Improved Algorithms for Fast Spatio-Temporal Interpolation of Photochemical Air Quality Model Results

Sastry S. Isukapalli, Vikram M. Vyas, and Panos G. Georgopoulos • Computational Chemodynamics Laboratory (www.ccl.rutgers.edu)
Environmental and Occupational Health Sciences Institute, a Joint Institute of UMDNJ-Robert Wood Johnson Medical School and Rutgers University, Piscataway, NJ

Summary

Context:

- Ambient air quality information is often needed at high resolutions
  - Higher than the typical 4km x 4km or 2km x 2km resolution of the results from photochemical air quality models
  - Usually at the census tract level (less than 1km x 1km area)

Possible Approaches:

- Running local scale models using result of grid models
  - Requires fine scale details of meteorology and emissions
- Spatio-temporal interpolation techniques
  - Spatio-Temporal Random Field (STRF) theory (Vyas & Christakos, 1997)
  - Bayesian Maximum Entropy (BME) (Serre & Christakos, 1999)

Current Limitations:

- Computationally demanding/impractical to use
  - STRF and BME designed for "irregular data"
  - Therefore, potential optimizations for "grids" have not been used so far

Algorithmic Improvements:

- Utilizing the space and time symmetry properties of regular grids
- Caching and reusing prior calculations of neighbors and distances

Motivation for Spatio-Temporal Interpolation

- Census tract level concentrations are required for population exposure calculations using models such as MENTOR/SHEDS-1A
- In densely populated urban areas, census tracts are much smaller than 4 km x 4 km in area. Regional photochemical air quality models such as CMAQ are usually not run at more than 4 km x 4 km resolution
- Simultaneous treatment of spatial and temporal domains permits evaluation of spatial factors as well as temporal evolution in observed or modeled data
- Spatio-temporal interpolation techniques such as STRF (Spatio-Temporal Random Field) and BME (Bayesian Maximum Entropy) offer a promising solution for obtaining air quality concentration distributions down to the required census tract level

Basics of STRF Methodology

Spatiotemporal estimate from available data, as $\hat{X}(\mathbf{s}_k,t_k)=\sum_{i=1}^{m}\lambda_i X(\mathbf{s}_i,t_i)$

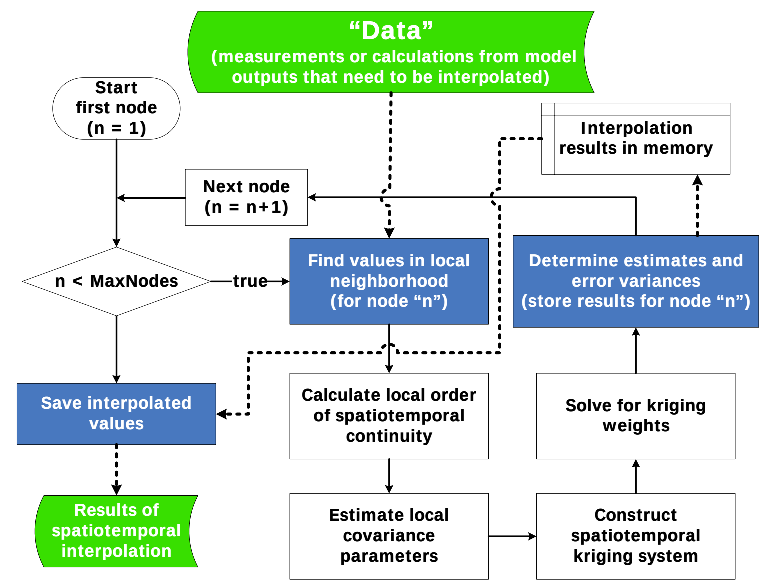
Subject to $\sum_{i=1}^{m}\lambda_i P(\mathbf{s}_i,t_i)=P(\mathbf{s}_k,t_k)$

Estimation error (or residual) vector : $\mathbf{Y}=\hat{\mathbf{X}}-\mathbf{X}$

Optimum estimates are obtained by minimizing error variance : $\sigma_X^2=E[\hat{\mathbf{X}}-\mathbf{X}]^2$

