

JGR Solid Earth



RESEARCH ARTICLE

10.1029/2021JB022323

Special Section:

Machine learning for Solid Earth observation, modeling and understanding

Machine Learning Approaches to Identifying Changes in Eruptive State Using Multi-Parameter Datasets From the 2006 Eruption of Augustine Volcano, Alaska

Grace F. Manley¹ , Tamsin A. Mather¹ , David M. Pyle¹ , David A. Clifton², Mel Rodgers³, Glenn Thompson³, and Diana C. Roman⁴ 

¹Department of Earth Sciences, University of Oxford, Oxford, UK, ²Department of Engineering Science, University of Oxford, Oxford, UK, ³School of Geosciences, University of South Florida, Tampa, FL, USA, ⁴Earth and Planets Laboratory, Carnegie Institution for Science, Washington, DC, USA

Key Points:

- Application of one-class and multi-class machine learning classification to seismic, gas and geodetic time series from the 2006 eruption of Augustine Volcano, Alaska
- We identify a transition from non-eruptive to eruptive activity in November 2005, coinciding with previous estimates of the initiation of dike intrusion
- Models formulated using all datasets (seismic, gas, and geodetic) distinguish between non-eruptive and eruptive data better than models trained on a single data type alone

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

G. F. Manley,
grace.manley@earth.ox.ac.uk

Citation:

Manley, G. F., Mather, T. A., Pyle, D. M., Clifton, D. A., Rodgers, M., Thompson, G., & Roman, D. C. (2021). Machine learning approaches to identifying changes in eruptive state using multi-parameter datasets from the 2006 eruption of Augustine Volcano, Alaska. *Journal of Geophysical Research: Solid Earth*, 126, e2021JB022323. <https://doi.org/10.1029/2021JB022323>

Received 28 APR 2021
Accepted 2 NOV 2021

Author Contributions:

Conceptualization: Grace F. Manley, Tamsin A. Mather, David M. Pyle, David A. Clifton, Mel Rodgers

© 2021. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Abstract Understanding the timing of critical changes in volcanic systems, such as the beginning and end of eruptive behavior, is a key goal of volcanic monitoring. Traditional approaches to forecasting these changes have used models motivated by the underlying physics of eruption onset, which assume that geophysical precursors will consistently display similar patterns prior to transition in volcanic state. We present a machine learning classification approach for detecting significant changes in patterns of volcanic activity, potentially signaling transitions during the onset or end of volcanic activity, which does not require a model of the physical processes underlying critical changes. We apply novelty detection, where models are trained only on data prior to eruption, to the precursory unrest at Augustine Volcano, Alaska in 2005. This approach looks promising for geophysically monitored volcanic systems which have been in repose for some time, as no eruptive data is required for model training. We compare novelty detection results with multi-class classification, where models are trained on examples of both non-eruptive and eruptive data. We contextualize the results of these classification models using constraints from petrological, satellite and visual observations from the 2006 eruption of Augustine Volcano. The transition from non-eruptive to eruptive behavior we identify in mid-November 2005 is in agreement with previous estimates of the initiation of dike intrusion prior to the 2006 eruption. We find that models which include multiple types of data (seismic, deformation, and gas emissions) can better distinguish between non-eruptive and eruptive data than models formulated on single data types.

Plain Language Summary Identifying the onset of volcanic activity is a central goal of volcanology. We apply machine learning approaches to seismic, gas and geodetic data in order to classify eruptive and non-eruptive activity. We consider models trained on only pre-eruptive data, and models trained on both non-eruptive and eruptive data. Our models begin to identify eruptive behavior in mid-November 2005, which coincides with previous estimates of the beginning of transport of magma to the surface. We compare the models constructed using all data types to those made with each data type individually. We find that models which include multiple types of data are better at identifying non-eruptive and eruptive activity.

1. Introduction

Volcanic eruptions are characterized by large-scale transitions in behavior: from repose to unrest; from unrest to eruption; and from eruption to quiescence. Identifying the transitions occurring at the beginning or end of volcanic activity is a fundamental goal of volcanology (NASEM, 2017; Sparks, 2003). The beginnings of volcanic eruptions are often marked by changes in geophysically observable indicators of unrest, including seismic, gas and ground deformation measurements; however, it is often difficult to interpret these signals and integrate multi-parameter observations to determine whether this unrest will lead to an eruption (Moran et al., 2011; Pritchard et al., 2019; Segall, 2013). Interpreting the physical meaning of changes in geophysical observations can be especially difficult in real-time. Figure 1 and Table 1 outline known physical processes that can generate various signals of unrest in seismic, gas and deformation data, and which operate across a range of depths within the volcanic system.

Geophysical parameters may not always show consistent changes leading up to eruption. For example, increasing volcano-tectonic (VT) seismicity is typically considered an important precursor to eruption (Cameron et al., 2018) at basaltic systems including Kilauea, Hawai'i and Piton de la Fournaise, Réunion Island. However,

Data curation: Grace F. Manley, Glenn Thompson

Formal analysis: Grace F. Manley

Investigation: Grace F. Manley

Methodology: Grace F. Manley

Software: Grace F. Manley

Supervision: Tamsin A. Mather, David

M. Pyle, David A. Clifton, Mel Rodgers

Visualization: Grace F. Manley

Writing – original draft: Grace F. Manley

Writing – review & editing: Grace F.

Manley, Tamsin A. Mather, David M.

Pyle, David A. Clifton, Mel Rodgers,

Glenn Thompson, Diana C. Roman

these precursory increases are not always present, and even if present do not consistently culminate in an eruptive episode (Chastin & Main, 2003; Collombet et al., 2003). An analysis by Pesicek et al., (2018) of seismically monitored eruptions spanning a variety of eruptive styles across Alaska found that only 30% were associated with a statistically significant increase in precursory event rate. At persistently restless volcanoes, such as Telica Volcano, Nicaragua, volcanic unrest is often prolonged and seismic event rates decrease prior to eruption due to weak sealing within the hydrothermal system (Geirsson et al., 2014; Rodgers et al., 2015; Roman et al., 2016, 2019). Compilations of global satellite observations of deformation illustrate that, although the majority of studied volcanoes which erupted were associated with deformation, there remain several eruptions for which there was no observable associated deformation prior to eruption, including Shishaldin and Cleveland volcanoes (Biggs et al., 2014; Ebmeier et al., 2013; Lu & Dzurisin, 2014). In the case of Cleveland volcano, this lack of observable deformation in both long-term and short-term Interferometric Synthetic Aperture Radar (InSAR) images is attributed to the “open-conduit” nature of the shallow magma storage system (Lu & Dzurisin, 2014). Gas emission rates including SO₂ emissions generally increase prior to eruption (e.g., Sparks, 2003). However, in some instances, such as the March 1993 eruption of Galeras Volcano, Colombia, a decreasing SO₂ emission rate was observed prior to eruption which was attributed to retention of gases due to sealing processes within the volcanic carapace (Fischer et al., 1994).

Previous approaches to identifying critical changes preceding eruptions have focused on incorporating physical understanding of potential drivers of eruption. For example, the Failure Forecast Method (FFM) is based on the assumption that an accelerating rate of a given observable geophysical parameter, such as the rate of seismicity, corresponds to the accumulation of subsurface damage (e.g., Kilburn et al., 2003; Voight, 1988; Voight & Cornelius, 1991). Eruptions may then be forecast by fitting a power-law relationship to the rate of occurrence of the observable parameter. However, physically motivated models rely on the assumption that the mechanics underlying the model are consistent, which may not always be the case. Application of a probabilistic formulation of the FFM to multiple run-ups to explosions in the same eruptive episode at Volcan de Colima, Mexico yielded variable results for run-ups in which the pattern of unrest was more complicated than a single acceleration of seismic rate (Boué et al., 2015). Other physically motivated models may be highly successful but are rather specific to one eruption or class of eruptions and limited in their applicability to a wide range of volcanic systems. For example, Anderson and Segall (2013) presented a physics-based extrusion model combined with a Monte-Carlo inversion approach, and demonstrated retrospectively that the final extruded volume of the Mt St Helens 2004–2008 dome-forming eruption could be successfully forecasted from up to two years prior to the end of the eruption (see also Segall, 2013).

Identifying eruption onset and end using machine learning classification approaches is a growing field. Dempsey et al. (2020) presented a machine learning approach to issuing short-term alerts from real-time seismic data at Whakaari, New Zealand, using supervised random forest classification on features derived from seismic tremor data and retrospectively applied the model to five episodes of eruption. Ren et al. (2020) applied both supervised and unsupervised classification to seismic data at Piton de la Fournaise, La Réunion and identified consistent tremor frequencies between eruptive episodes. Manley et al. (2020) identified transitions based on supervised classification of data from Telica and Nevado del Ruiz volcanoes, and showed that machine learning classification of features derived from seismic data alone held the potential to identify the end of eruptive sequences. Additionally, the timing of transition between eruptive and non-eruptive seismic activity at the end of eruption was found to be 2–4 months after the last visual activity observed at the volcano.

While seismic data has historically been the most common observable used to monitor volcanoes, complementary information is contained within other parameters, including gas and deformation data. Following on from our previous study (Manley et al., 2020) which was limited to seismic data, here we use multiple types of instrumental data, including seismicity, gas emission rate and GPS-derived deformation to classify behavior prior to and during a major eruption at Augustine volcano, Alaska, using a machine learning approach. We hypothesize that the range of physical phenomena captured from an ensemble of observables is much wider than would have been identified using a single data stream, and that the physical processes driving these phenomena will have operated across a range of depths within the volcanic system.

Augustine Volcano, Alaska is a stratovolcano located in the Cook Inlet region (Figure 2), ~280 km SW of Anchorage, and is one of the most historically active volcanoes in Alaska (Power et al., 2010), with major eruptions in 1976 and 1986. The most recent eruption of Augustine began in January 2006, and was documented in detail

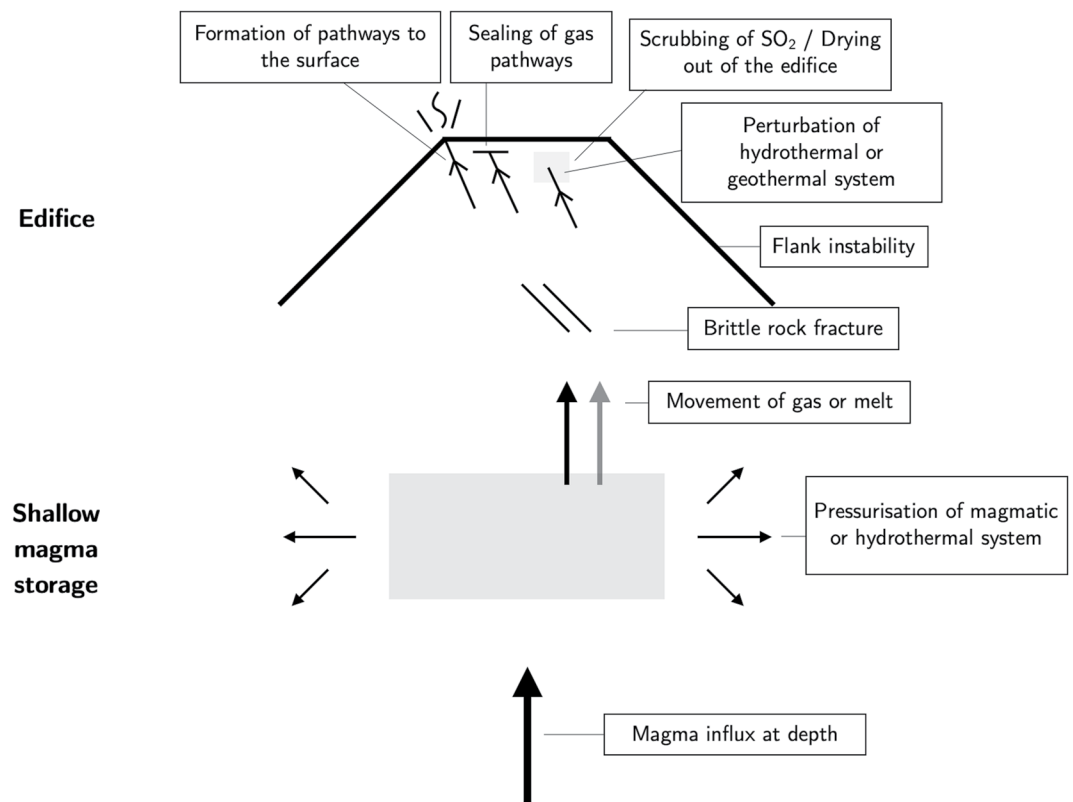


Figure 1. Illustration of the range and relative depths of physical processes associated with observation of seismicity, gas and geodetic measurements. Arrows are a conceptual representation of movement of material. Gray boxes are a generalised representation of magmatic or hydrothermal activity. Annotations denote the corresponding physical processes outlined in Table 1.

