

SHORT REPORT

Open Access



How big will the next eruption be?

Paul Colosi and Emily E. Brodsky*

Abstract

Anticipating the size of the next volcanic eruption in long-term forecasts is a major problem in both basic and applied volcanology. In this study, we investigate the extent to which eruption size is predictable based on historical and other attribute data. Data from the Smithsonian Global Volcanism Program (GVP) Catalog is used to determine the predictability of volcanic eruption size as quantified through the reported VEI (Volcano Explosivity Index). The numerical and categorical attributes from the global volcanic catalog were classified with trained random forest and simple prediction models to make a forecast of VEI that can be tested against the most recent eruption of each volcano. We compare these results to two different baseline predictability levels by: (a) selecting randomly from the global distribution of VEIs for the most recent eruptions to calculate a cohort baseline and (b) selecting the most frequent VEI for a given population to calculate a zero-rule baseline. We found that: (1) nearly any method that incorporates prior information on a specific volcano improves the prediction accuracy of the succeeding eruption VEI by at least 10 percentage points relative to the cohort baseline case, (2) incorporating attributes beyond previous VEIs can provide better accuracy and achieve up to 30 percentage point accuracy gains, (3) total accuracy of the VEI forecasting by these methods can be up to nearly 80% and (4) the zero-rule is an effective prediction method that is modestly outperformed (~ 5 percentage point gain) by random forest methods with multiple attributes on most datasets. We find no notable preference in accuracy based on volcano type. The results quantify the importance of volcano-specific information in long-term forecasting and may help practitioners assess their expected performance when anticipating future eruption size.

Keywords: VEI, Forecasting, Machine learning

Introduction

Volcanic eruption forecasting involves anticipating the location, timing, size, and type of eruption with as quantitative probabilities as possible (e.g., Decker 1986). Such efforts are normally informed by a combination of in situ monitoring, past eruptive history, and generalized understanding of the behavior of similar volcanoes. This blending of data types needs to be done on a case-by-case basis for both long-term and short-term forecasting. Predictive success has been evaluated retrospectively based on whether or not any eruption occurred when expected (Cameron et al. 2018; Caudron et al. 2020; Papale 2017; Poland & Anderson 2020; Winson et al. 2014). These

studies are an important stepping stone to quantifying the degree of certainty in the forecasts and improving the probabilities assigned to anticipated outcomes.

A different set of studies have evaluated the power of long-term forecasts based on the historical record and a small subset of these has focused specifically on the ability to forecast eruption size, rather than time (e.g., Bebbington 2014; Mendoza-Rosas & De la Cruz-Reyna 2008). Parameterized statistical models have been used to fit the observed distribution of eruption sizes and sometimes additional information has been incorporated such as repose time (e.g., Bebbington 2014). These studies to date have primarily focused on the handful of volcanoes on which sufficient data is present to fit the statistical models. This approach has had some success in evaluating physics-motivated hypotheses such as volume-predictability, however, it can only be applied to

*Correspondence: brodsky@ucsc.edu
Department of Earth & Planetary Sciences, University of California, Santa Cruz, CA 95060, USA



a very limited set of situations with sufficient information available. Such volcanoes are preferentially basaltic with short recurrence times. There is a gap in assessing long-term forecasting strategies of eruptive size based on the global catalog, which encompasses much more general cases than can currently be captured by individual focused studies.

As previous authors have noted, the parameterized approach has a second shortcoming in that it requires significant modification to incorporate additional information that might be relevant such as a volcano's morphology or typical petrology (Marzocchi & Bebbington 2012). A similar line of thinking underlies the VOLCANS method of identifying analog volcanoes (Tierz et al., 2019). Common practice might bound the expected Volcano Explosivity Index (VEI) in the next eruption based on these considerations. For instance, a basaltic shield volcano is much less likely to produce a VEI 6 eruption than a caldera with rhyolite flows. It would be helpful to incorporate this general knowledge of volcano behavior that is based on global patterns into a quantitative forecasting scheme.

Here we attempt to grapple with both of these gaps in forecasting volcanic size. We first aim to retrospectively forecast eruptive size in the common situation of data-poor volcanoes and assess the results. Based only on the admittedly problematic general-purpose catalogs available, how well can we apply knowledge from the global database to anticipate the size of the next volcanic eruption at a particular volcano? This is a practical question that influences the appropriate mitigation measures. It is also a question that could benefit from direct quantification to establish a point of reference for other, more sophisticated approaches. Secondarily, we aim to combine information from both the (limited) history of each volcano with other aspects of the volcano that are generally databased and might influence a professional's judgment of likely future behavior.

In this study, we specifically ask to what degree eruptive size as quantified by the Volcano Explosivity Index (VEI) can be anticipated based on the global eruption database as currently available in the most commonly utilized database which is the Global Volcanism Program Volcanoes of the World (Global Volcanism Program, 2013). We fully realize that for individual eruptions more information is commonly available, such as in situ monitoring. However, we also recognize that instrumentation is often limited. It is useful to understand to what degree future eruptive size can be anticipated based on past history and general volcano features in light of the trends captured in the global database.

