Computational Landscape Ecology: new opportunities for scaling up

Jeffrey A. Cardille
Associate Professor
Natural Resource Sciences
and McGill School of Environment

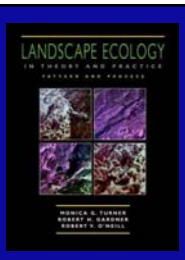
jeffrey.cardille@mcgill.ca

Biography

- Personal
  - Born in Western Pennsylvania
  - Undergraduate: Carnegie Mellon, Pittsburgh
  - Applied Math / Computer Science
- Bellcore Research in New Jersey
  - Evening course in remote sensing
- UW-Madison
  - MS, PhD in Remote Sensing/GIS program (J. Foley)
  - Postdoc: Simulation modeling in thousands of lakes in N. Wisconsin
- UQAM
  - Postdoc, lake modeling
- Université de Montréal, 2006 - 2012
  - Dépt de Géographie
- McGill, July 2012 - present
  - Dept of Natural Resource Sciences, McGill School of Environment

In a nutshell:

- *Computational Landscape Ecology*: Pattern vs Process. Large Areas. Ideal time to think big.
- Computer-intensive approaches to mapping/estimation problems across big areas
  - Estimating lake carbon with satellites
  - Modeling forest connectivity in all of Canada
  - Use of representatives in landscape ecology



Research Priorities in Landscape Ecology

- **Spatial heterogeneity and ecosystem processes**
- **Relating landscape metrics to ecological processes**
- **Sampling over large regions**
- **Causes and consequences of land-use change**
- **Thresholds, nonlinearities, and rules for scaling**
- **Feedbacks between ecosystems and organisms in space**

Some of my Approaches

Computational Landscape Ecology: Pattern vs Process. Large Areas.

One Lake

Data Mining, noise filtering

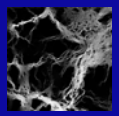
Many Lakes

Google Earth Engine: scaling way up, parallelization

Some of my Approaches

Computational Landscape Ecology: Pattern vs Process. Large Areas.

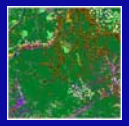
Connectivity Surface



Computer vision to detect features

Image Morphology algorithms to simplify and interpret

Real-world animal observations

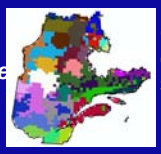


Optimization to see world from animals' perspective

Some of my Approaches

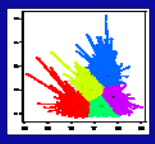
Computational Landscape Ecology: Pattern vs Process. Large Areas.

Representative Landscapes



Clustering landscape characteristics across space

Representative Lakes



Representativeness vs leverage of new data in estimation

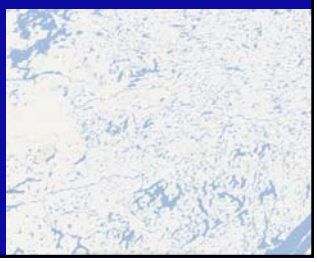
Greatly increased prospects in coming years

- **Free Landsat Data archive back to 1984 (started 2009)**
- **Google Earth: Historical imagery (started 2009)**
- **New working Landsat 8 (2013)**
- **European satellites (2014, 15)**
- **Google Earth Engine: access and tremendous power (2011)**
- **New algorithms for change detection (2013)**

Lake Carbon Content through Color

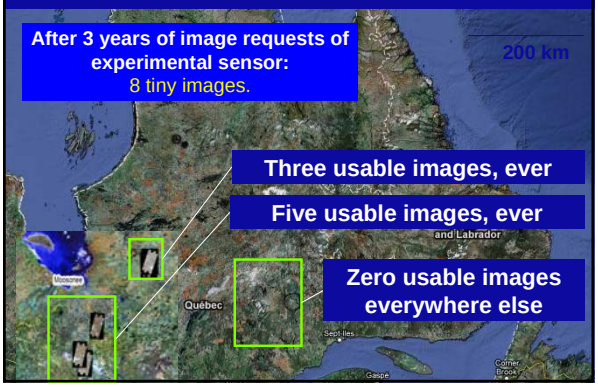
What is the carbon content of boreal lakes? How can we best merge existing field samples with satellite data?

