Homeland Security

Science and Technology

Data Needs in Computational Modeling and Simulation—An Industry Perspective

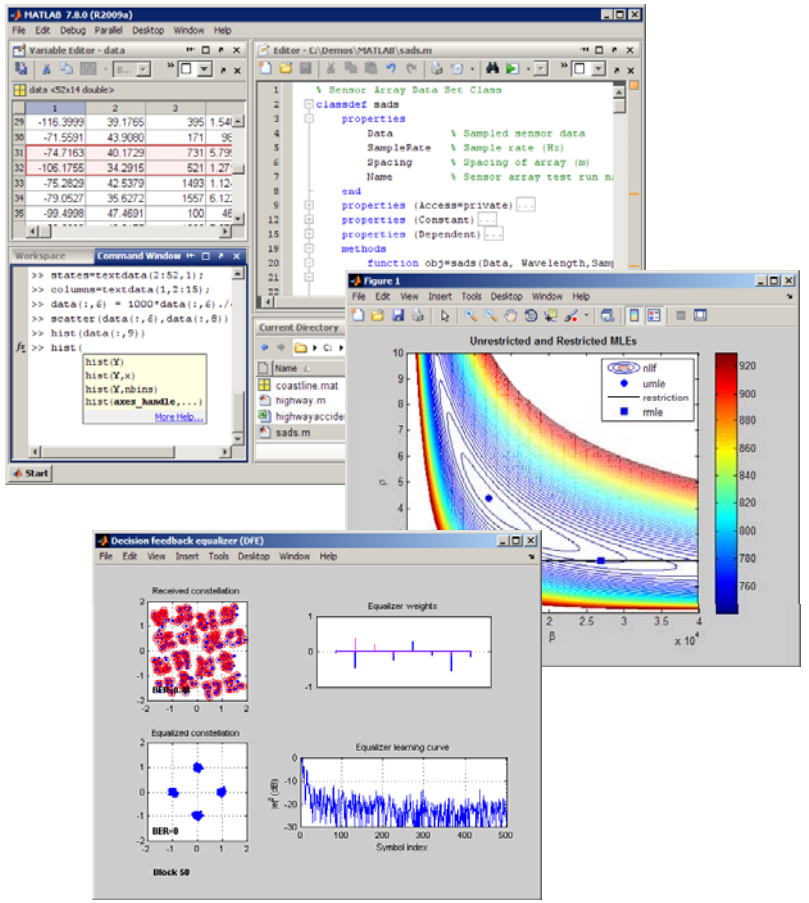
Pieter J. Mosterman
Senior Research Scientist, Design Automation
Adjunct Professor, School of Computer Science

Core MathWorks Products

MATLAB®

The leading environment for technical computing

- The *de facto* industry-standard, high-level programming language for algorithm development
- Numeric computation
- Data analysis and visualization
- Toolboxes for signal and image processing, statistics, optimization, symbolic math, and other areas
- Foundation of MathWorks products

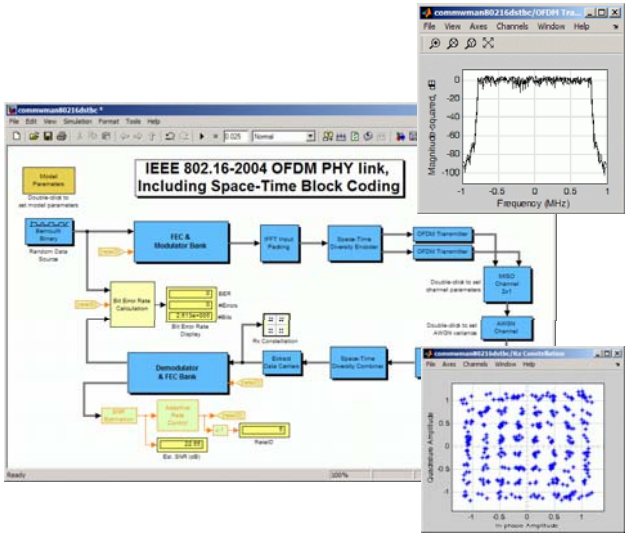


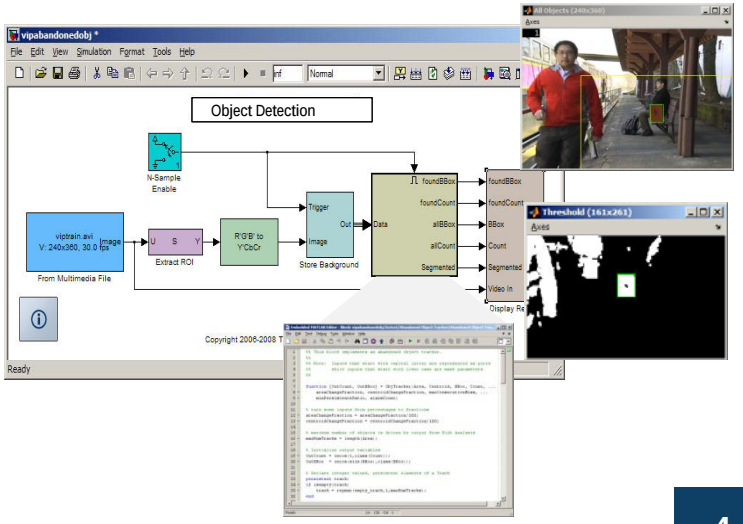
Core MathWorks Products

SIMULINK®

The leading environment for system-level modeling, simulation, and verification of communications and electronic systems

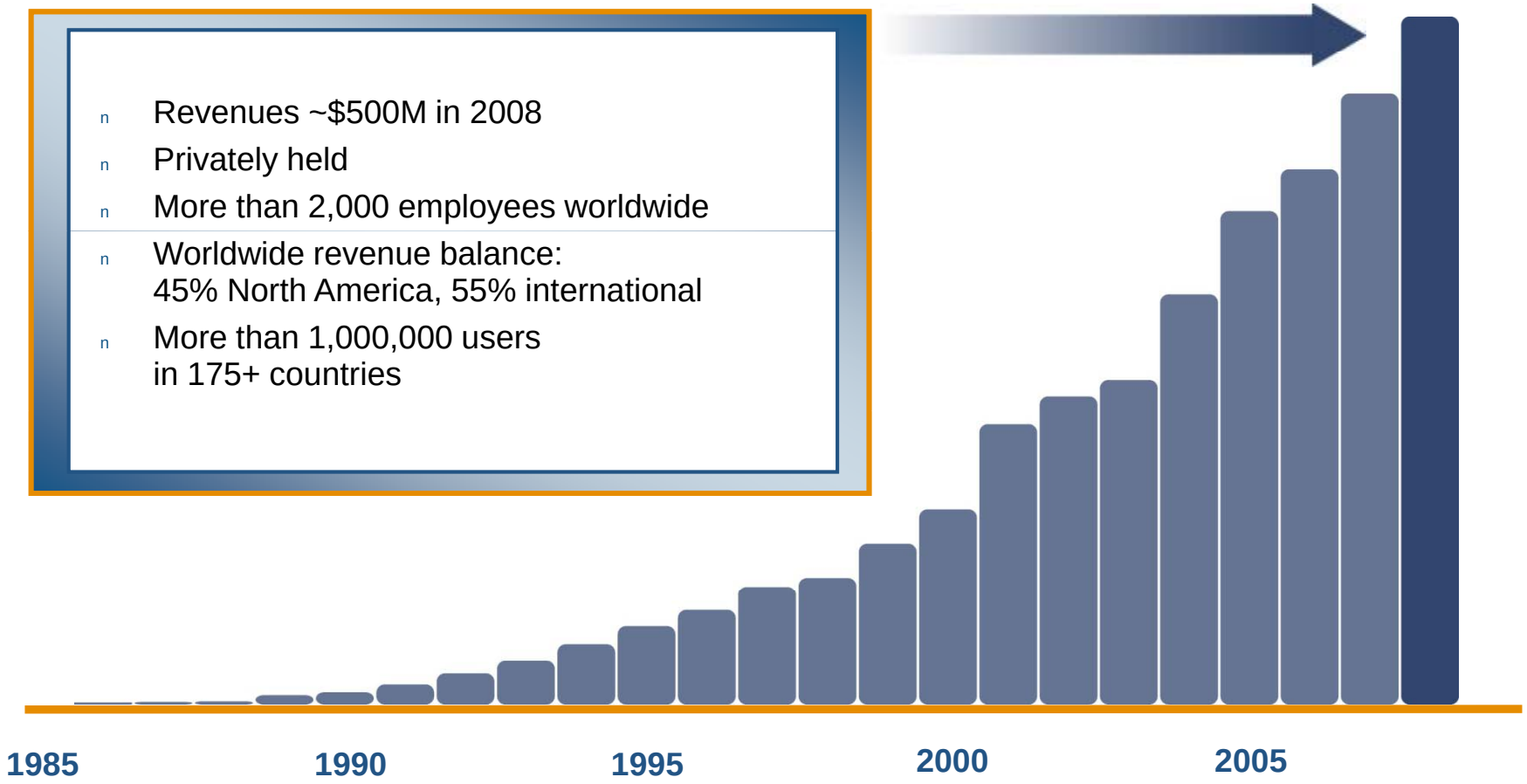
- Multidomain system-level design and verification
- Digital, analog, and mixed-signal simulation using discrete-time, continuous-time, state machine, and discrete event modeling
- Floating- and fixed-point algorithm development using MATLAB, Simulink blocks, or existing C code
- Blocksets for signal processing, video processing, communications, and RF
- Open architecture with links to third-party tools and development boards, and instrumentation
- C and HDL code generation for DSPs, embedded processors, and FPGAs



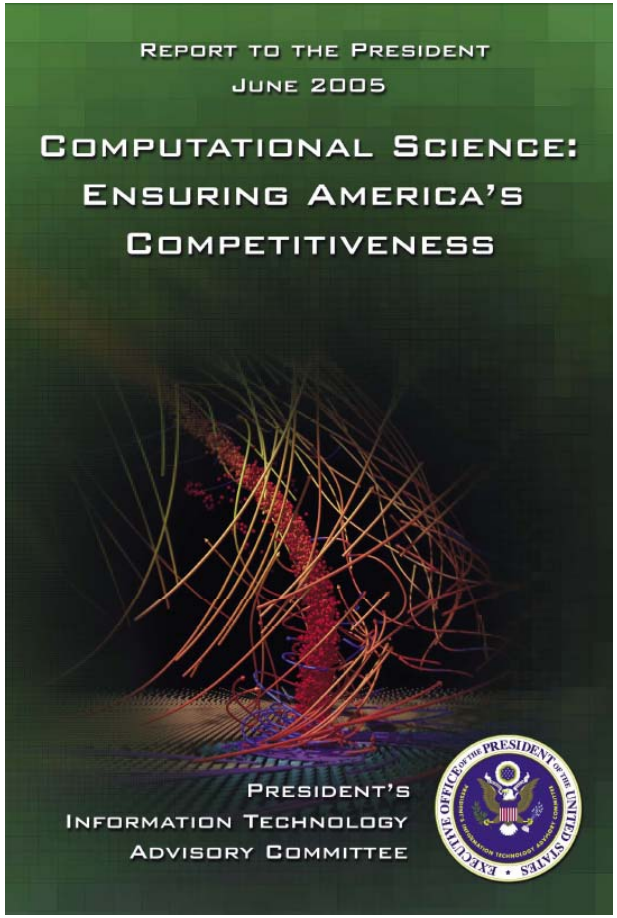


The MathWorks Today

- Revenues ~$500M in 2008
- Privately held
- More than 2,000 employees worldwide
- Worldwide revenue balance:
  45% North America, 55% international
- More than 1,000,000 users
  in 175+ countries



