Topics in Computational Sustainability

CS 325
Spring 2016

Lecture 1: Intro

Course information (Administrivia)
Examples of Computational Sustainability Projects

Course Information

Lectures: Tuesdays and Thursdays - 10:30 – 11:50

No Textbook

Website: http://cs.stanford.edu/~ermon/cs325/

Instructor

Stefano Ermon
ermon@cs.Stanford.edu
Office Location: 228 Gates Hall

There will also be several guest lectures

Computational Sustainability: Goals and Topics

1. Introduce students to sustainability notions, concepts, and challenges
2. Introduce students to computational models and algorithms, in the context of sustainability topics.

Sustainability topics:

Sustainable development, renewable resources, biodiversity and wildlife conservation, poverty mitigation, energy, transportation, and climate change.

Computational topics:

Machine learning (e.g., supervised and unsupervised learning), decision and optimization problems (e.g., linear and integer programming, dynamic programming), sequential decision making under uncertainty (markov decision processes), networks (e.g., graphs and network algorithms)

Background

- How many have taken an intro to Artificial Intelligence class (CS 221)?
- How many are familiar with Machine Learning (e.g., have taken CS 229 or CS 228)?
- How many are familiar with optimization problems (e.g., convex optimization)?
- How would you rate your programming skills? Beginner / Average / Good

```
P = [0 0; 0.5 1; 1 0];
X = zeros(50000,2);
for i=1:size(X,1)-1,
  X(i+1,:) = 0.5*X(i,:) + 0.5*P(ceil(3*rand),:);
end
plot(X(:,1), X(:,2), 'b.');
```

- **Prerequisites**: familiar with mathematical modeling, algebra, calculus, probability theory etc. Basic programming skills.

$$Ax - b + x = c,$$
$$x, b, c \in \mathbb{R}^n, \quad A \in \mathbb{R}^{n \times n}$$

- Coursework and grading (tentative)
  - **Project** (60%): proposal and final report. You are free to do something related to your research. Students can choose to work on their own or in a small team. Interdisciplinary teams encouraged!
  - **Reaction paper** (20%): critically summarize a sustainability-related problem and published solution approaches. It’s a good idea to use to use the reaction paper as background research for the project.
  - **Presentation** (20%): present 1) a paper concerning a computational approach to a sustainability topic, 2) a sustainability domain and its open challenges where computation can play a role, or 3) a computational technique, model or tool that can be used to address sustainability-related problems. More details on the logistics to come.
  - **Class participation** (up to extra 10%)

What is Computational Sustainability?

A new field of research that aims to develop computational methods to help solve some of the pressing challenges concerning sustainability.

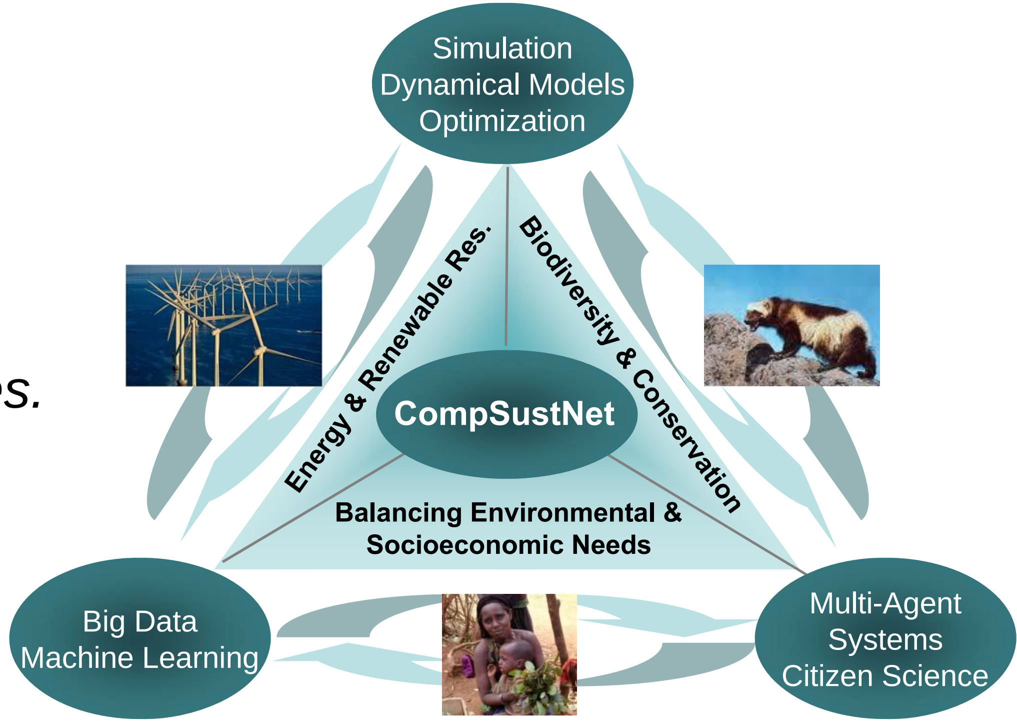
Research Themes

Core sustainability themes:

(1) Biodiversity and Conservation,

(2) Balancing Environmental and
Socio-economic Needs, and

(3) Energy and Renewable Resources.



Main computational thrusts:

(1) Big data and Machine Learning,

(2) Constraint Reasoning, Optimization,
Dynamic Control, and Simulation

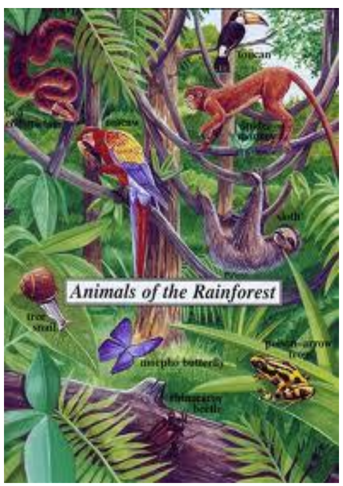
(3) Multi-Agent Systems and Citizen Science.

Examples of Computational Sustainability projects

I - Biodiversity and Species Distributions

Biodiversity or biological diversity

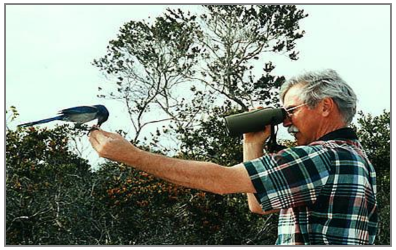
Degree of variety of life forms within a given species, ecosystem, or an entire planet.



