A newly born TIGR data set to meet the requirements of the high spectral resolution instruments : the TIGR-2020

J. Pernin[1], A. Chédin[1], C. Crevoisier[1], V. Capelle[1], R. Armante[1], L. Crépeau[1], N. Scott[1], M. Vrac[2]

[1]LMD, IPSL, CNRS, Ecole Polytechnique, Université Paris-Saclay, 91128, Palaiseau, France
[2]LSCE, IPSL, CEA, CNRS, UVSQ, Centre d'Etudes de Saclay, Ormes des Merisiers, Bât.701, 91191, Gif-sur-Yvette, France

Jérôme PERNIN
jerome.pernin@lmd.ipsl.fr

INTRODUCTION

- **What are Thermodynamic Initial Guess Retrieval (TIGR) data sets ?**

Successive versions of **size-limited** collections of **3D representative atmospheric thermodynamic states**, resulted from adequate sampling methods and classified into **air mass classes**, each situation described by **T, H2O, O3 vertical profiles** and surface T/P.

Associated with these profiles are **Radiances, Transmittances, Jacobians** TIGR data sets specifying the situations that satellite-based instrument (e.g., TOVS, ATOVS, AIRS, IASI, Modis, Seviri, IIR) would have observed under various in-flight conditions.

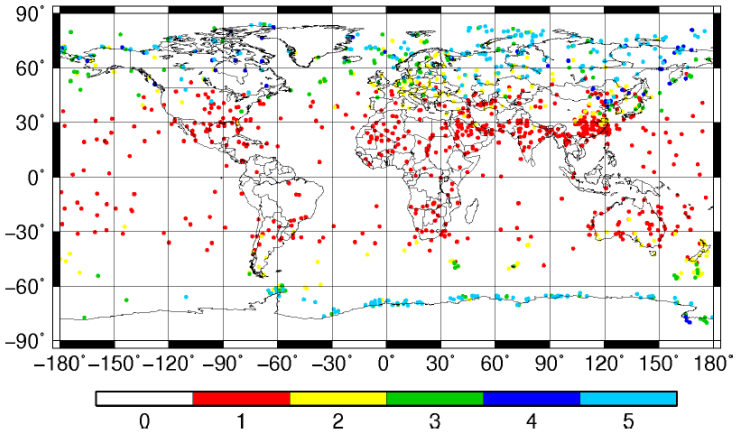
Widely used by our scientific community (300 versions distributed worldwide).

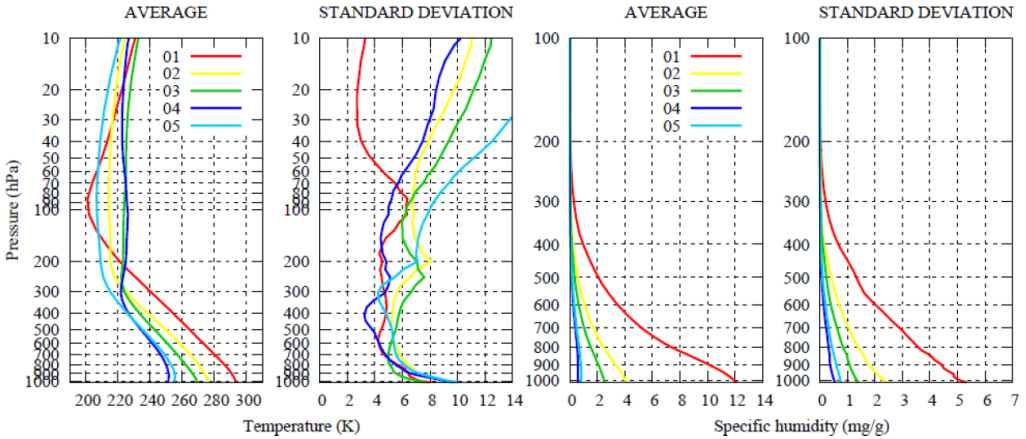
- **What for ?**

Used generally as pertinent **a priori information in inverse radiative transfer problems** (e.g., bayesian approach, training phase of Neural Networks) and **definition of space mission** (e.g., selection of channels, sensitivity studies).

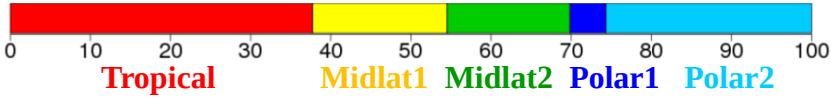
I – THE CURRENT TIGR 2000 AND ITS LIMITATIONS

Number of situations	2311	Number of airmass classes	5
Starting data sets to sample	NOAA/NESDIS, ECMWF for Europe, NCAR/ERL for Arctic, National Satellite Meteorological Center for China, New Zealand Meteorological Service for southern hemisphere + 420 000 profiles in January and July restituted via inversion.		
Vertical resolution of profiles	43 fixed pressure levels of 4A radiative transfer model : [1013 ; 0,0026] hPa		
Classification method	(T, Tv) Principal Component Analysis + Hierarchical Agglomerative Clustering (HAC) + subjective clustering into 5 classes		
Sampling method	Iterative accept/reject algorithm based on a metric estimating distance between profiles.		