Table 1

Summary of Observable Changes and Underlying Physical Processes Which Are Linked to Observations of Seismicity, Gas and Geodetic measurements

Type	Observed	Associated physical processes
Seismic	<ul style="list-style-type: none"> - Swarms of low frequency (LF) earthquakes - Volcanic tremor - Accelerating seismic rates for example, Cameron et al. (2018) or increase in real-time Seismic Amplitude Measurement (RSAM) - Declines in seismicity (if close to eruption) due to sealing Endo et al. (1996) 	<ul style="list-style-type: none"> - Pressurization of magmatic or hydrothermal system - Brittle rock fracture for example, Chouet and Matoza (2013) - Movement of gas or melt - Formation of pathways to surface
Gas	<ul style="list-style-type: none"> - Increased SO₂ flux approaching eruption - Decreased SO₂ flux due to sealing 	<ul style="list-style-type: none"> - Magma influx at depth modifying gas composition ratios for example, Aiuppa et al. (2007) - Formation of pathways to surface for example, Sparks (2003); Werner et al. (2011) - SO₂ scrubbing, or drying out of edifice Symonds et al. (2001)
Geodetic	<ul style="list-style-type: none"> - Deformation which exceeds the baseline recorded deformation - Changes in sign of deformation (e.g., deflation or inflation) if very close to eruption 	<ul style="list-style-type: none"> - Intrusion of new magma at depth - Magma migration - Perturbation of hydrothermal or geothermal system - Flank instability - Brittle rock fracture

Note. All of these physical processes are conceptually illustrated in Figure 1.

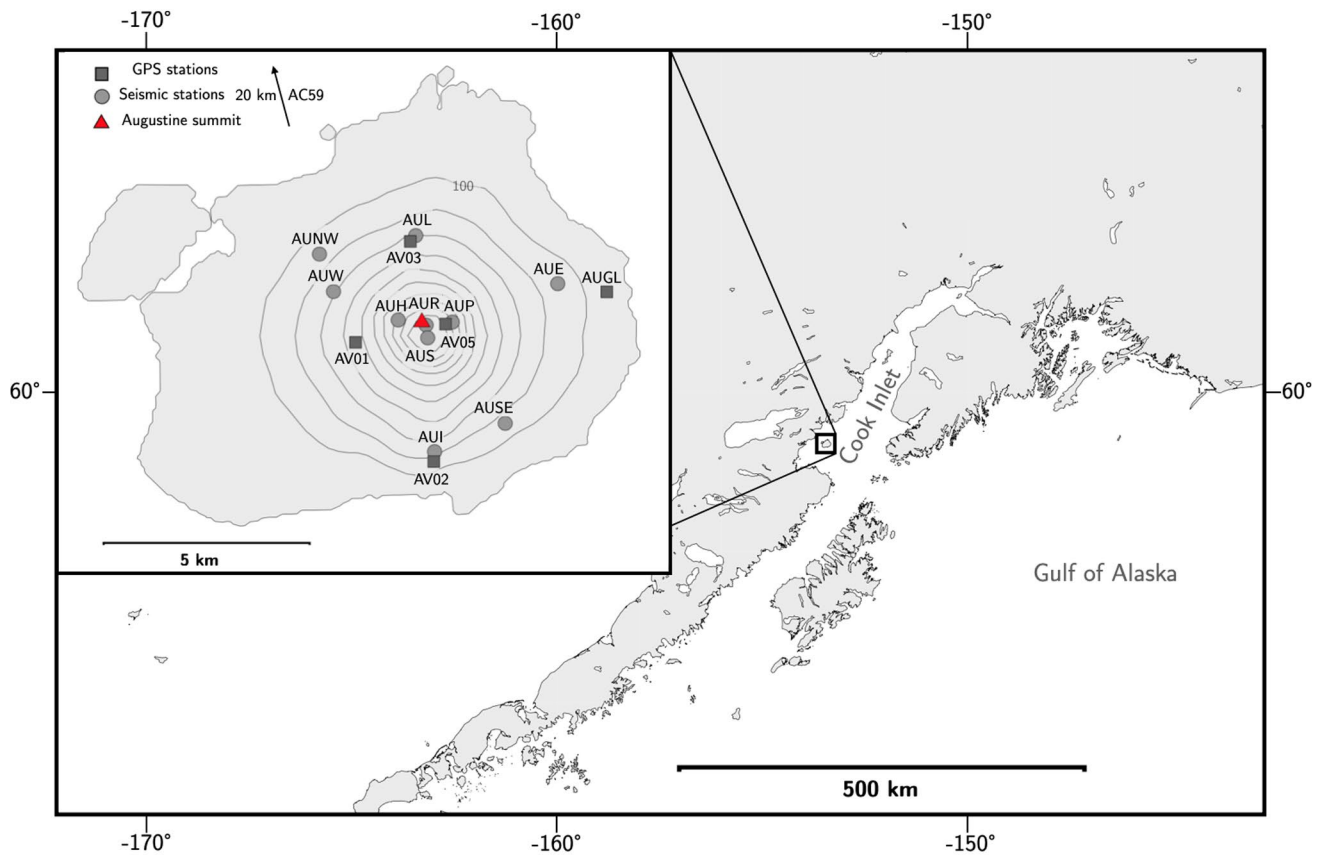


Figure 2. Seismic and GPS station map for Augustine Island, inset in a map of the Cook Inlet, Alaska (location indicated by latitude and longitude on axes) with contours at 100 m intervals. Squares denote the location of GPS stations, circles denote the location of seismic stations and the triangle represents the location of Augustine Volcano summit. Figure constructed using the Generic Mapping Tools version 6 (Wessel et al., 2019).

(as outlined in Section 2.1) (e.g., Power et al., 2010 and references therein). Coombs et al. (2010) described the chronology of deposits from the eruption, inferred from observations made during overflights, and satellite remote sensing. Timing of the large ash explosions which defined the explosive phase of the eruption were constrained using infrasound (Petersen et al., 2006), and these observations which extended over the course of the eruption provide an independent constraint on the timing of transitions in activity at Augustine, and of the physical interpretations of our modeling results detailed below. The 2006 eruption of Augustine represents an ideal case study as the eruption was large, and the availability of comprehensive geophysical and observational datasets constraining the activity provides context for the results of classification modeling.

In order to discuss the transitions between the state of volcanic activity at Augustine, we first define the states of behaviour which we adopt here, namely “repose,” “unrest,” and “eruption”. We use a definition of eruption after Siebert et al. (2015): eruption is characterized by the explosive ejection of fragmented new magma or older previously solidified magma, and/or the effusion of liquid lava. Unrest is defined as any non-eruptive activity above background which occurs outside the eruption, and the term repose is used for a volcano which is neither erupting nor displaying above-background signs of unrest. For the Augustine system specifically, the 2006 activity can also be divided into phases distinguished by their visual products erupted during these periods (Power et al., 2006), and according to the Alaska Volcano Observatory system of rating volcanic activity known as the Level of Concern Color Code system. These classification schemes are described in detail in Section 2.1.

In this paper we compare results from classifying volcanic state using a novelty detection or one-class approach, which does not incorporate information from future days of the time series, to a multi-class approach in which models are trained on examples of both non-eruptive and eruptive data. These models are formulated on features derived from seismic, gas and geodetic data. We present a novel approach for handling missing and irregularly sampled values in SO_2 emission rate time series. We compare the performance of novelty detection and

multi-class classification approaches using SO₂ emission rates estimated using either no future data points or with a probabilistic, retrospective approach formed using the full set of data points. We also compare models formulated with all datasets to models formulated using only seismic, gas or geodetic features.

2. Methods

2.1. Augustine Volcano

2.1.1. Eruption of 2006

The most recent explosive eruption of Augustine began in January 2006 and is recorded in the Smithsonian Global Volcanism Program (GVP) database as a VEI 3 event. The GVP start-date for activity is December 9, 2005 and the end-date is April 27, 2006, based on reports of changes in seismicity, though the exact days are marked with uncertainty in the GVP record (Global Volcanism Programme, 2006). Figure 3 illustrates the timeline of the visible eruption during the 2006 period of activity. The 2006 eruption of Augustine began with several small, phreatic explosions on December 2, 10, 12, and 15 (Power & Lalla, 2010), which were not thought to contain juvenile material. The main phase of the eruption began on the January 11, 2006, with two large explosions generating substantial ash-bearing plumes. January 11 was also the day of the eruption with the highest seismic event rate. Explosive activity continued until January 28, after which the eruption style mainly comprised low-level ash emission, and small pyroclastic flows (Vallance et al., 2010).

There are different ways to classify the stages of the event. At the time of its explosive eruption in 2006, Augustine volcano had a system of Concern Color codes which had 4 states: green, yellow, orange and red (Brantley, 1990). The beginning of the 2006 eruption was successfully forecast by the Alaska Volcano Observatory, with the Concern Color code being raised from Green to Yellow 3 days preceding the first phreatic explosions on December 2, 2005 (Figure 3, Cameron et al., 2018). In the Concern Color Code system in place at the time, the Yellow level implies that an eruption is possible within the next few weeks. Activity associated with the 2006 eruption itself was split into 4 phases by Power et al. (2010), and activity before the eruption is known as the precursory phase, whereas activity following the designated eruption end-date is known as the post-eruptive phase. The eruption itself comprises an explosive phase, continuous phase, hiatus and effusive phase. The classification schemes associated with the Augustine eruption of 2006 are summarized in Figure 3.

2.1.2. Physical Constraints on the Magmatic System at Augustine

There are two primary lithologies associated with the 2006 eruption (Larsen et al., 2010): High-Silica Andesite pumice (HSA) and Low-Silica Andesite Scoria (LSAS). The volcano summit was located at 1.2 km above sea level (Power & Lalla, 2010). The primary magma storage was located at 3–5 km depth below sea level, as constrained by post- and syn-eruptive earthquake locations (Power & Lalla, 2010). Melt inclusion analyses of HSA (Webster et al., 2010) and lack of significant SO₂ emissions before December 2005 indicated storage at >4 km depth below the edifice (Coombs et al., 2010). There was a primarily aseismic conduit linking the magma storage to the fracture zone; with few earthquakes located at these depths during the 1970s (Power & Lalla, 2010). The majority of earthquakes during the precursory swarm were located within a network of fractures from 100s of m above sea level to approximately 1 km below sea level (DeShon et al., 2010; Power & Lalla, 2010; Figure 3).

Cervelli et al., (2010) suggest that there was no active magma column below the edifice prior to the 2006 eruption, inferred due to low fumarolic temperatures at Augustine. However, Jacobs et al. (2010) report increases in the monthly temperature average from January 2005, measured at the location of the AUS seismic station, which are attributed to a volcanic origin. The conduit which fed the 2006 eruptive episode has been suggested to have formed in the preceding weeks–months to the eruption (Cervelli et al., 2006, 2010).

2.2. Data Processing

Before applying the machine learning methods (Section 2.3), we first extract a number of representative features from the seismic, gas and geodetic measurements. We derive a set of 40 features comprising 34 seismic features, 2 gas features and 4 geodetic features, which are outlined in Tables S3–S5 in Supporting Information S1. This full set of features is used to train and test the models described in Section 3.

