BRIDGING PHYSICS AND STATISTICAL LEARNING METHOLODOGIES FOR THE ACCURATE MODELING OF THE RADIATIVE PROPERTIES OF NON-UNIFORM ATMOSPHERIC PATHS

F. André[1], C. Delage[1], L. Guilmard[1], M. Galtier[1], C. Cornet[2]

[1]Univ. Lyon, CNRS, UMR 5008 – CETHIL, Lyon, France

[2]Univ. Lille, CNRS, UMR 8518 – LOA, Lille, France

Corresponding Authors: frederic.andre@insa-lyon.fr , cindy.delage@insa-lyon.fr

PROJET FINANCÉ PAR L'ANR
ANR
PROJECT FUNDED BY THE ANR

BRIDGING PHYSICS AND STATISTICAL LEARNING METHODOLODGIES FOR THE ACCURATE MODELING OF THE RADIATIVE PROPERTIES OF NON-UNIFORM ATMOSPHERIC PATHS

F. André[1], C. Delage[1], L. Guilmard[1], M. Galtier[1], C. Cornet[2]

[1]Univ. Lyon, CNRS, UMR 5008 – CETHIL, Lyon, France
[2]Univ. Lille, CNRS, UMR 8518 – LOA, Lille, France
Corresponding Author: frederic.andre@insa-lyon.fr.

PROJET FINANCÉ PAR L'ANR – PROJECT FUNDED BY THE ANR

INTRODUCTION (1/7)

The concept of transmissivity is fundamental in radiative transfer in participating media, including gases.

Transmissivities:

- Represent the fraction of an incident radiative intensity that travels a distance L inside a medium without being absorbed.
- Appear naturally in the integral form of the RTE (here averaged over a spectral band $\Delta\nu$ – no scattering):

$$I^{\Delta\nu}\left(L\right) = I_b^{\Delta\nu}\left(0\right)\cdot\tau^{\Delta\nu}\left(0,L\right) + \int_0^L \frac{\partial\tau^{\Delta\nu}\left(s',L\right)}{\partial s'}\cdot I_b^{\Delta\nu}\left(s'\right)\; ds'$$

INTRODUCTION (2/7)

As LBL data are usually provided in absorption coefficient form, calculation of transmissivities requires evaluating:

$$\tau^{\Delta\nu}\left(0,L\right)=\tau^{\Delta\nu}\left(L\right)=\frac{1}{\Delta\nu}\cdot\int_{\Delta\nu}\exp\left(-\kappa_{\nu}L\right)\ d\nu$$

which can be computationally expensive even in uniform isothermal cases (Ex: O_2 A-band, 10748 coefficients*).

*** Ex : POLDER O_2 A-band**

Used to characterize the macrophysical properties of cloud (altitude, geometrical length)
ex : Ferlay et al. 2010 ; Desmons et al. 2013
LBL =>10748 coefficients d'absorption

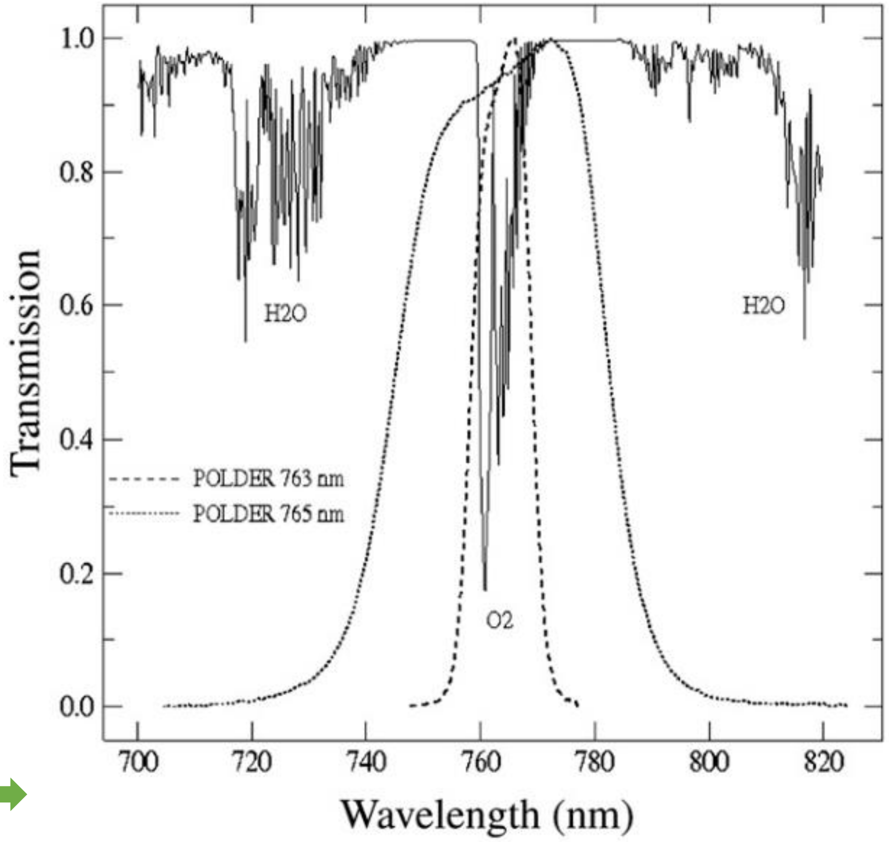


FIG. 1. Atmospheric transmission in the region of the oxygen A band at a resolution of 5 cm^{-1} ($\simeq$0.3 nm) and filter transmission in the narrow (10 nm) and wide (40 nm) POLDER bands centered at 763 and 765 nm, respectively.

INTRODUCTION (3/7)

Providing **accurate values of transmissivities** is usually said to be **difficult because gas spectra are made of many thin and overlapping spectral lines**.

This statement is **mostly irrelevant** because this problem was solved almost a century ago (even before the availability of data and codes for LBL calculations).

In reality, the **MAIN DIFFICULTY** concerns the **TREATMENT OF PATH NON-UNIFORMITIES** as encountered in almost all applications.

Standard non-uniform methods provide in reasonable cases an accuracy of a few percents that may be sufficient in some applications (in combustion for instance) but not in other ones (as in remote sensing).

THIS PROBLEM IS STILL UNSOLVED IN A GENERAL FRAME.

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR
ANR

INTRODUCTION (4/7)

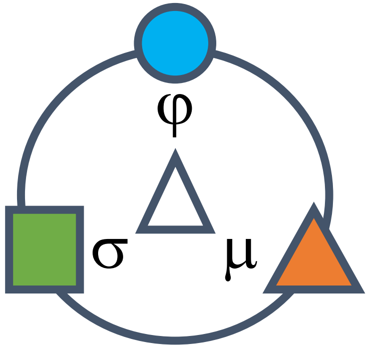
pp. 711-713*:

« ...radiance error magnification problem for the two-cell configuration, and highly nonuniform optical paths in general, is one of the biggest problems remaining in theoretical band model developments.

(...)

No solution to this problem is likely to be found STRICTLY within the framework of band models. »

Is it possible to propose a solution based on an appropriate combination of physics (φ), statistics (σ) and statistical learning (μ)?



* **Band model theory of radiation transport**, The Aerospace Press, 2013. ISBN: 978-1-884989-25-4.

INTRODUCTION (5/7)

For this purpose, we propose in this work to use the formalism introduced in the ℓ-distribution approach which is founded on the following property:

$$\frac{1}{\Delta\nu}\cdot\int_{\Delta\nu}\exp\left(-\kappa_\nu L\right)d\nu=\mathbb{P}\left[\ell\left(\xi\right)>L\right]=\int_0^1 H\left[\ell\left(\xi\right)-L\right]d\xi$$

$$\tau^{\Delta\nu}\left[\ell\left(\xi\right)\right]=\xi,\ \ \xi\in\left[0,1\right]$$

Used as the definition of ℓ

Path sampling strategies in Monte Carlo method are founded on the same result. But it has not been used apparently as the buidling block of methods other than Monte Carlo: ℓ-distribution theory fills this gap.

INTRODUCTION (6/7)

Modeling the inverses ℓ of transmissivities τ can be tricky*. However, their combination takes in some cases very simple forms:

$$\text{state 1} = \text{state 2: } \ell_1 \circ \tau_1^{\Delta\nu}\left(L\right) = L$$

$$\text{Scaled spectra } \kappa_{\nu,2} = u \cdot \kappa_{\nu,1} \text{ where } u \text{ is a constant: } \ell_1 \circ \tau_2^{\Delta\nu}\left(L\right) = u \cdot L$$

Question: what is the form of the function when (**definition of quasi-scaled spectra**)?

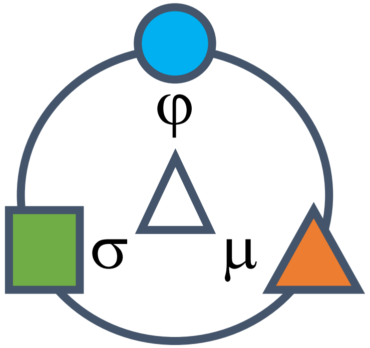
$$\kappa_{\nu,2} = u_\nu \cdot \kappa_{\nu,1} \text{ where: } \begin{cases} u_\nu \text{ and } \kappa_{\nu,1} \text{ are statistically independent} \\ 0 \le u_\nu - u_{\min} << u_{\min} \end{cases}$$

* **The L-distribution Method for modeling Non-Gray absorption in uniform and non-uniform gaseous media** March 2016 Journal of Quantitative Spectroscopy and Radiative Transfer 179
DOI: 10.1016/j.jqsrt.2016.02.034

INTRODUCTION (7/7) – Main result, where {1, 2} is a couple of states.