The strategy of this study is to avoid parameterization by utilizing two distinct methods to mimic and formalize

common geologic practice. We first use simple predictors based on the previous history, such as median previous VEI, to predict future behavior. We then also use a machine learning algorithm to determine the predictability of volcanic eruption VEI with groups of attributes consisting of historical data from eruptions and intrinsic properties of volcanoes. We compare the prediction accuracy of both methods to a baseline case of randomly selecting VEI from the historical database. This empirical, rather than analytical, cohort prediction allows empirical determination of the variability of the prediction, which is useful to assess the import of apparent differences between models. The cohort prediction baseline preserves catalog artifacts and allows us to evaluate the relative efficacy of approaches in the presence of the non-idealities in the data. We also examine an alternative baseline by predicting that all eruptions in a given population will have the modal VEI for that dataset. Since all eruptions in the population have the same prediction with no further decision-making necessary, this second baseline is known as a zero-rule approach (Witten et al., 2017). It is the more stringent standard for performance and turns out to be a very effective method in itself. Robustness of the results is assessed by comparing forecasting skill for various subsets of the data to both of the baselines and examining which trends appear to be robust to selection criteria.

Data

We use the Smithsonian Global Volcanism Program (GVP) database to provide numerical and nominal attributes for a global record of volcanoes and their eruption events during the Holocene (Global Volcanism Program, 2013). Data was downloaded in February 2019 and thus is complete through 2018. This data was then separated based on the available history. Volcanoes with more than 2 eruptions in the Holocene were used to train the model and provide the initial evaluation metrics. We call this group the historied volcanoes. Volcanoes with less than 2 Holocene eruptions documented are called unhistoried and reserved for a test data set for late in the study to evaluate both the robustness of the model and the importance of history in predicting behavior.

We use VEI as a measure of eruptive size (Newhall & Self 1982). VEI has its limitations. For instance, VEI is not designed to differentiate effusive eruptions that might have common-sense distinctions based on eruptive volume. VEI is also intrinsically a coarser measure than other magnitude and intensity scales (Houghton et al. 2013; Pyle 2015). The published catalog is incomplete for small eruptions, and under-recording varies regionally and temporally (Mead & Magill 2014; Sheldrake & Caricchi 2017; Wang et al. 2020). Catalog procedures also

affect the data. For instance, eruptions that are explosive but have otherwise insufficient data to determine an accurate VEI are recorded with a default value of VEI 2 leading to an over-representation of VEI 2 in the database (Siebert et al., 2010). Despite the limitations of both VEI and the global catalog, we utilize this database because the practical goal of this study is to quantify the degree of predictability based on the global patterns. The large database is essential for training any automated classifier method and no other standard measure of eruptive size can be currently utilized in this way on such a large dataset.

We attempt to mitigate the problems with VEI by considering subsets of the data which will have different completeness characteristics and other biases. By comparing results across these suites we can assess the robustness of various VEI-forecasting strategies. Therefore, we perform all analyses on the full dataset (minimum VEI 0), as well as subsets with minimum VEIs of 1, 2, and 3. We also consider a subset of the data with all VEIs other than 2 to specifically eliminate the default values. It should be noted that the thresholded subsets produced unhistoried volcanoes in the higher thresholds ($VEI \geq 1,2,3$) that were recorded as historied volcanoes in a lower threshold group. Thus the number of historied and unhistoried volcanoes varies with VEI selection criteria. This was most extreme for the highest completeness threshold of $VEI \geq 3$ because of the scarcity of large volcanic eruptions.

We also subset the data based on year. Although there are significant regional variations, ~1500 marks a major change in the VEI completeness (Mead & Magill 2014). This retrospective study tries to mimic modern efforts to anticipate future VEI and thus the post-1500 catalog is probably a more appropriate comparison than the full Holocene record. We therefore limit the primary results to volcanic eruptions that only occurred after the year 1500 but retain and present the full Holocene catalog results as a secondary result for reference.

The forecasts here use attributes that include three types of measures: (1) intrinsic features of a volcano such as petrology or morphology, (2) additional measures of volcanic history, i.e., the repose time since the previous eruption and eruption duration and (3) statistics based on the VEIs of prior eruptions at a volcano (Table 1).

The intrinsic attributes of volcanoes include both numeric and categorical data types. Dominant petrology is used as a numeric value by converting the named rock type into the silica content for use in a collective numerical attribute model (Le Bas et al. 1986). The original category of dominant petrology is also retained and used as a categorical variable in all other models. Morphology and tectonic settings are used as categorical (nominal) data

types in the classification procedure. Morphology categories are condensed from the original database into 6 categories as shown in Table 2 following Pesicek et al. (2021).

For the additional measures of volcano history, we calculate the repose time and previous eruption duration by differencing the reported start and end dates. These values therefore are only utilized when start and end dates are reported.

The statistical measures in the first category are the last, median, minimum, maximum, and mode VEI from the historical database excluding the most recent eruption, which is reserved as a target value to be forecast. Since VEI is limited to integer values, medians are rounded so as to yield realizable predictions.

Methods

For each volcano, we attempt to retrospectively forecast the VEI of the most recent recorded eruption with a suite of methods. Success is measured based on the fraction of volcanoes with the most recent eruption VEI successfully forecast by each method. Since we are only forecasting one eruption per volcano, we do not score performance on each volcano individually and only consider global measures.

These forecasts are achieved using two types of methods: simple predictions based on VEI history and random forest classifier, which is a standard machine-learning classification method described below.

Simple Forecasts Based on VEI History

The first method is a simple prediction based solely on the VEI-history of a particular volcano as recorded in the global catalog. This procedure is meant to mimic the common sense approach of simply asking if the next eruption is likely to be similar to the last eruption. More specifically, we compare the most recent eruption recorded at each volcano to five different statistics of the prior eruptions: last (most recent prior eruption), median, mode, minimum, and maximum VEI. If any of these statistics for that volcano are equal to the most recent recorded VEI, then a forecast based on that particular statistic is marked as correct.

