Computational Sustainability: Computing for a Better World

Carla P. Gomes
Institute for Computational Sustainability
Department of Computer Science
Cornell University

Expeditions in Computing (CISE)

Sustainability in the Digital Age
Resilience facing Global Changes
Montreal 2019

Dramatic Progress in AI

Rapid shift in AI research:

Academic world → Real-world

- **Machine perception is starting to work: finally! after "only" 50+ yrs of research…**

  →AI systems are starting to "*see*" and "*hear*" *(computer vision, speech recognition, natural language understanding)*

- **Our systems are finally becoming *grounded in (our) world.* Already:**
  - **super-human face recognition (Facebook)**
  - **super-human traffic sign recognition (Nvidia)**

- **Enabled by qualitative change in the field, driven in part by Big Data & Deep Learning but also other cumulative progress (reasoning, search, reinforcement learning, planning, decision theoretic methods, knowledge representation)**

Emergence of Intelligent Machines

Intelligent systems are radically transforming businesses, medicine, …

Wall Street: Autonomous Trading Systems

Automated Supply Chain

Assistive robotics Remote Robotic Surgery

Genome sequencing

And our daily lives ….

Unfortunately, the digital and AI revolution have done little for Sustainability

Our vision:

AlphaGo

Computer Science and AI can — and should — play a key role in helping address societal and environmental challenges in pursuit of a sustainable future,

while also **advancing computer science as a discipline.**

Thank you!

1st Expeditions in Computational Sustainability (2008)

Expeditions in Computing (CISE)

Conference,
Referred CompSust Tracks,
Workshops

- **To nucleate** the Computational Sustainability field
- **To identify** a number of **core research directions** for **maximal impact**, both in terms of **Computer Science** and **Sustainability.**

2nd Expeditions: Large-Scale Research Network for Expanding the Horizons of Computational Sustainability

Cornell University, Caltech, Carnegie Mellon University, Bowdoin, Georgia Tech, Howard University, The Ohio State University, OSU Oregon State University, Stanford University, Princeton University, UMass Amherst, USC University of Southern California, Vanderbilt University

and Gov and NGOs and several International Universities as collaborators

CompSustNet

300+ faculty, students, and collaborators!!!

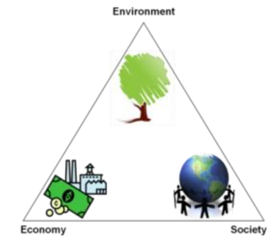
New interdisciplinary field that aims to develop computational methods for Sustainable Development.

Sustainable development *is development that meets the needs of the present without compromising future generations.*

United Nations, Our Common Future, 1987

Sustainable Development encompasses balancing

- *environmental,*
- *economic, and*
- *societal needs.*



Ultimate goal of Sustainable Development HUMAN WELL-BEING of current and future generations.

Expeditions in Computing (CISE)

2008/2016

https://sustainabledevelopment.un.org/sdgs (2015)

Sample of Interdisciplinary Research Projects @ Cornell

I Conservation and Biodiversity

Wildlife Corridors
Bird Conservation
Protecting Endangered Species

The Nature Conservancy

eBird

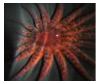
CORNELL LAB of ORNITHOLOGY

Bird Conservation

Wildlife Corridors
Social-Economic
Ecological Corridor
(Ecuador)

WWF

Elephant
Listening Project



II Balancing Environmental-and-Socio-Economic Needs

Impacts of Hydropower Dam Placement in the Amazon Basin

Protecting Migratory Herders in Africa

Socio-Econ-Environmental
Impacts of Dams in Amazon Basin

Herders in Africa

III Accelerating Discovery of Materials for Renewable Energy

Accelerating the Discovery of Solar Fuels

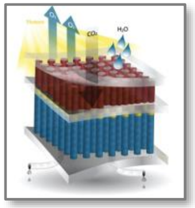
SARA: Scientific Autonomous Reasoning Agent for Materials Discovery

JCAP JOINT CENTER FOR ARTIFICIAL PHOTOSYNTHESIS

Caltech

CHESS

Cornell High
Energy Synchrotron



Solar Fuels

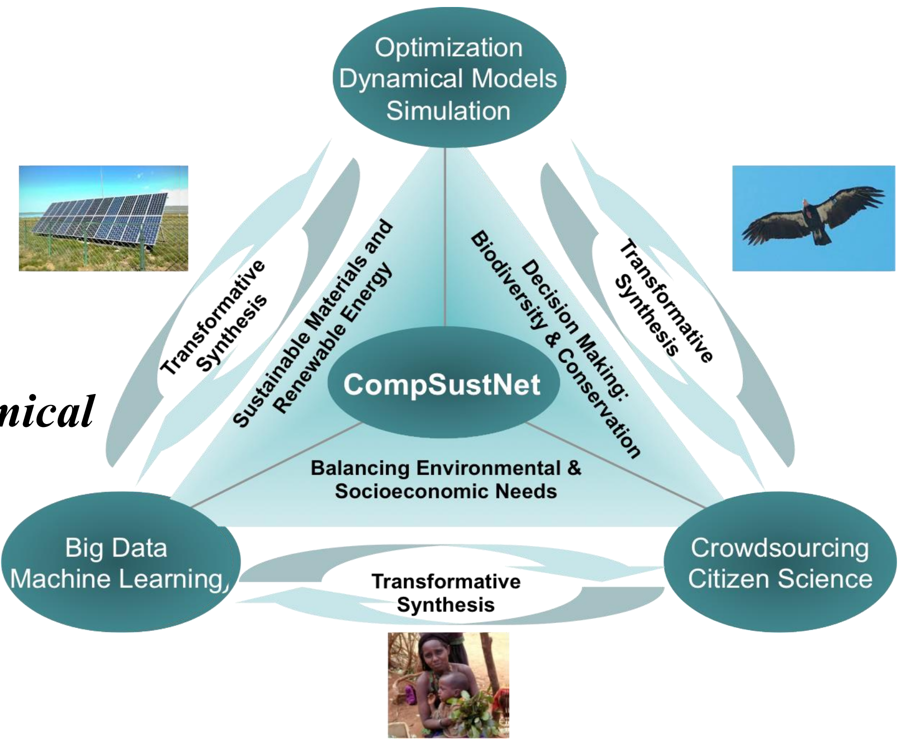
SARA

Wide range of sustainability applications covering
Cross-Cutting Core Computational Problems

3 Core Computational Thrusts

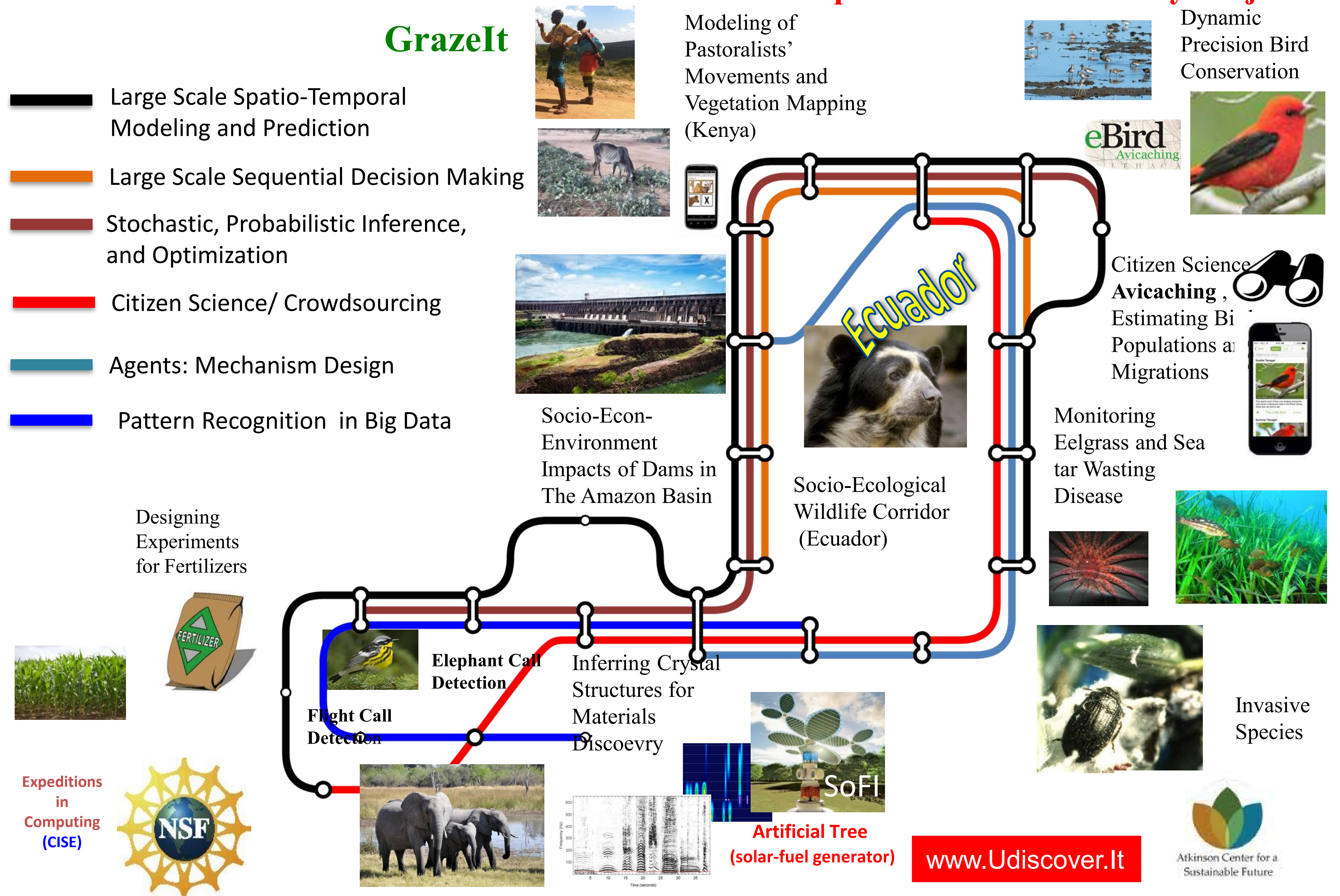
Main computational thrusts:

(1) Big data and Machine Learning
(2) Constraint Reasoning, Optimization, Dynamical Models and Simulation
(3) Multi-Agent Systems, Citizen Science, and Crowdsourcing



Interdisciplinary Research Projects (IRPs)
lead to transformative syntheses
across sustainability domains and computer science sub-areas

Subway Lines: Examples of Cross-Cutting Computational Themes and Interactions of some Computational Sustainability Projects



Outline

Impact of Hydropower Dam Placement in the Amazon Basin on Ecosystem Services

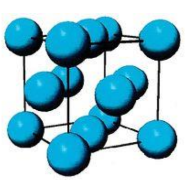
- Multifaceted “trust” in **decision support systems**
- Optimization with Multiple (and Conflicting) Objectives: Computing The Pareto Frontier

Species distributions

- Reducing Bias in Citizen Science Data:
  - Avicaching Game
  - Co-variate shift
- Multi-Entity Dependency Learning: Deep Multivariate Probit Model

Inferring Crystal Structures for Materials Discovery for Clean Energy

- Constrained Pattern Decomposition
- Human Computation for Speeding up Search



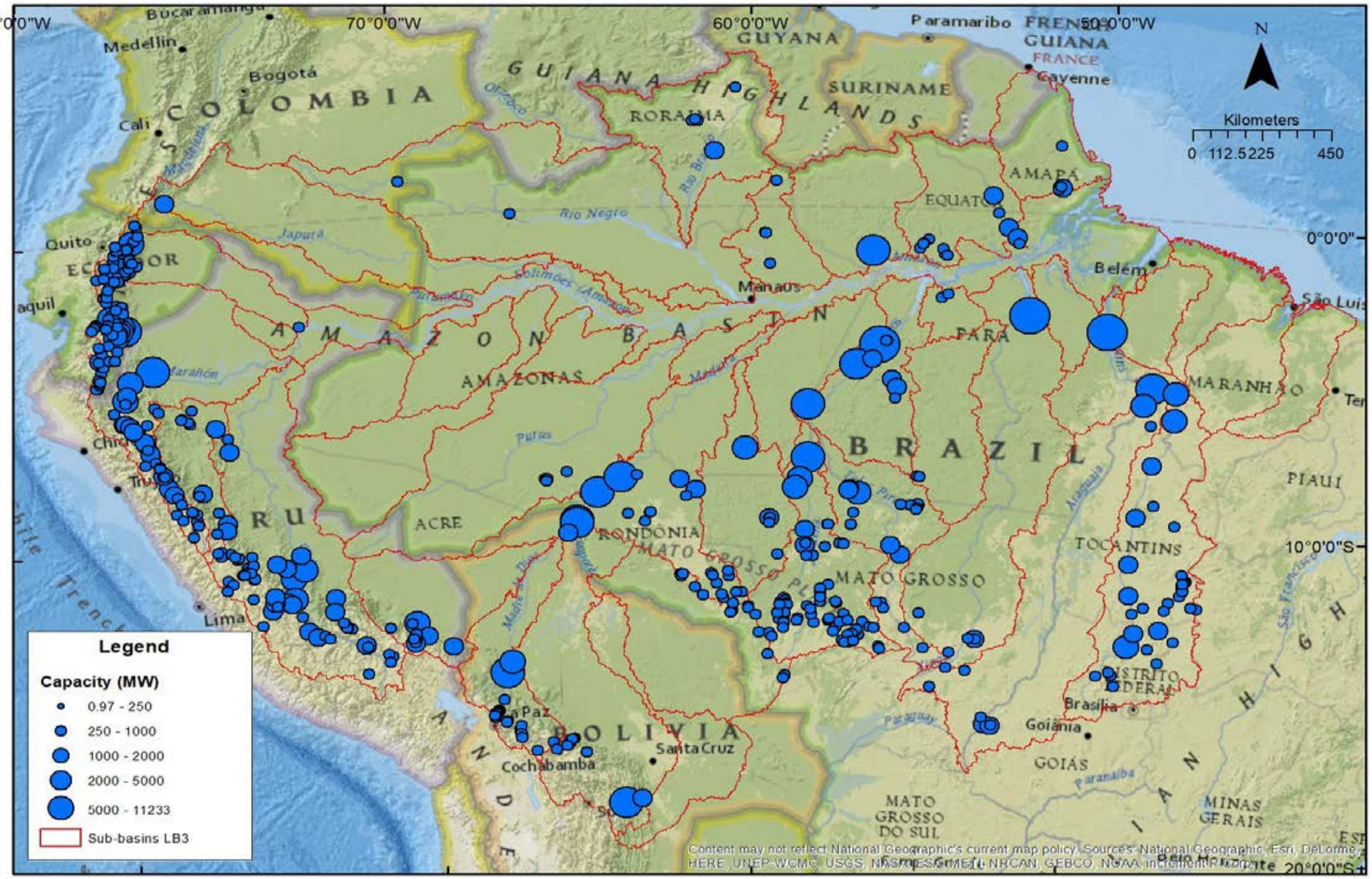
FCC Crystal Structure

Sustainability concerns balancing environmental, economic, and societal needs

Key Issue:

Understanding trade-offs of solutions wrt multiple (and often conflicting) objectives

Hydropower Dam Proliferation in the Amazon Basin



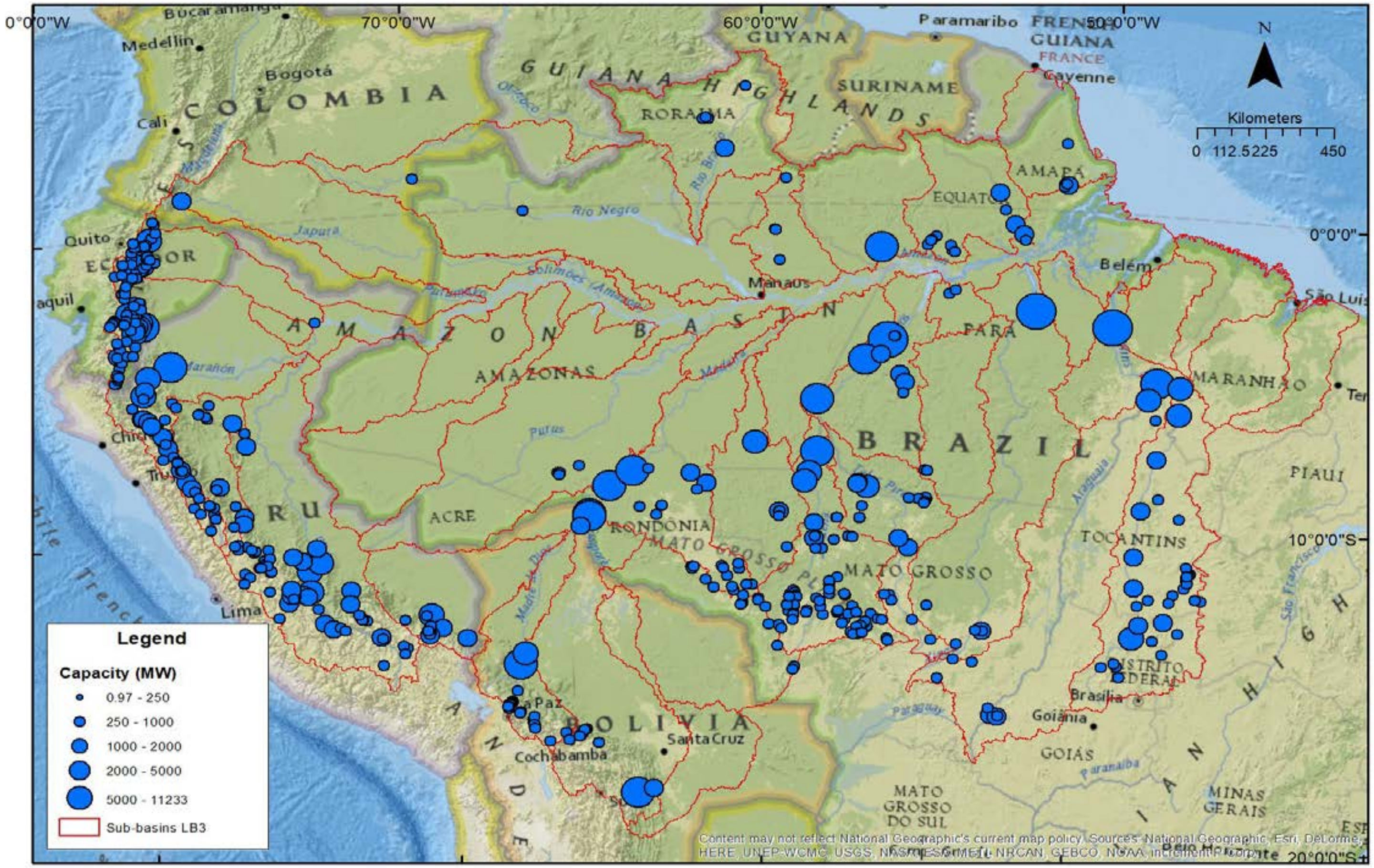
Hydropower Dam Proliferation in the Amazon Basin

