Modelling eruptive event sources in distributed volcanic fields

Elisabeth Gallant\textsuperscript{\textasteriskcentered \textasteriskcentered \textasteriskcentered \textasteriskcentered \textasteriskcentered a}, Lawrence Cole\textsuperscript{\textasteriskcentered \textasteriskcentered \textasteriskcentered \textasteriskcentered \textasteriskcentered \textasteriskcentered \textasteriskcentered \textasteriskcentered b}, Charles B. Connor\textsuperscript{\textasteriskcentered \textasteriskcentered \textasteriskcentered \textasteriskcentered \textasteriskcentered \textasteriskcentered \textasteriskcentered \textasteriskcentered \textasteriskcentered \textasteriskcentered γ}, Amy Donovan\textsuperscript{a}, Danielle Molisee\textsuperscript{γ}, Julie Morin\textsuperscript{a}, Rory Walshe\textsuperscript{a}, Paul Wetmore\textsuperscript{γ}

\textsuperscript{a}University of Cambridge, Department of Geography, Cambridge, UK.
\textsuperscript{b}Unaffiliated.
\textsuperscript{γ}University of South Florida, Department of Geosciences, Tampa, USA.

Abstract

High-quality model input data are a key component of a well-executed volcanic hazard assessment. Of equal importance is the quality of the conceptual models those data are fed into. Traditional vent opening hazard models have been built from datasets of all mapped vent structures within a given region. This approach often overlooks the fact that multi-vent eruptions, with sources linked at depth, are a common occurrence in areas of distributed volcanism. We present a new data-driven computational model to obtain more appropriate future source locations for distributed volcanic events. The code uses hierarchical clustering and linear cluster definition techniques to link vents in both time and space. Examples from Yucca Mountain, Nevada, USA; Craters of the Moon, Idaho, USA; and Pali-Aike, southern Patagonia, Argentina and Chile are provided to show how the code works for different datasets. The development of the algorithms used in the code is discussed, as are the strengths and limitations of our approach. The approach is best suited for datasets with full or partially complete age inventories, and we provide additional guidance on how to constrain the ages of undated units to improve performance. It is our hope that this work leads to further evaluation of conceptual models to improve volcanic hazard assessment methods.

Keywords: Monogenetic field; Hazard assessment; Numerical modelling; Spatial density; Volcanic vents;

1 Introduction

Distributed volcanism poses a serious threat to many locations across the globe (e.g. Auckland, New Zealand; La Palma, Spain; Managua, Nicaragua). Although distributed volcanic fields are common on Earth and other planetary bodies, their activity is rarely observed and volcanologists have limited experience in directly forecasting such eruptions. The most recent effusive eruptions not associated with a major edifice have been Geldingadalsgos (Iceland) in 2021, Paricutin (Michoacán-Guanajuato Volcanic Field, México) from 1943-1952 [Erlund et al. 2010], Carran-Los Venados (Chile) in 1907, 1955, and 1979 [Bertin et al. 2019, gand references therein], Shaitani (Chyulu Hills, Kenya) in 1855 [Scoon 2018], and potentially within the Harras of Dhamar (Yemen) in 1850 [Khalidi et al. 2010]. Hazards associated with distributed volcanism include ground deformation, phreatic explosions, gas emission, lava flows, pyroclastic density currents, shallow seismicity, tephra fallout, and ballistic impacts [Connor et al. 2015]. Due to the lack of modern observational data we must rely on the geologic record and computational resources to understand the potential threat of hazards from distributed volcanic fields.

Modern volcanic hazard assessments and forecasts, especially those focused on tephra dispersal, use estimation of eruption source parameters (ESPs) to forecast potential eruption impacts [Biass et al. 2014; Costa et al. 2016]. ESPs are model input variables or variable ranges that describe the magnitude and hazards associated with eruption forecasts. ESPs include vent location, proxies for eruption magnitude and intensity (e.g. volume of lava erupted or tephra grain size distribution), and many others [Spanu et al. 2016; Ogburn and Calder 2017; Freret-Lorgeril et al. 2021]. For many volcanoes the vent location is one of the most certain ESPs; this is not true for distributed systems (e.g. volcanic fields or rift zones on the flanks of large shield volcanoes), where new vents may open within an area hundreds to thousands of square kilometers in size [Valentine and Connor 2015].

Identifying the most likely source location of future activity is essential for modelling volcanic hazards, especially those strongly controlled by topography, such as pyroclastic density currents and lava flows. Small changes in vent location can have drastic impacts on inundation, especially in gentle terrain [Thompson et al. 2015; Hayes et al. 2018; Bilotta et al. 2019; Hayes et al. 2019; Becerril et al. 2021]. The complex vent structures associated with many eruptions can further complicate this effort because activity can be focused at various points along a fissure or across multiple fissures that span many kilometres to tens-of-kilometres over the
duration of an eruption [Richardson et al. 2017a; Neal et al. 2018]. Understanding how multi-fissure eruptions are preserved in the geologic record, how vents and fissures may have evolved over the course of an eruption, and ultimately how best to model these features is a key step for robust hazard assessment.

It is important to be critical of how we incorporate vent opening models into hazard assessment methods and which elements of the eruptive process we are most interested in, because not all vents produce the same type or intensity of hazard. Is it the mapped surface structures (vents and fissure) or the footprint of dyke that fed them which is most important to model for eruption sources? The answer to that question depends on which hazard is being modelled. For example, a community downslope from a vent might be most adversely impacted by lava flows (e.g. Meredith et al. Damage assessment for the 2018 Lower East Rift Zone lava flows of Kilauea volcano, Hawai‘i, in review), whereas critical energy infrastructure may be most negatively impacted by near-vent processes, such as extreme ground deformation [Wilson et al. 2014]. The source models needed to estimate near vent vs far vent hazards capture fundamentally different aspects of an eruption.