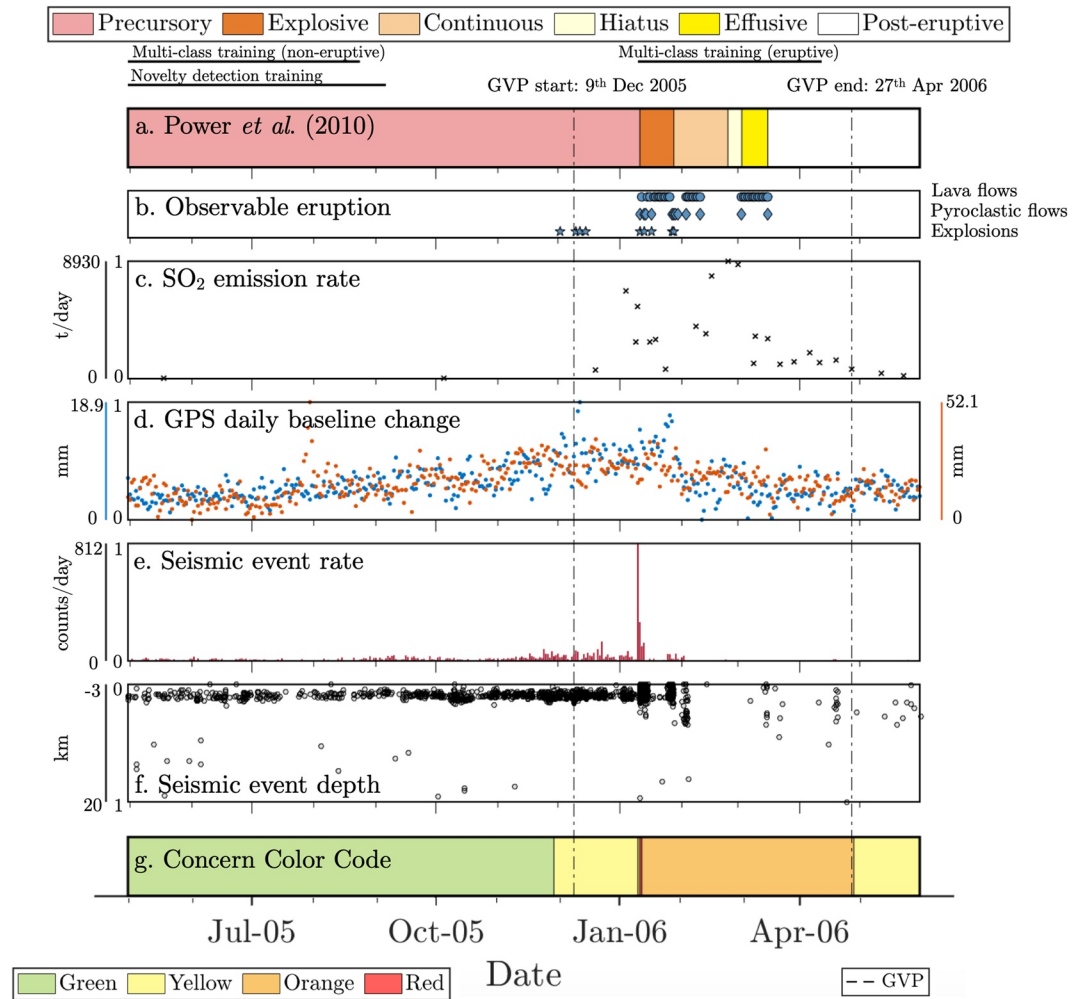


Figure 3. Timeline of the data period from April 2005 to May 2006 used in this paper and classification of the eruption timeline outlined in Section 2.2.1. Dashed vertical lines denote the GVP beginning and end dates for the eruptive period. Black horizontal lines above the top bar represent the non-eruptive and eruptive training periods for the multi-class model (upper) and non-eruptive training period for the novelty detection model (lower). The top bar (a) represents the classification of activity presented in Power et al. (2010) and the bottom bar (g) represents the Concern Color Code status of Augustine throughout the data period. In the observable eruption bar (b), blue symbols denote days on which visual eruption was confirmed: stars represent days in which explosions were observed, diamonds represent days in which pyroclastic flows were observed and circles represent days in which there was evidence of lava dome growth and/or effusion (compiled from Coombs et al., 2010). In the GPS panel (d), blue dots represent the baseline length changes of the N-components of AV02-AC59 and orange dots represent the AV01 vertical component, which are the station and component combinations used to derive geodetic features. Measurements of daily SO₂ emission rate (c), GPS baseline length changes (d), seismic event rate (which is one seismic feature) (e) and seismic event depth (f) are normalized between 0 and 1 and the true bounds of these quantities are contained within the second y-axes for each plot. Seismic event depths are not used to derive features.

2.2.1. Seismic Time Series

2.2.1.1. Catalog Curation

We create a catalog by performing a full network detection on the seismic waveform data from Augustine volcano between April 30, 2005 and May 31, 2006. We use a Short-Term Averaging/Long-Term Averaging (STA/LTA) detection, with detection parameters from DeRoin and McNutt (2012). We perform this detection on the full network of seismic stations (Figure 2), using the GISMO toolbox for MATLAB (Thompson & Reyes, 2018). The detection parameters for identifying a trigger were an STA window of 1s and an LTA window of 8s, with an STA/LTA threshold of 2 required for event triggering. The STA/LTA algorithm is used in “frozen” mode, meaning that the LTA window is fixed once an event is triggered. The detection was run twice over the data, once using a

low-frequency filter (0.5–10 Hz) and once using a high-frequency filter (3–18 Hz). For a detection to be included as an event we checked that it was detected on 3 different stations within 10 s of each other. To exclude nearby regional events, we checked the resulting detection times against the USGS earthquake catalog (USGS, 2021). We considered all events occurring in the data period of at least 2.5 magnitude within 500 km of Augustine Island and manually checked events in our detection which coincided with a travel time between 30 and 300s of the USGS catalog, excluding approximately 50 events.

The event rates from the network detection are in agreement with event rates quoted in Buurman and West (2010). A comparison of the two event rates highlights good agreement on days with no events and for days with the highest event rates during the explosive phase of the eruption (January 11 and 13, 2006). For days with medium event rates, our catalog yields a daily event count approximately 1 order of magnitude greater than the AVO catalog. Event rates in the explosive phase onwards were affected by the outages of summit seismic stations, resulting in fewer detected events meeting the criteria for inclusion in the catalog.

Alaska Volcano Observatory (AVO) published an analyst-reviewed catalog comprising 2,367 seismic events associated with the 2006 eruption (Dixon & Stihler, 2009). For an event to be included in the AVO catalog, it must have greater than 3 P phases and greater than 2 S phases in addition to locating an event hypocenter to within 15 km of the volcano and with standard hypocentral errors. Due to these criteria for inclusion, the AVO catalog represents a subset of the total events at Augustine. Buurman and West (2010) find that 39% of the full number of events they detect at Augustine are included in the AVO catalog, and that the criteria for an event to be included mean that small and emergent seismic events are underrepresented in the catalog.

2.2.1.2. Seismic Feature Extraction

Using event times from both the AVO catalog and our catalog, we derive a set of seismic features from the event waveforms at station AUI, a seismic station which was in place before the eruption and was not disabled during the eruption (listed in Tables S1 and S2 in Supporting Information S1). We use the number of seismic events detected per day, and any days with zero seismic events are omitted from the time series. Though the gaps in the time series could be filled with a substitute value, e.g., the mean value of the feature (Schafer & Graham, 2002), we suggest that no value should be substituted as no events were detected therefore it would not follow to interpolate a value for the event characteristics of non-existent events. Dominant frequency information is derived by recording the 5 peak frequencies from each event in a given day. The ratio of low to high frequency energy in a waveform can be expressed via the band ratio (Buurman & West, 2010; Rodgers et al., 2015) and is calculated by taking the base 2 log of the ratio of energy within two windows of low frequency (1–2 Hz) and high frequency (10–20 Hz). These windows have previously been applied to data from Augustine (Buurman & West, 2010). We calculate peak amplitude based on the maximum amplitude of the event normalized by its mean amplitude. We compute the fast Fourier transform of each waveform and calculate the spectral width of the largest peak of each event.

After calculating features for each event on a given day, we then derive the set of seismic features for each day by taking the mean, median, variance, minimum, maximum, 10th percentile, 90th percentile and change in mean from the previous day for all categories apart from event rate, for which the per-day value and change in value from previous day are used.

We derive two sets of features from the seismic data, and compare the two sets of features in Section 4.2.

1. Features from seismic events from our catalog
2. Features from seismic events from the AVO Augustine catalog

2.2.2. Gas Time Series

SO₂ emission rate measurements for Augustine are available from July 2002 to July 2008. Measurements of SO₂ column abundance were made via airborne traverses, using COSPEC (McGee et al., 2010). The SO₂ emission rates which we use for this analysis were processed by McGee et al. (2010), who processed the raw COSPEC traverses by calculating the area under the SO₂ traverse time series and divided by the approximate width. This average signal was then multiplied by the plume width and average wind speed to determine an SO₂ emission rate in units of t/day. The time series of SO₂ emission rate is irregular, reflecting the challenges of making these airborne measurements. The median interval between measurements is 8 days, and the mean interval between

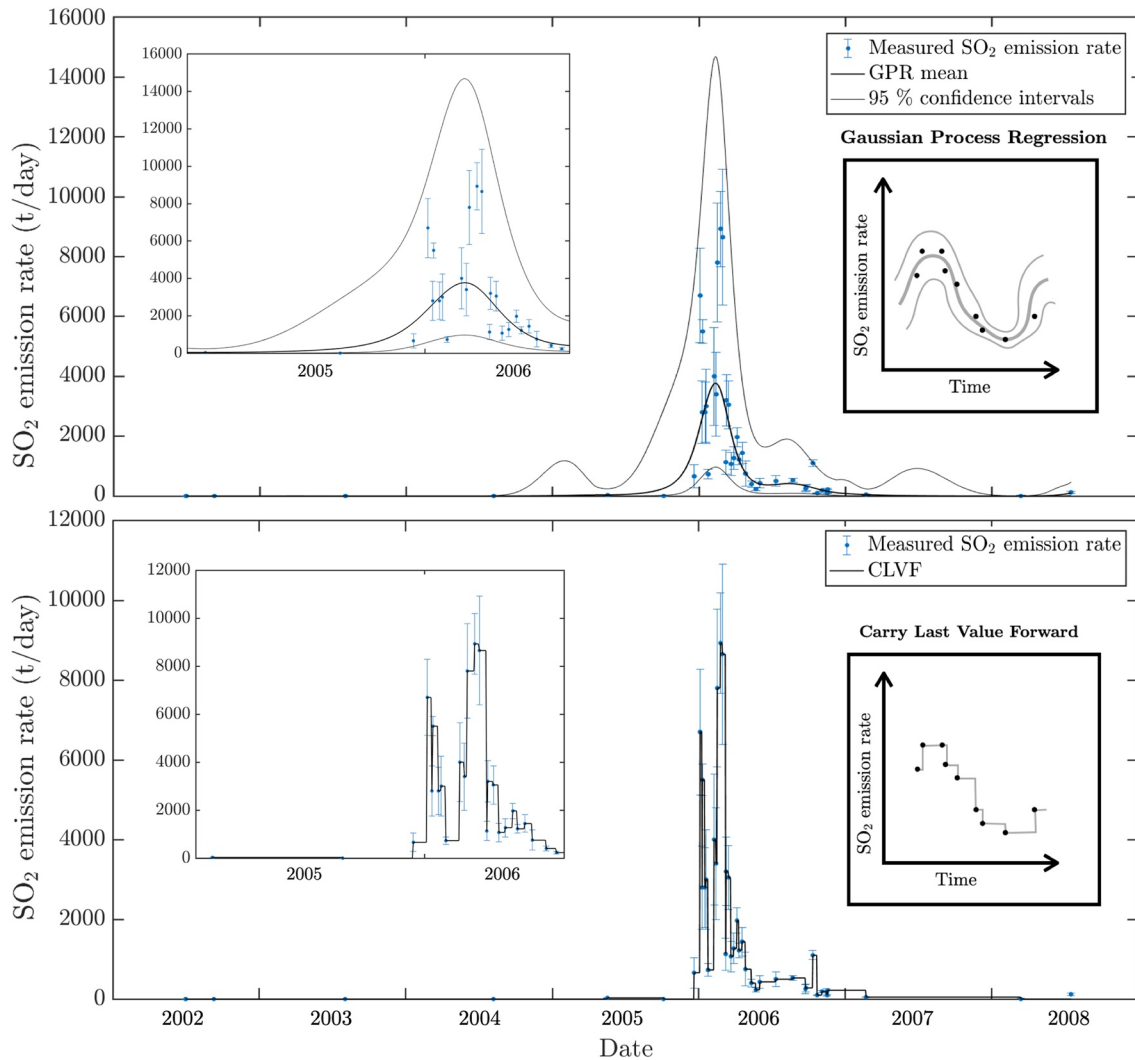


Figure 4. Daily SO₂ emission rate values for the measurements made at Augustine between 2002 and 2008 and values interpolated using Gaussian process regression (GPR; top panel) and Carry last value forward (CLVF; bottom panel). Blue points represent the values processed and reported in McGee et al. (2010) with bars denoting two standard deviations and vertical lines indicating the SO₂ emission rates during the study period. Inset figures within the panels represent an enlarged version of the SO₂ emission rates during the study period (left) and a conceptual illustration of each method (right).

measurements is 50 days. There are larger intervals between measurements during the pre-eruptive and post-eruptive phases than the eruptive phases as the measurement frequency increased when seismic unrest began in 2005.

