Scientific Reports

https://doi.org/10.1038/s41598-026-57822-6

Article in Press

Latent twins: a framework for scene recognition and fast radiative transfer inversion in FORUM all-sky observations

Received: 4 February 2026

Accepted: 9 June 2026

Published online: 19 June 2026

Cite this article as: Sgattoni C., Sgheri L., Chung M. *et al.* Latent twins: a framework for scene recognition and fast radiative transfer inversion in FORUM all-sky observations. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-57822-6

Cristina Sgattoni, Luca Sgheri, Matthias Chung & Michele Martinazzo

Latent Twins: A Framework for Scene Recognition and Fast Radiative Transfer Inversion in FORUM All-Sky Observations

Cristina Sgattoni[1,5,*], **Luca Sgheri**[2,5], **Matthias Chung**[3], **Michele Martinazzo**[4,*]

[1]Institute of BioEconomy (IBE), National Research Council (CNR), Via Madonna del Piano, 10, Sesto Fiorentino, 50019, Firenze, Italy.

[2]Institute of Applied Mathematics (IAC), National Research Council (CNR), Via Madonna del Piano, 10, Sesto Fiorentino, 50019, Firenze, Italy.

[3]Department of Mathematics, Emory University, 400 Dowman Drive, Atlanta, GA, USA.

[4]Department of Physics and Astronomy "Augusto Righi", University of Bologna, Viale Berti Pichat 6/2, Bologna, Italy.

[5]INdAM Research Group GNCS, P.le Aldo Moro, 5, 00185 Roma, Italy.

[*]Corresponding authors: `cristina.sgattoni@cnr.it`, `michele.martinazzo2@unibo.it`.

10 June 2026

Abstract. The FORUM (Far-infrared Outgoing Radiation Understanding and Monitoring) mission will provide, for the first time, systematic far-infrared spectral measurements of Earth's outgoing radiation, enabling improved understanding of atmospheric processes and the radiation budget. Retrieving atmospheric states from these observations constitutes a high-dimensional, ill-posed inverse problem, particularly under cloudy-sky conditions where multiple-scattering effects are present. In this work, we develop a data-driven, physics-aware inversion framework for FORUM all-sky retrievals based on *latent twins*: coupled autoencoders for atmospheric states and spectra, combined with bidirectional latent-space mappings. A lightweight model-consistency correction ensures physically plausible cloud variable reconstructions. The resulting framework demonstrates potential for retrievals of atmospheric, cloud and surface variables, providing information that can serve as a prior, initial guess, or surrogate for computationally expensive full-physics inversion methods. It also enables robust scene classification and near-instantaneous inference, making it suitable for operational near-real-time applications. We demonstrate its performance on synthetic FORUM-like data and discuss implications for future data assimilation and climate studies.

Keywords: deep learning, inverse problems, autoencoders, latent twins, remote sensing, atmospheric retrieval, scene recognition

Submitted to:

1. Introduction and Background

FORUM (Far-infrared Outgoing Radiation Understanding and Monitoring) [43] is the ninth Earth Explorer mission selected by the European Space Agency in 2019 and scheduled to launch in 2028. The mission will employ a Fourier Transform Spectrometer to measure Earth's outgoing spectral radiances over the spectral range from 100 to 1600 cm^{-1}. This will provide, for the first time, systematic and spectrally resolved observations in the far-infrared (FIR) region (100-667 cm^{-1}) from space [20]. FORUM's primary objective is to improve our understanding of Earth's infrared radiation budget and to characterize the role of FIR radiation in shaping the climate system.

Different components of the Earth system, including the surface, atmospheric gases, aerosols, and clouds, interact with electromagnetic radiation, producing distinct spectral signatures in the outgoing radiance. In the FIR, water vapor exhibits particularly strong absorption features, making FORUM measurements highly valuable for constraining atmospheric humidity and related climate processes [40]. In regions of low humidity, reduced water vapor absorption enables the retrieval of FIR surface emissivity from satellite observations [6]. FIR radiances are also strongly influenced by clouds and show greater sensitivity to ice particle scattering than those in the mid-infrared [32, 46, 72]. FORUM measurements will therefore help to characterize the impact of clouds on the Earth's radiation budget, with a particular focus on high-level ice clouds.

The propagation and interaction of electromagnetic radiation through the atmosphere are described by the radiative transfer equation. Analytic solutions exist only for idealized, homogeneous, non-scattering media. In the real atmosphere, which exhibits strong vertical variability in temperature, composition, and cloud properties, radiative transfer must be solved numerically [18].

In general, let $\mathbf{x} \in \mathbb{R}^n$ denote the atmospheric state, and let $\mathbf{y} \in \mathbb{R}^q$ denote the corresponding radiance then,

$$\mathbf{y} = \mathbf{F}^{\rightarrow}(\mathbf{x}) + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{F}^{\rightarrow} : \mathbb{R}^n \to \mathbb{R}^q$ represents the composition of the radiative transfer calculations and instrumental effects, and $\boldsymbol{\varepsilon} \in \mathbb{R}^q$ denotes additive measurement noise. This is known as *direct* or *forward* problem. In this work, we are primarily interested in the *inverse*, or *retrieval*, problem, namely, reconstructing the atmospheric state $\mathbf{x} \in \mathbb{R}^n$ from measurements $\mathbf{y} \in \mathbb{R}^q$.

Numerical inversion techniques, such as those based on Optimal Estimation (OE), are well established and widely used [45, 48, 49, 63]. In a variational setting, the optimal state is obtained by minimizing a cost functional that balances data fidelity and regularization,

$$\min_{\mathbf{x} \in \mathcal{X}} \quad \mathcal{D}\big(\mathbf{F}^{\rightarrow}(\mathbf{x}), \mathbf{y}\big) + \mathcal{R}(\mathbf{x}), \tag{2}$$

where $\mathcal{D}$ measures the discrepancy between the observations $\mathbf{y}$ and the model prediction $\mathbf{F}^{\rightarrow}(\mathbf{x})$, and $\mathcal{R}$ is a regularization term that penalizes implausible solutions [16, 62]. The forward operator can be expressed as the composition $\mathbf{F}^{\rightarrow} = \mathbf{p} \circ \mathbf{m}$, where $\mathbf{m}$ denotes the solution operator of the underlying physical model and $\mathbf{p}$ a projection onto the measurement space $\mathcal{Y}$.

While classical inversion techniques remain powerful and widely used, they are often hindered by several limitations, mainly related to the need for careful hyperparameter tuning, sensitivity to problem setup, intrinsic ill-posedness, curse of dimensionality, and the substantial computational cost associated with repeated forward simulations [10, 24]. In remote sensing retrievals, these challenges become particularly severe due to the high dimensionality of the problem and time-consuming radiative transfer models. For example, with the FORUM spectrometer expected to produce over 10,000 spectra daily, and each full-physics forward simulation requiring several minutes, brute-force inversion is infeasible for near-real-time (NRT) applications as required for weather and climate prediction [9]. Current approaches aim to mitigate these limitations by accelerating the forward model. Notable examples include fast radiative transfer codes such as σ-IASI/F2N (Far-to-Near Infrared) [3, 35, 37], RTTOV [41, 47] or PCRTM [29].

In response to these limitations, data-driven approaches have emerged as powerful alternatives. Rather than relying exclusively on explicit forward models, these methods take advantage of large-scale simulated and curated datasets to directly learn the mappings between measurements and unknown states. Deep neural networks, in particular, have demonstrated the ability to approximate highly nonlinear inverse operators and to accelerate retrieval by orders of magnitude compared with iterative variational schemes [4, 28]. For instance, autoencoders and related latent-variable models can be used to capture low-dimensional structure within high-dimensional data, providing surrogates that constrain reconstructions to physically plausible manifolds [5, 27].

In previous work, we developed a data-driven inversion approach restricted to clear-sky conditions, combining a pseudoinverse with weighted regularization implemented via neural networks [50]. While effective as a proof of concept, the approach did not fully exploit the expressive power of modern nonlinear deep learning models, nor did it address the additional challenges posed by cloudy-sky conditions, where multiple scattering and radiative degeneracies make inversion particularly difficult.

In this work, we tackle these challenges by developing a data-driven atmospheric retrieval framework based on *latent twins* [14], designed for fully all-sky applicability and practical use across the FORUM mission. Building on recent advances in latent twins and paired autoencoders for inverse problems [12, 13, 21], we construct a data-driven yet physics-aware surrogate inversion framework for FORUM all-sky observations. The key idea is to learn coupled autoencoders for atmospheric states and spectra, together with (bi)-directional mappings between their latent spaces, so that the latent coupling acts as a surrogate forward and inverse operator, effectively creating a latent twin of the original forward and/or inverse map. On top of this latent surrogate, we introduce a lightweight model-consistency correction targeted at cloud-related variables, enforcing physically meaningful relationships among cloud microphysical properties. The resulting framework shows potential for the retrieval of atmospheric, cloud, and surface variables, providing information that can be used as a prior, initial guess, or surrogate for computationally expensive full-physics inversion methods. It also enables robust scene classification and, once trained, near-instantaneous inference, making it a promising candidate for operational near-real-time applications and future data assimilation efforts.

This paper is organized as follows. Section 2 presents the latent twins framework, detailing the coupled autoencoders, latent-space mappings, and the model-consistency corrections for cloud variables. In Section 3, we describe the construction of the training dataset, while Section 4 details the organization of the atmospheric state and observation vectors. Numerical results and computational costs are presented in Section 5. In particular, Section 5.1 describes technical details of the models, including network architectures, hyperparameters, and loss functions. Section 5.2 examines retrieval performance on atmospheric variables. Section 5.3 covers scene classification, including clear, cloudy, and cloud-type identification, as well as the localization of cloud layers. Finally, Section 5.4 assesses the impact of the retrievals on reconstructed radiances via the forward model, including an additional correction for clear-sky cases. We conclude with a discussion of implications and future directions in Section 6. Two appendices are also included. In Appendix A, we review the radiative transfer model, including the treatment of multiple scattering, and describe the σ-IASI/F2N forward model used to generate training data. Appendix B describes the preprocessing of selected variables and how their transformation influences the reconstruction of radiances from the retrieved quantities. A Supplementary Material document is also provided. The first section presents a detailed analysis of the performance of the two autoencoder components of the proposed architecture, the second section compares the proposed latent-twins framework with a simpler baseline architecture, and the third section reports retrieval results for cloud-related variables before the application of the correction models.

2. Latent Twin Radiative Transfer Inversion

Data-driven inversion techniques replace or augment classical iterative solvers by learning components of the inverse process directly from data. One class of methods trains a neural network to approximate the reconstruction operator $\mathbf{y} \mapsto \mathbf{x}$ from paired training data $\{(\mathbf{y}, \mathbf{x})\}$, enabling fast inference but often requiring additional mechanisms to ensure robustness and data consistency [4]. A second class, commonly referred to as *unrolled* (or *algorithm-unfolded*) methods, derives neural architectures by unrolling and truncating classical iterative schemes and learning iteration-dependent parameters or proximal operators, thereby combining interpretability with state-of-the-art performance [1, 19].

Alternative approaches focus on *prior learning*. Plug-and-play methods incorporate powerful learned denoisers as implicit regularizers within model-based optimization frameworks [2, 65], while generative models provide explicit low-dimensional priors by restricting reconstructions to the range of a trained generator [7]. More recently, diffusion and score-based generative models have emerged as flexible learned priors that can be combined with known measurement operators at test time, enabling sampling-based reconstructions and principled uncertainty quantification [58].

Latent twins framework. A promising recent direction in data-driven inversion is the development of *latent twins* as surrogate models for forward and inverse operators. Realized as paired autoencoders, latent twins learn forward and inverse mappings through coupled latent representations, thereby improving stability and interpretability in ill-posed problems; see, for example, likelihood-free estimation [13], Bayes-risk minimization [21], and the unifying perspective in *Good Things Come in Pairs* [12].

In this work, we adopt the latent twins framework as a data-driven yet physics-aware approach to atmospheric retrieval. Here, physics-awareness stems from restricting solutions to a learned physically plausible manifold and from additional model-consistency corrections introduced below. The central idea is to represent both the atmospheric state and the observed spectrum in low-dimensional latent spaces and to learn the forward and inverse relationships directly between these representations; see Figure 1. Let $\mathcal{X} \subseteq \mathbb{R}^n$ denote the space of atmospheric states and $\mathcal{Y} \subseteq \mathbb{R}^q$ the space of radiance spectra. We define two autoencoders: an atmospheric-state autoencoder $a_x = d_x \circ e_x$, mapping $\mathcal{X}$ to itself through a latent space $\mathcal{Z}_x$, and a spectral autoencoder $a_y = d_y \circ e_y$, mapping $\mathcal{Y}$ through a latent space $\mathcal{Z}_y$. The key component of latent twins is the introduction of *trainable* mappings between $\mathcal{Z}_x$ and $\mathcal{Z}_y$. A latent forward map $s^{\rightarrow} : \mathcal{Z}_x \rightarrow \mathcal{Z}_y$ approximates the action of the radiative transfer operator in reduced coordinates, while a latent inverse map $s^{\leftarrow} : \mathcal{Z}_y \rightarrow \mathcal{Z}_x$ serves as a latent retrieval operator. Combined with the encoders and decoders, these mappings define surrogate forward and inverse operators,

$$f^{\rightarrow}(\mathbf{x}) = (d_y \circ s^{\rightarrow} \circ e_x)(\mathbf{x}), \qquad f^{\leftarrow}(\mathbf{y}) = (d_x \circ s^{\leftarrow} \circ e_y)(\mathbf{y}), \tag{3}$$

which aim to approximate the true forward and inverse mappings on the data manifold, i.e., $f^{\rightarrow} \approx \mathbf{F}^{\rightarrow}$ and $f^{\leftarrow} \approx \mathbf{F}^{\leftarrow}$. In this work, our primary interest lies in retrieval, namely the construction of an accurate surrogate inverse operator $f^{\leftarrow}$.

This separation between representation learning (autoencoders) and operator learning (latent mappings) is a distinguishing feature of latent twins, providing a transparent interpretation of how observations and states are related while yielding improved stability over direct surrogate inversion networks in ill-posed settings [12, 13, 21].

