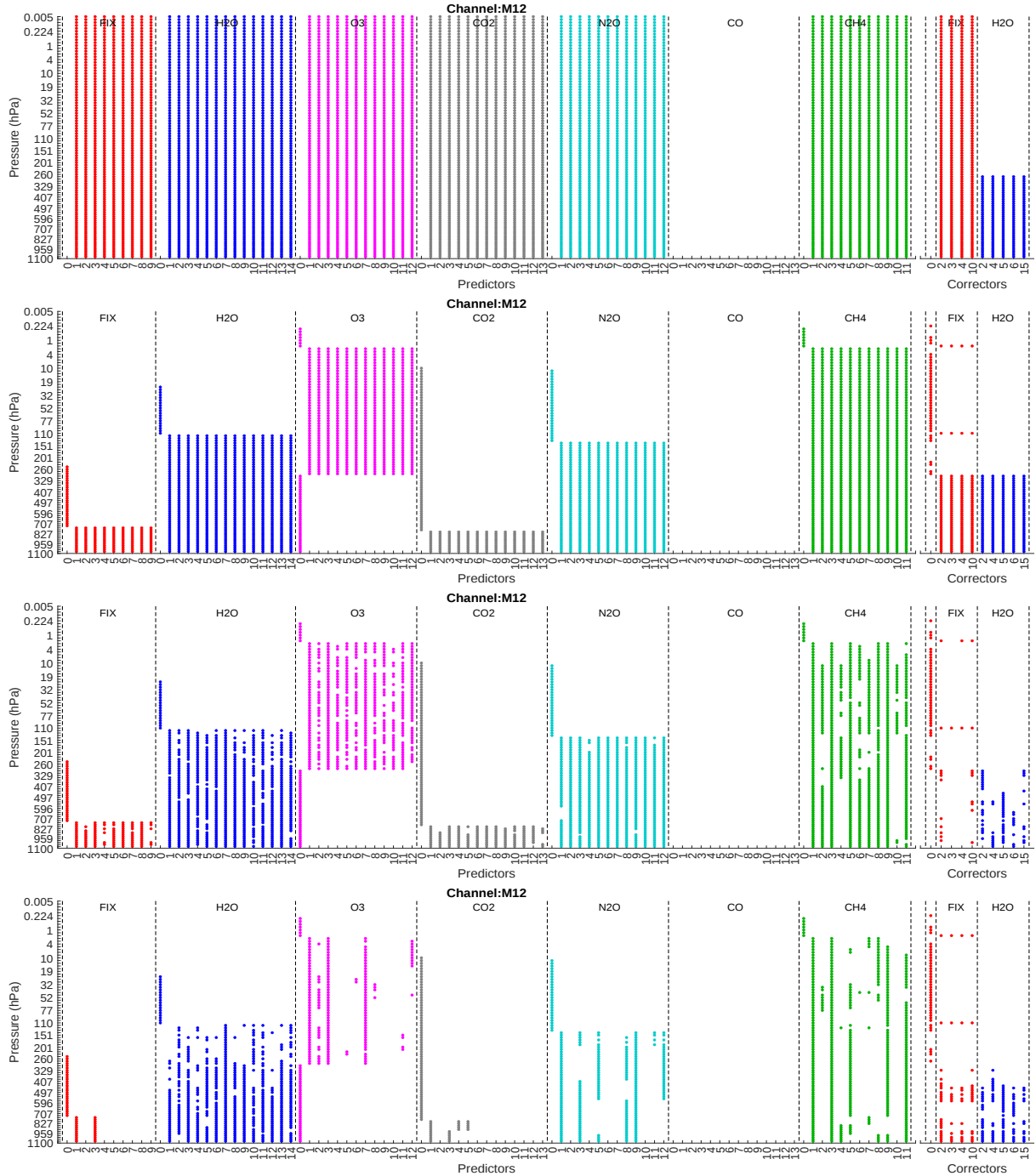
For channel M11 with SI model (upper-middle Fig. 4), gases O3, CO2, and CO are automatically discarded, and fixed gases only need one predictor. Meanwhile, gases H2O, N2O, and CH4 exhibit block-like sparsity patterns from surface pressure approximately to 200 hPa, 19 hPa, and 0.8 hPa, respectively, where concentrations of these gases are important and cause significant radiance absorption. For these gases with block-like sparsity patterns, replacing classical linear regression with L0+LASSO regression (bottom figure) clearly discards some predictors across all levels or shows them as less relevant, as seen in the sparsity patterns for CH4 and N2O. However, H2O still shows sparsity, but it is difficult for this channel to determine if any predictor can be discarded at all levels due to the importance of this gas and the strong non-linear relationship among the secant angle, temperature, and gas concentration in the predictors defined for it. Using BIC+LASSO regression (lower-middle figure) highlights less relevant predictors for CH4, but does not entirely discard it or any other gas predictor retained in the SI model. For the proposed models, no correction term is needed at all, showing that a good fit of the total transmittance is obtained by considering only the approximation of the individual gas transmittances.

400 For channel M12 with SI model (upper-middle Fig. 4), only CO is automatically discarded, which is expected since this gas has no absorption lines in this channel. The SI model still clearly reveals the block-like sparsity patterns of predictors and correctors for each gas at the pressure levels where they contribute to absorption (upper-middle figure). From the figure, CO<sub>2</sub> appears to be relevant at high pressures, approximately above 767 hPa, while O<sub>3</sub> seems relevant between about 2 hPa and 260 hPa. Using L0+LASSO regression (bottom figure) for these important pressure levels demonstrates that some predictors can be  
405 entirely discarded or downweighted, as seen for fixed gases, O<sub>3</sub>, N<sub>2</sub>O, and CH<sub>4</sub>. Similarly, the BIC+L1 model (lower-middle figure) highlights less relevant predictors but does not completely discard any predictor retained in the SI model, except in the corrector terms.

A similar analysis can be performed for each channel, as shown in the appendix, where Figs. A1 and A2 display the sparsity patterns for all channels using the L0+L1 model. These figures clearly indicate which gases are relevant in each channel,  
410 the pressure level ranges where they play a significant role, and which predictors are most important for reconstructing the transmittance of each gas.



**Figure 4.** Sparsity pattern for channel M11, comparing RTTOV13 (Top), SI (Upper-middle), BIC+L1 (Lower-middle), L0+L1 (Bottom) for  $\epsilon_1 = 10^{-6}$ .



**Figure 5.** Sparsity pattern for channel M12, comparing RTTOV13 (Top), SI (Upper-middle), BIC+L1 (Lower-middle), L0+L1 (Bottom) for  $\epsilon_1 = 10^{-6}$ .

### 4.3 Validation of transmittances

To validate the proposed RTTOV v13 variants, we calculated the root mean square error (RMSE) of the total transmittance for all atmospheric layers, vertical profiles, and viewing angles, as shown in the following formula:

$$415 \quad \text{RMSE} = \left( \frac{1}{LMN} \sum_{i=1}^L \sum_{j=1}^M \sum_{k=1}^N (\tau_{ijk}^{TOT} - \tilde{\tau}_{ijk}^{TOT})^2 \right)^{\frac{1}{2}},$$

where  $L = 100$ ,  $M = 83$ , and  $N = 6$ . Here,  $\tau_{ijk}^{TOT}$  and  $\tilde{\tau}_{ijk}^{TOT}$  represent the polychromatic transmittances calculated using LBLRTM optical depths and their corresponding approximations obtained from Eq. (10) using the training data. The results are shown in Table 8. The values in the table correspond to  $\text{RMSE} \times 10^4$ .

