Enhancing Atmospheric Transmittances Estimation for TOVs through Advanced Statistical Approaches

Jean-Marie Lalande †, James Hocking *, Jérôme Vidot †
NWPSAF Project Team

† CNRM/CEMS, Météo France – CNRS
* Met Office

May 9th, 2025

Motivations

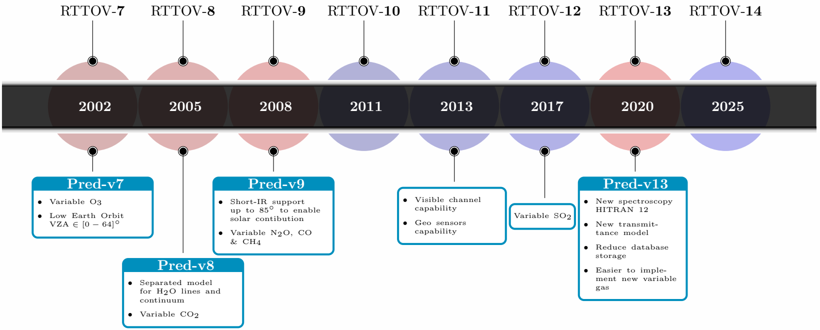
- A pressing need exists to improve our understanding of the RTTOV coefficients generation process
  ⇝ **Retired experts:** Marco Matricardi, Pascal Brunel, Roger Saunders, John Eyre, Hal Woolf

- Many of our users treat the so-called *RTTOV coefficients* as a black-box

- Conduct a thorough investigation of all aspects of the process, focusing on efficiency, stability, and accuracy
  ⇝ **This is still going on**

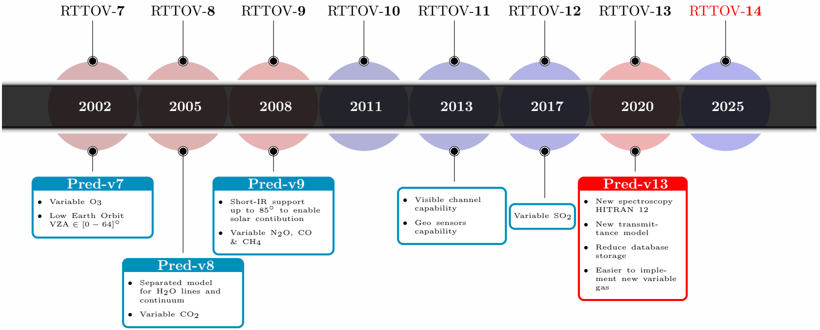
- Suggest potential improvements based on the findings

Overview of RTTOV gas absorption's parametric models



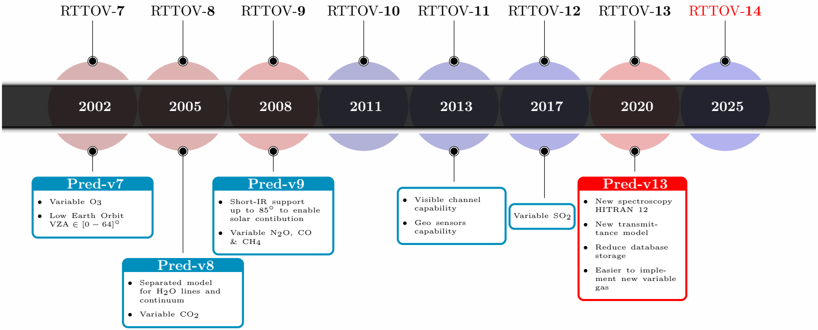
- **New** parametric **model version** are **named** after the **RTTOV version** in which **they have been introduced**.

Overview of RTTOV gas absorption's parametric models



- **New** parametric **model version** are **named** after the **RTTOV version** in which **they have been introduced**.
- RTTOV-v14 was released earlier this year
- Retirement of Pred 7, 8 & 9 anticipated for RTTOV-v15

Overview of RTTOV gas absorption's parametric models



- **New** parametric **model version** are **named** after the **RTTOV version** in which **they have been introduced**.
- RTTOV-v14 was released earlier this year
- Retirement of Pred 7, 8 & 9 anticipated for RTTOV-v15
- *McMillin et al., Applied Optics, 1976, 1977, 1979, 1995a, 1995b, 2003*
- *Hocking et al., GMD, 2022*

Theoretical fundation of the gas absorption parameteric model

Let's denote j the **channel index** and k the **layer index**.

