Comparing two Methods to exploit the Synergy between FORUM and IASI-NG

Institute for Applied Physics "Nello Carrara" of the Italian National Research Council

CNR-INO ISTITUTO NAZIONALE DI OTTICA CONSIGLIO NAZIONALE DELLE RICERCHE

Cecilia Tirelli[(1)], Marco Ridolfi[(2)] , Simone Ceccherini[(1)], Ugo Cortesi[(1)] , Luca Palchetti[(2)]

(1) Istituto di Fisica Applicata 'N. Carrara'- CNR Via Madonna del Piano 10, 50019, Sesto Fiorentino, Italy, Email: c.tirelli@ifac.cnr.it

(2) Istituto Nazionale di Ottica – CNR Via Madonna del Piano 10, 50019, Sesto Fiorentino, Italy

Synergystic use of remote sensing data

In the last few decades, the need to advance the knowledge of tropospheric and stratospheric chemical / physical processes and the availability of a large amount of data stimulated the use of synergistic approaches to exploit the complementarity of information contained in measurements provided by a great number of satellite missions.

Two classes of strategies can be used to combine multiple sets of independent measurements of the same air-mass / ground surface portion:

- the **Synergistic retrieval (SR).** The SR product is obtained fitting simultaneously the radiances acquired by the two instruments with forward model simulations

- the **Complete Data Fusion (CDF).** Is an a posteriori method that uses the results of the individual retrievals (vmr, CMs and AKMs) to provide a combined / enhanced final product.

In this work, we characterize the differences between the SR and CDF solutions for realistic conditions that may be encountered in the attempt to combine the complementary measurements of two forthcoming satellite missions: FORUM (Far Infrared Outgoing Radiation Understanding and Monitoring) and IASI-NG (Infrared Atmospheric Sounding Interferometer-New Generation).

The Synergistic Retrieval (SR)

The SR product is obtained from the simultaneous fit of the radiances measured by two instruments (FORUM and IASI-NG in this case) sounding the same (or nearly the same) air-mass / ground pixel, with forward model (FM) simulations. The SR solution x is obtained by minimizing the cost function:

$$\xi^2(\mathbf{x}) = \sum_{i=1}^{2} (\mathbf{y}_i - \mathbf{F}_i(\mathbf{x}))^t \, \mathbf{S}_{yi}^{-1} \, (\mathbf{y}_i - \mathbf{F}_i(\mathbf{x})) + (\mathbf{x}_a - \mathbf{x})^t \, \mathbf{S}_a^{-1} \, (\mathbf{x}_a - \mathbf{x}) \ ,$$

using Gauss-Newton iterative formula:

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \left[\sum_{i=1}^{2} \mathbf{K}_{i,k-1}^t \mathbf{S}_{yi}^{-1} \mathbf{K}_{i,k-1} + \mathbf{S}_a^{-1} \right]^{-1} \left[\sum_{i=1}^{2} \mathbf{K}_{i,k-1}^t \mathbf{S}_{yi}^{-1} (\mathbf{y}_i - \mathbf{F}_i(\mathbf{x}_{k-1})) + \mathbf{S}_a^{-1} (\mathbf{x}_a - \mathbf{x}_{k-1}) \right]$$

The SR solution is characterized by a **covariance matrix (CM)** and an **averaging kernel matrix (AKM)** given by:

$$\mathbf{S} = \left[\sum_{i=1}^{2} \mathbf{K}_i^t \mathbf{S}_{yi}^{-1} \mathbf{K}_i + \mathbf{S}_a^{-1} \right]^{-1}$$

$$\mathbf{A} = \left[\sum_{i=1}^{2} \mathbf{K}_i^t \mathbf{S}_{yi}^{-1} \mathbf{K}_i + \mathbf{S}_a^{-1} \right]^{-1} \sum_{i=1}^{2} \mathbf{K}_i^t \mathbf{S}_{yi}^{-1} \mathbf{K}_i \ .$$

In the equations: y_i are vectors including the spectral radiances of the instrument i=1,2 (FORUM or IASI-NG); x_i are the state vectors describing the atmospheres sounded by the instruments; S_{yi} the CMs; S_a the apriori CM of the apriori state vector x_a to constrain the SR; K_i the Jacobian of the FM.

