Simulation-informed optimization and techniques for big data mining

Urban Dynamics Institute
JICS Auditorium
Nov. 17, 2014

Joshua New, Ph.D.

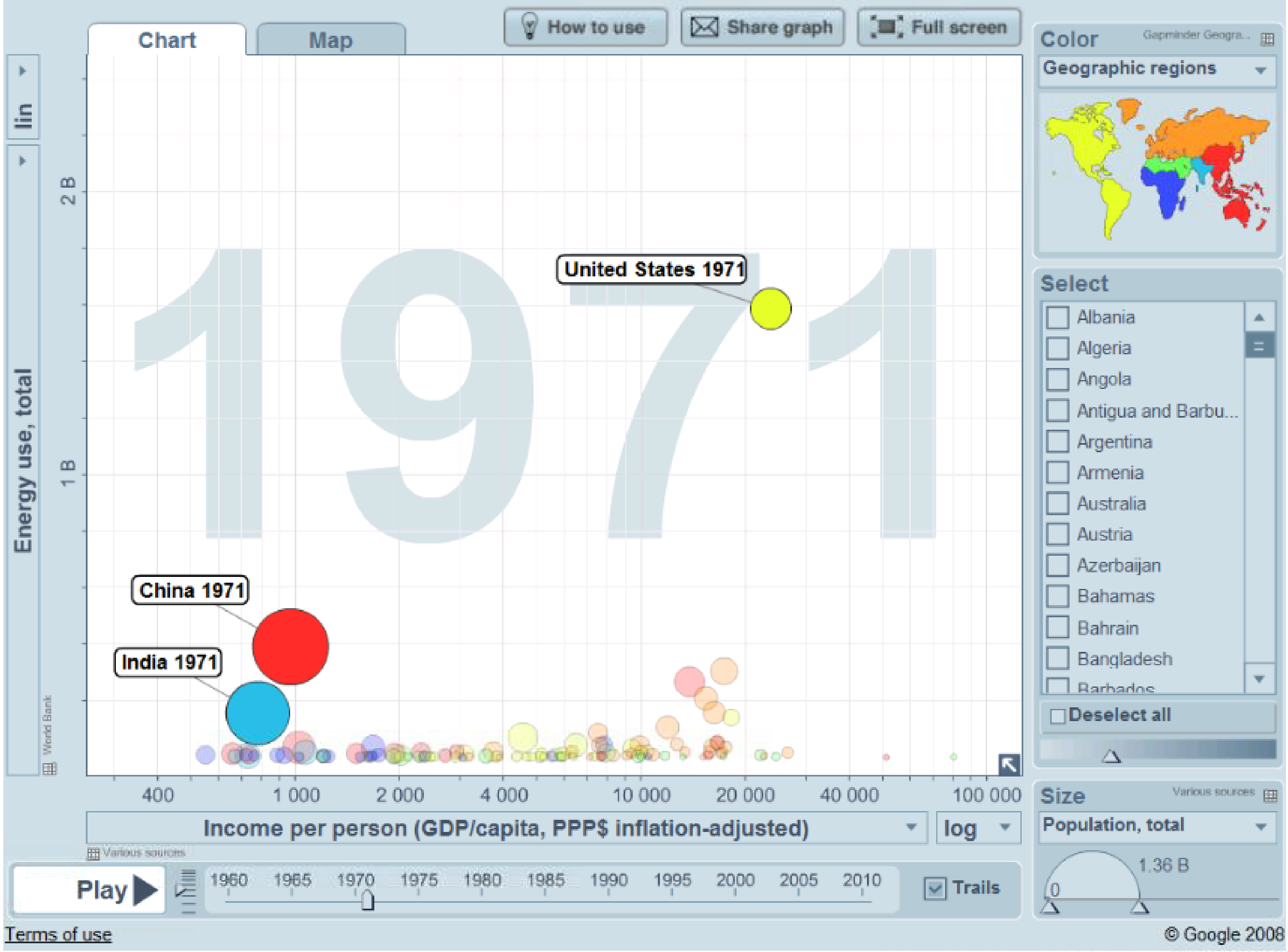
865-241-8783

newjr@ornl.gov

Urban dynamics and resource impact

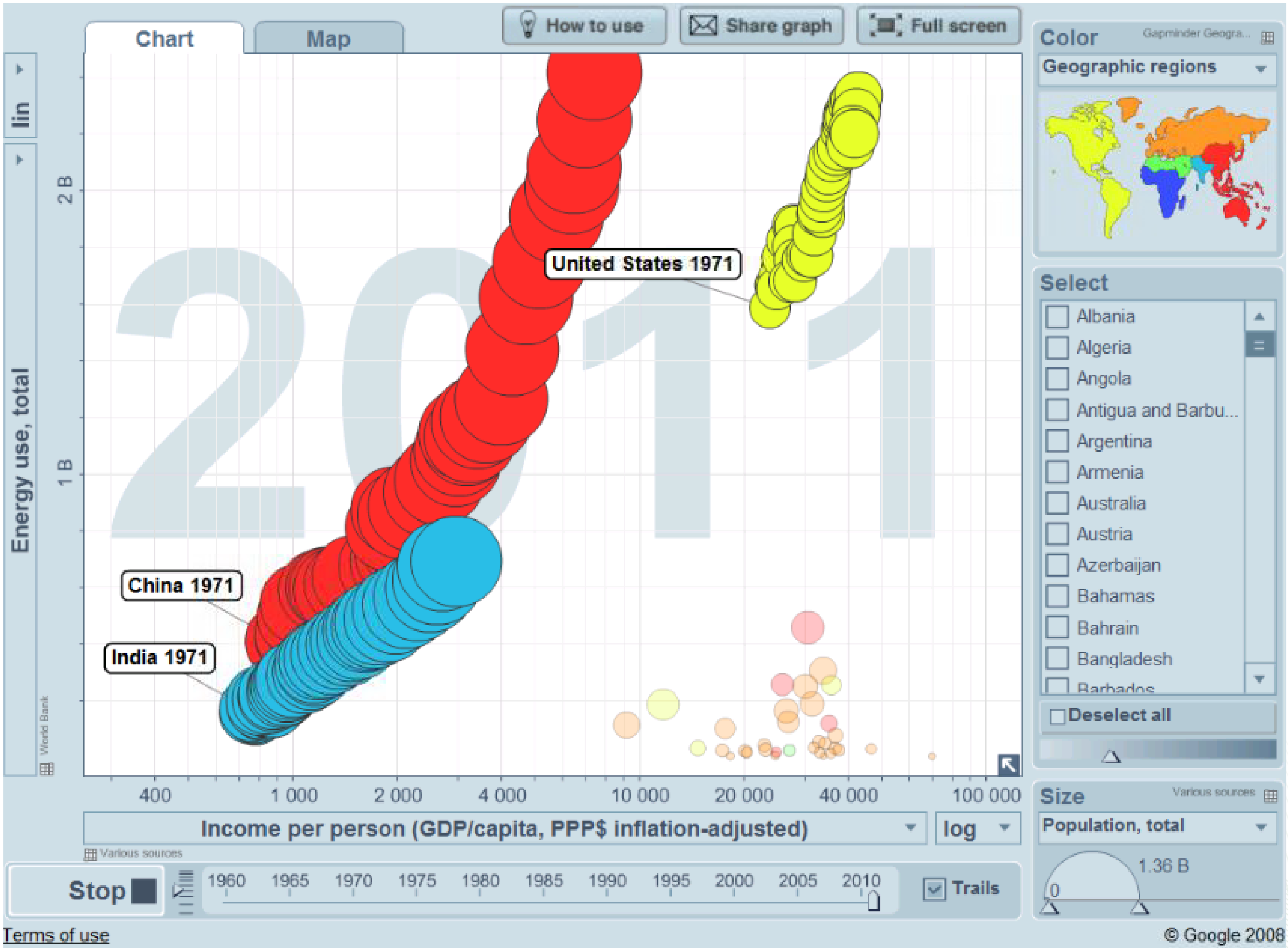
- Americans spend 90% of time indoors
  - U.S. Environmental Protection Agency. 1989. Report to Congress on indoor air quality: Volume 2. EPA/400/1-89/001C. Washington, DC.
- Internet of Things (IoT)
  - Anonymized cell phone records - instrumenting people
  - Cloud-connected wireless sensor networks - $100 billion market in 2018
    - OnWorld WSN report
  - IoT in smart Buildings (BIoT) - $85 billion market in 2020
    - Memoori report
  - 13% of US broadband homes have smart-home devices
  - Nest thermostat in over 1 million homes - 20-second data on occupancy
  - Health monitoring devices (Fitbit)
  - Self-learning home systems

A brief history of energy and life quality



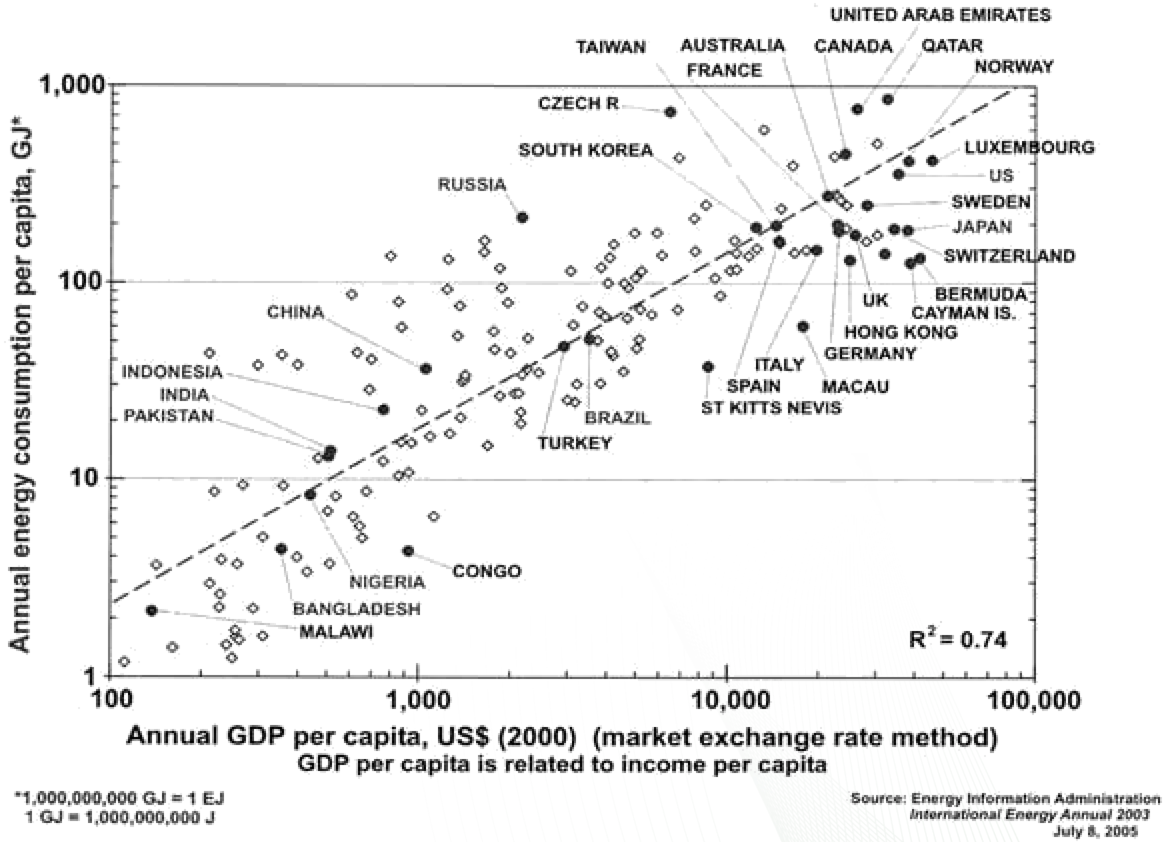
OAK RIDGE
National Laboratory

A brief history of energy and life quality



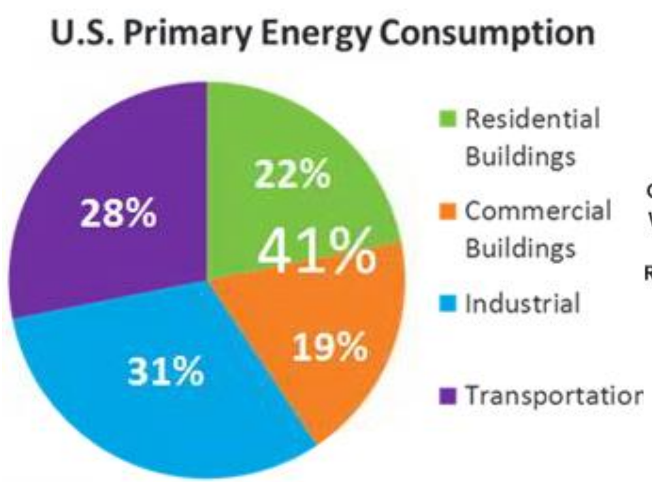
Sustainability is the defining challenge

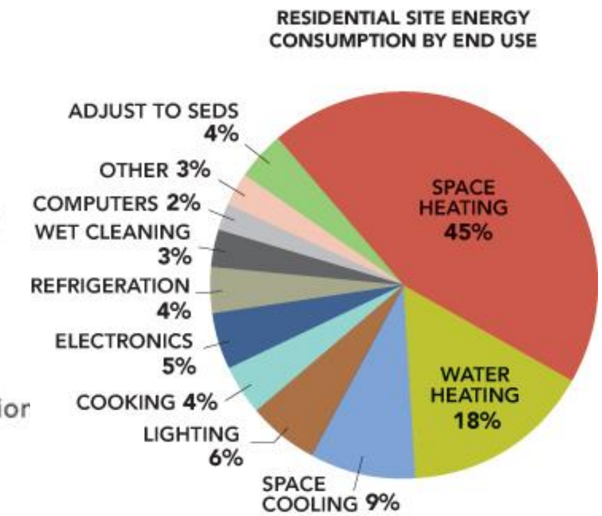
- Buildings in U.S.
  - 41% of primary energy/carbon 73% of electricity, 34% of gas
- Buildings in China
  - 60% of urban building floor space in 2030 has yet to be built
- Buildings in India
  - 67% of all building floor space in 2030 has yet to be built

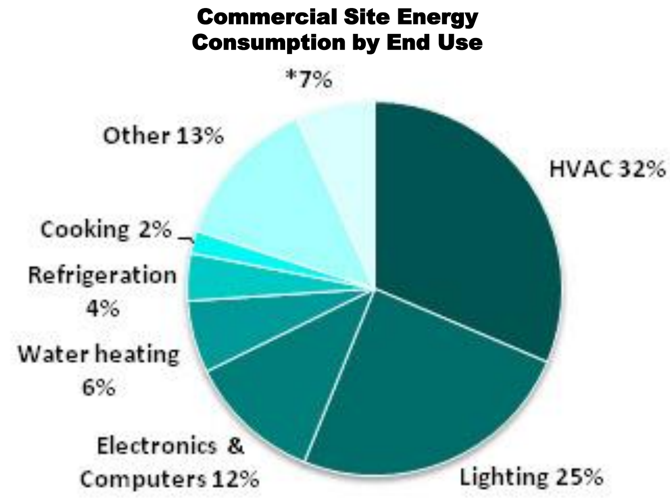


OAK RIDGE National Laboratory

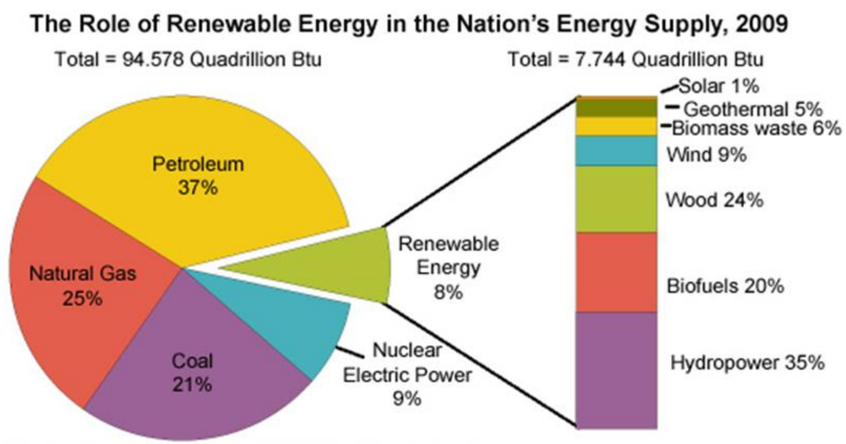
Energy Consumption and Production



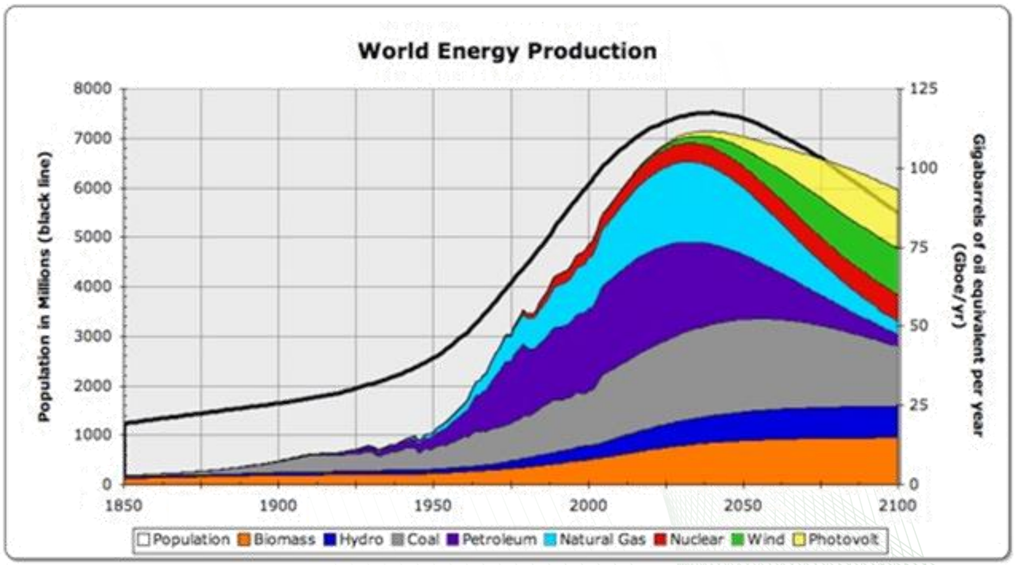




TN 2012 Electric Bill - $1,533



Note: Sum of components may not equal 100% due to independent rounding.
Source: U.S. Energy Information Administration, *Annual Energy Review 2009*, Table 1.3, Primary Energy Consumption by Energy Source, 1949-2009 (August 2010).



