Determining near real time surface ozone from TEMPO and machine learning

Dan Anderson

6 November 2025

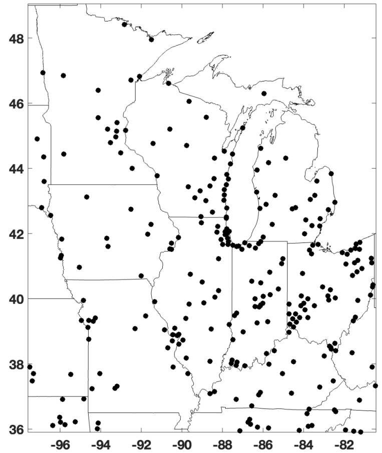
UMBC

GODDARD EARTH SCIENCES

NASA

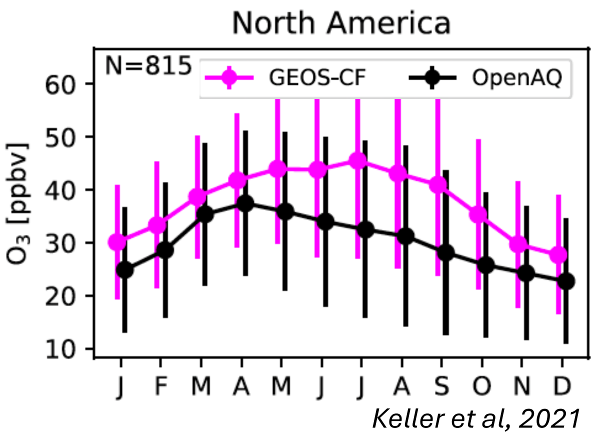
Motivation: Need better constraints on surface O_3 across the US

Surface observations are sparsely distributed



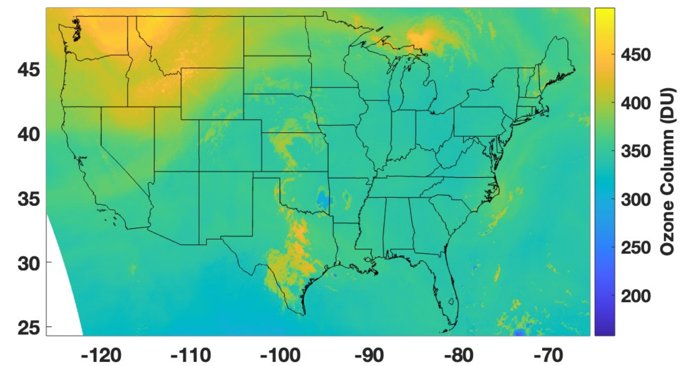
EPA O_3 monitoring sites

Air quality models often have significant biases



Evaluation of the NASA GEOS-CF model with surface observations

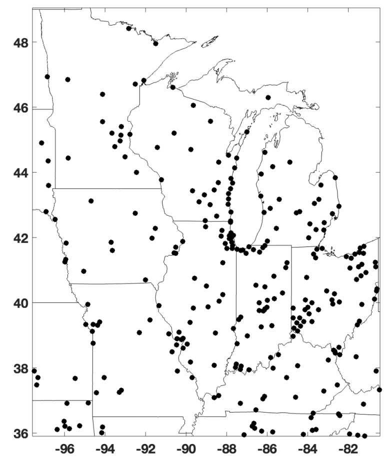
Satellites don't provide surface concentrations



TEMPO total column O_3 (15Z May 1, 2024)

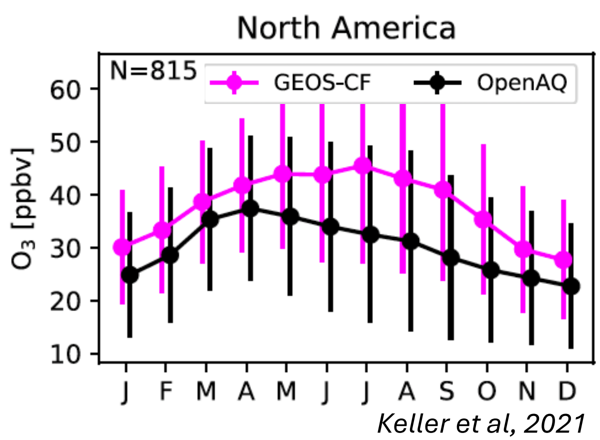
Motivation: Need better constraints on surface O_3 across the US