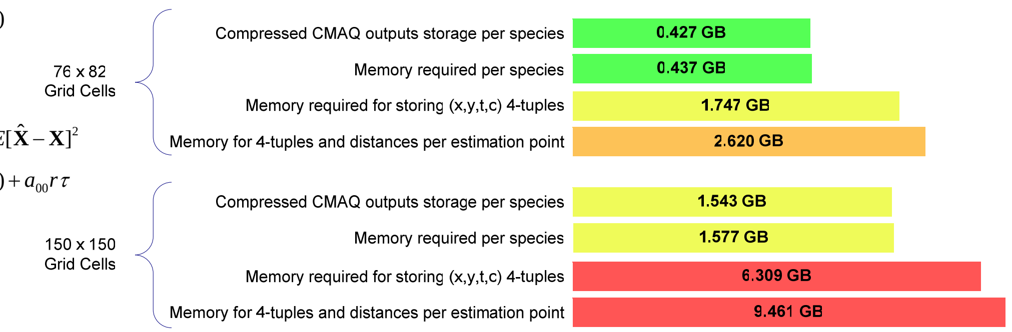
Polynomial covariance used here : $k_x(r,\tau)=c\delta(r)\delta(\tau)-a_0\tau\delta(r)-b_0 r\delta(\tau)+a_{00}r\tau$

Algorithm for Traditional STRF Implementation



The traditional algorithm for STRF is geared towards "irregular data" (in either space or time), thus the process of finding "neighbors" is repeated for each point in space and time ("node").

Space Requirements for Traditional STRF Implementation

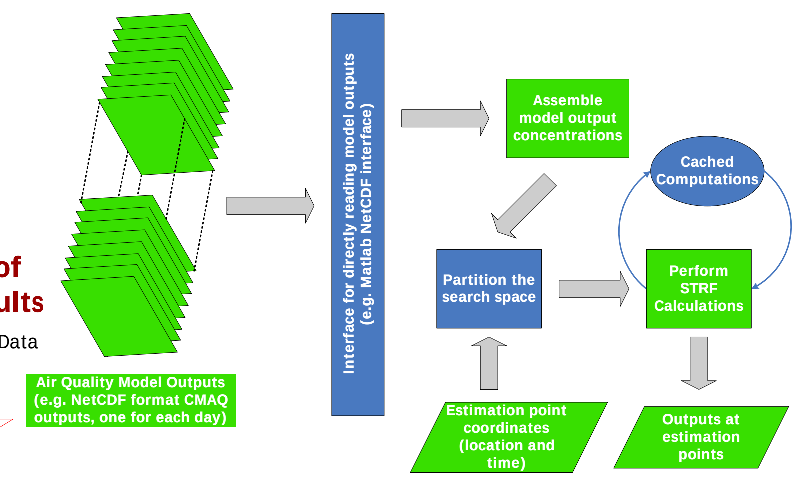


The space requirements for the traditional implementation of STRF are prohibitive, and for some data sets (e.g. an annual simulation results at hourly time resolution and 4 km x 4 km spatial resolution, corresponding to 76 x 82 cells), the STRF method cannot be applied as is.

As part of this work, in the initial stage, some minor optimizations were performed, which allowed the application of the STRF method, and the time requirements were of the order of several weeks.

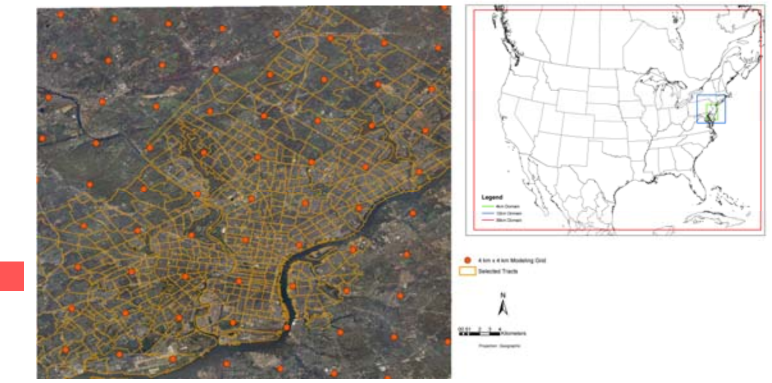
The method presented here incorporates those optimizations and includes additional optimizations that reduce the time requirements for the entire process to about 10 hours.

