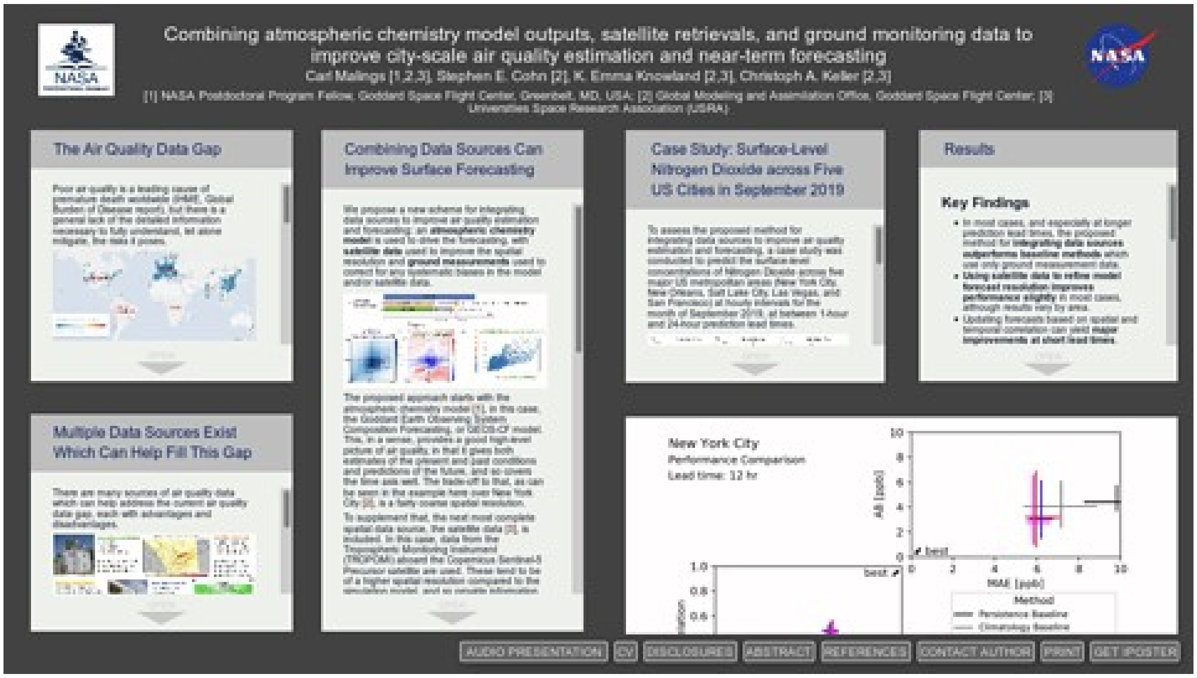
Combining atmospheric chemistry model outputs, satellite retrievals, and ground monitoring data to improve city-scale air quality estimation and near-term forecasting



Carl Malings [1,2,3], Stephen E. Cohn [2], K. Emma Knowland [2,3], Christoph A. Keller [2,3]

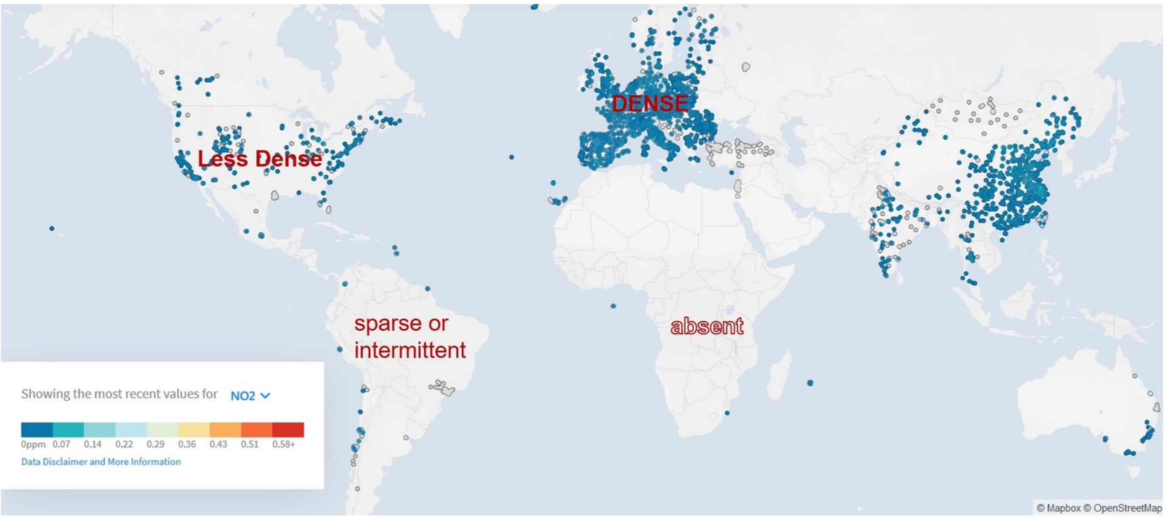
[1] NASA Postdoctoral Program Fellow, Goddard Space Flight Center, Greenbelt, MD, USA; [2] Global Modeling and Assimilation Office, Goddard Space Flight Center; [3] Universities Space Research Association (USRA)

