# SPATIAL DATA ANALYSIS TECHNICAL EXCHANGE WORKSHOP

Sponsored by:

U.S. EPA

at

Sheraton Imperial Hotel & Convention Center

Research Triangle Park, North Carolina

December 3-5, 2001

Network Design – Mixed Monitoring Technologies (Puget Sound) presented by:

Steven M. Bortnick

Battelle

SDAS

# Reminder

- $PM_{2.5}$ monitoring in Puget Sound
  - Mixed network:  FRMs, TEOMs, Nephelometers
  - Evaluate performance
    - – Look at annual $PM_{2.5}$ spatial process
    - – Assess network's MSPE summary statistics for that process
      - - current versus alternative designs
    - – Keep in mind:
      - - Federal regulations
      - - real-world bias adjustment of data

# Notation

- Z = data, as measured
- Y = true $PM_{2.5}$
- Asterisks (*) denote bias-adjusted data
- Will clarify more as we go along
- Will suppress subscripts/superscripts whenever possible !

# Basic Model Form

- Z* = Y + , * ® Data

- Y = X ( + * ® Process

- Written hierarchically, where:
  - X ( = deterministic large scale spatial trend
  - * = stochastic small scale variation
  - , = error in measurement

# Basic Model Assumptions

- *, , independent and normally distributed
- E [*] = E [ , ] = 0
- Var [*] = W → some spatial covariance matrix
  - explored several
  - spherical used in examples
- Var [ , *] = V*

Battelle

# More on V*

- With FRM-only network, V* = V = $F^2$ I

- With mixed network but no bias, V* = V = diag $\{\sigma_1^2, \sigma_2^2, \ldots\}$

- With mixed network, bias, and real-world bias correction, V* =
  - No longer diagonal !
  - Depends on measurement errors
  - Depends on uncertainty in bias correction

# Bias Correction

- Consider daily data, t denotes day

- In practice:
  - Assume $E\,[Z_F(t)] = " + \$E\,[Z_C(t)]$
  - $(", \$)$ = fixed area-wide bias correction
  - Suggests SOME uncertainty in bias adjustment, e.g., $se(\hat{\beta}) \approx \frac{\sigma_1}{\sqrt{n}}$

- Closer to reality:
  - $(", \$)$ = fixed area-wide bias correction
  - $("_s, \$_s)$ = random site-specific or monitor-specific bias
  - Suggests MORE uncertainty in bias adjustment, e.g., $se(\hat{\beta}) \approx \sqrt{\frac{\sigma_1^2}{n} + \sigma_2^2}$

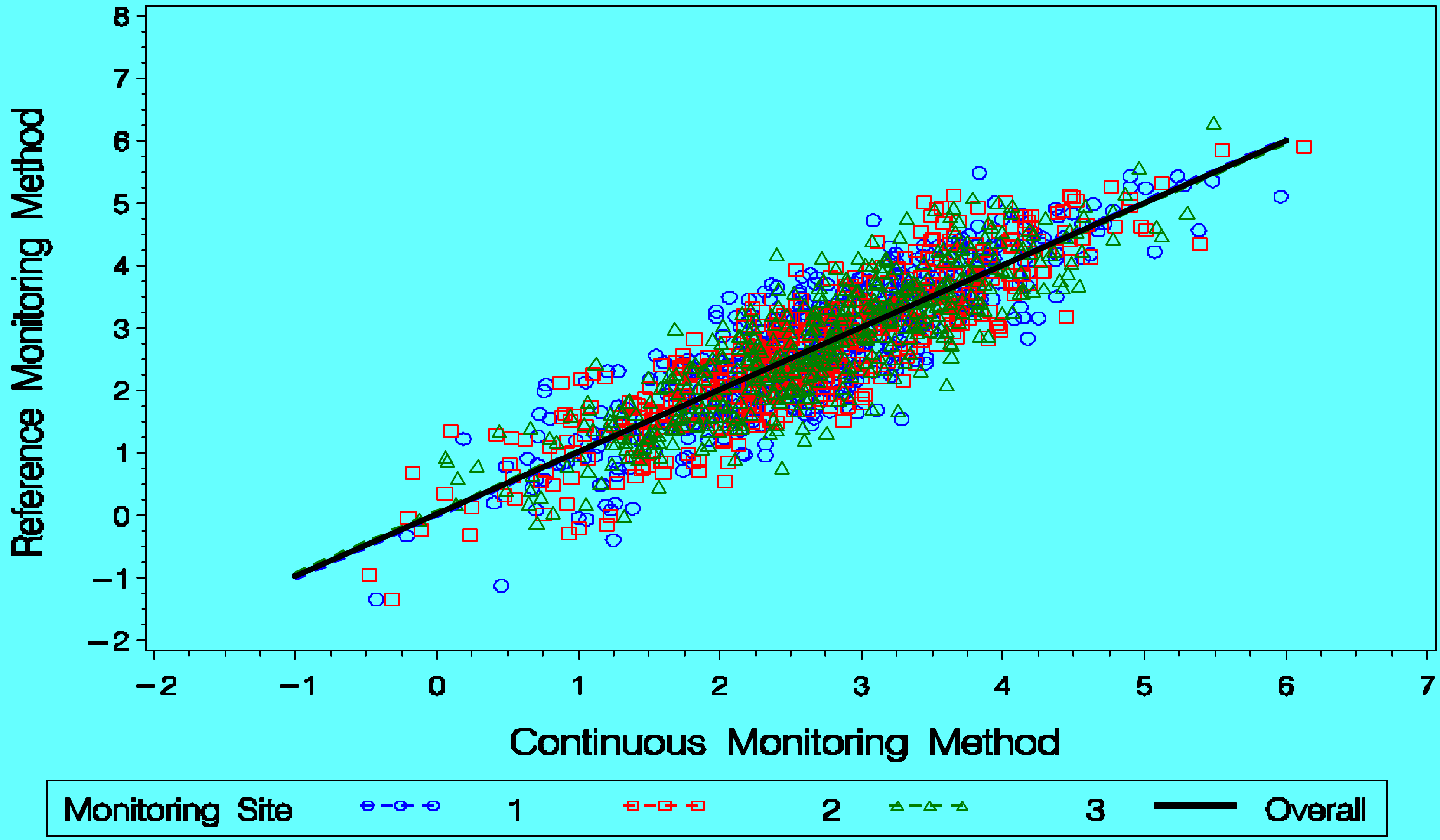
# Example of Uncertainty in Bias Adjustment

- Simulated three sites of co-located FRM and continuous data.
- Increased degree of between-site heterogeneity in continuous bias.
- Applied ordinary least squares, ignoring site effect.
- Compared $se(\hat{\beta})$ versus degree of heterogeneity.

