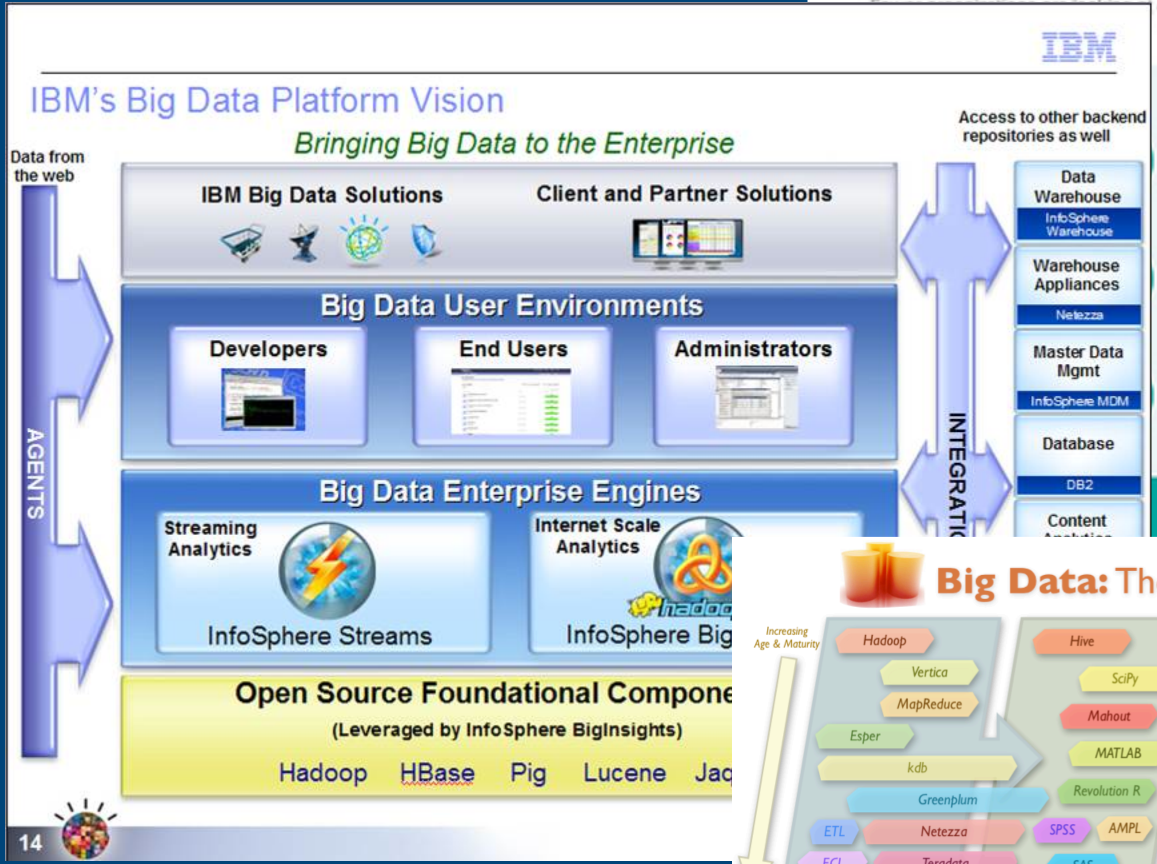
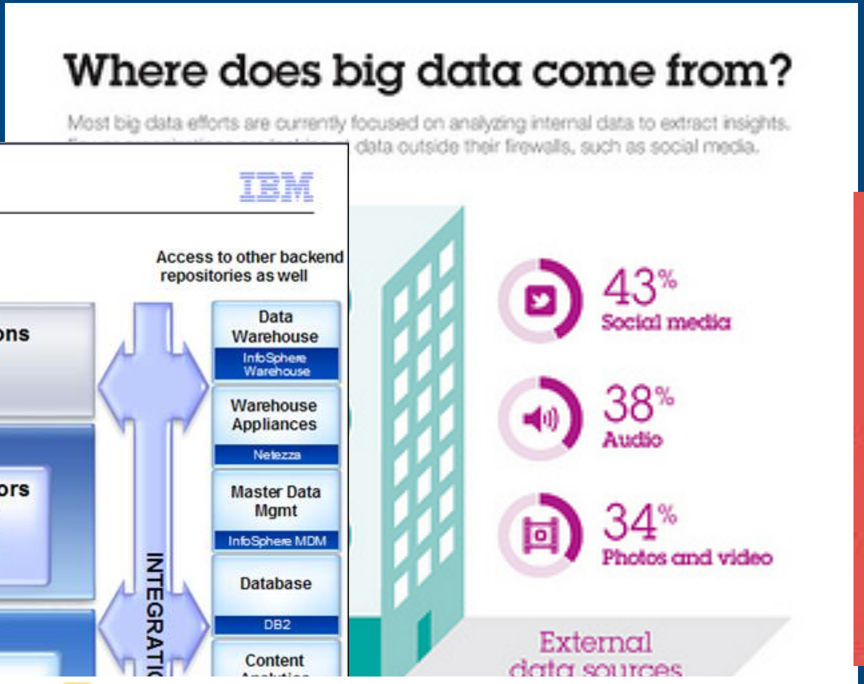
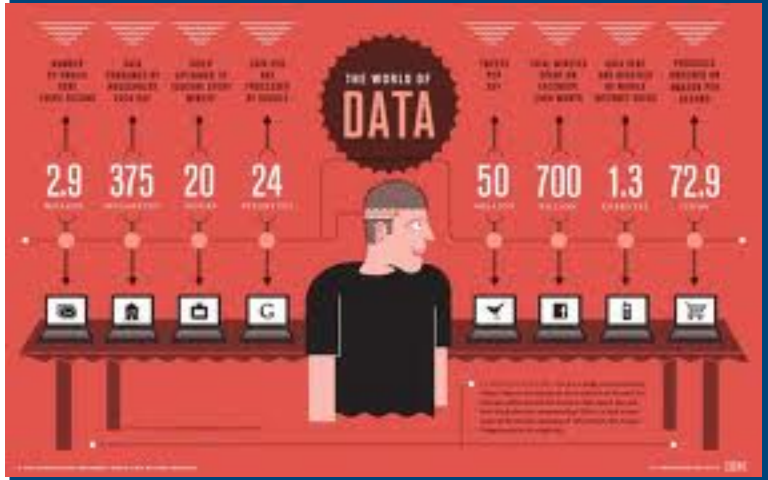
The Role of Models in Smarter Cities Solutions

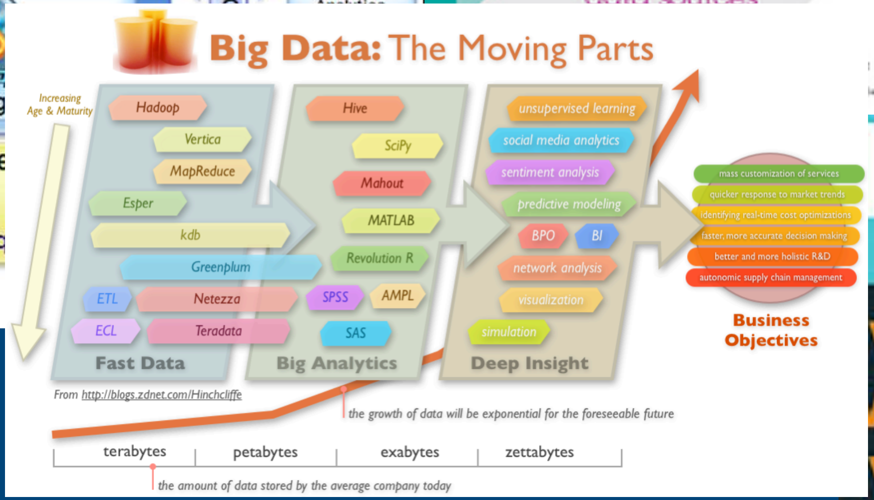
Sean A. McKenna, PhD, IBM Research, Dublin Ireland, seanmcke@ie.ibm.com

Big Data


IBM
IBM's Big Data Platform Vision
Bringing Big Data to the Enterprise
Data from the web
AGENTS
IBM Big Data Solutions
Client and Partner Solutions
Big Data User Environments
Developers
End Users
Administrators
Big Data Enterprise Engines
Streaming Analytics
InfoSphere Streams
Internet Scale Analytics
InfoSphere Big
Open Source Foundational Compone
(Leveraged by InfoSphere BigInsights)
Hadoop
HBase
Pig
Lucene
Jaq
Access to other backend repositories as well
Data Warehouse
InfoSphere Warehouse
Warehouse Appliances
Netezza
Master Data Mgmt
InfoSphere MDM
Database
DB2
Content
INTEGRATI
14


Where does big data come from?
Most big data efforts are currently focused on analyzing internal data to extract insights.
data outside their firewalls, such as social media.
43% Social media
38% Audio
34% Photos and video
External data sources


THE WORLD OF DATA
2.9
375
20
24
50
700
1.3
72.9


Big Data: The Moving Parts
Increasing Age & Maturity
Hadoop
Vertica
MapReduce
Esper
kdb
Greenplum
ETL
Netezza
ECL
Teradata
Fast Data
Hive
SciPy
Mahout
MATLAB
Revolution R
SPSS
AMPL
SAS
Big Analytics
unsupervised learning
social media analytics
sentiment analysis
predictive modeling
BPO
BI
network analysis
visualization
simulation
Deep Insight
mass customization of services
quicker response to market trends
identifying real-time cost optimizations
faster, more accurate decision making
better and more holistic R&D
autonomic supply chain management
Business Objectives
From http://blogs.zdnet.com/Hinchcliffe
the growth of data will be exponential for the foreseeable future
terabytes
petabytes
exabytes
zettabytes
the amount of data stored by the average company today


BIG DATA?
VOLUME
Large amounts of data.
VELOCITY
Needs to be analyzed quickly.
VARIETY
zynga
YouTube
LOL
Worldwide IP traffic will quadruple by 2015.
By 2015, nearly 3 billion people
will be online, pushing the data created and shared to nearly 8 zettabytes.

