On Using Model Performance Statistics in Applying Models

by
Akula Venkatram, Vlad Isakov* and Wenjun Qian
University of California, Riverside
* USEPA

Goal:

Convince you that model performance statistics should be used in the application of the model

- The model evaluation framework
- Application –near road concentrations of VOC
- Displaying model performance
- Conclusions

The Framework

$$C_o(\alpha,\beta) = C_p(\alpha) + \varepsilon(\alpha,\beta)$$

α = Model inputs

β = Variables not included in model

Need ε to simulate observations

Error is a component of the observation

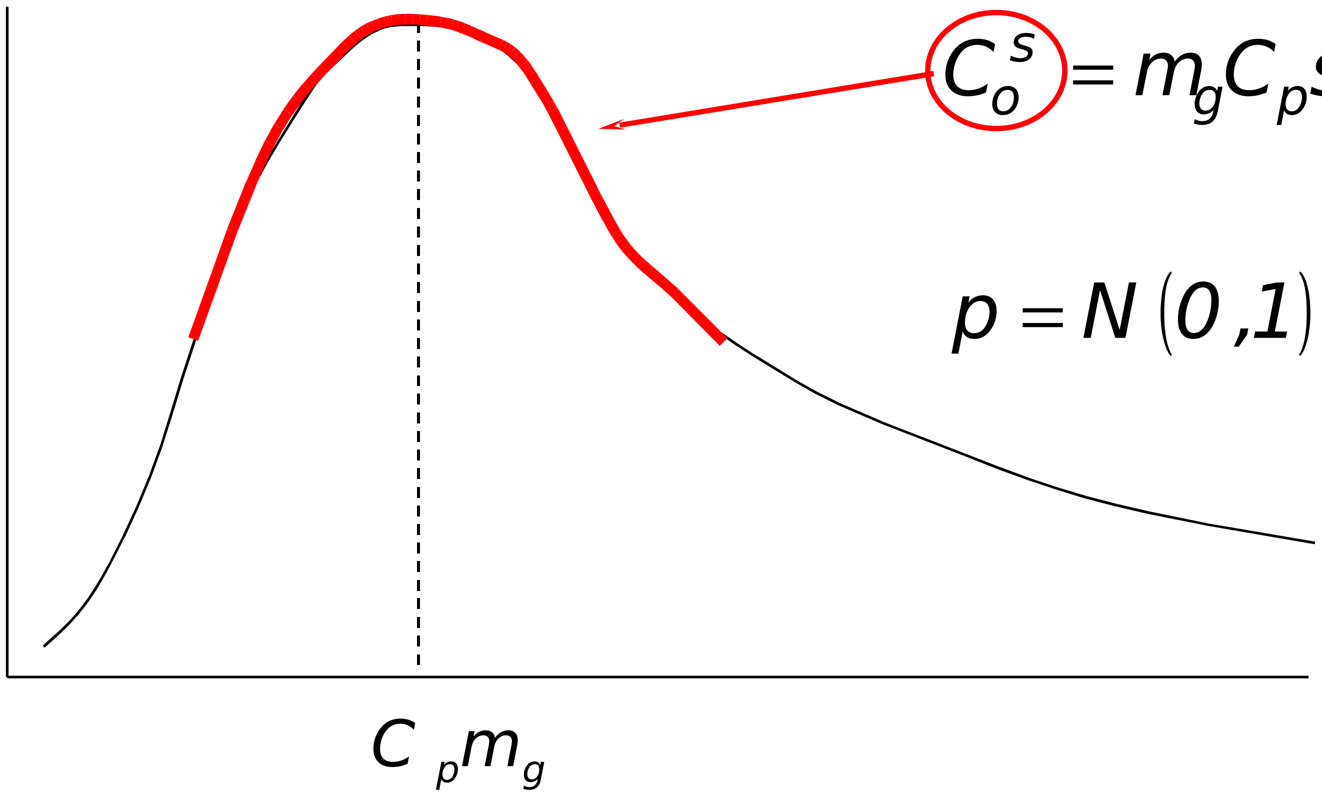
$$\ln(C_o) = \ln(C_p) + \varepsilon$$

$$m_g = \exp(\bar{\varepsilon})$$

$$s_g = \exp(\langle \varepsilon^2 \rangle^{1/2} - \bar{\varepsilon}^2)$$

The statistics m_g and s_g represent information that should be used explicitly in the application of the model.