Surface observations are sparsely distributed



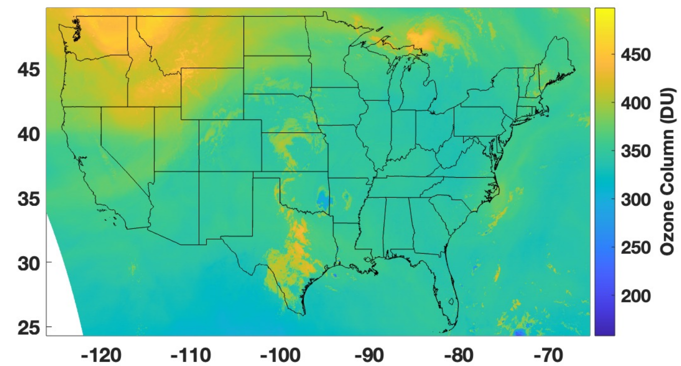
EPA O_3 monitoring sites

Air quality models often have significant biases



Evaluation of the NASA GEOS-CF model with surface observations

Satellites don't provide surface concentrations



TEMPO total column O_3 (15Z May 1, 2024)

- **Why Now**:
  - TEMPO satellite provides data at high spatial and temporal resolution.
  - Machine learning allows advances beyond previous techniques.

Project overview and goals

- **Project goal is to create 2 products**:
  - Satellite-constrained, near real time surface O_3 at hourly resolution derived from machine learning, satellite data, and in situ observations.
  - Short-term forecast (12 – 24 hours) of surface O_3 using a similar methodology.
  - Ideal is to provide both products at hourly resolution across the TEMPO domain.
- Near term aim is product development and validation.
- Potential applications for air quality forecasting, ozone exposure studies, air quality model validation, TEMPO product validation, etc.
- Today: Quick look at preliminary data to demonstrate product potential. **Results will likely change.**

Project team members

Science

Dan Anderson

Bryan Duncan

NASA Goddard

Kristen Okorn

NASA Ames

Gonzalo Gonzalez Abad

Harvard/SAO

Stakeholder Representatives

Mary Uhl

WESTAR
Western States Air Resources Council

Byeong Kim

ENVIRONMENTAL PROTECTION DIVISION

Alex Karambelas

NESCAUM

Zach Adelman

LADCO | LAKE MICHIGAN AIR DIRECTORS CONSORTIUM

General strategy for generating NRT surface O_3

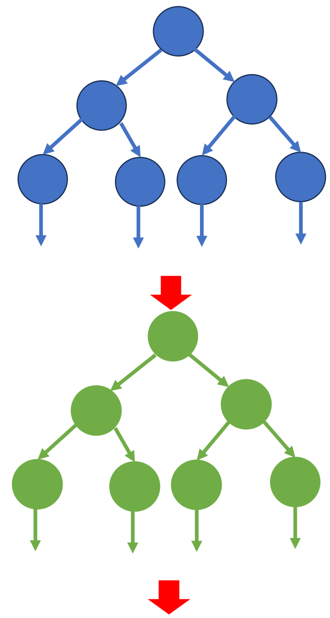
ML Model Inputs

TEMPO
- NO_2, HCHO, O_3
- Viewing Geometry
- Reflectivity

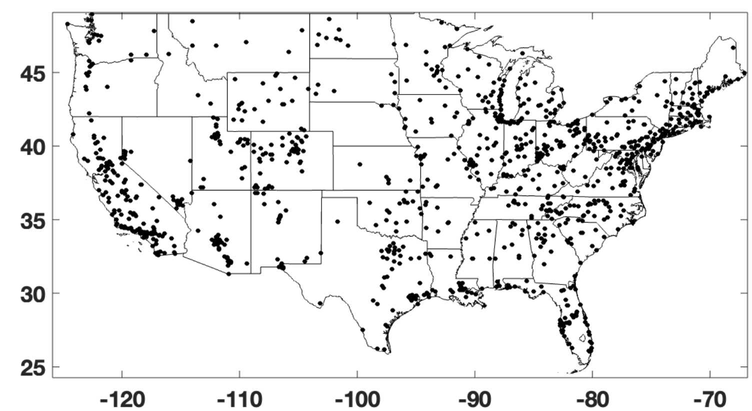
Met
- T, P, Humidity
- Wind speed
- Wind direction
- PBL Height
- Latitude