Lévy - Khintchine representation of $\ell_1 \circ \tau_2^{\Delta\nu}$

$$\ell_1 \circ \tau_2^{\Delta\nu}\left(L\right) \approx u_{\min} L + \int_0^1 \frac{1 - \exp\left[-s\left(0\right)\cdot v\left(\xi\right) L\right]}{s\left(0\right)} d\xi$$



●	φ	Physical component of the model
🟩	σ	Statistical component of the model
▲	μ	Statistical Learning component of the model

OUTLINE OF THE PRESENTATION

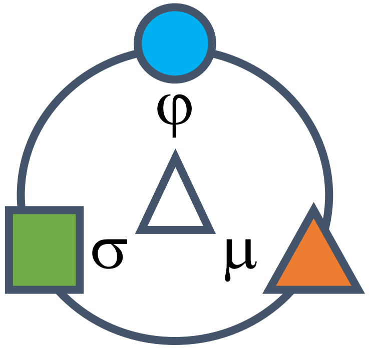
~~*I. The physical component of the model: the s function*~~ ●

II. Analysis of the two-layer problem ●

III. Generalization (ℓ-distribution and LBL training processes) ▲

III'. Case of the 3MI 7 O_2 A-band (C. Delage)

CONCLUSION

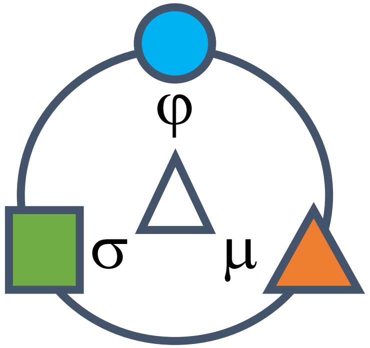


Comment: I have decided not to talk about the statistical component that relates to copula theory, due to limited time, but additional slides are available if you have questions. ■

I. The physical component of the model: the *s* function (1/5)

All (most of) the physics lies in the definition of $s(0)$ as:

$$s(0) = s(L=0) \text{ where } s(L) = \frac{\partial}{\partial L}\left(\ln\left[-\frac{1}{k_{P,2}} \frac{\partial \ln \tau_2^{\Delta\nu}(L)}{\partial L} \right] \right)$$



●	φ	Physical component of the model
■	σ	Statistical component of the model
▲	μ	Statistical Learning component of the model

II. Analysis of the two-layer problem (1/6)

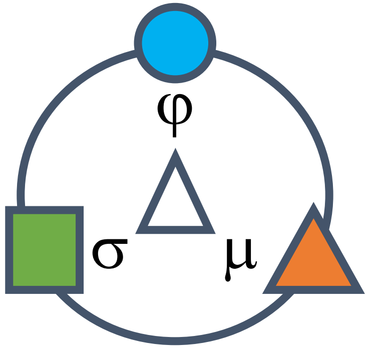
3 equivalent formulations (where $u_{\min}$, $\bar{u}$ and $v(\xi) = u(\xi)$-$u_{\min}$ are « scaling » coefficients):

$$\ell_1 \circ \tau_2^{\Delta v}(L) \approx u_{\min} L + \int_0^1 \frac{1 - \exp\left[-s(0)\cdot v(\xi) L\right]}{s(0)} d\xi$$

$$\ell_1 \circ \tau_2^{\Delta v}(L) \approx u_{\min} L + \left(\bar{u} - u_{\min}\right)\cdot \int_0^1 \frac{1 - \exp\left[-s(0)\cdot v^*(\xi) L\right]}{s(0)\cdot v^*(\xi)} d\xi$$

$$\ell_1 \circ \tau_2^{\Delta v}(L) = u(L)\cdot L$$

$$u(L) \approx u_{\min} + \left(\bar{u} - u_{\min}\right)\cdot \int_0^1 \int_0^1 \exp\left[-s(0)\cdot v(\xi) t\, L\right] dt d\xi$$



II. Analysis of the two-layer problem (2/6)

Step 1. integral is written in a discrete form:

$$\ell_1 \circ \tau_2^{\Delta\nu}(L) \approx u_{\min} L + \left(\bar{u} - u_{\min}\right) \cdot \int_0^1 \frac{1 - \exp\left[-s(0) \cdot v^*(\xi) L\right]}{s(0) \cdot v^*(\xi)} d\xi$$

$$\ell_1 \circ \tau_2^{\Delta\nu}(L) \approx u_{\min} L + \left(\bar{u} - u_{\min}\right) \cdot \sum_{i=1}^{N} \frac{\omega_i}{s(0) \cdot v^*(x_i)} \cdot \left(1 - \exp\left[-s(0) \cdot v^*(x_i) L\right]\right)$$

x_i and ω_i are the nodes and weights of a Gauss-Legendre quadrature over $[0,1]$.

The choice of this formulation for training is due to the existence of a simple method to initialize the model's coefficients (André et al, JQSRT, 2022).

II. Analysis of the two-layer problem (3/6)

Step 2. a loss function $\mathscr{L}$ is defined:

$$\mathscr{L} = \mathscr{L}\left[u_{\min}, v^*(x_1),..,v^*(x_N)\right] = \int_0^{+\infty} \left[\tau_2^{\Delta\nu}(L) - \tau_1^{\Delta\nu}\left(\ell_1 \circ \tau_2^{\Delta\nu}(L)\right)\right]^2 \, d\tau_2^{\Delta\nu}(L)$$

$$\ell_1 \circ \tau_2^{\Delta\nu}(L) \approx u_{\min} L + \left(\bar{u} - u_{\min}\right) \cdot \sum_{i=1}^{N} \frac{\omega_i}{s(0)\cdot v^*(x_i)} \cdot \left(1 - \exp\left[-s(0)\cdot v^*(x_i) L\right]\right)$$

x_i and ω_i are the nodes and weights of a Gauss-Legendre quadrature over $[0,1]$.

*__On the use of recurrent neural networks for fast and accurate non-uniform gas radiation modeling__ September 2022 Journal of Quantitative Spectroscopy and Radiative Transfer 293:108371
DOI: 10.1016/j.jqsrt.2022.108371

II. Analysis of the two-layer problem (4/6)

Step 3. the loss function $\mathcal{L}$ is discretized:

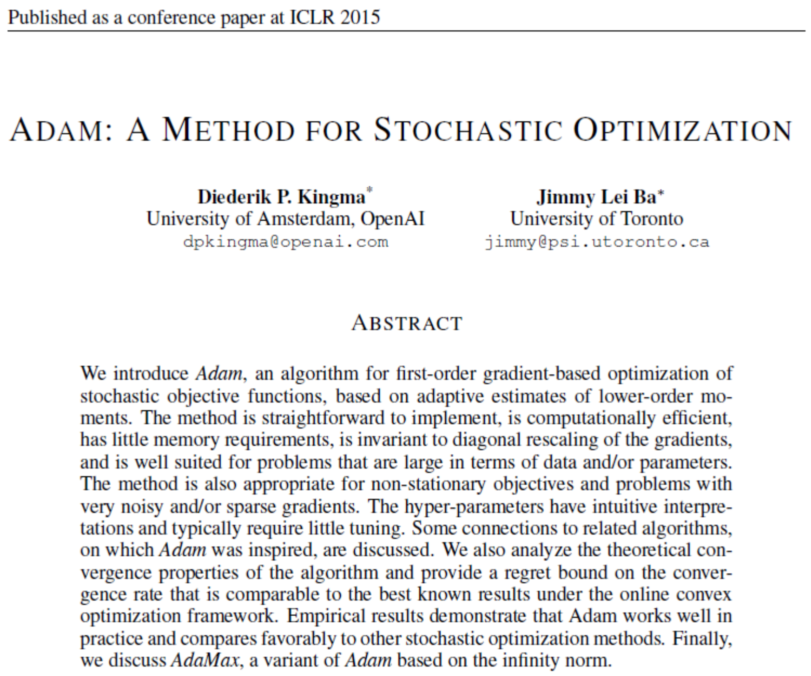
$$\mathcal{L} = \mathcal{L}\left[u_{\min}, v^*\left(x_1\right), .., v^*\left(x_N\right)\right] = \frac{1}{P}\cdot\sum_{p=1}^{P}\left[Y_p - \tau_1^{\Delta\nu}\left(\ell_1 \circ \tau_2^{\Delta\nu}\left(L_p\right)\right)\right]^2$$

$$Y_p = \frac{p}{P},\ \ L_p = \ell_2\left(Y_p\right) \text{ where } \ell_2 \circ \tau_2^{\Delta\nu}\left(L\right) = L$$

Important comment: in the logic of our methodology, the proposed developments come after an evaluation of both a CKD model and a « standard » ℓ-distribution model. The **corresponding model parameters** are thus known but the models are not considered accurate enough to justify an improvement stage (described in this work).

II. Analysis of the two-layer problem (5/6)

Step 4. the loss function $\mathscr{L}$ is minimized (using explicit gradients).



Published as a conference paper at ICLR 2015

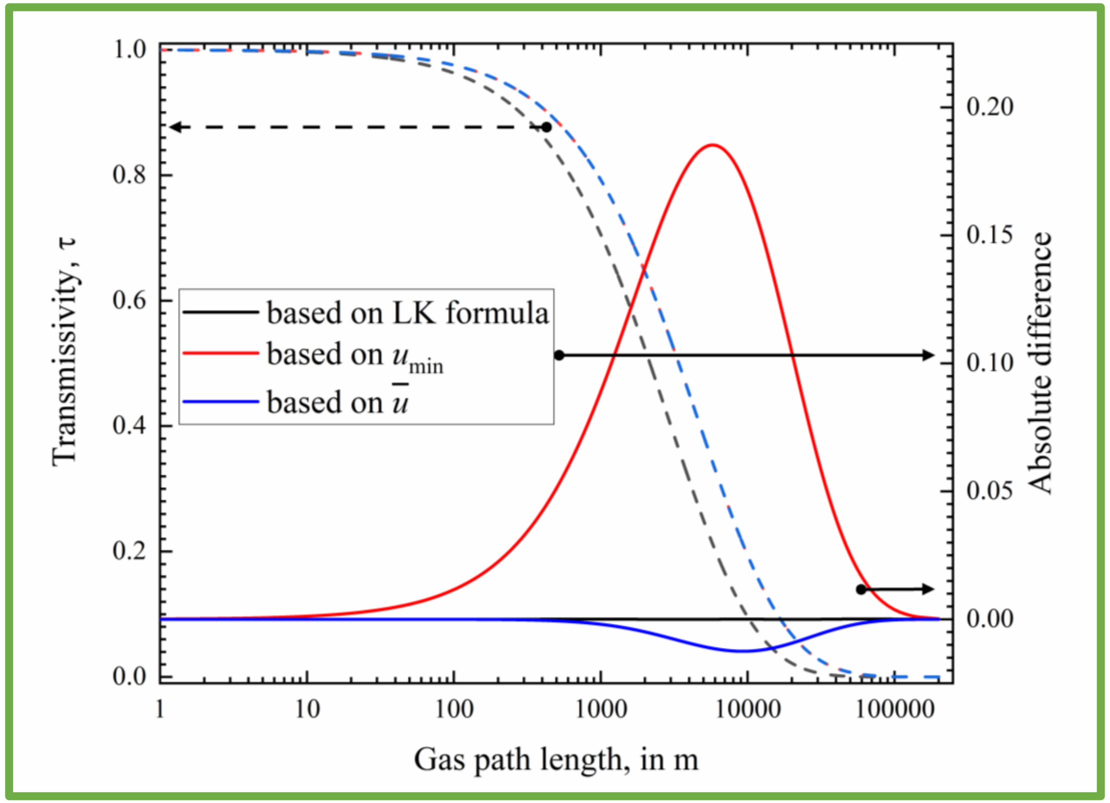
ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION

Diederik P. Kingma[*]
University of Amsterdam, OpenAI
dpkingma@openai.com

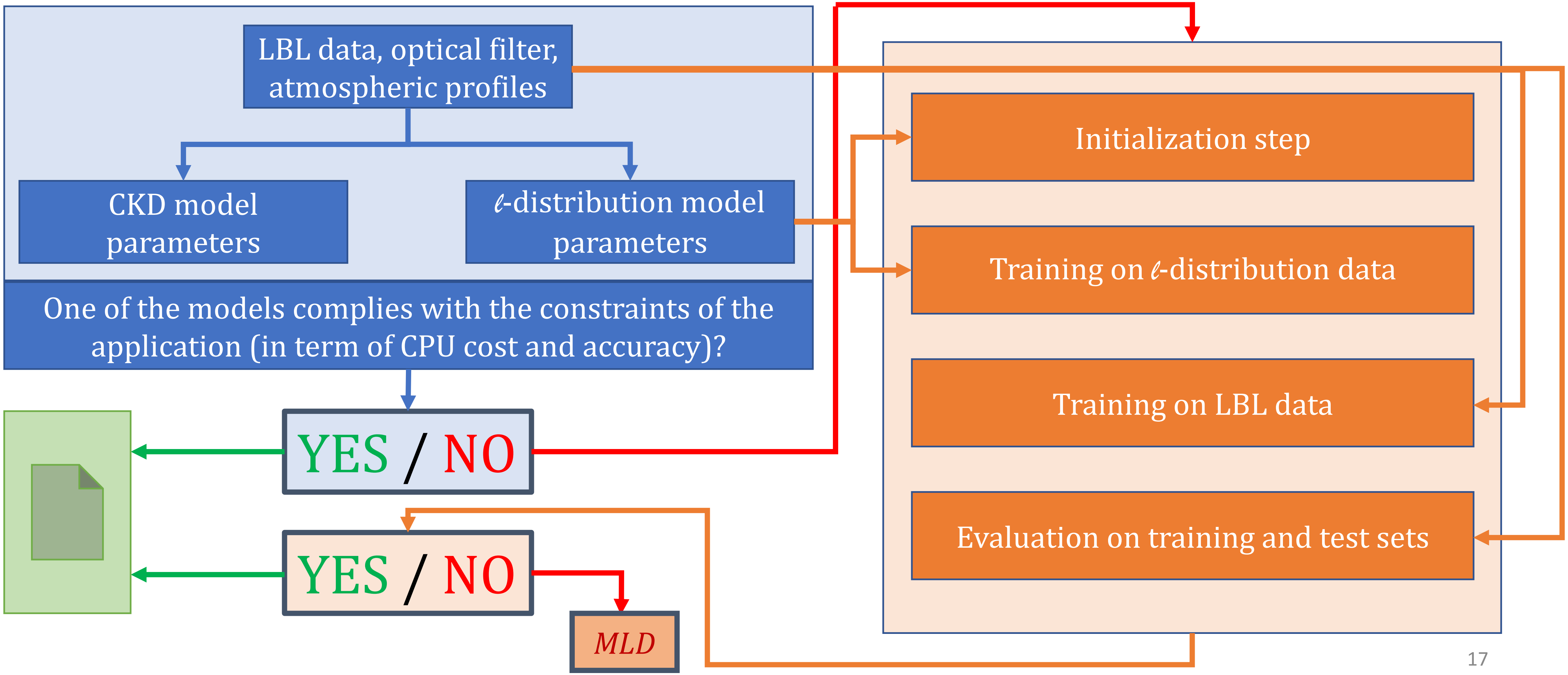
Jimmy Lei Ba[*]
University of Toronto
jimmy@psi.utoronto.ca

ABSTRACT

We introduce *Adam*, an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments. The method is straightforward to implement, is computationally efficient, has little memory requirements, is invariant to diagonal rescaling of the gradients, and is well suited for problems that are large in terms of data and/or parameters. The method is also appropriate for non-stationary objectives and problems with very noisy and/or sparse gradients. The hyper-parameters have intuitive interpretations and typically require little tuning. Some connections to related algorithms, on which *Adam* was inspired, are discussed. We also analyze the theoretical convergence properties of the algorithm and provide a regret bound on the convergence rate that is comparable to the best known results under the online convex optimization framework. Empirical results demonstrate that Adam works well in practice and compares favorably to other stochastic optimization methods. Finally, we discuss *AdaMax*, a variant of *Adam* based on the infinity norm.



III. Generalization (ℓ-distribution and LBL training processes)



LBL data, optical filter, atmospheric profiles

CKD model parameters

ℓ-distribution model parameters

One of the models complies with the constraints of the application (in term of CPU cost and accuracy)?

YES / NO

Initialization step

Training on ℓ-distribution data

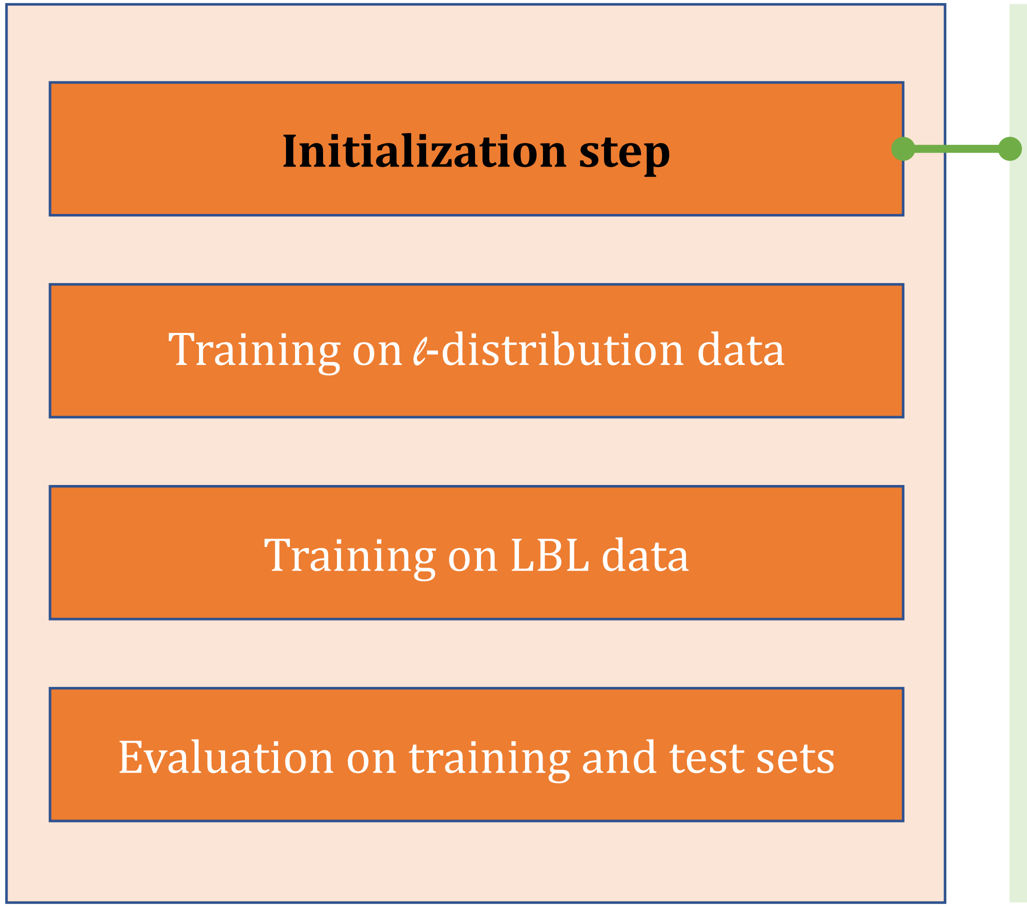
Training on LBL data

Evaluation on training and test sets

YES / NO

MLD

III. Generalization (ℓ-distribution and LBL training processes) (1/6)



Initialization step

Training on ℓ-distribution data

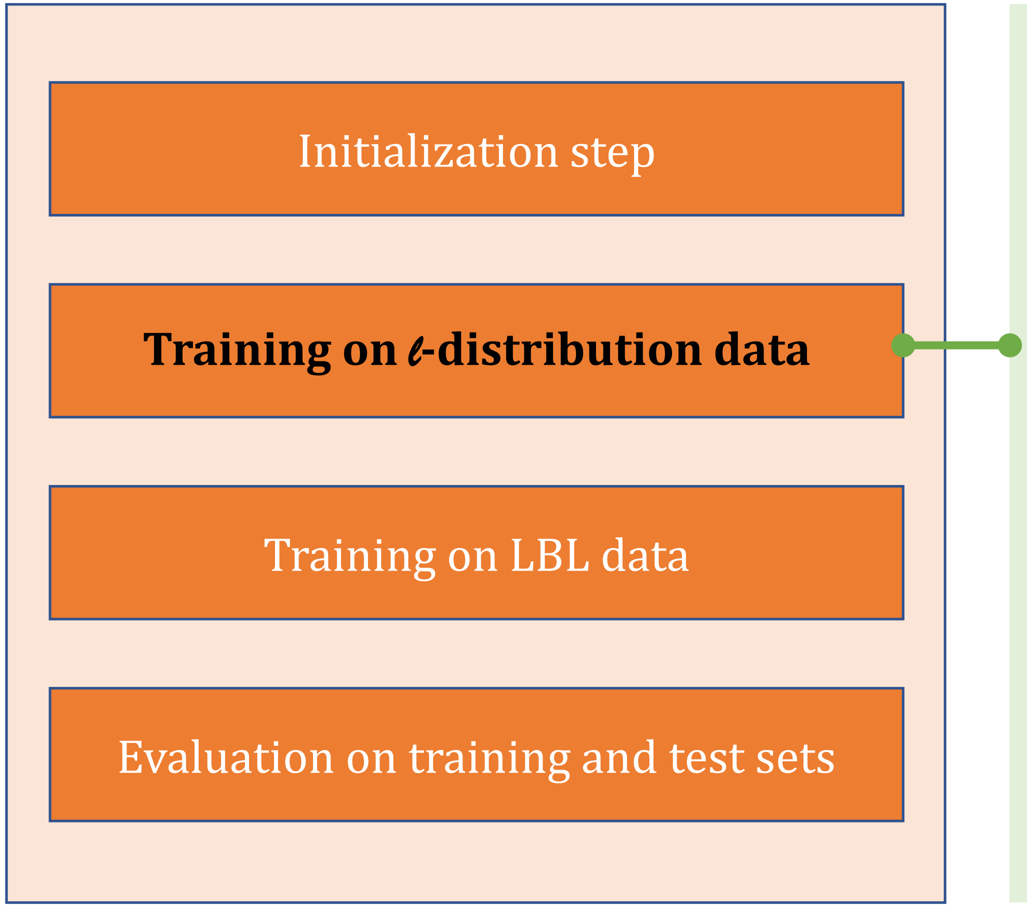
Training on LBL data

Evaluation on training and test sets

The first step consists of an analysis of two-layers combinations, as discussed previously.

