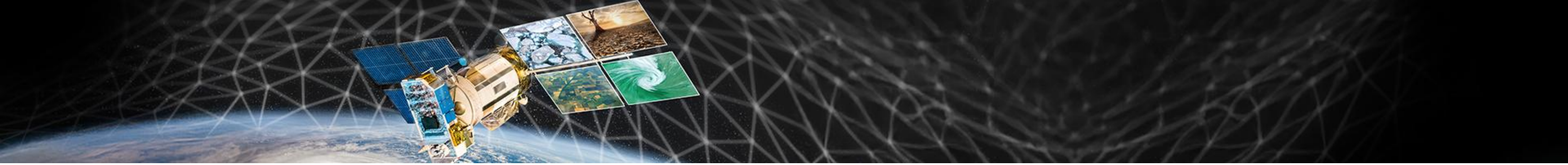
International workshop on AI Foundation Model for EO

5-7 May 2025 | ESA ESRIN - Frascati, Italy



Consiglio Nazionale delle Ricerche
Istituto per la BioEconomia

A physics-aware data-driven surrogate approach for fast atmospheric radiative transfer inversion

Sgattoni C.[1], Chung M.[2], Sgheri L.[3]

[1] *Institute of BioEconomy (IBE), National Research Council (CNR), Florence, Italy*

[2] *Department of Mathematics, Emory University, Atlanta, GA, USA*

[3] *Institute of Applied Mathematics (IAC), National Research Council (CNR), Florence, Italy*

Introduction – FORUM mission

- **FORUM**[1] (Far-infrared Outgoing Radiation Understanding and Monitoring) is a Fourier Transform Spectrometer selected as the ninth Earth Explorer mission by the European Space Agency in 2019.

- It will provide interferometric measurements in the Far-InfraRed (FIR) spectrum (100-1600 cm^{-1} region), constituting 50% of Earth's outgoing longwave flux.

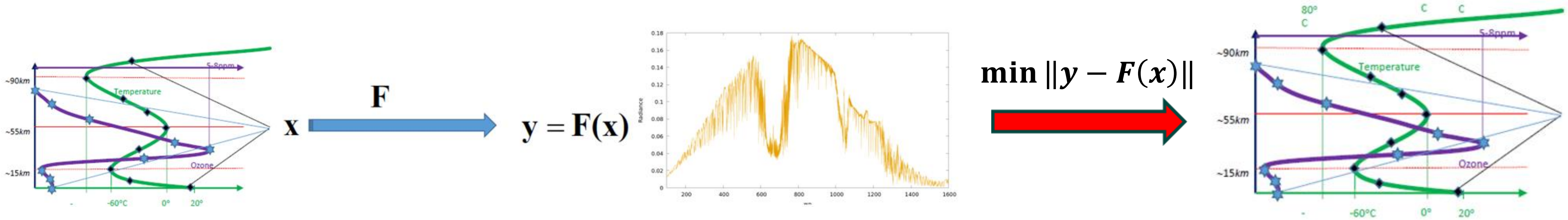
- Accurate Top Of the Atmosphere measurements in the FIR are crucial for improving climate models.

[1] 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/

Background

Direct problem: from the atmospheric status vector **x** find the simulated spectrum $\mathbf{y} = \mathbf{F}(\mathbf{x})$, with **F** known as **forward model.**

Inverse problem: from the measured spectrum **y** find the parameter vector **x** (**retrieval vector**) which minimizes $||\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{x})||$.



Retrieval – classical approach

- Find the atmospheric parameters $\mathbf{x}$ (**surface temperature, temperature, water vapor, ozone, surface spectral emissivity, clouds parameters**) that best reconstruct the measured spectrum $\mathbf{y}$.

! VERY ILL-CONDITIONED PROBLEM

- The problem is formulated as a Bayesian inference problem, solved using the **OPTIMAL ESTIMATION METHOD**[2]:

$$\mathbf{x}_{\mathrm{OE}} = \arg\min_{\boldsymbol{\xi}} \frac{1}{2}\left\|\mathbf{L}_{\mathbf{y}}\left(\mathbf{y} - \mathbf{F}(\boldsymbol{\xi})\right)\right\|_2^2 + \frac{1}{2}\left\|\mathbf{L}_{\mathrm{a}}(\boldsymbol{\xi} - \mathbf{x}_{\mathrm{a}})\right\|_2^2,$$

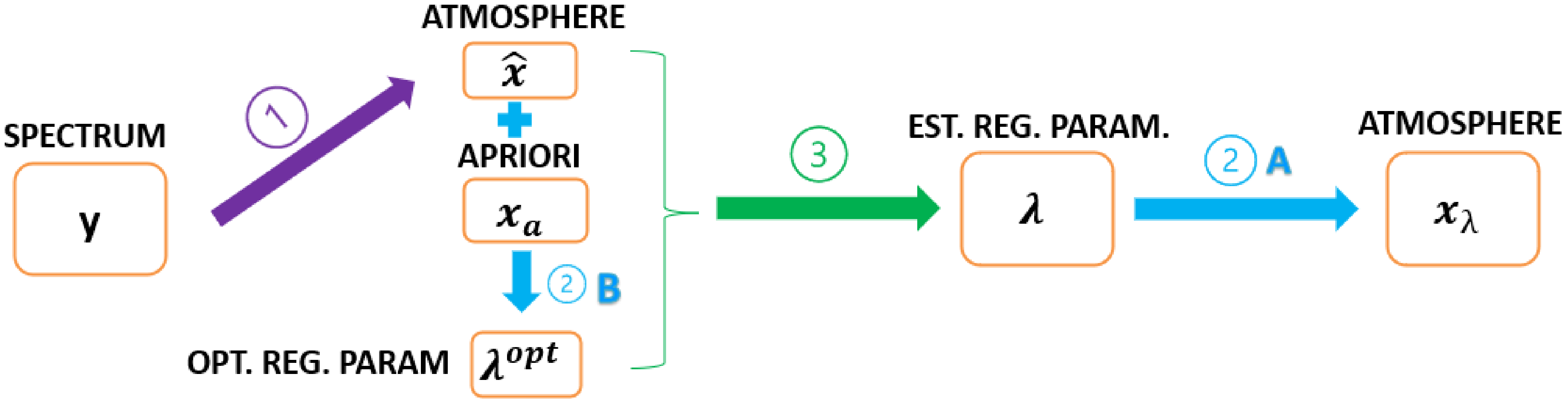
where $\mathbf{S}_{\mathbf{y}}^{-1} = \mathbf{L}_{\mathbf{y}}^{\mathrm{T}}\mathbf{L}_{\mathbf{y}}$ and $\mathbf{S}_{\mathrm{a}}^{-1} = \mathbf{L}_{\mathrm{a}}^{\mathrm{T}}\mathbf{L}_{\mathrm{a}}$ are the inverses of the covariance matrices of the measurements $\mathbf{y}$ and the a priori information $\boldsymbol{x}_a$, respectively.

