COMPARISON OF SULFATE CONCENTRATIONS SIMULATED BY TWO REGIONAL-SCALE MODELS WITH MEASUREMENTS FROM THE IMPROVE NETWORK

John S. Irwin[1], *Edith Gego*[2], *Christian Hogrefe*[3], *Jennifer M, Jones*[3] *and S. Trivikrama Rao*[1]

1 NOAA Atmospheric Sciences Modeling Division, RTP, NC, U.S.A
(On assignment to the U.S. Environmental Protection Agency).
2 University Corporation for Atmospheric Research, Idaho Falls, ID, U.S.A.
3 Atmospheric Sciences Research Center, University at Albany, Albany, NY. U.S.A.

9th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, June 1-4, 2004, Garmish-Partenkirchen, Germany

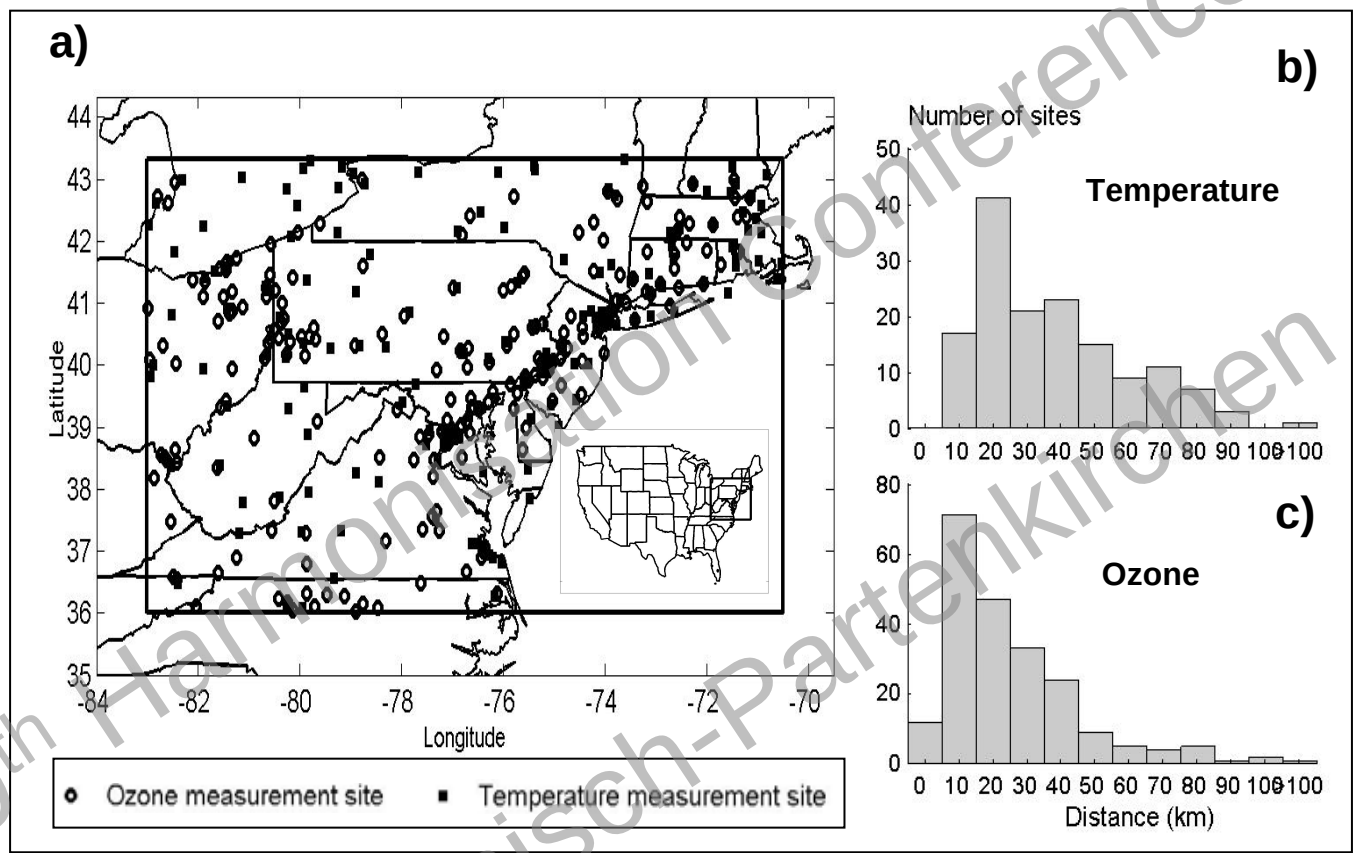
9th Harmonisation Conference Garmisch-Partenkirchen

Overview

- Some of the many issues that we worry with in assessing regional-scale air quality model performance.
- What are known model strengths and weaknesses.
- Is there a natural way to “group” data for analysis?
- Can we be objective in our model comparisons?

9th Harmonisation Conference
Garmisch-Partenkirchen

Spatial Resolution Problems

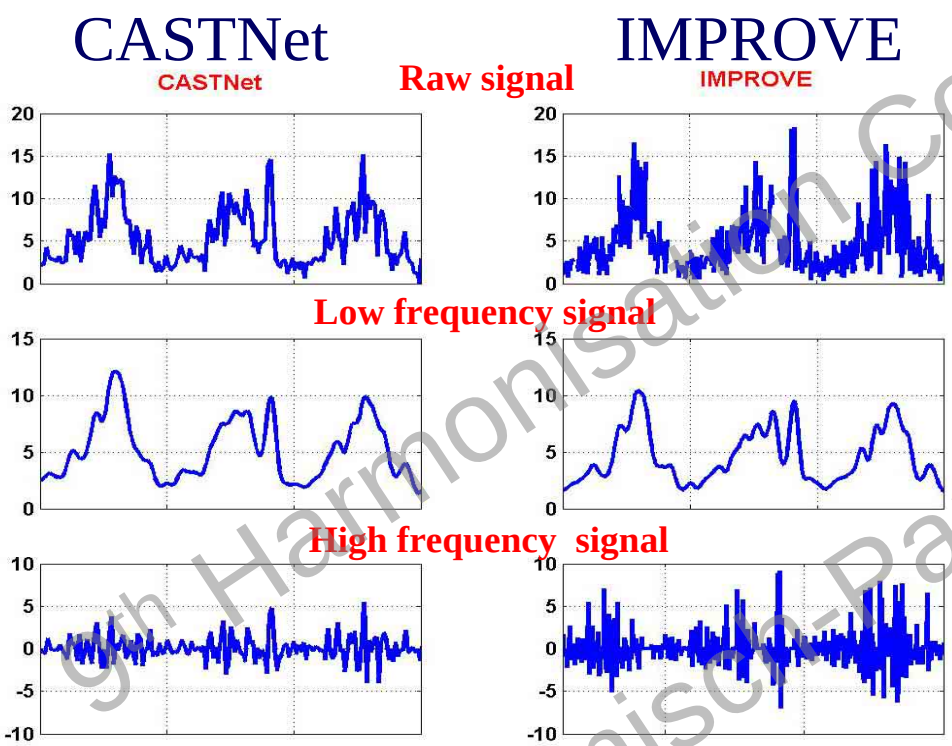


9th Harmonization Conference
Garmisch-Partenkirchen

Panel a, location of the analysis domain and of the temperature and ozone measurement sites. Panel b, histogram of the distance between temperature measurement sites and their nearest neighbors. Panel c, histogram of the distance between ozone monitoring sites and their nearest neighbors. Modal separation distance for temperature measurements is 20 km. Modal separation distance for ozone monitors is 10 km.

