Managing uncertainty when no one's in charge: Uncertainty propagation in distributed projects

Andy Packard, Michael Frenklach, Ryan Feeley, Pete Seiler

Mechanical Engineering

University of California

Berkeley, CA

Support from NSF-CTS Grant: CTS-0113985 and Network BioInformation Services.

Thanks to Matt Onsum, Arthur Hebert, Adam Arkin.

Connections, John Doyle 50th Birthday Celebration, July 2004

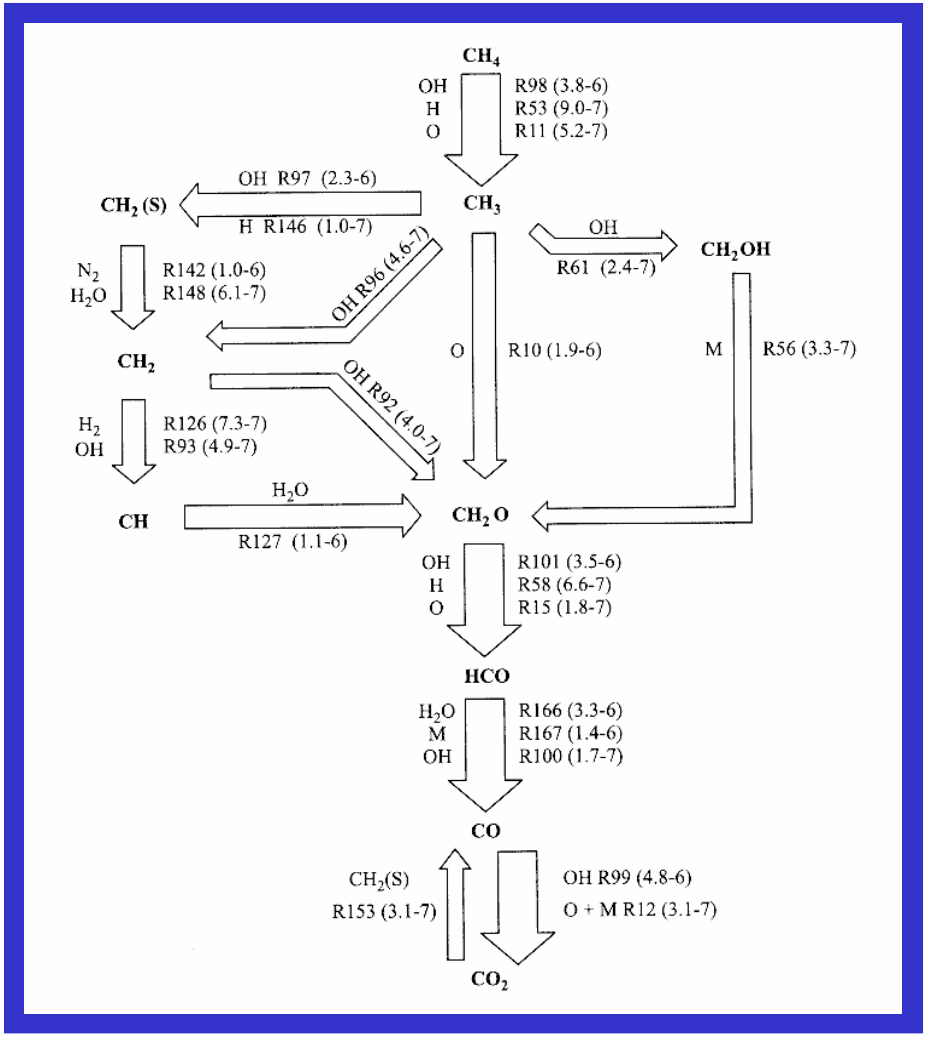
Methane Combustion: $CH_4 + 2\ O_2 \Rightarrow CO_2 + 2\ H_2O$

Methane reaction mechanisms have grown in complexity over time:

(~1970): Less than 15 elementary reactions with 12 species

(~1982): 75 elementary reactions (forward and reverse) with 25 species

(Today): GRI-Mech has 300+ elementary reactions, 53 Species, and 102 “active” parameters. The “gold standard” according to NIST.



Pathway diagram for methane combustion [Turns]

How did GRI-Mech come about?

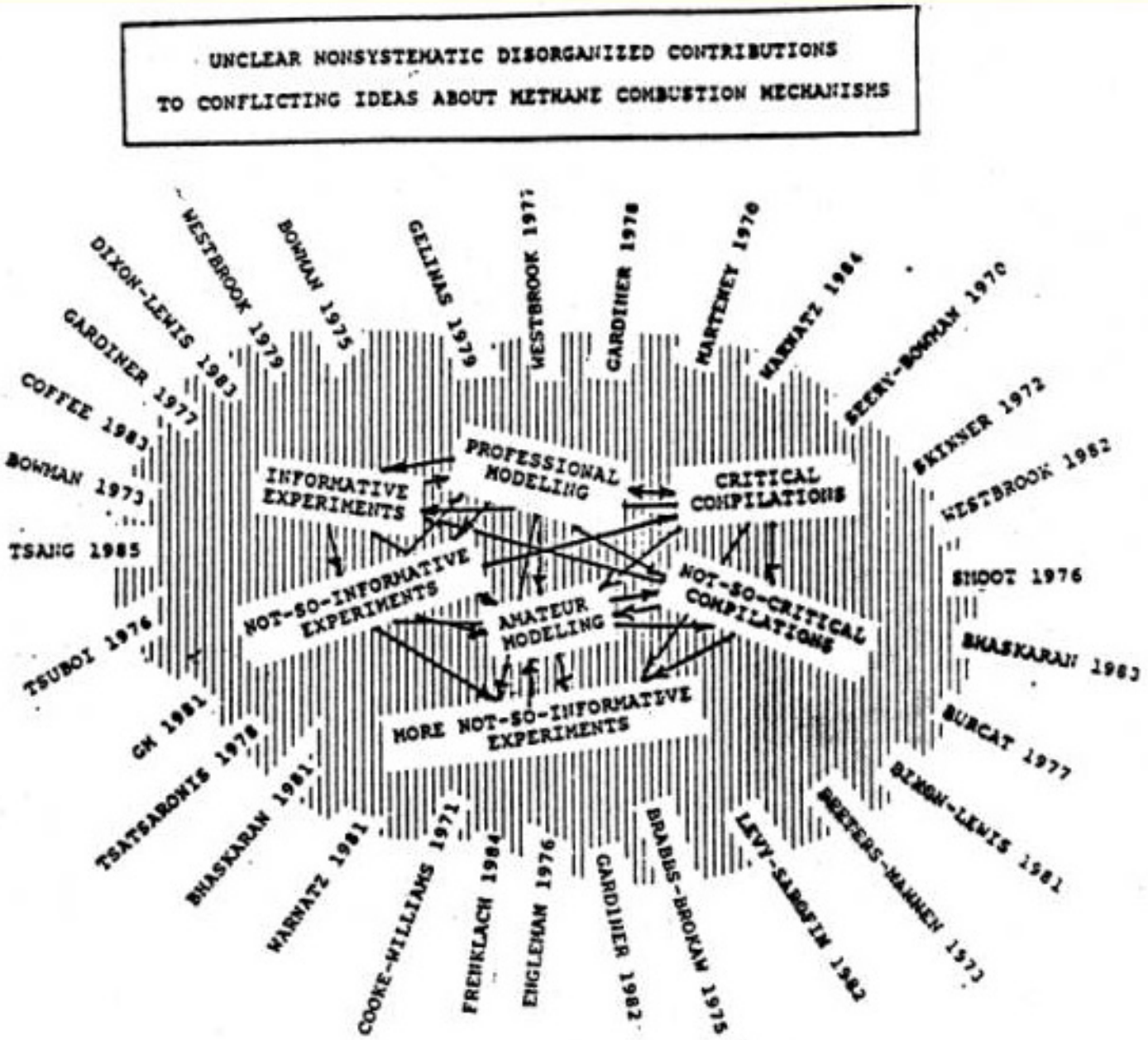
The gold standard GRI-Mechanism is truly the work of many people, but not explicitly working together. How did the successful collaboration occur?