The majority of studies that model vent opening represent vents as independent point features, use spatial density estimation methods to forecast the location of future eruptions, and do not discern between methods for vent proximal vs distal hazards (e.g. Wetmore et al. 2009; Marti and Felpeto 2010; Bebbington and Cronin 2011; Connor et al. 2012; Cappello et al. 2015; Connor et al. 2019; Nieto-Torres and Martin Del Pozzo 2019). In this work we take the next step by identifying the interdependence of these points using hierarchical clustering algorithms to relate vent distribution to magmatism (clustering together vents mapped at the surface that are linked by dykes or magmatic events). Our approach relates these models to processes of magma migration in the shallow crust in order to capture the strongest controls on vent opening hazards, which may vary from field to field. We present a workflow to accomplish this with examples from Yucca Mountain (Nevada, USA), Craters of the Moon (COM; Idaho, USA), and Pali-Aike (southern Patagonia, Argentina and Chile) to illustrate how our computational tool can be used for vent data sets that are complete, partial, and limited, respectively. The sensitivity of how event sources are defined is examined and compared against vent opening hazard models using spatial density estimation. We assert that this approach highlights the need to think about the different model input requirements for near vent vs far reaching hazards when forecasting future activity. This issue, in particular, is present where vent fields or their issued hazards cross geo-political boundaries (e.g. a vent issues a lava flow into another country) and hazard management is not coordinated across borders [Donovan and Oppenheimer 2019]. Ultimately, the intent of these methods is to improve our ability to forecast where the next activity will occur and to improve source location model inputs for long-reaching hazards, such as lava flows and tephra dispersal.

## 2 Background

### 2.1 Sub-surface controls on vent distribution

Most vent opening hazard models use the locations of previous eruptive sources to best approximate magma generation and ascent at depth. From source to surface, magma ascent is complex and a host of factors influence where volcanic vents will form [Németh and Keresztori 2015; Smith et al. 2021]. The movement of magma is controlled by the combination of buoyancy, overpressure of magma, and host-rock fracture toughness [Lister 1990; Rubin 1993; Rivalta et al. 2015; Kavanagh et al. 2017]. Imaging of the shallow subsurface structures of volcanic fields suggests that the associated plumbing systems are a complex network of transport and storage structures, as opposed to the direct tapping of deep magma reservoirs for each individual vent [McLean et al. 2017; Kugaenko and Volynets 2019]. Dykes may stall at shallow depths to feed a network of sills and then continue to the surface and erupt from a variety of geometries [Richardson et al. 2017a] (Figure 1). This is due, in part, to the spatial and temporal variation of magma properties during dike propagation, such as bubble segregation in conduits [Corazzato and Tibaldi 2006; Pioli et al. 2009; McGee and Smith 2016; Tedagucci et al. 2021]. Interaction with crustal structures and the local stress field also control the orientation and distribution of emplacement patterns [Gómez-Vasconcelos et al. 2020; Kösk et al. 2020; Morfulis et al. 2020]. The highest points of flux along the fissure will eventually focus activity, creating spatter cones, armored ramparts, or other small-scale features [Parcheta et al. 2013; Rivalta et al. 2015] (Figure 2). The post eruption landscape can be dramatically different than the initial phases of the eruption, further complicating interpretation of initial emplacement conditions. More developed vent geometries (e.g. cinder cones and shields) can develop if the duration of the eruption extends several months or years [Swanson et al. 1979; Németh and Keresztori 2015]. Ultimately, the post-eruption surface landscape consists of many linked structures that reflect the location of a dike or series of dikes, which in turn reflect conditions in the subsurface [Le Corvec et al. 2013b] (Figure 1 and Figure 2).

### 2.2 Statistical analyses of vent distribution

Various statistical approaches have been used to identify clustering and alignment trends to better understand underlying controls on vent distribution [e.g.
Magmatic storage and transit structures of the San Rafael Volcanic Field, Utah, USA. This figure illustrates the complicated relationship between sills (purple shading), dykes (black lines), and the surface expression of volcanism (mapped vents/conduits, shown as black dots). We note the broader distribution of subsurface features (dykes and sill) compared to the conduits, which fed eruptions at the surface. All figure coordinates are reported in degrees lat/long (WGS '84).

Connor [1990] employed uniform density fusion and single-linkage cluster analysis to group vents in the Trans-Mexican Volcanic Belt, and then investigated alignment trends using two-point azimuth, Hough transform, and 2-D Fourier analysis. A modified two-point azimuth method using kernel density estimation (KDE) was applied by Lutz and Gutmann [1995] to reveal alignments on different spatial scales in the Pinacate Volcanic Field (Mexico). Smethurst et al. [2009] present the development of a stochastic model for the temporal and spatial distributions of flank eruptions at Etna (Italy). Von Veh and Németh [2009] used nearest-neighbour azimuth for determining preferred orientations in vent distributions and Hough transform to identify linear alignments of vents across the Auckland Volcanic Field (New Zealand). Le Corvec et al. [2013b] used Poisson nearest neighbour analysis guided by user inputs to identify alignments of volcanic vents. Cappello et al. [2013] employed exhaustive statistical analysis and a nonhomogenous Poisson process model applied to flank eruptions of the last 400 years at Etna. Germa et al. [2013] integrated the approach of Le Corvec et al. [2013b] with the KDE analysis most recently outlined in Connor et al. [2019] to analyze several distributed volcanic fields on the Baja California peninsula (Mexico). Tadini et al. [2014] based their methods on the non-homogeneous Poisson models described in [Connor and Hill 1995] for vent clustering in the Lunar Crater Volcanic Field (USA).

2.3 Vent opening hazard assessments

Deterministic models are those that, if given the same inputs, will provide the exact same output every time. A deterministic hazard assessment relies on pre-set parameters from a system and does not incorporate randomness into the input selection process. For vent opening hazards this approach involves pre-selecting vent locations or zones of hazard based on epistemic knowledge (e.g. scenario-based forecasting, Hayes et al. [2019]). Hackett et al. [2002] presented a multi-hazard assessment for the eastern Snake River Plain (ESRP), United States, where the areas of highest hazard are designated by a set distance from a “recent” vent (within 8 km of a vent <400,000 years old). In a similar vein, Lirer et al. [2001] suggested a new vent is most likely to occur where it has in the recent past for Campi Flegrei, Italy (the eastern area based on the last 5,000 years of activity). Deterministic approaches generally do not account for events that are not preserved in the geologic record or those that are possible but have yet to occur, such as high impact-low probability events.

