Background on Statistical Modeling of Volcanic Vent Locations[1]

Chuck and Laura Connor, Aurélie Germa, Jacob Richardson

September 6, 2018

[1] Prepared for the CoV workshop on Volcanic Hazards for Nuclear Facilites

Hazards of new volcanic vents

The opening of a new volcanic vent is a geologically rare phenomenon, but one that can produce significant hazards. These hazards include the formation of lava flows, pyroclastic fall and flow, ballistic, and tephra fallout. Most volcanic systems are distributed, in the sense that new eruptions sometimes occur at a new location, which has not been the site of previous volcanic eruptions. Geologic mapping shows that over time a volcanic system, like the Mt. Lassen volcanic system, can form tens to hundreds of vents. This "diffusion of vents" means that the hazards of new vent formation need to be assessed.

Geologists Lyell and Desmarest both noted the dispersed nature of some volcanism, in Mexico and France respectively. Williams described the dispersed nature of volcanism around the newly formed Parícutin volcano (Mexico) in great detail. Rittman appears to have coined the term monogenetic volcanic field in his 1962 book *Volcanoes and Their Activity*. Nakamura [1977] appears to have been the first to use the term monogenetic volcanism in a scientifically reviewed publication.

Problems assessing new vent formation

We do not know where volcanic vents will form in the future. Even with the best available seismic data it has provided to be very difficult to forecast where vents will form – often just hours before they do erupt at the surface. So we have to guess. Statistical models are the most widely used method of guessing.

Estimating where new volcanic vents are likely to form has often proved contentious. Why the controversy? The underlying geologic processes controlling the distribution of these events are complex and incompletely understood. The frequency of such potentially catastrophic events is often low, so data used in these analyses are often sparse. The selection of specific statistical models to estimate spatial density is often subjective. These factors result in uncertainty.

Figure 1: A new vent forming. The 1975 eruption of Tolbachik volcano is one example of a large scoria cone built by eruptive activity during a period of only several weeks. Here, tephra fallout is spreading far downrange from the new vent. Photo attributed to Pavel M. Kartashov.

Hazards associated with the opening of new vents may be exasperated by the topography of volcanic systems, which is often complex and characterized by steep slopes. For example, small variations in vent location may cause lava to flow in a completely different direction down the flanks of a volcano. There is no doubt that probabilistic models of lava flow inundation are quite sensitive to models of vent location!

In addition, loci of activity may wax and wane with time, such that past vent patterns may not accurately forecast future volcano locations. It is important to determine if temporal patterns are present in the distribution of past events, so that an appropriate time interval can be selected for the analysis (i.e., use only those vents that repre-

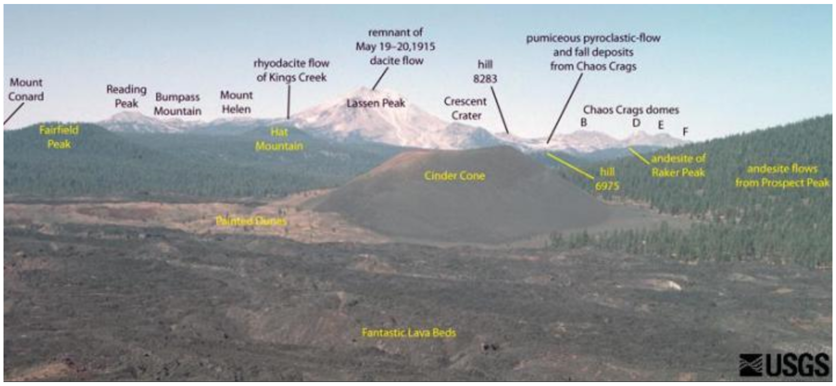

Figure 2: At Lassen, new eruptions sometimes produce new volcanic vents. The Lassen area includes approximately 300 volcanic vents formed in the last 2 Ma. Photo from USGS

sent likely future patterns of activity, not patterns that represent older volcano distributions).

These problems place a premium on statistical modeling of the likelihood of new vent formation. In spatial density estimation, the likelihood of new vent formation is treated as a probability density function. The probable location of a new vent, given that a new vent forms, varies across the map area. Statistical models attempt to guess what this probability density function looks like, using data, such as the locations of older vents.

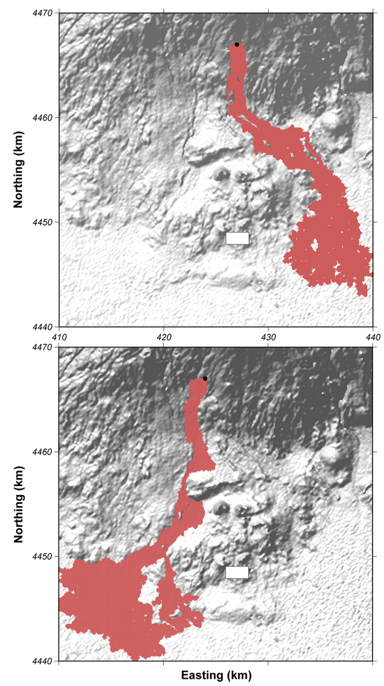

Figure 3: Small changes in vent location can result in large changes in areas impacted by volcanic hazard, illustrated here by a lava flow simulation for two different vents on the flanks of Aragats volcano, Armenia.

What is spatial density?

The reason to estimate spatial density is to determine possible locations of future geophysical events (volcanic eruptions, earthquakes, lahar source areas, sinkholes), or to estimate the probability of an event occurring at a specific location, given that such events occur within the region.