In the case of a **temporal/spatial mismatch** between the two measurements ($x_1 \neq x_2$) we still assume that both instruments are sounding the same atmospheric state x_1 but in the SR we assign to y_2 (radiances acquired by IASI-NG) the error CM S'_{y2} :

$$\mathbf{S}'_{y2} = \mathbf{S}_{y2} + \mathbf{K}_2 \mathbf{S}_M \mathbf{K}_2^t$$

i.e a larger error (that depends on the mismatch matrix S_M) as compared to the original one, described by S_{y2}

The IASI-NG retrieval obtained with S'_{y2} produces different state vector, CM and AKM with respect to those obtained with S_{y2}

The Complete Data Fusion (CDF)

The Complete Data Fusion (CDF), is an a posteriori algorithm to combine the individual retrievals obtained with the optimal estimation from independent measurements of the same airmass and/ or ground pixel into a single estimate. The CDF solution is obtained analytically (no iterations needed), minimizing the cost function:

$$\xi^2_{\mathrm{CDF}}(\mathbf{x}) = \sum_{i=1}^{2} (\boldsymbol{\alpha}_i - \mathbf{A}_i\mathbf{x})^t \mathbf{S}_{\mathrm{n},i}^{-1} (\boldsymbol{\alpha}_i - \mathbf{A}_i\mathbf{x}) + (\mathbf{x}_{\mathrm{a}} - \mathbf{x})^t \mathbf{S}_{\mathrm{a}}^{-1} (\mathbf{x}_{\mathrm{a}} - \mathbf{x})$$

with: $\boldsymbol{\alpha}_i = \hat{\mathbf{x}}_i - (\mathbf{I} - \mathbf{A}_i)\mathbf{x}_{\mathrm{a}i} = \mathbf{A}_i\mathbf{x}_i + \boldsymbol{\sigma}_i$,

by imposing its gradient equal to zero, and it is given by:

$$\mathbf{x}_{\mathrm{f}} = \left(\sum_{i=1}^{2} \mathbf{A}_i^t \mathbf{S}_{\mathrm{n},i}^{-1} \mathbf{A}_i + \mathbf{S}_{\mathrm{a}}^{-1} \right)^{-1} \left(\sum_{i=1}^{2} \mathbf{A}_i^t \mathbf{S}_{\mathrm{n},i}^{-1} \boldsymbol{\alpha}_i + \mathbf{S}_{\mathrm{a}}^{-1} \mathbf{x}_{\mathrm{a}} \right)$$

which is characterized by the following CM and AKM:

$$\mathbf{S}_{\mathrm{f}} = \left(\sum_{i=1}^{2} \mathbf{A}_i^t \mathbf{S}_{\mathrm{n},i}^{-1} \mathbf{A}_i + \mathbf{S}_{\mathrm{a}}^{-1} \right)^{-1}$$

$$\mathbf{A}_{\mathrm{f}} = \left(\sum_{i=1}^{2} \mathbf{A}_i^t \mathbf{S}_{\mathrm{n},i}^{-1} \mathbf{A}_i + \mathbf{S}_{\mathrm{a}}^{-1} \right)^{-1} \sum_{i=1}^{2} \mathbf{A}_i^t \mathbf{S}_{\mathrm{n},i}^{-1} \mathbf{A}_i$$

In the equations: x_i are the state vectors describing the atmospheres sounded by the instruments i=1,2 (FORUM or IASI-NG); $S_{n,i}$ the noise CMs; S_a, the apriori CM of the apriori state vector x_a used to constrain the CDF; Ai the AKM of the individual retrievals.

In the case of a **temporal/spatial mismatch** between the two measurements, we introduce a coincidence error (Ceccherini et al. 2018). In our case, we simply substituted in the equations above $S_{n,2}$ with:

$$\mathbf{S}'_{\mathrm{n},2} = \mathbf{S}_{\mathrm{n},2} + \mathbf{A}_2 \mathbf{S}_{\mathrm{M}} \mathbf{A}_2^t$$

adding the coincidence error to IASI-NG measurement.

In the CDF approach, only the CM of the IASI-NG retrieval is changed, leaving the state vector and the AKM equal to those obtained in the case of perfect matching

Differences between the two methods

The **SR** :

- rigorously combines complementary information of the measurements;
- requires to integrate into a single inversion system the radiative transfer models capable to simulate the measurements of all the sensors involved in the synergistic inversion;
- requires the simultaneous access to all the (Level 1) measurements used in the inversion, thus implying the need to handle relevant data volumes.

The **CDF**:

- combines the Level 2 products supplied by the individual retrieval processors of the independent measurements;
- requires a simple implementation
- has the capability to reduce the amount of data involved in the synergistic analysis.

