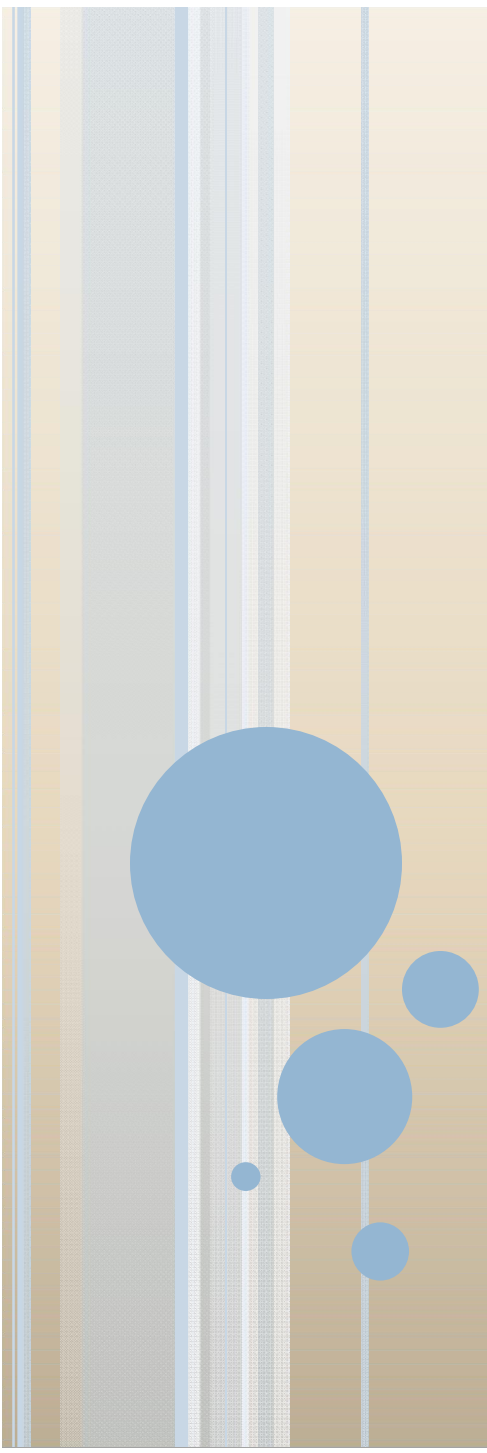




IDENTIFYING SOURCES OF UNCERTAINTY USING COVARIANCE ANALYSIS

Nicole Hyslop & Warren White

Crocker Nuclear Lab

University of California at Davis

IMPROVE Steering Committee Meeting

October 22, 2010



STN & IMPROVE NETWORKS

- Two networks collect PM_{2.5} samples and perform similar analyses
 - IMPROVE: ~170 sites operated by UCD
 - STN: ~50 sites operated by EPA
- Both networks operate collocated (duplicate) samplers at 6 to 7 sites
 - Collocated data used in past to estimate measurement errors for individual species
 - Extend error analysis to multivariate context