There is ambiguity in the literature regarding the use of the terms density and intensity. In the geosciences, variation in the number of events per unit area (say the number of volcanic vents or earthquake epicenters) is described using the term density. For example, one might report the density of volcanic vents in a region as the number of vents per 1000 km^2. Intensity, in geoscience contexts, often refers to the magnitudes of these events. The intensity of a volcanic eruption can be characterized in terms of its total mass of eruptive products or related indices [Pyle, 2000].

Density and intensity are defined differently in spatial statistics. In this context, spatial intensity refers to the expected number of events per unit area defined at a point, $\mathbf{s}$, a matrix containing the x and y coordinates of the location of the point [Silverman, 1978, Diggle, 1985]. Suppose there exists a set of events (e.g. volcano locations) that

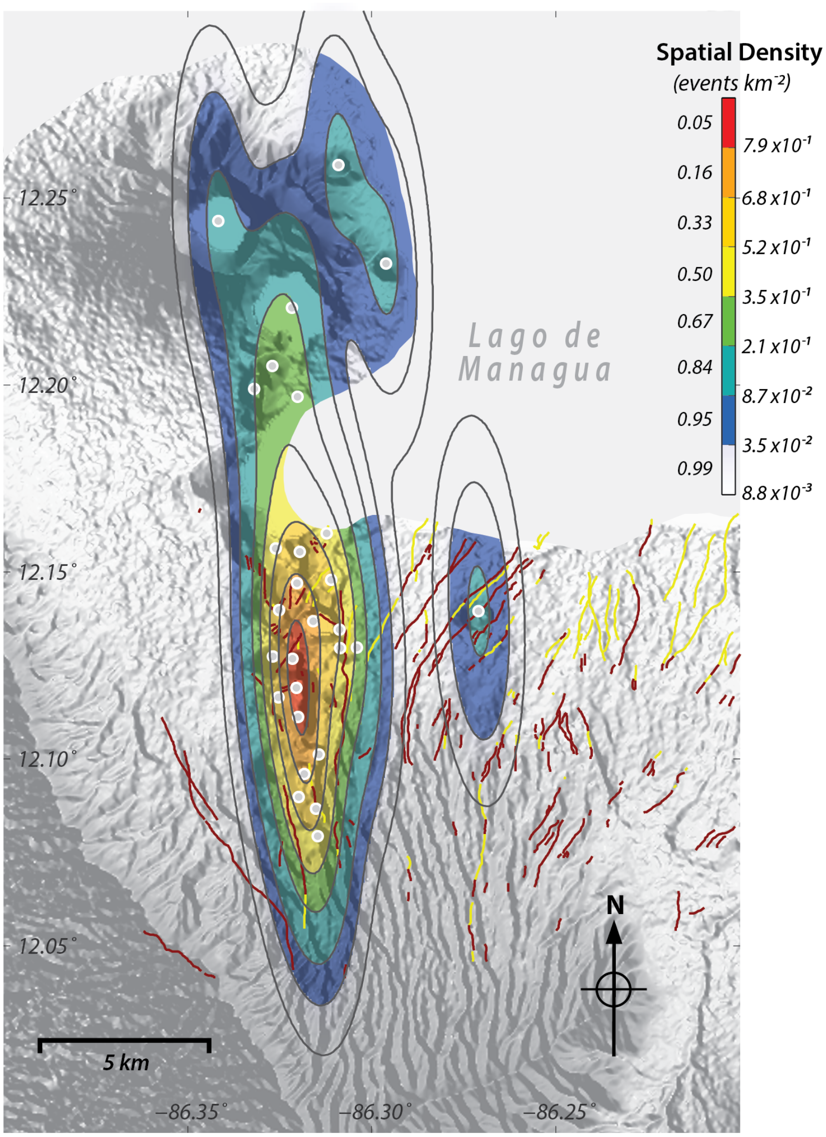

Figure 4: A statistical model of spatial density. White circles show volcanic vent locations on the western side of the Managua graben (Nicaragua), superimposed on a shaded-relief digital elevation model. Faults in the area are shown by red and yellow lines (red faults slipped during the 1972 earthquake that destroyed the city of Managua.

Areas most likely to experience future volcano vent formation are shown by colored areas, estimated using a elliptical kernel density function. There is a 95% chance that a future vent will form in the contour enclosing the blue-shaded area, with the highest probability zones shown in red.

One vent, in downtown Managua, appears to be an outlier and shows the shape of the kernel density function. From Connor et al. [2015], data courtesy of Jose Armando Saballos.

occur within a given region, R. These events can be designated as $\mathbf{x}_n(n = 1, 2, ..., N) \in R$ where N is the total number of events, each consisting of the spatial location, x and y of the event (possibly given in Easting and Northing coordinates, or, latitude and longitude). One way we can create a model of spatial intensity from these events is to imagine they are realizations of a random variable, $\mathbf{X}$, a function that describes the set of all possible realizations. For example, $\mathbf{X}$ might be the distribution of potential earthquakes or the distribution of potential volcanoes, from which a set of observed realizations (e.g. those found in the earthquake catalog or on a geologic map) are drawn. The spatial intensity is formally written as [Gatrell et al., 1996]

$$\lambda(\mathbf{s}) = \lim_{d\mathbf{s} \to \mathbf{0}} \left\{ \frac{E(\mathbf{X})}{d\mathbf{s}} \right\} \tag{1}$$

where $E(\mathbf{X})$ is the expected number of events that fall within a small area $d\mathbf{s}$ about the point $\mathbf{s}$ (hence, if the location, $\mathbf{s}$, is given as Easting and Northing with units of meters, then the units of $\lambda(\mathbf{s})$ are m^{-2}). At first glance it appears that the statistical definition of *intensity* is equivalent to the term *density* as commonly used in the geosciences. This is not quite true. The geological processes that result in a given event distribution are incompletely known. We can think of these geological processes as giving rise to a stochastic point process that describes the relationship between the set of events and the geological processes that led to their formation. As the stochastic point process is incompletely known, the true value of the local spatial intensity, $\lambda(\mathbf{s})$, is also unknown. That is, the observed distribution of events is only one realization of the underlying process that gives rise to these events. Our goals are to find an estimate of the spatial intensity, $\hat{\lambda}(\mathbf{s})$, that approximates the true but unknown value of spatial intensity, $\lambda(\mathbf{s})$, and to understand the uncertainty in this estimate.

