Using Emulators for Combining Deterministic & Stochastic Models for Hazard Assessment

Robert L Wolpert

Duke University
Department of Statistical Science
with
E.T. Spiller, A.K. Patra, E.B. Pitman,
M.J. Bayarri, J.O. Berger, E.S. Calder.

2017 Aug 02 : JETSCAPE : Columbus, OH

A Team Effort...

Along with many current and former students...

The Caribbean Island of Montserrat

Our Goal:

- Goal: quantify the hazard from Pyroclastic Flows (PFs)—
- For each location "x" on Montserrat, we would like to find the probability of a "catastrophic event" (*e.g.*, inundation to $\geq$1m) within T years— for $T = 1$, $T = 5$, $T = 10$, $T = 20$ years. Isoprobability curves $\Rightarrow$ "Hazard Maps".
- Reflecting all that is uncertain, including:
  - How often will PFs of various volumes occur?
  - What initial direction will they take?
  - How will the flow evolve?
  - How are things changing, over time?
  - What are the uncertain model parameters?
- This is *not* what is needed for short-term crisis and event management— here we consider only long-term hazard.

Our Goal:

- Goal: quantify the hazard from Pyroclastic Flows (PFs)—
- For each location "x" on Montserrat, we would like to find the probability of a "catastrophic event" (*e.g.*, inundation to $\geq$1m) within T years— for $T = 1$, $T = 5$, $T = 10$, $T = 20$ years. Isoprobability curves $\Rightarrow$ "Hazard Maps".
- Reflecting all that is uncertain, including:
  - How often will PFs of various volumes occur?
  - What initial direction will they take?
  - How will the flow evolve?
  - How are things changing, over time?
  - What are the uncertain model parameters?
- This is *not* what is needed for short-term crisis and event management— here we consider only long-term hazard.

Our Goal:

- Goal: quantify the hazard from Pyroclastic Flows (PFs)—
- For each location "x" on Montserrat, we would like to find the probability of a "catastrophic event" (*e.g.*, inundation to $\geq$1m) within T years— for $T = 1$, $T = 5$, $T = 10$, $T = 20$ years. Isoprobability curves $\Rightarrow$ "Hazard Maps".
- Reflecting all that is uncertain, including:
  - How often will PFs of various volumes occur?
  - What initial direction will they take?
  - How will the flow evolve?
  - How are things changing, over time?
  - What are the uncertain model parameters?
- This is *not* what is needed for short-term crisis and event management— here we consider only long-term hazard.

Our Goal:

- Goal: quantify the hazard from Pyroclastic Flows (PFs)—
- For each location "x" on Montserrat, we would like to find the probability of a "catastrophic event" (*e.g.*, inundation to $\geq$1m) within T years— for $T = 1$, $T = 5$, $T = 10$, $T = 20$ years. Isoprobability curves $\Rightarrow$ "Hazard Maps".
- Reflecting all that is uncertain, including:
  - How often will PFs of various volumes occur?
  - What initial direction will they take?
  - How will the flow evolve?
  - How are things changing, over time?
  - What are the uncertain model parameters?
- This is *not* what is needed for short-term crisis and event management— here we consider only long-term hazard.

Pieces of the Puzzle

Our team (geologist, volcanologist, applied mathematicians, statisticians, a stochastic processor) breaks the problem down into three parts: two Stochastic, one Deterministic.

- Try to model the volume, frequency, and initial direction of PFs at SHV, based on 15 years' data, using Prob & EVT;
- For specific volume V and initial direction θ (and other uncertain things like the basal friction angle ϕ), try to model the evolution of the PF, and max depth $M(x)$ at each x, using PDE-based TITAN-2D simulation model
- Try to model what TITAN-2D *would* say at untried locations, using GP Emulator

Pieces of the Puzzle

Our team (geologist, volcanologist, applied mathematicians, statisticians, a stochastic processor) breaks the problem down into three parts: two Stochastic, one Deterministic.

- Try to model the volume, frequency, and initial direction of PFs at SHV, based on 15 years' data, using Prob & EVT;
- For specific volume V and initial direction θ (and other uncertain things like the basal friction angle ϕ), try to model the evolution of the PF, and max depth $M(x)$ at each x, using PDE-based TITAN-2D simulation model
- Try to model what TITAN-2D *would* say at untried locations, using GP Emulator

Pieces of the Puzzle

Our team (geologist, volcanologist, applied mathematicians, statisticians, a stochastic processor) breaks the problem down into three parts: two Stochastic, one Deterministic.

- Try to model the volume, frequency, and initial direction of PFs at SHV, based on 15 years' data, using Prob & EVT;
- For specific volume V and initial direction θ (and other uncertain things like the basal friction angle ϕ), try to model the evolution of the PF, and max depth $M(x)$ at each x, using PDE-based TITAN-2D simulation model
- Try to model what TITAN-2D *would* say at untried locations, using GP Emulator

A Brute Force Bayesian Method

- Draw $V_j,\ \tau_j,\ \theta_j,\ \phi_j$ from some dist'n;
- Compute inundation height $M(x;\ V_j, \theta_j, \phi_j)$ at site x using TITAN-2D for each site x of interest;
- Use M/H to decide if you like the draw;
- Repeat about 10^6 times
- Consuming about 10^6 hours.

A Brute Force Bayesian Method

- Draw $V_j,\ \tau_j,\ \theta_j,\ \phi_j$ from some dist'n;
- Compute inundation height $M(x;\ V_j, \theta_j, \phi_j)$ at site x using TITAN-2D for each site x of interest;
- Use M/H to decide if you like the draw;
- Repeat about 10^6 times
- Consuming about 10^6 hours.

A Brute Force Bayesian Method

- Draw $V_j,\ \tau_j,\ \theta_j,\ \phi_j$ from some dist'n;
- Compute inundation height $M(x;\ V_j, \theta_j, \phi_j)$ at site x using TITAN-2D for each site x of interest;
- Use M/H to decide if you like the draw;
- Repeat about 10^6 times
- Consuming about 10^6 hours.

A Brute Force Bayesian Method

- Draw $V_j,\ \tau_j,\ \theta_j,\ \phi_j$ from some dist'n;
- Compute inundation height $M(x;\ V_j, \theta_j, \phi_j)$ at site x using TITAN-2D for each site x of interest;
- Use M/H to decide if you like the draw;
- Repeat about 10^6 times
- Consuming about 10^6 hours.

A Brute Force Method That Doesn't Work

- Draw $V_j,\ \tau_j,\ \theta_j,\ \phi_j$ from some dist'n;
- Compute inundation height $M(x;\ V_j, \theta_j, \phi_j)$ at site x using TITAN-2D for each site x of interest;
- Use M/H to decide if you like the draw;
- Repeat about 10^6 times
- Consuming about 10^6 hours.

A Brute Force Method That Doesn't Work

- Draw $V_j,\ \tau_j,\ \theta_j,\ \phi_j$ from some dist'n;
- Compute inundation height $M(x;\ V_j, \theta_j, \phi_j)$ at site x using TITAN-2D for each site x of interest;
- Use M/H to decide if you like the draw;
- Repeat about 10^6 times
- Consuming about 10^6 hours (> 114 years).

A Better Brute Force Method

After fitting a GP Emulator to T2D,

- Draw $V_j,\ \tau_j,\ \theta_j,\ \phi_j$;
- Compute $\mathrm{E}[M(x;\ V_j, \theta_j, \phi_j) \mid \text{Training Data}]$ for each x;
- Keep the ones you like (M/H);
- Repeat about 10^6 times.
- Fast enough now, but

  Still missing low-probability high-consequence tail behavior.

A Better Brute Force Method

After fitting a GP Emulator to T2D,

- Draw $V_j,\ \tau_j,\ \theta_j,\ \phi_j$;
- Compute $\mathrm{E}[M(x;\ V_j, \theta_j, \phi_j) \mid \text{Training Data}]$ for each x;
- Keep the ones you like (M/H);
- Repeat about 10^6 times.
- Fast enough now, but

  Still missing low-probability high-consequence tail behavior.

A Better Brute Force Method

After fitting a GP Emulator to T2D,

- Draw $V_j,\ \tau_j,\ \theta_j,\ \phi_j$;
- Compute $\mathrm{E}[M(x;\ V_j, \theta_j, \phi_j) \mid \text{Training Data}]$ for each x;
- Keep the ones you like (M/H);
- Repeat about 10^6 times.
- Fast enough now, but

  Still missing low-probability high-consequence tail behavior.

A Better Brute Force Method

After fitting a GP Emulator to T2D,

- Draw $V_j,\ \tau_j,\ \theta_j,\ \phi_j$;
- Compute $\mathrm{E}[M(x;\ V_j, \theta_j, \phi_j) \mid \text{Training Data}]$ for each x;
- Keep the ones you like (M/H);
- Repeat about 10^6 times.
- Fast enough now, but

  Still missing low-probability high-consequence tail behavior.

A Better Brute Force Method

After fitting a GP Emulator to T2D,

- Draw $V_j,\ \tau_j,\ \theta_j,\ \phi_j$;
- Compute $\mathrm{E}[M(x;\ V_j, \theta_j, \phi_j) \mid \text{Training Data}]$ for each x;
- Keep the ones you like (M/H);
- Repeat about 10^6 times.
- Fast enough now, but

  Still missing low-probability high-consequence tail behavior.

A Better Brute Force Method That Still Doesn't Work

After fitting a GP Emulator to T2D,

- Draw $V_j,\ \tau_j,\ \theta_j,\ \phi_j$;
- Compute $\mathrm{E}[M(x;\ V_j, \theta_j, \phi_j) \mid \text{Training Data}]$ for each x;
- Keep the ones you like (M/H);
- Repeat about 10^6 times.
- Fast enough now, but

  Still missing low-probability high-consequence tail behavior.

Finally, a New Idea

- Key idea: If some volume V with initiation angle θ inundates location x, then any higher volume $V' > V$ will too.
- SO: For each angle $0 \leq \theta < 2\pi$, identify the *smallest* volume $\Psi(\theta; x)$ that will reach x.
- Then: Find the probability of at least one $(V_j,\ \tau_j,\ \theta_j)$ with:

$$V_j \geq \Psi(\theta_j, x), \qquad \tau_j \in \left(t_0, t_0 + T\right]$$

- How do we find $\Psi(\theta, x)$?

Finally, a New Idea

- Key idea: If some volume V with initiation angle θ inundates location x, then any higher volume $V' > V$ will too.
- SO: For each angle $0 \leq \theta < 2\pi$, identify the *smallest* volume $\Psi(\theta; x)$ that will reach x.
- Then: Find the probability of at least one $(V_j,\ \tau_j,\ \theta_j)$ with:

$$V_j \geq \Psi(\theta_j, x), \qquad \tau_j \in \left(t_0, t_0 + T\right]$$

- How do we find $\Psi(\theta, x)$?

Finally, a New Idea

- Key idea: If some volume V with initiation angle θ inundates location x, then any higher volume $V' > V$ will too.
- SO: For each angle $0 \leq \theta < 2\pi$, identify the *smallest* volume $\Psi(\theta; x)$ that will reach x.
- Then: Find the probability of at least one $(V_j,\ \tau_j,\ \theta_j)$ with:

$$V_j \geq \Psi(\theta_j, x), \qquad \tau_j \in \left(t_0, t_0 + T\right]$$

- How do we find $\Psi(\theta, x)$?

Finally, a New Idea That Just Might Work

- Key idea: If some volume V with initiation angle θ inundates location x, then any higher volume $V' > V$ will too.
- SO: For each angle $0 \leq \theta < 2\pi$, identify the *smallest* volume $\Psi(\theta; x)$ that will reach x.
- Then: Find the probability of at least one $(V_j,\ \tau_j,\ \theta_j)$ with:

$$V_j \geq \Psi(\theta_j, x), \qquad \tau_j \in \left(t_0, t_0 + T\right]$$

- How do we find $\Psi(\theta, x)$?

Emulation

We could pick off $\Psi(\theta; x)$ from a grid of TITAN-2D outputs at each of:

- 100 Initiation Angles θ_i;
- 100 Volumes V_j;
- 100 Friction Angles ϕ_k;

Which would entail maybe $100 \times 100 \times 100 = 1\,000\,000$ runs of TITAN-2D...but we still don't have $1\,000\,000$ hours. Our Solution:

Build an Emulator for the slow T2D PDE flow model.

Emulation

We could pick off $\Psi(\theta; x)$ from a grid of TITAN-2D outputs at each of:

- 100 Initiation Angles θ_i;
- 100 Volumes V_j;
- 100 Friction Angles ϕ_k;

Which would entail maybe $100 \times 100 \times 100 = 1\,000\,000$ runs of TITAN-2D...but we still don't have $1\,000\,000$ hours. Our Solution:

Build an Emulator for the slow T2D PDE flow model.

Emulation

We could pick off $\Psi(\theta; x)$ from a grid of TITAN-2D outputs at each of:

- 100 Initiation Angles θ_i;
- 100 Volumes V_j;
- 100 Friction Angles ϕ_k;

Which would entail maybe $100 \times 100 \times 100 = 1\,000\,000$ runs of TITAN-2D...but we still don't have $1\,000\,000$ hours. Our Solution:

Build an Emulator for the slow T2D PDE flow model.

Emulation

We could pick off $\Psi(\theta; x)$ from a grid of TITAN-2D outputs at each of:

- 100 Initiation Angles θ_i;
- 100 Volumes V_j;
- 100 Friction Angles ϕ_k;

Which would entail maybe $100 \times 100 \times 100 = 1\,000\,000$ runs of TITAN-2D...but we still don't have $1\,000\,000$ hours. Our Solution:

Build an Emulator for the slow T2D PDE flow model.

Emulators

A GP Emulator is a (very fast):

- statistical model (based on Gaussian Processes) for our
- computer model (based on PDE solver) of the
- volcano.

- The emulator can predict (in seconds) what TITAN-2D would return (in hours), *with* an estimate of its accuracy;
- Based on Gaussian Stochastic Process (GP) Model. We:
  - Pick a few hundred "design points" (V_j, θ_j, ϕ_j, ...) (LHD);
  - Run TITAN-2D at each of them to find TITAN-2D Model Output $M(x;\ V_j, \theta_j, \phi_j)$;
  - Train the GP to return a statistical estimate

$$\mathrm{E}\big[M(x;\ V, \theta, \phi) \mid \{M(x;\ V_j, \theta_j, \phi_j)\}\big]$$

of model output M for site x at untried points (V, θ, ϕ).

Emulators

A GP Emulator is a (very fast):

- statistical model (based on Gaussian Processes) for our
- computer model (based on PDE solver) of the
- volcano.

- The emulator can predict (in seconds) what TITAN-2D would return (in hours), *with* an estimate of its accuracy;
- Based on Gaussian Stochastic Process (GP) Model. We:
  - Pick a few hundred "design points" (V_j, θ_j, ϕ_j, ...) (LHD);
  - Run TITAN-2D at each of them to find TITAN-2D Model Output $M(x;\ V_j, \theta_j, \phi_j)$;
  - Train the GP to return a statistical estimate

$$\mathrm{E}\big[M(x;\ V, \theta, \phi) \mid \{M(x;\ V_j, \theta_j, \phi_j)\}\big]$$

of model output M for site x at untried points (V, θ, ϕ).

