SAND2006-0939C

SANDIA NATIONAL LABORATORIES

HOMELAND SECURITY

Mapping Contaminants in Buildings

Sean A. McKenna and Patrick Finley

Sandia National Laboratories

Sandia National Laboratories

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94-AL-85000

Background

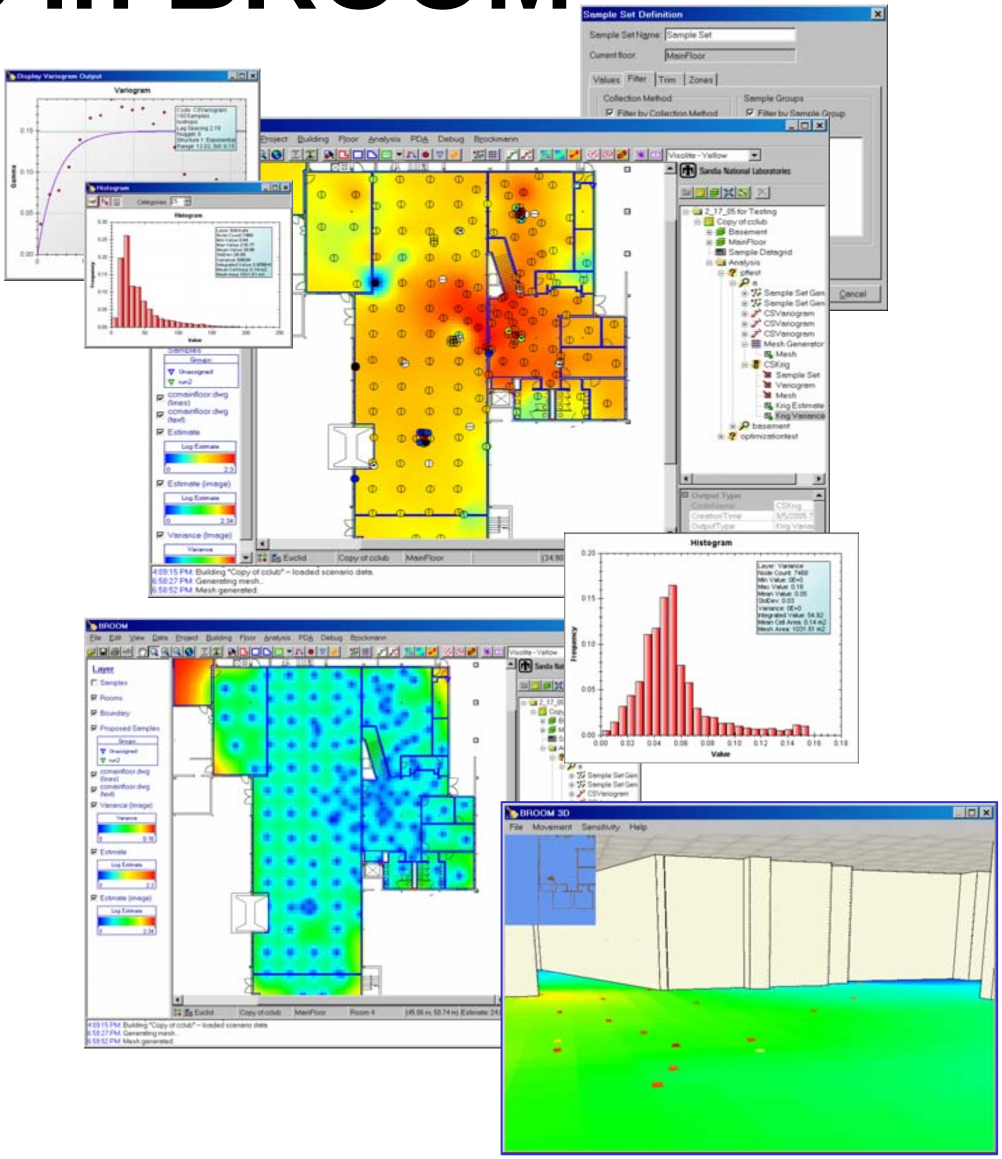
- **Buildings contaminated by chem/bio events need to have capability restored, <u>quickly</u>**
- **Notable Examples:**
  - **Hart Senate Bldg., Postal Facilities, AMI Bldg.**
- **Major Questions**
  - **Where did the contaminant originate and where is it now?**
  - **After decontamination, how do we have confidence that it is safe to reoccupy the building**

Motivation

- **Characterization Goals:**
  - **Define magnitude and extent of contamination**
- **Clearance Sampling Goals:**
  - **Provide confidence that building has been restored for safe use**
- **Constraint:**
  - **Do it quickly**
- **Here, examine *data-driven* approach to mapping**
  - **Contrary to *physics-based* numerical modeling approach**

Spatial Analysis in BROOM*

- Maps (2-D & 3-D) provide estimates of contaminant levels at unsampled locations and provide useful patterns for forensic analysis.
- Variance maps show level of confidence of contaminant estimates and provide basis for adaptive sampling plans based on variance reduction.
- Wide dynamic range: log and indicator transformation of sample sets.
- Integrated mass calculations give estimate of quantity of material released.
- Ability to incorporate effects of walls and doors into mapping without use of a CFD model



*BROOM = Building Restoration Operation Optimization Model

Geostatistics: Estimation

- **Basis of estimation is the simple kriging (SK) system**
  - **Minimum variance**
  - **Unbiased estimates**
  - **Linear combination of surrounding data**

$$Z_{SK}^{*}(x_o) = m + \sum_{i=1}^{N} \lambda_i (Z(x_i) - m)$$

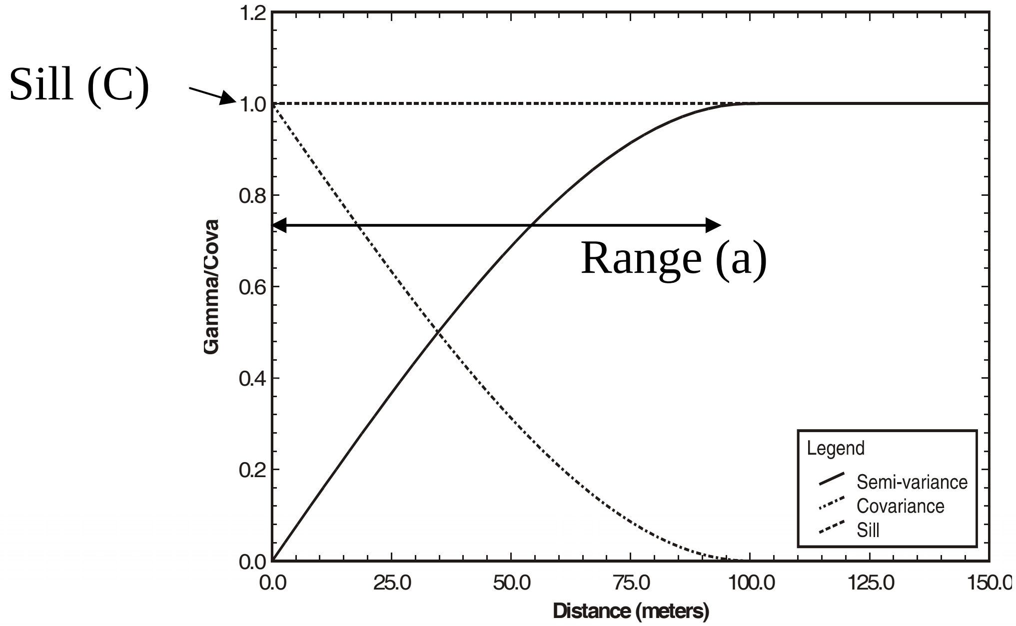
$$\sigma_{SK}^{2}(x_0) = Cov(0) - \sum_{i=1}^{N} \lambda_i Cov(x_0, x_i)$$

