THE PROOF OF CONCEPT FOR PROFILES RETRIEVALS FROM CrIS RADIANCES USING A DETERMINISTIC METHOD

Prabhat K. Koner, Member, IEEE and Andy Harris
CICS/NOAA, 5830 University Research Court, College Park, MD-20740, USA
Email: Prabhat.koner@noaa.gov

ABSTRACT

Different aspects of the remote sensing inverse problem in operational environment are thoroughly discussed and various ambiguities in prevalent stochastic method namely optimal estimation method (OEM) are pointed out. To overcome these obscurities, a deterministic method namely regularized total least squares (RTLS) is proposed. For demonstration purposes, water vapor profiles retrievals from simulated Suomi-NPP CrIS hyperspectral measurements are considered. Synthetic CrIS radiances are generated using a line-by-line radiative transfer model (GENSPECT) with realistic radio-sonde profiles and US standard atmosphere as inputs. Our findings show that the current stochastic method, even with additional deterministic constraints (truncated singular value decomposition) applied on top of OEM, is unable to produce useful retrieval results, i.e., *posterior* error is more than the *a priori* error. In contrast, RTLS is able to produce deterministically unique results commensurate with the available information content in the measurements, which could result in a paradigm shift in operational satellite inversion.

Index Terms—: Regularized total least squares, Hyperspectral infrared sounding, Ill-conditioned inverse, Suomi-NPP CrIS.

INTRODUCTION

There are two distinct schools of thoughts in physical inverse model: (i) deterministic and (ii) stochastic [1-2]. Deterministic methods assumes that there is a true value of all individual retrieved parameter but associated with error, and it can be derived at individual measurement points using an objective function representing data misfit (measurement error) that is minimized in a given norm. The stochastic method assumes that all retrieved parameter values are random variables and as a result these parameters are retrieved for a set of measurement using Bayesian probability theory or one dimension variation based estimation (1d-var) or regression based least squares solution. According to the literature survey, most of the operational satellite retrievals are based on the stochastic method (*e.g.* [3]), where information of the state space parameters are obtained in terms of the ensemble mean and a little significance is given at an individual pixel level, which is most desirable. To overcome this problem, we propose in this work a radically different deterministic inverse algorithm, which based on the Regularized Total Least Squares (RTLS), for profiles retrievals from the satellite hyper-spectral Infrared (IR) measurements.

Among the existing deterministic methods, RTLS is the only one which inherently accounts the optimal regularization strength to be applied to the normal equation 1st order Newtonian inverse using all of the noise terms embedded in the residual vector [4]. RTLS has a good heritage in medical imaging field (*e.g.* [5]), but not at all in earth observation science (EOS) field. On the contrary, any traditional stochastic method does not include any constraint to prevent noise enhancement in the state space parameters from the existing noise in measurement space for an inversion with an ill-conditioned Jacobian. To stabilize the noise propagation into parameter space due to ill-conditioned jacobian, recent trend to use truncated singular vector decomposition (TSVD) approximation top of the existing stochastic method as a Band-Aid approach [6]. This process increased complexity, numerical calculations/noises and ambiguities as well as information loss due to discarding of lower singular values based components from the inversion. On the other hand, without reduction of the model resolution, RTLS injects additional information into inversion using Laplacian first derivative operator (LFDP) as a stabilizer [7]. LFDP forced the solution as the adjacent atmospheric parameters in a profile are close, which is less harmful than use of an *a priori* of the dynamic atmospheric parameters. RTLS solution is totally independent of the initial guess parameter of targeted retrievals and regularization is data driven. RTLS algorithm does not require any constraints including *a priori* data, and measurement error and *a priori* covariance matrices [8-9], which are impossible to be obtained in an operational retrieval system because the variability of atmosphere and oceanographic parameters is very high. From the deterministic point of view, one can scientifically argue that the truth is known if *a priori* and error in *a priori* are known. These are the input parameters for currently implemented stochastic inverse methods in operation [7].

One of the primary objectives of this work is to establish the benefit of the physical deterministic method in overcoming many operational hurdles in satellite inverse problem over the traditional stochastic method, namely optimal estimation

method (OEM). For example, measurement error covariance is an input for the stochastic method and it is difficult to construct a properly representative measurement error covariance matrix (not required for RTLS) for any satellite retrievals because of errors in the instrument, the forward model, radiative transfer (RT) spectral coefficients, and due to nonlinearity. In addition, line-by-line calculations have become increasingly popular for forward modeling of the RT function and a numerical (finite difference) Jacobian is accepted for such problems to avoid high errors due to rounding-off, quantization and integration of the complete derivative of the RT equation, but will still inevitably be contaminated by error [8]. To overcome the basic deficiency of OEM, some *ad hoc* unscientific assumptions are considered, *e.g.* radiance level bias correction based on model minus observation, tuning the error covariance for the sake of cloud contamination, generating *a priori* profiles using regression of same measurements, tuning the reference dataset with averaging kernel of the inverse model [10-11].