A Reminder— What do we Emulate, and Why?

With this we try to evaluate a "threshold function" for each site x in Montserrat:

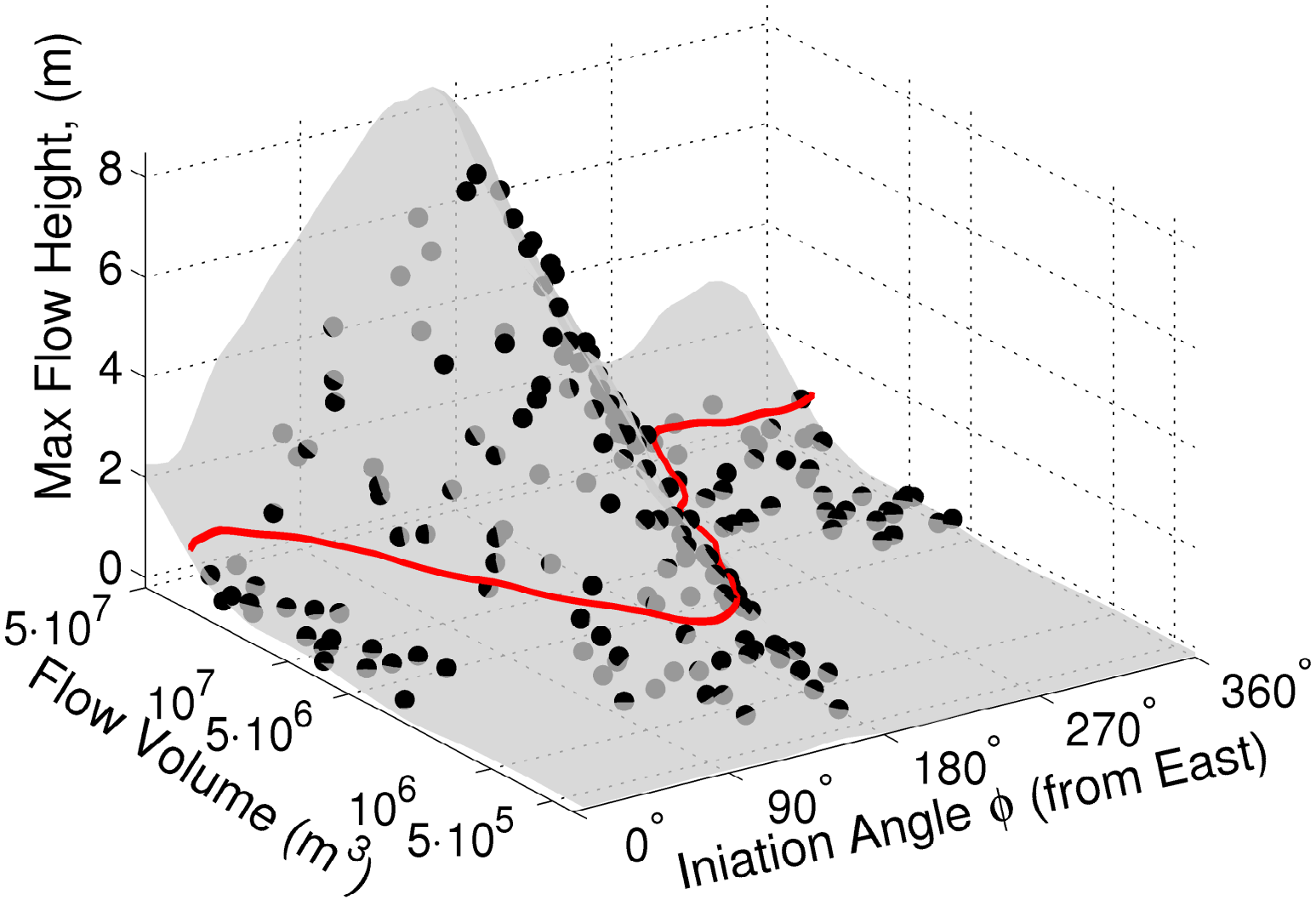
$$\Psi(\theta; x) = \text{Smallest volume } V \text{ that would inundate site } x \text{ if flow begins in direction } \theta$$

for each possible angle θ (0–2π, with $0 =$ East and $\pi/2 =$ North). Then we Quantify the Uncertain Hazard at x for T years as

$$\Pr\left\{ V_i \geq \Psi(\theta_i; x) \text{ for some PF } (V_i, \theta_i) \text{ in time } (t_0, t_0 + T] \right\}$$

which in turn we study with the probability models.

Finding $\Psi(\theta; x)$ In Pictures



Designs and Sub-designs

- We train the Emulator at "Design Points" $\{(V_i, \theta_i, \phi_i)\}$—
- *Not* random variables from a prior distribution— rather,
  - a deliberately chosen set of points, intended to be as widely-dispersed as we can manage in a 3-dim space.
- We use a maxi-min Latin Hypercube Design (LHD), with 2048 points in $[10^5, 10^9] \times [0, 2\pi) \times (8^\circ, 12^\circ)$
- . . . but we can't invert 2048×2048 matrices inside an MCMC loop, so we look for special SUBDESIGNS ($\approx$ 40–50 pts) for each x of the design points *important for that* x.

Designs and Sub-designs

- We train the Emulator at "Design Points" $\{(V_i, \theta_i, \phi_i)\}$—
- *Not* random variables from a prior distribution— rather,
  - a deliberately chosen set of points, intended to be as widely-dispersed as we can manage in a 3-dim space.
- We use a maxi-min Latin Hypercube Design (LHD), with 2048 points in $[10^5, 10^9] \times [0, 2\pi) \times (8^\circ, 12^\circ)$
- . . . but we can't invert 2048×2048 matrices inside an MCMC loop, so we look for special SUBDESIGNS ($\approx$ 40–50 pts) for each x of the design points *important for that* x.

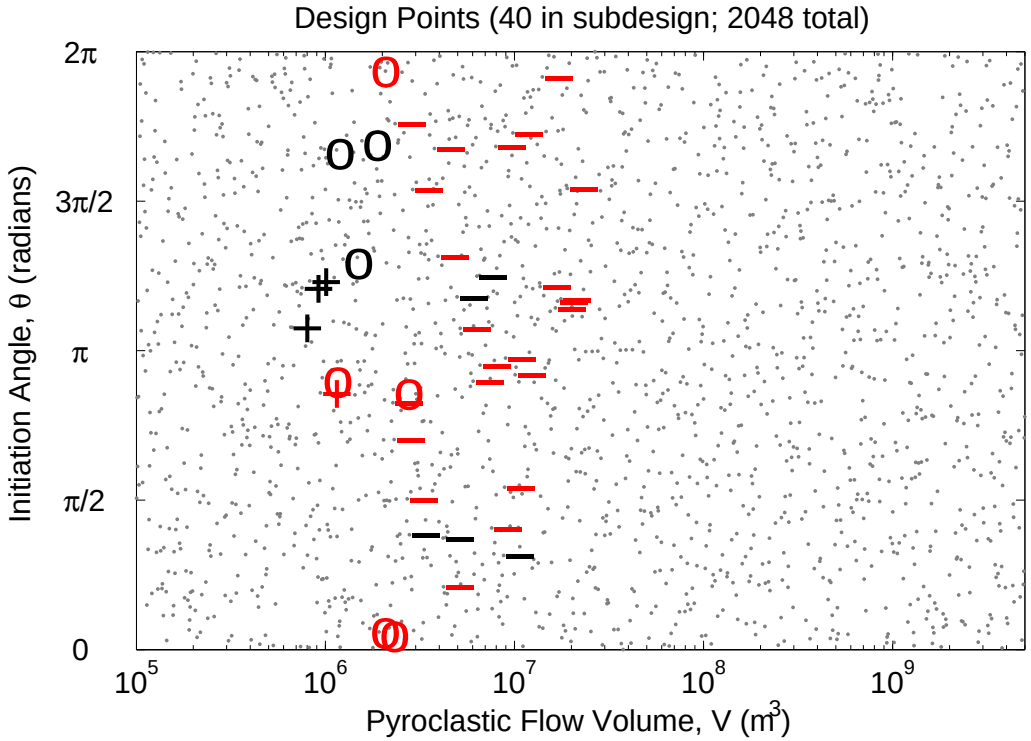
Designs and Sub-designs

- We train the Emulator at "Design Points" $\{(V_i, \theta_i, \phi_i)\}$—
- *Not* random variables from a prior distribution— rather,
  - a deliberately chosen set of points, intended to be as widely-dispersed as we can manage in a 3-dim space.
- We use a maxi-min Latin Hypercube Design (LHD), with 2048 points in $[10^5, 10^9] \times [0, 2\pi) \times (8^\circ, 12^\circ)$
- . . . but we can't invert 2048×2048 matrices inside an MCMC loop, so we look for special SUBDESIGNS ($\approx$ 40–50 pts) for each x of the design points *important for that* x.

Designs and Sub-designs

- We train the Emulator at "Design Points" $\{(V_i, \theta_i, \phi_i)\}$—
- *Not* random variables from a prior distribution— rather,
  - a deliberately chosen set of points, intended to be as widely-dispersed as we can manage in a 3-dim space.
- We use a maxi-min Latin Hypercube Design (LHD), with 2048 points in $[10^5, 10^9] \times [0, 2\pi) \times (8^\circ, 12^\circ)$
- . . . but we can't invert 2048×2048 matrices inside an MCMC loop, so we look for special SUBDESIGNS ($\approx$ 40–50 pts) for each x of the design points *important for that* x.

Design points V, θ

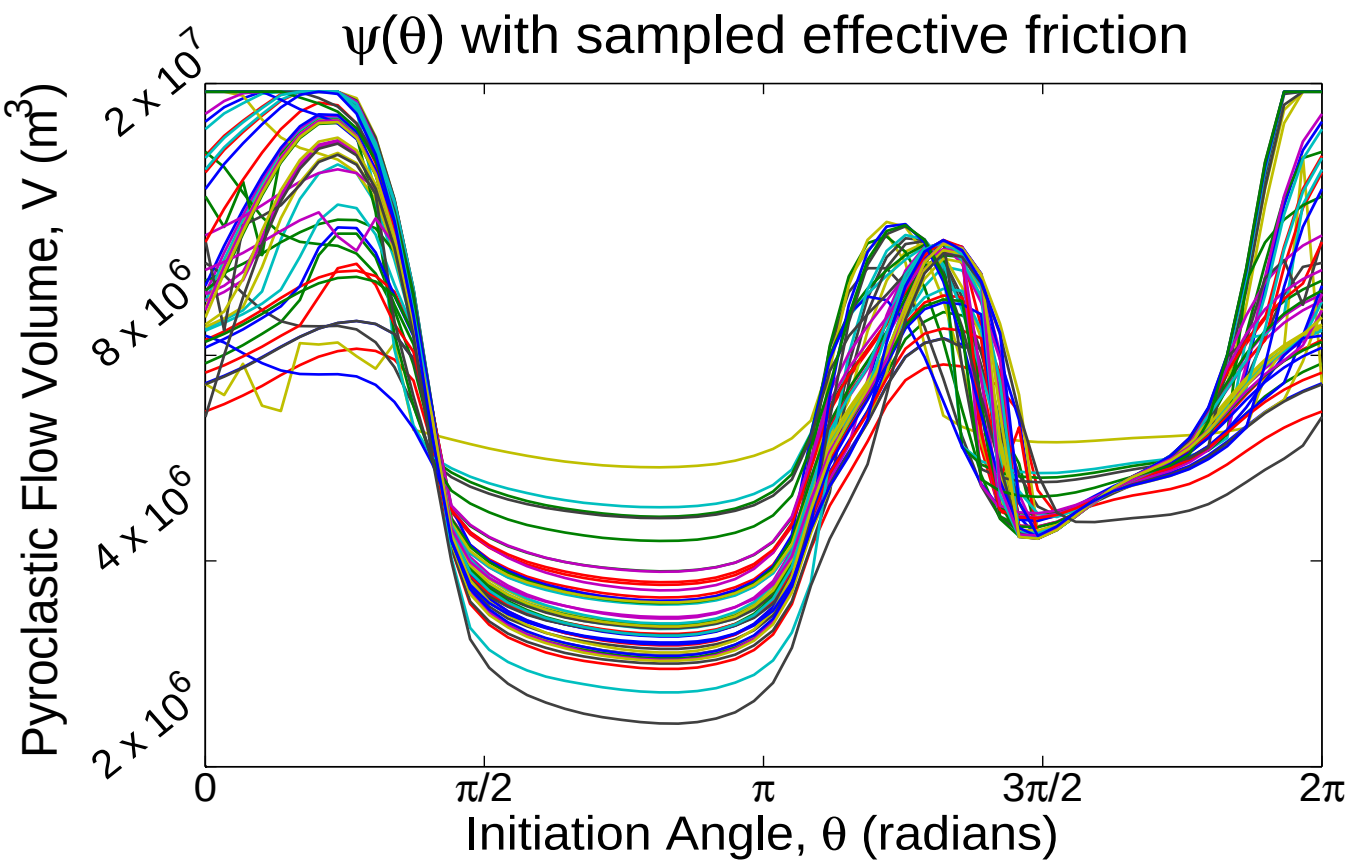


2048 Design Points total; 40 used for determining Ψ at $x=$ Level 4 Trigger Point: Dyers River Valley (head of Belham valley) .

Black $\Longleftrightarrow M(x;\ V,\theta,\phi)=0$ Red $\Longleftrightarrow M(x;\ V,\theta,\phi)>0$;

$-$: $8^\circ < \phi < 8.5^\circ$; $\bigcirc$: $8.5^0 < \phi < 10.5^\circ$; $+$: $10.5^\circ < \phi < 12^\circ$.

50 Simulations of $\Psi(\theta; x)$ for TP4, Dyers RV



Combining Det & Stoch Models

We have tried to separate our challenge into

- The Deterministic problem of identifying those "input values" V, θ, ϕ that lead to a catastrophe at x, using T2D, and
- The Stochastic problem of finding the probability of such an event.
- Use Det Model to identify the Hazard Region in parameter space;
- Use Stoch Model to estimate the Probability of disaster.

I say "*tried*" because we needed to Emulate T2D and emulation will entail more uncertainty to combine!!

Combining Det & Stoch Models

We have tried to separate our challenge into

- The Deterministic problem of identifying those "input values" V, θ, ϕ that lead to a catastrophe at x, using T2D, and
- The Stochastic problem of finding the probability of such an event.
- Use Det Model to identify the Hazard Region in parameter space;
- Use Stoch Model to estimate the Probability of disaster.

I say "*tried*" because we needed to Emulate T2D and emulation will entail more uncertainty to combine!!

Combining Det & Stoch Models

We have tried to separate our challenge into

- The Deterministic problem of identifying those "input values" V, θ, ϕ that lead to a catastrophe at x, using T2D, and
- The Stochastic problem of finding the probability of such an event.
- Use Det Model to identify the Hazard Region in parameter space;
- Use Stoch Model to estimate the Probability of disaster.

I say "*tried*" because we needed to Emulate T2D and emulation will entail more uncertainty to combine!!

