Scientific Visualization of Big Data

Gunther H. Weber

Computational Research Division
Lawrence Berkeley National Laboratory

Department of Computer Science
University of California, Davis

CRD
computational
research division

BERKELEY LAB
LAWRENCE BERKELEY NATIONAL LABORATORY

Introduction – Research Interests

- High performance visualization and analytics
  - Algorithms utilizing massively parallel computational resources

- Feature-based visualization, e.g., using topological analysis
  - Data abstraction and reduction
  - Comparison simulation/simulation and simulation/experiment

- Apply data analysis and visualization to wide range of data understanding problems
  - Simulations (cosmology, carbon sequestration, fusion, …)
  - Experimental data (carbon sequestration, brain imaging, background radiation, …)

- *In situ* analysis: Integrate data analysis into simulation

Accelerating Visual Analysis with Data Indexing

Work on query driven visualization by
K. Stockinger, J. Shalf, K. Wu, W. Bethel

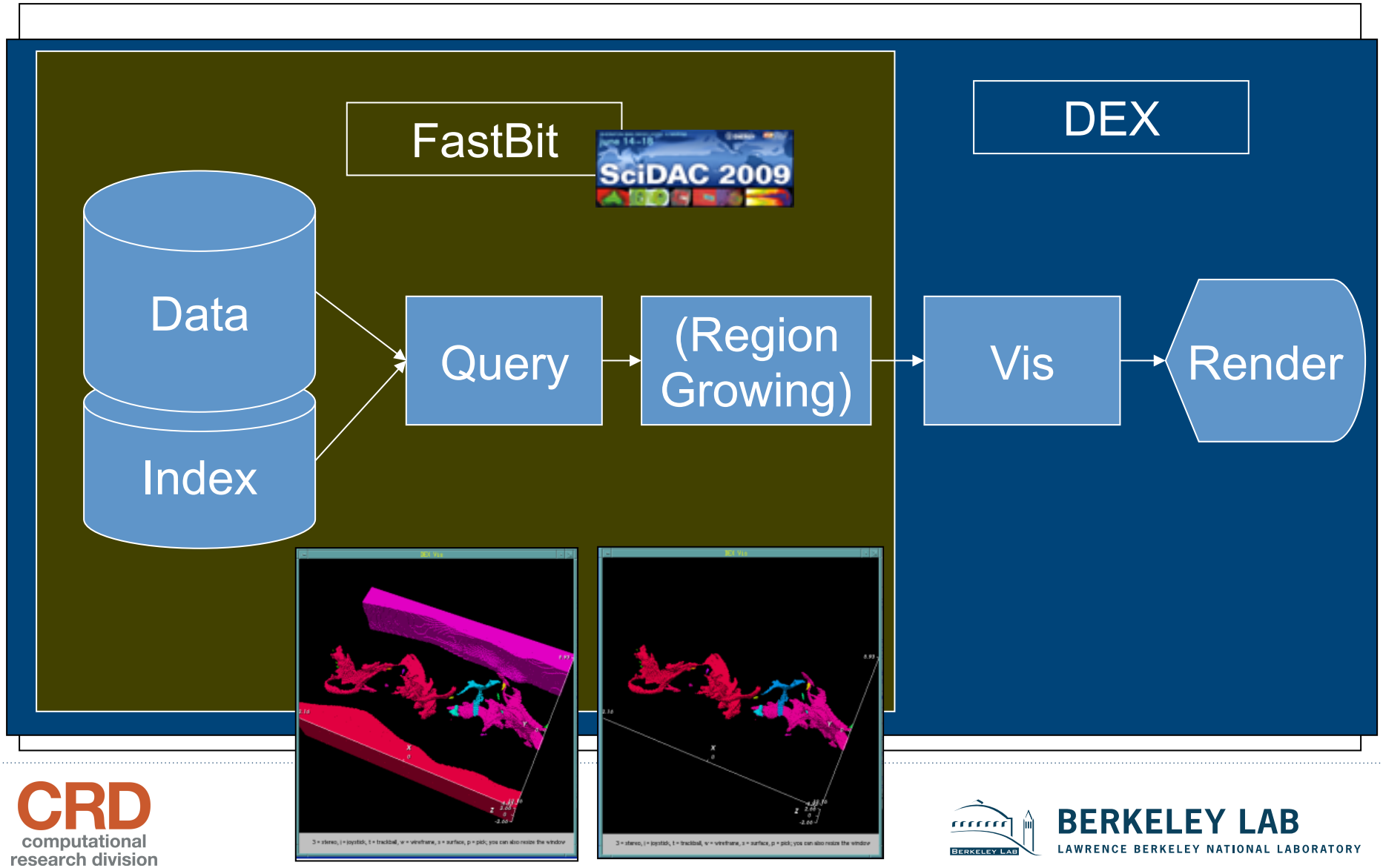
Work on ARES Project with
L. Ramakrishnan, M. Bandstra, D. Chivers, Y. Ong,
M. Quinlan, B. Quiter, K. Song

Query-Driven Visualization

- What is Query-Driven Visualization?
  - Find "interesting data" and limit visualization, analysis, machine and cognitive processing to that subset
- One way to define "interesting" is with compound boolean range queries.
  - E.g., ($CH_4 > 0.1$) AND ($T_1 < temp < T_2$)
- Quickly locate those data that are "interesting"
- Pass results along to visualization and analysis pipeline
- Quick computation of histograms (1D, 2D) provides additional information

Query-Driven Visualization

Query-Driven Visualization



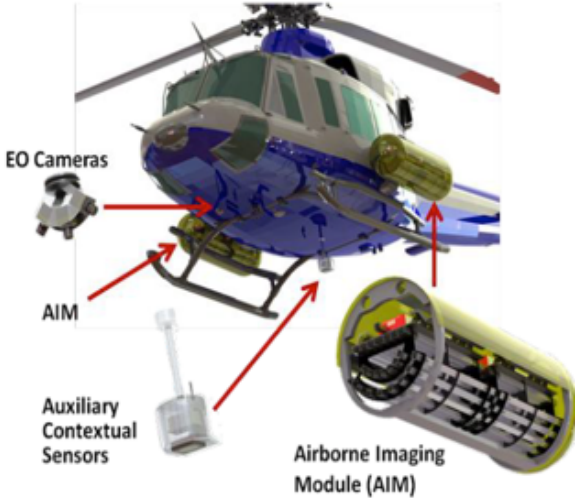
ARES – Helicopter-borne Radiological Search

Goal:

- Improved methods to detect and localize radiological sources

Data Collection:

- Multi-sensor systems for airborne platforms with highly sensitive radiation detectors
- Recorded data during flight
  - Position (GPS location, velocity)
  - Radar altitude
  - Radiation detection events: energy, time stamp, detector id
  - Video
- SQL database for metadata; HDF5 for data
- FastBit index accelerates data access



Data Browser

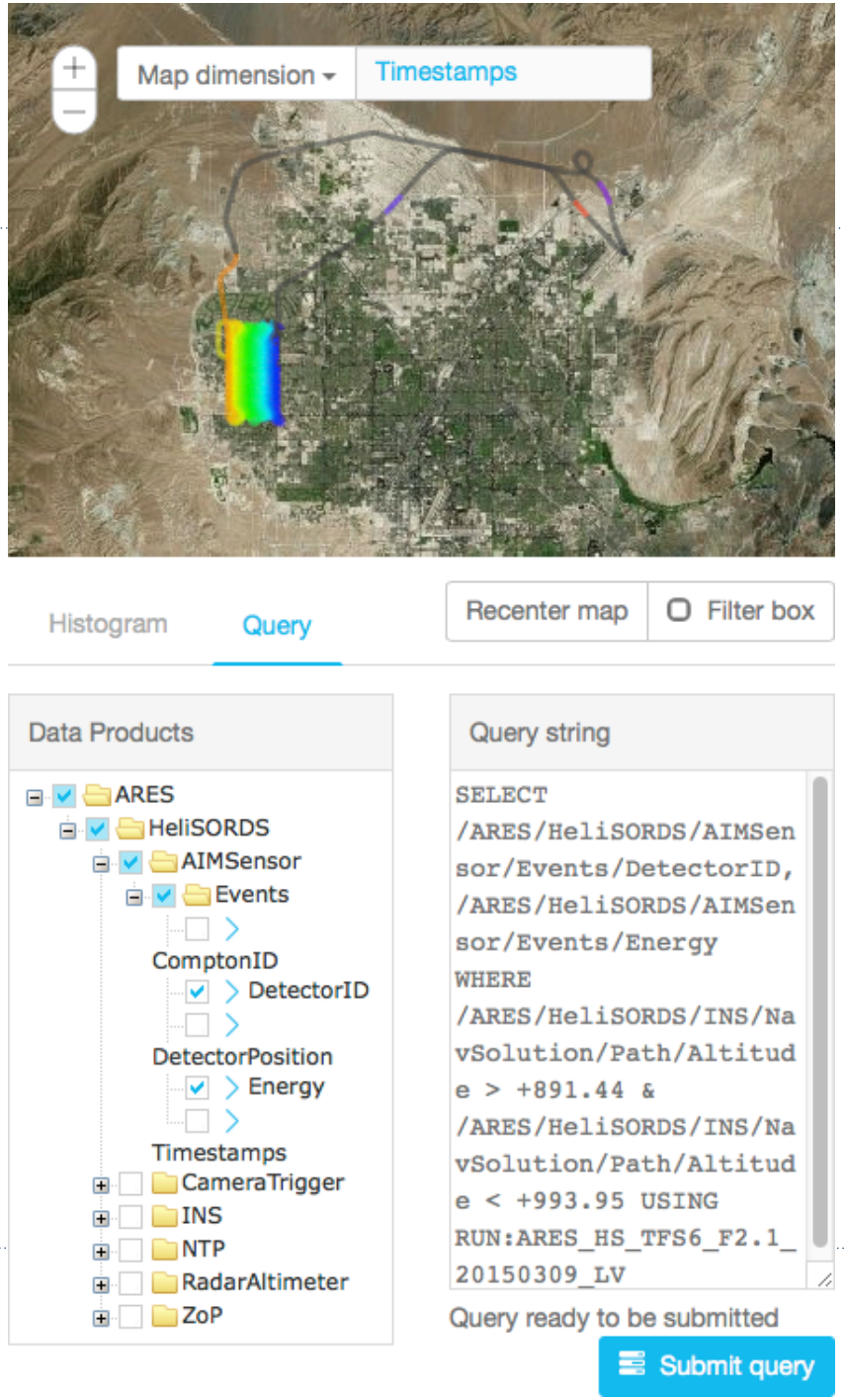
Goal:

- Identify relevant flight path / data set subsets for detailed analysis

Approach:

- Web-based UI; web server backend
- Multiple coordinated views
  - Map view with flight path
  - Histogram with data distribution

  show different data aspects
- Coloring flight path by data sets illustrates context
- Selecting ranges in histogram highlights corresponding map locations
- **FastBit accelerates queries**



