Air Quality Monitoring Network Design Using Pareto Optimality Methods for Multiple Objective Criteria

Paul D. Sampson [1]

with

Peter Guttorp [1] & **David M. Holland** [2]

[1] NRCSE, University of Washington

[2] U.S. Environmental Protection Agency

supported by a grant from the U.S. EPA

http://www.nrcse.washington.edu/

Outline

1. Spatial Monitoring Network Design Objectives in the Statistics Literature
2. Air Quality Monitoring Objectives
3. Multiple Objective Monitoring Network Design
   A. References to selected applications
   B. Pareto Optimality approach
4. Summary

Three components of an optimal spatial design problem (D.L. Zimmerman, Optimal spatial design.

In: *Encyclopedia of Environmetrics*)

1. Specification of a design space of candidate sites (finite, or in principle, continuous spatial domain).
2. Specification of a model for the existing observations (if any) and the potential observations at candidate sites.
3. Specification of **an** optimality criterion.

National and Regional Surveys

Program names link to pages with maps, measurements, and other program information.

Table of Contents

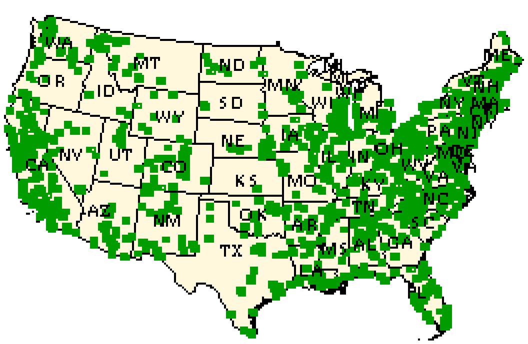
http://www.epa.gov/cludygxb/programs/index2.html

National

And

Regional

Surveys



http://www.epa.gov/cludygxb/programs/namslam.html

NAMS/SLAMS- National Air Monitoring Stations/ State and Local Air Monitoring Stations

Program Name (Acronym)	NAMS/SLAMS
Agency	EPA,State &loc agen ow
Year Initiated	1979
Measures	AIR-criteria pollutnts, visibility/fine particulates, toxics
Collection Source	
Point	Yes
Source	No
Transect	No
Other area	No
Locations for Data Collection	5000 samplrs,3150 sites
Temporal Interval	Hourly,Pb&PM10 variable
Sampling Design	**Selected**
Data Available	Yes
Accessible	EPA reg offices,AIRS
Extent for Reporting	Primarily urban,some rura
Annual Funding	FY 96-$36 M in fed fnds
Cost per Site for Installation	$5 K-$100 K per site
Cost per Site for Op. & Mgmt	Avrg $1 K per site/yrly
Partners	
International	No
Agency	EPA Regions
State	State agencies
Local	Local agencies,contrctrs
Authorities/Reason for Running Prg.	40CFR58
Users of Data per Year	450
Program Meets Metadata Standards	No
Expansion of Prog (Needed/Not)	Not needed
Contact Person	David Lutz
Phone #	919-541-5476

1. Spatial Monitoring Network Design Objectives:

Frameworks for optimal spatial design:

1. **Exploratory, random sampling, and/or space-filling designs.**
   Nychka, D. and Saltzman, N. (1998), Design of air quality networks. In *Case Studies in Environmental Statistics*
2. **Designs for estimating a regression function in a spatially correlated field (with known spatial covariance):**
   W.G. Müller (2000, *Collecting Spatial Data: Optimum Design of Experiments for Random Fields*)
   R.L. Smith (NSF-CBMS Lecture Notes, in prep).

3. **Designs for estimation of the spatial covariance or variogram:**
   W.G. Müller and D.L. Zimmerman (1999, *Environmetrics*)
   R.L. Smith (NSF-CBMS Lecture Notes, in prep).

4. **Designs for optimal spatial prediction**, including designs specifically concerned with **assessment of regulatory thresholds**: identification of sites or regions exceeding thresholds (for one or more pollutant measures) and/or the risk or expected cost of misclassifying sites according to a threshold (classifying a "contaminated" site as safe or vice versa).

Criteria for optimal spatial prediction design:

- Average kriging variance (ave sq pred error) over target grid *D*

$$\frac{1}{|D|}\int_D \sigma_K^2(s)\,ds$$

- Maximum kriging variance

$$\max_{x\in D}\{\sigma_K^2(s)\}$$

- Kriging variance for estimation of a spatial average
- Weighted spatial average of kriging variances

(Cressie, Gotway, and Grondana (1990), *Chem Intell Lab Syst*.)

$$w(s) = I\{E[Z(s) > K]\}\cdot I\left[\sigma_K^2(s) > L\right]$$

- Expected probability or cost (loss) for misclassification
- <u>Entropy</u>

Entropy

- Guttorp P., Le N.D., Sampson P.D., and Zidek J.V. (1993). Using entropy in the redesign of an environmental monitoring network. In: *Multivariate Environmental Statistics*, GP Patel and Cr Rao, eds., pp. 175-202.
- Le, N.D. and Zidek, J.V. (1994), Network designs for monitoring multivariate random spatial fields. In: *Recent Advances in Statistics and Probability*, M.L. Puri and J.P. Vilaplana, eds., pp. 191-206.
- Zidek JV, Sun WM, Le ND (2000) Designing and integrating composite networks for monitoring multivariate Gaussian pollution fields. *J Roy Stat Soc C-Applied Statistics* 49: 63-79.

- **<u>Entropy</u>: scenario of adding sites to an existing network**

Uncertainty in the vector of observations on a spatial process is

$$H(X) = E[-\log f(X)/h(X)]$$

$$X = \begin{pmatrix} U \\ G \end{pmatrix} = \begin{pmatrix} X^{rem} \\ X^{add} \\ X^{g} \end{pmatrix}$$

*rem*aining ungauged sites

sites to be *add*ed

current *g*auged sites

The <u>total uncertainty</u> about a future realization and model parameters is <u>fixed</u>:

$$H(X,\theta \mid D) = H(U \mid G,\theta,D) + H(\theta \mid G,D) + H(G \mid D)$$

or

$$TOTAL = PRED + MODEL + MEAS$$

Minimize predictive uncertainty by maximizing **:**

$$H(G|D) = H(X^{add}|X^{g}, D) + H(X^{g}|D)$$

Under a Gaussian model with inverted Wishart prior on the spatial covariance matrix,

$$H(G|D) = \frac{1}{2}\log\left|\Phi_{add|g}\right| + const$$

Where $\Phi_{add|g}$ is the residual (predictive) covariance matrix of X^{add} conditional on X^{g} .