Predictors v13 - 3 gases version - Model formulation

$$\widetilde{t}_{jk}^{\,\mathrm{total}} = \widetilde{t}_{jk}^{\,\mathrm{mix}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,cont}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,line}} \cdot \widetilde{t}_{jk}^{\,\mathrm{O_3}} \cdot \widetilde{t}_{jk}^{\,\mathrm{CO_2}} \cdot \widetilde{t}_{jk}^{\,\mathrm{c}}$$

Where the final correction term is

$$\widetilde{t}_{jk}^{\,\mathrm{c}} = \frac{\widetilde{t}_{jk}^{\,*,\mathrm{total}}}{\widetilde{t}_{jk}^{\,\mathrm{mix}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,cont}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,line}} \cdot \widetilde{t}_{jk}^{\,\mathrm{O_3}} \cdot \widetilde{t}_{jk}^{\,\mathrm{CO_2}}}$$

With $\widetilde{t}_{jk}^{\,*,\mathrm{total}}$ is the true total optical depth.

Theoretical fundation of the gas absorption parameteric model

Let's denote j the **channel index** and k the **layer index**.

Predictors v13 - 3 gases version - Model formulation

$$\widetilde{t}_{jk}^{\,\mathrm{total}} = \widetilde{t}_{jk}^{\,\mathrm{mix}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,cont}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,line}} \cdot \widetilde{t}_{jk}^{\,\mathrm{O_3}} \cdot \widetilde{t}_{jk}^{\,\mathrm{CO_2}} \cdot \widetilde{t}_{jk}^{\,\mathrm{c}}$$

Where the final correction term is

$$\widetilde{t}_{jk}^{\,\mathrm{c}} = \frac{\widetilde{t}_{jk}^{\,*,\mathrm{total}}}{\widetilde{t}_{jk}^{\,\mathrm{mix}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,cont}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,line}} \cdot \widetilde{t}_{jk}^{\,\mathrm{O_3}} \cdot \widetilde{t}_{jk}^{\,\mathrm{CO_2}}}$$

With $\widetilde{t}_{jk}^{\,*,\mathrm{total}}$ is the true total optical depth.

Where $\widetilde{t}_{jk}^{\,(\cdot)}$ are the channel integrated transmittances:

$$\widetilde{t}_{jk}^{\,(\cdot)} = \frac{\int_{\bar{\nu}} t_{jk}^{(\cdot)}(\nu')\mathrm{srf}_j(\nu')d\nu'}{\mathrm{Cst}}$$

Theoretical fundation of the gas absorption parameteric model

Let's denote j the **channel index** and k the **layer index**.

Predictors v13 - 3 gases version - Model formulation

$$\widetilde{t}_{jk}^{\,\mathrm{total}} = \widetilde{t}_{jk}^{\,\mathrm{mix}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,cont}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,line}} \cdot \widetilde{t}_{jk}^{\,\mathrm{O_3}} \cdot \widetilde{t}_{jk}^{\,\mathrm{CO_2}} \cdot \widetilde{t}_{jk}^{\,\mathrm{c}}$$

Where the final correction term is

$$\widetilde{t}_{jk}^{\,\mathrm{c}} = \frac{\widetilde{t}_{jk}^{\,*,\mathrm{total}}}{\widetilde{t}_{jk}^{\,\mathrm{mix}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,cont}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,line}} \cdot \widetilde{t}_{jk}^{\,\mathrm{O_3}} \cdot \widetilde{t}_{jk}^{\,\mathrm{CO_2}}}$$

With $\widetilde{t}_{jk}^{\,*,\mathrm{total}}$ is the true total optical depth.