170 dams already built or under construction

Hydropower Dam Proliferation in the Amazon Basin

500+ dams planned or proposed

Hydropower Dam Proliferation in the Amazon Basin

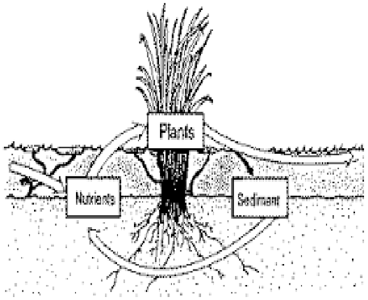


Ecosystem Services of River Networks

Energy

Fisheries

Transportation and navigation



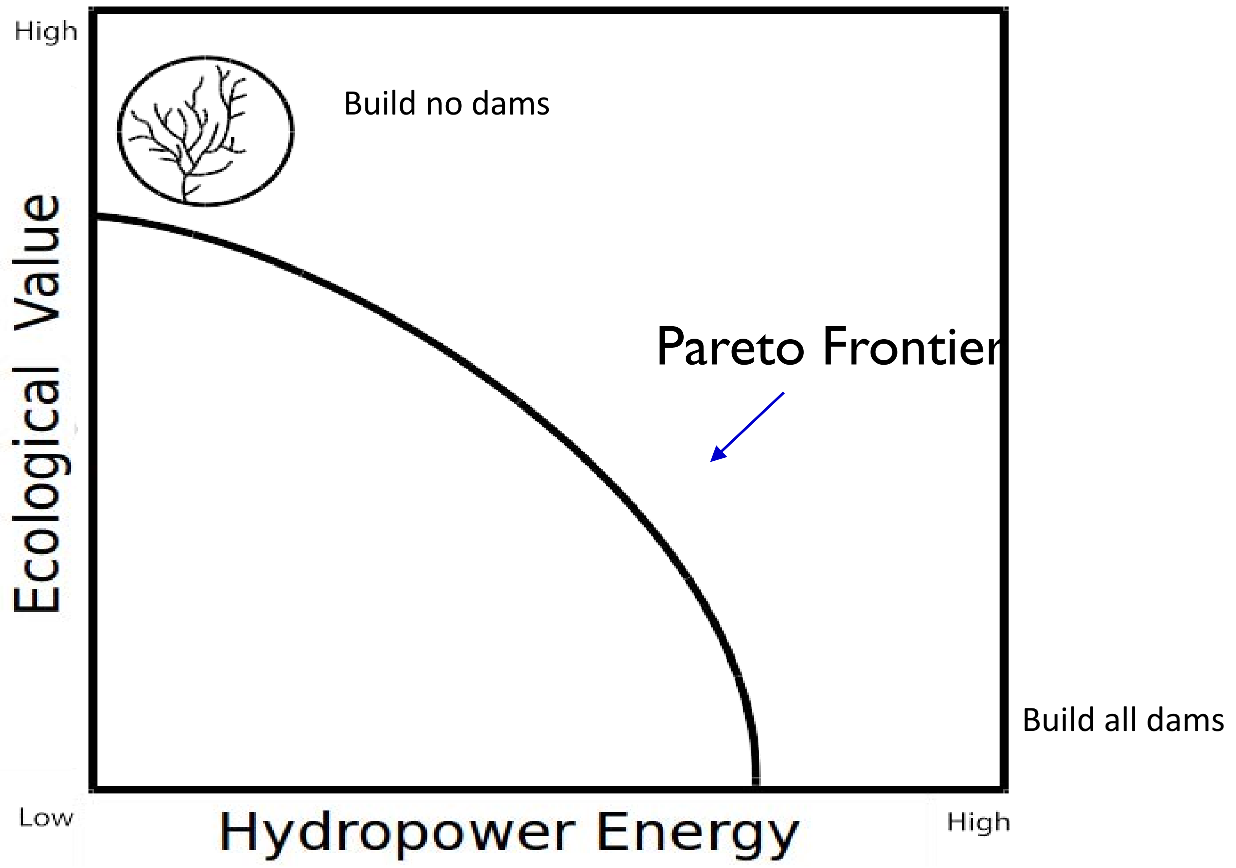
Sediments and Nutrients

Examples of Ecosystem Services

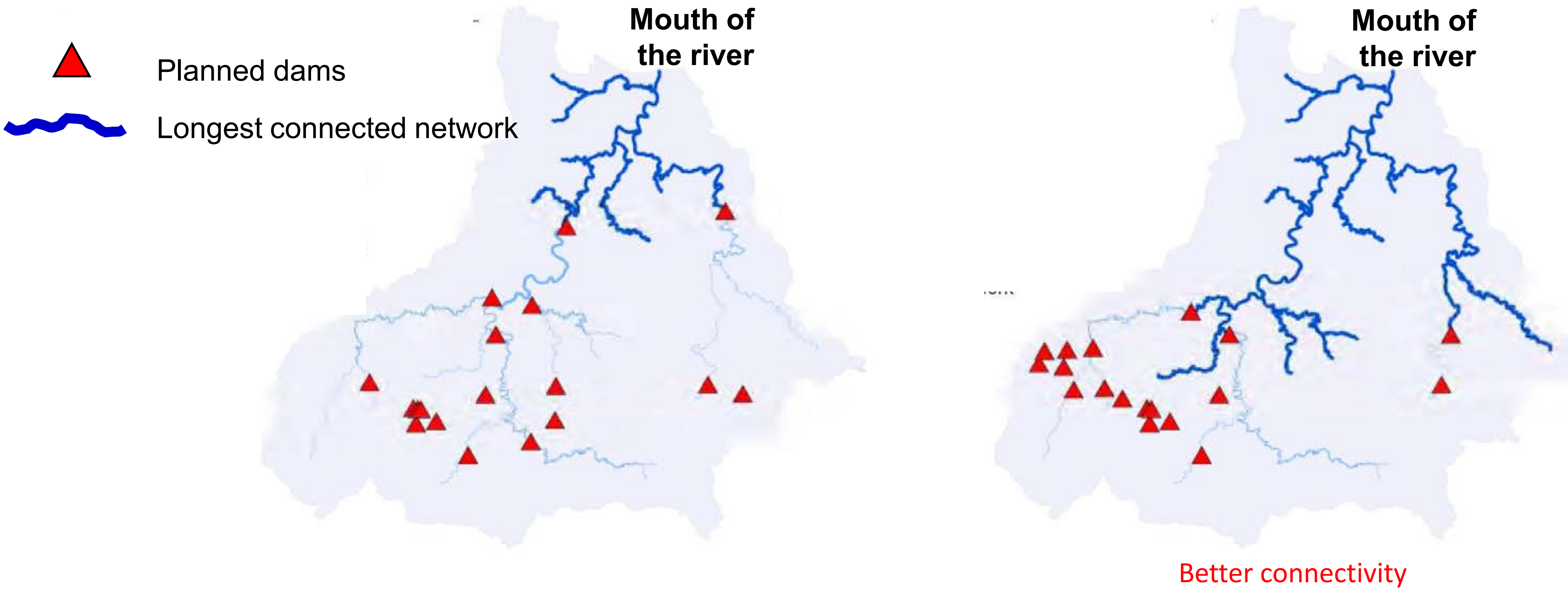
Computational Perspective:
Multi-objective Optimization Problem

Pareto frontier:
the **trade-offs** wrt to the different objectives **of different non-dominated solutions of dam portfolios**

Goal: Find Optimal Portfolios of Dams to Build



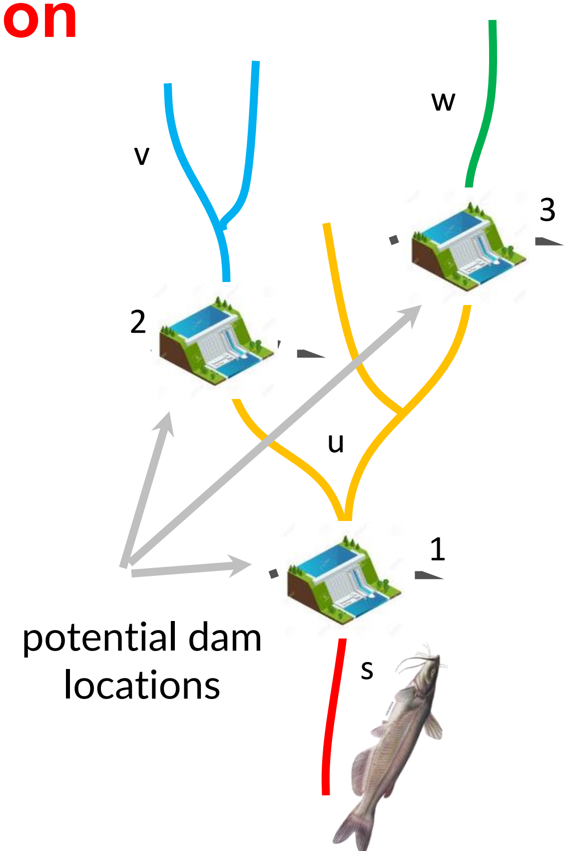
Example: Connectivity



Two dam network configurations with similar hydropower yields, but different degrees of river connectivity

Computing the Pareto Frontier
Problem Representation

River network



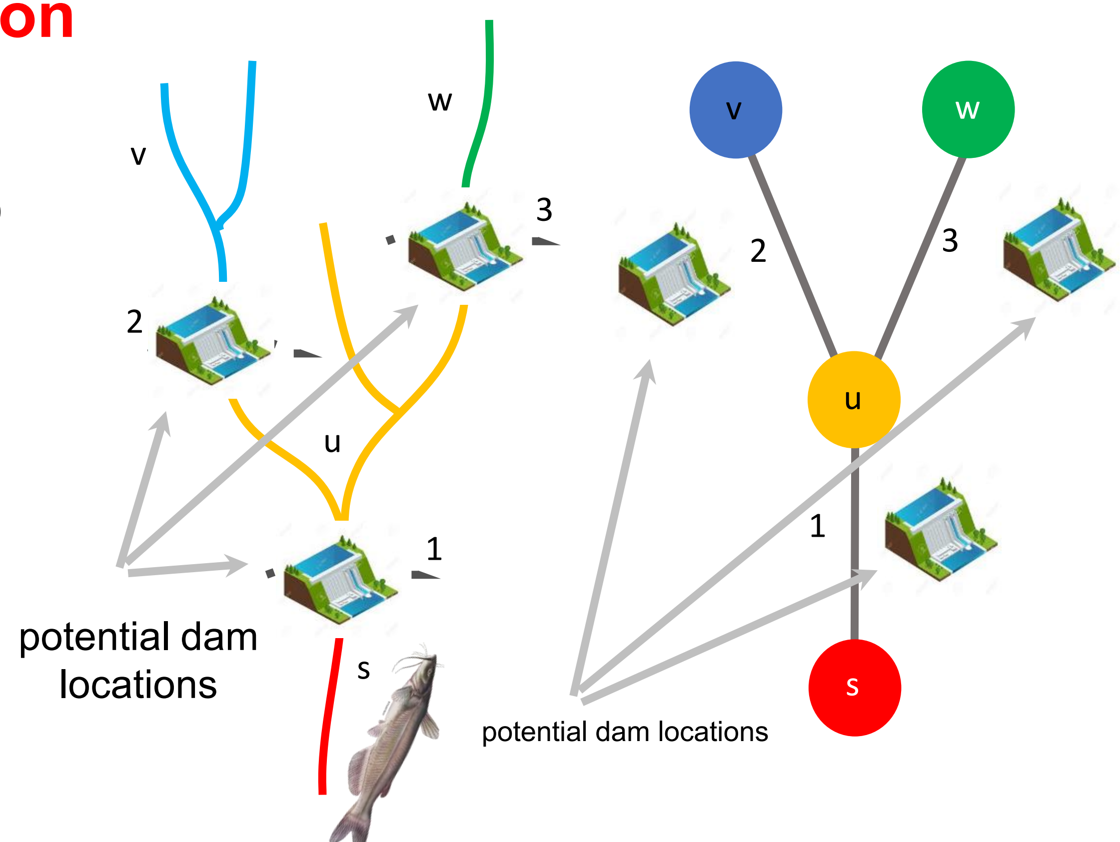
(a). River Network

(Amazon Basin has ~ **5M** river segments!)

Wu, Gomes-Selman, Shi, Xue, García-Villacorta, Anderson, Sethi, Steinschneider, Flecker, Gomes, AAAI18

Computing the Pareto Frontier
Problem Representation

River network (left) → Rooted tree (right)



(a). River Network

(Original Amazon network has
~ 5 M river segments!)

(b). Graph Representation

Compressed Amazon network:
~ 500 nodes/edges)

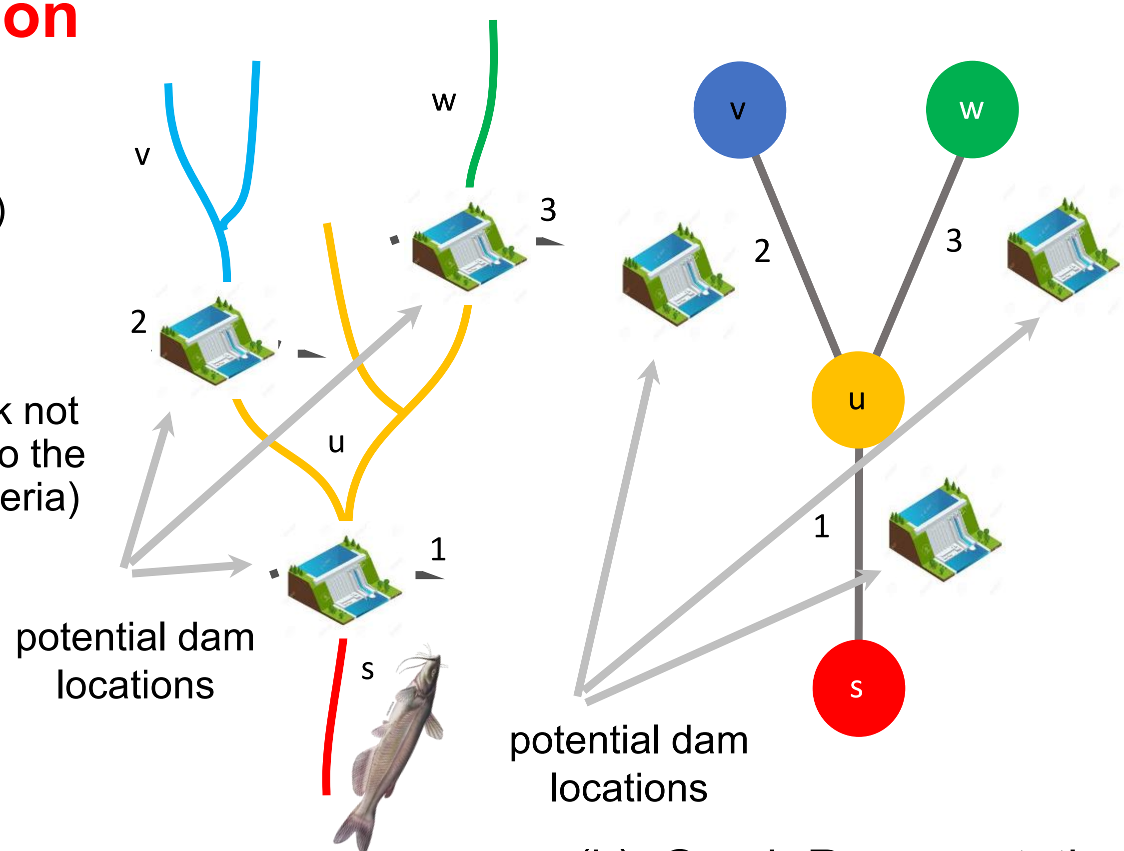
Wu, Gomes-Selman, Shi, Xue, García-Villacorta, Anderson, Sethi, Steinschneider, Flecker, Gomes, AAAI18

Computing the Pareto Frontier Problem Representation

River network (left) → Rooted tree (right)

Edge – potential dam location

Node – contiguous river sub-network not affected by a potential dam (assign to the node the utilities for the different criteria)



(a). River Network

(Original Amazon network has ~ 5 M river segments!)