The performance needs to be compared to meaningful baselines meant to capture the probability of a chance accurate forecast. Since the catalog contains artifacts and each of the subsets has a distinct distribution of VEIs, we measure the baseline case empirically two ways. The first is a cohort baseline for each group of data that was calculated by producing a cumulative density function (CDF) of VEI from the most recent eruption of each cataloged volcano. This empirical CDF was then randomly sampled and then tested against the most recent eruption VEI of each volcano to compute the baseline accuracy and

Table 1 Model Attributes. Attributes used for prediction. Type of attribute can be “Cat” for Categorical or “Num” for Numeric. For categorical data, the number of unique classes are listed for each attribute. To provide context on relative data size, the number of values (*N*) available for the post-1500 VEI ≥ 0 dataset are provided. This is an example dataset and the number of values available for the other dataset can be read from the white numbers in the bar graphs of Figs. 1–2

Attribute	Type	Number of classes	Models Using Attribute	<i>N</i> in post-1500 historied VEI ≥ 0 dataset	<i>N</i> in post-1500 unhistoried VEI ≥ 0 dataset
<i>Intrinsic Predictors</i>					
Dominant Rock Type	Cat	10	All Attributes Categorical Data Dominant Rock Type	424	133
Morphology	Cat	6	All Attributes Categorical Data Morphology	431	145
Tectonic Setting	Cat	10	All Attributes Categorical Data Tectonic Setting	431	145
<i>Additional History Predictors</i>					
Repose Time	Num	N/A	All Attributes All Numeric Data Repose Time	321	N/A
Eruption Duration (previous eruption)	Num	N/A	All Attributes All Numeric Data Eruption Duration	281	N/A
<i>Simple Predictors</i>					
Last VEI	Num	N/A	All Attributes All Numeric Data Last VEI	431	N/A
Median VEI	Num	N/A	All Attributes All Numeric Data Median VEI	431	N/A
Mode VEI	Num	N/A	All Attributes All Numeric Data Mode VEI	431	N/A
Minimum VEI	Num	N/A	All Attributes All Numeric Data Minimum VEI	431	N/A
Maximum VEI	Num	N/A	All Attributes All Numeric Data Maximum VEI	431	N/A

Table 2 Condensed Morphology. The submorphologies on the right are all condensed into the general heading on the left column for analysis following Pesicek et al. (2021)

General Morphology	Sub Morphology
Stratovolcano	Complex Compound Lava Cone Lava Dome
Field Volcano	Cone Crater Row Explosive Cater Maar Pyroclastic Cone Tuff Cone Tuff Ring Volcanic Field
Shield Volcano	Fissure Vent
Caldera	Pyroclastic Shield

its standard deviation. The second is a zero-rule baseline that predicts that all future eruption VEI will be the most frequent VEI in the population. For instance for populations that include VEI ≥ 0 , VEI ≥ 1 or VEI ≥ 2 , the most common VEI is 2 and thus the zero-rule baseline is VEI = 2 for these datasets.

We then report the overall accuracy and gain relative to the baseline cases for the full dataset as well as all of the data subsets. As will be seen below, the gains are more meaningful since both baselines perform relatively well in terms of total accuracy.

Machine Learning-Based Forecasts

Volcanoes include both categorical and numeric data that could potentially be helpful for VEI prediction and thus

a fairly general classification method might be useful. A random forest is a natural choice for a multiclass problem (Hastie et al, 2017). The algorithm finds an ensemble of decision trees. Each tree deterministically selects a VEI class by splitting the data based on the values of the predictor attributes. The random forest algorithm then uses a voting method to combine the classifications of the ensemble of trees and determine the predicted VEI of a given volcano. We use Matlab's implementation of the random forest through the classification ensemble learner and optimize the hyperparameters of method, learning cycles and learning rates (See Github repository for full code: <https://github.com/eebrodsky/VEI-Predict.git>). Before selecting the random forest approach, we also investigated alternative machine learning classifiers such as support vector machines (SVMs), naive Bayes and nearest neighbor algorithms. All algorithms performed nearly identically on this dataset and we confine ourselves to presenting the random forest results for brevity. We make no claims about having found the optimal classifier method and future work may indeed find an even better method than utilized here.

We trained the random forest on each of the datasets and report fivefold cross-validation estimates of generalization accuracy to guard against overfitting (Hastie et al. 2017). Because the goal of this study is to compare the predictive skill with various combinations of attributes rather than to determine the optimal prediction, we trained separate models for each individual attribute and as well as some combinations of attributes. The All Attributes model contains all available attributes. The Categorical Data model uses the tectonic setting, dominant petrology, and morphology of each volcano and can be viewed as a grouping that only includes intrinsic attributes without any parameterization of eruptive history. The All Numerical Attributes model uses all of the VEI statistics used by the simple prediction models as well as repose time and eruption duration.

Covariance between attributes clearly exists. For instance, we would expect morphology and petrology to be related. Because of such covariance, the performance of the models is interpreted strictly empirically rather than physically. We report the accuracy of the trained model on the historied volcano dataset and then subsequently test the model on the unhistoried volcanoes.

The unhistoried volcanoes are not used in the training process and thus provide an out-of-sample test. For the intrinsic attributes that are available regardless of history, the unhistoried volcanoes are likely the best measure of performance. However, there is the possibility that the unhistoried volcanoes may have other biases since they are preferentially volcanoes that erupt less frequently or have poorer quality information.

