Cognitive Internet of Things

Hendrik F. Hamann et al.
T.J. Watson Research Center, New York, USA

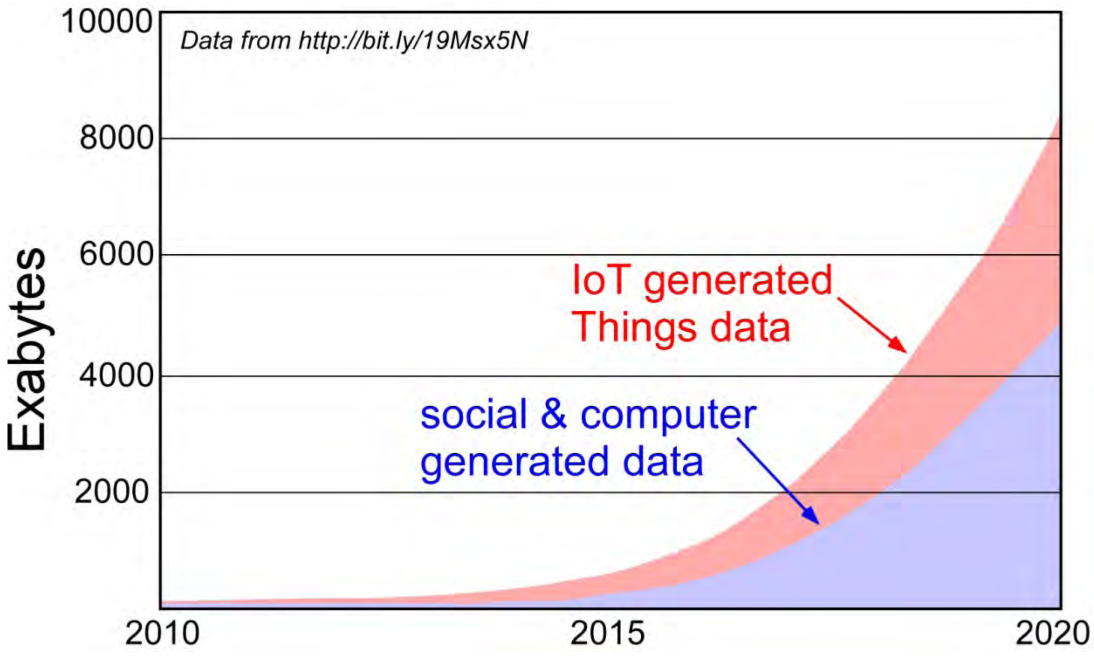
* Steve Lohr, NY Times, “Data-ism”

Outline

- Trends - Internet of Things
- Research experimental platform for Cognitive IoT

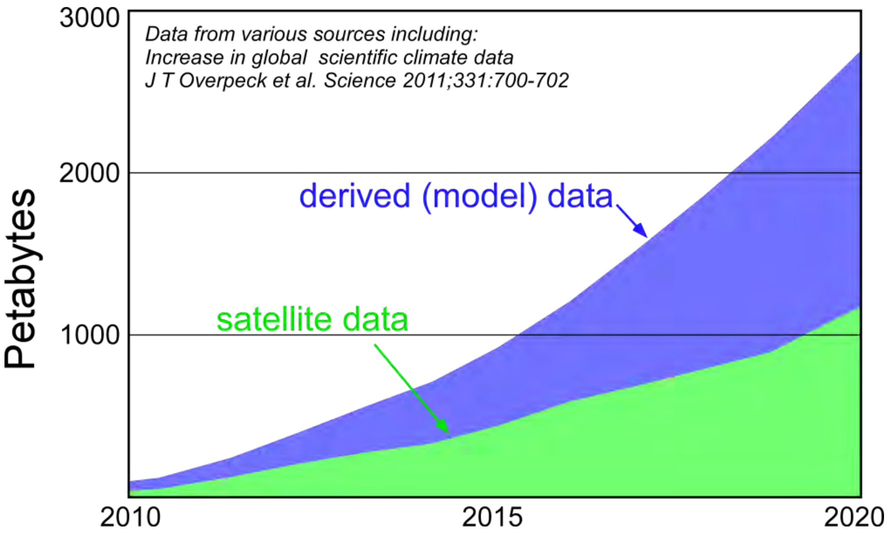
“Big” Data	Physical Analytics	Application & Solution
IoT sensor networks	Physical model complexity reduction	**Energy efficiency** in buildings / data centers
Simulation and model outputs	Situation-dependent, machine-learnt, multi-model blending	**Renewable energy forecasts /** Pollution managements
Satellite observations and IoT sensors	Statistical tuned evapo-transpiration model	**Water efficiency** and yields in **agriculture**

“Internet-of-Things” generated data soon bigger than social and transactional data



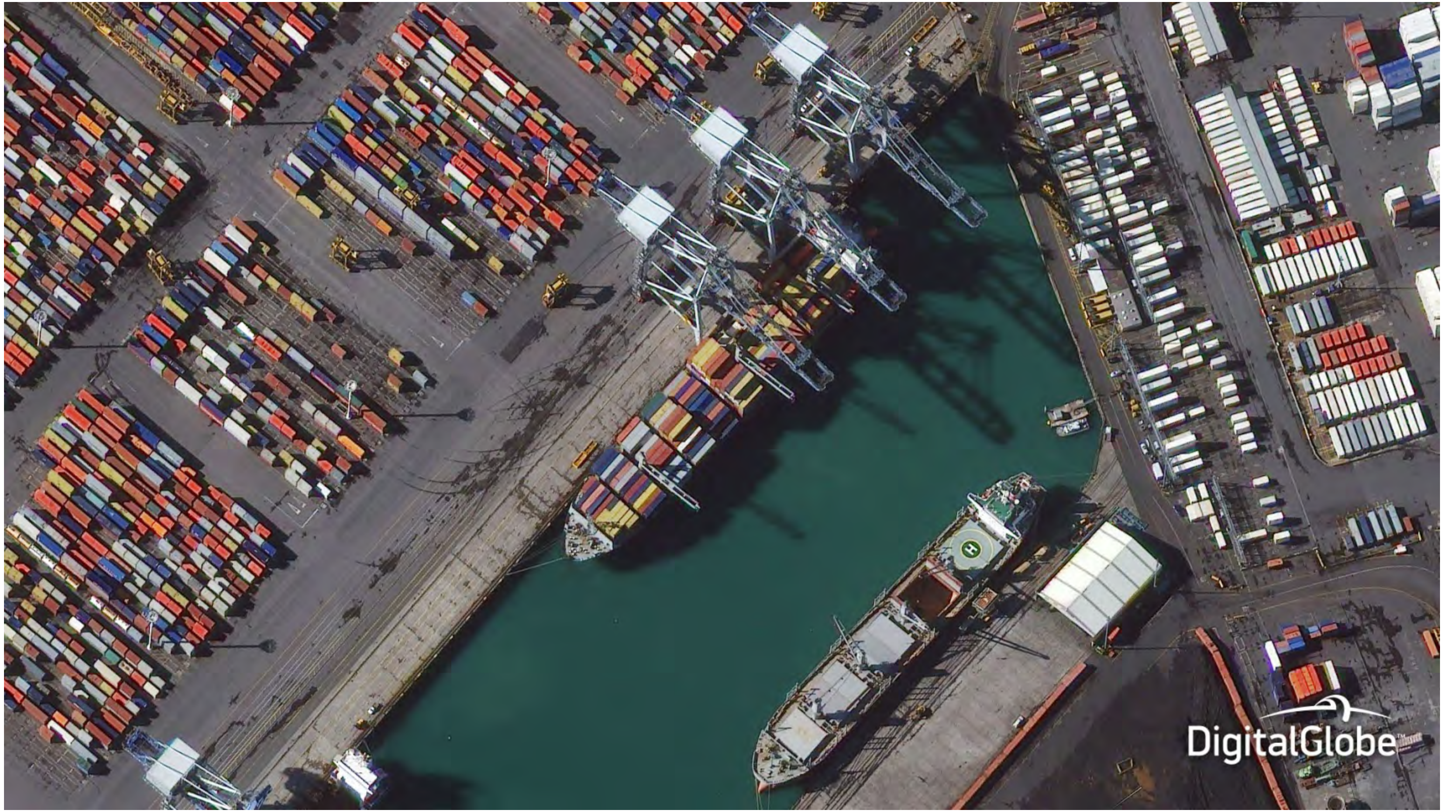
~ 2x more data growth (currently @ 44EB/month) than social & computer generated data

Data from remote sensors is exploding



- **Number of satellite launches are growing exponentially**
  - Many countries have re-started major satellite programs (China, Europe, India etc.)
  - Cube satellites, Nano-satellites
- **Drones, cell phone with hyper-spectral cameras**

The physical world is being digitized….



Auckland, New Zealand seen at 30 cm resolution. Photo: DigitalGlobe

Traditional (physical) industries are being transformed by (IoT) data

Industry	Past – Selling a Product	Future - Service
Healthcare	Diabetes pumps	Diabetes care
Agriculture	Seeds	Crop yields
Consumer	Packaged goods	Nutrition
Automotive	Cars	Transportation
…	…	…
IT Industry	Computers	Computation

McKinsey, GE, IBM, Cisco et al. estimate **hundreds of billions dollar** savings/efficiency improvements in the next 10 years