Network Comparison Problems:

Decomposition of Sulfate Time Series



IMPROVE data show more variability than their CASTNet counterparts

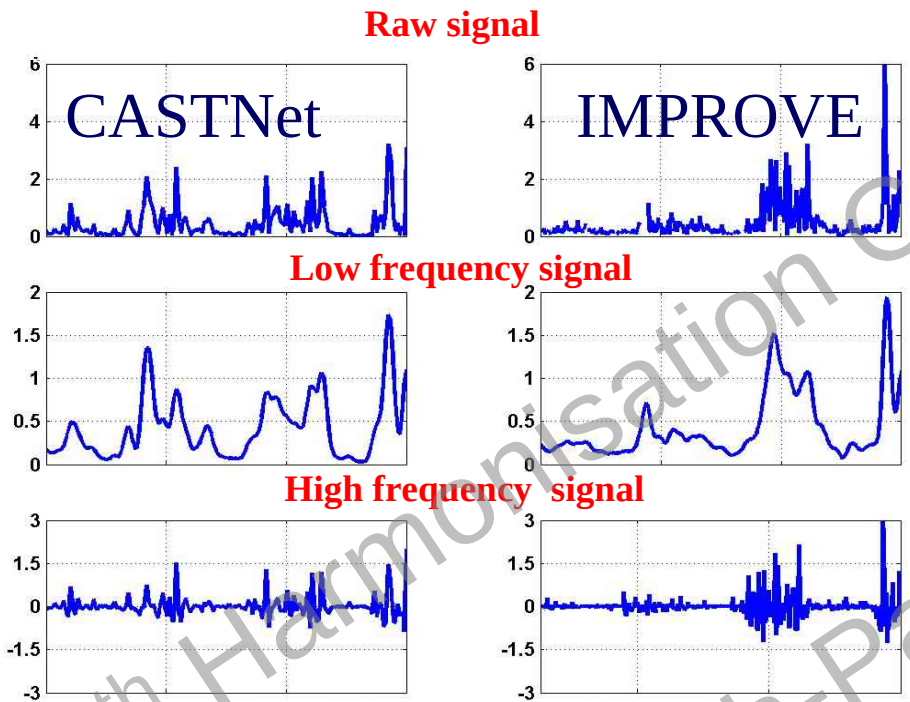
Low-frequency signals of both networks are very similar

Variance of high-frequency CASTNet signal smaller than variance of high-frequency IMPROVE signal

Probably because IMPROVE data are 24-h averages while CASTNet data are weekly averages

9th Harmonisation Conference Garmisch-Partenkirchen

Decomposition of Nitrate Time Series



IMPROVE data show more variability than their CASTNet counterparts

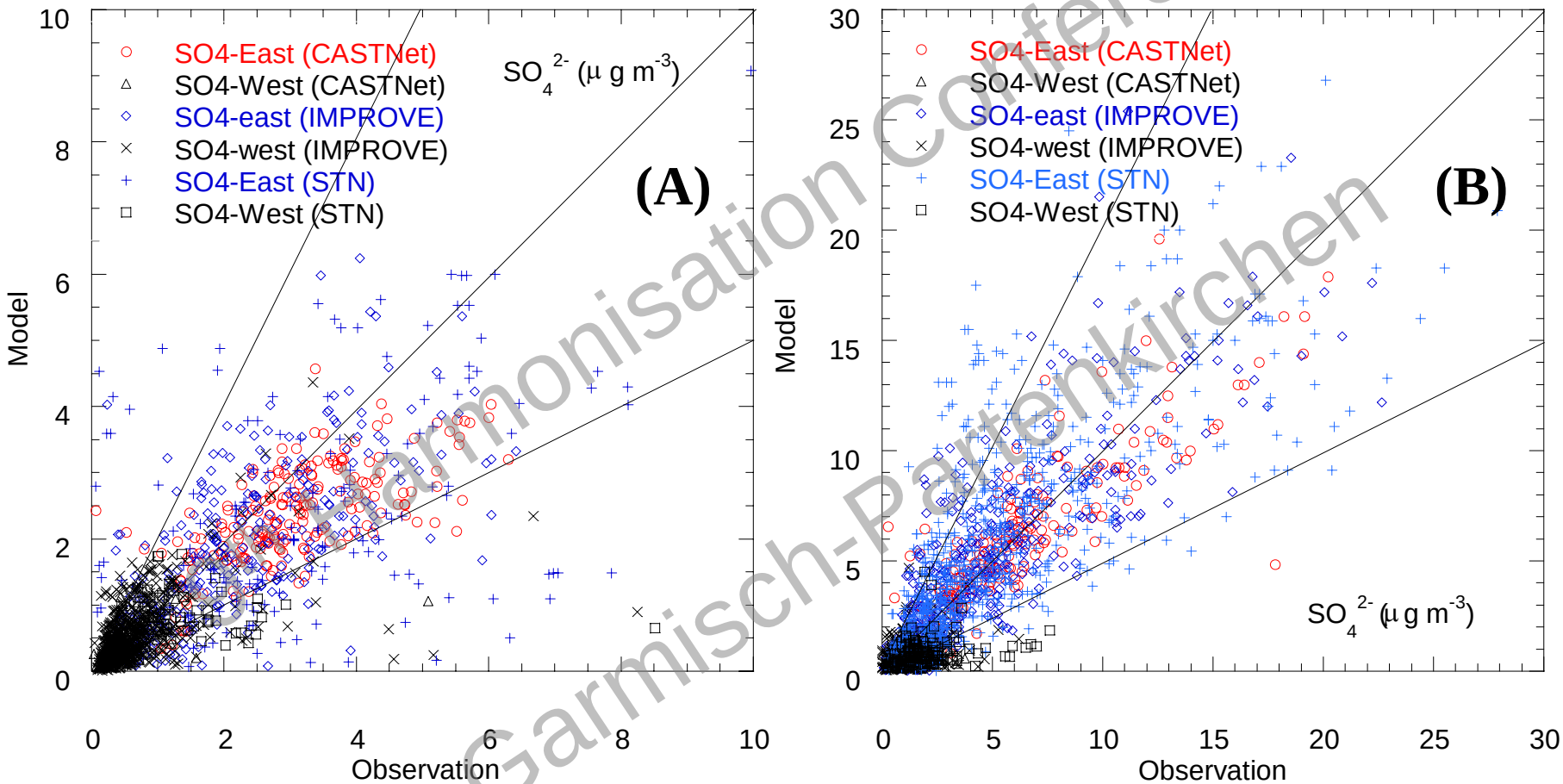
Differences between the two low-frequency signals are most pronounced during the winter season (high concentration season for nitrate)

Distinct sampling protocols may explain these dissimilarities. Nitrate is extracted from a Teflon filter in the CASTNet network while it is extracted from a nylon filter preceded by a HNO_3 denuder at IMPROVE sites.

9th Harmonisation Conference Garmisch-Partenkirchen

Communication of Model Performance Problems

Decomposing and deciphering how model performance varies in time and space.



Air Pollution Modeling and its Application XXIII Conference
Garmisch-Partenkirchen

Comparison of CMAQ model estimates of 24-hr average sulfate values with observations from three networks for (A) January 1-28, 2001 and (B) July 3-30, 2001.

So what do we conclude from the previous slides?

- The distance between monitors is too great to resolve fine scale features, nor can we confidently interpolate between monitors to define spatial patterns.

- We might combine measurements from monitors from different networks for sulfate, but cannot do so with nitrate.

