Bayesian Inference In Forecasting Volcanic Hazards:

An Example From Armenia

by

Jennifer N. Weller

A thesis submitted in partial fulfillment
of the requirement for the degree of
Master of Science
Department of Geology
College of Arts and Sciences
University of South Florida

Major Professor: Charles B. Connor, Ph.D.
Sarah Kruse, Ph.D.
Robert Watts, Ph.D.

Date of Approval:
November 9, 2004

Keywords: baye's theorem, anpp, kernel function, spatial denisty, gravity

Dedication

To Chuck: Undertake something that is difficult; it will do you good. Unless you try to do something beyond what you have already mastered, you will never grow. - Osborn

Acknowledgements

I would like to thank the Armenian National Academy of Sciences for excellent field and collaborative support. This work is supported by a NATO research grand and the USF foundation. Maps were created using GMT (Wessel and Smith).

In addition, I would like to thank my advisor, Chuck Connor, for his continued time, help and support in all aspects of this thesis and for constantly reminding me that I was capable of working through it. Thanks also to my committee members, Sarah and Rob, for their editorial comments and time. Furthermore, I would like to thank the following people for their immeasurable support, without which this thesis would not have been possible: Tim, Patty, Don, and Jamie.

Table of Contents

List of Figures

Bayesian inference in forecasting volcanic hazards:

An example from Armenia

Jennifer N. Weller

ABSTRACT

Scientists worldwide are increasingly faced with the need to assess geologic hazards for very infrequent events that have high consequence, for instance, in siting nuclear facilities for volcanic hazards. One of the methods currently being developed for such assessments is the Bayesian method. This paper outlines the Bayesian technique by focusing on the volcanic hazard assessment for the Armenia Nuclear Power Plant, (ANPP), which is located in a Quaternary volcanic field. The Bayesian method presented in this paper relies on the development of a probabilistic model based on the spatial distribution of past volcanic events and a geologic process model.

To develop the probabilistic model a bivariate Gaussian kernel function is used to forecast probabilities based on estimates of λ_t, temporal recurrence rate, and λ_s, spatial density. Shortcomings often cited in such purely probabilistic assessments are that it takes into account only known features and the event, new volcano formation, is rare and there is no opportunity for repeated experiments or uniform observations, the hallmarks of classical probability. One approach to improving such probabilistic models is to

incorporate related geological data that reflect controls on vent distribution and would improve the accuracy of subsequent models.

Geophysical data indicate that volcanism in Armenia is closely linked to crustal movement along major right lateral strike-slip fault systems that generates transtension across region. The surface expression of this transtension is pull-apart basins, filled with thick deposits of sediment, and antithetic normal faults. Volcanism in Armenia is concentrated in these deep sedimentary basins as is reflected in regional gravity data surveys. This means that low gravity anomalies are likely good indicators of future volcanic activity and therefore would improve probabilistic hazard models. Therefore, gravity data are transformed into a likelihood function and combined with the original probability model in quantitative fashion using Bayesian statistics. The result is a model that is based on the distribution of past events but modified to include pertinent geologic information. Using the Bayesian approach in this example increases the uncertainty, or range in probability, which reflects how well we actually know our probability estimate. Therefore, we feel it is appropriate to consider a range in probabilities for volcanic disruption of the ANPP, 1-4 x 10^{-6} per year (t=1 yr). We note that these values exceed the current International Atomic Energy Agency standard, 1 x 10^{-7} per year by at least one order of magnitude.

Chapter 1

Introduction

Scientists worldwide are increasingly faced with the need to assess hazards associated with point-like features such as volcanoes and earthquake epicenters on various temporal and spatial scales. Commonality among these phenomena exists because the analysis of their distribution and geologic setting can be used to estimate hazards quantitatively. Often, these geologic hazard assessments must evaluate the likelihood of very infrequent events that have high consequences (Haneberg 2000). For example, in the last two decades long-term probabilistic volcanic hazard assessment has increasingly been used in siting nuclear facilities worldwide (Crowe et al. 1982; Stamatakos and Ferrill 1996; Connor et al. 2000; McBirney et al. 2003; McBirney and Godoy 2003; Martin et al. 2004). Often, the central issue in these assessments is the likelihood of a new volcano forming by eruptions in close proximity to the facility. At such facilities, hazards with probabilities on the order of $10^{-6} – 10^{-8}$ per year are often considered high (Connor et al. 1995, Martin et al. 2004) because overall the risks associated with such facilities must be very low.

Geological hazard assessments for point-like features should present robust estimates of hazard rates, based on the frequency of past events and insights about the geological processes that control such events. One challenge associated with long-term

probabilistic assessment of future volcanism is that models of volcanic processes, such as the generation and ascent of magma, are inherently uncertain. One approach to making hazard assessments based on such models more robust is to modify probabilistic analyses by incorporating additional datasets through Bayesian inference (Von Mises 1957; Connor et al. 2000; Martin et al. 2004). Essentially, Bayesian inference allows us to combine two or more states of information (e.g., geophysical) to forecast the probability of volcanic events, such as formation of a new volcano, based on our understanding of volcanic systems, rather than solely based on the limited, and often incomplete, record of volcanic events. If we consider the frequency of volcanic events to be a physical property of a magmatic system, we are faced with the conclusion that the limiting value of the frequency of volcanic events is unknown. The event, formation of a new volcano, is rare and there is simply no opportunity for repeated experiments or uniform observations, the hallmarks of classical probability. Consequently, we are forced to update hazard forecasts using disparate observations of geologic and/or geophysical data that we believe impacts hazard forecasts. Bayesian inference provides a practical approach to incorporating such information.

In this paper, we analyze volcanological and geophysical data from Armenia with the goal of calculating the hazard associated with the disruption of the Armenian Nuclear Power Plant (ANPP) (Karakhanian et al. 2003), outline the technique, and illustrate the problems inherent to such analyses. We do this through the construction of an improved model that focuses on the probability of renewed volcanism that would impact the ANPP, by combining the probabilistic and geophysical models using Bayesian inference. The

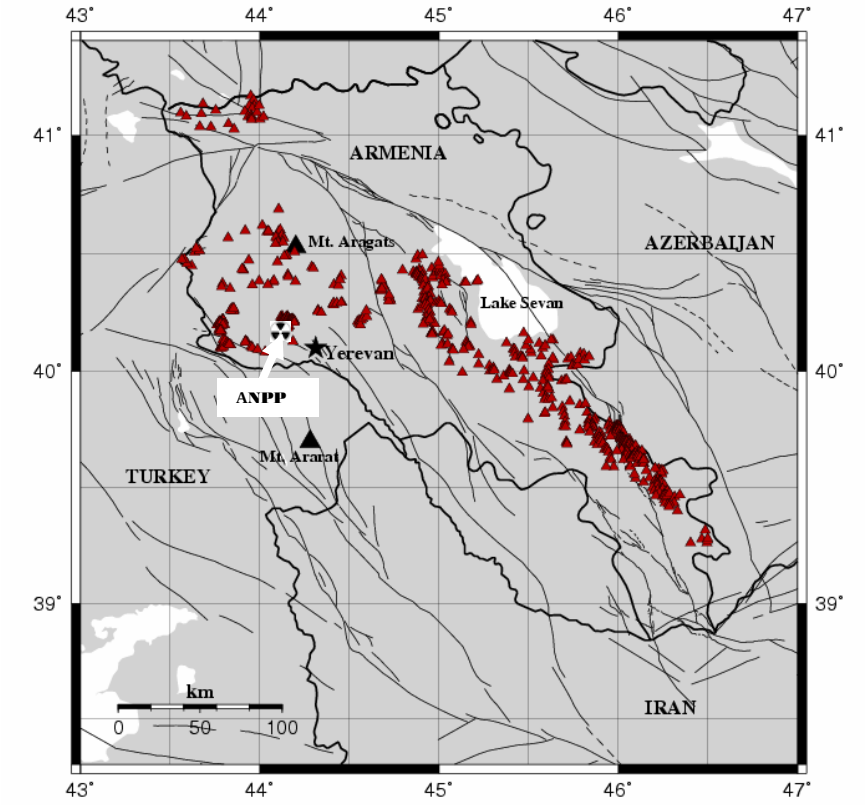


Figure 1. Location map of Armenia showing 554 Quaternary volcanoes used in study and faults. The rate of convergence just north of the ANPP is 18-19 mm/yr based on REVEL 2000 models. The ANPP is shown just south of a cluster of 38 cinder cones.

ANPP is located in the northwestern part of the Ararat Depression (a sedimentary basin between Mt. Aragats in the north and Mt. Ararat in the south) in close proximity to the town of Metzamor, and 28 km west of Yerevan, the capital of Armenia (Figure 1). The ANPP is a Chernobyl-style reactor that sits at the base of the southern foothills of Mt. Aragats, the largest composite volcano in Armenia. Mount Ararat, another large composite volcano in Turkey, is 55 km south of the ANPP. The ANPP is located on the Shamiram Volcanic Plateau and is only 1.3 - 6 km south of 38 small cinder cones arranged in four local clusters (Figure 2) (Karakhanian et al. 2003). An additional source of volcanic hazard for the ANPP and the capital city of Yerevan are the volcanoes of the Ghegam Ridge located 52 km to the east of the site and just west of

Figure 2. Photograph of the ANPP – clearly seen in the background are a cluster of cinder cones and to the far left the base of Mt. Aragats.

Lake Sevan (Figure 1). Some of these volcanoes have been dated as Holocene and their Late Pleistocene valley flow terminates 25 km east of the plant site. The most recent volcanic eruptions on the Ghegam Ridge have been dated between 4500 to 4400 $\pm$ 70 yr BP (Karakhanian et al. 2003).

Armenia is an appropriate choice for this type of analysis due to both its volcanic and tectonic setting. In the Quaternary (1.6 million years to the present), 554 basaltic to andesitic cinder cones (Savov et al. 2003) developed in response to mostly monogenetic activity. Monogenetic activity is characterized by the formation of a new volcano, such as a cinder cone or lava dome, and duration of volcanic activity at monogenetic volcanoes is thought to be typically less than 100 years (Connor and Conway 2000). After cessation of eruptive activity at any individual monogenetic volcano, renewed volcanism in the area builds a new monogenetic volcano. Thus, for this type of volcanism, the number of volcanoes reflects the number of volcanic events for which probabilistic forecasts are made. Because of the nature of this volcanic activity, the

volcanoes themselves can be considered point-like features (Figure 1) and the hazard assessment reduces to the problem of estimating the density distribution of these features. Examples of other hazard assessments for monogenetic volcanic activity include nuclear power plants and storage facilities (Connor et al. 1995; Karakhanian et al. 2003; Martin et al. 2004) and urban centers such as Auckland, New Zealand (Magill et al. 2004) and Mexico City (Bloomfield 1975; Martin del Pozzo 1982).

This distributed, monogenetic volcanism results from the complex tectonic history of the region that lies within a broad zone of deformation that forms part of the Alpine-Himalayan collision belt. Overall, volcanism describes an arc across Armenia (Figure 1) that is subparallel to this collision belt. Pull-apart basins can be delineated by mapping anomalies in the Earth's gravity field, caused by density variations between the sediments filling the basins and the surrounding crust (Tsuboi 1979). Presently, as a part of the Alpine fold belt, the uplift occurring across Armenia is a result of the northward motion of the Arabian plate with respect to Eurasia (Philip et al. 2001). The rate of convergence of these two plates is 18-19 mm/yr based on REVEL 2000 models (Dixon and Mao 2002). Volcanism across the region is linked to subduction and subsequent collision, and may result from slab steepening and breakoff which provides a viable mechanism for magma generation (Keskin 2003). In any case, volcanism is closely linked to N-S compression and E-W extension (Philip et al. 2001). The main geologic structures produced in this tectonic setting are north-west trending right-lateral strike-slip faults. These faults produce areas of transtension that create pull-apart basins within which volcanism is localized. Contrasts in crustal structure reflected in the distribution of