Forecasting probable locations for future volcanism in distributed fields solely on pre-existing points requires special consideration for the relationships between these points [Connor et al. 2000].
Eruptive event sources in distributed volcanic fields

Gallant et al. 2021

Studies have assumed that either vents are likely to open anywhere within the volcanic field with uniform probability [e.g. Le Corvec et al. 2013a; Sandri et al. 2018; Hayes et al. 2019] or that the past distribution of sources can help quantify the likelihood of where future vents may open within a field [Connor et al. 2019, references therein]. Methods that assign probabilities based on vent density do so in a variety of ways; points may be weighted [e.g. Cappello et al. 2012; Becerril et al. 2013; Cappello et al. 2013; El Difrawy et al. 2013; Bartolini et al. 2014; Bevilacqua et al. 2015; Bertin et al. 2019] or unweighted [e.g. Bebbington and Cronin 2011; Connor et al. 2012] and the probabilities around these points may dissipate in a uniform or non-uniform fashion.

Vent maps may be influenced by various structural and observational data to inform source ESPs. Felpeto et al. [2007] and Alcorn et al. [2013] built maps from a weighted scheme of existing vents, faults, fumaroles, and springs. Bartolini et al. [2013] also incorporated data from multiple input sources and allowed for different bandwidth and kernel functions to be integrated into the susceptibility map. Bevilacqua et al. [2017] used a Bayesian model for coupling new vents with pre-existing faults. Bertin et al. [2019] used structural and satellite data as inputs for a vent opening hazard model where users could select the bandwidth estimator and weight dated inputs with time. Vent opening models integrated within a wider hazard assessment model may allow for a variety of user-defined input geometries for vent selection (e.g. vent selection for the lava flow model of Mossoux et al. [2016] can occur along a linear track or a wide area for Monte Carlo simulation).

2.4 Event modelling

Vent opening models have traditionally treated each mapped volcanic vent as an independent eruptive source [e.g. Lutz and Gutmann 1995; Bebbington and Cronin 2011; Connor et al. 2012; Bevilacqua et al. 2015; Cappello et al. 2015; Connor et al. 2019]. The partitioning of lava between multiple locations is a common occurrence for basaltic eruptions [e.g. Coltelli et al. 2007; Runge et al. 2014; Kubanek et al. 2015; Neal et al. 2018; Vasconez et al. 2018], and therefore modelling vent opening hazards requires an augmented approach that takes into account the spatio-temporal relationships between eruptive structures [Gallant et al. 2018]. In an eruption where lava is issued from multiple sources, we can think of vents as being dependent on one another (i.e. these vents all result from the ascent of a single batch of magma and are therefore linked in space and time – a volcanic event).

Ho and Smith [1998] discussed eruptive events and present a numerical model to define them that relies on orientation of alignments. Connor et al. [2000] developed a model for events using a uniform random distribution for alignment length and number of vents along the alignment, with uniform random distribution for alignment orientation based on tectonic setting and fault dilation analysis. Martin et al. [2004] provided two event definitions for their study of the Tohoku Arc (Japan). Runge et al. [2014] took additional steps by suggesting an event is an eruption continuous in both time and space. In their study, visual analysis of high resolution photographs were coupled with expert elicitation and analogue field data to group vents into events for the Harrat Rahat, Saudi Arabia using Bayesian analysis and a priori distributions. Nieto-Torres and Martin Del Pozzo [2019] used volcano morphology and cone cluster age analysis to define events in the Younger Chichinautzin Monogenetic Field (Mexico). Bertin et al. [2019] incorporated thermal anomalies, structural data, pre-existing vents, and temporal analysis into probability calculations to define eruptive events through KDE, similar to the methods developed by Cappello et al. [2013]. Gallant et al. [2018] defined events using a computational approach based on age and spatial relationships for volcanism on the eastern Snake River Plain (ESRP). The work presented in the following sections expands upon the methods first described in Gallant et al. [2018] by updating the spatial event classification tool and describing different methods for improving input variables.

2.5 Conceptual differences between vent opening and event source parameters

When considering the hazards associated with the opening of a new vent, it is important to think about the temporal evolution of eruptive fissure systems. Ground deformation and increased shallow seismicity occur prior to the dike tip intersecting the ground surface. Variable amounts of spatter and gas may impact vent adjacent areas as fissures and fountains begin to establish [Houghton et al. 2020]. This activity (or portions of it) are present at the opening of each new vent. Of fundamental importance is that many of these vents do not persist to feed extended lava flows (i.e. the threat that each vent poses to surrounding areas is not equal). Therefore, when we model hazards that impact areas beyond the near-vent region we need to be critical about how and why their source inputs differ from near-vent hazards.

The 2018 lower East Rift Zone eruption of Kilauea (USA) is a well-documented example that illustrates the conceptual difference between the vent opening hazards and broader event sources. Twenty four vents opened as part of the total eruptive activity from May through August of 2018: ground deformation, seismicity, spattering, and gas emission occurred at each of these locations [Neal et al. 2018; Houghton et al. 2020]. However, only five of those vents drove substantial offrift lava flow inundation. The overwhelming majority of the 0.9–1.4 km³ total eruptive volume was sourced from a single vent (the Ahu‘ailâ‘au fissure, formerly
identified as fissure 8) [Dietterich et al. 2021]. Collectively, the entire eruptive event was controlled by a very small number of eruptive vents. The same concepts are also on display at the 2021 eruption on the Reykjanes peninsula (Iceland).