The sum of $\hat{\lambda}(\mathbf{s}) \times$ grid area on a spatial density map is equal to 1 (or close to 1 if some of the density function is beyond the bounds of the map). The sum of $\hat{\lambda}(\mathbf{s}) \times$ grid area on a spatial intensity map equals the total number of vents on the map (with the same caveat).

In hazard assessments, there is a further requirement, that this information be used to forecast the spatial distribution of possible future events. Often we consider spatial intensity in terms of the probable location of some future event, given that one occurs within our region of interest. This conditional probability can be estimated by:

$$\hat{f}(\mathbf{s}) = \frac{\hat{\lambda}(\mathbf{s})}{\int_R \hat{\lambda}(\mathbf{s}) d(\mathbf{s})}. \tag{2}$$

Integrating $\hat{f}(\mathbf{s})$ across the region of interest, R, gives unity, if R is sufficiently large. Since all values of $\hat{f}(\mathbf{s})$ within this region are greater than or equal to zero, this makes $\hat{f}(\mathbf{s})$ a probability density function and this function may be used in probabilistic hazard models. $\hat{f}(\mathbf{s})$ is referred to as one estimate of the spatial density, and one can consider the spatial density per unit area in terms of conditional probability (e.g. given a volcanic event in the region, what is the probability that the event will occur within some small area about the point **s**?). In addition, care is required in the selection of the region R, as external events located close to the border may have a non-negligible contribution to spatial density. A practical approach is to select R to be quite large compared to the region of specific interest (e.g. the volcanic system). In the following, we refer to spatial intensity and spatial density within the context of spatial statistics.

Spatial density is a type of probability density function, so the area under the curve must sum to 1.

Assumptions behind spatial density estimates

How does one develop a best estimate of spatial density? In the real world, there is only one realization of an underlying geologic process, the observed distribution of past events. Unfortunately, geology is not conducive to repeating the experiment in a natural system. For a given region there is just one earthquake catalog, or one geologic map of volcano distribution. Presumably, if there existed a complete geophysical model for these events, we would use this information to better forecast the locations of future events. For example, if we knew the distribution of melt in the asthenosphere and lithosphere, and if we knew the state of the lithosphere through which the magma rises, we might have a better sense of where volcanoes or volcanic vents are most likely to form next. Currently, we lack such a complete geophysical perspective. Some data sets give an idea of where partial melting of the mantle might occur, for example seismic tomographic models of "slowness" in the lithosphere and asthenosphere [Kiyosugi et al., 2010]. Other data, such as variations in gravity across a region [Connor et al., 2000, Deng et al., 2017], show some correlation with the existing distribution of volcanoes in some circumstances, but the mechanisms relating gravity anomalies to the origin of magmas are not completely understood.

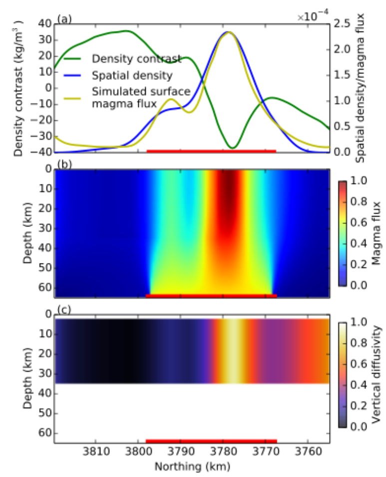

Figure 5: There have been recent attempts to build geophysical models to explain spatial density. Here the spatial density of vents along a profile in the Springerville volcanic field (AZ) is compared to a geophysical model that simulates magma flux, given crustal density contrast [Deng et al., 2017]. Magma originates in a uniform source zone (red bar) but the flux at the surface is altered by lithology variations in the crust. The idea is that these lithology variations cause observed variation in vent spatial density.

The reliance on the distribution of past events implies that these realizations are representations of some underlying random variable, **X**, that will govern the distribution of potential events in the future. This assumption immediately raises a fundamental question. Which are the past events that should be used to develop the spatial intensity estimate, $\hat{\lambda}(\mathbf{s})$, and density, $\hat{f}(\mathbf{s})$? Event datasets used to estimate the spatial density of future events need to be consistent with several features of geological processes.

First, any spatial intensity function for a geologic process must change with time. On time scales of tens of millions of years, plate boundaries change, volcanic arcs wax, wane, and migrate, and major fault systems reorganize. In very long term probabilistic hazard

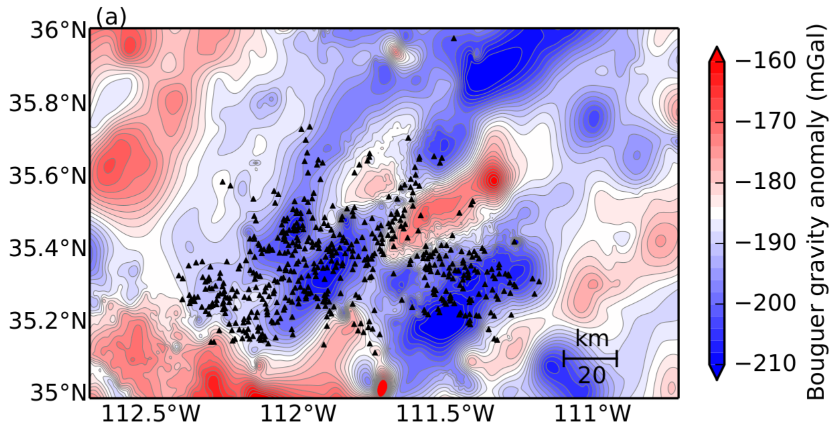

Figure 6: Volcanic vents are located in gravity lows, adjacent to gravity highs, in the San Francisco volcanic field (AZ). Dense, presumably rigid continental crust (indicated by red areas in the contoured gravity data) may hamper magma ascent, leading to the formation of volcano clusters over less rigid, less dense crust.