Fundamental question in biodiversity research:
How different species are distributed across landscapes over time.

eBird: Citizen Science at the Cornell Lab. Of Ornithology



- *Increase scientific knowledge*

  Gather meaningful data to answer large-scale research questions

- *Increase conservation action*

  Apply results to science-based conservation efforts

- *Increase scientific literacy*

  Enable participants to experience the process of scientific investigation and develop problem-solving skills

The Citizen Science project at the Lab of Ornithology at Cornell empowers everyone interested in birds to contribute to research by submitting bird observations to the eBird webportal.

Bird Distributions
Machine Learning and Citizen Science

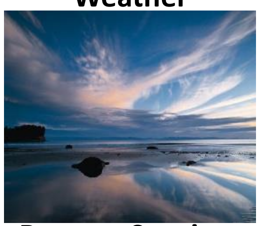
eBird
Citizen Science

Bird Observations

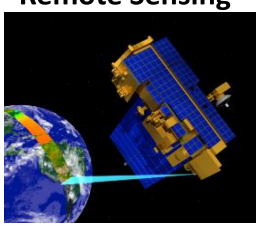
150,000+ volunteer birders	200,000,000+ bird observations	~1,500,000 hours of field work (170+years)

Land Cover

Weather



Remote Sensing



Environmental Data

TeraGrid

80,000+ CPU Hours (~ 10 Years!!!)

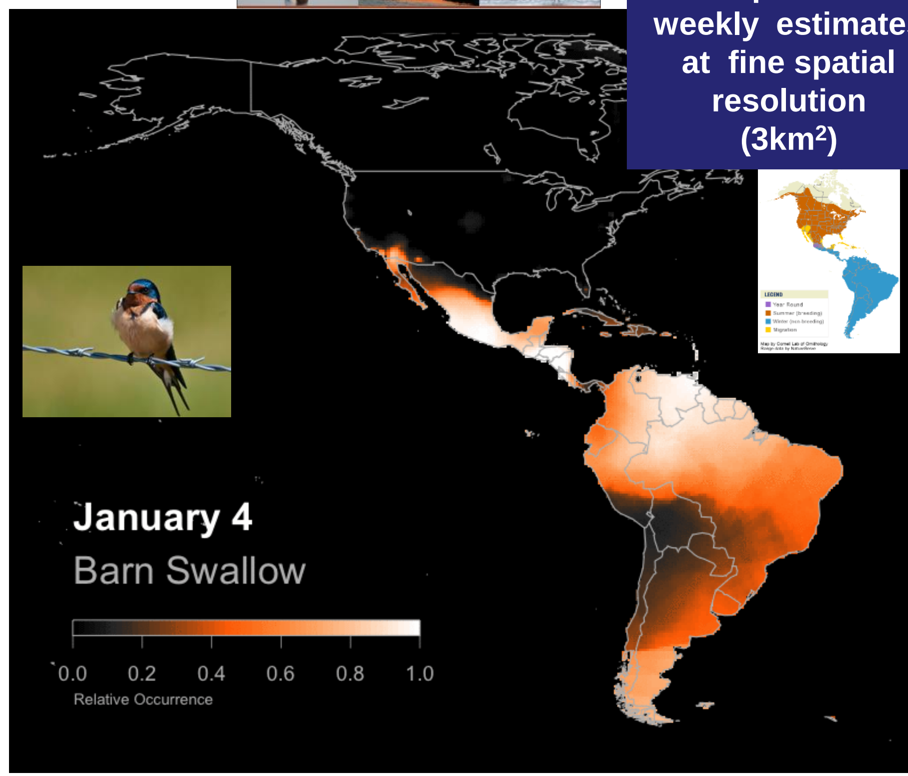
Adaptive Spatio-Temporal Machine Learning Models and Algorithms

Relate environmental predictors to observed patterns of occurrences and absences

State of the Birds Report
(officially released by Secretary of Interior)

Novel Approaches To Conservation Based on eBird Models

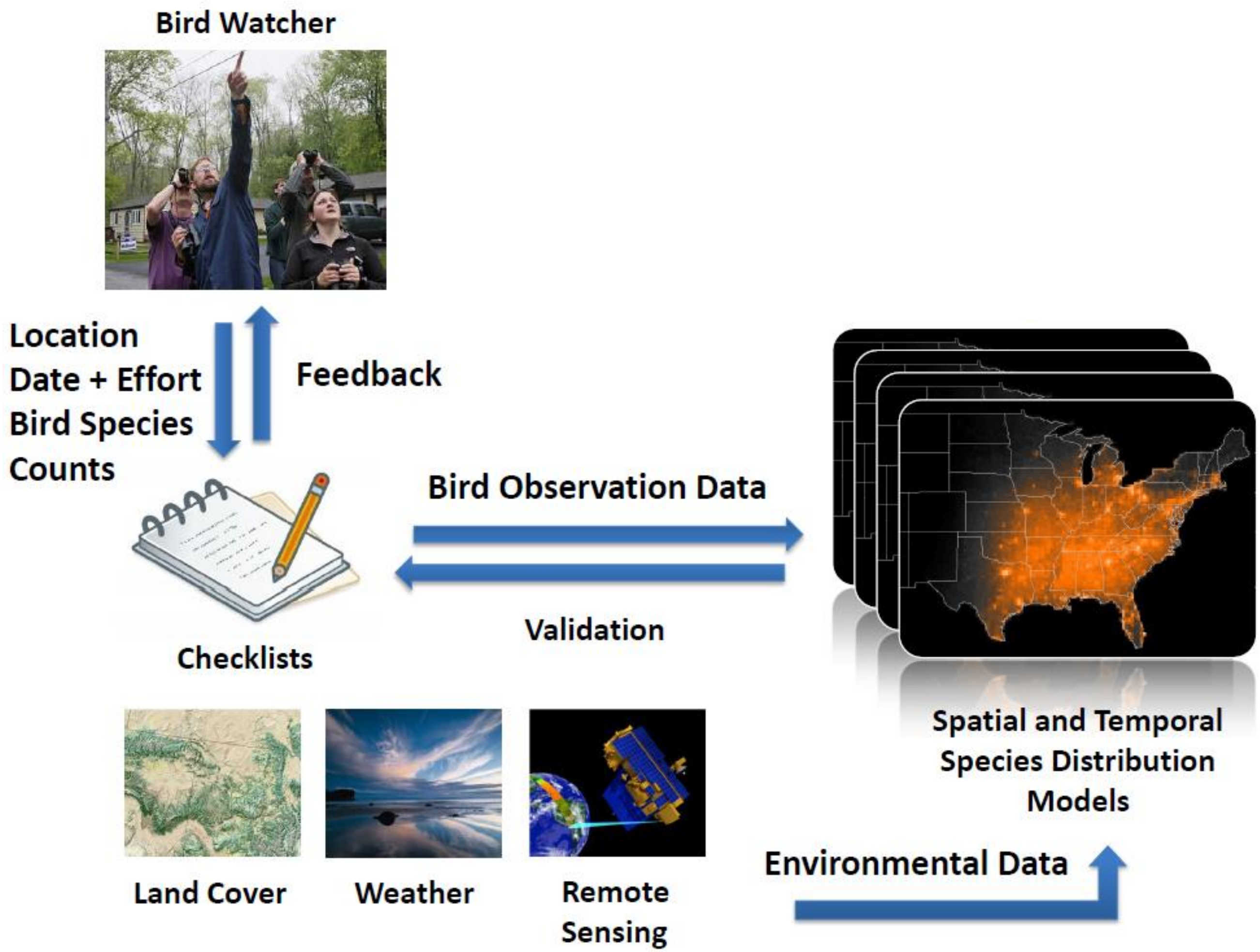
Distribution Models for 400+ species with weekly estimates at fine spatial resolution ($3km^2$)