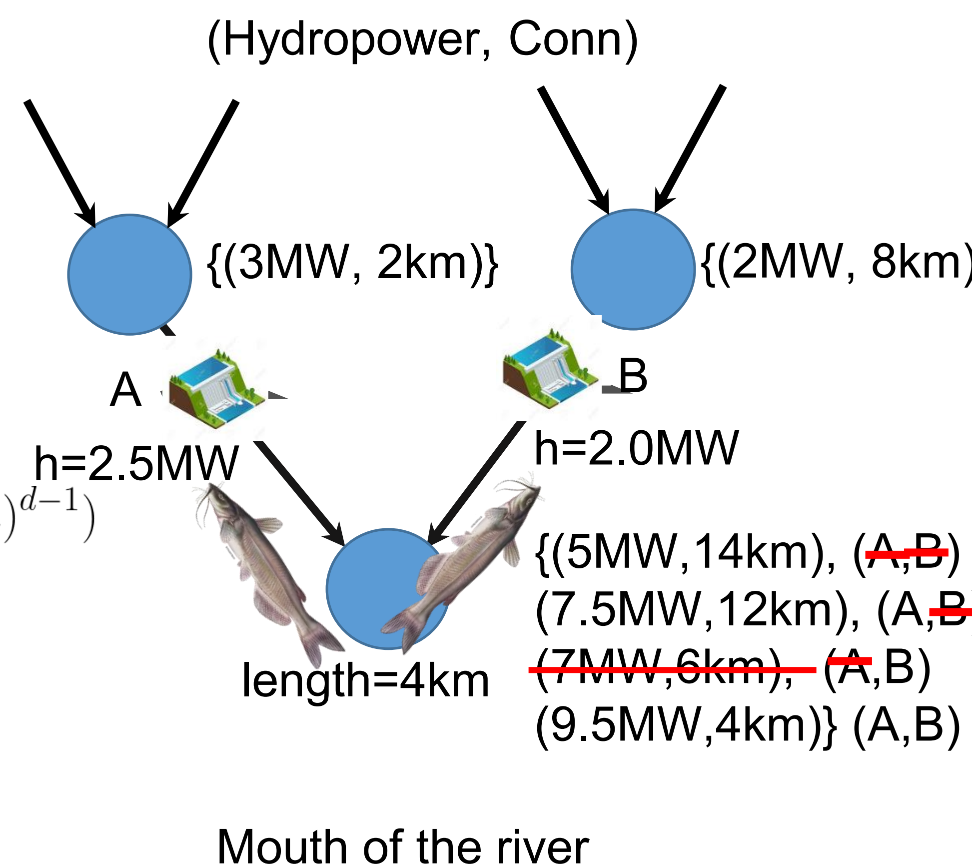
(b). Graph Representation

Compressed Amazon network: ~ 500 nodes/edges)

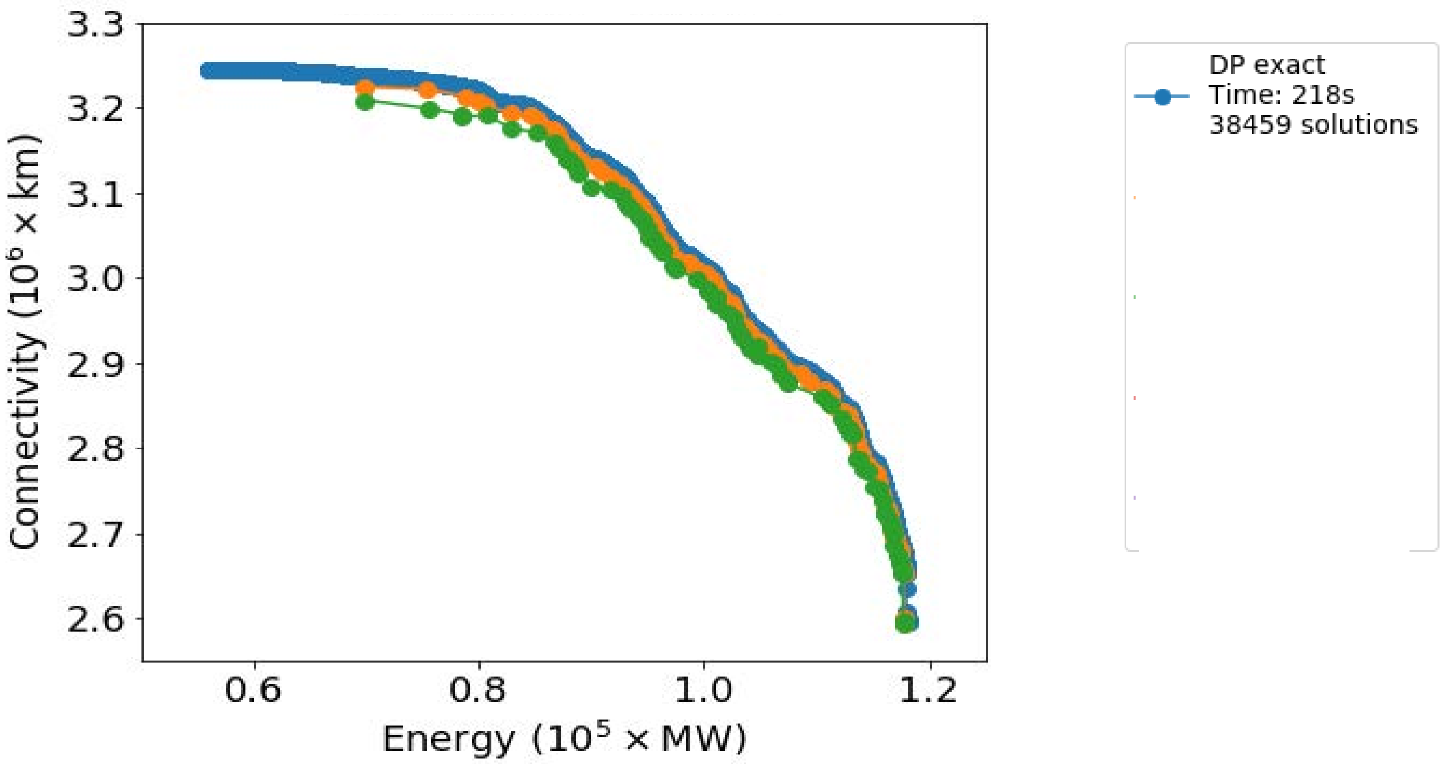
Wu, Gomes-Selman, Shi, Xue, García-Villacorta, Anderson, Sethi, Steinschneider, Flecker, Gomes, AAAI18

Computing the Pareto Frontier
Dynamic Programming Based Exact and Approximation

- Recursively compute the Pareto-frontier from leaves to root.
- Key Insight: Only need to keep Pareto-optimal partial solutions at each node.
- Fully polynomial-time approximation scheme (FPTAS) – rounding solutions guaranteeing accuracy of (1 – ε).
- Faster pruning dominated solutions. Improved from $O(n^2 d)$ to $O(n(\log n)^{d-1})$
- Other improvements to speed up algorithm (e.g., batching; imbalance binary tree)

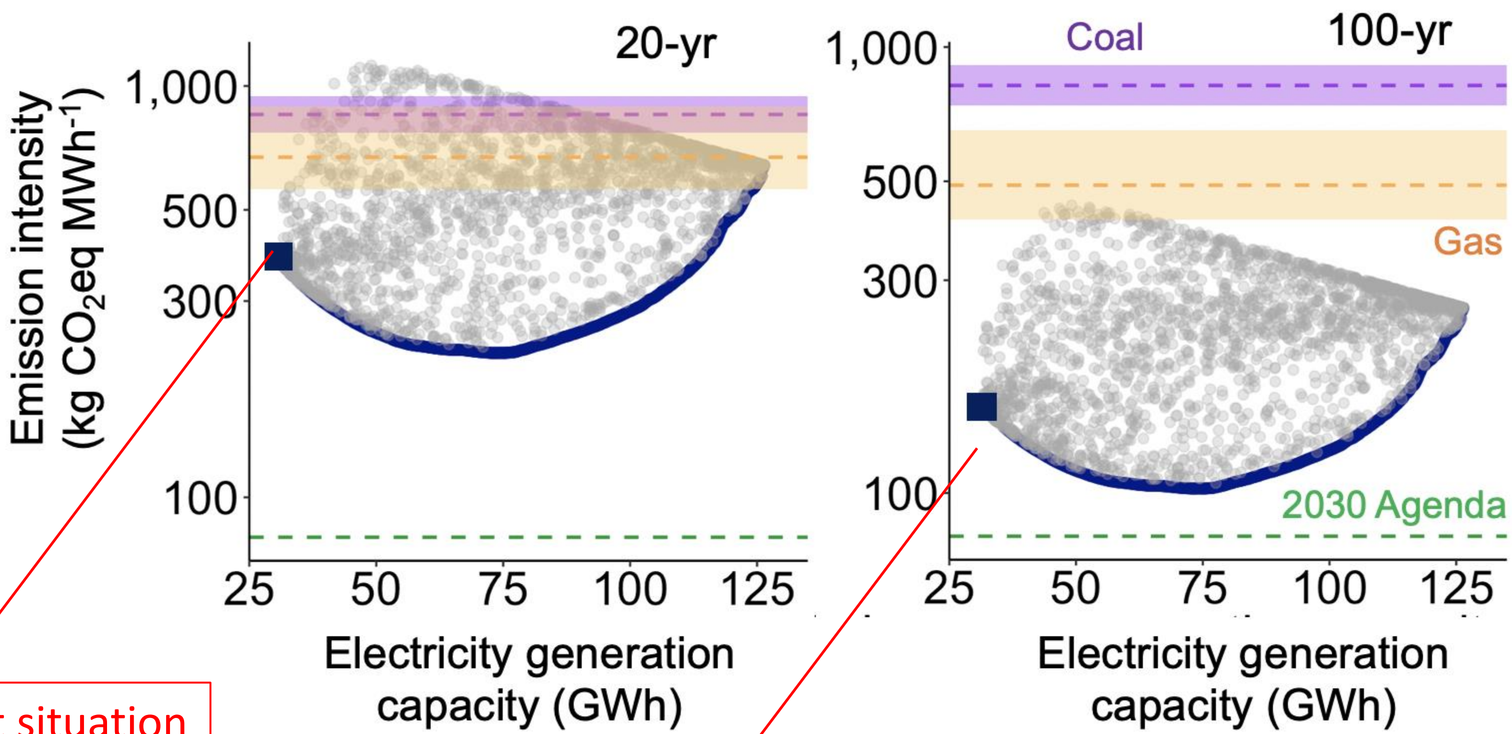


Approximation: in practice better than worst case guarantee



Entire Amazon Basin Two Criteria: Energy vs Connectivity

Greenhouse Gas Emissions



Current situation

Current situation

Rafael de Almeida et al 2019

We can now approximate the **Pareto frontier**
for **Entire Amazon basin (~5M river segments)**

Four criteria: energy, connectivity, sediment, and seismic risk

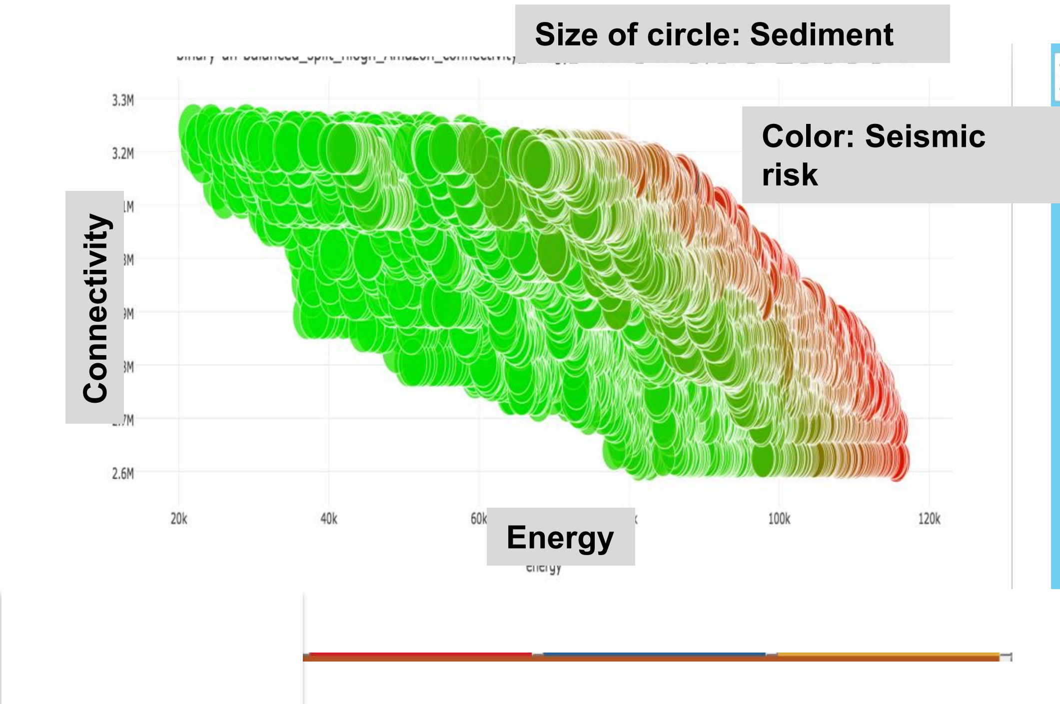
→Within 25% from true optimal Pareto frontier containing
~80K non-dominated solutions in ~ 6 minutes.

→Within 10% from true optimal Pareto frontier containing
~500K non-dominated solutions in ~ 6 hours.

→Within 5% from true optimal Pareto frontier containing
~2M non-dominated solutions in ~ 3 days.

Our approaches outperform other approaches (e.g., based on GAs).
We also provide guarantees

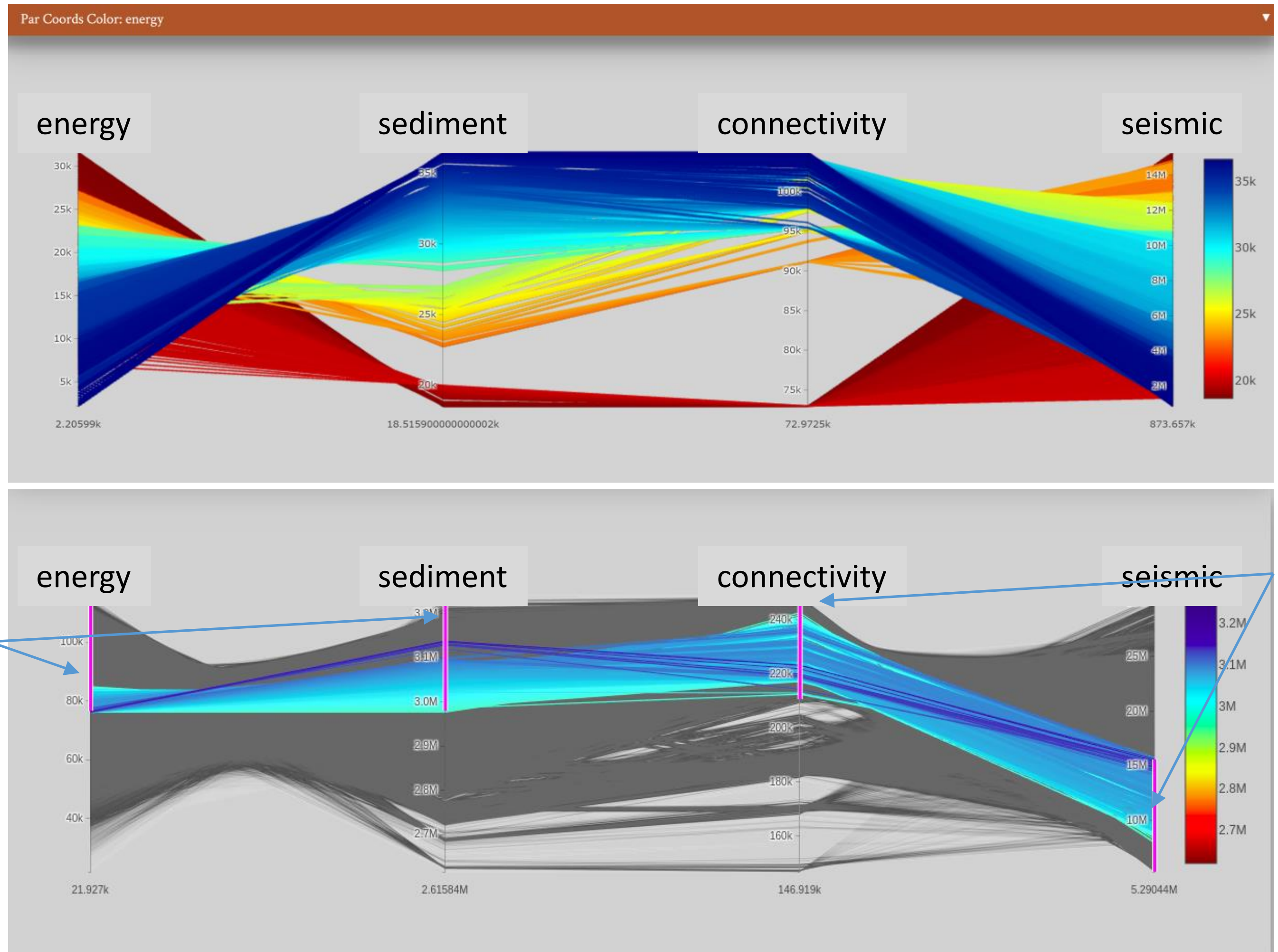
Interactive Visualizer: Parallel Coordinate Plots



Particular hydropower dam solution

Additional Criteria:
Green House Gas Emissions; biodiversity; impact on populations; etc