2. Air Quality Monitoring Objectives

Multiple scientific objectives are explicit in current guidelines for air quality monitoring networks. Three perspectives:

1. **Four general purposes for the ambient air monitoring program are** (http://www.epa.gov/oar/oaqps/qa/monprog.html):

- **to judge compliance with and/or progress made toward meeting ambient air quality standards;**
- **to activate emergency control procedures that prevent or alleviate air pollution episodes;**
- **to observe pollution trends throughout the region, including non-urban areas; and**
- **to provide data base for research evaluation of effects: urban, land-use, and transportation planning; development and evaluation of: abatement strategies and diffusion models.**

2. Specific objectives of monitoring sites in the SLAMS network according to U.S. Code of Federal Regulations, Part 58, Appendix D are:

- **to determine representative concentrations in areas of high population density;**
- **to determine highest concentrations expected to occur in the area covered by the network;**
- **to observe pollution trends throughout the region, including non-urban areas; and**
- **to determine general background concentration levels.**

See also EPA guidelines (U.S.EPA, 1998, EPA-454/R-98-002):

- **to determine the extent of air pollution transport into and out of an area.**

3. EPA OAQPS criteria aims for monitoring network assessments:

In general terms: What sampling coverage is required for

1. Compliance: attainment/non-attainment designation?

2. Exposure: to inform the public of exposure risks to the criteria air pollutants.

3. Trends: to evaluate progress on implementation of emission reduction strategies for the criteria pollutants? Progress could be viewed as the estimation of regional trends.

4. Emissions strategy development: to develop emission reduction strategies for the criteria pollutants?

Strategy: ranking of sites based on site-specific evaluation of:

- **Pollutant concentration** (NAAQS criterion values)
- **Estimation uncertainty** (estimation of site from 5 nearest sites by inverse distance weighting scheme (w/ declustering)
- **Deviation from NAAQS** value (sites well above or well below NAAQS ranked low)
- **Spatial coverage**: geographic area defined by Thiessen polygon around monitoring site)
- **Persons/Station**: number of people (sum of census tracts) in Thiessen polygon around monitoring site

Note: This assessment apparently provides site specific measures/rankings only, not any numerical assessment of the network as a whole, and no assessment of spatial estimation beyond the current network.

So, what does this suggest for numerical objective design criteria?

1. **Criterion associated with maps of <u>probability of exceeding standards</u> (**computed how?**)**
2. **Criteria assessing <u>spatial prediction accuracy</u>** (kriging error, entropy)
3. **Utility functions for other criteria:**
   - **network representation of <u>population</u>**
   - **network representation of <u>sources</u>**
   - **<u>Cost</u>**

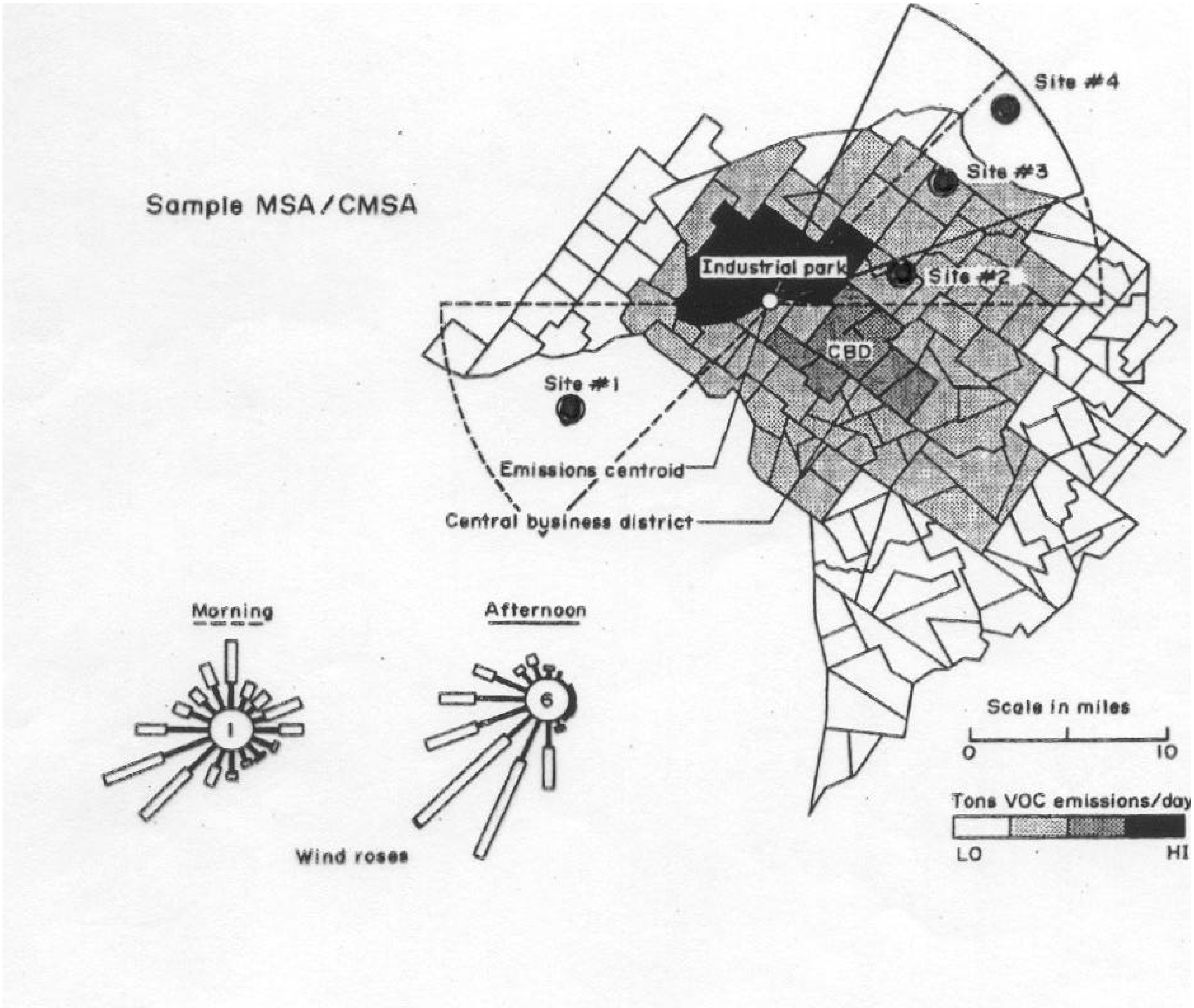
- **Remark:** 1 & 2 above are achievable using recent methods of spatial analysis

- **Many of these objectives assume some form of <u>prior information</u> regarding where:**
  - **people live,**
  - **pollutant sources are, and**
  - **high and background levels of pollutant concentrations are expected.**
- **Information about where high and background concentrations may be expected requires, probably, a <u>combination of available monitoring data and</u> <u>air quality model predictions</u> in network design calculations.**
- **Statistical network design methodology has apparently never been recommended to attempt to meet these objectives.**

From **"Guidance for Network Design and Optimum Site Exposure for $PM_{2.5}$ and PM_{10}"** (EPA-454/R-99-022):

Network Design Philosophies: statistical methods accounting for correlation, model-based methods, random sampling, systematic sampling, judgmental sampling, heterogeneous strategies.

