A data-driven approach for fast atmospheric radiative transfer inversion

OPEN ACCESS
IOP Publishing
Inverse Problems
Inverse Problems **41** (2025) 085006 (28pp)
https://doi.org/10.1088/1361-6420/adf3dd

A data-driven approach for fast atmospheric radiative transfer inversion

Cristina Sgattoni[1,4,*], **Luca Sgheri**[2] **and Matthias Chung**[3]

[1] Istituto per la BioEconomia (IBE), Consiglio Nazionale delle Ricerche (CNR), Via Madonna del Piano, 10, Sesto Fiorentino, 50019 Firenze, Italy
[2] Istituto per le Applicazioni del Calcolo (IAC), Consiglio Nazionale delle Ricerche (CNR), Via Madonna del Piano, 10, Sesto Fiorentino, 50019 Firenze, Italy
[3] Department of Mathematics, Emory University, 400 Dowman Drive, Atlanta, GA, United States of America
[4] INdAM Research Group GNCS, P.le Aldo Moro, 5, 00185 Roma, Italy

E-mail: cristina.sgattoni@cnr.it, luca.sgheri@cnr.it and matthias.chung@emory.edu

Received 30 October 2024; revised 13 June 2025
Accepted for publication 24 July 2025
Published 4 August 2025

Abstract

Far-infrared Outgoing Radiation Understanding and Monitoring (FORUM) was selected in 2019 as the ninth Earth Explorer mission by the European Space Agency. Its primary objective is to collect interferometric measurements in the far-infrared (FIR) spectral range, which accounts for 50% of Earth's outgoing longwave radiation emitted into space, and will be observed from space for the first time. Accurate measurements of the FIR at the top of the atmosphere are crucial for improving climate models. Current instruments are insufficient, necessitating the development of advanced computational techniques. FORUM will provide unprecedented insights into key atmospheric parameters, such as surface emissivity, water vapor, and ice cloud properties, through the use of a Fourier transform spectrometer. To ensure the quality of the mission's data, an end-to-end simulator was developed to simulate the measurement process and evaluate the effects of instrument characteristics and environmental factors. The core challenge of the mission is solving the retrieval problem, which involves estimating atmospheric properties from the radiance spectra observed by the satellite. This problem is ill-posed and regularization

* Author to whom any correspondence should be addressed.

techniques are necessary to stabilize the solution. In this work, we present a data-driven approach to approximate the inverse mapping in the retrieval problem, aiming to achieve a solution that is both computationally efficient and accurate. In the first phase, we generate an initial approximation of the inverse mapping using only simulated FORUM data. In the second phase, we improve this approximation by introducing climatological data as *a priori* information and using a neural network to estimate the optimal regularization parameters during the retrieval process. While our approach does not match the precision of full-physics retrieval methods, its key advantage is the ability to deliver results almost instantaneously, making it highly suitable for real-time applications. Furthermore, the proposed method can provide more accurate *a priori* estimates for full-physics methods, thereby improving the overall accuracy of the retrieved atmospheric profiles.

Keywords: radiative transfer inversion, data-driven inversion, regularization, parameters estimation, physics-guided machine learning

1. Introduction

The composition of Earth's atmosphere significantly influences how heat is transferred, making it crucial for accurate climate and weather predictions [10, 31, 59]. The *radiative transfer equation* models the intensity of electromagnetic radiation as it passes through a medium. For homogeneous media, solutions of this equation can be found analytically by combining Lambert–Beer and Planck's law. However, for heterogeneous media, like Earth's atmosphere, solutions require numerical integration techniques [11].

Interferometers, measuring radiation intensity as a function of wavelength (the spectrum), are operationally located on satellites. Along the instrument's line of sight, a high-resolution spectrum as a result of the radiative transfer is recorded. The objective is to use these spectra to reconstruct the composition of atmospheric gases and, subsequently, infer Earth's heat balance.

In September 2019, the European Space Agency announced the Far-infrared Outgoing Radiation Understanding and Monitoring (FORUM) mission as its ninth Earth Explorer mission. Scheduled for launch in 2027, FORUM will carry a Fourier transform spectrometer interferometer, which will point towards the nadir, i.e. vertically downward from the satellite to the Earth's surface, to collect spectral data for analyzing Earth's heat signature, see left panel of figure 1. Operating over the spectral range from 100 to 1600 cm^{-1}, the FORUM instrument will provide, for the first time, spectrally resolved radiances in the far-infrared band from space. While the Earth Explorer missions are relatively small projects, the FORUM spectrometer will generate more than 10 000 spectra per day, producing a substantial amount of data requiring processing and analysis.

For comparison, the state-of-the-art Infrared Atmospheric Sounding Interferometer-New Generation (IASI-NG) spectrometer [22], expected to be launched in August 2025, outputs 16 spectra every 800 ms, with short calibration pauses, amounting to approximately 1382 400 spectra per day. This sensor, which is also employed in weather forecasting, has nevertheless been allocated greater computational resources.

An end-to-end simulator for the FORUM spectrometer was developed during the early stages of mission development. This critical tool provides a proof-of-concept for the instruments and allows for a rigorous evaluation of how instrument characteristics and scene conditions will impact the quality of retrieved measurements, see [45, 57] for details.

The atmospheric spectrum depends on a set of factors that can be collectively referred to as the *atmospheric state* and includes surface temperature, surface emissivity, the vertical profiles of temperature, and concentrations of the atmospheric gases. After discretization, the atmospheric state can be denoted by $\mathbf{x} \in \mathbb{R}^n$, where n is the number of parameters considered. Let $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^q$ represent the combination of the radiative transfer and instrumental effects, $\boldsymbol{\epsilon} \in \mathbb{R}^q$ be some additive noise in the data acquisition process and $\mathbf{y} \in \mathbb{R}^q$ be the spectrum, with q number of spectral points, then the relationship between atmospheric state, radiative transfer, and observations can be expressed as

$$\mathbf{F}(\mathbf{x}) + \boldsymbol{\epsilon} = \mathbf{y}.$$

Two equally important tasks are associated with radiative transfer: the direct problem—determining $\mathbf{y}$ from $\mathbf{x}$—and the inverse problem, determining $\mathbf{x}$ from $\mathbf{y}$. Both are of interest for various applications. The direct problem (also referred to as the forward problem) is used both as a sub-problem of the inverse problem and for the assimilation of the radiances in climatological and meteorological models, see for instance [17, 44] and the bibliography therein. The inverse problem is also referred to as the *retrieval problem*, and it is used to monitor the atmosphere. On this subject, the bibliography is extensive. Good introductory monographs on the subject are [36, 60, 64, 66], while the classical book [52] focuses on error analysis and the solution of the inverse problem via regularization techniques. On the regularization, see also [19]. The direct problem, as well as the inverse problem (in which the direct problem is a sub-problem), can be solved using a full-physics method. However, the computational cost of this approach is too high to achieve near real-time data analysis. To solve this problem, different approaches have been proposed to speed up the computation of radiative transfer. The radiative transfer for TOVS (RTTOV) [53] (along with PC-RTTOV [41], the principal component version of RTTOV) and sigma-IASI [3, 37–40] are two candidates for fast radiative transfer codes. Both methods relay on a training set for the determination of preset parameters, which are then used to calculate the radiative transfer without the need of the spectroscopic database or the explicit calculation of the cloud scattering effect, which are the most time-consuming tasks of full-physics methods.

Recent advances in remote sensing have increasingly leveraged artificial intelligence techniques to enhance the efficiency and accuracy of forward models, particularly in solving the radiative transfer for the infrared spectrum at nadir viewing angles. We report here some of the extended bibliography. In 2012 Krishnan [34] developed an artificial neural network-based fast radiative transfer model for simulating IR sounder radiances, achieving high correlation coefficients exceeding 99% with traditional models while substantially reducing computational time. In 2015 Pyatnik [48] developed a fast radiative transfer model for hyperspectral IR satellite sounders. More recently [35] introduced a fully connected deep neural network algorithm integrated with the Community Radiative Transfer Model (CRTM). A recent paper on the CRTM is [32]. Stegmann [61] proposed a deep learning approach to accelerate radiative transfer calculations, demonstrating that neural networks can effectively emulate complex atmospheric transmittance computations. In [62] Ukkonen explored the emulation of a shortwave radiation scheme and its components using feed-forward and recurrent neural networks. Geared towards numerical weather prediction (NWP) is the rapid radiative transfer model for general circulation models—parallel [47]. In [63] a machine-learned gas optics parameterization was introduced in the model.

The inversion of the radiative transfer equation can undoubtedly be obtained by coupling any of the previous forward models with a minimization technique. There are however

some approaches who tackle directly the inverse problem. Zhang [67] introduced a retrieval method that combines an improved one-dimensional variational algorithm with artificial neural networks to retrieve atmospheric temperature profiles from hyperspectral data. In a recent study [8] physics-informed neural networks (PINNs) were used to solve the radiative transfer equation in atmospheric scenarios.

In this work, we develop a generalized data-driven approach to approximate the inverse mapping '$\mathbf{F}^{-1}$' to obtain a computationally fast approximation of the solution to the retrieval problem. We utilize the end-to-end simulated measurements from the FORUM spectrometer to generate training data. The key point of our work is the inclusion of climatological data. Since the inverse problem is ill-posed (see for instance, [52]), any approach that does not include regularization techniques is not able to retrieve accurate profiles. The usual approach to solve the inverse problem is to use climatological or correlative data as *a priori* information. Climatological data, however, depends on the particular case under analysis, so it is not possible to include this information in the training stage of a machine-learning approach. Our solution is to use an additional neural network to find the optimal Tikhonov parameters to be used in the regularization procedure. While this approach does not guarantee a solution of the same accuracy as full-physics retrieval methods, the solution is produced almost instantaneously. If applied to climatology, the provided solution can yield *a priori* estimates that are statistically closer to the true atmospheric state. Using *a priori* values that are closer to the true ones generally provides an advantage in the determination of the solution of the inverse problem [42, 58], because it reduces the bias in the solution introduced by errors in the *a priori*.

This approach can be naturally interpreted as a two-stage retrieval scheme: a fast data-driven inversion using a pseudoinverse operator, followed by a physics-aware regularization informed by climatological information. We refer to our approach as data-driven because the inverse operator is primarily derived from data, rather than being explicitly formulated from physical principles. The inverse operator is computed using the pseudoinverse approach, and it is learned from a training dataset consisting of scenario-spectrum pairs generated with a full-physics model. Once computed, this operator remains a static feature of the retrieval process.

By 'physics-aware', we mean that the entire retrieval framework includes regularization techniques that are informed by physical knowledge. These regularizations, which include *a priori* information such as climatological data and physical constraints, vary depending on the specific scenario, ensuring that the retrieval is consistent with physical reality. Hence, our approach aligns with methods like PINNs, which integrate physical laws, typically formulated as differential equations, into the loss function as a form of regularization [49].

This work is organized as follows. In section 2, we provide an overview of the physics behind the forward problem (section 2.1) and the classical inversion framework (section 2.2). In section 3, we introduce our surrogate inversion scheme, beginning with the purely data-driven phase (section 3.1), followed by the inclusion of *a priori* information (section 3.2), and the estimation of regularization parameters (section 3.3). In section 3.4, we summarize and illustrate the entire solution scheme, encompassing both the offline and online components. In section 4, we describe the structure of the data. In section 5, we present and analyze the results of our numerical experiments, covering the data-driven phase (section 5.1), the impact of incorporating *a priori* information (section 5.2), and a comparison with a full-physics approach (section 5.3). Finally, in section 6, we summarize our contributions, discuss further implications, and suggest directions for future work.

2. Background

2.1. Radiative transfer

In the following, we briefly review the underlying physics of the atmospheric remote sensing problem. For an atmospheric *homogeneous* layer, given a wavenumber ν, the radiative transfer equation can be written as

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}z}(z) = -\alpha_\nu(p,T,c)\,I_\nu(z) + \alpha_\nu(p,T,c)\,B_\nu(T), \qquad \text{with} \quad I_\nu(z_0) = I_{\nu_0}, \tag{1}$$

where z denotes the altitude, $I_\nu(z)$ is the intensity of radiation, B_ν represents the Planck function, and $\alpha_\nu(p,T,c)$ corresponds to the attenuation coefficient, depending on the layer pressure p, the layer temperature T, and the layer concentration of atmospheric gases c. The quantities p, T and c depend on z, but are assumed to be constant inside the layer. The attenuation coefficient can be obtained using a spectroscopic database, such as the high resolution transmission 2020 database, see [25]. Equation (1) is derived by combining both Lambert–Beer's and Planck's laws. The analytic solution of the initial value problem (1) is given by

$$I_\nu(z) = I_{\nu_0}\mathrm{e}^{-\alpha_\nu(p,T,c)(z-z_0)} + B_\nu(T)\left(1-\mathrm{e}^{-\alpha_\nu(p,T,c)(z-z_0)}\right). \tag{2}$$

Since the atmosphere is not a homogeneous medium, it is discretized in layers that are considered homogeneous, see right panel of figure 1. Let $z_0,\ldots,z_N$ the vertical discretization of the atmosphere, with z_0 being Earth level and z_N the altitude of the observer, or the atmosphere limit if the instrument is outside the atmosphere. In each layer $[z_{i-1},z_i]$, T_i is the Curtis–Godson [24] average temperature of the layer, α_i and $\tau_i = \alpha_i(z_i - z_{i-1})$ are the frequency-dependent attenuation coefficient and optical depth, respectively. For clarity, we drop the explicit frequency dependence in the notation from now on.