It is useful to contrast this construction with a direct inversion network that learns a single map $\mathbf{y} \mapsto \mathbf{x}$. Since neural networks are universal approximators, such a model can in

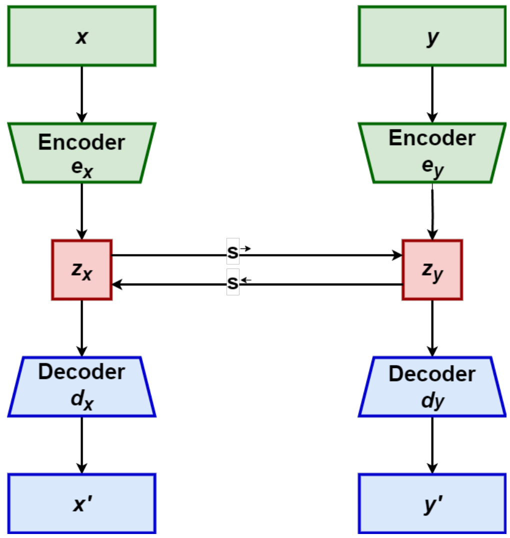

Figure 1. Schematic of the Latent Twin (paired autoencoder) architecture for inverse problems. Two autoencoders are trained on the parameter space $\mathcal{X}$ and the observation space $\mathcal{Y}$, producing low-dimensional latent representations $\mathbf{z}_x$ and $\mathbf{z}_y$, respectively. The inverse operator is learned directly in latent space through a trainable mapping $s^{\leftarrow} : \mathbf{z}_y \mapsto \mathbf{z}_x$, which is composed with the observation encoder and state decoder to form the full inverse surrogate/ latent twin $f^{\leftarrow} = d_x \circ s^{\leftarrow} \circ e_y$. By learning the inverse transformation in a shared latent coordinate system, the framework enables efficient and stable recovery of high-dimensional atmospheric states from observations, while preserving consistency between the data and model representations.

principle approximate the inverse map when the training data are sufficiently representative and the test samples remain in distribution. The purpose of the latent-twin formulation is therefore not to claim that a direct network cannot learn the retrieval map, but rather to impose additional structure that is valuable for ill-posed inverse problems. In a direct $\mathbf{y} \mapsto \mathbf{x}$ model, the inverse map is represented by one end-to-end transformation from radiances to atmospheric variables. By contrast, latent twins separate the task into interpretable components: e_y encodes the spectrum into a latent variable $\mathbf{z}_y$, $s^{\leftarrow} : \mathcal{Z}_y \to \mathcal{Z}_x$ represents the inverse radiative-transfer relation in latent space, and d_x decodes the atmospheric latent variable $\mathbf{z}_x$ back to physical variables. Thus, $\mathbf{z}_y$ and $\mathbf{z}_x$ represent compressed spectral and atmospheric states, while $s^{\leftarrow}$ is the surrogate inverse operator between them. Moreover, unlike a direct likelihood-free inversion network, the paired-autoencoder structure provides reconstruction and latent-consistency diagnostics that can help identify faulty or out-of-distribution observations through autoencoder reconstruction errors and latent mismatch measures [13, 14, 21].

An additional advantage of the latent twin construction is that it naturally induces a low-dimensional constraint on the inverse problem. In particular, $f^{\leftarrow}(\mathbf{y})$ may be interpreted not only as a direct retrieval operator, but also as defining an implicit prior or admissible manifold for atmospheric states. By restricting reconstructions to the form $d_x(\mathbf{z}_x)$ with $\mathbf{z}_x \in \mathcal{Z}_x$, the inversion is restricted to states representable by the learned atmospheric latent space, thus discouraging nonphysical solutions that lie off the data manifold. From an inverse-problem perspective, the latent twin model may therefore be used as a standalone surrogate inversion, as a rapid initializer for physics-based retrieval, or as a regularizing reference; for example, in Equation (2) one may incorporate the penalty

$$\mathcal{R}(\mathbf{x}) = \lambda \|\mathbf{x} - f^{\leftarrow}(\mathbf{y})\|_2^2, \qquad \lambda \geq 0. \tag{4}$$

To further assess the benefit of the proposed formulation, we also consider a direct spectrum-to-state baseline model learning the mapping $\mathbf{y} \mapsto \mathbf{x}$ without latent decomposition, whose performance is reported in the Supplementary Material. The direct model generally yields higher reconstruction errors, physical inconsistency and less stable performance compared to the proposed latent-twin approach, as detailed therein.

To support stable and accurate inversion, the training objective must jointly promote faithful reconstruction within each domain and consistent, well-aligned mappings between the corresponding latent spaces. The latent mapping may be trained in the forward direction, the inverse direction, or jointly in both directions, as adopted here. Reconstruction accuracy is explicitly assessed for both autoencoders by comparing decoded outputs with the corresponding original inputs. Results, reported in the Supplementary Material, show reconstruction errors significantly smaller than the intrinsic variability of the dataset across all considered variables, supporting the assumption of stable latent representations.

Latent Twin training. Given representative parameter–observation pairs $\{(\mathbf{x}_j, \mathbf{y}_j)\}_{j=1}^J$, the autoencoders are trained using reconstruction losses

$$\tfrac{1}{J}\sum_{j=1}^{J} \mathcal{J}_x\big((d_x \circ e_x)(\mathbf{x}_j), \mathbf{x}_j\big), \qquad \text{and} \qquad \tfrac{1}{J}\sum_{j=1}^{J} \mathcal{J}_y\big((d_y \circ e_y)(\mathbf{y}_j), \mathbf{y}_j\big), \tag{5}$$

where $\mathcal{J}_x$ and $\mathcal{J}_y$ denote suitable discrepancy measures on the parameter and observation spaces, respectively.

For the latent mappings, a variety of loss formulations may be employed, including losses defined directly in latent space or full surrogate losses composed with the encoders and decoders [13]. In this work, we adopt the latter and train the latent mappings using full surrogate objectives,

$$\tfrac{1}{J}\sum_{j=1}^{J} \mathcal{J}_{xy}\big(f^{\rightarrow}(\mathbf{x}_j), \mathbf{y}_j\big) \qquad \text{and} \qquad \tfrac{1}{J}\sum_{j=1}^{J} \mathcal{J}_{yx}\big(f^{\leftarrow}(\mathbf{y}_j), \mathbf{x}_j\big), \tag{6}$$

which enforce consistency between atmospheric states and observations through the learned forward and inverse surrogate operators. $\mathcal{J}_{yx}$ and $\mathcal{J}_{yx}$ are some other discrepancy measures on the observation and parameter spaces, respectively.

We optimize all components jointly by minimizing a weighted combination of reconstruction and surrogate losses,

$$\min_{\theta} \quad \frac{1}{J}\sum_{j=1}^{J}\Big(\alpha_x\,\mathcal{J}_x\big((d_x\circ e_x)(\mathbf{x}_j),\mathbf{x}_j\big)+\alpha_y\,\mathcal{J}_y\big((d_y\circ e_y)(\mathbf{y}_j),\mathbf{y}_j\big)$$
$$+\,\alpha_{xy}\,\mathcal{J}_{xy}\big(f^{\rightarrow}(\mathbf{x}_j),\mathbf{y}_j\big)+\alpha_{yx}\mathcal{J}_{yx}\big(f^{\leftarrow}(\mathbf{y}_j),\mathbf{x}_j\big)\Big), \tag{7}$$

where θ collects all trainable parameters of the encoders, decoders, and latent mappings, and $\alpha_x, \alpha_y, \alpha_{xy}, \alpha_{yx} > 0$ balance the individual loss terms. If only a forward or inverse surrogate is required, the corresponding mapping term may be omitted.

Empirically, we observe that joint optimization leads to better-aligned latent representations and simplifies the latent mappings, often allowing accurate surrogate operators to be represented by relatively low-complexity networks. In contrast to two-stage approaches that first train the autoencoders independently and then learn latent-space maps, joint training encourages the latent spaces to organize in a manner that is inherently compatible with the desired forward and/or inverse operator [12, 13, 21].

Model-consistent inversion with latent twins. Due to the lack of rigid physical constraints, latent-twin architecture can in some cases deviate from physical consistency, particularly for cloud-related variables, which are among the most challenging atmospheric quantities to retrieve and are prone to cross-variable inconsistencies. To address this while retaining a fully data-driven workflow, we augment the latent twin framework with targeted consistency correction models that operate directly in physical variable space. For each variable requiring correction, an auxiliary neural network is trained independently on the same dataset as the latent twin and applied selectively during inference only when inconsistencies are detected, thereby avoiding the cost of a full-physics-based inversion while preserving efficiency and interpretability.

Formally, we introduce corrective maps of the form $g : \mathcal{X} \to \mathcal{X}$, trained as

$$\min_{\theta_g} \quad \frac{1}{J}\sum_{j=1}^{J}\Bigg(\gamma_{\mathrm{inc}}\,\mathcal{J}_{\mathrm{corr}}\Big(\mathbf{m}_{\mathrm{inc}}\odot\Big(g\big(f^{\leftarrow}(\mathbf{y}_j)\big),\mathbf{x}_j\Big)\Big)+\gamma_{\mathrm{con}}\,\mathcal{J}_{\mathrm{corr}}\Big(\mathbf{m}_{\mathrm{con}}\odot\Big(g\big(f^{\leftarrow}(\mathbf{y}_j)\big),\mathbf{x}_j\Big)\Big)$$
$$+\,\mathcal{J}_{\mathrm{p}}\Big(g\big(f^{\leftarrow}(\mathbf{y}_j)\big)\Big)\Bigg), \tag{8}$$

where, θ_g collects all trainable parameters of the map g, $\mathcal{J}_{\mathrm{corr}}$ measures discrepancies between corrected predictions and reference states over complementary subsets of the state components. These subsets are defined through binary masks $\mathbf{m}_{\text{inc}}$ and $\mathbf{m}_{\text{con}}$, with $\mathbf{m}_{\text{inc}} \in \{0,1\}^n$ and $\mathbf{m}_{\text{con}} = \mathbf{1} - \mathbf{m}_{\text{inc}}$, identifying physically inconsistent and consistent elements, respectively. $\mathbf{m}_{\text{inc}}$ and $\mathbf{m}_{\text{con}}$ are computed from the first-stage model outputs via deterministic thresholding operations and are treated as fixed gating variables during backpropagation. Therefore, they are not trainable and do not participate in gradient computation. The functional $\mathcal{J}_{\text{p}}$ is a consistency metric measuring mismatches between paired state variables. The positive weights $\gamma_{\text{inc}}, \gamma_{\text{con}} > 0$ balance the contributions of the different loss terms.

3. Dataset generation

A dataset for the inversion problem is composed by a set of J pairs $\{\mathbf{x}_j, \mathbf{y}_j\}_{j=1}^{J}$. In our application, the vector $\mathbf{x}_j$ represents the atmospheric state, including atmospheric gases, surface properties such as Earth surface temperature and spectral emissivity, and cloud microphysical variables such as cloud mass content and cloud particle effective radius. The vector $\mathbf{y}_j$ denotes the simulated spectrum generated from $\mathbf{x}_j$ by applying the radiative transfer model, the convolution with the Instrument Spectral Response Function (ISRF), and the addition of noise according to the instrument variance-covariance matrix. This vector may also include ancillary information such as geolocation, acquisition time, and the atmospheric pressure of the layers. Details of the underlying physical formulation of the radiative transfer model are provided in Appendix A, along with relevant references, including [11, 37, 61].

Meteorological data for this study were sourced from the ERA5 reanalysis dataset provided by ECMWF [22]. Additional concentrations of atmospheric gases, including CO_2, CO, CH_4, NO_2, HNO_3, and SO_2, were retrieved from the Copernicus Atmosphere Monitoring Service (CAMS) global Greenhouse Gas reanalysis dataset [15]. The remaining gas data were drawn from the Initial Guess 2 (IG2) database [44]. Surface emissivity was obtained from the Huang dataset [23] using a tailored approach that matches emissivity spectra, expressed as a function of wavenumber, to specific times and locations [56], exploiting the Combined ASTER MODIS Emissivity over Land (CAMEL) database [8, 17]. The CAMEL data is available over land. On the sea, to provide some variability anyway to the emissivity profile, we applied the Masuda emissivity modification [39] due to wind, using the Cross-Calibrated Multi-Platform Wind Vector Analysis Product [67]. While the modification is minimal for nadir observations, it still prevents the generation of a large number of sample points with a constant emissivity profile, which could be easily detected by data-driven approaches.

In the left panel of Figure 2, we show the average January emissivity profile over water, with the shaded region indicating the corresponding standard deviation. For comparison, the right panel displays the emissivity profiles for points whose fields of view are dominated by deciduous land cover. In this case, the variability reflects the application of the method

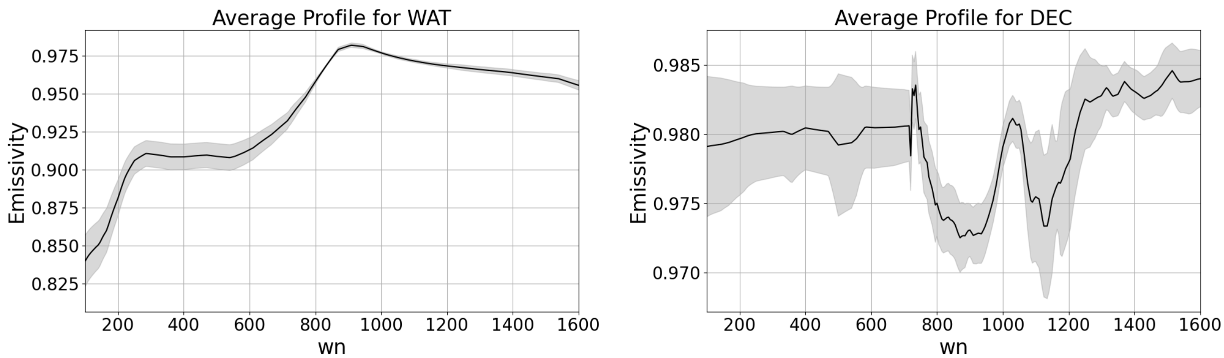

Figure 2. Average January emissivity profiles with standard deviation (shaded) for ocean (WAT) points (left panel) and land points dominated by deciduous (DEC) land cover (right panel).

proposed by [56].

For cloudy-sky scenes, the cloud ice and liquid water contents are taken from the ECMWF database. The average particle dimensions required for the multiple-scattering calculations are modeled using the Wyser model [69] for ice particles and the Martin model [34] for liquid water droplets.

We define a scene to be cloudy if the total optical depth due to clouds at 900 cm^{-1} exceeds 0.03, a threshold consistent with minimal cloud detectability, as supported by prior studies [51, 55].

We restrict our analysis to homogeneous fields of view. The ECMWF database provides *cloud fraction* at each atmospheric pressure level, corresponding to the fraction of the scene covered by clouds within a given grid cell. In this work, we treat each grid cell as homogeneous and characterize clouds directly using the cloud contents from the ECMWF dataset. While this assumption neglects sub-pixel cloud fraction, it preserves a realistic representation of the vertical cloud structure and associated optical properties.

Given the large volume of data required, a full-physics forward model such as the CLouds and Atmospheric Inversion Model (CLAIM), an advanced version of the inversion module of the FORUM End-to-End Simulator [54, 55, 66], is computationally prohibitive. In particular, cloudy-sky forward simulations using the accelerated DIScrete Ordinate Radiative Transfer (DISORT) solver [53] can require several hours per case, depending on the cloud's geometrical thickness and the characteristic particle size. Instead, we employ the σ-IASI/F2N code [3, 33, 35, 36], an extension of the σ-IASI framework that supports spectrometers operating across the infrared spectrum. The σ-IASI/F2N code parameterizes optical depths using low-order polynomials as functions of temperature and gas concentration. Under cloudy-sky conditions, multiple-scattering effects are incorporated via the Chou scaling method [11]. Finally, a tailored preprocessing code, hereafter referred to as ERA5 Code, assembles the surface and atmospheric data and interpolates them onto the grids required for the forward simulations.

