The Impact of Altering Emission Data Precision on Efficiency and Accuracy Regarding the Community Multiscale Air Quality Model

Michael Walters[1,2] and David C. Wong[1]
Environmental Protection Agency[1]
Oak Ridge Associated Universities[2]

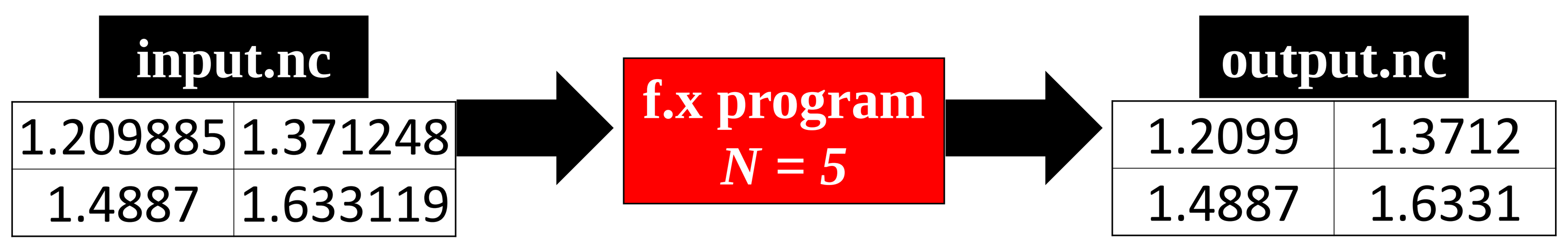
1. OBJECTIVES

Numeric simulations require a copious amount of disk space to archive input and output data. To *reduce disk space and improve storage costs*, this project applies a LOSSY compression algorithm on multiple emission datasets. Altered emission datasets are then ingested into the Community Multiscale Air Quality (CMAQ) model and simulations are run for 366 days for 2016. The impacts of such numerical manipulation on emission datasets is examined with respect to;

1) **Disk Space**
2) **Simulation Runtime**
3) **Numeric Accuracy for Particulate Matter 2.5 (PM2.5), Ozone, and Ammonia**

