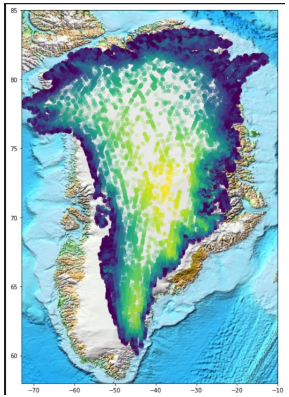


GAGA: A graph-theoretic approach to Greenland surface modeling

Noah Bergam
Advisor: Abani Patra
VERSEIM NSF REU 2023

Table of Contents



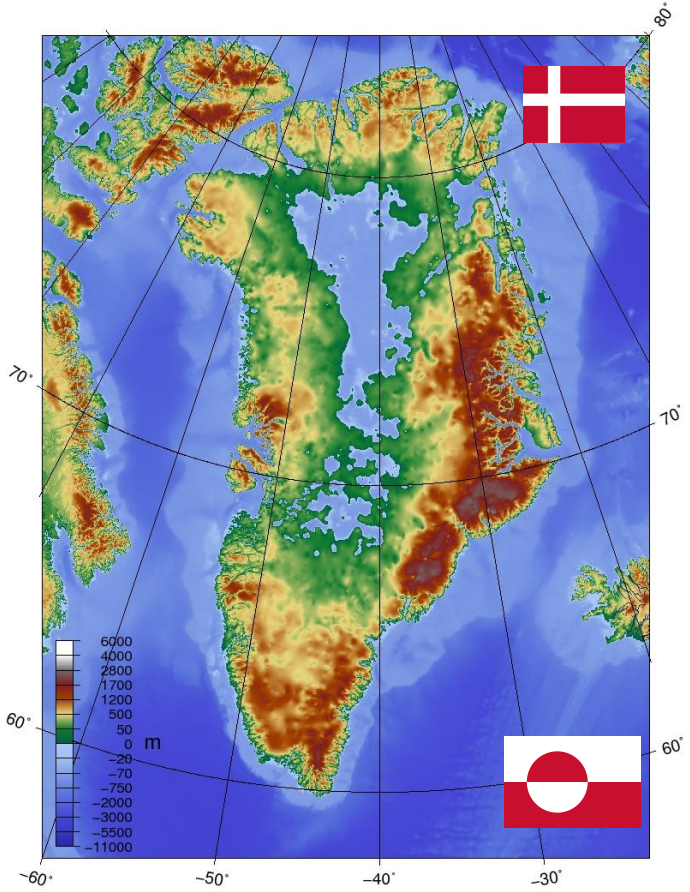
1. Background
    a. Greenland and remote sensing
    b. SERAC (2014): grid-based surface model
2. Contribution
    a. GAGA (2023): graph-based surface model
    b. Error Analysis
    c. Simple statistical model of IceSAT-like measurement

Notes on Greenland

2.1 million km^2 — *1.7 million km^2 is ice.*
Population: 56,000.
Autonomous territory of the Kingdom of Denmark

Melting, much faster than Antarctica.

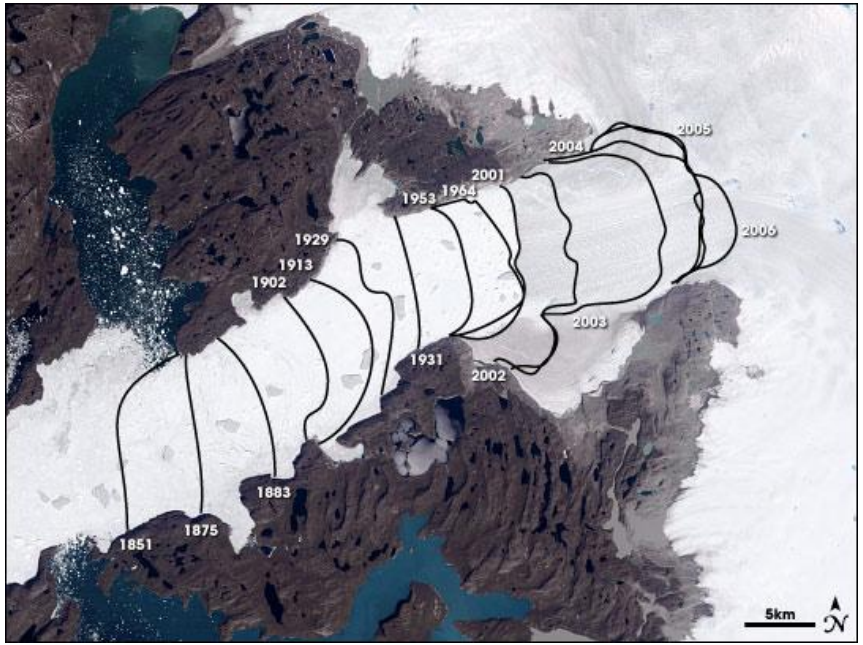
If fully melted, sea level would rise by 7 meters…

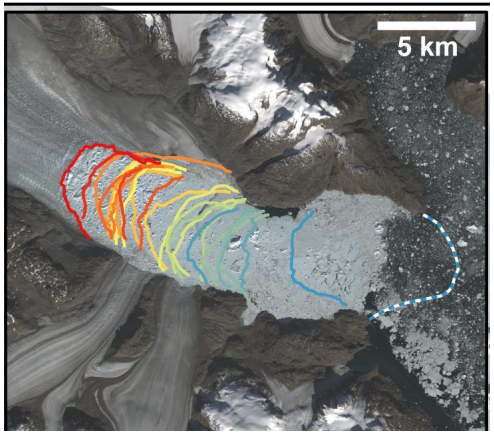


Major Glaciers

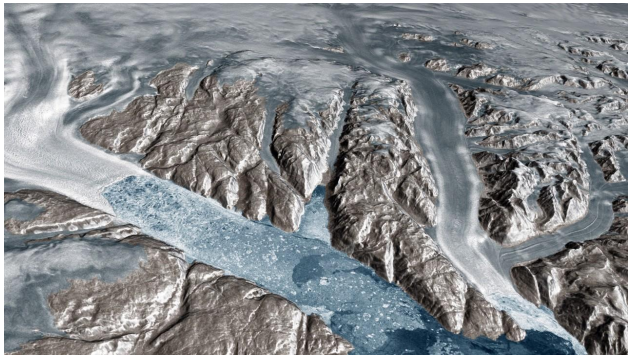
Jakobshavn Isbrae (Southwestern Greenland)