SIMULATION STUDIES

We will use simulated retrievals to show the *proof of concept*, because we have more control over the various parameters for numerical experimentation. Using such controlled tests one can eliminate different ambiguities associated with the operational environment to focus the attention on assessing the underlying characteristics of retrieval methods. A measurement can be considered equivalent to a model output if a compatible model exists. The model consists of a set of targeted parameters, which is embedded in a mathematical framework. Radiative transfer modeling (RTM) for atmospheric problem using line-by-line (LBL) calculation has reached a high level of maturity. First, we have calculated the spectra at a spectral resolution of 0.05cm^{-1} using the LBL model GENSPEC, for the standard atmospheric temperature and WV profiles, surface temperature of 300 K and surface emissivity of one. The CrIS channel radiance is given by a convolution of the ILS function with the monochromatic radiance at the entrance to the interferometer. It is a common practice to suppress ringing in the spectrum by apodizing the modulation function, artificially reducing the abruptness of the interferogram clipping by forcing the windowing function to tend smoothly to zero at the extremes. For the simulated study, it may be considered at the time of retrieval from CrIS along with Doppler shift and solar spectrum correction, simplified *sinc* ILS is considered to produce the equivalent CrIS measurement by convolving with the *sinc* ILS with simulated spectrum.

It is well known that many CrIS channels are subject to contamination by absorption due to various trace gases within any windows of WV retrievals. There are three alternatives in such a situation: (a) use all interfering trace gases in the forward model simulation, (b) model all gases but the approximated error is applied at the level of error covariance to which the other gases are uncertain, and (c) discard the channels that are affected by an amount more than assumed threshold value. It is very difficult to obtain correct profiles of all interfering gases at the time of measurement and computational cost increases if the first choice is considered. From a deterministic point of view, weighted measurement using variable values of elements in error covariance, as described in the second choice above, cannot remove channel errors where interfering gases increase the measurement error relative to the target parameter. In such a process, both ambiguities in retrieval and computational cost are increased. Thus, for this experiment, we have opted for a third choice, i.e., a threshold value of 100 for SNR is considered.

We searched the information-rich channels for WV profiles retrievals using the above-mentioned approach for different bands of CrIS measurements; i.e. 40 channels from long wave infrared (LWIR) region (700-960 cm^{-1}) for upper tropospheric WV information and 140 channels from MWIR region (1230-1625 cm^{-1}) for mid tropospheric WV information. We didn't consider any channels from the short wave IR (SWIR) region to avoid additional modeling of the effect of solar scattering during daytime. These selected channels may be somewhat incomplete for practical application but should be more than adequate for the intended purpose of comparing the inverse methods of RTLS and OEM.

Two different methods of RTLS and OEM and three different true profiles of TP-1 (realistic), TP-2 and TP-3 (sinusoidal profiles) are considered. We have used two initial guess (IG) profiles, one is constant (IG1) and one is realistic (close to TP-1, IG2) shown in black and magenta, respectively (Fig. 1). We used three different Monte Carlo noise realizations of 1% (SNR=100) in measurements, which is a somewhat conservative estimate for CrIS. Retrieval using OEM has been made considering 100% *a priori* variance, for the profiles solved from IG1 because all truths are within the range of 100% variance of the IG1. For TP1 solving from IG2, *a priori* variance of 30% is used. As is typical, the IG profile is considered the *a priori* profile and error covariance is calculated according to the noise added into simulated measurements. All three profiles are solved using RTLS from IG1 and OEM cannot solve any of these profiles from IG1 and also using a priori of IG2, which is close to truth. Second experiment was conducted by increasing the effective regularization strength of the OEM by changing the *a priori* variance to 1% instead of 100%, and keeping all other parameters fixed as previous experiment. OEM solutions for second experiment are within the domain of solution space due to high regularization, but are not at all good solutions (Fig. 2).