Results

Figure 1a shows the cross-validated accuracy gain relative to a cohort prediction for every forecast scheme considered for all different subsamples of VEI for the most robust period of data, i.e., after 1500 AD. The results show that nearly every approach on every subset does better than the random sampling baseline. The implication is that using global trends of common behavior for similar types of volcanoes or merely assuming consistent behavior with the past is a rational approach that provides a ~10–30 percentage point accuracy gain depending on the subset of VEI considered. The clearest outlier is the maximum VEI predictor. Forecasting the next VEI to be equal to the maximum of the previously recorded VEI may be an appropriately conservative choice from a mitigation standpoint, but is unsurprisingly a poor predictor since the VEI distribution is weighted towards smaller eruptions.

For most subsets of the data, the random forest performed as well as the simple predictors. In a few cases, the random forest significantly outperformed the other metrics. For instance, the random forest performs better for the datasets limited to high VEIs. For these small datasets with the highest threshold VEI the training process seems to be extracting useful information from the attributes of the volcano beyond the information encoded in the eruptive history.

Another view of the accuracy comes from the unhistoried volcanoes that were reserved for testing (Fig. 1b) which have a maximum accuracy gain of just over 15 percentage points. These gains on the test set in Fig. 1b are similar to the cross-validation accuracy for the same models in Fig. 1a and thus provide confidence that the gains in Fig. 1a should generalize beyond the training dataset. It is also worth noting that each model in Fig. 1b has similar accuracy gains for all populations. We infer that using the intrinsic characteristics of volcanoes for VEI predictions performs generally similarly over all the VEI thresholds.

The total accuracy of VEI forecasts ranges from ~20–75% (Fig. 1c), with most approaches yielding ~40–70% accurate forecasts. There exists an increase in the total accuracy with increasing VEI, which should be expected because increasing the minimum VEI considered reduces the number of possible VEI categories in the dataset. Even a random prediction with a small number of VEI values will be more accurate than one that has more options. In addition, catalog completeness likely increases with VEI, thus improving the data quality. Potentially the most complete group is the $VEI \geq 2$ subset which has a total accuracy of ~70% for most approaches with gains of ~15 percentage points for either the simple predictions or the random forest. In general, it should be

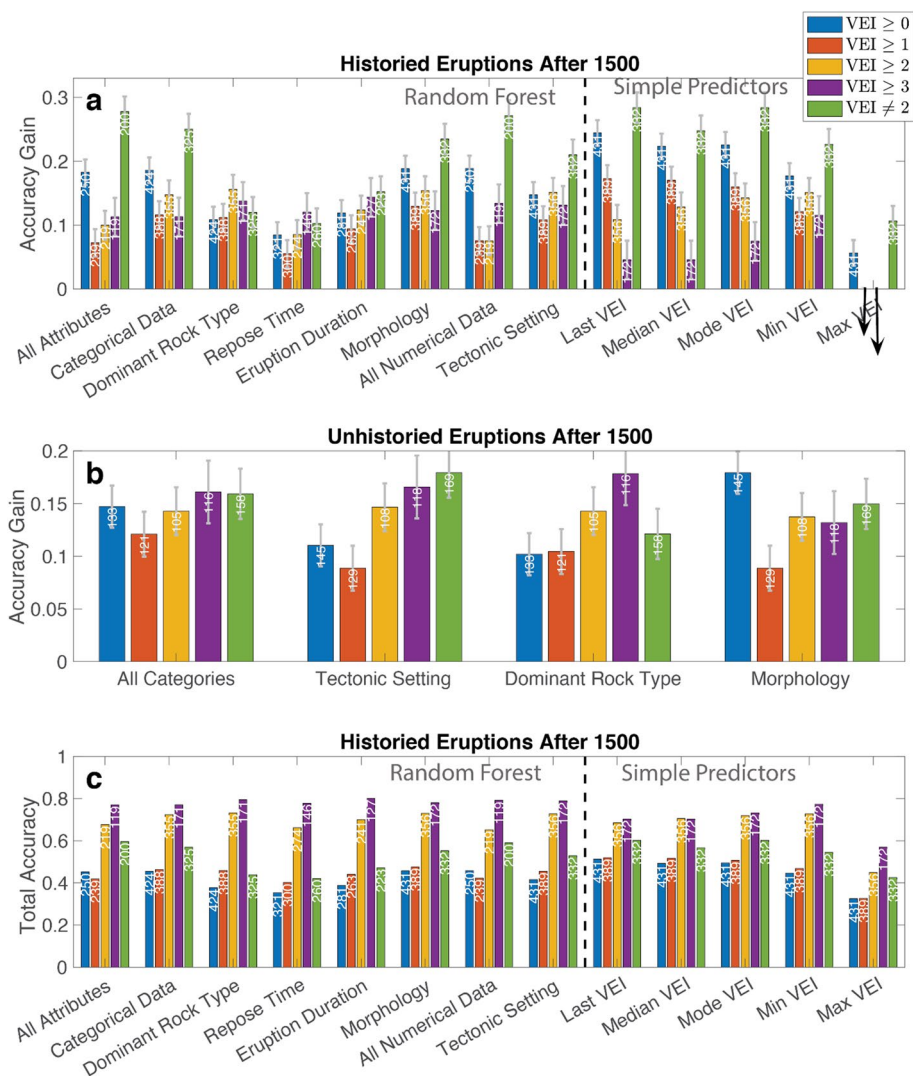


Fig. 1 Performance of forecasting models from 1500 through February 2019 for each VEI data subset (see legend). **(A)** Accuracy gain of historied volcanoes relative to cohort baseline accuracy. Forecasting accuracy is estimated based on fivefold cross-validation. Small white numbers are the total number of volcanoes (N) in each dataset. Error bars are 1 standard deviation of cohort baseline accuracy based resampling (see text), which maps directly into error on accuracy gain. Downward arrows indicate negative accuracy gains, i.e., an ineffective prediction method. **(B)** Accuracy gain relative to cohort baseline on the unhistoried volcanoes. This dataset cannot be used for attributes that demand a history of VEI, such as the simple statistical predictions. Error bars are 1 standard deviation of cohort baseline accuracy. **(C)** Total accuracy of all models as estimated by fivefold cross-validation

possible to have a long-term forecast of the VEI of the next eruption with these confidence levels, even in the absence of more detailed studies on a particular volcanic system.