Algorithmic Improvements in the Current Implementation



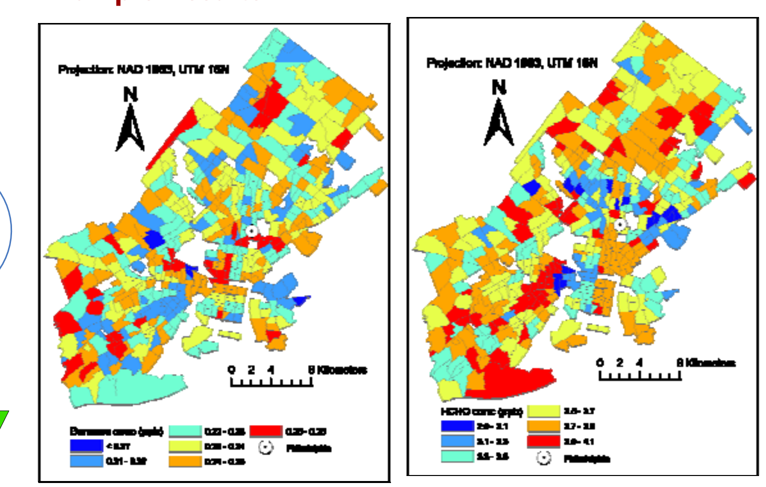
The "partitioning" of the neighborhood space and "caching" of neighborhood searches and distances resulted in savings of computational time from order of several weeks (in addition to several manual "pre-processing" steps) to about ten hours; in this process, most manual steps eliminated

Case Study: MENTOR/SHEDS-1A Application for the Philadelphia/Camden Region



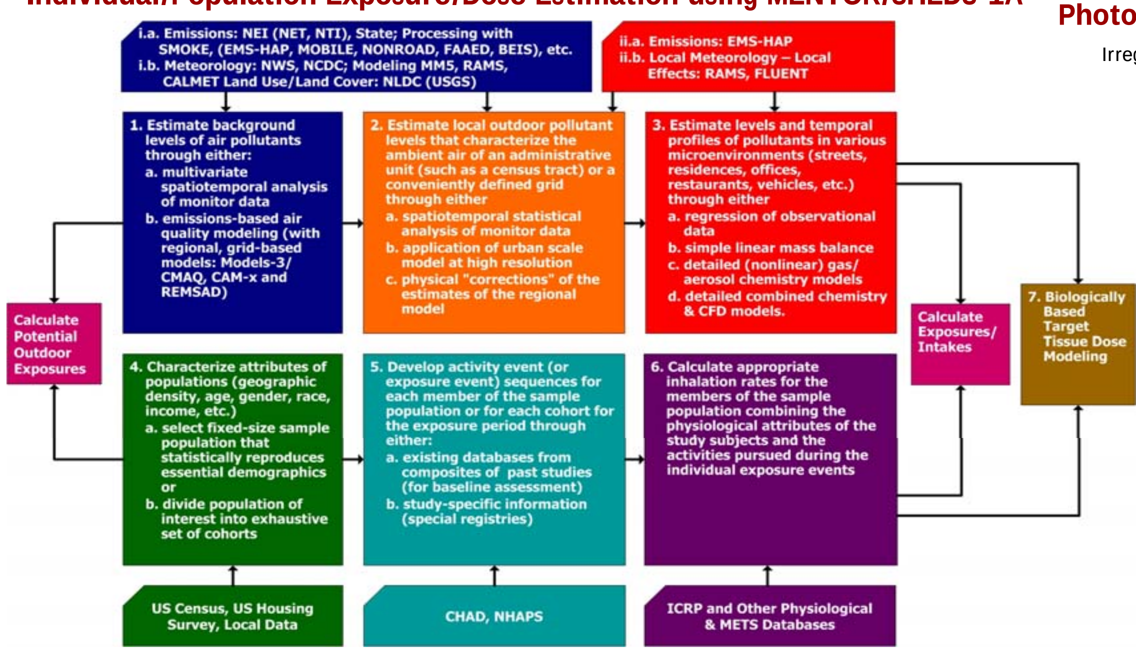
- Year long simulation for the year 2001 (performed by USEPA-ORD)
  - CMAQ model with a nested grid (36, 12, and 4 km resolutions)
- 4 km x 4 km resolution around the Philadelphia, PA region
- Grid Model: Emissions-based modeling of 20 air toxics in addition to ozone (CMAQ with the toxics version of the SAPRC99 mechanism)
- Need: Census-tract level concentrations for all species for entire year
- Method: interpolate 4 km x 4 km model results using the STRF