Simulating Observations



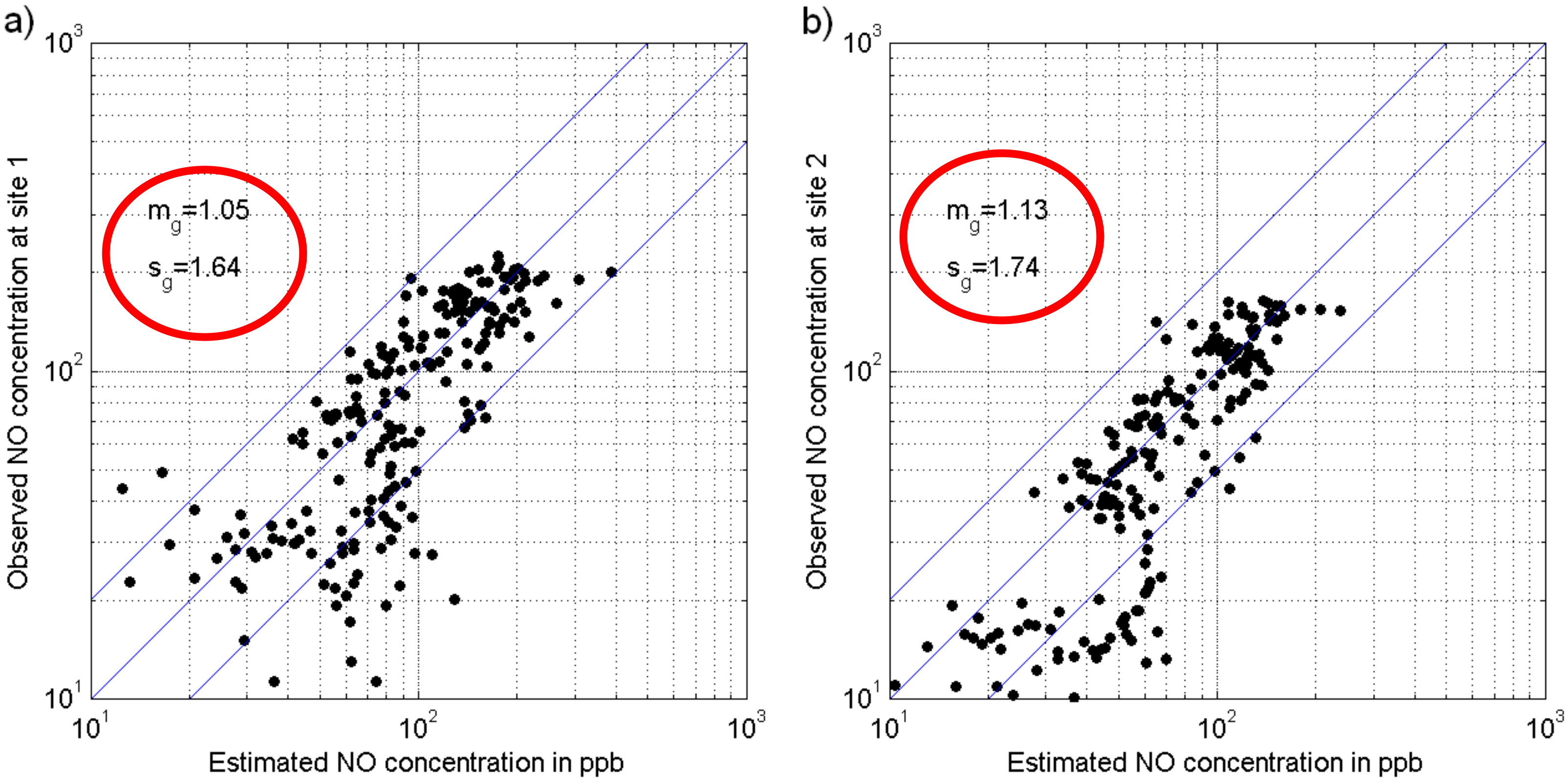
$$C_o^s = m_g C_p s_g^p$$

$$p = N(0,1)$$

Model estimates and error statistics are used to simulate observations

Model applied to estimate the contribution of traffic-generated emissions of VOCs – benzene, 1,3-butadiene, toluene – to concentrations at downwind receptors ranging from 10-m to 100-m from the edge of a major highway in Raleigh, North Carolina. during a field study conducted in August, 2006,

Model evaluated with NO concentrations measured at 7.6m and 17.6m from edge of road



Estimate contribution of highway to VOC concentrations

$$C_o - b = aC_o^s = am_g C_p s_g^p$$

$$b = Background\ concentration$$

$$a = Emission\ factor\ correction$$

The value of ***p*** represents lack of knowledge of variables that determines observed value.

We can only insist that mean and standard deviation of simulated observations are equal to those of actual observations