- Drains 6.5% of the Greenland ice sheet, produces 10% of its glaciers

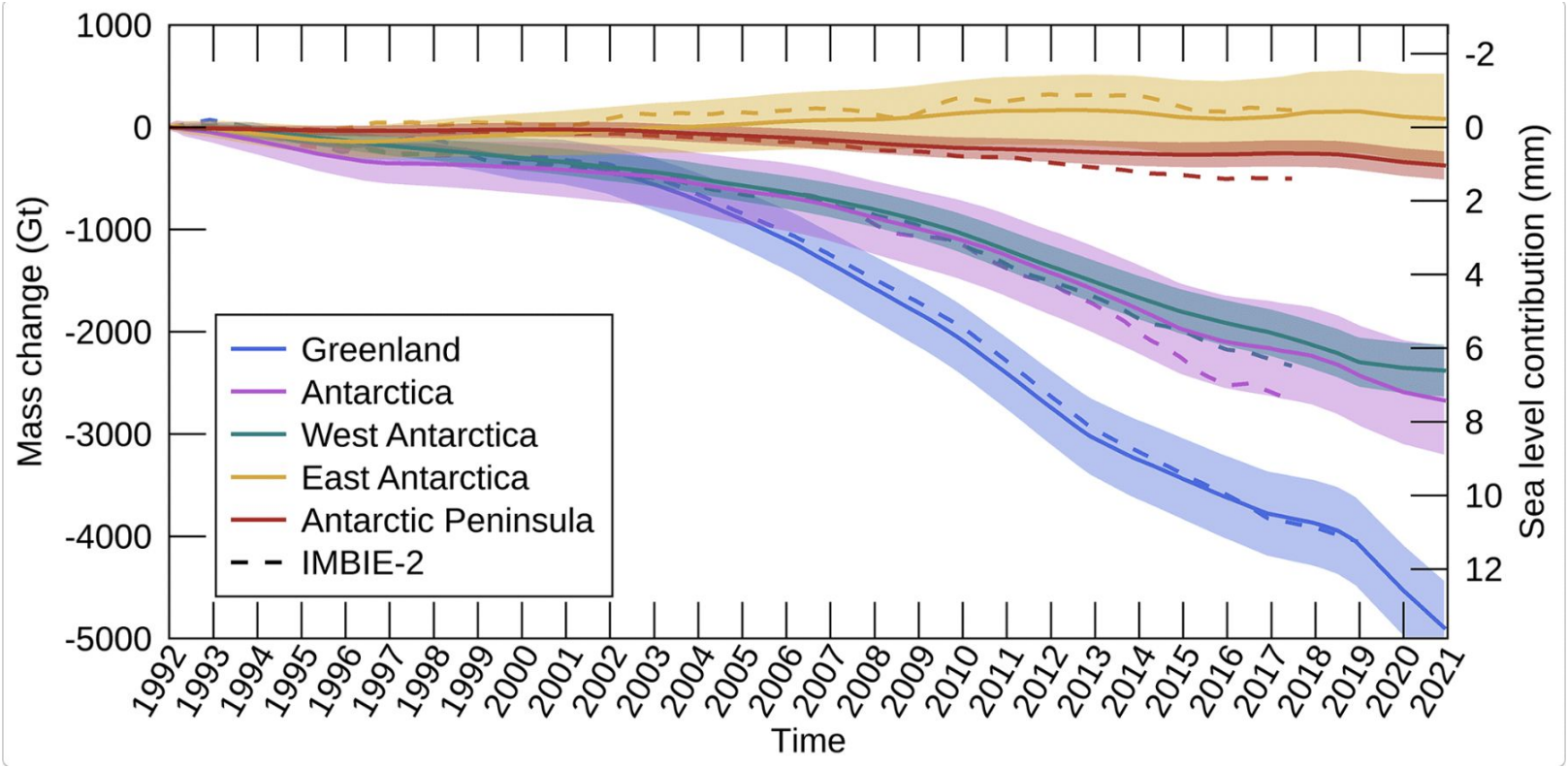




Kangerlussuaq Glacier (Southeast)



Helheim Glacier (Southeast): dynamic thinning, some thickening



International Mass Balance Inter-Comparison Exercise (IMBIE 2016)

Cross-validate many measurements (laser/radar altimetry, gravimetry, climate model)

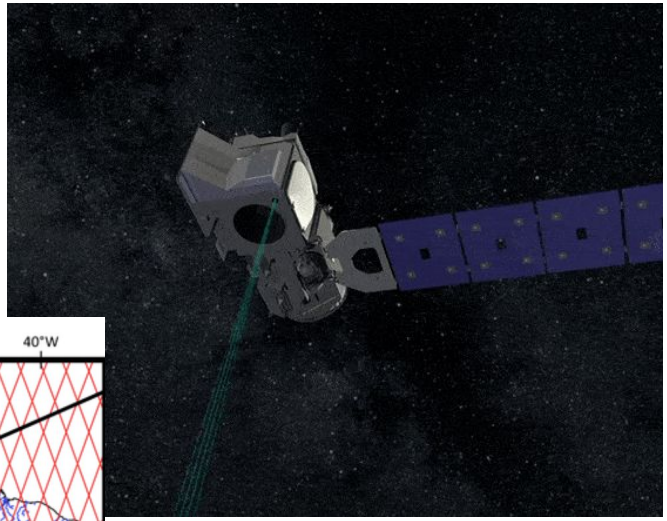
IceSAT and Laser Altimetry

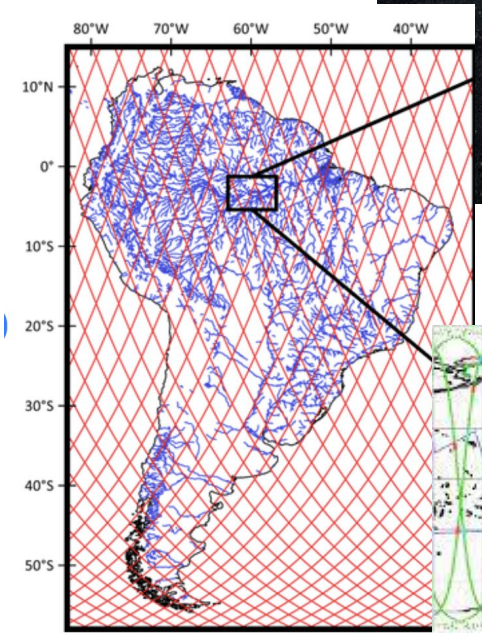
IceSAT-1: 2003 to 2010

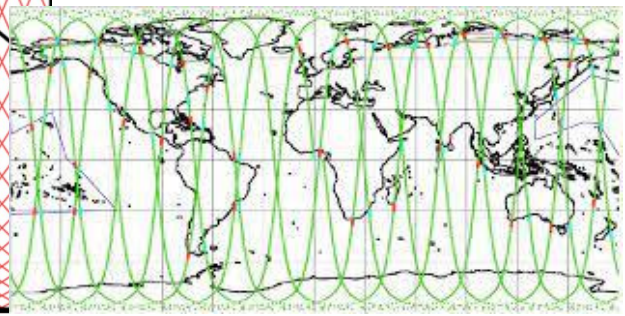
IceSAT-2: 2018 to present

- Shoots down laser, retrieves the signal. Probably the highest precision remote sensing tool (but not perfect!)
- 91-day cycle

QUESTION: How to turn these measurements into a surface model, for volume (and mass) estimation?





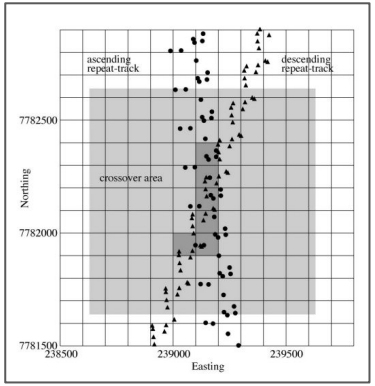


SERAC (2014) Methodology

Focus on crossover points (where satellite tracks intersect)

Treat area as a raster grid (2km by 2km). Estimate non-measured areas using Kriging interpolation.

Run polynomial regression for the time series at each grid point



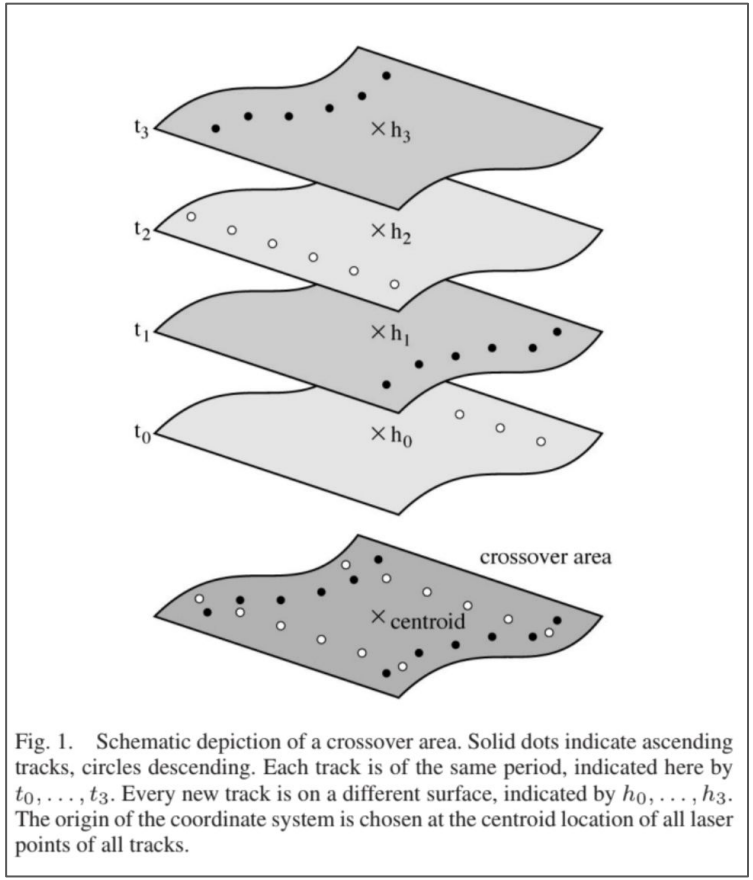


Fig. 1. Schematic depiction of a crossover area. Solid dots indicate ascending tracks, circles descending. Each track is of the same period, indicated here by $t_0, \ldots, t_3$. Every new track is on a different surface, indicated by $h_0, \ldots, h_3$. The origin of the coordinate system is chosen at the centroid location of all laser points of all tracks.