CRD
computational research division

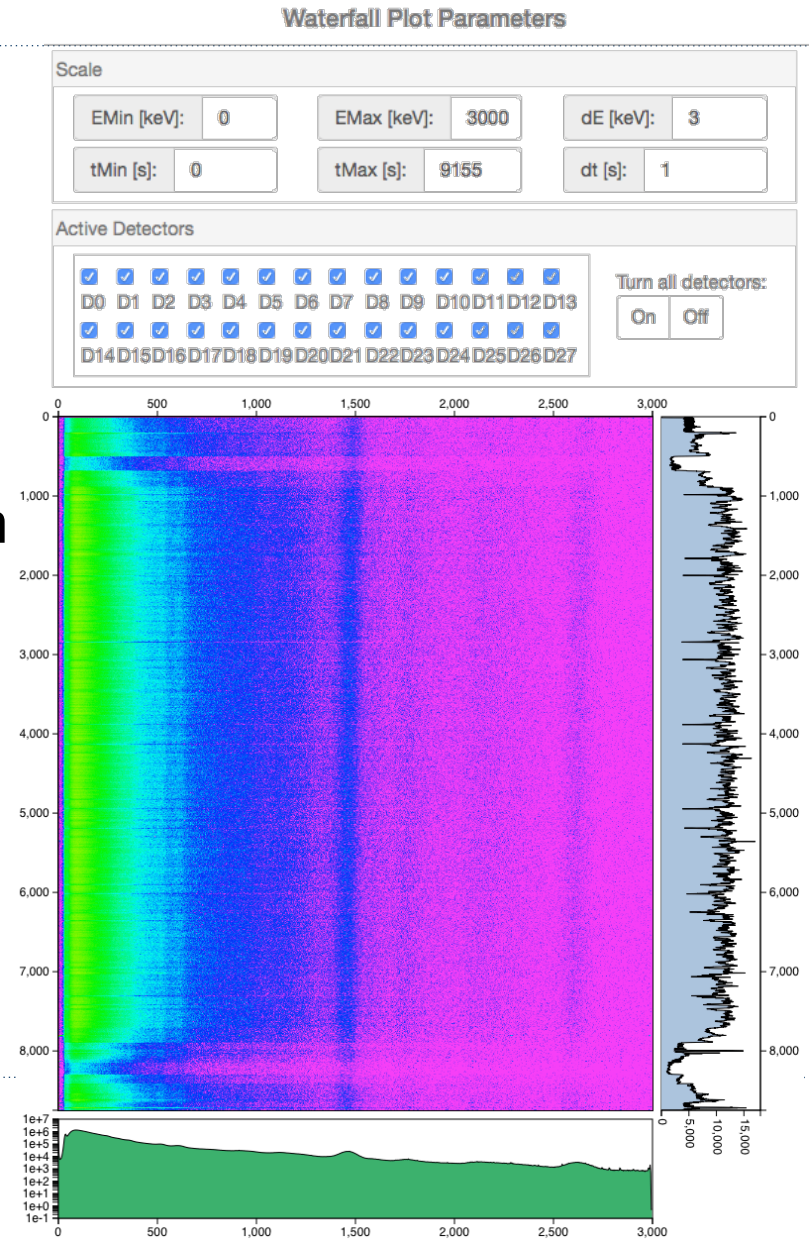
Spectrum Visualization with FastBit

Goal:

- Display spectrum for series of events (time, energy, detector id)
- Support changes to bin size for energy, time and active detectors

Approach:

- Use FastBit index to compute spectrum (= 2D histogram) on web backend
- FastBit reduces computation time to seconds instead of minutes
- No caching necessary; calculation for arbitrary parameter combinations
- Once spectrum downloaded, all interactions (zooming, …) occur locally



CRD
computational research division

Simulating Urban Environments for Energy Analysis

Work with H. Johansen, D. T. Graves, and T. J. Ligocki

Computational Research Division, Berkeley Lab, USA

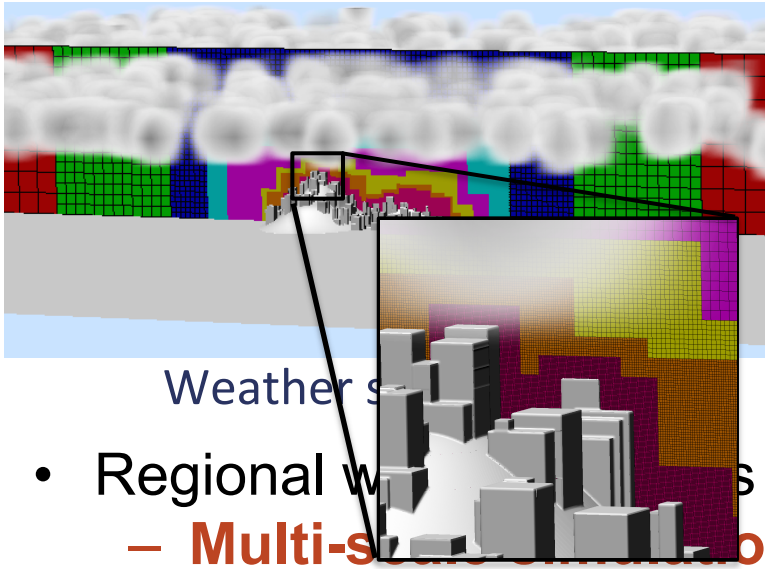
Complexity of Urban Power Supply and Demand Growing Rapidly

- **New technologies**, such as renewables (e.g., **solar and wind energy**), demand response, electric vehicles, combined heat/power systems and battery storage **must coexist with existing generation capacity**

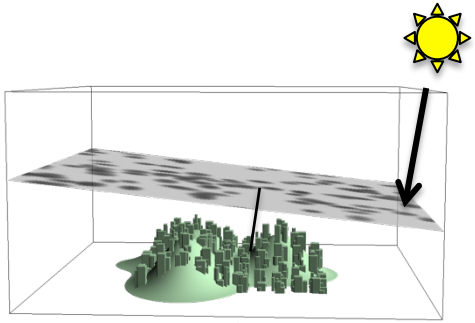
- **Factors driving energy balance,** including air temperature, air quality, humidity, precipitation, **incident sunlight, shade and cloud cover, dominated by uncertainty and day-to-day variation**

➜ Static peak views of system (solar angle, average hours sunlight per day) miss opportunities for optimizing supply/demand and solar energy usage

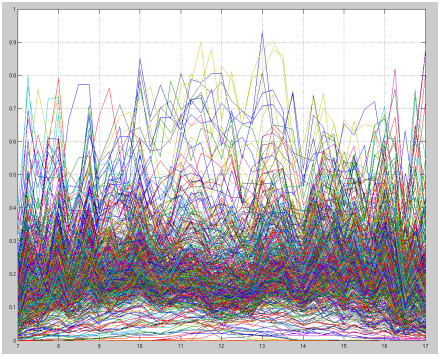
Optimal Design Tool for Urban Energy Planning Requires Regional Models



Weather s



Incident solar energy calculations

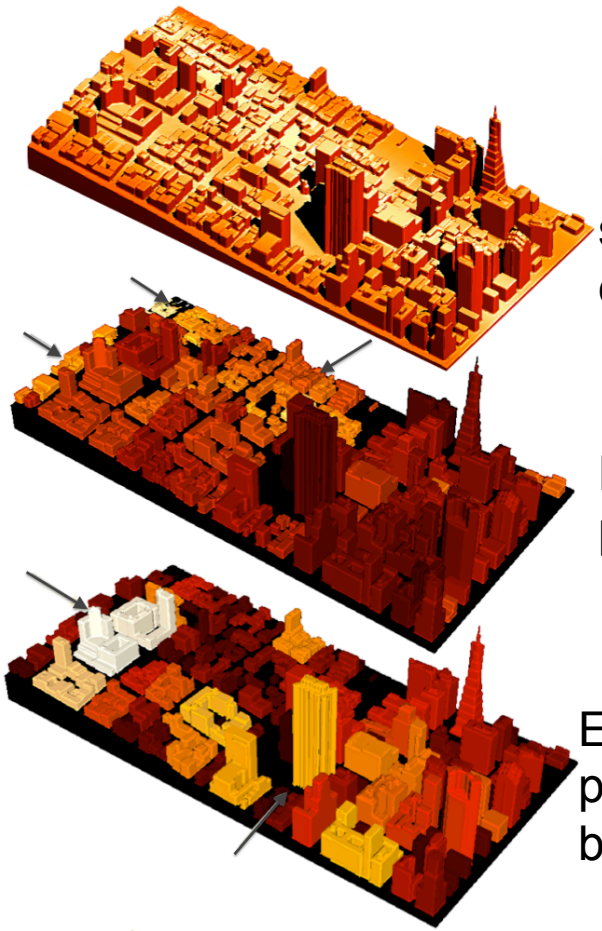


Per-building time series

- Regional w s in planning tool for urban energy usage
  - **Multi-s on** of key factors affecting incident solar radiation (e.g., **cloud cover**)
  - **Incident solar energy calculation** considering cloud cover, topography, angle of the sun, shadows, etc.
  - **Per-building** (per surface) **time series** of incident solar energy
- Use to evaluate potential for optimizing solar energy generation and electricity demand

Results – Variation Over Day

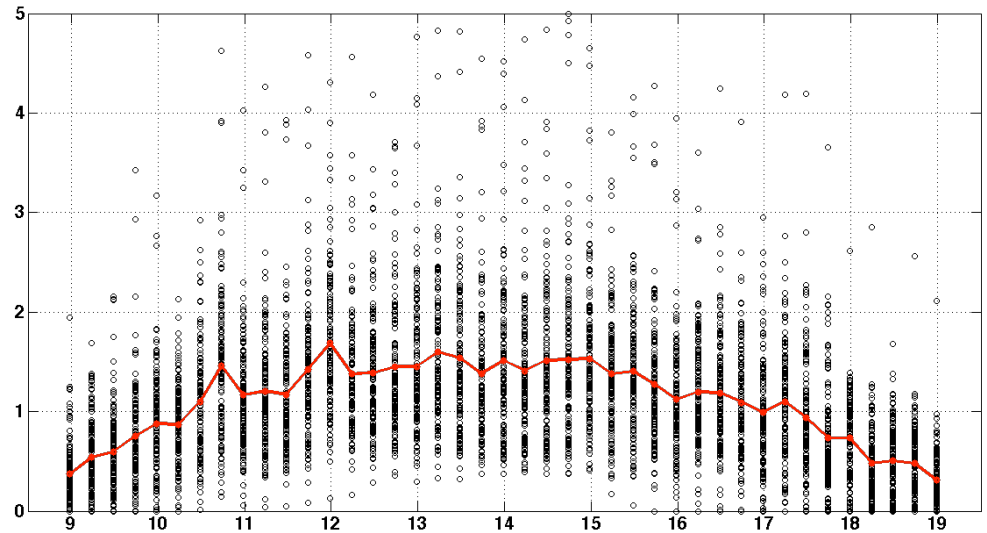
Results – Analysis



Incident solar energy

Energy per area

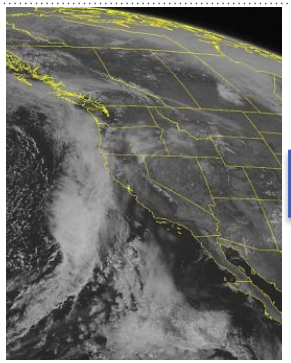
Energy per building



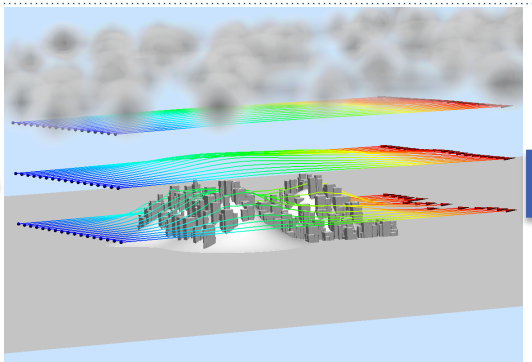
Incident solar energy per unit area vs. time

- Lighter color = more energy
- Arrows: Districts or building that might be focus of policy or demand initiatives

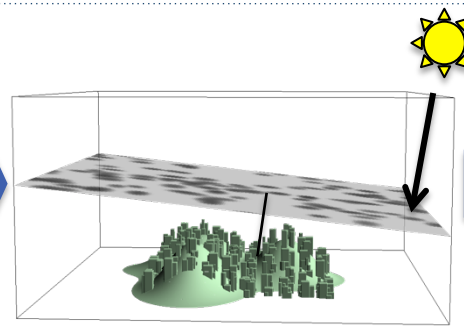
Regional Energy Modeling in the Future



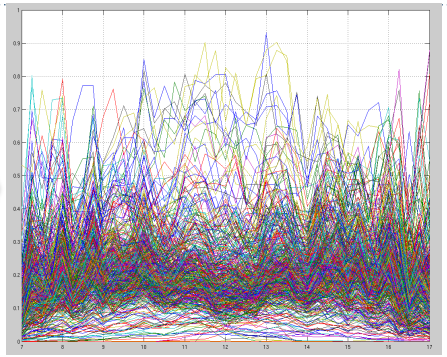
Data



Simulation



Models



Data assimilation and model reduction

Policy

Detailed Climate and Energy Available:

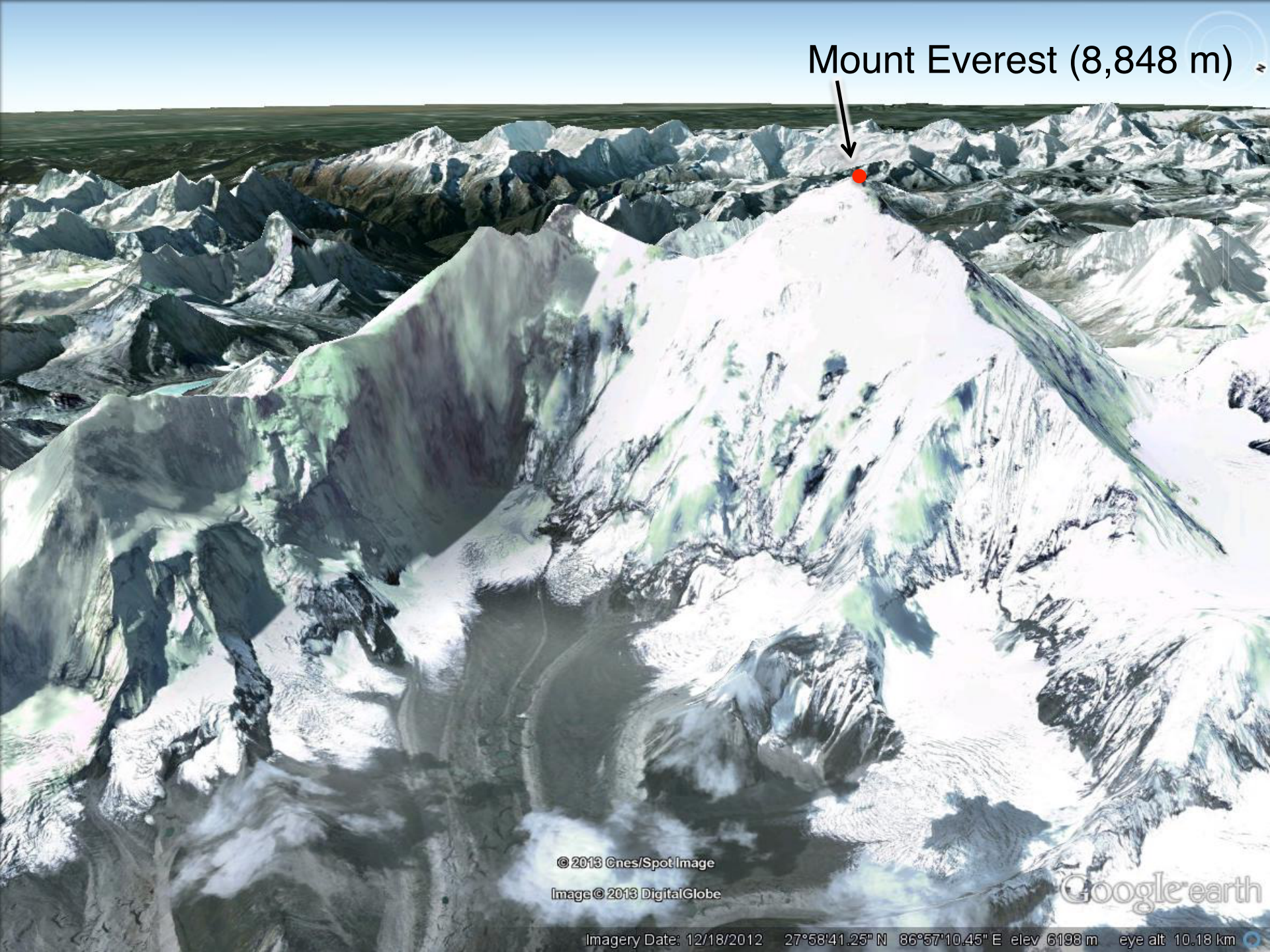
- Electricity grid/meter data becomes pervasive, detailed
- Climate/weather prediction improved long-term accuracy

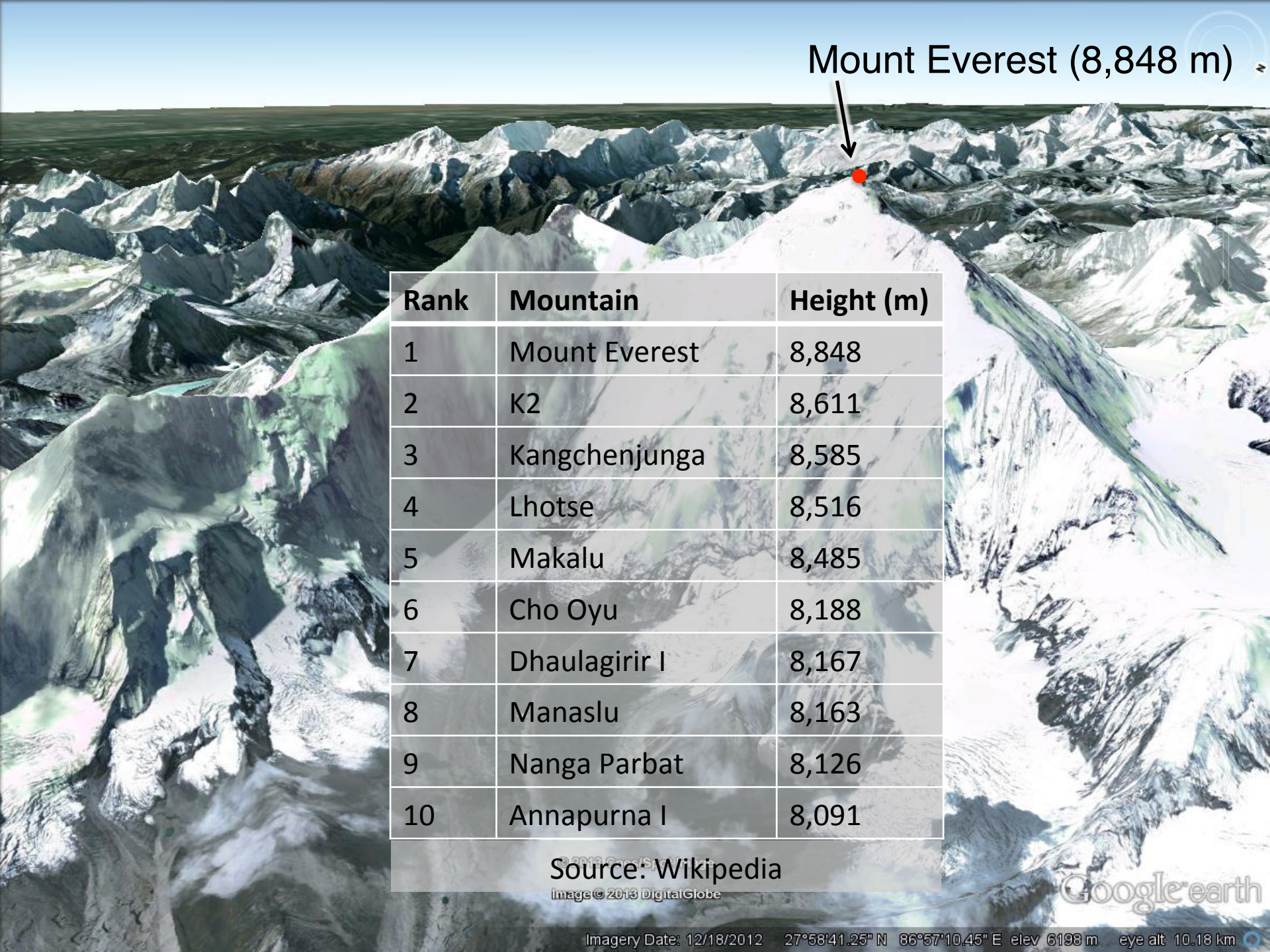
2025 Model:

- Real-time weather and energy usage data enables simulation and control of regional energy distribution networks; strategies to optimize efficiency, minimize risk
- Integrated data collection/simulation support detailed "what-if?" analyses and drive energy policy decisions

Topological Data Analysis

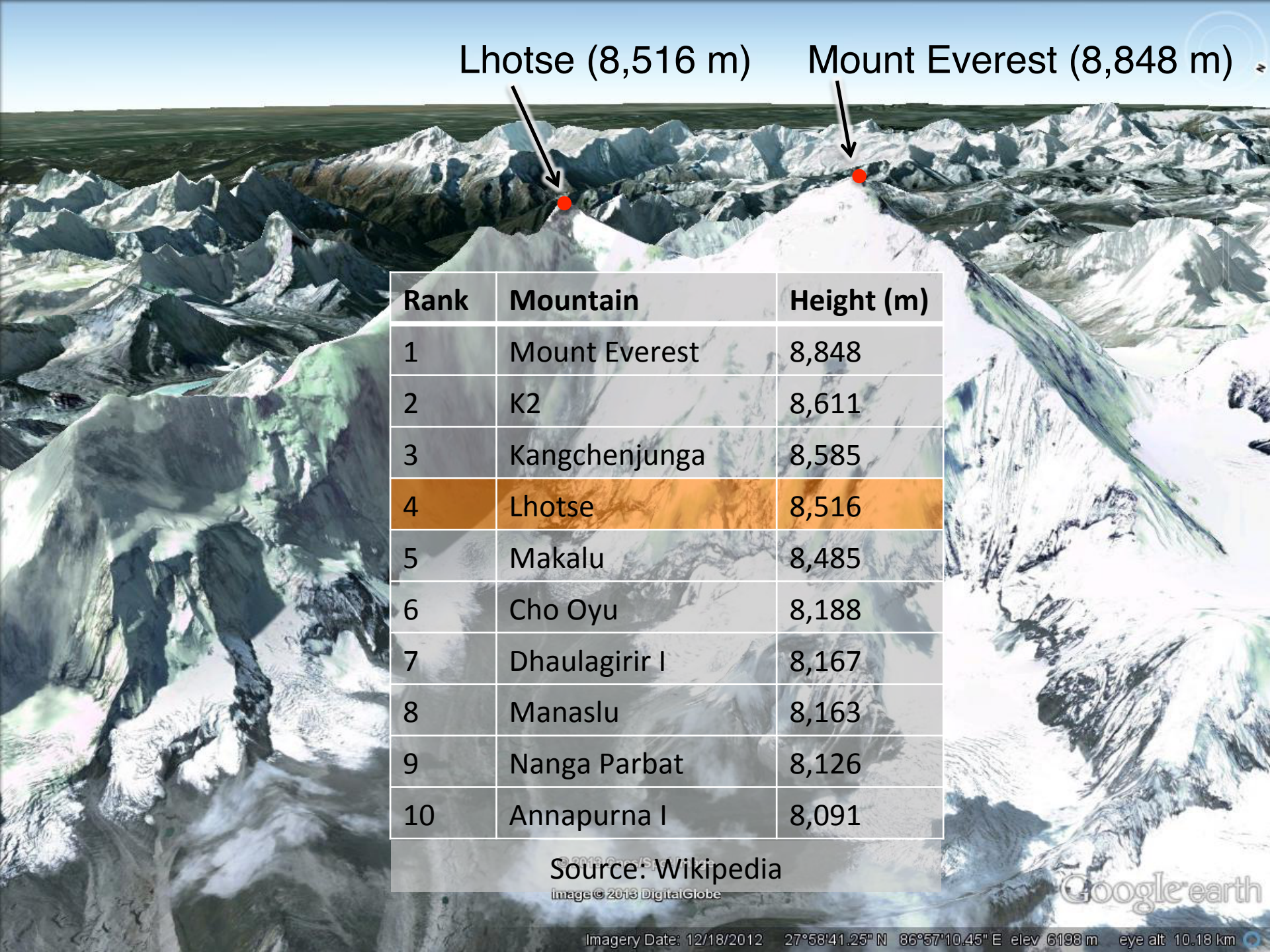
Work with D. Morozov


Mount Everest (8,848 m)
© 2013 Cnes/Spot Image
Image © 2013 DigitalGlobe
Google earth
Imagery Date: 12/18/2012 27°58'41.25" N 86°57'10.45" E elev 6198 m eye alt 10.18 km