Details

$$Compute\ C_p$$

$$Compute\ C_o^s(i) = m_g C_p s_g^{p(i)}\ ;\ p(i) = N(0,1)$$

$$mean(aC_o^s) = mean(C_o - b)\ and\ std(aC_o^s) = std(C_o)$$

$$leads\ to$$

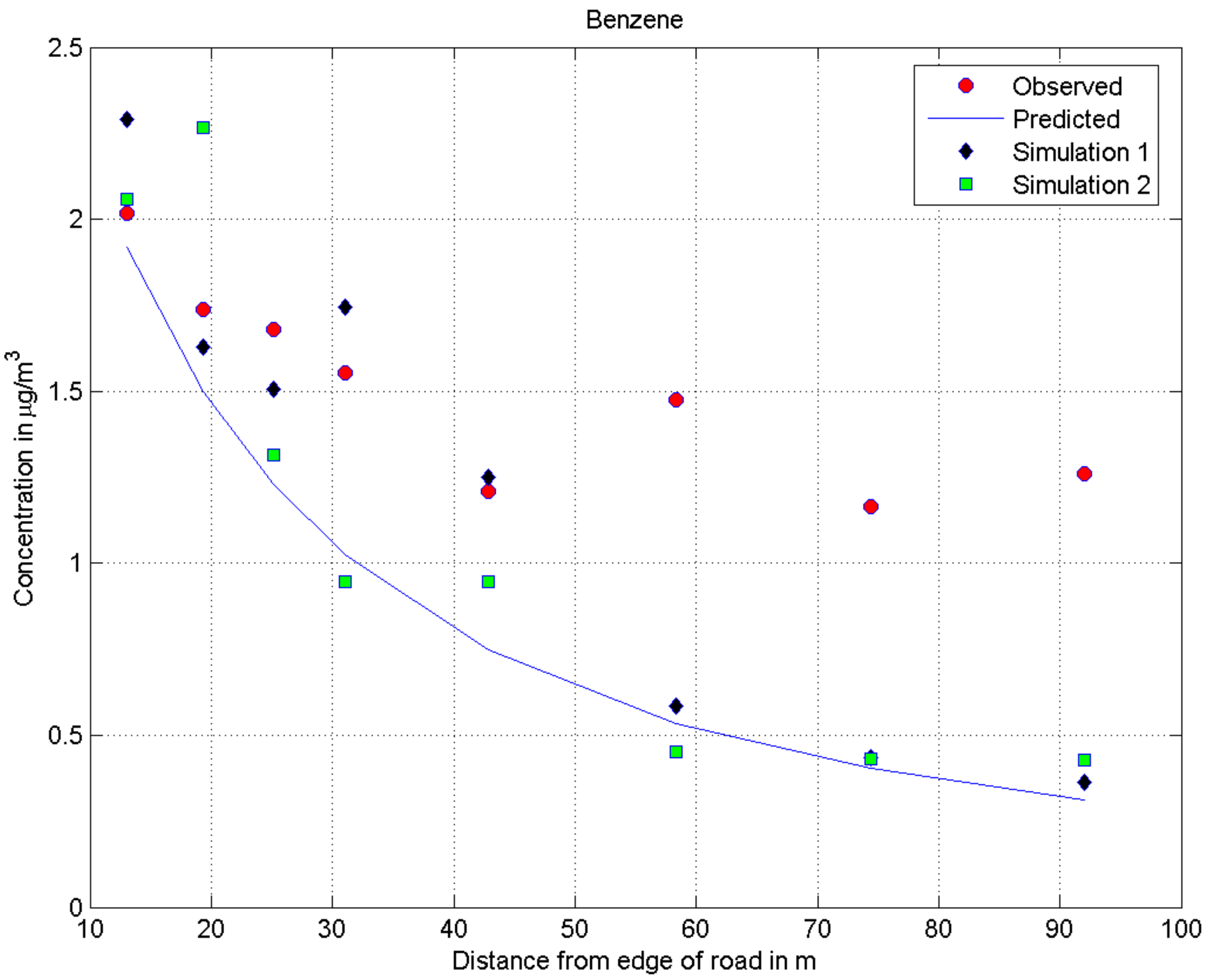
$$a = \frac{std(C_o)}{std(C_o^s)}\ and\ b = mean(C_o) - a\, s\, mean(C_o^s)$$