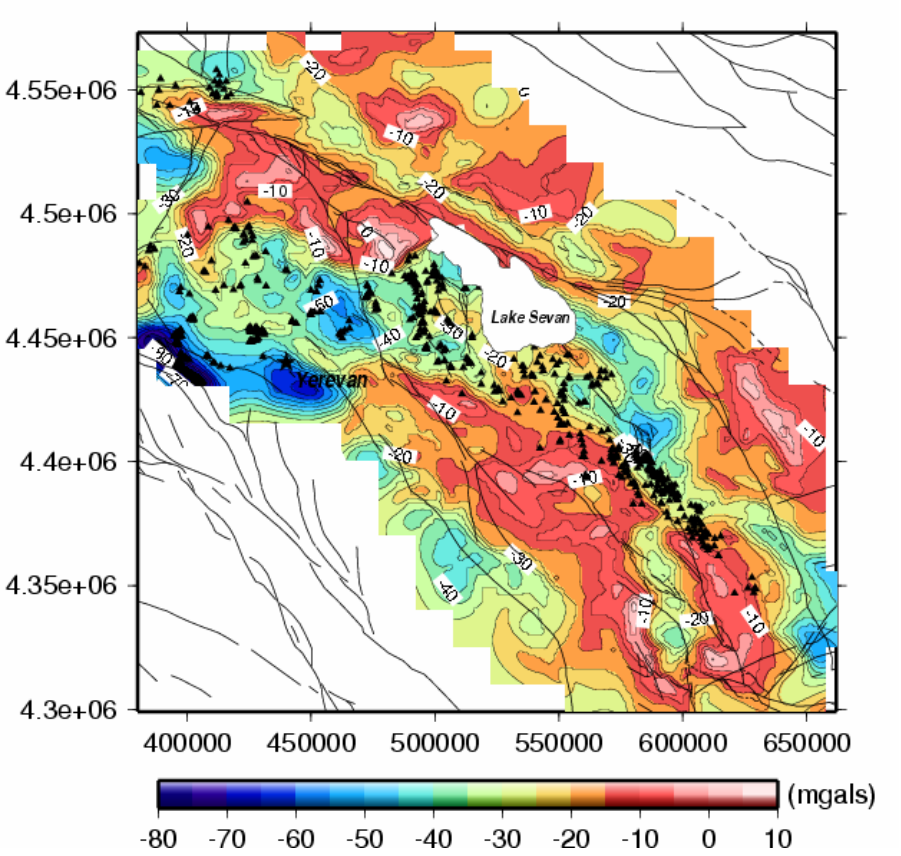


Figure 4. Gravity distribution across Armenia, coordinates are in UTM. Volcanoes typically occur in regions of broad gravity lows which correspond to deep sedimentary basins formed from extension and pull-apart mechanisms. The above observations point to a strong correlation between patterns of volcanism and structure.

gravity anomalies have been shown to influence both the generation and ascent of magma (Tamura et al. 2002) and monogenetic volcanism has been shown to preferentially occur in pull-apart basins and similar areas undergoing crustal extension (Connor and Conway 2000). This correlation between volcano distribution and gravity anomalies occurs in Armenia as well (Figure 3), and may result from magma generation by decompression melting of mantle previously enriched by subduction zone processes (Pearce 1990; Keskin 2003; Savov et al. 2003). Such a positive correlation between volcanism and gravity anomalies make it appropriate to consider gravity data in analysis of volcano distribution and preparation of volcanic hazard forecasts.

Chapter Two

Methods

In this paper we outline a method that allows geologic data to be cast in a way that permits systematic analysis of hazard at a site of interest. The simplest way to visualize the process is in chart form (Figure 4). The flowchart provides a step-by-step summary of the Bayesian process - from the initial step of defining databases to the final step of evaluating hazard at the location of interest. Each element of the flowchart relies on information about the region to construct a reasonable hazard model.

Defining the Database

Volcanic hazard assessments are only as good as the data used to create them. In this paper we use two data sets. The first dataset consists of the geographic coordinates of Armenian volcanoes. These coordinates were provided by the Armenia National Academy of Sciences and list all known Quaternary volcanoes in Armenia. The coordinates are provided in the universal transverse mercator (UTM) grid, WGS-84 datum. The second dataset comprises gravity data (Ohanissyan, 1985), also provided by the Armenia National Academy of Sciences. These gravity data consist of measurements of the relative change in the earth's gravity field, corrected for topographic effects. They

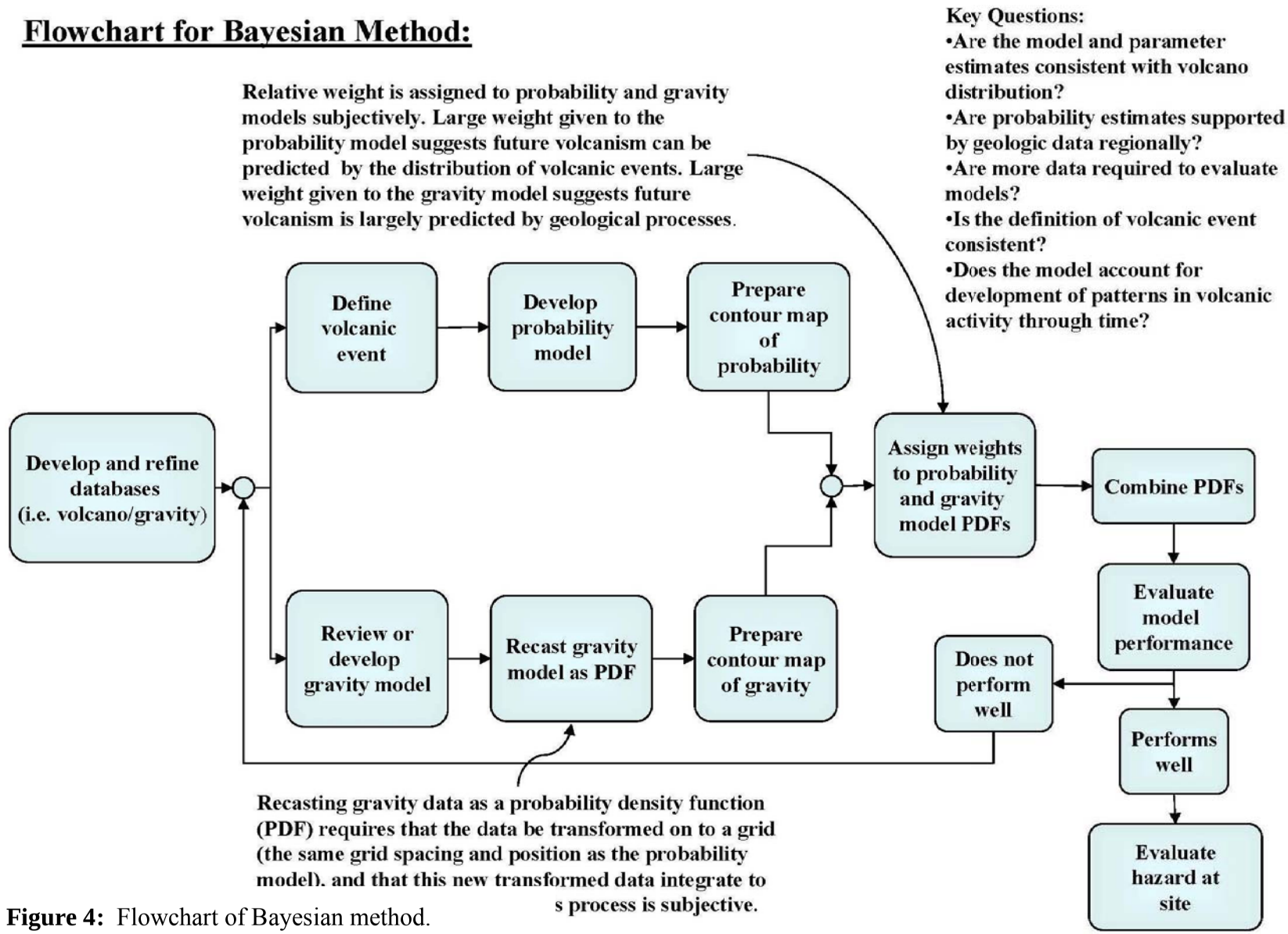


Figure 4: Flowchart of Bayesian method.

consist of geographic location, in the same UTM projection used in the volcano dataset, and relative change in gravity (mGal). Following this compilation, the two datasets follow somewhat different paths in the development of a probability model and a geophysical model until they are transformed so they can be combined using Bayesian methods (Figure 3). We will describe both these paths beginning with the development of the probability model.

Developing the Probability Model

The first and perhaps most important step in developing a probability model for volcanic hazard is to explicitly define events in the region of interest (Connor and Hill, 2000; Martin et al. 2004). Given the complexity of volcanic processes, what event is specifically forecast by the probability model? This is important for a number of reasons, not least of which is that the entire development of the probability analysis depends on the event being defined as a point process. Second, having defined the event as a point process, it is important to unambiguously define what constitutes an event and what does not. For instance, an event could be the epicenter of an earthquake, or the location of a sinkhole, or any other process that can be isolated to a precise location spatially. In the hazard assessments presented in this study, the definition of an event is limited to the formation of a new volcano, and is estimated directly from the distribution of existing Quaternary volcanoes and gravity anomalies. This assessment specifically does not look at the spatial or temporal distribution of volcanic eruptions at existing volcanoes or at the volcanic hazards associated with eruptions, such as lava flows, lahars, or pyroclastic flows. This is because the hazard assessment is intended to assess the probability of

eruptions from new volcanic vents in the area around the ANPP that would have a potential deleterious impact on operation of the nuclear facility, and containment of radionuclides (Crowe et al. 1982; McBirney and Godoy 2003).

The volcano dataset used in this study includes both monogenetic volcanoes (one eruptive sequence only) and polygenetic volcanoes (volcanoes that have more than one eruptive sequence). Volcano types include cinder cones, domes, composite volcanoes, maars, and calderas. Individual vents are treated equally in the analysis, regardless of volcano type, because we are modeling the probability of formation of a new volcanic vent. Estimates of the probability of dike injection, without volcano formation, are not considered in this hazard assessment but may be important in other types of hazard assessments, such as for high level waste repositories (Woods et al., 1999; Connor and Conway 2000). Formally, each volcanic event in this assessment is considered to be independent of the other volcanic events in the data set.

Alternative methods for defining volcanic events can be used in probabilistic volcanic hazard assessments. For example, polygenetic volcanoes may be weighted more heavily depending on the number of past eruptions (Martin et al. 2004). Also, closely spaced and similarly aged vents can be grouped together as a single event (Connor et al. 2000). Furthermore, only cones younger than a specific age, for example those formed in the last 100,000 years, might be included in the analysis as volcanic events or these could be weighted more heavily. Because vents may be grouped into single volcanic events in varying ways depending on their timing, distribution, and episodes of activity, data sets

can be defined in different ways which would change the probability of an event occurring. In this case, we use a simple definition in which each mapped volcano (Figure 1a) is a single event, partly because the additional data required by alternative definitions does not yet exist for Armenia.

Once the event is explicitly defined, we move on to the mathematical development of the probability model (Figure 3). Ultimately, the probability of an event occurring at a specific location is given by

$$\mathrm{P}[N \geq 1] = 1 - e^{-\lambda_t \lambda_s A t} \tag{1}$$

where A is the effective area of interest (e.g., the area within which volcanism would impact the ANPP, should it occur). In this study we define A as 50 km^2 around the power plant. The time period of interest is given by the parameter t, (e.g., the duration of expected operation of the ANPP). We estimate a 100 year time period for the expected operation of the ANPP. Once these site specific parameters are specified, we are left with the problem of how to estimate λ_s, spatial density (events per kilometer) and λ_t, temporal recurrence rate (events per year).

For Armenian volcanism, we use a simple estimate for temporal recurrence rate,

$$\lambda_t = \frac{N-1}{t_0 - t_y} \tag{2}$$

because the ages of individual volcanoes are not well constrained. In this equation, N is the total number of volcanoes, t_0 is the age of oldest event, and t_y is the age of the youngest event. We know that all the volcanoes used in this study are all of Quaternary age so we set t_0 equal to 1.6 million years and t_y to 0, or present. This gives λ_t a value of 3.5×10^{-4} years. That is, we expect one event every 2900 years. Where ages of volcanic events are better constrained, more detailed analyses have revealed time trends in volcanic activity and other methods for estimating λ_t have been employed (Ho 1991a,b; Condit and Connor 1996).