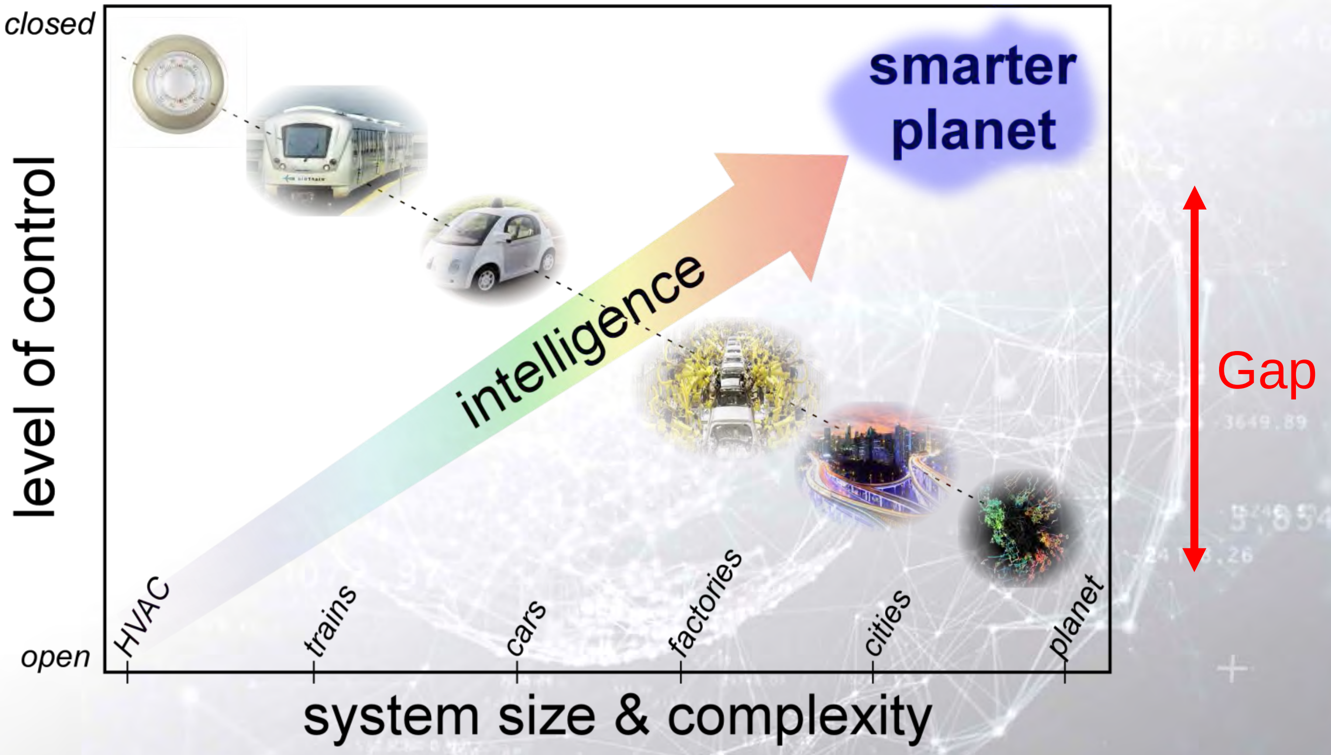
Watson IoT is a brand new IBM business unit headquartered in Munich, Germany

1. Watson Analytics
   - Physical Analytics
2. Security, Privacy and Trust
   - IBM Security 360
3. Platforms
   - Bluemix

Industry solutions

- IBM Research prototypes and develops new technologies and solutions
- My role…Pioneered Physical Analytics in Research the last 1o years

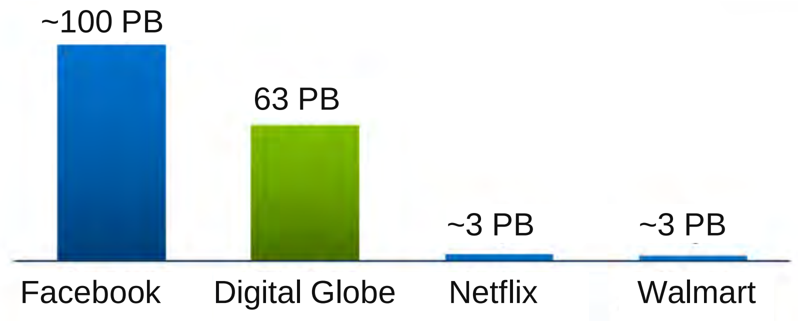
But "closing-the-loop" for complex systems will require more "physical" intelligence



Big data from the physical world is "special" and thus Analytics for IoT data should be different

- Data from the physical world is becoming soon **"mega" big.**
  - 44EByte / months corresponds to **~ 100M hard disks**
  - Processing 44 EByte with 100M servers **would still take more than one hour**
  - **Data throughput** (rather than processing) has become the limiting factor
- IoT has more **noise** and is more prone to **error.**
- Data **security**, **verification** etc even more important.
- Data **not always cheap**.

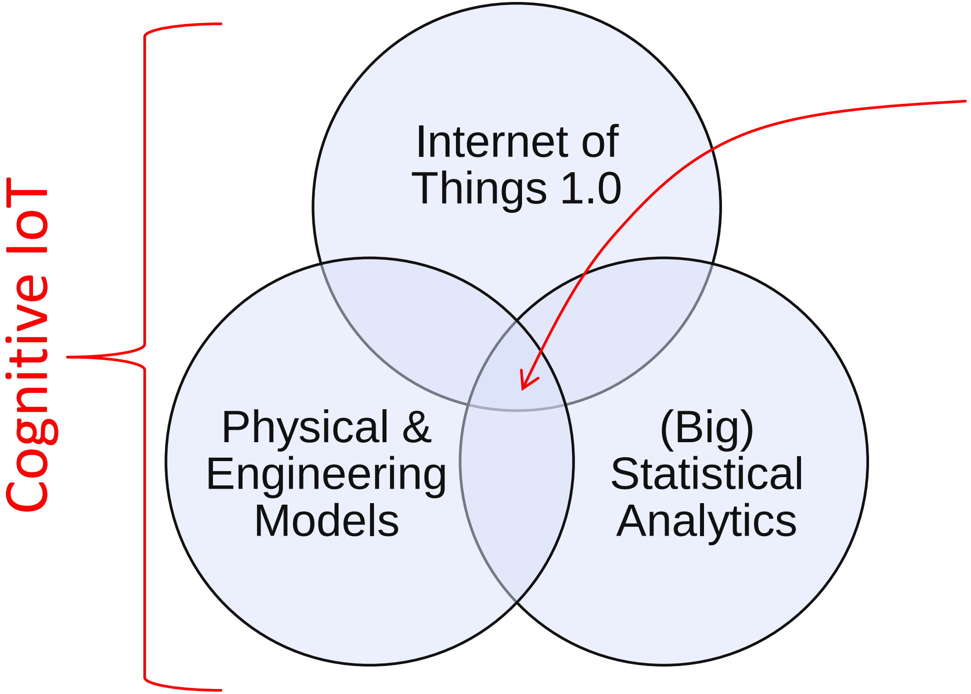
Data size owned by selected companies



Sources: Facebook IPO Propectus; Bloomberg May 2013; SAS 2012

Analyzing pressure sensor data from a **pipeline system** or
accoustic sensor data sensors from an engine or
vibration sensor data on a suspension bridge
Is very different than analyzing a chat log

Physical Analytics lies between IoT 1.0, physics-based modeling and big data analytics



Highly inter disciplinary with many interesting research topics

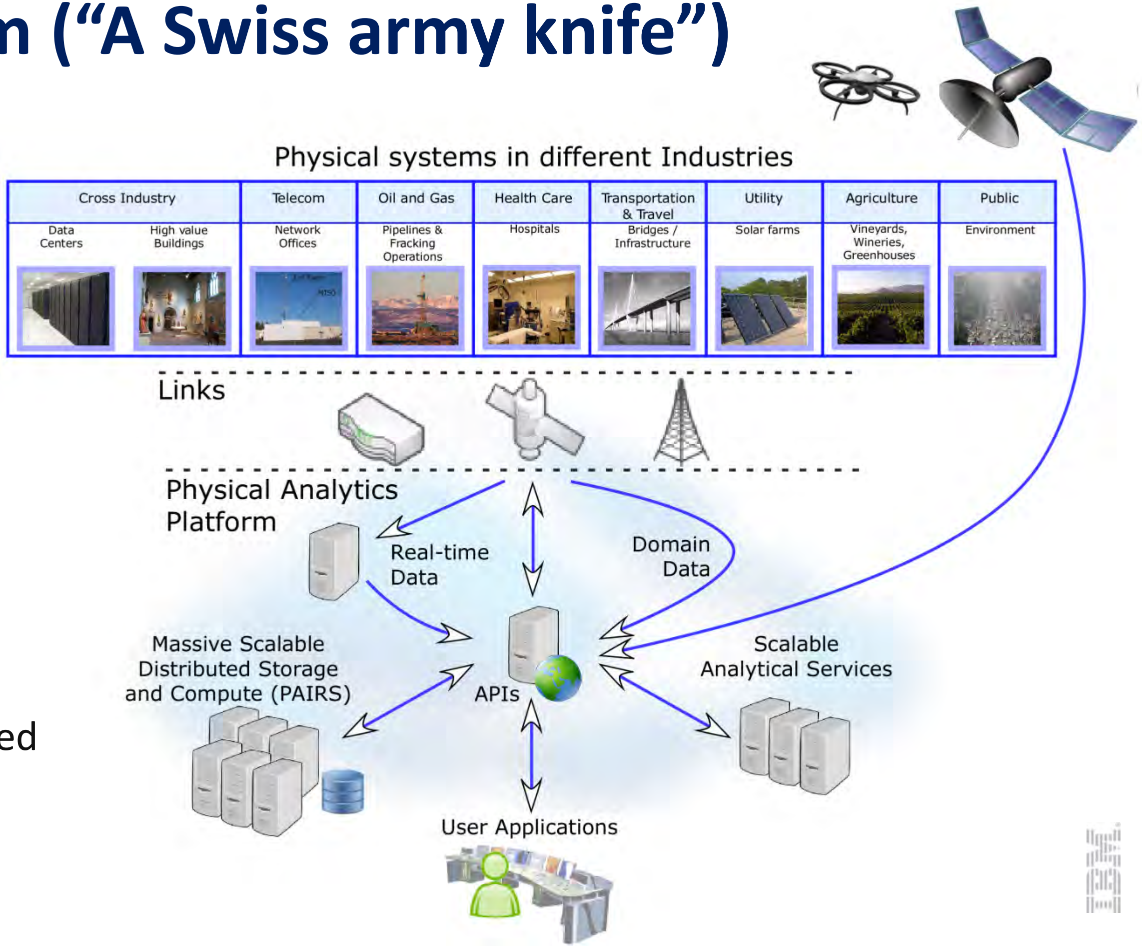
- Statistical learning under physical constraints
- Model complexity reduction
- Situation-dependent, machine-learnt multi-model blending
- Graph theory and statistical physics
- Parallelizing physical models for data-intensive computation
- …..

