Statistical Analysis of Distributed Volcanic Fields

Probabilistic Models of Volcanic Hazards Associated with Predominantly Monogenetic Volcanic Fields

PASI Workshop

7 May, 2013

Chuck Connor
Department of Geology, University of South Florida

Distributed volcanic fields

The Holocene Karqkar volcanic field, Armenia and Harat Khaybar, Saudi Arabia

- Renewed volcanic activity occurs from new vents, resulting in formation of tens to hundreds of vents over time

Harat Khaybar, Saudi Arabia, from NASA Earthobservatory.org

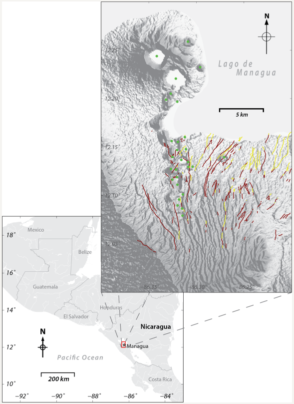
Distributed volcanic fields

Ararat volcano and the distributed volcanoes of the Yerevan basin, Armenia

- Distributed volcanic fields may or may not be associated with a larger volcanic system, such as calderas or composite volcanoes

Distributed volcanic fields

Nejapa – Apoyeque Alignment

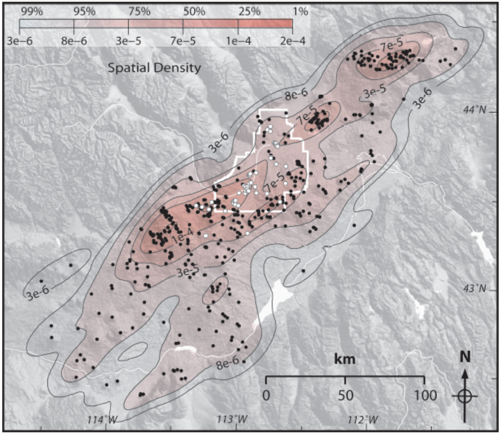


- The location, area, and shape of volcanic fields is often directly related to their tectonic settings

photo from INETER

Distributed volcanic fields

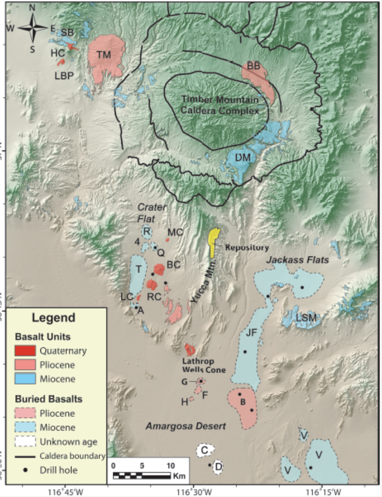
Eastern Snake River plain, USA



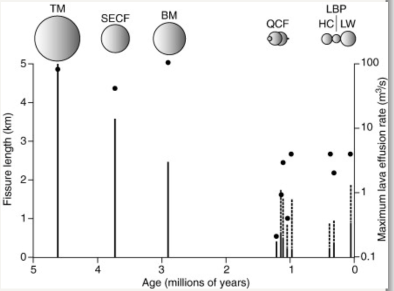
- The spatial intensity (or density) of volcanism (vents per unit area) and erupted volume reflects productivity over area and through time.

Distributed volcanic fields

Yucca Mountain volcanic field



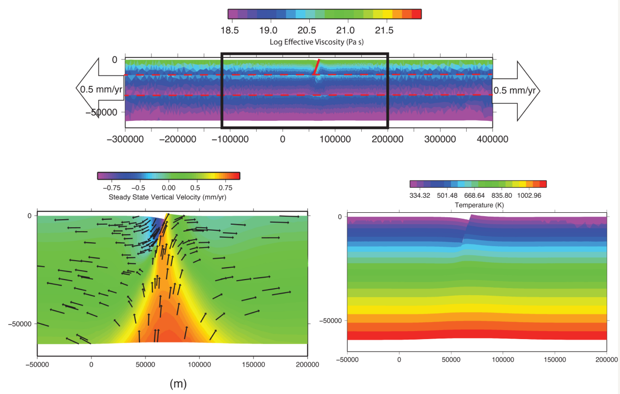
- Overall, recurrence rate of volcanic activity is low compared to individual composite volcanoes



from Valentine and Perry,2005

Productivity and recurrence rate

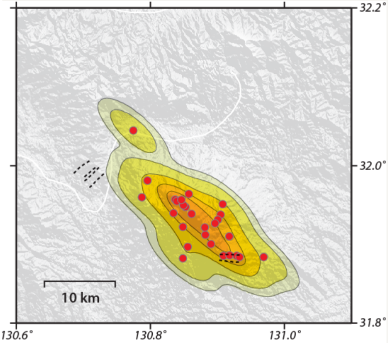
Yucca Mountain volcanic field



Numerical model of extension and focusing (from Rocco Malservisi), created using GTecton (Govers and colleagues). In this simplified model, slow extension of the crust with a through going fault creates vertical, slightly divergent motion and focusing in the fault region. Compare with Yucca Mountain (previous slide)