To use the gas time series data as input for our models alongside the daily time series from geodetic and seismic datasets, we use interpolation to estimate values of SO₂ emission rate between measurements made at Augustine. Figure 4 illustrates the two methods of interpolation used:

1. Gaussian Process Regression (GPR): a probabilistic approach in which a stochastic model is applied to provide estimates of intervening measurements with confidence intervals, in which no functional form is assumed.
2. Carry Last Value Forward (CLVF): until a new measurement of SO₂ emission rate is made, every day has the same value of emission flux.

To interpolate the gas measurements using the GPR, we fit the model to the measurements of SO₂ emission rate. The values are first logged as we do not want to fit a model which can permit negative values for SO₂ emission rate, and because the noise in the data set is multiplicative that is, the uncertainties in emission rate increase as the magnitude of the emission rate increases. Before the eruption, there are values of zero in the data set which

would be undefined in log space, so we choose a value an order of magnitude smaller than the lowest SO₂ emission rate as a “detection limit” of 10⁰ t/day. We use a matern covariance function (Rasmussen & Williams, 2006), which has previously been used to capture processes which are not smoothly varying in meteorological datasets (Guttorp & Gneiting, 2006).

These methods of interpolation each have advantages and disadvantages. CLVF is a conservative estimate of the daily emission rate, using no future emission rate measurements to interpolate the daily values. However, CLVF does not provide an estimate of uncertainty on the emission rate measurements, and also permits large jumps in the daily value of SO₂ emission rate. GPR does not assume an underlying functional form for the SO₂ emission rate time series and also yields an uncertainty bound on the interpolated daily values. The upper uncertainty bound broadly follows the form of the mean, apart from either side of the unrest and eruption (2004–2005 and 2007) where the uncertainty is higher due to an increase or decrease of measured emission rate respectively. This upper bound is not used as a feature in the models. GPR requires the use of all of the data points to estimate values for the intervening days and thus is a retrospective method. The total SO₂ flux estimated using CLVF interpolation is 4.3×10^5 t and the mean total SO₂ flux estimated using GPR is 4.5×10^5 t with 95% confidence interval bounds of 1.3×10^5 t and 2.0×10^6 t. Though neither method is preferable to a complete time series of measured daily values of SO₂ emission rate, the methods used for interpolating intervening daily values presented here represent two contrasting approaches to the problem of missing data in gas time series, which might find applications to estimating appropriate time-averaged SO₂ fluxes as well as for detecting changes in eruptive state as explored here.

Though measurements of CO₂ and H₂S were attempted at the same time as SO₂, there are under half as many values obtained for the emission rates of these species compared to SO₂ measurements (McGee et al., 2010). No value of CO₂ or H₂S emission rate is available for the continuous or hiatus phase during the eruption, and there are significant gaps in the precursory and post-eruptive phases of the eruption. Consistent measurements of SO₂ emission rate of 0 t/day during the precursory and post-eruptive phases could be due to the process of SO₂ scrubbing by groundwater (Symonds et al., 2001). As the precursory phase of the eruption occurred during the Alaskan winter, the presence of snowpack on Augustine could also have contributed to scrubbing and suppressed the quantity of SO₂ measured at the surface. We derive features only from the SO₂ emission rate, and take the data as reported with no modification to account for potential scrubbing processes. The features extracted from the full SO₂ emission rate time series are the daily estimated value and the change in daily value from the previous day.

2.2.3. Geodetic Measurements

A GPS network comprising 6 stations was in place before the 2006 eruption of Augustine (Figure 2; Cervelli et al., 2010). This network began to detect precursory deformation around July 2005, and the timing of the first dike intrusions is inferred from deformation measurements as occurring in November 2005 (Cervelli et al., 2006). Data availability from stations was variable during the eruption due to the destruction of stations by volcanic deposits during the later stages of eruption.

We use the daily GPS solutions provided by Nevada Geodetic Laboratory (Blewitt et al., 2018) in the International GNSS Service 14 reference frame (Rebischung & Schmid, 2016). We choose stations AV01, AV02, and AC59, which had a complete record of measurements before, during and after the eruption. For station AV01 we extract the daily vertical position. The difference between the maximum and minimum value of AV01 vertical component during the study period is 52.1 mm, a variation which exceeds the typical seasonal signal observed at AV01. For each day we also calculate the variation with respect to the previous day. For station pair AV02-AC59 we compute the daily change of the N-S component of the baseline vector with respect to a fixed chosen day. For the geometry of these two stations the baseline variation is a proxy for N-S extension or shortening across the volcano (e.g., Cervelli et al., 2010; Dixon, 1991; Saballos et al., 2014). The magnitude of change for the AV02-AC59 N-S component of the baseline is 18.9 mm. For each day we also calculate the variation with respect to the previous day. The features we use are daily vertical position, daily change in the N-S component of the baseline, and the difference with respect to the previous day of both the vertical position and the N-S component of the baseline.

2.3. Machine Learning Methods

Classification methods aim to categorize input observations into a finite number of categories which are representative of different types of activity of a physical system (Bishop, 2006). The methods we use are supervised,

meaning that the model is trained on a selected set of training examples with assigned labels and then subsequently tested on unseen data. During the model training process, some training points (known as validation data) are held back and used to test the model in order to maximize the training accuracy. This validation data is distinct from the testing data which is subsequently used to evaluate the effectiveness of the models in classifying eruptive state. The models presented in Section 3 are trained on the seismic, gas and geodetic features described in Section 2.2. Classification methods used here can be split into one-class and multi-class classification, which are described in further detail below.

2.3.1. Novelty Detection Classification

Novelty detection (also known as one-class classification) methods involve training the model on a single class of observation, and then testing subsequent observations against this model to identify patterns which are anomalous or novel (Pimentel et al., 2014). In this approach, typically the training or “normal” class is better sampled than the abnormal or “novel” data with the objective of identifying the onset of the “novel” data. A novelty detection approach is advantageous for systems in which there may be several possible “abnormal” modes, and in which those abnormal modes are not well characterized. Existing applications of novelty detection include medical diagnostics (e.g., Quinn & Williams, 2007) and jet engine condition monitoring (Clifton et al., 2007). In the geosciences, novelty detection has been previously applied to remote sensing for the purposes of classifying the occurrence of a single type of land use (Li et al., 2010; Muñoz-Marí et al., 2010).

Use of novelty detection in volcanology is novel. Volcanic systems are conceptually well-suited to a novelty detection approach, as there is a broad spectrum of volcanic styles (thus failure types) associated with eruption. Moreover, larger eruptions are associated with longer repose times (Passarelli & Brodsky et al., 2012; Pyle, 1998) and thus previous eruptions may not have been fully instrumented (and therefore well-characterized). We use a nearest-neighbor-based approach to novelty detection, using the NDtool novelty detection toolbox (Clifton, 2007; Pimentel et al., 2014). For a nearest neighbor approach, the Euclidean distance between a test sample and its nearest neighbor within the training data is computed. The novelty score is then calculated by comparing this distance to the maximum distance between that nearest neighbor within the training data and its nearest neighbor in the entire training set. Hence, a larger novelty score is achieved where the test sample lies far outside the region of the training data as the distance between the testing sample and its nearest neighbor will greatly exceed the distance between that nearest neighbor and any point in the training data. The novelty threshold is defined as the maximum novelty score calculated within the training data set. Novelty detection models require a model training period which is sufficiently long to capture the distribution of the normal class, and we select the minimum possible training length for successful model training in order to test the model over the months preceding the eruption. We select a training period for the novelty detection model which begins on April 30, 2005, the first day of the data period, and ends on September 6, 2005. We hold back 1/5 of the training points to use as validation data for the model. This training period represents approximately 30% of the total number of days spanned by the data set.

2.3.2. Multi-Class Classification

Multi-class classification methods involve training models on two or more classes of observation, and then testing on unseen data. In this study we apply binary classification, with two classes: “non-eruptive” data and “eruptive” data, and we select representative training examples from both classes. For the non-eruptive training data, we select days at the beginning of the data period in order to train as far away from the hypothesized timing of the start of eruption, from April 30, 2005–August 24, 2005. For the eruptive training period, although the GVP start-date was a month earlier, visible eruption was limited to a short period of steam explosions and therefore we choose a training period from the main explosive phase of the eruption onwards, from January 10, 2006–April 12, 2006. This training period represents approximately 40% of the total number of days spanned by the data set.

We apply a random forest method, in which an ensemble of decision trees is averaged to produce a classification, wherein the decision represents the conversion from model inputs to final classification. Individual decision trees represent a series of operations which compare the features to randomly selected thresholds (Hastie et al., 2001). The aggregation of decision trees means that the final models have good generalization. We use standard 5-fold cross validation in order to test the model performance during the training process (Hastie et al., 2001). Once a final aggregated tree is formulated, the model is then tested on the unseen data. This method was successfully applied in previous work classifying volcanic state from features derived from seismic data (Manley et al., 2020).

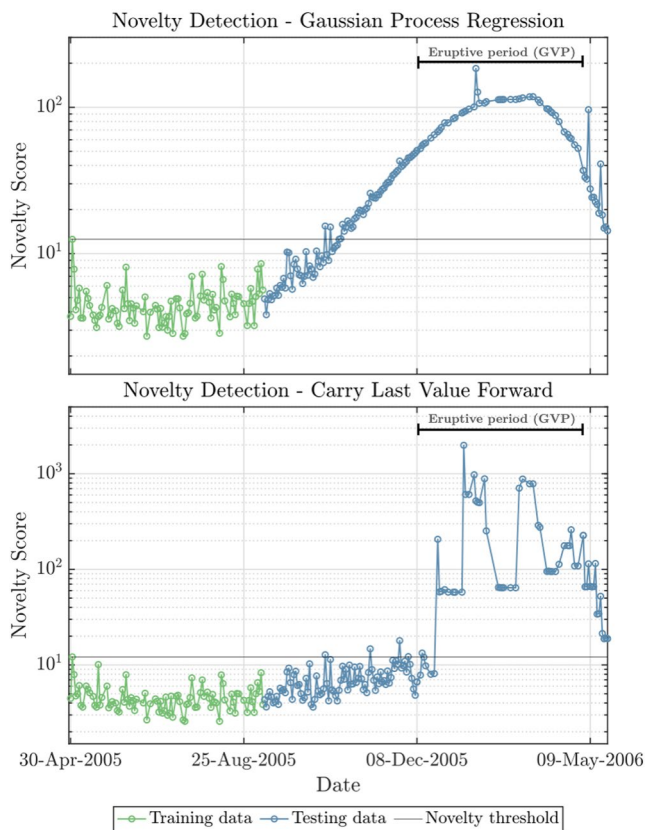


Figure 5. Results from novelty detection classification applied to the features derived from Augustine data. Green line denotes the training data and blue line denotes the testing data. The gray horizontal line represents the novelty threshold determined during model training. The black bar at the top of each panel represents the dates on which eruption was occurring according to the GVP database. Novelty score is dimensionless, therefore the y-axis has no associated units but note the different y-axis scales between the top and the bottom panels.

2.4. Assumptions

The classification approaches applied here involve the underlying assumption that the variables used as input are Independent and Identically Distributed (IID). This assumption implies that features on a given day are mutually independent from features from other days, and features are drawn from the same underlying distribution between days (Cover & Thomas, 2006). Data are rarely perfectly IID in practice; however, we hypothesize that this approximation holds to a first-order and observe whether this hypothesis is supported by the resulting models distinguishing between previously unseen eruptive and non-eruptive data. This assumption was shown to be acceptable to a first order in previous work involving these classification methods (Manley et al., 2020) and we therefore retain this assumption for the seismic features. In this paper, we also assume that the individual gas and geodetic changes will be approximately IID for similar reasons to the seismic data: the generative process which produces measured geodetic and gas emission rates is assumed to not have memory of values on previous days. This assumption is less constrained for the interpolated GPR and CLVF values of SO_2 emission rate as the values interpolated for the missing days depend on the nearest measured values. Nonetheless, interpolation of the missing gas data is preferable to excluding the days with no measurements altogether. Gill et al., (2007) establish a similar methodology to fill missing values in hydrological datasets prior to applying machine learning algorithms, and find significantly better performance for models with interpolated values compared to models formulated with missing values omitted from the data set.