SERAC = Surface Elevation Reconstruction And Change detection

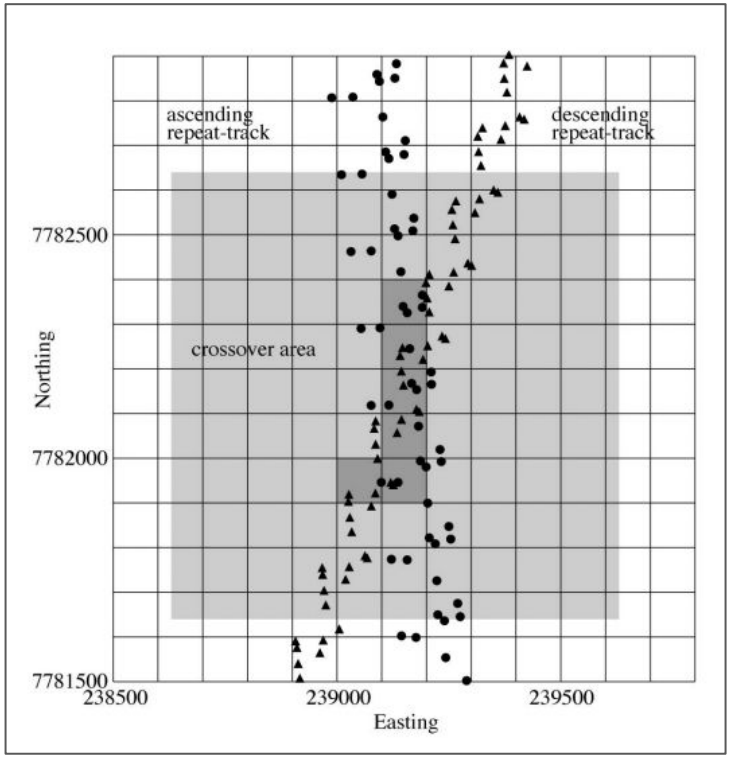
SERAC - Analysis

Benefits:

- Automatic crossover detection →
- Regular. Explainable. Relatively accurate.

Drawbacks:

- Rigid. The grid pixel size is arbitrary.
- Polynomial interpolation (for the time series) is fragile.
- Could use more strategic error minimization.



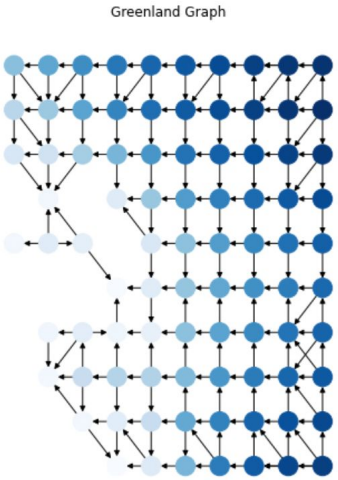
Method by which SERAC detects crossover points

Introducing GAGA (2023)

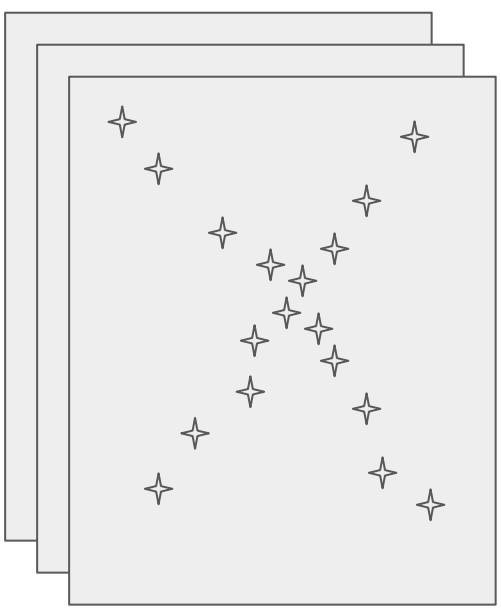
Proposition: Instead of grid, use a *graph data structure.*

Motivation:

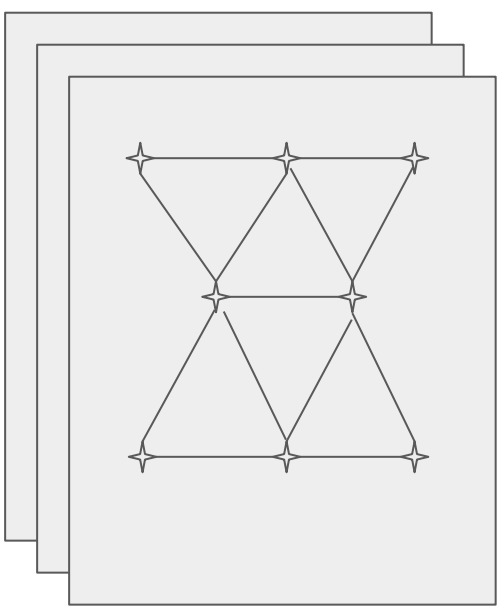
- Flexible. Localized. Easy to densify / prune.
- Lower error.
- Easy to implement the following:
  - Second-order change detection (Laplacian)
  - Crossover detection
  - Graph traversal, flow modeling



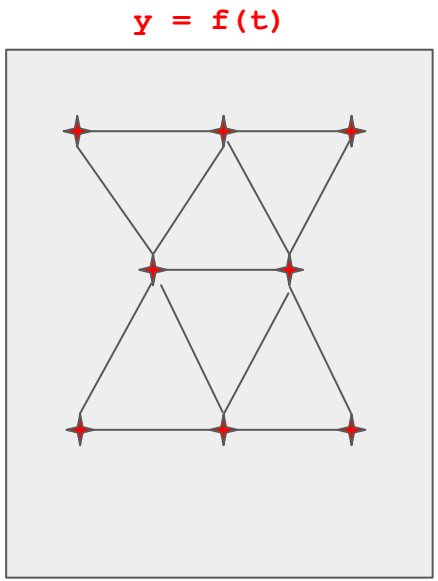
Pipeline: Combine Kriging, triangulation, and ALPS for a stronger, continuous-time surface model.



Original, scattered measurements



“Synthetic measurements” taken via Kriging interpolation.



Time series at individual grid points, using ALPS

A Brief Introduction to Kriging

and the mindset of geostatistics

(chalkboard)

GAGA Algorithm Description

Granularity epsilon is hyperparameter

Great for visualization

Used for volume change estimation

Algorithm 1 GAGA/SERAC++ surface model

Require: $M = \{(x_i, y_i, h_i, t_i)\}_{i \in N}$ bounded measurements; in particular, say $t_i \in [0,1]$ and $(x_i, y_i) \in S \subset [0,1]^2$.

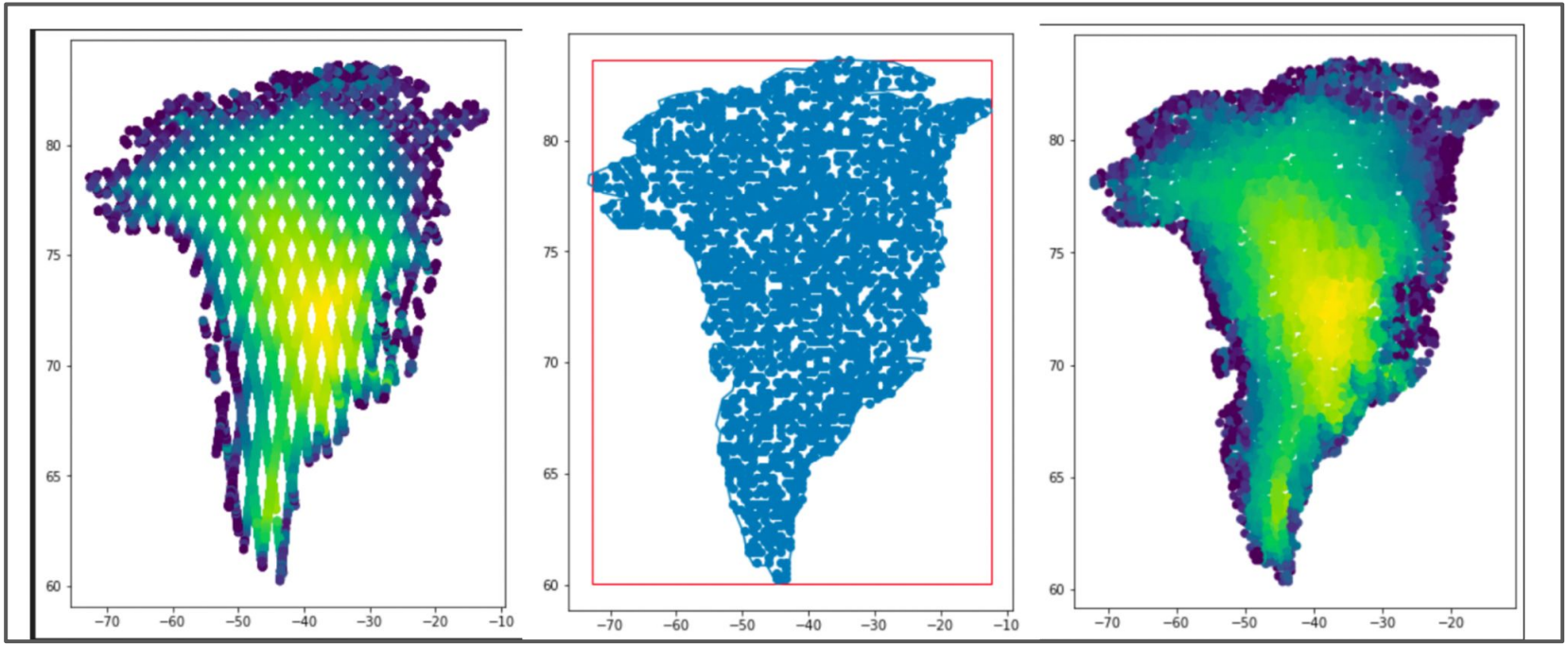
1: Select time steps $(t_0, ..., t_m)$ with $t_0 = 0$ and $t_m = 1$.
2: Select ϵ-net of S, denoted S_ϵ.
3: Construct Delaunay triangulation of S_ϵ with centroids D.
4: **for** $[t_i, t_{i+1})$ with $i \in \{0, 1, ..., N\}$ **do**
5: Krige-interpolate $z_i(d)$ for each centroid $d \in D$.
6: Run `ALPS` (penalized B-spline regression) over each centroid time series $\{z_i(d)\}_{i \in [N]}$.
7: Return the following:

$\{(x_t, y_t, h_t)\}_{t \in [0,1]}$ (continuous-time surface model)

$\left(\frac{dh}{dt}, \sigma_h^2\right)(d, t_i)$ for $d \in D, i \in \{1, ..., N\}$

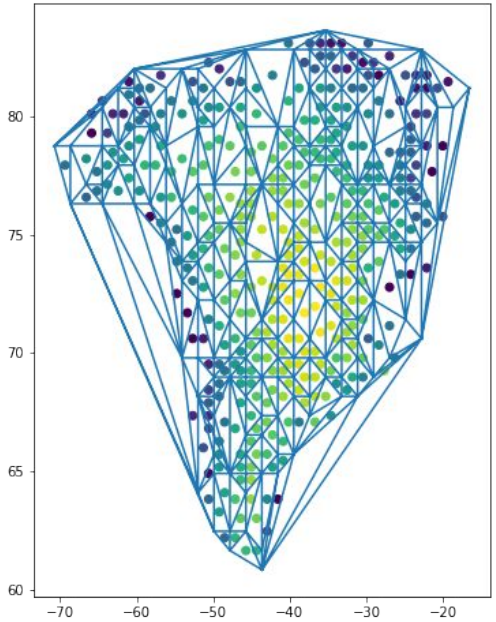
Would be great to show convergence properties!

An illustration of this process



Original measurements (left)
Placement of synthetic measurements, at centroids of Delaunay (middle)
Actual synthetic measurements, taken with nearest neighbors (right)

Triangulation Details



Instead of a grid we use the centroids of a Delaunay triangulation to discretize.

Easy to calculate the **area of a triangle** on the surface of the earth

(Current Approx) Use Haversine and Heron's Formula.

(Better Approx) Use Girard's Theorem.

Methods for calculating triangle area on earth.

Haversine-Heron (current)

(1) Calculate the side lengths of triangle using Haversine formula.

(2) Get (approximate) area via side lengths with Heron's formula:

$$A = \sqrt{s(s-a)(s-b)(s-c)}$$

Girard's Theorem (better)

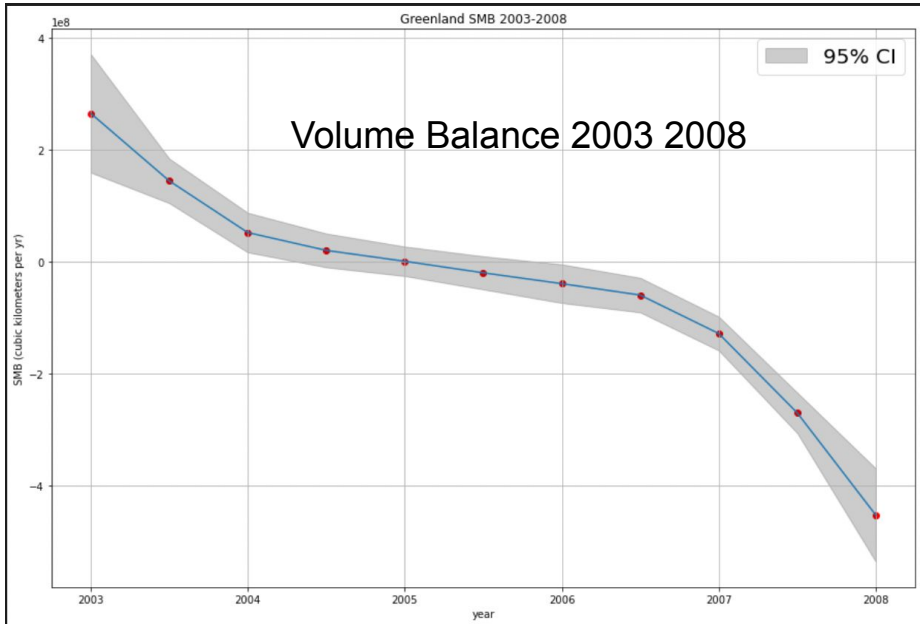
(1) Calculate interior angles of triangle, using law of sines.

(2) Apply Girard's formula:

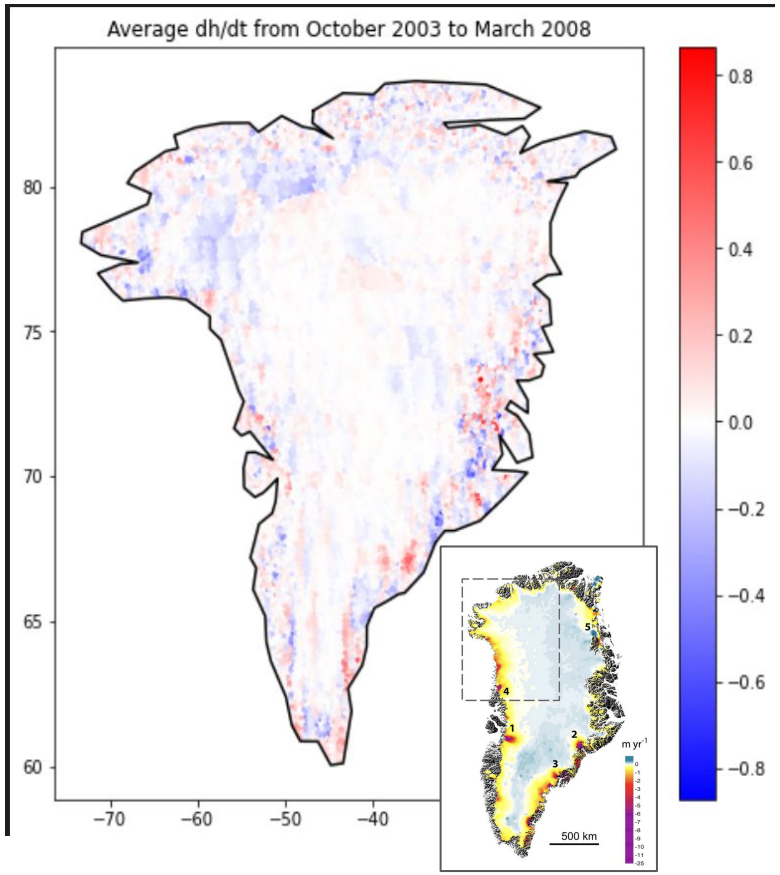
$$A = R^2(\alpha + \beta + \gamma - \pi)$$

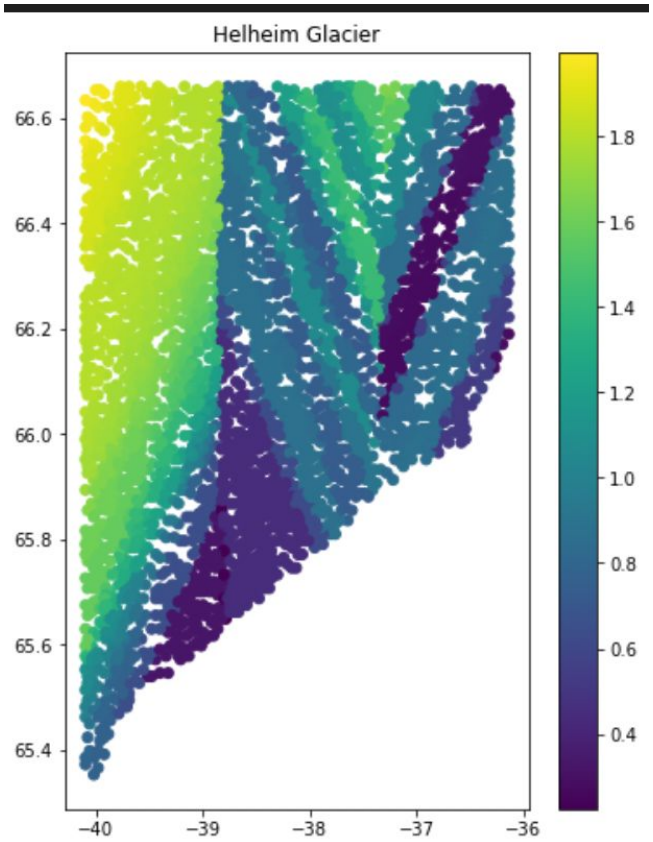
Remark: this formula follows almost immediately from the Gauss-Bonnet Theorem