Big Data

30 fps, ~50MB/minute, geo-located, network connected (social or provider)

8 Years!

All the data in the world cannot tell you what will happen next

“You cannot Google the future”

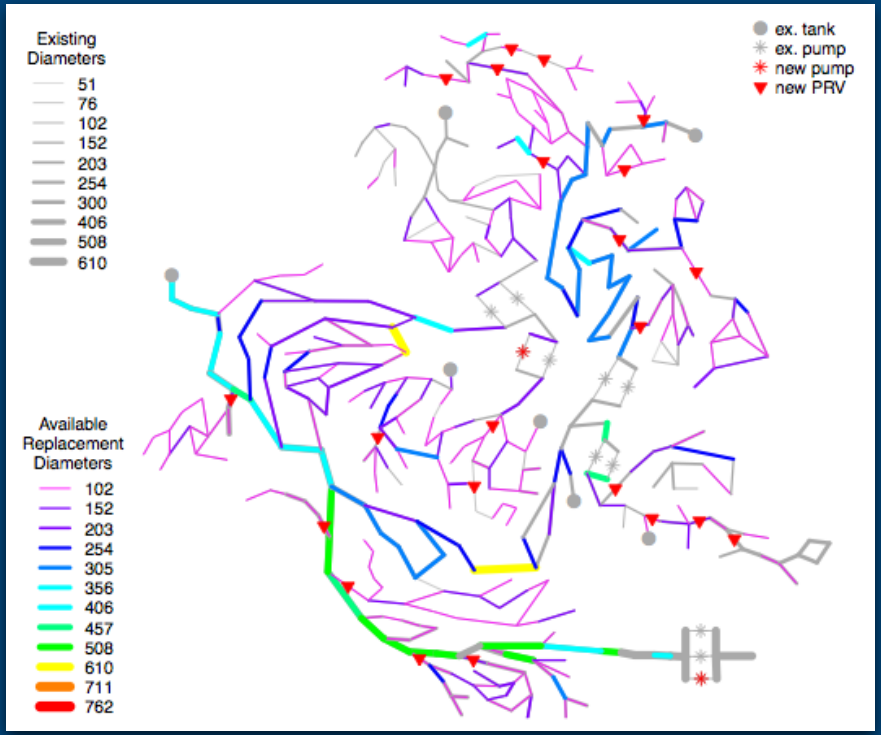
William Gibson

Models + Data

- Physical Processes: Calibrate model parameters using observed data to minimize difference between model outputs and observations
- Statistical: Rely on distribution theory to estimate/predict new observations from collected data (e.g., regression)
- Machine Learning: Non-linear process for learning from observations for classification or prediction

Example 1: Leak Localization

- Pipe networks supply fresh water to cities
- As much as 70% of that water is lost to leaks
  - –Costs of energy and of chemicals
- Traditional approach
  - –Listen for leaks during quiet times
- Newer approaches
  - –Acoustic sensors
  - –High frequency pressure waves
- *What about Data + Models* ?
  - –Pressure and Flow

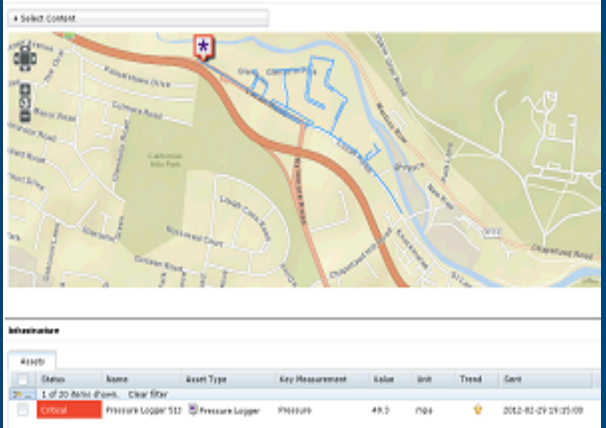


Leak Localization at sub-DMA Level

- Goal: localize leaks within a single District Metered Area (DMA)
  - –Medium resolution: Determine subset of DMA with leak(s)
  - –Use results to limit search area for crews/sensors with acoustic monitoring
- Approach:
  - –Use sensors (flow and/or pressure) within the DMA
  - –Use measured and estimated demands at network nodes
  - –Employ state estimation and correlation of residuals
- Requirements:
  - –Calibrated network model of DMA
  - –Continuous monitoring of flow and/or pressure data
  - –Measured or estimated demands at nodes
- Result:
  - –Determine subset of DMA for exhaustive search for leak

Finding leaks in Water Distribution Networks (WDNs)

- What has been **measured** & modeled?



- High-frequency sensors
- Large deployment
- Extensive coverage of WDN

Where to **dig**?

- ***Acoustic methods***
  - *High precision of detection and localization*
  - *Sensors must be relatively close to leak (depending on material)*
- ***Transient methods***
  - *Analyze hydraulic transients in high-freq signals*
  - *Suited for transmission (vs distribution) systems*

- Hydraulic model
- Demand profiles
- Low-freq hydraulic sensors
- Sparse coverage of WDN

•Where to **look**?

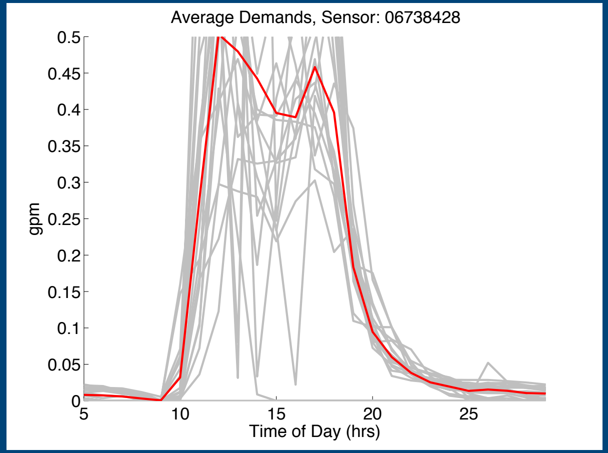
•http://www.vivax-metrotech.com/UploadFiles/image/HL10(2).JPG

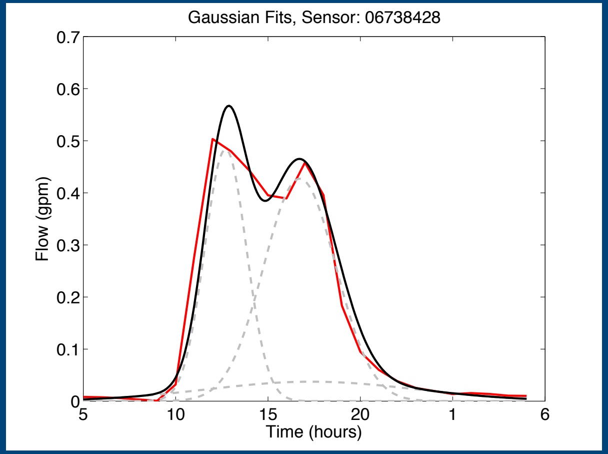
- ***Analytics and Optimization***
  - Use hydraulic model and pressure/flow sensors
  - Need knowledge of demands
  - Estimate leaks by matching model predictions with sensor readings

- Suspicious areas
- Targeted use of high-freq sensing equipment

Leak Localization

- Estimate demands at service connections
  - Census data, Billing records, Mapping data
- Use, or develop, calibrated network model
  - Total supply and demand amounts
  - Pressure and/or flow measurements
- State Estimation
  - Demands as the estimated state
- Residual Analysis
  - Large negative residuals indicate non-revenue water
  - Correlated residuals due to all demands impacted by leak
  - Cluster network into areas impacted by leak

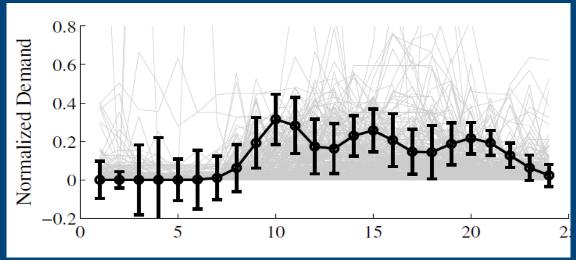




- Fusco, Eck and McKenna, 2014, Bad Data Analysis with Sparse Sensors for Leak Localisation in Water Distribution Networks, *International Conference on Pattern Recognition*
- Fusco, Eck and McKenna, 2014, Identifying leakage likelihood using state estimation and bad data identification methods, *Water Loss Conference*
- McKenna, Fusco and Eck, 2013*, Water Demand Pattern Classification From Smart Meter Data,* Computing and Control for the Water Industry

Our solution:
Demand Modeling, State Estimation and Residual Analysis (2)

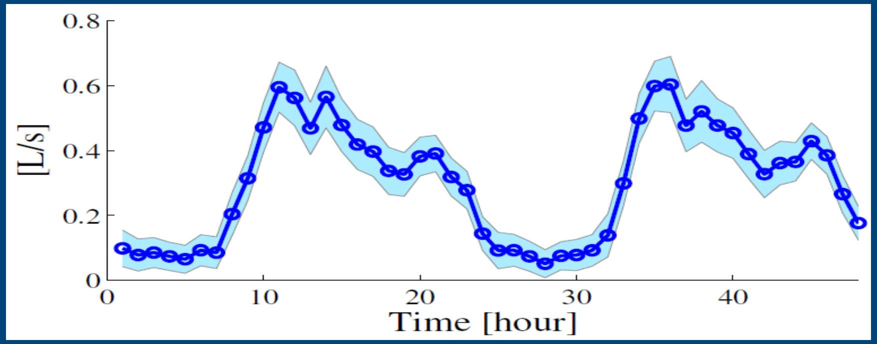
Network Model
Sensor Data
(pressures, flows,...)