Battelle

SDAS

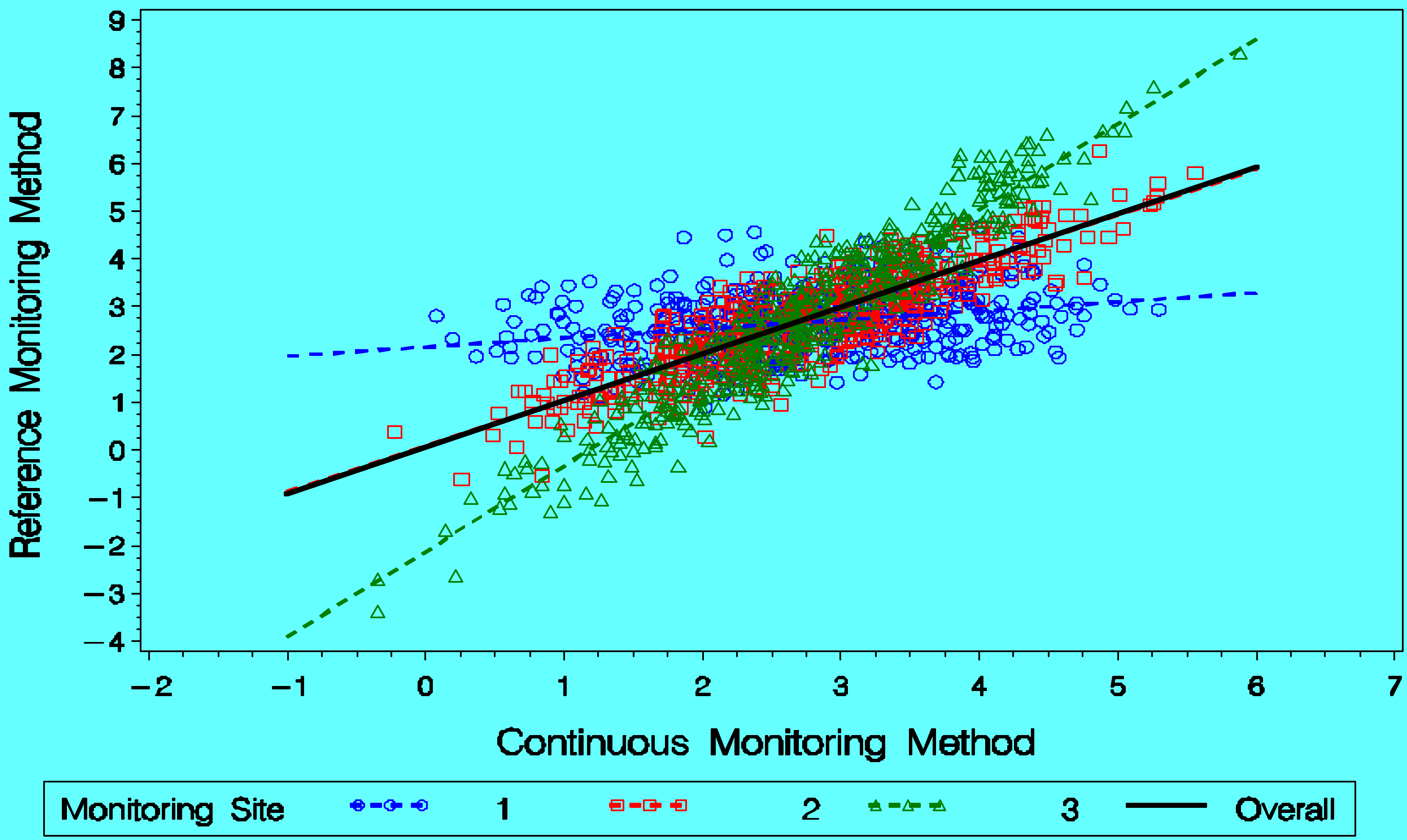
# Complete Spatial Homogeneity




Battelle

SDAS

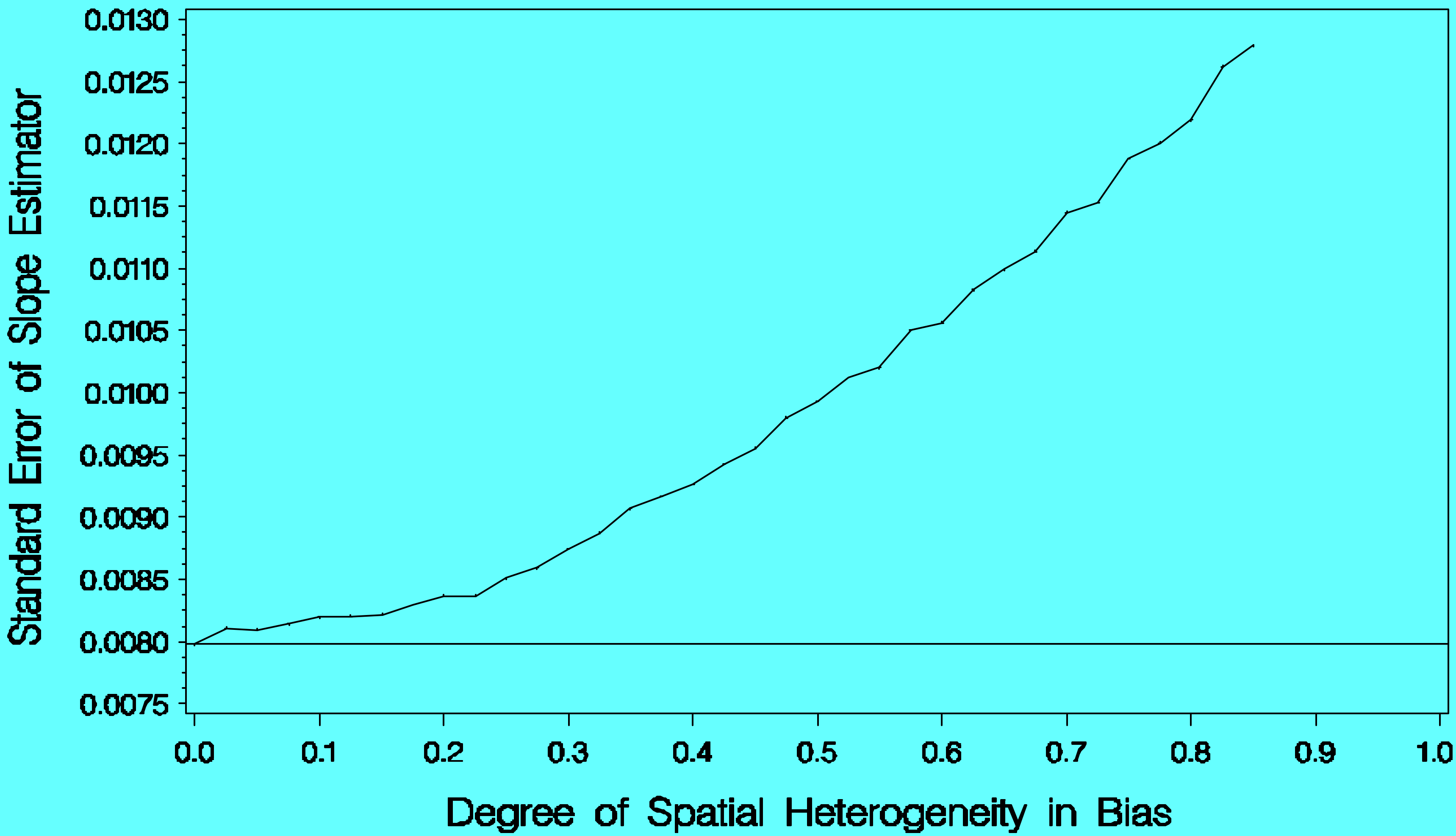
# Spatial Heterogeneity in the Bias




Battelle

SDAS

# Variability in Slope Estimate Versus Degree of Spatial Heterogeneity




Battelle

SDAS

# Back to V*

- Recall, in practice, data from all co-located FRM, continuous sites pooled to estimate (", $).

- This gives:
  - $Z_C^*(s) = \alpha + \beta Z_C(s)$ for continuous annual average $PM_{2.5}$
  - $Z_F^*(s) = Z_F(s)$ for FRM annual average $PM_{2.5}$

- Include some assumptions:
  - Day-to-day independence of 24-hour integrated $PM_{2.5}$ data
  - No significant seasonal or spatial trends
  - Monitor-to-monitor independence, conditional on Y

# Back to V* (continued)

- Along with assumptions, some **Very Careful Bookkeeping** gives V* as:

  - Diagonal elements:
    - $Var(Z^{*}_{iF}) = \sigma^2_F \{1/n_{iF}\}$, for FRMs
    - $Var(Z^{*}_{iC}) = \sigma^2_F \{ 1/M + s [ 1/M + 1/n_{iC} - 2/M (m_{iC}/n_{iC}) ] \}$ , for continuous

  - Off-diagonal elements:
    - $Cov(Z^{*}_{iC}, Z^{*}_{jC}) = \sigma^2_F \{ 1/M + s [ 1/M - 1/M (m_{iC}/n_{iC} + m_{jC}/n_{jC}) ] \}$ , for continuous i and j