"Monitoring networks for criteria pollutants always use judgmental sampling strategies that consider where source emissions are in relation to populations and which way the wind blows. … Most of this guidance is based on judgmental network design, though it is expected that networks will involve more of the hybrid approach as they are evaluated as future $PM_{2.5}$ measurements and improved aerosol modeling techniques are developed."



Sample MSA/CMSA
Site #4
Site #3
Industrial park
Site #2
CBD
Site #1
Emissions centroid
Central business district
Morning
Afternoon
1
6
Wind roses
Scale in miles
0
10
Tons VOC emissions/day
LO
HI

Table 4-1. Ozone and ozone precursor monitoring site types with corresponding monitoring objectives.

Type of Site	Relevant Pollutants	Monitoring Objective	Spatial-Scale	Notes
Maximum Downwind Concentration	Ozone	Regulatory Compliance	Urban to Regional	This site is required as part of the NAMS network and is designed to measure the maximum ozone concentration due to an urban area.
Maximum Exposure	Ozone	Regulatory Compliance	Neighborhood to Urban	This site is required as part of the NAMS network and is designed to measure the highest concentration in a heavily populated area.
Maximum Emissions	NO_x, VOC	Control Strategy Development	Middle-Scale	This site is designed to measure the concentration of NO_x and VOC in proximity to a source. This data would be used in modeling ozone formation.
Upwind	Ozone, NO_x, VOC	Control Strategy Development	Regional	This site is designed to measure the ozone and ozone precursor concentrations entering an urban area from an upwind source region.
Exposure	Ozone	Data for Health Studies	Neighborhood to Urban	This site provides additional (i.e., more than the maximum exposure site required for NAMS) exposure data for health studies.
Exposure	Ozone	Data for Vegetation Studies	Urban to Regional	This site is used to quantify the exposure of vegetation to ozone to assess the deleterious impact on the vegetation.
Research	Ozone, NO_x, VOC	Research	Middle-Scale to Regional	This site is established for a specific research purpose independent of regulations. Often these sites will operate only temporarily (i.e., during a single summer season).

For all MSAs or CMSAs with more than 200,000 people, two ozone NAMS sites are required, the maximum downwind site and the maximum exposure site. If the MSA/CMSA is a serious, severe, or extreme nonattainment area; up to five sites (PAMS) are required. (Note that PAMS sites are based on population - see 40 CFR Part 58, Appendix D, Table 2.)

[b]EPA requires PAMS VOC monitoring for serious, severe, and extreme ozone nonattainment areas. Up to two of the NAMS monitors may be part of the PAMS program in these areas. Appendix B to this document includes references for additional information on PAMS monitoring.

Additional downwind, exposure, and emission sites can be installed as necessary, or as resources permit. Definitions of the spatial scale and type of monitoring site are found in 40 CFR Part 58, Appendix D, ``Network design for SLAMS, NAMS, and PAMS.''

Multivariate/multi-pollutant monitoring goals:

"Conceptual Strategy for Ambient Air Monitoring," (draft document, www.epa.gov/ttn/amtic/cpreldoc.html) notes that, although many monitoring networks were designed for single pollutants, a multivariate perspective is necessary now as agencies attempt to optimize networks by measuring multiple pollutants whenever practicable. E.g., recommended that: PAMS and CASTNET networks be upgraded to measure NO_x in order to track effects of emission reductions programs; that the resources of the PAMS and air toxics programs be combined to optimally address the objectives of both programs. In particular, it was suggested that PAMS might focus more on VOC trends and NO_x reductions with less emphasis on air quality model evaluation.

➔ **network design for criteria for multiple pollutants**

3. Multiple Objective Monitoring Network Design:

A. References to selected applications in publications

B. Pareto optimality approach

A. Selected applications in publications

Cieniawski, Wayland Eheart, & Ranjithan, 1995. Using genetic algorithms to solve a multiobjective groundwater monitoring problem. Water Resources Research 31(2), 399-409.

Dutta, Das Gupta, & Ramnarong, 1998. Design and optimization of a ground water monitoring sytem using GIS and multicriteria decision analysis. Ground Water Monitoring and Remediation 18(1), 139-147.

Trujillo-Ventura & Ellis, 1991. Multiobjective air pollution monitoring network design. *Atmos Environ* 25A: 469-479.

Example: Trujillo-Ventura A, Ellis JH, 1991. Multiobjective air pollution monitoring network design. *Atmos Environ* 25A: 469-479.

Objectives combined for optimization:

1. "Spatial coverage": spatial integral of kriging estimate of pollutant concentration multiplied by kriging error, emphasizing regions of high estimated concentration.

2. Detection of violation of standards: sum over sites of probability of standard violation (assuming a random field model).

3. "Data validity": measure of spatial regularity of network (optimal for a triangular network)

4. **"Cost":** number of monitoring sites (considered as a constraint rather than an objective criterion)

B. Pareto optimal designs:

Alternative to optimization of a composite design criterion is to investigate the space of “Pareto optimal” designs.

Given a vector of n design criteria, $X_1,\ldots,X_n$, a design having attained numerical criteria values $a_1,\ldots,a_n$, is said to dominate another design attaining values $b_1,\ldots,b_n$, if $a_i \leq b_i$ for all i, and for at least one j, $a_j < b_j$.

A design that is not dominated by any other is said to be Pareto optimal, and the Pareto optimal set or Pareto frontier is the set of all Pareto optimal designs.

Consideration of the Pareto optimal set will allow better understanding (compared with optimization of a single criterion) of the trade-offs necessary to obtain greater relative efficiency on given criteria.