State Estimation

$$\min_x J(x) = [y - h(x)]^T W [y - h(x)]$$

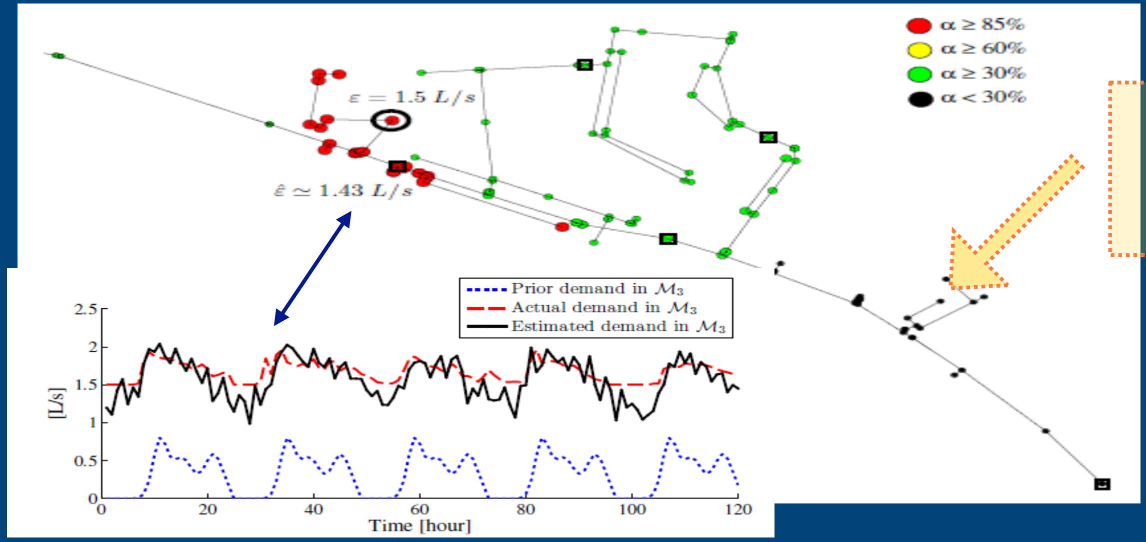
Estimated Demand



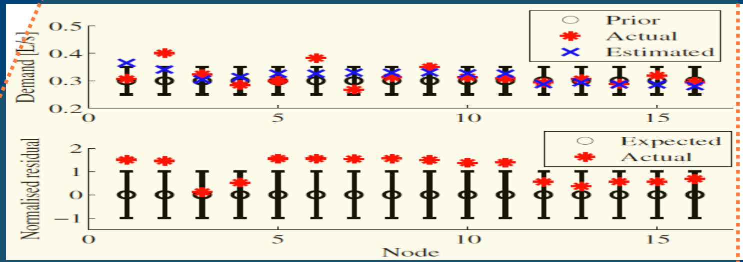
Demand Modeling
(Prior/expected demand)

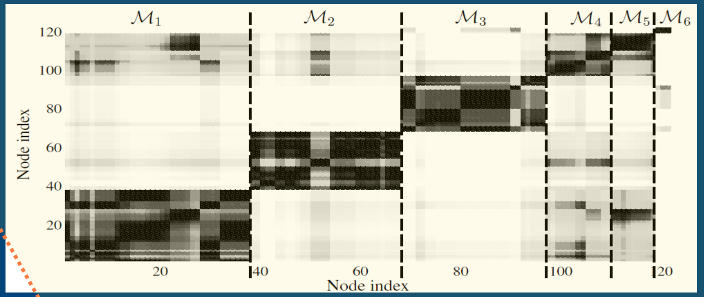
+
-

Demand Residuals



Statistical Analysis of Residuals



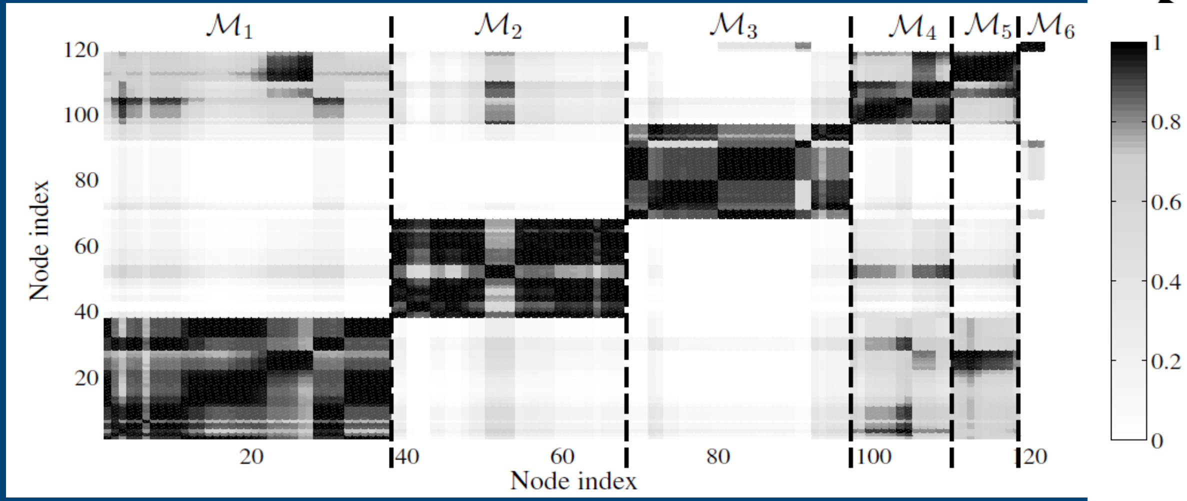


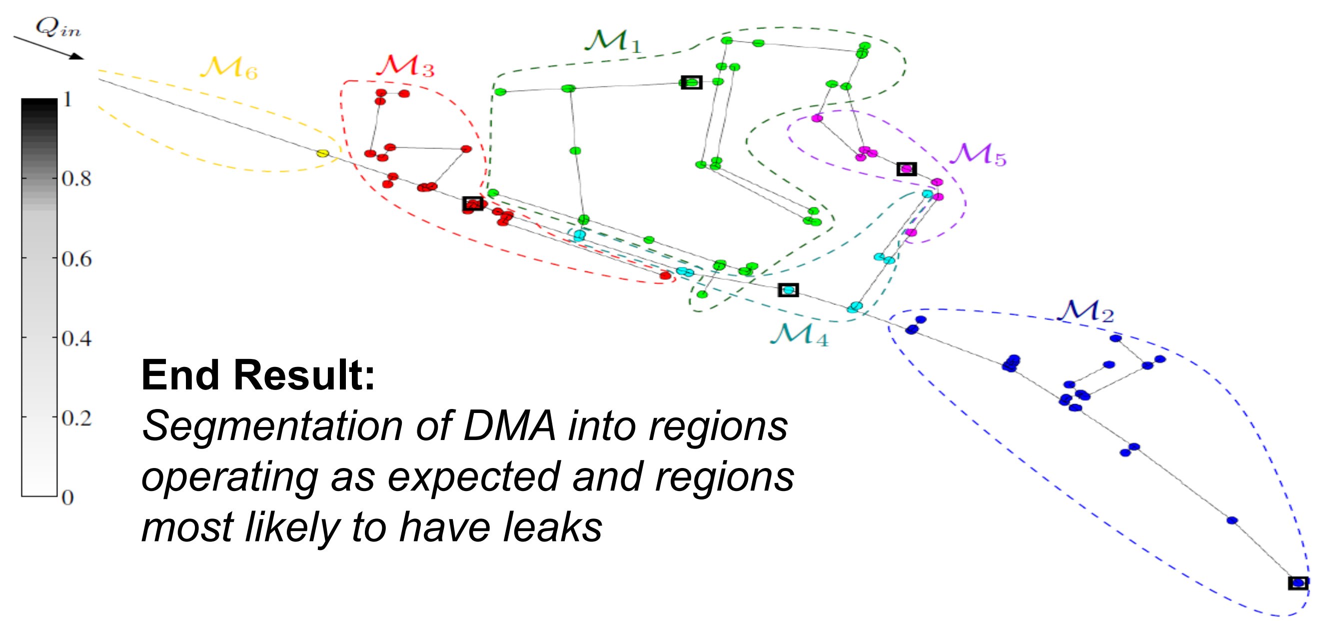
Residual Analysis (1): Identify network areas

