



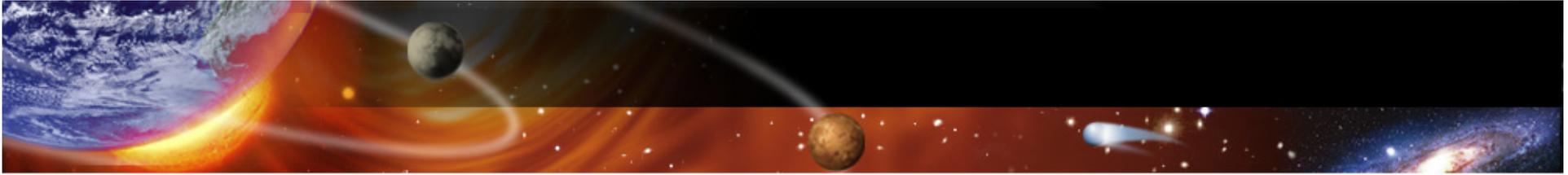
Welcome to CESS 2011

Michael Way & Catherine Naud
NASA/Goddard Institute for Space Studies

<http://www.giss.nasa.gov/meetings/cess2011>

CESS 2011





Why are we here?

Interdisciplinary Research!

Mike's example: Gaussian Process Regression:

Inverting large non-sparse matrices with San Jose State, Mathematics Department

Domain product?: Predicting Photometric Redshifts for Cosmology

Leveraged originally from work by Srivastava, Nemani & Oza:

Earth Imaging from Space: Predicting AVHRR channels from MODIS



Science Magazine: 11 Feb 2011

“Dealing with Data”

<http://www.sciencemag.org/site/special/data>



What did I learn? We should have invited Biologists & Chemists!!

Work on microarray data alone could be heavily leveraged.

(“The Elements of Statistical Learning” Hastie, Tibshirani, Friedman)

What did I *not* learn? That we have data challenges in all fields

Computational, Storage (Data Standards!), Viz

Climate Science, Genomic Data, Ecology, etc...

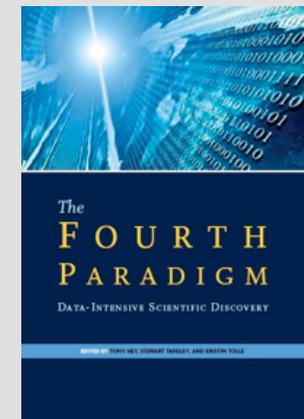
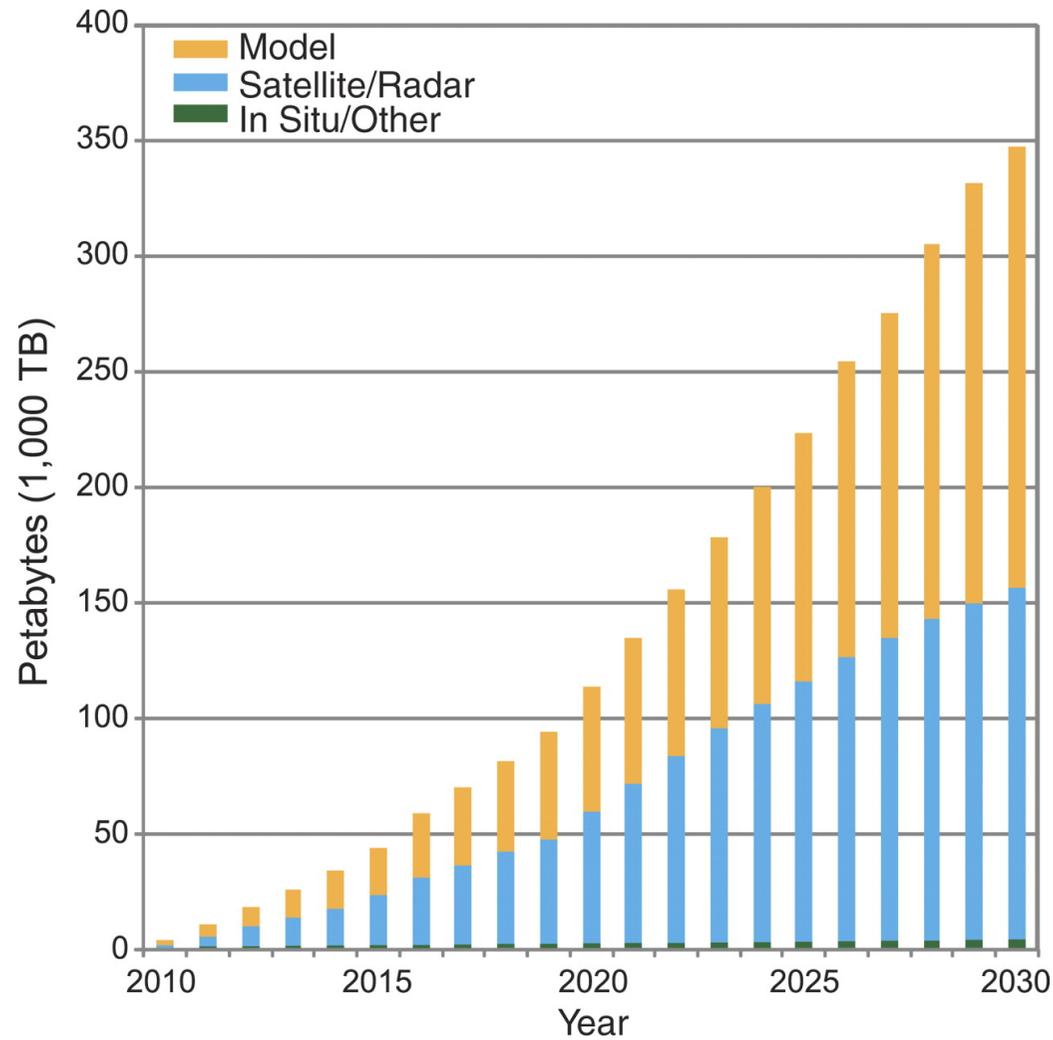


