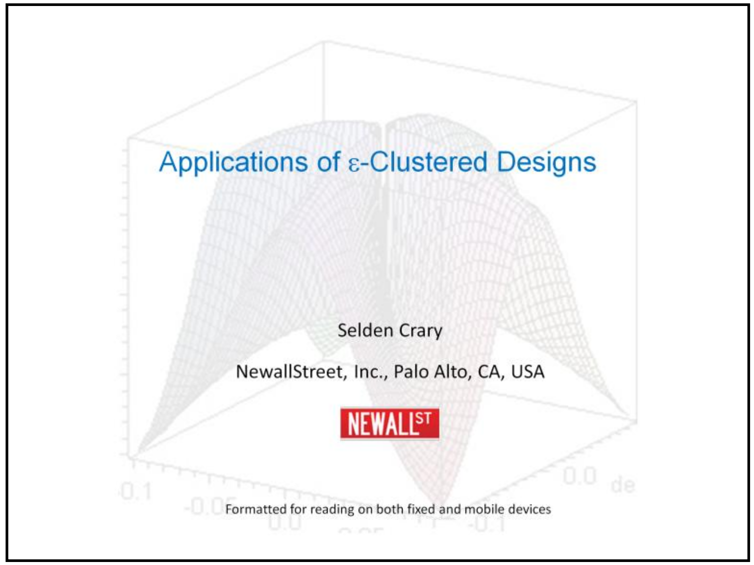
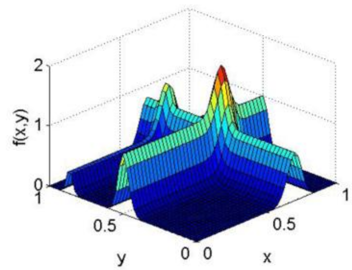
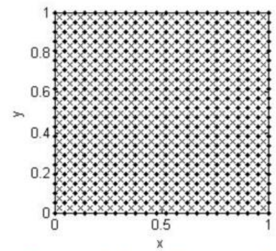


Thanks go to Chuck Boiler (SAS) for introducing us to Tom Donnelly (SAS), who informed us about both this Conference and the generally excellent research based on the University of Wisconsin, Madison dissertation of Benjamin Haaland (now Office of Clinical Sciences, Centre for Quantitative Medicine, Duke-NUS Graduate Medical School, Singapore, Department of Statistics and Applied Probability, National University of Singapore), and his dissertation advisor, Peter Qian.

Prelude 1: From Qian, Haaland, & Xu talk, this morning:

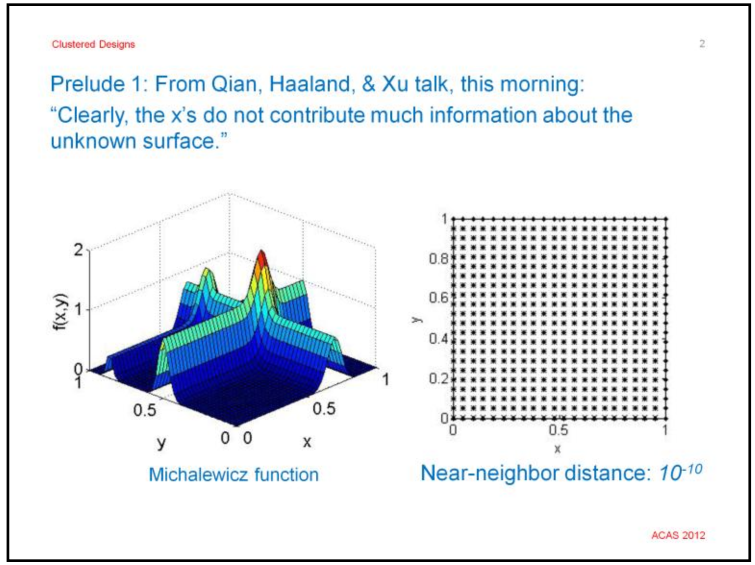


Michalewicz function

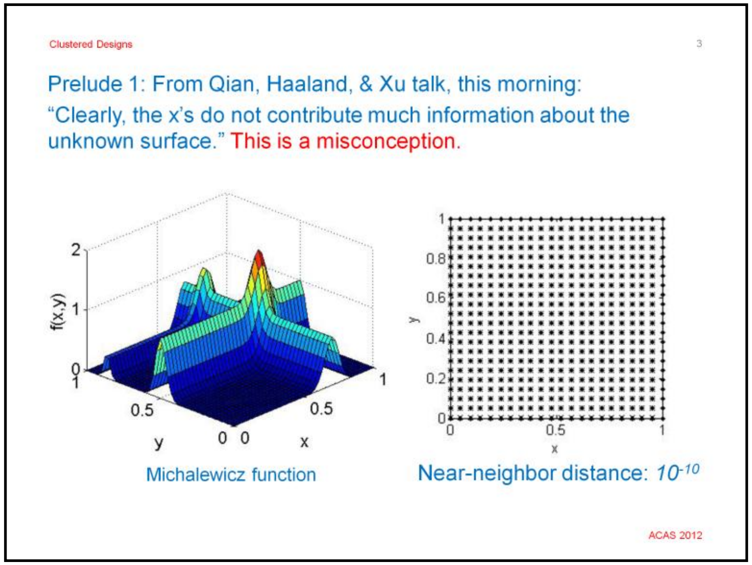


Near-neighbor distance: *0.034*

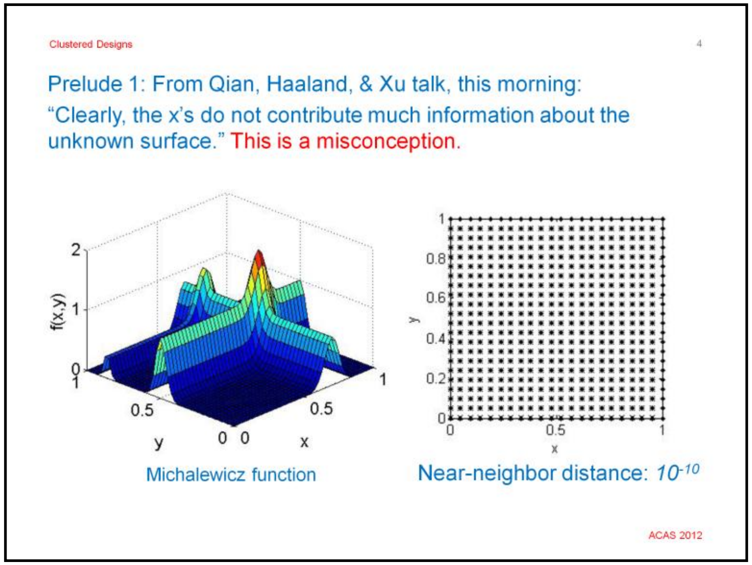
We'll start this presentation, in workshop mode, by commenting on the presentation this morning by Prof. Qian. Our hope is to stimulate discussion. Qian pointed out that, when there are evaluations of a function at a grid of points given by the filled diamonds in the figure, above, that additional information can be added by taking points on a second grid of points at the locations of the x's. This is sensible.



He then said that if the second grid is proximal to the first, as shown above, where the second grid is offset in the x-direction by just 10^{-10}, over a full-scale range of unity, “Clearly, the x’s do not contribute much information about the unknown surface.” It is tempting to take this as plausible, as when the distance is zero, for computer experiments, information from the second grid is non-informative.