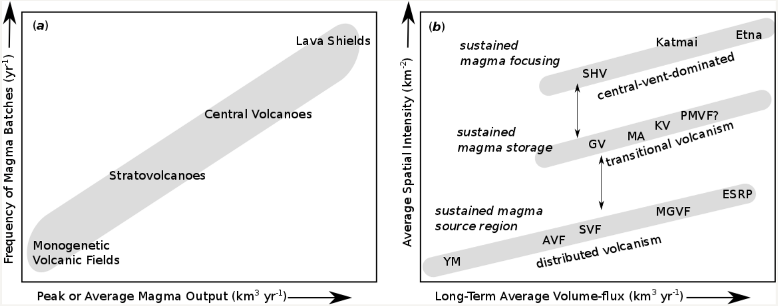
Transitional volcanism

Kirishima volcano



Some volcanic systems appear to be transitional between composite volcanoes and distributed volcanoes, like Kirishima, Japan. Some volcanic systems appear to change through time, becoming more distributed or more dominated by a central vent.

Summing up spatial and temporal trends



We may need to change our view of volcano classification from the "traditional" picture at left (from G.P.L. Walker, 2000) to a classification based on spatial intensity of volcanism (volcanoes per unit area - spatial distribution) in addition to long term output (some measure of magma productivity).

Distributed volcanic hazards

Auckland volcanic field, Auckland, NZ

Image of possible eruption, from Auckland Council website

- Where will new vents form?
- How frequently will volcanism occur?
- What will be the effects of volcanism?
- How does a conceptual model of volcanism help address these questions?

Examples of explosive activity

Cerro Negro, Nicaragua

- Eruption column from Cerro Negro volcano
- Eruption column heights from Cerro Negro commonly reach 7 km
- Active cinder cone for last 150 yr

Examples of explosive activity

Tolbachik, Kamchatka, Russia

- 1975 eruption of Cone 2 at Tolbachik volcano
- Incandescent jet reached about 800 m
- Eruption column height reached 18 km
- Tephra dispersed to 500 km

Examples of effusive activity

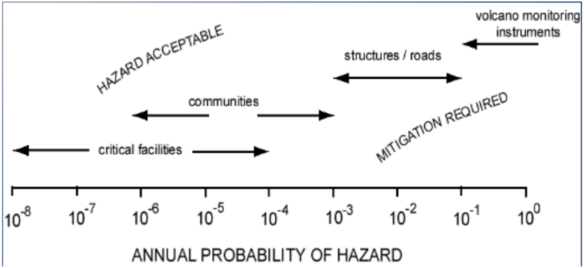
Parícutin, Mexico

Effusion of lavas from the 1943–1952 eruption of Parícutin had the most devastating impact

Tolbachik, Kamchatka, Russia, 2012

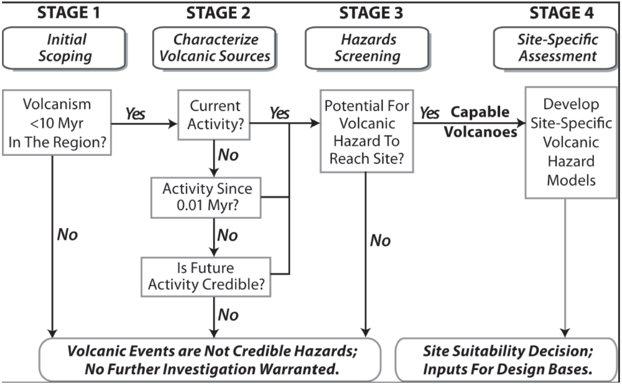
Lava flows of approximately 0.5 km^3 reaching 20 km from vent

How do we compare hazard?



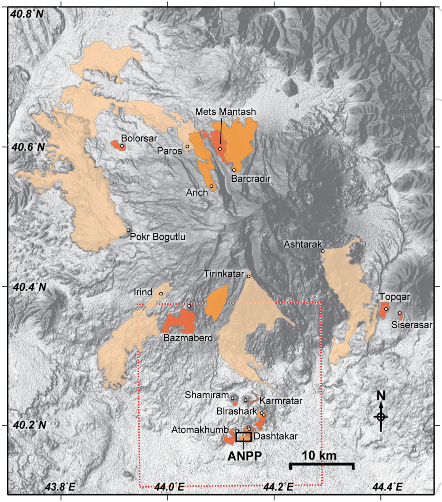
- Terms like "low" hazard or "low" probability are relative and sometimes misleading
- For critical facilities, probabilities of $10^{-4} - 10^{-8}$ per year may be considered "high"
- Goal is to bound probability within one order of magnitude