3. Results

We present results from one-class and multi-class classification models trained and tested independently on time series of features derived from the Augustine seismic, geodetic and gas data. We compare classification results using two end-member approaches for interpolation of the SO_2 emission rate: CLVF and GPR, introduced in Section 2.2.1.

3.1. Novelty Detection Results

Figure 5 illustrates the results from novelty detection methods using features derived from seismic, geodetic and both gas data interpolation methods, which are also summarized in Figure 6c and Table S6 in Supporting Information S1. These models were both formulated using a nearest-neighbor approach to novelty detection. Each day of the model is classified independently, therefore there is no influence of previous or future novelty scores on the current day. There are several observations which can be made from the novelty detection results (Figure 5):

1. Both models begin to consistently classify days as eruptive prior to the phreatic explosions in December 2005. The novelty scores begin to exceed the novelty threshold in mid-November 2005 for both gas interpolation methods (GPR and CLVF). This classification of abnormal behavior precedes the change in Concern Color code at Augustine (Cameron et al., 2018).
2. The overall shape of the novelty score is comparable to the method of SO_2 emission rate interpolation, thus the CLVF novelty scores appear step based and GPR novelty scores are smoothly varying.
3. There is agreement between the models on the end-date of eruption—GPR novelty scores and CLVF novelty scores do not return below novelty threshold within the data period end-date of May 2006.
4. Precursory classification of abnormal behavior in the CLVF model occurs whilst SO_2 emission rates remain at 0 t/day, indicating that changes in the geodetic and seismic features are primarily responsible for the novelty scores.

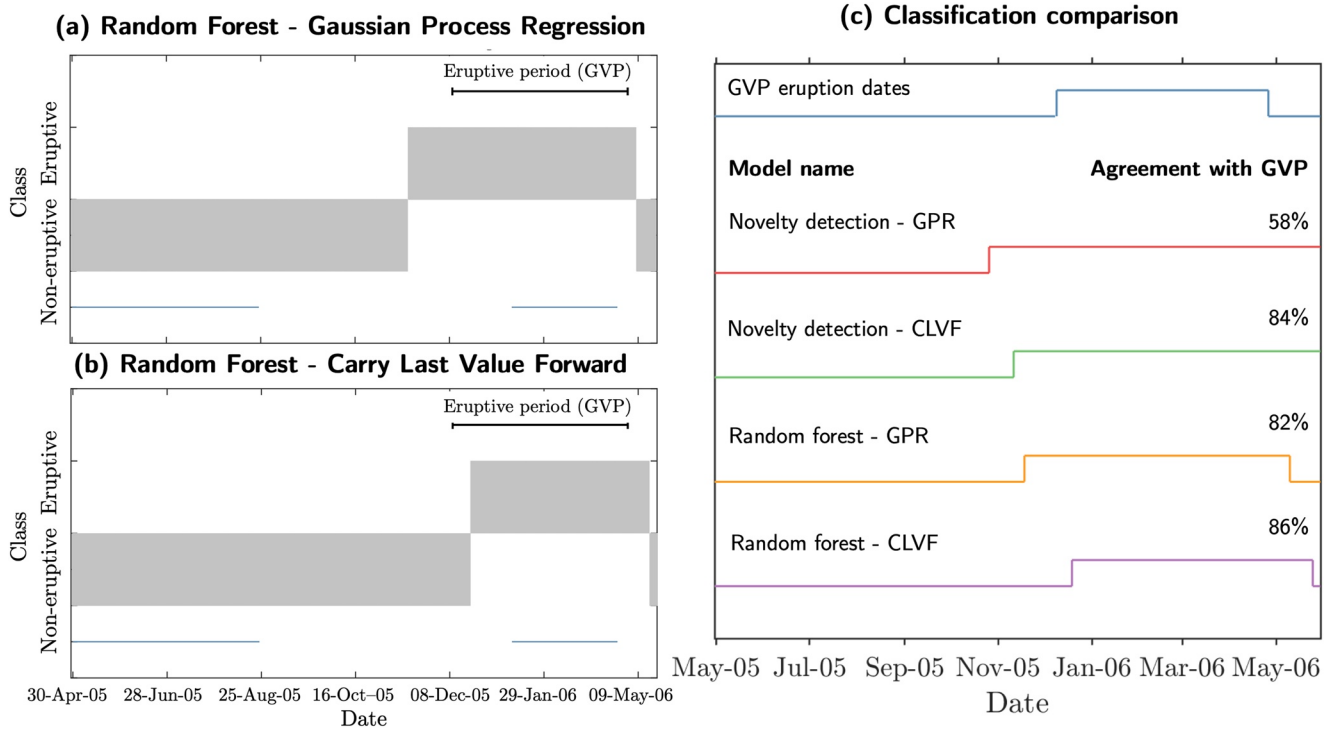


Figure 6. Results from the multi-class results from the random forest classification applied to the seismic, geodetic and gas features derived from Augustine data with (a) Gaussian Process Regression and (b) Carry Last Value Forward interpolation of gas time series data. The gray squares in panels (a) and (b) represent the classification output of the model corresponding to either non-eruptive or eruptive. The horizontal blue line beneath the classification results represents the training period of the model. The black bar at the top of each panel represents the dates on which eruption was occurring according to the GVP database. (c) Model comparison of the classification results from novelty detection and random forest methods with GVP dates. Each line represents the eruptive state (non-eruptive or eruptive) in GVP or as classified by the model, with the eruptive state beginning when a model begins to classify data as eruptive and ending when a model begins to classify data as non-eruptive. For the novelty detection models, the novelty score does not come below the threshold before the end of the study period. Model agreement with GVP is calculated by dividing the number of classified days which match the GVP classifications by the total number of classified days, expressed as a percentage. Please refer to Table S6 in Supporting Information S1 for more details.

We compare the novelty detection model results to the GVP labels and find 84% testing agreement for the CLVF-based model compared to GVP labels and 58% agreement for GPR-based model compared to GVP labels. The misclassifications with respect to the GVP labels are entirely false positives, that is, more days are classified as eruptive compared to the dates recorded in GVP. The model agreements reported here are computed relative to the labels recorded in GVP, where the beginning of the eruption is recorded as December 9, 2005 and the end-date is April 27, 2006. Labels from GVP are defined by the presence or absence of visual evidence of volcanic activity. However, as discussed in Section 1 and by Manley et al., (2020), a range of subsurface processes are sampled by seismic, gas and geodetic observations which are not necessarily accompanied by surface activity and therefore reflected in the GVP labels.

3.2. Multi-Class Results

The results from random forest classification are illustrated in Figure 6 and outlined in Table S6 in Supporting Information S1. As in the one-class models, each day of the model is classified independently thus the classification on any given day is not influenced by the classification on another day. The following key observations can be made from the two random forest approaches:

1. The model formulated using seismic, geodetic and GPR-interpolated gas features (Figure 6a) begins to classify days as eruptive in mid-November 2005, whereas the model based on seismic, geodetic and CLVF-interpolated gas features (Figure 6b) begins to classify days as eruptive in mid-December 2005.

2. There is some agreement between the models on the end-date of eruption—the CLVF-based model returns to non-eruptive classification in early mid May 2006, whereas the GPR-based model end-date is slightly earlier, in early May.

The agreement of the random forest models with GVP labels is higher for both interpolation approaches than the one-class approach. Overall model agreement for the model with CLVF-interpolated gas features is 86% and model agreement for the model with GPR-interpolated gas features is 82%. The multi-class model with CLVF-interpolated gas features begins to classify days as eruptive later than all other novelty detection or multi-class models. However, the classification of eruptive activity still begins approximately a month before the main explosive phase of the eruption, in which the first juvenile material associated with the eruption was ejected.

4. Discussion

4.1. Physical Interpretation of Model Results

Both CLVF-based and GPR-based novelty detection models and GPR-based multi-class models begin to classify days as eruptive during the precursory phase of the Augustine eruption, around November 2005. For the GPR-based novelty detection model, the novelty score does not return below the novelty threshold for the remainder of the study period whereas the CLVF-based novelty scores have isolated peaks above the novelty threshold. Previous qualitative models of activity during the precursory phase of eruption are consistent with transition at this time. On November 17, 2005, there was a marked offset in the GPS-derived offsets between AV04 and AV05, which Cervelli et al. (2010) attribute to a crack opening at a depth near sea level, and visual observations of surface cracks at the summit. A day later on November 18, the rate of edifice inflation also increased (Cervelli et al., 2010).

Previous modeling of ascent duration based on 40–100 micron reaction rims in amphiboles from the LSA magma is consistent with a ~20–57 days duration of ascent (Larsen et al., 2010). This petrologically derived constraint is consistent with dike initiation around November 17 and 18, as the shortest ascent duration would be associated with an eruption date around December 9, 2005 (consistent with phreatic explosions) and the longest ascent duration would be associated with an eruption date around January 12, 2006 (consistent with beginning of the explosive phase). Therefore, the transition which we identify using machine learning approaches in the precursory phase could be associated with physical changes associated with dike initiation.

Earthquakes in the early precursory phase of activity at Augustine, prior to dike initiation, are attributed to the accumulation of gas below the conduit (Larsen et al., 2010). The depth of earthquakes which would be generated by this mechanism is consistent with the model of Huppert and Sparks (2016), who model the pressure distribution associated with the rise of compressible vapor through a porous conduit. In this model, the gas pressure scales with the square root of depth, whereas the lithostatic pressure scales linearly with depth. Therefore, a critical overpressure is thought to occur approximately 2/3 of the distance between the source depth (i.e., magma storage zone) and the surface (i.e., the edifice) when the gas pressure exceeds the lithostatic pressure and causes rock fracture, thereby nucleating VT events. The observed depths of precursory VTs during the unrest at Augustine are consistent with this model, as the events are concentrated in a fracture zone 3–4 km above the primary magma storage depth. A similar process has been observed at Soufrière Hills Volcano, Montserrat, where VT swarms occurred in an overpressure zone ahead of the ascending magma (Rodgers et al., 2016). The eruptive transition we identify from our classification modeling could correspond to the source mechanism of precursory earthquakes changing as dike intrusion began. This hypothesis is supported by Syracuse et al. (2011) who relocated seismic events at Augustine with an improved velocity model, and found that from November 2005 onwards seismic event locations began to align in a NW-SE trend consistent with intrusive volumes modeled by Cervelli et al. (2010) and the regional stress field (Buurman et al., 2014).

The transport of magma to the surface via dike intrusion is also associated with changes in gas measurements and ground deformation (e.g., Sparks et al., 2012). Increased emission of SO₂ can represent shallow magmatic ascent of gas-rich magma, as the degassing of SO₂ occurs at lower pressures than CO₂ (Moran et al., 2011). Where gas scrubbing processes have suppressed the surface emission of SO₂ prior to eruption, increases in SO₂ emission rate can represent the boiling and subsequent removal of a hydrothermal system to allow previously scrubbed SO₂ emissions to reach the surface (Pritchard et al., 2019). Magmatic ascent is often associated with deformation

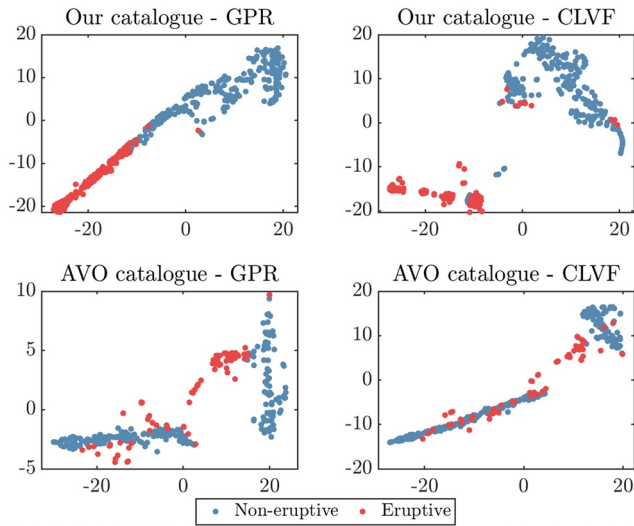


Figure 7. Two-dimensional visualisation of the seismic, gas flux and geodetic features derived from Augustine using the t-distributed stochastic nearest neighbor embedding (t-SNE; van der Maaten & Hinton, 2008). Blue points represent non-eruptive days and red points represent eruptive days according to the Global volcanism program (GVP) database. The top two panels, which include seismic features from our seismic catalog, display a greater separation between non-eruptive and eruptive days than the bottom two panels, which include seismic features from the Alaska volcano observatory (AVO) analyst-reviewed catalog. A three-dimensional visualisation of these datasets using t-SNE is plotted in Figure S2 in Supporting Information S1. The x and y axes represent non-linearly scaled distances projected into a two-dimensional space and thus have no units, and are not comparable across axes.

which, in the case of the Augustine eruption, was recorded as an increase in edifice inflation rate (Cervelli et al., 2010).