Other
- Population
- Month, Hour
- Road Density
- Day of Week
- NDVI

Gradient Boosted Regression Tree Model Training



Training Targets (Surface O_3)



EPA Ozone Monitoring Sites

Test model performance at independent O_3 monitoring sites

General strategy for generating NRT surface O_3

ML Model Inputs

TEMPO

- NO_2, HCHO, O_3
- Viewing Geometry
- Reflectivity

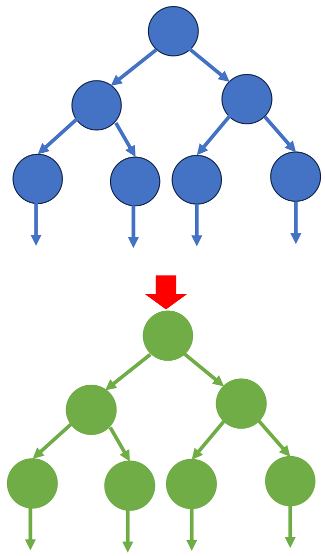
Met

- T, P, Humidity
- Wind speed
- Wind direction
- PBL Height
- Latitude

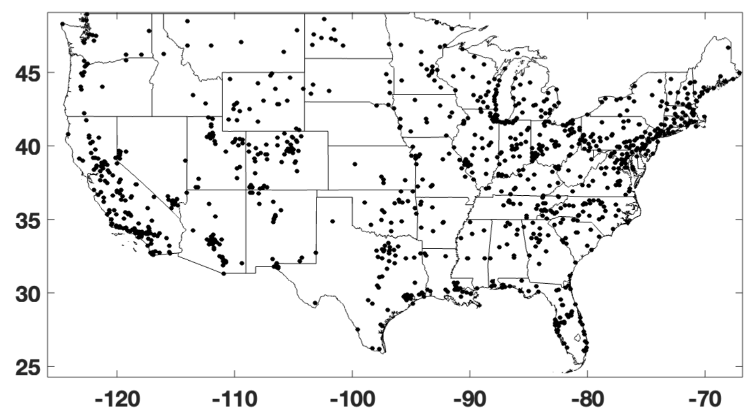
Other

- Population
- Month, Hour
- Road Density
- Day of Week
- NDVI

Gradient Boosted Regression Tree Model Training



Training Targets (Surface O_3)