- Residuals are difference between measured (expected) and estimated hydraulic quantities
  - Leakage produces ***demand residuals*** (“unexpected demand”) →

$$r \triangleq y - h(\hat{x})$$

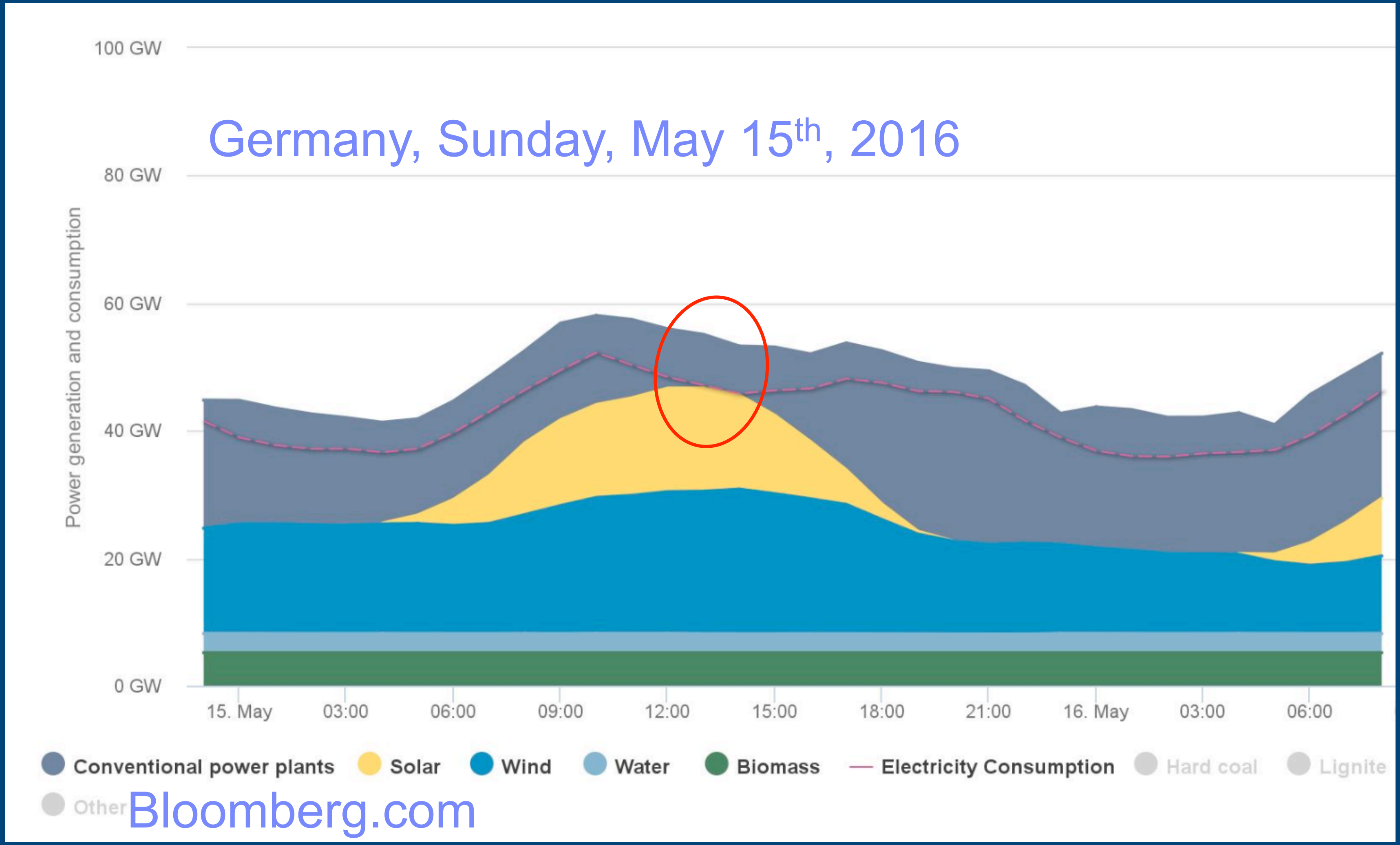
- Depending on the position/number of sensors and on the level of uncertainty
  - Groups of residuals may be strongly correlated
    - *Leaks at different locations look “similar” with respect to sensor data and produce “similar” residual patterns*
  - Advanced statistical analysis of residuals identifies these groups (network areas)





End Result:
Segmentation of DMA into regions operating as expected and regions most likely to have leaks

Example 2: Distributed Renewable Energy Forecasting



http://www.bloomberg.com/news/articles/2016-05-16/germany-just-got-almost-all-of-its-power-from-renewable-energy

Energy

"There will be more changes in the next 10 years than there have been in the last 100!

We are in an age of discovery; we must challenge the boundaries of prevailing wisdom!"

– Jeff Martin, CEO of San Diego Gas and Electric, DistribuTech keynote, February 2015

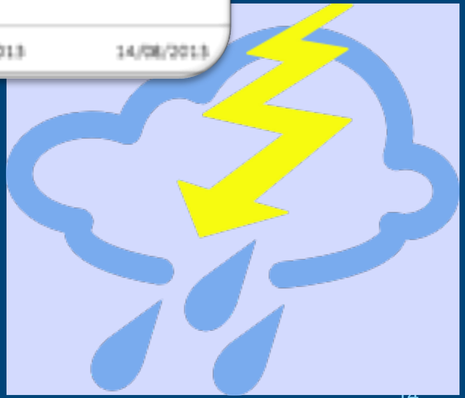
Outline

Trends

- Rise of Renewables
- IOT and Big Data
- Uncertainty

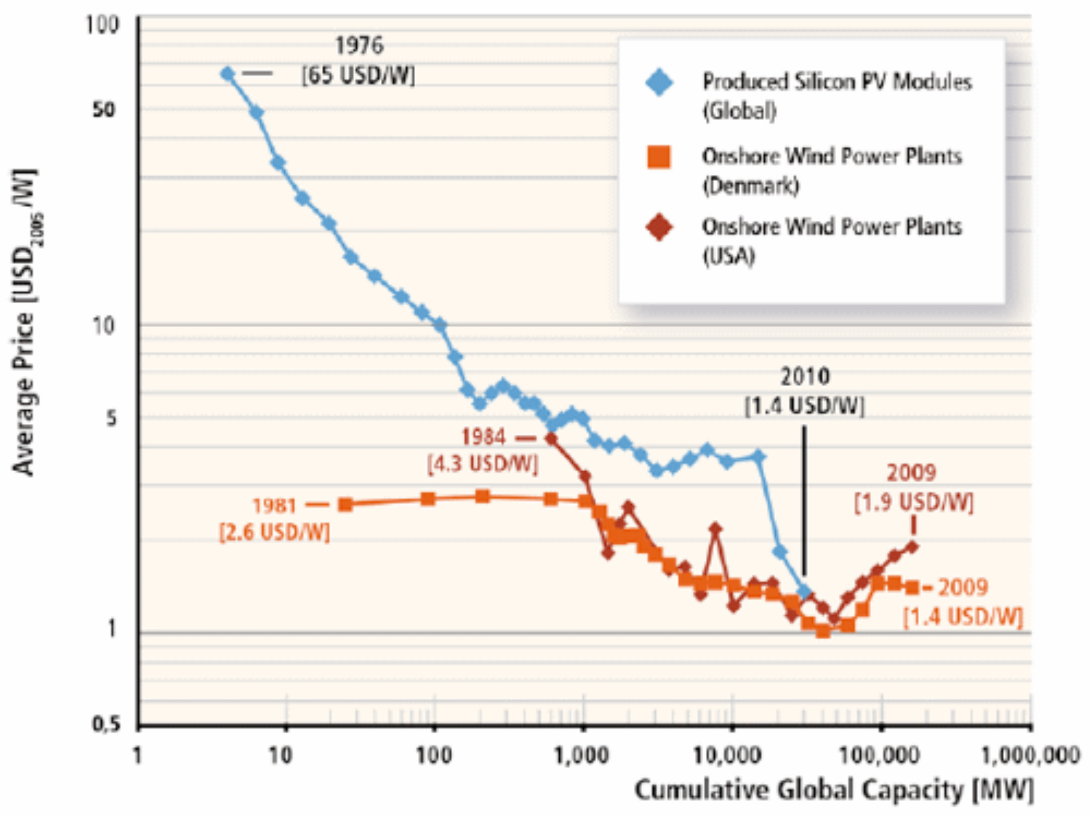
Examples of Data-Driven Operations

What’s next?