The general importance of computation

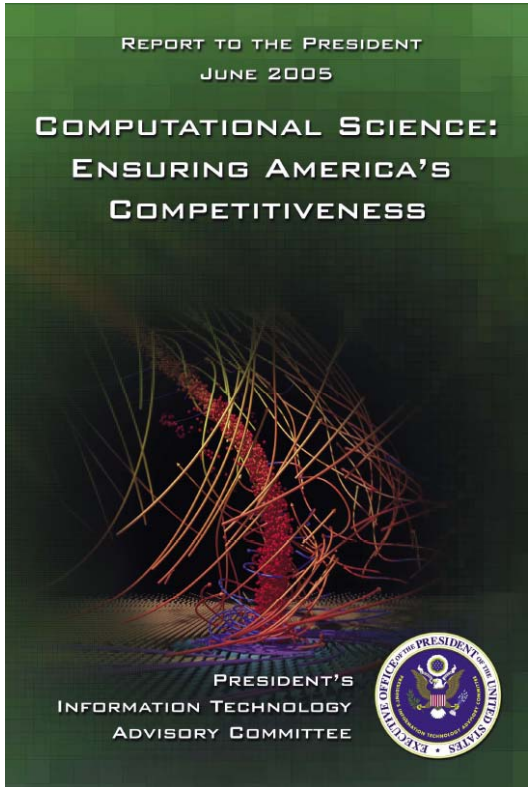


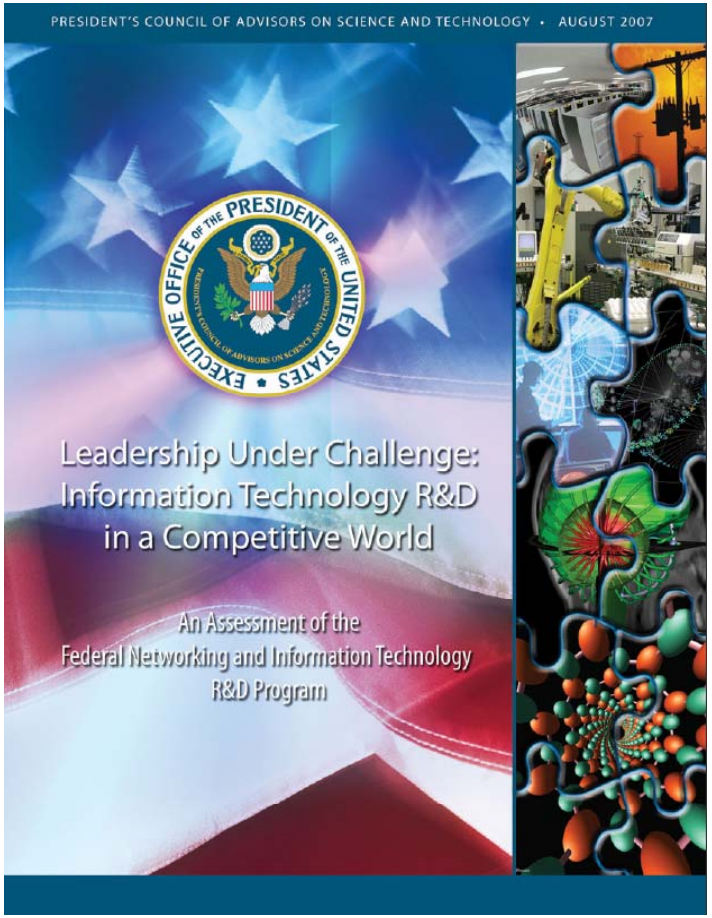
The general importance of computation

Together with theory and experimentation, computational science now constitutes the “third pillar” of scientific inquiry,

In industry, computational science provides a competitive edge by transforming business and engineering practices.

The general importance of computation



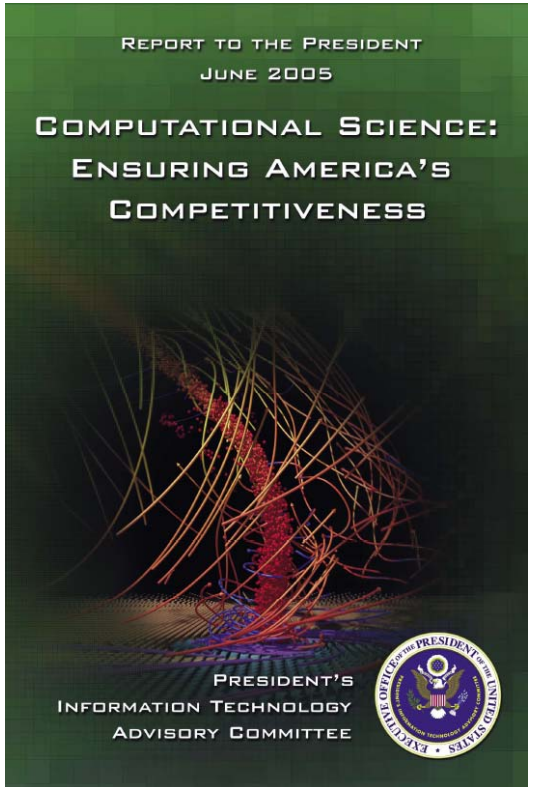


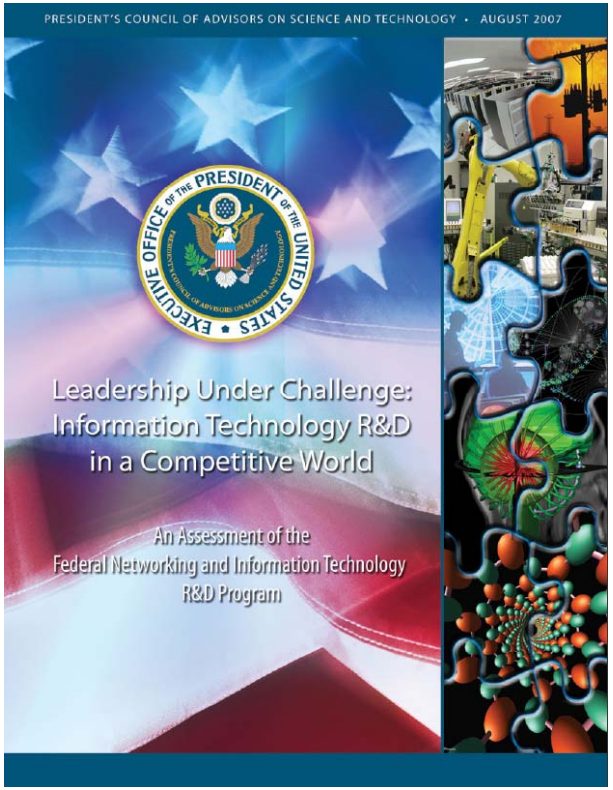
The general importance of computation

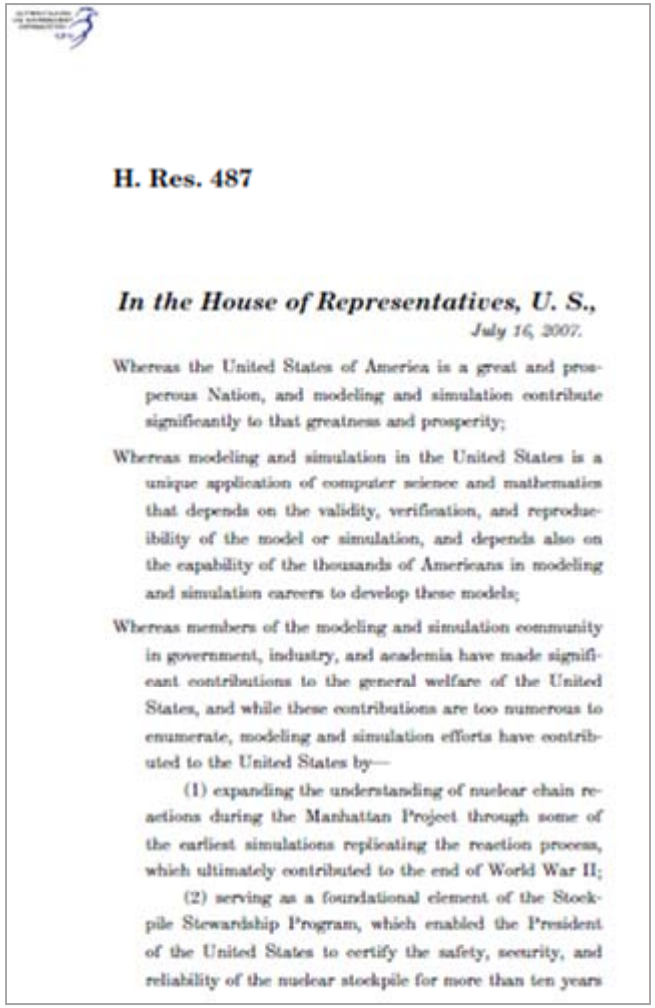
As new funding becomes available, the following four areas should receive disproportionally larger increases […]

- *NIT Systems Connected with the Physical World* (which are also called embedded, engineered, or cyber-physical systems)
- […]
- *Digital Data*: The Interagency Working Group on Digital Data, in cooperation with the NITRD Subcommittee, should develop a national strategy and develop and implement a plan to assure the longterm preservation, stewardship, and widespread availability of data important to science and technology.

The general importance of computation







H. Res. 487

In the House of Representatives, U. S.,
July 16, 2007.

Whereas the United States of America is a great and prosperous Nation, and modeling and simulation contribute significantly to that greatness and prosperity;

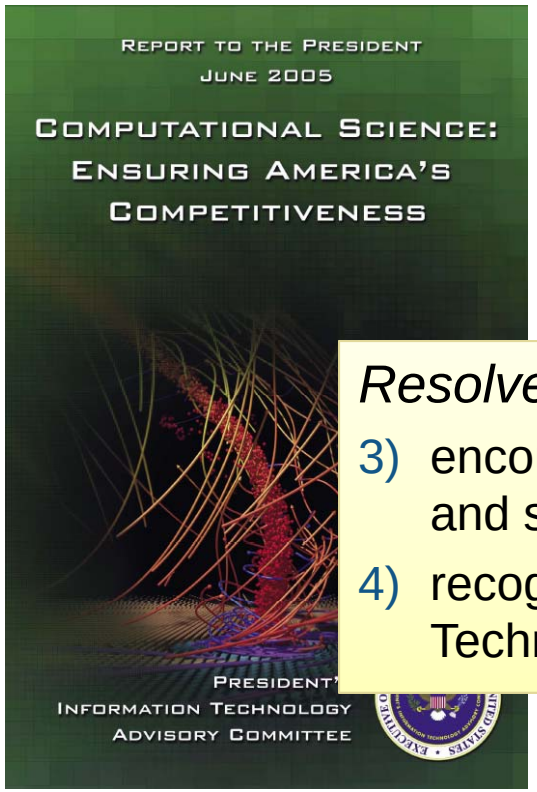
Whereas modeling and simulation in the United States is a unique application of computer science and mathematics that depends on the validity, verification, and reproducibility of the model or simulation, and depends also on the capability of the thousands of Americans in modeling and simulation careers to develop these models;

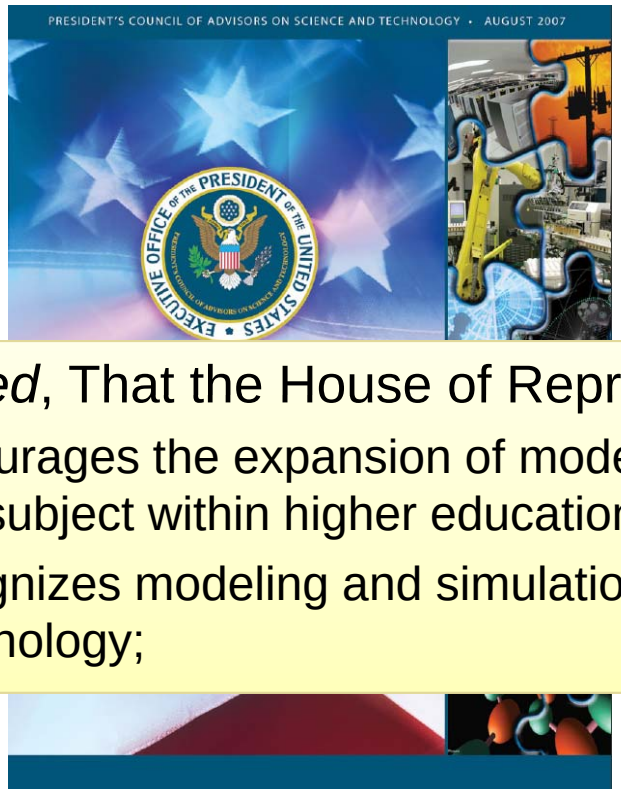
Whereas members of the modeling and simulation community in government, industry, and academia have made significant contributions to the general welfare of the United States, and while these contributions are too numerous to enumerate, modeling and simulation efforts have contributed to the United States by—