- Feed learning back to improve understanding of the underlying physics

An end-to-end IBM Research Physical Analytics (or cognitive IoT) platform ("A Swiss army knife")

- Sensor / Actuators
- Communication, Security
- Cloud-based IoT SW Platform
  - Server, Agents, Apps, Clients
- Physical analytics
  - Physical models
  - Reduction of Model Complexity
  - Machine-learnt, multi-model model blending
- (Big) data analytics platform
- Automation and Controls

- Platform has been successfully applied in various Industries

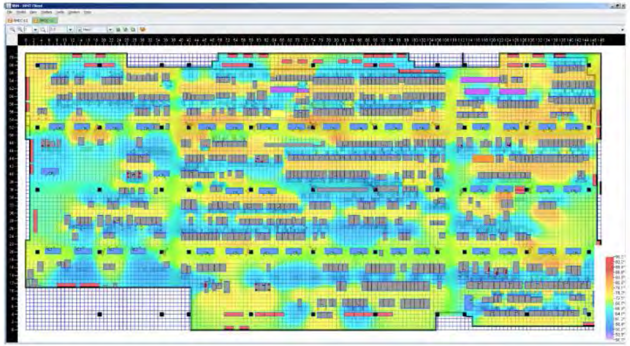


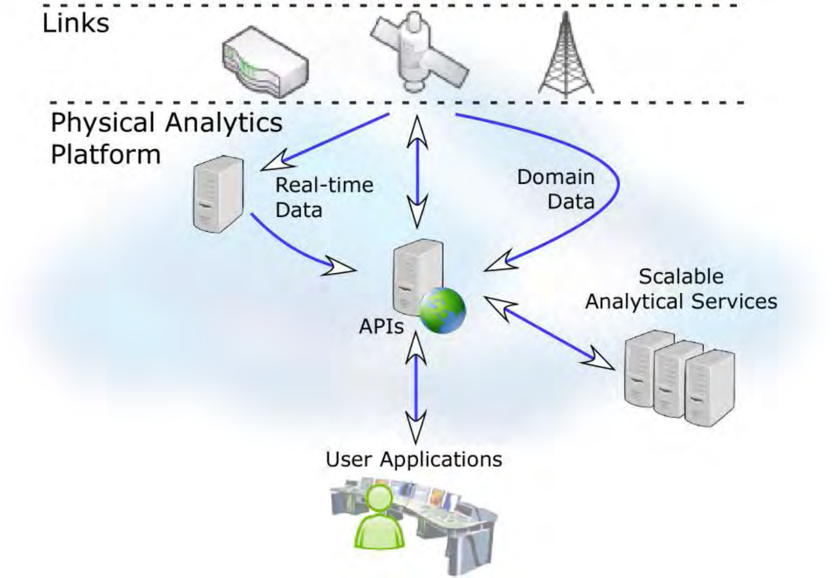
Energy management in data centers and buildings

Partners: Georgia Tech, AT&T, IBM GTS, DoE-EERE

PA platform applied to energy management in data centers and buildings

- Sensor / Actuators
- Communication, Security
- Cloud-based SW Platform
  - Server, Agents, Apps, Clients
- Physical analytics
  - Physical models
  - Reduction of Model Complexity
  - Machine-learnt, multi-model model blending
- (Big) data analytics platform
- Automation and controls



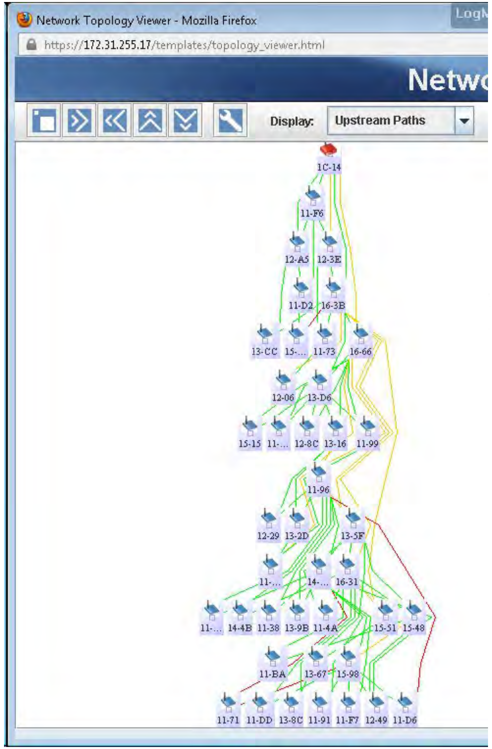


Data Centers are energy hogs

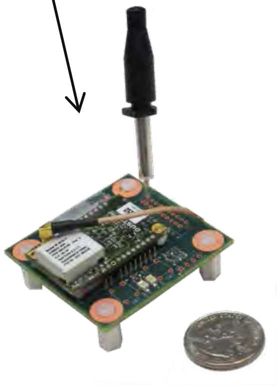
- DCs consume ~ 3% of US electricity
- **Annual growth of 15% is unsustainable (doubles every 5 years)**
- **Up to 50% of all energy in data centers is used for cooling**

- DC energy demand is surging because of
  - o insatiable IT demand, which has outpaced power-performance improvements
  - o power-limited core technology

Large-scale sensor & actuator mesh networks deployed around the world



IBM “Mote” provides wireless mesh network communication of sensors and actuators



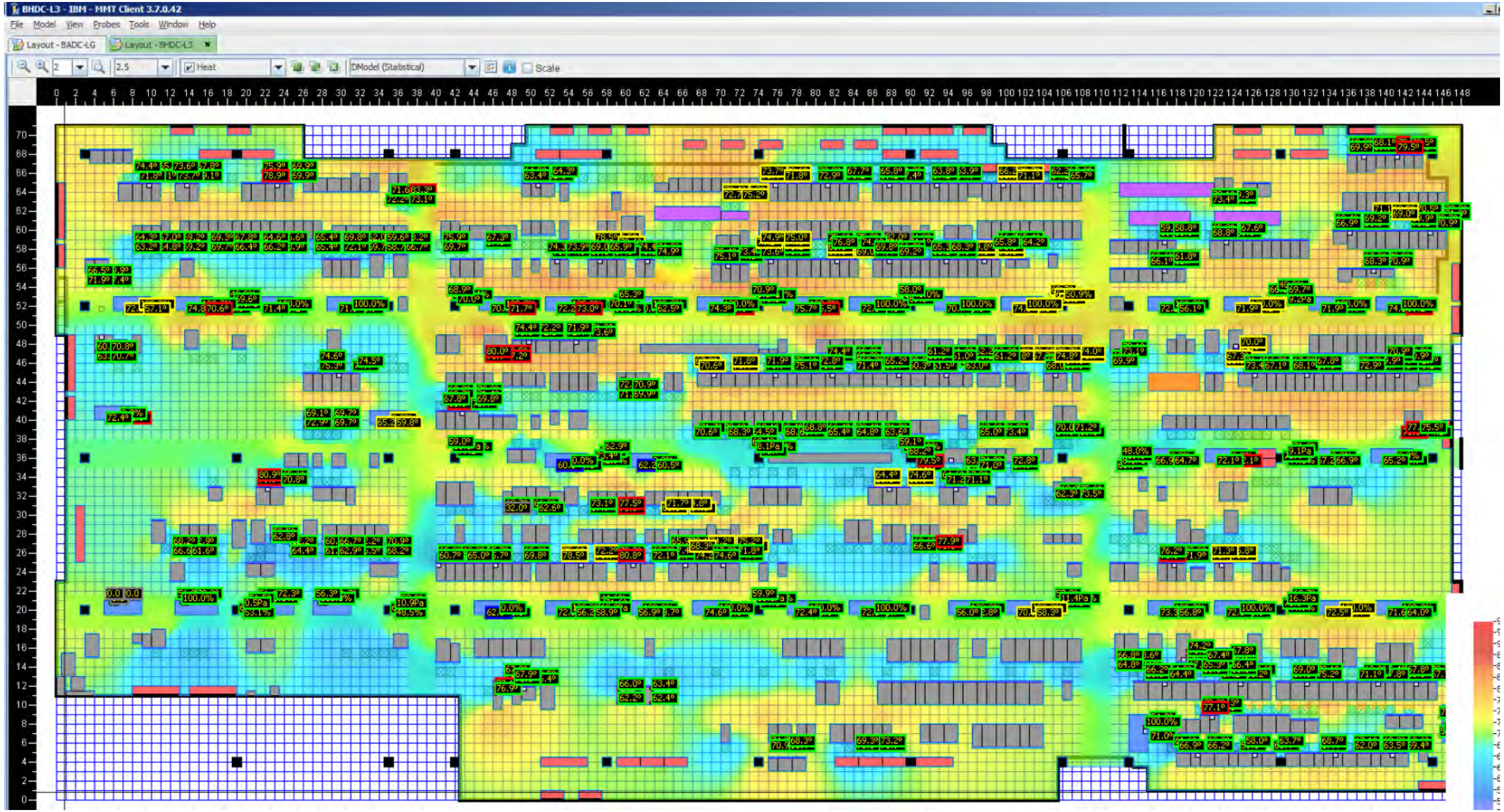
- **Deployed in >180 data centers world-wide**
- **Almost 1M sensors supported**
- **Example deployment***
  - 4300 thermal sensors
  - 250 pressure sensors
  - 612 flow sensors
  - 300 humidity sensors
  - 1200 power/current sensors
  - 80 air condition unit controllers

*for a 10000 m^2 data centers

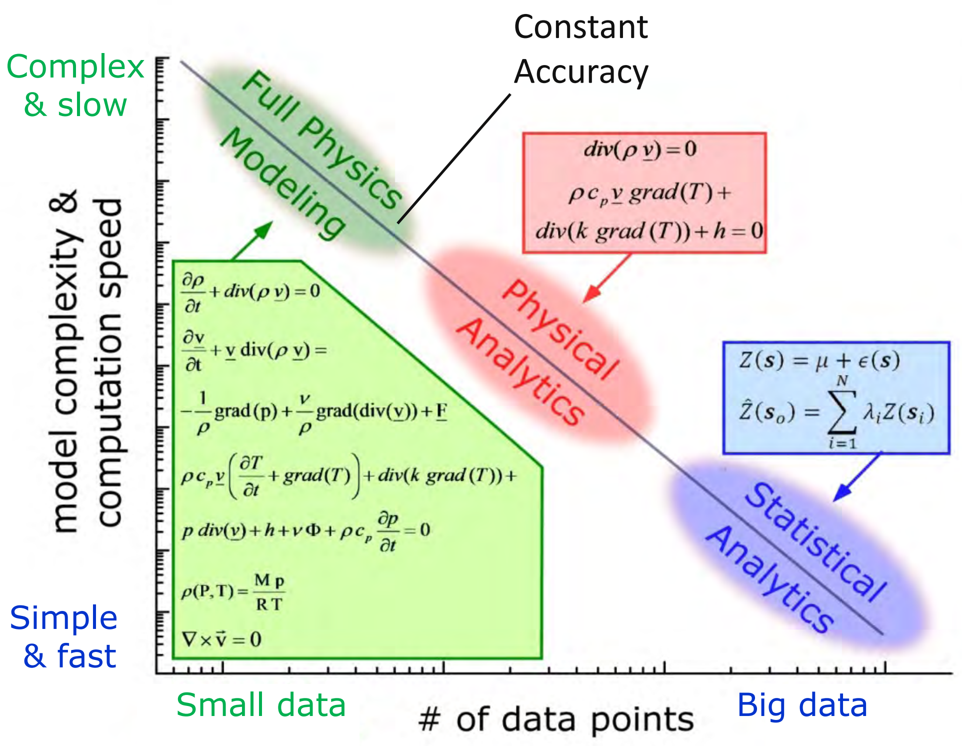
Energy efficiency improvements require real-time 3D heat distributions? How to get from here….