Ceccherini et al. 2015 show that CDF and SR provide the same solution (equal error estimates and number of DOFs) under:

1) **linear approximation** of the forward model of each measurement in the range of variability between the solutions of the single retrievals and of the synergistic retrieval;

2) assumption of **perfect spatial and temporal matching** between measurements.

Instruments

FORUM (Far-infrared Outgoing Radiation Understanding and Monitoring) experiment has been selected to be the ninth Earth Explorer mission of the European Space Agency (ESA) and it is scheduled for launch on a polar orbiting satellite in 2027. The core instrument will be a Far-InfraRed (FIR) spectrometer.

FORUM will fly in loose formation with the MetOp-SG satellite, hosting the **Infrared Atmospheric Sounding Interferometer – New Generation (IASI-NG)**. Since IASI-NG will measure the Middle-InfraRed (MIR) part of the upwelling atmospheric spectrum (from 645 to 2760 cm-1), the matching FORUM FIR spectra can be effectively used to synergistically complement the IASI-NG measurements.

	Instruments	
Characteristics	**IASI-NG**	**FORUM**
Spectral coverage	**645- 2760** cm^{-1}	**100-1600** cm^{-1}
Spectral sampling	**0.125** cm^{-1}	**0.36** cm^{-1} (goal)
Spectral resolution	**0.25** cm^{-1}	**0.5** cm^{-1} (FWHM)
Measurement mode	Step and stare (azimuth scanning)	Step and stare (no azimuth scanning)
Ground pixel (diameter at nadir)	**12 km**	**15 km**
Satellite	On board two sun-synchronous polar orbiting satellites in loose formation (MetOp-SG-A1 for IASI-NG)	

Simulated measurements

To generate the synthetic measurements of FORUM and IASI-NG we considered the instrumental specifications of the two instruments. We carried out two sets of test retrievals emulating a situation in which both FORUM and IASI-NG measure, with perfect matching or with a realistic spatial/temporal mismatch, the same portion of the Antarctic Plateau surface.

In particular, we base our tests on a reference clear-sky atmospheric scenario corresponding to winter conditions over the Antarctic plateau (82° S,71° E, 3600 m a.s.l, 20 June 2007) covered by coarse snow. This **reference scenario $\mathbf{x_0}$** was selected from the outputs of the NWP model of ECMWF.

The retrieved state vector contains: **temperature, surface temperature, water vapour, ozone** and **surface emissivity;**

To compute the outgoing spectral radiances in the interval from 80 to 2780 cm^{-1} (step 0.01 cm^{-1}) we used the fast monochromatic, parametrized forward **model σ-RTM** (developed at University of Basilicata*).

(*) Amato, U., Masiello, G., Serio, C., and Viggiano, M.: The σ-IASI code for the calculation of infrared atmospheric radiance and its derivatives, Environmental Modelling & Software, 17, 651–667, https://doi.org/https://doi.org/10.1016/S1364-8152(02)00027-0, 2002.

Mismatch

The objective of SR and CDF is to obtain the best estimate of the atmospheric and surface state corresponding to the air mass and ground pixel **sounded by FORUM**, with the help of IASI-NG.
If IASI-NG is not sounding the same airmass/ground pixel, **we introduce a MISMATCH error to the IASI-NG spectrum for SR and to the state vector retrieved from IASI-NG measurements when used in the CDF approach.** We assume the worst case of 1 min time lag and 26 km distance between the closest FORUM and IASI-NG soundings to evaluate the mismatch effect.

The mismatch error assigned to IASI-NG state vector is a block-diagonal CM, each block is associated to a retrieval target:

$$\mathbf{S}_{\mathrm{M}} = \begin{pmatrix} \mathbf{S}_{\mathrm{T}} & 0 & 0 & 0 & 0 \\ 0 & \mathbf{S}_{\mathrm{Ts}} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{S}_{\mathrm{H}} & 0 & 0 \\ 0 & 0 & 0 & \mathbf{S}_{\mathrm{O_3}} & 0 \\ 0 & 0 & 0 & 0 & \mathbf{S}_{\mathrm{e}} \end{pmatrix}$$

$\mathbf{S}_{O3}$ is considered=0.
$\mathbf{S}_{T}$, $\mathbf{S}_{Ts}$ **and** $\mathbf{S}_{H}$ are estimated considering the variability obtained from ERA5 atmospheric and surface fields.
$\mathbf{S}_{e}$ is estimated from a set of 19 surface emissivity models from Huang et al. 2016.