Estimates and Visualizations

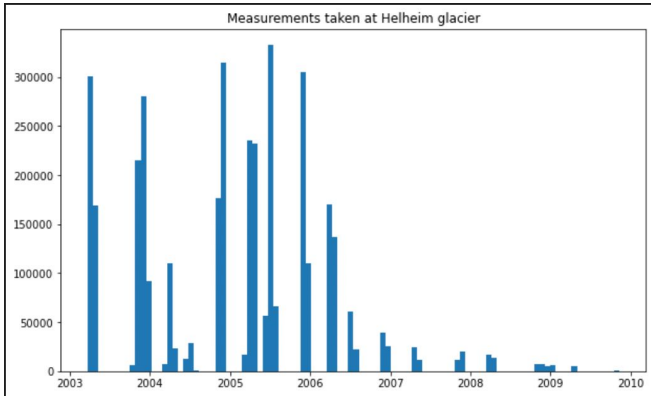


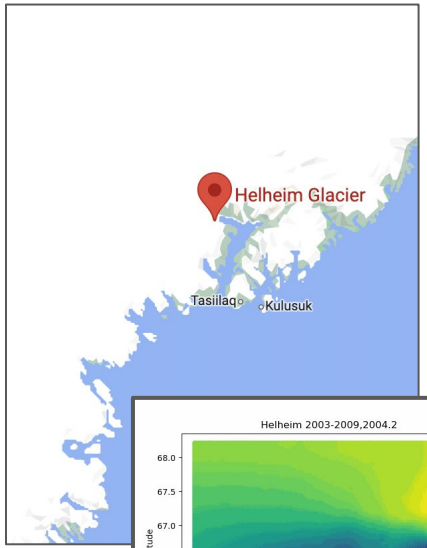
$$\frac{dV}{dt} = \sum_{i=1}^{N} \frac{dh_i}{dt} A_i$$



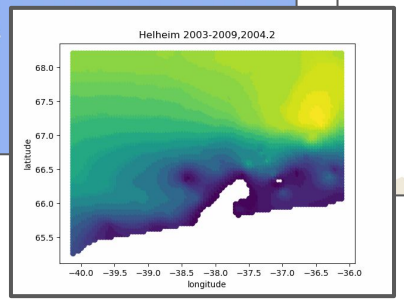

Helheim Glacier
66.6
66.4
66.2
66.0
65.8
65.6
65.4
−40
−39
−38
−37
−36
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4

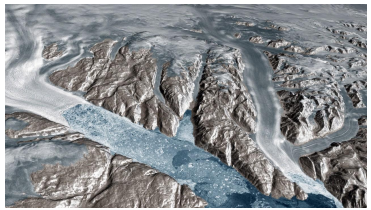
Helheim


Measurements taken at Helheim glacier
300000
250000
200000
150000
100000
50000
0
2003
2004
2005
2006
2007
2008
2009
2010

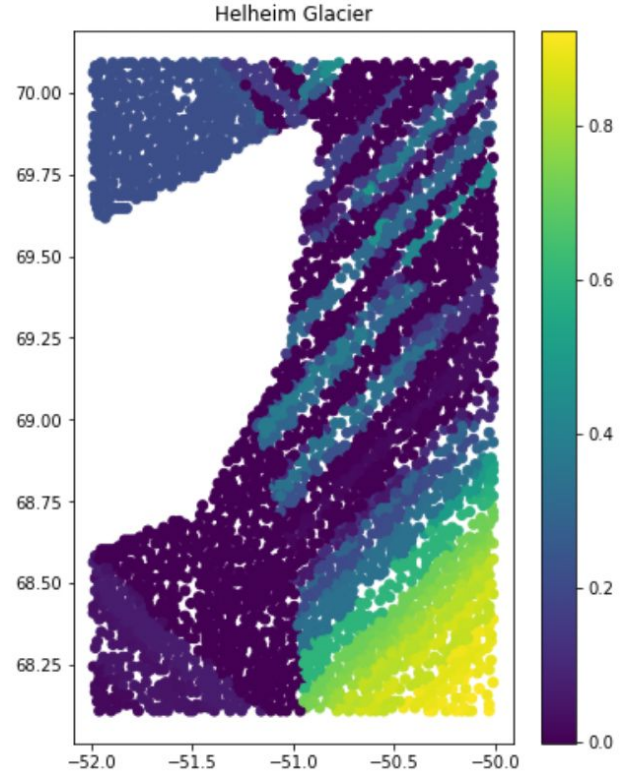

Helheim Glacier
Tasiilaq
Kulusuk

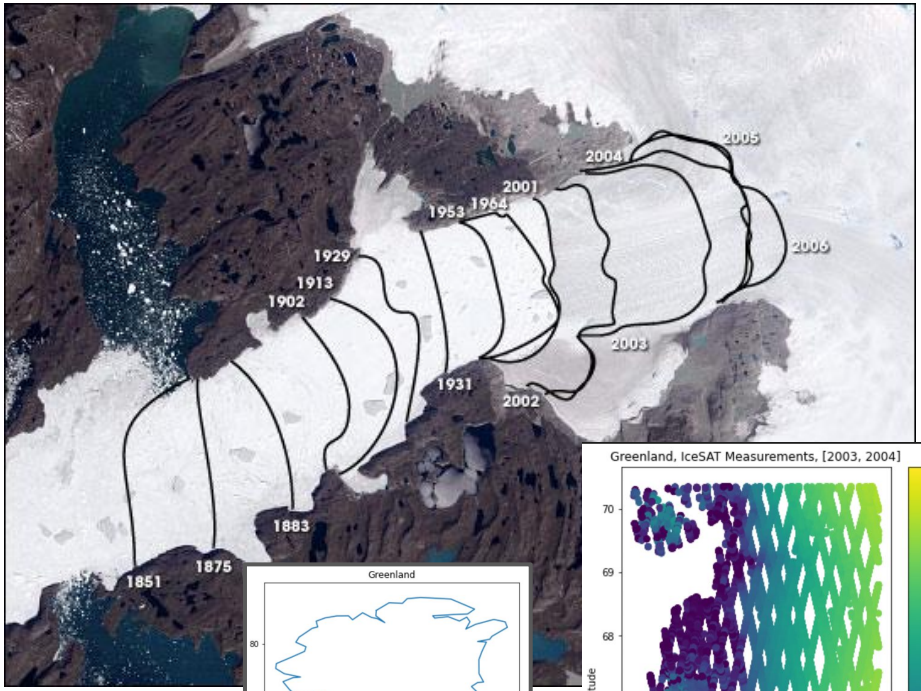

Helheim Glacier
Paarnakajiit

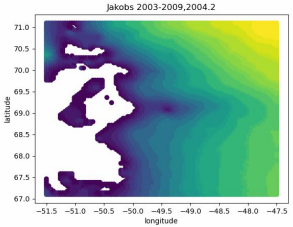

Helheim 2003-2009,2004.2
68.0
67.5
67.0
66.5
66.0
65.5
latitude
−40.0
−39.5
−39.0
−38.5
−38.0
−37.5
−37.0
−36.5
−36.0
longitude

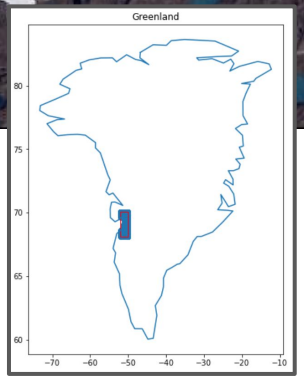


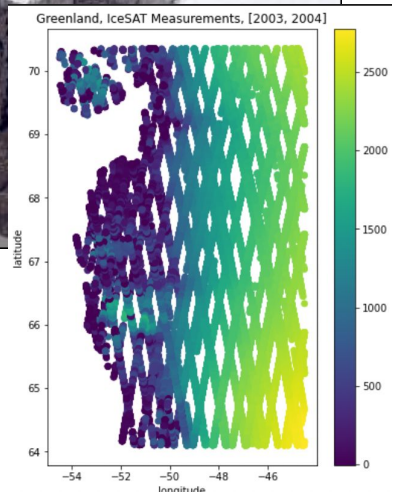
Jakobshavn Glacier







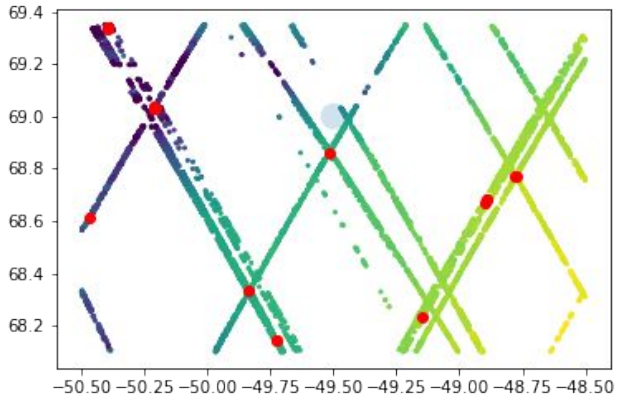




Graph-model benefits

(1) An alternative method for crossover detection, using nearest-neighbors

(2) Graph sparsification / densification



Crossover Detection (beta):

(1) Take intersection of k-nearest neighbor and radius ball graph

(2) If a point has significantly more neighbors than its neighbors, then mark it as a crossover.