Pomac-Evolve* evolutionary computation program for Pareto optimization:

- "population" of parameterizations = proposed networks
- evaluate performance of each network on a set of criteria
- nondominated sorting to assign parameterization (network) "fitness" ("niched" Pareto genetic algorithm)
- evolutionary iterations select parents (networks) for future generations by cross-over (exploration) and mutation (exploitation), updating current estimate of Pareto Frontier

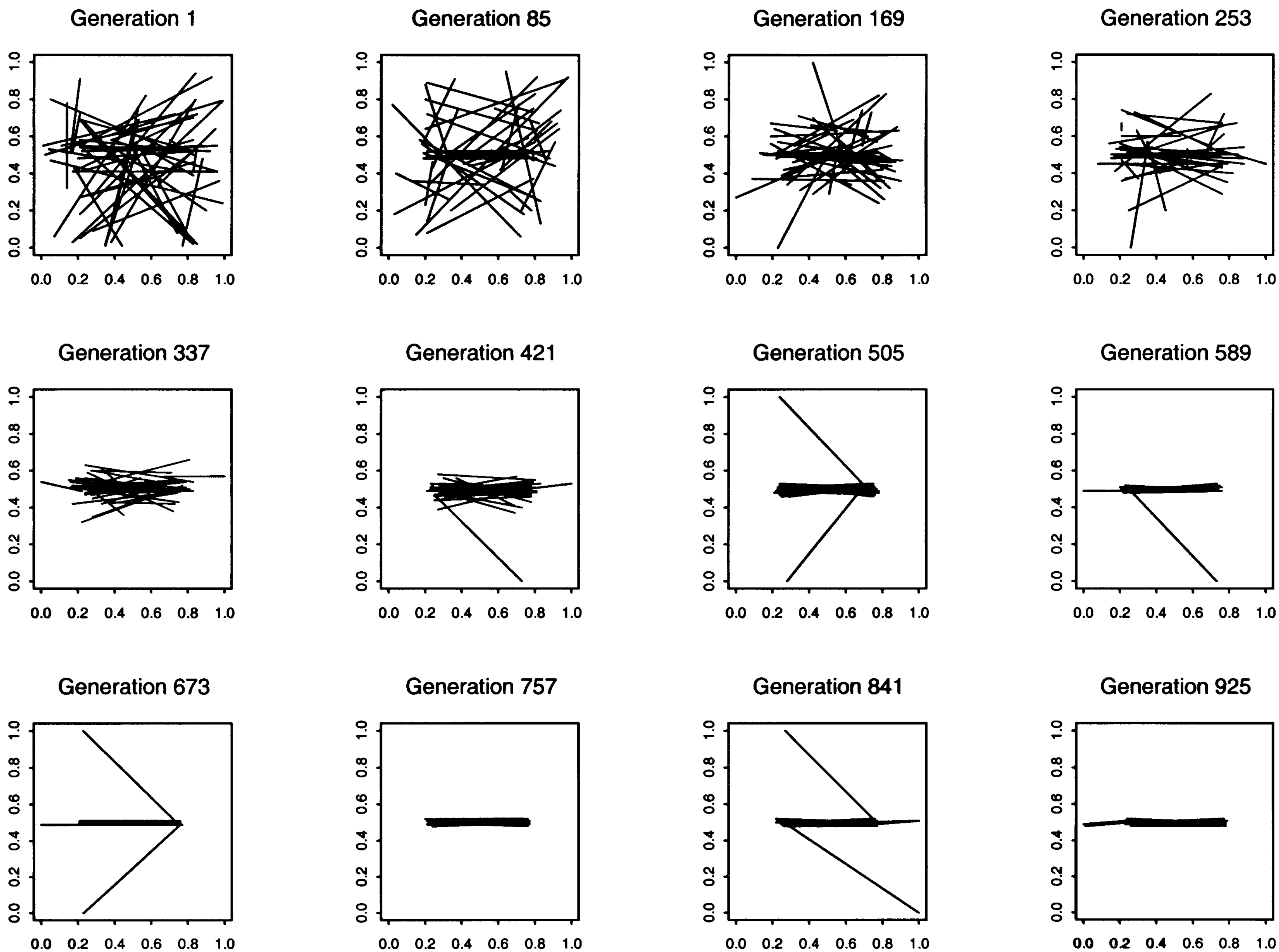
*Reynolds & Ford (1999) Multicriteria assessment of ecological process models. *Ecology* 80, 538-553.

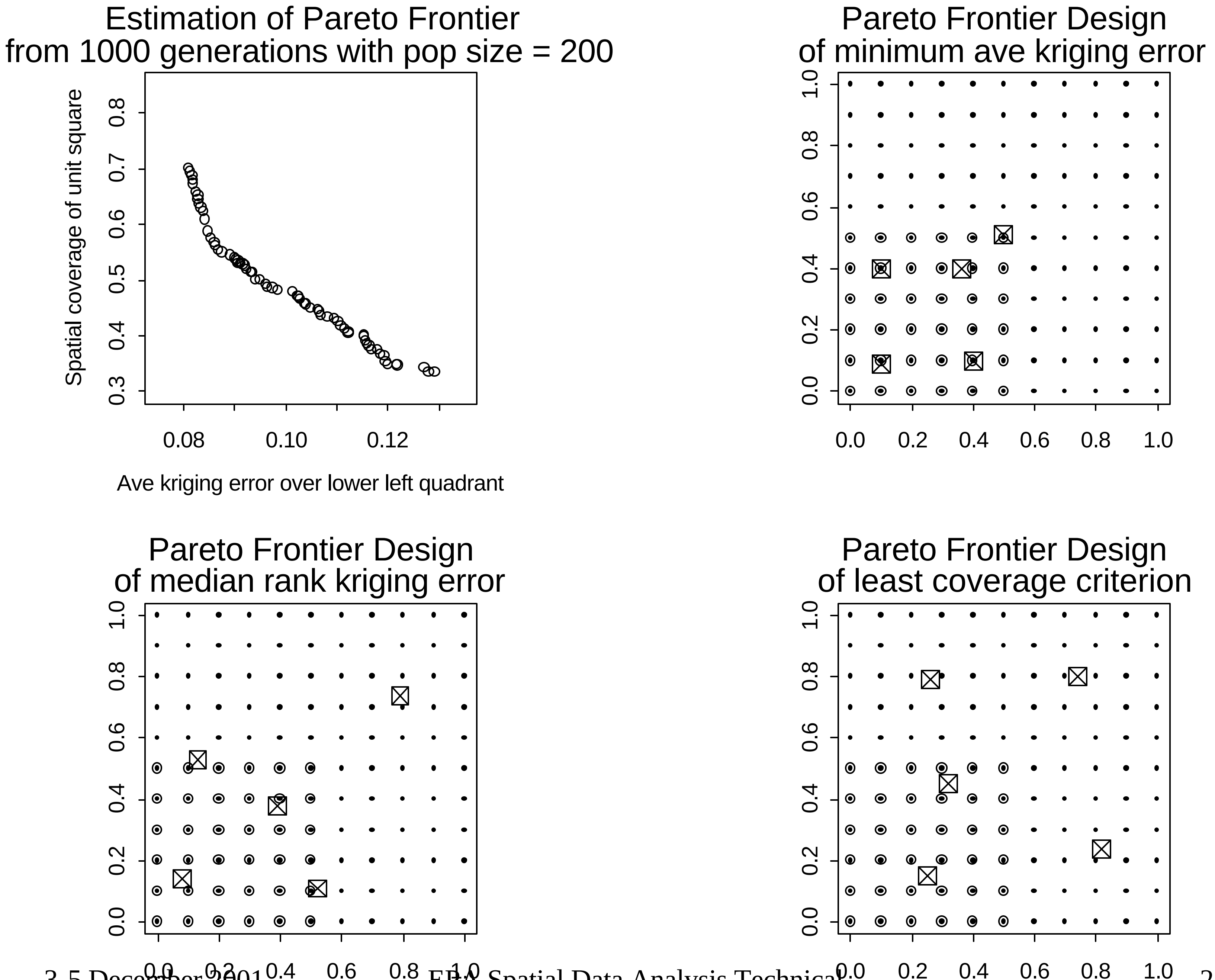
*Ford, Turley & Reynolds (2000) Pareto-Evolve Users Manual: The Pareto Optimal Model Assessment Cycle Using Evolutionary Computation, www.nrcse.washington.edu/software.