Patterns of occurrence of the Barn Swallow for different months of the year Source: Daniel Fink

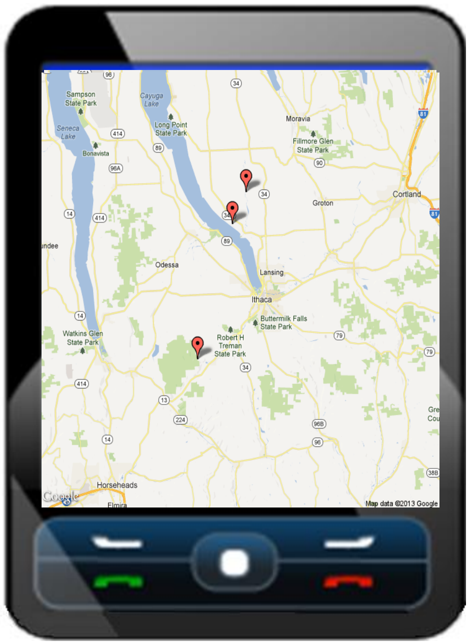
1st Time
Hemisphere Scale Bird Distribution Models, Revealing, at a fine resolution, Species' Habitat Preferences

How to Engage Citizen Scientists?
Bird-Watcher Assistant



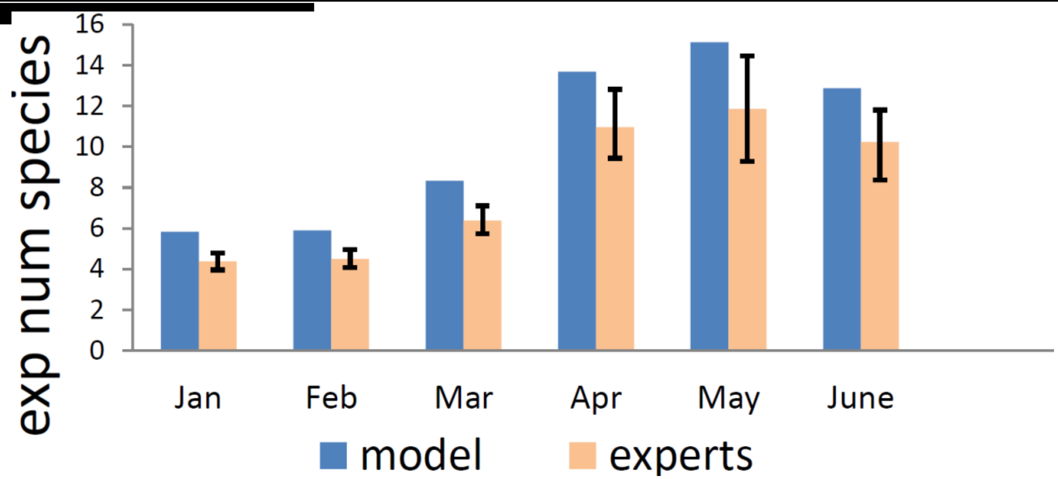
Xue et. al., HumComp 2013

Recommending Interesting sites to Birders Within a Region



Find Best Places to visit

Secondary criterion: Bird-Watcher Assistant suggests places which are not frequently visited previously, but are potentially interesting.



Suggesting interesting birding places

– *Optimization problem:*

Objective function: maximize # of different species seen

Constraint on the # of sites to visit

More species to observe compared with experts' suggestions

II - Protecting Species: Wildlife Corridor Design

Key causes of biodiversity loss:

Habitat Loss and Fragmentation

Conservation and Biodiversity : Wildlife Corridors

Wildlife Corridors

Preserve wildlife against land fragmentation

Link core biological areas, allowing animal movement between areas.

Limited budget; must maximize environmental benefits/utility

Protecting Species: Wildlife Corridors

Wildlife Corridors link core biological areas, allowing animal movement between areas.

Typically: low budgets to implement corridors.

Example:

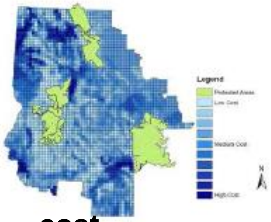
Goal: preserve grizzly bear populations in the U.S. Northern Rockies by creating wildlife corridors connecting 3 reserves:

- Yellowstone National Park
- Glacier Park / Northern Continental Divide
- Salmon-Selway Ecosystem

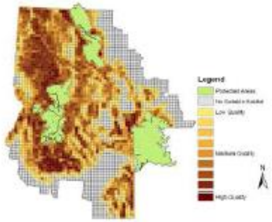
Glacier Park

Salmon-Selway

Yellowstone



cost



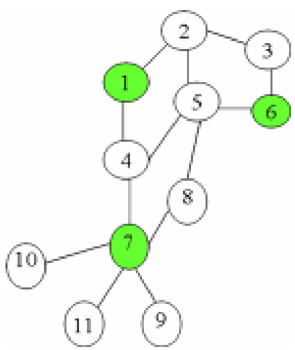
Habitat suitability

Turning a Conservation Problem Into a Computational One..

