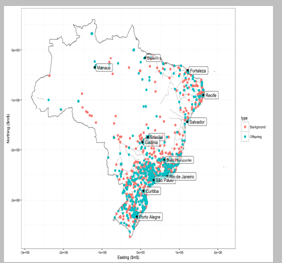
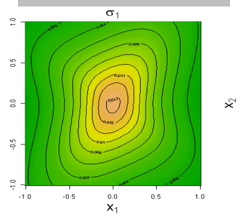
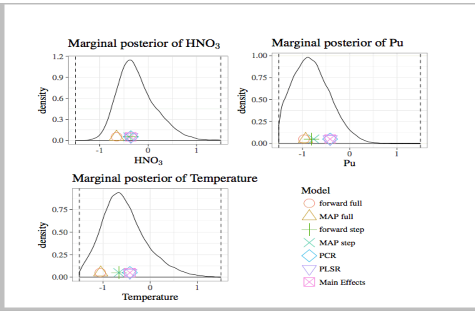
SAND2017-10461PE







Some Statistical Applications at Sandia

University Of Illinois Urbana-Champaign

Recruiting Visit 09/27/2017

John R. Lewis, PhD

Senior Member of Technical Staff

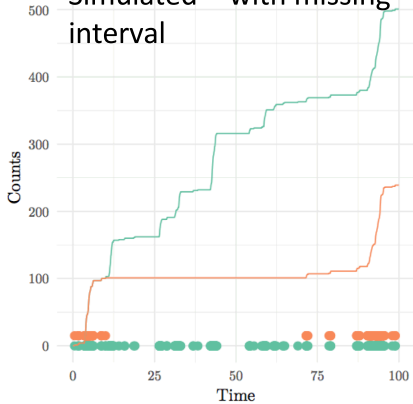
Statistical Sciences

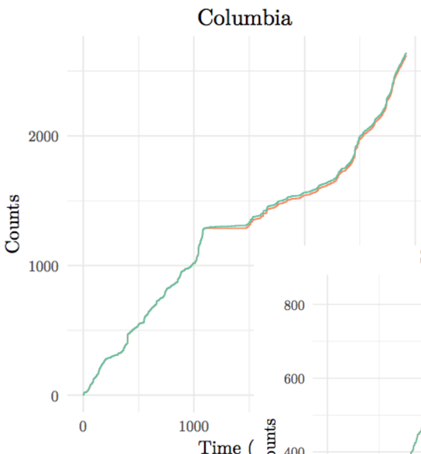
Outline

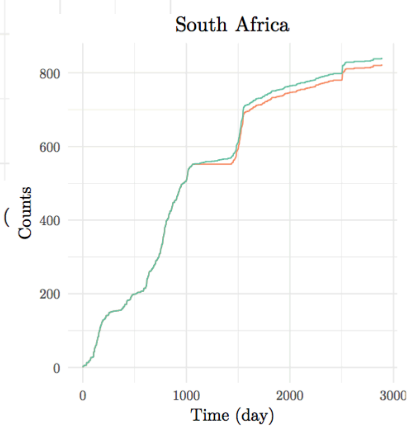
- Accounting for missing time/space histories in marked point-process models
- Inverse prediction for nuclear forensics applications

Point-Process Examples

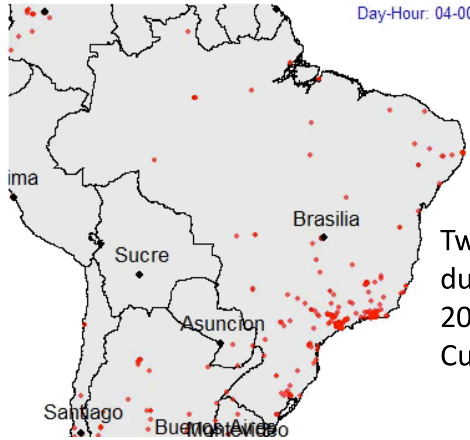
Simulated – with missing interval



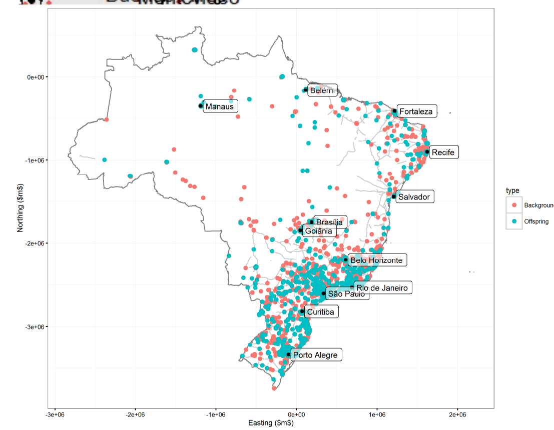


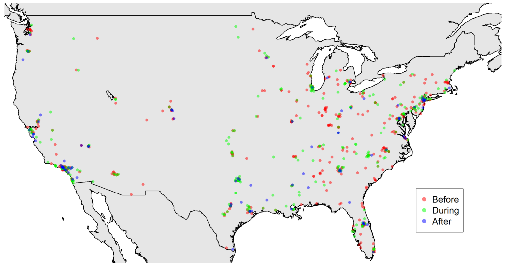


Terrorism events from GTD



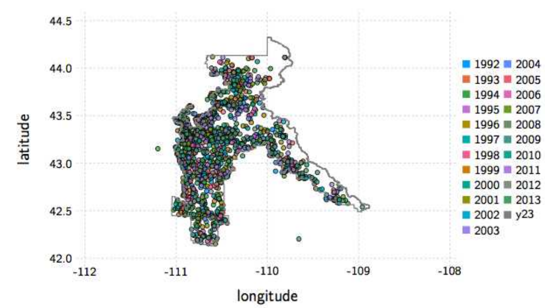
Tweets during the 2014 World Cup





Tweets by location shown for the four hours before, the four hours during, and the four hours after the Paris attacks on 11/13/2015.

All forest fires in Bridger-Teton National Forest from 1992-2013



Common Theme – Clustering of points in space and time

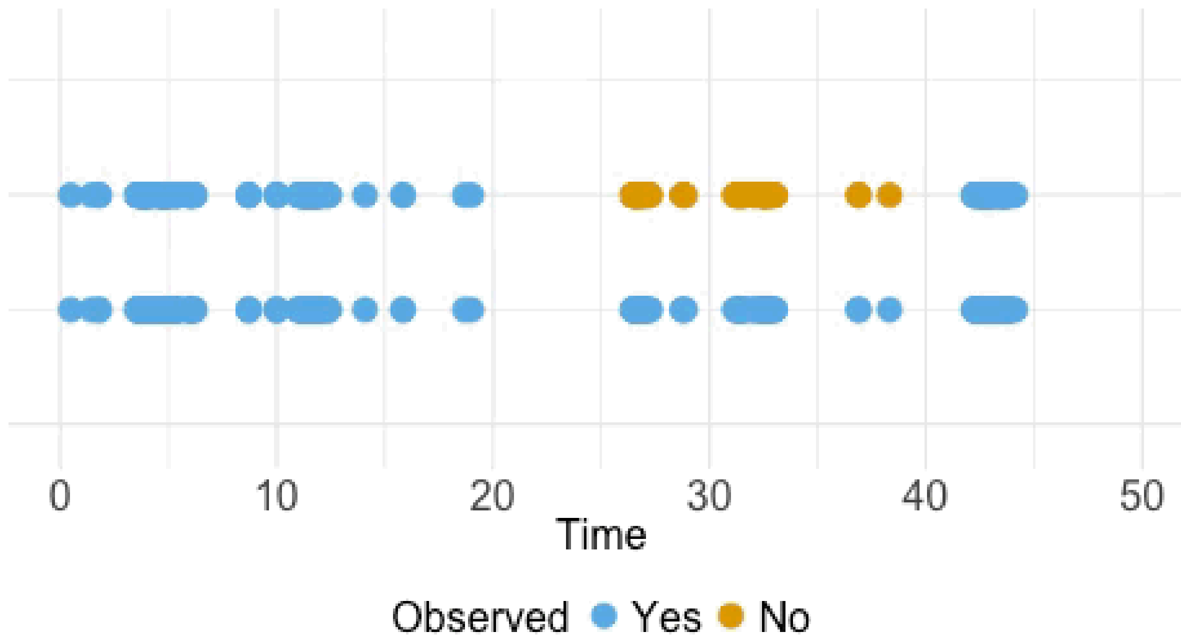
Motivation

Many surveillance applications require human interaction to interpret events

Exploitation elusive – large data sets with missing time (or space) histories

- Seismic sensors down
- Missing records of terrorism events

Simulated temporal point processes with sensors down from 20 to 40 seconds

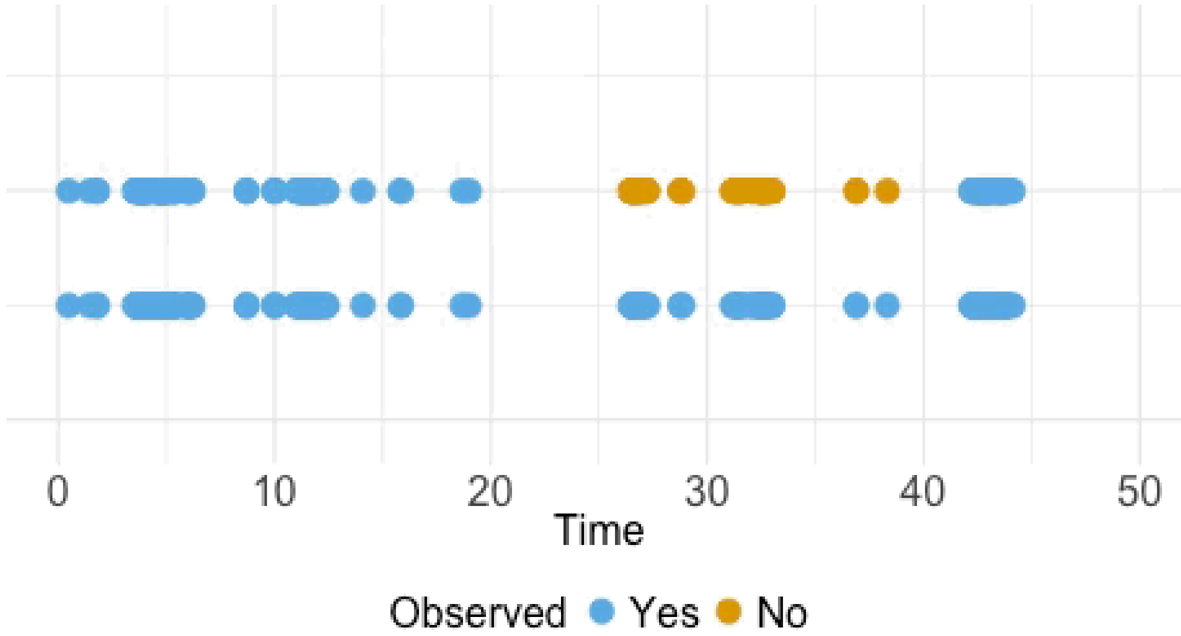


Point Processes with Missing Data

Can we model and correlate events that happen in a self-exciting process with missing time histories?