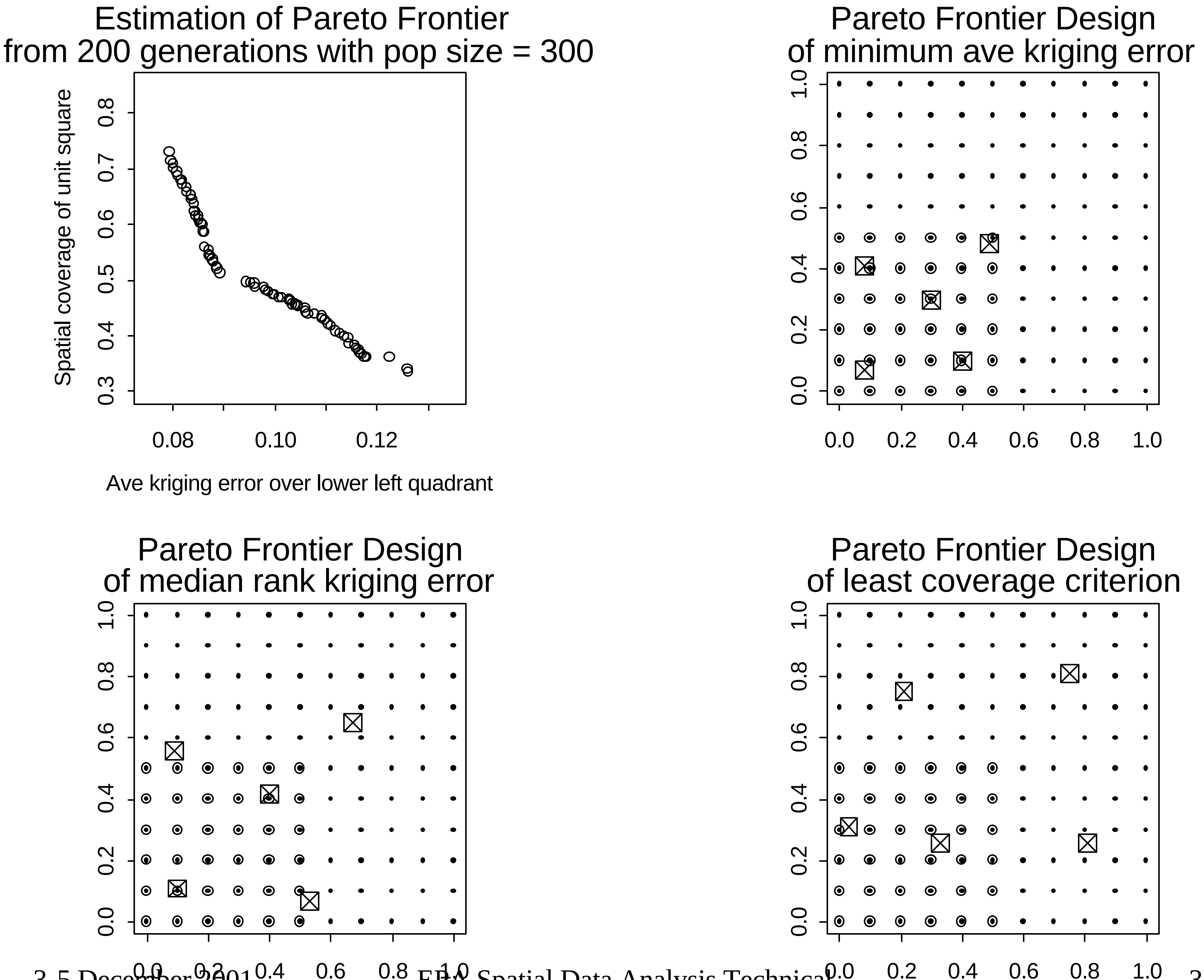
1. **Illustration of evolutionary computation: optimization of 2 sites for a single minimax spatial coverage criterion over a square grid:**
2. **Demonstration of estimation of Pareto Frontier for a 2-criterion problem:**
   - **minimax spatial coverage over the unit square**
   - **average kriging error over lower left quarter of unit square**

