Reduced Dimensionality Analysis of TEMPO Ozone (O_3) Profile Retrievals Using the Compact Phase Space Retrieval (CPSR) Algorithm

Arthur P. Mizzi[1,2,3], Junsung Park[4], Xiong Liu[4], Aaron R. Naeger[5], Michael Newchurch[6], Barry Lefer[7]

1 NASA Ames Research Center/BAERI, Moffett Field, CA USA
2 NOAA Chemical Systems Laboratory/CIRES, Boulder, CO USA
3 University of Colorado - Boulder, Department of Mechanical Engineering, Boulder, CO, USA
4 Harvard Smithonian Center for Astrophysics, Cambridge, MA USA
5 NASA Short-term Prediction Research and Transition Center, Huntsville, AL USA
6 University of Alabama - Huntsville, Department of Atmospheric Sciences, Huntsville, AL USA
7 NASA Headquarters, Washington D.C. USA

arthur.p.mizzi@nasa.gov

1. Chem-DART: An Introduction

Chem-DART is a community resource for realtime atmospheric composition data assimilation and emissions estimation research and operations. It extends NCAR's Data Assimilation Research Testbed (DART) to include assimilation of a broad range of atmospheric chemistry *in situ* (surface, tower, and profiler) and remote (profiler and satellite) observations. Chem-DART's capabilities have been well documented. At the NASA Ames Research Center, we are expanding/applying Chem-DART to:

- Integrate Chem-DART into the non-linear, non-Gaussian version of DART called the Quantile Conserving Ensemble Filtering Framework (QCEFF);
- Interface Chem-DART with NASA JPL's MOMO-Chem/LETKF, NASA GMAO's GEOS-CF and NU-WRF-Chem; and SJSU's WRF-SFIRE systems;
- Assimilate GOES LFCs, AOD total column retrievals, and JSAHs; VIIRS PM and AOD total column retrievals; and TROPOMI AOCHs.
- Prepare of a 20-yr Tropospheric Regional Atmospheric Composition and Emissions Reanalysis (2005-2024) (TRACER-I) for the continental US during the ozone and wildfire seasons (April - September).
- Determine whether assimilation of satellite observations (including height-dependent AOD) can recover wildfire plume injection heights as well as other emissions parameters.

2. Compact Phase Space Retrievals (CPSRs)

Mizzi et al. (2016) introduced the CPSR algorithm for efficient storage and assimilation of profile retrievals. The CPSR algorithm uses: (i) a 'compression transform' to remove redundant information from the profile retrieval; and (ii) a 'rotation transform' to normalize the compressed averaging kernel and account for error cross-correlations from the observation system error covariance. The mathematical formalism follows:

The retrieval equation is

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

where $\mathbf{y_r}$ is the profile retrieval (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 profile retrieval prior (dimension n), ε is the observational system error in retrieval space (dimension n) with error covariance $\mathbf{E_m}$ (dimension $n \times n$), and $\mathbf{y_t}$ is the true atmospheric profile (unknown; dimension n).

After subtracting the retrieval prior and error terms ($(\mathbf{I}-\mathbf{A})\mathbf{y_a}$ and ε) from both sides of Eq. 1, we get the 'quasi-optimal' form of the retrieval equation

$$\mathbf{y_r} - (\mathbf{I}-\mathbf{A})\mathbf{y_a} - \varepsilon = \mathbf{A}\mathbf{y_t}. \quad (2)$$

Let the singular value decomposition (SVD) of $\mathbf{A}$ be $\mathbf{A} = \mathbf{\Phi}\mathbf{\Delta}\mathbf{\Theta}^T$ and set the singular vectors associated with the zero singular values to zero. Then, we can transform Eq. 2 with $\mathbf{\Phi}^T$ to get