| Fast-RT |           | M7     | M8     | M9       | M10    | M11    | M12    | M13    | M14    | M15    | M16    |
|---------|-----------|--------|--------|----------|--------|--------|--------|--------|--------|--------|--------|
| RTTOV13 |           | 0.0126 | 0.0224 | 102.4458 | 0.0133 | 0.1341 | 0.7334 | 0.1128 | 0.8514 | 0.3340 | 0.7794 |
| SI      | $10^{-6}$ | 0.0160 | 0.0332 | 102.4442 | 0.0245 | 0.1320 | 0.7298 | 0.1130 | 0.8504 | 0.3325 | 0.7779 |
|         | $10^{-7}$ | 0.0128 | 0.0229 | 102.4455 | 0.0134 | 0.1336 | 0.7334 | 0.1155 | 0.8513 | 0.3339 | 0.7795 |
|         | $10^{-8}$ | 0.0126 | 0.0228 | 102.4457 | 0.0133 | 0.1340 | 0.7334 | 0.1128 | 0.8514 | 0.3340 | 0.7794 |
|         | $10^{-9}$ | 0.0126 | 0.0228 | 102.4458 | 0.0133 | 0.1341 | 0.7334 | 0.1128 | 0.8514 | 0.3340 | 0.7794 |
| BIC+L1  | $10^{-6}$ | 0.0160 | 0.0333 | 95.1889  | 0.0245 | 0.1321 | 0.7329 | 0.1132 | 0.8606 | 0.3345 | 0.7792 |
|         | $10^{-7}$ | 0.0128 | 0.0231 | 95.1962  | 0.0136 | 0.1341 | 0.7372 | 0.1155 | 0.8610 | 0.3355 | 0.7842 |
|         | $10^{-8}$ | 0.0127 | 0.0230 | 95.1988  | 0.0135 | 0.1345 | 0.7372 | 0.1128 | 0.8610 | 0.3356 | 0.7842 |
|         | $10^{-9}$ | 0.0127 | 0.0230 | 95.1977  | 0.0135 | 0.1345 | 0.7372 | 0.1127 | 0.8610 | 0.3356 | 0.7842 |
| L0+L1   | $10^{-6}$ | 0.0173 | 0.0343 | 165.5327 | 0.0262 | 0.1398 | 0.7384 | 0.1149 | 0.8758 | 0.3392 | 0.7829 |
|         | $10^{-7}$ | 0.0143 | 0.0246 | 165.5344 | 0.0175 | 0.1420 | 0.7406 | 0.1183 | 0.8767 | 0.3399 | 0.7854 |
|         | $10^{-8}$ | 0.0142 | 0.0245 | 165.5347 | 0.0174 | 0.1423 | 0.7407 | 0.1155 | 0.8767 | 0.3400 | 0.7853 |
|         | $10^{-9}$ | 0.0142 | 0.0245 | 165.5348 | 0.0174 | 0.1423 | 0.7407 | 0.1156 | 0.8767 | 0.3400 | 0.7854 |

**Table 8.** RMSE of total transmittance for each channel, scaled by  $10^4$ , for the proposed RTTOV v13 variants. The second column indicates the statistical threshold  $\epsilon_1$  used for each variant.

In Table 8, the RMSE for transmittance errors generally ranges between  $O(10^{-6})$  and  $O(10^{-5})$  across all Fast-RT methods and channels, except for channel M9, where errors are larger, in the range  $O(10^{-2})$  to  $O(10^{-3})$ . All three proposed models slightly degrade the precision of RTTOV13, but this degradation diminishes as the statistical threshold decreases. Comparing RTTOV13 with the SI model, the error difference reduces from  $O(10^{-7})$  to  $O(10^{-9})$  on average across channels, again except for M9. With BIC+LASSO, the difference remains around  $O(10^{-7})$ , while for channel M9 it is  $O(10^{-4})$ . Similarly, with L0+L1 the difference is about  $O(10^{-7})$  for most channels, but  $O(10^{-3})$  for M9. Among the three, the L0+L1 model shows the lowest precision, as expected due to its more aggressive sparsity, yet the errors remain comparable in order of magnitude to RTTOV13.

Overall, these results indicate that including statistical thresholds in RTTOV v13 has minimal impact on the transmittance approximation. Values remain very close to the standard RTTOV13 configuration for statistical threshold tolerances below

10<sup>-6</sup> (Table 8). Combining thresholds with LASSO regression in a bilevel framework for parameter selection, using either  
430 BIC-based or  $\ell_0$ -regularization, slightly modifies the approximation, improving or worsening it, but variations remain small.  
The approximated transmittances closely match those from LBLRTM, with the added benefit of a significant runtime reduction.

#### 4.4 Validation of brightness temperatures

To achieve a higher level of validation for the proposed transmittance parametrization, the brightness temperatures of the  
profiles used for training are calculated. The approximated brightness temperatures at the top of the atmosphere were calculated  
435 using polychromatic radiances from Eq. (3), applying the approximate transmittances provided by the RTTOV v13 scheme and  
the proposed variants, separately. To compare these results, brightness temperatures at the top of the atmosphere were calculated  
using the polychromatic radiances with Eq. (2), using the monochromatic radiances calculated with LBLRTM. In all cases, the  
integrals were approximated using composite trapezoidal formulas, with the spacing determined by the pressure levels of the  
data. In each case, the resulting brightness temperatures were averaged over all profiles and viewing angles. The relative errors  
440 in BT obtained with the Fast-RT models and those obtained with LBLRTM were then calculated, which are shown in Table 9  
( $\times 10^4$ ). The maximum relative error for brightness temperature, determined for each profile and viewing angle, is presented in  
Table 10 ( $\times 10^3$ ).

| Fast-RT |                  | M7     | M8     | M9     | M10    | M11    | M12    | M13    | M14    | M15    | M16    |
|---------|------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| RTTOV13 |                  | 3.8967 | 0.3357 | 7.5875 | 2.1803 | 0.5966 | 1.0472 | 0.6429 | 0.5221 | 0.7386 | 0.6715 |
| SI      | 10 <sup>-6</sup> | 4.7551 | 0.5896 | 7.6048 | 2.2527 | 0.5833 | 1.0181 | 0.6672 | 0.5088 | 0.7521 | 0.7512 |
|         | 10 <sup>-7</sup> | 4.7488 | 0.4203 | 7.5838 | 2.1820 | 0.5985 | 1.0419 | 0.6421 | 0.5198 | 0.7414 | 0.6723 |
|         | 10 <sup>-8</sup> | 4.3021 | 0.3429 | 7.5878 | 2.1799 | 0.5970 | 1.0470 | 0.6430 | 0.5218 | 0.7389 | 0.6716 |
|         | 10 <sup>-9</sup> | 3.8726 | 0.3348 | 7.5875 | 2.1803 | 0.5966 | 1.0472 | 0.6429 | 0.5221 | 0.7386 | 0.6715 |
| BIC+L0  | 10 <sup>-6</sup> | 4.7552 | 0.5896 | 7.6185 | 2.2526 | 0.5832 | 1.0220 | 0.6672 | 0.5109 | 0.7509 | 0.7557 |
|         | 10 <sup>-7</sup> | 4.7490 | 0.4203 | 7.5975 | 2.1819 | 0.5980 | 1.0492 | 0.6421 | 0.5229 | 0.7425 | 0.6742 |
|         | 10 <sup>-8</sup> | 4.3022 | 0.3429 | 7.6035 | 2.1797 | 0.5964 | 1.0549 | 0.6426 | 0.5257 | 0.7411 | 0.6730 |
|         | 10 <sup>-9</sup> | 3.8726 | 0.3348 | 7.6027 | 2.1802 | 0.5962 | 1.0550 | 0.6425 | 0.5258 | 0.7407 | 0.6730 |
| L0+L1   | 10 <sup>-6</sup> | 4.7572 | 0.5896 | 7.8953 | 2.2525 | 0.5819 | 1.0270 | 0.6568 | 0.5145 | 0.7575 | 0.7399 |
|         | 10 <sup>-7</sup> | 4.7503 | 0.4202 | 7.8756 | 2.1824 | 0.5974 | 1.0509 | 0.6309 | 0.5190 | 0.7434 | 0.6632 |
|         | 10 <sup>-8</sup> | 4.3030 | 0.3425 | 7.8858 | 2.1799 | 0.5962 | 1.0570 | 0.6324 | 0.5199 | 0.7414 | 0.6623 |
|         | 10 <sup>-9</sup> | 3.8738 | 0.3343 | 7.8860 | 2.1800 | 0.5959 | 1.0572 | 0.6324 | 0.5201 | 0.7406 | 0.6626 |

**Table 9.** Average Relative Errors in Brightness Temperature (K), scaled by 10<sup>4</sup>, between the Fast-RT and LBLRTM models. The second  
column indicates the statistical threshold  $\epsilon_1$  used for each variant.

