Results from a Quick Look at Applying the 'Compact Phase Space Retrieval' (CPSR) Analysis to Synthetic O_3 Retrieval Profiles

Arthur P. Mizzi[1], Matthew S. Johnson[2], Aaron R. Naeger[3], Daven K. Henze[4], and Jeffrey L. Anderson[5]

1 USRA NPP - NASA Ames Research Center, Earth Science Division, Mountain View, CA, USA
2 NASA Ames Research Center, Earth Science Division, Mountain View, CA, USA
3 UA-Huntsville/NASA Short-Term Predition Research and Transition Center, Huntsville, AL, USA
4 UC-Boulder, Department of Mechanical Engineering, Boulder, CO, USA
5 NCAR/CISL - Data Assimilation Research Section, Boulder, CO,USA

arthur.p.mizzi@nasa.gov

1. WRF-Chem/DART: An Introduction

WRF-Chem/DART is a community resource for realtime chemical weather forecasting/data assimilation research and operations. It couples the Weather Research and Forecasting model (WRF) with online chemistry (WRF-Chem) and the Data Assimilation Research Testbed (DART). DART has been modified to include assimilation of *in situ* and remote/satellite observations of atmospheric composition. Specifically, WRF-Chem/DART:

- Assimilates MOPITT and IASI CO total/partial column and/or profile retrievals;
- Assimilates IASI O_3 total/partial column and/or profile retrievals;
- Assimilates OMI O_3, NO_2, and SO_2 total/partial column and/or profile retrievals;
- Assimilates TROPOMI CO, O_3, NO_2, and SO_2 total/partial column and/or profile retrievals;
- Assimilates TEMPO O_3 and NO_2 total/partial column and/or profile synthetic retrievals;
- Assimilates MODIS AOD total column retrievals;
- Assimilates AirNOW *in situ* observations;
- Assimilates Retrieval profiles as raw retrievals (RETRs) or "compact phase space retrievals" (CPSRs);
- Uses state variables localization;
- Constrains emissions with the State Augmentation Method (SAM); and
- Includes a real-time scripting system;

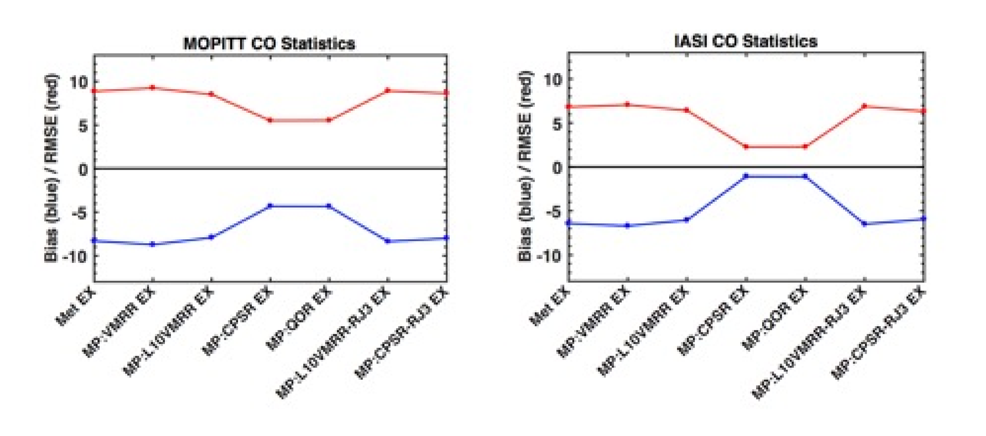

Figure 1: *Forecast skill scores for meteorological assimilation only (Met EX), Met EX with assimilation of raw retrievals (MP:VMRR EX and MP:L10VMRR EX) and Met EX with assimilation of CPSR (CPSR EX). The remaining experiments are not relevant to this presentation.*