Rise of Renewables

- **Growth**. In the OECD countries renewables account for virtually all net additions to power capacity
- **Solar**. Further decline in the cost of PV technology will drive a $3.7 trillion surge in investment in solar, both large-scale and small-scale.
- **Distributed**. Some $2.2 trillion of this investment will go on local PV systems, enabling consumers and businesses the ability to generate and store electricity



Sources:
- 2105 Bloomberg New Energy report: http://www.bloomberg.com/company/new-energy-outlook/
- **Executive Summary of IEA Renewable Energy Medium Term Market Report 2015**
- **Graphic: http://www.skepticalscience.com/ipcc-report-renewable-energy.html**

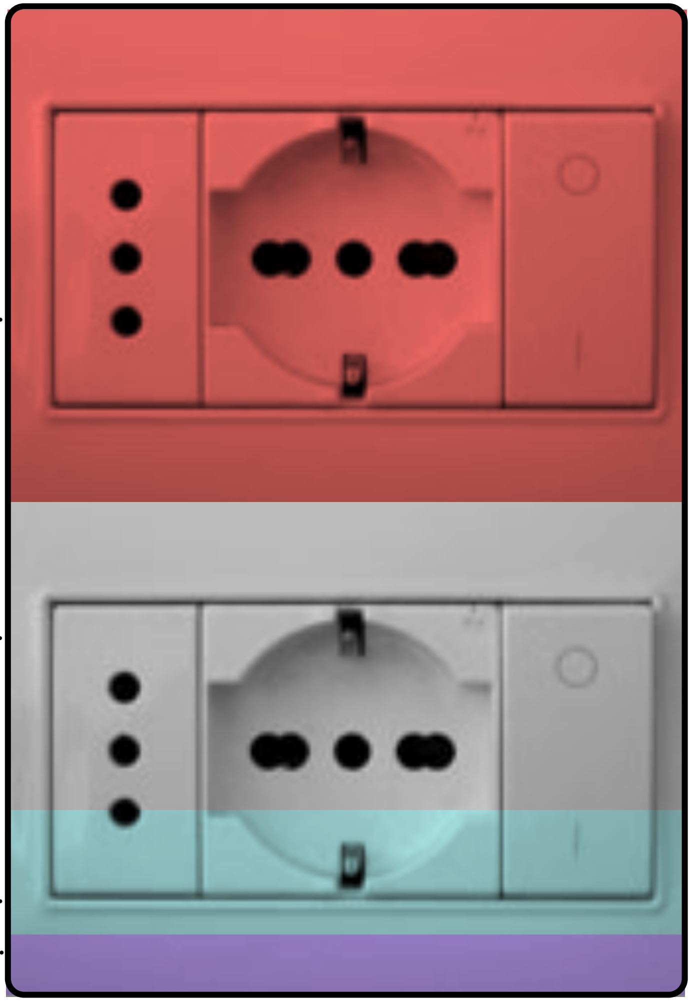
Data explosion dwarfs “traditional” data sources

Utility grid IoT data
200 Petabytes
Annually

Hyper-local weather data
+ weather-driven supply
and demand data
150 PB
Annually

Asset records
50 TB

Traditional customer records
5 TB



- Future business model depends on making use of this data rather than selling kWHrs
- We need demand to follow supply by exploiting flexibility in the demand

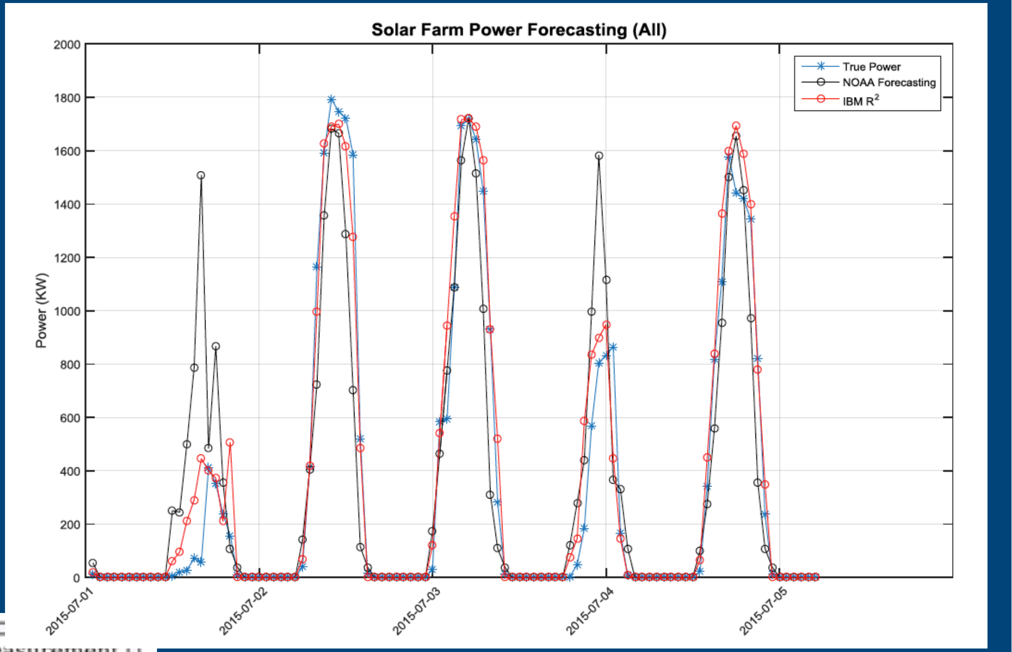
Uncertainty: Demand

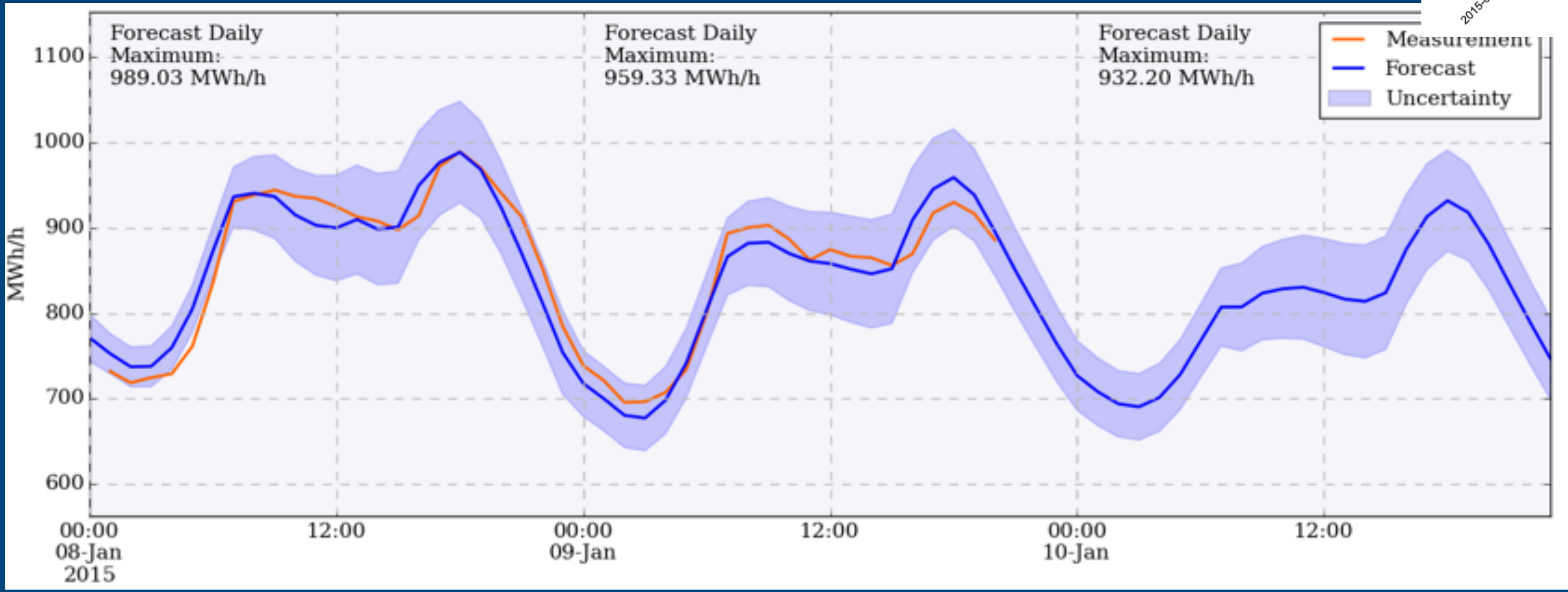
$$p(x) = \int e(t)\,dt \quad \Rightarrow \quad p(x) = \int e(t) \cdot c(t)\,dt$$

Increasingly distributed nature of electricity production and consumption leads to increasingly distributed (personalized) pricing

Uncertainty and Predictive Models

- For Distributions System Operators and Transmission System Operators, uncertainty in power generation and demands are significant under distributed generation scenarios
- Predictive models must also quantify uncertainty