    - $Cov(Z^{*}_{iF}, Z^{*}_{jF}) = 0$ , for FRM i and j
    - $Cov(Z^{*}_{iF}, Z^{*}_{jC}) = \sigma^2_F \{ 1/M (m_{iF}/n_{iFC}) \}$ , for FRM i and continuous j

# Back to V* (continued)

- Where
  - $F^2_F$ = FRM measurement error
  - $n_{iF}$ = number of daily observations in annual average of FRM at i-th site
  - $n_{iC}$ = number of daily observations in annual average of continuous at i–th site
  - $m_{iF}$ = number of daily observations used from FRM at i-th site in bias correction
  - $m_{iC}$ = number of daily observations used from continuous monitor at i-th site in bias correction
  - M = total number of observations used in bias correction
  - s = ($F^2_b$ + $\$^2$) $r^2$, where $ is slope of bias correction, $F_b$ is standard error of slope estimate, and r is ratio $F_c/F_f$, where $F^2_c$ is continuous measurement error
- Recall, $\sigma_b^2 = \frac{\sigma_1^2}{M} + \sigma_2^2$

# V* Example

- Consider 4 FRM and 4 continuous monitors
- Assume daily data collected as:

| | is.frm | n.mean | n.reg |
|---|---|---|---|
| F-1 | 1 | 120 | 0 |
| F-2 | 1 | 120 | 120 |
| F-3 | 1 | 60 | 0 |
| F-4 | 1 | 60 | 60 |
| C-1 | 0 | 240 | 0 |
| C-2 | 0 | 240 | 120 |
| C-3 | 0 | 120 | 0 |
| C-4 | 0 | 120 | 60 |