Interactive Visualizer: Parallel Coordinate Plots



Bounding different criteria

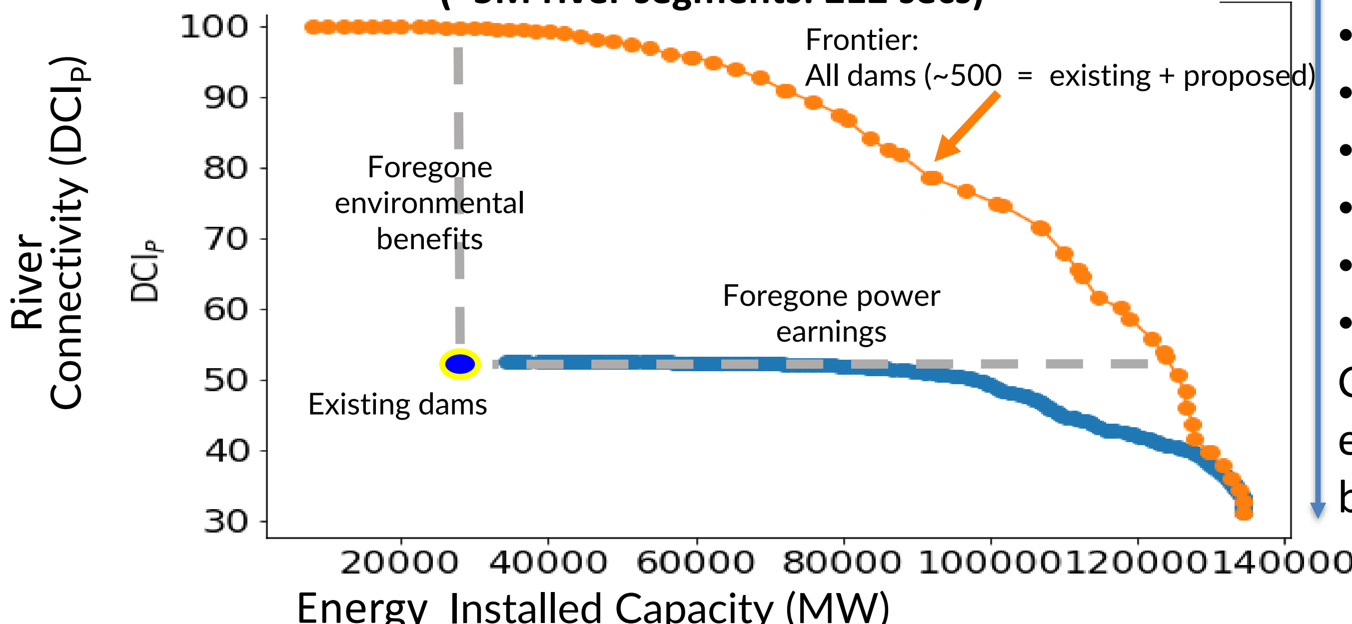
Bounding different criteria

Message to Policy Makers: The cost of inefficient planning

Challenge of Interdisciplinary Projec

Key challenge:
How to effectively establish the
large-scale interdisciplinary projects and collaborations.

Exact Pareto Frontier Amazon Basin
(Energy vs River Connectivity)
(~5M river segments: 212 secs)



Getting, vetting data and expertise for different criteria!

- Energy
- Connectivity
- Sediment
- Seismic risk
- Biodiversity
- Green House Gases
- Populations affected
- Cost
- …

Group **40+ collaborators**: ecologists, hydrologists, biologists, social scientists, …

Efficiently Approximation the Pareto Frontier: Hydropower Dam Placement in the Amazon Basin

Students

Qinru

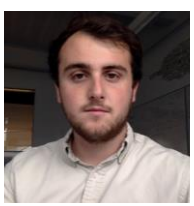
Roos

Collaborators

Xiaojian

Jonathan

Yexiang

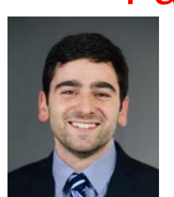


Brendan

Erin

Rafa

Faculty



The Nature Conservancy
Wildlife Conservation Society
UTEC
SEI
Michigan State University
IRD Institut de Recherche pour le Développement France
INPA Instituto Nacional de Pesquisas da Amazônia
FIU
Pontificia Universidad Javeriana
Universidad San Francisco de Quito
Escuela Politécnica Nacional
University of Nebraska Lincoln
Colorado State University

Outline

Impact of Hydropower Dam Placement in the Amazon Basin on Ecosystem Services

- ➢ Optimization with Multiple (and Conflicting) Objectives: Computing The Pareto Frontier

Species distributions (briefly)

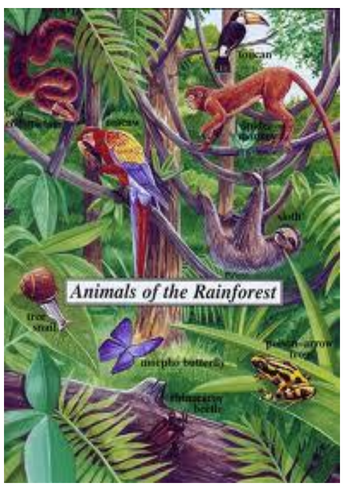
- ➢ Reducing Bias in Citizen Science Data: Avicaching Game
- ➢ Multi-Entity Dependency Learning: Deep Multivariate Probit Model

FCC Crystal Structure

Inferring Crystal Structures for Materials Discovery:

- ➢ Unsupervised learning with combinatorial constraints

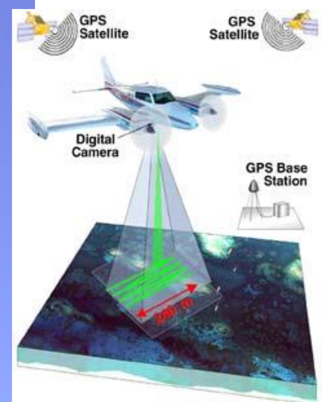
Biodiversity or Biological Diversity

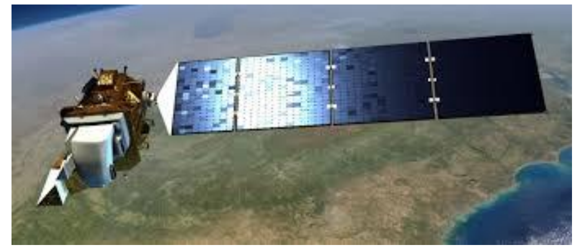


Fundamental question in biodiversity research:
How different species
are distributed
across landscapes over time.

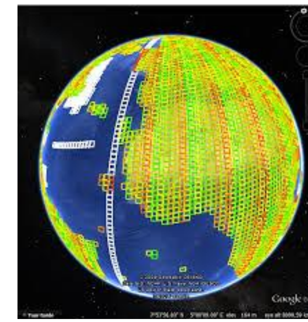
Sensors, sensor networks, and remote sensing

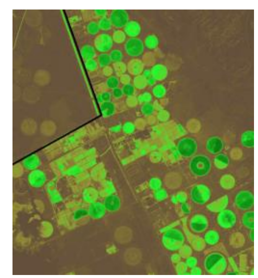
Photo courtesy of www.carboafrica.net

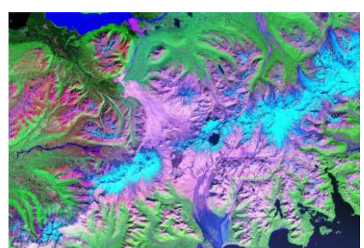




LandSat
~50years old







LandSat
images

Very sophisticated sensor

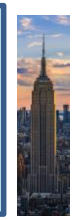
Species distributions

eBird
Citizen Science

Bird Observations

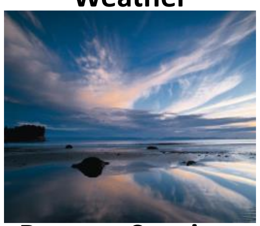
300,000+ volunteer birders	300,000,000+ bird observations	22,000,000+ **hours of field work** (2500+years)

3X

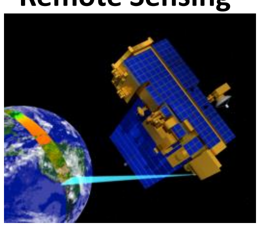


Land Cover

Weather



Remote Sensing

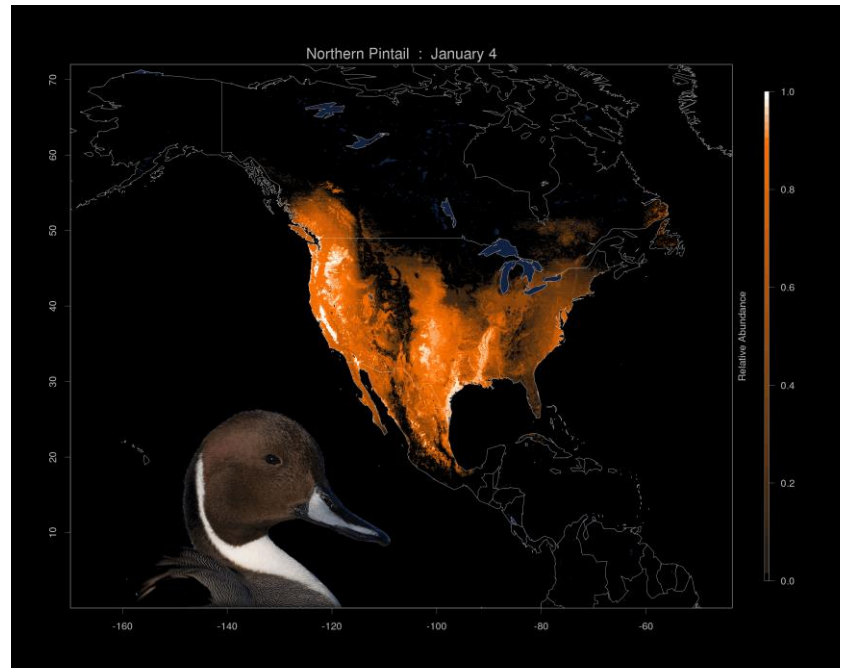


Environmental Data

$$F(X,s,t) = \frac{1}{n(s,t)} \sum_{i=1}^{m} f_i(X,s,t) I(s,t \in \theta_i)$$

Adaptive Spatial and Temporal Machine Learning Models & High-Performance Computing

Relate environmental predictors to observed patterns of occurrences and absences of the species



Patterns of occurrence of Northern Pintail for different months of the year Source: Daniel Fink

The models reveal the habitat preferences of the birds, at a fine resolution, Allowing for High-Precision Bird Conservation

Species Distributons

eBird
Citizen Science

Bird Observations

300,000+ volunteer birders	300,000,000+ bird observations	22,000,000+ hours of field work (2500+years)

3X

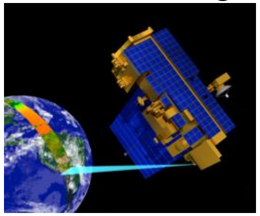
State of the Birds Report
(officially released by Secretary of Interior)

Novel Approaches To Conservation Based on eBird Models

Land Cover

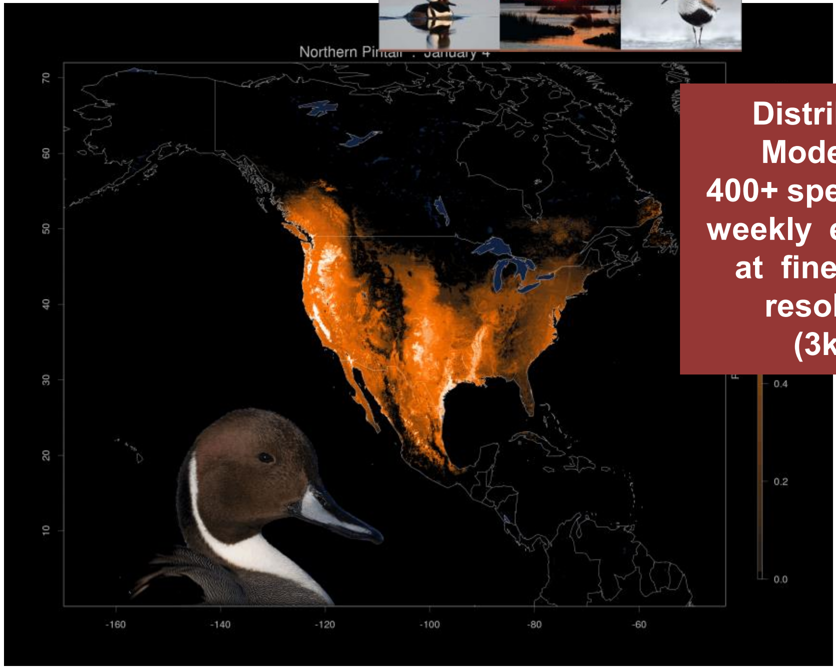
Weather

Remote Sensing



Environmental Data

Adaptive Spatial and Temporal Machine Learning Models & High-Performance Computing



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

Relate environmental predictors to observed patterns of occurrences and absences of the species

Patterns of occurrence of Northern Pintail for different months of the year Source: Daniel Fink

The models reveal the habitat preferences of the birds, at a fine resolution, Allowing for High-Precision Bird Conservation

High-Precision Bird Conservation

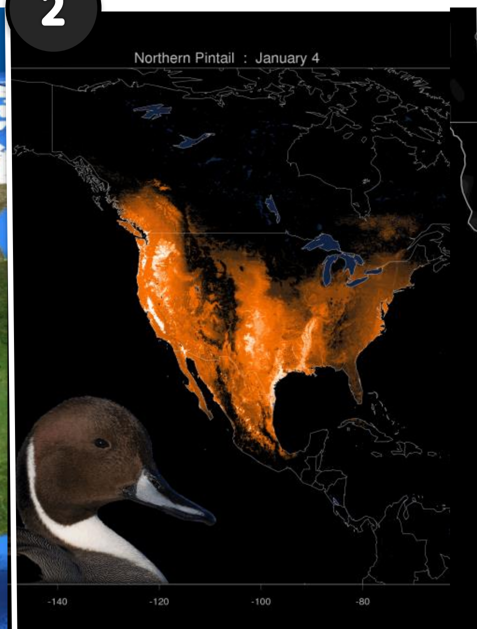
The Bird Returns Program

Protecting Migratory WaterBirds in California Against Drought

1

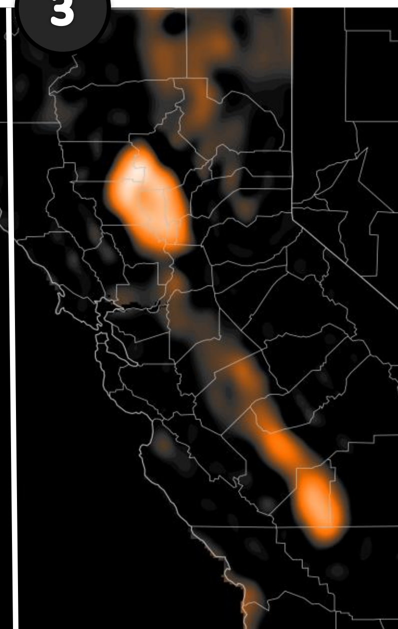
Pacific Migration Flyway

2



eBird Models

3



Target Areas

4

Farmers Submit Bids

Bids selected based on targeted estimates

Reverse Auction Bid Selection

Combinatorial Reverse Auctions

Sacramento Valley, CA

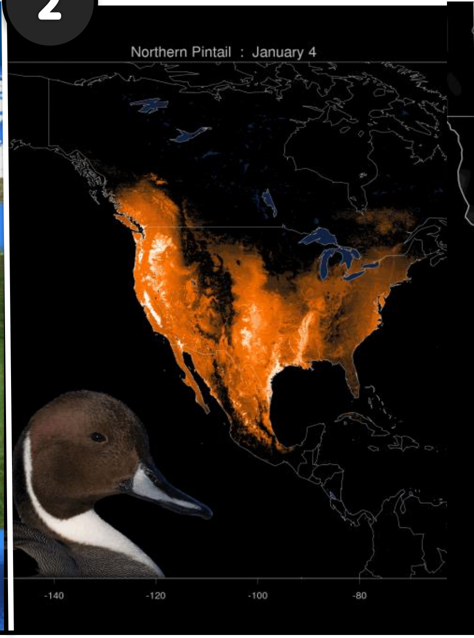
Farmers submit bids to keep the target rice fields flooded during short periods of bird migration in California.

Data-Science, Game-Theory, and Market-based approach

1

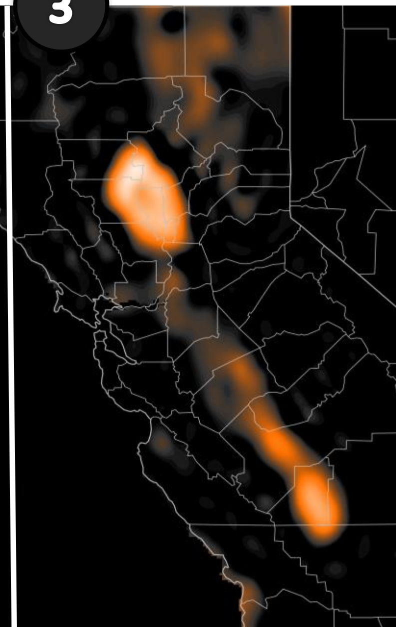
Pacific Migration Flyway

2



eBird Models

3



Target Areas

4

Farmers Submit Bids

The Nature Conservancy
Protecting nature. Preserving life.

Bids selected based on targeted estimates

Reverse Auction Bid Selection

Combinatorial Reverse Auctions

Sacramento Valley, CA

The Nature Conservancy

Over 30,000 acres
of additional habitat for waterbirds in California

Radically novel way of doing bird conservation.
Possible only because of advanced computational methods for high precision conservation.