Where $\widetilde{t}_{jk}^{\,(\cdot)}$ are the channel integrated transmittances:

$$\widetilde{t}_{jk}^{\,(\cdot)} = \frac{\int_{\bar{\nu}} t_{jk}^{(\cdot)}(\nu')\mathrm{srf}_j(\nu')d\nu'}{\mathrm{Cst}}$$

Where,

$$\widetilde{t}^{\,(\cdot)} \in \mathbf{\Omega} = \{\widetilde{t}_{jk}^{\,\mathrm{mix}},\ \widetilde{t}_{jk}^{\,\mathrm{wv,cont}},\ \widetilde{t}_{jk}^{\,\mathrm{wv,line}},\ \widetilde{t}_{jk}^{\,\mathrm{O_3}},\ \widetilde{t}_{jk}^{\,\mathrm{CO_2}},\ \widetilde{t}_{jk}^{\,\mathrm{c}}\}$$

Theoretical fundation of the gas absorption parameteric model

Let's denote j the **channel index** and k the **layer index**.

Predictors v13 - 3 gases version - Model formulation

$$\widetilde{t}_{jk}^{\,\mathrm{total}} = \widetilde{t}_{jk}^{\,\mathrm{mix}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,cont}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,line}} \cdot \widetilde{t}_{jk}^{\,\mathrm{O_3}} \cdot \widetilde{t}_{jk}^{\,\mathrm{CO_2}} \cdot \widetilde{t}_{jk}^{\,\mathrm{c}}$$

Where the final correction term is

$$\widetilde{t}_{jk}^{\,\mathrm{c}} = \frac{\widetilde{t}_{jk}^{\,*,\mathrm{total}}}{\widetilde{t}_{jk}^{\,\mathrm{mix}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,cont}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,line}} \cdot \widetilde{t}_{jk}^{\,\mathrm{O_3}} \cdot \widetilde{t}_{jk}^{\,\mathrm{CO_2}}}$$

With $\widetilde{t}_{jk}^{\,*,\mathrm{total}}$ is the true total optical depth.

Where $\widetilde{t}_{jk}^{\,(\cdot)}$ are the channel integrated transmittances:

$$\widetilde{t}_{jk}^{\,(\cdot)} = \frac{\int_{\bar{\nu}} t_{jk}^{(\cdot)}(\nu')\mathrm{srf}_j(\nu')d\nu'}{\mathrm{Cst}}$$

Where,

$$\widetilde{t}^{\,(\cdot)} \in \boldsymbol{\Omega} = \{\widetilde{t}_{jk}^{\,\mathrm{mix}},\ \widetilde{t}_{jk}^{\,\mathrm{wv,cont}},\ \widetilde{t}_{jk}^{\,\mathrm{wv,line}},\ \widetilde{t}_{jk}^{\,\mathrm{O_3}},\ \widetilde{t}_{jk}^{\,\mathrm{CO_2}},\ \widetilde{t}_{jk}^{\,\mathrm{c}}\}$$

RTTOV coefficients are **estimated** for each terms of $\boldsymbol{\Omega}$, each channels and each layers:

$$N_{tot} = N_j \times N_k \times \mathrm{Card}(\boldsymbol{\Omega})$$

Theoretical fundation of the gas absorption parameteric model

Let's denote j the **channel index** and k the **layer index**.

Predictors v13 - 3 gases version - Model formulation

$$\widetilde{t}_{jk}^{\,\mathrm{total}} = \widetilde{t}_{jk}^{\,\mathrm{mix}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,cont}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,line}} \cdot \widetilde{t}_{jk}^{\,\mathrm{O_3}} \cdot \widetilde{t}_{jk}^{\,\mathrm{CO_2}} \cdot \widetilde{t}_{jk}^{\,\mathrm{c}}$$

Where the final correction term is

$$\widetilde{t}_{jk}^{\,\mathrm{c}} = \frac{\widetilde{t}_{jk}^{\,*,\mathrm{total}}}{\widetilde{t}_{jk}^{\,\mathrm{mix}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,cont}} \cdot \widetilde{t}_{jk}^{\,\mathrm{wv,line}} \cdot \widetilde{t}_{jk}^{\,\mathrm{O_3}} \cdot \widetilde{t}_{jk}^{\,\mathrm{CO_2}}}$$

With $\widetilde{t}_{jk}^{\,*,\mathrm{total}}$ is the true total optical depth.