λ= kriging weight

m = global mean concentration

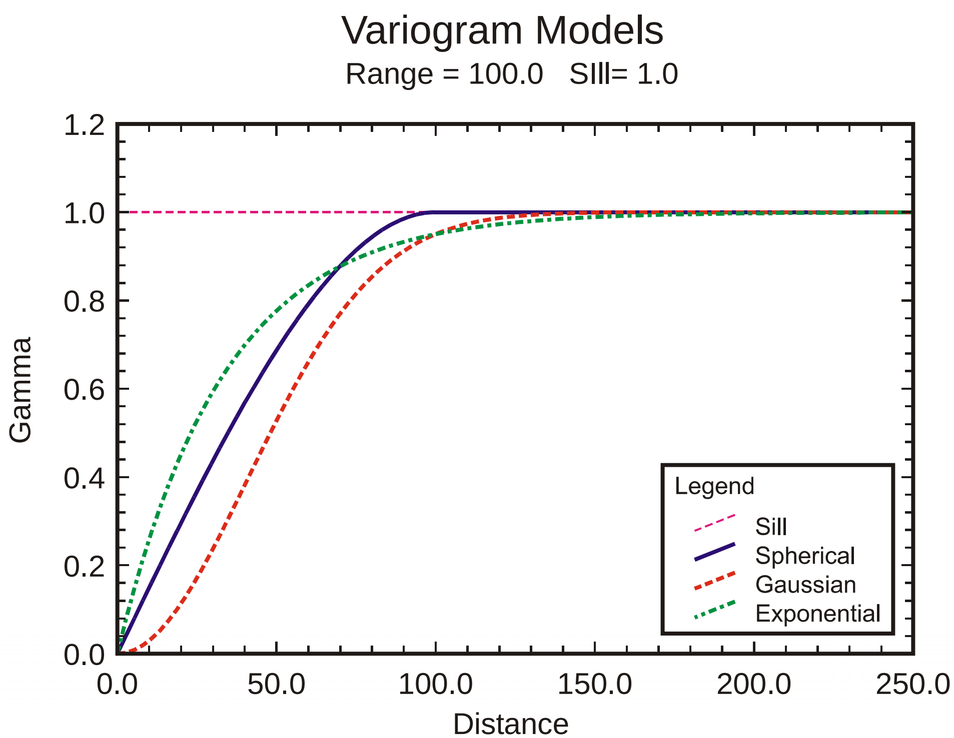
Geostatistics: Spatial Correlation

- Typically measure increase in variability as a function of sample separation (variogram)
  - The complement of spatial covariance under assumption of 2nd order stationarity



Variogram Models

Three commonly used varigoram models are examined here



Spherical

$$\mathbf{h<a:}\quad \gamma(h) = C \cdot \left[1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3 \right]$$

$$\mathbf{h \geq a:}\quad \gamma(h) = C$$

Exponential

$$\gamma(h) = C \cdot \left[1 - e^{-\frac{3h}{a}} \right]$$

Gaussian

$$\gamma(h) = C \cdot \left[1 - e^{\left(-\frac{(3h)^2}{a^2}\right)} \right]$$

Ordinary Kriging

C • λ = D

$$\begin{bmatrix} C_{11} & C_{12} & \vdots & C_{1N} \\ C_{21} & C_{22} & \vdots & C_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ C_{N1} & C_{N2} & \vdots & C_{NN} \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \end{bmatrix} = \begin{bmatrix} C_{01} \\ C_{02} \\ \vdots \\ C_{0N} \end{bmatrix}$$

Unique solution to SK system requires that covariance matrix be positive definite

Solve for weight vector

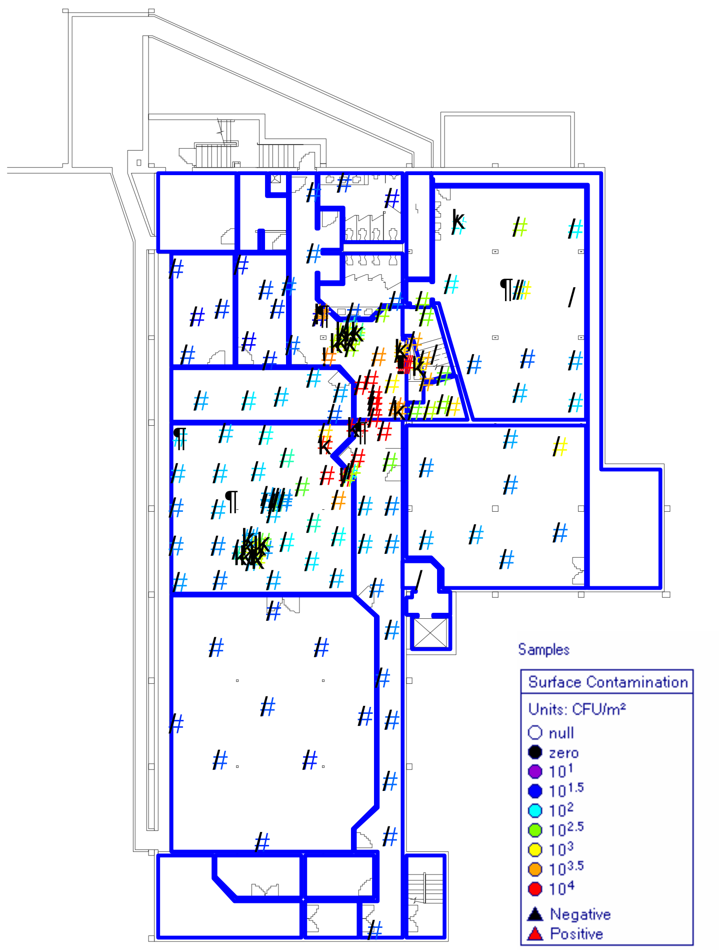
$$\lambda = C^{-1} \cdot D$$

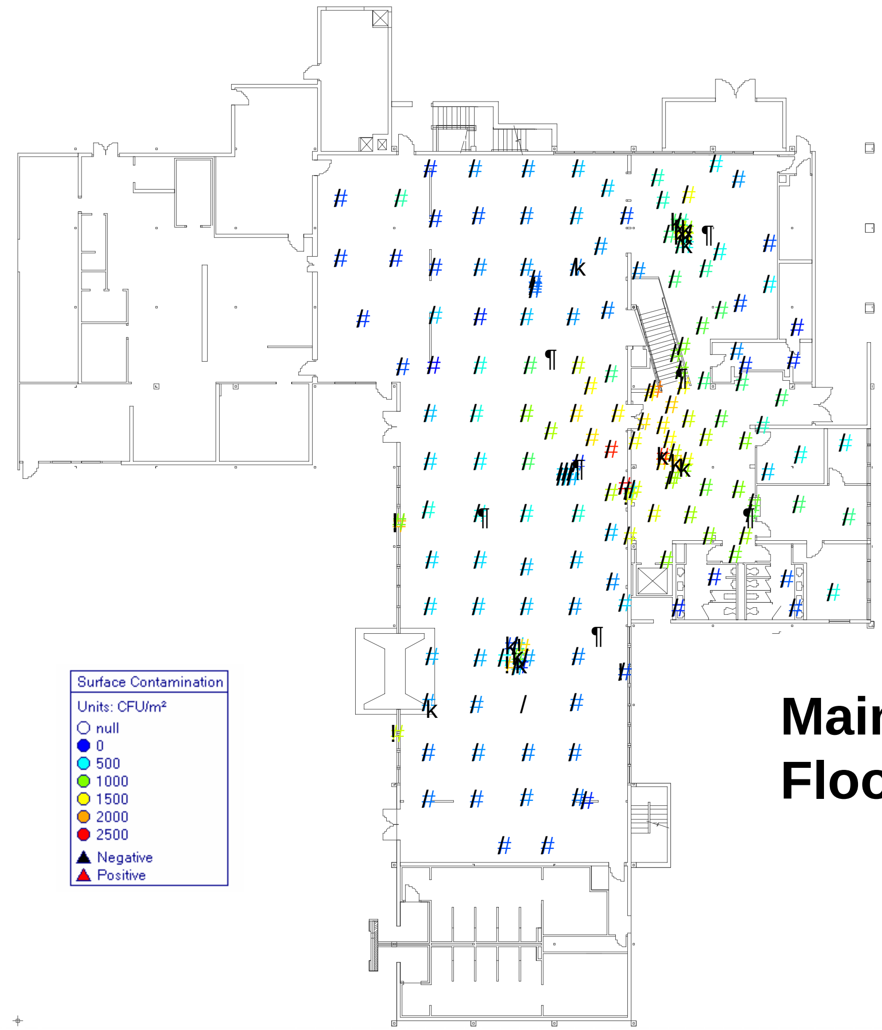
Coronado Club Testing

- Two different visiolite tracer tests: "yellow" and "pink"
  - Focus here on yellow results