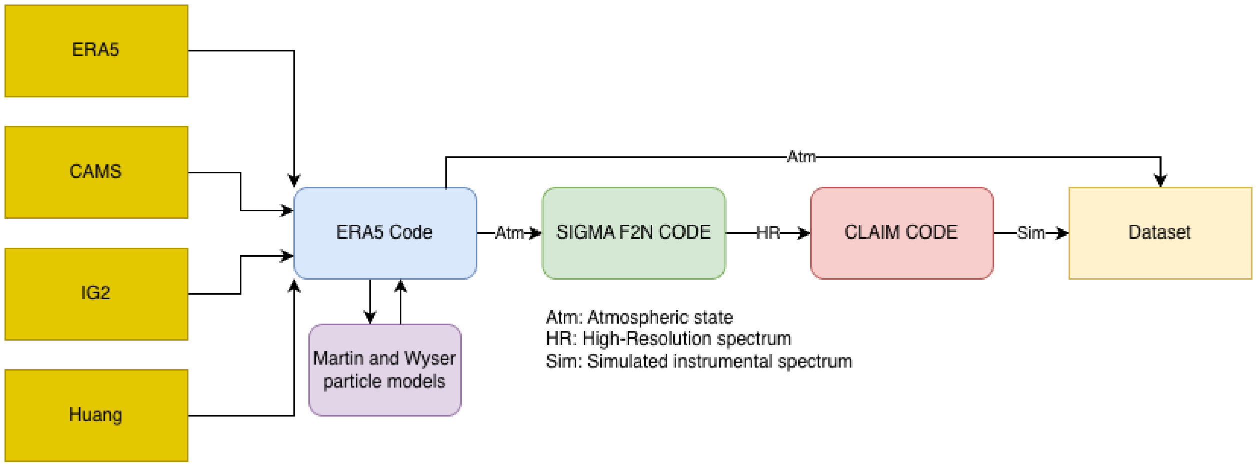

Figure 3. Workflow of the algorithm for the generation of the dataset. Atmospheric input data are retrieved from source databases (yellow, left) and processed through models and preprocessing codes (blue and purple) before being ingested by the forward model (green). The resulting high-resolution spectra are then convolved and corrupted with noise (red) to generate the final simulated spectral database (yellow, right).

The CLAIM code is used to perform the spectral convolution and to add instrument noise. For FORUM, we adopt a spectral resolution of 0.36 cm^{-1}, the Norton–Beer strong ISRF [42], and Noise-Equivalent Spectral Radiance (NESR) levels taken from one of the PhaseA instrument studies [55]. The overall procedure used to generate the dataset is summarized in Figure 3.

4. Data organization

We define an atmospheric state (or scenario) as a vector $\mathbf{x} \in \mathbb{R}^n$, with $n = 722$, collecting parameters that describe atmospheric gas constituents, surface properties, and cloud characteristics, and that fully determine the simulated radiance spectra. In the proposed framework, $\mathbf{x}$ serves as both the input and the reconstruction target of the first autoencoder. The subset of variables actively varied in this study is defined as

$$\mathbf{x} = \begin{bmatrix} T^0 & \mathbf{T} & \mathbf{w}_{\text{vap}} & \mathbf{o} & \varepsilon & \mathbf{c}_{\text{liq}} & \mathbf{r}_{\text{liq}} & \mathbf{c}_{\text{ice}} & \mathbf{r}_{\text{ice}} \end{bmatrix}^\top, \tag{9}$$

where each component corresponds to one of the key physical quantities listed in Table 1. All remaining parameters are fixed by the scene configuration and are not treated as retrieval variables in the neural network. Among the variables, Earth's surface temperature is a scalar, $T^0 \in \mathbb{R}$, whereas the surface spectral emissivity is a wavenumber-dependent vector, $\boldsymbol{\varepsilon} \in \mathbb{R}^{301}$, defined over the spectral interval from 100 to 1,600 cm^{-1}, and represented via linear interpolation on a regular grid with spacing 5 cm^{-1}. The remaining variables are vertical profiles defined on atmospheric pressure layers. While the vertical grid is fixed across the dataset, the lowest layer may be truncated to account for variations in surface altitude across geolocations. Consequently, the number of vertical layers may vary across samples, complicating the construction of a uniformly dimensioned input matrix. To address this issue, we adopt the approach of [50], which can be interpreted as a form of vertical stretching. Specifically, starting from the forward model pressure grid, each atmospheric column is first re-parameterized onto a normalized vertical coordinate $t \in [0, 1]$ (associating 0 to the top of the atmosphere level and 1 to the surface level), and then interpolated in 60 layers uniformly spaced in the same unit interval. This process is performed for all the atmospheric vertical profiles given to the the machine learning architecture. This ensures a consistent dimensionality across samples, such that $\mathbf{T}$, $\mathbf{w}_{\text{vap}}$, $\mathbf{o}$, $\mathbf{c}_{\text{liq}}$, $\mathbf{r}_{\text{liq}}$, $\mathbf{c}_{\text{ice}}$, and $\mathbf{r}_{\text{ice}} \in \mathbb{R}^{60}$. The water vapor and ozone profiles are represented in a transformed space, as detailed in Appendix B, where the motivation and precise formulation are provided. The cloud-related variables $\mathbf{c}_{\text{liq}}$, $\mathbf{r}_{\text{liq}}$,$\mathbf{c}_{\text{ice}}$, and $\mathbf{r}_{\text{ice}}$ are identically zero under clear-sky conditions or in atmospheric layers without clouds. For completeness, Table 1 also reports two additional quantities, namely the liquid and ice cloud optical depths, τ_{liq} and τ_{ice}, defined at 900 cm^{-1}.These are not included as inputs or outputs of the neural network, but are diagnostically derived from the retrieved profiles. Specifically, they are computed from the pairs $(\mathbf{c}_{\text{liq}}, \mathbf{r}_{\text{liq}})$ and $(\mathbf{c}_{\text{ice}}, \mathbf{r}_{\text{ice}})$ using standard formulations [70, 71]. For ice clouds, the optical depth is given by

$$\tau_{\text{ice}} = \frac{3}{2} \sum_i \frac{Q^i_{\text{ext}}\, c^i_{\text{ice}}\, \rho_{\text{air}}}{r^i_{\text{ice}}\, \rho_{\text{ice}}}\, \Delta z^i, \tag{10}$$

where c^i_{ice} and r^i_{ice} denote the cloud ice water content and cloud ice particle effective radius in the i-th atmospheric layer, respectively; ρ_{air} and ρ_{ice} are the air and ice densities; Δz^i is the geometric thickness of layer i; and Q^i_{ext} is the extinction efficiency [59, 64]. The latter depends on the cloud particle effective radius r^i_{ice}, the assumed cloud microphysical properties and the target wavenumber (here 900 cm^{-1}). An analogous expression is used to compute the liquid cloud optical depth τ_{liq}. The quantities τ_{liq} and τ_{ice} provide integrated measures of the optical thickness of liquid and ice clouds, respectively. Their sum, $\tau = \tau_{\text{liq}} + \tau_{\text{ice}}$, is used, for example, to support scene classification into clear- or cloudy-sky conditions.

The radiance-related variables are collected in the vector $\mathbf{y} \in \mathbb{R}^q$, with q=4233, and are detailed in Tables 2 and 3. The vector is defined as

$$\mathbf{y} = \begin{bmatrix} y_{\text{lon}} & y_{\text{lat}} & d_{\text{loc}} & h_{\text{loc}} & \mathbf{p} & \mathbf{s} \end{bmatrix}^\top. \tag{11}$$

This vector includes the geolocation of the observation, given by longitude ($y_{\text{lon}} \in \mathbb{R}$) and latitude ($y_{\text{lat}} \in \mathbb{R}$); the time of acquisition, expressed as month ($d_{\text{loc}} \in \mathbb{N}$) and local solar time ($h_{\text{loc}} \in \mathbb{R}$); the pressure of the layers $\mathbf{p} \in \mathbb{R}^{60}$; and the simulated FORUM noisy radiance spectrum $\mathbf{s} \in \mathbb{R}^{4,169}$. All measurements are assumed to be acquired at the same UTC time. Consequently, the local solar time, expressed in decimal hours, is fully determined by the longitude and can be computed as $h_{\text{loc}} = 12 + \frac{12}{180}\, y_{\text{lon}}$. In our model configuration, the pressure grid comprises 60 atmospheric layers. Since the surface pressure $p^0 \in \mathbb{R}$ is provided, we replace the lowest pressure layer p^1 with p^0 and interpolate all remaining variables accordingly. The spectra are provided across 4,169 spectral points within the wavenumbers range from 100 cm^{-1} to 1,600.48 cm^{-1} with steps of 0.36 cm^{-1} and are generated using the code architecture described in Section 3 giving as an input the corresponding atmospheric scenarios. In the proposed framework, $\mathbf{y}$ serves as both the input and the reconstruction target of the second autoencoder. All its components are processed jointly within the architecture; however, since the radiance spectrum $\mathbf{s}$ is the primary quantity of interest, results and discussion are reported in terms of $\mathbf{s}$ only. Accordingly, when referring to $\mathbf{y}$, we implicitly focus on its spectral component, unless otherwise stated. In Tables 1, 2, and 3, a detailed and comprehensive description of all the variables involved is provided, categorized into atmospheric scenario variables, measurement variables, and location description variables, respectively. For each variable, we report the notation, description, unit of measurement, and the range of values spanned by the entire dataset. The values of $\mathbf{p}, \mathbf{T}, \mathbf{w}_{\text{vap}}$ and $\mathbf{o}$, vary significantly across layers, so their ranges are shown in Figure 4.

Variable	Description	Unit	Range
T^0	Earth surface temperature	K	[206.55, 338.32]
T^i	air temperature at layer i	K	Figure 4
w^i_{vap}	water vapor concentration at layer i	g/kg	Figure 4
o^i	ozone concentration at layer i	ppv	Figure 4
ε^k	surface spectral emissivity at wavenumber k	1	[0.84, 1]
c^i_{liq}	cloud liquid water content in layer i	kg/kg	$[0,\ 9 \times 10^{-4}]$
c^i_{ice}	cloud ice water content in layer i	kg/kg	$[0,\ 5 \times 10^{-4}]$
r^i_{liq}	cloud liquid particle effective radius in layer i	μm	[0, 10.32]
r^i_{ice}	cloud ice particle effective radius in layer i	μm	[0, 249.8]
τ_{liq}	liquid cloud optical depth	1	≥ 0
τ_{ice}	ice cloud optical depth	1	≥ 0

Table 1. Variables associated with the atmospheric scenarios, presented with their notation, description, unit of measurement and corresponding value ranges.

For our analysis, we collect atmospheric scenario data from January and July 2021 at 12:00 UTC, covering the entire globe using a grid with 2° increments in both latitude and longitude for a total amount of $J_{\text{all}} = 31{,}862$ cases.

Variable	Description	Unit	Range
s^k	radiance at wavenumber k	$\mathrm{W/(m^2 \cdot sr \cdot cm^{-1})}$	≥ 0

Table 2. Variables associated with the measurements, presented with their notation, description and unit of measurement and corresponding value ranges.

Variable	Description	Unit	Range
p^0	Earth surface pressure	hPa	[519.98, 1038.52]
p^i	air pressure at layer i	hPa	Figure 4
y_{lon}	longitudinal position	degrees ($\circ$)	-178 to 178 in steps of 2
y_{lat}	latitudinal position	degrees ($\circ$)	-88 to 88 in steps of 2
d_{loc}	month	1	{1,7}
h_{loc}	local time	1	[0.1, 23.9].

Table 3. Variables related to the measurement locations, including their notation, description and unit of measurement and corresponding value ranges.

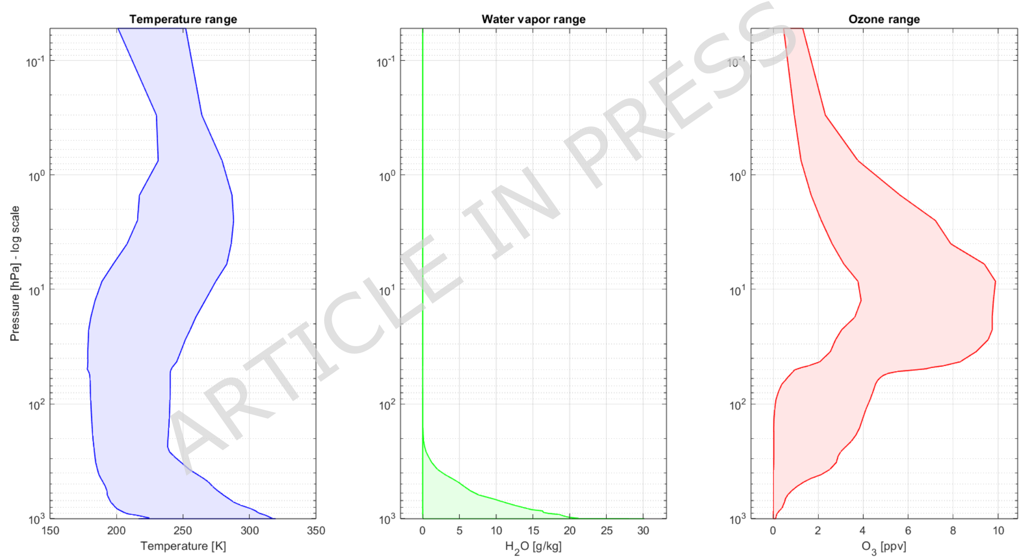

Figure 4. Variables associated with atmospheric scenarios, presented with their ranges across the layers. The lines indicate the minimum and maximum values in the database, while the shaded area shows the variability between them.

To train and evaluate the models, the dataset is partitioned into training and test sets. Of the total $J_{\mathrm{all}} = 31{,}862$ cases, $J_{\mathrm{train}} = 27{,}000$ are randomly assigned to the training set (approximately 85%), while the remaining $J_{\mathrm{test}} = 4{,}862$ cases (approximately 15%) form the test set. Prior to training, all features are normalized using min–max scaling to map their values to the interval $[0, 1]$. To avoid data leakage, the test set is normalized using the same minimum and maximum values computed from the training set, ensuring that the model is evaluated on previously unseen data. Some test samples may therefore fall slightly outside the interval $[0, 1]$ after normalization; however, this minor overflow is naturally handled by the network's activation functions.

5. Numerical results

5.1. Latent twin design

Network architecture. For simplicity, we employ standard multilayer perceptron (MLP) architectures for all neural network components of the latent twin. Both autoencoders employ fully connected symmetric encoder-decoder structures with ReLU (Rectified Linear Unit) activation functions and latent dimensions $n_{\text{latent}} = 512$ and $q_{\text{latent}} = 512$ for the state and observations spaces, respectively. The encoder architectures are $n \to (n + n_{\text{latent}})/2 = 617 \to 512$ for $\mathbf{x}$ and $q \to (q + q_{\text{latent}}) = 2372 \to 512$ for $\mathbf{y}$. In both cases, each decoder first maps the latent representation through a linear layer of the same dimensionality and then mirrors the encoder structure. An additional ReLU activation is applied after the final decoder layer to enforce non-negativity of the reconstructions.

Trainable linear mappings are introduced between the latent spaces of the two autoencoders to model the forward ($s^{\to} : \mathbf{z}_x \mapsto \mathbf{z}_y$) and inverse ($s^{\leftarrow} : \mathbf{z}_y \mapsto \mathbf{z}_x$) relationships. The choice of using linear latent mappings is consistent with related latent-coordinate approaches, including Koopman-based representations, where nonlinear mappings are represented through simpler transformations in learned coordinates [26, 30]. These mappings are implemented as single fully connected layers without activation functions. They enable the propagation of latent representations between the two domains, yielding surrogate latent representations $f^{\to}$ and $f^{\leftarrow}$ for both the forward and inverse processes.

Training of the latent twin. The latent twin architecture is trained jointly by minimizing a composite loss function consisting of four terms, as defined in Equation (7). Each term is formulated using the mean-squared error (MSE),

$$\mathcal{J}(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \mathbf{b}\|_2^2, \tag{12}$$

and measures either reconstruction fidelity or cross-domain consistency in latent space. Specifically, the loss comprises: (i) reconstruction losses for the atmospheric-state and radiance autoencoders; (ii) a forward surrogate loss comparing decoded radiances obtained by mapping atmospheric latent variables into the radiance latent space; and (iii) an inverse surrogate loss comparing decoded atmospheric states obtained by mapping radiance latent variables into the atmospheric latent space. The individual loss terms are combined linearly with weights

$$\alpha_x = \alpha_y = \alpha_{yx} = 1, \qquad \alpha_{xy} = 0.05, \tag{13}$$

reflecting our primary emphasis on accurate inversion from radiance observations to atmospheric states rather than on forward radiance prediction. This asymmetric weighting encourages faithful reconstruction of atmospheric variables while maintaining sufficient coupling between the two latent spaces to ensure stable and physically meaningful representations. The value $\alpha_{xy} = 0.05$ was selected empirically to keep the forward radiance-prediction term sufficiently small so that it does not dominate the optimization, while retaining a nonzero forward coupling that helps align the atmospheric and spectral latent spaces.