Journal Paper mode?

- informal, “read my paper” form of collaboration

No. Didn’t/Doesn’t work – community tried it, but predictive capability of model did not reliably improve as more high-quality experiments were done. Intrinsically bad uncertainty management and propagation…

- Papers tend to lump modeling and theory, experiments, analysis and convenience assumptions into a concise text-based conclusion
- Paper conclusions are conditioned on additional assumptions necessary to make the conclusion concise.
- Impossible to anonymously “collaborate" since the convenience assumptions are unique to each paper.
- Goals of one paper are often the convenience assumptions of another.
- Difficult to trace the pedigree of a conclusion reached sequentially across papers


UNCLEAR NONSYSTEMATIC DISORGANIZED CONTRIBUTIONS
TO CONFLICTING IDEAS ABOUT METHANE COMBUSTION MECHANISMS
PROFESSIONAL MODELING
INFORMATIVE EXPERIMENTS
CRITICAL COMPILATIONS
NOT-SO-INFORMATIVE EXPERIMENTS
AMATEUR MODELING
NOT-SO-CRITICAL COMPILATIONS
MORE NOT-SO-INFORMATIVE EXPERIMENTS
WESTBROOK 1979
BOWMAN 1975
GELINAS 1979
WESTBROOK 1977
GARDINER 1978
MARTENEY 1970
WARNATZ 1984
SEERY-BOWMAN 1970
SKINNER 1972
WESTBROOK 1982
SHOOT 1976
BHASKARAN 1983
BURCAT 1977
DIXON-LEWIS 1981
BEETERS-MARMEN 1977
LEVY-SAROFIM 1982
BRABBS-BROKAW 1975
GARDINER 1982
ENGLEMAN 1976
FRENKLACH 1984
COOKE-WILLIAMS 1971
WARNATZ 1981
BHASKARAN 1981
TSATSAROWIS 1978
GM 1981
TSUBOI 1976
TSANG 1985
BOWMAN 1973
COFFEE 1983
GARDINER 1977
DIXON-LEWIS 1983
WESTBROOK 1979

Lessons to take from the “Journal paper” mode

Journal model not suited to high-dimensional, distributed sysid

- The effort of researchers yields complex, intertwined, factual assertions about the possible values of the model parameters
- Handbook style of {*parameter*, *nominal*, *range, reference*} will not work
- Each individual assertion is usually not illuminating in the problem’s natural coordinates. Concise individual conclusions are rare.
- Machines must do the heavy lifting.
  - Managing lists of assertions
  - Reason and infer
- Useful role of journal paper is to document methodology leading to assertion

The GRI-Mech approach departed from the journal paper mode

- used all of the same information (but none of the “conclusions”)
- in a distributed fashion, successfully derived a model.

Alternate model for Dissemination and Analysis

Separate **asserted facts** from **analysis**

- Assertions
  - Web-based assertion of *models* of physical processes (“this parametrized math models nature”)
  - Web-based assertion of *measured outcomes* of physical processes (“nature behaves in this manner”)

  Together, these form constraints in "world"-parameter space of physical constants.

- Analysis on the assertions
  - Check consistency of collections, sensitivity of consistency to changes
  - Predict outcome of an asserted physical process model
  - Draw inferences
  - Generate samples which satisfy assertions
  - Discover informative (or suspect) assertions

Required Infrastructure

Within Project Domain

–Global parameter list management

–Posting/Publishing assertions

–Retrieving assertions

–Managing groups of assertions (and groups of groups, and so on…)

–Analyzing groups of assertions

–Posting traceable inferences

–Un-posting an assertion, blocking further use

Across project boundary

–Merging???

Partially in place at www.cmcs.org (Collaboratory Multi-Scale Chemical Science)