In the SR, we use eq $\mathbf{S}'_{\mathrm{y2}} = \mathbf{S}_{\mathrm{y2}} + \mathbf{K}_2\mathbf{S}_{\mathrm{M}}\mathbf{K}_2^t$. to map the error SM onto the IASI-NG spectrum.

In the CDF, we use eq. $\mathbf{S}'_{\mathrm{n},2} = \mathbf{S}_{\mathrm{n},2} + \mathbf{A}_2\mathbf{S}_{\mathrm{M}}\mathbf{A}_2^t$ to add the mismatch error to the noise CM of IASI-NG.

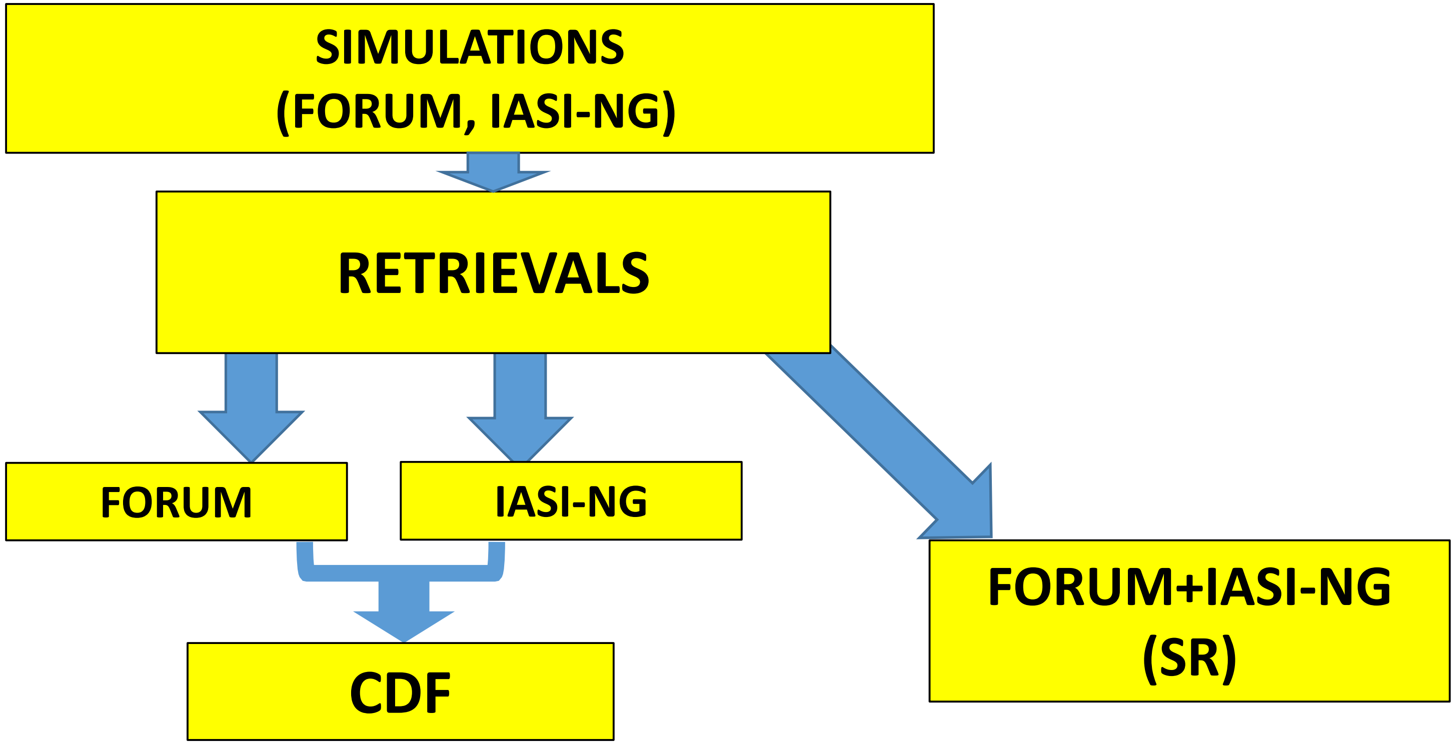
Retrieval Set-Up

We considered two cases: **no mismatch** (EXP1) and **realistic mismatch** (EXP2) between measurements.