OAK RIDGE National Laboratory

Presentation summary

- Scientific Paradigms
- Roof Savings Calculator
- Visual Analytics
- Knowledge Work
- Autotune
- Publications

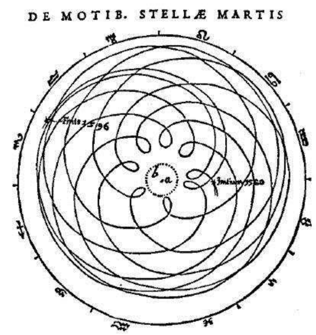
Presentation summary

- Scientific Paradigms (context)
- Roof Savings Calculator
- Visual Analytics
- Knowledge Work
- Autotune
- Publications

4th Paradigm –
The Science behind the Science

- Empirical – guided by experiment/ observation
  - In use thousands of years ago, natural phenomena
- Theoretical – based on coherent group of principles and theorems
  - In use hundreds of years ago, generalizations
- Computational – simulating complex phenomena
  - In use for decades
- Data exploration (eScience) – unifies all 3
  - Data capture, curation, storage, analysis, and visualization
  - Jim Gray, free PDF from MS Research

Tycho Brahe



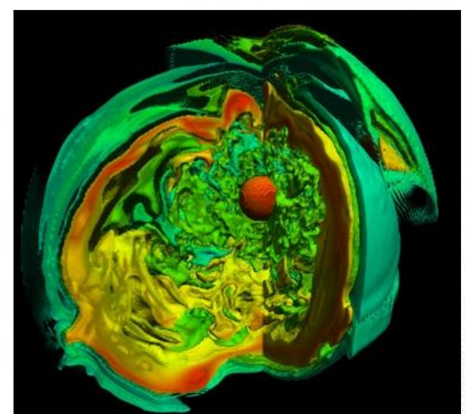
Johannes Kepler

$$\oint \mathbf{E} \cdot d\mathbf{A} = q / \varepsilon_0$$

$$\oint \mathbf{B} \cdot d\mathbf{A} = 0$$

$$\oint \mathbf{E} \cdot d\mathbf{S} = -d\Phi_B / dt$$

$$\oint \mathbf{B} \cdot d\mathbf{S} = \mu_0 i + \mu_0 \varepsilon_0 d\Phi_E / dt$$



Oak Ridge National Laboratory

9

Presentation summary

- Scientific Paradigms
- Roof Savings Calculator
- Visual Analytics
- Knowledge Work
- Autotune

10

Urban Heat Island Effect and Albedo Engineering

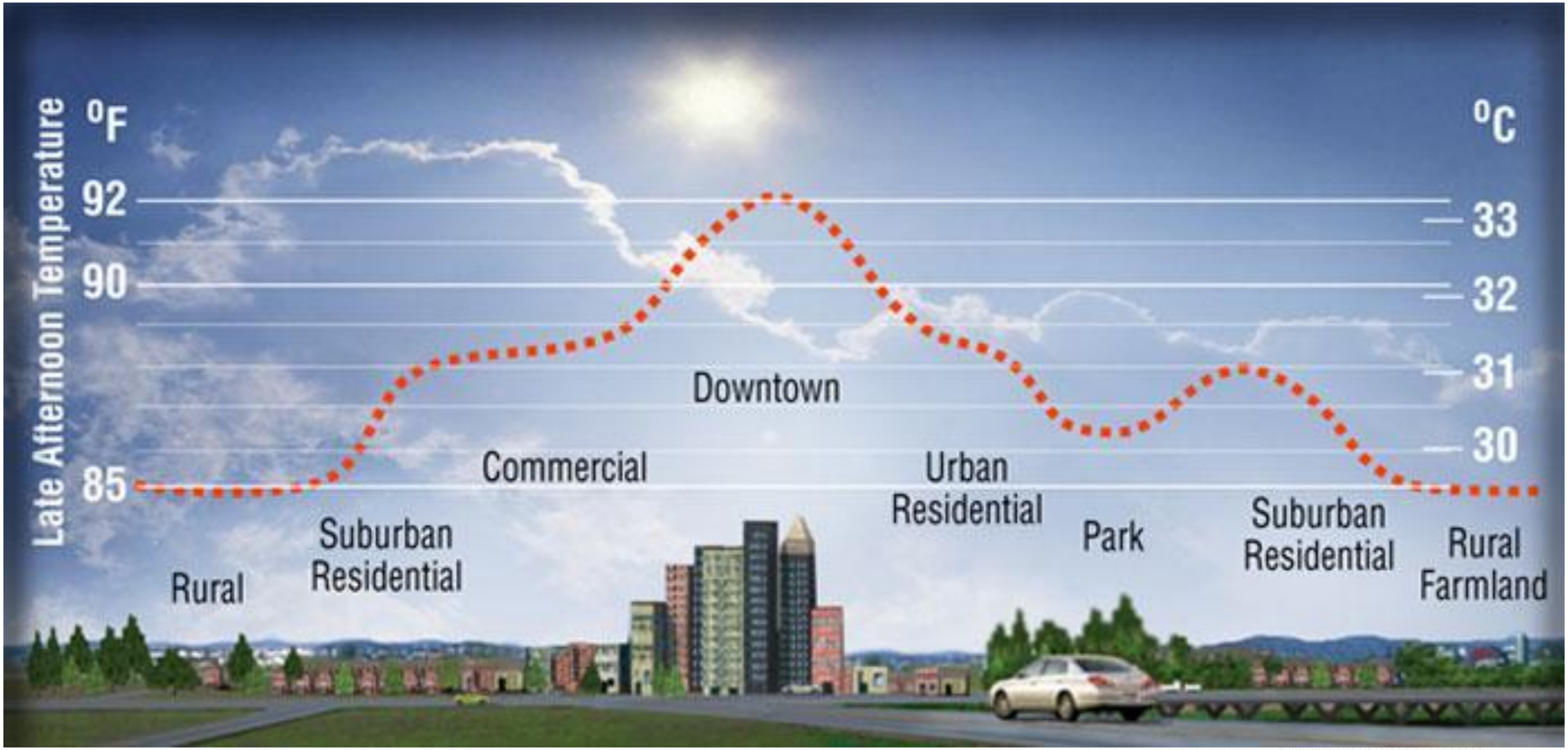
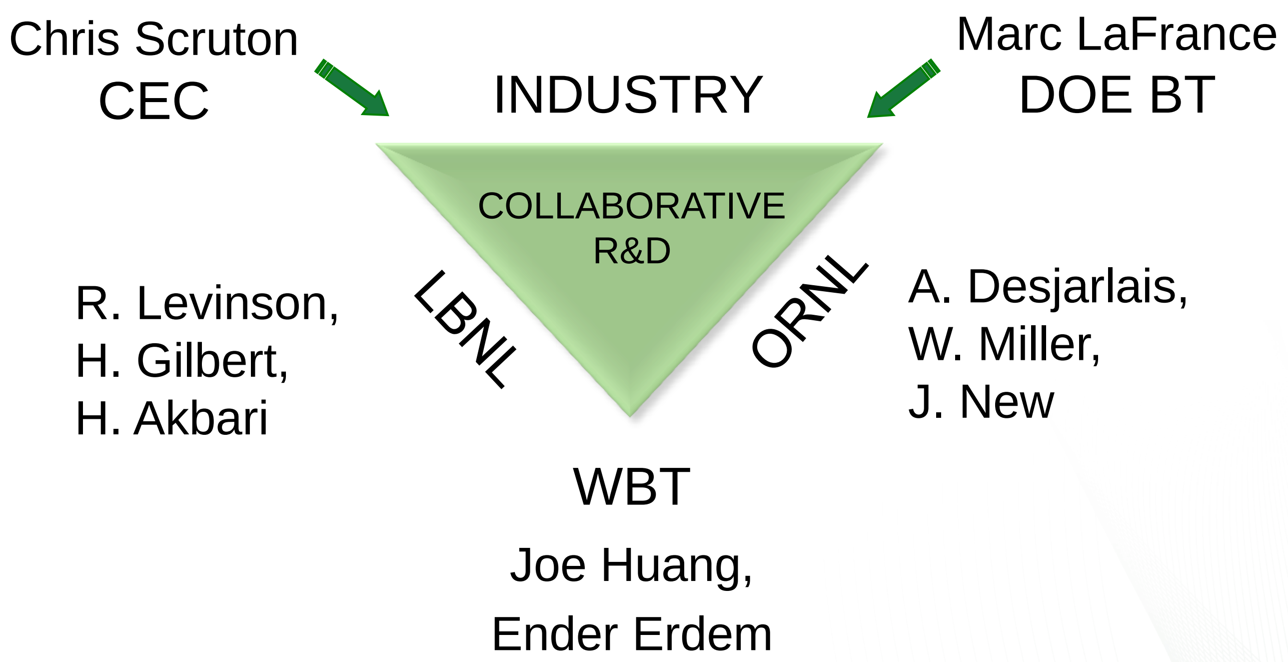


Image from Lawrence Berkeley National Laboratory

Computer tools for simulating cool roofs

Roof Savings Calculator (RSC)



Roof Savings Calculator

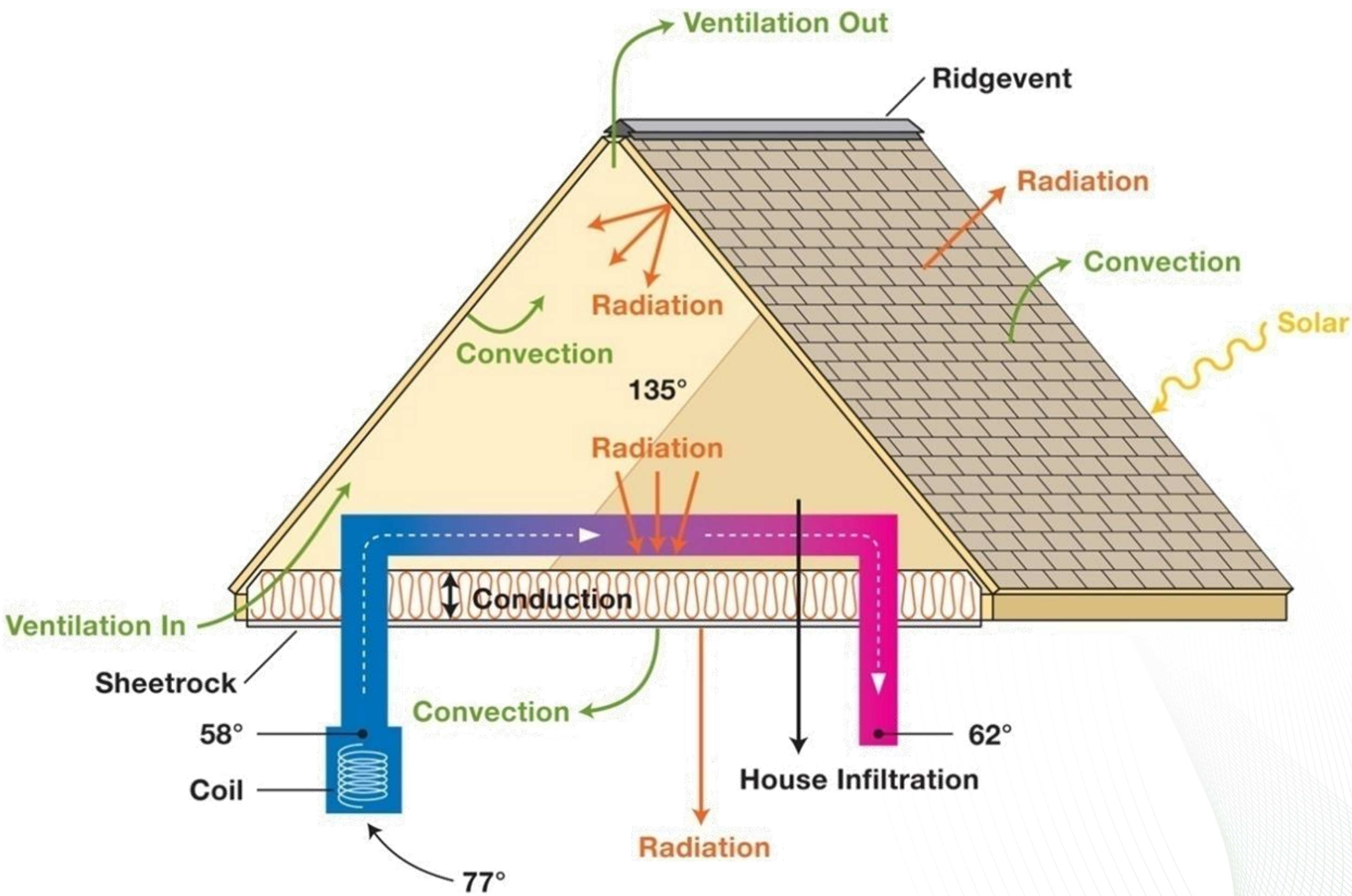
- Replaces:
  - EPA Roof Comparison Calc
  - DOE Cool Roof Calculator
- Minimal questions (<20)
  - Only location is required
  - Building America defaults
  - Help links for unknown information

Calculator Input Comparison Chart

	RSC[1]	PAC Slides[2]	PAC QRpt[3]	EPA[4]	DOE[5]
Building Type	✓	✓	✓	✓	
Location	✓	✓		✓	✓
Days of Operation per week		✓	✓	✓	
Building stock	✓	✓		✓	
Cooling system efficiency (SEER)	✓	✓	✓	✓	✓
Type of heating	✓	✓	✓	✓	✓
Heating system efficiency	✓	✓	✓	✓	✓
Duct location	✓	✓	✓		
Level of roof/ceiling insulation	✓	✓	✓	✓	✓
Above-sheathing ventilation	✓	✓			
Radiant barrier	✓	✓			
Roof thermal mass	✓	✓			
Roof solar reflectance	✓	✓	✓	✓	✓
Roof solar reflectance (black compare)	✓		✓	✓	
Roof thermal emittance	✓	✓	✓		✓
Roof thermal emittance (black compare)	✓		✓		
Internal load		✓			
Conditioned space under roof		✓			
Gas and electricity costs	✓	✓	✓	✓	✓
Inclination / Roof Area	✓			✓	
HVAC Schedule			✓		
Conditioned space (ft^2)	✓			✓	
Number of floors	✓				
Window-to-wall ratio	✓				

RSC = AtticSim + DOE-2.1E

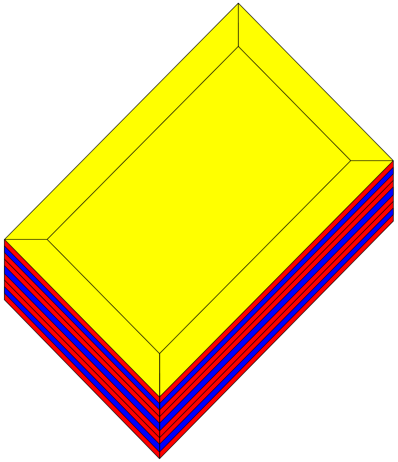
AtticSim - ASTM C 1340 Standard For Estimating Heat Gain or Loss Through Ceilings Under Attics