Self-Exciting? Events cluster in time and space (a.k.a. Hawkes process)

Simulated temporal point processes with sensors down from 20 to 40 seconds



Bayesian Approach to Missing Data

- Observed event times t_{obs}, parameters ϕ
- Missing Data t_{miss} = latent parameters
- Want the posterior:

$$\pi(\phi|t_{obs}) \propto \pi(\phi)p(t_{obs}|\phi)$$

(With implicit conditioning on known unobserved interval(s))

- Two Steps in Gibbs Sampler of $\pi(\phi, t_{miss}|t_{obs})$

1. $\pi(\phi|x)$ $\xrightarrow{x=(t_{miss},t_{obs})}$ Complete-data posterior – either using branching structure or conditional intensity
2. $\pi(t_{miss}|\phi, t_{obs})$ $\rightarrow$ Missing data step – propose missing data, accept/reject

1. Complete-Data Hawkes Process

- A temporal point process $N(t)$ is characterized by its conditional intensity

$$\lambda(t) = \lim_{\Delta t \downarrow 0} (E[N\{(t, t + \Delta t)\} | \mathcal{H}_t] / (\Delta t))$$

- Simplified 'Hawkes' process form with exponential decay:

$$\lambda(t) = \mu + \alpha \sum_{k: t_k < t} g(t - t_k)$$

In general:

$$\mu(t) + \sum_{t_k < t} \alpha(\kappa_i) g(t - t_i; \kappa_i)$$

- Parameters:
  - μ - Immigrant Intensity
  - α - Total offspring intensity
  - $g(t) = \beta \exp(-\beta t)$ - Normalized offspring intensity

1. Complete Data Likelihood

- Observed data $x = (t_1, \ldots, t_n)$ on $[0, T)$

$$p(x|\phi) = \left(\prod_{i=1}^{n} \lambda(t_i|\mathcal{H}_{t_i}) \right) \exp(-\Lambda^*(T))$$

$$\Lambda^*(t) = \int_0^t \lambda^*(s|\mathcal{H}_s)\, ds = M(t) + \alpha \sum_{k:t_k<t} G(t - t_k), \quad M(t) = \int_0^t \mu(s)\, ds$$

- Known as the 'conditional intensity formulation' of likelihood
- MLE is numerically unstable (Veen and Schoenberg 2008)
- Rasmussen (2013): Complete-data Bayesian models via MCMC

2. Missing Data Step

- Assume $[T_1, T_2]$ is the unobserved interval (WLOG)
- Proposal for missing data
  - Conditional distribution of data given history up to time T_1

$$p(t_{miss}|\phi, x_{T_1}) \propto p(x_{T_2}|\phi)$$

- Simulated using thinning method developed by (Ogata 1981)
- MH ratio for missing data

$$H_t = \frac{p(\tilde{x}|\phi)}{p(x|\phi)} \frac{p(x_{T_2}|\phi)}{p(\tilde{x}_{T_2}|\phi)}$$

$x = (t_{miss}, t_{obs})$ t_{miss} - current missing data

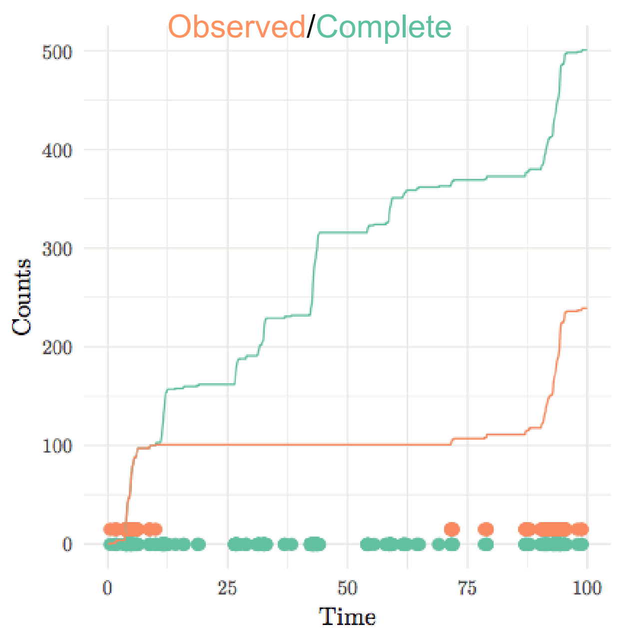
$\tilde{x} = (\tilde{t}_{miss}, t_{obs})$ $\tilde{t}_{miss}$ - proposed missing data

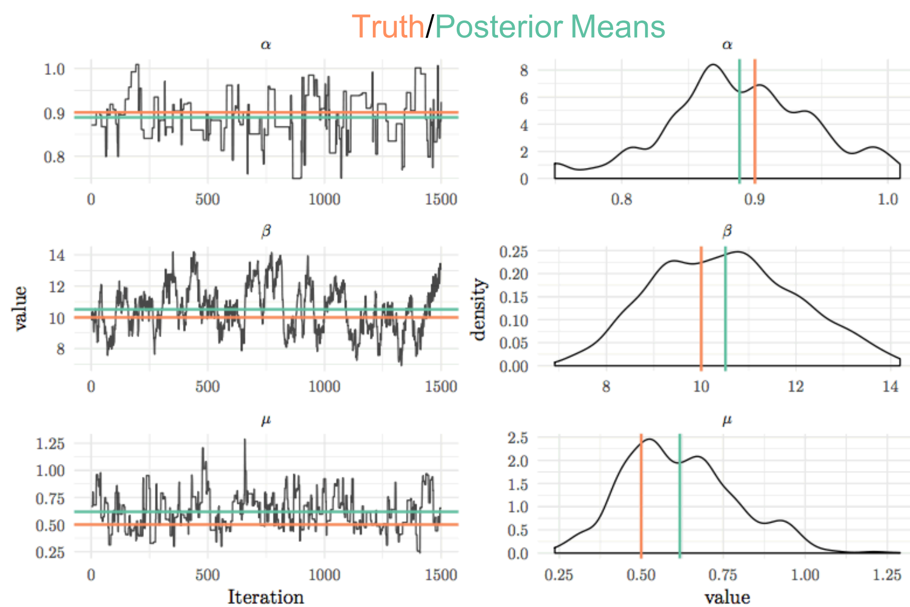
x_{T_j}, $\tilde{x}_{T_j}$ Current, proposed data up to time T_j

Simulated Results

- Complete-data posterior:
  - Likelihood known, specify priors, apply MH-within-Gibbs

$$\mu \sim Gamma(\alpha_\mu, \beta_\mu), \quad \alpha \sim U(l_\alpha, u_\alpha), \quad \beta \sim U(l_\beta, u_\beta)$$



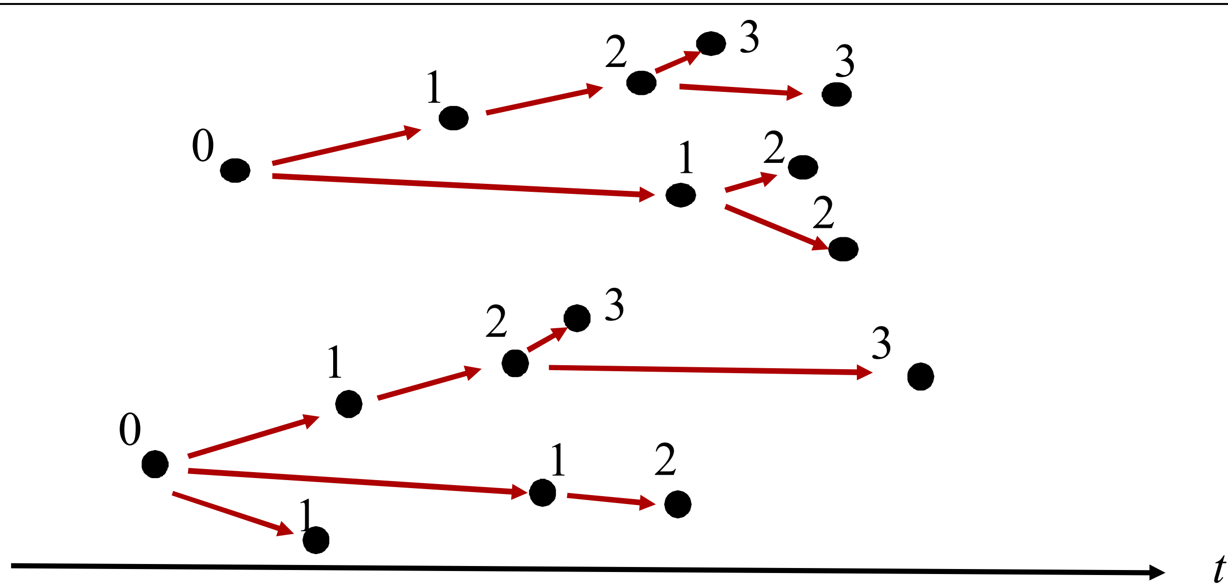


- Augment with missing data step
- Efficiency gains by considering the branching structure of the process (Rasmussen 2013)

Branching Structure

1. The parents I follow a Poisson process with intensity μ
2. Each parent $t_i \in I$ generates a cluster, C_i, where the clusters are assumed to be independent
3. A cluster C_i consists of points of offspring with the following structure: Generation 0 consists of the parents. Recursively, each t_i in generation l generates offspring of generation $l + 1$ from a Poisson process with intensity function $\alpha g(t - t_i)$
4. The process, is the union of all the clusters