Wildlife Corridors link core biological areas, allowing animal movement between areas;

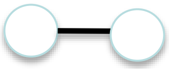
Typically: low budgets to implement corridors.

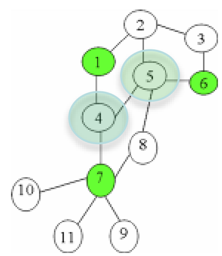
Map → "Graph"

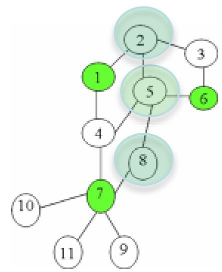


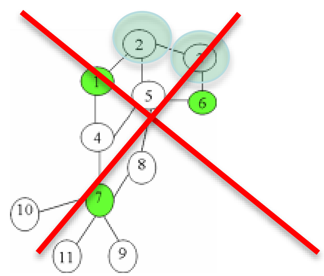
 = land patch

 = reserve

 If you can move between two patches







Connection Sub-graph Problem

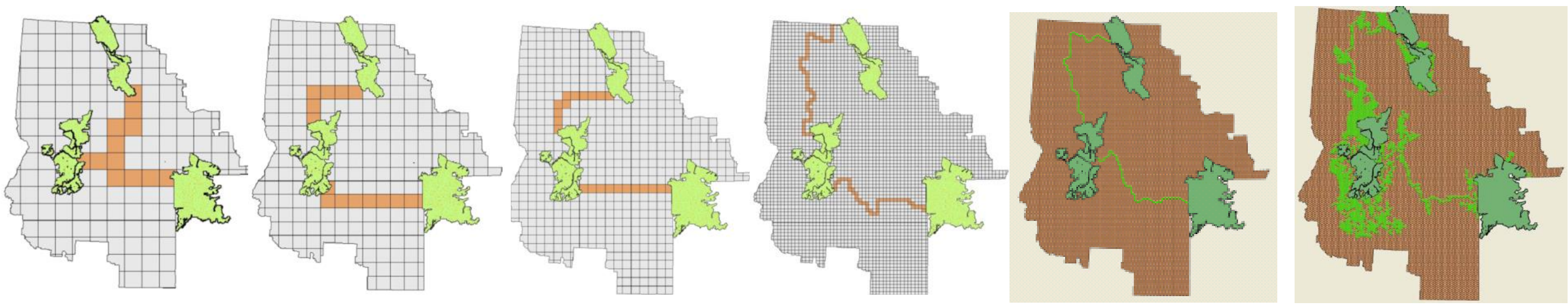
Given a graph G with a set of reserves:

Find a group of patches that:

- contains the reserves;
- is connected;
- with cost below a given budget;

(and with maximum habitat suitability)

Minimum Cost Corridor for the Connected Sub-Graph Problem



50x50 grid	40x40 grid	25x25 grid	10x10 grid	25 km^2 hex	25 km^2 hex
167 Cells	242 Cells	570 Cells	3299 Cells	1288 Cells	Extend with
$1.3B	$891M	$449M	$99M	$7.3M	2xB=$15M
<1 sec	<1 sec	<1 sec	10 mins	2 hrs	10x in Util

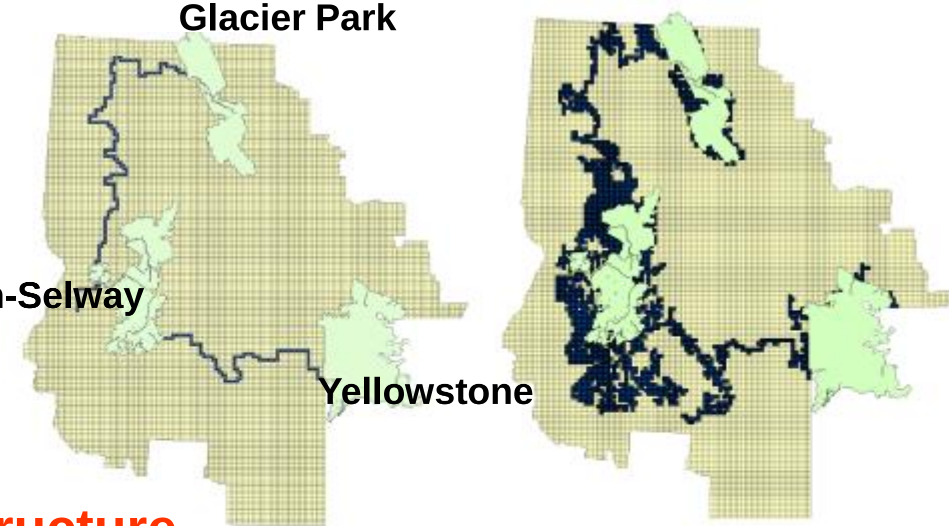
Need to solve problems large number of cells! →Scalability Issues

Real world instance:

Corridor for grizzly bears in the Northern Rockies, connecting:

- **Yellowstone**
- **Salmon-Selway Ecosystem**
- **Glacier Park**

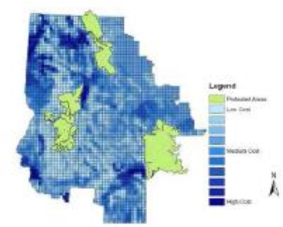
(12788 parcels) → $2^{12788} \sim 2.4 \times 10^{3726}$



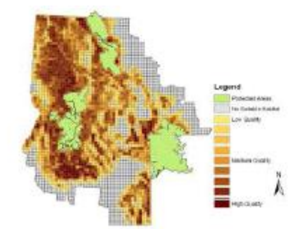
Scaling up Solutions by Exploiting Structure

Typical Case Analysis
Identification of Tractable Sub-problems
Streamlining for Optimization
Static/Dynamic Pruning

5 km grid
(12788 land parcels):
minimum cost solution
$8M



5 km grid
(12788 land parcels):
+1% of min. cost



Approach allows us to find optimal or near-optimal solutions (with guarantees) for large-scale problem instances and reduce corridor cost dramatically.

UN's Global Goals for Sustainable Development

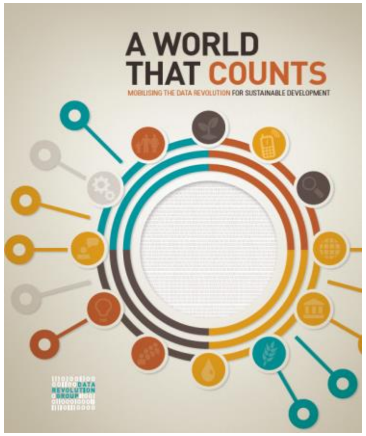
The 2030 Development Agenda (*Transforming our world)*

1. End extreme poverty
2. Fight inequality & injustice
3. Fix climate change

How measurable are these goals? How do we monitor progress?