PRESENTED AT:

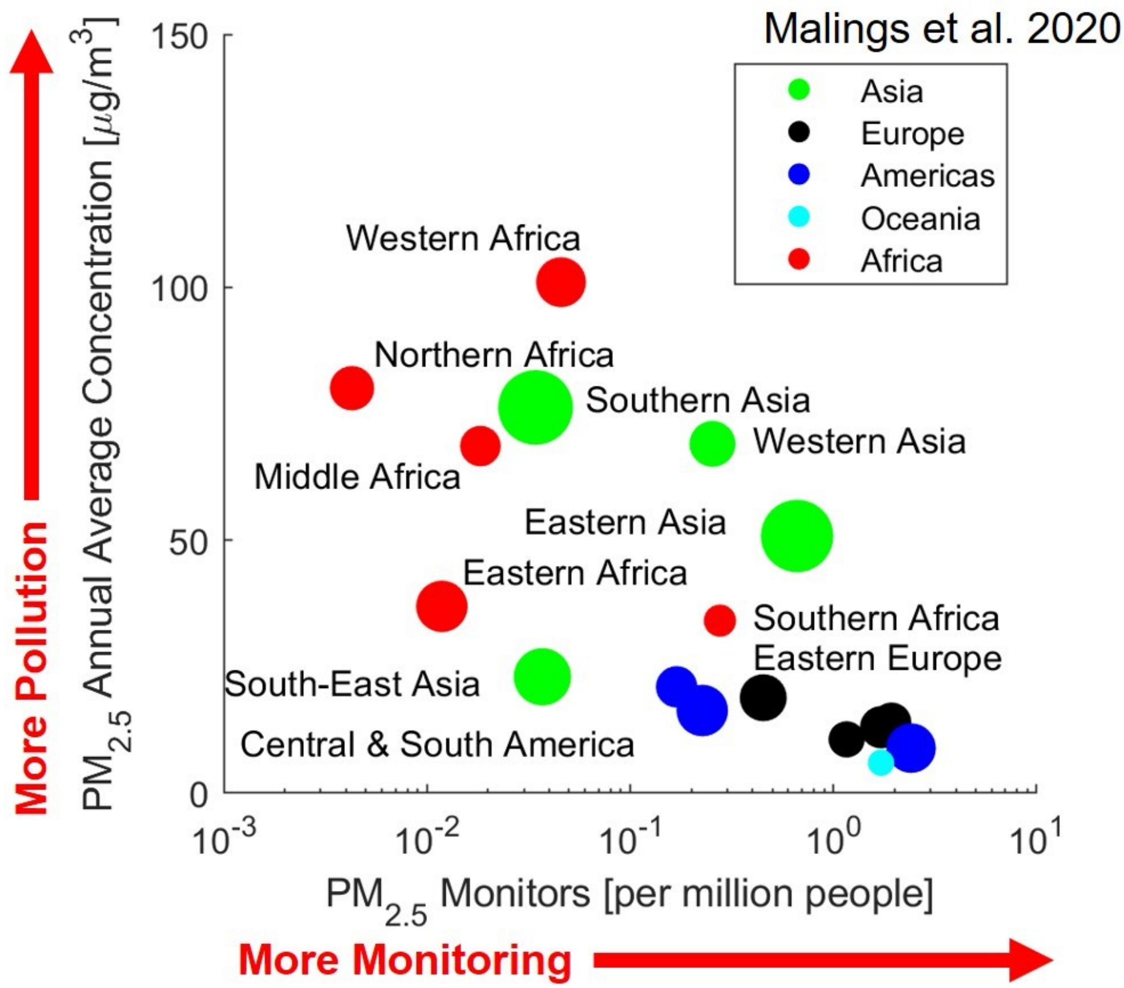
AGU FALL MEETING
Online Everywhere | 1–17 December 2020

THE AIR QUALITY DATA GAP

Poor air quality is a leading cause of premature death worldwide (IHME, Global Burden of Disease report), but there is a general lack of the detailed information necessary to fully understand, let alone mitigate, the risks it poses.