Depiction of Branching Structure with 2 parents



MCMC with Branching Structure

- Let $Y = \{y_i\}$ denote the branching structure

$y_i = 0$ means t_i is a parent

$y_i = j$ means t_i is an offspring of t_j

- Partition the arrival times

$$S_j = \{t_i; y_i = j\},\ 0 \leq j < n$$

- 'Cluster process formulation' of likelihood

$$p(x|\phi, Y) = \exp(-\mu T)\mu^{|S_0|} \prod_{i=1}^{n} \left(\exp(-\alpha G(T - t_i|\beta))\alpha^{|S_i|} \prod_{t_j \in S_i} g(t_i - t_j|\beta) \right)$$

MCMC with Branching Structure

- All full conditionals now include conditioning on branching structure
- Include a step to sample the branching structure:
  - Assume uniform prior on branching structure

$$p(Y_i = j|x, \phi) = \begin{cases} \frac{\mu}{\lambda(t)} & \text{if } j = 0 \\ \frac{\alpha g(t_j)}{\lambda(t)} & \text{if } j \in 1, 2, \dots, i-1 \end{cases}$$

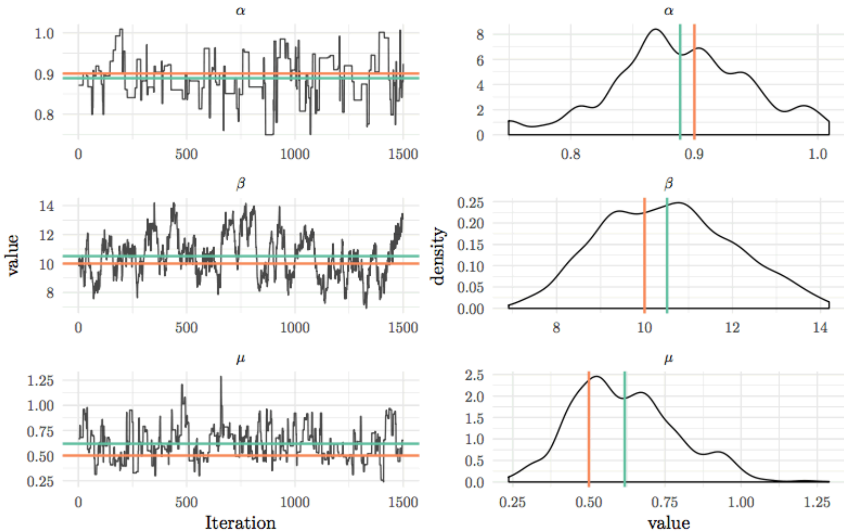
- Advantages:
  - Reduced likelihood computational burden in MH ratios
  - Inference on the latent branching structure (like missing data)
  - Numerical stability of likelihood (Veen and Schoenberg 2008)
  - More efficient convergence

Simulated Results

- Posterior given complete data:
  - Likelihood known, specify priors, apply MH-within-Gibbs

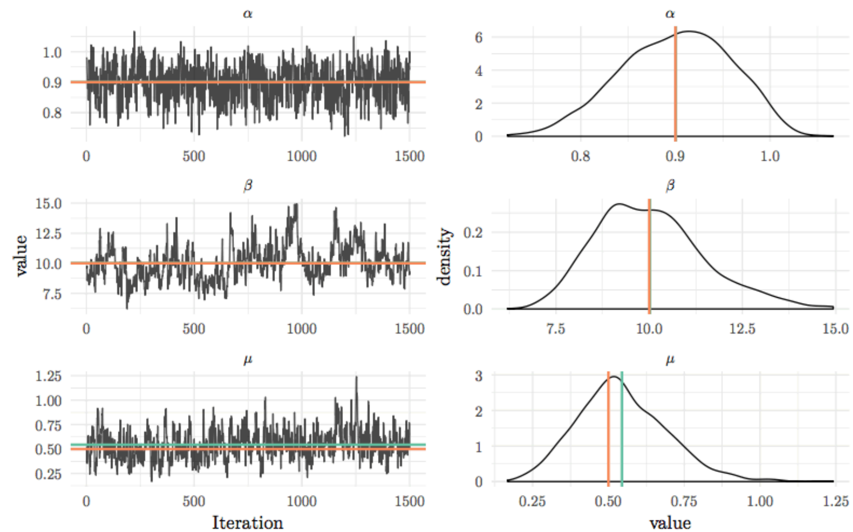
$$\mu \sim Gamma(\alpha_\mu, \beta_\mu), \quad \alpha \sim U(l_\alpha, u_\alpha), \quad \beta \sim U(l_\beta, u_\beta)$$

Truth/Posterior Means



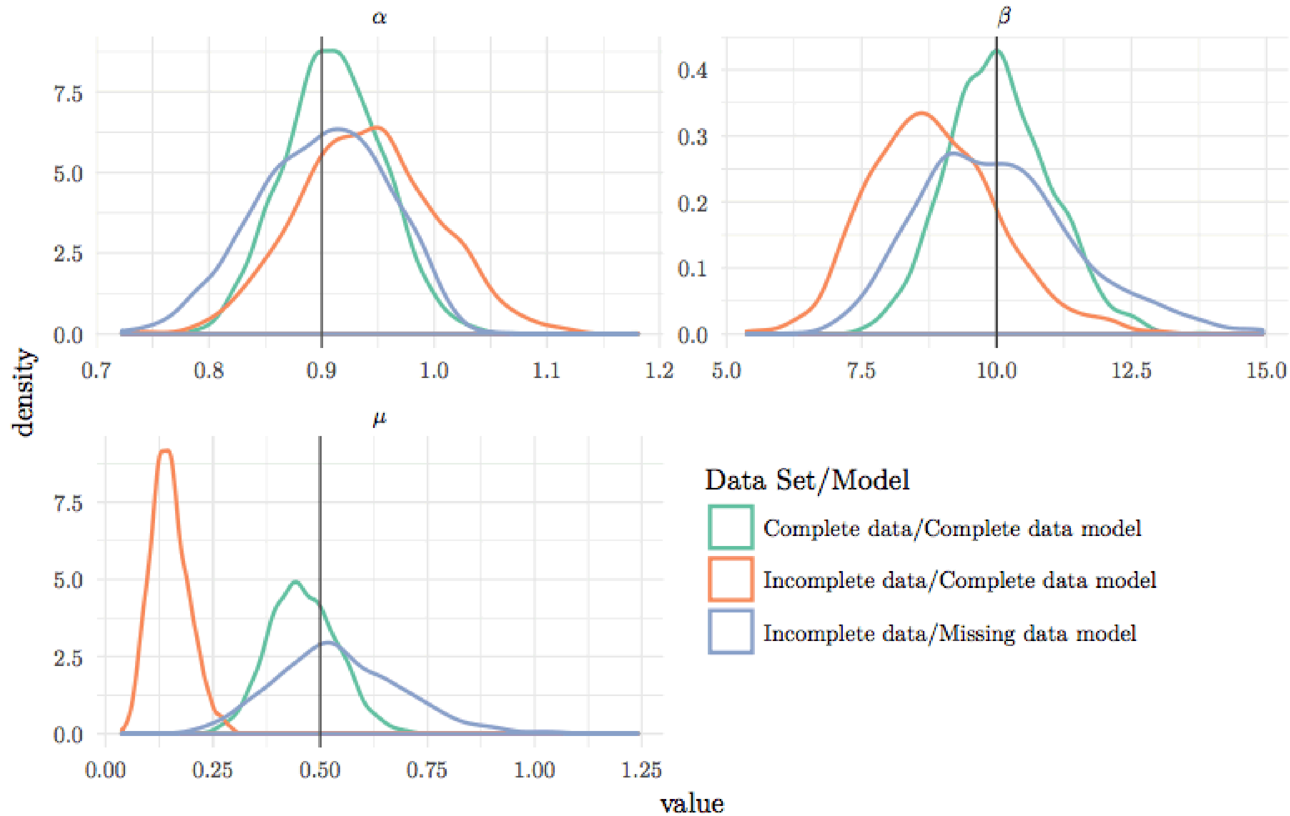
Using conditional intensity formulation

Truth/Posterior Means



Using cluster process formulation

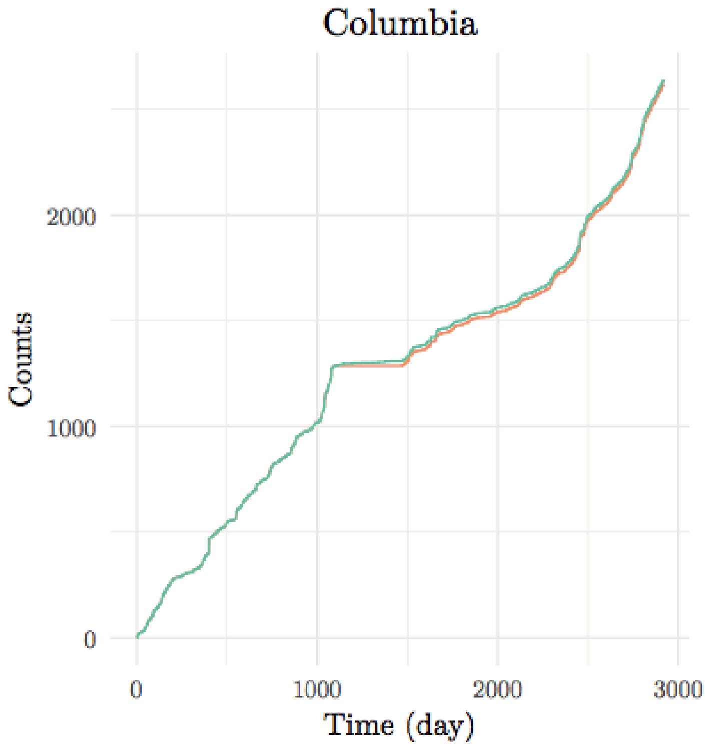
Ignore missing data?