Event models are likely to underestimate the near-vent hazards in areas with a high vent density. In situations where integration into additional hazard models is desired, vent models may over-estimate the likelihood of lava flow sources in those same areas. All events involve the same hazards as vent opening, but not all vent openings will involve broader event hazards. Another way to think about this is that vent opening models address direct hazards at the surface, whereas event models capture deeper processes associated with magma generation and ascent. For this reason, we assert that event models are more appropriate for integration with forecasts for far-reaching hazards (such as lava flows or tephra dispersal modeling) and present a method for obtaining the event sources.

3 Methods

We introduce a process that defines eruptive events using spatio-temporal relationships and provide a Matlab code to accomplish this (see Data Availability Statement). The purpose of this process is to provide a data-driven approach to defining eruptive events. We first describe the structure and function of the code and include algorithms for each step (Figure 3). We then describe methods for obtaining the required and suggested inputs: X-Y vent coordinates (required), the spatial template (required), and ages for each vent (suggested). We finish by briefly describing methods for visualizing these data as future event location hazard maps. Case studies from Yucca Mountain, Craters of the Moon, and Pali-Aike are used to illustrate this process in Section 4. These three examples were chosen because they show how datasets of varying size and completeness respond to the process. The development of this workflow and the evolution of the spatial clustering algorithm are discussed in Section 5.1.

3.1 Event modelling code

3.1.1 Step 1: Map age relationships within the field

Temporal clusters are ultimately controlled by the relationships between ages present in the data set (see Figure 4 for an example of the distribution of a synthetic dataset). We use the unique ages (i.e. no duplicate ages) for this first step and take into account age density for the next segment of this process. The unique ages are mapped into a dendrogram (tree diagram) with an agglomerative (bottom up) hierarchical clustering algorithm (see Figure 5 for a simplified example and Listing 1 for the pseudo-code). The temporal "distance"
between each point is measured, the two points with the smallest distance between them are linked (the next step up in the tree), and this process continues until all points are linked (the top of the tree, Figure 5). The output of this process provides the organisational foundation from which we are able to temporally cluster the data, which occurs in Step 2.

**Listing 1: Temporal Agglomerative Hierarchical Clustering.**

```
begin initialise: c, c ← n, T_i ← vent, i = 1, ..., n

begin initialise: k = 1, n ← 0

begin initialise: T = average age of the ith cluster

do c ← c − 1

find nearest datum, T_a and T_b

determine age of new cluster

as (T_a + T_b)/2

until c = 0

return c clusters

end

end

end
```

where; c = desired number of final clusters (1 in our case) and T = time (unique ages of data)

3.1.2 Step 2: Cluster by age

Temporal clustering is achieved through a divisive (top down) hierarchical clustering algorithm (Listing 2). This divisive algorithm uses the Mahalanobis distance between the ages of each vent as a grouping mechanism. The Mahalanobis distance is a measure of the distance between a point and a distribution (effectively, a univariate age z-score). For our temporal analysis, the age of a vent is the point, and the number of standard deviations a vent age is from the mean vent age of a cluster is the distribution. A user-defined Mahalanobis distance cut-off is implemented to control the range of ages represented by each cluster. This process starts by assuming every vent is initially part of the same event (the top of the tree shown in Figure 5). The mean age and standard deviation for all unique ages are first calculated. If ages reside outside of the cut-off, the process steps down the tree, recalculates the mean age and Mahalanobis distances for the branches, and applies the cut-off criteria again. This process repeats until all branches are within the established cut-off boundary (Figure 6). We recommend starting with a cut-off of 1 and adjusting as necessary (i.e. a cut-off of 1 standard deviation gives a randomly sampled point a 32% chance of being incorrectly excluded from its cluster, whereas a cut-off of 2 would only give a 5% probability but may include a larger range of ages than is desired for each cluster). We provide the inputs used for the analyses within the comments of the code to provide additional guidance.

**Listing 2: Temporal Divisive Clustering Algorithm.**

```
begin initialise: k = 1, n ← 0

begin initialise: T_i = ages within the ith cluster

begin initialise: \mu_i = average age of the ith cluster

begin initialise: \sigma_i = STD of the ith cluster

begin initialise: C = [c_1, c_2, ..., c_n]

begin initialise: \mu_i, \sigma_i for each unique cluster

find age in C

if any |T_i \mu_i|/\sigma_i > k

continue

else

return C, \mu_i

end
```

where; T_i = ages within the ith cluster, \mu_i = average age of the ith cluster, \sigma_i = Mahalanobis distance cut-off, and k = Mahalanobis distance cut-off.
3.1.3 **Step 3: Map spatial clusters**

Spatial clustering is established through linear cluster definition of each individual temporal cluster defined in Section 3.1.2. This method allows the definition of a cluster to potentially include a line segment, as opposed to a single point. We provide a descriptive algorithm in the supplement—see Data Availability Statement.

A set of vents that comprise a non-linear cluster are defined as a point if any of the following is true: 1) the maximum distance between the mean location and any vent within the temporal cluster is less than a user-defined spatial threshold, 2) the number of vents is less than N (this threshold can be changed in the code to reflect different trends for a given volcanic field, we used 5 for the analysis because it broadly qualified across the 3 fields analyzed), 3) the lower eigenvalue of the covariance is 0, and 4) the ratio of the larger and lower eigenvalues of the covariance is less than 1.5.

The definition of a linear cluster is most meaningful if there is significant spread in vent locations. If vents are tightly packed, the original point definition is more appropriate. Criterion 2 is needed to ensure there are enough vents in the cluster to assess linearity. Criteria 3 and 4 are applied to test the appropriateness of the linear definition. If none of the listed criteria are met, then the line segment definition is used. In this case the cluster is defined by a line segment centered at the mean location. The orientation of the line segment is along the dominant eigenvector of the covariance, as determined by principal component analysis. The length of the line segment is 3 times the average distance between the mean and all the vents in the cluster up to a user defined maximum.