Why? Earth's carbon budget

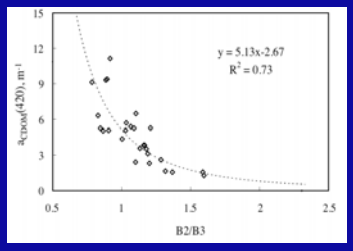


1. **State of the Art**
2. **Pilot Project :**
   - ALI sensor, a **prototype** of Landsat 8
3. **Current (2013):**
   - Landsat 8

We have always worked in a data-poor environment



Experimental ALI, and now Landsat: Green / Red

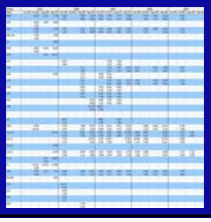


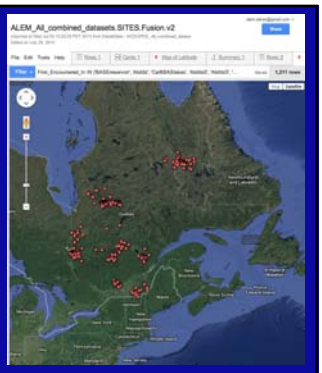
Kutser et al. 2005

Dark Lakes → **Clear Lakes**

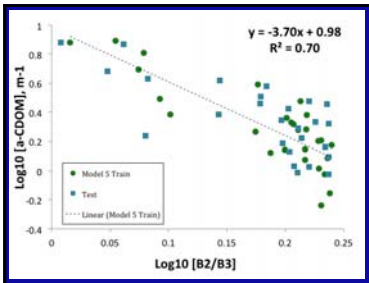
Legacy database of fieldwork in Quebec: can we mine it?

- 1400+ samples
- 400 lakes
- Some repeat visits
- Sparse information, temporally & spatially





Home-grown environmental data parser
Stored: Google Fusion Tables



Yes, we can mine this older data

Increased available information for model fitting/testing by 2 orders of magnitude

Next Question:

Can we use Landsat 8 for this same task?

- near-identical specifications
- better signal:noise ratio

How rare is the launch of a Landsat?

- **Landsat 7 : launched *14 years ago***
- **Landsat 5 : launched *29 years ago***

Pretty much all Landsat work you've seen in the past uses one of these two satellites

We're now suddenly facing:
Facing "big-data" research questions

- What should we do when we have 10 or 20 or 50 new satellite images of varying quality?
- What should we do when we get new field data?

How can this be best answered?

- Old Strategy: lake-by-lake work scraping data from one or two hard-won images
- New Strategy: overwhelm the noise

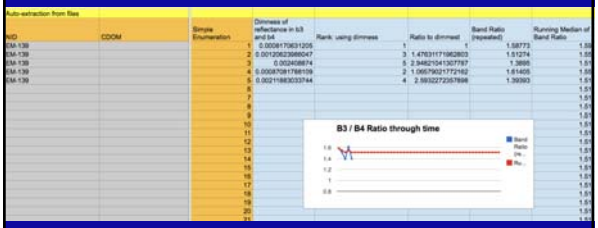
How to do this?