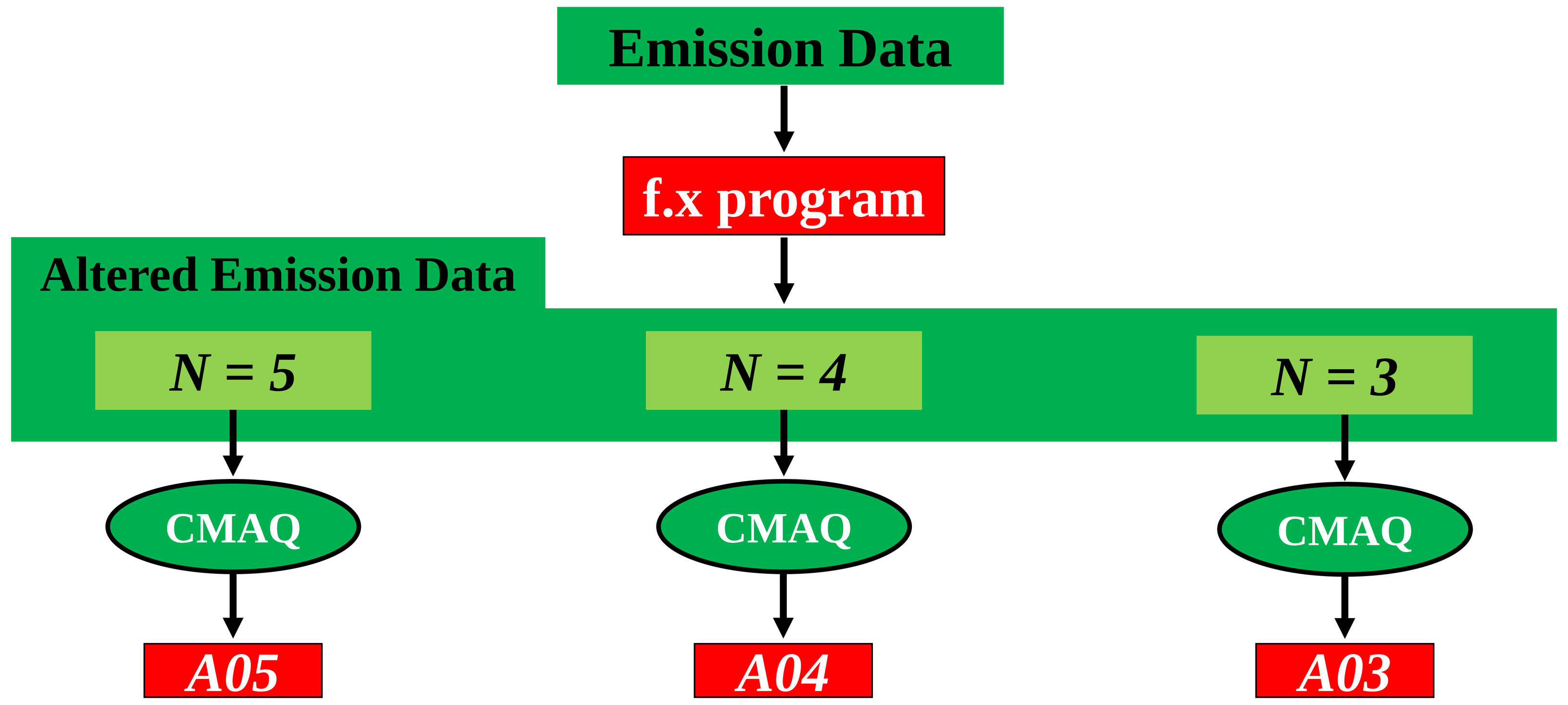
2. APPLIED LOSSY COMPRESSION ALGORTIHM

An *in-house* program (called; *f.x*) developed by David Wong: The program simply rounds all numeric (netCDF) files to *N* significant digits.



3. ALTERED SIMULATIONS

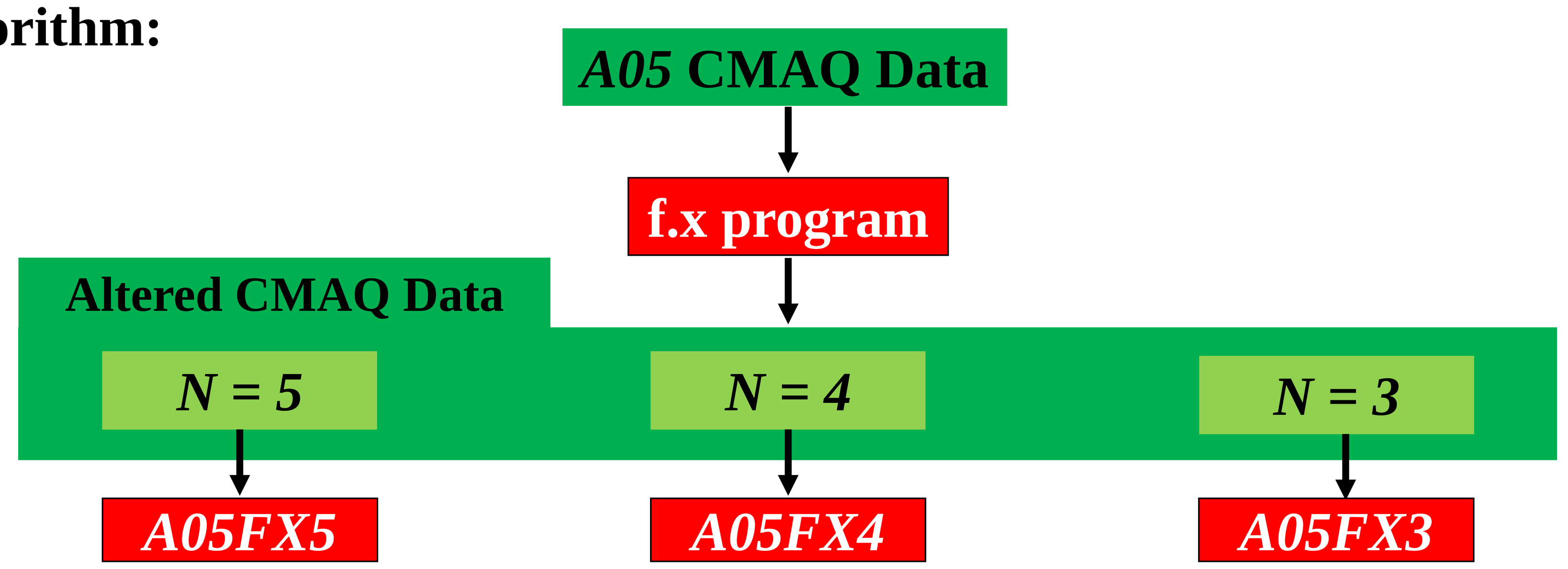
Simulations run with *altered emission data*:



Three altered simulations referred to as *A05*, *A04*, and *A03*.

