Synergistic Retrieval and Complete Data Fusion applied to FORUM and IASI-NG Simulated Measurements

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

[1] *Istituto Nazionale di Ottica del Consiglio Nazionale delle Ricerche, Via Madonna del Piano 10, 50019 Sesto Fiorentino (Firenze), Italy*

[2] *Istituto di Fisica Applicata "Nello Carrara" del Consiglio Nazionale delle Ricerche, Via Madonna del Piano 10, 50019 Sesto Fiorentino (Firenze), Italy*

INTRODUCTION

In the last two decades, there has been a strong focus on the development of innovative techniques to exploit all the available information from remote sensing measurements of the same portion of the atmosphere and/or Earth surface ground pixel. In this study, we compare two alternative approaches to determine atmospheric and surface state parameters by exploiting simultaneously both FORUM (Far-Infrared Outgoing Radiation Understanding and Monitoring) and IASI-NG (Infrared Atmospheric Sounding Interferometer – New Generation) simulated measurements. We examine the synergistic retrieval of state parameters from the simultaneous inversion of both measurements, and the Complete Data Fusion of state parameters obtained from the inversion of the individual measurements. We perform test retrievals based on synthetic clear-sky measurements in which both FORUM and IASI-NG measure, either with perfect matching or with a mismatch in time and space, over the Antarctic Plateau surface covered by coarse snow. For both cases we characterized the quality of synergistic and fused products by means of their error evaluated both from the covariance matrix of the retrieval and the statistics of the differences between retrieved and true state parameters. We finally evaluated the differences between fused and synergistic products. The retrieval products considered in this study are obtained with the optimal estimation method and include the profiles of temperature, surface temperature, water vapour, ozone and spectral emissivity from 100 to 2300 cm^{-1}.

THE SYNERGISTIC RETRIEVAL (SR)

The synergistic retrieval (SR) is a commonly used method to rigorously combine complementary information from data provided by sensors measuring in different spectral ranges.

The SR product is obtained from the simultaneous fit of the radiances measured by two instruments (FORUM and IASI-NG in this case) with the forward model (FM) simulations. The SR solution $\hat{\boldsymbol{x}}_{sr}$ 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

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

modified with the Levenberg-Marquardt method. The SR solution is characterized by a covariance matrix (CM) and an averaging kernel matrix (AKM) given by:

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

In the equations above: 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; Sa the apriori CM of the apriori state vector x_a to constrain the SR; K_i the Jacobian of the FM.

In case of a temporal/spatial mismatch between the two measurements we assume that both instruments are sounding the same atmospheric state and we add to IASI-NG a mismatch CM describing the variability of the measured atmospheric states. In the equations above S_{y2} is replaced by:

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

THE COMPLETE DATA FUSION (CDF) METHOD

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), minimizing the cost function:

$$\xi_{CDF}^2(\mathbf{x}) = \sum_{i=1}^{2} (\boldsymbol{\alpha}_i - \mathbf{A}_i\mathbf{x})^t \mathbf{S}_{n,i}^{-1} (\boldsymbol{\alpha}_i - \mathbf{A}_i\mathbf{x}) + (\mathbf{x}_a - \mathbf{x})^t \mathbf{S}_a^{-1} (\mathbf{x}_a - \mathbf{x}) \quad \text{with} \quad \boldsymbol{\alpha}_i = \hat{\mathbf{x}}_i - (\mathbf{I} - \mathbf{A}_i)\mathbf{x}_{ai} = \mathbf{A}_i\mathbf{x}_i + \boldsymbol{\sigma}_i \ ,$$

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

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

which is characterized by the following CM and AKM:

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

In the equations above: x_i are the state vectors describing the atmospheres sounded by the instruments i=1,2 (FORUM or IASI-NG); Sn,i the CMs; Sa, the apriori CM of the apriori state vector x_a 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 substitute in the equations above Sn,2 with:

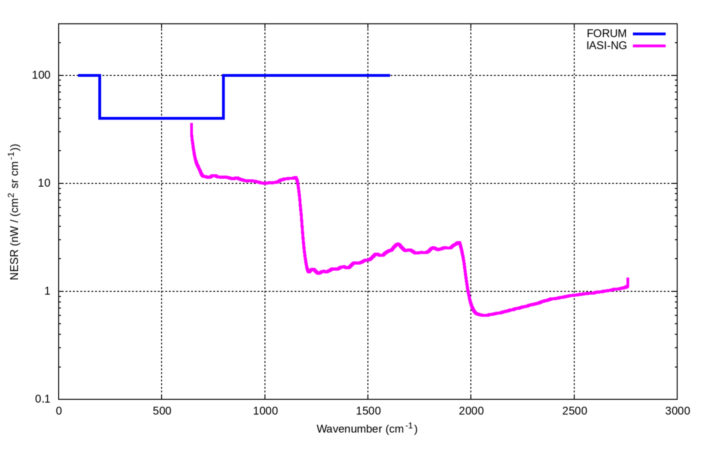
$$\mathbf{S}'_{n,2} = \mathbf{S}_{n,2} + \mathbf{A}_2 \mathbf{S}_M \mathbf{A}_2^t \ .$$

adding the coincidence error to IASI-NG measurement as in SR.

INSTRUMENTS:

	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}**
Spectral resolution	**0.25 cm^{-1}**	**0.5 cm^{-1}**
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-1A for IASI-NG)	

Summary of IASI-NG and FORUM main characteristics



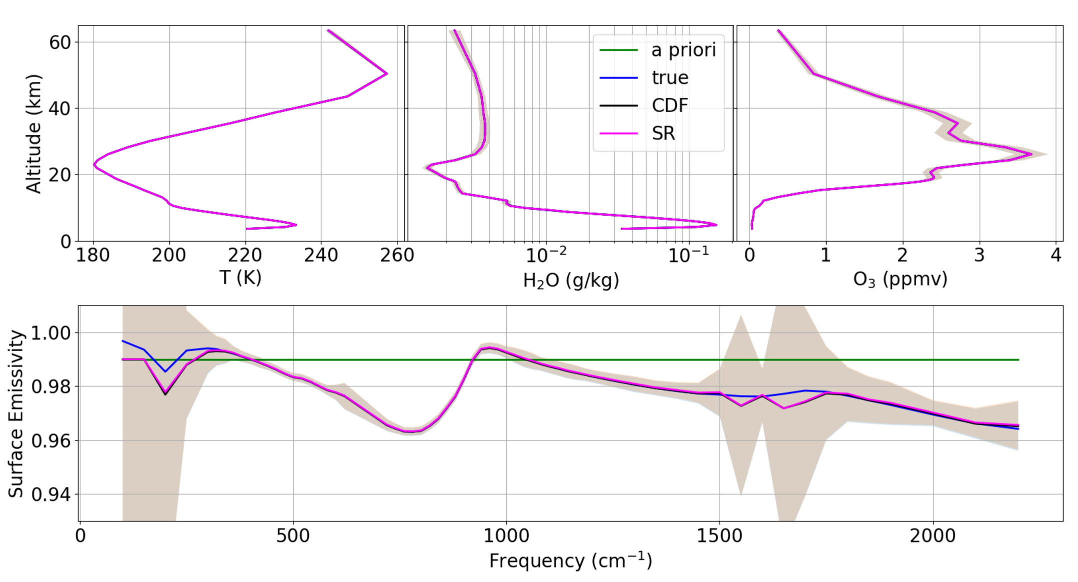
Summary of the assumed FORUM and IASI-NG Noise Equivalent Spectral Radiance (NESR)