9th Harmonisation Conference Garmisch-Partenkirchen

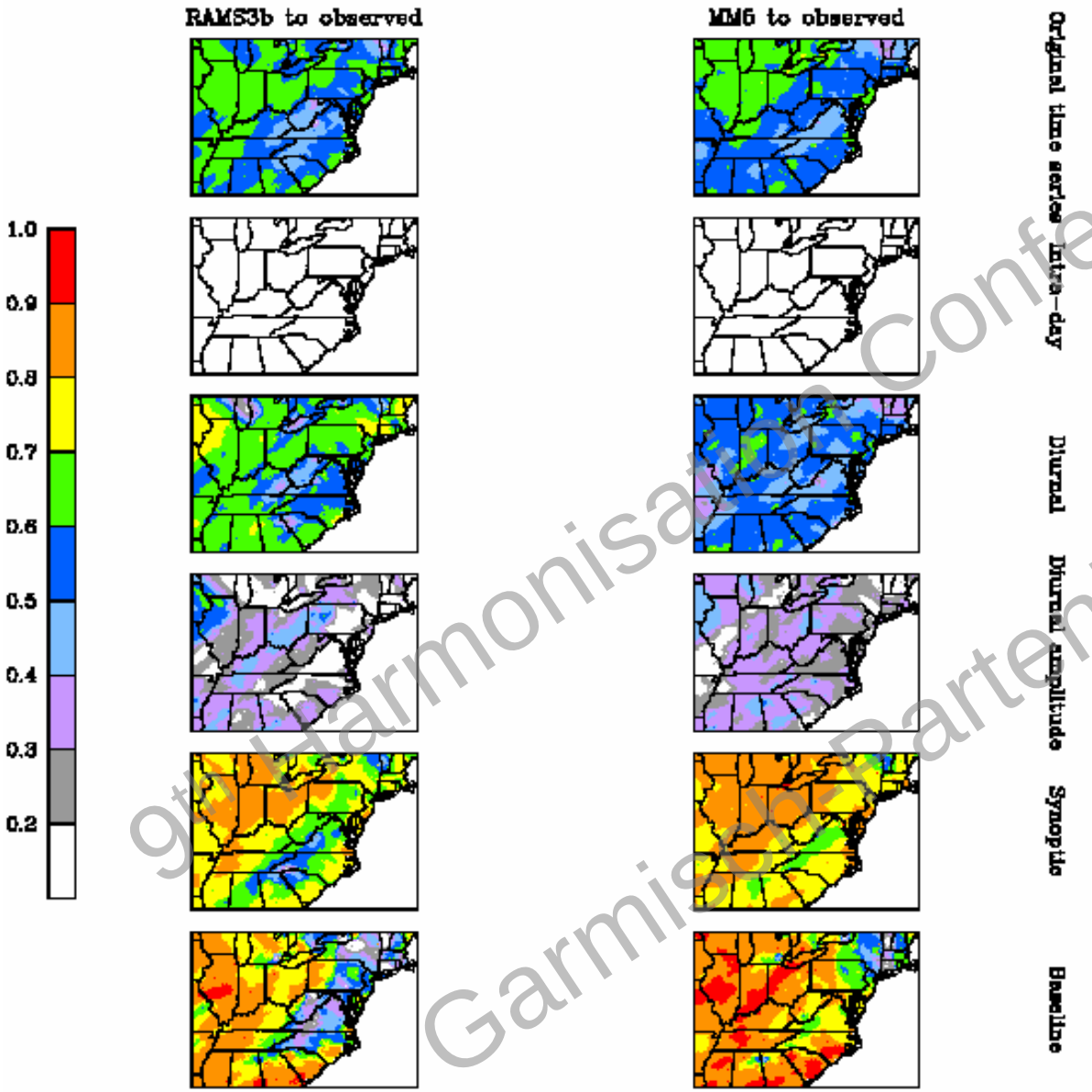
Other Issues to Work Around

- Models are estimates of what will happen “in general and on average” (i.e., spatial patterns and longer-term variations), what we see is something in the “particular.”

- Hogrefe et al., (2001ab) suggests that grid-based models are most skillful in simulating longer-term variations in time and space, as they lack the resolution and physics to simulate finer-scale variations. (Next two slides)

- Gego et al., (2003) detected through the use of a principal component analysis (PCA) of observations of sulfate there are locally contiguous regions where monitoring results have similar temporal behavior. (Next six slides thereafter)

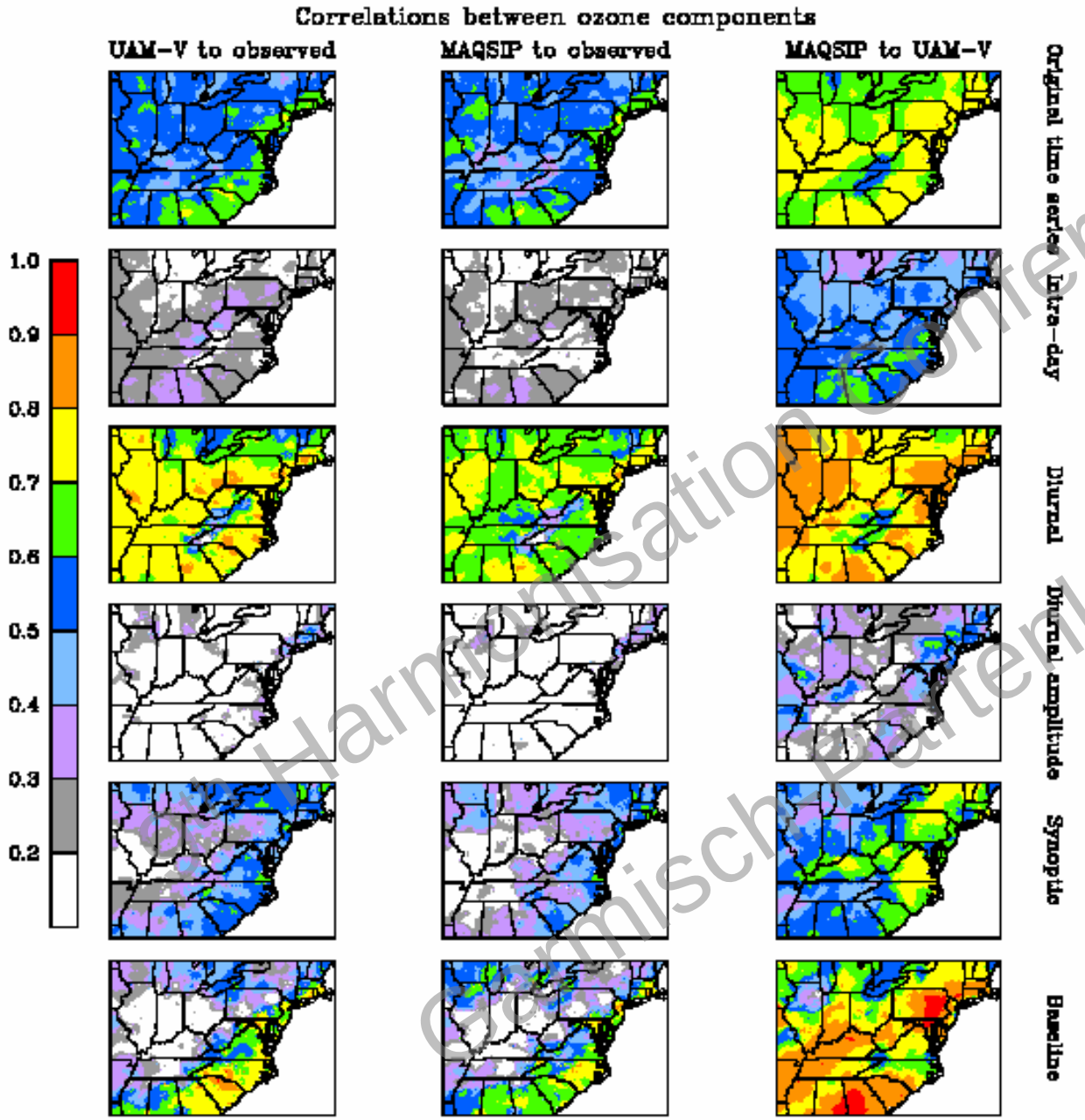
9th Harmonisation Conference
Garmisch-Partenkirchen



9th Harmonisation Conference
Garmisch-Partenkirchen

Correlation between wind speed component time series from observations and the RAMS3b and MM5 models, June – August, 1995.