4. ALTERED CASES

To determine the sensitivity of additional altered cases, *direct CMAQ output* was post-processed using the applied (*f.x*) lossy compression algorithm:



This process is repeated for the A04 and A03 simulations. In total, 1 simulation was run with unaltered emission data (*orig* – the "benchmark" simulation), 3 simulations were run with altered emission datasets (*A05*, *A04*, and *A03*), and 9 cases were created/processed from altered simulation output (*A05FX05*, *A05FX04*, *A05FX03*, *A04FX05*, *A04FX04*, *A04FX03*, *A03FX05*, *A03FX04*, and *A03FX03*)

5. IMPACT ON DISK SPACE

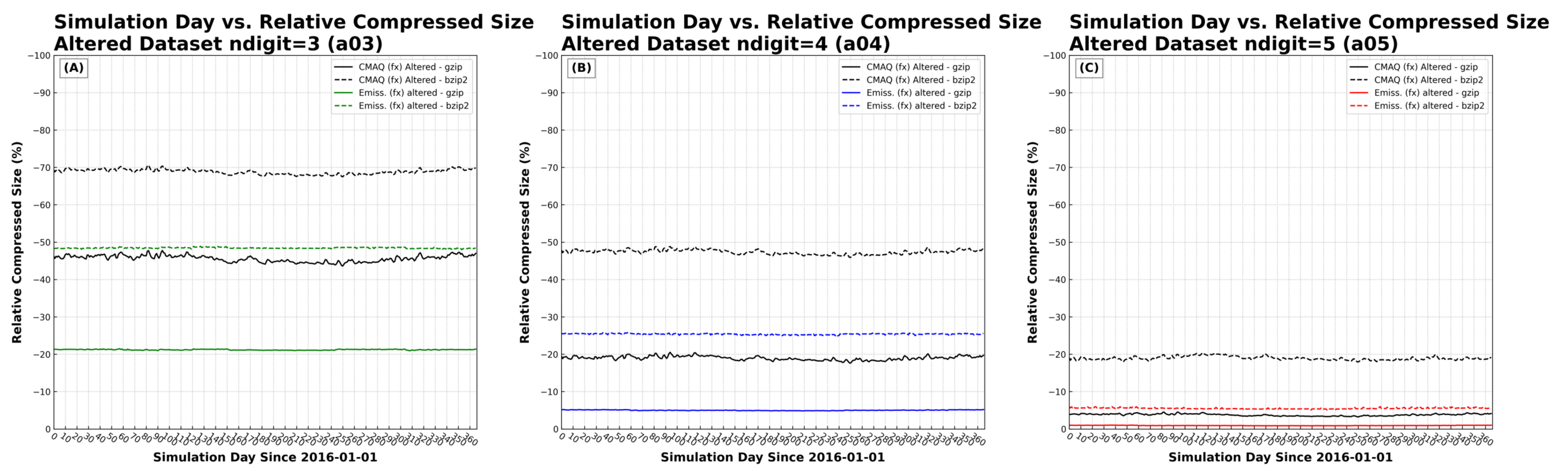


Figure 1. Simulation day (x) vs. relative compressed size of altered emission and CMAQ files with respect to the unaltered datasets for a) *N = 5*, b) *N = 4*, and C) *N = 3* datasets for two different lossless compression utilities (gzip and bzip2). **Using *f.x* in tandem with bzip2 significantly improves disk space.**