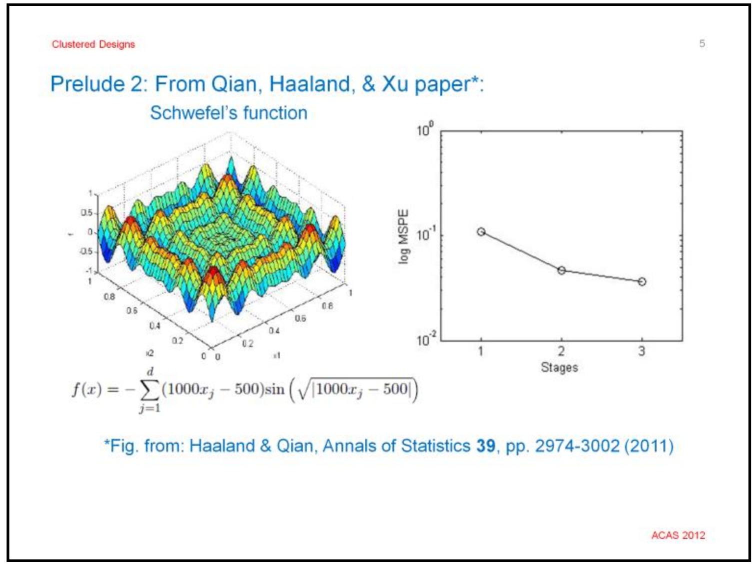
However, we should be cautious, as, we will argue, the second grid potentially can provide a wealth of information about the function. One way of seeing this is that a pair of proximal points provide both a function evaluation, as well as a directional derivative, in the vicinity of the pair of points. As reported earlier by the present author and his collaborators (SBC, “New Research Directions in Computer Experiments: epsilon-Clustered Designs,” SRC 2012 Proceedings, Statistical Computing Section, Alexandria, VA, USA: ASA, pp. pp. 5692-5706, and references therein. Revised editions are available from the author), such information can be highly informative. In fact, optimal N-point designs exist with points that are specified to be taken as closely together as possible, given the computational resources available. That is, a design with a pair of twin points is sometimes more informative than any design without a twin point. The discovery of these so-called “twin points” or “epsilon-clustered points” demonstrates that we should not hold rigidly to the view that designs for computer experiments should be space-filling in the usual sense of min discrepancy, min max distance between nearest-neighbor points, or max min Voronoi cell volumes. ... continues on next page…



… continued from last page …

Rather, we should be more open-minded, keeping in mind, from elementary calculus, that an analytic function can be determined to any desired accuracy by a Taylor series based on a function evaluation and an evaluation of all derivatives, at a single point. In this latter view, optimal twin-point designs simply tell us that the optimal design is sometimes neither space-filling nor an evaluation of derivatives at just one point.

We have emphasized, on the slide, that there is a misconception that needs to be overcome for the field to move forward properly. There are significant opportunities for innovative research related to epsilon-clustered designs.

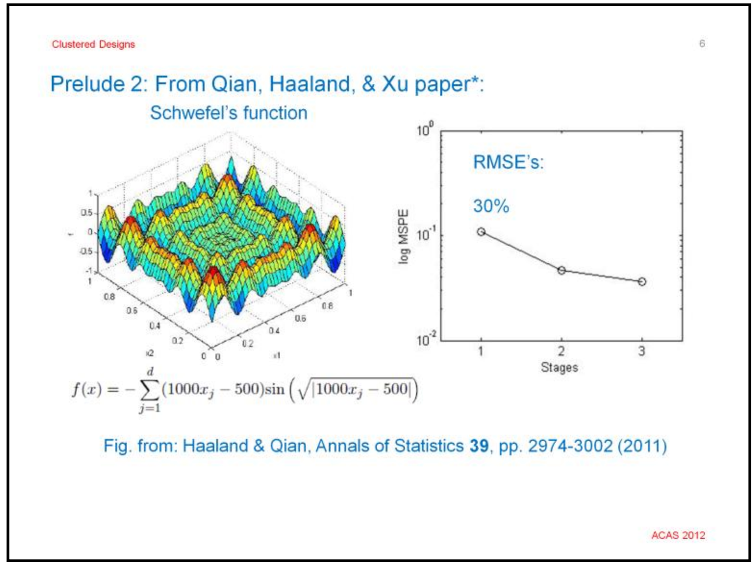


We now make a comment on the final slides in Haaland and Qian's recent Annals paper. For the complex* function shown, the nested-Kriging method proposed in the paper gave a log (mean-squared prediction error) of ~0.1, as shown in the figure. However, such MSPE is actually little better than the MSPE found if one just took the predictor to be the average of the responses.

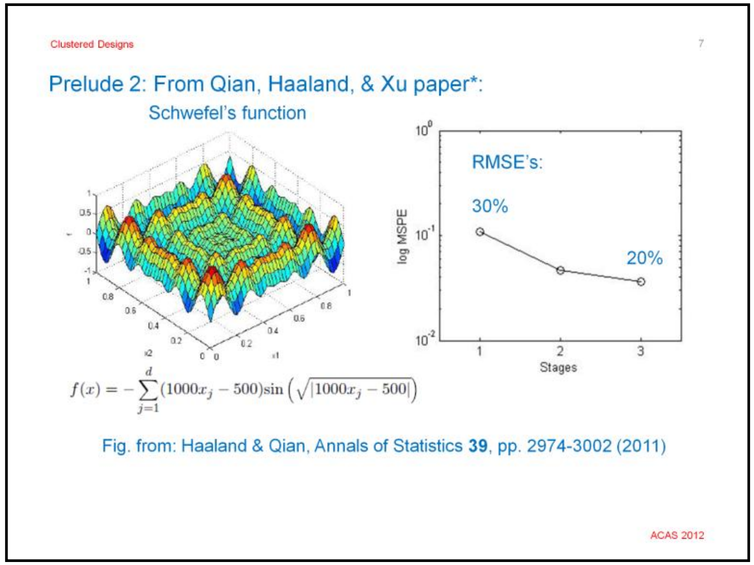
Ben Haaland mentioned this limitation of the nested approach, in his September 9, 2011 presentation: B. Haaland, "Accurate emulators for large-scale computer experiments," Isaac Newton Institute for Mathematical Studies, Cambridge, UK. (PDF and video available at URL:

http://www.newton.ac.uk/programmes/DAE/seminars/20110909 12001.html).

*Unless symmetry is allowed to be recognized or discovered in this function, the function is simply too complex to be adequately approximated by the given design. This is the principal reason for the poor fit.

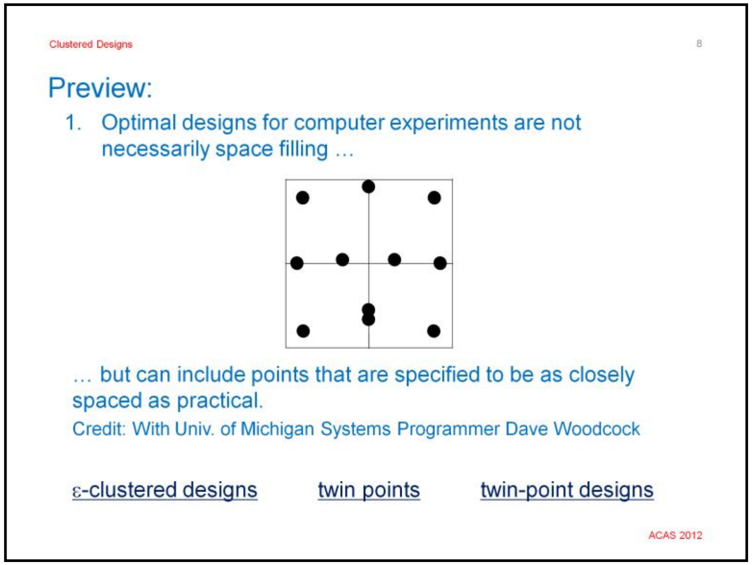


In fact, our analysis shows that the Stage 1 RMSE is ~30%.