As indicated in the map above (a plot of open-source data on Nitrogen Dioxide concentrations from regulatory-grade monitoring stations around the world, obtained on 22 September 2020 from openaq.org), the monitoring of relevant air pollutants is conducted at different intensities in different areas of the world, ranging from fairly dense monitoring networks in places like Europe to very sparse or practically absent monitoring for much of the developing world. In fact, there is a general trend of decreased monitoring density in more heavily polluted areas (Malings et al., 2020).



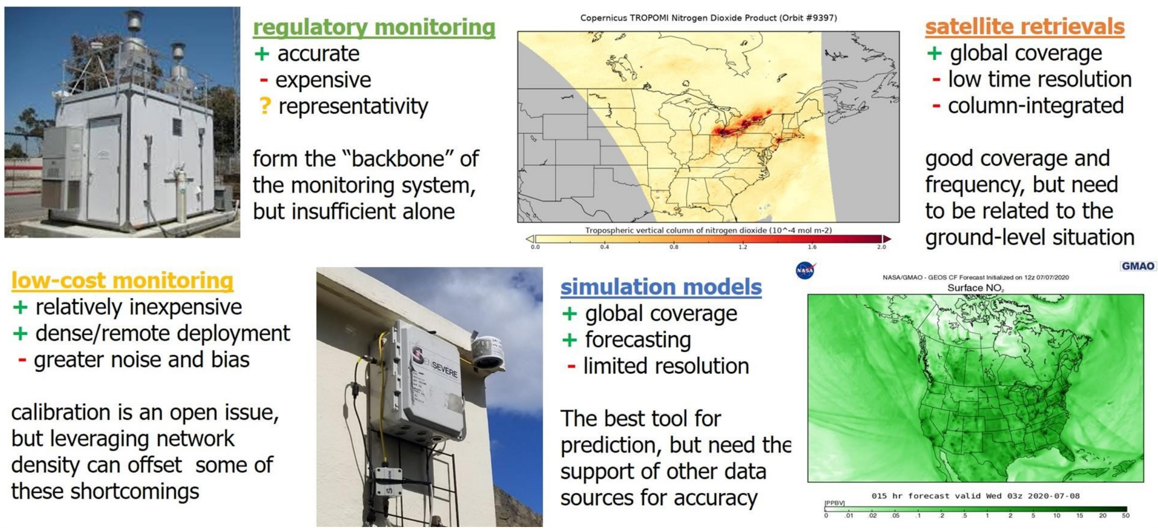
This problem makes it difficult to draw even general conclusions about global air quality, such as which city on earth has the highest level of annual average concentrations of a given pollutant (Martin el at., 2019), let alone provide the specific data needed to support air-quality-related policy.

(panel narration below):

0:00 / 1:05

MULTIPLE DATA SOURCES EXIST WHICH CAN HELP FILL THIS GAP

There are many sources of air quality data which can help address the current air quality data gap, each with advantages and disadvantages.



Regulatory Monitors form the baseline of air quality measurement data, with highly accurate and reliable instruments. However, as indicated above, these monitors are fairly sparsely deployed around the world, due to their high cost for purchase and operation. Their low density of deployment also leads to representativity issues, i.e., do the relatively few monitoring stations in an area fully capture the range of concentrations experienced by people living there?

Low-cost Sensors, as the name implies, are less expensive than regulatory monitors, and so can be deployed in greater numbers to increase surface monitoring density and expand it into previously unmonitored areas. However, the trade-off for this lower cost is higher levels of noise and higher biases compared to regulatory monitors, and these low-cost sensors must be carefully calibrated against available regulatory-grade instruments in their deployment areas to provide reliable data.