Sensors Reporting

Energy efficiency improvements require real-time 3D heat distributions? …to there



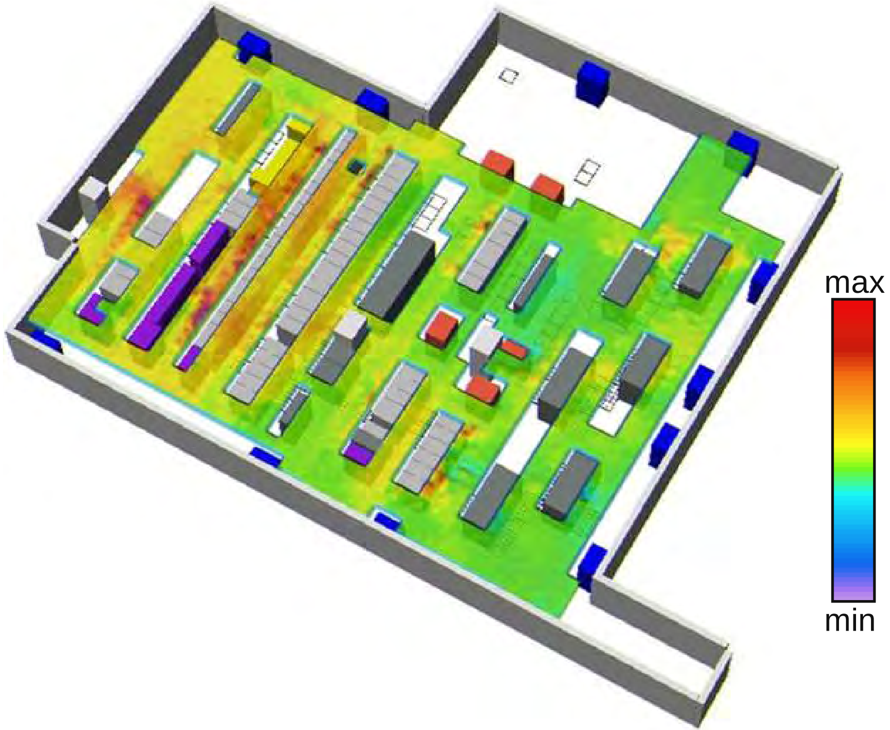
Trading complexity against data



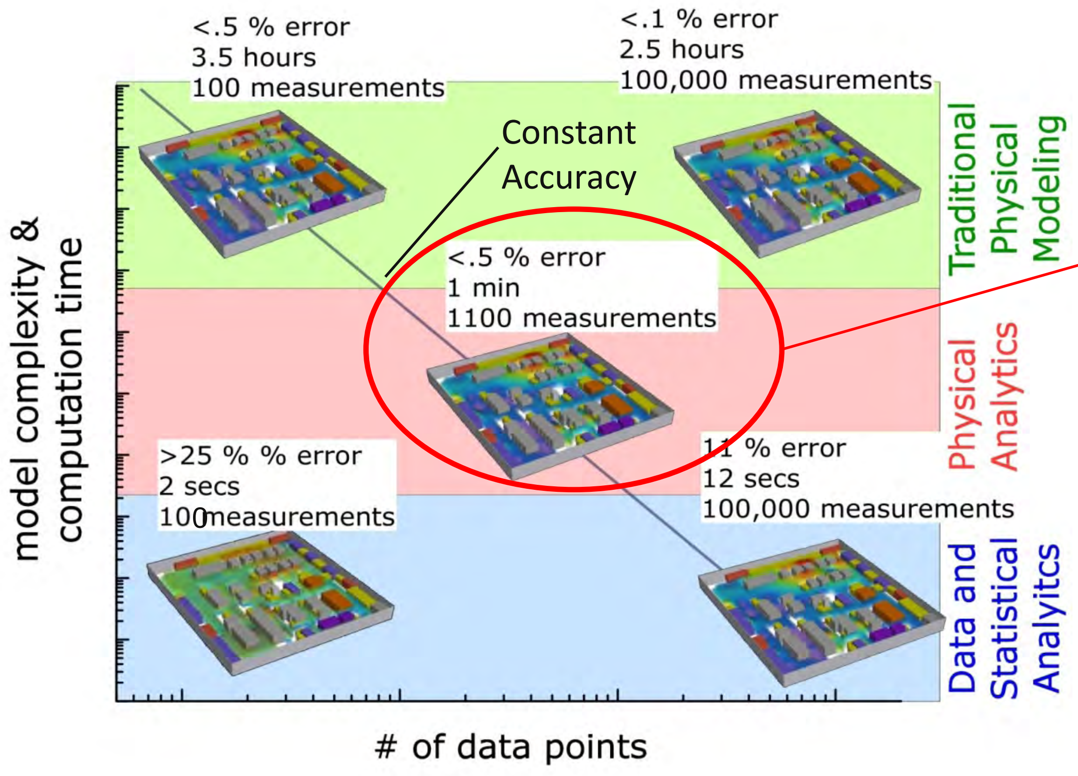
Ground truth established with more than 400,000 measurements

<u>*Robotic 3D temperature mapping tool*</u>

<u>*3D temperature results*</u>



RMS errors using different modeling approaches

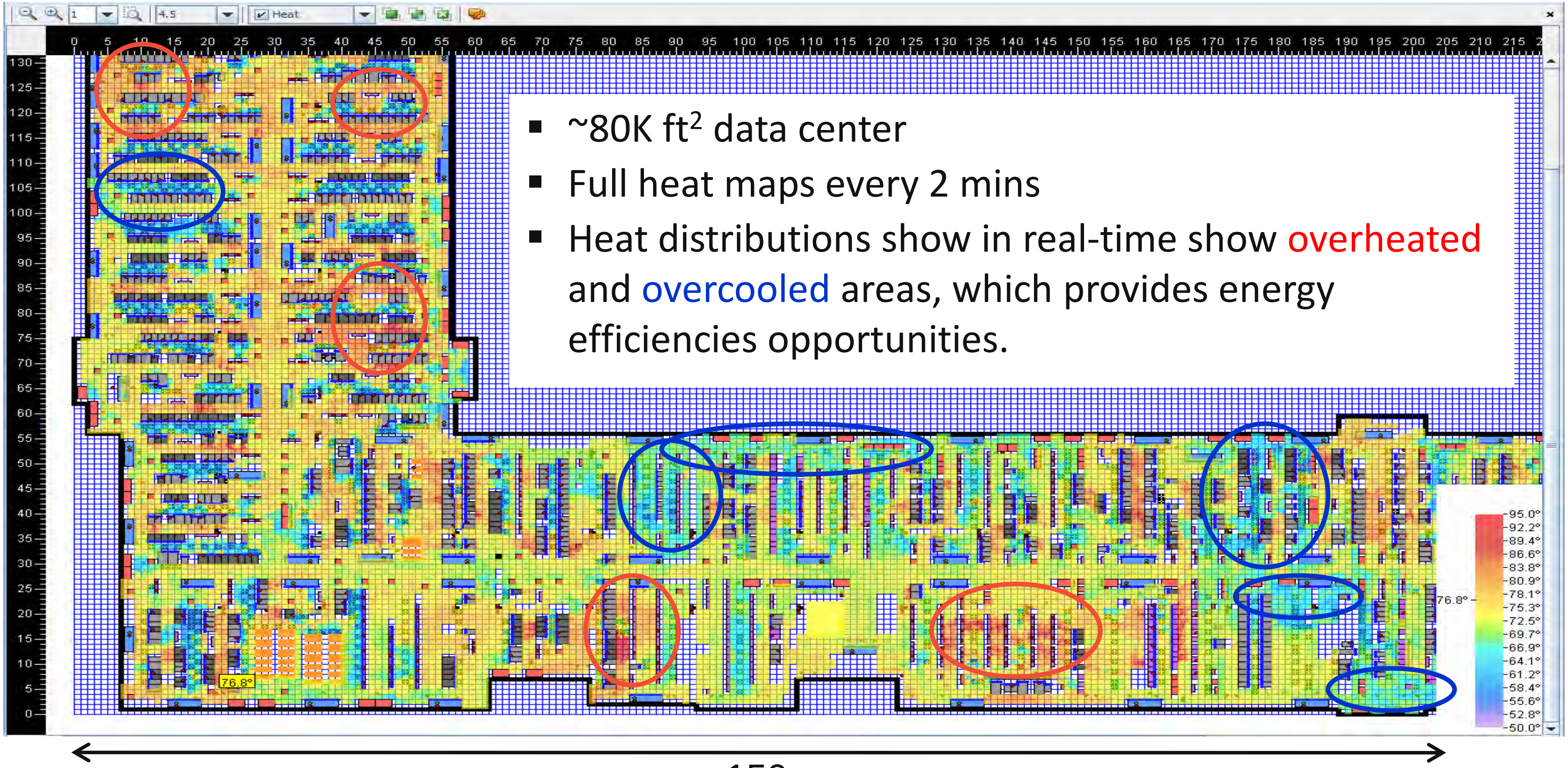


Model complexity can be very effectively reduced and computation time increased using sensor data as boundary conditions.