$$\bar{a} = mean(a) = emission\ correction\ factor$$

$$\bar{b} = mean(b) = background\ concentration$$

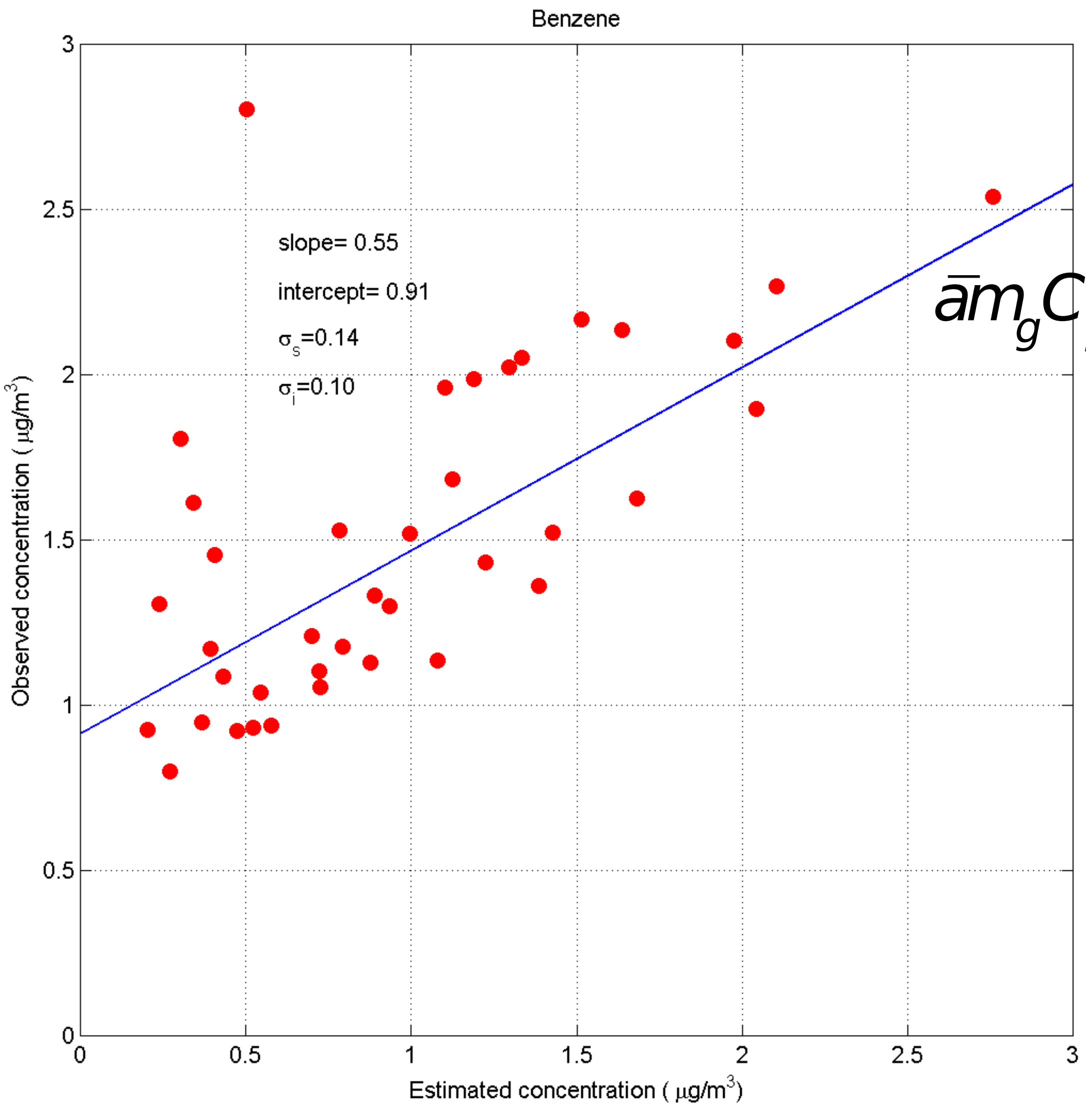
$$Contribution = \frac{\bar{a} m_g C_p}{\bar{a} m_g C_p + \bar{b}}$$