Where $\widetilde{t}_{jk}^{\,(\cdot)}$ are the channel integrated transmittances:

$$\widetilde{t}_{jk}^{\,(\cdot)} = \frac{\int_{\bar{\nu}} t_{jk}^{(\cdot)}(\nu')\mathrm{srf}_j(\nu')d\nu'}{\mathrm{Cst}}$$

Where,

$$\widetilde{t}^{\,(\cdot)} \in \mathbf{\Omega} = \{\widetilde{t}_{jk}^{\,\mathrm{mix}},\ \widetilde{t}_{jk}^{\,\mathrm{wv,cont}},\ \widetilde{t}_{jk}^{\,\mathrm{wv,line}},\ \widetilde{t}_{jk}^{\,\mathrm{O_3}},\ \widetilde{t}_{jk}^{\,\mathrm{CO_2}},\ \widetilde{t}_{jk}^{\,\mathrm{c}}\}$$

RTTOV coefficients are **estimated** for each terms of $\mathbf{\Omega}$, each channels and each layers:

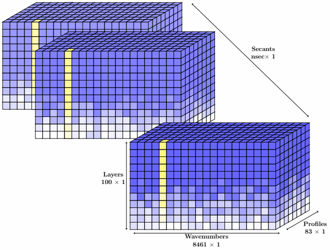
$$N_{tot} = N_j \times N_k \times \mathrm{Card}(\mathbf{\Omega})$$

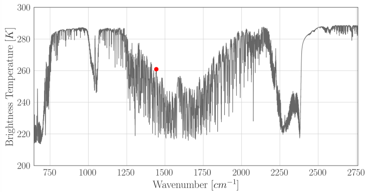
As an example...

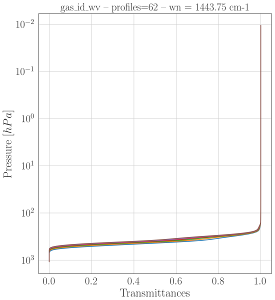
IASI pred 13 with 3 variable gases

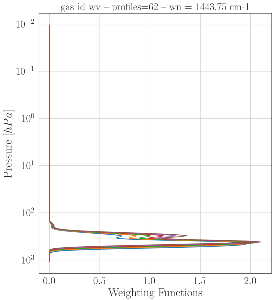
$$N_{iasi} = 8461 \times 100 \times 6 = \mathbf{5,076,600}$$

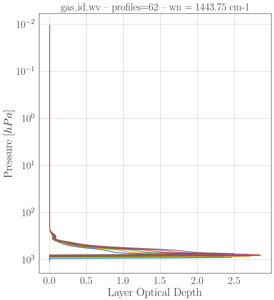
Channel Integrated transmittances datacubes



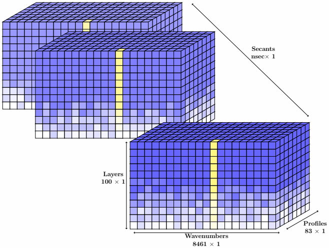


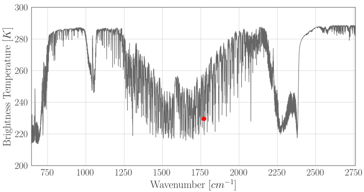


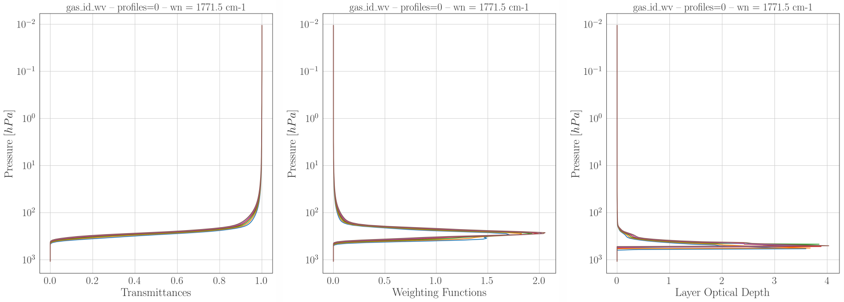




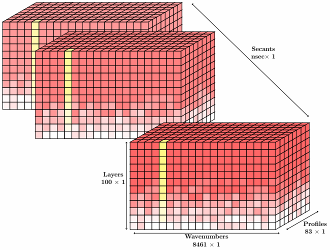
Channel Integrated transmittances datacubes