Satellites provide the opportunity to measure pollutant concentrations over large swaths of the globe at a time. Their temporal resolution is limited, however, to specific overpass times during the day for polar-orbiting instruments. Even the new generation of geostationary satellites coming online in the near future will be limited to daytime observations over select portions of the globe. Also, satellites measure the integrated concentrations of pollutants in an atmospheric column, rather than only the surface-level concentrations that are of immediate interest for air quality.

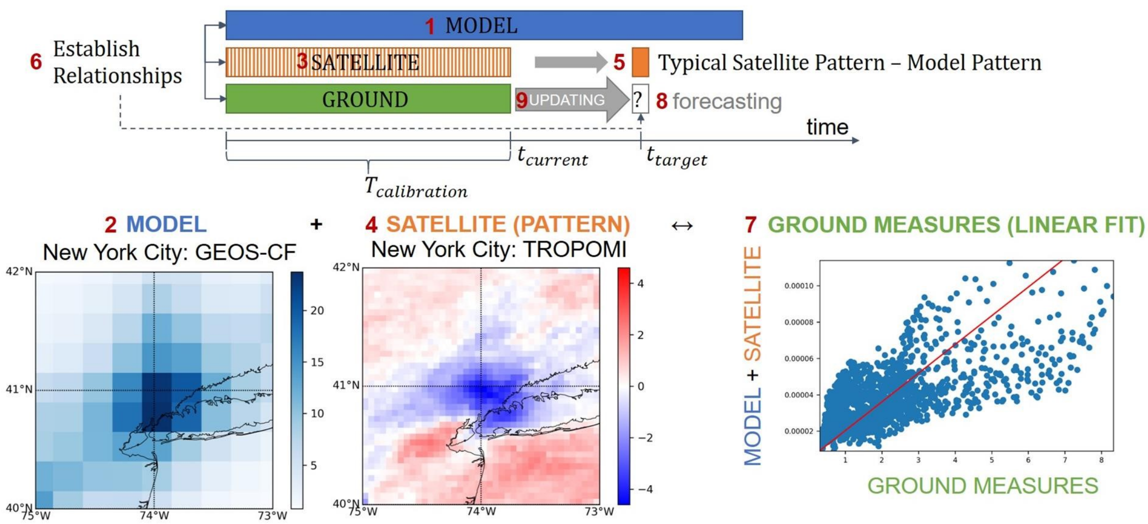
Atmospheric Chemistry Models have the capacity to simulate the concentrations of pollutants at a global scale, and also to make forecasts into the future. However, their spatial and temporal resolutions are limited by the computational effort needed to run these simulations. Also, simplifications in the modeling of atmospheric processes and uncertainties in the knowledge of pollutant emissions can lead to biased results, and so verification of model outputs against in-situ data is necessary to ensure accuracy.

(panel narration below):

0:00 / 1:53

COMBINING DATA SOURCES CAN IMPROVE SURFACE FORECASTING

We propose a new scheme for integrating data sources to improve air quality estimation and forecasting: an **atmospheric chemistry model** is used to drive the forecasting, with **satellite data** used to improve the spatial resolution and **ground measurements** used to correct for any systematic biases in the model and/or satellite data.



The proposed approach starts with the atmospheric chemistry model [1], in this case, the Goddard Earth Observing System Composition Forecasting, or GEOS-CF model. This, in a sense, provides a good high-level picture of air quality, in that it gives both estimates of the present and past conditions and predictions of the future, and so covers the time axis well. The trade-off to that, as can be seen in the example here over New York City [2], is a fairly coarse spatial resolution.

To supplement that, the next most complete spatial data source, the satellite data [3], is included. In this case, data from the Tropospheric Monitoring Instrument (TROPOMI) aboard the Copernicus Sentinel-5 Precursor satellite are used. These tend to be of a higher spatial resolution compared to the simulation model, and so provide information on some of the finer details of the spatial distribution of pollutants. However, unlike the model, these are not continuous estimates, but rather occur sporadically over time as satellite positions and cloud conditions allow. Therefore, instead of using them directly, the difference of the satellite "snapshots" and the model's estimates of pollution at the same time are compiled for some designated calibration period, for example, the previous week. These then represent the "typical pattern" in the satellite-to-model differences [4], an estimate of the systematic spatial bias of the model due to its lower resolution. It is assumed that this pattern will hold true throughout the day between satellite passes and into the near future [5]. With future geostationary satellites, a separate pattern might be extracted e.g. for each hour of the daytime, and so the assumptions of a constant pattern would not need to be as strict.