Benzene



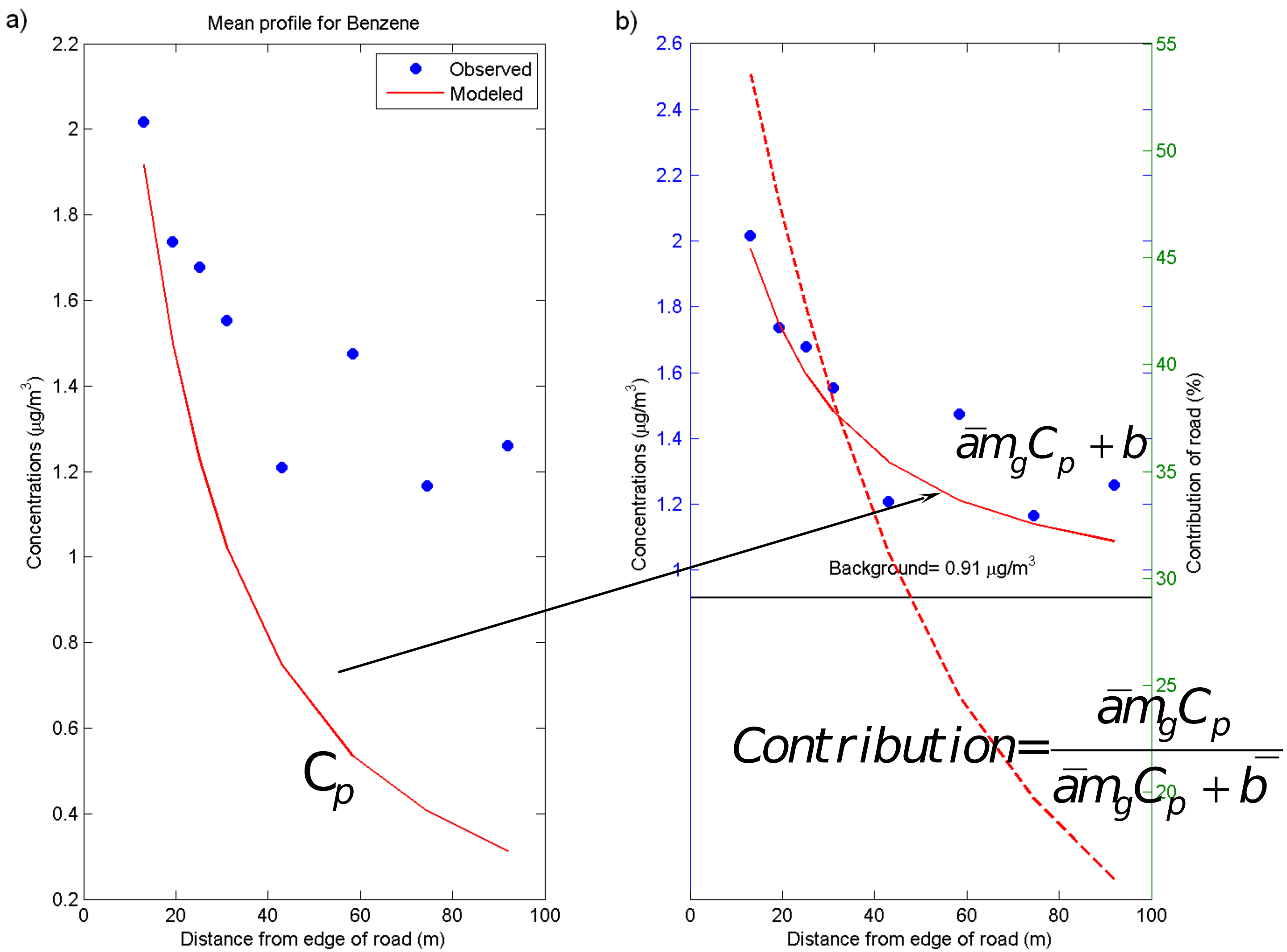
Benzene
Observed
Predicted
Simulation 1
Simulation 2
Concentration in μg/m3
Distance from edge of road in m
2.5
2
1.5
1
0.5
0
10
20
30
40
50
60
70
80
90
100