Compared with λ_t, estimation of spatial density, λ_s, is more difficult. However, pioneering work by Diggle (1985) and Silverman (1986) has led to the development of a kernel estimation technique that has been used in several studies to estimate spatial density of volcanism, including in the Pinacate Volcanic Field, Mexico (Lutz and Gutmann 1994), the Yucca Mountain region, USA (Connor and Hill 1995), and the Tohoku region, Japan (Martin et al. 2004). In this technique, spatial variation in λ_s is a function of distance to nearest-neighbor volcanoes and a smoothing parameter, h. The kernel function is a probability density function that is symmetric about the origin and spreads probability away from the event (Diggle 1985). Different kernel functions can be used including the Cauchy kernel (Martin et al. 2004), the Epanechnikov kernel (Lutz and Gutmann 1994), and the Gaussian kernel (Connor and Hill 1995). It is widely agreed that the shape of kernel function chosen in this type of analysis generally has a trivial

impact on probability calculations compared to other parameters (Connor and Hill 1995; Lutz and Gutmann 1994). The Gaussian function was chosen for the Armenian problem because volcanoes are treated as discrete events in time and space and the Gaussian model responds well to the patterns generally recognized in volcano distributions, such as clustering of vents. The probability surfaces generated by this model are continuous, as opposed to consisting of abrupt changes in probability that must be introduced in spatially homogeneous models (Connor and Hill 1995). Continuous probability surfaces allow for relatively easy comparison to other empirical data sets, e.g. gravity data that shed light on volcano distribution. The bivariate Gaussian kernel is given by:

$$\lambda_s(x,y) = \frac{1}{2\pi N h^2} \sum e^{-\frac{1}{2}\left(\frac{d_i}{h}\right)^2} \qquad (3)$$

where d_i is the distance from the point *x,y*, where λ_s is estimated, to the i^{th} volcano location, *h* is a smoothing parameter, and *N* is the number of volcanoes (points) that are used to estimate $\lambda_s(x,y)$. Due to the fact that *N* occurs in the denominator, the integral of $\lambda_s(x,y)$ across the map will be unity. Therefore the spatial density, λ_s, is a bivariate probability density function. Probability estimates made using equations 1 and 3 depend on the value chosen for *h*. Using a bivariate Gaussian kernel, events will have a high estimated probability in proximity to existing volcanoes if the value chosen for *h* is small, but low estimated probability away from the volcano. On the other hand, a large value of *h* will yield a more uniform estimate of probability distribution across the region. In the Gaussian kernel, the smoothing factor is equivalent to the standard deviation of a

symmetric, bivariate Gaussian distribution. Therefore, the kernel function depends on the assumption that the smoothing parameter is estimated in a geologically and/or statistically significant way (Connor and Hill 1995). In this study a wide range of smoothing constants are used (Figure 4); in addition, the range of reasonable smoothing constants is further constrained by use of the Clark-Evans (CE) spatial cluster analysis (Clark and Evans 1954). This analysis shows that volcanic events in Armenia are clustered across a variety of scales with >99% confidence (Blyth and Ripley 1980; Cressie 1991). Applying these tools, we use h=3000 m. As mentioned above, an additional assumption in the kernel estimation technique is the shape of kernel function chosen. This model gives an estimate of spatial density based on positions of volcanic vents for Armenia (Figure 5). The kernel model is useful because (1) probability maps can be made allowing ease of comparison with other geologic information; (2) there is no need to define zones of volcanic activity, as is required in a homogenous approach (Margulies et al. 1992); and (3) uncertainty in the distribution of individual events is easy to assess. The final step in developing the probability model is to prepare a contour map of spatial density (Figure 6). The map is made using a UTM projection for equal area measurements and the units of probability are typically converted into their logarithms because volcano vent density commonly varies by orders of magnitude across a region. Also, testing the suitability of the smoothing parameter chosen is appropriate at this point (Figure 6). The maximum probability of an event occurring is 3.4×10^{-4} where h has a value of 4000 m. More importantly, this plot demonstrates that probability is sensitive to the value of h chosen and varies rapidly from a maximum probability of 3.4×10^{-4} at

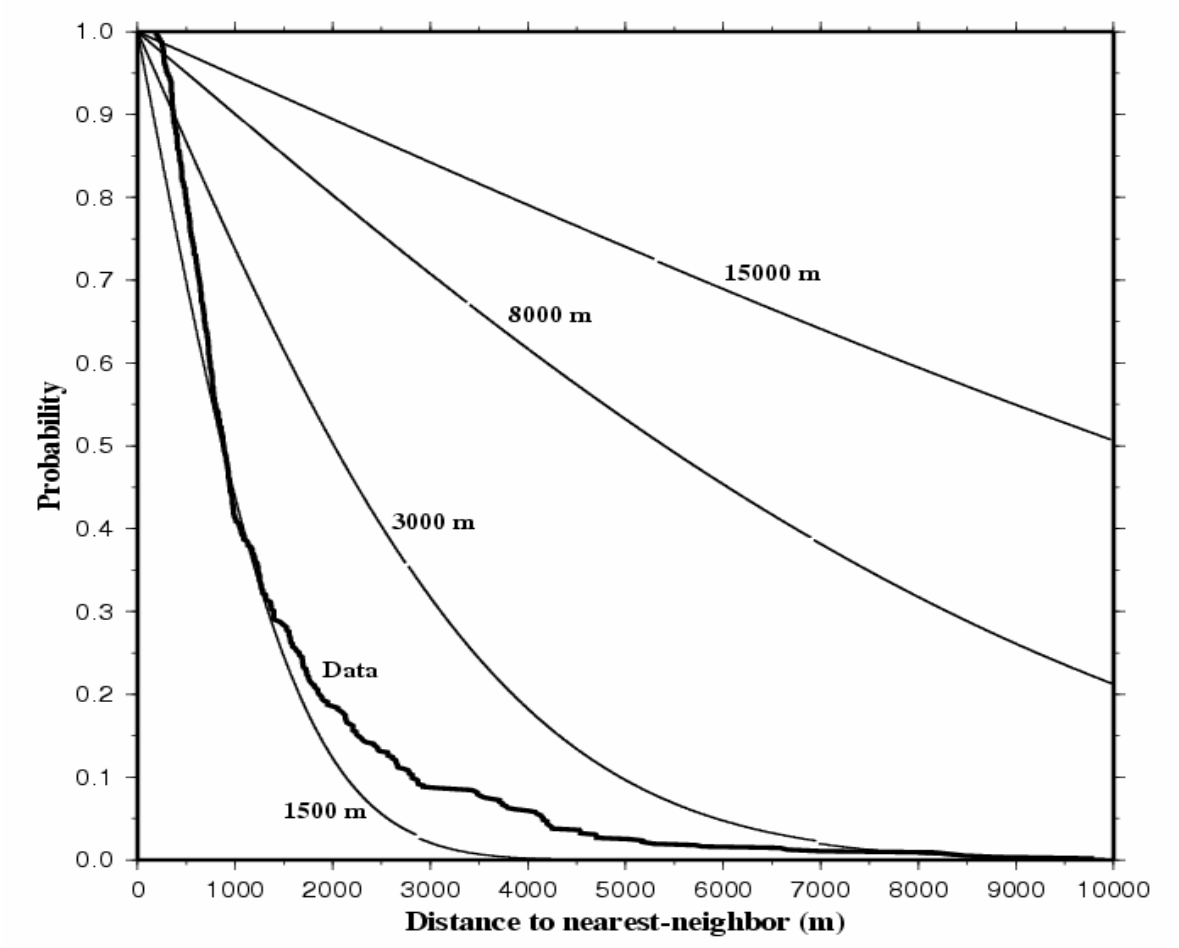


Figure 5. Modeling the data. Complimentary cumulative distribution function showing distance of nearest-neighbor volcano versus its cumulative probability of occurrence for the original data set and various values of smoothing parameter *h*. Notice that while *h*=1500 appears to fit the data well overall, it is a very poor fit in the tail of the distribution. Conversely, *h*=3000 fits the data well in the tail of the distribution.

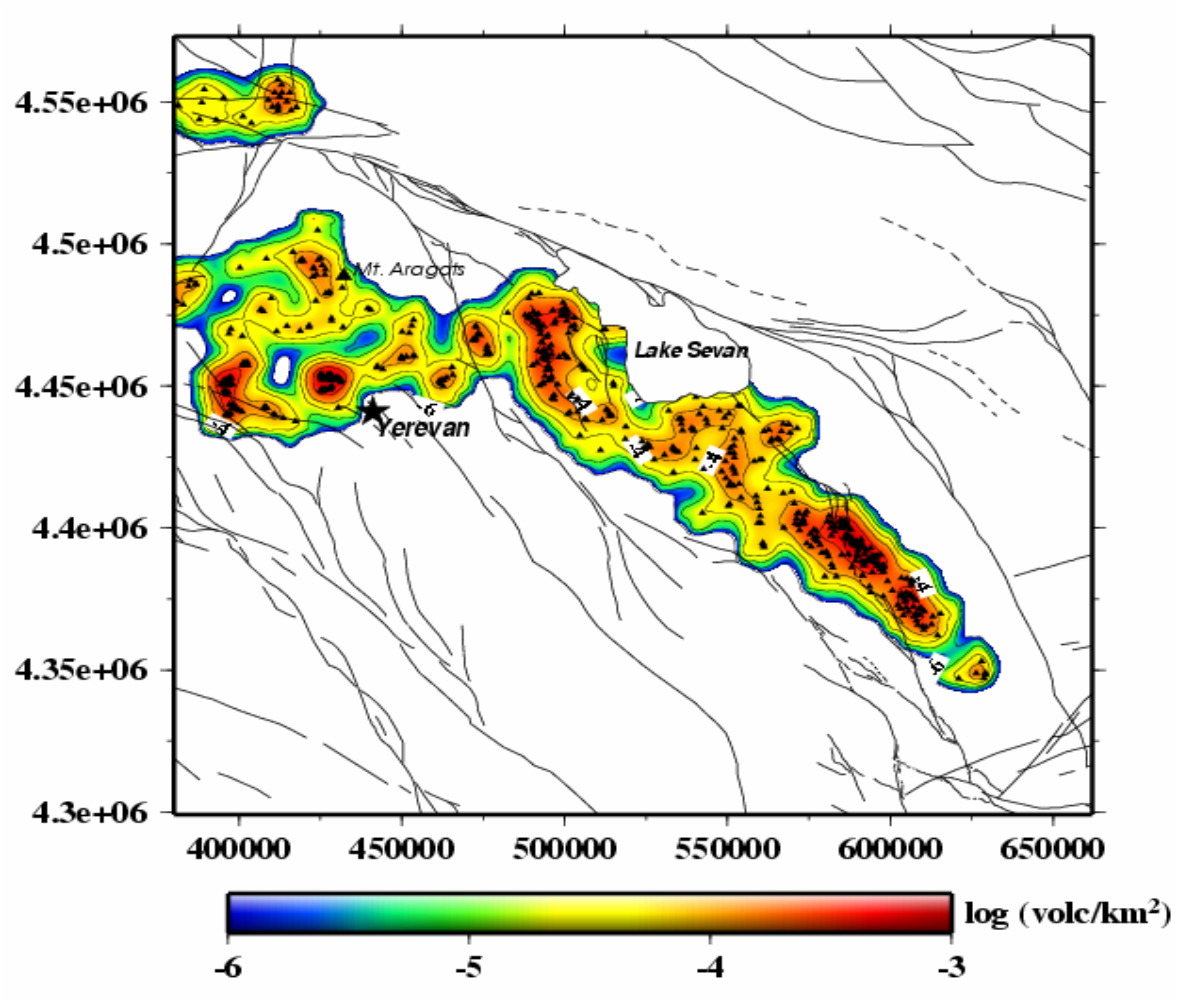


Figure 6. Map of spatial recurrence rate, coordinates are in UTM. Quaternary vent density is greatest in two volcano clusters in the Ararat Depression, including the region of the ANPP. Vent density varies by approximately three orders of magnitude across the area of interest.

4000 m to 7.5 x 10^{-5} at approximately 25,000 m. This sensitivity of the probability to the smoothing parameter is important to consider because we don't have a robust method for estimating *h* precisely. Using the above defined parameters, the acceptable range in probability of an event occurring that would disrupt the ANPP and using equation 1 is 1 to 3.4 x 10^{-4} for 1800 m < *h* < 17000 m (Figure 7).

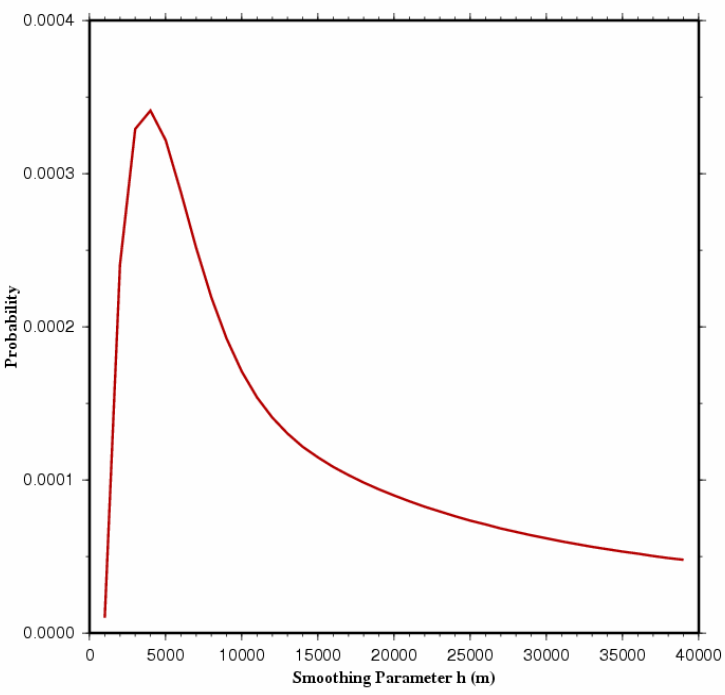


Figure 7. Plot showing values of *h* ranging from 1,000 m to 40,000 m versus the probability of an event occurring at the ANPP. The greatest hazard exists when *h* = 4,000 m.