IAEA guidelines

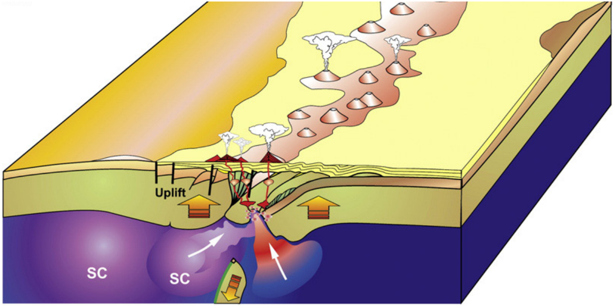


International Atomic Energy Agency guidelines follow a hierarchical approach to volcanic hazard assessments

Location map of ANPP



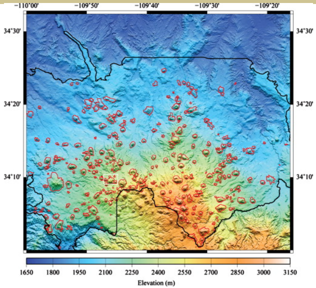
Conceptual model of volcanism in Armenia



Slab-breakoff and concurrent uplift associated with volcanism
Intensive errosion of the uplifted terranes.

Subduction ceased about 10 Ma. Slab break-off may result in hot–dry magmas produced and erupted at very low rates. Modified from *Keskin et al., 2008*. Last Aragats activity 0.4 Ma?

What is the potential distribution of vents?



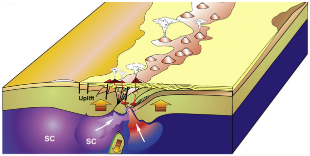
The observed (mapped) distribution of vents is one realization of the "potential" vent distribution. That is, the geological processes that gave rise to the mapped vent distribution could also give rise to vents in other locations (randomness).

Springerville volcanic field, USA (left) from Condit, 2005

Conceptual model (below) from Keskin (2008)

To understand the likelihood of a new vent forming in a given location, a statistical model of vent distribution is required, based on factors like:

- Distribution of mapped vents
- tectonic features that seem to influence volcano distribution,
- temporal trends in volcanism (shifts in location through time)
- geophysical data, and more!



Slab-breakoff and concurrent uplift associated with volcanism
Intensive errosion of the uplifted terranes.

The main requirement is a "conceptual model" of volcanism.

The concept of spatial distribution

The actual (mapped) distribution of vents is one realization of the potential distribution of vents. What is the probability density function that describes the potential distribution of vents?
Vent intensity: a statistical model of the potential number of vents per unit area (integrates to N where N is the total number of vents
Vent density: the probability density fuction that represents the potential distribution of vents (integrates to 1, as all pdfs do!).

Uniform random vent distribution

Wells-Gray-Clearwater volcanic field, BC, from Goward and Hickson (1995)

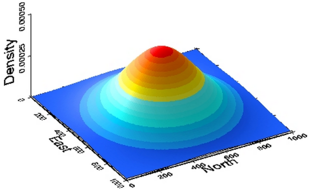
Using a uniform random statistical model, the spatial intensity is:

$$\hat{\lambda}(\mathbf{s}) = N/A$$

The local spatial intensity estimate, $\hat{\lambda}(\mathbf{s})$, depends only on the area, A of the volcano field and the number of vents, N:

$$N = \int \hat{\lambda}(\mathbf{s}) dA$$

Isotropic kernel density function



Gaussian kernel

$$\hat{\lambda}(\mathbf{s}) = \frac{1}{2\pi h^2} \sum_{i=1}^{N} \exp\left[-\frac{1}{2}\left(\frac{d_i}{h}\right)^2\right]$$

The local spatial intensity estimate, $\hat{\lambda}(\mathbf{s})$, depends on its distance, d_i, to each event location, and the smoothing bandwidth, h.

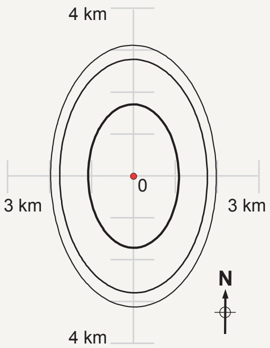
Anisotropic kernel density function

$$\hat{\lambda}(\mathbf{s}) = \frac{1}{2\pi N\sqrt{|\mathbf{H}|}} \sum_{i=1}^{N} \exp\left[-\frac{1}{2}\mathbf{b}^{\mathbf{T}}\mathbf{b}\right]$$

where $|\mathbf{H}|$ is the determinant of the bandwidth matrix, $\mathbf{b} = \mathbf{H}^{-\mathbf{1/2}}\mathbf{d}$ and $\mathbf{b}^{\mathbf{T}}$ is the transform of $\mathbf{b}$. $\mathbf{d}$ is a 1×2 distance matrix, and N is the total number of volcanic vents.