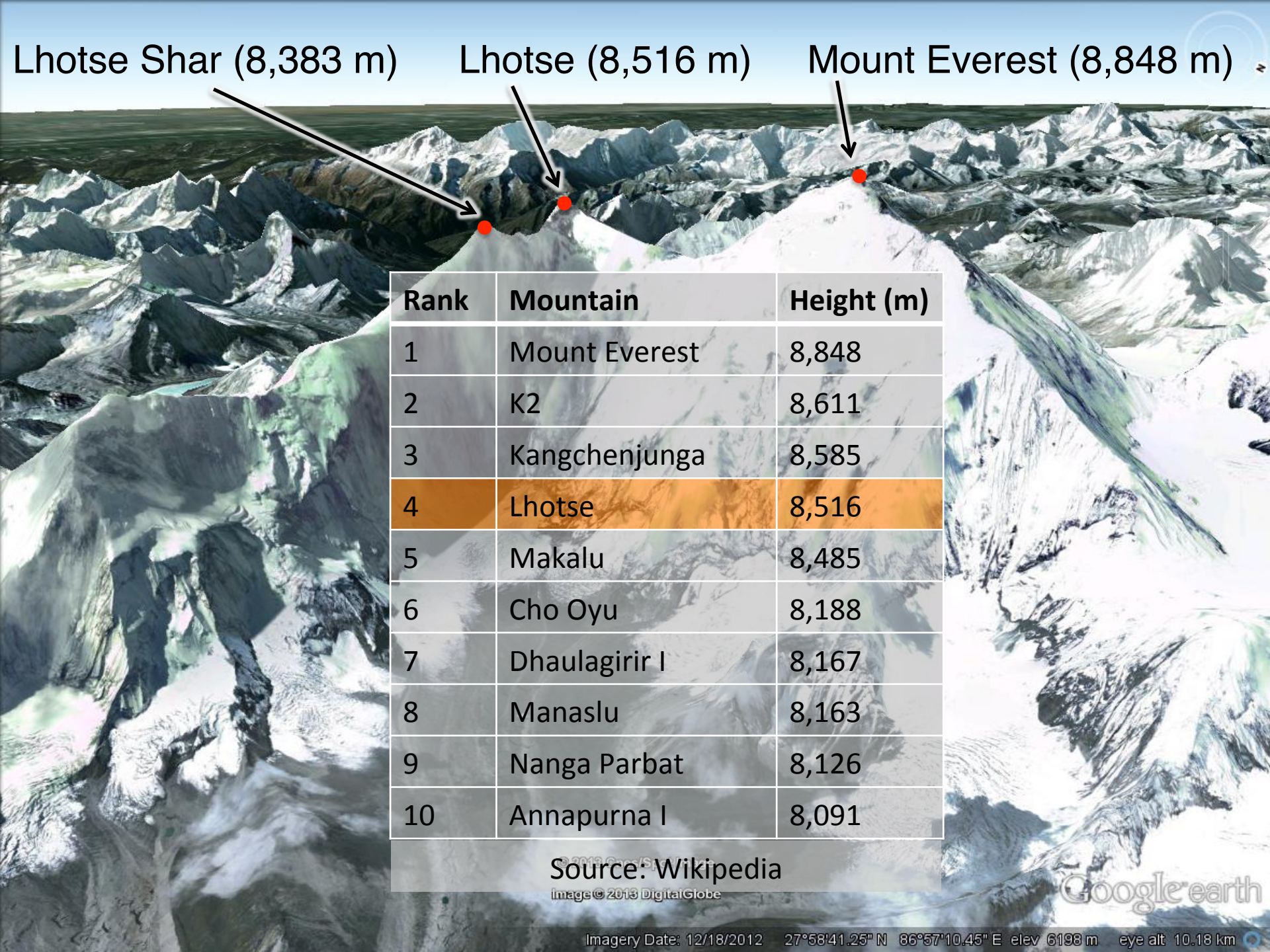
Rank	Mountain	Height (m)
1	Mount Everest	8,848
2	K2	8,611
3	Kangchenjunga	8,585
4	Lhotse	8,516
5	Makalu	8,485
6	Cho Oyu	8,188
7	Dhaulagirir I	8,167
8	Manaslu	8,163
9	Nanga Parbat	8,126
10	Annapurna I	8,091

Source: Wikipedia



Rank	Mountain	Height (m)
1	Mount Everest	8,848
2	K2	8,611
3	Kangchenjunga	8,585
4	Lhotse	8,516
5	Makalu	8,485
6	Cho Oyu	8,188
7	Dhaulagirir I	8,167
8	Manaslu	8,163
9	Nanga Parbat	8,126
10	Annapurna I	8,091

Source: Wikipedia

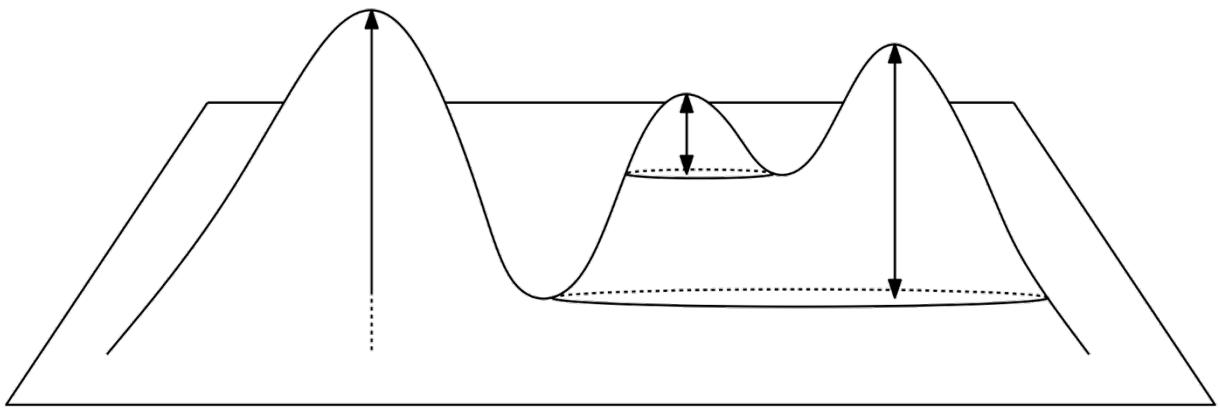


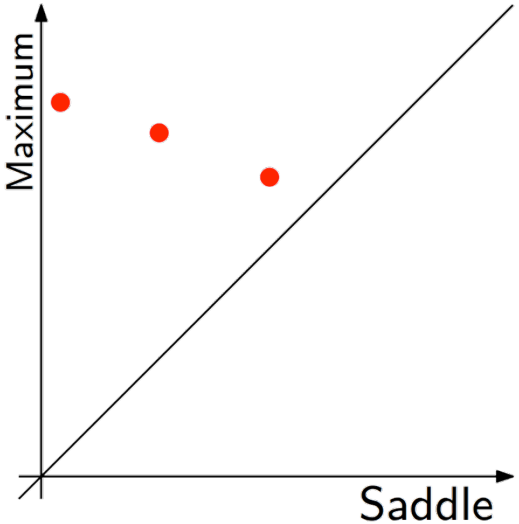
Rank	Mountain	Height (m)
1	Mount Everest	8,848
2	K2	8,611
3	Kangchenjunga	8,585
4	Lhotse	8,516
5	Makalu	8,485
6	Cho Oyu	8,188
7	Dhaulagirir I	8,167
8	Manaslu	8,163
9	Nanga Parbat	8,126
10	Annapurna I	8,091

Source: Wikipedia

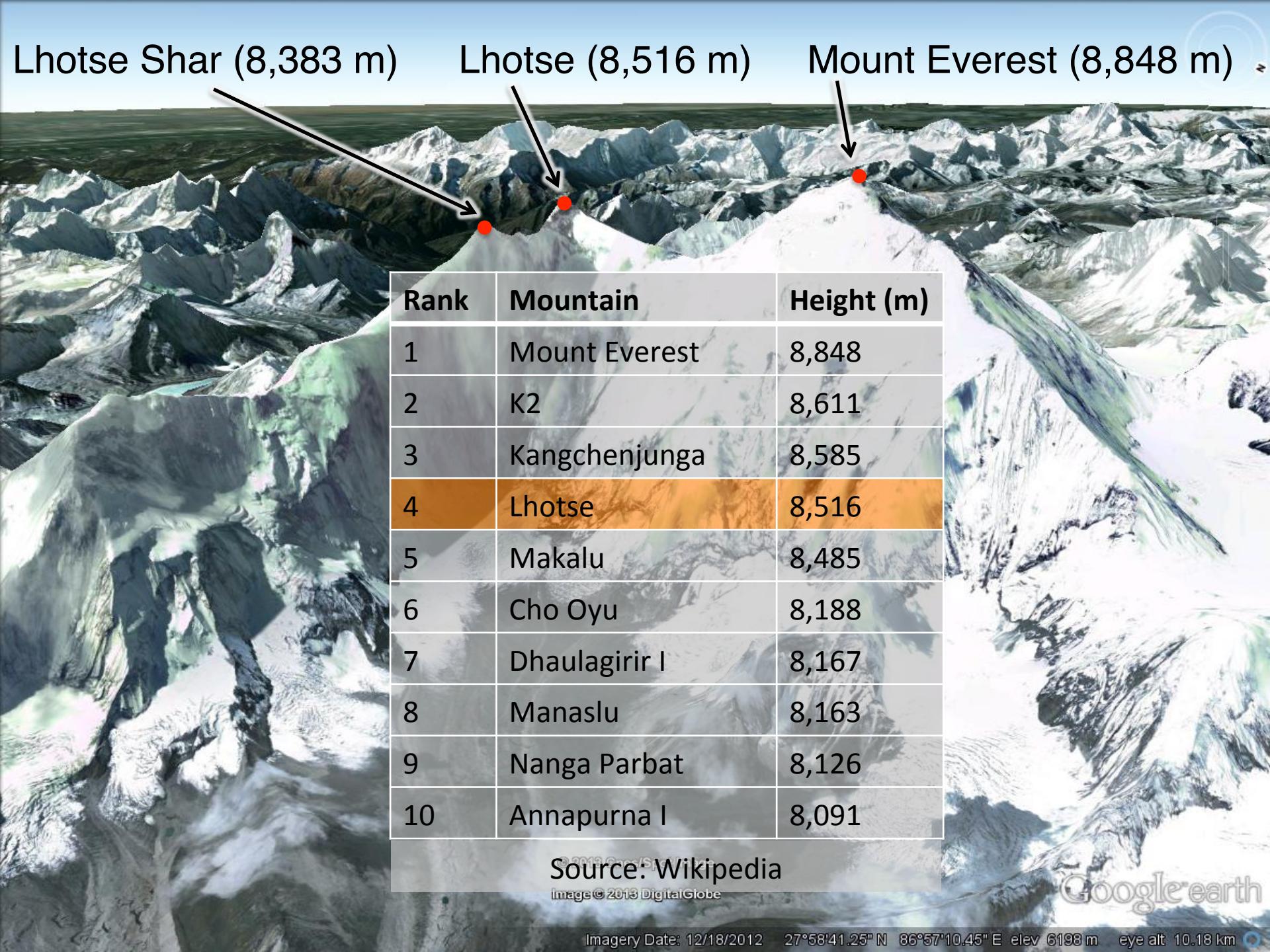
Topographic Prominence

- **Prominence** of a peak is the height of the peak's summit above the lowest contour line encircling it and no higher summit
- Prominence is also called **persistence**





- **Persistence diagram** records for each peak its value on the vertical axis, and the value of the saddle where it merges into a higher peak on the horizontal axis

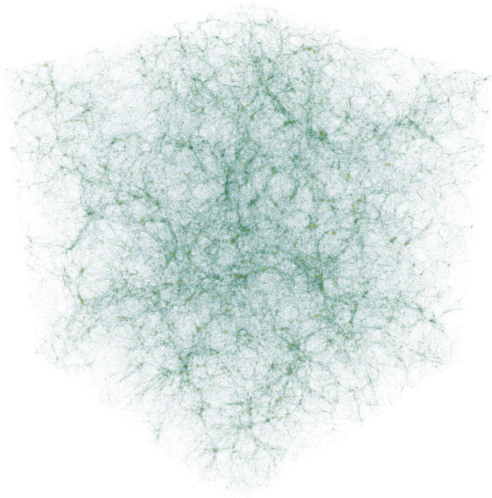


Rank	Mountain	Height (m)
1	Mount Everest	8,848
2	K2	8,611
3	Kangchenjunga	8,585
4	Lhotse	8,516
5	Makalu	8,485
6	Cho Oyu	8,188
7	Dhaulagirir I	8,167
8	Manaslu	8,163
9	Nanga Parbat	8,126
10	Annapurna I	8,091