- The minimization is carried out using Gauss Newton + Levenberg-Marquardt technique.

[2] Rodgers, C. D.: Inverse Methods for Atmospheric Sounding, World Scientific, https://doi.org/10.1142/3171, 2000.

New method scheme[3] – clear sky

1) Approximation of the RT inverse operator using a linear operator, trained on a database of FORUM simulated measurements (**completely data-driven phase**).
2) Incorporation of a priori information into the data-driven solution (**regularization technique**).
3) Estimation of the optimal regularization parameters using a neural network, trained on a database of pre-computed optimal parameters (**second training phase**).



[3] Sgattoni, C., Chung, T., Sgheri, L.: A physics-aware data-driven surrogate approach for fast atmospheric radiative transfer inversion, submitted to Inverse Problems, 2024 http://arxiv.org/abs/2410.22609

1. Data-driven model – clear sky

- ❑ Approximation of the RT inversion with a linear operator **Z** trained on simulated FORUM measurements
- ❑ Training set 1 (January and July 2021, 12:00, **clear sky**, 1708 cases all over the globe):
  - ➢ **X** = $[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_\mathrm{N}]$ → N atmospheric scenarios (dim: **425**x1708),
  - ➢ **Y** = $[\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_\mathrm{N}]$ → N simulated FORUM spectra (dim: **4049**x1708).

$$\arg\min_{\mathbf{Z}} \mathrm{f}(\mathbf{Z}) = \arg\min_{\mathbf{Z}} \|\mathbf{X} - \mathbf{Z}\mathbf{Y}\|_\mathrm{F}^2$$

$$\frac{\delta \mathrm{f}}{\delta \mathbf{Z}} = -2\mathbf{L}_\mathbf{y}^\mathrm{T}\mathbf{L}_\mathbf{y}\mathbf{X}\mathbf{Y}^\mathrm{T} + 2\mathbf{L}_\mathbf{y}^\mathrm{T}\mathbf{L}_\mathbf{y}\mathbf{Z}\mathbf{Y}\mathbf{Y}^\mathrm{T}$$

A minimizer $\hat{\mathbf{Z}}$ of f solves $\hat{\mathbf{Z}}\mathbf{Y}\mathbf{Y}^\mathrm{T} = \mathbf{X}\mathbf{Y}^\mathrm{T}$.

We can express:

$$\hat{\boldsymbol{Z}} = \mathbf{X}\mathbf{Y}^+ \;\Rightarrow\; \hat{\mathbf{x}} = \hat{\mathbf{Z}}\mathbf{y}.$$

