# Statistics - Definitions and Issues
# Deriving "Unbiased Symmetric" Metrics

**Shaocai Yu*, Brian Eder*++, Robin Dennis*++,**
**Shao-Hang Chu**, Stephen Schwartz*****

*Atmospheric Sciences Modeling Division, NERL
** Office of Air Quality Planning and Standards
U.S. EPA, RTP, NC 27711.

***Brookhaven National Laboratory, Upton, NY 11973
++ On assignment from Air Resources Laboratory, NOAA

# Introduction

Introduction

- What are problems with commonly used metrics for model evaluation?

- Operational evaluation (EPA, 1991; Russell and Dennis, 2000 )
  - Determine a model's degree of acceptability and
  - Usefulness for specific task

- Commonly used metrics (EPA, 1991 )

❶ Difference between model and obs
  - Mean Bias ($B_{MB}$)
  - Mean Absolute Gross Error ($E_{MAGE}$), RMSE

❷ Relative difference (normalized by Obs)
  - Mean Normalized Bias ($B_{MNB}$)
  - Mean Normalized Gross Error ($E_{MNGE}$)

See Table 1 for other metrics

Table 1. Commonly Used Metrics of Model Performance

| Metrics | Mathematical Expression | Range |
|---|---|---|
| **(1) Difference** | | |
| Mean Bias | $B_{MB} = \frac{1}{N}\sum_{i=1}^{N}(M_i - O_i) = \overline{M} - \overline{O}$ | $-\overline{O}$ to $+\infty$ |
| Mean Absolute Gross Error | $E_{MAGE} = \frac{1}{N}\sum_{i=1}^{N}\lvert M_i - O_i\rvert$ | 0 to $+\infty$ |
| Root Mean Square Error | $E_{RMSE} = [\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2]^{\frac{1}{2}}$ | 0 to $+\infty$ |
| **(2) Relative difference** | | |
| Mean Normalized Bias | $B_{MNB} = \frac{1}{N}\sum_{i=1}^{N}(\frac{M_i - O_i}{O_i}) \times 100\% = (\frac{1}{N}\sum\frac{M_i}{O_i} - 1) \times 100\%$ | -100% to $+\infty$% |
| Mean Normalized Gross Error | $E_{MNGE} = \frac{1}{N}\sum_{i=1}^{N}(\frac{\lvert M_i - O_i\rvert}{O_i}) \times 100\%$ | 0% to $+\infty$% |
| Normalized Mean Bias | $B_{NMB} = \frac{\sum_{i=1}^{N}(M_i - O_i)}{\sum_{i=1}^{N}O_i} \times 100\% = (\frac{\overline{M}}{\overline{O}} - 1) \times 100\%$ | -100% to $+\infty$% |
| Normalized Mean Error | $E_{NME} = \frac{\sum_{i=1}^{N}\lvert M_i - O_i\rvert}{\sum_{i=1}^{N}O_i} \times 100\% = \frac{E_{MAGE}}{\overline{O}} \times 100\%$ | 0% to $+\infty$% |
| Fractional Bias | $B_{FB} = \frac{1}{N}\sum_{i=1}^{N}\frac{(M_i - O_i)}{\frac{(M_i + O_i)}{2}}$ | -2 to +2 |
| Fractional Gross Error | $E_{FGE} = \frac{1}{N}\sum_{i=1}^{N}\frac{\lvert M_i - O_i\rvert}{\frac{(M_i + O_i)}{2}}$ | 0 to 2 |

# Introduction

- What are problems with commonly used metrics ? (Continued)

- Two problems with metrics in Table 1
  1. Asymmetry for underpredictioin and overprediction
     - Mean Bias: - $\overline{O}$ to $+\infty$
     - Mean Normalized Bias, NMB: -100% to $+\infty$%

  2. Biased because of small numbers in the denominator
     - Mean Normalized Bias:

$$B_{MNB} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{M_i - O_i}{O_i}\right)\times 100\% = \frac{1}{N}\left(\sum\frac{M_i}{O_i} - 1\right)\times 100\%$$

- Problem with Fractional Bias
  1. Against both Obs and Model
  2. Seriously compressed beyond ±1 to ±2
  3. Unclear meaning: 0.60 ?