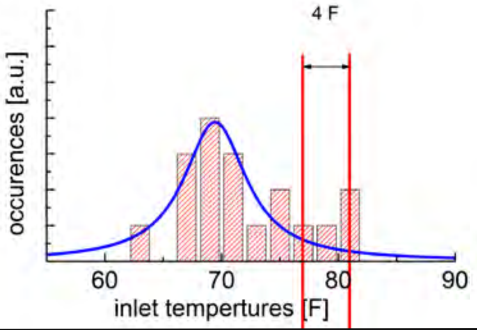
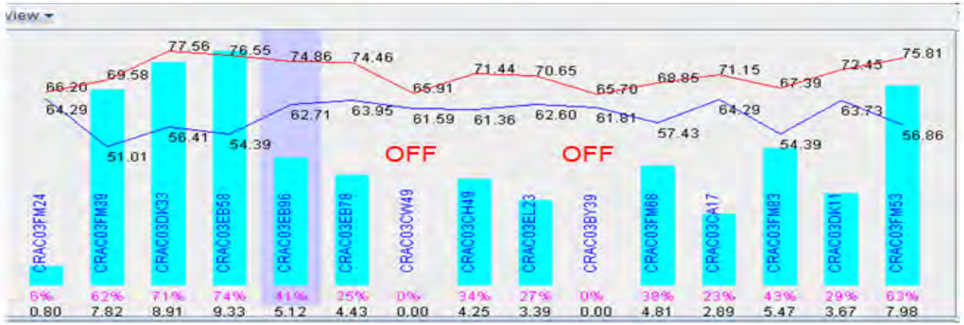
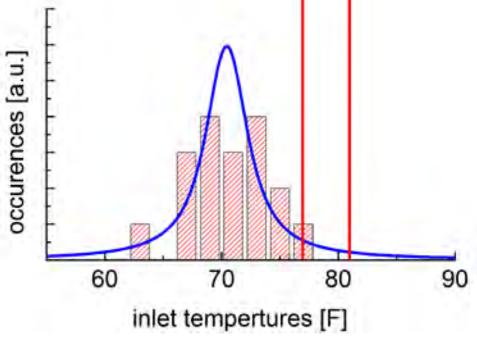
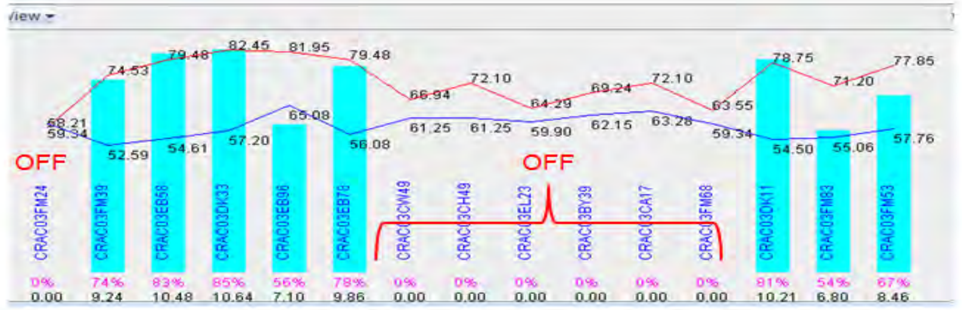
All modeling with the same computational infrastructure

Operational CFD for (near) real-time heat distributions

- ~80K ft^2 data center
- Full heat maps every 2 mins
- Heat distributions show in real-time show overheated and overcooled areas, which provides energy efficiencies opportunities.



Automatic controlling of cooling saves energy

	Temperature Distributions	Cooling Unit Status and Utilizations
Before: 9 am		
After: 5 pm		

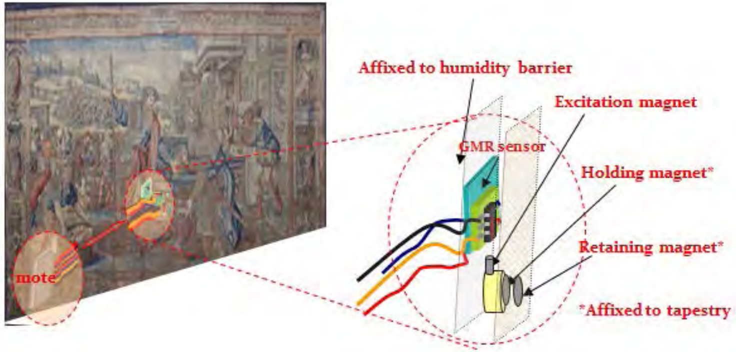
To date: > 900 M kWhour annually savings in WW deployments

Extending the technology to smarter buildings…. museums, infrastructure, hospitals, teleco facilities

<u>Environmental sensing in Late Gothic Hall @ NY Metropolitan</u>

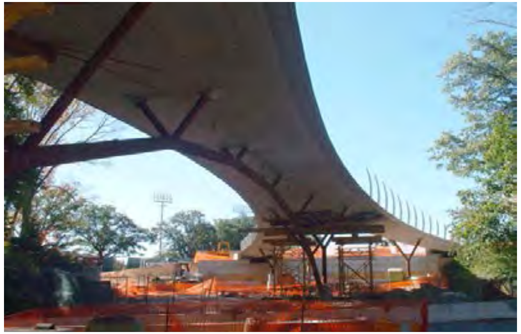
MET – Large mote network monitors temperature, humidity, light levels, dew point, numbers of visitors in galleries to optimally control HVAC system.

<u>Displacement sensors for tapestries @ Hampton Court, London</u>



Hampton Court – Displacement sensors measure additionally environmental response of works of art to control numbers of visitors in the galleries.

<u>Structural health monitoring @ Princeton University</u>



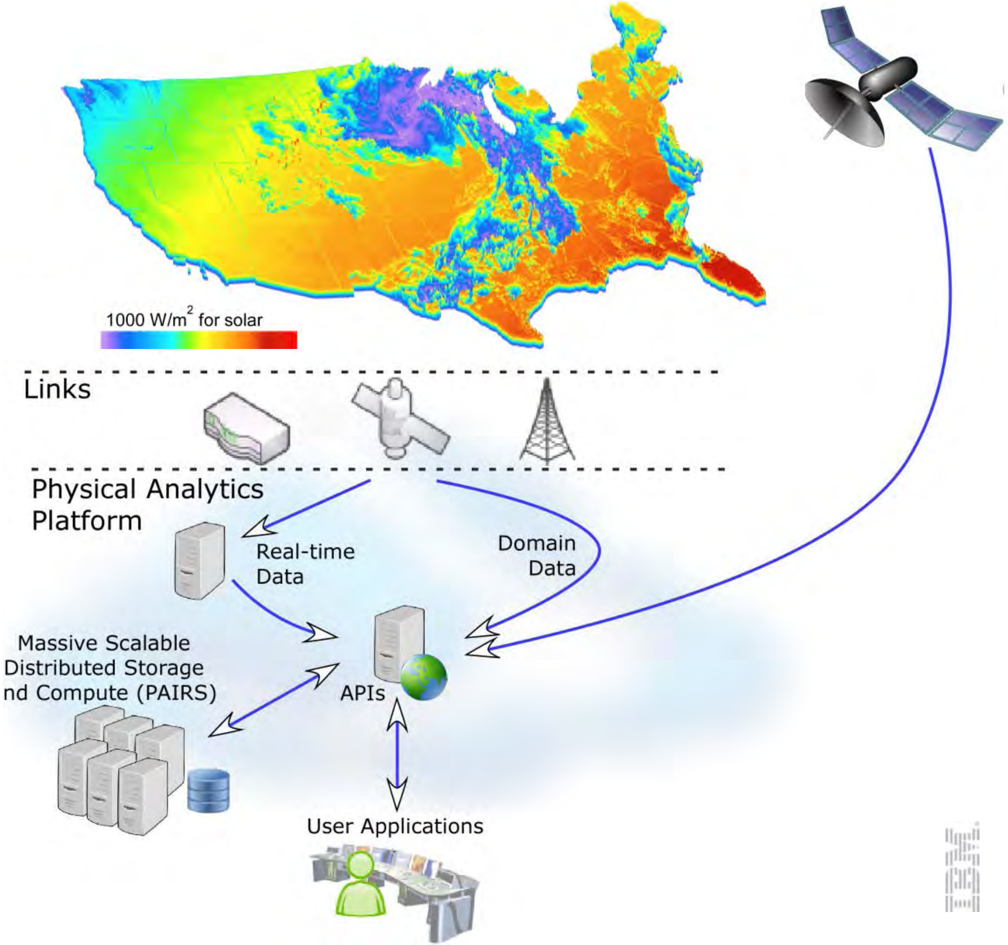
Princeton – Embedded sensors monitor vibration and internal corrosion inside of a bridge.

Forecasting renewable energy generation

Partners: NREL, ANL, Northeastern, ISO-NE, DoE-Sunshot

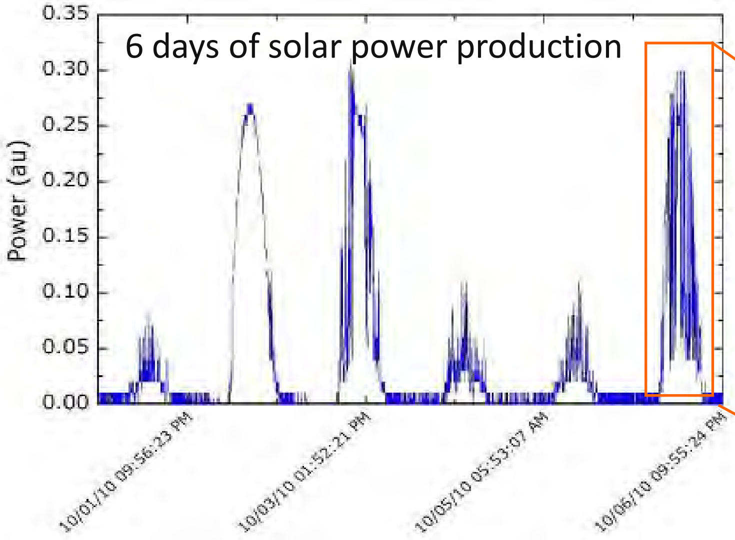
PA platform applied to forecasting and managing of renewable energy

- Sensor / Actuators
- Communication, Security
- Data Management
- Cloud-based SW Platform
  - Server, Agents, Apps, Clients
- Physical analytics
  - Physical models
  - Reduction of Model Complexity
  - Machine-learnt, multi-model model blending
- (Big) data analytics platform
- Automation and controls