*** Moore-Penrose pseudoinverse**

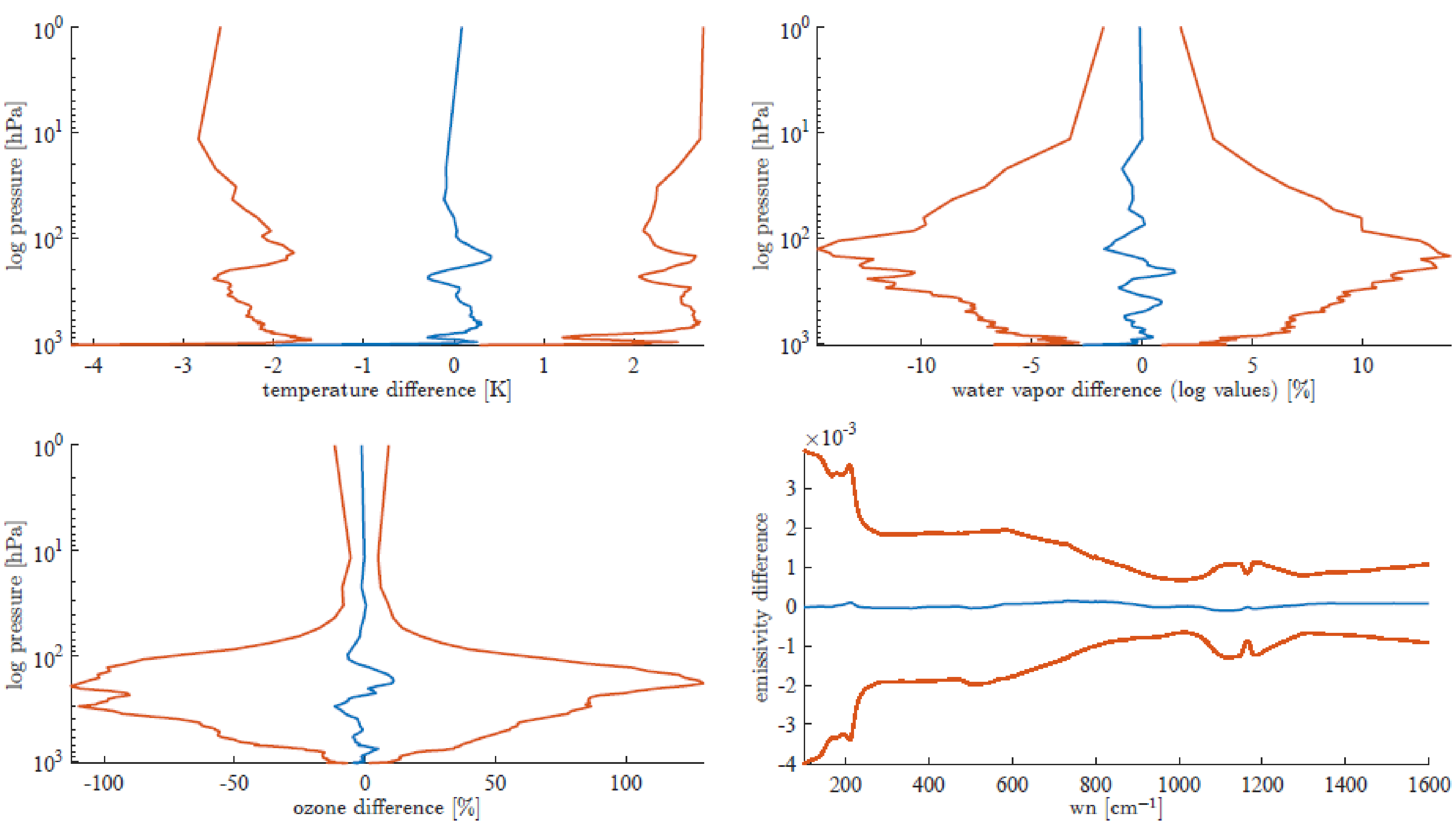
Let **M** be a matrix of rank k with singular value decomposition $\mathbf{M} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\mathrm{T}$, the Moore-Penrose pseudoinverse of **M** is given by

$$\mathbf{M}^+ = \mathbf{V}\,\tilde{\boldsymbol{\Sigma}}\,\mathbf{U}^\mathrm{T},$$

$$\tilde{\boldsymbol{\Sigma}} = \mathrm{diag}\left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \dots, \frac{1}{\sigma_k}, 0, \dots, 0\right).$$

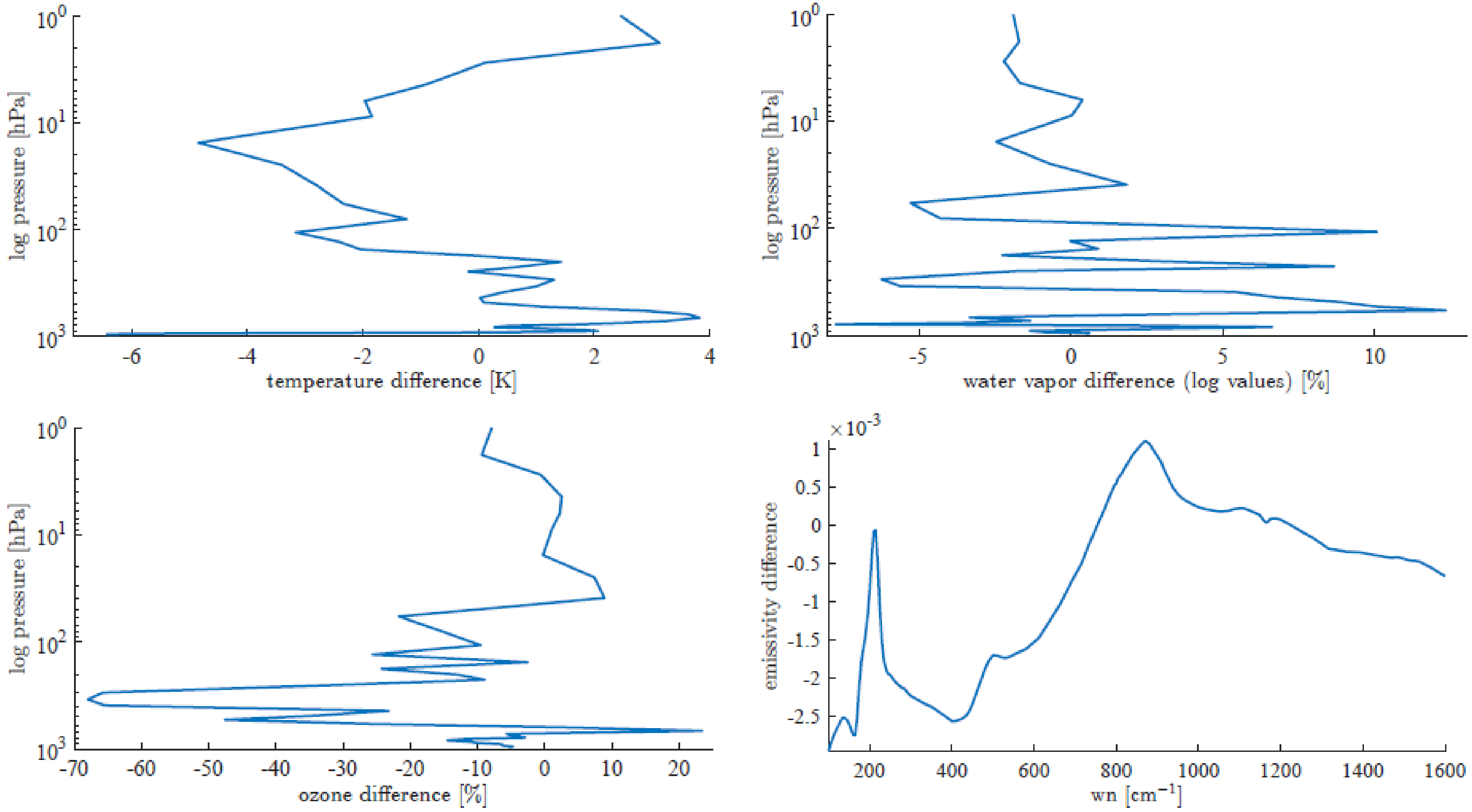
1. Data-driven results – clear sky

Mean signed (blue) and unsigned (orange) errors for Test Set 1 (396 cases) globally:



1. Data-driven results – clear sky

Signed error for case 25 in Test Set 1:



2. Tikhonov regularization – clear sky

A. Bilevel Optimization Problem: Inner Problem

- **Additional a priori information**:

$$\mathbf{x}_{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{\xi}} \frac{1}{2}\left\|\mathbf{L}_{\mathbf{x}}\left(\boldsymbol{\xi} - \hat{\mathbf{Z}}\mathbf{y}\right)\right\|_2^2 + \frac{1}{2}\|(\boldsymbol{\Lambda}\,\mathbf{L}_{\mathrm{a}}(\boldsymbol{\xi} - \mathbf{x}_{\mathrm{a}})\|_2^2, \text{ with}$$

- $\boldsymbol{\Lambda}$ is a diagonal regularization matrix, where $\boldsymbol{\lambda} = \mathrm{diag}(\boldsymbol{\Lambda})$,
- $\mathbf{S}_{\mathbf{x}}^{-1} = \mathbf{L}_{\mathbf{x}}^{\mathrm{T}}\mathbf{L}_{\mathbf{x}}$ is the inverse of the experimental covariance matrix,
- $\mathbf{x}_{\mathrm{a}}$ has been generated from covariance matrix $\mathbf{S}_{\mathrm{a}}$[4], with $\mathbf{S}_{\mathrm{a}}^{-1} = \mathbf{L}_{\mathrm{a}}^{\mathrm{T}}\mathbf{L}_{\mathrm{a}}$.

B. Bilevel Optimization Problem: Outer Problem

- **Optimal Regularization Parameters**: for each of the J cases in Test Set 1 (now renamed Training Set 2), compute :

$$\boldsymbol{\lambda}_{\mathrm{j}}^{\mathrm{opt}} = \arg\min_{\boldsymbol{\lambda}} \frac{\left\|(\mathbf{x}_{\boldsymbol{\lambda}})_{\mathrm{j}} - \mathbf{x}_{\mathrm{j}}\right\|_2}{\left\|\mathbf{x}_{\mathrm{j}}\right\|_2}, \quad \mathrm{j} = 1, \cdots, \mathrm{J}.$$

- **Optimization Method**: interior points method.

[4]defined by the UK MetOffice for assimilation of IASI products into the operational Numerical Weather Prediction (NWP) system.

3. Regularization parameter estimation clear sky

- Assume there exists a well-defined mapping $\widetilde{\boldsymbol{\Phi}}(\hat{\boldsymbol{x}} - \boldsymbol{x}_a) = \boldsymbol{\lambda}$.
- Set a NEURAL NETWORK $\boldsymbol{\Phi}$ parametrized by $\boldsymbol{\theta}$ to approximate $\widetilde{\boldsymbol{\Phi}}$.
- Given training data $\left((\hat{\mathbf{x}}_j - (\mathbf{x}_a)_j), \boldsymbol{\lambda}_j^{\mathrm{opt}}\right)_{j=1}^{J}$ the following equation is solved:

unique network for all 5 components

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\eta}} \frac{1}{J} \sum_{j=1}^{J} \left\| \boldsymbol{\Phi}\left((\hat{\mathbf{x}}_j - (\mathbf{x}_a)_j), \boldsymbol{\eta}\right) - \boldsymbol{\lambda}_j^{\mathrm{opt}} \right\|^2$$

!