By Stage 3 the RMSE has been reduced to ~20%.

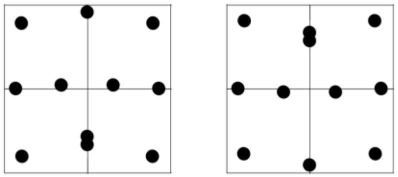
Aside: The Annals paper contains a possible typo. To be correct, the first "1000" should be "1.000," and the first "500" should be "0.500." Correcting this error makes the reanalysis, here, sensible.



We now quickly preview eleven numbered points we wish to make in this presentation. Underlined words are recently introduced or new terms to the field of computer experiments.

Preview:

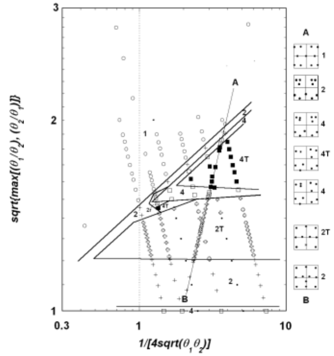
2. Each optimal design has a group of symmetries leaving the objective function unchanged.



phases

Preview:

3. Designs fall into contiguous regions of identical phase.
   There are phase transitions between the phases.



phase transitions

Preview:

4. Theory: The IMSE objective function is a low-degree-truncated, rational polynomial.

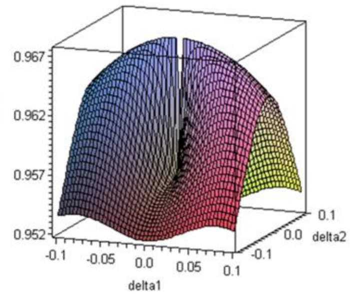
$$IMSE = \frac{a\delta_1^2 + b\delta_1\delta_2 + c\delta_2^2 \quad + \quad d\delta_1^3 + e\delta_1^2\delta_2 + f\delta_1\delta_2^2 + g\delta_2^3 \quad + \quad \cdots}{h\delta_1^2 + i\delta_1\delta_2 + j\delta_2^2 \quad + \quad k\delta_1^3 + l\delta_1^2\delta_2 + m\delta_1\delta_2^2 + n\delta_2^3 \quad + \quad \cdots}$$

low-degree-truncated rational polynomial

Preview:

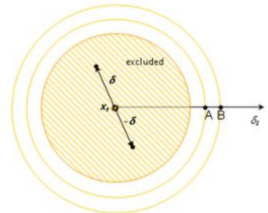
5. The IMSE objective function, as a function of the separation of a pair of twin points, is an essential discontinuity.

essential discontinuity



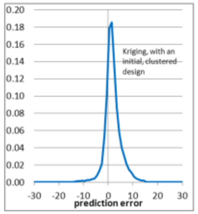
Preview:

6. Using the theory, previously incomputable quantities can be determined via extrapolation.



Preview:

7. Borehole example starting with a nonuptal-point design.



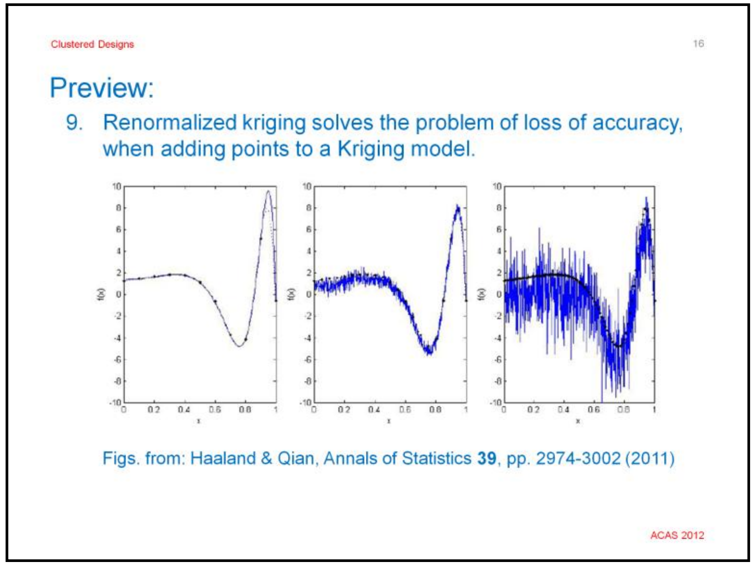
Preview:

8. With three, equispaced, collinear points, Kriging gives a faulty far-field predictand, …

$$\hat{Y} = \frac{1}{6\theta}\left[1 - e^{-\theta x^2}\right] y''(0) - \frac{1}{6} x^2 e^{-\theta x^2} + y(0) + \left[y'(0)x + \frac{1}{2} y''(0) x^2\right] e^{-\theta x^2} + O(\delta^2)$$

… but this can be cured by dropping the offending terms.

Renormalized Kriging



The function to be approximated by Kriging is shown with a (faint) dotted line. The function evaluations at N=11, 21, and 81 design points, are shown as black dots, in the left-hand, center, and right-hand panels, respectively. The Kriging fits are shown with dark blue lines and demonstrate increasing ill-conditioning as N increases.

Preview: HIGHLY SPECULATIVE

9. A new microscopic theory of Kriging is necessary. Is there a path via 3-way tensor covariance? υ-Kriging
10. nugget:kriging::strings:QED?
11. Can statistics come to the aid of physics?

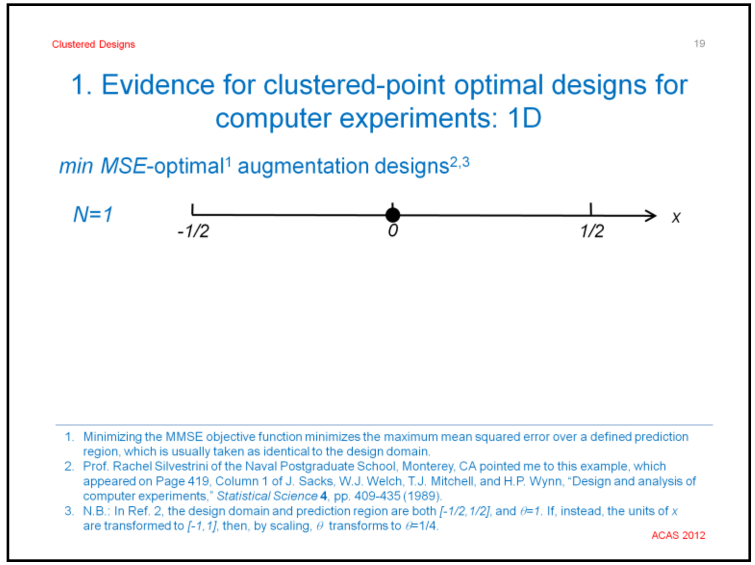
Ordinary-kriging set-up

k-dimensional design domain: $[-1,1]^k$, *except as noted*
Model function: $Y=\beta_0+Z(x)$
Gaussian process
Gaussian covariance:
 $cov[Z(s_1, s_2), Z(t_1, t_2)] = \sigma_Z^2 \exp\{-[\theta_1(s_1-t_1)^2 + \theta_2(s_2-t_2)^2]\}$
 fixed σ_Z, over design domain
 known, fixed $\theta=(\theta_1,\theta_2)$, over design domain
Optimality criterion: Minimum expected *IMSE*
 N-point design
 Integral over $[-1,1]^k$

$$\min_{\omega_N} \int_{-1}^{1} \cdots \int_{-1}^{1} E\{[\hat{Y}(x) - Y(x)]^2\} dx_1 \cdots dx_k$$

 Exact (integer number of design points at each location)

This is a review of the notation used in this presentation.