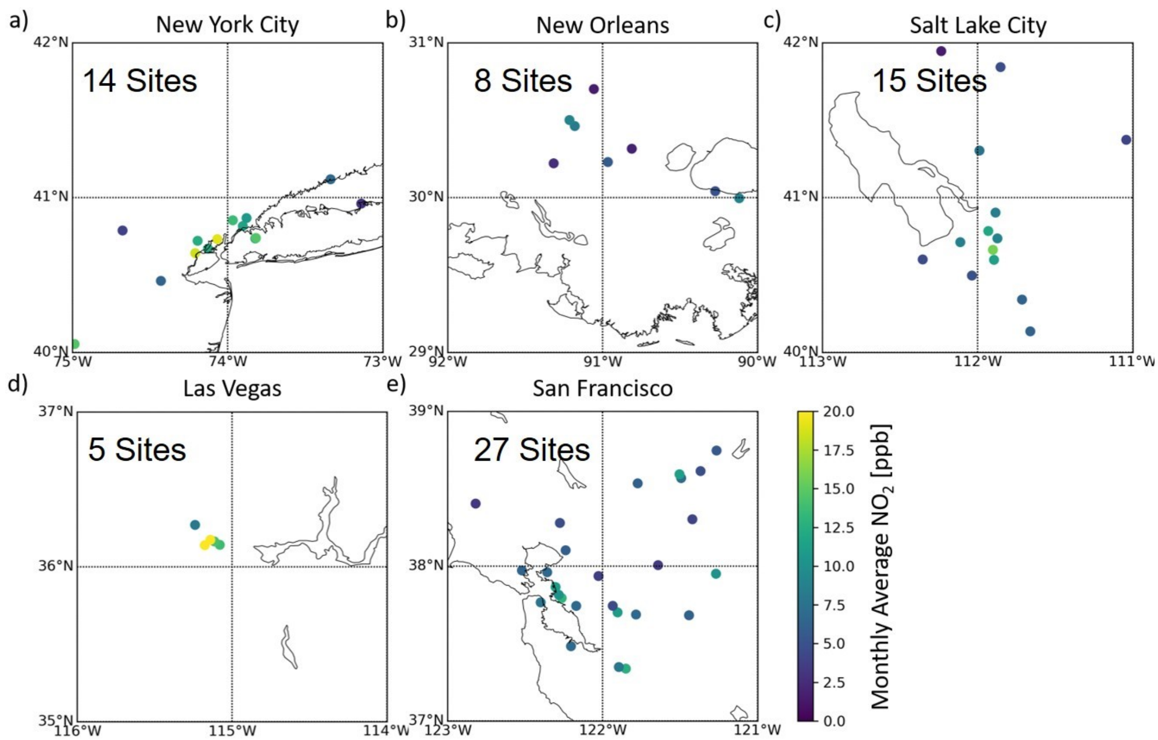
Using the ground-level monitoring information collected during the calibration period, relationships [6] can be established between, on the one hand, the spatial pollutant patterns predicted by the model combined with the typical pattern map from the satellites during the calibration period, and on the other hand, the ground-truth surface monitor data for the same period. For this work, ground-level monitoring data provided by the US Environmental Protection Agency (EPA) network are used. Currently, these relationships are established via linear regression [7] but other methods might also be applied to this problem in future work. Assuming that the relationships identified for the calibration period hold into the near future, a forecast of the future surface air quality can be made [8]. Finally, this forecast might be updated [9] using available ground measurement data together with information on the spatial and temporal correlation of pollutant concentrations, using techniques such as spatio-temporal kriging on residuals.

(panel narration below):

0:00 / 2:36

CASE STUDY: SURFACE-LEVEL NITROGEN DIOXIDE ACROSS FIVE US CITIES IN SEPTEMBER 2019

To assess the proposed method for integrating data sources to improve air quality estimation and forecasting, a case study was conducted to predict the surface-level concentrations of Nitrogen Dioxide across five major US metropolitan areas (New York City, New Orleans, Salt Lake City, Las Vegas, and San Francisco) at hourly intervals for the month of September 2019, at between 1-hour and 24-hour prediction lead times.



To assess the performance of the proposed methods, data from all but one EPA ground station in an area are used as inputs to the method, which is used to forecast surface concentrations at the left-out ground station site. All sites are rotated through, resulting in one set of performance metrics assessed per ground station site in a given area.