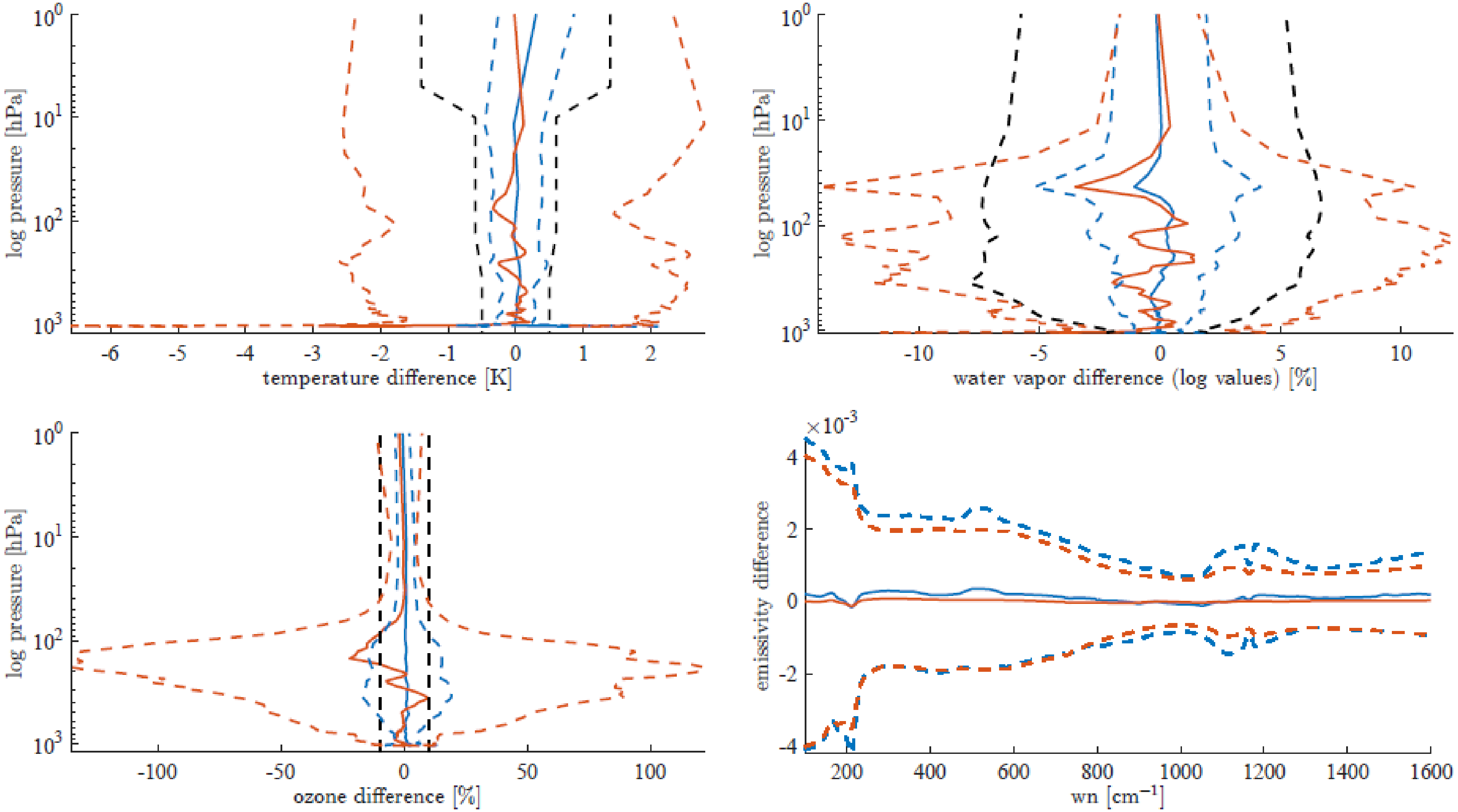
Neural Network INPUT: $\hat{\mathbf{x}} - \mathbf{x}_a$

Neural Network OUTPUT (prediction): $\log(\boldsymbol{\lambda}^{\mathrm{nn}})$

Neural Network ARCHITECTURE: 3 hidden layers **[425] → [15] → [10] → [5] → [5]**

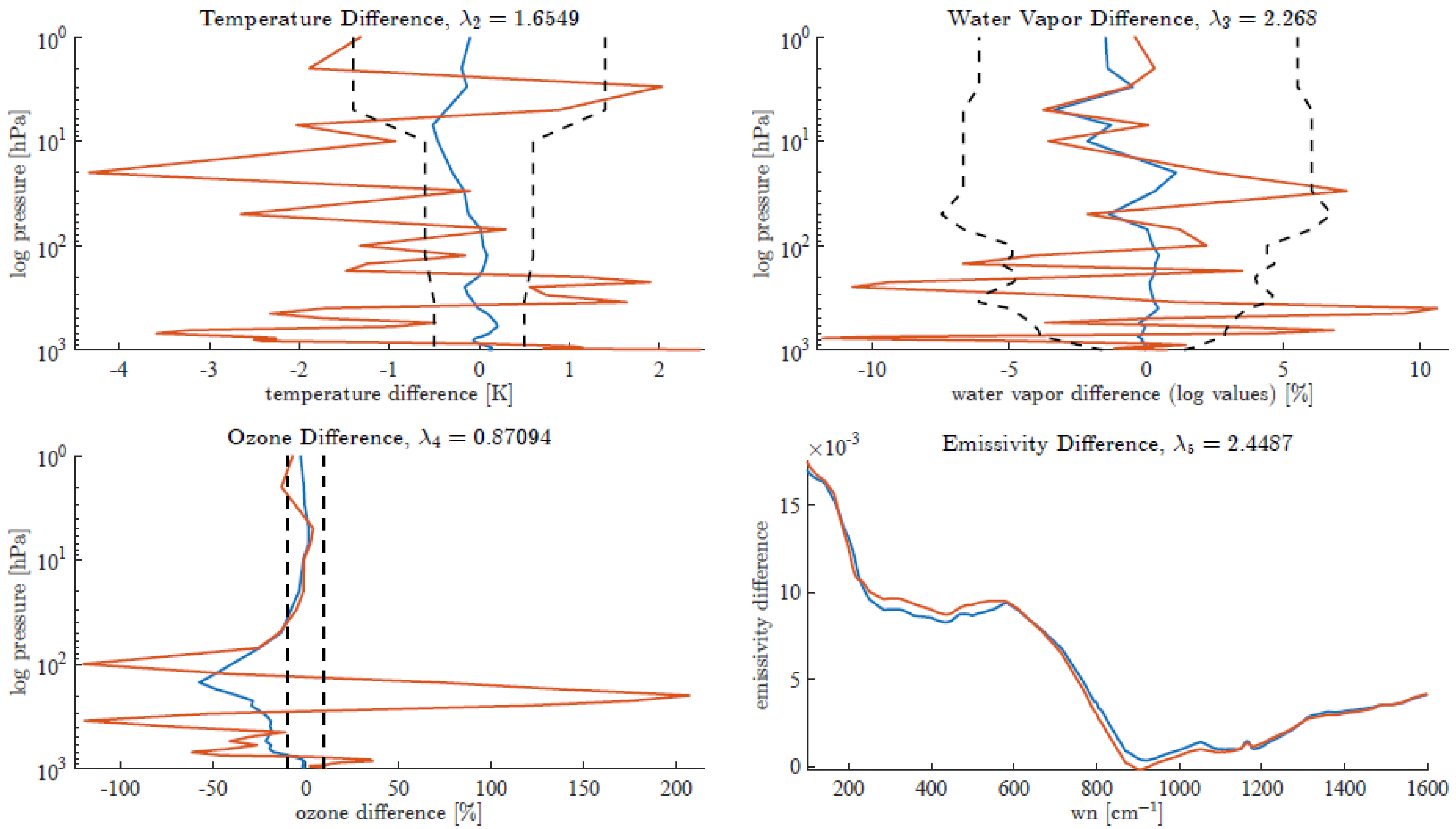
Results - clear sky

Mean signed (bold) and unsigned (dashed) errors for Test Set 2 (402 cases) globally related to:
data-driven solution $\widehat{x}$, regularized solution x_λ, a priori errors