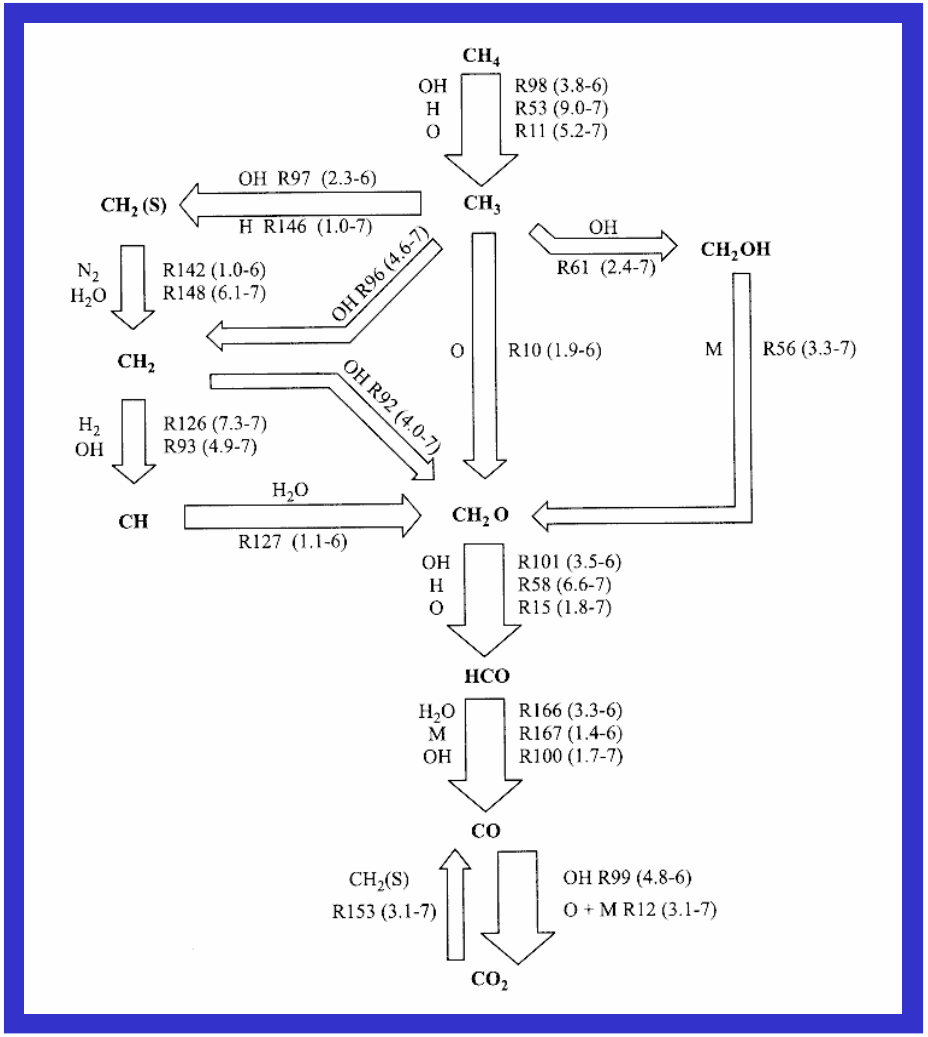
Methane Combustion: $CH_4 + 2\ O_2 \Rightarrow CO_2 + 2\ H_2O$

Methane reaction mechanisms have grown in complexity over time:

(~1970): Less than 15 elementary reactions with 12 species

(~1982): 75 elementary reactions (forward and reverse) with 25 species

(Today): GRI-Mech has 300+ elementary reactions, 53 Species, and 102 “active” parameters. The “gold standard” according to NIST.

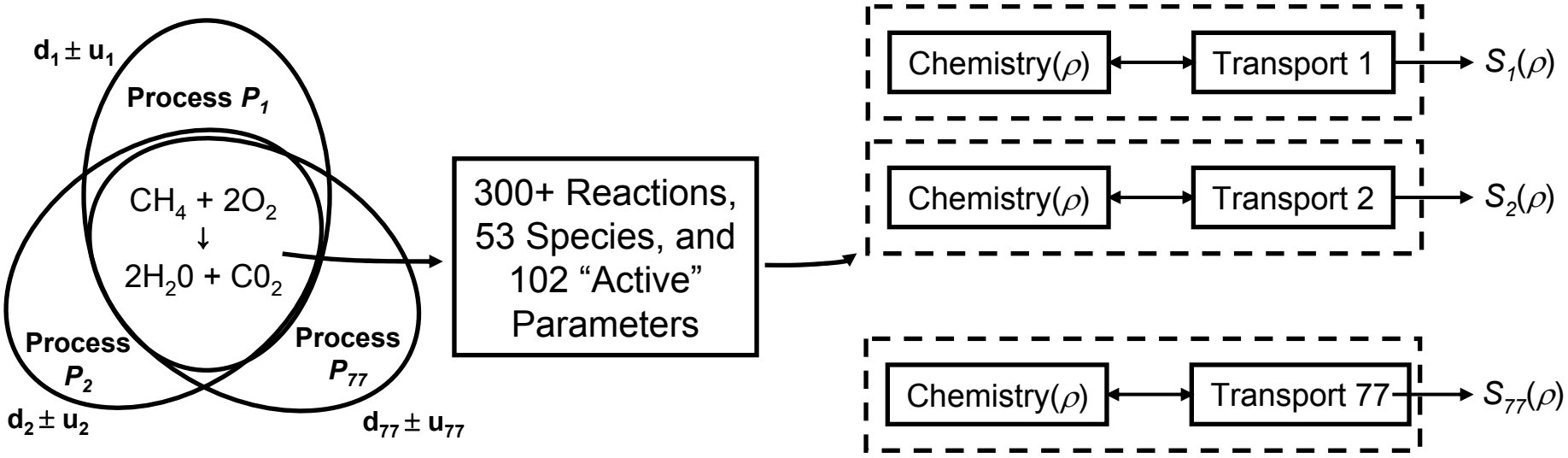


Pathway diagram for methane combustion [Turns]

GRI DataSet

The GRI-Mech (www.me.berkeley.edu/gri_mech) DataSet is collection of 77 experimental reports, consisting of models and ``raw'' measurement data, compiled/arranged towards obtaining a complete mechanism for $CH_4 + 2O_2 \rightarrow 2H_2O + CO_2$ capable of accurately predicting pollutant formation. The DataSet consists of:

- **Reaction model:** 53 chemical species, 325 reactions (nonlinear, but nonemergulent).
- **Processes (P_i)**: 77 widely trusted, high-quality laboratory experiments, all involving methane combustion, but under different physical manifestations, and different conditions.
- **Measured Data (d_i,u_i)** data and measurement uncertainty from 77 peer-reviewed papers reporting above experiments.
- **Unknown parameters (ρ):** 102 normalized parameters, typically derived from rate constants.
- **Prior Information:** Each normalized parameter is known to lie between -1 and 1.
- **Process Models (M_i):** 77 1-d and 2-d numerical PDE models, coupled with the reaction model.
- **Surrogate (reduced) Models (S_i):** 77 polynomials in 102 variables.



The prior information, models and measured data are assertions about possible parameter values.

