Bayesian Earthquake Modeling

Sarah Minson
USGS Seattle

Featuring contributions from…

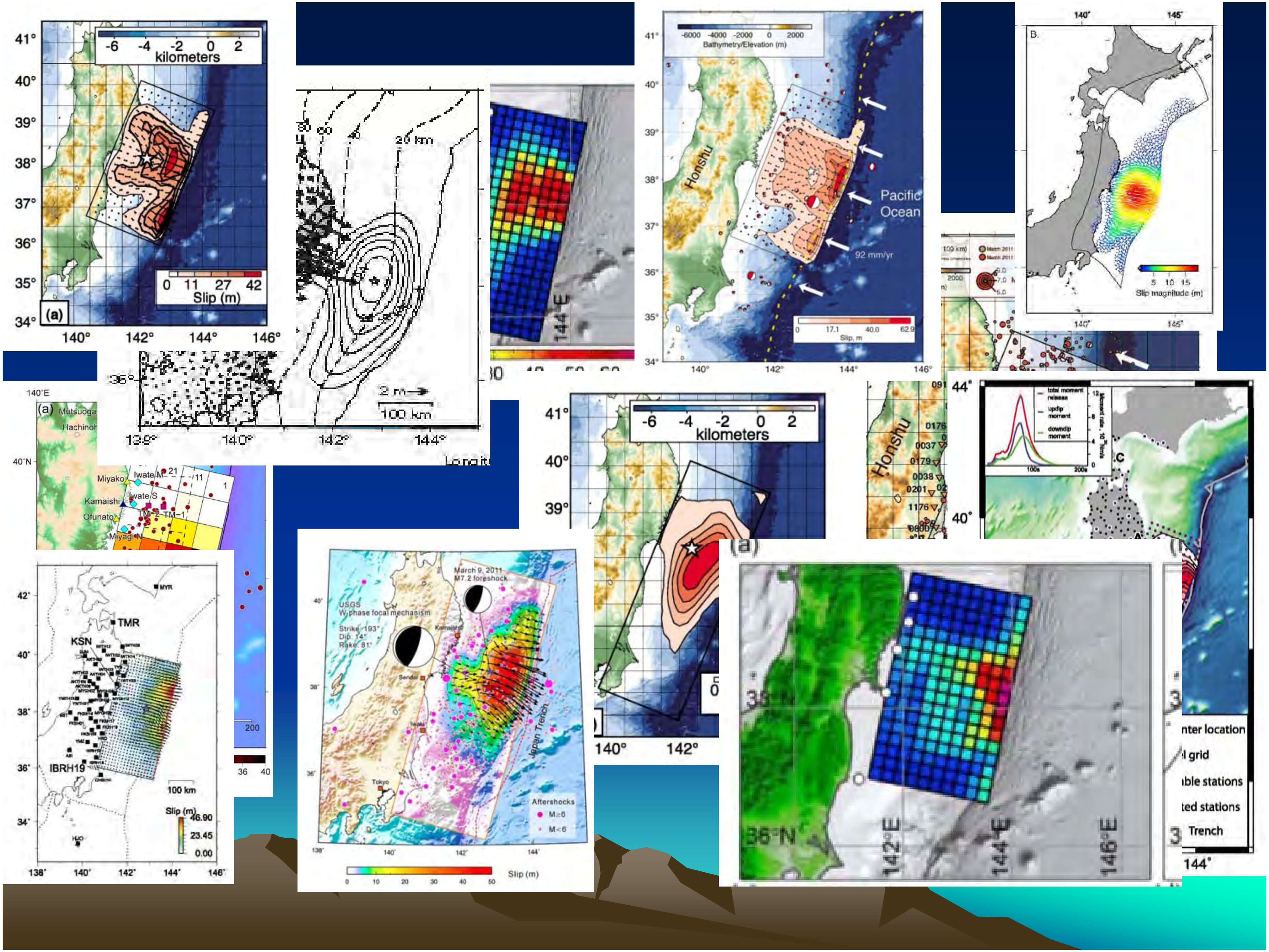
Mark Simons
James L. Beck
Junle Jiang
Francisco H. Ortega
Asaf Inbal
Susan E. Owen
Anthony Sladen

Overview

1. Some arguments for using Bayesian analysis
2. How to implement Bayesian methods
3. Applications:
   a) Bayesian source model for the great 2011 Tohoku-Oki earthquake
   b) Bayesian detection of slow slip events

Caveat

- This talk is mostly about earthquakes…
- …But the beauty of Bayesian methods is that the specifics of the model don't matter



kilometers
Slip (m)
(a)
20 km
2 m
100 km
Longitude
Bathymetry/Elevation (m)
Honshu
Pacific Ocean
92 mm/yr
Slip, m
Slip magnitude (m)
Mutsuoga
Hachinohe
Miyako
Kamaishi
Ofunato
Iwate M
Iwate S
Miyagi N
MYR
TMR
KSN
IBRH19
100 km
Slip (m)
March 9, 2011
M7.2 foreshock
USGS
W-phase focal mechanism
Strike: 193°
Dip: 14°
Rake: 81°
Sendai
Tokyo
Japan Trench
Aftershocks
M≥6
M<6
total moment release
updip moment
downdip moment
36°N
142°E
144°E
146°E
Trench

Geophysics is filled with under-determined inverse problems:

- What is plate locking near trench?
- We know moment or potency, but what is the length and width of rupture?
- We know surface rupture of a vertical strike-slip earthquake, but what happened at depth?
- …Plus many, many more

Uncertainty begets uncertainty

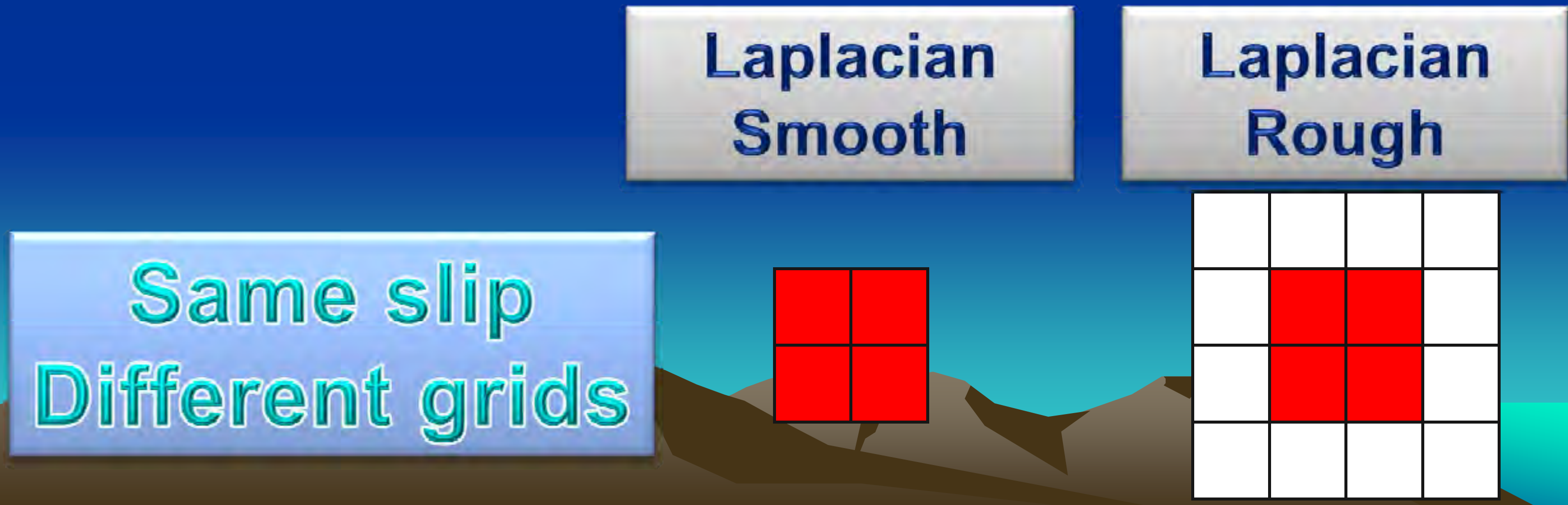
- E.g., uncertainty in spatial slip distribution yields uncertainty in inferred static stress changes
- E.g., uncertainty in STF and V_r yields uncertainty in rupture dynamics
- The problem isn't just that we can't constrain our models, but that we then try to draw conclusions from our non-unique models…
  - Seismic gaps, spatial relationship between coseismic/postseismic/interseismic slip

What to do?

- Hide uncertainty
- Embrace uncertainty

Hide uncertainty

- We do this a lot whether we mean to or not
- Regularized inverse picks ***one*** solution which fits data ***and*** regularization scheme
  - Choice of regularization is often based on **convenience** not **physics**
  - Regularization doesn't always mean what we think it means:



Embrace uncertainty

- When faced with a lack of model resolution, the best answer to the inverse problem is to identify the ensemble of all **plausible** models which **satisfy the data**
- "Plausible" describes what you believe model should be in the absence of data
  - These are your prior assumptions

Making explicit assumptions is a feature not a bug

Some possible prior assumptions for earthquakes

- Earthquakes are elastic
  - Limit on the maximum spatial gradient of slip
- Earthquakes rupture causally
- Known tectonic setting → known direction of slip
- Spatial relationship between coseismic/interseismic/postseismic slip, ETS, etc.

Making explicit assumptions is a feature not a bug

What should we do with under-determined inverse problems?

- Bayesian inference:
  1. Start with whatever prior assumptions you may have about your model (prior PDF)
  2. Gather some observations (data likelihood)
  3. Update the PDF describing possible models in light of these data (posterior PDF)
     - What is the posterior PDF?
- It's the distribution of all **plausible** models which **satisfy the data**

Bayes’ Theorem (1763)

- For inverse problems:

$$e^{-\frac{1}{2}[D-G(\theta)]^T \cdot C^{-1} \cdot [D-G(\theta)]}$$

Posterior

Prior

Data Likelihood

$$\mathrm{P}(\theta \mid D) \propto \mathrm{P}(\theta) \cdot \mathrm{P}(D \mid \theta)$$

$$D = \text{Data}$$

$$\theta = \text{Model}$$

$$C = \text{errors in data} + \text{errors in physics}$$

Inverse problems without inversion

$$\mathrm{P}(\theta \mid D) \propto \mathrm{P}(\theta) \cdot e^{-\frac{1}{2}[D-G(\theta)]^T \cdot C^{-1} \cdot [D-G(\theta)]}$$

- Only the forward problem is evaluated!
- Notice that the only matrix inverse is C^{-1}
  - Covariance matrices are symmetric positive definite
    - Inverse exists
  - Regularization is not required even for ill-conditioned problems

Why choose Bayesian analysis?

Optimization	**Bayesian**
One solution	Distribution of solutions
Converges to one minimum	Multi-peaked solution spaces OK
Regularized (lots of decisions)	No *a priori* regularization required
Limited choice of *a priori* constraints	Generalized *a priori* constraints
Error analysis hard for nonlinear problems	Error analysis comes free with solution
Sensitive to model parameterization (model covariance leads to trade-offs)	Insensitive to model parameterization (if model covariance is estimated)

Q. Bayesian analysis is the bee's knees, but how do I do it?

- A. Two options:
  1. If you choose your priors wisely, you can get an analytical form of $P(\theta|D)$
     - One way to do this is with a conjugate prior: e.g.,
       - $P(D|\theta)$ Gaussian & $P(\theta)$ Gaussian $\rightarrow$ $P(\theta|D)$ Gaussian
  2. You can choose whatever the heck priors and data likelihood you want and draw samples of $P(\theta|D)$ by Monte Carlo simulation
- Option 1 requires brain power and possible compromise. Option 2 requires a whole lotta computer power.

Curse of Dimensionality

- Monte Carlo simulation of PDFs requires drawing enough samples to fill the model space
  - Huge numbers of samples required for high-dimensional problems
    - One sample = One forward model evaluation
- The total numerical cost is:
  - Number of samples X Time to compute forward model
    - If you have a large number of model parameters, you need a very fast forward model
    - If you have a slow forward model, you must have a low-dimensional model

A few samplers you could use

- Rejection method: parallel but inefficient
- Metropolis algorithm: more efficient, but MCMC (i.e., a random walk) is serial
- Various tempering, annealing, and transitioning algorithms
- Build your own
  - CATMIP (efficient and parallel)

More on Monte Carlo simulation

- All algorithms for Monte Carlo simulation of an unknown target PDF are similar:
  1. Generate a candidate sample ($\theta_{proposed}$) from a proposal PDF (usually Gaussian)
  2. Decide whether to keep that sample
     - e.g., for the Metropolis algorithm (random walk):
     - Let $r = P(\theta_{proposed}|D) / P(\theta_{current}|D)$
       – If r > 1, accept candidate sample
       – Otherwise draw u~U(0,1), a random sample from the uniform distribution between 0 and 1
         - If r > u, accept candidate sample
         - Otherwise reject candidate sample

Cascading Adaptive Transitional Metropolis In Parallel: CATMIP

- Transitioning AKA Tempering AKA Simulated Annealing*

$$P(\theta \mid D) \propto P(\theta) \cdot P(D \mid \theta)^{\beta} \qquad 0 \le \beta \le 1$$

  - Dynamic cooling schedule**
- Parallel Metropolis
- Simulation adapts to model covariance**
- Simulation adapts to rejection rate***
- Resampling**
- Cascading