In both cases, we repeat 900 times this procedure:

- we generate two true state vectors $\mathbf{x_1}$ and $\mathbf{x_2}$ using T(p), Ts, x_{H2O}(p) and x_{O3}(p) obtained by applying a stochastic perturbation to x_0 consistent with $\mathbf{S_M}$/2 diagonal values. We assume the surface spectral emissivity of coarse snow (from Huang et al 2016) for both measurements in EXP1. In EXP2 the medium snow is assumed for IASI-NG;
- we added **a noise** to observations consistent with the respective noise error CMs S_{y1} and S_{y2};
- we generated the **a priori state** vector $\mathbf{x_a}$ applying a random perturbation to x_0 consistent with $\mathbf{S_a}$ (United Kingdom MetOffice). For surface emissivity $\mathbf{e}_a$ is constant and =0.99 and the a priori error is 0.1 in the spectral range covered by the measurements and 10^{-4} elsewhere. The a priori is used as initial guess of the retrieval iterations. The same a priori data are used to process a given pair of measurements;

Finally, we carry out the retrievals from **FORUM-only**, **IASI-NG-only** and **FORUM+IASI-NG** (synergistic retrieval) measurements and compute the **CDF** result starting from FORUM-only and IASI-NG only retrieved state vectors.



NO MISMATCH $x_1=x_2$	
CDF $\hat{x}_f$	**SR $\hat{x}_{sr}$**
Input state vectors and matrices: $\hat{x}_1$ and $\hat{x}_2$, A_1 and A_2, $S_{n,1}$ and $S_{n,2}$	x_{sr}, S_{y1} and S_{y2} are the same used in the individual retrievals.

MISMATCH $x_1 \neq x_2$	
CDF $\hat{x}_f$	**SR $\hat{x}_{sr}$**
Input state vector and matrices: $\hat{x}_1$ and $\hat{x}_2$, A_1 and A_2, $S_{n,1}$ and $S'_{n,2}$ $\mathbf{S}'_{n,2} = \mathbf{S}_{n,2} + \mathbf{A}_2 \mathbf{S}_M \mathbf{A}_2^t$	x_{sr}, S_{y1} is the same of no mismatch case, S'_{y2} is calculated as: $\mathbf{S}'_{y2} = \mathbf{S}_{y2} + \mathbf{K}_2 \mathbf{S}_M \mathbf{K}_2^t$

Results:

- Experiment 1: NO MISMATCH

- Experiment 2: MISMATCH

Analysis of the results

For T, H_2O, O_3 and surface emissivity, in the two experiments of 900 trials, we evaluated:

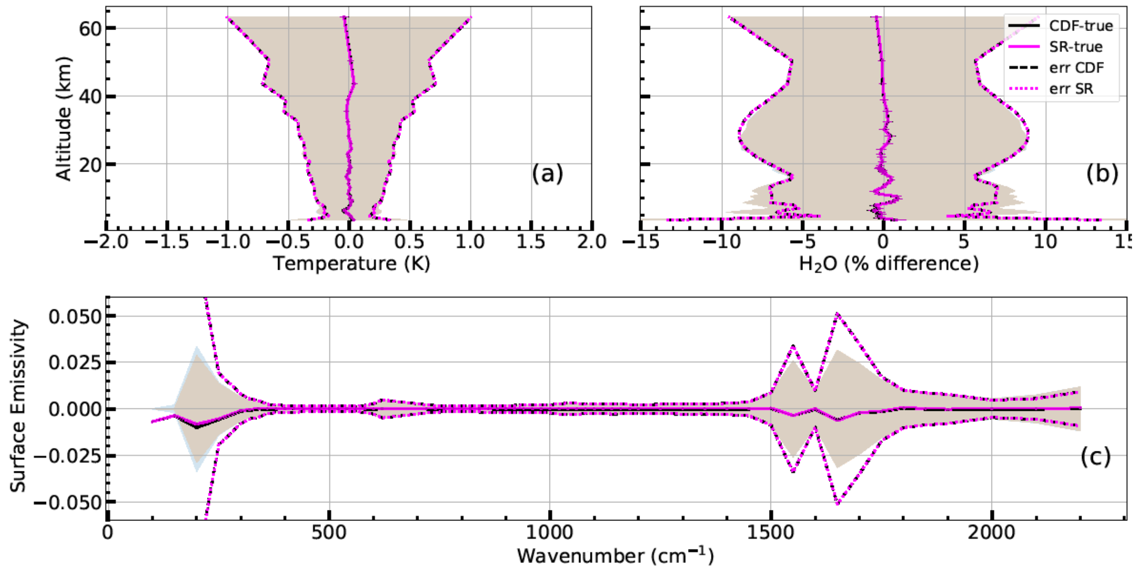
- the **average differences between CDF and SR products and the true values**
- the **average differences between CDF and SR products**

In both cases, we considered: the average error of CDF and of SR (as evaluated from the error CMs S and S_f), the standard deviations of the differences (shaded areas in the plots) and the standard error of the average differences (error bars in the plots).