According to (2), the electromagnetic radiation passing through each layer i is attenuated by a factor $\mathrm{e}^{-\tau_i}$. This energy is absorbed and re-emitted as $(1-\mathrm{e}^{-\tau_i})B(T_i)$ isotropically according to Planck law. The same mechanism is also valid for Earth's surface. The coefficient ϵ that measures the fraction of radiation absorbed is called spectral emissivity and depends on the frequency and the type of soil on the surface. Earth's surface absorbs a fraction ϵ of the radiation I_D reaching it and re-emits it according to Planck law $\epsilon B(T_\mathrm{E})$, where T_E is Earth's surface temperature. There are two main models for the Earth's surface. In the specular model (suitable for water bodies) the remaining fraction $(1-\epsilon)I_D$ is reflected back into space, so that the source upwelling radiation becomes: $I_U = \epsilon B(T_\mathrm{E}) + (1-\epsilon)I_D$. In the Lambertian model (suitable for land surfaces) the surface emission is assumed to be isotropic. While a full Lambertian treatment would require integrating the downwelling radiance over the entire hemisphere, a widely used approximation [13] introduces a fixed *diffusivity angle* $\theta_d \sim 53.13^\circ$ corresponding to $\cos(\theta_d) = \mu_d = 2/3$, which is the average of $\cos(\theta)$ over the hemisphere. This effectively accounts for the angular distribution of radiation without performing the full angular integration. To obtain the reflected intensity at nadir, the downwelling flux is calculated using this effective angle and then divided by π to obtain the intensity, consistent with Lambert's cosine law. The reflected radiation must factor the contributions of the emission of all the layers, each attenuated by passing through the lower layers, resulting in $\frac{1}{c(\mu)}\sum_{i=1}^{N} B(T_i)(1-\mathrm{e}^{-\tau_i/\mu})\mathrm{e}^{-\sum_{j=1}^{i-1}\tau_j/\mu}$, where $\mu = 1$ for specular surfaces (water), and $\mu = 2/3$ for Lambertian surfaces (land). The constant $c(\mu)$ is 1 if $\mu = 1$ and π if $\mu = 2/3$.

Analogously, to calculate the upwelling radiation reaching the instrument we must factor the attenuation of the source radiation I_U by all the atmospheric layers. To this expression, we must add the emission of each layer attenuated by the passage from the upper layers. The final

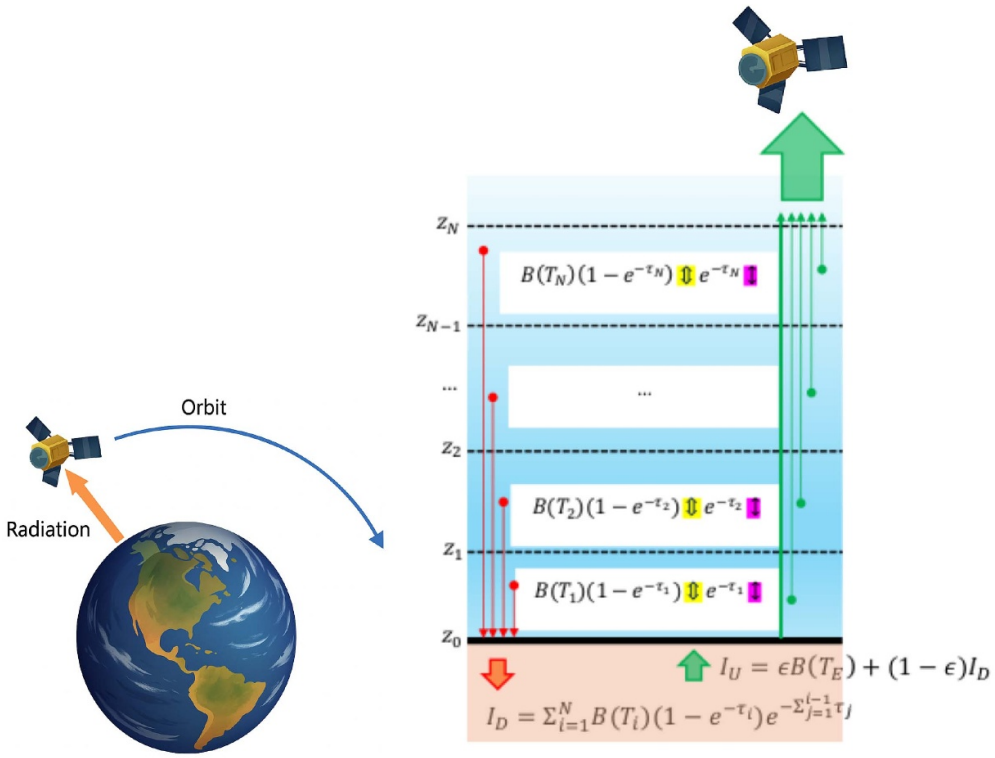

Figure 1. Left panel: FORUM orbit with upwelling radiation seen by the instrument. Right panel: radiative transfer and the discretization of the atmosphere. The downwelling radiation is indicated in red and the upwelling radiation is in green. In each atmospheric layer, the attenuation factor is highlighted with the purple arrow, while the thermal emission is highlighted with the yellow arrow. The specular model is presented. Earth image credit: © Visible Earth and © NASA.

discretized form of the radiative transfer equation is [66]:

$$I(z_N) = \left(\epsilon\, B(T_{\mathrm{E}}) + \frac{1-\epsilon}{c(\mu)} \left(\sum_{i=1}^{N} B(T_i) \left(1 - \mathrm{e}^{-\tau_i/\mu}\right) \mathrm{e}^{-\sum_{j=1}^{i-1}\tau_j/\mu} \right) \right) \mathrm{e}^{-\sum_{i=1}^{N}\tau_i} + \sum_{i=1}^{N} B(T_i) \left(1 - \mathrm{e}^{-\tau_i}\right) \mathrm{e}^{-\sum_{j=i+1}^{N}\tau_j}. \tag{3}$$

In the inverse problem related to (3), which is to determine T_{E}, T_i and τ_i for $i = 1, \ldots, N$, from the knowledge of I, the atmospheric unknowns appear in the exponential attenuation coefficients making the retrieval of atmospheric gases a challenging, ill-conditioned problem.

2.2. Inversion problem

In this remote sensing application, the goal is to retrieve the atmospheric state vector $\mathbf{x}$ from a set of observed radiances $\mathbf{y}$. The state vector $\mathbf{x}$ includes 1 scalar variable (Earth surface temperature), 3 vertical profiles (air temperature, water vapor, ozone), each defined over the $N = 41$

pressure levels, and surface spectral emissivity, defined on 301 spectral channels, resulting in a total of $n = 425$ parameters to be retrieved for each case. The observation vector $\mathbf{y}$ consists of radiance values measured at $q = 4049$ selected spectral channels. The forward operator $\mathbf{F}$ represents the radiative transfer model discussed in section 2.1, discretized over a set of vertical pressure levels ($N = 41$) and evaluated at different spectral channels ($q = 4049$). It maps the atmospheric state vector $\mathbf{x}$ into top-of-atmosphere radiances $\mathbf{y}$.

There are three main categories of techniques for solving this inverse problem. The first includes methods such as principal component analysis (PCA) (e.g. [2, 6, 26, 29, 46] and the book [33]), a compression method that reduces dimensionality by transforming the original variables into linear combinations that capture the signal variability while leaving out the error. A second approach utilized iterative methods and regularization is imposed by early stopping criteria [27]. The third category involves introducing prior information into the merit function, either through a Bayesian framework or via Tikhonov regularization, to guide the solution toward physically plausible results, especially when the measurements provide limited sensitivity to certain parameters. For nadir infrared sounders, this issue primarily affects profiles in the upper stratosphere and mesosphere, where the signal is weak due to the low atmospheric density. In this paper, we adopt the latter approach.

The retrieval problem can be stated as a Bayesian inference problem where the aim is to obtain a posterior distribution of the atmospheric parameters $\mathbf{x}$, given observations $\mathbf{y}$ and prior information of the atmospheric parameters $\mathbf{x}_a$. The posterior is given by

$$\pi(\mathbf{x} \mid \mathbf{y}) = \frac{\pi(\mathbf{y} \mid \mathbf{x})\ \pi(\mathbf{x})}{\pi(\mathbf{y})} \propto \pi(\mathbf{y} \mid \mathbf{x})\ \pi(\mathbf{x}).$$

In this remote sensing problem, it is commonly assumed that the measurement noise and the prior follow a Gaussian distribution [52]. More precisely, we assume $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{F}(\mathbf{x}) \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_\mathbf{y})$ and $\mathbf{x} \sim \mathcal{N}(\mathbf{x}_a, \mathbf{S}_a)$ with symmetric positive definite covariance matrices $\mathbf{S}_\mathbf{y}$ and $\mathbf{S}_a$. The matrix $\mathbf{S}_\mathbf{y}$ can be explicitly estimated from the instrumental characteristics, see for instance [16] for the IASI-NG sounder. The mean $\mathbf{x}_a$ can be determined for instance from correlated measurements or climatology. In both cases, it is also possible to obtain an estimate of $\mathbf{S}_a$, that represents either the correlated measurements error or the atmospheric variability. In this remote sensing application, the maximum *a posteriori* estimate is given by $\mathbf{x}_{\text{map}} \in \arg\max_\mathbf{x} \pi(\mathbf{x} \mid \mathbf{y})$ and can be computed by solving the optimization problem

$$\mathbf{x}_{\text{map}} = \arg\min_\mathbf{x} \chi^2(\mathbf{x}), \qquad \text{where} \tag{4}$$

$$\chi^2(\mathbf{x}) = \tfrac{1}{2}(\mathbf{y} - \mathbf{F}(\mathbf{x}))^\top \mathbf{S}_y^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{x})) + \tfrac{1}{2}(\mathbf{x}_a - \mathbf{x})^\top \mathbf{S}_a^{-1}(\mathbf{x}_a - \mathbf{x}). \tag{5}$$

Details on its derivations can be found for instance in [7, 9]. This radiative transfer inversion problem using full-physics methods equation (4) may also be interpreted as a nonlinear Tikhonov regularized variational inverse problem and is referred to as the optimal estimation (OE) method in the field of atmospheric retrieval problems and introduced by Rodgers [20, 52]. Here, other related techniques include methods based on PCA [3, 41]. By defining $\mathbf{L}_\mathbf{y}$ and $\mathbf{L}_a$ as symmetric decompositions of the precision matrices $\mathbf{S}_\mathbf{y}^{-1} = \mathbf{L}_\mathbf{y}^\top \mathbf{L}_y$ and $\mathbf{S}_a^{-1} = \mathbf{L}_a^\top \mathbf{L}_a$, e.g. by Cholesky or eigenvalue factorization, definition (5) can be restated as :

$$\chi^2(\mathbf{x}) = \tfrac{1}{2}\|\mathbf{L}_\mathbf{y}(\mathbf{F}(\mathbf{x}) - \mathbf{y})\|_2^2 + \tfrac{1}{2}\|\mathbf{L}_a(\mathbf{x} - \mathbf{x}_a)\|_2^2. \tag{6}$$

Solving equation (4) using full-physics methods is computationally demanding because the attenuation coefficients α_ν from equation (3) must be evaluated for each gas, each layer, and

each wavelength using the spectroscopic database. As a result, the evaluation of $\mathbf{F}$ is time-consuming [13, 15]. This process must be repeated at each iteration because the α_i depends both on temperature and the concentrations of the atmospheric gases, that are normally targets of the retrieval. To solve equation (4), Gauss–Newton-type methods are often employed because they normally need a limited number of forward model evaluations. However, due to the non-convexity and heterogeneous sensitivities of χ^2 convergence problems may arise. To address this, Levenberg–Marquardt techniques, which have regularizing properties, are frequently applied [18]. Still, full-physics methods are not adequate to cope with the amount of data produced by modern instruments, making the development of alternative fast and accurate inversion techniques essential.

3. Surrogate inversion scheme

Due to the computational challenges and bottlenecks laid out in the previous section, we propose a physics-aware data-driven approach to overcome these challenges in the atmospheric retrieval problem. In the first step, we directly approximate the inverse model '$\mathbf{F}^{-1}$' using a linear surrogate model, while in the second step, we propose to utilize deep learning methods to improve our predictions by incorporating additional physical information. In section 3.1, we introduce a fully data-driven method that is subsequently incorporated with *a priori* information in section 3.2. Following this, in section 3.3 we propose a procedure for estimating the regularization parameters using a neural network. In the final section 3.4 we summarize the entire solution chain.