Results - clear sky

Signed errors for case 15 in Test Set 2 related to: data-driven solution $\widehat{x}$, regularized solution x_λ, a priori errors



Computational time – clear sky

- ❑ **Training 1 → 0.178 seconds (offline stage)**
- ❑ **Training 2 → 128.984 seconds (offline stage)**
- ❑ **Testing - data-driven phase → about 0.0008 seconds for 1 case**
- ❑ **Testing - reg. par. estimation → about 0.007 seconds for 1 case**
- ❑ **Testing - entire solution scheme → about 0.05 seconds for 1 case**

New variables – all sky

Retrieval variables: $\mathbf{x} = \left[\, \mathrm{T}^0, \mathbf{T}, \mathbf{w}_{\mathrm{vap}}, \mathbf{o}, \mathbf{e}, \mathbf{c}_{\mathrm{liq}}, \mathbf{c}_{\mathrm{ice}}, \mathbf{r}_{\mathrm{liq}}, \mathbf{r}_{\mathrm{ice}} \right]^{\mathrm{T}} \in \mathrm{R}^{722}$

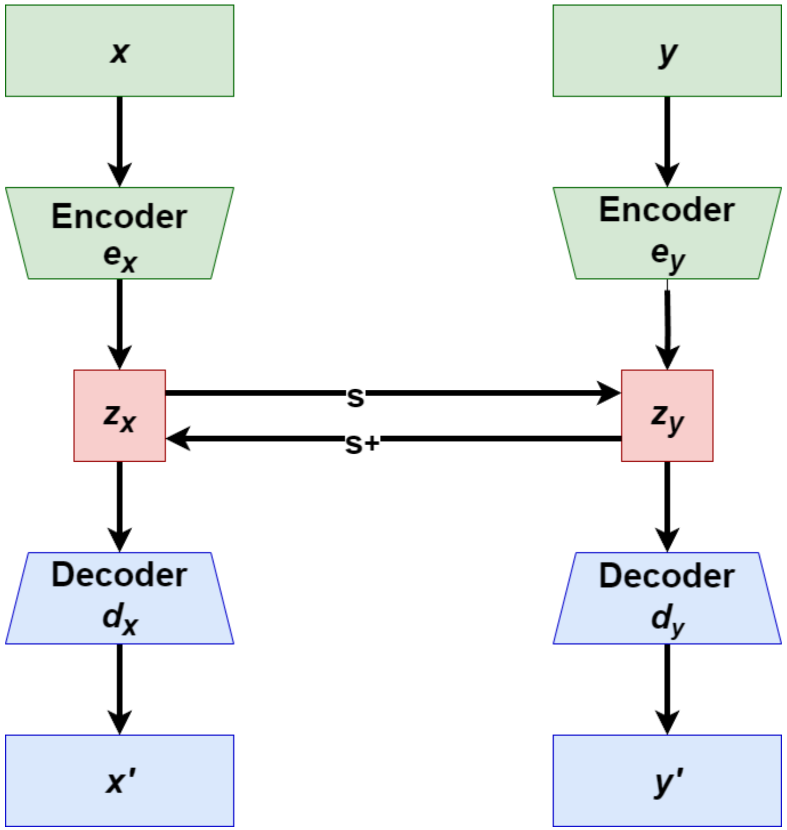
expansion of X dimension from 425 to 722

Input variables: $\mathbf{y} = [\mathrm{y}_{\mathrm{lon}}, \mathrm{y}_{\mathrm{lat}}, \mathrm{d}_{\mathrm{loc}}, \mathrm{h}_{\mathrm{loc}}, \mathbf{p}, \mathbf{s}]^{\mathrm{T}} \in \mathrm{R}^{4233}$

expansion of Y dimension from 4049 to 4233

- $T^0, y_{lon}, y_{lat}, d_{loc}, h_{loc,} \tau_{liq}, \tau_{ice} \in R,$
- $\boldsymbol{T}, \boldsymbol{w}_{vap}, \mathbf{o}, \mathbf{c}_{\mathrm{liq}}, \mathbf{c}_{\mathrm{ice}}, \mathbf{r}_{\mathrm{liq}}, \mathbf{r}_{\mathrm{ice}}, \mathbf{p} \in R^{60}$
- $\boldsymbol{e} \in R^{301}$ (from 100 cm^{-1} to 1600 cm^{-1} with stepsize 5 cm^{-1}),
- $\boldsymbol{s} \in R^{4169}$ (from 100 cm^{-1} to 1600 cm^{-1} with stepsize 0.36 cm^{-1}).