As already noticed, the use of results from (André et al, JQSRT, 2022) allows simplifying the initialization of the **regression process on two-layers systems**.

III. Generalization (ℓ-distribution and LBL training processes) (2/6)



Initialization step

Training on ℓ-distribution data

Training on LBL data

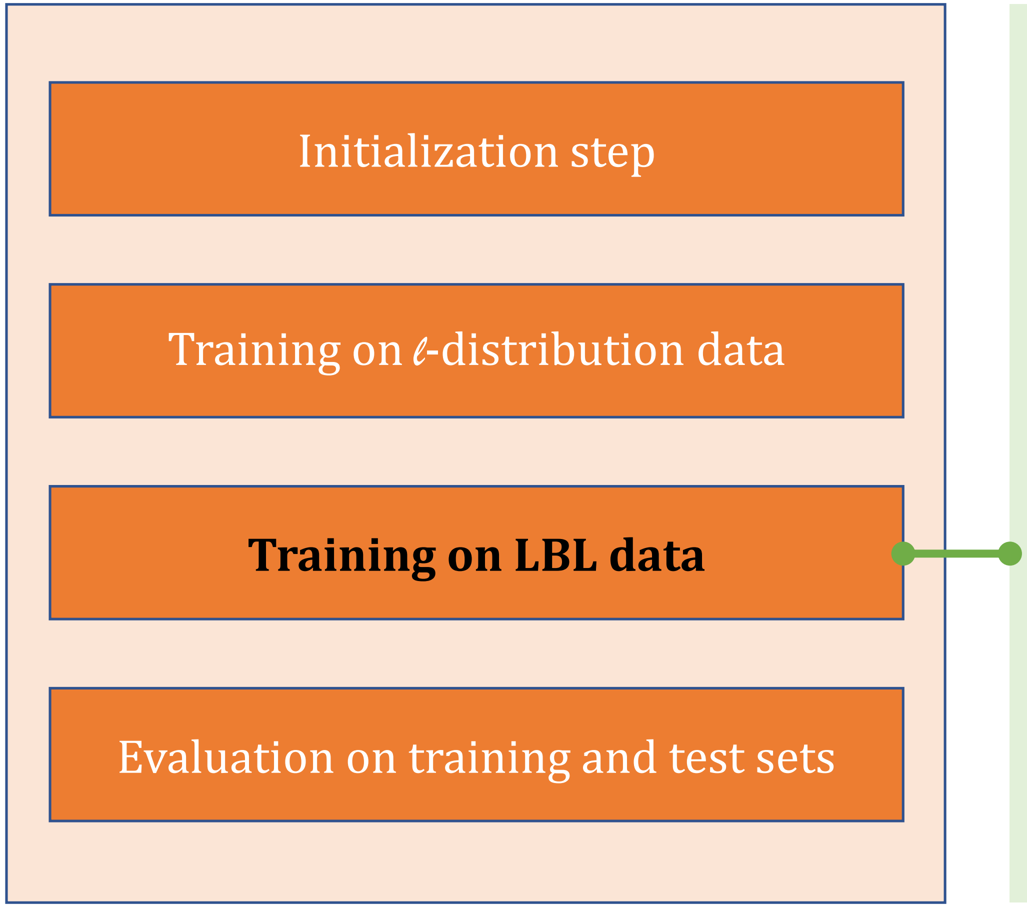
Evaluation on training and test sets

The second step consists of an analysis of the non-uniform ℓ-distribution solution.

In order to gain CPU time, the parameters for the two layer problems are not fully optimized (only improved compared to initialization).

This stage is used to correct partially this problem (**case dependent**). **The full atmosphere is considered for the training**.

III. Generalization (ℓ-distribution and LBL training processes) (3/6)



Initialization step

Training on ℓ-distribution data

Training on LBL data

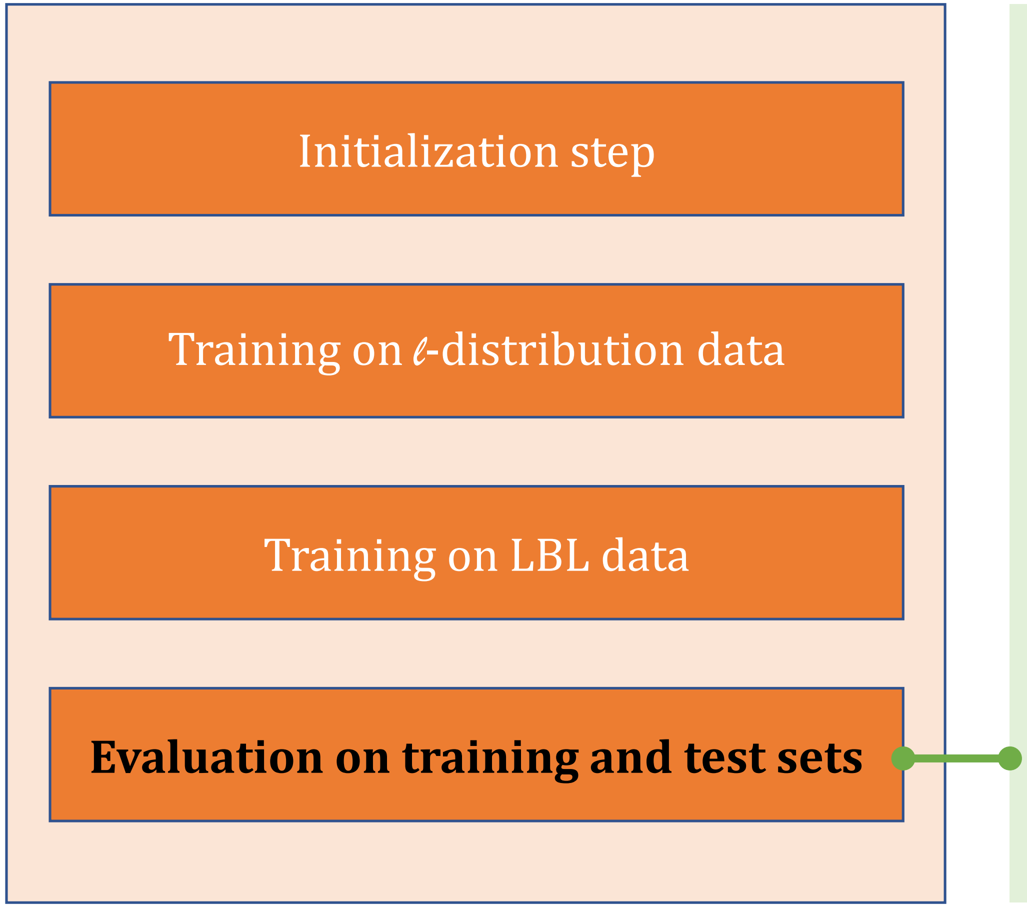
Evaluation on training and test sets

The third step consists of an analysis of non-uniform LBL solutions (**case dependent**).

This step is roughly the same as the previous one but consists of an improvement of the coefficients due to an **adjustment on LBL solutions**.

The full atmosphere is considered for the training.

III. Generalization (ℓ-distribution and LBL training processes) (4/6)



Initialization step

Training on ℓ-distribution data

Training on LBL data

Evaluation on training and test sets

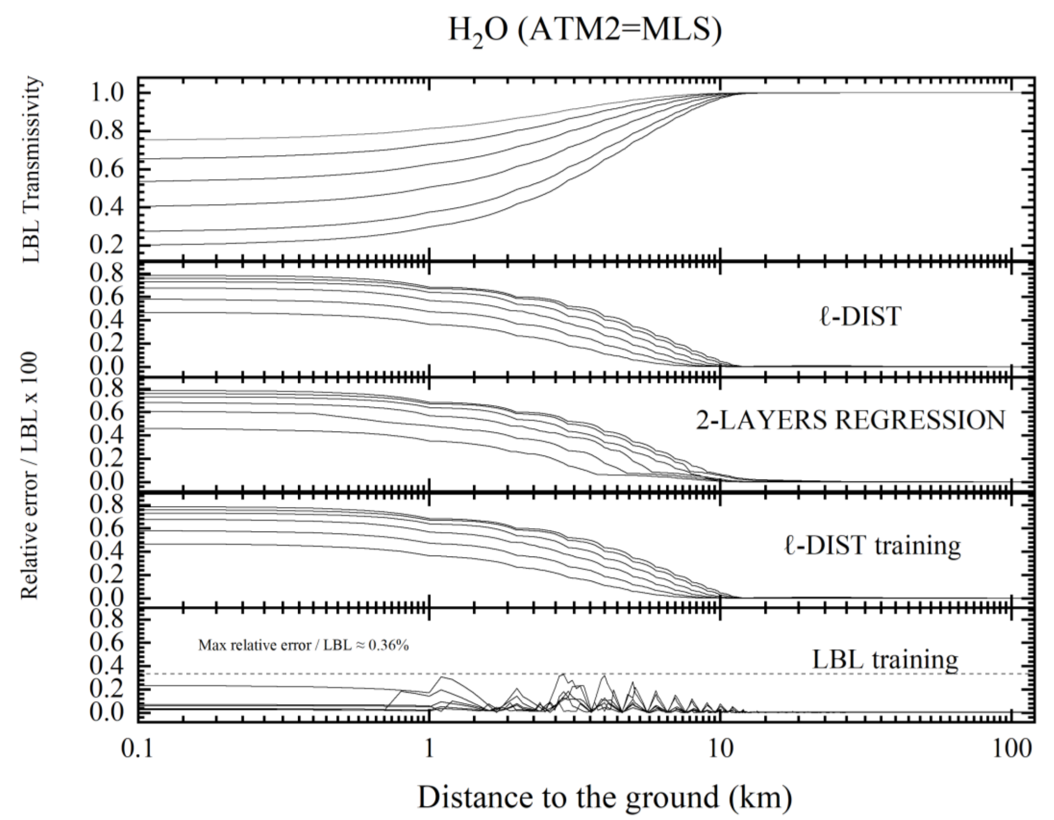
The fourth step consists of a **set of tests**.