Development of the Gravity Model

Although a probabilistic volcanic hazard assessment could be based solely on the methods outlined above, there are clear shortcomings. First, the distribution of volcanoes in the region may incompletely sample the possible distribution of volcanoes. In other words, because our sample of events is comparatively small, volcanism might yet occur in areas that have no record of previous activity. Second, the kernel estimation technique does not attempt to account for additional geologic information that might influence our assessment of the distribution of future volcanoes. Here, we take gravity

data and cast gravity anomalies as another probability density function in an effort to develop a more geologically realistic probability estimate, the ultimate goal being to modify λ_s with gravity information. However, to use the gravity data in a probability model, the gravity observations must be transformed into a likelihood function. This likelihood function, like λ_s, is a probability density function. However, before the likelihood function can be developed a relationship between volcano distribution and gravity must be shown to exist. A powerful tool for assessing the relationship between two distributions such as volcano distribution and gravity anomalies is the Kolmogorov-Smirnov (KS-test) goodness-of-fit test (Chakravarti and Roy 1967). In this case, gravity at each specific volcano location is

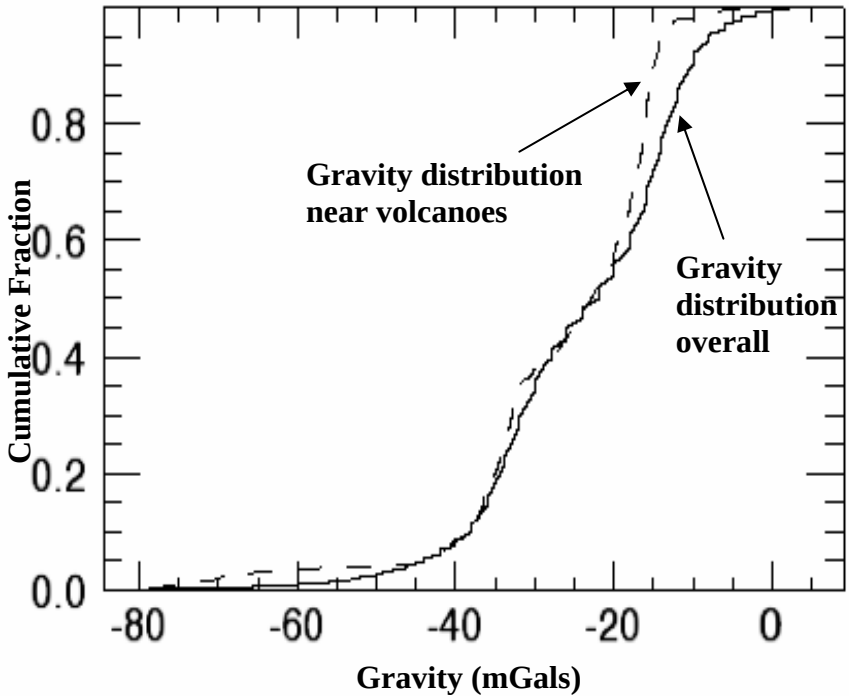


Figure 8. KS-test comparison cumulative fraction plot. Comparison of gravity distribution over the area of study and the gravity distribution near volcanoes shows that volcanoes tend to occur in areas of anomalously low gravity. A K-S test, based on the maximum difference in the two distributions, indicates that the tendency for volcanoes to occur in gravity lows is statistically significant (>99% confidence).

compared to gravity data collected across the entire region. Both distributions are plotted in cumulative form (Figure 8) and the maximum difference between the two distributions, the Kolmogorov-Smirnov statistic, (KS-statistic), is measured. For Armenian gravity data, the KS-statistic is 0.202 and the two distributions are different (P-value is ~ 0.0). This indicates volcanoes are clustered in low gravity regions, and this correlation suggests that gravity anomalies may be an indicator of the distribution of future volcanic activity. In other words, we need to consider how λ_s might be modified by the presence or absence of gravity anomalies.

Now that a relationship has been identified between volcano location and gravity using the K-S test, the next step is to transform the gravity data into a likelihood function. There is currently no standard method for developing the likelihood function and is therefore somewhat subjective as it is not purely based on statistics and relies to a certain extent on the expertise of those conducting the analysis (Martin et al. 2004). The first step in developing the likelihood function is to make some general observations about the relationship between gravity anomalies and volcano distribution. One observation is that there is an inverse relationship between the two distributions – that is, volcano density is high in areas with low gravity anomalies. Expounding on this observation, volcanoes only occur in areas where gravity is less than -10 mGal. This leads to the conclusion that the probability of volcanoes occurring in areas with gravity values greater than -10 mGal should be correspondingly small. Further observation shows that most volcanoes (about 90%) occur where gravity values are less than -15 mGal. This suggests that the gravity anomalies indicate a threshold in vent distribution. That is, volcanoes are equally likely

to occur anywhere gravity anomalies are less than some threshold value (such as -15 mGal) and not likely to occur at all in areas with gravity values greater than -10 mGal. In this case, we develop a step function to transformation to map gravity values into the likelihood function using Boolean logic. For example:

if ($G(x,y)$ > -10 mGal) then $W(x,y)$ = 0.001
if (-15 mGal < G(x,y) < -10 mGal) then $W(x,y)$ = 0.01
if ($G(x,y)$ < -15 mGal) then $W(x,y)$ = 0.1

where $G(x,y)$ is the measured value of gravity at map location x,y, usually after interpolation onto a grid, and $W(x,y)$ is the weight assigned to gravity values based on the observed distribution of volcanoes with respect to gravity anomalies (Figure 8). Adding additional if-then statements smoothes the mapped transformation. Using the weights derived from the above process, gravity values are transformed into a likelihood function:

$$L(\theta \mid x, y) = \frac{G(x, y)W(x, y)}{\sum_{XY} G(x, y)W(x, y)} \tag{4}$$

where $L(\theta \mid x, y)$ is the likelihood function, which integrates to unity over the region of interest XY, and θ is the set of weighted gravity values. Clearly, the way in which $L(\theta \mid x, y)$ is calculated is a reflection of the geologic interpretation of the data and the experimenter's belief about geologic processes governing magma generation, ascent, and eruption.

Finally, having developed and confirmed the validity of the gravity model, it is then normalized, recast as a probability density function, and set to the same grid spacing as the probability model. As with the probability model, the final step in developing the gravity model is to prepare a contour map of normalized gravity. The map is made using

a UTM projection for equal area measurements and the units of gravity are in mGal (Figure 2).

Bayesian Model

The final step in the Bayesian process is to combine the probability PDF (a priori function) of spatial density and gravity PDF (likelihood function) into a joint probability density function, or a posteriori PDF for weighted spatial density. The product (or intersection) of these two states of information is given by

$$\lambda_s\ (x, y \mid \theta) = \frac{\lambda_s(x, y)L(\theta \mid x, y)}{\iint_R \lambda_s(x, y)L(\theta \mid x, y)dxdy} \tag{5}$$

where x and y represent grid point locations in the region of interest A, θ is the gravity data, $\lambda_s(x,y)$ is the spatial density, which is modified by $L(\theta|x,y)$, the gravity PDF, or likelihood function. The posteriori PDF is normalized to unity by integrating over the entire volcanic field so that cumulative probability does not change, but the shape of the 2-D surface distribution will be modified by the gravity PDF (Figure 9) (Martin et al., 2004).

A limitation with the standard Bayes rule above in equation (5) is the inability to weight the respective PDFs. Further, since it is conditional probability, probabilities will always be negligible or zero in regions where there are no or extremely sparse volcanic events, irrespective of how irrefutably geophysical or other information show the potential of new volcanic events to form. We therefore modify equation (5) by combining PDFs through addition (or union) and assign weights:

$$\lambda_s(x, y \mid \theta) = \frac{a\lambda_s(x, y) + (1-a)L(\theta \mid x, y)}{\int\int_R a\lambda_s(x, y) + (1-a)L(\theta \mid x, y)dxdy} \quad (6)$$

where *a*, is the weighting function of factor. Relative weight is assigned to the probability and gravity PDF's subjectively. Giving a large weight to the probability model suggests that future volcanism can be predicted almost exclusively by the distribution of volcanic events. A large weight applied to the gravity model suggests that future volcanism is better predicted by gravity anomalies. Setting $a = 0.50$, gives both PDF's equal weight and produces the same result as achieved using equation 4 (Figure 8). Two problems arise from using this method. First, the aforementioned subjectivity in assigning weights to the PDF's and second, a step function must be introduced to determine how much weight should be given to individual gravity measurements. This process is also somewhat subjective but can be given credibility by looking at the graph of volcano fraction versus gravity (Figure 8). The observation that volcanoes cluster in regions under 20 mGals lends credence to weighting the step function heavily in areas with gravity values under this threshold. For values over this threshold gravity values are given increasingly less weight in the step function. When the parameter *a* is given equal weight in equation 6, we achieve the same result as using the Baye's Theorem equation (5). Now, the new value $\lambda_s(x,y|\theta)$ is used in equation 1, rather than the unmodified value $\lambda_s(x,y)$. For the ANPP the probability of volcanic disruption is 1×10^{-4} using Bayesian inference. Interestingly, this is lower than the probability calculated using the unmodified value $\lambda_s(x,y)$ of 4×10^{-4}.

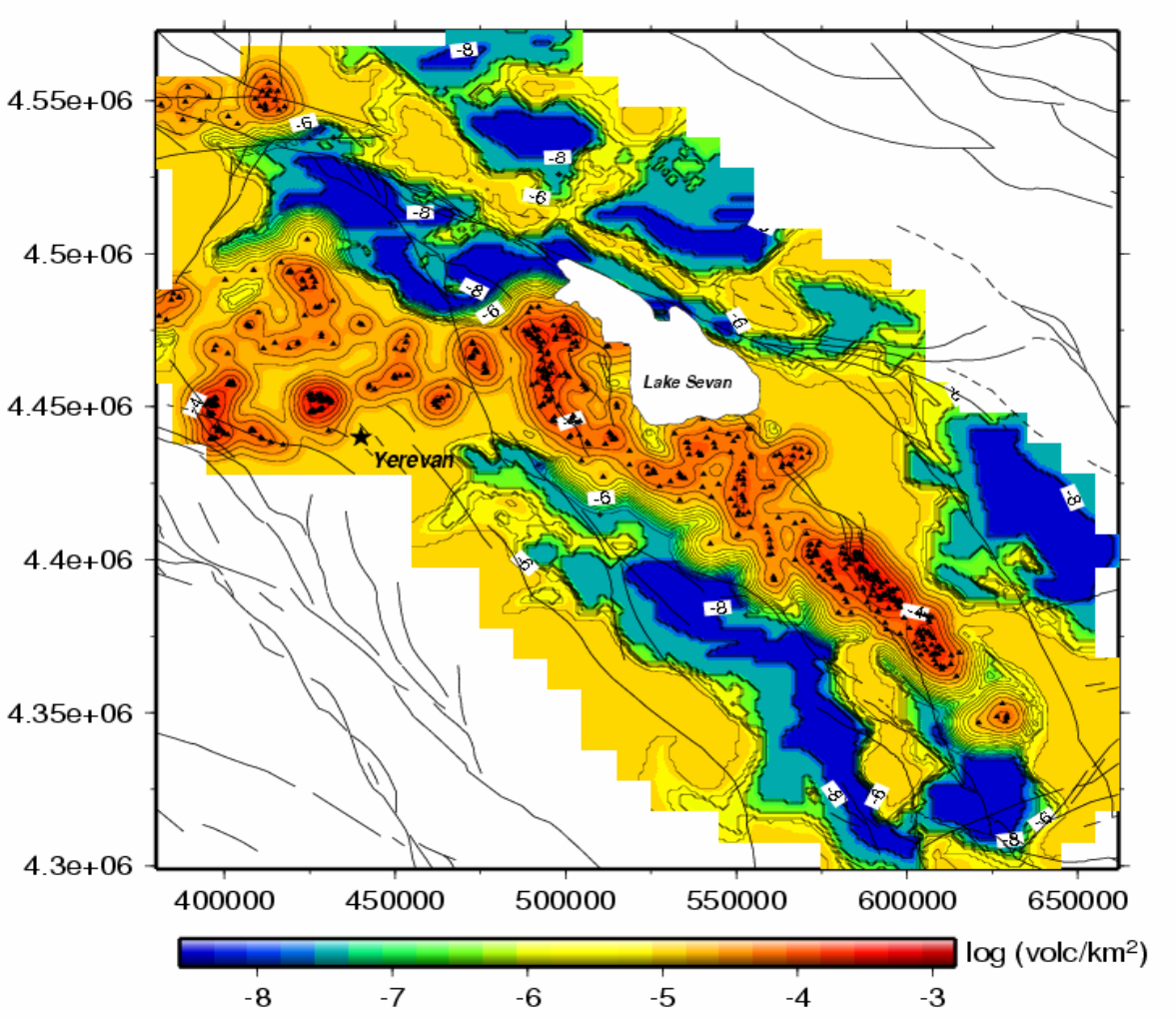


Figure 9. Bayesian output. Map showing results from equation 5 and equation 6 (parameter *a* is 0.50) where the spatial recurrence rate PDF and the gravity likelihood function are combined.