3.1. Data-driven surrogate model

Reduced order models have been well-established for dynamical systems and there exists a vast literature on approaches to build reliable surrogate forward models. We point the interested reader to the book [4] and references within. Another obvious approach is to use a linear approximation, e.g. using a Taylor expansion $\mathbf{F}(\mathbf{x}) \approx \mathbf{F}(\mathbf{x}_0) + (\mathrm{D}_{\mathbf{x}}\mathbf{F}(\mathbf{x}_0))\mathbf{x}$ for a given $\mathbf{x}_0$ to simplify the model simulation. Here, we follow a different approach. Since we are interested in the inverse problem, we aim *not* to approximate $\mathbf{F}$ but to approximate the action of its inverse mapping $\mathbf{y} \mapsto \mathbf{x}$. We aim to find a linear operator $\mathbf{Z}$ that minimizes the *Bayes risk*, i.e.

$$\min_{\mathbf{Z}} \quad \mathbb{E}_{\mathbf{x},\mathbf{y}} \ \|\mathbf{x} - \mathbf{Z}\mathbf{y}\|^2_{\mathbf{S}_x^{-1}}, \tag{7}$$

where $\mathbb{E}_{\mathbf{x},\mathbf{y}}$ denotes the expectation over $\mathbf{x}$ and $\mathbf{y}$ and for symmetric positive matrices $\mathbf{\Gamma}$ we define $\|\mathbf{x}\|_{\mathbf{\Gamma}} = (\mathbf{x}^\top \mathbf{\Gamma}\mathbf{x})^{1/2}$. Even if the distribution of $\mathbf{x}$ is known, obtaining the distribution $\mathbf{y} = \mathbf{F}(\mathbf{x}) + \boldsymbol{\epsilon}$ is computationally intractable. Assume we are given data pairs $\{\mathbf{x}_j, \mathbf{y}_j\}_{j=1}^m$, with m number of observations. We obtain through a simulated forward propagation $\mathbf{y}_j = \mathbf{F}(\mathbf{x}_j) + \boldsymbol{\epsilon}_j$, then with the notation $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_m]$, $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_m]$, the *empirical Bayes risk* is given by

$$f(\mathbf{Z}) = \tfrac{1}{m} \|\mathbf{L}_{\mathbf{y}}(\mathbf{X} - \mathbf{Z}\mathbf{Y})\|^2_{\mathrm{F}}, \tag{8}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. A minimizer of (7) can found by determining stationary points of f, i.e. $\frac{\partial f(\mathbf{Z})}{\partial \mathbf{Z}} = \mathbf{0}$. Noticing that

$$mf(\mathbf{Z}) = \mathrm{tr}\big(\mathbf{X}^\top \mathbf{L}_{\mathbf{y}}^\top \mathbf{L}_{\mathbf{y}} \mathbf{X}\big) - 2\mathrm{tr}\big(\mathbf{X}^\top \mathbf{L}_{\mathbf{y}}^\top \mathbf{L}_{\mathbf{y}} \mathbf{Z}\mathbf{Y}\big) + \mathrm{tr}\big(\mathbf{Y}^\top \mathbf{Z}^\top \mathbf{L}_{\mathbf{y}}^\top \mathbf{L}_{\mathbf{y}} \mathbf{Z}\mathbf{Y}\big) \tag{9}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix, we may compute its derivative as

$$m\frac{\partial f}{\partial \mathbf{Z}} = -2\mathbf{L}_{\mathbf{y}}^\top \mathbf{L}_{\mathbf{y}} \mathbf{X}\mathbf{Y}^\top + 2\mathbf{L}_{\mathbf{y}}^\top \mathbf{L}_{\mathbf{y}} \mathbf{Z}\mathbf{Y}\mathbf{Y}^\top \tag{10}$$

and a minimizer $\widehat{\mathbf{Z}}$ of f solves $\widehat{\mathbf{Z}}\mathbf{Y}\mathbf{Y}^\top = \mathbf{X}\mathbf{Y}^\top$. The (minimal norm) solution to this equation is given by

$$\widehat{\mathbf{Z}} = \mathbf{X}\mathbf{Y}^\dagger, \tag{11}$$

where † is the Moore–Penrose pseudoinverse. For a full row-rank matrix $\mathbf{Y}$ we obtain: $\mathbf{Y}^\dagger = \mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T)^{-1}$. This classical approach, that amounts to the minimization of the least-squares functional (8), has been applied for instance in [5], which introduces a novel variant of the iteratively regularized Landweber iteration for solving linear and nonlinear ill-posed inverse problems, or in [21] in the contest of microwave imaging. Further discussions on optimal regularized inverse matrices can be found in [12].

Let us provide a couple of remarks. Certainly, the empirical Bayes risk approach can be used to obtain an approximation on the forward operator $\widehat{\mathbf{A}}\mathbf{x} \approx \mathbf{F}(\mathbf{x})$, however, due to the ill-posed nature of the inverse process we observe that using $\widehat{\mathbf{A}}^\dagger$ is inferior to directly approximating the inverse operation $\widehat{\mathbf{Z}}\mathbf{y}$. In large-scale and severely ill-posed settings further rank constraints on $\mathbf{Z}$ may be considered as developed in [12], however, in the remote sensing application at hand $\mathbf{Z}$ is of moderate dimension and rank constraints are unnecessary. Here the fast repeated application of $\widehat{\mathbf{Z}}$ is of more concern.

Assume we have computed $\widehat{\mathbf{Z}}$ given training data $\{\mathbf{x}_j, \mathbf{y}_j\}_{j=1}^m$ in an off-line phase. If new data $\mathbf{y}$ becomes available we just need a fast matrix-vector multiplication to obtain a prediction of the atmospheric state $\widehat{\mathbf{x}}$, i.e.

$$\widehat{\mathbf{x}} = \widehat{\mathbf{Z}}\mathbf{y}. \tag{12}$$

Notice this purely data-driven approach does not include any prior knowledge of the atmospheric state, hence our next step is to improve upon the atmospheric state estimate to include the prior in the form of regularization.

3.2. *Data-driven approach with a priori correction*

Data-driven methods like neural networks have revolutionized science and our daily lives. While powerful for reconstruction and prediction, data-driven methods can also struggle to predict broader system dynamics or generate physically unrealistic parameters.

New approaches integrate underlying physics and prior knowledge into data-driven prediction processes. These approaches have demonstrated improving reliability and predictions [23, 49]. Focusing on the inverse problem at hand, we incorporate the prior information to improve upon the data-driven prediction. We assume that superior reconstruction can be obtained by balancing between the data-driven prediction $\widehat{\mathbf{x}}$ and the prior $\mathbf{x}_a$, i.e.

$$\widehat{\mathbf{x}}_{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{\xi}} \quad \tfrac{1}{2}\left\|\mathbf{L}_{\mathbf{x}}\left(\boldsymbol{\xi} - \widehat{\mathbf{Z}}\mathbf{y}\right)\right\|_2^2 + \tfrac{1}{2}\left\|\boldsymbol{\Lambda}\mathbf{L}_a\left(\boldsymbol{\xi} - \mathbf{x}_a\right)\right\|_2^2, \tag{13}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\lambda})$ is a positive diagonal weighting matrix representing the importance we decide to assign to the *a priori* contribution, with $\boldsymbol{\lambda}$ indicating its diagonal, and $\mathbf{L}_{\mathbf{x}}$ is the symmetric decomposition of the precision matrix inverse $\mathbf{S}_{\mathbf{x}}^{-1} = \mathbf{L}_{\mathbf{x}}^\top \mathbf{L}_{\mathbf{x}}$. Normally, the covariance matrix $\mathbf{S}_{\mathbf{x}}$ of the unknowns is estimated by linear propagation of errors [52]. In our case, we obtain:

$$\mathbf{S}_{\mathbf{x}} = \mathbb{E}\left[\left(\widehat{\mathbf{x}} - \mathbb{E}\left[\widehat{\mathbf{x}}\right]\right)\left(\widehat{\mathbf{x}} - \mathbb{E}\left[\widehat{\mathbf{x}}\right]\right)^\top\right] = \mathbb{E}\left[\left(\widehat{\mathbf{Z}}\mathbf{y} - \mathbb{E}\left[\widehat{\mathbf{Z}}\mathbf{y}\right]\right)\left(\widehat{\mathbf{Z}}\mathbf{y} - \mathbb{E}\left[\widehat{\mathbf{Z}}\mathbf{y}\right]\right)^\top\right] = \tag{14}$$

$$= \widehat{\mathbf{Z}}\,\mathbb{E}\left[\left(\mathbf{y} - \mathbb{E}\left[\mathbf{y}\right]\right)\left(\mathbf{y} - \mathbb{E}\left[\mathbf{y}\right]\right)^\top\right]\widehat{\mathbf{Z}}^\top = \widehat{\mathbf{Z}}\mathbf{S}_{\mathbf{y}}\widehat{\mathbf{Z}}^\top \tag{15}$$

where $\mathbf{S_y}$ is the known covariance matrix of the measurements.

The unique solution of (13) is given by

$$\widehat{\mathbf{x}}_{\boldsymbol{\lambda}} = \left(\mathbf{L_x} + \mathbf{L}_a^\top \boldsymbol{\Lambda}^2 \mathbf{L}_a\right)^{-1} \left(\mathbf{L_x}\widehat{\mathbf{Z}}\mathbf{y} + \mathbf{L}_a^\top \boldsymbol{\Lambda}^2 \mathbf{L}_a \mathbf{x}_a\right). \tag{16}$$

Note, when $\boldsymbol{\lambda} = \mathbf{0}$, the solution in (16) coincides with the purely data-driven solution in (12). Typically, a regularization technique includes a term of the form $\left\|\mathbf{L_y}(\widehat{\mathbf{A}}\mathbf{x} - \mathbf{y})\right\|_2^2$, which is computed in the measurement space and may be affected by the ill-posedness of the inverse problem. In contrast, the minimization of equation (13) is performed in the parameter space, which in this application has a much lower dimensionality. This minimization balances the solution's distance from the data-driven result with its distance from the prior.

Equation (13) follows a Bayesian approach, so in principle there is no need of the additional weight $\boldsymbol{\Lambda}$, since the covariance matrices already normalize the data according to their statistical properties. However, we achieve better results by introducing $\boldsymbol{\lambda}$, which adjusts the weighting between the two OE terms. Here, we select $\boldsymbol{\lambda}$ by using a neural network described in section 3.3 and discussed in section 5. The idea of modifying weights in an OE framework has been previously explored, for instance in [57], where a larger error on the emissivity profile is used to mitigate the anti-correlation between surface emissivity and surface temperature in the retrieval process. However, to our knowledge, the use of a neural network to determine such weights is introduced here for the first time.

3.3. Regularization parameters estimation

There are well-established statistical techniques to determine the regularization vector $\boldsymbol{\lambda}$, such as the *unbiased predictive risk estimator*, the *discrepancy principle*, *generalized cross validation*, see [7]. While these methods perform well in practice, they all share a common drawback: the associated computational cost often exceeds that of solving the inverse problem itself. This issue is exacerbated when multiple regularization parameters are involved, as in our case. To address this, we propose a data-driven approach for estimating regularization parameters. Specifically, following the method developed in [1], we train a deep neural network to predict these parameters using a training set of m_{t_1} cases. An important part of the training data consists of the optimal regularization parameters $\{\boldsymbol{\lambda}_j^{\text{opt}}\}_{j=1}^{m_{t_1}}$ for the m_{t_1} training points. To obtain these, we employ a deterministic approach. In particular, we formulate a bilevel optimization problem. The outer problem is as follows,

$$\boldsymbol{\lambda}_j^{\text{opt}} = \underset{\boldsymbol{\lambda} \geqslant \mathbf{0}}{\arg\min} \left\|(\mathbf{x}_{\boldsymbol{\lambda}})_j - \mathbf{x}_j\right\|_2 \tag{17}$$

where $(\mathbf{x}_{\boldsymbol{\lambda}})_j$ is the solution of the inner problem described in equation (16) for case j in the training set.

The remaining training data consist of the prior estimates $\{(\mathbf{x}_a)_j\}_{j=1}^{m_{t_1}}$ and the data-driven solutions $\{\widehat{\mathbf{x}}_j\}_{j=1}^{m_{t_1}}$. We assume there exists a well-defined mapping from generic vectors $\widehat{\mathbf{x}}$ and $\mathbf{x}_a$ to $\boldsymbol{\lambda}$. Specifically, we define the mapping as $\widetilde{\boldsymbol{\Phi}} : \mathbb{R}^n \to \mathbb{R}^n$, where $\widetilde{\boldsymbol{\Phi}}(\widehat{\mathbf{x}} - \mathbf{x}_a) = \boldsymbol{\lambda}$. To reduce the complexity of the mapping, we condense the first two pieces of information by using the difference between $\widehat{\mathbf{x}}$ and $\mathbf{x}_a$. We then approximate this mapping with a neural network $\boldsymbol{\Phi} : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^r$ parameterized by $\boldsymbol{\theta}$, i.e.

$$\boldsymbol{\Phi}\left((\widehat{\mathbf{x}} - \mathbf{x}_a), \boldsymbol{\theta}\right) = \boldsymbol{\lambda}. \tag{18}$$

Figure 2. Solution scheme where the different steps are represented with arrows of different colors with the related section indication: purple for data-driven step, blue for regularization step and green for regularization parameters computation and estimation.