$$B_{FB} = \frac{1}{N}\sum_{i=1}^{N}\frac{(M_i - O_i)}{\frac{(M_i + O_i)}{2}}$$

# *Objective*

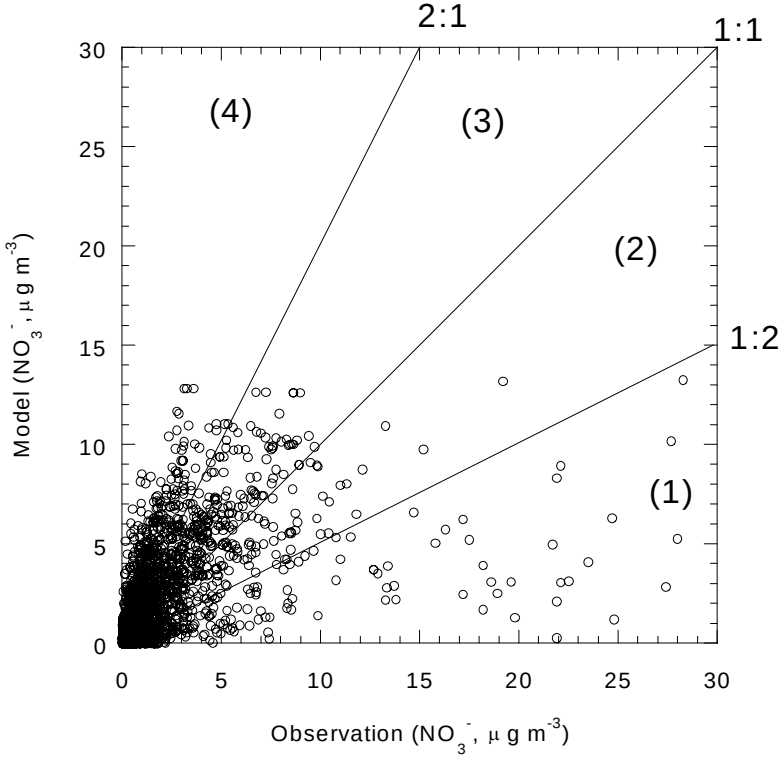
❶ **Propose new unbiased symmetric metrics on the basis of concept of factor**

❷ **Test new metrics and other metrics, and apply the new metrics in the CMAQ evaluation**

# New Metrics Description

- **Normalized mean Bias Factor ($B_{NMBF}$), Normalized mean error factor ($E_{NMEF}$)**

- **Concept of Factor (symmetry)**
  - **For Model>Obs (overprediction):**
  - **For Model<Obs (underprediction):**




$$Factor = \frac{Model}{Obs}$$

$$Factor = \frac{Obs}{Model}$$

Symmetry: overprediction and underprediction are treated proportionately

# New Metrics Description (Continued)

- $B_{NMBF}$ and $E_{NMEF}$

- **Normalized Mean Bias Factor ($B_{NMBF}$)**
  - **For $\overline{M} \geq \overline{O}$ (overprediction):**

$$B_{NMBF} = \left(\frac{\sum_{i=1}^{N} M_i}{\sum_{i=1}^{N} O_i} - 1\right) = \left(\frac{\overline{M}}{\overline{O}} - 1\right)$$

  - **For $\overline{M} < \overline{O}$ (underprediction):**

$$B_{NMBF} = \left(1 - \frac{\sum_{i=1}^{N} O_i}{\sum_{i=1}^{N} M_i}\right) = \left(1 - \frac{\overline{O}}{\overline{M}}\right)$$

$B_{NMBF}$: symmetry ,
(Range) $-\infty$ to $+\infty$,
$+$ is overprediction
$-$ is underprediction

$$B_{NMBF} = \frac{\sum_{i=1}^{N} M_i - \sum_{i=1}^{N} O_i}{\left|\sum_{i=1}^{N} M_i - \sum_{i=1}^{N} O_i\right|} \left[\exp\left(\left|\ln \frac{\sum_{i=1}^{N} M_i}{\sum_{i=1}^{N} O_i}\right|\right) - 1\right]$$

# New Metrics Description (Continued)

- $B_{NMBF}$ and $E_{NMEF}$

- **Normalized Mean Error Factor ($E_{NMEF}$)**
  - **For $\overline{M} \geq \overline{O}$ (overprediction):**

$$E_{NMEF} = \frac{\sum_{i=1}^{N} |M_i - O_i|}{\sum_{i=1}^{N} O_i} = \frac{E_{MAGE}}{\overline{O}}$$