Architecture[5] – all sky



- **Joint learning**
- **Batch learning** (batch_size = 64)
- **Epochs** (10k)
- **Training data:** shuffled randomly
- **Testing data:** original order
- **Loss function:**

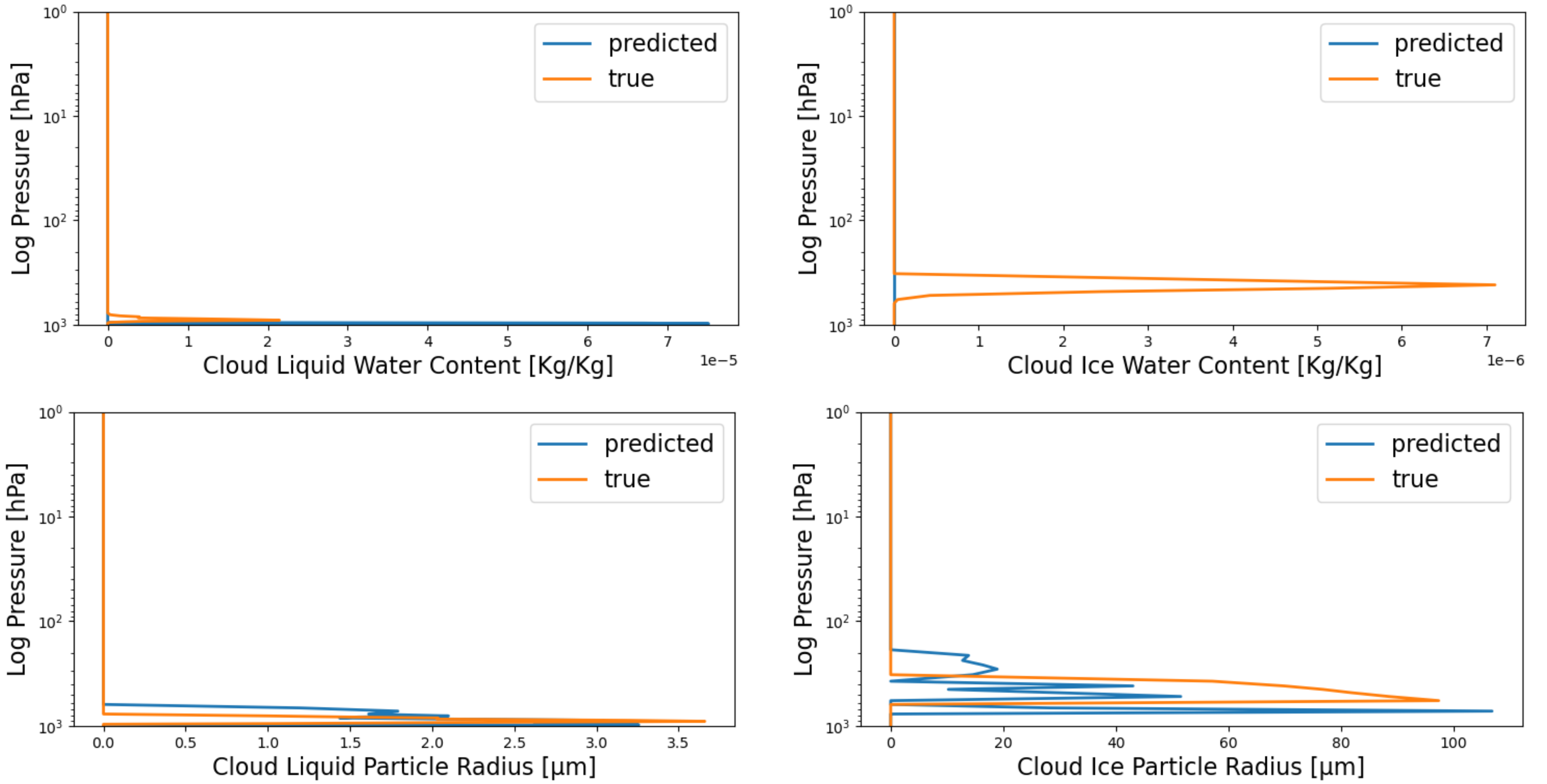
$$\mathrm{E} = \mathrm{w_{xx}}\mathrm{E}_{\mathbf{xx}} + \mathrm{w_{yy}}\mathrm{E}_{\mathbf{yy}} + \mathrm{w_{xy}}\mathrm{E}_{\mathbf{xy}} + \mathrm{w_{yx}}\mathrm{E}_{\mathbf{yx}},$$ with $\mathrm{w_{xx}} = \mathrm{w_{yy}} = \mathrm{w_{yx}} = 1$ and $\mathrm{w_{xy}} = 0.5$

- **ADAM (ADAptive Moment Estimation) Optimization**

[5] Matthias Chung *et al* 2024 *Mach. Learn.: Sci. Technol.* **5** 045055 DOI: 10.1088/2632-2153/ad95dd

Results – all sky

Case 6 in Test Set



Physical coherence – all sky

Retrieved $\mathbf{c}_{\mathrm{liq}}, \mathbf{c}_{\mathrm{ice}}, \mathbf{r}_{\mathrm{liq}}, \mathbf{r}_{\mathrm{ice}}$ are not physically consistent → **second training phase!**

The following training is activated at each layer l only when the content and the radius are not consistent!

4 Feedforward neural networks (MLP, fully-connected):

[4] → [16] → [8] → [1]

LINEAR + ReLU | **LINEAR + ReLU** | **LINEAR + ReLU**

INPUT: $(\mathbf{c}^{\mathrm{l}})^{\mathrm{inp}}$ or $(\mathbf{r}^{\mathrm{l}})^{\mathrm{inp}}$, $\mathbf{p}^{\mathrm{l}}$, $\mathbf{T}^{\mathrm{l}}$, $\mathbf{w}^{\mathrm{l}}_{\mathrm{vap}}$ OUTPUT: $(\mathbf{c}^{\mathrm{l}})^{\mathrm{new}}$ or $(\mathbf{r}^{\mathrm{l}})^{\mathrm{new}}$