Strongest correlations for both meso-scale meteorological models is for the synoptic and baseline (seasonal) variations. Both are weakest in replicating the hourly fluctuations that are superimposed on the diurnal variation.



Correlations between modeled and observed ozone.

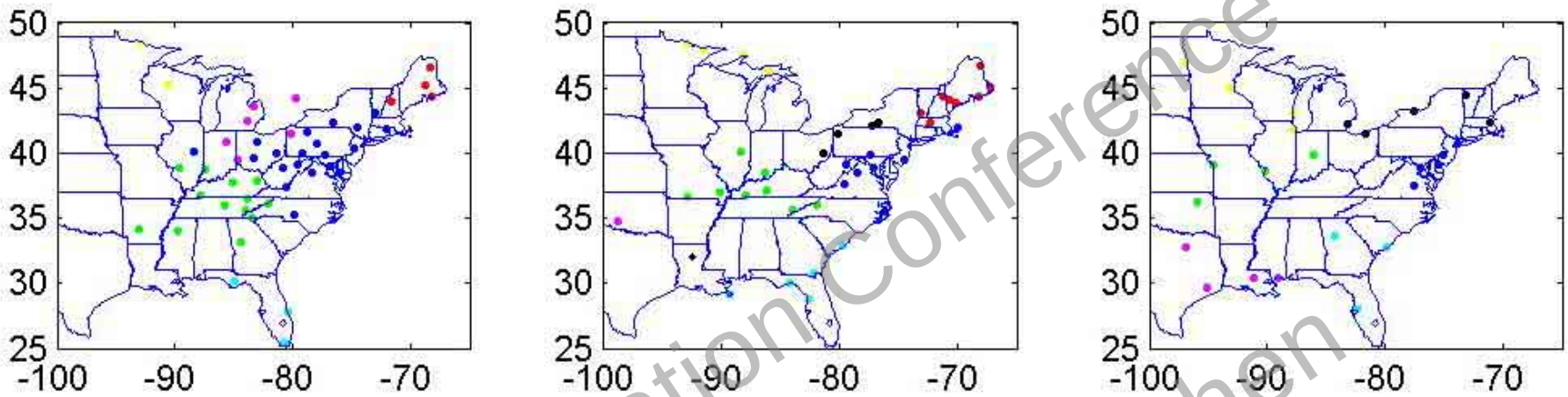
Models consistently correlate with each other better than they correlate with observations.

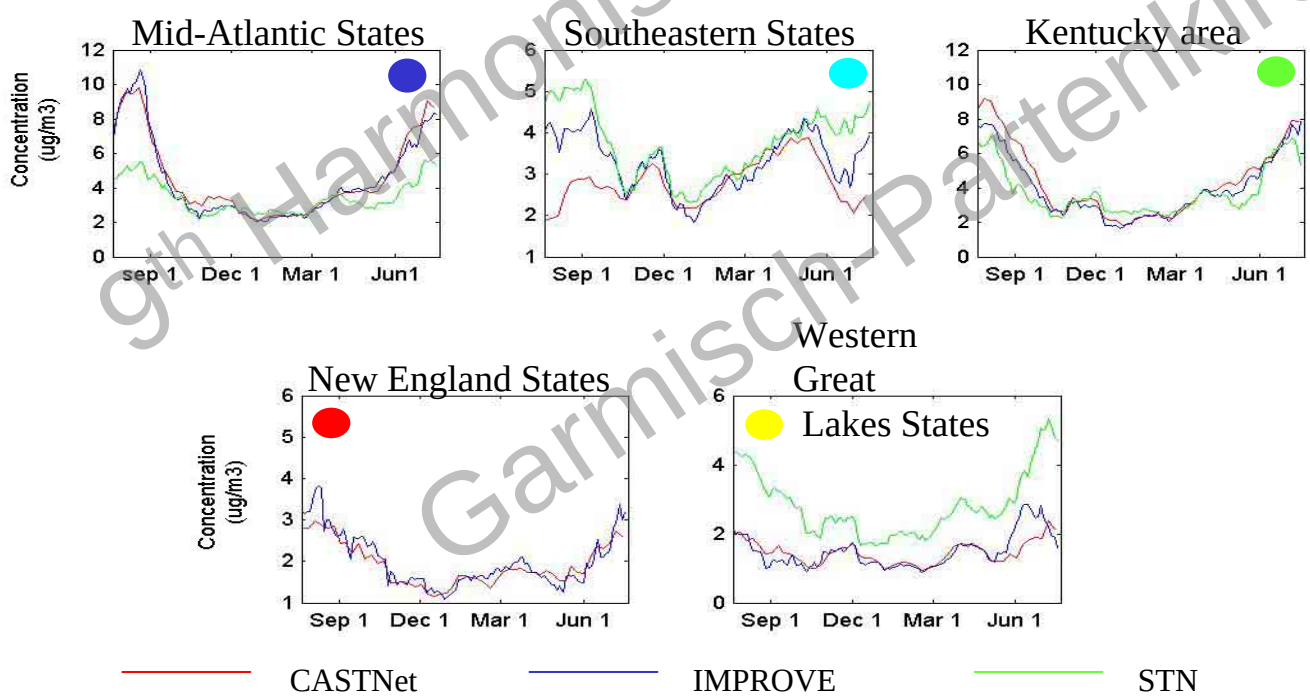
Models best predict the diurnal, synoptic and baseline variations.

Models do poorly in replicating the inter-hourly variations and the diurnal amplitude.

Harmonisation Conference
Garmisch-Partenkirchen

Sulfate **CASTNet** **IMPROVE** **STN**





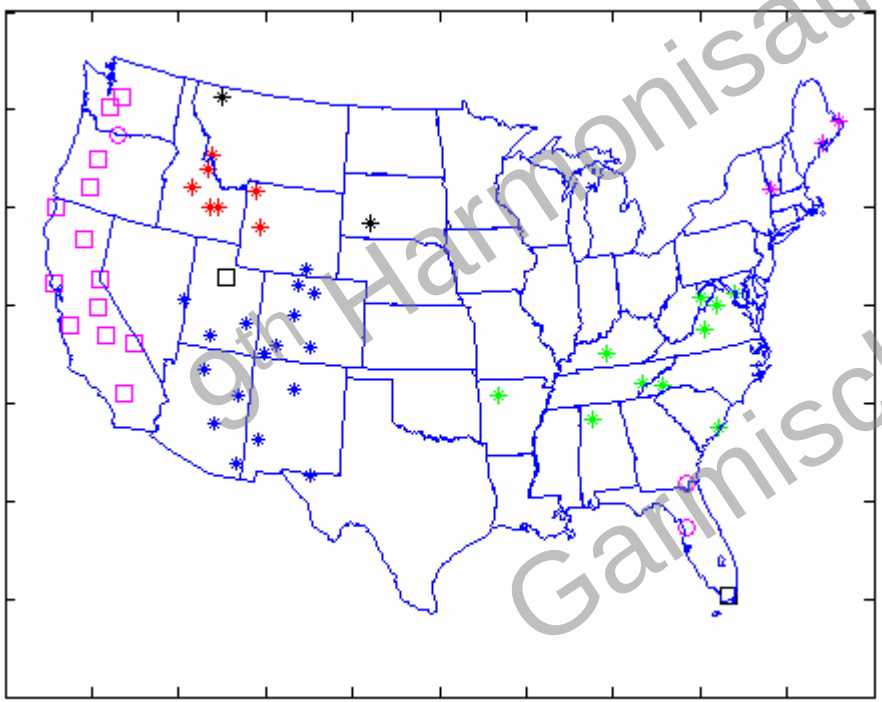
PCA results for sulfate at sites located east of -100º longitude (eastern U.S.), from July 1st 2001 to July 31st 2002, for those sites with less than 20 % missing values.

Similar groups are formed by the three networks, although some differences are evident in the time series within the groups.