Features / classes	1	2	3	4	5
Precipitable water (cm)	3.78 ± 1.5	1.22 ± 0.6	0.75 ± 0.36	0.24 ± 0.12	0.31 ± 0.20
[800-425] hPa average temperature (K)	249.4 ± 2.6	241.6 ± 5.7	242.8 ± 4.9	235.3 ± 4.2	230.9 ± 5.0
Surface temperature (K)	293.5 ± 8.0	276.6 ± 9.5	269.5 ± 7.7	251.2 ± 9.8	254.6 ± 10.0
Number of situations	872	388	354	104	593
Percentage of situations (%)	37.7	16.8	15.3	4.5	25.7



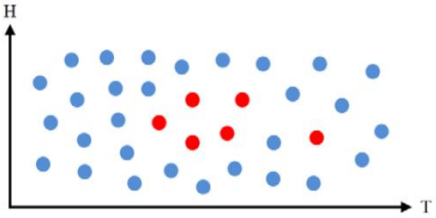
- **Limitations VS high vertical and spectral resolutions and low noise of recent sounding instruments**

(i) Thermodynamic state sampling
(ii) Thermodynamic profile vertical resolution
(iii) Air mass classification

II – HOW TO BUILD A NEW TIGR DATA SET ?

1. Choice of the starting data sets to sample

TIGR 2000	TIGR 2020
NOAA/NESDIS, ECMWF for Europe, NCAR/ERL for Arctic, Centre de Météorologie Spatiale de Pékin for China, New Zealand Meteorological Service for southern hemisphere + 420 000 profiles in January and July restituted via inversion.	Analyzed RadioSoundings Archive (ARSA) for 2008-2014 (more than 1 000 000 situations) + ERA-Interim reanalyses for 2008



Main data
Ancillary data

2. Choice of the profile vertical resolution

TIGR 2000	TIGR 2020
43 fixed pressure levels of 4A : [1013 ; 0,0026] hPa	70 hybrid sigma-pressure levels from surf. to 0,0026 hPa

3. Choice of the airmass classification method to classify starting data sets

TIGR 2000	TIGR 2020
5-class classification based mainly on PCA and HAC	8-class probabilistic classification method

4. Choice of the sampling method and the number of situations

Method selecting situations regularly in the thermodynamic space, instead of preserving statistics of the starting data sets. For each airmass class, iterative sampling method based on a metric *d(i,j)* estimating the difference between T and H2O profiles of 2 situations *i* and *j*. Example for TIGR 2020 :

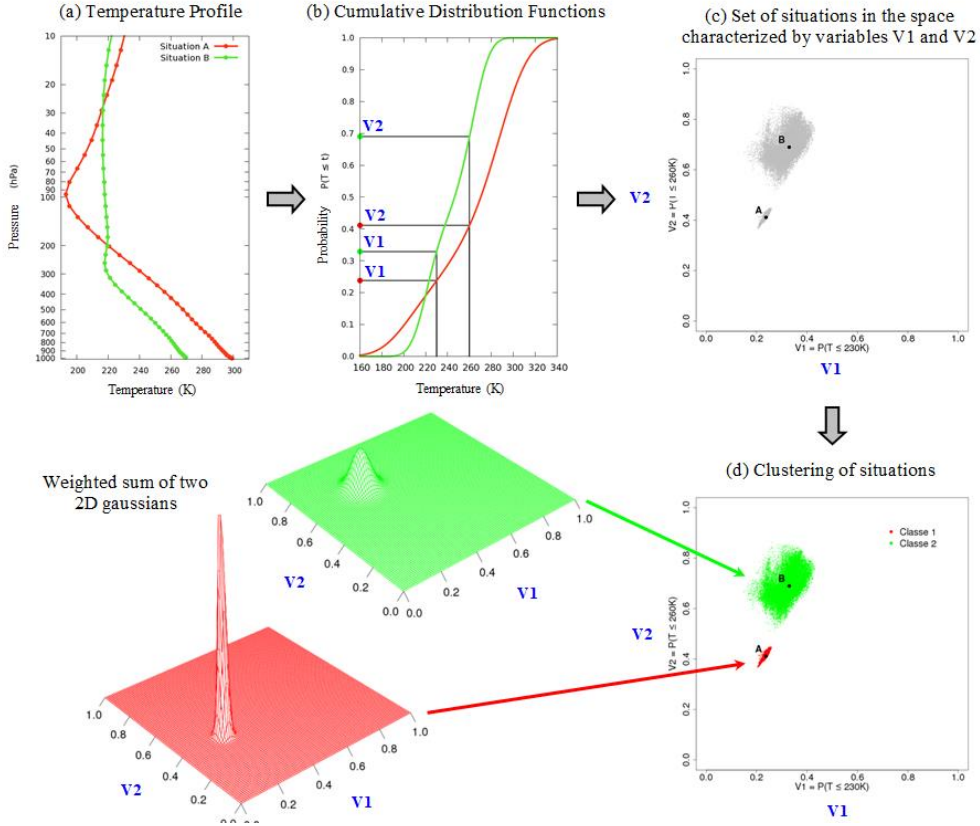
$$d(i,j) = \sqrt{\left(\frac{D_H}{D_T + D_H}\right)\sum_{d=1}^{D_T}\frac{\left(T(d,i) - T(d,j)\right)^2}{\sum_{d=1}^{D_T}\sigma_T^2(d,c_{i,j})} + \left(\frac{D_T}{D_T + D_H}\right)\sum_{d'=1}^{D_H}\frac{\left(Q(d,i) - Q(d,j)\right)^2}{\sum_{d=1}^{D_Q}\sigma_Q^2(d,c_{i,j})}}$$