All models are implemented in PyTorch and trained using mini-batch optimization with a batch size of 512. Training is conducted over 10,000 epochs, where each epoch corresponds to a full pass through the training dataset. Network parameters are optimized using the Adam optimizer with an initial learning rate of $\ell_r = 10^{-3}$. A step-based learning rate scheduler is applied, reducing the learning rate by a factor of 0.75 every 500 training iterations. Training data are randomly shuffled at the beginning of each epoch to improve generalization and prevent the network from learning spurious correlations induced by sample ordering, while test data are processed without shuffling to ensure reproducibility. Training is carried out on a GPU-enabled Apple MacBook Pro with an M1 processor and 16 GB of memory with training times of about 12 hours.

Additional training Despite promoting a physically plausible latent manifold, residual inconsistencies may still arise, particularly for cloud-related variables.

In some retrievals, for example, the inferred cloud particle effective radius is non-zero in atmospheric layers where the corresponding cloud content is zero, or vice versa, which is physically inconsistent. Although this behavior occurs only in a limited subset of cases, it motivates the introduction of an additional correction stage, as anticipated in Section 2. This correction is intentionally focused on cloud-related inconsistencies, while other possible forms of physical inconsistency (e.g., dynamically unstable temperature profiles or other unrealistic atmospheric states) are not explicitly addressed in this work and remain open directions for future research. In the following, we describe the configuration of this post-processing stage, including the network architecture, loss formulation, and training procedure.

We use two corrective maps g (see Equation (8)), implemented as two small, independent multilayer perceptrons: one for liquid-phase variables (cloud liquid water content and cloud liquid particle effective radius) and one for ice-phase variables (cloud ice water content and cloud ice particle effective radius). Each network consists of three fully connected hidden layers with 4, 3, and 2 neurons, respectively, and employs Softplus activation functions throughout. The input to each network comprises five predictors: the preliminary retrievals of cloud content and cloud particle effective radius, the retrieved temperature and water vapor, and the true atmospheric pressure. These predictors are chosen because cloud content and particle size are primarily controlled by local thermodynamic conditions and humidity. The networks output corrected, physically consistent estimates of the cloud microphysical variables.

The corrective networks are trained independently, while all parameters of the main autoencoders and latent mapping networks are kept frozen. Training is performed on the full training set. During inference, the corrective models are applied only to atmospheric layers in which physical inconsistencies are detected after the first-stage retrieval, leaving already consistent layers unchanged. Thus, the correction stage should be interpreted as a targeted consistency-enforcement step for cloud variables rather than as a general retrieval-improvement step, since applying it to already consistent layers may unnecessarily perturb physically plausible "first-stage" estimates.

The training is performed by minimizing a composite loss function, as defined in Equation (8). The first two terms are based on the MSE between predicted and reference values,

$$\mathcal{J}_{\mathrm{corr}}(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \mathbf{b}\|_2^2, \tag{14}$$

measuring the reconstruction accuracy separately over atmospheric layers identified as physically inconsistent and consistent after the first-stage retrieval. These terms are weighted by coefficients γ_{inc} and γ_{con}, respectively, allowing differential emphasis on correcting inconsistent retrievals while preserving already consistent ones.

In addition, a consistency metric is defined as the number of inconsistencies between the zero/non-zero states of cloud content and cloud particle effective radius:

$$\mathcal{J}_{\mathrm{p}}(\mathbf{a}) = \left\| \mathbf{1}_{\{\mathbf{a}_{I_c}=0\}} - \mathbf{1}_{\{\mathbf{a}_{I_r}=0\}} \right\|_1, \tag{15}$$

where I_c and I_r denote subsets of indices corresponding to the variables describing cloud content and particle size, respectively, within the same atmospheric state vector.

Due to its dependence on binary indicator functions, $\mathcal{J}_{\mathrm{p}}$ is non-differentiable with respect to the model parameters and does not contribute to gradient-based optimization. It is therefore used exclusively as a diagnostic measure to monitor physically inconsistent configurations during training. Overall, only the weighted regression terms based on $\mathcal{J}_{\mathrm{corr}}$ drive the optimization of the corrective networks.

Training is carried out using mini-batches of size 512 and the Adam optimizer, with an initial learning rate of 10^{-4} and a weight decay of 10^{-5} for regularization. A learning rate scheduler reduces the learning rate by a factor of 0.5 when the validation loss does not improve for five consecutive epochs. The networks are trained for 500 epochs with $\gamma_{\mathrm{con}} = 1$ and $\gamma_{\mathrm{inc}} = 5$, thereby emphasizing the correction of physically inconsistent predictions. In the subsequent 500 epochs, the weighting scheme is inverted by swapping the coefficients, i.e., $\gamma_{\text{con}} = 5$ and $\gamma_{\text{inc}} = 1$, in order to reinforce accuracy on physically consistent layers while preserving overall physical coherence. At initialization, the training set contains 409,705 inconsistent layers, with 71,822 inconsistencies in the test set. After training, all such inconsistencies are fully resolved. The total loss decreases from 5.41 at initialization to 0.04 at convergence. Owing to the small size of the corrective networks, this post-processing stage is computationally inexpensive and can be efficiently executed on standard CPU hardware.

5.2. Retrieval errors on atmospheric variables

Given a test sample $\mathbf{y}_j$ with corresponding reference atmospheric state $\mathbf{x}_j$ and its prediction $\hat{\mathbf{x}}_j = f^{\leftarrow}(\mathbf{y}_j)$, we define the residual vector as

$$\boldsymbol{\xi}_j = f^{\leftarrow}(\mathbf{y}_j) - \mathbf{x}_j. \tag{16}$$

Over the full test set, we evaluate two error metrics: the *mean bias error* (**MBE**), which captures systematic over- or underestimation, and the *mean absolute error* (**MAE**), which quantifies the average magnitude of deviations irrespective of sign,

$$\mathbf{MBE} = \frac{1}{J_{\text{test}}} \sum_{j=1}^{J_{\text{test}}} \boldsymbol{\xi}_j, \qquad \mathbf{MAE} = \frac{1}{J_{\text{test}}} \sum_{j=1}^{J_{\text{test}}} |\boldsymbol{\xi}_j|. \tag{17}$$

For the results presented in this section, and to enable consistent aggregation across the entire test set, all atmospheric profiles are interpolated onto a common linear pressure grid consisting of 100 levels, spanning pressures from 1,054 hPa to 1 hPa. This interpolation is used exclusively for computing and visualizing the **MBE** and **MAE** profiles. Individual-case plots, however, retain the original 60-layer pressure grids, which vary between cases.

Figure 5 illustrates the retrieval results for selected atmospheric state variables: temperature (top-left), water vapor (top-right), ozone (bottom-left), and surface spectral emissivity (bottom-right). In each panel, the blue solid line denotes the **MBE**, while the blue dashed lines indicate the interval defined by **MBE** $\pm$ **MAE**. The red dashed lines represent the intrinsic variability of the true atmospheric state, quantified as $\pm$ **MAD**, where the *mean absolute deviation* (**MAD**) is defined as

$$\mathbf{MAD} = \frac{1}{J_{\text{test}}} \sum_{j=1}^{J_{\text{test}}} |\mathbf{x}_j - \bar{\mathbf{x}}|, \tag{18}$$

and $\bar{\mathbf{x}}$ denotes the mean atmospheric state computed over all test samples. Results for the surface temperature are reported separately in the figure caption.

Figure 6 presents the same analysis for the cloud-related parameters: cloud liquid water content (top-left), cloud ice water content (top-right), cloud liquid particle effective radius (bottom-left), and cloud ice particle effective radius (bottom-right). As before, solid lines

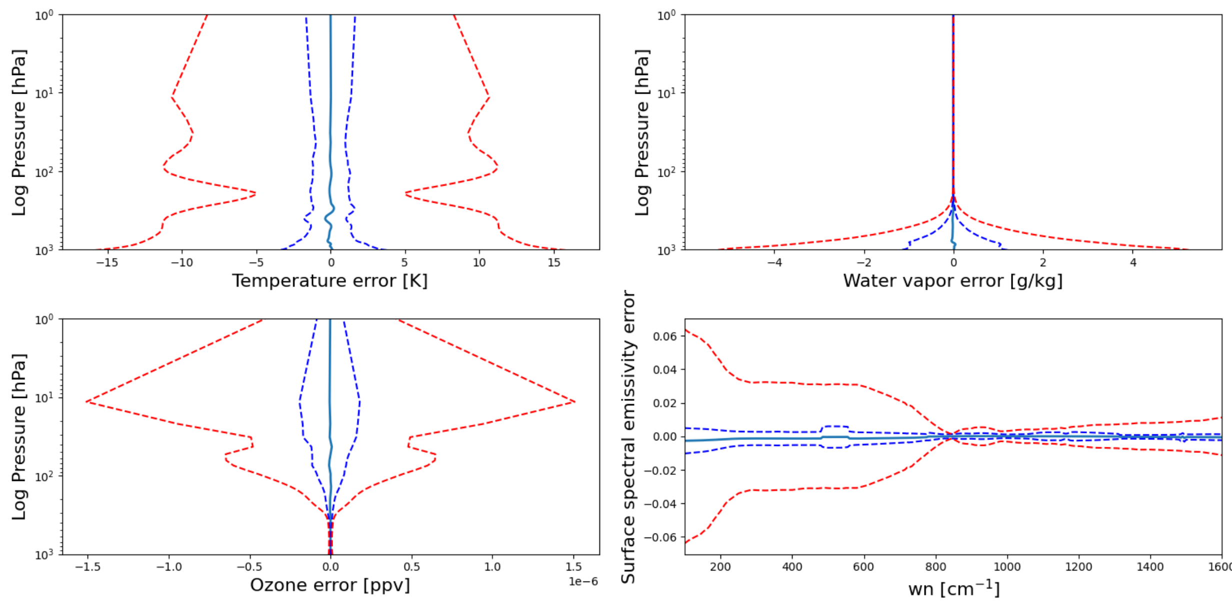

Figure 5. Retrieval results for atmospheric state variables: temperature (top-left), water vapor (top-right), ozone (bottom-left), and spectral emissivity (bottom-right). The blue solid line shows the **MBE**, while the blue and red dashed lines indicate **MBE** $\pm$ **MAE** and $\pm$ **MAD**, respectively. For surface temperature, the mean bias error is 0.035 K, the mean absolute error is 2.649 K, and the mean absolute deviation of the scenes is 17.413 K.

correspond to the **MBE**, while blue and red dashed lines indicate **MBE** $\pm$ **MAE** and $\pm$ **MAD**, respectively. This representation simultaneously conveys the retrieval bias and an envelope of the expected absolute error across the test set, relative to the intrinsic variability of the atmospheric state.

In addition, Figures 7, 8, 9, and 10 present the retrieval results for a randomly selected test case (case number 3,102). In Figures 7, 9, and 10, the first row shows the retrieved profiles (blue) alongside the corresponding reference profiles (orange), while the second row displays the associated residuals (black) with opposite sign with respect to Equation 16. Figure 7 reports, in order, temperature, water vapor, and ozone profiles, with the surface temperature indicated in the caption. Figures 9 and 10 show cloud liquid and ice water content, and cloud liquid and ice particle effective radius, respectively. Finally, Figure 8 presents the retrieved spectral emissivity (dashed blue line) together with the reference profile (orange).

Figures 6, 9 and 10 describe the final retrieval results. For completeness, the corresponding figures showing the results before the cloud model correction are reported in the Supplementary Material.

Once trained, the latent twin plus correction is able to invert the spectra for the entire

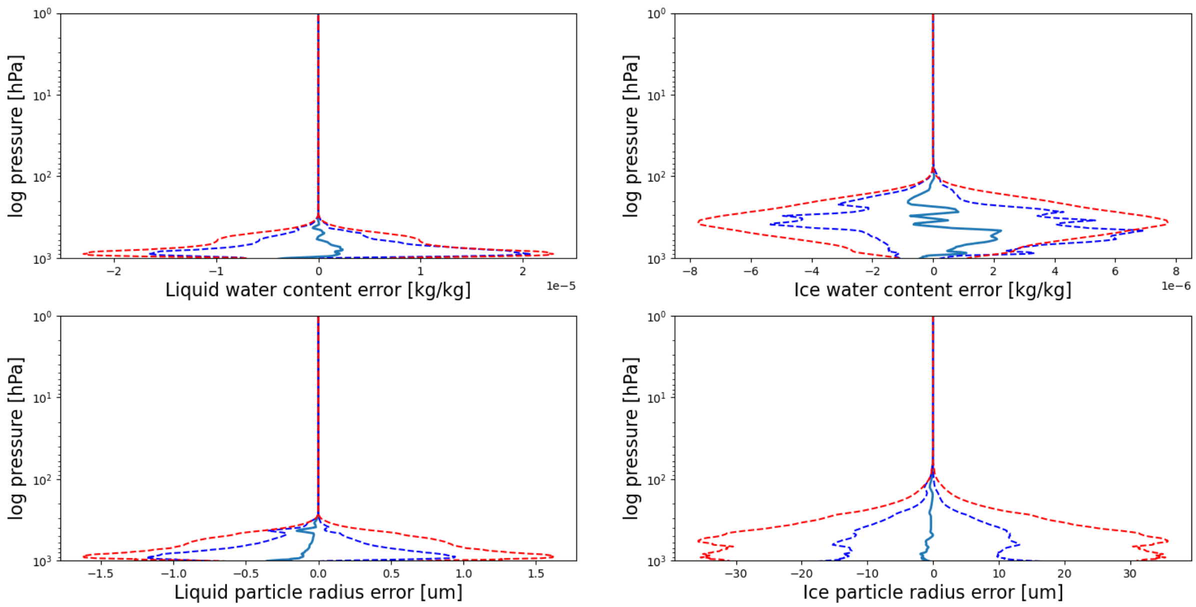

Figure 6. Retrieval results for cloud-related parameters: cloud liquid water content (top-left), cloud ice water content (top-right), cloud liquid particle effective radius (bottom-left), and cloud ice particle effective radius (bottom-right). The solid lines represent the **MBE**, while the dashed blue and red lines indicate $\mathbf{MBE} \pm \mathbf{MAE}$ and $\pm\mathbf{MAD}$, respectively.

test set in under 2.18 seconds on a standard laptop, corresponding to an average of 0.22 seconds per batch. This remarkable speed highlights the suitability of the method for large-scale applications and near-real-time processing scenarios.

5.3. Scene Classification Performance

As an additional validation of the retrieval performance, we analyze the reference total cloud optical depth τ, computed at 900 cm^{-1}. This quantity, introduced in Section 3 and defined in Section 4, serves as an effective discriminator between clear- and cloudy-sky scenes. In this study, as already mentioned, a threshold value of $\tau = 0.03$ is adopted to classify scenes as clear ($\tau \leq 0.03$) or cloudy ($\tau > 0.03$). FORUM simulations [55] indicate that clouds with optical depths below this threshold are generally undetectable.