The results are reinforced by probing the entire Holocene eruption catalog (Fig. 2). The highest accuracy gains of 30 percentage points were achieved for the $VEI \neq 2$ group using composite models including multiple attributes (All Attributes or All Numeric) and the simple predictor of the last VEI. In general, the random forest

accuracy gain for the entire Holocene is slightly better than for the dataset limited to post-1500. Most notably repose time, which provides very little accuracy gain for the post-1500 eruptions has a greater accuracy gain for the whole Holocene dataset.

The zero-rule usually outperforms the cohort prediction and thus the accuracy gains relative to the zero-rule are smaller (Figs. 3–4). For several of the random forest models, the algorithm recreated the zero-rule procedure as can be seen by the lack of an accuracy gain relative to



Fig. 2 Performance of forecasting models for the Holocene for each VEI data subset (see legend). **(A)** Accuracy gain of historied volcanoes relative to cohort baseline accuracy. Forecasting accuracy is estimated based on fivefold cross-validation. Small white numbers are the total number of volcanoes (N) in each dataset. Error bars are 1 standard deviation of cohort baseline accuracy based resampling (see text), which maps directly into error on accuracy gain. Downward arrows indicate negative accuracy gains, i.e., an ineffective prediction method. **(B)** Accuracy gain relative to cohort baseline on the unhistoried volcanoes. This dataset cannot be used for attributes that demand a history of VEI, such as the simple statistical predictions. Error bars are 1 standard deviation of cohort baseline accuracy. **(C)** Estimated total accuracy of all models based on fivefold cross-validation