Multi-Entity Dependence Learning

Species dependencies

- Competition, cooperation, not captured in most previous models (boosted random forests)

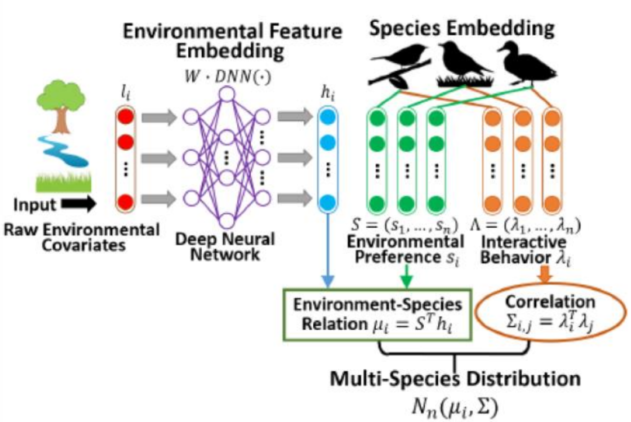
Deep Multi-Entity joint Embedding
[Chen, Xue, Fink, Gomes 16]

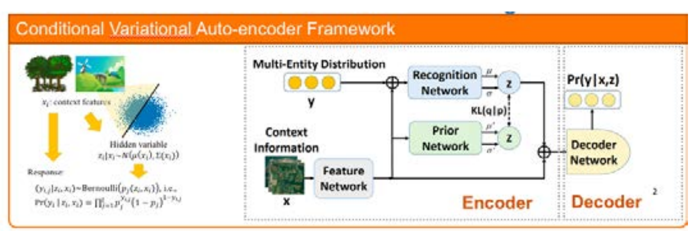
→

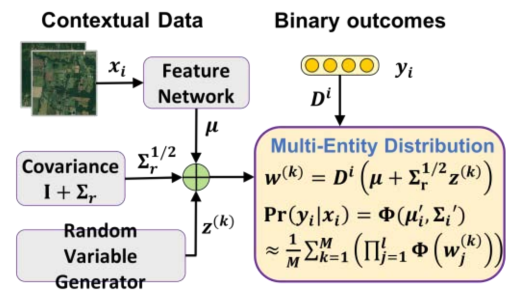
Multi-Entity Dependence Learning with Rich Context via Conditional Variational Auto-encoder
[Luming, Chen, Xue, Gomes, AAAI 18]

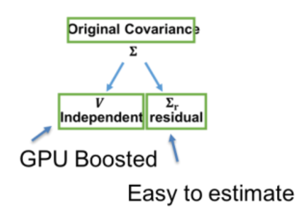
→

End-to-End GPU Boosted Learning for the Deep Multivariate Probit Model (DMVP)
[Chen, Xue, Gomes, ICML 18]





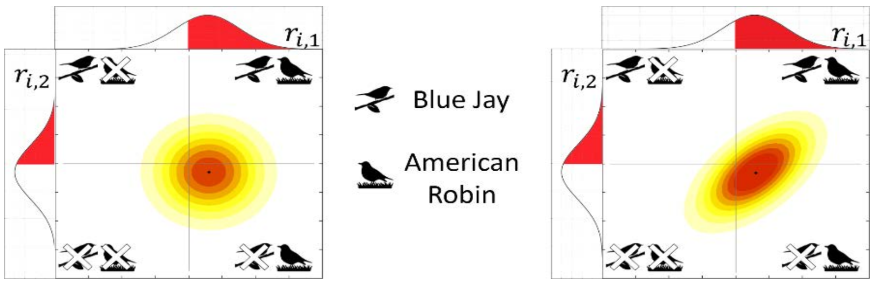




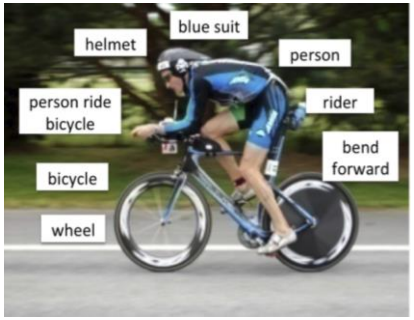
Multi-Entity Dependence Learning

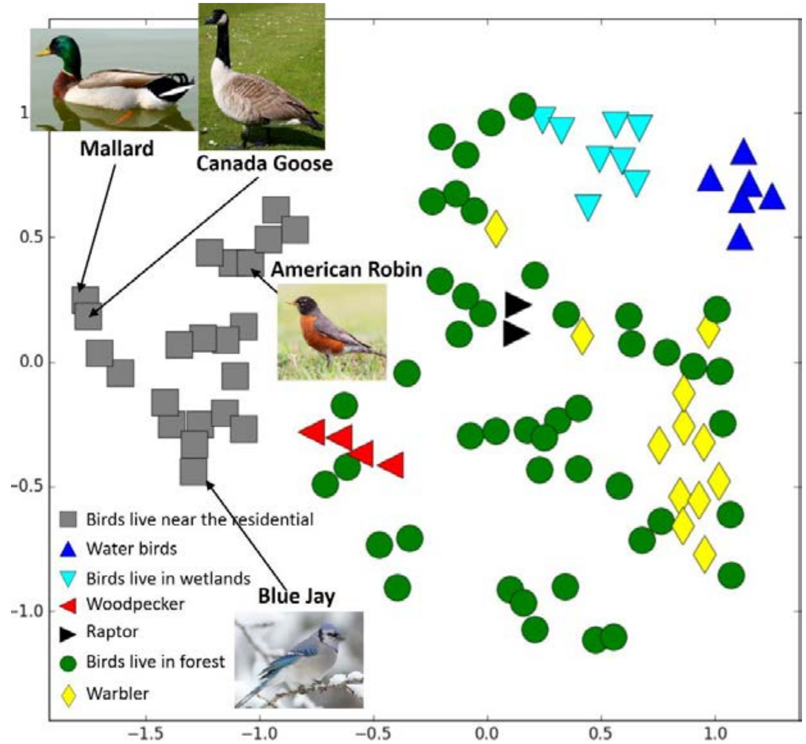
Species dependencies

- Competition, cooperation, etc.



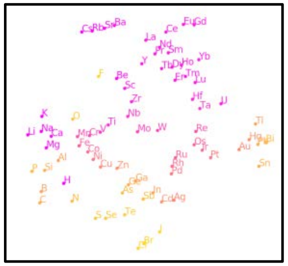
Multi-object detection
Computer vision





DMVP embeddings show elemental group and period trends correctly

Chemical Elements

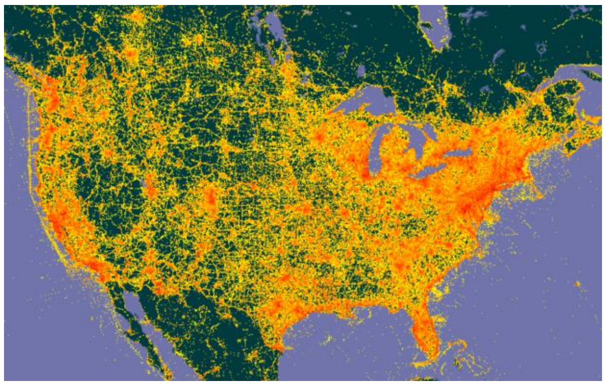


Correlations of bird species

The embedding of the multi-species interactions learned from DMVP.

Avicaching:
A Two Stage Game for Incentivizing **Bias Reduction in Citizen Science**

Data Bias Problem

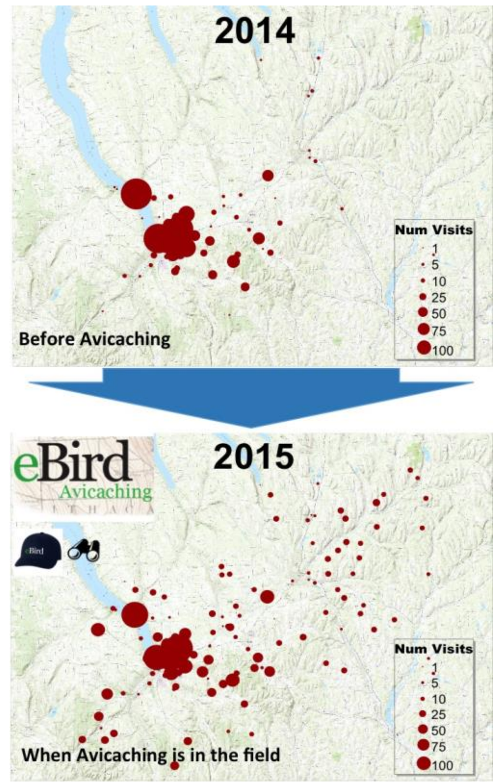


Distribution of eBird Observations in the US

Prevalent problem in citizen science
Collected data are often aligned with the participants' preferences rather than scientific objectives.

How to incentivize Citizens to visit under-sampled areas?

Principal-Agent Framework



Very Successful in Two US Counties (19% shift to undersampled areas in a 6 month period)

Field: Pilot Program

eBird Avicaching ITHACA

Incentivize eBirders to visit undersampled locations.

Incentives:

- Avicaching points,
- leaderboards
- Lotteries (e.g. binoculars.)

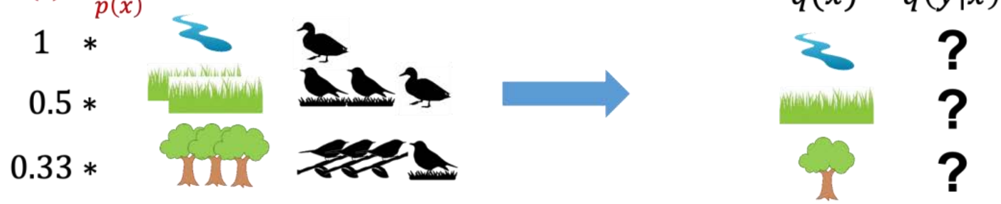
Yexiang Xue, Ian Davies, Daniel Fink, Christopher Wood, Carla P. Gomes, AAMAS 2015, CP 2016, NIPS 2016

Bias Reduction via End-to-End Shift Learning: Application to Citizen Science

Covariate Shift: re weighting the data points based on the features

Training: $p(x,y) = p(y|x)p(x)$

$w(x) = \frac{q(x)}{p(x)}$ $\quad p(x) \quad p(y|x) \quad\quad\quad q(x) \quad q(y|x)$

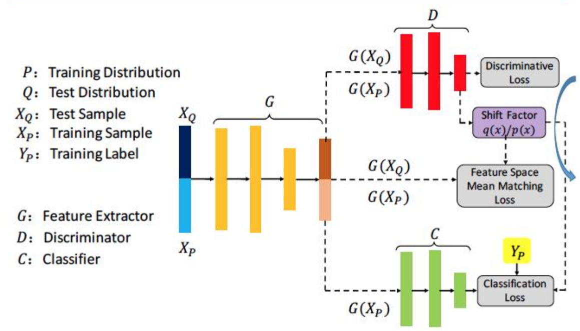


If $q(y|x) = p(y|x)$

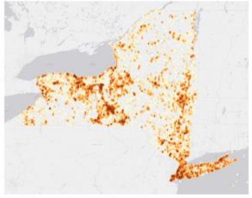
$$loss: \mathrm{E}_{(x,y)\sim P}\left[\ell(f(x),y)\frac{q(x)}{p(x)}\right] = loss: \mathrm{E}_{(x,y)\sim Q}[\ell(f(x),y)]$$

shift factor

Shift Compensation Network (SCN)



How does SCN reduce the data bias?



Before:
The original data distribution

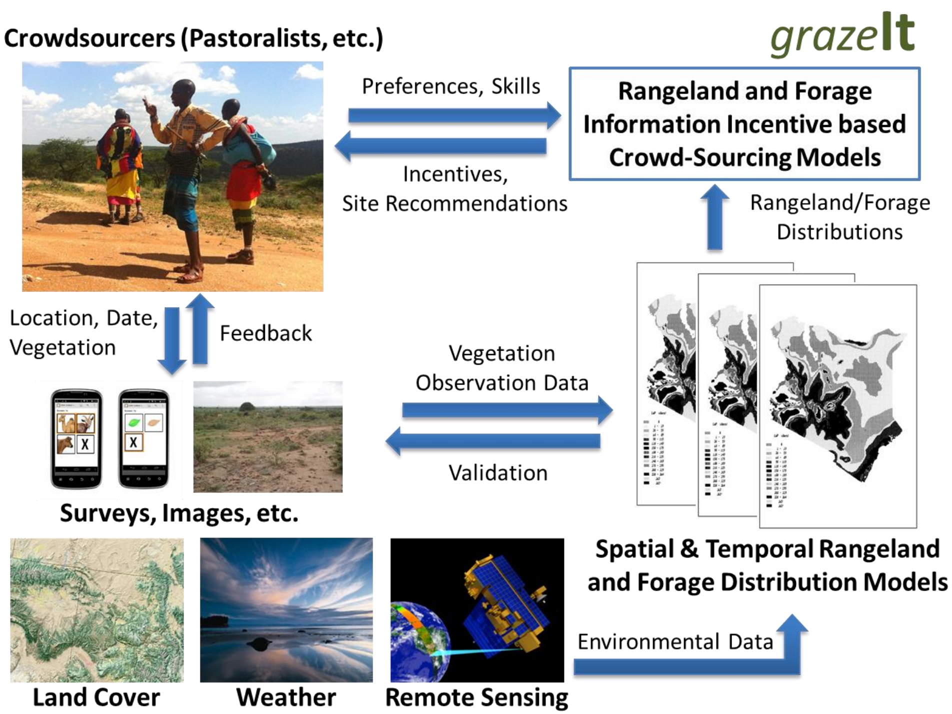
After:
The re-weighted data distribution (using SCN)

Also applicable to other bias reduction situations (loans etc)

Big Data for Africa

Improving Forage Maps in Africa

to protect farmers and herders



grazeIt

Herders Submit Vegetation Images and Surveys with Smartphones: incentives: real money (small for us, good money for pastoralists)

3 month Pilot project:
→ 100,000+ surveys

Africa is very poorly sensed

(limited environmental data, vegetation maps, only a few reliable weather stations)

Outline

Impact of Hydropower Dam Placement in the Amazon Basin on Ecosystem Services

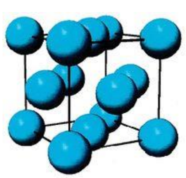
- Optimization with Multiple (and Conflicting) Objectives: Computing The Pareto Frontier

Species distributions (briefly)

- Avicaching Game: Mechanism design for Reducing Bias in Citizen Science Data

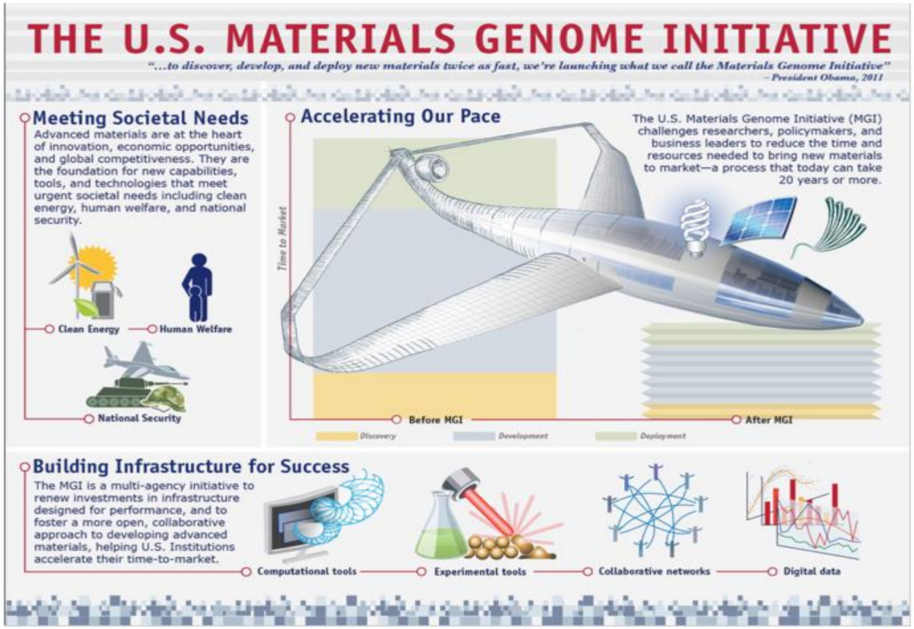
Inferring Crystal Structures for Accelerating Materials Discovery (very briefly):

- Constrained Matrix Factorization
- Human Computation for Speeding up Search

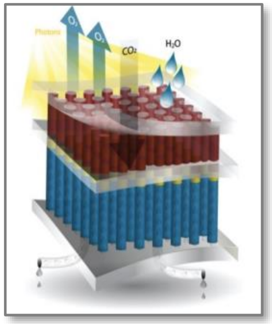


FCC Crystal Structure

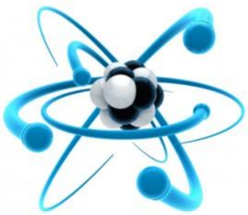
Accelerating the Discovery of New Energy Materials



Goal: Accelerate the pace and reduce the cost of discovery of new materials (Obama 2010)



Solar fuel Cell



Catalysts

.....