Static | Kinematic

$$P(\theta \mid D) \propto P(\theta_s) \cdot P(D_s \mid \theta_s)^{\beta} \cdot P(\theta_k) \cdot P(D_k \mid \theta_s, \theta_k)^{\alpha}$$

$$0 \le \beta \le 1, \qquad \alpha = 0$$

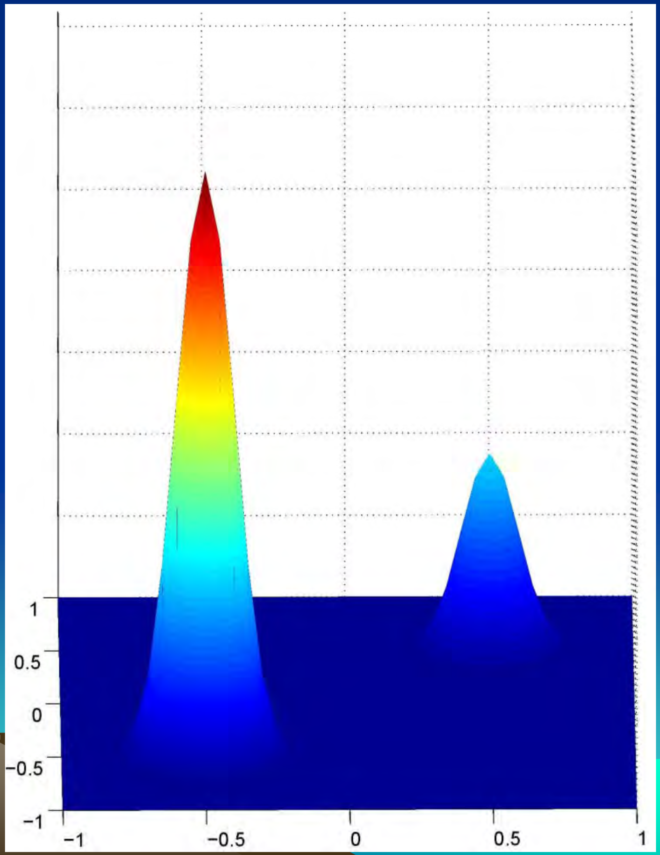
$$\beta = 1, \qquad 0 \le \alpha \le 1$$

* Marinari and Parisi (1992)
** Ching and Chen (2007)
*** Matt Muto

Example: Mixture of Gaussians

- Target distribution:

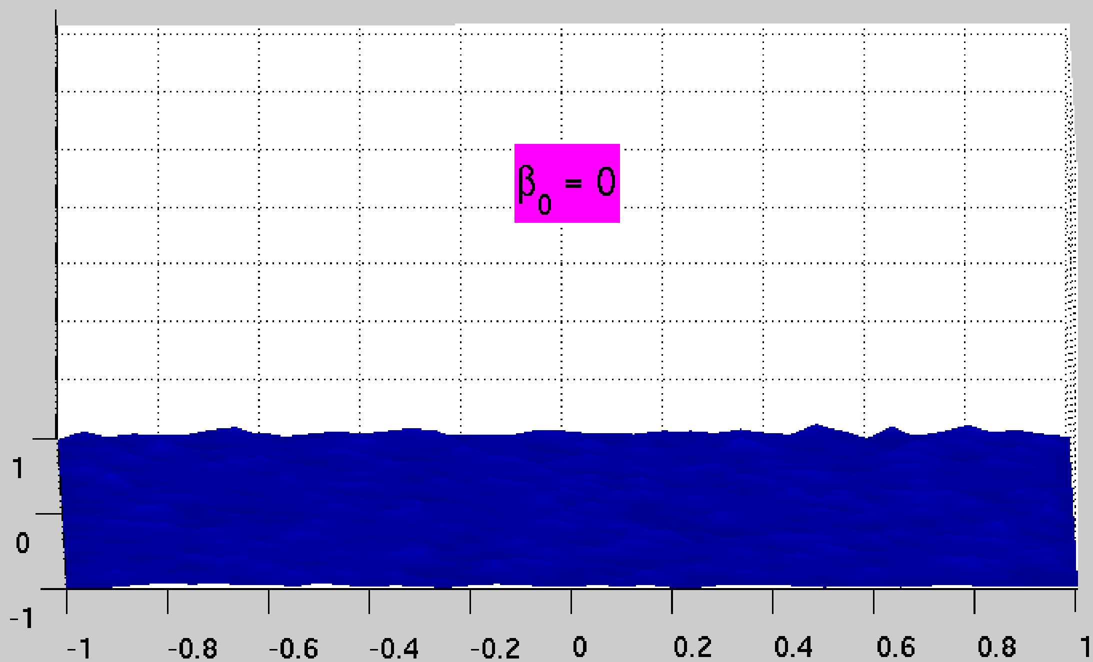
$$N(\mu_1, \sigma_1^2) + 3 \cdot N(\mu_2, \sigma_2^2)$$



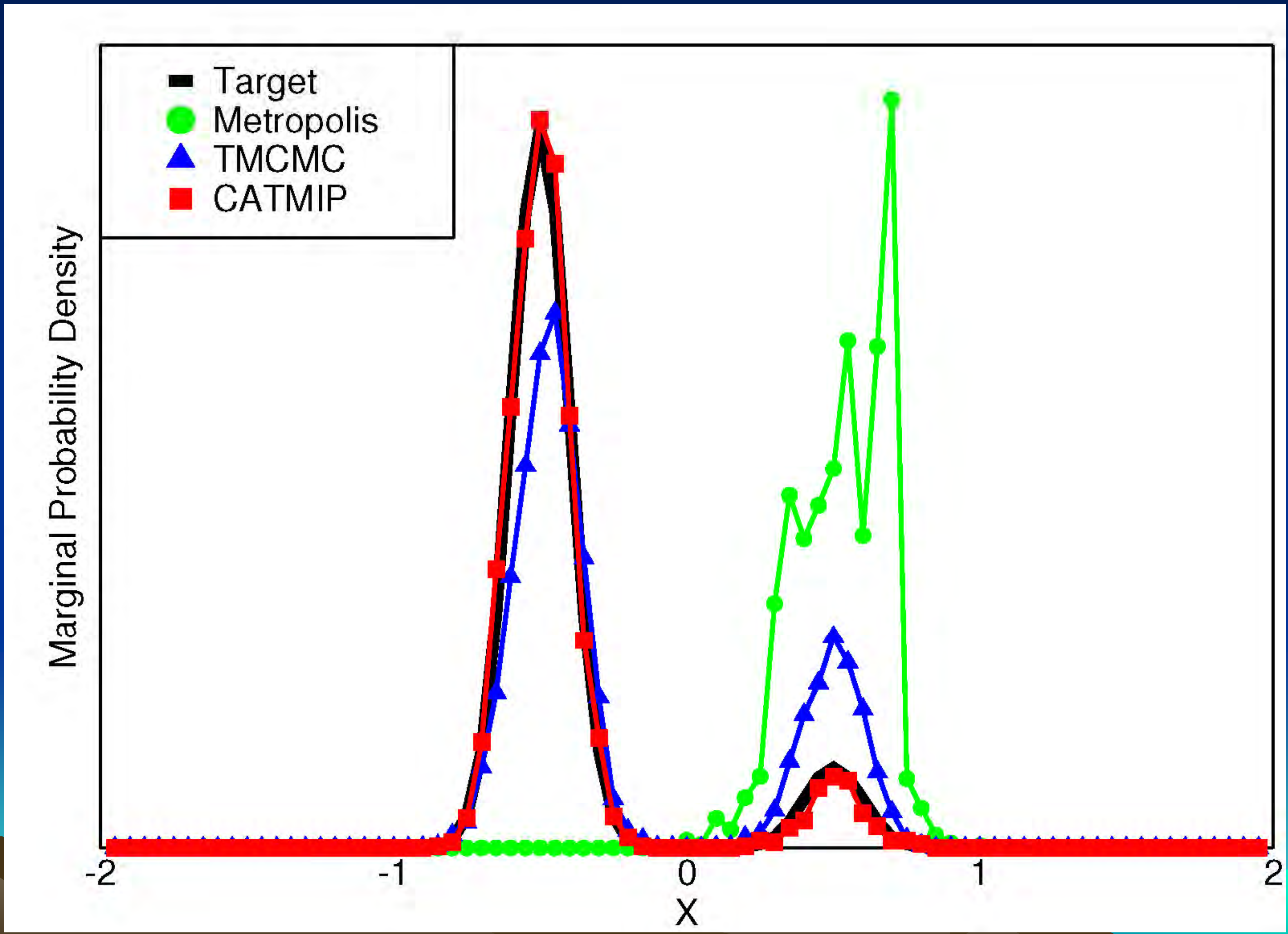
CATMIP

1. Sample P(θ)
2. Calculate β
3. Resample
4. Metropolis algorithm in parallel
5. Collect final samples
6. Go back to Step 2, lather, rinse, and repeat until cooling is achieved

CATMIP



Sampler performance comparison



Step-by-step example 1

TOHOKU-OKI RUPTURE MODEL: THE HIT IT WITH A BIG NUMERICAL HAMMER APPROACH

$$\mathrm{P}(\theta \mid D) \propto \mathrm{P}(\theta) \cdot \mathrm{P}(D \mid \theta)$$

- The steps:
  1. Design a forward model to stick into data likelihood, P(D|θ)
  2. Choose prior assumptions to define prior PDF, P(θ)
  3. Use Monte Carlo sampling to generate lots of random models and eventually fill in posterior PDF, P(θ|D)

Forward model

- Define a finite fault mesh, hypocenter and elastic structure
  - For each patch, we solve for four parameters:
    1. Strike-slip motion
    2. Dip-slip motion
    3. Slip duration of triangular STF
    4. Rupture velocity
  - For a mesh with N patches, θ is a vector with 4*N elements
- For GPS offsets, seafloor geodesy, and tsunami waveforms, data predictions are:

$$\begin{bmatrix} [\mathbf{G}_{ss}] \\ [\mathbf{G}_{ds}] \end{bmatrix} \cdot \begin{bmatrix} [\mathbf{u}_{ss}] \\ [\mathbf{u}_{ds}] \end{bmatrix} = \begin{bmatrix} [\mathbf{G}_{ss}] \\ [\mathbf{G}_{ds}] \end{bmatrix} \cdot \begin{bmatrix} u^{1}{}_{ss} \\ \vdots \\ u^{N}{}_{ss} \\ u^{1}{}_{ds} \\ \vdots \\ u^{N}{}_{ds} \end{bmatrix}$$

- For 1-Hz kinematic GPS, data predictions are synthetic seismograms calculated by:
  1. Scaling GFs for each patch by slip on that patch
  2. Convolving with the appropriate STF
  3. Time-shifting each waveform according to rupture velocity field
  4. Summing all of the resulting per-patch synthetics