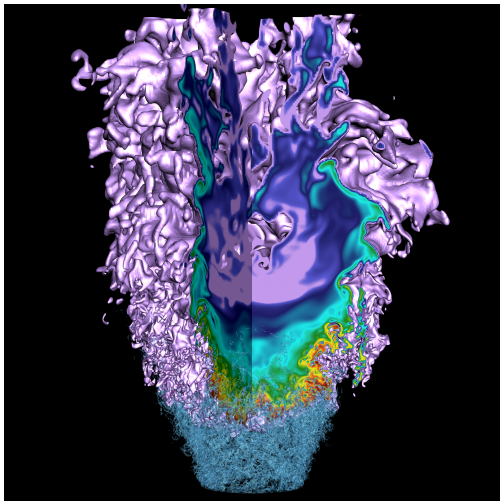
Source: Wikipedia

Why Do We Care? Feature Detection With Topological Data Analysis

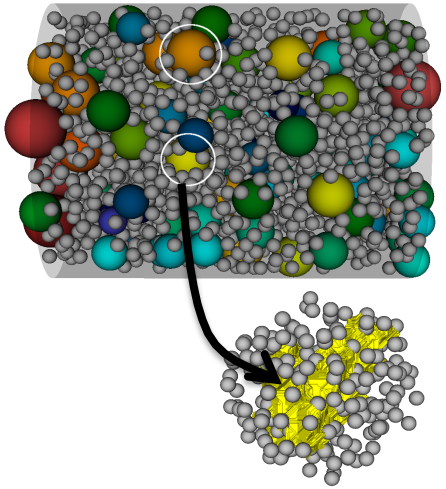
- Natural phenomena modeled as scalar functions, $f : \mathbb{X} \to \mathbb{R}$
  - density of stars, galaxies or clusters of galaxies (halos)
  - fuel consumption rate during combustion (burning regions)
  - geometry of a material encoded in its distance function (pockets)



Cosmology



Combustion

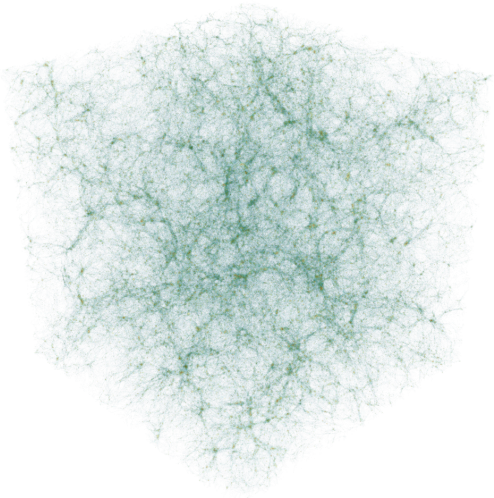


Materials

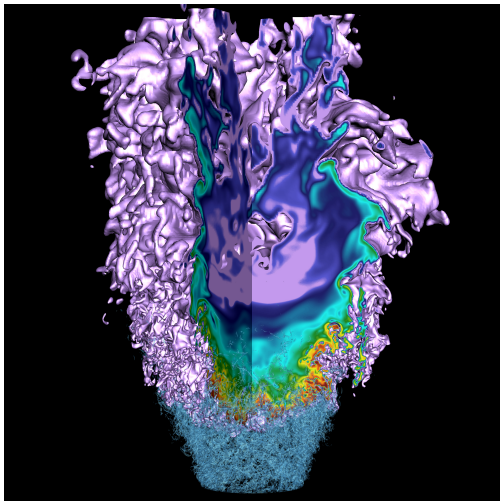
Needed: Method to detect and extract salient features for analysis

Why Do We Care? Feature Detection With Topological Data Analysis

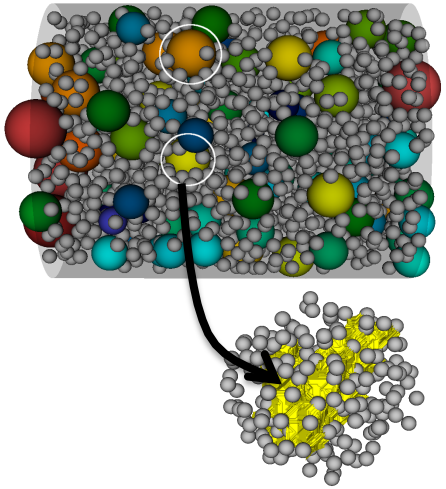
- Natural phenomena modeled as scalar functions, $f : \mathbb{X} \to \mathbb{R}$
  - density of stars, galaxies or clusters of galaxies (halos)
  - fuel consumption rate during combustion (burning regions)
  - geometry of a material encoded in its distance function (pockets)



Cosmology



Combustion

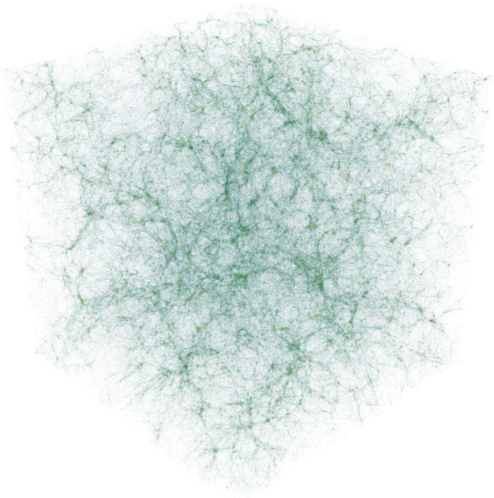


Materials

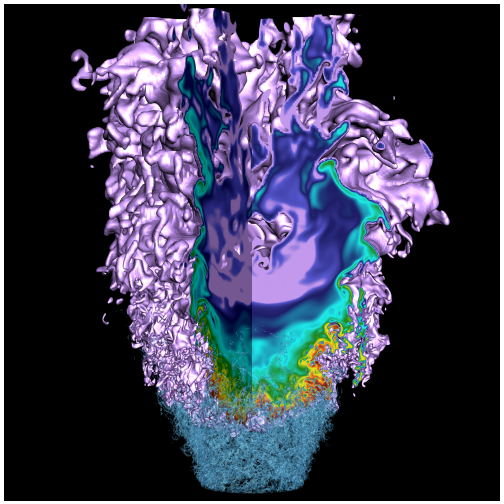
Needed: Method to detect and extract salient features for analysis

Why Do We Care? Feature Detection With Topological Data Analysis

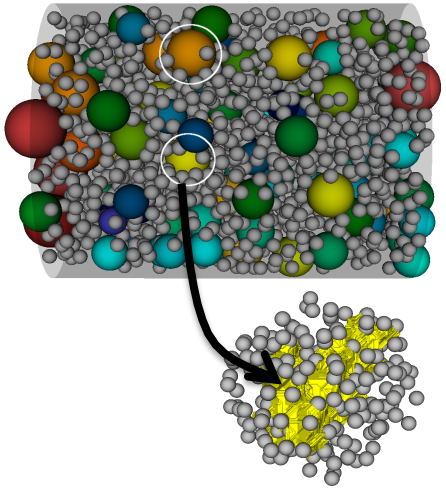
- Natural phenomena modeled as scalar functions, $f : \mathbb{X} \to \mathbb{R}$
  - density of stars, galaxies or clusters of galaxies (halos)
  - fuel consumption rate during combustion (burning regions)
  - geometry of a material encoded in its distance function (pockets)



Cosmology



Combustion

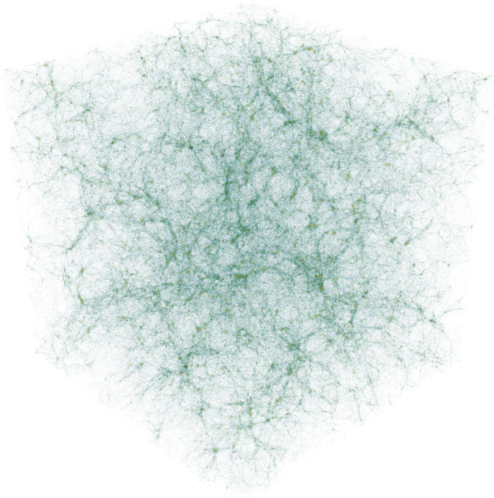


Materials

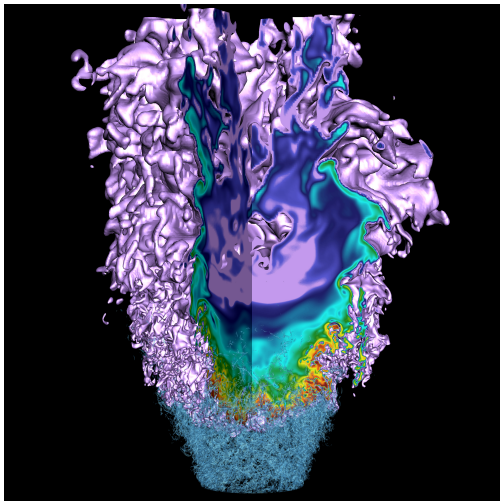
Needed: Method to detect and extract salient features for analysis

Why Do We Care? Feature Detection With Topological Data Analysis

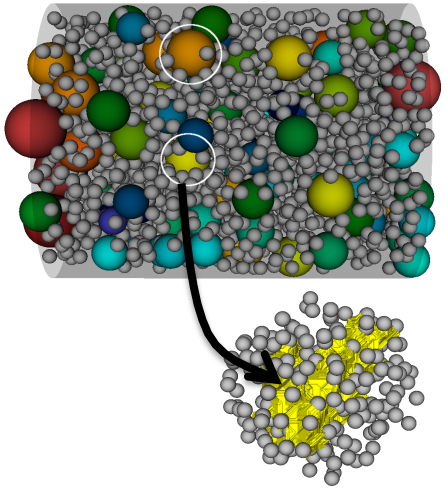
- Natural phenomena modeled as scalar functions, $f : \mathbb{X} \to \mathbb{R}$
  - density of stars, galaxies or clusters of galaxies (halos)
  - fuel consumption rate during combustion (burning regions)
  - geometry of a material encoded in its distance function (pockets)