- First 4 lines correspond to FRMs
- F-2, C-2 co-located and F-4, C-4 co-located

# V* Example (continued)

- Suppose $=1, $se(\hat{\beta}) = 0.05$, and $(F_C / F_F) = 1$ then correlation matrix from V* is:

| | F-1 | F-2 | F-3 | F-4 | C-1 | C-2 | C-3 | C-4 |
|---|---|---|---|---|---|---|---|---|
| F-1 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| F-2 | 0.00 | 1.00 | 0.00 | 0.00 | 0.49 | 0.62 | 0.44 | 0.52 |
| F-3 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| F-4 | 0.00 | 0.00 | 0.00 | 1.00 | 0.35 | 0.44 | 0.31 | 0.36 |
| C-1 | 0.00 | 0.49 | 0.00 | 0.35 | 1.00 | 0.68 | 0.64 | 0.57 |
| C-2 | 0.00 | 0.62 | 0.00 | 0.44 | 0.68 | 1.00 | 0.61 | 0.48 |
| C-3 | 0.00 | 0.44 | 0.00 | 0.31 | 0.64 | 0.61 | 1.00 | 0.51 |
| C-4 | 0.00 | 0.52 | 0.00 | 0.36 | 0.57 | 0.48 | 0.51 | 1.00 |

- F-i, F-j un-correlated
- F-2 and F-4 correlated with C-i
- C-i, C-j all correlated

# V* Example (continued)

- Suppose $F_F = 0.1$, then V* diagonals are:

| F-1 | F-2 | F-3 | F-4 | C-1 | C-2 | C-3 | C-4 |
|---|---|---|---|---|---|---|---|
| 0.00913 | 0.00913 | 0.0129 | 0.0129 | 0.0124 | 0.00987 | 0.014 | 0.0118 |

- F-diagonals vary due to different number of daily observations ONLY
- C-diagonals vary due to different number of daily observations AND whether or not co-located.

# V* Properties

- As desired, penalizes for:
  - Smaller number of daily observations
  - Smaller number of co-located FRM sites
  - Continuous monitor not being co-located with FRM (0-1)
  - TOTAL error in bias correction

- Accounts for correlation induced by real-world bias correction process

# Summary

- Interested in MSPE of BLUP for network evaluation
- MSPE = f(W, V*)
- W = variability in true spatial process
- V* = uncertainty in measurement process
  = f (sample size, measurement errors, true area-wide bias, $\sigma_b^2$)
- $\sigma_b^2$ = error in bias estimation/adjustment
  = f (sample size, measurement errors, heterogeneity in site-to-site or monitor-to-monitor bias)

# Prediction

- To predict Y-process at new site, S', BLUP of $Y_{s'}$ is:

  $$p_{s'} = x`_{s'}\, g + c`_{s'}\, E^{-1} (Z - Xg)$$

- Where $E = V^* + W$

- MSPE of BLUP can be written as:

- $$MSPE(p_{s'}) = Var(p_{s'} - Y_{s'}) = Var(Y_{s'}) + Var(p_{s'}) - 2\, Cov(P_{s'}, Y_{s'})$$
  $$= Var(Y_{s'}) + P_{s'}\, E\, P`_{s'} - 2\, P`_{s'}\, c_{s'}$$
  $$= Var(Y_{s'}) + P_{s'}\, V^*\, P`_{s'} + P_{s'}\, W\, P`_{s'} - 2\, P`_{s'}\, c_{s'}$$

# Parameter Estimation (FRM, TEOM)

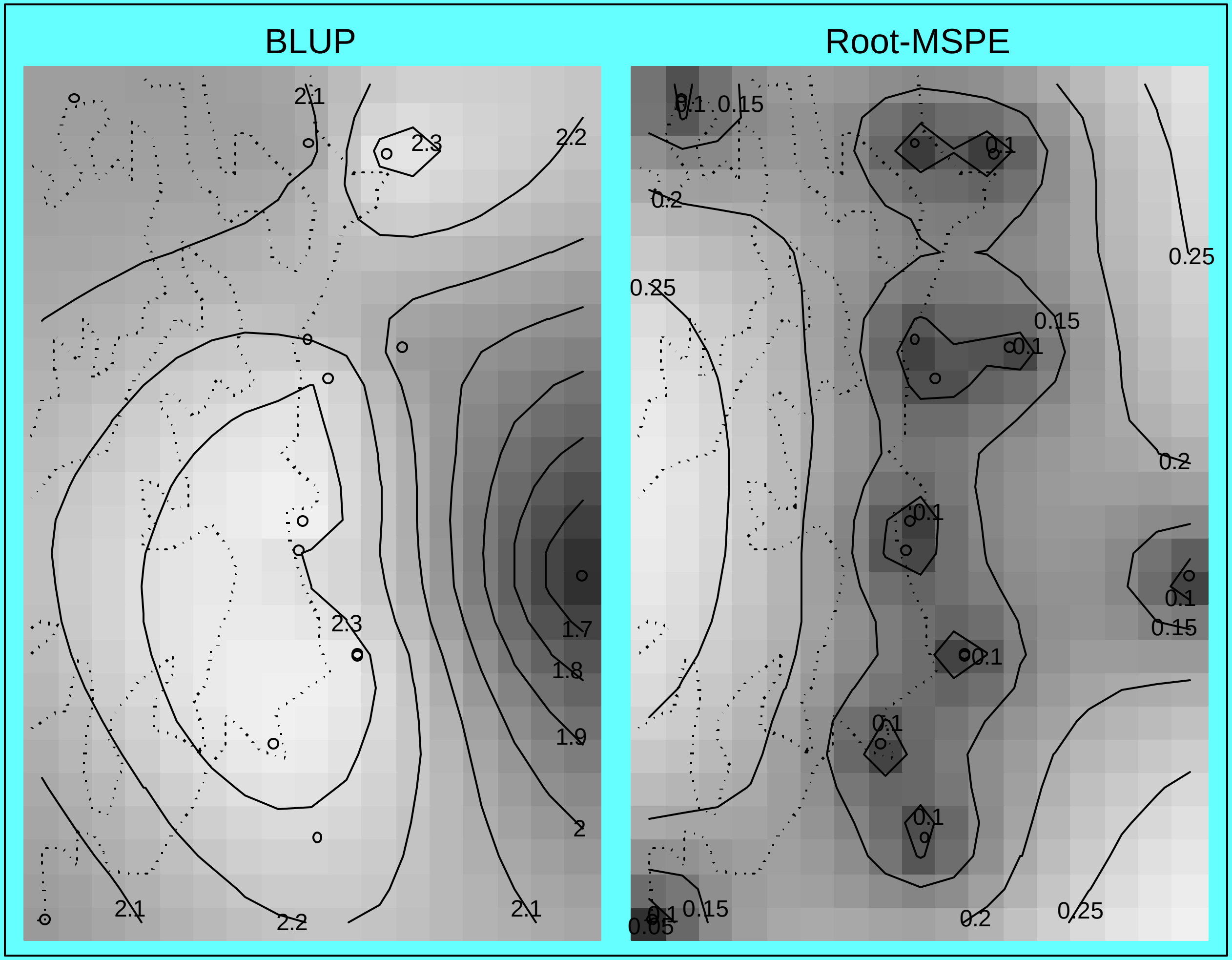
- Use data at hand to estimate unknown parameters in W and V*:
  - Restricted maximum likelihood (REML) in $S^+$ applied to 14 sites of FRM annual averages yields $\sigma_F^2$ and W (spherical spatial correlation, no nugget effect)
  - Linear mixed model fit by REML in $S^+$ applied to co-located FRM, TEOM daily data yields $\alpha$, $\beta$, and $\sigma_b^2$
  - Can obtain $\sigma_C^2$ a number of ways
  - Combined with a given network specification, yields MSPE for a mixed "network" of FRMs and continuous