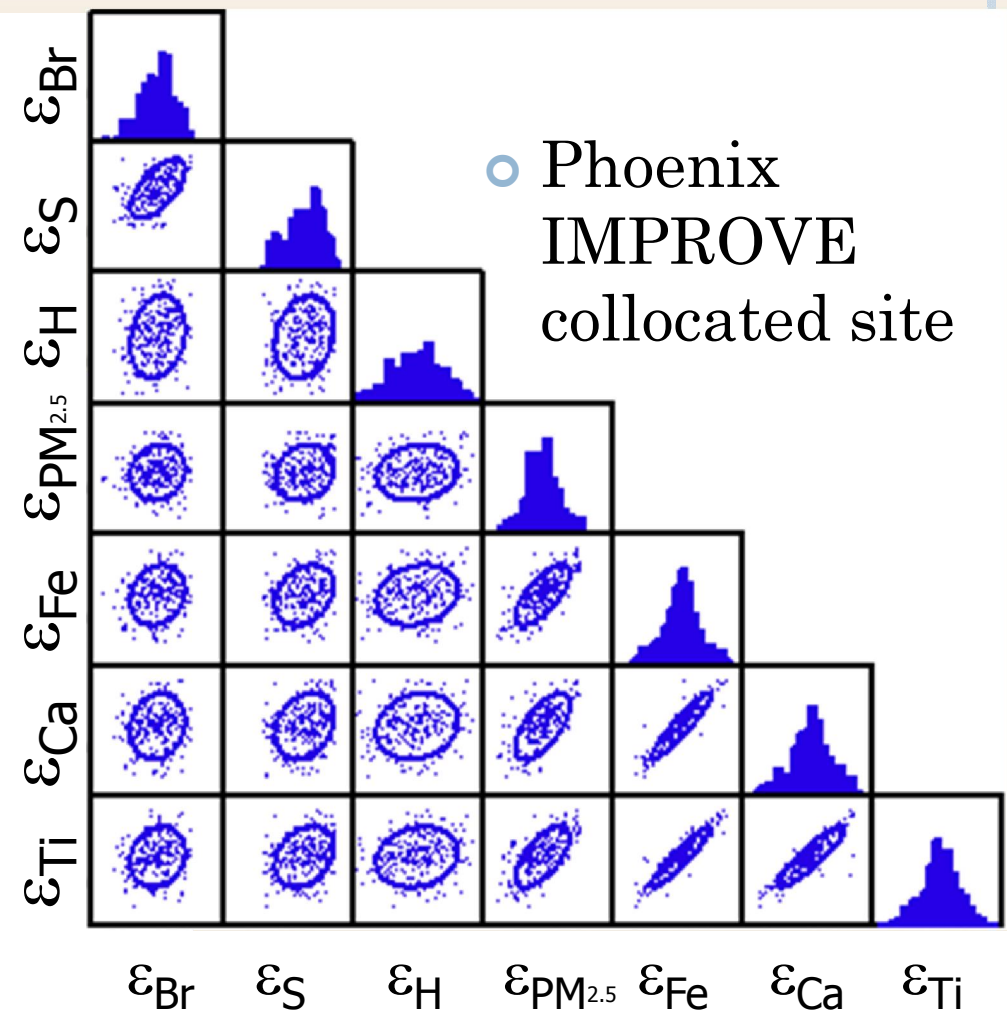
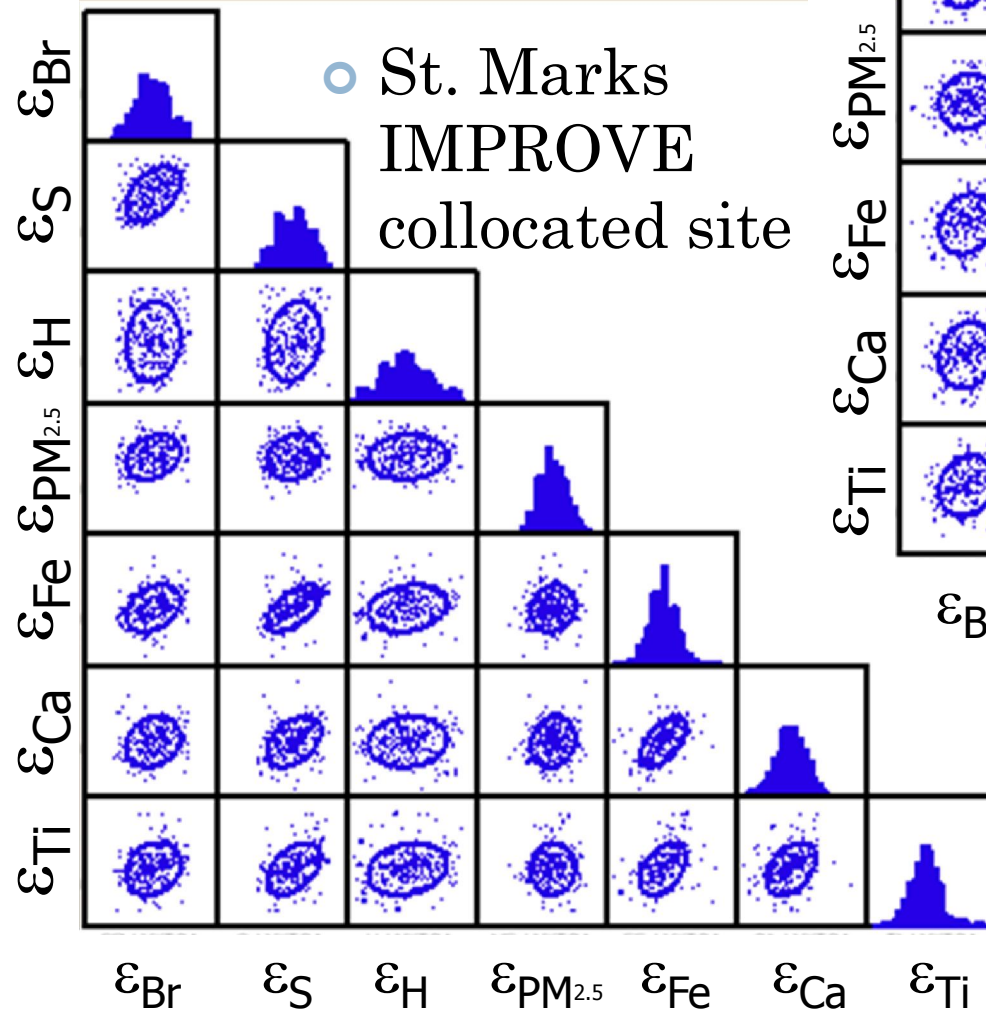