Shape of the kernel density function

Shamiram Plateau kernel density function

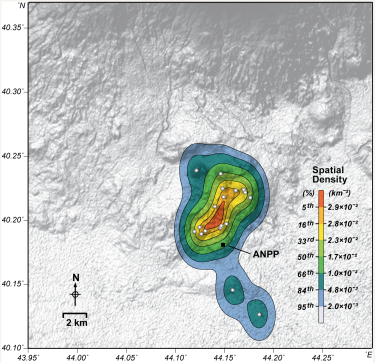


Generated by the SAMSE bandwidth estimation algorithm *from Duong 2007*

$$\sqrt{\mathbf{H}} = \begin{bmatrix} 0.92 & -0.005 \\ -0.005 & 1.5 \end{bmatrix},$$

which shows the semi-major axis is about 3 km long and oriented N–S

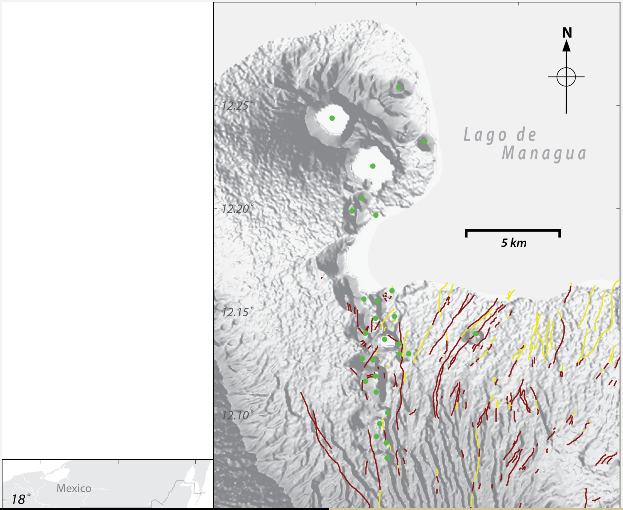
Shamiram Plateau kernel density model



A spatial density model for the Shamiram Plateau generated using an anisotropic kernel density function and using SAMSE bandwidth optimization

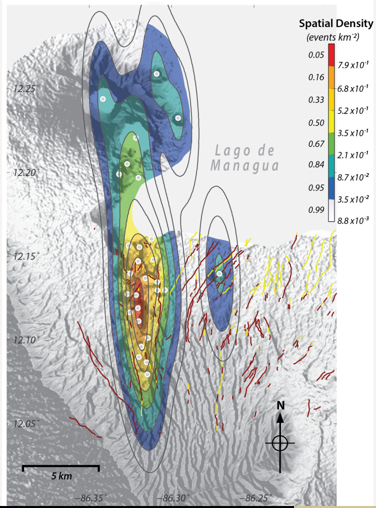
Shape of the kernel density function

Nejapa – Apoyeque kernel density model



Shape of the kernel density function

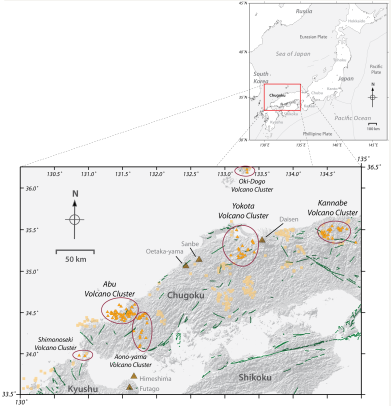
Nejapa – Apoyeque kernel density model



A spatial density model for the Nejapa – Apoyeque alignment generated using an anisotropic kernel density function and using SAMSE bandwidth optimization

Tectonic features and distributed volcanism

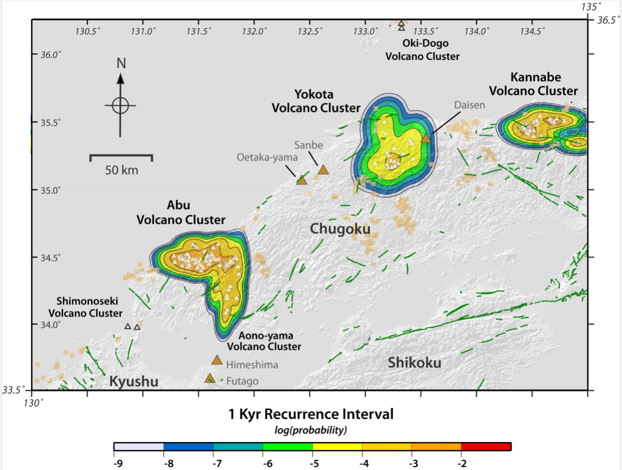
Chugoku, SW Honshu, Japan



- Several distributed volcanic fields and composite volcanoes
- relatively flat subduction of the Philippine Sea plate with leading plate edge under Chugoku (?)
- left-lateral strike slip motion in upper plate with book-shelf (?) faulting
- alkaline basaltic volcanism with adakites in Aono-yama