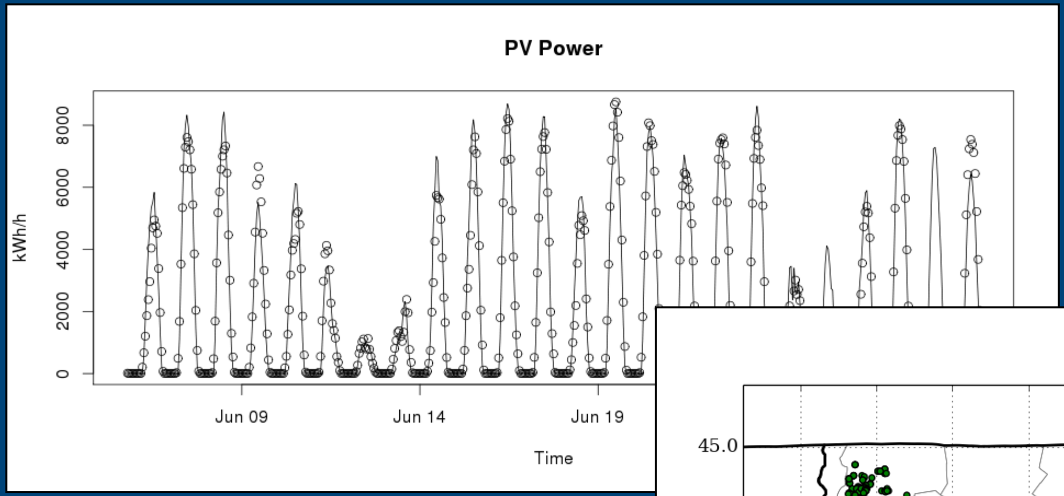


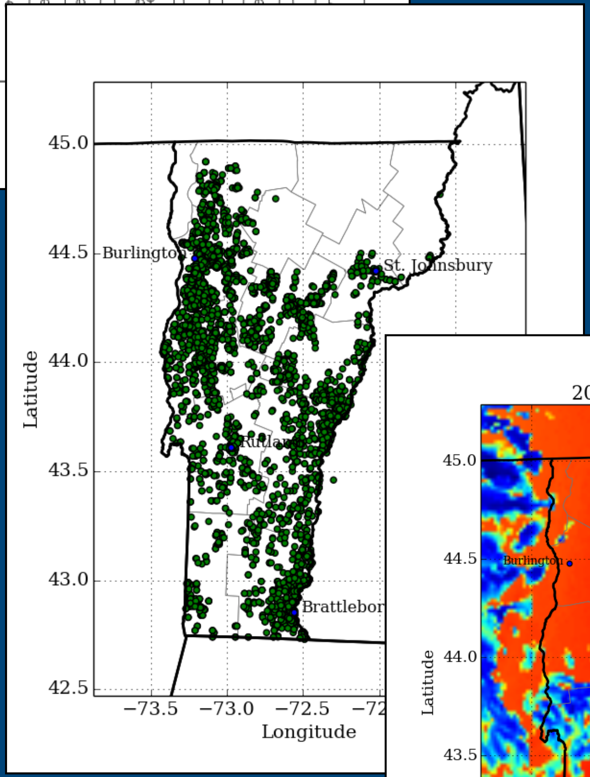
Example 2: Distributed Renewable Energy Forecasting

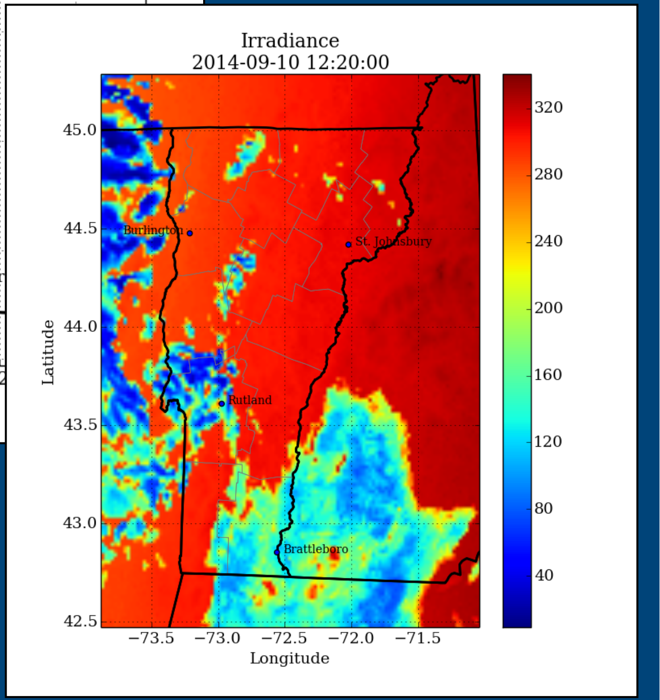
- Work with a regional transmission utility in the US
  - Serving 17 distribution utilities
- Strong government incentives to deploy rooftop solar (PV)
  - As much as 10 percent of electricity from solar (2015)

- Questions regarding rooftop PV
  - Where is it in the grid?
  - How much is there?
  - How much energy will be produced tomorrow?
  - What is the uncertainty in that forecast?

Solar Forecasting Example (State Level)





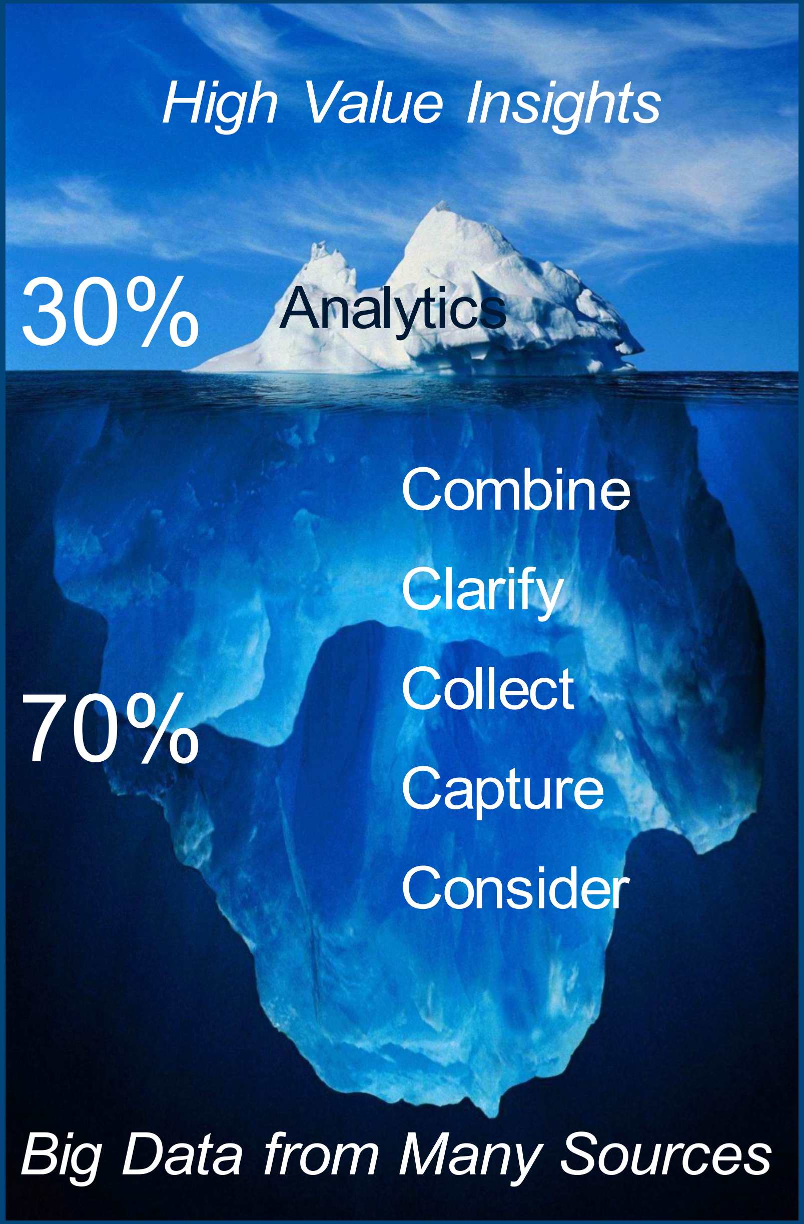


We combine physical models with machine learning to model PV generation at town, county and state level.

The physical model uses a solar position algorithm, irradiance and temperature data and an Irradiance-to-power model.

This physical model is combined with a Generalized Additive Model (GAM), trained on measured PV power data.

Data Driven



Data Curation

Bottom Up Approach

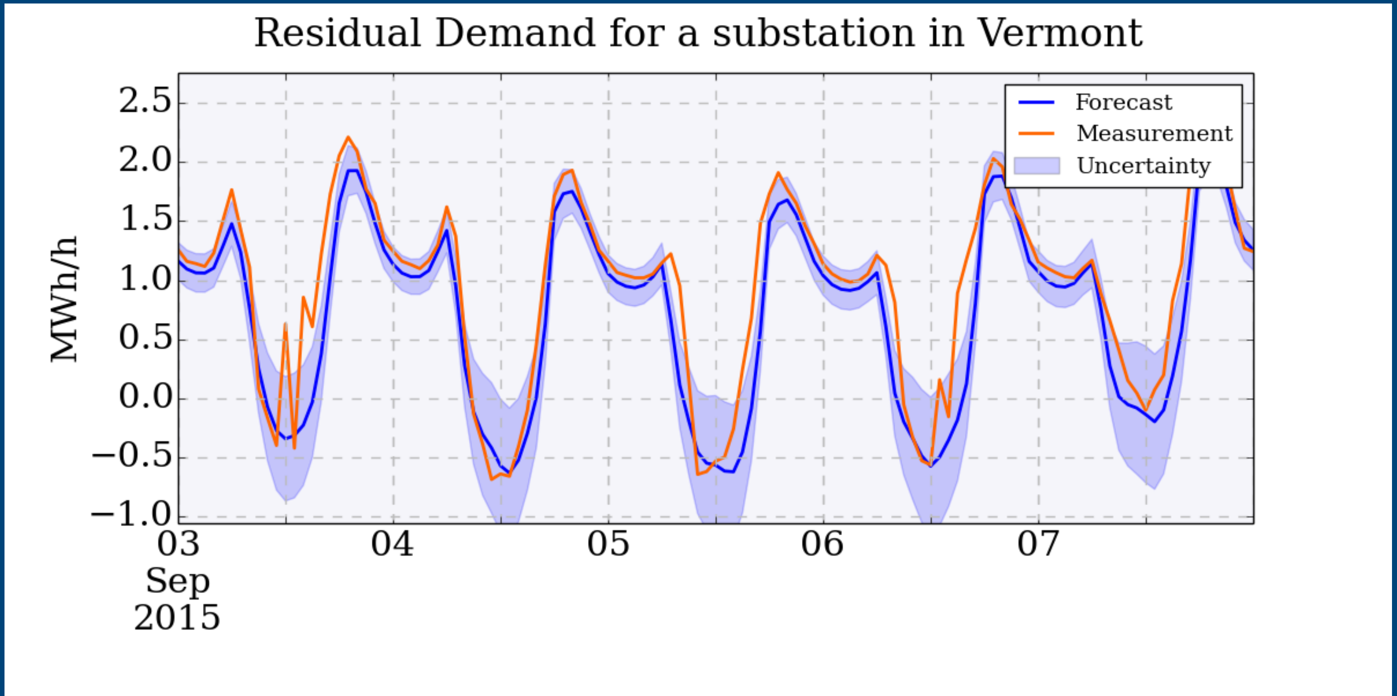
Predicting Demands (Substation Scale)

The contribution of PV generation can be significant at the substation level.