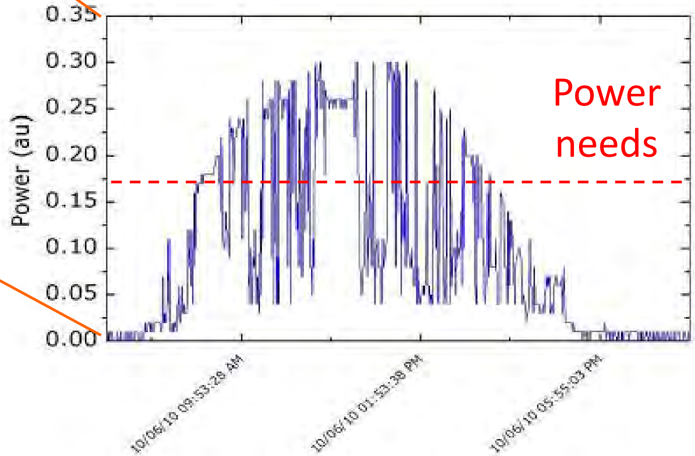
Intermittency of renewable energy poses a major problem for the power grid

<u>*Example: Solar*</u>



Solar power variability:

- Minute by minute fluctuating
- Mainly affected by local cloud coverage



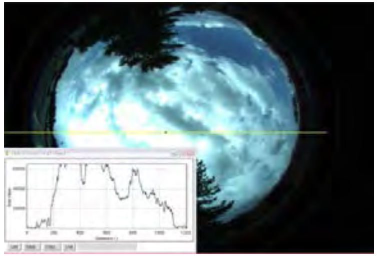
Forecasting and managing of renewable energy are critical for stability of the power grid.

Short-term forecasting with a sky-camera

IBM Cloud Imaging System without mechanical shutter

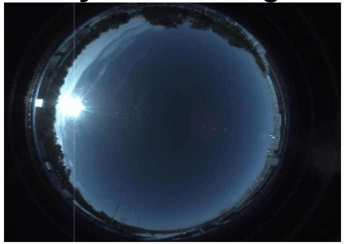
"Fish eye" lens

24 bit camera with several gain stages

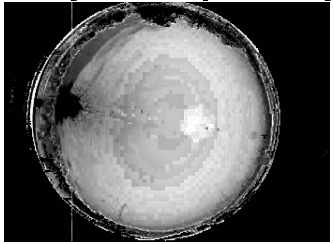


- Sky camera with fish eye lenses detects arrival incoming clouds
  - Field of view ~ 2 miles, no mechanical parts
- Multiple sky cameras increases prediction horizons and allow cloud height detection

SkyCam Image

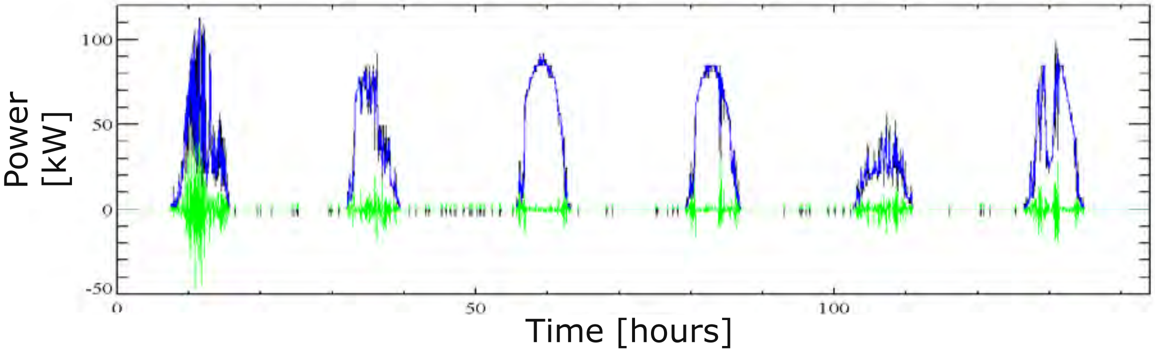


Sky Transparency



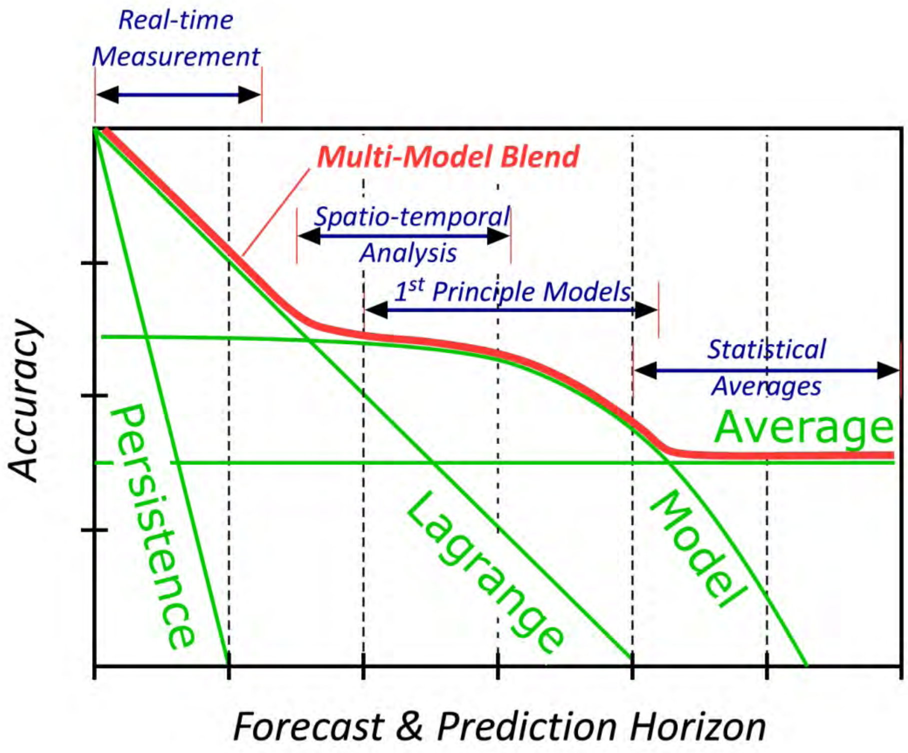
Measured power
Forecasted power
Difference / Error

6 min forecasts for 6 consecutive days



Expanding the forecast horizon - Big Data from various sources and models

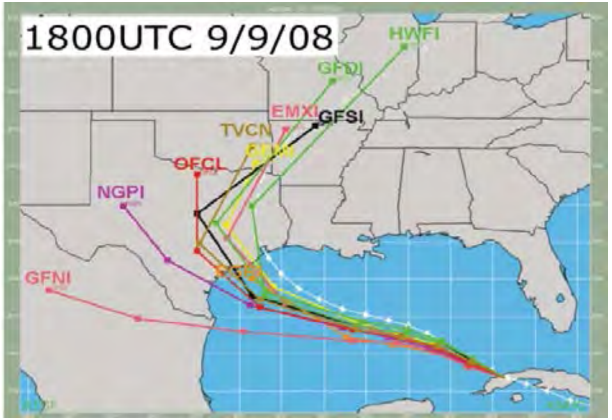
- **Persistence:**
  - Real-time power data
  - Weather station data
- **Lagrangrian Forecast Models:**
  - Sky camera model
  - Satellite-based (GOES), advection models
  - Time-series models
- **Weather Forecast Models:**
  - Rapid Refresh (RAP)
  - Hi-Resolution Rapid Refresh (HRRR)
  - Short-Range Ensemble Forecast (SREF)
  - North American Mesoscale Forecast (NAM
  - Global Forecast System (GFS)
  - European Center for Medium range Weat
    Forecasting (ECMWF)
- **Climate Models:**
  - Climate Forecasting System (CFS)



Key Idea: Situation-dependent, machine-learning based multi-model blending creates a “super” model

- Different forecasting models provide varying accuracies depending on time horizon, location, weather situation etc.
- Apply deep machine learning / “adaptive mixture of experts” techniques to dynamically blend different forecasts as a function of time horizon, location, weather situation etc
- System continuously learns, adjusts and improves.

Hurricane Ike path forecasts from 9 different weather models*



*M.J. Brennan, S.J. Majumdar, Weather and Forecasting 26, 848 (2011)
An Examination of Model Track Forecast Errors for Hurricane Ike (2008) in the Gulf of Mexico

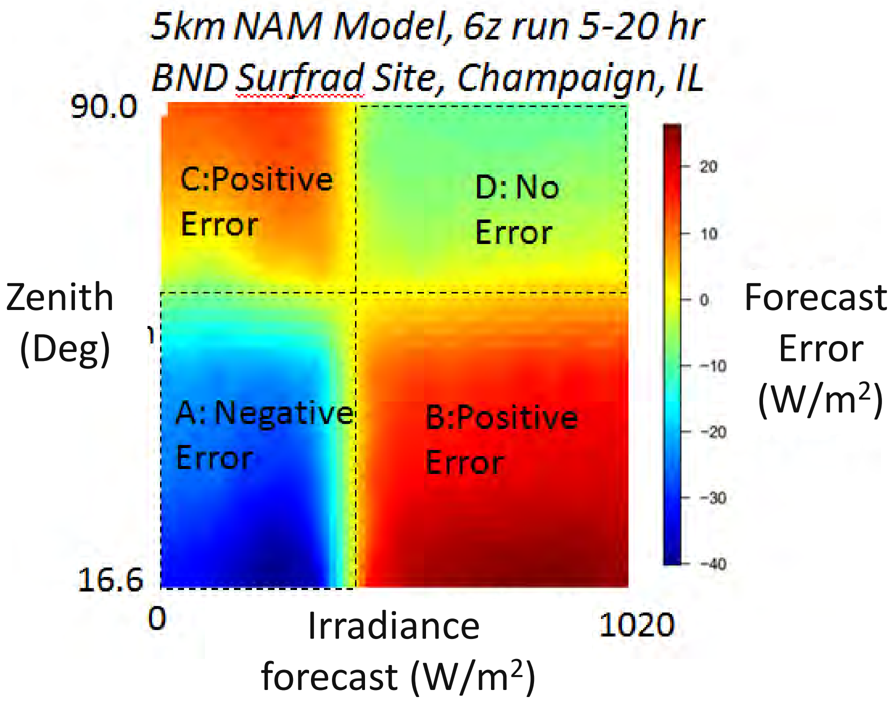
Improving accuracy using situation dependent, machine-learnt, multi-model blending