| Fast-RT |           | M7     | M8     | M9     | M10    | M11    | M12    | M13    | M14    | M15    | M16    |
|---------|-----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| RTTOV13 |           | 2.2108 | 0.3612 | 1.4245 | 0.8689 | 0.4965 | 0.4082 | 0.2817 | 0.2182 | 0.3501 | 0.2602 |
| SI      | $10^{-6}$ | 5.9256 | 2.0416 | 1.4218 | 1.3972 | 0.4918 | 0.4143 | 0.2967 | 0.1852 | 0.3549 | 0.2787 |
|         | $10^{-7}$ | 5.9259 | 1.2130 | 1.4243 | 0.8691 | 0.4946 | 0.4066 | 0.2819 | 0.2171 | 0.3501 | 0.2602 |
|         | $10^{-8}$ | 3.5565 | 0.4896 | 1.4245 | 0.8689 | 0.4967 | 0.4082 | 0.2817 | 0.2182 | 0.3500 | 0.2602 |
|         | $10^{-9}$ | 2.3083 | 0.3655 | 1.4245 | 0.8689 | 0.4965 | 0.4082 | 0.2817 | 0.2182 | 0.3501 | 0.2602 |
| BIC+L0  | $10^{-6}$ | 5.9256 | 2.0416 | 1.3104 | 1.3972 | 0.4918 | 0.4197 | 0.2969 | 0.1802 | 0.3559 | 0.2807 |
|         | $10^{-7}$ | 5.9259 | 1.2130 | 1.3127 | 0.8687 | 0.4943 | 0.4120 | 0.2820 | 0.2115 | 0.3515 | 0.2660 |
|         | $10^{-8}$ | 3.5565 | 0.4896 | 1.3129 | 0.8686 | 0.4962 | 0.4125 | 0.2816 | 0.2131 | 0.3513 | 0.2650 |
|         | $10^{-9}$ | 2.3085 | 0.3654 | 1.3129 | 0.8687 | 0.4960 | 0.4126 | 0.2817 | 0.2131 | 0.3514 | 0.2650 |
| L0+L1   | $10^{-6}$ | 5.9256 | 2.0415 | 1.5482 | 1.3993 | 0.4891 | 0.4348 | 0.3036 | 0.1998 | 0.3629 | 0.2851 |
|         | $10^{-7}$ | 5.9259 | 1.2130 | 1.5535 | 0.8686 | 0.4912 | 0.4255 | 0.2892 | 0.2246 | 0.3579 | 0.2692 |
|         | $10^{-8}$ | 3.5565 | 0.4882 | 1.5532 | 0.8684 | 0.4932 | 0.4270 | 0.2891 | 0.2256 | 0.3575 | 0.2683 |
|         | $10^{-9}$ | 2.3116 | 0.3636 | 1.5531 | 0.8684 | 0.4930 | 0.4270 | 0.2892 | 0.2258 | 0.3573 | 0.2681 |

**Table 10.** Maximum Relative Errors in Brightness Temperature (K), scaled by  $10^3$ , between the Fast-RT and LBLRTM models. The second column indicates the statistical threshold  $\epsilon_1$  used for each variant.

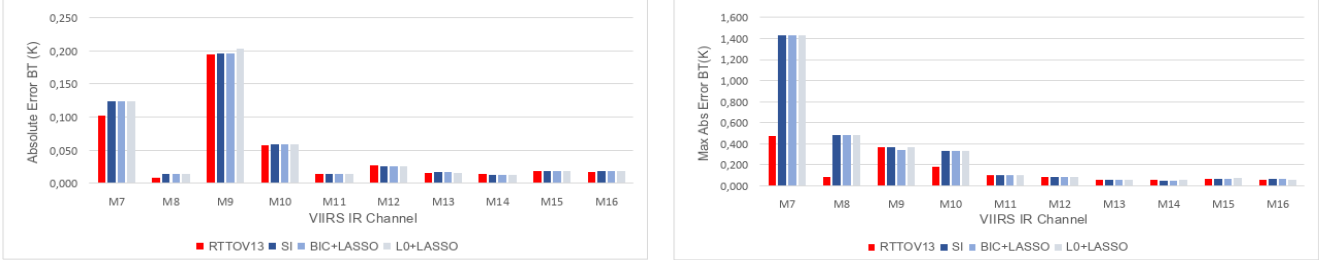
In Table 9, a similar behavior is observed in the errors when approximating transmittances. The average relative error of brightness temperature generally ranges from  $O(10^{-5})$  to  $O(10^{-4})$  across all channels and Fast-RT methods. The order of magnitude of the average relative error remains consistent when comparing the four methods by channel. The differences in average relative BT errors between RTTOV13 and the SI model decrease from  $O(10^{-5})$  to  $O(10^{-7})$  when lowering the statistical threshold tolerance. Similarly, the differences between RTTOV13 and the BIC+L1 model decrease in the same manner. For the L0+L1 model, the differences decrease from  $O(10^{-5})$  to  $O(10^{-6})$ .

Turning to the maximum errors, for all channels the sparse approximations of optical depth for RTTOV13 show minimal deviation from the BT results of standard RTTOV13 when  $\epsilon_1 \leq 10^{-7}$ . Table 10 shows maximum relative BT errors ranging from  $O(10^{-4})$  to  $O(10^{-3})$  across all channels and Fast-RT methods. Comparing the maximum absolute error by channel for the four methods, errors remain of the same order of magnitude for M7, M9, and M11–M16 ( $\epsilon_1 \leq 10^{-6}$ ), M8 ( $\epsilon_1 \leq 10^{-8}$ ), and M10 ( $\epsilon_1 \leq 10^{-7}$ ); in other cases, standard RTTOV13 may yield up to one order of magnitude lower errors.

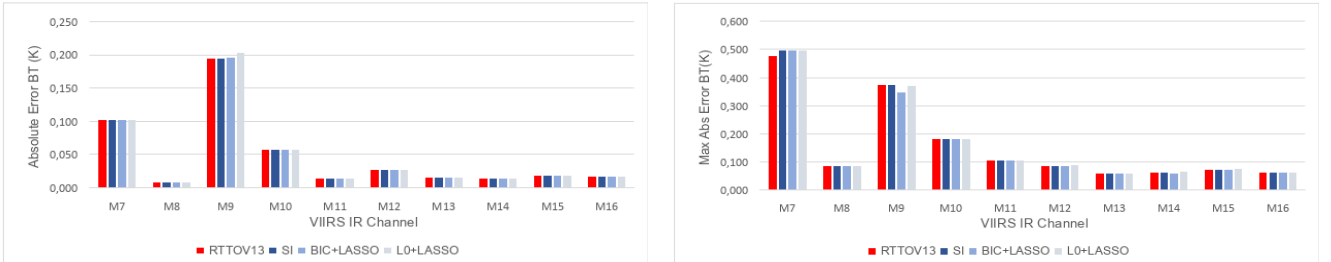
Observe in Table 9 that, for some channels, the errors with the proposed methods are slightly lower than those of RTTOV13. With the L0+L1 model at  $\epsilon_1 = 10^{-9}$  this happens for channels M7, M8, M10, M11, M13, M14, and M16, and with the BIC+L1 model at the same tolerance for channels M7, M8, M10, M11, and M13. Also note that, although the BIC+L1 model gives a better transmittance fit than RTTOV13 for channel M9, its brightness temperature error is not improved. These findings suggest that using merit functions based on radiances or BT, together with model complexity penalization, instead of relying only on optical depth fitting, could improve the results of Fast-RT models within the RTTOV13 framework.

Figure 6 (left) shows the average absolute BT error between the LBLRTM model and the Fast-RT models for  $\epsilon_1 = 10^{-6}$ , while Fig. 6 (right) shows the maximum absolute error across all profiles and viewing angles. The average brightness temperature shows some degradation in the proposed methods compared to RTTOV v13: in the worst case, 0.021 K for M7, 0.008 K

for M8, 0.20 K for M9, while the other channels remain below 0.003 K for all proposals. For the maximum absolute error per profile and viewing angle, the worst cases are 0.961 K for M7, 0.405 K for M8, and 0.375 K for M9, with the other channels below 0.15 K. These variations are not significant in relative terms, as shown in Table 9, and decrease with a lower statistical threshold, illustrated in Fig. 7 for  $\epsilon_1 = 10^{-9}$ . Under this setting, the average BT error worsens by only  $5.7 \times 10^{-4}$  K for M7,  $3.7 \times 10^{-5}$  K for M8, and  $7.5 \times 10^{-3}$  K for M9, while the others remain below  $3.2 \times 10^{-4}$  K. The maximum error increases by  $2.1 \times 10^{-2}$  K for M7,  $1.0 \times 10^{-3}$  K for M8, and  $1.0 \times 10^{-3}$  K for M9, with the other channels remaining below  $1.5 \times 10^{-3}$  K.



**Figure 6.** Average Absolute Errors (left) and Maximum Absolute Errors (right) in Brightness Temperature (K) between the Fast-RT and LBLRTM models for  $\epsilon_1 = 10^{-6}$ .



