Transforming Data into Knowledge: Process Informatics for Combustion Chemistry

M. Frenklach, A. Packard, Z.M. Djurisic, R. Feeley, T. Russi (UCB)
D.M. Golden, C.T. Bowman (Stanford)
W.H. Green, Jr., G.J. McRae (MIT)
P.J. Smith, A. Sarofim (U of Utah)
H. Wang (USC)
G. Smith (SRI)
P.R. Westmoreland (U of Mass)
T.C. Allison (NIST)
J. G. Frey, M.J. Pilling (Leeds)

Supported by:
NSF Chemistry Division (Cyber-infrastructure)
“The mere accumulation of facts, even an extremely extensive collection, ... does not constitute scientific method; it provides neither a direction for further discoveries nor does it even deserve the name of science in the higher sense of that word. The cathedral of science requires not only material, but a design, harmony ... a design ... for the harmonic composition of parts and to indicate the pathway, by which the most fruitful new material might be generated.”
Is there a problem?
A TYPICAL SEQUENCE OF EVENTS

Experimenter A

Model B

Experimenter B

Model A

Experimenter C
AND CHAOS OF DATA

Model A

Thermo Source 25

Rxn Rate Source 29

Rxn Rate Source 36

Model B

modified

Thermo Source 15

modified

Thermo Source 16

modified

Thermo Source 21

copied

copied

copied
PROBLEMS

• multitude of sources
• conflicting data in/among sources
• poor documentation of data
• no uncertainty reporting or analysis
• not much focus on integration of data
prior knowledge
• an “expert choice” of parameter values is like a needle in a haystack

• a set of individual uncertainties does not represent the total compound uncertainty
• models are not additive
• data are not additive
• need a *system* for synthesis and transformation of data into knowledge
Process Informatics

- Mechanism understanding through experimentation and modeling for the purpose of prediction
- Integrated system of data and tools – System Science
- Collaborative science
- Applications to:
  - chemical kinetics
  - combustion
  - atmospheric chemistry
  - astrophysics
  - system biology
  - ...

PrIME
The community builds PrIME, PrIME builds the community.
Sometime in 2008 ...

**Scientist to SYSTEM:**
What impact will my experiment have on...?

**SYSTEM to Scientist:**
Only if you control error within 2 % ...
Sometime in 2010 ...

Engineer to SYSTEM:

I need a reliable prediction for conditions...

SYSTEM to Engineer:

The best one will give error bars of 30 % ... $$ can drop it to 10 %, and $$$$ to 3 %
Sometime in 2020 ...

Policymaker to SYSTEM:

How much longer will there be an Antarctic ozone hole?

SYSTEM to Policymaker:

...
PrIMe INFRASTRUCTURE

- prime
  - warehouse
  - schema
  - tools
  - website (primekinetics.org)
    - depository
      - library (best current dataset)
PrIME INFRASTRUCTURE

- Data management
- Web services
- Model generation
- Numerical solution (flames, flow reactors, ...)
- Model reduction
- Optimization
- Error analysis/propagation
- ...
DATA COLLABORATION

- model/data sharing and automated reasoning to extract the totality of information in the community data

- Dataset: a set of \{Observation, Uncertainty, Model\}
- Prior Information on parameters
- Assertions/prediction over the feasible set
- Use global optimization methods, from control systems analysis, convex analysis, and algebraic geometry
PREDICTION ON THE FEASIBLE SET

Information Gain:

\[ I = 1 - \frac{Posterior\ Range}{Prior\ Range} \]

Including Experimental Observations

Range of \( Y \)

feasible set

prior knowledge
Pair-wise consistency test of GRI-Mech 3.0 Dataset
Sensitivity of uncertainty in predicting $Y$ to uncertainty in observation $Y_j$.
PrIMe organization:

practical realization of collaborative data management
data provider – experimenter – theorist

PrIMe

new data on PAH

data

submission

Data Depository

Submission Forms

Bibliography

Elements

Species

Reactions

Experiments

Models
Submission Forms

Bibliography

Elements

Species

Reactions

Experiments

Models
data provider
-experimenter
-theorist

PrIME
PAH work group members

Automatic notification
data submission

Data Depository
thermochemistry
transport reaction rates
experimental data

Best Current Dataset

Analysis

decision

update automatic verification

Tools:
sensitivity
optimization
error propagation
DataCollaboration
PrIME

Best Current Dataset

PAH work group members

Analysis

total team

data set

thermochemistry
transport
reaction rates
experimental data

decision

PAH work group members
PrIME

thermochemistry
transport
reaction rates
experimental

model building

make request a model

- detailed
- reduced

1 – 100 atm

select targets:

- flame: 1 atm, ...
- flame: 5 atm, ...
- shock tube: ignition, ...
- shock tube: OH conc, ...
- flow reactor: ...

Tools:
assembly
optimization
reduction
...

model user
-CFD researcher
-industrial user
-policy maker

request a model
BENEFITS of PrIME

• establishes/advances system science
  – synthesis of disparate data into shared knowledge
  – crosses and links all physical scales
  – expandable and extendable, modular and parallelizable
  – identifies and ranks problem areas and tests possible action

• addresses group sociology
  – opportunity and incentive for scientific collaboration within and among communities (“virtual” collaboration)
  – means for reaching broad community consensus
  – inclusive, yet “filters out” bad science
  – respects and protects individual territory

• enables technology transfer
  – continuous/instantaneous, in parallel with science
  – solid base for technology development
  – responsiveness to (individual) technological needs
  – assures privacy if so desired

• leverages resources
  – distributed funding — funding of individual Work Groups
MODULARITY, EXTENDIBILITY, ...

combustion modeling - quantum chemistry

reactions - thermo - molecular structure - spectra - absorption coefficient

thermosciences - diagnostics
CURRENT STATUS OF PrIMe

• grass-roots initiative
• open membership
• launched on April 21-22, 2006
• data Depository
  – NIST kinetic database
  – GRI-Mech 3.0
  – Stanford “Volume 1”
  – USC flames, EU/Leeds kinetic database, ...
• 20 Work Groups – all working in parallel
  – reaction subsets, thermodynamics, transport
  – experimental data
  – cyber-numerics (Web services, Grid)
  – industrial
• 88 registered members (as of December 6) from 14 counties
CURRENT STATUS OF PrIMe

• data was moved to “24/7” service
• initial release of PrIMe Data Management software is within days
• next infrastructure milestone: Web services
  – Grid access (UofUtah, Argonne)
  – Data Collaboration (UCB)
  – automated model generation (MIT)
  – Active Thermodynamic Tables (Argonne)
  – Sensitivity analysis (Hungary)
• first science “release”: C0-2 reaction system
  – communal collection and (re)evaluation of pertinent data
  – fully-automated, on-the-fly data analysis and model building
• IJCK/Wiley agreed to have a “direct” link to PrIMe
  – submission of data is tied to submission of manuscripts
  – automated assessment of newly submitted data, reporting to the author and reviewers