Data Mining, noise filtering, etc

Tracking/analysing data measures of data quality/depth over hundreds of thousands of lakes

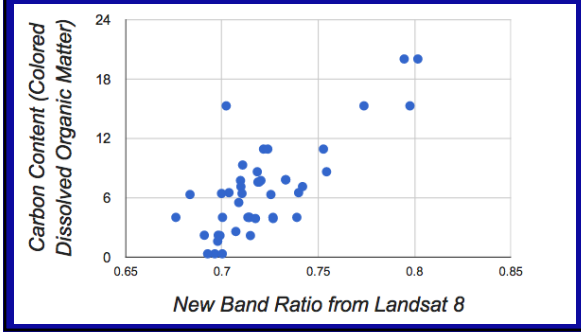
Goal: Distinguishing annual/seasonal cycles in carbon content from other variability

Summer 2013 Prototype for Estimation Engine for Landsat 8

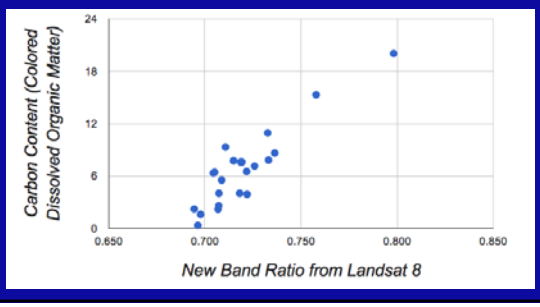


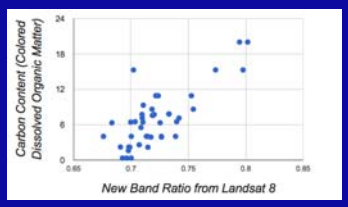
Question:
Can we use Landsat 8?

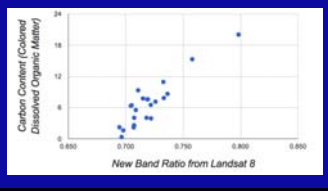
Carbon vs Landsat 8: all images: before median compression



Carbon vs Landsat 8: all images: after median compression







Carbon vs Landsat 8: all images: after median compression

So yes, we can probably use Landsat 8
But *much* more is waiting

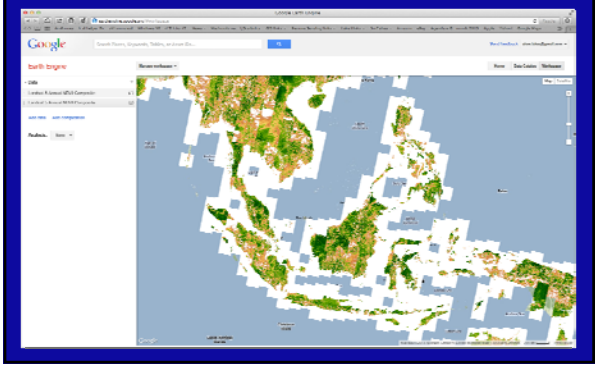
Great summer project

But now at limitations of download space, workstation speed.

- **Important longstanding questions,**
- **Now feasible to address since summer 2013**

Google Earth Engine: instantaneous calculations that took years

impossibly large data sets too big to download



***Science* Brendan Frey, U Toronto**

REPORTS

Clustering by Passing Messages Between Data Points

Brendan J. Frey* and Delbert Dueck

16 FEBRUARY 2007 VOL 315 SCIENCE www.sciencemag.org

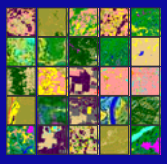
COMPUTER SCIENCE

Where Are the Exemplars?

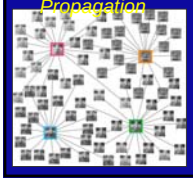
Marc Mézard

A fast way of finding representative examples in complex data sets may be applicable to a wide range of difficult problems.

Intercomparable landscapes



Affinity Propagation



What are the representative landscapes of the continental USA?