**Figure 7.** Average Absolute Errors (left) and Maximum Absolute Errors (right) in Brightness Temperature (K) between the Fast-RT and LBLRTM models for  $\epsilon_1 = 10^{-9}$ .

These findings confirm that the proposed methods achieve an accuracy level comparable to RTTOV v13 across most channels, with only minimal degradation observed in a few cases under stringent statistical threshold tolerances.

#### 4.5 Validation of Brightness Temperature Against Instrument Noise Characteristics

To evaluate the accuracy of the Fast RT model, we compare the brightness temperatures it generates with those from high-fidelity simulations using LBLRTM. A standard validation criterion requires that the absolute difference in brightness temperature remains below the instrument's noise level (Garand et al., 2001). Specifically, this involves comparing against the Noise Equivalent Delta Temperature (NEdT) for the thermal emissive bands (M12 to M16), and against the Noise Equivalent Delta

Radiance (NEdR) for the solar reflective bands (M7 to M11). For the VIIRS M-bands, Table 11 presents the NEdT values and the signal-to-noise ratios (SNR) used to compute the corresponding NEdR values, as reported in Table 1 of the manual Cao et al. (2017).

| Channel | SNR | $I_{\text{typ}}$ | Channel | NEdT  | $T_{\text{typ}}$ |
|---------|-----|------------------|---------|-------|------------------|
| M7 L    | 340 | 33.4             | M12     | 0.396 | 270              |
| M7 H    | 215 | 6.4              | M13 L   | 0.423 | 380              |
| M8      | 74  | 5.4              | M13 H   | 0.107 | 300              |
| M9      | 83  | 6.0              | M14     | 0.091 | 270              |
| M10     | 342 | 7.3              | M15     | 0.070 | 300              |
| M11     | 10  | 0.12             | M16     | 0.072 | 300              |

**Table 11.** SNR and NEdT Values for VIIRS IR M-Bands (L: Low Gain Mode, H: High Gain Mode).

For each channel from M7 to M11, the table reports the SNR at the reference radiance  $I_{\text{typ}}$  ( $\text{W}/\text{m}^2 \cdot \text{sr} \cdot \mu\text{m}$ ), and for channels  
480 M12 to M16, it reports the NEdT at the reference temperature  $T_{\text{typ}}$  (K). For a thermal emissive band, the NEdT at temperature  $T$  is defined as

$$\text{NEdT}(T) = \text{NEdT}(T_{\text{typ}}) \cdot \frac{B'(T_{\text{typ}})}{B'(T)}$$

where  $B'$  is the derivative of the Planck function with respect to temperature. For solar reflective bands, the Noise Equivalent Delta Radiance (NEdR) at radiance  $I$  is defined as

485 
$$\text{NEdR}(I) = \frac{I}{\text{SNR}}.$$

Let  $I_{ij}$  and  $\tilde{I}_{ij}$  denote the top of atmosphere polychromatic radiances obtained using LBLRTM and the Fast RT model, respectively, for atmospheric profile  $i$  and observation angle  $\theta_j$ , and let  $T_{ij}$  and  $\tilde{T}_{ij}$  be the corresponding brightness temperatures. For emissive bands, the following condition must be satisfied:

$$|T_{ij} - \tilde{T}_{ij}| \leq \text{NEdT}(T_{ij}),$$

490 and for solar reflective bands, we require:

$$|I_{ij} - \tilde{I}_{ij}| \leq \text{NEdR}(I_{ij}).$$

The percentage of atmospheric profiles for which these conditions are satisfied serves as a practical metric to evaluate the quality of the forward model. A high proportion of cases meeting the criterion indicates that the modeling error is smaller than the instrument noise, ensuring that the simulated radiances are sufficiently accurate for satellite retrievals and potentially  
495 suitable for data assimilation. Table 12 reports the percentage of cases, computed over 83 atmospheric profiles and 6 viewing angles, for which the corresponding noise threshold condition is met.

| Fast-RT |           | M7L  | M7H   | M8    | M9    | M10   | M11   | M12   | M13L  | M13H  | M14   | M15   | M16   |
|---------|-----------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| RTTOV13 |           | 6.63 | 10.84 | 100.0 | 97.99 | 63.45 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| SI      | $10^{-6}$ | 6.83 | 10.24 | 97.19 | 97.99 | 61.24 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 99.60 |
|         | $10^{-7}$ | 6.63 | 10.24 | 98.80 | 97.99 | 63.25 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
|         | $10^{-8}$ | 6.63 | 10.24 | 99.80 | 97.99 | 63.45 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
|         | $10^{-9}$ | 6.83 | 10.84 | 100.0 | 97.99 | 63.45 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| BIC+L1  | $10^{-6}$ | 6.83 | 10.24 | 97.19 | 98.39 | 61.24 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 99.60 |
|         | $10^{-7}$ | 6.63 | 10.24 | 98.80 | 98.39 | 63.25 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
|         | $10^{-8}$ | 6.63 | 10.24 | 99.80 | 98.19 | 63.45 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
|         | $10^{-9}$ | 6.83 | 10.64 | 100.0 | 98.19 | 63.45 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| L0+L1   | $10^{-6}$ | 6.83 | 10.24 | 97.19 | 97.79 | 61.24 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
|         | $10^{-7}$ | 6.63 | 10.24 | 98.80 | 97.59 | 63.25 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
|         | $10^{-8}$ | 6.63 | 10.24 | 99.80 | 97.59 | 63.86 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
|         | $10^{-9}$ | 6.83 | 10.64 | 100.0 | 97.59 | 63.86 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

**Table 12.** Percentage of absolute differences in radiance below the NEdR threshold for channels M7–M11, and percentage of absolute differences in brightness temperature below the NEdT threshold for channels M12–M16. The second column indicates the statistical threshold  $\epsilon_1$  used for each variant.

In the table 12, it can be observed that for channels M11 to M16, all methods fully satisfy the noise condition, and the proposed methods are comparable to standard RTTOV13 for a statistical tolerance threshold of  $\epsilon_1 \leq 10^{-6}$ . For channels M7 to M10, a stricter statistical tolerance threshold is required to achieve percentages comparable to RTTOV13. For channels M7 and M10, the fulfillment of the noise criterion is quite poor; we infer that this is due to the lack of solar radiation inclusion in the various Fast-RT methods. However, the results obtained with the proposed methods are similar to RTTOV13 for small statistical thresholds. For channel M9, the proposed BIC+L1 model slightly improve the percentage of profiles that meet the noise threshold compared to standard RTTOV13. It is clear that the proposed methods reproduce the results of standard RTTOV13 for large statistical thresholds in the emissive bands and for smaller statistical thresholds in the solar reflective bands, while offering the advantage of greater computational efficiency due to the induced sparsity.