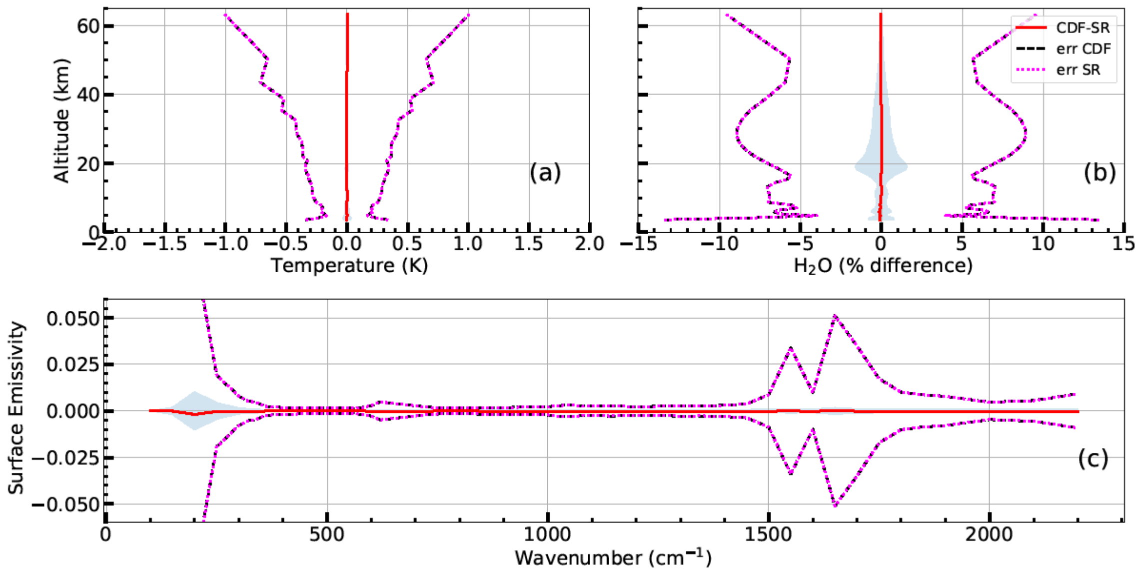
The average differences quantify the product's bias, while the standard deviation of the differences is an (ex-post) estimate of the product error which, in principle, should equal the product error estimated (ex-ante) with the error CMs. The standard error of the average (SD/30 in this case) is useful to evaluate whether the determined bias is statistically significant.

Experiment 1: NO MISMATCH

Average differences between CDF/SR and true profiles.



Average differences between CDF and SR profiles.



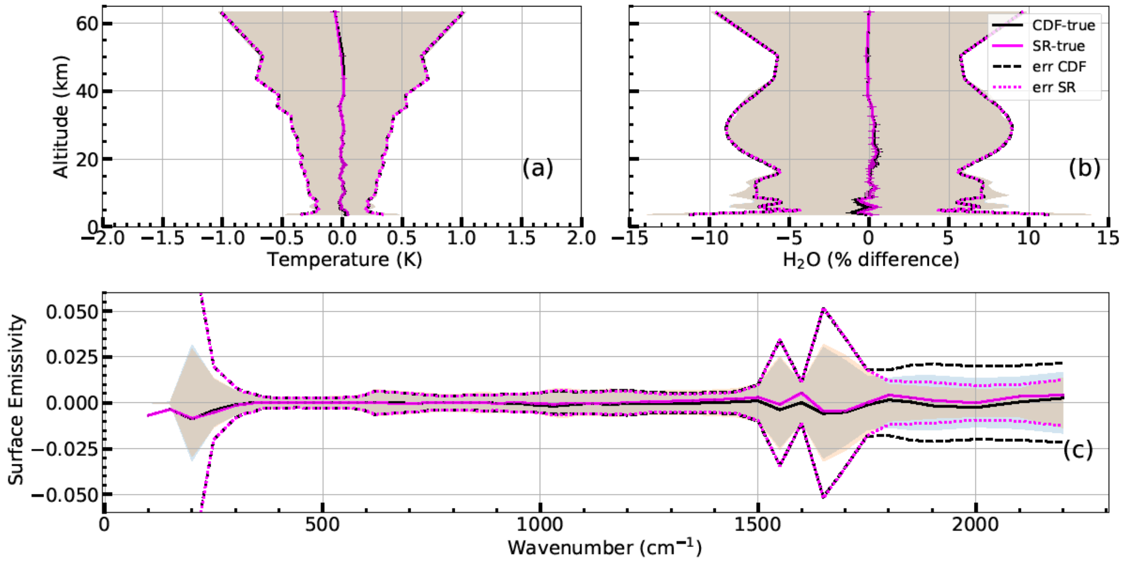
Dashed lines represent the average error of CDF (black) and of SR (magenta) as evaluated from the error CMs S and S_f. Shaded areas represent the standard deviations of the differences.