Chapter Three

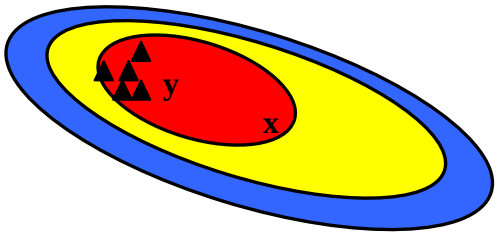
Discussion

Von Mises (1957), in his treatise on probability and statistics, shows that it is possible to develop an accurate model of spatial density, given enough events. That is, provided the experiment can be performed many times. In geologic hazard assessments we have only one spatial experiment to work with, in this case the observed distribution of volcanoes, and models developed to estimate hazards solely based on these data are inherently uncertain. The risk in using these models is that the distribution of past events may poorly reflect the distribution of potential future events. Any additional information that sheds light on the spatial distribution is therefore worth analyzing.

A model based solely on spatial distribution of past events cluster probability around known events and does not predict a future event at positions far from the cluster (Figure 10a). However, a geologic model, such as one based on gravity anomalies, allows probability of future events to be modified to account for our understanding of the geologic setting of volcanism. For example, we know that volcanoes in Armenia have not formed in the Quaternary where gravity values are higher than -10 mGal, but do occur through regions with gravity values below -15 mGal. Taking this into account modifies the probability model (Figure 10 b,c). By combining the statistical model with gravity

Spatial Density Model: $\lambda_s(x,y) = \frac{1}{2\pi N h^2} \sum e^{-\frac{1}{2}\left(\frac{d_i}{h}\right)^2}$

Gravity Model: $L(\theta \mid x,y) = \frac{G(x,y)W(x,y)}{\sum_{XY} G(x,y)W(x,y)}$



Bayesian Model:

$\lambda_s(x,y \mid \theta) = \frac{a\lambda_s(x,y) + (1-a)L(\theta \mid x,y)}{\int\int_R a\lambda_s(x,y) + (1-a)L(\theta \mid x,y)dxdy}$

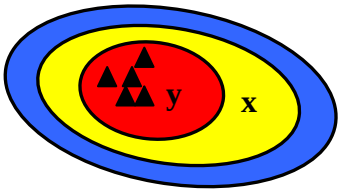


Figure 10 a,b,c. Stylized diagram showing (a) spatial density model; (b) gravity model; and (c) Bayesian model that combines the spatial and gravity models. Red areas correspond to a high probability of a future volcanic event, yellow areas to moderate probability of an event, and blue areas to regions of low probability of an event. Notice the probability associated with location *x* and location *y*. In the spatial density model (a) *x* falls in a region of low probability and *y* a region of high probability, i.e. an event would not be expected to occur at location *x* but would be expected at location *y*. In the gravity model (b) both *x* and *y* fall in a regions of low gravity (or high probability) and an event would be equally likely in either location. In the Bayesian model (c) an event is still likely at position *y* but there is a moderate probability of an event at location *x*. This model reflects the most information available for assessing the hazard at both points.

data through Bayesian inference we arrive at a different, and, if our geologic models are good, an improved volcanic hazard forecast.

There are several clear limitations to a purely statistical approach to hazard assessment. One being that this approach takes into account only known features. For instance, in Armenia volcanoes form in association with deep pull-apart basins that are a result of transtension and are associated with high rates of sedimentation. The majority of volcanoes in Armenia are monogenetic and hence do not have a great deal of topographic relief. It is conceivable that some Quaternary vents are buried beneath sediment in these basins, and are not accounted for in the estimate of spatial density, but should be. Furthermore, the distribution of volcanoes in the region may incompletely sample the possible distribution of volcanoes, i.e. because our sample of events is statistically small, volcanism could occur in areas that have no previous record of activity. Conversely, abrupt changes in geologic structure are common, especially at active plate margins, such as in Armenia. Estimates of spatial density from the point distribution may inadvertently indicate that areas of very different geology are equally likely to host future volcanic activity.

In this context, there are several advantages to the Bayesian approach. First, this approach steers us toward a geological basis for making probabilistic hazard assessments. In this case we use gravity anomalies as the geologic base from which we modify our probability model. We use gravity because (1) other studies have shown that distributed, monogenetic volcanism often correlates with gravity anomalies (e.g., Connor et al, 2000); (2) the tectonic setting of Armenia and resulting geologic features supports the idea that

this correlation should occur in Armenia; and (3) the K-S test statistically shows there is correlation between volcanic events and gravity. However, gravity is not unique – other pertinent geologic information could be used either alone or in combination with other data to enhance probabilistic hazard assessments. These include seismic tomographs (Martin et al., 2004), geochemical data (Condit and Connor, 1996), fault and other structural data, and magnetic data.

Second, this approach raises the issue that there is no standard method for development of the likelihood function. Transforming geophysical observations into PDF's and deciding how best to combine them is a subjective process that relies on the experimenter's interpretation of geological data and on general observations about the relationship between datasets.

Third, uncertainty in probability estimates may actually increase using the Bayesian approach. For instance, the range of probabilities may increase because of uncertainty in how to create the likelihood function, and in how much weight to assign the likelihood function. Since the posteriori function for $\lambda_s(x, y)$ is a PDF, increasing probability in regions of low gravity with no volcanoes decreases probability in the immediate vicinity of volcano clusters. The amount of change depends on the weighting of geologic models. Rather than being a negative outcome, increase in uncertainty more accurately reflects our understanding of the geology. In the case of the ANPP, the probability of volcanic disruption based solely on kernel estimation techniques is 3.4 x 10^{-4} for volcanic events impacting the site (A=50 km^2, t = 100 yr, h=3000 m). When the kernel estimation technique is modified with gravity using Bayesian inference, the

probability of volcanism decreases to 1 x 10^{-4}. If other models of volcano formation are used to assess volcanic hazards to the ANPP, the probability may change, depending on both the geologic model and probability model chosen. Perhaps a more useful estimate for this site is simply stating that a range of models based on spatial density of volcanoes and additional geologic information yield probabilities of volcanic eruptions impacting the ANPP site of 1-4 x 10^{-6} per year (t=1 yr). These values, whilst apparently low on human timescales, do in fact exceed the current International Atomic Energy Agency standard, 1 x 10^{-7} per year (McBirney and Godoy, 2003), by at least one order of magnitude.

Chapter 4

Conclusions

The approach presented here points to the applicability of assessing point-like geologic features, in this case volcanoes, through spatial statistics, specifically, Bayesian statistics. This study shows that geologic structure controls the distribution of volcanism in Armenia. Regionally, Quaternary volcanism is concentrated in pull-apart basins associated with low gravity anomalies, as a result of crustal extension and decompression melting of an enriched mantle. Gaussian kernel estimates of spatial density are greatest in two clusters in the Ararat Depression, including the cluster in proximity to the ANPP. Furthermore, vent density varies by three orders of magnitude across the region. These models lead to estimates of probability of volcanic disruption to the ANPP of between 1 x 10^{-4} and 3 x 10^{-4} in 100 years, which exceeds the current IAEA standard of 1 x 10^{-5} in 100 years. This variation indicates that modification of models through the incorporation of additional geologic information, may increase the range in probability estimates. Ultimately, this paper provides a pathway towards the incorporation of geologic process models in volcanic hazard assessments by allowing these models to be combined with more traditional probabilistic models through Bayesian inference.

References

Bloomfield, K. 1975, a late-Quaternary monogenetic volcanic field in central Mexico. *Geologische Rundshau*, v. 64, pp. 476-497.

Byth, K., and B.D. Ripley. 1980, On sampling spatial patterns by distance methods, *Biometrics*, v. 36, pp. 279-284.

Chakravarti, Laha and Roy. 1967, Handbook of Methods of Applied Statistics, Volume I. John Wiley and Sons, pp. 392-394.

Clark, P.J. and F.C. Evans. 1954, Distance to nearest-neighbor as a measure of spatial-relationships in populations. Ecology, v. 35, pp. 445-453.

Condit, C.D. and C.B. Connor. 1996, Recurrence rates of volcanism in basaltic volcanic fields: an example from the Springerville volcanic field, Az, USA. *Geological Society of America, Bulletin*, v. 108, pp. 1225-1241.

Connor, C.B. and F.M. Conway. 2000, Basaltic Volcanic Fields. Chapter in: Encyclopedia of Volcanology, Academic Press, 331-343.

Connor, C.B.; Stamatakos, J.A.; Ferrill, D.A.; Hill, B.E.; Ofoegbu, G.I.; Conway, M.F.; Sagar, B.; and J. Trapp. 2000, Geologic factors controlling patterns of small-volume basaltic volcanism: Application to a volcanic hazards assessment at Yucca Mountain, Nevada. *Journal of Geophysical Research*, v. 105, n. 1, pp. 417-432.

Connor, C.B. and B.E. Hill. 1995, Three nonhomogeneous Poisson models for the probability of basaltic volcanism: Application to the Yucca Mountain region. *Journal of Geophysical Research*, v. 100 (B6), n. 10, pp. 107-125.

Connor, C.B., Condit, C.D., Crumpler, L.S., and J,C, Aubele 1992, Evidence of regional structural controls on vent distribution, Springerville volcanic field, Arizona, *Journal of Geophysical Research*, v. 97, pp. 349-359.

Cressie, Noel. 1991, Statistics for Spatial Data. John Wiley and Sons, Inc., New York, 900.

Crowe, B.M.; Johnson, M.E.; and Beckman, R.J. 1982; Calculation of the probability of volcanic disruption of a high-level radioactive waste repository in southern Nevada, USA. *Radioactive Waste Management*, v. 3, pp. 167-190.

Diggle, P.J. 1985, A kernel method for smoothing point process data. *Applied Statistics*, v. 34, pp. 138-147.

Dixon, T. H. and A. Mao. 2002, REVEL: a model for recent plate velocities from space geodesy. *Journal of Geophysical Research*, v. 107, pp.1-32.

Haneberg, W. 2000, Deterministic and probabilistic approaches to geologic hazard assessment. *Environmental and Engineering Geoscience*, v. 6, n. 3, pp. 209-226.

Ho, Chih-Hsiang. 1991a, Time trend analysis of basaltic volcanism for the Yucca Mountain site. *Journal of Volcanology and Geothermal Research*, v. 46, pp. 61-72.

Ho, C-H. 1991b, Eruptive probability calculation for the Yucca Mountain site, USA: statistical estimation of recurrence rates. *Bulletin of Volcanology*, v. 54, pp. 50-56.

Karakhanian, A. Djrbashian, R., Trifonov, V., Philip, H., Arakelian, S., Avagian, A., Baghdassaryan, H., Davtian, V., and Y. Ghoukassyan. 2003, Volcanic hazards in the region of the Armenian Nuclear Power Plant. *Journal of Volcanology and Geothermal Research*, v. 126, pp. 31-62.

Keskin, M. 2003, Magma generation by slab steepening and breakoff beneath a subduction-accretion complex: an alternative model for collision-related volcanism in Eastern Anatolia, Turkey. *Geophysical Research Letters*, v. 30, n. 24, pp.?.

Lutz, T.M., and J.T. Gutmann. 1994, An improved method of determining alignments of point-like features and its implications for the Pinacate volcanic field, Mexico. *Journal of Geophysical Research*, in press.

Magill, C.; Blong, R.; and J. McAnency. 2004, Volcanic loss for the Auckland region. Abstract: Workshop on Statistics in Volcanology, Bristol, England, March 23, 2004.

Margulies, T., Lancaster, L.,Eisenberg, N., and L. Abramson. 1992, Probabilistic analysis of magma scenarios for assessing geologic waste repository performance, Rep. 92-WA/SAF-11. *American Society of Mechanical Engineering*, New York.