# Results – FRMs Only

- Fit model to 14 sites of FRM annual averages
- Yields estimates:
  - $\sigma^2_F$ = 0.034
  - $\sigma^2_W$ = 0.078
  - Range = 70.5 kilometers
- Summary Statistics

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|---|---|---|---|---|---|---|
| BLUP | 1.60000 | 2.066 | 2.1580 | 2.1460 | 2.2520 | 2.375 |
| root-MSPE | 0.04632 | 0.151 | 0.1827 | 0.1838 | 0.2203 | 0.287 |

# Results – FRMs Only (continued)



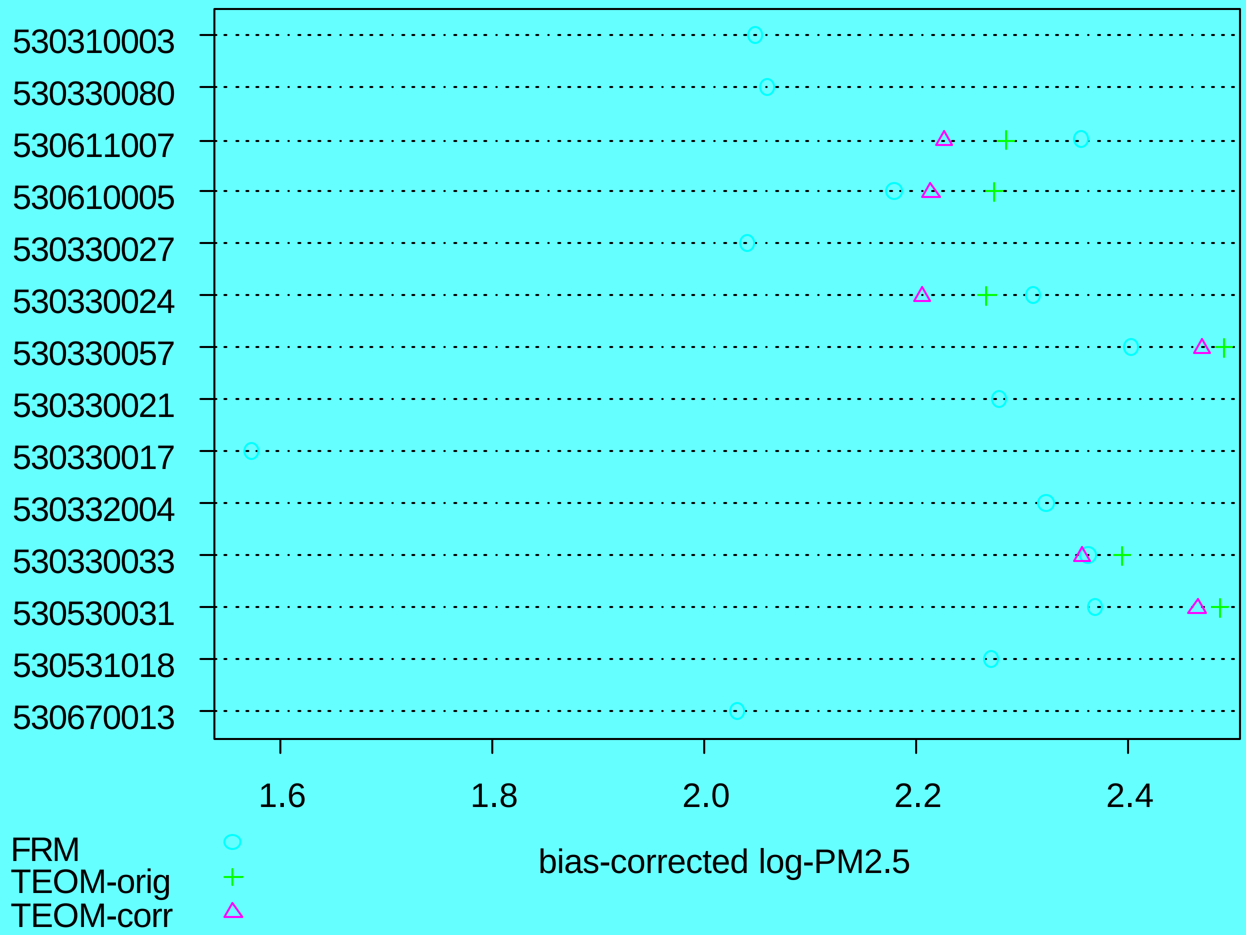

Battelle

SDAS