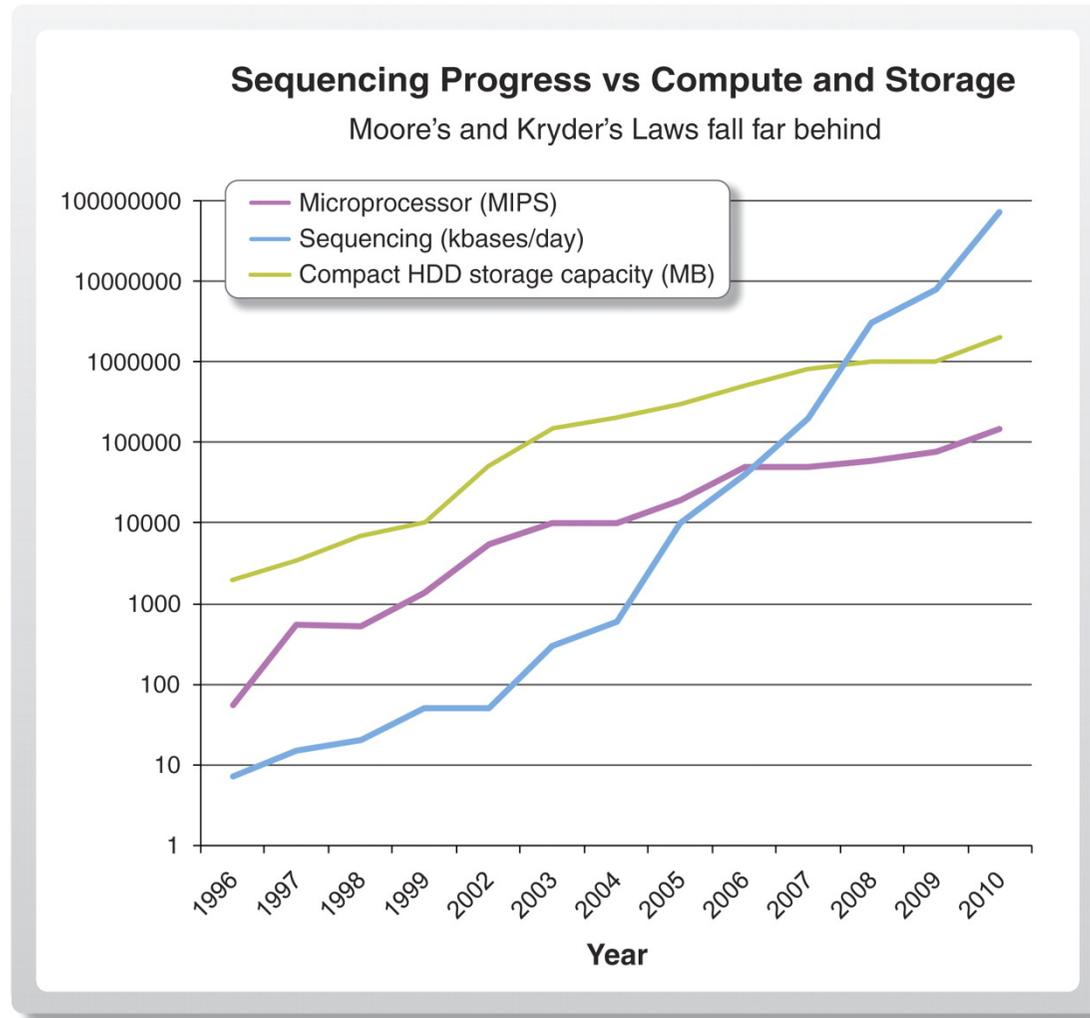
Fig. 2 The volume of worldwide climate data is expanding rapidly, creating challenges for both physical archiving and sharing, as well as for ease of access and finding what's needed, particularly if you are not a climate scientist.