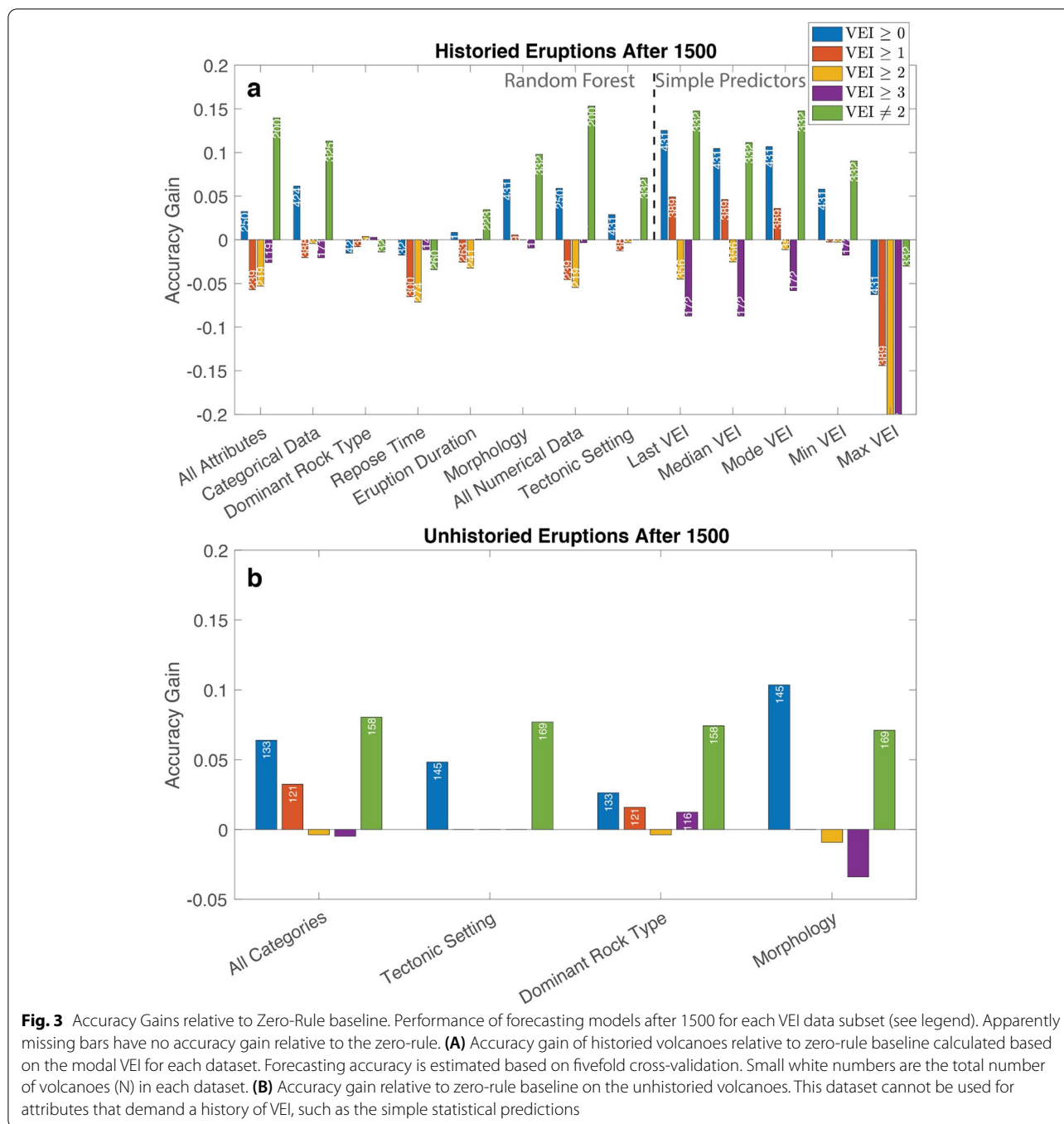


Fig. 3 Accuracy Gains relative to Zero-Rule baseline. Performance of forecasting models after 1500 for each VEI data subset (see legend). Apparently missing bars have no accuracy gain relative to the zero-rule. **(A)** Accuracy gain of historied volcanoes relative to zero-rule baseline calculated based on the modal VEI for each dataset. Forecasting accuracy is estimated based on fivefold cross-validation. Small white numbers are the total number of volcanoes (N) in each dataset. **(B)** Accuracy gain relative to zero-rule baseline on the unhistoried volcanoes. This dataset cannot be used for attributes that demand a history of VEI, such as the simple statistical predictions

the zero-rule in Figs. 3–4. The implication is that fairly good predictions can be achieved by simply predicting that the next eruption at a given volcano will be the most common VEI in the global dataset. Note that this procedure is distinct from predicting the modal VEI for a given volcano, which can perform differently (Fig. 3a, 4a). Only certain procedures on certain datasets outperform the zero-rule baseline. The composite models (All Attributes

and All Numeric) can provide accuracy gains of up to 20 percentage points on the $VEI \neq 2$ dataset. The performance on the unhistoried eruptions shows a modest predictive power of around 5 percentage points for the intrinsic attributes. However, the zero-rule is seldom outperformed for datasets limited to the largest eruptions ($VEI \geq 3$). In fact, the simple predictors do much worse than the zero-rule for these datasets and practitioners

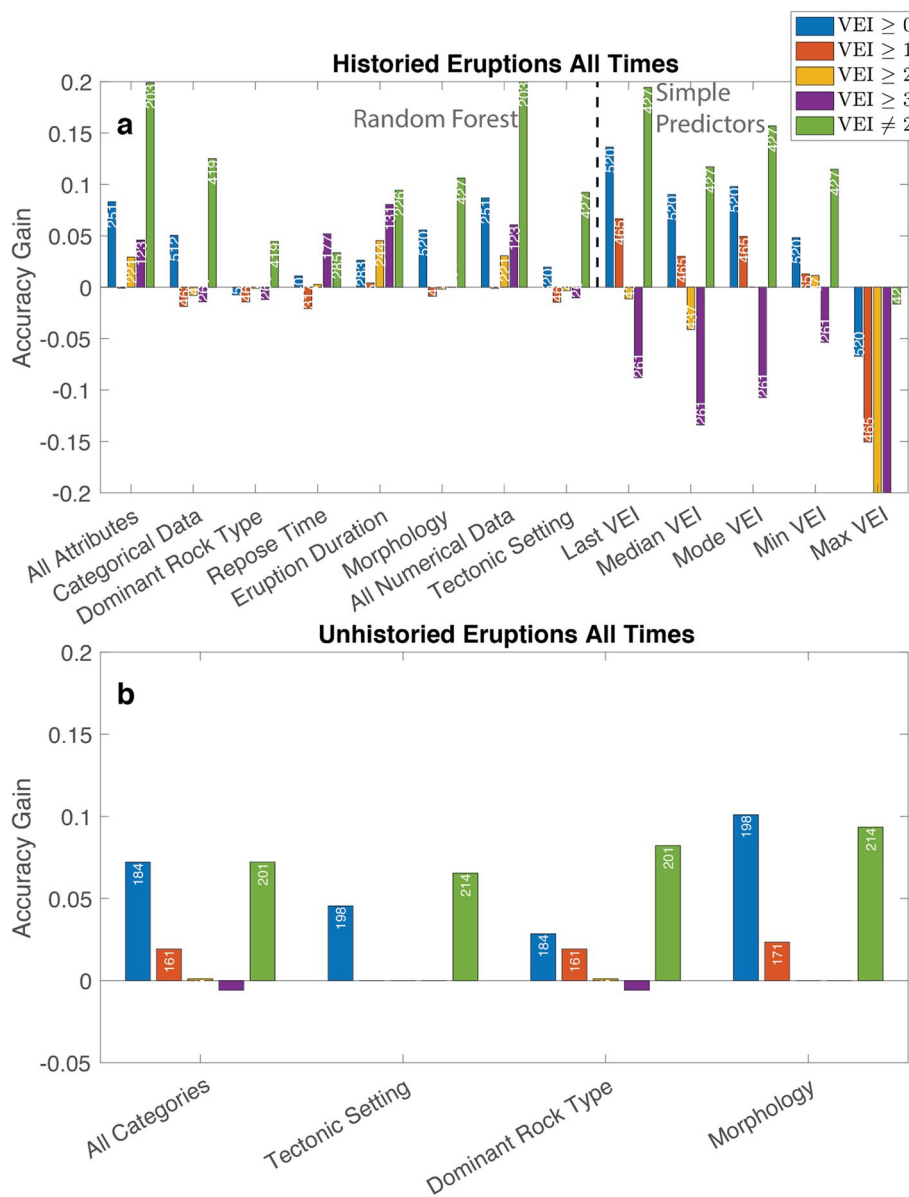
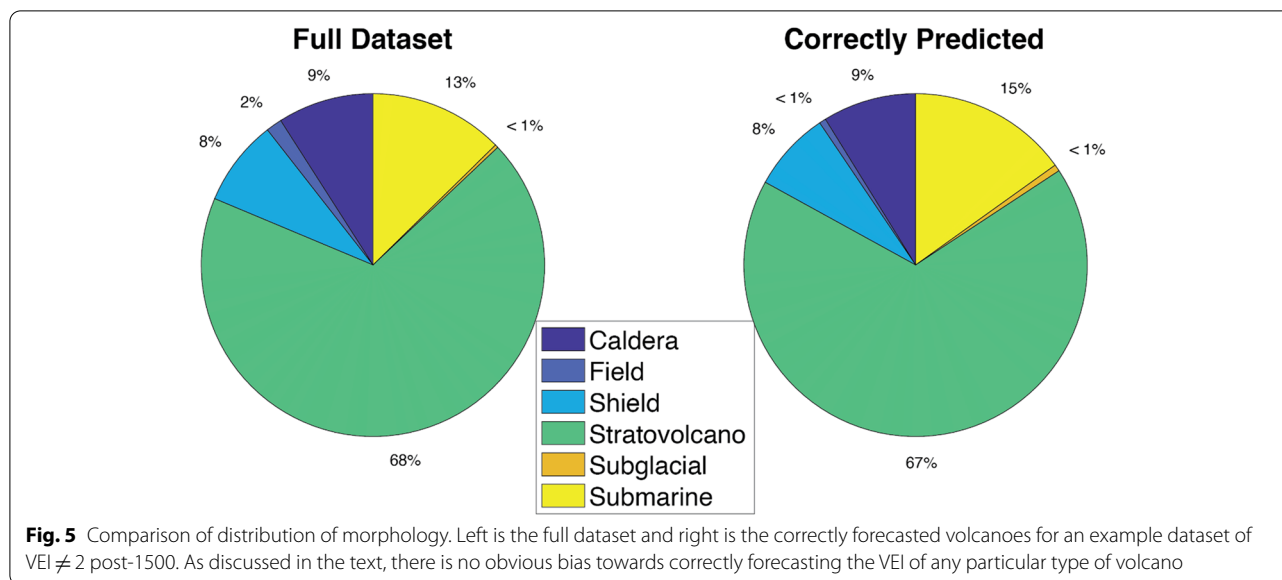