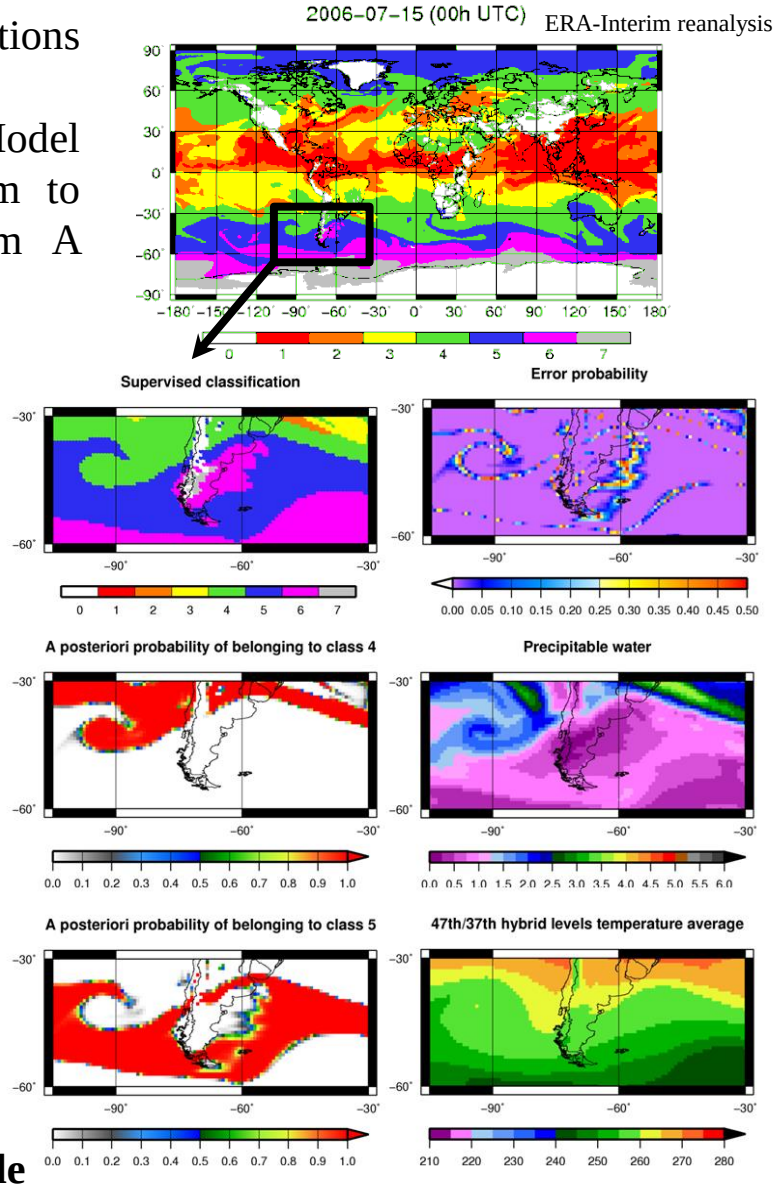
D_T = 52 T (temperature) values
D_H = 12 Q (precipitable water witin a given pressure layers) values
c : airmass class

For each *i* from starting data sets, if $\forall j \in$ TIGR, $d(i,j) \geq dmin$: $i \in$ TIGR. $dmin \propto 1$ / number of situations. First step to build a first TIGR with 100 « desert situations », and a second step to complete.

III – A PROBABILISTIC POINT OF VIEW FOR CLASSIFICATION

1. **Probabilistic input data** : cumulative Distribution Functions (CDFs) to aggregate T/Tdp vertical profiles.
2. **Probabilistic classification method** : Gaussian Mixture Model (GMM) with the Expectaction-Maximization (EM) algorithm to estimate the parameters of the mixture and the Maximum A Posteriori (MAP) to assign each situaton to its nearest class