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

This is the 'compressed' 'quasi-optimal' retrieval equation where $\mathbf{\Delta}\mathbf{\Theta}^T$ is the 'compressed' form of the averaging kernel, and the 'compressed' form of the error term ($\mathbf{E_m}$) is $\mathbf{\Phi}^T\mathbf{E_m}\mathbf{\Phi}$. Next, let the SVD of the 'compressed' error covariance be $\mathbf{\Phi}^T\mathbf{E_m}\mathbf{\Phi} = \mathbf{\Omega}\mathbf{\Sigma}\mathbf{\Psi}^T$. Again, set the singular vectors associated with the zero singular values to zero. Finally, transform Eq. 3 with $\mathbf{\Omega}^T$ and scale the result by the inverse square root of the associated singular values to get

$$\mathbf{\Sigma^{-1/2}}\mathbf{\Omega}^T\mathbf{\Phi}^T(\mathbf{y_r} - (\mathbf{I}-\mathbf{A})\mathbf{y_a} - \varepsilon) = \mathbf{\Sigma^{-1/2}}\mathbf{\Omega}^T\mathbf{\Delta}\mathbf{\Theta}^T\mathbf{y_t}. \quad (4)$$

Eq. 4 is the 'compressed', 'rotated', and 'normalized' form of the 'quasi-optimal' retrieval equation. The analogous form of $\mathbf{E_m}$ is the identify matrix.

For this presentation, we apply the CPSR analysis to TEMPO and TROPOMI O_3 profile retrievals from March 29, 2024. For TEMPO, we used all files from S0006. For TROPOMI, we used all available files. The data was filtered for coverage of the continental United States (CONUS). Here we present an analysis of the latitudinal dependence of the averaging kernel characteristics. We plan to study the TEMPO temporal dependencies in future work.

Due to the rank deficiency of $\mathbf{A}$, the number of observations to be assimilated can be 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 significant. Figure 1 shows degrees of freedom for signal (DOFs) histograms for TEMPO and TROPOMI O_3 profile retrievals.

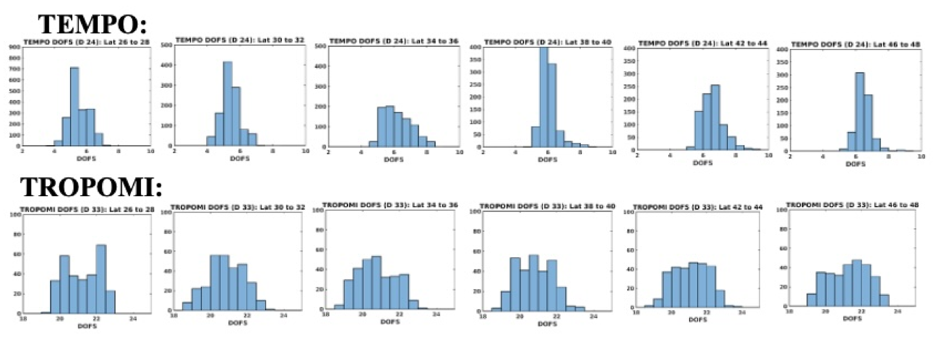

Figure 1: *Ozone profile retrieval averaging kernel degrees of freedom for signal (DOFs) histograms as a function of latitude band. TEMPO is in the upper row, and TROPOMI is in the lower row.*

Notice how the TEMPO DOFS ranges between 4 and 5 for the southern latitudes and increase northward reaching maxima between 6 and 7 for the northern latitudes. The TROPOMI DOFS ranges between 20 and 21 independent of the latitude. The TROPOMI DOFS is unusually large due to the mode one singular values. These values are suspect and under investigation.

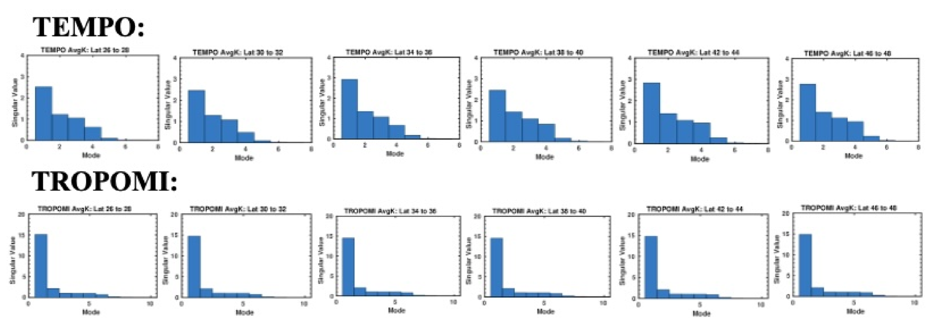

Figure 2: *Ozone profile retrieval averaging kernel mean singular values as a function of mode number and latitude. TEMPO is in the upper row, and TROPOMI is in the lower row.*