Combining Det & Stoch Models

We have tried to separate our challenge into

- The Deterministic problem of identifying those "input values" V, θ, ϕ that lead to a catastrophe at x, using T2D, and
- The Stochastic problem of finding the probability of such an event.
- Use Det Model to identify the Hazard Region in parameter space;
- Use Stoch Model to estimate the Probability of disaster.

I say "*tried*" because we needed to Emulate T2D and emulation will entail more uncertainty to combine!!

Our Stochastic Model

For us a PF (V_j, θ_j, τ_j) is *catastrophic* for x if $V_j \geq \Psi(\theta_j; x)$ and $t_0 < \tau_j \leq t_0 + T$.

SO, we need a joint model for points $\{(V_j, \theta_j, \tau_j)\} \subset \mathbb{R}^3$.
Let's do it in that order: first Volumes, then Angles, then Times.

Our Stochastic Model

For us a PF (V_j, θ_j, τ_j) is *catastrophic* for x if $V_j \geq \Psi(\theta_j; x)$ and $t_0 < \tau_j \leq t_0 + T$.

SO, we need a joint model for points $\{(V_j, \theta_j, \tau_j)\} \subset \mathbb{R}^3$.

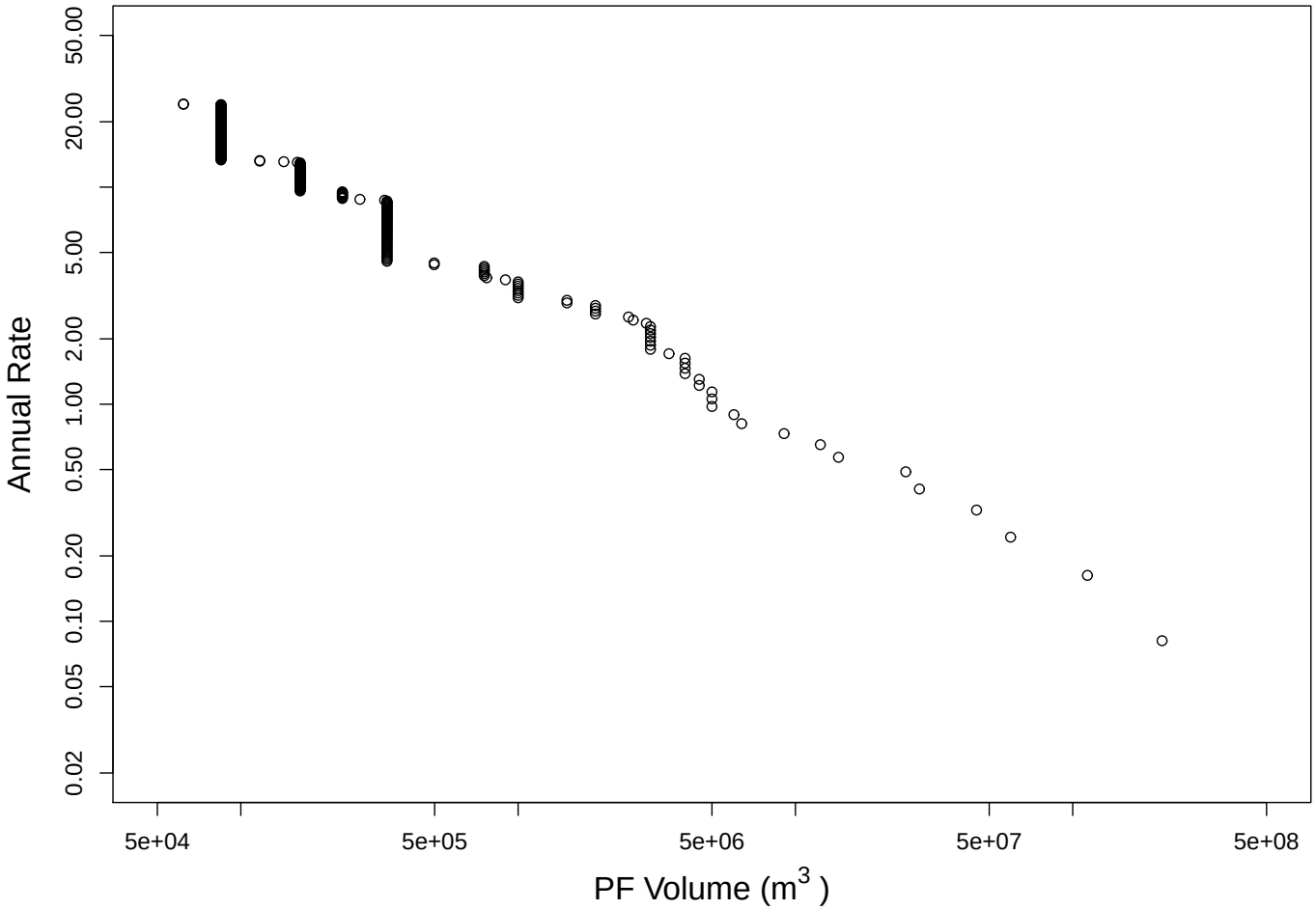
Let's do it in that order: first Volumes, then Angles, then Times.

Our Stochastic Model

For us a PF (V_j, θ_j, τ_j) is *catastrophic* for x if
$V_j \geq \Psi(\theta_j; x)$ and $t_0 < \tau_j \leq t_0 + T$.

SO, we need a joint model for points $\{(V_j, \theta_j, \tau_j)\} \subset \mathbb{R}^3$.
Let's do it in that order: first Volumes, then Angles, then Times.

PF Volumes



Seems kind of linear, on log-log scale... remember that. rtn

PF Frequencies (1990's–2010)

- PF's exceeding 10^4 m^3: Daily, 0–1.6 km runout
- PF's exceeding 10^5 m^3: Fortnightly, 2–3.0 km runout
- PF's exceeding 10^6 m^3: Quarterly, 4–6.0 km runout
- PF's exceeding 10^7 m^3: Yearly, 6.0+ km runout
- PF's exceeding 10^8 m^3: Just two, unknown runout
- PF's exceeding 10^9 m^3: None YET, but ...

PF Frequencies (1990's–2010)

- PF's exceeding 10^4 m^3: Daily, 0–1.6 km runout
- PF's exceeding 10^5 m^3: Fortnightly, 2–3.0 km runout
- PF's exceeding 10^6 m^3: Quarterly, 4–6.0 km runout
- PF's exceeding 10^7 m^3: Yearly, 6.0+ km runout
- PF's exceeding 10^8 m^3: Just two, unknown runout
- PF's exceeding 10^9 m^3: None YET, but ...

Flow Volume Context

How much *is* $10^9\,\mathbf{m}^3 = 1\,\mathbf{km}^3$ **of flow?**

V	Cube Width	Radius 1 m deep
10^3	10 m (32.8')	18 m
10^6	100 m (328')	564 m
10^9	1000 m (0.6 mi)	17.8 km

Montserrat is 16 km × 10 km.

Flow Volume Context

How much *is* $10^9\,\mathbf{m}^3 = 1\,\mathbf{km}^3$ **of flow?**

V	Cube Width	Radius 1 m deep
10^3	10 m (32.8')	18 m
10^6	100 m (328')	564 m
10^9	1000 m (0.6 mi)	17.8 km

Montserrat is 16 km × 10 km.

Power Laws

Remember how the log-log Freq. *vs.* Volume plot was nearly linear?
Linear log-log plots of Magnitude *vs.* Frequency are a hallmark of the Pareto probability distribution $\mathrm{Pa}(\alpha, \epsilon)$,

$$\mathrm{P}[V > v] = (v/\epsilon)^{-\alpha}, \qquad v > \epsilon.$$

Which is kind of bad news.

Power Laws

Remember how the log-log Freq. *vs.* Volume plot was nearly linear?
Linear log-log plots of Magnitude *vs.* Frequency are a hallmark of the Pareto probability distribution $\mathrm{Pa}(\alpha, \epsilon)$,

$$\mathrm{P}[V > v] = (v/\epsilon)^{-\alpha}, \qquad v > \epsilon.$$

Which is kind of bad news.

Power Laws

Remember how the log-log Freq. *vs.* Volume plot was nearly linear?
Linear log-log plots of Magnitude *vs.* Frequency are a hallmark of the Pareto probability distribution $\mathrm{Pa}(\alpha, \epsilon)$,

$$\mathrm{P}[V > v] = (v/\epsilon)^{-\alpha}, \qquad v > \epsilon.$$

Which is kind of bad news.

The Pareto Distribution

The negative slope seems to be about $\alpha \approx 0.64$ or so.
The Pareto distribution with $\alpha < 1$ has:

- Heavy tails;
- Infinite mean $\mathrm{E}[V] = \infty$, infinite variance $\mathrm{E}[V^2] = \infty$;
- No Central Limit Theorem for sums (α-Stable instead);
- Significant chance that, in the future, we will see volumes larger than any we have seen in the past. Like $V > 10^9 \mathrm{m}^3$.
- The Pareto comes up all the time in the Peaks over Threshold (PoT) approach to the Statistics of Extreme Events— related to Fisher/Tippett Three Types Theorem.

The Pareto Distribution

The negative slope seems to be about $\alpha \approx 0.64$ or so.
The Pareto distribution with $\alpha < 1$ has:

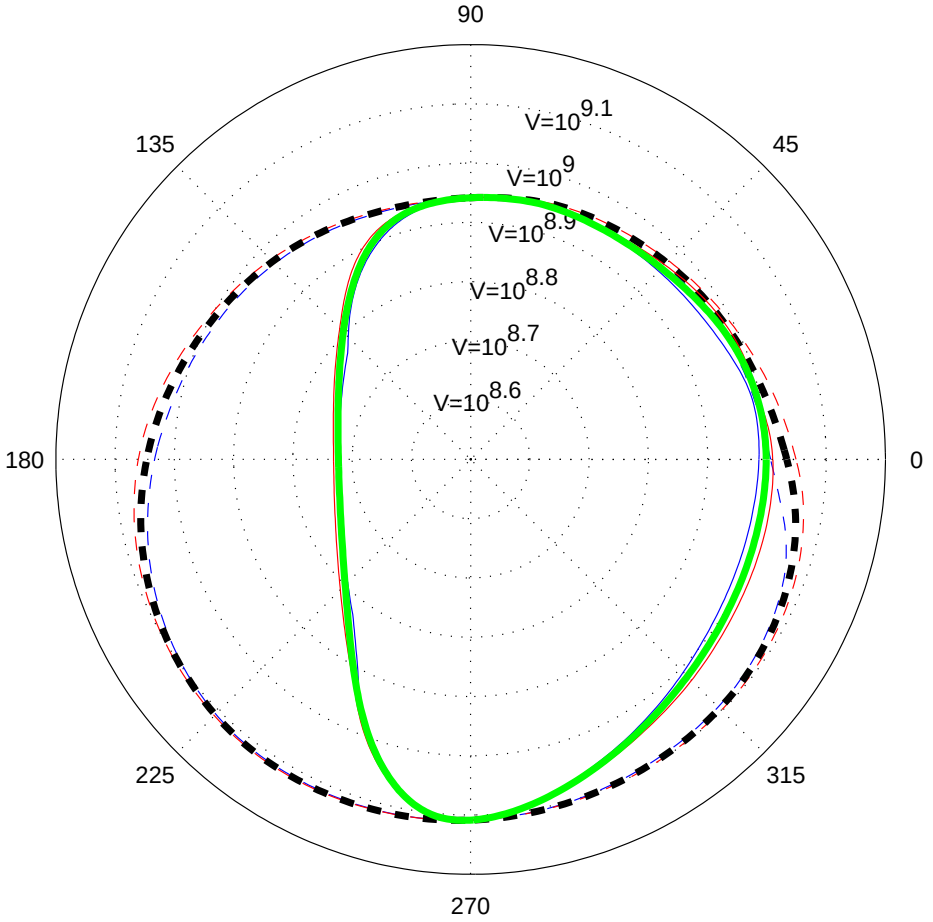
- Heavy tails;
- Infinite mean $\mathrm{E}[V] = \infty$, infinite variance $\mathrm{E}[V^2] = \infty$;
- No Central Limit Theorem for sums (α-Stable instead);
- Significant chance that, in the future, we will see volumes larger than any we have seen in the past. Like $V > 10^9 \mathrm{m}^3$.
- The Pareto comes up all the time in the Peaks over Threshold (PoT) approach to the Statistics of Extreme Events— related to Fisher/Tippett Three Types Theorem.

The Pareto Distribution

The negative slope seems to be about $\alpha \approx 0.64$ or so.
The Pareto distribution with $\alpha < 1$ has:

- Heavy tails;
- Infinite mean $\mathrm{E}[V] = \infty$, infinite variance $\mathrm{E}[V^2] = \infty$;
- No Central Limit Theorem for sums (α-Stable instead);
- Significant chance that, in the future, we will see volumes larger than any we have seen in the past. Like $V > 10^9 \mathrm{m}^3$.
- The Pareto comes up all the time in the Peaks over Threshold (PoT) approach to the Statistics of Extreme Events— related to Fisher/Tippett Three Types Theorem.

How big is bad?

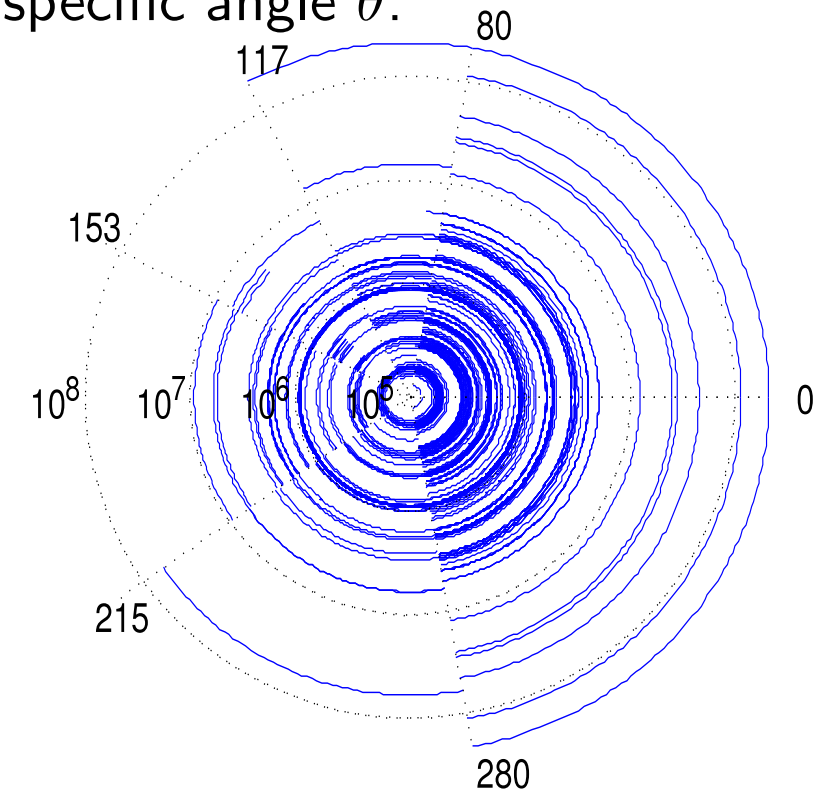


Kinda depends which way it goes...