Benzene

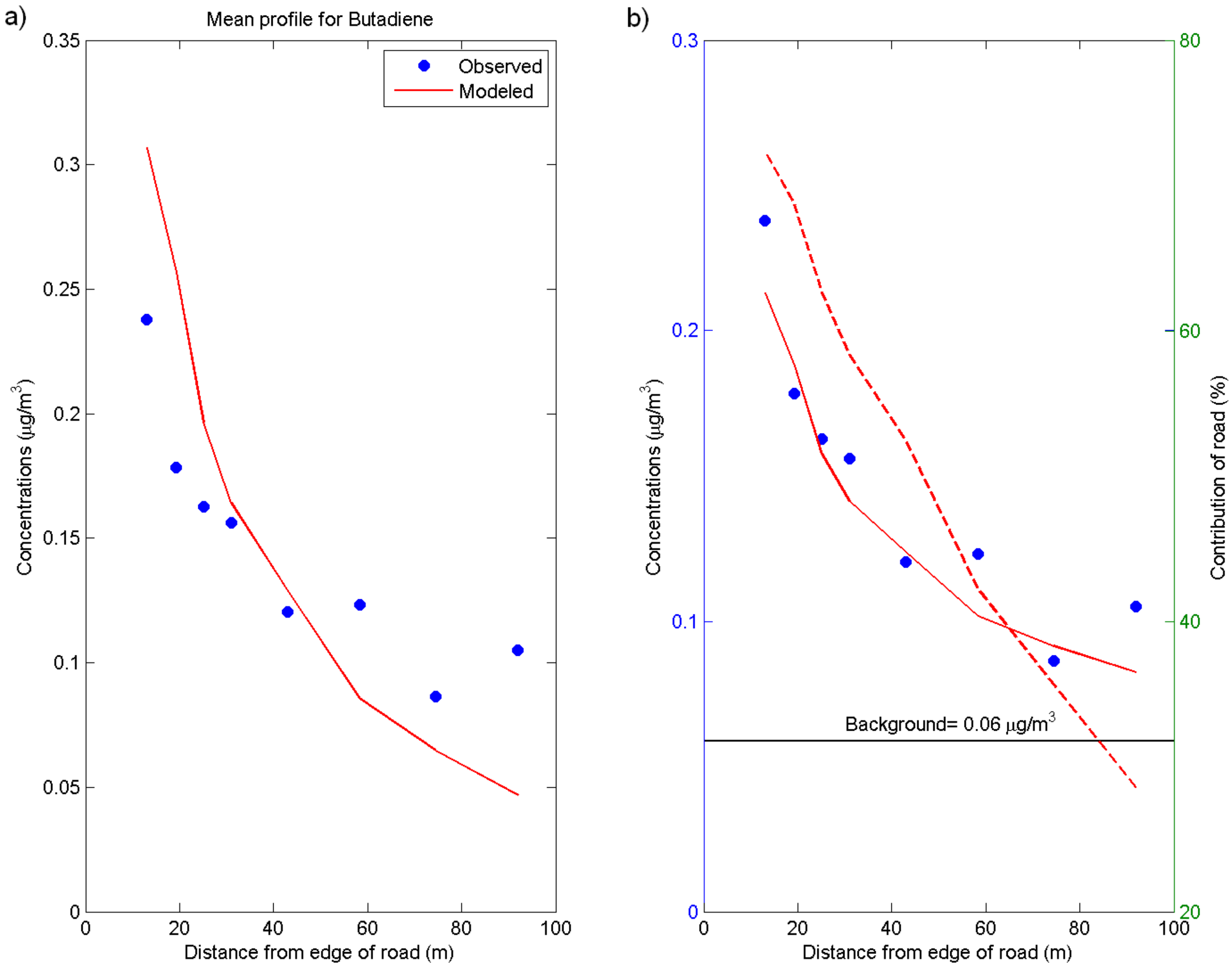


$$\bar{a} m_g C_p + \bar{b}$$

Contribution



1-3 Butadiene

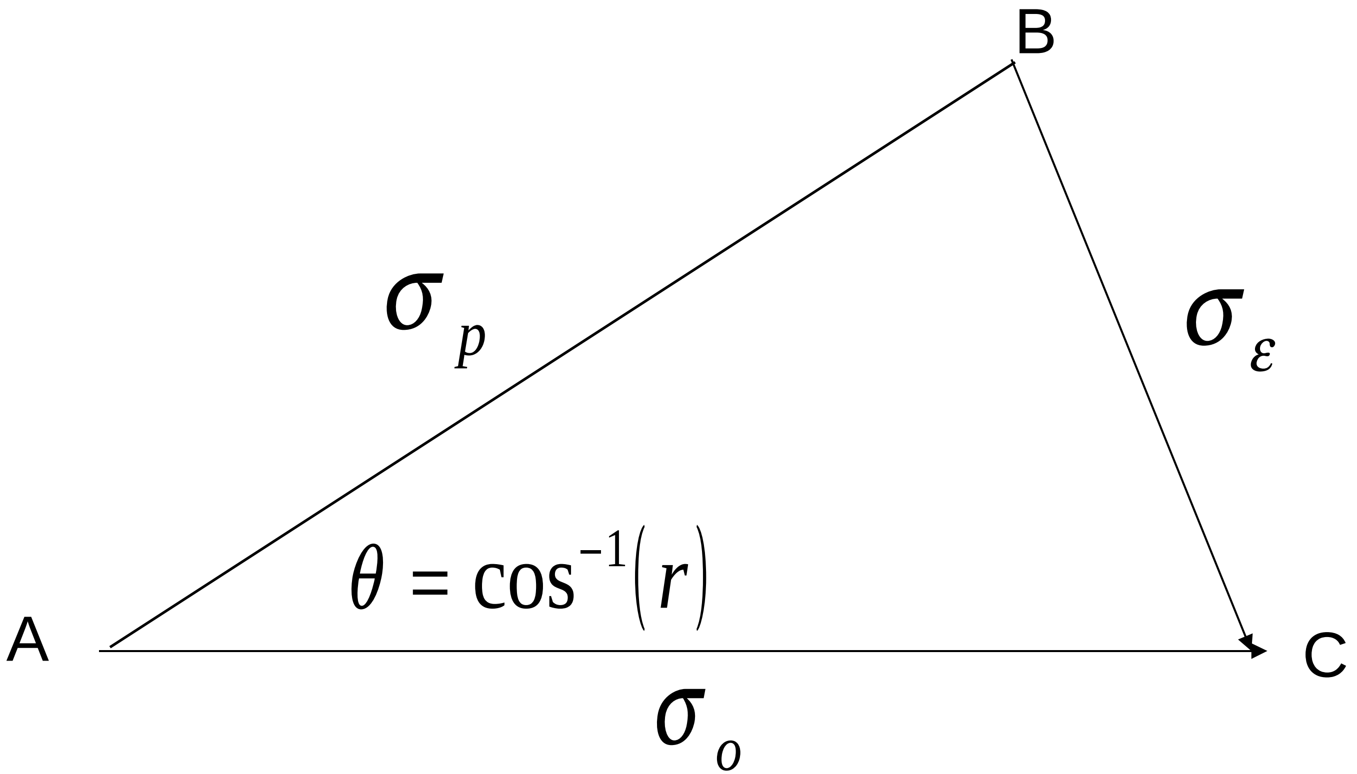


Summary

Model application is simulating observations

- Simulated observations can be used to estimate background concentrations and emission correction factors and their uncertainties
- Background concentrations and emission factors are consistent with independent measurements