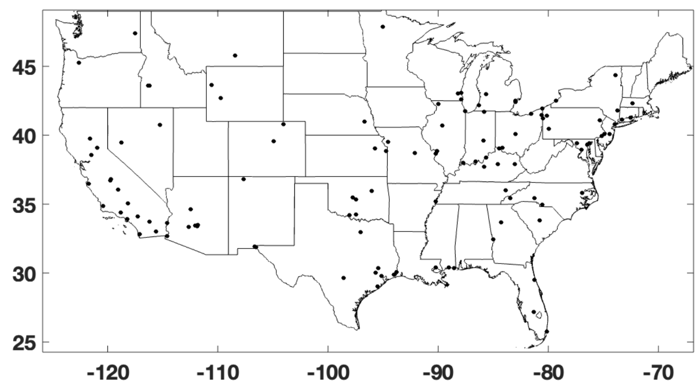


EPA Ozone Monitoring Sites

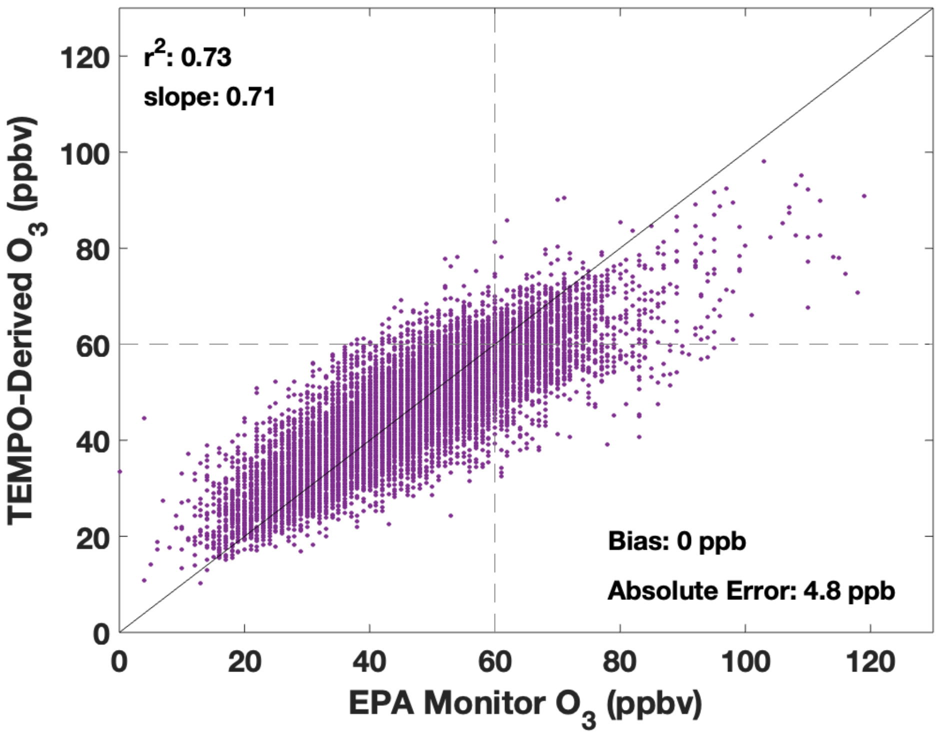
- **Note**: Today's preliminary results determined using a limited subset (May – June 2024) of Version 3 TEMPO data.
- Final model will incorporate all available O_3 season data using latest TEMPO retrievals.

TEMPO product captures the domain-wide observed variability with low bias

Validation sites independent of the training data

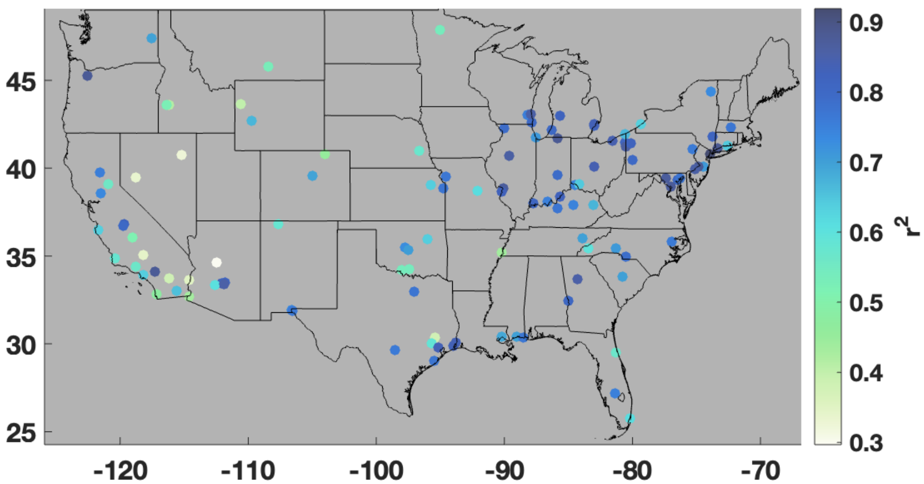


- Comparison of 1-hour averaged O_3 at all validation sites for all times in May & June 2024.
- Overestimates at low end of range and underestimates at high end.



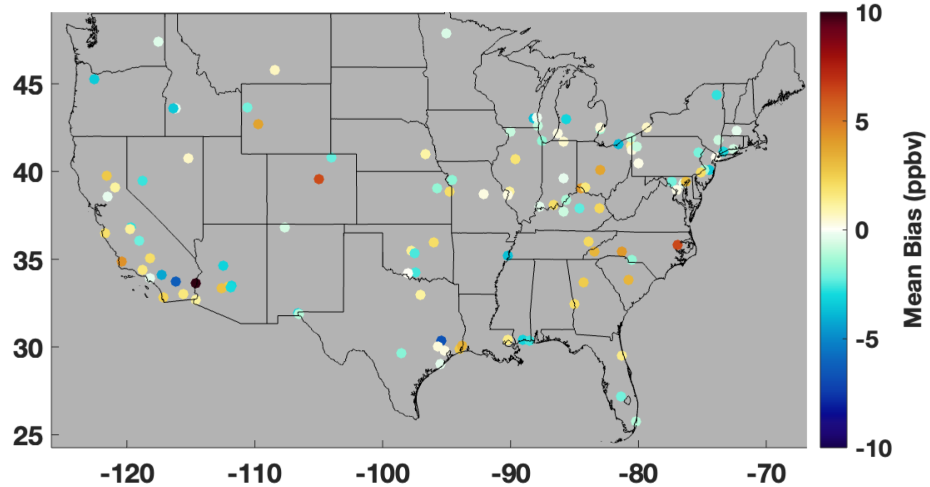
Preliminary results demonstrate regionally-dependent accuracy

r^2 of hourly O_3 between EPA monitors and TEMPO-derived product at validation sites