The spatial density model should be sensitive to geologic boundaries that may influence volcano vent distribution (and the possible locations of volcanic vents that may form in the future). From Deng et al. [2017].

assessments for high-level waste repositories, which may have 10^6 a performance periods, these factors have to be considered in weighing the validity of using specific data in developing spatial intensity models. For processes like volcanism, where a geologic record of past events usually persists for tens of millions of years, consideration needs to be given to which events best represent the distribution of future volcanism. For example, the distribution of Miocene volcanoes in a given area might be much less relevant than the distribution of Pliocene and Quaternary volcanoes. Thus, in order to develop an estimate of the spatial intensity, a model of the geologic evolution of the system is required. This geological model is used to justify the inclusion of some geological features in the event dataset, and the exclusion of others.

Second, it is necessary to assess the completeness of the geologic record. In seismology, it is particularly clear that short earthquake catalogs carry the risk of biasing estimates of spatial intensity. That is, the record of earthquakes in a given region collected on a short time scale might give an incomplete picture of the unknown distribution of potential earthquakes, $\lambda(\mathbf{s})$. Even volcanic events might be missed in initial geological investigations, as volcanic vents might be buried in sediment or otherwise obscured.

Third, geological events, even when they are all identified, may be so rare as to present an incomplete picture of the underlying process. Consider an earthquake as a single event, $\mathbf{x}_n$, one realization of the random variable, $\mathbf{X}$. If, for example, $\mathbf{X}$ can be characterized by a uniform random distribution, then it is likely that the observed set of realizations will have a spatially random distribution within the region of interest, R. However, the underlying density usually has additional structure, causing independent realizations to cluster. For example, earthquake epicenters tend to cluster along plate bound-

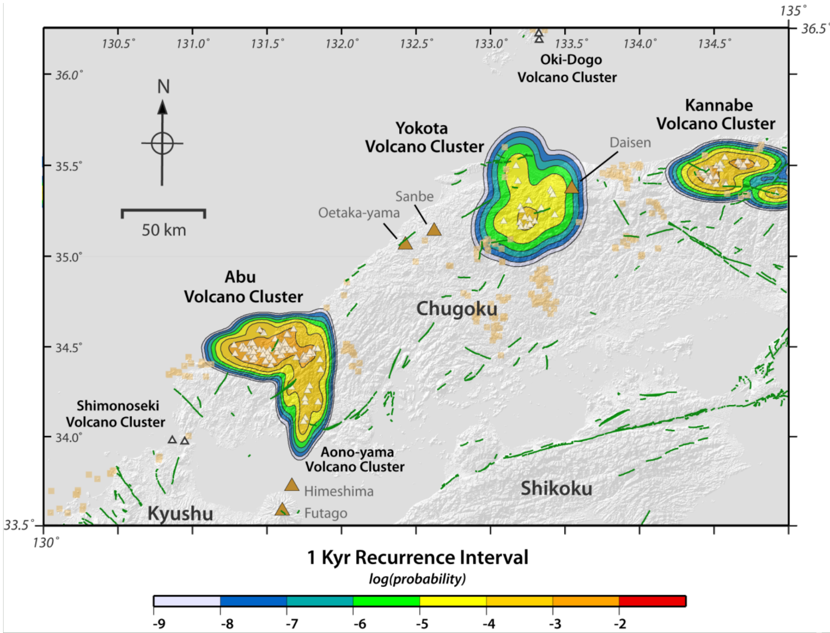

Figure 7: Volcanic vents in SW Honshu (Japan) migrate with time. Quaternary vent clusters (white triangles) do not have the same distribution as Pliocene volcanic vents (tan squares), or composite volcanoes (tan large triangles).

Because of this change in the location with time, the spatial density model includes some events and not others. Alternatively a spatio-temporal density model might be developed [Bebbington and Cronin, 2011].

On this map, the log(probability) is contoured by multiplying the spatial density of volcanism in each volcanic field by its estimate recurrence rate.

aries and volcanoes cluster above zones of partial melting in the mantle. For random variables with a great deal of statistical structure, such as many modes in spatial intensity, a great number of events might be required to identify the statistical structure of the random variable.

Figure 8: Volcanic vents in the Springerville volcanic field (AZ) are one realization of a random variable.

Fourth, it is critical to ascertain which geologic features are actually independent events. The true statistical structure of the random variable, $\mathbf{X}$, might be obscured if some events included in the event dataset are not independent. For example, great earthquakes are followed by aftershocks. An earthquake aftershock, however, is not a random sample of the random variable "*spatial distribution of great earthquakes*", because these aftershocks are not realizations of this particular random variable. Rather, they are independent realizations of another random variable, say "*spatial distribution of aftershocks about a great earthquake*". So, the distribution of aftershocks does not necessarily give the best sense of the spatial intensity of great earthquakes, although these two random variables are correlated.