INDEPENDENCE OF ERRORS

- Analysts assume errors are independent among analytes
 - Independence means correlation/covariance is zero
 - Assumption is often not completely valid
- Prior analyses indicate some severe violations of this assumption may exist
- Use log ratios to estimate errors, $\ln\left(\frac{[c_{x,2}]}{[c_{x,1}]}\right)$



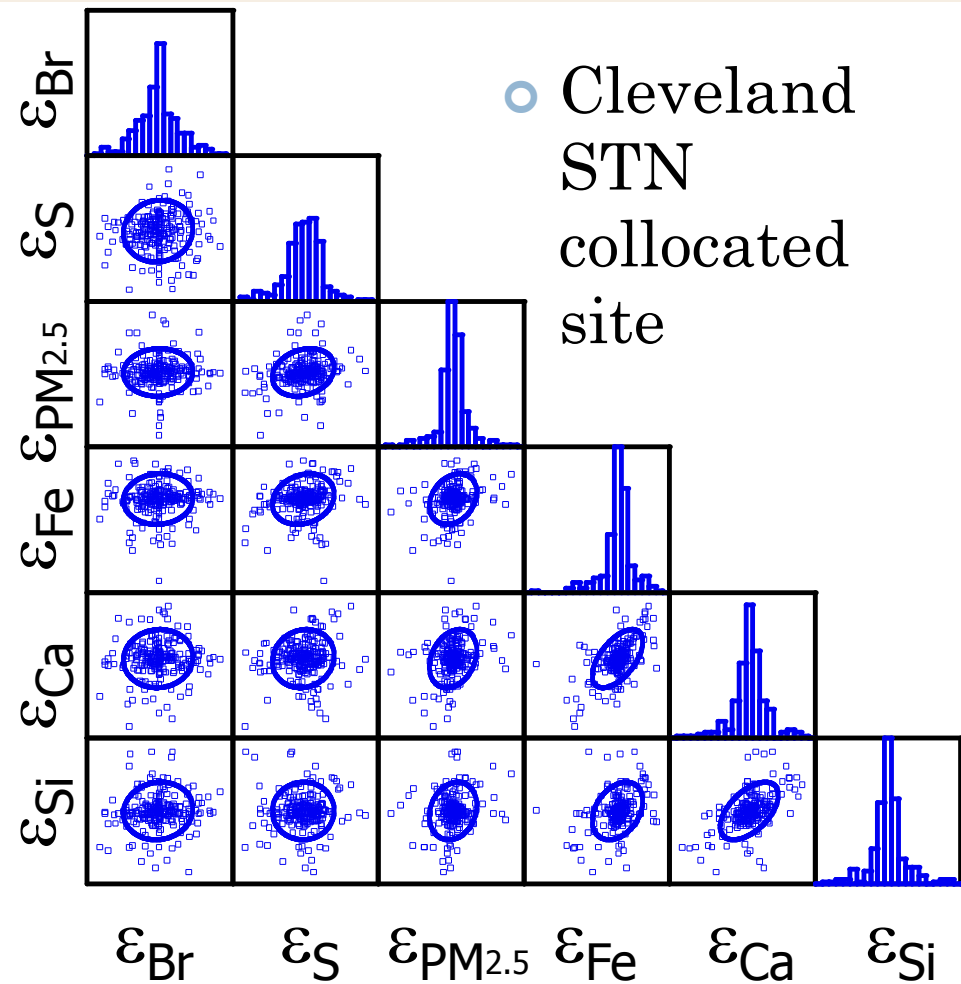