Martin, Andrew J., Umeda, Koji; Connor, Charles B.; Weller, Jennifer N.; Zhao, Dapeng; and Masaki Takahashi. 2004, Modeling long-term volcanic hazards through Bayesian inference: an example from the Tohoku arc, Japan. *Journal of Geophysical Research*, v. 109, B10208, pp. 1-20.

Martin, del Pozzo, A.L. 1982, Monogenetic volcanism in Sierra Chichinautzin, Mexico. *Bulletin Volcanologique*, v. 45, pp. 9-24.

McBirney, A., Serva, L.,Guerra, M., and C.B. Connor. 2003, Volcanic and seismic hazards at a proposed nuclear power site in central Java. *Journal of Volcanology and Geothermal Research*, 126, pp. 11-30.

McBirney, A. and A. Godoy. 2003, Notes on the IAEA guidelines for assessing vocanic hazards at nuclear facilities. *Journal of Volcanology and Geothermal Research*, 126, pp. 1-9.

Ohanissyan, Sh. S., 1985. Structure of the Earth Crust in Armenia , Geophysical fields and structure of the Earth crust in the Trans-Caucasus, Nauka, Moscow, pp. 43-63 [in Russian]

Pearce, J.A., Bender, J.F., DeLong, S.E., Kidd, W.S.F., Low, P.G., Guner, Y., Saroglu, F., Yilmaz, Y., Moorbath, S., and J.G. Mitchell. 1990, Genesis of collision volcanism in Eastern Anatolia, Turkey. *Journal of Volcanology and Geothermal Research*, v. 44, pp. 189-229.

Philip, H., Ritz, J.F., S. Rebai. 2001, Estimating slip rates and recurrence intervals for strong earthquakes along an intracontinental fault: example of the Pembak-Sevan-Sunik fault (Armenia). *Tectonophysics*, v. 343, n. 3-4, pp. 205-232.

Savov, I.P., Connor, C., D'Antonio, M., and Y. Jrbashian. 2003, Volcano-tectonic interactions in NE Armenia – results from petrology, geochemistry, and Quaternary volcano and fault distribution. Presented at the *RIDGE 2000-Norvolk* Volcano-Tectonic Interaction workshop in Iceland.

Sheridan, M.F., Hubbard, B., Bursik, M.I., Siebe, C, Abrams, M., Macias, J.L., and H. Delgado. 2001, Short-term potential volcanic hazards at volcan Popocatepetl, Mexico, *EOS, American Geophysical Union*, 82:185-189.

Silverman, B.W. 1986, Density Estimation for Statistics and Data Analysis. 175 pp., Chapman and Hall, New York.

Stamatakos,, J.A. and D.A. Ferrill. 1996, Tectonic processes in the central Basin and Range region, in NRC High-Level Radioactive Waste Research at CNWRA. July-December 1995, CNWRA 95-02S, edited by B. Sagar, Cent. for Nucl. Waste Regul. Anal., 61 – 6-25, San Antonio, Texas.

Tamura, Y.; Tatsumi, Y.; Zhao, D., Kido, Y.; and H. Shukuno. 2002, Hot fingers in the mantle wedge: new insights into magma genesis in subduction zones. *Earth and Planetary Science Letters*, v. 197, pp. 105-116.

Tsuboi, Chuji. Gravity. George Allen and Unwin, 1979. London, UK.

Von Mises, R. 1957, Probability, statistics, and truth. Dover Publications, Inc. NY.

Winkler, K., and A. Nur. 1979, Pore fluids and seismic attenuation in rocks. *Geophysical Research Letters*, v. 6, pp. 1-4.

Appendices

Appendix A: Armenia Volcano Location Data

43.56	41.09	43.91	40.43	44.16	40.21
43.59	41.08	43.90	40.44	44.17	40.21
43.68	41.13	43.90	40.44	44.17	40.22
43.67	41.09	43.79	40.38	44.17	40.22
43.67	41.04	43.79	40.36	44.17	40.22
43.73	41.04	43.83	40.35	44.18	40.22
43.75	41.11	43.62	40.45	44.18	40.22
43.82	41.05	43.59	40.46	44.18	40.22
43.86	41.03	43.58	40.48	44.19	40.22
43.95	41.17	43.56	40.48	44.19	40.22
43.96	41.15	43.64	40.51	44.19	40.22
43.94	41.13	43.66	40.51	44.19	40.22
43.97	41.13	43.66	40.52	44.19	40.21
43.96	41.11	43.64	40.53	44.15	40.19
43.92	41.10	44.10	40.69	44.15	40.19
43.93	41.08	44.29	40.44	44.16	40.19
43.95	41.09	44.30	40.44	44.16	40.20
43.95	41.08	44.41	40.37	44.15	40.19
43.95	41.07	44.44	40.37	44.12	40.18
44.00	41.07	44.45	40.41	44.13	40.18
43.99	41.10	44.45	40.39	44.13	40.18
44.02	41.08	44.32	40.26	44.13	40.19
44.00	41.13	44.32	40.26	44.12	40.18
43.82	40.57	44.34	40.25	44.13	40.18
43.92	40.60	44.41	40.29	44.13	40.19
44.01	40.62	44.41	40.28	44.79	40.43
44.05	40.59	44.41	40.29	44.87	40.48
44.04	40.59	44.43	40.29	44.88	40.49
44.05	40.60	44.46	40.29	44.91	40.50
44.11	40.60	44.45	40.29	44.68	40.38
44.11	40.60	44.55	40.21	44.69	40.37
44.12	40.58	44.55	40.21	44.70	40.36
44.13	40.57	44.55	40.22	44.69	40.37
44.09	40.57	44.56	40.22	43.84	40.26
44.09	40.57	44.56	40.20	43.85	40.26
44.10	40.55	44.57	40.23	43.85	40.27
44.13	40.55	44.60	40.23	43.85	40.26
44.09	40.54	44.59	40.26	43.85	40.26
44.14	40.49	44.12	40.24	43.84	40.26
44.15	40.49	44.12	40.23	43.84	40.27
44.13	40.49	44.12	40.23	43.85	40.27
44.10	40.47	44.12	40.22	43.79	40.21
44.19	40.51	44.12	40.22	43.79	40.22
44.16	40.41	44.13	40.22	43.79	40.22
44.16	40.40	44.13	40.22	43.79	40.22
44.07	40.38	44.15	40.23	43.79	40.22
44.08	40.39	44.16	40.23	43.80	40.22
44.04	40.37	44.15	40.23	43.76	40.20
44.00	40.38	44.15	40.23	43.77	40.20
43.96	40.37	44.15	40.22	43.77	40.21
43.94	40.47	44.16	40.23	43.77	40.21

Appendix A (Continued)

43.78	40.21	44.87	40.41	44.96	40.16
43.78	40.20	44.87	40.42	45.02	40.12
43.78	40.20	44.87	40.42	45.01	40.11
43.79	40.20	44.88	40.42	45.05	40.12
43.79	40.20	44.88	40.41	45.07	40.10
43.79	40.20	44.89	40.40	45.08	40.09
43.78	40.21	44.89	40.40	45.05	40.04
43.80	40.20	44.90	40.40	45.02	40.17
43.79	40.20	44.91	40.41	45.03	40.17
43.77	40.18	44.90	40.40	45.04	40.15
43.77	40.17	44.91	40.39	45.05	40.16
43.78	40.17	44.91	40.38	44.96	40.44
43.79	40.17	44.92	40.39	44.96	40.40
43.78	40.18	44.91	40.36	44.97	40.41
43.78	40.10	44.92	40.34	44.99	40.42
43.78	40.10	44.94	40.36	44.99	40.44
43.78	40.11	44.94	40.35	45.00	40.42
43.79	40.10	44.95	40.33	45.00	40.43
43.79	40.10	44.94	40.32	45.00	40.43
43.80	40.10	44.90	40.31	45.00	40.43
43.80	40.12	44.92	40.31	45.00	40.44
43.80	40.13	44.92	40.29	45.01	40.42
43.79	40.14	44.92	40.31	45.00	40.45
43.80	40.13	44.92	40.31	45.01	40.41
43.82	40.12	44.93	40.30	45.03	40.42
43.84	40.11	44.93	40.30	45.04	40.39
43.82	40.12	44.94	40.29	45.02	40.39
43.83	40.12	44.95	40.30	45.04	40.38
43.93	40.12	44.88	40.26	45.01	40.37
43.92	40.13	44.90	40.27	45.06	40.27
43.92	40.13	44.92	40.28	45.07	40.26
43.92	40.13	44.94	40.28	45.07	40.25
44.18	40.13	44.93	40.25	45.22	40.03
43.95	40.11	44.94	40.26	45.22	40.07
43.96	40.10	44.95	40.25	45.15	40.13
43.99	40.09	44.93	40.23	45.16	40.12
44.03	40.08	44.95	40.23	45.16	40.11
44.67	40.35	44.94	40.22	45.17	40.11
44.68	40.34	44.98	40.29	45.16	40.10
44.68	40.35	45.00	40.30	45.14	40.09
44.68	40.35	45.00	40.29	45.18	40.21
44.72	40.32	45.01	40.29	45.18	40.20
44.73	40.31	44.92	40.21	45.14	40.38
44.73	40.30	44.92	40.21	45.15	40.38
44.72	40.30	44.91	40.20	45.21	40.38
44.72	40.33	44.97	40.20	45.22	40.39
44.89	40.43	44.96	40.17	45.00	40.47
44.90	40.43	44.95	40.16	45.38	40.12
44.89	40.42	44.96	40.16	45.41	40.10
44.86	40.42	44.96	40.16	45.42	40.09

Appendix A (Continued)

45.42	40.08	45.61	39.99	45.91	39.72
45.42	40.06	45.61	39.97	45.89	39.71
45.46	40.06	45.58	39.96	45.90	39.71
45.46	40.06	45.56	39.93	45.91	39.70
45.46	40.06	45.58	39.91	45.91	39.69
45.51	40.08	45.60	39.94	45.92	39.67
45.51	40.09	45.61	39.92	45.94	39.67
45.48	40.11	45.61	39.93	45.94	39.67
45.57	40.06	45.68	39.96	45.95	39.65
45.64	40.05	45.70	39.96	45.95	39.63
45.58	40.10	45.71	39.96	45.96	39.62
45.47	40.16	45.73	39.87	45.97	39.59
45.50	40.13	45.49	39.80	45.93	39.59
45.56	40.12	45.59	39.89	46.14	39.53
45.59	40.11	45.59	39.88	46.09	39.54
45.64	40.13	45.61	39.88	46.09	39.56
45.63	40.14	45.59	39.82	46.06	39.59
45.63	40.06	45.59	39.82	46.07	39.59
45.61	40.03	45.63	39.84	45.95	39.78
45.61	40.01	45.63	39.89	45.94	39.77
45.72	40.05	45.69	39.85	45.96	39.76
45.73	40.06	45.69	39.83	45.95	39.75
45.75	40.03	45.70	39.81	45.97	39.75
45.74	40.02	45.70	39.78	45.97	39.73
45.77	40.03	45.79	39.86	45.96	39.72
45.77	40.03	45.78	39.75	45.99	39.77
45.79	40.08	45.69	39.76	45.99	39.77
45.81	40.07	45.70	39.76	45.99	39.76
45.78	40.05	45.71	39.70	46.00	39.76
45.81	40.05	45.71	39.69	46.01	39.77
45.28	40.02	45.71	39.69	46.00	39.76
45.28	40.00	45.82	39.86	45.99	39.74
45.30	40.01	45.82	39.82	45.99	39.76
45.30	40.01	45.83	39.79	45.99	39.75
45.31	40.00	45.84	39.79	45.99	39.75
45.39	40.00	45.87	39.80	46.01	39.76
45.47	40.00	45.86	39.79	46.01	39.75
45.13	40.00	45.83	39.78	46.01	39.75
45.32	39.97	45.86	39.77	46.02	39.75
45.36	39.98	45.86	39.77	46.03	39.74
45.39	40.00	45.86	39.76	46.01	39.73
45.40	39.99	45.85	39.76	46.01	39.72
45.40	39.99	45.84	39.76	46.01	39.72
45.41	39.99	45.83	39.76	46.01	39.71
45.47	39.96	45.84	39.74	46.02	39.71
45.39	39.93	45.88	39.73	46.02	39.72
45.50	39.93	45.89	39.74	46.02	39.71
45.52	39.97	45.90	39.74	46.02	39.71
45.59	40.00	45.87	39.70	46.03	39.71
45.59	39.99	45.88	39.69	46.03	39.70

Appendix A (Continued)

46.03	39.70	46.12	39.66	46.20	39.49
46.03	39.69	46.12	39.65	46.22	39.49
46.03	39.70	46.10	39.64	46.25	39.49
46.04	39.69	46.10	39.60	46.26	39.49
46.06	39.71	46.13	39.66	46.23	39.48
46.05	39.69	46.12	39.63	46.27	39.49
46.05	39.70	46.12	39.62	46.21	39.47
46.05	39.70	46.13	39.62	46.26	39.45
46.05	39.70	46.11	39.63	46.26	39.44
46.06	39.70	46.12	39.63	46.25	39.46
46.06	39.69	46.13	39.61	46.25	39.47
46.06	39.69	46.14	39.61	46.26	39.47
46.06	39.69	46.14	39.60	46.27	39.47
46.05	39.67	46.14	39.62	46.28	39.48
46.07	39.68	46.14	39.62	46.31	39.46
46.08	39.67	46.14	39.62	46.34	39.47
46.10	39.68	46.22	39.58	46.31	39.44
46.07	39.66	46.24	39.56	46.28	39.43
46.02	39.66	46.24	39.58	46.29	39.43
45.99	39.66	46.25	39.57	46.28	39.42
46.05	39.66	46.24	39.57	46.33	39.40
46.05	39.65	46.17	39.54	46.48	39.32
46.04	39.64	46.21	39.53	46.46	39.28
46.03	39.62	46.20	39.53	46.40	39.26
46.06	39.64	46.22	39.54	46.49	39.26
46.07	39.64	46.23	39.53	46.50	39.28
46.08	39.64	46.23	39.53	46.50	39.28
46.07	39.63	46.24	39.53	44.20	40.38
46.07	39.63	46.24	39.52	46.16	39.57
46.09	39.64	46.21	39.52	46.20	39.58
46.09	39.64	46.22	39.52	45.84	40.06
46.09	39.65	46.22	39.51	44.95	40.22
46.10	39.66	46.25	39.51	44.68	40.39
46.11	39.65	46.19	39.50		

Appendix B: Computer Codes

All codes for this project were written in the PERL computing language.