Slides 19-42 are drawn from the author's presentation* at the Spring Research Conference 2012 and are shown here, in rapid succession, as background material.

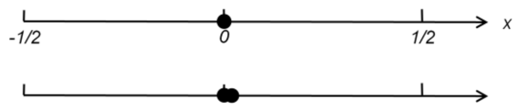
*See reference on Slide 3.

Evidence for clustered-point optimal designs for computer experiments: 1D

MMSE-optimal[1] augmentation designs[2,3]

N=1

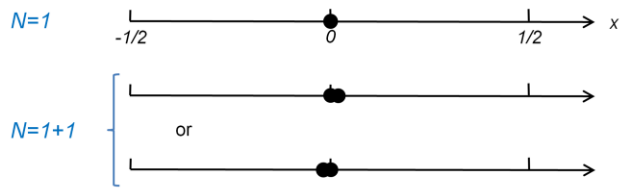
N=1+1



1. Minimizing the MMSE objective function minimizes the maximum mean squared error over a defined prediction region, which is usually taken as identical to the design domain.
2. Prof. Rachel Silvestrini of the Naval Postgraduate School, Monterey, CA pointed me to this example, which appeared on Page 419, Column 1 of J. Sacks, W.J. Welch, T.J. Mitchell, and H.P. Wynn, "Design and analysis of computer experiments," *Statistical Science* **4**, pp. 409-435 (1989).
3. N.B.: In Ref. 2, the design domain and prediction region are both $[-1/2, 1/2]$, and $\theta=1$. If, instead, the units of x are transformed to $[-1, 1]$, then, by scaling, θ transforms to $\theta=1/4$.

Evidence for clustered-point optimal designs for computer experiments: 1D

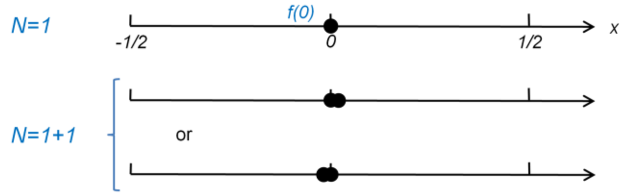
MMSE-optimal[1] augmentation designs[2,3]



1. Minimizing the MMSE objective function minimizes the maximum mean squared error over a defined prediction region, which is usually taken as identical to the design domain.
2. Prof. Rachel Silvestrini of the Naval Postgraduate School, Monterey, CA pointed me to this example, which appeared on Page 419, Column 1 of J. Sacks, W.J. Welch, T.J. Mitchell, and H.P. Wynn, "Design and analysis of computer experiments," *Statistical Science* **4**, pp. 409-435 (1989).
3. N.B.: In Ref. 2, the design domain and prediction region are both $[-1/2, 1/2]$, and $\theta=1$. If, instead, the units of x are transformed to $[-1, 1]$, then, by scaling, θ transforms to $\theta=1/4$.

Evidence for clustered-point optimal designs for computer experiments: 1D

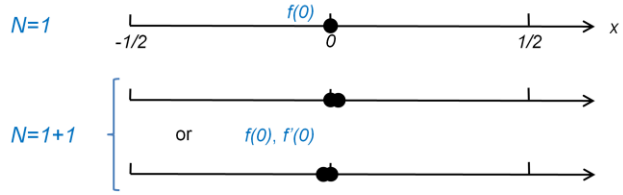
MMSE-optimal[1] augmentation designs[2,3]



1. Minimizing the MMSE objective function minimizes the maximum mean squared error over a defined prediction region, which is usually taken as identical to the design domain.
2. Prof. Rachel Silvestrini of the Naval Postgraduate School, Monterey, CA pointed me to this example, which appeared on Page 419, Column 1 of J. Sacks, W.J. Welch, T.J. Mitchell, and H.P. Wynn, "Design and analysis of computer experiments," *Statistical Science* **4**, pp. 409-435 (1989).
3. N.B.: In Ref. 2, the design domain and prediction region are both $[-1/2, 1/2]$, and $\theta=1$. If, instead, the units of x are transformed to $[-1, 1]$, then, by scaling, θ transforms to $\theta=1/4$.

Evidence for clustered-point optimal designs for computer experiments: 1D

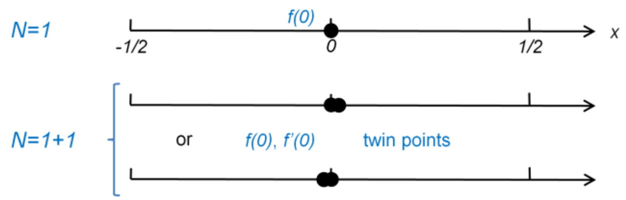
MMSE-optimal[1] augmentation designs[2,3]



1. Minimizing the MMSE objective function minimizes the maximum mean squared error over a defined prediction region, which is usually taken as identical to the design domain.
2. Prof. Rachel Silvestrini of the Naval Postgraduate School, Monterey, CA pointed me to this example, which appeared on Page 419, Column 1 of J. Sacks, W.J. Welch, T.J. Mitchell, and H.P. Wynn, "Design and analysis of computer experiments," *Statistical Science* **4**, pp. 409-435 (1989).
3. N.B.: In Ref. 2, the design domain and prediction region are both $[-1/2, 1/2]$, and $\theta=1$. If, instead, the units of x are transformed to $[-1, 1]$, then, by scaling, θ transforms to $\theta=1/4$.

Evidence for ε-clustered optimal designs for computer experiments: 1D

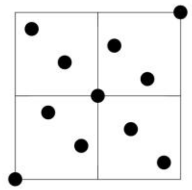
MMSE-optimal[1] augmentation designs[2,3]



1. Minimizing the MMSE objective function minimizes the maximum mean squared error over a defined prediction region, which is usually taken as identical to the design domain.
2. Prof. Rachel Silvestrini of the Naval Postgraduate School, Monterey, CA pointed me to this example, which appeared on Page 419, Column 1 of J. Sacks, W.J. Welch, T.J. Mitchell, and H.P. Wynn, "Design and analysis of computer experiments," *Statistical Science* **4**, pp. 409-435 (1989).
3. N.B.: In Ref. 2, the design domain and prediction region are both $[-1/2, 1/2]$, and $\theta=1$. If, instead, the units of x are transformed to $[-1, 1]$, then, by scaling, θ transforms to $\theta=1/4$.

Evidence … 2D

N=11, θ_1*=0.128*, θ_2*=0.069*, design domain *[-1,1]*, *abscissa is* x_1, …

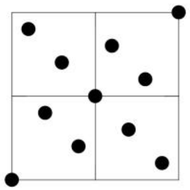


Typical LHC[4]

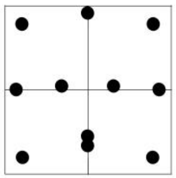
4. W.J. Welch, R.J. Buck, J. Sacks, H.P. Wynn, and T.J. Mitchell, and M.D. Morris, “Screening, Predicting, and Computer Experiments,” *Technometrics* **34**, pp. 15-25 (1992).

Evidence … 2D

N=11, θ_1*=0.128*, θ_2*=0.069*, design domain *[-1,1]*, *abscissa is* x_1, …



Typical LHC[4]



IMSE-optimal[5-9]

4. W.J. Welch, R.J. Buck, J. Sacks, H.P. Wynn, and T.J. Mitchell, and M.D. Morris, "Screening, Predicting, and Computer Experiments," *Technometrics* **34**, pp. 15-25 (1992).