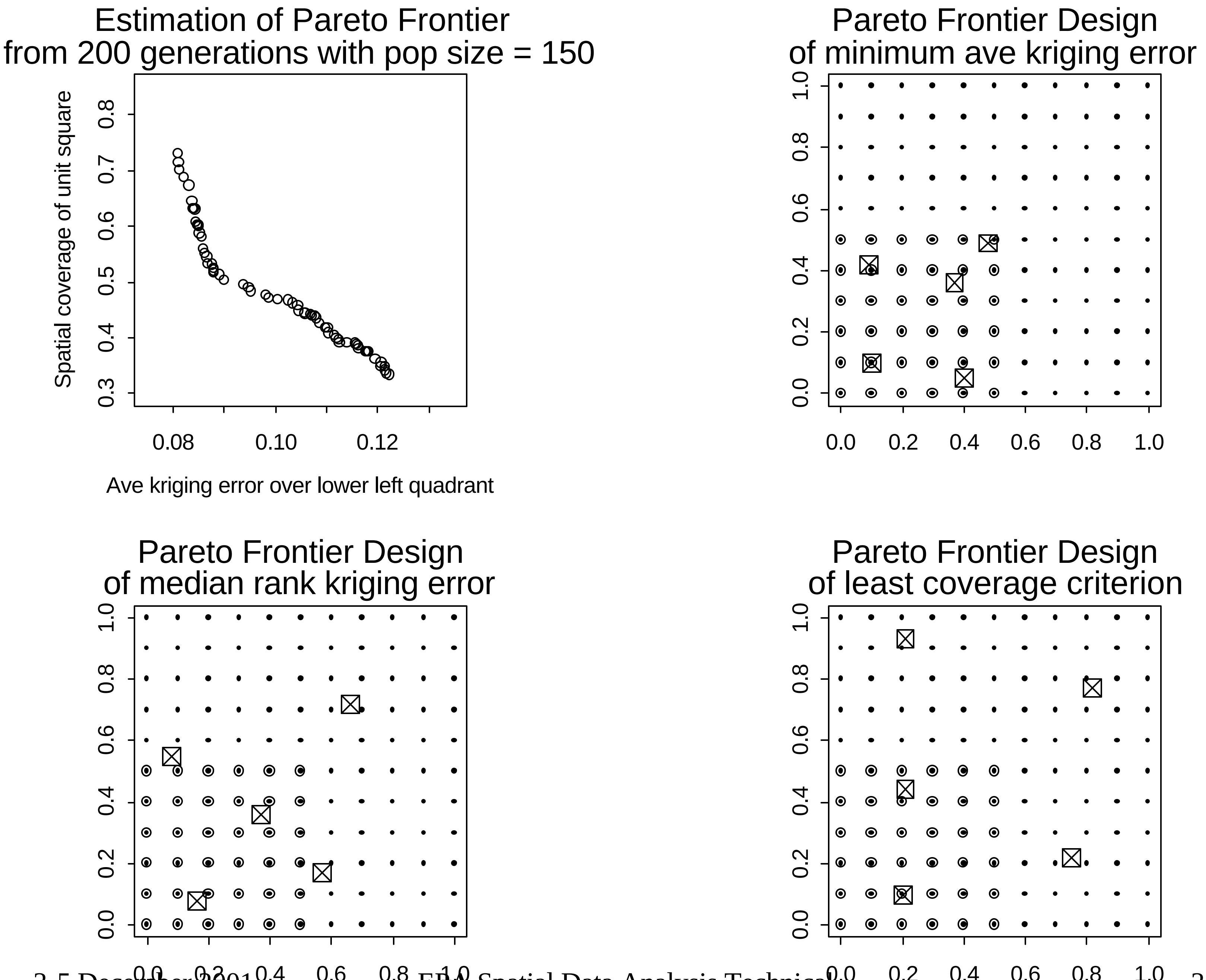
(Varying population size and number of evolutionary generations)

Selected generations for evolutionary computation
optimization of the minimax spatial coverage criterion with 2 sites