Evaluation Metrics

Correlation (r): This measures the linearity of the relationship between the predicted (x) and true (y) measurements. The metric ranges from -1 to 1, with higher values being more desirable.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Root Mean Square Error (RMSE): This measures the accuracy of the predictions (x) of true values (y), with larger errors given a greater relative weight. The metric ranges from 0 upwards, with lower values being more desirable.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}$$

Mean Absolute Error (MAE): This measures the accuracy of the predictions (x) of true values (y), with all errors given equal weight. The metric ranges from 0 upwards, with lower values being more desirable.

$$MAE = \frac{i}{n}\sum_{i=1}^{n}|x_i - y_i|$$

Absolute Bias (AB): This measures the absolute value of the average bias in predictions (x) of true values (y). The metric ranges from 0 upwards, with lower values being more desirable.

$$AB = |\frac{1}{n}\sum_{i=1}^{n}(x_y - y_i)|$$

Methods

Persistence Baseline: The concentration at any location and time is assumed to be equal to the most recently measured concentration at the nearest ground measurement site.

Climatology Baseline: The concentration at any location and time is assumed to be equal to the average concentration at that time of day during the calibration period at the nearest ground measurement site.

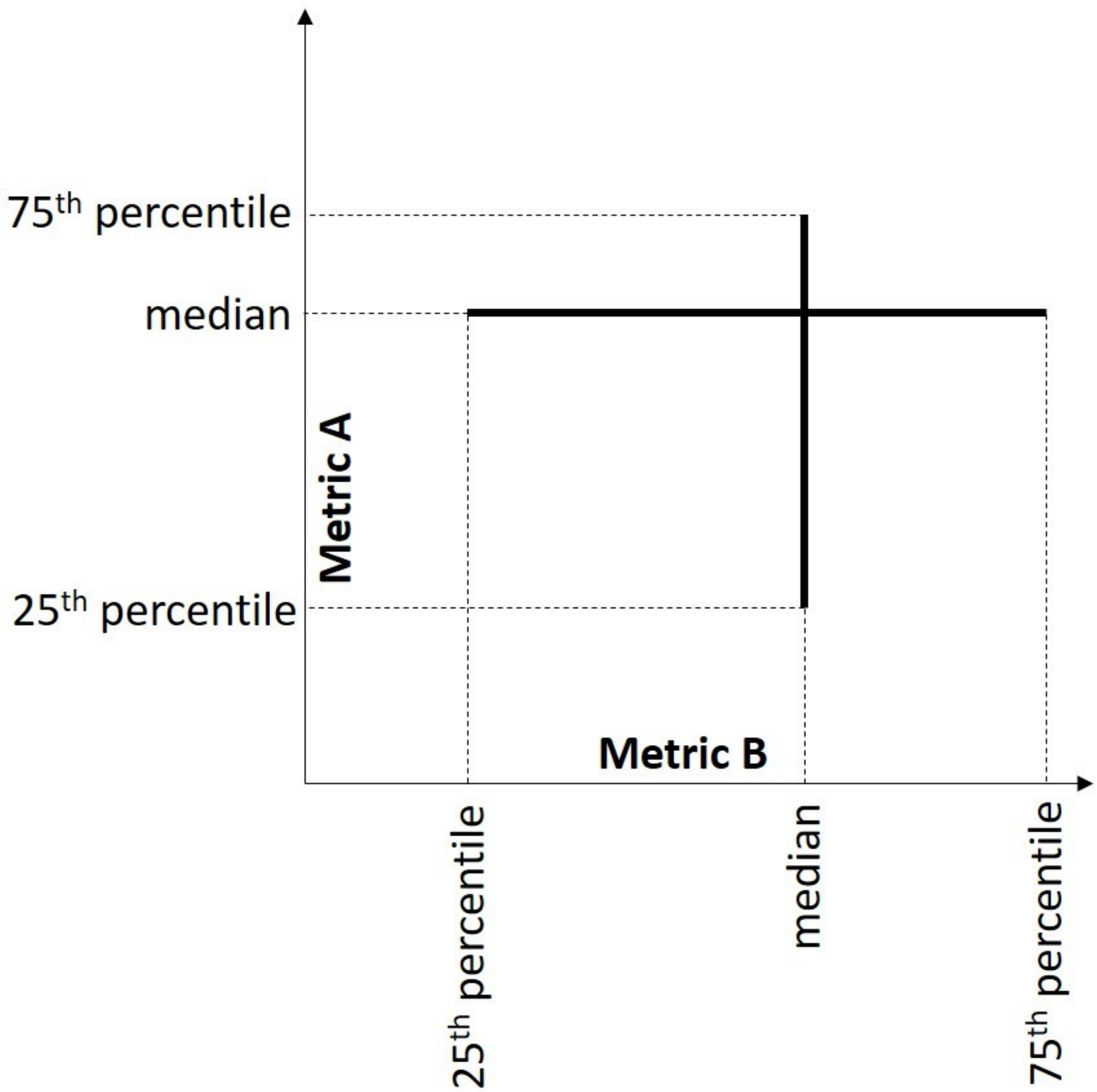
GEOS-CF: This is the proposed method using applied without the satellite (TROPOMI) data included; that is, only GEOS-CF model data and ground data are used.