Given training data $\left\{\left(\widehat{\mathbf{x}}_j-(\mathbf{x}_a)_j\right),\boldsymbol{\lambda}_j^{\text{opt}}\right\}_{j=1}^{m_{t_1}}$, we determine the network parameters $\boldsymbol{\theta}$ by solving

$$\boldsymbol{\theta}=\arg\min_{\boldsymbol{\zeta}}\frac{1}{m_{t_1}}\sum_{j=1}^{m_{t_1}}\left\|\boldsymbol{\Phi}\left(\left(\widehat{\mathbf{x}}_j-(\mathbf{x}_a)_j\right),\boldsymbol{\zeta}\right)-\boldsymbol{\lambda}_j^{\text{opt}}\right\|^2. \tag{19}$$

A key feature of our approach compared to [1] is that we do not learn a mapping from the data $\mathbf{y}_j$ to the regularization parameters $\boldsymbol{\lambda}_j$ but rather from the estimated parameter $\widehat{\mathbf{x}}_j$ to the regularization parameter so that again we operate within the parameters domain. The advantage of this approach is the reduced parameter size of the neural network which is easier to solve and results in shorter computational times, as demonstrated by the tests.

To improve the numerical stability and allow the network to predict a wider range of values more effectively, the neural network output is not the $\boldsymbol{\lambda}$ vector itself, but its logarithm.

3.4. Solution scheme

Here we summarize the complete solution scheme, as described in the previous sections and depicted in figure 2.

The solution process begins with a fully data-driven phase, represented by the purple arrow in the figure, and explained in section 3.1. In this phase, we determine a linear operator that approximates the inverse radiative transfer process using training data. This step forms the first offline phase. Given an input spectrum $\mathbf{y}$, the operator, denoted as $\widehat{\mathbf{Z}}$, produces an estimate of the atmospheric scenario $\widehat{\mathbf{x}}$ using equation (12). This phase is entirely data-driven and provides a preliminary solution to the atmospheric scenarios. The second step, shown by the blue arrow, incorporates *a priori* information $\mathbf{x}_a$ into the data-driven solution $\widehat{\mathbf{x}}$ through a Tikhonov regularization technique, as explained in section 3.2, resulting in $\widehat{\mathbf{x}}_{\boldsymbol{\lambda}}$. The inclusion of *a priori* information improves the initial estimate by applying additional constraints based on prior knowledge. The regularization parameters for this phase are estimated using a neural network, represented by the horizontal green arrow and discussed in section 3.3.

To train the neural network, we generated a training set of pre-computed optimal regularization parameters, as indicated by the vertical green arrow and detailed in section 3.3. This calculation forms the second offline phase. Together, these steps offer a comprehensive solution to the inverse problem, effectively combining data-driven methods with *a priori* information to produce estimates of atmospheric scenarios.

4. Data organization

For the definition of our datasets we need the input variables describing the atmospheric scenarios and the output variables representing the corresponding FORUM simulated spectra.

The atmospheric state was build using the European Center for Medium range Weather Forecast Re-Analysis ERA5 database [28]. Additional atmospheric gases concentrations of CO_2, CO, CH_4, NO_2, HNO_3, SO_2 were taken from the Copernicus Atmosphere Monitoring Service global Greenhouse Gas reanalysis [14]. The remaining atmospheric gases were obtained from the initial guess 2 database [50]. The surface emissivity profiles were taken from the Huang database [30], with an ad-hoc procedure to associate the most suitable profile to the time and geolocation [58]. Finally, the *a priori* data were generated by perturbing the true profiles according to a precision matrix $\mathbf{S}_a$, constructed based on error estimates provided by the UK Met Office for routine assimilation of IASI products into their operational NWP system.

For the spectral simulation, the radiative transfer problem was solved using the CLouds and Atmospheric Inversion Module (CLAIM), an advanced version of the inversion code used in [57], based on the line by line radiative transfer model (LBLRTM) forward model [13]. For the inversion, the CLAIM code uses the combination of the OE approach, the Gauss–Newton minimization method with the Levenberg–Marquardt technique and a final Tikhonov regularization step with the iterative variable strength method [51, 56]. The FORUM instrument characteristics were taken from one of the two instrument concepts available in phase A/B1 of the FORUM E2E project.

Among the atmospheric variables, we focus on five key parameters for our reconstruction purposes: Earth surface temperature, air temperature, water vapor, ozone, and surface spectral emissivity. Earth surface temperature is a scalar, surface spectral emissivity is a vector depending on the wavenumber, normally approximated with a linear spline, while the other parameters are vertical profiles that can be defined in terms of atmospheric pressure levels, using a fixed grid with 41 levels. However, the surface altitude of the geolocation determines the lowest pressure level for the radiative transfer, so that the actual number of pressure levels can be smaller. To standardize the data for our study, we require a uniform data dimension. We map the pressure range of each point onto a set of 41 pressure levels and then interpolate all other vertical profiles onto this grid. This ensures that while the actual pressure levels for each point may vary, the total number remains constant. Consequently, the atmospheric state $\mathbf{x}$ is composed of $n = 425$ components: one value for surface temperature, 41 components for the profiles of air temperature, water vapor, and ozone, and 301 values for surface spectral emissivity, defined over the spectral interval from 100 to 1600 cm^{-1} and interpolated on a 5 cm^{-1} regular grid. Furthermore, the output variables are represented by the simulated FORUM spectra, which are given on $q = 4049$ spectral points within the same spectral interval of 100–1600 cm^{-1}. These spectra are generated by running LBLRTM on the corresponding input atmospheric scenarios.

For the purpose of our analysis, we collect information on atmospheric scenarios in January and July 2021 at 12:00 all over the world, focusing only on clear sky conditions, employing the same selection process established in [55]. To provide further details, our dataset encompasses information spanning the entire globe, utilizing a grid system with longitude and latitude increments of 10° and 5° respectively. To manage the dataset's size, we restrict the data to the initial 20 days of both July and January and for each geographical location, we identify the first and second available clear sky days in the database. Consequently, not all geographical points in our dataset feature a clear sky scenario.

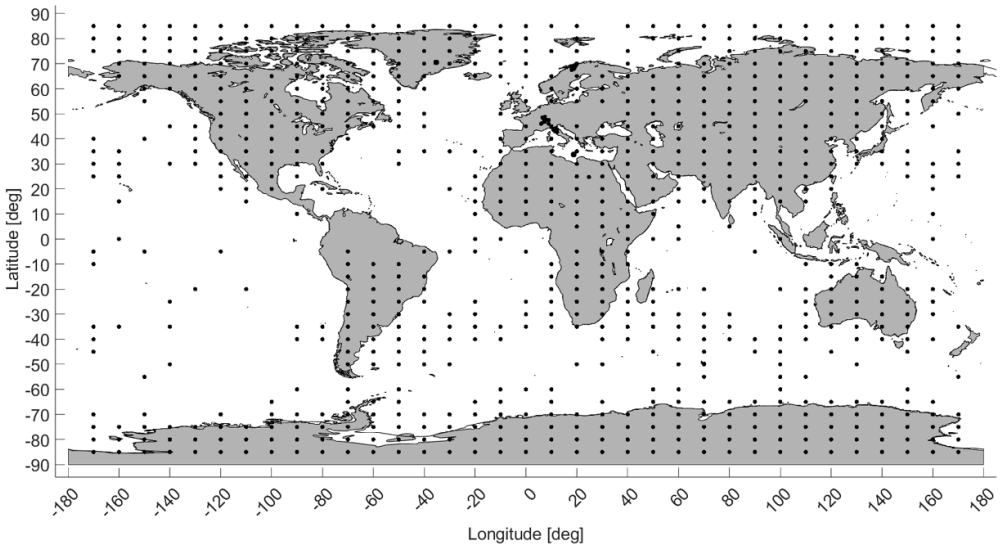

Figure 3. Geolocations of cases in TN1, represented by black dots for both January and July.

We organize these data into different sets. First, we create training set 1 (TN1), containing all the data related to the first clear sky day of both January and July. We enrich this set with additional points corresponding to well-known and typical locations with different climates and soil types, i.e. the Sahara desert, the city of Florence, the Black Forest, Greenland, Finland, Sweden, and the Mediterranean Sea. TN1 is represented by the input set $\mathbf{X} \in \mathbb{R}^{n\times m}$ of atmospheric scenarios and the output set $\mathbf{Y} \in \mathbb{R}^{q\times m}$ of spectra, with $m = 1708$ observations illustrated in figure 3.

Secondly, we use the remaining points corresponding to the second clear sky days in both January and July to construct two separate test sets, TS1 and TS2. We divide these points based on their indices, assigning odd indices to TS1 and even indices to TS2.

As a result, TS1 and TS2 consist of the input sets $\mathbf{X}_{\text{test1}} \in \mathbb{R}^{n\times m_{t_1}}$ and $\mathbf{X}_{\text{test2}} \in \mathbb{R}^{n\times m_{t_2}}$, respectively, and the output sets $\mathbf{Y}_{\text{test1}} \in \mathbb{R}^{q\times m_{t_1}}$ and $\mathbf{Y}_{\text{test2}} \in \mathbb{R}^{q\times m_{t_2}}$, where, $m_{t_1} = 396$ and $m_{t_2} = 402$ represent the number of observations for each test set, respectively. The composition of TS1 and TS2 is depicted figure 4

Finally, we generate *a priori* information, as described at the beginning of this section, and associate it with the input variables in TS1 and TS2. This information is represented by $(\mathbf{X}_a)_{\text{test1}}$ and $(\mathbf{X}_a)_{\text{test2}}$, respectively, and is linked to the corresponding precision matrices $(\mathbf{S}_a)_{\text{test1}}$ and $(\mathbf{S}_a)_{\text{test2}}$.

In our inversion scheme, detailed in section 3, we use TN1 in the first offline, fully data-driven training phase described in section 3.1. Then, TS1 is initially used to test this phase and later as a training set for the second offline training phase, as described in sections 3.2 and 3.3. Finally, TS2 is used to test the entire process chain. Starting from the preliminary data-driven results in TS1 we compute the experimental precision matrix $\mathbf{S}_\mathbf{x}$ used in the subsequent phases.

In general, the variables used in our simulations exhibit significant differences in magnitude. To ensure they are on a comparable scale during the computation of the data-driven solution, we apply a normalization procedure, as described in appendix A. This normalization

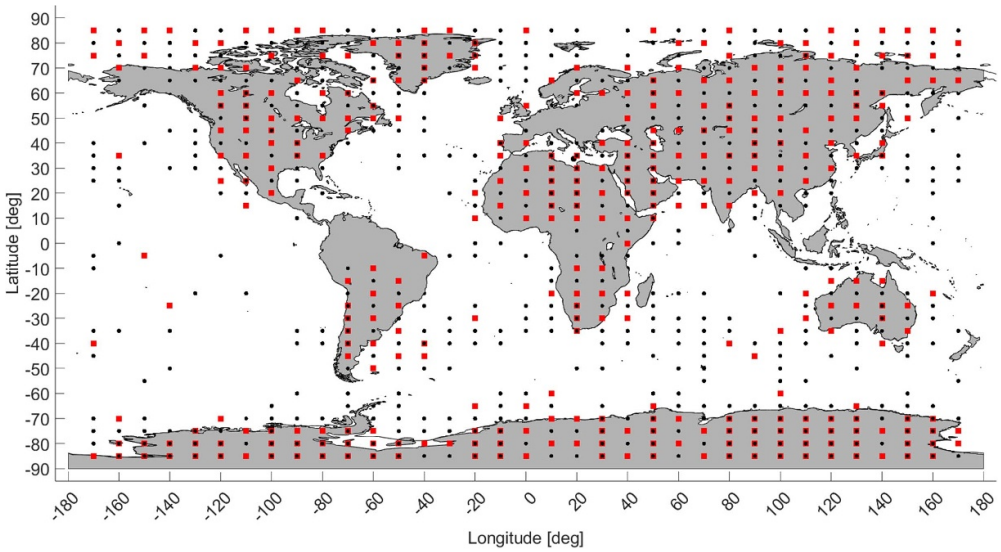

Figure 4. Geolocations of cases in TS1 and TS2 for both January and July, with red squares for TS1 and black dots for TS2.

is computed exclusively from the training data and then applied to the test data during evaluation. After obtaining the data-driven solution, we denormalize it to enable regularization using precision matrices as weights.

5. Results

In this section, we present the results from the various stages of our solution scheme. The experiments evaluate the proposed approach across each phase, specifically addressing the data-driven phase, section 5.1, and the integration of *a priori* information with the estimation of regularization parameters, section 5.2. A detailed analysis of the experimental findings follows, highlighting the efficiency of our method across different scenarios. These results demonstrate the contribution of each component to the overall performance in atmospheric scenario reconstruction. Finally, in section 5.3 we compare our solution with a full-physics procedure.