- The **biases** of both the CDF and SR solutions **are much smaller than the average profile errors;**
- the **standard deviation** of the differences (ex-post error estimator) **agree very well with the average errors of CDF and SR** (ex-ante error estimator).

- On average, the **differences between CDF and SR solutions are far smaller than the error estimated by the CMs**;
- the standard deviation of the differences is much smaller than the error (**differences between the CDF and SR solutions are very small also in the individual test runs**, i.e. the FM linear approximation used in the CDF is very accurate, for our case)

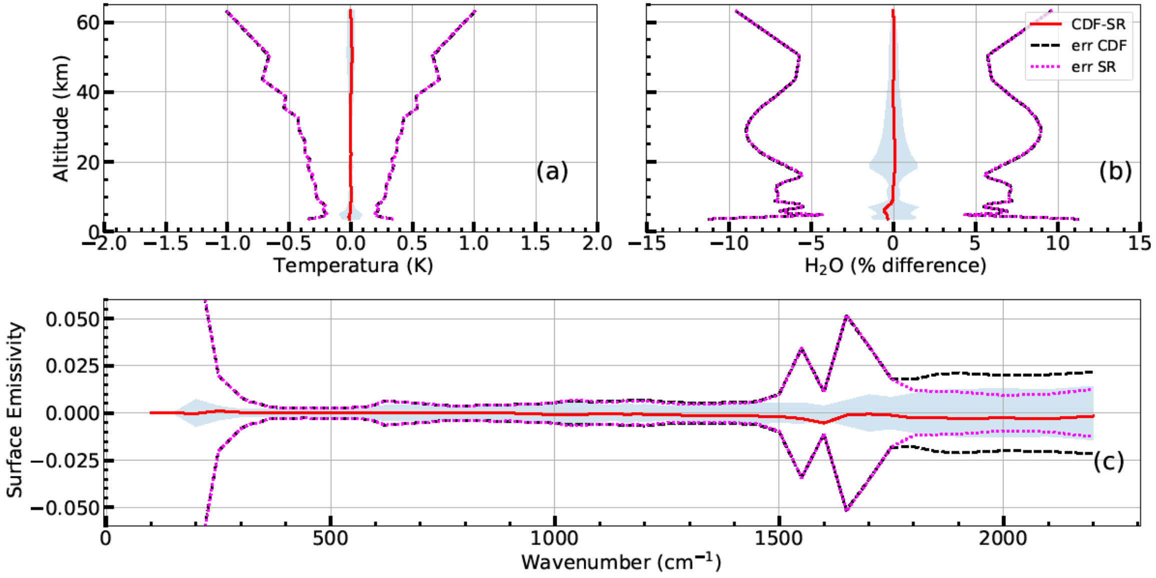
Experiment 2: MISMATCH

Average differences between CDF/SR and true profiles.



- The **bias** of both CDF and SR solutions is still much **smaller than the estimated error**.
- Estimated **error** (dashed lines) is **slightly increased** as compared to the case of perfectly matching measurements (especially for spectral emissivity).
- Above 1700 cm^{-1}, the **emissivity** error of the CDF solution is slightly larger than that of the SR (different handling of the mismatch error in the CDF and SR approaches)

Average differences between CDF and SR profiles.



- The **average differences** between the CDF and SR solutions are **much smaller than their estimated error**, even in the presence of a mismatch.

- the **standard deviation of the differences** in spectral emissivity between CDF and SR may **be as large as the estimated error** (in each individual test run, the difference between the CDF and SR solutions may be as large as the error estimated from the CMs).

Conclusions

In this study, we characterized, for a specific test scenario, the differences between SR and CDF techniques used to generate synergistic products of FORUM and IASI-NG, forthcoming missions that will fly in loose formation. We analysed the statistics of 900 simulated observations (and retrievals) in two cases: considering a perfect matching between measurements and evaluating a realistic time and space mismatch.

The objective of the study was to characterize the differences between the results obtained from the SR and CDF approaches.