- Bias posteriors (especially for the mean) if missing data is ignored (Orange)
- Larger uncertainty with missing data compared to complete data

Global Terrorism Database

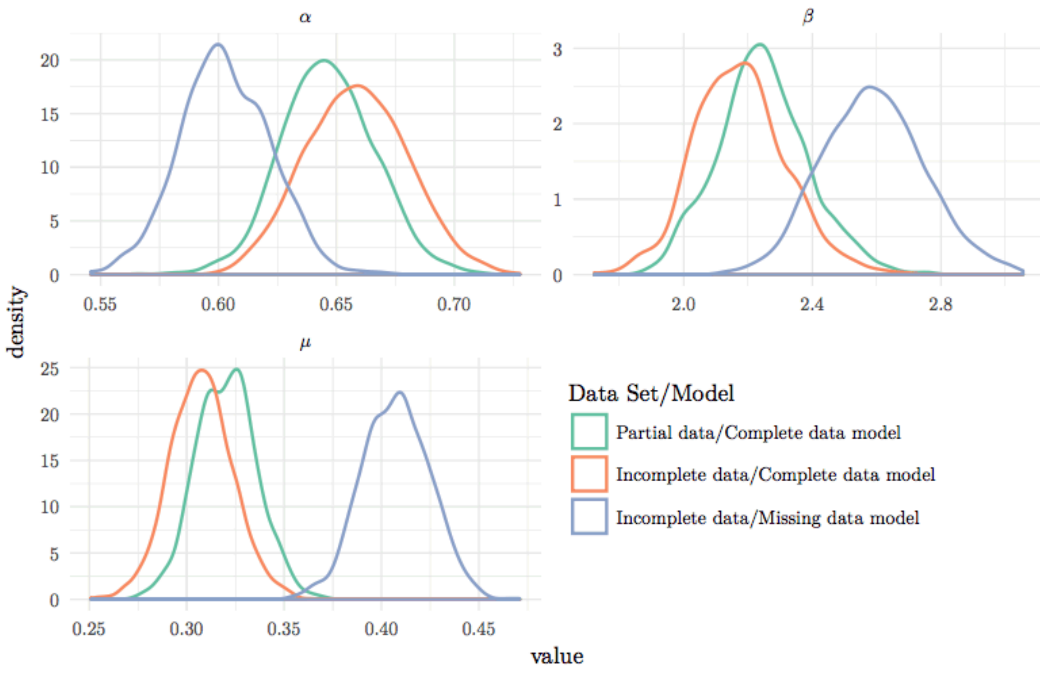
- The Global Terrorism Database (2017) (GTD) is an open-source database including information on terrorism events around the world from 1970-2015
- Look at 1990-1997 in Columbia - multiple problems with guerrillas, paramilitaries, and narcotics
- The database is missing records for the entire year of 1993



- A partial recovery of 21 events during 1993 is available (green)
- Safe to assume there were many more events

Global Terrorism Database

- Parameter estimates accounting for missing data increase, number of estimated events on order of those recovered in the data set



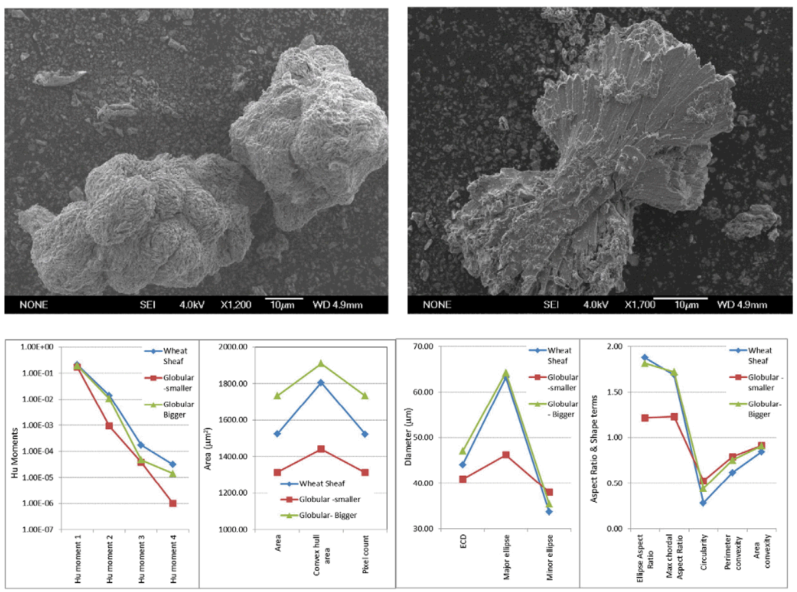
- Number of events in 1993
  - 95% CI: (69,201)
- Slightly below database's estimate of 225 events

NUCLEAR FORENSICS RESEARCH

Nuclear Forensics

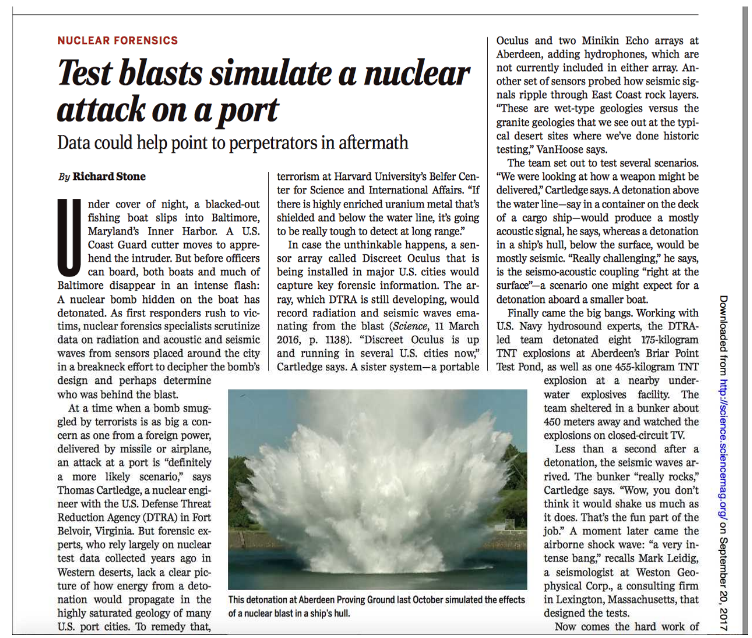
- U.S. Government is conducting research in nuclear forensics
- Two main objectives

Attribute interdicted material to a source.
E.g. Understand how it was produced



C. Anderson-Cook et al. / Chemometrics and Intelligent Laboratory Systems 149 (2015) 107–117

Understand the detonation device



NUCLEAR FORENSICS

Test blasts simulate a nuclear attack on a port

Data could help point to perpetrators in aftermath

By **Richard Stone**

Under cover of night, a blacked-out fishing boat slips into Baltimore, Maryland's Inner Harbor. A U.S. Coast Guard cutter moves to apprehend the intruder. But before officers can board, both boats and much of Baltimore disappear in an intense flash: A nuclear bomb hidden on the boat has detonated. As first responders rush to victims, nuclear forensics specialists scrutinize data on radiation and acoustic and seismic waves from sensors placed around the city in a breakneck effort to decipher the bomb's design and perhaps determine who was behind the blast.

At a time when a bomb smuggled by terrorists is as big a concern as one from a foreign power, delivered by missile or airplane, an attack at a port is "definitely a more likely scenario," says Thomas Cartledge, a nuclear engineer with the U.S. Defense Threat Reduction Agency (DTRA) in Fort Belvoir, Virginia. But forensic experts, who rely largely on nuclear test data collected years ago in Western deserts, lack a clear picture of how energy from a detonation would propagate in the highly saturated geology of many U.S. port cities. To remedy that,

terrorism at Harvard University's Belfer Center for Science and International Affairs. "If there is highly enriched uranium metal that's shielded and below the water line, it's going to be really tough to detect at long range."

In case the unthinkable happens, a sensor array called Discreet Oculus that is being installed in major U.S. cities would capture key forensic information. The array, which DTRA is still developing, would record radiation and seismic waves emanating from the blast (*Science*, 11 March 2016, p. 1138). "Discreet Oculus is up and running in several U.S. cities now," Cartledge says. A sister system—a portable

Oculus and two Minikin Echo arrays at Aberdeen, adding hydrophones, which are not currently included in either array. Another set of sensors probed how seismic signals ripple through East Coast rock layers. "These are wet-type geologies versus the granite geologies that we see out at the typical desert sites where we've done historic testing," VanHoose says.

The team set out to test several scenarios. "We were looking at how a weapon might be delivered," Cartledge says. A detonation above the water line—say in a container on the deck of a cargo ship—would produce a mostly acoustic signal, he says, whereas a detonation in a ship's hull, below the surface, would be mostly seismic. "Really challenging," he says, is the seismo-acoustic coupling "right at the surface"—a scenario one might expect for a detonation aboard a smaller boat.

Finally came the big bangs. Working with U.S. Navy hydrosound experts, the DTRA-led team detonated eight 175-kilogram TNT explosions at Aberdeen's Briar Point Test Pond, as well as one 455-kilogram TNT explosion at a nearby underwater explosives facility. The team sheltered in a bunker about 450 meters away and watched the explosions on closed-circuit TV.

Less than a second after a detonation, the seismic waves arrived. The bunker "really rocks," Cartledge says. "Wow, you don't think it would shake us much as it does. That's the fun part of the job." A moment later came the airborne shock wave: "a very intense bang," recalls Mark Leidig, a seismologist at Weston Geophysical Corp., a consulting firm in Lexington, Massachusetts, that designed the tests.

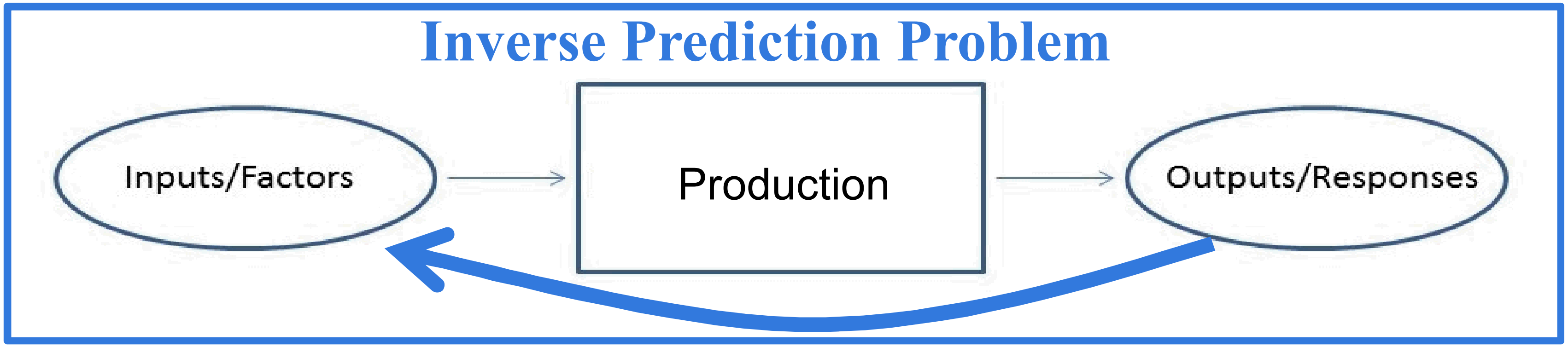
Now comes the hard work of

This detonation at Aberdeen Proving Ground last October simulated the effects of a nuclear blast in a ship's hull.