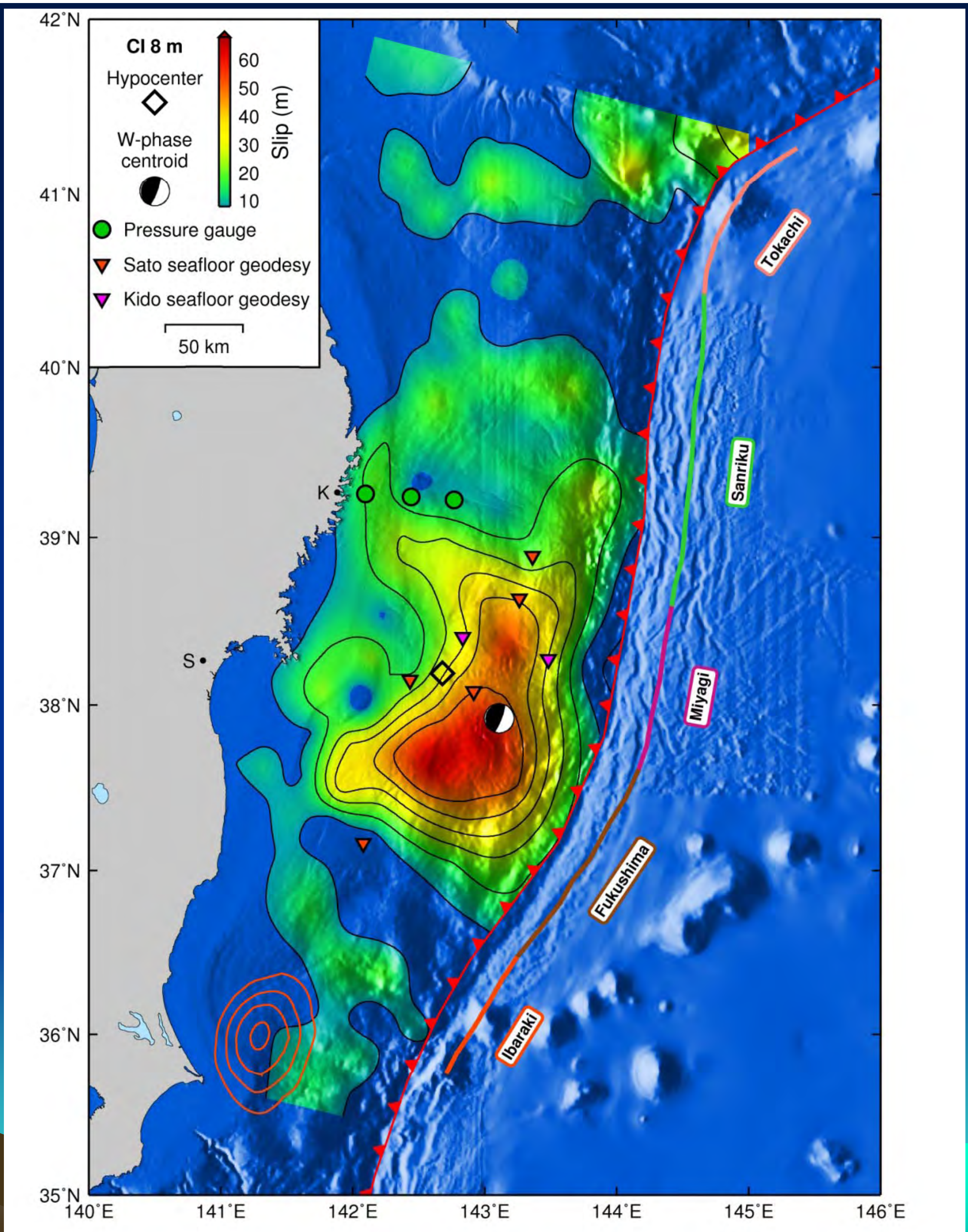
Prior PDF

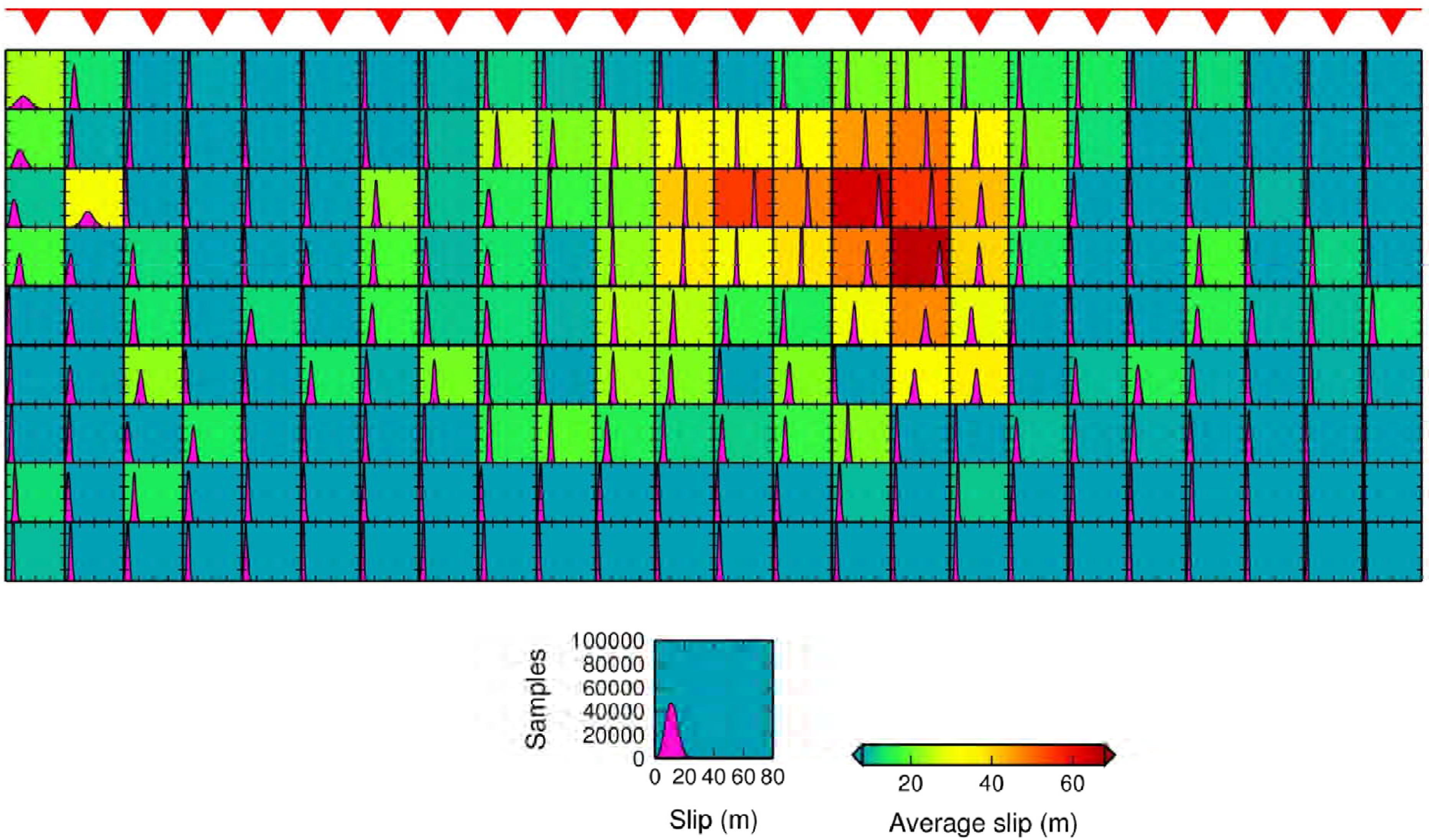
- Broad priors
  - What does the data resolve and what isn't resolved?
- Strike-slip prior is zero-mean Gaussian
  - On average, slip is thrust
- Dip-slip prior is uniform ~ $U(u_{min}<0,\infty)$
  - All dip-slip motion is equally likely except large amounts of normal motion is forbidden
- Priors on rise time and rupture velocity are uniform and broad

Sample

- For i=1…Gazillion
  1. Draw a random source model, θ
  2. Assign a prior probability to that model, P(θ)
  3. Run your forward model, G(θ)
  4. Evaluate P(D| θ)
  5. Evaluate the unnormalized posterior probability, P(θ|D)=P(D|θ)P(θ)

Average posterior slip

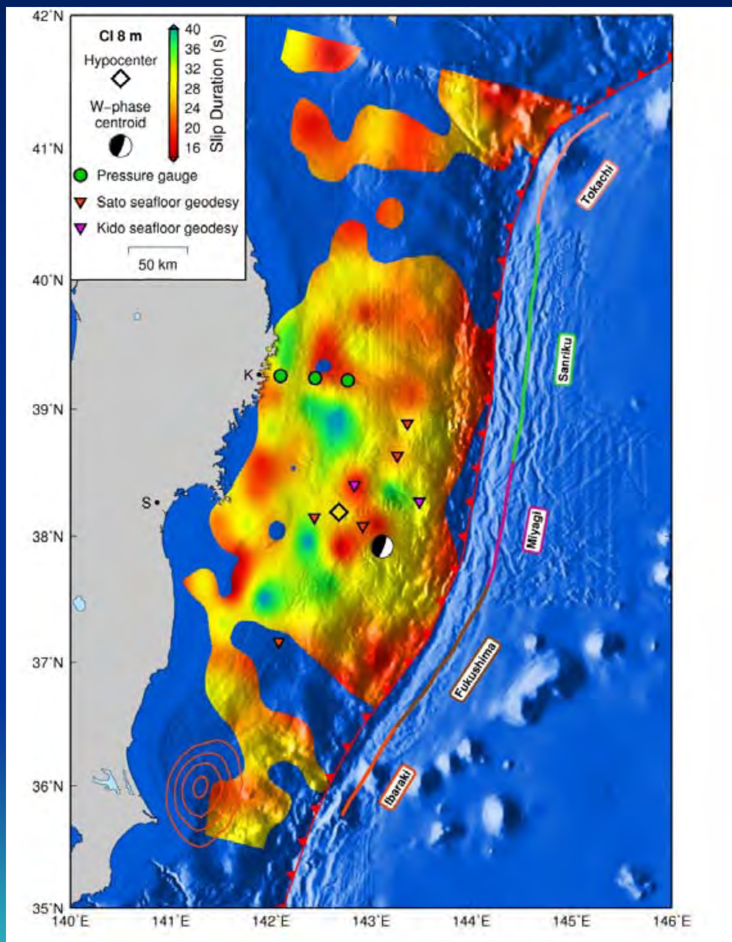


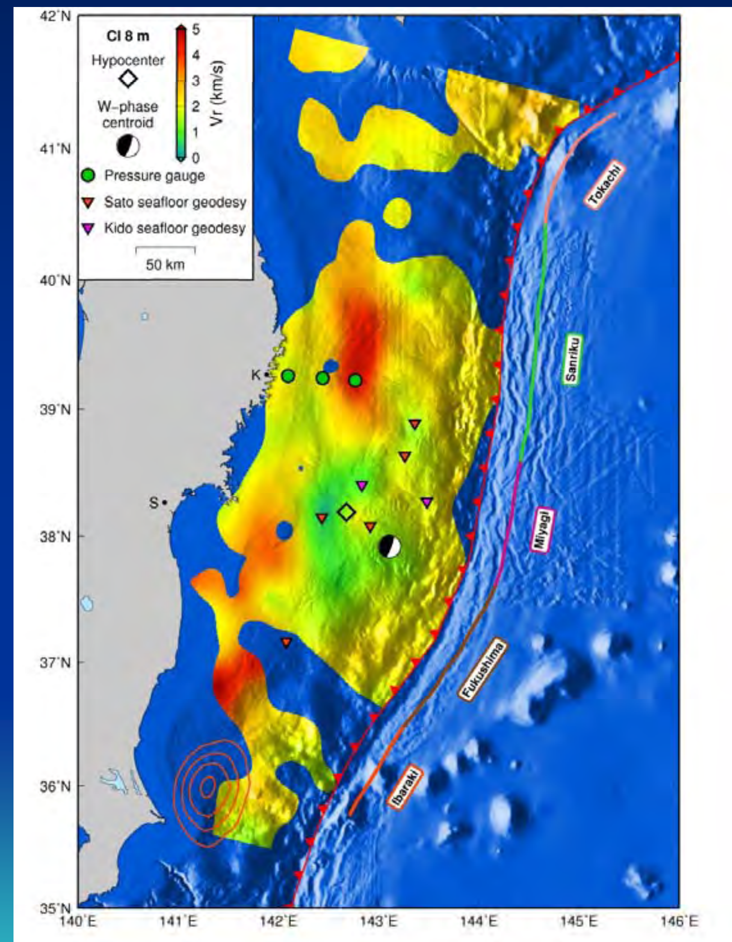


Samples
100000
80000
60000
40000
20000
0
0 20 40 60 80
Slip (m)
20 40 60
Average slip (m)

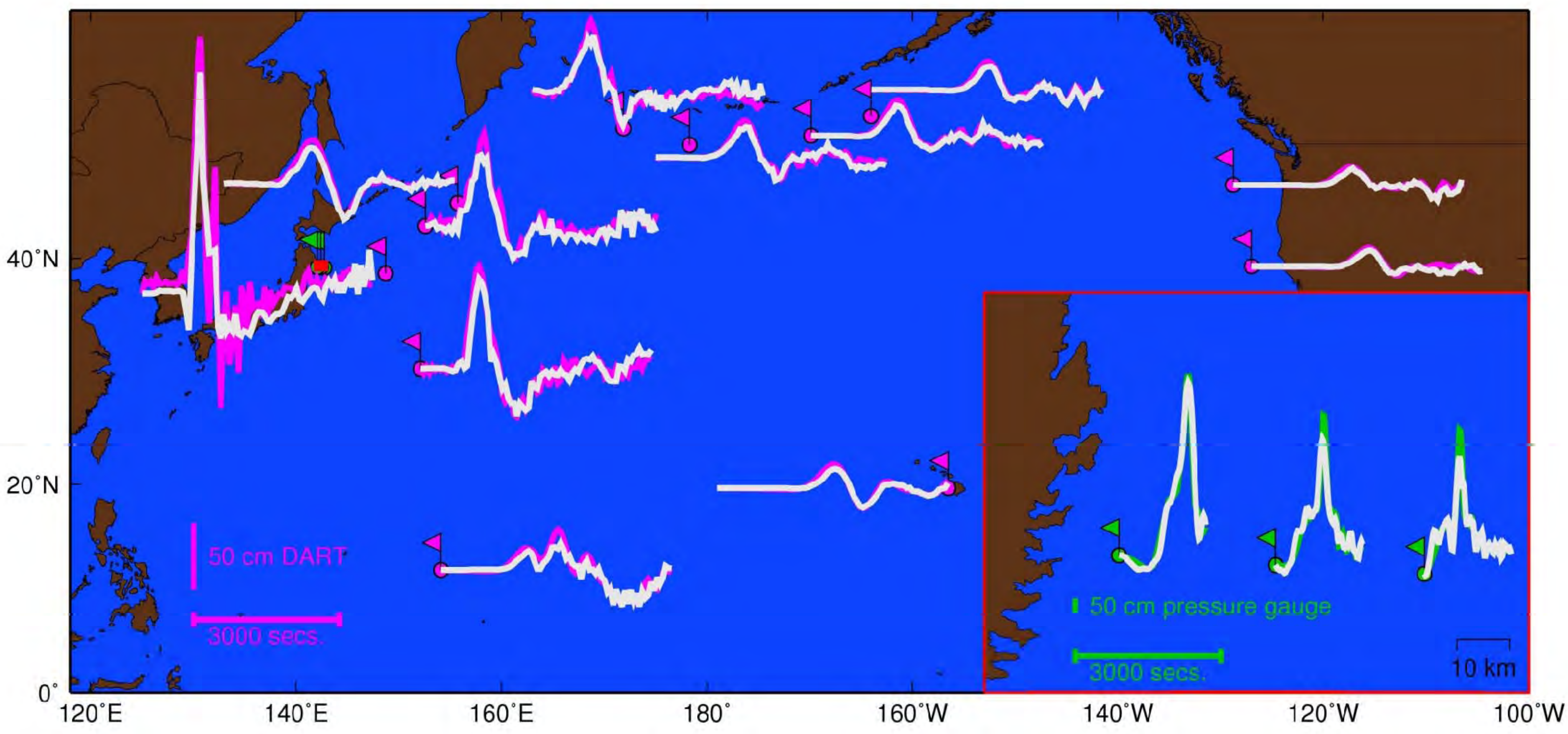
Posterior mean (one statistic of many)