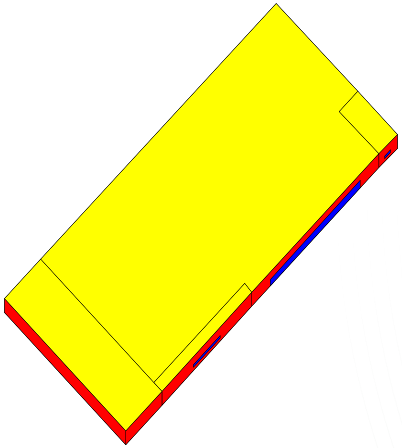


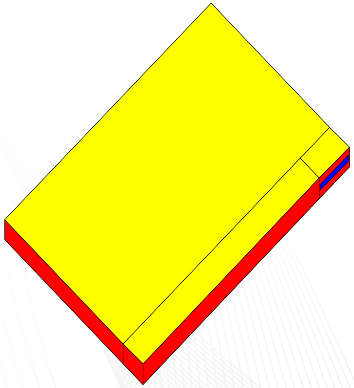
OAK RIDGE National Laboratory
14

Commercial building types

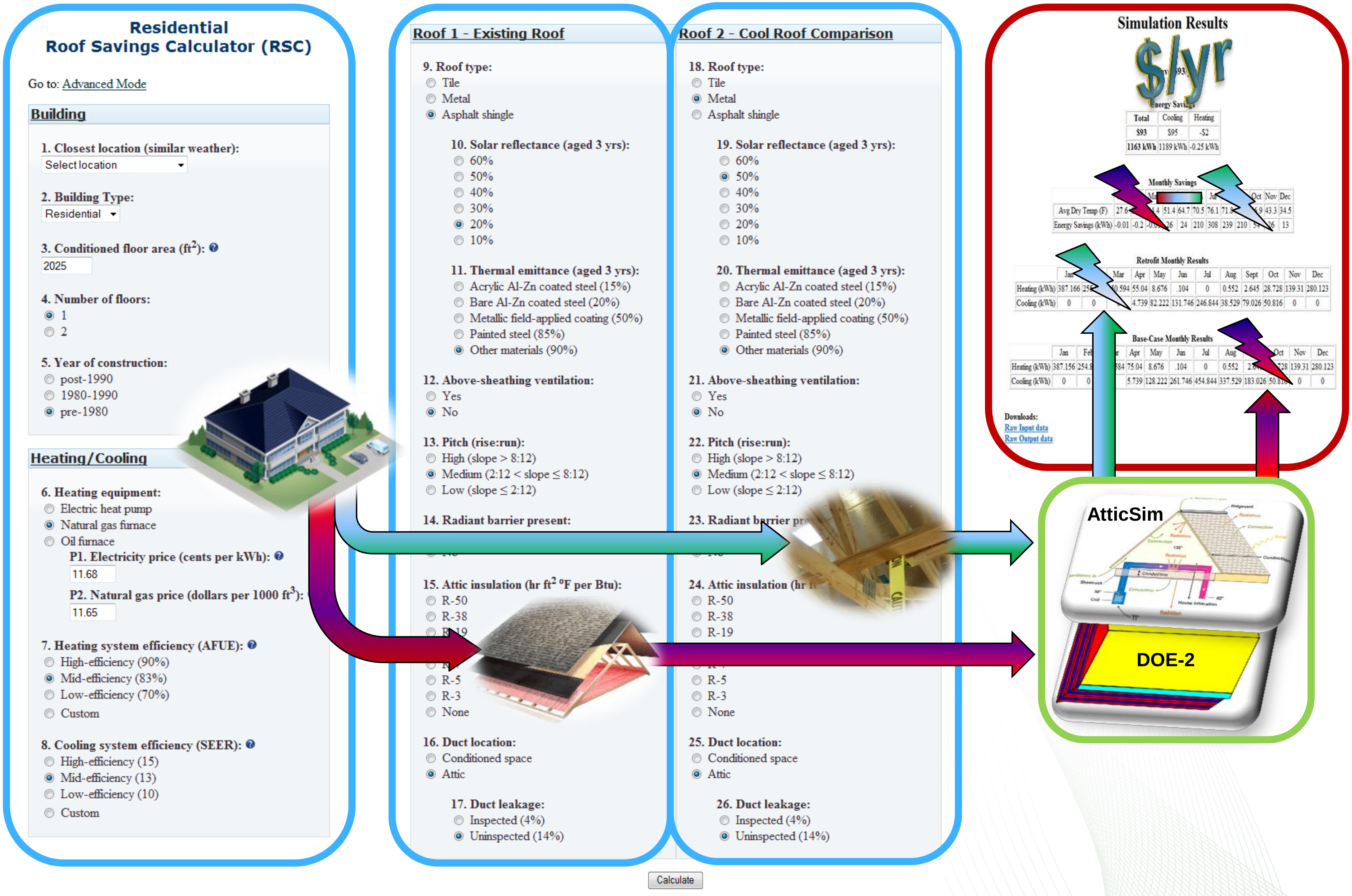
Office	"Big Box" Retail	Warehouse







Torcellini et al. 2008, "DOE Commercial Building Benchmark Models", NREL/CP-550-43291, National Renewable Energy Laboratory, Golden CO.



Residential
Roof Savings Calculator (RSC)
Go to: Advanced Mode
Building
1. Closest location (similar weather):
Select location
2. Building Type:
Residential
3. Conditioned floor area (ft²):
2025
4. Number of floors:
1
2
5. Year of construction:
post-1990
1980-1990
pre-1980
Heating/Cooling
6. Heating equipment:
Electric heat pump
Natural gas furnace
Oil furnace
P1. Electricity price (cents per kWh):
11.68
P2. Natural gas price (dollars per 1000 ft³):
11.65
7. Heating system efficiency (AFUE):
High-efficiency (90%)
Mid-efficiency (83%)
Low-efficiency (70%)
Custom
8. Cooling system efficiency (SEER):
High-efficiency (15)
Mid-efficiency (13)
Low-efficiency (10)
Custom
Roof 1 - Existing Roof
9. Roof type:
Tile
Metal
Asphalt shingle
10. Solar reflectance (aged 3 yrs):
60%
50%
40%
30%
20%
10%
11. Thermal emittance (aged 3 yrs):
Acrylic Al-Zn coated steel (15%)
Bare Al-Zn coated steel (20%)
Metallic field-applied coating (50%)
Painted steel (85%)
Other materials (90%)
12. Above-sheathing ventilation:
Yes
No
13. Pitch (rise:run):
High (slope > 8:12)
Medium (2:12 < slope ≤ 8:12)
Low (slope ≤ 2:12)
14. Radiant barrier present:
15. Attic insulation (hr ft² °F per Btu):
R-50
R-38
R-5
R-3
None
16. Duct location:
Conditioned space
Attic
17. Duct leakage:
Inspected (4%)
Uninspected (14%)
Roof 2 - Cool Roof Comparison
18. Roof type:
Tile
Metal
Asphalt shingle
19. Solar reflectance (aged 3 yrs):
60%
50%
40%
30%
20%
10%
20. Thermal emittance (aged 3 yrs):
Acrylic Al-Zn coated steel (15%)
Bare Al-Zn coated steel (20%)
Metallic field-applied coating (50%)
Painted steel (85%)
Other materials (90%)
21. Above-sheathing ventilation:
Yes
No
22. Pitch (rise:run):
High (slope > 8:12)
Medium (2:12 < slope ≤ 8:12)
Low (slope ≤ 2:12)
23. Radiant barrier pr
24. Attic insulation (hr ft
R-50
R-38
R-19
R-5
R-3
None
25. Duct location:
Conditioned space
Attic
26. Duct leakage:
Inspected (4%)
Uninspected (14%)
Calculate
Simulation Results
$/yr
Monthly Savings
Retrofit Monthly Results
Base-Case Monthly Results
Downloads:
Raw Input data
Raw Output data
AtticSim
DOE-2

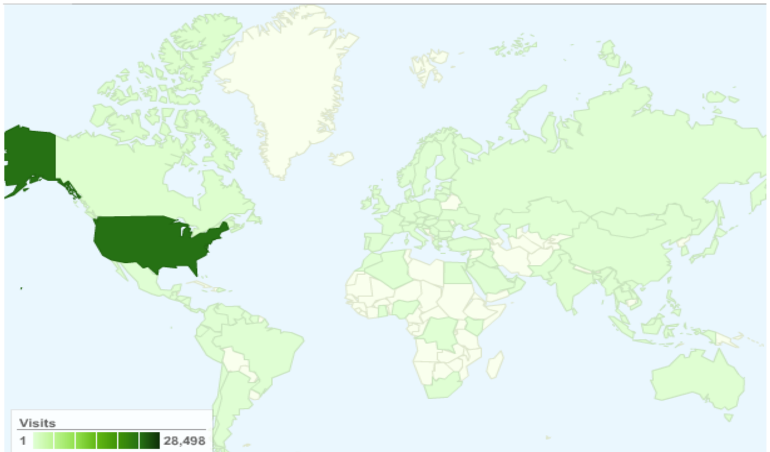
RoofCalc.com impact

Dashboard

Apr 20, 2010 - Feb 28, 2011

100,000+ visitors, 200+ user feedback,

Average: ~81 visitors/day



30,752 visits came from 112 countries/territories

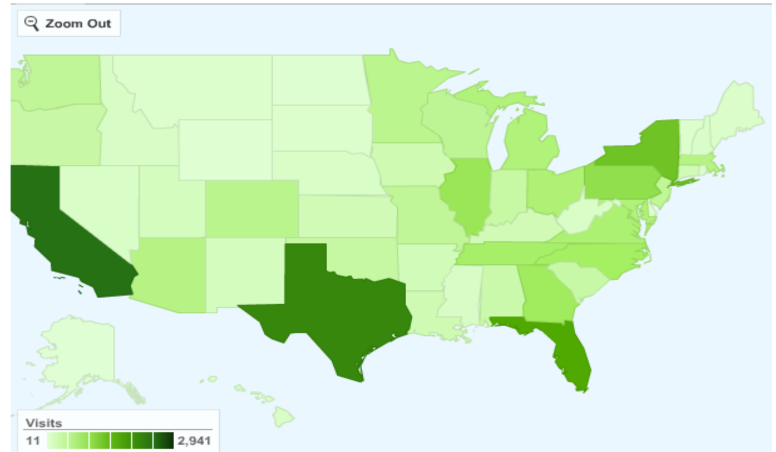
Detail Level: City | Country/Territory | Sub Continent Region | Continent Dimension: None

Site Usage | Goal Set 1 Views:

Visits	Pages/Visit	Avg. Time on Site	% New Visits	Bounce Rate
30,752 % of Site Total: 100.00%	1.42 Site Avg: 1.42 (0.00%)	00:01:25 Site Avg: 00:01:25 (0.00%)	88.26% Site Avg: 88.23% (0.04%)	70.34% Site Avg: 70.34% (0.00%)

	Detail Level: Country/Territory	Visits ↓	Pages/Visit	Avg. Time on Site	% New Visits	Bounce Rate
1.	United States	28,498	1.42	00:01:25	88.35%	70.34%
2.	Canada	483	1.36	00:01:05	91.30%	73.08%
3.	India	156	1.42	00:01:08	80.77%	73.72%
4.	Australia	129	1.66	00:01:42	82.17%	66.67%
5.	United Kingdom	94	1.39	00:01:13	94.68%	65.96%
6.	South Korea	79	1.52	00:01:07	70.89%	68.35%
7.	Italy	66	1.61	00:01:33	89.39%	63.64%

Zoom Out



This country/territory sent 28,498 visits via 52 regions

Detail Level: City | Region Dimension: None

Site Usage | Goal Set 1 Views:

Visits	Pages/Visit	Avg. Time on Site	% New Visits	Bounce Rate
28,498 % of Site Total: 92.67%	1.42 Site Avg: 1.42 (-0.09%)	00:01:25 Site Avg: 00:01:25 (0.96%)	88.35% Site Avg: 88.23% (0.14%)	70.34% Site Avg: 70.34% (-0.00%)

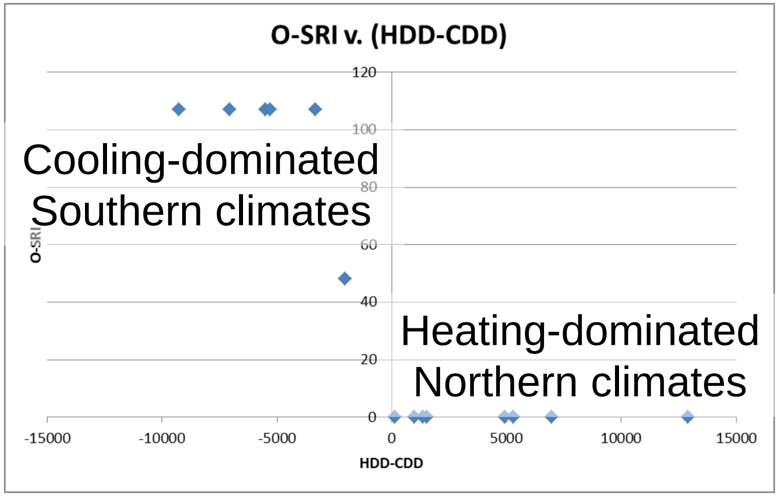
	Detail Level: Region	Visits ↓	Pages/Visit	Avg. Time on Site	% New Visits	Bounce Rate
1.	California	2,941	1.37	00:01:21	82.66%	73.95%
2.	Texas	2,558	1.43	00:01:26	90.30%	68.22%
3.	Florida	1,965	1.47	00:01:43	89.52%	68.09%
4.	New York	1,608	1.35	00:01:09	91.42%	73.45%
5.	Pennsylvania	1,206	1.39	00:01:20	91.04%	71.72%
6.	Illinois	1,114	1.36	00:01:12	89.41%	73.79%
7.	Georgia	1,032	1.40	00:01:18	90.50%	69.09%

OAK RIDGE National Laboratory

Nationwide results

Cost savings for offices - 14 cities, local utility prices, 22 roof types

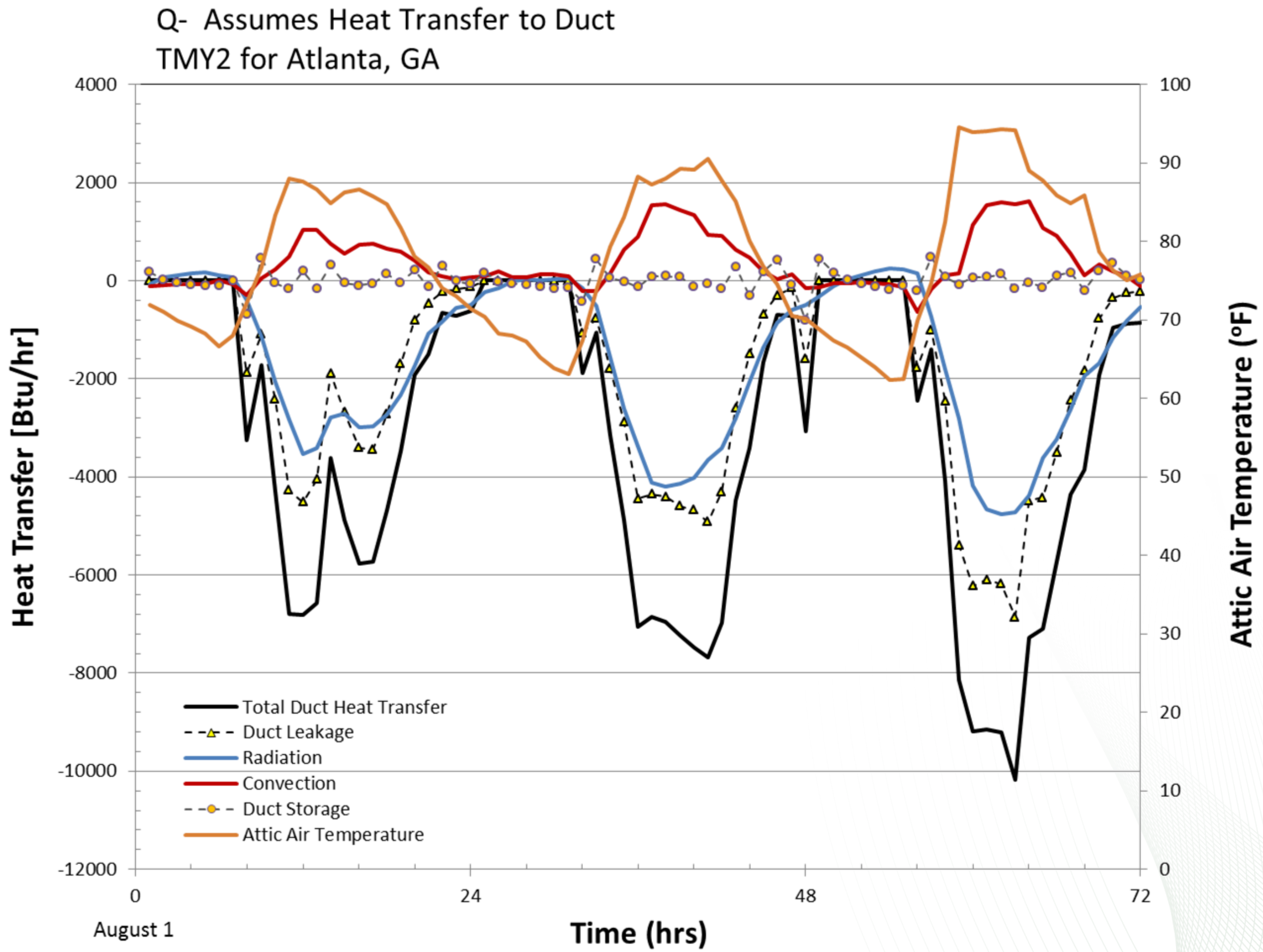
Description	Reflectance	Emissivity	SRI	Houston $ saved	…13
BUR No Coating	10	90	6	42	
Mineral Mod Bit	25	88	25	103	
Single Ply	32	90	35	230	
Mineral Mod Bit	33	92	35	197	
Metal	35	82	35	60	
Aluminum Coating	43	58	35	279	
Mineral Mod Bit	45	79	55	291	
Coating over BUR	49	83	55	433	
Metal	49	83	55	208	
…14					



Location	Trend Desired SRI	Maximum Observed Savings, $	Best Observed System	Related SRI	Slope Difference
Atlanta	107	1080	Aluminum Coating over BUR	65	Reversed
Austin	107	2680	Coating over BUR (White)	107	Same
Baltimore	107	1000	Single Ply /Coating over BUR	103.5	Reversed
Chicago	64.95	360	Aluminum Coating over BUR	48	Same
Fairbanks	42.68	680	Aluminum Coating over BUR	48	Same
Fargo	40.58	160	Aluminum Coating over BUR	48	Same
Houston	107	1840	Coating over BUR (White)	107	Same
Kansas City	107	800	Coating over BUR (White)	107	Reversed
Los Angeles	107	440	Aluminum Coating over BUR	65	Same
Miami	107	4440	Coating over BUR (White)	107	Same
Minneapolis	47.05	360	Aluminum Coating over BUR	48	Same
New York	107	560	Aluminum Coating over BUR	65	Reversed
Phoenix	107	3000	Coating over BUR (White)	107	Same
San Francisco	39.31	200	Aluminum Coating over BUR	48	Same

Mellot, Joseph W., New, Joshua R., and Sanyal, Jibonananda. (2013). "Preliminary Analysis of Energy Consumption for Cool Roofing Measures." In *RCI Interface Technical Journal*, volume 31, issue 9, pp. 25-36, October, 2013.