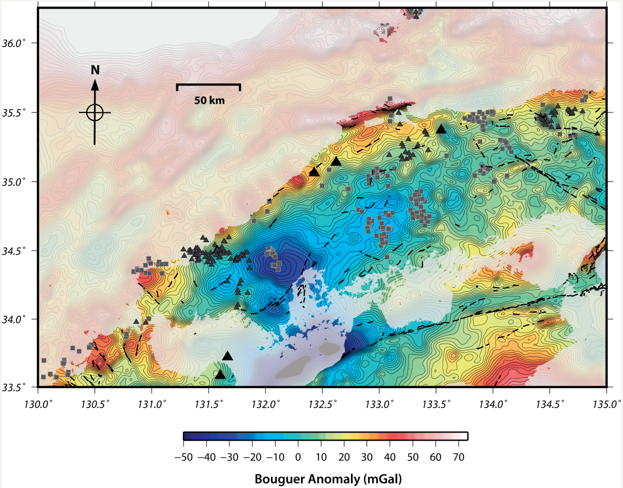
Tectonic features and distributed volcanism

Chugoku, SW Honshu, Japan



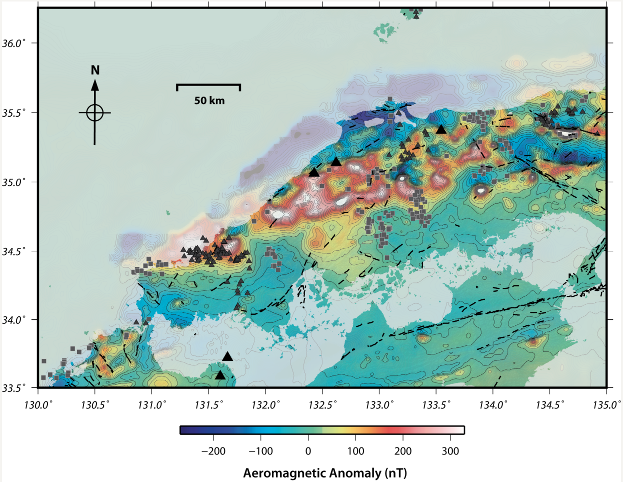
Tectonic features and distributed volcanism

Chugoku, SW Honshu, Japan



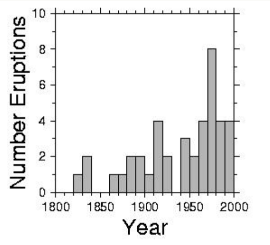
Tectonic features and distributed volcanism

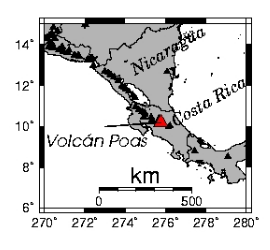
Chugoku, SW Honshu, Japan

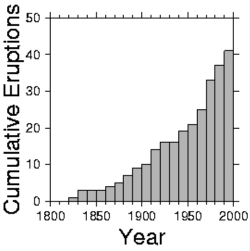


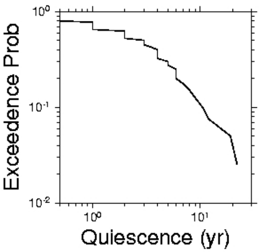
Look at the time series data!

Relatively recent eruptions of Poás volcano, Costa Rica









Stationarity

Simple univariate statistical models can be developed for data sets that are stationary. In Stationary time series, the mean and variance or repose intervals do not change substantially with time.

Consider a time series of N events (earthquakes, eruptions) that starts and time S and ends at time T. The cumulative number of events that have occurred by time t is $M(t)$, and the cumulative density of events, or fraction of events, that have occurred by t is:

$$F(t) = M(t)/N$$

and the expected number of events, $E(t)$ that have occurred by time t is:

$$E(t) = \frac{t - S}{T - S}$$

Stationarity

Is $F(t)$ significantly different from $E(t)$? The upper confidence bound on the expected number of events at time t is:

$$H(t) = E(t) + \frac{1.36}{\sqrt{N}}$$

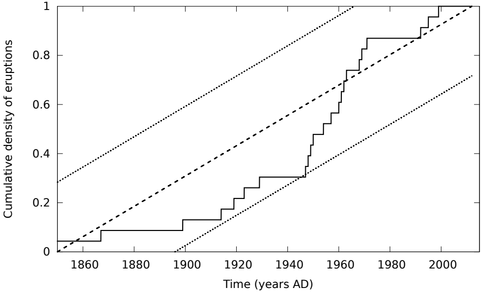
and the lower bound at time t is:

$$L(t) = E(t) - \frac{1.36}{\sqrt{N}}$$

at 95% confidence. We can also use a Kolmogorov-Smirnov test, based on the maximum difference between E(t) and F(t)

Stationarity

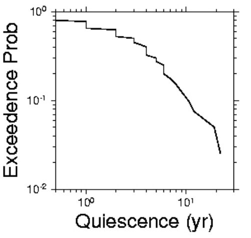
Example: history of eruptions of Cerro Negro volcano since its formation in 1850



Cerro Negro volcano eruptions are non-stationary since 1850, at the 95% confidence level. Instead, the rate of eruptions appears to have increased after 1940 and may have decreased after 1971.