# Results – Bias Correction

- Fit mixed model to daily co-located FRM, TEOM data
- Yields Estimates:
  - " = -0.474, $ = 1.182
  - $\sigma_C^2$ (unadjusted) = 0.011
  - $\beta^2\sigma_C^2$ (bias adjusted, $ known) = 0.015 (se **8**18%)
  - $\left(\sigma_b^2+\beta^2\right)\sigma_C^2$ (bias adjusted, $ unknown) = 0.016 (se **8**22%)
  - $\left(\sigma_b^2+\beta^2\right)\sigma_C^2$ (bias adjusted, $ unknown, site-to-site variation) = 0.024 (se **8**48%)

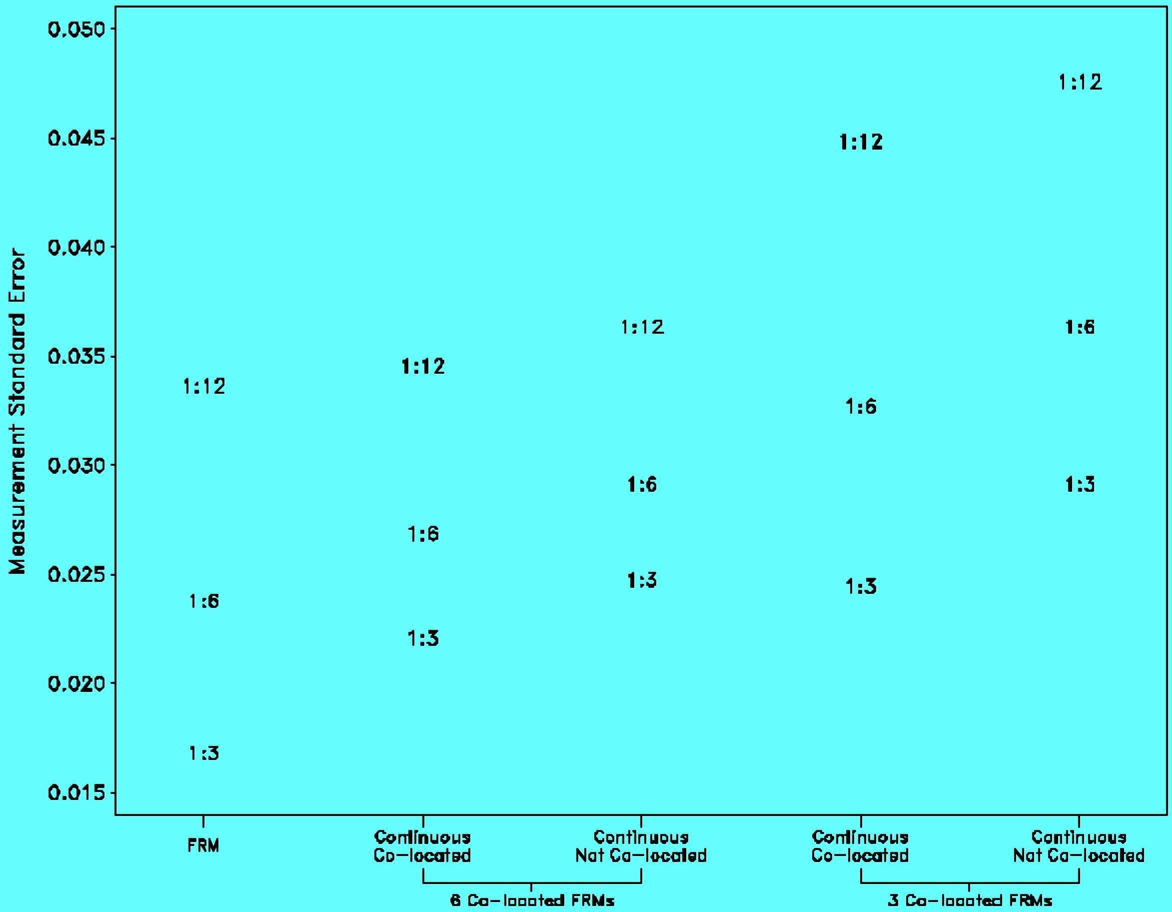
# Results – Bias Correction (continued)