Figure 2 shows that TEMPO has four to five dominant modes with singular values ranging between 1 and 3. TROPOMI has one very dominant mode with a singular value of ∼15 (which is suspect), but a total of five to six dominant modes. The TEMPO singular values increase from south to north while the TROPOMI singular values remain relatively constant. The number of dominant modes equals the rank of the associated averaging kernel. Therefore, dimensional reduction can provide an ∼77% savings for TEMPO and an ∼81% savings for TROPOMI. We can also leverage the averaging kernel rank deficiencies for dimensionally reduced emissions estimation which is part of our future work.

3. Dimensional Reduction of the TEMPO and TROPOMI Averaging Kernels

In the previous section, we introduced the CPSR algorithm and documented the rank deficiency of the TEMPO and TROPOMI averaging kernels. Following Mizzi et al. (2016; 2018), we can remove the linear dependencies and make the transformed averaging kernels full rank. The advantage of the associated transformation is that it identifies the 'true' vertical sensitivities and saves computation time and storage space.

Figures 3 through 7 show TEMPO in the upper row and TROPOMI in the lower row. The columns show southern latitude bands on the left, middle latitude bands in the center, and northern latitude bands on the right. Figures 3 and 4 show vertical profiles of the 'raw' averaging kernels. Figure 3 shows Ak-1 to Ak-6, and Fig. 4 shows Ak-7 to Ak-12. Ak-1 is for the retrieval element closest to the surface. Note that the TEMPO vertical grid is ordered from top to bottom so Ak-24 is closest to the surface. Our discussion will use the TROPOMI ordering, i.e., we will refer to Ak-1 as being closest to the surface.

Figure 3 shows that TEMPO has peak sensitivities in the lower troposphere (500 - 700 hPa) for the southern latitudes. In the middle and northern latitudes, the sensitivities strengthen, have a bimodal peak, and move upwards with extrema in the mid- to upper troposphere (200 - 500 hPa) and the upper stratosphere (20 - 50 hPa). TROPOMI is similar in that the sensitivities increase with increasing latitude. However, in the southern latitudes, TROPOMI has a bimodal peak with a primary peak at 50 - 100 hPa and a secondary peak at 400 - 700 hPa. Moving northward, the TROPOMI peaks merge, descend, and strengthen with single peak at 150 - 300 hPa. Figure 4 shows similar changes for TEMPO, but TROPOMI has a single peak (20 - 10 hPa) in the southern latitudes which becomes bimodal in the middle and northern latitudes with peaks at 20 - 70 hPa and 150 - 300 hPa.