- k^{th} assertion associated with prior info: $\rho_k^2 - 1 \le 0$
- Assertions associated with i^{th} dataset unit:

$$S_i(\rho) - d_i - u_i \le 0$$
$$d_i - u_i - S_i(\rho) \le 0$$

GRI-Mech: Project, DataSet and Mechanism

GRI-Mech (1994-present) addresses the collaborative data processing for methane

- Methane reaction model: 53 Species/300+ Reactions/102 Uncertain constants
  - Chemistry(ρ)
- 77 peer-reviewed, published Experiments/Measured Outcomes
  - Processes P_i, measured outcomes d_i, measurement uncertainties u_i
- Models of (Experiments/Measured Outcomes)
  - Math models involving heat transfer/chemistry/CFD/other phenomena
- Surrogate models of (Experiments/Measured Outcomes)
  - Factorial design of computer experiments, leading to quadratic $S_i(\rho)$
- Optimization to get "best" fit: single parameter vector (identification step). This is the "GRI-Mechanism"
- Validation

Features (www.me.berkeley.edu/gri_mech)

- Only "raw" assertions are used - none of the potentially erroneous conclusions. Heading towards "*give me your information, not your conclusions*…"
- Treats the models/experiments as information, and combines them all.
- Addresses the "lack-of-collaboration" in the post experimental data processing.
- Success! GRI Mechanism is a chemical kinetics model for methane combustion, referred to as "the gold standard" by NIST.
- Didn't arise naturally, 4 groups had to collect/reconstruct the assertions of many.

Surprise, surprise…

As expected, there were criticisms of the GRI-Mech approach. Typically, the objections read like

- "It's too early -- some fundamental knowledge is still lacking"
- "I am unwilling to rely on flame measurements and optimization to extract the value of some fundamental reaction's rate properties -- I prefer to do that in isolation"
- "Not all relevant data was included"
- "The result (one particular number) is different from mine"
- "No one can analyze my data better than me."
- "Are you trying to do away with experimentation?"

Root cause for objection seems mostly psychological:

- distributed effort dilutes any one specific contribution
- protection of individual's territory
- ownership
- required information technology

Complex (ie., non cube-like) geometry of feasible set is unappreciated

Beyond GRI-Mech: CMCS (www.cmcs.org)

- Collaboratory for MultiScale Chemical Science

Traditional Reporting of Experimental Results

The canonical structure of a technical report (a paper) is:

- Description of experiment: apparatus, conditions, measured observable
  - flow-tube reactors, laminar premixed flames, ignition delay, flame speed
- Care in eliminating unknown biases, and assessing uncertainty in outcome measurement
- Transport and chemistry models that involve uncertain parameters
  - momentum, diffusion, heat transfer
  - 10-100's reactions, uncertainty in the rate constant parameters

    $k(T)=AT^n \exp(-E/RT)$
- Sensitivities of modeled outcome to parameters in chemistry model
  - evaluate sensitivities at nominal parameter values
- Focus on parameter(s) resulting in high sensitivities on the outcome
- Convenience assumptions on parameters not being studied
  - freeze low-sensitivity parameters at nominal values
- Predict one or two parameter values/ranges

Terminology: Experiments

To be clear: an experiment consists of:

- Measured observable, *d*
- Experimental tolerance in measuring observable, *u*
- Mathematical Model, $M(\rho)$, showing dependency on active variables $\rho \in \Re^n$

(taking a deterministic, worst-case view) The experiment actually asserts an inequality constraint among the active variables: $|M(\rho) - d| < u$.

The mathematical model, $M(\rho)$, of the process is usually physics-based

- Parametrized chemical reaction ODEs
- Energy, momentum, and mass balance law

Now, recall last 3 steps of the paper

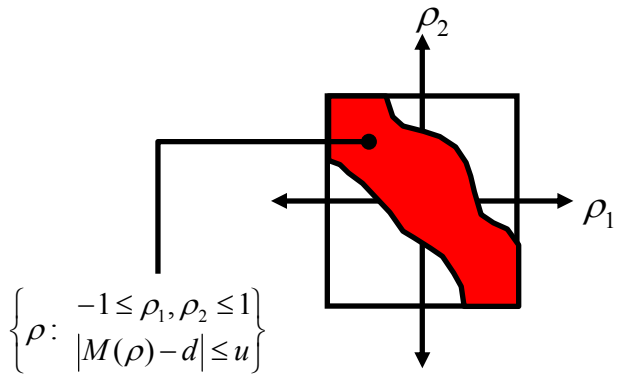
- Focus on parameter(s) resulting in high sensitivities on the outcome
- Convenience assumptions on parameters not being studied
  - freeze low-sensitivity parameters at nominal values
- Predict one or two parameter values/ranges

Pitfalls: Mixing facts with convenience assumptions

Given:

- A priori knowledge: $-1 \le \rho_k \le 1 \quad \forall k \le n$.
- An dataset unit: (M, d, u) with $\rho \in \Re^n$

From this, all that can be concluded is $|M(\rho)-d|<u$.



But, typically the procedure is:

- Freeze all parameters except one (the "most" influential), at the nominal:

  $\rho_k=0$ for $k \neq k_0$

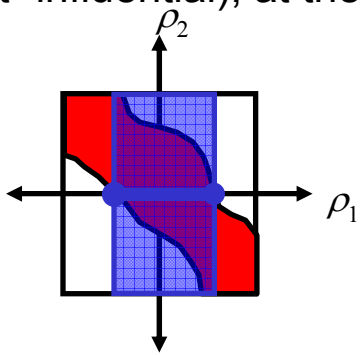
- Find range of the unfrozen parameter:

  max/min ρ_{k0}

  subject to: $\rho_k=0$ for $k \neq k_0$

  $-1 \le \rho_{k0} \le 1$

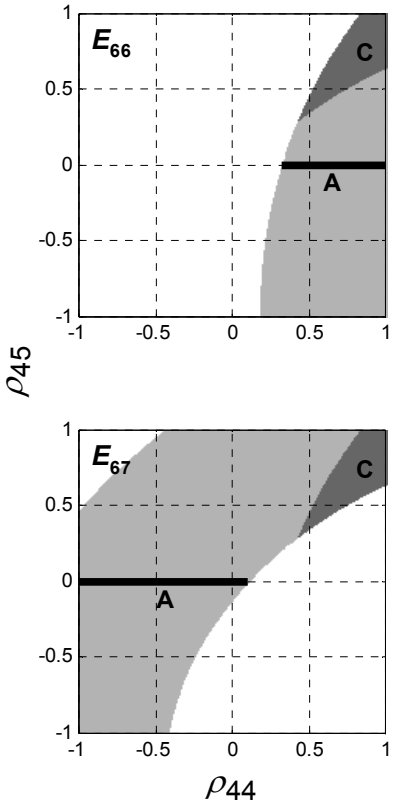
  $|M(\rho)-d|<u$



In the interest of a concise conclusion, everything useful in (M, d, u) is compressed into a potentially incorrect conclusion.