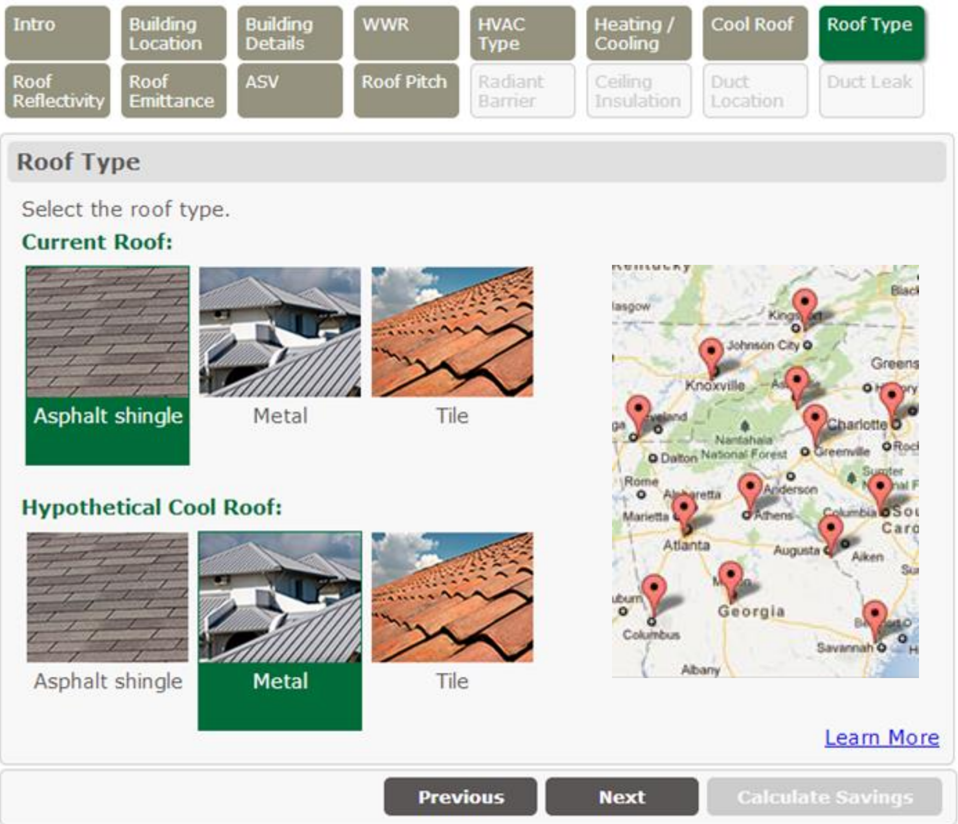
Summer operation of HVAC duct in ASHRAE climate zone 3



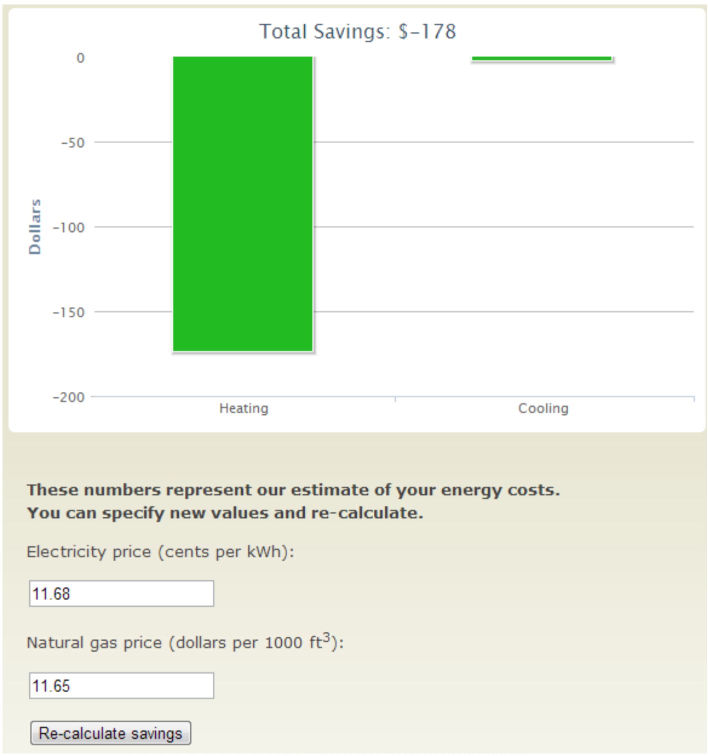
OAK RIDGE National Laboratory

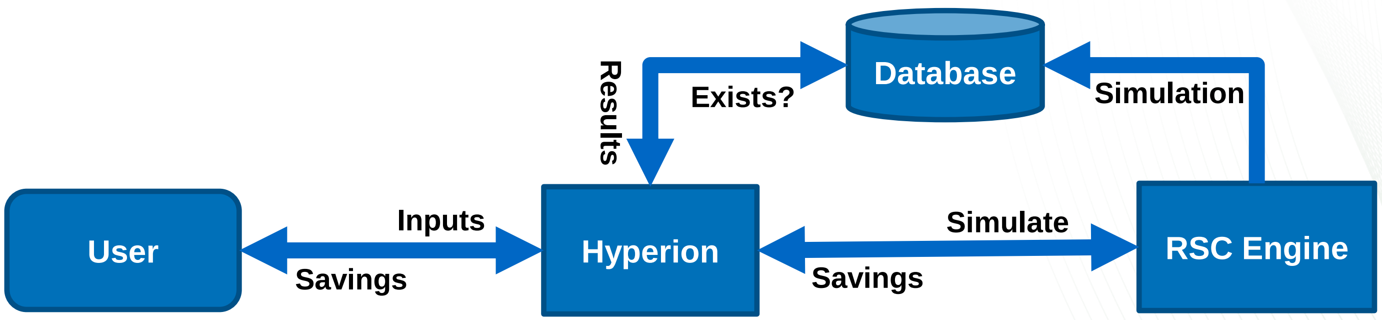
Enhanced RSC Site

Input Parameter GUI



Result Output





Oak Ridge National Laboratory

Quote

"We speak piously of … making small studies that will add another brick to the temple of science. Most such bricks just lie around the brickyard."

–J.R. Platt, Science 1964, 146:347-53

21

RSC Service Example (Python)

```
client = suds.client.Client('URL/TO/WEB/SERVICE/rsc.wsdl')
print(client)

sm = client.factory.create('schema:soapmodel')
load_soap_model_from_xml('../examplemodel.xml', sm)
sr = client.service.simulate(sm)
print(sr)

sm = client.factory.create('schema:soapmodel')
load_soap_model_from_xml('../examplemodel.xml', sm)
print(sm)
contents = client.service.test(sm)
with open('pytest.zip', 'wb') as outfile:
    outfile.write(base64.b64decode(contents))
```

…download example building and batch script from rsc.ornl.gov/web-service.shtml

22

Update 1 line of code to change servers

```python
import base64
import suds
import xml.dom.minidom
import logging


def load_soap_model_from_xml(xmlfilename, soapmodel):

def load_soap_results_from_xml(xmlfilename, soapresults):


logging.basicConfig()

test_type = ['simulate', 'test', 'upload', 'download']

print ("hello there, initializing client")
client = suds.client.Client('http://evenstar.ornl.gov/RSC/service/rsc.wsdl')
print ("printing client")
print(client)
raw_input('Press Enter to continue...'+'\n')
```


Oak Ridge National Laboratory

Millions of simulations visualized for DOE's Roof Savings Calculator and deployment of roof and attic technologies through leading industry partners

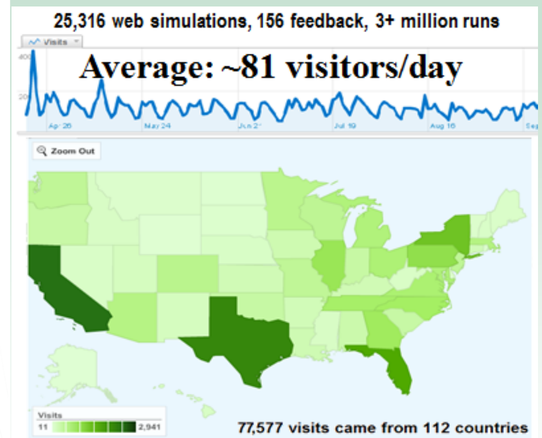
DOE: Office of Science | **CEC & DOE EERE: BTO** | **Industry & Building Owners**

Engine (AtticSim/DOE-2) debugged using HPC Science assets enabling visual analytics on $3x(10)^6$ simulations

Roof Savings Calculator (RSC) web site/service developed and validated [estimates energy and cost savings from roof and attic technologies]

CENTIMARK
Innovative Roofing and Flooring Solutions

CentiMark, the largest nation-wide roofing contractor (installs 2500 roofs/mo), is integrating RSC into their proposal generating system (20+ companies now interested)



Leveraging HPC resources to facilitate deployment of building energy efficiency technologies

OAK RIDGE National Laboratory

Personal story behind one of DOE's RSC images

OAK RIDGE
National Laboratory

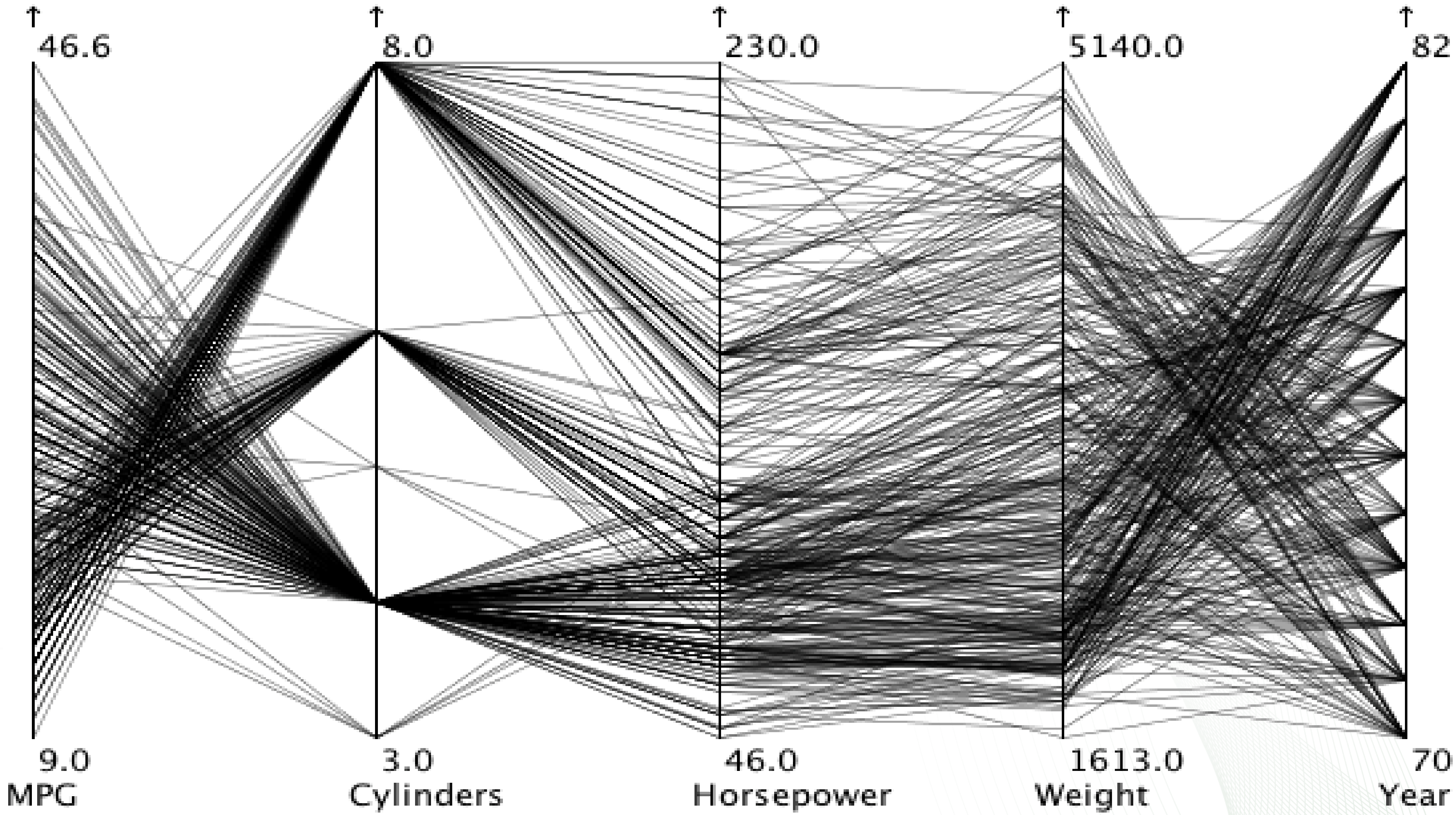
25

Presentation summary

- Scientific Paradigms
- Roof Savings Calculator
- Visual Analytics
- Knowledge Work
- Autotune

26

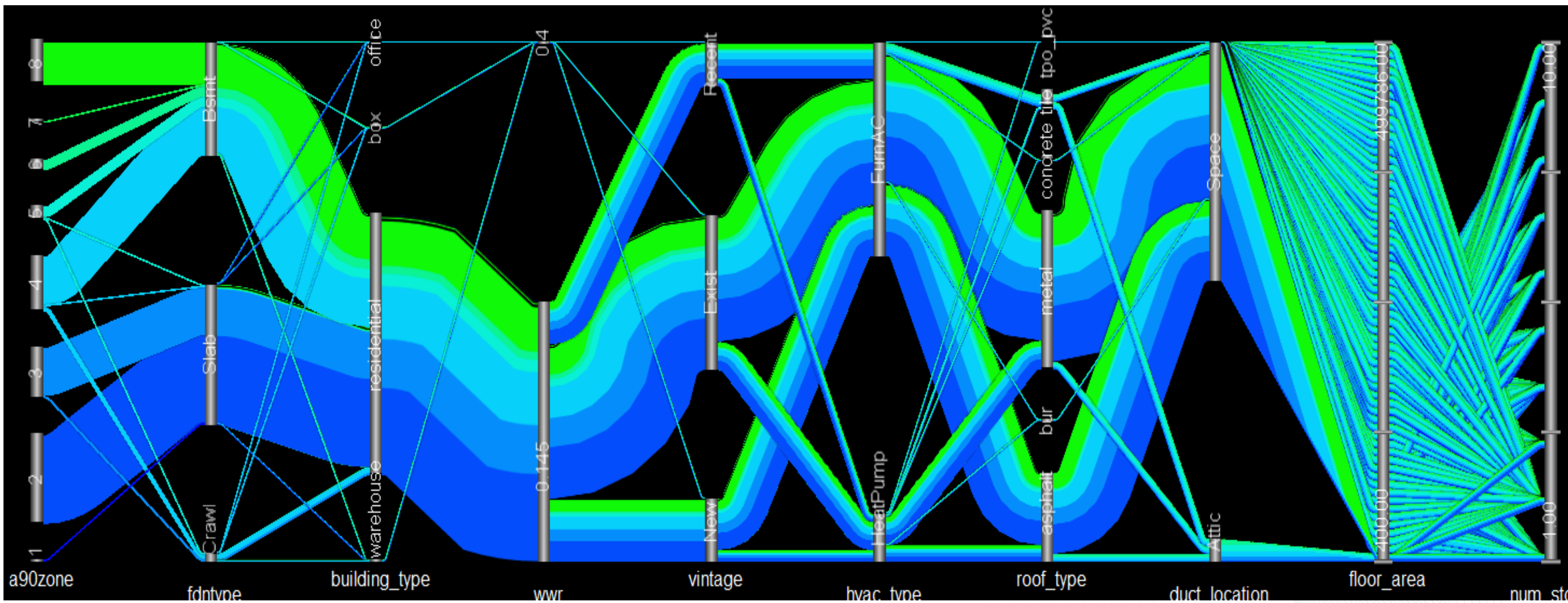
PCP - car data set



Oak Ridge National Laboratory

PCP bin rendering (data)

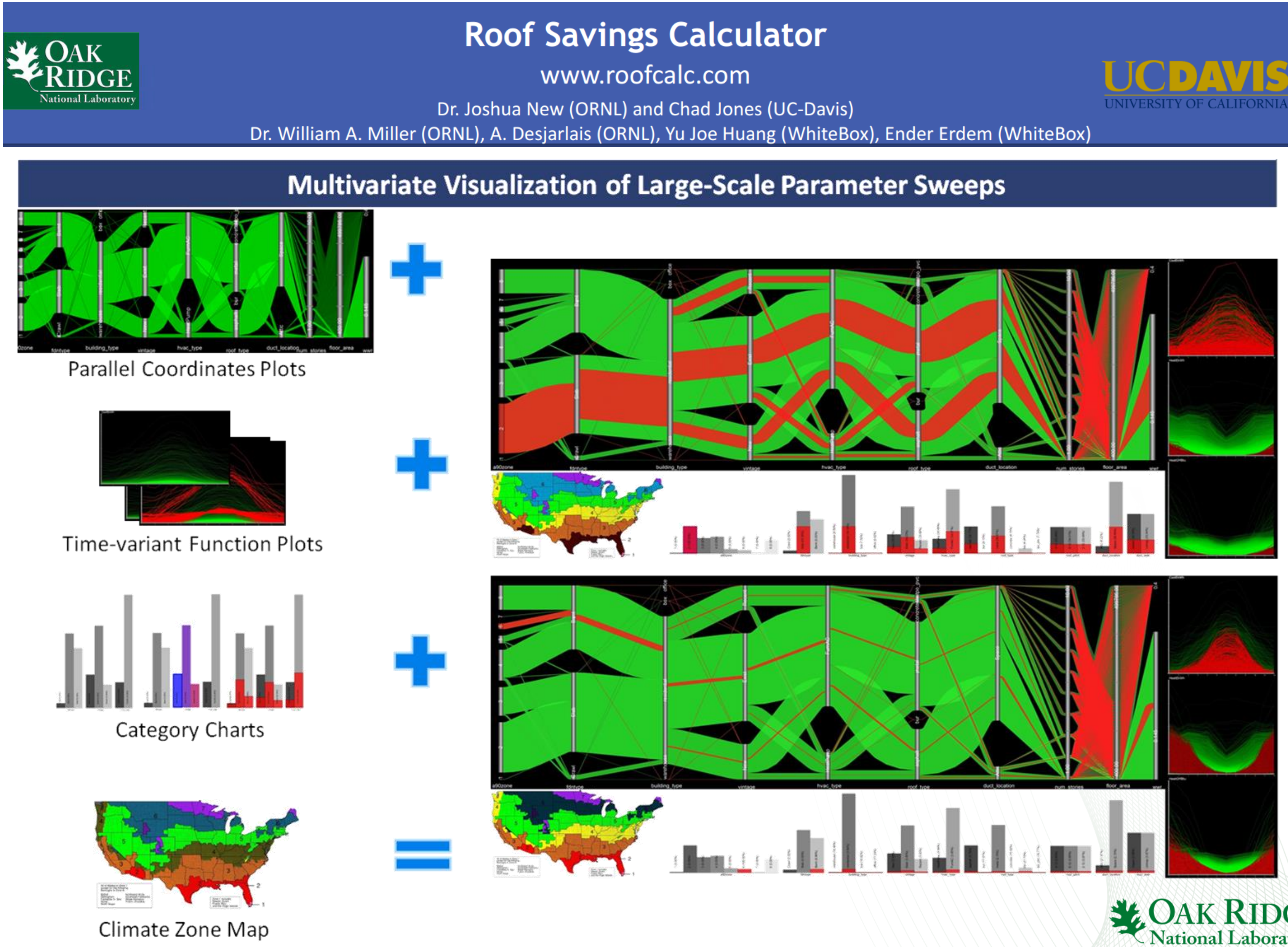
- Transfer function coloring:
  - Occupancy or leading axis