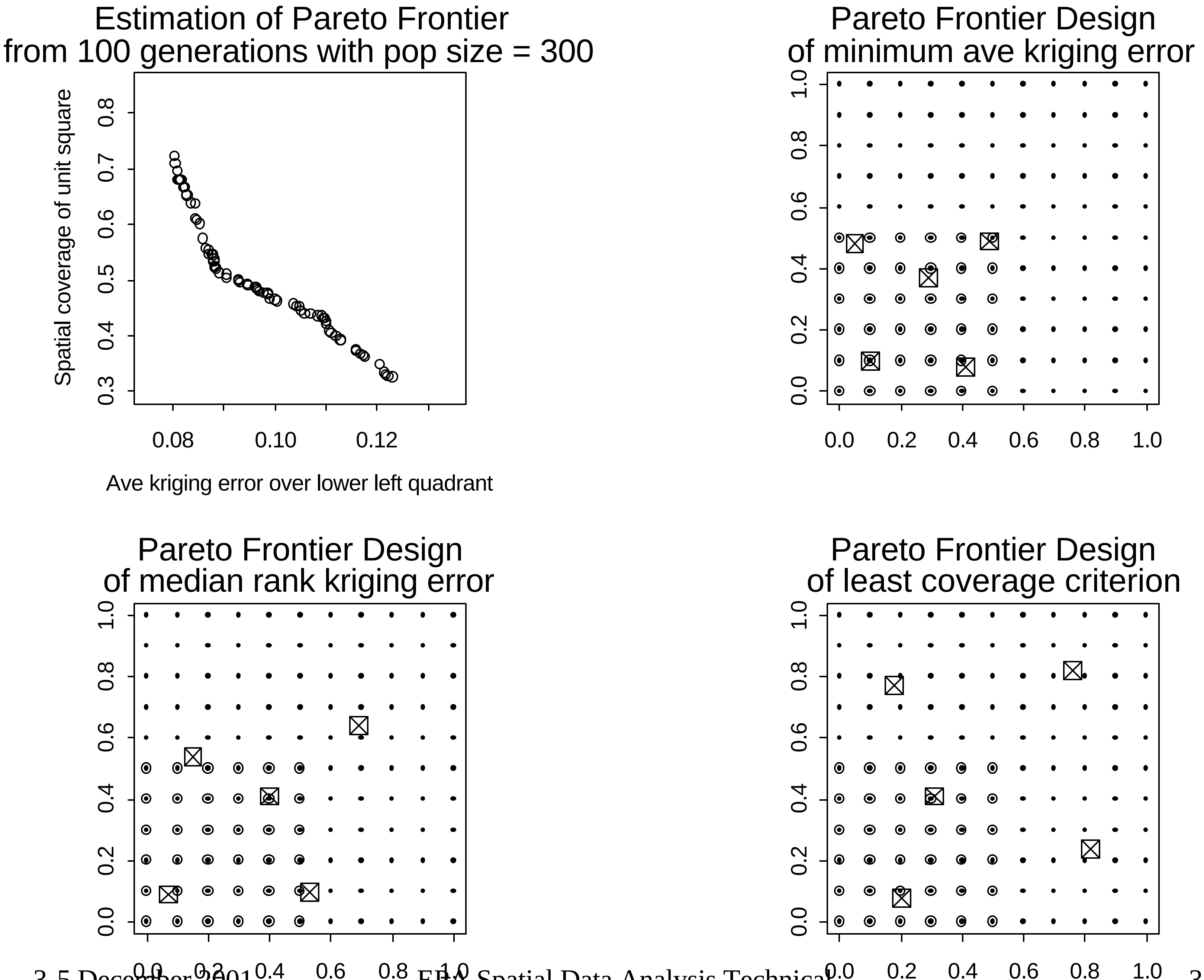


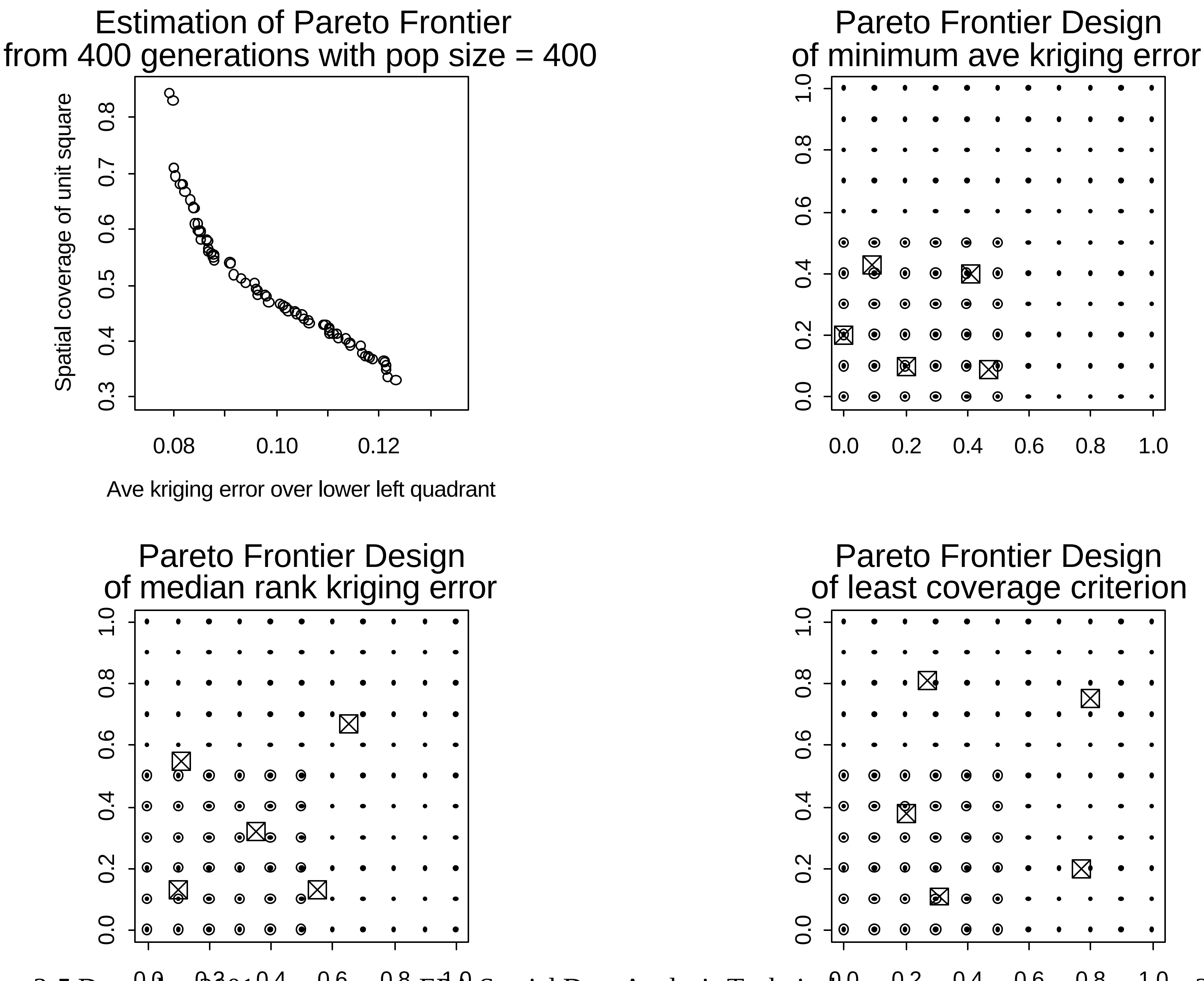

Estimation of Pareto Frontier
from 1000 generations with pop size = 200
Spatial coverage of unit square
Ave kriging error over lower left quadrant
0.3 0.4 0.5 0.6 0.7 0.8
0.08 0.10 0.12
Pareto Frontier Design
of minimum ave kriging error
Pareto Frontier Design
of median rank kriging error
Pareto Frontier Design
of least coverage criterion
0.0 0.2 0.4 0.6 0.8 1.0

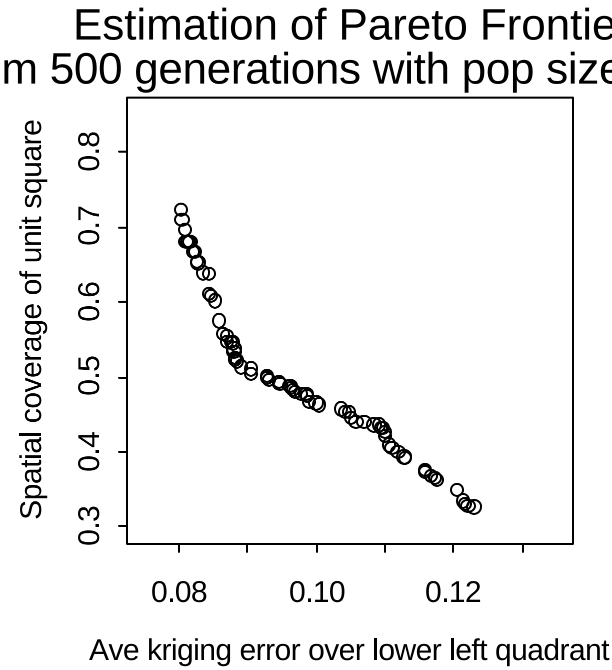

Estimation of Pareto Frontier
from 200 generations with pop size = 300
Spatial coverage of unit square
Ave kriging error over lower left quadrant
Pareto Frontier Design
of minimum ave kriging error
Pareto Frontier Design
of median rank kriging error
Pareto Frontier Design
of least coverage criterion