Using this criterion, we examine the composition of the training and test datasets. In the training set, 3,743 cases out of 27,000 (approximately 13.86%) are classified as clear-sky. Similarly, in the test set, 685 cases out of 4,862 (approximately 14.09%) are clear-sky, indicating a consistent proportion between the two datasets. The fraction of clear-sky scenes is lower than the global average of approximately 33% [25]. This discrepancy arises because, as discussed previously, cloud fraction information is not explicitly used in our framework,

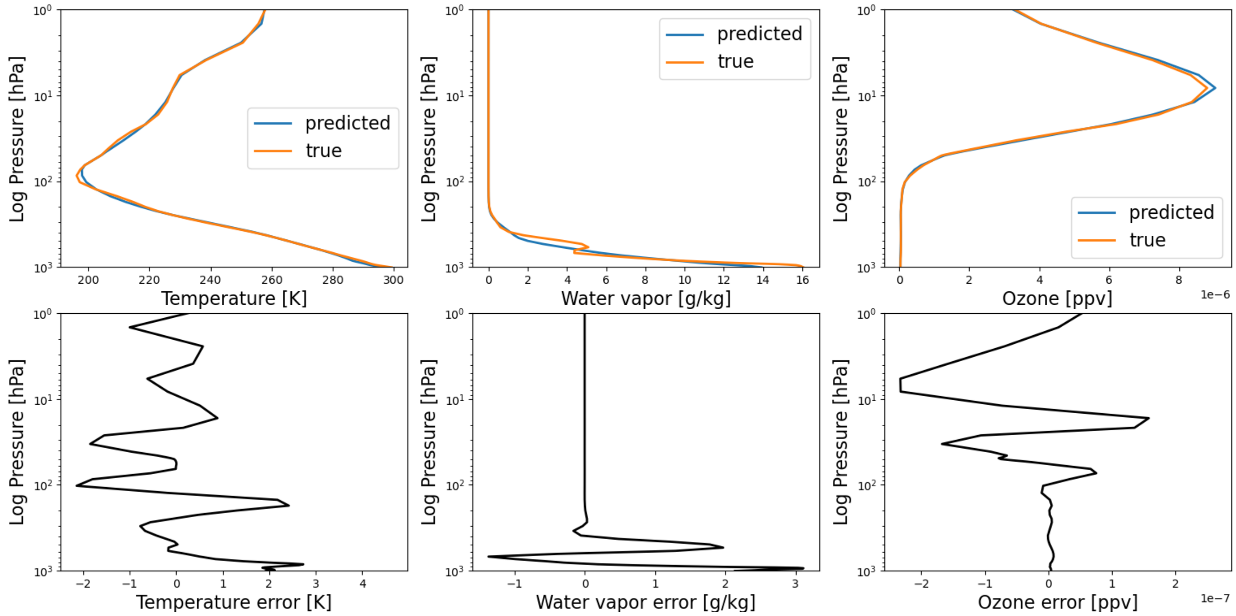

Figure 7. Retrieved atmospheric profiles for a randomly selected test case (case number 3102). The first row shows the retrieved profiles (blue) compared to the reference profiles (orange), while the second row displays the corresponding residuals (black). For surface temperature, the predicted value is 299 K and the reference value is 301 K.

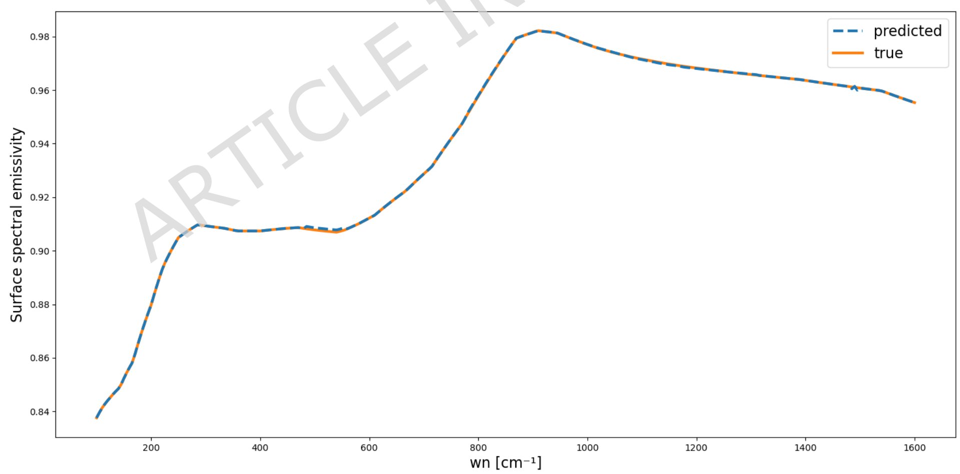

Figure 8. Retrieved spectral emissivity for the same randomly selected test case (case number 3102). The retrieved values are shown as a dashed blue line, while the reference values are shown in orange. The apparent small "arrow-like" feature visible on the dashed line on the top right is a visual artifact due to local numerical fluctuations and rendering effects.

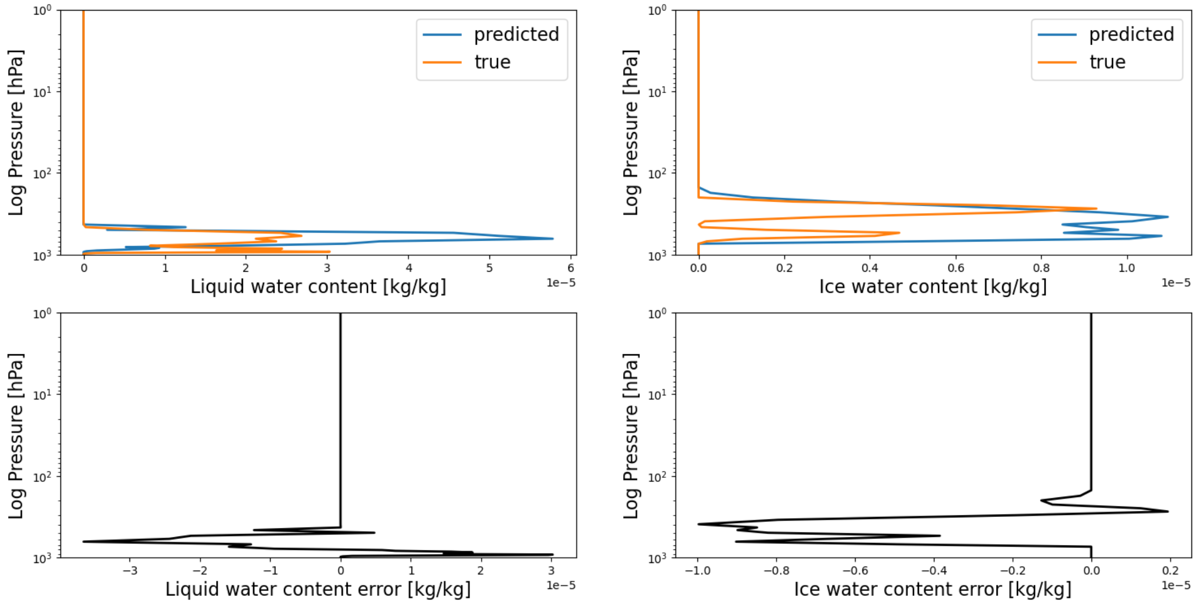

Figure 9. Retrieved vertical profiles for cloud contents for the same randomly selected case (case number 3,102) after the additional model correction. The top row shows the true profiles (orange) and the retrieved profiles (blue) for cloud specific liquid and ice water contents, while the bottom row displays the corresponding residual errors (black).

and each ERA5 tile is instead characterized by its cloudiest atmospheric column.

Focusing on the test set and excluding clear-sky cases, the distribution of cloudy scenes is shown in the histogram in Figure 11.

We next compare these reference values with the retrieved ones. The retrieved total cloud optical depth, denoted by τ^{pred}, is not a direct output of the retrieval architecture. Instead, it is analytically derived a posteriori from the retrieved cloud variables using the formulation in (10), by summing the contributions from the liquid and ice components, as described in Section 4.

Figure 12 shows the distribution of the residuals $\tau^{\text{pred}} - \tau$. The distribution is represented using a kernel density estimate, which provides a smooth approximation of the residual frequency. A vertical dashed line at zero indicates the ideal unbiased case, enabling a visual assessment of potential systematic deviations. This analysis provides complementary insight into the reliability and consistency of the retrievals across a wide range of total cloud optical depths.

Using the same metrics introduced above, we obtain on the entire test set an MBE $= 3.10$ and an MAE $= 17.17$. The scatterplot in Figure 13 illustrates the overall agreement between retrieved and reference values across the test set. Total cloud optical depth values are shown

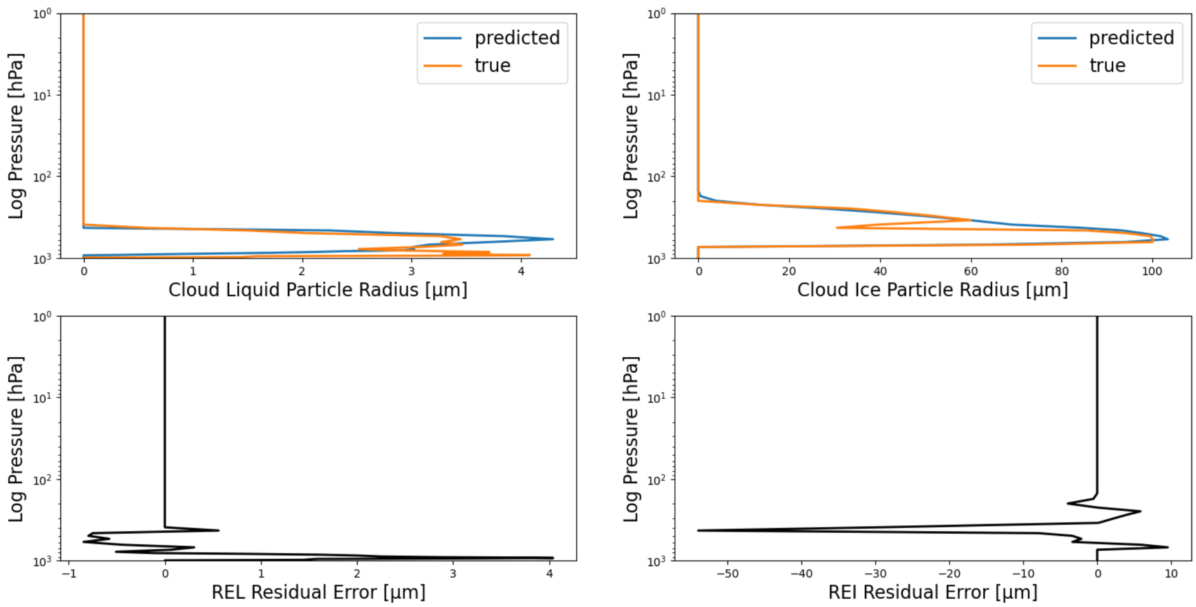

Figure 10. Retrieved vertical profiles for cloud particle effective radii for the same randomly selected case (case number3,102) after the additional model correction. The top row shows the true profiles (orange) and the retrieved profiles (blue) for cloud liquid and ice particle effective radii, while the bottom row displays the corresponding residual errors (black).

on a logarithmic scale to better visualize the full dynamic range. For visualization purposes, cases with zero optical depth (either retrieved, reference, or both) were excluded from the scatterplot. Colors highlight regions where points are more densely clustered.

As shown in Figure 12, the retrieval of the total cloud optical depth is not exact. In particular, the performance degrades for optically thin clouds. As the optical depth decreases, the radiance perturbation induced by the cloud becomes progressively less distinguishable from instrument noise, leading to a loss of sensitivity in the retrieval architecture. Consequently, the correlation between retrieved and reference optical depth weakens for thin clouds, with noticeable degradation already appearing for $\tau < 1$. This behavior is physically expected and also affects classical Optimal Estimation retrieval schemes. In the present machine-learning approach, the effect is amplified by the absence of an explicit physics-based constraint linking the atmospheric state to the observed radiance. As a result, the latent twin architecture is often able to identify the presence of a thin cloud, while still struggling to accurately estimate its exact optical depth. Conversely, for optically thick clouds ($\tau \gtrsim 10$), further increases in optical depth have only a marginal impact on the observed spectrum, since such clouds effectively block the radiance originating from below the cloud top. As a result, pointwise errors in the retrieved optical depth are not

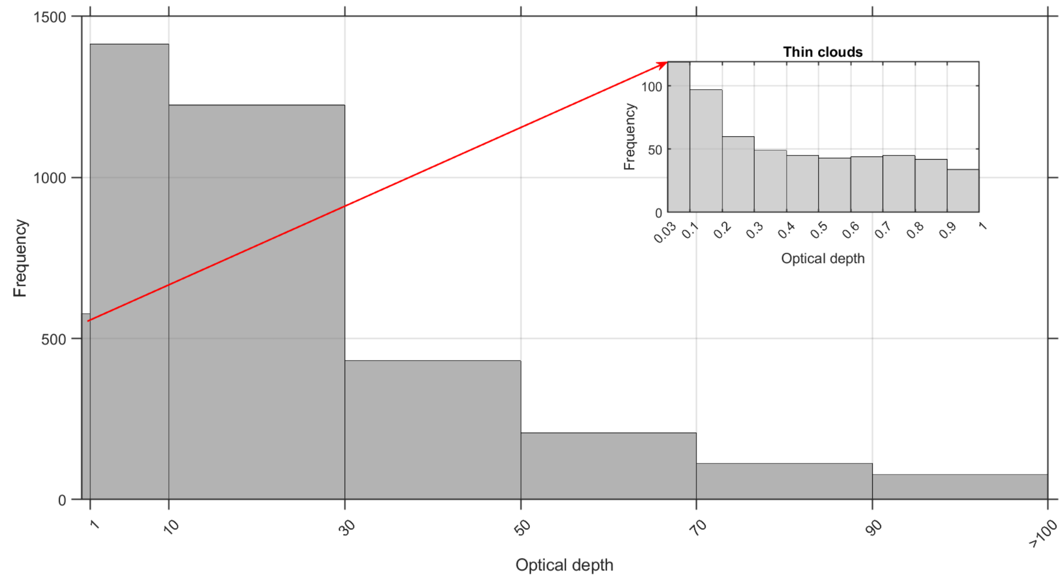

Figure 11. Distribution of cloudy cases in the test set as a function of the total cloud optical depth. The inset highlights the low–optical-depth range from 0.03 (the threshold used to define clear-sky conditions) to 1, providing a detailed view of the thin-cloud regime in the test set.

unexpected in these regimes, and the residual distribution should be interpreted with this inherent saturation behavior in mind.

Despite these limitations, the Spearman rank correlation coefficient computed over the test set reaches 0.84, indicating that the retrieval reliably captures the monotonic relationship between predicted and reference optical depth values, even when absolute deviations are present.

When used to discriminate between clear and cloudy scenes, the proposed architecture achieves an overall accuracy of 97.37% on the test set. The corresponding confusion matrix is shown in Figure 14, where the dominant diagonal entries indicate the high fraction of correctly classified cases. In particular, 98.2% of clear-sky scenes and 97.2% of cloudy scenes are correctly identified.