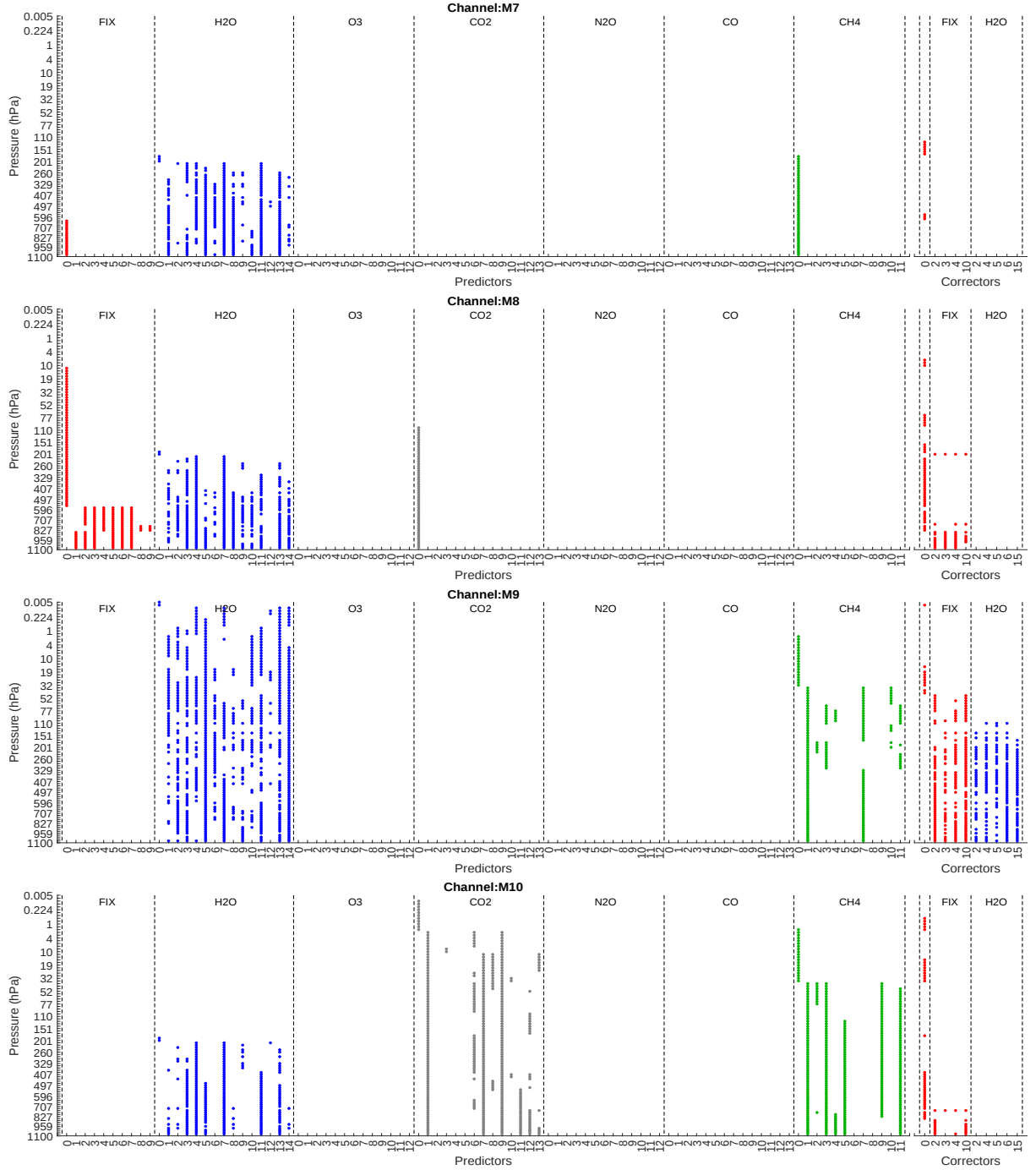
## 5 Conclusions

This study presents an automatic and sparse optical depth parametrization method for the RTTOV v13 model, aimed at optimizing parameter adjustment. The method applies statistical thresholding across different pressure levels, followed by LASSO regression, instead of the traditional least squares approach in the RTTOV v13 framework. A bilevel optimization approach is used to select the optimal regularization parameter, employing different model validation criteria: one based on  $\ell_0$  regression and another on the Bayesian Information Criterion (BIC). These alternatives enforce significant sparsity across all optical depth regression parameters, substantially reducing the computational cost of the Fast-RT model without compromising accuracy, demonstrating strong potential for satellite data assimilation applications.

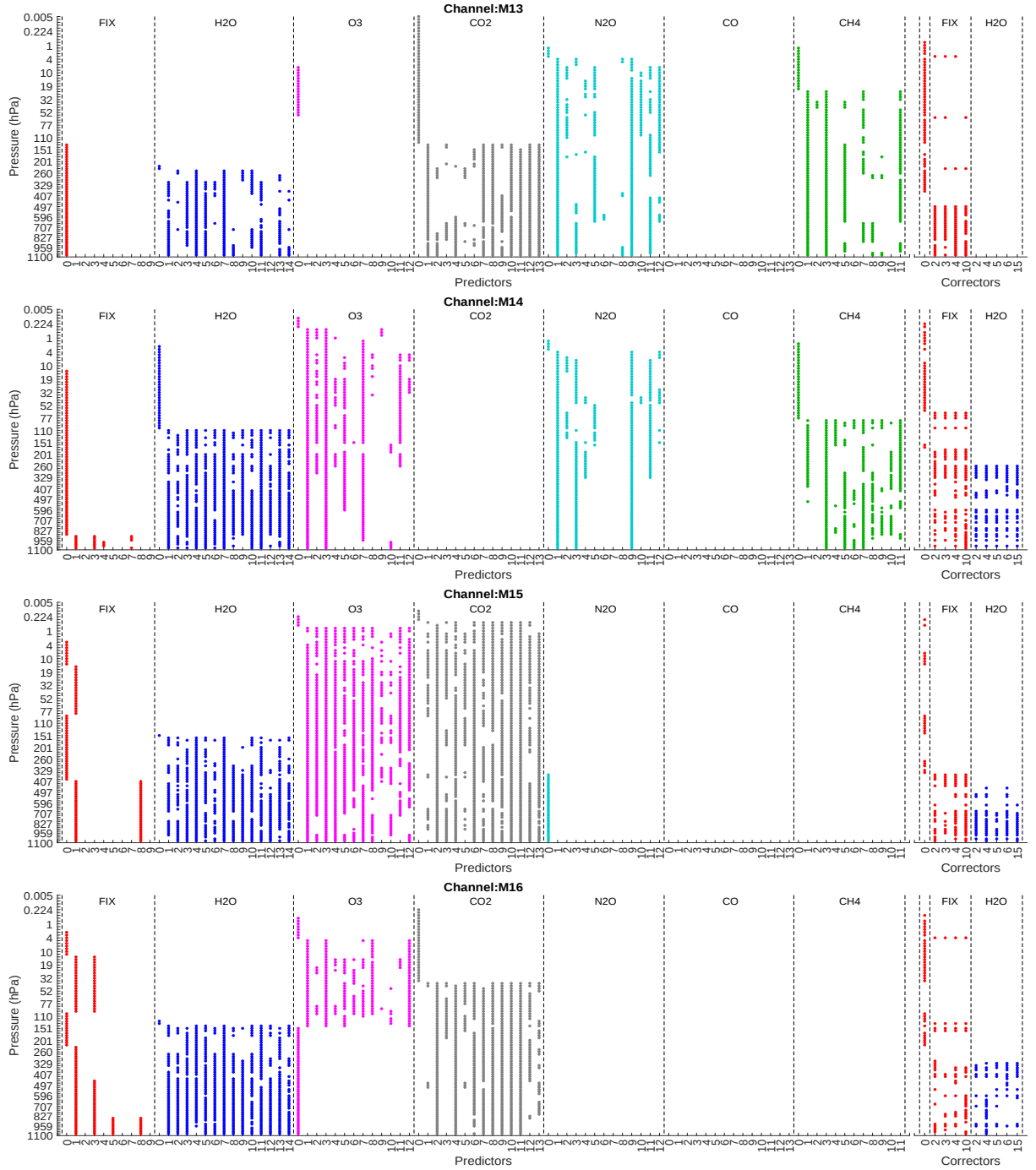
Validation experiments were conducted on the infrared channels of the M-bands for the VIIRS instrument. Different validation criteria were considered, including transmittance fitting against LBLRTM transmittance, brightness temperature fitting against LBLRTM transmittance, and the difference between brightness temperature and the instrument's Noise Equivalent Delta Temperature. The results show consistency with RTTOV v13, while providing improved runtime performance in the evaluation of parameterized transmittances.

The induced sparsity automatically excludes gases with negligible absorptivity in a channel, identifies pressure levels where gases significantly absorb radiance, highlights the most relevant predictors for each gas type, and classifies gases as either fixed or variable. This technique is particularly advantageous for multispectral instruments where multiple gases exhibit strong correlations in radiance absorption, especially in large-scale variable retrievals for inverse problems. The proposed method may be extended to other Fast-RT models, such as CRTM, and to other satellite instruments, such as the Advanced Technology Microwave Sounder (ATMS) and the Cross-track Infrared Sounder (CrIS), to enhance both the computational efficiency of radiative transfer models and the accuracy of retrieved atmospheric profiles.