J T Overpeck et al. Science 2011;331:700-702



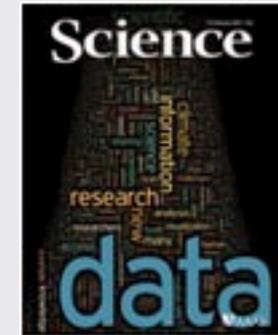
Fig. 1 A doubling of sequencing output every 9 months has outpaced and overtaken performance improvements within the disk storage and high-performance computation fields.



S D Kahn Science 2011;331:728-729



Changing the Equation on Scientific Data Visualization (Fox & Hendler)



“... the creation of visualizations for complex data remains more of an art form than an easily conducted practice.”

Of course people have tried to rectify this sometimes even with inexpensive solutions →

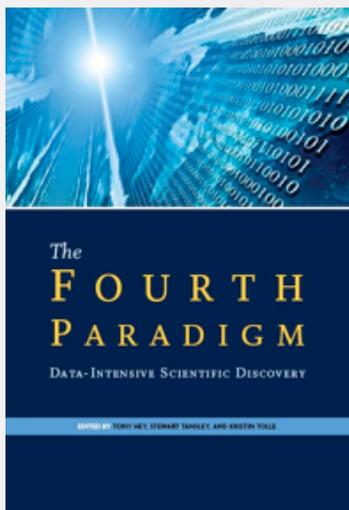
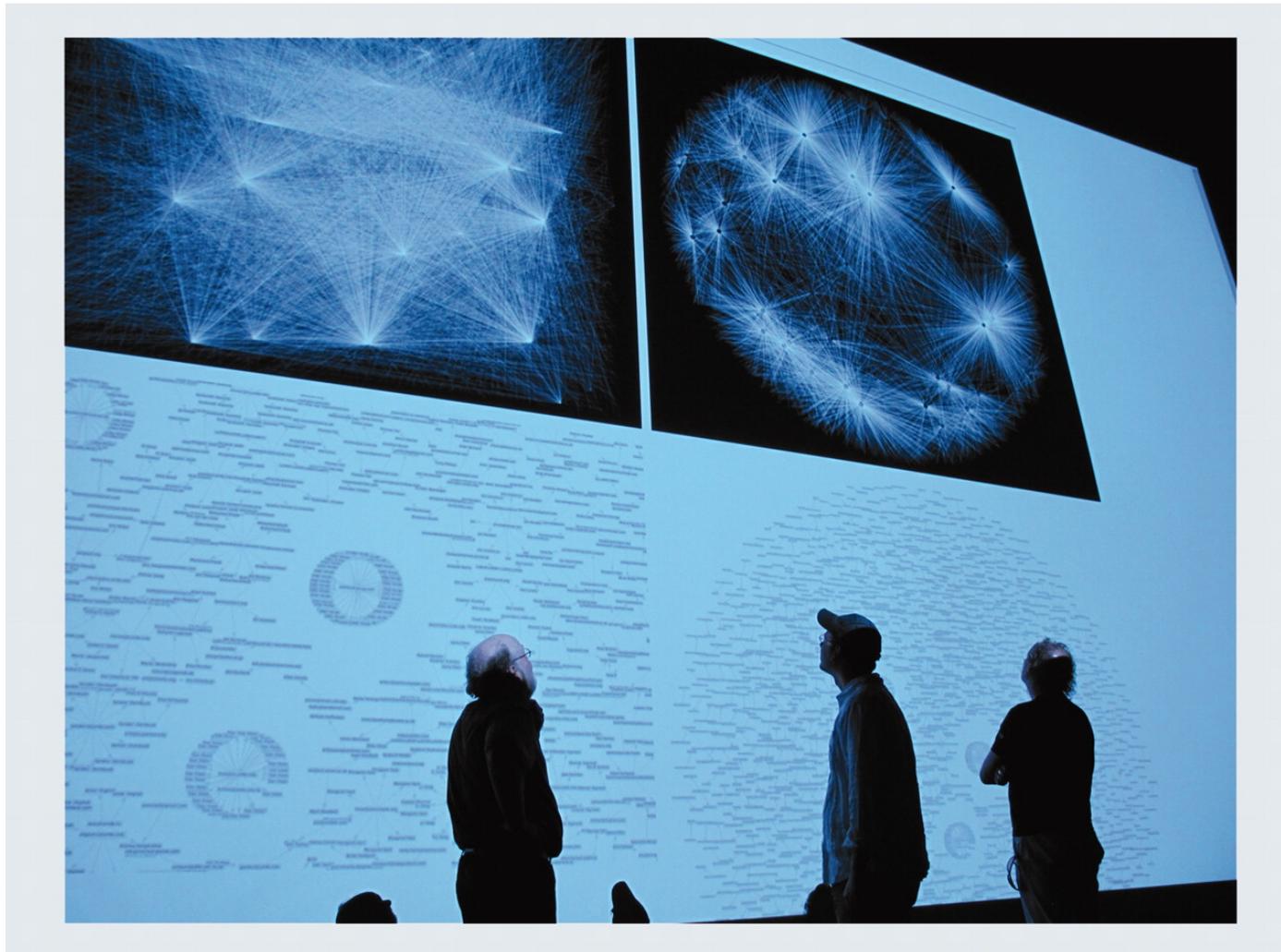