Fig. 4 Accuracy Gains relative to Zero-Rule baseline. Performance of forecasting models from the Holocene for each VEI data subset (see legend). Apparently missing bars have no accuracy gain relative to the zero-rule. **(A)** Accuracy gain of historied volcanoes relative to zero-rule baseline calculated based on the modal VEI for each dataset. Forecasting accuracy is estimated based on fivefold cross-validation. Small white numbers are the total number of volcanoes (N) in each dataset. **(B)** Accuracy gain relative to zero-rule baseline on the unhistoried volcanoes. This dataset cannot be used for attributes that demand a history of VEI, such as the simple statistical predictions

would be well-advised to simply use the global modes rather than volcano-specific information if only large eruption data is available.

Lastly, we probed to determine if any particular type of volcano was better forecast than the others by examining the distribution of intrinsic attributes in the successful forecasts compared to the full dataset. For instance, for the All Categories model, Fig. 5 shows the distribution

of morphologies for volcanoes in the training set of the VEI ≠ 2 post-1500 which were predicted correctly. Figure 5 also shows the distribution of morphologies in the full dataset. There is no obvious difference in the distribution and thus we conclude that there is no obvious preference for accurate predictions of certain volcano morphologies. Similar null results were found for tectonic setting, dominant petrology, and other attributes.



Caveats

This study is limited by design to the uniformly reported global database as available in 2019. The global database is limited in both the number of eruptions and attributes available. In the future, an increase in the collection and observation of volcanic eruptions of all VEIs is critical to establishing under what circumstances (if any) a greater level of predictability can be achieved. Machine learning algorithms require large datasets so it might be expected that this approach will become better constrained as the global database improves.

It is also worth noting that the focus of this study is on a conditional probability, i.e., we are studying the VEI of the next eruption assuming that the eruption happens. This is distinct from efforts to predict the timing of the next eruption.

Future work should address the clear skew in the distribution of volcanic eruptions of VEI 2. The $VEI \neq 2$ group was designed to eliminate eruptions that merely had a default value applied with insufficient data. The results in Figs. 1a and 2a clearly show that when VEI 2 eruptions have been removed there is an increase in accuracy gain in several models. Using VEI 2 as for under-documented explosive eruptions seems to have obscured some otherwise quantifiable trends in the data. Of course, this procedure eliminates many true VEI 2 eruptions and refinement would be helpful in future work.

There are some other differences across VEI-based subsets worth noting. The statistical attributes have the highest accuracy gain of any single attribute predictor when including low VEI eruptions, however, for the higher VEI threshold models, the random forest performs better. As discussed above, the total accuracy increases with

the limited VEI datasets simply due to the small number of target classes and thus the decrease in accuracy gain may be to some extent an effect of the rather large baseline accuracy of the null hypothesis (random prediction). However, the intrinsic properties of volcano models perform similarly for all subsets of the data and are not subject to such biases. Combining attributes gives the best performance.

Conclusions

The overarching question of this study is whether or not the size of the current eruption can be well-predicted from the global database. We conclude that the answer is yes. Nearly all models used in this study have a level of predictability of over 10 percentage point accuracy gain from the cohort baseline. The total accuracy was found to range from 30 to 80% across the VEI thresholds and the two time periods. Of the models that were used during this study, the random forest models perform as well or better than the simple prediction models. Multiple attribute models utilizing the random forest algorithm had the highest level of predictability. The All Attributes and All Numeric models have the highest accuracy gain with values of 30 percentage points above the cohort baseline and 20 percentage points above the zero-rule baseline. There is no preference in forecasting based on volcano type. These trained models and the code generating them are now available on github (<https://github.com/eebrodsky/VEI-Predict.git>) for practitioners to use as they endeavor to quantify forecasts on volcanoes that may or may not have instrumental monitoring or other indicators of their future behavior.