It can be followed by an **update of the mapping functions** in the standard ℓ-distribution formalism **to minimize the CPU cost of the method**.

New mapping functions are then constructed with the help of the optimized solutions obtained after LBL training.

The training set consists of full atmospheres for RAM = 1, 2, 4, 8, 19 and 24.
The test set adds RAM = 3, 5, 6, 7, 10, 12, 14 and 20.

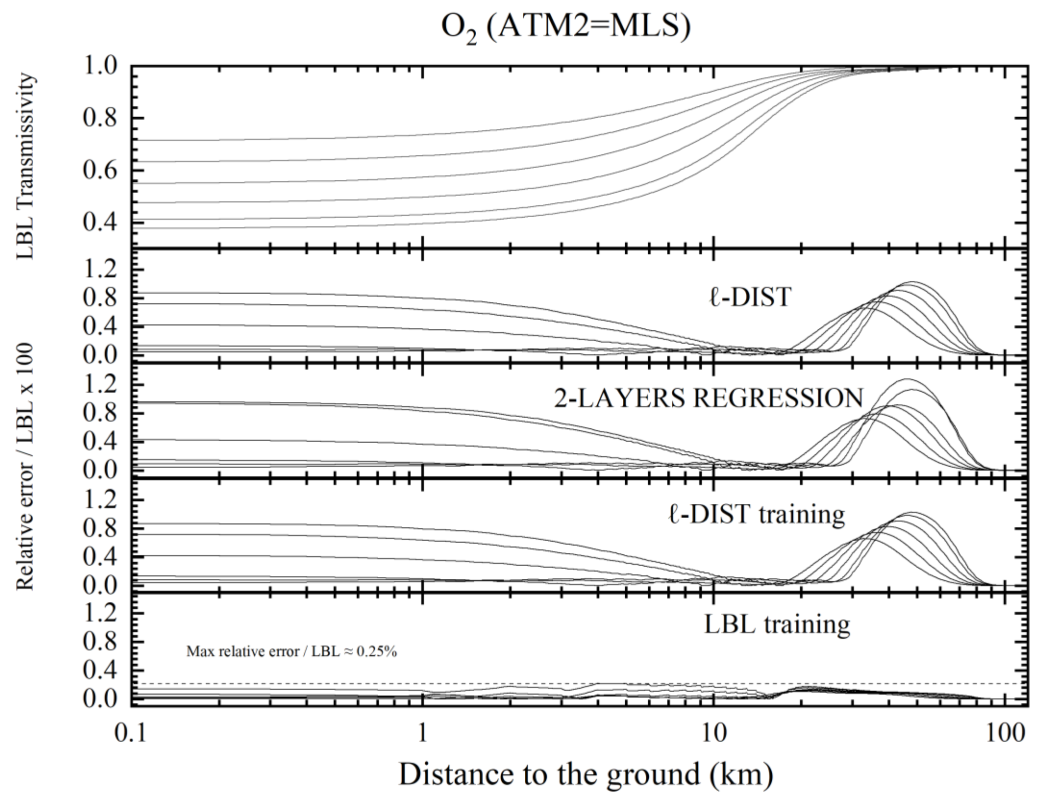
III. Generalization (ℓ-distribution and LBL training processes) (5/6)



This case corresponds to the one **treated in the RAD paper** (limited to the first two steps).

Clearly, adding a LBL training stage (lowest plot) allows improving significantly the accuracy of the method (at fixed CPU cost – see “poster”).

III. Generalization (ℓ-distribution and LBL training processes) (6/6)

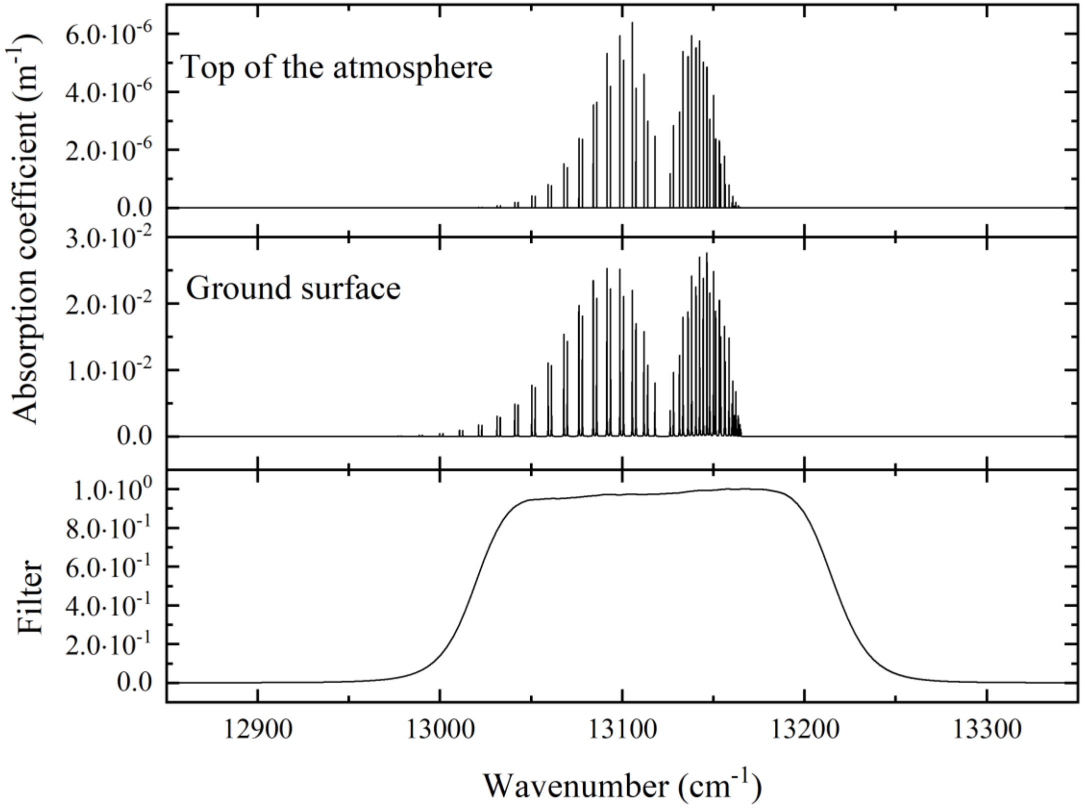


This case is treated in depth with additional technical details hereafter.

Clearly, adding a LBL training stage (lowest plot) allows improving significantly the accuracy of the method (at fixed CPU cost – see poster).

III'. Case of the 3MI 7 O_2 A-band (A/G)

Figure: O_2 Midlatitude summer (MLS) Profile

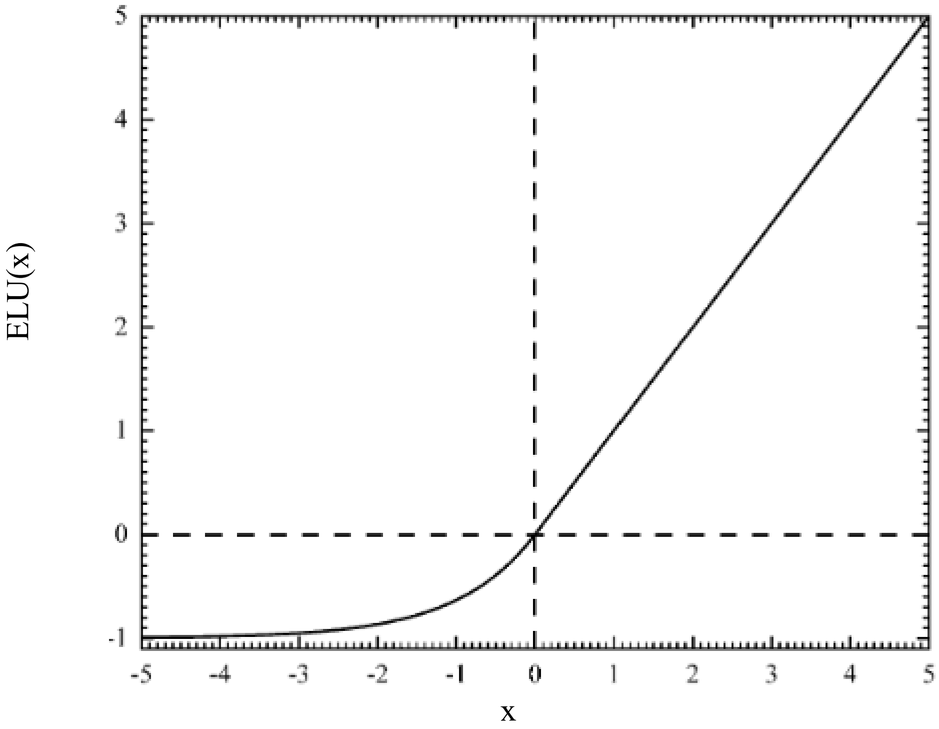


We focus here on one particular application test, widely used in cloudy atmospheres: the **O_2 A-band.** Channel 7 of the **3MI instrument** is treated. This channel is designed to study, among other, cloud top heights.

This case illustrates the **main steps** of the methodology.

III'. Case of the 3MI 7 O_2 A-band (B/G)

Figure: The Exponential Linear Unit (ELU)



The **ELU (Exponential Linear Unit)** is rather natural in our case since it appears in the Lévy-Kintchine formula.