Similarly, volcanoes are complex geologic structures. The spatial distribution of polygenetic volcanoes reflects processes of magma generation and rise through the crust. The distribution of small vents (sometimes referred to as parasitic or adventive cones) does not necessarily reflect the distribution of polygenetic volcanoes, so a spatial intensity estimate that includes all vents as events would not correctly model the underlying random variable. Furthermore, in monogenetic volcanic fields alignments of volcanic cones develop in response to single magmatic events, episodes of magma rise through the shallow crust. This is because single igneous dikes ascending through the crust might form segments and rotate within the shallow crust, each segment feeding a separate vent and each building a volcanic cone. If the goal of analysis is to forecast the distribution of future magmatic events, each of which might produce more than one monogenetic volcano, geological data must be gathered and volcanoes formed by the same magmatic event must be somehow grouped as single events [Runge et al., 2014].

Figure 9: The Laki (Iceland) fissure eruption. Many vents...one event.

Independence of events is not necessarily easy to determine. Rather than simply counting volcanoes on a geologic map, one must make a geologic assessment of the independence of these data. For volcanoes, this is generally accomplished through detailed analyses of radiometric age determinations, stratigraphic correlations, and related geologic data. Often, even detailed analyses do not resolve whether or not specific features should be grouped as single events or treated as separate, independent events.

Consequently, a major task in preparing a spatial intensity estimate is defining the dataset of events to be used. Certainly a major expense in hazard assessment for a community is data gathering to support interpretation of geological features as events. Hazard assessments often consider alternative event datasets and account for the affect of these varying datasets on spatial density estimates. This strategy will be employed in the following examples.

Estimating spatial intensity with kernel methods

Spatial intensity models based on the distribution of past volcanic or seismic events might be parametric or nonparametric. Parametric models involve fitting a distribution, usually one from a common set of distributions (e.g. uniform random or bivariate Gaussian) to the distribution of events throughout a region or within zones (i.e. subsets of the region of interest). This estimate yields a set of parameters (e.g. mean location of the volcanic field in Northing and Easting coordinates, variance in Northing and Easting coordinates, or rotation). Uncertainty in the distribution fit, and uncertainty in parameter estimates of spatial intensity, can be calculated using maximum likelihood estimation. A significant drawback of these parametric methods is that they assume *a priori* that the distribution of volcanoes is explained by the parametric distribution, for example that volcano distribution is reasonably described as a bivariate Gaussian density. This is not necessarily the case. In fact, it has been shown repeatedly that volcanoes cluster within volcanic fields. Such clustering may be completely smoothed by simple parametric models.

A nonparametric approach for estimating the spatial intensity involves kernel density estimation [2]. With this technique, the observed event locations are used to estimate the spatial intensity at any point in the region using a kernel function. For example,

[2] Bernard W Silverman. *Density Estimation for Statistics and Data Analysis*, volume 26. CRC press, 1986

$$\hat{\lambda}(\mathbf{s}) = \frac{1}{2\pi h^2} \sum_{i=1}^{N} \exp\left[-\frac{1}{2} \left(\frac{d_i}{h} \right)^2 \right] \tag{3}$$

is a 2D radially-symmetric kernel function where the spatial intensity decreases with distance from events based on a bivariate Gaussian function. The local spatial intensity estimate, $\hat{\lambda}(\mathbf{s})$, depends on its distance, d_i, to each event location, and the smoothing bandwidth, h. The rate of change in spatial intensity with distance from events depends on the size of the bandwidth, which, in the case of a Gaussian kernel function, is equivalent to the standard deviation of the kernel. In this example, the kernel is radially symmetric, that is, h is constant in all directions. Nearly all kernel estimators used in geologic hazard assessments have been of this type [Lutz and Gutmann, 1995, Connor and Hill, 1995, Condit and Connor, 1996, Connor et al., 2000]. The bandwidth is selected using some criterion, often visual smoothness of the resulting spatial intensity plots, and the spatial intensity function is calculated using this bandwidth. Alternatively, an adaptive kernel function can be used, in which the spatial intensity varies as a function of event spatial intensity. These adaptive kernel functions are also radially symmetric.

Divide the right-hand side of equation 3 by N to obtain the spatial density.

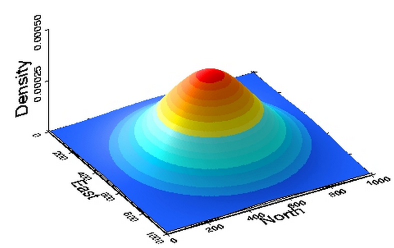

Figure 10: Shaded-relief map of a 2D radially-symmetric Gaussian kernel function drawn about a single point. The point is located at the highest spatial density.

Here is a snippet of PERL code that illustrates the implementation of equation (3) in code:

```
$sumn=0;
for ($i=0; $i<$N; $i++) {
     # here is the distance squared formula
     $dist1=($x-$volcanoes[$i][0])*($x-$volcanoes[$i][0])
          + ($y-$volcanoes[$i][1])*($y-$volcanoes[$i][1]);

     #now calculate and sum the kernel
     $dist2 = $dist1/($h*$h);
     $kuu = 1/(2*3.14159) * exp(-0.5*$dist2);
     $sumn += 1.0/($h*$h) * $kuu;
}
```

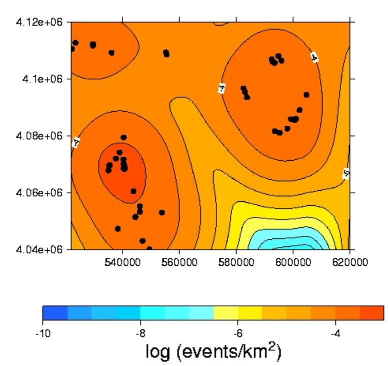

Figure 11: Spatial density about volcanic vents calculated using the computer code at left.

