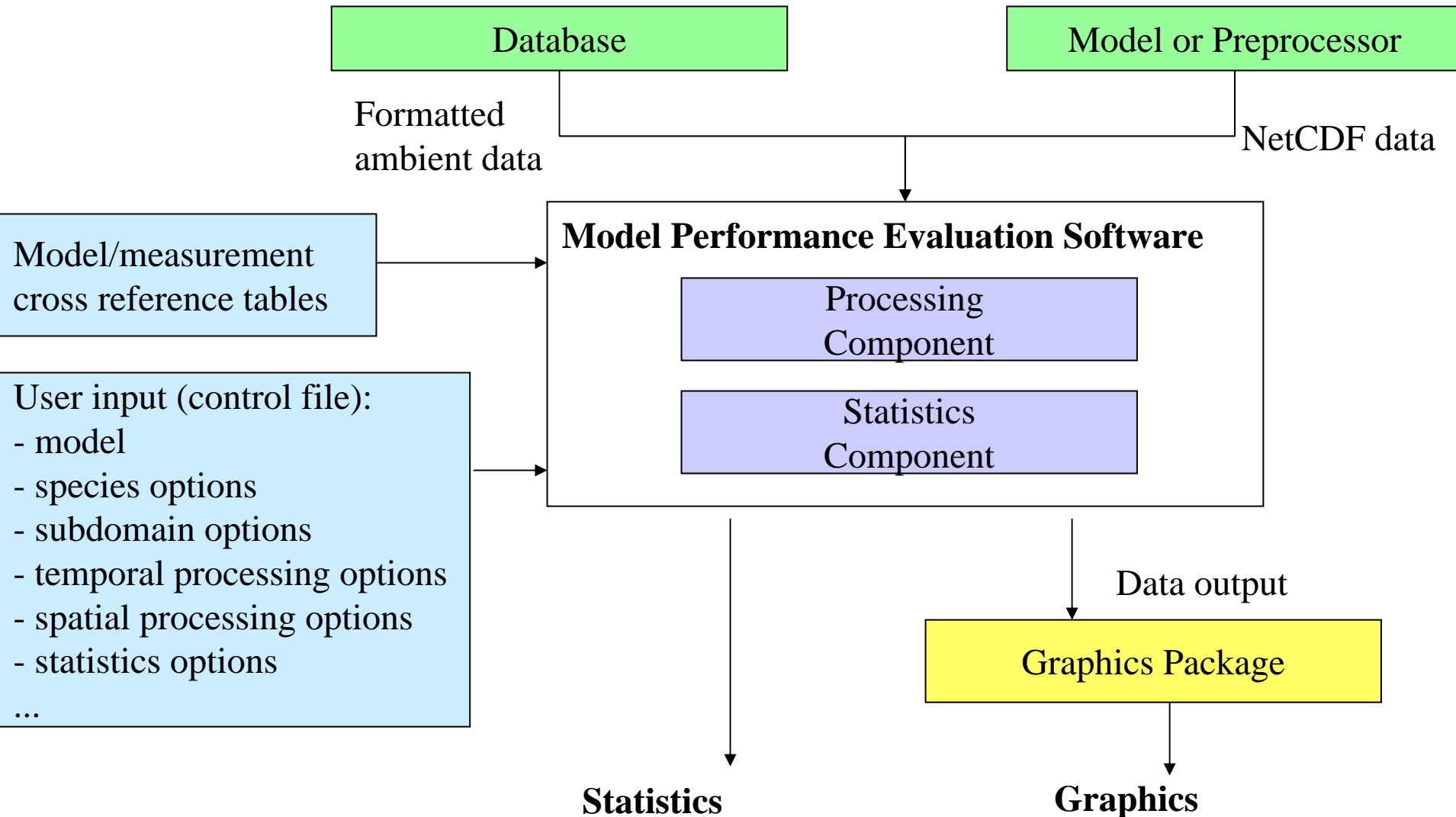


Model Performance Evaluation Software: Description and Application to Regional Haze Modeling

Betty K. Pun, Kristen Lohman, Shiang-Yuh Wu, Shu-Yun Chen,
Christian Seigneur
AER, San Ramon, CA

Presentation at the CMAS Conference, Chapel Hill, NC
27 September 2005

Funding for the development of the MPE database and software was provided by
CENRAP (Modeling Umbrella Contract 02-00307-RP-005 Work Order 1)



Data for PM Model Performance Evaluation

- IMPROVE: 24 hour $PM_{2.5}$ and component* data, one in three days
- CASTNet: 7-day sulfate, nitrate, ammonium, SO_2 , and HNO_3 (may be problematic)
- AQS: hourly to daily PM_{10} , $PM_{2.5}$, $PM_{2.5}$ speciation*, O_3 , NO , NO_2 , NO_y , VOC (ppb or ppbC), SO_2
- NADP: wet deposition fluxes, concentrations of sulfate, nitrate, ammonium in precipitation
- Other

*PM fraction can be evaluated if both total mass and component mass are available at the same site

How to Consolidate, Store, and Retrieve Ambient Data for MPE?