## Appendix A: Sparsity Pattern for RTTOV13+SI+LASSO



**Figure A1.** Sparsity pattern for channels M7 to M11 in L0-L1 for  $\epsilon_1 = 10^{-6}$ .



**Figure A2.** Sparsity pattern for channels M13 to M16 in L0-L1 for  $\epsilon_1 = 10^{-6}$ .

## Appendix B: RTTOV v3 Predictors

| N <sup>0</sup> | FIX                                  | H <sub>2</sub> O                   | O <sub>3</sub>                                  | CO <sub>2</sub>                 |
|----------------|--------------------------------------|------------------------------------|-------------------------------------------------|---------------------------------|
| 1              | $\sec(\theta)$                       | $(\sec(\theta)W_r)^2$              | $\sec(\theta)O_{3_r}$                           | $\sec(\theta)CO_{2r}$           |
| 2              | $\sec^2(\theta)$                     | $\sec(\theta)W_w$                  | $\sqrt{\sec(\theta)O_{3_r}}$                    | $T_r^2$                         |
| 3              | $\sec(\theta)T_r$                    | $(\sec(\theta)W_w)^2$              | $\sec(\theta)O_{3_r}\delta T$                   | $\sec(\theta)T_r$               |
| 4              | $\sec(\theta)T_r^2$                  | $\sec(\theta)W_r\delta T$          | $\sec(\theta)O_{3_r}/O_w$                       | $\sec(\theta)T_r^2$             |
| 5              | $T_r$                                | $\sqrt{\sec(\theta)W_r}$           | $(\sec(\theta)O_{3_r})^2$                       | $T_r$                           |
| 6              | $T_r^2$                              | $\sqrt[4]{\sec(\theta)W_r}$        | $\sec(\theta)O_{3_r}^2O_w$                      | $\sec(\theta)T_w$               |
| 7              | $\sec(\theta)T_w$                    | $\sec(\theta)W_r$                  | $\sqrt{\sec(\theta)O_{3_r}O_{3_r}/O_w}$         | $(\sec(\theta)CO_{2w})^2$       |
| 8              | $\sec(\theta)T_r^3$                  | $(\sec(\theta)W_w)^{1.5}$          | $\sec(\theta)O_{3_r}O_w$                        | $\sec(\theta)T_w\sqrt{T_r}$     |
| 9              | $\sec(\theta)\sqrt{\sec(\theta)T_r}$ | $(\sec(\theta)W_r)^{1.5}$          | $(\sec(\theta)O_{3_w})^{1.75}$                  | $\sqrt{\sec(\theta)CO_{2r}}$    |
| 10             | 1                                    | $(\sec(\theta)W_r)^{1.5}\delta T$  | $\sec(\theta)O_{3_r}\sqrt{\sec(\theta)O_{3_w}}$ | $T_r^3$                         |
| 11             | -                                    | $\sqrt{\sec(\theta)W_r}\delta T$   | $(\sec(\theta)O_{3_w})^2$                       | $\sec(\theta)T_r^3$             |
| 12             | -                                    | $(\sec(\theta)W_w)^{1.25}$         | $\sqrt{\sec(\theta)O_{3_w}^2}\delta T$          | $\sqrt{\sec(\theta)T_r^2T_w^3}$ |
| 13             | -                                    | $\sec(\theta)W_r^2/W_w$            | $\sec(\theta)O_{3_w}$                           | $T_r^2T_w^2$                    |
| 14             | -                                    | $\sqrt{\sec(\theta)W_rW_r/W_{wt}}$ | -                                               | $\sec(\theta)CO_{2w}$           |
| 15             | -                                    | $\sec(\theta)\sqrt{W_w}$           | -                                               | -                               |
| <b>OD</b>      | 1–9                                  | 1–14                               | 1–12                                            | 1–13                            |
| <b>COR</b>     | 2, 3, 4, 10                          | 2, 4, 5, 6, 15                     | 13                                              | 14, 8, 9                        |

| N <sup>0</sup> | N <sub>2</sub> O                         | CO                                   | CH <sub>4</sub>                                |
|----------------|------------------------------------------|--------------------------------------|------------------------------------------------|
| 1              | $\sec(\theta)N_2O_r$                     | $\sec(\theta)CO_r$                   | $\sec(\theta)CH_{4_r}$                         |
| 2              | $\sqrt{\sec(\theta)N_2O_r}$              | $\sqrt{\sec(\theta)CO_r}$            | $\sqrt{\sec(\theta)CH_{4_r}}$                  |
| 3              | $\sec(\theta)N_2O_r\delta T$             | $\sec(\theta)CO_r\delta T$           | $\sec(\theta)CH_{4_r}\delta T$                 |
| 4              | $(\sec(\theta)N_2O_r)^2$                 | $(\sec(\theta)CO_r)^2$               | $(\sec(\theta)CH_{4_r})^2$                     |
| 5              | $N_2O_r\delta T$                         | $\sqrt{\sec(\theta)CO_r\delta T}$    | $CH_{4_r}\delta T$                             |
| 6              | $\sqrt[4]{\sec(\theta)N_2O_r}$           | $\sqrt[4]{\sec(\theta)CO_r}$         | $\sqrt[4]{\sec(\theta)CH_{4_r}}$               |
| 7              | $\sec(\theta)N_2O_w$                     | $\sec(\theta)CO_r\delta T \delta T $ | $\sec(\theta)CH_{4_{wt}}$                      |
| 8              | $\sec(\theta)N_2O_{wt}$                  | $\sec(\theta)CO_r^2/CO_w$            | $CH_{4_{wt}}$                                  |
| 9              | $\sqrt{\sec(\theta)N_2O_rN_2O_r/N_2O_w}$ | $\sqrt{\sec(\theta)CO_rCO_r/CO_w}$   | $(\sec(\theta)CH_{4_w})^2$                     |
| 10             | $(\sec(\theta)N_2O_{wt})^2$              | $\sec(\theta)CO_r^2/\sqrt{CO_w}$     | $\sec(\theta)CH_{4_w}$                         |
| 11             | $(\sec(\theta)N_2O_{wt})^3$              | $(\sec(\theta)CO_w)^{0.4}$           | $\sqrt{\sec(\theta)CH_{4_r}CH_{4_r}/CH_{4_w}}$ |
| 12             | $\sec^2(\theta)N_2O_{wt}\delta T$        | $\sqrt[4]{\sec(\theta)CO_{wt}}$      | $(\sec(\theta)CH_{4_w})^{1.25}$                |
| 13             | -                                        | $\sec^2(\theta)CO_rCO_w$             | -                                              |
| 14             | -                                        | $\sec(\theta)CO_w$                   | -                                              |
| 15             | -                                        | $\sec(\theta)CO_{wt}$                | -                                              |
| 16             | -                                        | $(\sec(\theta)CO_w)^2$               | -                                              |
| <b>OD</b>      | 1–12                                     | 1–13                                 | 1–11                                           |
| <b>COR</b>     | 7, 8, 10, 11, 12                         | 12, 14, 15, 16                       | 7, 9, 10, 12                                   |

**Table B1.** Predictors for RTTOV v13, (Saunders et al., 2017).

$$p\delta p(l) = p(l+1)(p(l+1) - p(l)), \quad p\delta p(0) = p(1)(p(2) - p(1)),$$

$$T(l) = \frac{1}{2}(T^{\text{prof}}(l) + T^{\text{prof}}(l+1)), \quad T^*(l) = \frac{1}{2}(T^{\text{ref}}(l) + T^{\text{ref}}(l+1)), \quad T_r(l) = \frac{T(l)}{T^*(l)},$$

$$T_w(l) = \frac{\sum_{i=1}^l p\delta p(l-1)T(l)}{\sum_{i=1}^l p\delta p(l-1)T^*(l)}, \quad \delta T(l) = T(l) - T^*(l),$$

$$G(l) = \frac{1}{2}(G^{\text{prof}}(l) + G^{\text{prof}}(l+1)), \quad G^*(l) = \frac{1}{2}(G^{\text{ref}}(l) + G^{\text{ref}}(l+1)), \quad G_r(l) = \frac{G(l)}{G^*(l)},$$

$$G_w(l) = \frac{\sum_{i=1}^l p\delta p(l-1)G(l)}{\sum_{i=1}^l p\delta p(l-1)G^*(l)}, \quad G_{wt}(l) = \frac{\sum_{i=1}^l p\delta p(l-1)T(l)G(l)}{\sum_{i=1}^l p\delta p(l-1)T^*(l)G^*(l)}.$$

Where  $p(l)$  is the pressure (hPa) at level  $l$ ,  $T^{\text{prof}}(l)$  is the temperature (K) at level  $l$  of the input profile,  $T^{\text{ref}}(l)$  is the temperature (K) at level  $l$  of the reference profile which is the mean over the training profile set,  $G \in \{W = H_2O, O_3, CO_2, N_2O, CO, CH_4\}$  represents gas concentration (ppmv over dry air),  $G^{\text{prof}}(l)$  are the gas concentrations at level  $l$  of the input profile and  $G^{\text{ref}}(l)$  are the gas concentrations at level  $l$  of the reference profile which is the mean over the training profile set.