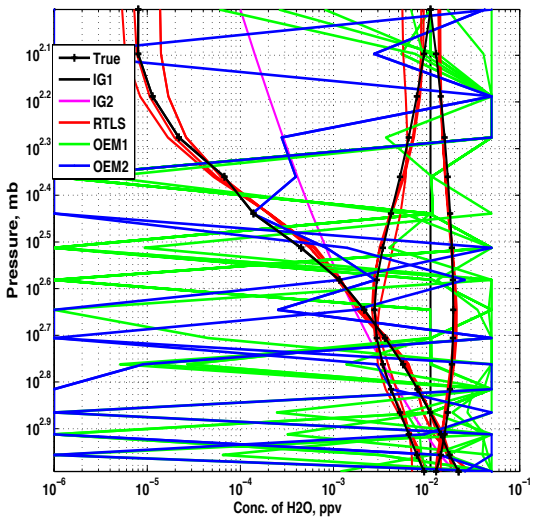

Fig. 1 *Three true profiles (+solid black); RTLS solution (red solid) for all true profiles from initial guess of IG1(black solid); OEM solution: OEM1(green solid) from IG1 for all true profiles and OEM2(blue) using IG2(magenta) for realistic profile.*

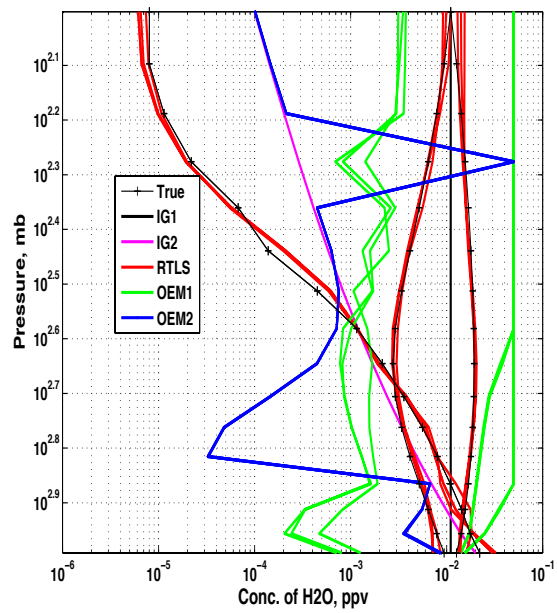

Fig. 2 *Same as Fig. 1 by tuning a priori variance of 1% instead of 100 or 30% in OEM solution.*

To address any potential concerns due to selection of unrealistic profiles in abovementioned experiments, another study is conducted using the Forecast Systems Laboratory (FSL) radio-sonde database (http://www.esrl.noaa.gov/raobs/intl/fsl_format-new.cgi) representative of the earth's atmosphere, in conjunction with other collocated in situ parameters. We have collected more than 400 profiles from this database to perform this simulation.

The simulation has been made on the grid of the individual radio-sonde profiles. The grid spacing is less than 100m in some profiles, which is difficult to solve due to limited information from such measurement. Thus, the problem is also solved using fixed coarser grids of 12. Once the synthetic measurements (using finer grid) are obtained, for inverse purposes, the RT calculation has been made at coarser grids, which also allows us to gain confidence in RTLS retrieval in the presence of forward model error.

As is observed in the abovementioned experiment, an ill-posed inversion is dependent regularization strength, not due to selection of *a priori* covariance. Even without understanding the root cause of the problem, a recent trend has developed to improve OEM retrievals by applying additional regularizations using truncated SVD on top of the stochastic methods (OEM-T) to reduce the oscillations in the resultant solutions. Thus, OEM is replaced by OEM-T in this study using 5 highest singular values of the inverted matrix and 10% of *a priori* covariance. Fig. 3 shows the double averaging error of the weighted RMSE of RTLS and OEM-T for the finer and coarser grids. As the values of the WV profile varies by more than two orders of magnitude we have considered the percentage of error as ($\delta x/x$) for individual points as described in report [12]. The first averaging is made along the altitudinal grid of 2km and then the second averaging is made over more than 400 profiles for individual 2km retrieval grid similar to the report [13].

DISCUSSIONS

Fig. 1 illustrates that OEM is not an appropriate method for inherently ill-conditioned satellite retrieval problems. This can be explained from the deterministic viewpoint by arguing that using a combination of *a priori* and measurement error covariances is unable to regularize the inverse problem adequately. The underlying cause is its inability to reduce the noise propagation from measurement space to state space for such an ill-conditioned inversion. Put another way, without an optimally regularized inverted matrix, even choosing *a priori* top to the truth may yield out-of-range solutions for a highly ill-conditioned inversion. On the other hand, RTLS shows that there exist unique solutions for such problems using deterministic inverse methods, within the limit of information content. We purposely used simulated retrievals to demonstrate the comparative performance of our proposed method over prevalent method because simulation eliminates all other operational ambiguities. Simulations have the benefit of allowing us to either exclude or include regulated operational problems (calibration, forward model error, cloud detection, etc.) and concentrate on the performance of the inversion methods itself.