Note that h has to be specified elsewhere in the code, the array *volcanoes* contains the spatial information (Easting, Northing) about the point distribution, and N is the total number of volcanoes in the dataset for which spatial density is estimated. In this code, spatial density is estimated at the point x, y. If there are many such points (say the calculation is done on a grid, then this snippet of code must be executed repeatedly.

Equation (3) is a simplification of the more general case, whereby the amount of smoothing by the bandwidth, h, varies in magnitude depending on direction. A two-dimensional elliptical kernel with a direction varying bandwidth is given by

$$\hat{\lambda}(\mathbf{s}) = \frac{1}{2\pi\sqrt{|\mathbf{H}|}} \sum_{i=1}^{N} \exp\left[-\frac{1}{2}\mathbf{b}^{\mathbf{T}}\mathbf{b}\right] \tag{4}$$

where,

$$\mathbf{b} = \mathbf{H}^{-\mathbf{1/2}}\mathbf{d}. \tag{5}$$

The bandwidth, $\mathbf{H}$, is a 2×2 element matrix that is positive and definite (important because the matrix must have a square root), $|\mathbf{H}|$ is the determinant of this matrix and $\mathbf{H}^{-\mathbf{1/2}}$ is the inverse of its square root. $\mathbf{d}$ is a 1×2 distance matrix (i.e. the x-distance and y-distance from $\mathbf{s}$ to an event), $\mathbf{b}$ is the cross product of $\mathbf{d}$ and $\mathbf{H}^{-\mathbf{1/2}}$, and $\mathbf{b}^{\mathbf{T}}$ is its transpose. The resulting spatial intensity at each point location, $\mathbf{s}$, is usually distributed on a grid which has total extent that defines the region, R.

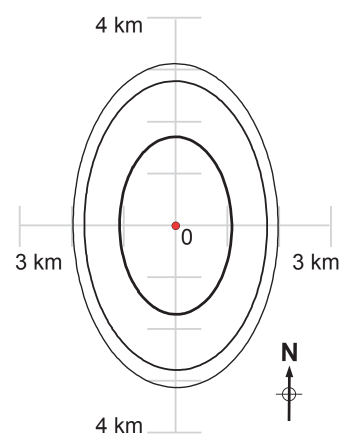

Figure 12: An elliptical bivariate Gaussian kernel density function. Closed contours are 1,2, and 3 standard deviations.

One difficulty with elliptical kernels is that all elements of the bandwidth matrix must be estimated. Several methods have been developed for estimating an optimal bandwidth matrix based on the locations of the event data, summarized in the statistics literature by Duong et al. [2007]. Here we utilize two techniques, a modified asymptotic mean integrated squared error (AMISE) method, developed by Duong and Hazelton [2003], called the SAMSE pilot bandwidth selector, and the smoothed cross-validation (SCV) method of Hall et al. [1992], to optimally estimate the smoothing bandwidth for our Gaussian kernel function. These bandwidth estimators are found in the freely-available *R* statistical package.

Uncertainty in spatial density estimates

Uncertainty exists in the estimates of spatial density. This uncertainty stems from: *(i)* ambiguity in event data sets used to develop kernel estimates, *(ii)* application of the kernel density function, *(iii)* uncertainty in the bandwidth estimate used in the kernel density estimation, and *(iv)* few event data, a common problem in hazard assessment. Each of these in considered in the following, with particular emphasis on the treatment of uncertainty arising from sparse data.

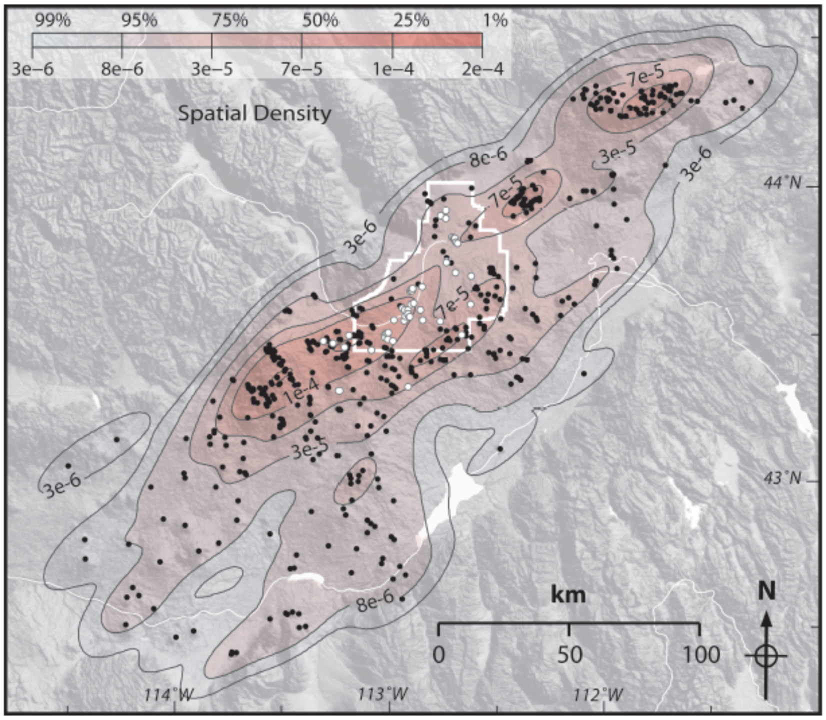

Figure 13: An elliptical bivariate Gaussian kernel density function is used to construct this spatial density model for volcanic vents of the Eastern Snake River Plain, Idaho. The Idaho National Laboratory (INL), a sprawling nuclear facility, is outlined in white.

