ARISTOTLE CLOUD FEDERATION

**Use Case 3: High Fidelity Modeling and Analytics for Improved Understanding of Climate-Relevant Aerosol Properties**

S. Pryor (Cornell) with R. Levy (NASA Goddard), F. Yu (SUNY Albany), A. Hodzic (NCAR), P. Crippa (University of Newcastle), H. Matsui (Japan Agency for Marine-Earth Science and Technology)

***Motivation:*** Atmospheric aerosol particles (herein 'aerosols') cause regional and global climate change [1] and excess human mortality and morbidity, particularly under 'extreme aerosol events' (i.e., unusually high aerosol loading in a specific place and time) [2]. However, understanding of the predictability, length scales and covariance structures of aerosol properties at regional and sub-regional scales is incomplete and can be best improved by integration of remote sensing observations [3] and high-fidelity modeling. Relatively few researchers are equipped to overcome the associated technical challenges including; (1) integration of multi-scale data streams (currently in many disparate files with different structures, requiring pre-processing for georeferencing) to generate a holistic observationally derived view of aerosol particle populations, (2) the need for efficient WRF-Chem [4] simulations at sufficient resolution to match the observations and conduct sensitivity analyses for forensic diagnostics, and (3) the need for new 'smart' data analytics for model-observation cross evaluation and diagnostics.

***Activity:*** The proposed work comprises the expert pre-processing and integration of existing datasets into a consistent dataset of high value to the research community, coupled with optimization of a workflow for modeling with WRF-Chem [4] and model diagnostics. These data are already publicly available but the demands of storing and curating a multi-TB dataset of hundreds of thousands or millions of files, the sparse matrix nature of the data, and the heterogeneous nature of data access and data and metadata formats is a barrier to wider engagement of the science community with these data and has greatly constrained their use for diagnostic evaluation of atmospheric chemistry models such as WRF-Chem. We propose to lower those barriers. Examples of the remote sensing products that we intend to leverage are:

- MODerate resolution Imaging Spectroradiometer (MODIS) (on Aqua and Terra). Level 2 Aerosol Optical Depth (AOD = the column-integrated aerosol extinction coefficient $\propto$ columnar aerosol particle burden) and Angstrom exponent (AE, $\propto$ dominant particle size mode), (daily, 2/25/2000 – Present) (resolution: 10 km $\times$ 10 km). 122,000 HDF files (160GB) for the contiguous US.
- Multi-angle Imaging SpectroRadiometer (MISR) (on Terra). Level 2 AOD and AE (daily, 2/25/2000 – Present) (resolution: 17.6 km $\times$ 17.6 km). About 30,000 data files (400GB) for the contiguous US, plus ancillary geographic coordinate files for geolocating the observations.
- Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP Lidar on CALIPSO). Level 2 aerosol layer properties (once daily, 6/13/2006 – Present), 5 $\times$ 5 km horizontal resolution and a nominal vertical resolution of 60 m. This totals about 1.2TB of HDF data files, plus there are ancillary geographic coordinate files for geolocating the observations.
- Ground-based sun-photometer measurements of AOD and AE at about 384 irregularly spaced stations within North America in the AErosol RObotic NETwork (AERONET)



#1541215

(multiple times per day, start date varies with station). Data are stored as a single text file for each site.

- We also envisage leveraging the NASA MERRA-2 reanalysis produced at ~0.5×0.5° horizontal resolution, output at 1 to 3-hour intervals, which will be released in summer 2015. NASA currently intends the new reanalysis to be available for a 30-year period and that approximately 100 variables will be made publicly available (resulting in about 9 million files).

We will generate a consistent dataset from these heterogeneous sources, ready for processing by the research community, in partnership with iGlobe (Use Case 1) and provide a pipeline which can be run on the proposed cloud resource to process the data, significantly enhancing the utility of these heterogeneous datasets which are currently separated from processing resources. We will also implement a flexible, but efficient, workflow that facilitates high-resolution simulations with WRF-Chem and enables users to leverage the product of the remote sensing data integration in novel diagnostic model evaluation.

## References

[1] Stocker, T., Qin, D., Plattner, G., Tignor, T., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V. & Midgley, P. (2013). Summary for policymakers, climate change. *Phys. Sci. basis. Contrib. Work. Gr. I to Fifth Assess. Rep. Intergov. Panel Climate Change*, 1–27. Retrieved from: http://www.climatechange2013.org/images/report/WG1AR5_SPM_FINAL.pdf.
[2] Brook, R., et al. (2010). Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association, Circulation, 121(21), 2331–2378. Retrieved from: http://circ.ahajournals.org/content/121/21/2331.full.
[3] Kahn, R. (2012). Reducing the uncertainties in direct aerosol radiative forcing. *Surv. Geophys.*, 33(3-4), 701–721. Retrieved from: http://link.springer.com/article/10.1007%2Fs10712-011-9153-z.
[4] Grell, G., Fast, J., Gustafson Jr., W., Peckham, S., McKeen, S., Salzmann, M. & Freitas, S. (2011). On-line chemistry within WRF: description and evaluation of a state-of-the-art multiscale air quality and weather prediction model. *Integrated Systems of Meso-Meteorological and Chemical Transport Models*, 41–54. Retrieved from: http://link.springer.com/chapter/10.1007%2F978-3-642-13980-2_3.



#1541215