GEOS-CF + TROPOMI: This is the proposed method with the satellite (TROPOMI) data incorporated, but without the final updating based on correlations with the most recent ground measurement data.

GEOS-CF + TROPOMI & Kriging: This is the full proposed method, with GEOS-CF, TROPOMI, and ground data being utilized, as well as a final updating step based on correlations in space and time to the most recent ground monitoring data.

Presentation of Results

In the plots below, two sub-plots are used to describe performance across four metrics, with values closer to the center of the figure overall indicating better performance. In the upper-left corner, the city and lead time of the results are indicated, and in the lower-right corner there is a legend for the colors denoting different methods. In each plot, the crosses indicate the relative performance of each method according to the different metrics, with the position of the cross indicating the median in metric values across evaluation sites in the city, and the size of the cross denoting the inter-quartile range (25^{th} to 75^{th} percentiles) in the metric values (see example diagram below).



(panel narration below):

0:00 / 2:42

RESULTS

Key Findings

- In most cases, and especially at longer prediction lead times, the proposed method for **integrating data sources outperforms baseline methods** which use only ground measurement data.
- **Using satellite data to refine model forecast resolution improves performance slightly** in most cases, although results vary by area.
- Updating forecasts based on spatial and temporal correlation can yield **major improvements at short lead times**.