Application of spatial clustering algorithms require new distance definitions to incorporate the linear segment model. There are 2 types of distance metrics required: 1) distance between a defined cluster and a vent location and 2) distance between two defined clusters. The distance between a vent and a cluster defined by a point is the Euclidean distance between the two. If the cluster is defined by a line segment, then the distance between the cluster and a vent is the average distance between the vent and the line segment.

The definition of distance between 2 clusters depends on whether the clusters are represented by a point or line segment. If both are represented by points, then the Euclidean distance is used. If one cluster is a point and the other a line segment, then the average distance between the point and line segment is used. If both are line segments, then the average distance between both line segments is used.

A hierarchical algorithm was used to cluster vent data according to the distance metrics described above. The hierarchical algorithm does not indicate the optimal choice of the number of clusters to represent the vents, and so the Davies-Bouldin index (DBI) was used to quantify the quality of the clustering solution [Davies and Bouldin 1979]. Ideally, clusters are tight and well separated. DBI quantifies two desirable qualities, minimizing distance between vents assigned to a cluster (tightness), and maximizing distance between clusters (separability). Lower DBI values (more optimal) indicate that these characteristics are more present in the clustering solution, although we note that this “best fit” may not always represent the best real-world solution. The final outputs for the full spatio-temporal clustering process are XY coordinates for events without any linear structure and line length, orientation, and centre point coordinates for events defined from aligned sources.

3.2 **Input data selection and processing**

3.2.1 **Vent selection**

The quality of vent data (XY coordinates) is the foundation of this process. Geologic maps are one of the most readily available sources of this data, though they are interpretations of the local geology and will have associated errors. For this reason we suggest using a combination of geologic maps, satellite imagery, digital elevation models, and field observations when possible to obtain vent data. We recommend the mapping scheme of Hauber et al. [2009] to identify effusive structures, such as fissures and scoria cones, when geologic maps are not available as an initial guide or the scale is too large for the required level of detail. Older maars and tuff cones may be more difficult to map due to sedimentation and questions about their plumbing structures (e.g. are they rootless?). Features like rootless cones and hornitos should be excluded because they lack subsurface pathways and therefore are not indicative of the location of magma generation at depth.

3.2.2 **Ages**

Unit ages are one of the most common, although incomplete, datasets available for volcanic fields [Wilson 2016]. Ages may be assigned relatively via stratigraphic sequence, absolutely through various radiometric dating techniques, or most commonly as a combination of both. One method to leverage relative age data and radiometric dates to enhance eruptive histories of sparsely dated volcanic fields is to incorporate stochastic computational models [Richardson et al. 2017b].

For a dataset with incomplete age data, we recommend the use of the Volcanic Event Age Model (VEAM) to better constrain the ages of vents without radiometric ages because it improves the quality of output by constraining ages in some undated vents, which reduces uncertainty in the age clustering. VEAM uses a-priori age data (radiometric ages and stratigraphic relationships) to constrain probable ages to undated units through Monte Carlo simulation (see Wilson [2016] and
Richardson et al. [2017b] for information on how to access and run VEAM). See the COM case study for more information on how VEAM is integrated into the event source modelling process.

When sparse age data are all that exist (e.g. on the order of geologic epoch), a median date for the epoch can be used for the vent age to facilitate reasonable clustering. See the Pali-Aike case study for more information on how we accomplished event source definition with limited available ages. If no ages are available, vents can still be clustered spatially, as described in Section 3.1.3.

### 3.3 Modelling vent opening hazards and event sources

The outputs of the event modelling process (X-Y coordinates of unaligned vent clusters and length, orientation, and centre point of linear features) are sufficient for deterministic or scenario-based hazard assessments, but need to be processed further for integration into probabilistic approaches. Kernel density estimation is the most common method employed for this purpose [e.g. Connor et al. 2012; Bartolini et al. 2013; Germa et al. 2013; Gallant et al. 2018; Bertin et al. 2019]. We use the methods described in Connor et al. [2019] to model the vent opening hazards and event source location probabilities for comparison. We note that the Connor et al. [2019] model requires point source inputs (X-Y coordinates), and so we use the linear centre points as the input for aligned features and the cluster coordinates to create the event source probability models. Cappello et al. [2013] and Bertin et al. [2019] present additional methods for incorporating linear features.

### 4 Case Studies: Results

#### 4.1 Yucca Mountain, Nevada (USA)

The geology of the Yucca Mountain area has been studied extensively because this location has been proposed as a geologic repository for high-level radioactive waste [Connor and Hill 1995; Potter et al. 2002; Parsons et al. 2006; Levich and Stuckless 2007]. Yucca Mountain lies within the Great Basin, an area characterised by extensive North–South trending mountain ranges and associated alluvial basins. Volcanism in the region has a long history, with a series of caldera-forming eruptions emplacing silicic tuffs and lava flows between 15 and 8 Ma. Quaternary volcanism is dominated by small-volume, distributed cinder cones and lava flows that span in age from ~10 Ma to 77 ka [Connor and Hill 1995; Potter et al. 2002]. These cones form within the extensional basins and along the structural alignments that transfer strain between the dominant normal faults of the region [Parsons et al. 2006] (Figure 7). Extensive sedimentation has buried many of these eruptive centers, particularly those at the southern end of the region [George et al. 2015]. Yucca Mountain was selected as a case study because it illustrates how the process works when every vent in a field has an assigned age, the optimal scenario for data availability.

![Figure 7: Map showing vent and event locations for the Yucca Mountain area, Nevada, USA (modified from Connor and Hill [1995]). The purple and yellow dots indicate the locations and relative ages of mapped vents, while the white dots show the location of calculated events. Map background data: Google, Landsat/Copernicus.](image)

#### 4.1.1 Input data

Vent coordinates and age inputs were obtained from Connor and Hill [1995]. Each of the 41 vents has an associated radiometric age, and therefore additional analysis using VEAM was not required to further clarify age relationships between vents.

#### 4.1.2 Event results and hazard maps

We have distilled the original vent catalog of 41 vents into 21 events (Figure 7). Figure 8 shows the vent opening hazard map and the event ESP input. An event list (including the output X-Y coordinates and information about which vents comprise the event) is included in the data repository.