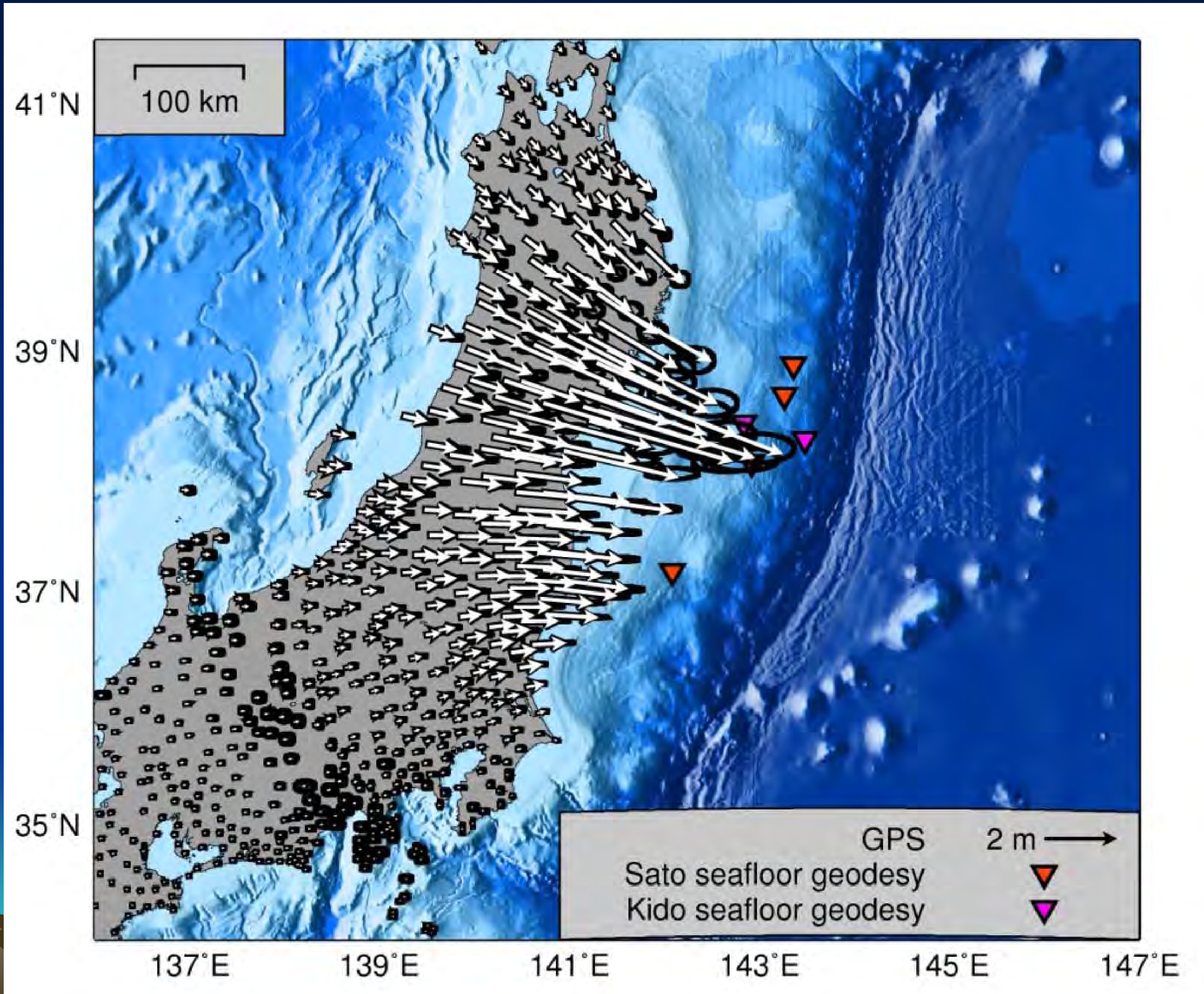
Slip duration *Rupture velocity*



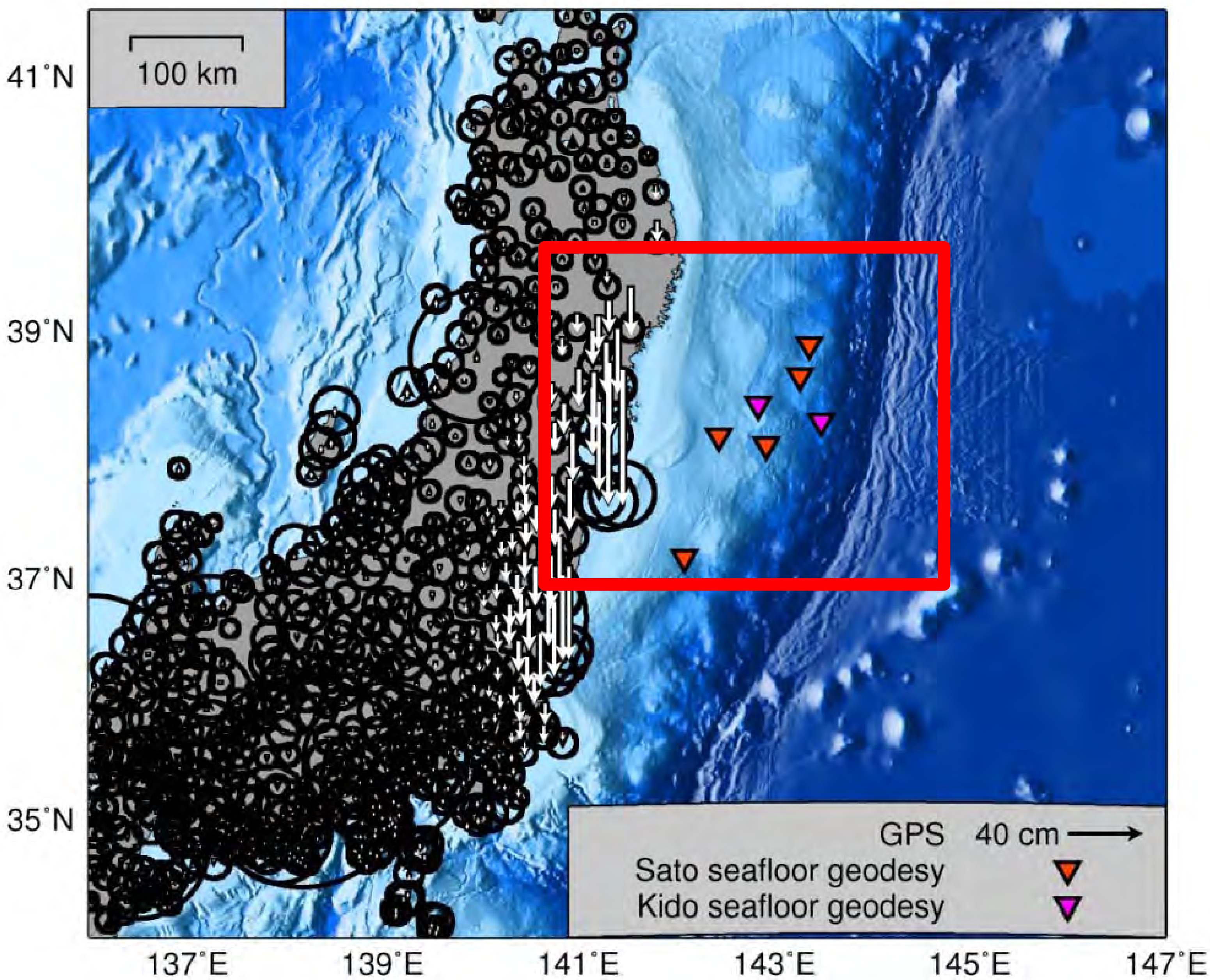


Tsunami data

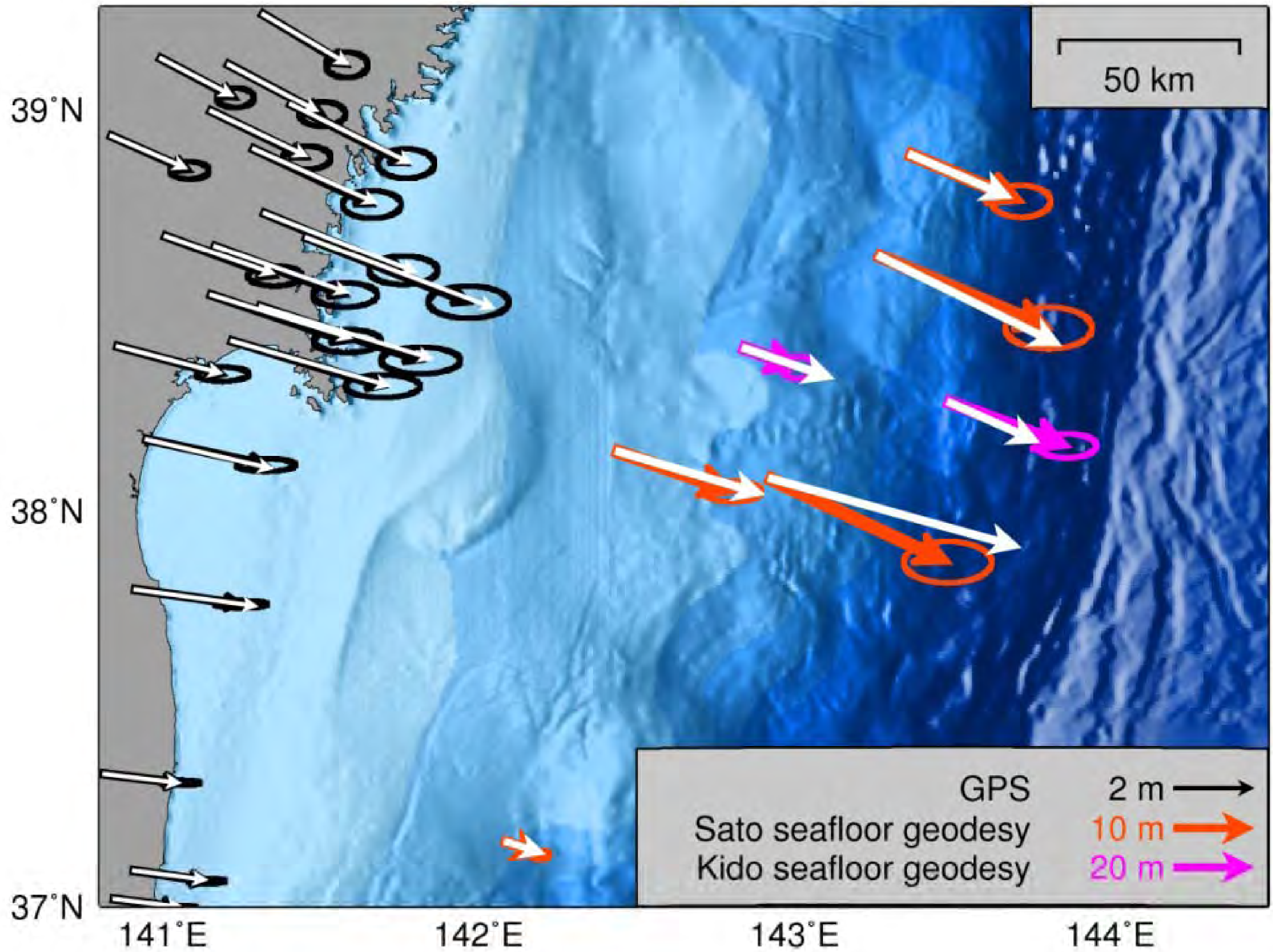




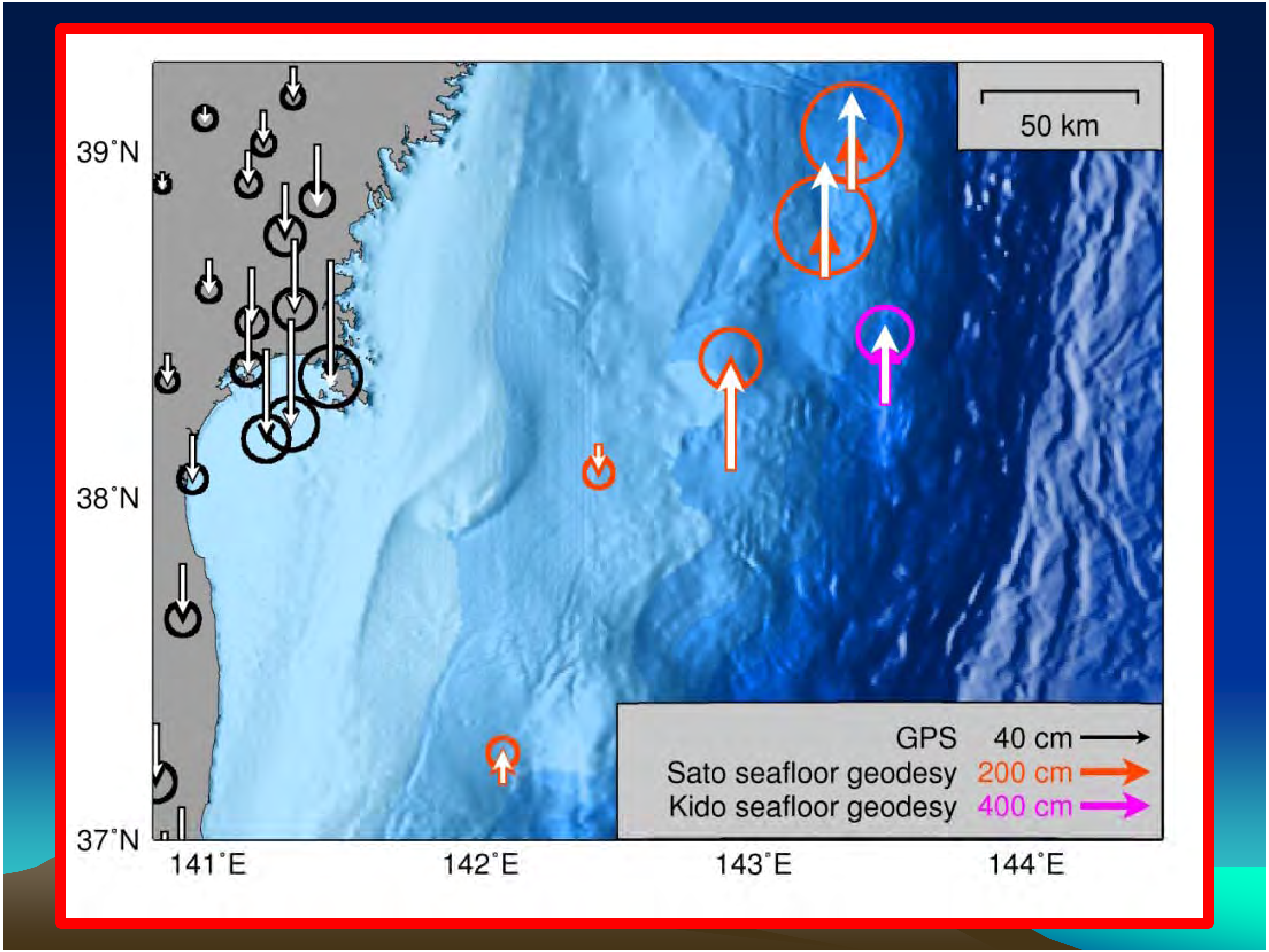
100 km
41°N
39°N
37°N
35°N
137°E
139°E
141°E
143°E
145°E
147°E
GPS
2 m
Sato seafloor geodesy
Kido seafloor geodesy



100 km
41˚N
39˚N
37˚N
35˚N
137˚E
139˚E
141˚E
143˚E
145˚E
147˚E
GPS 40 cm
Sato seafloor geodesy
Kido seafloor geodesy