TEST SETTINGS		INDIVIDUAL RETRIEVAL		SYNERGYSTIC METHODS	
		FORUM	IASI-NG	SR	CDF
Test		Assumptions to generate synthetic observations			
No mismatch (900 cases) **$x_1=x_2$**	State vector x_i	• $\mathbf{x_1}$ **(for T,Ts,H2O,O3)** true state vectors obtained applying a stochastic perturbation to their reference value X_0 consistent with $\mathbf{S_M}$/2 diagonal values. • $\mathbf{E_{S,i}}$ generated from the snow emissivity profile of Huang et al 2016			Input state vectors: $\hat{x}_1$ and $\hat{x}_2$
	A priori x_a	• $\mathbf{x_{ai}}$ **(for T,Ts,H2O,O3)** a priori state vector obtained applying a random perturbation to their reference value X_0 consistent with $\mathbf{S_a}$ • $\mathbf{E_{s,ai}}$ a priori=0.99 constant			$\mathbf{S_a}$
	Errors CMs used in the retrieval	$\mathbf{S_{y1}}$	$\mathbf{S_{y2}}$	$\mathbf{S_{y1}}$ **and** $\mathbf{S_{y2}}$	$\mathbf{S_f}$
	Retrieved $\hat{x}_i$	$\hat{x}_1$	$\hat{x}_2$	$\hat{x}_{sr}$	$\hat{x}_f$