The shell script to run the Bayesian code is as follows:

```
#!/usr/bin/csh

#make your grid spacing = to one kilometer!
#the file small_lat_long_5000.xyz is output
#from, for example, the Gaussian kernel function code
# It is the output grid of coordinates and might be
# in UTM at some spacing like 5000 m
# the script pow.pl converts the log of
#probabilities to "normal" probability numbers
#then feed the results to pow.out

echo "perl pow.pl redone_volc_gauss_3000.dat > pow.out"
perl pow.pl redone_volc_gauss_3000.dat > pow.out

#the Bayesian method is all about "normalizing"
#so this means the map has to integrate to 1 (unity)
#run "does_it_sum_to_1.pl to figure this out

echo "perl does_it_sum_to_1.pl pow.out"
perl does_it_sum_to_1.pl pow.out

#go ahead and normalize the grid file of probability values to integrate to
#unity across the map (grid) region

echo "perl normalize_data.pl pow.out > normalized_pow.out"
perl normalize_data.pl pow.out > normalized_pow.out

#just check one more time
#remember this just prints the value of the summation

echo "perl does_it_sum_to_1.pl normalized_pow.out"
perl does_it_sum_to_1.pl normalized_pow.out

#develop the weighting function. Use the input gravity data here
#instead of small_grav001.dat

echo "perl grav_wt_function.pl detrended_grav_Xyz.dat > normalized_grav.dat"
```

```
perl grav_wt_function.pl grav_Xyz.dat > normalized_grav.dat
perl does_it_sum_to_1.pl normalized_grav.dat

echo "perl two_files.pl normalized_grav.dat normalized_pow.out >
normalized_grav_x_pow.out"
perl two_files.pl normalized_grav.dat normalized_pow.out >
normalized_grav_x_pow.out

echo "perl does_it_sum_to_1.pl normalized_grav_x_pow.out"
perl does_it_sum_to_1.pl normalized_grav_x_pow.out

echo "perl normalize_product.pl normalized_grav_x_pow.out > normalize_product.out"
perl normalize_product.pl normalized_grav_x_pow.out > normalize_product.out

echo "perl does_it_sum_to_1.pl normalize_product.out"
perl does_it_sum_to_1.pl normalize_product.out

echo "take log for plotting..."
perl log_normalize_product.pl normalize_product.out > log_normalize_product.out

echo "./bayes.gmt"
./bayes.gmt
```

The following are the individual scripts that make up the shell script above:

Script: pow.pl

```
#!/usr/bin/perl

while (<>) {

($a, $b, $data) = split;
$newdata = 10.0**$data;
print "$a, $b, $newdata\n";
}
```

Script: does_it_sum_to_one.pl

```
#!/usr/bin/perl

open (INPUT, $ARGV[0])||die "cannont read file!/n";
#ARGV[0] means read from the first file after .pl listed on command line, @data creates
an array and assigns variables to each column
```

```
while (<INPUT>) {


        $line = $_;
        $N=$N+1;



@data = split(" ", $line);

        $x=$data[0];
        $y=$data[1];
        $z=$data[2];

        $sum_z=$sum_z+$z;


}
print "$sum_z\n";
```

Script: normalize.pl

```
#!/usr/bin/perl


$sum_z=0; $N=0;
#initializes variables

open (INPUT, $ARGV[0])||die "cannont read file!/n";
#ARGV[0] means read from the first file after .pl listed on command line, @data creates
an array and assigns variables to each column
while (<INPUT>) {

        $line = $_;
        $N=$N+1;

        @data = split(" ", $line);

        $x=$data[0];
        $y=$data[1];
        $z=$data[2];

        $sum_z=$sum_z+$z;
```

```
}
#print "$sum_z\n";


close (INPUT);


open (INPUT, $ARGV[0])||die "cannot read file!/n";
while (<INPUT>)


{

        $line=$_;
        if ($line=~m/\d/)
        {
                @data= split (" ", $line);

                $x=$data[0];
                $y=$data[1];
                $z=$data[2];


                $w=$z/$sum_z;

                print "$x $y $w\n";

        }
        else
        {
                print "[$line]\n";

        }
}
```

Script: log_normalize_product.pl

```
#!/usr/bin/perl

while (<>) {

($a, $b, $data) = split;
```

```
$newdata = log($data)/log(10.0);
print "$a $b $newdata\n";

}
```

Script: grav_wt_function.pl

```
#!/usr/bin/perl

$sum_z=0; $N=0;
#initializes variables

open (INPUT, $ARGV[0])||die "cannont read file!/n";

#ARGV[0] means read from the first file after .pl listed on command line, @data creates
an array and assigns variables to each column
while (<INPUT>) {

        $line = $_;
        $N=$N+1;

        @data = split(" ", $line);

        $x=$data[0];
        $y=$data[1];
        $z=$data[2];


        if ($z > -10){$new_z = 0.01};
        if ($z < -12){$new_z = 0.1};
        if ($z < -16){$new_z = 1.0};
        if ($z < -18){$new_z = 5.0};
        if ($z < -20){$new_z = 20.0};
        if ($z < -22){$new_z = 30.5};
        if ($z < -24){$new_z = 35.0};
        if ($z < -26){$new_z = 40.0};
        if ($z < -28){$new_z = 45.0};
        if ($z < -30){$new_z = 50.0};


        $sum_new_z += $new_z;
```

Appendix B (Continued)

```
}
#print "$sum_z\n";

close (INPUT);


open (INPUT, $ARGV[0])||die "cannot read file!/n";
while (<INPUT>)
{

        $line=$_;
        if ($line=~m/\d/)
        {
                @data= split (" ", $line);

                $x=$data[0];
                $y=$data[1];
                $z=$data[2];

        if ($z > -10){$new_z = 0.01};
        if ($z < -12){$new_z = 0.1};
        if ($z < -16){$new_z = 1.0};
        if ($z < -18){$new_z = 5.0};
        if ($z < -20){$new_z = 20.0};
        if ($z < -22){$new_z = 30.5};
        if ($z < -24){$new_z = 35.0};
        if ($z < -26){$new_z = 40.0};
        if ($z < -28){$new_z = 45.0};
        if ($z < -30){$new_z = 50.0};


                $w=$new_z/$sum_new_z;

                print "$x $y $w\n";

        }
        else
        {
                print "[$line]\n";
```

```
        }
}
```

Script: two_files.pl

```
#!/usr/bin/perl

open(FILE1, $ARGV[0]) || die "Cannot open <$ARGV[0]> for input: [$@]";
open(FILE2, $ARGV[1]) || die "Cannot open <$ARGV[1]> for input: [$@]";
```

Appendix B (Continued)

```
$a=0.50;
while (<FILE1>) {

  ($a1, $b1, $data1) = split;
  $line = <FILE2>;
  ($a2, $b2, $data2) = split " ", $line;
  $total =($a* $data1)+((1-$a)*$data2);
  print "$a1  $b1  $total\n";

}

close FILE1;
close FILE2;
```

Script: Bayes.gmt (Note: This is a GMT code for mapping the output log_normalize_product.out.

```
#!/bin/bash

gmtset FRAME_PEN 2.0p
gmtset HEADER_FONT 5
gmtset LABEL_FONT 5
gmtset HEADER_FONT_SIZE 20
gmtset LABEL_FONT_SIZE 16
gmtset ANOT_FONT_SIZE 14

# surface does the interpolation with a minimum curvature algorithm
# -I2  means the grid spacing is 2  units in each direction
# -Gfilename indicates the output file
# -R specifies the west,east,south, and north bounds of the map
```

Appendix B (Continued)

```
surface -I1000 -V log_normalize_product.out -Gnormalized_prod.grd -
R380000/662000/4299000/4573000

# makecpt creates a color table (to color shade contours)
# -Cseis specifies the basic color table
# -T specifies the range and interval
makecpt -Cseis -T-8/-3.5/.25 -V -I > normalized_grav.cpt

#psmask clips or masks area of no data on a map
#-R gives range of data
#-B gives tickmark info
#-I grid spacing

psmask grav_utm.dat -R -I10000 -Jx0.000022 -B25000a50000/WSne -K -V -P -Y1.5 >
bayesian_product_50.ps


# grdimage plots the image (color map)
# -JX6.0i is the scale (must match the following)
# -Clat_long_5000.cpt is the color scale created with makecpt
# -P portait mode (must match the following)
# -E is the dpi (dots per inch) of the color shading
# -K more postscript to be appended to cn.ps in the following
grdimage normalized_prod.grd -Jx0.000022 -Cnormalized_grav.cpt -P -E100 -K -V -O

>> bayesian_product_50.ps

#grdcontour draws the contours from the grid
# -JX6.0i means the map will be 6 inches wide
# -C250 means there is a 250 nT contour interval
# -A500 means the contours are annotated every 500 nT
# -B25a50f5/WSne draw the frame, 25 m tick with 50 m label, add 5 m ticks, label on
#                west and south side only
# -W0.25p set line width
# -P draw in portrait mode
# -O overlay contours on the image
grdcontour normalized_prod.grd -Jx0.000022 -C.5 -A2 -L-8 -W0.25p -P -O -K -V >>
bayesian_product_50.ps


psmask -C -O -K -P -V >> bayesian_product_50.ps
```

Appendix B (Continued)

```
# psxy plots volcanic vents as solid black triangles
# -S specifies the size and shape
# - G0 makes the triangles solid
psxy volcano_type.dat -R -Jx0.000022 -St0.05i -G0 -O -P -K -V >>
bayesian_product_50.ps

#add a color scale to show the range of lat_long_5000 intensities plotted
psscale -D2.65/-.5/5/0.2h -Cnormalized_grav.cpt -B1/:"log (volc/km@+2@+)": -O -P -I -
V -K >> bayesian_product_50.ps


psxy armfaults_utm.dat -Jx0.000022 -R -W2.0 -O -M -V -K -P

>>bayesian_product_50.ps

psxy lake_sevan_utm.dat -Jx0.000022 -R -W1.25 -P -G255 -V -O -K
>>bayesian_product_50.ps

pstext -R -Jx0.000022 -G0 -O -P -V -P -K <<EOF>> bayesian_product_50.ps
522500 4456000 9 0 24 BL Lake Sevan
EOF

pstext -R -Jx0.000022 -G0 -O -P -V -K <<EOF>> bayesian_product_50.ps
443000 4430000 12 0 24 BL Yerevan
EOF

psxy -R -Jx0.000022 -O -P -G0/0/0 -Sa0.15i -W0.5 -V <<EOF>>
bayesian_product_50.ps
440000 4440000
EOF
```

The following script is used to calculate the recurrence rate which is then mapped in GMT.

```
#!/usr/bin/perl
#this is a perl script by Jenn Weller
#created on Feb. 24, 2003
#Purpose: this script reads volcano location data from a file and calculates the spatially
nonhomgeneous recurrence rate
#using a guassian kernal function.
```