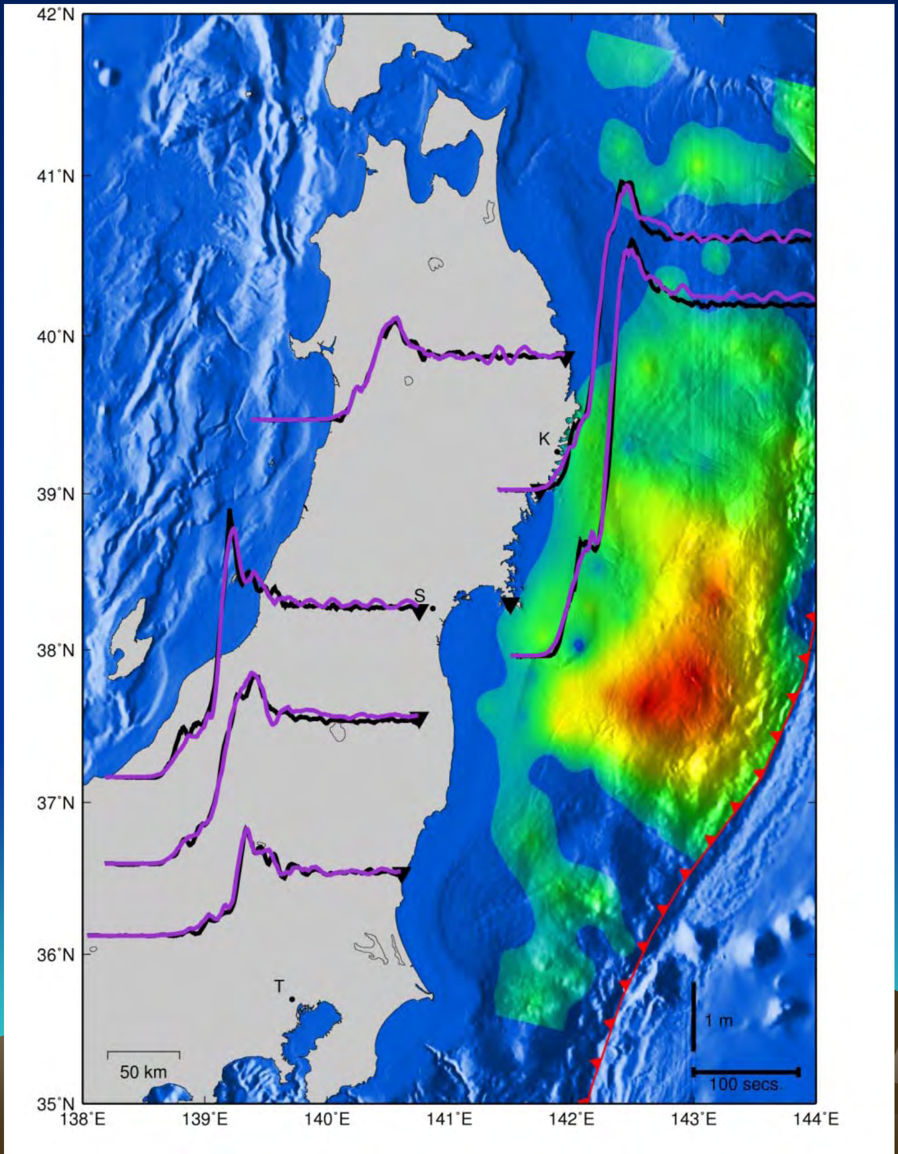
50 km
39˚N
38˚N
37˚N
141˚E
142˚E
143˚E
144˚E
GPS 2 m
Sato seafloor geodesy 10 m
Kido seafloor geodesy 20 m

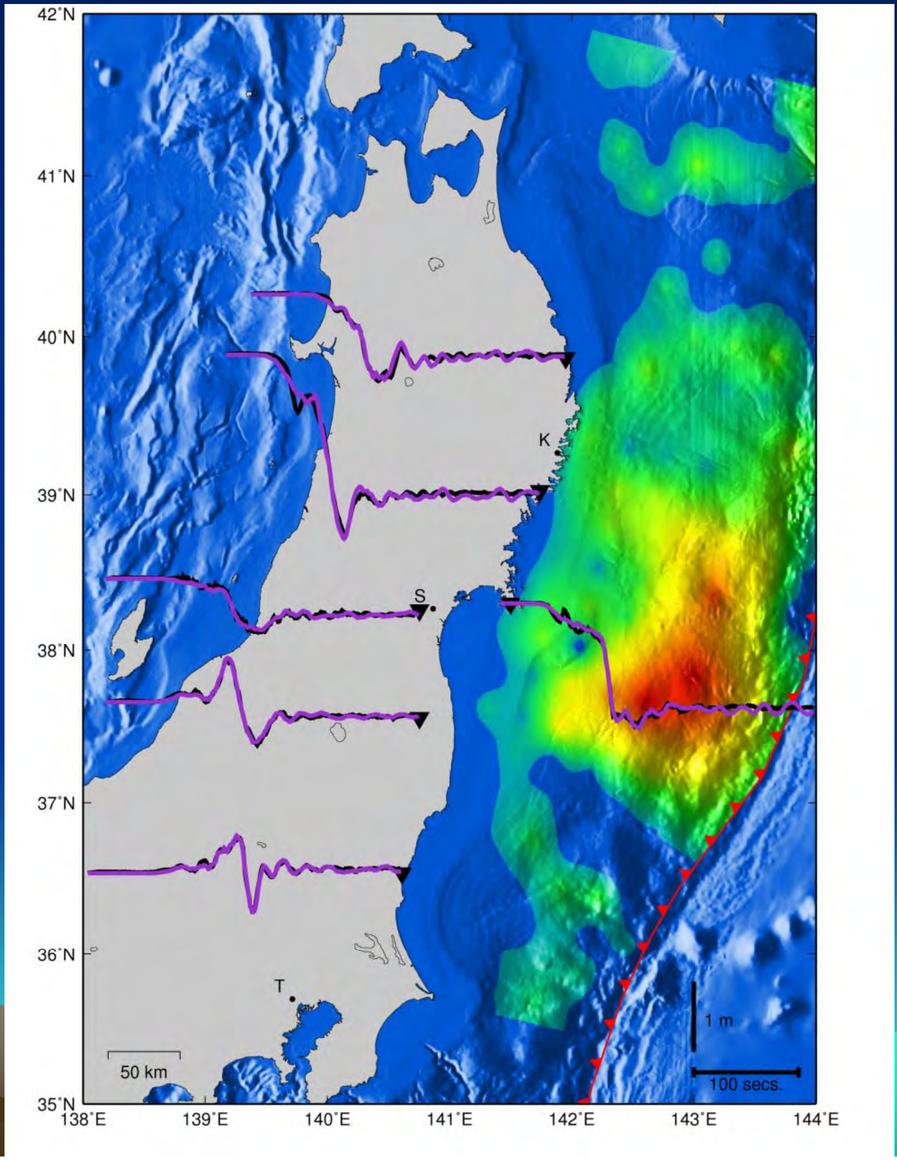


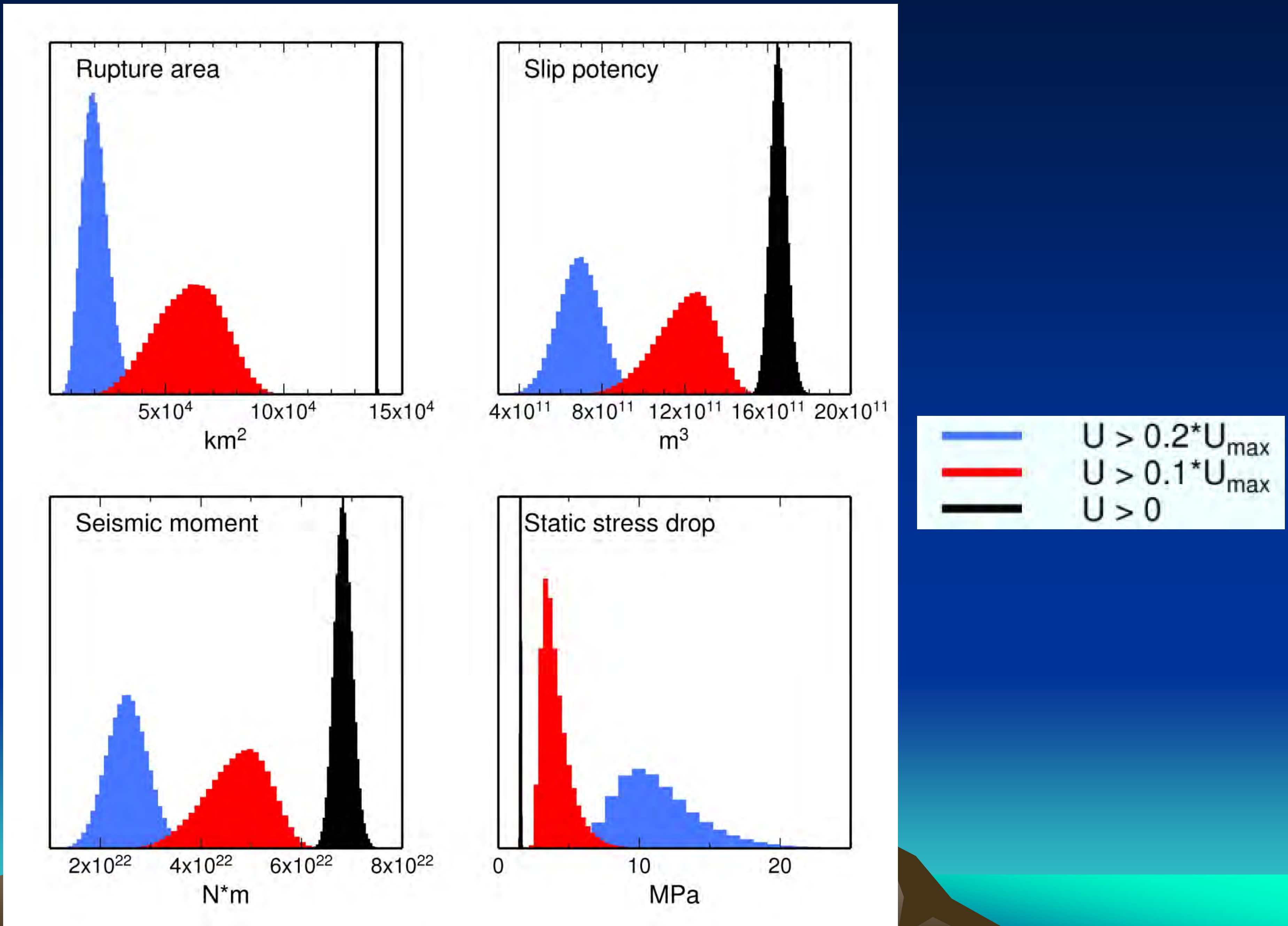
39˚N
38˚N
37˚N
141˚E
142˚E
143˚E
144˚E
50 km
GPS 40 cm
Sato seafloor geodesy 200 cm
Kido seafloor geodesy 400 cm

1 Hz GPS

East *North*

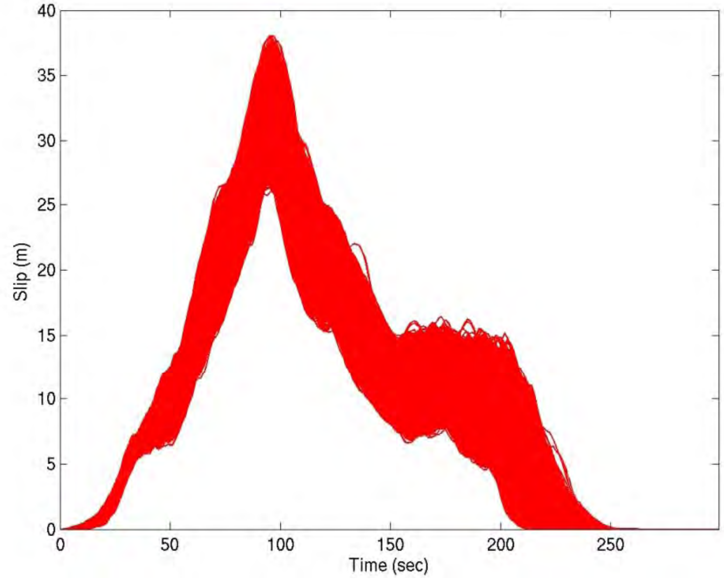


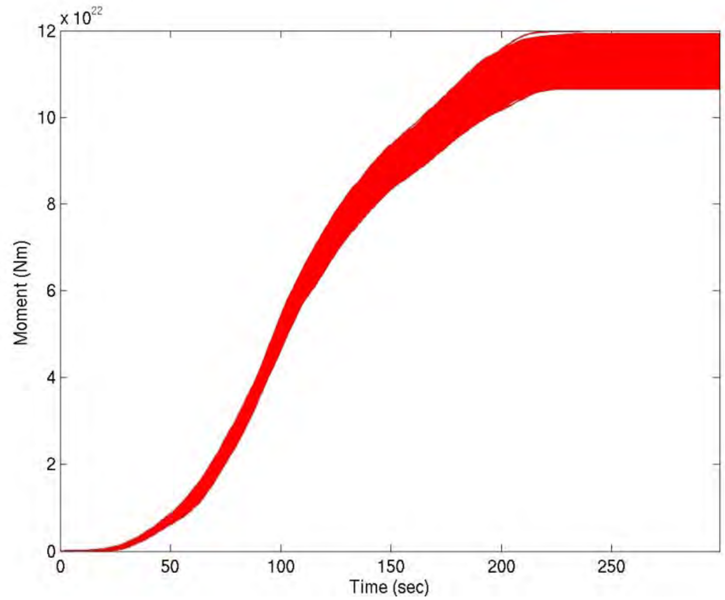


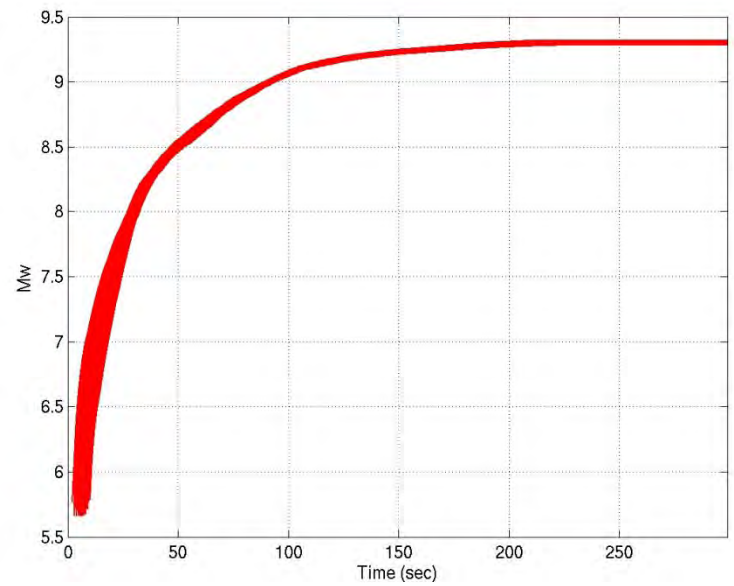


Rupture area
5x10^4
10x10^4
15x10^4
km^2
Slip potency
4x10^11
8x10^11
12x10^11
16x10^11
20x10^11
m^3
Seismic moment
2x10^22
4x10^22
6x10^22
8x10^22
N*m
Static stress drop
0
10
20
MPa
U > 0.2*U_max
U > 0.1*U_max
U > 0