Question: **Which model is more accurate, when, where, under what weather situation?**

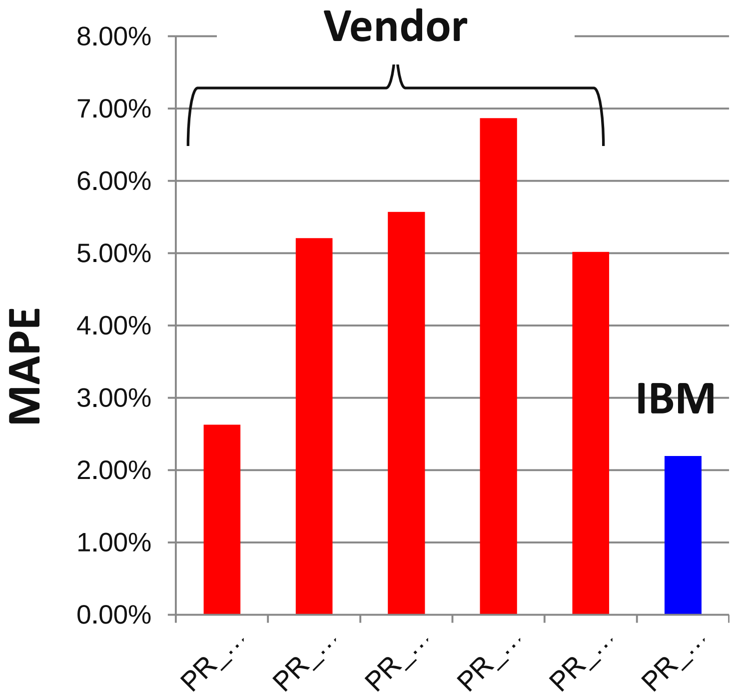
- Apply functional analysis of variance to understand 1st,2nd,3rd, ….order errors
- Model accuracy can depend strongly on **“weather situation”** category.
- “Weather situation” is determined using a set of parameters including forecasted ones on which model error depends on strongly.

Example, NAM solar irradiance forecast

- Depends strongly GHI and solar zenith angle.
- The two parameters create four categories of situations below.



High forecasting accuracy using situation dependent, machine-learnt, multi-model blending

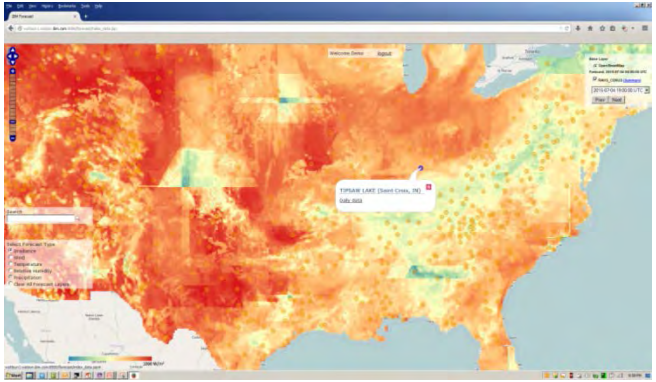


Italy; 1.935 MW Fixed Array

- In average more than 35% more accurate than next best input model
- Forecast for wind, solar, hydro from minutes to months ahead
- Regional, local and probabilistic forecasts

Extending the technology to nation-wide irradiance forecasting, the world solar car race, and pollution management

Irradiance forecast portal



- Daily nation-wide irradiance forecasts with real-time validation at over 1600 sites

Forecasting for solar car

- Provided highly accurate forecast for world solar car race for U Michigan in 2015
- U Michigan achieved its best performance in 15 years of racing

PM2.5 Sensor & Forecast portal

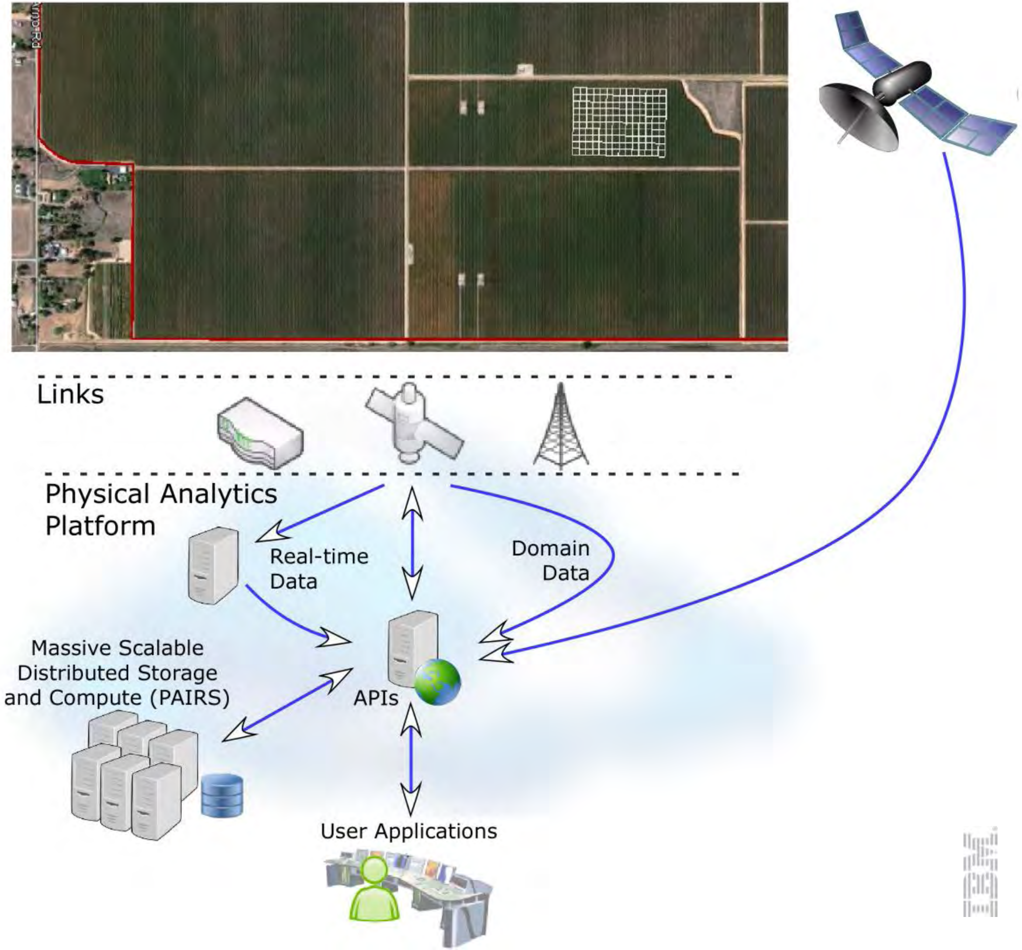
- Pollution forecast system developed for Beijing leveraging IBM's PM2.5 sensors.

Precision Agriculture

Partners: Gallo Wineries, Netafim, KSU

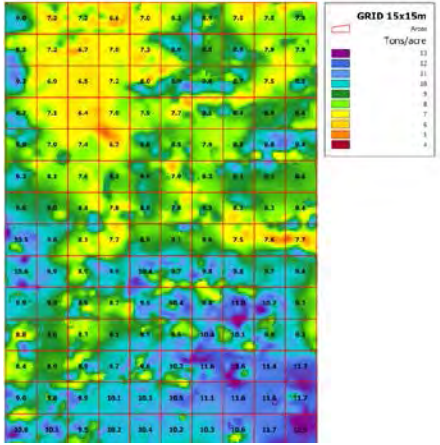
PA platform applied to precision agriculture

- Sensor / Actuators
- Communication, Security
- Data Management
- Cloud-based SW Platform
  - Server, Agents, Apps, Clients
- Physical analytics
  - Physical models
  - Reduction of Model Complexity
  - Machine-learnt, multi-model model blending
- (Big) data analytics platform
- Automation and controls



Spatio-temporal intra-field variability limits yield, water efficiency and quality

10 acre yield map

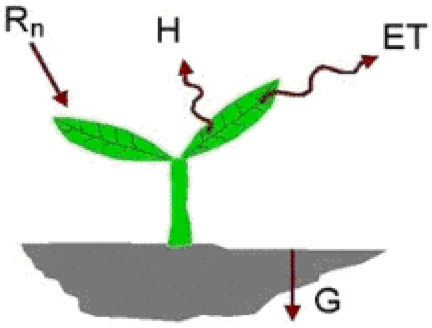


Harvester with yield measurement and GPS

- Large intra-field variability (measured during harvest) “limits” total crop yield
- Changes by 2x within less of 20 meters

If the low performing parts of field can be improved to the “current” average, then yield, water efficiency and quality can be drastically improved.

Evapo-transpiration modeling enables optimal irrigation

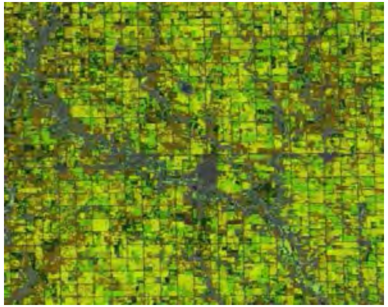


Energy Balance model:

$$ET = R_n - H - G$$

ET-Evapo transpiration
R_n-Net radiation Flux (W/m^2)
H-Sensible heat Flux (W/m^2)
G-Soil heat Flux (W/m^2)

Vegetation index from different satellite bands



Net Radiation:

$$R_n \approx (1-\alpha)R_s + (\varepsilon L_{in} - L_{out})$$

L_{in} incoming long wave radiation
L_{out} outgoing long wave radiation
R_s solar radiation
ε emissivity
a surface albedo