Abbreviations

VEI: Volcanic Explosivity Index; GVP: Global Volcanism Program.

Acknowledgements

We thank S. Ogburn for helpful conversations and the categorization of the volcanoes by morphology. Reviews by Jonathan Rougier and an anonymous reviewer significantly improved this paper. We are grateful to the Kathryn D. Sullivan award for supporting the undergraduate research at the core of this project.

Authors' contributions

PC wrote the initial codes, analyzed the data and wrote the initial draft of the manuscript. EB designed the study, guided the analysis, revised the codes and revised the manuscript. All author(s) read and approved the final manuscript.

Funding

Kathryn D. Sullivan Research Impact Award, UC Santa Cruz.

Availability of data and materials

The data in this study comes from the Global Volcanism Program online database at <http://volcano.si.edu/> accessed 13 Feb. 2019. The Matlab codes used for the study are available at: <https://github.com/eebrodsky/VEI-Predict.git>

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 2 June 2021 Accepted: 9 February 2022

Published online: 25 March 2022

References

- Bebbington MS (2014) Long-term forecasting of volcanic explosivity. *Geophys J Int* 197(3):1500–1515
- Cameron CE, Prejean SG, Coombs ML, Wallace KL, Power JA, Roman DC (2018) Alaska Volcano Observatory Alert and Forecasting Timeliness: 1989–2017. *Front Earth Sci* 6:86
- Caudron C, Chardot L, Girona T, Aoki Y, Fournier N (2020) Editorial: Towards Improved Forecasting of Volcanic Eruptions. *Front Earth Sci* 8:45
- Decker RW (1986) Forecasting Volcanic Eruptions. *Annu Rev Earth Planet Sci* 14(1):267–291
- Global Volcanism Program, 2013. *Volcanoes of the World*, v. 4.7.5 (21 Dec 2018). Venzke, E (ed.). Smithsonian Institution. Downloaded 13 Feb 2019. <https://doi.org/10.5479/si.GVPVOTW4-2013>.
- Hastie T, Tibshirani R, Friedman J (2017) *The Elements of Statistical Learning* Vol. 1, 2nd edn. Springer, New York
- Houghton BF, Swanson DA, Rausch J, Carey RJ, Fagents SA, Orr TR (2013) Pushing the Volcanic Explosivity Index to its limit and beyond: Constraints from exceptionally weak explosive eruptions at Kilauea in 2008. *Geology* 41(6):627–630
- Le Bas MJ, Maitre RWL, Streckesken A, Zabetin B (1986) A Chemical Classification of Volcanic Rocks Based on the Total Alkali-Silica Diagram. *J Petrol* 27(3):745–750
- Marzocchi W, Bebbington MS (2012) Probabilistic eruption forecasting at short and long time scales. *Bull Volcanol* 74(8):1777–1805
- Mead S, Magill C (2014) Determining change points in data completeness for the Holocene eruption record. *Bull Volcanol* 76(11):874
- Mendoza-Rosas AT, De la Cruz-Reyna S (2008) A statistical method linking geological and historical eruption time series for volcanic hazard estimations: Applications to active polygenetic volcanoes. *J Volcanol Geoth Res* 176(2):277–290
- Newhall CG, Self S (1982) The volcanic explosivity index (VEI) an estimate of explosive magnitude for historical volcanism. *J Geophys Res* 87(C2):1231–1238
- Papale P (2017) Rational volcanic hazard forecasts and the use of volcanic alert levels. *J Appl Volcanol* 6(1):1–13
- Pesicek J D, Ogburn S E, Prejean S G (2021) Indicators of volcanic eruptions revealed by global M4+ earthquakes. *J Geophysical Res: Solid Earth* 126:e2020JB021294. <https://doi.org/10.1029/2020JB021294>
- Poland MP, Anderson KR (2020) Partly cloudy with a chance of lava flows: forecasting volcanic eruptions in the twenty-first century. *J Geophysical Res: Solid Earth* 125(1):e2018JB016974
- Pyle DM (2015). Sizes of volcanic eruptions. In *The Encyclopedia of Volcanoes* Cambridge: Academic Press. (pp. 257-264).
- Sheldrake T, Caricchi L (2017) Regional variability in the frequency and magnitude of large explosive volcanic eruptions. *Geology* 45(2):111–114
- Siebert L, Simkin T, Kimberley P (2010) *Volcanoes of the World*. Univ. California Press, Berkeley
- Tierz P, Loughlin SC, Calder ES (2019) VOLCANS: an objective, structured and reproducible method for identifying sets of analogue volcanoes. *Bull Volcanol* 81(12):1–22
- Wang T, Schofield M, Bebbington M, Kiyosugi K (2020) Bayesian modelling of marked point processes with incomplete records: volcanic eruptions. *J Roy Stat Soc: Ser C (appl Stat)* 69(1):109–130
- Winson AEG, Costa F, Newhall CG, Woo G (2014) An analysis of the issuance of volcanic alert levels during volcanic crises. *J Appl Volcanol* 3(1):1–12
- Witten IH, Frank E, Hall MA, Pal CJ, Mining D (2017) *Practical Machine Learning Tools and Techniques*, 4th edn. Elsevier, Amsterdam

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