5. D.M. Woodcock and S.B. Crary, "Navigating Smooth IMSE Landscapes Using High Precision Floating Point," (oral only) Joint Statistical Meetings, Indianapolis, IN, 13-17 Aug 2000.

6. S.C., D.M. Woodcock, and A. Hieke, "Designing efficient computer experiments for metamodel generation," Published in the Proceedings of the Fourth International Conference on Modeling of Microsystems, MSM 2001, Hilton Head, SC, March 19-21, 2001, pp. 132-135.

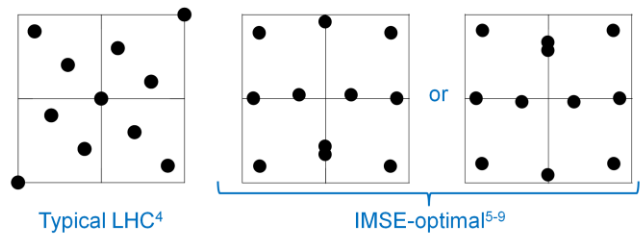
7. S.C., "Statistical design and analysis of computer experiments for the generation of parsimonious metamodels," Published in Design, Test, Integration, and Packaging of MEMS/MOEMS 2001, B. Courtois, J.M. Karam, S.P. Levitan, K.W. Markus, A.A.O. Tay, and J.A. Walker, Editors, Proceedings of SPIE Vol. 4408, pp. 29-39 (2001).

8. S.C., "Design of computer experiments for metamodel generation," Analog Integrated Circuits and Signal Processing **32**, pp. 7-16 (2002).

9. S.C. and R. Johnson, "Validation of the Twin-Point-Design Concept in the Design of Computer Experiments," Section on Statistical Computing – JSM 2011, pp. 5495-5505.

Evidence … 2D

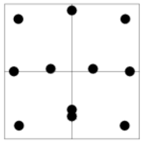
N=11, $\theta_1=0.128$, $\theta_2=0.069$, design domain *[-1,1]*, *abscissa is* x_1, …



4. W.J. Welch, R.J. Buck, J. Sacks, H.P. Wynn, and T.J. Mitchell, and M.D. Morris, "Screening, Predicting, and Computer Experiments," *Technometrics* **34**, pp. 15-25 (1992).
5. D.M. Woodcock and S.B. Crary, "Navigating Smooth IMSE Landscapes Using High Precision Floating Point," (oral only) Joint Statistical Meetings, Indianapolis, IN, 13-17 Aug 2000.
6. S.C., D.M. Woodcock, and A. Hieke, "Designing efficient computer experiments for metamodel generation," Published in the Proceedings of the Fourth International Conference on Modeling of Microsystems, MSM 2001, Hilton Head, SC, March 19-21, 2001, pp. 132-135.
7. S.C., "Statistical design and analysis of computer experiments for the generation of parsimonious metamodels," Published in Design, Test, Integration, and Packaging of MEMS/MOEMS 2001, B. Courtois, J.M. Karam, S.P. Levitan, K.W. Markus, A.A.O. Tay, and J.A. Walker, Editors, Proceedings of SPIE Vol. 4408, pp. 29-39 (2001).
8. S.C., "Design of computer experiments for metamodel generation," Analog Integrated Circuits and Signal Processing **32**, pp. 7-16 (2002).
9. S.C. and R. Johnson, "Validation of the Twin-Point-Design Concept in the Design of Computer Experiments," Section on Statistical Computing – JSM 2011, pp. 5495-5505.

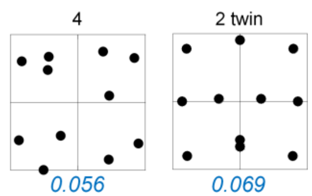
2. Phases: For all designs on this page: $\theta_1=0.128$

2 twin

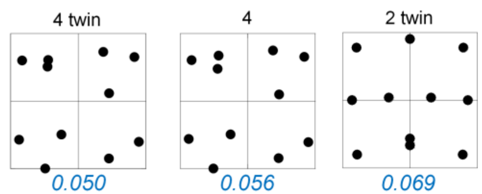


0.069

2. Phases: For all designs on this page: θ_1=*0.128*

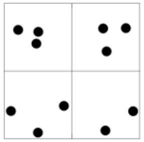


2. Phases: For all designs on this page: θ_1=*0.128*



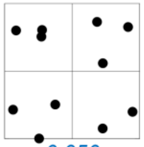
2. Phases: For all designs on this page: θ_1=*0.128*

4



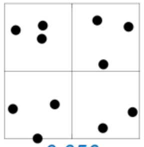
0.038

4 twin



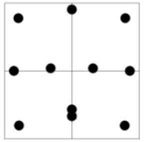
0.050

4



0.056

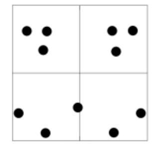
2 twin



0.069

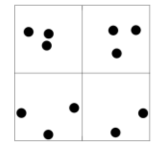
2. Phases: For all designs on this page: θ_1=0.128

4



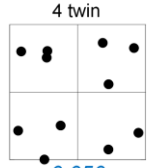
0.037

4



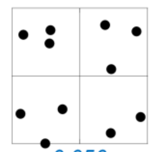
0.038

4 twin



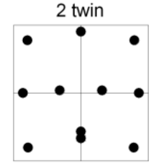
0.050

4



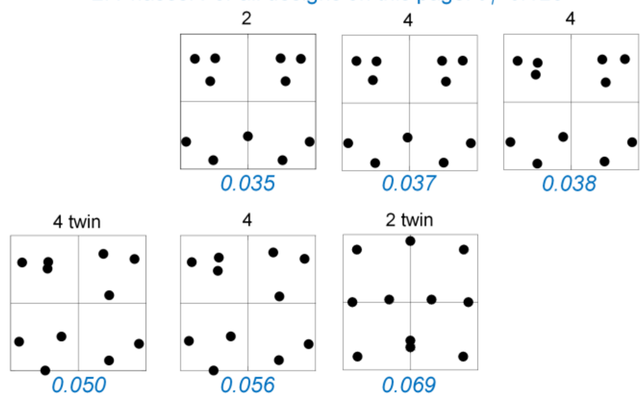
0.056

2 twin

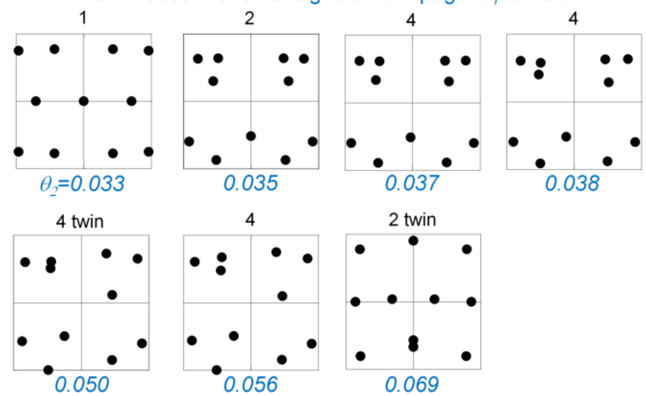


0.069

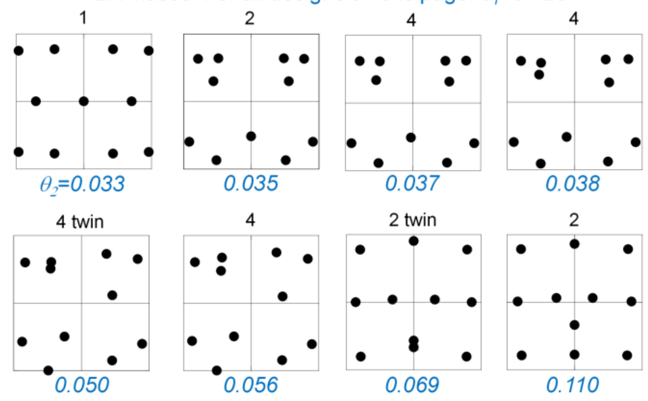
2. Phases: For all designs on this page: θ_1=0.128