Skip Model

PF Initiation Angles

Our data on angles is quite vague— we only know which of 7 or 8 *valleys* were reached by a given PF, from which we can infer a *sector* but not a specific angle θ:



Any nonuniformity for θ? Any dependence of θ on Volume V?

Angle/Volume Data (cont)

We need a *joint* density function

$$f(V,\ \theta) = f(V)\ f(\theta \mid V)$$

given by:

$$V, \theta \sim \alpha \epsilon^{\alpha}\ V^{-\alpha-1} \mathbf{1}_{\{V>\epsilon\}} \pi_{\kappa}(\theta)$$

where $\pi_{\kappa}(\theta)$ is the von Mises distribution with pdf

$$f(\theta \mid \kappa, \mu) = \frac{e^{\kappa \cos(\theta-\mu)}}{2\pi I_0(\kappa)},$$

centered at $\theta \approx \mu$ close to zero (East) with concentration κ that might depend on V if the data support that.

Angle/Volume Data (cont)

We need a *joint* density function

$$f(V,\ \theta) = f(V)\ f(\theta \mid V)$$

given by:

$$V, \theta \sim \alpha \epsilon^{\alpha}\ V^{-\alpha-1} \mathbf{1}_{\{V>\epsilon\}} \pi_{\kappa}(\theta)$$

where $\pi_{\kappa}(\theta)$ is the von Mises distribution with pdf

$$f(\theta \mid \kappa, \mu) = \frac{e^{\kappa \cos(\theta - \mu)}}{2\pi I_0(\kappa)},$$

centered at $\theta \approx \mu$ close to zero (East) with concentration κ that might depend on V if the data support that.

PF Times

If we observe λ PFs per year of volume $V > \epsilon$, then what is the probability that such a PF will occur in the next 24 hours?

For short-term predictions, it may be important to note this can depend on:

- How high is the dome just now?
- Any seismic activity suggesting dome growth and instability?
- Has it rained recently?
- How long a time interval since last PF?

But, for long-term predictions all these factors average out and we assume that:

- PF occurrences in disjoint time intervals are independent
- PF rates are constant over time, neither rising nor falling.

PF Times

If we observe λ PFs per year of volume $V > \epsilon$, then what is the probability that such a PF will occur in the next 24 hours?
For short-term predictions, it may be important to note this can depend on:

- How high is the dome just now?
- Any seismic activity suggesting dome growth and instability?
- Has it rained recently?
- How long a time interval since last PF?

But, for long-term predictions all these factors average out and we assume that:

- PF occurrences in disjoint time intervals are independent
- PF rates are constant over time, neither rising nor falling.

PF Times

If we observe λ PFs per year of volume $V > \epsilon$, then what is the probability that such a PF will occur in the next 24 hours?
For short-term predictions, it may be important to note this can depend on:

- How high is the dome just now?
- Any seismic activity suggesting dome growth and instability?
- Has it rained recently?
- How long a time interval since last PF?

But, for long-term predictions all these factors average out and we assume that:

- PF occurrences in disjoint time intervals are independent
- PF rates are constant over time, neither rising nor falling.

PF Times

If we observe λ PFs per year of volume $V > \epsilon$, then what is the probability that such a PF will occur in the next 24 hours?
For short-term predictions, it may be important to note this can depend on:

- How high is the dome just now?
- Any seismic activity suggesting dome growth and instability?
- Has it rained recently?
- How long a time interval since last PF?

But, for long-term predictions all these factors average out and we assume that:

- PF occurrences in disjoint time intervals are independent
- PF rates are constant over time, neither rising nor falling.

PF Times (cont)

Under those assumptions (which we will revisit), the number of PFs in any time interval of length Δt (like 1/365, for a single day) has an expected number of $\lambda \Delta t$ PFs and a *probability* of

$$\text{P[At least one PF during time } \Delta t \text{]} = 1 - \exp(-\lambda \Delta t)$$

Or about $1 - e^{-22/365} \approx 5.8\%$ for one day with our data.

PF Times (cont)

Under those assumptions (which we will revisit), the number of PFs in any time interval of length Δt (like 1/365, for a single day) has an expected number of $\lambda \Delta t$ PFs and a *probability* of

$$\text{P[At least one PF during time } \Delta t \text{]} = 1 - \exp(-\lambda \Delta t)$$

Or about $1 - e^{-22/365} \approx 5.8\%$ for one day with our data.

A Summary of our Stochastic Model:

The data suggest a (provisional!) model in which:

1. PF Volumes are iid from a Pareto $\mathrm{Pa}(\alpha, \epsilon)$ distribution for some shape parameter $\alpha (\approx 0.64)$ and minimum flow ϵ $(\approx 5 \cdot 10^4\ \mathrm{m}^3)$; and
2. PF Initiation Angles have a von Mises distribution on $[0, 2\pi)$; and
3. PFs arrival times follow a stationary Poisson process at some rate λ (≈ 22)/yr.

These have the beautiful and simplifying consequence that the number of PFs in *any* region of three-dimensional $(V \times \theta \times \tau)$ space has a Poisson probability distribution— so, we can evaluate:

A Summary of our Stochastic Model:

The data suggest a (provisional!) model in which:

1. PF Volumes are iid from a Pareto $\mathrm{Pa}(\alpha, \epsilon)$ distribution for some shape parameter $\alpha (\approx 0.64)$ and minimum flow ϵ $(\approx 5 \cdot 10^4\,\mathrm{m}^3)$; and
2. PF Initiation Angles have a von Mises distribution on $[0, 2\pi)$; and
3. PFs arrival times follow a stationary Poisson process at some rate λ (≈ 22)/yr.

These have the beautiful and simplifying consequence that the number of PFs in *any* region of three-dimensional $(V \times \theta \times \tau)$ space has a Poisson probability distribution— so, we can evaluate:

A Summary of our Stochastic Model:

The data suggest a (provisional!) model in which:

1. PF Volumes are iid from a Pareto $\mathrm{Pa}(\alpha, \epsilon)$ distribution for some shape parameter $\alpha (\approx 0.64)$ and minimum flow ϵ $(\approx 5 \cdot 10^4\,\mathrm{m}^3)$; and
2. PF Initiation Angles have a von Mises distribution on $[0, 2\pi)$; and
3. PFs arrival times follow a stationary Poisson process at some rate λ $(\approx 22)/\mathrm{yr}$.

These have the beautiful and simplifying consequence that the number of PFs in *any* region of three-dimensional $(V \times \theta \times \tau)$ space has a Poisson probability distribution— so, we can evaluate:

A Summary of our Stochastic Model:

The data suggest a (provisional!) model in which:

1. PF Volumes are iid from a Pareto $\mathrm{Pa}(\alpha, \epsilon)$ distribution for some shape parameter $\alpha (\approx 0.64)$ and minimum flow ϵ $(\approx 5 \cdot 10^4\ \mathrm{m}^3)$; and
2. PF Initiation Angles have a von Mises distribution on $[0, 2\pi)$; and
3. PFs arrival times follow a stationary Poisson process at some rate λ $(\approx 22)/\mathrm{yr}$.

These have the beautiful and simplifying consequence that the number of PFs in *any* region of three-dimensional $(V \times \theta \times \tau)$ space has a Poisson probability distribution— so, we can evaluate:

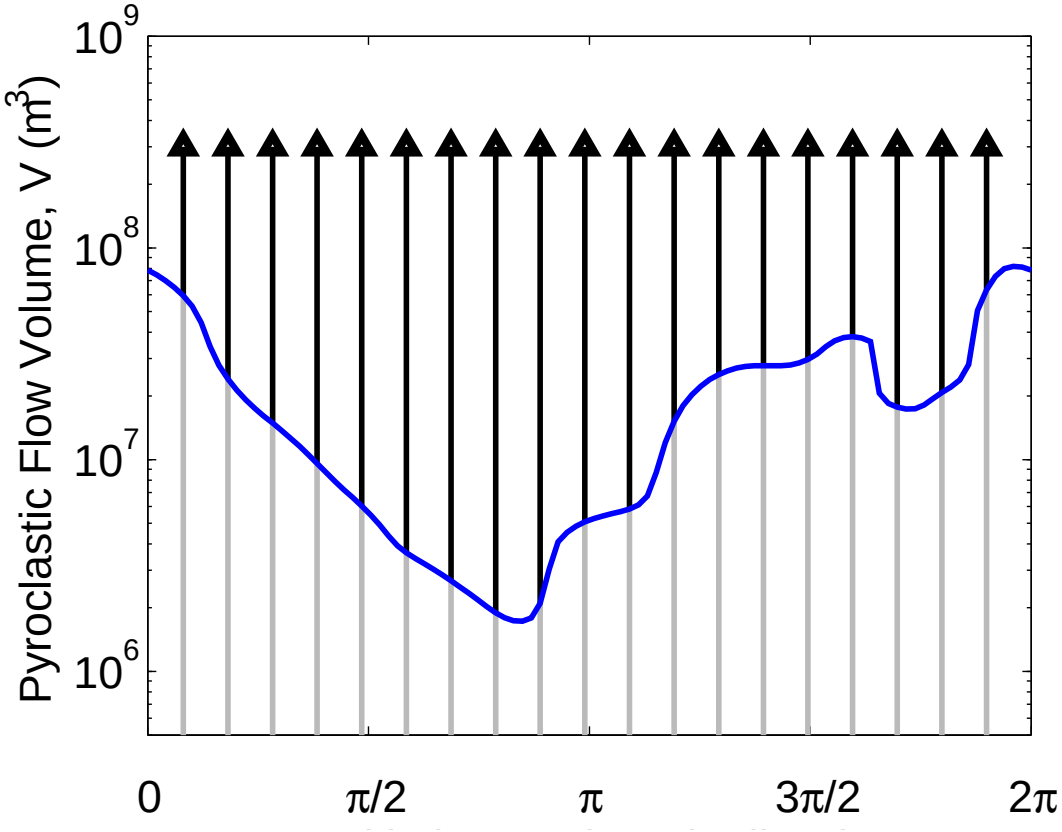
Hazard (in formulas)

The hazard prediction we wish to make is

$$
\begin{aligned}
&\mathrm{P}[\text{ Catastrophy within } t \text{ Years }] \\
&\quad = 1 - \mathrm{P}[Y_t = 0] \qquad \text{(where } Y_t \text{ is the } \textit{number} \text{ of catastrophes)} \\
&\quad = 1 - \exp\big(-\mathrm{E}\,Y_t\big) \\
&\quad = 1 - \exp\left(-t\,\lambda_\epsilon \int_{V > \Psi(\theta;x)} \epsilon^\alpha \alpha V^{-\alpha-1}\, dV\, \pi_\kappa(d\theta)\right)
\end{aligned}
$$

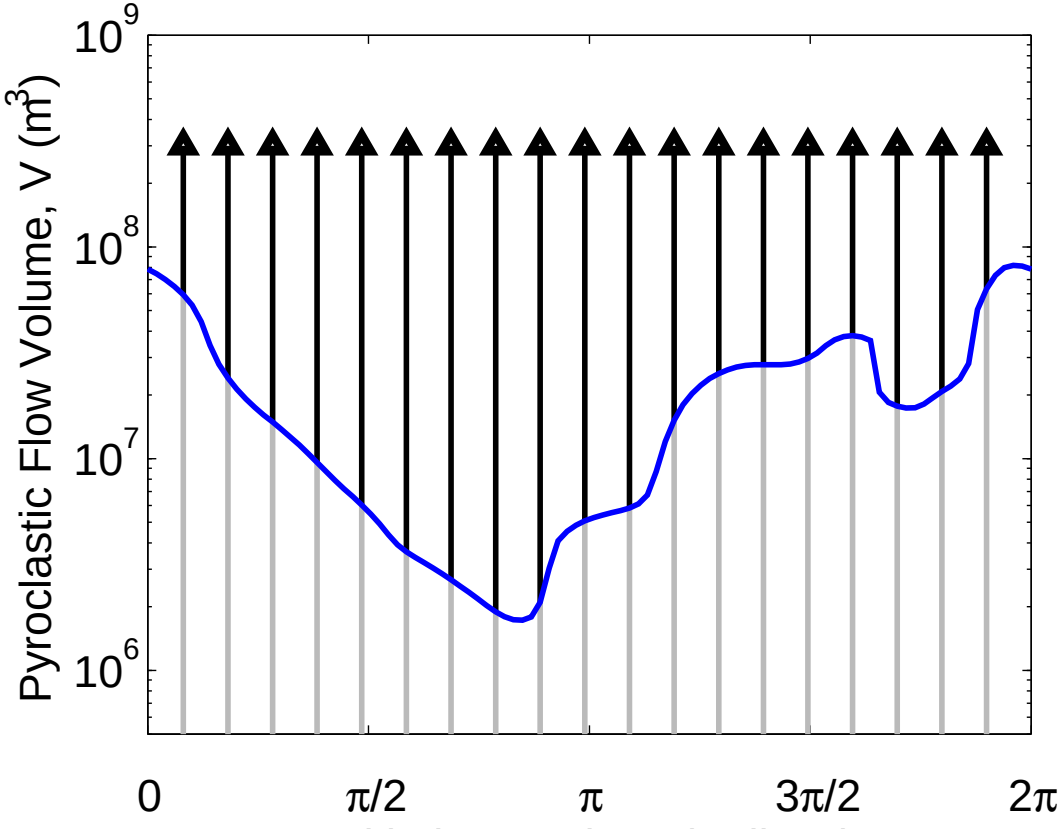
Which we can compute pretty easily on a computer:

Hazard (in pictures)



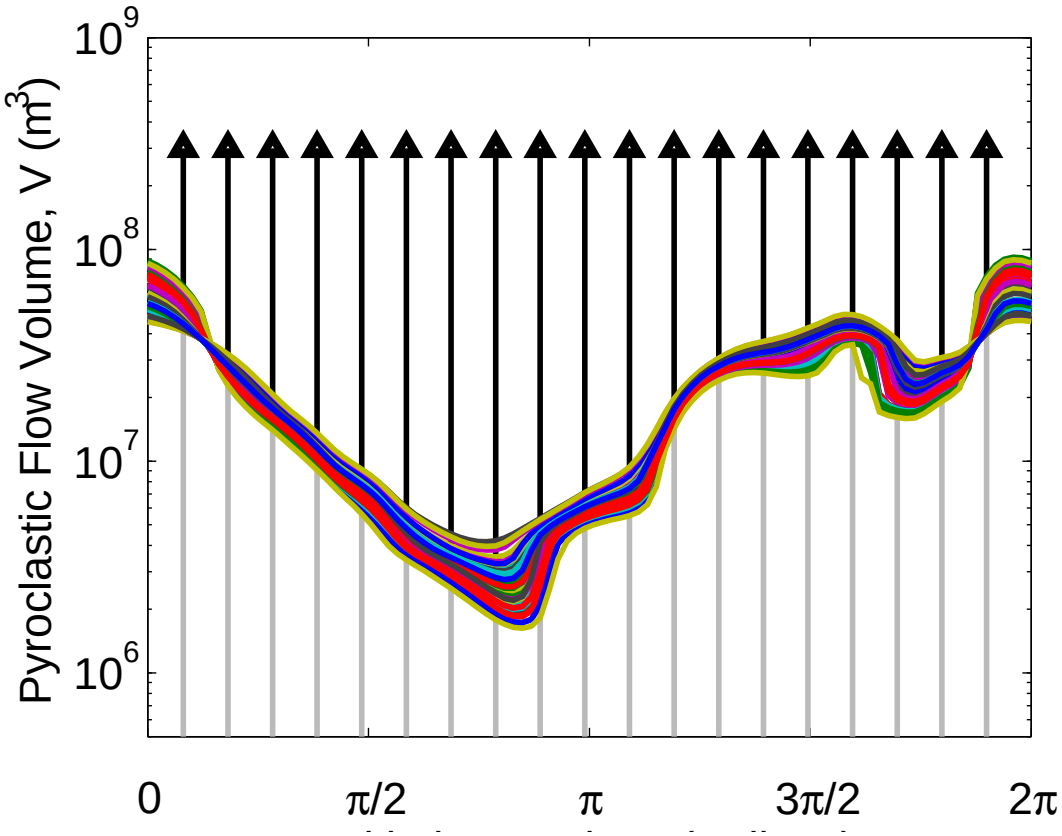
If we knew the (uncertain) values of the Pareto parameter $\alpha \approx 0.64$, rate parameter $\lambda_\epsilon \approx 22/\text{yr}$, basal friction $\phi \approx 8$–11°.