What we have:

- Many data sources
- Different formats
- Supporting information sometimes separate from data and sometimes difficult to find

MySQL database

- upload data (measurement, site)
- compile other relevant information
- use query to retrieve data in consistent format

What we need:

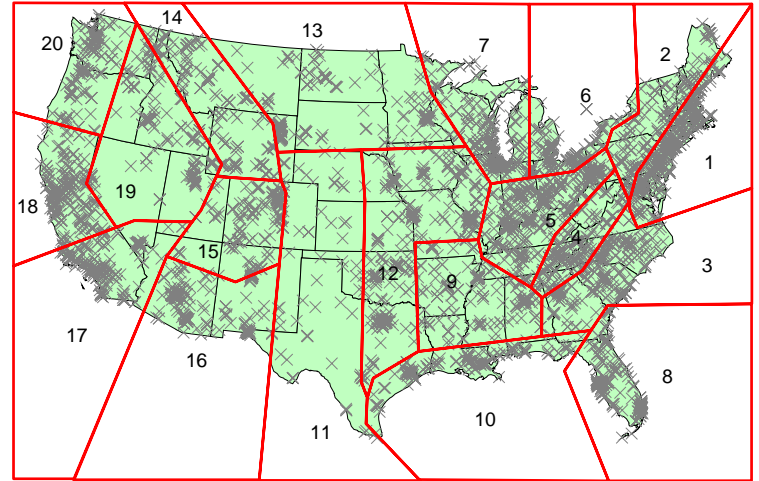
- Preferably one data source
- Consistent format
- Supporting information
 - site location
 - sample start time including time zone
 - sample duration
 - units

- Read observations
 - perform time zone changes
 - average to longer periods
 - perform unit conversions (may require meteorology)
- Extract modeling data (NetCDF)
 - calculate grid cell corresponding to latitude/longitude
 - extract model data at grid cell(s)
 - sum components of species (look up table)
 - perform unit conversions

Subdomain and Temporal Options

Subdomains

- All sites
- A list (e.g., a city)
- Ranges
 - latitude and longitude
 - cell numbers in x and y
- Performance evaluation zones

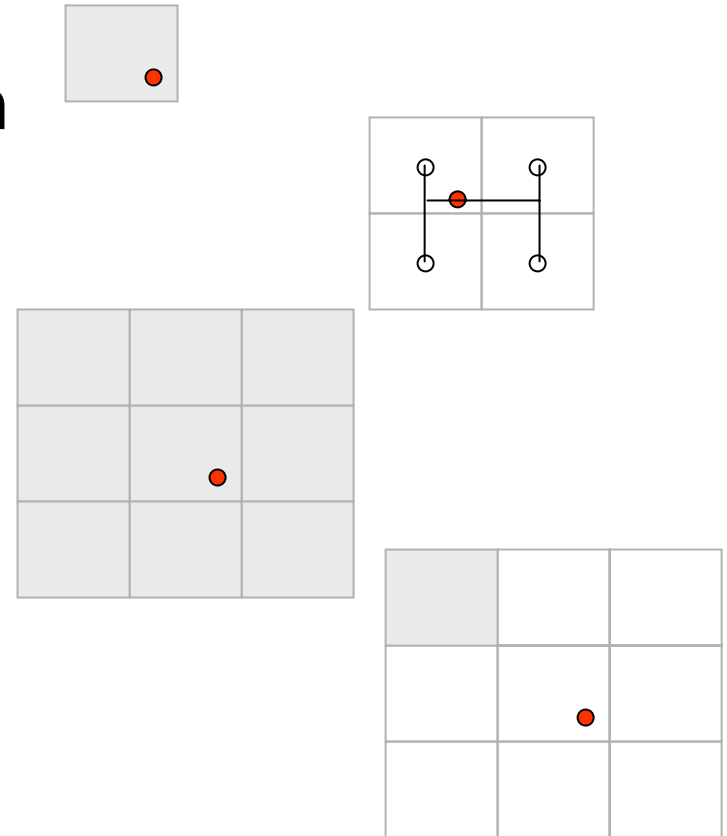


Temporal

- Subperiods
- Aggregation of high frequency measurements (data completeness criterion)

Spatial Processing Options

- Model value at grid cell corresponding to site location
- Linear interpolation using 4 closest grid cell
- Average of all cells within a user specified window
- Best estimate with a user specified window
- Distance-weighting method within user specified window



$$V(x, y) = \sum_{i=1}^n w_i V_i; \text{ where } w_i = \frac{h_i^{-2}}{\sum_{j=1}^n h_j^{-2}} \text{ and } h_i = \sqrt{(x - x_i)^2 + (y - y_i)^2}$$

- Threshold to be selected by user
- Default output metrics
 - accuracy of peak (unpaired in time): paired and unpaired in space
 - mean observed and modeled value
 - Gross and normalized bias and error
 - Coefficient of correlation
 - Normalized root mean square error
- Optional output metrics
 - ratio of means, fractional bias and error, r^2 , index of agreement, site specific root mean square error, normalized mean bias and error, slope and intercept

An MPE Database and Software Designed for Community Use

- Comprehensive Processor
 - data base; binary to NetCDF converter; software
 - output compatible with common graphics software
- Versatility
 - PM, PM components, gases, deposition fluxes
- User-friendly Design
 - Community input for software features
 - Software engineering standard
 - Based on Fortran 90 and MySQL (free!)
- Documentation

Example Results (BRAVO)