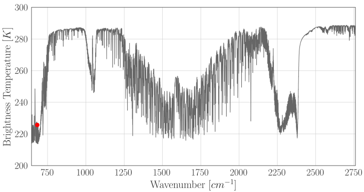


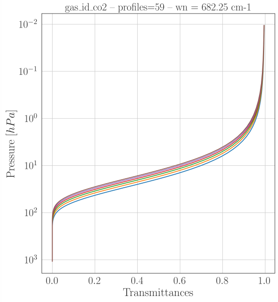


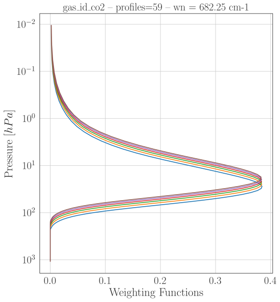


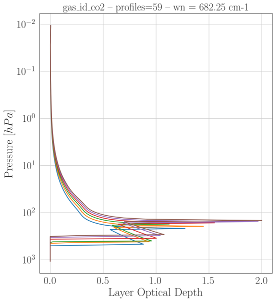
Channel Integrated transmittances datacubes











Theoretical fundation of the layer optical depth parameteric model

Layer optical depths $\widetilde{\boldsymbol{\tau}}^{*\,(.)}$ are modeled as a multilinear regression:

$$\widetilde{\boldsymbol{\tau}}^{*\,(.)}_{jk} = \boldsymbol{X}_k \boldsymbol{c}_{jk} + \boldsymbol{\epsilon}_{jk}$$

where,

- $\widetilde{\boldsymbol{\tau}}^{*\,(.)}_{jk}$: true layer optical depth: $[(\mathbf{N_{prof}} \times \mathbf{N_{sec}}) \times \mathbf{1}]$
- $\boldsymbol{c}^{(.)}_{jk}$: model parameters (RTTOV Coefficients): $[\mathbf{N_{pred}} \times \mathbf{1}]$
- $\boldsymbol{X}^{(.)}_{k}$: set of predictors depending on atmospheric model state and a specific gas: $[\mathbf{N_{pred}} \times (\mathbf{N_{prof}} \times \mathbf{N_{sec}})]$
- $\boldsymbol{\epsilon}^{(.)}_{jk}$: additive noise (i.e. gaussian distributed): $[(\mathbf{N_{prof}} \times \mathbf{N_{sec}}) \times \mathbf{1}]$

Predictors/Features for the water vapour lines

The **water vapour lines** predictors are:

$$\boldsymbol{x}_k^{\mathrm{wv,line}} = \begin{bmatrix} \left(\sec(\theta) W_{r,k}\right)^2 \\ \sec(\theta) W_{w,k} \\ \left((\sec\theta) W_{w,k}\right)^2 \\ \sec(\theta) W_{r,k}\, \delta T_k \\ \sqrt{\sec(\theta) W_{r,k}} \\ \sqrt[4]{\sec(\theta) W_{r,k}} \\ \sec(\theta) W_{r,k} \\ (\sec(\theta) W_{w,k})^{1.5} \\ (\sec(\theta) W_{r,k})^{1.5} \\ (\sec(\theta) W_{r,k})^{1.5}\, \delta T_k \\ \sqrt{\sec(\theta) W_{r,k}}\, \delta T_k \\ (\sec(\theta) W_{w,k})^{1.25} \\ \sec(\theta) W_{r,k}^2 / W_{w,k} \\ \sqrt{\sec(\theta) W_{r,k}}\, W_{r,k} / W_{wt,k} \\ \sec(\theta) \sqrt{W_{w,k}} \end{bmatrix}$$

- W_r, W_w, W_{wt} quantities depending on Water Vapour concentration
- δT depends on distance to the mean Temperature profile

Formulation of the coefficients estimation problem

Omitting j and k indexes.