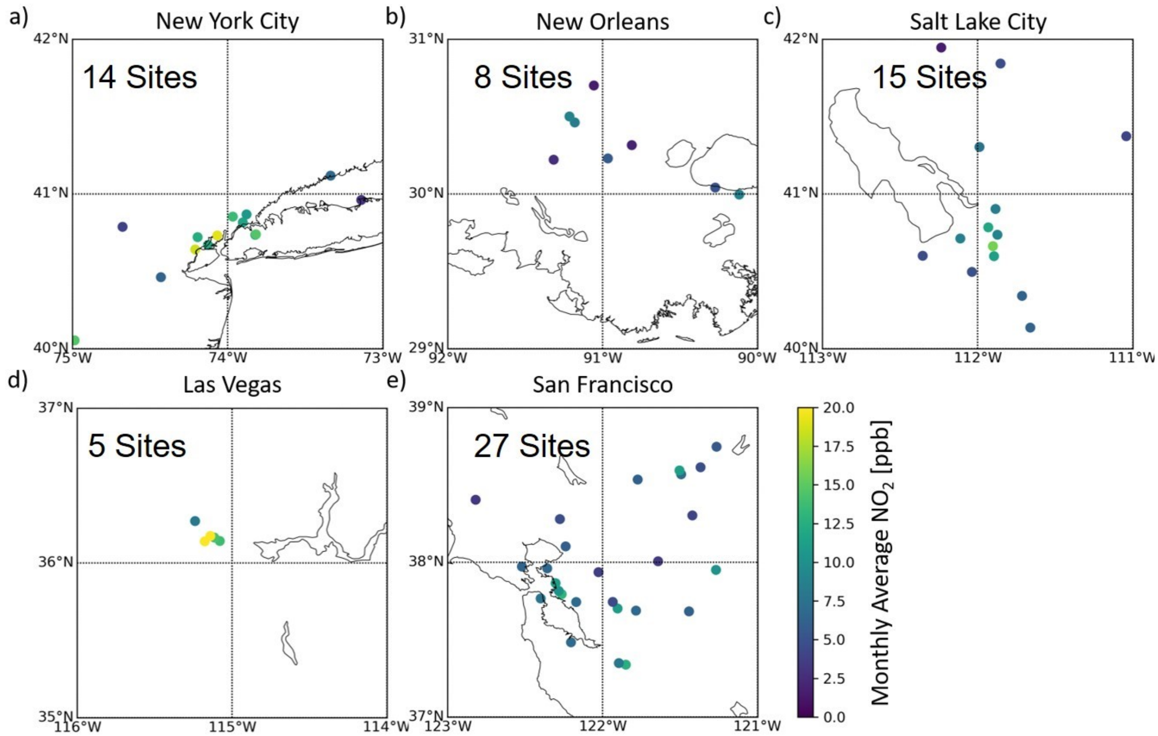
Ongoing & Future Work

- Sensitivity analysis, especially with respect to the number/density of ground measurement sites.
- Integrating data from low-cost sensor networks.
- Applying similar methods to other key air pollutants (Carbon Monoxide, Ozone, Sulfur Dioxide).
- Validating this method in areas of the world with sparse regulatory sensor networks.

(panel narration below):

0:00 / 1:37

A CASE STUDY FOR NITROGEN DIOXIDE IN MAJOR US CITIES (SEPTEMBER 2019)



DISCLOSURES

Carl Malings is supported by an appointment to the NASA Postdoctoral Program at the Goddard Space Flight Center, administered by Universities Space Research Association (USRA) through a contract with NASA.

AUTHOR INFORMATION

Carl Malings is a NASA Postdoctoral Fellow at the Goddard Space Flight Center in Greebelt, Maryland. Previously, he worked as a postdoctoral researcher for the "Make Air Quality Great Again" project in Paris, France, funded by the French government and at the Center for Atmospheric Particle Studies at Carnegie Mellon University, where he worked to calibrate and deploy low-cost air quality sensors.

ABSTRACT

Poor air quality is a major challenge for human health globally, especially in urban centers where emissions and exposure are both concentrated. Air quality monitoring has traditionally relied on ground-based measurements from a relatively small number of highly accurate but expensive regulatory-grade instruments, leading to limited spatial data coverage. More recently, these have been supplemented with other data sources, including remote satellite observations of pollutants in the atmospheric column, regional- and global-scale atmospheric chemistry model simulations, and less expensive in-situ monitoring systems allowing for denser spatial data collection at the expense of relatively lower accuracy compared to regulatory-grade monitors. Each of these air quality data sources individually have their own benefits and drawbacks, and therefore there is an opportunity to combine these disparate information sources while respecting their relative strengths and weaknesses in order to generate a more comprehensive and detailed picture of local air quality. This presentation investigates a potential approach to such a combination, with a focus on producing higher spatial resolution estimates and near-term forecasts of Nitrogen Dioxide in urban areas in the United States. Primary data sources include the GEOS Composition Forecasting (GEOS-CF) system, TROPOMI tropospheric NO_2 data products, ground measurements from the EPA regulatory monitoring network, and basic information on land usage and population density. Demonstrations of the system performance are presented for several urban areas in the United States, including New York City, New Orleans, San Francisco, Las Vegas, and Salt Lake City. Performance assessments are conducted across multiple seasons, and compared with simple persistence and climatological forecasts made using in-situ data only. Investigations of the relative benefits of different methods of combining the data sources are also conducted.

REFERENCES

Forouzanfar, M.H., Alexander, L., Anderson, H.R., Bachman, V.F., Biryukov, S., Brauer, M., Burnett, R., Casey, D., Coates, M.M., Cohen, A., et al., 2015. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. The Lancet 386, 2287–2323. https://doi.org/10.1016/S0140-6736(15)00128-2

Malings, C., Westervelt, D.M., Hauryliuk, A., Presto, A.A., Grieshop, A., Bittner, A., Beekmann, M., R. Subramanian, 2020. Application of low-cost fine particulate mass monitors to convert satellite aerosol optical depth to surface concentrations in North America and Africa. Atmos. Meas. Tech. 13, 3873–3892. https://doi.org/10.5194/amt-13-3873-2020

Martin, R.V., Brauer, M., van Donkelaar, A., Shaddick, G., Narain, U., Dey, S., 2019. No one knows which city has the highest concentration of fine particulate matter. Atmospheric Environment: X 3, 100040. https://doi.org/10.1016/j.aeaoa.2019.100040