A Data Divide is Emerging

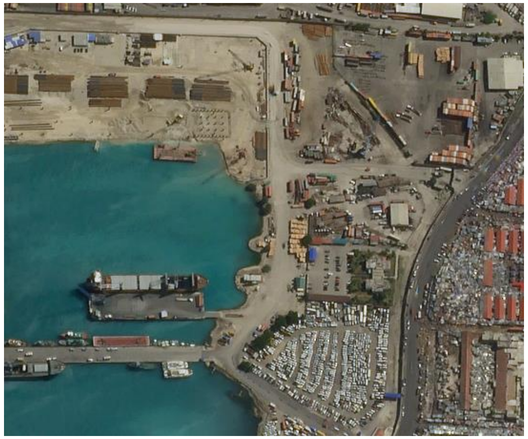
"Data are the lifeblood of decision-making and the raw material for accountability. Without data … designing, monitoring and evaluating policies becomes almost impossible"



- **Emerging data divide**: rich countries are flooded with data (*Big Data*), while developing countries are suffering from **data drought**
  - We have sensors in phones, watches, cars, thermostats, …
  - Afghanistan is still using census figures from 1979 (a count cut short after census-takers were killed by mujahideen)
  - Nearly 230 million births have gone unrecorded in the last 5 years
  - Botswana's poverty figure is extrapolated from data collected in 1993

Remotely Sensed Data

Remote sensing (e.g., satellite imagery) is among the few cost-effective technologies able to provide data at a global scale

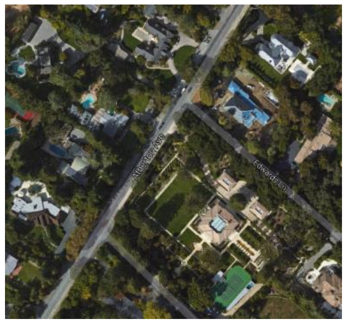


Becoming **increasingly accurate** and **cheap** (SpaceX, PlanetLabs, SkyBox, …). New opportunities for modeling global-scale phenomena.

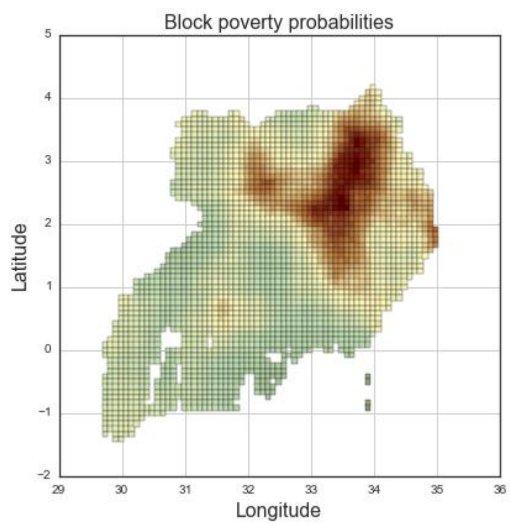
Is it possible to infer socioeconomic indicators (poverty, child mortality, etc.) from large-scale remotely sensed data?

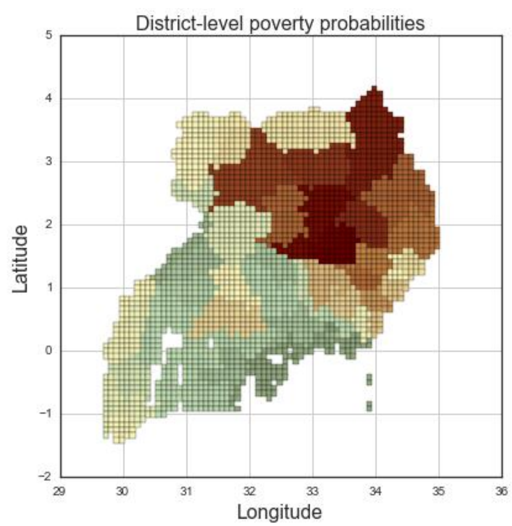
Focus on Poverty

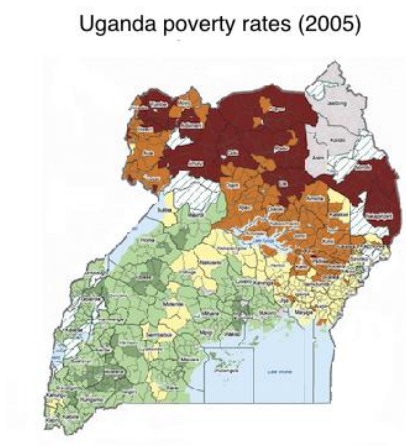
First step: infer household income and poverty from satellite imagery

Example:  vs. 