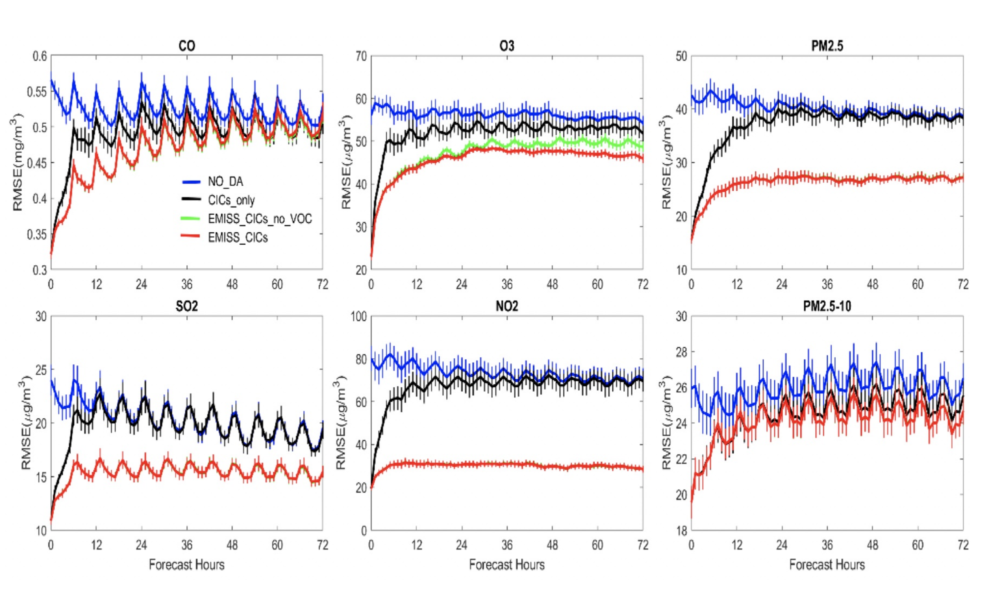

Figure 2: *Forecast skill scores for meteorological assimilation only (NO DA), assimilation of in situ observations (CICs-only), and CICs-only with emissions adjustment (EMISS-CICs). EMISS-CICs-no-VOC is not relevant to this presentation.*

Figure 1 from Mizzi et al. (2016) shows that using WRF-Chem/DART with assimilation of MOPITT CO CPSRs in MP:CPSR EX produces ~ 40% improvement in forecast skill compared to not using CPSRs in MET EX, MP:VMRR EX, and MP:L10VMRR EX. That improvement occurs due to the filtering effect of the CPSR compression transform on the retreival errors. Figure 2 from Ma et al. (2019) shows that constraining emissions with WRF-Chem/DART and assimilation of *in situ* measurements of criteria pollutants in EMISS-CICs improves forecast skill and increases the predictability time from 6-12 hrs to 48-72 hrs compared to not adjusting the emissions in CICs-only.

Together Figs. 1 and 2 illustrate the type of chemical weather forecast skill improvment we expects to see from assimilating retrievals from satellites like TEMPO.

2. The Compact Phase Space Retrieval (CPSR) Algorithm

Mizzi et al. (2016) introduced the CPSR algorithm for efficient storage and assimilation of full retrieval profiles. The CPSR algorithm uses: (i) a "compression transform" to remove redundant information in the retrieval profile; and (ii) a "rotation transform" to account for error cross-correlations in the observation error covariance matrix. Those transforms are based on a singular value decomposition (SVD) of the retrieval equation averaging kernel, and an SVD of the compressed observation error covariance. The mathematical derivation is as follows:

The quasi-optimal form of the retrieval equation is

$$\mathbf{y_r} - (\mathbf{I} - \mathbf{A})\mathbf{y_a} - \varepsilon = \mathbf{A}\mathbf{y_t}. \quad (1)$$

where $\mathbf{y_r}$ is the retrieval profile (dimension n), $\mathbf{I}$ is the identity matrix (dimension $n \times n$), $\mathbf{A}$ is the averaging kernel (dimension $n \times n$), $\mathbf{y_a}$ is the retrieval prior profile (dimension n), ε is the measurement error in retrieval space (dimension n) with measurement error covariance $\mathbf{E_m}$ (dimension $n \times n$), and $\mathbf{y_t}$ is the true atmospheric profile (unknown; dimension n). Let the SVD of $\mathbf{A}$ be $\mathbf{A} = \Phi\Delta\Theta^T$. Set the singular vectors associated with the zero singular values to zero. Then transform Eq. 3 with Φ^T to get

$$\Phi^T(\mathbf{y_r} - (\mathbf{I} - \mathbf{A})\mathbf{y_a} - \varepsilon) = \Delta\Theta^T\mathbf{y_t}. \quad (2)$$