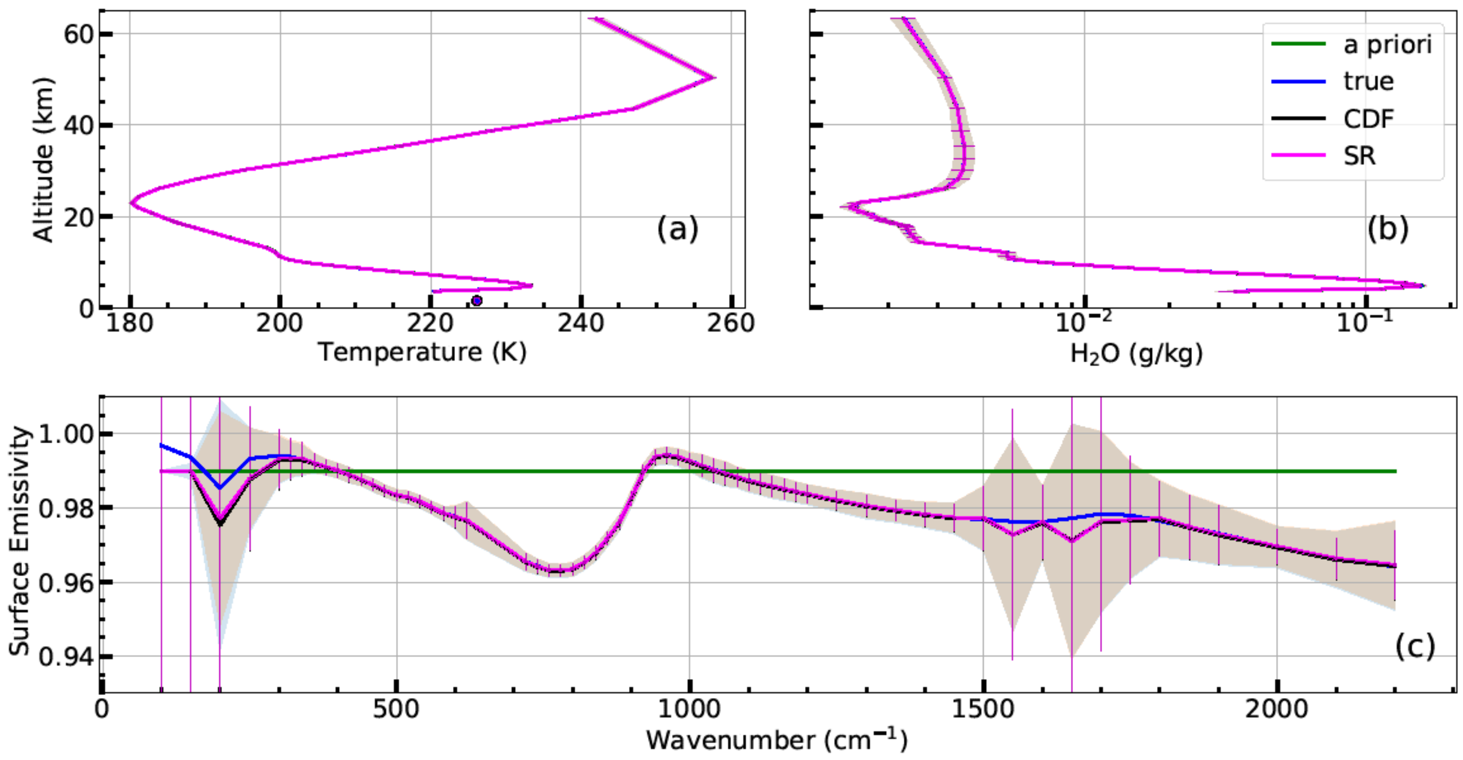
We found that:

- in case of perfectly matching measurements, SR and CDF actually provide results that differ by less than 1/10 of their associated noise retrieval error;

- in case of a realistic mismatch between the measurements, the two methods provide results that differ more, the differences, however, are still within the retrieval error;

- **the differences between SR and CDF results are mainly due to the different treatments of the mismatch in the two methods and not to the non-linearities of the problem.**

Paper (AMTD) : Ridolfi et al., Synergistic retrieval and Complete Data Fusion methods applied to FORUM and IASI-NG simulated measurements.



Average on the 900 trials, of a priori, true, CDF and SR **profiles** in case of **perfectly matching** measurements.
Error bars represent the average profile errors as evaluated from the error CMs.
Shadowed areas represent the SR and CDF profiles standard deviation.