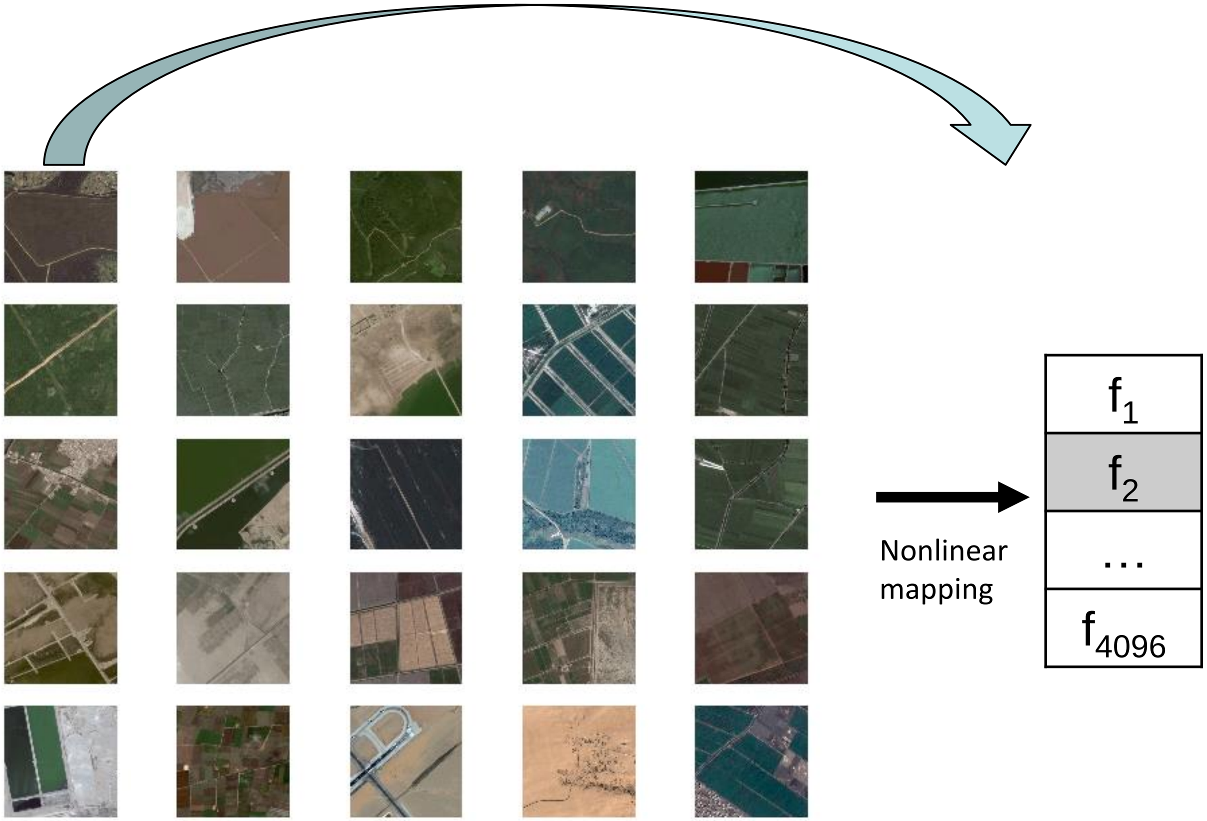
Do this **at scale, accurately** and with **unprecedented spatial resolution**:





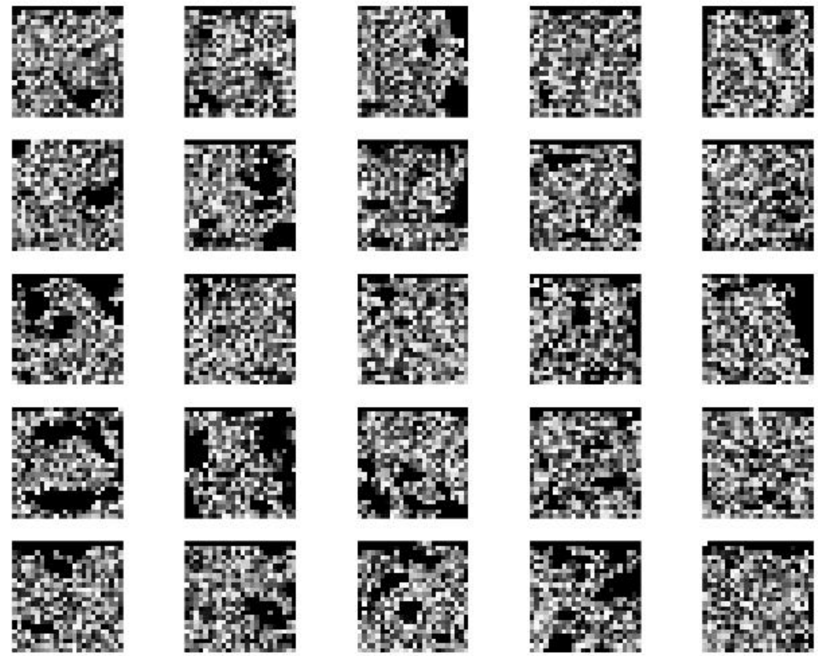


Learned Features: Roads



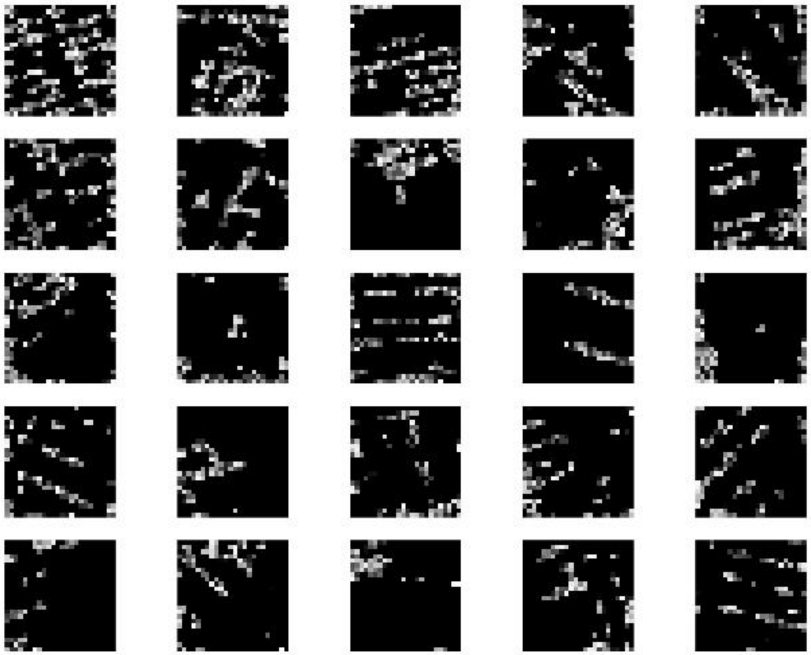
25 Maximally activating images

Learned Features: Urban Areas



Maximally activating images

Learned Features: Farmland



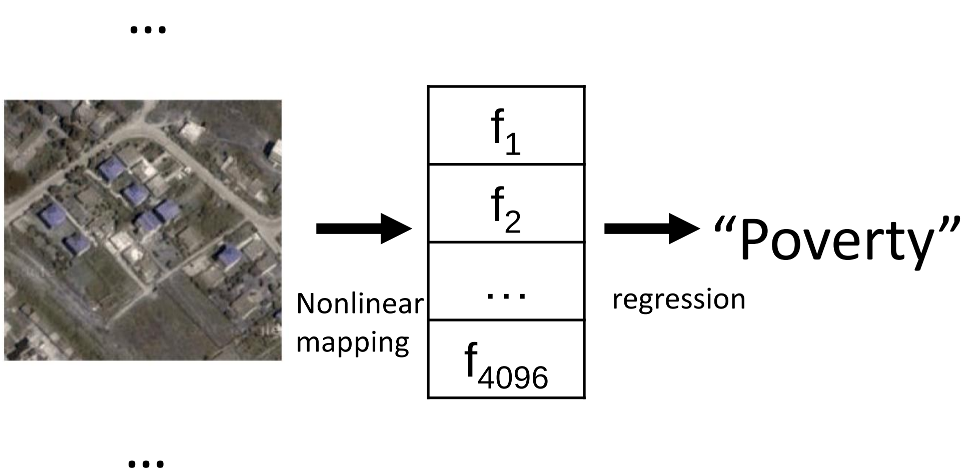
Maximally activating images

Can we use this knowledge for poverty estimation?