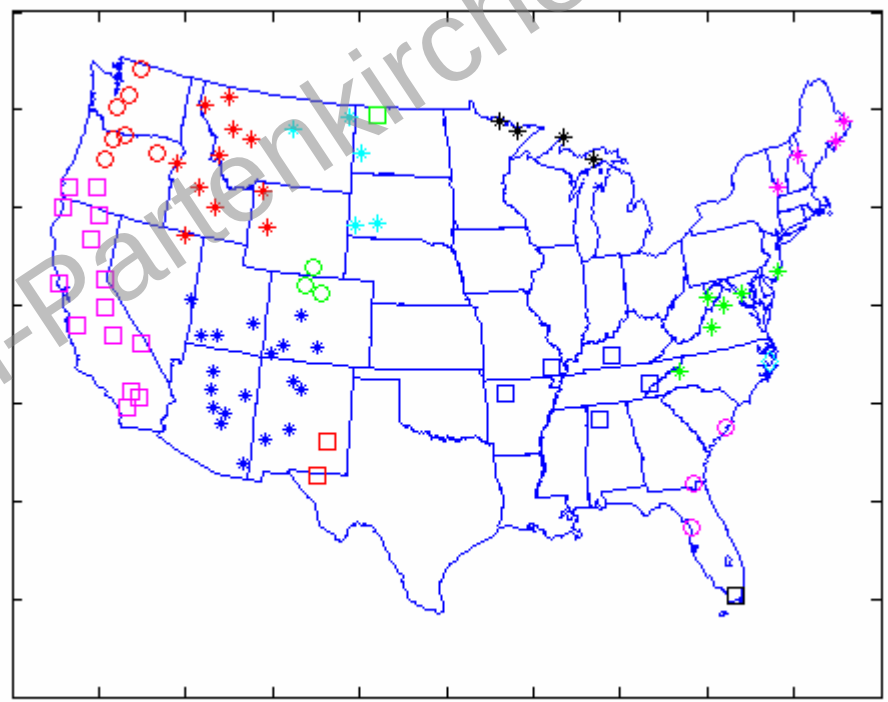
9th Harmonisation Conference Garmisch-Partenkirchen

Comparison of PCA classification for sulfate using IMPROVE sites, similar on the large scale with minor year-to-year differences in the details.

