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

• 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
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
Energy Consumption and Production

U.S. Primary Energy Consumption

- Residential Buildings: 28%
- Commercial Buildings: 41%
- Industrial: 19%
- Transportation: 31%

Commercial Site Energy Consumption by End Use

- HVAC: 32%
- Lighting: 25%
- Other 7%
- Water heating: 6%
- Electronics & Computers: 12%
- Refrigeration: 4%
- Cooking: 2%
- Wet cleaning: 3%
- Refrigeration: 4%
- Electronics: 5%
- Cooking: 4%
- Lighting: 6%
- Space cooling: 9%
- Water heating: 19%
- Space heating: 45%

The Role of Renewable Energy in the Nation's Energy Supply, 2009

- Total = 94.578 Quadrillion Btu (Petroleum 37%, Natural Gas 25%, Coal 21%, Nuclear Electric Power 9%, Renewable Energy 8%)
- Total = 7.744 Quadrillion Btu (Geothermal 5%, Biomass waste 6%, Wind 9%, Wood 24%, Biofuels 20%, Hydropower 35%, Solar 1%)

World Energy Production

- Coal, Petroleum, Natural Gas, Nuclear, Wind, Hydro, Biomass, Solar, Photovoltaic
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
Presentation summary

• Scientific Paradigms
• Roof Savings Calculator
• Visual Analytics
• Knowledge Work
• Autotune
Urban Heat Island Effect and Albedo Engineering

Image from Lawrence Berkeley National Laboratory
Computer tools for simulating cool roofs

Roof Savings Calculator (RSC)

Chris Scruton
CEC

R. Levinson,
H. Gilbert,
H. Akbari

Marc LaFrance
DOE BT

A. Desjarlais,
W. Miller,
J. New

INDUSTRY

COLLABORATIVE
R&D

LBNL

ORNL

WBT

Joe Huang,
Ender Erdem
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
RSC = AtticSim + DOE-2.1E

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

Office  “Big Box” Retail  Warehouse

Residential Roof Savings Calculator (RSC)

Building
1. Closest location (similar weather):
   Select location
2. Building Type: Residential
3. Conditioned floor area (ft²): 2025
4. Number of floors: 1

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 (80%)
   - Low-efficiency (70%)
   - Custom
8. Cooling system efficiency (SEER):
   - High-efficiency (15)
   - Mid-efficiency (13)
   - Low-efficiency (10)
   - Custom
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-5
    - R-3
    - None
16. Duct location:
    - Conditioned space
    - Attic
17. Duct leakage:
    - Inspected (4%)
    - Uninspected (14%)

Simulation Results

$\text{$/yr}$

Energy Savings
- Total: $1600.00 (31.6%)
- Cooling: $500.00 (28.6%)
- Heating: $1100.00 (4.6%)

Monthly Savings

Retract Monthly Results

Monthly Results

Download:
- Run Input Data
- Run Output Data

AtticSim

DOE-2
RoofCalc.com impact

Dashboard

100,000+ visitors, 200+ user feedback,

Average: ~81 visitors/day

30,752 visits came from 112 countries/territories

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

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

<table>
<thead>
<tr>
<th>Description</th>
<th>Reflectance</th>
<th>Emisivity</th>
<th>SRI</th>
<th>Houston $ saved</th>
<th>...13</th>
</tr>
</thead>
<tbody>
<tr>
<td>BUR No Coating</td>
<td>10</td>
<td>90</td>
<td>6</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>Mineral Mod Bit</td>
<td>25</td>
<td>88</td>
<td>25</td>
<td>103</td>
<td></td>
</tr>
<tr>
<td>Single Ply</td>
<td>32</td>
<td>90</td>
<td>35</td>
<td>230</td>
<td></td>
</tr>
<tr>
<td>Mineral Mod Bit</td>
<td>33</td>
<td>92</td>
<td>35</td>
<td>197</td>
<td></td>
</tr>
<tr>
<td>Metal</td>
<td>35</td>
<td>82</td>
<td>35</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Aluminum Coating</td>
<td>43</td>
<td>58</td>
<td>35</td>
<td>279</td>
<td></td>
</tr>
<tr>
<td>Mineral Mod Bit</td>
<td>45</td>
<td>79</td>
<td>55</td>
<td>291</td>
<td></td>
</tr>
<tr>
<td>Coating over BUR</td>
<td>49</td>
<td>83</td>
<td>55</td>
<td>433</td>
<td></td>
</tr>
<tr>
<td>Metal</td>
<td>49</td>
<td>83</td>
<td>55</td>
<td>208</td>
<td></td>
</tr>
</tbody>
</table>