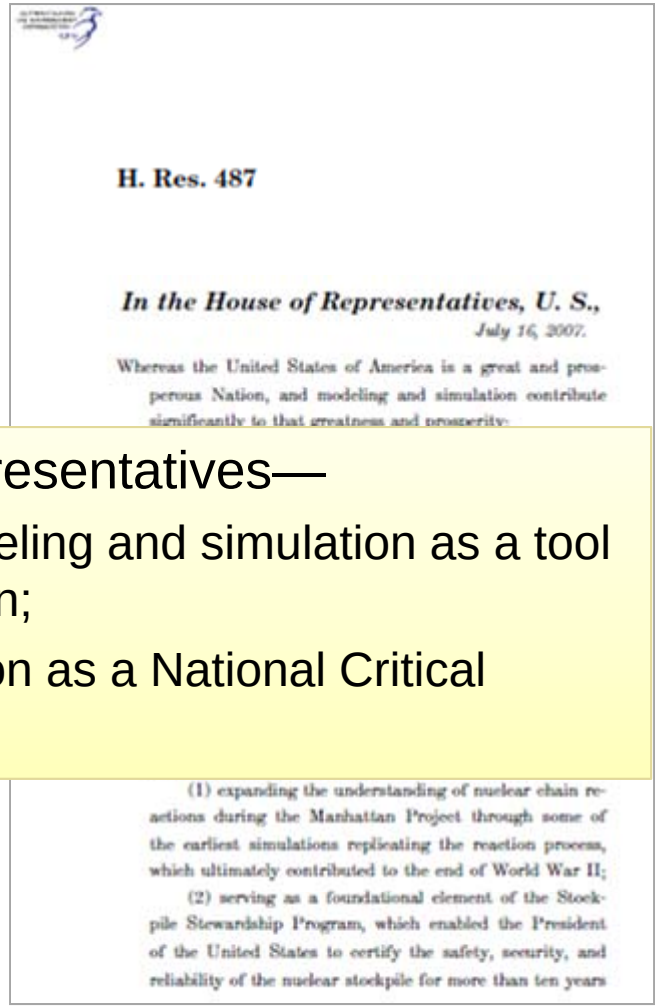
(1) expanding the understanding of nuclear chain reactions during the Manhattan Project through some of the earliest simulations replicating the reaction process, which ultimately contributed to the end of World War II;

(2) serving as a foundational element of the Stockpile Stewardship Program, which enabled the President of the United States to certify the safety, security, and reliability of the nuclear stockpile for more than ten years

The general importance of computation







H. Res. 487

In the House of Representatives, U. S.,
July 16, 2007.

Whereas the United States of America is a great and prosperous Nation, and modeling and simulation contribute significantly to that greatness and prosperity;

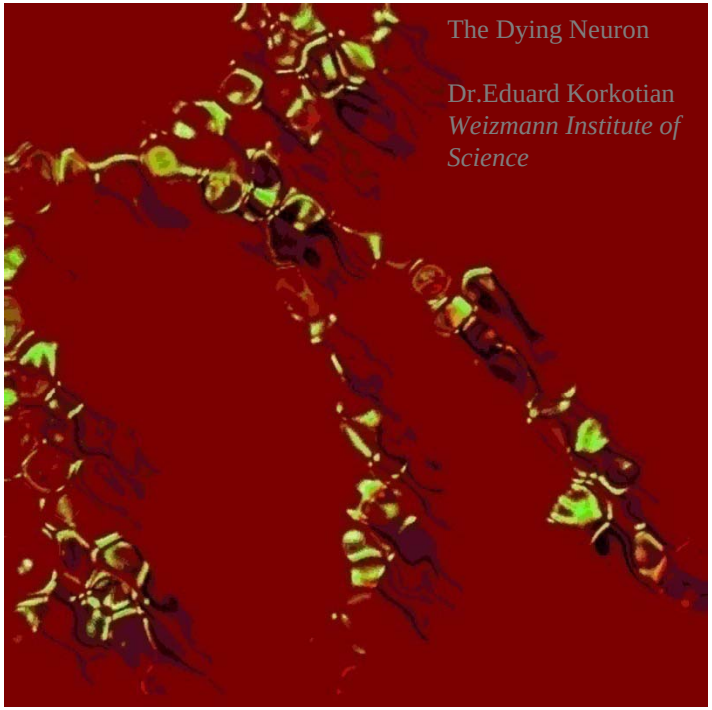
(1) expanding the understanding of nuclear chain reactions during the Manhattan Project through some of the earliest simulations replicating the reaction process, which ultimately contributed to the end of World War II;

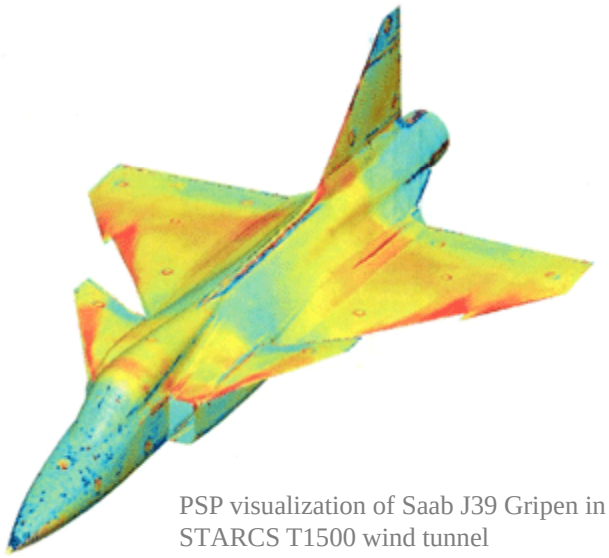
(2) serving as a foundational element of the Stockpile Stewardship Program, which enabled the President of the United States to certify the safety, security, and reliability of the nuclear stockpile for more than ten years

Resolved, That the House of Representatives—

3) encourages the expansion of modeling and simulation as a tool and subject within higher education;
4) recognizes modeling and simulation as a National Critical Technology;

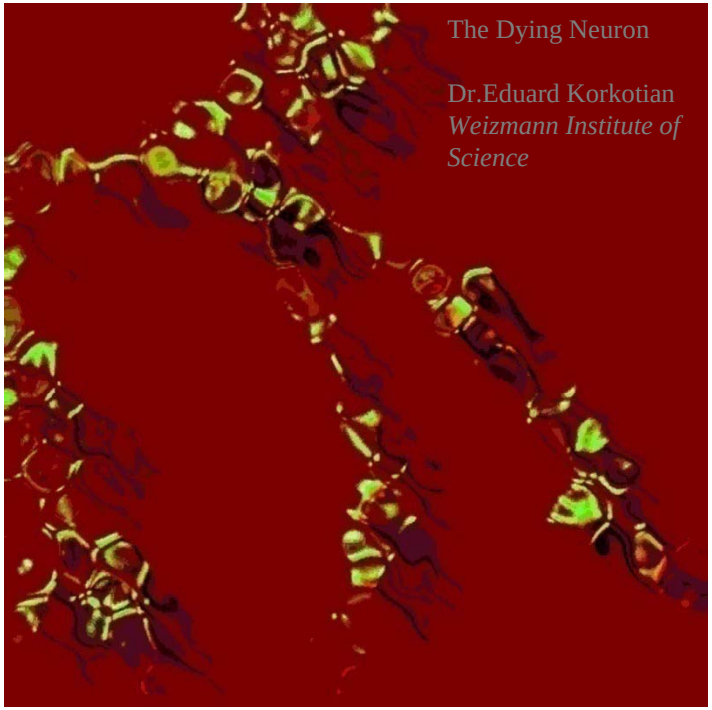
Computation for simulation and visualization

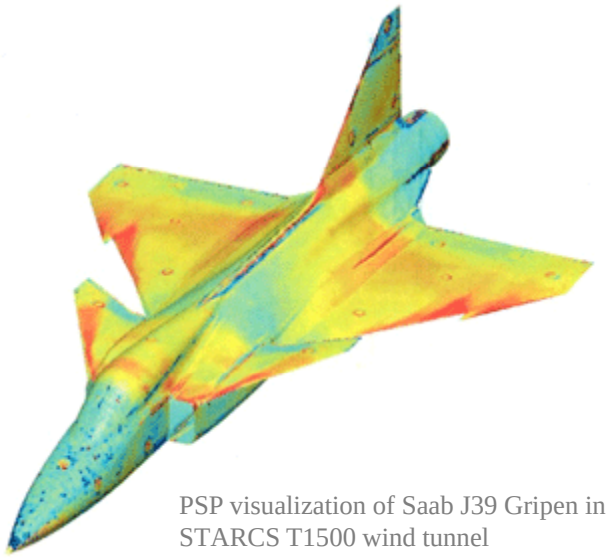




PSP visualization of Saab J39 Gripen in STARCS T1500 wind tunnel
http://www.starcs.se/advanced_methods.aspx

Computation for simulation and visualization





PSP visualization of Saab J39 Gripen in STARCS T1500 wind tunnel
http://www.starcs.se/advanced_methods.aspx

How do we know how 'good' these models are?
Verification & Validation (V&V)!

Common verification

- Compare computation with an exact result
- Assess error convergence against increased precision
- Monitor domain constraints
  - conservation of energy
  - symmetries
- Compare with computed results of related (smaller) problems

Common validation

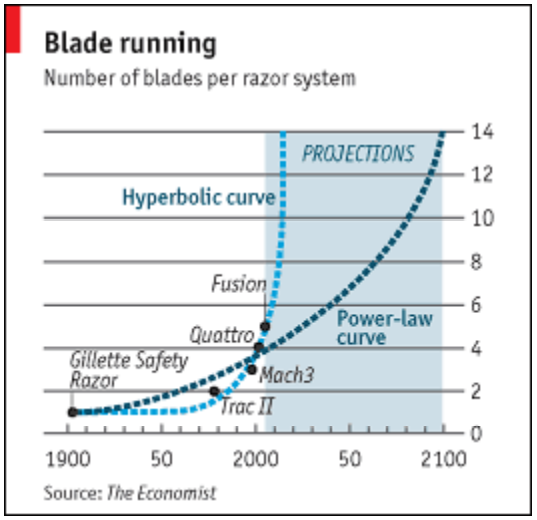
- Measurements of modeled system
- Controlled experiments to investigate principles
- Experiments to certify performance
- Experiments to validate specific computations

Common validation

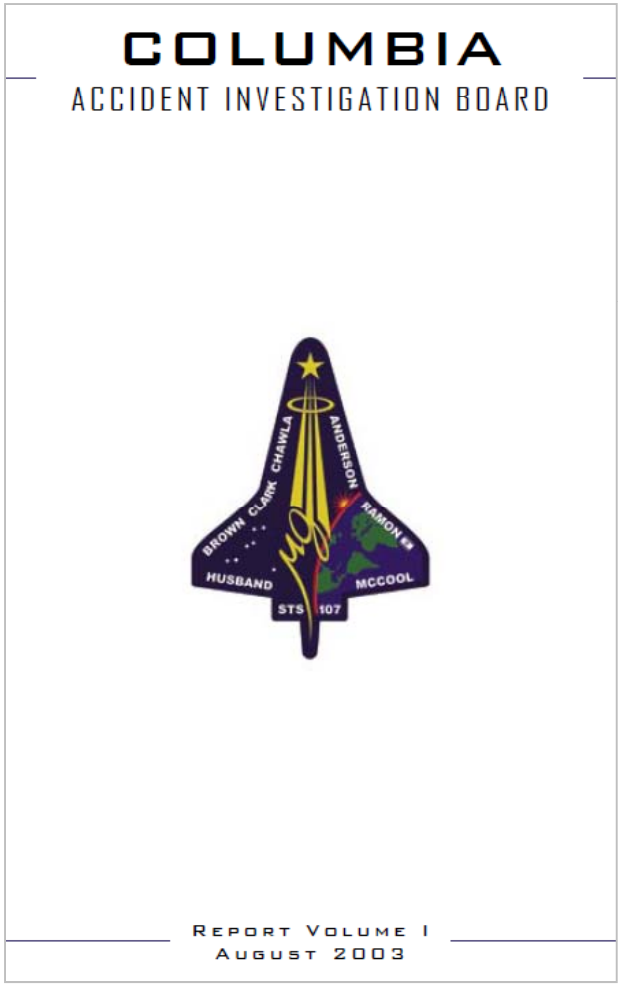
- Measurements of modeled system
- Controlled experiments to investigate principles
- Experiments to certify performance
- Experiments to validate specific computations

So, what is the problem?