  - **For $\overline{M} < \overline{O}$ (underprediction):**

$$E_{NMEF} = \frac{\sum_{i=1}^{N} |M_i - O_i|}{\sum_{i=1}^{N} M_i} = \frac{E_{MAGE}}{\overline{M}}$$

$E_{NMEF}$: 0 to $+\infty$

$$E_{NMEF} = \frac{\sum_{i=1}^{N} |M_i - O_i|}{(\sum_{i=1}^{N} O_i)^{[\frac{\sum_{i=1}^{N} M_i - \sum_{i=1}^{N} O_i}{|\sum_{i=1}^{N} M_i - \sum_{i=1}^{N} O_i|} + 1]/2} \quad (\sum_{i=1}^{N} M_i)^{[1 - \frac{\sum_{i=1}^{N} M_i - \sum_{i=1}^{N} O_i}{|\sum_{i=1}^{N} M_i - \sum_{i=1}^{N} O_i|}]/2}}$$

# New Metrics Description (Continued)

- **Normalized Mean Bias Factor :**

Unbiased: avoid undue influence of small numbers in denominator

- **For $\overline{M} \geq \overline{O}$ (overprediction):**

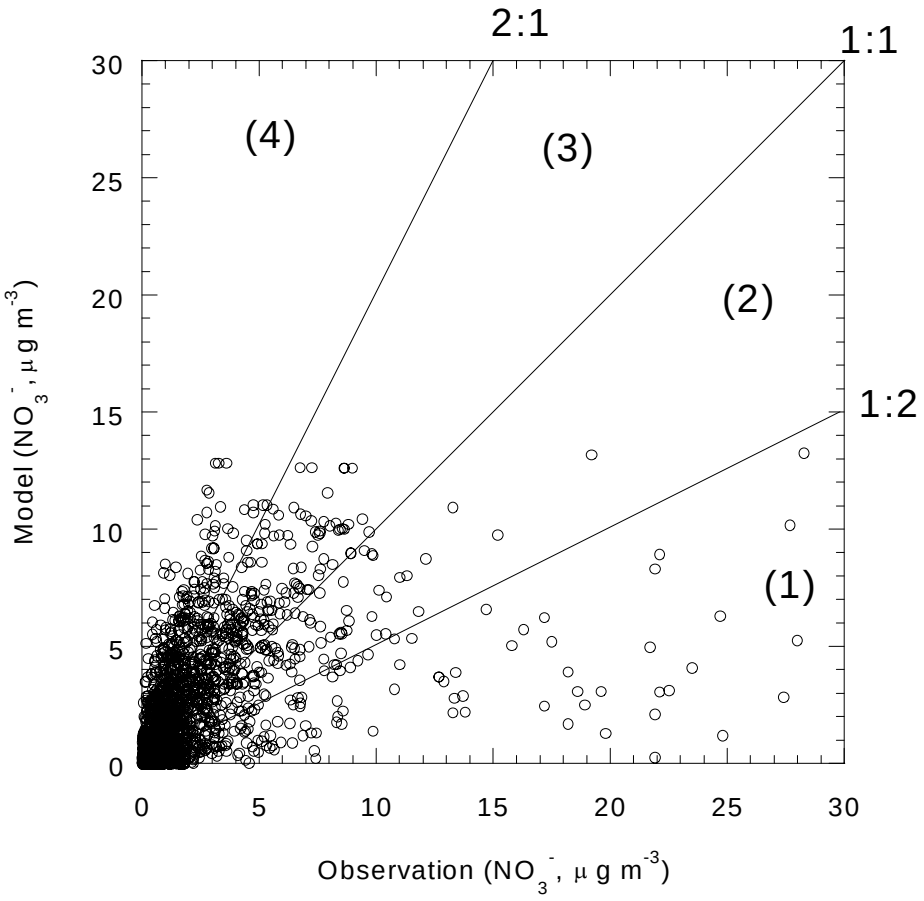
$$B_{NMBF} = \frac{\sum_{i=1}^{N} M_i}{\sum_{i=1}^{N} O_i} - 1 = \frac{\sum_{i=1}^{N} (M_i - O_i)}{\sum_{i=1}^{N} O_i} = \sum_{i=1}^{N} \left[ \frac{O_i}{\sum_{i=1}^{N} O_i} \frac{(M_i - O_i)}{O_i} \right]$$