Mistakes and Artificial Controversies

The most influential (linearized, at nominal) parameter for models 66 and 67 happens to be ρ_{44} (2nd most influential is ρ_{45}). Look at slices of the feasible set for experiments 66 and 67 (all other parameters set to nominal).



Following the simplistic paradigm

E_{66} reports $0.3 \le \rho_{44} \le 1.0$.

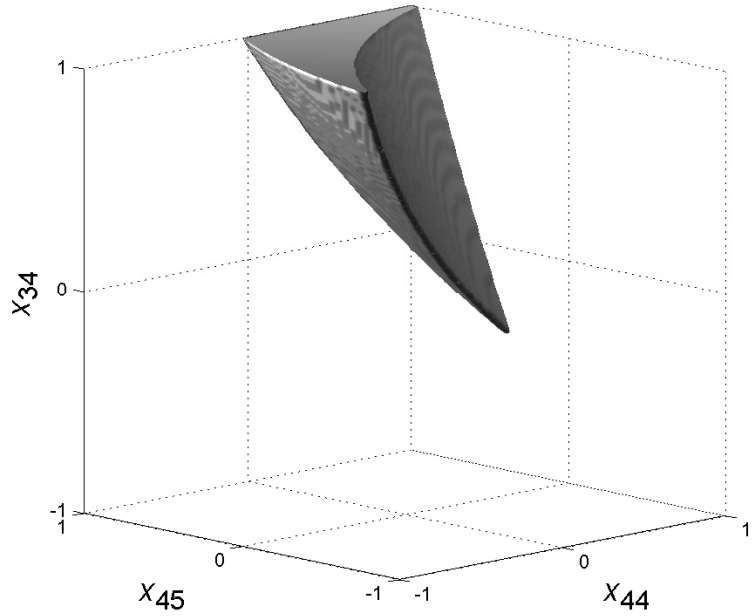
Following the simplistic paradigm

E_{67} reports $-1.0 \le \rho_{44} \le 0.15$, which is a direct conflict…

or, perhaps E_{67} considers both ρ_{44} and ρ_{45} and then reads report E_{66}. After doing so

E_{67} reports $0.3 \le \rho_{45} \le 1.0$…

or, perhaps noting that both (ρ_{44}=ρ_{45}=1) and (ρ_{44}=ρ_{45}=-1) are consistent with the data, E_{67} reports nothing!



In any case, all such reports are wrong!

A higher dimensional slice (but still a slice!) including parameter ρ_{34} illustrates the inaccuracy.

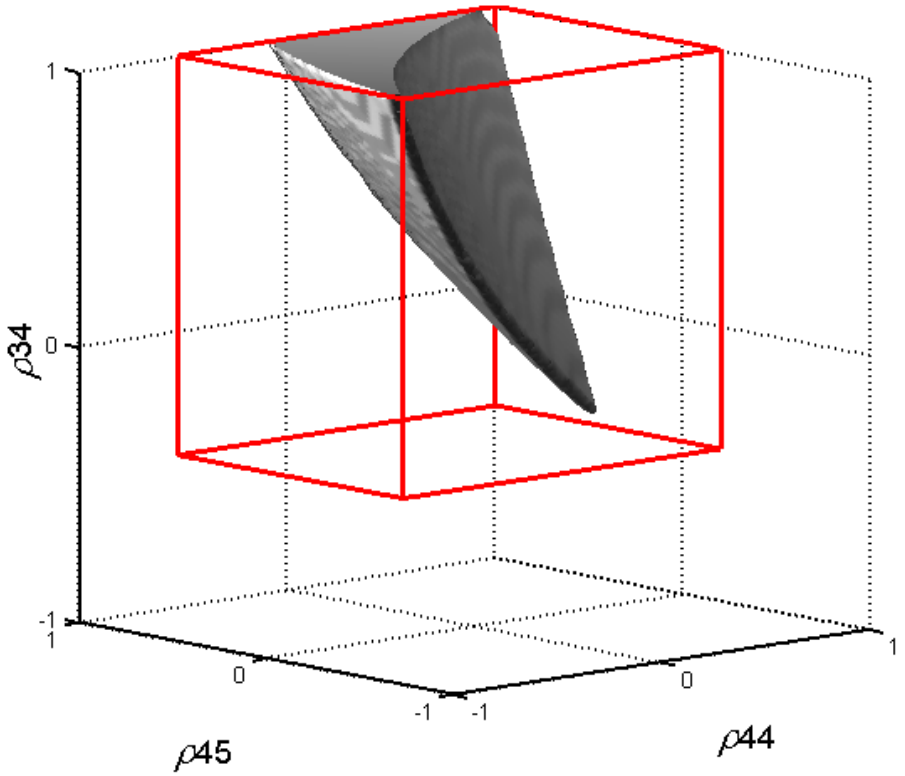
Manual management of uncertainty propagation

- Manual mode requires uncertainty description be efficient (linear in number of model parameters, say).
- Its also easy to do it wrong…
- How about consistent, but simple?

For this, use “CRC-Handbook” type description:

  - parameter values
  - plus/minus uncertainty (coordinate-aligned cube-like uncertainty)

requiring cube to contain feasible set.



High price of low cost uncertainty description

Computational exercise: assess capability of 76 assertions in predicting the outcome of the 77th model. Quantify information "lost" by doing more conservative analysis without the benefit of collaboration at the raw data level.