JCAP JOINT CENTER FOR ARTIFICIAL PHOTOSYNTHESIS Caltech

SLAC Stanford Linear Accelerator

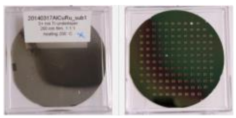
TOYOTA RESEARCH INSTITUTE

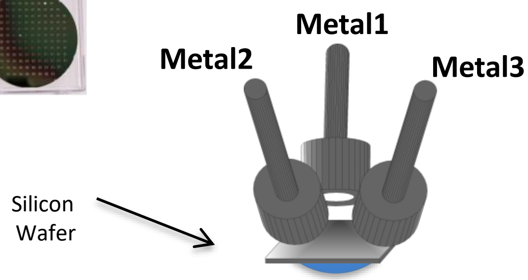
CHESS Cornell High-Energy Synchrotron

Crystal Structure Phase Mapping from Experimental Data: A Computational Perspective

Crystal Structure Mapping Problem from Hugh-Throughput Experiments

Co-sputtering
(similar to atomic spray painting)



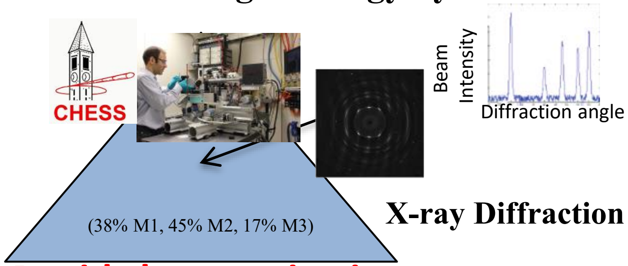


Simultaneous synthesis of thousands of materials

10^2 – 10^3 materials/day

High-Throughput Materials Discovery

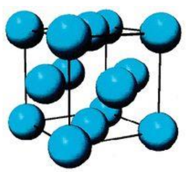
Cornell High-Energy Synchrotron



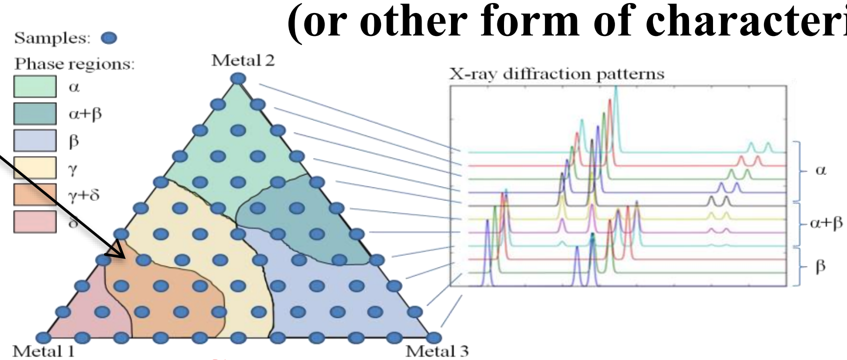
Rapid characterization of thousands of materials

10^3 – 10^5 materials/day

How to infer the crystal structure of the materials, based on the X-ray diffraction patterns (or other form of characterization, e.g., Raman)?



FCC Crystal Structure



Crystal Str)ucture Map Problem:

Infer the crystal structures of the materials from the X-ray diffraction patterns

Source Separation Problem

Mainly manual task requiring expert knowledge!

Only a few systems a year

Goal:
Achieve High-Throughput Crystal Structure Identification

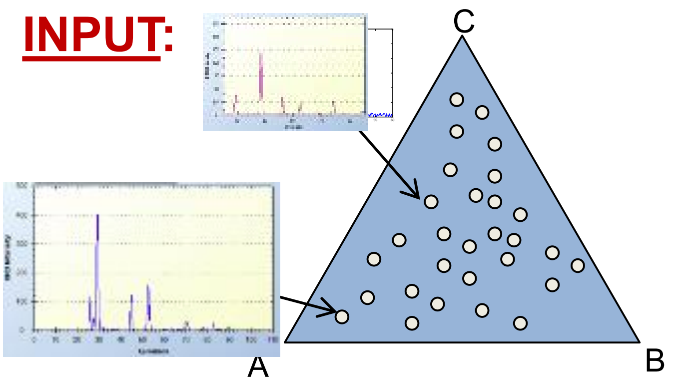
Difficulty: Often X-ray diffraction patterns correspond to a mixture of crystal structures

Challenging to un-mix the X-ray diffraction patterns

JCAP JOINT CENTER FOR ARTIFICIAL PHOTOSYNTHESIS
Caltech

Phase Map Identification Problem

INPUT:



Collection of XRD Patterns

Possible Additional Physical Requirements:

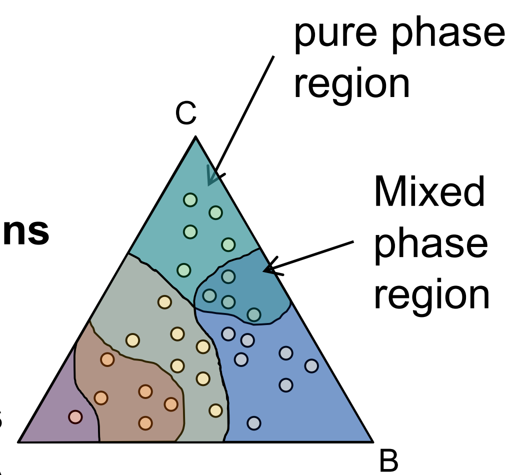
- Phase Connectivity
- Gibb’s Rule:
  Mixtures of at most 3 pure phases
- Peaks shift by ~15% within a region
  - Continuous and Monotonic
- Small peaks might be discriminative
- Peak locations matter,
  more than peak intensities

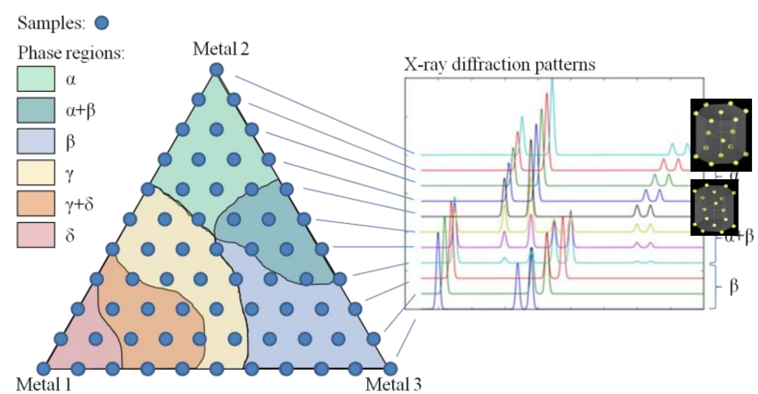
OUTPUT:

m **phase regions**

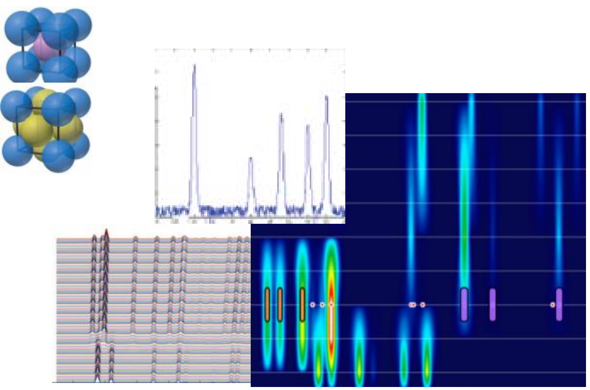
- ***k*** **pure regions**
- ***m-k*** **mixed regions**

XRD pattern characterizing pure crystal phases

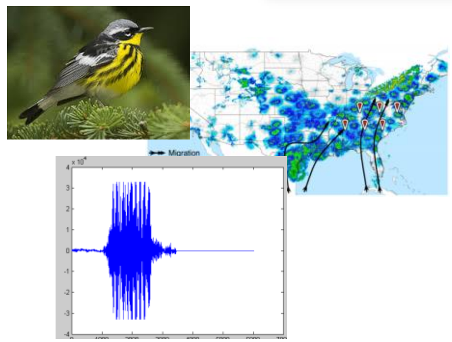




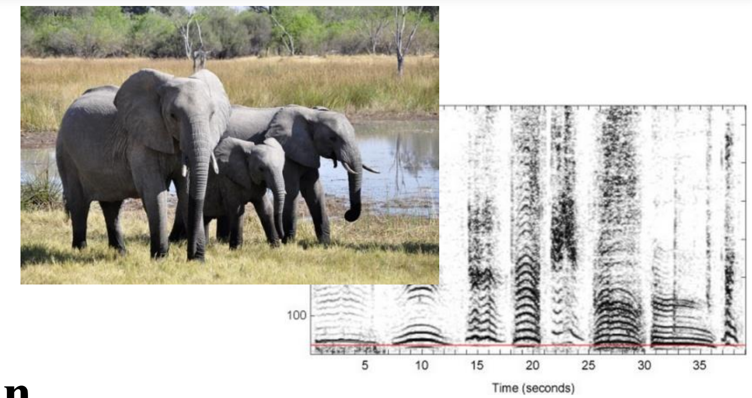
Related Problems: Pattern (Factor) Decomposition or Source Separation



Materials Discovery:
Phase Map Identification

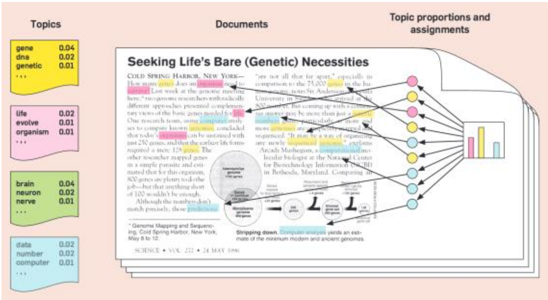


Flight Call Detection
for Bird Conservation



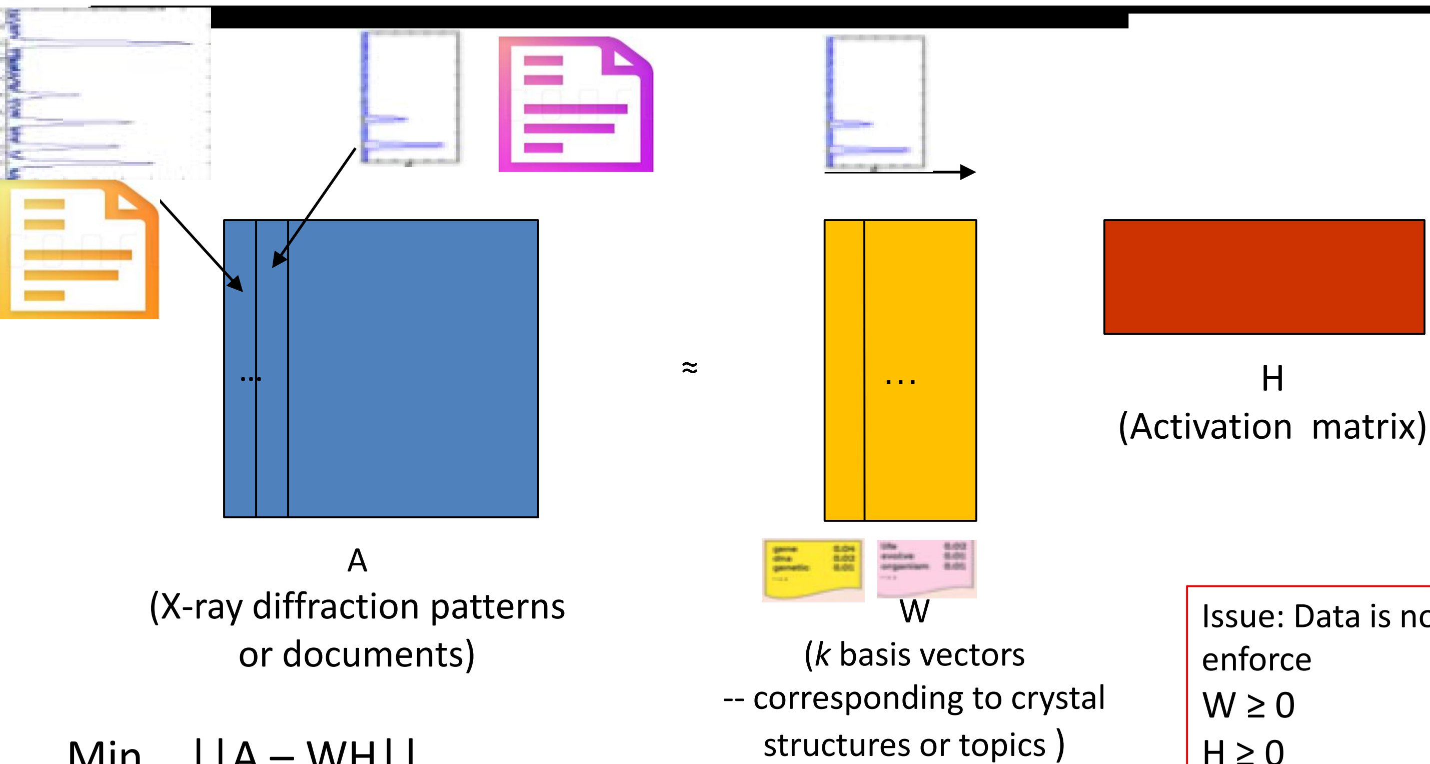
Elephant Listening Project;
Elephant Call Detection

Music source separation:
Extracting and identifying each single instrument sound in a



Topic Modeling:
Identifying the Key Topics of a collection of articles (or an article)
Blei, ACM 2012

Matrix factorization With Combinatorial Constraints



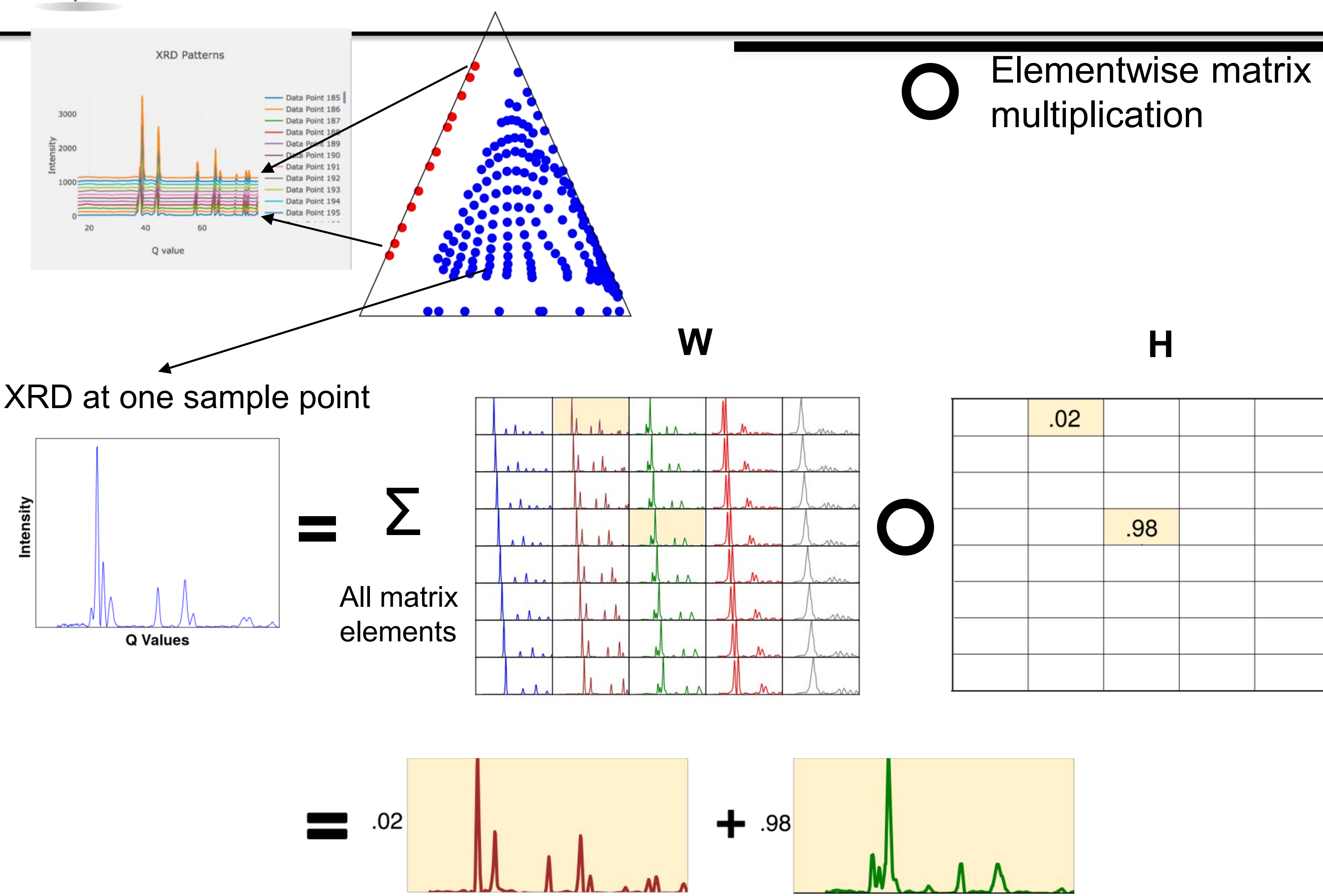
Issue: Data is non-negative. Need to enforce
$W \geq 0$
$H \geq 0$
And additional combinatorial constraints

$$\min_{W,H} \; \| A - WH \|_p$$

Subject to:

Combinatorial constraints to encode laws of physics – e.g shifts, Gibbs Rule, etc

Phase Mapping as A Matrix Factorization Problem



Crystal Structure Map: Computational Challenges

→Unsupervised learning – No labeled data

(ML success depends on large amounts of labeled data)

→ Need to deal with noise, incomplete data, uncertainty

→ Standard ML techniques:

fail to capture the underlying physics of the phenomena

Required: Rich set of combinatorial constraints
to capture the physics of the phenomena

↓

Integration multiple knowledge sources and reasoning mechanisms

1. XRD data (also Raman, optical, others)
2. Materials databases prior knowledge (Materials Project, OQMD, etc.)
3. Quantum physics (DFT calculations)
4. Human expertise