- **For $\overline{M} < \overline{O}$ (underprediction):**

$$B_{NMBF} = 1 - \frac{\sum_{i=1}^{N} O_i}{\sum_{i=1}^{N} M_i} = \frac{\sum_{i=1}^{N} (M_i - O_i)}{\sum_{i=1}^{N} M_i} = \sum_{i=1}^{N} \left[ \frac{M_i}{\sum_{i=1}^{N} M_i} \frac{(M_i - O_i)}{M_i} \right]$$

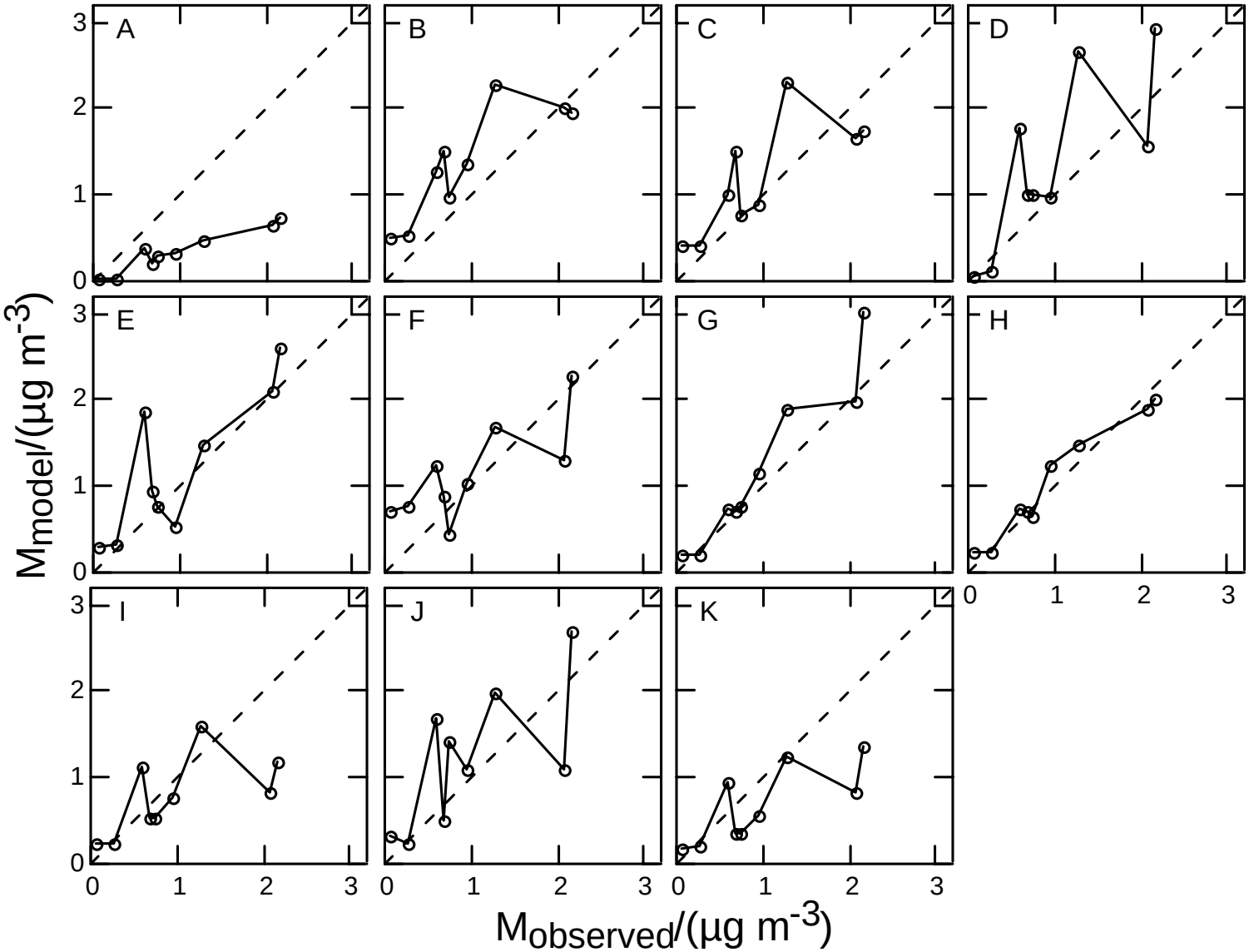
$B_{NMBF}$: result of sum of indiv. factor bias with obs (or model) conc. as a weighting function