Coronado Club Sampling Design

Basement





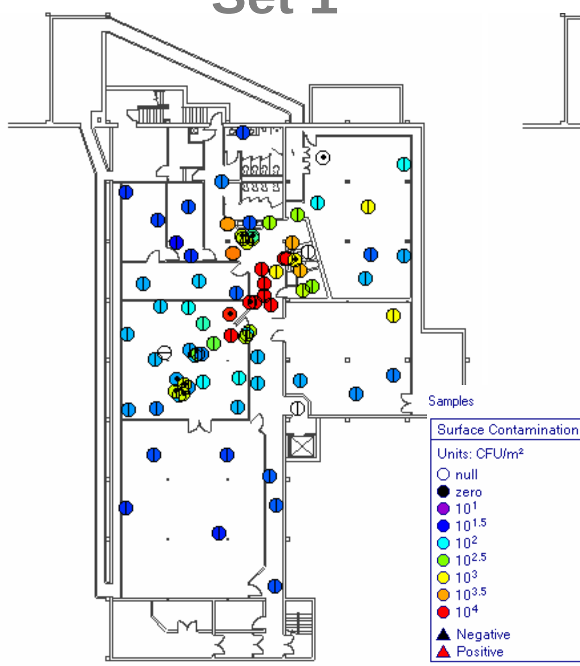
Main Floor

Limit presentation of results to basement

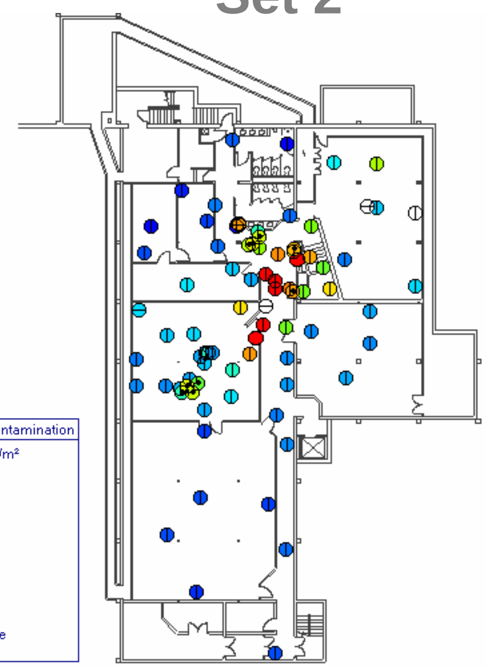
Jacknife Analysis

Random split of data into two sets with 65 samples each. Each set is used for estimation and results are compared against measured values in other set

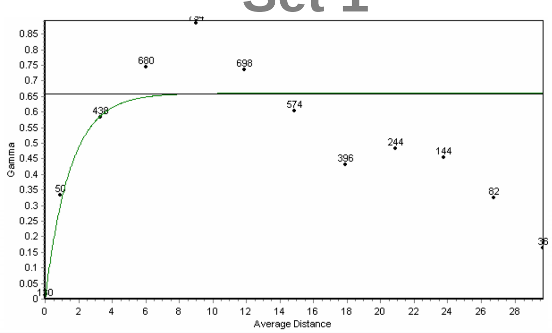
Set 1



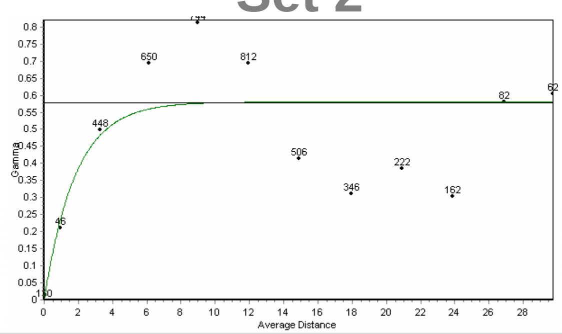
Set 2



Set 1



Set 2

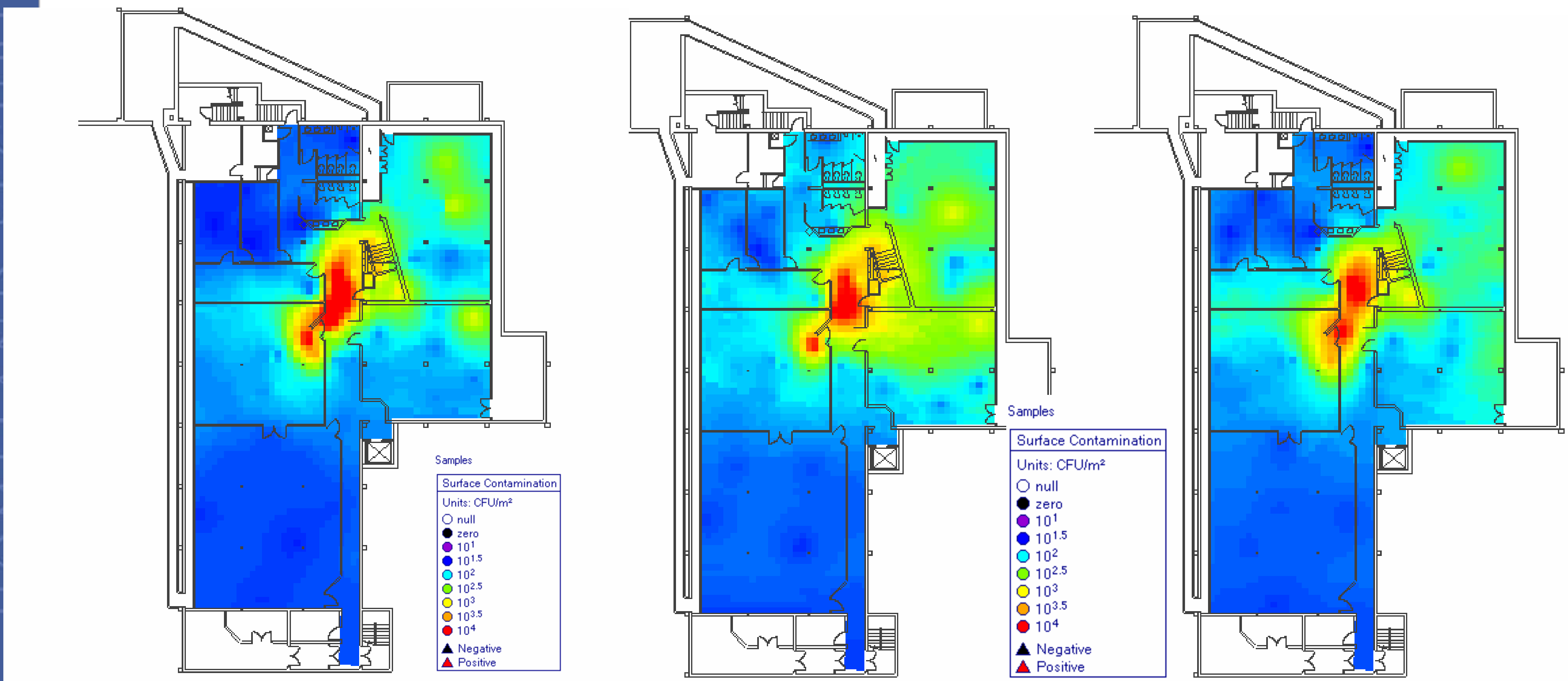


Exponential variogram models fit to each data set with ranges of 4.5 and 5.5 meters