All models



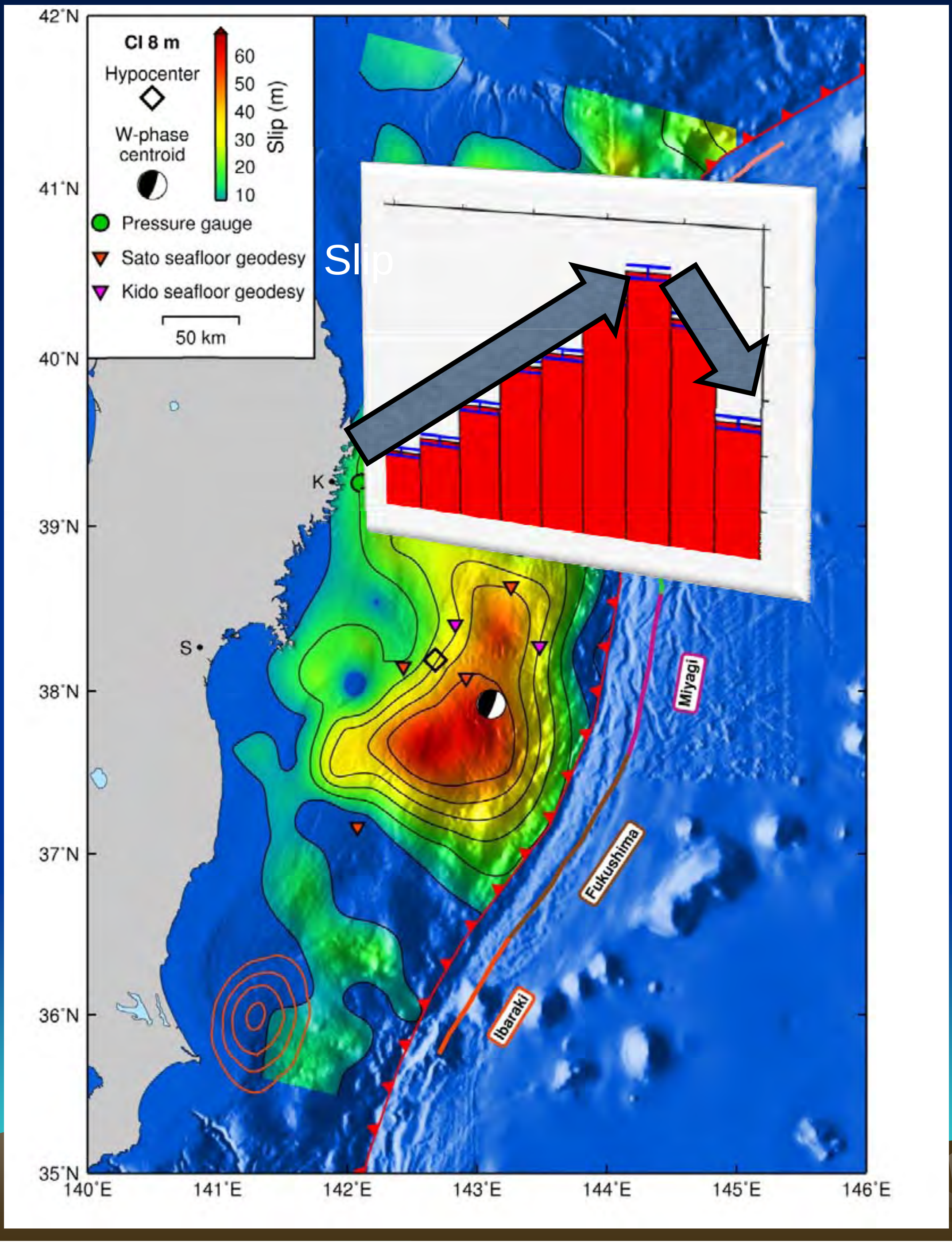


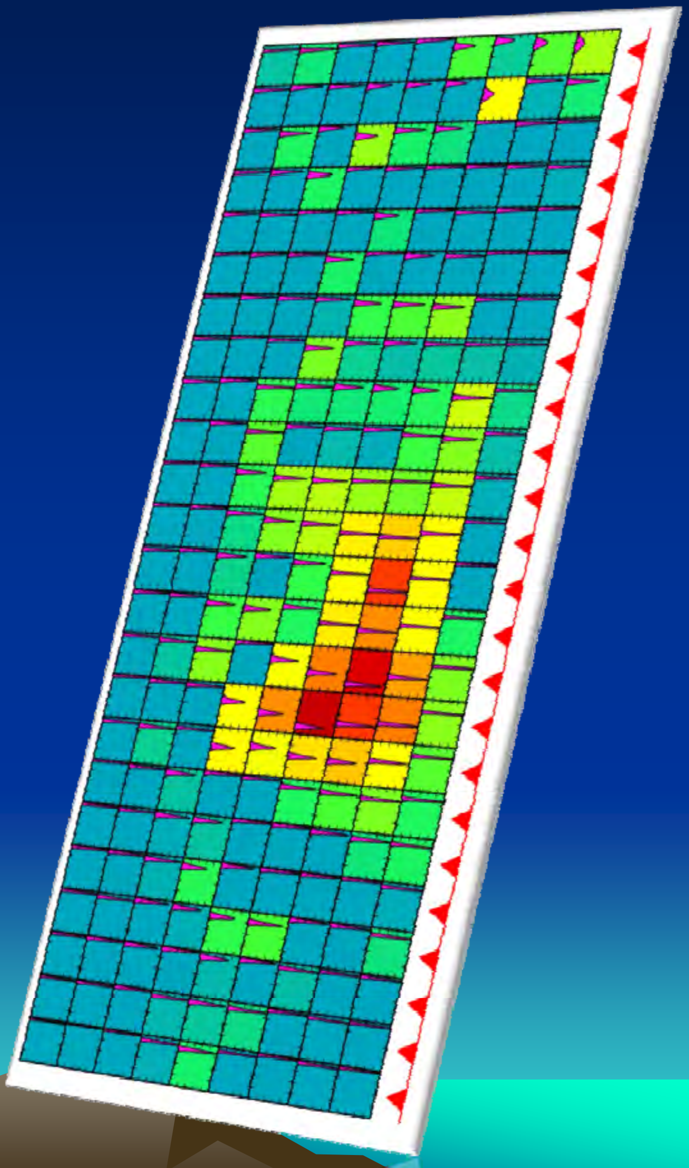


Slip is less well constrained as rupture evolves

But total moment magnitude is well constrained

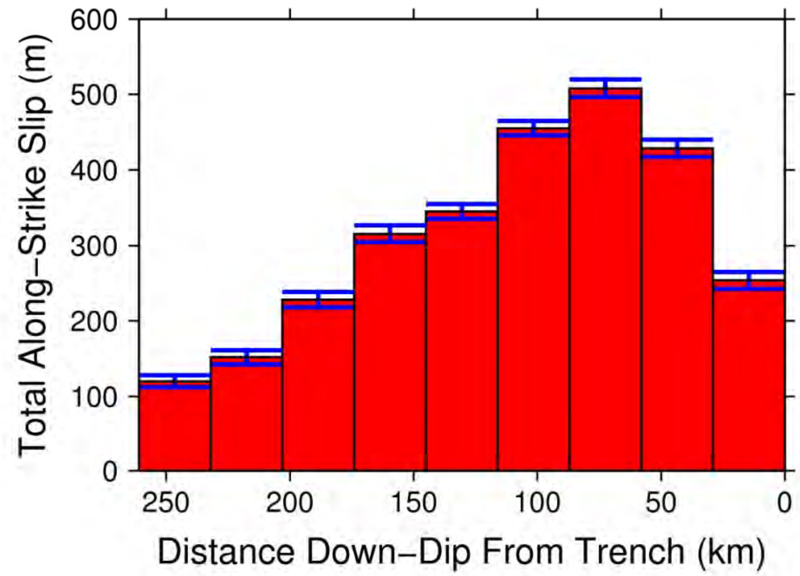
Average slip

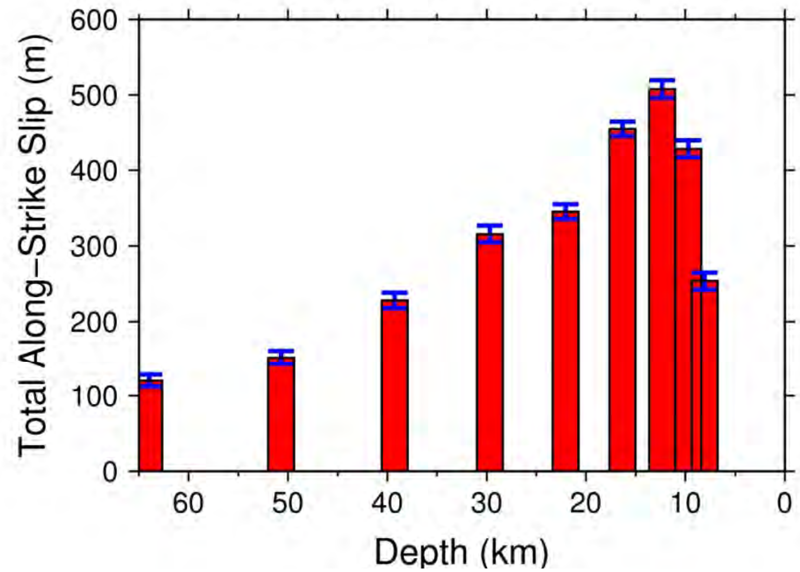




Along-strike integrated slip

- Significant near-trench fault slip
  - But, on average, peak slip is not at the trench
- Decrease in slip amplitude near trench is recoverable because the model is not regularized
- Localized zones of high slip may exist near trench





Conclusions

- Fully Bayesian kinematic rupture model for the Tohoku-Oki earthquake
  - Significant slip near trench, but peak large-scale slip feature is not at trench
  - Uncertainty in slip model is essential to understanding the subduction zone

There's more you can do with Bayes' theorem

- Error updating:

$$P(D\,|\,\theta) = e^{-\frac{1}{2}[D-G(\theta)]^T \cdot C^{-1} \cdot [D-G(\theta)]}$$

$$C = \text{errors in data} + \text{errors in physics}$$

$$D = G(\theta) + \varepsilon_D + \varepsilon_{G(\theta)}$$

- We don't know what $\varepsilon_{G(\theta)}$ is, but we can sample for $\varepsilon_{G(\theta)}$ just like any other variable

There's more you can do with Bayes' theorem

- Model class selection: which model design fits the data best?

$$P(\theta \mid D) = \frac{P(D \mid \theta) P(\theta)}{\int P(D \mid \theta) P(\theta)} = \frac{P(D \mid \theta) P(\theta)}{P(D \mid \mathcal{M})}$$

- Denominator is the evidence in favor of model, $\mathcal{M}$

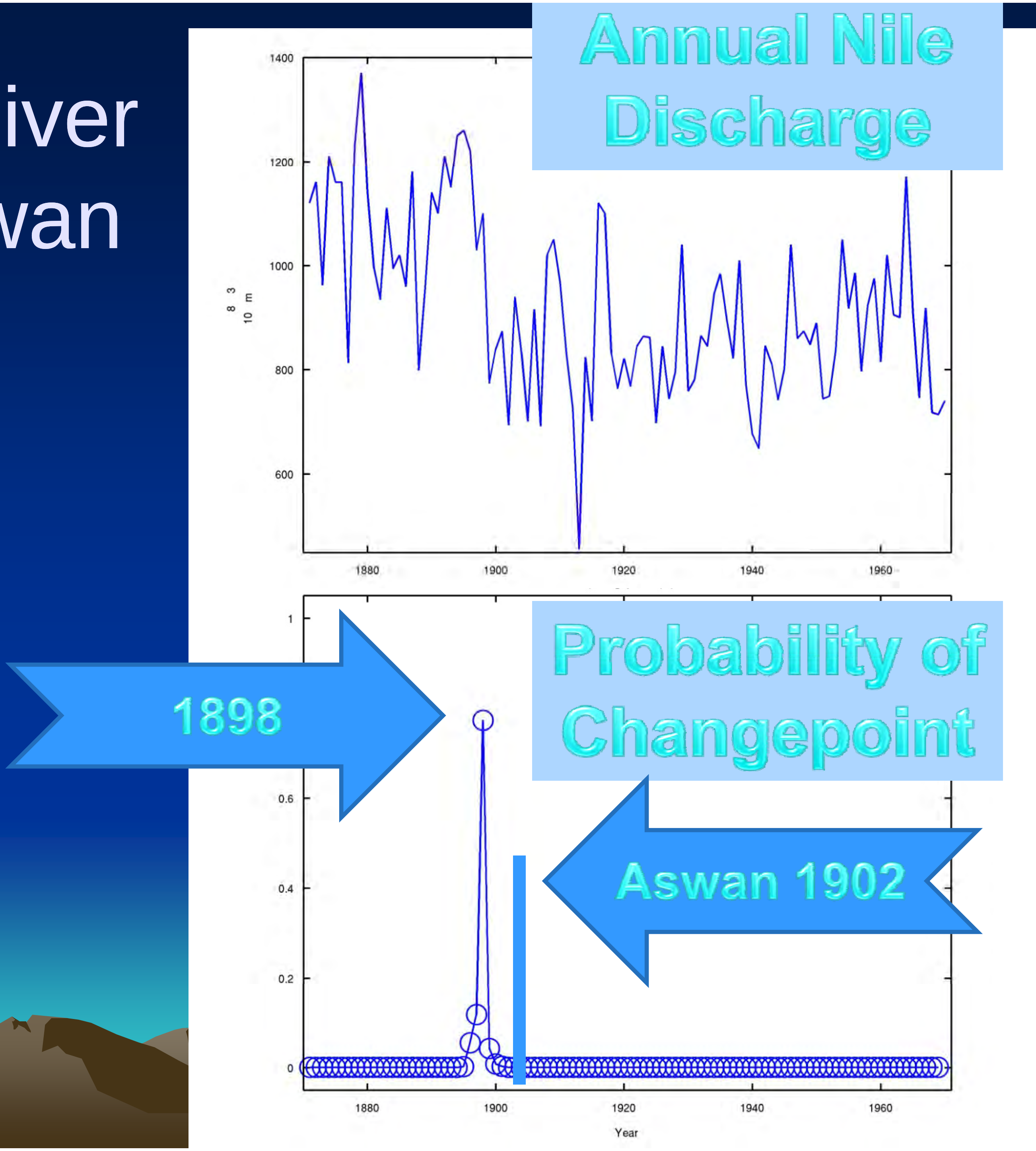
Step-by-step example 2

CASCADIA SLOW-SLIP EVENTS: THE RIDICULOUSLY CHEAP ANALYTICAL SOLUTION APPROACH

Bayesian changepoint detection

- Changepoint: the time that at least one model parameter changes
- Can use Bayes' theorem to compute the probability of a changepoint as a function of time, P(changepoint=t|D)
- Can also use Bayes' theorem to assess significance of potential changepoints