Fig. 2 **Projecting** the results of network-visualization tools onto a large screen allows a group of scientists to explore the relations among a large number of data elements without the specific need for expensive visualization tools.

P Fox, J Hendler Science 2011;331:705-708

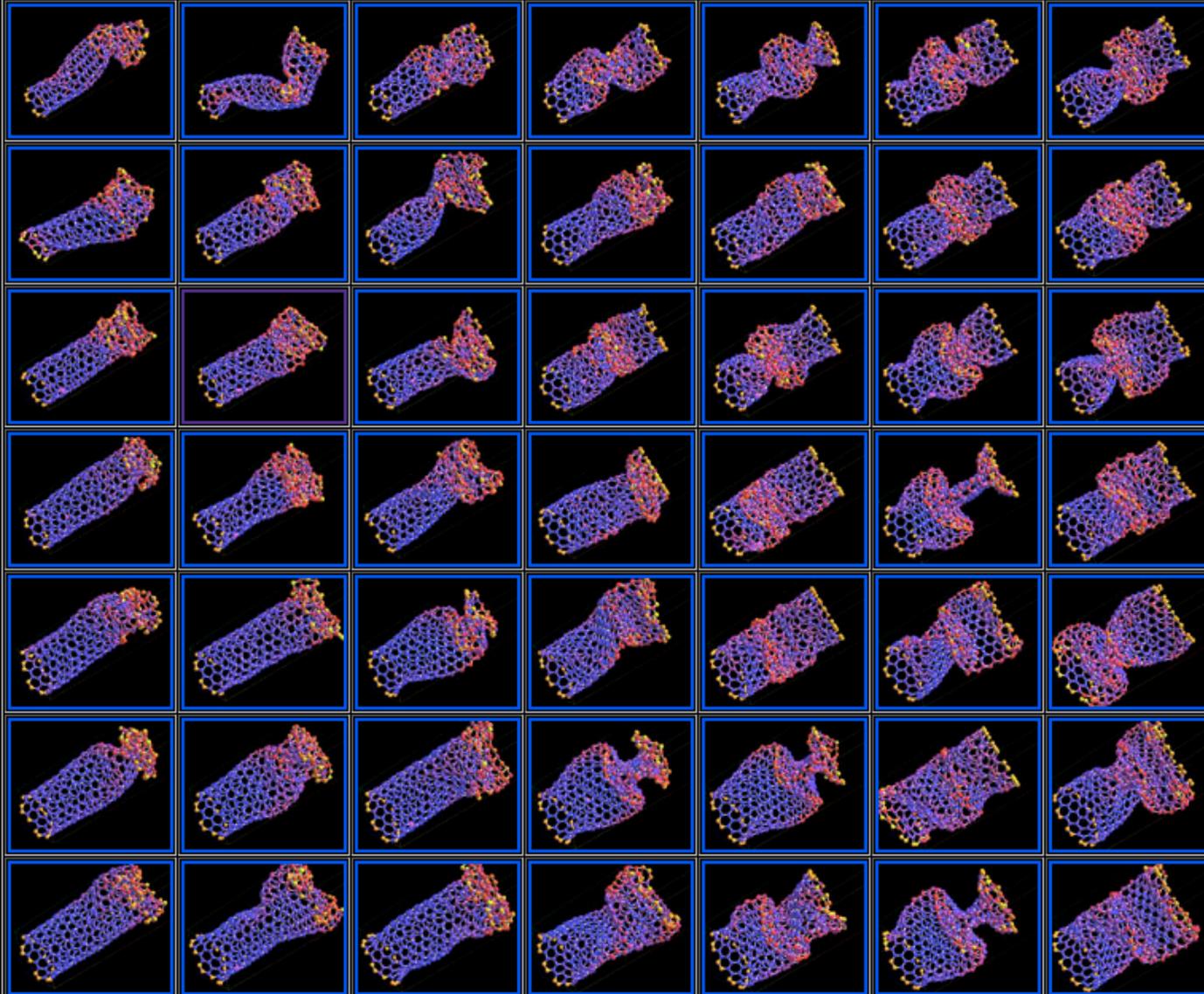


The more expensive way is easier to find → →

NASA/Ames Hyperwall

Nanotubes

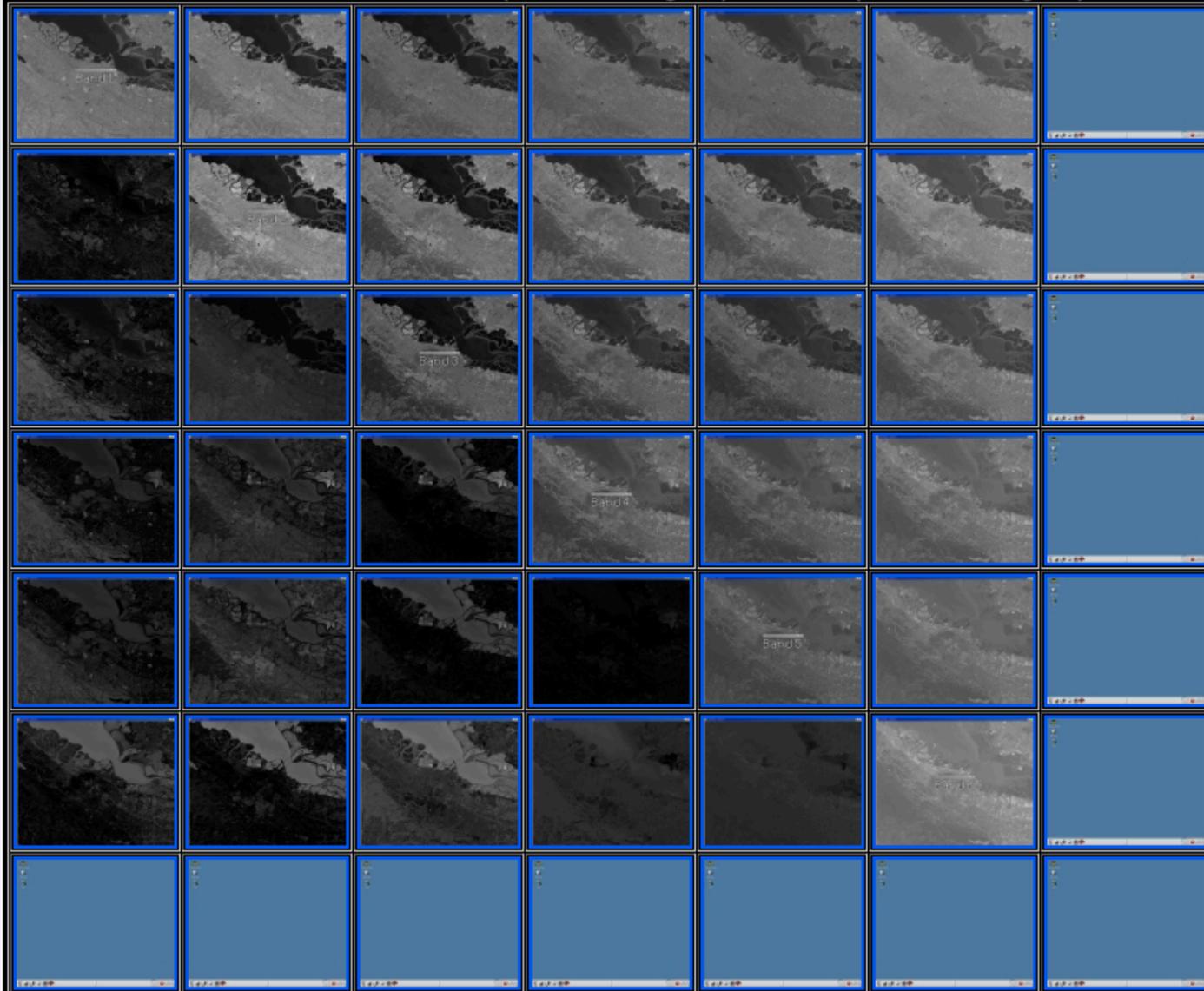
49 different nanotubes after identical compression experiments.