Displaying Model Performance
Taylor diagram



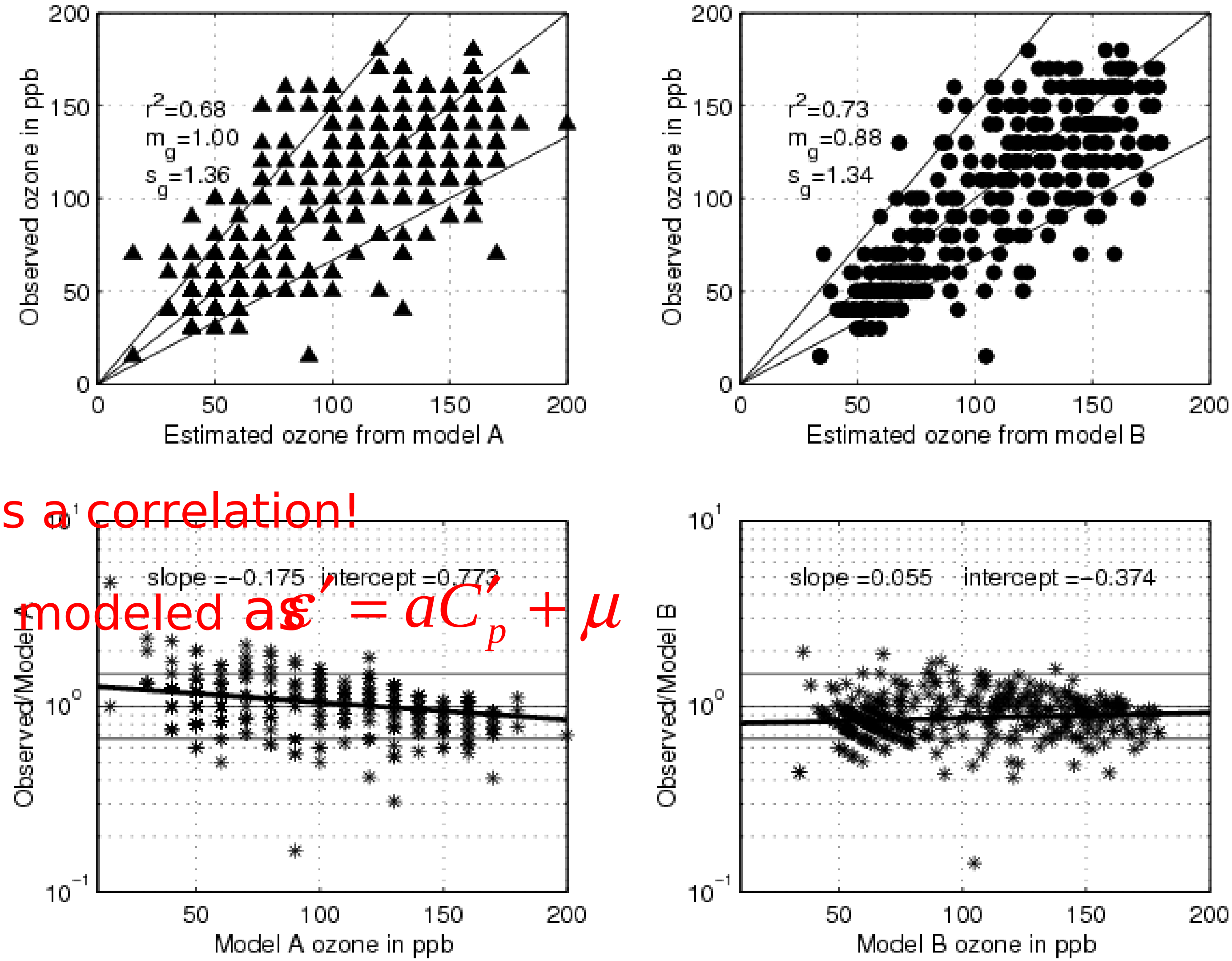
$$C'_o = C'_p + \varepsilon'$$

or

$$\varepsilon' = C'_o - C'_p$$

$$\sigma_\varepsilon^2 = \sigma_p^2 + \sigma_o^2 - 2r\sigma_p\sigma_o$$

Error correlation with model estimate is undesirable



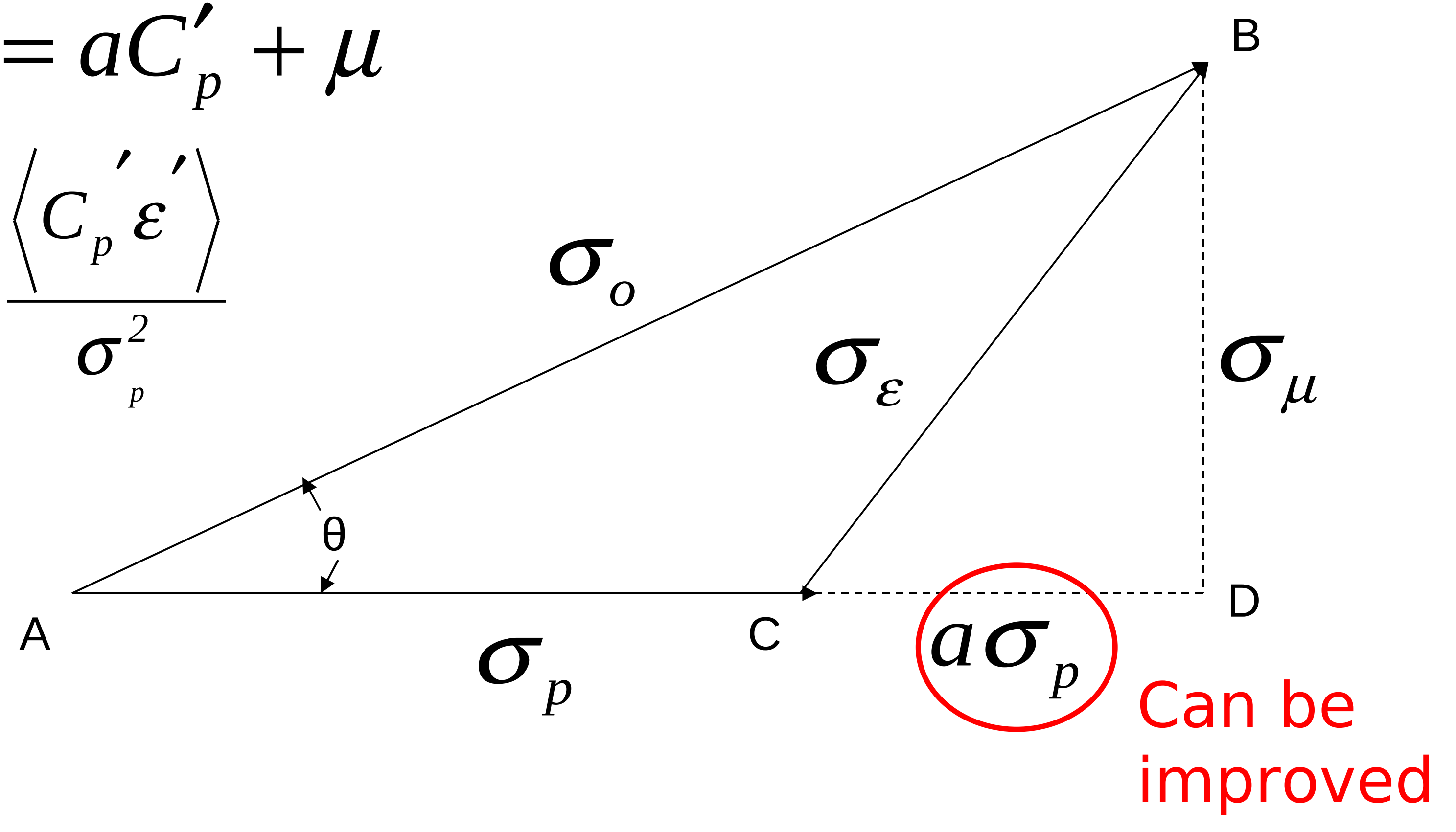
There is a correlation!