The compressed form of $\mathbf{E_m}$ is $\Phi^T\mathbf{E_m}\Phi$. Next, let the SVD of the compressed error covariance be $\Phi^T\mathbf{E_m}\Phi = \Omega\Sigma\Psi^T$. Set the singular vectors associated with the zero singular vlues to zero. Finally, transform Eq. 3 with Ω^T and scale the result by the inverse square of the associated singular values to get

$$\Sigma^{-1/2}\Omega^T\Phi^T(\mathbf{y_r} - (\mathbf{I} - \mathbf{A})\mathbf{y_a} - \varepsilon) = \Sigma^{-1/2}\Omega^T\Delta\Theta^T\mathbf{y_t}. \quad (3)$$

Eq. 3 is the compressed and rotated form of the quasi-optimal retrieval equation. The analogous form of $\mathbf{E_m}$ is the identify matrix.

Due to the rank deficient nature of $\mathbf{A}$, the number of observations to be assimilated from Eq. 3 is reduced by $1 - \frac{r}{n}$ where r is the rank of $\mathbf{A}$. Since $r << n$, the storage and computational saving from using CPSRs can be substantial.

3. Application of CPSRs to TEMPO

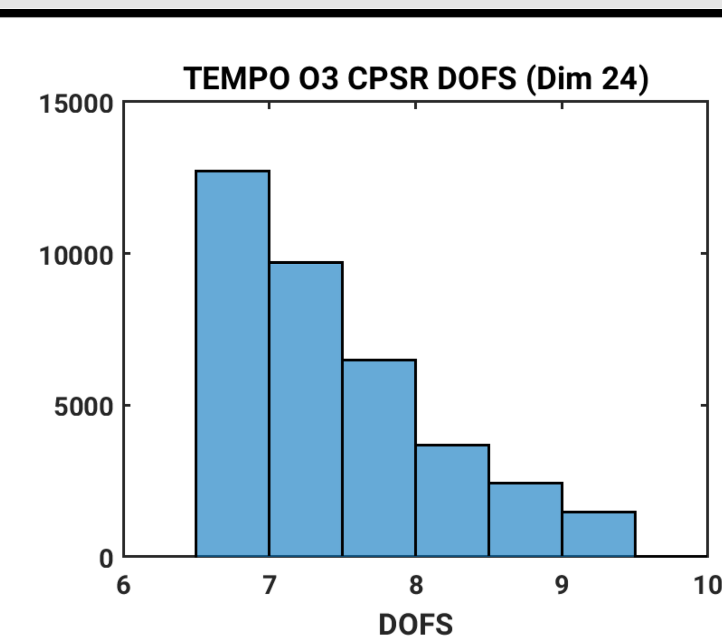

Figure 3: *Histogram of the synthetic TEMPO O_3 SVD-based DOFS.*

For this analysis, we use one synthetic TEMPO retrieval granule based on a GEOS-CF run for July 10, 2020. The granule contains 512 xtrack points of which we use the center 358 points, 123 mirror step points of which we use the center 102 points, and 24 vertical layers (35,516 profiles). Since the CPSR benefits are related to the rank of the averaging kernel, Fig. 3 shows a histogram of the SVD-based degrees of freedom of signal (DOFS) for the subject granule. Since the DOFS mode is between 6.5 and 7.0, and the storage/computational savings is proprtional to $1 - \frac{r}{n}$, the potential CPSR savings is ~ 70% for a DOFS of 7.0.

As an illustration, Figure 4(A) shows the mean reported averaging kernel profiles for the six lowest elements of the retrieval profile, and Fig. 4(B) shows the same mean averaging kerenl profiles after the forward and reverse CPSR compression transform. The two figure are identical, demonstrating that no information is lost by storing and/or assimilating the retrieval profiles in phase space. We obtain the same results for all other averaging kernel profiles.