Battelle

SDAS

# Results – Bias Correction (continued)



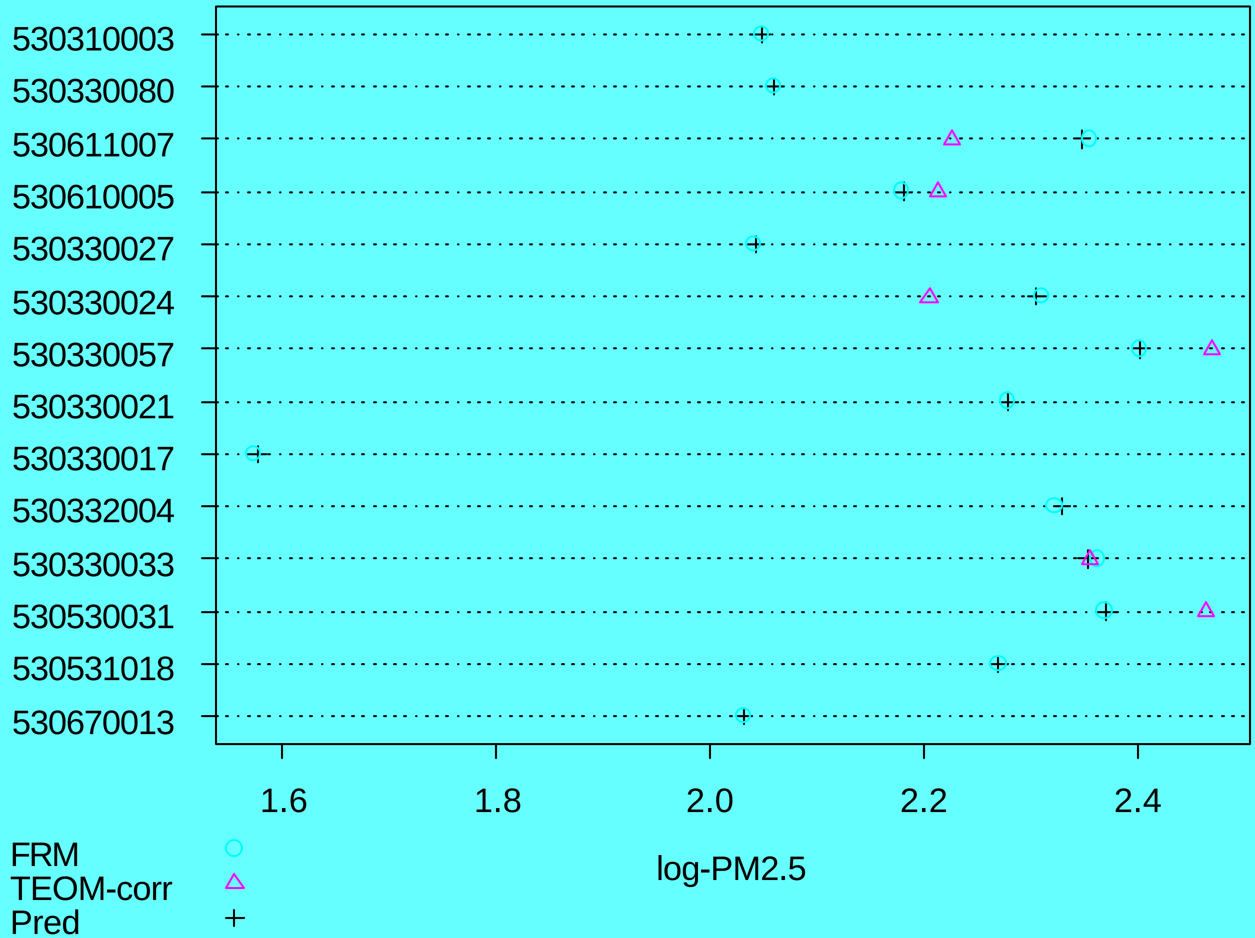

Battelle

# Results – FRMs and TEOMs

- See results from FRM-only model (virtually identical)

- Why is that ?
  - Propagation of error (good !)
  - Continuous versus FRM sample sizes (not representative)

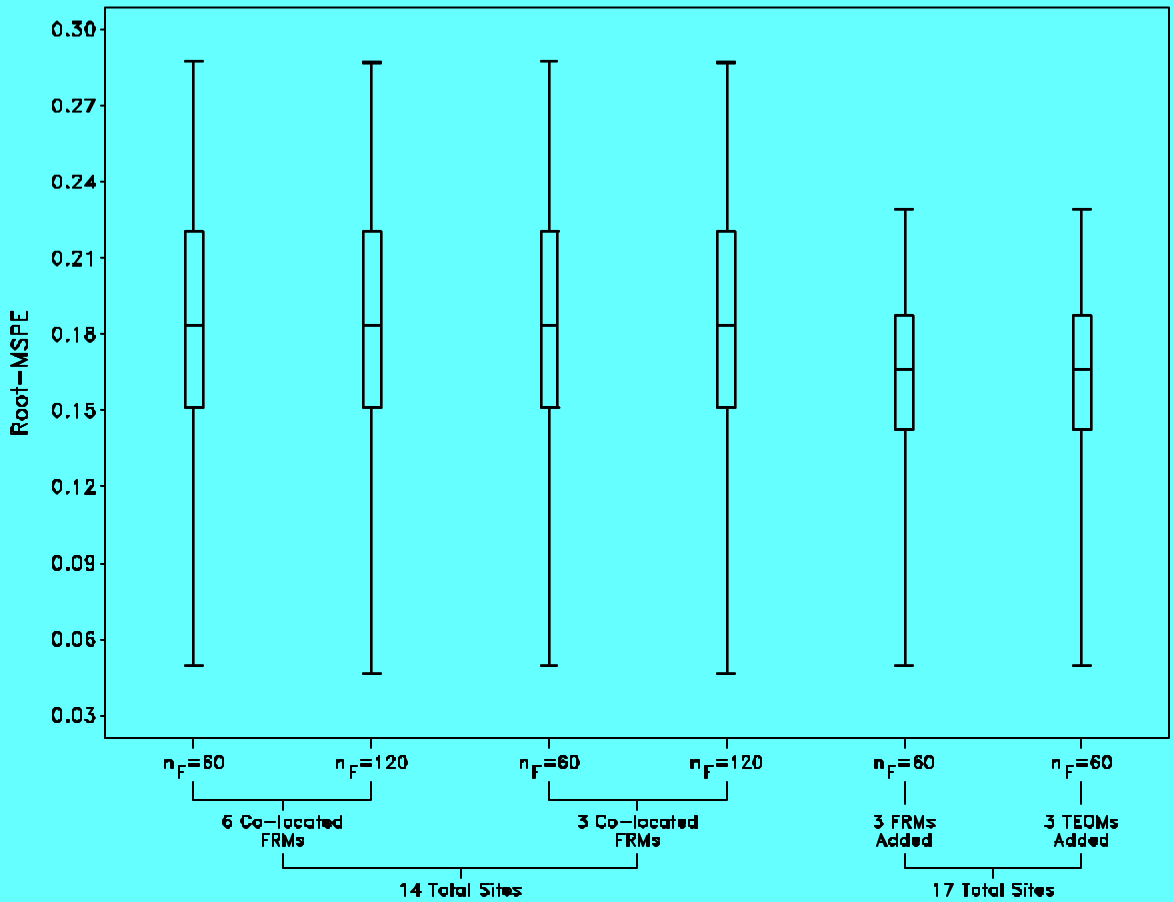
# Results – FRMs and TEOMs (continued)