4.2. Importance of Feature Groups

4.2.1. Comparison of Seismic Catalogs

In this study, we derive features from seismic, gas and geodetic data. Figure 7 is a 2-dimensional visualisation of the full set of seismic, gas flux and geodetic features (with 3-dimensional representation in Figure S2 in Supporting Information S1). This visualisation is performed using the t-Distributed Stochastic Nearest Neighbor Embedding (t-SNE), which is a probabilistic machine learning algorithm designed to project high-dimensional data—the set of 40 features—into lower dimensional space such as 2- or 3-dimensions (van der Maaten & Hinton, 2008). This process is performed by calculating a joint probability for points in the high-dimensional space using the conditional probability that a given datapoint would pick another as its neighbor, given a probability distribution described by a Gaussian centered at that point. The low-dimensional joint probability is calculated using the Student t-distribution with one degree of freedom. The final mapping is calculated by minimizing the Kullback-Leibler divergence between these two joint probabilities, such that points which are close in high-dimensional space remain close in the two-dimensional representation. We apply the t-SNE to four sets of features: the two sets of seismic and geodetic features used for classification in Section 3 (with GPR or CLVF derived daily values for gas), and two sets of features derived from the AVO seismic catalog (with the same geodetic and GPR or CLVF derived daily values for gas flux).

Comparison of the t-SNE projections for the set of seismic, gas and geodetic features containing seismic features derived from our seismic event catalog shows that the non-eruptive and eruptive days are clustered in separate areas

of the projected space, with little overlap between the two clusters. The transition between non-eruptive and eruptive data is smoother for GPR-based interpolation than CLVF-based interpolation, which reflects the continuous nature of GPR-based interpolation. Novelty detection classification methods rely on the ability to fully characterize the training class of data, so that outliers can be reliably identified. The distinction between clusters in the projection of the Augustine data set illustrates that there are systematic differences between the eruptive and non-eruptive days of the data set which can be exploited by a one-class approach. Comparison of the projections of features from our catalog and AVO catalog-derived features indicates that the differences in eruptive and non-eruptive data are less distinct for the AVO catalog, with no clear separation of data into clusters on this basis. The increased separability of features from our catalog suggests that the smaller, more emergent events not included in the AVO analyst-reviewed catalog contain meaningful information which contributes to distinguishing between eruptive and non-eruptive data.

4.2.2. Comparison of Multi-Parameter and Single-Parameter Models

For comparison against the full sets of features used for models in Section 3, we trained and tested three additional novelty detection models using features derived from only the seismic, gas or geodetic data with the same training dates (Figure S1 in Supporting Information S1). The models formulated using seismic data and geodetic data alone have a similar trend to the model using the full set of features in the precursory months, with novelty scores from both models exceeding the novelty threshold in November 2005. However, both seismic and geodetic models fail to consistently classify days from the main eruptive sequence above the novelty threshold. In the case of the seismic data, we propose that this model failure arises due to outages in summit seismic stations during the explosive phase of the eruption, which means that smaller events are underrepresented in our catalog during the later stages of the eruption. For the geodetic data, the inflation signal was concentrated in the precursory phase and, once the main phase of the eruption commenced, the daily displacement decreased until similar to pre-inflation values. Models formulated using gas features with a GPR interpolation classified behavior as eruptive as

soon as model training ended, whereas models formulated using gas with a CLVF interpolation fail to identify an appropriate novelty threshold as the interpolated SO₂ emission rates (hence novelty scores) remained very low until significant emissions were measured on December 20, 2005. The GPR failure likely results from the smoothly varying nature of the interpolated time series where there is no stable normal behavior to characterize during model training, whereas the CLVF failure is due to the low surface emission of SO₂ during the precursory phase.

The results of this single data modeling indicate that, in the case of the 2006 Augustine eruption, use of multiple types of data for classification is able to discern between non-eruptive and eruptive data better than single data-streams alone for novelty detection models. When all of the features are included in the model, the precursory increasing trend in novelty score remains similar to the individual seismic and geodetic models and eruptive days are better identified due to the inclusion of the SO₂ emission rate features. Moreover, the single data analysis indicates that the inclusion of multiple datasets can compensate for data quality issues due to instrument failures during a large eruption. Applying multi-class models to single data-streams was not attempted due to the issues identified above pertaining to isolating suitable eruptive and non-eruptive training periods for each of the data-streams separately.

4.3. Applications of the Method

4.3.1. Forecasting Eruptions

Both one-class and multi-class methods, explored in this paper, have potential utility in forecasting. Poland and Anderson (2020) note that machine learning approaches represent an as-yet underutilized resource in eruption forecasting, which is in part due to limited quantities of available training data. The choice to apply one type of machine learning model over another involves consideration of the past volcanic activity and/or monitoring infrastructure available at a given volcanic system, as these factors affect the data available for training models. Here, we discuss possible applications of novelty detection versus multi-class approaches to the detection of eruptive transitions in different volcanic scenarios.

Novelty detection approaches—those in which models are trained only on non-eruptive data—would be applicable to monitored volcanoes which have been in repose for some time (e.g., 10s years), as these systems would not have eruptive data available for training but a large amount of data from repose to characterize the non-eruptive class. This class of volcanoes would include many of those that have experienced major eruptions in recent decades (Pyle & Barclay, 2020). Other potential applications include volcanoes which have displayed several types of activity in the past or have highly variable manifestations of eruptive activity. Here, a novelty detection approach would not assume a consistent mode of eruption. Additionally, volcanoes where the monitoring infrastructure has significantly changed since the previous eruption are a good candidate for novelty detection, as the eruptive data from the previous eruption would not necessarily characterize the eruptive data on the updated monitoring network.

Multi-class approaches—those in which models are trained on both non-eruptive and eruptive data—could be more useful at volcanoes which display consistent signals associated with eruption. Here, eruptive training data available from past eruptions would well-characterize the eruptive behavior of the system in order to identify future eruptive signals. For example, Piton de la Fournaise volcano has a consistent tremor signal that is correlated with eruption (Battaglia & Aki, 2003) and most eruptive episodes are preceded by precursory VT swarms (De Barros et al., 2013), which has made it a good candidate for previous multi-class studies (Ren et al., 2020). Volcanoes with short repose time between eruptions could be better-suited to a multi-class approach, as there would be insufficient time to characterize the non-eruptive class with a one-class approach.

For the purposes of forecasting eruptions, consideration of which methods are used to interpolate values for missing data points are necessary. CLVF-based interpolation involves no future information to fill in missing values, therefore is well-suited to a forecasting context. GPR-based interpolation uses the full set of data points to construct the model from which the daily values for the training period are taken and therefore is a purely retrospective analysis in this study. However, Colopy et al., (2018) demonstrate that Gaussian Processes can be successfully applied in healthcare for forecasting patient vital signs beyond the model training periods. Future work could take Gaussian Processes into the domain of real-time volcanic forecasting.

4.3.2. Elucidating Timings of Transition in Volcanic State

Future applications of classification could extend the two-class approach here to investigate intra-eruptive transitions in activity. Prolonged eruptions may undergo intra-eruptive transitions; for example, dome-forming eruptions often switch between effusive and explosive behavior (e.g., Edmonds & Herd, 2007; Preece et al., 2016). Characterizing and understanding the nature and timing of changes in eruptive intensity and style, and anticipating future transitions in behavior, is critical for managing hazards during eruptive crises (e.g., Barclay et al., 2019; Fearnley & Beaven, 2018). For systems in which the transitions between activity are sufficiently long to generate enough training data, one-class or multi-class methods trained during eruption could help to distinguish intra-eruptive transitions.

Consideration of the methods for interpolating gaps in irregularly sampled time series is necessary. As discussed in Section 3, we observe different end-dates for the 2006 eruption depending on the SO₂ interpolation methods used. We find that models formulated based on GPR interpolation of the SO₂ time series yield an earlier estimate of eruption end-date than models formulated based on CLVF-based interpolation. This observation is explained due to the 'carry forward' nature of CLVF: at the end of the eruption where the SO₂ emission rate is consistently decreasing between observations, CLVF represents an overestimate of the daily emission rate and therefore data is classified as eruptive for a longer period. Consideration of interpolation methods becomes more important as the sampling frequency of the time series decreases, as more points need to be estimated for sparser time series.

Application of the approach presented in this paper is not limited to the data streams used here. Additional ground-based data streams from which features could be derived include infrasound or thermal measurements. Satellite-derived observations of ground displacement (e.g., InSAR) are being made in increasing frequency, and could be incorporated into a machine learning framework. Indeed, there is a growing body of literature examining the automatic detection of unrest in InSAR time series (Anantrasirichai et al., 2018; Biggs et al., 2019; Gaddes et al., 2019; Valade et al., 2019).

5. Conclusions

Identifying transitions in multi-parameter time series is a challenge in volcanology. We apply novelty detection (one-class) and multi-class machine learning methods to features derived from seismic, gas and geodetic data from the 2006 eruption of Augustine Volcano, Alaska. We present the first study applying a one-class or novelty detection approach to classify volcanic state, which does not require training examples characterizing eruptive behavior. We suggest a framework for processing multi-parameter data to produce daily representative features, including two distinct approaches toward interpolating missing values in SO₂ emission rate time series. These approaches include Carry Last Value Forward (CLVF) interpolation in which the daily value remains constant since the previous measurement of emission rate, and Gaussian Process Regression (GPR) which is a probabilistic approach using the full available data to estimate intervening values with uncertainty bounds. We compare the results of models formulated using seismic, geodetic and both CLVF and GPR estimates of daily gas emission rates. Both novelty detection and multi-class random forest models identify a key change-point from non-eruptive to eruptive state in mid-November. This change-point is in agreement with previous conceptual models at Augustine which suggest the initiation of dike intrusion from the shallow magmatic storage at this time. We find that novelty detection models formulated using multiple types of data are able to distinguish between non-eruptive and eruptive data better than models formulated on single data-streams alone. We suggest possible applications of one-class novelty detection and multi-class modeling and discuss considerations in applying these approaches in a monitoring or retrospective context. Novelty detection methods show promising application for monitoring volcanoes where previous eruptions have not been observed, or where the monitoring infrastructure has significantly changed since previous eruptions.

Data Availability Statement

Augustine seismic data was collected by the AV network (AVO/USGS, 1988). Augustine geodetic data was taken from the continuous GPS network installed as part of the National Science Foundation Plate Boundary Observatory (Cervelli et al., 2006) and can be obtained from Blewitt et al. (2018). Augustine gas data can be obtained

from McGee et al. (2010). We used the Generic Mapping Tools version 6 (Wessel et al., 2019). The features and code to run the models from this paper can be found at Manley (2021).

Acknowledgments

We thank R. Malservisi for guidance on the geodetic data analysis. This research has been supported by a Natural Environment Research Council (NERC) studentship (NE/L002612/1). DAC was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). We thank David Dempsey and two anonymous reviewers for their constructive comments on an earlier version of the manuscript.