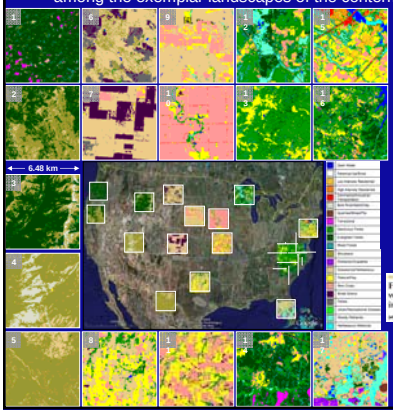
Define criteria
~100 landscape metric values for each
i.e., a big table of characteristics

Define similarities
Euclidean distance among values of these characteristics

In a big set of landscapes
~ 10,000 landscapes of USA

Algorithm finds groups, then selects a representative for each

A clear human signature is nearly ubiquitous
among the exemplar landscapes of the conterminous United States

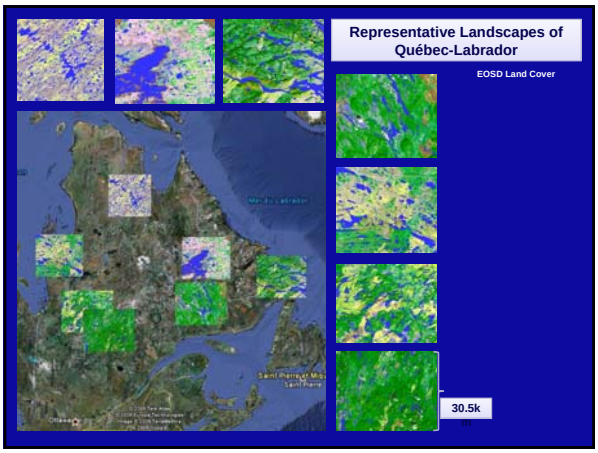


Of 17 exemplar landscapes for the US, 12 contained readily visible agriculture

a powerful way to summarize a data set objectively, efficiently, and clearly.

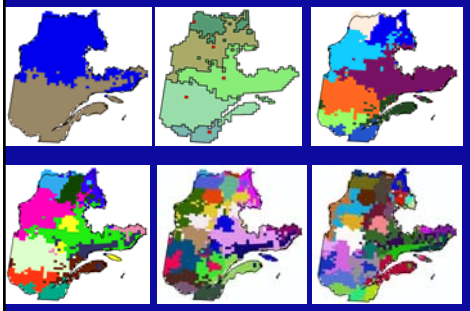
Individual Exemplar landscapes summarize the groups they represent

***Frontiers in Ecology and the Environment*: Vol. 8, No. 3, pp. 130-134.**



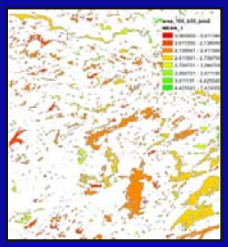
Clusters have a remarkable spatial consistency,

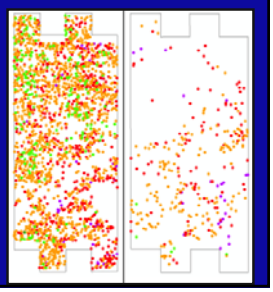
even though only patterns and not coordinates are used to estimate similarity

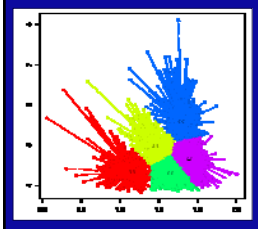


Idea: Representative lakes for greater sampling efficiency

Why? Visiting lakes is very costly (think 'helicopter time')

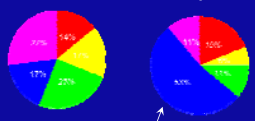






Observed

Sampled



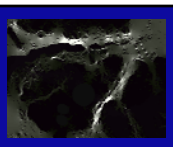
Too many big light lakes have been visited

We can identify representative lakes for future sampling

In a nutshell:

- *Computational Landscape Ecology*: Pattern vs Process. Large Areas. Ideal time to think big.
- Computer-intensive approaches to mapping/estimation problems across big areas
  - Estimating lake carbon with satellites
  - Modeling forest connectivity in all of Canada
  - Use of representatives in landscape ecology

Computational Landscape Ecology: new opportunities for scaling up



Jeffrey A. Cardille
McGill University
Natural Resource Sciences
and McGill School of Environment

jeffrey.cardille@mcgill.ca