Coronado Club Estimation

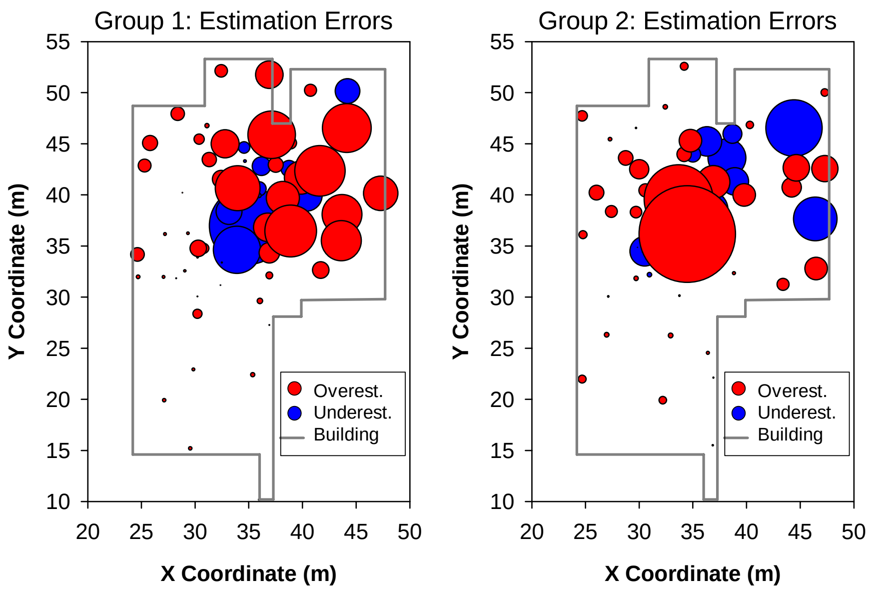
Kriging estimates of basement concentration (color scale in log10 space)

All Data | Data Set 1 | Data Set 2



Coronado Club Cross-Validation

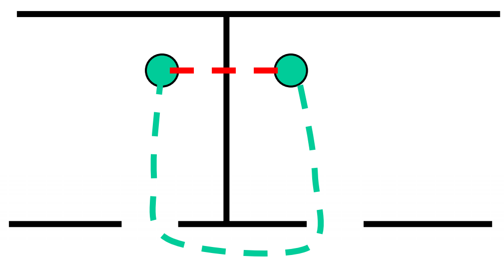
Cross validation results of estimation with log10 data. Largest errors in both sets are near room boundaries (walls). Relatively unbiased estimates for both sets (mean and median near zero)



Parameter	Group 1 Estimation Errors	Group 2 Estimation Errors
Mean	0.14	0.07
Std. Dev.	0.40	0.48
Median	0.08	0.09
Minimum	-1.39	-1.16
Maximum	0.93	1.73
Number	65	65

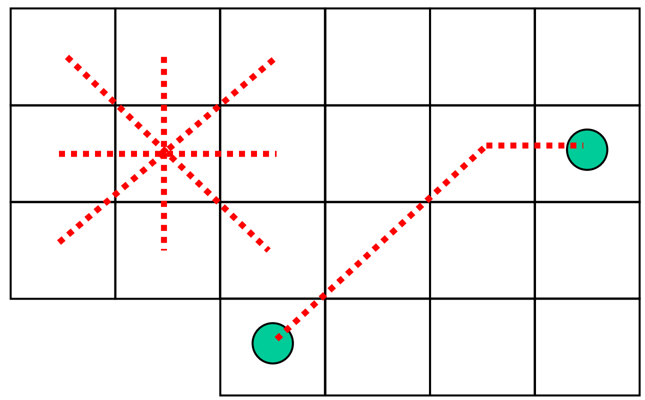
Improved Mapping Approach

- Variogram/covariance is built using differences between sample values separated by a straight line distance
- Building architecture often precludes straight-line paths between samples
- Can non-Euclidean distances improve the mapping process?



Calculating Non-Euclidean Distances

- Grid-based approach to distance calculation



- Dikstra's algorithm

- Also examining other approaches to distance calculation

Positive Definite: Geostats Spin

- **Solution of linear SK system must exist (non-singular) and be unique**
- **Positive definite definition:**

$$\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j\phi(\mathbf{x}_i\mathbf{x}_j) \geq 0$$

$\boldsymbol{x}$'s are coordinate vectors of points in an n-dimensional Euclidean space

- **For the SK system:**

$$\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j Cov(\mathbf{x}_i\mathbf{x}_j) \geq 0$$

$$\sigma^2(\mathbf{x}_0) = \sum_{i=0}^{m}\sum_{j=0}^{m}\lambda_i\lambda_j Cov(\mathbf{x}_i\mathbf{x}_j) \qquad \text{where } \lambda_0 = \text{-1}$$

Positive Definite: Linear Algebra Spin

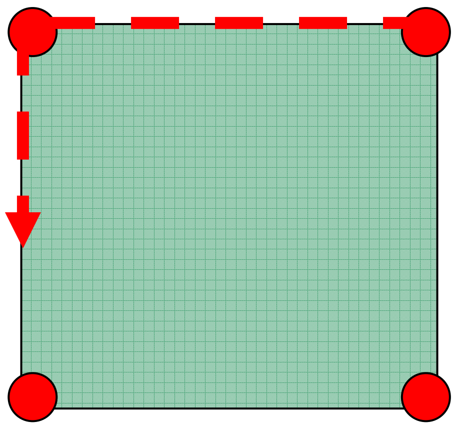
- **A symmetric matrix is always positive definite if all eigenvalues are positive**
- **The determinant of a positive definite matrix is always positive (i.e., always non-singular)**

$$\begin{vmatrix} C_{11} & C_{12} & : & C_{1N} & C_{1N} \\ C_{21} & C_{22} & : & C_{2N} & C_{2N} \\ : & : & : & : & : \\ C_{N1} & C_{N2} & : & : & : \\ C_{N1} & C_{N2} & : & : & C_{NN} \end{vmatrix} > 0$$

Main difference is that this criteria only applies to the single covariance matrix where the positive variance criteria applies to every estimation point

Positive Definite Test

- **When do non-Euclidean distance measures produce a positive definite covariance matrix?**



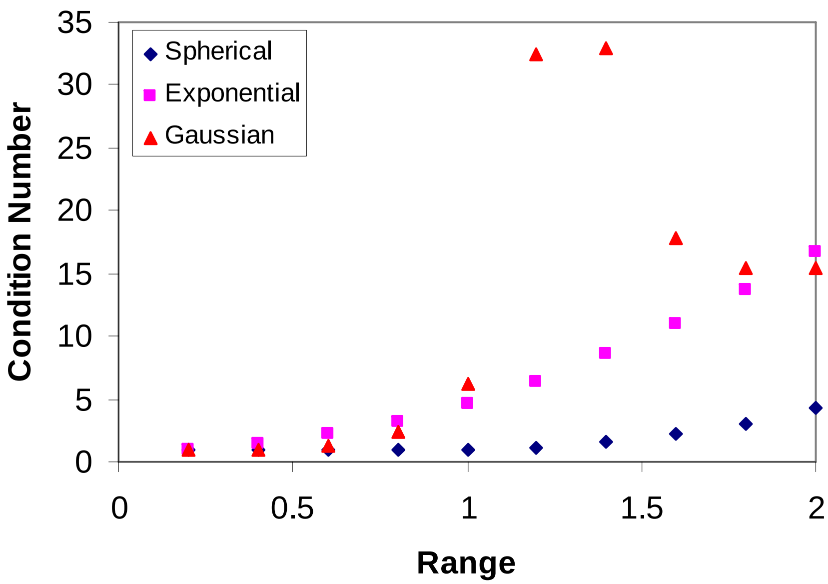
Classic test: City-Block distances around a square for different λ/side values.