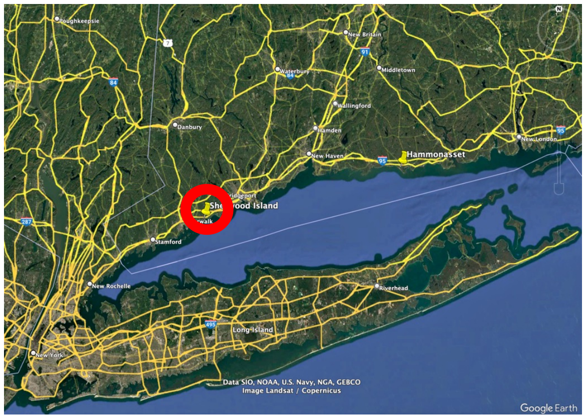
- TEMPO product consistently captures variability in the Midwest and mid-Atlantic but struggles in some western locations.

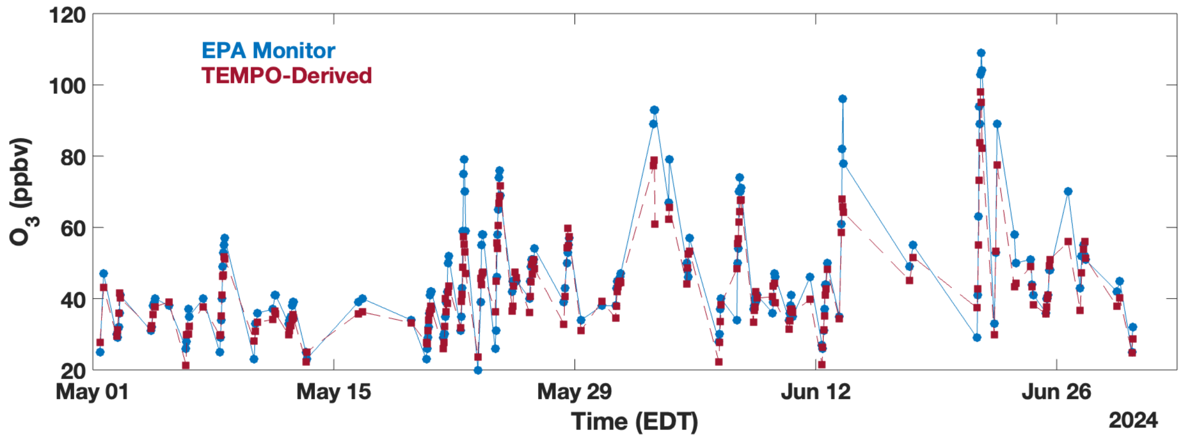
Mean bias of hourly O_3 for TEMPO-derived product for May – June 2024

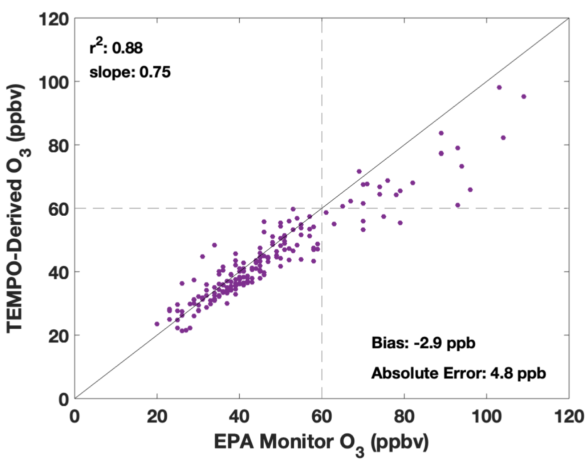


- Bias at individual sites is generally less than 5 ppbv, with some notable exceptions.