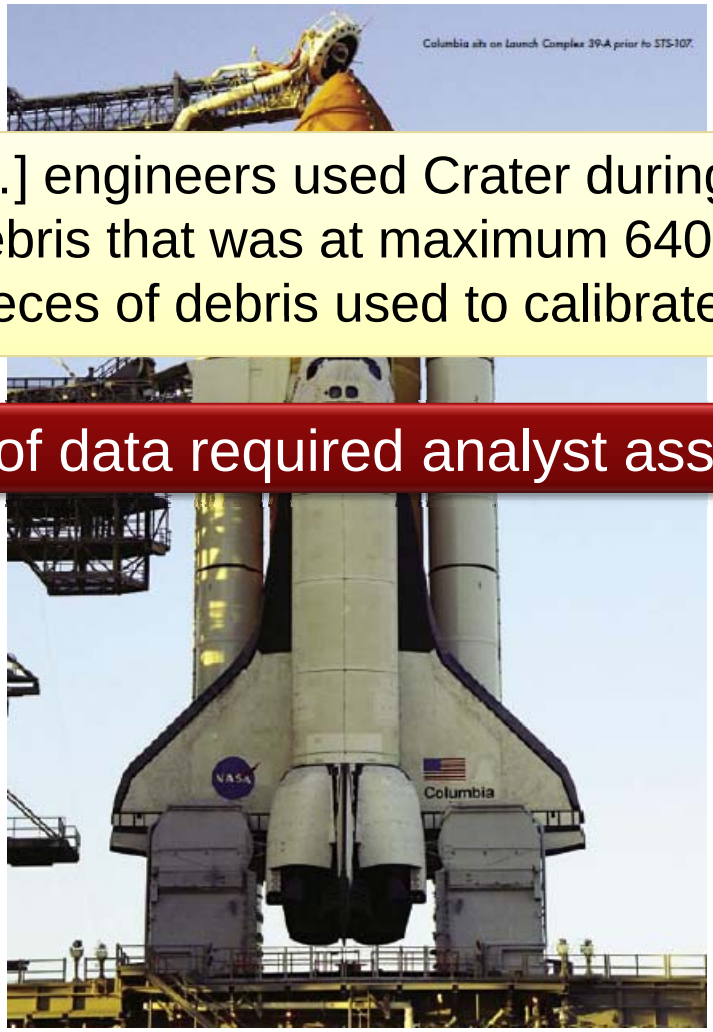
Lack of coverage from data



Lack of coverage from data

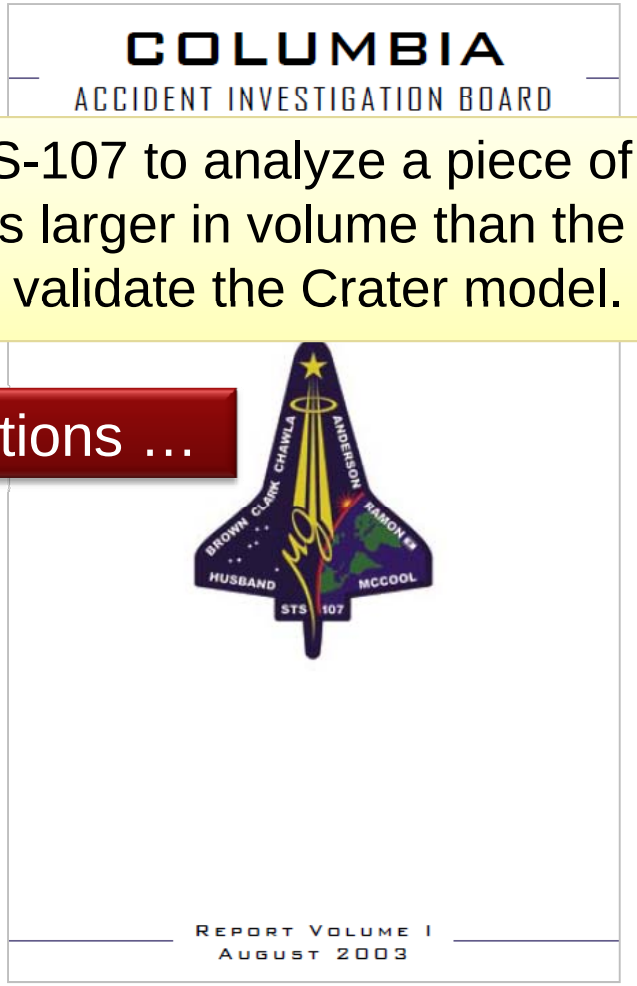


Lack of coverage from data



[…] engineers used Crater during STS-107 to analyze a piece of debris that was at maximum 640 times larger in volume than the pieces of debris used to calibrate and validate the Crater model.

Lack of data required analyst assumptions …



Unknown data needs

Unknown data needs

Ariane 501 Inquiry Board report

Paris, 19 July 1996

ARIANE 5

Flight 501 Failure

Report by the Inquiry Board

The Chairman of the Board :

Prof. J. L. LIONS

Unknown data needs

These simulations have faithfully reproduced the chain of events leading to the failure of the inertial reference systems.

Ariane 501 Inquiry Board report

ARIANE 5

Flight 501 Failure

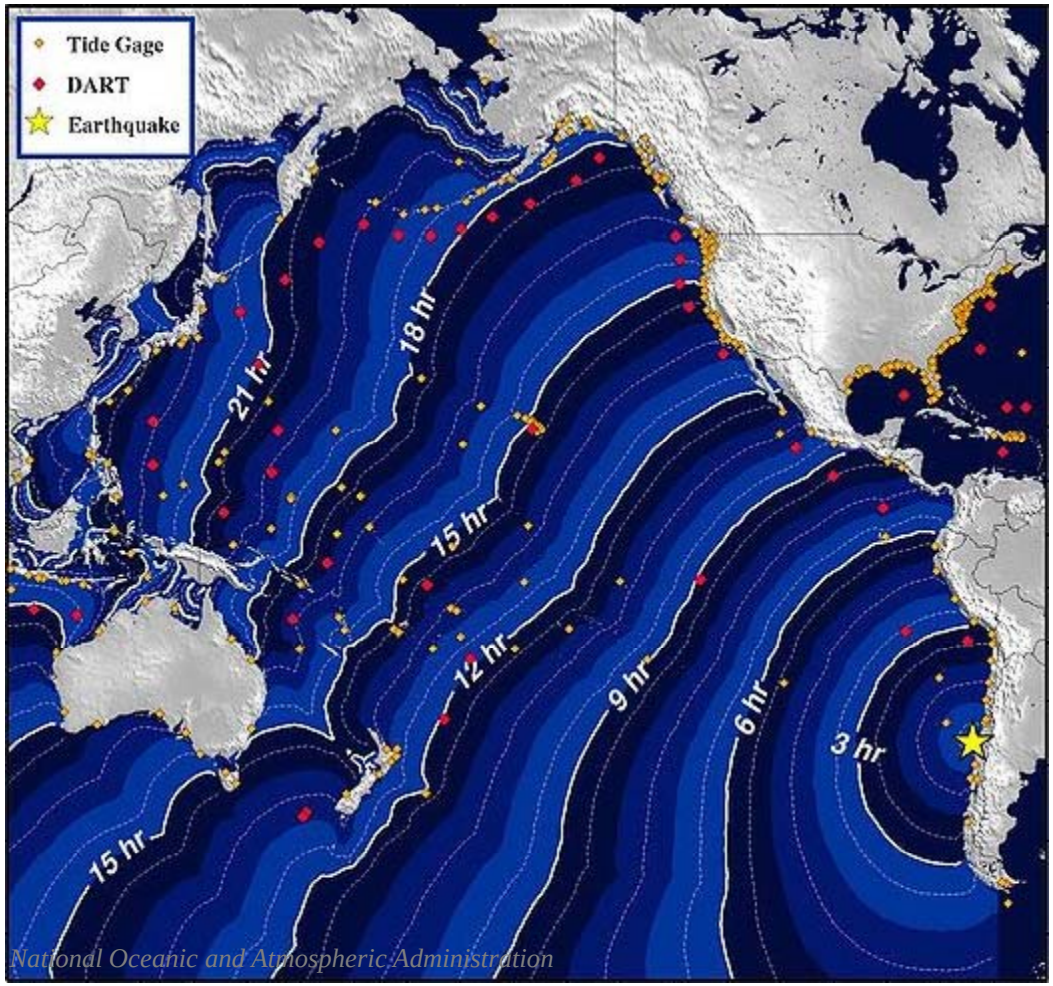
Report by the Inquiry Board

The Chairman of the Board :

Prof. J. L. LIONS

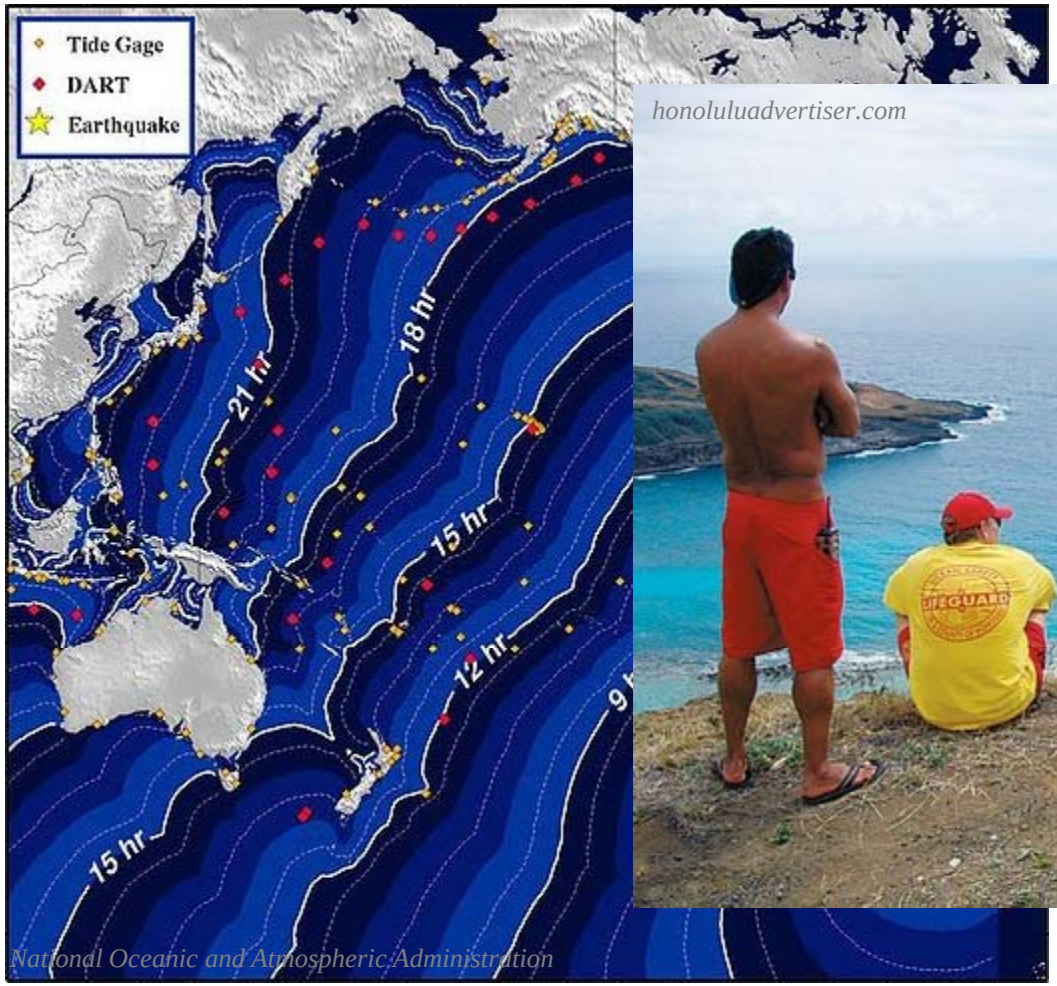
Implicit assumptions obscured (synthetic) data needs …