Stone, Richard. *Science* **355** (6328), 897.

Attribution of Material

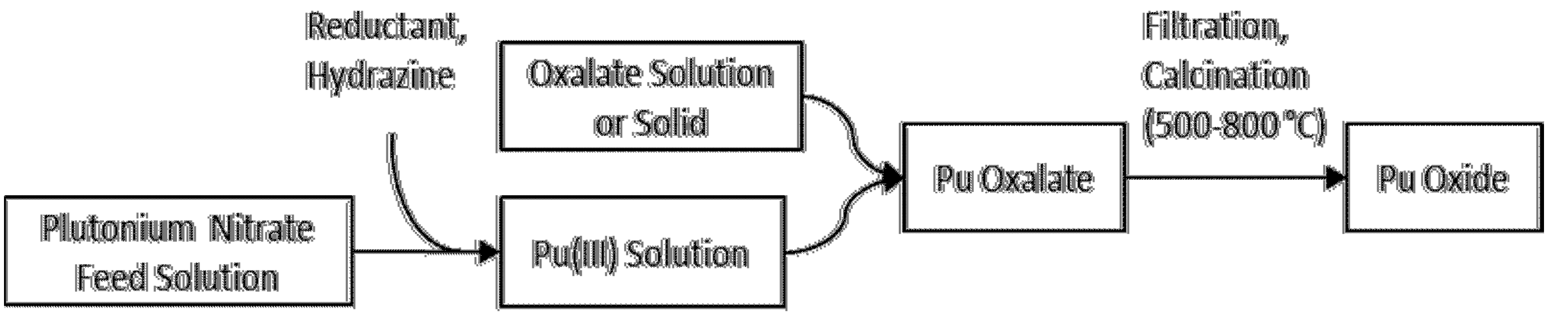
- Mining of historical production databases
  - Mostly U.S., limited variation, missing data
  - Machine learning – where was each piece of material produced?
- New Experiments
  - How do variations in inputs affect outputs? Production data doesn't help much.
  - DOE for inverse prediction?
  - Inverse prediction methods



Pu Signatures Project

- Objective:
  - Produce Pu oxalate, measure characteristics, predict processing conditions

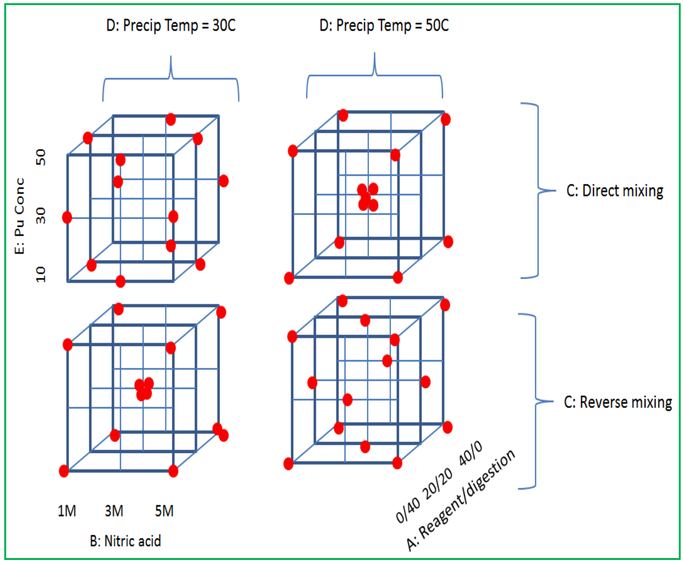
Example: Pu Signatures project – production of Pu(III) oxalate

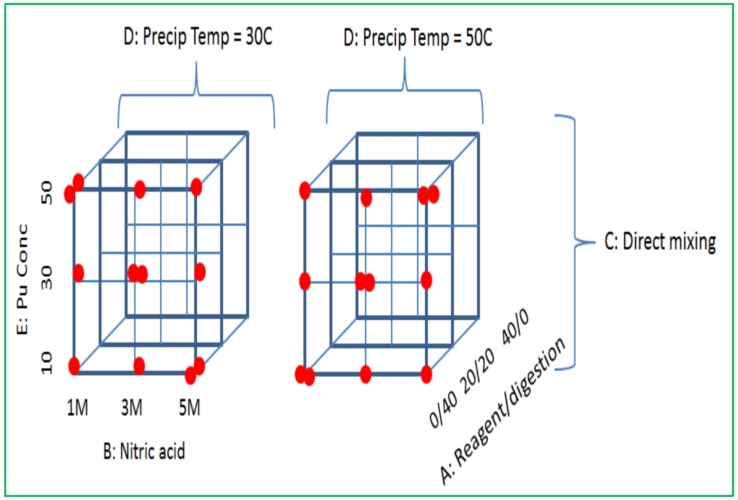


Experimental Design

- Little/no research for inverse problem
- Philosophy: Span factor space of interest, allow for accurate forward models, provide rich training data set for direct inverse models
  - I-optimality, considering span, replication, feasibility

Representation of Design: 6 factors, 2 separate designs, some settings not feasible





Approaches for Inverse Modeling

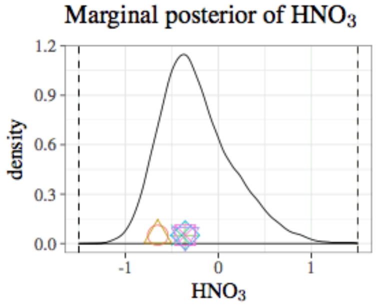
Factors = Processing conditions: $X_1, X_2, \ldots, X_p$
Responses = Measurements of processed material: $Y_1, Y_2, \ldots, Y_q$
Signature = Complete set of responses, $Y = \{Y_1, Y_2, \ldots, Y_q\}$

- **Causal Modeling Approach**
  - Forward models
    $$Y_i = f_i(X_1, X_2, \ldots, X_p)$$
  - 'Invert' using new signature to "predict" factor values

Example: classical – minimize an objective function

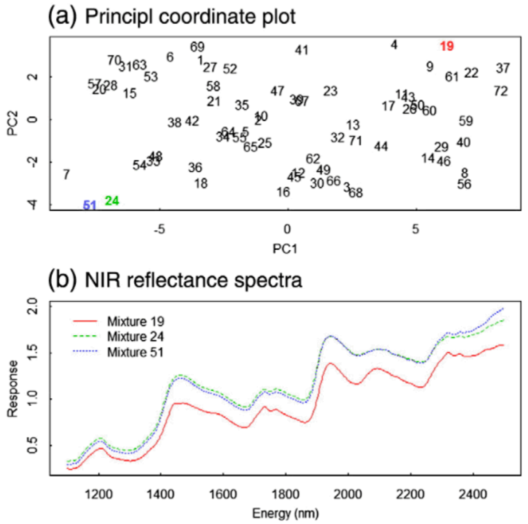
$$\widehat{X}^* = \text{argmin}_X \sum_{j=1}^{q} (\widehat{Y}_j - Y_j^*)^2$$

Example: Bayesian – posterior for X^*



Inverse "Soft" Modeling Approach

- Build supervised learning models
  $$X_j = g_j(Y)$$
- Predict value of j^{th} factor directly
  $$\hat{X}_j^* = g_j(Y^*)$$



C. Anderson-Cook et al. / Chemometrics and Intelligent Laboratory Systems 149 (2015) 107–117

Inverse Prediction Research

Two Main Objectives (so far)

1. Down-select a set of responses
   - Could take many measurements
   - Costly, time-consuming (after all, it's radioactive material)
   - Limited amount of material? We don't know what we'll get
   - Want the most informative/discriminating set of responses
2. Confidence in predictions
   - Idea is to inform criminal investigations
   - Large number of ways to produce material – training data is limited
   - How can we tell if the predictions will be informative in an actual interdiction?

Down-selecting a set of Responses

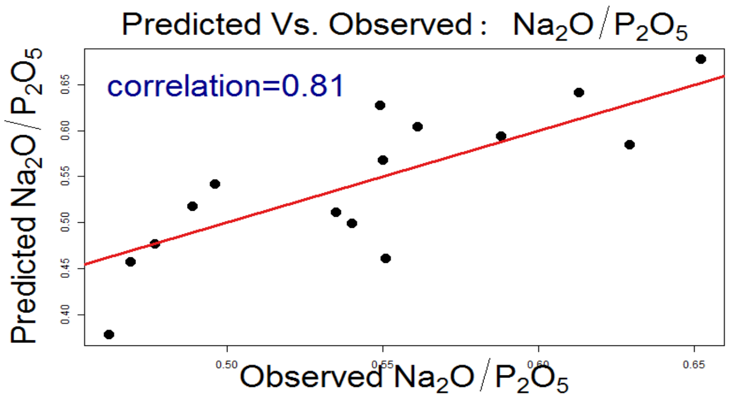
- Strategy to down-select an *informative/discriminating* subset of responses among a candidate set
  - Responses can yield accurate, precise, and unconfounded predictions of the factors
- High level idea (Details if time)
  - Fit forward models $Y_i = f_i(\beta_i; X) + \epsilon_i, i = 1, 2, \ldots, q$
  - Need to estimate X^* given new Y^*
  - Swith role of β and X using local linearity to estimate prediction variance $V(X)$
  - Choose a set with a
    - reasonable number of responses
    - small prediction variance across space of interest.