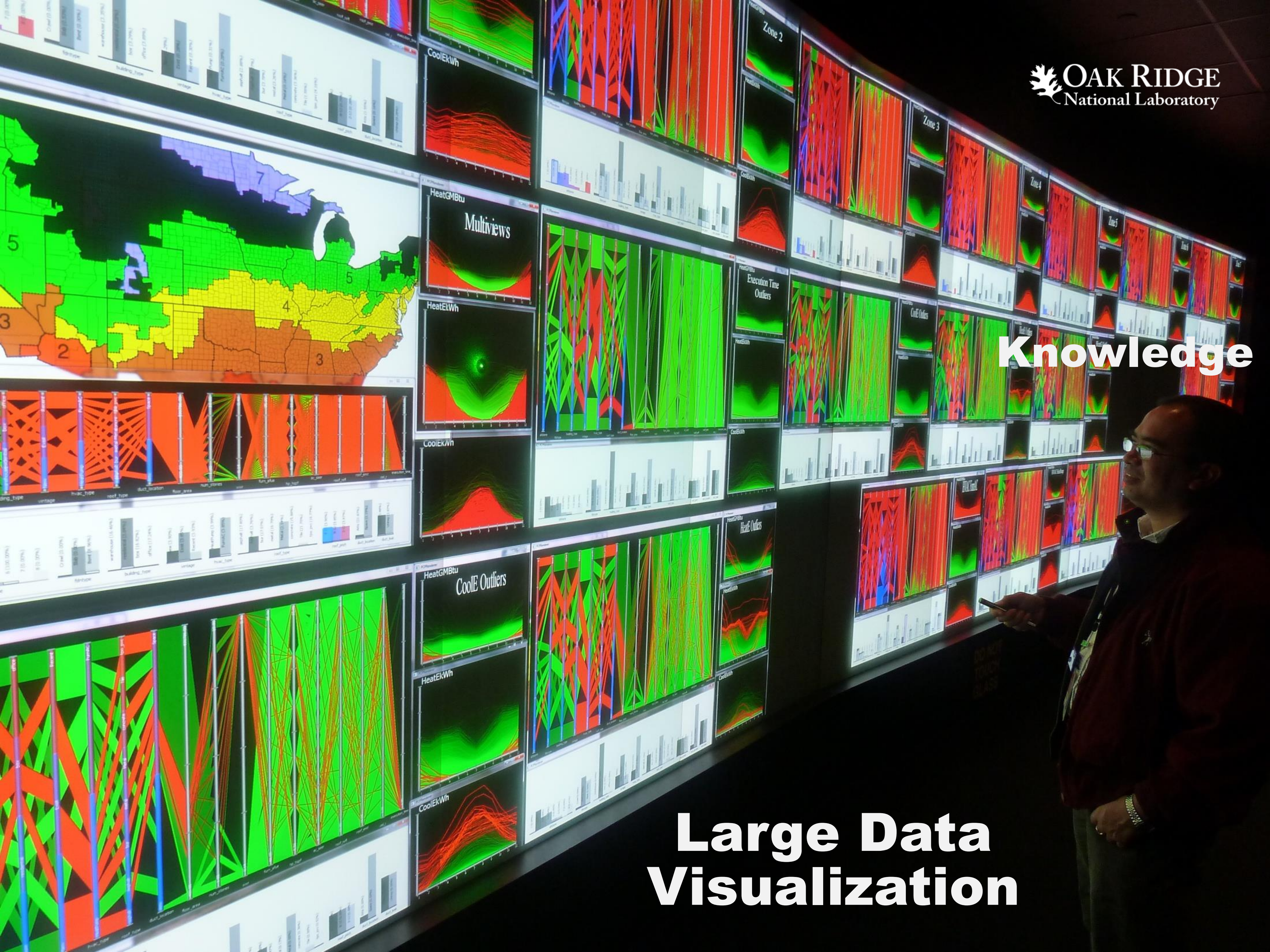
Oak Ridge National Laboratory

28

The power of "and" – linked views (info)



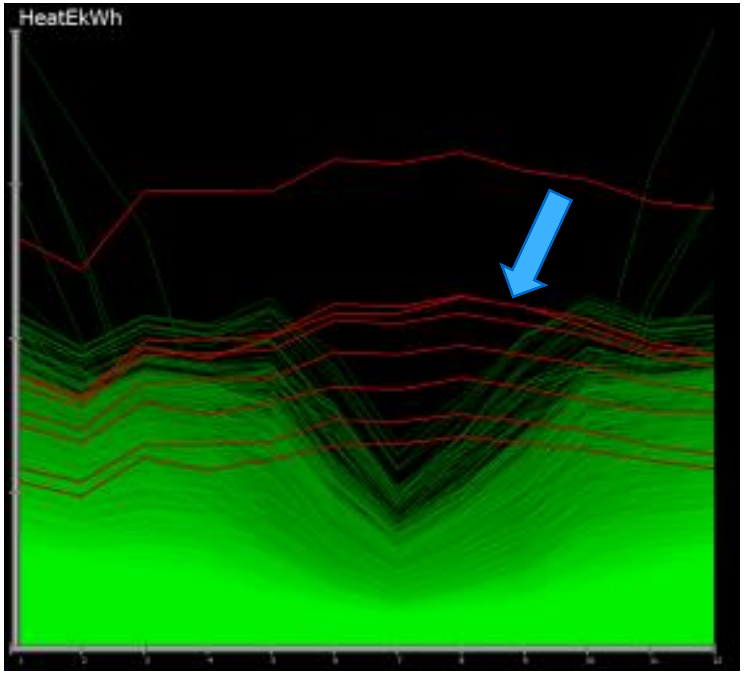
Oak Ridge National Laboratory

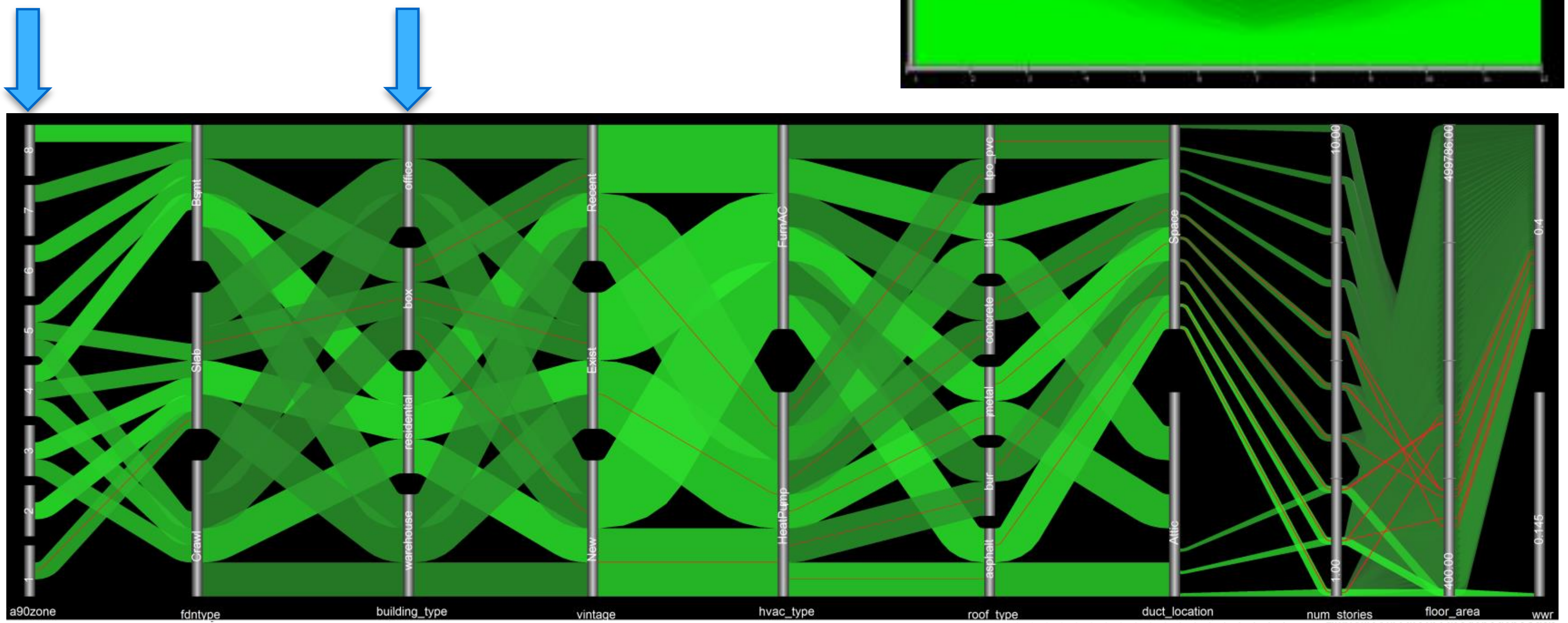


Large Data Visualization

Outliers (wisdom)

- Selection of heating outliers
- Find all have box building type and in Miami





OAK RIDGE National Laboratory

Impact – RSC and Visual Analytics

12 Publications, 20+ organizations interested in licensing

New, Joshua R., Huang, Yu (Joe), Levinson, Ronnen, Mellot, Joe, Sanyal, Jibonananda, Miller, William A., and Childs, Kenneth W. (2013). "Analysis of DOE's Roof Savings Calculator with Comparison to other Simulation Engines" ORNL internal report ORNL/TM-2013/501, November 1, 2013, 63 pages.

Mellot, Joseph W., Sanyal, Jibonananda, and New, Joshua R. (2013). "Preliminary Analysis of Energy Consumption for Cool Roofing Measures." Presented at the International Reflective Roofing Symposium, the American Coating Association's (ACA) conference, and in *Proceedings of the ACA's Coating Regulations and Analytical Methods Conference*, Pittsburgh, PA, May 14-15, 2013.

Jones, Chad, New, Joshua R., Sanyal, Jibonananda, and Ma, Kwan-Liu (2012). "Visual Analytics for Roof Savings Calculator Ensembles." In *Proceedings of the 2nd Energy Informatics Conference*, Atlanta, GA, Oct. 6, 2012.

Cheng, Mengdawn, Miller, William (Bill), New, Joshua R., and Berdahl, Paul (2011). "Understanding the Long-Term Effects of Environmental Exposure on Roof Reflectance in California." In *Journal of Construction and Building Materials*, volume 26, issue 1, pp. 516-26, August 2011.

New, Joshua R., Miller, William (Bill), Desjarlais, A., Huang, Yu Joe, and Erdem, E. (2011). "Development of a Roof Savings Calculator." In *Proceedings of the RCI 26th International Convention and Trade Show*, Reno, NV, April 2011.

Miller, William A., New, Joshua R., Desjarlais, Andre O., Huang, Yu (Joe), Erdem, Ender, and Levinson, Ronnen (2010). "Task 2.5.4 - Development of an Energy Savings Calculator." California Energy Commissions (CEC) PIER Project, ORNL internal report ORNL/TM-2010/111, March 2010, 32 pages.

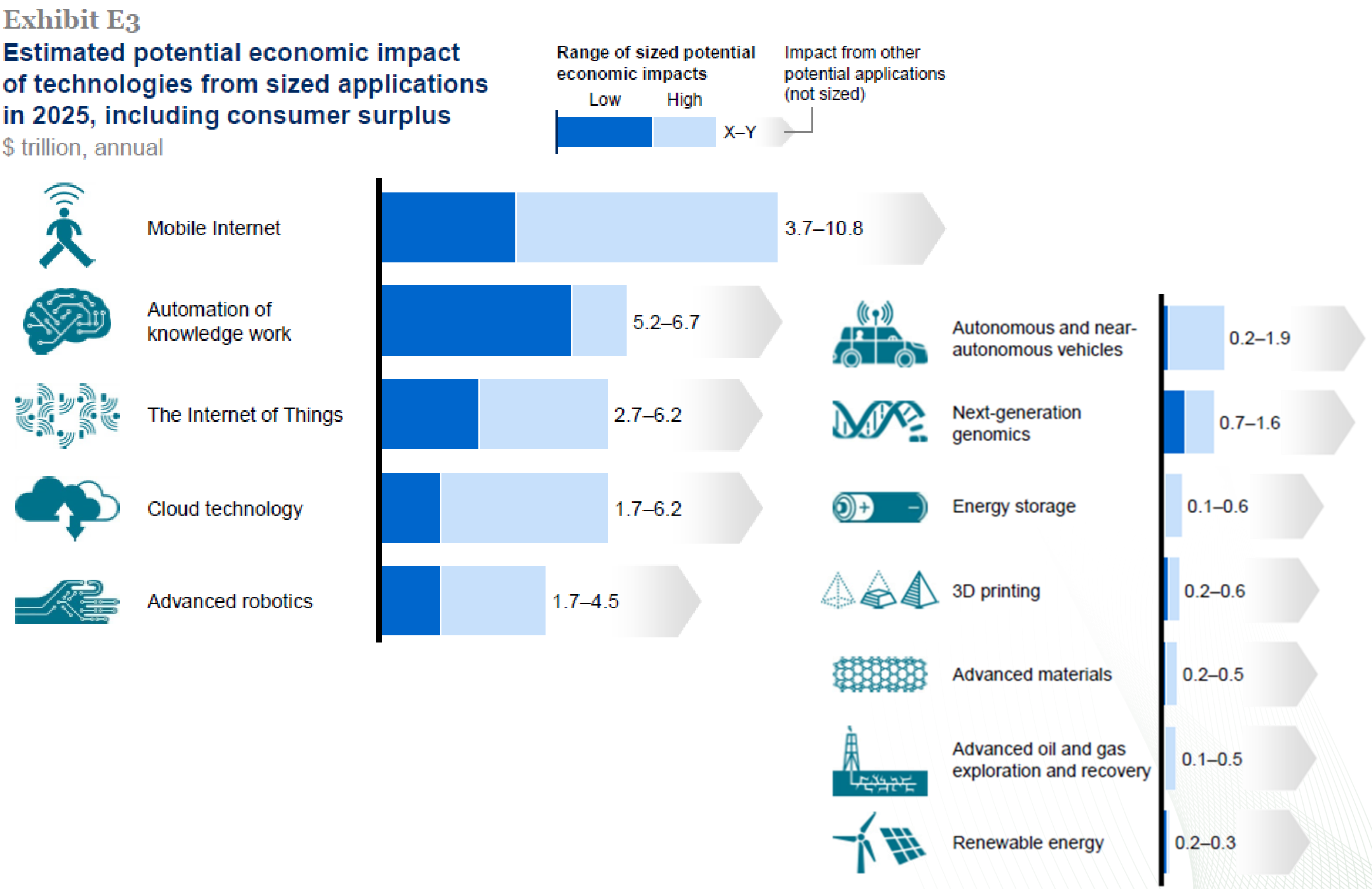
OAK RIDGE National Laboratory

Presentation summary

- Scientific Paradigms
- Roof Savings Calculator
- Visual Analytics
- Knowledge Work (context)
- Autotune