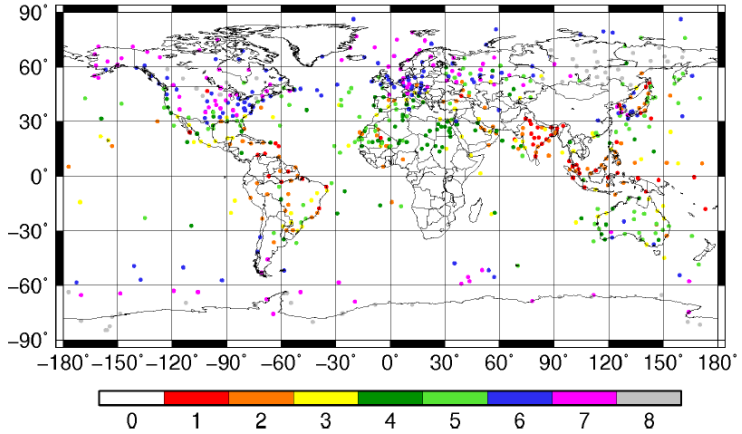


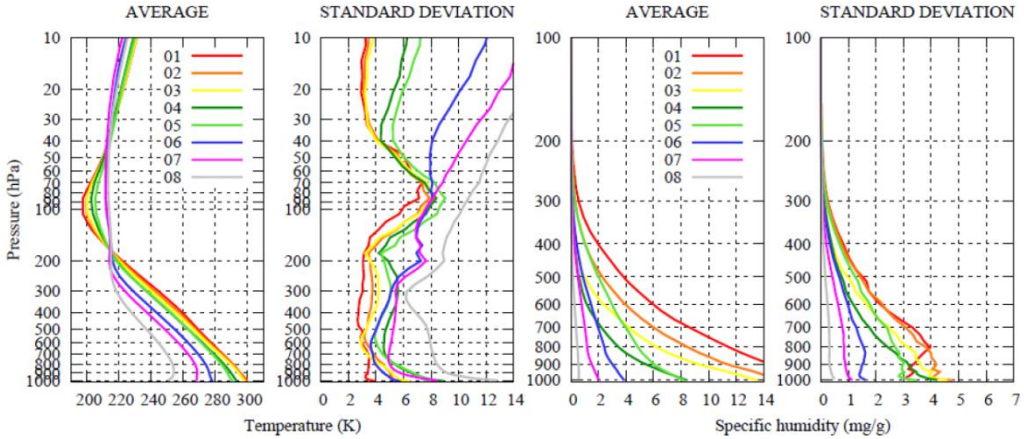
Unserpervised mode on a training data set **then supervised mode**

Pernin J., Vrac M., Crevoisier C. and Chédin A. : « Mixture model-based atmospheric air mass classification: a probabilistic view of thermodynamic profiles. », *Adv. Stat. Clim. Meteorol. Oceanogr.*, 2, 115–136, doi:10.5194/ascmo-2-115-2016 (2016)

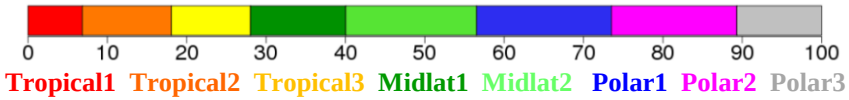
IV – A NEW DATA SET : TIGR 2020

Number of situations	3335	Number of airmass classes	8
Starting data sets to sample	Analyzed RadioSoundings Archive (ARSA) + ERA-Interim reanalyses as ancillary data		
Vertical resolution of profiles	70 hybrid sigma-pressure levels from surface to 0,0026 hPa		
Classification method	(T, Tdp) Cumulative Distribution Functions (CDF) + Gaussian Mixture Model (GMM) based on Expectation-Maximization (EM) algorithm and Maximum A Posteriori (MAP) estimation		
Sampling method	Iterative accept/reject algorithm based on a metric estimating distance between profiles, one step to select 100 « desert situations », followed by a second step to complete		



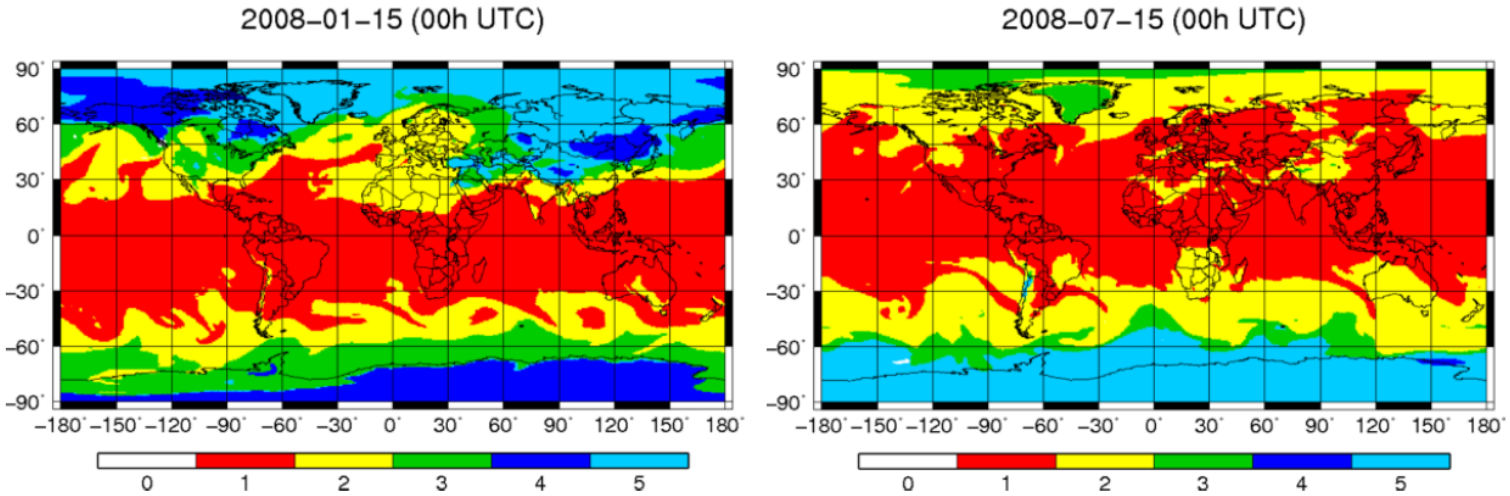


Features / classes	1	2	3	4	5	6	7	8
Precipitable water (cm)	5.67 ± 1.11	4.17 ± 1.11	3.09 ± 0.71	1.75 ± 0.54	2.58 ± 0.83	1.23 ± 0.34	0.64 ± 0.19	0.22 ± 0.10
[800-425] hPa average temperature (K)	273.8 ± 2.6	272.4 ± 3.1	271.2 ± 3.0	267.7 ± 4.5	267.3 ± 4.1	258.5 ± 3.5	252.3 ± 4.7	240.4 ± 7.0
Surface temperature (K)	300.1 ± 3.9	299.4 ± 5.7	298.4 ± 6.5	293.2 ± 8.9	290.9 ± 7.9	278.2 ± 5.7	268.5 ± 8.5	250.7 ± 12.8
Number of situations	228	375	332	400	550	567	527	356
Percentage of situations (%)	6.8	11.2	10.0	12.0	16.5	17.0	15.8	10.7