How to plot an empirical survivor function

Relatively recent eruptions of Poás volcano, Costa Rica



The empirical survivor function: The plot shows the likelihood or repose, or "quiescence" at this volcano exceeding some period of time.
This model is empirical, rather than statistical, becuse it is based entirely onthe history of eruptions at this volcano.

How to plot an empirical survivor function

Steps in calculating the values shown on the graph:
Gather the repose interval data (define what you mean by *repose*, so that you have N repose intervals.
Sort the repose intervals in ascending order (shortest repose intervals first):

$$t_i \leq t_{i+1} \leq t_{i+2} \leq ... \leq t_{N-1},\ 0 \leq i < N$$

These sorted values are the x coordinate values of the plot.
Calculate the empirical survior function:

$$\hat{S}(t) = \frac{N-i}{N}$$

These are the y coordinate values of your plot.

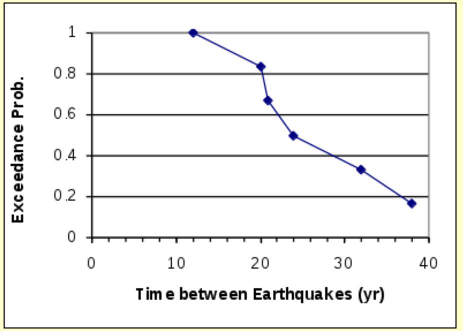
How to plot an empirical survivor function

Parkfield earthquake recurrence

Given: years with magnitude M=6.0 or greater earthquakes near Parkfield, California, calculate the empirical survivor function. Earthquakes in: 1857, 1881, 1901, 1922, 1934, 1966, 2004

How to plot an empirical survivor function

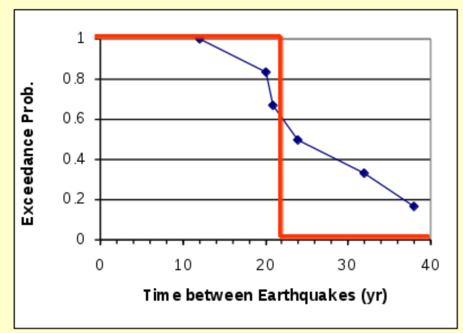
Empirical survivor function for earthquakes M$\geq$6 for Parkfield, CA, USA



Note the slight tendency for earthquakes to cluster around the median recurrence interval.

Is this periodic behavior?

Earthquakes M≥6 for Parkfield, CA, USA are periodic



Note the poor fit to a periodic model – some other statistical model might work better! This result dictates that we try another approach!

Statistical model selection and evaluation

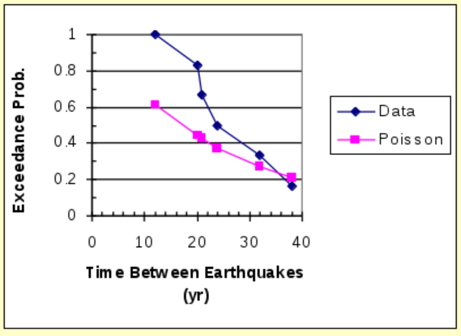
The Second Model: Recurrence intervals are the independent realizations of an exponential distribution. This model is OFTEN used, because it naturally describes the recurrence interval for Poisson Processes.
That is, the fact that an earthquake (or volcanic eruption for that matter!) has not happened yet, tells us nothing about how much time will elapse before it does happen.

$$S(t) = \exp[-t/\mu] \tag{1}$$

Our estimate of the mean is 24 yr. Using this statistical model, calculate the expected values for $S(t)$ at 21 yr, 32 yr, and 100 yr recurrence intervals. Verify that your calculations are correct. You should get $S(t) = 0.41, 0.26, 0.015$.

Statistical model selection and evaluation

Earthquakes M≥6 for Parkfield, CA, USA are random in time



The simplest (but often inadequate!) way to validate the model is by visual inspection. Visual inspection suggests the fit is not good!

We conclude that the Parkfield earthquakes are not modeled by a Poisson Process, they do not occur randomly in time.