Cosmology



Combustion

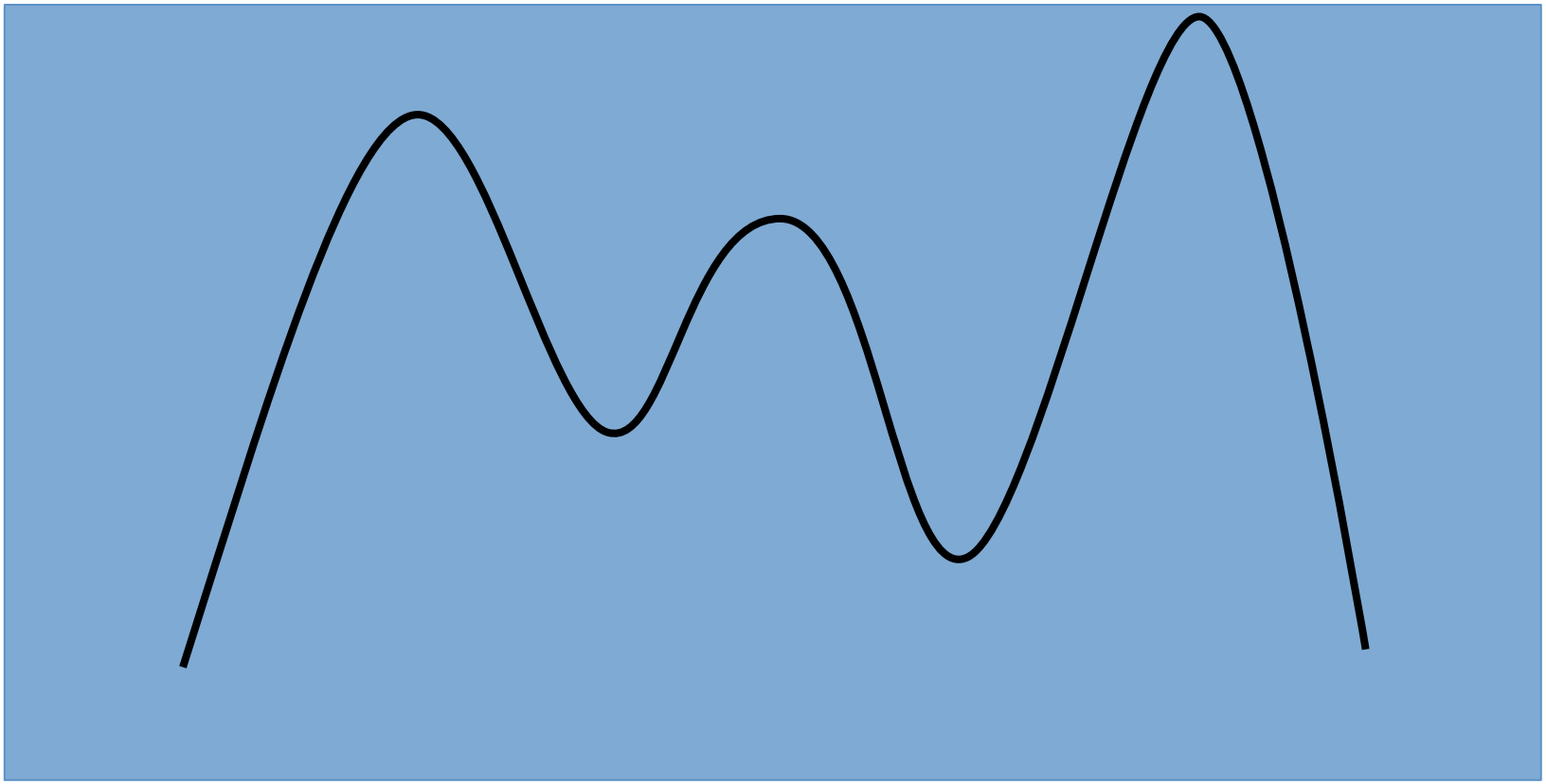


Materials

Needed: Method to detect and extract salient features for analysis

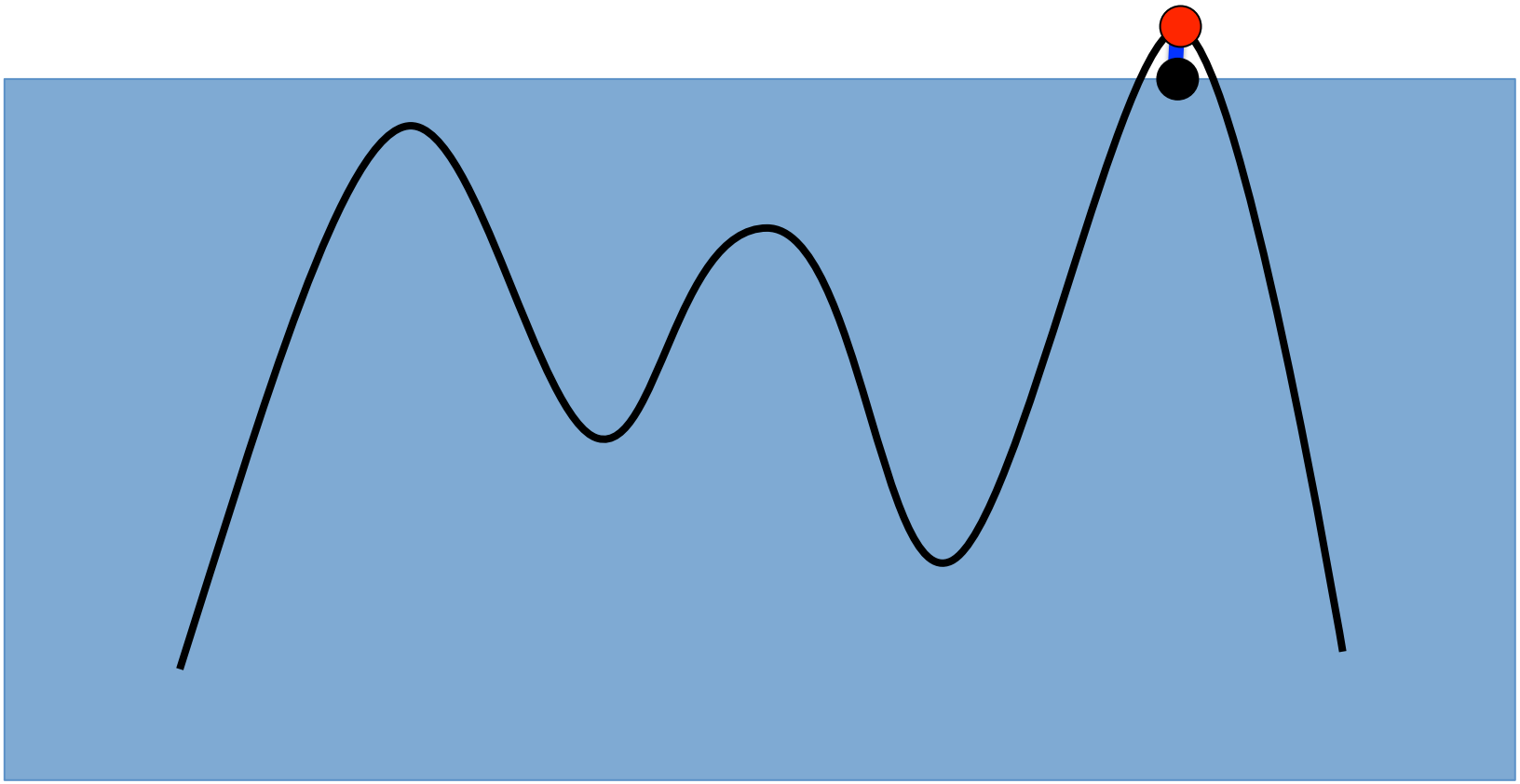
Merge Tree Construction – 1D Example

- Record connectivity of superlevel set components



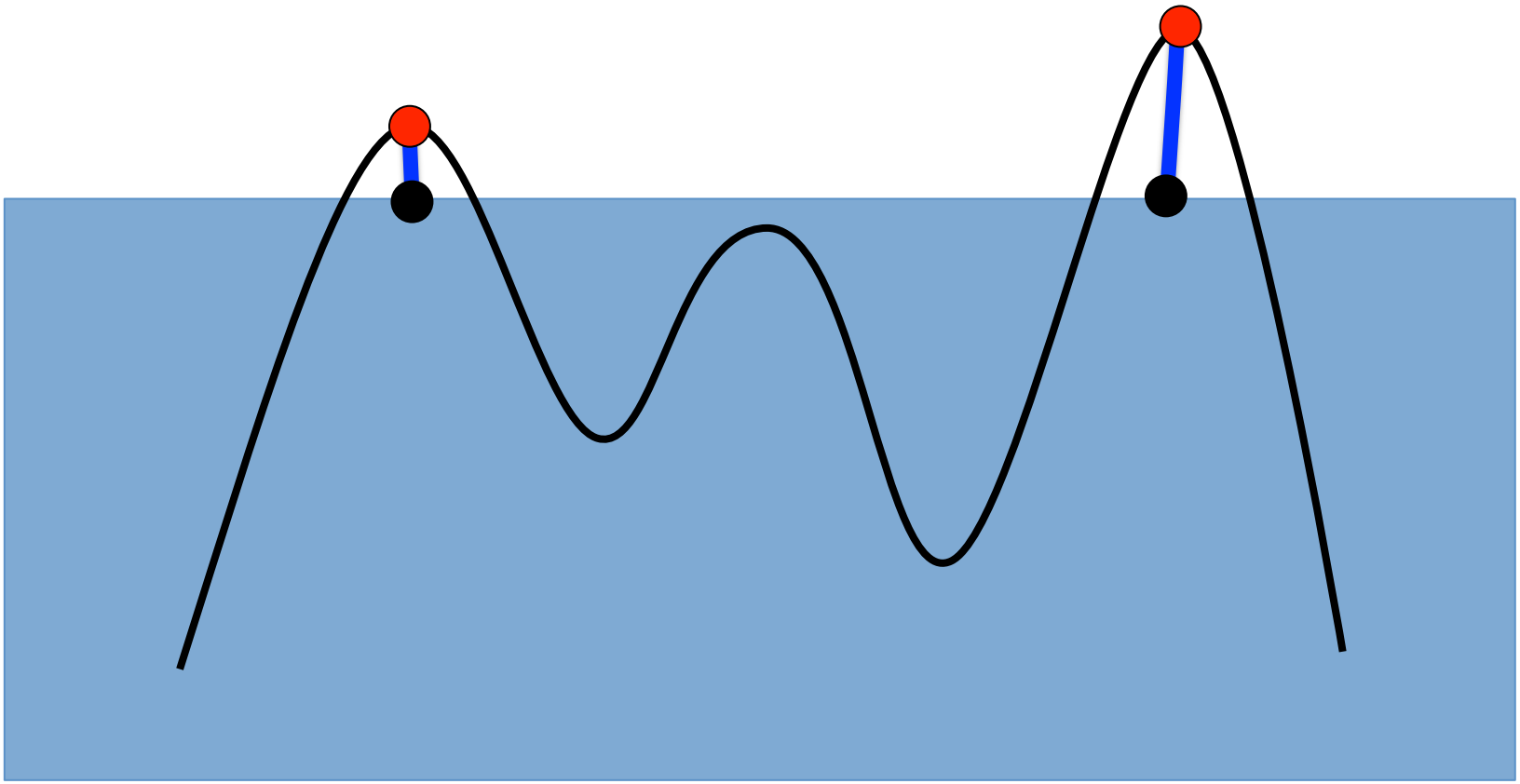
Merge Tree Construction – 1D Example

- Record connectivity of superlevel set components



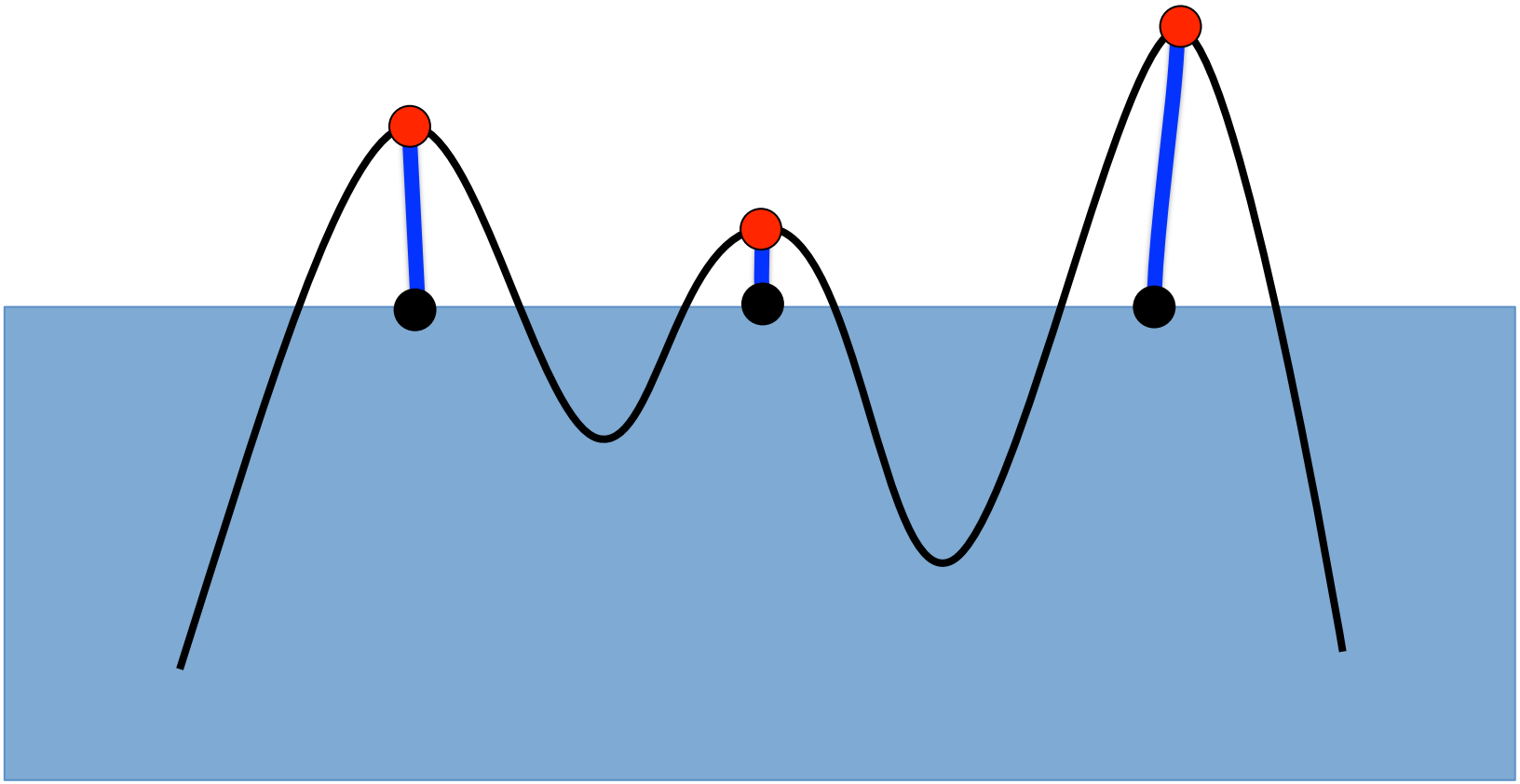
Merge Tree Construction – 1D Example

- Record connectivity of superlevel set components



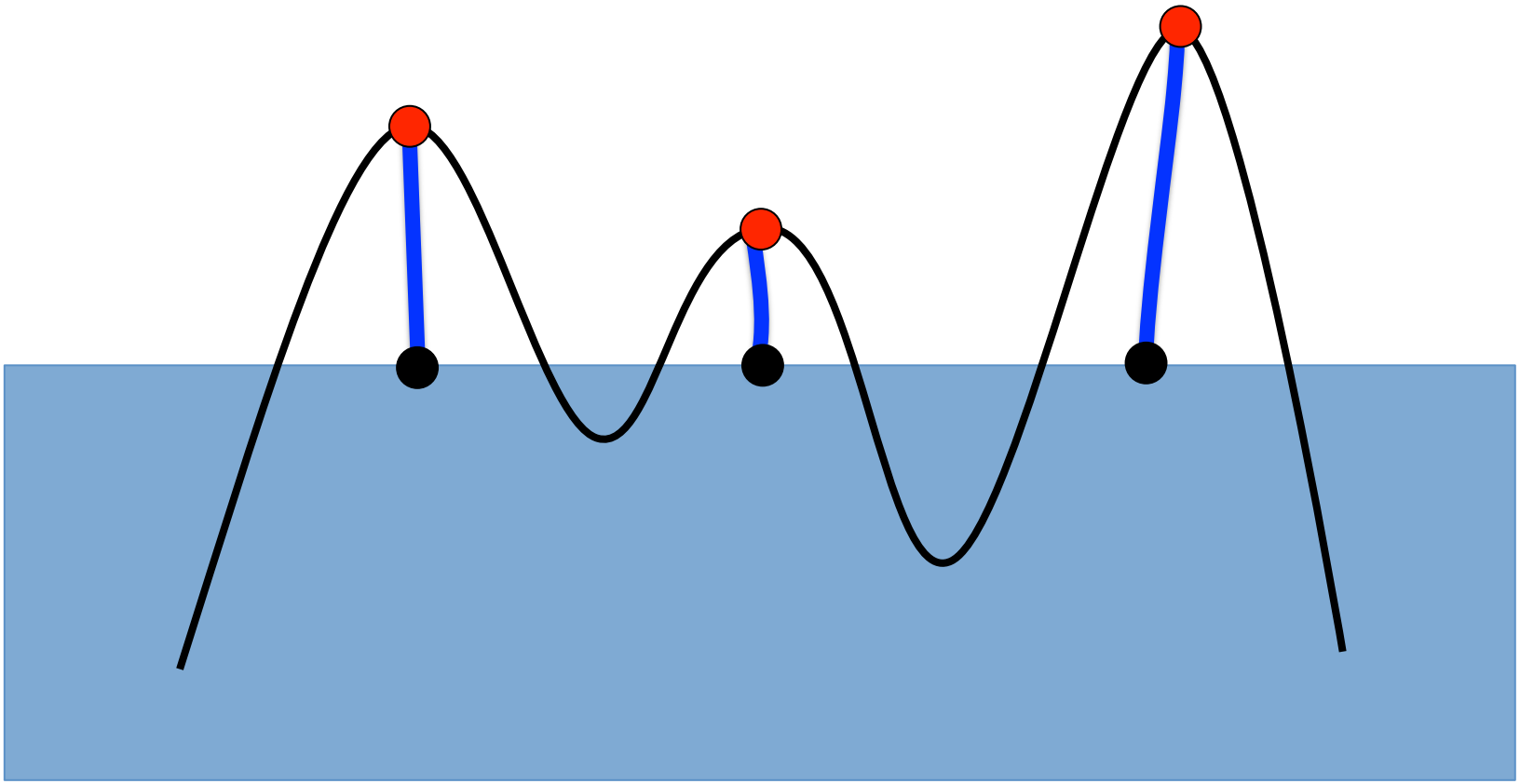
Merge Tree Construction – 1D Example

- Record connectivity of superlevel set components



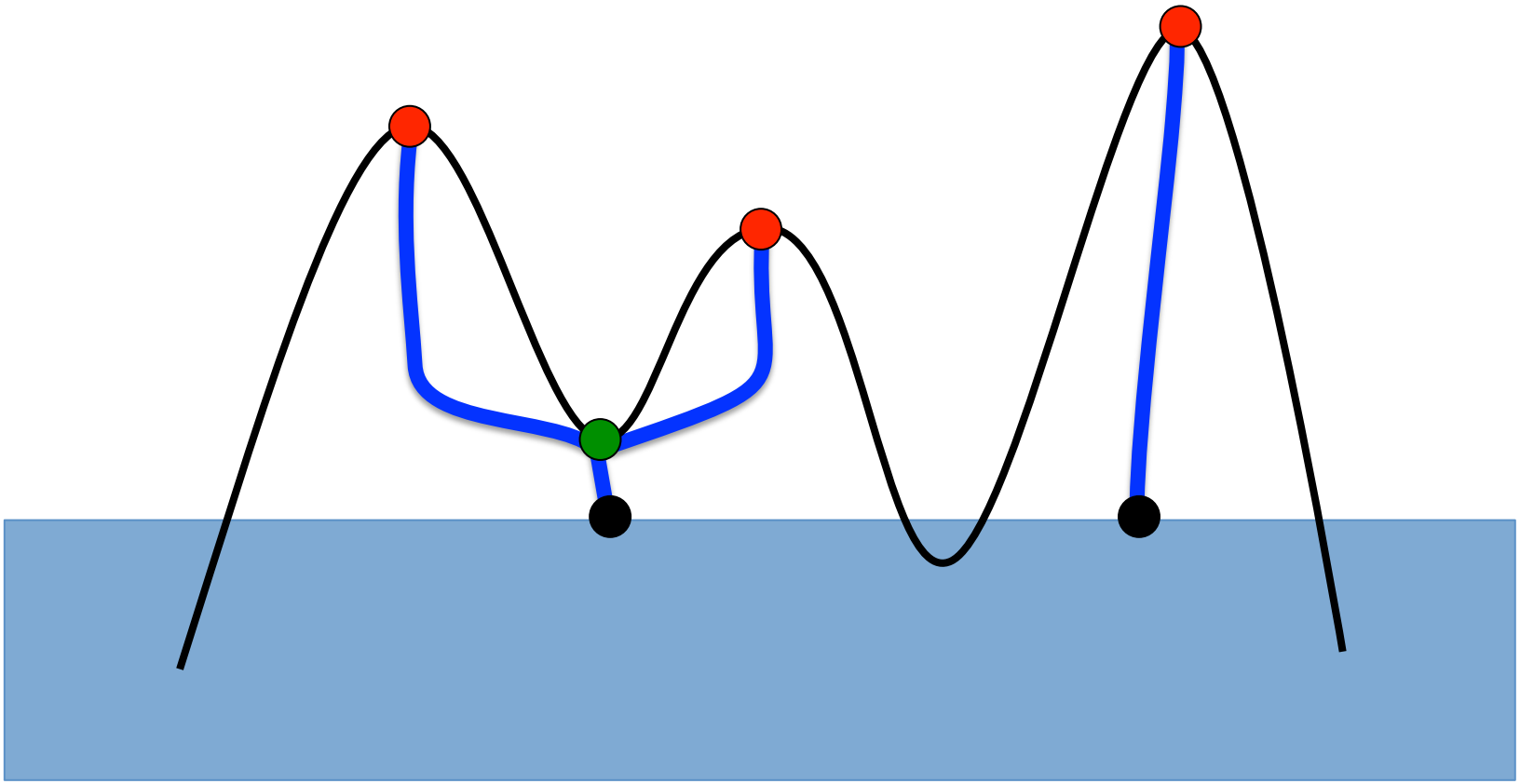
Merge Tree Construction – 1D Example

- Record connectivity of superlevel set components



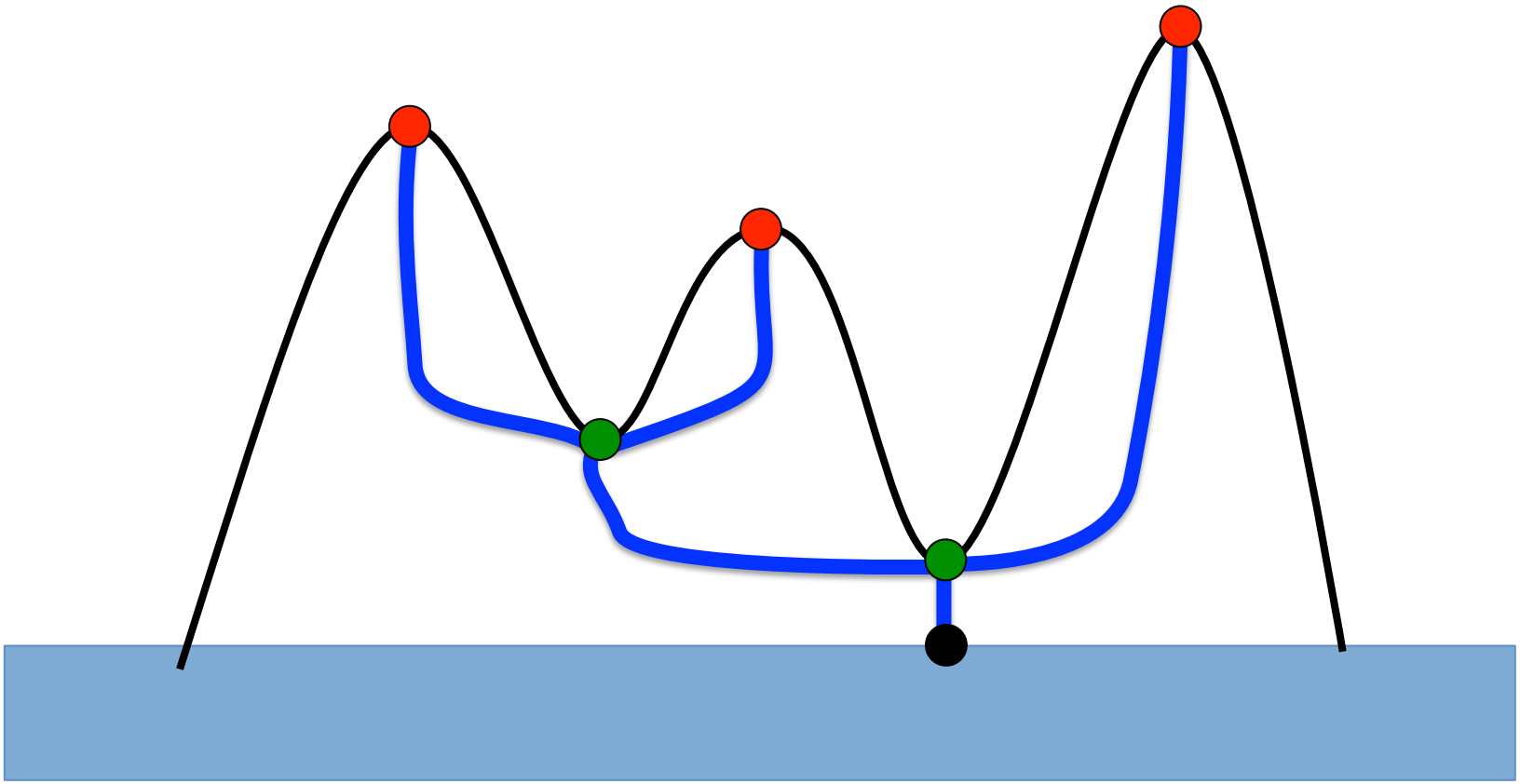
Merge Tree Construction – 1D Example

- Record connectivity of superlevel set components



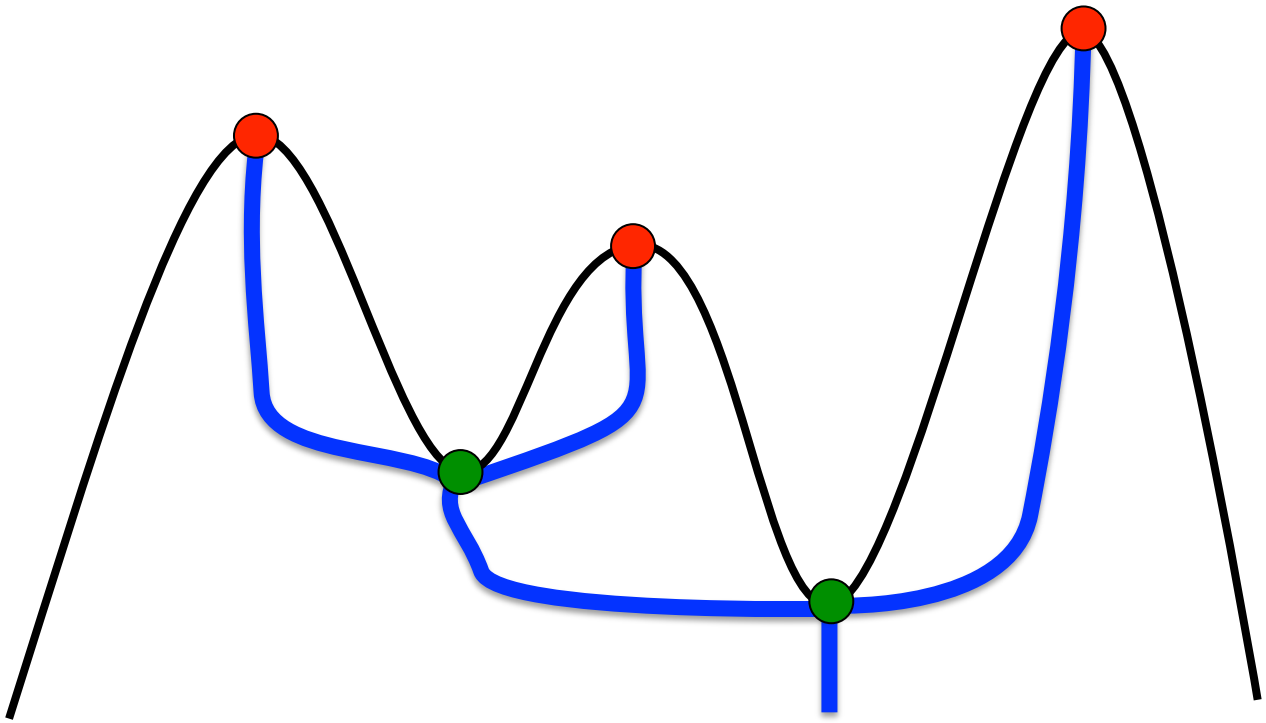
Merge Tree Construction – 1D Example

- Record connectivity of superlevel set components



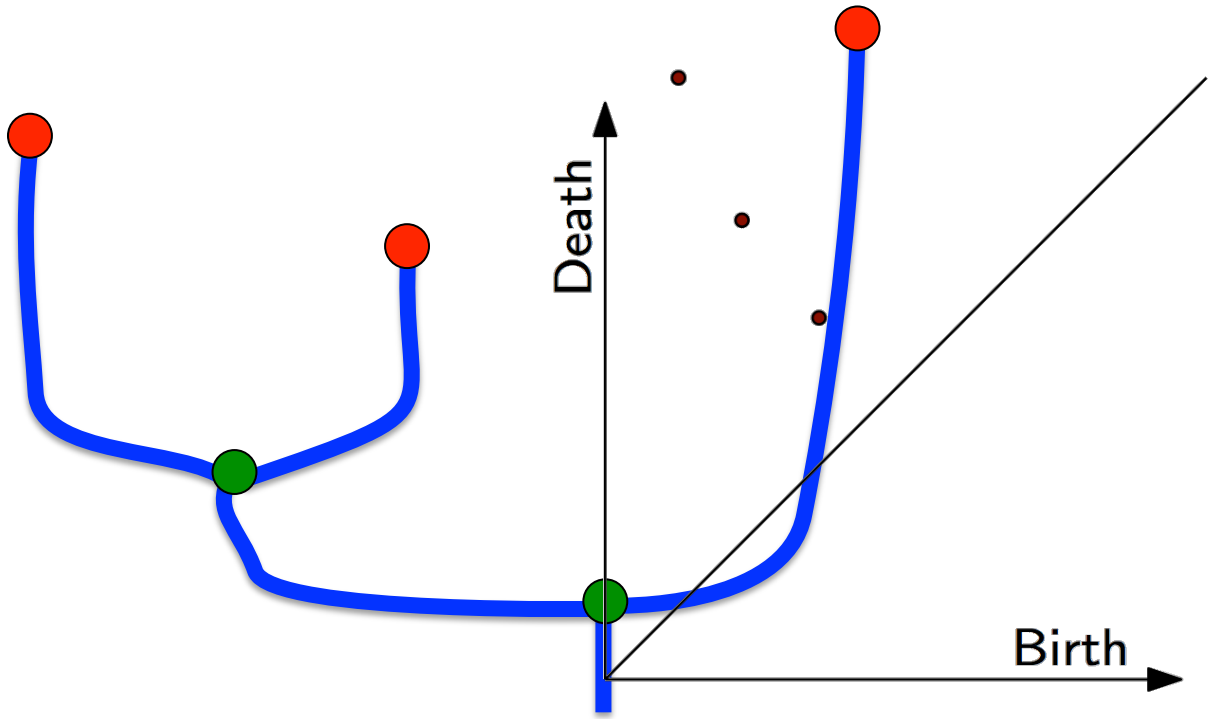