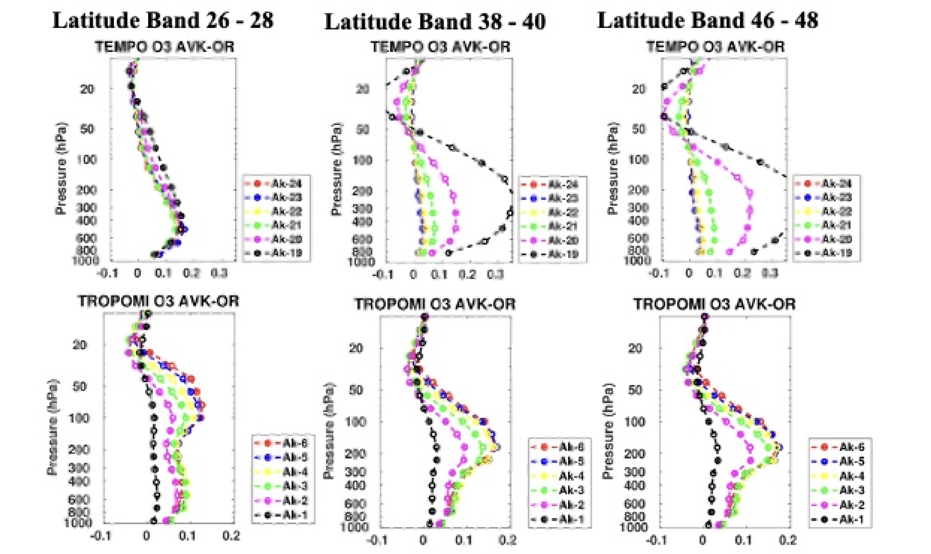

Figure 3: *TEMPO (upper row) and TROPOMI (lower row) raw averaging kernel profiles (Ak-1 to Ak-6) displayed as a function of latitude bands. The southern most bands are on the left and the northern most are on the right. Ak-1 corresponds to the retrieval element closest to the surface. Note that the TEMPO kernels are labeled as Ak-24 to Ak-19 because the TEMPO vertical grid is from top to bottom. Therefore, TEMPO Ak-24 corresponds to TROPOMI Ak-1, TEMPO Ak-23 to TROPOMI Ak-2, and so forth. Our discussion uses the TROPOMI labeling convention.*

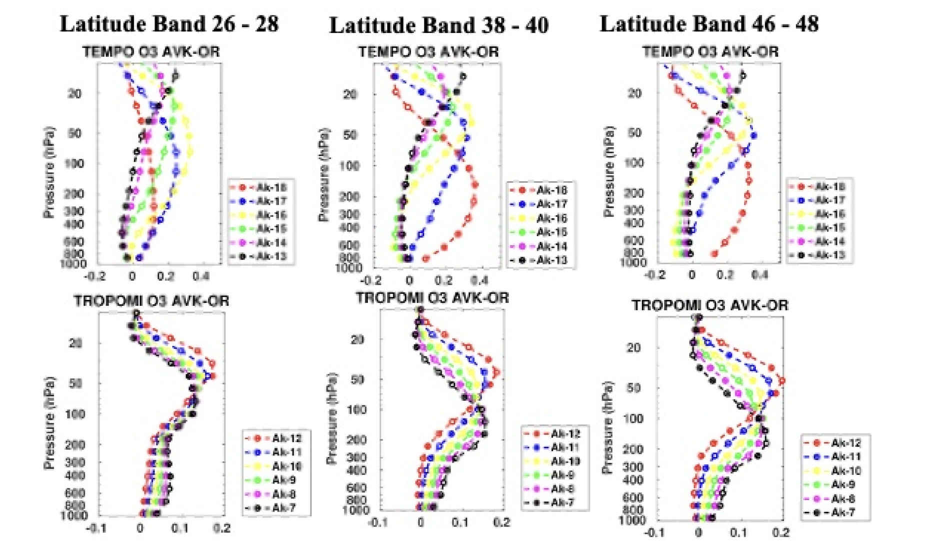

Figure 4: *Same as Fig. 3 except for TEMPO Ak-18 to Ak-13 and TROPOMI Ak-7 to Ak-12. See Fig. 3 caption for an explanation of kernel labeling convention.*

As explained previously, the raw averaging kernel rank deficiencies makes interpreting their sensitivities difficult due to the linear dependencies. We can use the left singular vectors of the averaging kernels to transform Eq. (2) to remove those dependencies. The 'compressed' averaging kernel in Eq. (3) shows that actual vertical sensitivities.