Method **P**: Use only the prior information (ρ 2 H) on parameters; gives the prediction interval

$$\mathcal{I}_P := \left[\min_{\rho \in \mathcal{H}} S_{77}(\rho) \quad \max_{\rho \in \mathcal{H}} S_{77}(\rho) \right]$$

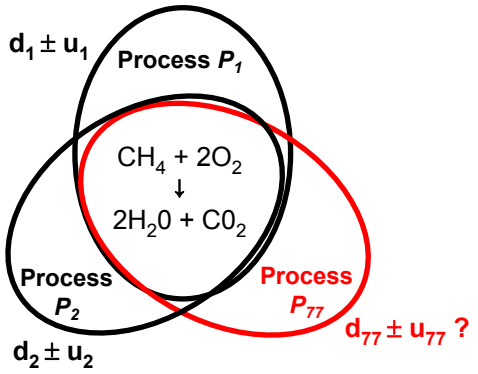
Method **C**: Community "pools" prior information and 76 assertions, but for simplicity, chooses only to reduce parameter uncertainty, maintaining a cube description (i.e., intervals for each parameter -- easy to publish, easy to share and easy to think about). The prediction interval on the 77th model using the new improved cube is

$$\mathcal{I}_C := \left[\min_{\rho \in Q_{\mathcal{F}}} S_{77}(\rho) \quad \max_{\rho \in Q_{\mathcal{F}}} S_{77}(\rho) \right]$$

Method **A**: The prediction directly uses the raw model/data pairs from all 76 experiments, as well as the prior information.

$$\mathcal{I}_A := \left[\min_{\rho \in \mathcal{F}} S_{77}(\rho) \quad \max_{\rho \in \mathcal{F}} S_{77}(\rho) \right]$$

Repeat computation 77 times…



Parameter values consistent with the assertions associated with P_i

$$\mathcal{F}_i := \{\rho \in \mathcal{H} : |S_i(\rho) - d_i| \leq u_i\}$$

Parameter values consistent with assertions associated with P_1, P_2, … P_{76}

$$\mathcal{F} := \bigcap_{i=1}^{76} \mathcal{F}_i$$

Denote set of Coordinate-Aligned Cubes by Q. Smallest coordinate aligned cube containing

$$Q_{\mathcal{F}} := \bigcap_{\substack{Q \in \mathbb{Q} \\ Q \supset \mathcal{F}}} Q$$

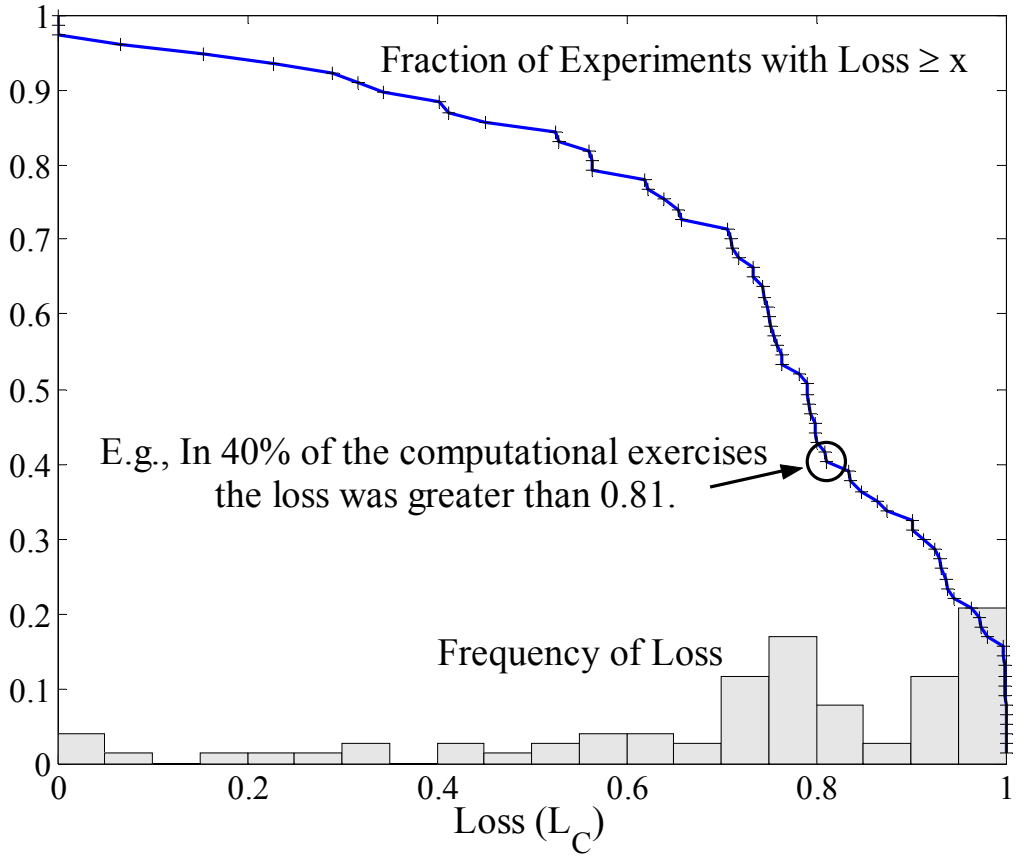
Loss using consistent, coordinate-aligned cube

How much information is lost when one resorts to method **C** instead of **A**? Define the "loss in using method **C**" as

$$L_c := \frac{\text{length}\mathcal{I}_C - \text{length}\mathcal{I}_A}{\text{length}\mathcal{I}_P - \text{length}\mathcal{I}_A}$$

No loss (L_C=0) if prediction by **C** is as tight as that achieved by **A**.

Complete loss (L_C=1) occurs if prediction by **C** is no better than method **P** (only using prior info). In such case, the experimental results are effectively wasted.



Method **C** pays a significant price for its crude representation of the assertions.

Clear: need (and payoff) for **fully collaborative environments** in which **models and data can be shared**, allowing sophisticated global optimization-based tools to **reason quantitatively with the community information**.

Consistency of data and models

- In this deterministic setting, a collection of assertions regarding models and data may be considered “consistent” if F is nonempty
- Consider this modification of the standard constraint feasibility question

$$\mathcal{C} = \text{ maximum value of } \gamma$$

$$\text{subject to the constraints: } \begin{cases} |\rho_j| \leq 1 & j = 1, 2, \ldots, n \\ |M_i(\boldsymbol{\rho}) - d_i| \leq u_i - \gamma & i = 1, 2, \ldots, m. \end{cases}$$

- Clear that C ¸ 0 implies consistency
- Refer to C as a “consistency measure”
- Bounds on parameters and experimental error may be tentative (more on this later)

Recall: Lagrange Dual Problem

C = objective; α_0 = nominal constraint bounds

λ^* = optimal Lagrange multipliers

Primal:

$$C(\boldsymbol{\alpha}) = \max_{\boldsymbol{x}} f(\boldsymbol{x})$$

$$\text{s.t. } g_i(\boldsymbol{x}) \leq \alpha_i, \quad i = 1, \ldots, N$$

$$\mathcal{C} = \max \gamma$$

$$\text{s.t.} \begin{cases} |\rho_j| \leq 1 & j = 1, 2, \ldots, n \\ |M_i(\boldsymbol{\rho}) - d_i| \leq u_i - \gamma & i = 1, 2, \ldots, m. \end{cases}$$