Insufficient quality of data



National Oceanic and Atmospheric Administration

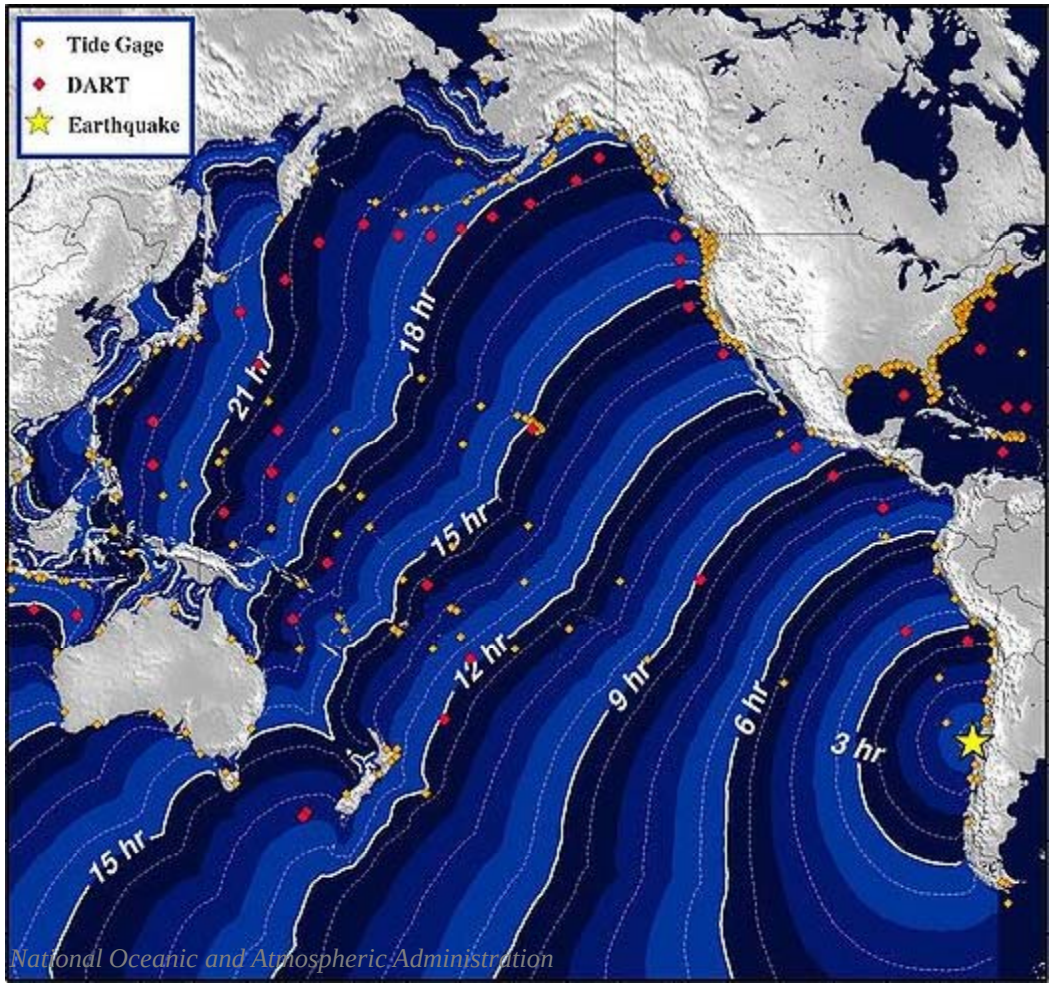
Insufficient quality of data



National Oceanic and Atmospheric Administration

honoluluadvertiser.com

Insufficient quality of data



National Oceanic and Atmospheric Administration

boston.com

THIS STORY HAS BEEN FORMATTED FOR EASY PRINTING

Scientists say tsunami models should be tested

AP Associated Press

By Herbert A. Sample, Associated Press Writer | March 2, 2010

HONOLULU --In the coming months and years, scientists will pore over reams of data from what turned out to be the minuscule tsunami that reached Hawaii on Saturday.

But already, some scientists are saying there is less need for additional measuring equipment of the kind that was placed in the Pacific Ocean after the devastating tsunami that killed 230,000 people around the Indian Ocean in 2004.

Instead, they say there should be a rigorous examination of long-standing assumptions within computer-generated models that are used to estimate the strength and impact of tsunamis.

"Our main problem right now is that we have unsubstantiated assumptions built into our warning system and we really have to check those," said Gerard Fryer, a geophysicist at the Pacific Tsunami Warning Center in Hawaii and formerly a professor at the University of Hawaii.

Had he been asked a week ago whether a magnitude 8.8 earthquake in Chile would cause a destructive tsunami in Hawaii, "I would have said, 'Unquestionably. It's going to be a bad scene,'" Fryer added. "Well, it wasn't. And we have to figure out why it wasn't."

The small tsunami generated by Saturday's quake in Chile also may provide an impetus for the Pacific Tsunami Warning Center to more fully adopt a forecasting system developed by another National Oceanic and Atmospheric Administration agency.

From the data gathered so far, the system designed by the Center for Tsunami Research in Seattle appears to have accurately estimated the severity of the tsunami that reached Hawaii.

"Our forecast played out pretty well," said Vasily Titov, the center's director. Because the system is still in development, its results were initially shared only with Pacific Tsunami Warning Center officials and not the public or news media, he added.

One assumption that Fryer said should be reassessed presumes that the Chilean quake occurred in deeper waters than actually happened. A rupture in deeper seas would have displaced more water and thus resulted in a larger tsunami, Fryer said.

Another assumption says tsunami waves travel at about the same speed, and it does not emphasize the intervals between waves, he said. But wave speed and intervals can affect how tsunamis interact with coastal zones, particularly bays and harbors, Fryer added.

The 1960 Chile quake that spawned huge waves that killed dozens on the Big Island and in Japan featured a longer interval between waves, about 30 minutes, than did the tsunami that lapped at Hawaii's coasts on Saturday, which were about 20 minutes apart. But the current models do not sufficiently take intervals into account, Fryer said.

The models also did not calculate "dispersion," which reduces the strength of tsunami waves as they spread out over the vastness of the Pacific Ocean, Fryer said. However, adding that factor into the models would greatly increase computational costs, he added.

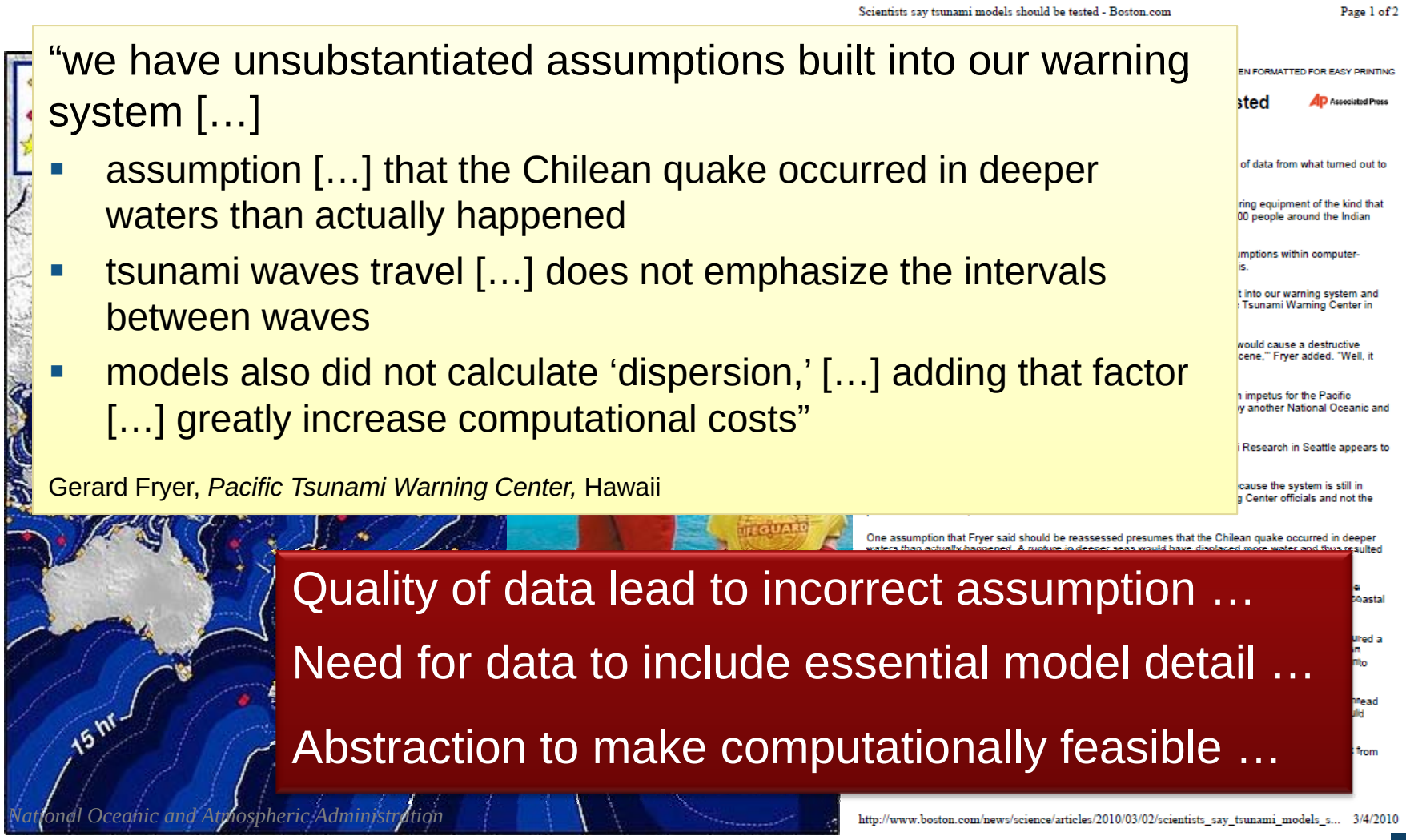
That is not to say that tsunami warnings should not have been issued, he said. Some data such as that from deep sea gauges off the Peruvian coast indicated a destructive tsunami was in the offing, Fryer said.

Insufficient quality of data

"we have unsubstantiated assumptions built into our warning system […]

- assumption […] that the Chilean quake occurred in deeper waters than actually happened
- tsunami waves travel […] does not emphasize the intervals between waves
- models also did not calculate 'dispersion,' […] adding that factor […] greatly increase computational costs"

Gerard Fryer, *Pacific Tsunami Warning Center,* Hawaii



Quality of data lead to incorrect assumption …

Need for data to include essential model detail …

Abstraction to make computationally feasible …

Challenges?