PCA Results for 1996
IMPROVE Measurements

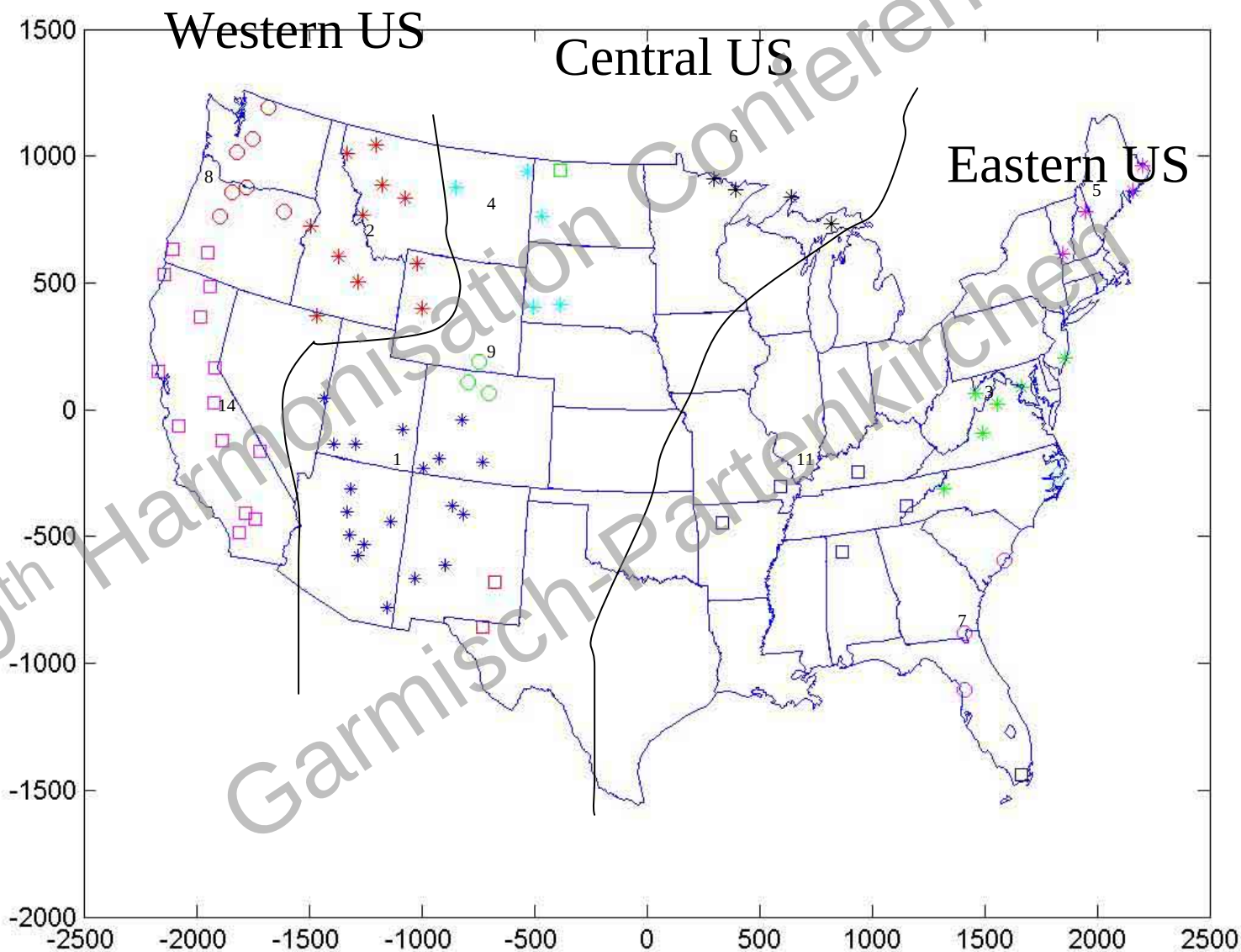


PCA Results for 2001
IMPROVE Measurements



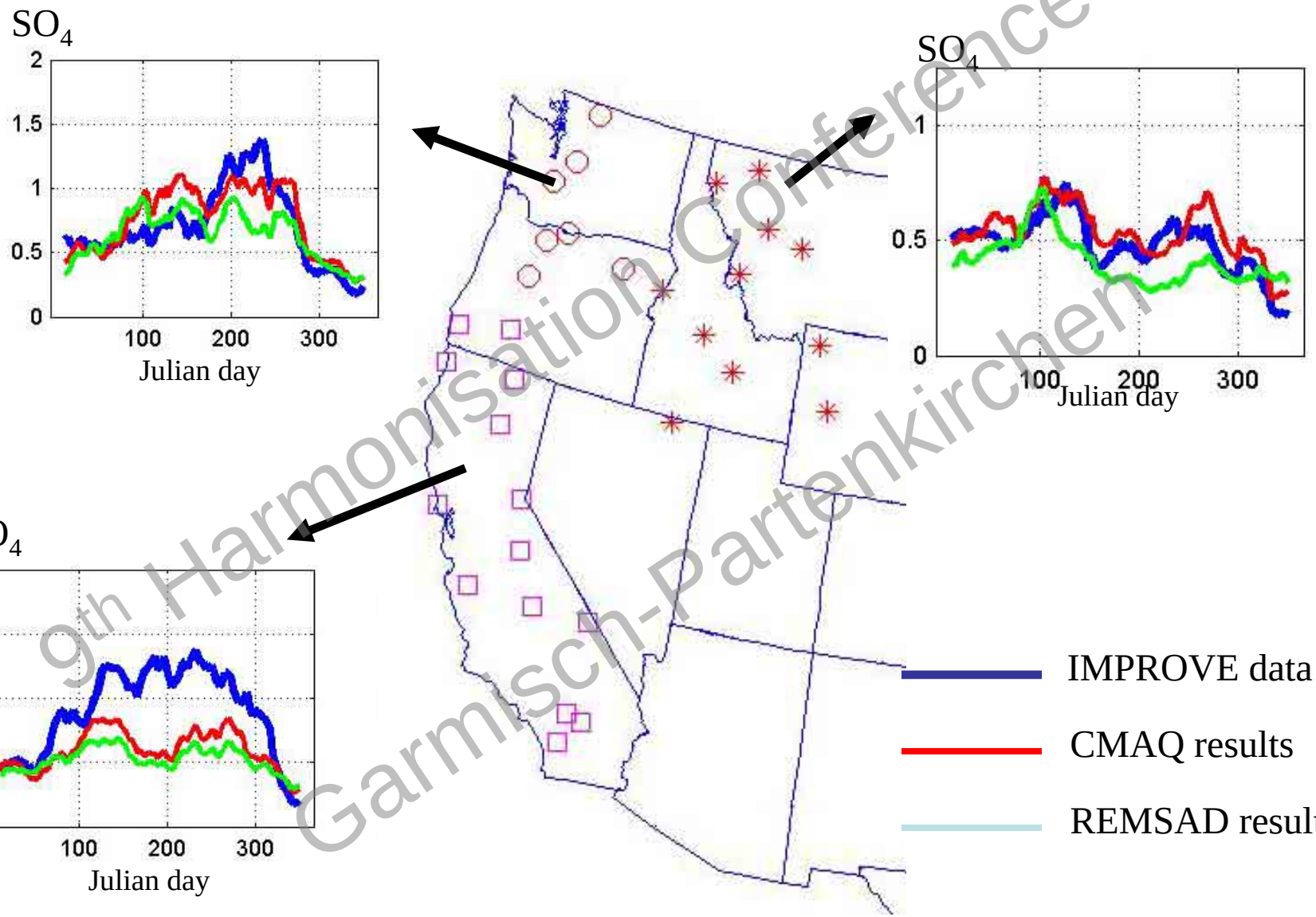
9th Harmonisation Conference
Garmisch-Partenkirchen

For next three slide we divide the country into 3 regions



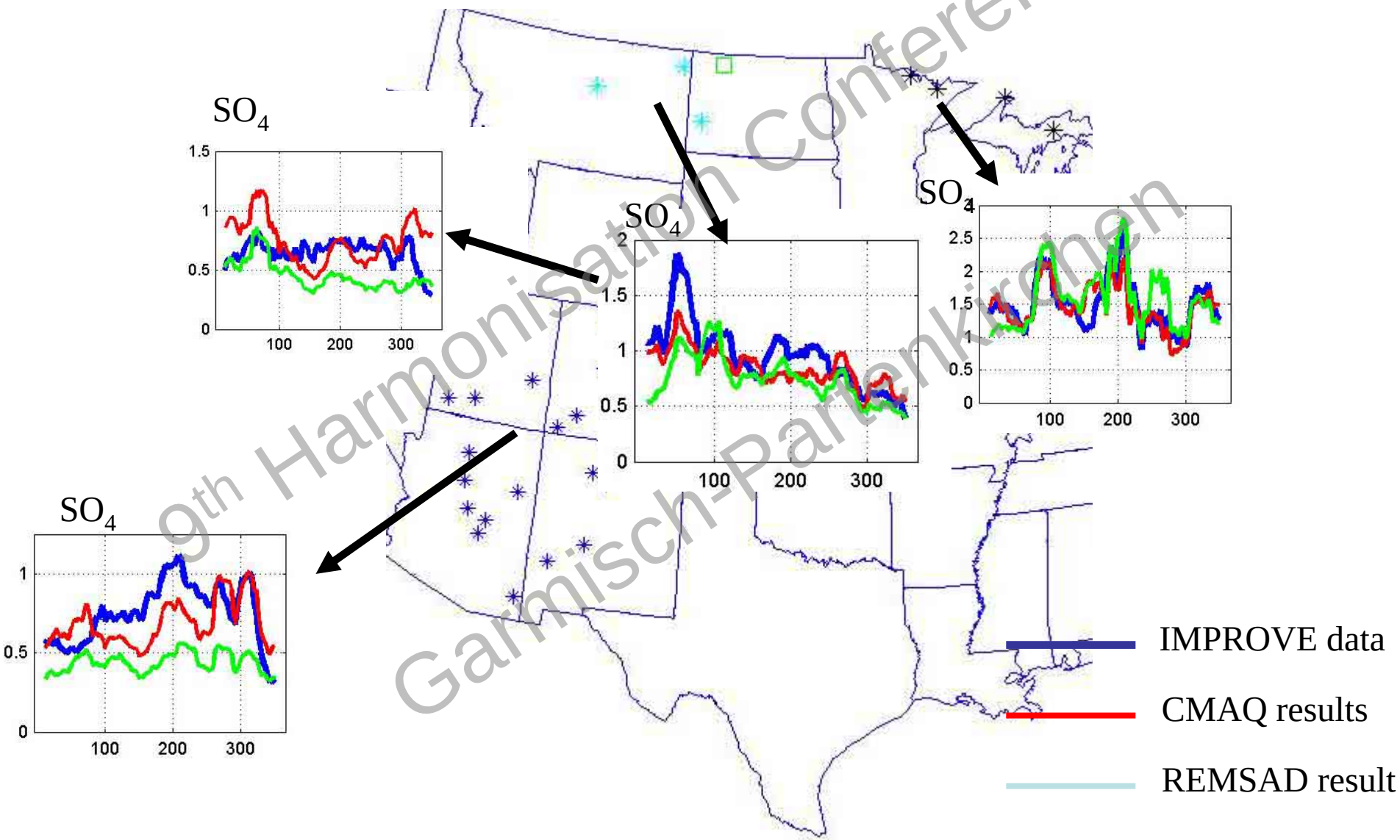
9th Harmonisation Conference Garmisch-Partenkirchen

Time series of the spatially average long term signals (observations, CMAQ or REMSAD signals) of all sites grouped in the same cluster.