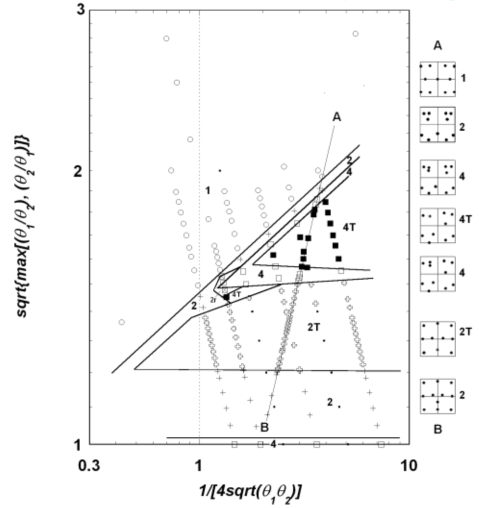
2. Phases: For all designs on this page: θ_1=0.128

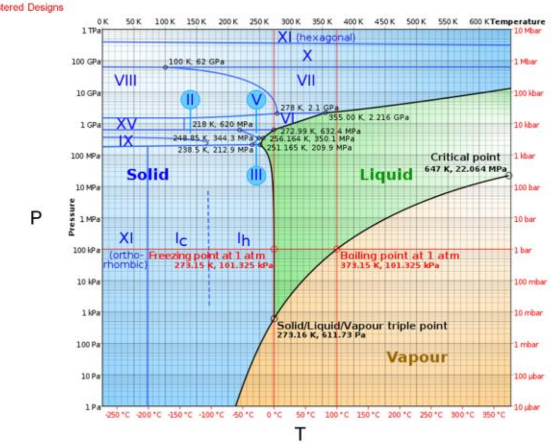


2. Phases: For all designs on this page: θ_1=0.128

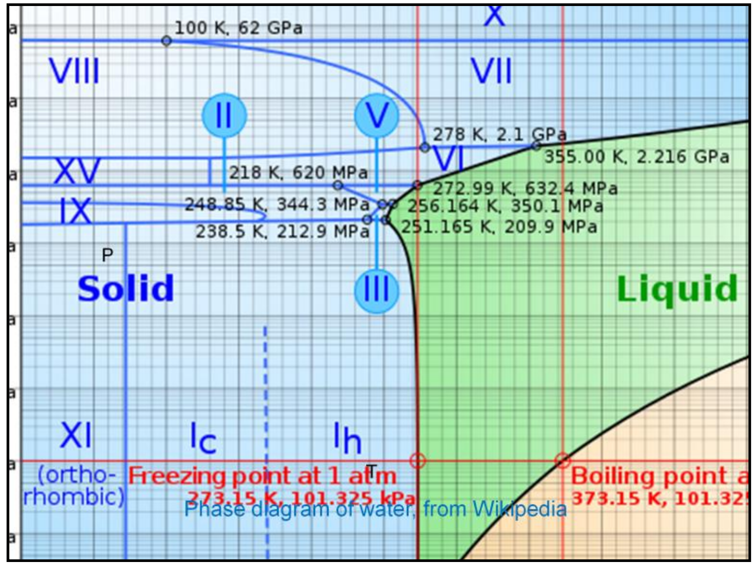


3. Phase transitions: IMSE optimal designs





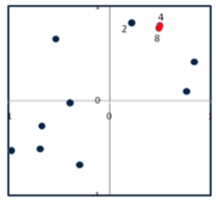
Phase diagram of water, from Wikipedia



Phase diagram of water, from Wikipedia

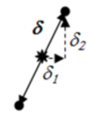
4. IMSE is a low-degree-truncated rational function

k=2, one pair twin points



Key ideas:

- Change variables

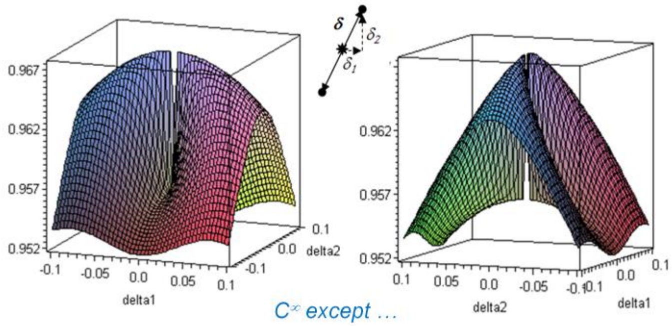


- Expand in powers of δ_1 and δ_2
- Assumes power laws exist and converge appropriately

$$IMSE = \frac{a\delta_1^2 + b\delta_1\delta_2 + c\delta_2^2 \quad + \quad d\delta_1^3 + e\delta_1^2\delta_2 + f\,\delta_1\delta_2^2 + g\delta_2^3 \quad + \quad \cdots}{h\delta_1^2 + i\delta_1\delta_2 + j\delta_2^2 \quad + \quad k\delta_1^3 + l\delta_1^2\delta_2 + m\delta_1\delta_2^2 + n\delta_2^3 \quad + \quad \cdots}$$

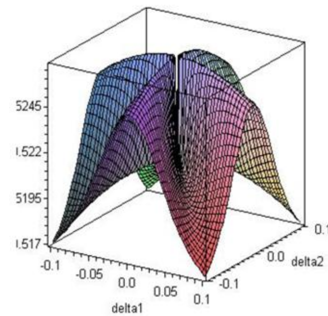
k=1 is a special case. Above equation generalizes for any $k>1$. Proof available, upon request.

5. *IMSE*, in the vicinity of a pair of twin points[10-11], *is an essential discontinuity*: *[-1,1]*^{2}, *N=2,* center of twins at *(0.0,0.6)*



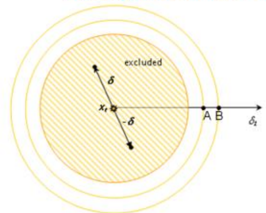
C^∞ except …

10. These figures are from SC. and R. Johnson, "Validation of the Twin-Point-Design Concept in the Design of Computer Experiments," Section on Statistical Computing – JSM 2011, pp. 5495-5505.
11. In the figures, *delta1*=2δ_1 and *delta2*=2δ_2.



Looking to the future: Here is the *IMSE* of three, equispaced, collinear points centered on the origin of $[-1,1]^2$.
As before, *delta1*=$2\delta_1$ and *delta2*=$2\delta_2$.

6. Previously incomputable quantities
Example: Accurate Computation of IMSE



There is severe ill-conditioning, when $\delta < 10^{-6}$ …

… but the form of the *IMSE* is known, *viz.*,

$$\lim_{\substack{\delta_2=0 \\ \delta_1\to 0}} IMSE/\sigma_z^2 \cong \alpha_0 + \alpha_1\delta_1^2 + \cdots$$

so, fit and extrapolate

Design-point label	δ_t	$x=\delta_t^2$	$IMSE/\sigma_z^2$
A	0.00120	0.0000144	0.7460920868…
B	0.00140	0.0000196	0.7460912887…
x_t via extrapolation	0	0	0.7460942972…
x_t via symbolic analysis	0	0	0.7460942972…

It works!