Poverty Estimation

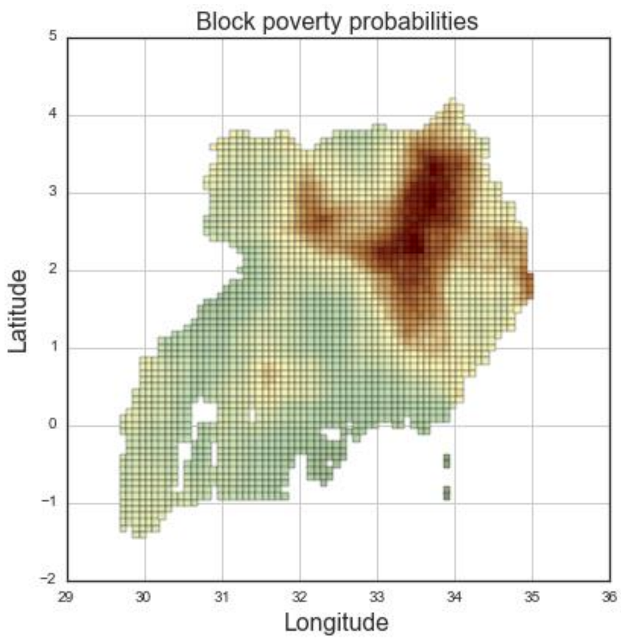
- Living Standards Measurement Survey (LSMS) data in Uganda (World Bank):
  - ~700 data points (enumeration areas)
  - Expenditures, above/below poverty line, coordinates

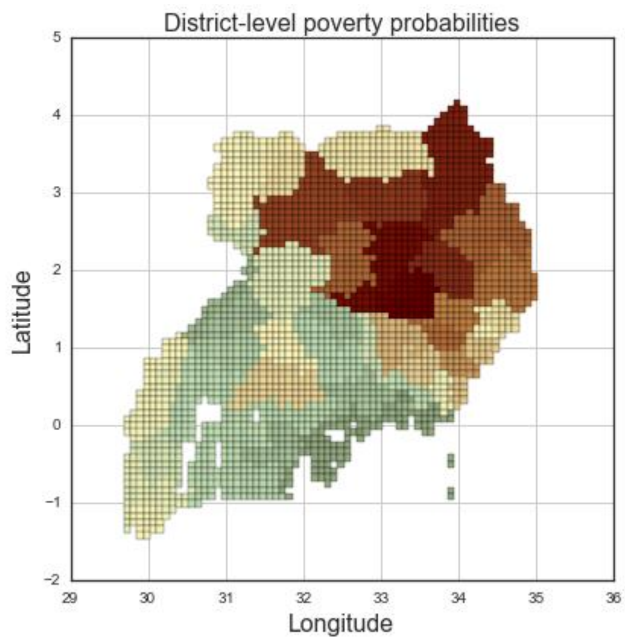
- **Task**: predict if the majority of households in an enumeration area are above or below the poverty line (from corresponding images)

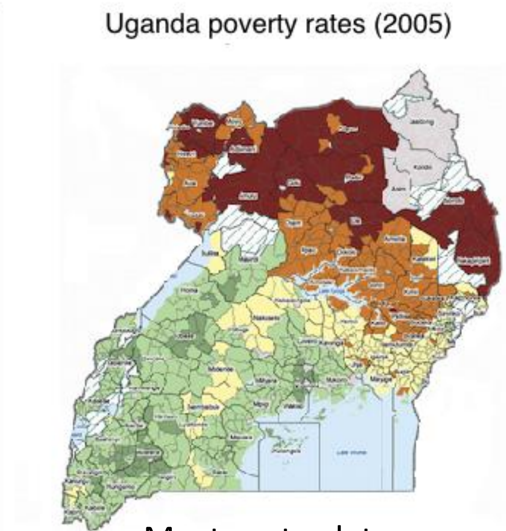


High Resolution Poverty Maps

Run the model on about 500,000 images from Uganda:





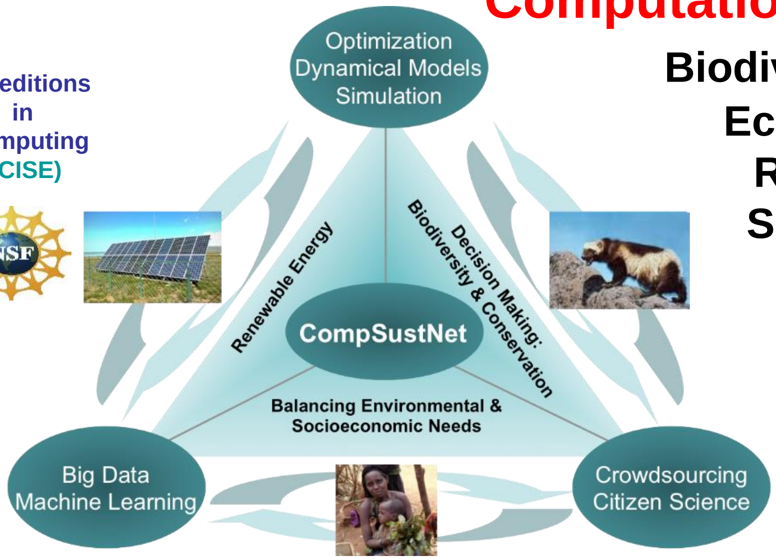


Most up-to-date map

Scalable and inexpensive approach to generate high resolution maps.

Computational Sustainability Network

Expeditions in Computing (CISE)



Biodiversity and Conservation
Economic Development
Renewable Energy &
Sustainable Materials

Cross-Cutting Computational Models/Algorithms
Leveraging them across Applications

Materials Discovery

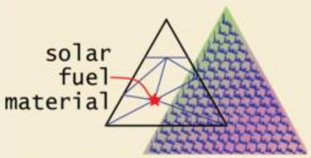


Photo: John Gregoire (JCAP/Caltech)

What: Rapid characterization of crystal structures from high-throughput X-ray diffraction experiments.
Why: Identify new materials for fuel cells, energy storage, and solar fuel generation.
How: Pattern decomposition, constraint and probabilistic reasoning, crowdsourcing.