5.1. Fully data-driven results

In this section, we present the results obtained with the fully data-driven approach in section 3.1, by testing it on synthetic data. The inverse operator $\widehat{\mathbf{Z}}$ is trained using the TN1 database, which includes $m = 1708$ spectra $\mathbf{y}_j \in \mathbb{R}^q$, each associated with a parameter vector $\mathbf{x}_j \in \mathbb{R}^n$. We recall that in our setting, $q = 4049$ is the number of spectral channels and $n = 425$ is the number of physical parameters to be retrieved, as detailed in section 4. As shown in equation (11), we apply the pseudoinverse operator with a tolerance of 10^{-6}, below which singular values would be treated as zero. Figure 5 shows the singular values of the matrix $\mathbf{Y}$, whose columns contain the m spectra from TN1. As illustrated in the figure, in our specific case, all singular values remain above the threshold, indicating that the chosen tolerance is conservative and that no components are discarded.

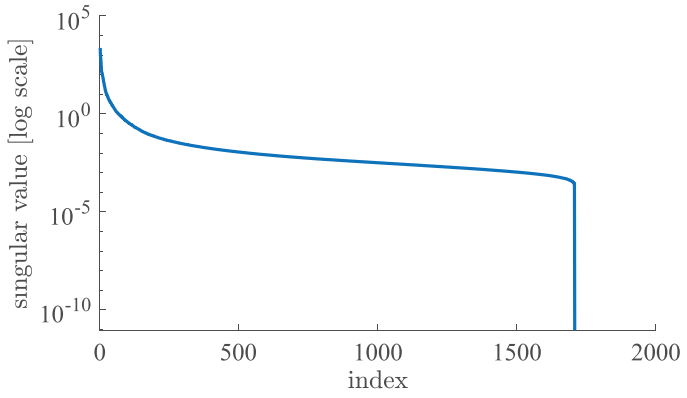

Figure 5. Singular values of matrix $\mathbf{Y}$ in decreasing magnitude. The plot shows a rapid decay in the singular values, suggesting that most of the information is concentrated in the leading components. However, in our specific case, all singular values remain above the chosen tolerance, and no dimensionality reduction is applied.

We use the resulting operator $\widehat{\mathbf{Z}}$ to test the $m_{t_1} = 396$ cases in TS1 database, as shown in equation (12). To evaluate the quality of the data-driven results, we apply the following quantifiers.

Given a single test spectrum $\mathbf{y}_j \in \mathbf{Y}_{\text{test1}}$, in order to evaluate the performance of the inverse problem in reconstructing the corresponding parameter vector $\mathbf{x}_j$, we compute the error vector $\mathbf{e}_j$, whose components are given by

$$e_j^i = x_j^i - (x_p)_j^i, \quad i = 1, \ldots, n,$$

where the prediction $(x_p)_j^i = \widehat{x}_j^i$ as in (12). In order to evaluate any possible systemic bias in the error behavior, an analysis of the whole m_{t_1} aggregated cases is carried out, computing the mean signed error $\mathbf{E}$ and the mean unsigned error $\mathbf{uE}$, whose components are:

$$E^i = \frac{1}{m_{t_1}} \sum_{j=1}^{m_{t_1}} e_j^i, \quad i = 1, \ldots, n,$$

$$(uE)^i = \frac{1}{m_{t_1}} \sum_{j=1}^{m_{t_1}} |e_j^i|, \quad i = 1, \ldots, n.$$

The first one is useful for evaluating a possible systematic bias in the error behavior, for example, whether a quantity is consistently over- or underestimated. The second one is more useful as a measure of overall accuracy.

Figure 6 presents the aggregated results for the 396 test cases in TS1. For each target variable, the plot displays the mean signed error $\mathbf{E}$ as a blue line, while the two orange lines represent the interval defined by $\mathbf{E} - \mathbf{uE}$ and $\mathbf{E} + \mathbf{uE}$, where $\mathbf{uE}$ is the mean unsigned error. The four panels refer to atmospheric temperature, water vapor, ozone, and surface spectral emissivity, respectively, following a clockwise arrangement from the top left. The first three variables are vertical profiles, reconstructed over 41 pressure levels varying across the test cases. To enable consistent aggregation, all profiles have been interpolated onto a common grid of 100 pressure levels, uniformly spaced between 1054 hPa and 1 hPa using a linear distribution. This

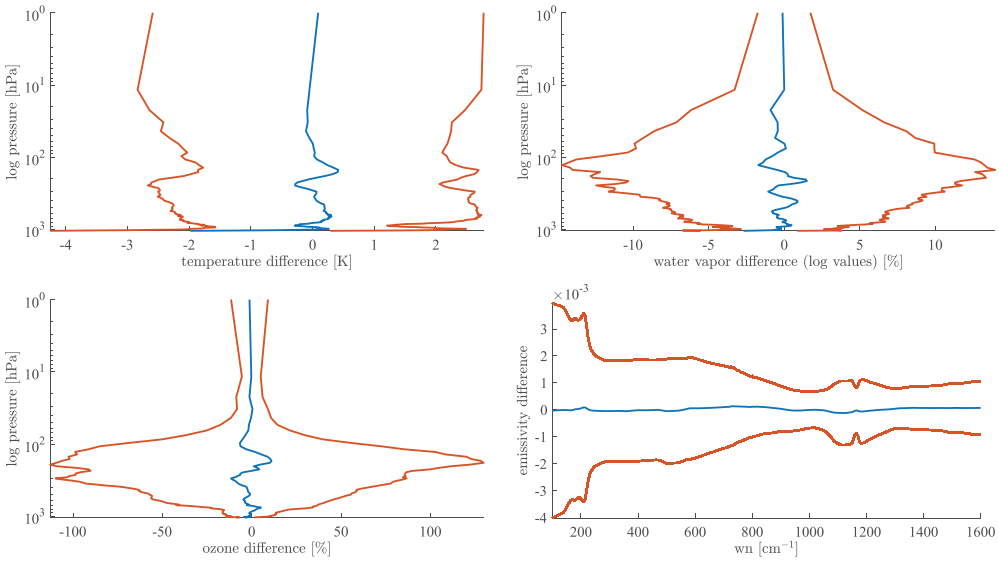

Figure 6. Results from the purely data-driven approach. Mean signed error for the entire TS1 is shown in blue and the two orange curves correspond to the signed error $\pm$ unsigned error. The panels, in order, correspond to temperature, water vapor, ozone, and surface spectral emissivity, with 'difference' referring to the discrepancy between the reconstructed and true value for each indicated quantity. The surface temperature difference is $5.901 \cdot 10^{-4} \pm 0.063$ K.

grid ensures full vertical coverage of the atmosphere, from the surface to the upper layers. The fourth variable, surface spectral emissivity, is defined over 301 spectral channels. A scalar value for surface temperature is also retrieved, but its reconstruction error is reported directly in the figure caption. We remark that, logarithmic values are employed for water vapor due to the substantial variations across pressure levels, a strategy that improves the training process. Overall, the signed errors tend to approach zero, remaining well within the unsigned error bars, which suggests that our method is generally unbiased across the different variables.

In figure 7, we present the pointwise results for a specific case in TS1, selected using a random number generator. As no aggregation is required, the original pressure levels of the selected case are used instead of the common interpolated grid adopted in figure 6. The four panels, matching the structure of figure 6, display the vertical profiles of atmospheric variables over their native 41 pressure levels, along with surface spectral emissivity across 301 spectral channels. A scalar surface temperature is also retrieved, with its reconstruction error reported in the figure caption. The figure shows reasonable signed error profiles for the quantities of interest. However, some unphysical oscillations and unusual local behavior are observed, particularly in the lower atmospheric layers for ozone and water vapor, and to a lesser extent, for temperature. While these oscillations remain within an acceptable range, their frequency and amplitude suggest non-physical behavior. In summary, based on the previous figures, the fully data-driven method generally provides high-quality reconstructions, assuming the training set is representative of the test set. However, the method fails in some individual cases. In other instances, strange local behavior and irregularities appear within otherwise accurate reconstructions. This fully data-driven approach is already capable of providing useful results

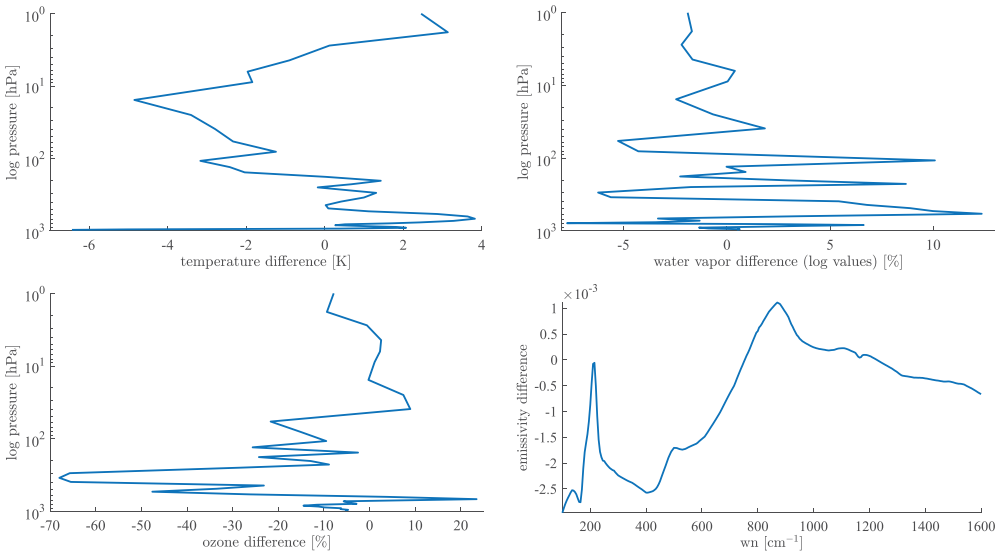

Figure 7. Results from the purely data-driven approach. Reconstruction signed error for case $j = 25$ in TS1. The panels, in order, correspond to temperature, water vapor, ozone, and surface spectral emissivity, with 'difference' referring to the discrepancy between the reconstructed and true value for each indicated quantity. The surface temperature difference is $-2.244 \cdot 10^{-2}$ K.

in most cases, though it could be improved. Incorporating *a priori* information, as suggested in section 3.2, may be beneficial.

5.2. Data-driven results with a priori correction

A priori information consists of atmospheric data known from climatology or external sources related to a specific geolocation and time. These data have been specifically selected for each test case and contain information for surface temperature, temperature, water vapor, ozone, and surface spectral emissivity, as described in section 4. This information is collected in the vector $\mathbf{x}_a \in \mathbb{R}^n$, with $n = 425$, in equation (16) representing an estimate of our solution. The atmospheric variability of these variables is represented in the related precision matrix $\mathbf{S}_a \in \mathbb{R}^{n\times n}$ constructed from the standard deviation and the correlation length of those variables provided by the UK Met Office for routine assimilation of IASI products into their operational NWP system.

At this point, the challenge is to determine the appropriate influence of the *a priori* information on our solution, which corresponds to selecting good values for $\boldsymbol{\lambda} \in \mathbb{R}^n$ in the minimization problem given by (13). It is important to note that the influence of the *a priori* information may vary for each of the five variables of interest, as well as for each component within these variables. In our experiments, we opt for a constant regularization parameter vector for each variable, resulting in a final minimization problem that involves determining five constant concatenated vectors. To address this, we analyze all $m_{t_1} = 396$ cases in TS1 and, for each case, we use equation (16) to solve equation (17) as detailed in sections 3.2 and 3.3. For this purpose, we employ an optimization algorithm based on the interior-point method, using a tolerance of 10^{-5} for both the objective function in equation (17) and the solution,

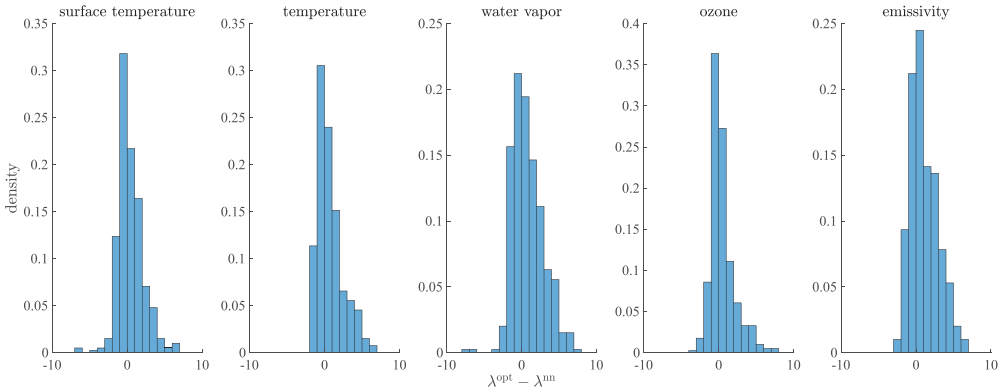

Figure 8. Difference between the optimal regularization parameters $\boldsymbol{\lambda}^{\text{opt}}$ and the estimated ones $\boldsymbol{\lambda}^{\text{nn}}$. The panels correspond, in order, to the surface temperature component, temperature component, water vapor component, ozone component, and surface spectral emissivity component.

i.e. the optimal regularization parameters. We initialize the algorithm with random vectors, each component ranging between 0 and 1. Specifically, we use Matlab's constrained optimization method `fmincon` with default settings, which is commonly used to find the minimum of constrained nonlinear multivariable functions.