Hazard (in pictures)



If we knew the (uncertain) values of the Pareto parameter $\alpha \approx 0.64$, rate parameter $\lambda_\epsilon \approx 22/\text{yr}$, basal friction $\phi \approx 8$–$11°$.

Hazard (in pictures)



The necessary integrals in V and λ_ϵ are available in closed form; we accommodate (posterior uncertainty) about α and ϕ using MCIS.

Computing the Posterior

P[Catastrophy within t Years $\mid \lambda_\epsilon,\ \alpha,\ \phi$]

$$= 1 - \exp\left(-t\,\lambda_\epsilon \int_{V>\Psi(\theta;x)} \epsilon^\alpha \alpha V^{-\alpha-1}\, dV\, \pi_\kappa(d\theta)\right)$$

$$= 1 - e^{-t\,\lambda_\epsilon I_\epsilon(\alpha,\phi)} \quad \text{where} \quad I_\epsilon(\alpha,\phi) = \int_0^{2\pi} \{\Psi(\theta;x)/\epsilon\}^{-\alpha}\ \pi_\kappa(d\theta)$$

P[No Catastrophy $\mid$ Data]

$$= Z^{-1} \int e^{-t\,\lambda_\epsilon I_\epsilon(\alpha,\phi)}\, L(\alpha,\lambda,\phi)\, \pi(d\lambda\, d\alpha\, d\phi)$$

$$= Z^{-1} \int \left\{1 + \frac{t}{Tb} I_\epsilon(\alpha,\phi)\right\}^{-J_\epsilon - a} L(\alpha,\phi)\, \pi(d\alpha\, d\phi)$$

$$\approx \frac{1}{T} \sum \left\{1 + \frac{t}{Tb} I(\alpha;\phi)\right\}^{-J_\epsilon - a}$$

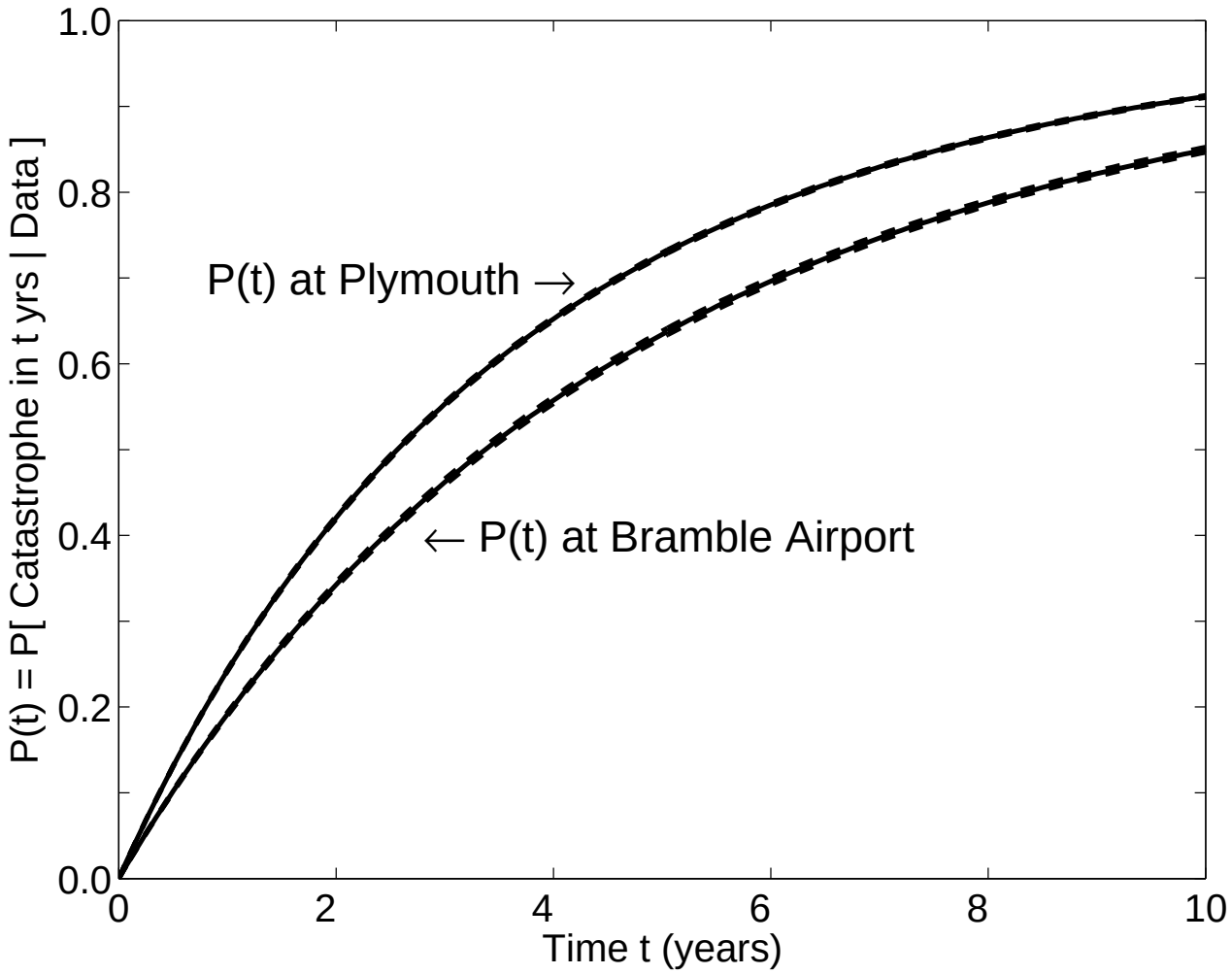
Computing the Posterior

$$
\begin{aligned}
&\text{P[No Catastrophy } \mid \text{ Data]} \\
&\qquad = Z^{-1} \int e^{-t\,\lambda_\epsilon I_\epsilon(\alpha,\phi)}\, L(\alpha,\lambda,\phi)\,\pi(d\lambda\, d\alpha\, d\phi) \\
&\qquad = Z^{-1} \int \left\{1 + \frac{t}{Tb} I_\epsilon(\alpha,\phi)\right\}^{-J_\epsilon - a} L(\alpha,\phi)\,\pi(d\alpha\, d\phi) \\
&\qquad \approx \frac{1}{N} \sum \left\{1 + \frac{t}{Tb} I_\epsilon(\alpha_i,\phi_i)\right\}^{-J_\epsilon - a}.
\end{aligned}
$$

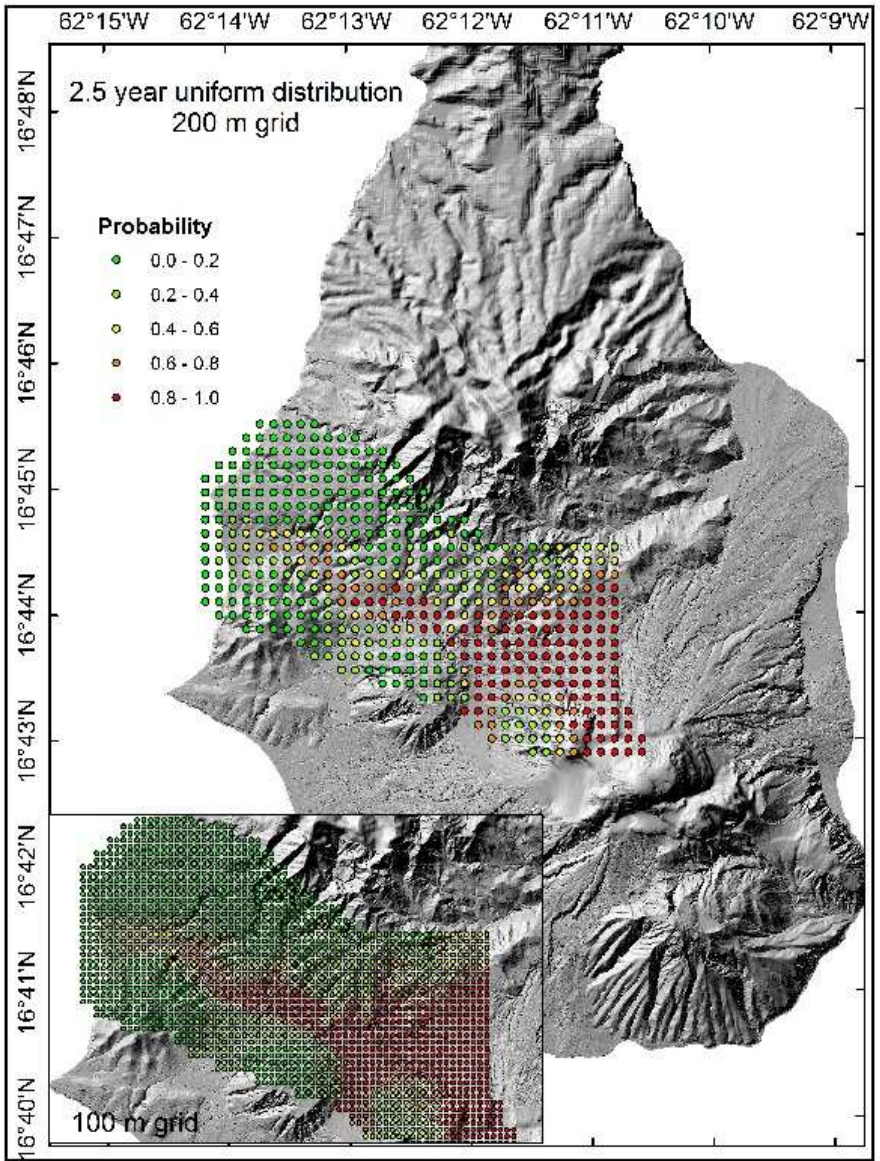
Computing the Posterior

$$
\begin{aligned}
&\text{P[No Catastrophy } \mid \text{ Data]} \\
&\qquad = Z^{-1} \int e^{-t\,\lambda_\epsilon I_\epsilon(\alpha,\phi)}\, L(\alpha,\lambda,\phi)\,\pi(d\lambda\, d\alpha\, d\phi) \\
&\qquad = Z^{-1} \int \left\{1 + \frac{t}{Tb} I_\epsilon(\alpha,\phi)\right\}^{-J_\epsilon - a} L(\alpha,\phi)\,\pi(d\alpha\, d\phi) \\
&\qquad \approx \frac{1}{N} \sum \left\{1 + \frac{t}{Tb} I_\epsilon(\alpha_i,\phi_i)\right\}^{-J_\epsilon - a}.
\end{aligned}
$$

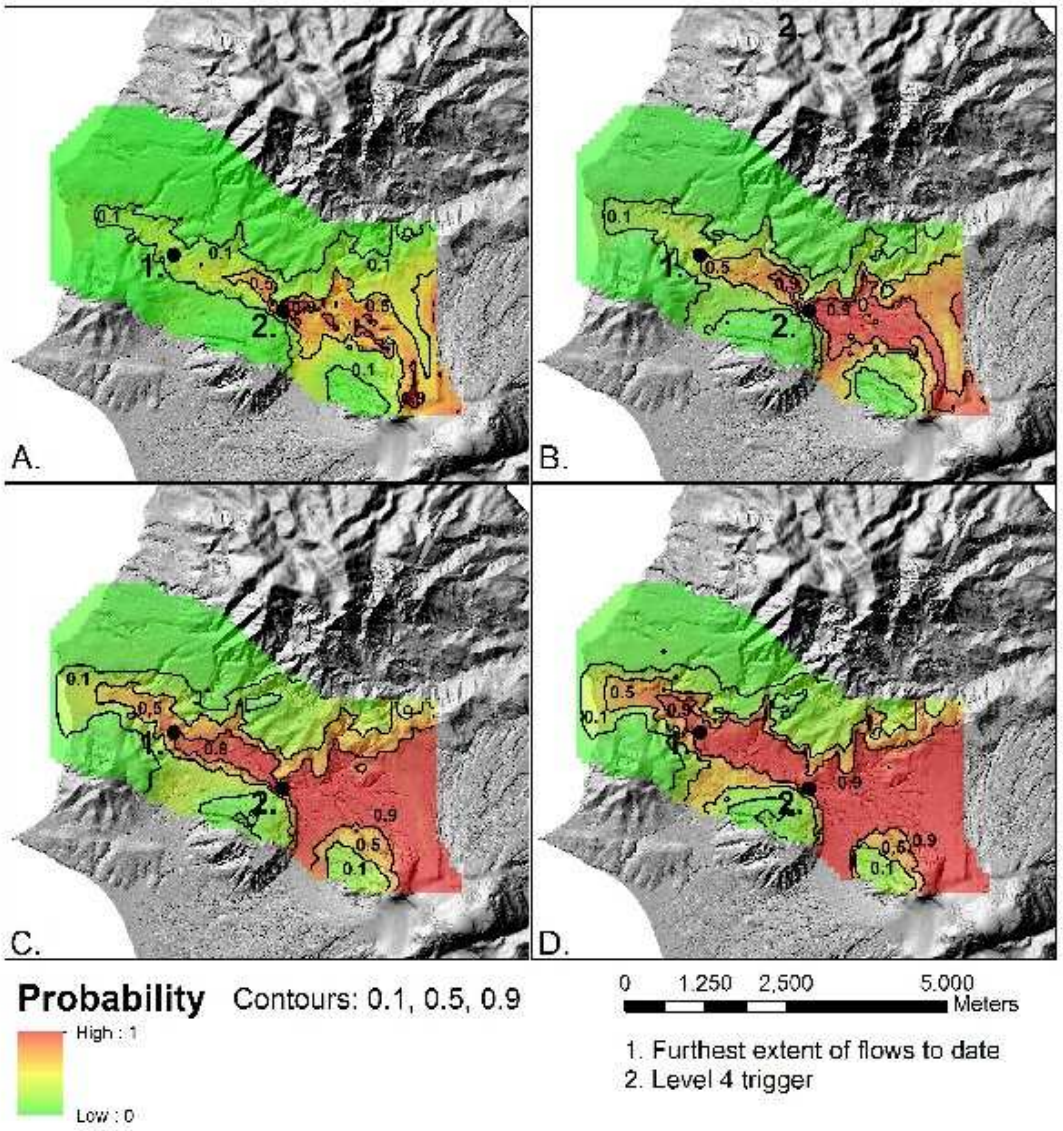
Hazard Over Time at Plymouth & Bramble



Hazard Over Space, for 2.5 Yrs



A few more Hazard Maps



Current Official Hazard Map...