Error Analysis

The Model

Assume that we can partition Greenland and estimate height change at each part.

$$\frac{dV}{dt} = \sum_{i=1}^{N} \frac{dh_i}{dt} A_i$$

The estimated height derivatives are random variables. Their variances add.
(assume zero covariance)

$$\sigma_V^2 = \sum_{i=1}^{n} \sigma_{h_i}^2 A_i^2$$

$$\operatorname{Var}(aX + bY) = a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y) + 2ab \operatorname{Cov}(X, Y)$$

Argument

How to reduce this error?

- Incorporate more measurements.
- Assign areas strategically
- Use a better time interpolation strategy (using ALPS!)

$$\sigma_V^2 = \sum_{i=1}^{n} \sigma_{h_i}^2 A_i^2$$

$$\text{s.t. } \sum_{i=1}^{n} A_i = A_{GrIS}$$

$$\sigma_{h_i}^2 = \frac{||y - B\hat{\theta}||^2}{n - 2tr(H) + tr(HH^T)} \left(B_t'(B^T B + \hat{P})^{-1} B_t'^T \right)$$

$$B_i'(t;p) = \frac{p}{t_{i+p} - t_i} B_i(t;p-1) + \frac{p}{t_{i+p+1} - t_{i+1}} B_{i+1}(t;p-1)$$

Heuristics for

1. (SERAC) Assume constant error and area size in estimate.

2. (GAGA) Accept that error varies. Match high error with small area, lower error with large area.

$$\sigma_V^2 = \sum_{i=1}^{n} \sigma_h^2 A_i^2 = \sigma_h^2 A_{GrIS}^2 / n$$

$$\sigma_V = \frac{\sigma_h A_{GrIS}}{\sqrt{n}}$$

$$\sigma_V^2 = \sum_{i=1}^{n} \sigma_{h_i}^2 A_i \leq \frac{1}{n}\Big(\sum_{i=1}^{n} A_i\Big)\Big(\sum_{i=1}^{n} \sigma_{h_i}^2\Big)$$

$$\sigma_V \leq \frac{\sqrt{\Big(A_{GrIS}\Big)\Big(\sum_{i=1}^{n} \sigma_{h_i}^2\Big)}}{\sqrt{n}}$$

* rough estimate, via Chebyshev sum inequality

Key Insight

In some sense, we have: $\sigma_V = \Omega(1/\sqrt{n})$

(Note that the area of Greenland is fixed, and associated error of the derivative is not directly proportional to the number of points)

Error Gambit: increase n using Kriging interpolation (“synthetic measurements”)

- This will add error, but we can control (upper bound) this as a function of the largest radius around which we would interpolate.

Next steps

Aiming for a paper in the coming months. Still need to:

(a) Incorporate more data (ATM, IceSAT-2)
(b) Refine error analysis and “error gambit.” Compare more closely with SERAC.
(c) Proof-of-concept with graph algorithms (esp. edge pruning)
(d) Collaborate with (Csatho, Schenk), the SERAC authors

A statistical model of IceSAT-like measurements

Model of IceSAT-like Surface Reconstruction

Say we have a time varying Gaussian Markov random field Z(x,y,t).

- x, y are spatial coordinates. t is time

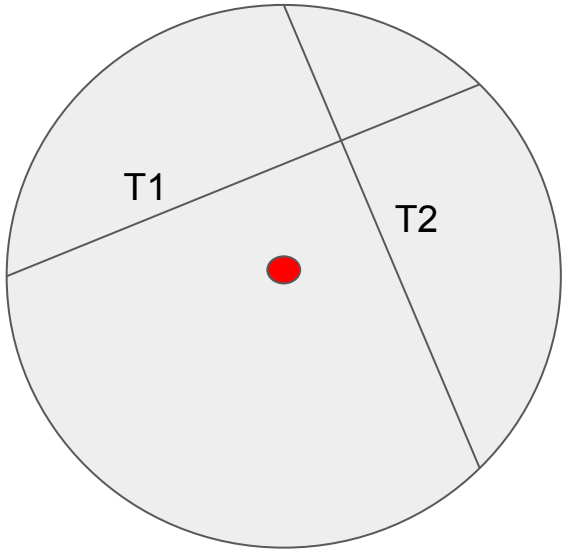
Fix some **query point q** and ball of radius R around q

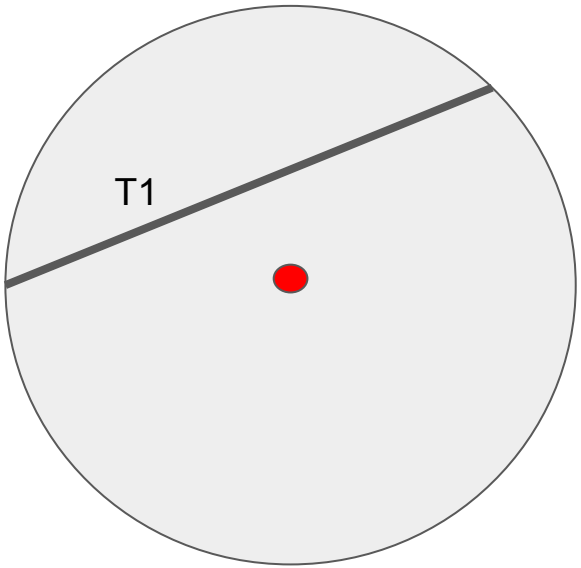
Say we pick **random chords** in this circle at a frequency f. This is a "measurement" of the field at a particular time.

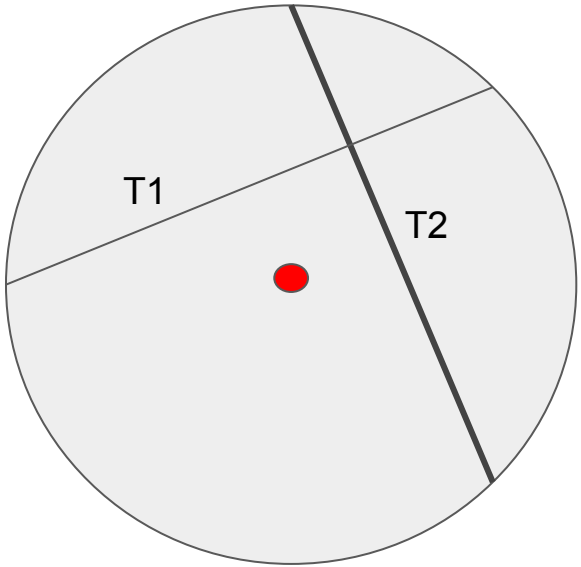
- Condition: *successive chords must intersect!*

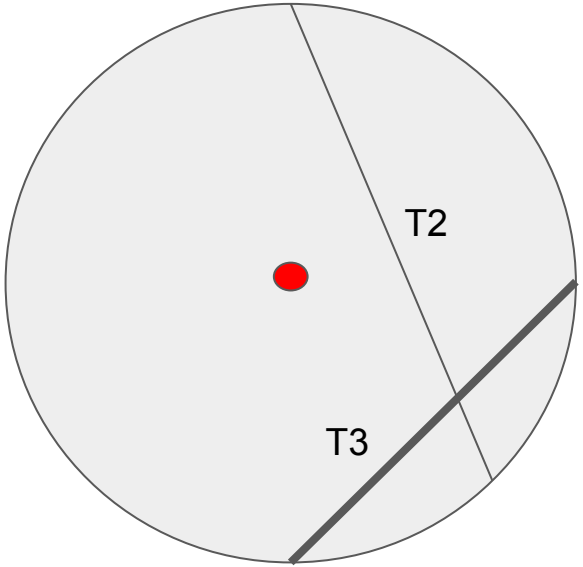
Questions:

- What is the best way to reconstruct the time series of this query point from the chords?
- Can we guarantee convergence to the true function?