It is important to note that the optimization process using the interior-point algorithm is repeated multiple times, specifically 10 times, with different initial conditions. This approach is known as the multi-start method and helps mitigate the risk of convergence to local minima and increases the likelihood of finding the global optimum. By running the algorithm multiple times, we can ensure the robustness of the solution and verify that the results are not overly dependent on the choice of initial values. For a detailed discussion of the bilevel optimization strategy for the five regularization variables and the aggregation of results via singular value decomposition (SVD), please refer to appendix B.

At this point, for cases in TS1, given the data-driven solutions $\widehat{\mathbf{x}}_j \in \mathbb{R}^n$, the corresponding *a priori* data $(\mathbf{x}_a)_j \in \mathbb{R}^n$, and the optimal regularization parameters $\boldsymbol{\lambda}_j^{\text{opt}} \in \mathbb{R}^n$, for $j = 1, \ldots, m_{t_1}$, we can train a neural network as detailed in equations (18) and (19) in section 3.3. We employ a feedforward neural network for all five variables, consisting of three hidden layers with sizes of 15, 10, and 5 units, respectively. In the hidden layers, we use a hyperbolic tangent activation function. The training is performed using the Levenberg–Marquardt optimization technique. Figure 8 presents histograms illustrating the network's performance in estimating the regularization parameters for the $m_{t_1} = 396$ cases in TS1. The figure is composed of five panels, each corresponding to one of the five variables. It shows the difference between the previously computed optimal regularization parameters $\boldsymbol{\lambda}^{\text{opt}}$ and the ones estimated by the network, denoted by $\boldsymbol{\lambda}^{\text{nn}}$.

To closely examine the neural network's performance, we focus on a single representative case from TS1. Figure 9 refers to case $j = 256$ and contains four panels corresponding to the vertical profiles of temperature, water vapor and ozone over their native 41 pressure levels, and surface spectral emissivity across 301 spectral channels. Once again, information on the surface temperature is provided directly in the figure caption. The panels show the signed errors between the true solution and the following: the fully data-driven solution (in red), the *a priori* solution (in purple), the regularized solution using the optimal regularization parameters (in

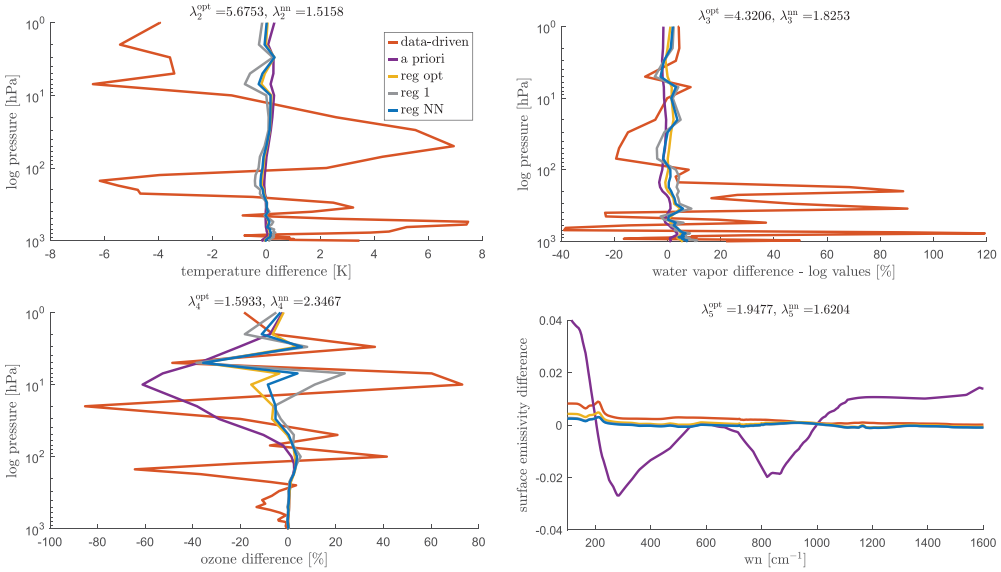

Figure 9. Comparison for case $j = 256$ between the signed reconstruction errors relative to the data-driven solution (red), the *a priori* solution (purple), the regularized solution using the optimal parameters (yellow), the regularized solution using each parameter fixed to 1 (gray), and the regularized solution using the estimated parameters (blue). The panels, in order, correspond to temperature, water vapor, ozone, and surface spectral emissivity, with 'difference' referring to the discrepancy between the reconstructed and true value for each indicated quantity. The respective regularization parameter values are reported at the top of each panel. For surface temperature, $\lambda_1^{opt} = 5.6753$ and $\lambda_1^{nn} = 2.8512$.

yellow), the regularized solution with each regularization parameter fixed to 1 (in gray), and the regularized solution using the estimated regularization parameters (in blue). The exact values of the regularization parameters are reported at the top of each panel. For the sake of clarity, we refer to them, using the notation λ_i, where the subscript $i = 1, \ldots, 5$, does not refer to the case index but to the variable index. Each λ_i represents a constant scalar value, corresponding to the regularization strength for the ith variable. As shown, the quality of the solution does not degrade when using the neural network's estimated regularization parameters.

At this point, we have all the elements to test the entire solution scheme on a set of new cases. In particular, we consider the $m_{t_2} = 402$ points in TS2. Figures 10 and 11 show a comparison between the fully data-driven phase and the complete solution procedure. Specifically, figure 10 presents the aggregated results for the 402 test cases in TS2. For each target variable, the mean signed error $\mathbf{E}$ is shown as a bold line, with the dotted lines indicating the interval defined by $\mathbf{E} - \mathbf{uE}$ and $\mathbf{E} + \mathbf{uE}$, where $\mathbf{uE}$ is the mean unsigned error. The fully data-driven solution is represented in orange and the regularized solution in blue. The four panels correspond to the usual four variables in the standard order. As in figure 6, the first three variables are aggregated onto a common grid of 100 pressure levels, ensuring consistent evaluation across the cases. Surface spectral emissivity is evaluated across 301 spectral channels, and the scalar surface temperature retrieval error is provided in the figure caption. Once again, we observe that the full method is unbiased and that regularization produces a significant improvement in the solution.

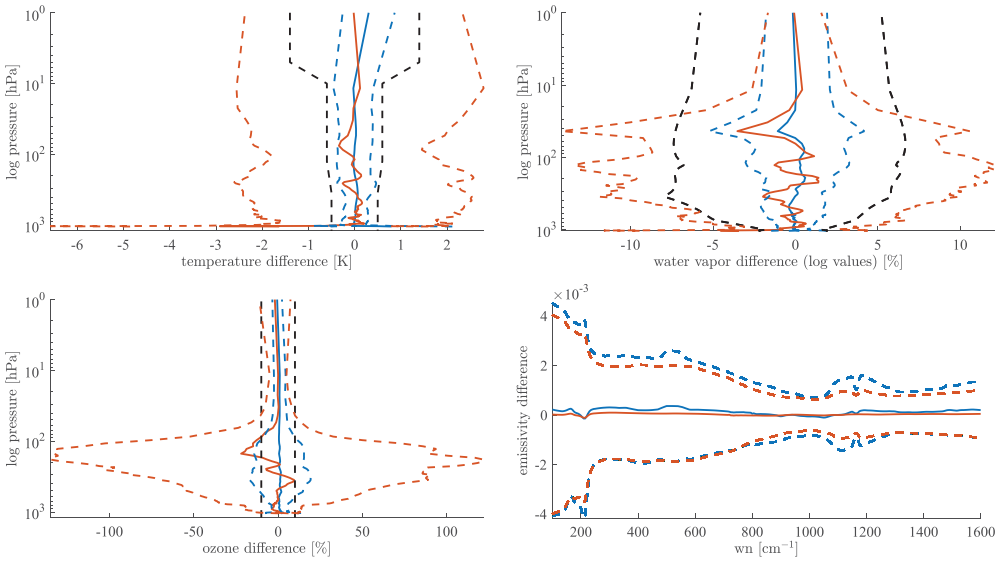

Figure 10. Mean signed error for the entire TS2 is shown using bold colored lines and dashed colored lines correspond to the signed error ± unsigned error. Errors related to the data-driven solution are represented in orange, while those for the regularized solution are in blue. Black dashed lines indicate the *a priori* errors. Each panel corresponds, in order, to atmospheric temperature, water vapor, ozone, and surface spectral emissivity, with 'difference' referring to the discrepancy between the reconstructed and true value for each indicated quantity. The *a priori* error ±0.05 for surface emissivity is omitted for clarity. The surface temperature difference for the data-driven solution is -0.002 ± 0.057 K and for the regularized solution is -0.01 ± 0.0714 K.

Next, we examine individual cases to assess whether the method now performs better even at a pointwise level. Figure 11 shows, for a single random case in TS2, the signed errors for the data-driven solution in orange, the regularized solution with each parameter fixed to 1 in gray, and our regularized solution in blue. It contains four panels corresponding to the vertical profiles of temperature, water vapor and ozone over their native 41 pressure levels, and surface spectral emissivity across 301 spectral channels. Once again, information on the surface temperature is provided directly in the figure caption. The values of the estimated regularization parameters are reported at the top of each panel using the same notation as in figure 9. As you can see, the regularization greatly improves the results, leading to a method that is quite accurate even for individual retrievals.

To conclude, we want to highlight that the optimal values of the regularization parameters generally differ from 1, typically ranging between 1.6 and 5.7 across the dataset. This variability highlights the necessity of adapting the regularization strength to the specific retrieval context, rather than adopting a fixed value. In fact, retrieval experiments performed with uniform $\lambda = 1$, as illustrated in figures 9 and 11, show degraded performance compared to those using the estimated values, confirming the effectiveness and relevance of this scenario-specific tuning.

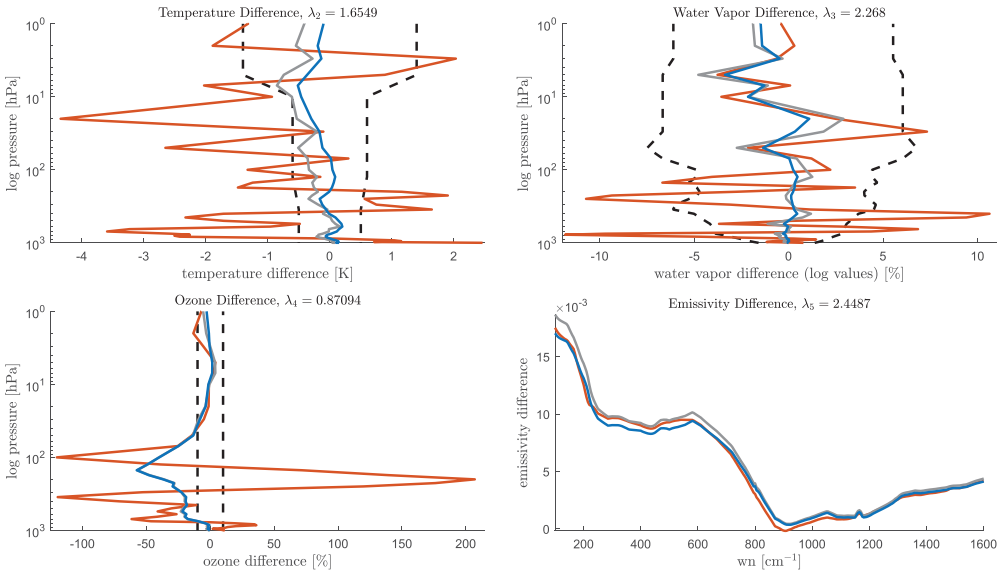

Figure 11. Signed errors for case 15 in TS2, with data-driven solutions shown in orange, regularized solutions with each parameter fixed to 1 in gray and our regularized solutions in blue. The dashed lines represent the *a priori* errors. The panels, in order, correspond to atmospheric temperature, water vapor, ozone, and surface spectral emissivity, with 'difference' referring to the discrepancy between the reconstructed and true value for each indicated quantity. The *a priori* error ± 0.05 for surface emissivity is omitted for clarity. The surface temperature difference for the data-driven solution is 0.019 K, for the regularized solution with each parameter fixed to 1 is -0.020 K, and for our regularized solution is -0.021 K. The values of the estimated regularization parameter components are reported at the top of each panel, with $\lambda_1 = 1.0993$.

5.3. Comparison with a full-physics method

Next, we compare our method with a full-physics approach, specifically the CLAIM, as described in section 4. In terms of computational time, our method outperforms the traditional approach. The CLAIM method requires approximately between 1 and 2 h to perform a single retrieval, depending on the number of Gauss–Newton iterations needed for convergence. The bottlenecks are the cross-section evaluation via the spectroscopic database and the convolution with the ISRF that must be repeated for the Jacobian of each retrieval parameter, or about 450 times for each Gauss–Newton step. In contrast, our method takes 0.178 s for the initial training phase and 128.984 s to train the neural network during the offline stage. The online stage requires about 0.0008 s for the fully data-driven solution of a single case, approximately 0.007 s to estimate the regularization parameters for a single case, and around 0.05 s for the entire solution scheme for a single case.