Dual:

$$\overline{C}(\boldsymbol{\alpha}) = \min_{\boldsymbol{\lambda} \succeq 0} \max_{\boldsymbol{x}} \Big(f(\boldsymbol{x}) - \sum_{i=1}^{N} \lambda_i (g_i(\boldsymbol{x}) - \alpha_i) \Big)$$

$$\forall \boldsymbol{\alpha} \quad C(\boldsymbol{\alpha}) \leq \overline{C}(\boldsymbol{\alpha}); \; C(\boldsymbol{\alpha}) \leq \overline{C}(\boldsymbol{\alpha}_0) + \boldsymbol{\lambda}^{\star T} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)$$

Implications:

Suppose the "gap" $\overline{C}(\boldsymbol{\alpha}_0) - C(\boldsymbol{\alpha}_0)$ is small.

i^{th} component of $\boldsymbol{\lambda}^\star$ small $\Rightarrow$ increasing α_i yields (at best) a modest increase in the optimal.

i^{th} component of $\boldsymbol{\lambda}^\star$ large $\Rightarrow$ decreasing α_i will cause optimal to decrease greatly.

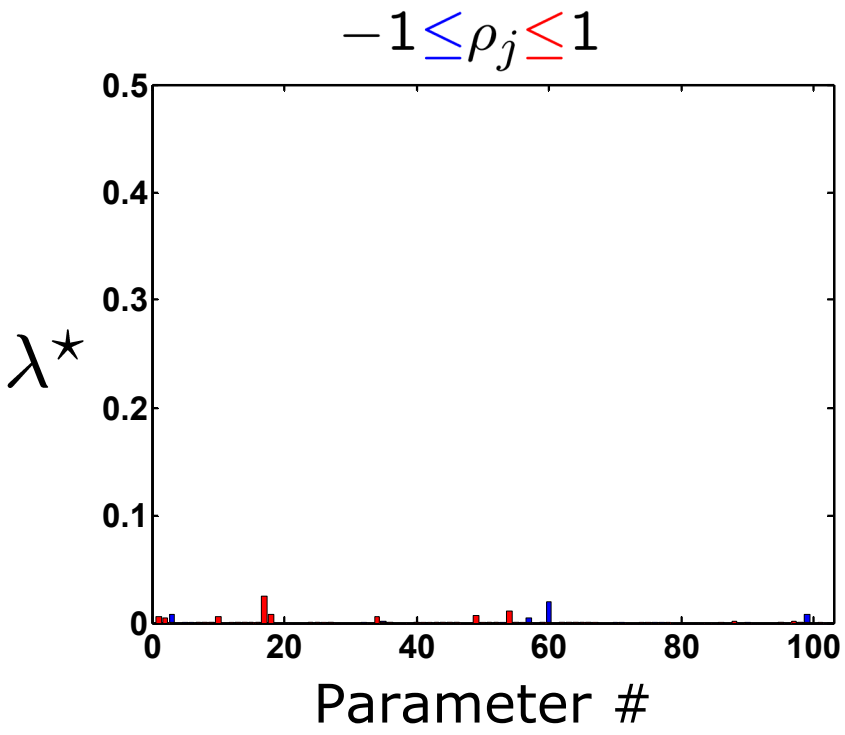
Sensitivity of C to uncertainty perturbations

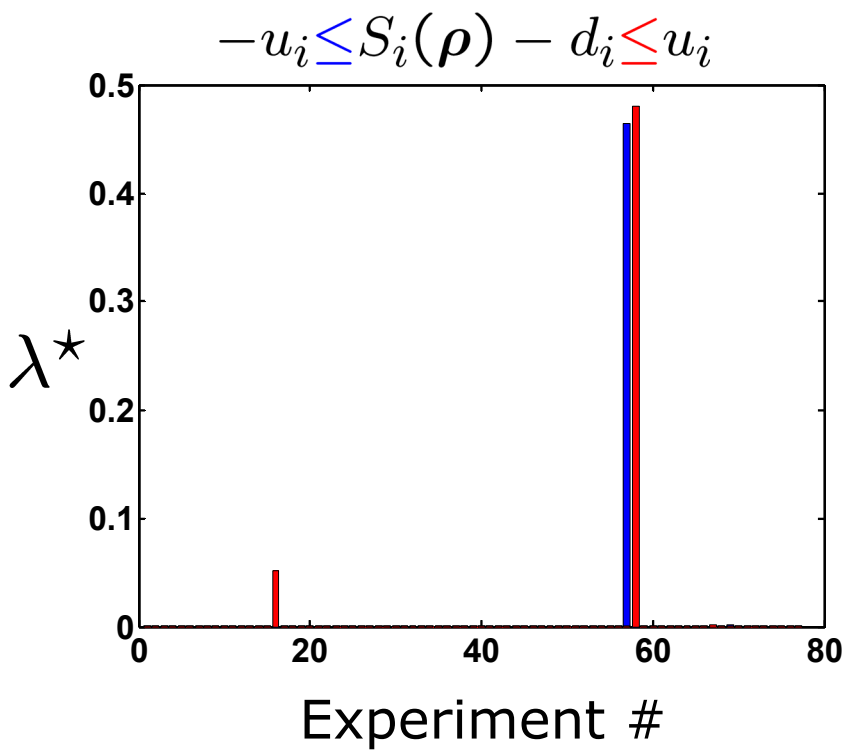
- Bounds on parameters and experimental error may be tentative.
- We use the relation $C(\boldsymbol{\alpha}) \leq \overline{C}(\boldsymbol{\alpha}_0) + \boldsymbol{\lambda}^{\star T}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)$ to investigate the impact their values have on the consistency measure.
- When $\overline{\mathcal{C}} < 0$, models, data, and error bounds involved in constraints having large multipliers should be verified

Suppose a collection of assertions is provably inconsistent ($\overline{\mathcal{C}} < 0$). Eliminate each assertion that corresponds to a component of $\boldsymbol{\lambda}^{\star}$ which equals zero. The resulting subcollection of assertions is still inconsistent so must include an erroneous claim. Note this does not imply the eliminated assertions are correct.