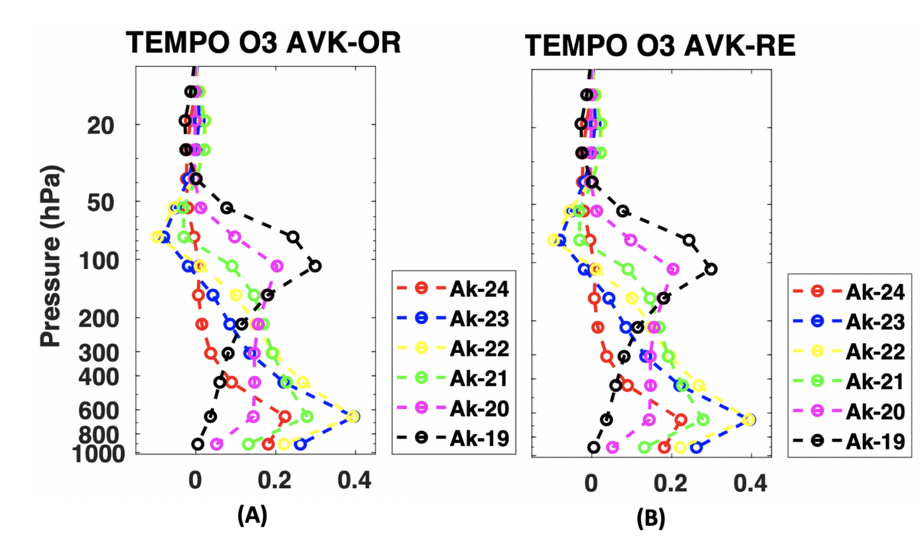

Figure 4: *Mean averaging kernels for the six lowest elements of the retrieval profile. Panel (A) is as reported. Panel (B) is after the forward and reverse CPSR compression transform.*

Figure 4 suggests that the greatest sensitivities are in the vicinity of 100 hPa and 700 hPa. However, due to the rank deficient nature of the averaging kernel, the vertical sensitivity structures in Fig. 4 can be misleading.

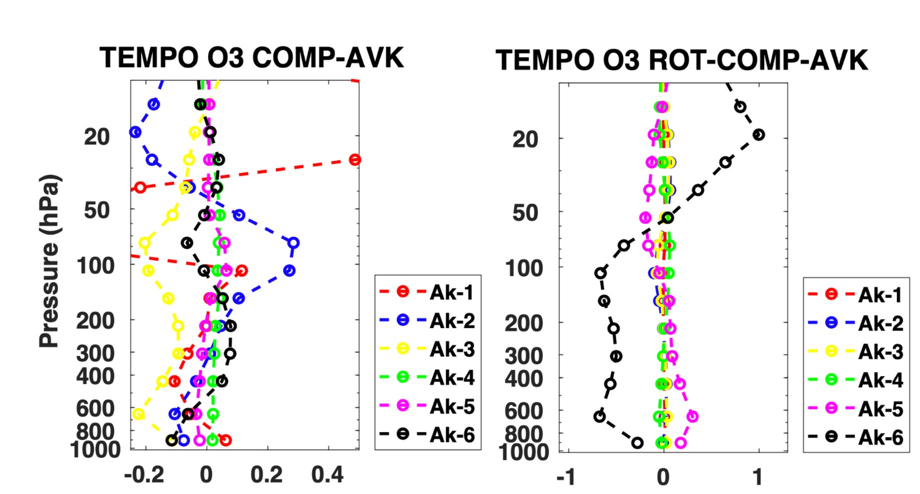

Figure 5: *Mean averaging kernel profiles after the CPSR transform. Panel (A) shows the compressed profiles, i.e., the profiles after removing the redundant information. Panel (B) shows the compressed/rotated profiles, i.e., the profiles after removing redundant information and accounting for the associated retrieval error.*

As explained in Mizzi at al. (2016; 2018), one should examine the compressed/rotated averaing kernel to understand the true sensitiviies, i.e., the sensitivities after removing redundant information and accounting for the retrieval error associated with each profile.