Of course, such a fast method cannot compete in terms of precision and accuracy with the full-physics method. To evaluate the performances of our method on the full TS2, we use the chi-square statistics, which measure the goodness of the spectrum reconstructed with our solution versus the full-physics solution [52].

More precisely, we use the reduced χ_r^2 statistics calculated as follows:

$$\chi_r^2 = \frac{(\mathbf{y} - \mathbf{F}(\mathbf{x}))^\top \mathbf{S}_\mathbf{y}^{-1} (\mathbf{y} - \mathbf{F}(\mathbf{x}))}{q}, \tag{20}$$

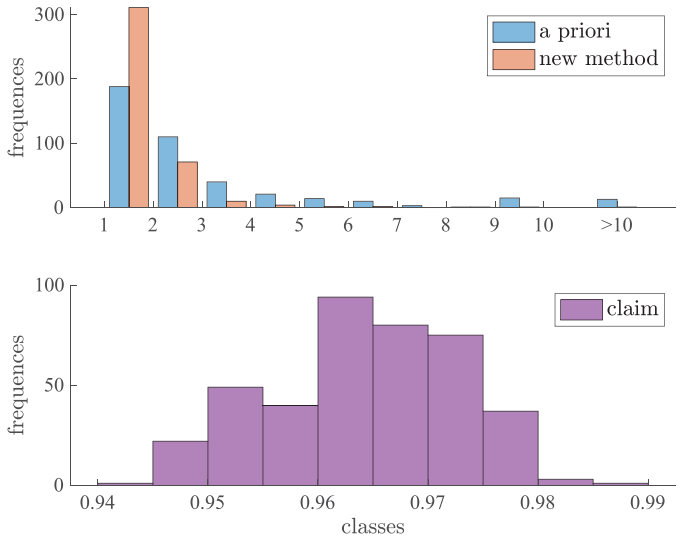

Figure 12. Binned distribution of the reduced chi-square statistic (χ^2_r), used to evaluate the goodness-of-fit between reconstructed and measured spectra in the TS2 dataset. Top panel: comparison between the initial *a priori* estimates (blue) and the results obtained using our proposed method (orange). Bottom panel: performance of the full-physics retrieval method (CLAIM). Lower values of χ^2_r indicate better agreement with the measured spectrum, with values around 1 expected for solutions consistent with instrumental noise. The figure shows that our method significantly improves upon the *a priori*, although it does not reach the accuracy of the full-physics solution.

where, as introduced in the previous sections, $\mathbf{y}$ is the measured spectrum, $\mathbf{F}(\mathbf{x})$ the simulated spectrum, $\mathbf{S_y}$ the measurements precision matrix and q the dimension of the spectrum. The expected value of the χ^2_r for a spectrum that is compatible with the instrumental noise is 1, and $(q-m)/q \sim 0.9$ for the reconstruction after the minimization. Of course, the full-physics method performs better, as it was expected. However, we also calculated the χ^2_r statistics obtained when using the *a priori* as a solution. The results are presented in figure 12. From the figure, we see that our solution improves the given *a priori*, though the *a priori* were calculated as a background error, hence with a reasonably small $\mathbf{S}_a$. It is important to get accurate *a priori* also as a starting point for the full-physics solution because the bias of a solution depends on the distance between the *a priori* and the truth [42, 58]. If the *a priori* were taken from the climatology, the average error would be much larger.

6. Conclusions

In this work, we present a data-driven approach for approximating the inverse mapping '$\mathbf{F}^{-1}$' in the retrieval problem, aiming for a solution that is both computationally efficient and robust. The use of simulated FORUM spectrometer measurements as training data, combined with the integration of climatological information, allows us to address the ill-posed nature of the inverse problem. A key feature of our method is the initial purely data-driven phase, where we approximate '$\mathbf{F}^{-1}$' without relying on any prior information, ensuring flexibility and

generalization across different scenarios. However, since the inverse problem is inherently ill-posed, approaches that do not apply regularization techniques are unable to accurately recover the profiles. The second key innovation lies in the use of a neural network to estimate the optimal regularization parameters during the retrieval process to incorporate prior information. Although our approach does not match the precision of full-physics retrieval methods, it offers near-instantaneous results, making it a valuable tool for applications where speed is crucial. Additionally, the proposed method's ability to provide *a priori* values that are statistically closer to the true atmospheric state suggests that it can reduce the bias in the inverse problem solution typically introduced by errors in the *a priori* data. Overall, our method provides a practical alternative to traditional retrieval approaches, especially in scenarios where computational efficiency and the ability to handle large datasets are essential.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.5281/zenodo.14010305 [54].

Acknowledgments

INdAM-GNCS supported the first author under Bando di concorso a n.30 mensilità di Borse di studio per l'estero A.A. 2022-2023. NRRP (National Recovery and Resilience Plan) supported the first and second authors under the Project EMM (Earth-Moon-Mars, Mission 4, Component 2, Investment 3.1, Project IR000038, CUPC53C22000870006). This work was partially supported by the National Science Foundation (NSF) under Grant [DMS-2152661] for M Chung. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Author contributions

Cristina Sgattoni 0000-0001-5734-0856
Conceptualization (equal), Data curation (equal), Formal analysis (equal), Funding acquisition (equal), Investigation (equal), Methodology (equal), Project administration (equal), Software (equal), Validation (equal), Writing – original draft (equal), Writing – review & editing (equal)

Luca Sgheri 0000-0002-6014-9363
Conceptualization (equal), Data curation (equal), Formal analysis (equal), Investigation (equal), Methodology (equal), Software (equal), Validation (equal), Writing – original draft (equal), Writing – review & editing (equal)

Matthias Chung 0000-0001-7822-4539
Conceptualization (equal), Formal analysis (equal), Investigation (equal), Methodology (equal), Software (equal), Validation (equal), Writing – original draft (equal), Writing – review & editing (equal)

Appendix A. Data normalization

Consider our matrix of input data $\mathbf{X} \in \mathbb{R}^{n\times m}$ and output data $\mathbf{Y} \in \mathbb{R}^{q\times m}$. The columns of $\mathbf{X}$ and $\mathbf{Y}$ correspond to different observations, and the rows to the different variables in each observation.

The variables in the training sets $\mathbf{X}$ and $\mathbf{Y}$ may differ significantly in magnitude. Consequently, it is often helpful to normalize the data so that all quantities of interest have a similar scale, allowing us to better exploit the correlations between each variable.

We employ the following *center-and-scale* normalization procedure, as follows:

$$\widehat{\mathbf{X}} = \operatorname{diag}\left(\sigma\left(\mathbf{X}\right)\right)^{-1}\left(\mathbf{X} - \overline{\mathbf{x}}\cdot\mathbf{1}^{\top}\right), \qquad \widehat{\mathbf{Y}} = \operatorname{diag}\left(\sigma\left(\mathbf{Y}\right)\right)^{-1}\left(\mathbf{Y} - \overline{\mathbf{y}}\cdot\mathbf{1}^{\top}\right), \tag{A.1}$$

where $\overline{\mathbf{x}}$, $\overline{\mathbf{y}}$ and $\sigma(\mathbf{X})$, $\sigma(\mathbf{Y})$ denote the vectors containing the mean and standard deviation of each row of $\mathbf{X}$, $\mathbf{Y}$ respectively, and $\mathbf{1}$ is the appropriately-sized vector containing all ones.

One may un-normalize the observations in a natural way, with the inverse maps given by:

$$\mathbf{X} = \operatorname{diag}\left(\sigma\left(\mathbf{X}\right)\right)\widehat{\mathbf{X}} + \overline{\mathbf{x}}\cdot\mathbf{1}^{\top}, \qquad \mathbf{Y} = \operatorname{diag}\left(\sigma\left(\mathbf{Y}\right)\right)\widehat{\mathbf{Y}} + \overline{\mathbf{y}}\cdot\mathbf{1}^{\top}. \tag{A.2}$$

In practice, we define these maps during the training phase, using the training data only. The maps are then saved and re-used, with no changes, on the test data during the testing phase. For this normalization procedure to be effective, it is, therefore, necessary that the space spanned by the training sets is sufficiently rich to represent the data contained in the test sets.

Appendix B. Methodology for independent minimization and aggregation of regularization variables

A key aspect of our optimization approach in section 5.2 is addressing the varying magnitudes of the five regularization variables. We perform the outer minimization of (17) independently for each variable, while still allowing the others to vary in each case, as they are not fully independent. More precisely, we change the target function each time for each variable, i.e. the outer problem in (17), while keeping the inner problem in (16) fixed at its original dimension. At the end of each optimization process, we obtain five distinct vectors for each case. Each vector contains five components, representing the optimized values for all five variables, but with the minimization focused specifically on one variable at a time. To aggregate the results and obtain a single vector that performs well across all variables we collect the five vectors into a square matrix $\mathbf{W}$. We then perform the SVD on this matrix and derive the optimal solution using a linear combination of the first two singular values σ_1 and σ_2 and the corresponding left singular vectors, where the coefficients are chosen empirically:

$$\mathbf{W} = \mathbf{U}\mathbf{S}\mathbf{V}^{\top}, \quad \boldsymbol{\lambda}^{\text{opt}} = \left|\tfrac{1}{4}\sigma_1\mathbf{u}_1 + \sigma_2\mathbf{u}_2\right|. \tag{B.1}$$

To justify this choice, figure B1 shows the typical decrease of the singular values of matrix $\mathbf{W}$. Similar patterns are observed in all cases, where the first two singular values capture the most informative components of the data.

In order to better understand the formulation, we analyze the relationship between matrix $\mathbf{W}$ and $\mathbf{W}\mathbf{W}^{\top}$.

By definition, all elements of $\mathbf{W}$ are non-negative, so this also applies to $\mathbf{W}\mathbf{W}^{\top}$. Since $\mathbf{W}\mathbf{W}^{\top}$ is non-negative, by the weak Perron–Frobenius theorem (see, for instance, chapter 8 of [43]), the eigenvector $\mathbf{u}_1$ corresponding to the largest eigenvalue μ_1 of $\mathbf{W}\mathbf{W}^{\top}$ can be chosen so that all its components are not negative. Moreover, since we verified that μ_1 is simple and that positive left and right eigenvectors exist for μ_1, by the Gantmacher theorem [65], $\mathbf{W}\mathbf{W}^{\top}$

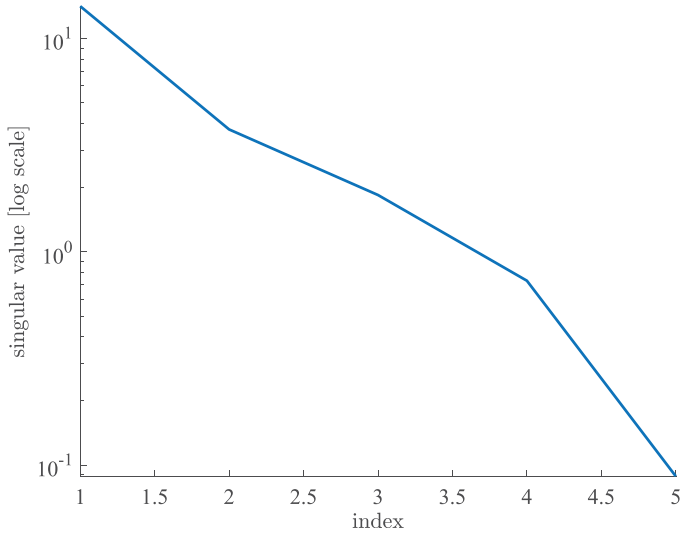

Figure B1. Typical $\mathbf{W}$ singular values behavior. The singular values, arranged from largest to smallest, reflect how much each component contributes to explaining the variance within the original matrix. The steep drop-off in singular values indicates that the first two components contain the bulk of the information, making it possible to reduce dimensionality with minimal data loss. Smaller singular values likely represent noise and can be omitted to achieve data compression.

is also irreducible, so we can apply the strong Perron–Frobenius theorem that guarantees that $\mathbf{u}_1$ has strictly positive components. By definition, $\mathbf{u}_1$ is also the left singular vector of $\mathbf{W}$ corresponding to the singular value $\sigma_1 = \sqrt{\mu_1}$.

Using again the Perron–Frobenius theorem, we know that $\mathbf{u}_1$ is the only eigenvector with all positive components, meaning that $\mathbf{u}_2$ is not entry-wise nonnegative. Given that the $\mathbf{u}_i$ vectors have unit norm and that σ_1 is larger than σ_2 (figure B1), the linear combination of the first two singular vectors, each weighted by its respective singular value, will statistically tend to have positive components. However, we observed empirically that better results are achieved by adding a coefficient of $1/4$ to $\mathbf{u}_1$. Therefore, as positive values are required for $\boldsymbol{\lambda}^{\text{opt}}$, we included the absolute value in equation (B.1).