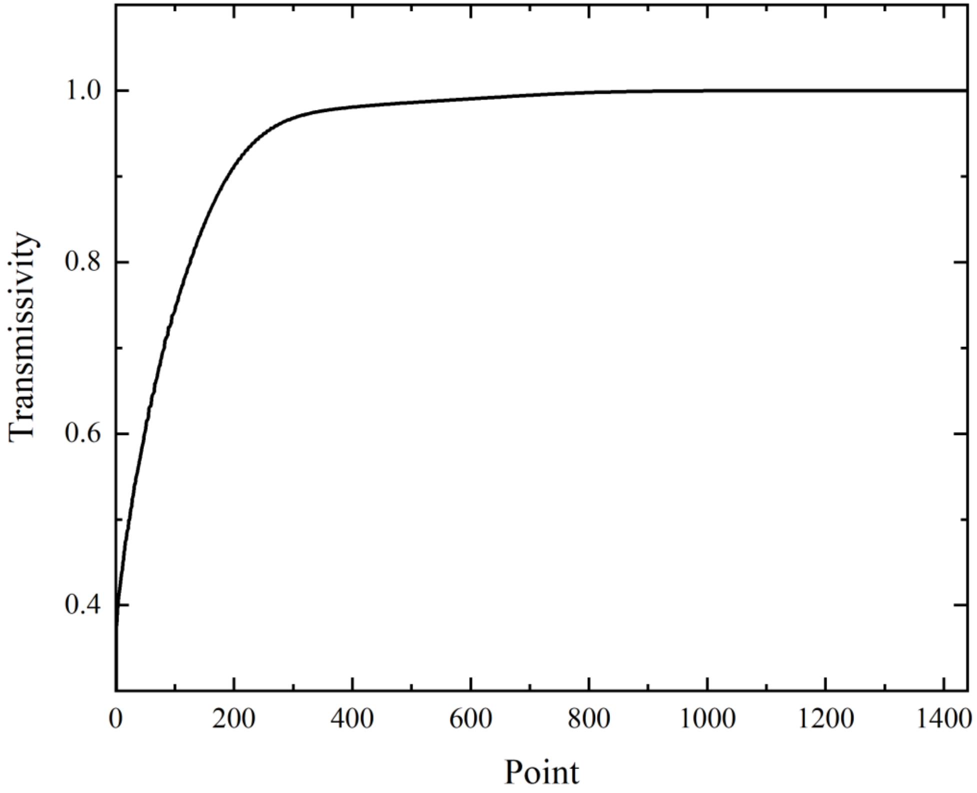
$$\varphi_\alpha(x) = \begin{cases} x \text{ if } x > 0 \\ \alpha\left(\exp(x) - 1\right) \text{ if } x \leq 0 \end{cases}$$

A **recurrent network** is also natural considering the propagative scheme used in ℓ-distribution model.

$$\begin{cases} L_{nn} = L_n \\ L_{i..n} = L_i + \ell_i \circ \tau_{i+1}^{\Delta\nu}\left(L_{i+1..n}\right) \end{cases}$$

III'. Case of the 3MI 7 O_2 A-band (C/G)

Figure: Training data for the O_2 A-band (MLS) Channel 7 3MI



The training process is made using ℓ-distribution (2nd step) and LBL (3rd step) data.

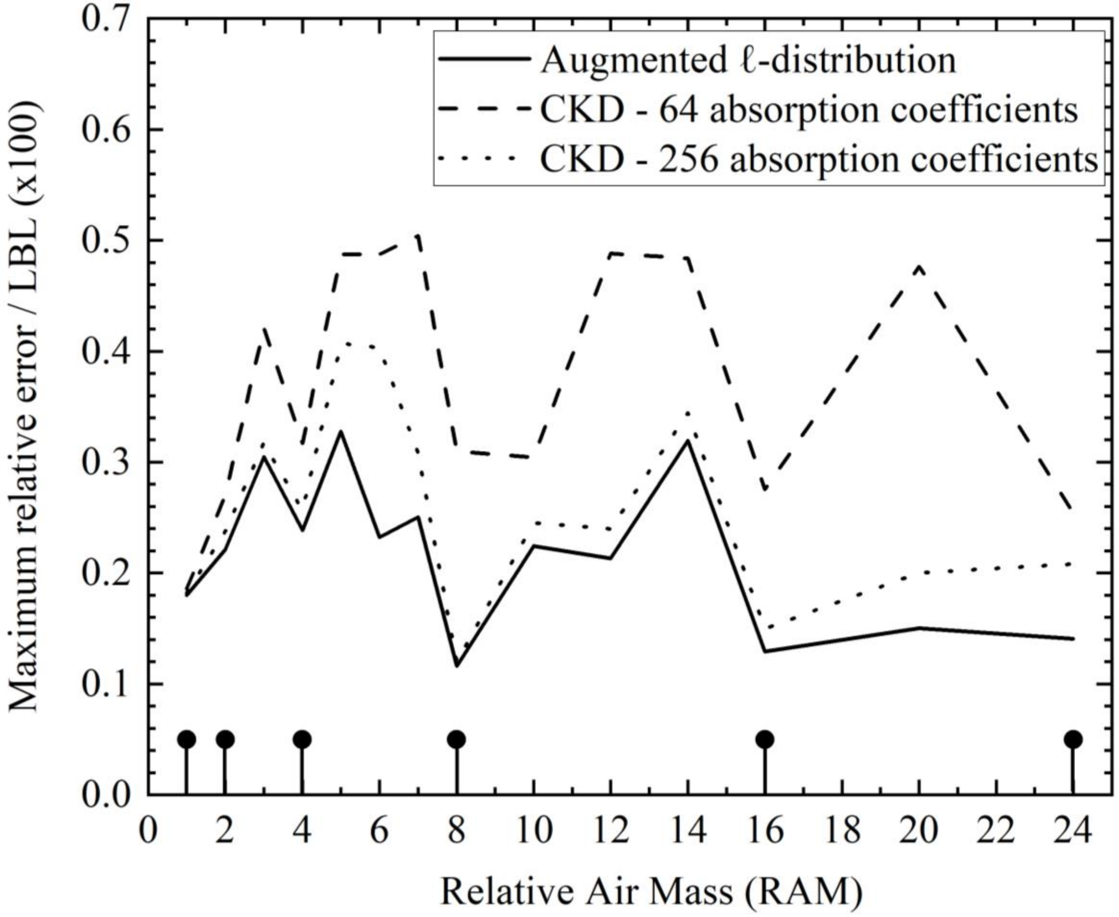
RAM 1, 2, 4, 8, 16, 24 are used as inputs for the training process.

For this purpose, transmissivities are plotted with respect to « Points », refering to data obtained through a combination of lengths and RAMs.

In order to reduce memory and computationnal costs, **one point every 500 meters** is used for the training process.

III'. Case of the 3MI 7 O_2 A-band (D/G)

Figure: O_2 A-Band, Augmented ℓ-distribution

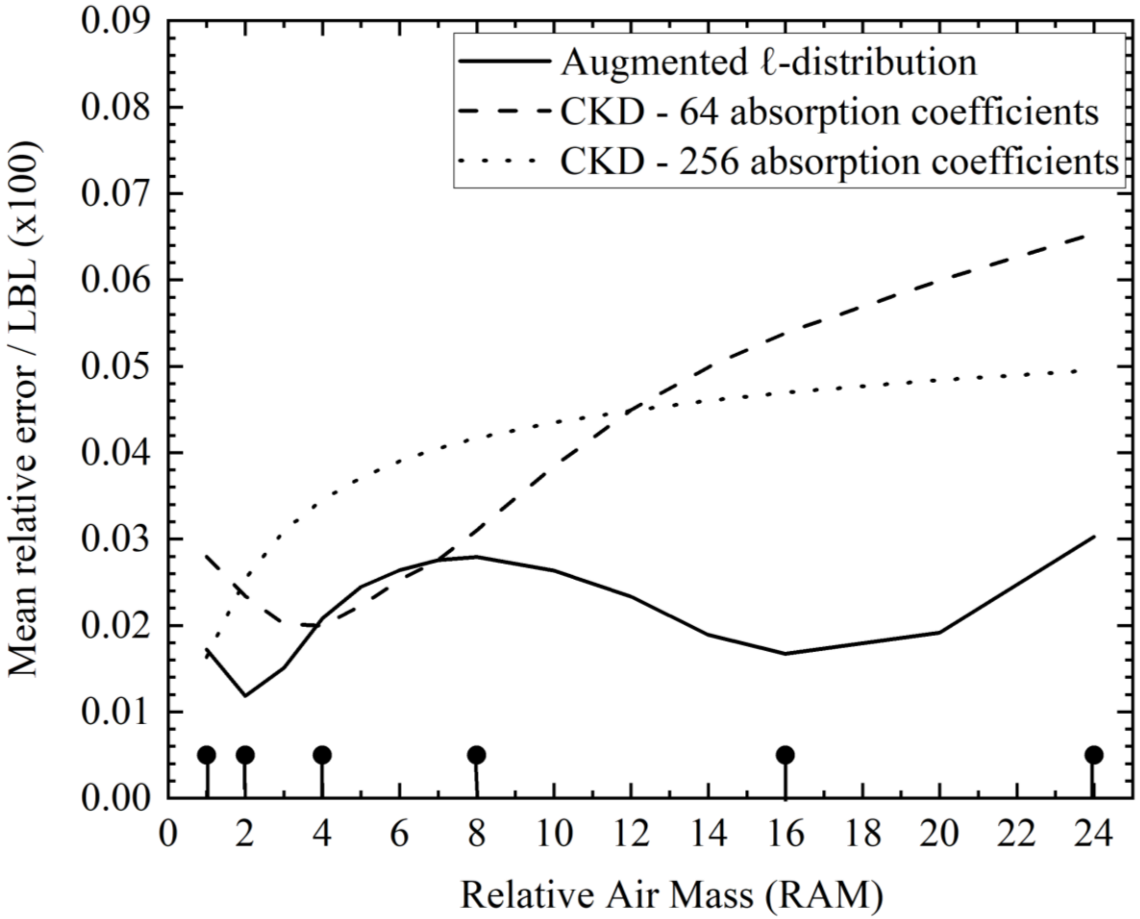


The main purpose of the Augmented ℓ-distribution is to **reduce maximum relative errors of the ℓ-distribution model, at fixed CPU cost**.

In the case of the O_2 A-Band, maximum relative errors for the standard ℓ-distribution model is about 0.75%. With augmented model, this error can be **divided by at least two**.

III'. Case of the 3MI 7 O_2 A-band (E/G)

Figure: O_2 A-Band, Augmented ℓ-distribution



Since the Augmented ℓ-distribution are based on a minimisation of a loss function writen in a summation or mean form, its main result is the **reduction of the mean relative error** of the model.

In the case of the O_2 A-Band, mean relative errors for the standard ℓ-distribution model is between 0.1 and 0.2%. With augmented version, this error can be **divided by a factor from three up to ten**.