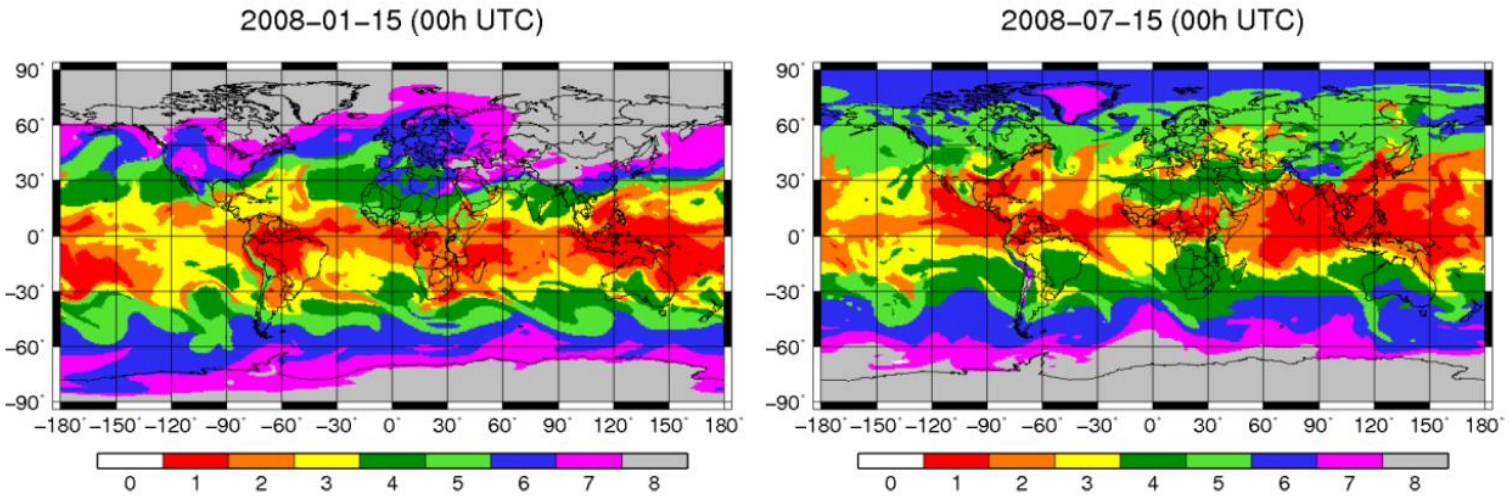
V – AIRMASS CLASSIFICATION

TIGR-2000 type



5-class supervised classification of ECMWF reanalyses, after unsupervised multivariate GMM on TIGR-2000