References

- Aiuppa, A., Moretti, R., Federico, C., Giudice, G., Gurrieri, S., Liuzzo, M., et al. (2007). Forecasting Etna eruptions by real-time observation of volcanic gas composition. *Geology*, *35*(12), 1115–1118. <https://doi.org/10.1130/g24149a.1>
- Alaska Volcano Observatory/USGS. (1988). *Alaska Volcano Observatory*. International Federation of Digital Seismograph Networks. <https://doi.org/10.7914/SN/AV>
- Anantrasirichai, N., Biggs, J., Albino, F., Hill, P., & Bull, D. (2018). Application of machine learning to classification of volcanic deformation in routinely generated InSAR Data. *Journal of Geophysical Research: Solid Earth*, *123*(8), 6592–6606. <https://doi.org/10.1029/2018jb015911>
- Anderson, K., & Segall, P. (2013). Bayesian inversion of data from effusive volcanic eruptions using physics-based models: Application to Mount St. Helens 2004–2008. *Journal of Geophysical Research: Solid Earth*, *118*(5), 2017–2037. <https://doi.org/10.1002/jgrb.50169>
- Barclay, J., Few, F., Armijos, M. T., Phillips, J. C., Pyle, D. M., Hicks, A. J., et al. (2019). Livelihoods, wellbeing and the risk to life during volcanic eruptions. *Frontiers of Earth Science*, *7*, 205. <https://doi.org/10.3389/feart.2019.00205>
- Battaglia, J., & Aki, K. (2003). Location of seismic events and eruptive fissures on the Piton de la Fournaise volcano using seismic amplitudes. *Journal of Geophysical Research*, *108*(B8), 2364. <https://doi.org/10.1029/2002JB002193>
- Biggs, J., Albino, F., Dogru, F., Yip, S., Yu, C., Li, Z., & Anantrasirichai, N. (2019). Automated methods for detecting volcanic deformation using Sentinel-1 InSAR time series. *AGU Fall Meeting Abstracts* (Vol. 2019, p. G22A-01).
- Biggs, J., Ebmeier, S. K., Aspinall, W. P., Lu, Z., Pritchard, M. E., Sparks, R. S. J., & Mather, T. A. (2014). Global link between deformation and volcanic eruption quantified by satellite imagery. *Nature Communications*, *5*(1), 1–7. <https://doi.org/10.1038/ncomms4471>
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Blewitt, G., Hammond, W. C., & Kreemer, C. (2018). Harnessing the GPS data explosion for interdisciplinary science. *Eos Transactions American Geophysical Union*, *99*. <https://doi.org/10.1029/2018EO104623>
- Boué, A., Lesage, P., Cortés, G., Valette, B., & Reyes-Dávila, G. (2015). Real-time eruption forecasting using the material Failure Forecast Method with a Bayesian approach. *Journal of Geophysical Research: Solid Earth*, *120*(4), 2143–2161.
- Brantley, S. R. (1990). *The eruption of Redoubt Volcano, Alaska, December 14, 1989–August 31, 1990* (Vol. 1061, p. 33). US Geological Survey Circular.
- Buurman, H., Nye, C. J., West, M. E., & Cameron, C. (2014). Regional controls on volcano seismicity along the Aleutian arc. *Geochemistry, Geophysics, Geosystems*, *15*(4), 1147–1163. <https://doi.org/10.1002/2013gc005101>
- Buurman, H., & West, M. (2010). Seismic precursors to volcanic explosions during the 2006 eruption of Augustine Volcano. In J. Power, M. Coombs, & J. Freymueller (Eds.), *The 2006 eruption of Augustine Volcano*. U.S. Geological Survey Professional Paper 1769, Alaska (U.S. Geological Survey Professional Paper 1769).
- Cameron, C. E., Prejean, S. G., Coombs, M. L., Wallace, K. L., Power, J. A., & Roman, D. C. (2018). Alaska volcano observatory alert and forecasting timeliness: 1989–2017. *Frontiers of Earth Science*, *6*, 86. <https://doi.org/10.3389/feart.2018.00086>
- Cervelli, P. F., Fournier, T., Freymueller, J., & Power, J. A. (2006). Ground deformation associated with the precursory unrest and early phases of the January 2006 eruption of Augustine Volcano, Alaska. *Geophysical Research Letters*, *33*(18). <https://doi.org/10.1029/2006gl027219>
- Cervelli, P. F., Fournier, T. J., Freymueller, J. T., Power, J. A., Lisowski, M., & Pauk, B. A. (2010). Geodetic constraints on magma movement and withdrawal during the 2006 eruption of Augustine Volcano: Chapter 17. In J. A. Power, M. L. Coombs, & J. T. Freymueller (Eds.), *The 2006 eruption of Augustine Volcano* (pp. 427–452) Alaska. U.S. Geological Survey. <https://doi.org/10.3133/pp176917>
- Chastin, S. F., & Main, I. G. (2003). Statistical analysis of daily seismic event rate as a precursor to volcanic eruptions. *Geophysical Research Letters*, *30*, 13. <https://doi.org/10.1029/2003GL016900>
- Chouet, B. A., & Matoza, R. S. (2013). A multi-decadal view of seismic methods for detecting precursors of magma movement and eruption. *Journal of Volcanology and Geothermal Research*, *252*, 108–175. <https://doi.org/10.1016/j.jvolgeores.2012.11.013>
- Clifton, D. A., Bannister, P. R., & Tarassenko, L. (2007). A framework for novelty detection in jet engine vibration data. In: *Key Engineering Materials* (Vol. 347, pp. 305–310). Trans Tech Publications Ltd. <https://doi.org/10.4028/www.scientific.net/kem.347.305>
- Clifton, L. A. (2007). *Multi-channel novelty detection and classifier combination*. The University of Manchester.
- Collombet, M., Grasso, J. R., & Ferrazzini, V. (2003). Seismicity rate before eruptions on Piton de la Fournaise volcano: Implications for eruption dynamics. *Geophysical Research Letters*, *30*(21), 2099. <https://doi.org/10.1029/2003GL017494>
- Colopy, G. W., Roberts, S. J., & Clifton, D. A. (2018). Bayesian optimization of personalized models for patient vital-sign monitoring. *IEEE Journal of Biomedical and Health Informatics*, *22*(2), 301–310. <https://doi.org/10.1109/jbhi.2017.2751509>
- Coombs, M. L., Bull, K. F., Vallance, J. W., Schneider, D. J., Thoms, E. E., Wessels, R. L., & McGimsey, R. G. (2010). Timing, distribution, and volume of proximal products of the 2006 eruption of Augustine Volcano, chapter 8. In J. A. Power, M. L. Coombs, & J. T. Freymueller (Eds.), *The 2006 eruption of Augustine Volcano*. U.S. Geological Survey Professional. Retrieved from https://pubs.usgs.gov/pp/1769/chapters/p1769_chapter08
- Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Wiley-Interscience.
- De Barros, L., Bean, C. J., Zecevic, M., Brenguier, F., & Peltier, A. (2013). Eruptive fracture location forecasts from high-frequency events on Piton de la Fournaise Volcano. *Geophysical Research Letters*, *40*(17), 4599–4603. <https://doi.org/10.1002/grl.50890>
- Dempsey, D. E., Cronin, S. J., Mei, S., & Kempa-Liehr, A. W. (2020). Automatic precursor recognition and real-time forecasting of sudden explosive volcanic eruptions at Whakaari, New Zealand. *Nature Communications*, *11*(1), 1–8. <https://doi.org/10.1038/s41467-020-17375-2>
- DeRoin, N., & McNutt, S. R. (2012). Rockfalls at Augustine Volcano, Alaska: The influence of eruption precursors and seasonal factors on occurrence patterns 1997–2009. *Journal of Volcanology and Geothermal Research*, *211*, 61–75. <https://doi.org/10.1016/j.jvolgeores.2011.11.003>
- DeShon, H. R., Thurber, C. H., & Power, J. A. (2010). Earthquake waveform similarity and evolution at Augustine Volcano from 1993 to 2006: Chapter 5. In J. A. Power, M. L. Coombs, & J. T. Freymueller (Eds.), (Eds.) *The 2006 eruption of Augustine Volcano* (pp. 103–118). U.S. Geological Survey. <https://doi.org/10.3133/pp17695>
- Dixon, J. P., & Stihler, S. D. (2009). *Catalog of earthquake hypocenters at Alaskan Volcanoes: January 1 through December 31, 2008* (Vol. 467, p. 88). U.S. Geological Survey Data Series.
- Dixon, T. H. (1991). An introduction to the Global Positioning System and some geological applications. *Reviews of Geophysics*, *29*(2), 249–276. <https://doi.org/10.1029/91rg00152>