Vents shown as black circles are mapped at the surface. Vents shown as white dots are only known from borehole data, because these vents are buried by subsequent lava flows from younger vents (The INL is located in a shallow valley that is gradually filling with lava flows). Including these "hidden vents" in the spatial density model increases the conditional probability of new vent formation estimated in the INL area. From [Wetmore et al., 2009].

Event definition

Event definition affects the total number of events used to estimate density, and may introduce bias in density estimates. This tends to diminish the weight associated with events in the center of the distribution in this particular case, as these events all formed multiple volcanoes.

Kernels functions

Spatial density estimates made using kernel functions, as opposed to hazard zonation models or parametric models, are explicitly data driven. A basic advantage of this approach is that any spatial density estimate will be consistent with the known data. Equations (3) and (4) are bivariate Gaussian kernels. Numerous authors have shown that the use of other kernels, such as the Epanechnikov kernel or the Cauchy kernel has little impact on the final density estimate. In hazard assessment, kernel functions with infinite tails (e.g. Gaussian) are preferred, as the probability is positive and real everywhere, albeit very small at locations far from past events. A picture of a kernel function, contoured around a single point, is shown in Figure 12. A potential disadvantage of these kernel functions is that they are not inherently sensitive to geologic boundaries. One might hope that a complete understanding of the geology would result in a modification of the density estimate derived from a mathematical function.

Kernel bandwidth

Bandwidth selection is a key feature of kernel density estimation, and is particularly relevant to volcanic hazard studies [Bebbington, 2013, Jaquet et al., 2008]. Bandwidths that are narrow focus density near past events. Conversely, a large bandwidth may over-smooth the density estimate, resulting in unreasonably low density estimates near clusters of past events, and overestimate density far from past events. This dependence on bandwidth can create ambiguity in the interpretation of spatial density if bandwidths are arbitrarily selected.

Bivariate bandwidth selectors like the SCV and SAMSE methods appear to be very promising, because although they are mathematically complex, they find optimal bandwidths using the actual data locations, removing subjectivity from the process. The bandwidth selectors used in this chapter provide global estimates of density, in the sense that one bandwidth or bandwidth matrix is used to describe variation across the entire region. An alternative method is to use adaptive kernel estimates, in which case the bandwidth changes with event density. These adaptive bandwidths are calculated assuming radially symmetric kernel functions. Future research will likely involve developing bandwidth selectors that are adaptive across the map region.

Sparse event data

Often in hazard assessment there is a "problem" that there are few data available from which to forecast future events. That is, often hazard assessments are needed for places where events are not so frequent that the geologic hazards are completely obvious. Instead, hazard analysis is most often required were few geologically hazardous events have occurred in the past. This is paradoxical because, by definition, uncertainty in hazard assessments must be comparatively high in these regions. If a spatial density is estimated using thousands of earthquakes or hundreds of volcanoes, we can assume that the true density is well-represented by this model. Conversely, if the spatial intensity estimate is based on a handful of events, we might expect high uncertainty in the estimate. For example, the discovery of a single additional volcano, buried in sediment, might alter the shape of the estimated regional spatial density.

A spatial density model for Lassen volcano

Volcanoes in the Lassen system and adjacent areas are broadly classified as “silicic" or “mafic" based on the chemistry and mineralogy of the lavas. We performed a spatial density analysis on these two groups separately, then added the maps together, renormalizing. This only works if the spatial density is weighted by the recurrence rate of volcanism, which is thought to be different for the two groups.