Statistical model selection and evaluation

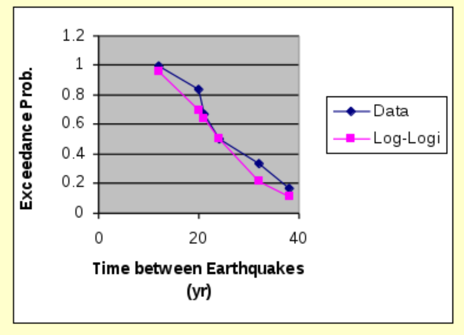
The Third Model: The recurrence intervals are quasi-periodic, but have large standard deviation. We need a model that can realistically capture the large standard deviation in the recurrence intervals, and perhaps add a thicker "tail" to the distribution, compared to "normal". The Log-logistic function does this:

$$S(t) = \frac{1}{1 + (t/\mu)^{\beta}} \tag{2}$$

Use $\mu = 24$ yr and $\beta = 4.0$. Using this new statistical model, recalculate the expected values for $S(t)$ at 21 yr, 32 yr, and 100 yr recurrence intervals. Verify that your calculations are correct. You should get $S(t) = 0.63, 0.24, 0.0033$.

Statistical model selection and evaluation

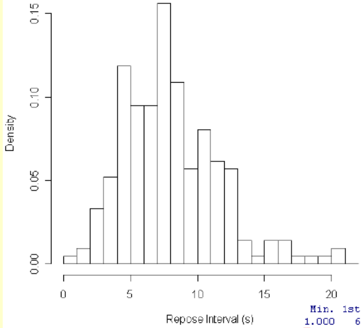
The log-logistic model actually provides a better fit to the Parkfield data (at least visually a better fit). Validation would vastly improve by actually testing the goodness-of-fit



Although the data remain very few indeed (so validation is not powerful!), one would better use the log-logistic model than periodic or Poisson models to forecast earthquakes in Parkfield.

Statistical model selection and evaluation

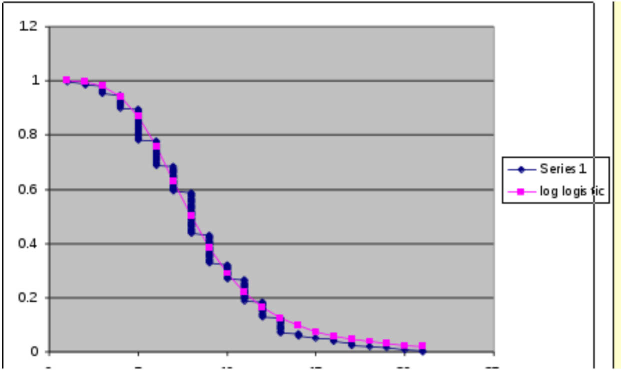
Another Example: Repose Intervals at Cerro Negro during 1995 eruption



Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	6.000	8.000	8.531	11.000	21.000

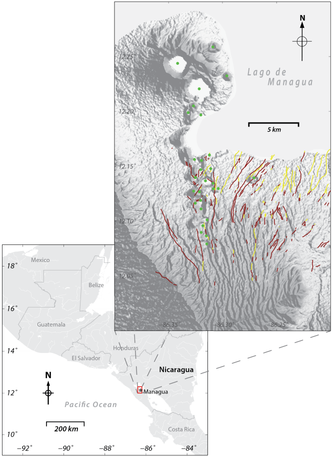
Note: This graph and summary statistics produced in the R statistical modeling package (free and open source).

Modeling Cerro Negro Repose Intervals



Modeling the repose interval (in seconds), between explosions at Cerro Negro volcano (1995). log-logistic distribution gives a reasonably good fit!

Exercise on temporal trends

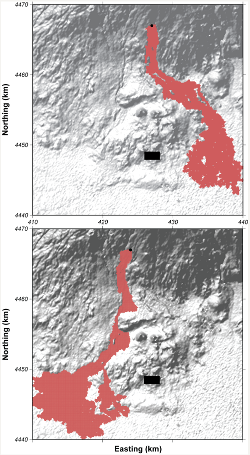


Given the distribution of past eruptions in the Nejapa – Apoyeque alignment, what is the probability of future eruptions?
12 Eruptions in Years: -7430 (7430 BC), -7300, -5350, -5230, -4390, -4160, -3050, -2550, -1050, -550, -50, 1060

- Is the time series stationary?
- Plot the empirical survivor function for the 11 repose intervals
- Try fitting a random (Poissonian) statistical model

Lava flow simulation

Example simulation



Lava flows from Aragats volcano (volume $= 0.5\,\mathrm{km}^3$). Note change in area inundated with slight change in vent location.