Extensions & Related matters...

Some challenges along the way:

- TITAN-2D relies on Digital Elevation Models for its flow predictions. How can we quantify uncertainty in the DEM and its effect on hazard?
- Our team is building "elastic" models, in collaboration with UB geologist Beá Csathó;
- MVO is a spectacular and rare resource, giving unrivaled wealth of data for SHV. How can we use Seismic Data at other volcanoes to improve hazard prediction?
- We are building models to link seismic & volcanic activity.

Extensions & Related matters...

Some challenges along the way:

- TITAN-2D relies on Digital Elevation Models for its flow predictions. How can we quantify uncertainty in the DEM and its effect on hazard?
- Our team is building "elastic" models, in collaboration with UB geologist Beá Csathó;
- MVO is a spectacular and rare resource, giving unrivaled wealth of data for SHV. How can we use Seismic Data at other volcanoes to improve hazard prediction?
- We are building models to link seismic & volcanic activity.

Extensions & Related matters...

Some challenges along the way:

- TITAN-2D relies on Digital Elevation Models for its flow predictions. How can we quantify uncertainty in the DEM and its effect on hazard?
- Our team is building "elastic" models, in collaboration with UB geologist Beá Csathó;
- MVO is a spectacular and rare resource, giving unrivaled wealth of data for SHV. How can we use Seismic Data at other volcanoes to improve hazard prediction?
- We are building models to link seismic & volcanic activity.

Extensions & Related matters...

Some challenges along the way:

- TITAN-2D relies on Digital Elevation Models for its flow predictions. How can we quantify uncertainty in the DEM and its effect on hazard?
- Our team is building "elastic" models, in collaboration with UB geologist Beá Csathó;
- MVO is a spectacular and rare resource, giving unrivaled wealth of data for SHV. How can we use Seismic Data at other volcanoes to improve hazard prediction?
- We are building models to link seismic & volcanic activity.

Wrap Up

Model Improvements Underway

All three of our Preliminary Assumptions are wrong, of course:

1. PFs arrive as stationary Poisson proc. at rate $\lambda(\approx 22)/\text{yr}$;

In fact the rate λ_t varies over time, and sometimes seems to vanish for a while at times. We're building dynamic models for this, which feature possibility of $\lambda_t = 0$ for some time intervals.

No cyclicity yet (7–11 hr, 6–7 wk, *etc.*),
No dynamic changes in θ dist'n,
No Survival Curve (yet!).

Model Improvements Underway

All three of our Preliminary Assumptions are wrong, of course:

1. PFs arrive as stationary Poisson proc. at rate $\lambda(\approx 22)/\text{yr}$;

In fact the rate λ_t varies over time, and sometimes seems to vanish for a while at times. We're building dynamic models for this, which feature possibility of $\lambda_t = 0$ for some time intervals.

No cyclicity yet (7–11 hr, 6–7 wk, *etc.*),
No dynamic changes in θ dist'n,
No Survival Curve (yet!).

Model Improvements Underway

All three of our Preliminary Assumptions are wrong, of course:

1. PFs arrive as stationary Poisson proc. at rate $\lambda(\approx 22)/\text{yr}$;

In fact the rate λ_t varies over time, and sometimes seems to vanish for a while at times. We're building dynamic models for this, which feature possibility of $\lambda_t = 0$ for some time intervals.

No cyclicity yet (7–11 hr, 6–7 wk, *etc.*),
No dynamic changes in θ dist'n,
No Survival Curve (yet!).

Model Improvements Underway

All three of our Preliminary Assumptions are wrong, of course:

1. PFs arrive as stationary Poisson proc. at rate $\lambda(\approx 22)/\text{yr}$;

In fact the rate λ_t varies over time, and sometimes seems to vanish for a while at times. We're building dynamic models for this, which feature possibility of $\lambda_t = 0$ for some time intervals.

No cyclicity yet (7–11 hr, 6–7 wk, *etc.*),
No dynamic changes in θ dist'n,
No Survival Curve (yet!).

Model Improvements Underway (pt 2)

2. PF Volumes are iid from a Pareto $\mathrm{Pa}(\alpha, \epsilon)$ distribution for shape parameter α (≈ 0.64) and min flow ϵ ($\approx 5 \cdot 10^4\ \mathrm{m}^3$);

At extremely high volumes the Pareto has unrealistic tails, suggesting a 0.1% chance that Sufrière Hills would have flow exceeding $10^{12}\mathrm{m}^3$ within a century.

Nobody believes that; we're building tempered Pareto models for more realistic extreme tails for V.

Model Improvements Underway (pt 2)

2. PF Volumes are iid from a Pareto $\mathrm{Pa}(\alpha, \epsilon)$ distribution for shape parameter α (≈ 0.64) and min flow ϵ ($\approx 5 \cdot 10^4\ \mathrm{m}^3$);

At extremely high volumes the Pareto has unrealistic tails, suggesting a 0.1% chance that Sufrière Hills would have flow exceeding $10^{12}\mathrm{m}^3$ within a century.

Nobody believes that; we're building tempered Pareto models for more realistic extreme tails for V.

Model Improvements Underway (pt 2)

2. PF Volumes are iid from a Pareto $\mathrm{Pa}(\alpha, \epsilon)$ distribution for shape parameter α (≈ 0.64) and min flow ϵ ($\approx 5 \cdot 10^4\ \mathrm{m}^3$);

At extremely high volumes the Pareto has unrealistic tails, suggesting a 0.1% chance that Sufrière Hills would have flow exceeding $10^{12}\mathrm{m}^3$ within a century.

Nobody believes that; we're building tempered Pareto models for more realistic extreme tails for V.

Model Improvements Underway (pt 3)

3. PF Initiation Angles are distributed uniform on $[0, 2\pi)$ (or maybe w/unimodal von Mises dist), independently of V.

We've already seen signs that easterly flows are more frequent, and suggestions that V and θ may be dependent. Can extend to mixture of von Mises, regression model, *etc.*

Probably I'm out of time now, and it's time to Wrap Up; if not, I'll show how we're approaching one of these issues:

Model Improvements Underway (pt 3)

3. PF Initiation Angles are distributed uniform on $[0, 2\pi)$ (or maybe w/unimodal von Mises dist), independently of V.

We've already seen signs that easterly flows are more frequent, and suggestions that V and θ may be dependent. Can extend to mixture of von Mises, regression model, *etc.*

Probably I'm out of time now, and it's time to Wrap Up; if not, I'll show how we're approaching one of these issues:

Model Improvements Underway (pt 3)

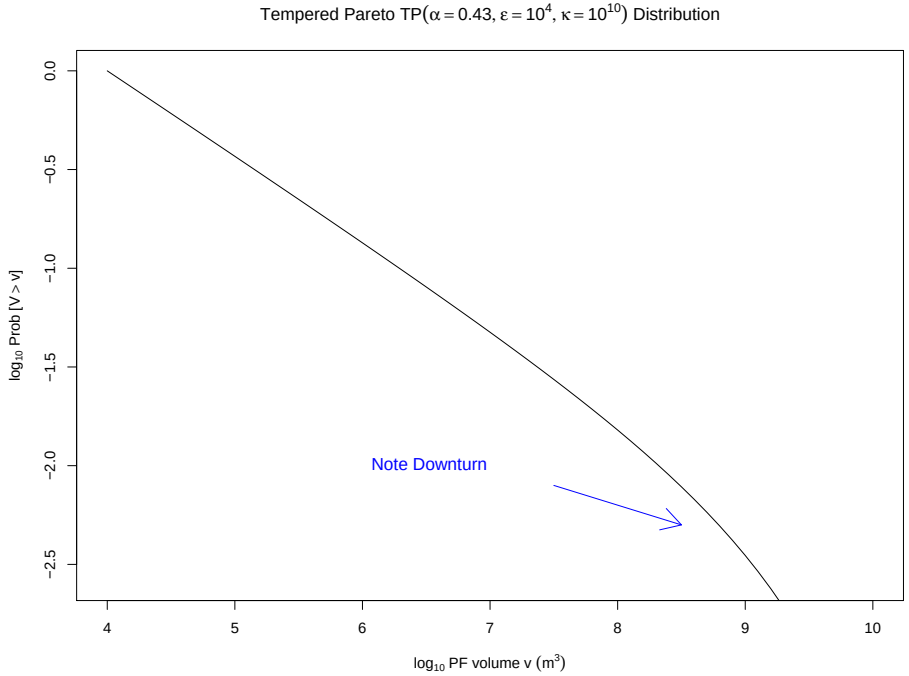
3. PF Initiation Angles are distributed uniform on $[0, 2\pi)$ (or maybe w/unimodal von Mises dist), independently of V.

We've already seen signs that easterly flows are more frequent, and suggestions that V and θ may be dependent. Can extend to mixture of von Mises, regression model, *etc.*
Probably I'm out of time now, and it's time to Wrap Up; if not, I'll show how we're approaching one of these issues:

Tempered Pareto Models

There's a slight suggestion of a downturn in the Freq-vs-Vol log-log plot; perhaps a better model is Tempered Pareto, with

$$P[V > v] = (v/\epsilon)^{-\alpha}\, e^{-\beta(v-\epsilon)}, \qquad v > \epsilon :$$

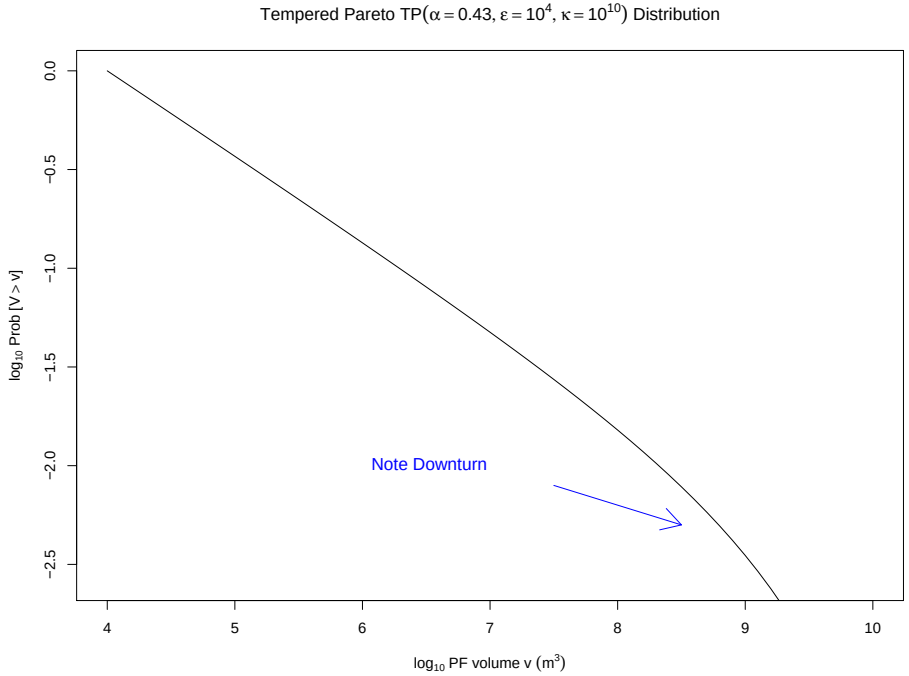


See the drop-off near 10^8 m^3 or 10^9 m^3? That's Good news!

Tempered Pareto Models

There's a slight suggestion of a downturn in the Freq-vs-Vol log-log plot; perhaps a better model is Tempered Pareto, with

$$P[V > v] = (v/\epsilon)^{-\alpha}\, e^{-\beta(v-\epsilon)}, \qquad v > \epsilon :$$



See the drop-off near 10^8 m^3 or 10^9 m^3? That's Good news!

Uncertainty

MLE and reference Bayesian posterior means ($\pm$ SD's) for the three parameters of this model are:

Shape	$\alpha \approx 0.64$	± 0.04	
Rate	$\lambda \approx 22.42$	± 1.3	/year
Bend	$\beta \approx 9.22 \times 10^{-9}$	$\pm 6.0 \times 10^{-9}$	/ m^3

Note how uncertain β is... only 1.5 SD above zero.
How does this affect hazard predictions?
How does Pareto ($\beta = 0$) model differ from
Tempered Pareto ($\beta > 0$)?

Uncertainty

MLE and reference Bayesian posterior means ($\pm$ SD's) for the three parameters of this model are:

Shape	$\alpha \approx 0.64$	± 0.04	
Rate	$\lambda \approx 22.42$	± 1.3	/year
Bend	$\beta \approx 9.22 \times 10^{-9}$	$\pm 6.0 \times 10^{-9}$	/ m^3

Note how uncertain β is... only 1.5 SD above zero.

How does this affect hazard predictions?

How does Pareto ($\beta = 0$) model differ from Tempered Pareto ($\beta > 0$)?

Uncertainty

MLE and reference Bayesian posterior means ($\pm$ SD's) for the three parameters of this model are:

Shape	$\alpha \approx 0.64$	± 0.04	
Rate	$\lambda \approx 22.42$	± 1.3	/year
Bend	$\beta \approx 9.22 \times 10^{-9}$	$\pm 6.0 \times 10^{-9}$	/ m^3

Note how uncertain β is... only 1.5 SD above zero.
How does this affect hazard predictions?
How does Pareto ($\beta = 0$) model differ from Tempered Pareto ($\beta > 0$)?