33

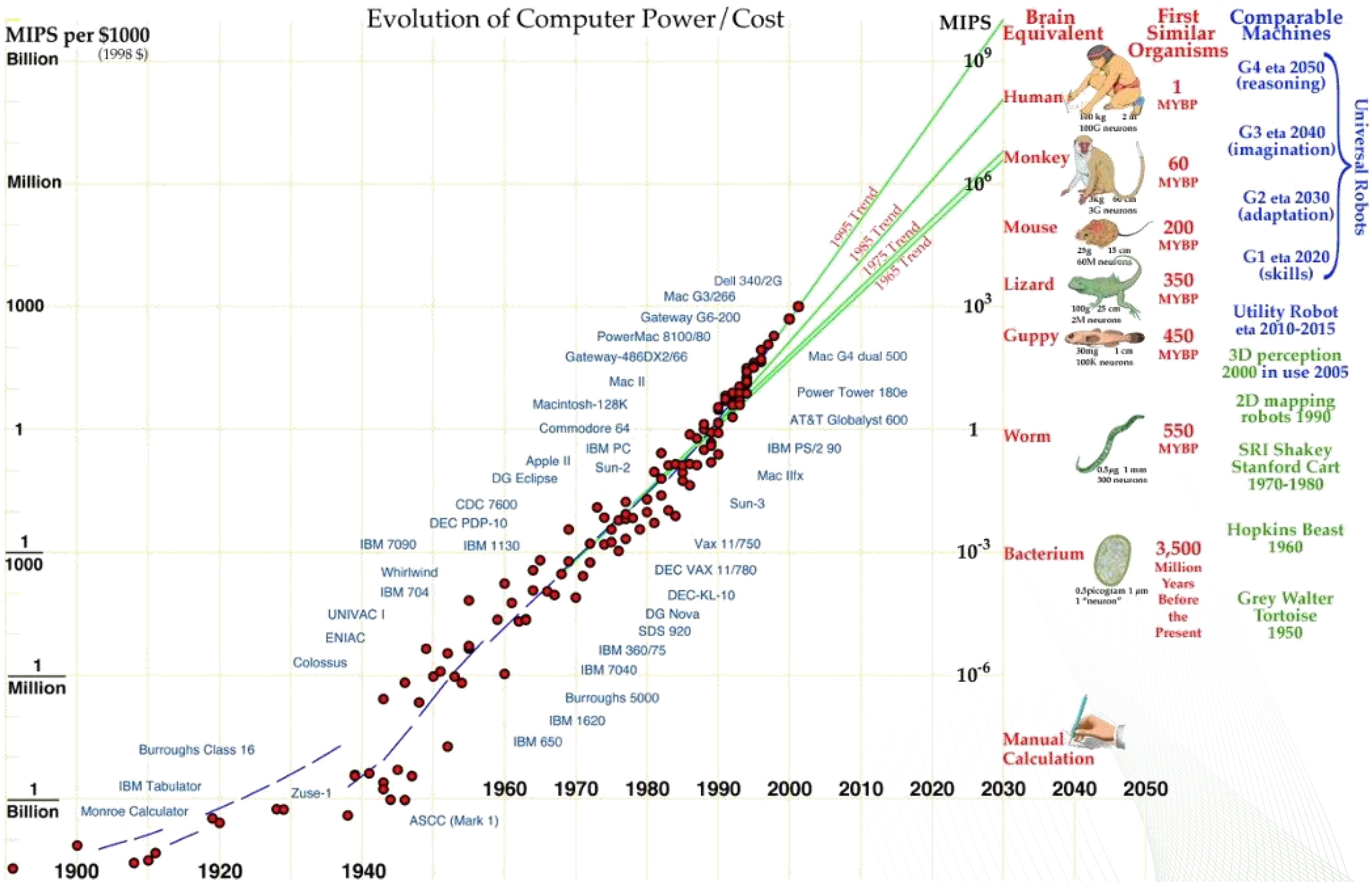
McKinsey Global Institute Analysis



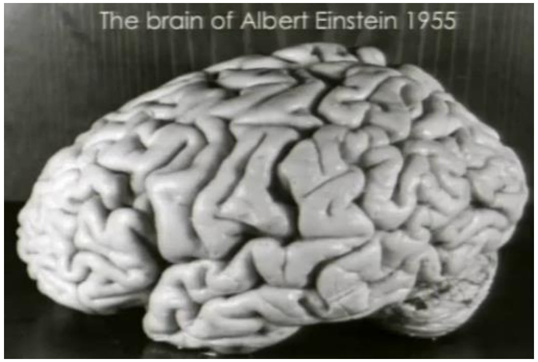
Source: McKinsey Global Institute analysis

OAK RIDGE National Laboratory

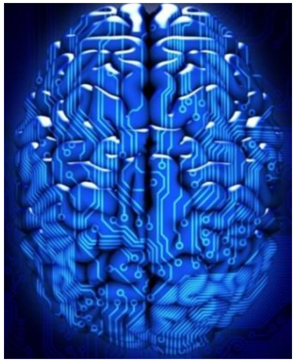
$1000 machine helping meat machines



Humans and computers



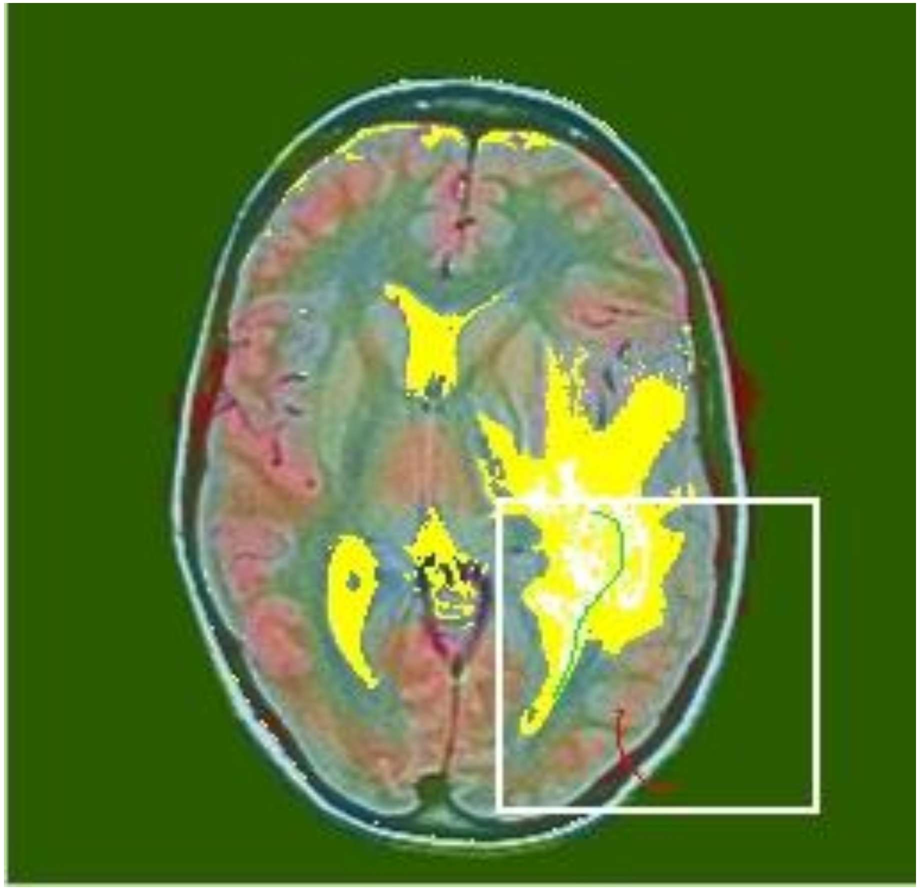
- 3 lbs (2%), 20 watts (20%)
- 120-150 billion neurons
- 100 trillion synapses
  - Firing time ~milliseconds
- 11 million bits/second input
  - Consciousness - 40 bits/second
- Working memory – 4-9 words
- Long-term memory – 1-1k TB
- Complex, self-organizing



- PC – 40 lbs, 500 watts
- 4 cores
- 3 billion Hz
  - Firing time ~nanoseconds
- 100 million bits/second
  - Not yet
- 62,500,000 words
- Disk – 3TB, perfect recall
- "Dumb", Artificial Intel.

36

Learning associations



Full Results

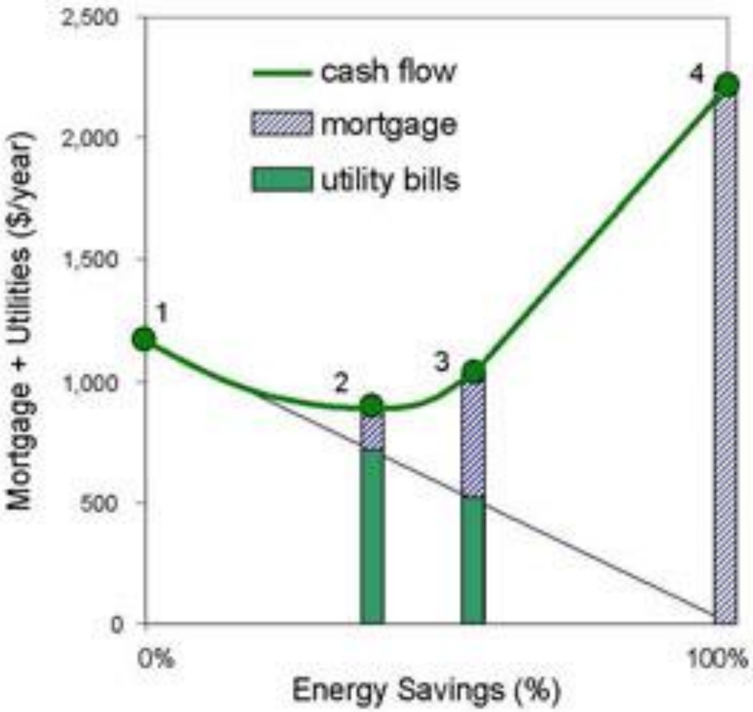
Detailed Results

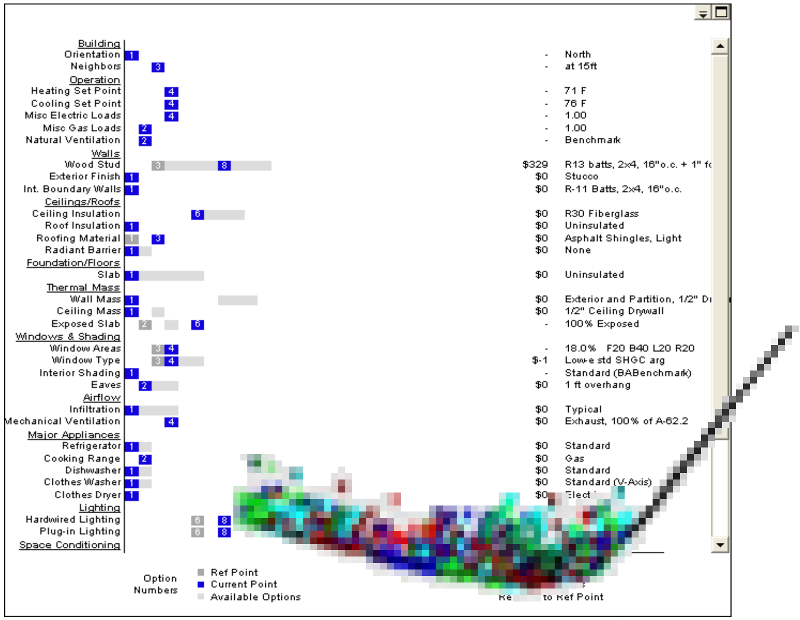
Oak Ridge National Laboratory

Presentation summary

- Scientific Paradigms
- Roof Savings Calculator
- Visual Analytics
- Knowledge Work (context)
- Autotune

Existing tools for retrofit optimization





Simulation Engine
DOE–$65M (1995–?)

EnergyPlus

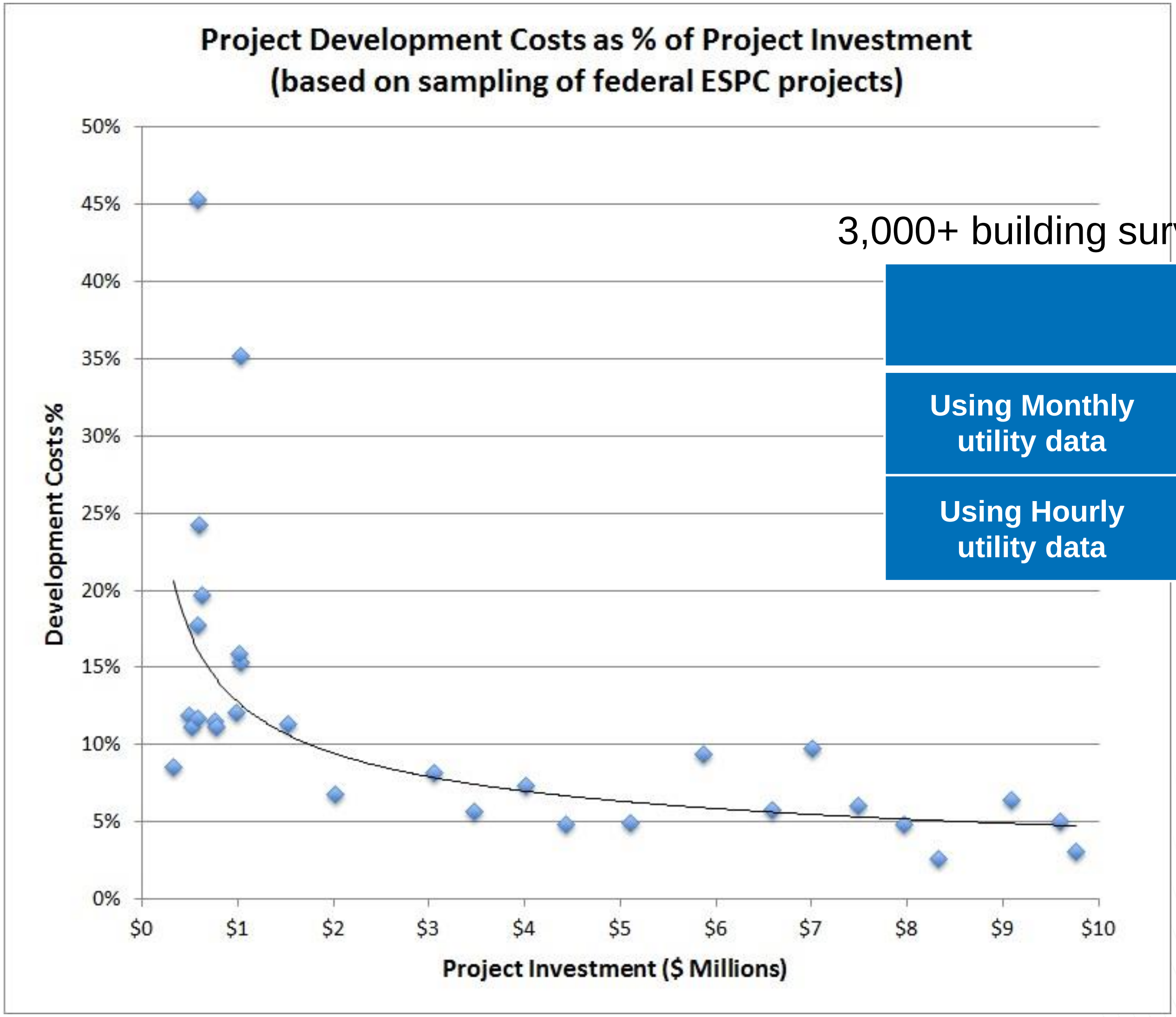
API

OpenStudio

OAK RIDGE National Laboratory

39

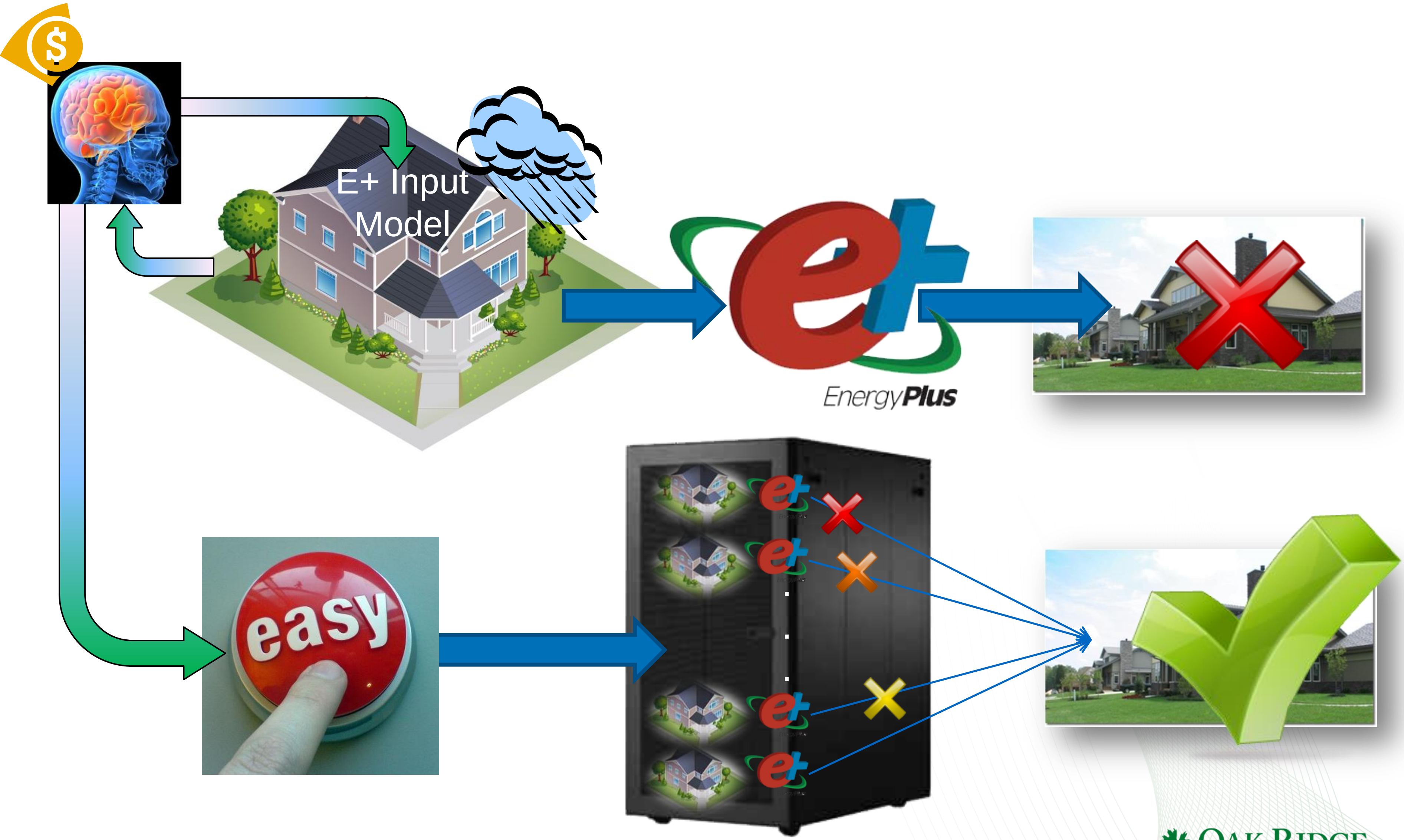
Business limitations for M&V



3,000+ building survey, 23-97% monthly error

		ASHRAE G14 Requires
Using Monthly utility data	CV(RMSE)	15%
	NMBE	5%
Using Hourly utility data	CV(RMSE)	30%
	NMBE	10%