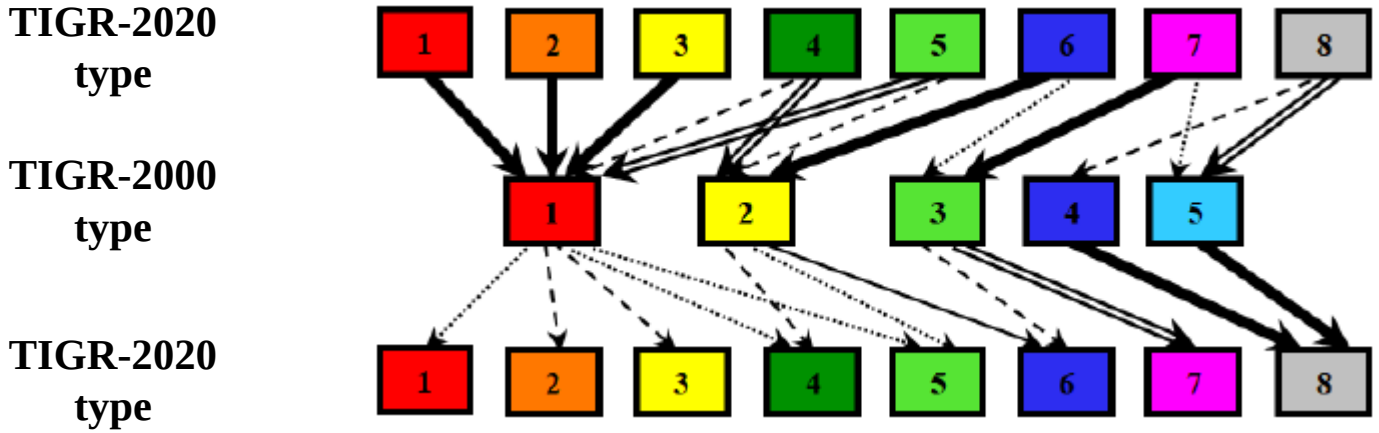
TIGR-2020 type



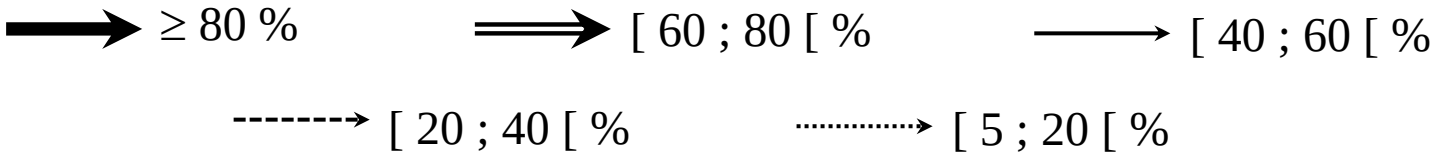
8-class supervised classification of ERA-Interim reanalyses, after unsupervised multivariate GMM on training dataset ERA-Interim reanalyses { 00h + 12h UTC of 15th of January, April, July, October between 2005 and 2009 }

V – AIRMASS CLASSIFICATION

Two different classifications :



Arrow diagram illustrating the correspondence between both classifications. The style of the arrows characterizes the % of atmospheric situations shared relative to the size of the starting class :



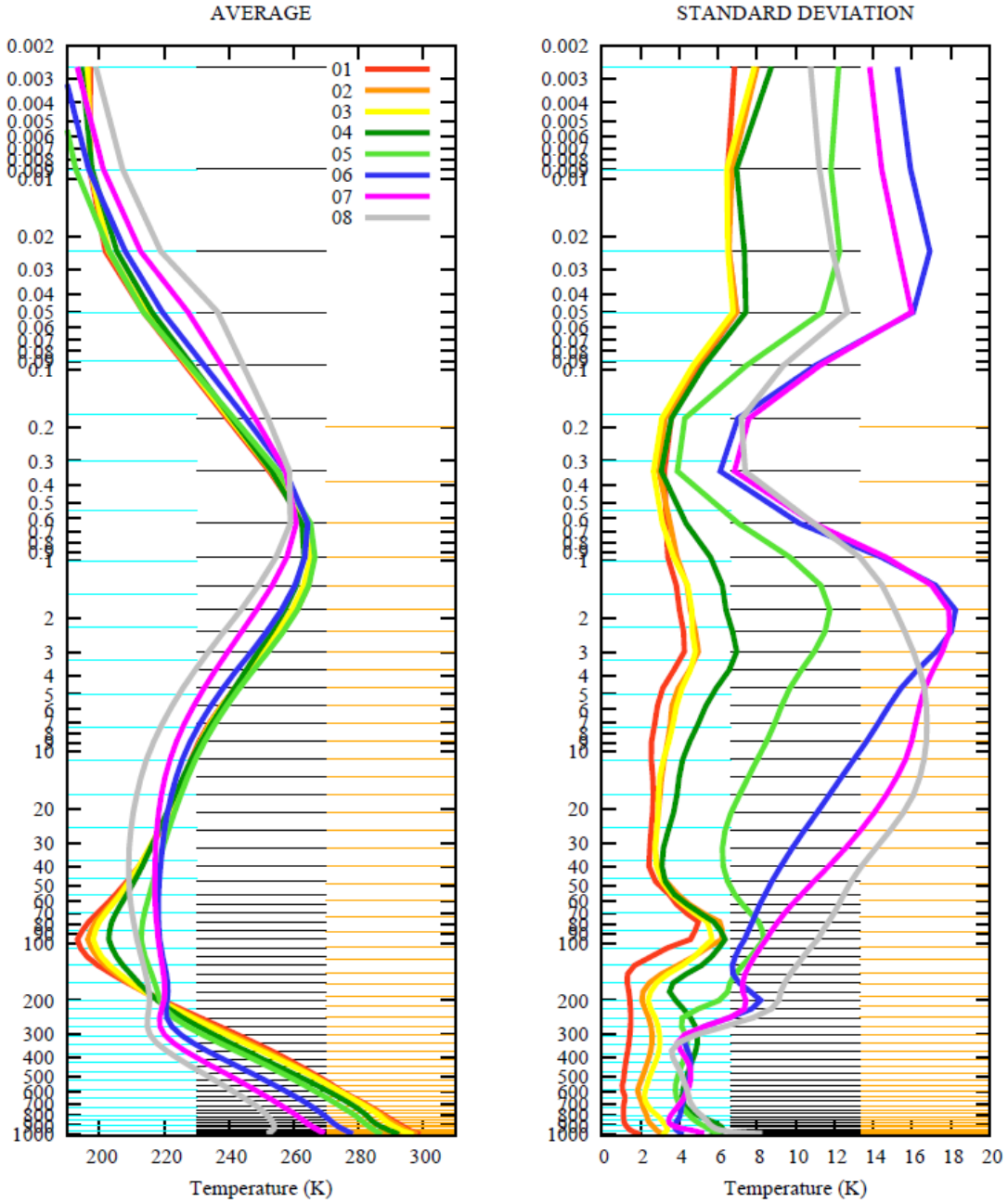
VI – VERTICAL RESOLUTION OF THERMODYNAMIC PROFILES

43 fixed pressure 4A levels (TIGR-2000)

70 LMD hybrid sigma-pressure levels (TIGR-2020)

60 ECMWF hybrid sigma-pressure levels

Finer representativity in the lower troposphere and near the tropopause.