# ❑Test of Metrics



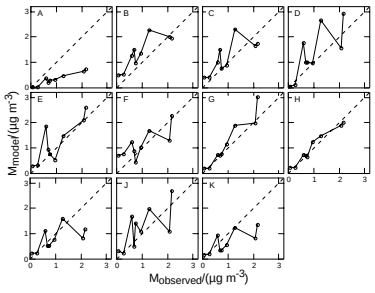

| Combination* | 1 | 2 | 3 | 4 | 1+2+3+4 |
|---|---|---|---|---|---|
| $\overline{O}$ | 1.92 | 2.15 | 2.11 | 0.88 | 1.72 |
| $\overline{M}$ | 0.42 | 1.58 | 2.94 | 2.88 | 1.88 |
| $N$ | 903 | 450 | 663 | 755 | 2771 |
| $r$ | 0.79 | 0.97 | 0.97 | 0.90 | 0.51 |
| **Difference** | | | | | |
| $B_{MB}$ | -1.50 | -0.57 | 0.83 | 1.99 | 0.16 |
| $E_{MAGE}$ | 1.50 | 0.57 | 0.83 | 1.99 | 1.32 |
| $E_{RMSE}$ | 4.25 | 1.07 | 1.29 | 2.70 | 2.91 |
| **Relative Difference** | | | | | |
| $B_{MNB}$ | -0.82 | -0.27 | 0.43 | 4.27 | 0.96 |
| $E_{MNGE}$ | 0.82 | 0.27 | 0.43 | 4.27 | 1.58 |
| $B_{NMB}$ | -0.78 | -0.26 | 0.39 | 2.25 | 0.09 |
| $E_{NME}$ | 0.78 | 0.26 | 0.39 | 2.25 | 0.77 |
| $B_{FB}$ | -1.43 | -0.33 | 0.33 | 1.12 | -0.13 |
| $E_{FGE}$ | 1.43 | 0.33 | 0.33 | 1.12 | 0.90 |
| $B_{NMBF}$ | **-3.58** | **-0.36** | **0.39** | **2.25** | **0.09** |
| $E_{NMEF}$ | **3.58** | **0.36** | **0.39** | **2.25** | **0.77** |