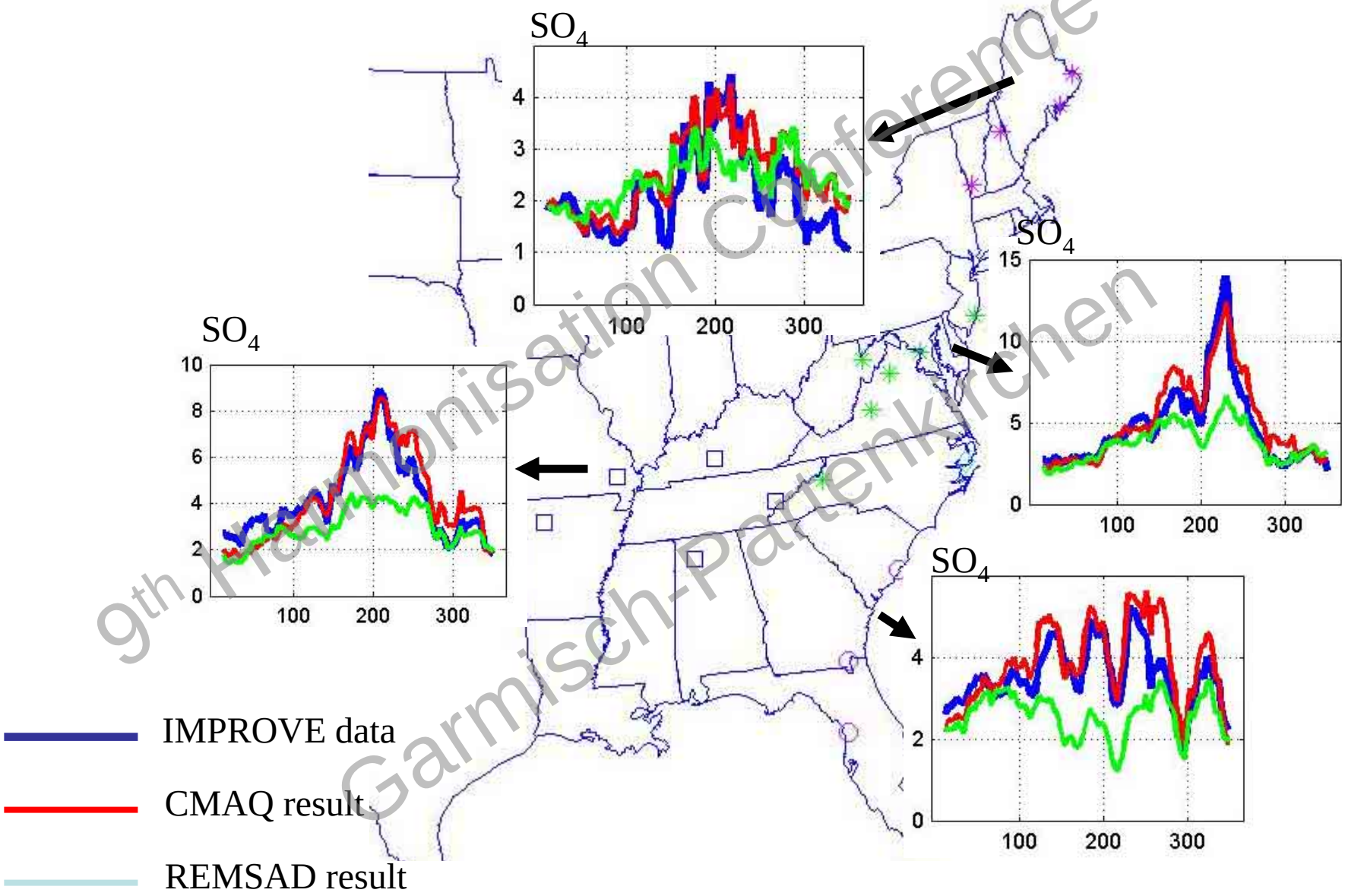
9th Harmonisation Conference Garmisch-Partenkirchen

Time series of the spatially average long term signals (observations, CMAQ or REMSAD signals) of all sites grouped in the same cluster.



9th Harmonisation Conference Garmisch-Partenkirchen

Time series of the spatially average long term signals (observations, CMAQ or REMSAD signals) of all sites grouped in the same cluster.



9th Harmonisation Conference Garmisch-Partenkirchen

One of my goals is to provide a method for objective comparisons of the observed and predicted modeling results.

There are various ways one can ask the question: which model compares best with the observations? For instance, you could perform compare scatter-plot linear regression statistics, or you could ask which model is “closest” (smallest sum of deviations squared).

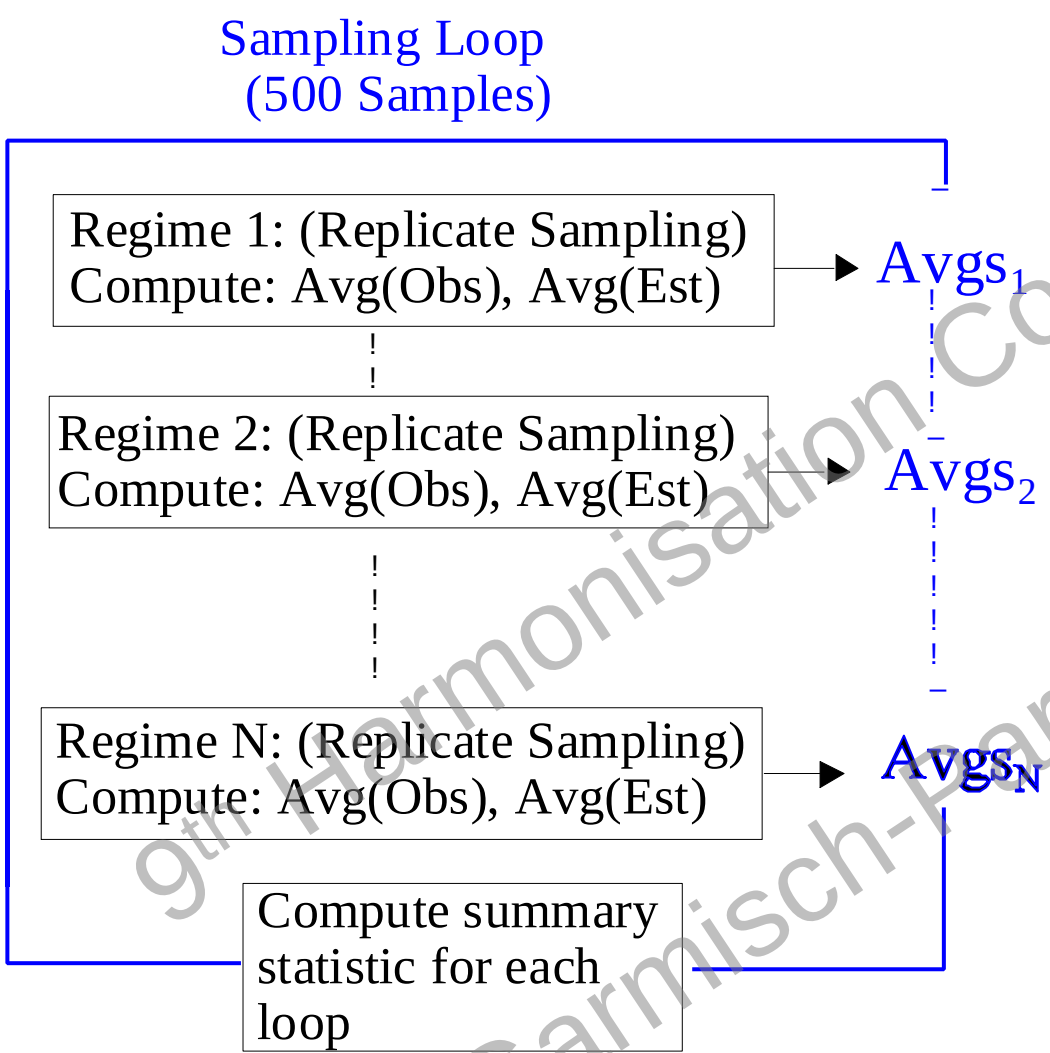
I have chosen the latter, it is simple, robust and intuitive to understand, whereas correlation analyses involve assumptions that may be violated if performed on values who themselves are averages.

9th Harmonisation Conference Garmisch-Partenkirchen

Objectively Measure Performance ASTM Standard Guide D 6589

- Group data (by time, arc, region)
- Determine average observed and modeled “patterns”
- Objectively compare observed and modeled average “patterns”
- Employ bootstrap resampling to ask whether statistics are significantly different by different models

9th Harmonisation Conference Garmisch-Partenkirchen



In this analysis "Regime" equals a 28-day lunar month. The IMPROVE data provides one 24-hr average every 3-days, providing at most 9 days in a lunar month.

When we select a day to include in the sample, we pull the observation and the corresponding CMAQ and REMSAD model estimates.

At the end of a loop, we compute our comparison statistics of the results for the 13 lunar months.

At the end of the complete sampling loop (500 samples), we compute a t-statistic to test whether the comparison statistics are significantly different.