Uncertainty

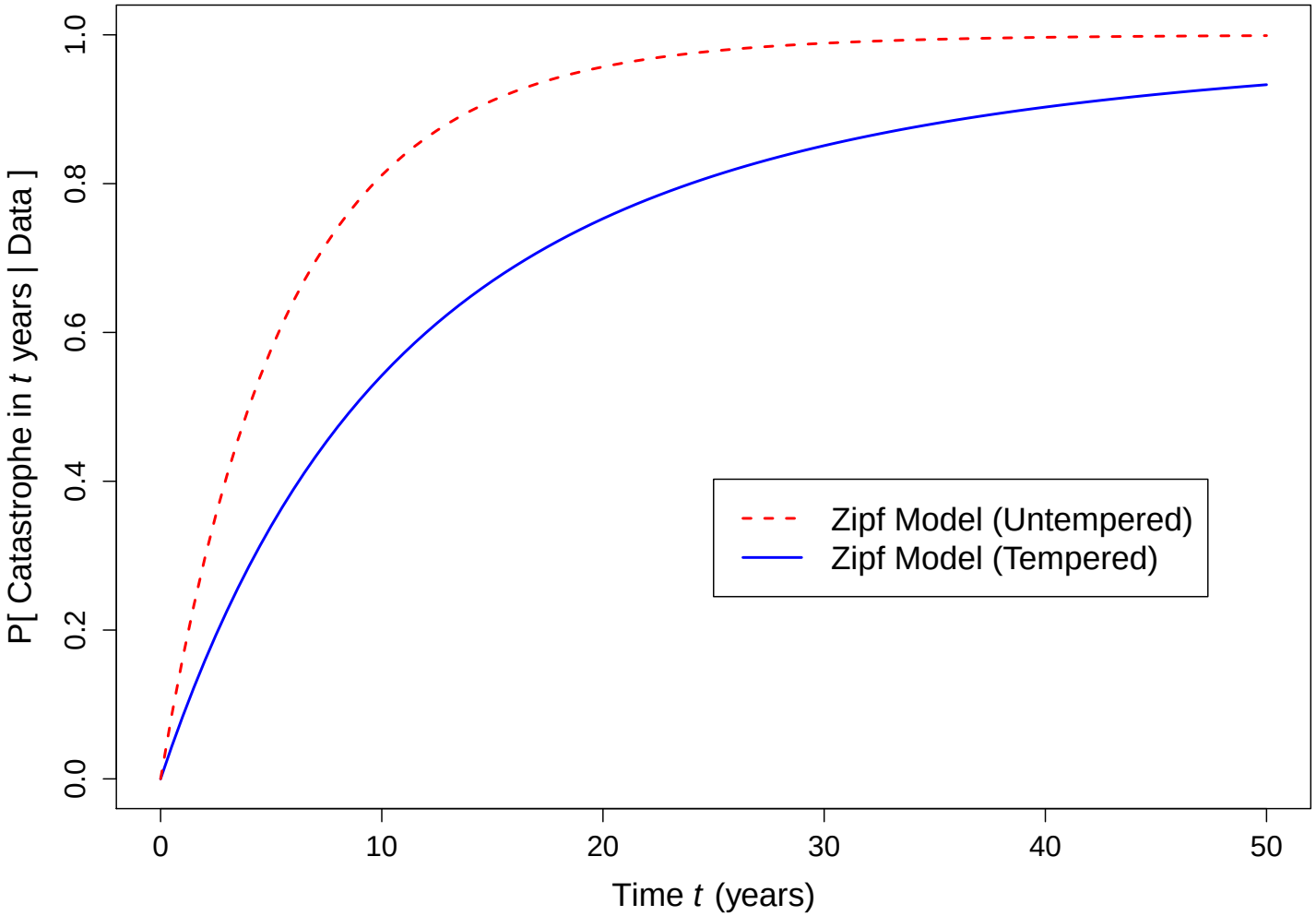
MLE and reference Bayesian posterior means ($\pm$ SD's) for the three parameters of this model are:

Shape	$\alpha \approx 0.64$	± 0.04	
Rate	$\lambda \approx 22.42$	± 1.3	/year
Bend	$\beta \approx 9.22 \times 10^{-9}$	$\pm 6.0 \times 10^{-9}$	/ m^3

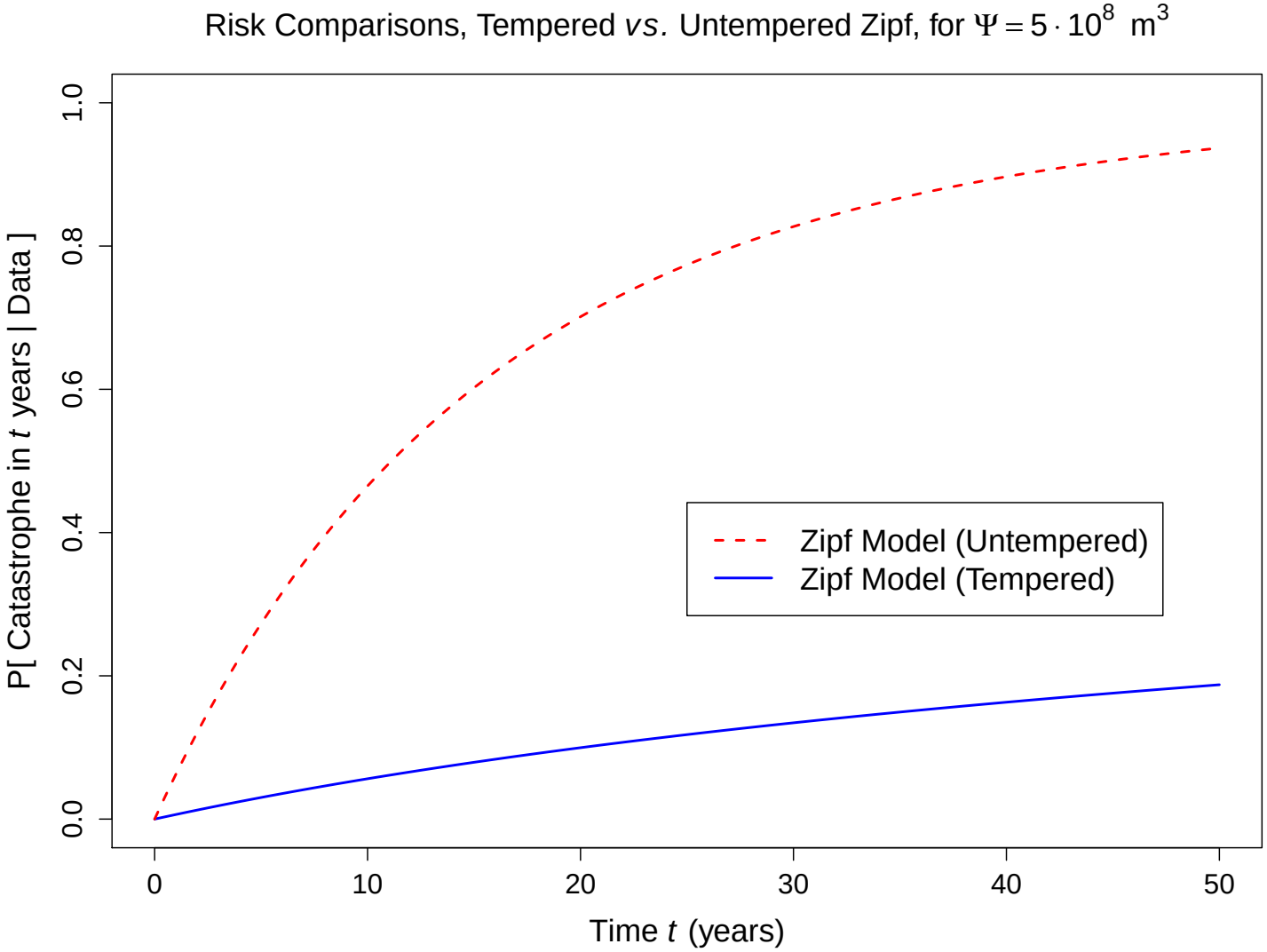
Note how uncertain β is... only 1.5 SD above zero.
How does this affect hazard predictions?
How does Pareto ($\beta = 0$) model differ from
Tempered Pareto ($\beta > 0$)?

Hazard for Fifty Years of $V > 10^8\,\mathrm{m}^3$

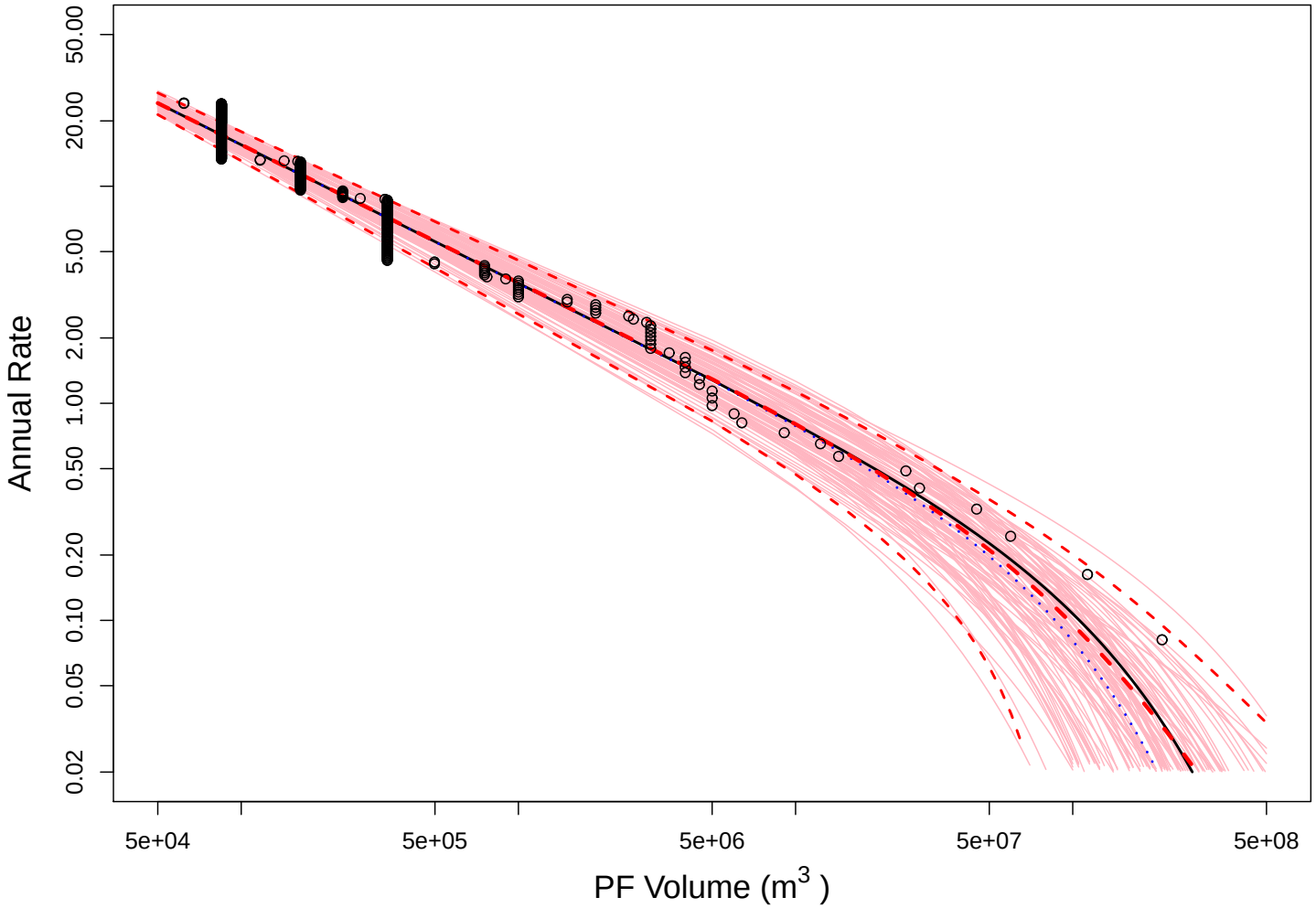
Risk Comparisons, Tempered *vs.* Untempered Zipf, for $\Psi = 10^8\ \mathrm{m}^3$



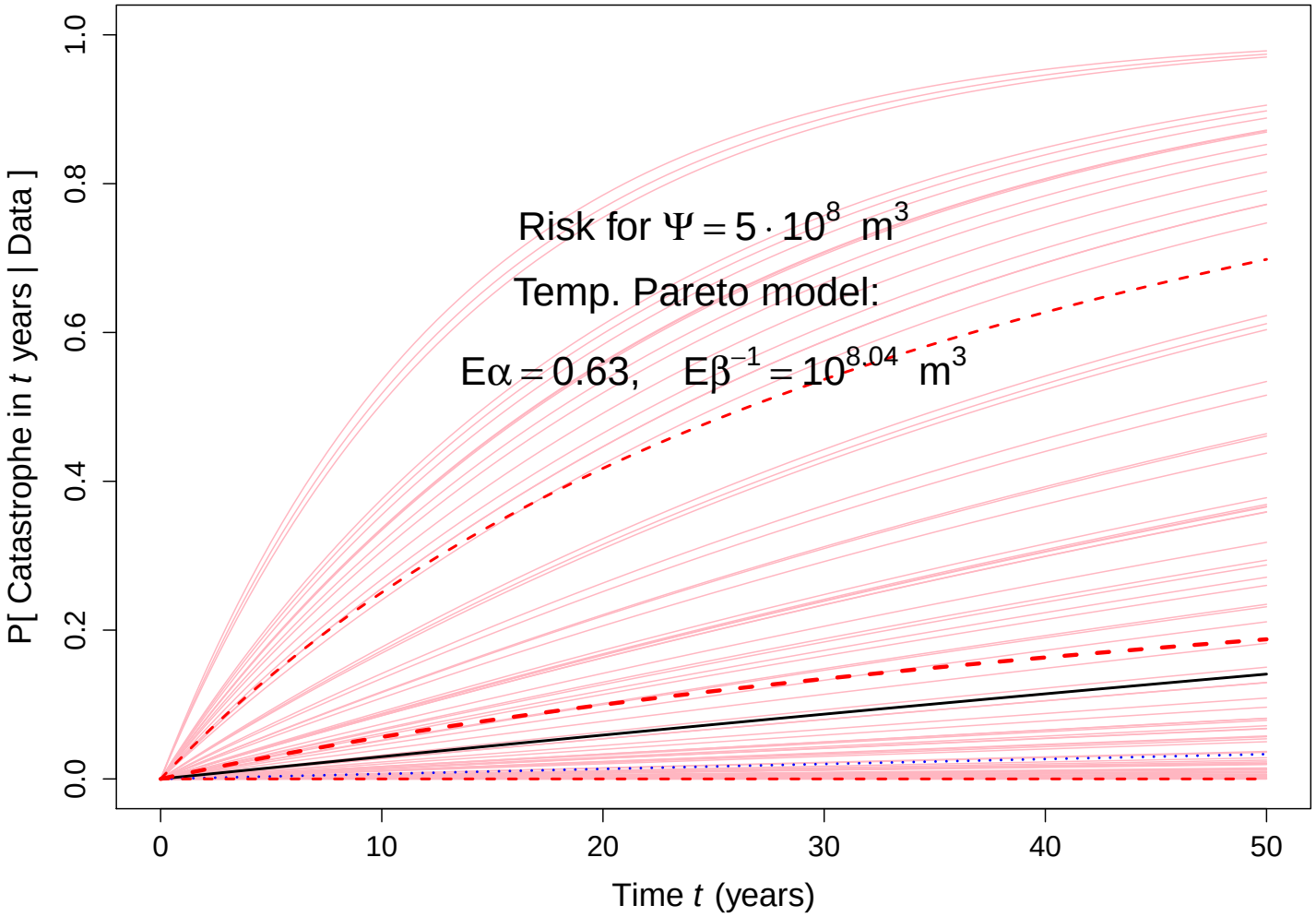
Hazard for Fifty Years of $V > 5 \times 10^8\ \mathrm{m}^3$