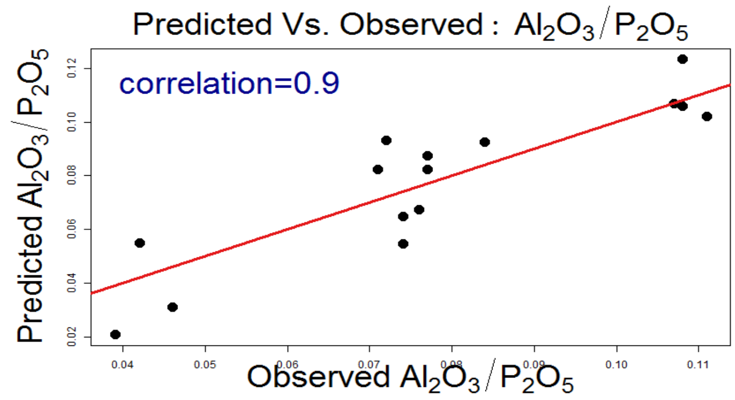
Glass Composition Example

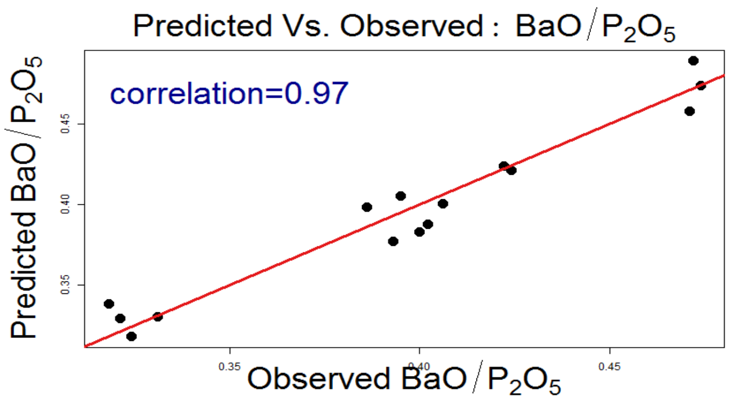
- Study to investigate how glass properties varied as a function of composition
  - Constituents are mole ratios: $X_1 = Na_2O/P_2O_5, X_2 = BaO/P_2O_5, \; X_3 = Al_2O_3/P_2O_5$
- ***Goal:*** Predict constituents based on six glass properties ($i = 1, \dots, 6$)
- Simple linear models: $\hat{Y}_i = \hat{\beta}_{i0} + \hat{\beta}_{i1}X_1 + \hat{\beta}_{i2}X_2 + \hat{\beta}_{i3}X_3$
- Best forward models (using R^2) are of **density** and **refraction**
  - Don't depend on Na_2O

Property: $i = 1, \dots, 6$	$\hat{\beta}_0$	$\hat{\beta}_1 \; (Na_2O)$	$\hat{\beta}_2 \; (Ba0)$	$\hat{\beta}_3 \; (Al_2O_3)$	$\hat{\sigma}_\varepsilon$	R^2
[redacted]						
2. Softening Temperature	**393(16)**	**-105(25)**	**----**	**695(63)**	**5.7**	**0.93**
[redacted]						
4. Crystallization Temperature	**571(29)**	**-220(489)**	**----**	**710(147)**	**14.7**	**0.74**
5. Density	**2.5(0.02)**	**----**	**1.1(0.05)**	**0.5(0.12)**	**0.01**	**0.97**
6. Refraction	**1.5(0.003)**	**0.01(0.004)**	**0.08(0.005)**	**0.1(0.012)**	**0.001**	**0.97**

Prediction of Glass Composition





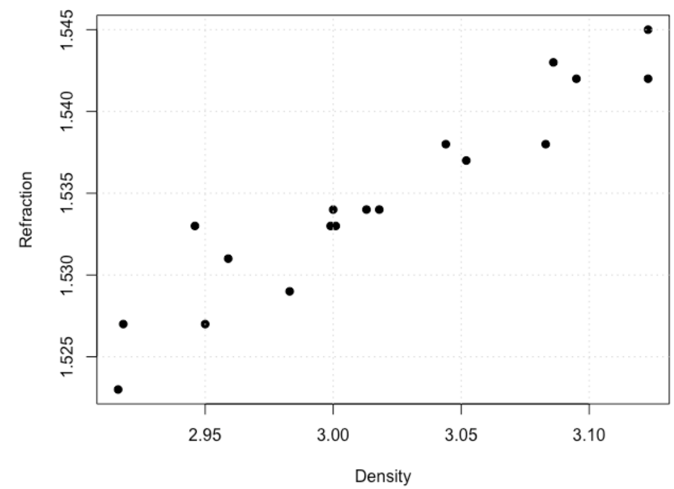


- Good for prediction $Ba0$ and Al_2O_3, not as good for Na_20
  - Strongest models don't depend on Na_20
  - Intuition: need strong forward models for inverse prediction
- Density is responsible for the precise predictions of BaO – Barium is very dense compared to other constituents

Average Prediction Variance

- Switch roles: $Y^* - \hat{B}_0 = \hat{\mathrm{B}}\, X^*$
- Estimated prediction variance: $\left(\hat{\mathrm{B}}\hat{V}^{-1}\hat{\mathrm{B}}\right)^{-1}$ - Can be (approximately) generalized

Subset	$\sqrt{Var_{avg}(X_1)}$ (Na_2O)	$\sqrt{Var_{avg}(X_2)}$ (BaO)	$\sqrt{Var_{avg}(X_3)}$ (Al_2O_3)
[redacted]	[redacted]	0.013	[redacted]
Excluding Density	0.08	0.02	0.019
[redacted]	[redacted]	[redacted]	[redacted]



- Excluding density results in ~1.5 times increase in the root prediction variance of X_2
- Multivariate response is less informative for predicting X_2 if density is excluded
- Excluding refraction is not as detrimental – despite a good forward model (redundancy)

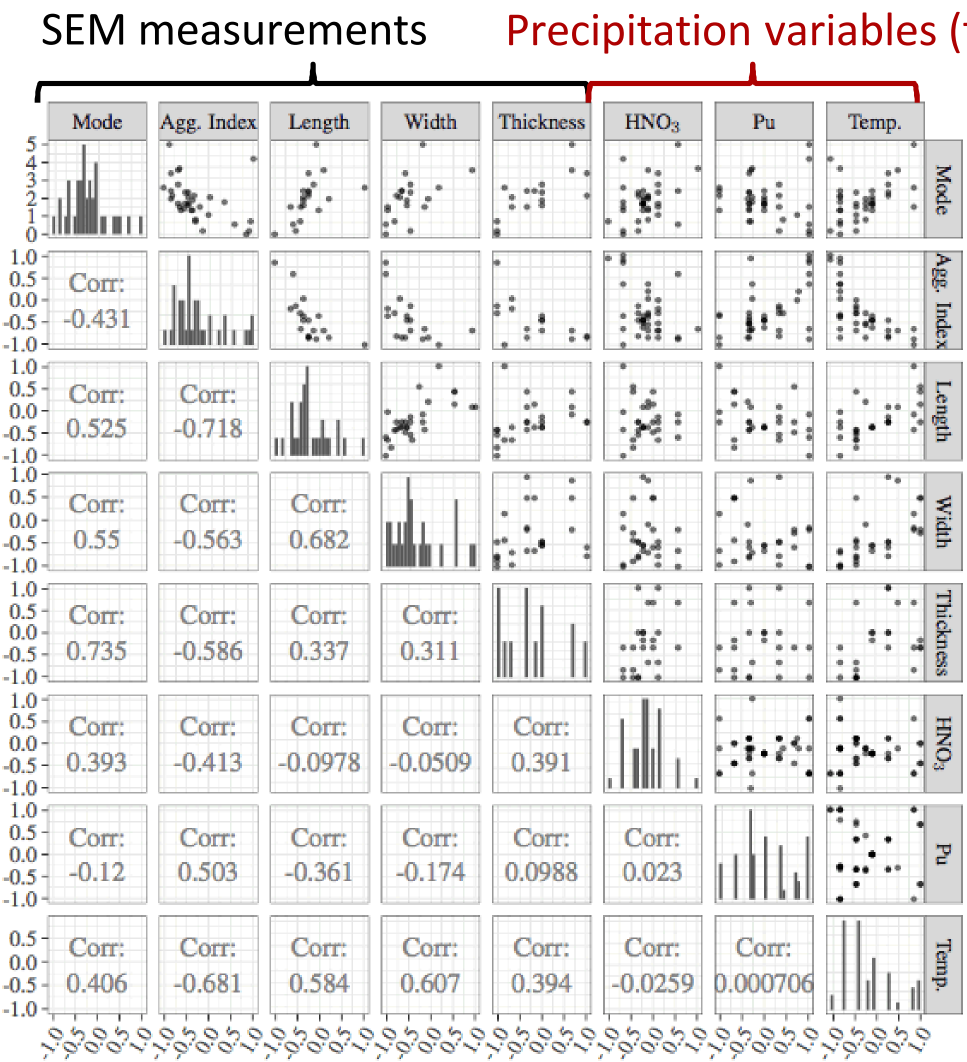
Confidence in Predictions

- Idea is to inform criminal investigations
- Large number of ways to produce material – training data is limited
- How can we tell if the predictions we make trained on the data we have will be informative in an actual interdiction?

<u>Main Idea</u>

- Try many methods
- Don’t just use the best one – but look for consistency
- Consistency builds confidence predictions are robust to the various assumptions of each method

Pu Production Data

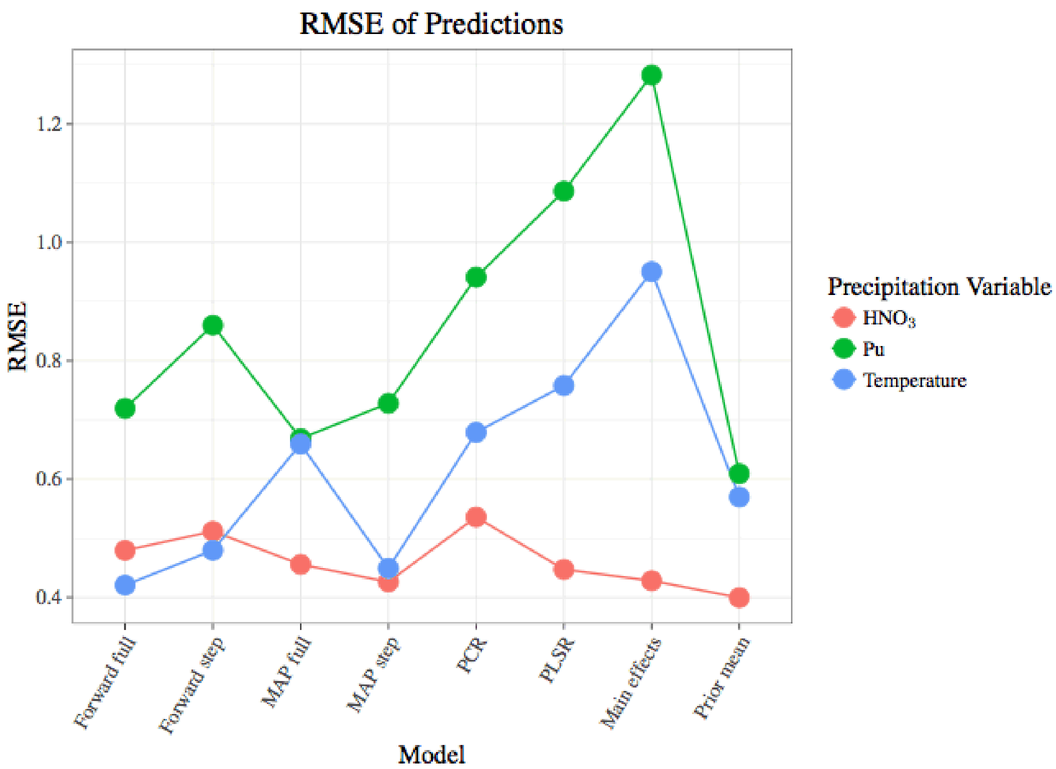


Historical Pu production data

- 3 precipitation variable
- 5 Scanning Electron Microscope measurements
  - Ave. length, width, height, etc. of particle size distributions
- Mostly weak relationships between factors and responses
- Strong correlations between responses
- Poor predictive performance likely
- Shows in large variation of performance between several methods

Burney, G. A. and Smith, P. K. (1984), “Contolled PuO$_2$ Particle Size from Pu(III) Oxalate Precipitation,”
Tech. rep., Savannah RiverLaboratory Technical Report DP-1689.