To the best of our knowledge, there are no directly comparable studies on scene classification based on FORUM-simulated measurements using deep learning approaches. While the cloud detection algorithm embedded in the FORUM End-to-End simulator [31, 55] showed promising results, it was evaluated on a relatively limited test set and does not rely on a deep learning framework. Another important infrared interferometer is the Infrared Atmospheric Sounding Interferometer (IASI) [57]. In this context, Whitburn et al. [68] reported an agreement of 87% between a supervised neural network and the operational

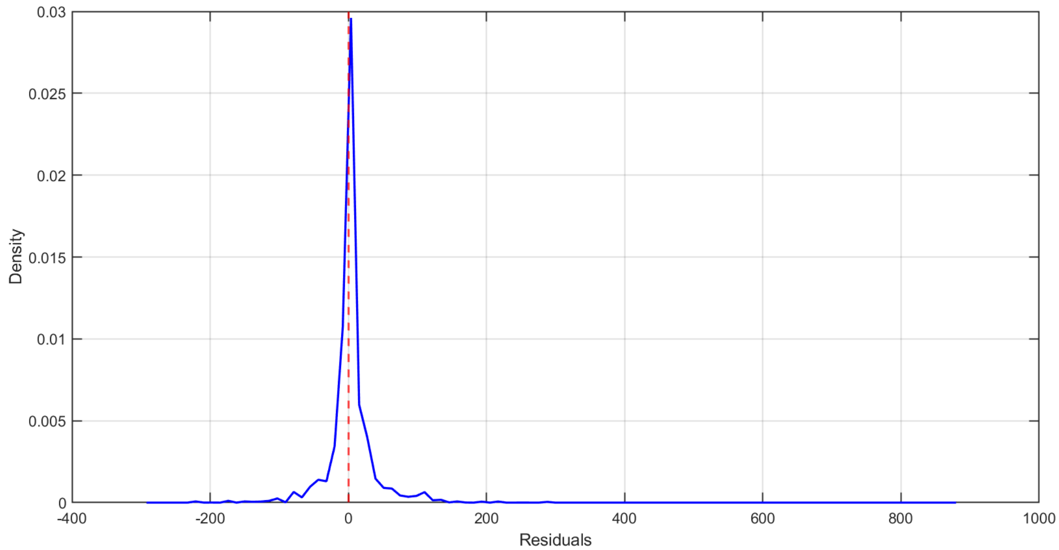

Figure 12. Distribution of residuals $\tau^{\text{pred}} - \tau$ in the test set, shown through a kernel density estimate. The vertical dashed red line at zero represents the ideal case of no bias.

IASI Level 2 cloud product using a subset of 45 spectral channels. Higher agreement rates have been achieved for IASI measurements by restricting the analysis to specific geographical regions; for example, Mastro et al. [38] reported a 93% agreement over Eastern Europe and tropical regions. In comparison, the results presented here demonstrate that the proposed latent twin framework can achieve competitive, and in some cases superior, classification performance using FORUM-like broadband spectral information over a global dataset.

Furthermore, if we refine the classification into three categories, clear-sky ($\tau \leq 0.03$), thin clouds ($0.03 < \tau \leq 1$), and thick clouds ($\tau > 1$), the percentage of correctly recognized cases in the test set is 87.70%.

Finally, to assess whether the reconstructed profiles preserve the vertical structure of the clouds, we perform a Fourier sine decomposition of both the true profiles $\mathbf{c}_{\text{liq}}$, $\mathbf{c}_{\text{ice}}$, $\mathbf{r}_{\text{liq}}$, $\mathbf{r}_{\text{ice}}$ and the corresponding predicted ones, denoted with an asterisk. The goal is not to analyze the content per se, but to verify whether the model captures the same spatial vertical variability patterns as the reference data. Given a profile $\mathbf{c}_{\text{liq}}$, defined in N layers, the coefficients of the sine series, representing the contribution of each harmonic to the profile structure, are computed as

$$b_k \approx \frac{2}{p^N - p^1} \sum_{i=1}^{N} \omega^i c_{\text{liq}}^i \sin\left(\frac{k\pi(p^i - p^1)}{p^N - p^1}\right), \tag{19}$$

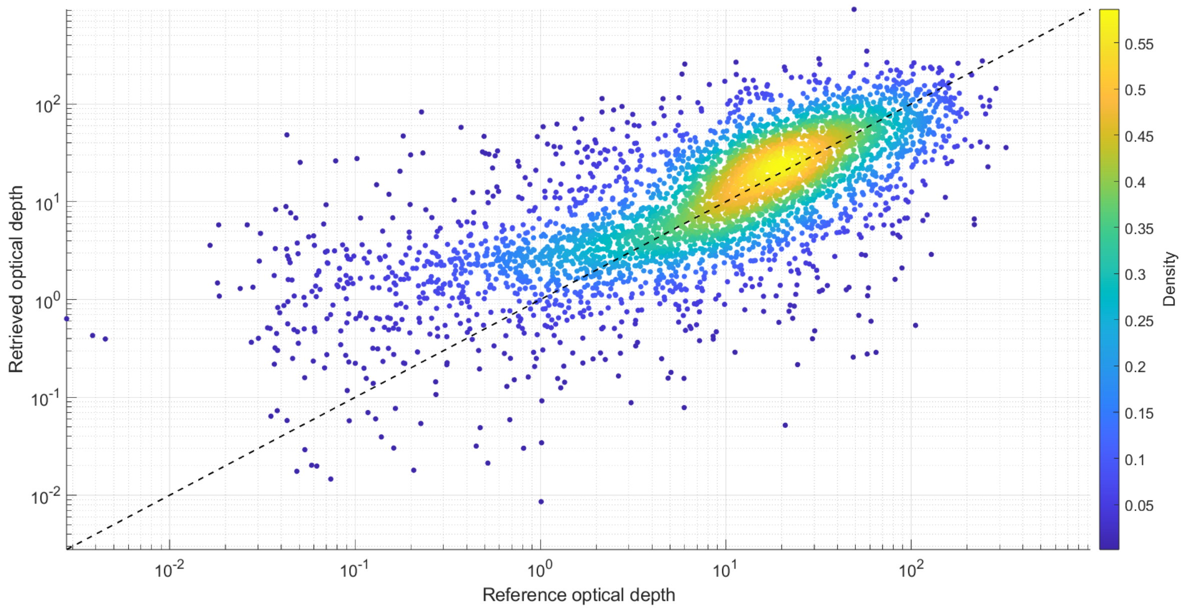

Figure 13. Scatterplot of retrieved versus reference total cloud optical depth for the entire test set. Each dot represents a single scene, while the dashed line corresponds to the ideal one-to-one relationship. Values are shown on a logarithmic scale and the color indicates the density of points, as represented by the colorbar on the right.

where $k = 1, \ldots, K$ is the harmonic index, ω^i are trapezoidal integration weights

$$\omega^i = \begin{cases} \frac{1}{2}(p^2 - p^1), & \text{for } i = 1, \\ \frac{1}{2}(p^{i+1} - p^{i-1}), & \text{for } i = 2, \ldots, N-1, \\ \frac{1}{2}(p^N - p^{N-1}), & \text{for } i = N. \end{cases} \tag{20}$$

To remove the amplitude dependence and highlight the similarity in the vertical structure of the profiles, the Fourier coefficients are normalized by their maximum absolute value:

$$\mathbf{b} = [b_1, \ldots, b_K], \qquad \tilde{\mathbf{b}} = \frac{\mathbf{b}}{\max_i |b_i|}. \tag{21}$$

Similarly, we compute the Fourier coefficients $\mathbf{b}^*$ for the predicted variable $\mathbf{c}^*_{\text{liq}}$.

Finally, for each test case, we compute the cosine similarity between the true and reconstructed coefficient vectors $\mathbf{b}$ and $\mathbf{b}^*$ as:

$$\cos_{\text{sim}} \beta = \frac{(\tilde{\mathbf{b}}^*)^\top \cdot \tilde{\mathbf{b}}}{\|\tilde{\mathbf{b}}^*\| \cdot \|\tilde{\mathbf{b}}\|} \tag{22}$$

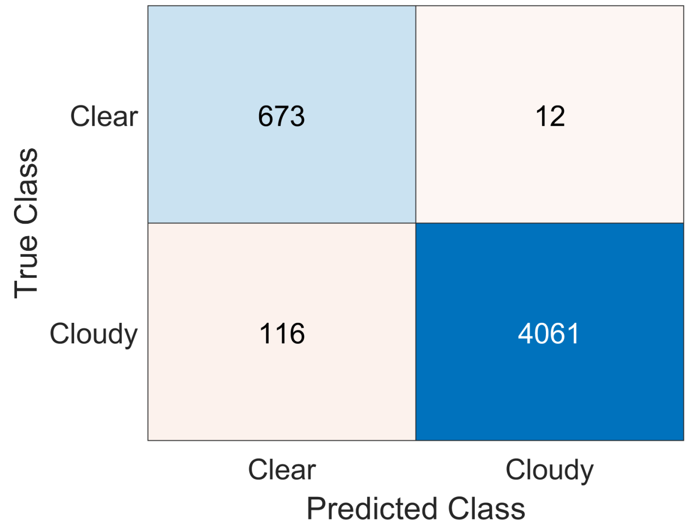

Figure 14. Confusion matrix for the classification of clear and cloudy scenes in the test set. Rows correspond to the true classes and columns to the predicted classes. Overall, the classifier correctly identifies 97.37% of the cases. Class-wise accuracies are 98.2% for clear-sky scenes and 97.2% for cloudy scenes.

This metric quantifies the degree of alignment between the spectral representations of the true and predicted profiles, with values close to 0 indicating no similarity and values close to 1 denoting an identical structural shape. The same analysis is carried out for the remaining three variables.

Across the test dataset, the cosine similarity averages around 0.75 for water-cloud variables and 0.67 for ice-cloud variables, indicating that the model correctly reproduces the vertical position and general structure of cloud layers, even when the amplitude of the retrieved profiles is not perfectly reconstructed.

5.4. Radiance Reconstruction

To assess radiance consistency, we apply the radiative transfer to the reconstructed profiles in order to evaluate the resulting radiance error. We use the reduced χ_r^2 statistics, defined as:

$$\chi_r^2 = \frac{1}{q}\big(\mathbf{y} - \mathbf{F}^{\rightarrow}(\mathbf{x})\big)^{\top}\mathbf{S}_{\mathbf{y}}^{-1}\big(\mathbf{y} - \mathbf{F}^{\rightarrow}(\mathbf{x})\big) \tag{23}$$

where q denotes the number of degrees of freedom of the system and $\mathbf{S}_{\mathbf{y}} \in \mathbb{R}^{q\times q}$ is the covariance matrix of the measurements. As explained in Section 3, this symmetric positive-definite matrix depends on the instrumental NESR and the apodization kernel used, and can be computed explicitly. The expected value of χ_r^2 for a spectrum that is compatible with the instrumental noise is 1. For all points in the dataset, applying the forward model to the true atmospheric state produces spectra with $\chi_r^2 \equiv 1$ when evaluated against the noisy spectra.

In order to improve the agreement between the model and reconstructed spectra for cases classified as clear, we applied a simple post-processing step. Specifically, we set the reconstructed surface temperature T^0 as the solution of the equation $s^{900} = \epsilon^{900} B(T^0)$, where B is the Planck function, s^{900} is the radiance at 900 cm^{-1}, and ϵ^{900} is the reconstructed emissivity at the same wavenumber. This approach assumes negligible atmospheric absorption at 900 cm^{-1}, which lies within the atmospheric window. This simple procedure significantly improves the agreement of the reconstructed spectra.

First of all, in our results we observe that the χ_r^2 error is strongly influenced by the stretching procedure described in Section 4. To illustrate this effect, Figure 15 shows the χ_r^2 obtained by applying only the stretching and de-stretching procedure to the true dataset, demonstrating that this operation alone already introduces a deviation from the expected value discussed above.

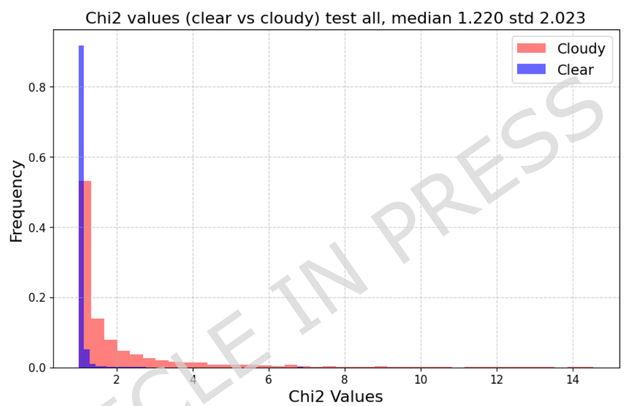

Figure 15. χ_r^2 obtained by applying the stretching and de-stretching procedure to the true atmospheric states over the entire test set. Clear-sky cases are shown in blue, while cloudy cases are shown in red. The median value and the corresponding standard deviation are reported in the figure title.

In Figure 16 we show the results of the reconstruction, which therefore also include the error introduced by the stretching procedure discussed above. In the top row, we present the results for the cases recognized as clear by the system: the left panel shows the reconstruction without the postprocessor, and the right panel shows the reconstruction with the postprocessor. In the bottom row, the left panel displays the results for the cases recognized as cloudy, while the right panel shows the combined histogram for all cases in the test set.

It is evident that for most cases the χ_r^2 is not satisfactory for a real retrieval, still the values obtained are not a bad estimate to be used as an initial guess or an a priori for a refinement retrieval. For comparison, in Figure 17 we show the combined χ_r^2 distribution

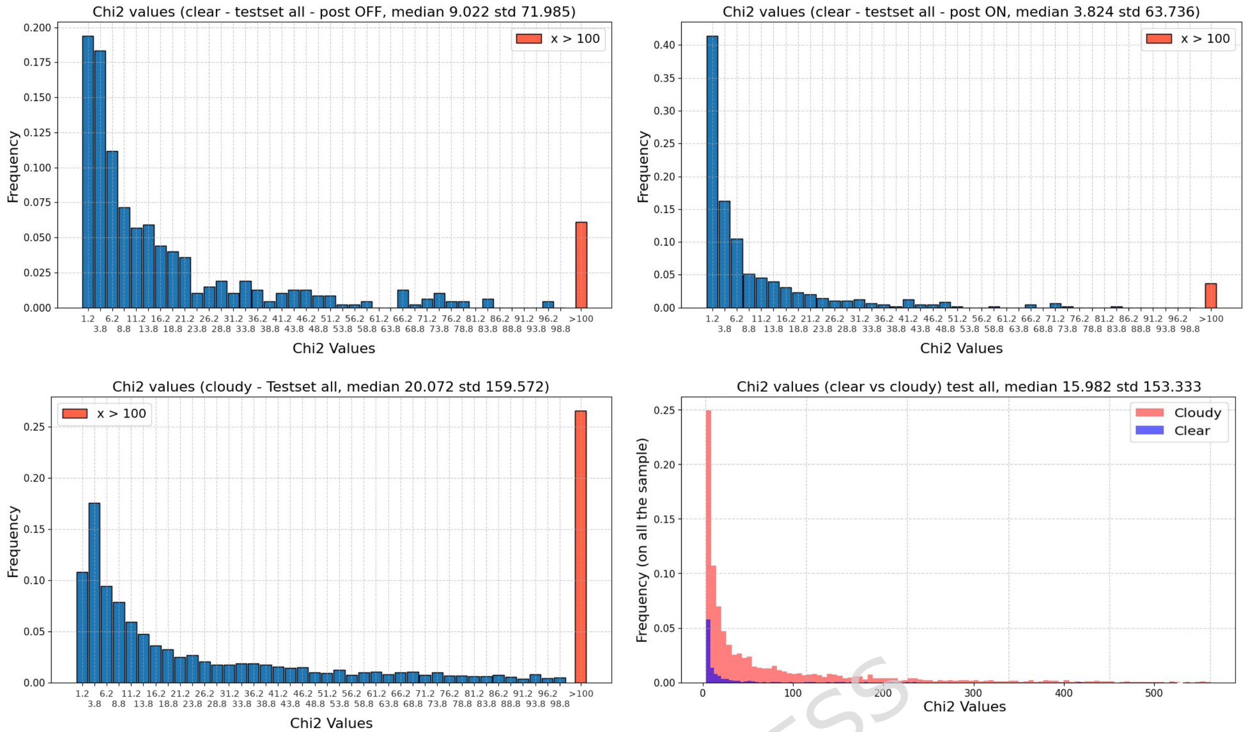

Figure 16. χ_r^2 obtained after reconstruction. The top row displays clear-sky cases, without (left) and with (right) the postprocessor, while the bottom-left panel shows cloudy cases only, and the bottom-right panel shows the combined histogram for all test cases. All quantitative information, including the median value and its standard deviation, is reported in the figure title. In the bottom-right panel, clear-sky cases are shown in blue and cloudy cases in red. In the other panels, red only highlights higher χ_r^2 values.

obtained perturbing the atmospheric state according to atmospheric variability, obtained from the IG2 estimates for clear sky variables, and determined from the dataset itself for the cloud variables. Overall, these results show that our reconstruction provides a significantly better starting point than simply using the climatological a priori.