Summer operation of HVAC duct in ASHRAE climate zone 3
Enhanced RSC Site

Input Parameter GUI

Result Output

Database

Total Savings: $-178

These numbers represent our estimate of your energy costs. You can specify new values and re-calculate.

Electricity price (cents per kWh):

11.68

Natural gas price (dollars per 1000 ft³):

11.65

Re-calculate savings

User

Hyperion

Savings

Inputs

Results

Exists?

Simulation

Savings

RSC Engine
“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.”
RSC Service Example (Python)

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

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

DOE: Office of Science

CEC & DOE EERE: BTO

Industry & Building Owners

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

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.
Personal story behind one of DOE’s RSC images

14. Radiant barrier present:
- Yes
- No

15. Attic insulation (hr):
- R-50
- R-38
- R-19
- R-11
- R-7
- R-5
- R-3
- None

16. Duct location:
- Conditioned space
- Attic

17. Duct leakage:
- Inspected (4%)
- Uninspected (14%)

Duct Leakage
Leaky ducts in unconditioned spaces are effectively costing you money to condition the planet, not your house. Commercial buildings have typical leakage rate of 10–20%; likewise, residential buildings typically have duct leakage rates near 14%. The CEC’s Title 24 target leakage rate for inspected ducts is 4% and requires no greater than 6%. This calculator supports duct leakage rates of 4% and 14%.

Leaky Connection  | Damaged Duct  | Sealed Ducts

[Image of duct connection types]

OAK RIDGE National Laboratory
Presentation summary

• Scientific Paradigms
• Roof Savings Calculator
• Visual Analytics
• Knowledge Work
• Autotune
PCP - car data set
PCP bin rendering (data)

- Transfer function coloring:
  - Occupancy or leading axis
The power of “and” – linked views (info)

Roof Savings Calculator
www.roofcalc.com
Dr. Joshua New (ORNL) and Chad Jones (UC-Davis)
Dr. William A. Miller (ORNL), A. Desjarlais (ORNL), Yu Joe Huang (WhiteBox), Ender Erdem (WhiteBox)

Multivariate Visualization of Large-Scale Parameter Sweeps

Parallel Coordinates Plots

Time-variant Function Plots

Category Charts

Climate Zone Map
Large Data Visualization

Knowledge

Large Data Visualization
Outliers (wisdom)

- Selection of heating outliers
- Find all have box building type and in Miami
Impact – RSC and Visual Analytics

12 Publications, 20+ organizations interested in licensing


Presentation summary

- Scientific Paradigms
- Roof Savings Calculator
- Visual Analytics
- Knowledge Work (context)
- Autotune
McKinsey Global Institute Analysis

Exhibit E3
Estimated potential economic impact of technologies from sized applications in 2025, including consumer surplus
$ trillion, annual

- Mobile Internet: 3.7–10.8
- Automation of knowledge work: 5.2–6.7
- The Internet of Things: 2.7–6.2
- Cloud technology: 1.7–6.2
- Advanced robotics: 1.7–4.5

Source: McKinsey Global Institute analysis
$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
Learning associations

Full Results

Detailed Results
Presentation summary

• Scientific Paradigms
• Roof Savings Calculator
• Visual Analytics
• Knowledge Work (context)
• Autotune
Existing tools for retrofit optimization

Simulation Engine
DOE—$65M (1995–?)
Business limitations for M&V

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

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

Acknowledgment: Jibo Sanyal, ORNL R&D Staff
Computational complexity

Problems/Opportunities:
Domain experts chose to vary 156
Brute-force = $5 \times 10^{52}$ simulations

E+ parameters

LoKU
13.75 billion years
Need $4.1 \times 10^{28}$ LoKU
No database technology sufficient?