Can be modeled as $\varepsilon' = aC_p' + \mu$

Modified Diagram

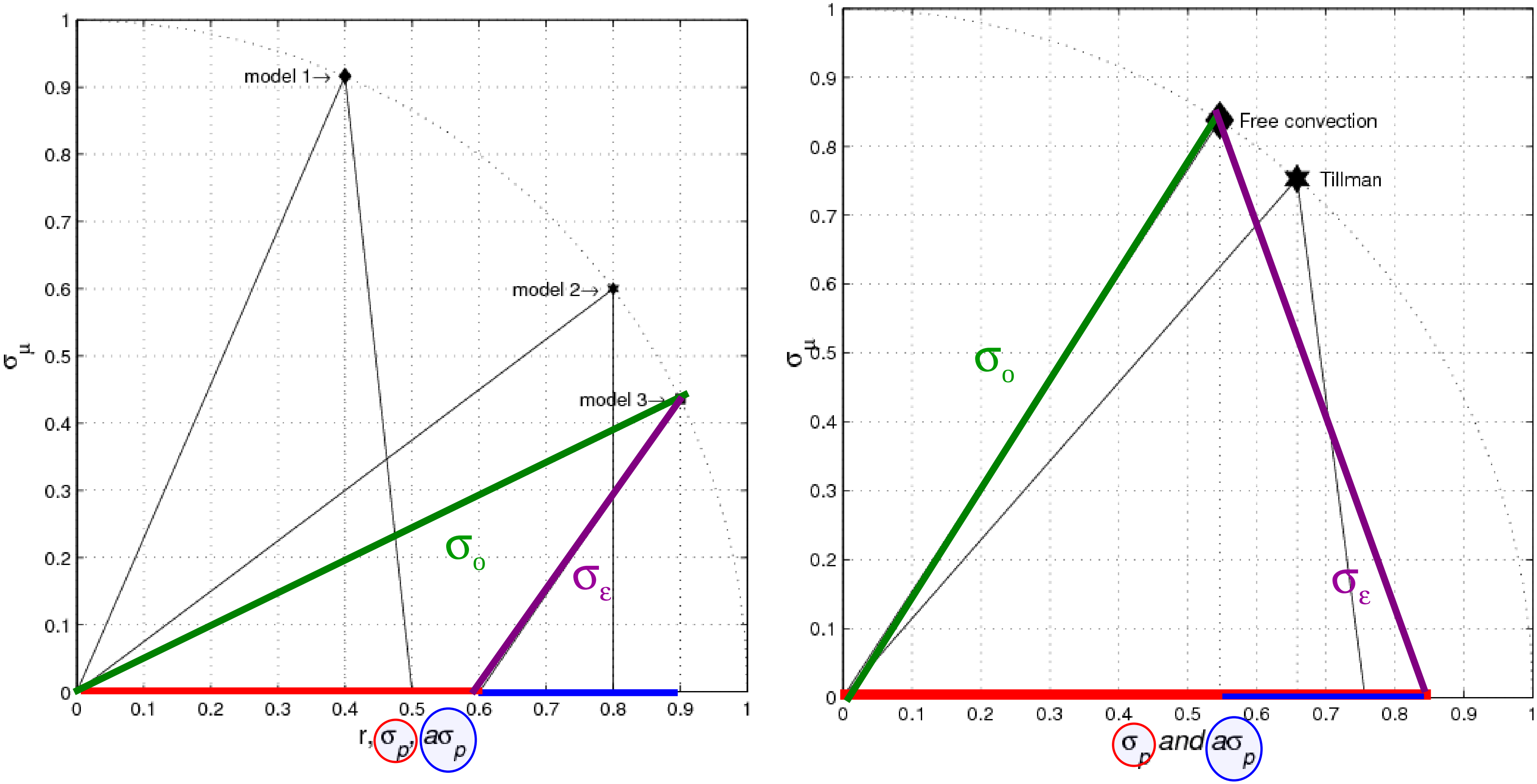
$$\varepsilon' = aC_p' + \mu$$

$$a = \frac{\langle C_p' \varepsilon' \rangle}{\sigma_p^2}$$



σ_μ is the inherent error

Comparing Model Performance



Summary

- “Error” statistics are an integral part of the model estimate
- The goal of model application is to simulate observations using the model estimate and error statistics
- Framework relating observation to model estimate can be translated to a vector diagram that displays properties of model performance