III’. Case of the 3MI 7 O_2 A-band (F/G)

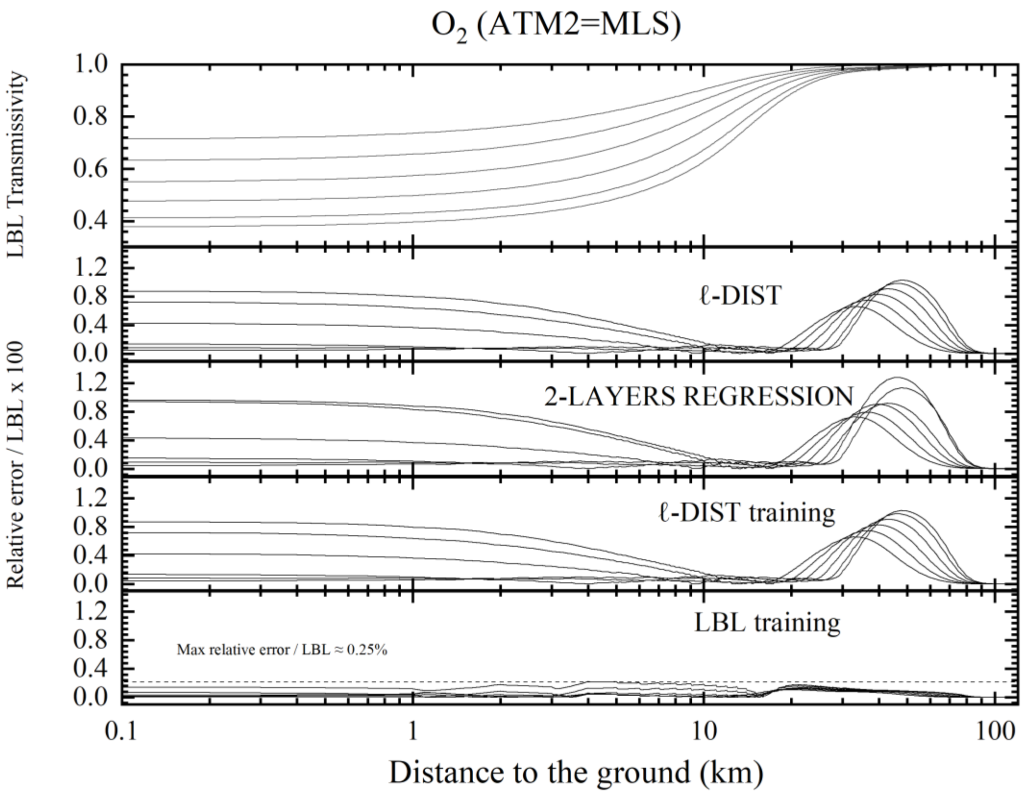


Figure : LBL training <span>without update of mapping functions</span>

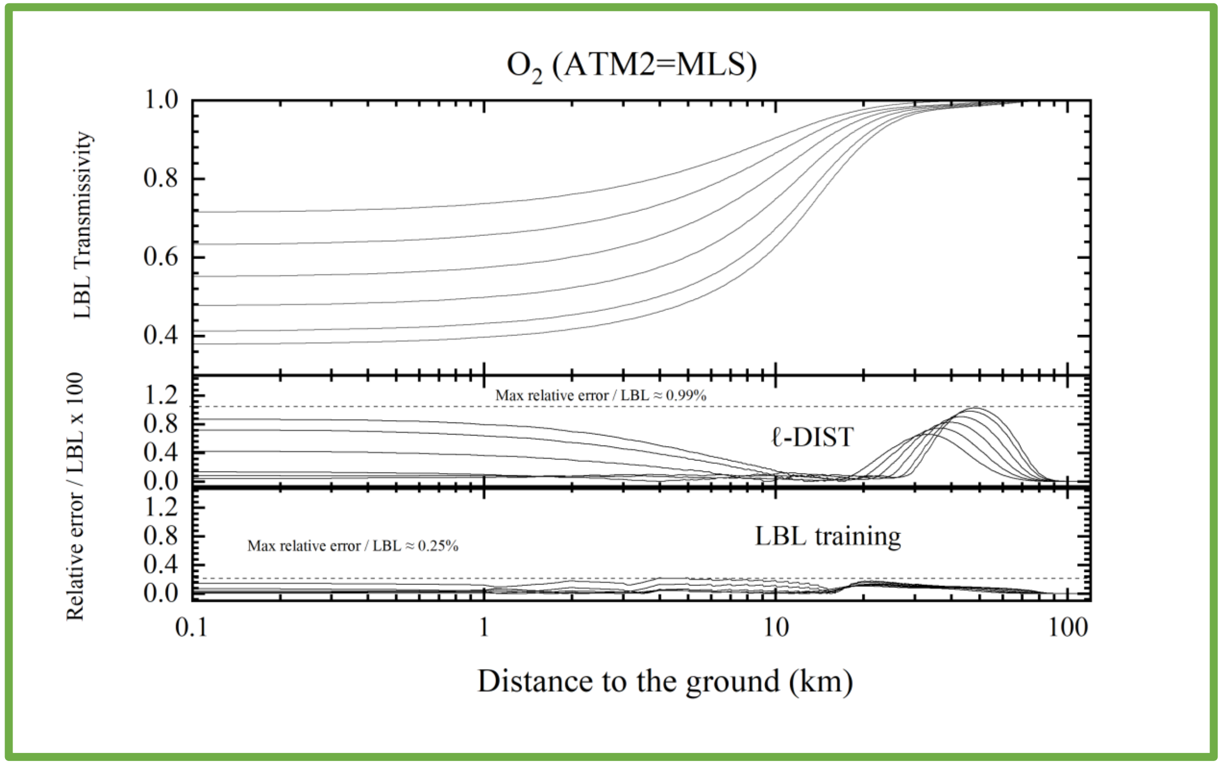


Figure : LBL training with update of mapping functions

III'. Case of the 3MI 7 O_2 A-band (G/G)

Once optimization is complete, the mapping functions are updated. This step ensures a **gain in terms of CPU cost** compared to the model trained on LBL data**, while preserving the accuracy** of the LK formulation.

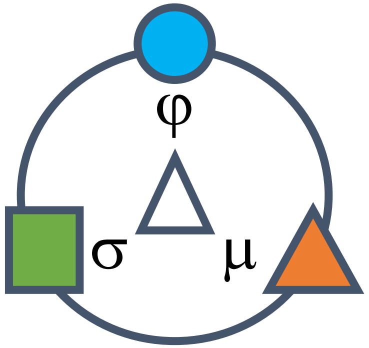
Original ℓ-distribution model **provides a result in 0,1 ms for a full atmosphere** calculation (1200 values of non-uniform path transmissivities).

Augmented ℓ-distribution leads to a **numerical gain of 1 % (not significant but the model remains highly competitive in terms of CPU cost)**.

CONCLUSION (1/3)

The present work is dedicated to the description of a method that combines physics (φ), statistics (σ) and statistical learning (μ) to produce accurate transmissivities of non-uniform atmospheric paths.

Each component of the model is equally important to the whole methodology, but is used to treat a distinct part of the model.



Lévy - Khintchine representation of $\ell_1 \circ \tau_2^{\Delta\nu}$

$$\ell_1 \circ \tau_2^{\Delta\nu}\left(L\right) \approx u_{\min} L + \int_0^1 \frac{1 - \exp\left[-s\left(0\right)\cdot v\left(\xi\right)L\right]}{s\left(0\right)} d\xi$$

CONCLUSION (2/3)

More than the result itself (it obviously works otherwise I would not be here today…), an **interesting part of the work is its philosophy**.

Indeed, **coupling band model theory with statistical learning** has required:

- Identifying **what are the key component(s)** in **band model theory** required to obtain a solution / bridge.
- Identifying in **band model theory, why it fails** at some point.
- **Once the diagnostics is made, delete / add components** to **treat the sources of failure** (this requires ensuring that what you add complies with existing theory i.e. that your new components contain existing methods as particular cases – this part is the most time consuming).
- **Use SL to learn the components you have added from LBL data**.

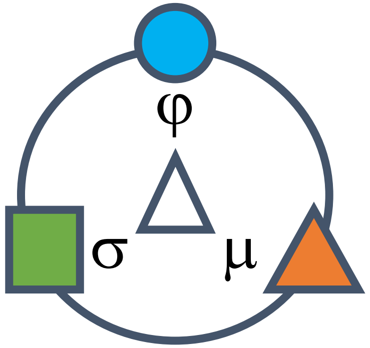
CONCLUSION (3/3)

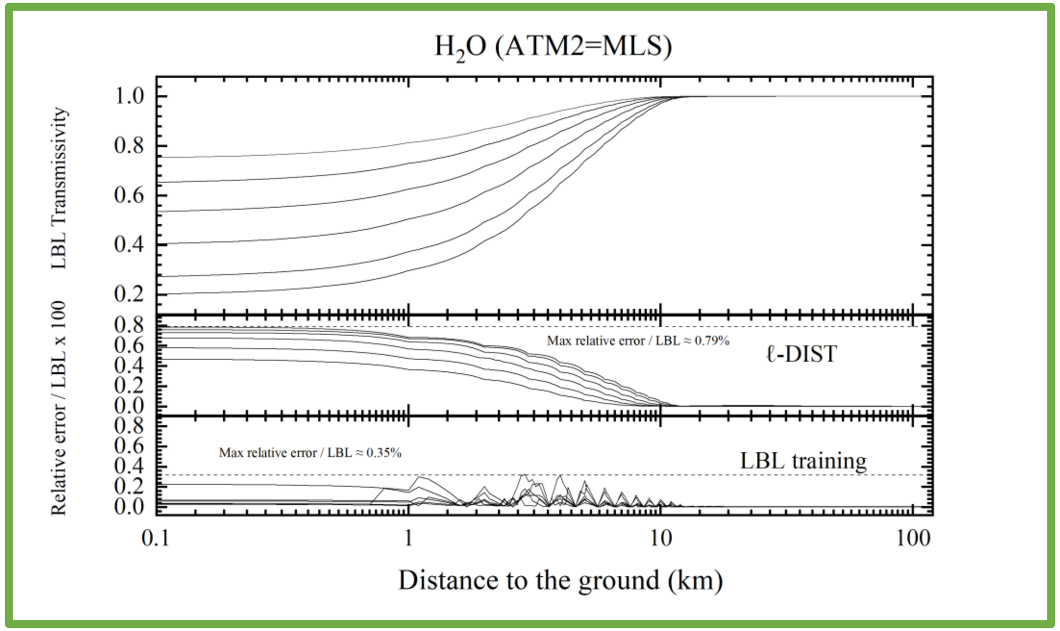
But this is however not the only interest:

- **First general solution of Godson / scaled-*k* implicit equation** (even if some restrictions due to quasi-scaling are added).
- **First formal proof that Godson's method (1953) actually provides relevant approximations of non-uniform path transmissivities** (up to know, only verifications).
- Indirectly, **may modernize the (mostly dying) field of band model theory**, by opening it to modern numerical methods.