Relational, Columnar, NoSQL, different compression and partitioning

<table>
<thead>
<tr>
<th>MyISAM</th>
<th>InnoDB</th>
</tr>
</thead>
<tbody>
<tr>
<td>• No ACID</td>
<td>• ACID compliant</td>
</tr>
<tr>
<td>• No foreign keys</td>
<td>• Foreign keys</td>
</tr>
<tr>
<td>• Speed at scale</td>
<td>• Slower adding data</td>
</tr>
<tr>
<td>– 0.71 s on LOAD DATA</td>
<td>– 2.3 s on LOAD DATA</td>
</tr>
<tr>
<td>• Better compression</td>
<td>• Poorer compression</td>
</tr>
<tr>
<td>– 10.27 MB</td>
<td>– 15.4 MB</td>
</tr>
<tr>
<td>– Read only compression:</td>
<td></td>
</tr>
<tr>
<td>6.003 MB</td>
<td></td>
</tr>
<tr>
<td>• $2^{32}$ rows maximum</td>
<td></td>
</tr>
</tbody>
</table>

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

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
MLSuite example

- EnergyPlus – 2-10 mins for an annual simulation

```plaintext
!- ALL OBJECTS IN CLASS
Version, 7.0;  !- Version

!- SIMULATION CONTROL
Simulation Control, No, !-Do Zone Sizing Calc No, !-Do System Sizing Calc
...
```

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

30x LS-SVM variants
(train/test and input order)
Applications of machine learning

• Linear Regression predicting whole building energy use

House 1
(House 2 is similar)

House 3

House 3
Next hour

• 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):

\[ MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - p_i}{y_i} \right| \]

Coefficient of Variance (CV):

\[ CV = \frac{RMSE}{\bar{y}} \times 100 \]

Mean Bias Error (MBE):

\[ MBE = \frac{1}{N} \sum_{i=1}^{N} (y_i - p_i) \times 100 \]

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

Best Four 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?**

<table>
<thead>
<tr>
<th></th>
<th>Thickness</th>
<th>Conductivity</th>
<th>Density</th>
<th>Specific Heat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bldg1</td>
<td>0.022</td>
<td>0.031</td>
<td>29.2</td>
<td>1647.3</td>
</tr>
<tr>
<td>Bldg2</td>
<td>0.027</td>
<td>0.025</td>
<td>34.3</td>
<td>1402.5</td>
</tr>
<tr>
<td>(1+2)_1</td>
<td>0.0229</td>
<td>0.029</td>
<td>34.13</td>
<td>1494.7</td>
</tr>
<tr>
<td>(1+2)_2</td>
<td>0.0262</td>
<td>0.024</td>
<td>26.72</td>
<td>1502.9</td>
</tr>
</tbody>
</table>

- Average each component
- Add Gaussian noise
- … “AI inside of AI”
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

$r = 0.94$

Monthly Electrical Usage

$r = 0.96$

Hourly Electrical Usage
Evolutionary combination

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

Features:
- Works with “any” software
- Tunes 100s of variables
- Customizable distributions
- Matches 1+ million points

ASHRAE G14 Requires

<table>
<thead>
<tr>
<th>Results</th>
<th>Monthly utility data</th>
<th>Hourly utility data</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVR</td>
<td>15%</td>
<td>30%</td>
</tr>
<tr>
<td>NMBE</td>
<td>5%</td>
<td>10%</td>
</tr>
</tbody>
</table>

Commercial Buildings

Residential home

<table>
<thead>
<tr>
<th>Tuned input avg. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within 30¢/day (actual use $4.97/day)</td>
</tr>
<tr>
<td>Hourly – 8%</td>
</tr>
<tr>
<td>Monthly – 15%</td>
</tr>
</tbody>
</table>

Leveraging HPC resources to calibrate models for optimized building efficiency decisions
HPC-informed algorithmic reduction... to commodity hardware

LoKU
13.75 billion years

Need $4.1 \times 10^{28}$ 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