# ❑Test of Metrics (Continued) :11 models from IPCC (2001) (nss-$SO_4^{2-}$)



# ❑Test of Metrics (Continued) :11 models for nss-$SO_4^{2-}$

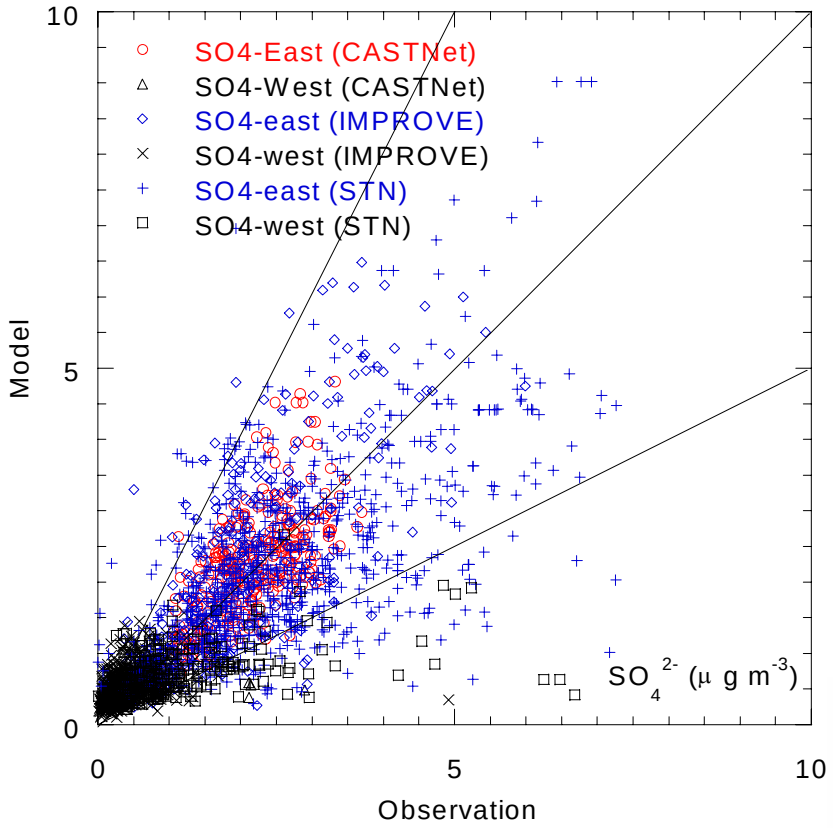
| Models | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\overline{O}$ | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| $\overline{M}$ | 0.35 | 1.37 | 1.19 | 1.34 | 1.22 | 1.16 | 1.19 | 1.02 | 0.79 | 1.23 | 0.67 | 0.00 | 1.96 | +∞ |
| ***N*** | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 |
| ***r*** | 0.959 | 0.840 | 0.737 | 0.777 | 0.839 | 0.769 | 0.953 | 0.977 | 0.609 | 0.692 | 0.767 | 0.00 | 1.00 | 0.00 |
| **Difference** | | | | | | | | | | | | | | |
| $B_{MB}$ | -0.63 | 0.40 | 0.21 | 0.37 | 0.24 | 0.18 | 0.21 | 0.05 | -0.19 | 0.25 | -0.31 | -0.98 | +0.98 | +∞ |
| $E_{MAGE}$ | 0.63 | 0.46 | 0.42 | 0.52 | 0.34 | 0.42 | 0.24 | 0.14 | 0.42 | 0.52 | 0.41 | 0.98 | +0.98 | +∞ |
| $E_{RMSE}$ | 0.79 | 0.55 | 0.52 | 0.70 | 0.49 | 0.48 | 0.37 | 0.16 | 0.58 | 0.63 | 0.55 | 0.98 | +0.98 | +∞ |
| **Relative Difference** | | | | | | | | | | | | | | |
| $B_{MNB}$ | -0.65 | 1.23 | 0.91 | 0.38 | 0.70 | 1.40 | 0.34 | 0.33 | 0.19 | 0.75 | -0.06 | -1.00 | +1.00 | +∞ |
| $E_{MNGE}$ | 0.65 | 1.26 | 1.01 | 0.60 | 0.80 | 1.58 | 0.39 | 0.39 | 0.59 | 0.94 | 0.52 | 1.00 | +1.00 | +∞ |
| $B_{NMB}$ | -0.64 | 0.41 | 0.22 | 0.38 | 0.25 | 0.18 | 0.21 | 0.05 | -0.20 | 0.26 | -0.32 | -1.00 | +1.00 | +∞ |
| $E_{NME}$ | 0.64 | 0.47 | 0.43 | 0.53 | 0.34 | 0.43 | 0.25 | 0.15 | 0.44 | 0.53 | 0.42 | 1.00 | +1.00 | +∞ |
| $B_{FB}$ | -1.00 | 0.53 | 0.37 | 0.16 | 0.30 | 0.35 | 0.22 | 0.16 | -0.04 | 0.30 | -0.24 | -2.00 | +0.67 | +∞ |
| $E_{FGE}$ | 1.00 | 0.56 | 0.48 | 0.45 | 0.43 | 0.56 | 0.27 | 0.24 | 0.47 | 0.53 | 0.53 | 2.00 | +0.67 | +∞ |
| $B_{NMBF}$ | -1.80 | 0.41 | 0.22 | 0.38 | 0.25 | 0.18 | 0.21 | 0.05 | -0.24 | 0.26 | -0.46 | -∞ | +1.00 | +∞ |
| $E_{NMEF}$ | 1.80 | 0.47 | 0.43 | 0.53 | 0.34 | 0.43 | 0.25 | 0.15 | 0.54 | 0.53 | 0.61 | +∞ | +1.00 | +∞ |