Example Results



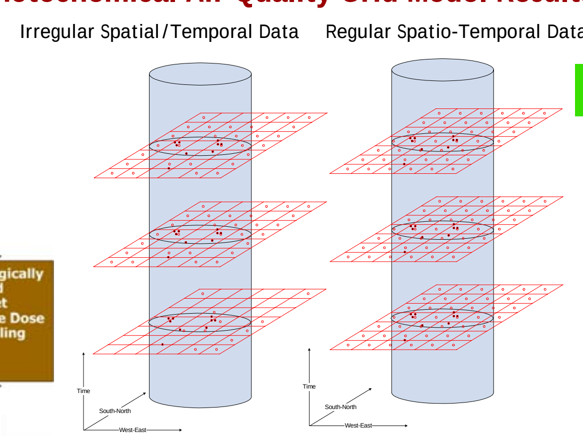
Census tract level estimates of formaldehyde and benzene concentrations for 10 am on Jul 18, and Jan 24, 2001, respectively using the STRF method. The results from the CMAQ model run at 4 km x 4 km resolution were used to obtain these estimates.

Individual/Population Exposure/Dose Estimation using MENTOR/SHEDS-1A



Steps in estimating individual/population exposures using MENTOR/SHEDS-1A [Modeling ENvironment for TOtal Risk studies (MENTOR) using the Stochastic Human Exposure and Dose Simulation (SHEDS) approach in a "One Atmosphere" (1A) setting] (Georgopoulos et al., 2005): Step 2 involves estimation of local level outdoor concentrations. The STRF method can be used for spatio-temporal interpolation of model results.

Utilizing Spatial and Temporal Regularity of Photochemical Air Quality Grid Model Results



Spatio-temporal interpolation for cases involving irregular spatial/temporal data versus regular spatio-temporal data. When regular data are available (as in the case of outputs of models such as CMAQ), algorithmic optimizations can be utilized in order to perform spatio-temporal interpolations in a practical time frame. Empty circles denote measurements, and filled circles (black) denote the points at which the concentration values are needed.

Ongoing Testing and Evaluation of the new SRTF Implementation

Ongoing work evaluates utility of STRF in interpolating coarser scale model runs as an alternative to model runs at finer scales

- CMAQ model runs at 36 km, 12 km, and 4 km resolutions are being used for this evaluation; current improvements in STRF facilitate such evaluations with large sets of model results

Sensitivity analysis of the STRF algorithm

- for model results using different sizes of neighbor sets
  - for different neighbor distribution (e.g. a minimum of 12 points are needed to cover all quadrants in spatio-temporal space around the "node")
- for cases involving missing neighbors

Case studies will focus on small spatio-temporal domains for multiple species

- Initial set will involve ozone, formaldehyde, benzene, and carbon monoxide

Acknowledgements

Base Funding for the Ozone Research Center is Provided by the State of New Jersey Department of Environmental Protection. The work presented here has been funded by a University Partnership Agreement (UPA) between the USEPA and EOHSI (Cooperative Agreement #831625).

References

P.G. Georgopoulos, S.W. Wang, V.M. Vyas, Q. Sun, J. Burke, R. Vedantham, T. McCurdy, and H. Özkaynak, "A Source-To-Dose Assessment of Population Exposures to Fine PM and Ozone in Philadelphia, PA, During a Summer 1999 Episode," Journal of Exposure Analysis and Environmental Epidemiology 15, 439-457 (2005).

M. Serre and G. Christakos, "Modern Geostatistics: Computational BME in the Light of Uncertain Physical Knowledge -- The Equus Beds Study," Stochastic Environmental Research and Risk Assessment 13, 1-26 (1999).

V.M. Vyas and G. Christakos, "Spatiotemporal Analysis and Mapping of Sulfate Deposition Data Over Eastern USA," Atmospheric Environment 31, 3623--3633 (1997).