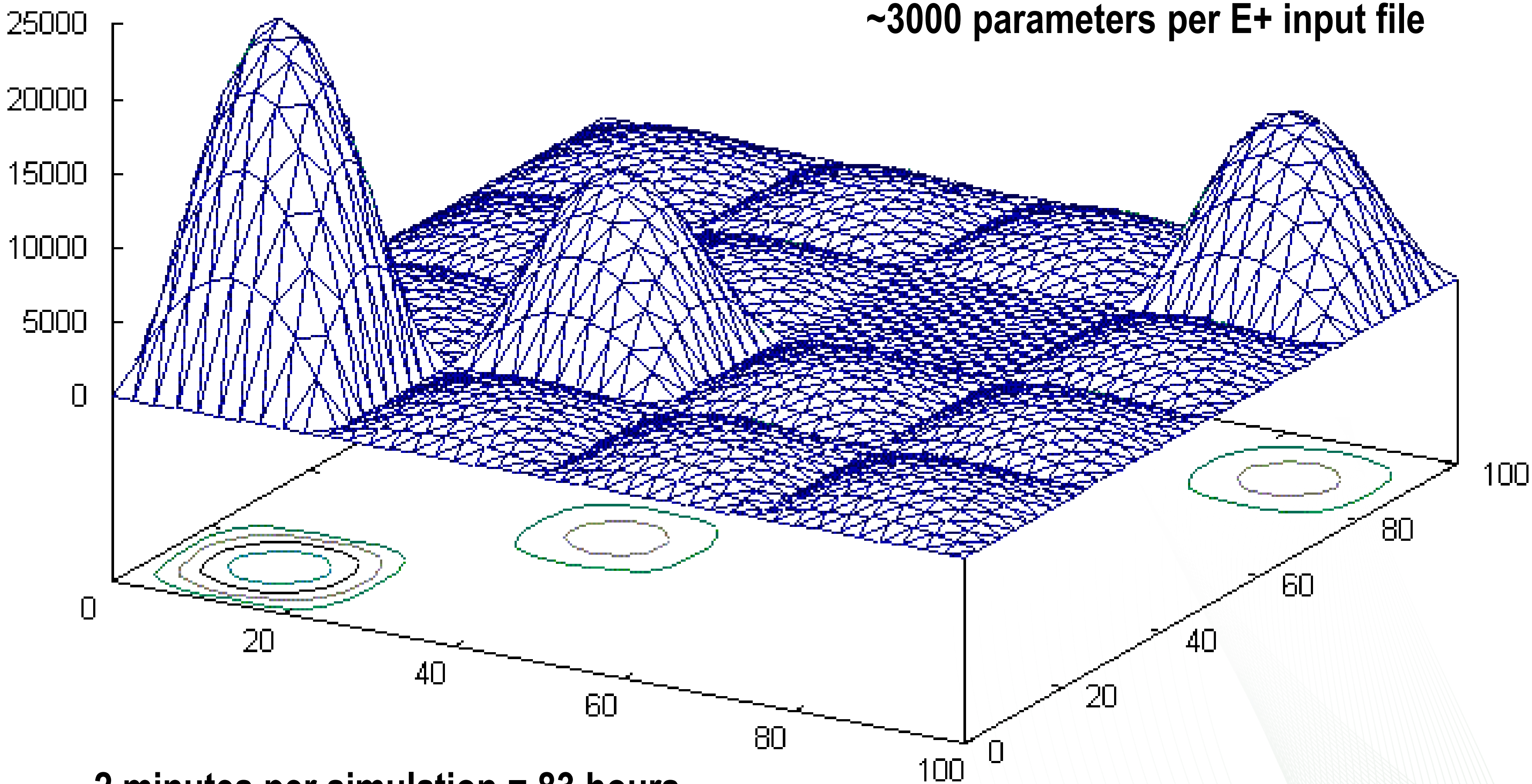
The Autotune Idea
Automatic calibration of software to data



The search problem

Problem/Opportunity:

~3000 parameters per E+ input file



2 minutes per simulation = 83 hours

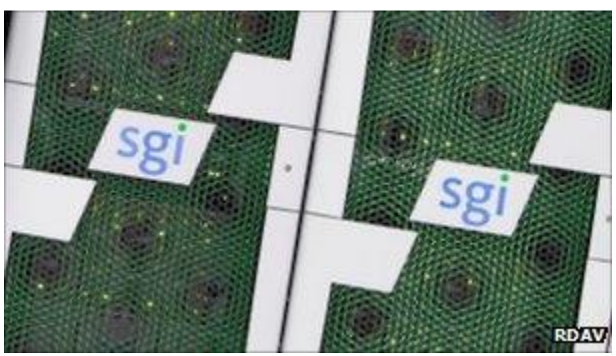
OAK RIDGE
National Laboratory

ORNL High Performance Computing Resources

Titan:
299,008 CPU cores
18,688 GPU cores
710TB memory, distributed

Jaguar:
224,256 cores
360TB memory

Nautilus:
1024 cores
4TB shared-memory



Kraken:
112,896 cores

Gordon:
12,608 cores
SSD

OAK RIDGE
National Laboratory

43

HPC scalability for desktop software

- EnergyPlus desktop app
- Writes files during a run
- Uses RAMdisk
- Balances simulation memory vs. result storage
- Works from directory of input files & verifies result
- Bulk writes results to disk

8 Million EnergyPlus Simulations

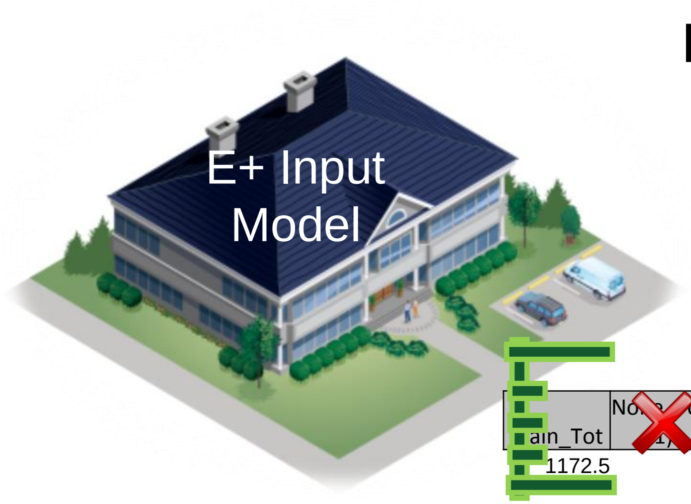
270TB dataset of energy data for buildings

No of Processors	E+ Tasks	Wall-clock Time (mm:ss)
64	256	18:34
128	512	18:22
256	1024	20:30
512	2048	20:43
1024	4096	21:03
2048	8192	21:11
4096	16384	20:00
8192	32768	26:14
16384	65536	26:11
65536	262144	44:52
131072	524288	45TB 68:08

Scalability on Titan

Acknowledgment: Jibo Sanyal, ORNL R&D Staff

Computational complexity

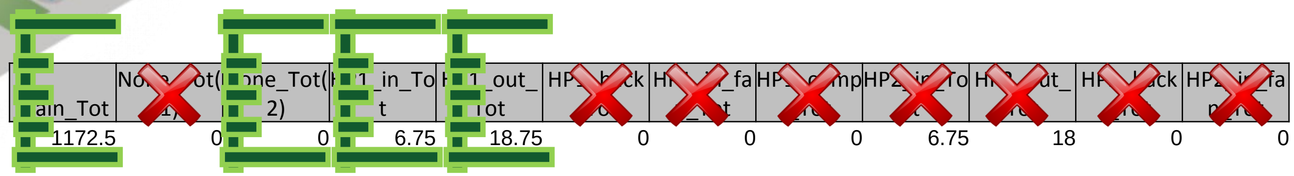


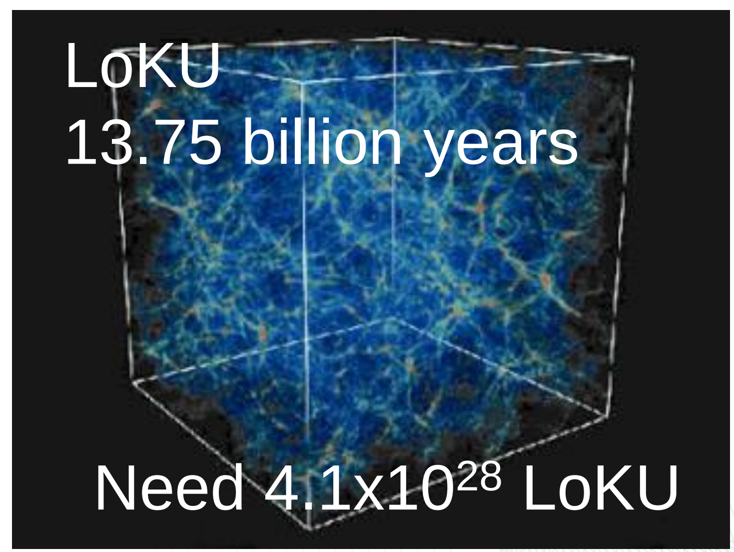
Problems/Opportunities:

Domain experts chose to vary 156

Brute-force = $5x10^{52}$ simulations

E+ parameters





No database technology sufficient?

Relational, Columnar, NoSQL, different compression and partitioning

MyISAM

- No ACID
- No foreign keys
- Speed at scale
  - 0.71 s on LOAD DATA
- Better compression
  - 10.27 MB
  - Read only compression: 6.003 MB
- 2^{32} rows maximum

InnoDB

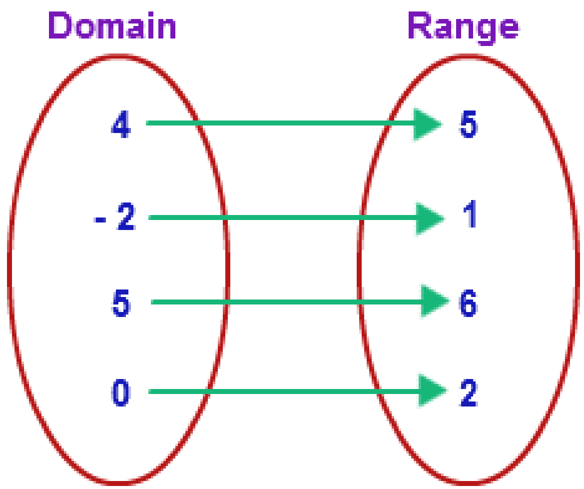
- ACID compliant
- Foreign keys
- Slower adding data
  - 2.3 s on LOAD DATA
- Poorer compression
  - 15.4 MB

Comparisons are based on inserting 200 csv output files, which is 7,008,000 records.

MS Azure DB almost hosted for free with Oakwood Systems, $512,237/month!

What is artificial intelligence?

- Give it (lots of) data
- It maps one set of data to another
- Paradigms
  - Unsupervised (clustering)
  - Reinforcement (don’t run into wall)
  - Supervised (this is the real answer)
- Methods for doing that… biologically motivated or not



act human	act rational
think human	think rational

OAK RIDGE National Laboratory
47

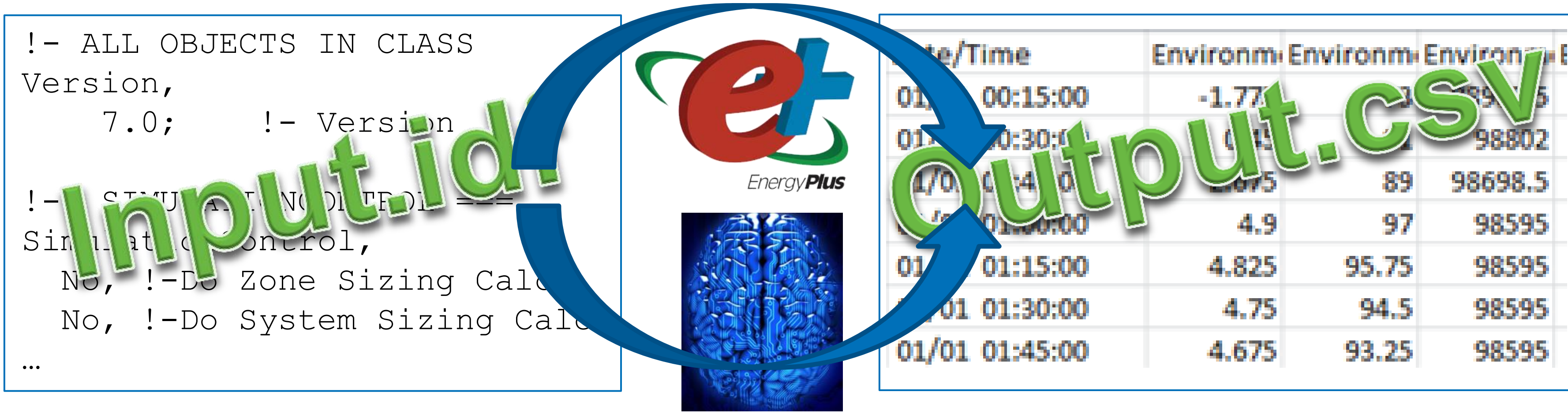
MLSuite: HPC-enabled suite of machine learning algorithms

- Linear Regression
- Feedforward Neural Network
- Support Vector Machine Regression
- Non-Linear Regression
- K-Means with Local Models
- Gaussian Mixture Model with Local Models
- Self-Organizing Map with Local Models
- Regression Tree (using Information Gain)
- Time Modeling with Local Models
- Recurrent Neural Networks
- Genetic Algorithms
- Ensemble Learning

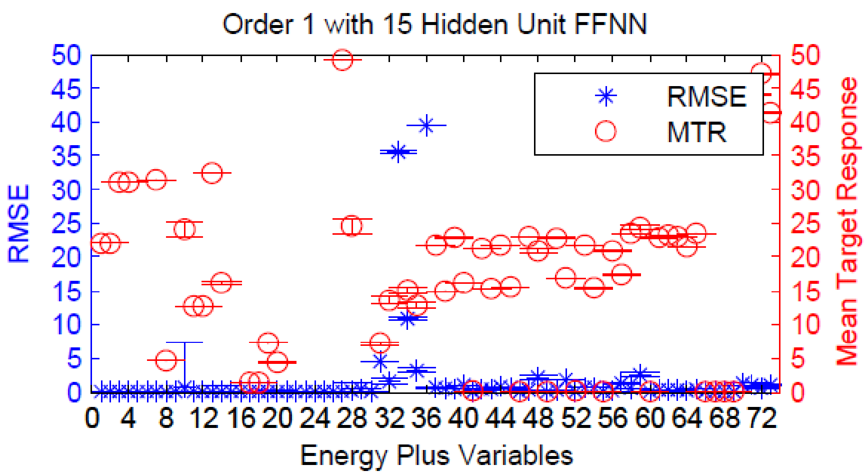
***Acknowledgment:* UTK computer science graduate graduate Richard Edwards, Ph.D. (advisor Dr. Lynne Parker); now Amazon**

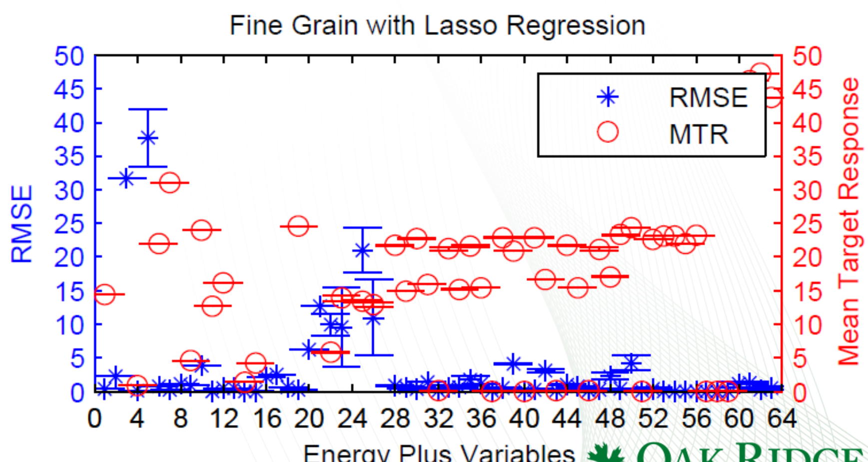
MLSuite example

- EnergyPlus – 2-10 mins for an annual simulation



- ~E+ - 4 seconds AI agent as surrogate model,
  90x speedup, small error, brittle





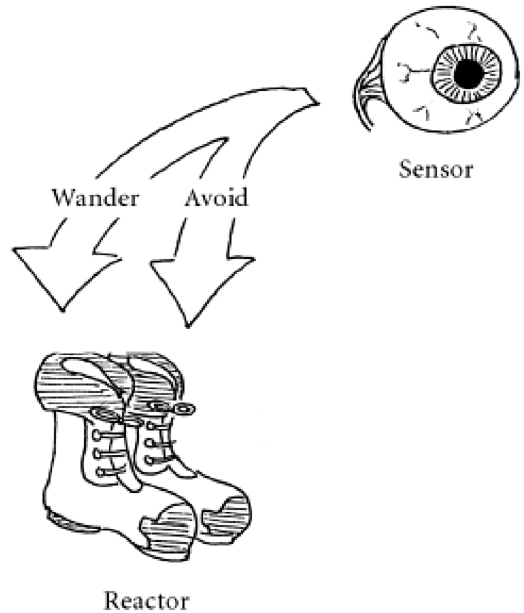
OAK RIDGE
National Laboratory

49

Quote

“the world is the best model of itself.”