This can cause reverse power flows.

Accurately predicted by our models.

- *Distributed Generation and Demand*
- *Bottom Up*
- *Data Driven*
- *~1500 models*

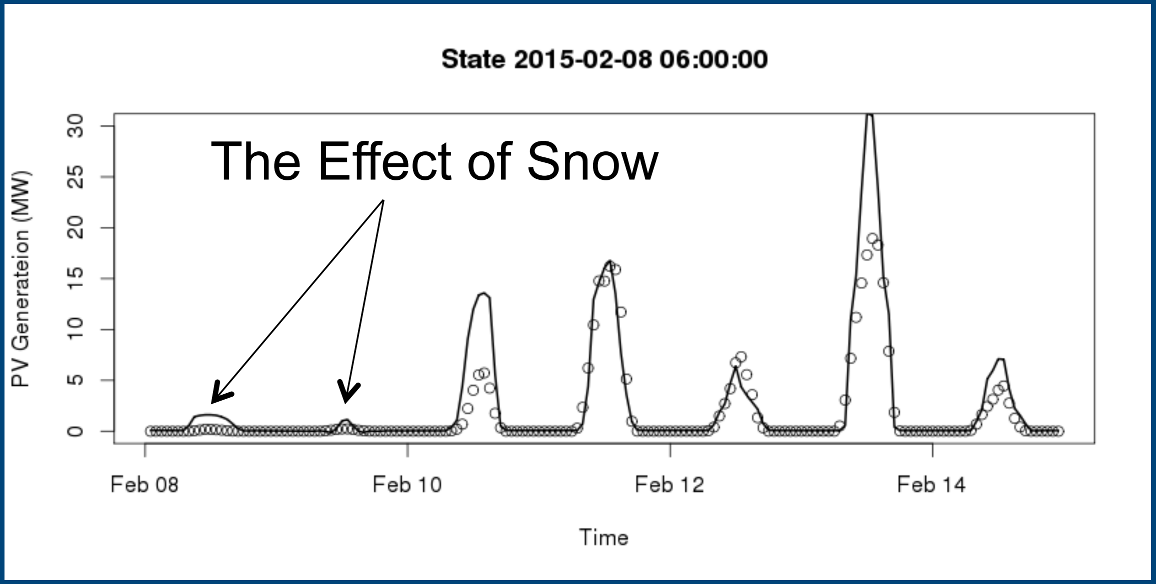


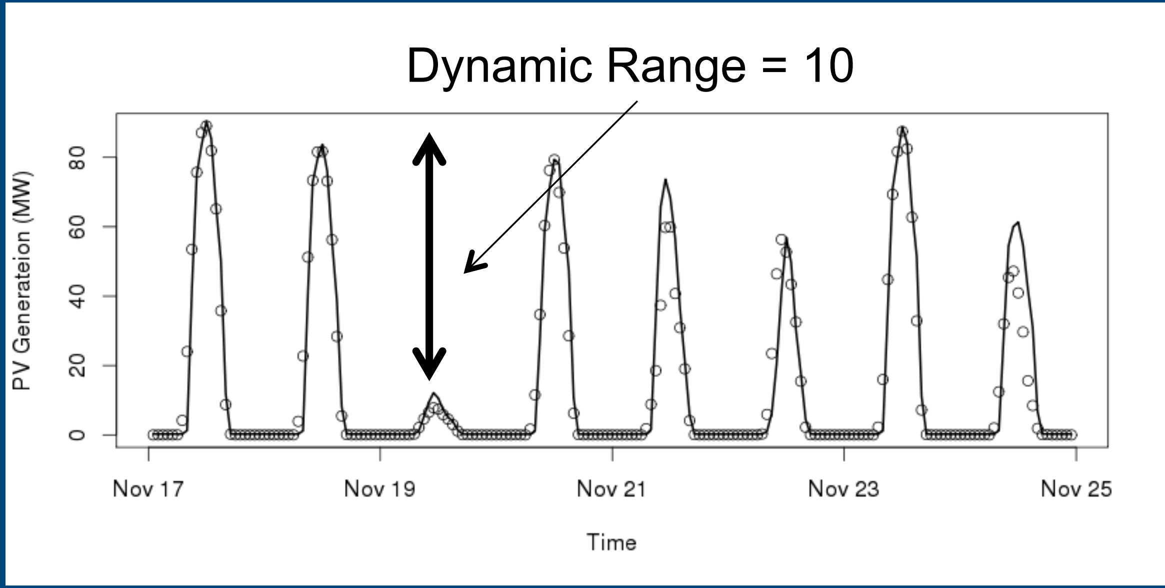
Residual Demand = Delivered - Returned

Prediction Complexity

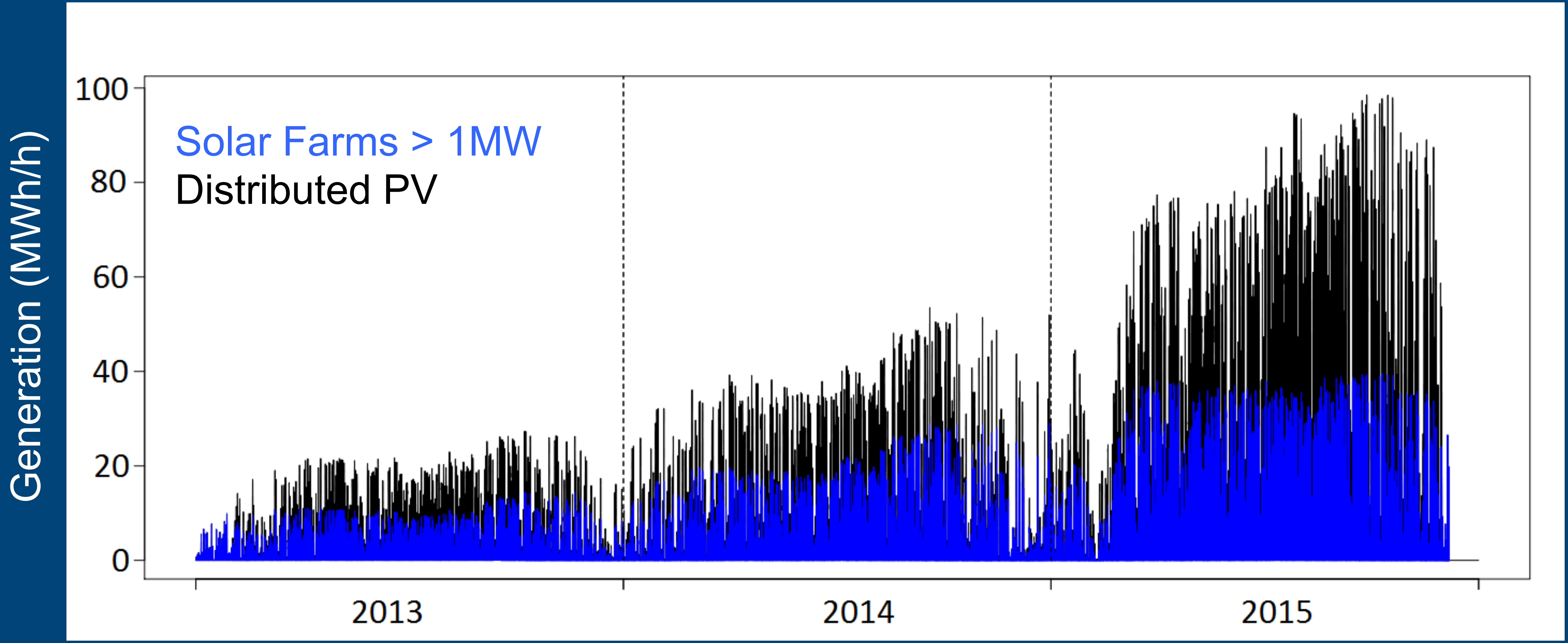
The impact of increases in generation capacity and snowfall are not well captured by physical models. Machine learning helps.