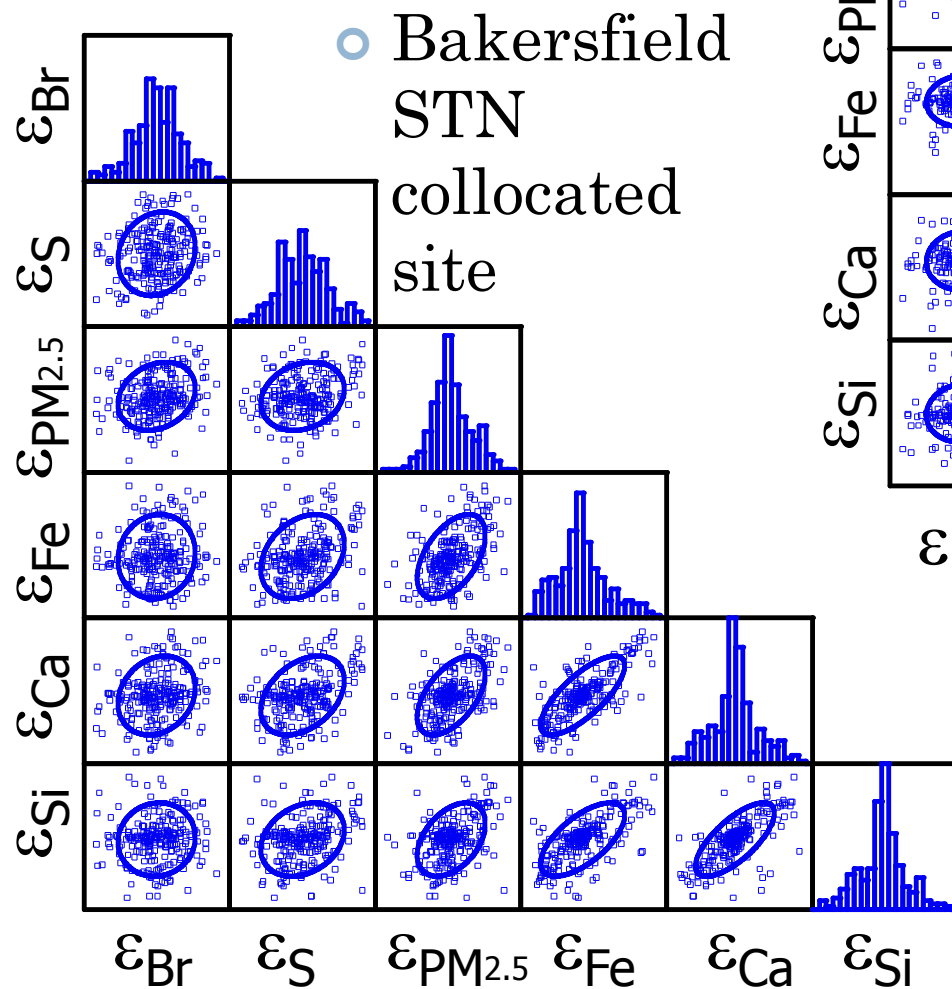
SCATTERPLOT MATRICES



$$\varepsilon_x \equiv \ln \left(\frac{[c_{x,2}]}{[c_{x,1}]} \right)$$



SCATTERPLOT MATRICES



$$\epsilon_x \equiv \ln \left(\frac{[c_{x,2}]_t}{[c_{x,1}]_t} \right)$$

MODELING INDIVIDUAL ERRORS

$$\frac{[c_x]_{2.5}}{[C_x]_{2.5}} = \frac{V}{v} \frac{s_x}{S_x} \frac{l_x}{L_x}$$
$$= \mathcal{E}_{vol} \mathcal{E}_{size} \mathcal{E}_{an,x}$$