Large Variation between methods



<u>Methods</u>

- Classical and Bayesian linear models, PCR, PLSR

<u>Observations</u>

- Prior mean (i.e. no modeling) predictions do well.
- Best performing method (lowest RMSE) is not consistent across precipitation variables.

Large Variation between methods

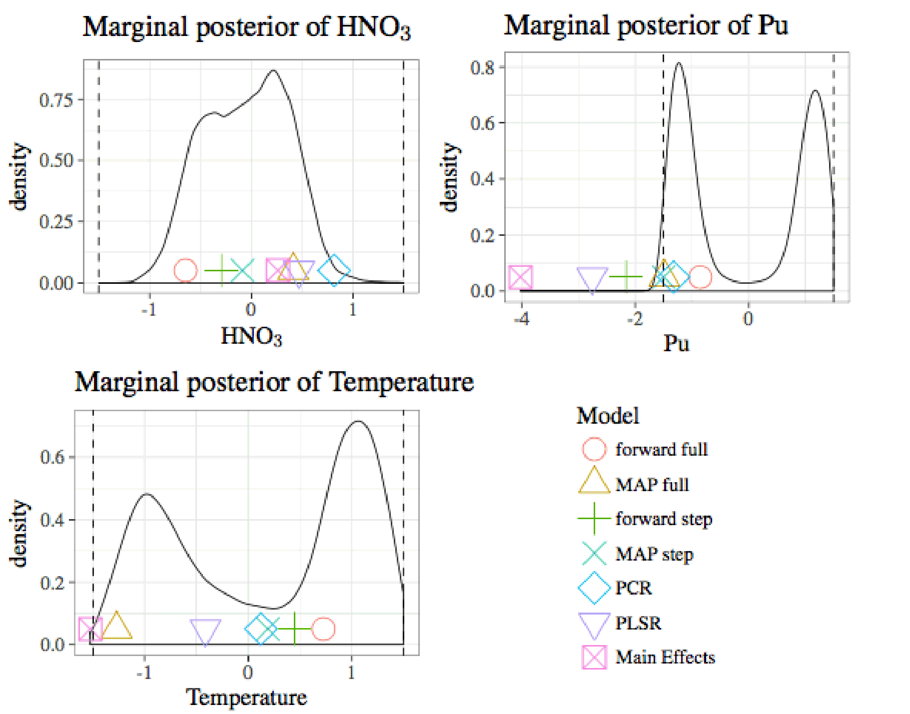


Figure 8: Marginal posteriors of each of the precipitation parameters. Predictions under each of the other models are indicated by points along the horizontal axis. The vertical dashed lines denote the range of the uniform prior distribution.

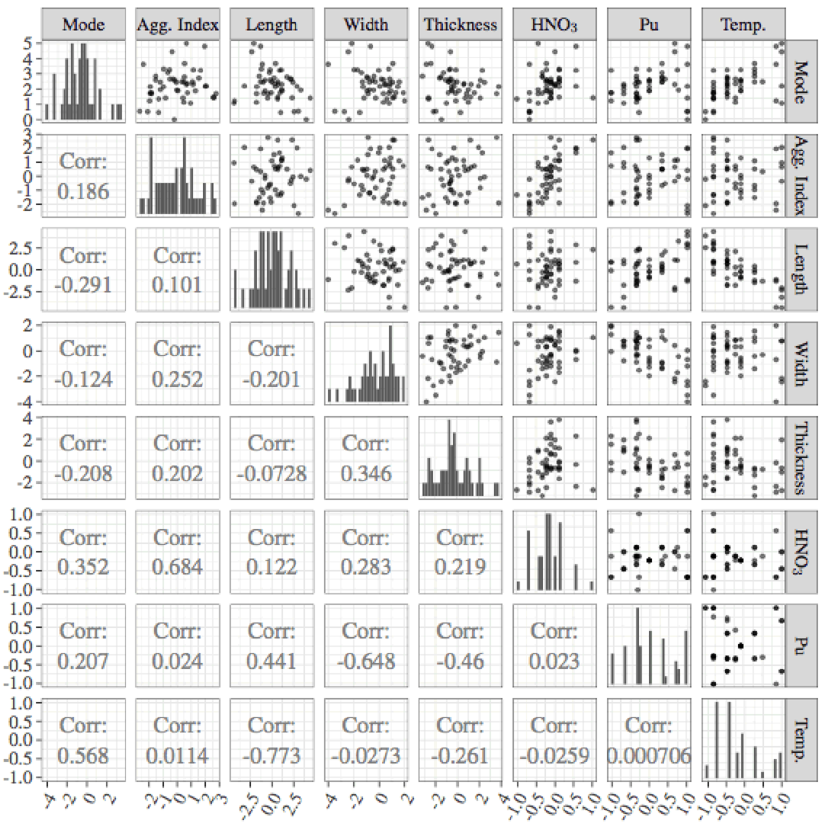
Predictions for a single holdout sample

Observations

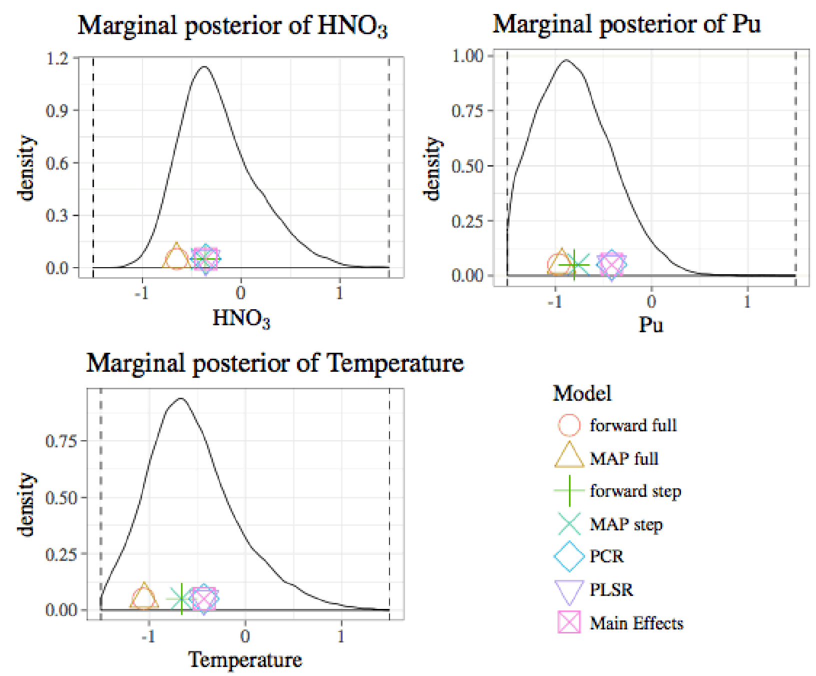
- Multimodal posteriors
- Large variation in point estimates
- Bottom line:
  - No agreement between various methods
  - Predictions are ambiguous
  - To improve prediction
    - Search for a better set of responses
    - Collect more data
      - mostly one-factor-at-a-time with many levels
      - fine for original purpose

What if…responses were better

Stronger relationships, less correlation between responses



Results in more agreement between the methods



Collaborators

- Point Process with missing Data
  - PI: J. Derek Tucker, Lyndsay Shand, Jonathan Lane, Kathy Simonson, John Rowe
- Nuclear Forensics
  - Christine Anderson-Cook, Adah Zhang, Edward Thomas

References

Sandia National Laboratories

Point Process

- Tucker, J.D., Lewis, John, and Lane, Jonathan, “Bayesian Modeling of Self-Exciting Point Processes with Missing Temporal Histories” *In Progress*
- Global Terrorism Database (2017). National Consortium for the Study of Terrorism and Responses to Terrorism (START). url: https://www.start.umd.edu/gtd.
- Ogata, Y. (1981). “On Lewis’ simulation method for point processes.” IEEE Transactions on Information Theory 27(1), 23–31.
- Rasmussen, J. G. (2013). “Bayesian Inference for Hawkes Processes.” Methodology and Computing in Applied Probability 15(3), 623–642.
- Veen, Alejandro, and Frederic P. Schoenberg (2008). "Estimation of space–time branching process models in seismology using an EM–type algorithm." *Journal of the American Statistical Association* 103 (482), 614-624.