The Nile River & The Aswan Dam



Bayesian linear regression

- In least squares, we solve D=G*m to get $m \pm C_m$
- This is equivalent to saying that our model is $\beta \sim \mathcal{N}(m, C_m)$, with m and C_m unknown
  - With the right choice of priors P(m) and P(C_m), we can stick this into Bayes' theorem and get an analytical solution

Bayesian changepoint detection = Bayesian piecewise linear regression

- Model: $$\begin{cases} \mathbf{G}_1(x)\cdot\boldsymbol{\beta}_1 & x < t \\ \mathbf{G}_2(x)\cdot\boldsymbol{\beta}_2 & x \geq t \end{cases}$$

- Goal: $P(\theta = t \mid D, \mathcal{M})$ as a function of t

Analytical solution with conjugate prior

$$P(\theta \mid \mathbf{D}) = \frac{P(\mathbf{D} \mid \theta)\, P(\theta)}{P(\mathbf{D} \mid t, \mathcal{M})} = \frac{P(\mathbf{D} \mid \boldsymbol{\beta}, \sigma^2)\, P(\boldsymbol{\beta}, \sigma^2)}{P(\mathbf{D} \mid t, \mathcal{M})} = \frac{P(\mathbf{D} \mid \boldsymbol{\beta}, \sigma^2)\, P(\boldsymbol{\beta} \mid \sigma^2)\, P(\sigma^2)}{P(\mathbf{D} \mid t, \mathcal{M})}$$

$$P(\mathbf{D} \mid \theta) = P(\mathbf{D} \mid \boldsymbol{\beta}, \sigma^2) = \mathcal{N}(\mathbf{D} - \mathbf{G}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$$

$$P(\boldsymbol{\beta} \mid \sigma^2) = \mathcal{N}(\mathbf{m}, \sigma^2 \mathbf{V})$$

$$P(\sigma^2) = \mathcal{IG}(a, b)$$

$$P(\theta \mid \mathbf{D}) \propto P(\mathbf{D} \mid \theta)\, P(\boldsymbol{\beta} \mid \sigma^2)\, P(\sigma^2) = \mathcal{NIG}(\mathbf{m}, \mathbf{V}, a, b)$$

$$\rightarrow \quad P(\theta \mid \mathbf{D}) = \mathcal{NIG}(\mathbf{m}^*, \mathbf{V}^*, a^*, b^*)$$

$$P(\mathbf{D} \mid t, \mathcal{M}) = \frac{P(\mathbf{D} \mid \theta)\, P(\theta)}{P(\theta \mid \mathbf{D})} = \frac{1}{(2\pi)^{n/2}} \frac{|\mathbf{V}^*|^{1/2}\, (b)^a\, \Gamma(a^*)}{|\mathbf{V}|^{1/2}\, (b^*)^{a^*}\, \Gamma(a)}$$

$$\mathbf{V}^* = \left(\mathbf{V}^{-1} + \mathbf{G}^T \mathbf{G}\right)^{-1}$$

$$a^* = a + \frac{n}{2}$$

$$b^* = b + \frac{1}{2}\left\{\mathbf{m}^T \mathbf{V}^{-1} \mathbf{m} + \mathbf{D}^T \mathbf{D} - (\mathbf{m}^*)^T (\mathbf{V}^*)^{-1} (\mathbf{m}^*)\right\}$$

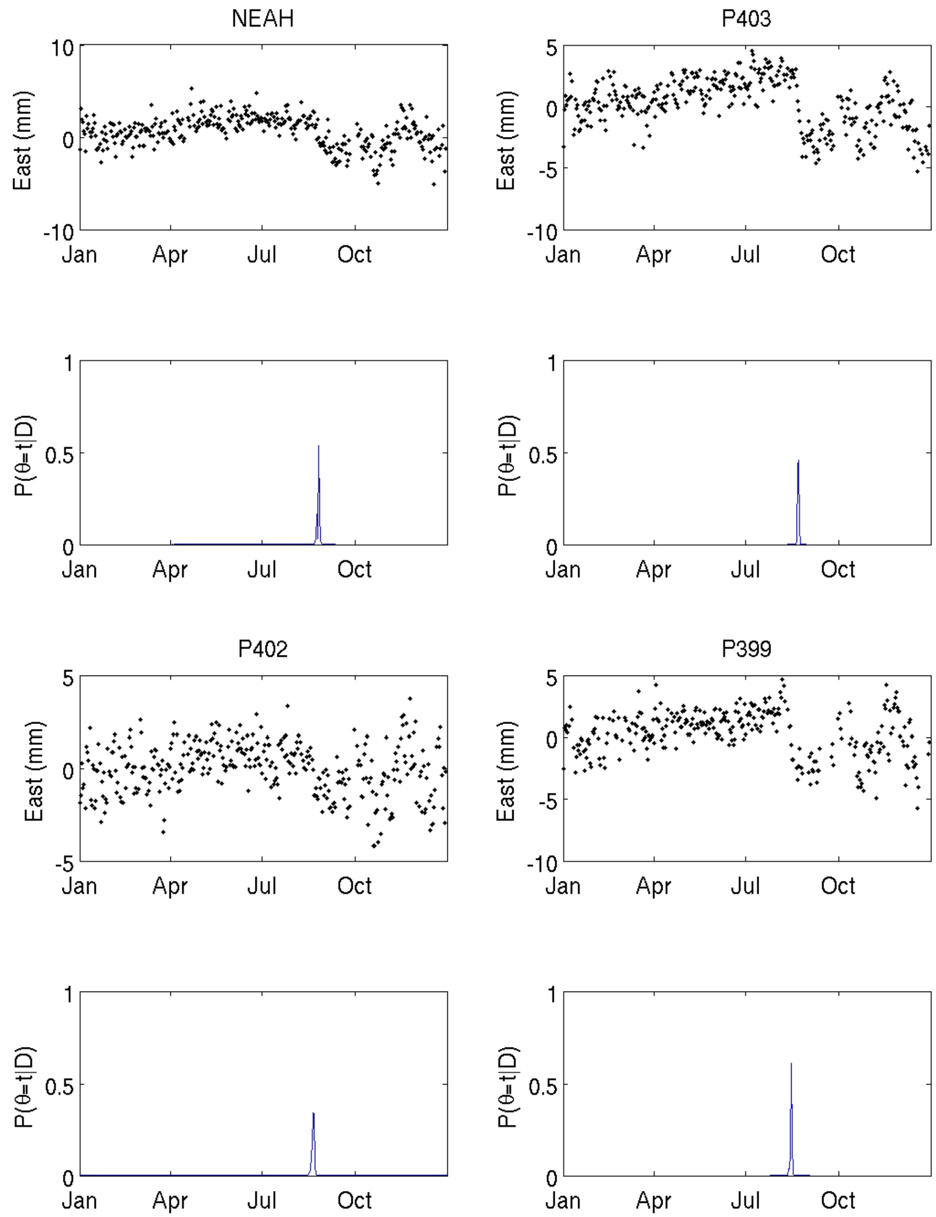
Long story short…

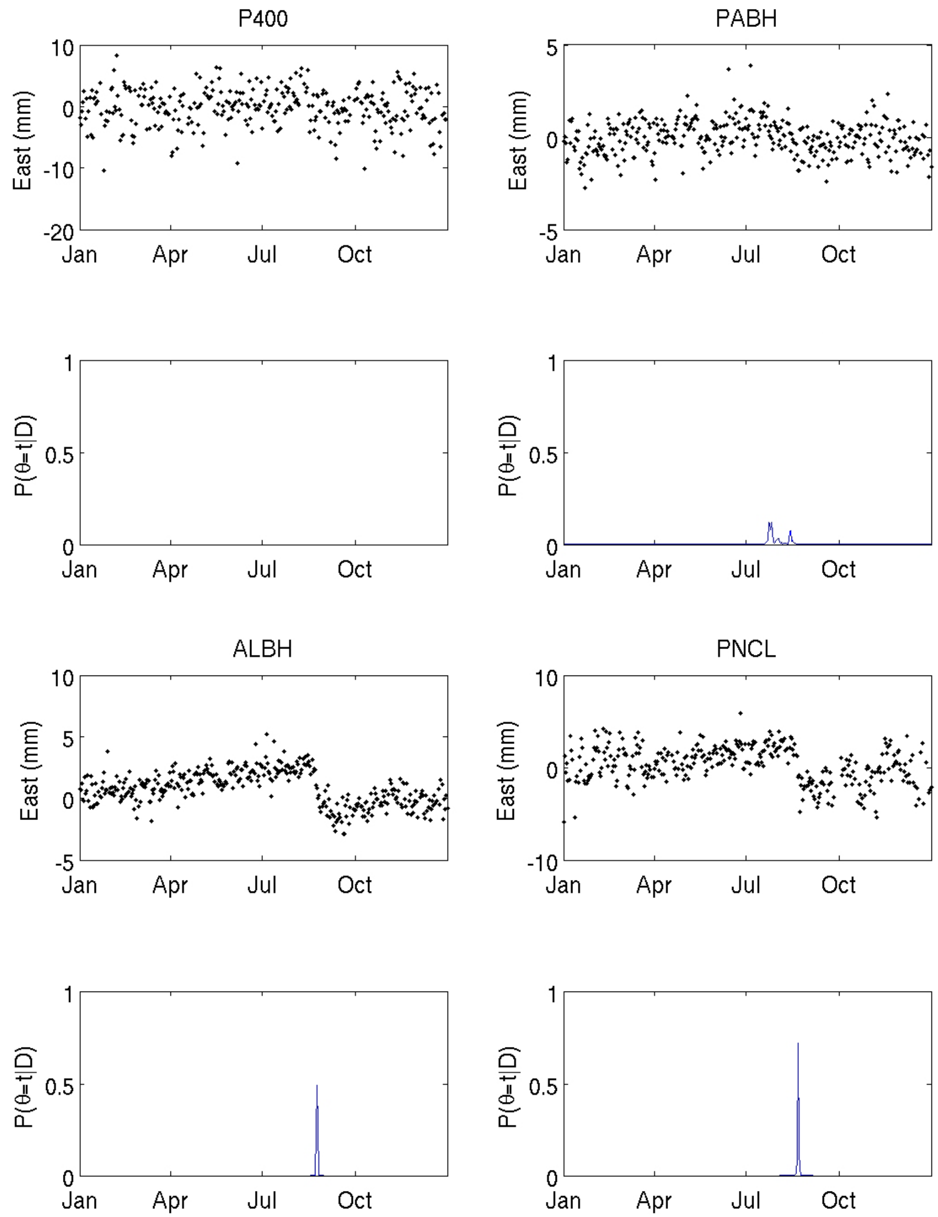
- The posterior probability for a changepoint as a function of time:

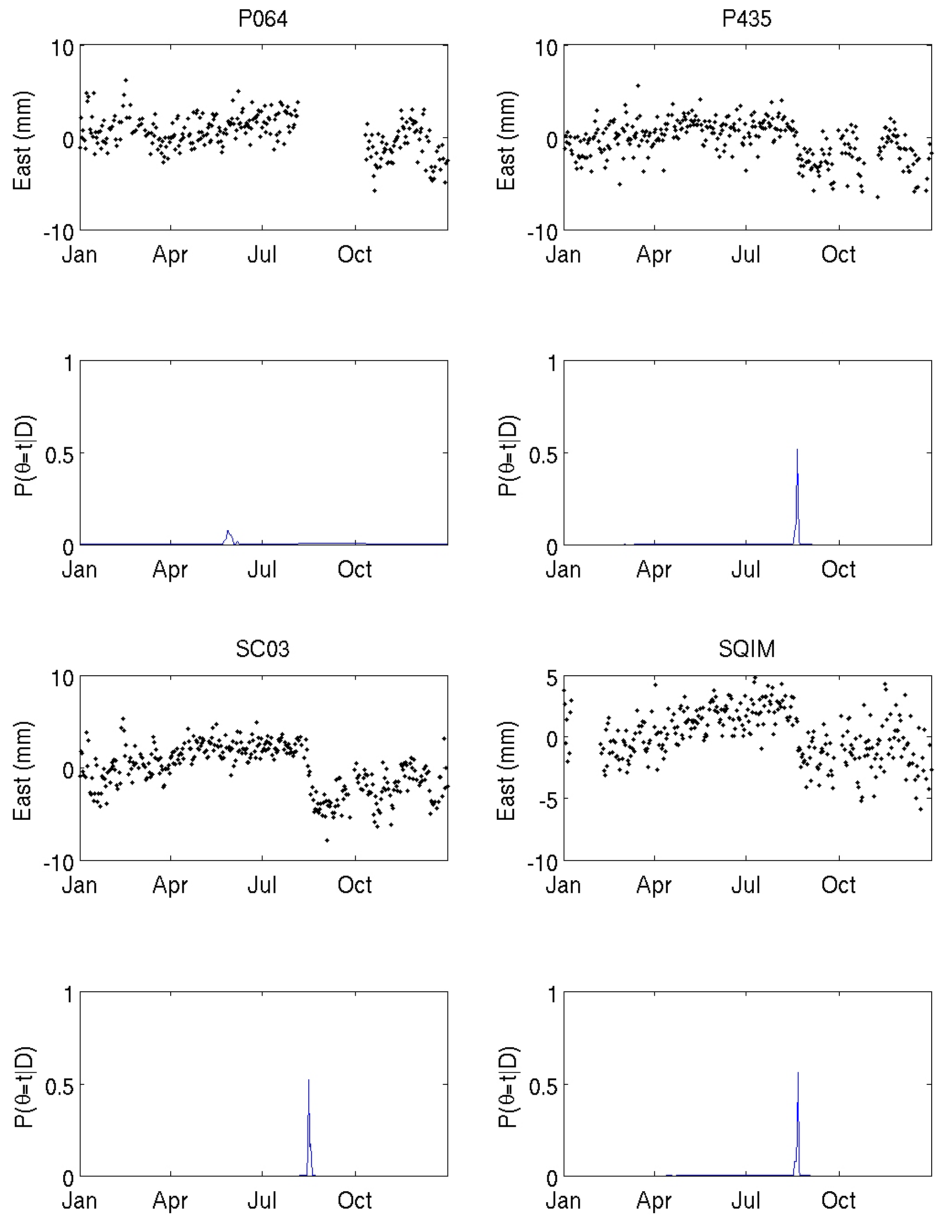
$$P(\theta = t \mid \mathbf{D}, \mathcal{M}) = \frac{P(\mathbf{D} \mid \theta = t, \mathcal{M}) P(\theta = t)}{\sum_{i=1}^{n} P(\mathbf{D} \mid \theta = t_i, \mathcal{M}) P(\theta = t_i)}$$