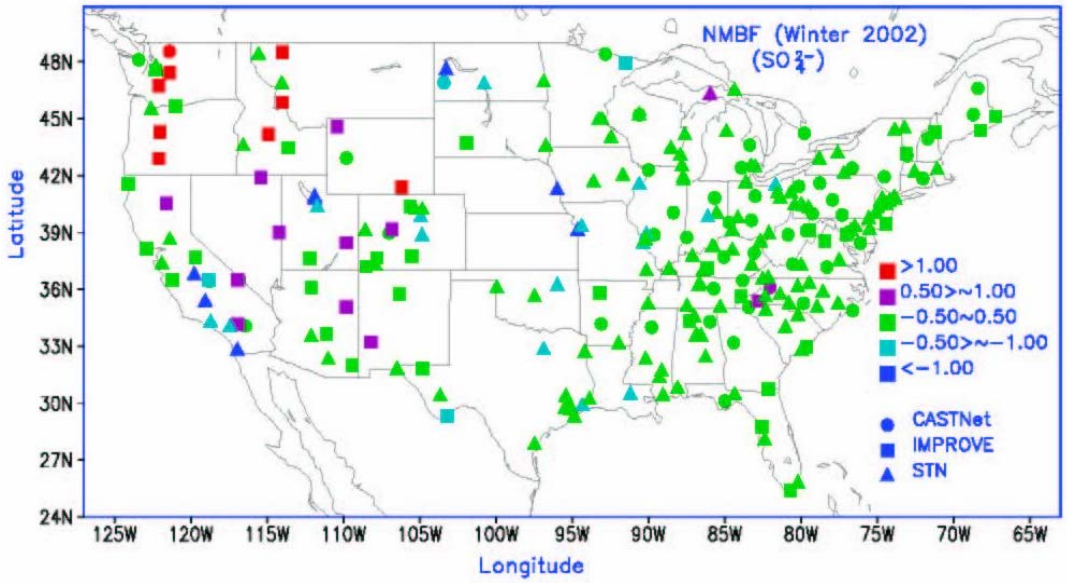

- Model H: best; Model A: worst
- Models E, G, H: acceptable

  If criteria: ±25% ($B_{NMBF}$), 35% ($E_{NMEF}$)