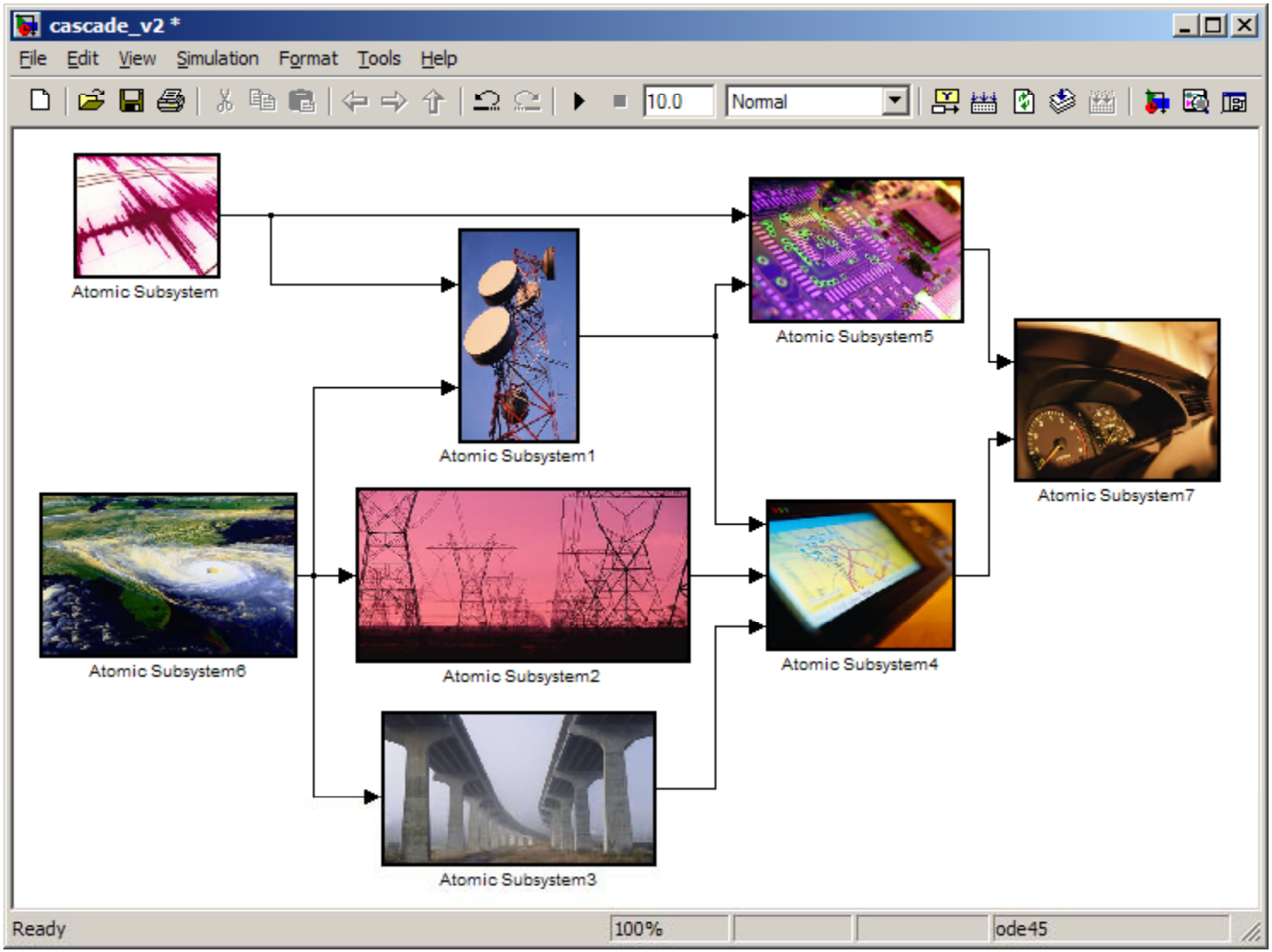
- How do we know the value of a simulation?
  - Is a simulation corroborated by data?
  - How well is it corroborated?
- How do we even know what data we need?
- At what level of abstraction?

Challenges?

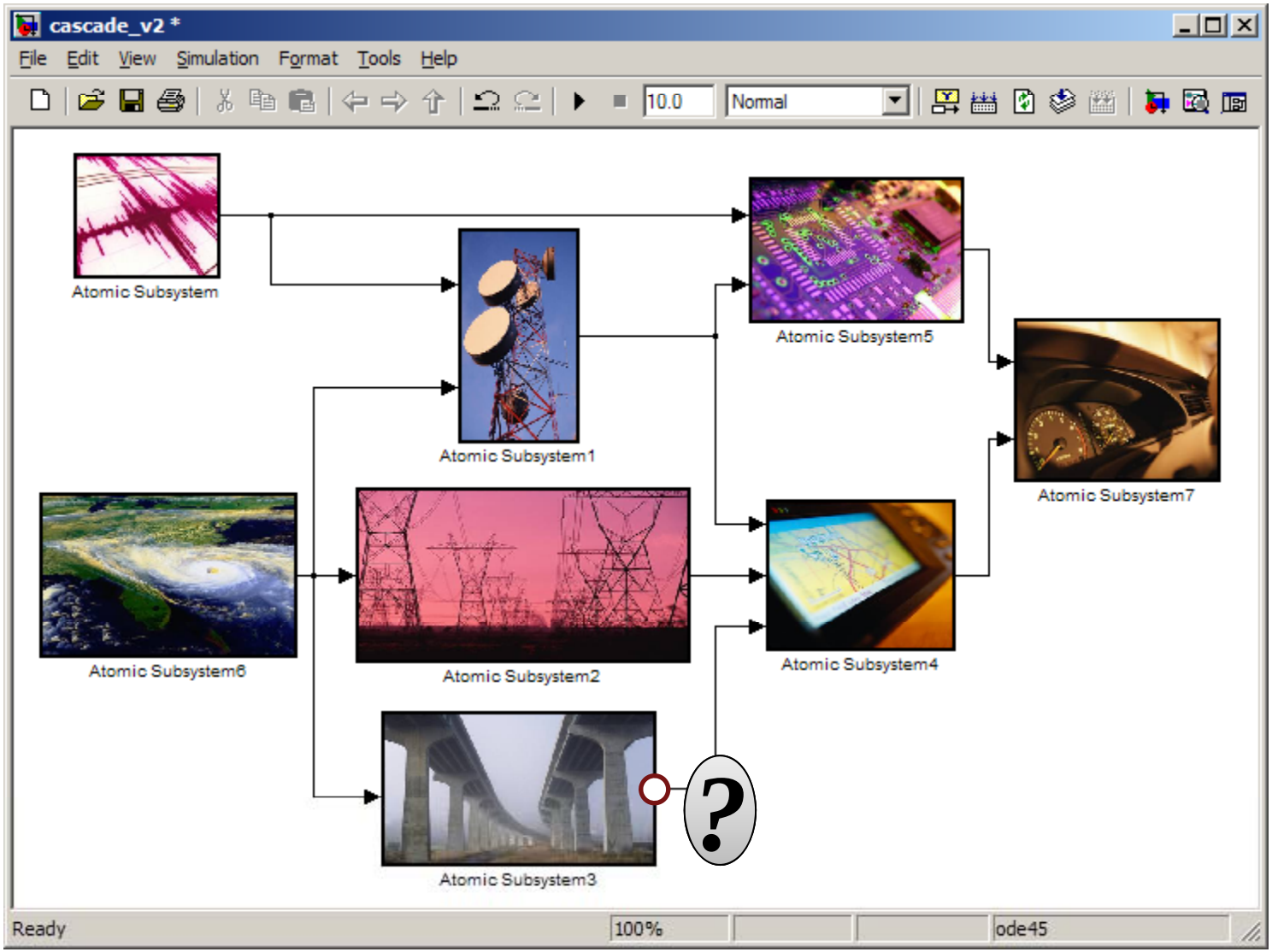
- How do we know the value of a simulation?
  - Is a simulation corroborated by data?
  - How well is it corroborated?
- How do we even know what data we need?
- At what level of abstraction?

And, that is just for a single model; cascading events in the infrastructure require federations of models!