This distance measure fails the positive definite test for some covariance functions (see Curriero working paper, 2005)

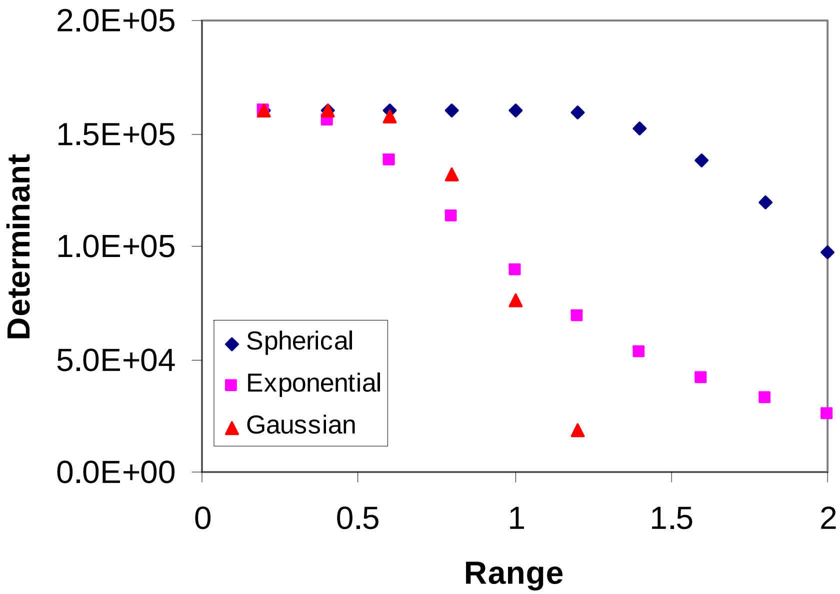
There is no analytical approach to a priori determine whether or not a given non-Euclidean distance measure will produce a positive definite covariance matrix

Positive Definite Test

- **Results for city-block distance test with 3 different covariance functions**



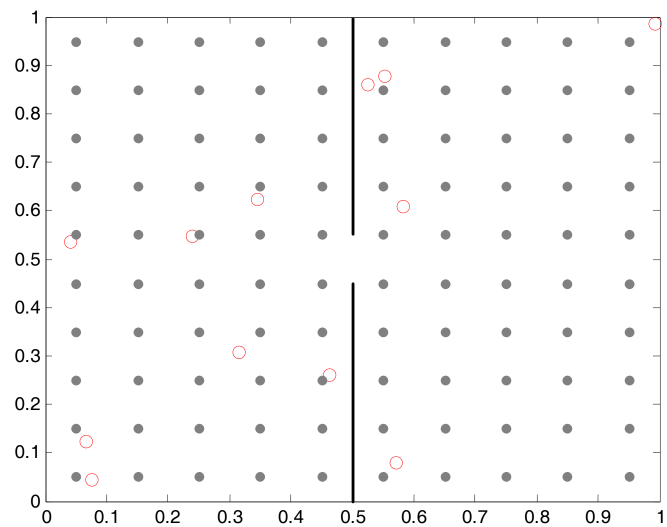
Condition numbers deviate from 1.0

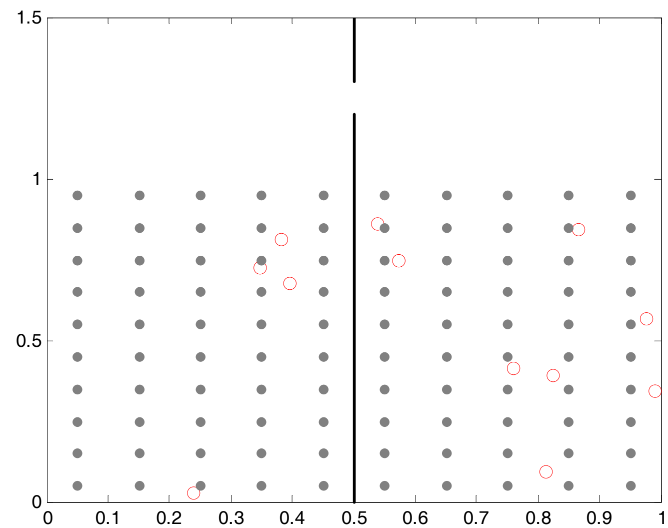


Determinant of Gaussian covariance matrix goes negative for range values > 1.2*side

Simulation Approach

- **Identify geometries encountered within building and simulate to identify when kriging can be applied**
  - **Vary height & width of passage between two rooms**





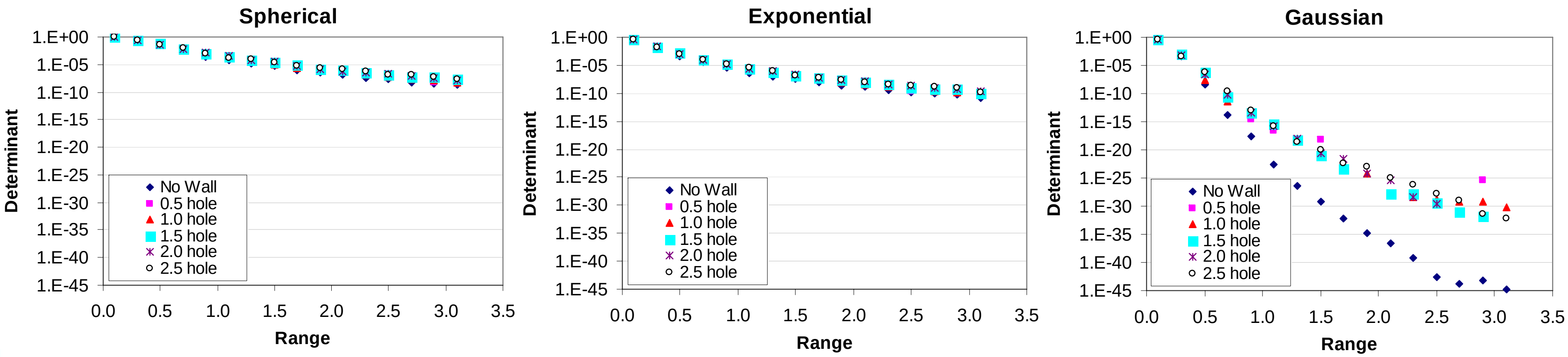
Red are randomly located sample data

Grey is the kriging estimation grid

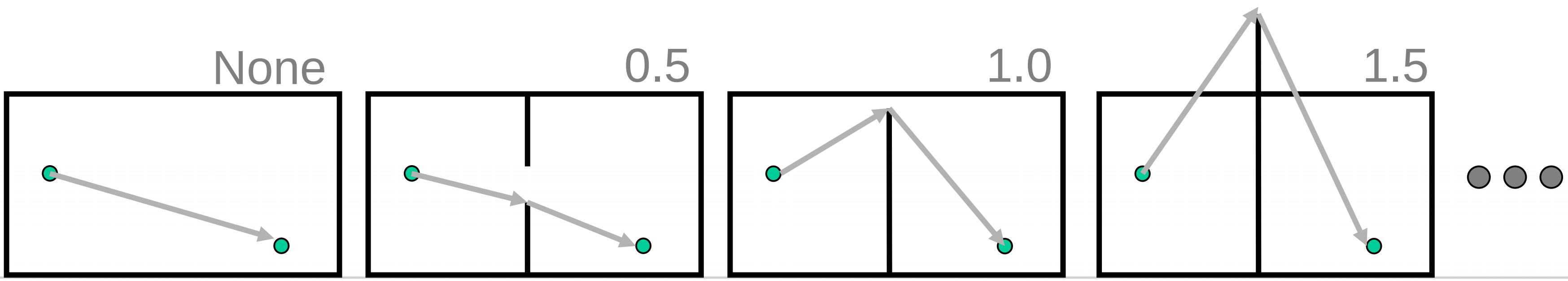
How to account for all possible building and data configurations?
Simulation test must be run on the full building architecture