Sensible Heat Flux

$$H \approx \rho_{air} c_p (a + bT_s)/r_{ah}$$

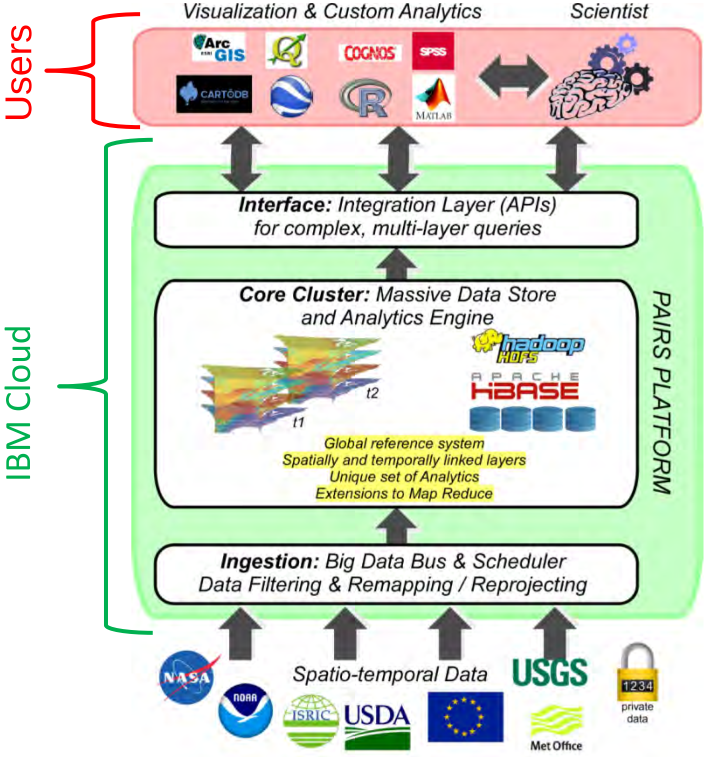
r_{air} density
c_p specific heat
a,b specific parameters
T_s surface temperature
r_{ah} transfer resistance

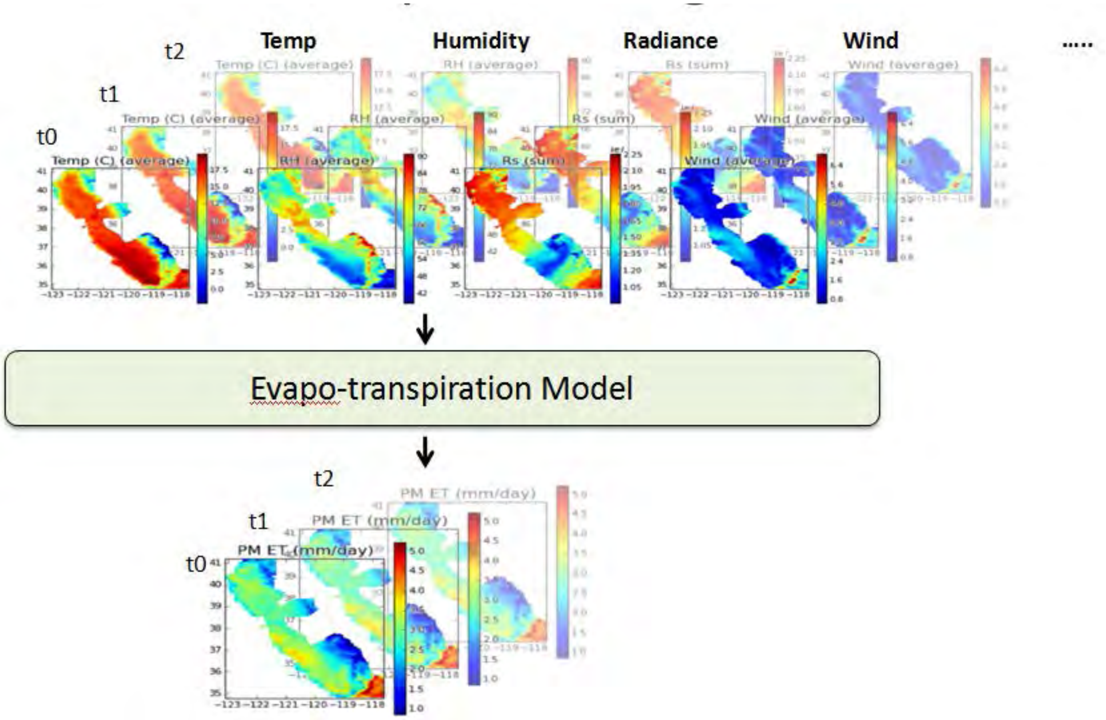
Soil Heat Flux

$$G \approx T_S(a + b\alpha)(1 - cNDVI^4)R_n$$

NDVI vegetation index
a,b,c specific parameters

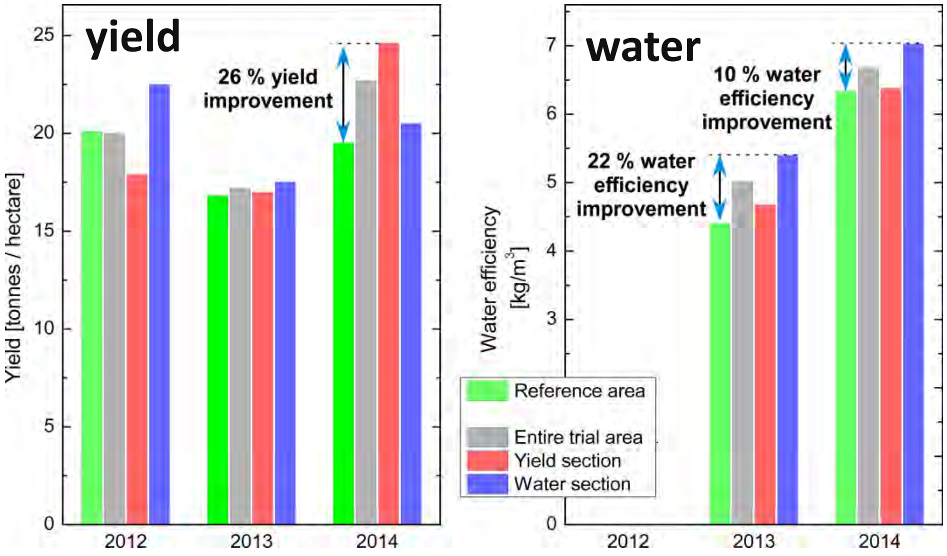
Multiple data sources are fused in big geo-spatial data platform to support scalable physical analytics

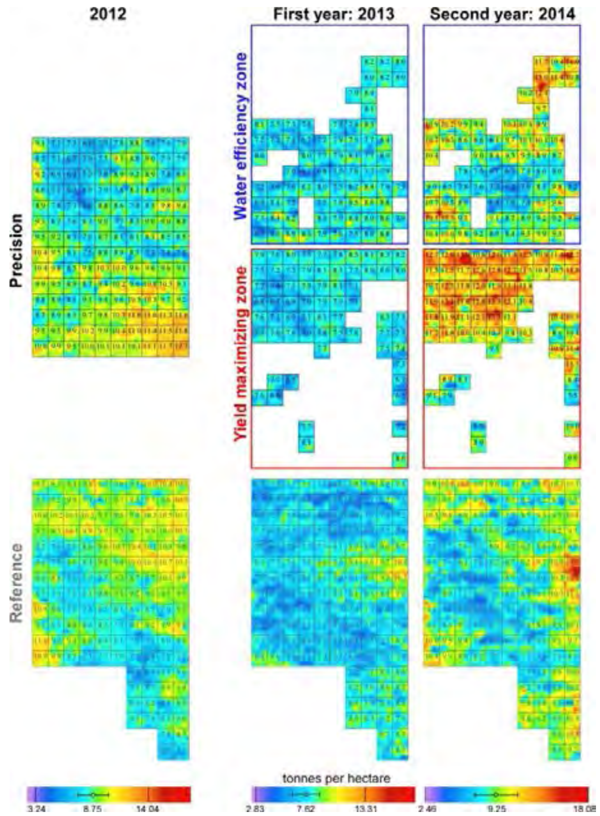




  27 April 2016

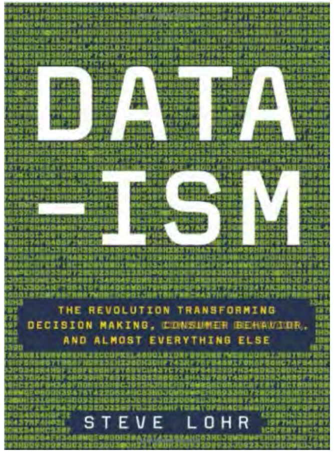
Yield maps after 2 years of closed loop precision irrigation show significant improvements





- 26% more yield
- 10-22 % higher water efficiency
- 50 % higher uniformity
- 2x improved quality index (Brix value)*

More about IBM's precision agriculture work



CONTENTS

FORTUNE

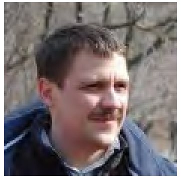
How IBM is Bringing Watson to Wine

VINTAGE REPORT • 2014 • INNOVATION AWARD

This year's winner was a collaborative experiment by **E. & J. Gallo and IBM**, whose approach used a variable-rate irrigation system across separate quadrants of a 31-acre Cabernet Sauvignon vineyard. The result decreased vineyard spatial variability and increased water-use efficiency without compromising quality during a period of historic drought.

Physical Analytics @ IBM Research


Conrad Albrecht
(Physics and Computation
Heidelberg PhD)


Levente Klein
(Physical Modeling)


Fernando Marianno
(Software Architect)

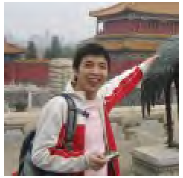

Michael Schappert
(Embedded System)


Marcus Freitag
(Precision Agriculture)

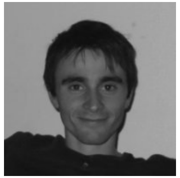

Hendrik Hamann
(Physical Analytics)


Vanessa Lopez
(Mathematics,)


Xiaoyan Shao
(Electrochemistry,
Data scientist)


Jun Song Wang
(Sensing)


Siyuan Lu
(Machine learning)


Bertrand Marchand
(Solar Forecasting)


Wang Zhou
(Robotics, drones)


Josphine Chang
(Sensor Platform, IoT)


Ramachandran Muralidhar
(Corrosion Science &
Pollution Modeling)


Theodore van Kessel
(Oil and Gas &
Instrumentation)


Bruce Elmegreen
(Astrophysics, Traffic)


Golnaz Badr
(Precision Agriculture)


Oki Gunawan
(Solar and Robotics)

and many more

Conclusion

Data from the physical world is growing faster than any other data source and will be soon the biggest data set.

Combination of big data and physical modeling provides unique opportunities:

- Simplifying and operationalizing physical models (example: building energy efficiency)
- Creating “super-models” to provide deeper understanding of physics/chemistry of models (example: renewable forecasting)
- Providing science-based decision support (example: precision agriculture)
- ….