NASA/Ames Hyperwall

Bay Area Multi-Spectral

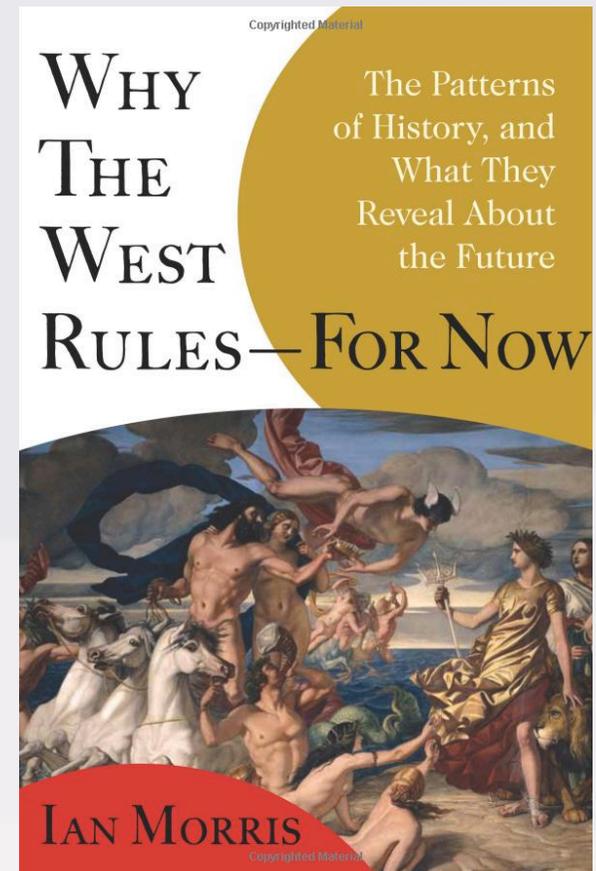
Sums (above main diagonal), Difference (below main diagonal).



The Humanities

So you don't think I'm a science chauvinist!
We know **data** is also allowing new
collaborations within the humanities
& even between science and the humanities:

- History
- Sociology
- Anthropology, Archaeology/Carbon Dating
- Geology
- Genetics



Today's Agenda

Time	Speaker	Title
10:00-10:20	Michael Way & Catherine Naud	Introductions, Coffee, Setup
10:20-10:45	Victor de la Pena	How long will it take. A historical approach to boundary crossing
10:45-11:30	Eyal Kazin (NYU)	Cosmology through the Large-Scale Structure of the Universe
11:30-12:15	William Heavlin (Google Inc)	On the shoulders of Gauss and Bessel: links, chunks, spheres, & conditional models
12:15-1:00	Kirk Borne (George Mason Univ)	Mining Citizen Science Data: Machine Learning Challenges
1:00-2:30	LUNCH BREAK -----	
2:30-3:15	Claire Monteleoni (Columbia/CS)	Tracking Climate Models: Advances in Climate Informatics
3:15-4:00	Kevin Knuth (SUNY-Albany)	Spectral Analysis Methods for Complex Source Mixtures
4:00-4:45	Joshua Peek (Columbia/Astronomy)	Beyond Objects: Using machines to understand the diffuse universe
4:45-5:00	Michael Way (NASA/GISS)	Viewpoints: A high-performance visualization and analysis tool