*Author contributions.* JC conceptualized the study, contributed to the development of the methodology, and supervised the research. FV developed the model code, performed the simulations, and designed and carried out the experiments. JC and FV validated the results, ensuring reproducibility. FV wrote the original draft, while JC was responsible for review and editing.

*Competing interests.* The authors declare that they have no conflict of interest.

*Code and data availability.* The code and data used in this study are available at <https://doi.org/10.5281/zenodo.14990818> ((Vargas Jiménez and De los Reyes, 2025)). Access to the code and data is available to anyone with the provided link, and there are no temporal embargoes or restrictions on access. Anyone who views or downloads the code and data from Zenodo does so anonymously. The software is still under development and is not finalized for end-user applications, but it is provided to allow for the reproduction of the results presented in the manuscript. The code and data is licensed under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license.

*Acknowledgements.* This research has been supported by Escuela Politécnica Nacional de Ecuador under award PIGR-22-01: *Asimilación de datos satelitales para el sistema de pronóstico meteorológico METEO: selección óptima de predictores y localización óptima de observaciones*. Franklin Vargas acknowledges partial support from the PhD Program in Applied Mathematics, Escuela Politécnica Nacional de Ecuador. This manuscript benefited from the use of artificial intelligence tools for style correction.

## References

- Belloni, A. and Chernozhukov, V.: L1-penalized quantile regression in high-dimensional sparse models, *The Annals of Statistics*, 39, 82–130, <https://doi.org/10.1214/10-AOS840>, 2011.
- 550 Bertsimas, D., King, A., and Mazumder, R.: Best Subset Selection via a Modern Optimization Lens, *The Annals of Statistics*, 44, 813–852, <https://doi.org/10.1214/15-AOS1388>, 2016.
- Cao, C., Xiong, X. J., Wolfe, R., DeLuccia, F., Liu, Q. M., Blonski, S., Lin, G. G., Nishihama, M., Pogorzala, D., Oudrari, H., and Hillger, D.: Visible Infrared Imaging Radiometer Suite (VIIRS) Sensor Data Record (SDR) User’s Guide (Version 1.3), NOAA technical report nesdis 142, NOAA, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, 2017.
- 555 Cao, D., Ma, Y., Sun, L., and Gao, L.: Fast observation simulation method based on XGBoost for visible bands over the ocean surface under clear-sky conditions, *Remote Sensing Letters*, 12, 674–683, <https://doi.org/10.1080/2150704X.2021.1925371>, 2021.
- Cardall, A. C., Hales, R. C., Tanner, K. B., Williams, G. P., and Markert, K. N.: LASSO (L1) regularization for development of sparse remote-sensing models with applications in optically complex waters using GEE tools, *Remote Sensing*, 15, 1670, <https://doi.org/10.3390/rs15061670>, 2023.
- 560 Chen, Y., Weng, F., Han, Y., and Liu, Q.: Validation of the Community Radiative Transfer Model by using CloudSat data, *Journal of Geophysical Research: Atmospheres*, 113, D00A03, <https://doi.org/10.1029/2007JD009561>, 2008.
- Clough, S. A. and Iacono, M. J.: Line-by-line calculation of atmospheric fluxes and cooling rates: 2. Application to carbon dioxide, ozone, methane, nitrous oxide and the halocarbons, *Journal of Geophysical Research*, 100, 16 519–16 535, <https://doi.org/10.1029/95JD01386>, 1995.
- 565 Clough, S. A., Iacono, M. J., and Moncet, J.-L.: Line-by-line calculations of atmospheric fluxes and cooling rates: Application to water vapor, *Journal of Geophysical Research*, 97, 15 761–15 785, <https://doi.org/10.1029/92JD01419>, 1992.
- Clough, S. A., Shephard, M. W., Mlawer, E. J., Delamere, J. S., Iacono, M. J., Cady-Pereira, K., Boukabara, S., and Brown, P. D.: Atmospheric radiative transfer modeling: a summary of the AER codes, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 91, 233–244, 2005.
- 570 De los Reyes, J. C.: Bilevel imaging learning problems as mathematical programs with complementarity constraints: Reformulation and theory, *SIAM Journal on Imaging Sciences*, 16, 1655–1686, 2023.
- De los Reyes, J. C. and Villacís, D.: Bilevel optimization methods in imaging, in: *Handbook of mathematical models and algorithms in computer vision and imaging: Mathematical imaging and vision*, pp. 1–34, Springer, 2022.
- DeSouza-Machado, S., Strow, L. L., Motteler, H., and Hannon, S.: kCARTA: A fast pseudo line-by-line radiative transfer algorithm with  
575 analytic Jacobians, fluxes, nonlocal thermodynamic equilibrium, and scattering for the infrared, *Atmospheric Measurement Techniques*, 13, 323–339, <https://doi.org/10.5194/amt-13-323-2020>, 2020.
- Edwards, D.: GENLN2: A General Line-by-Line Atmospheric Transmittance and Radiance Model, Version 3.0 Description and Users Guide, Technical note ncar/tn-367-str, NCAR, National Center for Atmospheric Research, Boulder, 1992.
- Fleming, H. E. and McMillin, L. M.: Atmospheric transmittance of an absorbing gas 2: a computationally fast and accurate transmittance  
580 model for slant paths at different zenith angles, *Applied Optics*, 16, 1366–1370, 1977.
- Garand, L., Turner, D. S., Larocque, M., Bates, J., Boukabara, S., Brunel, P., Chevallier, F., Deblonde, G., Engelen, R., Hollingshead, M., Jackson, D., Jedlovec, G., Joiner, J., Kleespies, T., McKague, D. S., McMillin, L., Moncet, J.-L., Pardo, J. R., Rayer, P. J., Salathé, E., Saunders, R., Scott, N. A., Delst, P. V., and Woolf, H.: Radiance and Jacobian intercomparison of radiative transfer models applied to

- HIRS and AMSU channels, *Journal of Geophysical Research: Atmospheres*, 106, 24 017–24 031, <https://doi.org/10.1029/2000JD000184>, 2001.
- 585 Gordon, I. E., Rothman, L. S., Hill, C., Kochanov, R. V., Tan, Y., Bernath, P. F., Birk, M., Boudon, V., Campargue, A., Chance, K. V., Drouin, B. J., Flaud, J.-M., Gamache, R. R., Hodges, J. T., Jacquemart, D., Perevalov, V. I., Perrin, A., Shine, K. P., Smith, M.-A. H., and Zak, E. J.: The HITRAN2016 molecular spectroscopic database, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 203, 3–69, <https://doi.org/10.1016/j.jqsrt.2017.06.038>, 2017.
- 590 Han, Y., van Delst, P., Liu, Q., Weng, F., Yan, B., Treadon, R., and Derber, J.: JCSDA Community Radiative Transfer Model (CRTM)—Version 1, Technical report nesdis 122, NOAA, NOAA, Camp Springs, MD, 2006.
- Heilemann, J., Marx, A., Klassert, C., Boeing, F., Samaniego, L., Klauer, B., Thober, S., and Gawel, E.: Projecting impacts of extreme weather events on crop yields using LASSO regression, *Weather and Climate Extremes*, 46, 100 738, <https://doi.org/10.1016/j.wace.2024.100738>, 2024.
- 595 Hocking, J., Vidot, J., Brunel, P., Roquet, P., Silveira, B., Turner, E., and Lupu, C.: A new gas absorption optical depth parameterisation for RTTOV version 13, *Geoscientific Model Development*, 14, 2899–2915, <https://doi.org/10.5194/gmd-14-2899-2021>, 2021.
- Hong, F. and Kong, Y.: Random Forest Fusion Classification of Remote Sensing PolSAR and Optical Image Based on LASSO and IM Factor, in: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 5048–5051, <https://doi.org/10.1109/IGARSS47720.2021.9553357>, 2021.
- 600 Joshi, D. R., Clay, S. A., Sharma, P., Rekabdarkolae, H. M., Kharel, T. P., Rizzo, D. M., Thapa, R., and Clay, D. E.: Artificial Intelligence and Satellite Based Remote Sensing can be used to Predict Soybean (*Glycine max*) Yield, *Agronomy Journal*, 2023.
- Kleespies, T. J., van Delst, P., McMillin, L. M., and Derber, J.: Atmospheric transmittance of an absorbing gas: OPTRAN status report and introduction to the NESDIS/NCEP community radiative transfer model, *Applied Optics*, 43, 3103–3109, <https://doi.org/10.1364/AO.43.003103>, 2004.
- 605 Krishnan, P., Srinivasa Ramanujam, K., and Balaji, C.: An artificial neural network based fast radiative transfer model for simulating infrared sounder radiances, *Journal of Earth System Science*, 121, 891–901, <https://doi.org/10.1007/s12040-012-0197-3>, 2012.
- Lavrentieva, N., Buldyreva, J., and Starikov, V.: Collisional line broadening and shifting of atmospheric gases: A practical guide for line shape modelling by current semi-classical approaches, Imperial College Press, ISBN 9781848165960, 184816596X, 2011.
- Li, D., Cui, F., Wang, A., Li, Y., Wu, J., and Qiao, Y.: Adaptive Detection Algorithm for Hazardous Clouds Based on Infrared Remote Sensing Spectroscopy and the LASSO Method, *IEEE Transactions on Geoscience and Remote Sensing*, 58, 8649–8664, 2020.
- 610 Liu, X., Smith, W. L., Zhou, D. K., Larar, A. M., Huang, H.-L., Ma, X., and Strow, L. L.: Retrieval of atmospheric profiles and cloud properties from IASI spectra using super-channels, *Atmospheric Chemistry and Physics*, 6, 255–265, <https://doi.org/10.5194/acp-6-255-2006>, 2006.
- Mairal, J. and Yu, B.: Complexity Analysis of the Lasso Regularization Path, in: Proceedings of the 29th International Conference on Machine Learning (ICML), pp. 1835–1842, Omnipress, Edinburgh, Scotland, UK, <https://dl.acm.org/doi/10.5555/3042573.3042807>, 2012.
- 615 Matricardi, M.: An inter-comparison of line-by-line radiative transfer models, Technical Memorandum 525, ECMWF, <https://doi.org/10.21957/b3amji4k>, 2007.
- Matricardi, M.: The generation of RTTOV regression coefficients for IASI and AIRS using a new profile training set and a new line-by-line database, Technical Memorandum 564, ECMWF, 2008.
- Matricardi, M.: A Principal Component Based Version of the RTTOV Fast Radiative Transfer Model, Tech. Rep. 136(652), *Quarterly Journal of the Royal Meteorological Society*, <https://doi.org/10.1002/qj.680>, 2010.
- 620