–Rodney Brooks, 1990, Elephants and nouvelle AI



Nouvelle AI. A robot should sense and then move according to simple rules such as “Avoid collisions” or “Wander.”

Source of Input Data

- 3 Campbell Creek homes (TVA, ORNL, EPRI)

- ~144 sensors/home, 15-minute data:
  - Temperature (inside/outside)
  - Plugs
  - Lights
  - Range
  - Washer
  - Radiated heat
  - Dryer
  - Refrigerator
  - Dishwasher
  - Heat pump air flow
  - Shower water flow
  - Etc.

MLSuite example

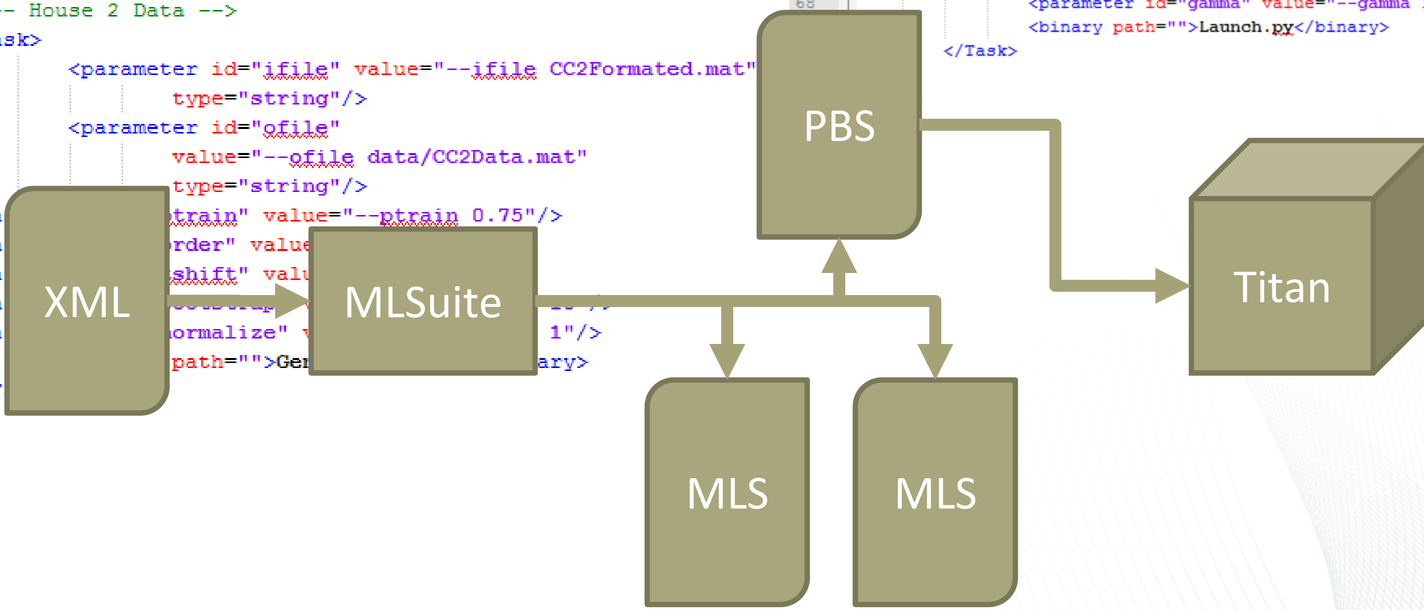
Data Preparation:

```
<Job>
        <!-- House 1 Data  -->
        <Task>
                <parameter id="ifile" value="--ifile CC1Formated.mat"
                        type="string"/>
                <parameter id="ofile"
                        value="--ofile data/CC1Data.mat"
                        type="string"/>
        <parameter id="ptrain" value="--ptrain 0.75"/>
        <parameter id="bootstrap" value="--bootstrap 10"/>
        <parameter id="normalize" value="--normalize 1"/>
                <binary path="">GenerateData.py</binary>
        </Task>


        <!-- House 2 Data -->
        <Task>
                <parameter id="ifile" value="--ifile CC2Formated.mat"
                        type="string"/>
                <parameter id="ofile"
                        value="--ofile data/CC2Data.mat"
                        type="string"/>
        <pa           train" value="--ptrain 0.75"/>
        <pa           rder" value
        <pa           shift" valu
        <pa                                        "/>
        <pa           ormalize" v                  1"/>
                      path="">Ge                   ary>
        </T

</Job>
```

30x LS-SVM variants

(train/test and input order)

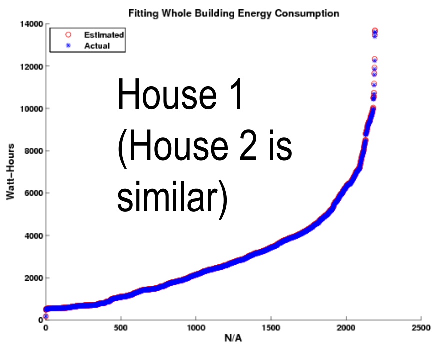
```
<Job>
        <Task>
                <parameter id="data" value="--data data/CC1Data.mat"/>
                <parameter id="output"
                        value="--ofile results/LSSVM/CC1/"
                        type="string"/>
                <parameter id="fold" value="--fold" type="string"/>
                <parameter id="foldy" value="1" maxvalue="10" step="1"
                    type="numeric"/>
                <parameter id="order" value="--order" type="string"/>
                <parameter id="ordery" value="1" maxvalue="3" step="1"
                    type="numeric"/>
                <parameter id="tshift" value="--tshift 1"/>
                <parameter id="learner" value="--learner EvaluateLSSVM"
                    type="string"/>
                <parameter id="C" value="--C 33.6134" type="string"/>
                <parameter id="gamma" value="--gamma 106.1072" type="string"/>
                <binary path="">Launch.py</binary>
        </Task>
```

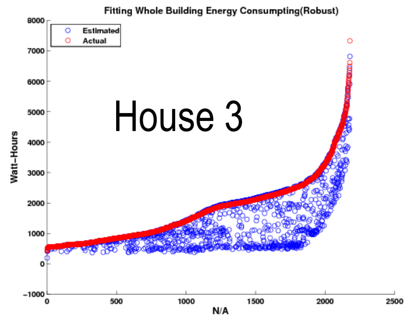


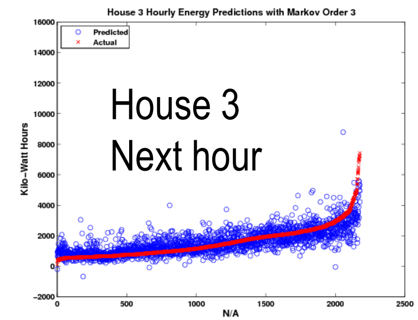
OAK RIDGE
National Laboratory

Applications of machine learning

- Linear Regression predicting whole building energy use







- Accuracy Metrics for best subset of sensors

▸ Root Mean Squared Error(RMSE):

$$RMSE = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - p_i)^2}$$

▸ Mean Absolute Percentage of Error(MAPE):

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - p_i|}{y_i}$$

▸ Coefficient of Variance(CV):

$$CV = \frac{RMSE}{y_{mean}} \times 100$$

▸ Mean Bias Error(MBE):

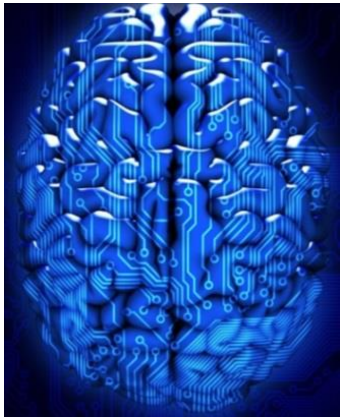
$$MBE = \frac{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - p_i)}{y_{mean}} \times 100$$

	HME FFNN	HME LS-SVM	SVR	FCM
RMSE(Watt-Hours)	569.96±50.13	582.61±33.97	603.85±40.55	581.87±41.67
MAPE(%)	17.07±1.19	15.94±0.92	15.48±0.87	17.37±1.02
CV(%)	20.14±1.65	20.59±1.12	21.32±1.32	20.56±1.37
MBE(%)	0.42±1.17	-0.07±0.89	-1.50±0.80	0.01±0.99

	Best Four Sensors	Best Model	Top 10 Sensors
RMSE	1127.88±33.00	942.25±26.14	1129.04±32.38
MAPE	41.17±1.12	30.53±1.03	40.4483±1.29
CV	39.76±1.02	33.21±0.73	39.80±0.96
MBE	-0.04±0.90	-0.06±0.92	-0.05±1.05
ICOMP(IFIM)	2166.3±1.54	1845.88±21.25	2125.50±2.72

MLSuite: HPC-enabled Suite of Machine Learning algorithms

- Linear regression
- Feedforward neural network
- Support vector machine regression
- Non-linear regression
- K-means with local models
- Gaussian mixture model with local models



- Self-organizing map with local models
- Regression tree (using information gain)
- Time modeling with local models
- Recurrent neural networks
- Genetic algorithms
- Ensemble learning

Evolutionary computation

How are offspring produced?

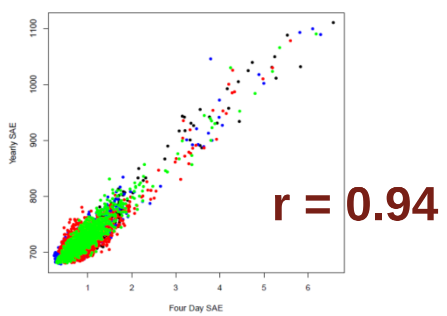
	Thickness	Conductivity	Density	Specific Heat
Bldg1	0.022	0.031	29.2	1647.3
Bldg2	0.027	0.025	34.3	1402.5
$(1+2)_1$	0.0229	0.029	34.13	1494.7
$(1+2)_2$	0.0262	0.024	26.72	1502.9

- Average each component
- Add Gaussian noise
- … "AI inside of AI"

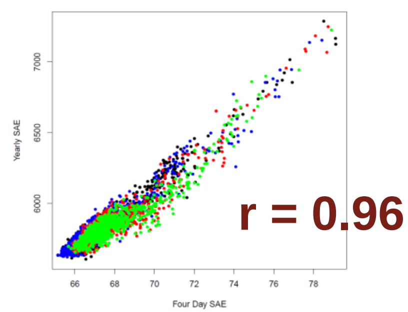
OAK RIDGE National Laboratory

Getting more for less

- EnergyPlus is slow
  - Full-year schedule
  - 2 minutes per simulation
- Use abbreviated 4-day schedule instead
  - Jan 1, Apr 1, Aug 1, Nov 1
  - 10 – 20 seconds per simulation

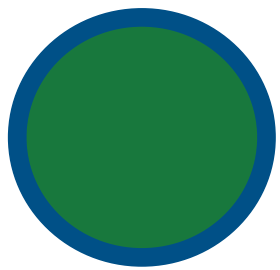


Monthly Electrical Usage

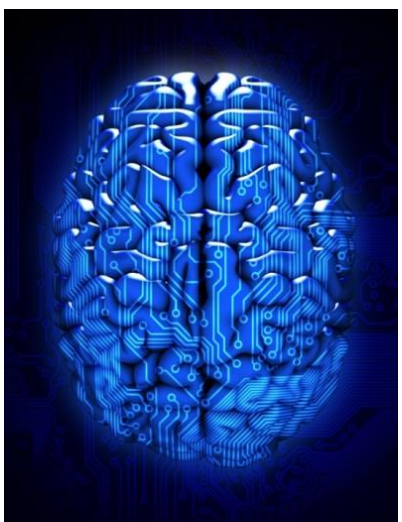


Hourly Electrical Usage

Evolutionary combination

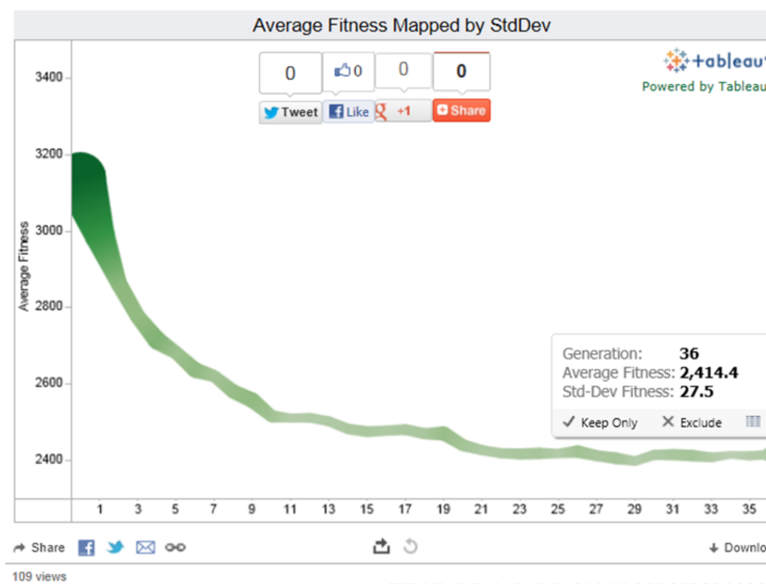


Island Hopping



4 of 19 experiments

1. Surrogate Modeling
2. Sensor-based Energy Modeling (sBEM)
3. Abbreviated Schedule
4. Island-model evolution



Automated M&V process Autotune calibration of simulation to measurements

XSEDE and DOE Office of Science → **DOE-EERE BTO** → **Industry and building owners**

Nautilus

No of Processors	No of E+ sims	Wall-clock Time (h:mm:ss)	Time/E+ task (mm:ss)
32	32	0:02:08	2:08
64	64	0:03:04	3:04
128	128	0:04:11	4:11
128	1024	0:34:24	4:18
256	2048	1:25:17	10:40
512	1024	0:18:05	9:02

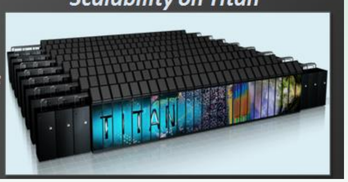
Scalability on Nautilus

Gordon

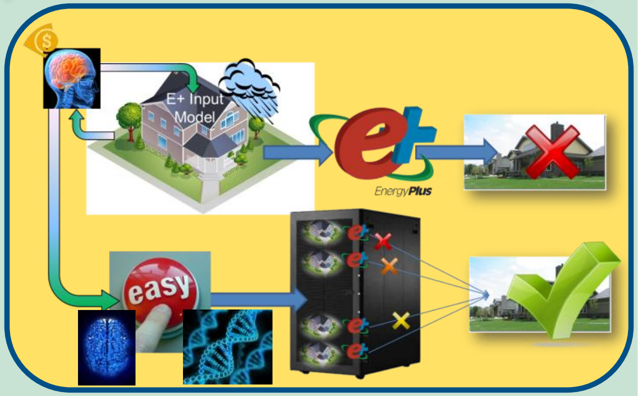
Kraken

No of Processors	E+ Tasks		Wall-clock Time (mm:ss)
64	256		18:34
128	512		18:22
256	1024		20:30
512	2048		20:43
1024	4096		21:03
2048	8192		21:11
4096	16384		20:00
8192	32768		26:14
16384	65536		26:11
65536	262144		44:52
131072	524288	45TB	68:08

Scalability on Titan



8 million building sims
270TB dataset!
(0.5M, 45TB, 1hr)



Features:

- Works with "any" software
- Tunes 100s of variables
- Customizable distributions
- Matches 1+ million points

Commercial Buildings

		ASHRAE G14 Requires
Monthly utility data	CVR	15%
	NMBE	5%
Hourly utility data	CVR	30%
	NMBE	10%

Residential home	Tuned input avg. error
Within 30¢/day (actual use $4.97/day)	Hourly – 8% Monthly – 15%

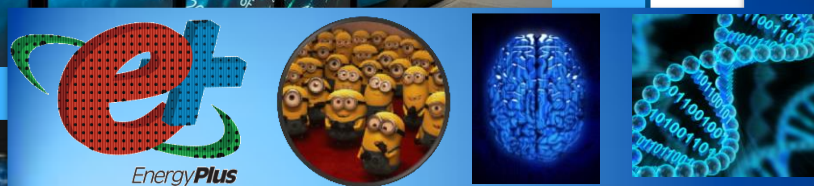
10+ companies interested

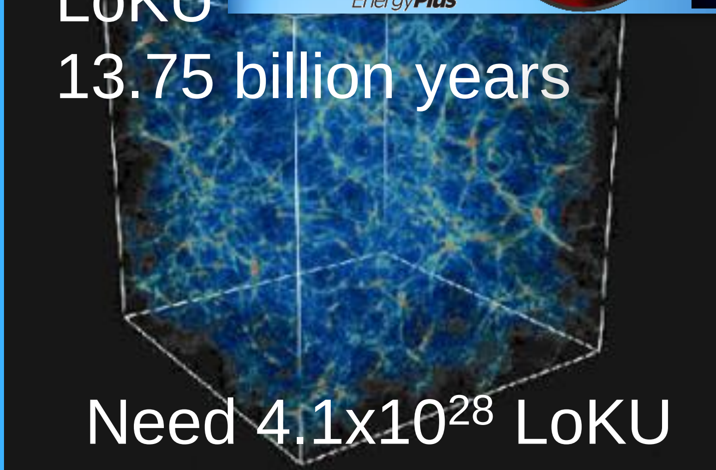
Leveraging HPC resources to calibrate models for optimized building efficiency decisions

OAK RIDGE National Laboratory

58

HPC-informed algorithmic reduction… to commodity hardware





LoKU
13.75 billion years

Need 4.1x10^{28} LoKU

1 hour

Discussion

Oak Ridge National Laboratory
EESD – Martin Keller
ETSD – Johney Green
BTRIC – Patrick Hughes &
Ed Vineyard
WBCI – Melissa Lapsa

Joshua New, Ph.D.
newjr@ornl.gov