References

[1] Afkham B M, Chung J and Chung M 2021 Learning regularization parameters of inverse problems via deep neural networks *Inverse Problems* **37** 105017
[2] Amato U, Antoniadis A and De Feis I 2006 Dimension reduction in functional regression with applications *Comput. Stat. Data Anal.* **50** 2422–46
[3] Amato U, Masiello G, Serio C and Viggiano M 2002 The σ-IASI code for the calculation of infrared atmospheric radiance and its derivatives *Environ. Modelling Softw.* **17** 651–67
[4] Antoulas A C, Beattie C A and Güğercin S 2020 *Interpolatory Methods for Model Reduction* (SIAM)
[5] Aspri A, Banert S, Öktem O and Scherzer O 2020 A data-driven iteratively regularized Landweber iteration *Numer. Funct. Anal. Optim.* **41** 1190–227
[6] Bak J, Liu X, Spurr R, Yang K, Nowlan C R, Miller C C, Abad G G and Chance K 2021 Radiative transfer acceleration based on the principal component analysis and lookup table of corrections: optimization and application to UV ozone profile retrievals *Atmos. Meas. Tech.* **14** 2659–72

[7] Bardsley J M 2018 *Computational Uncertainty Quantification for Inverse Problems: An Introduction to Singular Integrals* (SIAM)
[8] Biswal P, Avdijaj J, Parente A and Coussement A 2025 Physics informed neural networks to solve radiative transfer equation in absorbing-scattering media *J. Quant. Spectrosc. Radiat. Transf.* **344** 109509
[9] Calvetti D and Somersalo E 2007 *An Introduction to Bayesian Scientific Computing: ten Lectures on Subjective Computing* vol 2 (Springer)
[10] Carbon dioxide now more than 50% higher than pre-industrial levels (available at: www.noaa.gov/news-release/carbon-dioxide-now-more-than-50-higher-than-pre-industrial-levels) (Accessed 12 February 2024)
[11] Chandrasekhar S 1960 *Radiative Transfer* (Courier Corporation)
[12] Chung J and Chung M 2017 Optimal regularized inverse matrices for inverse problems *SIAM J. Matrix Anal. Appl.* **38** 458–77
[13] Clough S, Shephard M, Mlawer E, Delamere J, Iacono M, Cady-Pereira K, Boukabara S and Brown P 2005 Atmospheric radiative transfer modeling: a summary of the AER codes *J. Quant. Spectrosc. Radiat. Transfer* **91** 233–44
[14] Copernicus Atmosphere Monitoring Service (CAMS) 2020 CAMS global greenhouse gas reanalysis (EGG4) (available at: https://atmosphere.copernicus.eu/cams-global-greenhouse-gas-reanalysis) (Accessed 23 September 2024)
[15] Cortesi U, Bianco S D, Gai M, Laurenza L, Ceccherini S, Carli B, Barbara F and Buchwitz M 2014 Sensitivity analysis and application of KLIMA algorithm to GOSAT and OCO validation - KLIMA-IASI, final report of project ESA-ESRIN/contract n. 21612/08/i-ol *Technical Report* IFAC-TSRR (IFAC-CNR) 1–153
[16] Crevoisier C *et al* 2014 Towards IASI-New Generation (IASI-NG): impact of improved spectral resolution and radiometric noise on the retrieval of thermodynamic, chemistry and climate variables *Atmos. Meas. Tech.* **7** 4367–85
[17] Della Fera S, Fabiano F, Raspollini P, Ridolfi M, Cortesi U, Barbara F and von Hardenberg J 2023 On the use of Infrared Atmospheric Sounding Interferometer (IASI) spectrally resolved radiances to test the EC-earth climate model (v3. 3.3) in clear-sky conditions *Geosci. Model Dev.* **16** 1379–94
[18] Dinelli B M *et al* 2021 The ESA MIPAS/Envisat level2-v8 dataset: 10 years of measurements retrieved with ORM v8.22 *Atmos. Meas. Tech.* **14** 7975–98
[19] Doicu A, Daun K and Trautmann T 2010 *Numerical Regularization for Atmospheric Inverse Problems* (Springer)
[20] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* vol 375 (Springer)
[21] Estatico C, Schenone V, Fedeli A and Randazzo A 2024 Application of a mild data-driven technique to Lippmann–Schwinger inverse scattering in variable-exponent lebesgue spaces for microwave imaging *Inverse Problems* **40** 065007
[22] EUMETSAT 2023 EPS-SG IASI-NG (available at: www.eumetsat.int/eps-sg-iasi-ng) (Accessed 20 February 2023)
[23] Genedy R A, Chung M, Shortridge J E and Ogejo J A 2024 A physics-informed long short-term memory (LSTM) model for estimating ammonia emissions from dairy manure during storage *Sci. Total Environ.* **912** 168885
[24] Godson W L 1953 The evaluation of infra-red radiative fluxes due to atmospheric water vapour *Q. J. R. Meteorol. Soc.* **79** 367–79
[25] Gordon I *et al* 2022 The HITRAN2020 molecular spectroscopic database *J. Quant. Spectrosc. Radiat. Transfer* **277** 107949
[26] Guarnieri A M 2002 SAR interferometry and statistical topography *IEEE Trans. Geosci. Remote Sens.* **40** 2567–81
[27] Hansen P C 2010 *Discrete Inverse Problems: Insight and Algorithms* (SIAM)
[28] Hersbach H *et al* 2020 The ERA5 global reanalysis *Q. J. R. Meteorol. Soc.* **146** 1999–2049
[29] Huang H-L, Revercomb H E, Thom J, Antonelli P, Osborne B, Tobin D, Knutson B, Garcia R, Dutcher S and Li J 2001 Principal component analysis, compression and retrieval – application to high spectral resolution infrared measurements *Optical Remote Sensing* (Optica Publishing Group) (available at: https://opg.optica.org/abstract.cfm?URI=ORS-2001-OWA3), p OWA3
[30] Huang X, Chen X, Zhou D K and Liu X 2016 An observationally based global band-by-band surface emissivity dataset for climate and weather simulations *J. Atmos. Sci.* **73** 3541–55

[31] IPCC sixth assessment report (available at: www.ipcc.ch/report/ar6/wg1/) (Accessed 12 February 2024)
[32] Johnson B T, Dang C, Stegmann P, Liu Q, Moradi I and Auligne T 2023 The community radiative transfer model (CRTM): community-focused collaborative model development accelerating research to operations *Bull. Am. Meteorol. Soc.* **104** E1234–51
[33] Jolliffe I T 2002 *Principal Component Analysis* 2nd edn (Springer)
[34] Krishnan P, Ramanujam K S and Balaji C 2012 An artificial neural network based fast radiative transfer model for simulating infrared sounder radiances *J. Earth Syst. Sci.* **121** 891–901
[35] Liang X and Liu Q 2021 Applying deep learning to clear-sky radiance simulation for VIIRS with community radiative transfer model—part 1: develop AI-based clear-sky mask *Remote Sens.* **13** 222
[36] Liou K 2002 (ed) *An Introduction to Atmospheric Radiation* (*International Geophysics*) vol 84 (Academic)
[37] Maestri T, Martinazzo M, Cossich W, Serio C, Masiello G and Venafra S 2024 Fast radiative transfer in multiple scattering atmospheres at far and mid infrared wavelengths *AIP Conf. Proc.* **2988** 030003
[38] Martinazzo M, Magurno D, Cossich W, Serio C, Masiello G and Maestri T 2021 Assessment of the accuracy of scaling methods for radiance simulations at far and mid infrared wavelengths *J. Quant. Spectrosc. Radiat. Transfer* **271** 107739
[39] Masiello G *et al* 2024 Developments of the σ-IASI/F2N radiative transfer model: a new linear-in-T approach for accurate treatment of clouds *Proc. SPIE* **13193** 1319303
[40] Masiello G, Serio C, Maestri T, Martinazzo M, Masin F, Liuzzi G and Venafra S 2024 The new σ-IASI code for all sky radiative transfer calculations in the spectral range 10 to 2760 cm^{-1}: *sigma*-IASI/F2N *J. Quant. Spectrosc. Radiat. Transfer* **312** 1–19
[41] Matricardi M 2010 A principal component based version of the RTTOV fast radiative transfer model *Q. J. R. Meteorol. Soc.* **136** 1823–35
[42] Merchant C J, Saux-Picart S and Waller J 2020 Bias correction and covariance parameters for optimal estimation by exploiting matched *in-situ* references *Remote Sens. Environ.* **237** 111590
[43] Meyer C D 2010 *Matrix Analysis and Applied Linear Algebra* (Society for Industrial and Applied Mathematics)
[44] Noh Y-C, Sohn B-J, Kim Y, Bell W and Suaders R 2017 A new infrared atmospheric sounding interferometer channel selection and assessment of its impact on Met Office NWP forecasts *Adv. Atmos. Sci.* **34** 1265–81
[45] Palchetti L *et al* 2020 FORUM: unique far-infrared satellite observations to better understand how Earth radiates energy to space *Bull. Am. Meteorol. Soc.* **101** 1–52
[46] Paul M, Aires F, Prigent C, Trigo I F and Bernardo F 2012 An innovative physical scheme to retrieve simultaneously surface temperature and emissivities using high spectral infrared observations from IASI *J. Geophys. Res.: Atmos.* **117** D11302
[47] Pincus R, Mlawer E J and Delamere J S 2019 Balancing accuracy, efficiency and flexibility in radiation calculations for dynamical models *J. Adv. Model. Earth Syst.* **11** 3074–89
[48] Pyatkin V P, Rublev A N, Rusin E V and Uspenskii A B 2015 A fast radiative transfer model for hyperspectral IR satellite sounders *Pattern Recognit. Image Anal.* **25** 514–6
[49] Raissi M, Perdikaris P and Karniadakis G E 2017 Physics informed deep learning (part I): data-driven solutions of nonlinear partial differential equations (arXiv:1711.10561)
[50] Remedios J J, Leigh R J, Waterfall A M, Moore D P, Sembhi H, Parkes I, Greenhough J, Chipperfield M P and Hauglustaine D 2007 MIPAS reference atmospheres and comparisons to V4.61/V4.62 MIPAS level 2 geophysical data sets *Atmos. Chem. Phys. Discuss.* **7** 9973–10017
[51] Ridolfi M and Sgheri L 2011 Iterative approach to self-adapting and altitude-dependent regularization for atmospheric profile retrievals *Opt. Express* **19** 26696–709
[52] Rodgers C D 2000 *Inverse Methods for Atmospheric Sounding* (World Scientific) (available at: www.worldscientific.com/doi/abs/10.1142/3171)
[53] Saunders R *et al* 2018 An update on the RTTOV fast radiative transfer model (currently at version 12) *Geosci. Model Dev.* **11** 2717–37
[54] Sgattoni C, Sgheri L and Chung M 2024 Training set and test set for fast atmospheric radiative transfer inversion - clear sky (versione v1) *Zenodo* (available at: https://doi.org/10.5281/zenodo.14010305)
[55] Sgattoni C, Ridolfi M, Zugarini C and Sgheri L 2024 Characterization of surface spectral emissivity retrieved from EE9-FORUM simulated measurements *Remote Sens. Earth Syst. Sci.* **7** 15–25

[56] Sgheri L, Raspollini P and Ridolfi M 2022 Auto-adaptive tikhonov regularization of water vapor profiles: application to FORUM measurements *Appl. Anal.* **101** 3733–43
[57] Sgheri L *et al* 2022 The FORUM end-to-end simulator project: architecture and results *Atmos. Meas. Tech.* **15** 573–604
[58] Sgheri L, Sgattoni C and Zugarini C 2025 Determination of emissivity profiles using a bayesian data-driven approach *Math. Comput. Simul.* **229** 512–24
[59] Simmons A 2010 Monitoring atmospheric composition and climate (https://doi.org/10.21957/vqcvmlyg) (available at: www.ecmwf.int/node/17465)
[60] Stamnes K, Thomas G E and Stamnes J J 2017 *Radiative Transfer in the Atmosphere and Ocean* 2nd edn (Cambridge University Press)
[61] Stegmann P *et al* 2022 A deep learning approach to fast radiative transfer *J. Quant. Spectrosc. Radiat. Transfer* **281** 107971
[62] Ukkonen P 2022 Exploring pathways to more accurate machine learning emulation of atmospheric radiative transfer *J. Adv. Model. Earth Syst.* **14** e2021MS002875
[63] Ukkonen P and Hogan R J 2023 Implementation of a machine-learned gas optics parameterization in the ECMWF integrated forecasting system: RRTMGP-NN 2.0 *Geosci. Model Dev.* **16** 3241–61
[64] Wallace J M and Hobbs P V (eds) 2006 *Atmospheric Science* 2nd edn (Academic)
[65] Weiss G 1960 Applications of the theory of matrices. f. r. gantmacher. translated and revised by J. L. Brenner, D. W. Bushaw and S. Evanusa. interscience, New York, 1959. ix + 317 pp. $9 *Science* **131** 405–6
[66] Zdunkowski W, Trautmann T and Bott A 2007 *Radiation in the Atmosphere: A Course in Theoretical Meteorology* (Cambridge University Press)
[67] Zhang X, Zhang Y, Zhang Y, Liu Y, Wang Y, Xue X and Guo Y 2021 An improved method combining ANN and 1D-Var for the retrieval of atmospheric temperature profiles from FY-4A/GIIRS hyperspectral data *Remote Sens.* **13** 481