#### 4.2 Craters of the Moon, Idaho (USA)

COM is one of the largest lava fields in the United States and represents a small fraction of volcanism on the ESRP, a distributed volcanic field associated...
Spatial Density

<table>
<thead>
<tr>
<th>km⁻²</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.7 × 10⁻⁴</td>
<td>5%</td>
</tr>
<tr>
<td>4.0 × 10⁻⁴</td>
<td>16%</td>
</tr>
<tr>
<td>2.9 × 10⁻⁴</td>
<td>33%</td>
</tr>
<tr>
<td>2.0 × 10⁻⁴</td>
<td>50%</td>
</tr>
<tr>
<td>1.2 × 10⁻⁴</td>
<td>66%</td>
</tr>
<tr>
<td>6.8 × 10⁻⁵</td>
<td>84%</td>
</tr>
<tr>
<td>2.7 × 10⁻⁵</td>
<td>95%</td>
</tr>
</tbody>
</table>

Figure 8: Vent opening hazard map and event eruptive source parameter input map for Yucca Mountain. Map background data: Google, Landsat/Copernicus.

with the passage of the Yellowstone hotspot across southern Idaho [Pierce and Morgan 1992; Kuntz et al. 2007]. The ESRP is comprised of ~4000 km³ of basaltic lava, issued from thousands of vents over the last 10 Ma [Kuntz et al. 1992]. Distribution of the ESRP’s >500 mapped surface vents trends in the northeast-southwest direction, parallel to the track of the hotspot [Wetmore et al. 2009].

COM has been studied and mapped extensively by Kuntz et al. [1982, 1986, 1992, 2007], Putirka et al. [2009], and many others. The 30 km³ of COM lava covers ~1 % of Idaho and was erupted in over 60+ flows from 25 tephra cones and eight eruptive fissure systems during the last 15 ka [Kuntz et al. 1992]. Common features in the region include cinder cones (up to 250 m tall), spatter cones, explosive pits, and lava fields covering up to 280 km² [Kuntz et al. 2007] (Figure 9). The majority of the vents that feed these fields occur along the northern 50 km of the northwest-southeast trending Great Rift, perpendicular to the direction of least principal compressive stress in the surrounding Basin-and-Range. Several lava flows are present with source vents that have been inundated by younger flows [Kuntz et al. 2007].

4.2.1 Input data

COM has been mapped extensively by Kuntz et al. [2007] and provides an example for how a robust dataset with incomplete ages can be processed. We obtain vent and fissure locations from the map of Kuntz et al. [2007] and cross-reference those data with GoogleEarth imagery. A COM vent catalogue is available in the GitHub repository linked in the Data Availability Statement.

We use VEAM to help constrain the ages of undated vents to improve clustering performance. Inputs for this separate code included stratigraphic relationships and radiometric ages (when known) from each cone and lava flow from Kuntz et al. [2007]. We generate 10,000 potential eruption chronologies that honour data constraints using 43 eruptive units, 19 radiometrically dated units, and 125 stratigraphic relationships. This analysis allowed us to provide an estimated date for an additional 14 vents, leaving only 10 vents with no age estimate.

In addition to improved age estimation from VEAM analysis, we were able to use the output to help guide the selection of our Mahalanobis distance cut-off value for the temporal clustering (Figure 10). The number of peaks identified from the VEAM analysis can be used...
as a proxy for the number of age clusters to aim for in volcanic fields with a high density of different aged vents over a relatively short period of time, or when the uncertainty on the radiometric ages overlaps. A Mahalanobis cut-off value of 0.7 was sufficient to obtain a number of age clusters similar to the number of peaks observed from the VEAM analysis.

4.2.2 Event results and hazard maps

We distill the original vent catalogue of 74 entries into 32 events (Figure 9). Figure 11 shows the vent opening hazard map and the event ESP input. An event list (including the output X-Y coordinates and information about which vents comprise the event) is included in the data repository.

4.3 Pali-Aike, Patagonia (Argentina and Chile)

Pali Aike, active from 3.8 Ma through the Holocene, is the youngest and southern-most of the Cenozoic back-arc Patagonian Plateau volcanic fields and is thought to be driven by subduction of a slab-window [D'Orazio et al. 2000; Corbella 2002]. The overall stratigraphic sequence is broken down into three phases; Unit 1) old plateau-like basal lavas, Unit 2) dissected old cones, ramparts, tuff rings, maars, and lava flows that sit atop the basal lavas and are covered by eolian soil, and Unit 3) young lava flows and scoria cones erupted ~5500 years ago. Units 1, 2, and 3 occupy 83, 15, and 2 % of the 4500 km² area of the volcanic field, respectively [D'Orazio et al. 2000]. No vents from the oldest activity (U1) are preserved at the surface, although silt, dike, and vent structures are exposed within the canyon of Rio Gallegos River. The middle unit (U2) has the highest number of preserved vents, accounting for over 90 % of the mapped structures. U2 features can be further sub-divided into Pleistocene or Pliocene activity [Corbella 2002; Zolitschka et al. 2006]. The young cones, fissures, and lava flows of unit U3 are present only in the southwestern portion of the field and account for the remaining <10 % of mapped vent structures [Corbella 2002]. Vents from U2 and U3 are predominantly aligned along NW–SE trending orientations, with minor alignments in the ENE–WSW direction [Mazzarini and D’Orazio 2003].