6. Conclusion

In this work, we introduced a data-driven, physics-correcting retrieval method for FORUM all-sky observations based on the *latent twins* framework. By coupling paired autoencoders for atmospheric states and radiance spectra through trainable latent-space mappings, the proposed approach learns a surrogate inverse operator that is both computationally efficient and constrained to a physically plausible manifold. This design has the potential to enable fast, stable retrievals for a highly ill-posed and high-dimensional inverse problem, while retaining a clear interpretation in terms of forward and inverse radiative transfer surrogates.

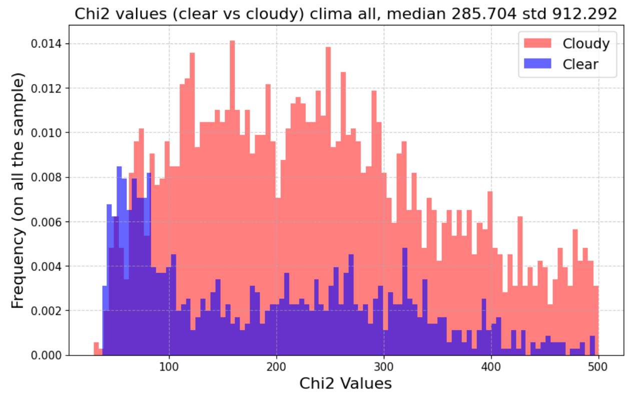

Figure 17. χ_r^2 obtained using climatological a priori for the full test set. All quantitative information, including the median and its standard deviation, is reported in the figure title. Clear-sky cases are shown in blue and cloudy cases in red.

The proposed framework shows promise along two main directions. First, it demonstrates strong performance in distinguishing clear-sky from cloudy conditions, achieving an overall classification accuracy of 97.37% over a global dataset. The results are consistent with, and complementary to, previous studies based on FORUM-simulated measurements, such as Maestri et al. [31], one of the few available references using FORUM simulations. Second, the framework provides a viable basis for the retrieval of atmospheric properties under all-sky conditions. In particular, it enables the retrieval of temperature, water vapor, ozone, and surface emissivity profiles with minimal bias relative to their natural variability. Retrieval of cloud-related variables remains inherently more challenging; however, the method captures key cloud features, including the vertical location of cloud layers and their phase, and provides informative estimates of cloud ice and liquid water content as well as cloud particle effective radius. These results indicate that the framework offers a promising basis for future improvements in cloud property retrieval. Moreover, the model-consistency correction introduced for cloud variables resolves physically inconsistent states, such as non-zero cloud particle effective radii in cloud-free layers.

A key strength of the proposed approach is its computational efficiency. Once trained, the latent twin model performs retrievals orders of magnitude faster than full-physics variational inversion methods, enabling near-real-time inference on standard hardware. Radiance reconstruction experiments further show that the retrieved atmospheric states provide substantially improved spectral consistency compared with climatological priors, making them well suited as initial guesses, adaptive priors, or proposal states for downstream physics-based refinement and data assimilation workflows.

Several important extensions remain for future work. These include training on more diverse seasonal and diurnal conditions, incorporating additional trace gases, and integrating the latent twin retrieval into hybrid inversion schemes that combine learned surrogates with iterative physics-based solvers. Extending the framework to explicitly quantify uncertainty, for example through Bayesian latent twins or ensemble-based formulations, is another promising direction. Further investigation of broader forms of physical inconsistency beyond the specific cloud-variable mismatches currently considered may also be valuable. A more comprehensive analysis of latent-space interpretability remains an additional important direction, since the present results primarily demonstrate a structured latent-space inversion framework. Finally, application to real observations from current infrared sounders and, ultimately, to FORUM flight data will be essential to fully assess robustness and generalization.

Taken together, these results demonstrate that the proposed methodology provides a computationally efficient, physically consistent, and flexible framework that lays the foundation for the inversion of complex remote-sensing observations, while enabling the exploitation of large and heterogeneous datasets and ensuring robust generalization across diverse cloudy conditions. By combining modern machine learning with radiative transfer physics, the approach opens new perspectives for operational retrievals, large-scale data exploitation, and future assimilation of far-infrared observations into weather and climate models.

7. Acknowledgments/Fundings

INdAM-GNCS supported the first author under Bando di concorso a n.45 mensilità di Borse di studio per l'estero A.A. 2023-2024. 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, CUP: C53C22000870006). ASI (Italian Space Agency) supported the second author under the project CASIA (CAirt and Sinergy with IAsi-ng), CUP: F93C23000430001.

8. Data availability

The training and test datasets used in this study are available for public access at `https://doi.org/10.5281/zenodo.17969539` [52].

Bibliography

[1] J. Adler and O. Öktem. "Solving ill-posed inverse problems using iterative deep neural networks". In: *Inverse Problems* 33.12 (2017), page 124007 (cited on page).

[2] B. M. Afkham, J. Chung, and M. Chung. “Learning regularization parameters of inverse problems via deep neural networks”. In: *Inverse Problems* 37.10 (2021), page 105017 (cited on page).

[3] U. Amato et al. “The sigma-IASI code for the calculation of infrared atmospheric radiance and its derivatives”. In: *Environmental Modelling and Software* 17.7 (2002), pages 651–667. ISSN: 1364-8152. DOI: https://doi.org/10.1016/S1364-8152(02)00027-0. URL: https://www.sciencedirect.com/science/article/pii/S1364815202000270 (cited on page).

[4] S. R. Arridge et al. “Solving inverse problems using data-driven models”. In: *Acta Numerica* 28 (2019), pages 1–174 (cited on page).

[5] P. Baldi and K. Hornik. “Neural networks and principal component analysis: Learning from examples without local minima”. In: *Neural Networks* 2.1 (1989), pages 53–58 (cited on page).

[6] M. Ben-Yami et al. “Emissivity retrievals with FORUM’s end-to-end simulator: challenges and recommendations”. In: *Atmospheric Measurement Techniques* 15.6 (2022), pages 1755–1777. DOI: 10.5194/amt-15-1755-2022. URL: https://amt.copernicus.org/articles/15/1755/2022/ (cited on page).

[7] A. Bora et al. “Compressed Sensing using Generative Models”. In: *Proceedings of the 34th International Conference on Machine Learning (ICML)*. 2017 (cited on page).

[8] E. E. Borbas et al. “The Combined ASTER MODIS Emissivity over Land (CAMEL) Part 1: Methodology and High Spectral Resolution Application”. In: *Remote Sensing* 10.4 (2018). ISSN: 2072-4292. DOI: 10.3390/rs10040643. URL: https://www.mdpi.com/2072-4292/10/4/643 (cited on page).

[9] M. Burton et al. “The FORUM mission: measuring Earth’s far-infrared spectrum”. In: *Bulletin of the American Meteorological Society* (2023) (cited on page).

[10] D. Calvetti and E. Somersalo. *An introduction to Bayesian scientific computing: ten lectures on subjective computing*. Volume 2. Springer Science & Business Media, 2007 (cited on page).

[11] M.-D. Chou et al. “Parameterization for cloud longwave scattering for use in atmospheric models”. In: *J. Climate* 12 (1999), pages 159–169 (cited on page).

[12] M. Chung, B. Peters, and M. Solomon. *Good Things Come in Pairs: Paired Autoencoders for Inverse Problems*. book chapter, arXiv preprint arXiv:2505.06549. 2025 (cited on page).

[13] M. Chung et al. “Paired autoencoders for likelihood-free estimation in inverse problems”. In: *Machine Learning: Science and Technology* 5.4 (2024), page 045055 (cited on page).

[14] M. Chung et al. "Latent Twins". In: *Machine Learning: Science and Technology* 7 (2026), page 025039. DOI: 10.1088/2632-2153/ae538e (cited on page).

[15] Copernicus Atmosphere Monitoring Service (CAMS). *CAMS Global Greenhouse Gas Reanalysis (EGG4)*. https://atmosphere.copernicus.eu/cams-global-greenhouse-gas-reanalysis. Accessed: 2024-09-23. 2020 (cited on page).

[16] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Volume 375. Springer Science & Business Media, 1996 (cited on page).

[17] M. Feltz et al. "The Combined ASTER MODIS Emissivity over Land (CAMEL) Part 2: Uncertainty and Validation". In: *Remote Sensing* 10.5 (2018). ISSN: 2072-4292. DOI: 10.3390/rs10050664. URL: https://www.mdpi.com/2072-4292/10/5/664 (cited on page).

[18] R. M. Goody and Y. L. Yung. *Atmospheric radiation : theoretical basis*. Oxford University Press, 1989 (cited on page).

[19] K. Gregor and Y. LeCun. "Learning Fast Approximations of Sparse Coding". In: *Proceedings of the 27th International Conference on Machine Learning (ICML)*. 2010 (cited on page).

[20] J. Harries et al. "The Far-infrared Earth". In: *Reviews of Geophysics* 46.4, RG4004 (Dec. 2008), RG4004. DOI: 10.1029/2007RG000233 (cited on page).

[21] E. Hart, J. Chung, and M. Chung. "A paired autoencoder framework for inverse problems via bayes risk minimization". In: *SIAM Journal on Scientific Computing* 48.2 (2026), pages C385–C414 (cited on page).

[22] H. Hersbach et al. "The ERA5 global reanalysis". In: *Quarterly Journal of the Royal Meteorological Society* 146.730 (2020), pages 1999–2049. DOI: https://doi.org/10.1002/qj.3803 (cited on page).

[23] X. Huang et al. "An Observationally Based Global Band-by-Band Surface Emissivity Dataset for Climate and Weather Simulations". In: *Journal of the Atmospheric Sciences* 73.9 (2016), pages 3541–3555. DOI: 10.1175/JAS-D-15-0355.1. URL: https://journals.ametsoc.org/view/journals/atsc/73/9/jas-d-15-0355.1.xml (cited on page).

[24] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*. Springer, 2006 (cited on page).

[25] M. D. King et al. "Spatial and Temporal Distribution of Clouds Observed by MODIS Onboard the Terra and Aqua Satellites". In: *IEEE Transactions on Geoscience and Remote Sensing* 51.7 (2013), pages 3826–3852. DOI: 10.1109/TGRS.2012.2227333 (cited on page).

[26] B. O. Koopman. "Hamiltonian systems and transformation in Hilbert space". In: *Proceedings of the National Academy of Sciences* 17.5 (1931), pages 315–318 (cited on page).

[27] M. A. Kramer. "Nonlinear principal component analysis using autoassociative neural networks". In: *AIChE Journal* 37.2 (1991), pages 233–243 (cited on page).

[28] Y. LeCun, Y. Bengio, and G. Hinton. "Deep learning". In: *Nature* 521.7553 (2015), pages 436–444 (cited on page).

[29] X. Liu et al. "Development of a fast and accurate PCRTM radiative transfer model in the solar spectral region." In: *Applied Optics* 55.29 (Oct. 2016), pages 8236–8247. DOI: https://doi.org/10.1364/AO.55.008236 (cited on page).

[30] B. Lusch, J. N. Kutz, and S. L. Brunton. "Deep learning for universal linear embeddings of nonlinear dynamics". In: *Nature communications* 9.1 (2018), page 4950 (cited on page).

[31] T. Maestri, W. Cossich, and I. Sbrolli. "Cloud identification and classification from high spectral resolution data in the far infrared and mid-infrared". In: *Atmospheric Measurement Techniques* 12.7 (2019), pages 3521–3540. DOI: 10.5194/amt-12-3521-2019. URL: https://amt.copernicus.org/articles/12/3521/2019/ (cited on page).

[32] T. Maestri and R. Rizzi. "A study on infrared diabatic forcing of ice clouds in the tropical atmosphere". In: *Journal of Geophysics Research* 108 (2003), page 4139. DOI: https://doi.org/10.1029/2002JD002146 (cited on page).

[33] T. Maestri et al. "Fast radiative transfer in multiple scattering atmospheres at far and mid infrared wavelengths". In: *AIP Conference Proceedings* 2988.1 (Jan. 2024), page 030003. ISSN: 0094-243X. DOI: 10.1063/5.0183019. eprint: https://pubs.aip.org/aip/acp/article-pdf/doi/10.1063/5.0183019/18570192/030003_1_5.0183019.pdf. URL: https://doi.org/10.1063/5.0183019 (cited on page).

[34] G. Martin, D. Johnson, and A. Spice. "The Measurement and Parametrization of Effective Radius of Droplets in Warm Stratocumulus Clouds". In: *Journal of Atmospheric Sciences* 51 (13 1994), pages 1823–1842. DOI: https://doi.org/10.1175/1520-0469(1994)051<1823:TMAPOE>2.0.CO;2 (cited on page).

[35] M. Martinazzo et al. "Assessment of the accuracy of scaling methods for radiance simulations at far and mid infrared wavelengths". In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 271 (2021), page 107739. ISSN: 0022-4073. DOI: https://doi.org/10.1016/j.jqsrt.2021.107739. URL: https://www.sciencedirect.com/science/article/pii/S0022407321002326 (cited on page).

[36] G. Masiello et al. "Developments of the σ-IASI/F2N radiative transfer model: a new linear-in-T approach for accurate treatment of clouds". In: *Proceedings of the SPIE conference: Remote Sensing of Clouds and the Atmosphere XXIX*. 2024 (cited on page).

[37] G. Masiello et al. "The new σ-IASI code for all sky radiative transfer calculations in the spectral range 10 to 2760 cm-1: σ-IASI/F2N". In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 312 (2024), page 108814. ISSN: 0022-4073. DOI: https://doi.org/10.1016/j.jqsrt.2023.108814. URL: https://www.sciencedirect.com/science/article/pii/S0022407323003321 (cited on page).

[38] P. Mastro et al. "Cloud detection from IASI hyperspectral data: a statistical approach based on neural networks". In: *Remote Sensing of Clouds and the Atmosphere XXV*. Edited by A. Comerón et al. SPIE, Sept. 2020, page 11. DOI: 10.1117/12.2573326. URL: http://dx.doi.org/10.1117/12.2573326 (cited on page).

[39] K. Masuda, T. Takashima, and Y. Takayama. "Emissivity of pure and sea waters for the model sea surface in the infrared window regions". In: *Remote Sensing of Environment* 24 (2 1988), pages 313–329. DOI: 10.1016/0034-4257(88)90032-6 (cited on page).

[40] M. Matricardi and R. Saunders. "Fast radiative transfer model for simulation of infrared atmospheric sounding interferometer radiances". In: *Appl. Opt.* 38.27 (Sept. 1999), pages 5679–5691. DOI: 10.1364/AO.38.005679. URL: http://ao.osa.org/abstract.cfm?URI=ao-38-27-5679 (cited on page).

[41] M. Matricardi et al. "An improved general fast radiative transfer model for the assimilation of radiance observations". In: *Quarterly Journal of the Royal Meteorological Society* 130.596 (2004), pages 153–173. DOI: https://doi.org/10.1256/qj.02.181. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1256/qj.02.181. URL: https://onlinelibrary.wiley.com/doi/abs/10.1256/qj.02.181 (cited on page).