Nuclear Forensics

- Thomas, Edward V., Lewis, John R., Anderson-Cook, Christine, Burr, Tom, and Hamada, Michael S (2017). “Selecting an Informative/Discriminating Multivariate Response for Inverse Prediction.” Journal of Quality Technology 49 (3), 228-243.
- Lewis, John, Zhang, Adah, Anderson-Cook. “Comparing Multiple Statistical Methods for Inverse Prediction in Nuclear Forensics Applications” *Under Revision at Chemometrics and Intelligent Laboratory Systems*

Extra Slides

Down-selecting a set of Responses

- Strategy to down-select an *informative/discriminating* subset of responses among a candidate set
  - Responses can yield accurate, precise, and unconfounded predictions of the factors
- Main Idea
  - Estimate prediction variance given a set of responses
  - Choose a set with a
    - reasonable number of responses
    - small prediction variance.
- Depends on an assumed 'forward model' for each of q responses related to p causal factors

$$Y_i = f_i(\beta_i; X) + \epsilon_i, \qquad i = 1, 2, \dots, q$$

Y_i – i^{th} response, β_i – model parameters, X – factors, ϵ_i – mean zero error

Predicting X: Least Squares Accounting for Errors

- Estimate each model: $Y_i \approx f_i(\hat{\beta}_i; X), i = 1, \dots, q$
- A new observed multivariate response ($Y^* = (Y_1^*, \dots, Y_q^*)^\top$) is used to predict unknown levels of factors X^*

$$Y_i^* = f_i(\beta_i, X^*) + \epsilon_i^*$$

- Find an "optimal" solution $\hat{X}^*$ such that $\hat{Y}_i^* \approx Y_i^*$, $i = 1, \dots, q$ where $\hat{Y}_i^* = f_i(\hat{\beta}_i, \hat{X}^*)$
- Prediction error at candidate solution $\hat{X}$: $d_i = \hat{Y}_i - Y_i^*$ where $\hat{Y}_i = f_i(\hat{\beta}_i, \hat{X})$

$$\hat{X}^* = \text{argmin}_X \, D^\top V^{-1} D$$

$$D = (d_1, \dots, d_q)^\top, \;\; V = \text{cov}(D) \text{ (function of } \hat{X})$$

- Solved iteratively – requires $\hat{\beta}_i$, initial $\hat{X}$, and $\hat{V}$

Accounting for Errors using V

- To estimate V first decompose d_i:

$$d_i = \hat{Y}_i - Y_i^* = f_i(\hat{\beta}_i, \hat{X}) - f_i(\beta_i, X^*) - \epsilon_i^* =$$
$$\lambda_i + \omega_i - \epsilon_i^*$$
$$\lambda_i = f_i(\hat{\beta}_i, \hat{X}) - f_i(\beta_i, \hat{X}) \text{ and } \quad \omega_i = f_i(\beta_i, \hat{X}) - f_i(\beta_i, X^*)$$

- Interpretation of components of d_i
  - $\lambda_i = f_i(\hat{\beta}_i, \hat{X}) - f_i(\beta_i, \hat{X})$: error due to uncertainty in model parameters
  - $\omega_i = f_i(\beta_i, \hat{X}) - f_i(\beta_i, X^*)$: error due to uncertainty in the candidate solution $\hat{X}$
- Assuming properly specified models and unbiased solutions: $E(d_i) = 0$ and

$$V := \text{cov}(D) = V_\lambda(\hat{X}) + V_\omega(\hat{X}) + 2cov_{\lambda\omega}(\hat{X}) + V_\epsilon$$

- V_λ, V_ω - first order approximations, V_ϵ - use residuals
- Simplifying assumptions – each component diagonal, covariance 0.

Variance-Covariance of Prediction

- Assume forward models are
  1. Continuous functions of the factors
  2. Not highly non-linear
- First-order linear approximation to $Y_i^* = f_i(\beta_i, X^*)$ near X^*

$$Y_i^* = f_i(\beta_i, X^*) \approx f_i\left(\beta_i; \hat{X}^*\right) + \sum_{j=1}^{p} J_{ij}\left(\hat{X}_j^*\right)\left(\hat{X}_j^* - X_j^*\right), \text{ where } J_{ij}\left(\hat{X}_j^*\right) = \frac{\partial}{\partial x_j} f_i\left(\beta_i; \hat{X}_j^*\right).$$

- Local linear regression of $\boldsymbol{Y}^*$ on $\hat{J}_{ij}(X^*)$ leads to an estimate of the covariance of $\hat{X}^*$

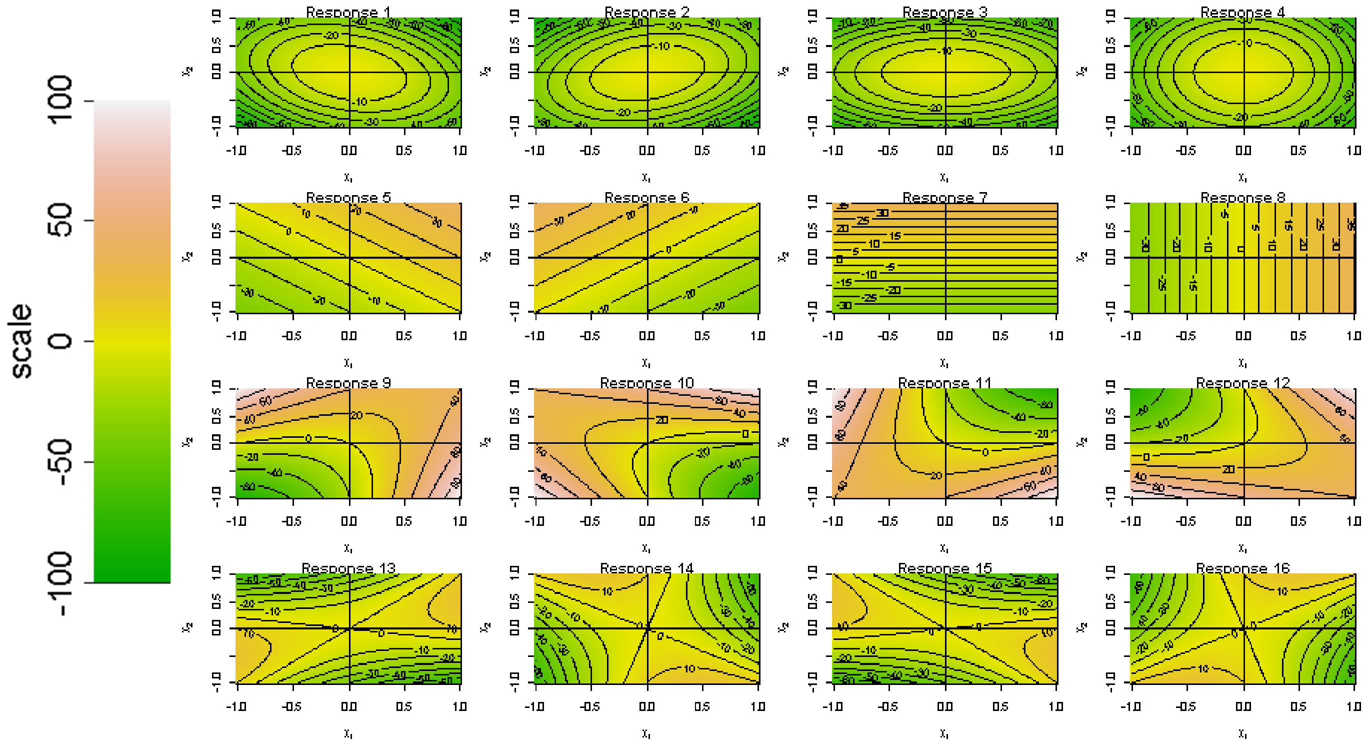
$$\hat{C}_{\hat{X}^*} = \left(\hat{J}^T\left(\hat{X}^*\right)\hat{V}^{-1}\hat{J}\left(\hat{X}^*\right)\right)^{-1}, \quad \text{where } \hat{J}_{ij}\left(\hat{X}^*\right) = \frac{\partial}{\partial x_j} f_i\left(\hat{\beta}_i; \hat{X}^*\right).$$

- Multivariate response is
  - *Informative* if diagonal elements are sufficiently small
  - *Discriminating* if the off diagonal elements are sufficiently small
- Covariance depends on location in the design space

Synthetic Example: 16 Response Surfaces

Goal: Choose a subset of the 16 response surfaces that is ***informative*** (small prediction variance) and ***discriminating*** (not redundant) for prediction of X_1 and X_2

16 Response surfaces



1-4: peaks

5-8 : hillsides

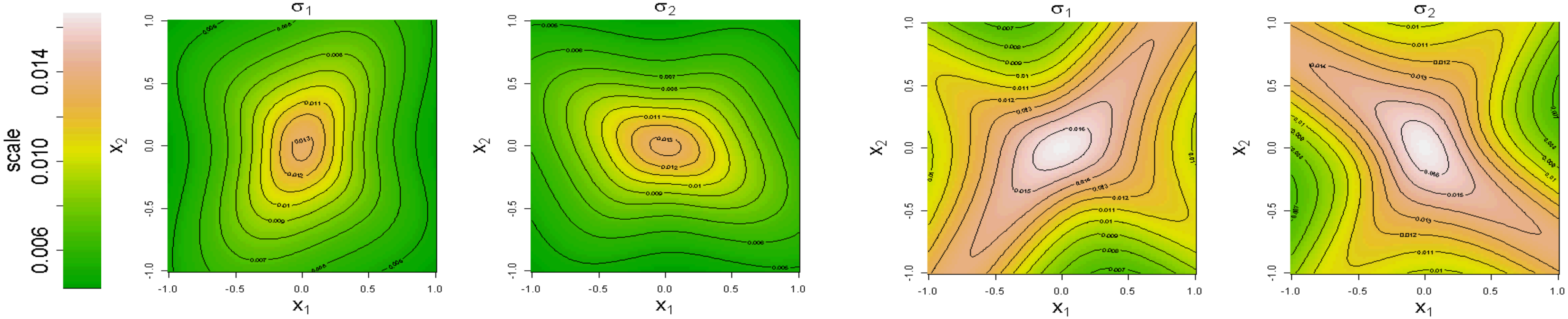
9-12: rising ridges

13-16: saddles

Prediction Variance

- Analytical results using $\hat{C}_{\hat{X}^*} = \left(\hat{J}^T(\hat{X}^*)\hat{V}^{-1}\hat{J}(\hat{X}^*)\right)^{-1}$
- Two candidate sets of responses: $S = \{1,2,\ldots,16\}$ and $S = \{9,10,11,12\}$

$S = \{1,2,\ldots,16\}$ $S = \{9,10,11,12\}$



- Value depends on X^*. Smaller standard deviation across design space when using all 16 responses
- Relative increase using just four responses is small across the design space

Average Prediction Variance

Subset	$\sqrt{Var_{avg}\ \hat{X}_1^*}$	$\sqrt{Var_{avg}\ \hat{X}_2^*}$
[redacted]		
{7,8}	0.0286	0.0286
[redacted]		
{9,10,11,12}	0.0121	0.0121

~ 4x larger (for $\hat{X}_1^*$)

~ 1.6x larger for ¼ of responses

- Set {9,10,11,12} is a good choice for prediction across the space of interest if constraints exist in obtaining new measurements
- Responses in this set complement each other well – i.e. steep contours are present in one or more of the responses throughout the range of interest
- Ideal combination: responses with strong difference across input space, and several responses with different shaped relationships