# Results – Alternative Designs

- Current network of 14 FRMs with 6 co-located TEOMs
  - $n_F$ = 60 (1:6), $n_C$ = 360 **6** 14 total sites
  - $n_F$ = 120 (1:3), $n_C$ = 360 **6** 14 total sites
- Remove 3 of the co-located FRMs
  - $n_F$ = 60 (1:6), $n_C$ = 360 **6** 14 total sites
  - $n_F$ = 120 (1:3), $n_C$ = 360 **6** 14 total sites
- Add 3 new monitoring sites
  - 3 FRMs, $n_F$ = 60 (1:6), $n_C$ = 360 **6** 17 total sites
  - 3 TEOMs, $n_F$ = 60 (1:6), $n_C$ = 360 **6** 17 total sites

# Alternative Designs (continued)

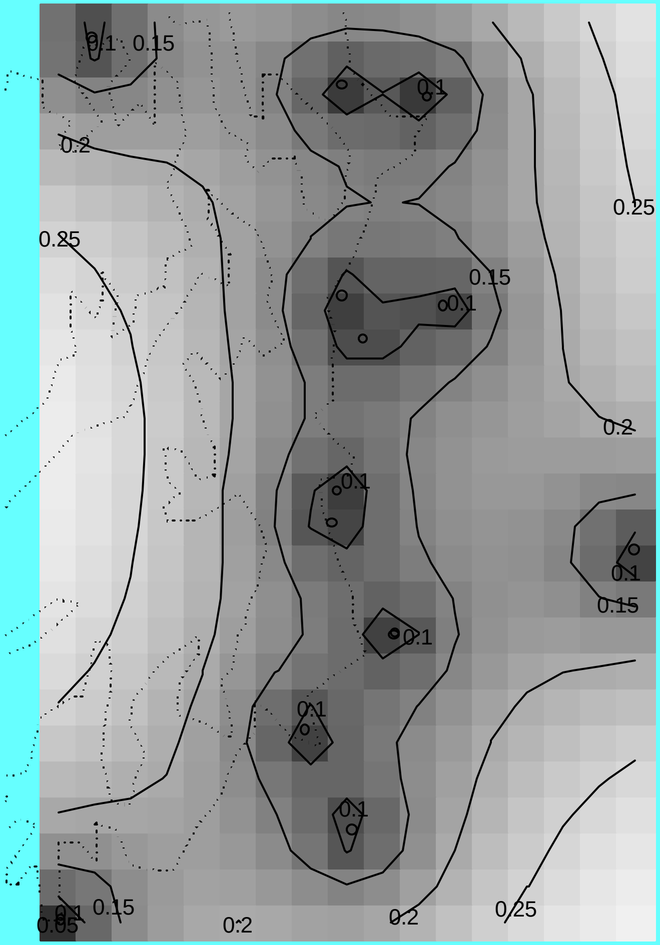



Battelle

SDAS

# Alternative Designs (continued)

Root-MSPE

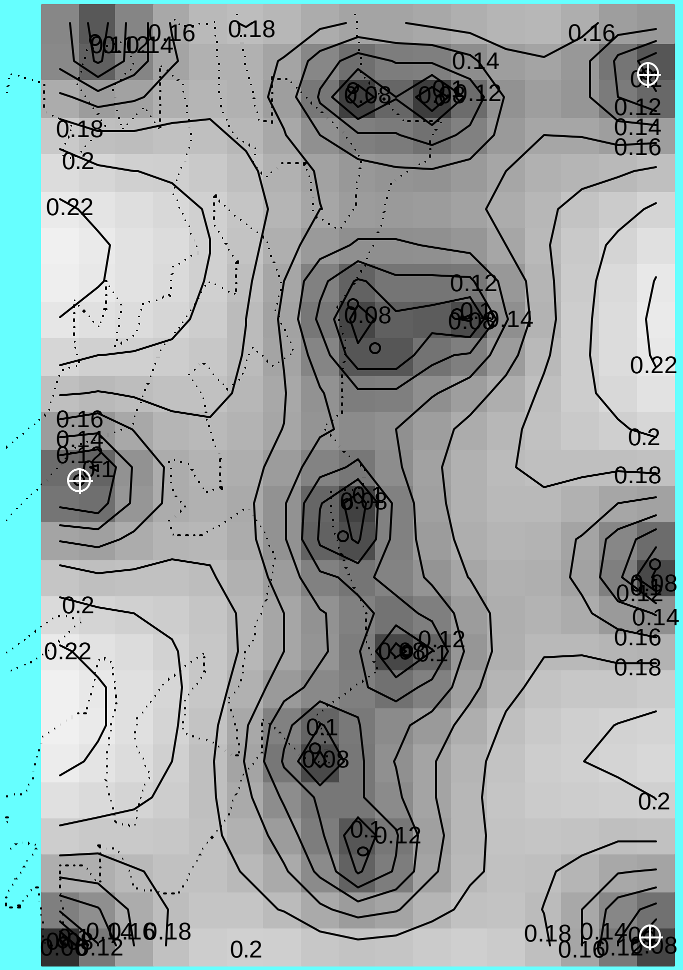



**Root-MSPE for 14 sites**

Battelle

SDAS

# Alternative Designs (continued)




**Root-MSPE for 17 sites**

# Wrap-Up

- The machine is built.
- Tweaks to consider:
  - Approach to parameter estimation
  - Network of multiple continuous types
- Scenarios to explore:
  - Many network alternatives
  - Numerous sensitivity analyses
  - Quarterly average time scale
  - Cost implications
  - Areas other than Puget Sound