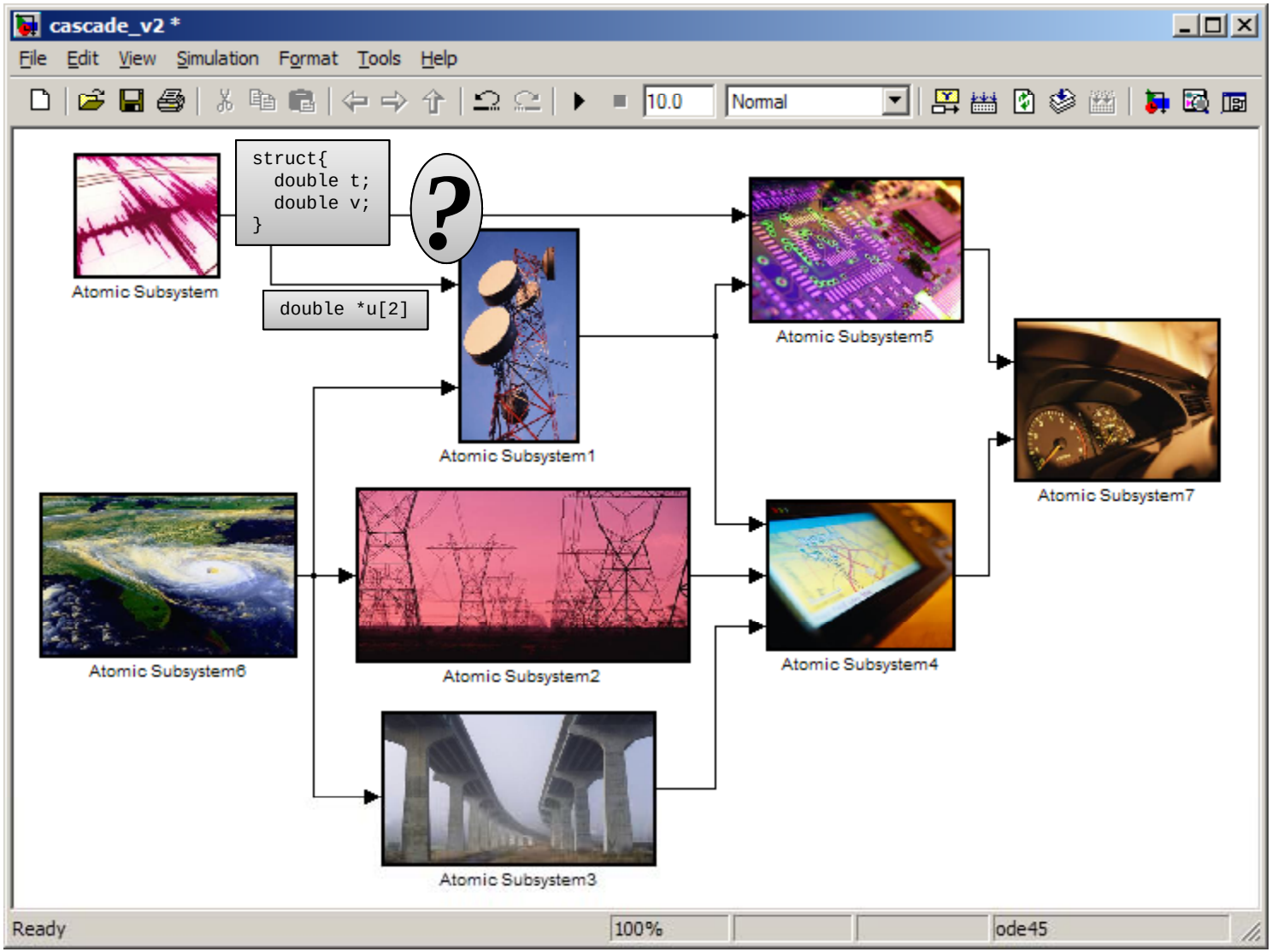
Would not it be nice if, …



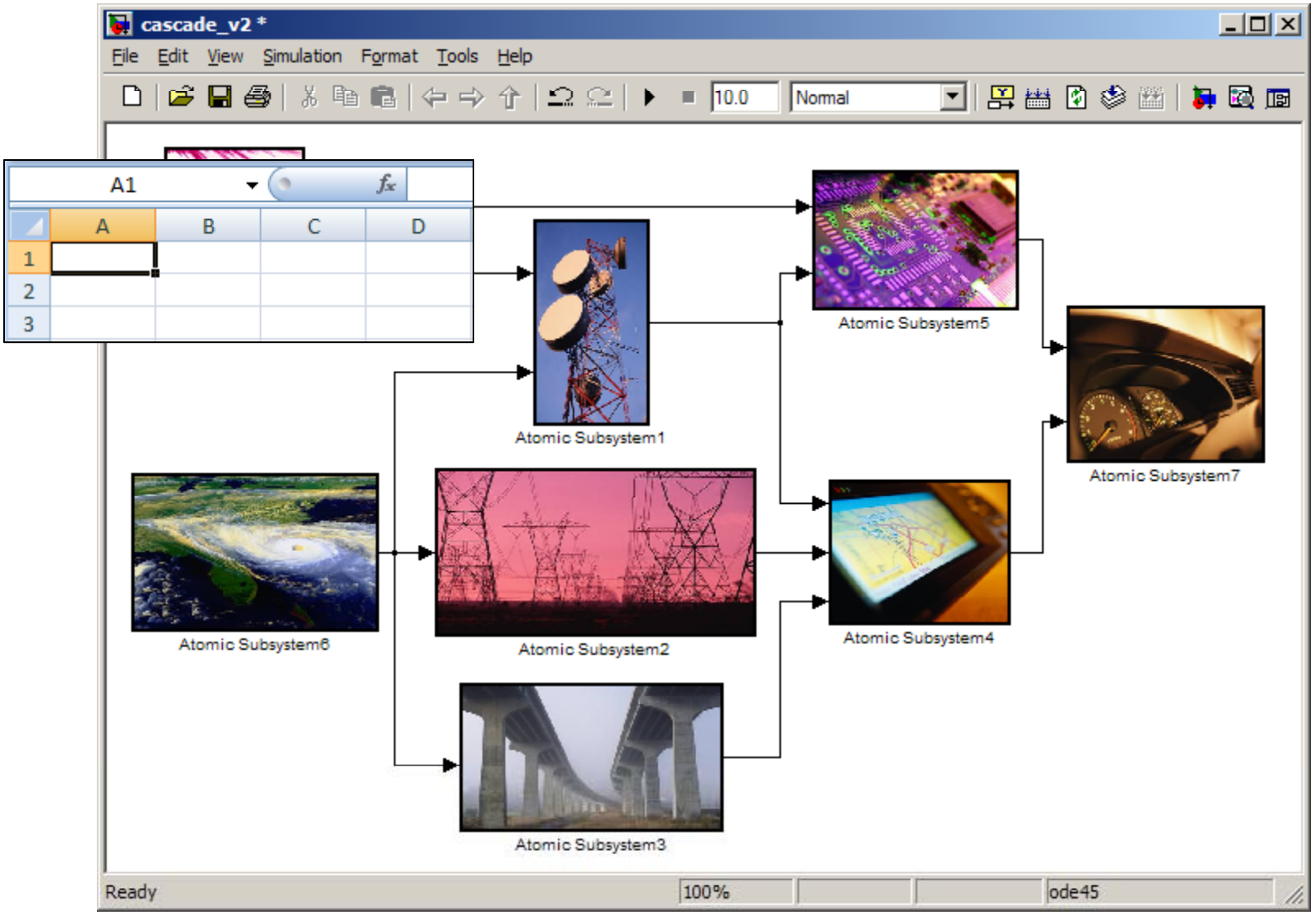
Would not it be nice if, …



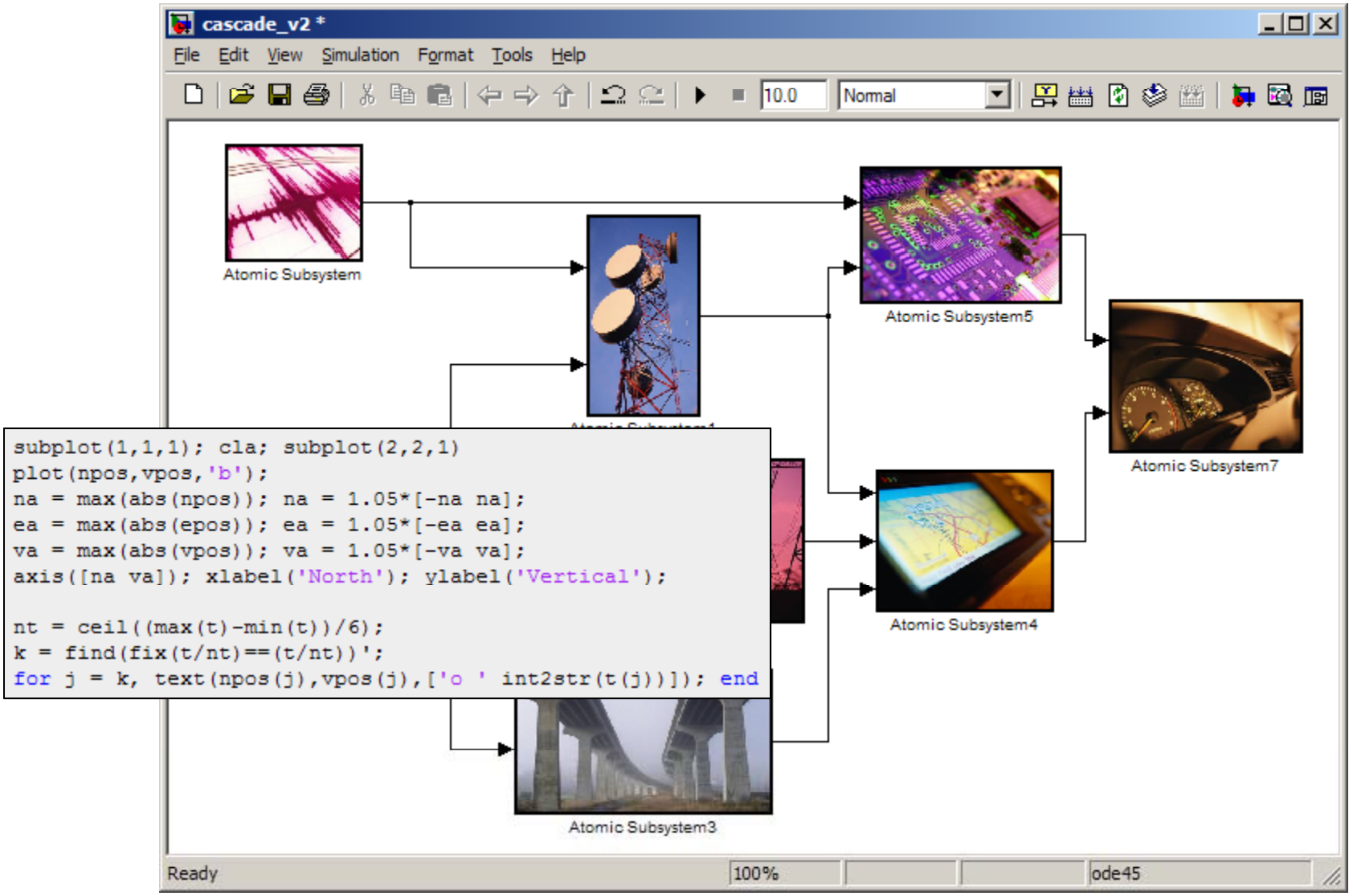
Would not it be nice if, …



Would not it be nice if, …



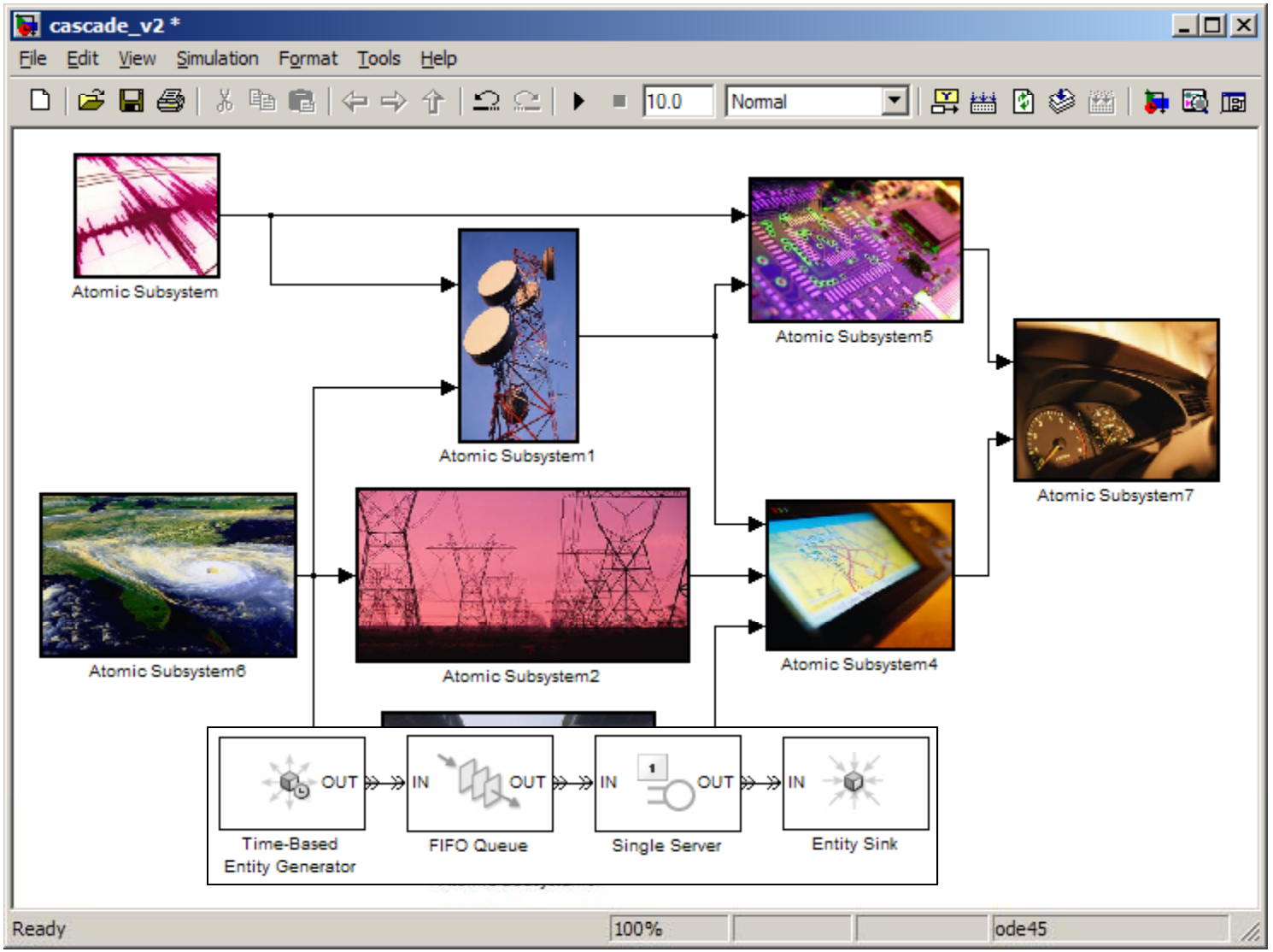
Would not it be nice if, …



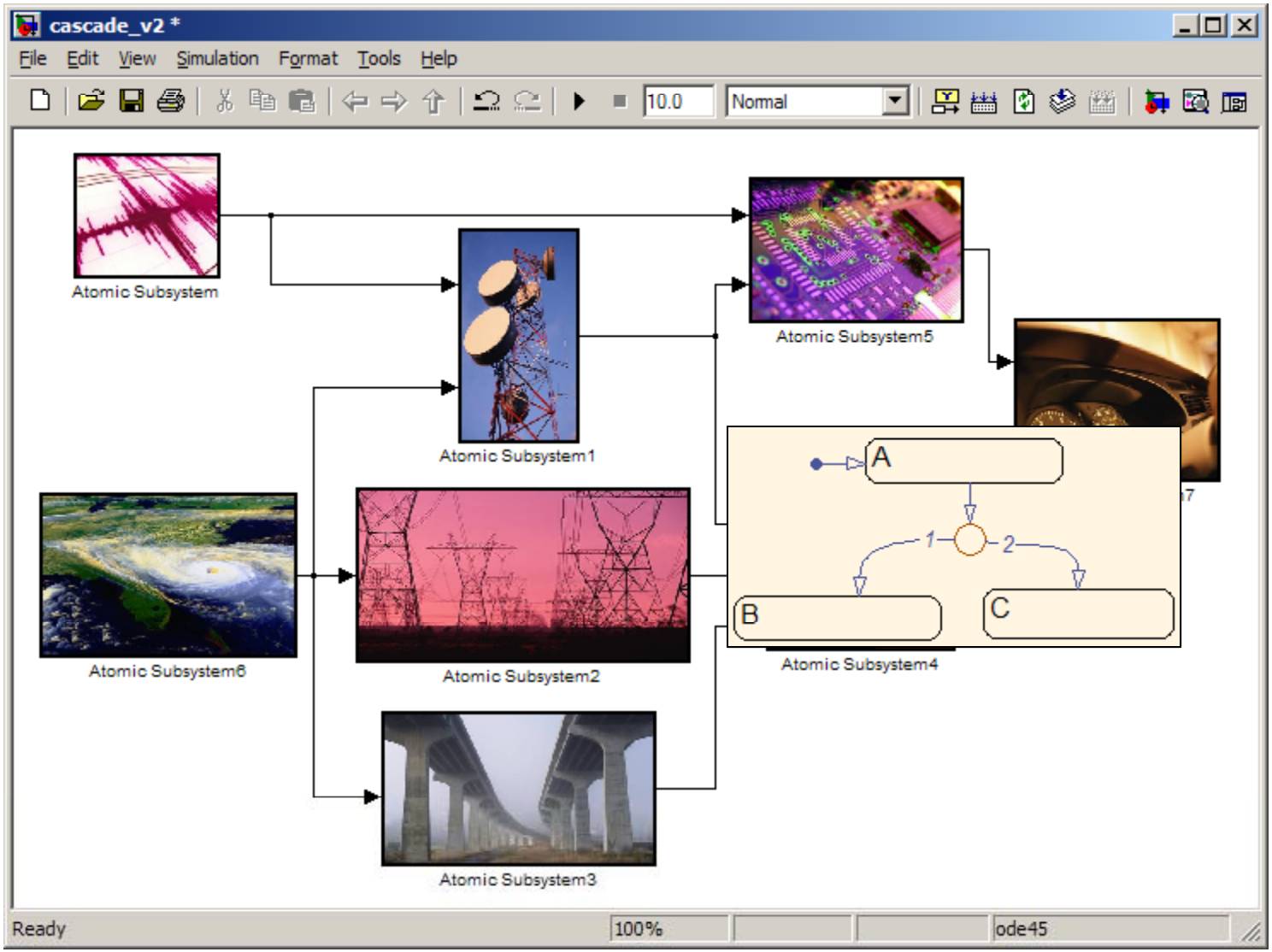

```
subplot(1,1,1); cla; subplot(2,2,1)
plot(npos,vpos,'b');
na = max(abs(npos)); na = 1.05*[-na na];
ea = max(abs(epos)); ea = 1.05*[-ea ea];
va = max(abs(vpos)); va = 1.05*[-va va];
axis([na va]); xlabel('North'); ylabel('Vertical');

nt = ceil((max(t)-min(t))/6);
k = find(fix(t/nt)==(t/nt))';
for j = k, text(npos(j),vpos(j),['o ' int2str(t(j))]); end
```

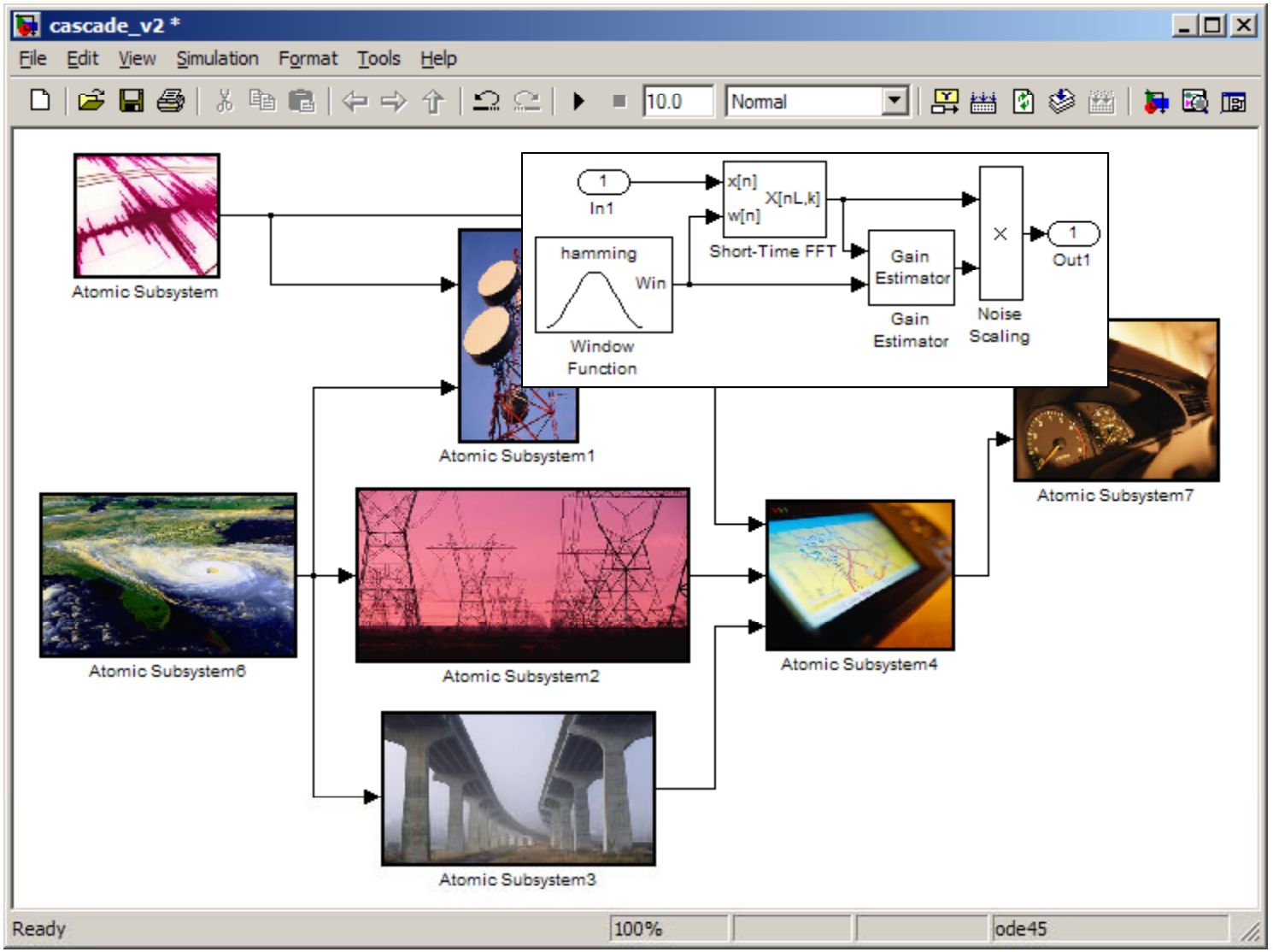
Would not it be nice if, …