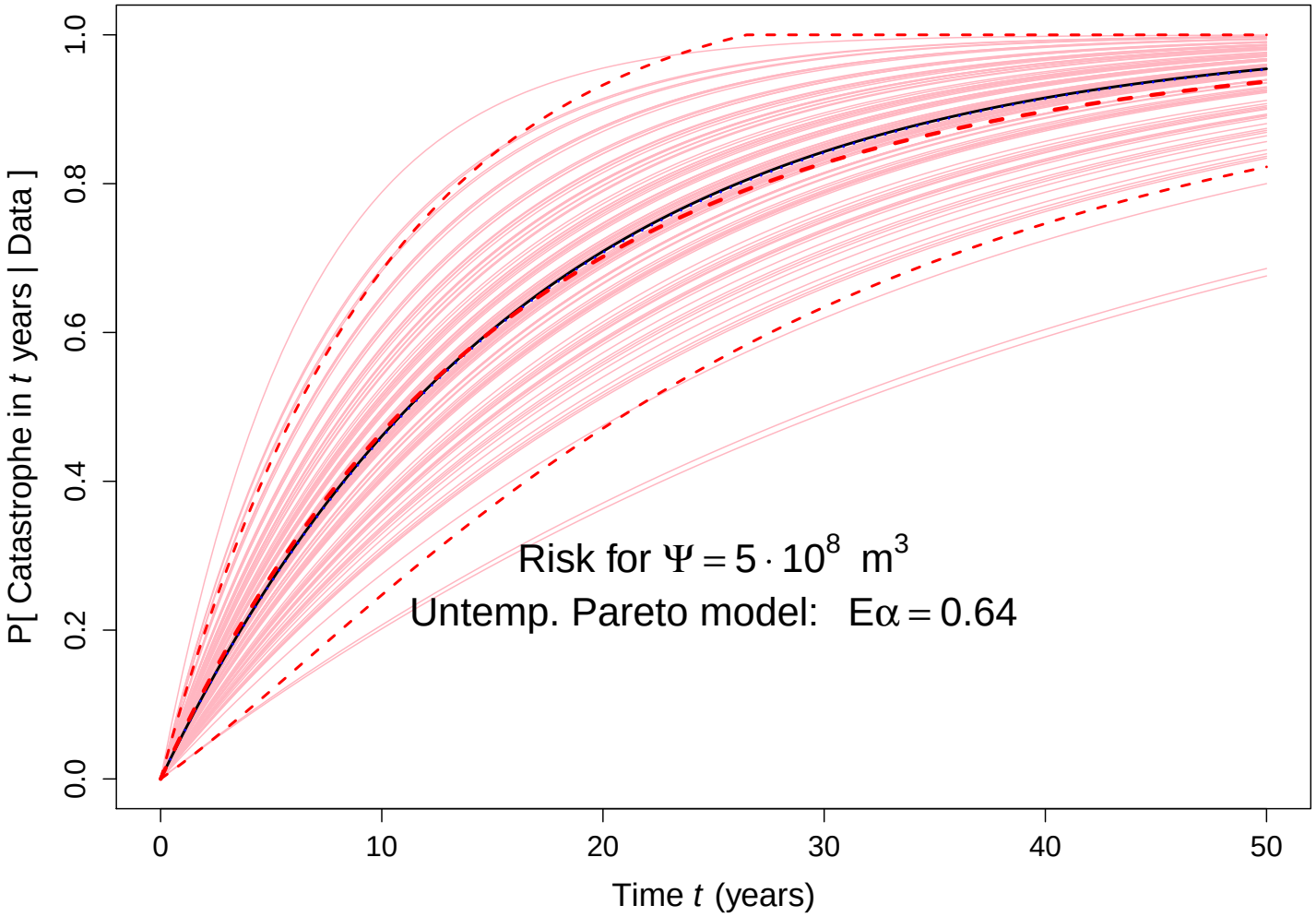
Why the wide uncertainty?



How does that affect estimated hazard?

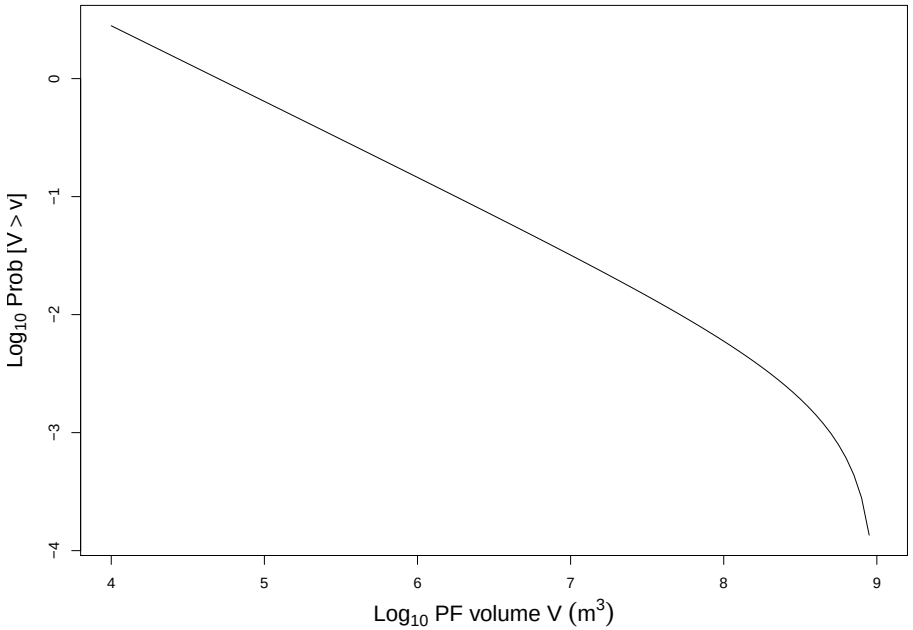


What if there is no tempering?



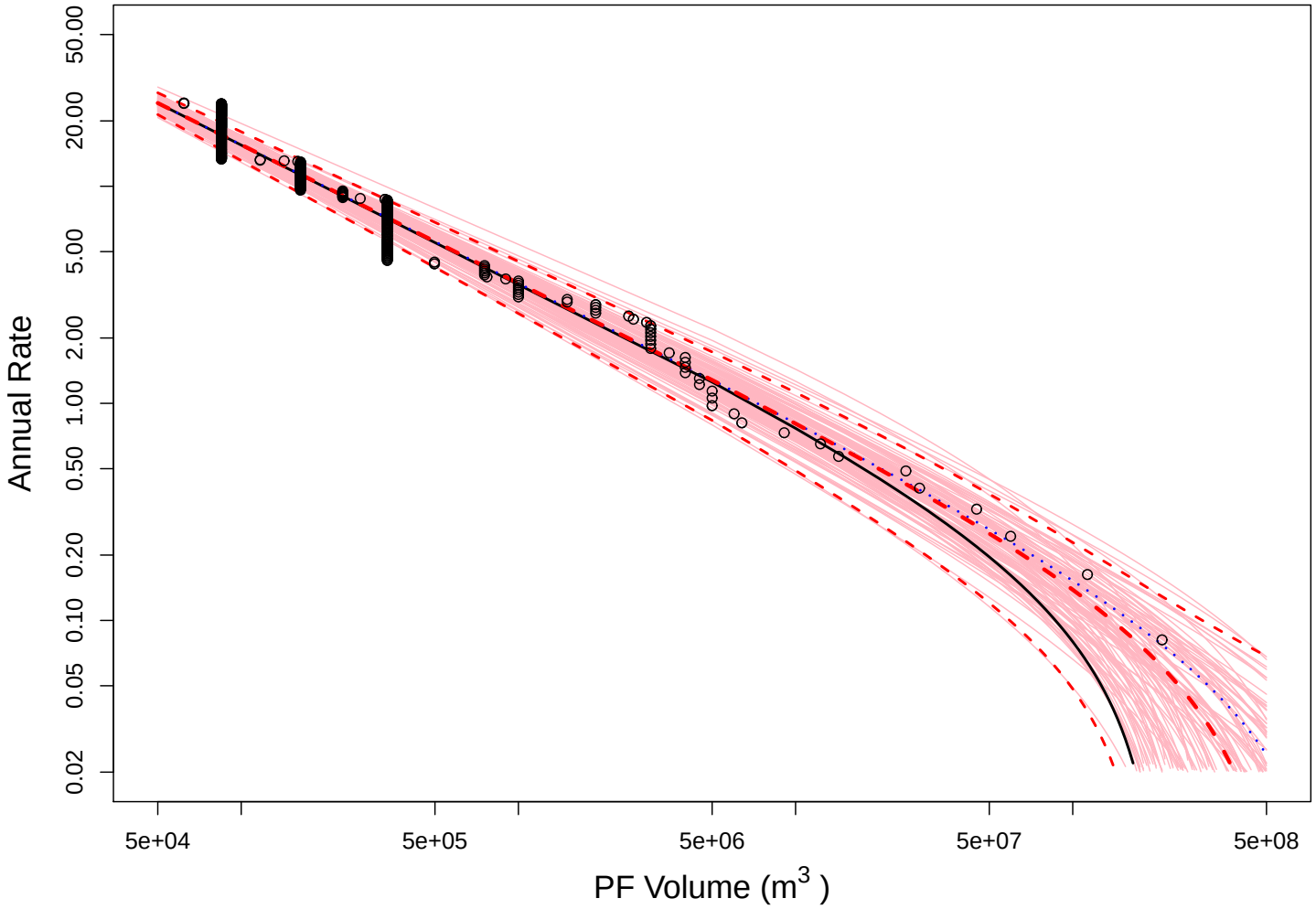
A little more extreme— Truncated Pareto

$$\mathrm{P}[V > v] = \frac{v^{-\alpha} - L^{-\alpha}}{\epsilon^{-\alpha} - L^{-\alpha}} \qquad \epsilon \leq v \leq L:$$

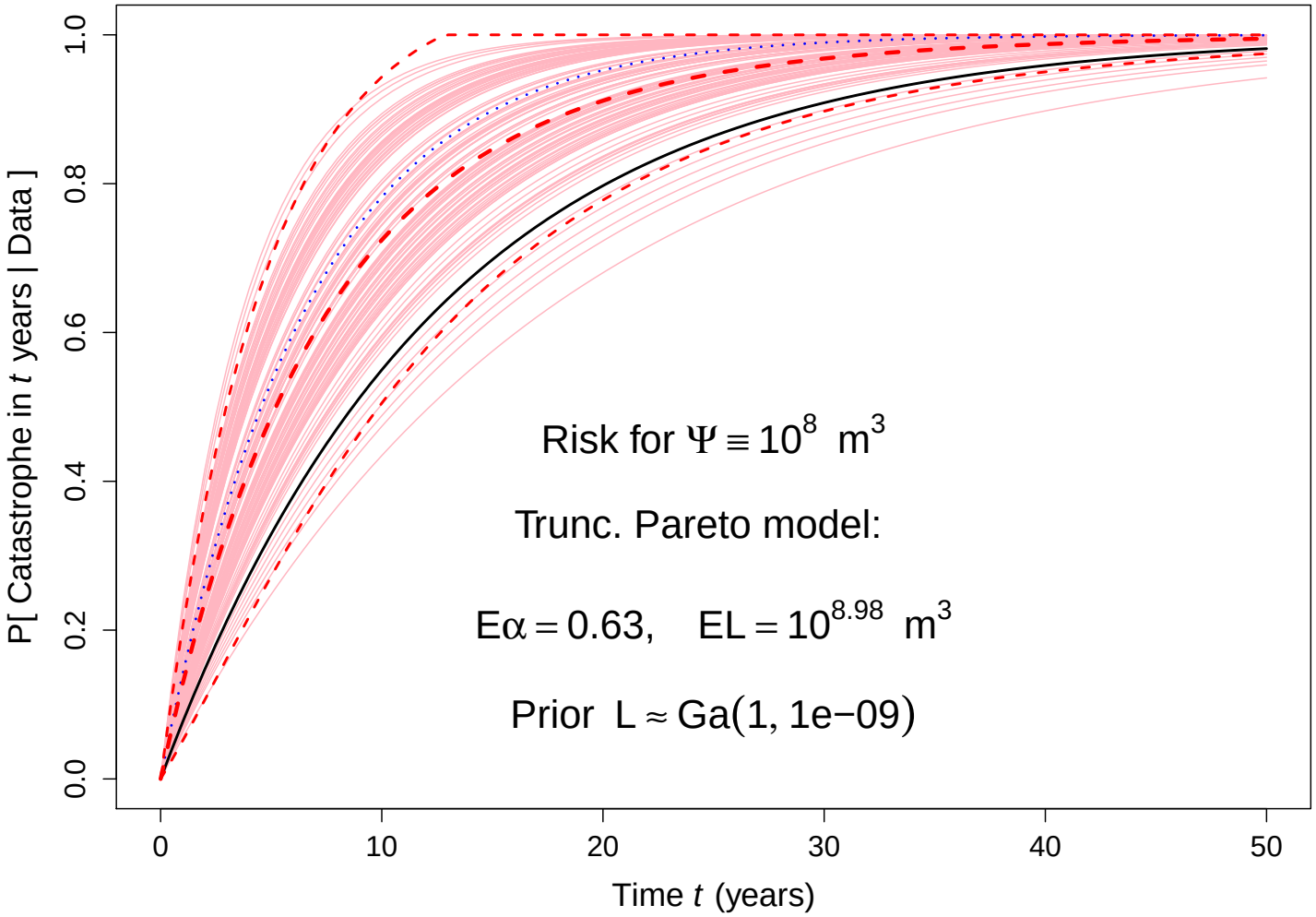


Now the drop-off approaching 10^9 m^3 is fast!

A little more extreme— Truncated Pareto



A little more extreme— Truncated Pareto



Morals

- It's hard to tell about tail behaviour (extremes) from data;
- Clearer picture of tempering would help—
  working to elicit expert opinion about physical limits
- All our model assumptions benefit from closer scrutiny.

Morals

- It's hard to tell about tail behaviour (extremes) from data;
- Clearer picture of tempering would help—
  working to elicit expert opinion about physical limits
- All our model assumptions benefit from closer scrutiny.

Morals

- It's hard to tell about tail behaviour (extremes) from data;
- Clearer picture of tempering would help—
  working to elicit expert opinion about physical limits
- All our model assumptions benefit from closer scrutiny.

Morals

- It's hard to tell about tail behaviour (extremes) from data;
- Clearer picture of tempering would help—
  working to elicit expert opinion about physical limits
- All our model assumptions benefit from closer scrutiny.

Conclusions

- Probability, Statistics, Computational PDE, and physical Sciences (Geo, Phys, Ast, Bio, Chem, ...) are a good mix!
- Useful to Separate the Stochastic & Deterministic aspects
  - Use Deterministic models to identify region $\mathcal{R}_C$ of input parameter space for Catastrophes
  - Use Stochastic models to identify posterior probability

$$\mathrm{P}[\mathcal{R}_C \mid \text{ Data }]$$

Ongoing Work & Extensions

- Studying dynamic models for variable rate of eruption process and changing direction of flows ($\Rightarrow$ short-term predictions);
- Exploring dependence between angle θ and flow volume V;
- Estimating possible "tempered Pareto" models,

$$\mathrm{P}[V > v] = (v/\epsilon)^{-\alpha} e^{-\beta(v-\epsilon)}, \qquad v > \epsilon$$

- Non-stationary variants;
- Finite lifetimes ("dormancy" and "extinction").

Thanks!

More details (references, this talk in .pdf, related work) are available on request from

rlw@duke.edu.

Thanks to SAMSI, SIAM, MVO and to the NSF[1]!

[1] ACI-1550225, DMS-0757549, DMS-1228317, DMS-1622403, SES-1521855.