mismatch (900 cases) **$x_1 \neq x_2$**	State vector x_i	observations simulated assuming $x_1 \neq x_2$, with x_1 and x_2 obtained as in the case of no mismatch			Input state vectors: $\hat{x}_{1,msm}$ and $\hat{x}_{2,msm}$
	A priori x_a	Simulated as in 'no mismatch' case			$\mathbf{S_a}$
	Errors CM used in the retrieval	$\mathbf{Sy_1}$	$\mathbf{Sy_2}$	$\mathbf{S_{y1}}$ **and** $\mathbf{S'_{y2} = S_{y2} + K_2 S_M K_2^t}$	$\mathbf{S_{f,msm}}$ **($\mathbf{S'_{n,2}}$ for IASI-NG)**
	Retrieved $\hat{x}_{i,\,msm}$	$\hat{x}_{1,msm}$	$\hat{x}_{2,msm}$	$\hat{x}_{sr,msm}$	$\hat{x}_{f,\,msm}$

RESULTS

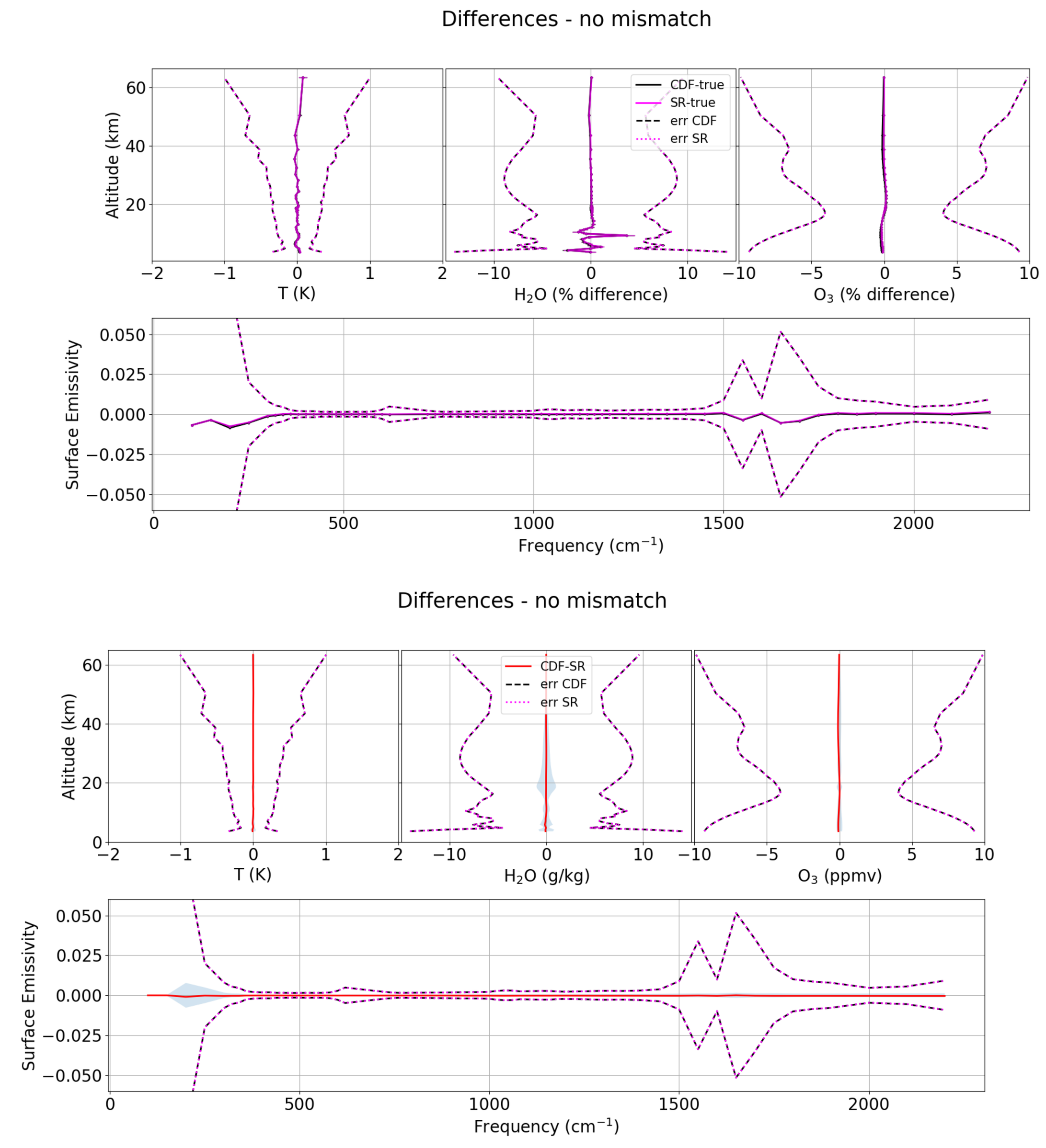
As explained in the Table above, in our study we carry out two sets of test retrievals emulating an idealized situation in which both FORUM and IASI-NG measure, with **perfect matching** and with a **mismatch in time and space**, for 900 times, the same portion of the Antarctic Plateau surface covered by coarse snow. At each measurement of the two tests the atmosphere and the surface temperature change slightly and randomly with respect to the reference scenario x_0.