Simulation Results

- **Effect of covariance function is stronger than that of geometry**
  - **Plots show average determinant over 50 realizations for each range and doorway geometry**

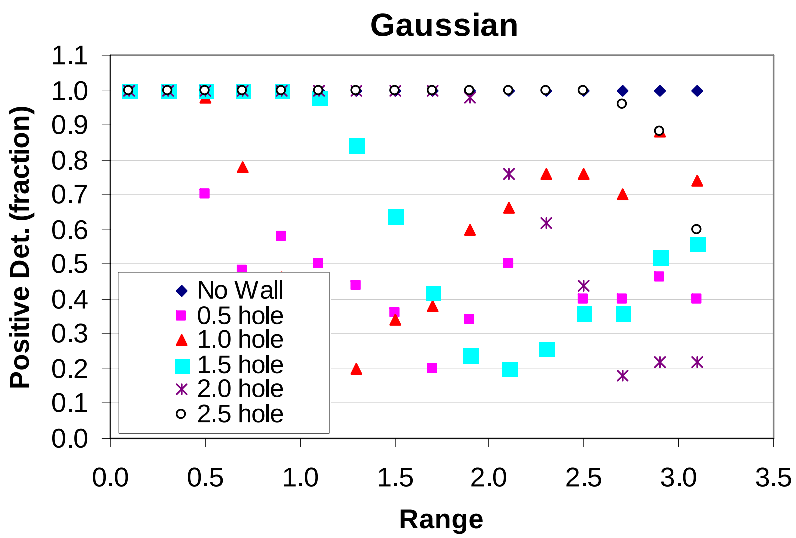


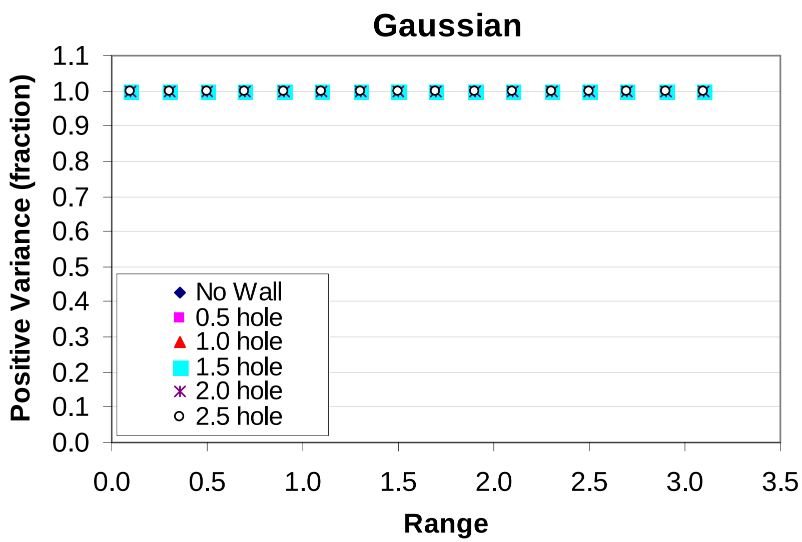
Only Gaussian function results have negative determinants



Positive Definite Criteria Comparison

- **Both criteria show that for all combinations tested, only the Gaussian covariance function produces results that fail**

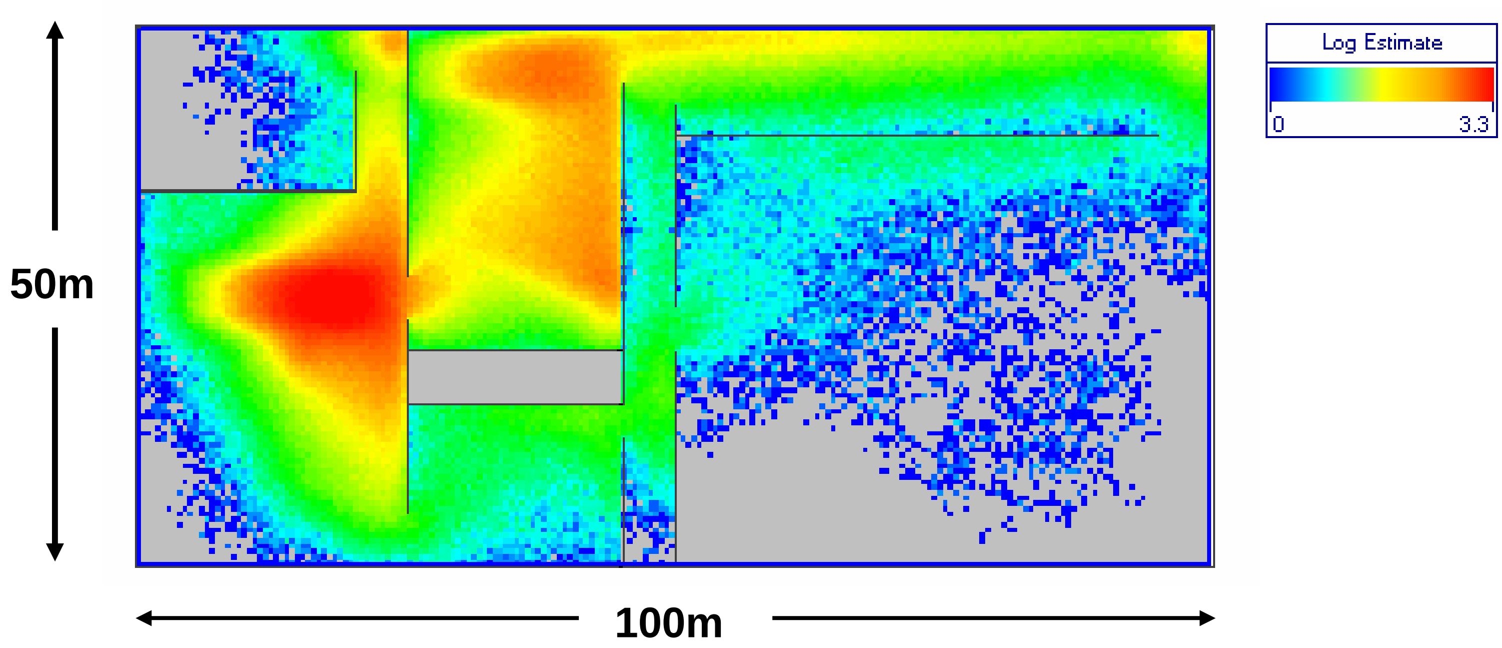




Considerable proportion of negative determinants can still produce positive kriging variance values – use negative determinant as criteria for further calculations