4.3.1 Input Data

We use the vent database presented in D’Orazio et al. [2000] and Le Corvec et al. [2013b] (n=467) as a foundation for locating vents. Vent identification in this field is more difficult than for COM due to the age of structures and the high degree of weathering and erosion that has occurred since emplacement. We reassessed this vent catalogue because higher resolution images are now available that were not at the time of original mapping. There are sparse radiometric age determinations and very few areas where relative stratigraphy is easy to discern. As a result, we focus heavily on the quality of vent locations. We apply a similar mapping scheme to Hauber et al. [2009] for effusive structures, such as fissures and scoria cones. The maars that have been studied by Mazzarini and D’Orazio [2003] and Ross et al. [2011] at Pali-Aike have mostly linear alignments, while rootless vents tend to cluster [e.g. Fagents et al. 2002; Boreham et al. 2018], and so we included all mappable vent structures in the catalogue. We adjust the pre-existing catalogue and map an addi-
52.00°S 51.50°S
-70.5°W -70.0°W -69.5°W -69.0°W

Figure 12: Map showing vent and event locations for Pali-Aike volcanic field, Argentina-Chile Border [modified from D’Orazio et al. 2000]. The purple dots indicate the U2 aged vents and the yellow U3 aged vents (as described in D’Orazio et al. [2000]), while the white dots show the location of calculated events. The blue lines show the alignment of linear clusters defined by the event modelling code. Map background data: Google, CNES/Airbus, Maxar Technologies.

4.3.2 Event results and hazard maps

The initial vent catalog of 519 points was distilled into 233 events (Figure 12). The 44 U3 vents were clustered into 4 events (yellow dots on Figure 12), while the remaining 477 U2 vents were clustered into 229 events. Figure 13 shows the vent opening hazard map and the event ESP input map. An event list (including the output X-Y coordinates and information about which vents comprise the event) is included in the data repository (Data Availability Statement). These maps were created using unweighted points in order to show the direct differences. We note that this dataset, in particular, is well suited for weighting due to the tight grouping of the youngest vents and direct users to for examples on how to determine weighting criteria.

5 Discussion

5.1 Algorithm development

This workflow has been developed over several years and has explored a number of different approaches. The temporal algorithm has remained consistent, but the spatial clustering approach has undergone a series of changes. Describing the evolution of the spatial clustering algorithm is important because it provides insight into the choices we have made with the code we present and allows future work to benefit from our development trial-and-error. The following versions of the spatial clustering were all tested independently from the temporal clustering to identify the failure points of each approach in the spatial domain. The spatial clustering of well-dated datasets is fairly straightforward and the outputs were similar across the range of approaches we tried, but we recognize that most volcanic fields are not comprehensively dated and we wanted to identify the most robust method for less than ideal conditions.

The first iteration of this code utilized an elliptical template for spatial clustering based on the footprint of an intruding dike and was described in Gallant et al. [2018]. The dimensions and orientation of this ellipse were determined by the user for their specific volcanic field. This approach was successful for volcanic fields with a single dominant alignment of vents, such as COM (ESRP), but problematic for areas where perpendicular alignments exist, such as Pali-Aike. The
prevalence of multiple alignment orientations in some volcanic fields [e.g. Le Corvec et al. 2013b] necessitates that the spatial clustering not rely on a single, fixed template. For this reason, we explored other approaches.

We then explored the use of Hough transforms, similar to Von Veh and Németh [2009] but encountered issues with fields where the density of data is sparse. The vent map is turned into a grid, which is then used to generate the a binary image. If the pixels are too large, then multiple vents are included in a single pixel and the linear feature is effectively smoothed out. If the pixels are too small, then a binary image is more speckled than structured and the Hough transform treats the vents like noise and does not extract the desired linear feature(s).

We considered exploring density-based spatial clustering but quickly abandoned that approach because it works under the assumption that the density of vents is equal across each event in a volcanic field, which not the case (i.e. in the COM data set we have events with only a single vent and others with up to 10). We settled upon the methods described earlier in the paper because they were able to capture variations across the field without being overly reliant on difficult to obtain user inputs.

5.2 Strengths and Limitations

Strengths: Our work is a step forward in improving model source locations for long term hazard assessments because it looks at eruptions as a whole unit instead of as individual mappable features and can be tailored to each volcanic field with a few easily obtained parameters (approximate length of dike in the field, X-Y vent coordinates, and temporal cut-off). The time component of our approach is particularly useful because it adapts to the distribution of the age data and does not apply a uniform block of time to temporally group vents; the temporal range (the Mahalanobis distance) will change across each age cluster (e.g. how the widths of clusters varies in Figure 10). The utility of this feature is more powerful in datasets where a higher number of vents are assigned an age (e.g. COM and Yucca Mountain) and where time boundaries between activity may be ambiguous (either due to the nature of the eruptions or the precision of the age dating). This work also is the first to provide segment information (both the length and orientation) of linear event features for volcanic events.

The more versatile an output, the more purposes it can serve. The components of the output (X-Y coordinates of non-linear clusters and the orientation, length, and center-point of linear clusters) allow for integration into various KDE models. As noted in Section 3.3, not all models are equipped to handle linear data components. KDE models can then be used as input location sources for models of future volcanic hazards, such as lava flows, pyroclastic currents, and tephra fallout. Additionally, information about each cluster (e.g. coordinates of each vent included and the ages of those vent) is provided as an additional output, and so combinations of analyses can be undertaken with this information if needed (weighting of events based on number of points, different KDE applications for each cluster, etc).

Limitations: Our approach operates under the constraints of "garbage in, garbage out" and therefore its effectiveness degrades as data quality decreases. The lack of age dating and dearth of apparent stratigraphic relationships from Pali-Aike made it difficult to sub-divide the data to further improve the spatial clustering accuracy. This was further complicated by the multiple alignment orientations. We suspect that Pali-Aike represents an end member of difficulty due to these complicating factors.