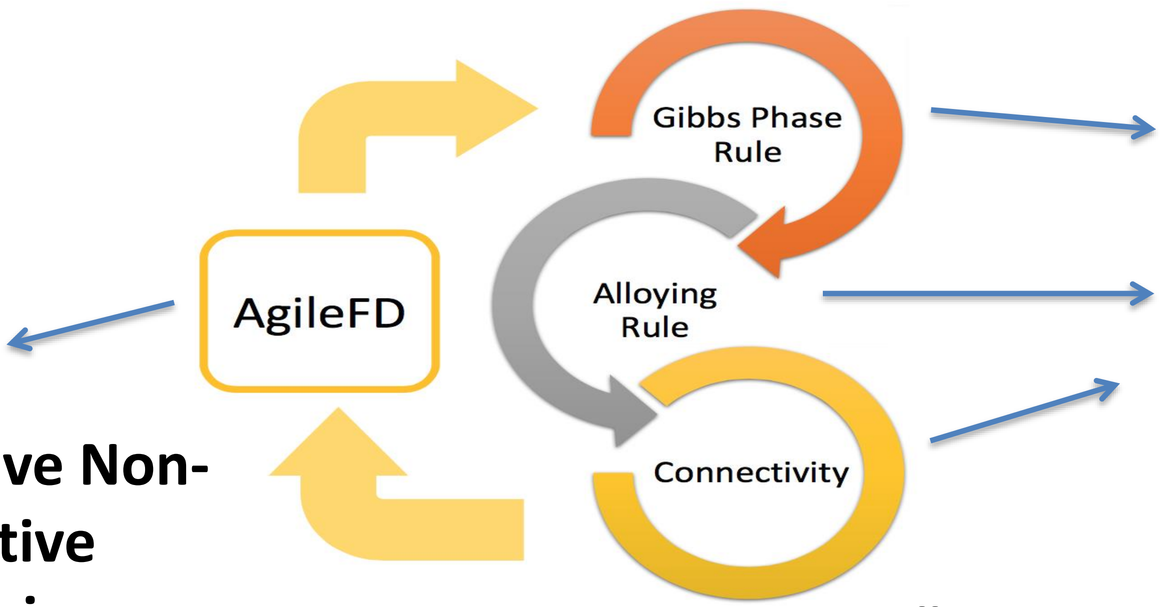
Computational Synthesis:
Integration of machine learning techniques with constraint and probabilistic reasoning, sampling, and optimization techniques

Phase Mapping as Constrained Matrix Factorization

INTERLEAVED AGILE FACTOR DECPOMPOSITION (IAFD)

Relaxation and Projection Methods for Constrained Matrix Factorization Problems → <u>producing physically meaningful solution</u>

IAFD

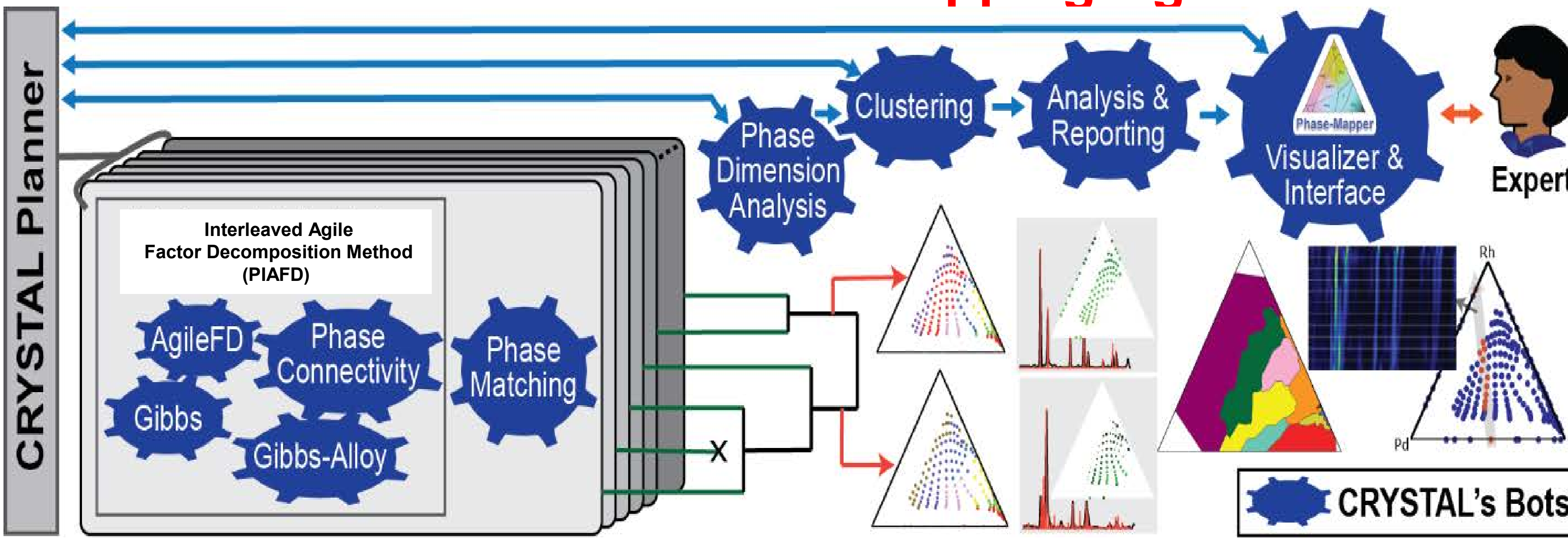


Specialized constraint reasoning algorithms to enforce physical constraints

Convolutive Non-Negative Matrix Factorization

We can now automatically generate a physically meaningful phase-diagram in ~5 min!!!!

Crystal:
Phase Mapping Agent Ensemble



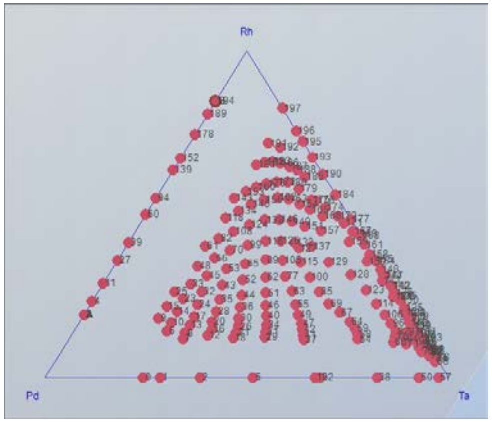
Crystal's Brain:
Interleaved Agile Factor Decomposition

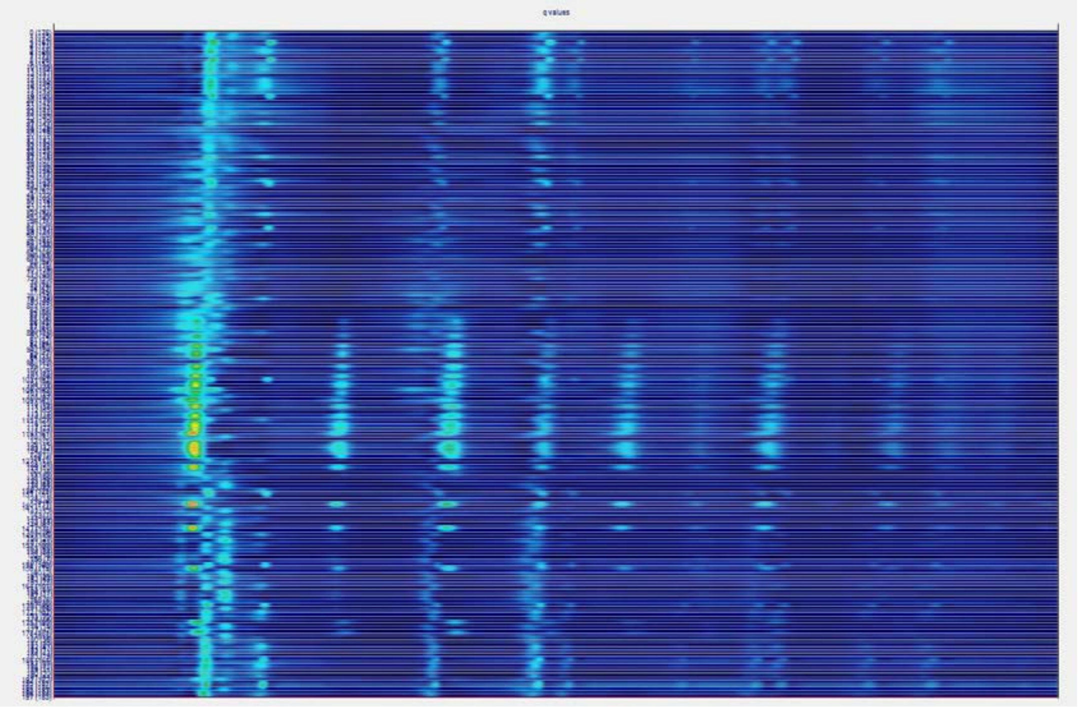
Crystal is a multi-agent system that encapsulates a diverse collection of fast and specialized algorithms with different types of knowledge and computational capabilities for Crystal Structure Phase Mapping

Gomes, Bai, Xue, Bjorck, Rappazzo, Ament, Bernstein, Kong, Suram, van Dover, Gregoire, 2019

Pd-Ta-Rh system

198 synchrotron
x-ray diffraction patterns

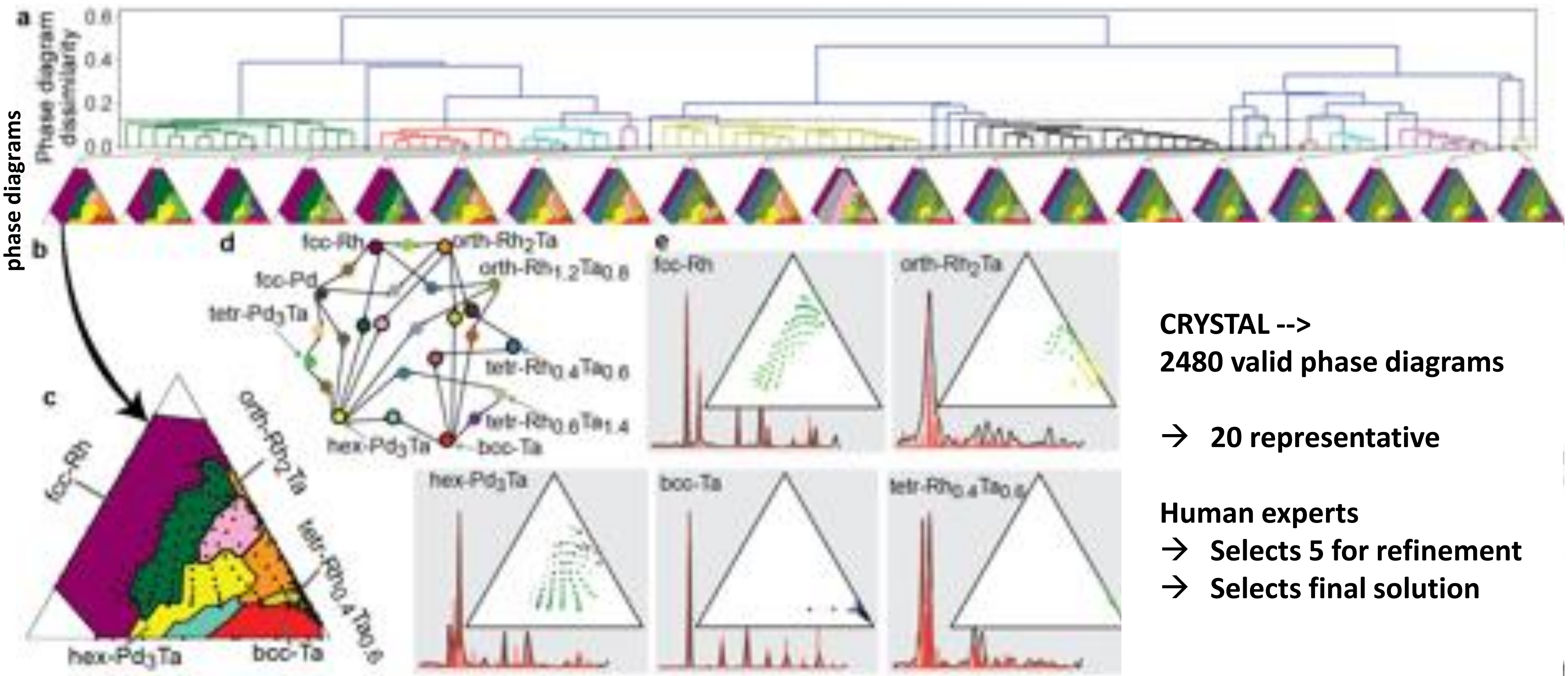




~2500 valid phase diagrmas

CRYSTAL: Pd-Rh-Ta System

20 representative phase diagrams



CRYSTAL -->
2480 valid phase diagrams

→ **20 representative**

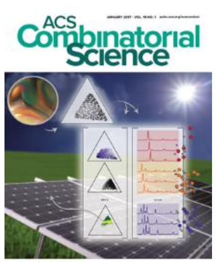
Human experts
→ **Selects 5 for refinement**
→ **Selects final solution**

a) Clustering of the **2480** phase diagrams produced by **Crystal** for the Pd-Rh-Ta system **b) 20** clusters and respective representative phase-diagrams were identified. Out of the 20 phase diagrams, the human expert ruled out 15, based on subtleties not enforced by Crystal. **c)** From the remaining 5 phase diagrams, the human expert selected the enlarged phase diagram, based on additional metrics, characterized by the phases represented in e). The five-phase solution selected by the human expert for the Pd-Rh-Ta system. **d)** Color scheme for the phase fields. **e)** The basis patterns are plotted along with the ICSD basis patterns. Each phase map is shown as a composition plot where the size represents the phase fraction and the color denotes the relative lattice constant, compared to the respective basis pattern, aligned to the best match of the ICSD pattern.

CRYSTAL enabled the discovery of
a mixed-intermetallic methanol oxidation electrocatalyst: $Pd_{0.17}Rh_{0.33}Ta_{0.5}$

CRYSTAL's Interactive Phase Mapper (to be made publicly available)

Interactive Phase Mapper backend powered by IAFD which provides the functionality of producing physically meaningful solutions.



Innovative AI Award

SARA

Scientific Autonomous Reasoning Agent (SARA): Integrating Materials Experiment, Theory, and Computation