The layer optical depth model parameters are estimated such as:

$$\widehat{\boldsymbol{c}}^{\,(\cdot)} = \arg\min_{\boldsymbol{c}} \mathcal{J}(\boldsymbol{c})$$

where $\mathcal{J}$ is a weighted quadratic function:

$$\mathcal{J}(\boldsymbol{c}) = \underbrace{\left(\widetilde{\boldsymbol{\tau}}^{*} - \widetilde{\boldsymbol{\tau}}^{(\cdot)}(\boldsymbol{c})\right)^{t} \boldsymbol{W} \left(\widetilde{\boldsymbol{\tau}}^{*} - \widetilde{\boldsymbol{\tau}}^{(\cdot)}(\boldsymbol{c})\right)}_{\text{weighted fit to target value}} + \underbrace{\alpha\, \boldsymbol{c}^{t}\boldsymbol{c}}_{\text{regularization}}$$

Where:

- $\widetilde{\boldsymbol{\tau}}^{*}$ are the true layer optical depth (target simulated with LBLRTM)
- $\widetilde{\boldsymbol{\tau}}^{(\cdot)}(\boldsymbol{c})$ are the model layer optical depth
- $\boldsymbol{W}$ is a diagonal matrix of weights
- The L_2 regularization term penalizes the "high values" of coefficients and is controlled by an hyperparameter α

Formulation of the coefficients estimation problem

Omitting j and k indexes.

The layer optical depth model parameters are estimated such as:

$$\widehat{\boldsymbol{c}}^{\,(\cdot)} = \arg\min_{\boldsymbol{c}} \mathcal{J}(\boldsymbol{c})$$

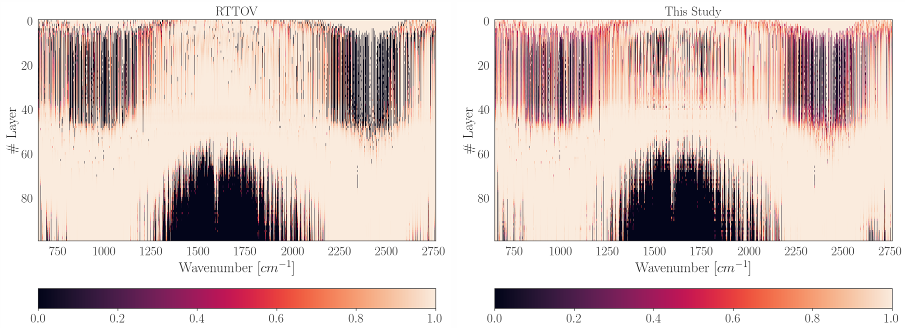
where $\mathcal{J}$ is a weighted quadratic function:

$$\mathcal{J}(\boldsymbol{c}) = \underbrace{(\widetilde{\boldsymbol{\tau}}^* - \boldsymbol{X}\boldsymbol{c})^t\,\boldsymbol{W}\,(\widetilde{\boldsymbol{\tau}}^* - \boldsymbol{X}\boldsymbol{c})}_{\text{weighted fit to target value}} + \underbrace{\alpha\;\boldsymbol{c}^t\boldsymbol{c}}_{\text{regularization}}$$

Where:

- $\widetilde{\boldsymbol{\tau}}^*$ are the true layer optical depth (target simulated with LBLRTM)
- $\widetilde{\boldsymbol{\tau}}^{(\cdot)}(\boldsymbol{c})$ are the model layer optical depth
- $\boldsymbol{W}$ is a diagonal matrix of weights
- The L_2 regularization term penalizes the "high values" of coefficients and is controlled by an hyperparameter α