- Ebmeier, S. K., Biggs, J., Mather, T. A., & Amelung, F. (2013). On the lack of InSAR observations of magmatic deformation at Central American volcanoes. *Journal of Geophysical Research: Solid Earth*, 118(5), 2571–2585. <https://doi.org/10.1002/jgrb.50195>
- Edmonds, M., & Herd, R. (2007). A volcanic degassing event at the explosive-effusive transition. *Geophysical Research Letters*, 34, L21310. <https://doi.org/10.1029/2007gl031379>
- Endo, E. T., Murray, T. L., & Power, J. A. (1996). A comparison of preeruption real-time seismic amplitude measurements for eruptions at Mount St. Helens, Redoubt Volcano, Mount Spurr, and Mount Pinatubo. *Fire and Mud*, 233–247.
- Fearnley, C. J., & Beaven, S. (2018). Volcano alert level systems: Managing the challenges of effective volcanic crisis communication. *Bulletin of Volcanology*, 80(5), 46. <https://doi.org/10.1007/s00445-018-1219-z>
- Fischer, T. P., Morrissey, M. M., Calvache, V. M. L., Gomez, M. D., Torres, C. R., Stix, J., & Williams, S. N. (1994). Correlations between SO₂ flux and long-period seismicity at Galeras volcano. *Nature*, 368(6467), 135–137. <https://doi.org/10.1038/368135a0>
- Gaddes, M. E., Hooper, A., & Bagnardi, M. (2019). Using machine learning to automatically detect volcanic unrest in a time series of interferograms. *Journal of Geophysical Research: Solid Earth*, 124(11), 12304–12322. <https://doi.org/10.1029/2019jb017519>
- Geirsson, H., Rodgers, M., LaFemina, P., Witter, M., Roman, D., Muñoz, A., et al. (2014). Multidisciplinary observations of the 2011 explosive eruption of Telica volcano, Nicaragua: Implications for the dynamics of low-explosivity ash eruptions. *Journal of Volcanology and Geothermal Research*, 271, 55–69. <https://doi.org/10.1016/j.jvolgeores.2013.11.009>
- Gill, M. K., Asefa, T., Kaheil, Y., & McKee, M. (2007). Effect of missing data on performance of learning algorithms for hydrologic predictions: Implications to an imputation technique. *Water Resources Research*, 43(7). <https://doi.org/10.1029/2006wr005298>
- Global Volcanism Program. (2006). Report on Augustine (United States). In R. Wunderman (Ed.), *Bulletin of the Global Volcanism Network* (Vol. 31, p. 1). Smithsonian Institution. <https://doi.org/10.5479/si.GVP.BGVN200601-313010>
- Guttorp, P., & Gneiting, T. (2006). Studies in the history of probability and statistics XLIX on the Matérn correlation family. *Biometrika*, 93(4), 989–995. <https://doi.org/10.1093/biomet/93.4.989>
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. Springer.
- Huppert, H. E., & Sparks, R. S. (2016). Compressible vapour flow in conduits and fractures. *Journal of Fluid Mechanics*, 802, 750–759. <https://doi.org/10.1017/jfm.2016.298>
- Jacobs, K. M., McNutt, S. R., & Power, J. A. (2010). Using seismic b-values to interpret seismicity rates and physical processes during the preeruptive earthquake swarm at Augustine Volcano 2005–2006. *The 2006 Eruption of Augustine Volcano, Alaska* (Vol. 1769, pp. 1–17). US Geol. Surv. Prof. Pap.
- Kilburn, C. R. (2003). Multiscale fracturing as a key to forecasting volcanic eruptions. *Journal of Volcanology and Geothermal Research*, 125(3–4), 271–289. [https://doi.org/10.1016/s0377-0273\(03\)00117-3](https://doi.org/10.1016/s0377-0273(03)00117-3)
- Larsen, J. F., Nye, C. J., Coombs, M. L., Tilman, M., Izbekov, P., & Cameron, C. (2010). *Petrology and geochemistry of the 2006 eruption of Augustine Volcano, chapter 15*. In J. A. Power, M. L. Coombs, & J. T. Freymueller (Eds.), *The 2006 eruption of Augustine Volcano*. U.S. Geological Survey Professional Paper 1769, p. 335–382 and spreadsheets. https://pubs.usgs.gov/pp/1769/chapters/p1769_chapter15
- Li, W., Guo, Q., & Elkan, C. (2010). A positive and unlabeled learning algorithm for one-class classification of remote-sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 49(2), 717–725.
- Lu, Z., & Dzurisin, D. (2014). InSAR imaging of Aleutian volcanoes. In *InSAR imaging of Aleutian volcanoes* (pp. 87–345). Berlin: Springer. https://doi.org/10.1007/978-3-642-00348-6_6
- Manley, G. (2021). Machine learning approaches to identifying changes in eruptive state using multi-parameter datasets from the 2006 eruption of Augustine Volcano, Alaska (Code repository). *Zenodo*. <https://doi.org/10.5281/zenodo.5226248>
- Manley, G. F., Pyle, D. M., Mather, T. A., Rodgers, M., Clifton, D. A., Stokell, B. G., et al. (2020). Understanding the timing of eruption end using a machine learning approach to classification of seismic time series. *Journal of Volcanology and Geothermal Research*, 401, 106917. <https://doi.org/10.1016/j.jvolgeores.2020.106917>
- McGee, K. A., Doukas, M. P., McGimsey, R. G., Neal, C. A., & Wessels, R. L. (2010). Emission of SO₂, CO₂, and H₂S from Augustine Volcano 2002–2008, chapter 26. In J. A. Power, M. L. Coombs, & J. T. Freymueller (Eds.), *The 2006 eruption of Augustine Volcano* (pp. 609–627). U.S. Geological Survey Professional Paper 1769. <https://doi.org/10.3133/pp176926>
- Moran, S. C., Newhall, C., & Roman, D. C. (2011). Failed magmatic eruptions: Late-stage cessation of magma ascent. *Bulletin of Volcanology*, 73(2), 115–122. <https://doi.org/10.1007/s00445-010-0444-x>
- Múnoz-Marí, J., Bovolo, F., Gómez-Chova, L., Bruzzone, L., & Camp-Valls, G. (2010). Semisupervised one-class support vector machines for classification of remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 48(8), 3188–3197.
- National Academies of Sciences, Engineering, and Medicine. (2017). *Volcanic eruptions and their repose, unrest, precursors, and timing*. The National Academies Press. <https://doi.org/10.17226/24650>
- Passarelli, L., & Brodsky, E. E. (2012). The correlation between run-up and repose times of volcanic eruptions. *Geophysical Journal International*, 188(3), 1025–1045. <https://doi.org/10.1111/j.1365-246x.2011.05298.x>
- Pesicek, J. D., Wellik, J. J., Prejean, S. G., & Ogburn, S. E. (2018). Prevalence of seismic rate anomalies preceding volcanic eruptions in Alaska. *Frontiers of Earth Science*, 6, 100. <https://doi.org/10.3389/feart.2018.00100>
- Petersen, T., De Angelis, S., Tytgat, G., & McNutt, S. R. (2006). Local infrasound observations of large ash explosions at Augustine Volcano, Alaska, during January 11–28, 2006. *Geophysical Research Letters*, 33(12). <https://doi.org/10.1029/2006gl026491>
- Pimentel, M. A., Clifton, D. A., Clifton, L., & Tarassenko, L. (2014). A review of novelty detection. *Signal Processing*, 99, 215–249. <https://doi.org/10.1016/j.sigpro.2013.12.026>
- Poland, M. P., & Anderson, K. R. (2020). Partly cloudy with a chance of lava flows: Forecasting volcanic eruptions in the twenty-first century. *Journal of Geophysical Research: Solid Earth*, 125(1), e2018JB016974. <https://doi.org/10.1029/2018jb016974>
- Power, J. A., Coombs, M. L., & Freymueller, J. T. (Eds.). (2010). *The 2006 eruption of Augustine Volcano, Alaska* (Vol. 1769, p. 667). U.S. Geological Survey. p., 1 plate, scale 1:20,000, and data files.
- Power, J. A., & Lalla, D. J. (2010). Seismic observations of Augustine Volcano, 1970–2007: Chapter 1. In J. A. Power, M. L. Coombs, & J. T. Freymueller (Eds.), *The 2006 eruption of Augustine Volcano* (pp. 3–40). U.S. Geological Survey Professional Paper 1769. https://pubs.usgs.gov/pp/1769/chapters/p1769_chapter01.pdf
- Power, J. A., Nye, C. J., Coombs, M. L., Wessels, R. L., Cervelli, P. F., Dehn, J., et al. (2006). The reawakening of Alaska's Augustine volcano. *EOS. Transactions - American Geophysical Union*, 87(37), 373–377. <https://doi.org/10.1029/2006eo370002>
- Preece, K., Gertisser, R., Barclay, J., Charbonnier, S. J., Komorowski, J. C., & Herd, R. A. (2016). Transitions between explosive and effusive phases during the cataclysmic 2010 eruption of Merapi volcano, Java, Indonesia. *Bulletin of Volcanology*, 78(8), 54. <https://doi.org/10.1007/s00445-016-1046-z>

- Pritchard, M. E., Mather, T. A., McNutt, S. R., Delgado, F. J., & Reath, K. (2019). Thoughts on the criteria to determine the origin of volcanic unrest as magmatic or non-magmatic. *Philosophical Transactions of the Royal Society A*, 377(2139), 20180008. <https://doi.org/10.1098/rsta.2018.0008>
- Pyle, D. M. (1998). Forecasting sizes and repose times of future extreme volcanic events. *Geology*, 26(4), 367–370. [https://doi.org/10.1130/0091-7613\(1998\)026<0367:fsarto>2.3.co;2](https://doi.org/10.1130/0091-7613(1998)026<0367:fsarto>2.3.co;2)
- Pyle, D. M., & Barclay, J. (2020). Historical records of volcanic eruptions deserve more attention. *Nature Reviews Earth and Environment*, 1, 183–184. <https://doi.org/10.1038/s43017-020-0044-z>
- Quinn, J. A., & Williams, C. K. (2007). June. Known unknowns: Novelty detection in condition monitoring. In *Iberian Conference on Pattern Recognition and Image Analysis* (pp. 1–6). Berlin: Springer.
- Rasmussen, C. E., & Williams, C. K. (2006). *Gaussian processes for Machine learning, vol. 1* (Vol. 39, pp. 40–43). MIT press.
- Rebuschung, P., & Schmid, R. (2016). *IGS14/igs14. Atx: A new framework for the IGS products*. AGU Fall Meeting.
- Ren, C. X., Peltier, A., Ferrazzini, V., Rouet-Leduc, B., Johnson, P. A., & Brenguier, F. (2020). Machine learning reveals the seismic signature of eruptive behavior at piton de la fournaise volcano. *Geophysical Research Letters*, 47(3), e2019GL085523. <https://doi.org/10.1029/2019GL085523>
- Rodgers, M., Roman, D. C., Geirsson, H., LaFemina, P., McNutt, S. R., Muñoz, A., & Tenorio, V. (2015). Stable and unstable phases of elevated seismic activity at the persistently restless Telica Volcano, Nicaragua. *Journal of Volcanology and Geothermal Research*, 290, 63–74. <https://doi.org/10.1016/j.jvolgeores.2014.11.012>
- Rodgers, M., Smith, P. J., Mather, T. A., & Pyle, D. M. (2016). Quiescent-explosive transitions during dome-forming volcanic eruptions: Using seismicity to probe the volcanic processes leading to the 29 July 2008 Vulcanian explosion of Soufrière Hills Volcano, Montserrat. *Journal of Geophysical Research: Solid Earth*, 121(12), 8453–8471. <https://doi.org/10.1002/2016jb013180>
- Roman, D. C., LaFemina, P. C., Bussard, R., Stephens, K., Wauthier, C., Higgins, M., et al. (2019). Mechanisms of unrest and eruption at persistently restless volcanoes: Insights from the 2015 eruption of Telica Volcano, Nicaragua. *Geochemistry, Geophysics, Geosystems*, 20. <https://doi.org/10.1029/2019gc008450>
- Roman, D. C., Rodgers, M., Geirsson, H., LaFemina, P. C., & Tenorio, V. (2016). Assessing the likelihood and magnitude of volcanic explosions based on seismic quiescence. *Earth and Planetary Science Letters*, 450, 20–28. <https://doi.org/10.1016/j.epsl.2016.06.020>
- Saballos, J. A., Conde, V., Malservisi, R., Connor, C. B., Alvarez, J., & Munoz, A. (2014). Relatively short-term correlation among deformation, degassing, and seismicity: A case study from Concepción volcano, Nicaragua. *Bulletin of Volcanology*, 76(8), 1–11. <https://doi.org/10.1007/s00445-014-0843-5>
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147–177. <https://doi.org/10.1037/1082-989x.7.2.147>
- Segall, P. (2013). Volcano deformation and eruption forecasting. *Geological Society, London, Special Publications*, 380(1), 85–106. <https://doi.org/10.1144/sp380.4>
- Siebert, L., Cottrell, E., Venzke, E., & Andrews, B. (2015). Earth's volcanoes and their eruptions: An overview. In *The Encyclopedia of Volcanoes* (pp. 239–255). Academic Press. <https://doi.org/10.1016/b978-0-12-385938-9.00012-2>
- Sparks, R. S. J. (2003). Dynamics of magma degassing. *Geological Society, London, Special Publications*, 213(1), 5–22. <https://doi.org/10.1144/gsl.sp.2003.213.01.02>
- Sparks, R. S. J., Biggs, J., & Neuberg, J. W. (2012). Monitoring volcanoes. *Science*, 335(6074), 1310–1311. <https://doi.org/10.1126/science.1219485>
- Symonds, R. B., Gerlach, T. M., & Reed, M. H. (2001). Magmatic gas scrubbing: Implications for volcano monitoring. *Journal of Volcanology and Geothermal Research*, 108(1–4), 303–341. [https://doi.org/10.1016/s0377-0273\(00\)00292-4](https://doi.org/10.1016/s0377-0273(00)00292-4)
- Syracuse, E. M., Thurber, C. H., & Power, J. A. (2011). The Augustine magmatic system as revealed by seismic tomography and relocated earthquake hypocenters from 1994 through 2009. *Journal of Geophysical Research*, 116(B9). <https://doi.org/10.1029/2010jb008129>
- Thompson, G., & Reyes, C. (2018). *GISMO - A seismic data analysis toolbox for MATLAB (Version 1.20b) [software package]*. Retrieved from <http://geoscience-community-codes.github.io/GISMO/>
- U.S. Geological Survey. (2021). *Earthquake Lists, Maps, and Statistics*. Retrieved from <https://www.usgs.gov/natural-hazards/earthquake-hazards/lists-maps-and-statistics>
- Valade, S., Ley, A., Massimetti, F., D'Hondt, O., Laiolo, M., Coppola, D., et al. (2019). Towards global volcano monitoring using multisensor sentinel missions and artificial intelligence: The MOUNTS monitoring system. *Remote Sensing*, 11(13), 1528. <https://doi.org/10.3390/rs11131528>
- Vallance, J. W., Bull, K. F., & Coombs, M. L. (2010). Pyroclastic flows, lahars, and mixed avalanches generated during the 2006 eruption of Augustine Volcano: Chapter 10. In J. A. Power, M. L. Coombs, & J. T. Freymueller (Eds.), *The 2006 eruption of Augustine Volcano* (pp. 219–267). U.S. Geological Survey Professional Paper 1769. <https://doi.org/10.3133/pp176910>
- van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(11).
- Voight, B. (1988). A method for prediction of volcanic eruptions. *Nature*, 332(6160), 125–130. <https://doi.org/10.1038/332125a0>
- Voight, B., & Cornelius, R. R. (1991). Prospects for eruption prediction in near real-time. *Nature*, 350(6320), 695–698. <https://doi.org/10.1038/350695a0>
- Webster, J. D., Mandeville, C. W., Goldoff, B., Coombs, M. L., & Tappen, C. (2010). Augustine Volcano - The influence of volatile components in magmas erupted A.D. 2006 to 2,100 years before present: Chapter 16. In J. A. Power, M. L. Coombs, & J. T. Freymueller (Eds.), *The 2006 eruption of Augustine Volcano* (pp. 383–423). U.S. Geological Survey Professional Paper 1769. <https://doi.org/10.3133/pp176916>
- Werner, C. A., Doukas, M. P., & Kelly, P. J. (2011). Gas emissions from failed and actual eruptions from Cook Inlet Volcanoes, Alaska, 1989–2006. *Bulletin of Volcanology*, 73(2), 155–173. <https://doi.org/10.1007/s00445-011-0453-4>
- Wessel, P., Luis, J. F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H. F., & Tian, D. (2019). The generic mapping tools version 6. *Geochemistry, Geophysics, Geosystems*, 20(11), 5556–5564. <https://doi.org/10.1029/2019gc008515>