6. IMPACT ON RUNTIME

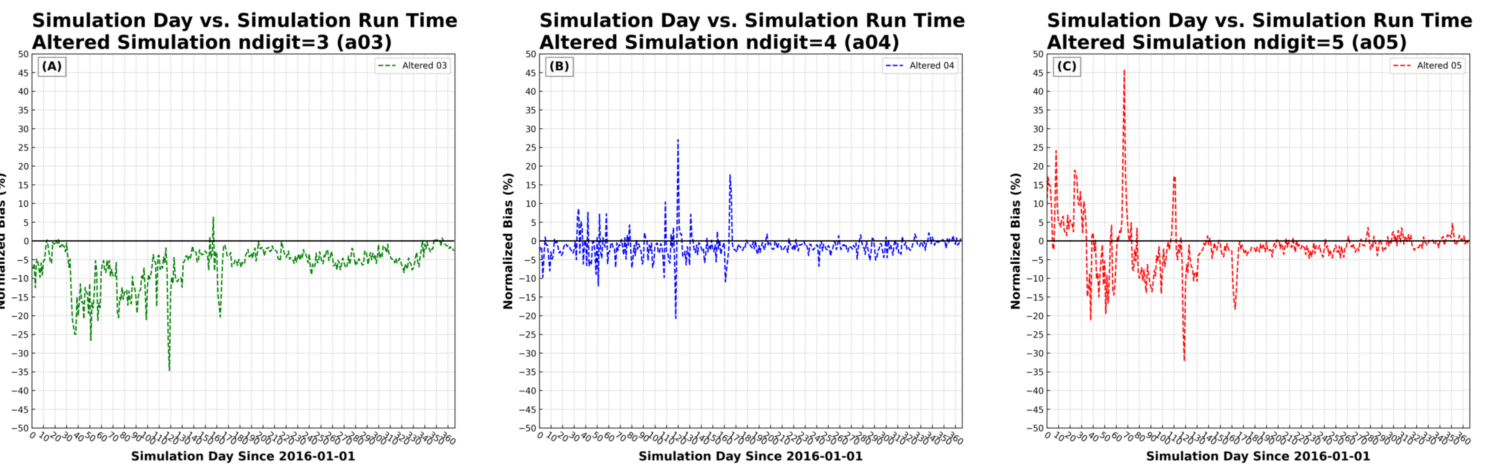


Figure 2. Simulation day (x) vs. relative compressed size of altered emission and CMAQ files with respect to the unaltered datasets for a) *N = 5*, b) *N = 4*, and C) *N = 3* datasets for two different lossless compression utilities (gzip and bzip2). Using *f.x* in tandem with bzip2 significantly improves disk space. **Runtime is slightly improved on a non-dedicated server.**

7. NUMERIC ACCURACY

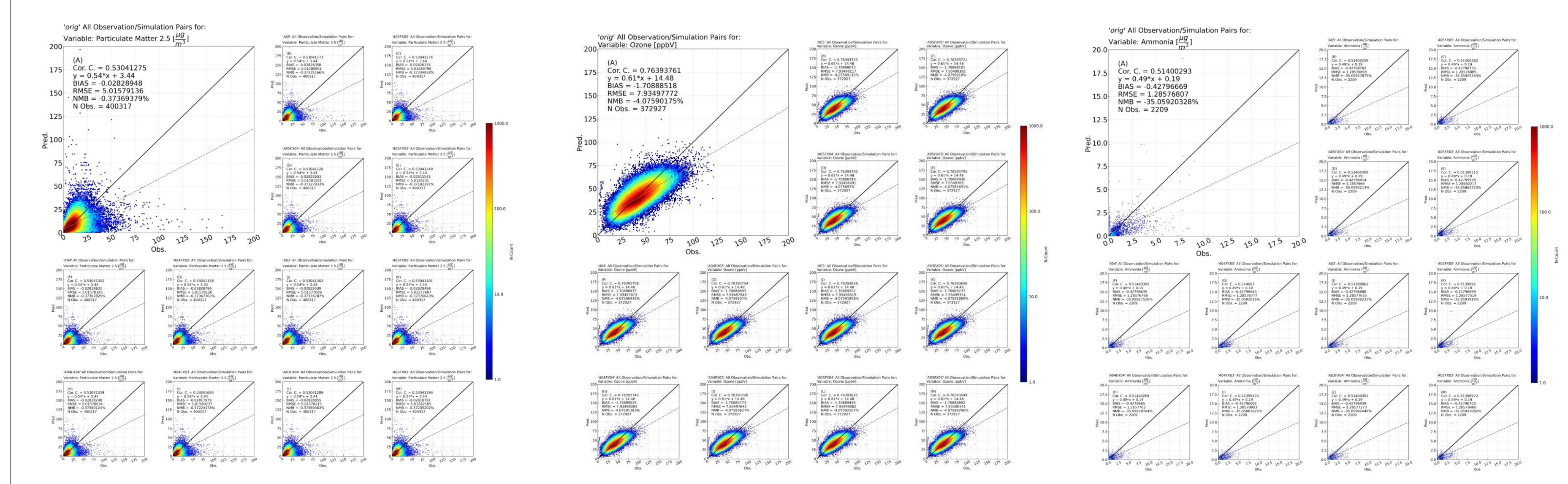


Figure 3, 4, and 5. In-situ observations (x) versus predicted (y) concentrations of daily averaged PM2.5 (left Fig.), maximum 8-hour ozone (center Fig.), and daily averaged ammonia (right Fig.) for all simulations and cases. **Bulk statistical metrics are not impacted.**