# ❑Application of new Metrics for CMAQ evaluation

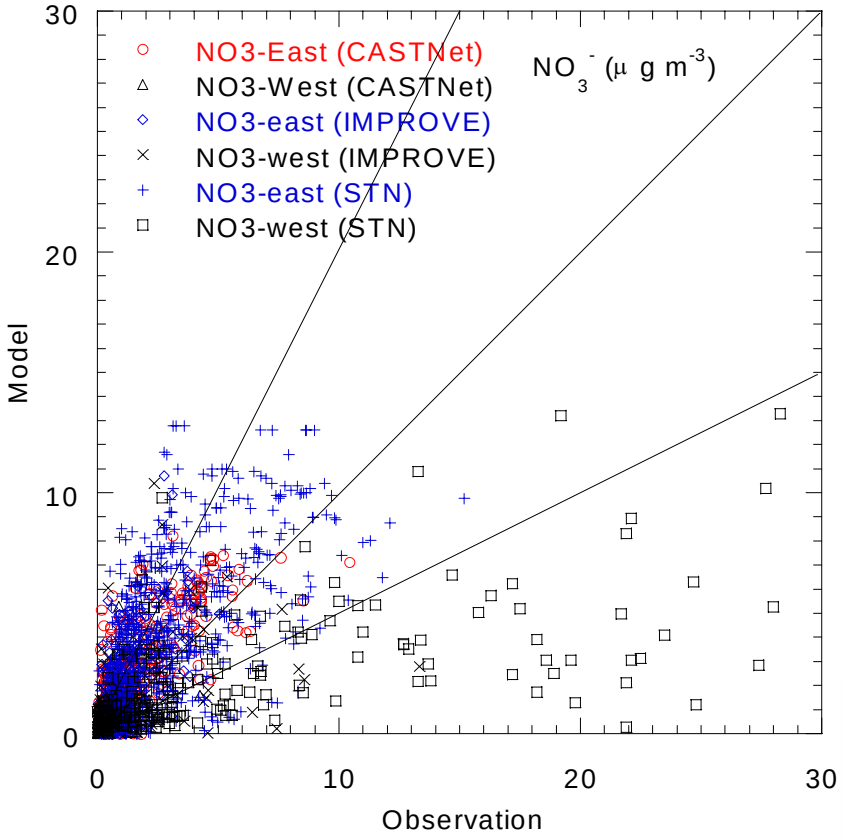



| Network | CASTNet | IMPROVE | STN |
|---|---|---|---|
| $SO_4^{2-}$ ($\mu g\ m^{-3}$) | | | |
| Mean Model ($\overline{M}$) | 1.76 | 1.13 | 1.88 |
| Mean OBS ($\overline{O}$) | 1.71 | 0.93 | 2.23 |
| *N* | 413 | 729 | 1149 |
| *r* | 0.841 | 0.860 | 0.670 |
| $B_{MB}$ | 0.06 | 0.194 | -0.344 |
| $E_{MAGE}$ | 0.41 | 0.427 | 0.793 |
| $B_{NMBF}$ | 0.03 | 0.22 | -0.19 |
| $E_{NMEF}$ | 0.24 | 0.46 | 0.42 |

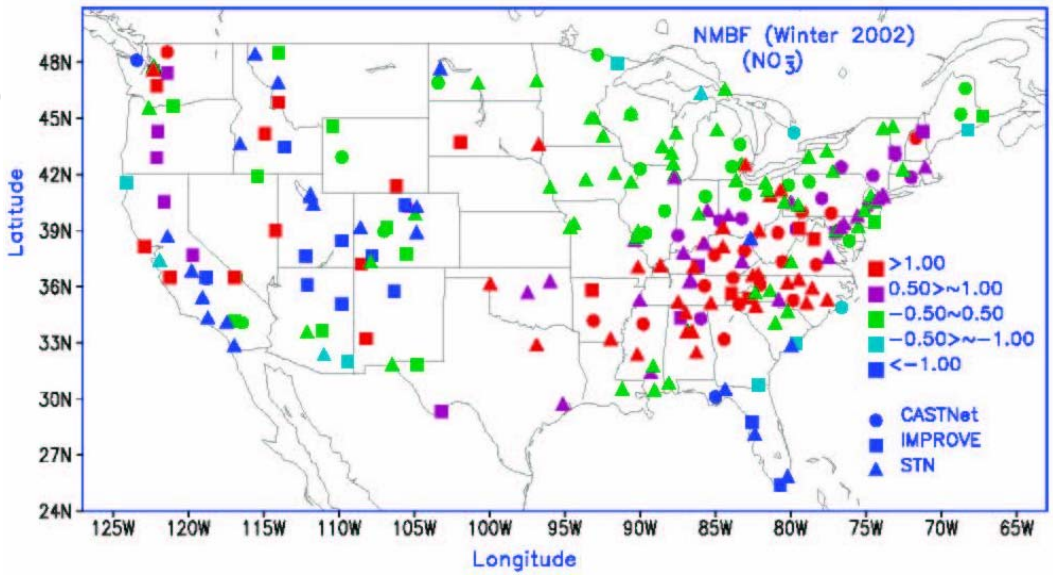



Jan. 8 to Feb. 18, 2002

# ❑Application of new Metrics (Continued)




| Network | CASTNet | IMPROVE | STN |
|---|---|---|---|
| $NO_3^-$ (μg $m^{-3}$) | | | |
| Mean Model ($\overline{M}$) | 2.19 | 0.904 | 3.38 |
| Mean OBS ($\overline{O}$) | 1.38 | 0.683 | 3.35 |
| *N* | 4 15 | 689 | 1044 |
| *r* | 0.758 | 0.540 | 0.360 |
| $B_{MB}$ | 0.811 | 0.221 | 0.033 |
| $E_{MAGE}$ | 1.11 | 0.676 | 2.428 |
| $B_{NMBF}$ | 0.59 | 0.32 | 0.01 |
| $E_{NMEF}$ | 0.80 | 0.99 | 0.72 |




Jan. 8 to Feb. 18, 2002

# Conclusions

- Normalized mean bias factor and normalized mean error factor are proposed to quantify the relative departure between model and obs .
- The newly proposed metrics are:
  - Symmetric: overprediction and underprediction are treated proportionately
  - Unbiased: avoid undue influence of small numbers in the denominator
- Tests show that the newly proposed metrics are useful, their meanings are clear and easy to explain.
- To represent the whole performance of the model:
  - Mean (model, obs), r, Number, difference ($B_{MB}$, $E_{MAGE}$), relative difference ($B_{NMBF}$, $E_{NMEF}$)
    - Values of relative differences depend on the units of model prediction and obs !!!!