7. Borehole example starting with a nonuptal-point design.

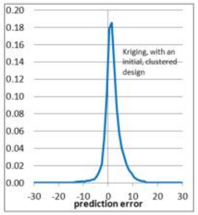
$$y = \frac{2\pi T_u (H_u - H_l)}{ln(r/r_w)\left[1 + \frac{2LT_u}{ln(r/r_w) r_w^2 K_w} + \frac{T_u}{T_l}\right]}$$

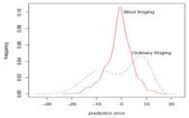
Stage 1: Nonuptal-point design at origin:

- antisymmetric dependence can be identified
- a method of screening
- dimensional analysis can be used

Stage 2: IMSE-optimal, 9-point, augmentation design, using JMP V.7

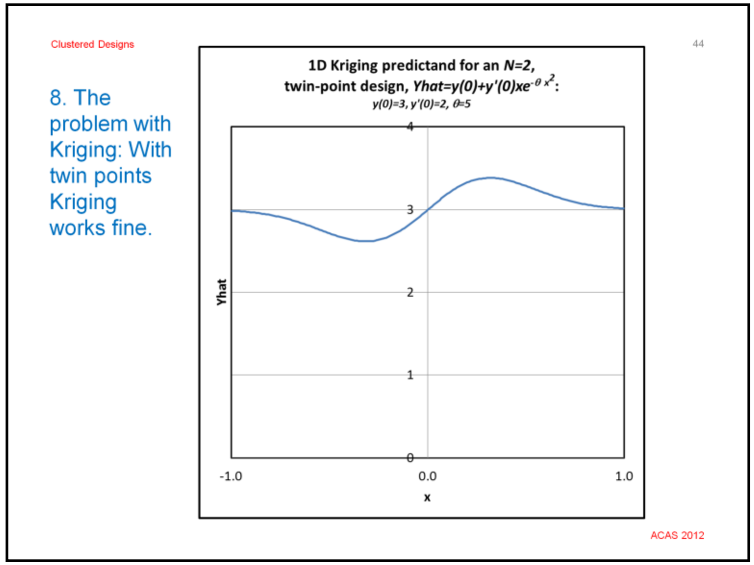
Stage 3: Final 9-point augmentation



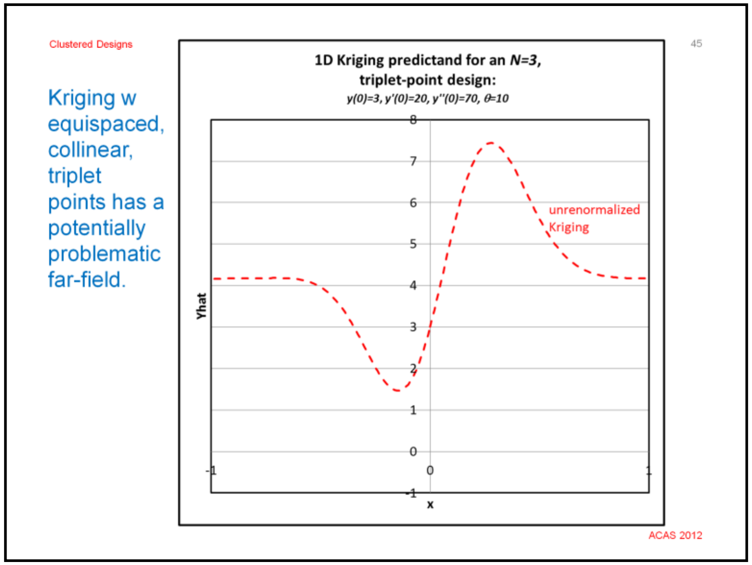


Ordinary and blind kriging from

Joseph, et al., ASME J. Mech. Des. **130**, 03112-1-8



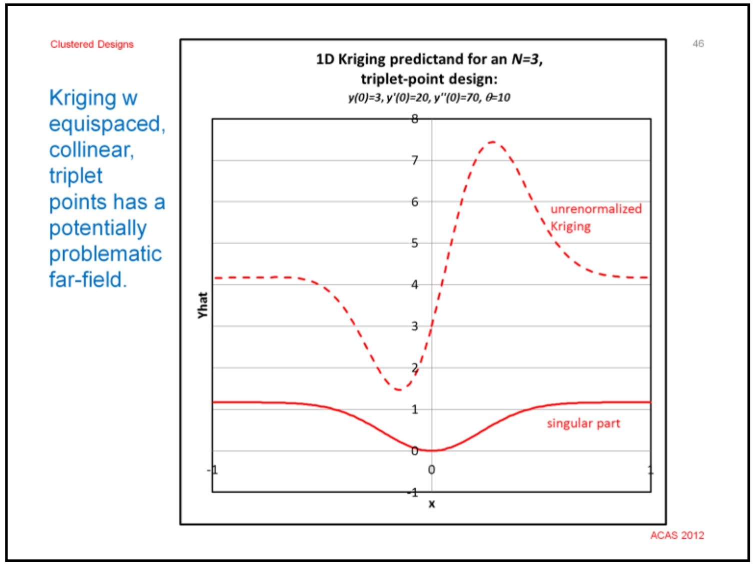
Slides 44 through 48 demonstrate an application of the twin-point perspective to Kriging. For N=2 points, the Kriging predictand can be expressed in simple algebraic form. We have shown that, in the limit of zero distance between the points (N.B.: This is not the same as repeated points), the predictand can be expressed as the formula for Yhat in the title of the plot, viz., as the value of the response at the location of the twin points (y(0) or alternatively as the average of the responses at the points) plus the derivative of the responses (y'(0) or alternatively determined by a simple, two-point difference formula) times the independent variable x times a decay function exp(-theta*x^2). This is sensible. It captures the data perfectly, and then, in the far-field, decays to the average of the data.



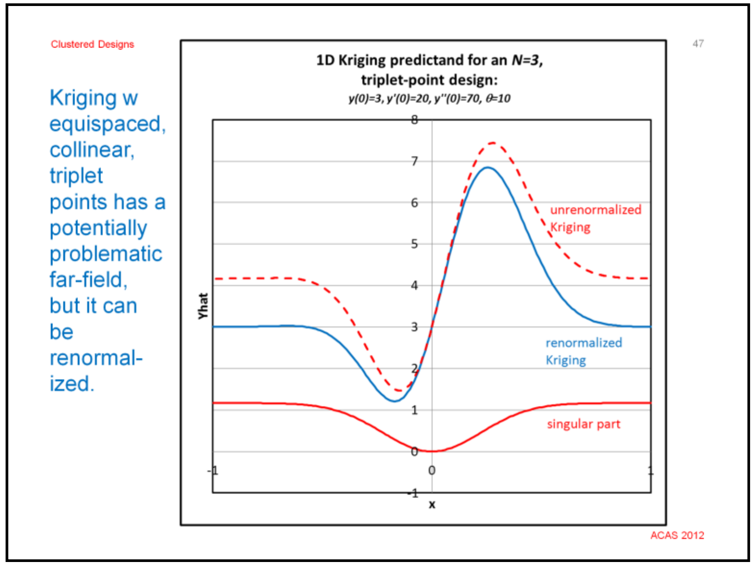
A similar analysis for three proximal points, i.e., triplet points, in the limit as the distances between them approach zero, gave a far-field that was distinctly different from the average of the three responses. (Detail is provided on Slide 51.) The function has the values of y(0), y'(0), and y''(0) given in the title, so the average of the three responses, in the limit, is just 3, but the far-field specified by Kriging is greater than 4. The Kriging predictand is sensible, local to the data, but raises an interesting issue in the far-field.