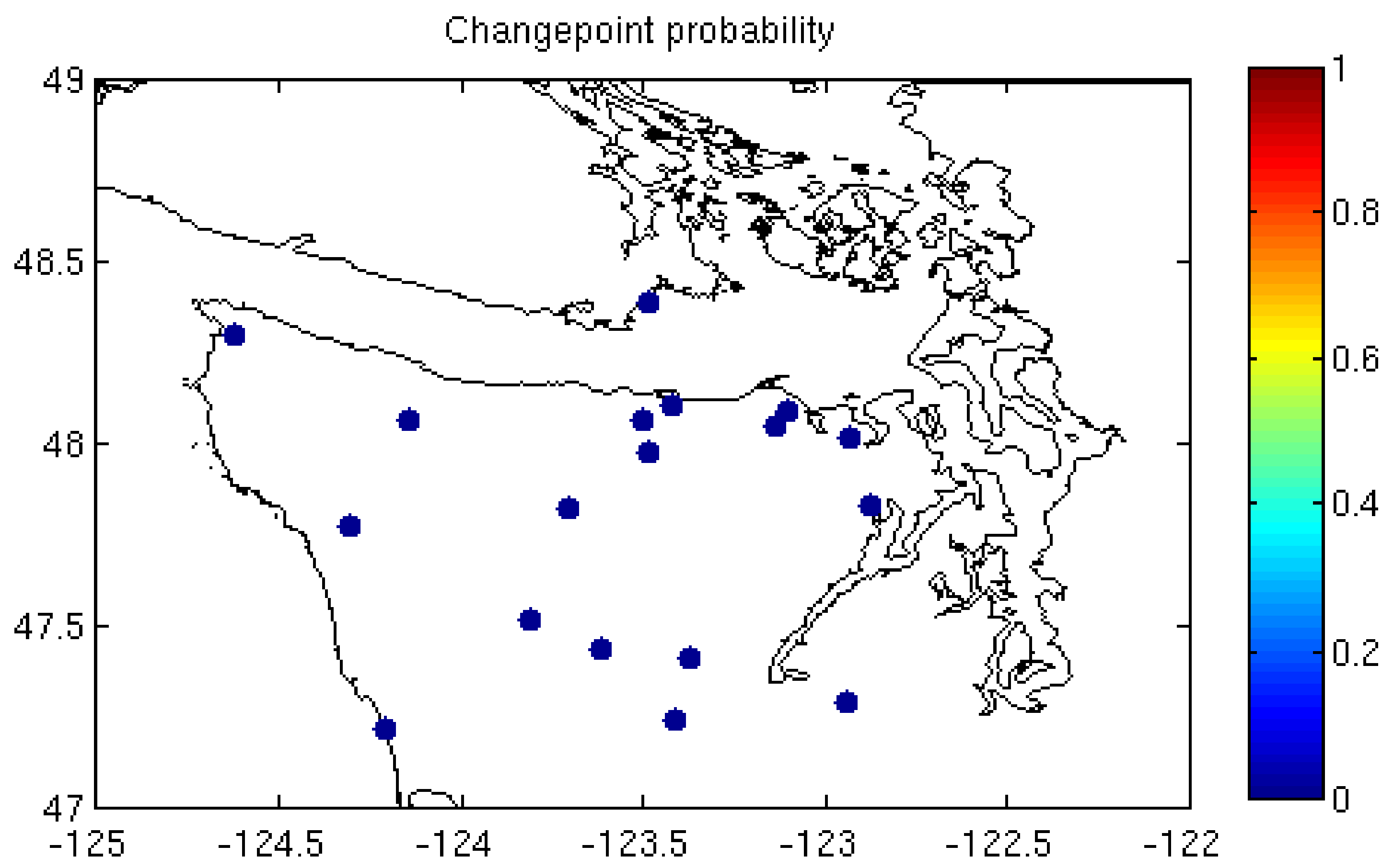
- With

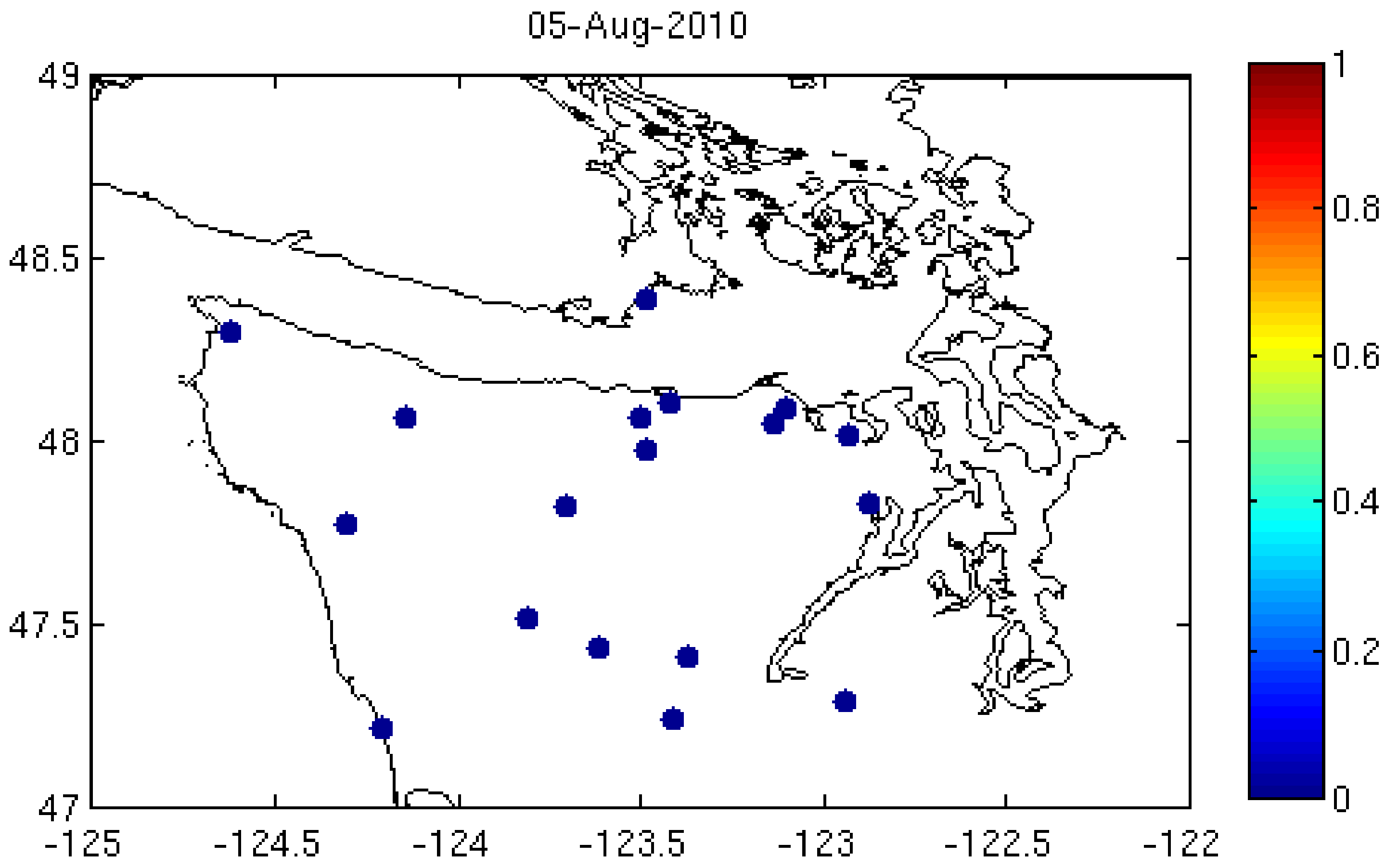
$$P(\mathbf{D} \mid \theta = t, \mathcal{M}) = \frac{1}{(2\pi)^{n/2}} \frac{|\mathbf{V}^*|^{1/2} (b)^a \Gamma(a^*)}{|\mathbf{V}|^{1/2} (b^*)^{a^*} \Gamma(a)}$$

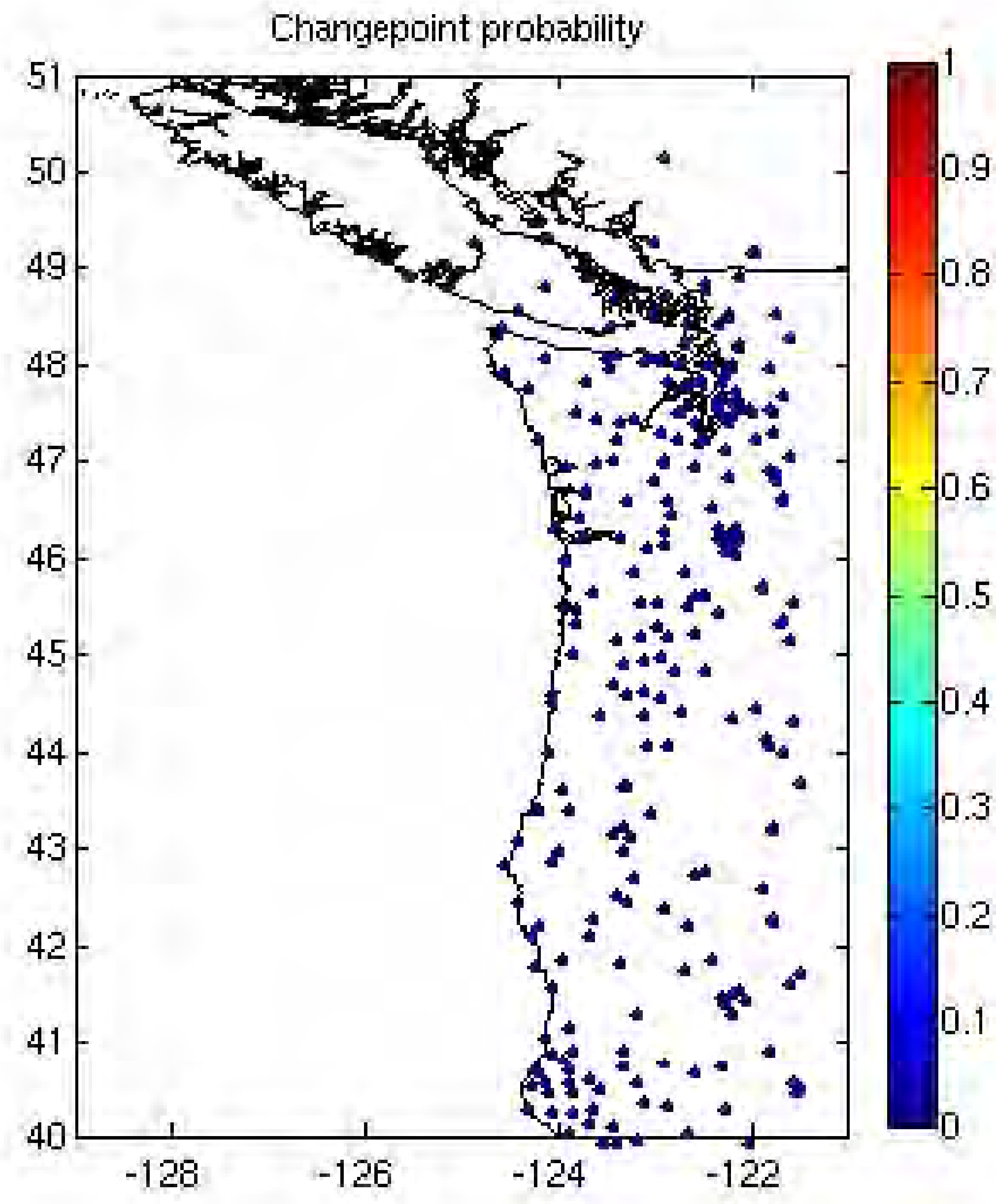

NEAH
P403
P402
P399
East (mm)
P(θ=t|D)
Jan
Apr
Jul
Oct
10
5
0
-5
-10
1
0.5


P400
PABH
ALBH
PNCL
East (mm)
P(θ=t|D)
Jan
Apr
Jul
Oct
10
0
-10
-20
5
-5
1
0.5


P064
P435
SC03
SQIM
East (mm)
P(θ=t|D)
10
0
-10
5
-5
1
0.5
Jan
Apr
Jul
Oct


Changepoint probability
49
48.5
48
47.5
47
-125
-124.5
-124
-123.5
-123
-122.5
-122
1
0.8
0.6
0.4
0.2
0


05-Aug-2010
49
48.5
48
47.5
47
-125
-124.5
-124
-123.5
-123
-122.5
-122
1
0.8
0.6
0.4
0.2
0


Changepoint probability
51
50
49
48
47
46
45
44
43
42
41
40
-128
-126
-124
-122
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

Conclusions

- Bayesian analysis has many advantages over traditional optimization solutions
  - Solve under-determined inverse problems without regularization
  - Produces ensemble of all acceptable solutions
- Given sufficient computational resources, any Bayesian solution can be formed by Monte Carlo simulation
- For some problems, the solution is analytical and just as computationally cheap and easy as traditional LSQ