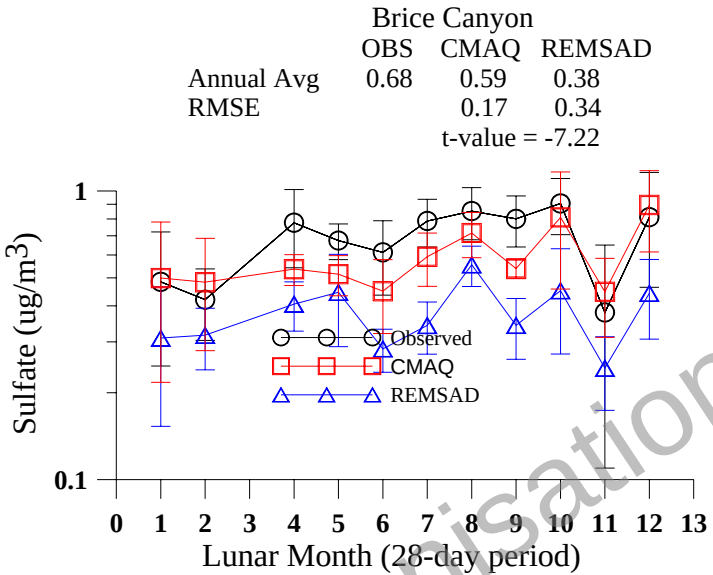
9th Harmonisation Conference Garmisch-Partenkirchen

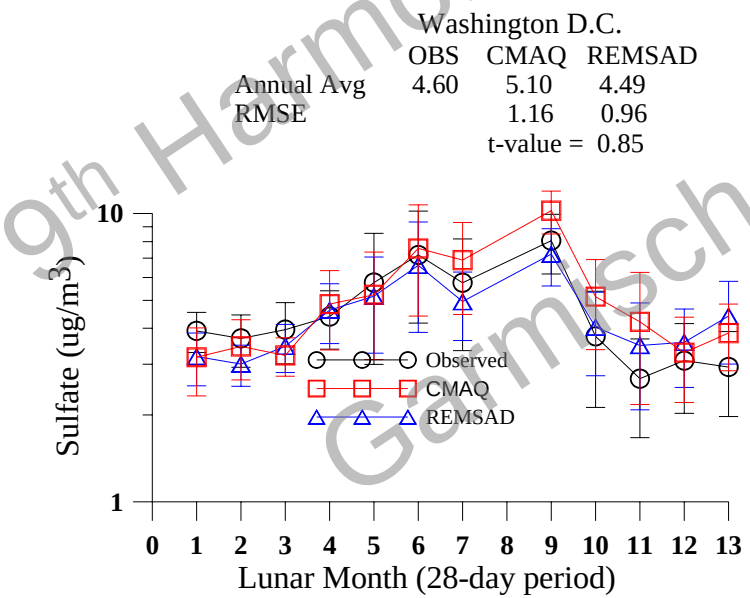
Objectively Measuring “Closeness”

Currently, I am recommending the RMSE be used in assessments where we are attempting to objectively assess the statistical significance of differences in model skill with other models, as it appears to be the most robust (not sensitive to near-zero values) of those tested and it is simple to understand.

So let us see what happens if we objectively compare CMAQ and REMSAD model performance using the RMSE and the IMPROVE data.

9th Harmonisation Conference Garmisch-Partenkirchen





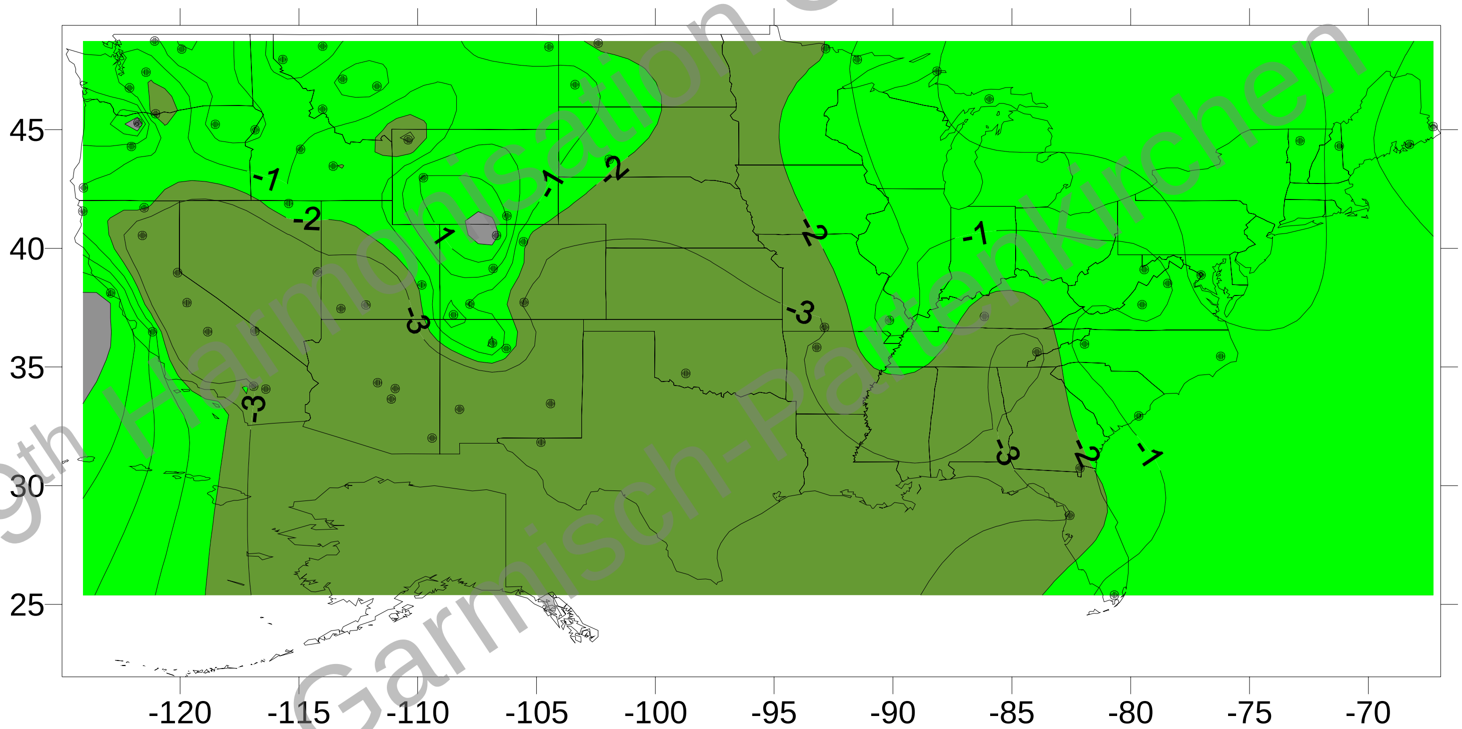
Bryce Canyon is in the southwest US. The t-value is -7.22, which means that the CMAQ results are significantly "closer" to the observed values than REMSAD's results.

Washington DC is in the mid-Atlantic states of the US east coast. Here the CMAQ and REMSAD results are too similar to judge one in closer in agreement with the observed values.

9th Harmonisation Conference Garmisch-Partenkirchen

Comparison of CMAQ and REMSAD performance in prediction of 28-day average IMPROVE sulfate concentration values for 2001. Contours are student-t test values derived from 500 bootstrap samples

(<0 means CMAQ's RMSE is less than REMSAD's RMSE)



9th Harmonisation Conference Garmisch-Partenkirchen

Light Green: Both have similar skill. Dark Green: CMAQ has significantly better skill than REMSAD

Summary

- The philosophy articulated in the ASTM D 6589 Model Evaluation Guide can be applied successfully to assess regional scale model performance.
- RMSE is robust and provides an adequate means for assessing which model is “closest” in tracking the monthly variation in concentration.
- Bootstrap resampling provides a means for objectively assessing whether differences in model performance are meaningful.
- Recommend objective methods like this be used to assess differences in performance between versions of CMAQ and between CMAQ and other models.

9th Harmonisation Conference Garmisch Partenkirchen