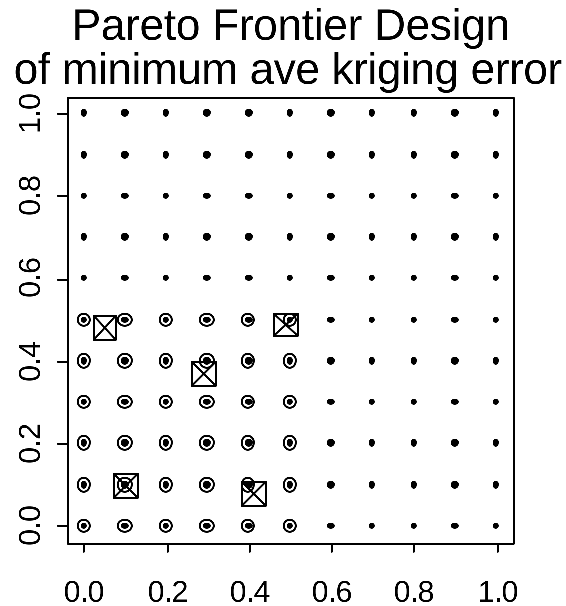


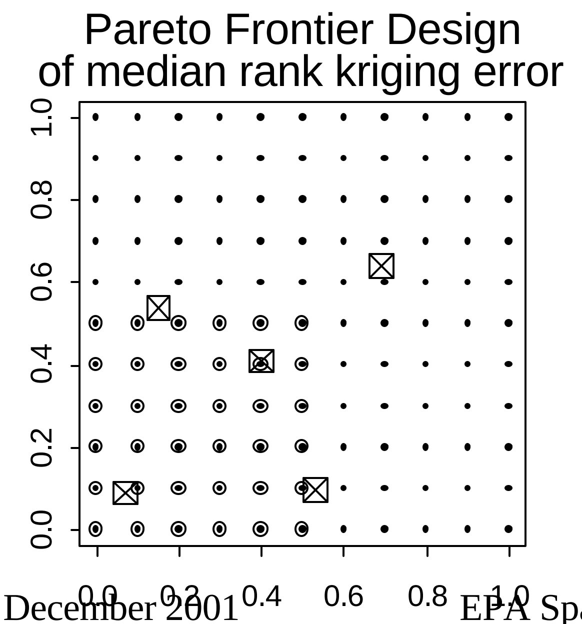
Estimation of Pareto Frontier
from 200 generations with pop size = 150
Spatial coverage of unit square
0.3 0.4 0.5 0.6 0.7 0.8
0.08 0.10 0.12
Ave kriging error over lower left quadrant
Pareto Frontier Design
of minimum ave kriging error
0.0 0.2 0.4 0.6 0.8 1.0
Pareto Frontier Design
of median rank kriging error
0.0 0.2 0.4 0.6 0.8 1.0
Pareto Frontier Design
of least coverage criterion
0.0 0.2 0.4 0.6 0.8 1.0

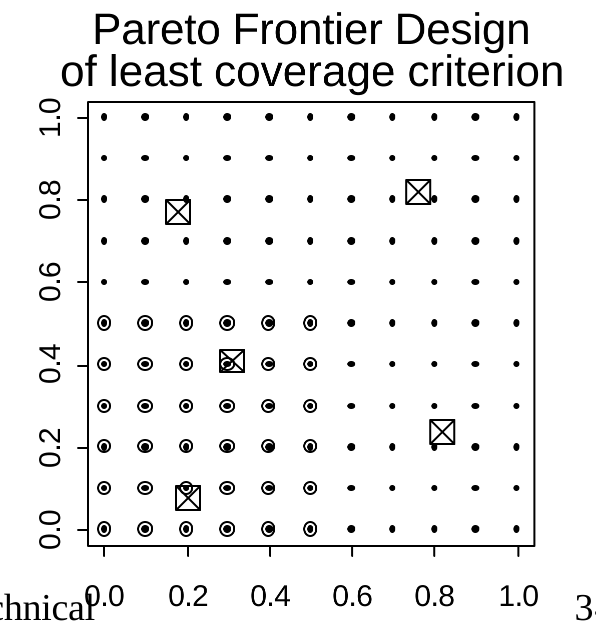

Estimation of Pareto Frontier
from 100 generations with pop size = 300
Spatial coverage of unit square
Ave kriging error over lower left quadrant
0.3
0.4
0.5
0.6
0.7
0.8
0.08
0.10
0.12
Pareto Frontier Design
of minimum ave kriging error
Pareto Frontier Design
of median rank kriging error
Pareto Frontier Design
of least coverage criterion
0.0
0.2
0.4
0.6
0.8
1.0


Estimation of Pareto Frontier
from 400 generations with pop size = 400
Spatial coverage of unit square
Ave kriging error over lower left quadrant
0.3 0.4 0.5 0.6 0.7 0.8
0.08 0.10 0.12
Pareto Frontier Design
of minimum ave kriging error
0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.2 0.4 0.6 0.8 1.0
Pareto Frontier Design
of median rank kriging error
0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.2 0.4 0.6 0.8 1.0
Pareto Frontier Design
of least coverage criterion
0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.2 0.4 0.6 0.8 1.0


Estimation of Pareto Frontier
from 500 generations with pop size = 150
Spatial coverage of unit square
Ave kriging error over lower left quadrant
0.3
0.4
0.5
0.6
0.7
0.8
0.08
0.10
0.12


Pareto Frontier Design
of minimum ave kriging error
0.0
0.2
0.4
0.6
0.8
1.0


Pareto Frontier Design
of median rank kriging error
0.0
0.2
0.4
0.6
0.8
1.0


Pareto Frontier Design
of least coverage criterion
0.0
0.2
0.4
0.6
0.8
1.0

4. Summary

- Fact: Multiple air quality monitoring objectives
- Some objectives require incorporation of prior information
  - In models and model-based estimates of errors for spatial estimation
  - In particular objectives of air monitoring that concern effects of pollutant sources & transport, and effects on human health
- Pareto optimal design calculations provide an effective way to make decisions in the context of multiple objectives.
- Evolutionary computation algorithms provide feasible tools for Pareto optimization.

- To do:
  - Redesign code for efficiency
  - Implement spatial estimation criteria based on nonstationary spatial models
  - Extend models, criteria and application to multiple pollutants
  - Write code for network assessment/reduction scenario (not just current addition/relocation of sites)