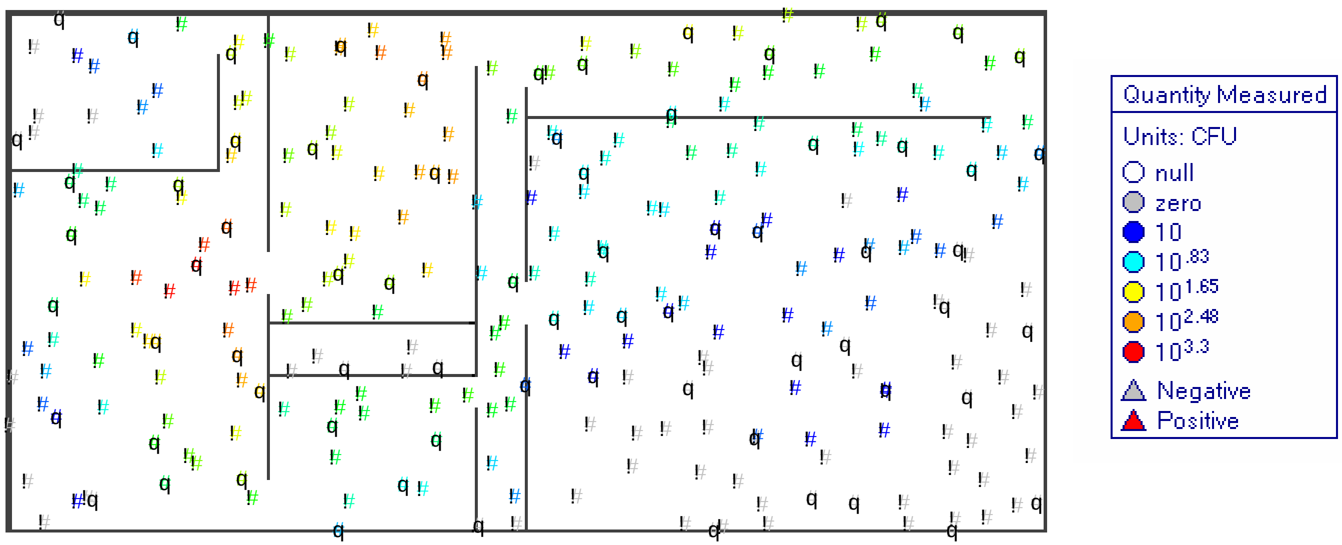
Hypothetical Example

- **Hypothetical contaminant release in test building (ground truth)**



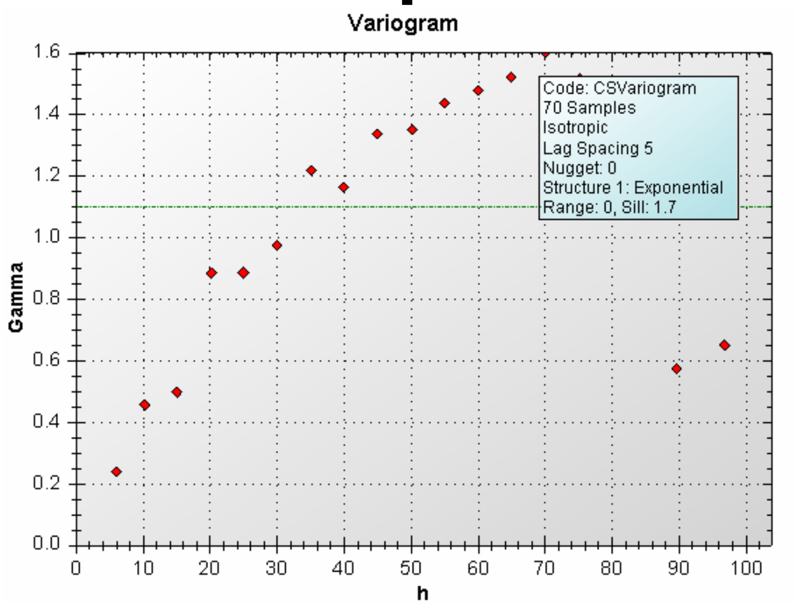
Example Sample Data

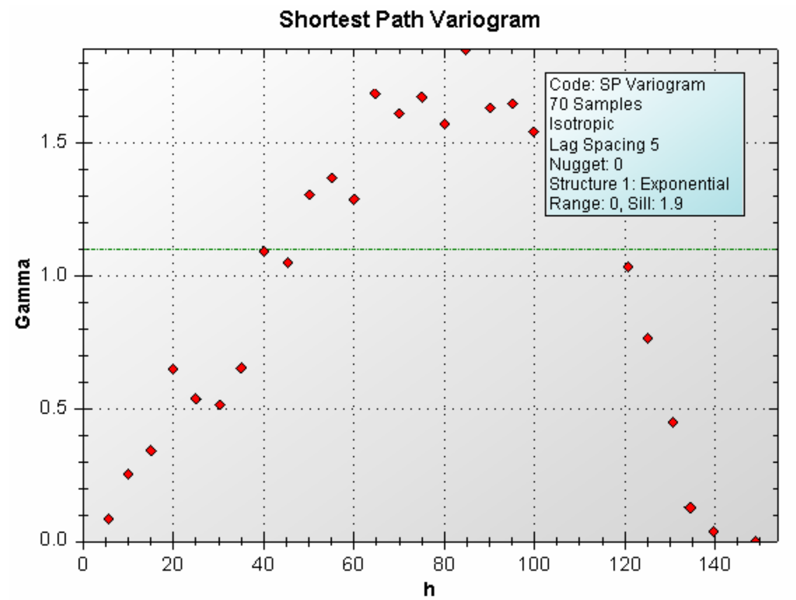
- **270 samples split into two groups of 200 (o's) and 70 (x's)**



Spatial Variation

- **70 Sample Data**

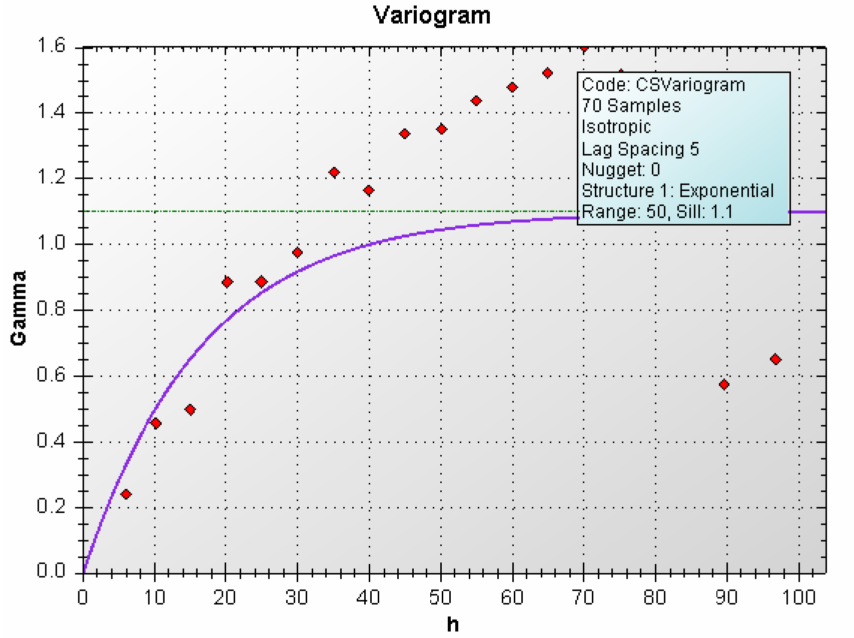


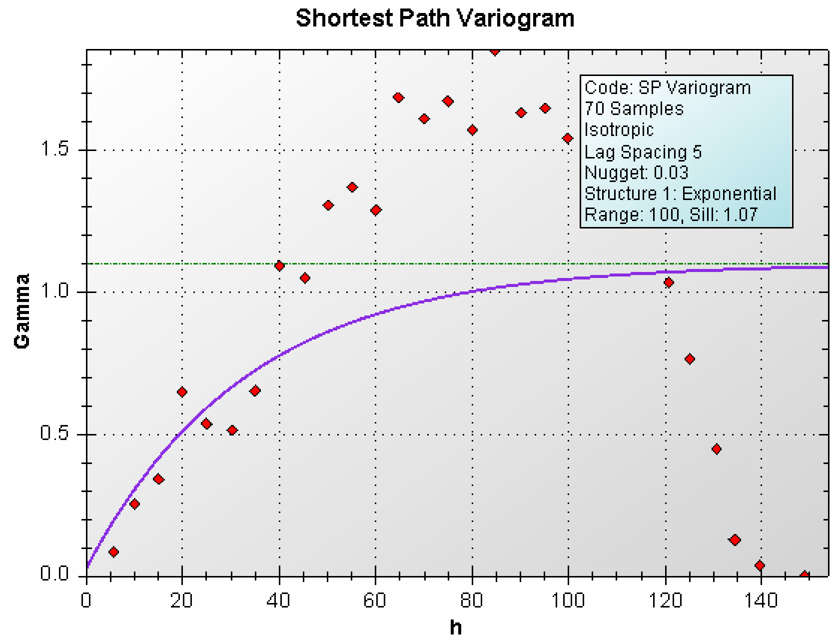


Effect of rooms lengthens distances so that maximum variance occurs from 60-100 meters vs 55-75 meters for Euclidean case, initial sill occurs at 20 meters in both cases –characteristic room size, but is more pronounced for non-Euclidean due to better separation of disparate values