Appendix B (Continued)

```
#The location of the point density estimate and the value of the point density estimate is
output (x y z);
#this output can be contoured in GMT.

#This code requires the input file to contain two columns of numbers giving location of
volcanoes.

#initialize variables used in calculations
$n=0; $x=0; $y=0; $h=3000;
@volcanoes=();

open (INPUT, $ARGV[0])||die "cannot access file!/n";

while (<INPUT>)
{


        $line = $_;
        if ($line=~m/\d/)
        {

                @data = split(" ", $line);

                $volcanoes[0][$n]=$data[0];
                $volcanoes[1][$n]=$data[1];
                #print "$volcanoes[0][$n]  $volcanoes[1][$n]\n";
                $n=$n+1;
        }
        else
        {


print "[$line]\n";
}
}
for ($y=4.573e+06; $y>4.29899e+06; $y-=1000){
        #steps through all y points on grid


                for ($x=380000; $x<662001; $x+=1000){
                #steps through all x points on grid
```

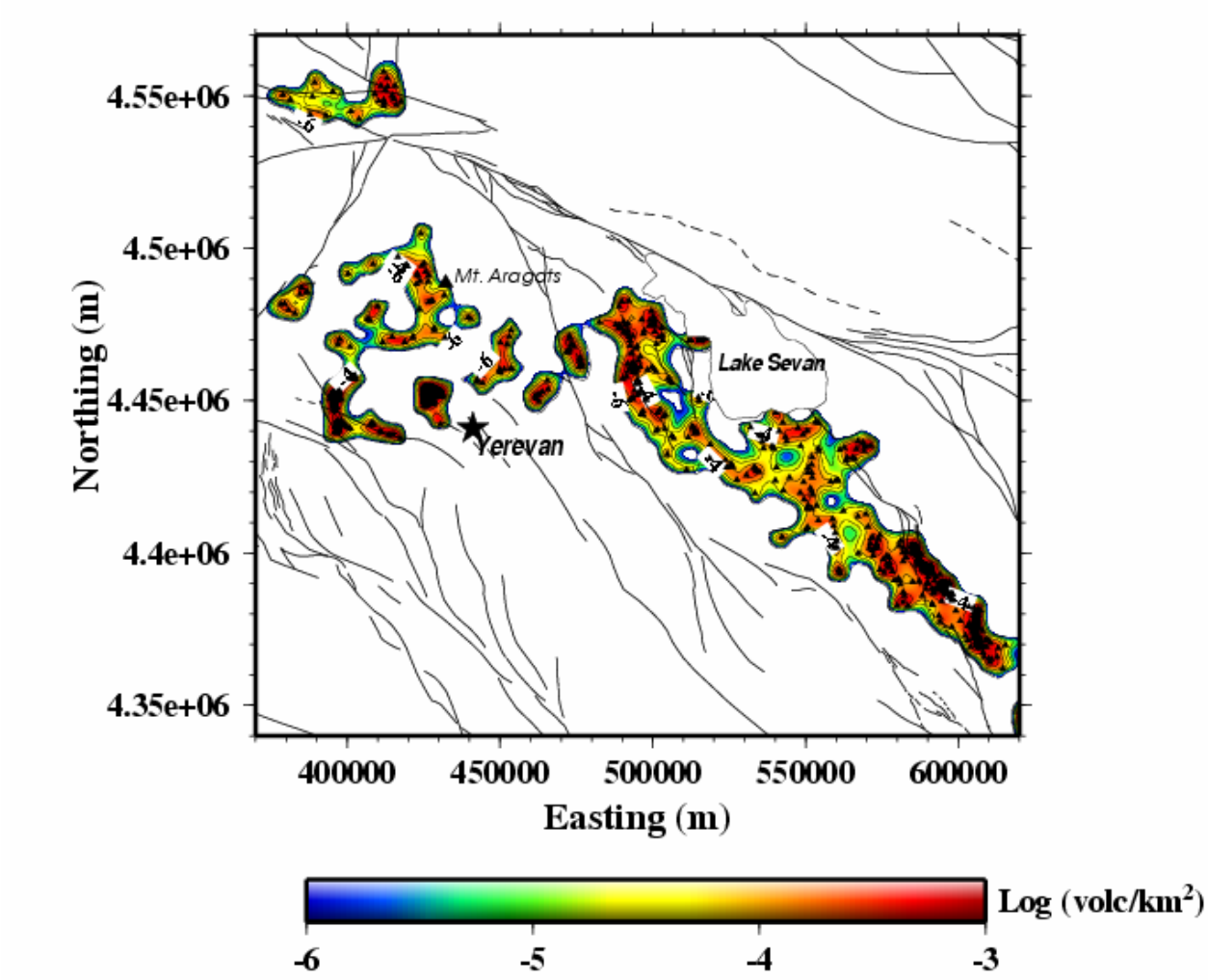
Appendix B (Continued)

```
$sum_k=0;

for ($ct=0; $ct<$n; $ct++){
        #steps through all data points

        $dist=sqrt(($volcanoes[0][$ct]-$x)**2 + ($volcanoes[1][$ct]-
$y)**2);
        #calculates the distance from point x,y to point $ct
        #print "$dist\n";

        $kernel=1/(2*3.14159)* exp(-0.5*$dist**2/$h**2);
        $sum_k=$sum_k + $kernel;
        #print "$kernel\n";

        #calculates the kernel and the sum respectively
}
$lambda=1/($n*$h**2)*$sum_k;
#lambda is calculated
```

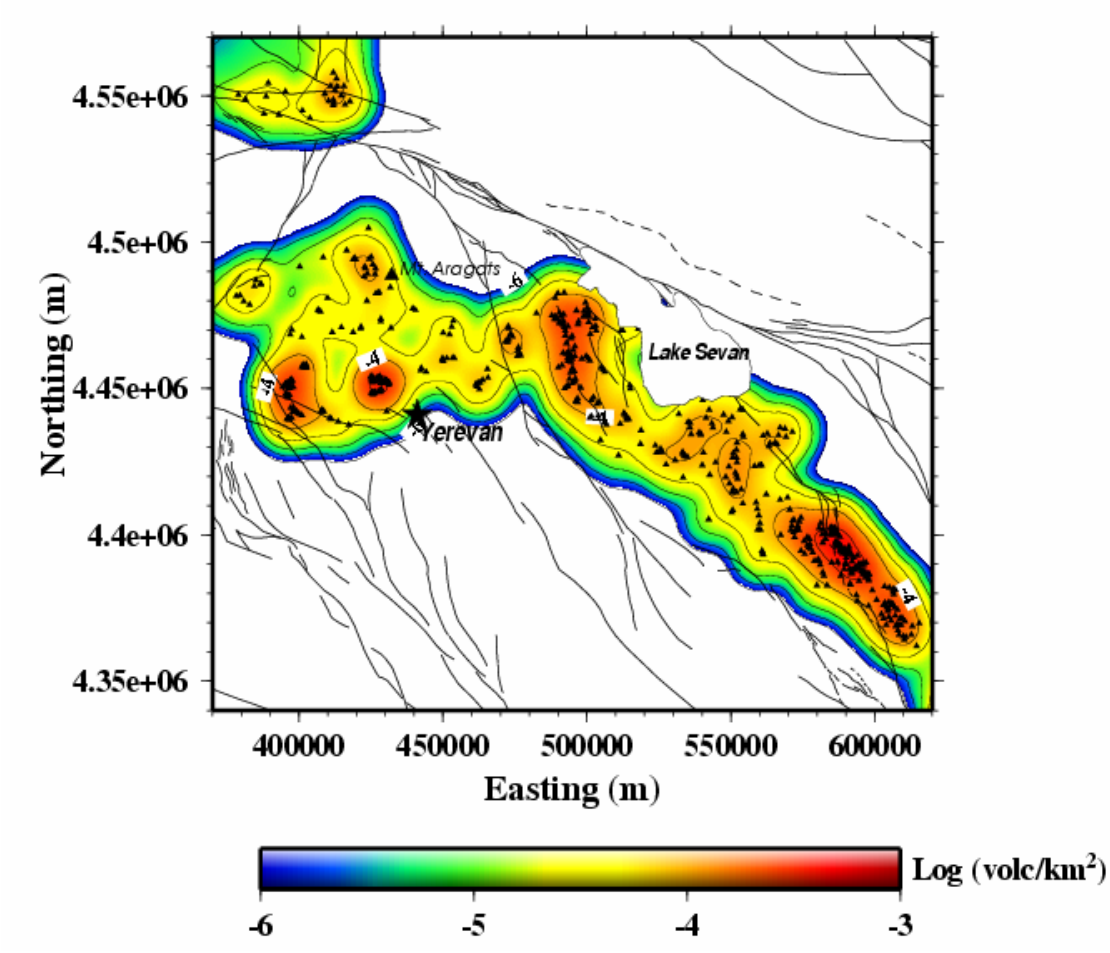
Appendix B (Continued)

```
                if ($lambda>=0.00000000000001){
                $lambda2=log($lambda*1e6)/log(10);
                #the log of lambda is calculated for ease of use in contouring
                #note that "log" in perl is ln so divid by log(10)
                #also multiple by 1e6 to report answer in terms of volcanoes/km^2 rather
than m^2
        }
        else{
        $lambda2=-13;
        }
                print "$x   $y     $lambda2\n";
        }
}
close (INPUT);
```

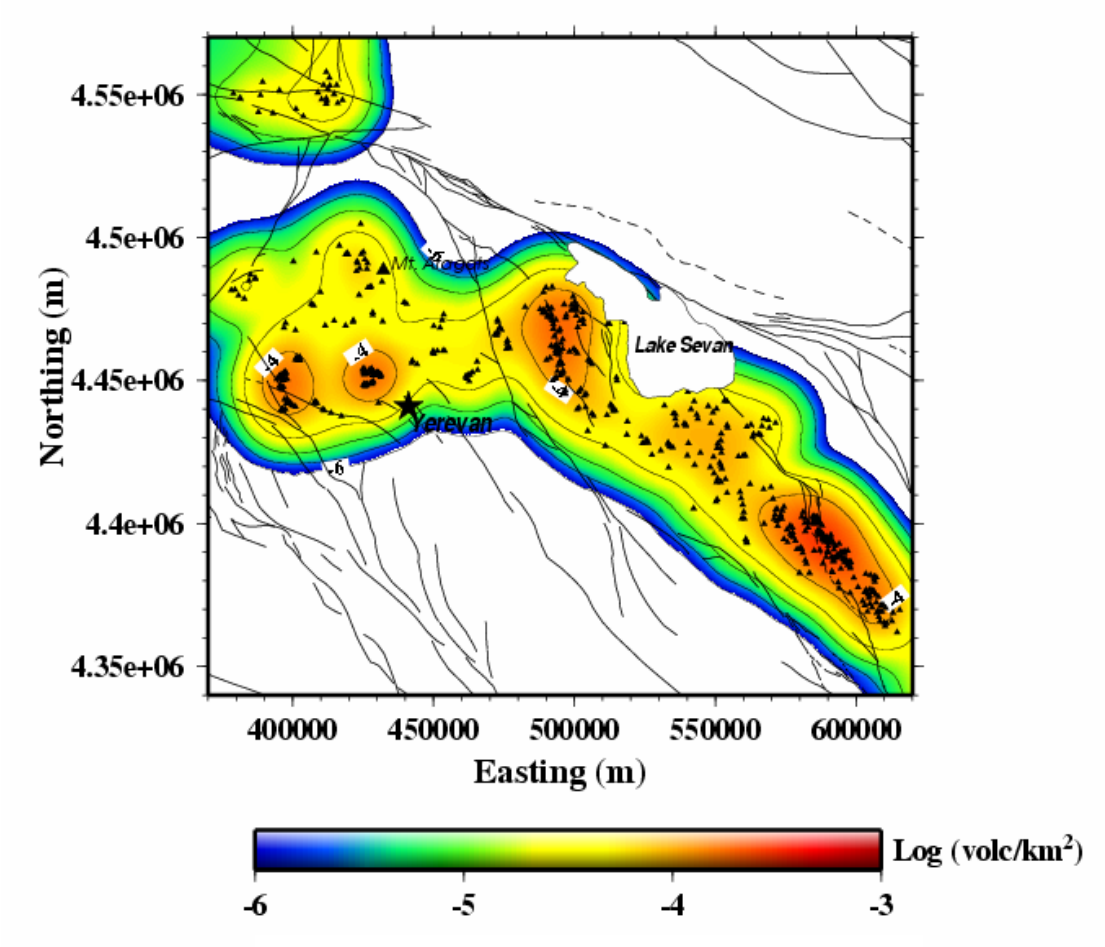
Appendix C: Additional Maps



Map of spatial density h=1300 m.



Map of spatial denstity h=6000



Map of spatial denstity h=8000