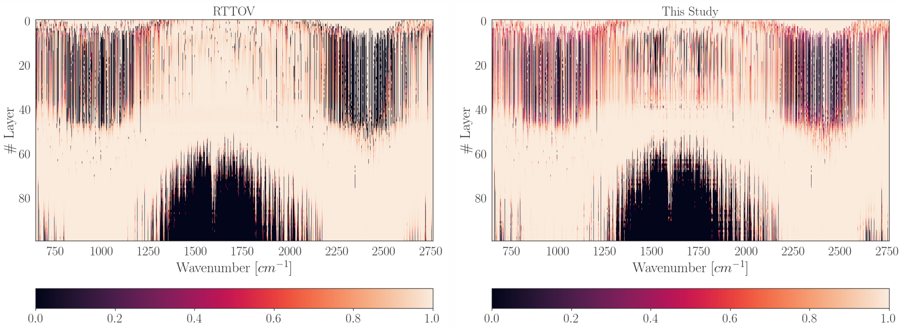
Coefficients estimation – Goodness of fit on Water Vapour datacube



Coefficient of determination of RTTOV versus using Singular Value Decomposition

- RTTOV (Direct computation of an inverse matrix) & Singular Value Decomposition shares some similar features in the overall metrics
- What are the differences ? We *theoretically* solve the same inverse problem...

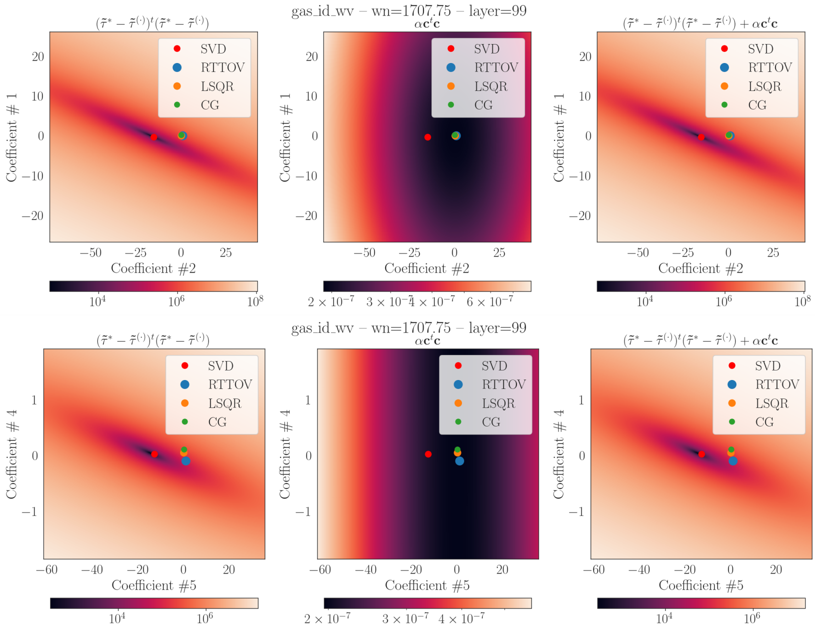
Coefficients estimation – Goodness of fit on Water Vapour datacube



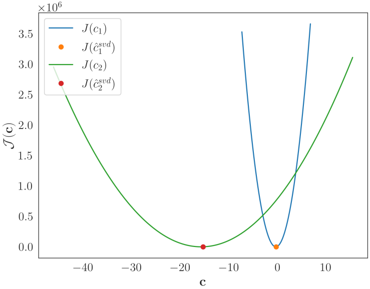
Coefficient of determination of RTTOV versus using Singular Value Decomposition

- RTTOV (Direct computation of an inverse matrix) & Singular Value Decomposition shares some similar features in the overall metrics
- What are the differences ? We *theoretically* solve the same inverse problem...
- It turns out that we don't really but that is not the full story!

Coefficients estimation – The impact of solver

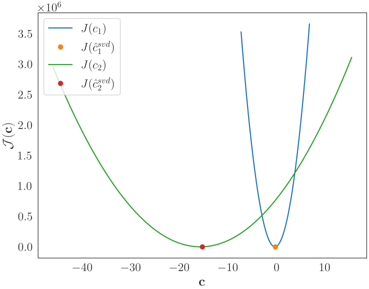


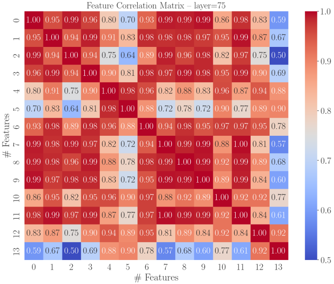
Coefficients estimation – An ill-conditioned problem