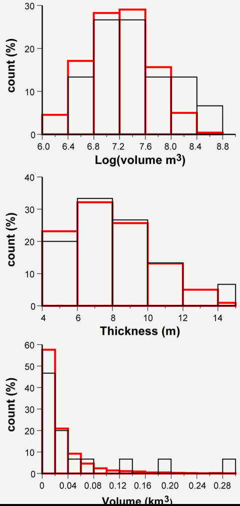
Lava flow simulation

Using Geologic data to constrain inputs to the lava flow simulation

Volcano *(source)*	Area *(km²)*	Thickness *(m)*	Volume *(km³)*	Length *(km)*
Arich	16.3	8	0.130	9.48
Atomakhumb	3.9	6	0.020	3.43
Barcradir	32.9	9	0.296	12.10
Bazmaberd	13.1	14	0.184	6.34
Blrashark-1	1.6	6	0.010	2.49
Blrashark-2	2.5	7	0.018	3.13
Bolorsar	2.2	6	0.013	2.72
Dashtakar-1	2.1	10	0.021	4.44
Dashtakar-2	1.6	6	0.009	3.66
Karmratar	0.7	4	0.003	3.61
Mets Mantash	8.9	9	0.080	8.47
Shamiram	1.0	4	0.004	1.41
Siserasar	0.8	11	0.009	1.72
Tirinkatar-2	13.3	4	0.053	6.54
Topqar	2.9	9	0.026	3.07

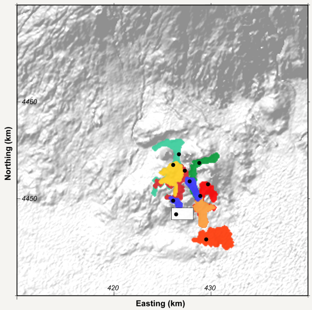
Lava flow simulation

Inputs to the lava flow simulation



- Model inputs include lava flow volume and model thickness
- Tuned using observed flows on the Shamiram Plateau
- black – observed flow parameters
- red – simulated flow parameters

Volume-limited lavas of the Shamiram Plateau

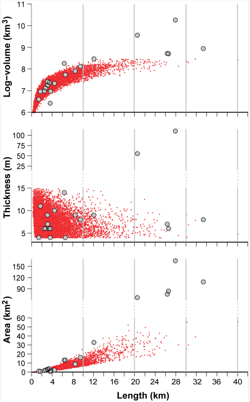


Lava flow model inputs

- vent location, $X \sim$ spatial density(SAMSE bandwidth,N)
- modal flow thickness, $T \sim \log N_T(\mu_t, \sigma_t)$
- total flow volume, $V \sim \log N_T(\mu_v, \sigma_v)$
- lava pulse volume (volume / code step)
- grid of elevations (DEM)
- area of interest (AOI)

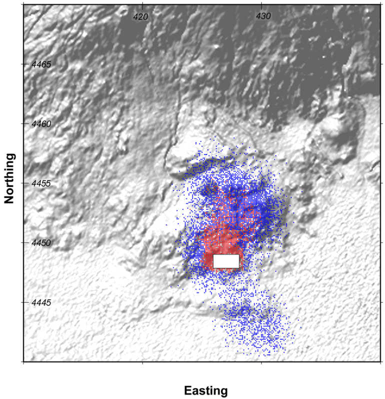
Lava flow simulation

Model performance



Comparison of observed lava flow parameters and simulated parameters for 10 000 simulations

Lava flow simulation results



- red dots – lava flow inundates the site (white box)
- blue dots – lava flow misses the site

Conclusions

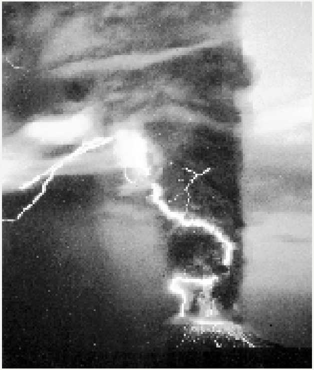
In monogenetic volcanic fields, hazards are best assessed probabilistically:

- **Spatial density** is estimated using anisotropic kernel density functions. These have the advantages of being (*1*) nonparametric, (*2*) continuous, (*3*) estimated quantitatively using bandwidth optimization algorithms.
- **Lava flows** seem to be well-represented by a simple inundation model tuned to mapped lava flow features (volume, thickness, effusion rate).
- **Analysis** yields a conditional probability of lava flow inundation. In the Armenia NPP case:

$$Pr[N \geq 1 \mid \text{new volcano}] = 0.249$$

Additional points

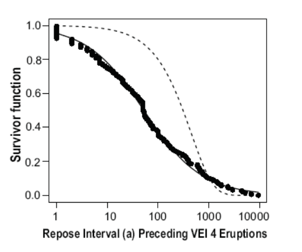
Tephra fallout



- **tephra fallout** hazards may also be substantial in monogenetic volcanic fields.
- Like the probability of lava inundation, tephra fallout can be assessed using numerical simulation, for example using the code *Tephra2*

Additional points

Recurrence rate of volcanism



- **Recurrence rate** is usually also factored into monogenetic volcanic hazard assessment
- Recurrence rate models might be log-logistic, clustered, or modeled non-parametrically *Bebbington and Cronin, 2010*
- recurrence rate of new vent formation in the Shamiram plateau is on order of 5×10^{-6} per year, yielding:

$$Pr[N \geq 1] = 1.3 \times 10^{-6}$$