Nice property: This a Markov process (next step only depends on the previous)

Approach

Algorithm 1 (nearest):

- At T_i assign query value to closest point on chord
- Interpolate those estimates through time

Algorithm 2 (crossover):

- Interpolate the i-th crossover point over $[T_i, T_{i+1}]$.
- Piece together these time

Question: Which one is better

1) as radius decreases
2) frequency of measurement increases

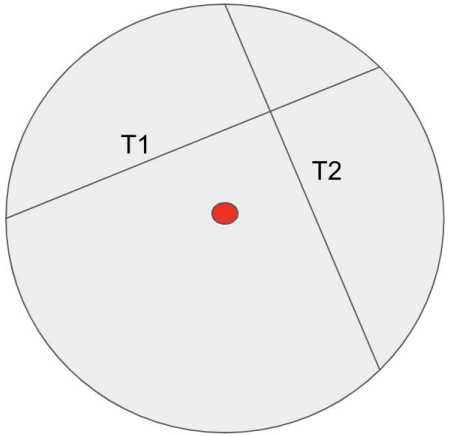
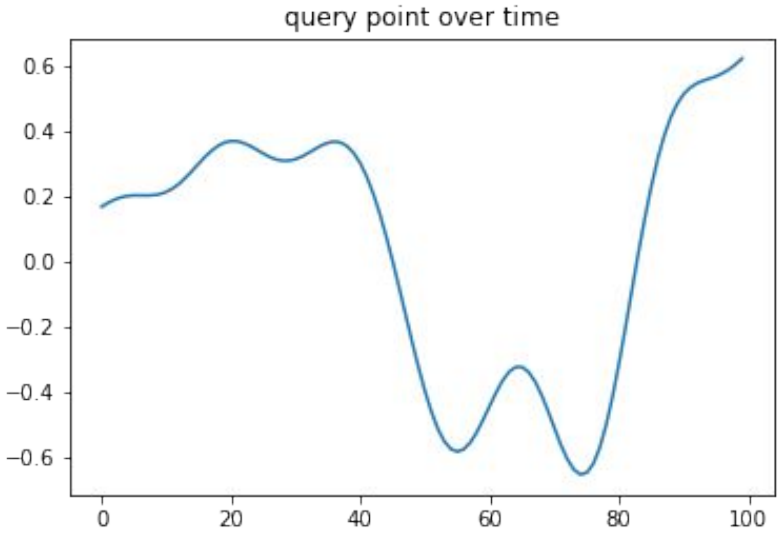
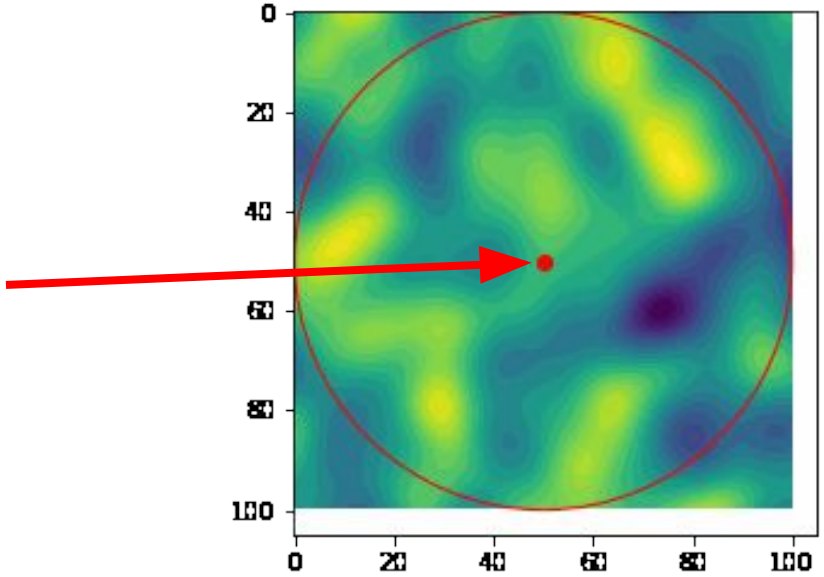


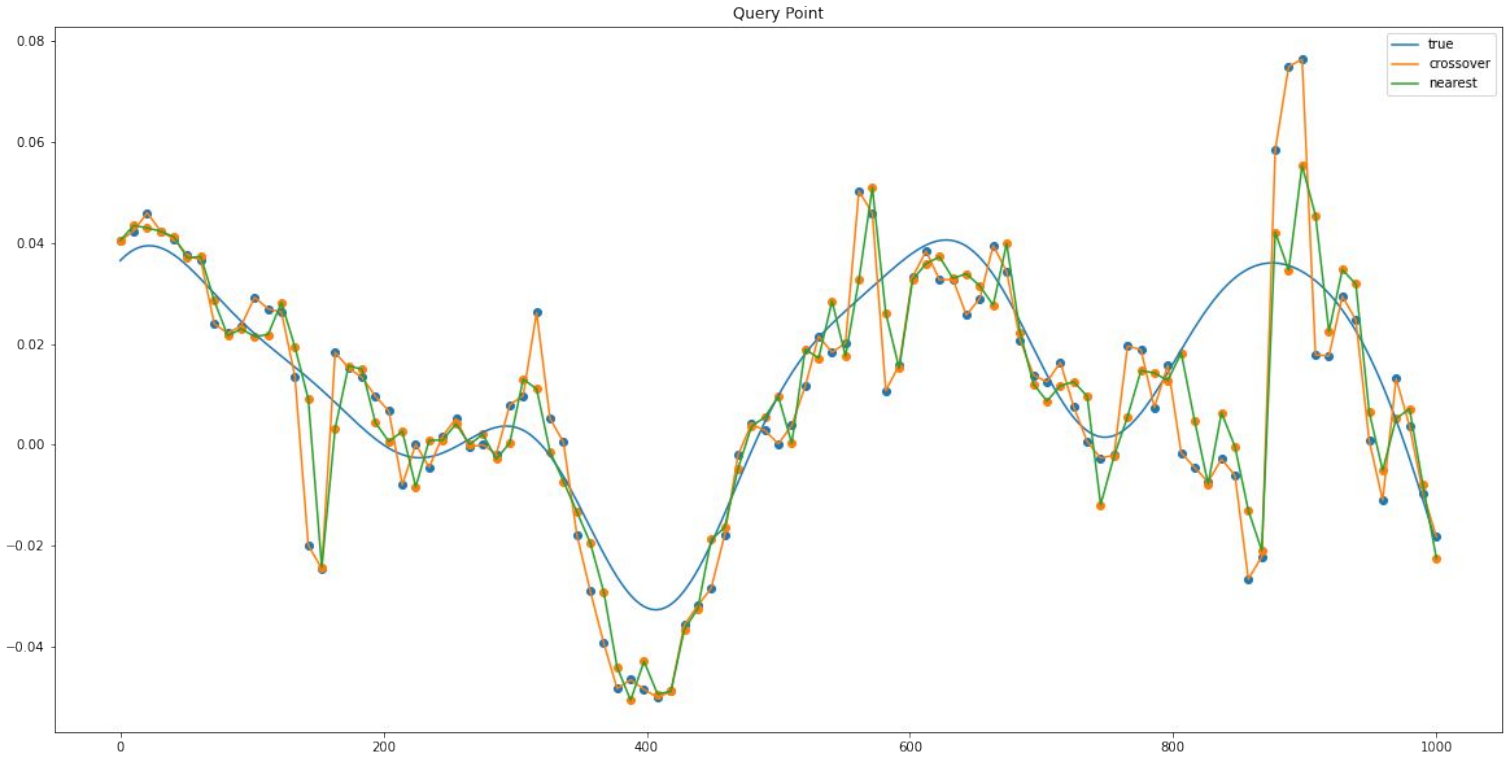
Illustration of the problem.



We want to estimate this time series using the chord-measurements!