Gordon A. Fenton, "Estimation for Stochastic Soil Models," *ASCE Journal of Geotechnical and Geoenvironmental Engineering*, **125** (6), pp. 470-485, 1999 noted this effect, but he wrote , "Note that this estimator is generally not very different from the usual estimator obtained by simply averaging the observations."

We note the fact that, for triplet points, the far-field differs from the average of the responses by exactly y''(0)/(6*theta), and this can achieve any value, for fixed theta, depending upon the data contributing to the second derivative.



If we consider that the desirable part of the predictand is y(0)+[y'(0)*x+(1/2)*y''(0)*x^2]*exp(-theta*x^2), then the predictand has additional terms, which we clump together as the "singular part."



If this singular part is removed, we get the desired far-field. The resulting predictand is shown by the blue line. This is a sensible predictand, as it has the correct y(0), y'(0), and y''(0) at the triplet-point, and the far-field decays to the average of the responses.

However, it should be pointed out that, as the limit is approached, the predictand does not pass through the data perfectly. But to emphasize: everything is fine in the limit.

For true triplet-point designs, this indicates we might choose to abandon customary Kriging, as it gives a possibly undesirable far-field. But, what if we are in a regime where the distance between the points is small but finite? Then we have to make a choice. I suggest this is an interesting and important area for research.

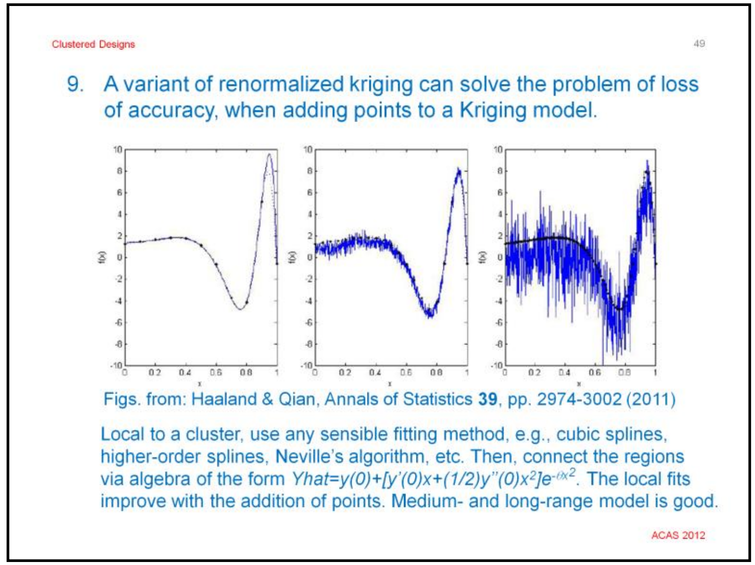
Renormalized Kriging

Kriging with three or more points can give a potentially problematic far-field.

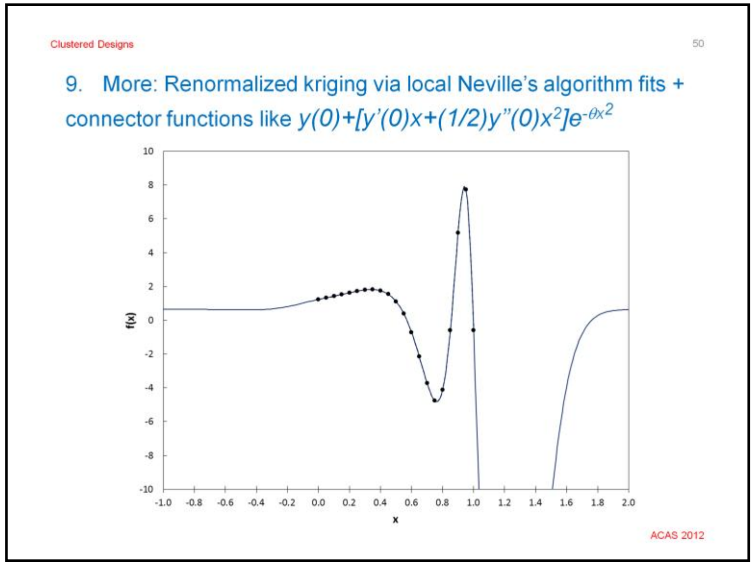
The potentially offending terms can be removed, leading to renormalized Kriging: $Yhat=y(0)+[y'(0)x+(1/2)y''(0)x^2]e^{-\theta x^2}$

Renormalized Kriging is extensible to m-uplet-point designs.

There are many outstanding research questions.



A variant of renormalized Kriging is to take any sensible fit local to the data at hand and then to extrapolate to other local regions, or to the far-field, using a bridging function of the general form exp(-theta*x^2). Instead of the fit getting worse, as shown by the solid-blue lines, when data are added, the fit improves.



For the function given in the Haaland and Qian paper, a local fit via Neville's algorithm can be used and then connected to the far-field. The resulting fit is shown by the solid-blue line, for a fixed value of theta.

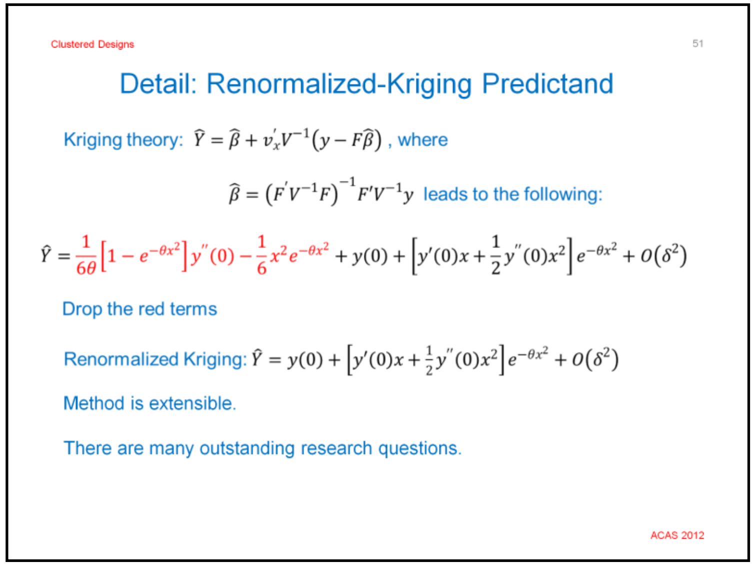
We have also found that the undesirable effects of ill-conditioning, in this problem, can be removed by use of the linear-exponential covariance function (see, e.g., Erik Vanmarcke, Random Fields: Analysis and Synthesis, World Scientific, revised edition, 2010, p. 132):

[(sigma_z)^2]*[1+alpha*abs(tau)]*exp[-alpha*abs(tau)].

We found that this function provided a much smoother interpolant just the exponential covariance function

[(sigma_z)^2]*exp[-alpha*abs(tau)].

Our research on condition numbers and solutions to the ill-conditioning for this problem is on-going, and we will provide a fuller report, in the near future.



In summary, twin points and their m-uplet-point extensions pose both conceptual and interesting research challenges. As a community, we should be open-minded to designs with clusters of proximal points, and we should learn how to exploit the information they provide, even if it comes in the form of derivative information. Analysis of Kriging with these concepts in mind has led to the possibility of a renormalized Kriging that neglects certain terms that arise in customary Kriging.

The author thanks the organizers of ACAS for their hospitality and organizational efforts. He welcomes others to collaborate with him, or to start their own research programs, on the interesting subject of epsilon-clustered designs.