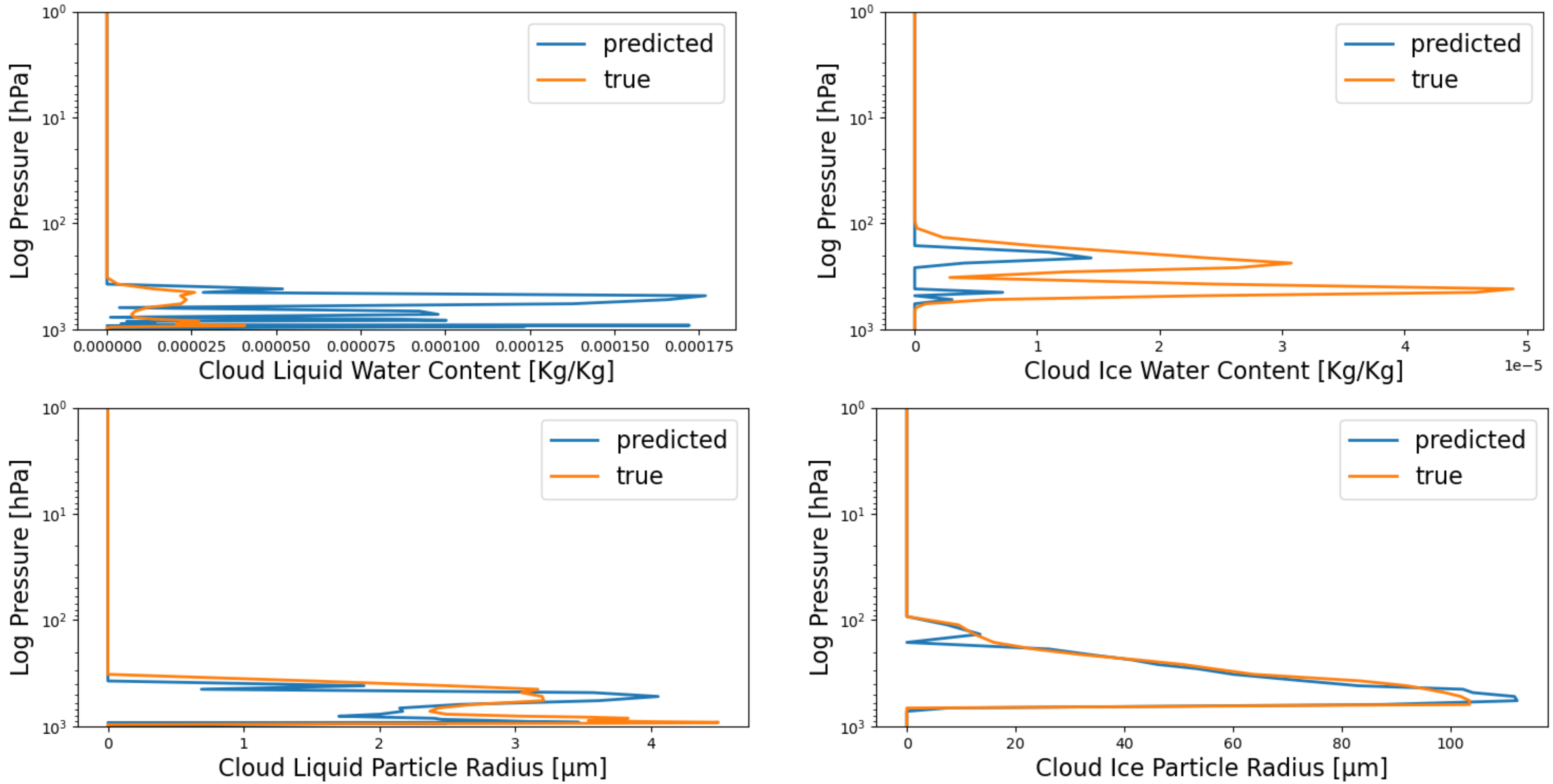
We have new parameters keeping the others fixed → same procedure and settings as in the first training.

Loss function (with a Coherence term): $\mathrm{E} = \mathrm{MSE}_1 + \mathrm{MSE}_2 + \mathrm{MSE}_3 + \mathrm{MSE}_4 + \mathrm{w}_{\mathrm{coh}}\mathrm{C}$

$$\mathrm{C} = \frac{1}{N}\sum\left|\boldsymbol{\sigma}(10\, \boldsymbol{c}_{liq}) - \boldsymbol{\sigma}(10\, \boldsymbol{r}_{liq})\right| + \frac{1}{N}\left|\boldsymbol{\sigma}(10\, \boldsymbol{c}_{ice}) - \boldsymbol{\sigma}(10\, \boldsymbol{r}_{ice})\right|, \quad w_{coh} = 100,\ \boldsymbol{\sigma} \text{ sigmoid function.}$$

Results – all sky

Results - Case 6 in Test Set



Computation times - all sky

Training the first 4 models → it depends on the epochs and laptops → some hours

Training the last 4 networks → it depends on the epochs and laptops → some minutes

Testing → **Total time taken for predictions (2504): 1.61 seconds**

Average time per batch (64): 0.0403 seconds

*The computational times refer to my personal computer, which is equipped with an AMD Ryzen 5 7530U processor with Radeon Graphics.

Take-home messages

- **Instantaneous Method for Retrieval in Clear-Sky:**
  - First approximation using a machine-learning approach (pseudoinverse operator).
  - Incorporation of a priori physical information through regularization.
  - Machine-learning approach for estimating regularization parameters (feedforward neural network).
  - Applied to FORUM simulated measurements; data generated with line-by-line code.

- **Instantaneous Method for Retrieval in All-Sky:**
  - Machine-learning architecture combining autoencoders and linear mapping.
  - Introduction of physical constraints via additional neural networks (feedforward NN) and a coherence-forcing term.
  - Applied to FORUM simulated measurements (data produced using a fast radiative transfer code) and real data from other instruments.

Thank you for your attention!