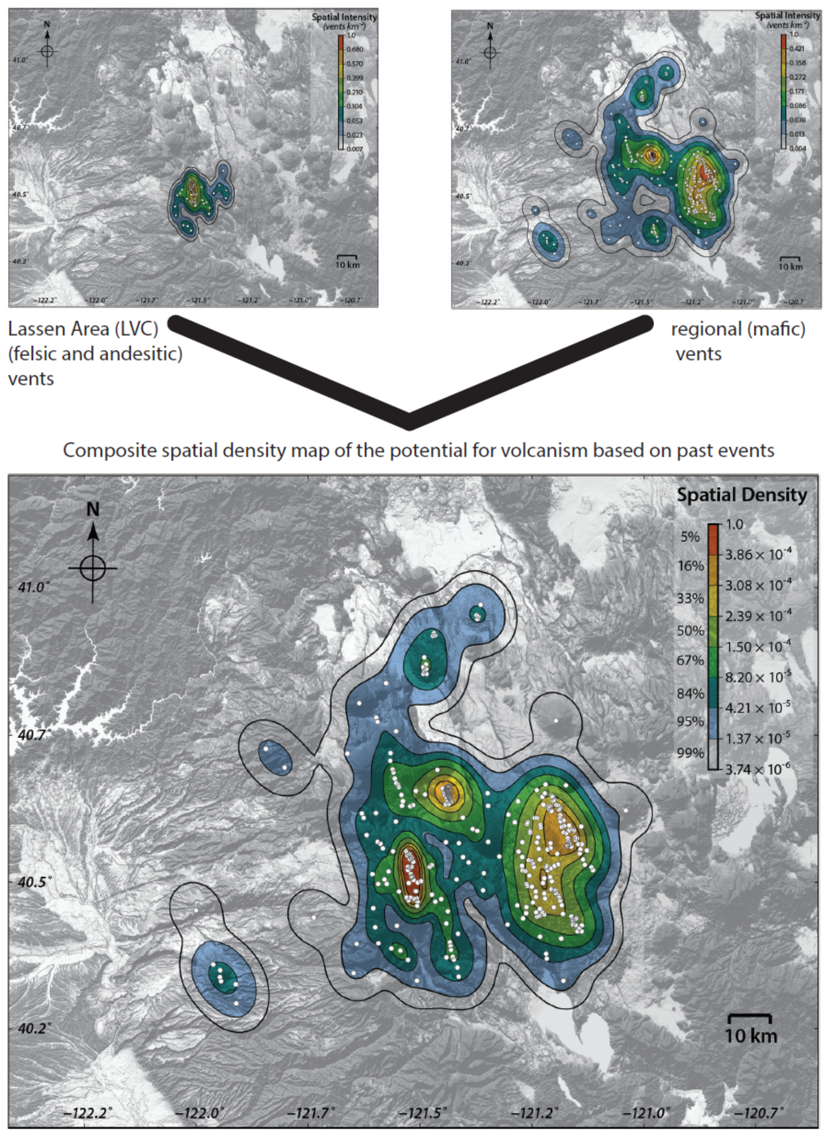

Figure 14: A spatial density model for Lassen volcano and the surrounding region. Felsic and andesitic volcanic vents tightly cluster in the area of Lassen Peak and Chaos Crags. Mafic volcanism is more widely distributed, and most clustered in the area of the Poison Lake Chain of craters, also known as the Caribou volcanic field, East of Lassen Peak. How to weight different types of volcanoes, perhaps with different magma source regions and different hazards, is an important area of research.

References

Mark S Bebbington. Assessing spatio-temporal eruption forecasts in a monogenetic volcanic field. *Journal of Volcanology and Geothermal Research*, 252:14–28, 2013.

Mark S Bebbington and Shane J Cronin. Spatio-temporal hazard estimation in the Auckland Volcanic Field, New Zealand, with a new event-order model. *Bulletin of Volcanology*, 73(1):55–72, 2011.

CD Condit and CB Connor. Recurrence rates of volcanism in basaltic volcanic fields: An example from the Springerville volcanic field, Arizona. *Geological Society of America Bulletin*, 108(10):1225–1241, 1996.

C Connor, M Bebbington, and W Marzocchi. Probabilistic volcanic hazard assessment. In *The Encyclopedia of Volcanoes*, pages 897–910. Springer, 2015.

CB Connor, J Stamatakos, D Ferrill, BE Hill, G Ofoegbu, and FM Conway. Volcanic hazards at the proposed Yucca Mountain, Nevada, high-level radioactive waste repository. *J. Geophys. Res*, 105:417–432, 2000.

Charles B Connor and Brittain E Hill. Three nonhomogeneous Poisson models for the probability of basaltic volcanism: application to the Yucca Mountain region, Nevada. *Journal of Geophysical Research: Solid Earth*, 100(B6):10107–10125, 1995.

Fanghui Deng, Chuck Connor, Rocco Malservisi, Laura Connor, Jeremy White, Aurelie Germa, and Paul Wetmore. A geophysical model for the origin of volcano vent clusters in a Colorado Plateau volcanic field. *Journal of Geophysical Research*, 122:in press, 2017.

Peter Diggle. A kernel method for smoothing point process data. *Applied Statistics*, pages 138–147, 1985.

Tarn Duong and Martin Hazelton. Plug-in bandwidth matrices for bivariate kernel density estimation. *Journal of Nonparametric Statistics*, 15(1):17–30, 2003.

Tarn Duong et al. ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. *Journal of Statistical Software*, 21(7):1–16, 2007.

Anthony C Gatrell, Trevor C Bailey, Peter J Diggle, and Barry S Rowlingson. Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British geographers*, pages 256–274, 1996.

Peter Hall, JS Marron, and Byeong U Park. Smoothed cross-validation. *Probability Theory and Related Fields*, 92(1):1–20, 1992.

Olivier Jaquet, Charles Connor, and Laura Connor. Probabilistic methodology for long-term assessment of volcanic hazards. *Nuclear Technology*, 163(1):180–189, 2008.

K. Kiyosugi, C. B. Connor, D. Zhao, L. J. Connor, and K. Tanaka. Relationships between temporal-spatial distribution of monogenetic volcanoes, crustal structure, and mantle velocity anomalies: An example from the Abu monogenetic volcano group, southwest Japan. *Bulletin of Volcanology*, 71:doi:10.1007/s00445–009–0316–4, September 2010.

Timothy M Lutz and James T Gutmann. An improved method for determining and characterizing alignments of pointlike features and its implications for the Pinacate volcanic field, Sonora, Mexico. *Journal of Geophysical Research: Solid Earth*, 100(B9):17659–17670, 1995.

Kazuaki Nakamura. Volcanoes as possible indicators of tectonic stress orientation – principle and proposal. *Journal of Volcanology and Geothermal Research*, 2(1):1–16, 1977.

DM Pyle. Sizes of volcanic eruptions. *Encyclopedia of Volcanoes*, 1: 263–269, 2000.

Melody G Runge, Mark S Bebbington, Shane J Cronin, Jan M Lindsay, Catherine L Kenedi, and Mohammed Rashad H Moufti. Vents to events: determining an eruption event record from volcanic vent structures for the Harrat Rahat, Saudi Arabia. *Bulletin of Volcanology*, 76(3):804, 2014.

Bernard W Silverman. *Density Estimation for Statistics and Data Analysis*, volume 26. CRC press, 1986.

BW Silverman. Choosing the window width when estimating a density. *Biometrika*, 65(1):1–11, 1978.

Paul H Wetmore, SS Hughes, Laura J Connor, and ML Caplinger. Spatial distribution of eruptive centers about the Idaho National Laboratory. In C. B. Connor, N. A. Chapman, and L. J. Connor, editors, *Volcanic and Tectonic Hazard Assessment for Nuclear Facilities*, pages 229–256. Cambridge University Press, 2009.