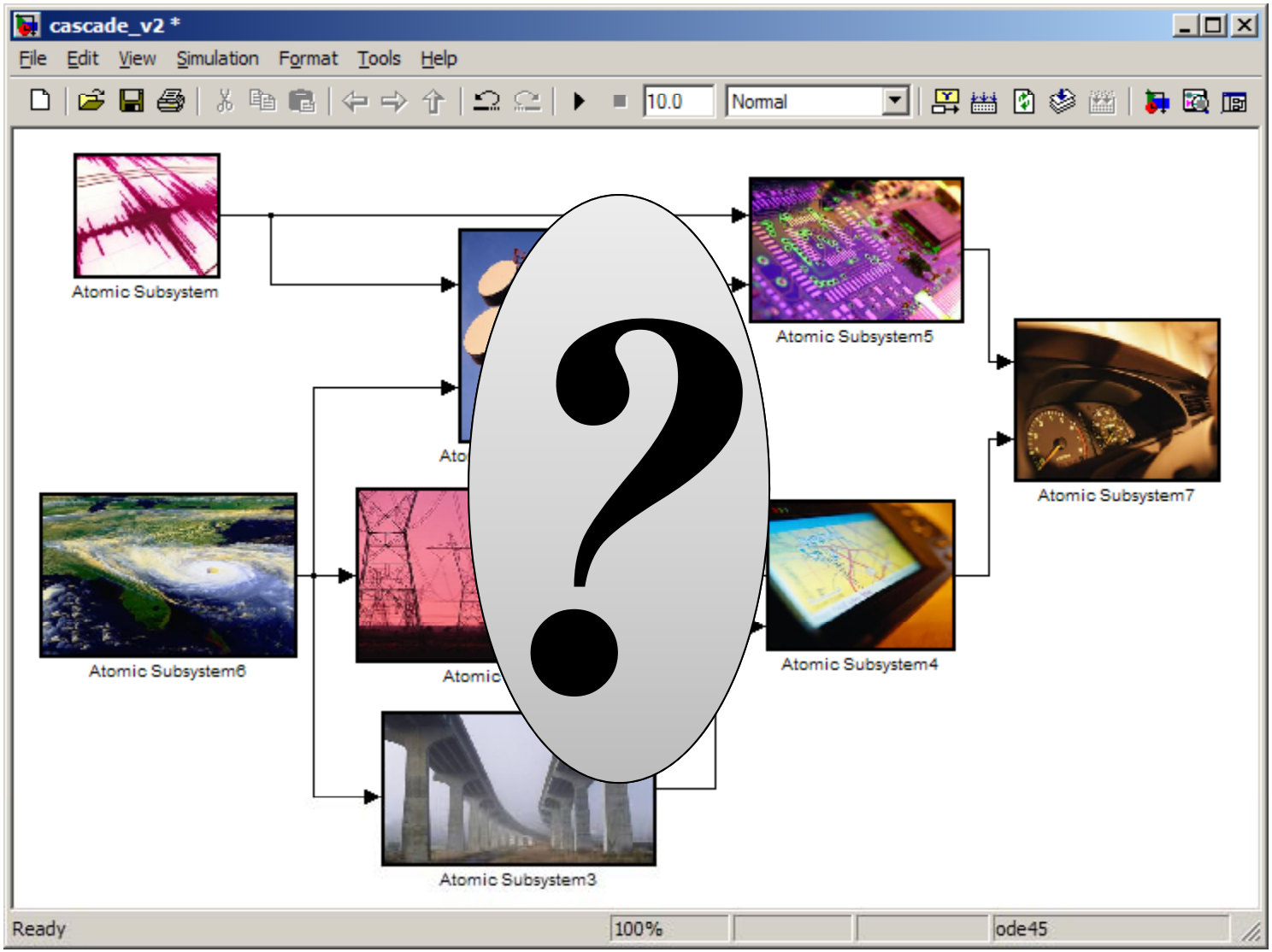
Would not it be nice if, …



Would not it be nice if, …



Would not it be nice if, …



Data in federated models

- Interfaces
  - Number
  - Type
  - Meaning
- Models
  - Formalisms
  - Paradigms
  - Abstraction

Some questions—how do we?

- Scope the problem to determine data needs
- Make data available
  - technologically
  - organizationally
- Assess the uncertainty of data
  - quantitative
  - qualitative
- Infer data at different levels of detail
- Determine coverage
  - which operational regions
  - how well

Panelists

- Mohamed Belkhayat, *Northrop-Grumman*
- Judith C Spering, *Boeing*
- Zubin Wadia, *Civiguard*
- Philip C. Cooley, *RTI International*
- Trevor Ament, *Australian Reinsurance Pool Corporation*

Homeland Security

Science and Technology