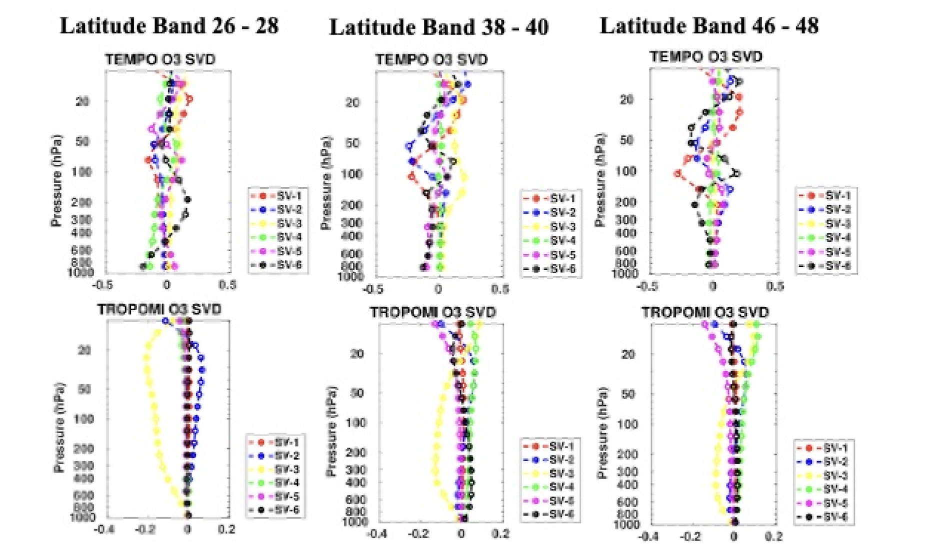

Figure 5: *Same as Fig. 3 except for the averaging kernel left singular vectors.*

Figures 5 and 6 show the averaging kernel left singular vectors (Fig. 5) and 'compressed' averaging kernels (Fig. 6). The organization of these figures is the same as in Figs. 3 and 4. Figure 5 shows that the TEMPO left singular vectors have structure throughout the troposphere while the TROPOMI averaging kernels are uniform with height except in the mid- to upper stratosphere. Figure 6 shows that the 'compressed' averaging kernel sensitivities generally increase from south to north. In the middle and northern latitudes, TEMPO shows sensitivities for Ak-1 to Ak-3 throughout the troposphere. TROPOMI Ak-1 shows peak sensitivities at 20 hPa and 70 hPa with weak sensitivities throughout the troposphere (250 - 800 hPa).

The 'compressed' averaging kernels in Figure 6 are full rank, but they do not consider the retrieval solution uncertainties. Therefore, the associated sensitivities can be misleading because the effective sensitivities could be more dependent on 'compressed' averaging kernel elements with moderate sensitivity and low uncertainty (as opposed to elements with high sensitivity and high uncertainty). The 'rotated and compressed' averaging kernels in Fig. 7 account for the uncertainties.