The information content for RTLS at finer grids is very poor in high altitude (see Fig. 3), which makes it impossible to solve at this region but the corresponding error is still not

more than *a priori* error. It is observed from this study that the weighted altitudinal errors of coarser grids RTLS ('red circle') is mostly similar or higher than the same at finer grid ('red') at high altitude. This implies that the forward model error for such retrievals has significant impact on retrieval error as compared to the advantages of using a coarser grid due to less degree of freedom as mentioned earlier.

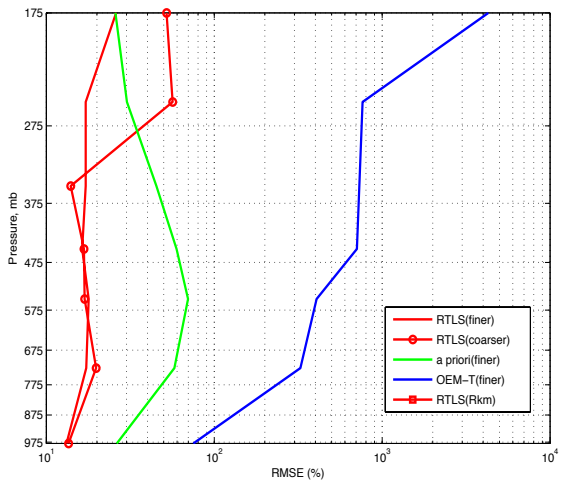

Fig. 3 *The weighted retrieval error along the altitude (average of 2km grid) of more than 400 profiles for RTLS with finer grid (red), RTLS with coarser grid (red circle), IG with finer grid (green) and OEM-T with finer grid (blue) .*

CONCLUSIONS

This study concludes that satellite retrieval or nonlinear RT inverse does not follow the covariance based stochastic inverse theory. As is observed that OEM solutions are improved (within domain of solution space) if a lower *a priori* covariance is used than the expected value. Based on the theoretical discussions and simulated experiments, we conclude that RTLS is one of the most suitable inverse methods, which can be used for any highly non-linear remote sensing problem for extraction of quantitative information. On the other hand, stochastic inversion techniques are unable to produce unambiguous scientific information and often yield more errors than in the initial guess.

REFERENCES

1. J.L. Mead, "Parameter estimation: A new approach to weighting a priori information," Inv. Ill-Posed Problems, vol. 15, pp.1-21, 2007.
2. R. Zhang, M. K Sen, S. Phan and S. Srinivasan, "Stochastic and deterministic seismic inversion methods for thin-bed resolution," J. Geophys. Eng., vol. 9, pp. 611–618 , 2012.
3. T. Clarmann and U. Grabowski, "Elimination of hidden a priori information from remotely sensed profile data," Atmos. Chem. Phys., vol. 7, pp. 397–408, 2007.
4. P.K. Koner, A.R. Harris and E. Maturi. "A physical deterministic inverse method for operational satellite remote sensing: an application for sea surface temperature retrievals," In press, IEEE Trans. Geosci. Remote Sens.
5. V.Z. Mesarovic, N.P. Galatsanos and A.K. Katsaggelos, "Regularized constrained total least squares image restoration," IEEE Trans. On Image Process., vol. 4, pp. 1096-1108, 1995.
6. ATBD for CrIS, Vol II, EDR, GSFC JPSS CMO, November 16, 2012.
7. P.K. Koner and J.R. Drummond, "A comparison of regularization techniques for atmospheric trace gases retrieval," JQSRT, vol. 109, pp. 514-26, 2008.
8. P.K. Koner and J.R. Drummond, "Atmospheric trace gases profile retrievals using nonlinear regularized total least squares method," JQSRT, vol. 119, pp. 2045-59, 2008.
9. P.K. Koner, A. Battaglia and C. Simmer, "A Rain Rate retrieval algorithm for Attenuating Radar Measurement," J. Appl. Meteo & Clima., vol. 49, pp. 381-93, 2010.
10. P.K. Koner and A.R. Harris, "A deterministic inversion technique for sea surface temperature retrieval from MODIS radiances," Proc. SPIE 9459, Ocean Sensing and Monitoring VII, 94590Y (May 19, 2015); doi: 10.1117/12.2179868.
11. P.K. Koner, A.R. Harris and P. Dash. "A deterministic method for profiles retrievals from hyperspectral satellite measurements," undergoing review, IEEE Trans. Geosci. Remote Sens..
12. E.S. Maddy and C.D. Barnet, "Vertical resolution estimates in version 5 of AIRS operational retrievals," IEEE Trans on Geoscience and Remote Sens., vol. 46, pp. 2375-84, 2008.
13. N.R. Nalli, C. D. Barnet, A. Reale, et. al., "Validation of satellite sounder environmental data records: Application to the croos-track Infrared Microwave sounder suite," J. Geophys. Res. Atmos., vol. 118, pp. 628-43, 2013.