- Mauceri, S., O'Dell, C. W., McGarragh, G., and Natraj, V.: Radiative Transfer Speed-Up Combining Optimal Spectral Sampling With a Machine Learning Approach, *Frontiers in Remote Sensing*, 3, 932 548, <https://doi.org/10.3389/frsen.2022.932548>, 2022.
- McMillin, L. M. and Fleming, H. E.: Atmospheric transmittance of an absorbing gas: a computationally fast and accurate transmittance model for absorbing gases with constant mixing ratios in inhomogeneous atmospheres, *Applied Optics*, 15, 358–363, 1976.
- 625 McMillin, L. M., Fleming, H. E., and Hill, M. L.: Atmospheric transmittance of an absorbing gas. 3: a computationally fast and accurate transmittance model for absorbing gases with variable mixing ratios, *Applied Optics*, 18, 1600–1606, 1979.
- McMillin, L. M., Crone, L. J., Goldberg, M. D., and Kleespies, T. J.: Atmospheric transmittance of an absorbing gas. 4. OPTRAN: a computationally fast and accurate transmittance model for absorbing gases with fixed and variable mixing ratios at variable viewing angles, *Applied Optics*, 34, 6269–6274, 1995.
- 630 McMillin, L. M., Xiong, X., Han, Y., Kleespies, T. J., and van Delst, P.: Atmospheric transmittance of an absorbing gas. 7. Further improvements to the OPTRAN 6 approach, *Applied Optics*, 45, 2028–2034, <https://doi.org/10.1364/AO.45.002028>, 2006.
- Miller, A.: *Subset Selection in Regression*, Chapman and Hall/CRC, 2nd edn., <https://doi.org/10.1201/9781420035933>, 2002.
- Osborne, M., Presnell, B., and Turlach, B.: A new approach to variable selection in least squares problems, *IMA Journal of Numerical Analysis*, 20, 389–403, <https://doi.org/10.1093/imanum/20.3.389>, 2000.
- 635 Pak, A., Rad, A. K., Nematollahi, M. J., et al.: Application of the Lasso regularisation technique in mitigating overfitting in air quality prediction models, *Scientific Reports*, 15, 547, <https://doi.org/10.1038/s41598-024-84342-y>, 2025.
- Saunders, R., Hocking, J., Rundle, D., Rayer, P., Havemann, S., Matricardi, M., Geer, A., Lupu, C., Brunel, P., and Vidot, J.: RTTOV v12 science and validation report, Tech. rep., NWP SAF, EUMETSAT, [https://www.nwpsaf.eu/site/download/documentation/rtm/docs\\_rttov12/rttov12\\_svr.pdf](https://www.nwpsaf.eu/site/download/documentation/rtm/docs_rttov12/rttov12_svr.pdf), 78 pp., 2017.
- 640 Saunders, R., Hocking, J., Turner, E., Rayer, P., Rundle, D., Brunel, P., Vidot, J., Roquet, P., Matricardi, M., Geer, A., Bormann, N., and Lupu, C.: An update on the RTTOV fast radiative transfer model (currently at version 12), *Geoscientific Model Development*, 11, 2717–2737, <https://doi.org/10.5194/gmd-11-2717-2018>, 2018.
- Saunders, R., Hocking, J., Turner, E., Havemann, S., Geer, A., Lupu, C., Vidot, J., Chambon, P., Köpken-Watts, C., Scheck, L., Stiller, O., Stumpf, C., and Borbas, E.: RTTOV v13 Science and Validation Report, Tech. rep., NWP SAF, [https://nwp-saf.eumetsat.int/site/download/documentation/rtm/docs\\_rttov13/rttov13\\_svr.pdf](https://nwp-saf.eumetsat.int/site/download/documentation/rtm/docs_rttov13/rttov13_svr.pdf), 2020.
- 645 Schwarz, G.: Estimating the Dimension of a Model, *The Annals of Statistics*, 6, 461–464, <https://doi.org/10.1214/aos/1176344136>, 1978.
- Smith, H. D., Dubeux, J. C. B., Zare, A., and Wilson, C. H.: Assessing transferability of remote sensing pasture estimates using multiple machine learning algorithms and evaluation structures, *Remote Sensing*, 15, 2940, <https://doi.org/10.3390/rs15112940>, 2023.
- Stegmann, P. G., Johnson, B., Moradi, I., Karpowicz, B. M., and McCarty, W.: A Deep Learning Approach to Fast Radiative Transfer, *Journal of Quantitative Spectroscopy and Radiative Transfer*, <https://doi.org/10.1016/j.jqsrt.2022.108088>, 2022.
- 650 Su, M. Y., Liu, C., Di, D., Le, T., Sun, Y., Li, J., Lu, F., Zhang, P., and Sohn, B. J.: A Multi-Domain Compression Radiative Transfer Model for the Fengyun-4 Geosynchronous Interferometric Infrared Sounder (GIIRS), *Advances in Atmospheric Sciences*, 40, 1844–1858, <https://doi.org/10.1007/s00376-023-2293-5>, 2023.
- Turner, E., Rayer, P., and Saunders, R.: AMSUTRAN: A microwave transmittance code for satellite remote sensing, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 227, 117–129, <https://doi.org/10.1016/j.jqsrt.2019.02.013>, 2019.
- 655 Vargas Jiménez, F. and De los Reyes, J. C.: Automatic Optical Depth Parametrization in Radiative Transfer Models via LASSO-Induced Sparsity for Satellite Data Assimilation: Code, <https://doi.org/10.5281/zenodo.14990818>, zenodo, 2025.

- Wang, P., Tan, S., Zhang, G., Wang, S., and Wu, X.: Remote Sensing Estimation of Forest Aboveground Biomass Based on Lasso-SVR, *Forests*, 13, 1597, <https://doi.org/10.3390/f13101597>, 2022a.
- 660 Wang, S., Wu, Y., Li, R., and Wang, X.: Remote sensing-based retrieval of soil moisture content using stacking ensemble learning models, *Land Degradation & Development*, 34, 911–925, 2022b.
- Weinreb, M., Fleming, H., McMillin, L., and Neuendorffer, A.: NOAA Technical Report NESS 85: Vertical Sounder, Technical report, NOAA, Washington, D.C., 1981.
- Xiong, X. and McMillin, L. M.: Alternative to the effective transmittance approach for the calculation of polychromatic transmittances in  
 665 rapid transmittance models, *Applied Optics*, 44, 67–76, 2005.