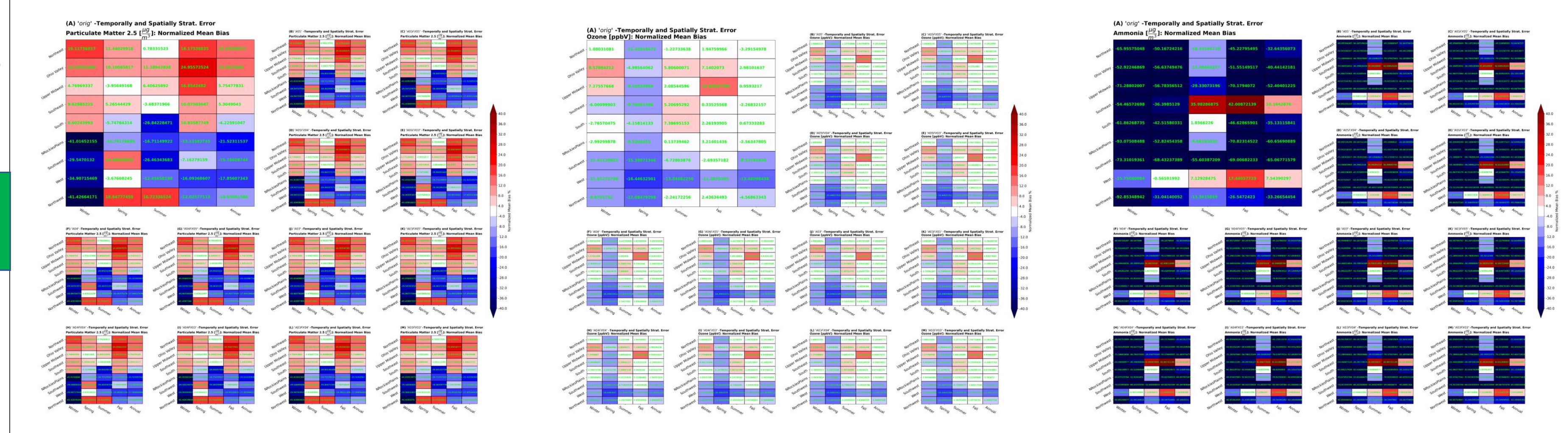


Figure 6, 7, and 8. Mean bias (color shade) of daily averaged PM2.5 (left Fig.), maximum 8-hour ozone (center Fig.), and daily averaged ammonia (right Fig.) for all simulations and cases stratified by region (y) and by season (x). **Results do not vary based on space and time.**

8. CONCLUSION

1) **Disk Space usage can be significantly reduced by utilizing the *f.x* program then applying bzip2 to emision data (Fig. 1).**
2) **Albeit small (and unexpected), runtime can be shortened (Fig. 2).**
3) **Numeric accuracy was not impacted by netCDF manipulation of emision data based on bulk and stratified statistics (Fig. 3 to Fig. 8).**

9. STATISTICAL METRICS

Equations: In which *X* denotes observed values, *Y* denotes predicted values, and *N* denotes the number of observations -

$$MB = \frac{1}{N} \cdot \sum(Y_i - X_i), \quad \text{Eq. 1}$$

$$CC = \frac{1}{N-1} \cdot \frac{\sum((X_i - \bar{X}) \cdot (Y_i - \bar{Y}))}{\sigma_X \cdot \sigma_Y}, \quad \text{Eq. 2}$$

$$RMSE = \sqrt{\frac{\sum(Y_i - X_i)^2}{N}}, \quad \text{Eq. 3}$$

$$NMB = \frac{\sum(Y_i - X_i)}{\sum(X_i)} \cdot 100\%, \quad \text{Eq. 4}$$

10. OBSERVATIONS

Particulate Matter 2.5: United States Environmental Protection Agency's Air Quality System (AQS)
Ozone: United States Environmental Protection Agency's Air Quality System (AQS)
Ammonia: United States Environmental Protection Agency's Ammonia Monitoring Network (AMON)
Obs.-Pre. Pairing: Completed using the Atmospheric Modeling Evaluation Toolkit (AMET)

Contact: wong.david-c@epa.gov walters.michael.s@epa.gov