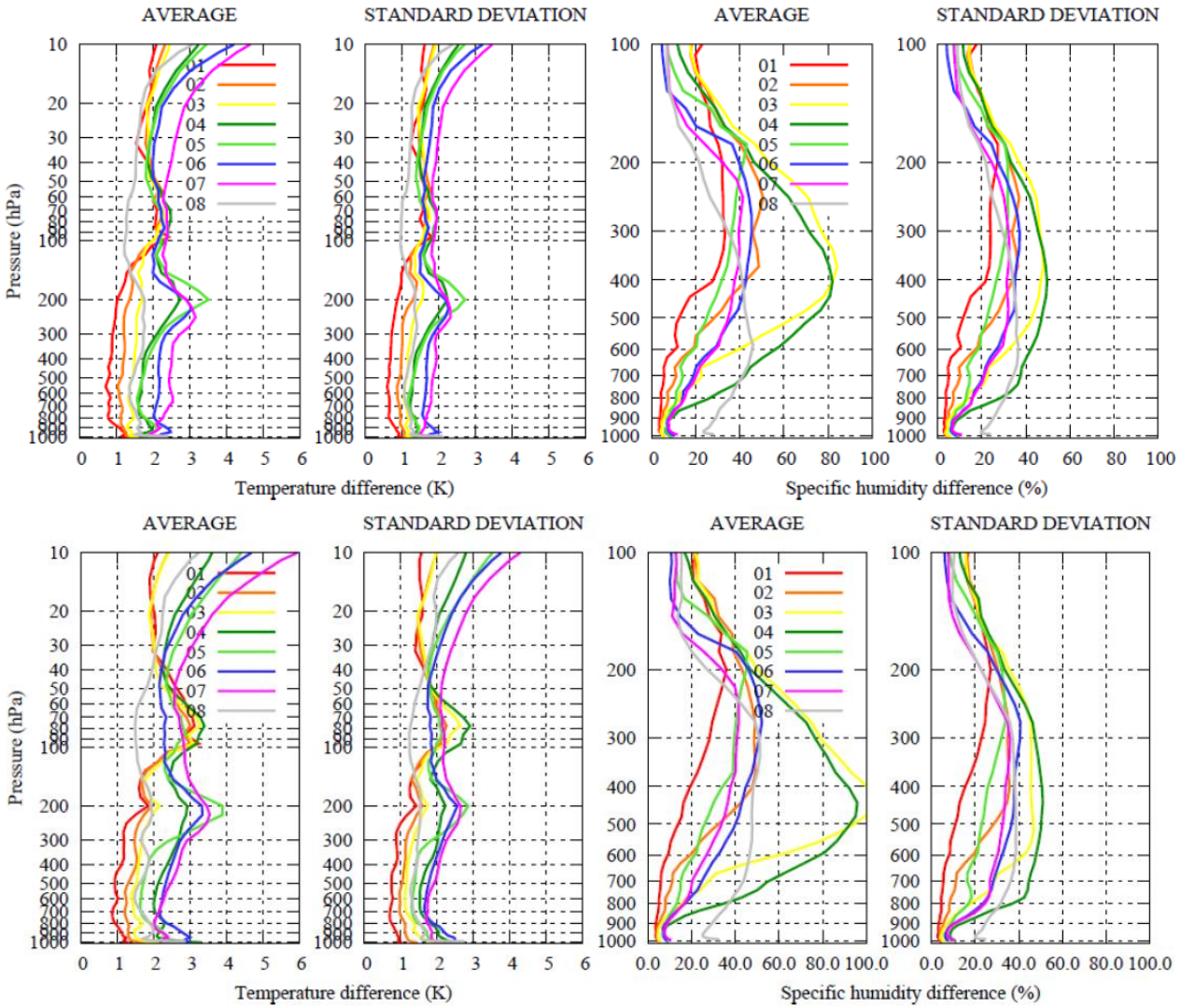


VII – VALIDATION IN THERMODYNAMIC SPACE

TIGR 2020

Finer representativity of 3D thermodynamic states
(= finer mesh in multivariate thermodynamic space)