- Sensitivity to different coefficients can vary substantially
- Some Predictors have less to no resolving power with respect to the target values

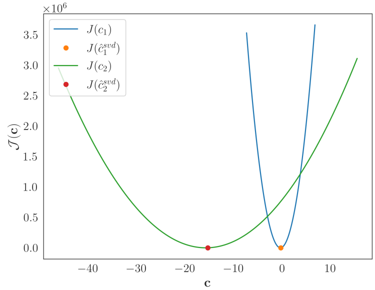
Coefficients estimation – An ill-conditioned problem

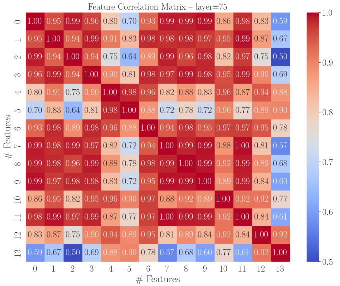




- Sensitivity to different coefficients can vary substantially
- Some Predictors have less to no resolving power with respect to the target values
- Predictors are overally strongly correlated

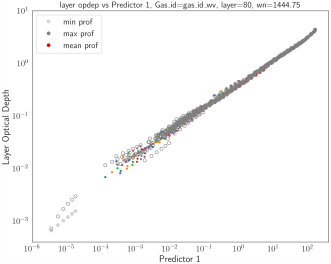
Coefficients estimation – An ill-conditioned problem



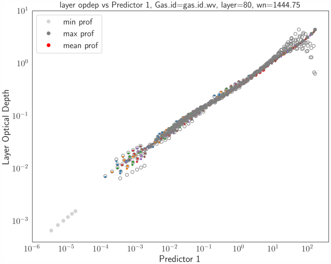


- Sensitivity to different coefficients can vary substantially
- Some Predictors have less to no resolving power with respect to the target values
- Predictors are overally strongly correlated
- That is one of the reasons we need preconditioning (normalization, whitening...)

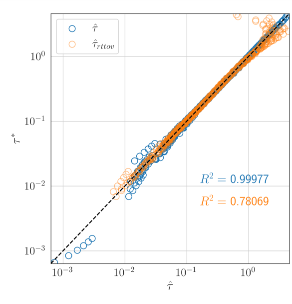
Coefficients estimation - How does the fit look in the layer optical depth space?

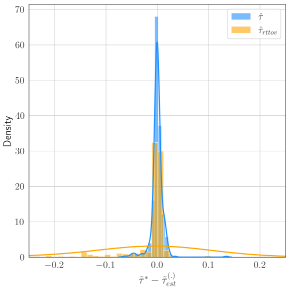


Fit with new coefficients.

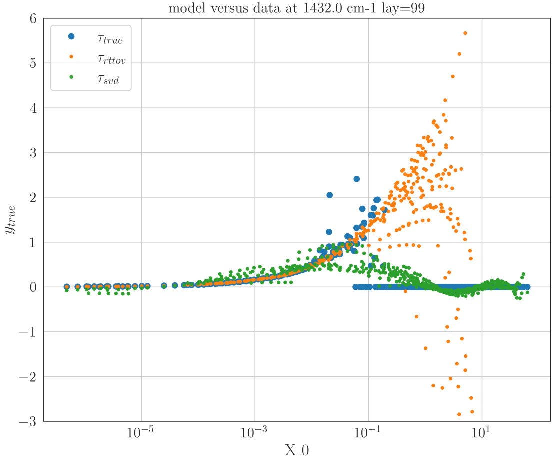


Fit with RTTOV coefficients.





A notorious problem of RTTOV – breakdown of the linear model



The model is predicting negative optical depth

Take home messages

- This is still an ongoing study which raises more questions than it answers yet!
- Our current inverse problem is strongly ill-conditioned
  $\rightsquigarrow$ Mostly because of colinear features
- Is their room for improvement without changing the philosophy of the gas absorption parametrization?
  $\rightsquigarrow$ I strongly believe so
- Hybrid approach classification/regression could leverage current model limitations
  $\rightsquigarrow$ Especially the nonlinearities
- Machine learning approach are probably flexible enough to fit the observed layer optical depth