Coastal Site 1: Sherwood Island (Westport, CT)

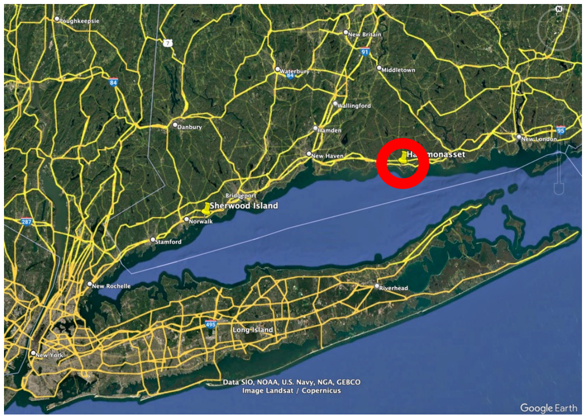


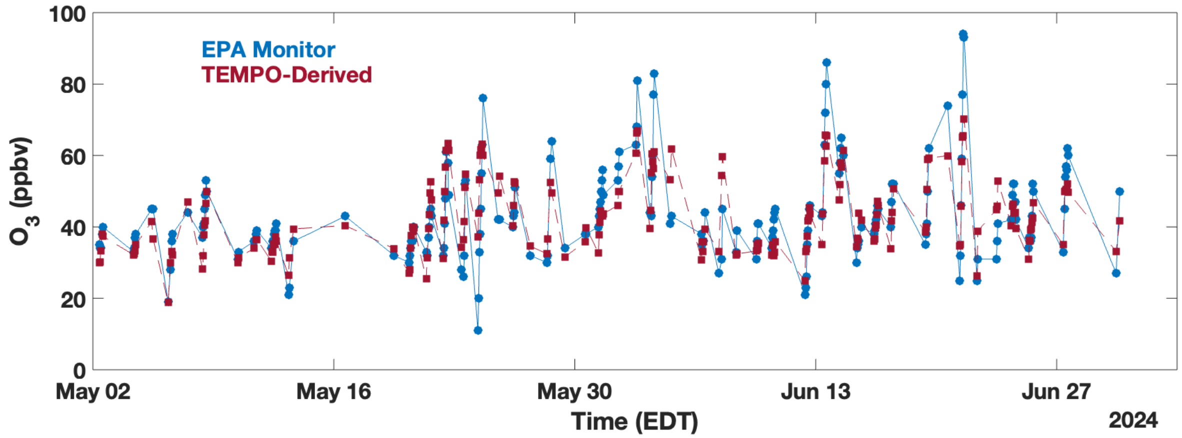


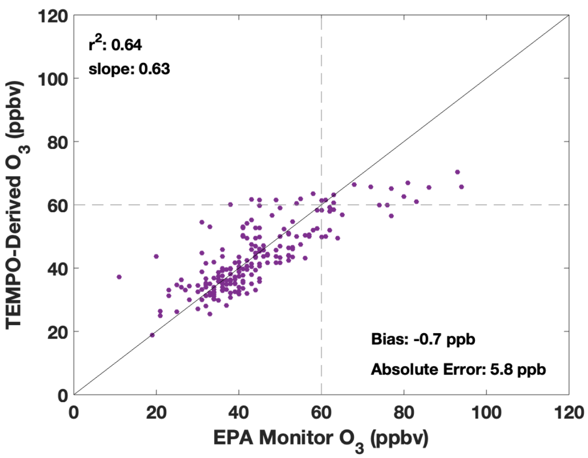


- TEMPO-derived product captures observed variability of hourly-averaged O_3 with high fidelity at a coastal site.
- Average of 4 observations/day for TEMPO-derived product.
- Frequent underestimate of maximum O_3, although correctly captures high O_3 events 79% of the time.

Coastal Site 2: Hammonasset Beach







- Decrease in accuracy as compared to Westport site, despite geographic proximity.
- Accuracy better in first half of May.
- Frequent underestimate of maximum O_3 and only correctly captures high O_3 events 50% of the time.

Conclusions and Future Work

- Preliminary results suggest eventual NRT product could provide valuable constraints on surface O_3 across broad regions of the US.
- Near term goals:
  - Update to TEMPO v4 data and include more months in training
  - More thorough model training and input variable testing
  - Uncertainty quantification
  - Case studies at individual sites
  - Methodology to infer O_3 in cloudy conditions
- Contact info: daniel.c.anderson@nasa.gov