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR
ANR

Thank you for your attention!





After update of the mapping functions, CPU cost is minimum (0.1 ms / atm) but accuracy is high!

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR
ANR

STATISTICAL INTERMEZZO (1/4)

🟩 σ

From a set of models for the transmissivities and their inverses, various methods can be proposed based on:

$$\tau^{\Delta\nu}\left(L_1,..,L_n\right) = \frac{1}{\Delta\nu}\cdot\int_{\Delta\nu}\exp\left(-\kappa_\nu^1 L_1 - .. - \kappa_\nu^n L_n\right)\ d\nu = C_{1..n}\left[\tau_1^{\Delta\nu}\left(L_1\right),..,\tau_n^{\Delta\nu}\left(L_n\right)\right]$$

where (notice that here we have an exact calculation):

$$C_{1..n}\left(X_1,..,X_n\right) = \frac{1}{\Delta\nu}\cdot\int_{\Delta\nu}\exp\left[-\kappa_\nu^1\ell_1\left(X_1\right) - .. - \kappa_\nu^n\ell_n\left(X_n\right)\right]\ d\nu$$

This function $C_{1..n}$ has some interesting properties

STATISTICAL INTERMEZZO(2/4)

$$\boxed{C_{1..n}\left(X_1,.,X_i=0,.,X_n\right)=0}$$

This property ensures the proper model behavior at the « ***optically thick limit »***.

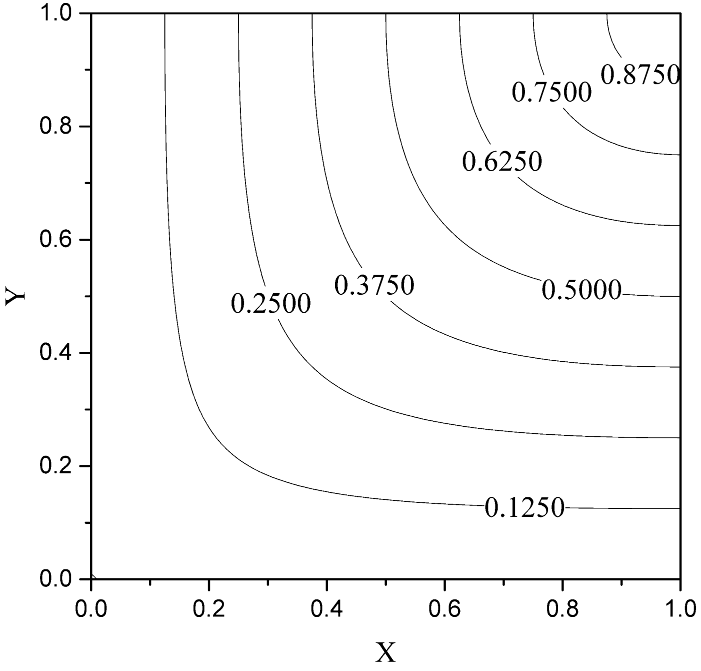
$$\boxed{C_{1..n}\left(1,1,.,X_i,.,1\right)=X_i}$$

This property ensures the proper model behavior at the « ***optically thin limit »***.

$$\boxed{\left(-1\right)^n\frac{\partial^n C_{1..n}\left(X_1,.,X_i,.,X_n\right)}{\partial X_1 \dots \partial X_n}\geq 0}$$

This property is closely related to the ***sign of net exchanges*** between distinct elements along a non-uniform path.

Function $C_{1..n}$ is mathematically called a **copula.**



$$C\left(X,Y\right)$$

More details in: F. André, C. Cornet, M. Galtier, Ph. Dubuisson, "Radiative transfer in the O_2 A-band – a fast and accurate forward model based on the ℓ-distribution approach", *J. Quant, Spectrosc. Radiat. Transfer*, vol. 260, 107470, 2020.

STATISTICAL INTERMEZZO (3/4)

In the ℓ-distribution method (the same technique is used here), the true copula is approximated by a hierarchical structure (called Archimedean, HAC):

$$\tau^{\Delta\nu}\left(L_1,..,L_n\right) \approx C_{11}\left[\tau_1^{\Delta\nu}\left(L_1\right), C_{22}\left(\tau_2^{\Delta\nu}\left(L_2\right), C_{33}\left(\tau_3^{\Delta\nu}\left(L_3\right),..\right)\right)\right]$$

where:

$$C_{ii}\left(X,Y\right) = \frac{1}{\Delta\nu}\cdot\int_{\Delta\nu}\exp\left[-\kappa_\nu^i\ell_i\left(X\right)-\kappa_\nu^i\ell_i\left(Y\right)\right]\ d\nu = \tau_i^{\Delta\nu}\left[\ell_i\left(X\right)+\ell_i\left(Y\right)\right]$$

It can be shown that the hierarchical structure is a copula (as the true one) if the generators of the HAC are compatible. This is the case if (sufficient nesting condition):

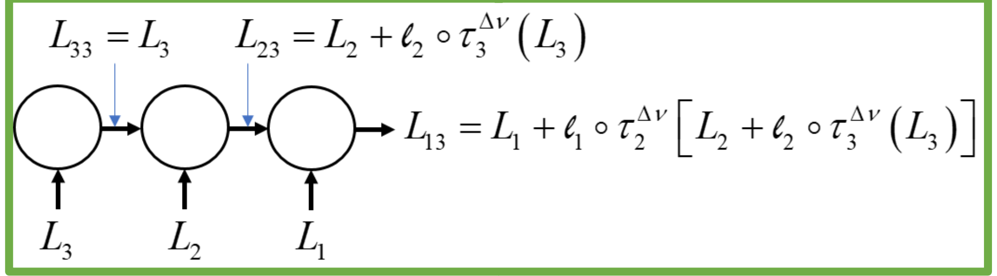
$$\ell_i\circ\tau_{i+1}^{\Delta\nu}\left(L\right) \approx u_{\min,i}L + \int_0^1 \frac{1-\exp\left[-s_{i+1}\left(0\right)\cdot v_{i,i+1}\left(\xi\right)L\right]}{s_{i+1}\left(0\right)}\,d\xi$$

The HAC is in this case called Lévy-subordinated (LS-HAC).

PROJET FINANCÉ PAR L'ANR
ANR
PROJECT FUNDED BY THE ANR

STATISTICAL INTERMEZZO (4/4)

The hierarchical structure of the previous slide can be equivalently formulated as a recurrent structure (similar to a recurrent neural network with ELU activation functions) – See André et al, JQSRT, 2022 for more details:



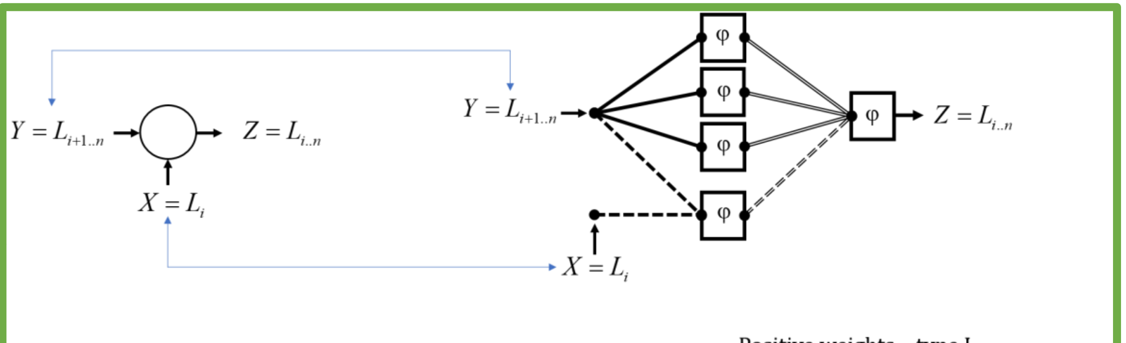
$L_{33} = L_3$ $\quad L_{23} = L_2 + \ell_2 \circ \tau_3^{\Delta\nu}\left(L_3\right)$

$L_{13} = L_1 + \ell_1 \circ \tau_2^{\Delta\nu}\left[L_2 + \ell_2 \circ \tau_3^{\Delta\nu}\left(L_3\right)\right]$

$$L_3 = \ell_3 \circ C_{33}\left[0, \tau_3^{\Delta\nu}\left(L_3\right)\right]$$

$$L_{23} = \ell_2 \circ C_{22}\left[\tau_2^{\Delta\nu}\left(L_2\right), \tau_3^{\Delta\nu}\left(L_3\right)\right] = L_2 + \ell_2 \circ \tau_3^{\Delta\nu}\left(L_3\right)$$

$$L_{13} = \ell_1 \circ C_{11}\left[\tau_1^{\Delta\nu}\left(L_1\right), \tau_2^{\Delta\nu}\left(L_{23}\right)\right] = L_1 + \ell_1 \circ \tau_2^{\Delta\nu}\left(L_{23}\right)$$

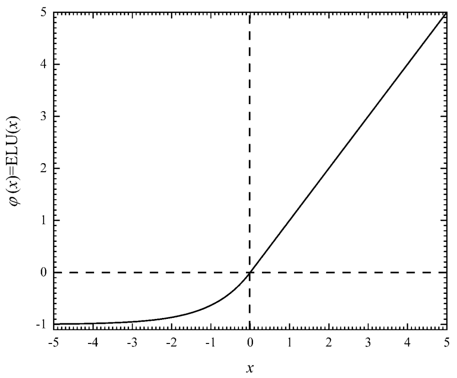


●- - - - - -● Positive weights – type I

●= = = = = =● Positive weights – type II

●———————● Negative weights – type III

●═══════● Positive or Negative weights – type IV



$$\varphi(x) = \begin{cases} x \text{ if } x \geq 0 \\ \exp(x) - 1 \text{ if } x \leq 0 \end{cases}$$

$$\ell_i \circ \tau_{i+1}^{\Delta\nu}\left(L\right) \approx u_{\min,i} L + \int_0^1 \frac{1 - \exp\left[-s_{i+1}(0) \cdot v_{i,i+1}(\xi) L\right]}{s_{i+1}(0)} d\xi$$

$$= \varphi\left(u_{\min,i} L\right) - \frac{1}{s_{i+1}(0)} \cdot \int_0^1 \varphi\left[-s_{i+1}(0) \cdot v_{i,i+1}(\xi) L\right] \, d\xi$$