Merge Tree Construction – 1D Example

- Record connectivity of superlevel set components



Merge Tree Construction – 1D Example

- Record connectivity of superlevel set components

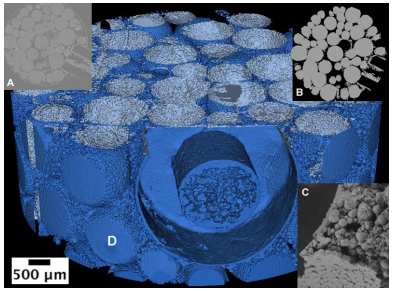


Augmented Topological Descriptors of Pore Networks for Material Science

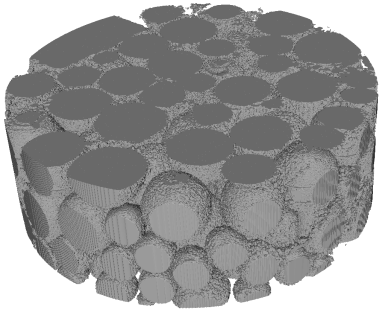
Work with D. Ushizima, D. Morozov, A. G. C. Bianchi, J. A. Sethian, E. W. Bethel

Computational Research Division, Berkeley Lab, USA

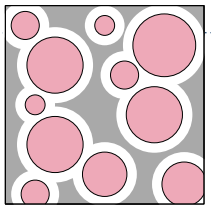
Augmented Topological Descriptors of Pore Networks for Material Science

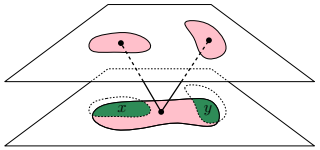


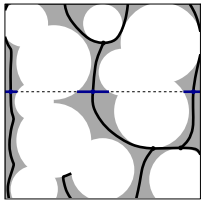
MicroCT imaging

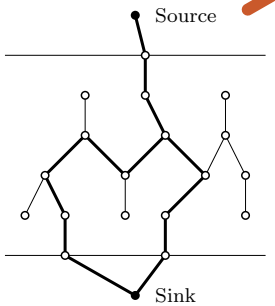


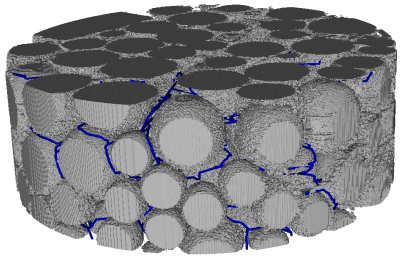
Segmentation



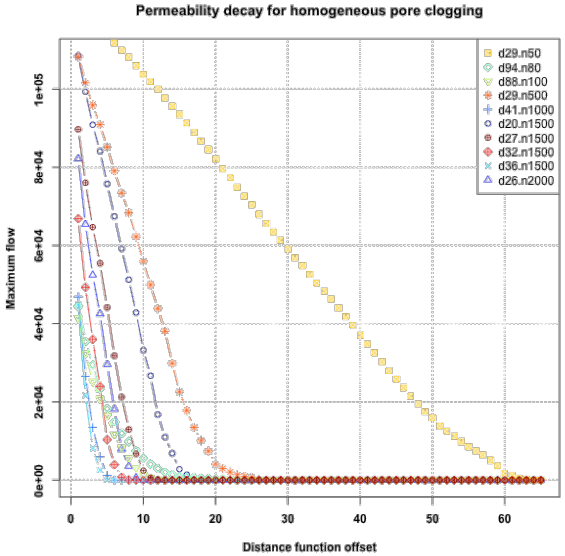




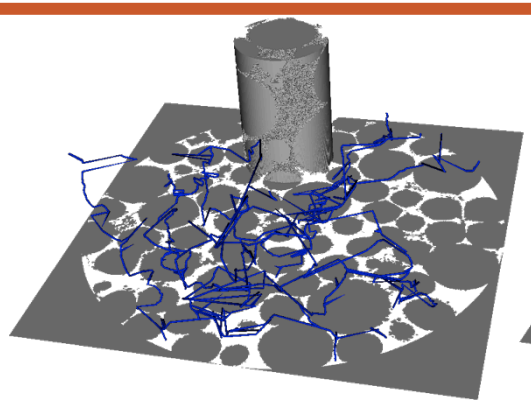


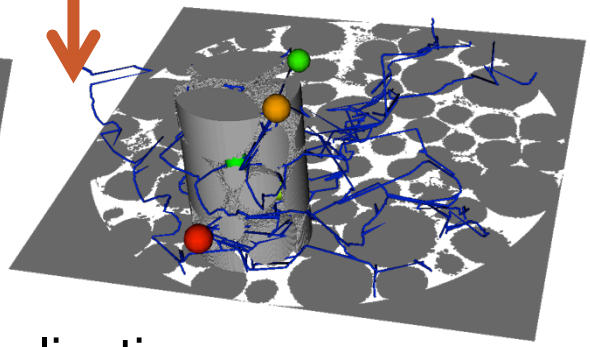


Augmented topological Descriptors



Analysis





Visualization

Conclusions

- Presented two case studies and one key technology
  - Collection of background radiation data -> FastBit indexing
  - Urban environments
  - Topological data analysis

- Urban energy modeling has closest connection to topic here

- Indexing and topological analysis have broad applications and useful for applications discussed here

Acknowledgements

This work is supported by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 through

- the “Scientific Data Management and Analysis at Extreme Scale” program (LAB 10-256), program manager Lucy Nowell
- the Scientific Discovery through Advanced Computing programs: VACET SciDAC (combustion application), SciDAC-e (carbon sequestration application).
- the use of resources at the National Energy Research Scientific Computing (NERSC) Center

We thank the members of the

- DOE Energy Frontier Research Center (EFRC) for Nanoscale Control of Geologic CO_2 (NCGC), in particular, Donald J DePaolo and J. Ajo-Franklin
- LBNL Vis Group
- LBNL Applied Numerical Algorithms Group (ANAG)
- LBNL Center for Computational Sciences and Engineering (CCSE)
- Scalable Data Management, Analysis, and Visualization (SDAV) SciDAC
- ARES project

And the VisIt developers, particularly Eric Brugger (X-ray query help)

Questions?