An AFOSR MURI launched Feb. 2018

Materials Scientists

PI Bruce van Dover

Dept. Materials Science and Engineering
Cornell University

Co-PI Mike Thompson

Co-PI Chris Wolverton

Materials Science and Engineering
Univ. of Northwestern

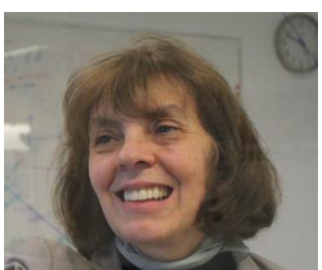
Co-PI John Gregoire

Joint Center for Artificial Photosynthesis
Caltech

Co-PI Alex Zunger

Materials Science and Engineering
Univ. of Colorado Boulder

Computer Scientists



Co-PI Carla Gomes

Dept. of Computer Science
Cornell University

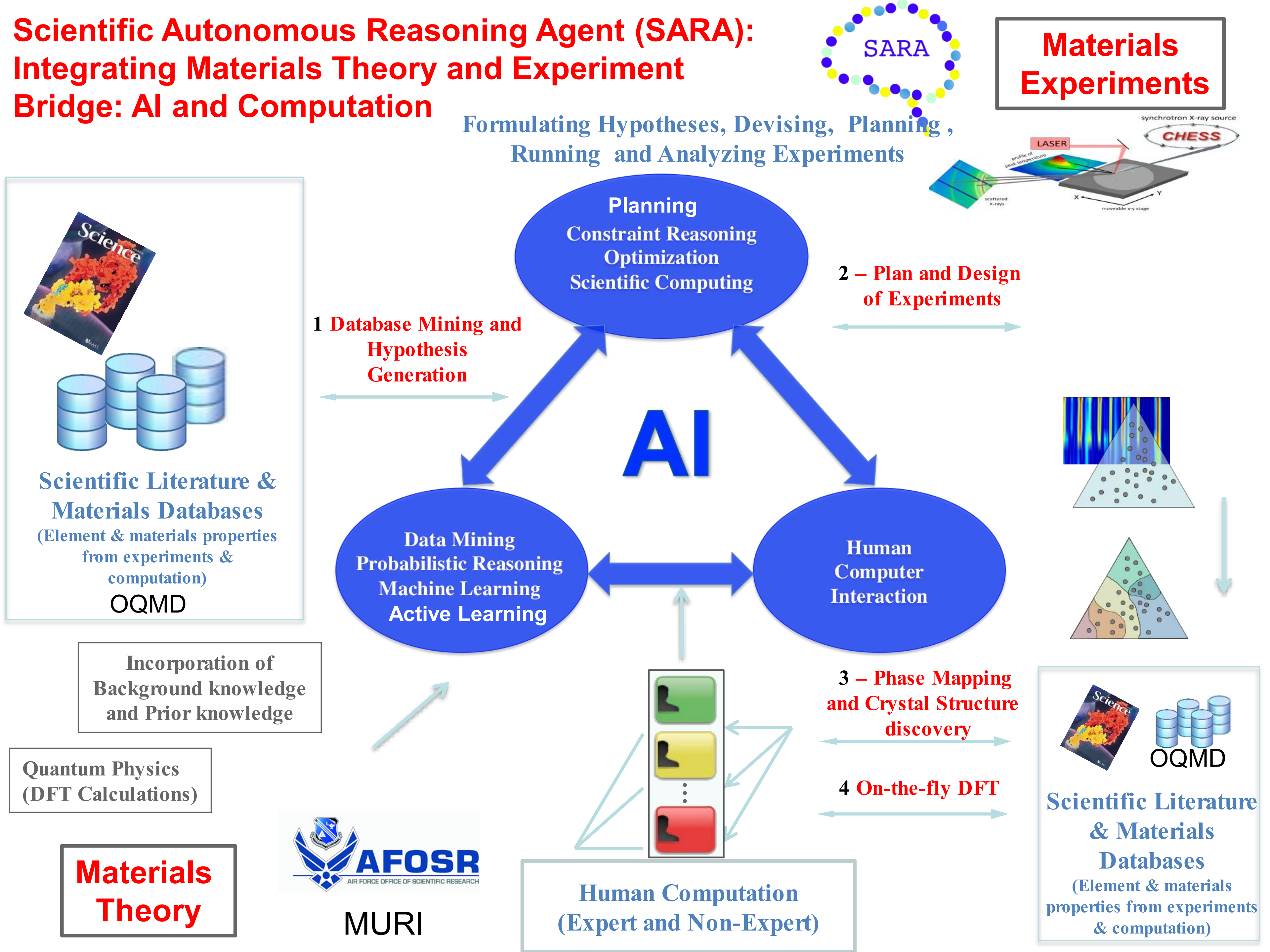
Co-PI Bart Selman

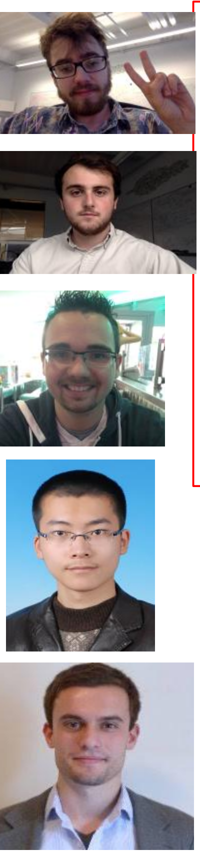
Cornell University

Northwestern University

University of Colorado Boulder

Caltech


Scientific Autonomous Reasoning Agent (SARA):
Integrating Materials Theory and Experiment
Bridge: AI and Computation
SARA
Materials Experiments
Formulating Hypotheses, Devising, Planning, Running and Analyzing Experiments
synchrotron X-ray source
CHESS
LASER
Planning
Constraint Reasoning
Optimization
Scientific Computing
2 – Plan and Design of Experiments
1 Database Mining and Hypothesis Generation
AI
Scientific Literature & Materials Databases
(Element & materials properties from experiments & computation)
OQMD
Data Mining
Probabilistic Reasoning
Machine Learning
Active Learning
Human Computer Interaction
Incorporation of Background knowledge and Prior knowledge
Quantum Physics (DFT Calculations)
3 – Phase Mapping and Crystal Structure discovery
4 On-the-fly DFT
Science
OQMD
Scientific Literature & Materials Databases
(Element & materials properties from experiments & computation)
AFOSR
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
MURI
Human Computation (Expert and Non-Expert)
Materials Theory

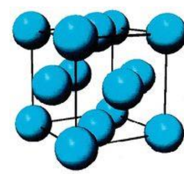


My Philosophy:
Students should work on similar computational problems in different domains.
Grad students working on materials discovery also work on the elephant and flight call, eelgrass problems and music.

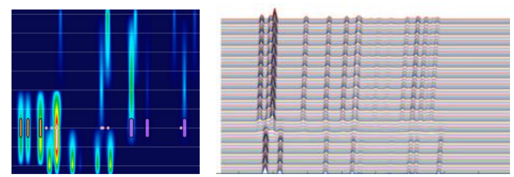
Pattern Decomposition in Big Data

Dimensionality Reduction,
Source Separation,
and Segmentation
with Complex Constraints

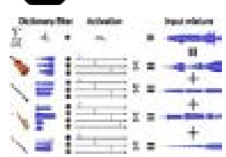
Crystal Phase Mapping from X-Ray Diffraction Data



FCC Crystal Structure



Eelisa

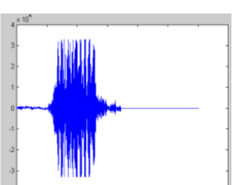


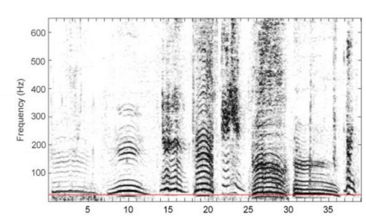
Separating instruments in music

Identifying wasting disease lesions in eelgrass

Flight Call Detection from audio recordings

Elephant Call Detection (from audio recordings)






Computational Sustainability
@ Cornell
Gomes Lab
GrazeIt
Modeling of Pastoralists' Movements and Vegetation Mapping (Kenya)
eBird
Avicaching
Dynamic Precision Bird
Pattern Decomposition in Big Data
Citizen Science/ Crowdsourcing
Agents: Mechanism Design
Large Scale Spatio-Temporal Modeling and Prediction
Stochastic, Probabilistic Inference, and Optimization
Large Scale Sequential Decision Making
Ecuador
Citizen Science
Avicaching ,
Estimating Bi
Populations a
Migrations
Socio-Economic-Environment
Impacts of Dams in The Amazon Basin
Socio-Ecological Wildlife Corridor (Ecuador)
Monitoring Eelgrass and Sea tar Wasting
Designing Experiments for Fertilizers
FERTILIZER
Music source separation
Elephant Call Detection
Inferring Crystal Structures from X-Ray Diffracti
Identifying wasting disease lesions in eelgrass
Materials Discovery
Flight Call Detection
Invasive Species
Expeditions in Computing (CISE)
NSF
www.Udiscover.It
Atkinson Center for a Sustainable Future

Conclusions

Computational Sustainability

Computational Sustainability aims to advance computational methods to help balance economic, environmental, and societal needs for sustainable development.

1. New challenging problems

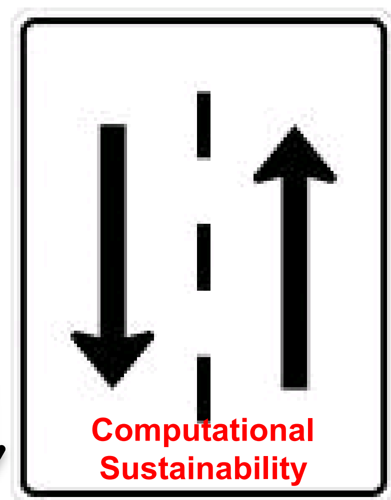
2. New formalisms and concepts from other disciplines

→ New Core Paradigmatic problems in Comp. Sci.

Foundational Contributions to CS

Broaden Scope & Diversity of CS

Sustainability Questions



Computational Sciences

Computational Thinking providing new
insights,
methodologies,
and solutions
to sustainability problems

→ Societal Impact

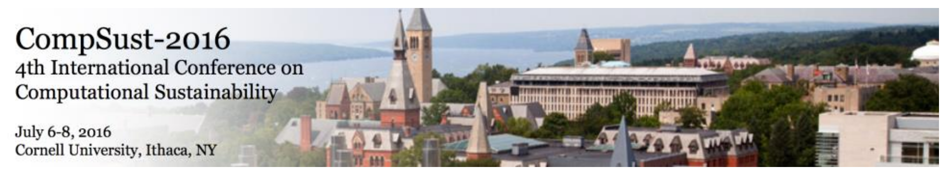
Thank you! Computational Sustainability Vibrant Research Community

CompSust Virtual Seminar

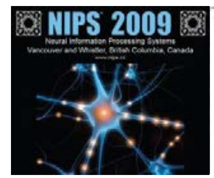
www.compsust.net

CompSust Conference Series:

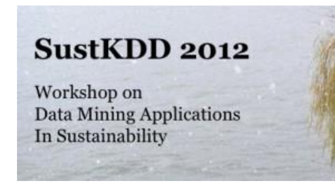
(international researchers from several disciplines and institutions (universities, labs, government)



Workshops at Conferences



Neural Information Processing Systems Foundation



STOC 2012 – 44th ACM SYMPOSIUM ON THEORY OF COMPUTING

CompSust Virtual Seminar

Tracks at Established Conferences

IJCAI

Theme of IJCAI-2013 (CHINA)
AI and Computational Sustainability

IJCAI-13

PERVASIVE 2012

COMPUTING COMMUNITY CONSORTIUM

ICML 2012 EDINBURGH

CROCS: Constraint Reasoning and Optimization for Computational Sustainability

CROCS at **CP**-09, **CPAIOR**-10 **CP**-10 and **CP**-12

CoompSust@AAAI-2018 CoompSust@AAAI-2019

And many other related conferences and journals

Expeditions in Computing (CISE)

NSF

Thank you!

Cornell University · Caltech · Carnegie Mellon University · Bowdoin · Georgia Tech · Howard University · The Ohio State University · OSU Oregon State University · Stanford University · Princeton University · UMass Amherst · USC University of Southern California · Vanderbilt University

and Gov and NGOs and several International Universities as collaborators

CompSustNet

125+ faculty, students, and collaborators!!!

Probabilistic Matrix Factorization + Exponentially-Modified-Gaussian Mixture Model: Multi-Component Background Learning Automates Inference of Background

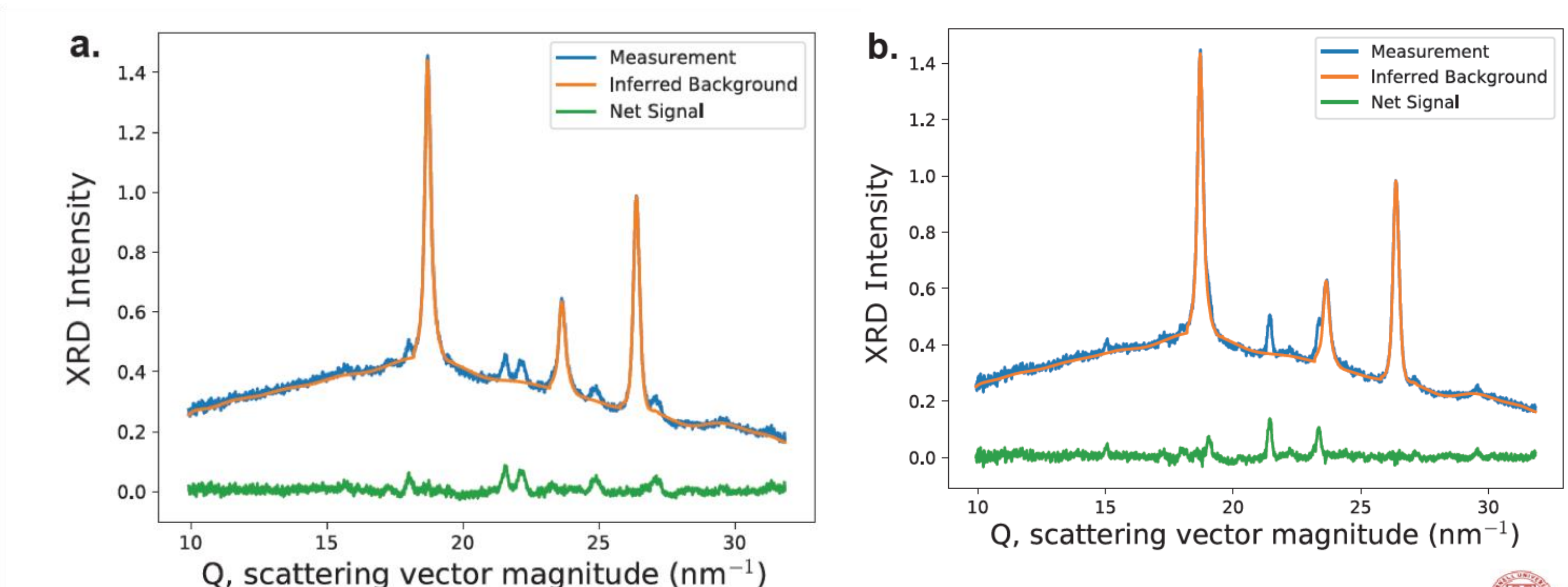
Collective learning for a set of measurements

Fully automated algorithm, no human selected parameters

Multiple signal sources: noise; substrate signal; signal for deposited materials

Applicable to different spectroscopy data.

XRD



Ament, Gregoire, Gomes, 2019

Cornell University Gomes Lab

Gregoire Lab

Probabilistic Matrix Factorization + Exponentially-Modified-Gaussian Mixture Model: Multi-Component Background Learning Automates Inference of Background and Signal of Interest for Spectroscopic Data

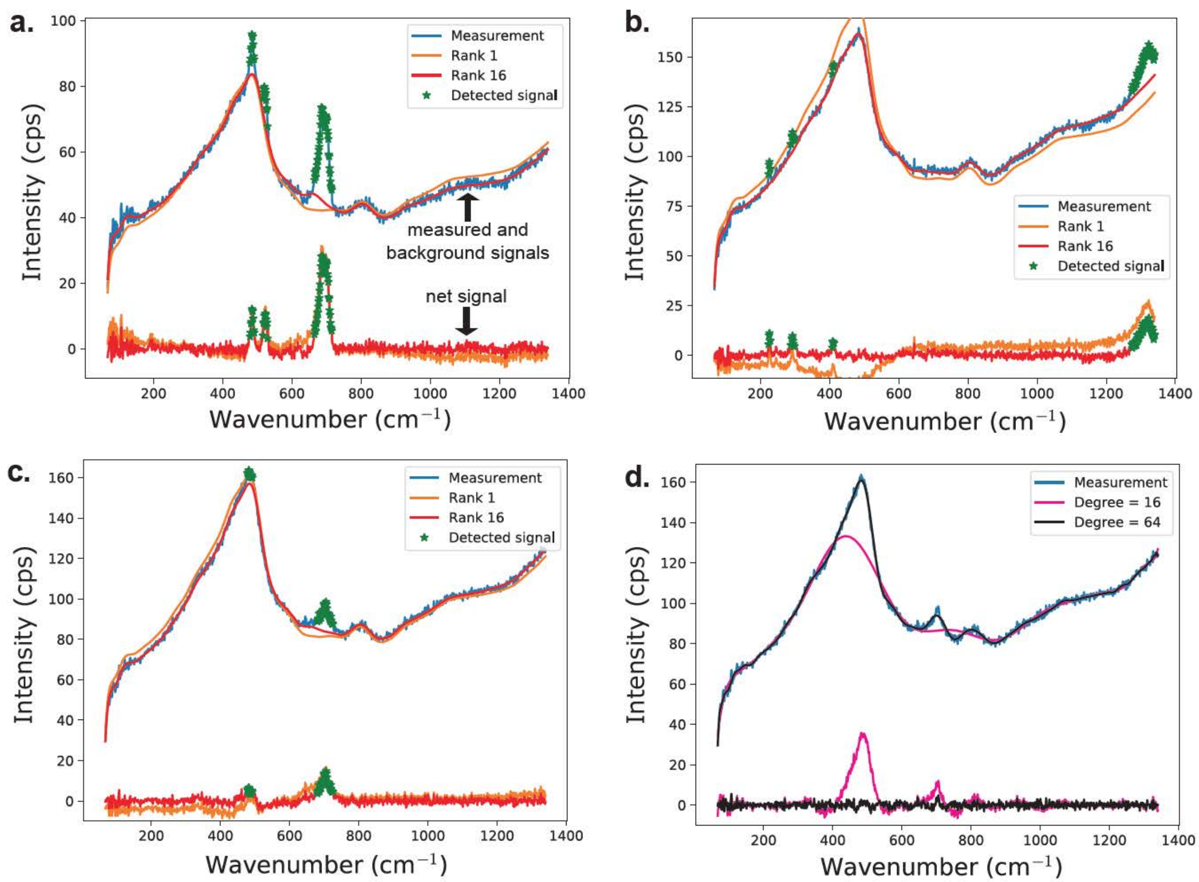
PROBABILISTIC MATRIX FACTORIZATION +
EXPONENTIALLY-MODIFIED-GAUSSSIAN MXTURE MODELS
to automatically infer background and signal of interest

Fully automated algorithm, no human selected parameters

Collective learning for a set of measurements

Multiple signal sources: noise; substrate signal; signal for deposited materials

RAMAN

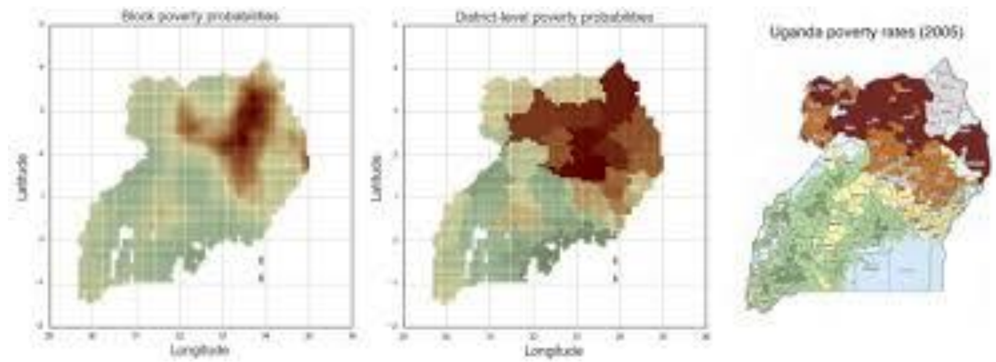


Cornell University
Gomes Lab
Gregoire Lab

Ament, Gregoire, Gomes, 2019

Poverty Mapping: Combining satellite imagery and machine learning to predict poverty

Deep learning and transfer learning



Science AAAS

Neal Jean, Marshall Burke, Michael Xie, W. Matthew Davis4, David B. Lobell, Stefano Ermon,

Science 19 Aug 2016:
Vol. 353, Issue 6301, pp. 790-794
DOI: 10.1126/science.aaf7894

Stefano Ermon: IJCAI Computer & Thought Award –