From an operational perspective, it may be confusing to have two different maps, one of which was created using "real" data (the vents) while the other is a conceptual representation of sub-surface magma generation (the events). Additionally, probabilistic data can be easily misinterpreted, regardless of the format in which they are conveyed [Doyle et al. 2014; Thompson et al. 2015], and the probabilistic event ESP map represents a multiply difficult to communicate concept that may be (mis)used for political purposes [e.g. Donovan 2017]. Best practice suggests that a table translating the probabilistic data into numerical and verbal phrases should accompany these efforts when used for operational purposes [Doyle et al. 2014]. During a crisis situation it is also recommended that conversations with civil protection and additional stakeholders take place to understand what information they value and need the most (i.e. which map would help them to make the most informed decision) [Thompson et al. 2017; Marrero et al. 2019; Barsotti 2020; Charlton et al. 2020]. We note that the event ESP maps are most useful for long-term hazard assessments and in a crisis situation, monitoring data (such as seismicity or deformation) would indicate the possible source location of an event with less uncertainty [e.g. Biggs et al. 2013; Orr et al. 2015; Lengliné et al. 2021]. For event ESP maps to be effective and understood tools for emergency managers, training and regular communication between those groups and scientists is needed to create a successful exchange between these groups (e.g. the effective use of probabilistic maps at Piton de la Fournaise, Réunion [Chevrel et al. 2020].)

5.3 Broader Applications

We argue that the most important part of this entire process is the act of evaluating existing model input parameters, challenging what they represent, and improving hazard assessment methods. The application of these methods is not limited to the examples provided; they can be modified to a number of different data types outside of those described in the previous
sections (e.g. geochemical data in lieu of or in addition to age). Any numerical data that can be treated as a Euclidian distance can be used with the code. Event clustering can also be used outside of a hazard assessment framework and can provide insight into the spatio-temporal evolution of volcanic fields over time, especially when combined with elements of VEAM not introduced in this paper (e.g. magma generation rates over time as a function of eruption rate for fields with tephra or lava flow volume estimates). The broad and unrefined ages make a magma generation rate over time calculation speculative at best, but partial or complete datasets (e.g. COM and Yucca Mountain) would be excellent candidates.

Our efforts also highlight the importance of obtaining additional high resolution and abundant radiometric age determinations in distributed volcanic fields. It is fair to say that much of the previous literature concentrated on spatial relationships among vents because there just was insufficient age dating to emphasize temporal relationships, but these temporal relationships are very important. Abundant age determinations in the Yucca Mountain volcanic field, for example, showed that volcanism shifted from East to West over millions of years. It is possible that temporal models, like VEAM, and spatio-temporal models (as we propose) will prompt additional age determinations in distributed volcanic fields. These methods may also help individuals identify the most important units for dating, which is particularly vital when financial factors limit the amount of ages that can be obtained within an operational budget. New modelling with more robust age determinations will improve our understanding of temporal evolution of these volcanic fields and have a direct impact on hazard assessments. It is important, for instance, to distinguish between volcano clusters which are truly monogenetic, with all vents forming in a geologically brief period of time, from clusters that are repeatedly active over long periods of time with activity separated by long hiatus [e.g. Németh and Kereszturi 2015].

This work brings into focus the need to improve the ability to assess the independence of volcanic vents. That is, if 10 vents are mapped in a cluster, do these vents represent one magmatic event, three, or ten? This assessment has a direct bearing on recurrence interval calculations and the types of statistical models that might be applied to forecast hazards. While the proposed method does not give a definitive way to determine independence, which may always require additional assumptions, it does provide a way to assess spatio-temporal relationships that can be used to justify alternative models. Ultimately, the goal is to better constrain and quantify uncertainty in models and hazard assessments to help reduce the impact of future events that occur or to preemptively mitigate the hazard-societal interaction before it occurs (i.e. avoidance through appropriate site-selection for critical infrastructure by identifying the most likely places to host volcanic activity in the future [Connor et al. 2012; Gallant et al. 2018]). The event source models are meant to describe subsurface magma migration and may also be a useful tool during non-crisis times for planning instrumentation deployments. By shifting the focus of source input models from the mapped vent data to one that captures dynamic processes at depth, we are able to expand the applications of these models.

6 Conclusions and recommendations

We challenge the idea that mapped vents represent the most appropriate location sources for modelling distal volcanic hazards and have developed a computational approach to improve the the process. Application of the model is illustrated by three examples from volcanic fields with varying levels of data quality (Craters of the Moon, Yucca Mountain, and Pali-Aike). Most importantly, we encourage folks to explore the capabilities of the code and have fun with it; it is our hope that interest in this approach quickly renders our code obsolete and we invite collaboration on its continued improvement.

Acknowledgements

We thank Daniel Bertin, Álvaro Amigo, Jan Lindsay, and Nicolas Le Corvec for their assistance locating the original Pali-Aike vent data. Brit Hill is thanked for providing the Yucca Mountain vent catalogue. We also thank Frances Boreham, Jacob Richardson, Patrick Whelley, and Sarah Sutton for their insight into maar classification and vent mapping best practices. We thank editor Mike James and two anonymous reviewers for their thorough and considerate suggestions, which greatly improved the quality of this manuscript. Reviewing during a pandemic is no small task and we are especially appreciative of their time and effort during such chaos. We also thank David Hyman for his comments on the text. This work was supported by ERC IMAGINE grant 804162 (PI Amy Donovan).

Author contributions

Each author contributed in meaningful ways to the writing and editing of this manuscript. EG, CC, and PW developed the early drafts of this work as part of the PhD of EG. EG, LC, and CC all contributed to the development and implementation of the event modelling algorithm. DM and EG ran the VEAM outputs. AD, JM, RW, and EG helped frame and build the broader context of this work.

Data availability

https://github.com/elisabeth-gallant/volcanic_event_source: This repository contains the Matlab code and required functions to run the analyses described in this manuscript, a .xls file that contains the vent lists used and the event outputs, and
digital elevation models for plotting. The model algorithms are also provided. The version associated with this manuscript is v2021.12.28, available at: https://github.com/elisabeth-gallant/volcanic_event_source/releases/tag/v2021.12.28.

**Copyright notice**

© The Author(s) 2021. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

**References**


Eruptive event sources in distributed volcanic fields

Gallant et al. 2021


Morfulis, M., W. Báez, S. Retamoso, L. Bardelli, R. Filipovich, and C. A. Sommer (2020). “Quantitative spatial distribution analysis of mafic monogenic volcanism in the southern Puna, Argentina: Implications for magma production rates and structural con-
Eruptive event sources in distributed volcanic fields