- Assumptions:

- multiplicative (slope) errors dominate
 - Additive (offset) errors are insignificant
- 3 sources of error considered
 - Volume error: V, v (m³)
 - Size characterization error : S_x, s_x (μg)
 - Chemical analysis error : L_x, l_x (μg/cm²)



COVARIANCE MATRIX MODEL

$$\begin{array}{c}
 \varepsilon_a \\
 \varepsilon_b \\
 \varepsilon_c
 \end{array}
 \begin{array}{c}
 \varepsilon_a \quad \varepsilon_b \quad \varepsilon_c \\
 \left[\begin{array}{ccc}
 \text{var}(\varepsilon_a) & & \\
 \text{cov}(\varepsilon_a, \varepsilon_b) & \text{var}(\varepsilon_b) & \\
 \text{cov}(\varepsilon_a, \varepsilon_c) & \text{cov}(\varepsilon_b, \varepsilon_c) & \text{var}(\varepsilon_c)
 \end{array} \right]
 \end{array}$$

Note
 $\text{cov}(\varepsilon_a, \varepsilon_b) = \text{cov}(\varepsilon_b, \varepsilon_a)$

$$\begin{array}{c}
 \text{HAZE} \\
 \varepsilon_a \\
 \text{DUST} \\
 \varepsilon_b \quad \varepsilon_c
 \end{array}
 \begin{array}{c}
 \varepsilon_a \quad \varepsilon_b \quad \varepsilon_c \\
 \left[\begin{array}{ccc}
 \text{var}(\varepsilon_{an,a}) + \text{var}(\varepsilon_{vol}) & & \\
 \text{var}(\varepsilon_{vol}) & \text{var}(\varepsilon_{an,b}) + \text{var}(\varepsilon_{vol}) + \text{var}(\varepsilon_{size}) & \\
 \text{var}(\varepsilon_{vol}) & \text{var}(\varepsilon_{vol}) + \text{var}(\varepsilon_{size}) & \text{var}(\varepsilon_{an,c}) + \text{var}(\varepsilon_{vol}) + \text{var}(\varepsilon_{size})
 \end{array} \right]
 \end{array}$$

INDIVIDUAL AND SHARED ERROR ESTIMATES

- St. Marks IMPROVE collocated site

	Br	S	H	PM_{2.5}	Fe	Ca	Si
Br	6%						
S	3%	3%					
H	3%	3%	7%				
PM_{2.5}	2%	1%	2%	3%			
Fe	3%	3%	3%	2%	6%		
Ca	3%	3%	2%	1%	5%	7%	
Si	4%	4%	4%	2%	6%	5%	11%

- PM_{2.5} lowest shared errors
- “Dust” elements have higher errors



INDIVIDUAL AND SHARED ERROR ESTIMATES

- Phoenix IMPROVE collocated site

	Br	S	H	PM_{2.5}	Fe	Ca	Si
Br	5%						
S	4%	4%					
H	3%	4%	8%				
PM_{2.5}	3%	3%	5%	7%			
Fe	4%	5%	7%	9%	14%		
Ca	4%	5%	7%	10%	14%	15%	
Si	4%	5%	7%	9%	14%	14%	14%

- “Dust” elements have much higher individual and shared errors than other species
- PM_{2.5} and “dust” element shared errors high at Phoenix

INDIVIDUAL AND SHARED ERROR ESTIMATES

- Bakersfield STN collocated site

	Br	S	PM_{2.5}	Fe	Ca	Si
Br	25%					
S	3%	4%				
PM_{2.5}	3%	2%	7%			
Fe	4%	4%	5%	13%		
Ca	6%	4%	5%	9%	14%	
Si	5%	2%	4%	9%	11%	15%

- Common error of 2-4% amongst all species
- “Dust” element individual errors similar to Phoenix but shared errors are lower



CONCLUSIONS

- IMPROVE and STN measurement errors are not independent amongst species
 - Strength of the dependencies varies by species and site
 - Sample volume error is observable
 - Size discrimination error affects “dust” elements – not accounted for by current uncertainty estimates
- Dependencies may bias data analysis results
- Ultimately need to modify sampling techniques to reduce errors



ACKNOWLEDGMENTS

- U.S. National Park Service