Smart Grid

What: Power grid modeling, control, and energy storage.
Why: Managing the power system with increasing use of renewable sources of electricity.
How: Stochastic optimization, sequential decision making, pattern decomposition.

Photo: DOE

Big Data for Africa

Photo: Frank Annor (TAHMO)

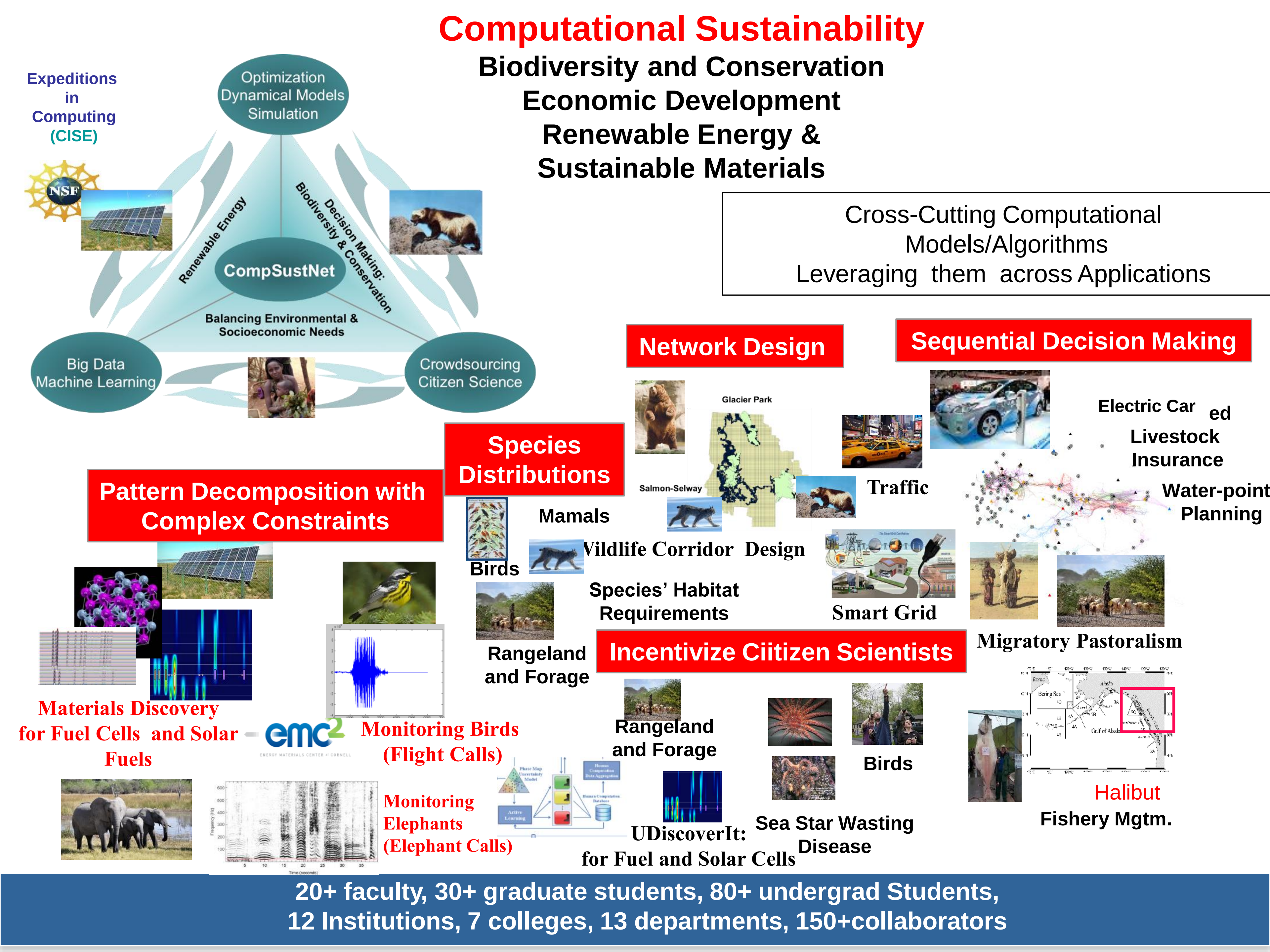
What: Deploy 20,000 low-cost weather stations across Africa.
Why: Improve weather predictions, which is directly related food security.
How: Optimal placement, bayesian networks, multi-scale probabilistic modeling.

Landscape-Scale Conservation

What: Socio-ecological corridor in the Ecuadorian Andes.
Why: Protect endangered Andean bear and other species in a significant biodiversity hotspot, while improving livelihoods of local communities.
How: Spatial capture-recapture, stochastic optimization, spatio-temporal modeling.

Photo: Santiago Molina

20+ faculty, 30+ graduate students, 80+ undergrad Students,
12 Institutions, 7 colleges, 13 departments, 150+collaborators



Computational Sustainability
Biodiversity and Conservation
Economic Development
Renewable Energy &
Sustainable Materials
Expeditions in Computing (CISE)
NSF
Optimization
Dynamical Models
Simulation
Renewable Energy
Decision Making: Biodiversity & Conservation
CompSustNet
Balancing Environmental & Socioeconomic Needs
Big Data
Machine Learning
Crowdsourcing
Citizen Science
Cross-Cutting Computational Models/Algorithms
Leveraging them across Applications
Network Design
Sequential Decision Making
Glacier Park
Salmon-Selway
Traffic
Electric Car
Livestock Insurance
Water-point Planning
Species Distributions
Pattern Decomposition with Complex Constraints
Mamals
Birds
Wildlife Corridor Design
Species' Habitat Requirements
Smart Grid
Migratory Pastoralism
Rangeland and Forage
Incentivize Ciitizen Scientists
Materials Discovery for Fuel Cells and Solar Fuels
emc2
Monitoring Birds (Flight Calls)
Rangeland and Forage
Birds
Halibut
Fishery Mgtm.
Monitoring Elephants (Elephant Calls)
UDiscoverIt: for Fuel and Solar Cells
Sea Star Wasting Disease
20+ faculty, 30+ graduate students, 80+ undergrad Students,
12 Institutions, 7 colleges, 13 departments, 150+collaborators

More examples