We can predict PV generation with an error of better than 10% over a dynamic range of factor 10





Nature of Distributed Energy

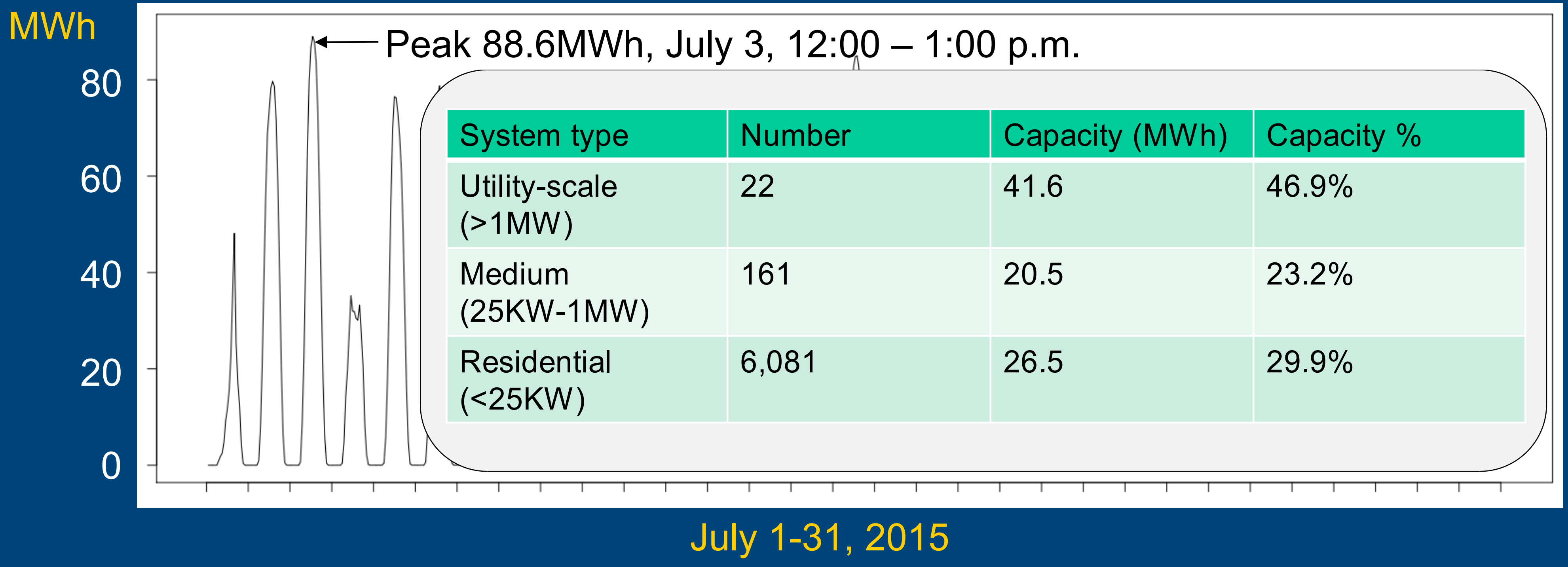


Hourly Measurements

Smart Meter Analytics

Total PV capacity in Vermont:

- Based on AMI and available SCADA
- Missing: approx. 5-10 MW distributed PV



System type	Number	Capacity (MWh)	Capacity %
Utility-scale (>1MW)	22	41.6	46.9%
Medium (25KW-1MW)	161	20.5	23.2%
Residential (<25KW)	6,081	26.5	29.9%

Example 3: Weather Forecasting

- Weather forecasts for critical business operations
  - Renewable energy
  - Insurance
  - Aviation
  - Agriculture

- Impacts of weather
  - Decision making from forecasts
  - Multi-dimensional vs. One-dimensional impact

IBM Acquired The Weather Company in January 2016
TWC is a *data* company
Internet of Things (IoT)
Spatial-Temporal Data
Global forecasting capability
Business models: B2C and B2B

High-Resolution Weather Forecasting

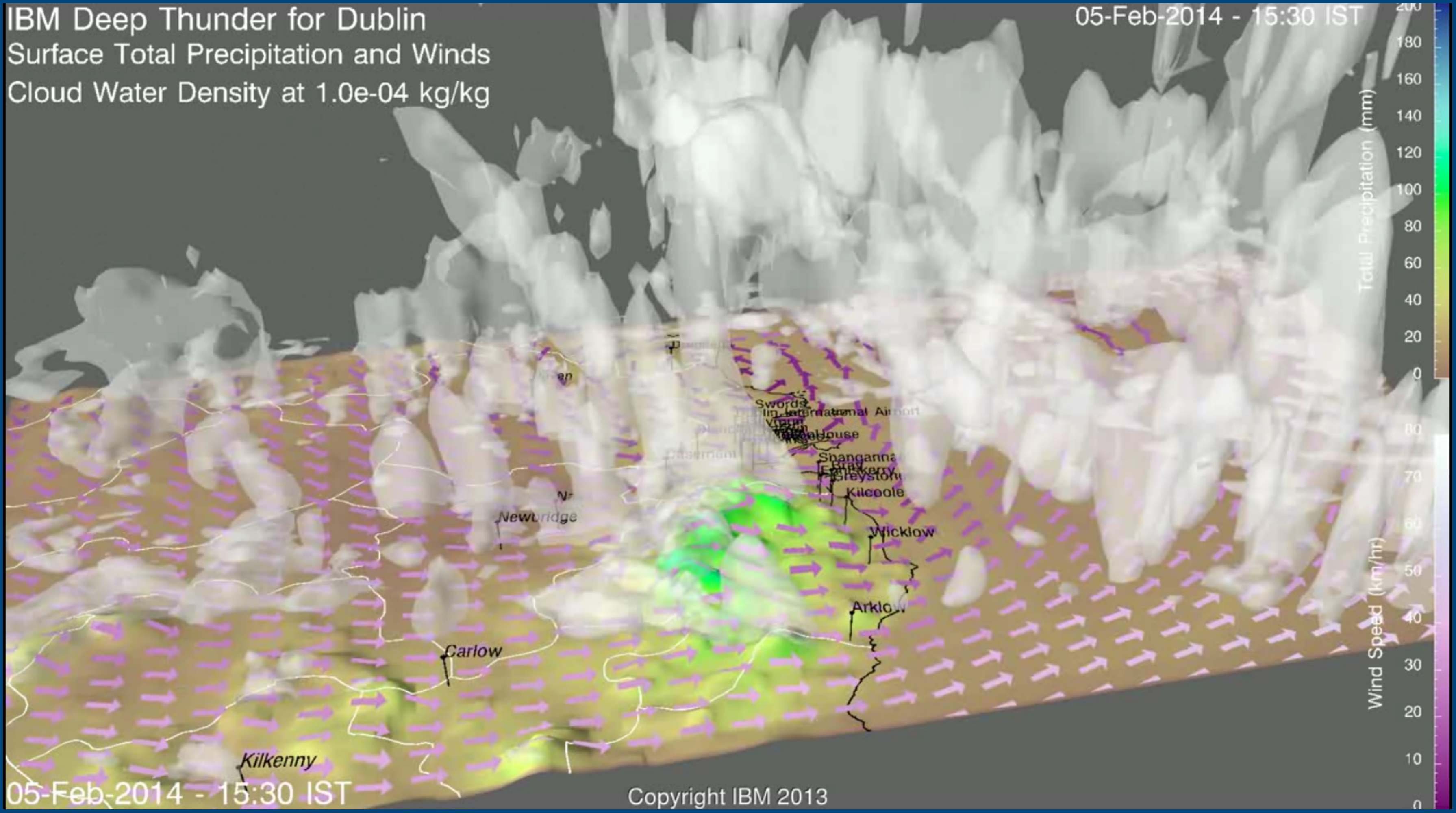
- Goal: High-resolution weather forecasting at specific time and location
  - Spatial scales of 1x1 km and temporal scale of 10 minute time steps
  - Forecasts out to 72 hours in advance

- Approach:
  - 3D physics model of weather
  - Assimilation of real time observational data into forecasts
  - Cloud-based implementation

- Requirements:
  - Coarse-scale forecasts (e.g., NOAA, ECMWF)
  - Observational data

- Result:
  - High resolution, on-demand forecasts anywhere in the world
  - Applications for flood forecasting and storm water runoff

Deep Thunder Example: Dublin Ireland



Forecast of clouds

Color scale shows cumulative rainfall on ground

Arrows show direction of wind and velocity (color scale)

IBM

Cognitive Computing

Individualized, ranked, evidence based treatment options at the point of care. Watson extracts patient case attributes and provides supporting rationale for the treatment options.

Other innovative recipes conjured up by Watson include coconut-flavored Caribbean Snapper Fish & Chips, Belgian Bacon Pudding, and the Austrian Chocolate Burrito with lean ground beef and two ounces of dark chocolate.

Cognitive Computing: Urban Systems

Leakage Rates

Road Construction

Asset Condition

Traffic Jam

Population Growth

How much snow?

Interest Rates

How do I best operate the city today?

How do I make decisions today to improve operations in the future

Consumer Demands

Citizen Complaints

Network Flows

Rainfall

Regulatory Constraints

Fuel Costs

Electricity Prices

Tank Levels

River Flows

Maintenance Schedules

Security

Storm Water Drains

Effluent Quality

Bus Driver Strike

Summary

- Models are required for predictions
  - Combining different classes of models for improved predictions
- Data + (Physics) Models
  - Water: Aging infrastructure, more data, extend lifetime
  - Energy: More solar, more data, uncertainty remains
  - Weather: Storms impact multiple infrastructure systems
- Cognitive (Self-learning) Models

Sean A. McKenna
seanmcke@ie.ibm.com