TIGR 2000

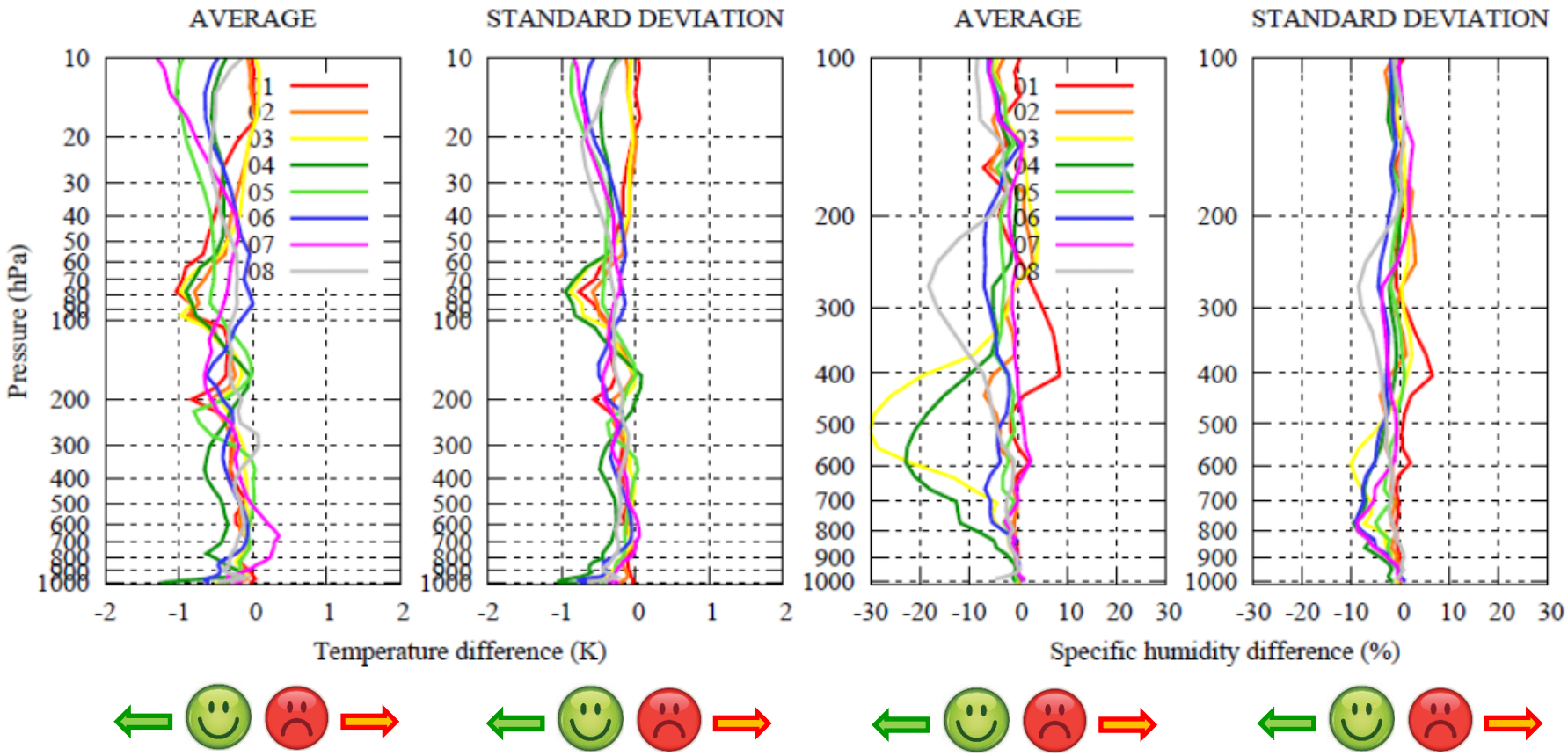


Average and standard deviation, by pressure level and airmass class, projected on 8 classes of $| T_i - T_j |$ (in K) and $| Q_i - Q_j |$ (in %) with T : temperature, Q : specific humidity, and :
i : a given situation of a test data set (ARSA+ERA-Interim for 2015)
j : average of nearest situations in TIGR via pattern recognition within all airmass gathered

VII – VALIDATION IN THERMODYNAMIC SPACE

TIGR 2020 – TIGR 2000

(substraction of the previous figures)



A difference < 0 indicates a finer representativity of 3D atmospheric thermodynamic states

(= finer mesh in multivariate thermodynamic space)

CONCLUSION

- **A new TIGR data set : TIGR 2020**

(i) A similar multivariate sampling approach applied on up-to-date starting data sets (radiosoundings from ARSA + ERA-Interim as ancillary data), leading globally to a **finer representativity of 3D atmospheric states in the thermodynamic space**.

(ii) A **finer thermodynamic profile vertical resolution** (lower troposphere, tropopause).

(iii) A 8-airmass classification depicting **more accurately tropical temperature and humidity behaviours**.

- **Radiative TIGR 2020**

- The corresponding TIGR 2020 radiative dataset has been computed with 4A/OP (2016, v1.7) using GEISA2020. For each of the 3335 situations have been computed: **IASI spectrum, Jacobians of T and main gases, atmospheric transmissions**.

- Both the **thermodynamic and radiative TIGR 2020 will soon be available** on the AERIS website.

- **Perspectives**

- Updates of the various retrievals schemes developed with IASI and other instruments for climate variables.

- Specific studies for desert and polar situations, in particular, to evaluate the improvement in their representativity.

A newly born TIGR data set to meet the requirements of the high spectral resolution instruments : the TIGR-2020

Contact

Jérôme PERNIN (jerome.pernin@lmd.ipsl.fr)

References

Classification: Pernin J., Vrac M., Crevoisier C. and Chédin A. : « Mixture model-based atmospheric air mass classification: a probabilistic view of thermodynamic profiles. », *Adv. Stat. Clim. Meteorol. Oceanogr.*, 2, 115–136, doi:10.5194/ascmo-2-115-2016 (2016)

TIGR: https://ara.lmd.polytechnique.fr/index.php?page=tigr
ARSA: https://ara.lmd.polytechnique.fr/index.php?page=arsa
4A/OP: https://ara.lmd.polytechnique.fr/index.php?page=4a
GEISA: https://ara.lmd.polytechnique.fr/index.php?page=geisa-2
Publications: https://ara.lmd.polytechnique.fr/index.php?page=publications