Profiles - no mismatch

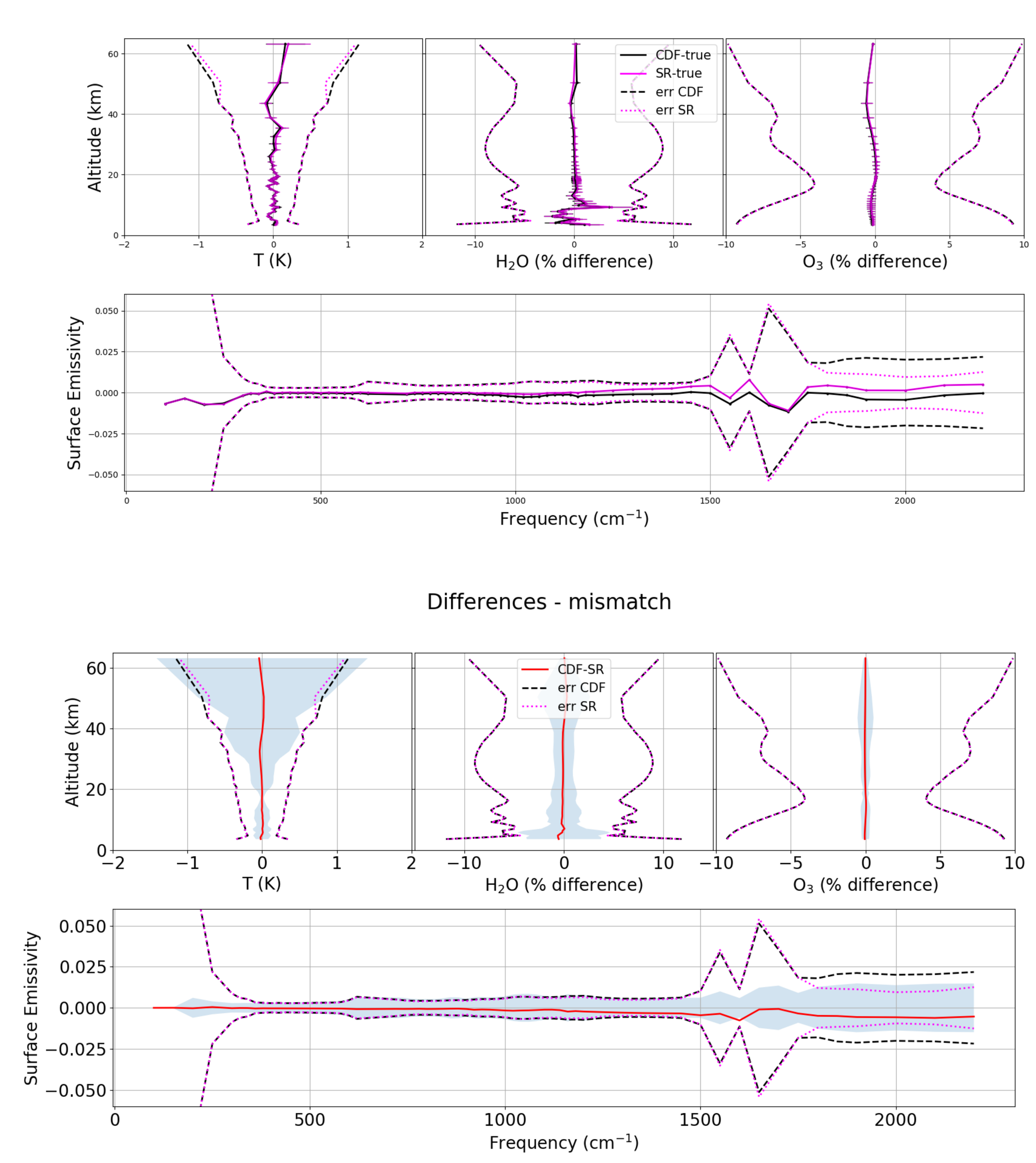


True (blue), average CDF (black), average SR (magenta), and a-priori (green) T, H_2O, O_3 and emissivity profiles for the no-mismatch test case. The shaded area represents the average total retrieval error.

After the 900 runs for both tests cases, we evaluate both the average of the differences between the synergistic/fused results and the true values used for the generation of synthetic observations. The average differences quantify the product's bias. The standard error of the average, i.e. the standard deviation of the differences divided by the square root of the number of trials ($\sqrt{900}$ = 30, in this case) is useful to evaluate whether the determined bias is statistically significant. We finally evaluate the average differences between fused and synergistic products and their standard deviation.



Average differences between CDF (black), SR (magenta) and true state parameters (top) and between CDF and SR products (bottom) for T, H2O, O3 and emissivity profiles in the no-mismatch test case. The dashed lines represent the CDF and SR average total errors, the error bars (top) represent the standard error of the average differences while the shaded areas (bottom) represent the standard deviation of the average differences .



Average differences between CDF (black), SR (magenta) and true state parameters (top) and between CDF and SR products (bottom) for T, H2O, O3 and emissivity profiles in the mismatch test case. The dashed lines represent the CDF and SR average total errors, the error bars (top) represent the standard error of the average differences while the shaded areas (bottom) represent the standard deviation of the average differences .

REFERENCES

S. Ceccherini, B. Carli and P. Raspollini, *Equivalence of data fusion and simultaneous retrieval*, Optics Express, **23**, 8476-8488 (2015).

S. Ceccherini, B. Carli, C. Tirelli, N. Zoppetti, S. Del Bianco, U. Cortesi, J. Kujanpääand R. Dragani, *Importance of interpolation and coincidence errors in data fusion*, Atmospheric Measurement Techniques, **11**, 1009-1017 (2018).

Huang, X., Chen, X., Zhou, D. K., and Liu, X.: *An Observationally Based Global Band-by-Band Surface Emissivity Dataset for Climate and Weather Simulations*, Journal of the Atmospheric Sciences, 73, 3541 – 3555,https://doi.org/10.1175/JAS-D-15-0355.1, 2016.

CONCLUSIONS

In this work we assess the performance of the SR and CDF techniques on the basis of synthetic measurements of the FORUM and the IASI-NG missions that are planned for launch in a few years, on two different polar satellites flying in loose formation.

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 more different results, the differences, however, are still within the retrieval error;
- from this study we can conclude that 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.