[42] R. H. Norton and R. Beer. "New apodizing functions for Fourier spectrometry". In: *J. Opt. Soc. Am.* 66.3 (Mar. 1976), pages 259–264. DOI: 10.1364/JOSA.66.000259. URL: https://opg.optica.org/abstract.cfm?URI=josa-66-3-259 (cited on page).

[43] L. Palchetti et al. "FORUM: unique far-infrared satellite observations to better understand how Earth radiates energy to space". In: *Bulletin of the American Meteorological Society* (Aug. 2020), pages 1–52. ISSN: 0003-0007. DOI: 10.1175/BAMS-D-19-0322.1. eprint: https://journals.ametsoc.org/bams/article-pdf/doi/10.1175/BAMS-D-19-0322.1/4990048/bamsd190322.pdf. URL: https://doi.org/10.1175/BAMS-D-19-0322.1 (cited on page).

[44] J. J. Remedios et al. "MIPAS reference atmospheres and comparisons to V4.61/V4.62 MIPAS level 2 geophysical data sets". In: *Atmospheric Chemistry and Physics Discussions* 7 (2007), pages 9973–10017. DOI: 10.5194/acpd-7-9973-2007. URL: https://www.atmos-chem-phys-discuss.net/7/9973/2007/ (cited on page).

[45] C. D. Rodgers. “Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation”. In: *Reviews of Geophysics* 14.4 (1976), pages 609–624. DOI: https://doi.org/10.1029/RG014i004p00609. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/RG014i004p00609. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/RG014i004p00609 (cited on page).

[46] M. Saito et al. “Spaceborn middle- and far-infrared observatiosns improving nighttime ice cloud properties retrievals.” In: *Geophysics Research Letter* 47 (2020). DOI: https://doi.org10.1029/2020GL087491 (cited on page).

[47] R. Saunders et al. “An update on the RTTOV fast radiative transfer model (currently at version 12)”. In: *Geoscientific Model Development* 11.7 (2018), pages 2717–2737. DOI: 10.5194/gmd-11-2717-2018 (cited on page).

[48] C. Serio et al. “Retrieval of foreign-broadened water vapor continuum coefficients from emitted spectral radiance in the H_2O rotational band from 240 to 590 cm-1”. In: *Optics Express* 16.20 (2008), page 15816. DOI: 10.1364/OE.16.015816 (cited on page).

[49] C. Serio et al. “Demonstration of a physical inversion scheme for all-sky, day-night IASI observations and application to the analysis of the onset of the Antarctica ozone hole: Assessment of retrievals and consistency of forward modeling”. In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 329 (2024), page 109211. ISSN: 0022-4073. DOI: https://doi.org/10.1016/j.jqsrt.2024.109211. URL: https://www.sciencedirect.com/science/article/pii/S0022407324003182 (cited on page).

[50] C. Sgattoni, L. Sgheri, and M. Chung. “A data-driven approach for fast atmospheric radiative transfer inversion”. In: *Inverse Problems* 41.8 (2025), page 085006. DOI: 10.1088/1361-6420/adf3dd (cited on page).

[51] C. Sgattoni et al. “Characterization of Surface Spectral Emissivity Retrieved from EE9-FORUM Simulated Measurements”. In: *Remote Sensing in Earth Systems Sciences* 7 (2024), pages 15–25. DOI: 10.1007/s41976-024-00102-6 (cited on page).

[52] C. Sgattoni et al. “FORUM All-Sky Simulated Observations Database for Scene Classification and Fast Radiative Transfer Inversion”. In: *Zenodo* (2025). URL: https://doi.org/10.5281/zenodo.17969539 (cited on page).

[53] L. Sgheri and E. Castelli. “Speeding up the DISORT solver: mathematical approach and application to radiance simulations of FORUM”. In: *ESA ATMOS Conference, Salzburg*. 2018. DOI: 10.5281/zenodo.8119188 (cited on page).

[54] L. Sgheri, P. Raspollini, and M. Ridolfi. “Auto-adaptive Tikhonov regularization of water vapor profiles: application to FORUM measurements”. In: *Applicable Analysis* 101.10 (2022), pages 3733–3743. DOI: 10.1080/00036811.2020.1751825 (cited on page).

[55] L. Sgheri et al. "The FORUM end-to-end simulator project: architecture and results". In: *Atmospheric Measurement Techniques* 15.3 (2022), pages 573–604. DOI: 10.5194/amt-15-573-2022. URL: https://amt.copernicus.org/articles/15/573/2022/ (cited on page).

[56] L. Sgheri, C. Sgattoni, and C. Zugarini. "Determination of emissivity profiles using a Bayesian data-driven approach". In: *Mathematics and Computers in Simulation* 229 (2025), pages 512–524. ISSN: 0378-4754. DOI: https://doi.org/10.1016/j.matcom.2024.10.015. URL: https://www.sciencedirect.com/science/article/pii/S0378475424004051 (cited on page).

[57] D. Simeoni, C. Singer, and G. Chalon. "Infrared atmospheric sounding interferometer". In: *Acta Astronautica* 40.2 (1997). Enlarging The Scope of Space Applications, pages 113–118. ISSN: 0094-5765. DOI: https://doi.org/10.1016/S0094-5765(97)00098-2. URL: https://www.sciencedirect.com/science/article/pii/S0094576597000982 (cited on page).

[58] Y. Song et al. "Score-Based Generative Modeling through Stochastic Differential Equations". In: *International Conference on Learning Representations (ICLR)*. 2021 (cited on page).

[59] K. Stamnes et al. "Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media". In: *Appl. Opt.* 27 (1988), pages 2502–2509. DOI: doi:10.1364/AO.27.002502 (cited on page).

[60] K. Stamnes, S.-C. Tsay, and I. Laszlo. *DISORT, a General-Purpose Fortran Program for Discrete-Ordinate-Method Radiative Transfer in Scattering and Emitting Layered Media: Documentation of Methodology*. Technical report. NASA, 2000. URL: https://www.libradtran.org/lib/exe/fetch.php?media=disortreport1.1.pdf (cited on page).

[61] G. Tang et al. "Improvement of the Simulation of Cloud Longwave Scattering in Broadband Radiative Transfer Models". In: *Journal of the Atmospheric Sciences* 75.7 (2018), pages 2217–2233. DOI: 10.1175/JAS-D-18-0014.1. URL: https://journals.ametsoc.org/view/journals/atsc/75/7/jas-d-18-0014.1.xml (cited on page).

[62] A. Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005 (cited on page).

[63] D. D. Turner and U. Lohnert. "Information Content and Uncertainties in Thermodynamic Profiles and Liquid Cloud Properties Retrieved from the Ground-Based Atmospheric Emitted Radiance Interferometer (AERI)". In: *Journal of Applied Meteorology and Climatology* 53.3 (2014), pages 752–771. DOI: 10.1175/JAMC-D-13-0126.1. URL: https://journals.ametsoc.org/view/journals/apme/53/3/jamc-d-13-0126.1.xml (cited on page).

[64] D. D. Turner et al. “Cloud Phase Determination Using Ground-Based AERI Observations at SHEBA”. In: *Journal of Applied Meteorology* 42.6 (2003), pages 701–715. DOI: 10.1175/1520-0450(2003)042<0701:CPDUGA>2.0.CO;2. URL: https://journals.ametsoc.org/view/journals/apme/42/6/1520-0450_2003_042_0701_cpduga_2.0.co_2.xml (cited on page).

[65] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg. *Plug-and-Play Priors for Model Based Reconstruction*. Technical report ECE Technical Reports 448. Purdue University, 2013 (cited on page).

[66] L. Warwick et al. “Retrieval of Tropospheric Water Vapor From Airborne Far-Infrared Measurements: A Case Study”. In: *Journal of Geophysical Research: Atmospheres* 127.7 (2022). e2020JD034229 2020JD034229, e2020JD034229. DOI: https://doi.org/10.1029/2020JD034229. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020JD034229. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020JD034229 (cited on page).

[67] F. J. Wentz et al. *Remote Sensing Systems Cross-Calibrated Multi-Platform (CCMP) Wind Vector Analysis Product*. Technical report. Remote Sensing Systems, 2015. URL: https://www.remss.com/measurements/ccmp/ (cited on page).

[68] S. Whitburn et al. “A CO_2-independent cloud mask from Infrared Atmospheric Sounding Interferometer (IASI) radiances for climate applications”. In: *Atmospheric Measurement Techniques* 15.22 (2022), pages 6653–6668. DOI: 10.5194/amt-15-6653-2022. URL: https://amt.copernicus.org/articles/15/6653/2022/ (cited on page).

[69] K. Wyser. “The Effective Radius in Ice Clouds”. In: *Journal of Climate* 11.7 (1998), pages 1793–1802. ISSN: 08948755, 15200442. DOI: https://doi.org/10.1175/1520-0442(1998)011<1793:TERIIC>2.0.CO;2 (cited on page).

[70] P. Yang et al. “Spectral signature of ice clouds in the far-infrared region: Single-scattering calculations and radiative sensitivity study”. In: *Journal of Geophysical Research* 108.D18 (2003), page 4569. DOI: 10.1029/2002JD003291 (cited on page).

[71] P. Yang et al. “Radiative properties of cirrus clouds in the infrared (8–13μm) spectral region”. In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 70.4 (2001). Light Scattering by Non-Spherical Particles, pages 473–504. ISSN: 0022-4073. DOI: https://doi.org/10.1016/S0022-4073(01)00024-3. URL: https://www.sciencedirect.com/science/article/pii/S0022407301000243 (cited on page).

[72] P. Yang et al. “Spectrally Consistent Scattering, Absorption, and Polarization Properties of Atmospheric Ice Crystals at Wavelengths from 0.2 to 100 µm”. In: *Journal of the Atmospheric Sciences* 70.1 (2013), pages 330–347. DOI: 10.1175/JAS-D-12-039.1. eprint: https://doi.org/10.1175/JAS-D-12-039.1. URL: https://doi.org/10.1175/JAS-D-12-039.1 (cited on page).

Appendix A. Forward Model

The forward mapping $\mathbf{F}^{\rightarrow}(\mathbf{x})$ represents a combination of the radiative transfer equation and the instrumental effects. Assuming a horizontal and azimuthal symmetry for the problem, the radiative transfer equation can be written as the following initial value problem:

$$\begin{cases} \mu \frac{dI(\tau,\mu)}{d\tau} = I(\tau,\mu) - \frac{\omega(\tau)}{2} \int_{-1}^{1} P(\tau,\mu,\mu') I(\tau,\mu') d\mu' - [1-\omega(\tau)]\, B(\tau), \\ I(\tau_0,\mu) = I_0(\mu), \end{cases} \tag{A.1}$$

where μ denotes the cosine of the zenith angle, τ the integrated total cloud optical depth from the top of the atmosphere to the layer of interest, ω the single scattering albedo of the layer at level τ, B the Planck function, and $P(\tau,\mu,\mu')$ the azimuthally averaged scattering phase function that describes scattering events for radiation entering the layer at angle μ' and exiting at μ. The integral term in equation A.1, describing the multiple-scattering interactions, substantially increases the computational cost of the solution. This is mainly due to the increased dimensionality of the problem in solvers such as the DIScrete Ordinate Radiative Transfer (DISORT) method [53, 59, 60]. A possible solution for mitigating the computational burden is to avoid the direct calculation of the multiple scattering term by scaling the absorption optical depth of the cloud. The fast radiative transfer code σ-IASI/F2N [37] exploits scaling methodologies to accelerate radiative transfer calculations while maintaining solution accuracy. Specifically, the code implements the Chou approximation [11] and the Tang adjustment routine [61]. In this work, to generate the training and test sets, the Chou solution is considered. This scaling methodology allows considering an apparent absorption optical depth $\tilde{\tau}$, defined as

$$\tilde{\tau} = \tau\big((1-\omega) + \omega b\big) \tag{A.2}$$

where b is called back-scattering coefficient and describes the mean fraction of radiation back-scattered from the medium. After this scaling, the radiative transfer equation can be written as a Schwarzschild-like form, and the radiative transfer problem becomes

$$\begin{cases} \mu \frac{dI(\tilde{\tau},\mu)}{d\tilde{\tau}} = I(\tilde{\tau},\mu) - [1-\omega(\tilde{\tau})]\, B(\tilde{\tau}), \\ I(\tilde{\tau}_0,\mu) = I_0(\mu), \end{cases} \tag{A.3}$$

Finally, instrumental effects are modeled by convolving the pseudo-monochromatic radiances produced by σ-IASI/F2N with the strong Norton-Beer apodized kernel, followed by the addition of NESR according to the expected FORUM noise level.

Appendix B. Input Preprocessing for Radiance Reconstruction

Studying how radiance responds to perturbations in atmospheric profiles provides valuable insight into the most appropriate pre-processing to apply to the input variables of the inversion model. Figure B1 shows the derivatives of the forward model with respect to the water vapor concentration $\frac{\partial s}{\partial w_{\text{vap}}}$, and ozone concentration $\frac{\partial s}{\partial o}$, respectively. These quantities are directly computed by σ-IASI/F2N considering the US standard atmosphere as input scenario.

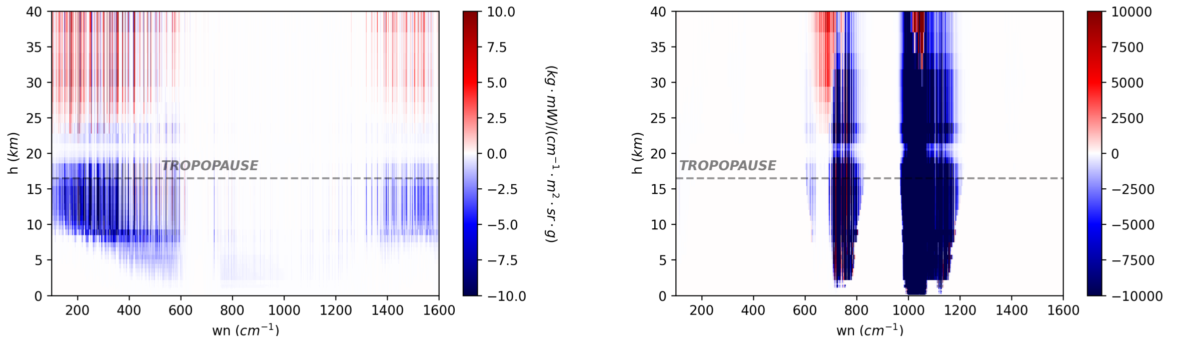

Figure B1. σ-IASI/F2N Jacobians with respect to the water vapor concentration (left panel) and ozone concentration (right panel). These quantities are calculated considering the US standard atmosphere and successively convolved to FORUM resolution.

The radiance shows a strong sensitivity to water vapor variations across the whole vertical profile, with a similar behavior observed for the ozone channels. Accurately reconstructing the vertical distributions of water vapor and ozone throughout the atmosphere is therefore essential to ensure consistency between the measured spectrum and the radiance simulated from the retrieved quantities.

In this work, the vertical concentration of water vapor and ozone are transformed using the inverse softplus function. Specifically:

$$\tilde{x} = \ln(e^{x \cdot c} - 1) \tag{B.1}$$

where x is the quantity of interest and c is a scaling constant, which is set to be $c = 1$ for water vapor and $c = 10^6$ for ozone. This transformation behaves approximately linearly for $x >> 1$ and approximately logarithmically for $x << 1$. This reduces the range of magnitudes spanned by the physical quantities and helps the model, during training, to give sufficient importance to regions where the concentration profile is very low, such as stratospheric water vapor and tropospheric ozone.