Consistency results for GRI-DataSet assertions

- Artificial but realistic uniform levels of experimental uncertainties were used
- With each u_i = 0.1, C 2 [0.013 0.017] (so consistent)
-) Consistent for u_i ¸ 0.087, inconsistent for u_i < 0.083





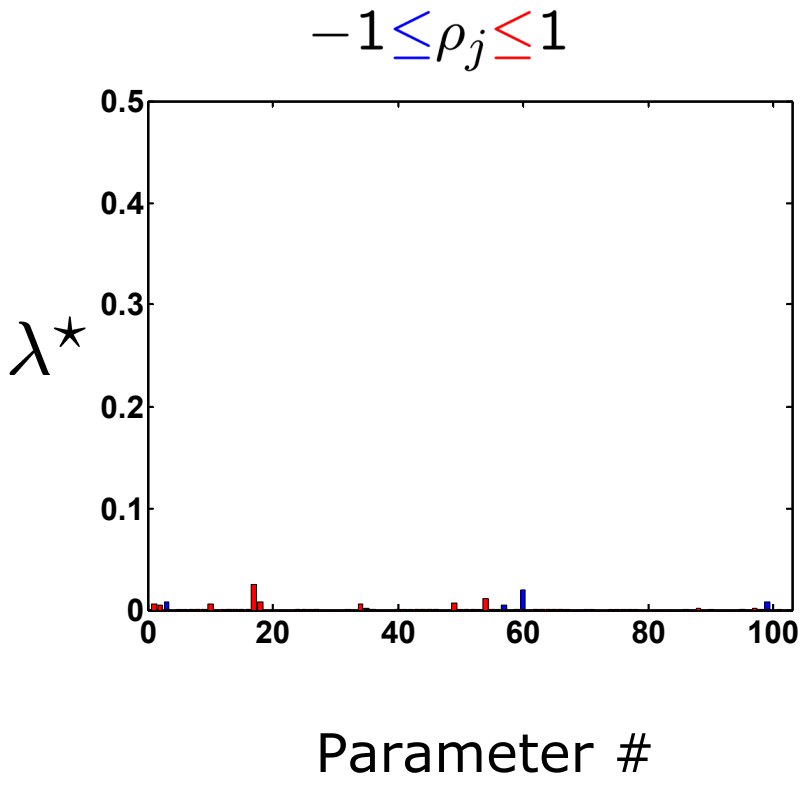
Modifications made to data

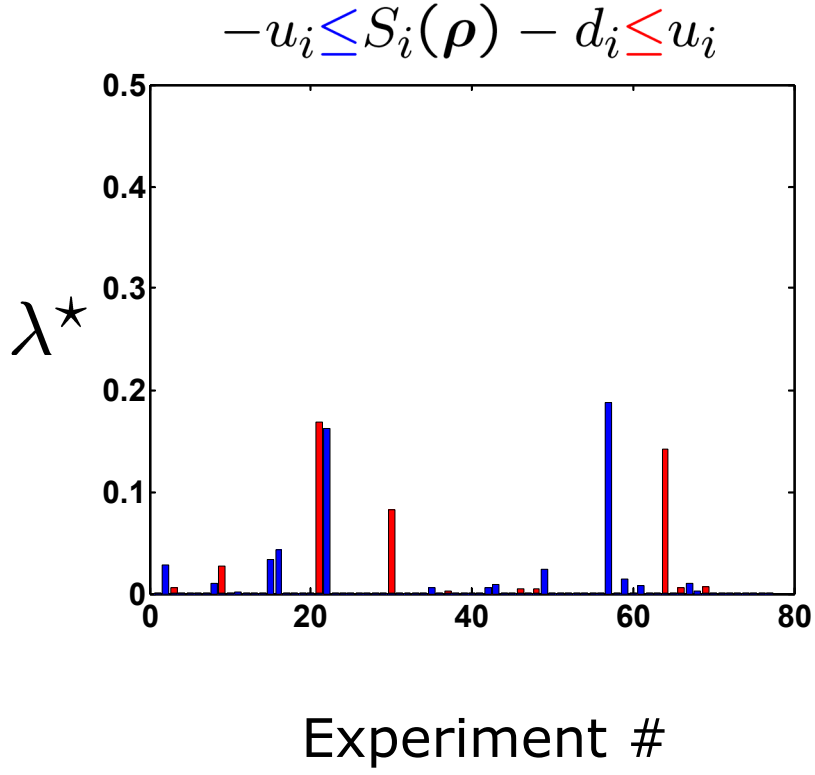
- Large peaks in expt 57 and 58 suggest that some data should be reexamined…
- The scientists involved in (57) and (58) rechecked calculations, and concluded that reporting errors had been made
- Both reports were updated -- one measurement value increased, one decreased -- exactly what the consistency analysis had suggested

TARGET	INITIAL	REVISED
OH.1a	970	700
OH.1b	218	255

Multipliers with revised dataset

- In both cases, data was altered in the direction that improved consistency
- Originally inconsistent at $u_j = 0.083$, now inconsistent at $u_j = 0.062$





How are we computing?

- Transforming real models to polynomial models
  - Large-scale computer "experimentation" on $M(\rho)$.
    - Random sampling and sensitivity calculations to determine active parameters
    - Factorial design-of-experiments on active parameter cube
  - Linear, Quadratic or Polynomial (stay in Sum-of-Squares hierarchy) fit
  - Assess the residuals, account for fit error in assertion
- Assertions become polynomial inequality constraints
- Most analysis is optimization subject to these constraints
- S-procedure, sum-of-squares (scalable emptyness proofs, outer bounds)
  - Outer bounds are also interpreted as solutions to the original problem when cost is an expected value, constraints are only satisfied on average, and the decision variable is a random variable.
- Branch & Bound (or increase order) to eliminate ambiguity due to fit errors
- Off-the-shelf constrained nonlinear optimization for inner bounds
  - Use stochastic interpretation of outer bounds to aid search

Papers

Michael Frenklach, Andrew Packard, Pete Seiler and Ryan Feeley, "Collaborative data processing in developing predictive models of complex reaction systems," *International Journal of Chemical Kinetics*, vol. 36, issue 1, pp. 57-66, 2004.

Michael Frenklach, Andy Packard and Pete Seiler, "Prediction uncertainty from models and data," *2002 American Control Conference*, pp. 4135-4140, Anchorage, Alaska, May 8-10, 2002.

Pete Seiler, Michael Frenklach, Andrew Packard and Ryan Feeley, "Numerical approaches for developing predictive models," submitted to *Engineering Optimization*, Kluwer December 2003.

Ryan Feeley, Pete Seiler, Andy Packard and Michael Frenklach, "Consistency of a reaction data set," submitted to *Journal of Physical Chemistry*, June 2004.