Figure 5 shows the compressed averaging kernel profiles in panel (A) and the compressed/rotated profiles in panel (B). The compressed profiles show that the dominant sensitivities are at 100 hPa and above. However, after considering the associated errors, Fig. 5 shows that the dominant modes have sensitivity throughout the troposphere (Ak-6) and in the lower troposphere at 700 hPa and below (Ak-5).

Finally, Fig. 6 shows the storage memory requirments for an example TEMPO granule with/without applying the CPSR transform.

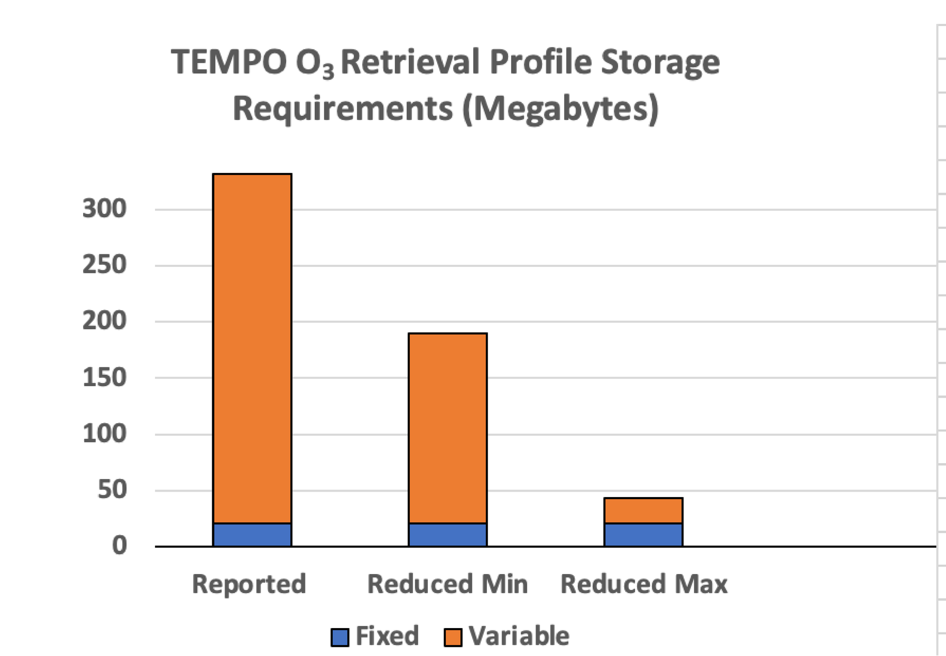

Figure 6: *Storage memory requirements in megabytes (MB) for 32-bit words and a TEMPO O_3 retrieval profile granule. 'Fixed' refers to variables whose memory requirements cannot be reduced. 'Variable' refers to variables whose memory requirements can be reduced. The 'Max Reduction' analysis assumes that the retrieval, apriori, and apriori error profiles and the averaging kernel, retrieval error covariance, and measurement error covariance matrices can be reduced. The 'Min Reduction' analysis assumes that only the retrieval profile and the averaging kernel matrix can be reduced*

From Fig. 6 we see that one TEMPO O_3 retrieval profile granule requires ~ 330 MB. With application of the CPSR transform, that can be reduced to between ~ 43 MB and ~ 190 MB.

4. More Information

For more information on WRF-Chem/DART. chemical data assimilation, CPSRs generally, or CPSRs as applied to TEMPO O_3 retrieval profiles, contact Dr. Arthur P. Mizzi by e-mail at arthur.p.mizzi@nasa.gov or by phone at 303-903-5544.

5. References

Mizzi, A. P., A. F. Arellano, D. P. Edwards, J. L. Anderson, and G. G. Pfister: 2016, Assimilating compact phase space retrievals of atmospheric composition with WRF-Chem/DART: A regional chemical transport/ensemble Kalman filter data assimilation system. *Geosci. Model Dev.*, **9**, 965-978.

Mizzi, A. P., D. P. Edwards, and J. L. Anderson: 2018, Assimilating compact phase space retrievals (CPSRs): comparison with independent observations (MOZAIC *in situ* and IASI retrievals) and extension to truncated retrieval profiles. *Geosci. Model Dev.*, **11**, 3727-3745.