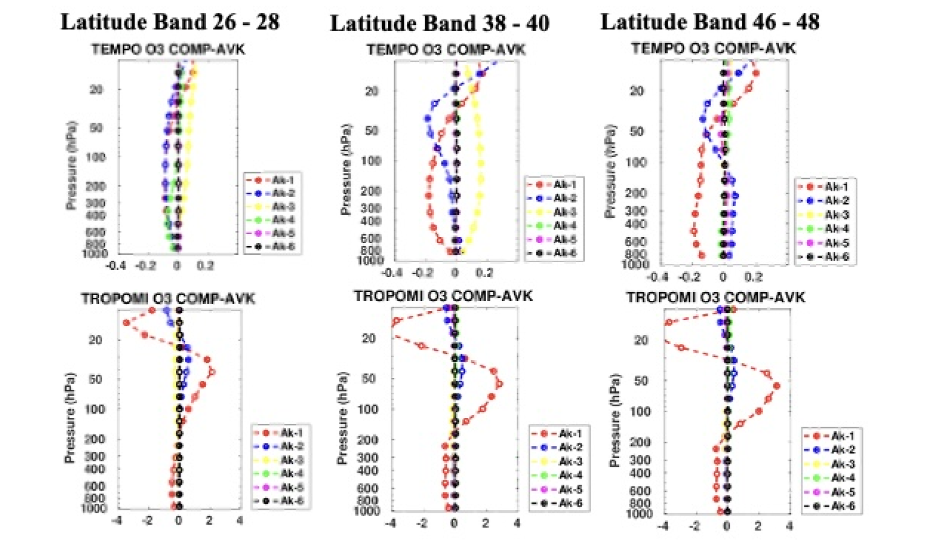

Figure 6: *Same as Fig. 3 except for the 'compressed' averaging kernels.*

Figure 7 shows that TEMPO and TROPOMI generally have the strongest sensitivities in the mid-latitudes due to the impact of uncertainties in the southern and northern latitudes. Notice how TEMPO Ak-1 and Ak-2 have sensitivities throughout the troposphere, but for TROPOMI, the tropospheric sensitivities seen on Fig. 6 have disappeared due to consideration of the uncertainties. This means that TEMPO is better able to resolve tropospheric O_3 profiles compared to TROPOMI. This type of fidelity was one of the primary goals of the TEMPO mission.

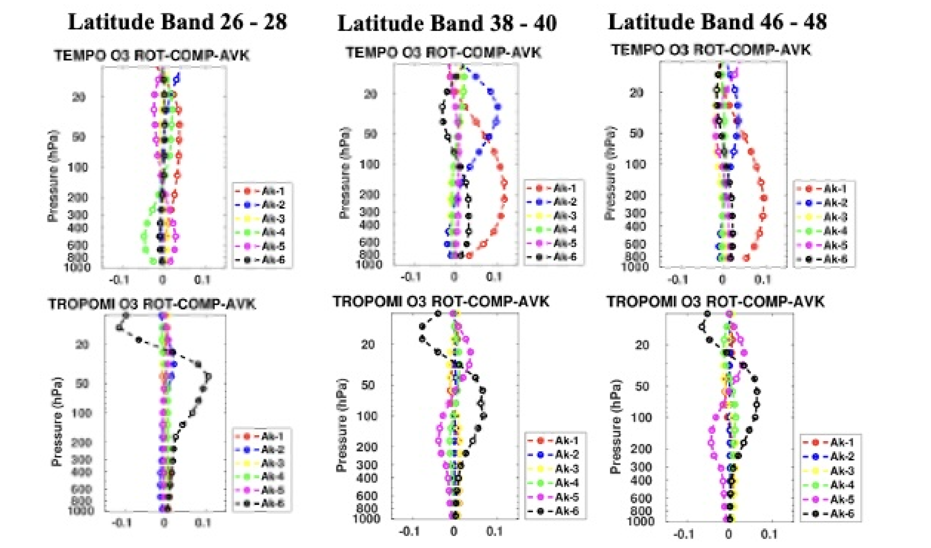

Figure 7: *Same as Fig. 3 except for the 'rotated and compressed' averaging kernels.*

4. Summary and Conclusions

In this poster, we presented an analysis of applying the CPSR dimensional reduction algorithm to TEMPO and TROPOMI O_3 profile retrievals. The results showed that:

- TEMPO and TROPOMI O_3 profile retrievals have rank deficient averaging kernels. TEMPO has four to five dominant modes with singular values ranging 1 – 3. TROPOMI has five to six dominant modes with singular values ranging 1 – 15 (the singular values of O(15) are suspect);
- Computational and storage cost for TEMPO can be reduced by ∼77% and for TROPOMI by ∼81%. Additionally, the rank deficiencies can be leveraged for reduced dimensional emission estimation. This one direction of our future work;
- The TEMPO left singular vectors in Fig. 5 have considerable vertical structure throughout the troposphere. TROPOMI left singular vectors are generally uniform throughout the troposphere. This means that the TEMPO averaging kernels should be better able to resolve the O_3 profile structure in the troposphere.
- The TEMPO 'compressed' (Ak-1) and 'rotated and compressed' (Ak-1) averaging kernels have sensitivities in the mid- to lower troposphere. The TROPOMI 'compressed' (Ak-1) averaging kernels have similar sensitivities, but the 'rotated and compressed' averaging kernels do not. These results suggest that even after considering 'solution uncertainties' TEMPO is better able to resolve the tropospheric O_3 profiles compared to TROPOMI.