Variogram Models

- **Variograms fit to 70 sample data, Exponential models**

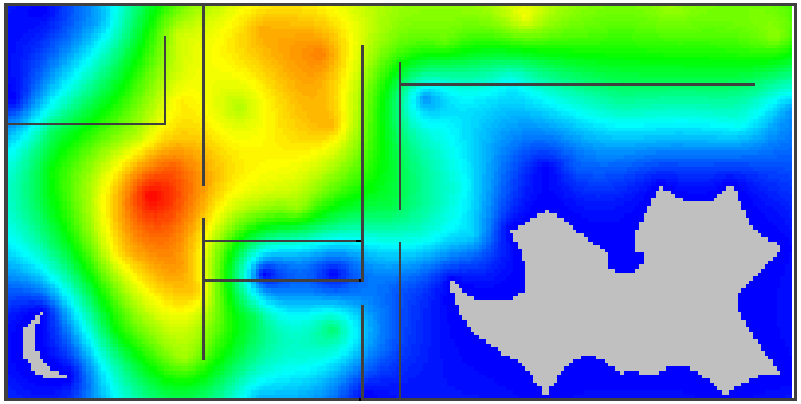




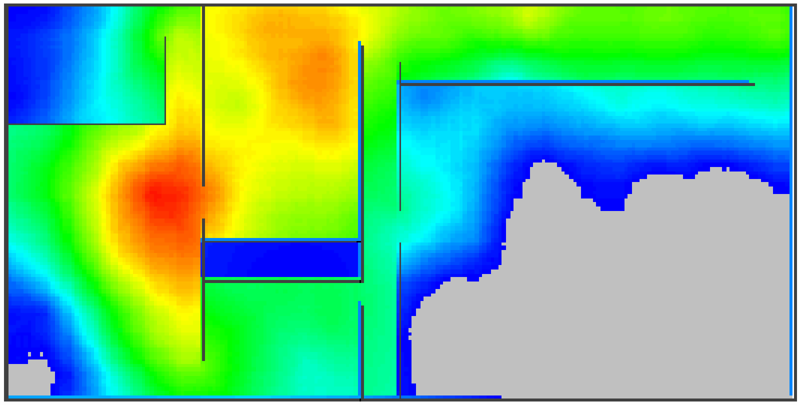
Kriging Estimates

- **Built with 70 samples**

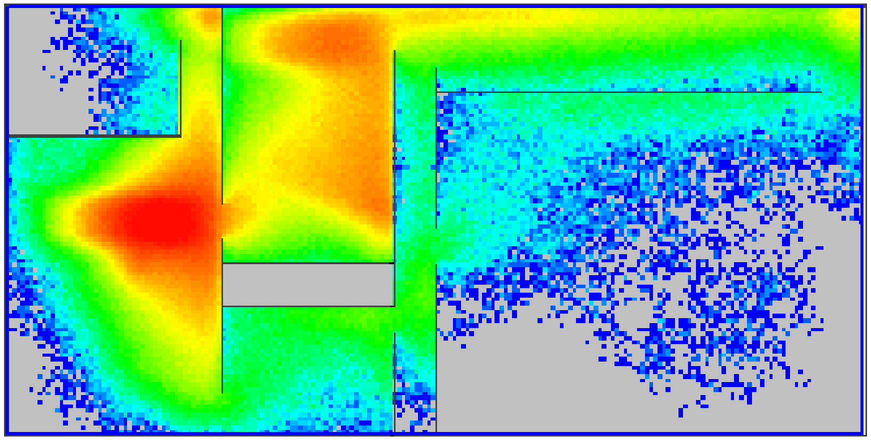
Euclidean Distances



Shortest-Path Distances

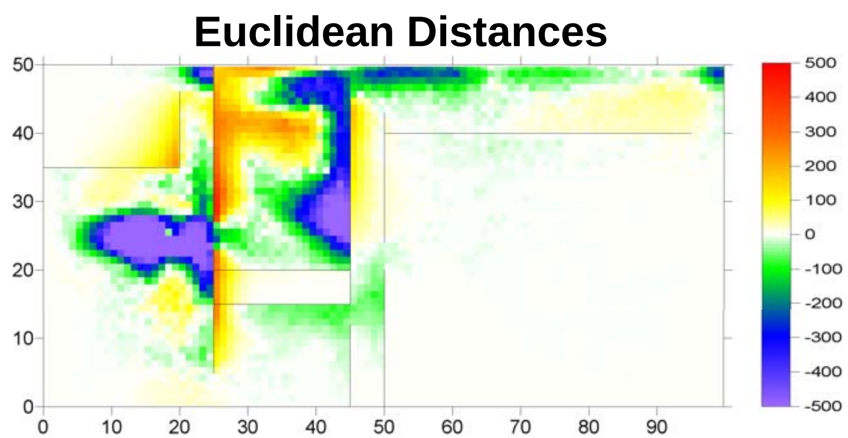


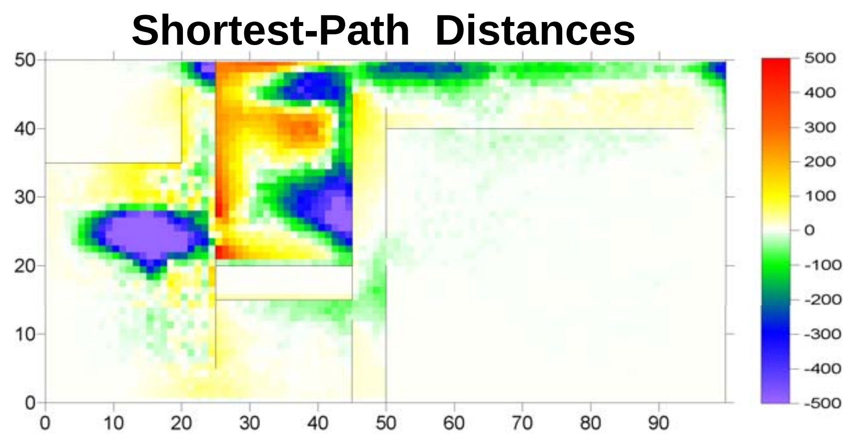
Ground Truth



Estimation Comparison

- **Comparison is done on raw concentration data, (error = estimate – true value)**





<u>Error Statistics:</u>

Mean: -25.0

Std. Dev.: 134.3

Median: 0.98

<u>Error Statistics:</u>

Mean: -19.9

Std. Dev.: 128.8

Median: 0.53

Summary

- Spatial statistics tools provide efficient and accurate means of mapping magnitude and extent of contamination
- Accurate representation of building geometry requires distance modification
  - Issue of what distance measures create positive definite kriging solutions is difficult
- Future work on stronger ties between data-driven models and physics-based models