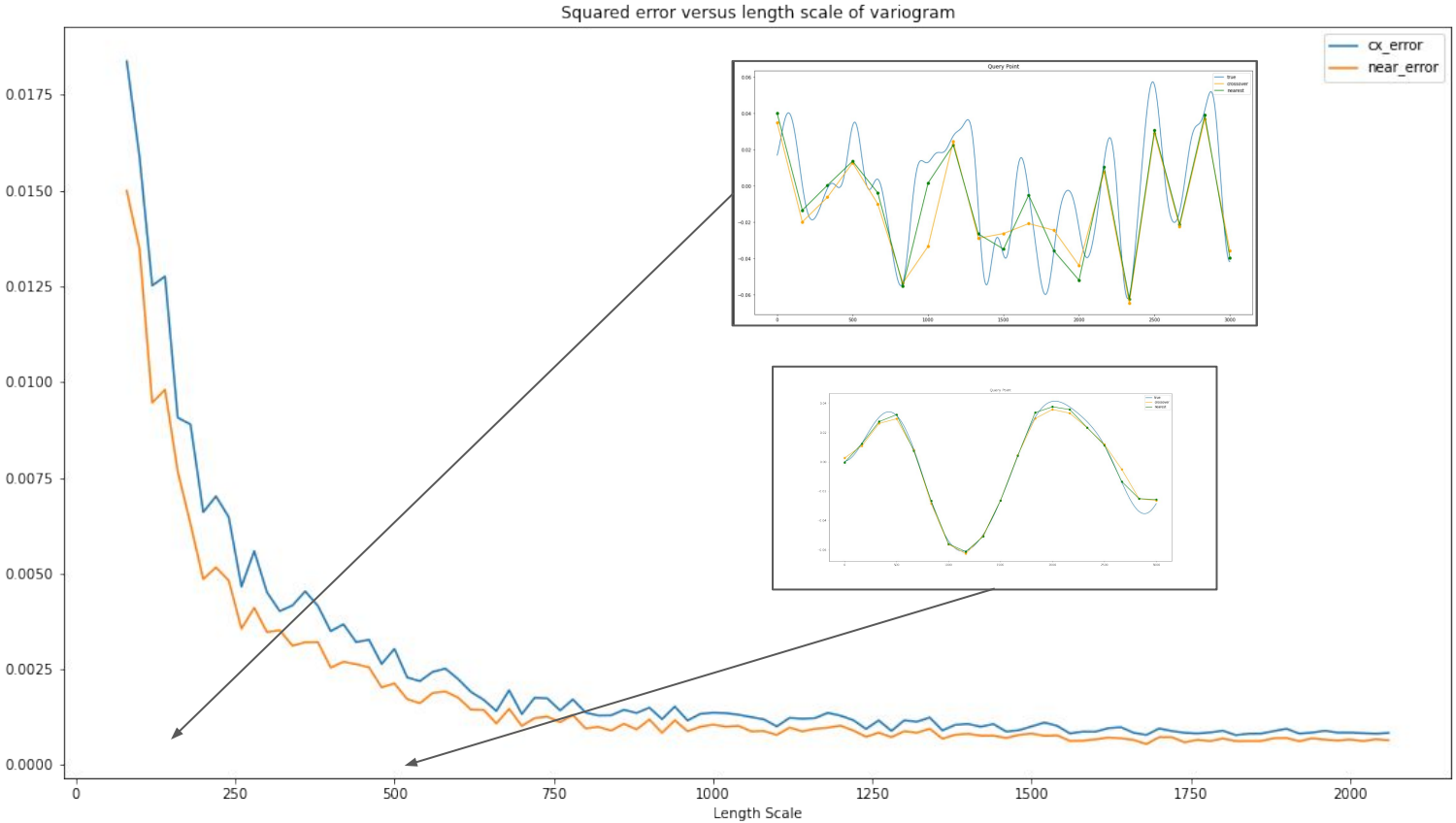
Gaussian random field.



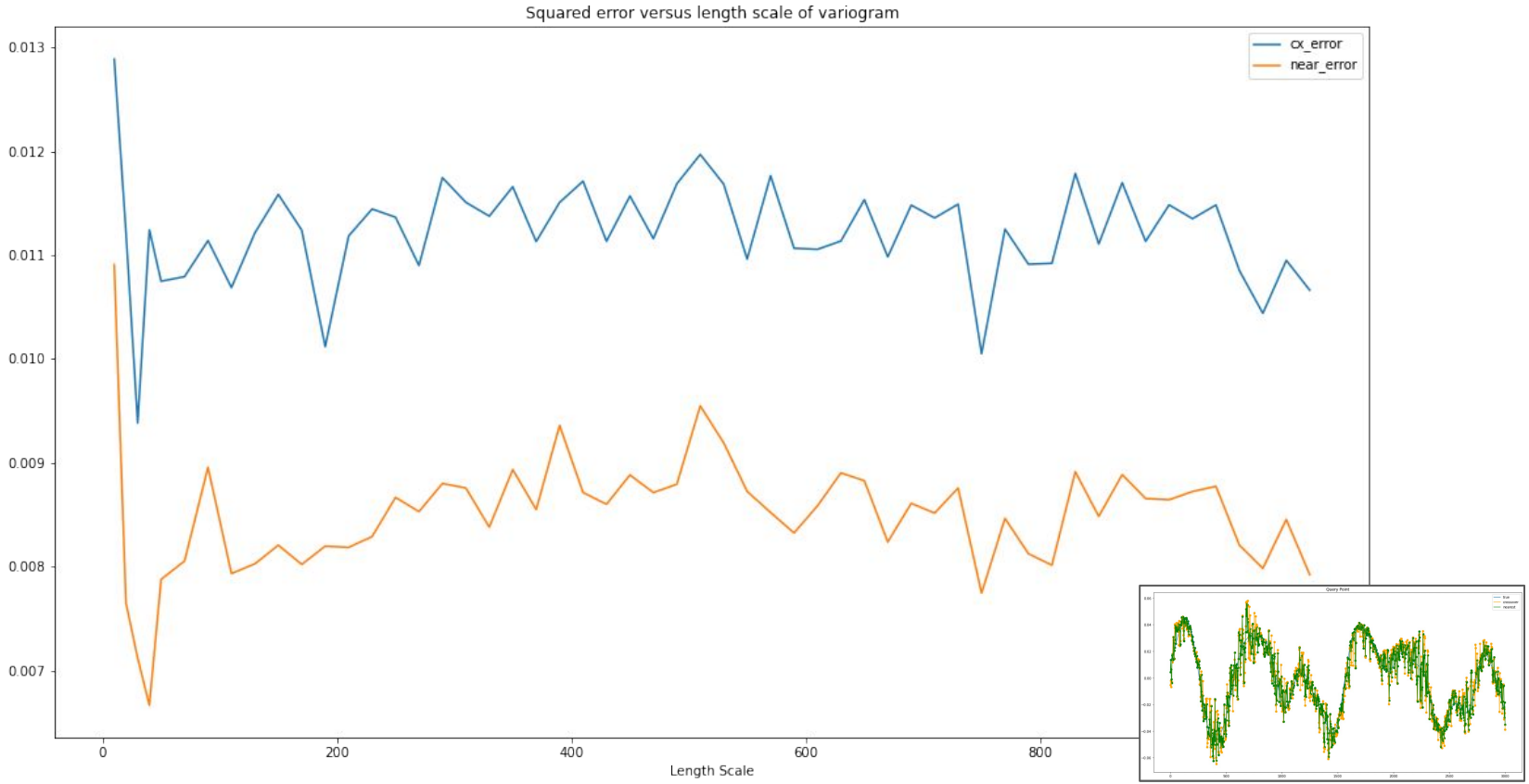


example estimations of query point.

(1) Nearest estimation beats crossover-based
(2) less variance means easier to minimize



Higher frequency measurements does not decrease the error after a certain point!



Discussion

Leveraging crossovers may not be ideal…

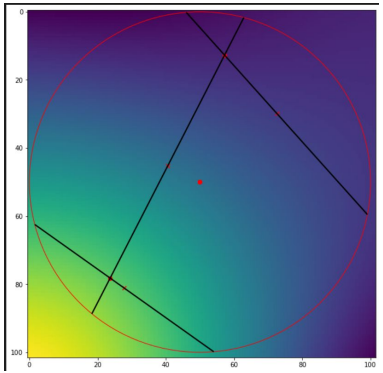
- Is there a (simple) way to leverage crossover points more effectively?

How would GAGA model fare?

- First steps toward a convergence guarantee

Given some field covariance structure, what is the optimal sampling rate?

- In a similar vein as Nyquist-Shannon sampling theorem, but for a Gaussian process…frequency is only in expectation.



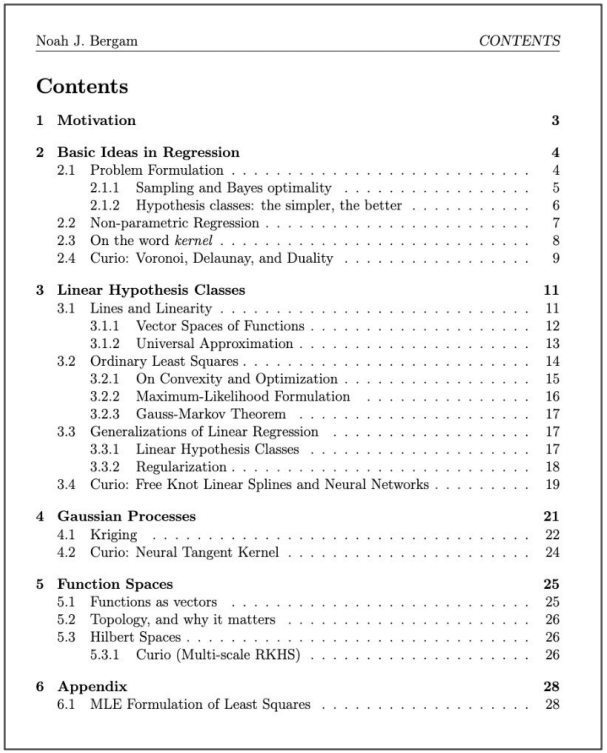
Conclusions

(1) A graph data structure can make for a much more flexible and efficient surface model.

(2) The problem of remote-sensing-based surface modeling motivates some interesting theoretical questions.

Acknowledgements: Abani Patra (advisor). Sami Choe, Mallory Kochanek, Ashley Jeon (group mates). Todd Quinto and Kasso Okoudjou (VERSEIM program coordinators). The National Science Foundation.

Thank you!



Noah J. Bergam *CONTENTS*

Contents

If you want to learn more, I am working on some lecture notes on mathematical modeling in glaciology, called *Regression on Ice.*